



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-10567-TDI/942/B

**SPATIAL DYNAMIC MODELING AS PLANNING TOOL:
SIMULATION OF URBAN LAND USE CHANGE IN BAURU AND
PIRACICABA (SP), BRAZIL**

Cláudia Maria de Almeida

A Thesis of the PhD Program in Remote Sensing, supervised by Drs. Antonio Miguel Monteiro and Gilberto Câmara Neto, approved on the 3rd of October, 2003.

528.711.7

ALMEIDA, C. M.

Spatial dynamic modeling as a planning tool:
simulation of urban land use change in Bauru and Piracicaba
(SP), Brazil / C. M. Almeida. – São José dos Campos: INPE,
2003.

321p. – (INPE-10567-TDI/942/B).

1.Cellular automata. 2 Urban planning. 3.Urban
modeling. 4.Land use change.5.Weights of evidence. 5
Logistic regression. 6.Regression analysis. I.Title.

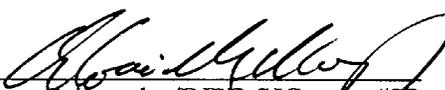
Aprovada pela Banca Examinadora em cumprimento a requisito exigido para a obtenção do Título de **Doutora em Sensoriamento Remoto.**

Dr^a Leila Maria Garcia Fonseca



Presidente/INPE-SJCampos/SP

Dr. Antônio Miguel Vieira Monteiro



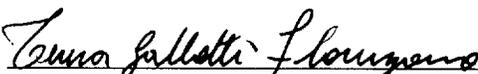
Orientador/INPE-SJCampos/SP

Dr. Gilberto Câmara Neto



Orientador/INPE-SJCampos/SP

Dr^a Teresa Gallotti Florenzano



Membro da Banca/INPE-SJCampos/SP

Dr. Britaldo Silveira Soares Filho



Membro da Banca
Convidado UFMG, Belo Horizonte/MG

Dr. Michael Batty



Membro da Banca
Convidado University College London, UK

Candidato (a): Cláudia Maria de Almeida

São José dos Campos, 03 de outubro de 2003.

*“The flame of knowledge is kept by
the continuous and tireless desire of surpassing limits
belonging to those who pursue it...”*

*“One does not discover new continents without
consenting to lose sight of the shore for a very long time.”*

ANDRE GIDE (1869-1951)

*To my parents,
Claudionor de Almeida (in memorian) and
Antonita de Almeida.*

ACKNOWLEDGEMENTS

I am grateful to all those who directly and indirectly contributed for the successful achievement of this work

The most heartfelt thanks go to my supervisors, Dr. Antonio Miguel Vieira Monteiro – head of the Division for Images Processing (DPI), and Dr. Gilberto Câmara Neto, head of the General Co-ordination for Earth Observation (OBT) of the Brazilian National Institute for Space Research (INPE). Their immeasurable support, constant incentive and guidance were crucial for the development of this thesis.

I am as well deeply thanked for the support and guidance provided to me by Prof. Michael Batty, director of the Centre for Advanced Spatial Analysis of the University College London (CASA-UCL), where part of this PhD research has been accomplished.

Special thanks are due to the São Paulo State Foundation for Research Support – FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), which sponsored my PhD program at INPE.

My sincere thanks also go to the Academic Co-ordination of the Post-Graduation Program in Remote Sensing of INPE, which granted me special funding for the digitalization of cartographic data and the acquisition of softwares and satellite images.

Acknowledgements must be given to the Brazilian National Foundation for the Undergraduates and Graduates Upgrade – CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), for sponsoring my PhD affiliate program held at CASA-UCL.

I am thanked to INPE for allowing me to attend the PhD program in Remote Sensing and for rendering available its physical infrastructure and technical support. I must as well address my honest thanks to the professors of the Postgraduate Program in Remote Sensing for all the shared knowledge and experience. I am also especially grateful to the INPE library team for their dedication and competence in helping me during the execution of this work.

My sincere and deep thanks go to Dr. Britaldo Silveira Soares-Filho, head of the Center for Remote Sensing of the Federal University of Minas Gerais (CSR-UFMG), who trained me to deal with the statistical methods and the simulation model used in this research. These thanks are extensive to the whole CSR team, especially to Gustavo, Araújo, Eliane, William and Leo.

Special words of thanks go to the DPI team, mainly to Miro, Júlio D'Alge, Mari, Maurano, Rosa, Silvia, Eymar, Diehl and Fabiano.

I also wish to thank the Water Supply and Waste Water Disposal Department (DAE) of Bauru, and the Bauru Planning Secretariat (SEPLAN) for providing the city maps.

These thanks are as well due to the Water Supply and Waste Water Disposal Department (DAE) of Piracicaba, and the Piracicaba Planning Secretariat (SEMUPLAN), for conceding the city maps of Piracicaba.

The “AJS Engenharia e Informática” team must be thanked for all their cooperative support in the digitalization of the paper plans.

I am grateful for the guidance given to me on logistic regression models by Rejane of the “Laboratory of Epidemiology and Statistics - Dante Pazzanese Cardiology Institute – São Paulo” (LEE), and by Alexandre Pantuzzo and Renato Assunção (UFMG). These thanks also go to Camila (LEE) for her consultancy on linear regression models and predictions.

The support and help given to me by the library team of the Brazilian Institute for Geography and Statistics (IBGE), the library team of Fundação Seade – São Paulo and by Dr. Mérida of the National Institute for Applied Economics (IPEA) ought to be acknowledged.

I must also truly thank the valuable help provided to me by the technical support team, especially Prof. Eliseu Weber, of IDRISI Resources Center of the Federal University of Rio Grande do Sul (CRIDRISI-UFRGS), and by Bartira Carvalho of Geosoft Latinoamérica Ltda.

Special words of thanks also go to Alfredo, for helping me to manage the S-Plus statistical package; to Missae, for teaching me how to operate the S-Carta module of SPRING; and to Flávia, for allowing me to use her computer to run my simulations.

I should also acknowledge the help, moral support and companionship of my friends at INPE, especially Rosely, Waldiza, Marcus, Nina, Bianca, Mariza, Ezequiel, Patrícia, Fred Ramos, Fred Angelis, Thalês, Viviane, Cacá, Isabela, Emerson, Norton, Fernando Ventura, Marcos, Ana Paula, Tiago, Paulina, Gilberto, Rovedy, Silvana, Bel, Karine, Lúbia, Marcelino, Taciana, Piedade, Gibotti, Giane, and Eduardo Pinheiro. Special acknowledgments go to Paulo Souza, who programmed the macro of this document.

And finally, my deepest and greatest thanks go to my parents for their constant support and encouragement throughout my academic life.

ABSTRACT

This work is committed to provide methodological guidelines for the simulation of urban land use dynamics. Modeling experiments of urban land use change were conducted for two medium-sized cities (Bauru and Piracicaba) located in the inland of São Paulo State over time series of approximately thirty-five years. Land use transitions were estimated using two different empirical probabilistic methods – the ‘weights of evidence’ approach, based on Bayes’ theory, and logistic regression. The thereof derived land use change probabilities drove a cellular automata model, built upon basis of stochastic land use allocation algorithms. Socioeconomic and infrastructural factors demonstrated to be the drivers of local land use change, whose logic is explained in light of economic theories of urban development and growth. The simulation outputs were statistically validated according to a multiple resolution fitting procedure. After the accomplishment of simulations for successive time periods along the whole time series, forecast simulations were carried out for stationary and non-stationary scenarios of transition trends. The former were assessed through the Markov chain, while the latter were obtained from linear regression models relating rates of land use change to demographic data and economic performance indicators. Both types of forecast scenarios were built for the short- and medium-term, respectively 2004 and 2007. And finally, a due attention was drawn to possible extensions of this work throughout.

**MODELAGEM DA DINÂMICA ESPACIAL COMO UMA FERRAMENTA
AUXILIAR AO PLANEJAMENTO: SIMULAÇÃO DE MUDANÇAS DE USO
DA TERRA EM ÁREAS URBANAS PARA AS CIDADES DE
BAURU E PIRACICABA (SP), BRASIL**

RESUMO

Este trabalho propõe-se a fornecer diretrizes metodológicas para a simulação de dinâmicas de uso do solo urbano. Experimentos de modelagem de mudanças de uso da terra foram realizados para duas cidades de médio-porte (Bauru e Piracicaba) localizadas no interior do Estado de São Paulo, ao longo de séries multitemporais de aproximadamente trinta e cinco anos. As transições de uso do solo foram estimadas através de dois diferentes métodos probabilísticos empíricos – ‘pesos de evidência’, baseado no teorema da probabilidade condicional de Bayes, e regressão logística. As probabilidades de mudança de uso do solo obtidas a partir daí alimentaram um modelo de autômatos celulares, construído com base em algoritmos de transição estocásticos. Aspectos sócio-econômicos e de infra-estrutura demonstraram ser variáveis forçantes de mudanças de uso do solo em nível local, cuja lógica pode ser explicada à luz das teorias econômicas de crescimento e desenvolvimento urbano. Os resultados das simulações foram validados espacialmente em função de um procedimento estatístico de ajuste por múltiplas resoluções. Após a conclusão das simulações para sucessivos ciclos de tempo ao longo das séries multitemporais, foram realizadas simulações de prognóstico de cenários estacionários e não-estacionários de tendências de transição. Os primeiros foram determinados através do modelo Markoviano, ao passo que os cenários não-estacionários foram obtidos a partir de modelos de regressão linear, relacionando taxas de transição do uso do solo a dados demográficos e indicadores de desempenho econômico. Ambos os tipos de cenários de prognóstico foram conjecturados para o curto e médio prazo, respectivamente 2004 e 2007. E por fim, devida atenção foi dispensada a possíveis aplicações e extensões deste trabalho.

CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

CHAPTER 1.....	41
INTRODUCTION	41
1.1 Overview of the Research Topic	41
1.2 General Research Goals.....	47
1.3 Specific Research Goals	48
1.4 Research Hypotheses.....	49
1.5 Detailed Outline of the Thesis	50
CHAPTER 2.....	53
THEORETICAL FOUNDATIONS ON URBAN MODELS	53
2.1 The Modeling Object.....	53
2.1.1 Definitions of Urban Settlements	53
2.1.2 The Modeling Object in the Present Work.....	56
2.2 A Brief Historical Perspective on Models of Urban Land Use Change.....	57
2.2.1 Introduction	57
2.2.2 Non-Dynamic Models of Urban Land Use Change	60
2.2.3 Early Dynamic Models of Urban Land Use Change.....	66
2.3 Spatial Dynamic Models of Urban Land Use Change	76
2.3.1 Space and Time in the Present Work.....	76
2.3.2 Cellular Automata and the Advent of Spatially Explicit Models of Land Use Change.....	77
2.3.2.1 Cellular Automata: Definition and Properties	77
2.3.2.2 Spatially Explicit or Spatial Dynamic Models of Urban Land Use Change	80
2.3.2.3 Theory-Driven x Data-Driven CA Models of Urban Land Use Change.....	85
2.3.3 Why Using Cellular Automata for Urban Modeling.....	88
2.4 Conclusions	89
CHAPTER 3.....	91
STUDY AREA	91
3.1 Location.....	91
3.2 Historical Background of the Urbanization Process.....	92
3.2.1 The Mining Cycle (1580 – 1730)	92
3.2.2 The Sugar Cane Cycle (1730 – 1822)	93
3.2.3 The Coffee Cycle (1822 – 1920).....	93
3.2.4 The Industrialization Process and Economic Dynamics of the XX Century	95

3.2.4.1 Years 1920 – 1929	95
3.2.4.2 Years 1930 – 1955	95
3.2.4.3 Years 1956 – 1969	96
3.2.4.4 Years 1970 – 1979	97
3.2.4.5 Years 1980 – 1989	98
3.2.4.6 Years 1990 – 1999	98
3.3 Spatial Configuration of the Urban Network in the Macroarea of São Paulo State Inland	99
3.4 Selected Cities for Analysis: Geographical Characteristics and Socioeconomic Peculiarities	103
3.4.1 Bauru	103
3.4.1.1 Geographical Characteristics	103
3.4.1.2 Socioeconomic Peculiarities	103
3.4.2 Piracicaba	107
3.4.2.1 Geographical Characteristics	107
3.4.2.2 Socioeconomic Peculiarities	107
3.5 Conclusions	109
CHAPTER 4.....	111
BUILDING THE GEOGRAPHIC DATABASE	111
4.1 Spatial Input Data: Cartographic, Aerophotogrammetric and Remote Sensing Data	111
4.1.1 Cartographic Data	113
4.1.2 Digital Aerial Photos	114
4.1.3 Remote Sensing Data	115
4.2 Non-Spatial Input Data: Socioeconomic Data	117
4.2.1 Demographic Data	118
4.2.2 Economic Data	118
4.3 Data Pre-Processing Operations	122
4.3.1 Digital Conversion of Cartographic Data	122
4.3.2 Georeferencing Techniques	122
4.3.2.1 Image to Map Georeferencing	122
4.3.2.2 Image to Image Registration	125
4.3.2.3 Image to Vector Data Registration	126
4.3.3 Spatial Data Analysis and Processing	126
4.3.4 Updating the Land Use Maps through Remote Sensing Data	128
4.3.5 Generalization Procedures Applied to the Land Use Maps	130
4.4 Conclusions	131
CHAPTER 5.....	133
METHODS FOR THE URBAN LAND USE DYNAMICS SIMULATION MODEL	133
5.1 Introduction	133
5.2 Simulation Methods	133
5.2.1 The Weights of Evidence Method	133
5.2.1.1 Introduction to the Weights of Evidence Method	133

5.2.1.2 Exploratory Analysis and Selection of Variables.....	140
5.2.1.3 Estimation of Global Transition Probabilities.....	144
5.2.1.4 Calculation of Local Transition Probabilities.....	145
5.2.1.5 Model Calibration.....	145
5.2.1.6 Statistical Validation Test.....	148
5.2.2 The Logistic Regression Method.....	150
5.2.2.1 Introduction to the Logistic Regression Method.....	150
5.2.2.2 Exploratory Analysis and Selection of Variables.....	157
5.2.2.3 Estimation of Global Transition Probabilities.....	159
5.2.2.4 Calculation of Local Transition Probabilities.....	159
5.2.2.5 Model Calibration.....	160
5.2.2.6 Statistical Validation Test.....	163
5.3 Forecast Methods.....	164
5.3.1 The Markov Chain and the Future Transition Probabilities Matrix.....	164
5.3.2 Linear Regression Models for the Parameterization of Future Land Use Transition Probabilities.....	168
5.3.2.1 Introduction to Linear Regression Models.....	168
5.3.2.2 Exploratory Analysis.....	169
5.3.2.3 Least Squares Estimators.....	171
5.3.2.4 Analysis of Variance.....	172
5.3.2.5 Analysis of Residuals.....	175
5.3.2.6 Conversion of Predicted Destination Areas into Transition Probabilities.....	176
5.3.3 Conceived Scenarios and Time Horizons for Land Use Change Forecasts.....	179
5.4 The Urban Land Use Dynamics Simulation Model.....	180
5.4.1 DINAMICA General Data Model.....	180
5.4.2 Software Structure and Input Parameters.....	182
5.4.3 Transition Algorithms.....	183
5.4.3.1 The <i>Expander</i> Function.....	184
5.4.3.2 The <i>Patcher</i> Function.....	184
5.5 Methodological Summary Flowchart.....	185
5.6 Conclusions.....	185
CHAPTER 6.....	189
RESULTS AND DISCUSSIONS.....	189
6.1 Bauru.....	189
6.1.1 Simulation Period: 1967 - 1979.....	189
6.1.2 Simulation Period: 1979 - 1988.....	203
6.1.2.1 Weights of Evidence Method.....	203
6.1.2.2 Logistic Regression Method.....	214
6.1.3 Simulation Period: 1988 - 2000.....	220
6.1.4 Yearly Simulations: 1967 - 2000.....	231
6.1.5 Short-Term Forecasts: 2000 - 2004.....	236
6.1.5.1 Stationary Forecasts.....	236
6.1.5.2 Non-Stationary Forecasts.....	237
6.1.5.2.1 "Non-Urban to Residential Use (nu_res)" Linear Regression Model.....	238
6.1.5.2.2 "Non-Urban to Industrial Use (nu_ind)" Linear Regression Model.....	241

6.1.5.2.3 "Non-Urban/Residential to Services Use (nu/res_serv)" Linear Regression Model.....	244
6.1.5.3 Forecasts Simulations Outputs	248
6.1.6 Medium-Term Forecasts: 2000 - 2007	250
6.1.6.1 Stationary Forecasts.....	250
6.1.6.2 Non-Stationary Forecasts	250
6.1.6.3 Forecasts Simulations Outputs	251
6.2 Piracicaba	253
6.2.1 Simulation Period: 1962 - 1985.....	253
6.2.2 Simulation Period: 1985 - 1999.....	266
6.2.3 Yearly Simulations: 1962 - 1999.....	277
6.2.4 Short-Term Forecasts: 2000 - 2004.....	283
6.2.4.1 Stationary Forecasts.....	283
6.2.4.2 Non-Stationary Forecasts	283
6.2.4.2.1 "Non-Urban to Residential Use (nu_res)" Linear Regression Model	284
6.2.4.2.2 "Non-Urban to Industrial Use (nu_ind)" Linear Regression Model.....	287
6.2.4.3 Forecasts Simulations Outputs	291
6.2.5 Medium-Term Forecasts: 2000 - 2007	293
6.2.5.1 Stationary Forecasts.....	293
6.2.5.2 Non-Stationary Forecasts	293
6.2.5.3 Forecasts Simulations Outputs	294
6.3 Conclusions	296
CHAPTER 7.....	301
FINAL REMARKS AND CONCLUSIONS	301
7.1 Dealing with Spatial Dynamic Models.....	301
7.2 Recent Advances and Potential New Paths for Models of Urban Land Use Change.....	302
7.3 Main Contributions of this Research.....	304
7.4 Possible Applicability of this Work	305
BIBLIOGRAPHIC REFERENCES	307

LIST OF FIGURES

2.1 – Weber’s classical triangle for industrial location.	61
2.2 – Hexagonal system of Christaller’s central place theory.	62
2.3 – Hexagonal nests of Lösch’s economic regions theory	62
2.4 – Generalized flow chart of the Lowry model.....	65
2.5 – Dynamic behavior curves in fixed resource systems.....	70
2.6 – Structure of activities and land uses in the model for Reading, UK	72
2.7 – The Game of Life ‘glider’ configuration - an example of emergence.....	79
2.8 – Common generalizations of cellular automata (CA).....	80
2.9 – A CA model integrating spatial micro and macroscales	83
2.10 – A CA model incorporating macroscale variables.....	84
3.1 – Location of the case study cities: Bauru and Piracicaba, inland of SP.....	91
3.2 – Current transport matrix in São Paulo State and urban network throughout the Tietê river watershed	94
3.3 – Piracicaba skyline with high-rise buildings in the background.....	102
3.4 – Bauru aerial view with high-rise buildings.....	105
3.5 – Evolution of Bauru urban area in the latest century	106
3.6 – Sectorial GDPs for the city of Bauru in the latest four decades.....	106
3.7 – View of Piracicaba city and its green areas.....	108
3.8 – Sectorial GDPs for the city of Piracicaba in the latest four decades	109
4.1 – Map of Piracicaba in 1999.....	113
4.2 – Mosaic of digital aerial photos for the city of Piracicaba - 1962	114
4.3 – TM – 5 image of Piracicaba, 1B_4R_7G, 222/76 – 16/07/99.....	117
4.4 – Reconstitution of Piracicaba city map in 1962.....	123
4.5 – Georeferenced image of Bauru with control points.....	124
4.6 – Coupling between the Piracicaba images of 1985 and 1999	125
4.7 – Image to vector data registration for the Bauru city plan of 2000.....	127
4.8 – Sequence of operations for generating a map of distance to commercial activities clusters	128
4.9a – Western sector of Bauru city map superimposed on the TM – 5 image.....	129
4.9b – Eastern sector of Bauru city map superimposed on the TM – 5 image.....	129
5.1 – Diagram to illustrate the weights of evidence method	134
5.2 – Illustrative image showing the visual analysis on the identification of factors determining the transition ‘residential to mixed use’ in Bauru: 1979-1988	143
5.3 – Example of visual comparative analysis for the model empirical calibration in relation to the transition “residential-services (res_serv)” in Bauru from 1979 to 1988	146
5.4 – Examples of scatter plots and respective trendlines for the relations between subcategories of evidences (<i>X axis</i>) and their corresponding positive weights of evidence (<i>Y axis</i>), considering different types of land use change.....	147
5.5 – Example of the multiple resolution method for a scene of size 10 x 10 pixels and with four classes	149

5.6	– Schematic plots showing the presence/absence of coronary heart disease (CHD) in relation to the individuals age (to the left), and the incidence frequency of CHD versus the midpoint of age intervals of ten years.....	151
5.7	– Illustrative plot of the <i>logits</i> under three different models showing the presence and absence of interaction	158
5.8	– Illustrative plots of statistical relations without and with a fitted linear regression model.....	168
5.9	– Illustrative plot of the autocorrelation function for the response variable: destination area - industrial use (destarea) in Bauru for the years 1970, 1975, 1980, 1985, 1990, 1995 and 2000	171
5.10	– Illustrative plot of standardized residuals versus an independent variable (industrial added value - `indav´) of a linear regression model for the industrial area in the city of Piracicaba (1970-2000)	176
5.11	– DINAMICA generic data model for urban applications	181
5.12	– Summary methodological flowchart	186
6.1	– Bauru official city map in 1967	190
6.2	– Bauru official city map in 1979	190
6.3	– Generalized land use map in Bauru in 1967 (left) and in 1979 (right).....	191
6.4	– Cross-tabulation map between Bauru land use maps of 1967 and 1979	192
6.5	– Independent variables used to explain the land use transitions in Bauru during the simulation period 1967-1979	193
6.6	– Example of a land use transition map (non-urban to leisure/recreation) for Bauru in the period: 1967-1979	196
6.7a	– Estimated transition probability surfaces and land use change for Bauru: 1967-1979	198
6.7b	– Estimated transition probability surfaces and land use change for Bauru: 1967-1979 (continued)	199
6.8	– The best simulations compared to the actual land use in Bauru in 1979.....	201
6.9	– Bauru official city map in 1979	204
6.10	– Bauru TM – 5 image and official city map in 1988	204
6.11	– Generalized land use map in Bauru in 1979 (left) and in 1988 (right).....	205
6.12	– Cross-tabulation map between Bauru land use maps of 1979 and 1988	206
6.13	– Independent variables used to explain the land use transitions in Bauru during the simulation period 1979-1988	207
6.14a	– Estimated transition probability surfaces and land use change for Bauru: 1979-1988, from the weights of evidence method	211
6.14b	– Estimated transition probability surfaces and land use change for Bauru: 1979-1988, from the weights of evidence method (continued).....	212
6.15	– The best simulations produced by the weights of evidence method compared to the actual land use in Bauru in 1988.....	213
6.16	– Boxplot of the transition `residential to mixed use´ versus social housing (left) and boxplot of the transition `non-urban to industrial use´ versus distances to industrial zones (right) - Bauru: 1979-1988	214
6.17	– Conversion of numerical raster grids into statistical database in MINITAB 13.0	216

6.18a– Estimated transition probability surfaces and land use change for Bauru: 1979-1988, from the logistic regression method.....	218
6.18b– Estimated transition probability surfaces and land use change for Bauru: 1979-1988, from the logistic regression method (continued).....	219
6.19 – The best simulations produced by the logistic regression method compared to the actual land use in Bauru in 1988.....	220
6.20 – Bauru TM – 5 image and official city map in 1988	221
6.21 – Bauru TM – 5 image and official city map in 2000	221
6.22 – Generalized land use map in Bauru in 1988 (left) and in 2000 (right).....	222
6.23 – Cross-tabulation map between Bauru land use maps of 1988 and 2000	223
6.24 – Independent variables used to explain the land use transitions in Bauru during the simulation period 1988-2000	224
6.25a– Estimated transition probability surfaces and land use change for Bauru: 1988-2000	228
6.25b– Estimated transition probability surfaces and land use change for Bauru: 1988-2000 (continued)	229
6.26 – The best simulations compared to the actual land use in Bauru in 2000.....	230
6.27a– Yearly simulation outputs for Bauru: 1967-1974	232
6.27b– Yearly simulation outputs for Bauru: 1975-1982.....	233
6.27c– Yearly simulation outputs for Bauru: 1983-1990	234
6.27d– Yearly simulation outputs for Bauru: 1991-1998.....	235
6.27e– Yearly simulation outputs for Bauru: 1999-2000	236
6.28 – ACF and partial ACF tests for the residential use area (destarea).....	239
6.29 – Correlation matrix scatter plots for the "nu_res" model: Bauru, 2000-2004.....	240
6.30 – Analysis of residuals for the "nu_res" model: Bauru, 2000-2004.....	241
6.31 – ACF and partial ACF tests for the industrial use area (destarea)	242
6.32 – Correlation matrix scatter plots for the "nu_ind" model: Bauru, 2000-2004	243
6.33 – Analysis of residuals for the "nu_ind" model: Bauru, 2000-2004	244
6.34 – ACF and partial ACF tests for the services use area (destarea)	245
6.35 – Correlation matrix scatter plots for the "nu/res_serv" model: Bauru, 2000-2004	246
6.36 – Analysis of residuals for the "nu/res_serv" model: Bauru, 2000-2004	247
6.37 – Stationary, optimistic and pessimist simulations for 2004 compared to the actual land use in Bauru in 2000	249
6.38 – Stationary, optimistic and pessimist simulations for 2007 compared to the actual land use in Bauru in 2000	252
6.39 – Piracicaba official city map in 1962 (reconstitution map)	254
6.40 – Piracicaba TM – 5 image and official city map in 1985	254
6.41 – Generalized land use map in Bauru in 1962 (left) and in 1985 (right).....	255
6.42 – Cross-tabulation map between Bauru land use maps of 1962 and 1985	256
6.43 – Independent variables used to explain the land use transitions in Piracicaba during the simulation period 1962-1985	258
6.44a– Estimated transition probability surfaces and land use change for Piracicaba: 1962-1985	263

6.44b– Estimated transition probability surfaces and land use change for Piracicaba: 1962-1985 (continued)	264
6.44c– Estimated transition probability surfaces and land use change for Piracicaba: 1962-1985 (continued)	265
6.45 – The best simulations compared to the actual land use in Piracicaba in 1985...	266
6.46 – Piracicaba TM – 5 image and official city map in 1985	267
6.47 – Piracicaba TM– 5 image and official city map in 1999	267
6.48 – Generalized land use map in Piracicaba in 1985 (left) and in 1999 (right).....	268
6.49 – Cross-tabulation map between Piracicaba land use maps of 1985 and 1999 ...	269
6.50 – Independent variables used to explain the land use transitions in Piracicaba during the simulation period 1985-1999	270
6.51 – Estimated transition probability surfaces and land use change for Piracicaba: 1985-1999	274
6.52 – The best simulations compared to the actual land use in Piracicaba in 1999...	276
6.53a– Yearly simulation outputs for Piracicaba: 1962-1969	278
6.53b– Yearly simulation outputs for Piracicaba: 1970-1977	279
6.53c– Yearly simulation outputs for Piracicaba: 1978-1985	280
6.53d– Yearly simulation outputs for Piracicaba: 1986-1993	281
6.53e– Yearly simulation outputs for Piracicaba: 1994-1999	282
6.54 – ACF and partial ACF tests for the residential use area (destarea).....	285
6.55 – Correlation matrix scatter plots for the "nu_res" model: Piracicaba, 2000-2004	286
6.56 – Analysis of residuals for the "nu_res" model: Piracicaba, 2000-2004	287
6.57 – ACF and partial ACF tests for the industrial use area (destarea)	288
6.58 – Correlation matrix scatter plots for the "nu_ind" model: Piracicaba, 2000-2004	289
6.59 – Analysis of residuals for the "nu_ind" model: Piracicaba, 2000-2004.....	290
6.60 – Stationary, optimistic and pessimist simulations for 2004 compared to the actual land use in Piracicaba in 1999	292
6.61 – Stationary, optimistic and pessimist simulations for 2007 compared to the actual land use in Piracicaba in 1999	295

LIST OF TABLES

1.1 – Trends and projections in urban population by region, 1950-2010.....	42
3.1 – Comparative increment evolution of urban areas between São Paulo metropolis and main cities of the State inland – 1974-1989.....	101
4.1 – Urban population, total and sectorial GDPs of Bauru – 1970-2000.....	121
4.2 – Urban population, total and sectorial GDPs of Piracicaba – 1970-2000.....	121
4.3 – Control points coordinates in the georeferencing of Bauru image.....	124
4.4 – Control points coordinates in the georeferencing of Piracicaba image.....	124
6.1 – Existent land use transitions in Bauru: 1967-1979.....	192
6.2 – Matrix of global transition probabilities for Bauru: 1967-1979.....	192
6.3 – Independent variables defining land use change in Bauru: 1967-1979.....	194
6.4 – Selection of variables determining land use change in Bauru: 1967-1979.....	194
6.5 – Associations between independent variables - Bauru: 1967-1979.....	195
6.6 – Values of W^+ for the selected independent variables - Bauru: 1967-1979.....	197
6.7 – DINAMICA internal parameters for the simulation of urban land use change in Bauru: 1967-1979.....	200
6.8 – Goodness-of-fit tests for the best land use change simulations of Bauru: 1967-1979.....	202
6.9 – Existent land use transitions in Bauru: 1979-1988.....	206
6.10 – Matrix of global transition probabilities for Bauru: 1979-1988.....	206
6.11 – Independent variables defining land use change in Bauru: 1979-1988.....	208
6.12 – Selection of variables determining land use change in Bauru: 1979-1988.....	208
6.13 – Associations between independent variables - Bauru: 1979-1988.....	209
6.14 – Values of W^+ for the selected independent variables - Bauru: 1979-1988.....	209
6.15 – DINAMICA internal parameters for the simulation of urban land use change in Bauru: 1979-1988.....	212
6.16 – Goodness-of-fit tests for the best land use change simulations produced by the weights of evidence method for Bauru: 1979-1988.....	213
6.17 – Correlations between independent variables - Bauru: 1979-1988.....	215
6.18 – Results of the logistic regression analyses for Bauru: 1979-1988.....	217
6.19 – Goodness-of-fit tests for the best land use change simulations produced by the logistic regression method for Bauru: 1979-1988.....	219
6.20 – Existent land use transitions in Bauru: 1988-2000.....	223
6.21 – Matrix of global transition probabilities for Bauru: 1988-2000.....	223
6.22 – Independent variables defining land use change in Bauru: 1988-2000.....	225
6.23 – Selection of variables determining land use change in Bauru: 1988-2000.....	225
6.24 – Associations between independent variables - Bauru: 1988-2000.....	226
6.25 – Values of W^+ for the selected independent variables - Bauru: 1988-2000.....	226
6.26 – DINAMICA internal parameters for the simulation of urban land use change in Bauru: 1988-2000.....	229
6.27 – Goodness-of-fit tests for the best land use change simulations of Bauru: 1988-2000.....	230
6.28 – Matrix of yearly transition probabilities for Bauru: 1967-1979.....	231

6.29 – Matrix of yearly transition probabilities for Bauru: 1979–1988	231
6.30 – Matrix of yearly transition probabilities for Bauru: 1988–2000	231
6.31 – Matrix of stationary transition probabilities for Bauru: 2000–2004	237
6.32 – Areas of destination land uses, urban population, total and sectorial GDPs (US\$): Bauru - 1970-2000	238
6.33 – Correlation matrix for the “nu_res” model: Bauru, 2000–2004.....	240
6.34 – Analysis of variance for the “nu_res” model: Bauru, 2000–2004.....	240
6.35 – Correlation matrix for the “nu_ind” model: Bauru, 2000–2004	243
6.36 – Analysis of variance for the “nu_ind” model: Bauru, 2000–2004	243
6.37 – Correlation matrix for the “nu/res_serv” model: Bauru, 2000–2004.....	246
6.38 – Analysis of variance for the “nu/res_serv” model: Bauru, 2000–2004.....	246
6.39 – Matrix of optimistic transition probabilities for Bauru: 2000–2004	248
6.40 – Matrix of pessimist transition probabilities for Bauru: 2000–2004	248
6.41 – Matrix of stationary transition probabilities for Bauru: 2000–2007	250
6.42 – Matrix of optimistic transition probabilities for Bauru: 2000–2007	251
6.43 – Matrix of pessimist transition probabilities for Bauru: 2000–2007	251
6.44 – Existent land use transitions in Piracicaba: 1962-1985.....	256
6.45 – Matrix of global transition probabilities for Piracicaba: 1962-1985	257
6.46 – Independent variables defining land use change in Piracicaba: 1962-1985.....	257
6.47 – Selection of variables defining land use change in Piracicaba: 1962-1985	259
6.48 – Associations between independent variables-Piracicaba: 1962-1985.....	259
6.49 – Values of W^+ for the selected independent variables-Piracicaba: 1962–1985 ..	260
6.50 – DINAMICA internal parameters for the simulation of urban land use change in Piracicaba: 1962–1985	265
6.51 – Goodness-of-fit tests for the best land use change simulations of Piracicaba: 1962–1985	266
6.52 – Existent land use transitions in Piracicaba: 1985-1999.....	269
6.53 – Matrix of global transition probabilities for Piracicaba: 1985-1999	269
6.54 – Independent variables defining land use change in Piracicaba: 1985-1999.....	271
6.55 – Selection of variables defining land use change in Piracicaba: 1985-1999	271
6.56 – Associations between independent variables-Piracicaba: 1985-1999	272
6.57 – Values of W^+ for the selected independent variables-Piracicaba: 1985–1999 ..	272
6.58 – DINAMICA internal parameters for the simulation of urban land use change in Piracicaba: 1985–1999	275
6.59 – Goodness-of-fit tests for the best land use change simulations of Piracicaba: 1985–1999	275
6.60 – Matrix of yearly transition probabilities for Piracicaba: 1962–1985	277
6.61 – Matrix of yearly transition probabilities for Piracicaba: 1985–1999	277
6.62 – Matrix of stationary transition probabilities for Piracicaba: 2000–2004.....	283
6.63 – Areas of destination land uses, urban population, total and sectorial GDPs (US\$): Piracicaba - 1970-2000	284
6.64 – Correlation matrix for the “nu_res” model: Piracicaba, 2000–2004.....	286
6.65 – Analysis of variance for the “nu_res” model: Piracicaba, 2000–2004.....	286
6.66 – Correlation matrix for the “nu_ind” model: Piracicaba, 2000–2004.....	289
6.67 – Analysis of variance for the “nu_ind” model: Piracicaba, 2000–2004	289
6.68 – Matrix of optimistic transition probabilities for Piracicaba: 2000–2004	291

6.69 – Matrix of pessimist transition probabilities for Piracicaba: 2000–2004.....	291
6.70 – Matrix of stationary transition probabilities for Piracicaba: 2000–2007.....	293
6.71 – Matrix of optimistic transition probabilities for Piracicaba: 2000–2007	294
6.72 – Matrix of pessimist transition probabilities for Piracicaba: 2000–2007.....	294

LIST OF EQUATIONS

- 2.1 $D_j = A \exp(-\beta d_j)$ 63
- 2.2 $P(t+1) = a_1 + b_1 E(t-1)$ 67
- 2.3 $S(t+1) = a_2 + b_2 P(t)$ 67
- 2.4 $E^b(t+1) = a_3 + b_3 X(t+1)$ 67
- 2.5 $E(t+1) = E^b(t+1) + E^l(t+1) + S(t+1)$ 67
- 2.6 $W_j^\circ = \varepsilon \sum_i \frac{W_j^a \exp(-\beta_i c_{ij})}{\sum_j W_j^a \exp(-\beta_i c_{ij})} e_i P_i - k W_j$ 73
- 2.7 $W_j^\circ = \varepsilon \sum_i \frac{[A_j (k + c_{ij}) - \beta_i]^a}{\sum_j [A_j (k + c_{ij}) - \beta_i]^a} e_i P_i - (k + c_{ij}) W_j$ 74
- 4.1 $\frac{GDP\ 1996\ (US\$ \text{ of } 1998)}{GDP\ 2000\ (US\$ \text{ of } 1998)} = \frac{GDP\ 1996\ (R\$ \text{ of } 2001)}{GDP\ 2000\ (R\$ \text{ of } 2001)}$ 119
- 4.2 $\frac{GDP\ 1996\ (US\$ \text{ of } 1998)}{GDP\ 1995\ (US\$ \text{ of } 1998)} = \frac{GDP\ 1996\ (R\$ \text{ of } 2001)}{GDP\ 1995\ (R\$ \text{ of } 2001)}$ 119
- 4.3 $AV\ (R\$ \text{ of } 2000) = AV\ (R\$ \text{ of } 2001) \times \frac{IPCA\ (2000)}{IPCA\ (2001)}$ 120
- 4.4 $Sectorial\ GDP_k\ (US\$ \text{ of } 1998) = Total\ GDP\ (US\$ \text{ of } 1998) \times \frac{Sectorial\ GDP/AV_k\ (R\$ \text{ 2000})}{Total\ GDP/AV\ (R\$ \text{ 2000})}$ 120
- 4.5 $Sectorial\ GDP_k\ (US\$ \text{ of } 1998) = Total\ GDP\ (US\$ \text{ of } 1998) \times Proportional\ Rate_k^{1990}$ 120
- 4.6 $Proportional\ Rate_k^{1990} = \frac{Proportional\ Rate_k\ (1989 + 1991)}{2}$ 120

- 4.7 $Proportional\ Rate_k^{1991} = \frac{Proportional\ Rate_k^{(1989 + 1993)}}{2}$ 120
- 4.8 $Proportional\ Rate_k^{1989} = \frac{Proportional\ Rate_k^{(1985 + 1993)}}{2}$ 121
- 4.9 $Proportional\ Rate_k^{1985} = \frac{Sectorial\ GDP_k^{1985}}{Total\ GDP^{1985}}$ 121
- 4.10 $Proportional\ Rate_k^{1993} = \frac{Sectorial\ AV_k^{1993}}{Total\ AV^{1993}}$ 121
- 5.1 $P\{R/S\} = \frac{P\{R \cap S\}}{P\{S\}}$ 134
- 5.2 $P\{R/S\} = \frac{N\{R \cap S\}}{N\{S\}}$ 134
- 5.3 $P\{S/R\} = \frac{P\{S \cap R\}}{P\{R\}}$ 135
- 5.4 $P\{R/S\} = \frac{P\{R\} \cdot P\{S/R\}}{P\{S\}}$ 135
- 5.5 $P\{R/\bar{S}\} = \frac{P\{R\} \cdot P\{\bar{S}/R\}}{P\{\bar{S}\}}$ 135
- 5.6 $\frac{P\{R/S\}}{P\{\bar{R}/S\}} = \frac{P\{R\} \cdot P\{S/R\}}{P\{\bar{R}/S\} \cdot P\{S\}}$ 135
- 5.7 $P\{\bar{R}/S\} = \frac{P\{\bar{R} \cap S\}}{P\{S\}} = \frac{P\{S/\bar{R}\} \cdot P\{\bar{R}\}}{P\{S\}}$ 136

$$5.8 \quad \frac{P \{R/S\}}{P \{\bar{R}/S\}} = \frac{P \{R\}}{P \{\bar{R}\}} \cdot \frac{P \{S\}}{P \{S\}} \cdot \frac{P \{S/R\}}{P \{S/\bar{R}\}} \quad 136$$

$$5.9 \quad O \{R/S\} = O \{R\} \cdot \frac{P \{S/R\}}{P \{S/\bar{R}\}} \quad 136$$

$$5.10 \quad \text{logit} \{R/S\} = \text{logit} \{R\} + W^+ \quad 136$$

$$5.11 \quad O \{R/\bar{S}\} = O \{R\} \cdot \frac{P \{\bar{S}/R\}}{P \{\bar{S}/\bar{R}\}} \quad 136$$

$$5.12 \quad \text{logit} \{R/\bar{S}\} = \text{logit} \{R\} + W^- \quad 137$$

$$5.13 \quad P \{R/S_1 \cap S_2\} = \frac{P \{R \cap S_1 \cap S_2\}}{P \{S_1 \cap S_2\}} \quad 137$$

$$5.14 \quad P \{R/S_1 \cap S_2\} = \frac{P \{S_1 \cap S_2/R\} \cdot P \{R\}}{P \{S_1 \cap S_2\}} \quad 138$$

$$= \frac{P \{S_1 \cap S_2/R\} \cdot P \{R\}}{P \{S_1 \cap S_2/R\} \cdot P \{R\} + P \{S_1 \cap S_2/\bar{R}\} \cdot P \{\bar{R}\}} \quad 138$$

$$5.15 \quad P \{S_1 \cap S_2/R\} = P \{S_1/R\} \cdot P \{S_2/R\} \quad 138$$

$$5.16 \quad P \{R/S_1 \cap S_2\} = P \{R\} \cdot \frac{P \{S_1/R\}}{P \{S_1\}} \cdot \frac{P \{S_2/R\}}{P \{S_2\}} \quad 138$$

$$5.17 \quad O \{R/S_1 \cap S_2\} = O \{R\} \cdot LS_1 \cdot LS_2 \quad 138$$

$$5.18 \quad \text{logit} \{R/S_1 \cap S_2\} = \text{logit} \{R\} + W^+_1 + W^+_2 \quad 139$$

$$5.19 \quad \text{logit} \{R/S_1 \cap \bar{S}_2\} = \text{logit} \{R\} + W^+_1 + W^-_2 \quad 139$$

$$5.20 \quad \text{logit} \{R/\bar{S}_1 \cap S_2\} = \text{logit} \{R\} + W^-_1 + W^+_2 \quad 139$$

$$5.21 \quad \text{logit} \{R/\bar{S}_1 \cap \bar{S}_2\} = \text{logit} \{R\} + W^-_1 + W^-_2 \quad 139$$

$$5.22 \quad O \{R/S_1 \cap S_2 \cap S_3 \cap \dots S_n\} = O \{R\} \cdot \prod_{i=1}^n LS_i \quad 139$$

$$5.23 \quad \text{logit} \{R/S_1 \cap S_2 \cap S_3 \cap \dots S_n\} = \text{logit} \{R\} + \sum_{i=1}^n W^+_i \quad 139$$

$$5.24 \quad T_{ij}^* = \frac{T_i \cdot T_j}{T_{..}} \quad 141$$

$$5.25 \quad X^2 = \sum_{i=1}^n \sum_{j=1}^m \frac{(T_{ij} - T_{ij}^*)^2}{T_{ij}^*} \quad 141$$

$$5.26 \quad V = \sqrt{\frac{X^2}{T_{..} M}} \quad 141$$

$$5.27 \quad H(A) = - \sum_{j=1}^m p_{.j} \ln p_{.j} \quad 142$$

$$5.28 \quad H(B) = - \sum_{i=1}^n p_{i.} \ln p_{i.} \quad 142$$

$$5.29 \quad H(A,B) = - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \ln p_{ij} \quad 142$$

$$5.30 \quad U(A,B) = 2 \left[\frac{H(A) + H(B) - H(A,B)}{H(A) + H(B)} \right] \quad 142$$

$$5.31 \quad C = W^+ - W^- \quad 144$$

$$5.32 \quad S^2(C) = \frac{1}{P \{S \cap R\}} + \frac{1}{P \{S \cap \bar{R}\}} + \frac{1}{P \{\bar{S} \cap R\}} + \frac{1}{P \{\bar{S} \cap \bar{R}\}} \quad 144$$

$$5.33 \quad P_{x,y} \{R/S_1 \cap S_2 \cap \dots S_n\} = \frac{O \{R\} \cdot e^{\sum_{i=1}^n W^+_{x,y}}}{1 + O \{R\} \cdot e^{\sum_{i=1}^n W^+_{x,y}}} \quad 145$$

$$5.34 \quad F_w = \frac{\sum_{s=1}^{tw} \left[1 - \sum_{i=1}^p \frac{|a_{i1} - a_{i2}|}{2w^2} \right]^s}{tw} \quad 148$$

$$5.35 \quad F_t = \frac{\sum_{w=1}^n F_w e^{-k(w-1)}}{\sum_{w=1}^n e^{-k(w-1)}} \quad 149$$

$$5.36 \quad L = \log \left[\frac{P_{i,j}(x,y)}{1 - P_{i,j}(x,y)} \right] = \beta_{0,ij} + \beta_{1,ij} \cdot V_{1,xy} + \dots + \beta_{k,ij} \cdot V_{k,xy} \quad 151$$

$$5.37 \quad P_{i,j}(x,y) = \frac{e^L}{1 + e^L} \quad 152$$

$$5.38 \quad g_1(x) = \log \left[\frac{P(Y=1/X_{1-p})}{P(Y=0/X_{1-p})} \right], \quad g_1(x) = \beta_{10} + \beta_{11} \cdot X_1 + \dots + \beta_{1p} \cdot X_p \quad 152$$

$$5.39 \quad g_2(x) = \log \left[\frac{P(Y=2/X_{1-p})}{P(Y=0/X_{1-p})} \right], \quad g_2(x) = \beta_{20} + \beta_{21} \cdot X_1 + \dots + \beta_{2p} \cdot X_p \quad 152$$

$$5.40 \quad P(Y=0/X_{1-p}) = \frac{1}{1 + e^{g_1(x)} + e^{g_2(x)}} \quad 153$$

$$5.41 \quad P(Y=1/X_{1-p}) = \frac{e^{g_1(x)}}{1 + e^{g_1(x)} + e^{g_2(x)}} \quad 153$$

$$5.42 \quad P(Y=2/X_{1-p}) = \frac{e^{g_2(x)}}{1 + e^{g_1(x)} + e^{g_2(x)}} \quad 153$$

$$5.43 \quad l(\beta) = \prod_{i=1}^n \{ \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \}, \quad 153$$

$$5.44 \quad \pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} \quad 153$$

$$5.45 \quad L(\beta) = \ln [l(\beta)] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln [1 - \pi(x_i)]\} \quad 153$$

$$5.46 \quad \sum_{i=1}^n [y_i - \pi(x_i)] = 0 \quad 154$$

$$5.47 \quad \sum_{i=1}^n x_i [y_i - \pi(x_i)] = 0 \quad 154$$

$$5.48 \quad l(\beta) = \prod_{i=1}^n [\pi_0(x_i)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}}] \quad 154$$

$$5.49 \quad L(\beta) = \sum_{i=1}^n y_{1i} g_1(x_i) + y_{2i} g_2(x_i) - \ln (1 + e^{g_1(x_i)} + e^{g_2(x_i)}) \quad 154$$

$$5.50 \quad \frac{\partial L(\beta)}{\partial \beta_{jk}} = \sum_{i=1}^n x_{ki} (y_{ji} - \pi_{ji}) \quad 154$$

$$5.51 \quad W = \frac{\hat{\beta}_i}{(\hat{SE}) \hat{\beta}_i} \quad 155$$

$$5.52 \quad G = -2 \{L(\beta) - [n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)]\} \quad 155$$

$$5.53 \quad \Delta_{A,B} = 1/N \sum_{i=1}^N (x_A(i) - \bar{x}_A(i)) (x_B(i) - \bar{x}_B(i)) \quad 157$$

$$5.54 \quad \alpha_{A,B} = \frac{\Delta_{A,B}}{\sqrt{\sigma_A^2 \cdot \sigma_B^2}} \quad 158$$

$$5.55 \quad P_{i,j}(x,y) = \frac{e^{\beta_0 + \gamma_{i,j} V(x,y)}}{1 + \sum_{i=1}^t e^{\beta_0 + \gamma_{i,j} V(x,y)}} \quad 160$$

$$5.56 \quad m_j \hat{\pi}_j = m_j (\exp [\hat{g}(x_j)] / \{1 - \exp [\hat{g}(x_j)]\}) \quad 161$$

$$5.57 \quad r(y_j, \hat{\pi}_j) = \frac{(y_j - m_j \hat{\pi}_j)}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}} \quad 161$$

$$5.58 \quad X^2 = \sum_{j=1}^J r(y_j, \hat{\pi}_j)^2 \quad 161$$

$$5.59 \quad d(y_j, \hat{\pi}_j) = \pm \left\{ 2 \left[y_j \ln \left(\frac{y_j}{m_j \hat{\pi}_j} \right) + (m_j - y_j) \ln \left(\frac{(m_j - y_j)}{m_j (1 - \hat{\pi}_j)} \right) \right] \right\}^{1/2} \quad 162$$

$$5.60 \quad d(y_j, \hat{\pi}_j) = - \sqrt{2m_j \left| \ln(1 - \hat{\pi}_j) \right|} \quad 162$$

$$5.61 \quad d(y_j, \hat{\pi}_j) = - \sqrt{2m_j \left| \ln(\hat{\pi}_j) \right|} \quad 162$$

$$5.62 \quad D = \sum_{j=1}^J d(y_j, \hat{\pi}_j)^2 \quad 162$$

$$5.63 \quad \hat{C} = \sum_{k=1}^g \frac{(o_k - n_k' \bar{\pi}_k)^2}{n_k' \bar{\pi}_k (1 - \bar{\pi}_k)} \quad 162$$

$$5.64 \quad o_k = \sum_{j=1}^{ck} y_j \quad 163$$

$$5.65 \quad \bar{\pi}_k = \sum_{j=1}^{ck} \frac{m_j \hat{\pi}_j}{n'_k} \quad 163$$

$$5.66 \quad \prod (t + 1) = P^n \cdot \prod (t) \quad 164$$

$$5.67 \quad P_{ij} = a_{ij} / \sum_j a_{ij} \quad 165$$

$$5.68 \quad P = H V H^{-1} \quad 166$$

$$5.69 \quad P^n = H V^n H^{-1} \quad 166$$

$$5.70 \quad P^\infty = H \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} H^{-1} \quad 166$$

$$5.71 \quad \sqrt{\quad} (t + 1) = P [f(t)] \cdot \sqrt{\quad} (t) \quad 167$$

$$5.72 \quad Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad 169$$

$$5.73 \quad E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1} \quad 169$$

$$5.74 \quad \gamma_k = \text{Cov}(Z_t, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu) \quad 170$$

$$5.75 \quad \rho_k = \frac{\text{Cov}(Z_t, Z_{t+k})}{\sqrt{\text{Var}(Z_t)} \sqrt{\text{Var}(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0} \quad 170$$

$$5.76 \quad \underset{nx1}{Y} = \underset{nxp}{X} \underset{px1}{\beta} + \underset{nx1}{\varepsilon} \quad 171$$

$$5.77 \quad E(Y) = X \beta \quad 172$$

$$5.78 \quad \sigma^2(Y) = \sigma^2 I \quad 172$$

$$5.79 \quad \underset{px1}{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_{p-1} \end{bmatrix} \quad 172$$

$$5.80 \quad \underset{pxp}{(X'X)} \underset{px1}{b} = \underset{pxn}{X'} \underset{nx1}{Y} \quad 172$$

$$5.81 \quad \underset{px1}{b} = \underset{pxp}{(X'X)^{-1}} \underset{px1}{X'Y} \quad 172$$

$$5.82 \quad \hat{Y}_{n \times 1} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \cdot \\ \hat{Y}_n \end{bmatrix} \quad 173$$

$$5.83 \quad e_{n \times 1} = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ e_n \end{bmatrix} \quad 173$$

$$5.84 \quad \hat{Y} = Xb \quad 173$$

$$5.85 \quad e = Y - \hat{Y} \quad 173$$

$$5.86 \quad SSTO = Y'Y - n\bar{Y}^2 \quad 173$$

$$5.87 \quad SSR = b'X'Y - n\bar{Y}^2 \quad 173$$

$$5.88 \quad SSE = e'e = Y'Y - b'X'Y \quad 173$$

$$5.89 \quad MSR = \frac{SSR}{p - 1} \quad 174$$

$$5.90 \quad MSE = \frac{SSE}{n - p} \quad 174$$

$$5.91 \quad F^* = \frac{MSR}{MSE} \quad 174$$

$$5.92 \quad R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} \quad 174$$

$$5.93 \quad \hat{Y}_h \pm t(1 - \alpha/2; n - p) S(Y_{h(new)}) \quad 175$$

$$5.94 \quad S^2(Y_{h(new)}) = MSE (1 + X'_h (X'X)^{-1} X_h) \quad 175$$

5.95	$zre_i = \frac{e_i}{\sqrt{MSE}}$	175
5.96	$P_{nu_ind} = \frac{ind_f - ind_i}{non\ urb_i}$	177
5.97	$P_{nu_res} = \frac{res_f - res_i}{non\ urb_i}$	177
5.98	$P_{nu_serv} = \frac{(serv_f - serv_i) - res_serv}{non\ urb_i}$	178
5.99	$P_{res_serv} = \frac{(serv_f - serv_i) - nu_serv}{non\ urb_i}$	178
5.100	$P_{nu_res} = \frac{(res_f - res_i) - [(serv_f - serv_i) - nu_serv] - (mix_f - mix_i)}{non\ urb_i}$	179
5.101	$Q_{ij} = r x$ (expander function) + $s x$ (patcher function)	183
5.102	If $n_j > 3$ then $P'_{ij}(x,y) = P_{ij}(x,y)$ else	184
	$P'_{ij}(x,y) = P_{ij}(x,y) \times (n_j)/4$	
6.1	$Y = 29,705.243 + 0.116 \cdot urbpop$	239
6.2	$Y = 0.008343 \cdot indgdp$	242
6.3	$Y = 1,590.457 + 0.02401 \cdot urbpop$	245
6.4	$Y = 0.008081 \cdot totgdp$	285
6.5	$Y = 1,105.895 + 0.001795 \cdot indgdp$	288

LIST OF ABBREVIATIONS

ACF	- Autocorrelation Function
A.D.	- Anno Domini (in the year of our Lord)
AI	- Artificial Intelligence
ANOVA	- Analysis of Variance
AV	- Added Value
B.C.	- Before Christ
c.	- circa, approximately
CA	- Cellular Automata
CASA	- Centre for Advanced Spatial Analysis of the University College London
CEPAM	- Center for Studies and Research on Municipal Governments
CESP	- São Paulo State Agency for Energy
CIAGRI	- Informatics Center - Agronomy School of the University of São Paulo
CITP	- Tietê-Paraná Intermunicipal Consortium
CP	- Control Point
DAEE	- São Paulo State Agency for Water and Electric Energy
EMPLASA	- São Paulo State Agency for Metropolitan Planning
EPUSP	- Engineering College of the University of São Paulo
ESA	- European Space Agency
ESALQ	- School of Agronomy “Luiz de Queiroz” of the University of São Paulo
GDP	- Gross Domestic Product
GIS	- Geographical Information Systems
GISc	- Geographical Information Science

IBGE - Brazilian Institute for Geography and Statistics

IEEE - Institute of Electrical and Electronic Engineers

INPE - National Institute for Space Research

JRC - Joint Research Center – European Commission

IPCA - Consumers Annual Prices Index

IPEA - National Institute for Applied Economics

ITE-FCEB - Toledo Educational Institute – Faculty of Economic Sciences – Bauru

LEGAL - Spatial Language for Algebraic Geocomputation

LSN - Department of Soils and Plants Nutrition of ESALQ

LUCC - Land Use and Land Cover Change

MAS - Multi-Agent Systems

MSS - Multispectral Scanner Sensor

NICs - Newly Industrialized Countries

NASA - National Aeronautics and Space Administration

NPP - Net Primary Production

ORB - Online Reference Book

PhD - Doctor of Philosophy

PND - National Development Plan

RS - Remote Sensing

SEPLAN - Planning Secretariat of Bauru Municipal Government

SOC - Self-Organized Criticality

SUDENE - National Agency for the Northeastern Region Development

SUDAM - National Agency for the Amazon Region Development

TOMM - Time-Oriented Metropolitan Model

TM - Thematic Mapper Sensor

TM - 5 - Thematic Mapper Sensor of Landsat 5

TNCs - Transnational Corporations

USP - University of São Paulo

UNESP - State University of São Paulo

UTM - Universal Transverse Mercator (Geographic Projection System)

WRS - Worldwide Reference System

CHAPTER 1

INTRODUCTION

1.1 Overview of the Research Topic

In the two last centuries, particularly in the latest decades, humankind has witnessed a drastic shift of population from rural to urban areas. According to a World Bank report, the proportion of the world population living in urban areas rose from 33.6% in 1960 to 46.6% in 1999, and this rate is expected to reach up to 75.8% in the newly industrialized (NICs) and developed countries (Easterly and Sewadeh, 2000). It is estimated that 55% of the global population will be urban dwellers by 2015 (United Nations, 1996).

One of the earliest and most important drivers for these changes in the urban environment in the developed world since the XIX and beginning of the XX century is the Industrial Revolution, which actually started to take place in the second half of the XVIII century. A population surplus, enhanced by continuous drops in mortality rates, in association with migratory flows of *expatriated* rural workers caused serious impacts on towns at that time. This population shift was accompanied by a growth of goods and services in urban areas, which was supported by gains in productivity in agricultural and industrial production, made possible by the rapid technological changes and economic development. Just to mention a few examples, the population of Manchester shifted from 12,000 in 1760 to over 400,000 inhabitants by the middle of the XIX century. London, which had already one million inhabitants by the end of the XVIII century, sheltered two and half million people around 1851 (Benevolo, 1983).

Similar processes of industrialization and rapid urbanization were experienced in the developing world in the XX century. In Latin America, Brazil and Argentina entered the era of large-scale industrial production from the 1950s onwards, whereas countries such as South Korea, Indonesia and Malaysia went through a transformation of moving from a basically rural into an industrial economy mainly in the late 1970s and 1980s.

Between 1950 and 1995, the urban population in Asia, Africa, Latin America and the Caribbean grew more than fivefold from 346 million to 1.8 billion. Although Asia and Africa still have more rural than urban dwellers (TABLE 1.1), they both have very large urban populations in terms of absolute figures. United Nations projections suggest that urban populations are growing so much faster than rural populations that 80 per cent of the growth in the world's population between 1990 and 2010 will be in urban areas and virtually all this growth will be in Africa, Asia and Latin America (United Nations, 1998).

TABLE 1.1 – Trends and projections in urban populations by region, 1950 – 2010.

<i>Percentage of Population Living in Urban Areas</i>					
<i>Region</i>	<i>1950</i>	<i>1965</i>	<i>1980</i>	<i>1995</i>	<i>2010*</i>
<i>Africa</i>	14.6	20.7	27.3	34.9	43.6
<i>Asia</i>	17.4	22.4	26.7	34.7	43.6
<i>Latin America and the Caribbean</i>	41.4	53.4	64.9	73.4	78.6
<i>Rest of the world</i>	55.3	64.1	70.5	74.2	78.0

* This is projected according to censuses held around 1990. Rest of the world includes all countries in Europe, North America and Oceania.

SOURCE: United Nations (1998, p. 25).

Two aspects of this rapid growth in urban population have been the increase in the number of large cities and the historically unprecedented size of the largest cities. Two centuries ago, there were only two 'million-cities' worldwide (i.e. cities with one million or more inhabitants) – London and Beijing. By 1950 there were 80 and by 1990 there were 293 'million-cities'. A large (and increasing) proportion of these million-cities are in Africa, Asia and Latin America, and many have populations that grew more than tenfold between 1950 and 1990 (Hardoy et al. 2001).

In Brazil, urban growth and development was for a very long time concentrated in coastal State capitals. Some of them were defined as metropolitan areas in the 1970s in face of their high demographic density. Given their strategic location and the fact that they were well supplied with technical infrastructure and social equipments, they could keep strong ties with their surrounding regions, organizing and centralizing the economic relationships within their catchment areas.

The development strategies adopted in Brazil as well as in many other Latin American countries were oriented towards foreign markets and aimed at economic efficiency and increasing competitiveness. This brought about the rising of external economies of agglomeration¹, leading to the above-mentioned sharply concentrated pattern of spatial development (Fernandes et al. 1977).

Nevertheless, this multi-centralized development framework of 'economies of agglomeration' had a number of serious drawbacks. After an initial period of good economic performance, metropolitan areas started to behave as "diseconomies of agglomeration", in face of problems including: increasing real estate prices; imminent collapse of the transportation, telecommunication and/or water supply systems; greater time and costs spent in displacement trips inside the metropolitan areas; higher wages; environmental problems; and criminality.

Policies to overcome these regional imbalances and the adverse effects of the concentrated spatial development have been established by the Brazilian federal government since the late 1950s and 1960s, which resulted in the creation of the Manaus tax free zone in the Amazonas State (1957); the foundation of the new federal capital, Brasília (1960); the creation of regional development institutions like the Agency for the Northeastern Region Development (SUDENE) in 1959, the Agency for the Amazon Region Development (SUDAM) in 1956, etc.; and in actions promoting the

¹ The term 'economies of agglomeration' is used in urban economics to describe the benefits that firms obtain when locating near each other. It is related to the idea of economies of scale and network effects, in that the more related firms that are clustered together, the lower the cost of production and the greater the market they can sell into (Wikipedia, 2003).

decentralization of the industrial development in the São Paulo metropolitan area, amongst other initiatives.

Of all these actions, the II National Development Plan (II PND), launched in 1974, was a most ambitious plan to relocate part of the industries situated in the São Paulo metropolitan area to other regions in Brazil. This plan included the creation of industrial development funds and of policies to foster export and agribusiness activities as well as to decentralize the production of basic industrial inputs. Its goals have been partially achieved, since part of the intended industrial relocation actually took place in the inland of São Paulo State. This region offered good infrastructure conditions, successful agricultural enterprises and an urban network already established since the expansion of export coffee plantations during the XIX century (Fundação Seade, 1992).

Two of the medium-sized cities located in the inland of São Paulo State that inherited part of the metropolis industrial development are the very foci of the present PhD research: Bauru and Piracicaba. In both cases, the challenges for studying them lie on the fact that they underwent urbanization booms and show a great diversity concerning their economic specialization profiles throughout the latest decades.

The importance in drawing special attention and conducting in-depth research on urban areas can be ascribed to the fact that not only they are bound to shelter the greatest part of the world's population, as previously exposed, but also and chiefly to the fact that they control the world's economy in the present era of globalization, managing the flows of financial resources, man-made and natural assets, human capital, information, technical and scientific knowledge and decision power. On a due and prudent command of their institutional and financial framework as well as on a skilful running of its physical structure will the success or failure of the majority of mankind's undertakings depend.

This is exactly the concern of local strategic planning, which strives for a better understanding of dynamic changes that occur in the physical sphere of urban areas, and also for developing tools and skills to foresee probable events that might take place in the urban environment in a near future.

Traditional planning has always endeavored to rationalize and clarify territorial changes in cities and anticipate possible transformations in the urban scene as well. In the same way, quantitative methods have for long been employed in urban issues, struggling to understand processes of urban change. Von Thünen's theory of concentric rings, Weber's classical triangular model of industrial location, Christaller's model of central places, and Lösch's theory of economic regions all date back from the last decades of the XIX century and beginning of the XX century, respectively 1826, 1909, 1933, and 1940 (Merlin, 1973).

With the advent of personal computers and the Quantitative Revolution in social sciences and related fields in the late 1950s and early 1960s, computerized urban models came into play. Initially, these models were cross-sectional in time and, with a few exceptions, largely sectorial, in the sense that different aspects of urban life (transportation, residential demand, retail location, etc.) were handled by different models. Clark's model of residential location (Clark, 1951), Lowry's model of transportation for the Pittsburgh area (Lowry, 1964), and Lakshmanan and Hansen's retail location model (Lakshmanan and Hansen, 1965) are worthy of mention.

The following generation of models attempted to model temporal dynamics and to provide a more integrative approach (Crecine, 1964; Batty, 1971, 1976; Allen et al. 1981; Wilson, 1981; Wegener et al. 1986). The refinements introduced by these models included: (a) addition of the time dimension in the quantitative analyses; (b) employment of sophisticated mathematical and theoretical tools (e.g. differential equations, Catastrophe and Bifurcation Theory, etc.); (c) analysis of the spatial interactions among different activities in a city. Although the improvements introduced, these models remained fairly non-spatial, especially in the sense that their results could not be spatially visualized.

Effective advances in the spatial representation of urban models occurred only by the end of the 1980s, when cellular automata (CA) models started to be widely applied. Cellular automata are simple – usually grid-based – formal systems, in which dynamic change is represented grid-cell by grid-cell, as a simple mapping from the current state

of a cell and its neighbors into the next state which the cell will change. A CA model of a city focuses on dynamics and can be used to investigate urban processes of change (Couclelis, 1985; Batty and Xie, 1996; Xie, 1996).

According to Batty et al. (1997), CA were somehow implicit in the wave of computer models designed for land use – transportation planning already in the early 1960s. Chapin and his colleagues at North Carolina in their modeling of the land development process articulated cell-space models where changes in state were predicted as a function of a variety of factors affecting each cell, some of which embodied neighborhood effects (Chapin and Weiss, 1968); while Lathrop and Hamburg (1965) proposed similar cell-based simulations for the development of western New York State.

“Strict CA models came into the urban field from another source – from theoretical quantitative geography. These were largely due to Waldo Tobler who during the 1970s worked at the University of Michigan where Arthur Burks and his Logic of Computers group were keeping the field alive. Tobler (1970) himself first proposed cell-space models for the development of Detroit but in 1974 formally began to explore the way in which strict CA might be applied to geographical systems, culminating in his paper “Cellular geography” (Tobler, 1979). At Santa Barbara, in the 1980s, Couclelis influenced by Tobler continued these speculations, until the late 1980s when applications really began to take off as computer graphics, fractal, chaos, complexity all generated conditions which have led to the currently seen advances” (Batty et al. 1997).

The 1990s experienced successive improvements in urban CA models, which started to incorporate environmental, socioeconomic and political dimensions, and were finally successful in articulating analysis factors of spatial micro and macroscale (Phipps and Langlois, 1997; White and Engelen, 1997; White et al. 1998).

There are now some twenty or more applications of CA to cities (Batty, 2000), such as intra-migration and social segregation (Portugali et al. 1997), retail location optimization (Benati, 1997), traffic network expansion (Batty and Xie, 1997), urban

growth (Clarke et al. 1997; Clarke and Gaydos, 1998), urban land use change (Phipps and Langlois, 1997; White and Engelen, 1997; White et al. 1998), etc.

It is worthy remarking that spatial dynamic models in a general way represent an immediate challenge for the coming generation of Geographical Information Systems (GIS). According to Burrough (1998), methods of open systems modeling of which CA is one of the best examples and which meet many of the general requirements for simulating dynamic processes quickly and efficiently, are rarely implemented in GIS. In Openshaw's opinion, "...GIS remains surprisingly narrowly focused, ..., it is largely devoid of much of the modeling and the simulation relevant to the modern world..." (Openshaw, 2000). All these opinions find support in the work of Câmara et al. (2003), for whom the current computational paradigms of knowledge representation are essentially static and unable to appropriately model the temporal dimension and the dynamic context-based relationships amongst entities and their attributes.

This PhD research is inserted in this context, and by carrying out simulations of land use change applied to two real cities, it seeks to provide evidence to the needs of the Geographical Information Science (GISc) community in conceiving techniques and abstractions capable of properly representing dynamic environmental phenomena.

1.2 General Research Goals

The general goals of this research are the following:

- To perform experiments of urban land use change for two cities – Bauru and Piracicaba, located in the western inland area of São Paulo State, Brazil – using a cellular automata model, driven by parameters obtained through two empirical statistical methods (weights of evidence and logistic regression), as well as by spatial information extracted from Landsat 5 – Thematic Mapper (TM – 5) satellite images; digital land use, occupation density, technical and social infrastructure² maps; and digital aerial photos.

² Technical infrastructure can be understood as the underlying or hard framework of a city, consisting of traffic and transport systems, energy supply systems, water supply and sewerage, telecommunications, etc. Social infrastructure,

- To identify, by means of a multi-temporal analysis of land use change processes observed in the two cities, the role of the biophysical and socioeconomic variables shaping urban form. The aim is to understand the dynamic interaction amongst these factors, and hence to elaborate possible future scenarios of land use arrangements in the short- and medium-terms for these cities.
- To collaborate, through the land use change modeling experiments and the thereof derived investigations and findings, to provide insights and answers to the needs of GISc in developing techniques and abstractions capable to suitably represent dynamic events of the real world.

1.3 Specific Research Goals

Concerning specific research goals, this PhD research should be proper to:

- Serve as a guidebook for the accomplishment of land use change simulations of cities either in Brazil or worldwide.
- Subsidize, in a general way, the decision-making processes of local and regional administrators related to urban and environmental planning in the cities of Bauru and Piracicaba as well as those related to regional planning working for the administrative regions of Bauru and Campinas, which respectively comprise the municipalities of Bauru and Piracicaba.
- Supply guiding information for specific sectorial plans at the intra-urban scale, such as:
 - industrial location plans (subdivided according the field of production);
 - residential settlements plans;

on its turn, constitutes the soft framework necessary for the functioning of a city, like educational and health care equipments, commercial and services activities, sports facilities, religious and public institutions, leisure areas, etc.

- social equipments location plans (schools, health care facilities, kindergarten, etc.);

- plans for the implementation of technical infrastructure (bus routes and stops, roads paving, extension of water supply and sewerage pipes, construction of landfills, expansion of the energy supply network, etc.).

1.4 Research Hypotheses

In face of what was exposed in the previous sections, this PhD research postulates the following hypotheses:

- a) Is it possible to run models of urban land use change in a cellular automata environment that simulate the patterns and processes of spatial transformations of land use in distinct medium-sized São Paulo inland cities that were subjected to urbanization booms?
- b) In case the first hypothesis is deemed affirmative, can these models be driven by spatial information obtained from digital maps, remote sensing (RS) images and digital aerial photos? Is it possible that the spatial variables identified in these cartographic products provide indication of how and where changes in the urban land uses are taking place, and therewith, allow the elaboration of maps showing the areas most susceptible to undergo changes?
- c) If this late hypothesis can be confirmed, what would be such variables? How to select them? In which way can they be combined for the generation of the above-mentioned maps?
- d) Moreover, what would be proper statistical methods to weight the different variables according to their contribution for rendering certain areas more or less prone to changes? What would be the comparative effectiveness of these methods in detecting land use transitions?

- e) Finally, can all the information obtained in the previous stages be used for conceiving forecast scenarios of land use change? What statistical methods would be suitable for estimating future land use change rates? Would regression analysis models relating past rates of transition and socioeconomic indicators be appropriate for this end?

The answers to these queries are presented at the end of Chapter 6, when the whole process of digital spatial data pre-processing, selection and weighting of variables, parameterization and calibration of the models, and simulations running will be concluded, validated and evaluated.

1.5 Detailed Outline of the Thesis

After an introductory overview on the thesis topics and the statement of the objective and specific goals of this research in Chapter 1, the theoretical foundations on urban models are approached in Chapter 2.

Firstly, an anthology of definitions on urban areas and related concepts is presented, followed by a brief historical perspective on urban models in the second part of the Chapter, showing the progressive advances achieved in the latest decades, particular with respect to their spatial representation.

Since spatial dynamic models are the central theme of this thesis, the third part of Chapter 2 is dedicated to considerations on how space and time is approached in the present work. The third part of Chapter 2 still contains a special section on Cellular Automata, for it best represents systems of dynamic modeling, and also a conclusive section justifying the usage of a CA model for the urban land use simulations carried out in this research.

Chapter 3 introduces the two case study cities adopted for the modeling experiments - Bauru and Piracicaba - regarding their location, historical background and socioeconomic particularities. Their recent economic specialization profile will be understood in view of the development process of São Paulo State and Brazil since early

colonial times. This Chapter attempts to clarify the reasons lying behind their high urbanization growth rates as well as their current status of regional and industrial development poles in São Paulo State.

Chapter 4 is committed to present both the spatial and non-spatial data that drive the urban land use simulation experiments. The first section introduces the spatial data, like maps of land use, aerial photos and satellite images, mentioning their characteristics such as source, scale, type of information, level of detailing and date of acquisition, whereas the non-spatial data, which are those concerning demographic information and economic indicators records and exclusively used for the simulation of land use change forecasts, are presented in the second section. A concluding section describes pre-processing operations undertaken in the geographical database.

Particular attention is drawn to methodological aspects in Chapter 5, which consists of two sections introducing the empirical statistical methods employed for the parameterization of the simulation models (weights of evidence and logistic regression), and of a conclusive section, designed to present the dynamic simulation software used in the modeling experiments.

Chapter 6 reports the results obtained for the different time periods of simulation, including the short-term (2004) and medium-term (2007) forecasts time horizons. Special attention is paid to justify the sets of variables selected to explain the respective land use transitions in each of the simulation periods in light of economic theories of urban growth and development.

Finally, an attempt is made in the concluding Chapter 7 to draw together the whole range of topics, findings and questions raised by all of the foregoing material. Recommendations as to the future paths of this research work are added to this discussion.

CHAPTER 2

THEORETICAL FOUNDATIONS ON URBAN MODELS

2.1 The Modeling Object

Once the present work deals with the precise topic of urban land use change modeling, it is essential to initially draw a due attention to the definition of this research object.

2.1.1 Definitions of Urban Settlements

The mankind, gregarious since its remote times, always lived in groups or communities. Human beings are social animals who for reasons such as mutual support (breeding, feeding, protection) and territoriality are inclined to form communities (ORB, 2002). Since their origins on Earth about 500,000 years ago and for a very long time henceforth (which corresponds to the Pleistocene epoch), they lived collecting their food and searching for shelter in the natural environment without modifying it in a deep and permanent way.

Around 10,000 years ago, after the glaciers thaw completion - the last deep environment transformation marking the transition from the Pleistocene to the Holocene – the inhabitants of the temperate zone learned how to produce their own food, cultivating plants and raising cattle, organizing the first stable settlements – the villages – located near to their work places. This is the Neolithic period, which for many peoples extended until the European colonization (for the Maoris in New Zealand, it lasted until the XIX century).

Approximately 5,000 years ago, on the flood plains of the Middle East, some villages turned into cities; the food providers were persuaded or forced to generate a surplus in order to sustain a class of specialized labor force: craftsmen, traders, warriors and priests. Since they were in a more favorable condition living in cities, they could keep an ascendance over the countryside. This social organization demanded the invention of

writing; and this is the very point in time where civilization and the written history begins in opposition to the prehistory. Henceforward, all the successive historical events will depend on the amount and distribution of this food surplus (Benevolo, 1983).

Researchers in this field distinguished within the aforementioned period a) the Brass Age, in which the metals used for tools and weapons were rare and expensive, being reserved to a restricted leading class, which absorbed all the available food surplus but at the same time also limited the population and production growth; and b) the Iron Age, which started around 1,200 B.C. with the diffusion of alphabetic writing, stamped coins and more accessible metallic tools, enlarging hence the leading class and allowing an increase in population. The Greco-Roman civilization developed under these precepts on a very large area – the Mediterranean Basin – but enslaved and impoverished the food providers, being in this way condemned to an economic collapse, what indeed happened from the IV A.D. century onwards.

Other historical events – the feudal and bourgeois civilizations – made way for the following historical transition: the usage of scientific methods directly in the production, what characterizes the industrial civilization. The ever-increasing and unlimited surplus was not necessarily reserved to a leading minority, but is distributed for the greater part of the population, and theoretically for the whole of it, which have no more economic hindrances to grow. In this new situation, the city (seat of the dominant classes) still opposes itself to the countryside (seat of the subordinate classes), but this dichotomy was not inevitable and could be overcome. And from this very daring possibility was born the modern city (Benevolo, 1983).

Regarding the central topic of this section, it is important to note that the term urban settlements encompasses a set of related concepts: urban centers, towns, cities. According to Hardoy et al. (2001), there is no general agreement among governments as to how define 'a town', an 'urban center' and a 'city'. In virtually all nations, urban centers include all settlements with 20,000 or more inhabitants, but governments differ in what smaller settlements they include as urban centers – from those that include all

settlements with a few hundred inhabitants as urban to those that only include settlements with 20,000 or more inhabitants.

Most governments define urban centers in one of four ways: through size thresholds (for instance all settlements with 2,000 or more inhabitants are urban); through size thresholds combined with other criteria (for instance for density or for proportion of the economically active population in non-agricultural activities); through administrative or political status (for instance all settlements designed as national, provincial or district capitals are urban centers); or through lists of settlements named in the census being 'the urban centers'.

The term 'city' and 'urban center' are often used interchangeably, and there is no consensus as to the definition of a city and how it might differ from the definition of an urban center. But few people would consider that settlements with 1,000 or 2,000 inhabitants are cities. There are thousands of settlements in Africa, Asia and Latin America that have only a few thousand inhabitants and are considered by their governments as urban, which lack the economic, administrative or political status that would normally be considered as criteria for classification as a city. In countries with long urban histories, there may be many historic 'cities' which achieved city status many decades or centuries ago because of their importance at that time (for instance by being political capitals or religious centers or key trading centers) and are still considered as cities, even though they are relatively small within their national urban system. The term 'town' implies a small urban center and might be taken to include all urban centers that were not cities, but again, there is no consensus on this (Hardoy et al. 2001).

These are official definitions, usually employed by some international and governmental institutions. Regarding researchers proceeding from the urbanistic and related fields, cities may however assume broader and more divergent conceptualizations. Merlin (1973), in the light of the historic materialism, defines the city as a place for the association of individuals and their activities, so as to enable more efficient production means. For Serra (1991), also a materialist theoretician, cities are

agglomerations whose supreme role is the surplus control. According to the ecological theory, the city is regarded as a man-made ecosystem, with its own energy periods and eco-cultural species (Odum, 1963; Watt, 1973). Mc Loughlin (1972), guiding himself through the Systems Approach, sees the city as a cybernetic machine or a complex system of activities and flows hierarchically arranged in special subsystems. For the contemporary theory of strategic planning and management, the city is a mega-organization, i.e. a wide set of multidimensional collective action (Fischer, 1996).

2.1.2 The Modeling Object in the Present Work

In Brazil, there is no official definition or classification of cities regarding their size in terms of population figures. The metropolitan areas in São Paulo State, for instance, are defined with no mention to population amount as "...the cluster of neighboring municipalities that assumes a remarkable national prominence, in view of its high demographic density, an outstanding conurbation pattern and a high degree of socioeconomic integration, diversity and specialization of its urban and regional functions so as to demand an integrated planning strategy and a permanent collective action of its public entities" (São Paulo, 1989).

The Brazilian Federal Constitution obliges every city with a minimum population of 20,000 inhabitants to have its own master plan (Brasil, 1988), what makes it implicit that urban settlements with less population than this threshold may also be acknowledged as cities. In the Amazon region, for instance, new municipalities are created to improve local administration, and their seat cities not rarely contain a population of just a few thousand inhabitants. According to some official cartographic studies, urban settlements containing between 5,000 and 20,000 inhabitants are regarded as "rural cities", and any human settlement with less than 5,000 inhabitants is considered either a hamlet or a village.

Any official regulation for the Brazilian cities size in function of their population would have to be continuously revised, given the rapid urbanization processes underwent by most cities. A city regarded as big in the first decades of last century, could be

considered as medium in the 1940s or 1950s and even as small in the beginning of this century.

Some population thresholds are however extra-officially adopted in some urban studies. Considering the current Brazilian urban network patterns, it is sensible to classify urban settlements with political-administrative autonomy sheltering more than 20,000 and less than 100,000 inhabitants as small cities; more than 100,000 and less than 500,000 inhabitants, as medium-sized cities; more than 500,000 and less than 1,000,000 inhabitants as relatively big cities; and those containing more than 1,000,000 inhabitants as big cities, which can eventually become seats of metropolitan areas. Some international definitions also employ the term “mega-cities” for those cities sheltering over 10,000,000 inhabitants.

Urban settlements of small size and detached from main urban agglomerations are regarded as urban districts and are politically subordinated to a municipality, which is the smallest political-administrative division in Brazil.

The modeling object in the present work concerns two medium-sized cities located in the western inland of São Paulo State: Bauru and Piracicaba, which contained an urban population respectively of 310,442 and 317,374 inhabitants in the year 2000 (IBGE, 2000). Only their main urban agglomerations and immediate non-urban surroundings are considered in the simulations, what implies that urban districts located farther than 10 km from the official urban boundaries were excluded from the modeling experiments.

2.2 A Brief Historical Perspective on Models of Urban Land Use Change

2.2.1 Introduction

Before introducing a short historical background on models of urban land use change, it becomes necessary to explain the terms herewith related. Various are the definitions for the terms land, land use and land use change, and they vary with the purpose of the application and the context of their use (Briassoulis, 2000).

Wolman (1987) cites Stewart's (1968) definition of land in the scope of natural sciences: "the term land is used in a comprehensive, integrating sense ... to refer to a wide array of natural resource attributes in a profile from the atmosphere above the surface down to some meters below the land surface. The main natural resource attributes are climate, land form, soil, vegetation, fauna and water."

In a more economical approach, Hoover and Giarratani (1984) state that land "first and foremost denotes space ... The qualities of land include, in addition, such attributes as the topographic, structural, agricultural and mineral properties of the site; the climate; the availability of clean air and water; and finally, a host of immediate environmental characteristics such as quiet, privacy, aesthetic appearance, and so on."

Land use, on its turn, denotes the human employment of land (Turner and Meyer, 1994). Skole (1994) states that "land use itself is the human employment of a land-cover type, the means by which human activity appropriates the results of net primary production (NPP) as determined by a complex of socioeconomic factors." Finally, FAO/IIASA (1993) states that "land use concerns the function or purpose for which the land is used by the local human population and can be defined as the human activities which are directly related to land, making use of its resources or having an impact on them."

According to Briassoulis (2000), land use change "... means quantitative changes in the areal extent (increases or decreases) of a given type of land use...". For Jones and Clark (1997), it may involve either a) conversion of one type of use into another or b) modification of a certain type of land use, such as changes from high-income to low-income residential areas (the buildings remaining physically and quantitatively unaltered), etc.

All these differences in definition reveal how different disciplines theorize on and model land use change (Briassoulis, 2000).

As to the particular term "model", it can be understood as the representation of a system, which can be achieved through different languages: mathematical, logical, physical, iconic, graphical, etc., and according to one or more theories (Novaes, 1981).

A system is a set of parts, presenting interdependence among its constituent components and attributes (Chadwick, 1973). A theory, on its turn, can be defined as a set of connected statements which, through logical constructs, supplies an explanation of a process, behavior, or other phenomenon of interest as it exists in reality (Chapin and Kaiser, 1979; Johnston et al. 1994).

According to Batty (1976), Popper's concept of science as "a process of conjecture, then refutation of problems, followed by tentative solutions, error-elimination and the redefinition of problems (Popper, 1972)" is reflected in the development of urban theories and models.

In a general way, models can be basically classified according to the following typologies (Echenique, 1968; Novaes, 1981):

- *descriptive model*: it aims at solely understanding the functioning of a system;
- *explorative model*: it is a descriptive model that involves the parametric analyses of several states, by means of variations in the systems elements and their relations, with no external interference upon it. These kind of models are meant for answering "what if" questions;
- *predictive model*: it is an explorative model that involves the time variable, comprising the projection of some basic elements;
- *operational model*: it renders available the interference of the modeler, who can introduce exogenous factors in the system components and relations in a way to modify its behavior³.

More detailed categorization of specifically land use and urban land use change models are proposed by several other authors. Merlin (1973), for instance, divides urban models into three categories - urban development, transportation, and urban resources - subdividing each of these three categories according to their goals and methods.

³ A branch of operational models would be the prescriptive or normative models, which attempt to change the system under analysis in some optimal way.

Explanatory, descriptive, and stochastic models are placed in the first category; models dealing with economic choice of transport means and geographical displacements (including gravitational models) are found in the second category; and finally, geographical investigation (including central place models), econometric and statistical models are approached in the category of urban resources models.

Perraton and Baxter (1974) and Novaes (1981) classify urban and regional models generically into empirical models, microeconomic or behaviorist models, macroeconomic or social physics models, and dynamic simulation models.

The most extensively detailed categorization of generic land use change models is presented by Briassoulis (2000). According to her, models can be classified in view of their functional and methodological aspects into statistical and econometric; spatial interaction models; optimization models (which include linear, dynamic, goal, hierarchical and non-linear programming as well as utility maximization models and multi-objective/multi-criteria decision making models); integrated models (comprising gravity, simulation integrated and input-output integrated models); natural sciences-based models; GIS-based models and Markov chain-based models.

The classification of urban (and regional) land use change models herein proposed focuses on basic conceptual and operational aspects of such models, like their ability in apprehending and dealing with the spatio-temporal representation of events, and will observe a chronological path regarding their creation.

2.2.2 Non-Dynamic Models of Urban Land Use Change

Theoretical and mathematical models have for long been created for purposes of urban studies, aiming at clarifying processes of urban and regional change. One of the oldest contributions in this sense is Von Thünen's theory of concentric rings, dated back to 1826 (Merlin, 1973). Von Thünen was actually concerned with agricultural location, but the urban use was decisive to his problem. He conceived a very simple economic model consisting of one city and its concentric surrounding regions. According to his rent-locational model, within a certain distance from the center, a particular crop will outbid

the others; in a second area, another crop will outbid the others, and so on. In this case, the most intensive use of land will be near the center, and the rent or land values will decrease outwards (Perraton and Baxter, 1974).

Another similar approach in economic theory is a model of industrial location proposed by Weber in 1909. He was committed to explain the location of a firm in a featureless plain. The final location will be at the point where the costs for transporting the raw material to the firm and for delivering the finished product to the market are minimum (Merlin, 1973; Perraton and Baxter, 1974). This has been called the classical Weber triangle (FIGURE 2.1).

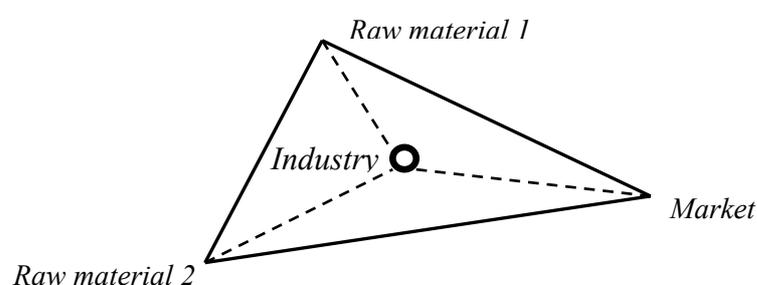


FIGURE 2.1 – Weber's classical triangle for industrial location.

