

THE APPLICATION OF GEOGRAPHIC INFORMATION  
SYSTEMS CELLULAR AUTOMATA BASED MODELS TO LAND  
USE CHANGE MODELLING OF LAGOS, NIGERIA

by

Onuwa Honey Stephen Okwuashi

A thesis submitted to Victoria University of Wellington  
in fulfilment of the requirement for the degree of  
*Doctor of Philosophy in Geography*

January 2011

School of Geography, Environment, and Earth Sciences  
Victoria University of Wellington  
Wellington, New Zealand



## ABSTRACT

The urban expansion of Lagos continues unabated and calls for urgent concern. This thesis explored the use of both the conventional and unconventional techniques for modelling land use change. Two conventional methods (ordinary least squares and geographically weighted regression) were based on geographic information systems, while four unconventional methods (logistic regression, artificial neural networks, and two proposed types of support vector machine) were based on cellular automata. These techniques were evaluated using three land use epochs: 1963-1978, 1978-1984, and 1984-2000.

The conventional methods make quite strong statistical assumptions, some of which are shown not to be met by the land use data at hand. Despite this, these methods do exhibit substantial agreement between observed and the predicted maps. The non cellular automata and cellular automata modelling were then implemented with the logistic regression, artificial neural network, support vector machine, and fuzzy support vector machine models, with model parameters set by  $k$ -fold cross-validation. The cellular automata predicted maps were more accurate than those of the non cellular automata.

The cellular automata modelling results from the proposed support vector machine and fuzzy support vector machine were compared with those from the geographic information systems based geographically weighted regression, logistic regression, and artificial neural network. The results from the geographic information systems based geographically weighted regression were the best, followed by those from the support vector machine and fuzzy support vector machine, followed by the artificial neural network, and logistic regression. This research demonstrated that the proposed support vector machine and fuzzy support vector machine based cellular automata models are promising tools for land use change modelling.

## ACKNOWLEDGEMENTS

With utmost humility, I express my profound gratitude to God for granting me the vision and knowledge needed to accomplish this thesis.

I remain very grateful to my supervisors: A/Prof. (Dr.) Jack McConchie, Dr. Stephen Marshall, and Prof. Peter Nwilo. My immense gratitude to the university administrative officers and lecturers whom I relied upon during my study: Mrs. Shona de Sain, Mrs. Liz Richardson, A/Prof. (Dr.) Mike Hannah, Prof. John Overton, Mr. Richard Willis, Dr. Marcus Freat, and Dr. Deborah Laurs.

I remain grateful to my relatives, friends, and well wishers: Mr. Felix Okwuashi, Surv. Etim Eyo, Miss Uduak Nnah, Mr. Samuel Ekundayo, Engr. Chiedu Enwenmadu, Dr. Alex Okwuashi, Mr. Ben Okwuashi, Mr. Lawrence Okwuashi, Engr. Festus Okwunwa, Mr. Patrick Onyemah, Mrs. Anthonia Nwachukwu, Surv. U. W. Obot, Builder Isaac Odesola, Builder Tony Ujene, Surv. Sola Daves, Surv. Chuks Azonobi, Surv. Aniekan Ekpo, Engr. Israel Abali, Mr. Emmanuel Ebilueye, Mr. Innocent Aimiator, Mr. Augustine Aimiator, Chief Edward Molokwu, Chief S.A.N. Onwudinjo, Miss Ene Afia, Mr. Cheriff Adara, Dr. Richard Juma, Miss Helen Walls, Mr. Godwin Utuka, and Miss Ekaette Ekpo; and so many others not mentioned.

## TABLE OF CONTENTS

ABSTRACT-----	i
ACKNOWLEDGEMENTS-----	ii
TABLE OF CONTENTS-----	iii
LIST OF FIGURES-----	viii
LIST OF TABLES-----	xiii
 CHAPTER 1 INTRODUCTION-----	 1
1.1 Research problem-----	1
1.2 Choosing predictive models and reasons for choosing models for Lagos-----	2
1.3 Research questions-----	3
1.4 Specific research aims and objectives-----	5
1.5 Thesis domain: the concept of geocomputation-----	7
1.6 Research contributions-----	8
1.7 Structure of thesis-----	9
 CHAPTER 2 GEOGRAPHIC INFORMATION SYSTEMS AND CELLULAR AUTOMATA BASED MODELS-----	 10
2.1 Preamble-----	10
2.2 Geographic information systems based models-----	10
2.2.1 Geographic information systems-----	10
2.2.2 Geographically weighted regression and ordinary least squares-----	11
2.3 Cellular automata based models-----	14
2.3.1 Cellular automata-----	14
2.3.2 Cellular automata calibration methods-----	17
2.3.2.1 Heuristic-based modelling-----	17
2.3.2.2 Agent-based modelling-----	18
2.3.2.3 Fractal-based modelling-----	19
2.3.2.4 Fuzzy logic based modelling-----	20
2.3.2.5 Spatial statistics based modelling-----	22
2.3.2.5.1 Markov chain based modelling-----	22
2.3.2.5.2 Logistic regression based modelling-----	24
2.3.2.6 Artificial neural network based modelling-----	25

2.4 Integration of geographic information systems and cellular automata and cellular automata models-----	27
2.5 Summary-----	28
CHAPTER 3 SUPPORT VECTOR MACHINES-----	29
3.1 Preamble-----	29
3.2 Statistical learning theory, structural risk minimisation, and Vapnik-Chervonenkis dimension-----	31
3.3 Support vector machine classification-----	33
3.4 Fuzzy support vector machine classification-----	39
3.5 Numerical examples-----	41
3.5.1 Classifying linearly separable data-----	41
3.5.2 Classifying nonlinearly separable data-----	46
3.5.3 Selecting an optimal value for the regularisation parameter $C$ , and kernel parameters for SVM classification and computing the accuracy of the SVM-----	51
3.6 Summary-----	54
CHAPTER 4 LAND USE DEVELOPMENT OF LAGOS-----	55
4.1 Preamble-----	55
4.2 Study area-----	55
4.3 Land use transition in Lagos-----	57
4.4 Selecting factors influencing land use development in Lagos and reasons for their consideration-----	59
4.4.1 Water-----	60
4.4.2 Residential structures-----	60
4.4.3 Industrial and commercial centres-----	61
4.4.4 Major roads-----	61
4.4.5 Railway-----	61
4.4.6 Lagos Island-----	62
4.4.7 International airport-----	62
4.4.8 International seaport-----	62
4.4.9 University of Lagos-----	63
4.4.10 Lagos State University-----	63
4.4.11 Income potential-----	63
4.4.12 Population potential-----	64
4.5 Land use planning of Lagos-----	65

4.5.1 Pre-independence planning legislations-----	65
4.5.2 Post-independence planning legislations-----	66
4.5.3 Appraisal of pre-independence and post-independence planning legislations in Lagos-----	67
4.6 Summary-----	69

CHAPTER 5 CONVENTIONAL METHODS: APPLICATION OF GIS-BASED ORDINARY LEAST SQUARES AND GEOGRAPHICALLY WEIGHTED REGRESSION MODELS TO LAND USE CHANGE MODELLING OF LAGOS-----	70
---	----

5.1 Preamble-----	70
5.2 Mathematical calibration for land use change modelling-----	70
5.3 Land use data-----	71
5.4 Land use data constraints-----	71
5.5 Land use data preparation-----	73
5.6 Extraction of land use variables-----	73
5.7 Modelling-----	76
5.7.1 Model significance-----	81
5.7.2 Assessing the significance of each independent variable in the model-----	82
5.7.3 Traditional statistical assumptions vis-à-vis GIS-based linear regression land use change modelling-----	86
5.7.3.1 Multicollinearity-----	87
5.7.3.2 Linearity-----	87
5.7.3.3 Normality-----	88
5.7.3.4 Autocorrelation/spatial autocorrelation-----	90
5.7.3.5 Homoscedasticity or stationarity-----	91
5.7.4 Prediction-----	92
5.7.5 Assessing the impact of excluding the insignificant variable from the OLS and GWR modelling results-----	103
5.7.6 The effect of excluding each independent variable from the GWR model-----	104
5.8 Summary-----	107

CHAPTER 6 COMMON UNCONVENTIONAL METHODS: APPLICATION OF LOGISTIC REGRESSION AND ARTIFICIAL NEURAL NETWORK BASED CELLULAR AUTOMATA MODELS TO LAND USE CHANGE MODELLING OF LAGOS-----	108
6.1 Preamble-----	108
6.2 LR-based modelling-----	108
6.2.1 LR-based CA calibration-----	108
6.2.2 Data and methodology-----	110
6.2.3 Non-CA modelling-----	112
6.2.4 CA modelling-----	124
6.3 ANN-based modelling-----	128
6.3.1 ANN-based CA calibration-----	128
6.3.2 Data and methodology-----	130
6.3.3 Non-CA modelling-----	132
6.3.4 CA modelling-----	138
6.4 Summary-----	141
CHAPTER 7 PROPOSED UNCONVENTIONAL METHODS: APPLICATION OF SUPPORT VECTOR MACHINE AND FUZZY SUPPORT VECTOR MACHINE BASED CELLULAR AUTOMATA MODELS TO LAND USE CHANGE MODELLING OF LAGOS-----	143
7.1 Preamble-----	143
7.2 SVM-based modelling-----	143
7.2.1 SVM-based CA calibration-----	143
7.2.2 Data and methodology-----	144
7.2.3 Non-CA modelling-----	147
7.2.4 CA modelling-----	157
7.3 FSVM-based modelling-----	165
7.3.1 FSVM-based CA calibration-----	165
7.3.2 Data and methodology-----	165
7.3.3 Non-CA modelling-----	166
7.3.4 CA modelling-----	176
7.4 The effect of the neighbourhood function on SVM and FSVM based CA modelling-----	185
7.5 Predicting future land use using SVM and FSVM based CA-----	187
7.6 Summary-----	189



CHAPTER 8 CONCLUSION-----	190
8.1 Appraisal of modelling results-----	190
8.1.1 GIS-based GWR, LR, and ANN versus the proposed SVM and FSVM-----	190
8.1.1.1 GIS-based GWR versus SVM and FSVM-----	190
8.1.1.2 LR versus SVM and FSVM-----	194
8.1.1.3 ANN versus SVM and FSVM-----	199
8.1.2 A comparison of the non-CA and CA predicted maps for LR, ANN, SVM, and FSVM models-----	204
8.2 Overview-----	206
8.2.1 Research questions and answers-----	206
8.2.2 The pros and cons of the proposed SVM and FSVM models-----	207
8.3 The benefit of the thesis to Lagos and Nigeria-----	208
8.4 The scientific benefit of the thesis-----	208
8.5 Recommendation for further research-----	209
REFERENCES-----	210

## LIST OF FIGURES

2.1 A hypothetical one-dimensional cellular automata-----	15
2.2 A hypothetical von Neumann neighbourhood image-----	15
2.3 A hypothetical Moore neighbourhood image-----	15
2.4 A single output neural network architecture-----	26
3.1 Hyperplane with the maximum separation for a linear data-----	29
3.2 The separating linear hyperplanes for a linear data-----	30
3.3 Separating non-separable data with linear separable hyperplanes-----	34
3.4 The process of classifying a nonlinearly separable data-----	37
3.5 2D linearly separable data-----	43
3.6 Results of classifying a linearly separable data with a linear kernel-----	45
3.7 2D nonlinearly separable data -----	47
3.8 Classification result from the polynomial and the RBF kernels-----	50
4.1 Lagos in relation to Nigeria-----	56
4.2 Land use change of Lagos between 1963 and 2000-----	58
5.1 Gridded land use maps of Lagos from 1963-2000-----	78
5.2 (a) A typical linear regression relationship between a dependent and an independent variable; (b) dependent variable plotted against distance to water-----	88
5.3 Plots for OLS residuals for periods 1963-1978, 1978-1984, and 1984-2000-----	90
5.4 Predicted GWR maps for 1963-1978, 1978-1984, and 1984-2000-----	95

5.5 Estimated GWR standard residuals and local $R^2$ for 1963-1978, 1978-1984, and 1984-2000-----	100
5.6 Estimated GWR model condition numbers for 1963-1978, 1978-1984, and 1984-2000-----	101
5.7 Normality plot for GWR model residuals for periods 1963-1978, 1978-1984, and 1984-2000-----	102
5.8 The effect of excluding each variable on the GWR estimated AIC and r-square results for 1963-1978-----	106
5.9 The effect of excluding each variable on the GWR estimated AIC and r-square results for 1978-1984-----	106
5.10 The effect of excluding each variable on the GWR estimated AIC and r-square results for 1984-2000-----	107
6.1 Flowchart for LR modelling-----	111
6.2 Calculated LR probabilities for 1963-1978, 1978-1984, and 1984-2000-----	115
6.3 Showing the effect of excluding each variable from the model for 1963-1978-----	117
6.4 Showing the effect of excluding each variable from the model for 1978-1984-----	117
6.5 Showing the effect of excluding each variable from the model for 1984-2000-----	118
6.6 LR modelling results for 1963-1978-----	121
6.7 LR modelling results for 1978-1984-----	122
6.8 LR modelling results for 1984-2000-----	122
6.9 LR predicted maps for periods 1963-1978, 1978-1984, and 1984-2000-----	123

6.10 LR modelling result: computed overall mean kappa and standard deviations for 200 designated iteration thresholds-----	126
6.11 LR-based CA predicted maps for 1963-1978, 1978-1984, and 1984-2000-----	126
6.12 LR: plotted mean sensitivity versus 1-specificity and standard deviations for periods 1963-1978, 1978-1984, and 1984-2000-----	128
6.13 Flowchart for ANN modelling-----	131
6.14 Computed mean kappa for each dataset based on designated ANN neurons (1963-1978)-----	133
6.15 Computed mean kappa for each dataset based on designated ANN neurons (1978-1984)-----	134
6.16 Computed mean kappa for each dataset based on designated ANN neurons (1984-2000)-----	134
6.17 The ANN overall mean kappa for 1963-1978, 1978-1984, and 1984-2000 based on the designated neurons-----	135
6.18 ANN non-CA predicted maps-----	135
6.19 Computed mean kappa coefficients from the cross-validation results from the LR and a one neuron ANN-----	138
6.20 ANN result: overall mean kappa and standard deviations for 200 designated iteration thresholds-----	139
6.21 ANN based CA predicted maps for 1963-1978, 1978-1984, and 1984-2000-----	140
6.22 ANN-based CA ROC analysis: plotted mean sensitivity versus 1-specificity and standard deviations for 1963-1978, 1978-1984, and 1984-2000-----	141
7.1 Flowchart for SVM and FSVM models-----	146

7.2 SVM-based non-CA: polynomial, RBF, and linear kernels cross-validation results for 1963-1978-----	148
7.3 SVM-based non-CA: polynomial, RBF, and linear kernels cross-validation results for 1978-1984-----	149
7.4 SVM-based non-CA: polynomial, RBF, and linear kernels cross-validation results for 1984-2000-----	150
7.5 SVM-based non-CA: selecting an optimal C value for periods 1963-1978, 1978-1984, and 1984-2000-----	151
7.6 SVM-based non-CA: cross-validation results for obtaining optimal values for $d$ and gamma-----	152
7.7 SVM-based non-CA: overall model accuracy for periods 1963-1978, 1978-1984, and 1984-2000-----	153
7.8 SVM-based non-CA predicted maps for 1963-1978, 1978-1984, and 1984-2000 using the polynomial, RBF, and Linear kernels-----	154
7.9 SVM-based non-CA: computed overall mean kappa and standard deviations for 200 designated iteration thresholds-----	158
7.10 SVM-based CA predicted maps for periods 1963-1978, 1978-1984, and 1984-2000-----	159
7.11 SVM based CA: model accuracy for periods 1963-1978, 1978-1984, and 1984-2000-----	163
7.12 SVM based CA: ROC plot for 1963-1978-----	163
7.13 SVM based CA: ROC plot for 1978-1984-----	164
7.14 SVM based CA: ROC plot for 1984-2000-----	164
7.15 FSVM-based non-CA polynomial, RBF, and linear kernels cross-validation results for 1963-1978-----	167

7.16 FSVM-based non-CA polynomial, RBF, and linear kernels cross-validation results for 1978-1984-----	168
7.17 FSVM-based non-CA polynomial, RBF, and linear kernels cross-validation results for 1984-2000-----	169
7.18 FSVM-based non-CA: selecting an optimal C value for periods 1963-1978, 1978-1984, and 1984-2000-----	170
7.19 FSVM-based non-CA: cross-validation results for obtaining an optimal value for $d$ and gamma-----	171
7.20 FSVM-based non-CA: overall model accuracies for periods 1963-1978, 1978-1984, and 1984-2000-----	171
7.21 FSVM-based non-CA predicted maps for 1963-1978, 1978-1984, and 1984-2000-----	173
7.22 FSVM-based CA: computed overall mean kappa and standard deviations for 200 designated iteration thresholds-----	177
7.23 FSVM-based CA predicted maps for periods 1963-1978, 1978-1984, and 1984-2000-----	179
7.24 FSVM-based CA model accuracy-----	183
7.25 FSVM-based CA: ROC plot for period 1963-1978-----	183
7.26 FSVM-based CA: ROC plot for period 1978-1984-----	184
7.27 FSVM-based CA: ROC plot for period 1984-2000-----	184
7.28 The effect of the Moore and Neumann neighbourhood functions on the SVM and FSVM based CA modelling-----	186
7.29 Predicted land use maps in 2015 and 2030-----	188

## LIST OF TABLES

1.1 Conventional and unconventional land use change models adopted for Lagos-----	4
3.1 2D linearly separable data (training data)-----	42
3.2 Test data for a linearly separable classification-----	43
3.3 Training results using a linear kernel-----	44
3.4 Test results using a linear kernel-----	45
3.5 2D nonlinearly separable data (training data)-----	46
3.6 Test data for a nonlinearly separable classification-----	47
3.7 Training results using a polynomial kernel-----	48
3.8 Test results using a polynomial kernel-----	49
3.9 Training results using an RBF kernel-----	49
3.10 Test results using an RBF kernel-----	50
3.11 T1: training data 1 for a $k$ -fold cross-validation experiment-----	52
3.12 T2: training data 2 for a $k$ -fold cross-validation experiment-----	52
3.13 T3: training data 3 for a $k$ -fold cross-validation experiment-----	53
3.14 $k$ -fold cross-validation results for selecting optimal values of $C$ and $\gamma$ -----	54
4.1 Estimated land use area, change, and rate of change from 1963 to 2000-----	59
4.2 Estimated population growth rate of Lagos from 1866-2015-----	64
5.1 Extracted land use variables-----	74

5.2 Joint Wald Statistic for 1963-1978, 1978-1984, and 1984-2000-----	81
5.3 Statistical results for assessing the significance of each independent variable in the model for 1963-1978-----	83
5.4 Statistical results for assessing the significance of each independent variable in the model for 1978-1984-----	84
5.5 Statistical results for assessing the significance of each independent variable in the model for 1984-2000-----	85
5.6 Jarque-Bera test for 1963-1978, 1978-1984, and 1984-2000-----	89
5.7 OLS modelling: spatial autocorrelation test for 1963-1978, 1978-1984, and 1984-2000 using ArcGIS Moran's I tool-----	91
5.8 Koenker (BP) Statistic for 1963-1978, 1978-1984, and 1984-2000-----	92
5.9 Calculated AIC values for GIS-based OLS and GWR regression models-----	93
5.10 GWR modelling: calculated $R^2$ values for 1963-1978, 1978-1984, and 1984-2000-----	94
5.11 GIS-based GWR modelling: confusion matrix for period 1963-1978-----	96
5.12 GIS-based GWR modelling: confusion matrix for period 1978-1984-----	96
5.13 GIS-based GWR modelling: confusion matrix for period 1984-2000-----	96
5.14 Calculated kappa statistic estimates for 1963-1978, 1978-1984, and 1984-2000-----	99
5.15 GWR modelling: spatial autocorrelation test for 1963-1978, 1978-1984, and 1984-2000 using ArcGIS Moran's I-----	103
6.1 Logistic regression results for 1963-1978-----	113
6.2 Logistic regression results for 1978-1984-----	114
6.3 Logistic regression results for 1984-2000-----	114



6.4 LR-based non-CA: confusion matrix for period 1963-1978-----	124
6.5 LR-based non-CA: confusion matrix for period 1978-1984-----	124
6.6 LR-based non-CA: confusion matrix for period 1984-2000-----	124
6.7 LR-based CA: confusion matrix for period 1963-1978-----	127
6.8 LR-based CA: confusion matrix for period 1978-1984-----	127
6.9 LR-based CA: confusion matrix for period 1984-2000-----	127
6.10 ANN-based non-CA: confusion matrix for period 1963-1978-----	136
6.11 ANN-based non-CA: confusion matrix for period 1978-1984-----	136
6.12 ANN-based non-CA: confusion matrix for period 1984-2000-----	136
6.13 Input weights from a one neuron neural network for periods 1963-1978, 1978-1984, and 1984-2000-----	137
6.14 ANN-based CA: confusion matrix for period 1963-1978-----	140
6.15 ANN-based CA: confusion matrix for period 1978-1984-----	140
6.16 ANN-based CA: confusion matrix for period 1984-2000-----	141
7.1 SVM-based non-CA: confusion matrix for polynomial 1963-1978-----	155
7.2 SVM-based non-CA: confusion matrix for RBF 1963-1978-----	155
7.3 SVM-based non-CA: confusion matrix for linear 1963-1978-----	155
7.4 SVM-based non-CA: confusion matrix for polynomial 1978-1984-----	155
7.5 SVM-based non-CA: confusion matrix for RBF 1978-1984-----	156
7.6 SVM-based non-CA: confusion matrix for linear 1978-1984-----	156

7.7 SVM-based non-CA: confusion matrix for polynomial 1984-2000-----	156
7.8 SVM-based non-CA: confusion matrix for RBF 1984-2000-----	156
7.9 SVM-based non-CA: confusion matrix for linear 1984-2000-----	157
7.10 SVM-based CA: confusion matrix for polynomial 1963-1978-----	160
7.11 SVM-based CA: confusion matrix for RBF 1963-1978-----	160
7.12 SVM-based CA: confusion matrix for linear 1963-1978-----	160
7.13 SVM-based CA: confusion matrix for polynomial 1978-1984-----	160
7.14 SVM-based CA: confusion matrix for RBF 1978-1984-----	161
7.15 SVM-based CA: confusion matrix for linear 1978-1984-----	161
7.16 SVM-based CA: confusion matrix for polynomial 1984-2000-----	161
7.17 SVM-based CA: confusion matrix for RBF 1984-2000-----	161
7.18 SVM-based CA: confusion matrix for linear 1984-2000-----	162
7.19 FSVM-based non-CA: confusion matrix for polynomial 1963-1978-----	174
7.20 FSVM-based non-CA: confusion matrix for RBF 1963-1978-----	174
7.21 FSVM-based non-CA: confusion matrix for linear 1963-1978-----	174
7.22 FSVM-based non-CA: confusion matrix for polynomial 1978-1984-----	174
7.23 FSVM-based non-CA: confusion matrix for RBF 1978-1984-----	175
7.24 FSVM-based non-CA: confusion matrix for linear 1978-1984-----	175
7.25 FSVM-based non-CA: confusion matrix for polynomial 1984-2000-----	175

7.26 FSVM-based non-CA: confusion matrix for RBF 1984-2000-----	175
7.27 FSVM-based non-CA: confusion matrix for linear 1984-2000-----	176
7.28 FSVM-based CA: confusion matrix for polynomial 1963-1978-----	180
7.29 FSVM-based CA: confusion matrix for RBF 1963-1978-----	180
7.30 FSVM-based CA: confusion matrix for linear 1963-1978-----	180
7.31 FSVM-based CA: confusion matrix for polynomial 1978-1984-----	180
7.32 FSVM-based CA: confusion matrix for RBF 1978-1984-----	181
7.33 FSVM-based CA: confusion matrix for linear 1978-1984-----	181
7.34 FSVM-based CA: confusion matrix for polynomial 1984-2000-----	181
7.35 FSVM-based CA: confusion matrix for RBF 1984-2000-----	181
7.36 FSVM-based CA: confusion matrix for linear 1984-2000-----	182
8.1 Calculated kappa coefficients for the GWR, SVM, and FSVM simulated maps-----	191
8.2 A typical 2-class confusion matrix-----	192
8.3 Computed confusion matrices, $z$ -scores, and $p$ -values respectively for the GWR versus SVM/FSVM simulated maps-----	193
8.4 A Non-CA modelling results: calculated kappa coefficients for the LR, SVM, and FSVM simulated maps-----	194
8.5 A CA modelling results: calculated kappa coefficients for the LR, SVM and FSVM Simulated maps-----	195
8.6 A Non-CA modelling: computed confusion matrices, $z$ -scores, and $p$ -values respectively for the LR versus SVM/FSVM simulated maps-----	196

8.7 CA modelling: computed confusion matrices, $z$ -scores, and $p$ -values respectively for the LR versus SVM/FSVM simulated maps-----	197
8.8 ROC analysis: calculated mean AUC and their respective standard deviations for LR versus SVM/FSVM simulated maps-----	198
8.9 Calculated $t$ -statistic and $p$ -values (in parenthesis) for LR versus SVM/FSVM simulated maps-----	199
8.10 Non-CA modelling: calculated kappa coefficients for the ANN, SVM, and FSVM Simulated maps-----	200
8.11 CA modelling: calculated kappa coefficients for the ANN, SVM, and FSVM simulated maps-----	200
8.12 Non-CA modelling: computed confusion matrices, $z$ -scores, and $p$ -values respectively for the ANN versus SVM/FSVM simulated maps-----	201
8.13 CA modelling: computed confusion matrices, $z$ -scores, and $p$ -values respectively for the ANN versus SVM/FSVM simulated maps-----	202
8.14 ROC results for the ANN, SVM, and FSVM simulated maps-----	203
8.15 Calculated $t$ -statistics and $p$ -values (in parenthesis) for ANN versus SVM/FSVM simulated maps-----	203
8.16 LR: computed confusion matrices, $z$ -scores, and $p$ -values respectively-----	204
8.17 ANN: computed confusion matrices, $z$ -scores, and $p$ -values respectively-----	205
8.18 SVM: computed confusion matrices, $z$ -scores, and $p$ -values respectively-----	205
8.19 FSVM: computed confusion matrices, $z$ -scores, and $p$ -values respectively-----	206





## CHAPTER 1

### INTRODUCTION

#### 1.1 Research problem

Changing land use globally is a topical issue of discussion. Land use change generally implies urban growth (or urban expansion), deforestation, desertification, change in land cover, soil erosion, and soil salination (Richards, 1990). This research focuses on urban expansion. Urbanisation is considered a vital social process that has had one of the most profound impacts on the local, regional, and global scales of the environment (Turner et al., 1990).

Urbanisation may be the benchmark for measuring economic growth and development, however in the case of sprawling cities of developing countries it has been accompanied by poverty, unemployment, environmental degradation, decaying infrastructure, and uncontrollable growth of informal settlements (Angotti, 1993). It was predicted that most African countries will be 54 percent urban by 2025 (United Nations, 2002). Sprawl in most African megacities is mainly due to rural-urban migration (Bernstein, 1995).

Urban sprawl in Lagos has put profound pressure on housing, infrastructure, and the environment (Braimoh & Onishi, 2007), and is generally viewed by most Nigerians as an intractable problem (Abiodun, 1974; Gandy, 2006). Urban planning in Lagos is practically unregulated (Abiodun, 1977). The use of spatial forecast models is lacking in the planning of Lagos. Spatial forecast models are vital for formulating sustainable planning strategies. The social and environmental repercussion of loosely planned urban cities like Lagos could be catastrophic especially in the present situation in Lagos that has constantly experienced remarkable urban expansion in a short period of time. Lagos has maintained an exponential population growth such that the population of Lagos is expected to attain 24.4 million people by 2015 (Barredo et

al., 2004). Technological methods, such as Geographic Information Systems (GIS), predictive models among other requirements for sustainable physical planning, are not presently utilised by urban planners in Lagos. Sustainable physical planning will remain a mirage in Lagos unless contemporary e-planning techniques are adopted (Oduwaye, 2009).

The author's interview with the Nigerian planning authorities corroborated the assertions made by Barredo et al. (2004) and Oduwaye (2009) that contemporary scientific predictive models are not being used for formulating planning programmes for Lagos. This planning defect has been identified as a problem to be addressed by this thesis.

Having identified the enormous gap in the lack of the use of (contemporary) scientific predictive tools for planning in Lagos, the next issue will be the search for predictive models capable of modelling a complex urban city like Lagos. This issue will be explored in the next section.

## **1.2 Choosing predictive models and reasons for choosing models for Lagos**

Conventional land use change modelling is done with the GIS-based Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) models. OLS and GWR models are both linear regression models which are subject to fundamental statistical assumptions, such as: linearity, normality, homoscedasticity, autocorrelation, and normality (Wheeler & Tiefelsdorf, 2005). The dependent variables of regression models can be represented with categorical variables (Menard, 1995); however, the assumptions of linear regression models are likely to be violated when categorical variables are used to represent the dependent variables (Pohlmann & Dennis, 2003). Based on the above reasons, the conventional GIS-based models cannot be solely relied upon for modelling land use change in Lagos.

In the search for robust predictive tools for modelling land use change in Lagos, it is important to note that urban systems are complex in nature and stochastic in behaviour; these



characteristics make them different from most social systems (Couclelis, 1988). Therefore, modelling an unregulated complex urban environment like Lagos may be unyielding without employing robust predictive tools that can realistically model their complexity, dynamism, and growth (Barredo et al., 2004). In this research, Cellular Automata (CA) have been adopted as the most appropriate tool for modelling land use change in Lagos. This is because CA models present the necessary structure for modelling complex adaptive systems like land use change (Torrens & O'Sullivan, 2001). Apart from the intrinsic ability of CA models to mimic macro scale urban environments from innumerable interactions among simple elements, their benefits also include: their easy visualisation of modelling results, their simplicity, their flexibility, their dynamic technique, as well as their affinity with the GIS and Remote Sensing (RS) (Torrens & O'Sullivan, 2001).