SOURCE: Perraton and Baxter (1974, p. 36).

Another related achievement based on economic theory is Christaller's model of central places of 1933 (Merlin, 1973). According to him, cities are central places hierarchically organized, whose fundamental role is the provision of goods and services. His basic idea was that every point in space should be less than one hour away (about 4 km) from a central place. This pretext has led him, by means of equilateral triangles, to organize the regional space in a regular hexagonal structure (FIGURE 2.2), which reflects the range of influence of each place, i.e. the extent of its catchment area.

Lösch developed Christaller's idea further and conceived his theory of economic regions in 1940. Whereas Christaller was concerned with the location of settlement and

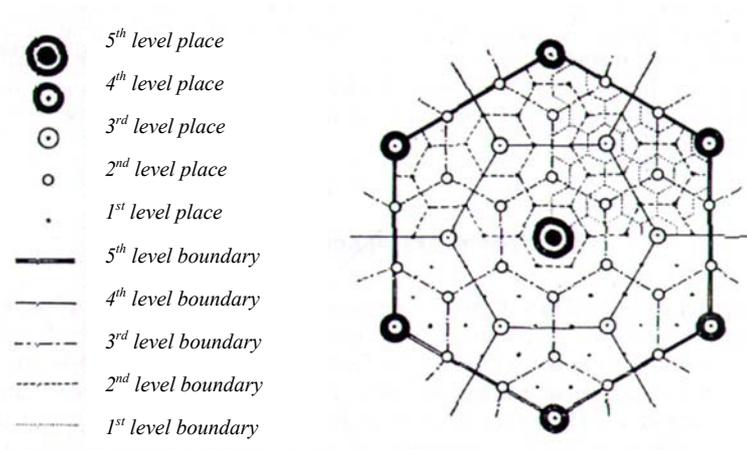


FIGURE 2.2 – Hexagonal system of Christaller’s central place theory.

SOURCE: Merlin (1973, p.152).

villages, Lössch’s theory considers retail or service location. According to it, the centers can be classified into hierarchical groups according to the type of service and size of the market area. If the population is distributed homogeneously throughout the considered area, there is an increase in the demand for a particular good near the center because of costs decreases to it; the demand curve drops away with distance from the center. Since this is a circular market area, the free entry of business in a perfect competition, produces a trade area of nested hexagons (FIGURE 2.3), because it comes to be the closest to the ideal of circle (Perraton and Baxter, 1974).

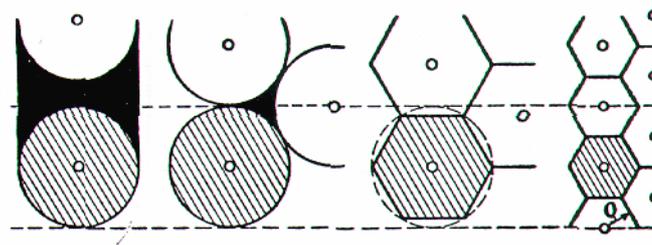


FIGURE 2.3 – Hexagonal nests of Lössch’s economic regions theory.

SOURCE: Merlin (1973, p.135).

In the sequence of these simple economic-oriented accomplishments in urban modeling, a new generation of computerized urban models came into play in the late 1950s and

early 1960s, immediately following the advances in computational facilities at that time and the advent of the Quantitative Revolution (a scientific revolution committed to introduce both rigor and quality into disciplines such as sociology, political science and urban studies).

The pioneer developments in urban modeling arose almost exclusively in North America, where the increasing car ownership during the 1940s and early 1950s led to the realization that cities with their traditional physical form could simply not cope with the new mobility needs. The first transportation studies involved forecasts of future trip generation and its spatial distribution and to meet these needs, trip generation was modeled using linear regression analysis, and distribution was modeled using the 'gravity model', so called because of its analogy with Newton's Law of Gravitation. These initial studies yet neglected many important questions concerning land use, but the then increasing academic and professional awareness towards the interrelationship between traffic and land use enabled the construction of land use models already by 1960 (Batty, 1976).

Parallel to these developments in transport planning, two important research projects concerning urban and location economics also had a meaningful impact on further modeling achievements of that time. The first of them is the more theoretically oriented economic model of residential location proposed by Alonso (1960), who took Von Thünen's work a stage further by setting the whole model within the microeconomic theory of consumer behavior based on utility maximization, where the consumer defines his/her place of living in function of trade-offs between housing and transport costs. The second project of impact is the intra-urban location model designed by Wingo (1961), also based on Von Thünen's work, which integrated detailed transport costs and explained population density. Relatively short before, Clark (1951) in a similar residential location model also gave an extension to Von Thünen's theory, describing the distribution of residential densities from the center of a town as an exponential which decays with distance (Equation 2.1).

$$D_j = A \exp(-\beta d_j) \quad . \quad (2.1)$$

where, D_j is the residential density at place j ; A is a constant; d_j is the distance to the town center and β is a parameter.

Some examples worthy of mention among these first modeling achievements are the Greensborough model (Chapin and Weiss, 1962) and the Baltimore and Connecticut models (Lakshmanan, 1964, 1968), all of them based on linear statistical techniques and employing a somewhat inductive approach to modeling with a little 'a priori' theory. On the other hand, inductive non-linear models have also been constructed, such as the Delaware Valley (Penn-Jersey) Activities Allocation model (Seidman, 1969).

Many models, however, were built around the gravity model, suggesting a more deductive approach in which specific mechanisms at work in the urban system were simulated. The Pittsburgh model (Lowry, 1964) and its successors are good examples of the gravity modeling approach, which was responsible for producing the most successful urban models during this time (Batty, 1976).

The Lowry model organizes the urban space-economy into activities (population, service employment and basic employment) on the one hand, and land uses (residential, service and industrial) on the other. It then allocates these activities to zones of the urban region. Population is allocated in proportion to the population potential of each zone, and service employment in proportion to the employment or market potential of each zone.

Having located the various activities in accordance with predetermined constraints, the model also tests the predicted distribution of population against the distribution used to compute potentials to find out whether the two distributions are coincident. This is necessary to secure consistency between these distributions because the model uses distributions of population and employment to calculate the potentials which indirectly affect the predicted location of these same variables. This is done by feeding back into the model predicted population and employment and reiterating the whole allocation procedure until the distributions input to the model are coincident with the outputs. A diagrammatic interpretation of these sequential operations is seen in FIGURE 2.4.

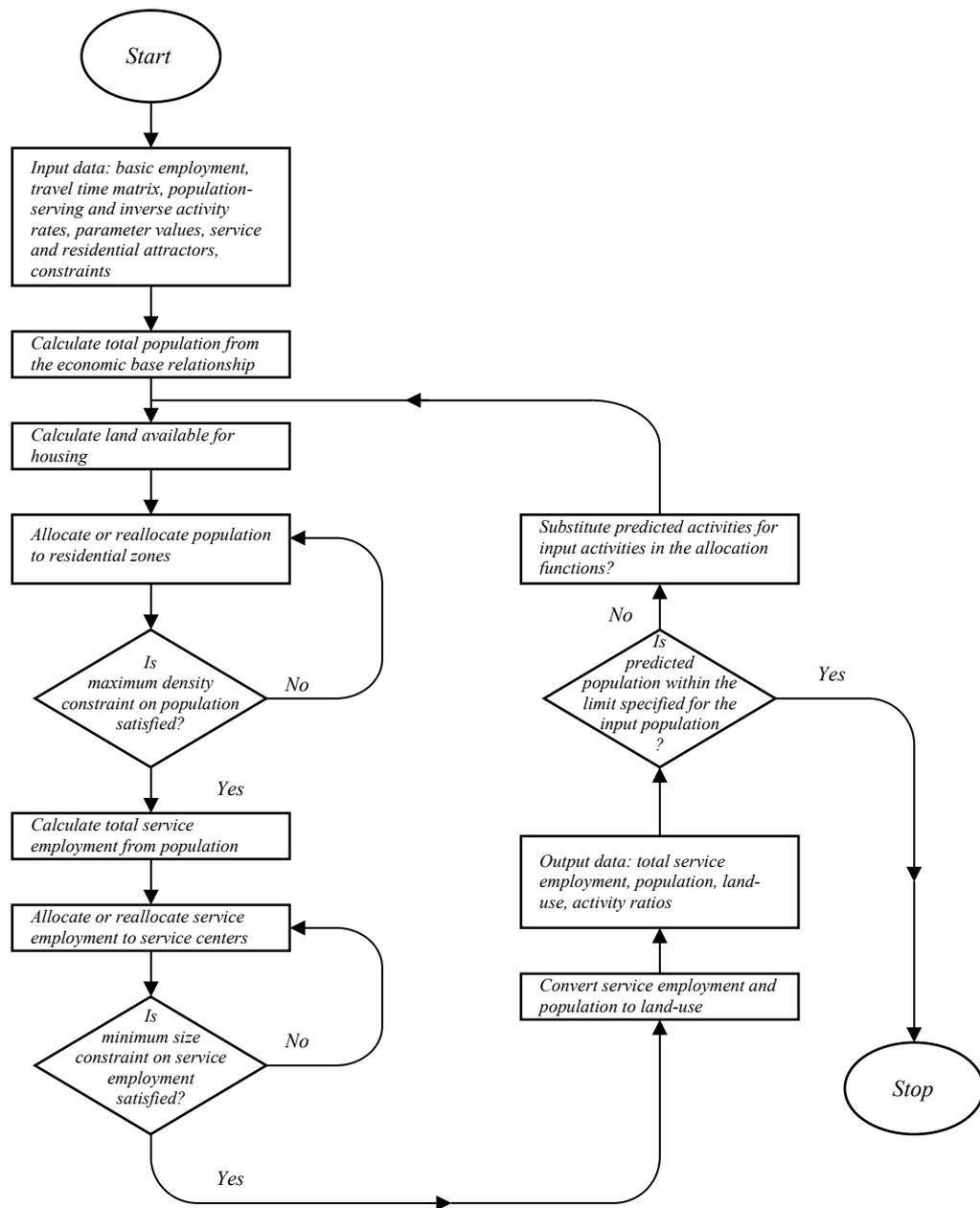


FIGURE 2.4 – Generalized flow chart of the Lowry model.

SOURCE: Batty (1976, p. 61).

Many are the drawbacks of these early models, and incisive critics have been directed to them, particularly in questions related to their size. Some of the models were so ambitious regarding their scope, data required and computer time and capacity needed,

that several were either abandoned or drastically reduced. Another problem concerns the fact that some models, although based on fairly well defined formal structures had to rely upon relatively sparse theory, thus often appearing arbitrary and mechanistic in structure. A third critic relates to the failure of many modelers in acknowledging inherent contextual limitations of the modeling object, in view of the complex set of variables intervening on urban systems, which cannot be analyzed simultaneously but just one at a time. And finally, most of these early models could only describe the urban structure at one cross-section in time, or at best, compare these static structures incorporating some long-term and often imputed equilibrium, what rendered them mere simulations of the static observable structure of cities (Batty, 1976).

2.2.3 Early Dynamic Models of Urban Land Use Change⁴

In an effort to overcome the shortcomings of the first generation urban models, a new bunch of modeling accomplishments commits itself, amongst other things, to work on a dynamic basis. Not all of these new models were fully dynamic, so it is important here to draw attention to some basic definitions in terms of dynamic modeling.

According to Wegener et al. (1986), a model is called *dynamic* as opposed to static, if it has an explicit time dimension, if its inputs and outputs vary over time, and if its states depend on its earlier states. A rudimentary form of dynamic model is a *comparative statics* one, which attempts to represent the static structure of urban systems at one cross-section in time without recourse to any explanation of the changes in structure over time which constitute system behavior⁵ (Batty, 1976).

A sequence of comparative statics models is called a *recursive* model, in which the end state of one time period serves as the initial state of the subsequent one (Wegener et al.

⁴ This section is fundamentally structured on the work of Batty (1976), for his book consists in the best literary revision on urban models of the early generations, presenting a broad overview of representatives belonging to this class of modeling endeavors in a concise and objective way.

⁵ According to the Systems Approach (Bertalanffy, 1951), the system structure cannot be interpreted without knowledge of system behavior and vice-versa. The Systems Approach postulates the idea of systems being described in terms of structure and behavior, in terms of input and output, and the notion of purposeful control of such systems in terms of negative and positive feedbacks.

1986). Batty (1976) makes a further distinction, stating that a model is called *quasi-dynamic*, if it contains static parts in a dynamic framework.

Perraton and Baxter (1974) present a comparative statics model analyzing the repercussions of alternative arrangements of transportation networks, land development regulations and public service facilities on the distribution of population. The work of Butler (1969) is a good example of a recursive model, where residential redistribution is approached by breaking down this process into two sub-models - which households move and where - with the new residential distribution being used to assess which families move in the next time period.

Quasi-dynamic models are usually built upon the spatial interaction location model as conceived by Lowry (1964). In this case, the spatial interaction model is generalized by incorporating zonal or network capacity or supply constraints in a multi-activity framework by means of constrained non-linear optimization. Good examples of this category of models are Boyce (1977), Coelho and Williams (1978), Leonardi (1981), etc.

Among the first dynamic models of linear type is the one proposed by Czamanski (1965), who has applied a simple time-oriented economic base model to the Baltimore region. This is a four-equation second-order model which can be stated as

$$P(t+1) = a_1 + b_1E(t-1), \quad (2.2)$$

$$S(t+1) = a_2 + b_2P(t), \quad (2.3)$$

$$E^b(t+1) = a_3 + b_3X(t+1), \quad (2.4)$$

$$E(t+1) = E^b(t+1) + E^l(t+1) + S(t+1), \quad (2.5)$$

where P is population; S is service employment; E^b is derived basic employment; X is exogenous basic employment and E^l is locationally-oriented basic employment which

includes X ; a_1 , a_2 , a_3 and b_1 , b_2 , b_3 are parameters to be estimated; the time notation is self evident. The model provides a simple approach to generating urban activities although there is no spatial dimension.

Another example of dynamic modeling is the EMPIRIC model, which is based on a system of first-order linear difference equations referring to different zones and activities. This model was designed by Hill (1965) for the Boston Regional Planning Project and underwent successive revisions that resulted in several versions. In contrast to Czamanski's model, the EMPIRIC model is spatially based and recognizes the simultaneous nature of urban interrelationships, adopting for this end formal solution methods, such as those based on two-stage least squares. The time interval adopted is ten years, and consequently the emphasis upon dynamics is implicit rather than explicit.

A basic problem with these aforementioned linear models concerns the fact that they do not attempt to distinguish the *mover pool* (the bunch of activities that are relocating in the city) from the stayers. Furthermore, there is a rather inductive bias in that the emphasis upon explanation is completely statistical, and being so, there are few guiding principles in the choice of time interval.

Defining the length of time interval is of great importance in dynamic modeling. For Forrester (1969), the simulation period should be short enough not to influence the behavior of the model, in particular, it should not be used to introduce implicit lags. Batty (1976) citing Broadbent (1969) states that the time interval should be small enough to detect the time-varying phenomena of interest; if the time interval is too large, then the dynamic model will become trivial and thus decomposable into a series of comparative static models, one for each time period.

Forrester (1969) suggests that the period length should be half or less of the shortest delay⁶ present in the system. For him, delays are essential for understanding complex

⁶ Delay is the time taken between the introduction of a certain stimulus and the manifestation of its impacts on the system under consideration.

systems, and this in line with the assertion of Bifurcation Theory⁷ that already small changes in the constellation of system variables can lead to significantly different paths of system behavior. Following this rule, most dynamic urban models with period lengths of five or more years are grossly inadequate to capture the dynamics of urban change (Wegener et al. 1986).

The Time-Oriented Metropolitan Model (TOMM), developed by Crecine (1964) for the Pittsburgh Community Renewal Program, attempted to turn an initial simple static model based upon the original Pittsburgh model (Lowry, 1964) into a more complex dynamic model by specific consideration of system behavior. An important distinction was made in the model between new locators and relocators (*mover pool*). TOMM has been improved in several ways since its first attempt. In its second version, a time interval of two years was used and a more realistic formulation of the measure of locational attraction, incorporating site rent, amenity and transport cost, was provided (Crecine, 1968). In its third version, questions of dynamics and mover behavior have been approached.

Even though this model presents some meaningful advances, there is no consistency between mobile activities and their relationship through the economic base. Although TOMM is organized around the concept of a mover pool, new locators and relocators are not separated out, thus the equilibrium properties of the model are difficult to trace (Batty, 1976).

A set of techniques designed for simulating industrial processes in firms, called collectively Systems Dynamics, has been employed in the modeling of hypothetical urban systems and world systems, adopting intervals of five years or more over a 250-year period (Forrester, 1969, 1971). This technique has its foundations in ideas from control engineering; the concept of system structure and behavior is conceived in terms of levels of stocks which are progressively altered through time by rates of change which are affected by positive and negative feedbacks within the system of interest.

⁷ The Bifurcation Theory is commonly mentioned in association with the Catastrophe Theory, for their interrelationships. The latter is basically concerned with processes undergoing sudden changes and consequently presenting a discontinuous behavior (Bak et al. 1989).

Many of these models are based on the notion that a system is fundamentally constrained by some fixed limit on resources, which affect the growth of the system through time. Typically, such a system grows explosively or exponentially at first and then as its resource limit is neared, the growth is damped and an equilibrium condition is eventually reached, usually with some oscillation around the steady state (FIGURE 2.5).

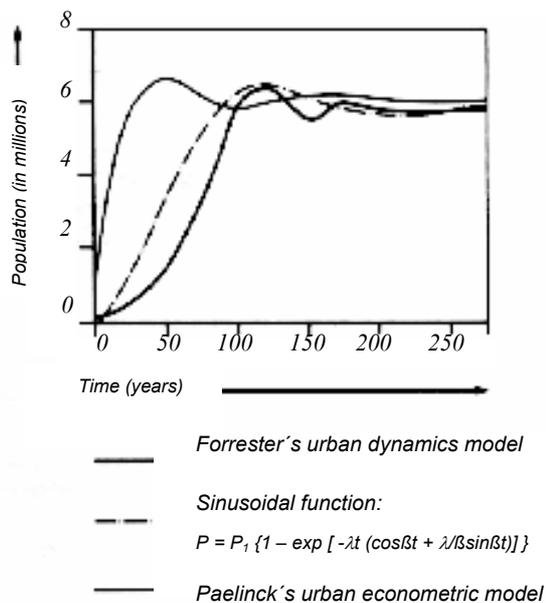


FIGURE 2.5 – Dynamic behavior curves in fixed resource systems: Forrester's urban dynamics model and Paelinck's urban econometric model.

SOURCE: Batty (1976, p. 305).

Forrester's models are not spatial and do not recognize that the structure of activities in a city can be explained in terms of spatial interaction (Batty, 1971). They also completely disregard the incorporation of well-demonstrated and accepted urban theory, and many of their hypotheses are either unproven or untestable (Batty, 1976).

Nevertheless, these models have been applied to two real situations. The first is the case of Harris County, Texas (Porter and Henley, 1972), which is not of much interest. The second case, which revealed to be a very promising experiment, concerns the Venice subregion (Costa and Piasentin, 1971). Venice was divided into three zones – Centro

Urbano, Estuario and Terra Ferma – and their model simulates the growth in population, employment and housing from 1951 to 1971. The appropriateness of this application lies in the fact that Venice subregion is highly constrained in spatial terms and thus quite unsuitable for activity allocation models.

Another class of dynamic urban models has been devised following the work of economists such as Paelinck (FIGURE 2.5), based on the specification of urban dynamics in terms of difference equations relating macroeconomic phenomena (Paelinck, 1970; Blokland et al. 1972). One of the features of models like these whose spatial component is implicit rather than explicit, regards the fact that many more hypotheses are needed for validation than in their spatial counterparts. Although these models tend to be richer in detail, they are frequently more difficult to calibrate to any real situation. On the other hand, their hypotheses are usually grounded on well established economic theories, and the estimation of their parameters is a quite well developed area in Econometrics, making them easier to apply to real situations than the Systems Dynamics models (Batty, 1976).

Batty (1971) developed a dynamic model, applied to the region of Reading, UK, that was mathematically formulated as a system of differential equations relating population, service employment and basic employment together through time by the economic base hypotheses, and through space by the gravity models which simulate the flows between the various activities. One of the main underlying assumptions is that the city is always in disequilibrium for at any instant in time there are still repercussions in the model which have not worked themselves out.

This model has been further detailed (Batty, 1976), differentiating locators from relocators. The spatial interactions derived in the model are oriented around the location of activities, divided into three major types: residential population, services and basic employment. Services are subdivided into consumer and producer-oriented groups and basic employment is broken down into employment dependent on existing employment, and unique locators, whose location cannot be forecast by the model.

As Batty (1976) explains, the unique locators provide the external stimuli to the model for although the total level of basic employment is exogenous, the other category of basic employment is distributed spatially using a linear model. Both population and service employment are allocated using production and attraction-constrained gravity models of the type derived by Wilson (1970). These models endeavor to simulate, though very coarsely, an equilibrium between demand and supply of activities while in the case of the residential location model, a further submodel has been designed to deal with the supply of residential land and floorspace. FIGURE 2.6 illustrates a schematic form for the model and the main relationships between the sectors.

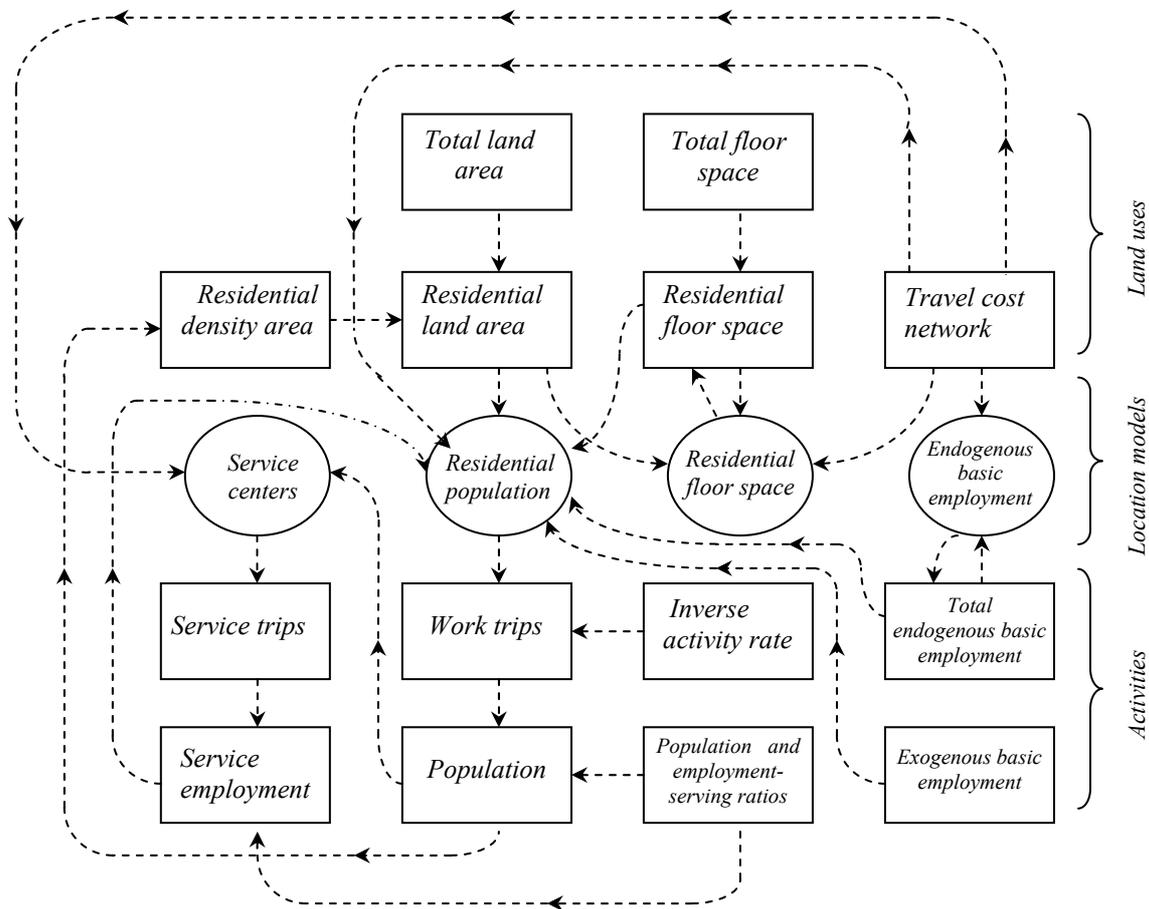


FIGURE 2.6 – Structure of activities and land uses in the model of urban dynamics for Reading, UK.

SOURCE: Batty (1976, p. 315).

Another way to arrive at fully dynamic urban models is to interpret the convergence to equilibrium of static models as an adjustment process over time. This permits the investigation of the time path of an urban system as a sequence of equilibrium-seeking steps which may or may not arrive at equilibrium depending on exogenous influences.

Beaumont et al. (1981) proposed a model in this sense, whose basic idea is very simple: it starts from the production-constrained shopping-trip model of the Lakshmanan-Hansen (1965) type and interprets its column sums, depending on their sign, as unsatisfied demand or excess supply and uses this information to drive the growth or decline of retail facilities (Equation 2.6).

$$W_j^\circ = \varepsilon \sum_i \frac{W_j^a \exp(-\beta_l c_{ij})}{\sum_j W_j^a \exp(-\beta_l c_{ij})} e_i P_i - k W_j \quad W_j, \quad (2.6)$$

where the W_j = retail facilities in retail zones j ; e_i = shopping expenditures of population P_i of zones i in j ; k = the costs of supplying retail services in j ; ε is an elasticity parameter determining the speed of the adjustment process; and β_l = the parameter of the deterrence function. In the absence of external stimuli, the model produces logistic growth of shopping centers up to a spatial equilibrium, but due to its non-linearities it displays a variety of bifurcations for different combinations of its parameters α , β_l , and ε , and for different initial distributions of P_i and W_j .

Another dynamic model envisaged to produce bifurcations is that of the Brussels group (Allen et al. 1981). Their model approaches the issue of dynamics in urban systems from a different direction. It is based on the concept of self-organization through random perturbations found on the molecular or genetic level in physical or biological systems. Allen and his team developed this set of ideas further in what they named “general dynamic model of the spatial evolution of urban systems”, which lied upon sophisticated concepts like open and adaptive systems, chaos and complexity (Allen et al. 1986).

The Brussels model is more elaborate than the previous one as it has four industrial sectors and, instead of housing, two population groups. The way for predicting growth and decline of retail facilities W_j in retail zones j is given below (Equation 2.7):

$$W_j^\circ = \varepsilon \sum_i \frac{[A_j (k + c_{ij}) - \beta_i]^a}{\sum_j [A_j (k + c_{ij}) - \beta_i]^a} e_i P_i - (k + c_{ij}) W_j \quad W_j, \quad (2.7)$$

where A_j is the attractiveness of zone j , and the remaining notations have the same meaning as in Equation 2.6.

The two models have different formulations of attractiveness. They also differ in the way they produce bifurcations. While in the first case bifurcations are systematically explored by parametric variations, in the Brussels model they are generated by random disturbances, although the dynamic behavior produced by the two models are very similar.

As in the precedent examples, these two models received several critics, but an important one concerns the fact that, like all spatial interaction models, they predict a slow process (establishing retail facilities) from a fast process (shopping trips). Wegener et al. (1986), the authors of these critics, proposed instead a dynamic urban model observing different time scales of urban change, i.e. a model being multilevel in its temporal dimension.

Wegener and his team classify urban change processes into three levels according to their speed of occurrence. Slow processes of change (Level 1) regard the construction of transport infrastructure, public industrial plants and public housing. These investments tend to be durable and involve the longest time lags between planning and completion. Medium-speed processes (Level 2) refer to economic, social and technological changes (aging, death, formation of new households, change of job, etc.), since these changes are visible only on a medium-term scale, while fast processes (Level 3) concern mobility, for they are the most volatile occurrences of urban change.

This multi-level temporal model comprises three modules (choice, transition and exogenous modules), and each of the categorized urban changes are assigned to one of these modules. Public investments are always included in the exogenous module; aging, death, retirement and alike processes, for being non-choice changes, are added to the transition module; and the remaining changes, to the choice module. These modules are linked by a file system and driven by a scheduler.

The model structure consists of a temporal framework of nested simulation periods of different lengths as multiples of a very short incremental period, a month or a quarter. If a process module is activated by the scheduler, it is told the number of incremental units of time it is supposed to remain active. The process module takes this information to calculate the number of microevents to be processed during that time interval. If the number is large, the process module can increase its sampling factor or resolution and thus save computer time. Given the model's detailed approach of the temporal dimension and microeconomic behavior, it was not fully implemented, presenting though most of its components operational. This model has been applied for the simulation of long-term change processes in the urban region of Dortmund, Germany (Wegener et al. 1986).

And to conclude this section, it can be finally stated that even though considerable refinements were introduced in this early generation of dynamic models in terms of coping with the complex and sometimes recursive spatial interactions among different activities in a city, of adding the (multi-level) time dimension in the quantitative analyses and employing sophisticated mathematical and theoretical tools - e.g. differential equations, Catastrophe and Bifurcation theory, etc. -, these models remained fairly non-spatial, especially in the sense that their results could not be spatially visualized.

2.3 Spatial Dynamic Models of Urban Land Use Change

2.3.1 Space and Time in the Present Work

Considering that the “purpose of any model is to simplify reality” (Batty, 1976, p. 353), discretization will be the tone of this work regarding both its space and temporal dimensions. Following this premise, space will be taken as an artifact represented by a regular two-dimensional grid framework, consisting of homogeneous square units – the cells – forming the so-called cellular space. Time, on its turn, will be considered upon an absolute point of view as an independent one-dimensional entity, and in the particular case of this research, will be regarded as made of successive discrete lag units, where each unit may account for a set of years in some cases and in others may be yearly defined.

It is important at this point to clarify that the adoption of a metric (Euclidian) space will not restrict the possibility of coping with non-contiguous spatial interactions. As it will be further demonstrated in Chapters 5 and 6, the empirical statistical methods employed to parameterize the land use change simulation models will be based in certain cases on ranges of distance to given physical variables, and in some situations, the farther the range, the greater its influential contribution for a certain type of land use transition.

Even in the cases where computational implementations are based on theoretical proximal models of space, it becomes evident that the adopted spatial framework lies within the realm of Euclidian spaces. Examples to illustrate this are the graph-CA based models (O’Sullivan, 2000, 2001a, 2001b), where graphs (inter-connected linear structures) operate over a cellular automata model to redefine functional relations as being those not exclusively derived from spatial contiguity, and the geo-algebra model (Couclelis, 1997; Takeyama, 1997; Takeyama and Couclelis, 1997), in which by means of several raster operations, executed within a CA environment, taking into account connectivity matrices, relational and metarelational maps, non-local influences are assessed and incorporated into the dynamics of the urban system at issue.

It is essential to highlight here the decisive role that the formulation of space as being a cellular entity represented to the endeavors in the field of urban simulation. In fact, this conception drastically impacted the whole scenario of dynamic modeling in the most diverse branches of scientific knowledge.

2.3.2 Cellular Automata and the Advent of Spatially Explicit Models of Land Use Change

2.3.2.1 Cellular Automata: Definition and Properties

According to what was exposed in the section on brief historical perspectives of models of urban land use change, namely Sections 2.2.1 through to 2.2.3, the urban models developed from the 1950s until the mid 1980s, in a fairly general sense, did not take the spatial dimension into account. When this happened, the urban space was decoupled into units (usually zones defined according to trip generation, census districts or other alike criteria), but the output of such models could not be spatially visualized. In fact, effective advances in the spatial representation of urban models occurred only by the end of the 1980s, when cellular automata (CA) models started to be largely applied.

Stephen Wolfram, one of the most renowned theoreticians on cellular automata defines them as:

“... mathematical idealizations of physical systems in which space and time are discrete, and physical quantities take on a finite set of discrete values. A cellular automaton consists of a regular uniform lattice (or ‘array’), usually infinite in extent, with a discrete variable at each site (‘cell’). The state of a cellular automaton is completely specified by the values of the variables at each site. A cellular automaton evolves in discrete time steps, with the value of the variable at one site being affected by the values of variables at sites in its ‘neighborhood’ on the previous time step. The neighborhood of a site is typically taken to be the site itself and all immediately adjacent sites. The variables at each site are updated simultaneously (‘synchronously’), based on the values of the variables

in their neighborhood at the preceding time step, and according to a definite set of ‘local rules’ (Wolfram, 1983, p. 603)”.

Cellular automata (CA) models have found applications in diverse fields, ranging from statistical and theoretical physics to land use and land cover change, traffic engineering and control, diseases spread, behavioral biology, amongst others. Wolfram’s (1983) paper refers to near 50 other papers concerned with possible applications of cellular automata.

John Conway’s Game of Life (Gardner, 1970), or ‘Life’ as it became later known, immortalized the concept of CA. In Life CA, the cellular space is composed of a regular 2-D square lattice, where its cell’s neighborhood consists of its eight immediate neighbors in the lattice, that is, four orthogonal neighbors and four diagonal neighbors⁸. Any cell can be alive (‘on’) or dead (‘off’), and there are only two simple rules for cells becoming alive or dying. A cell which is not alive becomes alive if there are exactly three live cells immediately adjacent to it. A cell remains alive if there are two or three live cells adjacent to it, otherwise it dies. This means that a cell may die from isolation (less than two live cells adjacent to it) or overcrowding (more than three live cells in its adjacencies). Albeit the simplicity of the rules, this game can afford the generation of countless patterns of dynamic behavior.

In fact, the Life CA has been demonstrated to be rich enough to support universal computation (Berlekamp et al. 1982). It is thus evident, that even in the simplest CA, complex global patterns can emerge directly from the application of local rules, and it is this very property of emergent complexity that makes CA so fascinating and their usage so appealing.

Another very interesting property of Life was the possibility of presenting certain configurations that would be self-perpetuating. A very well-known example of such a configuration in Life is the ‘glider’. A glider is made of five live cells, organized as shown in FIGURE 2.7, at time step $t = 0$. The sequence of four time steps results in the

⁸ This eight-cell neighborhood may be referred to as the ‘Moore’ neighborhood. A neighborhood of the four orthogonally adjacent cells only is the ‘von Neumann’ neighborhood.

changes shown, at the end of which the configuration has moved one cell up and to the right. As O’Sullivan (2000) states, an interesting aspect of a glider is that its behavior can be described without any knowledge of the underlying rules.

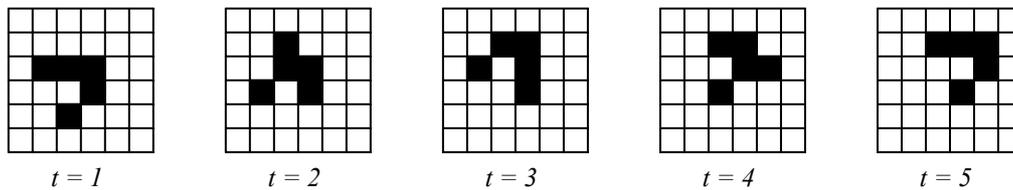


FIGURE 2.7 – The Game of Life ‘glider’ configuration – an example of emergence.

‘Live’ cells are in black.

SOURCE: O’Sullivan (2000, p. 58).

Despite the fact that CA can cope with emergence as seen in complex phenomena, they own a great flexibility for dealing with the most diverse dynamic processes, and reveal an amazing operational simplicity amongst other advantages, they are in principle limited as a result of constraints imposed by most of their theoretical assumptions.

In response to this, Couclelis (1997) proposes a set of improvements to be incorporated to cellular automata models, which represent a technical possibility to relax any or all of the assumptions of standard CA (FIGURE 2.8). She devised these proposals mainly as a way to meet the demands of urban and regional modeling.

According to her opinion, space no longer needs to be homogeneous either in its properties or its structure; neighborhoods need not be uniform across the space, and transition functions need not be universal (that is, equally applicable at every point). For her, distance-decay effects could be built into CA neighborhoods, the transition rules could be probabilistic rather than deterministic, and variable time steps could be used to fit some external schedule.

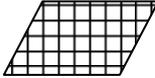
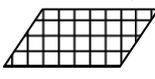
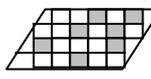
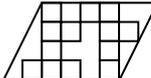
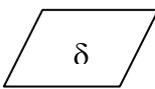
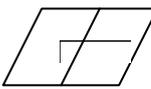
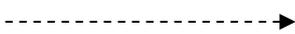
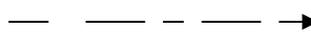
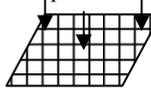
Space	Structure	regularity 	irregularity 
	Properties	uniformity 	non-uniformity 
Neighborhood		stationarity 	non-stationarity 
Transition function		universality 	non-universality 
Time		regularity 	non-regularity 
System closure		closed 	open 

FIGURE 2.8 – Common generalizations of cellular automata (CA).

SOURCE: Couclelis (1997, p. 168).

Some theoreticians of CA would argue that in this case the cellular automaton paradigm no longer holds, especially due to relaxations in the notion of local neighborhood as well as transition rules, and that these adapted models should be preferably called cell-space models (Albin, 1975; Batty, 2000).

2.3.2.2 Spatially Explicit or Spatial Dynamic Models of Urban Land Use Change

According to what was already stated in Section 1.1, CA were somehow implicit in the early generation of computer models in the 1960s, with the experiments of Chapin and Weiss (1968) and of Lathrop and Hamburg (1965), carried out for North Carolina and western New York State respectively. In the 1970s, Tobler (1970), influenced by the theoretical quantitative geography, first proposed cell-space models for the development of Detroit. Shortly after, in 1974, he formally began to explore the way in which strict CA might be applied to geographical systems, what resulted in his famous paper

“Cellular geography” (Tobler, 1979). Finally, in the late 1980s, CA started to be widely applied to urban and regional issues, also impelled by the parallel development in computer graphics and in theoretical branches of complexity, chaos, fractals and alike (Batty et al. 1997).

The 1990s experienced successive improvements in urban CA models, which started to incorporate environmental, socioeconomic and political dimensions, and were finally successful in articulating analysis factors of spatial micro and macroscale (Phipps and Langlois, 1997; White and Engelen, 1997; White et al. 1998).

Cellular automata can be regarded as a category of spatially explicit (or more generally, spatial dynamic) models. The term “spatially explicit” is not new but it is still recently employed to characterize models which attempt to describe and predict the evolution of environmental attributes in sub-units of area with distinct location and configuration (Baker, 1989). Briassoulis (2000) further subdivides these models into fully spatially explicit (geo-referenced) and incompletely spatially explicit (non-geo-referenced).

CA models may regard both theoretical and practical applications, where the formers concern abstract exercises, and the latter ones, experiments dealing with real case studies. There are now some twenty or more applications of CA to cities (Batty, 2000), including a vast repertoire at both thematic and methodological levels.

The first CA models applied to urban studies were not rarely based on very simple methodological procedures, such as the usage of neighborhood coherence constraints (Phipps, 1989) or Boolean rules (Couclelis, 1985) for the transition functions. Further on, successive refinements started to be incorporated into these models, like the adoption of dynamic transition rules (Deadman et al. 1993), which could change as conditions and policies within the township under study changed.

Other examples in this direction are the work of Wu (1996), who conceived transition rules to capture uncoordinated land development process based on heuristics and fuzzy sets theory, and the work of Ward et al. (1999), in which transition rules are modified in

accordance with the outcomes of the optimization of economic, social and environmental target thresholds associated with sustainable urban development.

Further improvements concern the embodiment of theoretical microeconomic approaches into CA, which led to the development of agent-based or multi-agent-based (MAS) systems. Just to cite a few examples in this sense, the work of Semboloni (1997), describing the patterns of major cities and surrounding towns according to four actors represented by different categories of population and activities, the works of Sanders et al. (1997) and Torrens (2003) as well as the works of Batty et al. (1998) and Schelhorn et al. (1999) on pedestrian flow in what is called 'microscopic' or 'atomic'⁹ models are to be mentioned.

CA transition functions have also been enhanced by the incorporation of decision support tools, including AHP or analytical hierarchy process-based techniques, what has been strongly enabled by the linkages between CA and GIS (Engelen et al. 1997; Wu, 1998). Besides supporting CA internal operations (Clarke and Gaydos, 1998; Batty et al. 1999; Li and Yeh, 2000), GIS have been as well useful in implementing cellular automata devices based on proximal models of space (Takeyama and Couclelis, 1997) and in articulating spatial analysis factors of micro and macroscale (Phipps and Langlois, 1997; White and Engelen, 1997; White and Engelen, 2000).

An illustrative example of CA models integrating analysis factors of spatial micro and macroscales is the work of White et al. (1998), in which demand for residential land use is estimated through a social subsystem that takes into account interregional migration flows, and where demand for economic activities (industrial, commercial, services) is obtained by means of regionalized subsystems that evaluate the performance of different economic sectors, providing parallel data on employment opportunities, which will be again used to reckon residential demand. This model is envisaged to meet the diverse land use demands also considering the environmental carrying capacity of concerned sites (natural subsystem) as well as the constraints imposed at the micro or

⁹ 'Microscopic' or 'atomic' models are those concerning objects represented at entity-level resolutions: pedestrian, households, vehicles, houses, etc. (Benenson and Torrens, 2003).

local level by functional, physical, institutional and infra-structural aspects (FIGURE 2.9).

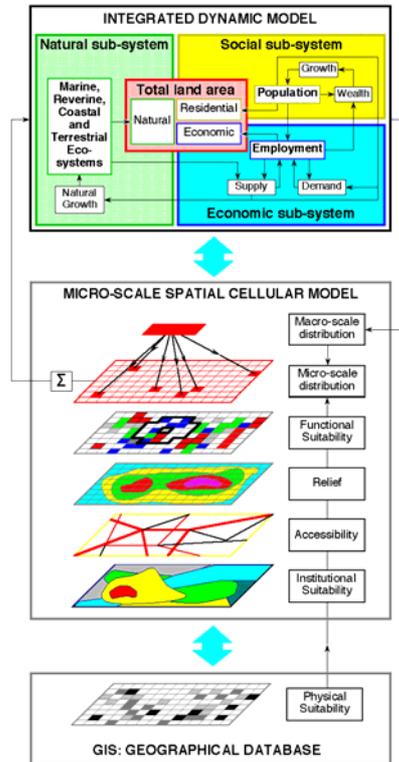


FIGURE 2.9 – A CA model integrating spatial micro and macroscales.

SOURCE: White et al. (1997, p. 237).

Phipps and Langlois (1997) also succeeded in estimating the impacts of analysis factors from the macro or national level, like planning policies restraining foreign migration, at the local level. Their data model conceives a 'Global Systems Dynamics (GSD)' module, which receives an input reflecting the spatial pattern macrostate at each time unit, and whose output also at each time unit alters a Markovian spatial transition probability matrix (STPM). This model has not been implemented with the aid of GIS, but through parallel computing (FIGURE 2.10).

Theoretical advances in the field of complex systems have been also added to cellular automata through the seminal work of Wolfram (1984). Research papers dealing with

fields related to the theme of complexity, like chaos¹⁰, fractals¹¹ and self-organized criticality (SOC)¹² started to be produced after Wolfram's work and are still recurrent in the CA scientific community (Couclelis, 1988; Batty and Longley, 1991; White and Engelen, 1993; Batty and Longley, 1994; Portugali et al. 1997; Portugali et al. 1999; Sobreira and Gomes, 2001; Batty, 2003).

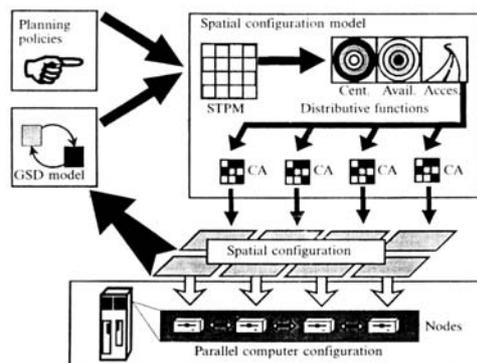


FIGURE 2.10 – A CA model incorporating macroscale variables.

SOURCE: Phipps and Langlois (1997, p. 196).

Leading theoretical progresses in the broader discipline of artificial intelligence (AI), such as expert systems, artificial neural networks and evolutionary computation, have been lately included in the scope of CA simulations. Moore (2000), citing Weiss and Kulikowski (1984), defines expert systems as “computer systems that advise on or help solve real-world problems requiring an expert’s interpretation using a computer model of expert human reasoning reaching the same conclusion the human expert would reach if faced with a comparable problem.” Neural networks, on their turn, attempt to simulate human reasoning (Moore, 2000) offering fault-tolerant solutions. According to Fischer and Abrahart (2000), these mechanisms are able to learn from and make decisions based

¹⁰ Chaos is formally defined as the study of complex non-linear dynamic systems, provided that the term complex involves a multitude of intervening variables and their interrelationships in a given phenomenon (Thompson, 1997).

¹¹ Fractals are object of any kind whose spatial form is nowhere smooth, hence termed ‘irregular’, and whose irregularity repeats itself geometrically across many scales. Fractals can be called the geometry of chaos (Batty and Longley, 1994).

¹² Self-organized criticality (SOC) refers to systems where its individual components follow their own local dynamics to a critical state where the emergent dynamics are global and communication flows freely throughout the entire system. Complexity is the consequence of self-organized criticality (Bak et al. 1989).

on incomplete, noisy and fuzzy information. And finally, evolutionary computing methods are an extremely flexible and intelligent search procedure which mimic the known mechanisms of natural evolution (Diplock, 2000).

As stated in Almeida et al. (2003), methods recently embedded in CA models like those employing contemporary pattern-fitting tools such as neural nets (Wu, 1998; Xia and Yeh, 2000) and evolutionary learning (Papini et al. 1998) are revealing to be amongst the most promising for the coming generation of urban CA modeling achievements.

2.3.2.3 Theory-Driven x Data-Driven CA Models of Urban Land Use Change

There is no consensus amongst researchers, specifically in the field of urban and general land use modeling, as to where the fine line between theory and models starts and ends, and even whether there is indeed this boundary line at all. For Batty (1976), “urban modelling is not just a reflection of urban theory formulated elsewhere; it is now an essential part of theory in the fields of urban economics, geography and planning.” This assertion was stated around the mid 1970s, when urban models were basically concentrated around econometric approaches, activities allocation and gravitational theories.

It is interesting to notice the shift in opinion in the course of the current decade, where the impacts of massive advances in computational tools, mainly grounded on logic and algorithmic routines, are pronouncedly felt. Torrens and O’Sullivan (2001) report that “one of the drawbacks of urban CA models is that in some cases they have done relatively little to inform theory. Claims are made that models explore various hypothetical ideas about the city, but the reported results are often more concerned with the fine details of model construction, at the expense of the theories that they set out to explore. Research in urban CA modeling is just that: research in modeling, and not research on urban dynamics and theory.”

Following an alike thought, Briassoulis (2000), who is concerned with general land use and cover change (LUCC) models, states that “although the use of theory in model building seems indispensable, of the several theories of land use change proposed, a

relatively small number has been used to support and guide operational model building. Some theories and models have been conceived simultaneously; hence, the use of the terms 'theory' and 'model' interchangeably...but the majority of theories are still without modelling (not necessarily mathematical) counterparts and the reverse is also true. Several models are devoid of theoretical foundations.”

The reasons for this gap in the relationship between theories and models are twofold (Briassoulis, 2000): a) the first regards the differing epistemological¹³ positions adopted by theory and model builders, where models usually move in the positivist¹⁴ tradition while theories cover a much broader spectrum of epistemologies; and b) a second reason concerns the fact that reality is highly complex, what leads to a dichotomy in the field of modeling: either the reduction and simplification of this real world results in a very crude representation of reality, or models endowed with a very complicated structure end up by being impossible to handle within the bounds of reasonable time and other resources to provide answers to practical problems.

This 'theory x modeling' hiatus finds exemplifying cases throughout the whole historical horizon of urban modeling: since the early generation of ambitious transport models in the mid 1960s (see Section 2.2.2, p. 49) that were abandoned or drastically pruned, until the relatively recent microeconomic approaches that attempt to be deeply acquainted with individual behavior (Wegener et al. 1986; Wegener, 2000), and thence render a fully computational implementation of its postulations nearly impossible. In summary, efforts to apprehend a given real context with an excessively great level of detailing have proved to be partially or totally impracticable from an operational point of view.

Theory-driven models can be understood as those whose bunch of assumptions, premises and derived equations (when existent) defining the system behavior is set 'a priori'. Data-driven models, on the other hand, rely on the available data to draw conclusions on the system 'a posteriori', employing statistical empirical methods or

¹³ Epistemology refers to the critical study of principles, hypotheses and results of established sciences, aiming at determining the logical foundations, value and objective scope of each one of them (Ferreira, 1986).

even simple deterministic rules. Theory-driven CA models of urban land use change may deal with both abstract/hypothetical cities and real cities, whereas data-driven models are strictly concerned with real urban areas.

Couclelis (1985), Portugali et al. (1997) and Semboloni (1997), just to mention a few examples, all work with theory-driven CA models for hypothetical cities. Batty and Xie (1997) also deal with abstract theory-driven CA models, but upon a more stochastic approach. Papini et al. (1998), on their turn, work with a theory-driven CA model, but applied to a real urban environment: the Rome metropolitan area. The work of Clarke et al. (1997), on the other hand, is a good example of a data-driven CA model, in which deterministic transition rules are established upon the observation of their study object, the San Francisco Bay area, and calibration is also executed taking into account the changing behavior of this real case study throughout the considered simulation time span.

The split between theory-driven and data-driven CA models should be carefully regarded, for it presents more blurry boundaries than strictly rigid delimitations. Some CA models could be called as hybrid ones, since their internal framework reconciles both theory and empiricism. The model proposed by White et al. (1998) exemplifies this situation, for it defines accessibility on the basis of gravitational theory, and estimates the suitability of a certain cell for changing its original land use into a new one through a probabilistic equation, empirically built according to a set of conditional factors.

The CA models to be presented in the following chapters could be considered hybrid as well. Although such models estimate their parameters directly from the data, there is indeed a robust theoretical body underlying the empirical statistical methods employed, concerning sets and probability theory, the Bayes' theorem, weights of evidence, multivariate linear and log-linear regression, etc. And also for the phenomena themselves being studied in this research - urban land use change – a whole set of economically-oriented theoretical arguments is adopted to explain them, taking into

¹⁴ Positivism regards a set of doctrines, which impose a scientific orientation in the philosophical thought, ascribing to science each and every progress in knowledge (Ferreira, 1986).

account notions of user's utility maximization and optimization of locational advantages.

In fact, choosing between a theory-driven or data-driven approach, or even adopting a hybrid solution, will always rely upon the modeler's sensible judgment, which will consider the particular profile of the study object as well as the specific purpose(s) envisaged by the modeling experiment. This ought to be done in order to avoid basic methodological problems in modeling identified by Harris (1975) already in the mid 1970s. For him, a great amount of urban models did not originate in the need to solve specific, practical urban problems, but instead, new approaches seemed to derive from the expansion, or refinement, of existing frameworks in the field.

And finally, the following recommendation raised by Oppenheim (1986) shall be left here as a guiding commandment for each and every model builder: "...modelling consists of developing a representation for a given situation and not of identifying a situation which fits a given model."

2.3.3 Why Using Cellular Automata for Urban Modeling

Cellular automata (CA) models have become popular largely because they are tractable, present an amazing operational simplicity, generate a dynamics which can replicate traditional processes of change through diffusion, but contain enough complexity to simulate unexpected and surprising change as observable in emergent phenomena. CA models are flexible in the sense that they provide a framework which is not overburdened with theoretical assumptions, and which is applicable to space represented as a raster or a grid. These models can thus be directly connected to raster data surfaces commonly used in GIS (Almeida et al. 2003).

CA are also regarded as flexible models because they can deal with the most diverse dynamics processes of the real world and can be coupled with a wide range of theories in the field of complexity: chaos, emergence, fractals, self-organized criticality, etc., as well as in the field of AI, like expert systems, artificial neural networks, evolutionary learning.

Considering all this, it is sensible to state that this kind of model representation is still one of the best techniques currently available to cope with the needs and interests of the queries raised by the present research investigation.

2.4 Conclusions

Chapter 2 was envisaged to provide an overview of both historical perspectives and current state of dynamic urban models. Although models in general have continuously been the target of severe criticism, mainly in view of their reductionism and constraints to fully capture the reality inherent complexity (Briassoulis, 2000), it can be argued that they ought to exist, for they offer an incomparable way of abstracting patterns, order and main dynamic trends of real world processes.

As stated in Batty (1976), "... pattern and order does exist and is fairly easy to identify at least on a superficial level in urban and regional systems. As to whether or not an individual agrees with the description of such patterns statistically is a matter of opinion, and ultimately of faith in the fundamental ideas."

Actually, urban models should be conceived, handled, applied and interpreted in a wise and critical way, so that modelers, practitioners, public and private decision-makers as well as citizens as a whole could take the best of what they can offer and sensibly acknowledge their limits. Again, all these ideas are very well synthesized by Batty (1976), who says that "there is a need for a more liberal perspective on the state of the art by all involved in urban modelling, thus fostering the view that models are aids to imagination in a wider process of design, problem-solving and decision-making in society at large."

CHAPTER 3

STUDY AREA

3.1 Location

The two case study cities, Bauru and Piracicaba, adopted as modeling objects of the present research are located in the western inland of the southeastern State of São Paulo, Brazil, within the Tietê river watershed. Bauru is located in the central portion of the State, and belongs to the Tietê lower midstream sub-basin, whereas Piracicaba is found eastwards of Bauru and belongs to the upper midstream stretch of the Tietê basin (FIGURE 3.1).

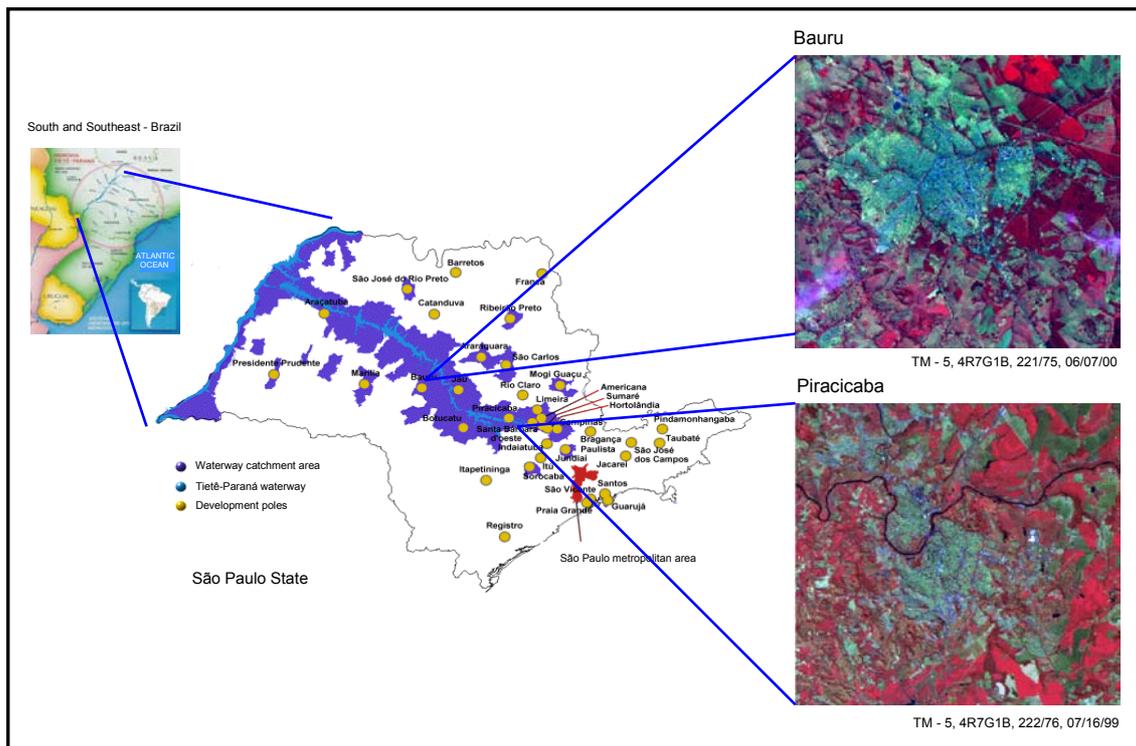


FIGURE 3.1 – Location of the case study cities: Bauru and Piracicaba, inland of SP.

SOURCE: CESP (1998); SEADE (1999); INPE (1999, 2000).

Bauru is delimited by the following geographic coordinates:

w 49° 09' 28"	s 22° 25' 03"	[Projection System: UTM Datum: Córrego Alegre.
w 48° 55' 43"	s 22° 15' 02"	

And Piracicaba, on its turn, by the geographic coordinates given below:

w 47° 43' 30"	s 22° 47' 59"	[Projection System: UTM Datum: Córrego Alegre.
w 47° 32' 48"	s 22° 39' 06"	

3.2 Historical Background of the Urbanization Process

A historical background of the urbanization process concerning the two research objects will not solely focus on these cities, but it will rather encompass the urbanization history of the Tietê river watershed and the inland of São Paulo State as a whole, for the understanding of their urban formation process cannot be detached from the historical context of the macroarea within which they are located.

3.2.1 The Mining Cycle (1580 – 1730)

The human occupation throughout the Tietê river valley dates back to the time Brazil was still a colony of Portugal. Between 1580 and 1640, the Piratininga village, as São Paulo city was known at that time, sheltered the departure point of many exploring expeditions (“*Bandeiras*”) along the Tietê river, which were in search of gold, precious stones and the conquest of new lands. Their members sought as well for Indians to be domesticated and enslaved, what brought about the extermination of several tribes in the inland of São Paulo State (Nóbrega, 1981).

According to Ohtake (1991), the mining cycle along the Tietê river actually started in 1640 and lasted until 1730, period during which the river acquired importance also as a transport route towards huge gold deposits recently discovered in the States of Goiás and Mato Grosso, central west of Brazil. Around 1720, the river passed through a new

period of expeditions, the so-called “*Monções*”, which lasted beyond the mining cycle. Such expeditions, in contrast to the “*Bandeiras*”, owned a commercial and colonizing character and had as its departure port the village of Araraitaguaba, current town of Porto Feliz.

In face of the great number of expeditions, villages and small settlements were founded along the Tietê river margins for the travellers’ provision. The Tietê has also worked as an induction vector of urban centers during the XVII and XVIII centuries (Ohtake, 1991).

3.2.2 The Sugar Cane Cycle (1730 – 1822)

By the year 1775, São Paulo State witnessed a late sugar cane apogee and of secondary importance in the colonial economic scenario. Beside the subsistence tillages and settlements spread throughout the Tietê river margins, large sugar-mill farms started to settle down from the first half of the XVIII century onwards, concomitantly with the decline of the mining cycle. The sugar cane culture found in the Tietê midstream valleys ideal conditions for its development, such as fertile soils, navigable stretches of the Tietê river for the sugar cane transport, the commercial activities brought with the “*Monções*”, the local socioeconomic conditions, and last but not least, governmental measures fostering exports and, consequently, the usage of the Santos port, on the Atlantic coast. In a very short time, the sugar cane farming had become the leading economic force of the São Paulo Captaincy (a political-administrative land division adopted during colonial times), what resulted in the emergence of important towns in this region (Ohtake, 1991).

3.2.3 The Coffee Cycle (1822 – 1920)

With the decline of the sugar cane culture by the time of the Empire demise (1822), the coffee ascends in the national scenario, becoming already in the first half of the XIX century the main Brazilian export product, and making São Paulo State early around 1890 the world leader in coffee production. Mostly covered by the red soil, which is highly suitable for the coffee culture, the Tietê river valley rendered favorable a quick

and concentrated occupation of the western portion of the State. Many of its cities developed in such a way that they became, as it is the case of “Itu”, political centers of major importance, which even sheltered some political movements that culminated with the Empire ruin and the institution of the Republican regime in 1889 (Ohtake, 1991).

Thus, the configuration of the current urban system in the inland of São Paulo State goes back to the time of the coffee cycle. The fast and intensive incorporation of lands to this culture, having as its fronts the railway lines, made way for the basic structure of the present urban network in this region (FIGURE 3.2). At the same time, some of the main cities of this network, mostly located either at the edge or crossing of railway tracks, started to play the role of regional centers, providing logistical support for the occupation process in the pioneer zones (Fundação Seade, 1992).

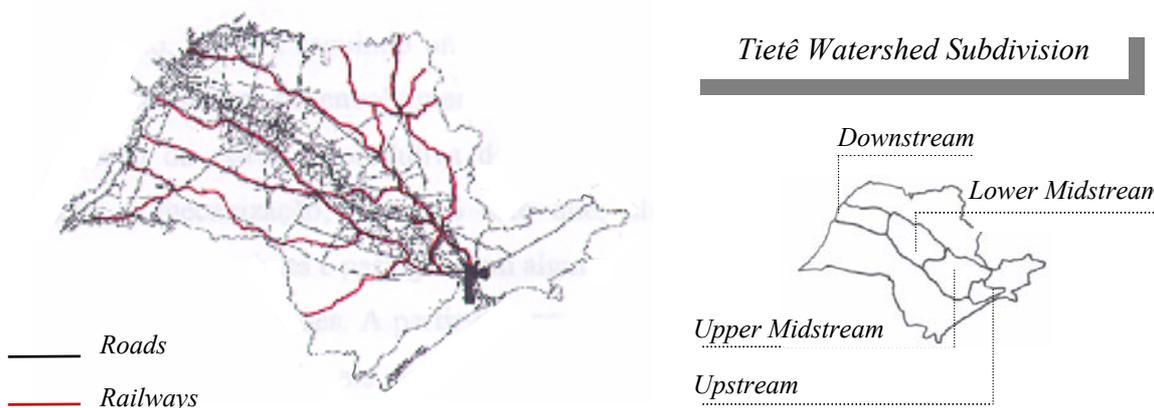


FIGURE 3.2 – Current transport matrix in São Paulo State and urban network (municipal division) throughout the Tietê river watershed.

SOURCE: CITP and CEPAM (1994).

The dynamics of the coffee economy had an essentially urban character. Industries, banks, offices, maintenance workshops, wholesale businesses, export and import commercial activities were required in this process, and the greater part of them were located in São Paulo city, seat of the State government and focal point of the transport system as well as of the distribution of the immigrants flow. The population of São

Paulo State rose from 1.4 million inhabitants in 1890 to 4.6 million in 1920 and 6.4 million in 1934. The occupation of the State inland meant not only the conquest of lands for the coffee farming but also the diversification of the agricultural activities in São Paulo (Fundação Seade, 1992).

3.2.4 The Industrialization Process and Economic Dynamics of the XX Century

3.2.4.1 Years 1920 –1929

The industrial activity, impelled by the coffee culture, tended to concentrate in the State capital, which already in the 1920s presented particular features that resembled a big city, sheltering the mass of urban workers employed either in the industrial sector or in the services activities. Agribusiness activities were mainly settled in the State inland, although some important industrial plants, notably textile, were also present in this region.

In the economic history of São Paulo State, the 1920s represent a transition stage. The coffee culture led to a surplus crisis that had to cope with the shrinking international demand from 1929 onwards. The industrial activities also passed through a similar process, experiencing a surplus crisis, especially in the cotton-textile sector. This transition culminated with a rupture represented by the world crisis of 1929 and the ensuing revolutionary movement of October 1930 (Fundação Seade, 1992).

3.2.4.2 Years 1930 – 1955

Recovery policies were implemented by the Brazilian national government during the years following the “world crack of 1929”, engendering the transition to a new accumulation standard: between 1930 and 1955, under the hegemony of the industrial capital, the industrialization process started to move forward, although with a restrained pattern, given that some sectors of the key industry were still incipient and would present a late development. From the 1950s on, the agricultural activities in São Paulo State underwent successive modernization processes (chemical fertilizers, pesticides, agricultural machinery), which in association with the gradual replacement of

permanent crops by annual crops and cattle raising in certain areas concurred for the rural migration towards the cities. Also from the 1950s on, inter-regional migration flows towards São Paulo started to become more intensive, which represented an important component for the urbanization process in the State at that time.

As explained by Fundação Seade (1992), these were the very processes that sustained urbanization during this period. The closer linkages between agriculture and industry as well as a greater diversity in their productive scope meaningfully impacted the tertiary sector in the urban areas, requiring as a counterpart the enlargement of the commercial and support services circuit. In the same way, the growth in population, mainly caused by migration displacements, increased the demand for personal and social services.

3.2.4.3 Years 1956 – 1969

By the mid and late 1950s, the intensity of the national economy growth, and the simultaneous advances in its productive structure, increasingly demanded the implementation of new economic branches and sectors. Already from the mid 1950s on, some sectors of the key industry gained a greater importance, and Brazil once again in face of the strong ties between its agriculture and industry, enhanced its presence in the international market of agricultural and agribusiness goods, strengthening technological progresses in the agricultural sector, incorporating new cultures and altering the relationships of the labor force engaged in primary activities (Fundação Seade, 1992).

These transformations in the primary and secondary productive sectors were reflected in the cities by an increasingly complex social structure. Typically urban activities as well as new ways of commerce, consistent with the new paradigms of mass consumption and changes in the individual habits, were greatly favored. At the national scale, these changes were markedly felt in São Paulo State, chiefly in the metropolitan area, but progressively, also in the remaining inland urban network.

3.2.4.4 Years 1970 – 1979

After the completion of the key industry initial development (1956 – 1962), the Brazilian economy entered in a new period of high growth rates regarding both income and production. This has allowed further advances in the industrialization process, what has also occurred through the expansion of transnational corporations (TNCs) branches dedicated to the production of lasting goods. In face of the inability of these corporations to promote a self-sustained growth and of the world oil crisis in 1973, associated with the fact that the Brazilian development strategy was dependent on external loans, Brazil ended up by sinking in a profound economic crisis.

Incautiously disregarding the new international scenario, the Brazilian government elaborated the II National Development Plan (II PND), launched in 1974, fostering the expansion of the capital goods industry and investments in the energy sector and infrastructure, widening thence its foreign debit, which has even worsened around 1978 – 1979, after the second world oil crisis. At this stage, the Brazilian economy definitely entered in an unprecedented crisis, which would last for the following ten years. In summary, the 1970s represented an immeasurable effort towards modernizing and strengthening the material bases of the national capitalism. Nevertheless, this decade witnessed an aggravation of the historical poverty and inequalities in the Brazilian society (Fundação Seade, 1992).

During this period, the endeavors of the II PND in relocating part of the ongoing industrial development in the São Paulo metropolitan area to other regions in Brazil were only partly successful. It was in São Paulo State that the national industrialization process reached its summit: its industrial structure was diversified as a result of the greater importance obtained by the lasting and capital goods industry; its agriculture was export-oriented and increasingly linked to the industrial sector; its commercial and services activities were differentiated by the remarking growth of their most dynamic sectors; its cities acquired giant dimensions, and the State absorbed almost three million immigrants.

3.2.4.5 Years 1980 – 1989

The 1980s were known as the “lost decade” and started with one of the most severe crisis in the urban history of Brazil. The economic recession, the increasing unemployment rates and the social crisis had their greatest impacts exactly in São Paulo State. The share of the industrial activities in the State gross domestic product (GDP) dropped from near 51% in 1980 to 43% in 1988. The industrial retraction, however, does not account for the whole set of internal transformations experienced by the national industry (especially amongst the export-oriented sectors) during the crisis, like productive and administrative rationalization as well as “reengineering” for increased gains in productivity and greater competitiveness. Yet some top-edge sectors, established in the preceding decade, had an extremely favorable performance, such as the processing agribusiness sector, the aeronautics and armament industries as well as the informatics, microelectronics and telecommunications industries.

Albeit the reduction in the subsidized financing, the agriculture in São Paulo State grew on average 2.1% between 1980 and 1988, smoothing the crisis effects. This is greatly due to the “Proálcool”¹⁵ program and to the enlargement of export cultures, particularly the orange, favored by its growing participation in the international market of concentrated juice. In this way, the most capitalized agricultural sectors, mainly those meant for export, assured high profits, requiring hence the growth of complementary tertiary activities, with clear repercussions in the densification of the State urban network (Fundação Seade, 1992).

3.2.4.6 Years 1990 – 1999

This decade represented a landmark concerning the evolution of entrepreneurial adjustment strategies initiated in the 1980s. With the release of the “Plano Collor” in 1990 - a national economic development plan based on liberalization, deregulation and privatization - conceived by President Fernando Collor for his four-year mandate, the

¹⁵ “Proálcool” is a Brazilian national program, designed to fund the production (planting, processing and refining) of sugar cane alcohol for vehicles fuelling purposes.

enterprises were generally forced to adopt severe defensive operational and organizational adjustments, favoring the better schooled and qualified employees (Fundação Seade, 1995).

In face of the globalization motto at the international level, and opening of markets and recession at the national level, the unemployment problem has worsened above all in São Paulo State, where the still ongoing process of agriculture modernization associated with the relocation of some important industries to other States ended up by stressing the social and urban problems of the previous decades. This economic recession together with a decline in the population growth rates¹⁶ produced a marked relenting in the urbanization rhythm of São Paulo metropolis and cities in the State inland. Such phenomenon sharply contrasts with the accelerated urbanization phenomena of the preceding decades.

3.3 Spatial Configuration of the Urban Network in the Macroarea of São Paulo State Inland

A meaningful proportion of the industrial development held during the 1970s occurred in the inland of São Paulo State, as an indirect result of federal initiatives. According to what was already stated in Section 1.1, the II PND included the creation of industrial development funds and of policies to foster export and agribusiness activities as well as to decentralize the production of basic industrial inputs. Its goals have been partially achieved, given that greater part of the intend industrial relocation took place in the inland of São Paulo State, for it presented good infrastructure conditions, successful agricultural enterprises and an incomparable urban network, established since the expansion of export coffee plantations during the XIX century (Fundação Seade, 1992).

In this sense, the inland of the State did not only receive more industries and more progress from São Paulo metropolis, but also a representative share of its urban and

¹⁶ According to Fundação Seade (1992), the population growth rate of São Paulo State declined from 2.7% in the 1980s to 1.9% in the 1990s.

social problems. The 1970s were marked by a chaotic and booming urbanization, above all in the metropolitan area, but also to a certain extent in main Brazilian cities and in cities of São Paulo State inland. Former medium-sized cities experienced drastic inequalities that characterize social life in mass consumption societies with late development: on the one hand, an increasingly complex and diverse social and material framework, and on the other hand, signs of social exclusion (slums and unequipped peripheral squatter settlements) that were no more exclusive of big cities (Fundação Seade, 1992).

The 1980s witnessed the sequence of the industrial relocation from São Paulo metropolis towards other areas in the country, especially the inner regions of this State. In the inland of São Paulo, the industries promoted an economic growth above the national average, what was followed by a good agricultural and agribusiness performance, leading to a more positive employment scenario, both in the industry and services sector. If the 1970s were characterized by the introduction of industrial development in São Paulo inland, the 1980s, on their turn, witnessed a truly inward advance of services. The implementation of big supermarkets; sophisticated franchising shops and malls; foreign banks; consultancy companies; marketing agencies; international standard hotels and TV broadcasting stations are typical phenomena of the 1980s in the main State inland cities.

Therefore, the 1980s concurred to sharpen the distinct economic behaviors of the metropolis and the inland, which were already observable by the end of the 1970s. In São Paulo metropolis, the small economic growth can be imputed to a poor performance of the tertiary sector, whereas in the State inland, the relocation of the industrial development and the consequent inward advance of services assured a comparatively improved economic efficiency. This conjecture can be better explained by the comparative evolution of urban areas between the metropolis and the inland (TABLE 3.1), accomplished through the usage of Landsat Multispectral Scanner Sensor (MSS) and Thematic Mapper sensor (TM) images series (Fundação Seade, 1992).

TABLE 3.1 – Comparative increment evolution of urban areas between São Paulo metropolis and main cities of the State inland – 1974 – 1989.

<i>AREA INCREMENT (THOUSAND HA) BY TIME PERIOD</i>			
<i>URBAN AREAS</i>	<i>1974 – 1980</i>	<i>1980 – 1984</i>	<i>1984 - 1989</i>
<i>São Paulo metropolis</i>	<i>46.9</i>	<i>21.0</i>	<i>18.4</i>
<i>Main cities of São Paulo State inland</i>	<i>27.0</i>	<i>29.4</i>	<i>18.1</i>

SOURCE: Adapted from FUNDAÇÃO SEADE (1992, p. 58).

With spatial particularities and distinct rhythms, the urbanization process of the latest decades assigned a very peculiar urban growth pattern to the State of São Paulo as a whole. During these years, the construction of high-rise buildings (FIGURE 3.3) and the increase in peripheral low-income settlements are detectable in most of its main cities, cases of conurbation are more and more frequent, and 'metropolization'¹⁷ in São Paulo State was no more an exclusive phenomenon of São Paulo capital: new metropolitan areas were established in São Paulo State, such as Santos in 1996, and Campinas in 2000 (EMPLASA, 2000).

The construction of high-rise buildings, with the derived densification of certain residential areas, can be ascribed to a set of factors, like the increase in the demand for housing, the search for personal external economies (resulted from the clustering of infrastructure and services in certain urban areas), speculations in the real estate market, the rising difficulties for mobility inside cities, the growing urban violence, etc.

The emergence of peripheral low-income settlements, on its turn, might also be explained by some of the aforementioned set of factors, but it is still enhanced by the ever-increasing urban densification process, which tends to valorize more central areas to the detriment of those presenting low-density of occupation, and also by the local

governments, which build social housing in relatively distant plots, thus permissively contributing for the implementation of low-income settlements in even farther and strictly peripheral areas, totally deprived of infrastructure facilities (Fundação Seade, 1992).



FIGURE 3.3 – Piracicaba skyline with high-rise buildings in the background.

SOURCE: CIAGRI-USP (2003).

In conclusion, in nearly all main cities of São Paulo State inland, the economic growth of the 1950s through 1980s allowed a fast and pronounced expansion of their urban areas. Those which historically had a well established urban framework and more diversified services became, already by the late 1950s and early 1960s, regional development poles (see FIGURE 3.1), organizing the urban network within their catchment areas and the web of socioeconomic relationships of smaller towns in their surroundings.

In the doorway of the XXI century, new economic possibilities arise for the western inland of São Paulo State. The accomplishment of the Tietê-Paraná waterway will ensure a greater economic dynamism amongst South American countries, mainly regarding the transport of grains. The drop of customs barriers enabled by the unified South American market ('Mercosul'), associated with cheaper freight transport means,

¹⁷ Other southeastern States of Brazil, like Minas Gerais, Paraná and Santa Catarina, also witnessed the foundation of metropolitan areas in cities of their inland (EMPLASA, 2000).

may bring about drastic changes in the agricultural profile of the State, especially within the waterway catchment area (CITP and CEPAM, 1994), what will certainly strengthen the inland developmental trend observed in the latest decades.

3.4 Selected Cities for Analysis: Geographical Characteristics and Socioeconomic Peculiarities

3.4.1 Bauru

3.4.1.1 Geographical Characteristics

Bauru is located in the Tietê lower midstream watershed, western inland of São Paulo State, and approximately 345 km away from the State capital – São Paulo city. This municipality comprises a population of 316,064 inhabitants, out of which 310,442 inhabitants live in urban areas (IBGE, 2000). The city of Bauru lies upon a geomorphological unit called 'Peripheric Depression', which dates back to the High Cretaceous Period, and is predominantly covered by 'cerrado' (bushy savannah). The city presents highland tropical climate, with mean temperatures ranging from 17° C to 29° C throughout the year, and average rainfall of 1,310 mm/year (DAEE, 1990).

3.4.1.2 Socioeconomic Peculiarities

Bauru is regarded as the biggest crossing point among railways, waterway and highways in Latin America. The city itself was born as a crossing point between railways during the inward advance of the coffee culture in the XIX century. Four major interregional roads pass through the city, connecting it to the State capital, to the farthest western regions of São Paulo State as well as to the States of Mato Grosso do Sul and Paraná.

The railway transport, through its south branch, connects the city to the markets of Argentina and Uruguay. Towards the west, it is possible to reach Paraguay and Bolivia, the north of Argentina, and through Chile, the Pacific Ocean. In the eastwards direction, the railways lead to the seaports of Santos and Paranaguá in the Atlantic coast, and also to metallurgical raw material zones, such as Volta Redonda, in the State of Rio de

Janeiro, and Cubatão, in São Paulo State. The municipality of Bauru is located only 25 km away from the freight intermodal terminal of the Tietê-Paraná waterway, situated in the neighboring municipality of Pederneiras. Along this waterway, it will be possible to reach fluvial ports in Uruguay and Argentina.

Bauru currently shelters a State airport, which offers regular flights to main cities of São Paulo State, like São Paulo, Campinas, Guarulhos and São José do Rio Preto. A second bigger airport is being presently built, and it will operate standard flights to the whole country and to South American countries as well, what will be crucial to the economic development in the region.

A gas pipeline extending from Bolivia towards the Brazilian Atlantic coast passes through the city of Bauru, supplying its population with natural gas. This route also supports optic fiber pipes, which connect Bauru to the State capital.

The city also shelters a very diversified industrial park, with an outstanding performance of the agribusiness sector. Around thirty companies settled in this municipality export products like notebooks, machinery, automobile batteries, processed bovine meat, powder juice, noodles, leather articles, and plastic utilities.

Regarding commercial facilities, the city is well supplied with several department stores, one hypermarket and a mall. A second mall is currently under construction. In terms of services, the city offers around forty bank agencies, international standard hotels, several TV broadcasting stations and internet providers. The city counts on a great number of health care facilities, some of which are top-ranking, like highly-specialized and university-based hospitals. There are as well plenty of tourism and leisure opportunities in the city, like the botanic garden, the Bauru aero club, the tree farm, and many other country clubs. And finally, Bauru is also known to be an educational city, sheltering two State universities campi (University of São Paulo – USP and State University of São Paulo – UNESP), two private universities and three private colleges (Fundação Seade, 2000a).

In order to present such a highly diversified urban framework, it is deducible that the city underwent a drastically fast urbanization process. Bauru together with the greater part of the inland western cities are those State cities which most grew in comparative terms during the 1980s, reproducing to a smaller extent the pattern of Latin American big cities, with high-rise buildings in central areas (FIGURE 3.4), non-occupied areas within the main urban agglomeration, resulted from speculative actions, and peripheral low-income settlements (Fundação Seade, 1992).

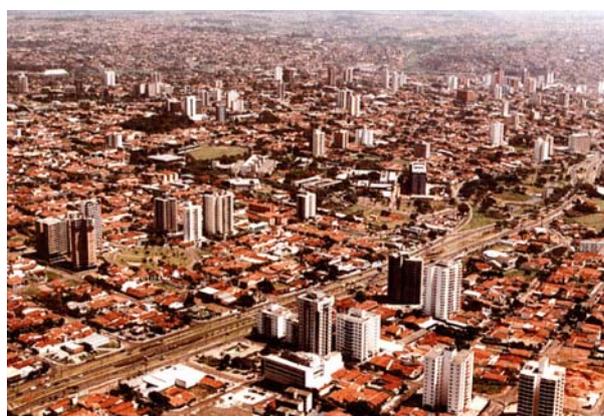
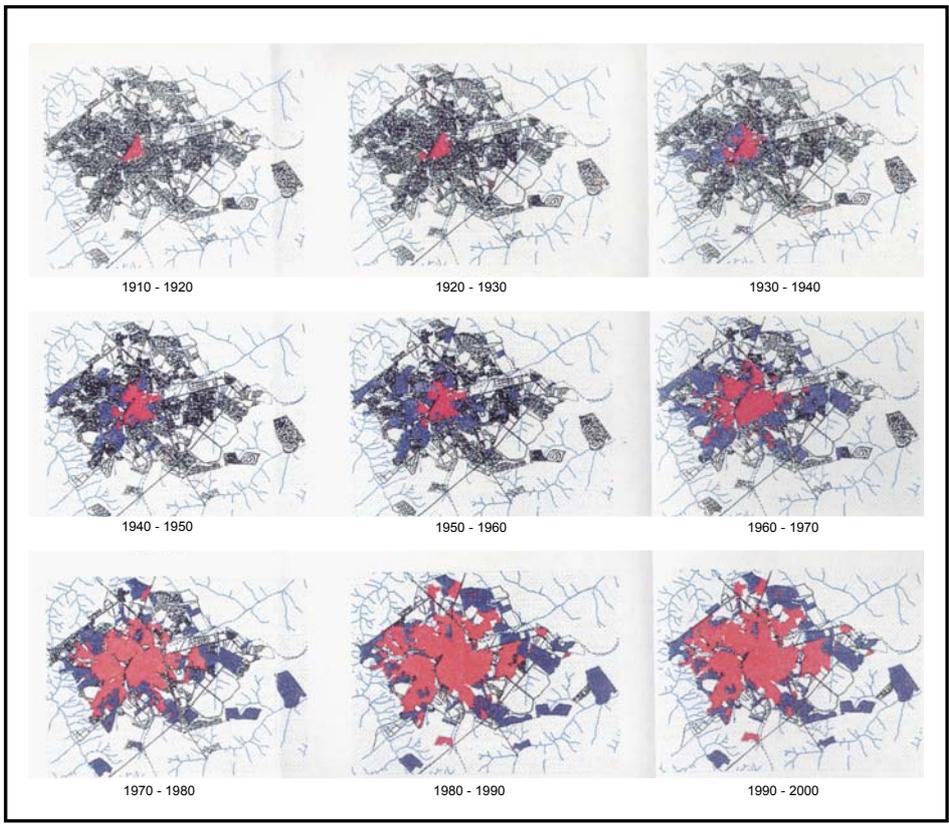


FIGURE 3.4 – Bauru aerial view with high-rise buildings.

SOURCE: ITE-FCEB (1998).

The urbanization boom process is reported in Bauru master's plan (SEPLAN – Bauru, 1997), where through a comparison of the urban evolution during the latest century, it has been assessed that its urban area grew over 300% in 50 years (FIGURE 3.5).

As already stated in Section 3.3, the industrial development in the State inland during the 1970s indirectly strengthened the tertiary activities, especially during the subsequent decades. Bauru is a typical case where the commercial and services activities are preponderant in relation to the industrial sector, what can be identifiable by the sectorial shares of the municipal GDP (FIGURE 3.6).



* The red color corresponds to the effectively occupied areas, and the blue, to legalized settlements.

FIGURE 3.5 – Evolution of Bauru urban area in the latest century.

SOURCE: SEPLAN - BAURU (1997).

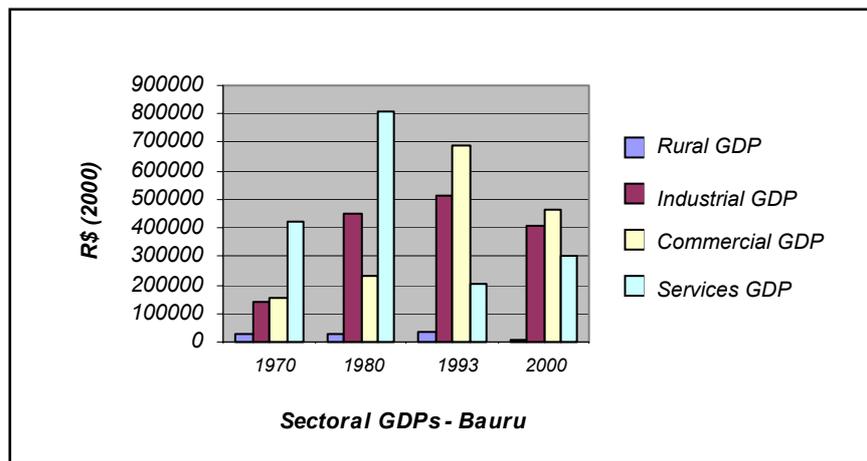


FIGURE 3.6 – Sectorial GDPs for the city of Bauru in the latest four decades.

SOURCE: Adapted from IPEA (2001); SEADE (2002).

3.4.2 Piracicaba

3.4.2.1 Geographical Characteristics

Piracicaba is located in the Tietê upper midstream watershed, western inland of São Paulo State, and approximately 152 km away from the State capital – São Paulo city. This municipality comprises a population of 329,158 inhabitants, out of which 317,374 inhabitants live in its seat city (IBGE, 2000). Piracicaba lies upon the 'Peripheric Depression', which dates back to the Pre-Cambrian Period. This geomorphological unit is partially covered by 'cerrado' (bushy savannah) and partially by the Atlantic rain forest. The city presents highland tropical climate, with mean temperatures ranging from 12° C to 22° C throughout the year, and average rainfall of 1,405 mm/year, with rains distributed into two seasons (DAEE, 1990).

3.4.2.2 Socioeconomic Peculiarities

Piracicaba is amongst the one hundred best Brazilian cities for entrepreneurial investments according to a ranking established by a renowned economic magazine – Exame - in 2001 (Fundação Seade, 2000b).

The city is served by two main roads: the first connects the city to two parallel highways that link the State capital to the metropolis of Campinas, and the other one, connects Piracicaba to a highway leading to the western regions of the State.

The city contains a local airport which allows the operation of medium-sized aircrafts during day and night. Nevertheless, the international airport of Viracopos, located in Campinas, is only 85 km away from the city.

Piracicaba will also benefit from the Tietê-Paraná waterway, with the construction of a dam and a canal lock in the neighboring municipality of Santa Maria da Serra, what will render the Piracicaba river navigable until Artemis, a district belonging to the municipality of Piracicaba. The construction of a railway terminal to be integrated to the Artemis waterway terminal is also envisaged.

The gas pipeline that extends from Bolivia towards the Brazilian Atlantic coast also passes through the city of Piracicaba, supplying its population and industries with natural gas.

Piracicaba has distinguished itself as an important sugar cane producer pole in São Paulo State, and gathers huge sugar cane and alcohol agribusiness complexes. The city also shelters a very diversified industrial park, with big-sized and high technology plants, and whose outstanding sectors are metallurgy, food processing, paper and textiles. Concerning commercial facilities, the city is well supplied with three commercial zones, several great supermarkets, a big-sized mall and three minor ones. In terms of services, the city offers tens of bank agencies, international standard hotels, five TV broadcasting stations and several internet providers. The city counts on a great number of health care facilities, with four hospitals and several health care centers.

A very strong economic sector in Piracicaba is tourism, since this fluvial city offers a greened and pleasant landscape (FIGURE 3.7); a set of remarkable historic and architectural assets, like museums, old mills, buildings and churches; an astronomic observatory; and beautiful leisure areas. Piracicaba is also distinguishable for being one of the first Brazilian cities to implement the local Agenda 21, a chart of environmental guidelines for sustainable development, created as a collective compromise of member-countries that attended the Earth Summit Conference – ECO 92, in Rio de Janeiro.



FIGURE 3.7 – View of Piracicaba city and its green areas.

Likewise Bauru, Piracicaba is also known to be an educational city, comprising excellence research centers and universities, being a national reference in the fields of agronomy and biotechnology. The city contains two public research centers (one dedicated to the study of nuclear energy in agriculture, and the other, to technologies supporting sugar cane by-products), besides a campus of the USP agronomy school, two public colleges, a private university and a private college (Fundação Seade, 2000b).

But in contrast with the economic profile of Bauru, the industrial sector in Piracicaba, chiefly driven by the sugar cane and alcohol agribusiness complexes, is prevalent over the commercial and services activities, as shown in the bar graphic of sectorial GDP values (FIGURE 3.8).

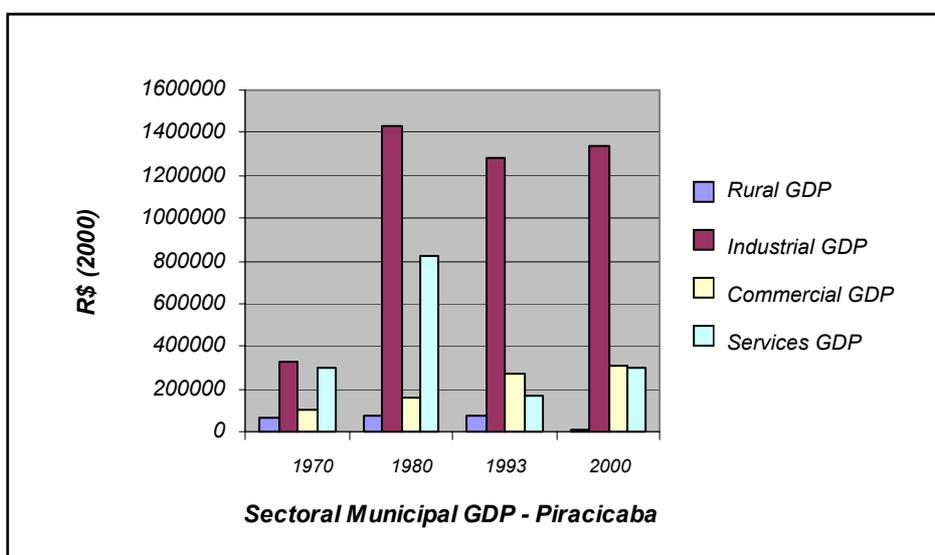


FIGURE 3.8 – Sectorial GDPs for the city of Piracicaba in the latest four decades.

SOURCE: Adapted from IPEA (2001); SEADE (2002).

3.5 Conclusions

The two cities adopted for the modeling experiments of the current PhD research present similarities regarding their historical background, for they were born along the Tietê river valley as a result of the human inward occupation with the first fluvial

exploring expeditions and mining activities in the XVII and XVIII centuries. The two of them grew in importance with the apogee of the coffee economy in the XIX and beginning of the XX century, and the concomitant arrival of railways and more diversified and sophisticated commercial and services activities. In the same way, these two cities have a renowned tradition in agribusiness and both of them inherited part of the metropolis industrial development from the 1970s onwards as a result of federal initiatives. And finally, Bauru and Piracicaba are regional development poles of São Paulo State, with a very sophisticated and complex urban framework, and a high level of educational infrastructure.

Nevertheless, these two cities differ in their urban structure and landscape. Bauru, which was born as a crossing point of railways, is still marked in its urban tissue by the transport system: the city is organized around four interregional roads and the railway track, which still passes through the city center.

Although the railway also reaches the center of Piracicaba, and the roads also play a certain role in structuring its urban area, the city lays a great emphasis on the river, which is a landmark for the social life of its inhabitants. The main central commercial zone borders the green areas established along the Piracicaba river margins, as do the huge areas dedicated for universities and colleges campi in the northern and northeastern portions of the city.

Another divergent aspect between these two cities concerns their economic vocation. While Bauru is typically distinguished by the remarkable presence of tertiary activities, the industrial sector is overwhelming in the local economy of Piracicaba.

In summary, in both cases the challenges for studying these cities lie on the fact that they underwent urbanization booms and show a great diversity concerning their economic specialization profiles throughout the latest decades.

CHAPTER 4

BUILDING THE GEOGRAPHIC DATABASE

4.1 Spatial Input Data: Cartographic, Aerophotogrammetric and Remote Sensing Data

As stated previously, the central topic of the current research focus on modeling of urban land use change. In fact, a variety of spatial data (city maps, satellite images, conventional and digital aerial photos) provide information on urban features like extent and shape of urban areas, urban greening, urban traffic network, etc. But to specifically depict information on urban land use with a high degree of precision and reliability, it is crucial to count on urban land use maps, commonly supplied by planning departments of local governments.

This work has been essentially based on this sort of information to generate the digital urban land use maps that fed the simulation model. And due to this reason, the starting and end dates of the consecutive simulation periods were defined in function of the existent land use maps for the two cities under study. For instance, the planning secretariat of the local government of Bauru issued four land use maps in the latest four decades: in 1967, 1979, 1988 and 2000. In this way, the time series adopted for the Bauru simulation experiments, comprising thirty-three years, was divided into three periods, namely 1967-1979, 1979-1988 and 1988-2000.

In the case of Piracicaba, only two land use maps were elaborated during the four latest decades: in 1985 and 1999. However, the first urban zoning legislation (without maps) released for the city dates from 1964 (Piracicaba, 1964), immediately after the completion of an aerophotogrammetric survey carried out over the city and surrounding regions in 1962. Thus, the urban land use change simulations for Piracicaba are based on a time series, comprising thirty-seven years and consisting of two consecutive periods: 1962-1985 and 1985-1999.

Forty years were deemed an ideal length for the simulation time series, since one of the chief aims of this study concerns the understanding of the relatively recent driving forces determining urban land use change during the time in which the two case study cities were subjected to urbanization booms. The 1970s and 1980s were marked by the strengthening of the inward industrial relocation process in São Paulo State (see Section 3.2.4.4), and the consequent fast urban expansion of its main cities. These two decades were included in the analysis, together with the preceding decade, once the rural out migration phenomena in Brazil, and especially in São Paulo State, were sharply enhanced around the 1960s and henceforth.

A further reason for limiting the analysis to the latest four decades regards the matter of data availability. Land use maps in the 1950s and earlier are rare (if not totally inexistent), and even more scarce are maps of technical and social infrastructure as well as of occupation density, which account for the input variables to drive the simulation model. In this sense, working with longer time series would prove to be unfeasible.

An important remark should be addressed to the delimitation of simulation periods. As previously said, their bounds (initial and final time) are confined to the available releases of urban land use maps. In fact, if these maps were issued more often, these periods could be set in accordance with homogeneous circumstantial contexts, such as a sequence of local and/or State governments with a similar political approach, stable macroeconomic scenarios, etc. Some works on deforestation processes modeling adopt this methodology for defining simulation periods (Soares-Filho, 1998; Mertens and Lambin, 1999; Mertens et al. 2000). This delimiting procedure is thus functional due to the fact that the set of driving forces determining the total amount of land use change tend to alter from time to time, as a result of exogenous forces like political-economic contexts. This suggestion remains as a methodological guidance for further studies, in case urban land use maps become more often available.

The following sections will be dedicated to introduce the spatial input data used in the modeling experiments, reporting their main technical characteristics and also their specific usefulness for the elaboration of the digital land use and variables maps.

4.1.1 Cartographic Data

All the cartographic data used in the modeling experiments were paper plans, concerning official city maps, land use maps, technical and social infrastructure maps as well as occupation density maps. Many of them, usually the most recent ones, existed in digital format, but due to copyright constraints, they were not rendered available by their authors: the planning and water supply & sewerage secretariats of the municipal governments in Bauru and Piracicaba.

These plans were mostly drawn at scales of 1:20,000 for both cities, with a few exceptions being the map of zoning for Bauru in 1967, drawn at scale 1:10,000 and the map of earth and paved interregional roads of Bauru in 1988 at scale 1:75,000. The oldest plans were made with conventional China ink on onionskin, whereas the recent ones are either plotter or deskjet printings. The most recent official city maps were initially scanned and then vectorized on screen. Details about this digital conversion process will be provided in Section 4.3.1. Below is an example of a paper plan, converted in digital format (FIGURE 4.1).

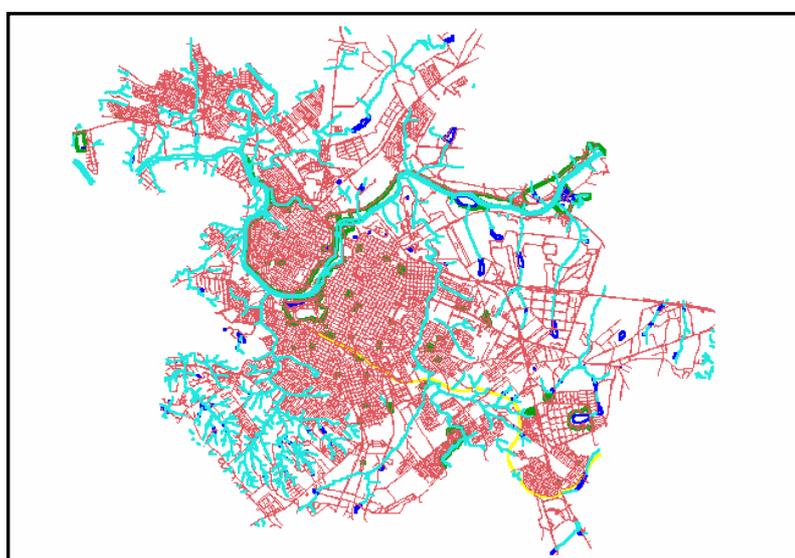


FIGURE 4.1 – Map of Piracicaba in 1999. Street blocks are in red, railway in yellow, leisure areas in green, rivers in light blue, and lagoons in dark blue.
SOURCE: SEMUPLAN (1999).

4.1.2 Digital Aerial Photos

As already stated in Section 4.1, an aerophotogrammetric survey was carried out over the city of Piracicaba and surrounding regions in 1962. This survey was undertaken by the enterprise 'Terrafoto S.A.', and the flight scale was 1:100,000.

The photos covering the city of Piracicaba have been converted to digital format through a scanning with the Scanjet HP 6350 - 1250 dpi, and their subsequent arrangement in a digital mosaic (FIGURE 4.2), accomplished with the software TNT@mips 6.6. This mosaic has been done by the team under the coordination of Prof. Dr. Gerd Sparovek of the Department of Soils and Plants Nutrition (LSN), belonging to the School of Agronomy "Luiz de Queiroz" of the University of São Paulo (ESALQ-USP). The urban area of Piracicaba has been visually delimited with polygons in vector format, drawn in orange color, and superimposed on the photos mosaic.

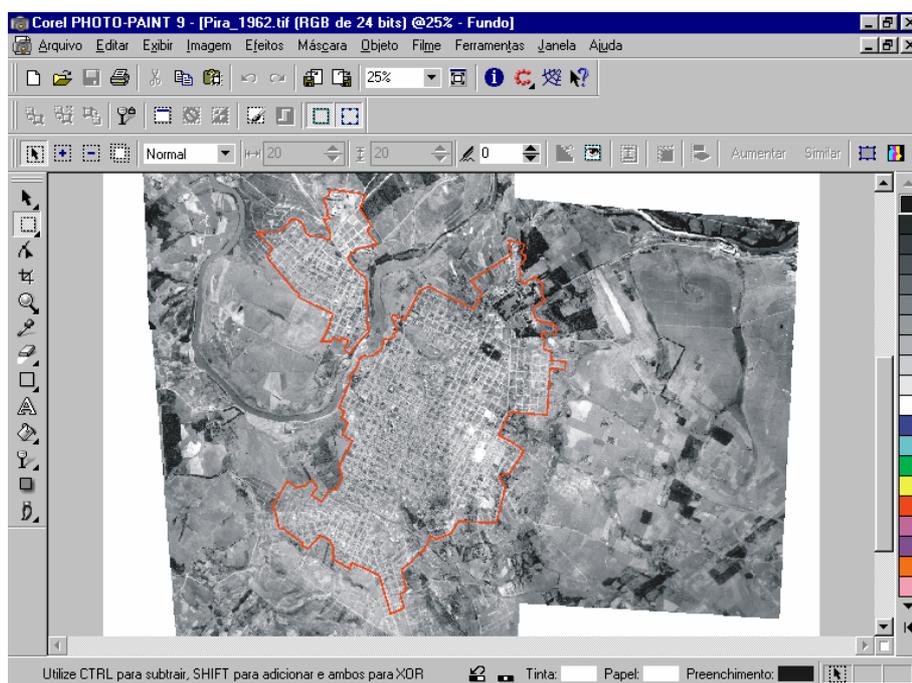


FIGURE 4.2 – Mosaic of digital aerial photos for the city of Piracicaba - 1962.
SOURCE: LSN - ESALQ-USP (2002).

4.1.3 Remote Sensing Data

Landsat satellite images have been used in this work to update the official city maps, inasmuch as illegal residential settlements are not shown on the latter. Their acquisition dates correspond to the bounds of the simulation periods (see Section 4.1) adopted for each of the two cities. A Landsat – 3 MSS image of 1979, and two TM - 5 images of 1988 and 2000 (INPE, 1979, 1988, 2000) were used to update the Bauru city maps. In the case of Piracicaba, two TM – 5 images of 1985 and 1999 (INPE, 1985, 1999) have been employed. The land use maps issued in the 1960s could not be updated by means of RS data, since satellite images started being available from the beginning of the 1970s.

Among the scientific community engaged in the applications of remotely sensed data to urban studies, there is no consensus as to which is the most appropriate TM bands composition for detecting urban areas and differentiating their intra-urban structures. Seevers et al. (1985) reports that the combination of bands 5, 4, 3 is ideal for visual interpretation of urban areas, and bands 5, 4, 1, for automated classification. Johnston and Watters (1996) have worked with an isodata¹⁸ classification on bands 2, 3, 4, 5 and 7 in order to manually interpret and group the derived sixteen classes into "paved" and "non-paved" ground cover types. Liu (2003) citing a study of the USGS EROS Data Center, states that for mapping urban features, the combinations 5, 4, 3; 4, 3, 2 and 7, 4, 3 are preferred for visual interpretation.

The heterogeneous nature of urban areas implies a great spectral complexity, what can be explained by the fact that they present materials with a very diversified spectral behavior and they often include non-urban uses, such as green areas, bare soil, etc. This diversity context creates thus problems for classification (Haack et al. 1987). Some alternative solutions have been proposed to tackle this problem. Gong and Howarth (1990) proposed a classification method designed to extract segments containing an intensive density of high spatial frequency, which they believe best characterizes

¹⁸ Richards (1995) citing Ball and Hall (1965) defines the isodata algorithm as the one based upon estimating some reasonable assignment of the pixel vectors into candidate clusters and then moving them from one cluster to another in such a way that the sum of squared errors measure of the preceding section is reduced.

urban areas. Solberg et al. (1990) developed an algorithm for the detection of urban areas based on an interference filtering. This method, restricted to high spatial resolution sensors only (10m or 20m), comprises the extraction of the traffic network on the image, which is then followed by a filtering operation, meant to detach areas with a high density of linear elements. The filtering algorithm creates thus interferences among the lines, making its answer enhanced when the lines density is greater. Although this method can detect urban areas with a high level of accuracy, its borders are uncertain and little detailed.

Jensen-Moller (1990) created an algorithm for the detection of urban areas based on different flexible classification rules, comprising both quantitative information and heuristic assumptions reflecting the criteria that would be commonly adopted by a human interpreter. The procedures sequence is the following: a) intelligent detection of roads and linear structures; b) processing of image attributes, like mean and variance values and texture; c) final classification upon basis of an expert systems approach.

An alternative method of describing urban land cover is the V-I-S (vegetation – impervious surface – soil) model proposed by Ridd (1995), which endeavors to assess biophysical composition of the urban landscape by assigning values for these three components. Briefly, this method deals with matters related to the mixture of or interference among targets at a sub-pixel scale.

The usage of remote sensing data in this work, however, was limited to a visual interpretation of color composition images, since detailed information on the urban structure of the cities under study was already available, and satellite images have been solely employed to identify eventually existent illegal settlements. To meet this end, digital city maps in vector format have been superimposed on the satellite images, which previously underwent radiometric and geometric corrections, georeferencing operations and contrast enhancement. Detailed information on the georeferencing techniques will be provided in Section 4.3.2.

For the color composition, the TM bands **1** (0.45 μm – 0.52 μm), **4** (0.76 μm – 0.90 μm) and **7** (2.08 μm – 2.35 μm) have been selected. Band 1 has been subjected to

atmospheric correction procedures. Built-up areas present a high response in bands 1 and 7 (Bowker et al. 1985), whereas vegetation presents a peak of response in band 4 (Richards, 1995), which is typically used for biomass assessment (Novo, 1988). This combination of bands offers therefore a good differentiation between urban and non-urban areas. A further advantage in working with this set of bands is that they present little correlation among each other. An example of a colour composition image (1B_4R_7G) is given below, which shows the city of Piracicaba in 1999 (FIGURE 4.3).

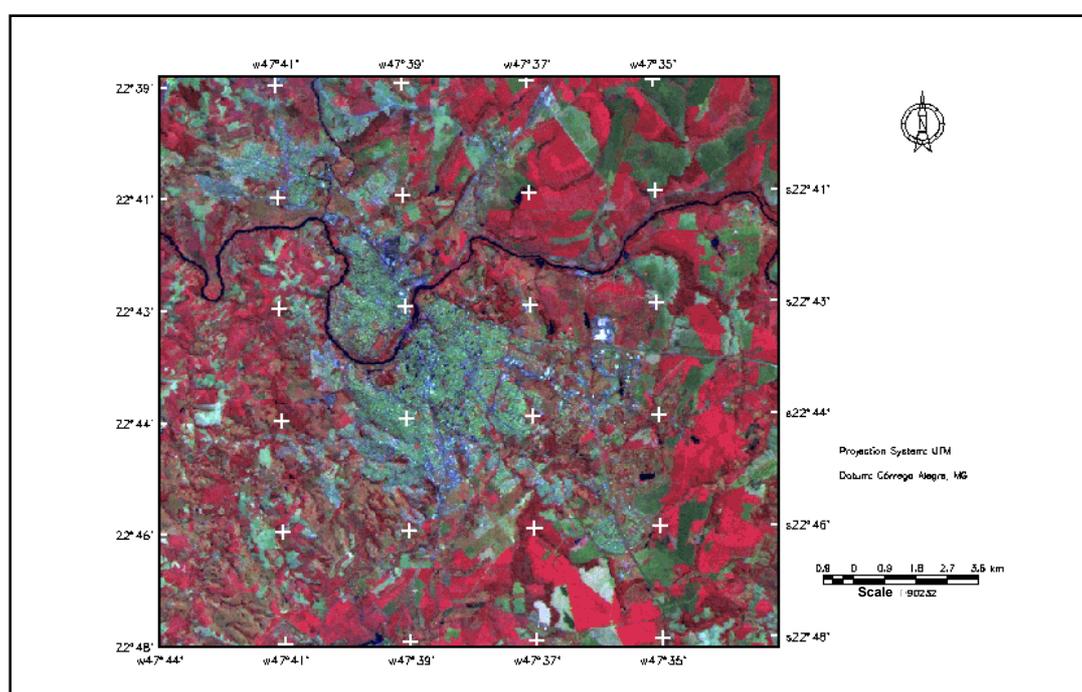


FIGURE 4.3 – TM – 5 image of Piracicaba, 1B_4R_7G, 222/76 – 07/16/99.

SOURCE: INPE (1999).

4.2 Non-Spatial Input Data: Demographic-Economic Data

The non-spatial data have been used in this work to parameterize the total demand of land use change in the forecast scenarios. These future scenarios were built upon basis of linear regression analysis, where the dependent variable corresponds to amounts (in hectares) of land use types, while demographic data and economic performance

indicators account for the independent variables. Further details on these parameterization methods of forecast scenarios will be given in Section 5.3.2. This section is basically committed to introduce these non-spatial data and their respective sources, and present the way conversions of economic data in different currencies were made.

4.2.1 Demographic Data

Demographic data were obtained for the greater part of the time series of Bauru and Piracicaba in spans of five years. This is due to the fact that economic performance indicators were available only in quinquennia starting from 1970. Urban population has been exclusively taken into account in the regression analysis, because this is the share of the municipal population effectively impacting the growth of urban areas.

According to Fundação Seade (2003), urban population refers to the counting of inhabitants living within the urban boundaries of a given municipality, which are defined by municipal legislation. Only the urban populations belonging to the municipal seat cities of Bauru and Piracicaba have been considered in the analysis.

The population data have been obtained through demographic censuses for the years 1970, 1980 and 2000 (IBGE, 1971, 1982, 2000). Estimates of population have been used to assess the urban population of Bauru and Piracicaba in the years 1985, 1990 and 1995 (IBGE, 1987, 1991, 1997). Since only total population is provided by these estimates, the urban populations have been calculated by interpolating the proportion of urban population identified in the two closest censuses. And finally, the urban populations of these cities in 1975 were estimated by obtaining the arithmetic mean between the population data of 1970 and 1980 (TABLES 4.1 and 4.2).

4.2.2 Economic Data

Total and sectorial municipal gross domestic products (GDP) have been selected as indicators to assess the economic performance of the cities under study. The total municipal GDP is defined as the value of goods and services produced in a given

municipality, from which the expenses with the inputs used in the production process during a year is deducted. It refers to the measure of the total gross added value generated by all economic activities (IBGE, 2002).

An equivalent term for the GDP is the added value (AV), which corresponds to the value that the economic activities add to the goods and services consumed in its own productive process. In brief, it is the contribution or gross domestic product generated by the different economic activities, obtained by the difference between the production value and the intermediate consumption absorbed by these activities (IBGE, 2002).

In fact, the GDP is the added value, and these terms are often used interchangeably. Total GDPs for the municipalities of Bauru and Piracicaba are available in US\$ of 1998 for the years 1970, 1975, 1980, 1985, 1990 and 1996 (IPEA, 2001). Total and sectorial (rural, industrial, commercial and services) GDPs for these municipalities are available in R\$ of 2000 for the years 1970, 1975, 1980 and 1985 (IPEA, 2003b). Total added values (AV) are available in R\$ of 2001 for the years 1993 through 2000, and sectorial added values are provided in R\$ of 2000 also for the years 1993 through 2000 (Fundação Seade, 2002).

In order to obtain standardized values of total and sectorial GDPs for the years 1970, 1975, 1980, 1985, 1990, 1995 and 2000, the following conversion procedures were adopted:

a) Estimation of the total GDP in 2000 (US \$ of 1998):

$$\frac{GDP\ 1996\ (US\$ \text{ of } 1998)}{GDP\ 2000\ (US\$ \text{ of } 1998)} = \frac{GDP\ 1996\ (R\$ \text{ of } 2001)}{GDP\ 2000\ (R\$ \text{ of } 2001)} \quad (4.1)$$

b) Estimation of the total GDP in 1995 (US \$ of 1998):

$$\frac{GDP\ 1996\ (US\$ \text{ of } 1998)}{GDP\ 1995\ (US\$ \text{ of } 1998)} = \frac{GDP\ 1996\ (R\$ \text{ of } 2001)}{GDP\ 1995\ (R\$ \text{ of } 2001)} \quad (4.2)$$

c) Conversion of the total AV in R\$ of 2001 into R\$ of 2000:

$$AV (R\$ \text{ of } 2000) = AV (R\$ \text{ of } 2001) \times \frac{IPCA (2000)}{IPCA (2001)} \quad (4.3)$$

These are all simple rules of proportion, since conversion of currencies are linear transformations.

d) Estimation of sectorial GDPs in US\$ of 1998:

$$Sectorial \text{ GDP}_k (US\$ \text{ of } 1998) = Total \text{ GDP} (US\$ \text{ of } 1998) \times \frac{Sectorial \text{ GDP}/AV_k (R\$ 2000)}{Total \text{ GDP}/AV (R\$ 2000)} \quad (4.4)$$

where k denotes a certain type of sectorial GDP: rural, industrial, commercial or services.

e) Calculation of sectorial GDPs for 1990 in US\$ of 1998:

$$Sectorial \text{ GDP}_k (US\$ \text{ of } 1998) = Total \text{ GDP} (US\$ \text{ of } 1998) \times Proportional \text{ Rate}_k^{1990} \quad (4.5)$$

And this proportional rate has been obtained through the following series of ratios:

$$Proportional \text{ Rate}_k^{1990} = \frac{Proportional \text{ Rate}_k^{(1989 + 1991)}}{2} \quad (4.6)$$

$$Proportional \text{ Rate}_k^{1991} = \frac{Proportional \text{ Rate}_k^{(1989 + 1993)}}{2} \quad (4.7)$$

¹⁹ IPCA (“Índice Anual Geral de Preços ao Consumidor”) refers to ‘Consumers Annual Prices Index’, and the values of IPCA in 2000 and 2001 are respectively R\$1,683.47 and R\$ 1,812.65 (IPEA, 2003a).

²⁰ GDP values were used in the years 1970, 1975, 1980, and 1985, while AV values were used in the years 1995 and 2000.

$$\text{Proportional Rate}_k^{1989} = \frac{\text{Proportional Rate}_k^{(1985 + 1993)}}{2}, \quad (4.8)$$

$$\text{Proportional Rate}_k^{1985} = \frac{\text{Sectorial GDP}_k^{1985}}{\text{Total GDP}^{1985}}, \quad (4.9)$$

$$\text{Proportional Rate}_k^{1993} = \frac{\text{Sectorial AV}_k^{1993}}{\text{Total AV}^{1993}}. \quad (4.10)$$

The final GDP values in US\$ of 1998 as well as the population data used in the parameterization of forecast scenarios are shown below (TABLES 4.1 and 4.2).

TABLE 4.1 – Urban population, total and sectorial GDPs (US\$): Bauru – 1970-2000.

<i>Years</i>	<i>Urban Pop.</i>	<i>Total GDP</i>	<i>Rural GDP</i>	<i>Ind. GDP</i>	<i>Com. GDP</i>	<i>Serv. GDP</i>
1970	61,592	526,500.428	24,884.128	124,826.722	136,439.859	376,832.414
1975	110,166	718,986.733	14,959.686	233,091.154	155,342.626	470,935.893
1980	159,926	983,887.317	23,486.596	344,767.749	176,902.442	1,572,421.062
1985	215,153	1,222,203.235	62,325.610	428,632.198	201,542.378	731,245.427
1990	237,954	1,358,236.390	46,922.487	472,677.454	479,003.725	419,780.872
1995	279,407	2,256,520.737	11,292.207	705,164.856	891,184.402	299,490.980
2000	310,442	1,906,359.257	6,512.305	627,732.044	713,024.547	467,496.231

SOURCE: Adapted from IPEA (2001, 2003a, 2003b) and FUNDAÇÃO SEADE (2002).

TABLE 4.2 – Urban population, total and sectorial GDPs (US\$):Piracicaba– 1970-2000.

<i>Years</i>	<i>Urban Pop.</i>	<i>Total GDP</i>	<i>Rural GDP</i>	<i>Ind. GDP</i>	<i>Com. GDP</i>	<i>Serv. GDP</i>
1970	73,153	666,029.934	60,056.683	314,733.774	95,101.568	291,239.467
1975	115,960	1,312,284.818	39,437.798	853,930.964	121,743.828	418,916.047
1980	158,708	2,126,207.943	66,581.222	1,307,157.079	147,927.333	752,469.633
1985	198,407	1,704,037.322	68,131.878	1,075,191.585	128,818.571	560,713.867
1990	218,590	2,424,143.275	96,842.165	1,615,544.478	289,475.481	432,548.144
1995	236,687	3,251,245.119	24,907.799	2,108,284.969	491,681.220	277,346.390
2000	317,374	3,335,315.985	17,102.944	2,170,367.463	494,371.845	492,749.832

SOURCE: Adapted from IPEA (2001, 2003a, 2003b) and FUNDAÇÃO SEADE (2002).

It is worth remarking that the sum of sectorial GDPs surpasses the total GDP due to the fact that the financial dummy activities are only deducted from the total GDP, since they cannot be accounted for each sector separately (IBGE, 1999, 2002).

4.3 Data Pre-Processing Operations

4.3.1 Digital Conversion of Cartographic Data

As previously stated in Section 4.1.1, the most recent city maps of Bauru (DAE, 2000) and Piracicaba (SEMUPLAN, 1999) were scanned in black and white at the Dutch scanner OCE – G6035S. The TIFF files thereby derived were converted in files with extension OLE or XREF through the Australian software DESKAN 4.5, and then exported to AUTOCAD 14.

These maps were vectorized on screen. Older city maps were recomposed directly upon the recent digital city maps through a visual comparison between the recent maps and the respectively older paper plans. The reconstitution of the Piracicaba city map in 1962, however, was based on a comparison between the 1985 digital city map and a digital mosaic of aerial photos (see Section 4.1.2) dating from 1962 (FIGURE 4.4). These procedures have been adopted aiming at a perfect adjustment in-between maps throughout the whole time series of the two cities.

Further information like technical and social infrastructure, occupation density and land use in diverse years was drawn in different layers, always using the respective city maps as a reference basis.

4.3.2 Georeferencing Techniques

4.3.2.1 Image to Map Georeferencing

Initially, TM – 5 satellite images of Bauru and Piracicaba, respectively acquired in 1984 and 1985 (INPE, 1984, 1985) were used for georeferencing procedures upon basis of topographic charts (IBGE, 1969, 1973) inside SPRING GIS (Câmara et al. 1996). Although the image of Bauru dating from 1984 was not used in the modeling processes

themselves, it has been useful in the georeferencing stage, for it presented the best contrast for visualization purposes.

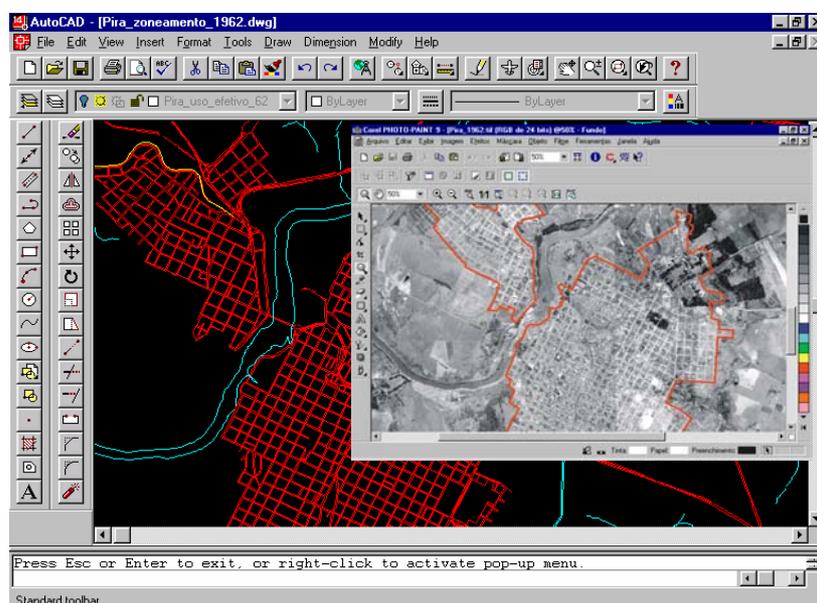


FIGURE 4.4 – Reconstitution of Piracicaba city map in 1962: recent digital city map and digital mosaic of aerial photos dating from 1962.
SOURCE: SEMUPLAN (1985); LSN – ESALQ-USP (2002).

In both cities, the georeferencing used five control points (CP) and a first order polynomial for the correction method. A minimum of four control points is needed when this method is adopted and overdetermined systems are to be employed (Richards, 1995), but given the fact that the images presented a high level of geometric correction and that the urban areas account for a very small portion of an image scene, five points were deemed convenient. In all cases, there was a struggle for scattering the control points all over the urban area as much as possible in order to improve the georeferencing results (FIGURE 4.5). In the case of Bauru, the total georeferencing error was 1.3 pixels, and in Piracicaba, 1.2 pixels, which are not surpassing the admissible threshold of 1.5 pixels (Machado e Silva and D'Alge, 1986; D'Alge, 1987). The control points coordinates for the two cities are presented in TABLES 4.3 and 4.4.

TABLE 4.3 – Control points coordinates in the georeferencing of Bauru image (TM – 5 221/75, 10/01/84).

<i>Control Points (CP)</i>	<i>Latitude (UTM, Datum: SAD - 69)</i>	<i>Longitude (UTM, Datum: SAD - 69)</i>
1	s 22° 16' 53.90"	w 49° 05' 10.97"
2	s 22° 21' 44.21"	w 49° 02' 25.74"
3	s 22° 18' 24.85"	w 49° 07' 22.76"
4	s 22° 19' 07.70"	w 49° 03' 11.47"
5	s 22° 18' 56.31"	w 49° 08' 20.56"

TABLE 4.4 – Control points coordinates in the georeferencing of Piracicaba image (TM – 5 220/76, 08/10/85).

<i>Control Points (CP)</i>	<i>Latitude (UTM, Datum: SAD - 69)</i>	<i>Longitude (UTM, Datum: SAD - 69)</i>
1	s 22° 42' 10.03"	w 47° 39' 00.98"
2	s 22° 42' 55.82"	w 47° 39' 28.20"
3	s 22° 42' 08.72"	w 47° 38' 52.95"
4	s 22° 42' 31.81"	w 47° 38' 42.38"
5	s 22° 40' 45.18"	w 47° 40' 30.28"

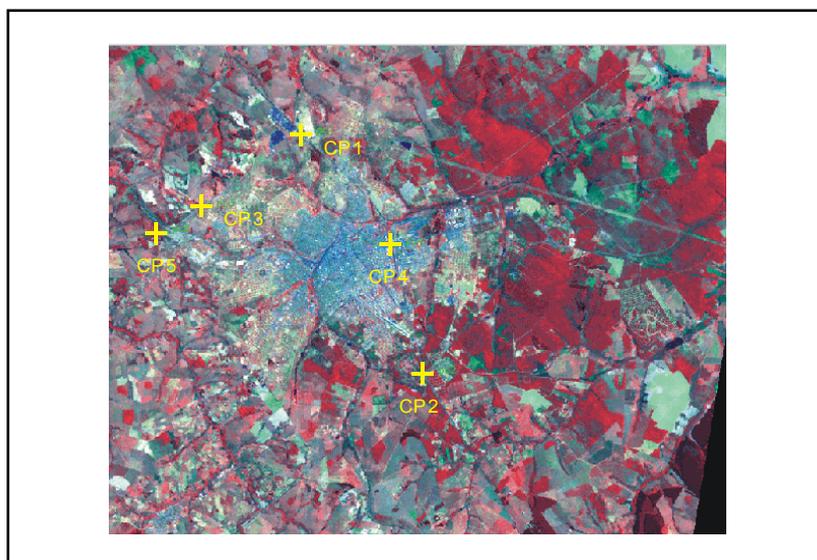


FIGURE 4.5 – Georeferenced image of Bauru with control points (TM – 5 221/75, 10/01/84).

SOURCE: INPE (1984).

4.3.2.2 Image to Image Registration

Image to image registration is necessary for the perfect match among different images of a time series. The geometric distortion of satellite images can be ascribed to a number of factors. These systematic distortions are mainly caused by alterations in the location, altitude, attitude and speed of the platform, sensor non-idealities as well as from the relief and curvature of the ground surface and the earth's revolution (Machado e Silva, 1989).

The 1999 image of Piracicaba was thus co-registered with the image of 1985, and the 1988 and 2000 images of Bauru were co-registered with the image of 1984. In the first case, the total error was around 0.3 pixel, and in the case of Bauru, the errors remained around 0.5 pixel. Three control points were used in both cases, in face of the high geometric correction level of the images. In order to check for the match between images, the SPRING coupling resource was used (FIGURE 4.6).

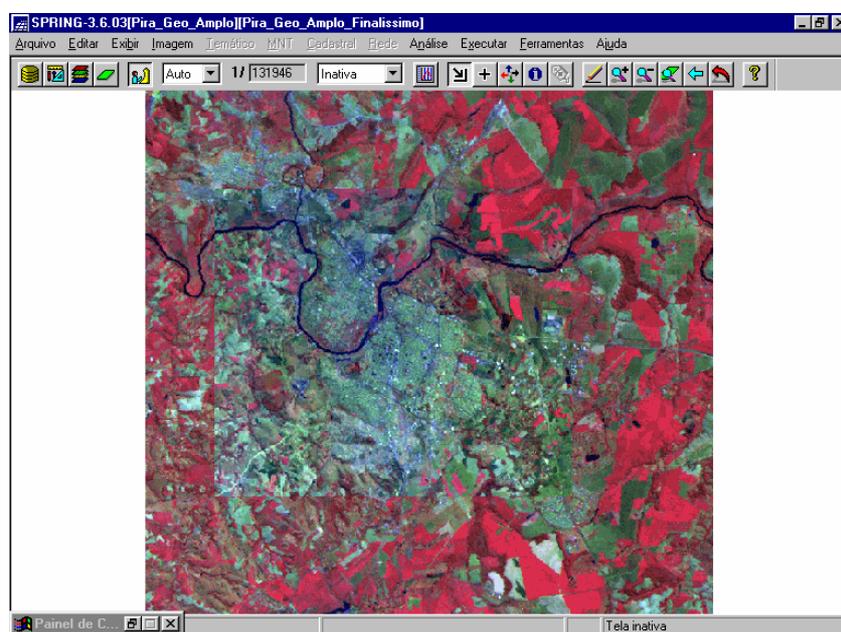


FIGURE 4.6 – Coupling between the Piracicaba images of 1985 and 1999 (TM – 5 220/76, 08/10/85 and 07/16/99).
SOURCE: INPE (1985, 1999).

4.3.2.3 Image to Vector Data Registration

All the vector data were exported as files with extension DXF, and imported to a SPRING project without geographic projection system within the database containing the georeferenced images.

The set of control points used in the image to map registration was also used for registering the vector data imported from AUTOCAD. For Bauru and Piracicaba, the vector data first employed in the registration procedures were the most recent city maps, showing the street blocks and traffic network, since all of the control points were located at crossing points of roads and eventually of roads and railways.

In SPRING, the main window is reserved for the georeferenced image, and in a secondary window, the vector data are retrieved. The control points coordinates, recorded with datum SAD-69, were digitized in the georeferencing command window, and subsequently, the control points were one by one placed in their respective positions on the city maps (FIGURE 4.7).

The set of three control points producing the best adjustment by means of a first order polynomial were kept, and the two other ones excluded. This set of final control points is recorded as a file with PRO extension. The recent city map as well as all the other remaining layers in vector format are finally imported into a definitive project inside SPRING presenting the Universal Transverse Mercator (UTM) geographic projection system. During this import process, the control points file is used in order to insert each of the maps in the correct position in relation to the adopted projection system.

4.3.3 Spatial Data Analysis and Processing

The georeferenced vector data underwent preliminary processing operations in SPRING, such as vector edition (elimination of duplicated and/or spurious lines), polygonalization and association of thematic classes to polygons. A further processing stage was undertaken to generate derivative raster maps from the edited vector data,

like maps of distances and spatial statistical analysis maps like the Kernel points density estimator (Bailey and Gatrell, 1996).

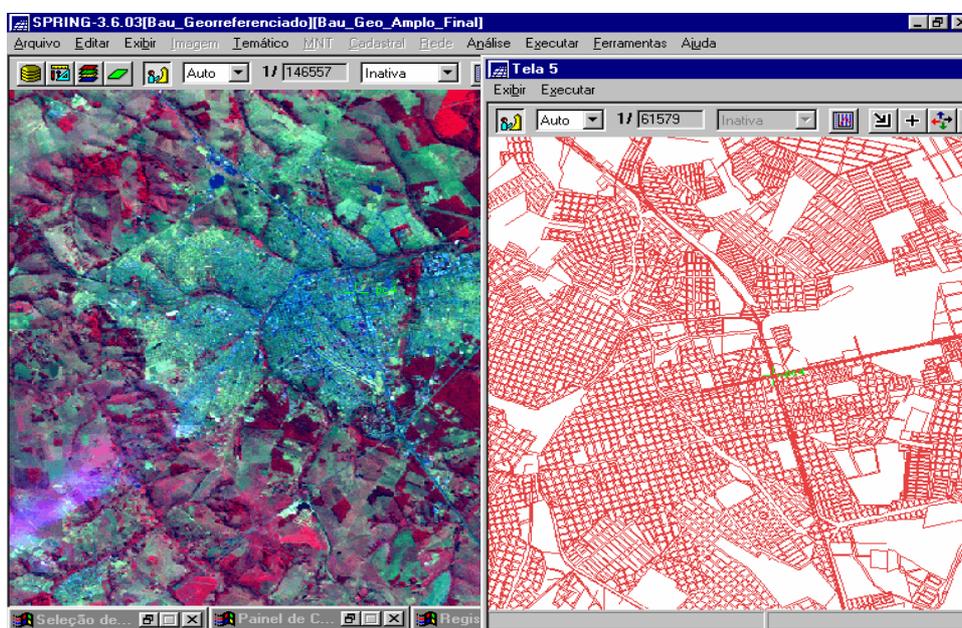


FIGURE 4.7 – Image to vector data registration for the Bauru city plan of 2000 (TM – 5 221/75, 06/07/00).

SOURCE: INPE (2000) and DAE (2000).

Maps of distance concerned both linear elements (roads, rivers, railways) and polygons (classes of land use, types of occupation density, etc). In some simulation periods, as we may check with more depth in Chapter 6, some input variables concerned maps of distance to the ranges of clusters density, defined upon basis of the Kernel points density estimator. This was the case, for instance, of the map of commercial activities in Bauru in 1979 (FIGURE 4.8). For the urban dweller, distances to main clusters of commercial activities are what really concern them, and not the individual distances to particular commercial establishments themselves.

Some maps were kept in binary format (presence/absence), like water supply, social housing and ranges of occupation density. In the latter case, each interval of occupation

density (0%-5%; 5%-10%; 10%-15%; 15%-25%; 25%-40%; 40%-70%; greater than 70%) was considered in different layers, i.e. different types of occupation density were represented in different maps.

Binary maps were converted to raster files with cells size 100 x 100 (m), which was also held in the maps of distances. The adopted resolution is about a city block size, what was deemed convenient for the purpose of urban land use analysis, for intra-city block variations in land use are disregarded.

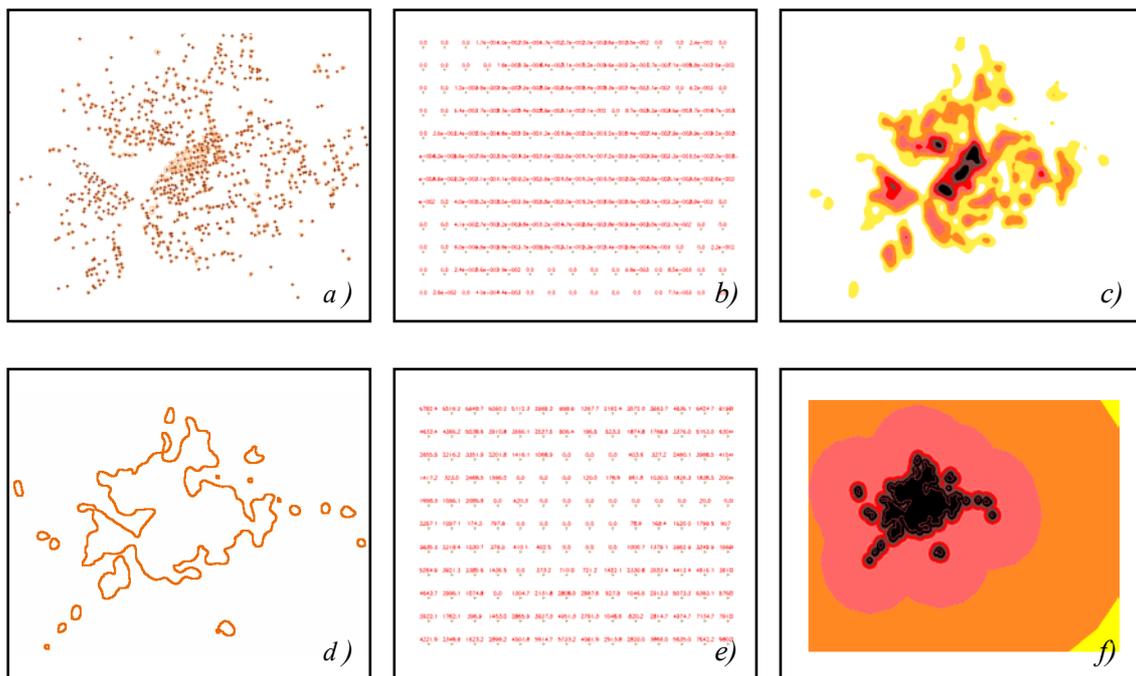


FIGURE 4.8 – Sequence of operations for generating a map of distance to commercial activities clusters: a) polygons of commercial activities and their centroids; b) Kernel density numerical grid (bandwidth = 5,000 m); c) Kernel density grid slicing; d) vector map of grouped slicing ranges; e) grid of distances to slicing ranges; f) map of distances slicing.

4.3.4 Updating the Land Use Maps through Remote Sensing Data

As previously stated in Section 4.1.3, the remotely sensed data were visually interpreted so as to update the official city maps, since illegal residential settlements are not shown

on the latter. This has been achieved by superimposing the city maps in vector format on the respective enhanced color composition images (FIGURE 4.9a and 4.9b). In fact, satellite imagery (and aerial photos) could have also been used to remove from the official city plans the legally approved settlements which are drawn but have not indeed been built. Given the fact that there is no ground truth (aerophotogrammetric survey) for the city of Bauru in 1967, this procedure has not been adopted due to standardization requirements throughout the time series. In the two cities, squatter settlements are very small and practically always inserted within regular residential areas.

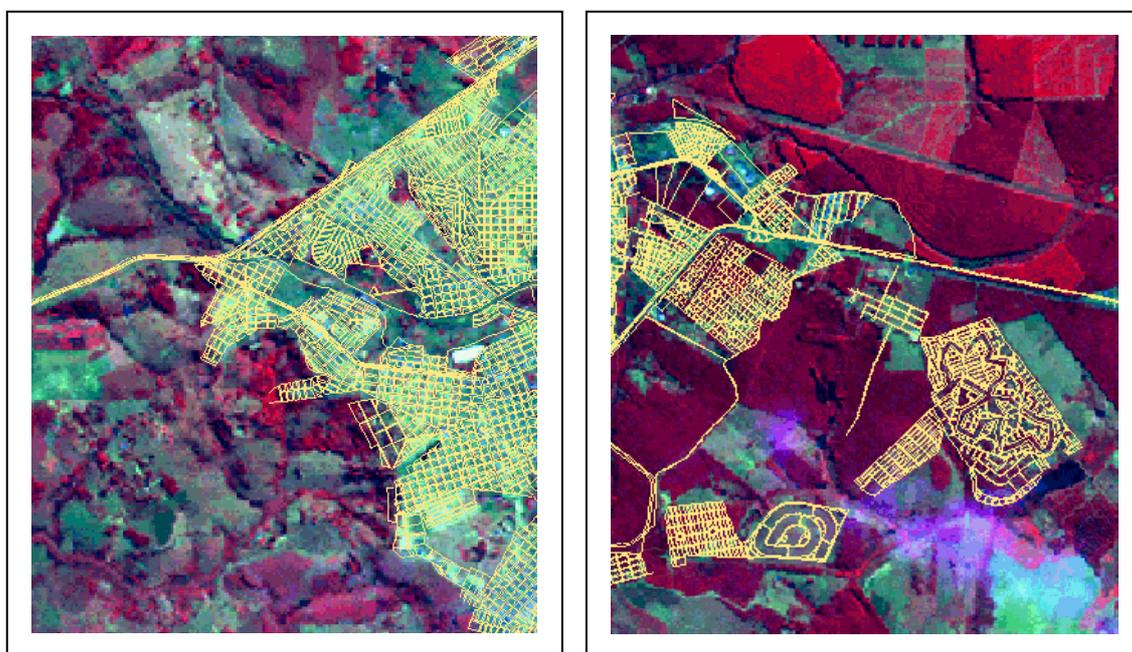


FIGURE 4.9a and 4.9b – Bauru city map superimposed on the TM – 5 image of Bauru, 1B_4R_7G, 221/75 – 06/07/00. Western city sector to the left (a) and eastern city sector to the right (b).

SOURCE: INPE (2000) and DAE (2000).

The city maps of Bauru contain little distortions in its western and eastern sectors, whereas Piracicaba city maps, as it can be checked in Chapter 6, present generalized distortions throughout the whole of its urban area. These deformities, however, are

moderate to subtle and do not harm the identification of illegal settlements, since this is not done automatically, but by visual analysis.

These distortions can be ascribed to integrity problems in the original cartographic data, such as:

- a) quality of stereoscopic procedures;
- b) low precision of conventional China ink drawings in the oldest city maps;
- c) expansions in the onionskins in face of moist and heat;
- d) updates of older city maps often accomplished by the mere addition of new aerial photos without the due stereoscopic accuracy care.

The inclusion of further control points well scattered through the image and the usage of higher orders polynomials could have yielded better matches between the images and their respective city maps to the detriment of undesirable distortions in the vector data.

4.3.5 Generalization Procedures Applied to the Land Use Maps

The maps provided by the Bauru and Piracicaba local authorities contained severe inconsistencies. Some urban zones refer to areas which are not yet occupied, and some other zones categories do not correspond to the prevailing use indeed encountered within their limits, reflecting just the local officials' intention for their future use. For instance, a zone designed as industrial may shelter in reality just a few small-sized industries, and be extensively occupied by residential settlements.

In this way, the following generalization procedures were applied to the Bauru and Piracicaba land use maps so as to render them agreeable with the reality to which they are related, and at the same time, workable by the computational model:

- a) reclassification of zones initially assigned by the cities local authorities according to their dominant and effectively existent use;
- b) reclassification of similar zones shown on official maps to only one category, e.g.: residential zones of different densities are all reclassified to simply

residential; special use zones and social infrastructure zones are reclassified to institutional zones only;

c) adoption of eight land use zone categories²¹: residential, commercial, industrial, services, institutional, mixed use zone, leisure/recreation, and non-urban use;

d) exclusion of districts segregated from the main urban agglomeration by more than 10 km from the official urban boundary; and,

e) disregard of minor non-occupied areas and traffic network in the simulations, for the latter is at a fine enough scale to be represented as a land use.

4.4 Conclusions

Building the geographic database was a complex task. The data preprocessing involved a great amount of operational proceedings, like standardization of non-spatial data, conversion of paper plans into digital data, georeferencing and digital processing of satellite images, cross-checking between remotely sensed and vector data, etc. Adjustments were made in the land use maps in order to make them more compatible with the indeed existing urban situation. In this case, loyalty to the reality observed in the land use contexts of the two cities has been preserved in a more generalized level.

All these procedures comply with the implicit reductionism of model building. Nevertheless, as stated by Briassoulis (2000), when models refuse to subject themselves to feasible generalization approaches, they are prone to be condemned either to total impracticability or ineffectiveness under the existent time and other resources constraints.

²¹ The mixed use zone basically comprises the residential, commercial, and services uses. The leisure/recreation zone includes parks, the city zoo and other public green areas. Institutional zones refer to areas sheltering great public equipments, like university campi, airports, railways support areas, and great hospital and other social infrastructure complexes.

CHAPTER 5

METHODS FOR THE URBAN LAND USE DYNAMICS SIMULATION MODEL

5.1 Introduction

This chapter will initially regard the two empirical statistical methods used to parameterize the land use change simulation model. Both of these methods – “weights of evidence” and “logistic regression” - are based on algebraic manipulations of the log function.

Statistical methods employed to determine the total amount of land use change in the forecast scenarios are presented in the third section. The stationary approach, in which urban land use transition rates are supposed to be constant throughout time, is given by the Markov chain or Markov model, whereas non-stationary procedures to estimate the global amount of change are provided by linear regression analyses, relating the rate of land use change to demographic trends and economic performance. These regression analyses are able to provide optimistic and pessimist forecast scenarios, upon basis of variations in the independent or explaining variables.

And finally, similarities and divergences between the “weights of evidence” and “logistic regression” as well as *pros* and *cons* in applying each of these two methods will be dealt with in the last section.

5.2 Simulation Methods

5.2.1 The Weights of Evidence Method

5.2.1.1 Introduction to the Weights of Evidence Method

The weights of evidence method is entirely based on the Bayes’ theorem of conditional probability (Bonham-Carter, 1994). Basically, this theorem concerns the favorability to

detect a certain event, which can be in the current case a given category of land use change (e.g. non-urban use to residential use), provided that an evidence (e.g. water supply area), also called explaining variable, has already happened (FIGURE 5.1).

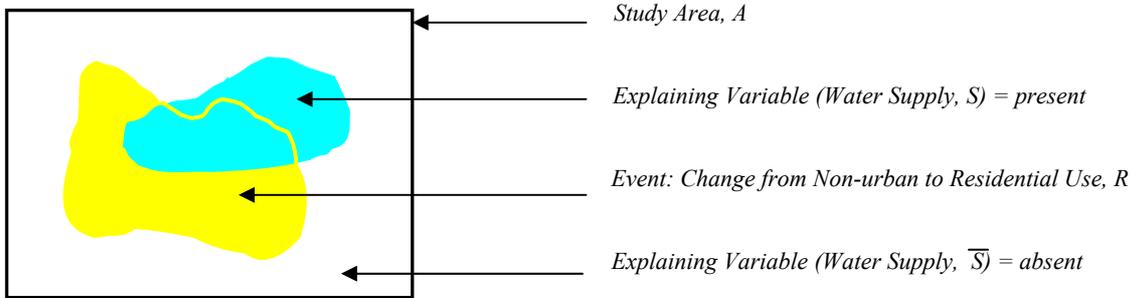


FIGURE 5.1 – Diagram to illustrate the weights of evidence method.

The favorability to find the event (change from non-urban to residential use) R given the presence of the evidence (water supply) S can be expressed by:

$$P \{R/S\} = \frac{P \{R \cap S\}}{P \{S\}} , \quad (5.1)$$

where $P \{R/S\}$ is the conditional probability of occurring the event R given the presence of the explaining variable or evidence S . But, $P \{R \cap S\}$ is equal to the proportion of the total area occupied by R and S together. Supposing N is the counting of map cells (area of an event or an evidence), then the above formula can be rewritten as:

$$P \{R/S\} = \frac{N \{R \cap S\}}{N \{S\}} . \quad (5.2)$$

In order to obtain an expression relating the *posterior* probability of the event R in terms of the *prior* probability and a multiplication factor, we note that the conditional

probability of being on the explaining variable map S , given the presence of the event R is defined as:

$$P \{S/R\} = \frac{P \{S \cap R\}}{P \{R\}} . \quad (5.3)$$

Since $P \{S \cap R\}$ is the same as $P \{R \cap S\}$, Equations (5.3) and (5.1) can be combined to solve for $P \{R/S\}$, satisfying the relationship:

$$P \{R/S\} = \frac{P \{R\} \cdot P \{S/R\}}{P \{S\}} . \quad (5.4)$$

A similar expression can be derived for the *posterior* probability of the event R occurring given the absence of the evidence \bar{S} . Thus,

$$P \{R/\bar{S}\} = \frac{P \{R\} \cdot P \{\bar{S}/R\}}{P \{\bar{S}\}} . \quad (5.5)$$

All the equations above can be expressed in an *odds* form. **Odds** are defined as a ratio of the probability that an event will occur to the probability that it will not occur. The weights of evidence method uses the natural logarithm of *odds*, known as *log odds* or **logits**. To clarify this approach, Equation (5.5) will be converted to *odds*. For this end, both sides will be divided by $P \{R/\bar{S}\}$, leading to:

$$\frac{P \{R/S\}}{P \{\bar{R}/S\}} = \frac{P \{R\} \cdot P \{S/R\}}{P \{\bar{R}/S\} \cdot P \{S\}} . \quad (5.6)$$

But from the definitions of conditional probability:

$$P \{ \bar{R}/S \} = \frac{P \{ \bar{R} \cap S \}}{P \{ S \}} = \frac{P \{ S/\bar{R} \} P \{ \bar{R} \}}{P \{ S \}} . \quad (5.7)$$

Substituting Equation (5.7) in Equation (5.6) yields the following:

$$\frac{P \{ R/S \}}{P \{ \bar{R}/S \}} = \frac{P \{ R \}}{P \{ \bar{R} \}} \cdot \frac{P \{ S \}}{P \{ S \}} \cdot \frac{P \{ S/R \}}{P \{ S/\bar{R} \}} . \quad (5.8)$$

Substituting *odds* into Equation (5.8) and canceling leads to the desired expression:

$$O \{ R/S \} = O \{ R \} \cdot \frac{P \{ S/R \}}{P \{ S/\bar{R} \}} , \quad (5.9)$$

where $O \{ R/S \}$ is the conditional (*posterior*) *odds* of R given S , $O \{ R \}$ is the *prior odds* of R and $P \{ S/R \} / P \{ S/\bar{R} \}$ is known as the *sufficiency ratio (LS)*. In weights of evidence, the natural logarithm of both sides of Equation (5.9) are taken, and $\log_e LS$ is the *positive weight of evidence* W^+ , which is calculated from the data. Then:

$$\text{logit} \{ R/S \} = \text{logit} \{ R \} + W^+ . \quad (5.10)$$

Similar algebraic manipulations lead to the derivation of an *odds* expression for the conditional probability of R given the absence of the evidence S , with the result being:

$$O \{ R/\bar{S} \} = O \{ R \} \cdot \frac{P \{ \bar{S}/R \}}{P \{ \bar{S}/\bar{R} \}} . \quad (5.11)$$

The term $P\{\bar{S}/R\} / P\{\bar{S}/\bar{R}\}$ is called the *necessity ratio* (LN). In weights of evidence, the *negative weight of evidence* W^- is the natural logarithm of LN , or $\log_e LN$. Thus in *logit* form, Equation (5.11) is:

$$\text{logit } \{R/\bar{S}\} = \text{logit } \{R\} + W^- . \quad (5.12)$$

LS and LN are also called *likelihood ratios*. When events and evidences are *positively correlated*, the value of LS is greater than 1, whereas LN is in the range $[0,1]$. However, if an evidence is *negatively correlated* with the events, LN would be greater than 1 and LS would be in the range $[0,1]$. If the evidence is uncorrelated with the events, then $LS=LN=1$, and the *posterior* probability equals the *prior* probability, and the probability of an event would be unaffected by the presence or absence of a certain evidence.

Similarly, W^+ is positive, and W^- is negative, due to the positive correlation between the evidences and the events. Conversely W^+ would be negative and W^- positive for the case where a very limited part of the event occur on the evidence area than would be expected due to chance. If the events are independent of whether the evidence is present or not, then $W^+ = W^- = 0$, and the *posterior* = the *prior*, as above (Bonham-Carter, 1994).

When the evidence from several maps is combined, the weights are calculated from each map independently, and then combined in a single equation. The conditional probability of an event occurring, given the presence of two predictive evidences, S_1 (water supply) and S_2 (sewerage supply) is:

$$P \{R/S_1 \cap S_2\} = \frac{P \{R \cap S_1 \cap S_2\}}{P \{S_1 \cap S_2\}} , \quad (5.13)$$

which can be written as

$$\begin{aligned}
P \{R/S_1 \cap S_2\} &= \frac{P \{S_1 \cap S_2/R\} \cdot P \{R\}}{P \{S_1 \cap S_2\}} \\
&= \frac{P \{S_1 \cap S_2/R\} \cdot P \{R\}}{P \{S_1 \cap S_2/R\} \cdot P \{R\} + P \{S_1 \cap S_2/\bar{R}\} \cdot P \{\bar{R}\}}. \quad (5.14)
\end{aligned}$$

This is Bayes' theorem. According to it, there are only two mutually exclusive hypotheses, R and \bar{R} , with $P \{R\} + P \{\bar{R}\} = 1$. The effects of interaction between S_1 and S_2 can be ignored by making an assumption of conditional independence. This provides a simplification, because it permits the effects of each evidence map to be evaluated individually and then combined by multiplying (or adding in the log-linear case) the factors for several maps together.

The conditional independence assumption can be stated as:

$$P \{S_1 \cap S_2/R\} = P \{S_1/R\} \cdot P \{S_2/R\}, \quad (5.15)$$

which allows Equation (5.14) to be simplified in the following form:

$$P \{R/S_1 \cap S_2\} = P \{R\} \cdot \frac{P \{S_1/R\}}{P \{S_1\}} \cdot \frac{P \{S_2/R\}}{P \{S_2\}}. \quad (5.16)$$

Using the *odds* formulation, the conditional or *posterior odds* can be expressed from:

$$O \{R/S_1 \cap S_2\} = O \{R\} \cdot LS_1 \cdot LS_2, \quad (5.17)$$

or with the log-linear weights of evidence from:

$$\text{logit } \{R/S_1 \cap S_2\} = \text{logit } \{R\} + W^+_{1} + W^+_{2} . \quad (5.18)$$

Whichever formulation of the model is used, there are four different ways of combining two evidence maps, the first being when both evidences are present (Equation 5.18); the other three ways are S_1 present and S_2 absent; S_1 absent and S_2 present; and both S_1 and S_2 absent. In the log-linear form, they can be written as:

$$\text{logit } \{R/S_1 \cap \bar{S}_2\} = \text{logit } \{R\} + W^+_{1} + W^-_{2} , \quad (5.19)$$

$$\text{logit } \{R/\bar{S}_1 \cap S_2\} = \text{logit } \{R\} + W^-_{1} + W^+_{2} , \quad (5.20)$$

$$\text{logit } \{R/\bar{S}_1 \cap \bar{S}_2\} = \text{logit } \{R\} + W^-_{1} + W^-_{2} . \quad (5.21)$$

With three evidences, there are 2^3 or 8 possible combinations and in general with n maps there are 2^n possible different combinations. The general expression for combining $i = 1, 2, \dots, n$ maps is either:

$$O \{R/S_1 \cap S_2 \cap S_3 \cap \dots S_n\} = O \{R\} \cdot \prod_{i=1}^n LS_i \quad (5.22)$$

for the likelihood ratios or

$$\text{logit } \{R/S_1 \cap S_2 \cap S_3 \cap \dots S_n\} = \text{logit } \{R\} + \sum_{i=1}^n W^+_{i} , \quad (5.23)$$

for the weights. In these general formulas, the LS becomes LN , and W^+ becomes W^- , if the i -th map pattern is absent instead of present. Where data is missing for a particular map layer in some locations, the likelihood ratio is set to 1, or the weight is set to 0. Equations (5.22) and (5.23) are the computing formulae for combining a set of binary maps with the Bayesian model.

According to Bonham-Carter (1994), the advantages of the Bayesian model are:

- the method is objective, and avoids the subjective choice of weighting factors;
- multiple maps of evidence can be combined with a model that is straightforward to program with a modeling language;
- input maps with missing data can be accommodated into the model;
- the possibility of incorporating multiclass input maps, where each map class is associated with a weight (or likelihood ratio);
- the modeling of uncertainty due to variances of weights or missing data.

And the disadvantages are:

- the combination of inputs maps assumes that the maps are conditionally independent of one another with respect to the response variable. The testing for conditional independence is only possible where the method is applied in a data driven mode, since it requires overlay data between pairs of evidence maps.
- weights of evidence, in common with other data-driven methods, is only applicable in regions where the response variable (event) is fairly well known.

5.2.1.2 Exploratory Analysis and Selection of Variables

Since the weights of evidence method is based on the Bayes' theorem of conditional probability as previously mentioned, the selection of variables for the modeling analysis should take into account the checking of independence amongst pairs of explaining variables or evidences chosen to explain the same category of land use change.

For this end, two methods were used: the Cramer's Coefficient (V) and the Joint Information Uncertainty (U). In both cases, it is necessary to obtain values from an area cross-tabulation between pairs of maps of variables under analysis. Let the area table between map A and map B be called matrix T , with elements T_{ij} , where there are $i = 1$,

2, ..., n classes of map B (rows of the table) and $j = 1, 2, \dots, m$ classes of map A (columns of the table). The marginal totals of T are defined as T_i for the sum of the i -th row, T_j for the sum of the j -th column, and $T_{..}$ for the grand total summed over rows and columns. If the two maps are independent of one another, with no correlation between them, then the expected area in each overlap category is given by the product between the marginal totals, divided by grand total. Thus the expected area T_{ij}^* for the i -th row and j -th column is:

$$T_{ij}^* = \frac{T_i \cdot T_j}{T_{..}} \quad (5.24)$$

Then, the **chi-square** statistic is defined as:

$$X^2 = \sum_{i=1}^n \sum_{j=1}^m \frac{(T_{ij} - T_{ij}^*)^2}{T_{ij}^*} \quad , \quad (5.25)$$

the familiar $(observed - expected)^2 / expected$ expression, which has a lower limit of 0 when the observed areas exactly equal the expected areas, and the two maps are completely independent. As the observed areas become increasingly different from the expected areas, chi-square increases in magnitude and has a variable upper limits. The Cramer's Coefficient (V) is then defined as (Bonham-Carter, 1994):

$$V = \sqrt{\frac{X^2}{T_{..} M}} \quad , \quad (5.26)$$

where M is the minimum of $(n-1, m-1)$.

The Joint Information Uncertainty (U) belongs to the class of **entropy** measures, which are also based on the area cross-tabulation matrix T , and can be used for measuring associations as well. Suppose that the T_{ij} values are transformed to area proportions, p ,

by dividing each area element by the grand total $T_{..}$. Thus, $p_{ij} = T_{ij}/T_{..}$, and the marginal proportions are defined as $p_{i.} = T_{i.}/T_{..}$ and as $p_{.j} = T_{.j}/T_{..}$. Therefore entropy measures, also known as *information statistics* can be defined using the area proportions as estimates of probabilities. Proportions are dimensionless, so entropy measures have the advantage over chi-squared measures of being unaffected by measurement units (Bonham-Carter, 1994).

Assuming that an area proportions matrix for map A and map B has been determined from T , then the *entropy* of A and B are defined as:

$$H(A) = - \sum_{j=1}^m p_{.j} \ln p_{.j} \quad \text{and} \quad (5.27)$$

$$H(B) = - \sum_{i=1}^n p_{i.} \ln p_{i.} \quad , \quad (5.28)$$

where \ln is the natural logarithm. The **joint entropy** of the combination, $H(A,B)$, is simply

$$H(A,B) = - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \ln p_{ij} . \quad (5.29)$$

Then the “**Joint Information Uncertainty**” of A and B, $U(A,B)$, can be used as a measure of association and is defined as

$$U(A,B) = 2 \left[\frac{H(A) + H(B) - H(A,B)}{H(A) + H(B)} \right], \quad (5.30)$$

which varies between 0 and 1. When the two maps are completely independent, then $H(A,B) = H(A) + H(B)$ and $U(A,B)$ is 0, and when the two maps are completely dependent, $H(A) = H(B) = H(A,B) = I$, and $U(A,B)$ is 1.

The criterion which is used to determine whether one factor is independent of another is to a large extent arbitrary as there is no large body of case results associated with the application of these methods (Almeida et al. 2003). Where this particular variant of *logit* modeling has been used in the geosciences, Bonham-Carter (1994) reports that values less than 0.5 for Cramer's Coefficient and the Joint Information Uncertainty suggest less association rather than more.

In practice, the variables selection routine also include empirical procedures in SPRING GIS, based on the visualization of distinct variables superimposed on the final land use map in vector format, so as to identify those more meaningful to explain the different types of land use change (FIGURE 5.2).

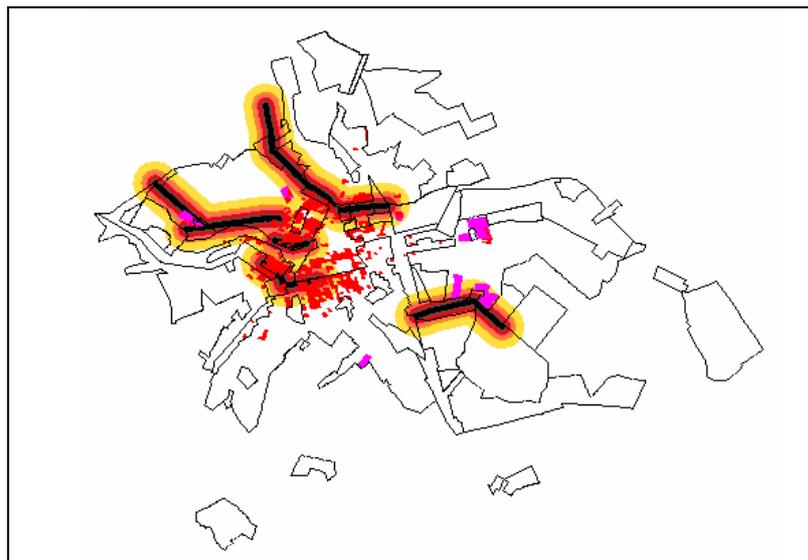


FIGURE 5.2 – Illustrative image showing the visual analysis on the identification of relevant factors determining the transition from residential to mixed use in Bauru from 1979 to 1988. The buffer bands are distance to planned roads; the red polygons are areas of medium-high density of occupation (40%-70% of built area per block); and the pink polygons correspond to social housing.

In the cases where high positive or negative correlations are identified, two evidence maps can be combined into one, through a Boolean operation, or one of them should be discarded, preferably the one with the smaller association with the response variable or event (Bonham-Carter, 1994).

A useful and commonly used measure of spatial association between an evidence (S) and a given event (R) is the contrast C , whose formula is given by:

$$C = W^+ - W^- \quad (5.31)$$

The question of determining whether the magnitude of the contrast is large enough to be statistically significant can be tested by the variance of the contrast, $S^2(C)$, estimated from the expression (Goodacre et al. 1993):

$$S^2(C) = \frac{I}{P \{S \cap R\}} + \frac{I}{P \{S \cap \bar{R}\}} + \frac{I}{P \{\bar{S} \cap R\}} + \frac{I}{P \{\bar{S} \cap \bar{R}\}} \quad (5.32)$$

5.2.1.3 Estimation of Global Transition Probabilities

The global transition probabilities refer to the total amount of change per type of land use transition in a given simulation period. Throughout a sufficiently long time series, the Markov chain can be used to estimate the global amount of change when data on urban land use in a certain period bound is missing, provided data on land use are available at least in the two preceding bounds (initial and final time of the former simulation period). The Markov chain will be approached in detail in Section 5.3.1.

Global transition rates, however, are estimated in the modeling experiments regarding past and present times by means of a cross-tabulation operation between the initial and final land use maps of each simulation period.

5.2.1.4 Calculation of Local Transition Probabilities

As exposed in Section 5.2.1.2, the *logit* of the conditional (*posterior*) probability of occurring a certain event (R) in face of “ n ” evidences (S) is given by a formula relating the *prior logit* of the event and the sum of the positive weights for each evidence or each class of evidence (Equation 5.23). In the simulation experiments of the current research, the local transition probabilities are calculated for each cell, represented by its x,y coordinates, and the equation used for this end converts the *logit* formula into a conventional conditional probability (Equation 5.33). The *odds* of R is purposely set to 1, so as to raise the final *posterior* probability value. In the denominator, values of e raised to the sum of positive weights of evidence related to other transitions (t) are included, in order to generate a more judicious value for the transition probabilities.

$$P_{x,y} \{R/S_1 \cap S_2 \cap \dots S_n\} = \frac{O \{R\} \cdot e^{\sum_{i=1}^n W_{x,y}^+}}{1 + O \{R\} \cdot \sum_{j=1}^t e^{\sum_{i=1}^n W_{x,y}^+}} \quad (5.33)$$

5.2.1.5 Model Calibration

Model calibration aims at the selection of the best set of input variables and internal software parameters, so as to produce the best fit between the empirical data and the observable reality. This is basically a twofold task. Firstly, a visual comparative analysis is carried out for each type of land use change, amongst the general trends of preliminary simulation results, the hints provided by maps showing transition probabilities and the exact area of land use transition, and the guideline information contained in the simultaneous overlay of different explaining variables maps upon the final land use map in vector format. This comparison envisages identifying those variables or evidences which are effectively concurring to explain the respective events from those which are just noise in the modeling.

An example of this visual comparative analysis is shown in FIGURE 5.3 for Bauru in relation to the “residential-services (res_serv)” land use change from 1979 to 1988. The transition probabilities map can be seen on the upper left corner; the land use transition map, on the upper right; a preliminary simulation result, on the lower left corner; the real final land use map, on the lower right corner; and a simultaneous overlay of the water supply map and the services axes distances map upon the borders of the final land use map, enabled by SPRING, is seen in the center. The elaboration of land use transition maps and transition probabilities maps will be further explained in Chapter 6.

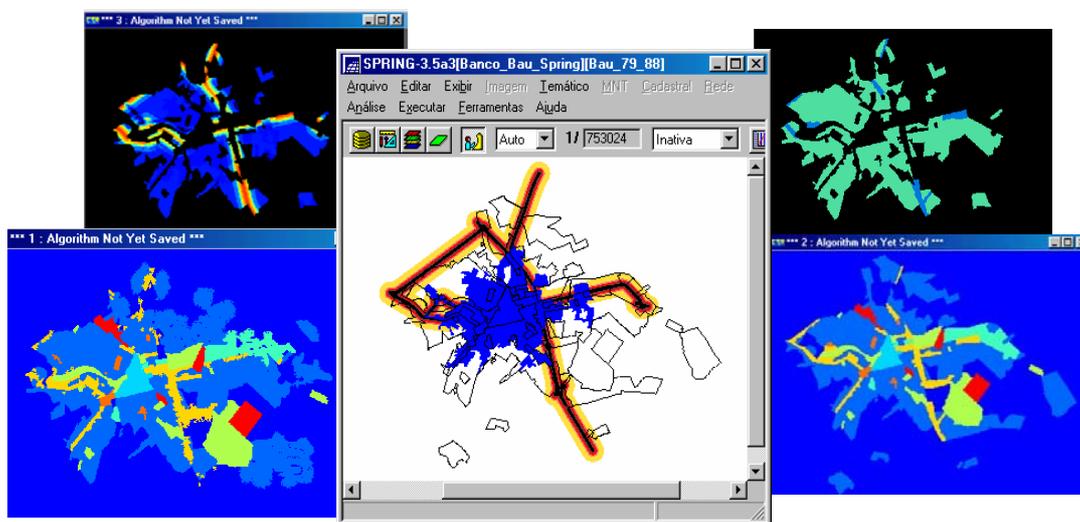


FIGURE 5.3 – Example of visual comparative analysis for the model empirical calibration in relation to the transition “residential-services (res_serv)” in Bauru, from 1979 to 1988.

The model calibration, on the other hand, is as well accomplished by the analysis of scatter plots relating subcategories of evidences (distances ranges), whenever they are available, with their respective positive weights of evidence. In a general manner, when the plots present a good fit of trendlines (which can assume different orders and types), i.e. when the lines do not demand very complex models for adjustment, the evidences to which they are associated are highly prone to be included in the model (FIGURE 5.4).

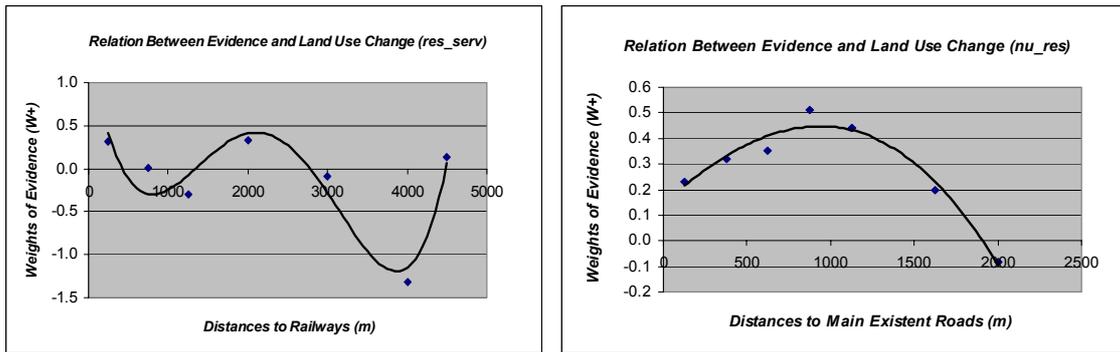


FIGURE 5.4 – Examples of scatter plots and respective trendlines for the relations between subcategories of evidences (*X axis*) and their corresponding positive weights of evidence (*Y axis*), considering different types of land use change.

The scatter plots above relate to the transitions “residential-services (res_serv)” and “non-urban-services (nu_serv)” held in the city of Bauru during the simulation period 1979 – 1988. The plot to the left hand side show a case of complex fit, and hence, of probable evidence exclusion. On the contrary, the plot to the right present a good adjustment of trendline, what implies the high probability of inclusion of such evidence in the urban land use dynamics model.

The scatter plots analysis is a mere ancillary tool in calibration and is thus not a decisive criterion in the variables selection. The final decision towards the inclusion or exclusion of a given variable or evidence will always rely upon a broad judgment, in which the environmental importance of the evidence and its coherence concerning the phenomenon (land use transition) being modeled are analyzed (Soares-Filho, 1998).

The second stage of model calibration concerns the setting of the simulation software internal parameters: size and variance of patches, number of iterations, proportions of the transitions algorithms. As it will be further approached in Section 5.4, heuristic techniques are employed to define these simulation model parameters upon basis of a visual comparative analysis amongst preliminary simulation results, the final land use map, transition probabilities and land use transition maps as shown in FIGURE 5.3.

5.2.1.6 Statistical Validation Test

Validation tests can be understood as procedures to verify whether or not the model results reflect reality to the desired degree (Batty, 1976). With the purpose of conducting statistical tests for the spatial validation of land use dynamics models, Constanza (1989) presents a procedure entitled “*Multiple Resolution Method*”, which can be applied to a wide variety of spatial resolutions through the change of size in a sampling window.

This sampling window moves over the entire images (FIGURE 5.5), and the average fit between two given scenes (the real and the simulated one) for a particular window size is calculated by the following expression:

$$F_w = \frac{\sum_{s=1}^{tw} \left[1 - \sum_{i=1}^p \frac{|a_{i1} - a_{i2}|}{2w^2} \right]_s}{tw}, \quad (5.34)$$

where F_w is the fit for the window of size $w \times w$; a_{i1} is the number of cells belonging to class i in scene 1 (simulated image) and a_{i2} is the number of cells belonging to class i in scene 2 (real image) in the sampling window; p refers to the number of different classes found in the sampling window and t_w , to the total number of windows sampled in a scene for a window size of $w \times w$.

For two identical scenes, a plot relating F_w and w will provide a straight line. But, if the scenes present the same proportion of land use classes with different spatial patterns, this line will gradually increase until F_w reaches value 1. When this happens, the sampling window will be identical to the scene under evaluation. However, if a reasonable patterns spatial fit exists, this curve will rapidly increase in an asymptotic way.

The total goodness of fit is then given by the equation below:

$$F_t = \frac{\sum_{w=1}^n F_w e^{-k(w-1)}}{\sum_{w=1}^n e^{-k(w-1)}}, \quad (5.35)$$

where F_t is the average of all fit measures obtained by the different window sizes employed in the analysis, F_w is the fit for sampling windows of linear dimension w , and k , a constant.

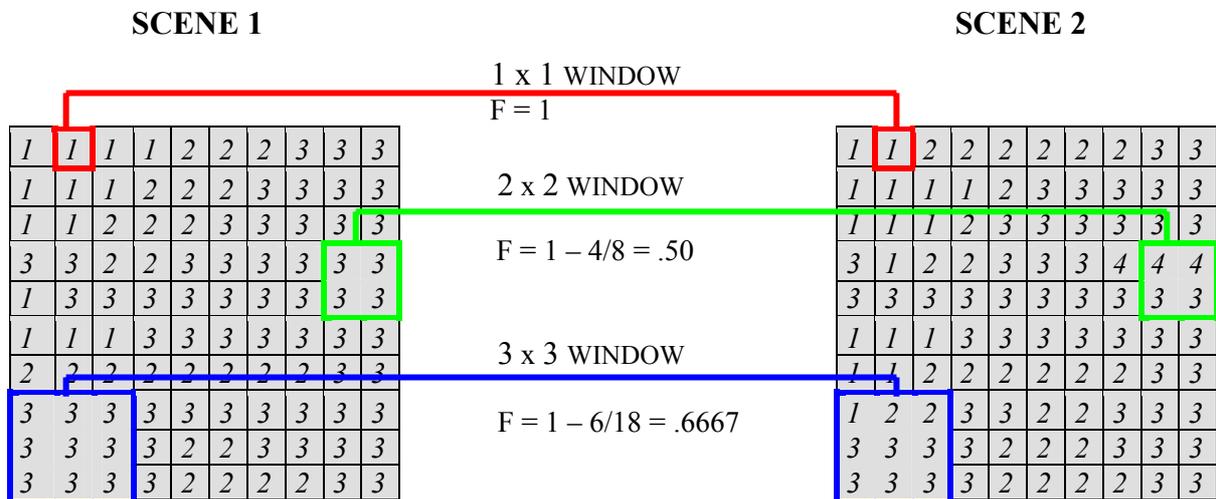


FIGURE 5.5 – Example of the multiple resolution method for a scene of size 10 x 10 pixels and with four classes. In this example, $k = 0.1$ and $F_t = 0.84$.

SOURCE: Constanza (1989).

When k is zero, all the window sizes have the same weight, whereas when $k=1$, only the bigger windows are important. According to Constanza (1989), the values of k can be adjusted in function of the model objective and the data quality.

This multiple resolution method was implemented in a UNIX environment program named FIT, developed by the Center for Remote Sensing of the Federal University of Minas Gerais, Brazil (CSR-UFMG).

Other spatial statistical validation models operating on multiple resolutions have been proposed by Pontius Jr. (2000, 2002), in which the error assessment is subdivided into quantification and location error. In his methods, the proportion of fit is separately calculated for areas that underwent change and for areas where no change took place in reality. Due to the fact that our simulation model contains an algorithmic routine that prevents areas of land use permanence to suffer change, Pontius's statistical validation methods could not have been applied.

5.2.2 The Logistic Regression Method

5.2.2.1 Introduction to the Logistic Regression Method

The logistic regression method is applicable to cases where the response or dependent variable is discrete, taking on two or more possible values, i.e. the variable owns a qualitative character. This method was originally conceived to respond to the needs of biomedical and public health sciences, in the sense that it aimed at modeling the drivers of pathologies in general.

A good example that illustrates a possible application of the logistic regression method is the incidence of coronary heart disease (CHD), held by Hosmer and Lemeshow (1989) for a sample group of 100 subjects with different ages. In this example, the outcome or response variable is binary or dichotomous, i.e. it presents two levels only, which are coded with a value of *zero* to indicate the CHD absence, or *1* to indicate that the disease is present in the individual. The conversion of a simple plot of presence/absence of CHD into a plot of the proportion of individuals with CHD versus the midpoint of each age interval, shows that the resulting curve presents the shape of an “s” (FIGURE 5.6). This is the so-called “logistic distribution” or “logistic function” curve, which shows constant or stationary trends in its extremes and a markedly linear intermediate behavior. It is noticeable that the occurrence frequency of CHD increases with age.

When the logistic regression method embodies more than two levels for the outcome variable, it is termed 'polytomous logistic regression model'. In this case, the outcome or response variable can assume n values ($n \in N$), i.e. $Y = 0, 1, 2, \dots, n$.

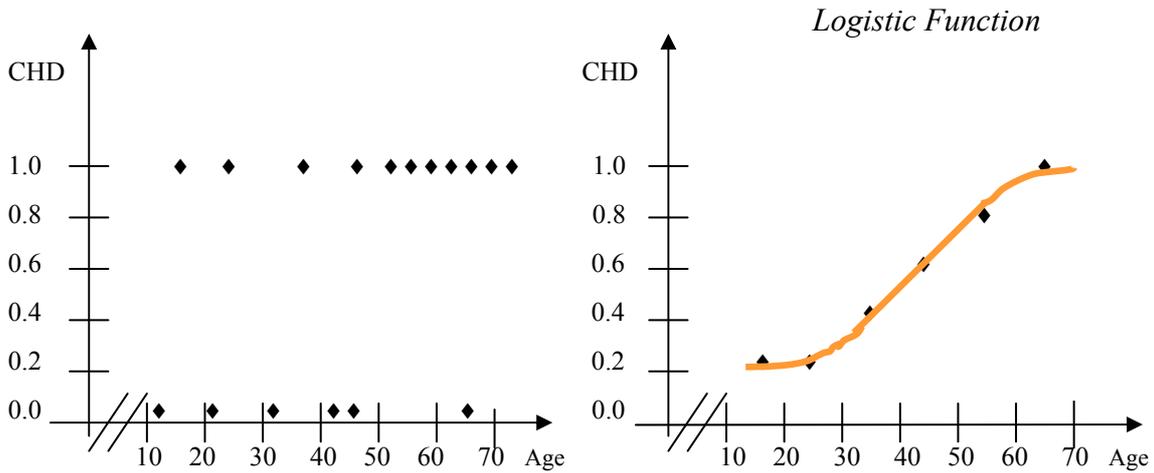


FIGURE 5.6 – Schematic plots showing the presence/absence of coronary heart disease (CHD) in relation to the individuals age (to the left), and the incidence frequency of CHD versus the midpoint of age intervals of ten years .
SOURCE: Adapted from Hosmer and Lemeshow (1989, p. 4, 5).