### **1.3 Research questions**

The research problem explained in section 1.1 will be tackled by exploring the conventional GIS-based methods and unconventional CA-based techniques. Table 1.1 presents a simple framework of the modelling methods that have been adopted in this thesis for modelling land use change in Lagos. These adopted methods for Lagos are: the GIS-based OLS and GWR; and the CA-based Logistic Regression (LR), Artificial Neural Networks (ANN), Support Vector Machine (SVM), and Fuzzy Support Vector Machine (FSVM) . The GIS-based OLS and GWR are the conventional methods of land use change modelling. The CA-based LR and ANN are common unconventional parametric and nonparametric methods respectively, while the CA-based SVM and FSVM are the proposed unconventional methods.

Table 1.1 *Conventional and unconventional land use change models adopted for Lagos*

Conventional land use change models (GIS-based)	Unconventional land use change models (cellular automata based)		
Ordinary Least Squares (OLS) & Geographically Weighted Regression (GWR)	Common parametric model	Common nonparametric model	The proposed nonparametric models
	Logistic Regression (LR)	Artificial Neural Networks (ANN)	Support Vector Machines (SVM) & Fuzzy Support Vector Machines (FSVM)

The main questions posed by this thesis in modelling land use change in Lagos are:

- (i) Does the conventional GIS-based GWR model perform better than the proposed SVM and FSVM models?
- (ii) Does the LR-based CA model yield better accuracies than the proposed SVM and FSVM models?
- (iii) Does the ANN-based CA model outperform the proposed SVM and FSVM models?
- (iv) Are the results of the CA based techniques better than those of the non-CA based techniques?

Note that the non-CA modelling mentioned on the list of research questions, is simply the classification application of any of the predictive models mentioned in Table 1.1 without involving the use of CA. CA modelling is usually implemented after the classification of the training data.

## 1.4 Specific research aims and objectives

The specific aims of this research in predicting land use change in Lagos are:

- The implementation of the GIS-based OLS and GWR modelling
- The implementation of the LR-based non-CA and CA modelling
- The implementation of the ANN-based non-CA and CA modelling
- The implementation of the proposed SVM-based non-CA and CA modelling
- The implementation of the proposed FSVM-based non-CA and CA modelling
- Comparing the results of the GIS-based GWR, LR-based CA, and ANN-based CA with the proposed SVM and FSVM based CA models
- The comparison of the CA-based results with those of the non-CA based results

The above stated specific aims will be accomplished with the following specific objectives:

### (i) GIS-based OLS and GWR modelling:

- To assess the effectiveness of the independent variables using the OLS Joint Wald Statistic
- To assess the statistical significance/contribution of each independent variable in the model and estimate  $R^2$  using the OLS model
- To investigate how the GIS-based models satisfy traditional statistical assumptions
- To predicting land use change in Lagos using the GWR model
- To assess the effect of excluding the insignificant independent variables from the OLS and GWR models
- To assess the effect of excluding each independent variable from the GWR model

## (ii) LR-based non-CA and CA modelling:

- To assess the significance of each independent variable in the model
- To assess the impact of excluding the insignificant variables from the model
- To assess the impact of excluding each independent variable one at a time, from the model
- To prediction land use change using the non-CA based model
- To prediction land use change using the CA based model

## (iii) ANN-based non-CA and CA modelling:

- To find an optimal number of neurons
- To prediction land use change using the non-CA based model
- To calculate the input weights of the independent variables using a one neuron ANN, and compare the calculated one neuron ANN weights to the calculated coefficients using the LR model
- To compare the cross-validation accuracies of the one neuron ANN with that of the LR model
- To prediction land use change using the CA based model

## (iv) SVM-based non-CA and CA modelling:

- To optimise SVM model parameters: penalty value ( $C$ ), polynomial order ( $d$ ), and gamma ( $\gamma$ )
- To prediction land use change using the non-CA based model
- To prediction land use change using the CA based model

- To assess the effect of the Moore and von Neumann neighbourhood functions on SVM and FSVM based CA modelling
  - To forecast future land use in 2015 and 2030
- (v) FSVM-based non-CA and CA modelling:
- Same as SVM
- (vi) To compare the results of the GIS-based GWR, LR-based CA, and ANN-based CA with the proposed SVM and FSVM based CA models:
- By using the McNemar's test to test whether there are significant differences between the GIS-based GWR and SVM/FSVM; LR and SVM/FSVM; and, ANN and SVM/FSVM predicted maps
  - By using the  $t$ -statistic to test whether there are significant differences between the calculated LR and SVM/FSVM; and ANN and SVM/FSVM Area Under Curve (AUC) indices
- (vii) To compare the CA-based with the non-CA based results:
- By using the McNemar test to test whether there are significant differences between the LR, ANN, SVM, and FSVM non-CA and CA predicted maps

## 1.5 Thesis domain: the concept of geocomputation

The domain of this thesis is geocomputation. Geocomputation is a relatively new discipline in geography. Geocomputaion is the process of applying computing technology to geographical problems (Rees & Turton, 1998, p. 1835). Geocomputation is also defined as the universe of computational techniques applicable to spatial problems (Couclelis, 1998, p.18).

Over the years, geographers have discovered the limitations in conventional spatial-based computer software and the need to develop their own computer models especially on a small scale. An example of the geocomputation effort is the IDRISI<sup>1</sup> software.

Contemporary sub-geography disciplines like the GIS, Remote Sensing (RS), and particularly geocomputation are groundbreaking subjects in the field of geography “so geography can emerge from its postmodern slumbers in a geocomputational world (Macmillan, 1998, p. 264). Geocomputation and GIS complement each other (Longley, 1998). Geocomputational outputs can be visualised and analysed in the GIS, while the GIS support the use of geocomputation subroutines to optimise their modelling tasks.

Some real-world modelling applications of geocomputation include: land use change (Okwuashi et al., 2008; Okwuashi et al., 2009a; Okwuashi et al., 2009b; Fkirin et al., 2009), fire (Li & Magill, 2001), and hazard (Erener et al., 2010).

## **1.6 Research contributions**

The main contributions of this research in modelling land use change in Lagos are:

- (i) The application of the GIS-based OLS and GWR modelling
- (ii) The application of the LR-based non-CA and CA modelling
- (iii) The application of the ANN-based non-CA and CA modelling
- (iv) The application of the SVM-based non-CA and CA modelling
- (v) The application of the FSVM-based non-CA and CA modelling

---

<sup>1</sup> IDRISI is a raster-based integrated GIS and RS software that was developed at the Department of Geography, Clark University, Worcester, MA, USA.

## 1.7 Structure of thesis

This thesis consists of eight chapters. Chapter 1 presents a general background of the thesis.

Chapter 2 discusses GIS and CA based models used for modelling land use change, and reviews on how these models were applied.

Chapter 3 explains the basic concepts of support vector machines including numerical examples.

Chapter 4 discusses land use development in Lagos and why the land use divers were selected. It also examines pre-independence and post-independence planning in Lagos.

Chapter 5 presents the application of the GIS-based OLS and GWR models to land use change modelling of Lagos. This chapter illustrates the conventional technique including some of their limitations.

Chapter 6 presents the application of the LR and ANN based CA and non-CA techniques to land use change modelling of Lagos. This chapter elucidates the virtues of using the LR model for land use change modelling. It also examines the relationship between a typical LR model and a one neuron ANN model.

Chapter 7 presents the applications of the proposed SVM and FSVM based CA and non-CA approaches to land use change modelling of Lagos. It also discusses the influence of the neighbourhood functions on SVM and FSVM results; and illustrates how future land use can be modelled using the SVM and FSVM models.

Chapter 8 presents the statistical comparison of results from the GIS-based GWR, LR, and ANN with respect to the proposed SVM and FSVM models. It also presents the statistical comparison of the results from CA and non-CA based modelling. This chapter also discusses the benefit of the research to Lagos and Nigeria, contribution of the research to science, and recommendations for further research.

## CHAPTER 2

# GEOGRAPHIC INFORMATION SYSTEMS AND CELLULAR AUTOMATA BASED MODELS

## 2.1 Preamble

This chapter presents the basic concepts of the GIS-based models for modelling land use change in Lagos, and also the known methods of CA calibration. This chapter will focus on: the basic concepts of these models, their merits and demerits, and whether they were considered for modelling land use change in Lagos or not. It will also discuss the rationale for integrating GIS with CA for land use change modelling.

## 2.2 Geographic information systems based models

### 2.2.1 Geographic information systems.

GIS are “powerful set of tools for collecting, storing, retrieving, at will, transforming, and displaying spatial data from the real world for a particular set of purposes” (Burrough & McDonnell, 1998, p. 11). A United States based conglomerate, Environmental Systems Research Institute (ESRI) is the sole provider of the universal GIS software called “ArcGIS.” ArcGIS is the most widely used GIS software in the world. ArcGIS was designed to handle all geo-based real world applications.

GIS describe real world data in two ways: (i) graphical/map interface and, (ii) database management system. The graphical interface (called layers) is an image representation of the data, while the database management system (called attributes) is used to store information about the image. GIS can be applied in banking, telecommunication, land use change, sea level rise, oil



spill, deforestation, election monitoring, hazards prediction, military, tourism, real estate, and agriculture. Spatial features in a GIS are represented as either in vector or raster formats. Vectors (called features) are used to represent lines and polygons. For example, a road network will be represented as a line, while land parcels are represented as polygons. A raster layer is a pixel-by-pixel representation of an image. The advantage of images in vector format unlike their raster counterpart is that, they can easily be edited on-screen. Raster files are more stable, and are usually used for performing numerical analysis. Vector files are less stable and can easily be affected by distortion. Remote sensing data and analogue maps are the major primary sources of data into the GIS (Ehlers et al., 1989; Lunetta et al., 1991; Burrough & McDonnell, 1998; Lo & Yeung, 2007).

### **2.2.2 Geographically weighted regression and ordinary least squares.**

Conventional land use change modelling is implemented in the ArcGIS environment using the OLS and GWR models. Even though ArcGIS maintains the reputation of being the de facto global model for land use change modelling, several local land use models have been developed across the globe. These models are adapted to specific areas of application. An example of a local land use model is the Conversion of Land Use and its Effects modelling framework (CLUE), specifically developed for applications in Asia and North America. It has been applied in China, Indonesia, and Ecuador (Veldkamp & Fresco, 1996; Verburg et al., 1999). Others are, the Metropolitan Integrated Land Use System (METROPILUS), which has been applied in the United States cities of Chicago, Los Angeles, and Atlanta (Putman, 1992; 1993; 1995); and Clarke's urban growth models (UGM and SLEUTH) which have been applied in Chicago, USA (Clarke & Gaydos, 1998) and Lisbon in Portugal (Silva & Clarke, 2002).

The GWR (Fotheringham et al., 2002) is the local equivalent of the global OLS. OLS and GWR are both linear regression models. The global OLS model can be expressed mathematically

as,

$$y_i = \beta_0 + \sum_p \beta_p x_{ip} + \varepsilon_i \quad (2.1)$$

where  $y_i$  are the dependent variables,  $\beta_0$  denotes the intercept,  $\beta_p$  are the slope coefficients for the  $p$  variable,  $x_{ip}$  denotes the value for a  $p$  th variable for  $i$  number of observations, and  $\varepsilon_i$  represents the error parameter. Unlike the GWR model, OLS assumes that  $\beta_p$  is stationary or homoscedastic. This is a major difference between the OLS and the GWR (Fotheringham et al., 2002; Mennis, 2006). The GWR model can be expressed mathematically as,

$$y_i = \beta_0(u_i, v_i) + \sum_q \beta_q(u_i, v_i) x_{iq} + \varepsilon_i \quad (2.2)$$

where  $\beta_0$  denotes the intercept,  $\beta_q$  represents the slope coefficients for the variable  $q$ ,  $x_{iq}$  denotes the value for a  $q$  th variable for  $i$  number of observations, and  $\varepsilon_i$  represents the error parameter, and  $(u_i, v_i)$  stands for the coordinates of the  $i$  th location for an  $i$  observations. GWR takes the effect of spatial dependency into consideration and assumes that  $\beta_q$  is non-stationary or heteroscedastic. The non-stationarity of  $\beta_q$  means that the solutions of  $\beta_q$  vary across the globe for the same values of  $x_{iq}$ . The solution of  $\beta_q$  is affected by the locations  $(u_i, v_i)$  where  $x_{iq}$  were actually observed (Fotheringham et al., 2002).

The GWR is less affected by spatial autocorrelation (Moran, 1948) than the OLS, because the presence of the local parameter  $(u_i, v_i)$  of the study area in the GWR mitigates the effect of spatial dependency in the model (Platt, 2004; Windle et al., 2010). OLS assumes that explanatory

variables are stationary across space. The GWR is a local linear regression model designed to mitigate spatial effects in the data. Unlike the OLS, the GWR assumes spatial non-stationarity across space (Brunsdon et al., 1998; Fotheringham et al., 2002). GWR incorporates geographical coordinates of observed points (Dormann et al., 2007), such that observed points closer to predicted points are assigned more weights than farther points. Comparisons done between GWR and OLS have shown that GWR furnishes better results than OLS because it is far less compromised by the effect of spatial dependency (Zhang & Gove, 2005; Shi et al., 2006; Kupfer & Farris, 2007; Osborne et al., 2007; Kimsey et al., 2008).

In spite of the merits of the GWR model over the OLS model, it is still subject to fundamental statistical assumptions (just like the OLS), that there is: (i) a normal distribution; (ii) a linear relationship between the dependent and independent variable; (iii) no multicollinearity between the independent variables or no exact correlation between the independent variable; (iv) no spatial autocorrelation; (v) homoscedasticity or that the variance between all the independent variables is equal (Leung et al., 2000; Fotheringham et al., 2002; Wheeler & Tiefelsdorf, 2005; Farber & Páez, 2007). The presence of spatial autocorrelation in linear regression results is a result of the effect of spatial dependency or the first law of Geography (Tobler, 1970). It is important to determine the presence of spatial autocorrelation in a model using the model residuals; which are the discrepancies between the observed and the predicted data. The ArcGIS Moran I tool can be used to test for spatial autocorrelation in the data. Three inferences are obtained from the autocorrelation test: clustered, dispersed, and random. Clustered indicates that the data are close together or concentrated. Dispersed implies that the data are far apart. Random means that the data are spatial-autocorrelation free. It is desirable that the data be random. Moran I values close to +1 indicate clustered; values close to -1 imply dispersed, while values close to zero mean that the data is random. The presence of significant

spatial autocorrelation in the data implies that at least one important explanatory variable might be missing from the model (Moran, 1948; Lo & Yeung, 2007).

The main advantages of the GIS-based OLS and GWR models is that they have tremendous explanatory properties, such that the significant contribution of the independent variables in the models can be explored, as well as tests for detecting multicollinearity, normality, non-homoscedasticity, spatial autocorrelation, and determining  $R^2$ . Another advantage that they have is that they yield high accuracy when used for modelling land use change (Noresah & Ruslan, 2009; Thapa & Murayama, 2009).

The disadvantage of the OLS and GWR models is that they are based on the principle of linear regression, therefore the basic assumptions that govern traditional statistical linear regression models must be satisfied before results from the OLS and GWR can be trusted. If these assumptions are not satisfied, the significance tests and estimated standard errors of the model parameters may not be valid (Anselin & Griffith, 1998; Fox et al., 2001).

In spite of the aforementioned problems with the GIS-based OLS and GWR models; for the purpose of comparison with the SVMs, they were adopted for modelling land use change in Lagos being the conventional techniques for modelling land use change.

The next section presents the concept of CA models and some of their major applications.

## **2.3 Cellular automata based models**

### **2.3.1 Cellular automata.**

CA are discrete dynamic mathematical systems which were introduced by Ulam (1952) and von Neumann (1966) who initially used cellular automata models to study mathematical compatibility theory and biological systems (Wolfram, 1983; Wolfram, 2002). The bottom-up simulation attribute of CA models makes them invaluable tools for dynamic process modelling. CAs are comprised of four basic components: cells, states, neighbourhoods, and transition rules (Webster

& Wu, 2001). Conceptually, there are five attributes that characterise CA systems: (i) a uniform regular lattice; (ii) a cell; (iii) the state of a cell; (iv) the neighbours of a cell; and (v) the transition rules. The uniform regular lattice can be in  $n$ -dimensions (O'Sullivan & Torrens, 2000).

Common examples are: one and two dimensional cellular automata (see Figures 2.1-2.3).



Figure 2.1 *A hypothetical one-dimensional cellular automata (yellow = urban cells; green cells = non-urban cells)*

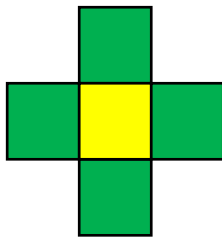


Figure 2.2 *A hypothetical von Neumann neighbourhood image (is a two-dimensional cellular automata; yellow = urban cell; green = non-urban cells; where each cell transition depends on four neighbours)*

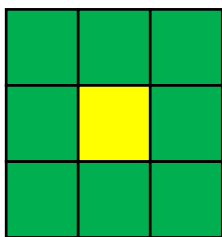


Figure 2.3 *A hypothetical Moore neighbourhood image (is a two-dimensional cellular automata; yellow = urban cell; green = non-urban cells; where each cell transition depends on eight neighbours)*

A cell is the individual unit contained in the uniform regular lattice. For land use, the state of each cell can be urban/non-urban or 0, 1, 2, 3, 4, 5, water, vegetation, and road. The neighbours of a cell are the cells surrounding each cell in the regular lattice. The von Neumann and Moore neighbourhoods are the two universally used two-dimensional cellular automata neighbourhoods (Maerivoet & De Moor, 2005). The von Neumann neighbourhood and Moore neighbourhood transition systems require four and eight surrounding cells respectively (see Figures 2.2 and 2.3).

The transition rules are predefined conditions that determine the future state of every cell in the lattice. These rules must be fulfilled before any cell transition is possible. A cellular automaton can be expressed mathematically as,

$$q_{ij}(t+1) = f(\Omega_{ij}(t), q_{ij}(t), T) \quad (2.3)$$

where  $q_{ij}(t+1)$  is the state of the cell at the time  $t+1$ ;  $q_{ij}(t)$  is the state of a cell at the time  $t$  (e.g.  $q_{ij}(t)=0$  may imply non-urban and  $q_{ij}(t)=1$  urban);  $\Omega_{ij}(t)$  represents the neighbourhood influence on a cell; and  $T$  denotes a set of transition rules governing the process (O'Sullivan, 2001).

CA can either be deterministic or nondeterministic. For deterministic CA, the same results are obtained given the same starting conditions. The commonest example of the deterministic CA is the Game of Life (GOL). GOL is a totalistic<sup>2</sup> binary CA based on the Moore neighbourhood. Every cell in the lattice has two states: Live (L) or Dead (D). The system is basically controlled by two transition rules: (i) Rule 1: any D cell that has 3 L neighbours

---

<sup>2</sup> A totalistic cellular automata (Wolfram, 1983) is a type of CA whose transition rules are based on the neighbourhood influence around each cell.

automatically transits to an L cell; and (ii) Rule 2: an L cell can only remain live if it has 2 or 3 live neighbours (Gardner, 1970; Waldrop, 1992; Batty, 1997; Holland, 1998).

Nondeterministic or stochastic CA depends on probabilities or random trials, such that, different outcomes are produced given the same starting conditions. The fact that real world phenomena are nondeterministic may make nondeterministic CA systems more desirable than the deterministic CA. The next section reviews the major models that have so far been applied to CA calibration.

### **2.3.2 Cellular automata calibration methods.**

#### ***2.3.2.1 Heuristics-based modelling.***

Heuristic CA models are essentially based on trial-and-error (Wu, 2002), using simple and flexible transition rules to capture complex real-world behaviours (O'Sullivan, 2001; Batty, 2005; Torrens & Benenson, 2005; Benenson, 2007). These rules are mainly IF, THEN, ELSE... (Batty, 1997). Heuristic CA models do not employ machine learning algorithms. In some cases formulas were used to represent real world phenomena (Almeida et al., 2003a; Barredo et al., 2004; Caruso, 2005); in other cases weighted probabilities (Sietchiping, 2004) were used.

There is no universal formula for representing a heuristic CA model. The commonest heuristic approach neither uses formulas nor weighted probabilities, but just simple IF... THEN...ELSE transition rules. Urban systems are complex, therefore finding a globally accepted formula may be unrealistic (Batty, 1997).

The advantage of heuristic CA models is that they can easily be adapted to finding optimal procedures to represent complex real-world patterns, through the use of simple rules, which makes their use very attractive.

But, despite their appealing methodologies, they have the disadvantage of lacking computational validity; their assumptions are biased, too flexibly defined, and their results cannot

be mathematically validated (Wu, 2002). These reasons make results from heuristic models a subject of academic debate. Due to the above reasons this technique was not considered for modelling land use change in Lagos.

### ***2.3.2.2 Agent-based modelling.***

Agent Based Models (ABMs) (Benenson, 1998; Nara & Torrens, 2005; Batty, 2005; Polhill et al., 2007) are models of autonomous actors involved in an activity. Agents can represent humans, robots, as well as real world activities/phenomena. Agents are, (i) autonomous: they do not require human intrusion to initiate any action; (ii) interdependent: an agent can mingle with other agents; (iii) governed by simple rules: agents explore simple rules to model complex patterns; (iv) adaptive: they are inventive, replicating, and capable of imitating or learning (Macy & Willer, 2002). ABMs are well adapted to cellular automata models. ABMs can be implemented in ArcGIS using the Agent Analyst. Other popular agent based models are: StarLogo, SWARM, REPAST, NetLogo, and ASCAPE (Nara & Torrens, 2005).

ABMs, also known as multi-agent models, are veritable tools for modelling complex relationships among actors unlike CA models they are independent of direct human intervention and can capture more complex relationships among actors both at micro and macro levels. Integrating agent based models with CA means that CA models can account for dynamic simulation of space, while agent based models handle decision making and complex relationships among agents (White & Engelen, 2000).

ABMs are relatively new to land use change modelling (Parker, 2005), the advantage of ABMs is that they are highly enriched algorithms that are well suited for modelling complex systems like urban systems.



The disadvantage of ABMs is their inability to evaluate model constraints and results, since the rules driving the activities of agents are flexible and complicated (O'Sullivan, 2004). For this reason ABMs were not considered for modelling land use change in Lagos.

### ***2.3.2.3 Fractal-based modelling.***

Fractals are used to represent the geometrical characteristics of real world objects; such that an object can be broken down or subdivided into several parts at reduced scales. These parts of the whole remain self-similar (like the whole), irrespective of the scale. Fractals are: (i) self-similar; (ii) irregular or fragmented no matter the scale used for the investigation; (iii) they are iterative; (iv) fractionally dimensioned; and (v) over a large range, fractals could contain elements that are distinct with varied scales (Mandelbrot, 1967; Mandelbrot, 1982). White and Engelen (1993) explained that the integration of CA with fractal models can be used simultaneously to evolve order and complexity in urban structure; and that fractal patterns of landforms can be generated using simple rules to exhibit complex spatial behaviour.

A fractal structure can be investigated by varying the scale and observing changes in the object form or by varying the object size while the scale is fixed (Batty & Longley, 1994). A fractal dimension can be calculated using either the box-counting or radial analysis method. Shen (1997) illustrated how a box counting method can be used to estimate the fractal dimension of 30 urban transportation networks in conjunction with the correlation between urban population and fractal dimension of space. Batty and Longley (1994) used a Diffusion Limited Aggregation (DLA) model to explain urban phenomenon. The DLA model was used to generate fractal structures that are analogous to real cities. Marske et al. (1998) criticised the DLA model for creating just one large cluster rather than several central spatially distributed hierarchical city patterns.

The advantage of fractals is that they have the ability to evaluate cellular automata generated similarity amid simulated spatial patterns (Yeh & Li, 2001). Another advantage of fractals is their ability to represent complexity with simple and coherent patterns (Brown & Witschey, 2003). Fractal modelling is still new to land use change research.

The disadvantage of fractal modelling is that they are sometimes very complicated to interpret because various spatial structures may be represented by the same fractal dimension (White & Engelen, 1993). For the above reason, fractal modelling was not considered for modelling land use change in Lagos.

#### ***2.3.2.4 Fuzzy logic based modelling.***

Fuzzy logic, being a multivalued logic, deviates from the Boolean logic by replacing categorical evaluations like true/false and high/low, with continuous evaluations such as: true, fairly true, less fairly true, not entirely false, false; and high, fairly high, fairly low, low respectively (Zadeh, 1962; 1965; 1971). The idea of fuzzy logic is basically to make computers make decisions like humans (Zadeh, 1984).

According to Zadeh (1965; 1971), a fuzzy set can be expressed as follows: Let  $X$  represent objects whose generic element is in  $x$ , therefore  $X = \{x\}$ . In that case, a fuzzy set  $A$  in  $X$  can be expressed as a set of ordered pairs,

$$A = \{(x, \mu_A(x)) \mid x \in X\} \quad (2.4)$$

where  $\mu_A(x)$  denotes the grade of membership  $x$  in  $A$ , where  $x$  is a real number in  $[0, 1]$ .

Wu (1996; 1998) modelled urban development using the crisp theory based on fuzzy logic controlled transition rules. Liu (2001) criticised the idea of Wu (1996; 1998) which represented stages of land use development with sharp boundaries like urban/non-urban; and instead continuous boundaries like urban/semi-urban/near-non-urban/urban were introduced. A corroborating postulate had earlier been stated by Herbert and Thomas (1997), that urban change is a continuous process that trends a logistic curve. Liu (2001) elicited fuzzy logic and fuzzy set based CA transition rules, to examine urban development in Sydney, Australia. Logistic regression was used to construct the fuzzy membership function used to describe the categories of urban development: very quick development, quick development, slow development, very slow development, and extremely slow development. The merit of this technique is its ability to determine the stages of land use development using continuous land use boundaries. Another merit is the introduction of stochasticity into the model to portray a more realistic random pattern of urban development. The challenge of the approach of Liu (2001) is the use of heuristics like very quick development, quick development, slow development, very slow development, and extremely slow development for determining the transition of non-urban cells to urban which may make it difficult to mathematically validate the model.

Mantelas et al. (2007; 2008) presented a deterministic method of eliciting CA transition rules by using variables consisting of partially overlapping qualitative fuzzy sets. Linguistic variables were used to obtain qualitative numerical information which were used to construct fuzzy membership functions; and IF... THEN... ELSE rules were then used to obtain various stages of urban development. The merit of this work is the use of flexible CA rules to determine cell transition. Its demerit is that the use of heuristics and qualitative data mining may produce results that overfit the data.

Al-Ahmadi et al. (2009) described a nondeterministic technique where two automated methods of calibration (genetic algorithm and simulated annealing) were used to mine fuzzy logic

driven CA transition rules. Heuristics (IF... THEN... ELSE rules) were used to elicit various urban development scenarios of Riyadh in Saudi Arabia. The merits of this technique are that the introduction of stochasticity in the calibration will help capture the randomness in real world land use development; while the use of machine learning algorithms for calibrating the fuzzy logic based model will help improve the validity of the calibration. The demerit of the experiment is that the use of heuristics may make the model overfit, and reduce its validity in making future predictions.

In conclusion, the main virtue of fuzzy logic based CA calibration is its ability to model uncertainty in real world data with continuous decision boundaries. Its major limitation in predicting future land use is the use of heuristics, which may produce models that are overfitted. Due to the aforementioned issues the fuzzy logic based CA was not considered for modelling land use change in Lagos.

### *2.3.2.5 Spatial statistics based modelling.*

#### *2.3.2.5.1 Markov chain based modelling.*

For a particular land use change process, a Markov chain model can be expressed as:

$$b^n = b^{n-1}P \quad (2.5)$$

$P$  is the transition matrix; it is a stochastic matrix whose elements must be non-negative and rows sum up to 1;  $n$  represents discrete time steps; and  $b^n$  is a resulting vector that bears the final probabilities of the land use distribution at the end of the process. For  $n = 0$ ,  $b^0$  is a vector that represents the distribution of land use states at the beginning of the process (1 = urban and 0 = nonurban); it is the initial condition.

For example, given a stochastic matrix  $P$  derived by observing historical land use change and  $b^0$ , let us compute  $n = 3$ :

$$b^0 = [1 \ 0 \ 0 \ 0], P = \begin{bmatrix} 0.10 & 0.4 & 0.2 & 0.3 \\ 0.6 & 0.1 & 0.2 & 0.1 \\ 0.11 & 0.19 & 0.4 & 0.3 \\ 0.05 & 0.35 & 0.2 & 0.4 \end{bmatrix},$$

$$b^1 = b^0 P = [0.1 \ 0.4 \ 0.2 \ 0.3],$$

$$b^2 = b^1 P = [0.2870 \ 0.2230 \ 0.2400 \ 0.2500],$$

$$b^3 = b^2 P = [0.2014 \ 0.2702 \ 0.2480 \ 0.2804].$$

Markov chain models are compatible with CA models for simulating urban systems (Clarke et al., 1997). Antoni (2001) explained that a Markov chain model can be used to generate transition and simulation matrices using historical land use data, but such data require the integration of CA to determine the land use category of each pixel. In that case, categories of land use cells will depend on the neighbouring cells surrounding each cell. Transition rules are eventually required to decide the category of each land use cell.

Ozah et al. (2010) used a Markov chain driven cellular automata to predict land use change in the Lake Chad Basin in Nigeria. Two land use epochs were used to predict land use in 2011 and 2013. The transition probabilities were obtained by observing the trend in land use using the two land use epochs. A deterministic CA model was used to mine Markov chain constrained transition rules, by incorporating the influence of neighbouring cells into the model. A similar methodology was described by Demirel and Cetin (2010) who predicted land use in 2017 for Istanbul, Turkey.

The merit of Markov chain models is that they offer the simple use of the observed historical change between land use epochs to generate transition probabilities that will represent

future land use forms. Another merit of Markov chain models is that they are free from the rigour of machine learning since no training examples are required to calibrate the model.