The binary logistic regression model ($Y=0$ or $Y=1$) consists in the extraction of the natural logarithm of the chance or *odds* in relation to the two levels of the outcome variable. As already defined in Section 5.2.1.1, *odds* are the ratio of the probability that an event will occur to its complementary probability, i.e. the probability that it will not occur. In the case of the binary regression, the *odds* are the ratio $P(1)/P(0)$. The *odds* logarithm or *logit* corresponds to a conventional linear uni or multivariate regression, which is then transposed to the logistic function. Equation (5.36) is an example of this log transformation for the case of a multivariate regression:

$$L = \log \left[\frac{P_{i,j}(x,y)}{1 - P_{i,j}(x,y)} \right] = \beta_{0,ij} + \beta_{1,ij} \cdot V_{1,xy} + \dots + \beta_{k,ij} \cdot V_{k,xy} , \quad (5.36)$$

$$P_{i,j}(x,y) = \frac{e^L}{1 + e^L}, \quad (5.37)$$

where i and j represent cell states or cell land use classes; x,y indicate a given cell in function of its location coordinates; and V accounts for the k independent variables selected to explain the transition from state i to state j .

The polytomous logistic regression model, on its turn, aggregates partial binary logistic regressions, where the *logit* always adopts as denominator one of the levels of the outcome variable, which is called reference level. The reference level is determined by the modeler and is usually a level whose behavior distinguishes itself from the behavior of the other levels. The reference level can also be chosen in function of advantages it may offer to the regression analysis in being compared with the remaining levels of the dependent variable.

Below are the formulas to calculate the probability in a polytomous logistic regression model, in which the dependent variable assumes three levels (0, 1 and 2), and where 0 is elected as the reference level. Equations (5.38) and (5.39) correspond to the *logits*, and Equations (5.40), (5.41) and (5.42) present the probability calculations properly speaking.

$$g_1(x) = \log \left[\frac{P(Y=1 / X_{1-p})}{P(Y=0 / X_{1-p})} \right], \quad g_1(x) = \beta_{10} + \beta_{11} \cdot X_1 + \dots + \beta_{1p} \cdot X_p \quad (5.38)$$

$$g_2(x) = \log \left[\frac{P(Y=2 / X_{1-p})}{P(Y=0 / X_{1-p})} \right], \quad g_2(x) = \beta_{20} + \beta_{21} \cdot X_1 + \dots + \beta_{2p} \cdot X_p \quad (5.39)$$

$$P (Y=0/X_{1-p}) = \frac{1}{1 + e^{g^1(x)} + e^{g^2(x)}} \quad (5.40)$$

$$P (Y=1/X_{1-p}) = \frac{e^{g^1(x)}}{1 + e^{g^1(x)} + e^{g^2(x)}} \quad (5.41)$$

$$P (Y=2/X_{1-p}) = \frac{e^{g^2(x)}}{1 + e^{g^1(x)} + e^{g^2(x)}} \quad (5.42)$$

The general method of estimation adopted for the logistic regression models is the maximum likelihood. In a very general sense, this method yields values for the unknown parameters (β_i) which maximize the probability of obtaining the observed set of data (Hosmer and Lemeshow, 1989). For a binary regression model, the likelihood function (l) is obtained from:

$$l (\beta) = \prod_{i=1}^n \{ \pi (x_i)^{y_i} [1 - \pi (x_i)]^{1-y_i} \}, \quad (5.43)$$

where

$$\pi (x_i) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}} \quad (5.44)$$

Since it is easier mathematically to work with the log of Equation (5.43), the log likelihood is defined as:

$$L(\beta) = \ln [l(\beta)] = \sum_{i=1}^n \{ y_i \ln[\pi(x_i)] + (1 - y_i) \ln [1 - \pi(x_i)] \} \quad (5.45)$$

To find the value of β that maximizes $L(\beta)$ we differentiate $L(\beta)$ with respect to β_0 and β_1 and set the resulting expressions equal to zero. These equations are as follows:

$$\sum_{i=1}^n \left[y_i - \pi(x_i) \right] = 0 \quad (5.46)$$

and

$$\sum_{i=1}^n x_i \left[y_i - \pi(x_i) \right] = 0 . \quad (5.47)$$

For a polytomous logistic regression model, where the response variable presents three levels (0, 1 and 2), the likelihood function (l) is obtained from:

$$l(\beta) = \prod_{i=1}^n \left[\pi_0(x_i)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}} \right]. \quad (5.48)$$

Taking the log and using the fact that $\sum y_{ji} = 1$ for each i , the log likelihood function is:

$$L(\beta) = \sum_{i=1}^n y_{1i} g_1(x_i) + y_{2i} g_2(x_i) - \ln(1 + e^{g_1(x_i)} + e^{g_2(x_i)}). \quad (5.49)$$

The likelihood equations are found by taking the first partial derivative of $L(\beta)$ with respect to each of the 2 ($p + 1$) unknown parameters. In order to simplify the notation, let $\pi_{ji} = \pi_j(x_i)$. The general form of these equations is as follows:

$$\frac{\partial L(\beta)}{\partial \beta_{jk}} = \sum_{i=1}^n x_{ki} (y_{ji} - \pi_{ji}), \quad (5.50)$$

for $j = 1, 2$ and $k = 0, 1, 2, \dots, p$; where $x_{0i} = 1$ for each observation. The maximum likelihood estimator, $\hat{\beta}$, is obtained by setting these equations equal to zero and

solving for β . In both cases of binary and polytomous logistic regression models, the solution of likelihood equations requires iterative methods, usually available in statistical packages.

The goal of any statistical regression model is to select the minimum and at the same time the best set of input variables to explain a given phenomena (Neter and Wasserman, 1974). In other words, the aim is to extract the most parsimonious set of independent variables to integrate the final regression model. To meet this end, the method “backward stepwise” has been adopted for selecting the final set of independent variables. The initial model included all variables and excluded the least significant variable at each step. Significance was based on the Wald chi-square test and the G statistics. The Wald test is obtained by comparing the maximum likelihood estimate of the slope parameter, $\hat{\beta}_i$, to an estimate of its standard error. The G statistics, on its turn, evaluates the model by comparing it with and without a given independent variable (Hosmer and Lemeshow, 1989). The formulas for these significance tests are given below:

$$W = \frac{\hat{\beta}_i}{(SE) \hat{\beta}_i} \quad (5.51)$$

and

$$G = - 2 \{L (\beta_i) - [n_1 \ln (n_1) + n_0 \ln (n_0) - n \ln (n)]\} \quad , \quad (5.52)$$

where

$$n_1 = \sum y_i \quad , \quad n_0 = \sum 1 - y_i \quad \text{and} \quad n = n_1 + n_0 .$$

The model is accepted when all independent variables are significant at the 0.05 level and the loss of the G statistics remains lower than 5% (Soares-Filho et al. 2001).

One of the first issues raised in conducting modeling experiments based on logistic regression concerns the way the outcome or response variable should be modeled. In the case of selecting a polytomous regression model, the land use permanences would be regarded as the reference level (and could be assigned value 0), and the different land use changes would assume natural values, ranging from 1 to n , given the n existent types of land use transitions observed in the simulation period under study.

A second alternative could be the adoption of partial polytomous models, where transitions owning identical origin states would be handled in the same model. In this way, a certain simulation period would present different polytomous models. For instance, land use transitions like 'non-urban to residential' and 'non-urban to industrial use' would account for different levels of the outcome variable in one of the models, whereas the permanence of non-urban areas as such would correspond to the reference level of this outcome variable. In another model relating to the same simulation period, transitions like 'residential to services' and 'residential to mixed use' would represent levels 1 and 2 of the response variable, while the permanence in the residential use would be regarded as the reference level, and so forth.

A third and last solution concerns the usage of binary models for each type of land use transition observed in a simulation period. According to this procedure, value 1 is assigned to the considered transition, and value 0 , to the respective land use permanence.

In the current research, the logistic regression model has been experimentally employed for the city of Bauru only in the simulation period 1979 to 1988. Applying this model to all simulation periods would be time-consuming and unnecessary for purposes of comparative analysis with the weights of evidence method. The binary modeling has been adopted due to the fact that, in each case, only the independent variables selected to explain the respective transition are considered. As a consequence, the insertion of noise is avoided by the fact that variables designed to explain other transitions are excluded from the model. Furthermore, the binary modeling complies with the

algorithmic logic of the simulation model, in which each land use transition has its calibration parameters individually adjusted.

In conclusion, it is worth remarking that logistic regression is integrated into the class of generalized linear models – GLM (McCullagh and Nelder, 1989). In these models, there is a natural parameter to be estimated, for which the asymptotic convergence of estimators is very fast. This natural parameter $[g(x)]$ is expressed as a linear function of covariates (independent variables). Thus, the logistic regression is regarded as a robust method, since it operates upon linear regressions.

5.2.2.2 Exploratory Analysis and Selection of Variables

In the logistic regression model as in linear regression methods in a general way, the selection of independent variables is based on an analysis of their pair-wise correlation indices with the outcome variable as well as on the correlation between pairs of independent variables themselves.

A useful measure to evaluate colinearity is the correlation matrix, which provides basic information on the model input data, indicating the degree of association between two independent variables and between an independent variable and the outcome variable. The correlation index is obtained from the concept of covariance ($A_{A,B}$), which measures the total mean of the sum of the products between deviations of variables belonging to two numerical data sets (A,B) in relation to their respective means:

$$\Delta_{A,B} = 1/N \sum_{i=1}^N (x_A(i) - \bar{x}_A(i)) (x_B(i) - \bar{x}_B(i)) \quad . \quad (5.53)$$

The correlation index ($\alpha_{A,B}$), on its turn, indicates association between two numerical data sets upon an absolute scale. In brief, it normalizes the covariance within the range $[-1, +1]$, where values close or equal to -1 indicate negative correlation, and values approaching or equal to $+1$, denote positive correlation. This index is calculated dividing the covariance by the square root of the product between the variances of the two data sets:

$$\alpha'_{A,B} = \frac{\Delta_{A,B}}{\sqrt{\sigma_A^2 \cdot \sigma_B^2}} \quad (5.54)$$

Another important aspect in the exploratory analysis regards the existence of interaction and confounding. Interaction is present when two independent variables, which present no correlation between each other, can together enhance the outcome variable. Graphically, the absence of interaction yields a model with two parallel lines, one for each independent variable. Similarly, the presence of interaction between two independent variables is identified by two non-parallel lines. FIGURE 5.7 is an illustrative plot of an independent variable relating to distance in the X axis, versus the *logits* of other independent variables (l_1 , l_2 and l_3).

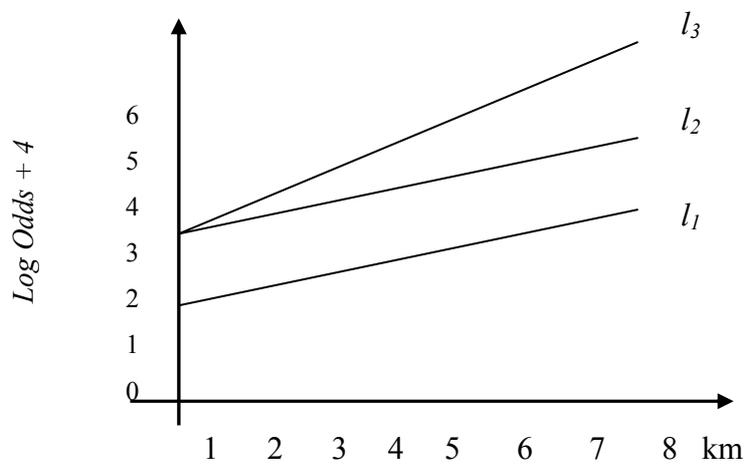


FIGURE 5.7 – Illustrative plot of the *logits* under three different models showing the presence (l_1 – l_3) and absence (l_1 – l_2) of interaction.

SOURCE: Adapted From Hosmer and Lemeshow (1989, p. 65).

The term confounder is used to describe a covariate that is associated with both the outcome variable of interest and an independent variable. Except in the cases where the

confounder and the independent variable associated with it are both decisive for explaining the response variable, one of them should be discarded from the model.

The right way of dealing with an interaction term is to include the two independent variables, provided at least one of them is significant to integrate the model, and also the product between these variables. And for the cases when the confounder and the independent variable associated with cannot be excluded from the model, a product between these two terms should be included as well.

5.2.2.3 Estimation of Global Transition Probabilities

Likewise the weights of evidence method, the Markov chain can be employed to estimate the global amount of change when data on urban land use in a certain period bound is missing, provided data on land use are available at least in the two preceding bounds (initial and final time of the former simulation period).

Global transition rates are estimated in the modeling experiments regarding past and present times by means of a cross-tabulation operation between the initial and final land use maps of each simulation period.

5.2.2.4 Calculation of Local Transition Probabilities

According to Equations (5.36) and (5.37) exposed in Section 5.2.2.1, the formula for estimating the land use transition probability of a given cell takes into account the logistic function and presents the linear regression model as the exponent of the term “ e ”.

Although the binary modeling has been adopted for the estimation of the regression parameters in the simulations of the present work, the formula employed for the calculation of cells transition probabilities (Equation 5.55) is very similar to the formula of the polytomous regression model, for it inserts in the denominator values of e raised to linear regression models referring to other land use transitions (i). This mathematical artifact helps in preventing probability values from ending in a tie and produces more sensible values for the transition probabilities.

$$P_{i,j}(x,y) = \frac{e^{\beta_0 + \gamma_{i,j} \cdot V_{(x,y)}}}{1 + \sum_{i=1}^t e^{\beta_0 + \gamma_{i,j} \cdot V_{(x,y)}}}, \quad (5.55)$$

where t refers to transitions other than i,j , like i,k ; i,l ; k,l ; etc.; $\gamma_{i,j}$ is a vector of parameters: $\gamma_{i,j} = [\beta_1, \beta_2, \dots, \beta_n]$; and $V_{(x,y)}$ a vector of independent variables:

$$V_{(x,y)} = \begin{bmatrix} V_1 \\ V_2 \\ \cdot \\ \cdot \\ V_n \end{bmatrix}.$$

5.2.2.5 Model Calibration

According to what was stated in Section 5.2.2.1, the Wald chi-square and the G statistics significance tests correspond to a sort of model calibration, for they contribute to define whether an independent variable remains in or is excluded from the final logistic regression model.

Other resources for defining as to maintaining or changing the set of selected variables are goodness-of-fit measures used in logistic models and usually available in statistical packages, such as the Pearson chi-square, the deviance and the Hosmer-Lemeshow tests.

Goodness-of-fit is assessed over the constellation of fitted values determined by the covariates (independent variables) in the model, not the total collection of covariates. For instance, supposing that a given fitted model contains p independent variables, $x' = (x_1, x_2, x_3, \dots, x_p)$, and let J denote the number of distinct values of x observed. If some

observations have the same value of x then $J < n$. The number of observations with $x = x_j$ will be denoted by $m_j, j = 1, 2, 3, \dots, J$. It follows that $\sum m_j = n$. Let y_i denote the number of positive responses, $y = 1$, among the m_j observations with $x = x_j$. It follows that $\sum y_j = n_1$, the total number of observations with $y = 1$. The distribution of the goodness-of-fit statistics is obtained by letting n become large. If the number of covariate patterns also increases with n then each value of m_j will tend to be small. Distributional results obtained under the condition that only n becomes large are said to be based on *n-asymptotics*. If $J < n$ is fixed and let n become large then each value of m_j will tend to become large. Distributional results based on each m_j becoming large are said to be based on *m-asymptotics* (Hosmer and Lemeshow, 1989).

The fitted values in logistic regression are calculated for each covariate pattern and depend on the estimated probability for that covariate pattern. Let \hat{y}_j be denoted as the fitted value. This fitted value equals:

$$m_j \hat{\pi}_j = m_j (exp [\hat{g}(x_j)] / \{1 - exp [\hat{g}(x_j)]\}), \quad (5.56)$$

where $\hat{g}(x_j)$ is the estimated *logit*. Two measures of the difference between the observed and the fitted values, the Pearson residual and deviance residual, will be considered. The Pearson residual is defined as follows:

$$r(y_j, \hat{\pi}_j) = \frac{(y_j - m_j \hat{\pi}_j)}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}}. \quad (5.57)$$

The summary statistic based on these residuals is the Pearson chi-square statistic:

$$X^2 = \sum_{j=1}^J r(y_j, \hat{\pi}_j)^2. \quad (5.58)$$

The deviance residual is defined as follows:

$$d(y_j, \hat{\pi}_j) = \pm \left\{ 2 \left[y_j \ln \left(\frac{y_j}{m_j \hat{\pi}_j} \right) + (m_j - y_j) \ln \left(\frac{(m_j - y_j)}{m_j (1 - \hat{\pi}_j)} \right) \right] \right\}^{1/2}, \quad (5.59)$$

where the sign is the same as the sign of $(y_j - m_j \hat{\pi}_j)$. For covariate patterns with $y_j = 0$ the deviance residual is:

$$d(y_j, \hat{\pi}_j) = - \sqrt{2m_j \left| \ln(1 - \hat{\pi}_j) \right|}. \quad (5.60)$$

and the deviance residual when $y_j = m_j$, is:

$$d(y_j, \hat{\pi}_j) = - \sqrt{2m_j \left| \ln(\hat{\pi}_j) \right|}. \quad (5.61)$$

The summary statistic based on the deviance residuals is the deviance:

$$D = \sum_{j=1}^J d(y_j, \hat{\pi}_j)^2. \quad (5.62)$$

The Hosmer-Lemeshow goodness-of-fit statistic, \hat{C} , is obtained by calculating the Pearson chi-square statistic from the $2 \times g$ table of observed and estimated expected frequencies. A formula defining the calculation of \hat{C} is as follows:

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n_k \bar{\pi}_k)^2}{n_k \bar{\pi}_k (1 - \bar{\pi}_k)}, \quad (5.63)$$

where n_k' is the total number of observations in the k^{th} group,

$$o_k = \sum_{j=1}^{c_k} y_j, \quad (5.64)$$

is the number of responses among the c_k covariate patterns, and

$$\bar{\pi}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n_k'}, \quad (5.65)$$

is the average estimated probability and c_k denotes the number of covariate patterns in the k^{th} decile.

The results provided from the above-mentioned statistics tests should be used in a critical and wise way. The heuristic procedures presented in Section 5.2.1.5 should be also used in the case of logistic regression modeling, given that visual analysis is decisive for identifying success and errors in model calibration.

Not only goodness-of-fit tests but also tests of statistical significance (W and G) should be regarded as non-exclusive criteria for the insertion or removal of independent variables in a logistic regression method. According to Hosmer and Lemeshow (1989, p.32), "...we must not base our models entirely on tests of statistical significance...there are numerous other considerations that will influence our decision to include or exclude variables from a model."

5.2.2.6 Statistical Validation Test

In some situations it may be possible to exclude a subsample of observations, develop a model based on the remaining data, and then test the model in the originally excluded observations. In other situations, it may be possible to obtain a new sample of data to assess the goodness-of-fit of a previously developed model.

Fitted models always perform in an optimistic manner on the developmental data set. This is the cause why it is preferable to apply tests where the fitted model is considered to be theoretically known, but no estimation has been performed yet (Hosmer and Lemeshow, 1989).

In the particular case of the experiments carried out in this research, this type of validation tests is not effective. The reason for this can be ascribed to the fact that this is a spatial simulation experiment, and validation tests must necessarily take into account the spatial dimension of the modeling outputs. In this way, only the multiple resolution approach proposed by Constanza (1989), introduced in Section 5.2.1.6, has been applied to the logistic regression model outputs for validation purposes.

5.3 Forecast Methods

5.3.1 The Markov Chain and the Future Transition Probabilities Matrix

A Markov chain is a mathematical model for describing a certain type of process that moves in a sequence of steps throughout a set of states (JRC and ESA, 1994). The Markov model can be expressed in matrix notation as (Baker, 1989):

$$\prod(t + 1) = P^n \cdot \prod(t), \quad (5.66)$$

where $\prod(t)$ is a column vector, with k elements, representing the fraction of land area in each of the s states at time t , n is the number of time steps between (t) and $(t+1)$. If n corresponds to six months, then n would be 2 in the above formula, considering that the addition in time corresponds to 1 year. $\prod(t+1)$ is a column vector showing the fraction of occupation of s states at time $t+1$, and P^n is a matrix whose elements are transition probabilities P_{ij} , accounting for the probability of a certain cell to change from state i to state j during the time interval $t \rightarrow t+1$.

The attractiveness of Markov chain analysis is that the model's parameters are easily estimated. The transition probabilities can be statistically estimated from a sample of transitions occurring during some time interval. Given data a_{ij} indicating transitions between pairs of states over some time interval, the transition probabilities P_{ij} are readily estimated as (JRC and ESA, 1994):

$$P_{ij} = a_{ij} / \sum_j a_{ij} . \quad (5.67)$$

In this way, a Markov chain only requires the determination of a finite number of states and that the transition probabilities be known. Although its relative simplicity, several constraints and assumptions are associated with the employment of Markov models to simulate land use change.

An important limitation of Markov chain models lies in the assumption that the probability of a particular set of outcomes depends only on the current distribution among states and on the transition probabilities – i.e. that Markov chain is a first-order process. However, according to JRC and ESA (1994), it is also possible to define chains whose dependency relationship involve more than one preceding state. A double dependence chain for example is dependent on two preceding states. If these two states are the two immediately preceding ones, the chain is a second-order chain. However, in that case, projecting future behavior would be much more difficult (Bell and Hinojosa, 1977).

Another assumption, which is not always suitable in the light of empirical knowledge of land use change phenomenon, is the stationarity of the transition matrix – i.e. temporal homogeneity (JRC and ESA, 1994). If this assumption holds, then consecutive iterations between the states column vector and the transition matrix n times would result in a vector representing the system states at time $(t+n)$. If this vector converges towards a limit probability distribution among the possible states of the system regardless of its initial condition, then the Markov chain is said to be stationary or ergodic (Facelli and Steward, 1990).

A Markov chain is thus ergodic if it presents a finite number of states, its dynamics is non-periodic and it has no absorbing states (when $P_{ij} = I$). In this case, it is possible to calculate the system state in a hypothetical future equilibrium. According to Bell and Hinojosa (1977), this can be obtained by the principal components method:

$$P = H V H^{-1} , \quad (5.68)$$

where H is the eigenvector matrix, H^{-1} is the transposed eigenvector matrix, and V is the eigenvalue matrix.

In this way, P can be decomposed in:

$$P^n = H V^n H^{-1} , \quad (5.69)$$

where n is the number of time steps. When the first eigenvalue is equal to I and the others smaller than I , for $n \rightarrow \infty$, it is obtained:

$$P^\infty = H \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} H^{-1} , \quad (5.70)$$

whose result corresponds to the proportion of equilibrium among the system states. As stated by Facelli and Steward (1990), the ratio of the first eigenvector to the second eigenvalue could still be used to foresee the time in which the system will reach equilibrium.

According to Baker (1989), the non stationarity alone does not preclude the use of a Markov chain approach, since “even if transitions are in reality non-stationary, stationarity can be assumed as a heuristic device.”

Markov chains can be adjusted to incorporate higher order effects, like the influence of endogenous and exogenous variables, spatial effects and heterogeneity. The

contribution of both endogenous and exogenous variables for the land use transitions, stationary or non-stationary, can be modeled using the following approach, in which Equation (5.66) is modified to (Baker, 1989):

$$\vec{X}(t + 1) = P [f(t)] \cdot \vec{X}(t) , \quad (5.71)$$

where P is a matrix with elements P_{ij} , with $P_{ij} = b_1X_1 + b_2X_2 + \dots + b_nX_n$ and $b_1 \dots b_n$ are the parameters that relate P_{ij} to the variables X_1, X_2, \dots, X_n . Following this line of thought, X_1, \dots, X_n can represent endogenous and exogenous variables. In the case of urban land use change modeling, endogenous variables correspond to the availability of technical and social infrastructure, types of urban occupation density, relief, urban zoning, etc., while exogenous variables refer to breakages in the general economic trends (economic or financial crises, energetic shortages, etc.); climatic disturbances which affect agricultural or tourism activities; local or regional policies that may impact the expected performance of the different economic sectors, etc.

In conclusion, it is important to restate here that the Markov chain is suitable for generating forecasts of urban land use change. It can also be used to estimate the global amount of change when data on urban land use in a certain period bound is missing, provided data on land use are available at least in the two preceding bounds (initial and final time of the former simulation period).

The principal components method applied to the Markov model is helpful in enabling a decomposition of the transition matrix probabilities, which are then estimated for narrower time lags, e.g. steps of one year or less. This has been used in the simulations in order to generate yearly land use maps for the whole time series of the two cities under study.

And finally, possible ways to overcome the Markov chain stationarity, as proposed by Baker (1989), will be approached in the next section.

5.3.2 Linear Regression Models for the Parameterization of Future Land Use Transition Probabilities

5.3.2.1 Introduction to Linear Regression Models

Regression analysis is a statistical tool which utilizes the relation between two or more quantitative variables so that one variable can be predicted from the other, in the case of univariate models, or from others, in the case of multivariate models (Neter and Wasserman, 1974).

A statistical relation, unlike a functional relation, is not a perfect one. In general, the observations for a statistical relation do not fall directly on the curve of relationship (FIGURE 5.8).

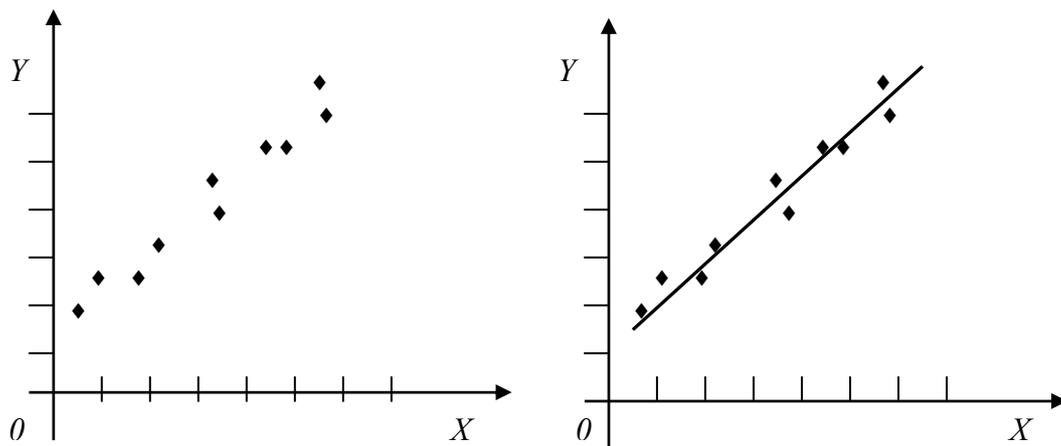


FIGURE 5.8 – Illustrative plots of statistical relations without and with a fitted linear regression model.

SOURCE: Adapted from Neter and Wasserman (1974, p. 24).

In multivariate models, it may be possible that the relationship between one or more independent variables (X_i) and the response variable (Y) are not linear. When this

happens, these relationships are linearized through mathematical transforms in order to fit them in the regression model.

The linear regression model in the multivariate case is obtained from the following general equation:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad , \quad (5.72)$$

where Y_i is the response obtained in the i -th observation; $\beta_0, \beta_1, \dots, \beta_{p-1}$ are parameters to be estimated; $X_{i1}, \dots, X_{i,p-1}$ are known variables; ε_i are the independent errors with normal distribution, with mean equals to zero and constant variance – $N(0, \sigma^2)$; i are the observations, $i = 1, 2, \dots, n$.

The response function of the model, which is the average of several observations, is given by:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1} \quad . \quad (5.73)$$

The parameter β_0 refers to the intercept of the regression plane. If none of the independent variables can assume value zero, β_0 loses its meaning. Yet, many statisticians prefer to include it in the final model for reasons of mathematical fit.

The parameter β_1 corresponds to the average change in the outcome variable $E(Y)$ for each increment unit in the value of X_1 , holding all other independent variables constant, and so on for the remaining parameters associated with explicative variables.

5.3.2.2 Exploratory Analysis

One of the first steps in conducting a linear regression analysis is the verification of independence among observations obtained for the outcome or response variable (Y_i). This is achieved by means of a test entitled autocorrelation function (ACF), which checks correlations of a series with lagged values of itself. Autocorrelations are calculated for lags of 1, 2, ..., up to a specified number. The ACF can be carried out in a

partial way, when it correlates the values of a series with the values lagged by l or more cases, after the effects of correlations at the intervening lags have been removed.

According to Wei (1990), for a stationary process $\{Z_t\}$, the mean is $E(Z_t) = \mu$ and variance $Var(Z_t) = E(Z_t - \mu)^2 = \sigma^2$, which are constant, and the covariances $Cov(Z_t, Z_s)$, which are functions only of the time difference $|t - s|$. Hence, in this case, the covariance between Z_t and Z_{t+k} is written as:

$$\gamma_k = Cov(Z_t, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu) \quad (5.74)$$

and the correlation between Z_t and Z_{t+k} as:

$$\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)} \sqrt{Var(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0} \quad (5.75)$$

where it is noticeable that $Var(Z_t) = Var(Z_{t+k})$. As functions of k , γ_k is called the autocovariance function and ρ_k is called the autocorrelation function (ACF) in time series analysis since they represent the covariance and correlation between Z_t and Z_{t+k} from the same process, separated only by k time lags.

It is recognizable that for a stationary process the autocovariance function γ_k and the autocorrelation function ρ_k have the following properties:

- a. $\gamma_0 = Var(Z_t)$; $\rho_0 = 1$;
- b. $|\gamma_k| \leq \gamma_0$; $|\rho_k| \leq 1$;
- c. $\gamma_k = \gamma_{-k}$ and $\rho_k = \rho_{-k}$, for all k , i.e., γ_k and ρ_k are even functions, and hence symmetric about the time origin, $k = 0$. This follows from the fact that the time difference between Z_t and Z_{t+k} and Z_t and Z_{t-k} are the same.

The correlation function, therefore, is often plotted only for the nonnegative lags (FIGURE 5.9). This plot is sometimes called a correlogram.

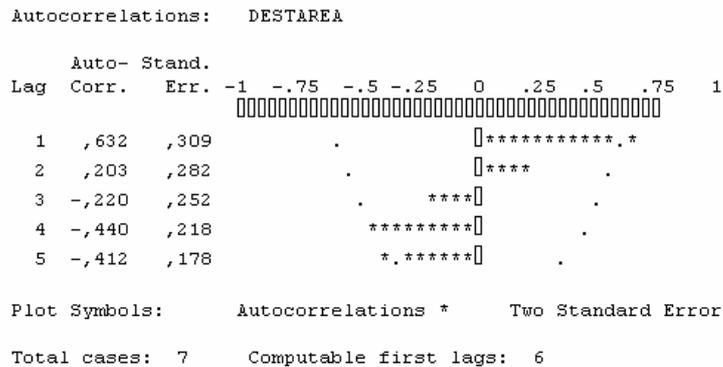


FIGURE 5.9 – Illustrative plot of the autocorrelation function for the response variable: destination area – industrial use (destarea) in Bauru for the years 1970, 1975, 1980, 1985, 1990, 1995 and 2000.

Other important steps in the exploratory analysis are boxplot analyses, correlation matrix analyses and the verification of interaction and confounding. Correlation, interaction and confounding as well as approaches to deal with these aspects have been already explained in Section 5.2.2.2.

5.3.2.3 Least Squares Estimators

In matrix terms, the general linear regression model (Equation 5.72) is:

$$Y_{n \times 1} = X_{n \times p} \beta_{p \times 1} + \varepsilon_{n \times 1}, \tag{5.76}$$

where Y is a vector of observations, β is a vector of parameters, X is a matrix of constants, ε is a vector of independent normal random variables with expectation $E(\varepsilon) = 0$ and variance-covariance matrix $\sigma^2(\varepsilon) = \sigma^2 I$.

Consequently, the random vector Y has expectation:

$$E(Y) = X\beta \quad (5.77)$$

and the variance-covariance matrix of Y is:

$$\sigma^2(Y) = \sigma^2 I \quad (5.78)$$

Let us denote the vector of estimated regression coefficients b_0, b_1, \dots, b_{p-1} as b :

$$b_{p \times 1} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_{p-1} \end{bmatrix} \quad (5.79)$$

The least squares normal equations for the general linear regression model are:

$$(X'X)_{p \times p} b_{p \times 1} = X'Y_{p \times n \quad n \times 1} \quad (5.80)$$

and the least estimators are:

$$b_{p \times 1} = (X'X)^{-1}_{p \times p} X'Y_{p \times 1} \quad (5.81)$$

These least squares estimators are also maximum likelihood estimators and have the properties of being minimum variance unbiased estimators, consistent and sufficient (Neter and Wasserman, 1974).

5.3.2.4 Analysis of Variance

Let the vector of the fitted values \hat{Y}_i be denoted \hat{Y} and the vector of the residual terms be denoted $e_i = Y_i - \hat{Y}_i$ be denoted e :

$$\hat{Y}_{n \times 1} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \cdot \\ \hat{Y}_n \end{bmatrix}, \quad (5.82)$$

$$e_{n \times 1} = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ e_n \end{bmatrix}. \quad (5.83)$$

The fitted values are represented by:

$$\hat{Y} = Xb \quad (5.84)$$

and the residual terms by:

$$e = Y - \hat{Y}. \quad (5.85)$$

The sum of squares for the analysis of variance is the following:

$$SSTO = Y'Y - n\bar{Y}^2, \quad (5.86)$$

$$SSR = b'X'Y - n\bar{Y}^2, \quad (5.87)$$

$$SSE = e'e = Y'Y - b'X'Y. \quad (5.88)$$

SSTO stands for *total sum of squares* and has $n - 1$ degrees of freedom associated with it. *SSE* denotes *error sum of squares* and has $n - p$ degrees of freedom associated with it since p parameters need to be estimated in the regression model (Equation 5.76). Finally, *SSR*, the *regression sum of squares*, has $p - 1$ degrees of freedom associated with it, representing the number of X variables X_1, \dots, X_{p-1} .

The means squares MSR (*regression mean square*) and MSE (*error mean square*) are given by:

$$MSR = \frac{SSR}{p - 1} , \quad (5.89)$$

$$MSE = \frac{SSE}{n - p} . \quad (5.90)$$

The expectation of MSE is σ^2 , as for simple regression. The expectation of MSR is σ^2 plus a quantity which is positive if any of the β_k ($k = 1, \dots, p - 1$) coefficients is not zero.

To test whether there is a relation between the dependent variable Y and the set of variables X_1, \dots, X_{p-1} , that is, to choose between the alternatives:

$$H_0: \beta_1 = \beta_2 \dots = \beta_{p-1} = 0$$

$$H_1: \text{not all } \beta_k \text{ (} k = 1, \dots, p - 1 \text{) equal } 0,$$

the following statistics is used:

$$F^* = \frac{MSR}{MSE} . \quad (5.91)$$

$$\text{If } F^* \leq F(1 - \alpha; p - 1; n - p), \text{ accept } H_0;$$

$$\text{If } F^* > F(1 - \alpha; p - 1; n - p), \text{ accept } H_1.$$

An important measure used to assess the fit of a linear regression model is the coefficient of multiple determination, denoted by R^2 , which is defined as follows:

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} . \quad (5.92)$$

R^2 lies within the interval [0,1] and assumes the value 0 when all $b_k = 0$ ($k = 1, \dots, p - 1$). R^2 takes on the value 1 when all observations fall directly on the fitted response surface, that is, when $Y_i = \hat{Y}_i$ for all i .

The existence of a regression relation, by itself, does not obviously assure that useful predictions can be made by using it. A prediction interval with $1 - \alpha$ confidence coefficient for a new observation $Y_{h(new)}$ corresponding to X_h (the specified values of the X variables) is:

$$\hat{Y}_h \pm t (1 - \alpha/2; n - p) S (Y_{h(new)}) \quad . \quad (5.93)$$

The prediction error variance (difference between the new observation and the estimated value) is:

$$S^2 (Y_{h(new)}) = MSE (1 + X'_h (X'X)^{-1} X_h) \quad . \quad (5.94)$$

5.3.2.5 Analysis of Residuals

The aptness of a linear regression model can be also tested by an analysis of residuals (Neter and Wasserman, 1974). For analytical convenience, residuals are also used in a standardized form (zre_i) according to the following formula:

$$zre_i = \frac{e_i}{\sqrt{MSE}} \quad . \quad (5.95)$$

A plot of the residuals e_i versus the fitted values \hat{Y}_i should provide scattered points presenting no correlation pattern at all. When an absence of correlation is identified in the plot, this indicates that (a) the linear regression model is well fit, (b) the assumption of variance homogeneity is met; (c) and the model contains no outliers. Instead of using e_i , the standardized residuals (zre_i) can also be used.

A plot of a standard normal distribution against the residuals e_i is appropriate for checking the assumption of errors normality. When this assumption holds, this plot should present a linear correlation pattern.

A plot of the residuals e_i against an independent variable is useful to determine both whether a linear regression function is appropriate and problems of order exist, and also to examine whether the variance of the error terms is constant. An independent variable can also be plotted against the standardized residuals (zre_i). In both cases, the plots should present no correlation pattern (FIGURE 5.10).

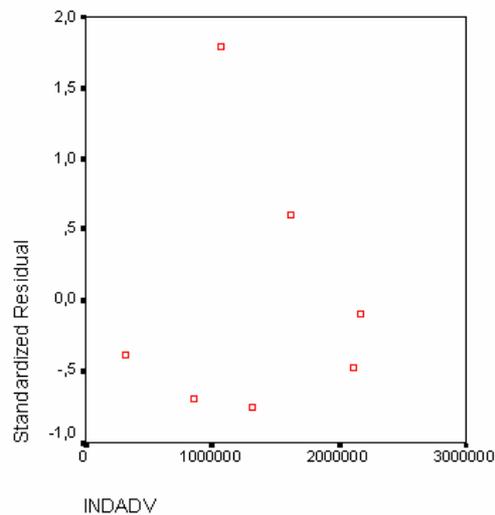


FIGURE 5.10 – Illustrative plot of standardized residuals versus an independent variable (industrial added value - *indadv*) of a linear regression model for the industrial area in the city of Piracicaba (1970–2000).

5.3.2.6 Conversion of Predicted Destination Areas into Transition Probabilities

Linear regression models relating the area of certain land uses with demographic data and economic indicators were built in order to allow predictions for future scenarios of land use. As population data and economic performance proved to be highly correlated, all the linear regression models designed to parameterize forecasts of land use change

ended up as univariate models, i.e. either population or an economic indicator (total or sectorial GDP) was elected as the independent variable in each one of these models.

In all regression models where the correlation of economic data surpassed the correlation of demographic data in relation to the response variable, the “total GDP” presented the highest correlation with this variable. Nevertheless, not in all of the cases the total GDP was chosen as the independent variable. When the response variable was the industrial use, the industrial GDP was elected as the explicative or independent variable, since this sectorial GDP presented a high correlation with such response variable.

In order to convert the land use area (outcome variable of linear regression models) into land use transition probabilities meant to feed up the simulation model, a set of formulas was used according to the respective case.

There are basically three main cases relating origin and destination uses in land use transitions. The first of them is when the origin use of a certain transition is also the origin use of other transitions. For instance, the transitions “non-urban to industrial use” (nu_ind) and “non-urban to residential use” (nu_res) have the non-urban use as their origin use. In this case, the following formulas are employed for the calculation of transition probabilities:

$$P_{nu_ind} = \frac{ind_f - ind_i}{non\ urb_i} \quad (5.96)$$

and

$$P_{nu_res} = \frac{res_f - res_i}{non\ urb_i} \quad (5.97)$$

where P_{nu_ind} is the probability for the transition “non-urban to industrial use”; ind_i and ind_f are respectively the industrial use areas in the initial and final time of simulation; $non\ urb_i$ is the non-urban use area in the initial time of simulation; P_{nu_res} is the

probability for the transition “non-urban to residential use”, and res_i and res_f are respectively the residential use areas in the initial and final time of simulation.

The second case regards the situations where the destination use in a certain transition is also the destination use in other transitions. For example, the transitions “non-urban to services use” (nu_serv) and “residential to services use” (res_serv) have the services use as their common destination use. In this case, the formulas below are used for the calculation of transition probabilities:

$$P_{nu_serv} = \frac{(serv_f - serv_i) - res_serv}{non_urb_i} \quad (5.98)$$

$$P_{res_serv} = \frac{(serv_f - serv_i) - nu_serv}{non_urb_i} \quad , \quad (5.99)$$

where P_{nu_serv} is the probability for the transition “non-urban to services use”; $serv_i$ and $serv_f$ are respectively the services use areas in the initial and final time of simulation; non_urb_i is the non-urban use area in the initial time of simulation; res_serv is the residential use area converted into services use, and nu_serv is the non-urban use area converted into services use. res_serv and nu_serv are both obtained through a *simple system of equations* built upon basis of *sets theory*.

Finally, the third case concerns the situations in which the destination use in a certain transition is the origin use in other transitions, and eventually the origin use of the considered transition is also the origin use of the remaining transitions existent in the simulation period. For instance, the set of transitions “non-urban to residential use” (nu_res), “residential to services use” (res_serv), “residential to mixed use” (res_mix) and “non-urban to services use” (nu_serv) exemplify this third case. The following formula should be employed in this situation for the probability estimation of the transition “non-urban to residential use” (nu_res):

$$P_{nu_res} = \frac{(res_f - res_i) - [(serv_f - serv_i) - nu_serv] - (mix_f - mix_i)}{non_urb_i}, \quad (5.100)$$

where P_{nu_res} is the probability for the transition “non-urban to residential use”; $serv_i$ and $serv_f$ are respectively the services use areas in the initial and final time of simulation; mix_i and mix_f are respectively the mixed use areas in the initial and final time of simulation; non_urb_i is the non-urban use area in the initial time of simulation, and nu_serv corresponds to the non-urban use area converted into services use, obtained by means of a *triple system of equations* built upon basis of *sets theory*. This system of equations besides nu_serv involves the unknown variables nu_res (non-urban use area converted into residential use) and res_serv (residential use area converted into services use). The remaining transition probabilities in this third case (P_{nu_serv} , P_{res_serv} , P_{res_mix}) are then calculated upon basis of the estimated values of nu_serv , nu_res and res_serv .

5.3.3 Conceived Scenarios and Time Horizons for Land Use Change Forecasts

As previously stated in Section 5.3.1, the Markov chain has been used to generate stationary forecasts of land use change. Optimistic and pessimist future scenarios of land use change were also conceived, with respectively slight over and underestimations of the independent variable. Both of them were based likewise the Markov chain on transitions exclusively observed in the last simulation period of the time series, i.e. on periods ranging from 1988 to 2000, for Bauru, and from 1985 to 1999, for Piracicaba.

These non-stationary scenarios were formulated through linear regression models, relating areas of certain land uses with population and economic data (see Section 5.3.2), where slight over- and under-estimations in the values of the independent variable respectively produced optimistic and pessimist scenarios of land use change.

The adoption of slight variations in the projections of population and economic data can be justified by the fact that the demographic and macroeconomic scenarios of Brazil in the latest years are expected to reproduce themselves in the current decade, in view of steadfastly decreasing population growth rates as well as of the administrative continuity trend demonstrated by the current federal government.

As to the delimitation of time horizons, there are no official definitions regarding short- and medium-terms for urban land use change phenomena. In a general way, it is sensible to define a short-term as being a time length of up to five years, and a medium-term being comprised within more than five and less than ten years. In the current research, the **short-term** horizon was set to the year **2004**, and the **medium-term** horizon, to **2007**.

Specifically regarding urban land use change modeling, it is unsuitable to deal with long-term forecasts, due to two main reasons. First, long-term land use changes are hardly foreseeable, and hence error-prone, in face of sudden alterations in the macroeconomic sphere and consequently in the land use change behavior that may eventually take place in the course of years in the later medium-term or early long-term. Second, within the scope of strategic town planning, only the short- and medium-terms are relevant for the priorities definition, resources allocation and decision-making processes.

5.4 The Urban Land Use Dynamics Simulation Model

5.4.1 DINAMICA General Data Model

The urban land use dynamics simulation model employed to carry out the simulation experiments in this research was the DINAMICA, developed by the Center of Remote Sensing of the Federal University of Minas Gerais (CSR – UFMG). The DINAMICA software was written in object-oriented C++ language and its present version runs on 32 bit Windows © system (Soares-Filho et al. 2002).

The DINAMICA is a cellular automaton model, implemented through empirical land use allocation algorithms. A generic data model for DINAMICA is presented in FIGURE 5.11. To operate it, an initial land use map and two data sets corresponding to the static and dynamic input variables are necessary. All these maps together with parameters obtained from either the weights of evidence or the logistic regression method will be used for the calculation of cells transition probabilities.

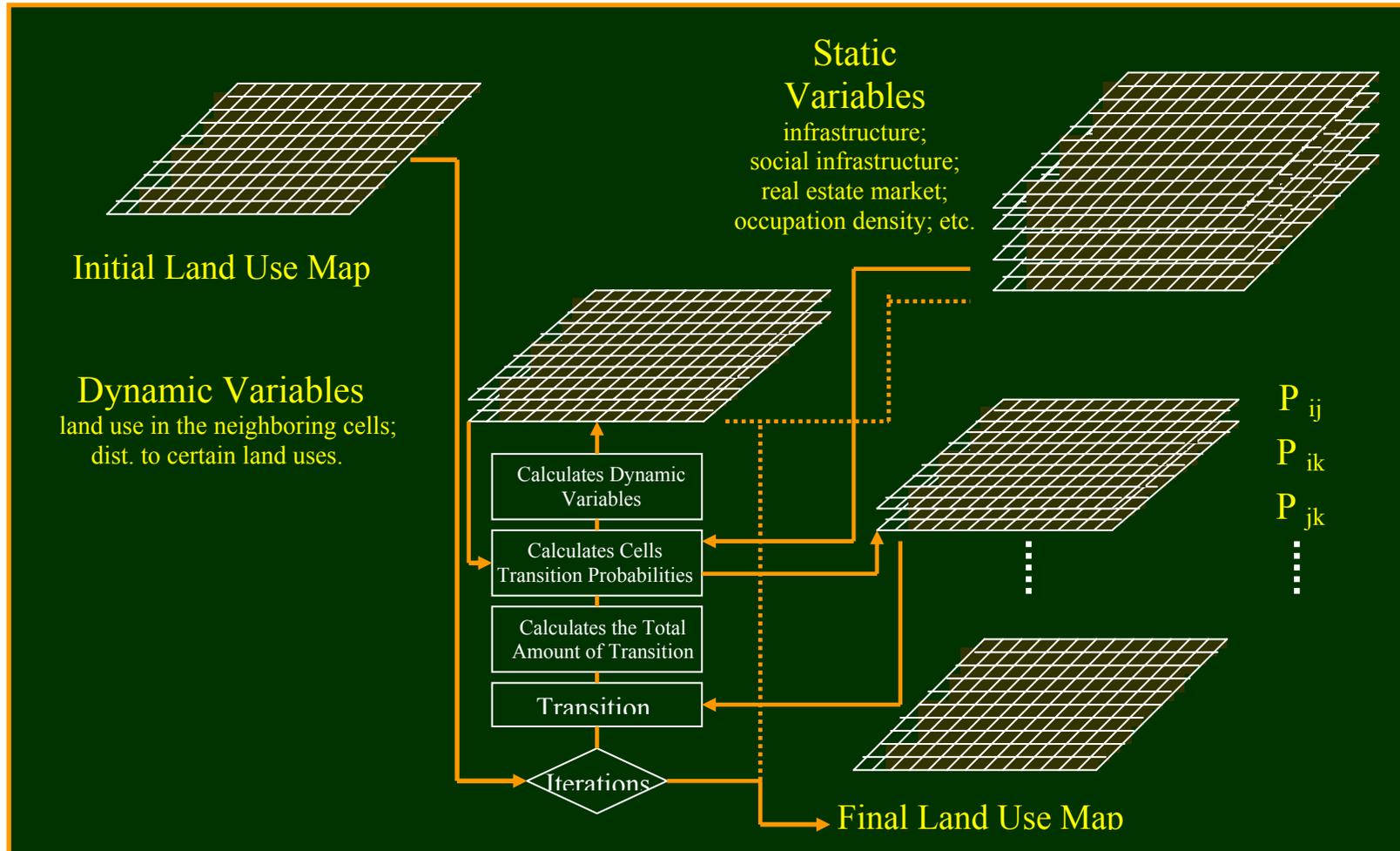


FIGURE 5.11 – DINAMICA generic data model for urban applications.

SOURCE: Adapted from SOARES-FILHO (1998, p.257).

The static variables refer to maps of technical and social infrastructure availability, types of urban occupation density, relief, urban zoning, etc. The dynamic variables refer to maps of distance to certain types of land use. These variables undergo changes in each program iteration, which are continuously updated so as to influence the calculation of transition probabilities in the next iteration. Thus, in each program iteration, changes take place in the state (land use) of given cells, changing thence the distance to these uses estimated from other cells.

From the estimated land use transition probabilities (P_{ij} , P_{ik} , etc.), the land use transition maps are elaborated, which will be used for heuristic procedures in the model calibration (see Section 5.2.1.5). Finally, with the cells transition probabilities calculated and the total amount of land use change estimated through cross-tabulation, the Markov chain or linear regression models, changes in the cells land use will occur in successive iterations, so as to produce a final urban land use map.

5.4.2 Software Structure and Input Parameters

DINAMICA is a CA model based on an eight cell Moore neighborhood approach implemented through empirical land use allocation algorithms. The DINAMICA parameters include:

- a) a file containing the values of the positive weights of evidence (W^+) or the parameters (β_0 , β_i) of the logistic regression models, depending on the estimation method adopted;
- b) the total transition probabilities for each of the land use transitions identified in the considered simulation period;
- c) the size and variance of patches for each transition;
- d) the proportion of the transition or allocation algorithms (*'expander function'* and *'patcher function'*) also considering each transition; and,
- e) the total number of iterations.

The latest version of DINAMICA allows the parameters setting by graphical interfaces, although for the current research, the DOS version of DINAMICA has been used.

5.4.3 Transition Algorithms

DINAMICA presents two land use transition (or land use allocation) algorithms: the *expander* and the *patcher* functions. The *expander* function accounts for the expansion of previous patches of a certain land use class. The *patcher* function, on its turn, is designed to generate new patches through a seedling mechanism (Soares-Filho et al. 2002). In summary, the *expander* function realizes transitions from a state i to a state j only in the adjacent vicinities of cells with state j . And the *patcher* function accomplishes transitions from a state i to a state j only in the adjacent vicinities of cells with state other than j .

These two processes can be merged into the following equation:

$$Q_{ij} = r \times (\text{expander function}) + s \times (\text{patcher function}) \quad , \quad (5.101)$$

where Q_{ij} corresponds to the total amount of transitions of type ij specified per simulation period, and r and s are respectively the percentage of transitions executed by each function, with $r + s = 1$.

According to Soares-Filho et al. (2002), both transition algorithms adopt a stochastic selecting mechanism. The applied algorithm consists in scanning the initial land use map to sort out the cells with the highest probabilities and then arrange them in a data array. Following this procedure, cells are selected randomly from top to bottom of the data array (the internal stochastic choosing mechanism can be loosened or tightened depending on the degree of randomization desired). In a final step, the land use map is again scanned to accomplish the selected transitions.

In the case that the *expander* function does not execute the amount of desired transitions after a fixed number of iterations, it passes on to the *patcher* function a residual number

of transitions, so that the total number of transitions always reaches an expected value (Soares-Filho et al. 2002).

5.4.3.1 The *Expander* Function

The *expander* algorithm is expressed by the following equation:

$$\begin{aligned} \text{If } n_j > 3 \text{ then } P'_{ij}(x,y) = P_{ij}(x,y) \text{ else} \\ P'_{ij}(x,y) = P_{ij}(x,y) \times (n_j)/4 \end{aligned} \quad , \quad (5.102)$$

where n_j corresponds to the number of cells of type j occurring in a window 3×3 . This method guarantees that the maximum P'_{ij} will be the original P_{ij} , whenever a cell type i is surrounded by at least 50% of type j neighboring cells.

5.4.3.2 The *Patcher* Function

The *patcher* function strives to simulate patterns of land use change by generating diffused patches and preventing at the same time the formation of single isolated one-cell patches. This function employs a device which searches for cells around a chosen location for a considered transition. This is achieved firstly by selecting the core cell of the new patch and then selecting a specific number of cells around the core cell, according to their P_{ij} transition probabilities.

The *expander* and *patcher* functions, as previously mentioned, incorporate an allocation device which is responsible for identifying cells with the highest transition probabilities for each ij transition. This allocation device stores the cells and organizes them for subsequent selection. In this process, each newly selected cell will form a core for a new patch or an expansion fringe, which still need to be further developed by using one of these two transition algorithms. The sizes of the new patches and the expansion fringes are set according to a lognormal probability distribution, whose parameters are determined as a function of the mean size and variance of each type of patch and expansion fringe to be generated (Soares-Filho et al. 2002).

5.5 Methodological Summary Flowchart

A flowchart that summarizes the methodological procedures for modeling urban land use dynamics is presented in FIGURE 5.12. In the initial block, which concerns the data input, remote sensing, urban cartographic and census data are gathered and duly processed to integrate a digital geographic database. In the next stage, an exploratory analysis of the input data is carried out, aiming at the variables selection. The aim of this stage is to extract the minimum and at the same time the best set of variables to explain the phenomena under study: the land use transitions.

After defining the final variables set for each type of transition, the modeling stage itself is approached. Firstly, transition rates are calculated through cross-tabulation or via the Markov chain. Next, cells transition probabilities are obtained either by logistic regression or the weights of evidence method.

With the transition rates and cells transition probabilities estimated, the simulations can be carried out at last. Simulations outputs are continuously calibrated until the obtainment of satisfactory results, which will then be validated afterwards.

Therefore, by getting acquainted with trends of land use change throughout a sufficiently long time series, the modeler is finally able to conceive scenarios for future transitions. In this case, transition rates can be stationary, and hence estimated by the Markov chain; or non-stationary, which can be obtained from linear regression models relating areas of certain land uses with demographic and economic data. In all cases, forecasts of urban land use change are generated for time horizons in the short- and medium-terms.

5.6 Conclusions

Chapter 5 presented the statistical techniques designed to conduct the simulations and forecasts modeling experiments of land use change in the cities of Bauru and Piracicaba. The weights of evidence method, based on the Bayes' theorem of conditional probability, has been extensively used in all simulation periods, while the logistic

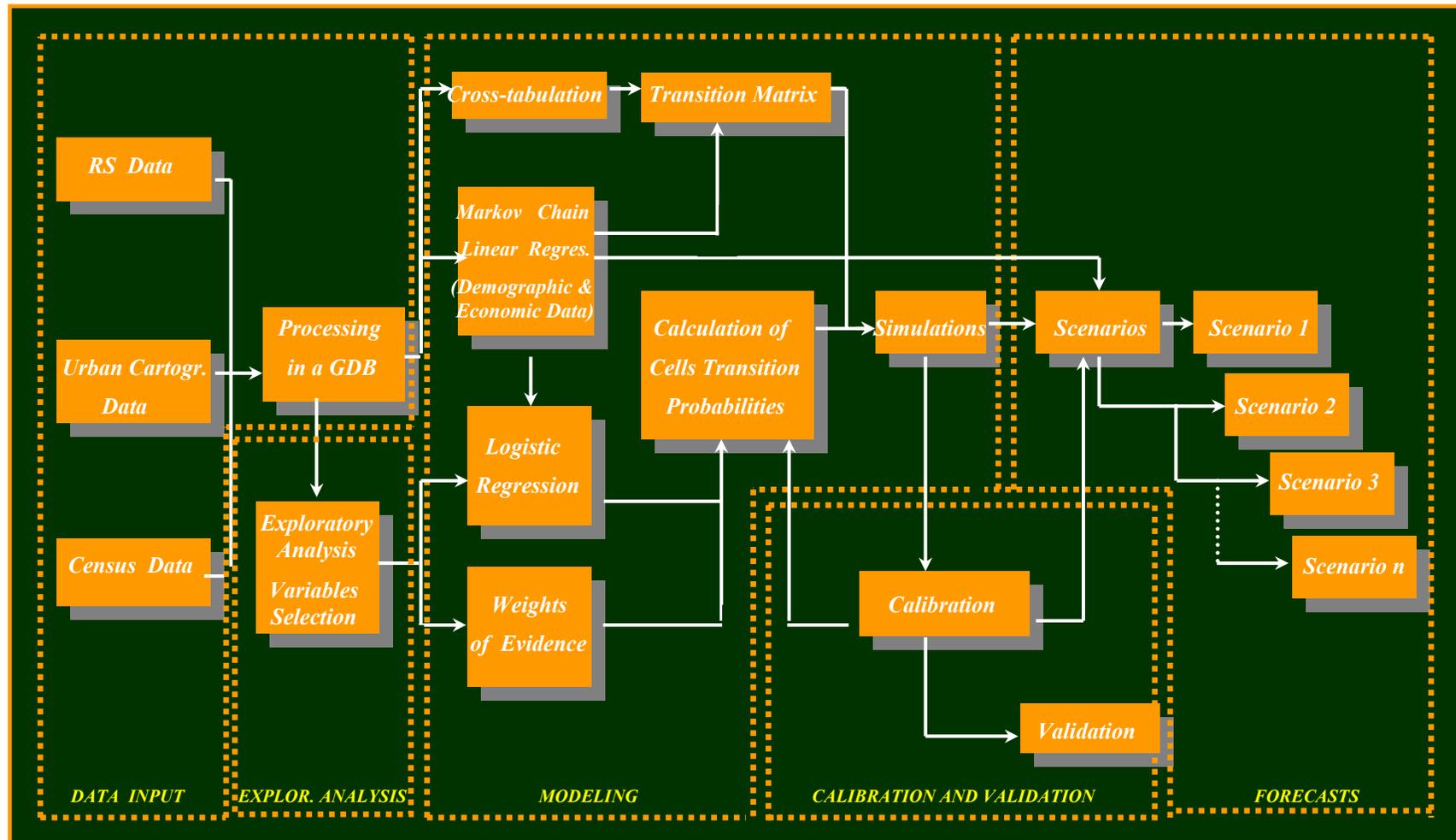


FIGURE 5.12 – Summary methodological flowchart.

regression method has been experimentally applied to the city of Bauru and only for the simulation period ranging from 1979 to 1988.

Both methods are based on *logits*, i.e. the natural logarithm of *odds*, defined as a ratio of the probability that an event will occur to the probability that it will not occur. The weights of evidence method calculates the parameters for each evidence (explicative variable) or class of evidence in a univariate form, i.e. each variable has its positive weight of evidence (W^+) and negative weight of evidence (W^-) individually determined. In the logistic regression method, on the other hand, the assessment of the regression parameters (β_0, β_i) is done simultaneously for all variables integrating the model.

According to Hosmer and Lemeshow (1989), “one problem with any univariate approach is that it ignores the possibility that a collection of variables, each of which is weakly associated with the outcome, can become an important predictor of outcome when taken together.”

Although the determination of parameters in the weights of evidence method is done individually for each variable, the probability formula aggregates all the information related to each of these variables.

The weights of evidence can be regarded as a much more intuitive method, for the modeler monitors the whole process of parameters assessment. On the contrary, parameters estimation in the logistic regression model, executed through continuously iterative methods, remains a black box for modelers. These two methods present very similar results. Nevertheless, the weights of evidence should be given preference of application in view of its transparency and operational simplicity, what together concur for a faster and more consistent model calibration.

One methodological aspect that should be highlighted here is the fact that in the simulation experiments that provide yearly outputs, the input variables should preferably be yearly updated as well, so as to allow for more sensible results. Unfortunately, most of the input variables used in urban land use change modeling are

very complex and detailed to be yearly updated from remote sensing or conventional cartographic data.

And finally, it is worth remarking that the non-stationary forecast simulation methods, which are based on linear regression models, ought to employ time series analysis for the independent variables estimation. As time series analysis requires a minimum of thirty observations (Wei, 1990), and as the yearly total and sectorial municipal GDPs of Bauru and Piracicaba in the latest three decades were not available by the time this research was finished, this statistical technique has not been adopted. Anyway, the usage of time series analysis in the forecast simulations will remain as a direction for future work.

CHAPTER 6

RESULTS AND DISCUSSIONS

6.1 Bauru

As previously stated in Section 3.4.1.2, Bauru was born as a crossing point between railways during the inward advance of the coffee culture in the XIX century. Still today, the city is mainly shaped by the transport system: its urban framework is organized around four interregional roads and the railway track, which still crosses the city core area.

Bauru is currently regarded as a dynamic regional development pole in the central-west portion of São Paulo State, with an outstanding performance of tertiary activities (commerce and services). In view of its strategic historic development conditions, the city underwent a drastically fast urbanization process. These urbanization booms were followed by speculative processes, leading to the formation of a discontinuous urban tissue, marked by the scattered presence of empty areas, low occupation densities in the outer areas and the existence of detached residential settlements orbiting around the city center, as it will be seen in the next sections.

In the sequence, the results for the land use change simulations throughout the time series ranging from 1967 to 2000 will be presented, followed by the forecasts simulations generated for the short- and medium-term, respectively 2004 and 2007.

6.1.1 Simulation Period: 1967 - 1979

There are no official data about the population of Bauru in 1967. The first demographic census in Brazil was held in 1970 (IBGE, 1971), and the total population of this municipality at that time was 64,859 inhabitants, out of which 61,592 inhabitants lived in urban areas. It can be approximately inferred that the population in Bauru in 1967

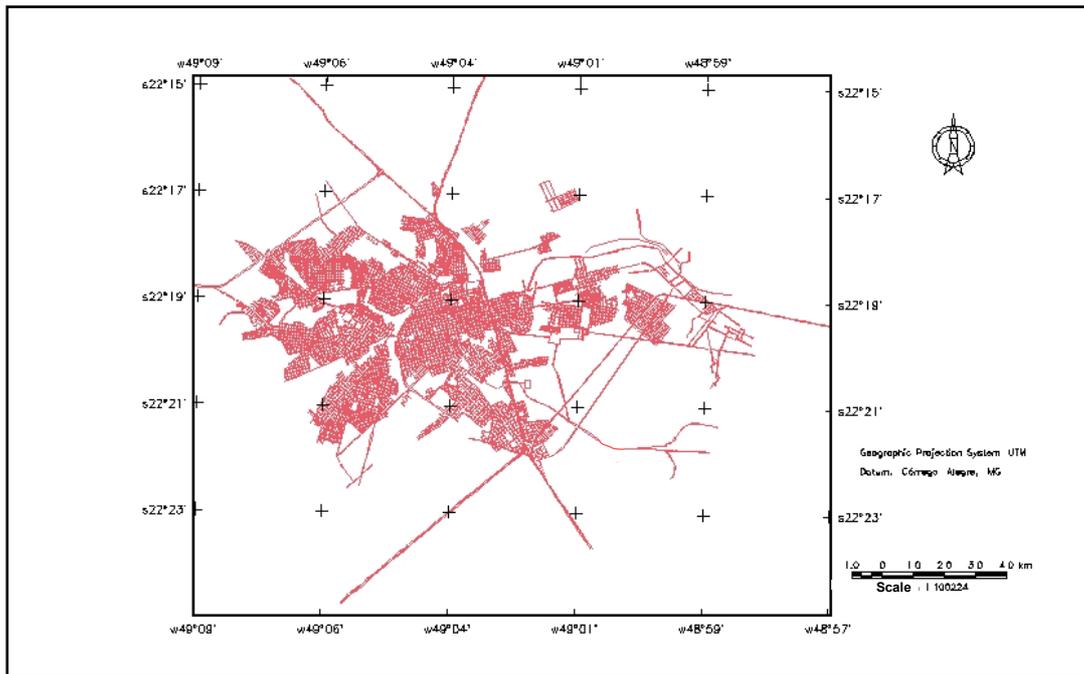


FIGURE 6.1 – Bauru official city map in 1967.

SOURCE: CEPEU-FAUUSP (1967a).

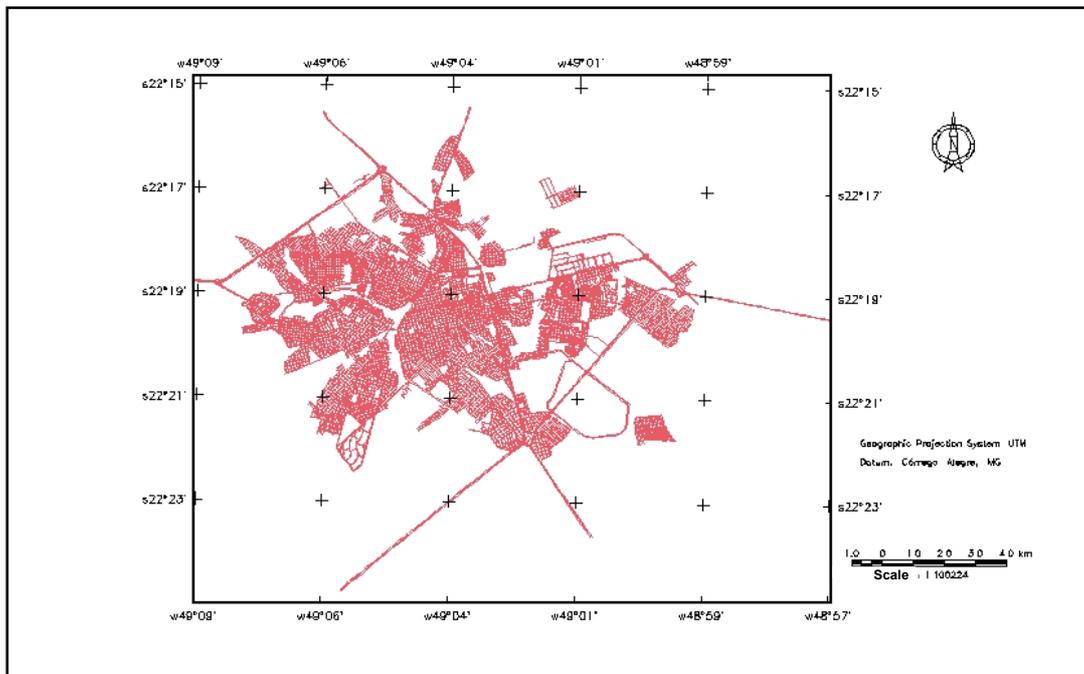


FIGURE 6.2 – Bauru official city map in 1979.

SOURCE: SEPLAN (1979a).

was about 57,000 inhabitants. In the national census of 1980 (IBGE, 1982), the total population of Bauru rose to 164,105 inhabitants, from which 159,926 people were urban inhabitants. The urban population growth rate within this period (1967-1979) was considerably high, around 2.81%. The shifts in the urban area can be seen in FIGURES 6.1 and 6.2, which present the city maps for the initial and final time of simulation.

The initial and final land use maps used in the simulation period 1967 – 1979 (FIGURE 6.3) were elaborated upon basis of the two official city maps previously shown, of generalization procedures applied to the original land use maps of 1967 (CEPEU-FAUUSP, 1967b) and 1979 (SEPLAN, 1979a) and of the printed satellite image of Bauru in 1979 (INPE, 1979).

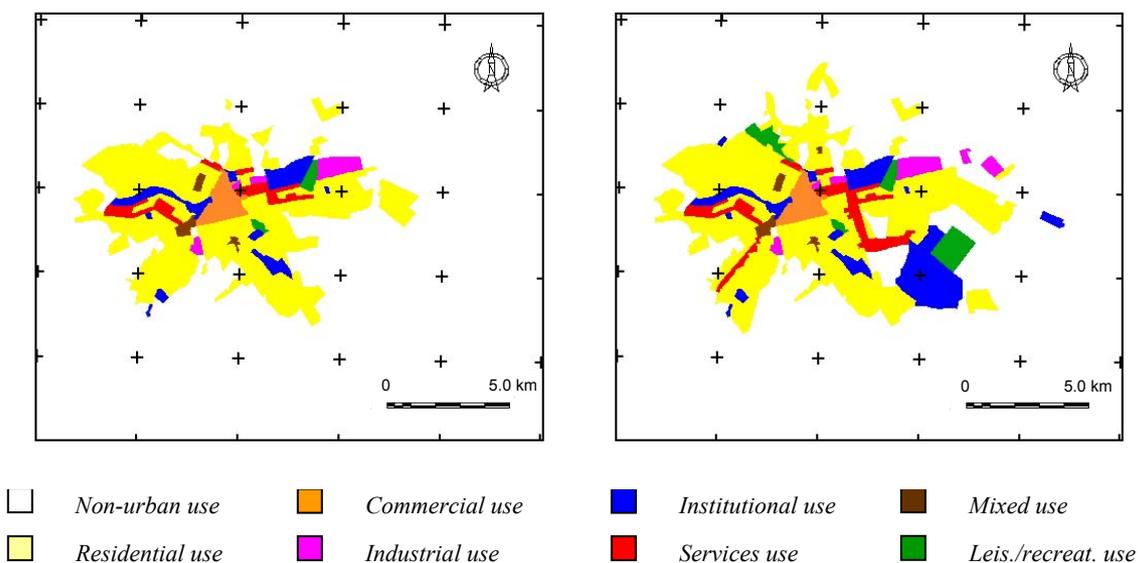


FIGURE 6.3 – Generalized land use map in Bauru in 1967 (left) and 1979 (right).

In order to calculate land use transition rates for the period 1967-1979, the initial and final land use maps, processed in SPRING, were converted to raster files with extension TIFF and resolution 100 x 100 (m), and then exported to the IDRISI Geographic Information System. A cross-tabulation operation was made between both land use maps (FIGURE 6.4) so as to generate transition percentages for the existent types of land use change (TABLES 6.1 and 6.2).

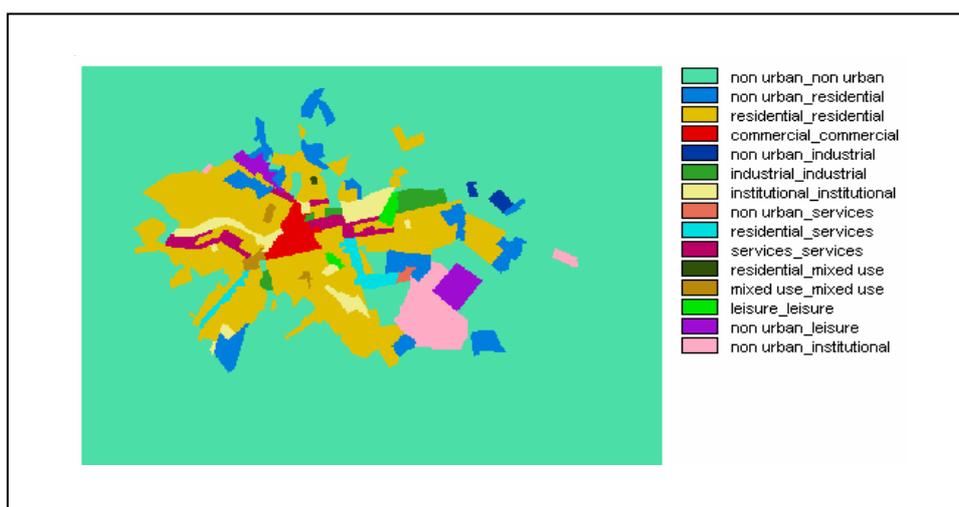


FIGURE 6.4 – Cross-tabulation map between Bauru land use maps of 1967 and 1979.

TABLE 6.1 – Existent land use transitions in Bauru: 1967–1979.

<i>NOTATION</i>	<i>LAND USE TRANSITION</i>
<i>NU_RES</i>	<i>Non-urban to residential</i>
<i>NU_IND</i>	<i>Non-urban to industrial</i>
<i>NU_INST</i>	<i>Non-urban to institutional</i>
<i>NU_SERV</i>	<i>Non-urban to services</i>
<i>NU_LEIS</i>	<i>Non-urban to leisure/recreation</i>
<i>RES_SERV</i>	<i>Residential to services</i>
<i>RES_MIX</i>	<i>Residential to mixed use</i>

TABLE 6.2 – Matrix of global transition probabilities for Bauru: 1967–1979.

<i>Land Use</i>	<i>Non-urban</i>	<i>Resid.</i>	<i>Comm.</i>	<i>Industr.</i>	<i>Instit.</i>	<i>Services</i>	<i>Mixed</i>	<i>Leis./Recr.</i>
<i>Non-urban</i>	0.9361	0.0315	0	0.0022	0.0199	0.0009	0	0.0094
<i>Resid.</i>	0	0.9498	0	0	0	0.0485	0.0016	0
<i>Comm.</i>	0	0	1	0	0	0	0	0
<i>Industr.</i>	0	0	0	1	0	0	0	0
<i>Instit.</i>	0	0	0	0	1	0	0	0
<i>Services</i>	0	0	0	0	0	1	0	0
<i>Mixed</i>	0	0	0	0	0	0	1	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1

After estimating the global transition probabilities, i.e. the total amount of land use change for Bauru in the period 1967 – 1979, it is necessary to determine the local transition probabilities (cells land use change probabilities). For this end, the set of variables designed to explain each of the land use transitions must be defined. Heuristic procedures and statistical tests are employed in this selection process as stated in Section 5.2.1.2. For the simulation period 1967–1979, twelve variables were selected (CEPEU-FAUUSP, 1967a, 1967b, 1967c), some of which are shown in FIGURE 6.5.

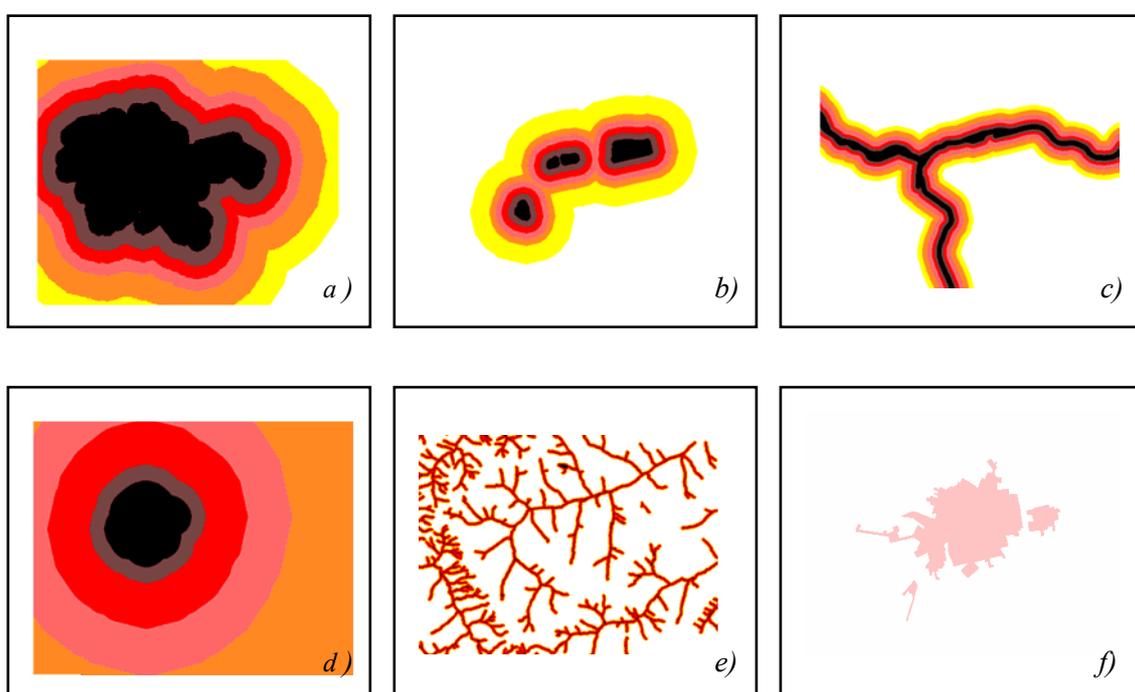


FIGURE 6.5 – Independent variables used to explain the land use transitions in Bauru during the simulation period 1967 – 1979: a) distances to residential zones; b) distances to industrial zones; c) distances to railways; d) distances to ranges of commercial activities clusters; e) distances to rivers and water bodies; f) water supply.

TABLE 6.3 shows the notations utilized for each map of variable employed in this simulation experiment; TABLE 6.4 indicates which variable was selected to explain each of the seven existent transitions; and finally, TABLE 6.5 presents the values

obtained for the Cramer's Coefficient (V) and the Joint Information Uncertainty (U) for the pairs of variables used to explain the same type of land use transition.

TABLE 6.3 – Independent variables defining land use change in Bauru: 1967–1979.

<i>NOTATION</i>	<i>PHYSICAL OR SOCIOECONOMIC LAND USE CHANGE VARIABLE</i>
<i>water</i>	<i>Area served by water supply.</i>
<i>dist_riv</i>	<i>Distances to river and water bodies.</i>
<i>dist_com</i>	<i>Distances to commercial zones in general.</i>
<i>com_kern</i>	<i>Distances to different ranges of commercial activities concentration, defined by the Kernel estimator.</i>
<i>dist_ind</i>	<i>Distances to industrial zones.</i>
<i>dist_res</i>	<i>Distances to residential zones.</i>
<i>dist_rail</i>	<i>Distances to railways.</i>
<i>trv_rds</i>	<i>Distances to transversal peripheral roads (sw-ne; se-nw).</i>
<i>exist_rds</i>	<i>Distances to main existent roads.</i>
<i>serv_axes</i>	<i>Distances to the services and industrial axes.</i>
<i>asph_rds</i>	<i>Distances to asphalted roads.</i>
<i>per_rds</i>	<i>Distances to peripheral roads, which pass through non-occupied areas.</i>

TABLE 6.4 – Selection of variables determining land use change in Bauru: 1967–1979.

<i>NOTATION</i>	<i>NU_RES</i>	<i>NU_IND</i>	<i>NU_INST</i>	<i>NU_SERV</i>	<i>NU_LEIS</i>	<i>RES_SERV</i>	<i>RES_MIX</i>
<i>water</i>							♦
<i>dist_riv</i>					♦		
<i>dist_com</i>			♦	♦			♦
<i>com_kern</i>	♦						
<i>dist_ind</i>		♦					
<i>dist_res</i>	♦		♦	♦			
<i>dist_rail</i>		♦					
<i>trv_rds</i>					♦		
<i>exist_rds</i>	♦			♦			
<i>serv_axes</i>		♦				♦	♦
<i>asph_rds</i>						♦	
<i>per_rds</i>			♦				

According to what was already stated in Section 5.2.1.2, Bonham-Carter (1994) reports that values less than 0.5 for Cramer's Coefficient and the Joint Information Uncertainty suggest less association rather than more. As none of the association measure values surpassed this threshold simultaneously for both indices, no variables preliminarily selected for modeling have been discarded from the analysis.

TABLE 6.5 – Associations between independent variables - Bauru: 1967–1979.

<i>VARIABLE A</i>	<i>VARIABLE B</i>	<i>CRAMER'S STATISTIC (V_{A,B})</i>	<i>UNCERTAINTY (U_{A,B})</i>
<i>dist_com</i>	<i>exist_rds</i>	0.3552	0.2674
	<i>dist_res</i>	0.3878	0.3189
	<i>per_rds</i>	0.1448	0.0580
	<i>serv_axes</i>	0.1650	0.0753
<i>com_kern</i>	<i>dist_res</i>	0.5334	0.3822
	<i>exist_rds</i>	0.4751	0.3039
<i>dist_res</i>	<i>per_rds</i>	0.1470	0.0461
	<i>exist_rds</i>	0.5774	0.4249
<i>serv_axes</i>	<i>water</i>	0.2446	0.0475
	<i>dist_ind</i>	0.1279	0.0497
	<i>asph_rds</i>	0.0548	0.0130
	<i>dist_rail</i>	0.1334	0.0426
<i>dist_ind</i>	<i>serv_axes</i>	0.1279	0.0497
<i>trv_rds</i>	<i>dist_riv</i>	0.0420	0.0025

With the final set of independent variables being defined, procedures to obtain the positive weight of evidence (W^+) are finally taken. In IDRISI, the land use cross-tabulation map of Bauru (1967-1979) was used to generate land use transition maps (FIGURE 6.6) for each of the seven types of land use change presented in TABLE 6.1. This was done through reclassification tables, on which the following rules were observed:

- all raster values corresponding to classes of land use permanence or transition whose initial land use was different from the initial land use category in the considered type of land use change were assigned value 0 (**black colour**);
- all raster values corresponding to classes of land use transition whose initial and final land use categories were equal to the initial and final categories of the land use change at issue were assigned value 2 (**blue colour**);
- all other remaining classes of land use permanence or transition were assigned value 1 (**green colour**).



FIGURE 6.6 – Example of a land use transition map (non-urban to leisure/recreation) for Bauru in the period: 1967-1979.

Once all possible types of land use transition maps were elaborated, they were then subjected to partial cross-tabulations with the respective independent variables maps, according to the relationships established on TABLE 6.4. The variables (evidences) maps, pre-processed in SPRING, were in the same manner as the initial and final land use maps converted to raster files with resolution 100 x 100 (m).

The partial cross-tabulations disregard the raster values 0 (black color) in the land use transition maps. The numerical values of cells proportions existing in the absence/presence of a binary evidence (e.g. water supply) or in the different ranges of distances maps and found to be overlying on either class 1 (green color) or 2 (blue color) of the land use transition maps are (for each cross-tabulation table) selectively transferred to files especially created for the calculation of the weights of evidence (Equations 5.9 and 5.10).

Using the values of the positive weights of evidence W^+ (TABLE 6.6) concerning the several evidences maps employed in the analysis of each category of land use change, the DINAMICA simulation model will then calculate the cells transition probabilities (Equation 5.33) for the seven types of land use transition.

TABLE 6.6 – Values of W^+ for the selected independent variables - Bauru: 1967–1979.

<i>Land Use Transition</i>	<i>Variable</i>	<i>Positive Weights of Evidence W^+</i>						
		<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>
<i>NU_RES</i>	<i>com_kern</i> ¹	2.876	2.126	0.846	-0.515	0	na	na
	<i>dist_res</i> ²	1.849	-0.102	-0.772	-0.407	-3.087	0	0
	<i>exist_rds</i> ³	1.383	0.739	-2.352	-1.887	-1.120	0	0
<i>NU_IND</i>	<i>dist_ind</i> ⁴	0	0	0	3.273	0.916	2.510	0
	<i>dist_rail</i> ⁵	2.247	2.425	2.450	1.315	0	0	0
	<i>serv_axes</i> ⁶	2.527	2.462	2.498	2.523	2.353	1.982	0
<i>NU_INST</i>	<i>dist_com</i> ⁷	0	0	0	0	-0.466	1.082	-2.235
	<i>dist_res</i> ²	0.366	1.283	0.623	-1.647	0	0	0
	<i>per_rds</i> ⁸	3.247	2.550	1.194	0	0	-1.818	0
<i>NU_SERV</i>	<i>dist_com</i> ⁷	0	0	0	0	3.259	0	0
	<i>dist_res</i> ²	2.067	0	0	0	0	0	0
	<i>exist_rds</i> ³	1.973	-3.784	0	0	0	0	0
<i>NU_LEIS</i>	<i>trv_rds</i> ⁹	3.850	2.188	0	0	0	0	0
	<i>dist_riv</i> ¹⁰	0.909	0.836	0.668	-0.730	na	na	na
<i>RES_SERV</i>	<i>serv_axes</i> ⁶	1.619	1.242	0.504	-0.305	-0.349	-0.536	-0.348
	<i>asph_rds</i> ⁶	3.069	1.324	-0.764	-2.504	0	0	-0.158
<i>RES_MIX</i>	<i>water</i>		<u>Presence 1.300</u>			<u>Absence -0.959</u>		
	<i>serv_axes</i> ⁶	0	0	1.591	1.691	2.294	-0.712	0
	<i>dist_com</i> ⁷	0	1.817	0	0	0	0	0

Note: Distance bands in meters

na : non available

¹ 1: 0 -3000; 2: 3000-6000; 3: 6000-14500; 4: 14500-23500; 5: > 23500

² 1: 0 -2000; 2: 2000-5000; 3: 5000-7500; 4: 7500-10000; 5: 10000-15000; 6: 15000-20000; 7: >20000

³ 1: 0 -2250; 2: 2250-5000; 3: 5000-7500; 4: 7500-10000; 5: 10000-15000; 6: 15000-20000; 7: >20000

⁴ 1: 0 -500; 2: 500-1500; 3: 1500-2500; 4: 2500-3500; 5: 3500-4500; 6: 4500-7000; 7: >7000

⁵ 1: 0 -500; 2: 500-1000; 3: 1000-1500; 4: 1500-2500; 5: 2500-3500; 6: 3500-4500; 7: > 4500

⁶ 1: 0 -250; 2: 250-500; 3: 500-750; 4: 750-1000; 5: 1000-1250; 6: 1250-2000; 7: > 2000

⁷ 1: 0 -2250; 2: 2250-4500; 3: 4500-6650; 4: 6650-8900; 5: 8900-11000; 6: 11000-19750; 7: > 19750

⁸ 1: 0 -1000; 2: 1000-3500; 3: 3500-5000; 4: 5000-7500; 5: 7500-10000; 6: 10000-23500; 7: > 23500

⁹ 1: 0 -1000; 2: 1000-3500; 3: 3500-5000; 4: 5000-7500; 5: 7500-10000; 6: 10000-23500; 7: > 23500

¹⁰ 1: 0 -50; 2: 50-350; 3: 350-500; 4: > 500

By means of the cells transition probabilities, DINAMICA will generate the respective transition probabilities maps (FIGURES 6.7a, 6.7b and 6.7c) for each of the seven types of land use change existing in Bauru from 1967 to 1979. These maps are seen in ERMAPPER, employed by the DOS version of DINAMICA for visualization purposes.