The demerit of Markov chain models is that the significance of each land use driver in the model cannot be assessed (unlike in OLS and logistic regression models), therefore they lack the power to explain the influences/relationships of the causal factors that drive change (Baker, 1989). Consequently, the Markov chain model was not considered for Lagos due to the above mentioned demerits of the model.

#### *2.3.2.5.2 Logistic regression based modelling.*

The Logistic Regression (LR) model is the most commonly used parametric model for modelling land use change (Verburg et al., 2004). A mathematical illustration of the LR model is presented in section 6.2.1. The earliest logistic regression based-CA calibration was done by Wu (2002). Wu (2002) extracted land use descriptors from historical land use data to train the logistic regression model. The sigmoid function was used to derive land use transition probabilities. Stochasticity was introduced into the model to capture the randomness inherent in real world data. The neighbourhood influence was used to obtain a binary map of the study area. Almeida et al. (2003b) presented a similar methodology for predicting land use change in Sao Paulo, Brazil.

The advantages of LR models are that relationships among land use drivers can be examined (McMillen, 1989). They can help identify the contributions of each independent variable in the model; they can assess the degree of confidence of each variable; they can be applied to model nonlinear relationships; and unlike OLS models they are not required to satisfy statistical assumptions for the model to be: normal, linear, homoscedastic, free of multicollinearity, and spatial autocorrelation free (Pohlmann & Dennis, 2003).

The disadvantages of LR models are that, unlike OLS models they do not furnish  $R^2$  value which makes it difficult to validate LR outputs; LR models require large amount of training data

to calculate their model coefficients before their results could be trusted (Pohlmann & Dennis, 2003).

Despite the above mentioned problems with the LR model and for the purpose of comparison with results from the proposed SVM and FSVM models, the LR model has been adopted in this research for modelling land use change in Lagos being the most widely used parametric model for modelling land use change (Verburg et al., 2004) and also its ability to model nonlinear urban systems.

#### ***2.3.2.6 Artificial neural network based modelling.***

The architecture of Artificial Neural Networks (ANNs) or neural networks consist of three main components the input layer, hidden layer, and the output layer (Figure 2.4). Each neuron in the input layer represents a land use variable. The hidden layer is the engine room of the neural network; it consists of  $n$  neurons ( $n = 1, 2, 3, \dots$ ). The output layer consists of just a single neuron (Almeida et al., 2008).

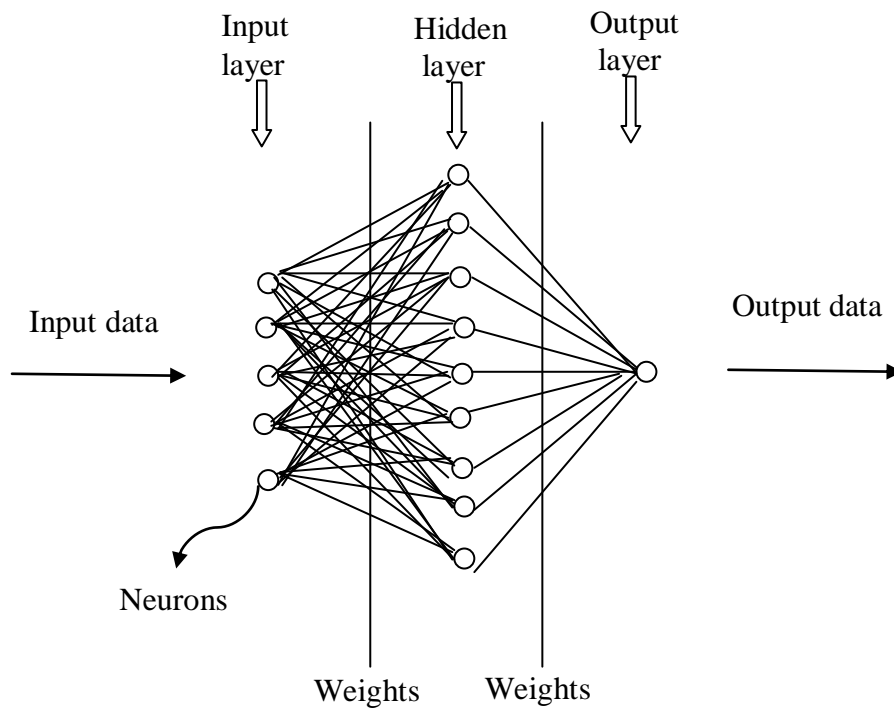


Figure 2.4 A single output neural network architecture. Adapted from Li and Yeh (2002, p. 328)

The input data are essentially scaled to keep the output in the range of either  $[0, 1]$  or  $[-1, 1]$ . This is done before the input data are fed into the network (Gong, 1996). The artificial neural network has been the most commonly used nonparametric model for CA calibration. The first ANN-based CA calibration was done by Li and Yeh (2002) after Clarke et al. (1997) recommended that the nonparametric ANN may offer viable potentials for modelling complex urban systems. The feed-forward backpropagation neural network is the most commonly used technique for CA calibration.

Li and Yeh (2002) introduced a novel approach for calibrating CA using ANN. Land use variables were extracted with the GIS and small sized training sets were used to train the network. Stochasticity was introduced in the model to reflect the randomness in actual land use development. A sigmoid function was used to generate stochastic probabilities of land use evolution. Nonlinear CA transition rules were based on probability thresholds that served as



benchmarks for determining different scenarios of urban development. This approach was replicated by Almeida et al. (2008) in Sao Paulo, Brazil.

The advantage of ANNs is that they are highly robust for modelling complex systems like urban systems (Li & Yeh, 2002). Another advantage ANN models have is that they are free from satisfying statistical requirements for the model to be: normal; linear; homoscedastic; free of multicollinearity; and spatial autocorrelation free (Olden & Jackson, 2001; Guan et al., 2005; Zhou & Kang, 2005).

Despite the ANN model's ability to handle complex nonlinear spatio-temporal data, they have the following disadvantages: (i) they are regarded as black box models: because they lack explanatory properties which means that the relationships and contributions of land use descriptors are obscured by the model; and (ii) unlike the SVMs, only a few ANN parameters can be optimised (Guan et al., 2005; Zhou & Kang, 2005).

In spite of these highlighted demerits of the ANN model and for the purpose of comparison with the SVMs, the ANN model has been adopted in this research for modelling land change in Lagos being the most commonly used nonparametric model for CA calibration; and also for its robustness in modelling complex urban systems.

## **2.4 Integration of geographic information systems and cellular automata models**

As previously explained in chapter one, CA models have been adopted for land use change modelling because of their simplicity, dynamic properties, and inventive bottom-up approach (Clarke & Gaydos, 1998). Another vital attribute of the CA is their compatibility with remote sensing and geographic information systems (Torrens & O'Sullivan). Conventional GIS models are not flexible enough (Wagner, 1997). They are not well suited for dynamic modelling like CA models; this makes them more or less static in operation (Longley & Batty, 2003). Another

limitation of conventional GIS models is their inability to be easily adjusted to perform complex numerical analysis (Wagner, 1997; Couclelis, 2002).

Coupling GIS and CA models has helped improve dynamic spatial modelling (Park & Wagner, 1997), especially in the domain of land use change research. A loose coupling technique can be used to synergise the analytical capabilities of CA models and the spatial attributes of the GIS (Bivand & Lucas, 2000; Almeida et al., 2002; Sietchiping, 2004; Kocabas & Dragicevic, 2007). The loose coupling synergy between the GIS and CA models is such that both systems can interact freely yet function independently. Couclelis (1997) highlighted that the integration of GIS and CA models is such that one of them makes up for the limitations of the other. GIS could be used to perform functions such as, data preparation; georeferencing; overlaying; scaling; rasterisation; digitisation; Euclidean distance estimation; data storage; data management; data retrieval; neighbourhood calculation; and visualization. CA models could perform functions that cannot easily be performed with the GIS. Functions such as, weighting of parameters; data sampling; training and testing; dynamic calibration; and iteration. Tight coupling GIS and CA models is also feasible. Local land use models (Clarke et al., 1997; Wadell, 2002) are based on a tight coupling strategy. A tightly coupled GIS and CA model implies that both models are embedded in a single system.

## **2.5 Summary**

This chapter presented the basic concepts and review of the conventional GIS-based OLS and GWR models; and also reviewed notable CA calibration techniques. The rationale for the integration of GIS and CA systems were also presented. The next section will present the mathematic illustrations of the SVM and FSVM models.

## CHAPTER 3

## SUPPORT VECTOR MACHINES

## 3.1 Preamble

The aim of this chapter is to create a better understanding of the SVM and FSVM experiments presented in chapter 7.

The concept of the Support Vector Machine (SVM) was introduced by Cortes and Vapnik (1995). SVMs constitute an assemblage of supervised learning algorithms initially applied to classification problems, but later extended to regression analysis (Vapnik, 1995; Gunn, 1998). SVM algorithms are based on statistical learning theory (Vapnik, 1995; Drucker et al., 1999). SVMs employ the principle of Structural Risk Minimization (SRM), which makes them robust and independent of underlying data distributions (Joachims, 1999).

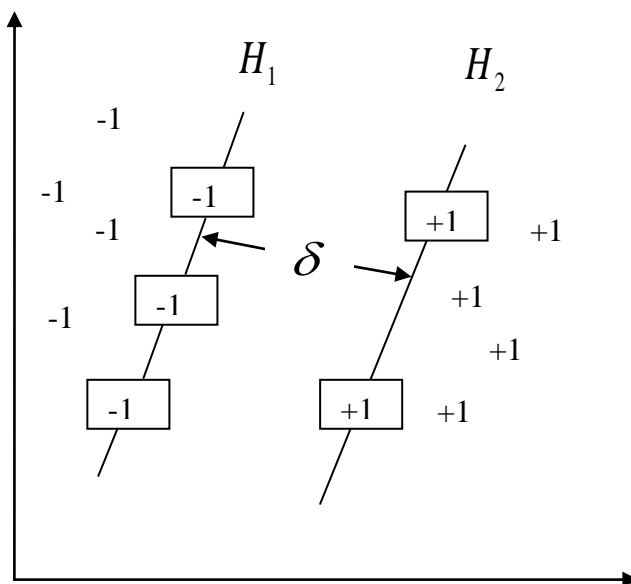


Figure 3.1 *Hyperplane with the maximum separation for a linear data. Adapted from Ivanciuc (2007, p. 318)*

SVMs are primarily binary classifiers. With reference to Figure 3.1, SVMs seek a unique hyperplane that has a maximum margin,  $\delta$ . In reality an infinite number of hyperplanes can be used to separate the data. Hyperplane  $H_1$  separates class -1 while  $H_2$  separates class +1 (Ivanciuc, 2007).

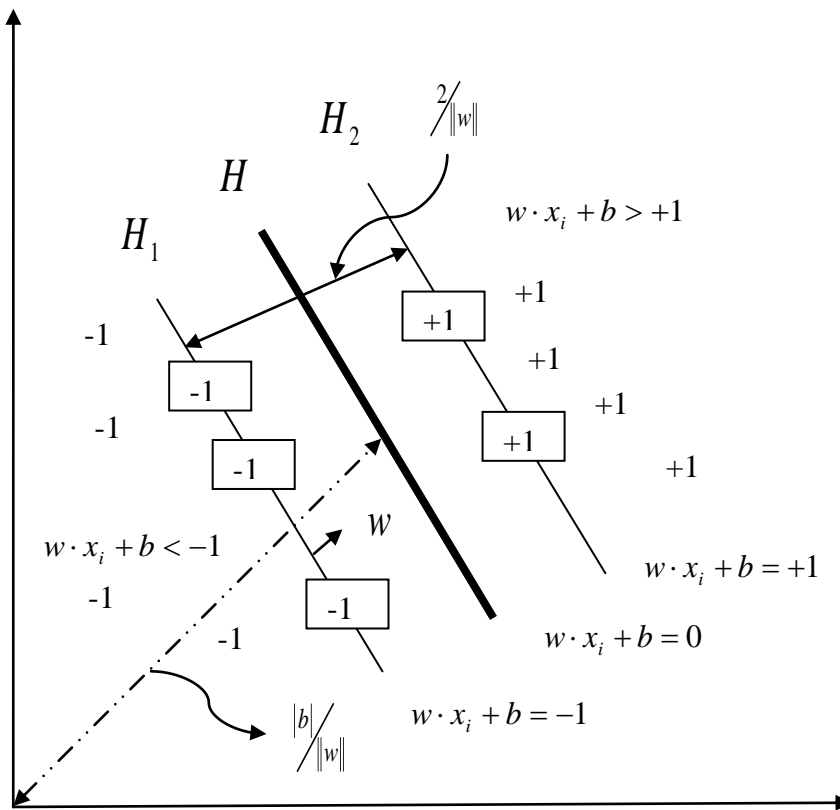


Figure 3.2 The separating linear hyperplanes for a linear data. Adapted from Ivanciuc (2007, p. 323)

The three points of -1 and two points of +1 in the rectangles located on hyperplanes  $H_1$  and  $H_2$  are called “support vectors” (see Figures 3.1 and 3.2). Looking at Figure 3.2, support vectors are the vectors that are nonzero that are defined by either hyperplane  $w \cdot x_i + b = -1$  for

support vectors that are in class -1 or hyperplane  $w \cdot x_i + b = +1$  for support vectors that are in class +1. The hyperplane H satisfies the linear classifier

$w \cdot x_i + b = 0$ . Other correctly classified points of -1 and +1 satisfy  $w \cdot x_i + b < -1$  and  $w \cdot x_i + b \geq 1$  respectively. The distance from the origin to hyperplane H is  $\frac{|b|}{\|w\|}$ . The maximum separation of  $\frac{2}{\|w\|}$  is required for a binary separation;  $b$  is a scalar; while  $w$  is a vector (Ivanciuc, 2007).

As mentioned in section 3.1, SVM is based on the principle of SRM; therefore the next section explains the principles of the statistical learning theory, SRM, and Vapnik Chervonenkis dimension; and the relationships amongst them.

### 3.2 Statistical learning theory, structural risk minimisation, and Vapnik-Chervonenkis dimension

Given observations  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , where  $x_i \in \mathfrak{R}^N$  and  $y_i \in \{-1, +1\}$ ;  $x_i$  are observed descriptors and  $y_i$  represents the target variables or decision space. Let us assume that the training data were drawn independently from probability distribution  $P(x, y)$ , which is not known but fixed. This forms the basic underlying assumption of Statistical Learning Theory (SLT). The objective of the statistical learning is therefore to seek a classifier  $y = f(x)$  that can classify/predict the test dataset that were also generated from  $P(x, y)$  (Vapnik, 1995).

For a further underlying understanding of the nature of SVMs, it is important at this juncture to introduce the concept of Vapnik-Chervonenkis dimension ( $d_{VC}$ ) as it relates to SRM. Let us map  $x \rightarrow y$  using a classifier  $f(x, w)$ . The goal of Vapnik-Chervonenkis dimension is to measure the capacity of  $f(x, w)$  to map  $x \rightarrow y$  based on the adjustment of the vector  $w$ . If the test and training data are identically distributed with cumulative probability distribution  $P(x, y)$  and drawn independently, the test set will have an expected risk (Vapnik, 2000):

$$R(w) = \int \frac{1}{2} |y - f(x, w)| dP(x, y) \quad (3.1)$$

Note that, conventionally the expected risk is usually replaced with the empirical risk (for a  $k$  training samples because of the unavailability of  $P(x, y)$  in typical cases):

$$R_{emp}(w) = \frac{1}{k} \sum_{i=1}^k |f(x_i, w)| \quad (3.2)$$

The objective of machine learning is to minimize the empirical risk,  $R_{emp}$ . If we rely on just minimizing the empirical risk for training the samples, we will end up having a learning machine with too much capacity, which will lead to overfitting. The way out of this problem is to introduce the principle of SRM. SRM seeks a classifier that finds a middle ground between low empirical risk and low capacity; given that  $0 \leq \eta \leq 1$ , with probability  $1 - \eta$ , therefore the following bound subsists for the expected risk (Vapnik, 2000):

$$R(w) = R_{emp}(w) + \sqrt{\left( \frac{d_{VC} [\log(2k / d_{VC}) + 1] - \log(\eta / 4)}{k} \right)} \quad (3.3)$$

The right-hand side of equation 3.3 is known as the “risk bound” and the second term is called “VC confidence interval”. If the number of training samples increases the VC confidence will decrease; and if the  $d_{VC}$  increases the VC confidence also increases. Therefore, the higher the dimensionality of the feature space ( $d_{VC}$ ), the higher the number of training points needed for a good classification for a given linear learning machine (Vapnik, 2000). By adopting the principle of SRM, SVMs offer a robust approach for pattern recognition, by minimising the problem of overfitting (Gunn, 1998).

In conclusion, two objectives are paramount when classifying a binary data: (i) the chosen classifier is the classifier that has the smallest empirical risk: this classifier must completely separate any two classes distinctly; and (ii) the ideal classifier should have the smallest VC dimension: this condition is realisable with a classifier with the maximum separation margin (Ivanciuc, 2007).

### 3.3 Support vector machine classification

Given a binary classification problem that belongs to classes -1 and +1 respectively; these two classes can be separated with a linear hyperplane (see Figure 3.3). To separate these two sets of objects, we need to choose a few training samples. Now, let us assume that our training set has  $n$ -training samples, that is,  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , where  $x_i \in \mathfrak{R}^N$  is an  $N$  dimensional vector that belongs to one of classes  $y_i \in \{-1, +1\}$ . The stated binary classification problem can only be separated using a linear decision function (Vapnik, 2000):

$$f(x) = w \cdot x + b \quad (3.4)$$

where  $w \in \mathfrak{R}^N$  is a vector that determines the orientation of our desired hyperplane required for the separation and  $b \in \mathfrak{R}$  is called the “bias.”

We can see from Figure 3.3 that our optimal hyperplanes needed to separate the two objects is,

$$y_i(w \cdot x + b) \geq 1 \quad (3.5)$$

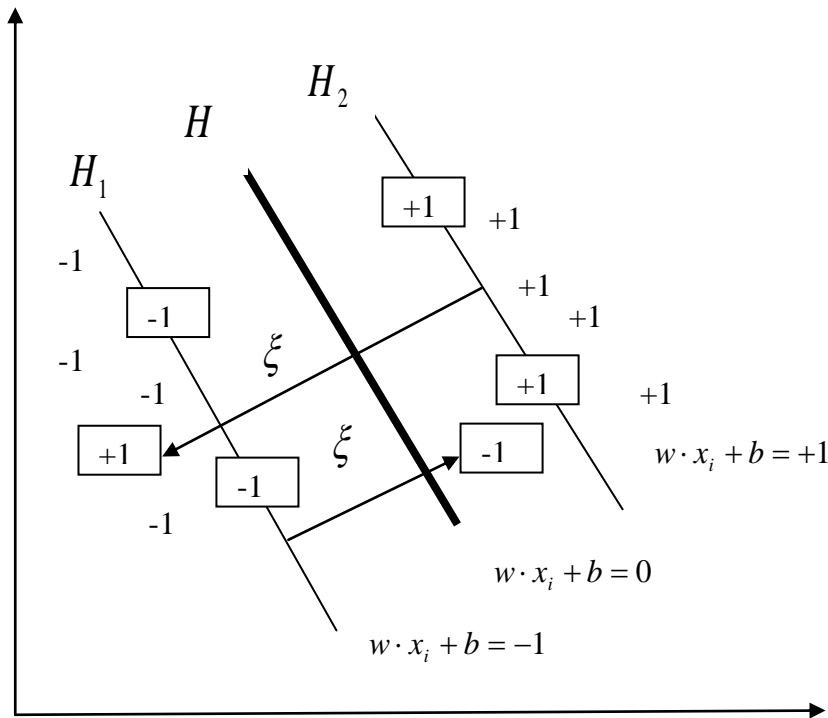


Figure 3.3 *Separating non-separable data with linear separable hyperplanes. Adapted from Ivanciuc (2007, p. 318)*

As shown in Figure 3.3, some of the objects that belong to the two datasets we are trying to classify may end up being misclassified. It is important to account for this misclassification in our modelling; in that case, we have to introduce variables  $\xi_i > 0$ ; they are called “slack variables;” they will be used to represent objects that were misclassified. Let us now incorporate the slack variable in equation 3.5; which can be revised as:

$$y_i(w \cdot x + b) \geq 1 - \xi_i \quad (3.6)$$

We can see from Figure 3.3 that our optimal hyperplane is  $f(x) = 0$ , which lies between classes +1 and -1; it is actually located at the point of maximum separation between classes +1 and -1, as well as the point of minimum error in course of the separation. At this point, the solution to



this problem can be found by solving the following constrained optimization problem (or primal problem) (Vapnik, 2000):

$$\text{Minimize } \frac{1}{2} w \cdot w + C \sum_{i=1}^n \xi_i \quad (3.7)$$

subject to:  $y_i(w \cdot x + b) \geq 1 - \xi_i$ ,  $\xi_i > 0$ , and for  $\forall i = 1, \dots, n$ ; where  $C$ ,  $0 < C < \infty$ , is called the penalty value or regularization parameter. According to Ivanciuc (2007),  $C$  is a trade-off between misclassified points and achieving the maximum margin during the training; it is usually chosen by trial-and-error. Though  $C = 100$  is usually assumed to be the standard value for most applications; it should not be taken for granted because the appropriate value of  $C$  may vary on a case to case basis; therefore it is necessary to carry out a cross-validation technique to determine the ideal value of  $C$  for a particular application.

According to Vapnik (2000), we can solve the primal problem given in equation 3.7 using the Lagrangian function,

$$L(w, b, \xi, \alpha, \beta) = \frac{1}{2} w \cdot w + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i (y_i (w \cdot x_i + b) - 1 + \xi_i) - \sum_{i=1}^n \beta_i \xi_i \quad (3.8)$$

and parameters  $L(w, b, \xi, \alpha, \beta)$  must satisfy:

$$\frac{\partial L(w, \xi, \alpha, \beta)}{\partial w} = w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \quad (3.9)$$

$$\frac{\partial L(w, \xi, \alpha, \beta)}{\partial b} = -\sum_{i=1}^n \alpha_i y_i = 0 \quad (3.10)$$

$$\frac{\partial L(w, \xi, \alpha, \beta)}{\partial \xi_i} = C - \alpha_i - \beta_i = 0 \quad (3.11)$$

The optimization problem or dual form resulting from the application of equations 3.9 – 3.11 to the primal problem given in equation 3.7 can be expressed as:

$$\text{Maximize: } \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (3.12)$$

subject to:  $\sum_{i=1}^n \alpha_i y_i = 0$ , and,  $0 \leq \alpha_i \leq C$ , for  $i = 1, \dots, n$ .

Therefore, the decision function for the linear case can be given as:

$$f(x) = \text{sign} \left[ \sum_{i=1}^n y_i \alpha_i^0 (x_i \cdot x) + b^0 \right] \quad (3.13)$$

where  $x_i$  are the training samples;  $y_i$  are the target labels of the training samples (such that,  $y_i \in \{-1, +1\}$ );  $\alpha_i^0$  are (the Lagrangian multipliers) called “support vectors;”  $b^0$  is known as the “bias;” while  $x$  denotes the test set.

Now we can consider a nonlinearly separable problem; that is a case where a linear hyperplane cannot separate the data without error or having some points misclassified. As shown in Figure 3.4, a nonlinearly separable problem can be separated using a nonlinearly separable hyperplane. Since the use of the linear boundary is inappropriate the SVM maps the input vector into a high dimensional space. SVM constructs an optimal hyperplane in the higher dimensional space, by choosing a linear mapping a priori (Gunn, 1998). This initiative utilises the work of Aizerman et al. (1964) which makes it possible to address the *curse of dimensionality* (Bellman, 1961). What actually happens is that points in the input space are mapped onto a higher dimensional feature space where these points are separated with a linear separating hyperplane.

The result is a nonlinear separation between classes +1 and -1 (see Figure 3.4). This task is done by using a nonlinear function  $\phi$  to map the data onto a higher dimensional feature space (see Figure 3.4). In that case a kernel function  $K$  is introduced, such that (Vapnik, 2000):

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \quad (3.14)$$

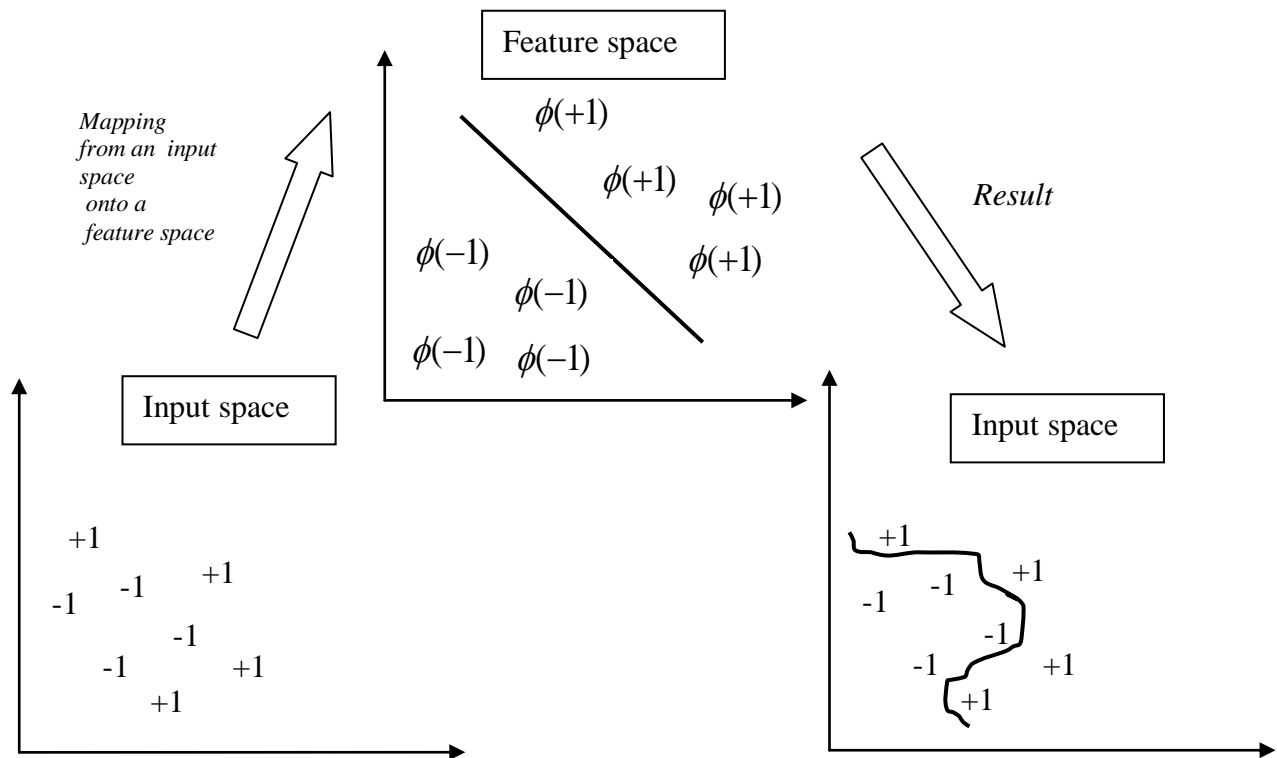


Figure 3.4 The process of classifying a nonlinearly separable data. Adapted from Ivanciuc (2007, p. 323)

The use of kernel functions in SVMs addresses the problem of *curse of dimensionality* (Gunn, 1998). Separation in the feature space does not require that  $\phi$  be determined explicitly; therefore it is more convenient to use the kernel function for our computation. The derivation of

$K(x_i, x_j)$  from  $\phi(x_i)^T \phi(x_j)$  is based on the Mercer's theorem (Mercer, 1909; Cristianini, 2000).

The optimization problem for the nonlinear case can be derived by replacing  $x \cdot x_i$  with  $K(x_i, x_j)$  in equation 3.12, and we can revise equation 3.12 as,

$$\text{Maximize: } \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (3.15)$$

subject to:  $\sum_{i=1}^n \alpha_i y_i = 0$ , and,  $0 \leq \alpha_i \leq C$ , for  $i = 1, \dots, n$ .

The decision function given in equation 3.13 can also be revised by replacing  $(x_i, x)$  with  $K(x_i, x_j)$ ; therefore the decision function for the nonlinear case can be written as,

$$f(x) = \text{sign} \left[ \sum_{i=1}^n y_i \alpha_i^0 K(x_i, x) + b^0 \right]. \quad (3.16)$$

Given two arbitrary support vectors  $x_A \in \text{class } A$  and  $x_B \in \text{class } B$ , the bias can be evaluated as:

$$b^0 = -\frac{1}{2} \sum_{i=1}^n y_i \alpha_i^0 [K(x_A, x_B) + K(x_A, x_B)] \quad (\text{Vapnik, 2000}). \quad (3.17)$$

The kernel  $K(x_i, x_j)$  can be any of the following common kernel functions: the linear kernel  $x \cdot x_i$ , polynomial kernel  $(x \cdot x_i + 1)^d$ , and Radial Basis Function (RBF) kernel

$$K(x_i, x_j) = \exp \left( -\frac{\|x_i - x_j\|^2}{2\gamma^2} \right) \quad (\text{Vapnik, 2000}).$$

The polynomial and RBF are nonlinear kernel functions. The parameters: gamma  $\gamma$  and polynomial order  $d$  control the shape of the separating hyperplane. Kernel functions are extremely important because each of these three

kernels functions (linear, polynomial, and RBF) may produce different results for a given classification problem. It is therefore crucial for the user to decide which particular kernel will best model a particular problem. It is also important that the values of the user defined parameters ( $d$  for polynomial function and  $\gamma$  for the RBF) usually chosen by trial-and-error, be chosen carefully. Different results may be obtained for different values of  $d$  and  $\gamma$ .

### 3.4 Fuzzy support vector machine classification

FSVM was introduced by Lin and Wang (2002). According to Lin and Wang (2002), FSVM simply implies the introduction of the fuzzy membership function into the standard SVM. The essence of introducing the fuzzy function into the standard SVM is to help reduce the effect of noise usually inherent in real world data and thereby improving the learning surface of the standard SVM.