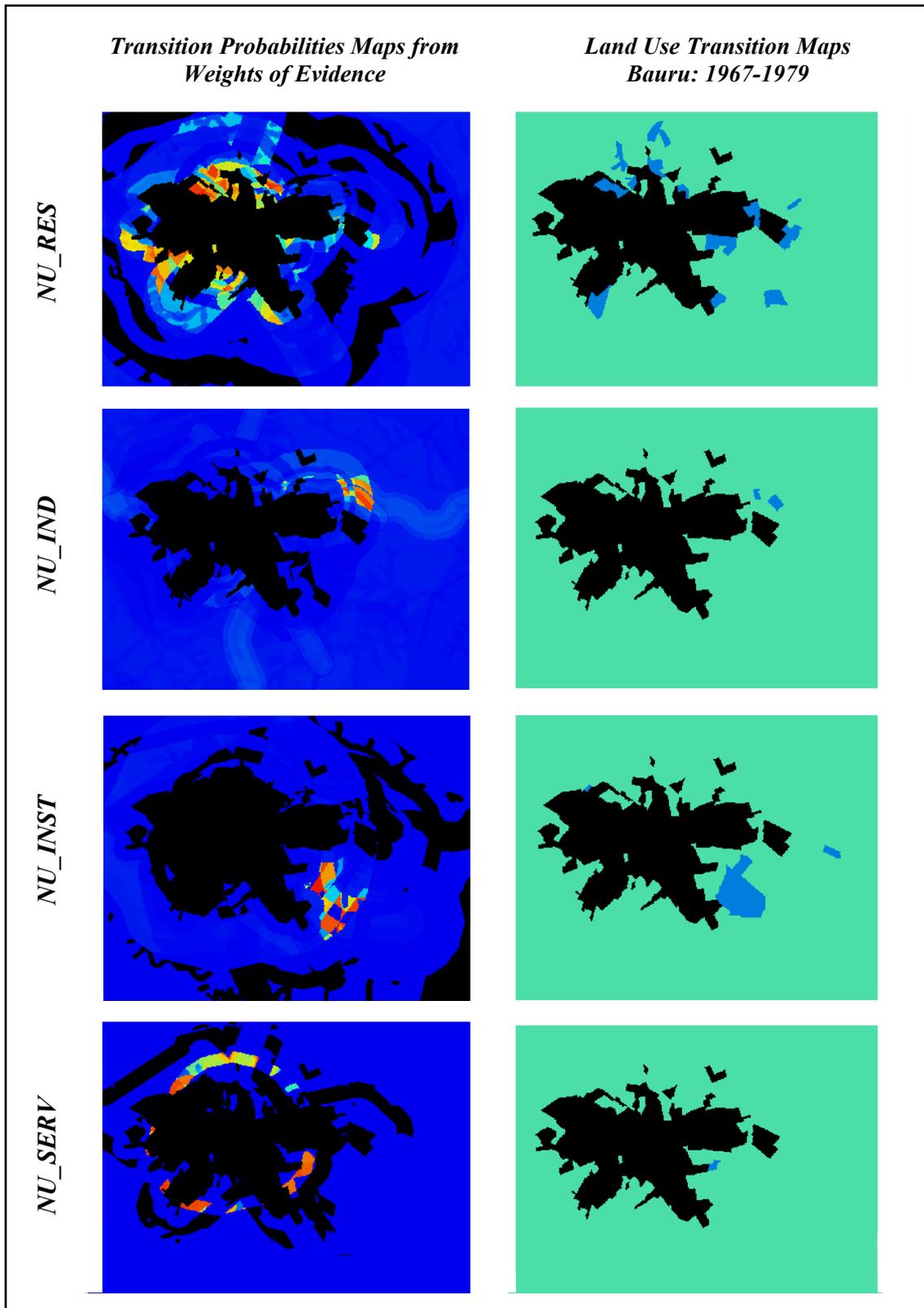
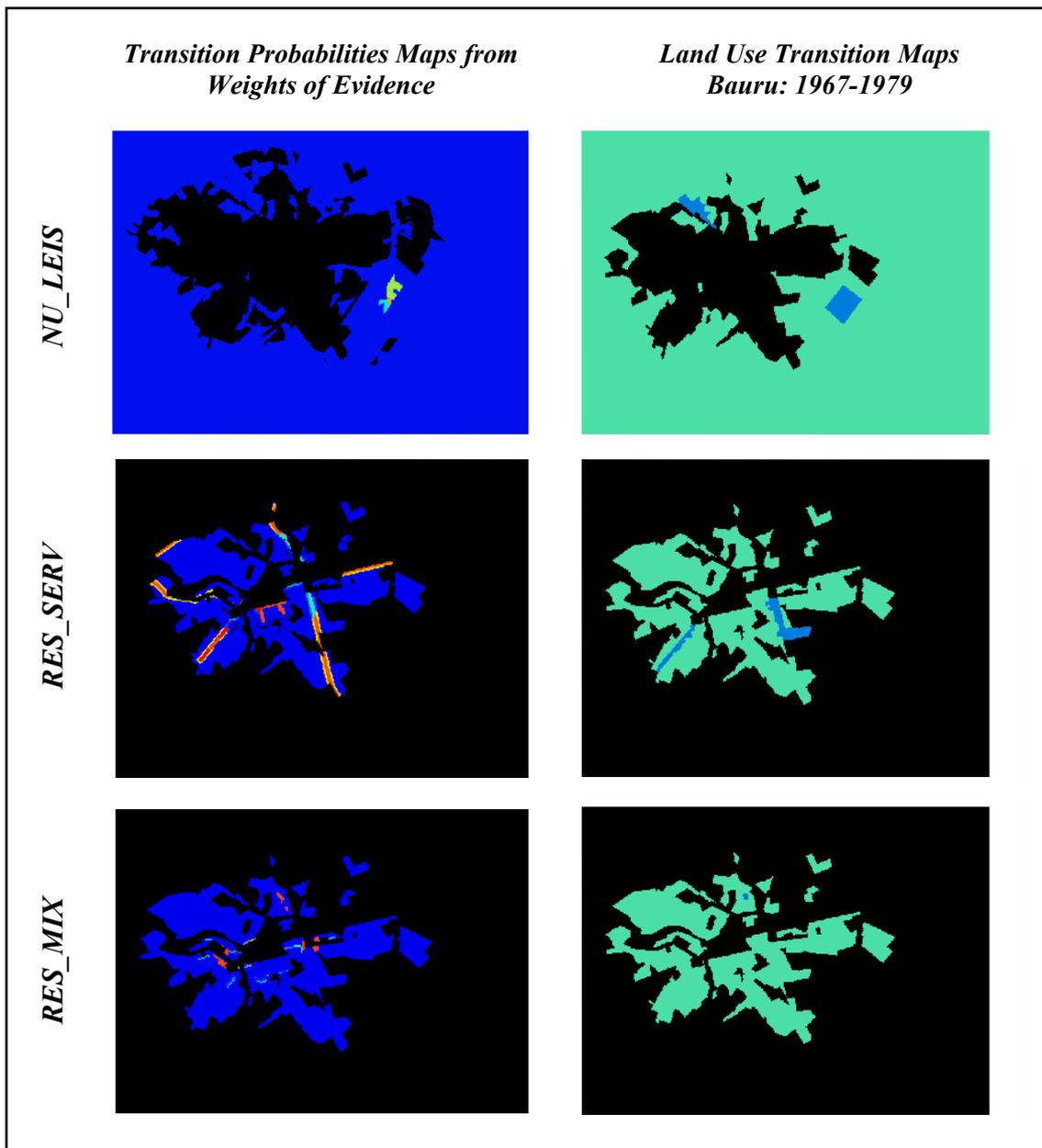


FIGURE 6.7a – Estimated transition probability surfaces and land use change for Bauru.



Legends:

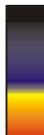
	$P=0$ null probability (increasing ranking) $P=1$ maximum probability	 areas where land use transitions occurred	 areas where no transitions occurred	 areas not considered for transition analyses
---	---	---	---	--

FIGURE 6.7b – Estimated transition probability surfaces and land use change for Bauru.

It is worth remarking the good ability of these probabilities maps to detect the transition areas (blue color) in the corresponding land use transition maps, for all the reddish

regions in the probabilities maps relate to the very areas owning the highest transition probabilities rates.

With the transition probabilities maps, it is possible to define the DINAMICA internal parameters that produce the best simulation results (TABLE 6.7). Due to the randomness of the DINAMICA transition algorithms, even though the same sets of evidences maps for each type of land use transition and the internal parameters are kept in different runs, distinct simulations results will be produced after each run of the model. In this way, the best urban land use simulation results for the city of Bauru in the period 1967–1979 are presented in FIGURE 6.8.

TABLE 6.7 – DINAMICA internal parameters for the simulation of urban land use change in Bauru: 1967–1979.

<i>Land Use Transition</i>	<i>Average Size of Patches</i>	<i>Variance of Patches Size</i>	<i>Proportion of 'Expander'</i>	<i>Proportion of 'Patcher'</i>	<i>Number of Iterations</i>
<i>NU_RES</i>	900	300	0.65	0.35	5
<i>NU_IND</i>	70	1	0	1.00	5
<i>NU_INST</i>	1000	0	0	1.00	5
<i>NU_SERV</i>	40	1	0	1.00	5
<i>NU_LEIS</i>	230	0	0	1.00	5
<i>RES_SERV</i>	25	1	0.10	0.90	5
<i>RES_MIX</i>	30	2	0	1.00	5

The *patcher* algorithm proved to be greatly suited to modeling residential settlements detached from the main urban agglomeration. Nevertheless, the shapes of these settlements in the modeling results do not strictly coincide to those observed in reality. This happens because these contours are associated with limits of real estate properties. Since actions for the merging or split of plots may occur at any time and drastically alter their form, such boundaries can be regarded as highly unstable factors, and thus, inappropriate for modeling.

The exact areas where new residential settlements occur are not always precisely identified. This is due to the fact that these new investments depend on landlords' and

real estate entrepreneurs' decisions, who define the areas to be invested to the detriment of other also advantageous locations. Nevertheless, the aim of modeling is not to reproduce reality as close as possible, but solely to detect main patterns and trends of land use change.

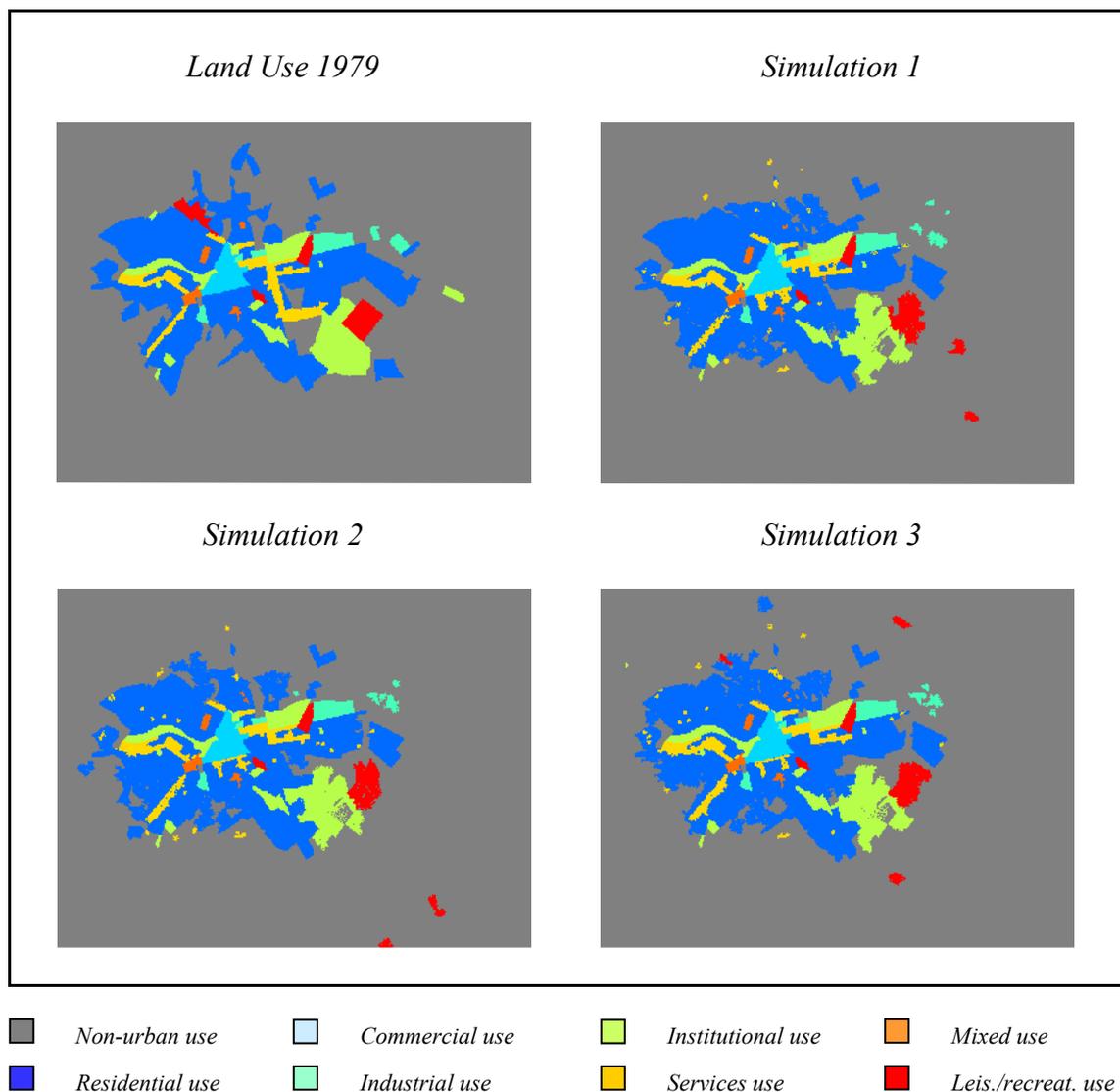


FIGURE 6.8 – The best simulations compared to the actual land use in Bauru in 1979.

These simulation results underwent statistical validation tests based on the multiple resolution fitting procedure proposed by Constanza (1989), for windows size of 3x3, 5x5 and 9x9, and $k = 0.5$ (TABLE 6.8).

TABLE 6.8 – Goodness-of-fit tests for the best land use change simulations of Bauru: 1967-1979.

<i>Simulations</i>	<i>Multiple Resolution Goodness-of-Fit (F)</i>
S_1	$F = 0.944541$
S_2	$F = 0.942551$
S_3	$F = 0.941809$

Upon basis of the carried out calibration process, it becomes evident that the probability of certain non-urban areas in the city of Bauru to shelter residential settlements (“nu_res land use transition”) largely depends on the previous existence of this type of settlements in their surroundings, because this implies the possibility of extending existing nearby infrastructure. It also depends on the greater proximity of these areas to commercial activities clusters as well as on the available accessibility to such areas.

As to the transition of “non-urban areas to industrial use (nu_ind)”, there are three great driving forces: the nearness of such areas to the previously existent industrial use and the availability of road and railway access. This can be explained by the fact that in the industrial production process, the output of certain industries represent the input of other ones, what raises the need of rationalization and optimization of costs by the clustering of plants interrelated in the same productive chain. Furthermore, plots in the vicinities of industrial areas are often prone to be devaluated for other uses, what makes them rather competitive for the industrial use.

Concerning the implementation of large institutional areas (“nu_inst”), it is observable that they arise in farther areas, relatively away from the central commercial zone, near peripheral roads, but at the same time, reasonably close to the demand areas (residential zones).

Regarding the transition of “non-urban areas to services use (nu_serv)”, three major factors are crucial: the proximity of these areas to clusters of commercial activities, their closeness to areas of residential use, and last but not least, their strategic location in

relation to the main urban roads of Bauru. The first factor accounts for the suppliers market (and in some cases also consumers market) of services; the second factor represents the consumers market itself; and the third and last factor corresponds to the accessibility for both markets related to the services use.

The creation of leisure and recreation zones (“nu_leis”), on its turn, takes place in outer areas with good accessibility, and sometimes along low and flat riverbanks, since these areas are floodable and hence unsuitable for sheltering other urban uses.

The transition “residential to services use (res_serv)” supposes the insertion of services into previously consolidated urban areas. In this way, since this transition type already takes place amid the suppliers and consumers markets, it will solely prioritize good accessibility conditions such as the proximity to major asphalted roads and a strategic location in relation to the N-S / E-W services axes of Bauru.

Finally, the last type of land use transition concerns the shift from “residential use to mixed use (res_mix)”. The mixed use zones, which actually play the role of urban sub-centers, constitute a sort of secondary commercial centers enhancement, which at a later stage start to attract services and social infrastructure equipments besides commercial activities themselves. This transition supposes the availability of water supply and the existence of good accessibility conditions. These areas also strive for locating not too far away from central commercial areas, for they depend on the specialized supply from these areas, but not too close to them either for competitiveness reasons.

6.1.2 Simulation Period: 1979 – 1988

6.1.2.1 Weights of Evidence Method

As previously stated in Section 6.1.1, the population of Bauru at the initial time of this simulation period was 164,105 inhabitants, from which 159,926 people were urban inhabitants. In 1988, the total population accounted 236,740, out of which 232,005 inhabitants lived in urban areas (IBGE, 1991). The population growth rate for this period

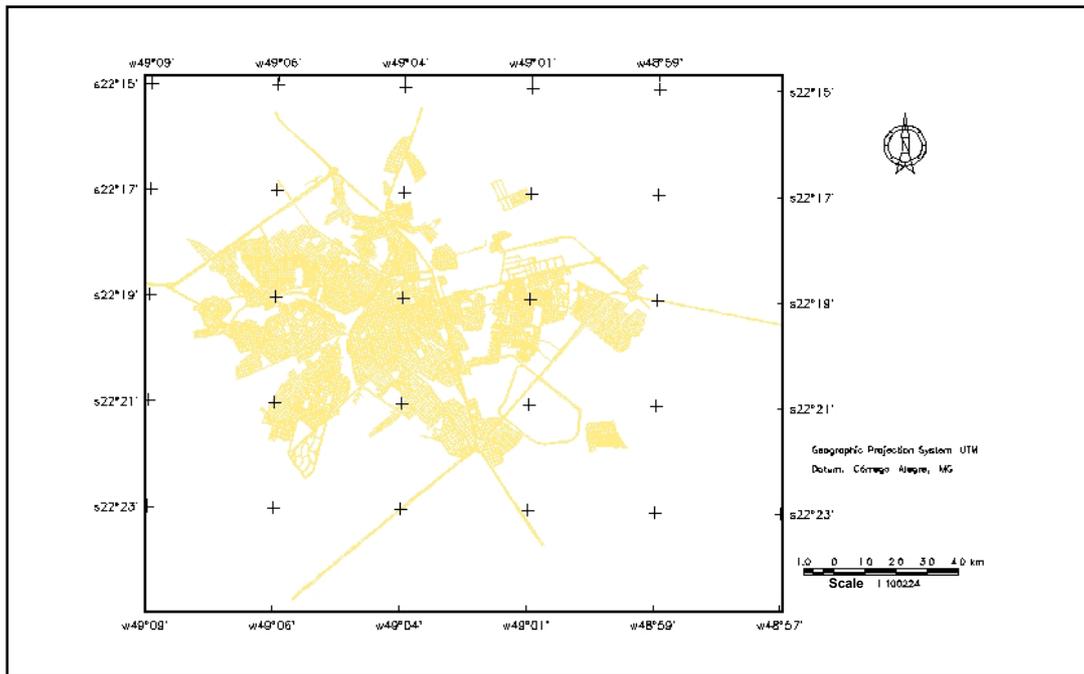


FIGURE 6.9 – Bauru official city map in 1979.
SOURCE: SEPLAN (1979a).

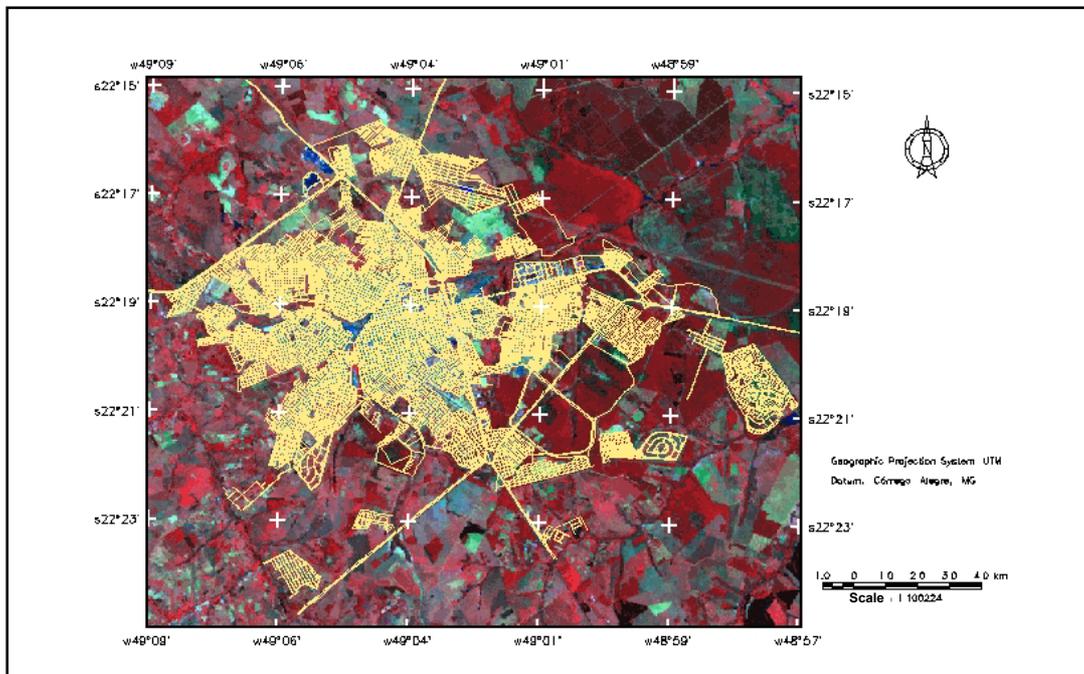


FIGURE 6.10 – Bauru TM – 5 image and official city map in 1988.
SOURCE: INPE (1988) and SEPLAN (1988a).

is comparatively lower than the preceding simulation period, remaining around 1.45%. The shifts in the urban area extension can be seen in FIGURES 6.9 and 6.10, which present the city maps for the initial and final time of simulation.

The initial and final land use maps used in the simulation period 1979 – 1988 (FIGURE 6.11) were elaborated upon basis of the two official city maps previously shown, of generalization procedures applied to the original land use map of 1979 (SEPLAN, 1979a) and 1988 (SEPLAN, 1988a), of the printed satellite image of Bauru in 1979 (INPE, 1979) and of the digital satellite image in 1988 (INPE, 1988).

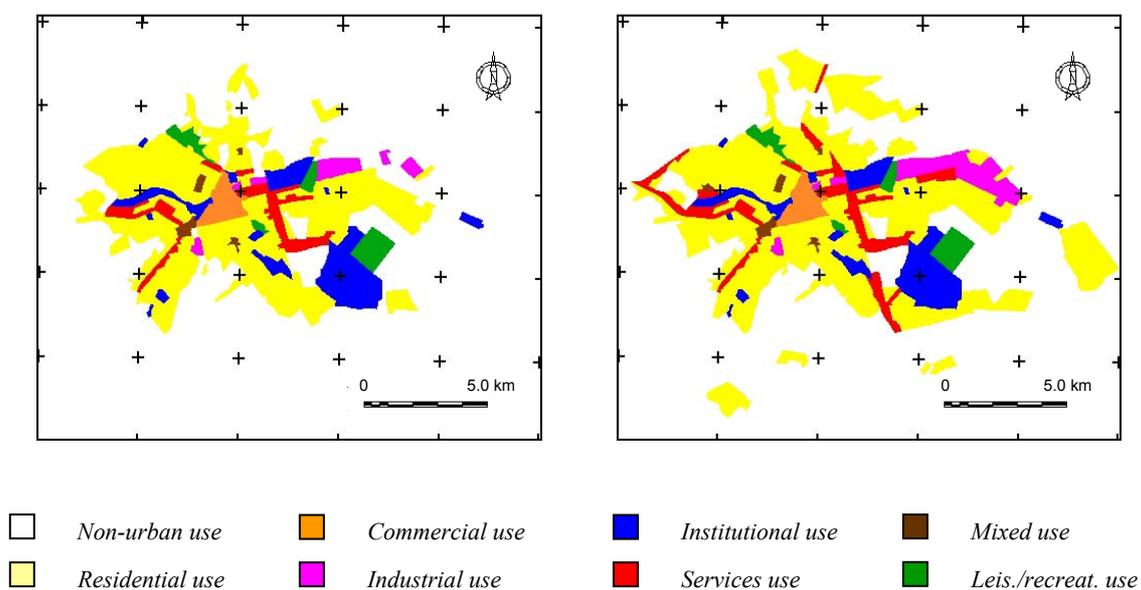


FIGURE 6.11 – Generalized land use map in Bauru in 1979 (left) and 1988 (right).

A cross-tabulation operation was made between both land use maps (FIGURE 6.12) so as to generate transition percentages for the existent types of land use change (TABLES 6.9 and 6.10).

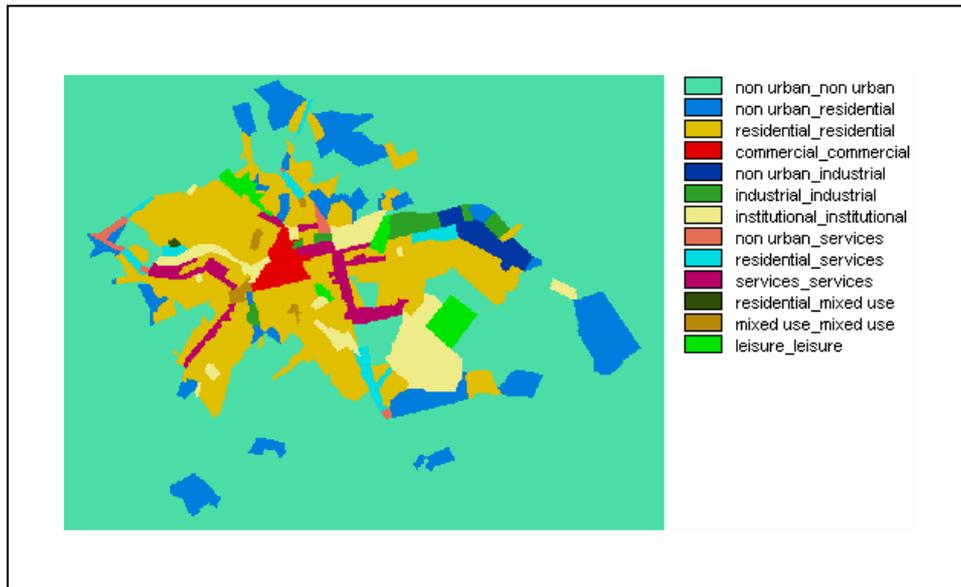


FIGURE 6.12 – Cross-tabulation map between Bauru land use maps of 1979 and 1988.

TABLE 6.9 – Existent land use transitions in Bauru: 1979–1988.

<i>NOTATION</i>	<i>LAND USE TRANSITION</i>
<i>NU_RES</i>	<i>Non-urban to residential</i>
<i>NU_IND</i>	<i>Non-urban to industrial</i>
<i>NU_SERV</i>	<i>Non-urban to services</i>
<i>RES_SERV</i>	<i>Residential to services</i>
<i>RES_MIX</i>	<i>Residential to mixed use</i>

TABLE 6.10 – Matrix of global transition probabilities for Bauru: 1979–1988.

<i>Land Use</i>	<i>Non-urban</i>	<i>Resid.</i>	<i>Comm.</i>	<i>Industr.</i>	<i>Instit.</i>	<i>Services</i>	<i>Mixed</i>	<i>Leis./Recr.</i>
<i>Non-urban</i>	0.9171	0.0698	0	0.0095	0	0.0036	0	0
<i>Resid.</i>	0	0.9380	0	0	0	0.0597	0.0023	0
<i>Comm.</i>	0	0	1	0	0	0	0	0
<i>Industr.</i>	0	0	0	1	0	0	0	0
<i>Instit.</i>	0	0	0	0	1	0	0	0
<i>Services</i>	0	0	0	0	0	1	0	0
<i>Mixed</i>	0	0	0	0	0	0	1	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1

For the simulation period 1979–1988, twelve variables have been selected (DAE, 1979a, 1979b, 1979c, 1979d, 1979d; SEPLAN, 1979a, 1979b), some of which are shown in FIGURE 6.13.

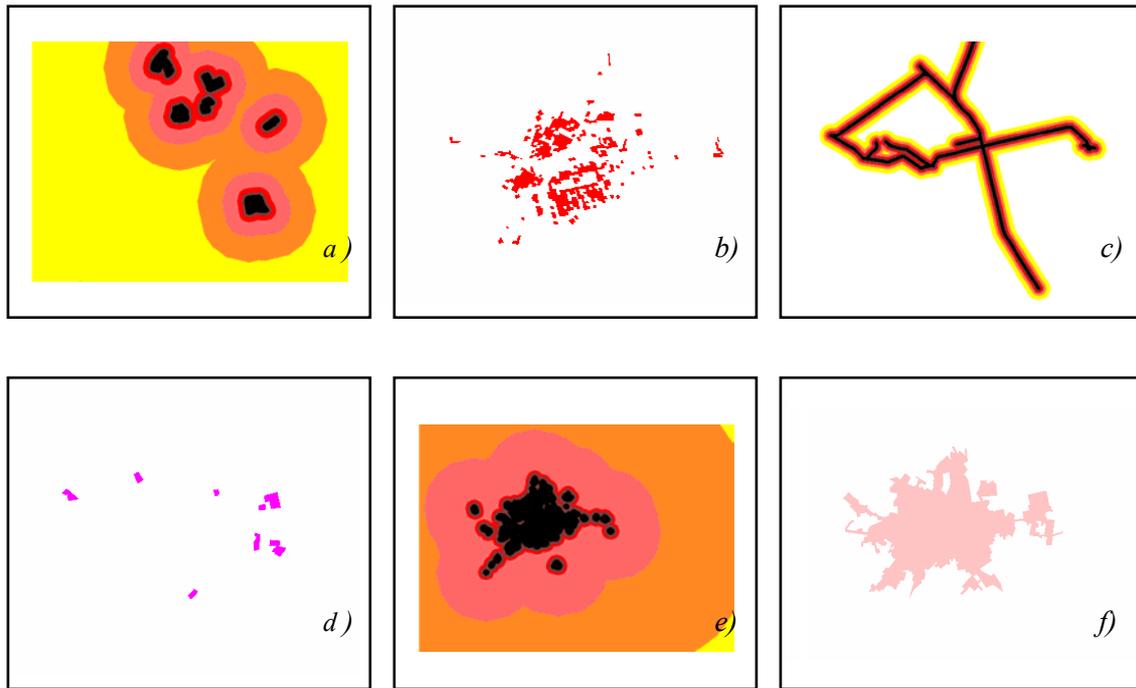


FIGURE 6.13 – Independent variables used to explain the land use transitions in Bauru during the simulation period 1979 – 1988: a) distances to detached residential settlements; b) medium-high density of occupation; c) distances to the services and commercial axes; d) existence of social housing; e) distances to ranges of commercial activities clusters; f) water supply.

TABLE 6.11 shows the notations utilized for each map of variable employed in this simulation experiment; TABLE 6.12 indicates which variable was selected to explain each of the five existent transitions; TABLE 6.13 presents the values obtained for the Cramer’s Coefficient (V) and the Joint Information Uncertainty (U) for the pairs of

variables used to explain the same type of land use transition; and finally TABLE 6.14 presents the values for the positive weights of evidence.

TABLE 6.11 – Independent variables defining land use change in Bauru: 1979–1988.

<i>NOTATION</i>	<i>PHYSICAL OR SOCIOECONOMIC LAND USE CHANGE VARIABLE</i>
<i>water</i>	<i>Area served by water supply.</i>
<i>mh_dens</i>	<i>Medum-high density of occupation (25% to 40%).</i>
<i>soc_hous</i>	<i>Existence of social housing.</i>
<i>com_kern</i>	<i>Distances to different ranges of commercial activities concentration, defined by the Kernel estimator.</i>
<i>dist_ind</i>	<i>Distances to industrial zones.</i>
<i>dist_res</i>	<i>Distances to residential zones.</i>
<i>per_res</i>	<i>Distances to peripheral residential settlements, isolated from the urban concentration.</i>
<i>dist_inst</i>	<i>Distances to social infrastructure (institutional use), isolated from the urban concentration.</i>
<i>exist_rds</i>	<i>Distances to main existent roads.</i>
<i>serv_axes</i>	<i>Distances to the services and commercial axes.</i>
<i>plan_rds</i>	<i>Distances to planned roads.</i>
<i>per_rds</i>	<i>Distances to peripheral roads, which pass through non-occupied areas.</i>

TABLE 6.12 – Selection of variables determining land use change in Bauru: 1979–1988..

<i>NOTATION</i>	<i>NU_RES</i>	<i>NU_IND</i>	<i>NU_SERV</i>	<i>RES_SERV</i>	<i>RES_MIX</i>
<i>water</i>				♦	
<i>mh_dens</i>					♦
<i>soc_hous</i>					♦
<i>com_kern</i>	♦		♦		
<i>dist_ind</i>		♦			
<i>dist_res</i>			♦		
<i>per_res</i>	♦				
<i>dist_inst</i>	♦				
<i>exist_rds</i>	♦				
<i>serv_axes</i>		♦	♦	♦	
<i>plan_rds</i>					♦
<i>per_rds</i>	♦				♦

Likewise the first simulation period, as none of the association measure values surpassed the threshold of 0.50 simultaneously for both indices, no variables preliminarily selected for modeling have been discarded from the analysis.

TABLE 6.13 – Associations between independent variables - Bauru: 1979–1988.

<i>VARIABLE A</i>	<i>VARIABLE B</i>	<i>CRAMER'S STATISTIC (V_{A,B})</i>	<i>UNCERTAINTY (U_{A,B})</i>
<i>water</i>	<i>serv_axes</i>	0.3257	0.0767
<i>mh_dens</i>	<i>soc_hous</i>	0.0460	0.0017
	<i>plan_rds</i>	0.2617	0.0701
	<i>per_rds</i>	0.0201	0.0003
<i>soc_hous</i>	<i>plan_rds</i>	0.1174	0.0188
	<i>per_rds</i>	0.0480	0.0047
<i>com_kern</i>	<i>dist_res</i>	0.4129	0.3447
	<i>per_res</i>	0.1142	0.0310
	<i>dist_inst</i>	0.1218	0.0520
	<i>exist_rds</i>	0.2685	0.1499
	<i>serv_axes</i>	0.2029	0.1099
	<i>per_rds</i>	0.0434	0.0064
	<i>dist_inst</i>	<i>serv_axes</i>	0.1466
<i>dist_res</i>	<i>serv_axes</i>	0.2142	0.1002
<i>per_res</i>	<i>dist_inst</i>	0.1487	0.0559
	<i>exist_rds</i>	0.0592	0.0078
	<i>per_rds</i>	0.1733	0.0553
<i>dist_inst</i>	<i>exist_rds</i>	0.0601	0.0108
	<i>per_rds</i>	0.0765	0.0238
<i>exist_rds</i>	<i>per_rds</i>	0.0239	0.0019
<i>plan_rds</i>	<i>per_rds</i>	0.0247	0.0029

TABLE 6.14 – Values of W^+ for the selected independent variables - Bauru: 1979–1988.

<i>Land Use Transition</i>	<i>Variable</i>	<i>Positive Weights of Evidence W⁺</i>						
		1	2	3	4	5	6	7
<i>NU_RES</i>	<i>com_kern</i> ¹	3.749	2.106	1.864	0.491	-0.323	0	na
	<i>per_res</i> ³	1.968	1.615	1.392	0.892	-0.626	-0.469	na
	<i>dist_inst</i> ⁴	0.003	0.600	1.254	0.727	-0.359	-0.089	na
	<i>exist_rds</i> ⁵	0.231	0.320	0.353	0.510	0.443	0.196	-0.085
	<i>per_rds</i> ⁶	2.377	2.269	2.068	1.984	1.444	0.857	-0.127
<i>NU_IND</i>	<i>dist_inst</i> ²	3.862	4.016	3.792	3.452	1.763	0	0
	<i>serv_axes</i> ⁵	2.722	2.799	2.676	2.625	2.525	1.727	-3.832
<i>NU_SERV</i>	<i>com_kern</i> ¹	3.412	4.469	2.912	0.878	0	0	na
	<i>dist_res</i> ³	2.144	1.523	0.621	-0.065	0	0	na
	<i>serv_axes</i> ⁵	3.508	3.321	2.917	1.869	0.450	0	0
<i>RES_SERV</i>	<i>water</i>	<u>Presence -0.6611</u>			<u>Absence 0.2883</u>			
	<i>serv_axes</i> ⁵	2.780	1.948	1.461	0.888	-0.297	-1.412	-3.284
<i>RES_MIX</i>	<i>mh_dens</i>	<u>Presence 0.6452</u>			<u>Absence -0.0635</u>			
	<i>soc_hous</i>	<u>Presence 2.4678</u>			<u>Absence -0.3214</u>			
	<i>plan_rds</i> ⁵	3.506	1.863	0	0	0	0	0
	<i>per_rds</i> ⁶	1.775	1.652	1.848	0.903	0	0	0

Note: Distance bands in meters

na : non available

¹ 1: 0 -500; 2: 500-1000; 3: 1000-1500; 4: 1500-10000; 5: 10000-30000; 6: > 30000

² 1: 0 -500; 2: 500-1000; 3: 1000-1500; 4: 1500-2000; 5: 2000-5000; 6: 5000-10000; 7: >10000

³ 1: 0 -500; 2: 500-1000; 3: 1000-2000; 4: 2000-5000; 5: 5000-10000; 6: > 10000

⁴ 1: 0 -500; 2: 500-1000; 3: 1000-3000; 4: 3000-8000; 5: 8000-15000; 6: > 15000

⁵ 1: 0 -250; 2: 250-500; 3: 500-750; 4: 750-1000; 5: 1000-1250; 6: 1250-2000; 7: > 2000

⁶ 1: 0 -250; 2: 250-500; 3: 500-750; 4: 750-1000; 5: 1000-1500; 6: 1500-2500; 7: > 2500

The maps of estimated transition probabilities surfaces, generated by DINAMICA upon basis of the values of the positive weights of evidence (W^+), together with their respective land use transition maps are seen in FIGURES 6.14 a and 6.14b.

The arguments employed to explain the transitions “nu_res”, “nu_ind” and “nu_serv” in the period 1967-1979 (Section 6.1.1) are also valid to justify these same transitions in the current simulation period. Slight differences concern the requirements for the transitions “res_serv” and “res_mix”.

The transition “residential to services use (res_serv)” from 1979 to 1988 supposes the insertion of services into residential areas. In this way, since this transition type already takes place amid the suppliers and consumers markets, it will prioritize the strategic location in relation to the N-S / E-W services axes of Bauru and will take place in the absence of water supply, considering the initial time of simulation. Note that in this case W^+ is negative. This means that areas close to the services axes but deprived of water supply in the initial time of simulation refer to those areas located in the immediate fringe of consolidated urban areas, exactly where new services zones are prone to occur.

The decisive factors for the conversion of “residential areas into mixed use zones (res_mix)” in the period 1979-1988 are (a) existence of medium-high density of occupation (higher density values only occur in the central commercial zone of the town or in the immediacies of already existent mixed use zones); (b) presence or proximity of social housing settlements (for they shelter the greatest occupational densities in more peripheral areas, and hence, greater consumers markets); (c) nearness to planned or peripheral roads, since new mixed use zones arise in farther areas of the town, so as not to compete with the central commercial zone.

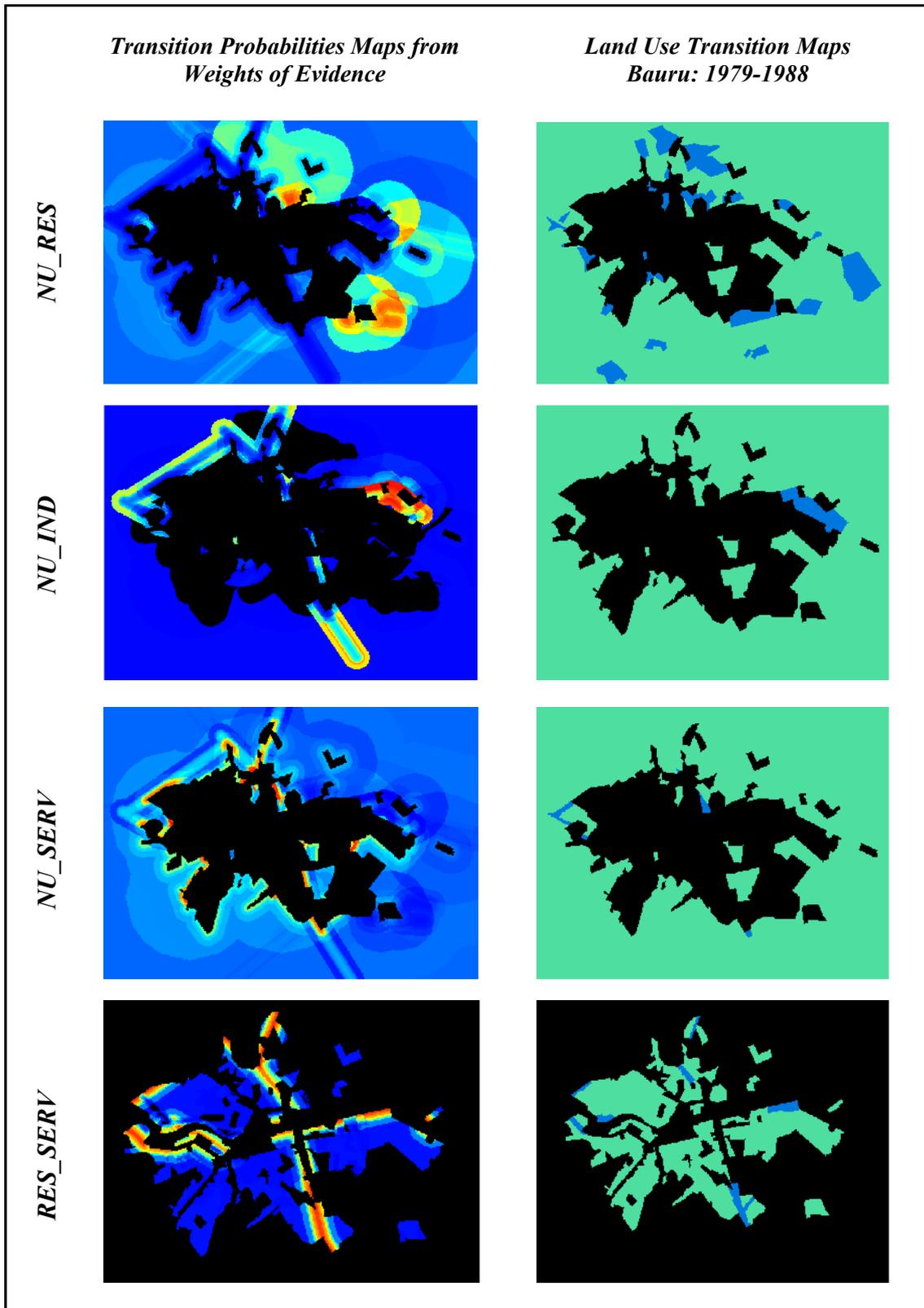


FIGURE 6.14a – Estimated transition probability surfaces and land use change - Bauru.

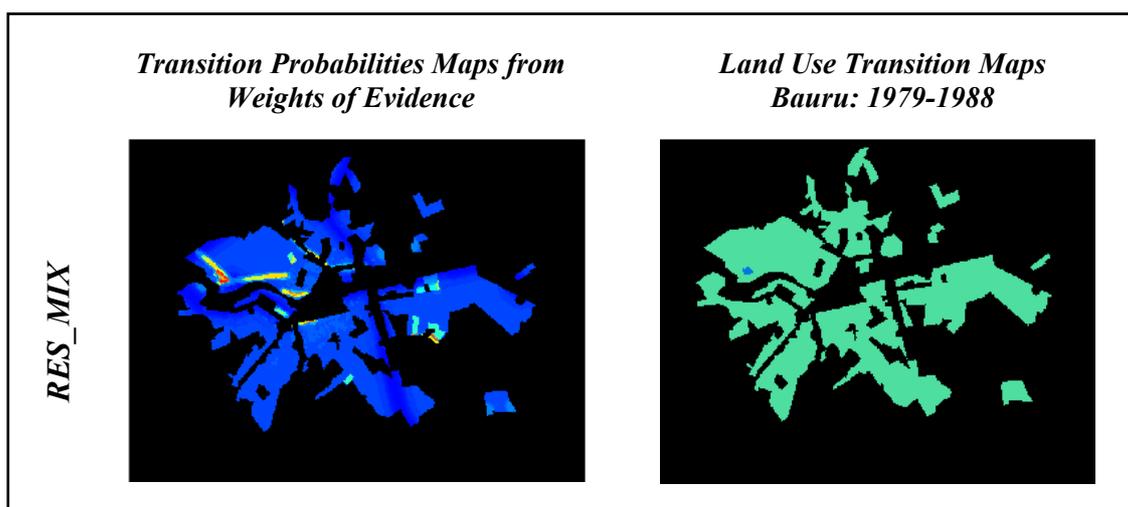


FIGURE 6.14b – Estimated transition probability surfaces and land use change – Bauru.

The three best simulation results produced for the period 1979-1988 are presented in FIGURE 6.15. The internal DINAMICA parameters associated with these optimal simulations are seen in TABLE 6.15, whose statistical validation tests for windows size of 3x3, 5x5 and 9x9, and $k = 0.5$ are listed on TABLE 6.16.

TABLE 6.15 – DINAMICA internal parameters for the simulation of urban land use change in Bauru: 1979–1988.

<i>Land Use Transition</i>	<i>Average Size of Patches</i>	<i>Variance of Patches Size</i>	<i>Proportion of 'Expander'</i>	<i>Proportion of 'Patcher'</i>	<i>Number of Iterations</i>
<i>NU_RES</i>	1100	500	0.65	0.35	5
<i>NU_IND</i>	320	1	1.00	0	5
<i>NU_SERV</i>	25	2	0.50	0.50	5
<i>RES_SERV</i>	25	2	0.10	0.90	5
<i>RES_MIX</i>	35	2	0	1.00	5

TABLE 6.16 – Goodness-of-fit tests for the best land use change simulations produced by the weights of evidence method for Bauru: 1979-1988.

<i>Simulations</i>	<i>Multiple Resolution Goodness-of-Fit (F)</i>
S_1	$F = 0.903103$
S_2	$F = 0.896722$
S_3	$F = 0.901703$

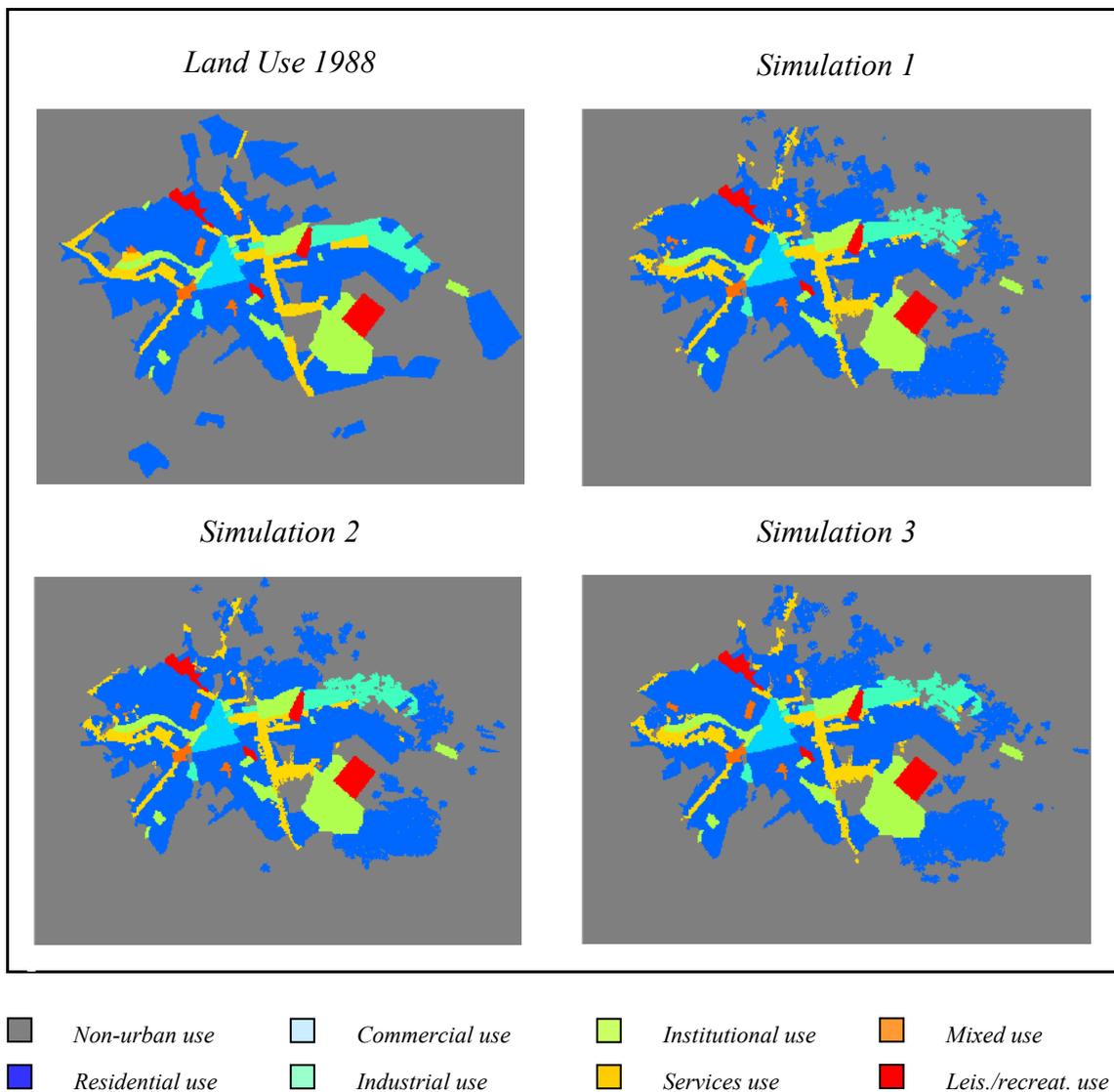


FIGURE 6.15 – The best simulations produced by the weights of evidence method compared to the actual land use in Bauru in 1988.

6.1.2.2 Logistic Regression Method

In the logistic regression method, all continuous variables, i.e. all independent variables related to maps of distances, were treated as categorical, owning thematic classes associated to each range of distance. This has been done in view of the fact that practically all of them presented a non-linear and/or multi-modal behavior in relation to the respective land use transitions. Handling this information as continuous variables would imply a great data heterogeneity, what would certainly bring about noise in the simulations and harm the model calibration.

According to what was already stated in Section 5.3.2.2, empirical procedures were used for the variables selection, like the visualization of distinct variables superimposed on the final land use map, what aimed at identifying the set of those ones more meaningful to explain the five different types of land use change. Another auxiliary method was the analysis of boxplots generated by each selected independent variable versus the respective land use transition. FIGURE 6.16 shows examples of boxplots. It is observable for the first case that the majority of the cells where the transition from residential to mixed use occurs coincide with the cells where social housing units are also found. In the second case, the boxplot shows that the transition from non-urban to industrial use takes place in the closest areas from the already existent industrial zones.

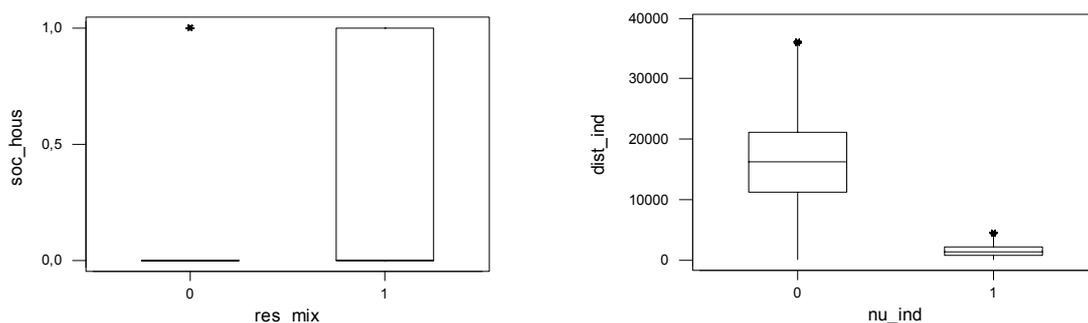


FIGURE 6.16 – Boxplot of the transition ‘residential to mixed use’ versus social housing (left) and boxplot of the transition ‘non-urban to industrial use’ versus distances to industrial zones (right) – Bauru: 1979-1988.

The exploratory analysis also took into account the correlation index (A), assessed for pairs of variables selected to explain the same type of land use transition. This proved to be a measure highly sensitive to the spatial autocorrelation between cells, for it is obtained from a pixel-based statistics. Thence, this index has not been used as an eliminatory criterion for the variables selection, and the association measures provided by the Cramer's Statistics (V) and the Joint Information Uncertainty (U) were prevailing also in the exploratory analysis for the logistic regression method. TABLE 6.17 presents the values obtained for the correlation index (A).

TABLE 6.17 – Correlations between independent variables - Bauru: 1979–1988.

<i>VARIABLE A</i>	<i>VARIABLE B</i>	<i>CORRELATION INDEX ($A_{A,B}$)</i>
<i>water</i>	<i>serv_axes</i>	-0.3060
<i>mh_dens</i>	<i>soc_hous</i>	0.0530
	<i>plan_rds</i>	-0.1600
	<i>per_rds</i>	-0.0560
<i>soc_hous</i>	<i>plan_rds</i>	-0.0760
	<i>per_rds</i>	-0.0440
<i>com_kern</i>	<i>dist_res</i>	0.9050
	<i>per_res</i>	0.1580
	<i>dist_inst</i>	-0.2930
	<i>exist_rds</i>	0.7670
	<i>serv_axes</i>	0.7060
<i>dist_ind</i>	<i>per_rds</i>	0.4490
	<i>serv_axes</i>	0.5630
<i>dist_res</i>	<i>serv_axes</i>	0.7900
<i>per_res</i>	<i>dist_inst</i>	0.7070
	<i>exist_rds</i>	0.1860
	<i>per_rds</i>	0.5380
<i>dist_inst</i>	<i>exist_rds</i>	-0.1190
	<i>per_rds</i>	0.2170
<i>exist_rds</i>	<i>per_rds</i>	0.3090
<i>plan_rds</i>	<i>per_rds</i>	0.0658

In order to estimate the parameters (β_i) of the logistic regression model, the independent variables and the land use transition maps were converted to numerical raster grids in SPRING. These transition maps were generated through a maps algebra operation in SPRING by means of a language intitled “Spatial Language for Algebraic Geocomputation” (‘LEGAL’ - *Linguagem Espacial para Geoprocessamento Algebrico*).

The numerical grids were converted into columns text files and then exported to the statistical package MINITAB 13.0, where separate database for each of the five types of land use change were built. FIGURE 6.16 synthesizes these conversion procedures.

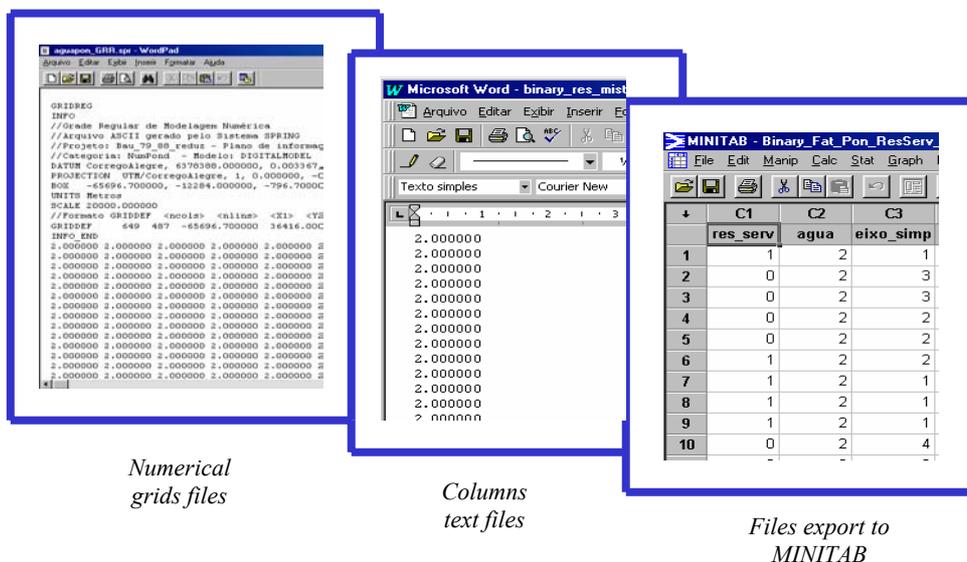


FIGURE 6.17 – Conversion of numerical raster grids into statistical database in MINITAB 13.0.

According to what was previously exposed in Section 5.2.2.1, the method “backward stepwise” was adopted for the parameters estimation, which can be seen in TABLE 6.18. Although the variables “*dist_res*” and “*mh_dens*” were not significant at the 0.05 level ($p < 0.05$), they were kept in the model in view of their effective contribution for explaining the transitions “*nu_serv*” and “*res_mix*”, respectively. According to Hosmer and Lemeshow (1989), a diversified set of criteria other than statistical significance tests must be taken into consideration for including or removing variables from a model.

Some variables presented mild support for being included in the model as cases of interaction and confounding. Nevertheless, when this was the case, the statistical package routines could not reach a convergence of parameters even after a high number of iterations (20,000). Consequently, the products of these variables have been suppressed from the model.

It is noticeable in TABLE 6.18 that in all models, except for the Hosmer-Lemeshow test concerning the transition “res_mix”, goodness-of-fit tests do not indicate statistical significance ($p > 0.05$). As stated in Section 5.2.2.5, these tests results should be regarded in a critical and wise way. In any case, the fitting of these five models was effectively evaluated by means of the statistical validation test proposed by Constanza (1989), where the models spatial dimension is duly and carefully taken into account.

TABLE 6.18 – Results of the logistic regression analyses for Bauru: 1979–1988.

VARIABLES	NU_RES		NU_IND		NU_SERV		RES_SERV		RES_MIX	
	β_k	P								
Constant (β_0)	7.646900	0.000	5.274530	0.000	4.865300	0.000	-1.551900	0.000	3.901200	0.000
water	#	#	#	#	#	#	1.708810	0.000	#	#
mh_dens	#	#	#	#	#	#	#	#	0.383300	0.232
soc_hous	#	#	#	#	#	#	#	#	-1.068800	0.000
com_kern	-0.924990	0.000	#	#	-1.461660	0.000	#	#	#	#
dist_ind	#	#	-1.048320	0.000	#	#	#	#	#	#
dist_res	#	#	#	#	0.027680	0.442	#	#	#	#
per_res	-0.392090	0.000	#	#	#	#	#	#	#	#
dist_inst	-0.405525	0.000	#	#	#	#	#	#	#	#
exist_rds	0.051476	0.000	#	#	#	#	#	#	#	#
serv_axes	#	#	-0.741110	0.000	-0.974470	0.000	-0.929550	0.000	#	#
plan_rds	#	#	#	#	#	#	#	#	-1.865200	0.000
per_rds	-0.309469	0.000	#	#	#	#	#	#	-0.521040	0.000
RESULTS FOR THE TESTS OF GOODNESS-OF-FIT										
Tests	Chi-square	P								
Pearson	41,202.475	0.000	13,639.316	0.000	938.120	0.000	338.064	0.000	422.206	0.000
Deviance	30,435.653	0.000	6,055.790	0.000	774.369	0.000	341.693	0.000	328.558	0.000
Hosmer-Lem.	613.082	0.000	258.618	0.000	44.667	0.000	247.916	0.000	1.653	0.438

With the parameters (β_i) estimated for each of the five types of land use change, DINAMICA will produce likewise in the weights of evidence method the maps of transition probabilities surfaces for each of the existent types of transition. These maps and their respective land use change maps are seen in FIGURES 6.17a and 6.17b.

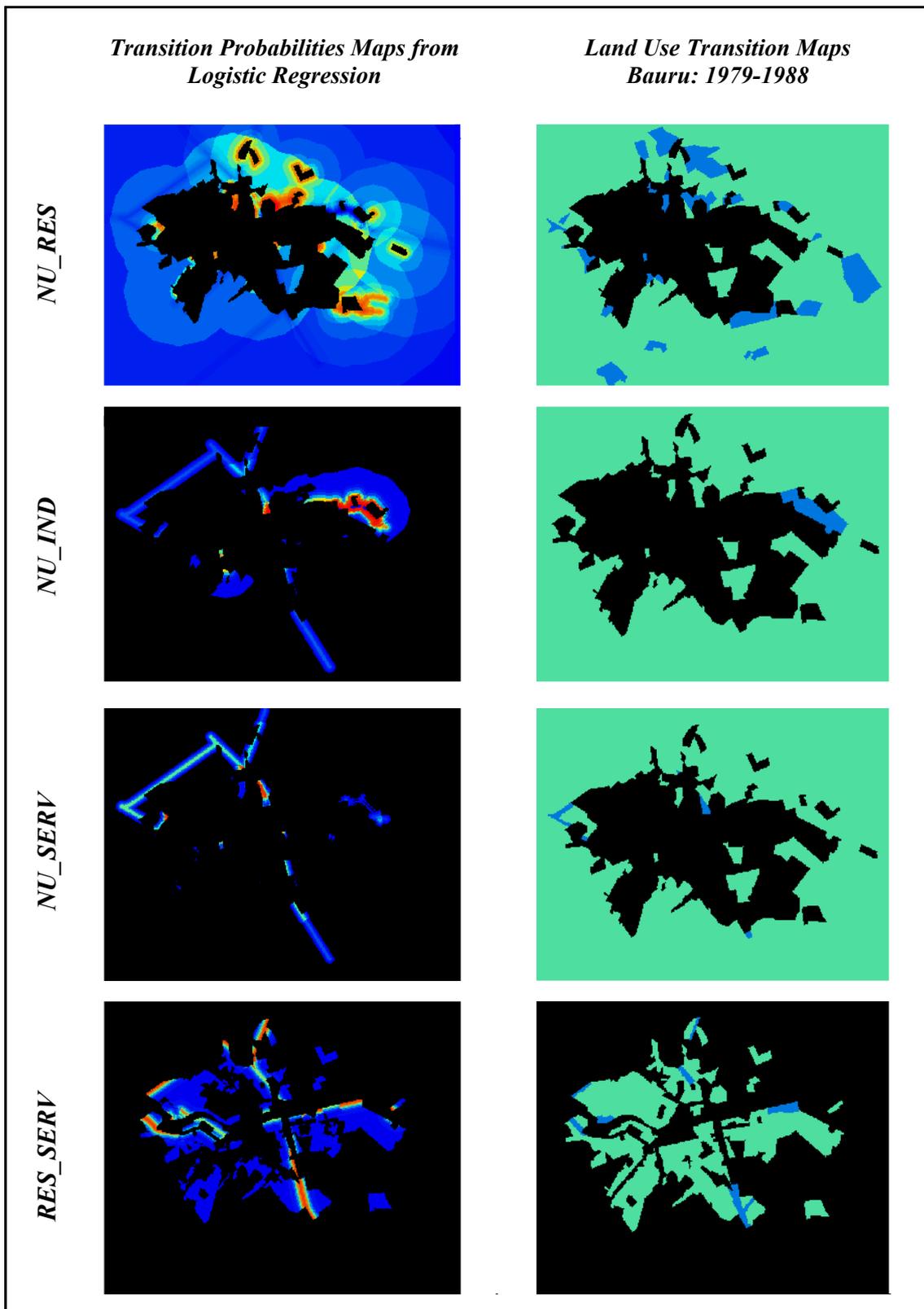


FIGURE 6.18a – Estimated transition probability surfaces and land use change - Bauru.

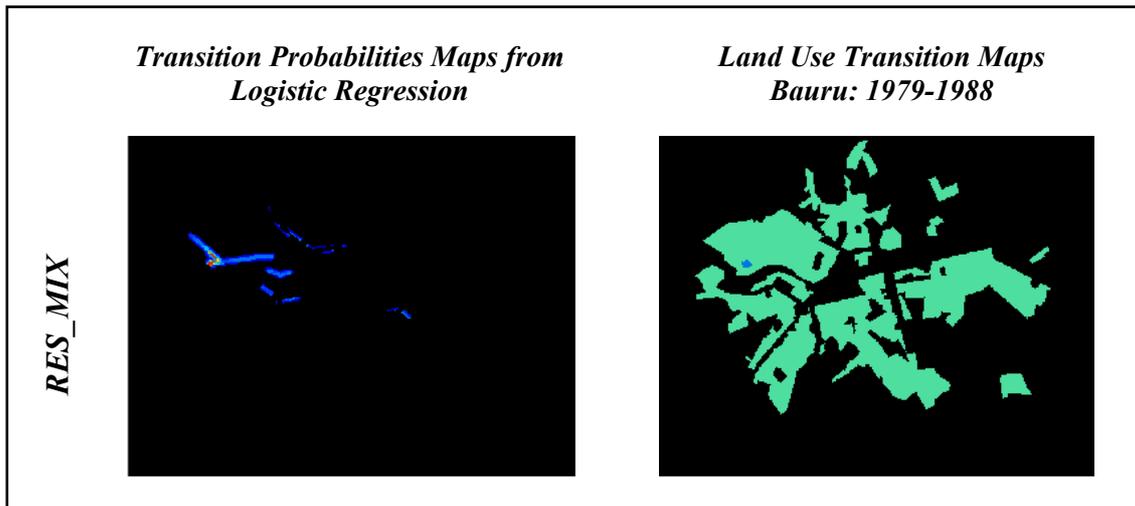


FIGURE 6.18b – Estimated transition probability surfaces and land use change – Bauru.

The three best simulation results produced for the period 1979-1988 are presented in FIGURE 6.19. The internal DINAMICA parameters associated with these optimal simulations are seen in TABLE 6.15, whose statistical validation tests for windows size of 3x3, 5x5 and 9x9, and $k = 0.5$ are listed on TABLE 6.19.

TABLE 6.19 – Goodness-of-fit tests for the best land use change simulations produced by the logistic regression method for Bauru: 1979-1988.

<i>Simulations</i>	<i>Multiple Resolution Goodness-of-Fit (F)</i>
S_1	$F = 0.905172$
S_2	$F = 0.907539$
S_3	$F = 0.907868$

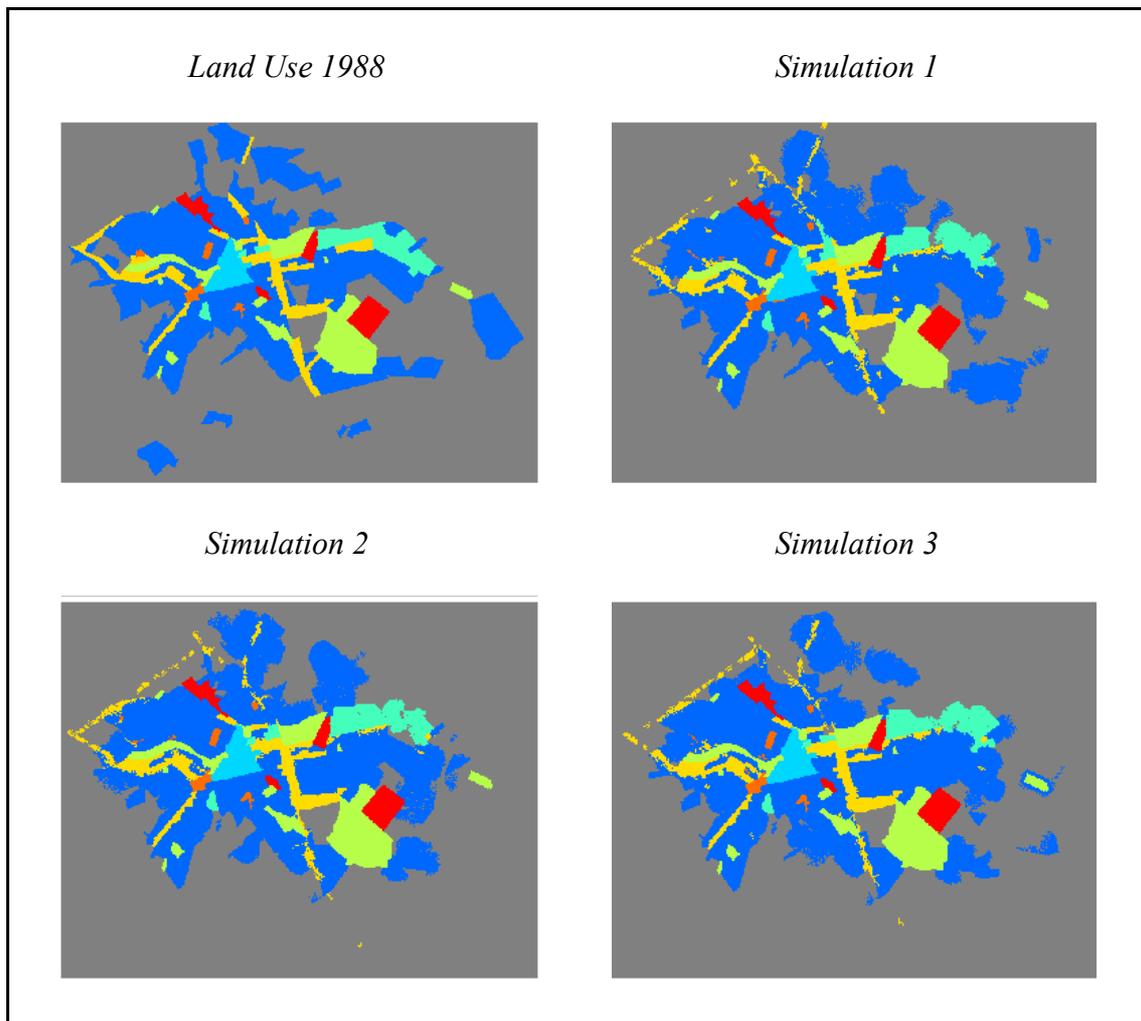


FIGURE 6.19 – The best simulations produced by the logistic regression method compared to the actual land use in Bauru in 1988.

6.1.3 Simulation Period: 1988 – 2000

According to what was stated in Section 6.1.2, the population of Bauru at the initial time of this simulation period was 236,740 inhabitants, from which 232,005 people were urban

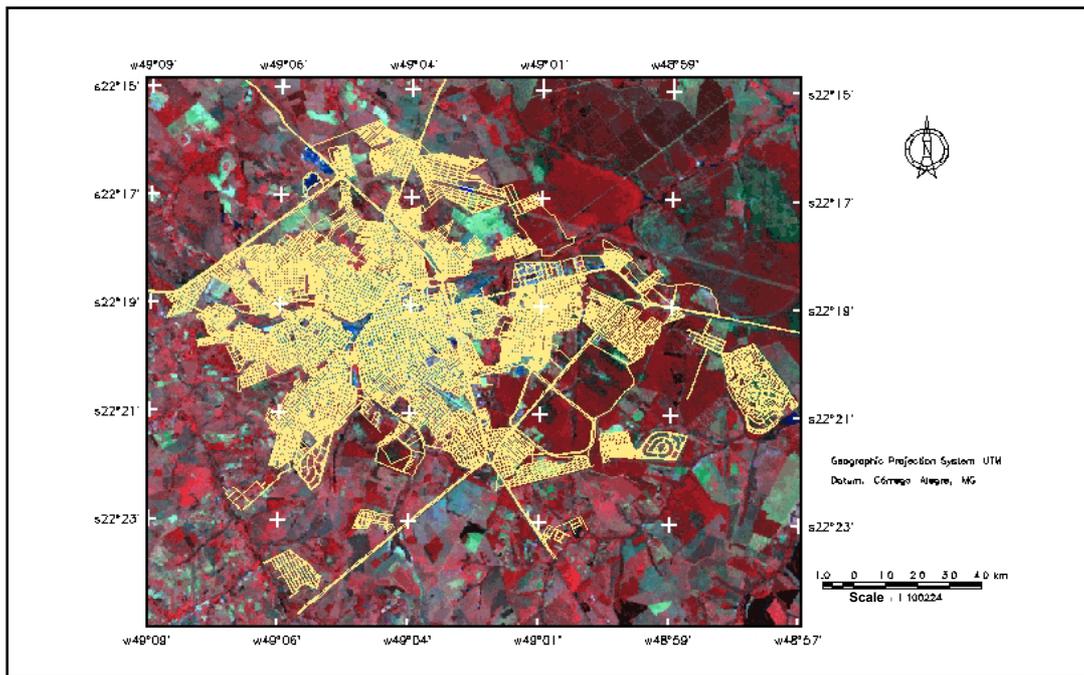


FIGURE 6.20 – Bauru TM – 5 image and official city map in 1988.
 SOURCE: INPE (1988) and SEPLAN (1988a).

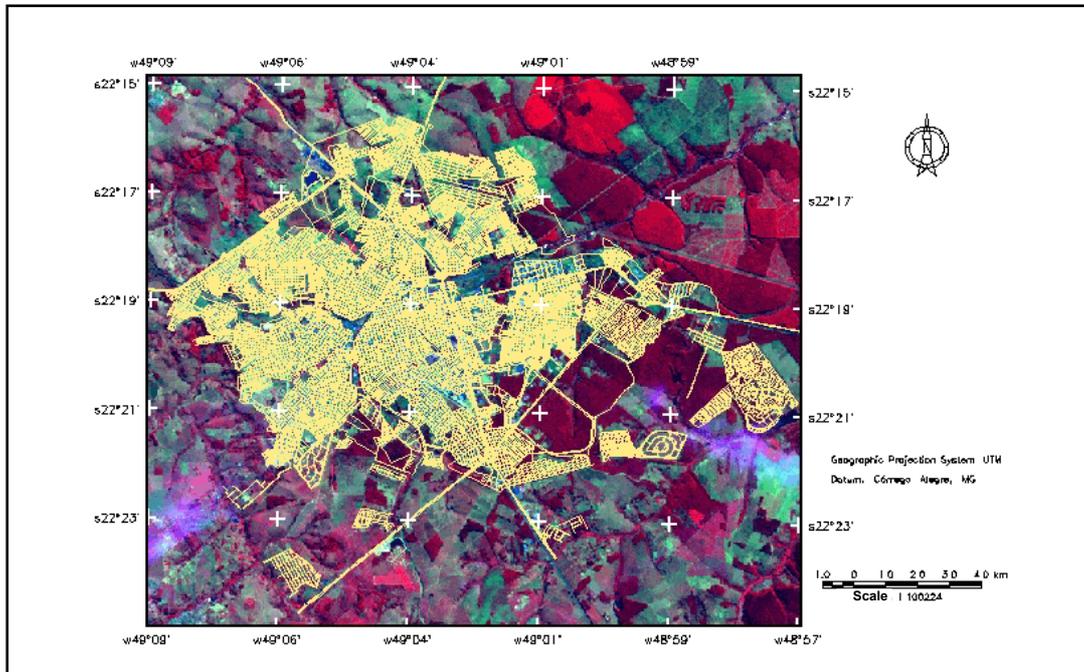


FIGURE 6.21 – Bauru TM – 5 image and official city map in 2000.
 SOURCE: INPE (2000) and DAE (2000).

inhabitants. In 2000, the total population rose to 316,064, out of which 310,442 inhabitants lived in urban areas (IBGE, 2000). The population growth rate for this period is slightly lower than the preceding simulation period, lying around 1.34%. The impacts of this population growth on the urban area extension can be seen in FIGURES 6.20 and 6.21, which present the city maps for the initial and final time of simulation.

The initial and final land use maps used in the simulation period 1988 – 2000 (FIGURE 6.22) were elaborated upon basis of the two official city maps previously shown, of generalization procedures applied to the original land use map of 1988 (SEPLAN, 1988a) and 2000 (DAE, 2000) and of the digital satellite images of Bauru in 1988 and 2000 (INPE, 1988, 2000).

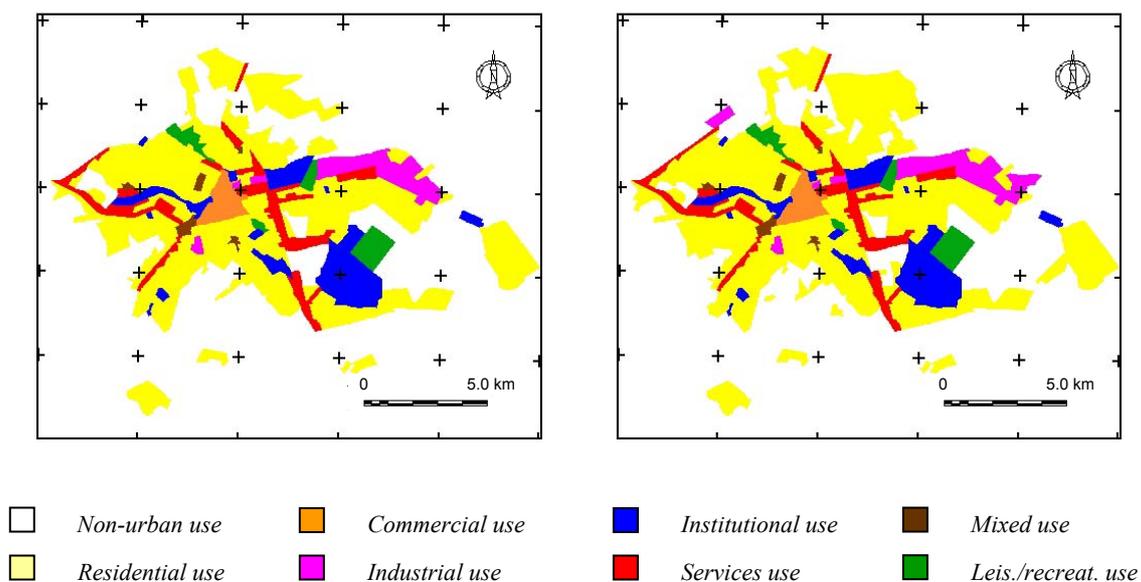


FIGURE 6.22 – Generalized land use map in Bauru in 1988 (left) and 2000 (right).

A cross-tabulation operation was made between both land use maps (FIGURE 6.23) so as to generate transition percentages for the existent types of land use change (TABLES 6.20 and 6.21).

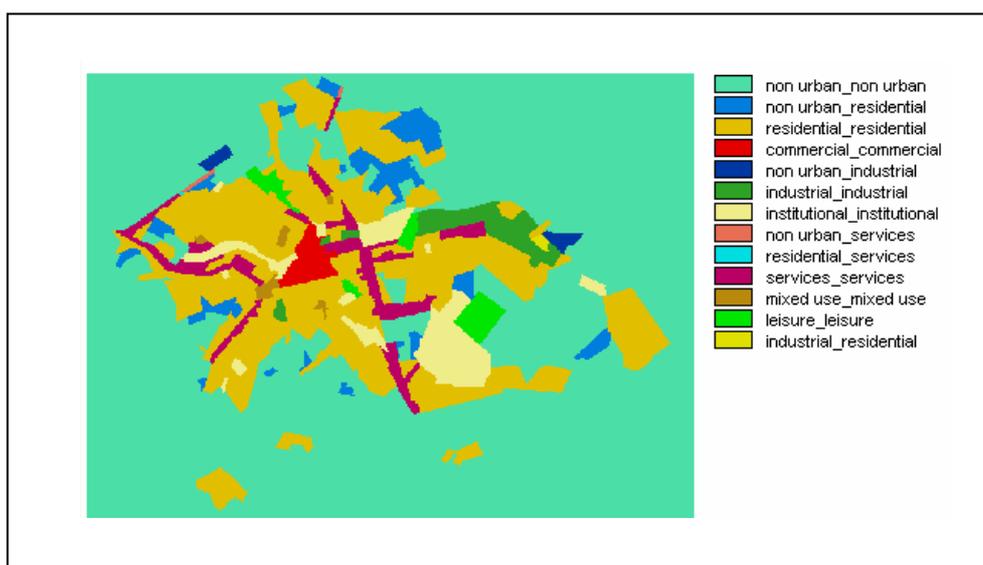


FIGURE 6.23 – Cross-tabulation map between Bauru land use maps of 1988 and 2000.

TABLE 6.20 – Existent land use transitions in Bauru: 1988–2000.

<i>NOTATION</i>	<i>LAND USE TRANSITION</i>
<i>NU_RES</i>	<i>Non-urban to residential</i>
<i>NU_IND</i>	<i>Non-urban to industrial</i>
<i>NU_SERV</i>	<i>Non-urban to services</i>
<i>RES_SERV</i>	<i>Residential to services</i>
<i>IND_RES</i>	<i>Industrial to residential</i>

TABLE 6.21 – Matrix of global transition probabilities for Bauru: 1988–2000.

<i>Land Use</i>	<i>Non-urban</i>	<i>Resid.</i>	<i>Comm.</i>	<i>Industr.</i>	<i>Instit.</i>	<i>Services</i>	<i>Mixed</i>	<i>Leis./Recr.</i>
<i>Non-urban</i>	0.9615	0.0333	0	0.0043	0	0.0009	0	0
<i>Resid.</i>	0	0.9997	0	0	0	0.0003	0	0
<i>Comm.</i>	0	0	1	0	0	0	0	0
<i>Industr.</i>	0	0.0438	0	1	0	0	0	0
<i>Instit.</i>	0	0	0	0	1	0	0	0
<i>Services</i>	0	0	0	0	0	1	0	0
<i>Mixed</i>	0	0	0	0	0	0	1	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1

For the simulation period 1988–2000, seven variables have been selected (SEPLAN, 1988a, 1988b), most of which are shown in FIGURE 6.24.

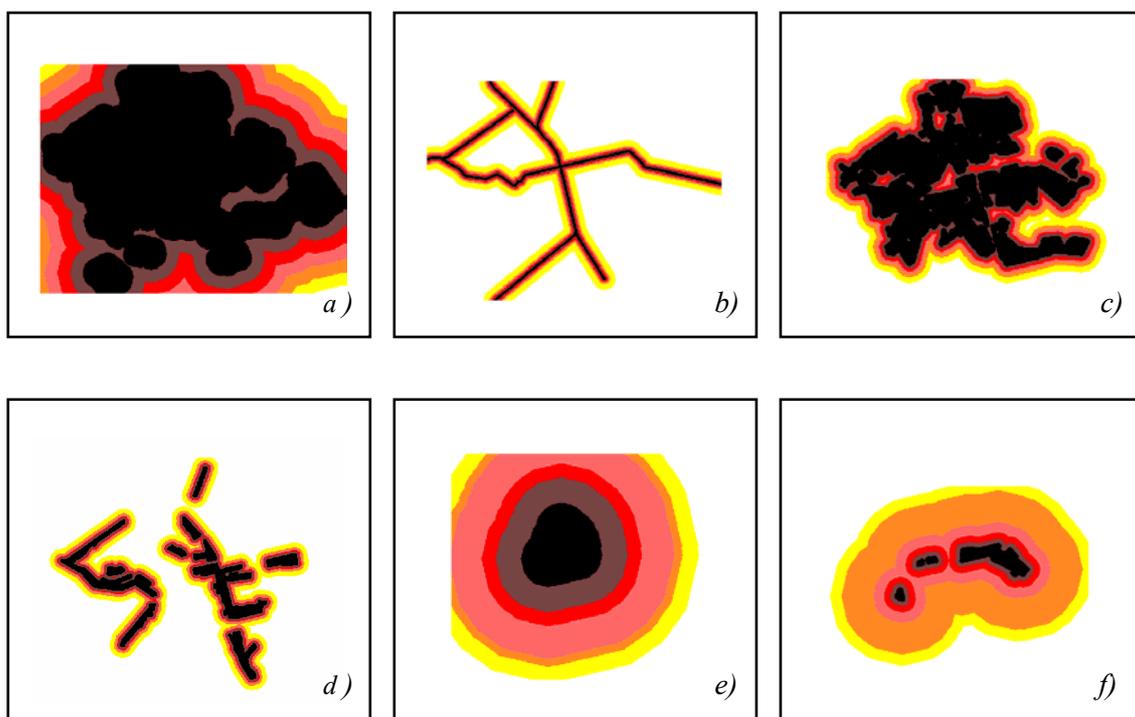


FIGURE 6.24 – Independent variables used to explain the land use transitions in Bauru during the simulation period 1988 – 2000: a) distances to residential zones; b) distances to the services and commercial axes; c) distances to residential areas belonging to the main urban agglomeration ; d) distances to services corridors; e) distances to the central commercial zone; f) distances to industrial zones.

TABLE 6.22 shows the notations utilized for each map of variable employed in this simulation experiment; TABLE 6.23 indicates which variable was selected to explain each of the five existent transitions; TABLE 6.24 presents the values obtained for the Cramer’s Coefficient (V) and the Joint Information Uncertainty (U) for the pairs of

variables used to explain the same type of land use transition; and finally, TABLE 6.25 presents the values for the positive weights of evidence.

TABLE 6.22 – Independent variables defining land use change in Bauru: 1988–2000.

<i>NOTATION</i>	<i>PHYSICAL OR SOCIOECONOMIC LAND USE CHANGE VARIABLE</i>
<i>dist_ind</i>	<i>Distances to industrial zones.</i>
<i>dist_res</i>	<i>Distances to residential zones.</i>
<i>dist_com</i>	<i>Distances to the central commercial zone.</i>
<i>main_res</i>	<i>Distances to residential areas belonging to the main urban agglomeration.</i>
<i>dist_serv</i>	<i>Distances to services corridors.</i>
<i>serv_axes</i>	<i>Distances to the services and commercial axes.</i>
<i>exist_rds</i>	<i>Distances to main existent roads.</i>

TABLE 6.23 – Selection of variables determining land use change in Bauru: 1988–2000.

<i>NOTATION</i>	<i>NU_RES</i>	<i>NU_IND</i>	<i>NU_SERV</i>	<i>RES_SERV</i>	<i>IND_RES</i>
<i>dist_ind</i>		♦			
<i>dist_res</i>	♦				
<i>dist_com</i>	♦	♦	♦		♦
<i>main_res</i>		♦	♦		
<i>dist_serv</i>				♦	
<i>serv_axes</i>		♦	♦	♦	♦
<i>exist_rds</i>	♦				

Likewise the preceding simulation periods, as none of the association measure values surpassed the threshold of 0.50 simultaneously for both indices, no variables preliminarily selected for modeling have been discarded from the analysis.

TABLE 6.24 – Associations between independent variables - Bauru: 1988–2000.

<i>VARIABLE A</i>	<i>VARIABLE B</i>	<i>CRAMER'S STATISTIC (V_{A,B})</i>	<i>UNCERTAINTY (U_{A,B})</i>
<i>dist_ind</i>	<i>serv_axes</i>	0.1324	0.0381
	<i>dist_com</i>	0.3007	0.1730
<i>serv_axes</i>	<i>main_res</i>	0.1442	0.0485
	<i>dist_serv</i>	0.2504	0.1512
<i>dist_res</i>	<i>dist_com</i>	0.2954	0.1920
	<i>exist_rds</i>	0.3733	0.2022
<i>main_res</i>	<i>dist_ind</i>	0.2638	0.1593
	<i>dist_com</i>	0.3554	0.2651
<i>dist_com</i>	<i>serv_axes</i>	0.1426	0.0462
<i>exist_rds</i>	<i>dist_com</i>	0.3072	0.1763

TABLE 6.25 – Values of W^+ for the selected independent variables - Bauru: 1988–2000.

<i>Land Use Transition</i>	<i>Variable</i>	<i>Positive Weights of Evidence W⁺</i>						
		1	2	3	4	5	6	7
<i>NU_RES</i>	<i>dist_com</i> ¹	3.0315	1.8660	1.4883	0.0488	0	0	-1.5975
	<i>dist_res</i> ²	1.0269	-3.0001	0	0	0	0	0
	<i>exist_rds</i> ³	0.2788	0.0470	0.0618	0.5859	0	0	0
<i>NU_IND</i>	<i>dist_com</i> ¹	0	-1.6787	1.9652	0	-1.0938	1.2762	-3.5038
	<i>dist_ind</i> ⁴	3.9186	3.4248	2.0878	-4.0570	0	1.1740	0
	<i>serv_axes</i> ⁵	1.5034	1.6862	1.8472	1.9530	2.1609	1.4820	-4.2138
	<i>main_res</i> ⁶	-0.1651	0.4635	1.1896	1.2176	1.5354	1.1978	0
<i>NU_SERV</i>	<i>dist_com</i> ¹	0	0	2.4844	-0.1735	0	0	0
	<i>serv_axes</i> ⁵	3.6826	2.6842	0	0	0	0	0
	<i>main_res</i> ⁶	1.4228	1.6990	1.7045	1.1503	-1.0949	-2.0000	-2.0000
<i>RES_SERV</i>	<i>serv_axes</i> ⁵	3.4313	2.8969	0	0	0	0	0
	<i>dist_serv</i> ⁷	2.2674	1.5067	1.0902	0	0	0	0
<i>IND_RES</i>	<i>dist_com</i> ¹	0	0	0	0	2.1914	-0.7973	0
	<i>serv_axes</i> ⁵	0.3286	0.5856	0.5630	0.4506	0.2105	-3.4915	0

Note: Distance bands in meters

¹ 1: 0 -5500; 2: 5500-11000; 3: 11000-14000; 4: 14000-19750; 5: 19750-22750; 6: 22750-26250; 7: >26250

² 1: 0 -2700; 2: 2700-5400; 3: 5400-8100; 4: 8100-10800; 5: 10800-13500; 6: 13500-16200; 7: >16200

³ 1: 0 -2500; 2: 2500-5000; 3: 5000-7500; 4: 7500-11150; 5: 11150-15000; 6: 15000-20000; 7: >20000

⁴ 1: 0 -500; 2: 500-1500; 3: 1500-3000; 4: 3000-5500; 5: 5500-13000; 6: 13000-15700; 7: >15700

⁵ 1: 0 -250; 2: 250-500; 3: 500-750; 4: 750-1000; 5: 1000-1250; 6: 1250-2500; 7: > 2500

⁶ 1: 0 -500; 2: 500-1000; 3: 1000-1500; 4: 1500-2000; 5: 2000-2700; 6: 2700-3850; 7: > 3850

⁷ 1: 0 -250; 2: 250-550; 3: 550-750; 4: 750-1000; 5: 1000-1250; 6: 1250-2000; 7: > 2000

The maps of estimated transition probabilities surfaces, generated by DINAMICA upon basis of the values of the positive weights of evidence (W^+), together with their respective land use transition maps are seen in FIGURES 6.25 a and 6.25b.

The arguments employed to explain the transitions “nu_res” and “nu_serv” in the period 1967-1979 (Section 6.1.1) are also valid to justify these same transitions in the current simulation period. Some differences concern the requirements for the transitions “nu_ind” and “res_serv”.

The expansion of the industrial zone (“nu_ind”) in the northeastern sector of Bauru still demands the nearness of such areas to the previously existent industrial use and the availability of road access. But from 1988 to 2000, a new industrial sector was effectively created in the northwestern section of the city, requiring proximity to the labor force supply centers (residential areas) and also a location not too distant from the central commercial zone, since industrial areas depend on commercial activities for logistical support.

The transition “residential to services use (res_serv)” occurs in a moderate rhythm from 1988 to 2000, basically as an extension of already established services corridors. In this sense, this type of land use change will request closeness to previously existent services areas and a strategic proximity to the N-S / E-W services axes of Bauru.

And finally, the transition “industrial to residential use (ind_res)” supposes good accessibility conditions and a location within a reasonable distance from the central commercial zone in face of the dwellers’ need of commuting to work places and shops in central areas.

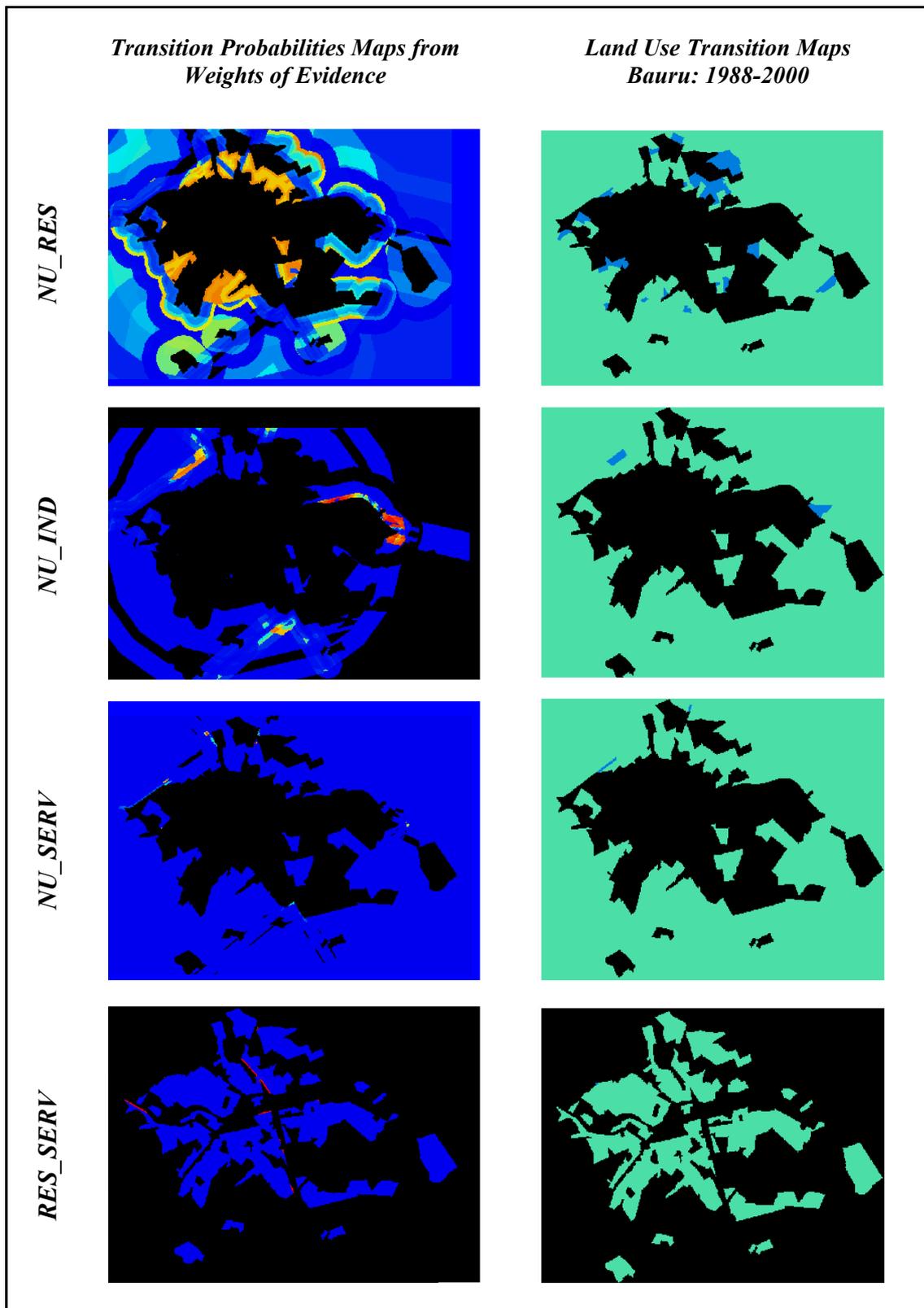


FIGURE 6.25a – Estimated transition probability surfaces and land use change - Bauru.

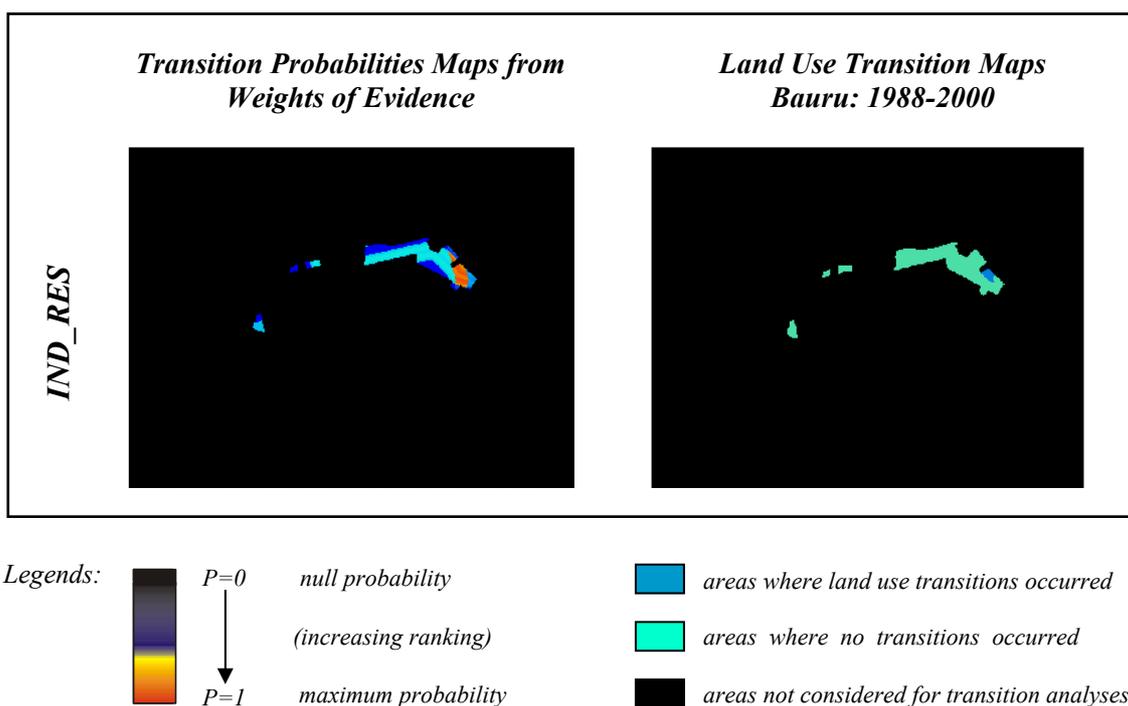


FIGURE 6.25b – Estimated transition probability surfaces and land use change – Bauru.

The three best simulation results produced for the period 1988-2000 are presented in FIGURE 6.26. The internal DINAMICA parameters associated with these optimal simulations are seen in TABLE 6.26, whose statistical validation tests for windows size of 3x3, 5x5 and 9x9, and $k = 0.5$ are listed on TABLE 6.27.

TABLE 6.26 – DINAMICA internal parameters for the simulation of urban land use change in Bauru: 1988–2000.

<i>Land Use Transition</i>	<i>Average Size of Patches</i>	<i>Variance of Patches Size</i>	<i>Proportion of 'Expander'</i>	<i>Proportion of 'Patcher'</i>	<i>Number of Iterations</i>
<i>NU_RES</i>	2000	500	0.60	0.40	300
<i>NU_IND</i>	150	2	0.45	0.55	300
<i>NU_SERV</i>	400	2	1.00	0	300
<i>RES_SERV</i>	40	1	1.00	0	300
<i>IND_RES</i>	100	1	1.00	0	300

TABLE 6.27 – Goodness-of-fit tests for the best land use change simulations produced by the weights of evidence method for Bauru: 1988-2000.

<i>Simulations</i>	<i>Multiple Resolution Goodness-of-Fit (F)</i>
S_1	$F = 0.954193$
S_2	$F = 0.956617$
S_3	$F = 0.956341$

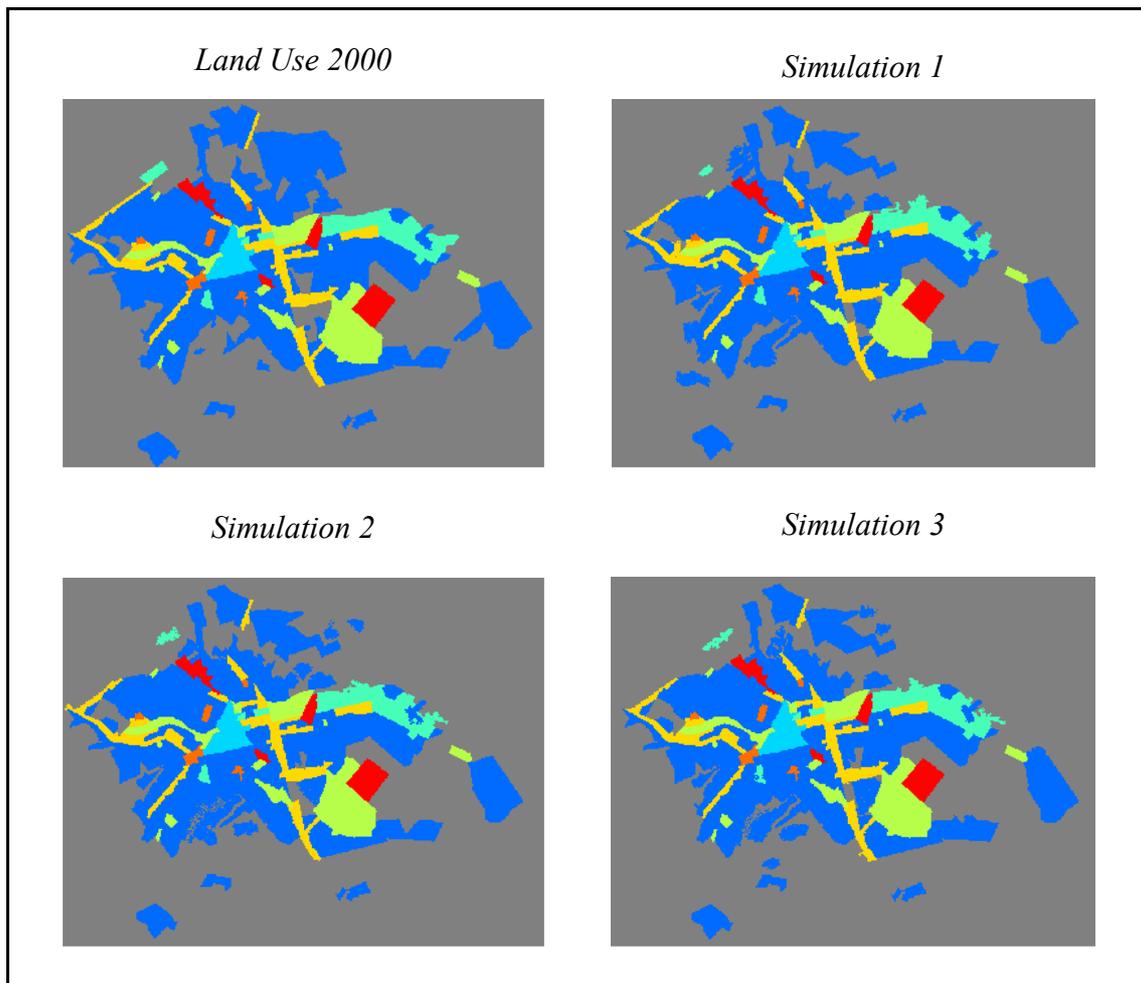


FIGURE 6.26 – The best simulations compared to the actual land use in Bauru in 2000.

6.1.4 Yearly Simulations: 1967 – 2000

According to Equation (5.69), the global matrix of transition can be decomposed in annual transition probabilities by the principal components method. This has been done separately for each simulation period (TABLES 6.28, 6.29 and 6.30). The yearly simulation outputs can be seen in FIGURES 6.27a, 6.27b, 6.27c, 6.27d and 6.27e.

TABLE 6.28 – Matrix of yearly transition probabilities for Bauru: 1967–1979.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Mixed Use</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.9452 \cdot 10^{-1}$	$2.7672 \cdot 10^{-3}$	0	$1.8591 \cdot 10^{-4}$	$1.7121 \cdot 10^{-3}$	$1.2074 \cdot 10^{-5}$	0	$8.0959 \cdot 10^{-4}$
<i>Residential</i>	0	$9.9572 \cdot 10^{-1}$	0	0	0	$4.1413 \cdot 10^{-3}$	$1.3996 \cdot 10^{-4}$	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Mixed Use</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

TABLE 6.29 – Matrix of yearly transition probabilities for Bauru: 1979–1988.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Mixed Use</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.9043 \cdot 10^{-1}$	$8.2863 \cdot 10^{-3}$	0	$1.1001 \cdot 10^{-3}$	0	$1.8758 \cdot 10^{-4}$	0	0
<i>Residential</i>	0	$9.9291 \cdot 10^{-1}$	0	0	0	$6.8296 \cdot 10^{-3}$	$2.5885 \cdot 10^{-4}$	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Mixed Use</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

TABLE 6.30 – Matrix of yearly transition probabilities for Bauru: 1988–2000.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Mixed Use</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.8414 \cdot 10^{-1}$	$2.5872 \cdot 10^{-3}$	0	$3.4581 \cdot 10^{-4}$	0	$7.1684 \cdot 10^{-5}$	0	0
<i>Residential</i>	0	$9.9997 \cdot 10^{-1}$	0	0	0	$2.5003 \cdot 10^{-3}$	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	$3.7187 \cdot 10^{-3}$	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Mixed Use</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

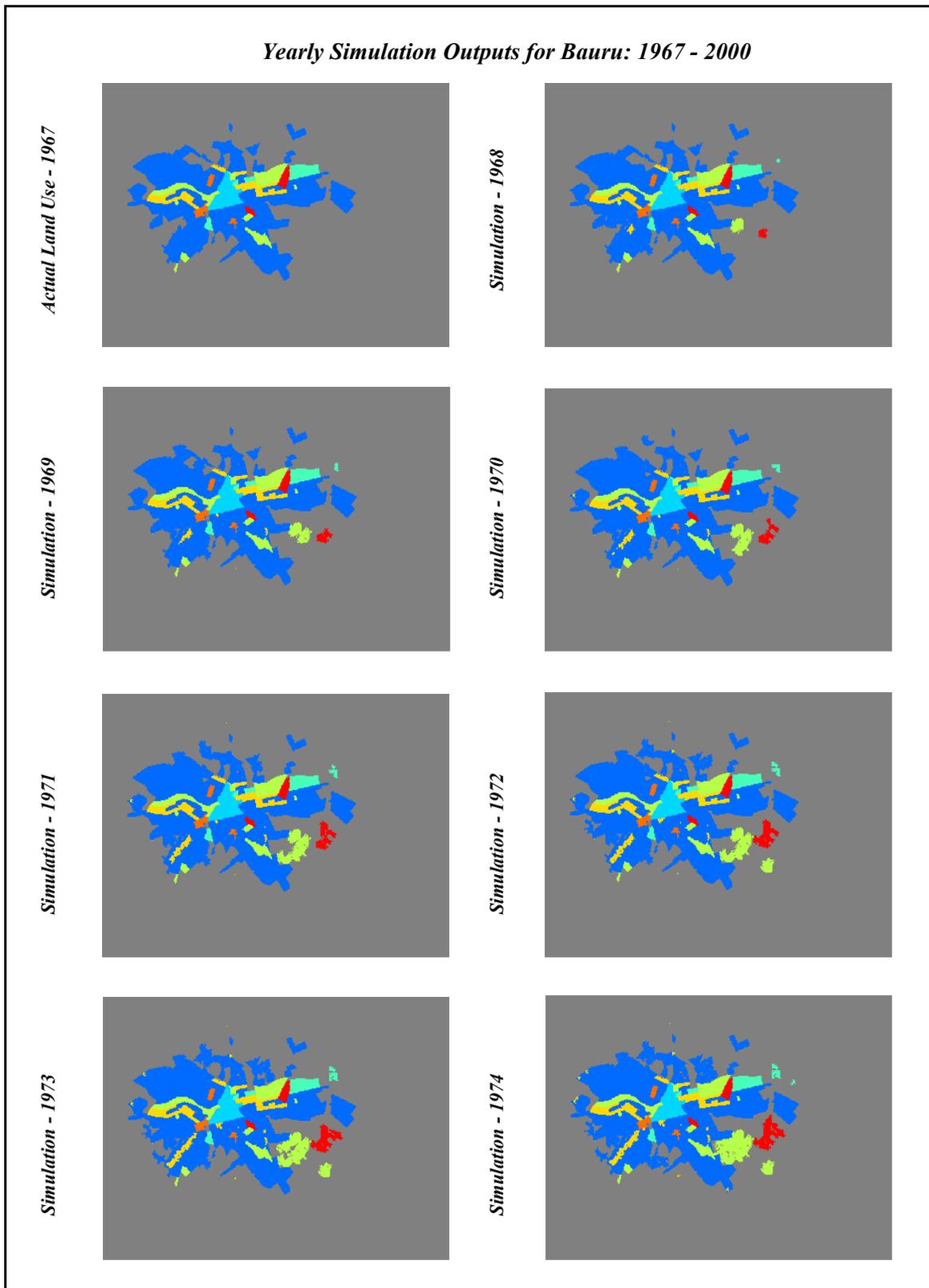


FIGURE 6.27a – Yearly simulation outputs for Bauru: 1967 - 1974.

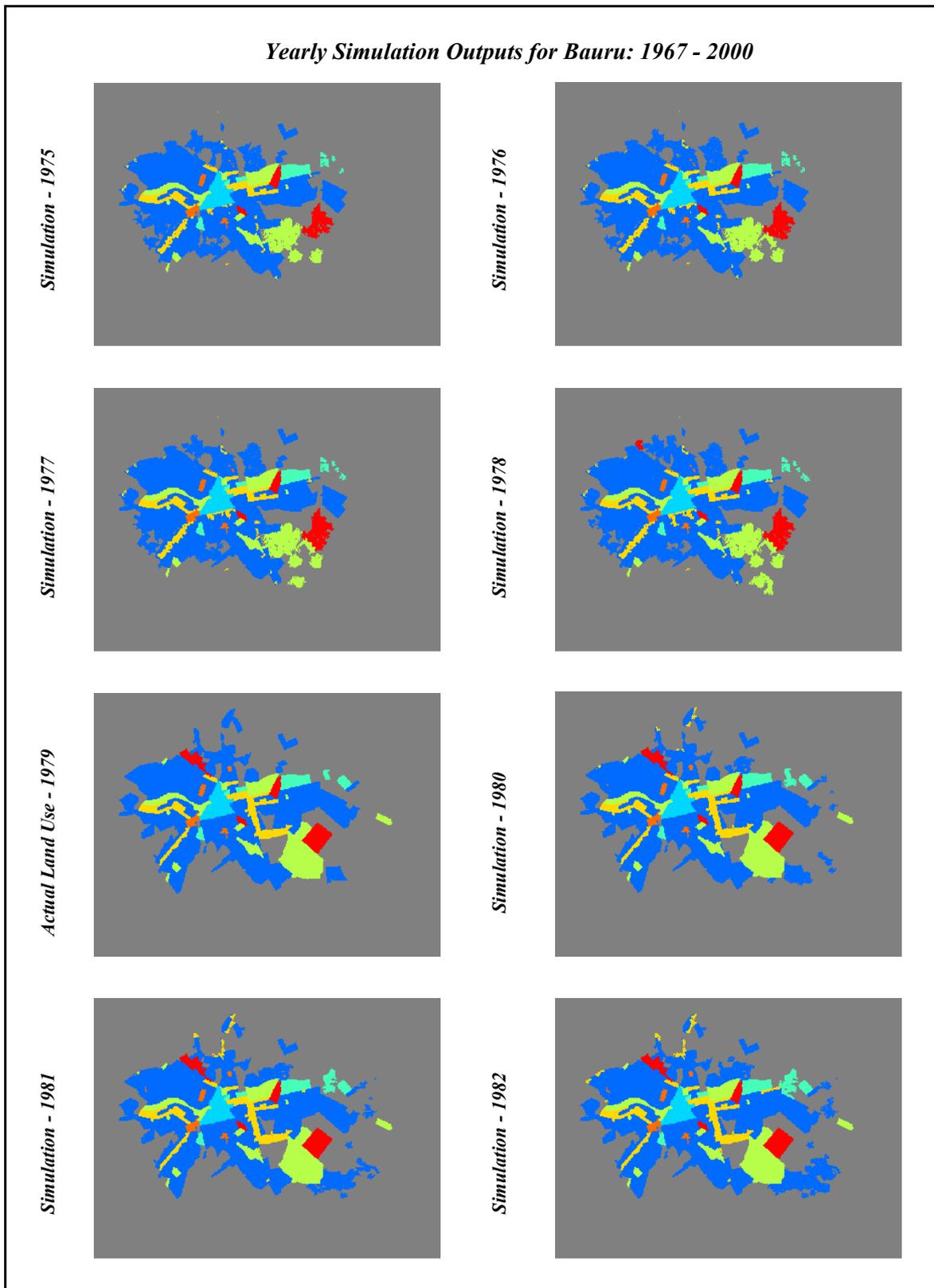


FIGURE 6.27b – Yearly simulation outputs for Bauru: 1975 - 1982.

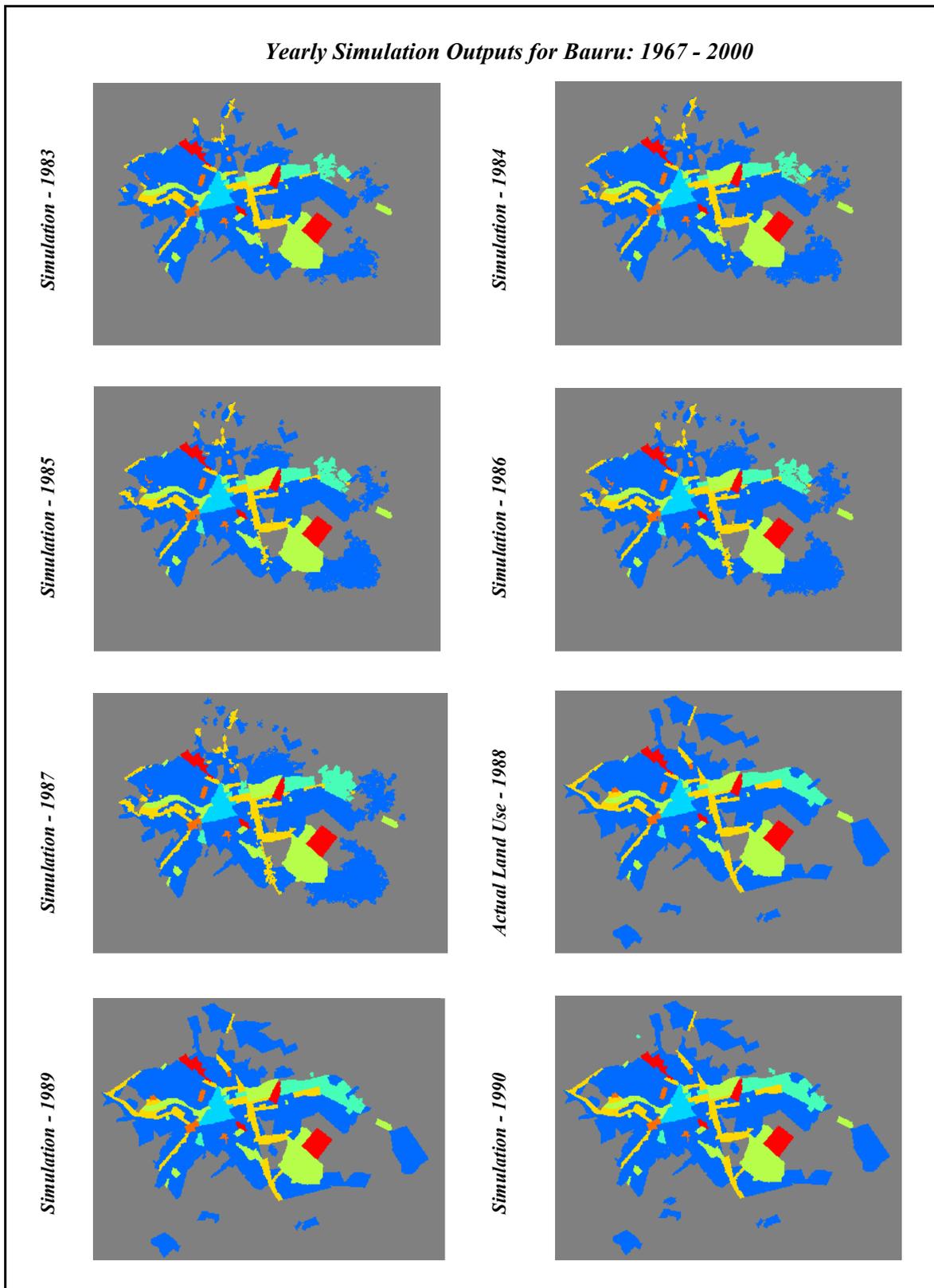


FIGURE 6.27c – Yearly simulation outputs for Bauru: 1983 - 1990.

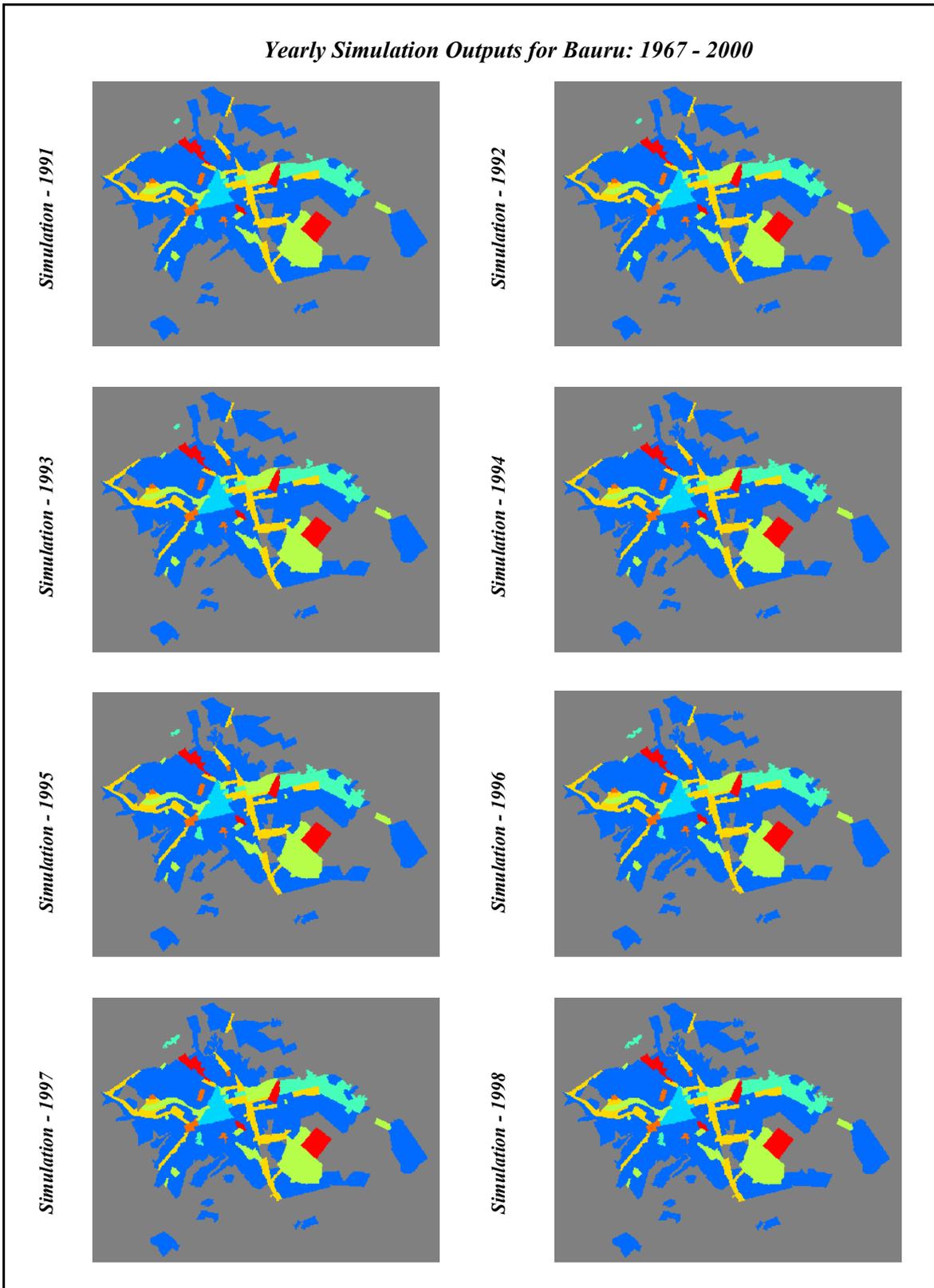


FIGURE 6.27d – Yearly simulation outputs for Bauru: 1991 - 1998.

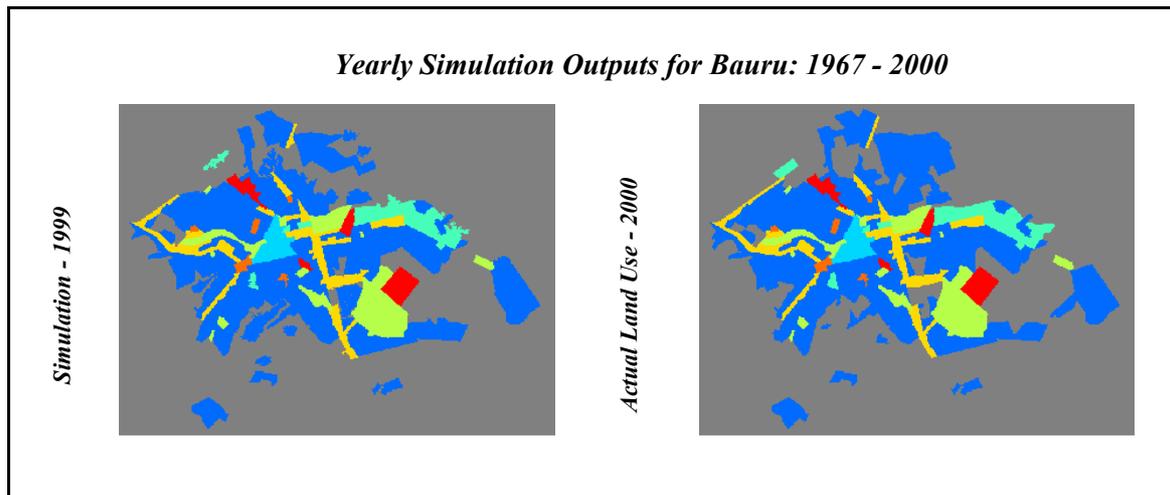


FIGURE 6.27e – Yearly simulation outputs for Bauru: 1999 - 2000.

6.1.5 Short-Term Forecasts: 2000 – 2004

The land use changes considered for forecasts were those observed in the last simulation period (1988-2000), excluding the transition “industrial to residential use (ind_res)”, for it is regarded as unusual taking into account the whole time series. In the next sections, the transition matrices as well as simulation outputs for stationary (Markovian) and non-stationary forecasts of land use change will be presented.

6.1.5.1 Stationary Forecasts

A stationary transition matrix for the year 2004 was obtained using the Markov chain (Equation 5.66) upon basis of the global transition matrix 1988-2000, in which the probability for the transition “ind_res” was purposely set to zero, due to reasons exposed in the previous section. These stationary transition probabilities are shown in TABLE 6.31.

TABLE 6.31 – Matrix of stationary transition probabilities for Bauru: 2000-2004.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Mixed Use</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.9437 \cdot 10^{-1}$	$5.3857 \cdot 10^{-3}$	0	$2.4048 \cdot 10^{-4}$	0	$1.7599 \cdot 10^{-10}$	0	0
<i>Residential</i>	0	$9.9999 \cdot 10^{-1}$	0	0	0	$5.8664 \cdot 10^{-11}$	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Mixed Use</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

6.1.5.2 Non-Stationary Forecasts

Non-stationary forecasts of land use change have been built with the aid of linear regression models relating the area of certain land uses with demographic data and economic indicators, as previously mentioned in Section 5.3.2.6, where the destination uses of land use transitions were regarded as the dependent variable.

For the transitions owning the same destination use, namely “nu_serv” and “res_serv”, only one linear regression model has been built, leading to the estimation of one common transition probability. The shares of transition to services originating from non-urban use and from residential use observed in the simulation period 1988-2000 were assigned to the common transition probability both in the short- and medium-term forecasts. In the particular case of Bauru, the transition “non-urban to services” accounts for 75% of the total transition into services use, and the remaining 25% corresponds to the transition “residential to services”.

The destination land uses had to be estimated for the quinquennia where the data on population and economic performance (see Section 4.2.2) were also available, i.e. for 1970, 1975, 1980, 1985, 1990, 1995 and 2000. This has been enabled by the yearly simulation outputs generated by means of the principal components method (Section 6.1.4). The simulation outputs produced inside DINAMICA were exported in TIFF format, and then imported in IDRISI. Inside IDRISI, the areas of the thematic classes

corresponding to the land uses of interest were assessed. The areas of these destination land uses together with the demographic and economic data used in the linear regression models are seen in TABLE 6.32.

TABLE 6.32 – Areas of destination land uses, urban population, total and sectorial GDPs (US\$): Bauru – 1970-2000.

<i>Years</i>	<i>Dest. Area: Residential (ha)</i>	<i>Dest.Area: Industrial (ha)</i>	<i>Dest. Area: Services (ha)</i>	<i>Urban Population</i>	<i>Total GDP (US\$ 1998)</i>	<i>Rural GDP (US\$ 1998)</i>	<i>Indust. GDP (US\$ 1998)</i>	<i>Comm. GDP (US\$ 1998)</i>	<i>Services GDP (US\$ 1998)</i>
1970	38,904	1,825	3,147	61,592	526,500.428	24,884.128	124,826.722	136,439.859	376,832.414
1975	41,642	2,064	3,984	110,166	718,986.733	14,959.686	233,091.154	155,342.626	470,935.893
1980	45,476	2,531	5,054	159,926	983,887.317	23,486.596	344,767.749	176,902.442	1,572,421.062
1985	53,808	3,865	6,938	215,153	1,222,203.235	62,325.610	428,632.198	201,542.378	731,245.427
1990	59,651	4,761	8,253	237,954	1,358,236.390	46,922.487	472,677.454	479,003.725	419,780.872
1995	62,657	5,057	8,338	279,407	2,256,520.737	11,292.207	705,164.856	891,184.402	299,490.980
2000	65,627	5,341	8,420	310,442	1,906,359.257	6,512.305	627,732.044	713,024.547	467,496.231

SOURCE: Adapted from IPEA (2001, 2003a, 2003b) and FUNDAÇÃO SEADE (2002).

6.1.5.2.1 “Non-Urban to Residential Use (nu_res)” Linear Regression Model

Initially, the tests for verifying the independence of observations (see Section 5.3.2.2) regarding the outcome variable Y_i (“residential use area – destarea”) showed partial acceptance for the autocorrelation function and total acceptance for the partial autocorrelation function (FIGURE 6.28).

The extrapolations in the autocorrelation function are very moderate, for the greater part of the observations lie within the confidence interval. Commonly, very rare events belonging to a time series in reality are totally independent. Even though there are just a few observations for Y_i , it is reasonable to suppose that the residential area observed for Bauru in lags of five years constitute independent observations. In any case, the partial autocorrelation function test provided support for considering the evolution of residential use area throughout the years as independent events.

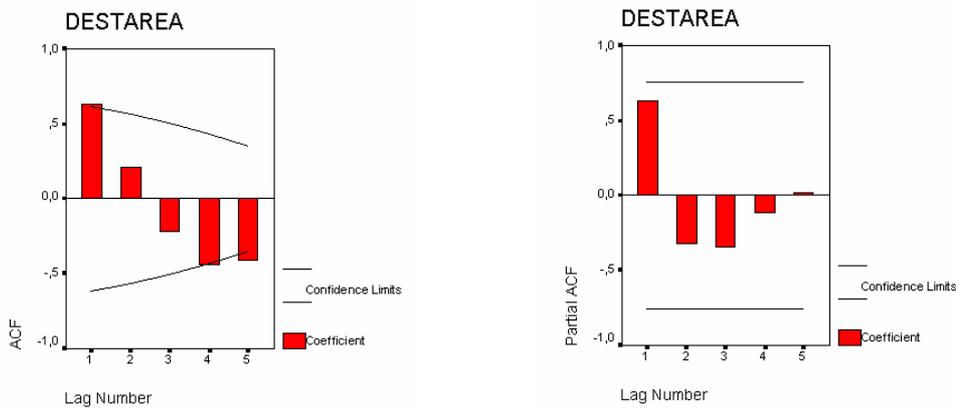


FIGURE 6.28 – ACF and partial ACF tests for the residential use area (destarea).

Still in the exploratory analysis, the independent variables “urban population (urbpop)”, “total GDP (totgdp)” and “industrial GDP (indgdp)” show a high correlation with the response variable “residential use area (destarea)”. Since these three independent variables are highly correlated amongst themselves (TABLE 6.33), only the variable “urban population” remains in the final regression model. The scatter plots concerning the correlation matrix for the “nu_res” model are seen in FIGURE 6.29.

The final equation for this univariate model is the following:

$$Y = \beta_0 + \beta_1 X$$

$$Y = 29,705.243 + 0.116 \cdot urbpop \quad , \quad (6.1)$$

whose R^2 is 0.972, and the p -value of the 95% confidence interval for β_0 and β_1 is 0.000 ($p < 0.05$).

In the analysis of variance (ANOVA), the sum of squares reflects to which extent Y is divergent from \hat{Y} . In this sense, for well fit models it is expected that the regression sum of squares \gggg residual sum of squares, and therefore that the regression sum of squares accounts for the majority of the total sum of squares (TABLE 6.34).

TABLE 6.33 – Correlation matrix for the “nu_res” model: Bauru, 2000-2004.

		DESTAREA	URBPOP	TOTGDP	RURALGDP	INDGDP	COMGDP	SERVGDP
DESTAREA	Pearson Correlation	1	,986**	,931**	-,042	,955**	,879**	-,295
	Sig. (2-tailed)		,	,002	,928	,001	,009	,521
	N	7	7	7	7	7	7	7
URBPOP	Pearson Correlation	,986**	1	,935**	-,063	,970**	,848*	-,165
	Sig. (2-tailed)	,000		,002	,893	,000	,016	,724
	N	7	7	7	7	7	7	7
TOTGDP	Pearson Correlation	,931**	,935**	1	-,254	,989**	,957**	-,271
	Sig. (2-tailed)	,002	,002		,583	,000	,001	,557
	N	7	7	7	7	7	7	7
RURALGDP	Pearson Correlation	-,042	-,063	-,254	1	-,168	-,409	,132
	Sig. (2-tailed)	,928	,893	,583		,718	,363	,778
	N	7	7	7	7	7	7	7
INDGDP	Pearson Correlation	,955**	,970**	,989**	-,168	1	,917**	-,192
	Sig. (2-tailed)	,001	,000	,000	,718		,004	,680
	N	7	7	7	7	7	7	7
COMGDP	Pearson Correlation	,879**	,848*	,957**	-,409	,917**	1	-,438
	Sig. (2-tailed)	,009	,016	,001	,363	,004		,326
	N	7	7	7	7	7	7	7
SERVGDP	Pearson Correlation	-,295	-,165	-,271	,132	-,192	-,438	1
	Sig. (2-tailed)	,521	,724	,557	,778	,680	,326	
	N	7	7	7	7	7	7	7

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

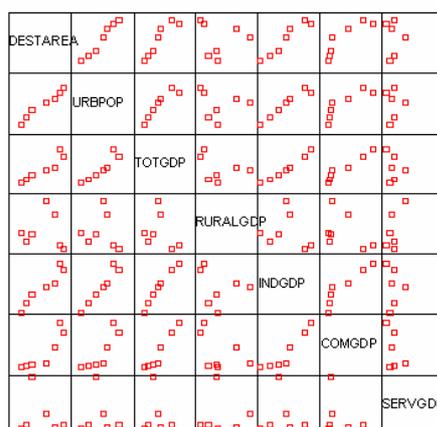


FIGURE 6.29 – Correlation matrix scatter plots for the “nu_res” model: Bauru, 2000-2004.

TABLE 6.34 – Analysis of variance for the “nu_res” model: Bauru, 2000-2004.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6,61E+08	1	661245024,7	172,565	,000 ^a
	Residual	19159342	5	3831868,441		
	Total	6,80E+08	6			

a. Predictors: (Constant), URBPOP

b. Dependent Variable: DESTAREA

Concerning the analysis of residuals, when the model is satisfactorily fit, the standardized residuals (see Section 5.3.2.5) and the studentized residuals (adjusted to a t-distribution) must lie within the interval $[-2,+2]$. For the current model, these values are $-1.442; +1.164$ and $-1.583; +1.493$ respectively for the standardized and studentized residuals. Plots of the standardized residuals versus the independent variable (“urbpop”) and versus the adjusted predicted value \hat{Y}_i are seen in FIGURE 6.30. As stated in Section 5.3.2.5, these plots should not present any kind of correlation pattern.

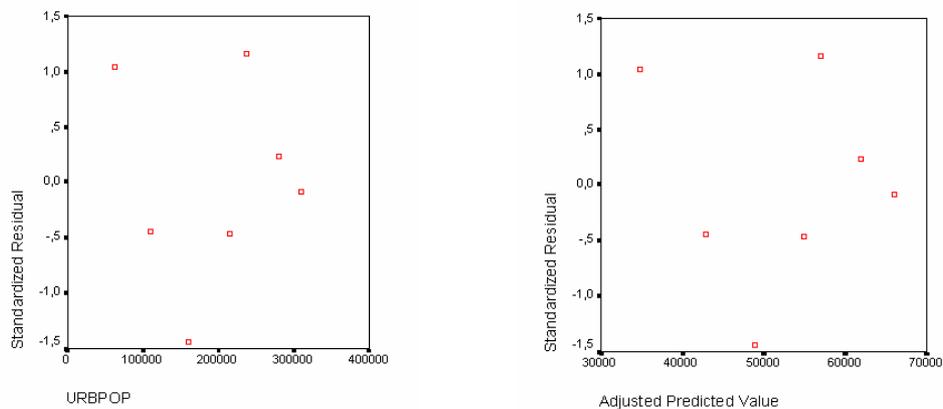


FIGURE 6.30 – Analysis of residuals for the “nu_res” model: Bauru, 2000-2004.

6.1.5.2.2 “Non-Urban to Industrial Use (nu_ind)” Linear Regression Model

Likewise the “nu_res” model, the tests for verifying the independence of observations regarding the outcome variable Y_i (“industrial use area – destarea”) showed partial acceptance for the autocorrelation function and total acceptance for the partial autocorrelation function (FIGURE 6.31). The observations of industrial use area throughout time were deemed as independent events for the same reasons exposed in Section 6.1.5.2.1.

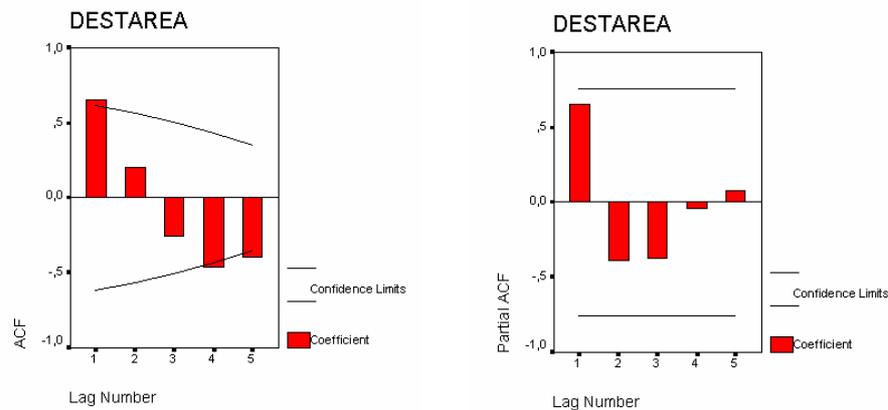


FIGURE 6.31 – ACF and partial ACF tests for the industrial use area (destarea).

The independent variables “urban population (urbpop)”, “total GDP (totgdp)” and “industrial GDP (indgdp)” show a high correlation with the response variable “industrial use area (destarea)”. Since these three independent variables are highly correlated amongst themselves (TABLE 6.35), only the variable “industrial GDP” remains in the final regression model. Although its correlation with the response variable is not the highest, it was judged to be the one which best explains the increases in the industrial area in the city in the latest three decades. The scatter plots concerning the correlation matrix for the “nu_ind” model are presented in FIGURE 6.32.

The final equation for this univariate model is the following:

$$Y = \beta_1 X$$

$$Y = 0.008343 \cdot indgdp \quad , \quad (6.2)$$

whose R^2 is 0.979, and the p -value of the 95% confidence interval for β_1 is 0.000 ($p < 0.05$). Since β_0 did not pass the significance test, it was removed from the model.

In the analysis of variance (ANOVA), the sum of squares reflects to which extent Y is divergent from \hat{Y} . In this sense, for well fit models it is expected that the regression sum of squares \gggg residual sum of squares, and therefore that the regression sum of squares accounts for the majority of the total sum of squares (TABLE 6.36).

TABLE 6.35 – Correlation matrix for the “nu_ind” model: Bauru, 2000-2004.

		DESTAREA	URBPOP	TOTGDP	RURALGDP	INDGDP	COMGDP	SERVGDP
DESTAREA	Pearson Correlation	1	,974**	,921**	,000	,943**	,873*	-,329
	Sig. (2-tailed)		,000	,003	1,000	,001	,010	,471
	N	7	7	7	7	7	7	7
URBPOP	Pearson Correlation	,974**	1	,935**	-,063	,970**	,848*	-,165
	Sig. (2-tailed)	,000		,002	,893	,000	,016	,724
	N	7	7	7	7	7	7	7
TOTGDP	Pearson Correlation	,921**	,935**	1	-,254	,989**	,957**	-,271
	Sig. (2-tailed)	,003	,002		,583	,000	,001	,557
	N	7	7	7	7	7	7	7
RURALGDP	Pearson Correlation	,000	-,063	-,254	1	-,168	-,409	,132
	Sig. (2-tailed)	1,000	,893	,583		,718	,363	,778
	N	7	7	7	7	7	7	7
INDGDP	Pearson Correlation	,943**	,970**	,989**	-,168	1	,917**	-,192
	Sig. (2-tailed)	,001	,000	,000	,718		,004	,680
	N	7	7	7	7	7	7	7
COMGDP	Pearson Correlation	,873*	,848*	,957**	-,409	,917**	1	-,438
	Sig. (2-tailed)	,010	,016	,001	,363	,004		,326
	N	7	7	7	7	7	7	7
SERVGDP	Pearson Correlation	-,329	-,165	-,271	,132	-,192	-,438	1
	Sig. (2-tailed)	,471	,724	,557	,778	,680	,326	
	N	7	7	7	7	7	7	7

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

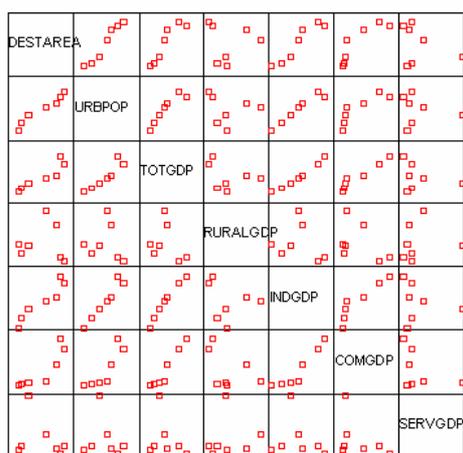


FIGURE 6.32 – Correlation matrix scatter plots for the “nu_ind” model: Bauru, 2000-2004.

TABLE 6.36 – Analysis of variance for the “nu_ind” model: Bauru, 2000-2004.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11764114	1	11764113,87	40,506	,001 ^a
	Residual	1452139	5	290427,798		
	Total	13216253	6			

a. Predictors: (Constant), INDGDP

b. Dependent Variable: DESTAREA

Regarding the analysis of residuals, when the model is satisfactorily fit, the standardized residuals (see Section 5.3.2.5) and the studentized residuals (adjusted to a t-distribution) must lie within the interval $[-2,+2]$. For the current model, these values are -1.366 ; $+1.353$ and -1.675 ; $+1.467$ respectively for the standardized and studentized residuals. Plots of the standardized residuals versus the independent variable (“indgdp”) and versus the adjusted predicted value \hat{Y}_i are seen in FIGURE 6.33. According to what was stated in Section 5.3.2.5, these plots should not contain any kind of correlation pattern.

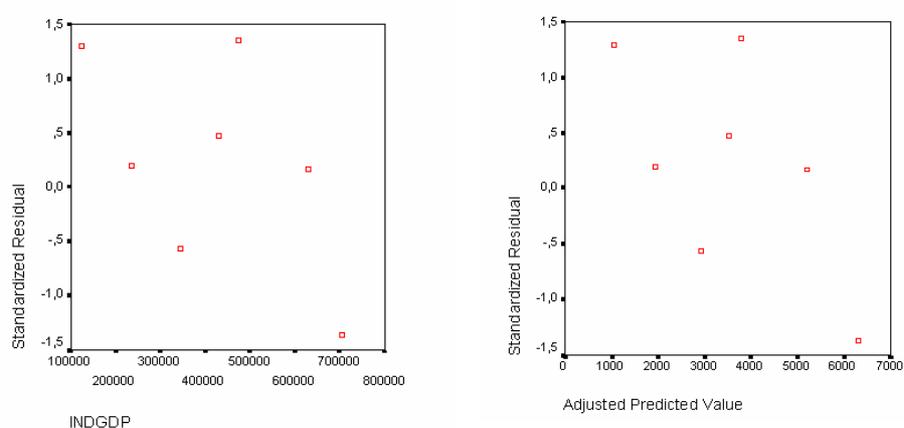


FIGURE 6.33 – Analysis of residuals for the “nu_ind” model: Bauru, 2000-2004.

6.1.5.2.3 “Non-Urban/Residential to Services Use (nu/res_serv)” Linear Regression Model

Likewise the “nu_res” and “nu_ind” models, the tests for verifying the independence of observations regarding the outcome variable Y_i (“services use area – destarea”) showed partial acceptance for the autocorrelation function and total acceptance for the partial autocorrelation function (FIGURE 6.34). The observations of services use area throughout the years were judged as independent events for the same reasons exposed in Section 6.1.5.2.1.

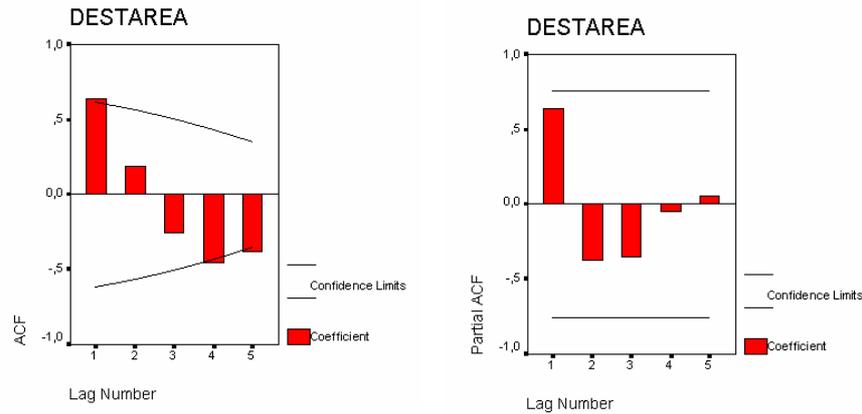


FIGURE 6.34 – ACF and partial ACF tests for the services use area (destarea).

In the same way as the preceding linear models, the independent variables “urban population (urbpop)”, “total GDP (totgdp)” and “industrial GDP (indgdp)” show a high correlation with the response variable “services use area (destarea)”. Since these three independent variables are highly correlated amongst themselves (TABLE 6.37), only the variable “urban population” remains in the final regression model. The scatter plots regarding the correlation matrix for the “nu/res_serv” model are presented in FIGURE 6.35.

The final equation for this univariate model is the following:

$$Y = \beta_0 + \beta_1 X$$

$$Y = 1,590.457 + 0.02401 \cdot urbpop \quad , \quad (6.3)$$

whose R^2 is 0.948, and the p -value of the 95% confidence interval for β_0 is 0.031 and for β_1 is 0.000 ($p < 0.05$).

In the analysis of variance (ANOVA), the sum of squares reflects to which extent Y is divergent from \hat{Y} . In this sense, for well fit models it is expected that the regression sum of squares \gggg residual sum of squares, and therefore that the regression sum of squares accounts for the majority of the total sum of squares (TABLE 6.38).

TABLE 6.37 – Correlation matrix for the “nu/res_serv” model: Bauru, 2000-2004.

		DESTAREA	URBPOP	TOTGDP	RURALGDP	INDGDP	COMGDP	SERVGDP
DESTAREA	Pearson Correlation	1	,974**	,899**	,092	,940**	,823*	-,233
	Sig. (2-tailed)		,000	,006	,844	,002	,023	,615
	N	7	7	7	7	7	7	7
URBPOP	Pearson Correlation	,974**	1	,935**	-,063	,970**	,848*	-,165
	Sig. (2-tailed)	,000		,002	,893	,000	,016	,724
	N	7	7	7	7	7	7	7
TOTGDP	Pearson Correlation	,899**	,935**	1	-,254	,989**	,957**	-,271
	Sig. (2-tailed)	,006	,002		,583	,000	,001	,557
	N	7	7	7	7	7	7	7
RURALGDP	Pearson Correlation	,092	-,063	-,254	1	-,168	-,409	,132
	Sig. (2-tailed)	,844	,893	,583		,718	,363	,778
	N	7	7	7	7	7	7	7
INDGDP	Pearson Correlation	,940**	,970**	,989**	-,168	1	,917**	-,192
	Sig. (2-tailed)	,002	,000	,000	,718		,004	,680
	N	7	7	7	7	7	7	7
COMGDP	Pearson Correlation	,823*	,848*	,957**	-,409	,917**	1	-,438
	Sig. (2-tailed)	,023	,016	,001	,363	,004		,326
	N	7	7	7	7	7	7	7
SERVGDP	Pearson Correlation	-,233	-,165	-,271	,132	-,192	-,438	1
	Sig. (2-tailed)	,615	,724	,557	,778	,680	,326	
	N	7	7	7	7	7	7	7

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

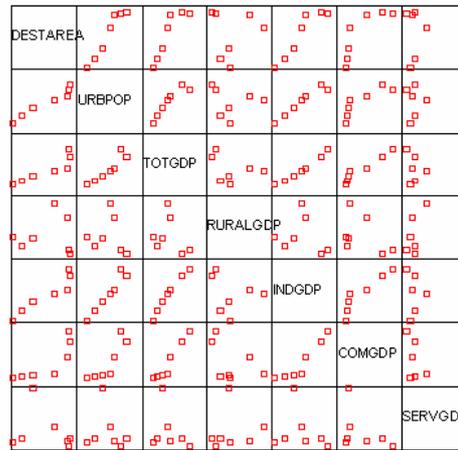


FIGURE 6.35 – Correlation matrix scatter plots for the “nu/res_serv” model: Bauru, 2000-2004.

TABLE 6.38 – Analysis of variance for the “nu/res_serv” model: Bauru, 2000-2004.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	28190577	1	28190577,19	91,758	,000 ^a
	Residual	1536136	5	307227,134		
	Total	29726713	6			

a. Predictors: (Constant), URBPOP

b. Dependent Variable: DESTAREA

As to the analysis of residuals, when the model is satisfactorily fit, the standardized residuals (see Section 5.3.2.5) and the studentized residuals (adjusted to a t-distribution) must lie within the interval $[-2,+2]$. For the current model, these values are -1.124 ; $+1.714$ and -1.462 ; $+1.891$ respectively for the standardized and studentized residuals. Plots of the standardized residuals versus the independent variable (“urbpop”) and versus the adjusted predicted value \hat{Y}_i are seen in FIGURE 6.36. According to what was stated in Section 5.3.2.5, these plots should not contain any kind of correlation pattern.

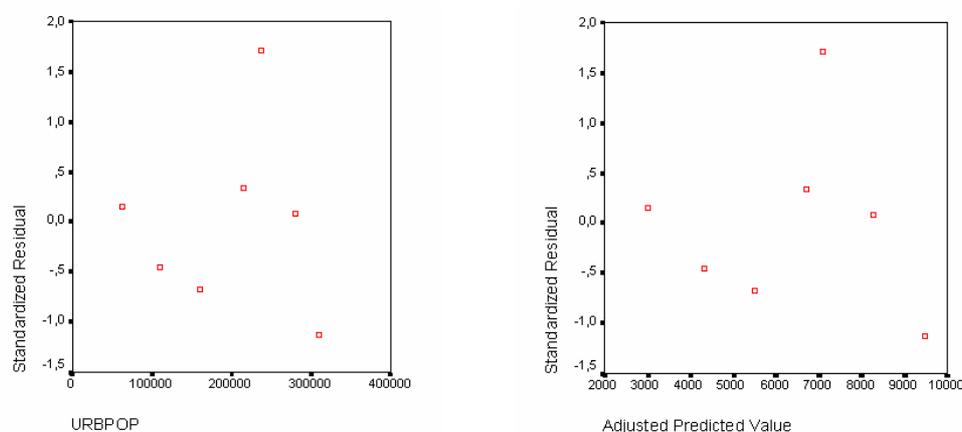


FIGURE 6.36 – Analysis of residuals for the “nu/res_serv” model: Bauru, 2000-2004.

Heuristic projections were made for the values of the independent variables upon basis of their past trends. According to what was exposed in Section 5.6, time series analysis should have been employed for this end. As time series analysis requires a minimum of thirty observations, and as the yearly total and sectorial municipal GDPs of Bauru in the latest three decades were not available by the time this research was finished, this statistical technique could have not been adopted.

Two types of non-stationary forecast scenarios of land use change were formulated: an optimistic and a pessimist one, with respectively slight over and underestimations of the independent variable, as already defined in Section 5.3.3. In the first case, the value of X concerning urban population was 346,934 inhabitants for the “nu_res” and “nu/res_serv” models; and US\$ 676,495.266 regarding industrial GDP for the “nu_ind”

model. In the pessimist scenario, X assumed the values of 325,369 inhabitants for the “nu_res” and “nu/res_serv” models; and US\$ 650,964.881 concerning industrial GDP for the “nu_ind” model. Using these values of X , transition probabilities were calculated for both scenarios in the short-term (TABLES 6.39 and 6.40), employing the conversion equations presented in Section 5.3.2.6.

TABLE 6.39 – Matrix of optimistic transition probabilities for Bauru: 2000-2004.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Mixed Use</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.8830 \cdot 10^{-1}$	$1.0726 \cdot 10^{-2}$	0	$9.7098 \cdot 10^{-4}$	0	$3.5050 \cdot 10^{-10}$	0	0
<i>Residential</i>	0	$9.9999 \cdot 10^{-1}$	0	0	0	$1.1683 \cdot 10^{-10}$	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Mixed Use</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

TABLE 6.40 – Matrix of pessimist transition probabilities for Bauru: 2000-2004.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Mixed Use</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.9762 \cdot 10^{-1}$	$2.3729 \cdot 10^{-3}$	0	$4.5373 \cdot 10^{-6}$	0	$7.7541 \cdot 10^{-11}$	0	0
<i>Residential</i>	0	$9.9999 \cdot 10^{-1}$	0	0	0	$2.5847 \cdot 10^{-11}$	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Mixed Use</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

6.1.5.3 Forecasts Simulation Outputs

Taking into account the same sets of independent variables selected for the last simulation period (1988-2000) and their respective weights of evidence (TABLE 6.25), stationary, optimistic and pessimist simulation outputs were generated for Bauru in the short-term (2004), using the TABLES 6.31, 6.39 and 6.40 to respectively set the total amount of land use change. These simulation outputs are presented in FIGURE 6.37.

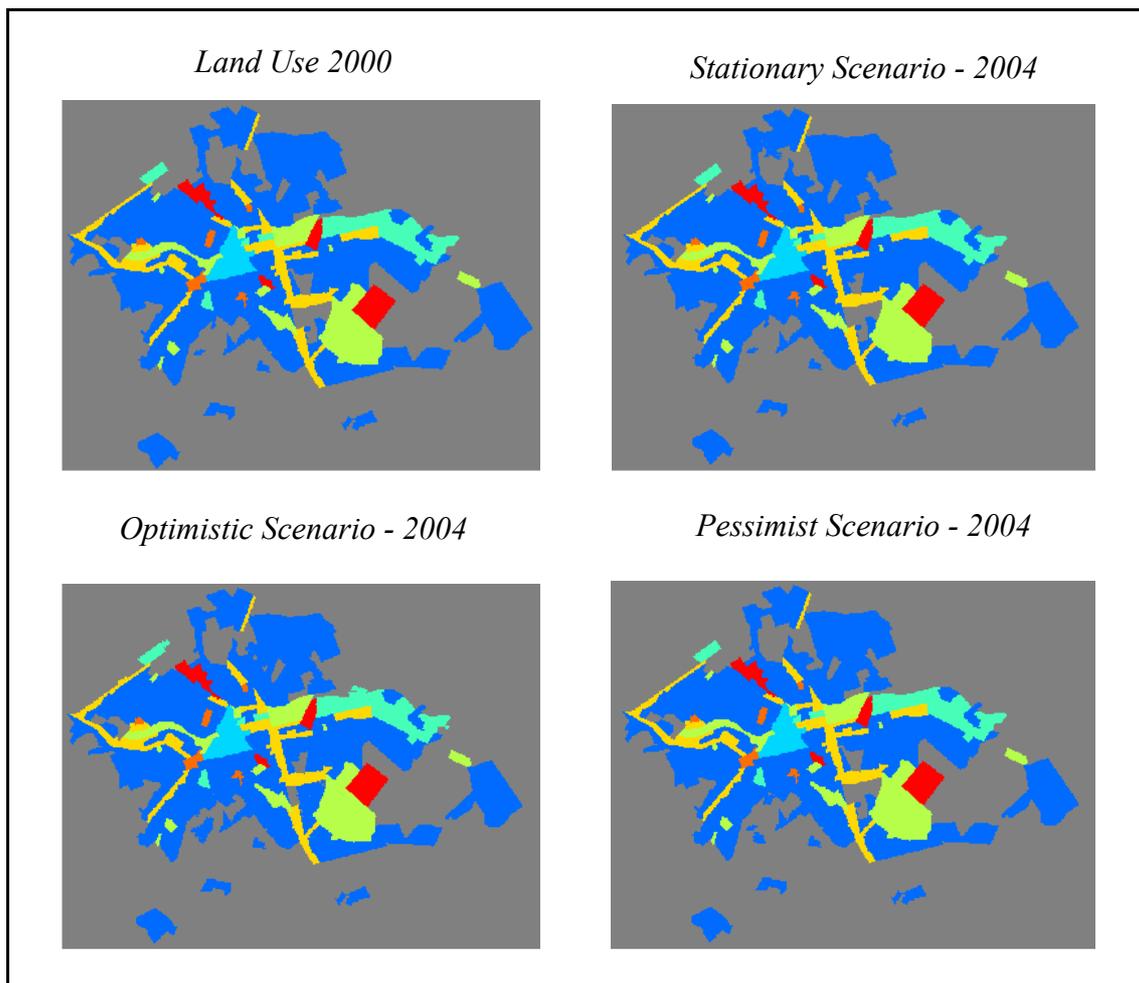


FIGURE 6.37 – Stationary, optimistic and pessimist simulations for 2004 compared to the actual land use in Bauru in 2000.

It is observable from the above forecast simulations that the expansion of the services areas is meaningless. This does not imply that there will be no economic growth in the services sector in the near term, but that the areas sheltering services will undergo densification processes instead.

The main areas where new residential settlements arise are located in the northern and southern sectors of Bauru, close to the N-S services and commercial axes. This

expansion pattern complies with trends observed in a recent official city map of Bauru, issued in 2003 (DAE, 2003).

As previously explained in Section 5.3.3, only mild variations in the projections of population and economic data were introduced. This is due to the fact that, in view of the administrative continuity trend demonstrated by the current federal government as well as of steadfastly decreasing population growth rates, the demographic and macroeconomic scenarios of Brazil in the latest years are expected to reproduce themselves in the current decade.

6.1.6 Medium-Term Forecasts: 2000 – 2007

6.1.6.1 Stationary Forecasts

A stationary transition matrix for the year 2007 was obtained using the Markov chain (Equation 5.66) upon basis of the global transition matrix 1988-2000, in which the transition “ind_res” was purposely set to zero. These stationary transition probabilities are shown in TABLE 6.41.

TABLE 6.41 – Matrix of stationary transition probabilities for Bauru: 2000-2007.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Mixed Use</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.9034 \cdot 10^{-1}$	$9.2448 \cdot 10^{-3}$	0	$4.1278 \cdot 10^{-4}$	0	$3.0210 \cdot 10^{-10}$	0	0
<i>Residential</i>	0	$9.9999 \cdot 10^{-1}$	0	0	0	$1.0070 \cdot 10^{-10}$	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Mixed Use</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

6.1.6.2 Non-Stationary Forecasts

The same linear regression models used for short-term predictions have been used in the medium-term forecasts. For the optimistic scenario, the value of X concerning urban population was 350,119 inhabitants for the “nu_res” and “nu/res_serv” models; and US\$ 694,833.993 regarding industrial GDP for the “nu_ind” model. In the pessimist

scenario, X assumed the values of 322,877 inhabitants for the “nu_res” and “nu/res_serv” models; and US\$ 651,084.742 concerning industrial GDP for the “nu_ind” model. Using these values of X , transition probabilities were calculated for both scenarios in the medium-term (TABLES 6.42 and 6.43), employing the conversion equations presented in Section 5.3.2.6.

TABLE 6.42 – Matrix of optimistic transition probabilities for Bauru: 2000-2007.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Mixed Use</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.7992*10^{-1}$	$1.8412*10^{-2}$	0	$1.6667*10^{-3}$	0	$6.0165*10^{-10}$	0	0
<i>Residential</i>	0	$9.9999*10^{-1}$	0	0	0	$2.0055*10^{-10}$	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Mixed Use</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

TABLE 6.43 – Matrix of pessimist transition probabilities for Bauru: 2000-2007.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Mixed Use</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.9592*10^{-1}$	$4.0732*10^{-3}$	0	$7.7884*10^{-6}$	0	$1.3310*10^{-10}$	0	0
<i>Residential</i>	0	$9.9999*10^{-1}$	0	0	0	$4.4367*10^{-11}$	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Mixed Use</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

6.1.6.3 Forecasts Simulation Outputs

Taking into account the same sets of independent variables selected for the last simulation period (1988-2000) and their respective weights of evidence (TABLE 6.25), stationary, optimistic and pessimist simulation outputs were generated for Bauru in the medium-term (2007), using the TABLES 6.41, 6.42 and 6.43 to respectively set the total amount of land use change. These simulation outputs are presented in FIGURE 6.38.

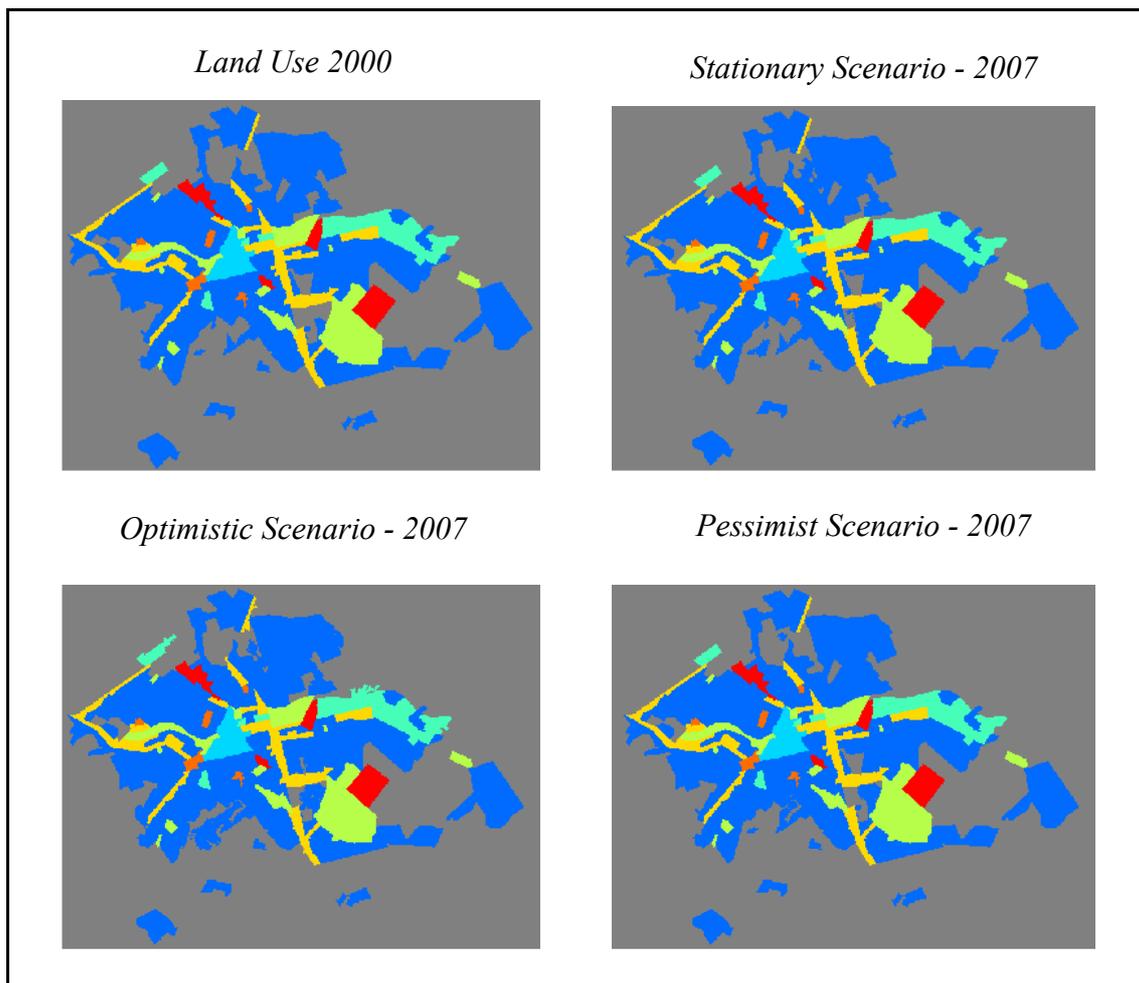


FIGURE 6.38 – Stationary, optimistic and pessimist simulations for 2007 compared to the actual land use in Bauru in 2000.

The medium-term forecasts simulations represent an enhancement of the residential expansion patterns observed in the official city map of 2003 (DAE, 2003), in which new residential areas are located in the northern and southern sectors of Bauru, close to the N-S services and commercial axes.

6.2 Piracicaba

As previously mentioned in Section 3.4.2.2, Piracicaba has distinguished itself as an important sugar cane producer pole in São Paulo State, and gathers huge sugar cane and alcohol agribusiness complexes. The city also shelters a very diversified industrial park, with big-sized and high technology plants.

The urban framework of Piracicaba is comparatively less fragmented than the one of Bauru. The city lays a great emphasis on the river, which is a landmark for the social life of its inhabitants. The main central commercial zone borders the green areas established along the Piracicaba river margins, as do the huge areas dedicated for universities and colleges campi in the northern and northeastern portions of the city.

In contrast to the concentrated pattern of industrial development of Bauru, industrial districts in Piracicaba are scattered throughout the urban area. Services axes, on their turn, show a more or less radio-concentric pattern, departing from the main central commercial zone. As the city expands across both margins of the Piracicaba river, commercial nuclei are found on both sides of the river.

In the next sections, the results for the land use change simulations throughout the time series ranging from 1962 to 1999 will be presented, followed by the forecasts simulations generated for the short- and medium-term, respectively 2004 and 2007.

6.2.1 Simulation Period: 1962 - 1985

There are no official data about the population of Piracicaba in 1962. As already stated in Section 6.1.1, the first demographic census in Brazil was held in 1970 (IBGE, 1971), and the total population of this municipality at that time was 76,439 inhabitants, out of which 73,153 inhabitants lived in urban areas. In the population estimates carried out in the year 1985 (IBGE, 1987), the total population of Piracicaba rose to 252,079 people, from which 198,407 were urban inhabitants.

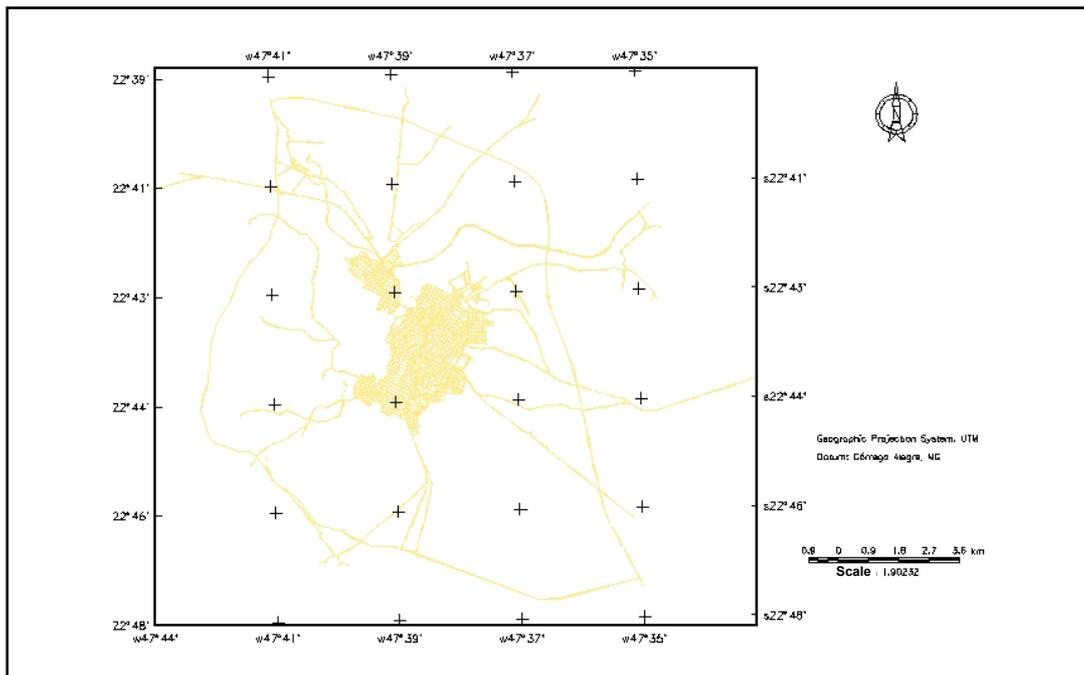


FIGURE 6.39 – Piracicaba official city map in 1962 (reconstitution map).
 SOURCE: LSN-ESALQ-USP (2003) and SEMUPLAN (1985).

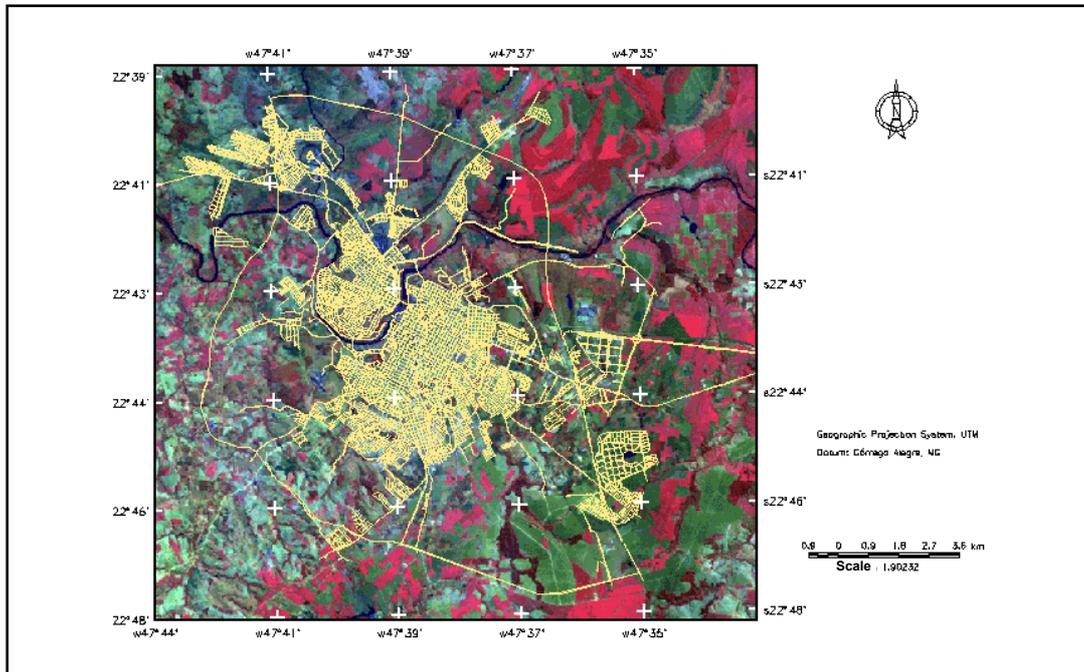


FIGURE 6.40 – Piracicaba TM – 5 image and official city map in 1985.
 SOURCE: INPE (1985) and SEMUPLAN (1985).

The impacts of this population growth on the urban area extension can be seen in FIGURES 6.39 and 6.40, which present the city maps for the initial and final time of simulation.

The initial and final land use maps used in the simulation period 1962 – 1985 (FIGURE 6.41) were elaborated upon basis of the two city maps previously shown, of generalization procedures applied to the zoning legislation issued in 1964 (Piracicaba, 1964) and to the original land use map of 1985 (SEMUPLAN, 1985) as well as upon basis of the digital satellite image of Piracicaba in 1985 (INPE, 1985).

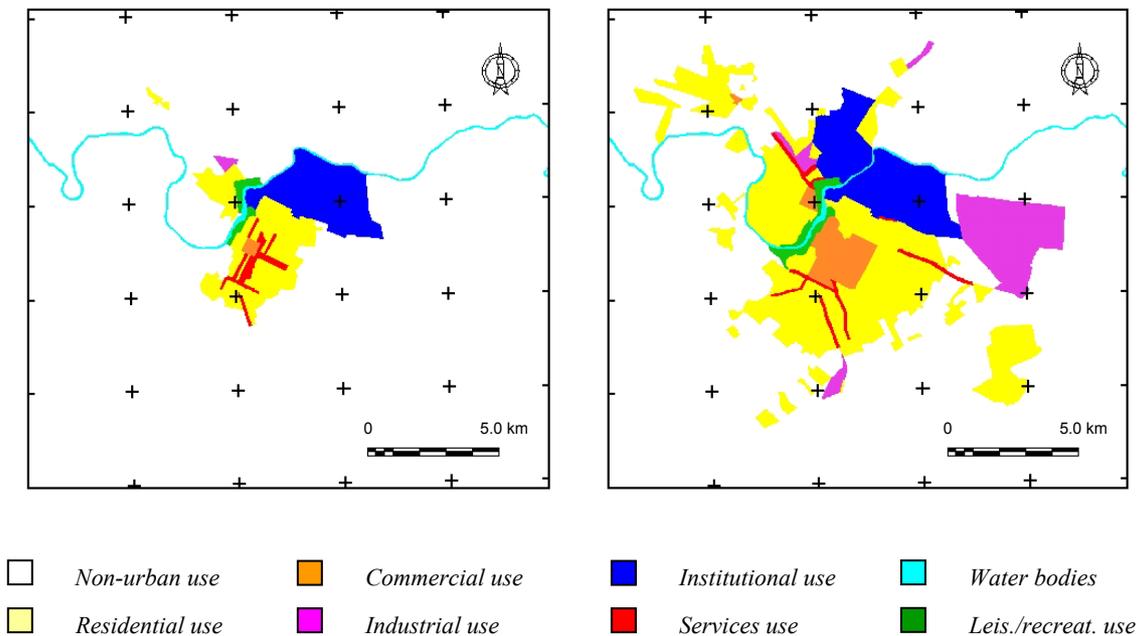


FIGURE 6.41 – Generalized land use map in Piracicaba in 1962 (left) and 1985 (right).