Similar to the concise illustration presented in section 3.3 for the SVM classification, let us present the FSVM version for a binary classification problem. The data we want to classify can be expressed as:  $(x_1, y_1, s_1), (x_2, y_2, s_2), \dots, (x_n, y_n, s_n)$ , where  $x_i \in \mathfrak{R}^N$  belongs to classes  $y_i \in \{-1, +1\}$  for  $s_i \in S$ . Where  $S$  denotes a set of fuzzy membership values described by  $0 \leq s_i \leq 1$ . We can construct a fuzzy grade of membership  $s_i$  as:

$$s_i = \begin{cases} 1 - \frac{\|m_+ - x_i\|}{r_+ + \delta} & \text{if } y_i = +1 \\ 1 - \frac{\|m_- - x_i\|}{r_- + \delta} & \text{if } y_i = -1 \end{cases} \quad (3.18)$$

where,

$$\begin{aligned} r_+ &= \max_{\{x_i: y=+1\}} \|m_+ - x_i\| \\ r_- &= \max_{\{x_i: y=-1\}} \|m_- - x_i\| \end{aligned} \quad (3.19)$$

where  $m_+$  denotes the mean of class +1;  $r_+$  represents the radius of  $m_+$ ;  $m_-$  denotes the mean of class -1;  $r_-$  represents the radius of  $m_-$ ;  $\delta$  is a constant; and  $x_i$  represents land use variables (Lin and Wang, 2002).

It is unnecessary to repeat all the introductory steps given in section 3.3, which also applies to the FSVM. For the linear case the constrained optimization problem can be expressed as (Lin and Wang, 2002),

$$\text{Minimize } \frac{1}{2} w \cdot w + C \sum_{i=1}^n s_i \xi_i \quad (3.20)$$

subject to:  $y_i(w \cdot x + b) \geq 1 - \xi_i$ ,  $\xi_i > 0$ , and for  $\forall i = 1, \dots, n$ ; where  $C$ ,  $0 < C < \infty$ .

Note that, the classification error  $\xi_i$  can now be represented as  $s_i \xi_i$ . The FSVM Lagrangian function can be expressed as (Lin and Wang, 2002),

$$L(w, b, \xi, \alpha, \beta) = \frac{1}{2} w \cdot w + C \sum_{i=1}^n s_i \xi_i - \sum_{i=1}^n \alpha_i (y_i (w \cdot x_i + b) - 1 + \xi_i) - \sum_{i=1}^n \beta_i \xi_i \quad (3.21)$$

and parameters  $L(w, b, \xi, \alpha, \beta)$  must satisfy the following conditions:

$$\frac{\partial L(w, \xi, \alpha, \beta)}{\partial w} = w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \quad (3.22)$$

$$\frac{\partial L(w, \xi, \alpha, \beta)}{\partial b} = -\sum_{i=1}^n \alpha_i y_i = 0 \quad (3.23)$$

$$\frac{\partial L(w, \xi, \alpha, \beta)}{\partial \xi_i} = s_i C - \alpha_i - \beta_i = 0 \quad (3.24)$$

The resulting optimization problem can be given as (Lin and Wang, 2002),

$$\text{Maximize: } \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (3.25)$$

subject to:  $\sum_{i=1}^n \alpha_i y_i = 0$ , and,  $0 \leq \alpha_i \leq s_i C$ , for  $i = 1, \dots, n$ .

The decision rule for the linear case is the same as the SVM (see equation 3.13).

For the nonlinear case, the optimization problem is (Lin and Wang, 2002),

$$\text{Maximize: } \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (3.26)$$

subject to:  $\sum_{i=1}^n \alpha_i y_i = 0$ , and,  $0 \leq \alpha_i \leq s_i C$ , for  $i = 1, \dots, n$ .

The decision rule for the nonlinear case is the same as that of the SVM (see equation 3.16). The same kernel functions used for the SVM also apply to the FSVM.

### 3.5 Numerical examples

For a better understanding on how SVM works, let us apply SVM to a few hypothetical data. All the experiments in this section were implemented in MATLAB<sup>3</sup> using Quadratic Programming (QP) (Cortes & Vapnik, 1995; Vapnik, 1995; Gunn, 1998; Vapnik, 2000).

#### 3.5.1 Classifying linearly separable data.

A linearly separable hypothetical data (Table 3.1) can easily be separated with the linear and nonlinear kernel functions the polynomial kernel and the RBF kernel. Since the data in Table 3.1

---

<sup>3</sup> MATLAB: Matrix Laboratory

can be separated easily with a linear kernel, there is no need to experiment with the nonlinear (polynomial and RBF) kernels in this subsection.

Table 3.1 consists of 10 training points; 5 points each belonging to classes +1 and -1 respectively. An additional 10 test points (5 points each belonging to classes +1 and -1 respectively) (see Table 3.2) were used to test the accuracy of the classification. The training and test data are depicted in Figure 3.5. From Figure 3.5, the black squares and the black triangles are the training data for classes +1 and -1 respectively, while the red squares and red triangles are the test data for classes +1 and -1 respectively.

Table 3.1 *2D linearly separable data (training data)*

Points	$x_1$	$x_2$	$y_i$ (Actual)
1	8	12	+1
2	4	9	+1
3	6	16	+1
4	5	4	+1
5	10	0	+1
6	-1	-13	-1
7	-6	-7	-1
8	-3	-16	-1
9	-7	-5	-1
10	-9	-1	-1



Table 3.2 Test data for a linearly separable classification

Points	$x_1$	$x_2$	$y_i$ (Actual)
1	6	10	+1
2	10	10	+1
3	8	15	+1
4	10	15	+1
5	10	20	+1
6	-10	-5	-1
7	-8	-10	-1
8	-8	-15	-1
9	-6	-15	-1
10	-8	-20	-1

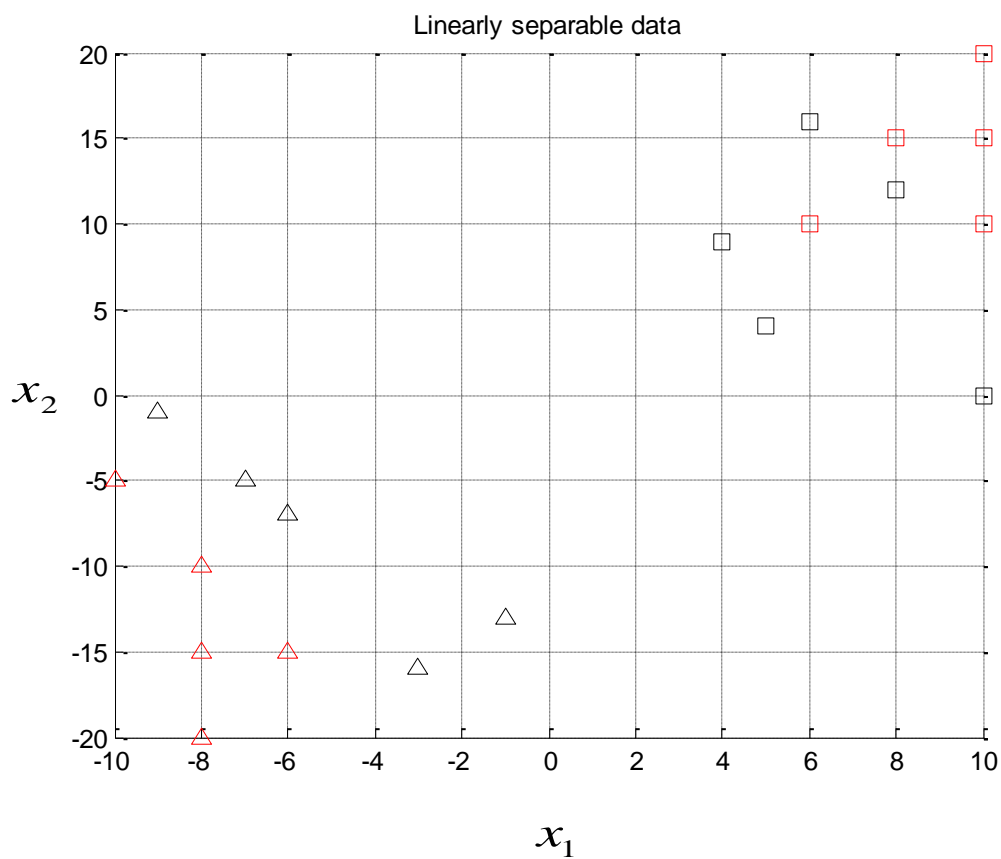


Figure 3.5 2D linearly separable data

The classification was implemented using a linear kernel and setting  $C = 100$ . The classification results (in Table 3.3 and Figure 3.6) show that the linear kernel used 3 support vectors for the separation. From Figure 3.6, the black thick dash-line represents the linear hyperplane that separates classes +1 and -1. From Table 3.3, the training points with nonzero  $\alpha_i$  values are the support vectors. The support vectors are points 4, 6, and 10 with coordinates (5, 4), (-1,-13), and (-9, -1) respectively. Another distinguishing characteristic of support vectors is that their predicted  $y_i$  for classes +1 and -1, must be +1 and -1 respectively. The results in Table 3.3 show that all the training points were correctly predicted by the SVM classifier. Also from Figure 3.6, the circled points are the support vectors. From Table 3.4, all the test points were accurately classified. The next subsection presents a nonlinearly separable case.

Table 3.3 *Training results using a linear kernel*

Points	$x_1$	$x_2$	$y_i$ (Actual)	$b^0$	$\alpha_i^0$	$y_i$ (Predicted)	Remark
1	8	12	+1	0.1154	0	1.9615	Correct
2	4	9	+1	„	0	1.2692	„
3	6	16	+1	„	0	2.0385	„
4	5	4	+1	„	0.0096	1.0000	„
5	10	0	+1	„	0	1.2692	„
6	-1	-13	-1	„	0.0024	-1.0000	„
7	-6	-7	-1	„	0	-1.1154	„
8	-3	-16	-1	„	0	-1.4615	„
9	-7	-5	-1	„	0	-1.0769	„
10	-9	-1	-1	„	0.0072	-1.0000	„

Table 3.4 Test results using a linear kernel

Points	$x_1$	$x_2$	$y_i$ (Actual)	$y_i$ (Predicted)	Remark
1	6	10	+1	1.5769	Correct
2	10	10	+1	2.0385	„
3	8	15	+1	2.1923	„
4	10	15	+1	2.4231	„
5	10	20	+1	2.8077	„
6	-10	-5	-1	-1.4231	„
7	-8	-10	-1	-1.5769	„
8	-8	-15	-1	-1.9615	„
9	-6	-15	-1	-1.7308	„
10	-8	-20	-1	-2.3462	„

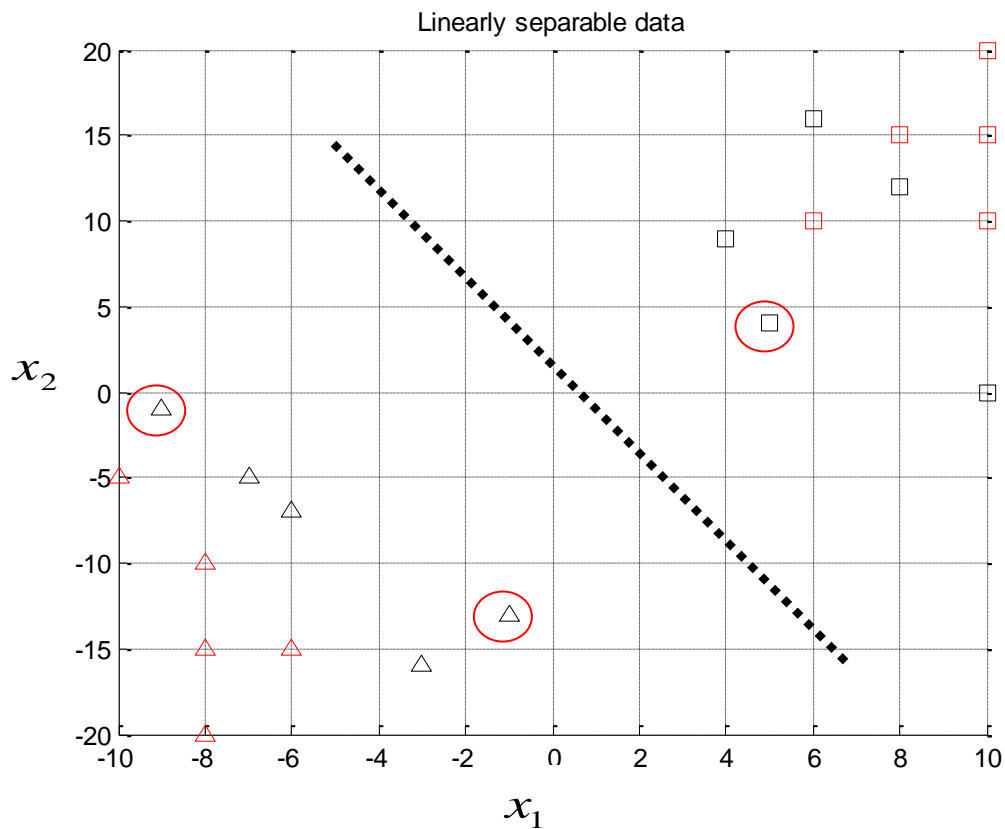


Figure 3.6 Results of classifying a linearly separable data with a linear kernel

### 3.5.2 Classifying nonlinearly separable data.

In this section, we will separate points that cannot be separated with a linear classifier without errors. The polynomial and RBF nonlinear kernels are used to classify the nonlinear data (see Table 3.5). The 10 points given in Table 3.6 will be used to test the accuracy of the classification. Figure 3.7 shows nonlinearly separable data, where the black squares and the black triangles are the training data for classes +1 and -1 respectively, while the red squares and red triangles are the test data for classes +1 and -1 respectively. It can easily be discerned from Figure 3.7 that a linear classifier cannot separate classes +1 and -1 without errors.

Table 3.5 *2D nonlinearly separable data (training data)*

Points	$x_1$	$x_2$	$y_i$ (Actual)
1	3.3	2.7	+1
2	5.1	5.8	+1
3	6.6	4.2	+1
4	5.3	5.4	+1
5	2.9	4.8	+1
6	1.8	2	-1
7	6	0.9	-1
8	8.4	4.2	-1
9	5.7	6.3	-1
10	2.3	6.7	-1

Table 3.6 Test data for a nonlinearly separable classification

Points	$x_1$	$x_2$	$y_i$ (Actual)
1	3.1	3.7	+1
2	4.8	5.2	+1
3	5.8	4.2	+1
4	4.7	3.7	+1
5	3.5	5.2	+1
6	6.3	6.4	-1
7	1.4	0.7	-1
8	7.6	1.2	-1
9	1.3	6.8	-1
10	4.3	6.8	-1

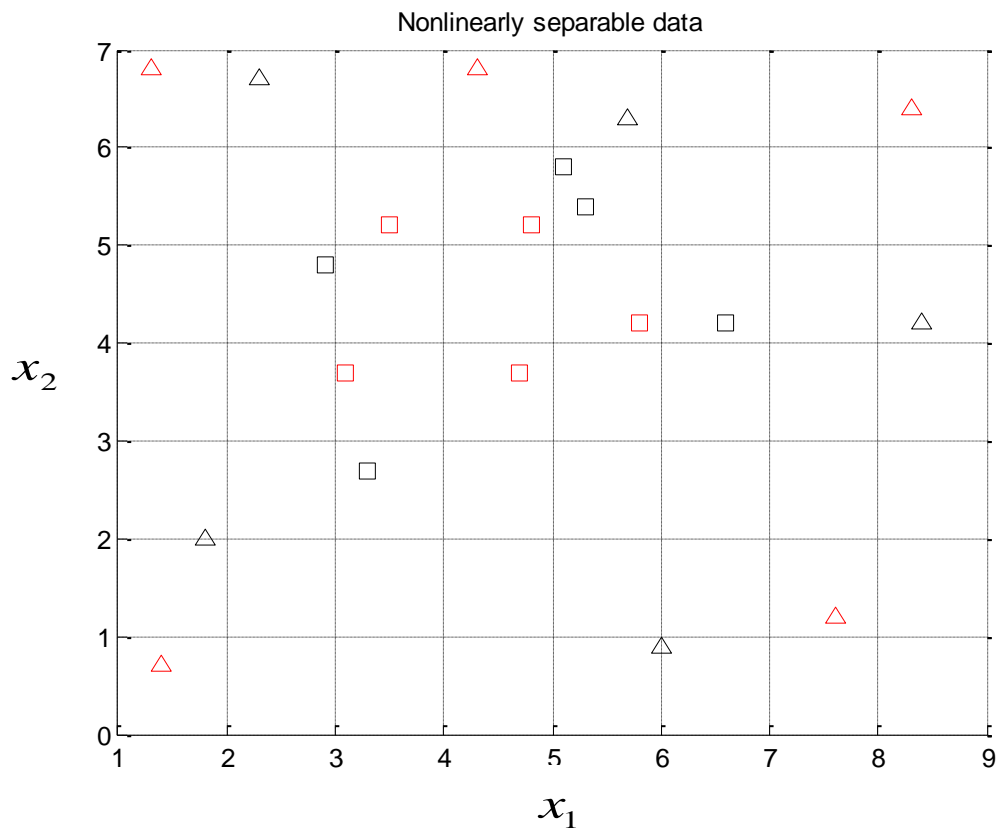


Figure 3.7 2D nonlinearly separable data

The experiment with the polynomial kernel was implemented with  $d=3$  and  $C=100$ , while the RBF kernel experiment was implemented with  $\gamma=0.5$  and  $C=100$ . From Figure 3.8, the points in the red and green circles are the support vectors yielded by the polynomial and the RBF kernels respectively. The thick black dash-line ring/circle is the hyperplane used to separate classes +1 and -1. The training and test results of the polynomial and the RBF kernels are given in Tables 3.7-3.8 and 3.9-3.10 respectively. From Tables 3.7 and 3.10, the polynomial kernel yielded 7 support vectors (points 1, 2, 3, 6, 7, 9, & 10), while the RBF yielded 5 support vectors (points 1, 6, 7, 9, & 10). Both the training and test points of the polynomial and RBF kernels were correctly classified. The next section will present an illustration on how to optimise SVM parameters.

Table 3.7 *Training results using a polynomial kernel*

Points	$x_1$	$x_2$	$y_i$ (Actual)	$b^0$	$\alpha_i^0$	$y_i$ (Predicted)	Remark
1	3.3	2.7	+1	-9.0949e-013	0.0305	1.0000	Correct
2	5.1	5.8	+1	„	0.5563	1.0000	„
3	6.6	4.2	+1	„	0.0247	1.0000	„
4	5.3	5.4	+1	„	0	1.4777	„
5	2.9	4.8	+1	„	0	1.3026	„
6	1.8	2	-1	„	0.6095	-1.0000	„
7	6	0.9	-1	„	0.0073	-1.0000	„
8	8.4	4.2	-1	„	0	-4.6882	„
9	5.7	6.3	-1	„	0.4147	-1.0000	„
10	2.3	6.7	-1	„	0.0081	-1.0000	„

Table 3.8 *Test results using a polynomial kernel*

Points	$x_1$	$x_2$	$y_i$ (Actual)	$y_i$ (Predicted)	Remark
1	3.1	3.7	+1	1.4191	Correct
2	4.8	5.2	+1	2.0366	„
3	5.8	4.2	+1	2.1356	„
4	4.7	3.7	+1	2.5380	„
5	3.5	5.2	+1	1.7201	„
6	6.3	6.4	-1	-2.7357	„
7	1.4	0.7	-1	-1.6207	„
8	7.6	1.2	-1	-2.7458	„
9	1.3	6.8	-1	-2.4249	„
10	4.3	6.8	-1	-0.6033	„

Table 3.9 *Training results using an RBF kernel*

Points	$x_1$	$x_2$	$y_i$ (Actual)	$b^0$	$\alpha_i^0$	$y_i$ (Predicted)	Remark
1	3.3	2.7	+1	-8.8818e-016	1.4487	1.0000	Correct
2	5.1	5.8	+1	„	0	9.0916	„
3	6.6	4.2	+1	„	0	10.9083	„
4	5.3	5.4	+1	„	0	11.046	„
5	2.9	4.8	+1	„	0	2.641	„
6	1.8	2	-1	„	2.5165	-1.0000	„
7	6	0.9	-1	„	0.1107	-1.0000	„
8	8.4	4.2	-1	„	0	-31.9733	„
9	5.7	6.3	-1	„	0.0069	-1.0000	„
10	2.3	6.7	-1	„	0.0116	-1.0000	„

Table 3.10 Test results using an RBF kernel

Points	$x_1$	$x_2$	$y_i$ (Actual)	$y_i$ (Predicted)	Remark
1	3.1	3.7	+1	1.9003	Correct
2	4.8	5.2	+1	10.0486	„
3	5.8	4.2	+1	11.3851	„
4	4.7	3.7	+1	6.5386	„
5	3.5	5.2	+1	5.0082	„
6	6.3	6.4	-1	-17.5453	„
7	1.4	0.7	-1	-1.2446	„
8	7.6	1.2	-1	-10.9616	„
9	1.3	6.8	-1	-3.1689	„
10	4.3	6.8	-1	-0.4317	„

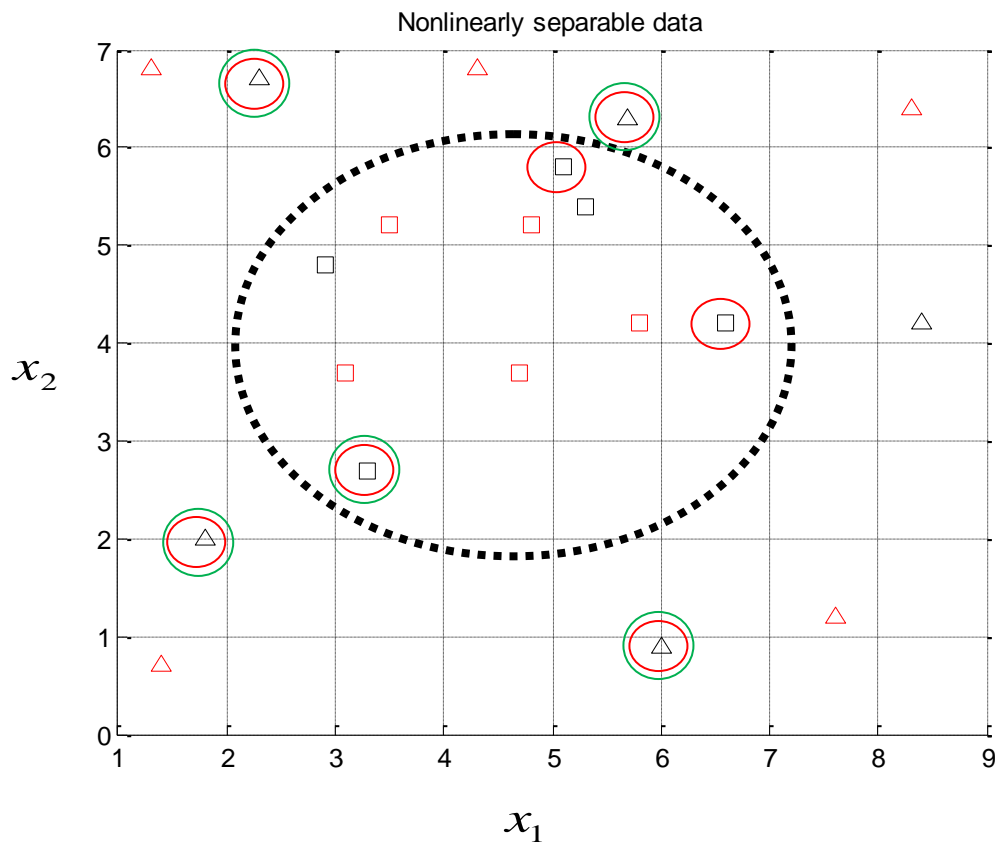


Figure 3.8 Classification result from the polynomial and the RBF kernels (the points in the red and green circles are the support vectors yielded by the polynomial and the RBF kernels respectively)



### 3.5.3 Selecting an optimal value for the regularisation parameter $C$ and kernel parameters for SVM classification and computing the accuracy of the SVM.

Selecting optimal SVM model parameters is usually done by trial-and-error based on a cross-validation approach (Ivanciuc, 2007). This section applies the  $k$ -fold cross-validation technique to numerically illustrate how an optimal penalty value  $C$  and RBF kernel parameter  $\gamma$  can be selected in SVM modelling. In the  $k$ -fold cross-validation technique, usually a certain amount of data is selected from the population and split into  $k$ -subsets or  $k$ -datasets of equal sample size. One of the  $k$ -subsets is first used as the test set while the remaining  $k-1$  subsets are assembled together to form the training set. The process is repeated  $k$  times such that each of the  $k$ -subsets must at one point be used for both training and testing (see Bhardwaj et al., 2005).

In this example,  $k = 3$ , meaning that a selected original sample was split into 3 datasets (T1, T2, and T3 given in Tables 311-3.13). The training sets T1, T2, and T3 (Tables 311-3.13) which consist of 10 points each of equal size were randomly generated for the purpose of illustration. The experiment consists of three stages. At each stage of the experiment, two datasets were put together and used for training while one dataset was kept for testing or validation. Eventually all the three datasets were involved in both training and testing.

Table 3.11 *T1: training data 1 for a k-fold cross-validation experiment*

Points	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$y_i$
1	0.7971	0.6813	0.5603	0.2866	0.4409	+1
2	0.2852	0.7145	0.8458	0.3548	0.8758	+1
3	0.0149	0.4598	0.2848	0.5358	0.8650	+1
4	0.0941	0.9193	0.6633	0.9908	0.3552	+1
5	0.3287	0.9889	0.6023	0.0282	0.6311	+1
6	0.3056	0.9326	0.6565	0.7095	0.8645	-1
7	0.0180	0.4615	0.3099	0.9052	0.0210	-1
8	0.1622	0.9049	0.3316	0.8658	0.0768	-1
9	0.4440	0.3865	0.1882	0.1192	0.3767	-1
10	0.7668	0.6030	0.1007	0.9553	0.1492	-1

Table 3.12 *T2: training data 2 for a k-fold cross-validation experiment*

Points	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$y_i$
1	0.7437	0.9998	0.5643	0.3090	0.7266	+1
2	0.3020	0.8643	0.4315	0.9508	0.5297	+1
3	0.0896	0.0369	0.3378	0.9820	0.8291	+1
4	0.8260	0.5447	0.7207	0.5136	0.5119	+1
5	0.3896	0.9976	0.0137	0.9926	0.5520	+1
6	0.7753	0.5110	0.3741	0.4558	0.2133	-1
7	0.1794	0.8735	0.9227	0.4260	0.5878	-1
8	0.1094	0.0702	0.5465	0.2132	0.1428	-1
9	0.9052	0.9875	0.4739	0.1932	0.0522	-1
10	0.8764	0.9227	0.4965	0.8328	0.6833	-1

Table 3.13 *T3: training data 3 for a k-fold cross-validation experiment*

Points	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$y_i$
1	0.8184	0.5678	0.7657	0.9744	0.9302	+1
2	0.5951	0.6518	0.7566	0.7264	0.0047	+1
3	0.5364	0.4911	0.8433	0.1480	0.6500	+1
4	0.3309	0.3985	0.7702	0.1479	0.6785	+1
5	0.4117	0.4775	0.9787	0.7048	0.2536	+1
6	0.7940	0.0666	0.1114	0.3810	0.8432	-1
7	0.3432	0.4110	0.3961	0.0764	0.2940	-1
8	0.4626	0.9691	0.4921	0.4108	0.0269	-1
9	0.3678	0.7807	0.2581	0.1430	0.0933	-1
10	0.6796	0.7290	0.0370	0.7989	0.7979	-1

The results of the illustration presented in Table 3.14 are the computed overall accuracies (overall accuracy is explained in section 5.7.4). From Table 3.14, the designated values of  $C$  were:  $10e-03$ ,  $10e0$ , and  $10e03$  while the designated values of  $\gamma$  were: 5, 8, and 12. Training sets T2 and T3 were first put together, trained, and validated with training set T1. The process was repeated until all the training sets were used for both training and testing (see Table 3.14). From Table 3.14, the mean overall accuracy for  $\gamma=5$  is 0.3667,  $\gamma=8$  is 0.4000, and  $\gamma=12$  is 0.4111. Therefore the optimal value for gamma  $\gamma$  is 12, since it has the highest mean overall accuracy. The optimal value of  $C$  is  $10e03$ , since it has the highest accuracy (see Table 3.14). The SVM model accuracy is 0.3926.

Table 3.14 *k*-fold cross-validation results for selecting optimal values of  $C$  and  $\gamma$ 

	Overall accuracy								
	C=10e-03			C=10e0			C=10e03		
	$\gamma=5$	$\gamma=8$	$\gamma=12$	$\gamma=5$	$\gamma=8$	$\gamma=12$	$\gamma=5$	$\gamma=8$	$\gamma=12$
(T2+T3) tested with T1	0.3	0.5	0.5	0.3	0.4	0.5	0.3	0.6	0.6
(T1+T3) tested with T2	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
(T1+T2) tested with T3	0.2	0.1	0.1	0.2	0.1	0.1	0.2	0.1	0.1
Mean accuracy	0.3667	0.4	0.4	0.3667	0.3667	0.4	0.3667	0.4333	0.4333
Accuracy for each $C$	0.3889			0.3778			0.4111		
Model accuracy	(0.3889+0.3778+0.4111)/3 = 0.3926								

### 3.6 Summary

This chapter presented the basic concepts of the SVM, and how the FSVM can be derived from the standard SVM. Hypothetical numerical data were used to illustrate how SVMs can be applied. The next chapter describes land use development in Lagos.

## CHAPTER 4

### LAND USE DEVELOPMENT OF LAGOS

#### 4.1 Preamble

This chapter presents an overview of the study area. Three land use epochs, 1963-1978, 1978-1984, and 1984-2000 are used to show land use transition in Lagos from 1963 to 2000. Causal factors considered responsible for land use change in Lagos and the reasons for their consideration will also be presented. Finally, a review of the urban planning regulations that have been applied in Lagos will be discussed.

#### 4.2 Study area

Lagos is a littoral environment, has a relatively flat terrain, and an area of about 2910km<sup>2</sup> (Braimoh & Onishi, 2007). Lagos is bounded by coordinates (515000E, 750000N), (568500E, 750000N), (515000E, 710000N), and (568500E, 710000N).

The rainy (April-October) and dry (November-March) seasons are the two main climatic seasons in Lagos. Its mean temperature and relative humidity hovers around 30°C and between 80-100% respectively (Braimoh & Onishi, 2007). Lagos was the seat of the Nigerian Government until 12<sup>th</sup> December 1991 when Abuja became the nation's capital. Lagos was originally Lagos Island, but presently encompasses, Lekki, Apapa, Ikeja, Festac, Ikeja, Ojota, Oshodi, Mushin, Yaba, and Iddo (see Figure 4.1); as a result of the creation of Lagos State on the 27<sup>th</sup> of May 1967 based on the need to expand the nation's capital territory. Despite the relocation of the seat of Nigerian Government from Lagos to Abuja, Lagos still maintains the reputation as the commercial and industrial nerve centre of Nigeria. The next section presents land use transition in Lagos from 1963 to 2000.

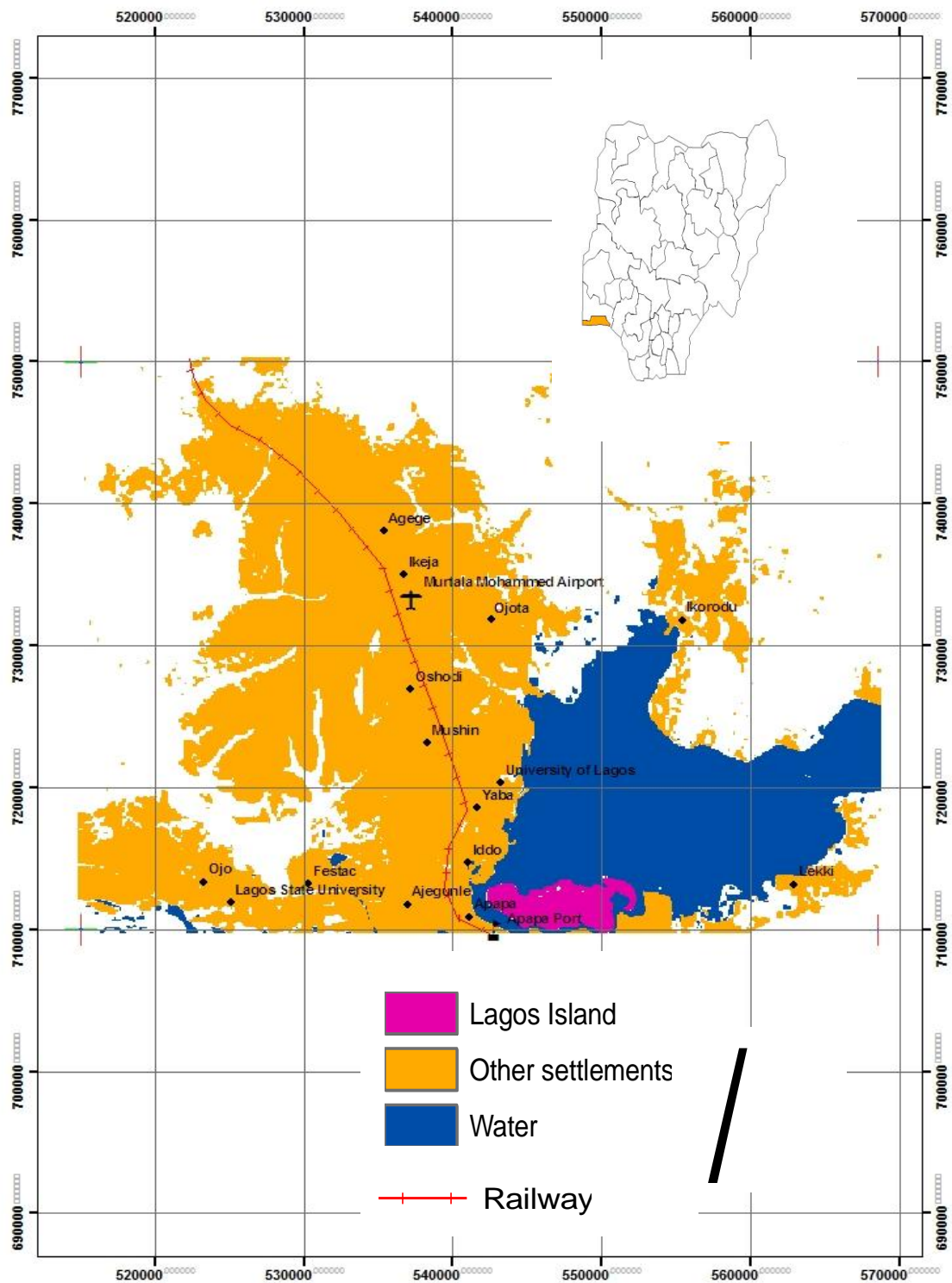


Figure 4.1 *Lagos in relation to Nigeria*

### **4.3 Land use transition in Lagos**

Substantial land use change has occurred in Lagos between 1963 and 2000 (see Figure 4.2). The estimated change in land use in km<sup>2</sup> (see Table 4.1) indicates a sharp increase in land use from 1963 to 2000. The highest change occurred in period 1984-2000, followed by period 1963-1978, while the lowest change occurred in period 1978-1984. The highest rate of change occurred in period 1978-1984, followed by period 1984-2000, while the lowest rate of change occurred in period 1963-1978. Some factors identified as being responsible for land use change in Lagos will be discussed in the following section.

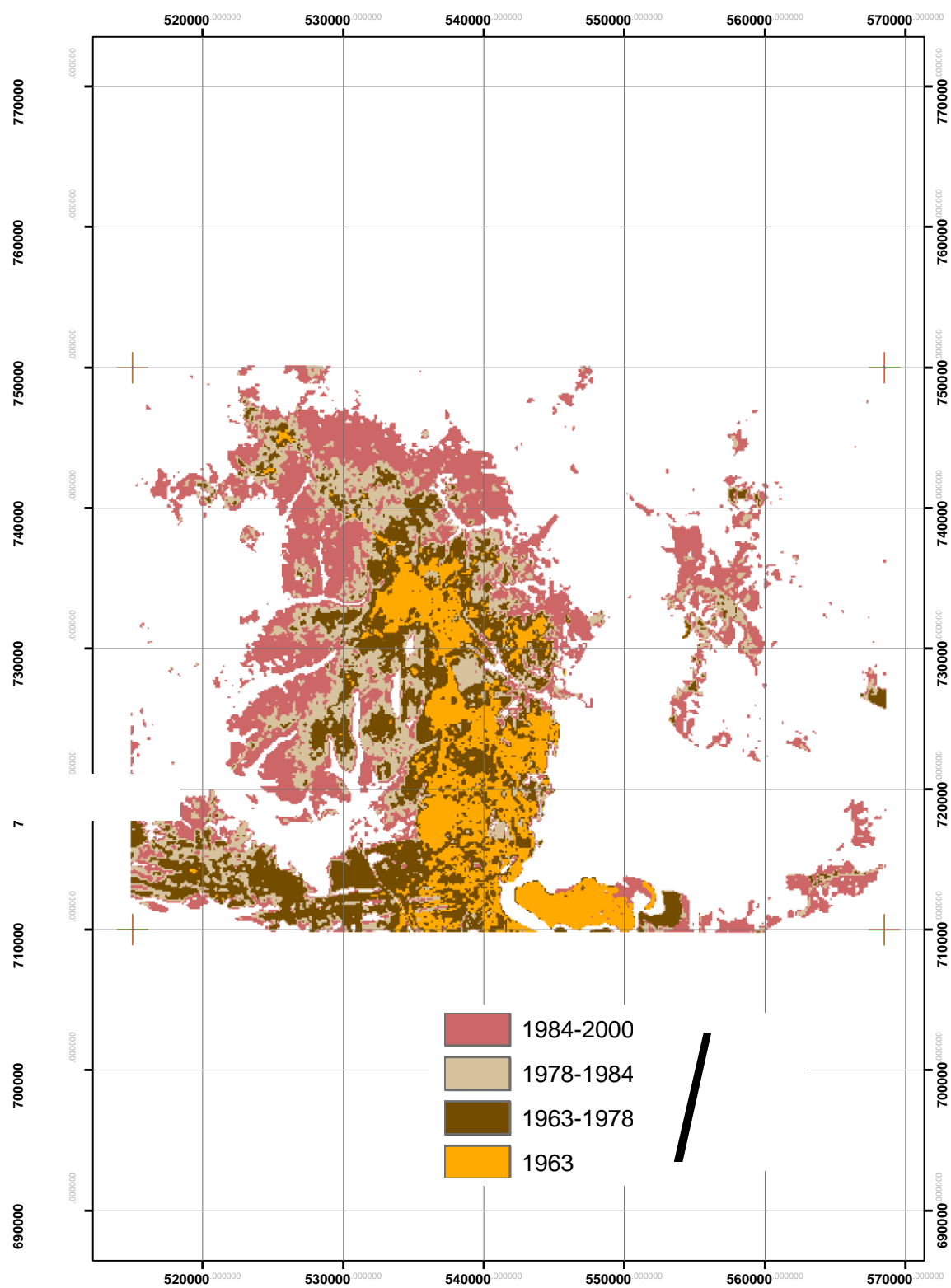


Figure 4.2 *Land use change of Lagos between 1963 and 2000*



Table 4.1 *Estimated land use area, change, and rate of change from 1963 to 2000*

Year	Area (km <sup>2</sup> )	Change (km <sup>2</sup> )	Rate of change (km <sup>2</sup> /yr)
1963	119.66		
1978	279.50	159.84 (1963-1978)	10.6560
1984	435.80	156.30 (1978-1984)	26.0500
2000	667.82	232.02 (1984-2000)	14.5013

#### 4.4 Selecting factors influencing land use development in Lagos, and reasons for their consideration

The selection of the forces driving land use is a crucial aspect of land use change modelling, because land use drivers are the main characteristics that can help to understand the processes of land use transition (from nonurban to urban). There is no hard-and-fast rule or known global formula for selecting land use drivers. The list of land use drivers could be endless. Land use drivers are usually chosen on a case-to-case basis. Land use drivers in one environment might not apply to another (Baker, 1989).

Urban systems are complex systems whose modelling may involve many spatial variables (Liu et al., 2007). If the number of training samples is proportionately lower than the number of spatial variables, the problem of curse of dimensionality will arise which may lead to the overfitting of the training data (Camps-Valls & Bruzzone, 2005). One way of mitigating the problem of curse of dimensionality, is by ensuring a reliable approximation of the data which requires that the amount of training data points be directly proportional to the exponential of the number of variables (Duch & Diercksen, 1995). Another way to mitigate the problem of curse of dimensionality is by reducing the dimensionality of the data; that is, by reducing the number of variables (Jimenez & Landgrebe, 1997). Therefore, twelve salient land use drivers were selected for Lagos. The selected land use drivers influencing land use change in Lagos are: water; residential structures; industrial and commercial centres; major roads; railway; Lagos Island; international airport; international seaport; University of Lagos; Lagos State University; income

potential; and population potential. Statistical methods can be applied to the selected land use drivers to determine their level of influence on the dependent variables, as well as their degree of confidence with respect to their contribution to the model (Lo & Hu, 2007).

The OLS (see chapter 5) and LR (see chapter 6); as well as the one neuron ANN (see chapter 6) will be applied to the selected land use drivers to determine their significance in the model. The next subsection describes the selected land use drivers.

#### **4.4.1 Water.**

Lagos waters encompass the ocean, sea, lagoon, rivers, and inland water ways (see Figure 4.1). Fishing is the main occupation of the indigenous people of Lagos. Sand mining is another activity done in Lagos waters. The glamorous Lagos beaches have attracted local and international tourists. Early civilization and development in Lagos were centred on the coastal areas. Shipping activities are also done in Lagos waters. Water is generally seen as the live-wire of Lagos. In Lagos, land parcels close to water are more desirable. A popular saying in Lagos is, “the closer you are to water the better.” Water is therefore considered an important factor controlling development in Lagos.

#### **4.4.2 Residential structures.**

Residential structures are considered vital determinants for land use change in Lagos; because a new building is more likely to be sited adjacent to an old one than otherwise. For modelling convenience, structures such as schools, hospitals, churches, and mosques are considered as residential structures.

#### **4.4.3 Industrial and commercial centres.**

The industrial and commercial areas of Lagos include towns like Apapa, Ikeja, Yaba, and Lagos Island. The Nigerian oil boom of the seventies and eighties resulted in rapid industrialisation and commercialisation in Lagos. Apart from the aforementioned prominent industrial and commercial towns, massive industrialisation and commercialisation now pervade remote areas such as Ojo, Ikorodu, and Ojota. As a result, these hitherto remote towns have transformed into huge urban centres as the sprawling continues unabated. Industrial and commercial centres are therefore considered a driving force for urban expansion in Lagos.

#### **4.4.4 Major roads.**

Lagos has the largest road network in Nigeria. Highly complex road networks that link Lagos with the rest of Nigeria and neighbouring African countries criss-cross in and around Lagos. Most of these roads were built in the seventies and eighties during the Nigerian oil boom; these vast road networks mainly transverse very remote areas of Lagos. Accessibility to major roads is a prerequisite that is considered by property-developers. For this reason, settlements are usually centred around major-roads. Major roads are therefore catalysts that are very important for urban expansion in Lagos.

#### **4.4.5 Railway.**

Only one railway links Lagos with the rest of the Nigeria. Heavy goods are usually conveyed by rail. Most industrial and commercial settlements were sited close to the railway in order to facilitate the movement of goods and services. The railway has also attracted dense urban settlements around it. Some examples are in towns such as Apapa, Iddo, and Yaba. The railway is therefore considered a propellant for the urban growth of Lagos.

#### **4.4.6 Lagos Island.**

Lagos Island is the central business district and commercial hub of Lagos. Lagos Island was the original seat of the Nigerian Federal Government. The majority of the inhabitants of Lagos commute to Lagos Island on a daily basis. It is home to administrative headquarters of most big corporations in Nigeria including public corporations, banks, and oil companies. Development in Lagos is centred on Lagos Island. The proximity of a parcel of land to Lagos Island invariably determines its value. For these reasons, Lagos Island was chosen as one of the major drivers of urban growth in Lagos.

#### **4.4.7 International airport.**

Murtala Mohammed International Airport, located in Ikeja is the only international airport in Lagos. It is open for business twenty-four hours each day and maintains the reputation as the busiest international airport in Nigeria. Since it was opened for operations on the 15<sup>th</sup> of March 1979, there has been rapid urban growth around it. On this basis, it was considered an important factor for the urban growth of Lagos.

#### **4.4.8 International seaport.**

Apapa Port was built in 1954. It is the largest and busiest international sea port in Nigeria. Apapa Port, or Apapa Quay as it is known, serves as the principal outlet for Nigeria's exports as well as imports, and has galvanised urban expansion around it. Apapa town where Apapa Port is located is home to one of the biggest industrial settlements in Lagos. These industries mainly serve as processing units for exported and imported goods. These reasons make Apapa Port a vital factor propelling urban growth in Lagos.

#### **4.4.9 University of Lagos.**

The University of Lagos was founded by the Nigerian government in 1962 in the Yaba area of Lagos, shortly after Nigeria's independence from Britain in 1960. It is generally abbreviated as UNILAG. UNILAG is located on the mainland and maintains the status as the largest university in Lagos. Since its establishment it has attracted huge settlements mainly around Yaba area, which is presently a highly cosmopolitan area. UNILAG is therefore considered a driving force for urban expansion in Lagos.

#### **4.4.10 Lagos State University.**

Lagos State University generally abbreviated as LASU is the second largest university in Lagos and was founded by the Lagos State government in 1983 in the Ojo area of Lagos. Since its establishment, it has attracted massive urban expansion stretching from Festac to Ojo areas. LASU is therefore considered as one of the factors responsible for urban expansion in Lagos.

#### **4.4.11 Income potential.**

There is presently no viable planning control on sprawl in Lagos. Certain parts of Lagos are more susceptible to sprawl than others. These more susceptible parts of Lagos are high income areas; in the big towns where commercial and industrial activities are prominent. Examples of such towns are Lagos Island, Yaba, Ikeja, Apapa, Festac, and Lekki. Places closer to the high income areas will be more susceptible to growth than places farther apart. Income potential is therefore a determinant for urban growth in Lagos.

#### 4.4.12 Population potential.

Population is unarguably one of the propelling factors of urban change in Lagos. Table 4.2 shows that there has been a steady rise in the population of Lagos from 1866 to 2015.

Table 4.2 *Estimated population growth rate of Lagos from 1866-2015 (Okude & Ademiluyi, 2006, p. 31)*

Census Year	Total population (Million)	Annual rate of growth (%)
1866	0.012	-
1871	0.028	2.70
1881	0.037	2.40
1891	0.032	-1.30
1901	0.042	2.50
1911	0.074	5.70
1921	0.100	3.10
1931	0.126	2.30
1950	0.354	3.30
1963	0.952	14.00
1975	3.300	20.50
1980	4.390	5.50
1990	7.740	5.50
1995	10.290	5.50
2000	13.400	5.05
2004	15.000	5.05
2015	20.200	3.61

As population increases in an environment, there is a natural propensity for more structures to be built in that environment. Highly populated areas of Lagos like Ajagunle, Mushin, Agege, and Oshodi have experienced rapid urban expansion; these towns are predominantly informal settlement areas, necessitated by the surging urban population. Population is therefore considered one of the main factors that determines land use change in Lagos.

## 4.5 Land use planning of Lagos

Significant factors propelling land use change in Lagos, and the reasons for their adoption were presented in the previous section. It is important to review planning regulations that the Nigerian Government has implemented in Lagos. This section presents how planning legislations have shaped the development of Lagos. Nigeria was administered as a British colony from 1854 to 1960. Sections 4.5.1 and 4.5.2 discuss pre-independence and post-independence planning legislations in Lagos. Pre-independence and post-independence legislations represent planning legislations implemented in Lagos from 1854-1960 and 1960-present date respectively.

### 4.5.1 Pre-independence planning legislations.

No significant planning policy was implemented for Lagos until 1873. One of the foremost planning legacies of Lagos could be traced to the effort made by the colonial surveyors, whom in 1873 gazetted that owners of both occupied and unoccupied lands should ensure their properties are kept clean. Property owners were also requested to ensure that their streets, roads, and entire environs were kept tidy and should be devoid of unwarranted bushes and waste dumps (Aduwo, 1999).

A proactive planning scheme was initiated by Governor MacGregor who established the Sanitary Board of Health in 1902. The role of the Board was to advise the Governor primarily on the improvement of public health, environmental sanitation, and potable water supply. Under Planning Ordinance of 1902 the Governor was mandated to establish European Reservations under a sub-board called ‘Local Board of Health.’ The Governor had the responsibility to advance the health and sanitary condition of the European Reservations. After the Planning Ordinance of 1902 came the Public Health Ordinance, which was promulgated in 1908 as an appendage of the Lagos Municipal Board of Health primarily for environmental and health monitoring. The Township Ordinance was promulgated in 1917. This legislation body became

the first conventional planning legislation in Nigeria; the primary objectives of this law include: the provision of infrastructure; public utilities; health; and environmental sanitation (Oduwaye, 2009).

As the population of Lagos continued to grow, there was a need to keep reviewing the planning regulations in order to keep pace with the growing needs of the public. The planning Ordinance of 1928 called ‘Lagos Town Planning Ordinance’ was established primarily for the Lagos Colony. The Lagos Executive Development Board (LEDB) was established under the Lagos Town Planning Ordinance. The roles of the LEDB are: the provision of housing, vetting of building plans, and enactment of bye-laws for land use planning; provision of infrastructure and public utilities; and land reclamation (Oduwaye, 2009).

#### **4.5.2 Post-independence planning legislations.**

The LEDB remained the regulatory planning body in Lagos until 1972 when the Lagos State Development and Property Corporation (LSDPC) was created. The objectives of the LSDPC include: the acquisition, development, holding, selling, leasing, and letting of both mobile and immobile properties within Lagos (Oduwaye, 2009).

By 1973, the government saw the need to harmonise all the existing planning laws affecting Lagos. This gave rise to the Lagos State Town Country Planning Law of 1973 (popularly called ‘Cap 133’ ). The Lagos Local Government (Power) Act of 1959 ( Cap 77), the Lagos Town Planning (Compensation) Act of 1964, the Lagos Executive Development Board (Power) Act of 1964, the Lagos Town Planning (Miscellaneous Provision) Decree of 1967, the Lagos State Town Planning (Miscellaneous Provision) Decree of 1967, the Western Region Law No. 41 of 1969, and the Town Planning Authority (Supervisory Power) Edict of 1971, were all reviewed under the Lagos State Town Country Planning Law of 1973. In 1982 came the Town and Country Planning (Building Plans) Regulation, which was succeeded by the Town and Country



Planning (Guidelines for Approval of Layouts) of 1983. The Town and Country Planning Edict and Town and Country Planning (Building Planning) Regulation were enacted in 1985 and 1986 respectively (Oduwaye, 2009).

In 1998, the Lagos State Urban and Regional Planning Edict No. 2 was enacted because there was need to review the subsisting Lagos State planning law to reflect the newly promulgated federal law known as ‘Nigerian Urban and Regional Law Decree 88 of 1992.’ The consequence of the Nigerian Urban and Regional Law was that it empowered each of the three levels of government in Nigeria (federal, state, and local) to make separate laws to meet their respective planning objectives in such a manner that the states derive their authority from the federal government, while the local governments derive their authority from their respective state governments (Oduwaye, 2009).

In 2005, a law called ‘Administration of Physical Planning and Development Agencies in Lagos State’ came into being (Lagos State Official Gazette, 2005). This new law modified the Lagos State Urban and Regional Planning Edict of 1998; first, to make it workable in a democracy (since the 1998 law was made by the military); second, to give more powers to the ministry in charge of planning rather than concentrating too much powers in the government (as was the case during the military regime); third, to bring planning nearer to the less developed parts of Lagos (Oduwaye, 2009).

### **4.5.3 Appraisal of pre-independence and post-independence planning legislations in Lagos.**

All the planning legislations from pre-independence times until the present were enacted with the intention of creating an enabling framework that can facilitate the development of good road networks, water supply, social amenities, agriculture, solid waste management, electricity supply, and environmental management.

The merit of the pre-independence planning laws of Lagos is that they laid the foundation for the building of roads and bridges, preparing Lagos as the capital city of Nigeria, health management, and the establishment of schools and other social infrastructure. The merit of post-independence planning laws is that they provided the enabling framework that opened up tremendous development in Lagos, especially in less developed areas. The oil boom of the seventies and eighties motivated the Nigerian government to facelift Lagos. The Lagos planning departments can be credited for the substantial transformation that has taken place in Lagos from the time of independence to the present date.

The main problem with Lagos pre-independence planning laws is that they did not target the development of native settlements in Lagos. This planning defect has led to informal settlements or slums that presently pervade native towns in Lagos. The poor workers in Lagos who could not afford the high cost of living in developed parts of Lagos find shelter in low profile housing in informal settlements (Oduwaye, 2009). Critics have persistently blamed post-independence planning programmes for the present deteriorating state of infrastructure and utilities in Lagos.

Most critics have posited that the main shortcomings of post-independence planning programmes in Lagos have been: (i) political instability and lack of continuity: Nigeria has been politically unstable since her independence from Britain; from civil war shortly after independence to successive military governments such that governments come and go, with every new government coming with new planning laws to replace that of the previous government; a new administration usually vilifies the previous one in order to justify itself; (ii) too much power concentrated in the government rather than in the planning ministry; (iii) poor funding; (iv) mismanagement; (v) the inability of the Nigerian Government to stem rural – urban drift by developing other parts of Nigeria aside Lagos; (vi) the neglect of science and technological knowhow. This last shortcoming is a very important factor yet and unfortunately

the least considered reason for low performance of government's planning programmes in Lagos; this is where this thesis finds relevance, in helping to complement the urban planning of Lagos.

#### **4.6 Summary**

Lagos is a mega-city, with a very high population density. Tremendous urbanisation has taken place in Lagos since Nigerian independence from Britain in 1960. This can be discerned from the significant change in the land use of Lagos (see Figure 4.2).

To date, Lagos is still described as a laissez-faire urban society. Informal settlements constitute the biggest problem with urban planning in Lagos. The pre-independence and post-independence planning laws and programmes described in this chapter were intended to equate Lagos with cities like London and New York, but poor planning policies have been the major setback. In the next chapter, the GIS-based GWR and OLS models will be used to investigate the dependence of land use change in Lagos on the twelve selected land use change drivers.

## CHAPTER 5

### CONVENTIONAL METHODS: APPLICATION OF GIS-BASED ORDINARY LEAST SQUARES AND GEOGRAPHICALLY WEIGHTED REGRESSION MODELS TO LAND USE CHANGE MODELLING OF LAGOS

#### 5.1 Preamble

This chapter presents the mathematical principles underlying land use change modelling, the methods used in acquiring the land use data, and few data constraints that were encountered. It will also describe how the land use data were prepared and extracted and will finally describe the application of the GIS-based OLS and GWR models to land use change modelling of Lagos, based on periods 1963-1978, 1978-1984, and 1984-2000.

#### 5.2 Mathematical calibration for land use change modelling

As stated in equation 2.1, a typical (global) linear regression equation for land use change modelling can be expressed as,  $y_i = \beta_0 + \sum_p \beta_p x_{ip} + \varepsilon_i$ . Let  $x_{ip} \in X$  and  $y_i \in Y$ , such that  $X$  represents the present year, while  $Y$  represents the future year;  $x_{ip}$  are the independent variables in the present year  $X$ , while  $y_i$  are the unknown dependent variables in the future year  $Y$ ;  $\beta_p$  are the slope coefficients;  $\beta_0$  is the intercept; while  $\varepsilon_i$  is the error term. In order to map the data in  $X$  onto  $Y$ ;  $\beta_p$ ,  $\beta_0$ , and  $\varepsilon_i$  must be calculated.

Empirical measurements are done in the  $X$  system, while corresponding points in the  $Y$  system are represented with discrete variables. When map  $X$  and  $Y$  are overlaid, for the purpose of predicting change between both systems three categories of land use are produced: (i) the change region, (ii) the developed region, and (iii) undeveloped region. In linear regression

modelling, the three categories of land use can be represented with discrete variables as: (i) undeveloped region = 1; (ii) change region = 2; and (iii) developed region = 3.

The mathematical relationship for predicting land use change between a present year and a future year expressed as equation 2.1 shows that only two maps (a present and a future year) are actually needed to calibrate the land use change equation. Empirical measurements are based on the present year for the purpose of validation a known data representing the future year is used for validating the experiment. However, the use of multiple maps for calibration is better than just using two maps. The advantages of using multiple maps of different epochs/time-slices include: (i) the ability to validate the model at various discrete time slices, and (ii) the historical trend of change in land use can be evaluated/captured over time.

In this research, four maps (consisting three epochs) will be used for modelling land use change in Lagos. Having presented the linear mathematical relationship between the dependent and the independent variables the next section briefly illustrates the sources of the land use data used in this research.

### **5.3 Land use data**

The land use data of Lagos consist of two remotely sensed Landsat Thematic Mapper (TM) images, acquired in 1984 and 2000 respectively; and two analogue base maps acquired in 1963 and 1978. The satellite data have a cell size of 100m x 100m. The base maps were sourced from the Lagos State Ministry of Lands. The next section explains some data constraints encountered in this research.

### **5.4 Land use data constraints**

The main data constraint experienced in this research is the limited amount of historical data for Lagos. The high cost of acquiring both analogue and satellite data makes it extremely difficult for

most developing countries (especially African countries) to acquire historical land use data on a yearly basis. The author's field investigation in Nigeria revealed that no comprehensive land use mapping of Nigeria has been done since independence from Britain in 1960. This is due to the high cost of acquiring land use data and significant political instability in Nigeria since independence. More pressing issues like housing, road network, education, poverty alleviation, and fostering political stability are of highest priority to the Nigerian government. The Nigerian government having recently recognised the importance of land use mapping (and to overcome the rigour and high cost of using classical methods of mapping) has now launched her own satellite programme which is still at the fledgling stage.

As explained in section 5.2, only two maps are basically required for land use change calibration. Land use change predictions by Wu (2002); Li and Yeh (2002); Braimoh and Onishi (2007); Kocabas and Dragicevic (2007); and (PhD thesis by) Sietchiping (2004) were all based on only two land use maps.

High quality remote sensing data for African countries are usually provided by the United States owned Landsat programme. Only the 1984 and 2000 high-resolution Landsat satellite images of Lagos currently available for civilian use can be reliably used in this research; this is the first data constraint in this research. The second data constraint is that very limited land use maps of Lagos have been produced. Only the land use maps produced in 1963 and 1978 are of adequate quality for harmonisation with the 1984 and 2000 satellite images. It is vital to harmonise the satellite and analogue data. The harmonisation process mainly involves four steps: (i) geo-referencing; (ii) overlaying (iii) reclassification; and (iv) editing. These procedures are explained in the next section.

## 5.5 Land use data preparation

The Landsat images acquired in 1984 and 2000 were classified with the  $k$ -means algorithm using MATLAB 7.8 software. The analogue base maps acquired in 1963 and 1973 were scanned and digitised on-screen using ArcGIS 9.3. The analogue and remote sensing data were geo-referenced to ensure both data were in the same coordinate system. Digitising analogue maps reduces their accuracy. It is therefore important to enhance the accuracy of the analogue data to ensure they approximate that of the satellite maps.

The enhancement of the analogue base maps was done in MATLAB, by first obtaining  $n$ -classifications of the satellite data and overlaying the resulting classified satellite data with the digitised base maps. A digital editing procedure was then used to remove errors from the digitised maps. The final images for 1963, 1963-1978, 1978-1984, and 1984-2000 are shown in Figure 4.2.