A cross-tabulation operation was made between both land use maps (FIGURE 6.42) so as to generate transition percentages for the existent types of land use change (TABLES 6.44 and 6.45).

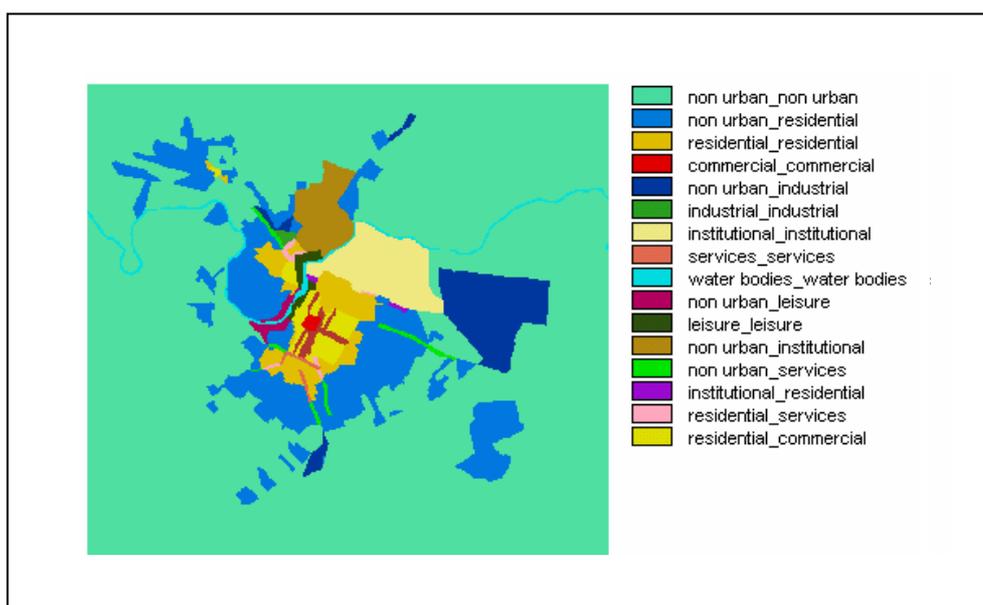


FIGURE 6.42 – Cross-tabulation map between Piracicaba land use maps of 1962 and 1985.

TABLE 6.44 – Existent land use transitions in Piracicaba: 1962–1985.

<i>NOTATION</i>	<i>LAND USE TRANSITION</i>
<i>NU_RES</i>	<i>Non-urban to residential</i>
<i>NU_IND</i>	<i>Non-urban to industrial</i>
<i>NU_INST</i>	<i>Non-urban to institutional</i>
<i>NU_SERV</i>	<i>Non-urban to services</i>
<i>NU_LEIS</i>	<i>Non-urban to leis./recreation</i>
<i>RES_COM</i>	<i>Residential to commercial</i>
<i>RES_SERV</i>	<i>Residential to services</i>
<i>SERV_COM</i>	<i>Services to commercial</i>
<i>INST_RES</i>	<i>Institutional to residential</i>

TABLE 6.45 – Matrix of global transition probabilities for Piracicaba: 1962–1985.

<i>Land Use</i>	<i>Non-urban</i>	<i>Resid.</i>	<i>Comm.</i>	<i>Industr.</i>	<i>Instit.</i>	<i>Services</i>	<i>Water B.</i>	<i>Leis./Recr.</i>
<i>Non-urban</i>	0.8332	0.1103	0	0.0341	0.0163	0.0035	0	0.0026
<i>Resid.</i>	0	0.6439	0.3122	0	0	0.0439	0	0
<i>Comm.</i>	0	0	1	0	0	0	0	0
<i>Industr.</i>	0	0	0	1	0	0	0	0
<i>Instit.</i>	0	0.0252	0	0	0.9745	0	0	0
<i>Services</i>	0	0	0.7505	0	0	0.2495	0	0
<i>Water Bodies</i>	0	0	0	0	0	0	1	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1

For the simulation period 1962–1985, eleven variables have been selected (LSN-ESALQ-USP, 2003; SEMUPLAN, 1985), some of which are presented in FIGURE 6.43. The notations utilized for each map of variable employed in this simulation experiment are shown in TABLE 6.46.

TABLE 6.46 – Independent variables defining land use change in Piracicaba: 1962–1985.

<i>NOTATION</i>	<i>PHYSICAL OR SOCIOECONOMIC LAND USE CHANGE VARIABLE</i>
<i>dist_riv</i>	<i>Distances to rivers.</i>
<i>dist_com</i>	<i>Distances to the main commercial zone.</i>
<i>dist_ind</i>	<i>Distances to the industrial zone.</i>
<i>dist_inst</i>	<i>Distances to institutional zones.</i>
<i>dist_res</i>	<i>Distances to residential zones.</i>
<i>dist_leis</i>	<i>Distances to leisure/recreation zones.</i>
<i>main_rds</i>	<i>Distances to main paved and non-paved interurban roads.</i>
<i>urb_ext</i>	<i>Distances to main urban roads and their extensions.</i>
<i>int_rds</i>	<i>Distances to roads interconnecting isolated residential settlements.</i>
<i>ewper_rds</i>	<i>Distances to peripheral e-w roads.</i>
<i>trv_rds</i>	<i>Distances to transversal (sw-ne) interurban roads.</i>

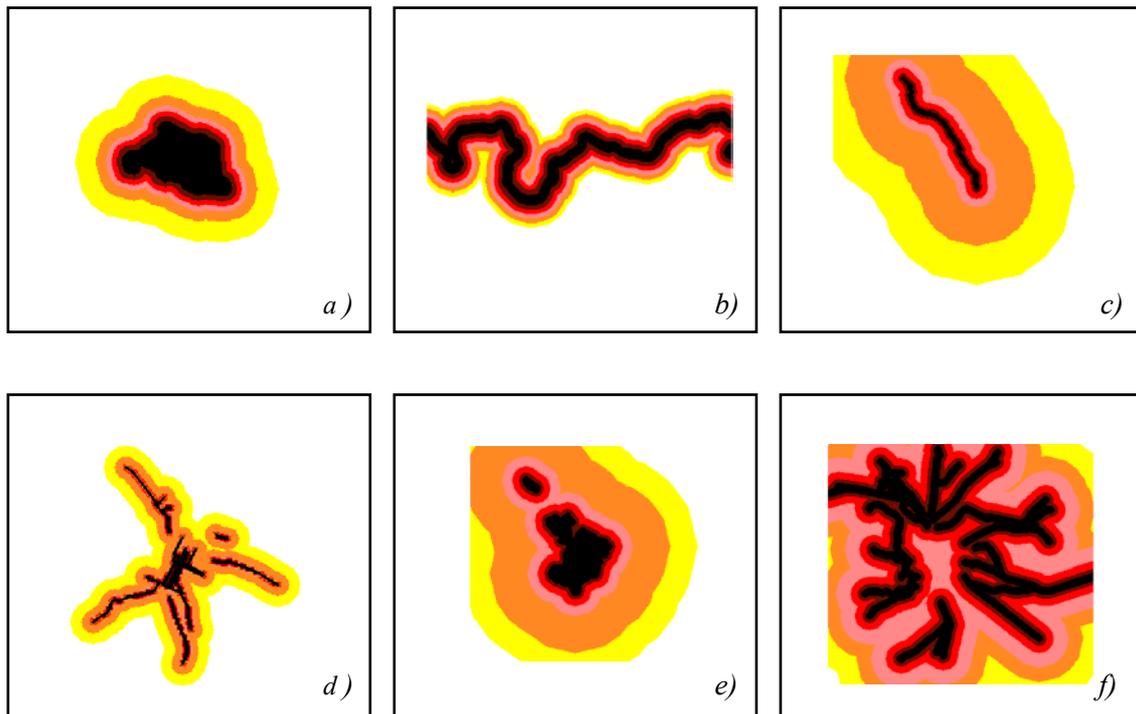


FIGURE 6.43 – Independent variables used to explain the land use transitions in Piracicaba during the simulation period 1962 – 1985: a) distances to institutional zones; b) distances to rivers; c) distances to roads interconnecting isolated residential settlements; d) distances to main urban roads and their extensions; e) distances to residential zones; f) distances to main paved and non-paved interurban roads.

TABLE 6.47 indicates which variable was selected to explain each of the nine existent transitions; TABLE 6.48 presents the values obtained for the Cramer's Coefficient (V) and the Joint Information Uncertainty (U) for the pairs of variables used to explain the same type of land use transition; and finally TABLE 6.49 presents the values for the positive weights of evidence.

TABLE 6.47 – Selection of variables defining land use change in Piracicaba: 1962-1985.

NOTATION	NU_	NU_	NU_	NU_	NU_	RES_	RES_	SERV_	INST_
	RES	IND	INST	SERV	LEIS	COM	SERV	COM	RES
<i>dist_riv</i>					♦				
<i>dist_com</i>		♦		♦		♦	♦	♦	♦
<i>dist_ind</i>			♦						
<i>dist_inst</i>		♦	♦						
<i>dist_res</i>	♦			♦	♦				♦
<i>dist_leis</i>					♦				♦
<i>main_rds</i>	♦								
<i>urb_ext</i>				♦			♦		
<i>int_rds</i>						♦			
<i>ewper_rds</i>		♦							
<i>trv_rds</i>			♦						

TABLE 6.48 – Associations between independent variables - Piracicaba: 1962–1985.

VARIABLE A	VARIABLE B	CRAMER'S STATISTIC ($V_{A,B}$)	UNCERTAINTY ($U_{A,B}$)
<i>dist_riv</i>	<i>dist_leis</i>	0.2071	0.1219
	<i>dist_res</i>	0.1367	0.0387
<i>dist_com</i>	<i>dist_inst</i>	0.2513	0.1630
	<i>dist_leis</i>	0.2967	0.1619
	<i>dist_res</i>	0.5375	0.4524
	<i>ewper_rds</i>	0.2152	0.1007
	<i>urb_ext</i>	0.3185	0.2243
<i>dist_ind</i>	<i>int_rds</i>	0.2708	0.1791
	<i>trv_rds</i>	0.1650	0.0920
<i>dist_inst</i>	<i>ewper_rds</i>	0.2343	0.1470
	<i>dist_ind</i>	0.1969	0.1279
	<i>trv_rds</i>	0.1440	0.0611
<i>dist_res</i>	<i>dist_leis</i>	0.2588	0.1208
	<i>main_rds</i>	0.2962	0.1607
	<i>urb_ext</i>	0.3229	0.2029

As none of the association measure values surpassed the threshold of 0.50 simultaneously for both indices, no variables preliminarily selected for modeling have been discarded from the analysis.

TABLE 6.49 – Values of W^+ for the selected independent variables - Piracicaba:
1962–1985.

<i>Land Use Transition</i>	<i>Variable</i>	<i>Positive Weights of Evidence W^+</i>						
		<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>
<i>NU_RES</i>	<i>dist_res</i> ¹	3.5372	2.8824	2.4818	1.2739	-0.7771	-0.7661	0
	<i>main_rds</i> ²	0.9624	0.7108	0.3230	-0.8406	0	0	0
<i>NU_IND</i>	<i>dist_com</i> ³	0	0	0	0	-0.3359	0.6469	-4.1709
	<i>dist_inst</i> ⁴	1.0527	1.7593	1.8492	1.7601	1.6944	1.6070	-0.9707
<i>NU_INST</i>	<i>ewper_rds</i> ²	0.4117	0.7181	1.1100	1.8514	1.9465	-0.6499	-2.0727
	<i>dist_inst</i> ⁴	2.3179	2.5599	2.5427	2.3438	2.2362	1.5974	0
<i>NU_SERV</i>	<i>dist_ind</i> ⁵	2.5498	3.3498	3.8420	3.7771	3.4886	2.0421	0
	<i>trv_rds</i> ⁶	1.7467	1.6864	1.9748	1.6190	1.7839	1.2866	-6.5520
<i>NU_LEIS</i>	<i>dist_res</i> ¹	2.5528	2.5745	1.5656	0.6994	-0.6624	0	0
	<i>dist_com</i> ³	0	0	2.0271	2.2408	1.2789	-2.0861	0
<i>RES_COM</i>	<i>urb_ext</i> ⁷	5.7160	1.9544	0.0494	0.5500	-0.0357	0	0
	<i>dist_res</i> ¹	2.4515	3.0108	2.2754	-1.4958	0	0	0
<i>RES_SERV</i>	<i>dist_riv</i> ⁸	3.0856	0.9169	-0.4063	0	0	0	0
	<i>dist_leis</i> ⁸	5.6992	4.8919	3.4233	3.5253	3.4364	1.2944	0
<i>SERV_COM</i>	<i>dist_com</i> ³	2.6250	1.4373	-0.8688	0	0	0.6390	0
	<i>int_rds</i> ⁹	0.2626	0.4668	-0.1211	-0.2062	0.2615	0	0
<i>IND_RES</i>	<i>dist_com</i> ³	0	-1.3292	-0.2857	1.5782	0	0	0
	<i>urb_ext</i> ⁷	3.2240	-0.2676	-2.8540	0	0	0	0
<i>IND_RES</i>	<i>dist_com</i> ³	0	0	-2.9027	0	0	0	0
	<i>dist_res</i> ¹	1.3750	1.1637	0.2279	0	0	0	0
<i>IND_RES</i>	<i>dist_com</i> ³	0	0	1.3783	0.2638	-2.1116	0	0
	<i>dist_leis</i> ⁸	3.1220	-1.1165	0	0	0	0	-0.3929

Note: Distance bands in meters

¹ 1: 0 -250; 2: 250-500; 3: 500-1000; 4: 1000-1750; 5: 1750-4500; 6: 4500-6500; 7: >6500

² 1: 0 -250; 2: 250-500; 3: 500-1000; 4: 1000-2000; 5: 2000-3000; 6: 3000-4500; 7: >4500

³ 1: 0 -500; 2: 500-1000; 3: 1000-2000; 4: 2000-3000; 5: 3000-4500; 6: 4500-8000; 7: >8000

⁴ 1: 0 -250; 2: 250-500; 3: 500-750; 4: 750-1000; 5: 1000-1500; 6: 1500-2500; 7: >2500

⁵ 1: 0 -250; 2: 250-500; 3: 500-750; 4: 750-1000; 5: 1000-1500; 6: 1500-3000; 7: > 3000

⁶ 1: 0 -50; 2: 50-100; 3: 100-150; 4: 150-200; 5: 200-500; 6: 500-1500; 7: > 1500

⁷ 1: 0 -50; 2: 50-100; 3: 100-150; 4: 150-200; 5: 200-500; 6: 500-1000; 7: > 1000

⁸ 1: 0 -250; 2: 250-500; 3: 500-750; 4: 750-1000; 5: 1000-1250; 6: 1250-1500; 7: > 1500

⁹ 1: 0 -100; 2: 100-250; 3: 250-550; 4: 550-1000; 5: 1000-3000; 6: 3000-5000; 7: > 5000

Considering the calibration process carried out for this experiment, it became clear that the probability of certain non-urban areas in the city of Piracicaba to shelter residential settlements (“nu_res land use transition”) largely depends on the previous existence of this type of settlements in their surroundings, because this implies the possibility of extending existing nearby infrastructure. It also depends on the available accessibility to such areas.

As to the transition of “non-urban areas to industrial use (nu_ind)”, there are three great driving forces. Firstly, industrial areas demand the proximity to institutional areas, for the latter are located in the outskirts, and therefore, in cheaper plots. Secondly, although new industrial settlements take place in more peripheral areas, they would rather be located within a reasonable distance from central clusters of commercial activities, which also account for the logistical support of industrial districts. And lastly, industrial areas request good accessibility conditions, in view of the continuously increasing need of raw material inflow and final production outflow.

Concerning the implementation of large institutional areas (“nu_inst”), it is observable that they arise close to industrial areas, in view of the need of large and cheap plots for their expansion. They are also located near peripheral roads and previously existent institutional areas, since they grow as extensions of already established institutional zones.

Regarding the transition of “non-urban areas to services use (nu_serv)”, three major factors are crucial: the proximity of these areas to commercial zones, their closeness to areas of residential use, and last but not least, their strategic location in relation to the main urban roads of Piracicaba. The first factor accounts for the suppliers market (and in some cases also consumers market) of services; the second factor represents the consumers market itself; and the third and last factor corresponds to the accessibility for both markets related to the services use.

The creation of leisure and recreation zones (“nu_leis”), on its turn, takes place adjacent to already existent zones of this type, since they are commonly created as extensions of previous leisure and recreation areas. These areas are created along low and flat

riverbanks, since they are floodable and hence unsuitable for sheltering other urban uses.

Leisure and recreation zones are as well strategically located in relation to their catchment area, i.e. near central residential areas, which are those sheltering higher population densities.

The transition “residential to services use (res_serv)” supposes good accessibility conditions and a location within a reasonable distance from the suppliers (and in some cases, also consumers) market, represented by the central commercial area.

The conversion of “residential use into commercial zones (res_com)” demonstrated that farther residential settlements are prone to develop their own commercial nuclei, which are likely to occur along roads of greater hierarchical importance, as a means to assure both strategic visibility and good accessibility conditions. Furthermore, residential settlements located very near the central commercial area are susceptible of giving place to commercial use.

Regarding the change of services into commercial use, this can be explained by the fact that services axes established very close to the central commercial area are prone to be “devoured” by the latter.

And finally, the last type of transition concerns the transformation of institutional areas into residential zones. This conversion regards institutional areas located in the immediacies of residential settlements, which are valorized by their proximity to leisure and commercial areas.

The maps of estimated transition probabilities surfaces, generated by DINAMICA upon basis of the values of the positive weights of evidence (W^+), together with their respective land use transition maps are seen in FIGURES 6.44 a, 6.44b and 6.44c.

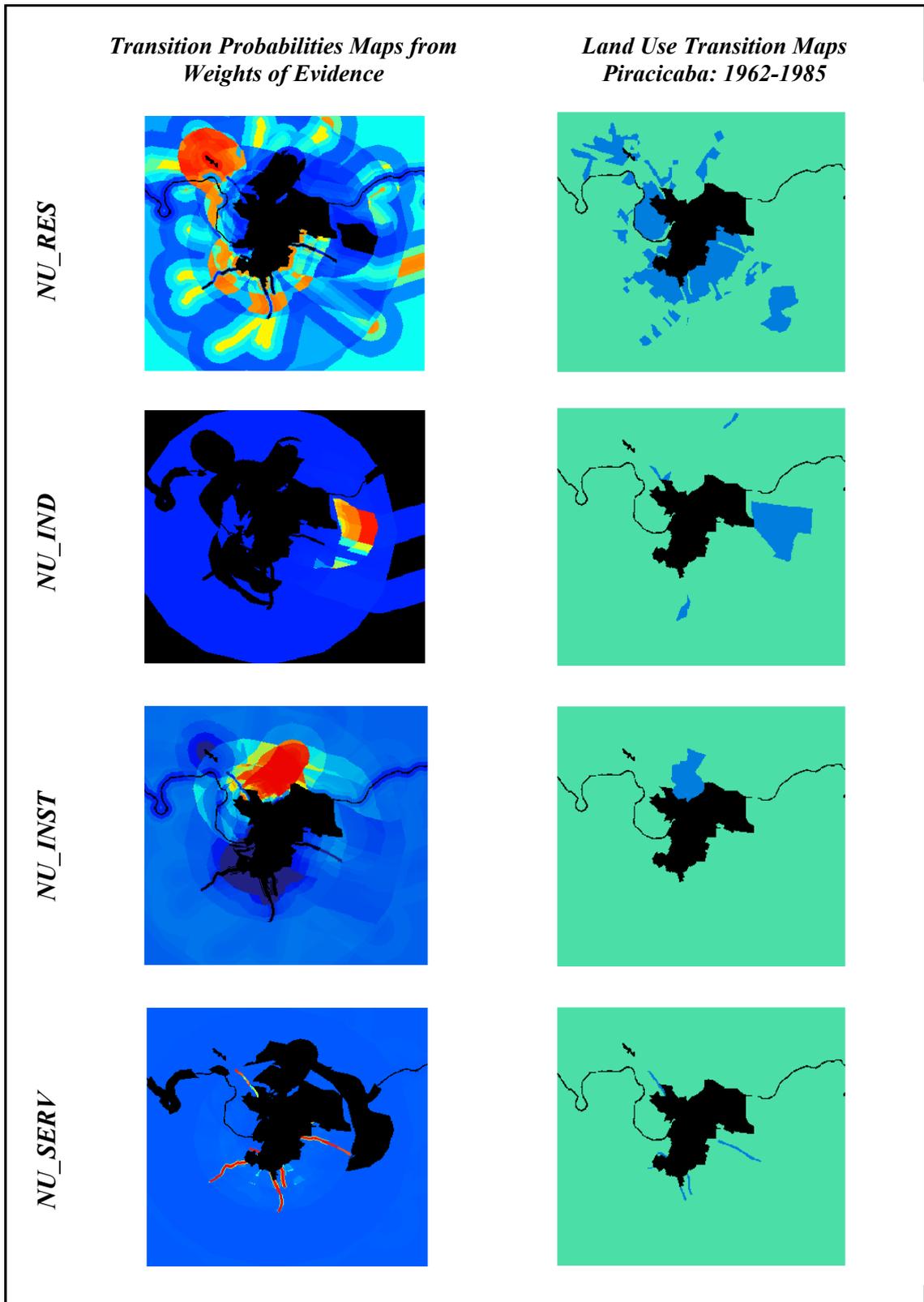


FIGURE 6.44a – Estimated transition probability surfaces and land use change - Piracicaba.

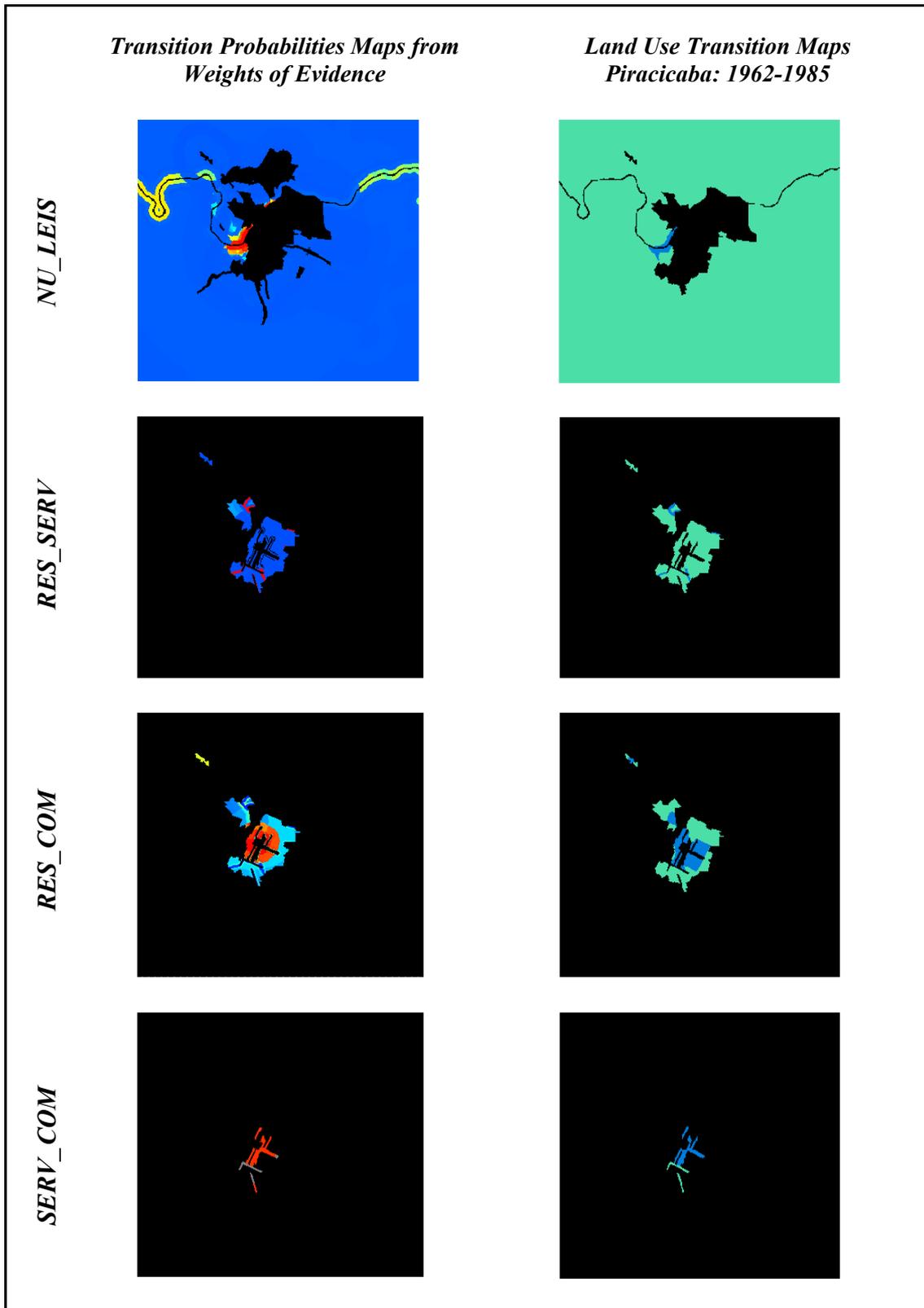


FIGURE 6.44b – Estimated transition probability surfaces and land use change - Piracicaba.

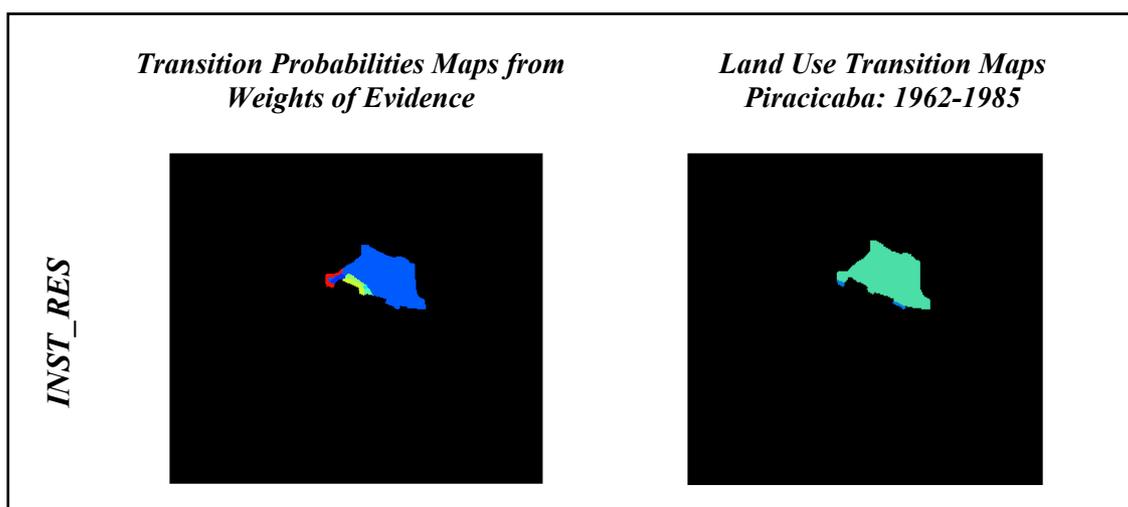


FIGURE 6.44c – Estimated transition probability surfaces and land use change - Piracicaba.

The three best simulation results produced for the period 1962-1985 are presented in FIGURE 6.45. The internal DINAMICA parameters associated with these optimal simulations are seen in TABLE 6.50, whose statistical validation tests for windows size of 3x3, 5x5 and 9x9, and $k = 0.5$ are listed on TABLE 6.51.

TABLE 6.50 – DINAMICA internal parameters for the simulation of urban land use change in Piracicaba: 1962–1985.

<i>Land Use Transition</i>	<i>Average Size of Patches</i>	<i>Variance of Patches Size</i>	<i>Proportion of 'Expander'</i>	<i>Proportion of 'Patcher'</i>	<i>Number of Iterations</i>
<i>NU_RES</i>	300	30	0.85	0.15	300
<i>NU_IND</i>	1500	1	0.10	0.90	300
<i>NU_INST</i>	2000	1	0	1.00	300
<i>NU_SERV</i>	5	0	0.05	0.95	300
<i>NU_LEIS</i>	100	1	0.15	0.85	300
<i>RES_COM</i>	300	10	0.20	0.80	300
<i>RES_SERV</i>	8	0	0.05	0.95	300
<i>SERV_COM</i>	100	1	0.65	0.35	300
<i>INST_RES</i>	20	1	0.50	0.50	300

TABLE 6.51 – Goodness-of-fit tests for the best land use change simulations of Piracicaba: 1962-1985.

<i>Simulations</i>	<i>Multiple Resolution Goodness-of-Fit (F)</i>
S_1	$F = 0.842271$
S_2	$F = 0.834702$
S_3	$F = 0.840050$

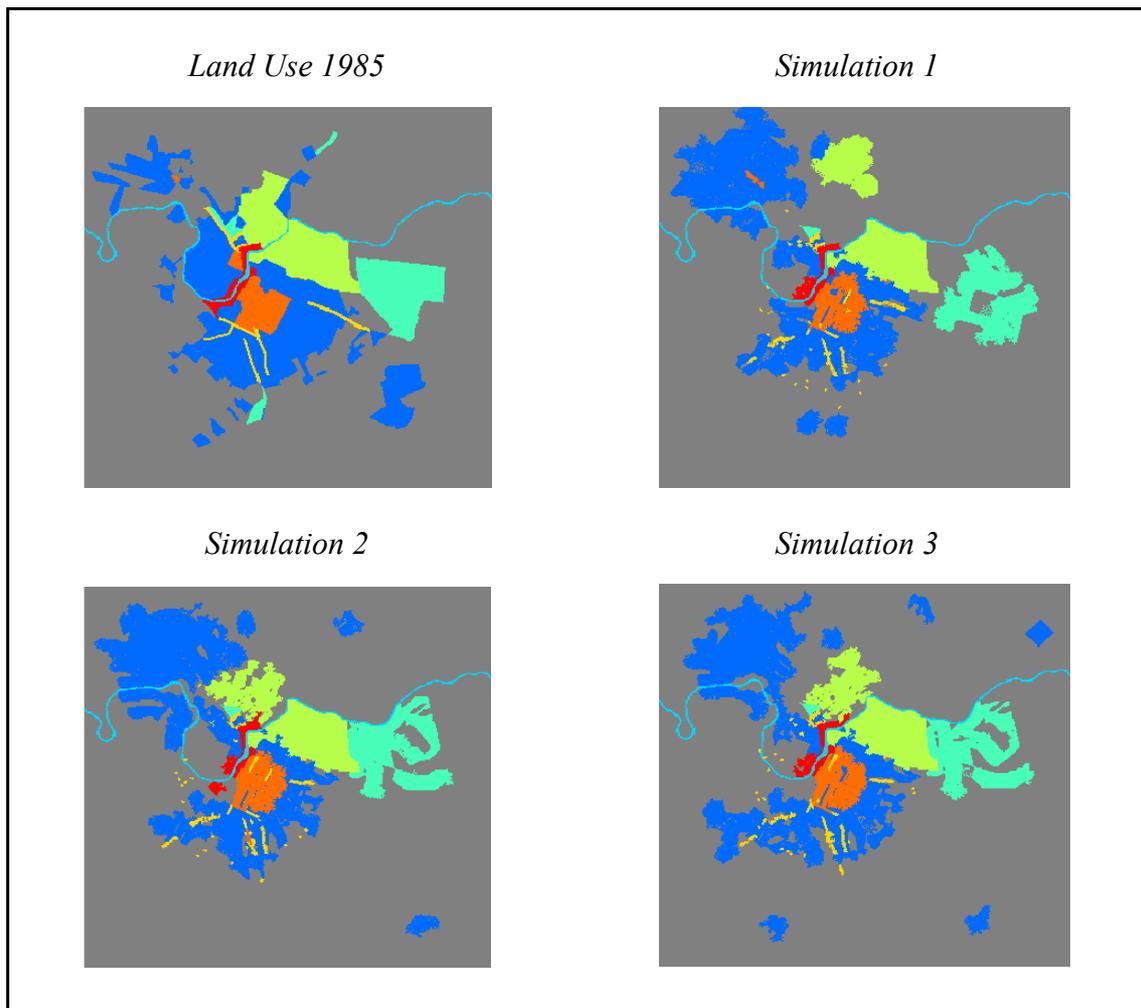


FIGURE 6.45 – The best simulations compared to the actual land use in Piracicaba in 1985.

FIGURE 6.45 – The best simulations compared to the actual land use in Piracicaba in 1985.

6.2.2 Simulation Period: 1985 - 1999

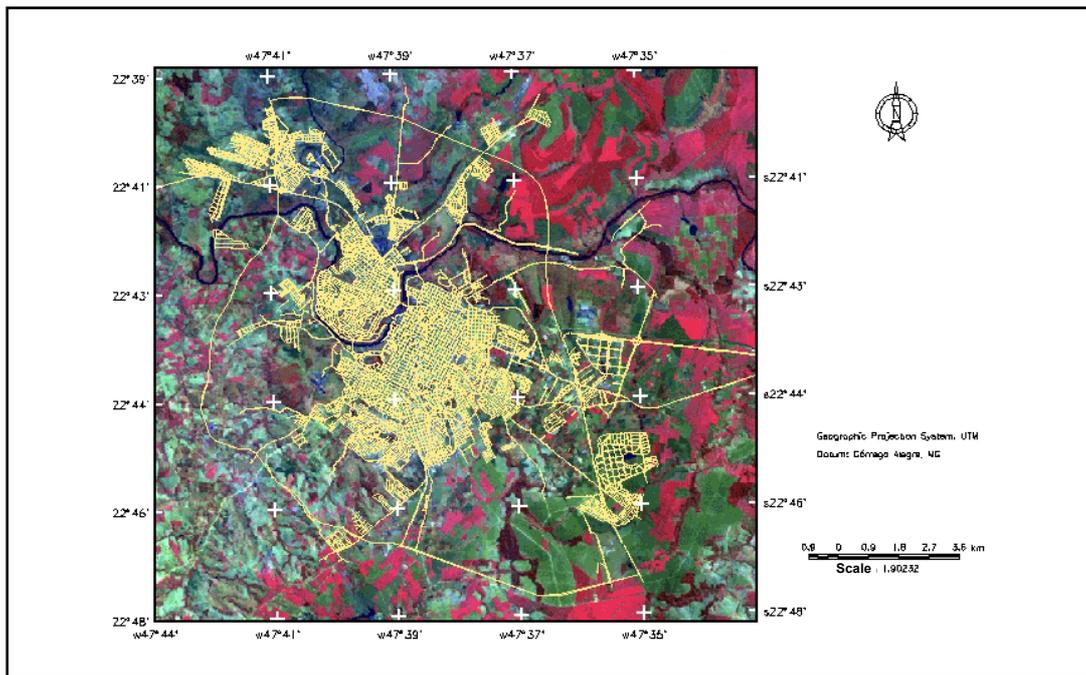


FIGURE 6.46 – Piracicaba TM – 5 image and official city map in 1985.
SOURCE: INPE (1985) and SEMUPLAN (1985).

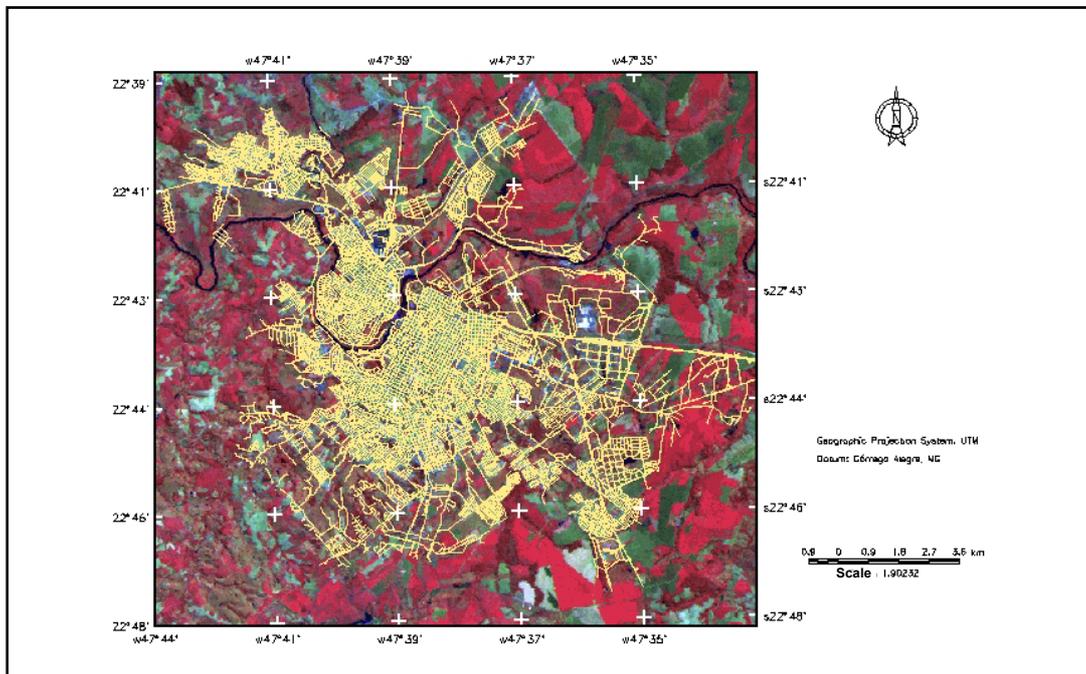


FIGURE 6.47 – Piracicaba TM – 5 image and official city map in 1999.
SOURCE: INPE (1999) and SEMUPLAN (1999).

As previously stated in Section 6.2.1, the population of Piracicaba at the initial time of this simulation period was 252,079 inhabitants, from which 198,407 people were urban inhabitants. In 1999, the total population accounted 319,104, out of which 309,531 inhabitants lived in urban areas (IBGE, 1999). The population growth rate for this period is around 1.56%. The effects of this population growth on the urban area extension can be seen in FIGURES 6.46 and 6.47, which present the official city maps for the initial and final time of simulation.

The initial and final land use maps used in the simulation period 1985 – 1999 (FIGURE 6.48) were elaborated upon basis of the two official city maps previously shown, of generalization procedures applied to the land use maps of 1985 and 1999 (SEMUPLAN, 1985, 1999) and of the digital satellite images of Piracicaba in 1985 and 1999 (INPE, 1985, 1999).

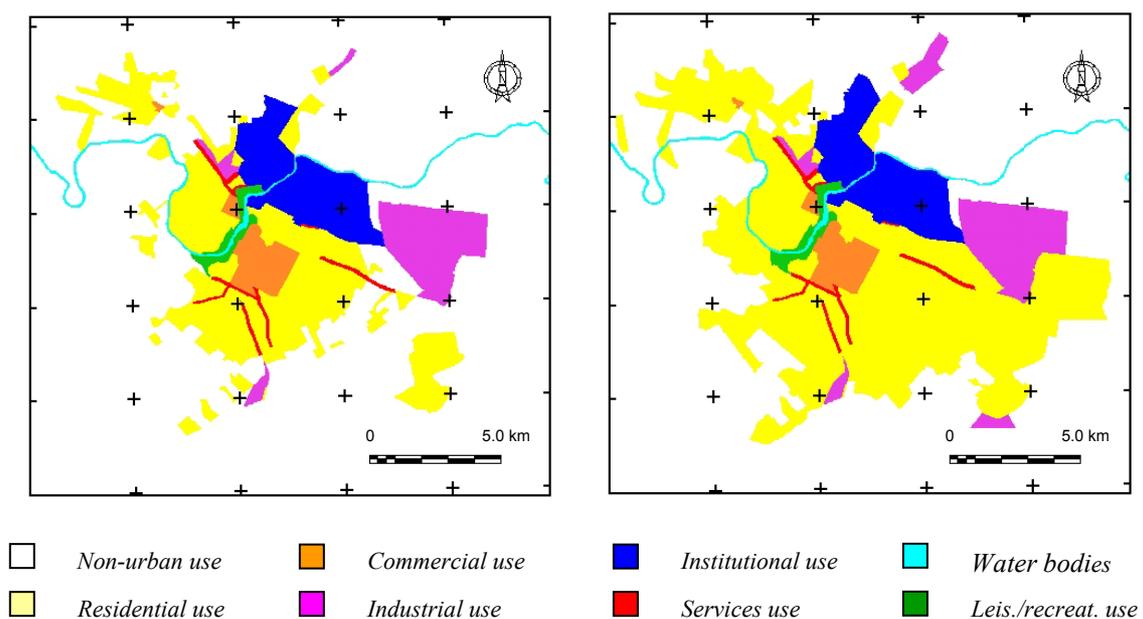


FIGURE 6.48 – Generalized land use map in Piracicaba in 1985 (left) and 1999 (right).

A cross-tabulation operation was made between both land use maps (FIGURE 6.49) so as to generate transition percentages for the existent types of land use change (TABLES 6.52 and 6.53).

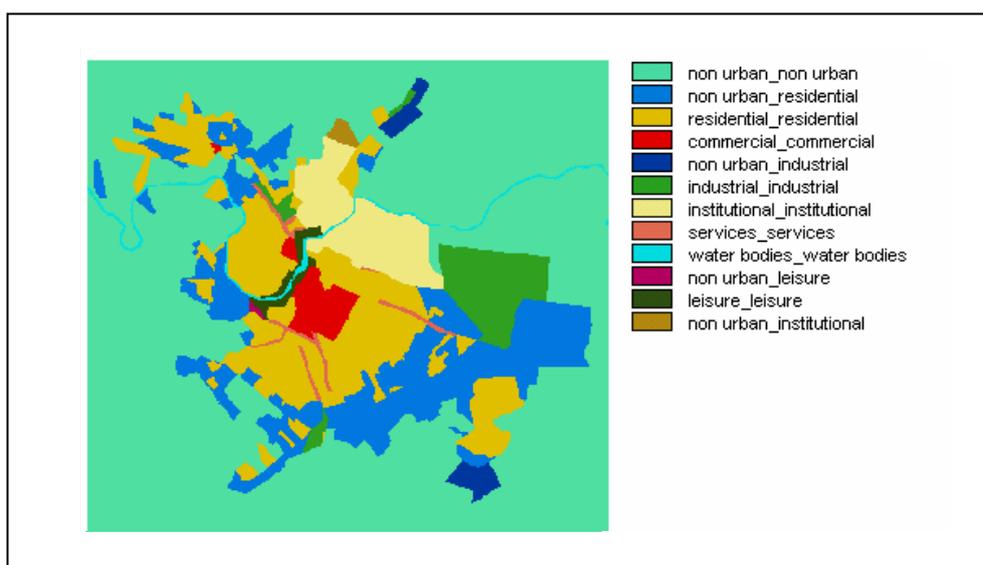


FIGURE 6.49 – Cross-tabulation map between Piracicaba land use maps of 1985 and 1999.

TABLE 6.52 – Existent land use transitions in Piracicaba: 1985–1999.

<i>NOTATION</i>	<i>LAND USE TRANSITION</i>
<i>NU_RES</i>	<i>Non-urban to residential</i>
<i>NU_IND</i>	<i>Non-urban to industrial</i>
<i>NU_INST</i>	<i>Non-urban to institutional</i>
<i>NU_LEIS</i>	<i>Non-urban to leisure/recreation</i>

TABLE 6.53 – Matrix of global transition probabilities for Piracicaba: 1985–1999.

<i>Land Use</i>	<i>Non-urban</i>	<i>Resid.</i>	<i>Comm.</i>	<i>Industr.</i>	<i>Instit.</i>	<i>Services</i>	<i>Water B.</i>	<i>Leis./Recr.</i>
<i>Non-urban</i>	0.8353	0.1501	0	0.0113	0.0028	0	0	0.0005
<i>Resid.</i>	0	1	0	0	0	0	0	0
<i>Comm.</i>	0	0	1	0	0	0	0	0
<i>Industr.</i>	0	0	0	1	0	0	0	0
<i>Instit.</i>	0	0	0	0	1	0	0	0
<i>Services</i>	0	0	0	0	0	1	0	0
<i>Water Bodies</i>	0	0	0	0	0	0	1	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1

For the simulation period 1985–1999, eight variables have been selected (SEMUPLAN, 1999), most of which are shown in FIGURE 6.50.

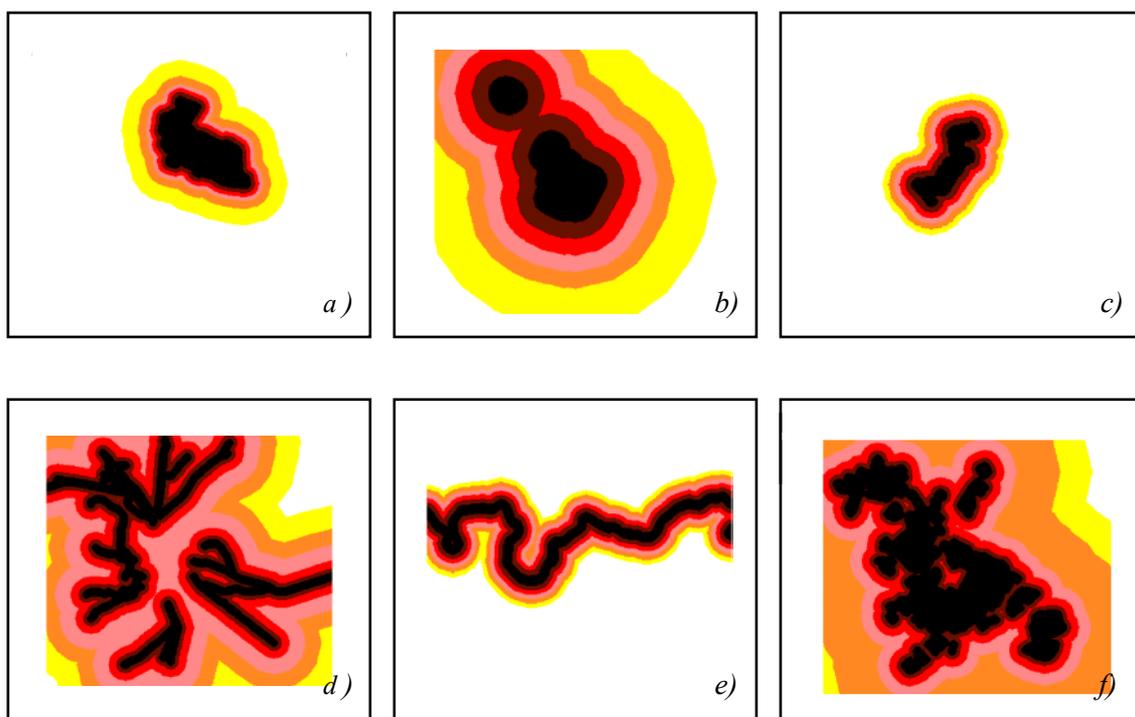


FIGURE 6.50 – Independent variables used to explain the land use transitions in Piracicaba during the simulation period 1985 – 1999: a) distances to institutional zones; b) distances to commercial zones; c) distances to leisure and recreation zones; d) distances to main paved and non-paved urban and interurban roads; e) distances to rivers; f) distances to residential zones.

TABLE 6.54 shows the notations utilized for each map of variable employed in this simulation experiment; TABLE 6.55 indicates which variable was selected to explain each of the five existent transitions; TABLE 6.56 presents the values obtained for the Cramer’s Coefficient (V) and the Joint Information Uncertainty (U) for the pairs of variables used to explain the same type of land use transition; and finally TABLE 6.57 presents the values for the positive weights of evidence.

TABLE 6.54 – Independent variables defining land use change in Piracicaba: 1985–1999.

<i>NOTATION</i>	<i>PHYSICAL OR SOCIOECONOMIC LAND USE CHANGE VARIABLE</i>
<i>dist_riv</i>	<i>Distances to rivers.</i>
<i>dist_com</i>	<i>Distances to commercial zones.</i>
<i>dist_ind</i>	<i>Distances to small-sized industrial zones.</i>
<i>dist_inst</i>	<i>Distances to institutional zones.</i>
<i>dist_res</i>	<i>Distances to residential zones.</i>
<i>dist_leis</i>	<i>Distances to leisure/recreation zones.</i>
<i>int_rds</i>	<i>Distances to main interurban roads.</i>
<i>main_rds</i>	<i>Distances to main paved and non-paved urban and interurban roads.</i>

TABLE 6.55 – Selection of variables defining land use change in Piracicaba: 1985–1999.

<i>NOTATION</i>	<i>NU_RES</i>	<i>NU_IND</i>	<i>NU_INST</i>	<i>NU_LEIS</i>
<i>dist_riv</i>				♦
<i>dist_com</i>		♦		
<i>dist_ind</i>		♦		
<i>dist_inst</i>			♦	
<i>dist_res</i>	♦	♦		♦
<i>dist_leis</i>				♦
<i>int_rds</i>		♦	♦	
<i>main_rds</i>	♦			

Likewise the preceding simulation period, as none of the association measure values surpassed the threshold of 0.50 simultaneously for both indices, no variables preliminarily selected for modeling have been discarded from the analysis.

TABLE 6.56 – Associations between independent variables - Piracicaba: 1985–1999.

<i>VARIABLE A</i>	<i>VARIABLE B</i>	<i>CRAMER'S STATISTIC</i> <i>(V_{A,B})</i>	<i>UNCERTAINTY</i> <i>(U_{A,B})</i>
<i>dist_res</i>	<i>dist_riv</i>	0.1120	0.0252
	<i>dist_com</i>	0.3955	0.2690
	<i>dist_leis</i>	0.1445	0.0674
	<i>dist_ind</i>	0.1518	0.0492
	<i>main_rds</i>	0.3141	0.1715
	<i>int_rds</i>	0.2458	0.1025
<i>dist_com</i>	<i>dist_ind</i>	0.1743	0.0537
	<i>int_rds</i>	0.1139	0.0243
<i>int_rds</i>	<i>dist_ind</i>	0.3339	0.1764
	<i>dist_inst</i>	0.1808	0.0870
<i>dist_leis</i>	<i>dist_riv</i>	0.2328	0.1368

TABLE 6.57 – Values of W^+ for the selected independent variables - Piracicaba: 1985–1999.

<i>Land Use Transition</i>	<i>Variable</i>	<i>Positive Weights of Evidence W⁺</i>						
		1	2	3	4	5	6	7
<i>NU_RES</i>	<i>dist_res</i> ¹	2.1009	1.3378	0.2675	-0.7339	-1.3935	0	0
	<i>main_rds</i> ²	1.2736	0.9527	0.4770	-0.8162	-3.7650	0	0
<i>NU_IND</i>	<i>dist_com</i> ³	0	0	0	0	-1.5333	1.0890	-2.8871
	<i>dist_ind</i> ⁴	2.8956	0	0	0	0	0.8313	0
	<i>dist_res</i> ¹	-0.5849	0.5347	1.1667	0.7213	0	0	0
<i>NU_INST</i>	<i>int_rds</i> ⁵	2.2310	0.9677	0.5135	0.9252	0	0	0
	<i>dist_inst</i> ⁶	4.0903	3.5881	3.0078	-0.4728	0	0	0
	<i>int_rds</i> ⁵	1.2368	1.7747	1.7116	0	0	0	0
<i>NU_LEIS</i>	<i>dist_riv</i> ⁷	1.2483	2.8570	1.2788	0	0	0	0
	<i>dist_res</i> ¹	2.2379	0.0639	0	0	0	0	0
	<i>dist_leis</i> ⁷	7.9451	4.0632	0	0	0	0	0

Note: Distance bands in meters

¹ 1: 0-250; 2: 250-500; 3: 500-1000; 4: 1000-1750; 5: 1750-4500; 6: 4500-6500; 7: > 6500

² 1: 0-250; 2: 250-500; 3: 500-1000; 4: 1000-2000; 5: 2000-3000; 6: 3000-4500; 7: > 4500

³ 1: 0-1000; 2: 1000-2000; 3: 2000-3000; 4: 3000-4000; 5: 4000-5000; 6: 5000-7500; 7: >7500

⁴ 1: 0-750; 2: 750-1500; 3: 1500-2500; 4: 2500-3500; 5: 3500-4500; 6: 4500-6500; 7: >6500

⁵ 1: 0-250; 2: 250-750; 3: 750-1250; 4: 1250-2250; 5: 2250-3250; 6: 3250-4250; 7: > 4250

⁶ 1: 0-250; 2: 250-500; 3: 500-750; 4: 750-1000; 5: 1000-1500; 6: 1500-2500; 7: > 2500

⁷ 1: 0-250; 2: 250-500; 3: 500-750; 4: 750-1000; 5: 1000-1250; 6: 1250-1500; 7: > 1500

Likewise the preceding period (1962-1985), the transition “non-urban to residential use” largely depends on the previous existence of residential settlements in their surroundings, because this implies the possibility of extending existing nearby infrastructure. It also depends on the available accessibility to such areas. In the same way as the former simulation period, the implementation of large institutional areas (“nu_inst”) takes place near peripheral roads and previously existent institutional areas, since they grow as extensions of already established institutional zones.

The expansion of industrial districts requires the proximity to previously existent industrial zones and the availability of road access. This transition also supposes the nearness to the labor force supply centers (peripheral residential areas) and also a location not too distant from commercial zones, since industrial areas depend on commercial activities for logistical support. And finally, the arguments used in the former simulation period to explain the conversion of non-urban use into leisure and recreation zones are also valid for the current simulation period.

The maps of estimated transition probabilities surfaces, generated by DINAMICA upon basis of the values of the positive weights of evidence (W^+), together with their respective land use transition maps are seen in FIGURE 6.51. The black color in the maps to the left correspond to areas of null probability, the blue refers to areas of low transition probability, whereas the red color accounts for areas with the highest transition probabilities. In the maps to the right, the blue color refers to areas where the considered land use transition took place, and the green color, to areas where the origin land use remained as such or changed to other uses. The black color in these maps relates to areas with an origin use different from the one of the transition under consideration.

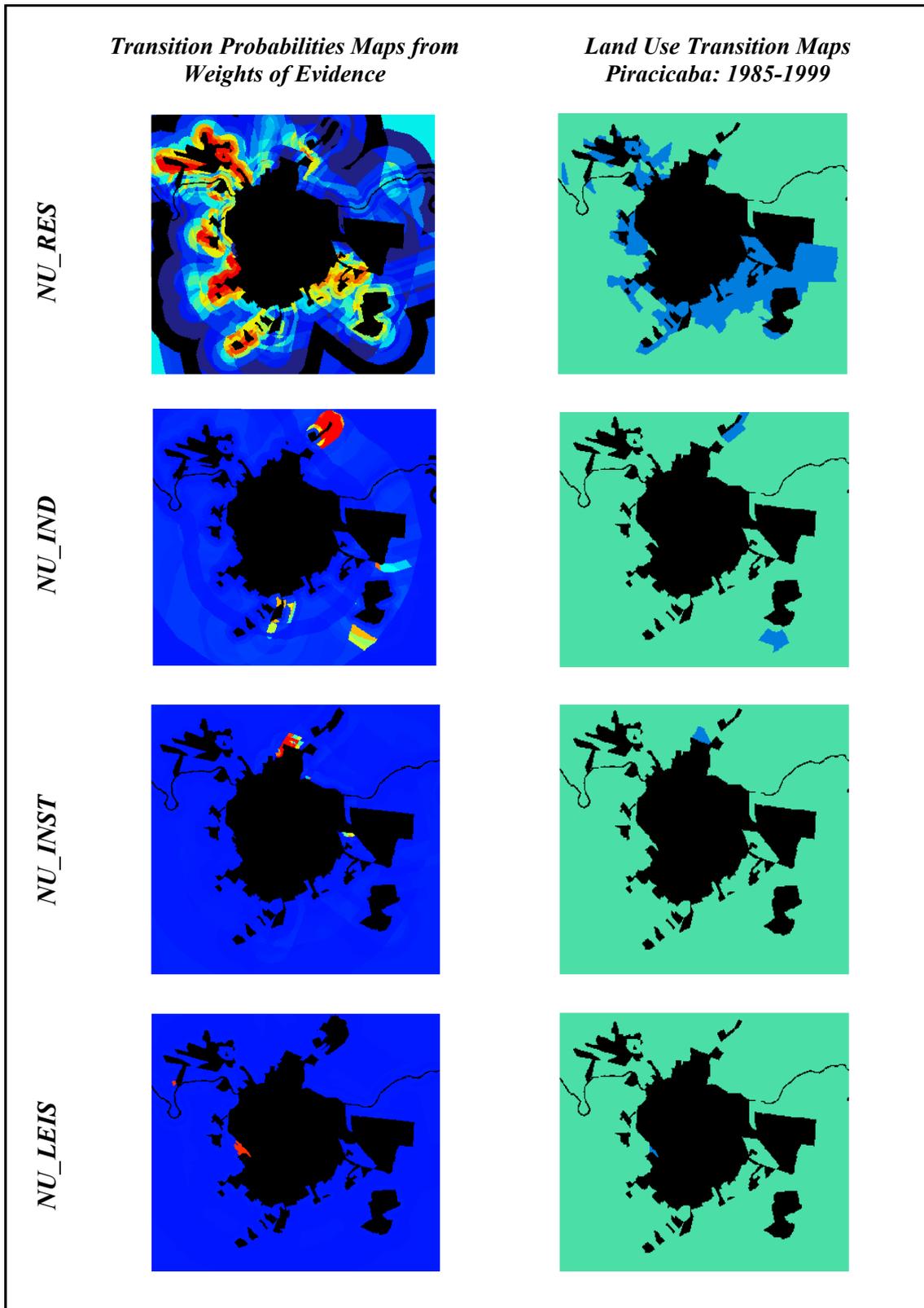


FIGURE 6.51 – Estimated transition probability surfaces and land use change - Piracicaba.



The three best simulation results produced for the period 1985-1999 are presented in FIGURE 6.52. The internal DINAMICA parameters associated with these optimal simulations are seen in TABLE 6.58, whose statistical validation tests for windows size of 3x3, 5x5 and 9x9, and $k = 0.5$ are listed on TABLE 6.59.

TABLE 6.58 – DINAMICA internal parameters for the simulation of urban land use change in Piracicaba: 1985–1999.

<i>Land Use Transition</i>	<i>Average Size of Patches</i>	<i>Variance of Patches Size</i>	<i>Proportion of 'Expander'</i>	<i>Proportion of 'Patcher'</i>	<i>Number of Iterations</i>
<i>NU_RES</i>	300	30	0.85	0.15	10
<i>NU_IND</i>	150	1	0.45	0.55	10
<i>NU_INST</i>	75	1	1.00	0	10
<i>NU_LEIS</i>	20	0	1.00	0	10

TABLE 6.59 – Goodness-of-fit tests for the best land use change simulations of Piracicaba: 1985-1999.

<i>Simulations</i>	<i>Multiple Resolution Goodness-of-Fit (F)</i>
S_1	$F = 0.862682$
S_2	$F = 0.864872$
S_3	$F = 0.864644$

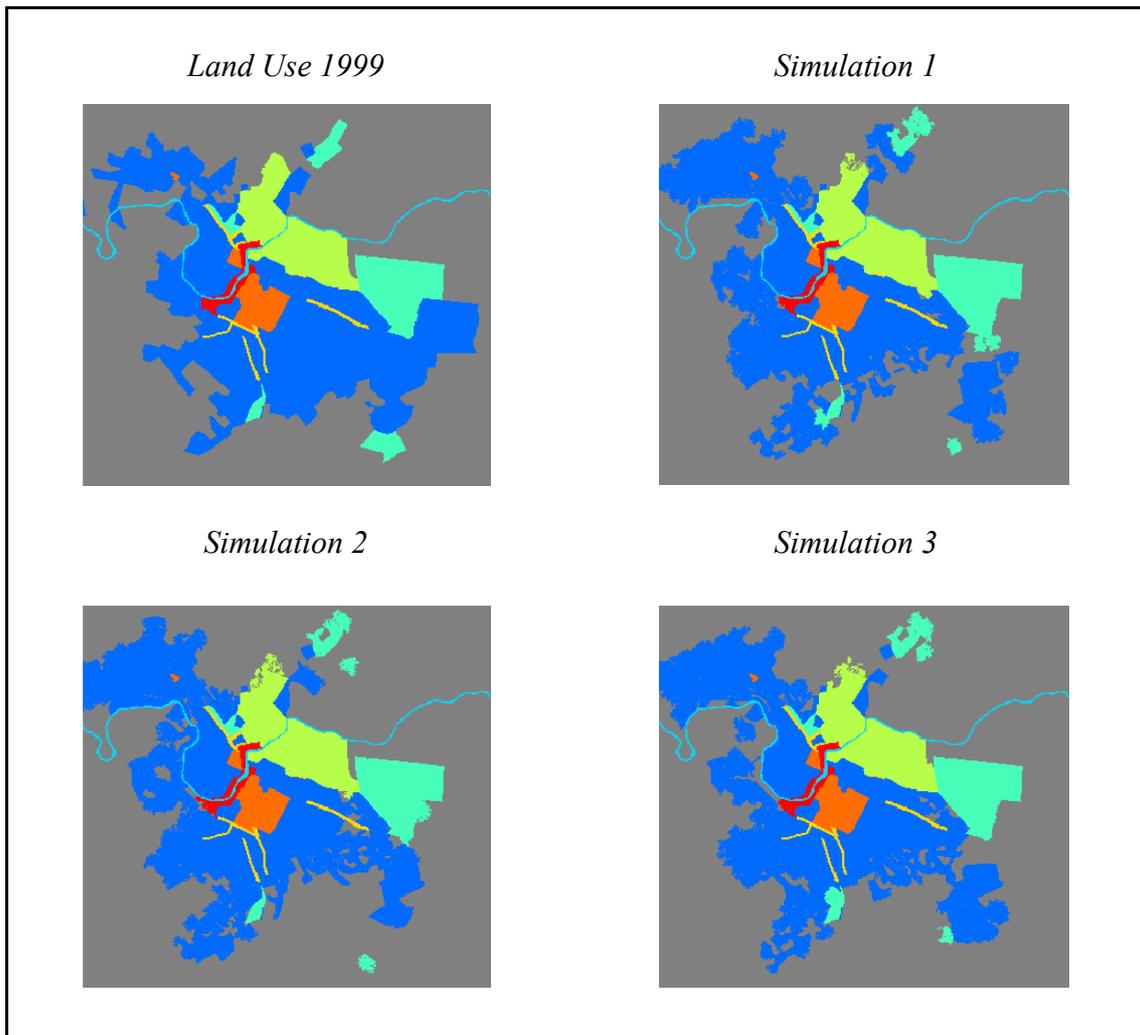


FIGURE 6.52 – The best simulations compared to the actual land use in Piracicaba in 1999.

6.2.3 Yearly Simulations: 1962 – 1999

According to Equation (5.69) in Section 5.3.1, the global matrix of transition can be decomposed in annual transition probabilities by the principal components method. This has been done separately for each simulation period (TABLES 6.60 and 6.61). The yearly simulation outputs can be seen in FIGURES 6.53a, 6.53b, 6.53c, 6.53d and 6.53e.

TABLE 6.60 – Matrix of yearly transition probabilities for Piracicaba: 1962–1985.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Water B.</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	9.9210×10^{-1}	6.4258×10^{-3}	0	1.6166×10^{-3}	7.8241×10^{-4}	4.2635×10^{-5}	0	1.2184×10^{-4}
<i>Residential</i>	0	9.8105×10^{-1}	1.4547×10^{-2}	0	0	4.4082×10^{-3}	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	1.3608×10^{-3}	0	0	9.9889×10^{-1}	0	0	0
<i>Services</i>	0	0	5.8574×10^{-2}	0	0	9.4143×10^{-1}	0	0
<i>Water B.</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

TABLE 6.61 – Matrix of yearly transition probabilities for Piracicaba: 1985–1999.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Water B.</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	9.8723×10^{-1}	1.1639×10^{-2}	0	8.7816×10^{-4}	2.1349×10^{-4}	0	0	4.2532×10^{-5}
<i>Residential</i>	0	1.00	0	0	0	0	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Water B.</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

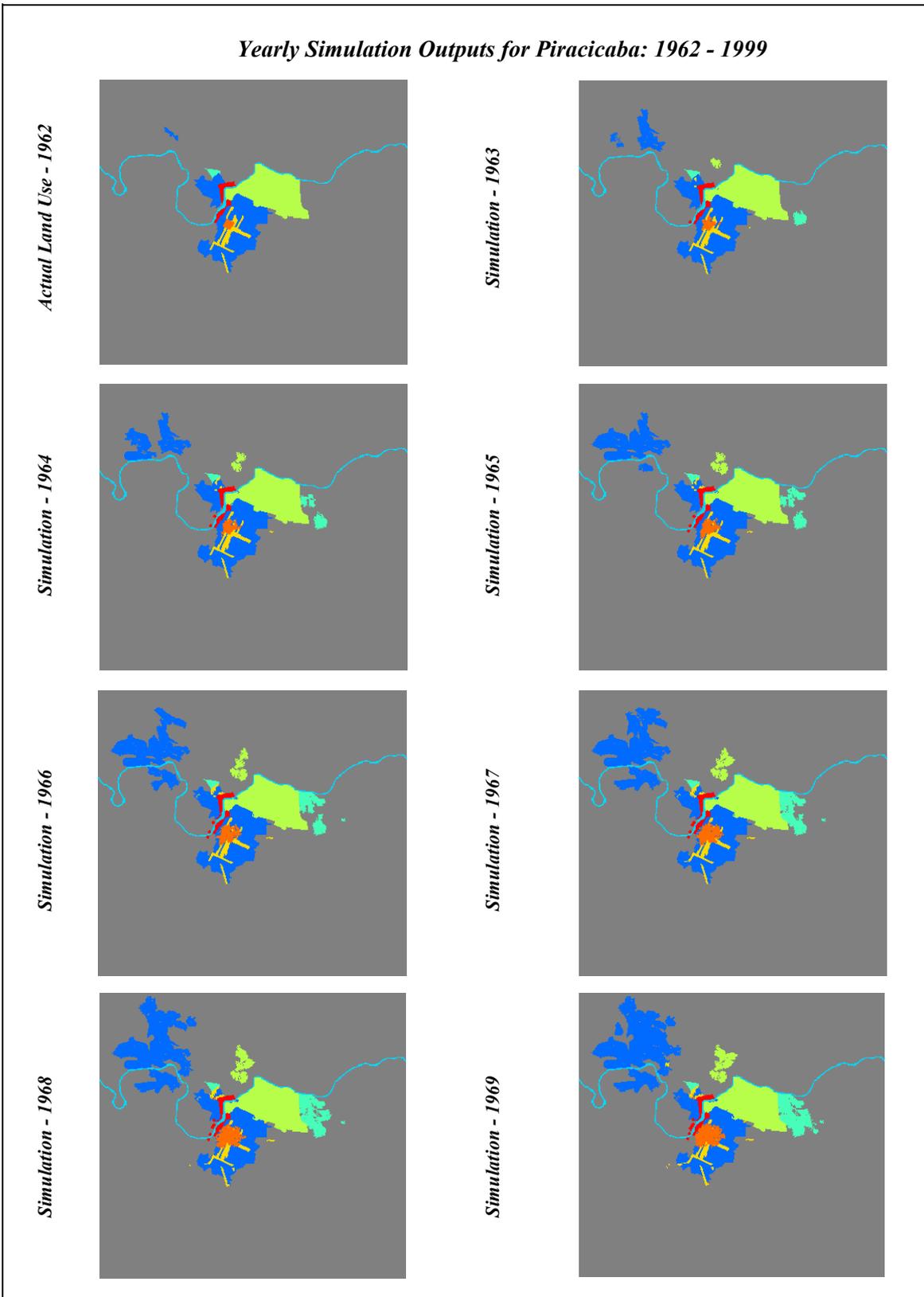


FIGURE 6.53a – Yearly simulation outputs for Piracicaba: 1962 - 1969.

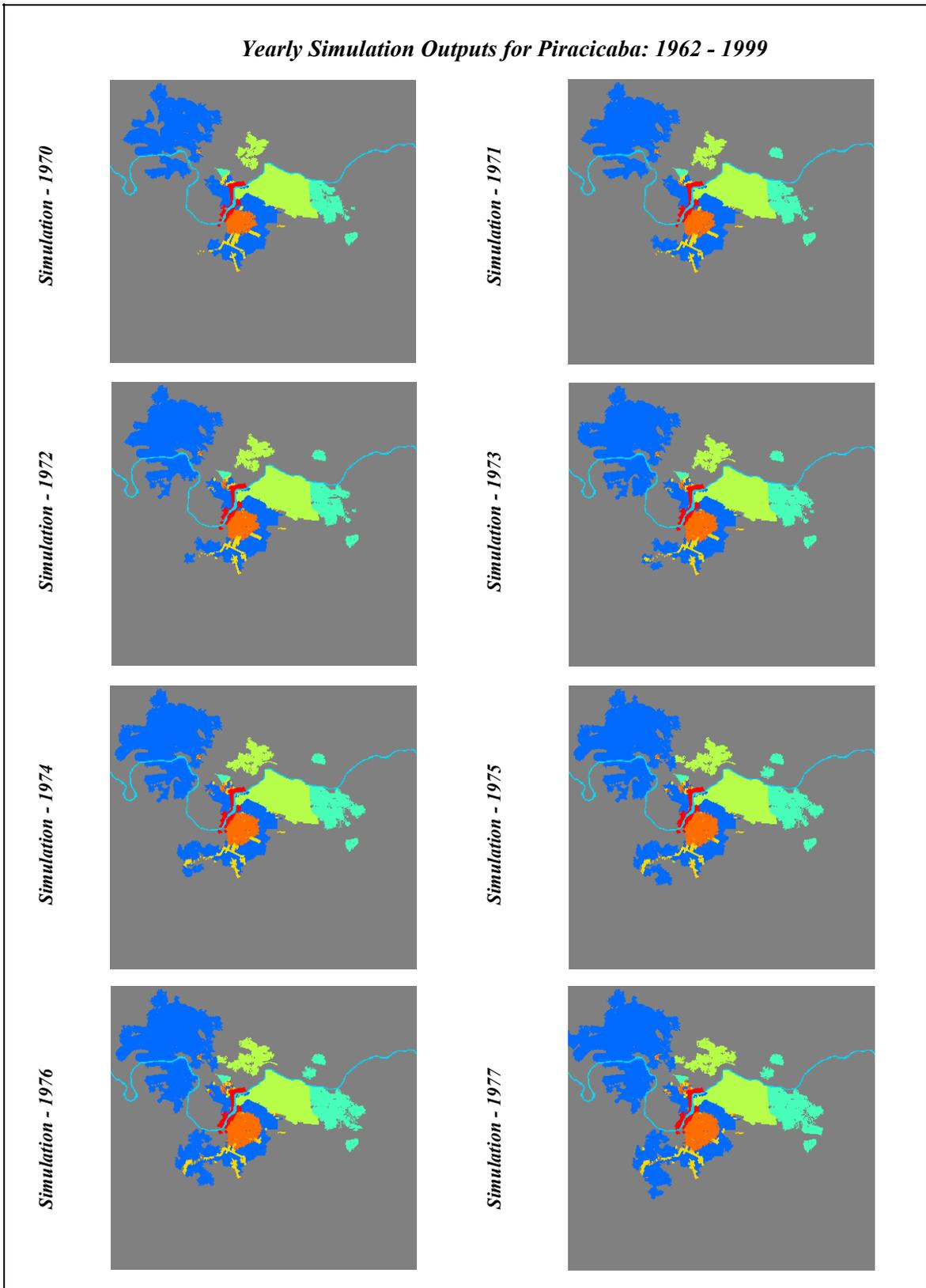


FIGURE 6.53b – Yearly simulation outputs for Piracicaba: 1970 - 1977.

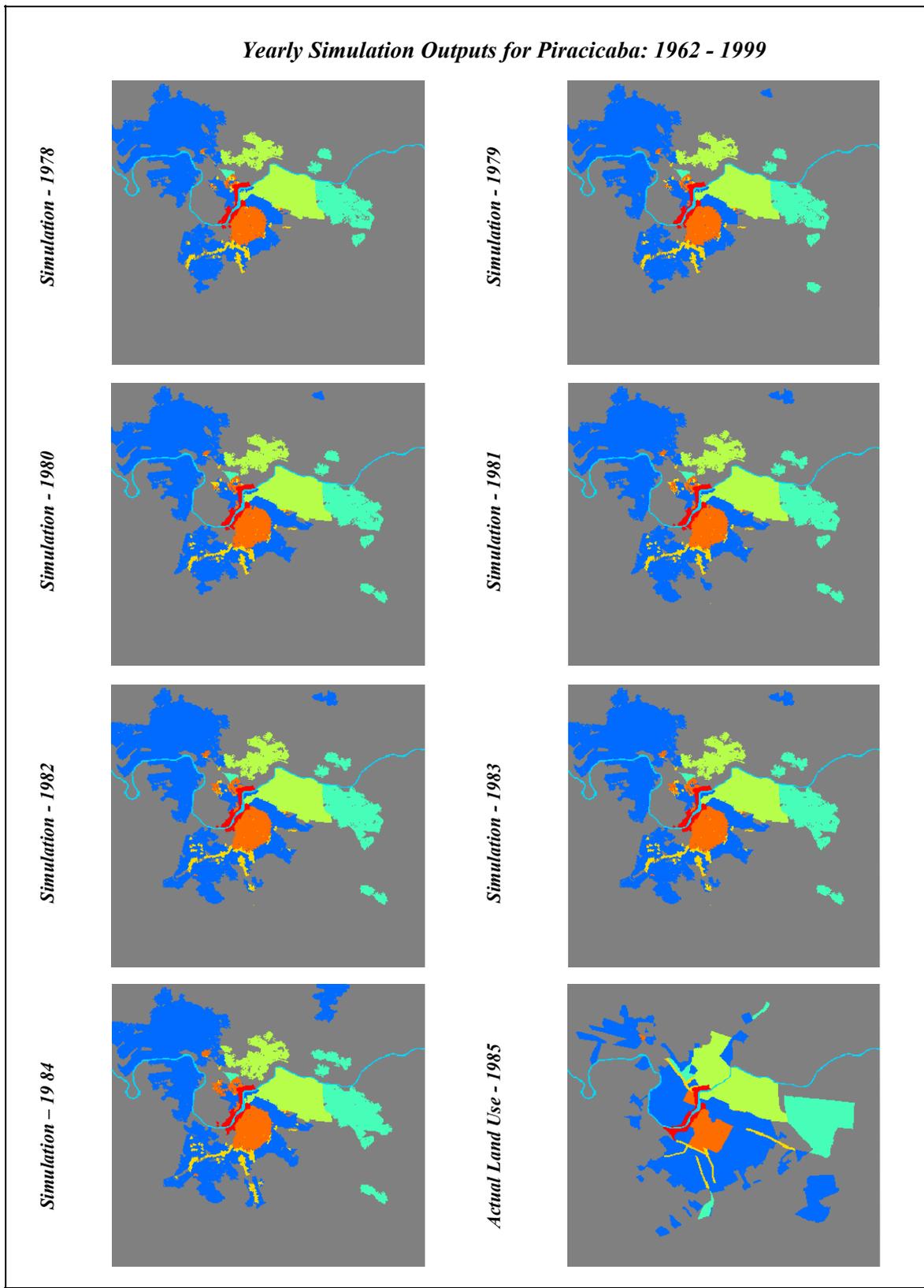


FIGURE 6.27c – Yearly simulation outputs for Piracicaba: 1983 - 1990.

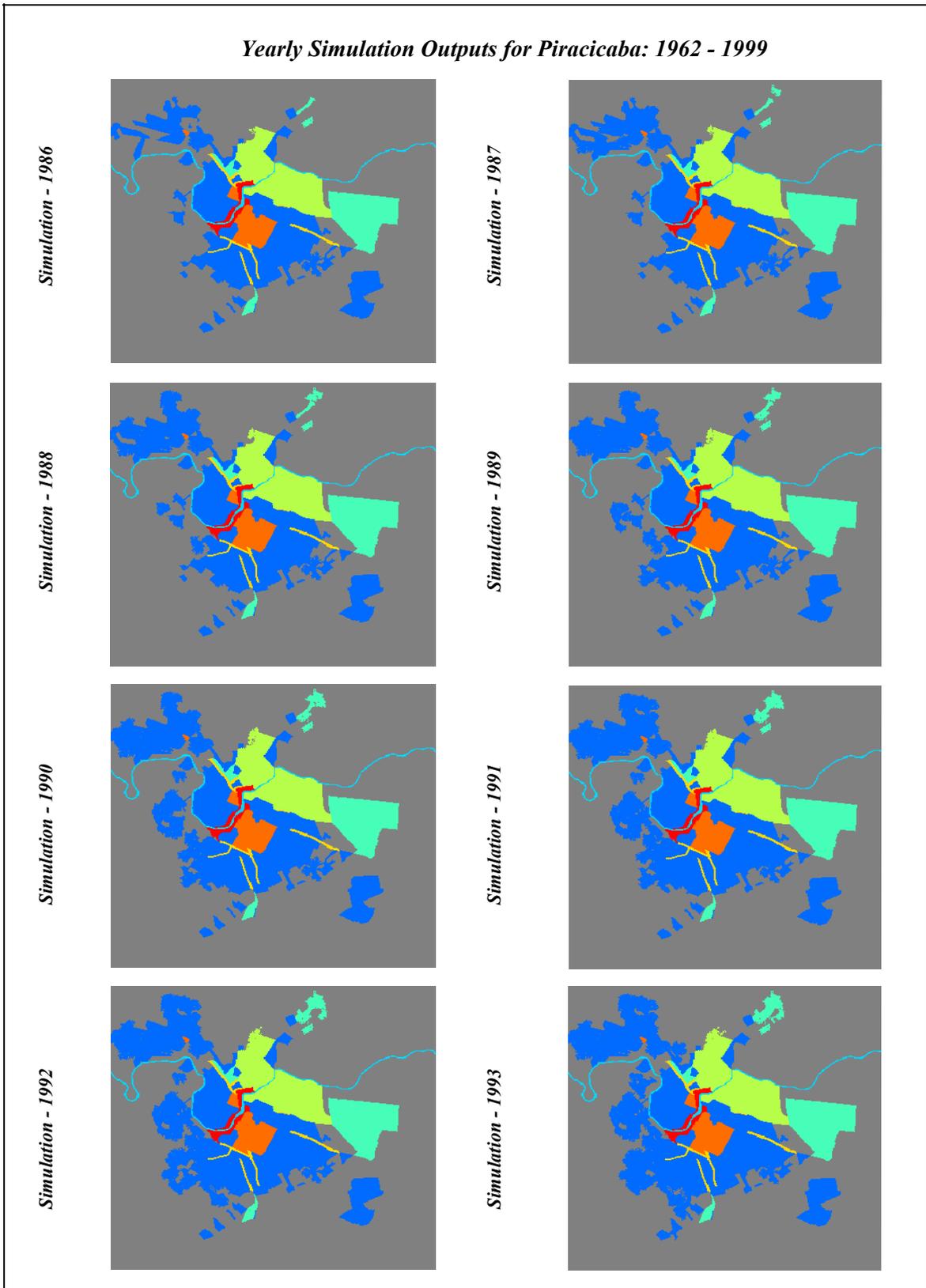
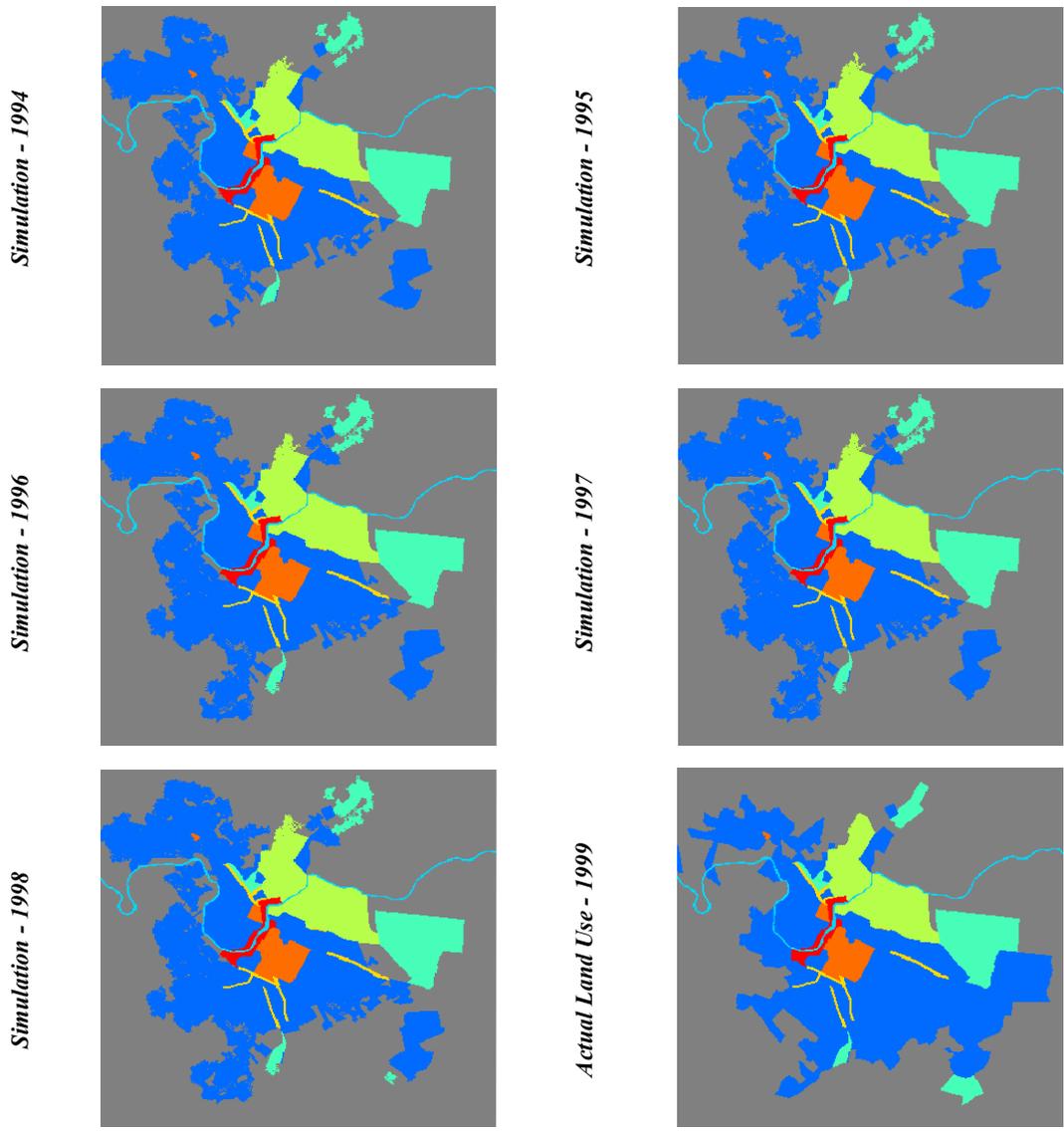


FIGURE 6.53d – Yearly simulation outputs for Piracicaba: 1986 - 1993.

Yearly Simulation Outputs for Piracicaba: 1962 - 1999



- | | | | |
|-----------------|----------------|-------------------|--------------------|
| Non-urban use | Commercial use | Institutional use | Water bodies |
| Residential use | Industrial use | Services use | Leis./recreat. use |

FIGURE 6.53e – Yearly simulation outputs for Piracicaba: 1994 - 1999.

6.2.4 Short-Term Forecasts: 2000 – 2004

The land use changes considered for forecasts were those observed in the last simulation period (1985-1999), excluding the transitions “non-urban to institutional (nu_inst)” and “non-urban to leisure (nu_leis)”, for they are deemed to be stabilized, and hence, not occurring in subsequent years. In the next sections, the transition matrices as well as simulation outputs for stationary (Markovian) and non-stationary forecasts of land use change will be presented.

6.2.4.1 Stationary Forecasts

A stationary transition matrix for the year 2004 was obtained using the Markov chain (Equation 5.66) upon basis of the global transition matrix 1985-1999, in which the transitions “nu_inst” and “nu_leis” were purposely set to zero. These stationary transition probabilities are shown in TABLE 6.62.

TABLE 6.62 – Matrix of stationary transition probabilities for Piracicaba: 2000-2004.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Water B.</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.9367 \cdot 10^{-1}$	$4.4801 \cdot 10^{-3}$	0	$1.8538 \cdot 10^{-3}$	0	0	0	0
<i>Residential</i>	0	1.00	0	0	0	0	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Water B.</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

6.2.4.2 Non-Stationary Forecasts

Non-stationary forecasts of land use change have been built with the aid of linear regression models relating the area of certain land uses with demographic data and economic indicators, as previously mentioned in Section 5.3.2.6, where the destination uses of land use transitions were regarded as the dependent variable.

As in the case of Bauru, the destination land uses had to be estimated for the quinquennia where the data on population and economic performance (see Section 4.2.2) were also available, i.e. for 1970, 1975, 1980, 1985, 1990, 1995 and 2000. This has been enabled by the yearly simulation outputs generated by means of the principal components method (Section 6.2.3). The simulation outputs produced inside DINAMICA were exported in TIFF format, and then imported in IDRISI. Inside IDRISI, the areas of the thematic classes corresponding to the land uses of interest were assessed. The areas of these destination land uses together with the demographic and economic data used in the linear regression models are seen in TABLE 6.63.

TABLE 6.63 – Areas of destination land uses, urban population, total and sectorial GDPs (US\$): Piracicaba – 1970-2000.