After the land use data have been prepared, the next step is to extract the land use variables. The methods of extracting the land use variables are described in the next section.

## 5.6 Extraction of land use variables

This section explains the processes of extracting the land use variables already explained in section 4.4. The land use variables (Table 5.1) are grouped into two categories: (i) proximity variables, and (ii) weighted variables.

Table 5.1 *Extracted land use variables*

Land use variables	
Proximity variables	$x_1$ : distance to water
„	$x_2$ : distance to residential structures
„	$x_3$ : distance to industrial and commercial centres
„	$x_4$ : distance to major roads
„	$x_5$ : distance to railway
„	$x_6$ : distance to Lagos Island
„	$x_7$ : distance to international airport (1984-2000 only)
„	$x_8$ : distance to international seaport
„	$x_9$ : distance to University of Lagos
„	$x_{10}$ : distance to Lagos State University (1984-2000 only)
Weighted variables	$x_{11}$ : income potential
„	$x_{12}$ : population potential

The proximity variables were extracted in ArcGIS 9.3 using the Spatial Analyst Tools. Distance to water was extracted by calculating the Euclidean distance of all locations to the nearest water body. Distance to residential structures was extracted by calculating the Euclidean distance of cells to the nearest residential structure. Distance to industrial and commercial centres was extracted by calculating the Euclidean distance of cells to the nearest industrial and commercial centre. Distance to major roads was extracted by calculating the Euclidean distance of all locations to the nearest major road. Distance to railway was extracted by calculating the Euclidean distance of cells to the nearest railway. Distance to Lagos Island was extracted by calculating the Euclidean distance of all cells to Lagos Island. Distance to international airport was extracted by calculating the Euclidean distance of cells to the location of the international airport. Distance to international seaport was extracted by calculating the Euclidean distance of all cells to the location of the international seaport. Distance to University of Lagos was extracted by calculating the Euclidean distance of all cells to the location of University of Lagos.



Distance to Lagos State University was extracted by calculating the Euclidean distance of all cells to the location of Lagos State University.

The weighted variables were derived with MATLAB 7.8. Income disposition in Lagos varies from place to place. This variable was estimated by ranking major towns in Lagos, using a ranking  $r_i$  of a production/service centre  $c_i$ . The ranking was normalized to weight  $w_i$  using the formula ( Braimoh & Onishi, 2007),

$$w_i = \frac{r_i}{\sum_{c_i=1}^n r_i} \quad (5.1)$$

Therefore the weighted inverse distance formula for calculating the income potential  $i_s$  of a location  $S$  is given as (Braimoh & Onishi, 2007),

$$i_s = \frac{\sum_{x=1}^n \frac{1}{|S - S_x|^2} w_i}{\sum_{x=1}^n \frac{1}{|S - S_x|^2}} \quad (5.2)$$

where  $x = 1, 2, \dots, n$  are settlements in the study area.

The final ranking was based on government documents, interviews, and personal knowledge of the author. The population data were made available by the Lagos State Government. The population estimate was obtained using the formula (Braimoh & Onishi, 2007),

$$P_p = \frac{P_r}{e^{r(Y_r - Y_p)}} \quad (5.3)$$

where,  $r$  = intercensal population growth rate;  $P_p$  = previous population;  $P_r$  = present population;  $Y_r$  = recent year; and  $Y_p$  = previous year. An inverse distance weighting formula for calculating an unknown population potential  $p_s$ , at a given location  $S$ , is given as (Braimoh & Onishi, 2007),

$$p_s = \frac{\sum_{x=1}^n \frac{1}{|S - S_x|^2} P(S_x)}{\sum_{x=1}^n \frac{1}{|S - S_x|^2}} \quad (5.4)$$

where  $P(S_x)$  is the population of a settlement located at  $S$ ;  $x = 1, 2, \dots, n$  are settlements in the study area.

Since all the land use variables have now been extracted, the next section will present the use of the extracted land use variables for modelling using the GIS-based OLS and GWR models.

## 5.7 Modelling

Spatial statisticians consider the OLS model an important tool for understanding the dynamics among spatial agents, even though land use data are usually inconsistent with basic underlying traditional statistical assumptions (Overmars et al., 2003; Aguiar, 2006). This section will explore the use of the GIS-based GWR and OLS models for predicting land use change in Lagos, and also examine how spatial models satisfy some fundamental OLS model assumptions of: multicollinearity among the explanatory variables; normality of model residuals; linearity between

the dependent and independent variables; absence of spatial autocorrelation in the model residuals; and homoscedasticity or stationarity among explanatory variables.

The GIS-based GWR model does not furnish information about the significance of the explanatory variables only the OLS model does. The GIS-based OLS and GWR models will serve different purposes. The OLS model will be used to assess the significance of the explanatory variables and other statistical tests that will be presented in this section. The GWR model will be used for the actual prediction, after using the OLS model to assess the significance of the explanatory variables and other statistical tests.

The land use prediction from 1984-2000 will be based on all the 12 explanatory variables, while 10 explanatory variables will be used for periods 1963-1978 and 1978-1984 (see Table 5.1). The reason being that, two of the selected land use drivers (international airport and Lagos State University) came into being after 1978.

The original land use maps (in 1963, 1978, 1984, and 2000) had a cell size of 100m x 100m (see Figure 4.2) before they were gridded to increase their cell size to 500m x 500m (see Figure 5.1). This implies that every 25 pixels in the original map became just a single pixel. Invariably, the value of each pixel in the gridded map became the average of every 25 pixels in the original map. The matrix size of the original maps were too large with very small cell size, as a result could not be modelled with the ArcGIS-based GWR and OLS models. Maps are usually gridded before being modelled with the GWR and OLS models (Erener et al., 2010). The maps were gridded in order to increase their cell sizes and ultimately reduce the amount of pixels in the entire maps. Rescaling may introduce errors but, nonetheless, it is a common phenomenon in spatial mapping and modelling usually applied to either increase or decrease the amount of pixels in the map.

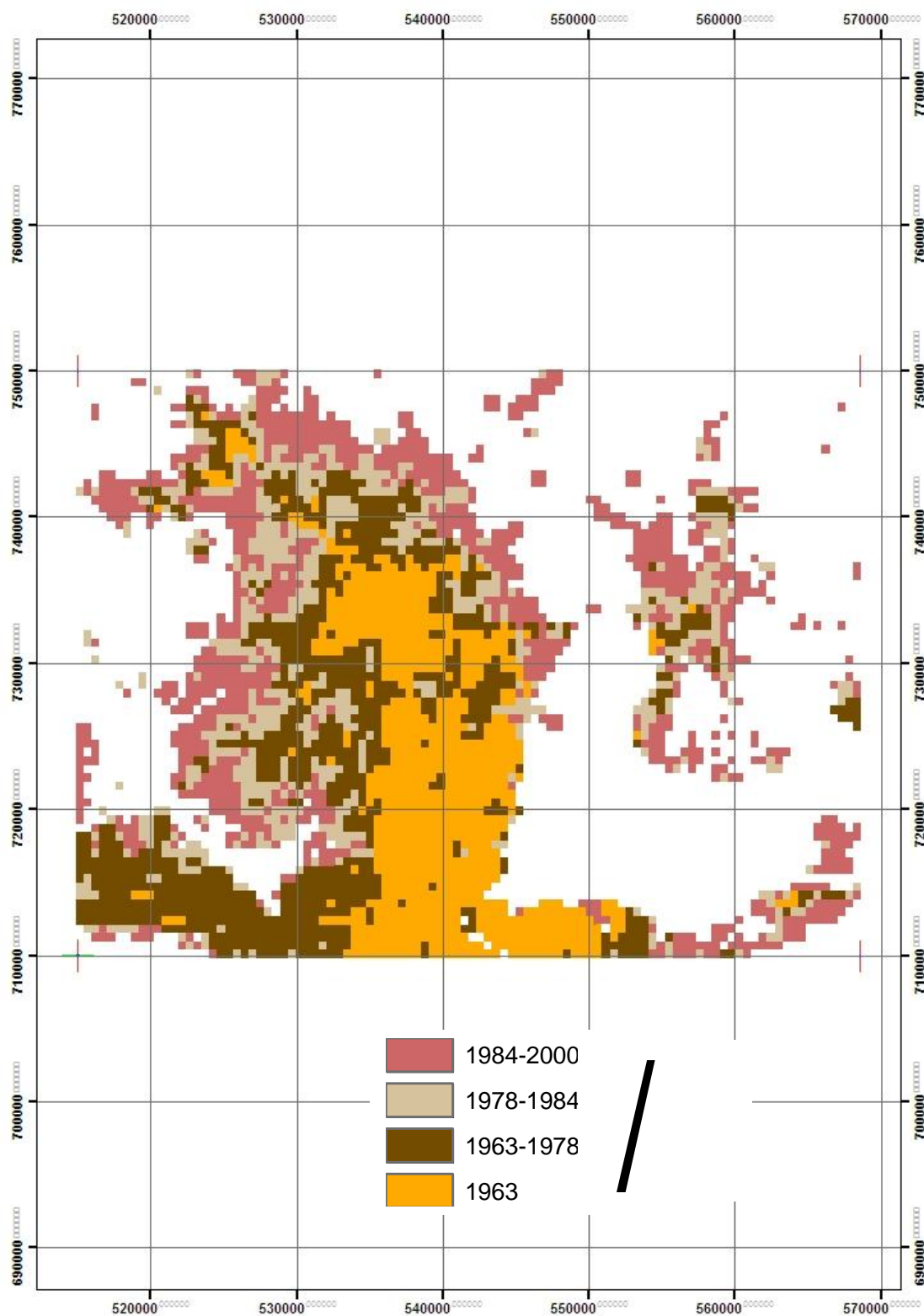


Figure 5.1 *Gridded land use maps of Lagos from 1963-2000*

The GIS-based GWR and OLS models do not need to train the data before the data is used for prediction. The amount of pixels after gridding the maps should not be too large. If the amount of (gridded) pixels is too large the GIS software will produce an error message. The observed data in 1963 were used to predict the 1978 land use; 1978 data were used to predict the 1984 land use; and the 1984 data were used to predict the 2000 land use. The present year is furnished by the independent variables, while the target year is furnished by the dependent variables. Empirical measurements are only done on the present year. The dependent variables are represented with discrete variables. The present year and target year maps were overlaid to determine the changed regions between the present and target years. The results of the overlay are three categories of land use of the target year: (i) undeveloped region; (ii) change region; and (iii) developed region. The resulting overlaid maps form the data for the dependent variable. The dependent variable was represented as follows: undeveloped region=1; changed region=2; and developed region=3.

The original values of the independent variables were scaled to [0, 1] using the transformation formula (Gong, 1996; Li & Yeh, 2002),

$$x'_i = (x_i - \min) / (\max - \min) \quad (5.5)$$

where  $x'_i$  is the scaled land use variable;  $\min$  is the lowest value in the land use vector;  $\max$  is the highest value in the land use vector; and  $x_i$  represents the land use variables.

This scaling technique is effective in ensuring that all the independent variables are equally weighted (Gong, 1996; Li & Yeh, 2002).

Equations 5.6-5.8 and equations 5.9-5.11 are the OLS and GWR equations for periods 1963-1978, 1978-1984, and 1984-2000 respectively;

$LUC_{1963-1978_i}$ ,  $LUC_{1978-1984_i}$ , and  $LUC_{1984-2000_i}$  represent Land Use Change (LUC) from 1963-1978, 1978-1984, and 1984-2000 respectively:

$$LUC_{1963-1978_i} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + e_i \quad (5.6)$$

$$LUC_{1978-1984_i} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + e_i \quad (5.7)$$

$$LUC_{1984-2000_i} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} + \beta_{12} x_{12} + e_i \quad (5.8)$$

$$LUC_{1963-1978_i} = \beta_0 + \beta_1 x_1(u_1, v_1) + \beta_2 x_2(u_2, v_2) + \beta_3 x_3(u_3, v_3) + \beta_4 x_4(u_4, v_4) + \beta_5 x_5(u_5, v_5) + \beta_6 x_6(u_6, v_6) + \beta_7 x_7(u_7, v_7) + \beta_8 x_8(u_8, v_8) + \beta_9 x_9(u_9, v_9) + \beta_{10} x_{10}(u_{10}, v_{10}) + e_i \quad (5.9)$$

$$LUC_{1978-1984_i} = \beta_0 + \beta_1 x_1(u_1, v_1) + \beta_2 x_2(u_2, v_2) + \beta_3 x_3(u_3, v_3) + \beta_4 x_4(u_4, v_4) + \beta_5 x_5(u_5, v_5) + \beta_6 x_6(u_6, v_6) + \beta_7 x_7(u_7, v_7) + \beta_8 x_8(u_8, v_8) + \beta_9 x_9(u_9, v_9) + \beta_{10} x_{10}(u_{10}, v_{10}) + e_i \quad (5.10)$$

$$LUC_{1984-2000_i} = \beta_0 + \beta_1 x_1(u_1, v_1) + \beta_2 x_2(u_2, v_2) + \beta_3 x_3(u_3, v_3) + \beta_4 x_4(u_4, v_4) + \beta_5 x_5(u_5, v_5) + \beta_6 x_6(u_6, v_6) + \beta_7 x_7(u_7, v_7) + \beta_8 x_8(u_8, v_8) + \beta_9 x_9(u_9, v_9) + \beta_{10} x_{10}(u_{10}, v_{10}) + \beta_{11} x_{11}(u_{11}, v_{11}) + \beta_{12} x_{12}(u_{12}, v_{12}) + e_i \quad (5.11)$$

The meaning of the explanatory variables ( $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}$ ) are given in Table 5.1. In equations 5.6-5.11,  $\beta_0$  is the intercept;  $\beta_1, \dots, \beta_{12}$  are the coefficients of the independent variables;  $u, v$  are the horizontal coordinates; and  $e_i$  is the error term. The data were prepared in MATLAB, and thereafter imported into ArcGIS 9.3 for modelling. The dependent

variables must be shapefiles consisting of discrete variables. The attribute tables of the dependent variables (to be predicted) must contain: unique or primary keys, all the attributes of the independent variables (explanatory variables), and discrete variables of the dependent variables. The dependent variables cannot be represented with continuous variables because the resulting map-overlay consists of only three land use types: undeveloped, change, and developed. The next subsection will assess the effectiveness of the GIS-based OLS model to estimate the regression coefficients.

### 5.7.1 Model significance.

The first stage of the modelling is to assess whether the explanatory variables can effectively estimate the regression coefficients. In ArcGIS 9.3, the OLS model furnishes the Joint Wald Statistic (see Table 5.2) used to test the overall significance of the model:

$H_0$ : the explanatory variables in the model are not effective at the 95%

Confidence Level (CL)

$H_1$ : the explanatory variables in the model are effective (reject  $H_0$  if  $p$ -value  $< 0.05$ )

Table 5.2 *Joint Wald Statistic for 1963-1978, 1978-1984, and 1984-2000 (\*significant at  $p < 0.05$ )*

Periods	Jarque-Bera Statistic	Degrees of freedom	$P$ -value
1963-1978	2420.836938	10	0.000000*
1978-1984	3228.506277	10	0.000000*
1984-2000	3627.033362	12	0.000000*

If the Koener (BP) Statistic (see Table 5.8) indicates statistical significance (as in this case), then the Joint Wald Statistic is used to assess the overall significance of the model. The Joint Wald Statistic results (Table 5.2) show that the overall model is significant for periods 1963-1978, 1978-1984, and 1984-2000. The Joint Wald Statistic significant results imply that the explanatory variables can effectively estimate the regression coefficients. The next subsection explores the use of the ArcGIS Spatial Statistics Tools to assess the importance of each independent/explanatory variable in the model.

### **5.7.2 Assessing the significance of each independent variable in the Model.**

The GIS-based OLS model was used to explore the significance of each explanatory variable. The results given in Tables 5.3-5.5 are ArcGIS 9.3 statistical outputs that show the significance of each explanatory variable in the model. The calculated multiple  $R^2$  values for periods 1963-1978, 1978-1984, and 1984-2000 were 0.318959, 0.360174, and 0.381849 respectively; while the calculated adjusted  $R^2$  values for periods 1963-1978, 1978-1984, and 1984-2000 were: 0.317984, 0.359258, and 0.380787 respectively. The hypothesis for assessing the significance of each explanatory variable can be stated as follows:

$H_0$ : the coefficients are zero at the 95% CL

$H_1$ : the coefficients are not zero (reject  $H_0$  if  $p$ -

value <5%)

The assessment of the explanatory variables was based on the 95% CL.  $P$ -values that were <5% based on a two-tailed test were considered significant in the model. The critical value of  $t$  was 1.96; therefore,  $t$ -statistic values >1.96 were considered significant in the model.



Table 5.3 *Statistical results for assessing the significance of each independent variable in the model for 1963-1978 (\*significant at  $p < 0.05$  or  $t > 1.96$ )*

Variable	Coefficient	Std Error	t-statistic	P-value	VIF [1]
Intercept	0.847101	0.018673	45.363946	0.000000*	- - - -
Distance to water	-0.117348	0.055944	2.097593	0.035964*	1.509747
Distance to residential	-1.124156	0.050378	22.314590	0.000000*	1.159984
Distance to industrial and commercial	-0.806965	0.045379	17.782810	0.000000*	1.342798
Distance to major roads	-0.421015	0.047971	8.776518	0.000000*	1.271059
Distance to railway	-0.716223	0.042507	16.849342	0.000000*	1.194308
Distance to Lagos Island	-0.397699	0.042050	9.457734	0.000000*	1.325239
Distance to international seaport	-0.174235	0.039471	4.414274	0.000013*	1.254065
Distance to University of Lagos	-0.062781	0.041584	1.509759	0.131166	1.274922
Income potential	0.520350	0.043490	11.964778	0.000000*	1.141347
Population potential	0.230385	0.035664	6.459927	0.000000*	1.141712

Table 5.4 *Statistical results for assessing the significance of each independent variable in the model for 1978-1984 (\*significant at  $p < 0.05$  or  $t > 1.96$ )*

Variable	Coefficient	Std Error	t-statistic	P-value	VIF [1]
Intercept	1.092893	0.022225	49.173422	0.000000*	-----
Distance to water	-0.208187	0.065870	3.160574	0.001597*	1.440754
Distance to residential	-1.919315	0.062447	30.735343	0.000000*	1.244771
Distance to industrial and commercial	-1.010004	0.059138	17.078679	0.000000*	1.410719
Distance to major roads	-0.927202	0.074672	12.417024	0.000000*	1.086095
Distance to railway	-0.655001	0.052848	12.393995	0.000000*	1.270764
Distance to Lagos Island	-0.308435	0.050105	6.155783	0.000000*	1.321559
Distance to international seaport	-0.140088	0.047237	2.965640	0.003041*	1.236370
Distance to University of Lagos	0.137918	0.049860	2.766108	0.005689*	1.261713
Income potential	0.746732	0.054971	13.584038	0.000000*	1.178389
Population potential	0.339526	0.057042	5.952223	0.000000*	1.182373

Table 5.5 *Statistical results for assessing the significance of each independent variable in the model for 1984-2000 (\*significant at  $p < 0.05$  or  $t > 1.96$ )*

Variable	Coefficient	Std Error	t-statistic	(P-value)	VIF [1]
Intercept	-2.290807	0.023219	98.658942	0.000000*	-----
Distance to water	-0.222466	0.075164	2.959756	0.003099*	1.176389
Distance to residential	-2.606189	0.077746	33.522009	0.000000*	1.104094
Distance to industrial and commercial	-2.113546	0.092167	22.931596	0.000000*	1.137527
Distance to major roads	-1.486506	0.106371	13.974726	0.000000*	1.063784
Distance to railway	-0.910023	0.058224	15.629605	0.000000*	1.050858
Distance to Lagos Island	-0.339675	0.056222	6.041727	0.000000*	1.093061
Distance to international airport	-0.035236	0.054862	0.642261	0.520725	1.102546
Distance to international seaport	-0.189215	0.062306	3.036886	0.002412*	1.093976
Distance to University of Lagos	0.147534	0.060228	2.449619	0.014312*	1.104240
Distance to Lagos State University	-0.719737	0.058858	12.228258	0.000000*	1.052826
Income potential	0.721885	0.063363	11.392814	0.000000*	1.154016
Population potential	0.414852	0.048891	8.485296	0.000000*	1.153289

The variables with  $p$ -values marked \* were found to be significant at the 95% CL. For 1963-1978 (see Table 5.3), 9 independent variables out of 10 were significant at the 95% CL. Only distance to University of Lagos was insignificant at the 95% CL. For 1978-1984 (see Table 5.4), all the 10 independent variables were significant at the 95% CL. For 1984-2000 (see Table 5.5), 11 independent variables out of 12 were significant at the 95% CL. Only distance to international airport was insignificant at the 95% CL.

The higher the value of the calculated  $t$ -statistic of an explanatory variable, the greater its influence in the model conversely the smaller the  $p$ -value of an explanatory variable, the greater its influence in the model. In like manner as the  $t$ -statistic, the higher the absolute value of the regression coefficient of an explanatory variable the greater its influence in the model. From

Tables 5.3 and 5.4 (that is for both periods 1963-1978 and 1978-1984) distance to residential had the highest influence while distance to University of Lagos had the lowest influence. From Table 5.5, distance to residential had the highest influence while distance to international airport had the lowest influence.

The signs of the calculated regression coefficients of the independent variables indicate their relationship with the predicted dependent variable. An increase in the values of the explanatory variables with positive coefficients will increase the number of developed cells in the model; while a decrease in the values of the explanatory variables with positive coefficients will decrease the number of developed cells in the model. An increase in the values of the explanatory variables with negative coefficients will decrease the number of developed cells; while a decrease in the values of the explanatory variables with negative coefficients will increase the number of developed cells. The last column in Tables 5.3-5.5 explains the effect of multicollinearity in the model. Multicollinearity in the model will be explained in section 5.7.3.1. The next section examines how the GIS-based linear regression land use change modelling conforms to traditional statistical assumptions underlying linear regression modelling.

### **5.7.3 Traditional statistical assumptions vis-à-vis GIS-based linear regression land use change modelling.**

This section assesses the spatial GIS-based linear regression OLS model with respect to some fundamental traditional statistical assumptions. This section assesses how the GIS-based OLS modelling results meet the following traditional statistical criteria: multicollinearity; linearity; normality; spatial autocorrelation; and homoscedasticity or stationarity.

### ***5.7.3.1 Multicollinearity.***

Traditional statistics assumes that the linear regression model is devoid of multicollinearity; which means that no exact correlation exists between the explanatory variables. In spatial statistics, the presence of a significant multicollinearity in the spatial linear regression model implies that the offending explanatory variable is redundant in the spatial model, and should be removed.

In ArcGIS, the Variance Inflation Factor (VIF) (see the last columns of Tables 5.3- 5.5) assesses the effect of multicollinearity in the spatial model. Explanatory variables with  $VIF > 7.5$  are considered redundant, and should be excluded from the model. The calculated VIF values for all the explanatory variables in the three experiments (1963-1978, 1978-1984, and 1984-2000) were  $< 7.5$  (see the last columns of Tables 5.3- 5.5). This implies that all the variables were important in the prediction, and should be retained in the model. The next section assesses whether the land use data meet the linearity criterion for linear regression modelling.

### ***5.7.3.2 Linearity.***

The scatter plot in Figure 5.2a depicts a typical relationship between a dependent and an independent variable in linear regression modelling. Based on empirical land use change data, Figure 5.2b is the plot of a dependent variable (1963-1978) against an independent variable (distance to water 1963-1978). Figure 5.2b shows that land use change data do not conform to a typical linear regression relationship between a dependent and an independent variable shown in Figure 5.2a.

Figure 5.2b shows that the relationship between the land use dependent variable (y-axis) and the independent variable (x-axis) is nonlinear as expected. Using discrete variables to represent the dependent variable will definitely produce a nonlinear relationship between the dependent

and explanatory variables, as in the case of Figure 5.2b. Therefore the land use change data do not meet the linearity condition expected of all linear regression modelling.

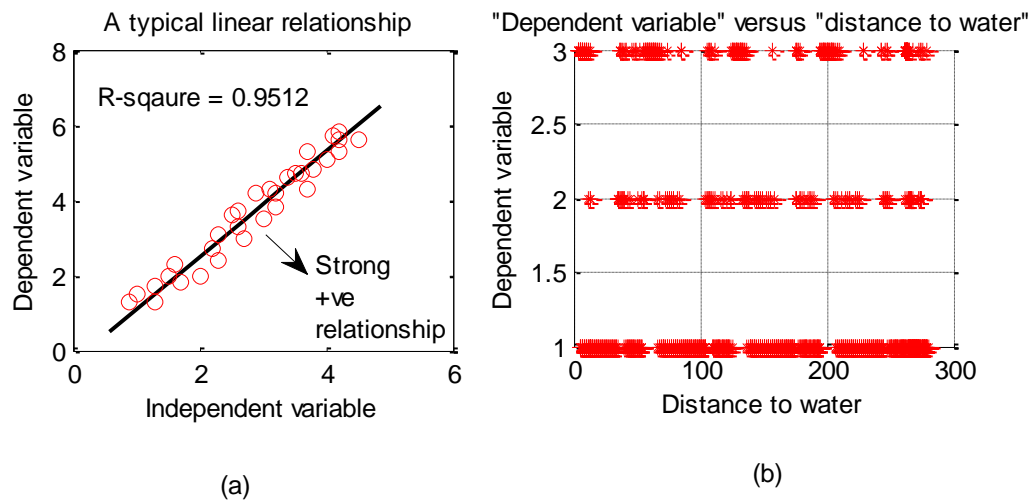


Figure 5.2 (a) A typical linear regression relationship between a dependent and an independent variable; (b) dependent variable plotted against distance to water

### 5.7.3.3 Normality.

Even though the use of binary or discrete variables to represent the dependent variable is synonymous with logistic regression, dependent variables in OLS models can also be represented with binary or discrete variables (Menard, 1995), as typical in land use change modelling using linear regression. The disadvantage of using discrete variables for the dependent variables is that the normality and homoscedasticity assumption of the OLS may be violated; nonetheless the results can still be useful for classification and statistical testing (Pohlmann & Dennis, 2003).

The Jarque-Bera Statistic is used for testing whether the model residuals are normal:

$H_0$ : the residuals are normally distributed at the 95% CL

$H_1$ : the residuals are not normally distributed (reject  $H_0$  if  $p$ -value  $< 0.05$ )

The results of the ArcGIS Jarque-Bera Statistic are given in Table 5.6.

Table 5.6 *Jarque-Bera test for 1963-1978, 1978-1984, and 1984-2000 (\*significant at  $p < 0.05$ )*

Periods	Jarque-Bera Statistic	Degrees of freedom	$P$ -value
1963-1978	706.116841	10	0.000000*
1978-1984	392.961529	10	0.000000*
1984-2000	318.762167	12	0.000000*

All the calculated  $p$ -values for 1963-1978, 1978-1984, and 1984-2000 were  $< 0.05$ ; which indicates that the residuals deviate from the normal distribution as expected. The three known methods of transforming a nonlinear model to a linear model are: the logarithmic ( $y = a + b \ln x$ ), exponential ( $y = ae^{bx}$ ,  $\ln y = \ln a + bx$ ) and power function ( $y = ax^b$ ,  $\ln y = \ln a + b \ln x$ ). As expected, the relationship in Figure 5.2b remained the same when these transformations were applied to the model. If the dependent variables are changed to continuous variables, the model will not be able to model the dichotomous land use data (which is either developed or undeveloped).

A histogram plot was used to further examine the normality of the OLS model residuals. The plots for periods 1963-1978, 1978-1984, and 1984-2000 given in Figure 5.3 show that the model residuals deviate from a normal distribution. That was the reason why the ArcGIS Jarque-Bera test was statistically significant for periods 1963-1978, 1978-1984, and 1984-2000.

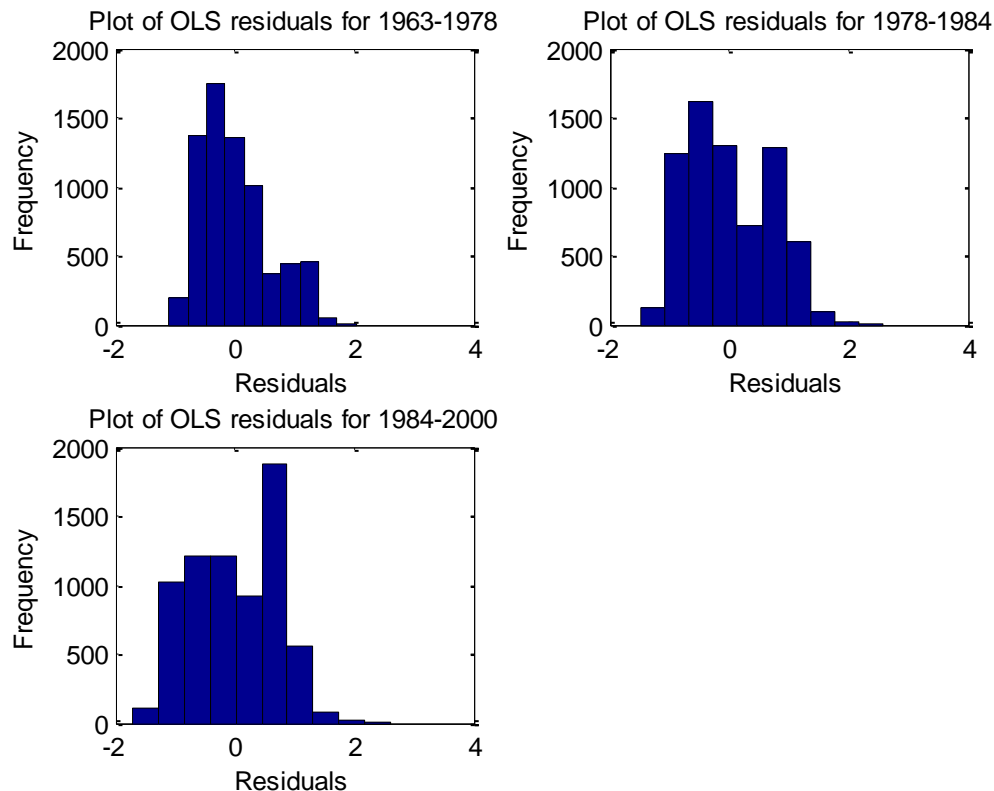


Figure 5.3 *Plots for OLS residuals for periods 1963-1978, 1978-1984, and 1984-2000*

#### ***5.7.3.4 Autocorrelation/spatial autocorrelation.***

Traditional statisticians and spatial statisticians have divergent views on the issue of autocorrelation in the model. In spatial statistics, autocorrelation is called ‘spatial autocorrelation.’ Spatial autocorrelation informs geographers on how spatial variables are interacting with their environment. A spatial variable is said to be spatially autocorrelated if there is a systematic pattern in the spatial distribution of the model residuals (the residual is the difference between the observed and predicted quantity). If a variable is influenced by nearby areas, then it is positively spatially autocorrelated. A variable is negatively autocorrelated if it is influenced by farther areas in the environment. When no positive or negative spatial autocorrelation exists in the model residual, then the pattern is said to be random.



For spatial data to conform to the requirement of traditional statistics, land use data should be autocorrelation free or random. Spatial statistics assumes that the presence of significant spatial autocorrelation in the data implies the absence of key explanatory variables from the model (Fotheringham et al. 2002). The ArcGIS GWR model is robust to spatial autocorrelation effect. The ArcGIS Spatial Autocorrelation Moran's I Tool is used to test for the presence of spatial autocorrelation in the model using the model residuals. As explained in section 2.2.2, Moran I values close to +1 indicate positive spatial autocorrelation; values close to -1 indicate negative spatial autocorrelation; while values close to zero indicate that the model residuals are random. The Moran's I tests (see Table 5.7) for 1963-1978, 1978-1984, and 1984-2000 yielded values close to zero. These results indicate that the model residuals were not spatially autocorrelated.

Table 5.7 *OLS modelling: spatial autocorrelation test for 1963-1978, 1978-1984, and 1984-2000 using ArcGIS Moran's I tool*

Periods	Moran' s I index For OLS
1963-1978	0.081073
1978-1984	0.073943
1984-2000	0.069293

#### ***5.7.3.5 Homoscedasticity or stationarity.***

Traditional statisticians assume that homoscedasticity or variance between the explanatory variables in a linear regression model is equal. On the other hand spatial statisticians assume that the global OLS model does not account for the variation in space; and as a result cannot deal with the problem of non-stationarity. Spatial statisticians have developed the GWR model to

deal with the spatial variations when estimating model coefficients (Fotheringham et al. 2002).

In ArcGIS, the Koenker (BP) Statistic is used for testing for non-stationarity:

$H_0$ : the model is stationary at the 95% CL

$H_1$ : the model is non-stationary (reject  $H_0$  if  $p$ -value  $< 0.05$ )

The GIS-based OLS model estimates for the Koenker (BP) Statistic are given in Table 5.8.

Table 5.8 *Koenker (BP) Statistic for 1963-1978, 1978-1984, and 1984-2000 (\*significant at  $p < 0.05$ )*

Periods	Koener (BP) Statistic	Degrees of freedom	$P$ -value
1963-1978	1237.735416	10	0.000000*
1978-1984	532.003124	10	0.000000*
1984-2000	74.834390	12	0.000000*

All the three results for periods 1963-1978, 1978-1984, and 1984-2000 indicate significant non-stationarity. This implies that the model is not homoscedastic and therefore violates traditional statistical requirements that expect OLS models to be homoscedastic. In ArcGIS, the GWR model becomes a veritable option for the prediction when the OLS results indicate significant non-stationarity.

#### 5.7.4 Prediction.

Spatial dependency in land use data makes them violate basic OLS model assumptions (Tang & Choy, 2000). The OLS model is a global model that does not consider the effect of variation in space or non-stationarity. OLS assumes homoscedasticity of the independent variables across

space; which means that the variance of the independent variables is constant across space. This assumption does not subsist with spatial data.

In spatial statistics, non-stationarity does not invalidate the result of the OLS model; the GWR model was specifically designed to mitigate this problem (Fotheringham et al. 2002, Platt 2004). As already mentioned in section 2.2.2, note that the GWR (just like the OLS) is still subject to the fundamental statistical assumptions being investigated in this chapter.

In the course of the prediction, it is important to assess the performances of both the global OLS model and the local GWR model even though it has been experimentally proven that the GWR model is more accurate (Fotheringham et al. 2002). Akaike's Information Criterion (AIC) measures model performance between or among regression models. The model that has a lower AIC value is considered better. The ArcGIS 9.3 AIC results (Table 5.9) show that the GWR model provided a better fit for the dependent variables, since its values were significantly lower than those of the OLS model.

Table 5.9 *Calculated AIC values for GIS-based OLS and GWR regression models*

Periods	AIC values for OLS	AIC values for GWR
1963-1978	12123.468750	4745.185736
1978-1984	14737.540971	7122.286379
1984-2000	15197.338192	8392.552088

The remaining parts of the prediction were based on the GWR model, since our goal from this point is to find the best fit for the dependent variables. Let us now assess the goodness-of-fit of the land use change prediction. The  $R^2$  was used to assess the goodness-of-fit for predicting land use change in the three periods (1963-1978, 1978-1984, and 1984-2000). The resulting high

$R^2$  and  $R^2$  adjusted values (Table 5.10) show that the GWR predictions for the three periods (1963-1978, 1978-1984, and 1984-2000) were highly accurate.

Table 5.10 *GWR modelling: calculated  $R^2$  values for 1963-1978, 1978-1984, and 1984-2000*

Periods	$R^2$	$R^2$ Adjusted
1963-1978	0.843159	0.792102
1978-1984	0.846021	0.805692
1984-2000	0.853782	0.799722

This result is an indication that the GWR model was able to mitigate the effects of non-stationarity and autocorrelation that usually compromise the goodness-of-fit of the dependent variable as in the case of the OLS model (the calculated  $R^2$  and adjusted  $R^2$  for the OLS model is given in section 5.7.2).

The GWR model predicted maps for 1963-1978, 1978-1984, and 1984-2000 (Figure 5.4) were evaluated with the kappa statistic (also known as kappa coefficient) and the overall accuracy methods of evaluations.

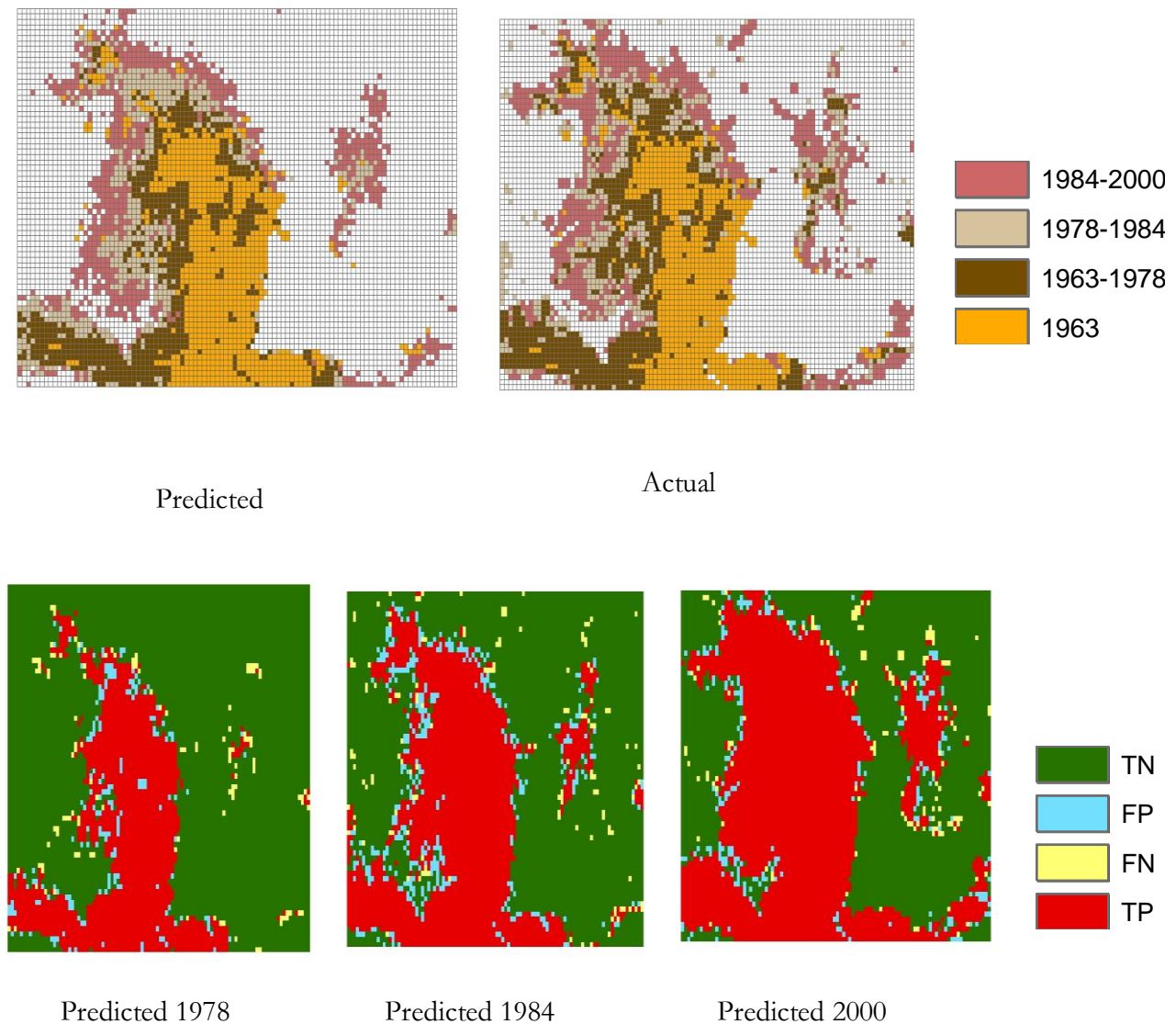


Figure 5.4 *Predicted GWR maps for 1963-1978, 1978-1984, and 1984-2000*

Kappa statistic and overall accuracy are calculated from the confusion matrix also known as error matrix. The cell-by-cell comparison of the GWR predicted maps and the reference data in 1978, 1984, and 2000 are given in the confusion matrices in Tables 5.11- 5.13.

A confusion matrix or error matrix is a square array of values usually used to express classification results. Conventionally, the rows of the matrix represent the values of the predicted data while the columns of the matrix represent the reference data (Lo & Yeung, 2007). Tables 5.11-5.13 are 2 x 2 matrices representing classification results. A confusion matrix is basically a tabular comparison of the predicted map with the actual or reference map. The confusion matrix is a data visualisation tool that facilitates the analysis of results.

Table 5.11 *GIS-based GWR modelling: confusion matrix for period 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1669 (TP)	185 (FP)
Undeveloped	123 (FN)	5023 (TN)

Table 5.12 *GIS-based GWR modelling: confusion matrix for period 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	2369 (TP)	399 (FP)
Undeveloped	139 (FN)	4093 (TN)

Table 5.13 *GIS-based GWR modelling: confusion matrix for period 1984-2000*

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	3282 (TP)	266 (FP)
Undeveloped	150 (FN)	3302 (TN)

The 2 x 2 confusion matrices (Tables 5.11-5.13) consist four elements: TP, FP, FN, and TN. Their meanings are:

*TN (True Negative) = Undeveloped cells that were correctly predicted as undeveloped cells*

*FP (False Positive) = Undeveloped cells that were wrongly predicted as developed cells*

*FN (True Negative) = Developed cells that were wrongly predicted as undeveloped cells*

*TP (False Positive) = Developed cells that were correctly predicted as developed cells*

Thus, overall accuracy is calculated from the error matrix as “the sum of the diagonal values divided by the total number of sample locations” (Lo & Yeung, 2007, p. 120). That is:

*Overall accuracy = the sum of the diagonal elements (TP + TN) ÷ the sum of all the elements in the matrix (TP + FP + FN + TN).*

Kappa statistic (k) or kappa coefficient was introduced by Cohen (1960). Kappa statistic can be expressed mathematically as (Ma & Redmond, 1995; Lo & Yeung, 2007):

$$k = \frac{P_o - P_c}{1 - P_c} \quad (5.12)$$

therefore,

$$P_o = \sum_{i=1}^m P_{ii} = \frac{1}{N} \sum_{i=1}^m n_{ii} \quad , \quad (5.13)$$

$$P_c = \sum_{i=1}^m P_{i+} P_{+i} = \frac{1}{N^2} \sum_{i=1}^m n_{i+} n_{+i} , \quad (5.14)$$

where,

$P_o$  = the proportion of agreement between the reference and sample

$n_{ii}$  = the total number of correctly classified points by class along the diagonal of the error matrix

$N$  = the total number of points checked (sampled)

$P_{ii}$  = the proportion of correctly classified sample points by class at the diagonal of the error matrix (i.e.  $n_{ii} / N$ )

$P_{i+}$  = the marginal distribution of the sample data (  $n_{i+} / N$  where  $n_{i+}$  is the row sum by class)

$P_{+i}$  = the marginal distribution of the reference data ( $n_{+i} / N$  where  $n_{+i}$  is the column sum of class)

$m$  = the total number of classes

The kappa statistic is a more reliable measure than the overall accuracy because it has the ability to evaluate the actual agreement and chance agreement (Fung & LeDrew, 1988). The kappa statistic estimates are usually lower than those of the overall accuracy (see Table 5.14). The calculated kappa statistic and overall accuracies from the GWR predicted maps are given in Table 5.14. The calculated kappa statistic estimates reveal (Table 5.14) that there was substantial agreement between the observed and predicted maps.



Table 5.14 *Calculated kappa statistic estimates for 1963-1978, 1978-1984, and 1984-2000*

Periods	Kappa statistic	Overall accuracy
1963-1978	0.8858	0.9560
1978-1984	0.8366	0.9231
1984-2000	0.8812	0.9406

It is desired that the standard residuals (that is under and over prediction) be randomly distributed. The standard residuals for the three periods shown in Figure 5.5 indicate that the residuals were random. This is an indication that no key independent variable was omitted from the model (Thapa & Murayama, 2009).

High local  $R^2$  values (Figure 5.5) indicate areas on the map where the GWR model predicted well while low local  $R^2$  values indicate areas on the map that were less well predicted by the GWR model.

Cond denotes condition number. Cond assesses local collinearity in the model. The results from cells with Cond values greater than 30 may not be reliable. All the predicted Cond maps for periods 1963-1978, 1978-1984, and 1984-2000 (Figure 5.6) yielded Cond values below 30. This indicates that the results from this prediction can be trusted.

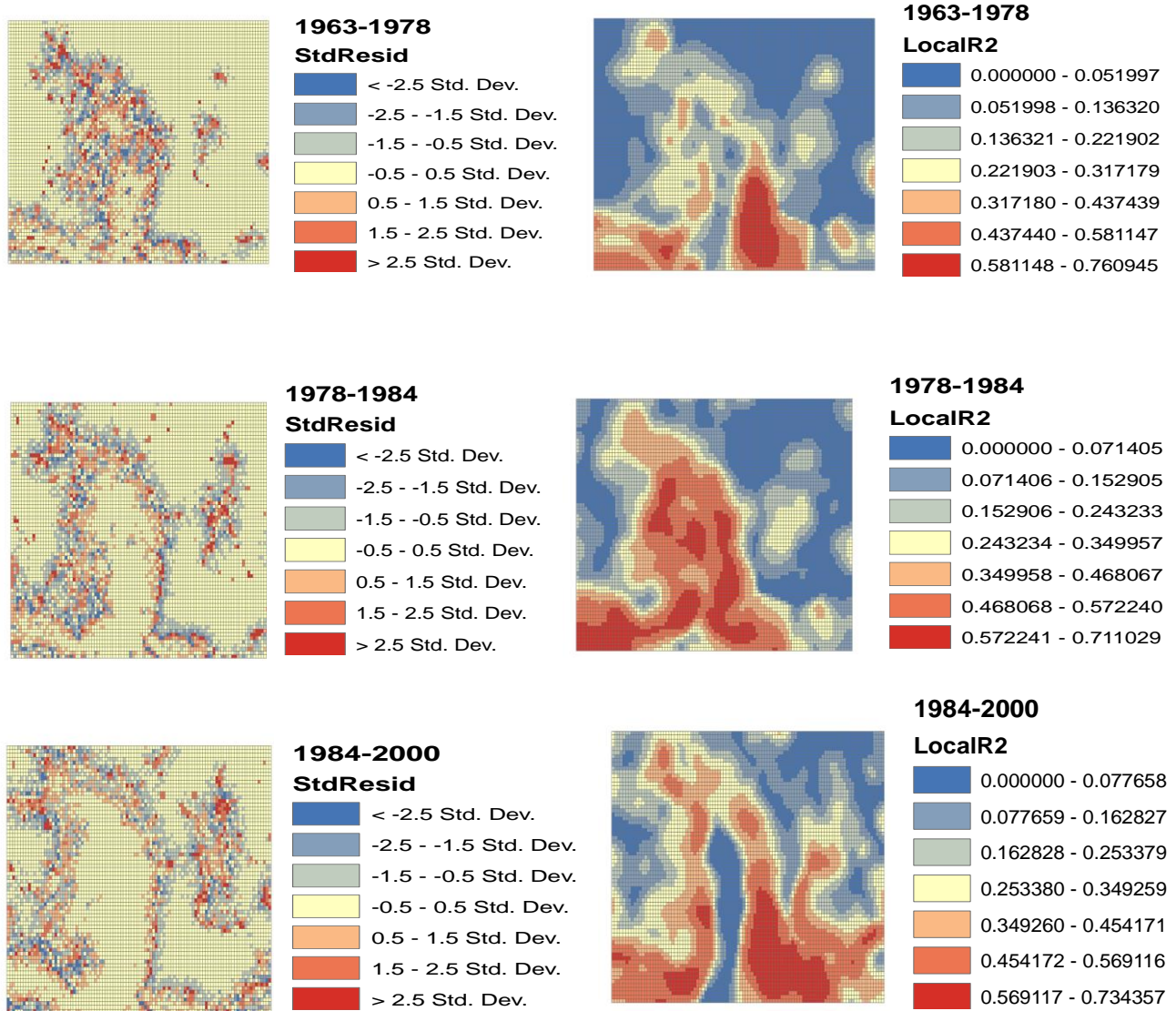


Figure 5.5 *Estimated GWR standard residuals and local  $R^2$  for 1963-1978, 1978-1984, and 1984-2000*

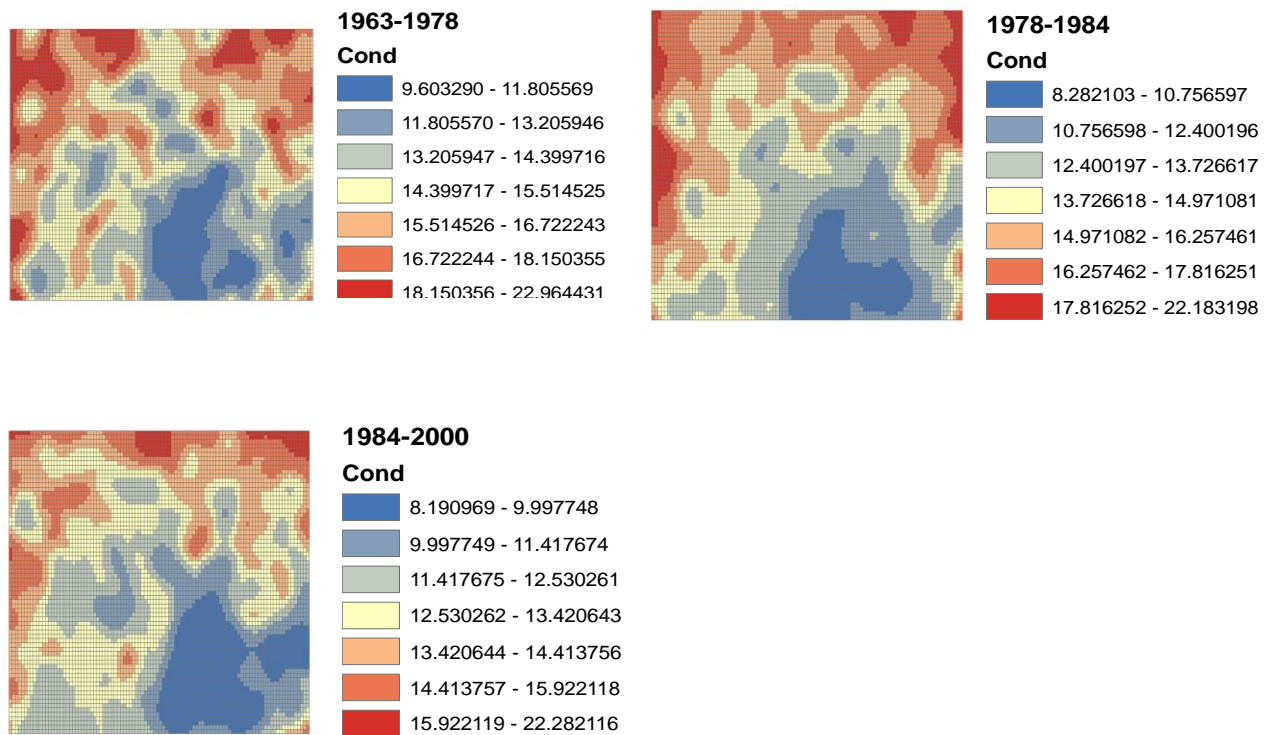


Figure 5.6 *Estimated GWR model condition numbers for 1963-1978, 1978-1984, and 1984-2000*

In ArcGIS, unlike the OLS model, the GWR model does not furnish the Jarque-Bera Statistic. Therefore, the only way to investigate the normality of the GWR model residuals will be to examine them graphically by plotting them. The histogram plot was used to assess the normality of the GWR model residuals. The plots (Figure 5.7) show that the residuals deviated slightly from a normal distribution. The GWR normal plots were more normally distributed than those of the OLS model displayed in Figure 5.3. The normality plot also showed that the GWR model was able to enhance the normality of the model when compared with the normal plots from the OLS model. A further test was done with the GWR model residuals, to investigate the presence of spatial autocorrelation in the data; even though the standard residual maps (see Figure 5.5) indicate that the model residuals were random.

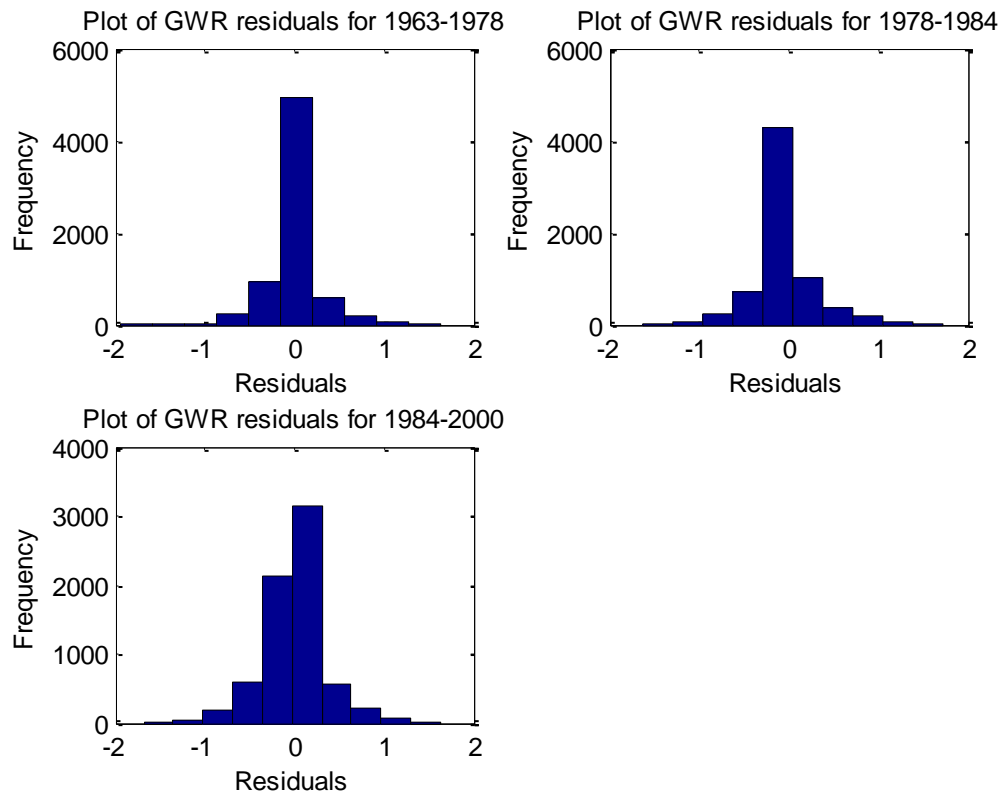


Figure 5.7 Normality plot for GWR model residuals for periods 1963-1978, 1978-1984, and 1984-2000

The ArcGIS Moran's I indices for 1963-1978, 1978-1984, and 1984-2000 (see Table 5.15) indicate that the model residuals were random or autocorrelation free because the Moran's I indices were close to zero. These results indicate that the GWR model reduced autocorrelation effect in the model when compared with results from the OLS model, since its results were closer to zero (see Table 5.7).

Table 5.15 *GWR modelling: spatial autocorrelation test for 1963-1978, 1978-1984, and 1984-2000 using ArcGIS Moran's I*

Periods	Moran's I index for GWR
1963-1978	0.004824
1978-1984	0.005082
1984-2000	0.006313

### 5.7.5 Assessing the impact of excluding the insignificant variable from the OLS and GWR modelling results.

All the independent variables in period 1978-1984 were significant in the model at the 95% CL (see Table 5.4) but each of the variables in periods 1963-1978 and 1984-2000 was not significant at the 95% CL (see Tables 5.3 and 5.5). This experiment aims at assessing the effect of excluding the insignificant variables from the GWR model.

The results of the experiment were as follows: for 1963-1978, OLS multiple  $R^2 = 0.318737$ ; adjusted  $R^2 = 0.317860$ ; and AIC = 12123.751337. GWR:  $R^2 = 0.8396$ ;  $R^2$  adjusted = 0.7934; and AIC = 4580.4994. For 1984-2000, OLS: multiple  $R^2 = 0.381812$ ;  $R^2$  adjusted = 0.380839; and AIC = 15195.751446. GWR:  $R^2 = 0.855443$ ;  $R^2$  adjusted = 0.803408; and AIC = 8230.301948.

For 1963-1978, using the OLS model, the previous multiple  $R^2$ , adjusted  $R^2$ , and AIC results obtained with all the independent variables were better (see previously obtained multiple  $R^2$ , adjusted  $R^2$ , and AIC values using all the independent variables in section 5.7.2; and Table 5.9, respectively). This implies that the independent variable that was insignificant at the 95% CL slightly aided the performance of the OLS model. Using the GWR model, the previous  $R^2$ ,  $R^2$  adjusted, and AIC results obtained with all the independent variables were better for the  $R^2$  value, but worse for the  $R^2$  adjusted and the AIC estimates (see the AIC,  $R^2$ , and  $R^2$  adjusted values

obtained with all the independent variables in Tables 5.9 and 5.10 respectively). This implies that the independent variable that were insignificant at the 95% CL slightly aided the performance of the GWR model with respect to the  $R^2$  estimate and conversely slightly hindered the GWR model with respect to the adjusted  $R^2$  and AIC estimations. The difference between the results using all the independent variable and only the significant variables is very negligible.

From Table 5.4, there was no insignificant independent variable resulting from period 1978-1984.

For 1984-2000, using the OLS model, the previous  $R^2$  obtained with all the independent variables was better but worse for the adjusted  $R^2$ , and AIC estimates. This implies that the independent variable slightly aided the OLS performance with respect to the calculated  $R^2$  estimate but slightly hindered the OLS model with respect to the adjusted  $R^2$ , and AIC estimates. For the GWR model, the computed AIC,  $R^2$ , and  $R^2$  adjusted values were all better than the previous results obtained by using all the explanatory variables. This implies that the insignificant variable slightly hindered the performance of the GWR model.

### **5.7.6 The effect of excluding each independent variable from the GWR model.**

Figure 5.8-5.10 depict the GWR model AIC and  $R^2$  results when each independent variable was excluded from the model one after the other.

From Figure 5.8, 1963-1978; 1=distance to water; 2=distance to residential; 3=distance to industrial and commercial; 4=distance to major roads; 5=distance to railway; 6=distance to Lagos Island; 7=distance to international seaport; 8=distance to University of Lagos; 9=income potential; and 10=population potential. The result in Figure 5.8 indicates that the performance of the GWR model was lowest when distance to residential structures was excluded from the model. This implies that distance to residential structures had the highest impact in the

prediction. Note that the lower the AIC value the better the performance of the model. The GWR model performance was highest when distance to railway was excluded from the model. This implies that distance to railway had the least impact in the prediction.

From Figure 5.9, 1978-1984, the independent variables 1 to 10 are the same as in Figure 5.8. The distance to residential structures had the highest impact on the performance of the model since the accuracy of the prediction was lowest when distance to residential structures was excluded from the model. The distance to international seaport had the lowest impact on the performance of the prediction since the model yielded the highest AIC and  $R^2$  values when distance to international seaport was excluded from the model.

From Figure 5.10, 1984-2000; 2=distance to water; 3=distance to residential; 4=distance to industrial and commercial; 5=distance to major roads; 6=distance to railway; 7=distance to Lagos Island; 8= distance to international airport; 9=distance to international seaport; 10=distance to University of Lagos; 11= distance to Lagos State University; 12=income potential; and 13=population potential. The distance to residential structures had the highest impact in the performance of model since the accuracies of the prediction were lowest when distance to residential structures was excluded from the model. The distance to railway had the lowest impact in the performance of model because the computed accuracies of the prediction were highest when distance to railway was excluded from the model.