<i>Years</i>	<i>Dest. Area: Residential (ha)</i>	<i>Dest. Area: Industrial (ha)</i>	<i>Urban Population</i>	<i>Total GDP (US\$ 1998)</i>	<i>Rural GDP (US\$ 1998)</i>	<i>Indust. GDP (US\$ 1998)</i>	<i>Comm. GDP (US\$ 1998)</i>	<i>Services GDP (US\$ 1998)</i>
1970	8,579	1,487	73,153	666,029.934	60,056.683	314,733.774	95,101.568	291,239.467
1975	11,329	2,304	115,960	1,312,284.818	39,437.798	853,930.964	121,743.828	418,916.047
1980	13,912	3,086	158,708	2,126,207.943	66,581.222	1,307,157.079	147,927.333	752,469.633
1985	14,714	3,906	198,407	1,704,037.322	68,131.878	1,075,191.585	128,818.571	560,713.867
1990	19,982	4,297	218,590	2,424,143.275	96,842.165	1,615,544.478	289,475.481	432,548.144
1995	24,923	4,662	236,687	3,251,245.119	24,907.799	2,108,284.969	491,681.220	277,346.390
2000	28,667	4,958	317,374	3,335,315.985	17,102.944	2,170,367.463	494,371.845	492,749.832

SOURCE: Adapted from IPEA (2001, 2003a, 2003b) and FUNDAÇÃO SEADE (2002).

6.2.4.2.1 “Non-Urban to Residential Use (nu_res)” Linear Regression Model

Likewise the linear regression models of Bauru, the tests for verifying the independence of observations regarding the outcome variable Y_i (“residential use area in Piracicaba – destarea”) showed partial acceptance for the autocorrelation function and total acceptance for the partial autocorrelation function (FIGURE 6.54). The observations of residential use area throughout time were judged to be independent events for the same reasons exposed in Section 6.1.5.2.1.

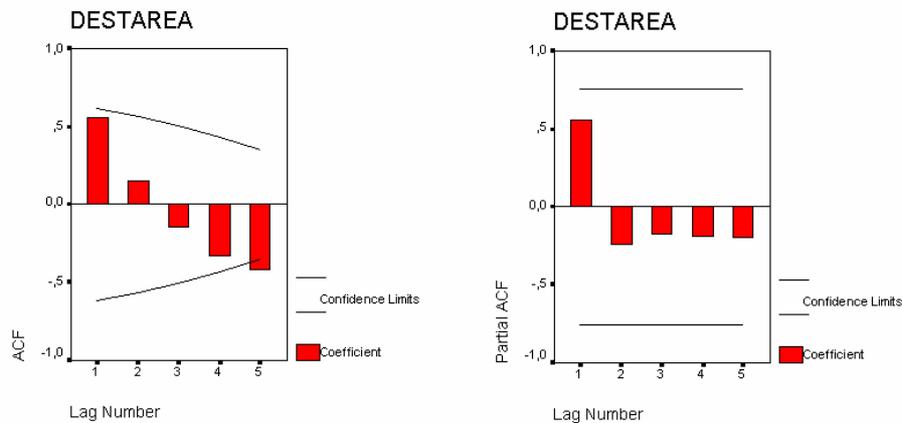


FIGURE 6.54 – ACF and partial ACF tests for the residential use area (destarea).

Still in the exploratory analysis, the independent variables “urban population (urbpop)”, “total GDP (totgdp)”, “industrial GDP (indgdp)” and “commercial GDP (comgdp)” show a high correlation with the response variable “residential use area (destarea)”. Since these four independent variables are highly correlated amongst themselves (TABLE 6.64), only the variable “total GDP” remains in the final regression model. The scatter plots concerning the correlation matrix for the “nu_res” model are seen in FIGURE 6.55.

The final equation for this univariate model is the following:

$$Y = \beta_1 X$$

$$Y = 0.008081 \cdot \text{totgdp} \quad , \quad (6.4)$$

whose R^2 is 0.989, and the p -value of the 95% confidence interval for β_1 is 0.000 ($p < 0.05$). Since β_0 did not pass the significance test, it was removed from the model.

In the analysis of variance (ANOVA), the sum of squares reflects to which extent Y is divergent from \hat{Y} . In this sense, for well fit models it is expected that the regression sum of squares \gggg residual sum of squares, and therefore that the regression sum of squares accounts for the majority of the total sum of squares (TABLE 6.65).

TABLE 6.64 – Correlation matrix for the “nu_res” model: Piracicaba, 2000-2004.

		DESTAREA	URBPOP	TOTGDP	RURALGDP	INDGDP	COMGDP	SERVGDP
DESTAREA	Pearson Correlation	1	,959**	,965**	-,465	,966**	,968**	-,087
	Sig. (2-tailed)		,001	,000	,293	,000	,000	,853
	N	7	7	7	7	7	7	7
URBPOP	Pearson Correlation	,959**	1	,926**	-,351	,930**	,859*	,117
	Sig. (2-tailed)	,001		,003	,440	,002	,013	,803
	N	7	7	7	7	7	7	7
TOTGDP	Pearson Correlation	,965**	,926**	1	-,412	,998**	,928**	,066
	Sig. (2-tailed)	,000	,003		,358	,000	,003	,889
	N	7	7	7	7	7	7	7
RURALGDP	Pearson Correlation	-,465	-,351	-,412	1	-,396	-,559	,278
	Sig. (2-tailed)	,293	,440	,358		,380	,192	,546
	N	7	7	7	7	7	7	7
INDGDP	Pearson Correlation	,966**	,930**	,998**	-,396	1	,925**	,054
	Sig. (2-tailed)	,000	,002	,000	,380		,003	,908
	N	7	7	7	7	7	7	7
COMGDP	Pearson Correlation	,968**	,859*	,928**	-,559	,925**	1	-,277
	Sig. (2-tailed)	,000	,013	,003	,192	,003		,547
	N	7	7	7	7	7	7	7
SERVGDP	Pearson Correlation	-,087	,117	,066	,278	,054	-,277	1
	Sig. (2-tailed)	,853	,803	,889	,546	,908	,547	
	N	7	7	7	7	7	7	7

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

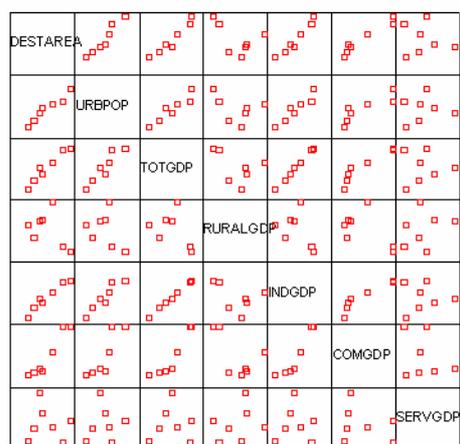


FIGURE 6.55 – Correlation matrix scatter plots for the “nu_res” model: Piracicaba, 2000-2004.

TABLE 6.65 – Analysis of variance for the “nu_res” model: Piracicaba, 2000-2004.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3,02E+08	1	302045078,1	68,038	,000 ^a
	Residual	22196909	5	4439381,863		
	Total	3,24E+08	6			

a. Predictors: (Constant), TOTGDP

b. Dependent Variable: DESTAREA

Concerning the analysis of residuals, when the model is satisfactorily fit, the standardized residuals (see Section 5.3.2.5) and the studentized residuals (adjusted to a t-distribution) must lie within the interval $[-2,+2]$. For the current model, these values are $-1.535; +1.500$ and $-1.638; +1.509$ respectively for the standardized and studentized residuals. Plots of the standardized residuals versus the independent variable (“totgdp”) and versus the adjusted predicted value \hat{Y}_i are seen in FIGURE 6.56. As stated in Section 5.3.2.5, these plots should not present any kind of correlation pattern.

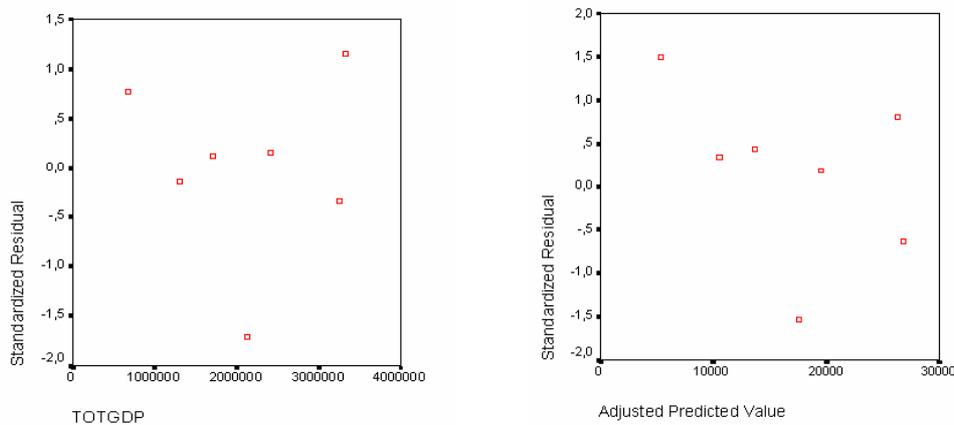


FIGURE 6.56 – Analysis of residuals for the “nu_res” model: Piracicaba, 2000-2004.

6.2.4.2.2 “Non-Urban to Industrial Use (nu_ind)” Linear Regression Model

In the same manner as the “nu_res” model, the tests for verifying the independence of observations regarding the outcome variable Y_i (“industrial use area in Piracicaba – destarea”) showed partial acceptance for the autocorrelation function and total acceptance for the partial autocorrelation function (FIGURE 6.57). The observations of industrial use area throughout time were deemed to be independent events for the same reasons exposed in Section 6.1.5.2.1.

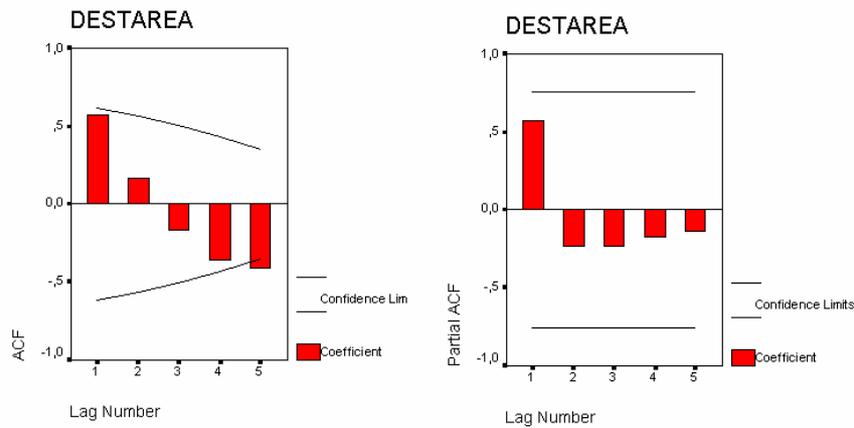


FIGURE 6.57 – ACF and partial ACF tests for the industrial use area (destarea).

The independent variables “urban population (urbpop)”, “total GDP (totgdp)” and “industrial GDP (indgdp)” show a high correlation with the response variable “industrial use area (destarea)”. Since these three independent variables are highly correlated amongst themselves (TABLE 6.66), only the variable “industrial GDP” remains in the final regression model. Although its correlation with the response variable is not the highest, it was judged to be the one which best explains the increases in the industrial area in the city in the latest three decades. The scatter plots concerning the correlation matrix for the “nu_ind” model are presented in FIGURE 6.58.

The final equation for this univariate model is the following:

$$Y = \beta_0 + \beta_1 X$$

$$Y = 1,105.895 + 0.001795 \cdot indgdp \quad , \quad (6.5)$$

whose R^2 is 0.882, and the p -value of the 95% confidence interval for β_0 is 0.053 and for β_1 is 0.002 ($p < 0.05$).

In the analysis of variance (ANOVA), the sum of squares reflects to which extent Y is divergent from \hat{Y} . In this sense, for well fit models it is expected that the regression sum of squares \gg residual sum of squares, and therefore that the regression sum of squares accounts for the majority of the total sum of squares (TABLE 6.67).

TABLE 6.66 – Correlation matrix for the “nu_ind” model: Piracicaba, 2000-2004.

		DESTAREA	URBPOP	TOTGDP	RURALGDP	INDGDP	COMGDP	SERVGDP
DESTAREA	Pearson Correlation	1	,965**	,930**	-,221	,939**	,831*	,097
	Sig. (2-tailed)		,000	,002	,633	,002	,021	,837
	N	7	7	7	7	7	7	7
URBPOP	Pearson Correlation	,965**	1	,926**	-,351	,930**	,859*	,117
	Sig. (2-tailed)	,000		,003	,440	,002	,013	,803
	N	7	7	7	7	7	7	7
TOTGDP	Pearson Correlation	,930**	,926**	1	-,412	,998**	,928**	,066
	Sig. (2-tailed)	,002	,003		,358	,000	,003	,889
	N	7	7	7	7	7	7	7
RURALGDP	Pearson Correlation	-,221	-,351	-,412	1	-,396	-,559	,278
	Sig. (2-tailed)	,633	,440	,358		,380	,192	,546
	N	7	7	7	7	7	7	7
INDGDP	Pearson Correlation	,939**	,930**	,998**	-,396	1	,925**	,054
	Sig. (2-tailed)	,002	,002	,000	,380		,003	,908
	N	7	7	7	7	7	7	7
COMGDP	Pearson Correlation	,831*	,859*	,928**	-,559	,925**	1	-,277
	Sig. (2-tailed)	,021	,013	,003	,192	,003		,547
	N	7	7	7	7	7	7	7
SERVGDP	Pearson Correlation	,097	,117	,066	,278	,054	-,277	1
	Sig. (2-tailed)	,837	,803	,889	,546	,908	,547	
	N	7	7	7	7	7	7	7

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

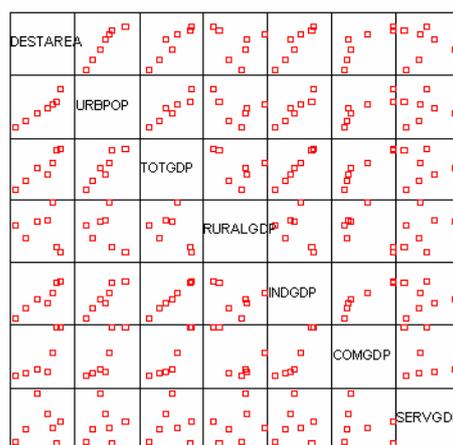


FIGURE 6.58 – Correlation matrix scatter plots for the “nu_ind” model: Piracicaba, 2000-2004.

TABLE 6.67 – Analysis of variance for the “nu_ind” model: Piracicaba, 2000-2004.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8748416	1	8748415,915	37,199	,002 ^a
	Residual	1175904	5	235180,760		
	Total	9924320	6			

a. Predictors: (Constant), INDGDP

b. Dependent Variable: DESTAREA

As to the analysis of residuals, when the model is satisfactorily fit, the standardized residuals (see Section 5.3.2.5) and the studentized residuals (adjusted to a t-distribution) must lie within the interval $[-2,+2]$. For the current model, these values are -0.757 ; $+1.793$ and -0.817 ; $+1.969$ respectively for the standardized and studentized residuals. Plots of the standardized residuals versus the independent variable (“indgdp”) and versus the adjusted predicted value \hat{Y}_i are seen in FIGURE 6.59. According to what was stated in Section 5.3.2.5, these plots should not contain any kind of correlation pattern.

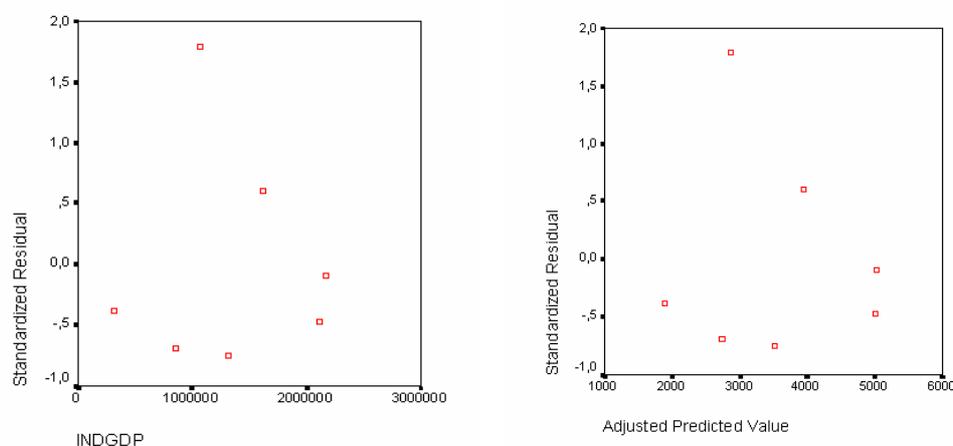


FIGURE 6.59 – Analysis of residuals for the “nu_ind” model: Piracicaba, 2000-2004.

Likewise Bauru, heuristic projections were made for the values of the independent variables concerning economic indicators of Piracicaba upon basis of their past trends.

Two types of non-stationary forecast scenarios of land use change were formulated: an optimistic and a pessimist one, with respectively slight over and underestimations of the independent variable, as previously defined in Section 5.3.3. In the first case, the value of X concerning total GDP was US\$ 3,623,190.199 for the “nu_res” model and US\$ 2,320,949.861 regarding industrial GDP for the “nu_ind” model. In the pessimist scenario, X assumed the values of US\$ 3,574,805.098 for the “nu_res” model and US\$ 2,210,211.052 concerning industrial GDP for the “nu_ind” model. Using these values of X , transition probabilities were calculated for both scenarios in the short-term (TABLES 6.68 and 6.69), employing the conversion equations presented in Section 5.3.2.6.

TABLE 6.68 – Matrix of optimistic transition probabilities for Piracicaba: 2000-2004.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Water B.</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.8808*10^{-1}$	$7.8788*10^{-3}$	0	$4.0443*10^{-3}$	0	0	0	0
<i>Residential</i>	0	1.00	0	0	0	0	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Water B.</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

TABLE 6.69 – Matrix of pessimist transition probabilities for Piracicaba: 2000-2004.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Water B.</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.9567*10^{-1}$	$2.8451*10^{-3}$	0	$1.4805*10^{-3}$	0	0	0	0
<i>Residential</i>	0	1.00	0	0	0	0	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Water B.</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

6.2.4.3 Forecasts Simulation Outputs

Taking into account the same sets of independent variables selected for the last simulation period (1985-1999) and their respective weights of evidence, indicated in TABLE 6.57, stationary, optimistic and pessimist simulation outputs were generated for Piracicaba in the short-term (2004), using TABLES 6.62, 6.68 and 6.69 to respectively set the total amount of land use change. These simulation outputs are presented in FIGURE 6.60.

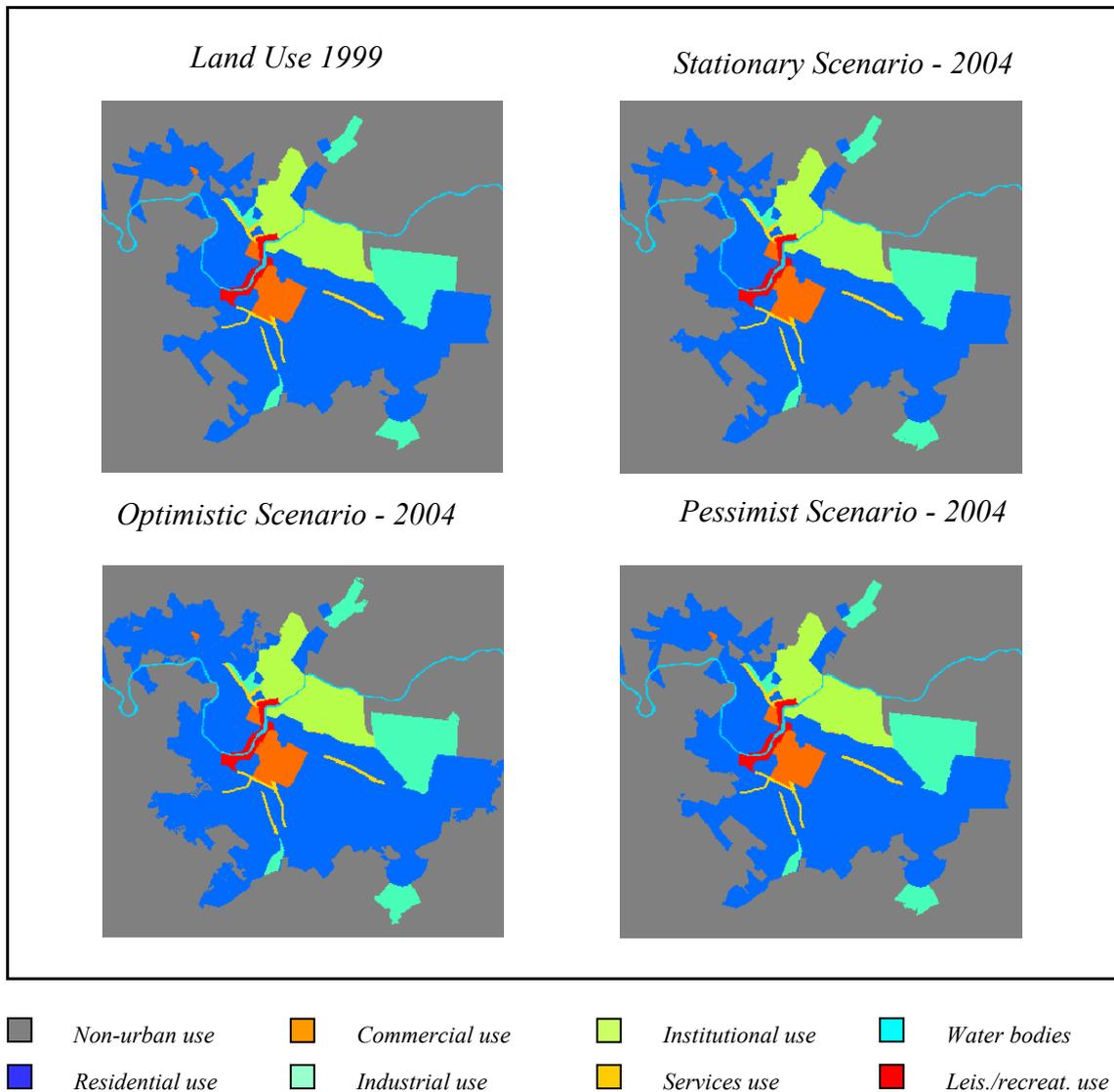


FIGURE 6.60 – Stationary, optimistic and pessimist simulations for 2004 compared to the actual land use in Piracicaba in 1999.

It is observable from the above forecast simulations that there is no increase in the services areas. This does not imply that there will be no economic growth in the services sector in the near term, but simply that the areas sheltering services are likely to go through densification processes.

The main areas where new residential settlements arise are located in the northern and western sectors of Piracicaba, whereas the increase in industrial area takes place mostly

in the southern district. This expansion pattern complies with trends observed in a recent official city map of Piracicaba, issued in 2003 (SEMUPLAN, 2003).

As previously explained in Sections 5.3.3 and 6.1.5.3, just smooth variations in the projections of population and economic data were introduced. This is due to the fact that, in view of the administrative continuity trend demonstrated by the current federal government as well as of steadfastly decreasing population growth rates, the demographic and macroeconomic scenarios of Brazil in the latest years are expected to reproduce themselves in the current decade.

6.2.5 Medium-Term Forecasts: 2000 – 2007

6.2.5.1 Stationary Forecasts

A stationary transition matrix for the year 2007 was obtained using the Markov chain (Equation 5.66) upon basis of the global transition matrix 1985-1999, in which the transitions “nu_inst” and “nu_leis” were purposely set to zero. These stationary transition probabilities are shown in TABLE 6.70.

TABLE 6.70 – Matrix of stationary transition probabilities for Piracicaba: 2000-2007.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Water B.</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.7237 \cdot 10^{-1}$	$2.4872 \cdot 10^{-2}$	0	$2.7550 \cdot 10^{-3}$	0	0	0	0
<i>Residential</i>	0	1.00	0	0	0	0	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Water B.</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

6.2.5.2 Non-Stationary Forecasts

The same linear regression models used for short-term predictions have been used in the medium-term forecasts. For the optimistic scenario, the value of X concerning total GDP was US\$ 3,824,481.866 for the “nu_res” model and US\$ 2,426,242.340 regarding

industrial GDP for the “nu_ind” model. In the pessimist scenario, X assumed the values of US\$ 3,751,365.720 for the “nu_res” model and US\$ 2,238,071.080 concerning industrial GDP for the “nu_ind” model. Using these values of X , transition probabilities were calculated for both scenarios in the medium-term (TABLES 6.71 and 6.72), employing the conversion equations presented in Section 5.3.2.6.

TABLE 6.71 – Matrix of optimistic transition probabilities for Piracicaba: 2000-2007.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Water B.</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.6471 \cdot 10^{-1}$	$2.8824 \cdot 10^{-2}$	0	$6.4708 \cdot 10^{-3}$	0	0	0	0
<i>Residential</i>	0	1.00	0	0	0	0	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Water B.</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

TABLE 6.72 – Matrix of pessimist transition probabilities for Piracicaba: 2000-2007.

<i>Land Use</i>	<i>Non-Urban</i>	<i>Residential</i>	<i>Commercial</i>	<i>Industrial</i>	<i>Institutional</i>	<i>Services</i>	<i>Water B.</i>	<i>Leis./Recr.</i>
<i>Non-Urban</i>	$9.7666 \cdot 10^{-1}$	$2.1216 \cdot 10^{-2}$	0	$2.1242 \cdot 10^{-3}$	0	0	0	0
<i>Residential</i>	0	1.00	0	0	0	0	0	0
<i>Commercial</i>	0	0	1.00	0	0	0	0	0
<i>Industrial</i>	0	0	0	1.00	0	0	0	0
<i>Institutional</i>	0	0	0	0	1.00	0	0	0
<i>Services</i>	0	0	0	0	0	1.00	0	0
<i>Water B.</i>	0	0	0	0	0	0	1.00	0
<i>Leis./Recr.</i>	0	0	0	0	0	0	0	1.00

6.2.5.3 Forecasts Simulation Outputs

Taking into account the same sets of independent variables selected for the last simulation period (1985-1999) and their respective weights of evidence (TABLE 6.57), stationary, optimistic and pessimist simulation outputs were generated for Piracicaba in the medium-term (2007), using the TABLES 6.70, 6.71 and 6.72 to respectively set the

total amount of land use change. These simulation outputs are presented in FIGURE 6.61.

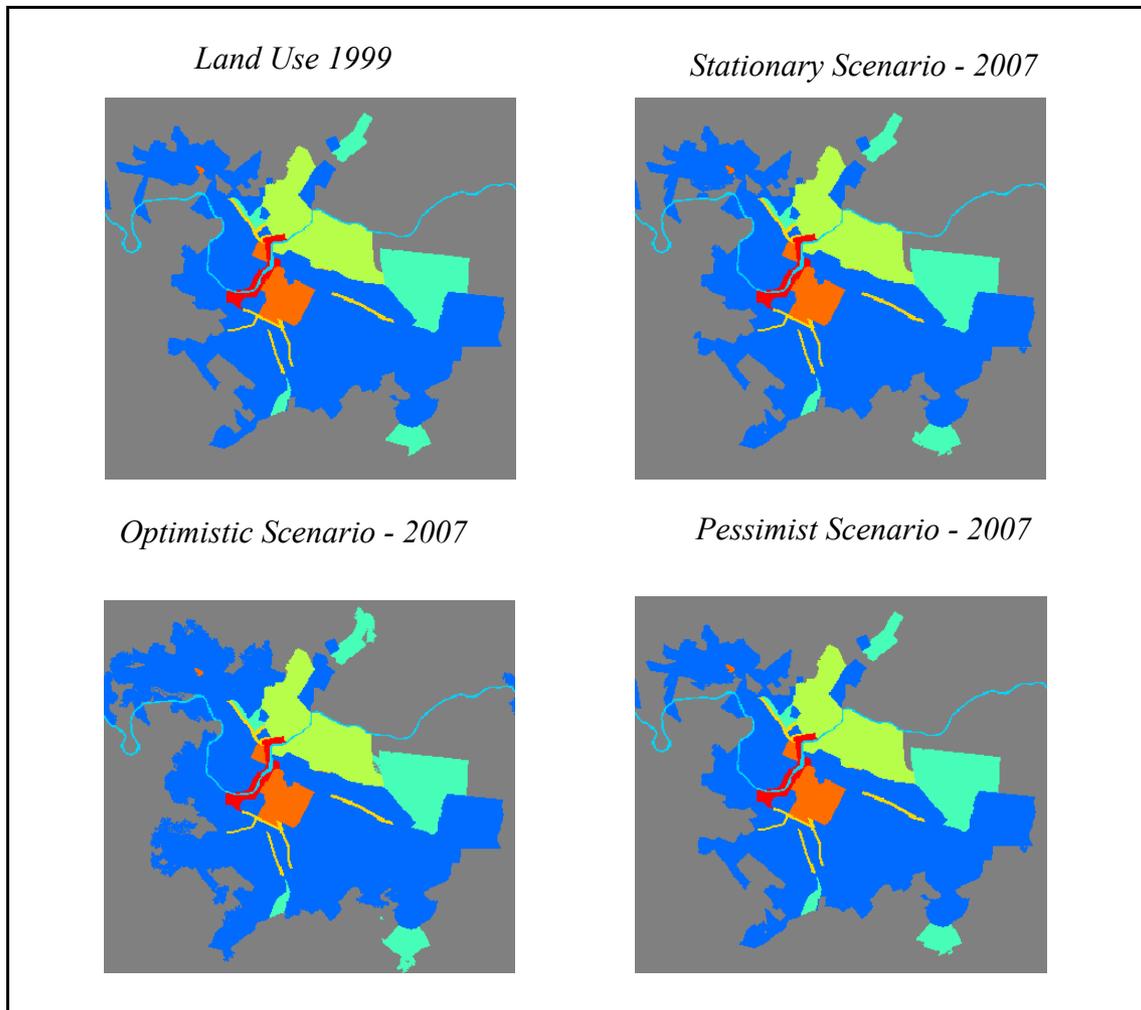


FIGURE 6.61 – Stationary, optimistic and pessimist simulations for 2007 compared to the actual land use in Piracicaba in 1999.

6.3 Conclusions

For both cities under analysis, the driving forces determining land use change remained practically constant throughout the time series, with slight differences from one simulation period to another. Upon basis of the generated simulation outputs, it was possible to infer that new residential settlements tend to occur in the vicinities of already existent settlements and in areas with good accessibility conditions. Industrial zones are likely to expand surrounding former industrial districts and/or close to main roads. Services corridors usually arise along main roads and close to residential and commercial areas. Institutional zones follow the same growth logic of industrial zones, and not rarely are located closed to the latter, in view of the need of large and cheap plots for their expansion.

Mixed use zones, on their turn, require the availability of good accessibility as well as the existence of higher population densities, what ensures a feasible consumers market, and hence, economic sustainability. And finally, leisure and recreation zones take place in regions with good accessibility, close to residential areas and usually along low and flat riverbanks, since these areas are floodable and therefore unsuitable for sheltering other urban uses.

It becomes thus evident that the land use transitions comply with economic theories of urban growth and change, grounded on the key concept of utility maximization (Black and Henderson, 1999; Medda et al. 1999; Papageorgiou and Pines, 2001; Zhou and Vertinsky, 2001), in which there is a continuous search for optimal location, able to assure:

- competitive real estate prices;
- good accessibility conditions;
- rationalization of transportation costs, and
- a strategic location in relation to suppliers and consumers markets.

In both case studies, Bauru and Piracicaba, the natural environment (soil, vegetation, relief, conservation areas) has not been regarded as a decisive driver for land use change. In other words, natural characteristics of the physical environment, excluding the Piracicaba river in the city of Piracicaba, have not been considered as impedances to urban growth at a more generalized level. The cities sites are relatively flat, with mild slopes, and present no outstanding condition regarding soil, vegetation and conservation areas constraints.

Although the current work does not concern agent-based simulation experiments, all the information regarding the behavior of social actors and shapers of urban form is implicitly aggregated in the models input variables. That is to say that the preferences and options adopted by local administrators, entrepreneurs and real estate landlords and investors in the decision-making processes affecting the urban structure are reflected in the way infrastructural and socioeconomic factors guide urban land use change.

As previously mentioned in Section 5.6, the simulation outputs provided by the weights of evidence and the logistic regression methods presented comparatively similar results (Section 6.1.2). In any case, the weights of evidence should be given preference of application in view of its transparency and operational simplicity, what together concur for a faster and more consistent model calibration.

Considering the simulation results produced for both cities, it is observable that in longer simulation periods, the modeling performance is reduced. This is due to the fact that a greater number of transitions take place in the meantime, and even for the same type of land use change, drivers may suffer slight alterations throughout time, with which longer simulation periods cannot cope.

As to the forecast scenarios, it can be acknowledged that the stationary ones overestimate current trends of land use change. This can be ascribed to the fact that both cities witnessed sharp decreases in urban growth rates in the present decade, and the transition probabilities for the short- and medium-terms (respectively 2004 and 2007) have been established upon basis of growth trends experienced in the late 1980s and 1990s, when the experiments regarding the last simulation periods were carried out.

It is worth restating here that as a means to conceive realistic optimistic and pessimist predictions, only slight variations were introduced in the projections of population and economic data meant to parameterize non-stationary forecasts of land use change. This can be justified by the fact that the demographic and macroeconomic scenarios of Brazil in the latest years are expected to reproduce themselves in the current decade, in face of steadfastly decreasing population growth rates as well as of the administrative continuity trend demonstrated by the current federal government.

And finally, some remarks concerning the DINAMICA transition algorithms are worthy of mention. Shifts from non-urban areas to residential use represented the most challenging category of land use transition in the modeling experiments of the present research. The reasons for the difficulties in detecting their shapes have been already noted in Section 6.1.1 but it should be highlighted here that 65% of this type of transition occur through the *expander* algorithm. An evident shortcoming of this algorithm which is being tackled at present by the CSR-UFGM team lies in the fact that after the random selection of a seed cell for transition, all neighboring cells are subject to transition regardless of their transition probability values, and this is too blunt an instrument to accurately mirror the prioritization of development in real situations (Almeida et al. 2003).

Another inherent drawback of the transition algorithms refers to their sequence arrangements. For instance, supposing that a certain non-urban cell presents a transition probability for the industrial use of 0.95 , and of 0.89 for the residential use. In the case that the transition “non-urban to residential use” is set to be executed first, this cell will be converted to residential use, even though its suitability to shelter the industrial use is higher. Devices to overcome this limitation ought to be developed. Further improvements to enhance the performance of such algorithms, like the incorporation of fractal parameters in the transition functions as well as the possibility to separately define patches average sizes and variances for the *expander* and *patcher* should be also envisaged.

Albeit all the aforementioned technical and methodological difficulties, the land use change simulation experiments undertaken for Bauru and Piracicaba proved to be consistent both in terms of “goodness-of-fit” tests and of adherence to the logic of economic theories on urban growth and development. This confirms the first research hypothesis.

The input data – digital maps, remote sensing images and digital aerial photos – used in the modeling experiments were able to provide information on spatial variables driving land use change in the cities under analysis, what confirms the second hypothesis. The employment of the weights of evidence and logistic regression methods to assess the relative contribution of these variables for the land use transitions observed in the two cities, and the thereof generated maps of transition probability surfaces confirm the third and fourth hypotheses. The comparative effectiveness of these methods has been presented in Section 5.6.

And lastly, the knowledge on transition trends gained with the simulation experiments carried out for long time series (of approximately thirty-five years) was crucial for conceiving forecast scenarios of land use change. The adoption of the Markov chain to generate stationary scenarios, and of linear regression models to develop non-stationary scenarios of future land use transitions validates the fifth and last hypothesis.

CHAPTER 7

FINAL REMARKS AND CONCLUSIONS

7.1 Dealing with Spatial Dynamic Models

It is unquestionable the advance brought by the incorporation of space representation in land use change models as well as by the evolution of static into dynamic models (see Section 2.2.3), particularly with respect to urban land use change models.

Departing from the elementary early generation of urban models, experiments in this field attempted to progressively add the spatio-temporal dimension in a wise and consistent manner. The first spatial models could not have their outputs spatially visualized, and the spatial aspect was approached in the form of spatial units (zones, census districts, etc.) designed for numerical information processing (Batty, 1971, 1976; Allen et al. 1981, 1986). The evolution of static into dynamic models was not a one-step process, and static models (Clark, 1951; Alonso, 1960; Lowry, 1964) firstly yielded place to recursive models (Butler, 1969), comparative statics (Perraton and Baxter, 1974) and quasi-dynamic models (Boyce, 1977; Coelho and Williams, 1978; Leonardi, 1981) before being converted into fully dynamic models (Beaumont et al. 1981; Wegener et al. 1986).

Effective advances in the spatial representation of urban models occurred only by the end of the 1980s, when cellular automata (CA) models started to be extensively applied (Coclelis, 1985, 1987; Deadman et al. 1993; Batty and Xie, 1997; Benati, 1997; Clarke et al. 1998; Papini et al. 1998; White et al. 1998; Portugali et al. 1999; White and Engelen, 2000). GIS, whose raster data can be easily associated with CA models, offered new possibilities in spatial modeling, including spatial data analysis tools.

The usage of the weights of evidence statistical method to parameterize the urban land use change simulation experiments of the present research could have not been able without the aid of GIS, which are crucial to provide the contingency tables figures

derived from cross-tabulation operations (Bonham-Carter, 1994). In the same way, the pre-processing of the input variables maps, like maps of distances, Kernel points density estimator and other procedures were accomplished using GIS. Even the empirical logistic regression technique, which is not a spatial method in principle, was successfully applied in the simulation experiments due to facilities available in the GIS environment.

Although all the above-mentioned advances represented by the advent of CA models and the facilities resulting from their coupling with GIS, one cannot deny the limitations and shortcomings implicit in current traditional implementations of CA. They mainly concern premises of space and time discretization, like space regularity, neighborhood stationarity and universality of transition functions. Further limitations ought to be mentioned, like the artificial order imposed to the different types of land use transition (when they actually take place simultaneously in reality), the generalizations applied to the land use maps in order to render them computationally workable, and lastly, the current standards of computational processing capacity which directly limit the spatial resolution of CA models. Most efforts to tackle such constraints still lie in the theoretical realm (see Section 2.3.2.1).

Yet spatial dynamic models, of which CA is one of the best representatives, are still the most promising means for rendering land use change simulation outputs communicable and transparent to politicians, planners and decision-makers in particular, and to the lay public in general.

7.2 Recent Advances and Potential New Paths for Models of Urban Land Use Change

As already stated in Section 2.3.2.2, successive refinements have been introduced to urban CA models in the latest decade, mainly derived from related advances in the fields of Artificial Intelligence (AI) and complex systems. Examples in this sense are the works based on contemporary pattern-fitting tools such as neural nets (Wu, 1998; Xia and Yeh, 2000) and evolutionary learning (Papini et al. 1998). Still along with this line of research, studies dealing with topics in complexity like chaos, fractals and self-

organized criticality are worthy of mention (Couclelis, 1988; Batty and Longley, 1991; White and Engelen, 1993; Batty and Longley, 1994; Portugali et al. 1997; Portugali et al. 1999; Sobreira and Gomes, 2001; Batty, 2003).

As exposed previously in Section 2.3.2.1, urban CA-based modeling moves towards a relaxation of any or all assumptions embedded in standard cellular automata. In this respect, instead of homogeneous cells, space could be subdivided into irregular structures (polygons) representing more realistic administrative or political divisions, like census districts, communes, origin-destination zones, etc. Neighborhoods could vary in view of the functionality associated with them, and transition functions could be selectively applied considering their effective extent context. For Couclelis (1997), variable time steps could be used to fit some external schedule, like seasonal variations in the growth rate of a certain type of land use.

A new thematic trend in urban CA models concerns the focus shift from issues related to the urban framework expansion to more subtle transformations held in intra-urban structures, like densification processes (Rabino, 2002; Yeh and Li, 2002). This represents an attempt of modelers and planners to cope with more pressing matters in the urban environment. In fact, there is an ongoing decline drift in the growth rates of cities nearly everywhere, and urban growth is bound to become a zero-sum game (Prud'Homme, 1989).

Urban growth is considerably decreasing in developed countries, but this decline has started to be experienced in many developing countries. The implications of this change are numerous and will certainly affect the dynamics of real estate markets, and consequently, the nature of urban policies. According to Prud'Homme (1989), competition between cities will probably be enhanced, and “each urban area will have to fight to retain people and jobs, and to be innovative and attractive to suit that purpose.” For him, even in the developing world, the management of urban areas, in contrast to planning, will gain in importance, and the scarcity of resources will drive policy-makers to search for flexible, low-cost, innovative, alternative, and temporary answers to the problems raised by urban growth.

7.3 Main Contributions of this Research

Technological innovations as well as scientific advance in a general way occur in a gradual form, step by step. Individual contributions in this sense are often not structural, but of secondary character, which when assembled all together produce meaningful increments to the theoretical and technological bodies of Science in the long run.

It can be stated that the original idea of CA was innovative, but its application for urban modeling purposes represents a mere secondary scientific contribution, since the branch of urban studies simply imported a theoretical resource and its technological counterpart already available in another field of knowledge.

In the context of CA modeling, there is a sort of library of scientific-technological tools, some of which were originally conceived for CA, and some others not. The latest contributions within the scope of urban CA models concern basically new manners of assembling such devices in inedited ways into the modeling process. For instance, fuzzy sets theory, AHP, neural nets and evolutionary learning are all innovative theoretical and methodological approaches proceeding mostly from the Computer Sciences, and all of them have been recently embedded into CA models dealing with urban topics (Wu, 1996, 1998; Xia and Yeh, 2000; Papini et al. 1998).

Along with this line of action, this research is mainly committed to explore the dynamics of land use change in two real medium-sized cities by employing two different empirical statistical methods for the estimation of cells transition probabilities: the weights of evidence and logistic regression. Considering the fact that the models handle a detailed categorization of urban land uses (eight categories of land use were taken into account for the simulation experiments), this work aims at opening up new methodological paths for further investigations in alike lines of research.

Urban land use change simulation experiments dealing with real cities and diverse types of land uses have been conducted by White and Engelen (1997, 2000) and White et al. (1998), but their parameterization is based on empirical multivariate linear regression methods. While their forecasts are generated upon exploratory techniques (by answering

“what if” questions), the forecasts in the present research are based upon linear regression models calibrated upon long time series. Similar simulation experiments handling with real cities and several land uses have also been carried out by Papini et al. (1998), but in opposition to the methods employed in the present thesis, their transition rules are set upon basis of evolutionary learning algorithms and Boolean logic.

In brief, the current work is entrusted with providing methodological guidance for future scientific investigations that deal with data-driven modeling experiments applied to real case studies, in which a detailed categorization of land uses is at issue.

7.4 Possible Applicability of this Work

The urban land use dynamics models demonstrate to be useful for the identification of main urban growth vectors and their general land use tendencies, what enables local planning authorities to manage and reorganize (if it comes into question) city growth according to the environmental carrying capacity of concerned sites and to their present and envisaged infrastructure availability.

The urban expansion forecasts provided by such models also help local authorities in general, like sub-majors, districts administrators and municipal ministers, to establish investments goals in terms of technical and social infrastructure equipments, such as the extension of roads, the enlargement of the water supply and sewerage catchment areas, the creation of new bus lines, the construction of kindergartens, schools, hospitals and health centers, etc.

Decision makers from the private sphere can as well benefit from the modeling output data, since companies of transportation, conventional and mobile phones, cable TV and internet, and others will have subsidies for defining priorities as to where and how intense to invest.

Also the organized civil society, either through NGOs or local associations, can profit from the modeling forecasts in order to enhance, by legal means, social movements demanding the implementation of social and technical infrastructure, since their

requests and respective arguments shall be based on realistic short- and medium-term urban growth trends.

And as final words, Batty (1976) concisely exposes the key ideas lying behind the applications and purposes of urban modeling when he says:

“There are many reasons for the development of such models: their role in helping scientists to understand urban phenomena through analysis and experiment represents a traditional goal of science, yet urban modelling is equally important in helping planners, politicians and the community to predict, prescribe and invent the urban future (Batty, 1976, p. xx).”

BIBLIOGRAPHIC REFERENCES

- Albin, P. S. **The analysis of complex socioeconomic systems**. Lexington, MA: D. C. Heath, 1975.
- Allen , P. M.; Engelen, G.; Sanglier, M. Towards a general dynamic model of the spatial evolution of urban systems. In: Hutchinson, B.; Batty, M. ed. **Advances in urban systems modelling**. Amsterdam: Elsevier, 1986. p. 199-220.
- Allen , P. M.; Sanglier, M; Boon, F. Intra-urban evolution - Part II. In: Kahn, D. ed. **Models of urban settlement and structure as dynamic self-organizing systems**. Washington, D. C.: US Department of Transportation, 1981.
- Almeida, C. M.; Batty, M.; Monteiro, A. M. V.; Câmara, G.; Soares- Filho, B. S.; Cerqueira, G. C.; Pennachin, C. L. Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation. **Computers, Environment and Urban Systems**, v. 27, n. 5, p. 481-509, 2003.
- Alonso, W. A general theory of the urban land market. **Papers and Proceedings of the Regional Science Association**, v. 6, p. 149-158, 1960.
- Bailey, T. C.; Gatrell, A. C. **Interactive spatial data analysis**. Essex, UK: Longman, 1996. 413p.
- Bak, P.; Chen, K.; Creutz, M. Self-organized criticality in the 'Game of Life'. **Nature**, v. 342, p. 780-782, 1989.
- Baker, W. L. A review of models of landscape change. **Landscape Ecology**, v. 2, n. 2, p. 111-133, 1989.
- Batty, M. Modeling cities as dynamic systems. **Nature**, v. 231, p. 425-428, June 1971.
- Batty, M. **Urban modelling: algorithms, calibrations, predictions**. Cambridge, UK: Cambridge University Press, 1976. 381 p.
- Batty, M. GeoComputation using cellular automata. In: Openshaw, S.; Abrahart, R. J. ed. **Geocomputation**. New York: Taylor & Francis, 2000. Chap. 5, p. 95-126.
- Batty, M. **The emergence of cities: complexity and urban dynamics**. 2003. Available at < http://www.casa.ucl.ac.uk/working_papers/paper64.pdf>. Accessed on 07/12/03.
- Batty, M.; Couclelis, H.; Eichen, M. Urban systems as cellular automata. **Environment and Planning B**, v. 24, n. 2, p. 159-164, 1997.
- Batty, M.; Jiang, B.; Thurstain-Goodwin, M. **Local movement: agent-based models of pedestrian flow**. 1998. Available at <http://www.casa.ucl.ac.uk/local_movement.doc>. Accessed on 08/03/02.
- Batty, M.; Longley, P. Cities as fractals: simulating growth and form. In: Crilly, T.; Earnshaw, R. A.; Jones, H. ed. **Fractals and chaos**. New York: Springer Verlag, 1991. p. 41-69.

- Batty, M.; Longley, P. **Fractal cities: a geometry of form and function**. London: Academic Press, 1994.
- Batty, M.; Xie, Y. Preliminary evidence for a theory of the fractal city. **Environment and Planning A**, v. 28, p. 1745-1762, 1996.
- Batty, M.; Xie, Y. Possible urban automata. **Environment and Planning B**, v. 24, n. 2, p. 175-192, 1997.
- Batty, M.; Xie, Y.; Sun, Z. Modeling urban dynamics through GIS-based cellular automata. **Computers, Environment and Urban Systems**, v.23, p. 205-233, 1999.
- Beaumont, J. R.; Clarke, M.; Wilson, A. G. The dynamics of urban spatial structure: some explanatory results using difference equations and bifurcation theory. **Environment and Planning A**, v. 13, p. 1473-1483, 1981.
- Bell, E. J.; Hinojosa, R. C. Markov analysis of land use change: continuous time and stationary processes. **Socio-Economic Planning Science**, v. 8, p. 13-17, 1977.
- Benati, S. A cellular automaton for the simulation of competitive location. **Environment and Planning B**, v. 24, n. 2, p. 205-218, 1997.
- Benenson, I; Torrens, P. M. Object-based modeling of urban phenomena: editorial. **Computers, Environment and Urban Systems**, 2003 (in press).
- Benevolo, L. **História da cidade**. São Paulo: Editora Perspectiva, 1983. 729 p.
- Berlekamp, E. R.; Conway, J. H.; Guy, R. K. **Winning ways for your mathematical plays**. New York: Academic Press, 1982.
- Bertalanffy, L. von. An outline of general systems theory. **British Journal of the Philosophy of Science**, v. 1, p. 134-165, 1951.
- Black, D.; Henderson, V. A theory of urban growth. **The Journal of Political Economy**, v. 107, n. 2, p. 252-284, 1999.
- Blokland, J.; Hendricks, A. J.; Paelinck, J. H. P. **Elements for the construction of a simple model for a stagnating town**. Amsterdam: Regional Science Association - Dutch Meeting, 1972.
- Bonham-Carter, G. F. **Geographic Information Systems for Geoscientists: Modelling with GIS**. Ontario: Pergamon, 1994. 305 p.
- Bowker, D. E.; Davis, R. E.; Myrick, D. L.; Stacy, K.; Jones, W. T. **Spectral reflectances of natural targets for use in remote sensing**. Hampton: National Aeronautics and Space Administration - NASA, 1985. (NASA Reference Publication, n. 1139).
- Boyce, D. E. **Equilibrium solutions to combined urban residential location, modal choice, and trip assignment models**. Philadelphia: Regional Science Research Institute, 1977. (RSRI Discussion Paper, n. 98).
- Brasil. **Constituição Federal** (CF. Título VII. Cap. II. art. 182. § 1º). 1988. Available at <<http://www.senado.gov.br/bdtextual/const88/Con1988br.pdf>>. Accessed on 07/18/03.

- Briassoulis, H. **Analysis of land use change: theoretical and modeling approaches.** Lesvos, Greece. PhD Thesis in Geography - University of Aegean, 2000. Available at <<http://www.rri.wvu.edu/WebBook/Briassoulis/contents.htm>>. Accessed on 10/28/02.
- Burrough, P. A. Dynamic modelling and geocomputation. In: Longley, P. A.; Brooks, S. M.; McDonnell, R.; MacMillan, B. ed. **Geocomputation: a primer.** Chichester: John Wiley & Sons, 1998. Chap. 9, p. 165-192.
- Butler, E. W. **Moving behavior and residential choice.** Washington, D.C.: Highway Research Board, 1969. (Report n. 81).
- Câmara, G.; Monteiro, A. M. V.; Medeiros, J. S. Representações computacionais do espaço: fundamentos epistemológicos da ciência da geoinformação. **Geografia**, v. 28, n.1, p. 83-96, 2003.
- Câmara, G.; Souza R.C.M.; Freitas, U. M.; Garrido, J. C. P. SPRING: Integrating remote sensing and GIS with object-oriented data modelling. **Computers and Graphics**, v.15, n.6, p. 13-22, 1996.
- Censo demográfico – São Paulo – 1970. Rio de Janeiro: IBGE, 1971. v. 1. 478 p.
- Censo demográfico – Dados distritais – São Paulo – 1980. Rio de Janeiro: IBGE, 1982. v. 1. 611 p.
- Censo demográfico – 2000 – Características da população e dos domicílios – Resultados do universo. Rio de Janeiro: IBGE, 2000. 519p.
- Centro de Informática do Campus “Luiz de Queiroz” – Universidade de São Paulo (CIAGRI-USP). 2003. **Piracicaba.** Available at <<http://www.ciagri.usp.br/piracica/pirap.htm>>. Accessed on 08/02/03.
- Centro de Pesquisa e Estudos Urbanísticos – Faculdade de Arquitetura e Urbanismo da Universidade de São Paulo (CEPEU-FAUUSP). **Plano diretor de Bauru – cidade de Bauru: rede de água.** Bauru: CEPEU-FAUUSP, 1967a. One map. Scale 1:20,000. Paper plan.
- Centro de Pesquisa e Estudos Urbanísticos – Faculdade de Arquitetura e Urbanismo da Universidade de São Paulo (CEPEU-FAUUSP). **Plano diretor de Bauru – cidade de Bauru: usos de solo.** Bauru: CEPEU-FAUUSP, 1967b. One map. Scale 1:20,000. Paper plan.
- Centro de Pesquisa e Estudos Urbanísticos – Faculdade de Arquitetura e Urbanismo da Universidade de São Paulo (CEPEU-FAUUSP). **Plano diretor de Bauru – cidade de Bauru: pavimentação.** Bauru: CEPEU-FAUUSP, 1967c. One map. Scale 1:20,000. Paper plan.
- Chadwick, G. **Una vision sistematica del planeamiento.** Barcelona: Editorial Gustavo Gili, 1973. 283 p.
- Chapin Jr., F. S.; Kaiser, E. J. **Urban land use planning.** Urbana: University of Illinois Press, 1979.
- Chapin Jr., F. S.; Weiss, S. F. **Factors influencing land development.** Chapel Hill: Institute for Research in Social Science - University of North Carolina, 1962.

- Chapin Jr., F. S.; Weiss, S. F. A probabilistic model for residential growth. **Transportation Research**, v. 2, p. 375-390, 1968.
- Clark, C. Urban population densities. **Journal of the Royal Statistical Society A**, v. 114, p. 490-496, 1951.
- Clarke, K. C.; Gaydos, L. J. Loose-coupling a cellular automaton model and GIS: long-term urban growth predictions for San Francisco and Baltimore. **International Journal of Geographic Information Science**, v. 12, p. 699-714, 1998.
- Clarke, K. C.; Hoppen, S.; Gaydos, L. J. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. **Environment and Planning B**, v. 24, n. 2, p. 247-261, 1997.
- Coelho, J. D.; Williams, H. C. W. L. On the design of land use plans through locational surplus maximization. **Papers of the Regional Science Association**, v. 40, p. 71-85, 1978.
- Consórcio Intermunicipal Tietê-Paraná (CITP) and Centro de Estudos e Pesquisas de Administração Municipal (CEPAM). **Tietê-Paraná - Projeto Calha**. São Paulo: Unida Artes Gráficas e Editora Ltda., 1994. 67 p.
- Constanza, R. Model goodness of fit: a multiple resolution procedure. **Ecological Modelling**, v. 47, p. 199-215, 1989.
- Costa, P.; Piasentin, U. **Un modello di simulazione dello sviluppo urbano di Venezia**. Venice: Symposium on Trends in Mathematical Modelling – UNESCO, 1971.
- Couclelis, H. Cellular worlds: a framework for modeling micro-macro dynamics. **Environment and Planning A**, v. 17, p. 585-596, 1985.
- Couclelis, H. Cellular dynamics: how individual decisions lead to global urban change. **European Journal of Operational Research**, v. 30, p. 347-355, 1987.
- Couclelis, H. Of mice and men: what rodent populations can teach us about complex spatial dynamics. **Environment and Planning A**, v. 20, p. 99-109, 1988.
- Couclelis, H. From cellular automata to urban models: new principles for model development and implementation. **Environment and Planning B**, v. 24, n. 2, p. 165-174, 1997.
- Crecine, J. P. **TOMM (Time oriented metropolitan model)**. Pittsburgh: Department of City Planning, 1964. (CRP Technical Bulletin, n. 6).
- Crecine, J. P. **A dynamic model of urban structure**. Santa Monica: RAND Corporation, 1968. (P-3803).
- Czamanski, S. A method of forecasting metropolitan growth by means of distributed lags analysis. **Journal of Regional Science**, v. 6, p. 35-49, 1965.
- D'Alge, J. C. L. Qualidade geométrica de imagens TM do LANDSAT-5. In: **XIII Congresso Brasileiro de Cartografia**. Brasília: SBC, 1987. INPE-8090-PRE/090.

Data ITE – Bauru. Bauru: Instituição Toledo de Ensino – Faculdade de Ciências Econômicas de Bauru (ITE – FCEB), 1998. 205p.

Deadman, P.; Brown, R. D.; Gimblett, P. Modelling rural residential settlement patterns with cellular automata. **Journal of Environmental Management**, v. 37, p. 147-160, 1993.

Departamento de Água e Esgoto da Prefeitura Municipal de Bauru (DAE – Bauru). **Comércio.** Bauru: DAE, 1979a. One map. Scale 1:20,000. Paper plan.

Departamento de Água e Esgoto da Prefeitura Municipal de Bauru (DAE – Bauru). **Conjuntos habitacionais.** Bauru: DAE, 1979b. One map. Scale 1:20,000. Paper plan.

Departamento de Água e Esgoto da Prefeitura Municipal de Bauru (DAE – Bauru). **Principais vias de acesso.** Bauru: DAE, 1979c. One map. Scale 1:20,000. Paper plan.

Departamento de Água e Esgoto da Prefeitura Municipal de Bauru (DAE – Bauru). **Rede de água.** Bauru: DAE, 1979d. One map. Scale 1:20,000. Paper plan.

Departamento de Água e Esgoto da Prefeitura Municipal de Bauru (DAE – Bauru). **Planta geral do município de Bauru - SP.** Bauru: DAE, 2000. One map. Scale 1:20,000. Paper plan.

Departamento de Água e Esgoto da Prefeitura Municipal de Bauru (DAE – Bauru). **Planta geral do município de Bauru - SP.** Bauru: DAE, 2003. One map. Scale 1:20,000. Paper plan.

Departamento de Solos e Nutrição de Plantas – Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo (LSN – ESALQ-USP). **Digital aerial photos mosaic for the city of Piracicaba – 1962.** Piracicaba: LSN – ESALQ-USP, 2002. Digital photos. 23.0 MB. Tiff. Scale 1:100,000.

Diplock, G. Genetic programming: a new approach to spatial model building. In: Openshaw, S.; Abraham, R. J. ed. **Geocomputation.** New York: Taylor & Francis, 2000. Chap. 9, p. 219-252.

Easterly, W.; Sewadeh, M. **Global development network growth database.** 2000. Available at <<http://www.worldbank.org/research/growth/GDNdata.htm>>. Accessed on 07/15/03.

Echenique, M. **Models:** a discussion. Cambridge, UK: University of Cambridge, 1968. (Land Use and Built Form Studies, WP 6).

Empresa Paulista de Planejamento Metropolitano S.A. (EMPLASA). **Regiões metropolitanas do Brasil.** 2000. Available at <http://www.emplasa.sp.gov.br/metropoles/rms_brasil/RM-BRASIL-LEGISLACAO.htm>. Accessed on 08/02/03.

Engelen, G.; White, R.; Uljee, I. Integrating constrained cellular automata models, GIS and decision support tools for urban planning and policy making. In: Timmermans, H. P. J. ed. **Decision support systems in urban planning.** London: E&FN Spon, 1997. Chap. 8, p. 125-155.

Estimativas da população residente em 1º de julho de 1985, segundo os municípios e a situação político-administrativa. Rio de Janeiro: IBGE, 1987. Valid from 03/16/87.

Estimativas da população residente em 1º de julho de 1990, segundo os municípios e a situação político-administrativa. Rio de Janeiro: IBGE, 1991.

Estimativas das populações residentes em 1º de julho de 1992 a 1º de julho de 1995, segundo as situações político-administrativas vigentes em 1º de julho de 1992 a 1995 e segundo os municípios. Rio de Janeiro: IBGE, 1997. Resolução n. 27. Valid from 08/27/96.

Estimativas da população residente em 1º de julho de 1999, segundo os municípios e a situação político-administrativa. Rio de Janeiro: IBGE, 1999.

Facelli, J. M.; Steward, T. A. P. Markovian chains and the role of history in succession. *Tree*, v.5, n. 1, p. 27-29, 1990.

FAO/IIASA. **Agro-ecological assessments for national planning: the example of Kenya**. Rome: FAO, 1993. (FAO Soils Bul, n. 67).

Fernandes, C. L. L.; Medeiros, C. M.; Mendes, A. G. O elemento regional no processo de planejamento do Brasil - Notas preliminares. In: Cintra, A. O.; Haddad, P. R. ed. **Dilemas do planejamento urbano e regional no Brasil**. Rio de Janeiro: Zahar Editores, 1977. Cap. II, p. 51-168.

Ferreira, A. B. H. **Novo dicionário da língua portuguesa**. Rio de Janeiro: Nova Fronteira, 1986. 1838 p.

Fischer, M. M.; Abrahart, R. J. Neurocomputing – tools for geographers. In: Openshaw, S.; Abrahart, R. J. ed. **Geocomputation**. New York: Taylor & Francis, 2000. Chap. 8, p. 187-217.

Fischer, T. **Gestão contemporânea: cidades estratégicas e organizações locais**. Rio de Janeiro: Editora da Fundação Getúlio Vargas, 1996. 208 p.

Forrester, J. W. **Urban dynamics**. Cambridge, Mass: MIT Press, 1969.

Forrester, J. W. **World dynamics**. Cambridge, Mass: Wright-Allen Press, 1971.

Fundação Seade. **Hidrovia Tietê-Paraná: um rio de negócios**. 1999. Available at <<http://www.seade.gov.br/negocios/snpor01b.html>>. Accessed on 09/23/00.

Fundação Seade. **Bauru: a cidade sem limites e a capital da terra branca**. 2000a. Available at <http://www.seade.gov.br/negocios/snpci_05-v2.html>. Accessed on 08/01/03.

Fundação Seade. **Piracicaba: a terra da cachaça boa**. 2000b. Available at <http://www.seade.gov.br/negocios/snpci_23-v2.html>. Accessed on 08/01/03.

Fundação Seade. **Informações dos municípios paulistas: valor adicionado fiscal total e setorial para Bauru e Piracicaba – 1993 - 2001**. 2002. Available at <http://www.seade.gov.br/cgi-bin/lingcv98/spd_01.ksh>. Accessed on 06/10/03.

Fundação Seade. **Informações dos municípios paulistas: demografia**. 2003. Available at <http://www.seade.gov.br/cgi-bin/lingcv98/spd_01.ksh>. Accessed on 08/10/03.

- Gardner, M. The fantastic combinations of John Conway's new solitaire game 'Life'. **Scientific American**, v. 223, p. 120-123, 1970.
- Gong, P.; Howarth, P. J. The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. **Photogrammetric Engineering & Remote Sensing**, v. 56, n. 1, p. 67-76, 1990.
- Goodacre, A. K.; Bonham-Carter, G. F.; Agterberg, F. P.; Wright, D. F. A statistical analysis of the spatial association of seismicity with drainage patterns and magnetic anomalies in western Quebec. **Tectonophysics**, v.217, p. 285-305, 1993.
- Haack, B.; Bryant, N.; Adams, S. An assessment of Landsat MSS and TM data for urban digital classification. **Remote Sensing of Environment**, v. 21, p. 201-213, 1987.
- Hardoy, J. E.; Mitlin, D.; Satterthwaite, D. **Environmental problems in an urbanizing world: finding solutions for cities in Africa, Asia and Latin America**. London: Earthscan Publications Ltd, 2001. 446 p.
- Harris, B. Bridging the gap between theory and practice in urban planning. In: Pack, J. ed. **Proceedings of the Conference on the Use of Models by Planning Agencies**. Philadelphia: University of Pennsylvania, 1975.
- Hidrovia Tietê-Paraná: hidrovia do Mercosul – Um rio de negócios**. São Paulo: CESP, 1998. 26 p.
- Hill, D. M. A growth allocation model for the Boston region. **Journal of the American Institute for Planners**, v. 31, p. 111-120, 1965.
- Hoover, E. M.; Giarratani, F. **An introduction to regional economics**. New York: Alfred K., 1984.
- Hosmer, D.; Lemeshow, S. **Applied logistic regression**. New York: John Wiley & Sons, Inc., 1989. 307p.
- Instituto Brasileiro de Geografia e Estatística (IBGE). **Carta topográfica do Brasil - Piracicaba**. Rio de Janeiro: IBGE, 1969. One map. Projection system: UTM. Horizontal datum: Córrego Alegre, MG. Vertical datum: marégrafo Imbituba, SC. Scale 1:50,000. Paper plan.
- Instituto Brasileiro de Geografia e Estatística (IBGE). **Carta topográfica do Brasil - Bauru**. Rio de Janeiro: IBGE, 1973. One map. Projection system: UTM. Horizontal datum: Córrego Alegre, MG. Vertical datum: marégrafo Imbituba, SC. Scale 1:50,000. Paper plan.
- Instituto Nacional de Pesquisas Espaciais (INPE). **Landsat 3 - MSS satellite image – WRS 237/75 – Acquisition date 06-22-79**. São José dos Campos: INPE, 1979. One satellite image. SOL: EL28 A2045 BB 190 LL N. Printed in paper.
- Instituto Nacional de Pesquisas Espaciais (INPE). **Landsat 5 - TM satellite image – WRS 221/75 – Acquisition date 10-01-84**. São José dos Campos: INPE, 1984. One satellite image. 384 MB. GeoTiff. DGI-LPE-0271. CD-ROM.

Instituto Nacional de Pesquisas Espaciais (INPE). **Landsat 5 - TM satellite image – WRS 220/76 – Acquisition date 08-10-85**. São José dos Campos: INPE, 1985. One satellite image. 383 MB. GeoTiff. DGI-LPE-0111. CD-ROM.

Instituto Nacional de Pesquisas Espaciais (INPE). **Landsat 5 - TM satellite image – WRS 221/75 – Acquisition date 11-29-88**. São José dos Campos: INPE, 1988. One satellite image. 384 MB. GeoTiff. DGI-LPE-0271. CD-ROM.

Instituto Nacional de Pesquisas Espaciais (INPE). **Landsat 5 - TM satellite image – WRS 220/76 – Acquisition date 07-16-99**. São José dos Campos: INPE, 1999. One satellite image. 235 MB. DAT. CCB-CCT-0002. CD-ROM.

Instituto Nacional de Pesquisas Espaciais (INPE). **Landsat 5 - TM satellite image – WRS 221/75 – Acquisition date 06-07-00**. São José dos Campos: INPE, 2000. One satellite image. 368 MB. GeoTiff. DGI-ERG-0208. CD-ROM.

Instituto de Pesquisa Econômica Aplicada (IPEA). **PIBs dos municípios brasileiros: 1970, 1975, 1980, 1985, 1990, 1996 (US\$ of 1998)**. 2001. Available at <<http://www.ipea.gov.br/pub/PIBsMu/pibs.htm>>. Accessed on 06/10/03.

Instituto de Pesquisa Econômica Aplicada (IPEA). **IPCA – geral (%) a.m. 2003a**. Available at <<http://www.ipeadata.gov.br/>>. Accessed on 06/10/03.

Instituto de Pesquisa Econômica Aplicada (IPEA). **PIBs dos municípios brasileiros: 1970, 1975, 1980, 1985 (R\$ of 2000)**. 2003b. Available at <<http://www.ipeadata.gov.br/>>. Accessed on 06/10/03.

Jensen-Moeller, L. Knowledge-based classification of an urban area using texture and context information in Landsat TM imagery. **Photogrammetric Engineering & Remote Sensing**, v. 56, n.6, p. 899-904, 1990.

Johnston, A. K.; Watters, T. R. Assessing spatial growth of the Washington metropolitan area using thematic mapper data. In: American Congress on Surveying and Mapping and American Society for Photogrammetry and Remote Sensing. **Proceedings**. Baltimore: ACSM - ASPRS, 1996.

Johnston, R. J.; Gregory, D.; Smith, D. M. **The dictionary of human geography**. Oxford: Blackwell, 1994.

Joint Research Centre – European Commission - Institute for Remote Sensing Applications (JRC); European Space Agency – Earthnet Programme Office (ESA-ESRIN). **Modelling Deforestation Processes – a review**. Luxembourg: ECSC-EC-EAEC, 1994. (Trees Series B, Research Report n.1).

Jones, A.; Clark, J. Driving forces behind European land use change: an overview. In: The user needs for more harmonized land use information at the national and EU level. **Report on the CLAUDE (Coordinating Land Use and Cover Data Analyses in Europe) Workshop**. Wageningen: CLAUDE, Concerted Action, EC, 1997. v. DGXII, p. 24-31.

Lakshmanan, T. R. An approach to the analysis of intra-urban location applied to the Baltimore region. **Economic Geography**, v. 40, p. 348-370, 1964.

- Lakshmanan, T. R. A model for allocating urban activities in a state. **Socio-Economic Planning Sciences**, v. 1, p. 283-295, 1968.
- Lakshmanan, T. R.; Hansen, W. G. A retail market potential model. **Journal of the American Institute of Planners**, v. 31, p. 134-143, 1965.
- Lathrop, G. T.; Hamburg, J. R. An opportunity-accessibility model allocating regional growth. **Journal of the American Institute of Planners**, v. 31, p. 95-103, 1965.
- Leonardi, G. A general accessibility and congestion-sensitive multi-activity spatial interaction model. **Papers of the Regional Science Association**, v. 47, p. 3-17, 1981.
- Li, X.; Yeh, A. G. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. **International Journal of Geographical Information Science**, v. 14, n. 2, p. 131-152, 2000.
- Liu, H. **Spectral signatures and multi-spectral image interpretation**. 2003. Available at <<http://geog.tamu.edu/~liu/courses/g489/note8.pdf>>. Accessed on 08/10/03.
- Lowry, I. S. **A model of metropolis**. Santa Monica: RAND Corporation, 1964. (RM-4035-RC).
- Machado e Silva, A. J. F. **Modelos de correção geométrica para imagens HRV-SPOT**. São José dos Campos. 223p. Published MSc Dissertation – Instituto Nacional de Pesquisas Espaciais - INPE, 1989. INPE-5001-TDL/395.
- Machado e Silva, A. J. F.; D'Alge, J. C. L. Avaliação da qualidade geométrica das imagens TM-LANDSAT. In: **Simpósio Latino-Americano de Sensoriamento Remoto**. Gramado: SELPER, 1986. INPE-8113-PRE/113.
- McCullagh, P.; Nelder, J. A. **General linear models**. New York: Chapman Hall, 1989. 511 p.
- Mc Loughlin, J. B. **Planification urbaine et regionale**. Paris: Dunod, 1972.
- Medda, F.; Nijkamp, P.; Rietveld, P. Urban industrial relocation: the theory of edge cities. **Environment and Planning B**, v. 26, p.751-761, 1999.
- Merlin, P. **Méthodes quantitatives et espace urbain**. Paris: Masson et Cie Éditeurs, 1973. 190 p.
- Mertens, B.; Lambin, E. Modelling land cover dynamics: integration of fine scale land cover data with landscape attributes. **Journal of Applied Geography**, v. 1, n.1, p. 48-52, 1999.
- Mertens, B.; Sunderlin, W. D.; Ndoye, O.; Lambin, E. Impact of macro-economic change on deforestation in South Cameroon: integration of household survey and remotely-sensed data. **World development**, v. 28, n.6, p. 983-999, 2000.
- Moore, T. Geospatial expert systems. In: Openshaw, S.; Abrahart, R. J. ed. **Geocomputation**. New York: Taylor & Francis, 2000. Chap. 6, p. 127-159.
- Neter, J.; Wasserman, W. **Applied linear statistical models**. Georgetown: Irving-Dorsey, 1974. 842 p.
- Nóbrega, M. **História do Rio Tietê**. Belo Horizonte: Editora Itatiaia Ltda., 1981. 245 p.

- Novaes, A. G. **Modelos em planejamento urbano, regional e de transportes**. São Paulo: Editora Edgard Blücher, 1981. 290 p.
- Novo, E. M. L. M. **Sensoriamento remoto – princípios e aplicações**. São Paulo: Editora Edgar Blücher Ltda, 1988. 308p.
- Ohtake, R. ed. **O livro do Rio Tietê**. São Paulo: CESP, 1991. 167 p.
- O’Sullivan, D. **Graph-based cellular automaton models of urban spatial processes**. London. 317p. Unpublished PhD thesis – Centre for Advanced Spatial Analysis of the University College London (CASA-UCL), 2000.
- O’Sullivan, D. Exploring spatial processes dynamics using irregular graph-based cellular automaton models. **Geographical Analysis**, v. 33, p. 1-18, 2001a.
- O’Sullivan, D. Graph-cellular automata: a generalised discrete urban and regional model. **Environment and Planning B**, v. 28, p. 687-705, 2001b.
- Odum, E. P. **Ecology**. New York: Holt, Rinehart and Winston, 1963.
- Openshaw, S. GeoComputation. In: Openshaw, S.; Abraham, R. J. ed. **Geocomputation**. New York: Taylor & Francis, 2000. Chap. 1, p. 1-31.
- Oppenheim, N. A critical survey of current developments in urban and regional modelling. In: Hutchinson, B.; Batty, M. ed. **Advances in urban systems modelling**. Amsterdam: Elsevier, 1986. p. 41-54.
- ORB - The online reference book for medieval studies. **Medieval English towns - Problems of definition**. 2002. Available at <http://www.trytel.com/~tristan/towns/townint2.html>. Accessed on 07/17/03.
- Paelinck, J. H. P. Dynamic urban growth models. **Papers of the Regional Science Association**, v. 24, p. 25-37, 1970.
- Papageorgiou, Y. Y.; Pines, D. An essay on urban economic theory. **Regional Science and Urban Economics**, v. 31, n. 5, p. 626-630, 2001.
- Papini, L.; Rabino, G. A.; Colonna, A.; Di Stefano, V.; Lombardo, S. (1998) Learning cellular automata in a real world: the case study of the Rome metropolitan area. In Bandini, S.; Serra, R.; Liverani, F. S. ed. Cellular automata: research towards industry. **ACRI’96: Proceedings of the Third Conference on Cellular Automata for Research and Industry**. London: Springer-Verlag, 1998. p. 165-183.
- Perraton, J.; Baxter, R. **Models, evaluations & information systems for planners**. Lancaster: MTP Construction, 1974. 305 p.
- Phipps, M. Dynamical behaviour of cellular automata under constraints of neighbourhood coherence. **Geographical Analysis**, v. 21, p. 197-215, 1989.
- Phipps, M.; Langlois, A. Spatial dynamics, cellular automata and parallel processing computers. **Environment and Planning B**, v. 24, n. 2, p. 193-204, 1997.
- Piracicaba (Prefeitura Municipal). **Código de obras do município de Piracicaba – Lei nº 1.297, de 22 de dezembro de 1964**. Piracicaba: Divisão de Obras Particulares e Urbanismo da Prefeitura Municipal, 1965.

Plano diretor de Bauru 1996. Bauru: Secretaria Municipal de Planejamento de Bauru (SEPLAN – Bauru), 1997. 200p.

Plano estadual de recursos hídricos – Primeiro Plano do Estado de São Paulo – 1990. São Paulo: Departamento de Águas e Energia Elétrica (DAEE), 1990.

Pontius Jr., G. Quantification error versus location error in comparison of categorical maps. **Photogrammetric Engineering & Remote Sensing**, v. 66, n. 8, p. 1011-1016, 2000.

Pontius Jr., G. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. **Photogrammetric Engineering & Remote Sensing**, v. 68, n. 10, p. 1041-1049, 2002.

Popper, K. R. **Objective knowledge.** Oxford, UK: Oxford University Press, 1972.

Porter, H. R.; Henley, E. J. Applications of the Forrester model to Harris County, Texas. In: Chen, K. ed. **Urban dynamics: extensions and reflections.** San Francisco: San Francisco Press Inc., 1972.

Portugali, J.; Benenson, I.; Omer, I. Spatial cognitive dissonance and sociospatial emergence in a self-organizing city. **Environment and Planning B**, v. 24, n. 2, p. 263-285, 1997.

Portugali, J.; Benenson, I.; Omer, I. City: the greens and the blues. In: Portugali, J. ed. **Self-organization and the city.** Berlin: Springer Verlag, 1999. Chap. 5.

Produtividade e ajuste na indústria paulista – 1986 - 1994. São Paulo: Fundação Seade, 1995. 120p. (Análises & Ensaios).

Prud'Homme, R. New trends in the cities of the world. In: Knight, R. V.; Gappert, G. ed. **Cities in a global society.** Newbury Park: SAGE Publications, Inc., 1989. Chap. 2, p. 44-57.

Rabino, G. A. Identification of cellular automata. In: **Workshop on Cellular Automata, Neural Networks and Multi-Agent Systems in Urban Planning.** Milan: Politecnico di Milano, 2002.

Richards, J. A. **Remote sensing digital image analysis – An introduction.** Berlin: Springer-Verlag, 1995. 340p.

Ridd, M.K. Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities. **International Journal of Remote Sensing**, v. 16, p. 2165-2186, 1995.

Sanders, L.; Pumain, D.; Mathian, H.; Guérin-Pace, F.; Bura, S. SIMPOP: a multi-agent system for the study of urbanism. **Environment and Planning B**, v. 24, n.2, p. 287-305, 1997.

São Paulo (Estado). **Constituição do Estado de São Paulo** (CE – SP. Cap. II. Seção II. art. 153. § 1º). 1989. Available at < http://folio.al.sp.gov.br/cgi-bin/om_isapi.dll?clientID=146938&infobase=constsp.nfo&softpage=Browse_Frame_Pg42>. Accessed on 07/18/03.

- São Paulo no limiar do século XXI - Cenários da urbanização paulista. São Paulo: Fundação Seade, 1992. v. 5. 106 p.
- Schelhorn, T.; O'Sullivan, D.; Haklay, M.; Thurstain-Goodwin, M. **Streets: an agent-based pedestrian model**. 1999. Available at < <http://www.casa.ucl.ac.uk/streets.pdf>>. Accessed on 07/31/02.
- Secretaria Municipal de Planejamento de Bauru (SEPLAN - Bauru). **Zoneamento – usos do solo**. Bauru: SEPLAN, 1979a. One map. Scale 1:20,000. Paper plan.
- Secretaria Municipal de Planejamento de Bauru (SEPLAN - Bauru). **Densidades de ocupação**. Bauru: SEPLAN, 1979b. One map. Scale 1:20,000. Paper plan.
- Secretaria Municipal de Planejamento de Bauru (SEPLAN - Bauru). **Planta da area urbana e arredores – zoneamento**. Bauru: SEPLAN, 1988a. One map. Scale 1:20,000. Paper plan.
- Secretaria Municipal de Planejamento de Bauru (SEPLAN - Bauru). **Planta da area urbana e arredores – principais obras viárias existentes e a construir**. Bauru: SEPLAN, 1988b. One map. Scale 1:20,000. Paper plan.
- Secretaria Municipal de Planejamento de Piracicaba (SEMUPLAN - Piracicaba). **Plano diretor de desenvolvimento de Piracicaba – planta de zoneamento**. Piracicaba: SEMUPLAN, 1985. One map. Scale 1:20,000. Paper plan.
- Secretaria Municipal de Planejamento de Piracicaba (SEMUPLAN - Piracicaba). **Planta geral do município de Piracicaba**. Piracicaba: SEMUPLAN, 1999. One map. Scale 1:20,000. Paper plan.
- Secretaria Municipal de Planejamento de Piracicaba (SEMUPLAN - Piracicaba). **Planta geral do município de Piracicaba**. Piracicaba: SEMUPLAN, 2003. One map. Scale 1:20,000. Paper plan.
- Seevers, P. M.; Johnston, D. C.; Feuquay, J. W. Band combinations selection for visual interpretation of thematic mapper data for resource management. In: Technical paper: Racing into tomorrow. American Congress on Surveying and Mapping and American Society for Photogrammetry and Remote Sensing. **Proceedings**. Fall Church, Indianapolis: ACSM - ASPRS, 1985.
- Seidman, D. R. **The construction of an urban growth model**. Philadelphia: Delaware Valley Regional Planning Commission, 1969. (DVRPC Plan - Report n. 1).
- Semoloni, F. An urban and regional model based on cellular automata. **Environment and Planning B**, v. 24, p. 589-612, 1997.
- Serra, G. **Urbanização e centralismo autoritário**. São Paulo: Livraria Nobel - EDUSP, 1991. 172 p.
- Sistema de contas nacionais – contas regionais do Brasil. Rio de Janeiro: IBGE, 1999. (Série Contas Nacionais, n. 6).
- Sistema de contas nacionais – Brasil – 1999-2001. Rio de Janeiro: IBGE, 2002. 205 p. (Série Contas Nacionais, n. 9).

- Skole, D. L. Data on global land-cover change: Acquisition, assessment, and analysis. In: Meyer, W. B.; Turner, B. L. II. Ed. **Changes in land use and land cover: a global perspective**. Cambridge, UK: Cambridge University Press, 1994. p. 437-471.
- Soares-Filho, B. S. **Modelagem da dinâmica de paisagem de uma região de fronteira de colonização amazônica**. São Paulo. 299p. Unpublished PhD thesis – Escola Politécnica da Universidade de São Paulo (EPUSP), 1998.
- Soares-Filho, B. S.; Assunção, R. M.; Pantuzzo, A. E. Modeling the spatial transition probabilities of landscape dynamics in an Amazonian colonization frontier. **BioScience**, v. 51, n. 12, p. 1059-1067, 2001.
- Soares-Filho, B. S.; Cerqueira, G. C.; Pennachin, C. L. DINAMICA – a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. **Ecological Modelling**, v. 154, p. 217-235, 2002.
- Sobreira, F.; Gomes, M. **The geometry of slums: boundaries, packing and diversity**. 2001. Available at <<http://www.casa.ucl.ac.uk/paper30.pdf>>. Accessed on 11/09/01.
- Solberg, R.; Fiskum, S.; Giffen, M. C. Detection of urban areas by interference filtering. In: Remote Sensing for the Nineties, Igarss'90. 10th Annual International Geoscience and Remote Sensing Symposium. **Proceedings**. Washington, D.C.: IEEE, 1990. v. II, p. 20-24.
- Takeyama, M. Building spatial models within GIS through geo-algebra. **Transactions in GIS**, v. 2, p. 245-256, 1997.
- Takeyama, M.; Couclelis, H. Map dynamics: integrating cellular automata and GIS through geo-algebra. **International Journal of Geographical Information Science**, v. 11, p. 73-91, 1997.
- Thompson, A. **Chaos theory** – definition of chaos explained. 1997. Available at <<http://www.galactic-guide.com/articles/8R85.html>>. Accessed on 07/31/03.
- Tobler, W. A computer movie simulating urban growth in the Detroit region. **Economic Geography**, v. 26, p. 234-240, 1970.
- Tobler, W. Cellular geography. In: **Philosophy in geography**. Dordrecht: D. Reidel Publishing Co., 1979. p. 379-386.
- Torrens, P. M. **Automata-based models of urban systems**. 2003. Available at <http://www.geogr.ku.dk/courses/phd/gis-model/Torrens_automata_based_models_of_urban_systems.pdf>. Accessed on 05/18/03.
- Torrens, P. M.; O'Sullivan, D. Cellular automata and urban simulation: where do we go from here? editorial. **Environment and Planning B**, v. 28, p. 163-168, 2001.
- Turner, B. L. II; Meyer, B. L. Global land use and land cover change: An overview. In: Meyer, W. B.; Turner, B. L. II. Ed. **Changes in land use and land cover: a global perspective**. Cambridge, UK: Cambridge University Press, 1994. p. 3-10.
- United Nations. **World urbanization prospects: the 1996 revision database**. New York: United Nations Population Division, 1996.

- United Nations. **World urbanization prospects: the 1998 revision database**. New York: United Nations Population Division, 1998.
- Ward, D. P.; Murray, A. T.; Phinn, S. R. **An optimized cellular automata approach for sustainable urban development in rapidly urbanizing regions**. 1999. Available at <http://www.geovista.psu.edu/sites/geocomp99/Gc99/025/gc_025.htm>. Accessed on 10/30/02.
- Watt, K. E. F. **Principles of environmental science**. New York: McGraw-Hill Book Co., 1973.
- Wegener, M. New spatial planning models. **Emerging Directions in Urban Planning Research Seminar**. Enschede: ITC, 2000.
- Wegener, M.; Gnad, F.; Vannahme, M. The time scale of urban change. In: Hutchinson, B.; Batty, M. ed. **Advances in urban systems modelling**. Amsterdam: Elsevier, 1986. p. 175-197.
- Wei, W. W. S. **Time Series Analysis**. California: Attison Wesley Publishing Co., 1990. 478 p.
- White, R.; Engelen, G. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land use patterns. **Environment and Planning A**, v. 25, p. 1175-199, 1993.
- White, R.; Engelen, G. Cellular automata as the basis of integrated dynamic regional modelling. **Environment and Planning B**, v. 24, n. 2, p. 235-246, 1997.
- White, R.; Engelen, G. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. **Computers, Environment and Urban Systems**, v. 24, p. 383-400, 2000.
- White, R.; Engelen, G.; Uljee, I. **Vulnerability assessment of low-lying coastal areas and small islands to climate change and sea level rise – Phase 2: Case study St. Lucia**. Kingston, Jamaica: United Nations Environment Programme: Caribbean Regional Co-ordinating Unit, 1998.
- Wikipedia – The Free Encyclopedia. **Economies of agglomeration**. 2003. Available at <http://www.wikipedia.org/wiki/Economies_of_agglomeration>. Accessed on 08/09/03.
- Wilson, A. G. **Entropy in urban and regional modeling**. London: Pion Press, 1970.
- Wilson, A. G. **Catastrophe theory and bifurcations: applications to urban and regional systems**. Berkeley: University of California Press, 1981.
- Wingo, L. **Transportation and urban land**. Washington, D. C.: Resources for the Future Inc., 1961.
- Wolfram, S. Statistical mechanics of cellular automata. **Review of modern physics**, v. 55, p. 601-643, 1983.
- Wolfram, S. Universality and complexity in cellular automata. **Physica**, v. 10D, p. 1-35, 1984.

Wolman, M. G. Criteria for land use. In: McLare, D. J.; Skinner, B. J. ed. **Resources and world development**. New York: John Wiley, 1987. p. 643-657.

Wu, F. A linguistic cellular automata simulation approach for sustainable land development in a fast growing region. **Computers, Environment and Urban Systems**, v. 20, n.6, p. 367-387, 1996.

Wu, F. Simland: a prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules. **International Journal of Geographical Information Science**, v. 12, n. 1, p. 63-82, 1998.

Xia, L.; Yeh, A. G. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. **International Journal of Geographic Information Science**, v. 14, p. 131-152, 2000.

Xie, Y. A generalized model for cellular urban dynamics. **Geographical Analysis**, v. 28, n. 4, p. 350-373, 1996.

Yeh, A. G.; Li, X. A cellular automata model to simulate development density for urban planning. **Environmental and Planning B**, v. 29, p. 431-450, 2002.

Zhou, D.; Vertinsky, I. Strategic location decisions in a growing market. **Regional Science and Urban Economics**, v. 31, n. 5, p. 523-533, 2001.