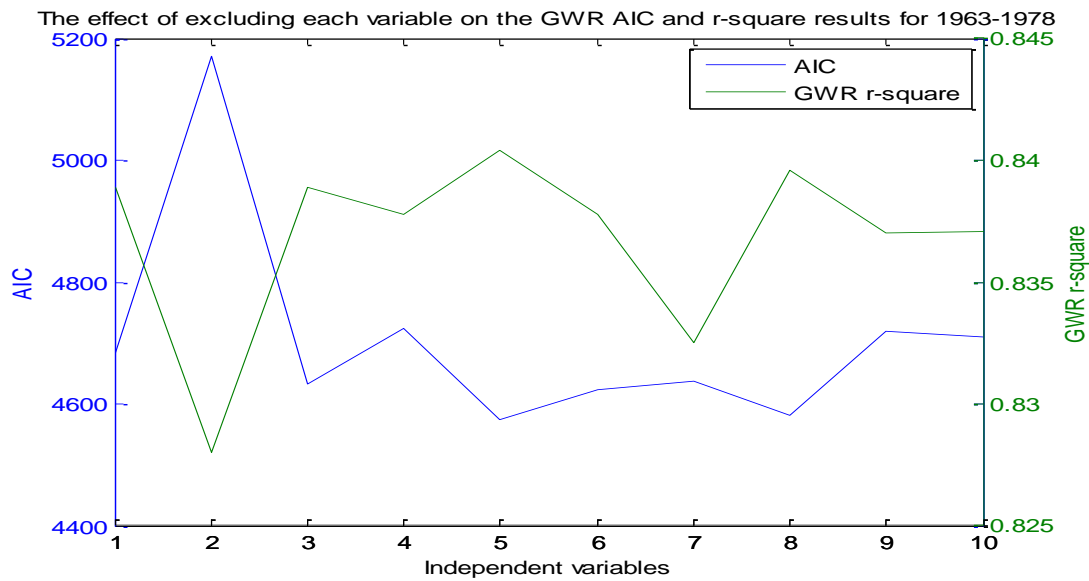


Figure 5.8 The effect of excluding each variable on the GWR estimated AIC and r-square results for 1963-1978

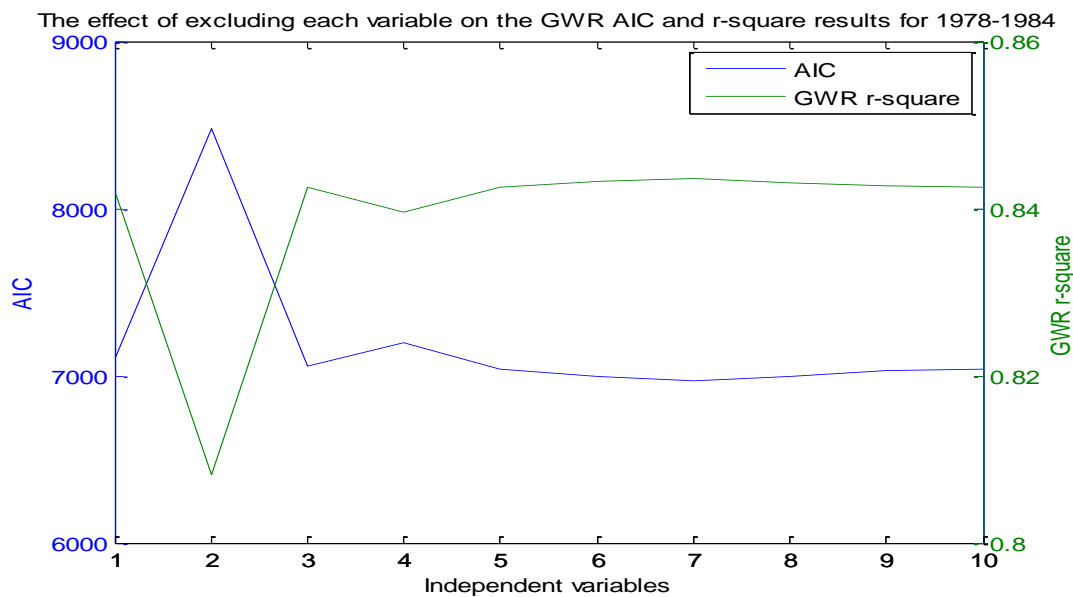


Figure 5.9 The effect of excluding each variable on the GWR estimated AIC and r-square results for 1978-1984



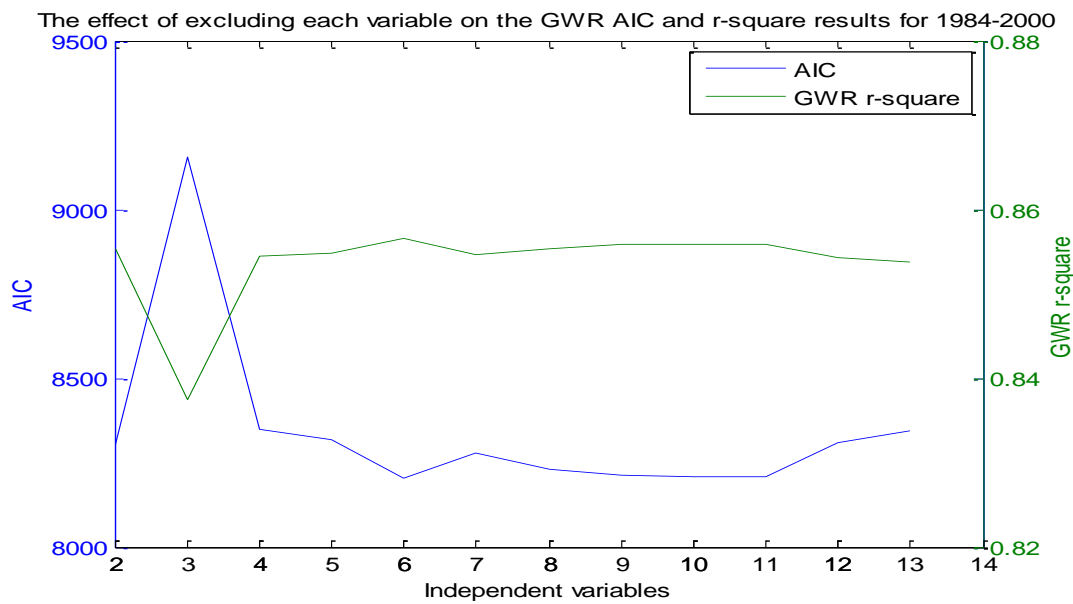


Figure 5.10 *The effect of excluding each variable on the GWR estimated AIC and r-square results for 1984-2000*

The results obtained by excluding each independent variable one at a time corroborate the OLS modelling results presented in Tables 5.3-5.5. The actual prediction was implemented using all the independent variables, due to the fact that all the independent variables had influence on the model based on the OLS results in Tables 5.3-5.5. For an unbiased comparison of the GIS-based results with those from the LR, ANN, SVM, and FSVM results, all the independent variables will be used for the LR, ANN, SVM, and FSVM based CA modelling.

## 5.8 Summary

The  $R^2$  estimates and kappa statistic of the GWR model were satisfactory. Nonetheless, the results from the GIS-based regression modelling indicate that only the multicollinearity and spatial autocorrelation criteria for linear regression models were fulfilled. Normality, linearity, and homoscedasticity conditions were not met. The next section explores the application of the unconventional LR and ANN based CA calibrations.

## CHAPTER 6

### COMMON UNCONVENTIONAL METHODS: APPLICATION OF LOGISTIC REGRESSION AND ARTIFICIAL NEURAL NETWORK BASED CELLULAR AUTOMATA MODELS TO LAND USE CHANGE MODELLING OF LAGOS

#### 6.1 Preamble

This chapter explores the use of Logistic Regression (LR) and Artificial Neural Network (ANN) models for land use change modelling of Lagos. Two simulation approaches will be applied to both models: the non-CA and the CA-based simulations.

#### 6.2 LR-based CA modelling

##### 6.2.1 LR-based CA calibration.

This section presents a brief mathematical illustration of how an LR-based CA model can be derived. Logistic Regression (LR) is the linear regression model usually used in cases where the dependent variable is dichotomous [0, 1]. In this experiment, developed cells = 1, while undeveloped cell = 0. Given a linear function,

$$q = \beta_0 + \sum_{i=1}^n \beta_i x_i, \quad (6.1)$$

where  $q$  is a binary dependent variable,  $\beta_0, \dots, \beta_i$  are logistic regression coefficients to be estimated, while  $x_i$  are independent variables. A logistic regression model can therefore be

expressed as:

$$q = \ln \left( \frac{P(y = 1/x)}{1 - P(y = 1/x)} \right), \quad (6.2)$$

where  $P$  is the probability that  $q = 1$ , given  $x_i$  independent variables;  $\frac{P(y = 1/x)}{1 - P(y = 1/x)}$  is called the odds, while  $\ln \left( \frac{P(y = 1/x)}{1 - P(y = 1/x)} \right)$  is called the logit. Therefore,

$$P = \frac{e^q}{1 + e^q} = \frac{e^{\beta_0 + \sum_{i=1}^n \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^n \beta_i x_i}} = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^n \beta_i x_i)}} \quad (\text{Pohlmann \& Dennis, 2003}) \quad (6.3)$$

Equation 6.3 is the LR-based non-CA model; and  $P$  is the development probability. By introducing the Moore neighbourhood function  $\Omega_{3 \times 3}$  (Wu, 2002), a coefficient  $Q$ , constraints contributions  $cons_{ij}$ , and a stochastic function  $1 + (-\ln \gamma)^\alpha$  (White & Engelen, 1993), equation 6.3 can be revised to derive the final development probability (Liu & Li, 2006; Liu et al., 2007):

$$P_{ij}^t = Q * \left( \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^n \beta_i x_i)}} \right) * \left( 1 + (-\ln \gamma)^\alpha \right) * \Omega_{3 \times 3}^{t-1} * \prod_{i=1}^m cons_{ij} \quad (6.4)$$

where  $\gamma$  is a uniform random variable within the range of  $[0, 1]$ ;  $\alpha$  is a constant that controls the magnitude of the perturbation;  $\Omega_{3 \times 3}^{t-1}$  is an updated function that determines the values of  $P_{ij}^t$  in each iteration;  $Q$  is a coefficient that ensures the values of  $P_{ij}^t$  are confined to  $[0, 1]$ ; and

$\prod_{i=1}^m cons_{ij}$  are the immutable cells that are not affected by the simulation (water and developed cells are considered immutable).

Equation 6.4 is the LR-based CA model. A threshold probability value ( $\psi$ ) is set as a benchmark for determining undeveloped cells that are eligible to transit to developed cells:

$$\begin{cases} P_{ij}^t \geq \psi & \text{developed} \\ \text{Otherwise} & \text{undeveloped} \end{cases} \quad (6.5)$$

$Q$  can also be used to regulate the value of  $P_{ij}^t$  with respect to  $\psi$ , in order to either decrease or increase the number of iterations required for the simulation. The next sub-section presents a brief description of the data used in the modelling and how they were prepared.

### 6.2.2 Data and methodology.

The same data (see section 5.3) used in the GIS-based modelling in chapter 5 were used in this chapter. How the land use data were prepared was described in section 5.5 of chapter 5. This sub-section describes the basic procedures that were undertaken in the course of the modelling (see Figure 6.1).

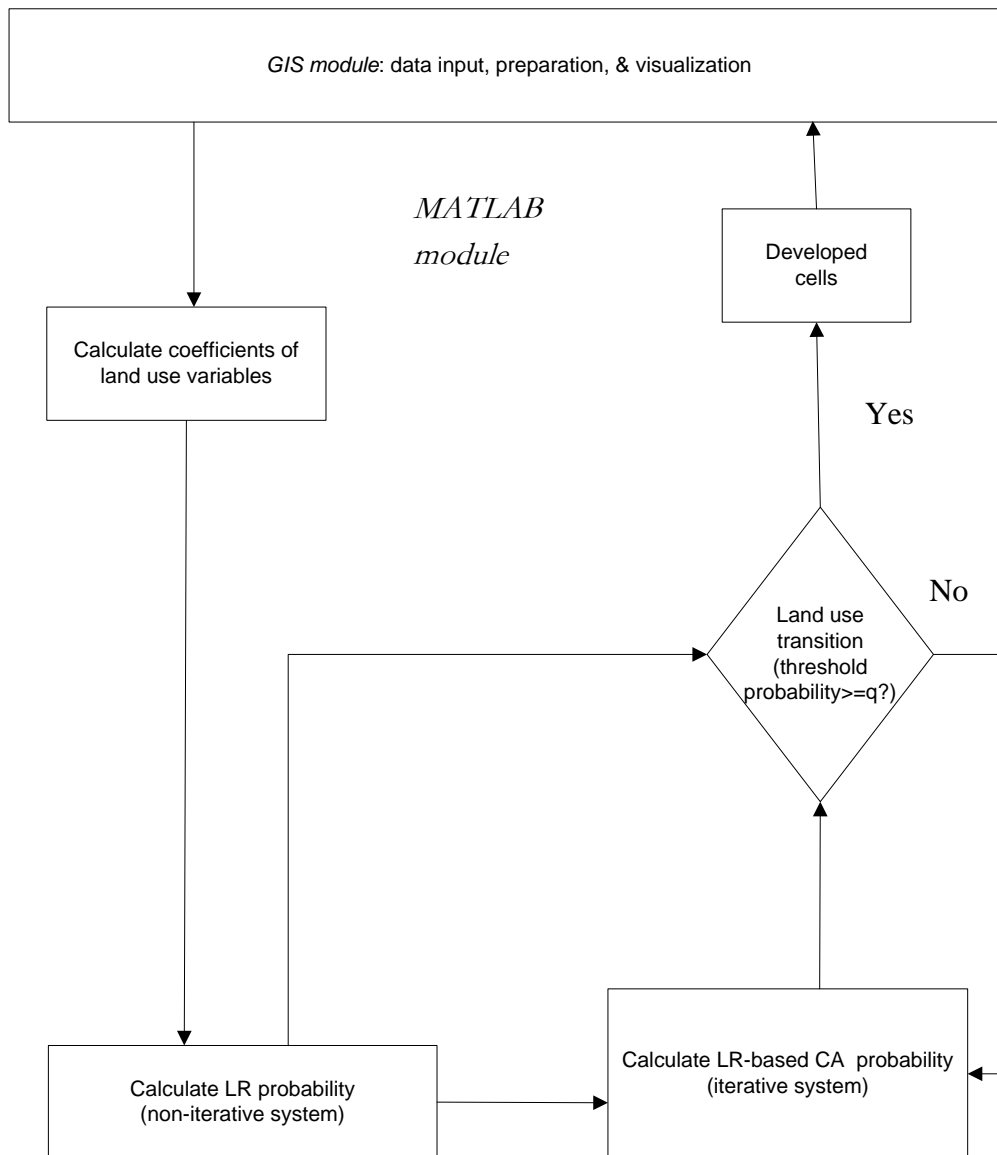


Figure 6.1 *Flowchart for LR modelling*

All the land use independent variables were prepared in the GIS using ArcGIS 9.3 before being imported into MATLAB. The coefficients of the land use variables were first determined. The LR and the OLS (described in chapter 5) can furnish information regarding both the coefficients of the explanatory variables and their significance in the prediction. Unlike the OLS,

LR model does not furnish information on multicollinearity; that is whether the explanatory variables were redundant or not in the simulation. As stated in section 6.1, two stages of land use development were used to assess the evolution of land use: (i) land use change modelling without the use of CA, and (ii) land use change modelling using CA. These two techniques were run through a land use transition module that finally determined the future state of the undeveloped cells. Only undeveloped cells can change their state to developed. The developed cells are immutable in the simulation.

The land use change module without CA is non-iterative while the land use change model using CA is iterative. For the non-CA modelling, the transition probability from undeveloped to developed is when undeveloped cells have development probabilities greater than 0.5 ( $P > 0.5$ ). For CA modelling, cells with probabilities greater or equal to the threshold probability became developed cells while cells with probabilities below the threshold probability were re-admitted into the land use processing module (see Figure 6.1). The final land use maps were visualised in the GIS.

### **6.2.3 Non-CA modelling.**

Unlike the OLS, LR model does not furnish  $R^2$  values (Pohlmann & Dennis, 2003); therefore the overall accuracy and kappa statistic are used as yardsticks for assessing the calibration. This subsection describes the modelling of land use change between 1963-1978, 1978-1984, and 1984-2000. Dummy variables were used to represent the land use dependent variables. The dependent variable is dichotomous (1=developed and 0=undeveloped). The independent variables consist of the explanatory variables for periods 1963-1978, 1978-1984, and 1984-2000 given in Table 5.1. As explained in section 5.7, all the data for the three periods 1963-1978, 1978-1984, and 1984-2000 were scaled to [0, 1].

Each pair of land use map was overlaid to determine the changed regions. The reasons for overlaying these maps were to: (i) extract all the developed and undeveloped points that are common to each pair, and (ii) ensure that the data from the change region were excluded from the training data. The extracted points formed the training data that were used to calculate the LR coefficients for periods 1963-1978, 1978-1984, and 1984-2000 respectively. The calculated LR coefficients for 1963-1978, 1978-1984, and 1984-2000 are presented in Tables 6.1-6.3. Using a two-tailed test, the significance of the explanatory variables was evaluated at 95% CL;  $p$ -values  $< 0.05$  or  $|t| > 1.960$  were considered significant in model. The hypothesis is:

$H_0$ : the logistic regression coefficients are zero at 95% CL

$H_1$ : the logistic regression coefficients are not zero (reject  $H_0$  if  $p$ -value  $< 0.05$ )

Table 6.1 *Logistic regression results for 1963-1978 (\* significant at  $< 0.05$ ;  $t_{critical} = \pm 1.960$ )*

Variable	Coefficient	Std Error	t-statistic	P-value
Intercept	3.707823	0.2647	14.00863	1.3800e-044*
Distance to water	-2.12069	0.7090	-2.99116	0.0028*
Distance to residential	-49.4663	3.8374	-12.8905	5.0900e-038*
Distance to industrial and commercial	-9.12891	0.7985	-11.4323	2.8800e-030*
Distance to major roads	-6.30084	1.4556	-4.32865	1.5000e-005*
Distance to railway	-15.4482	1.2007	-12.8663	6.9600e-038*
Distance to Lagos Island	-3.09282	0.5029	-6.14938	7.7800e-010*
Distance to international seaport	-0.88752	0.4315	-2.05701	0.0397*
Distance to University of Lagos	-1.51889	0.4622	-3.2859	0.0010*
Income potential	0.64486	0.4115	1.567077	0.1171
Population potential	0.191304	0.3566	0.536451	0.5916

Table 6.2 *Logistic regression results for 1978-1984 (\* significant at  $<0.05$ ;  $t_{critical} = \pm 1.960$ )*

Variable	Coefficient	Std Error	t-statistic	P-value
Intercept	4.1945	0.2718	15.4301	1.0300e-053*
Distance to water	-2.2895	0.5830	-3.9269	8.6000e-005*
Distance to residential	-197.2990	9.3164	-21.1775	1.5400e-099*
Distance to industrial and commercial	-4.2734	0.5908	-7.2334	4.7100e-013*
Distance to major roads	-2.2110	0.9224	-2.3969	0.0165*
Distance to railway	-2.5068	0.5290	-4.7371	2.1700e-006*
Distance to Lagos Island	-0.3284	0.4578	-0.7173	0.4732
Distance to international seaport	-0.7065	0.4010	-1.7618	0.0781
Distance to University of Lagos	0.0976	0.4180	0.2334	0.8155
Income potential	1.2384	0.4720	2.6251	0.0087*
Population Potential	-0.5500	0.5180	-1.0622	0.2881

Table 6.3 *Logistic regression results for 1984-2000 (\* significant at  $<0.05$ ;  $t_{critical} = \pm 1.960$ )*

Variable	Coefficient	Std Error	t-statistic	P-value
Intercept	3.9982	0.1983	20.2000	2.0281e-090*
Distance to water	-1.4906	0.4360	-3.4200	6.2950e-004*
Distance to residential	-95.4014	4.1323	-23.1000	6.3240e-118*
Distance to industrial and commercial	-15.1512	0.9895	-15.3000	6.3821e-053*
Distance to major roads	-5.6384	1.0141	-5.5602	2.6948e-008*
Distance to railway	-3.9565	0.3715	-10.6501	1.7422e-026*
Distance to Lagos Island	-1.3604	0.3485	-3.9035	9.4807e-005*
Distance to international airport	-0.5005	0.3218	-1.5553	0.1199
Distance to international seaport	-0.9869	0.3717	-2.6552	0.0079*
Distance to University of Lagos	-0.2014	0.3650	-0.5517	0.5812
Distance to Lagos State University	-3.1174	0.4002	-7.7888	6.7631e-015*
Income potential	0.5985	0.4134	1.4479	0.1477
Population potential	0.3483	0.3191	1.0913	0.2751



For 1963-1978, only income potential and population potential were not-significant at 95% CL. For 1978-1984, distance to Lagos Island, distance to international seaport, distance to University of Lagos, and population potential were not-significant at 95% CL. For 1984-2000, income potential and population potential were not-significant at 95% CL. Just like in the case of the OLS described in chapter 5, an increase in the values of the explanatory variables with positive coefficients will increase the number of developed cells; while a decrease in the values of the explanatory variables with positive coefficients will decrease the number of developed cells in the model; and vice versa. Figure 6.2 depicts the calculated probabilities from the LR modelling results given in Tables 6.1-6.3.

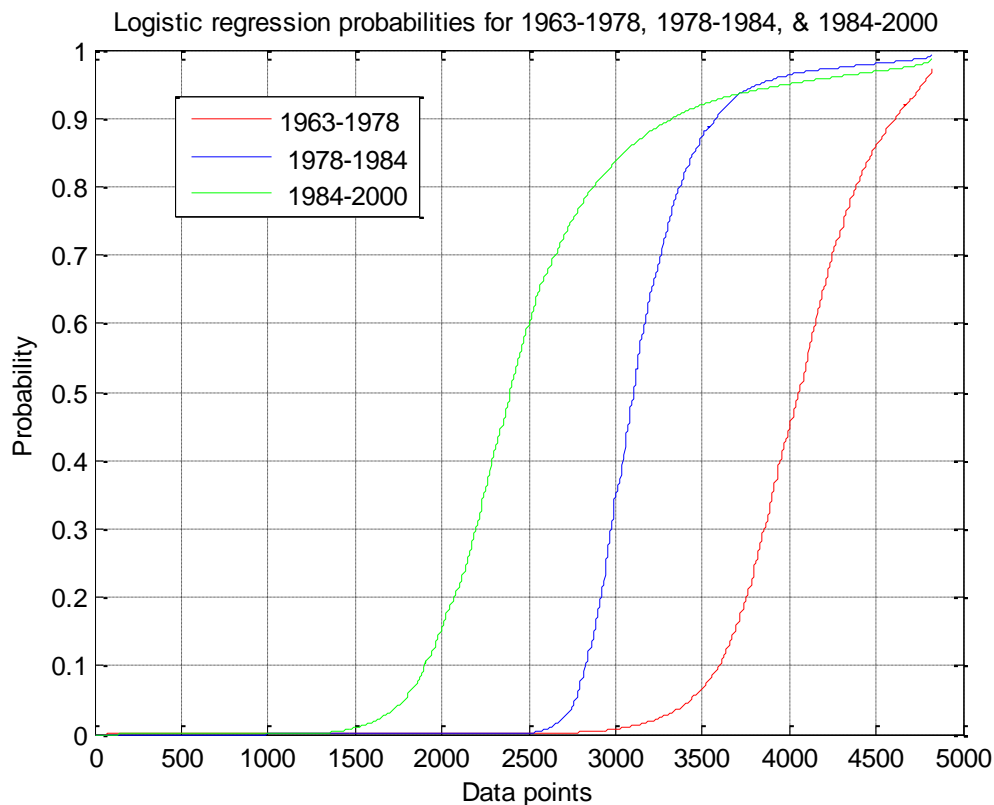


Figure 6.2 *Calculated LR probabilities for 1963-1978, 1978-1984, and 1984-2000*

From Tables 6.1-6.3, not all the independent variables were significant at the 95% CL therefore the leave-one-pair-out holdout model evaluation method (see Bhardwaj et al., 2005) was used to assess the effect of excluding any independent variable from the model. In this experiment, one thousand sample size data were selected, and split into two equal sizes, 500 points as the training data and the other 500 points as the test data. For 1963-1978, using all the independent variables, the calculated kappa coefficient was 0.8320, and the overall accuracy was 91.60%. Using only the significant variables, the calculated kappa coefficient was 0.8240, while the overall accuracy was 91.20%. This implies that the insignificant variables aided the accuracy of the model. For 1978-1984, using all the independent variables, the calculated kappa coefficient was 0.8880, and the overall accuracy was 94.40%. Using only the significant variables, the calculated kappa coefficient was 0.9000, while the overall accuracy was 95.00%. This implies that the insignificant variables compromised the accuracy of the model. For 1984-2000, using all the independent variables, the calculated kappa coefficient was 0.7920, and the overall accuracy was 89.60%. Using only the significant variables, the calculated kappa coefficient was 0.7720, while the overall accuracy was 88.60%. This implies that the insignificant variables compromised the accuracy of the model.

Each independent variable was then excluded from the model one at a time to assess the effect the exclusion of each independent variable will have on the accuracy of the LR model. The results of the exclusion of each independent variable from the LR are presented in Figures 6.3-6.5.

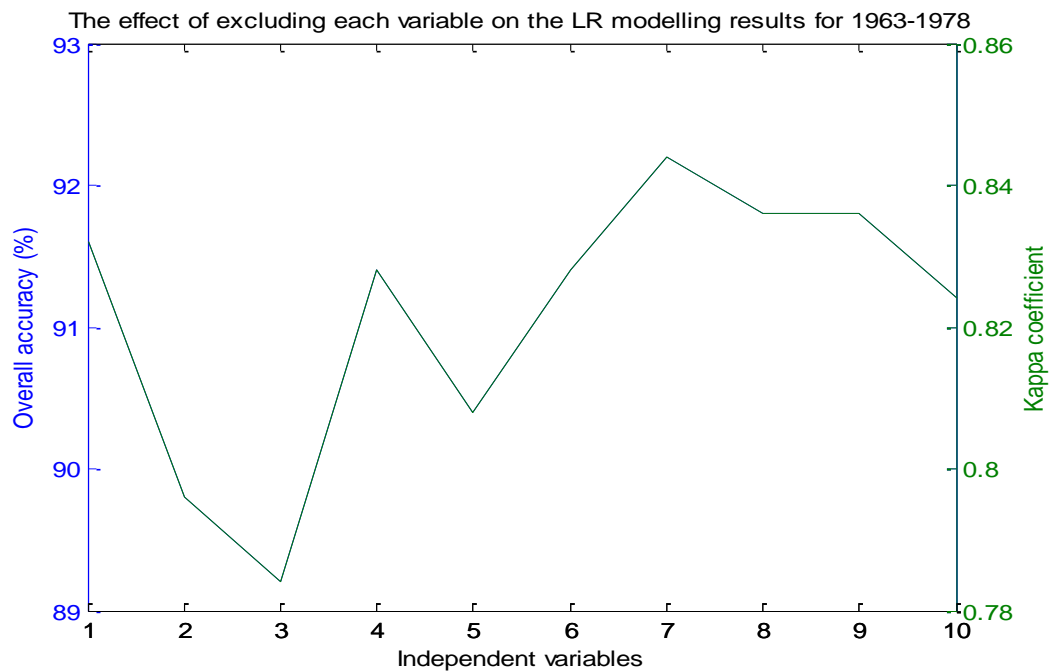


Figure 6.3 Showing the effect of excluding each variable from the model for 1963-1978

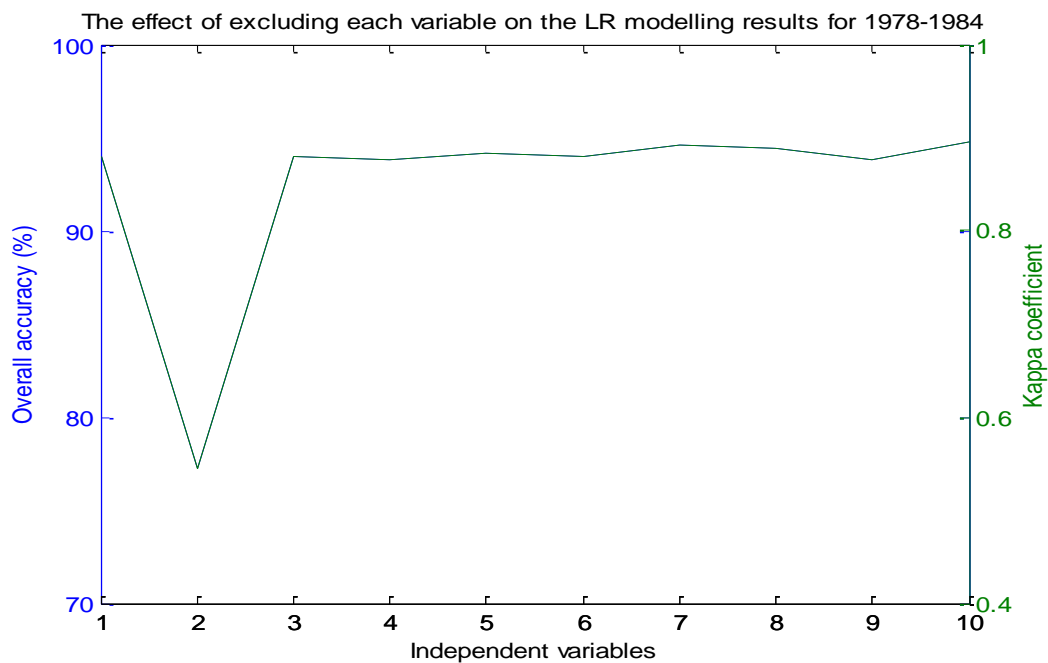


Figure 6.4 Showing the effect of excluding each variable from the model for 1978-1984

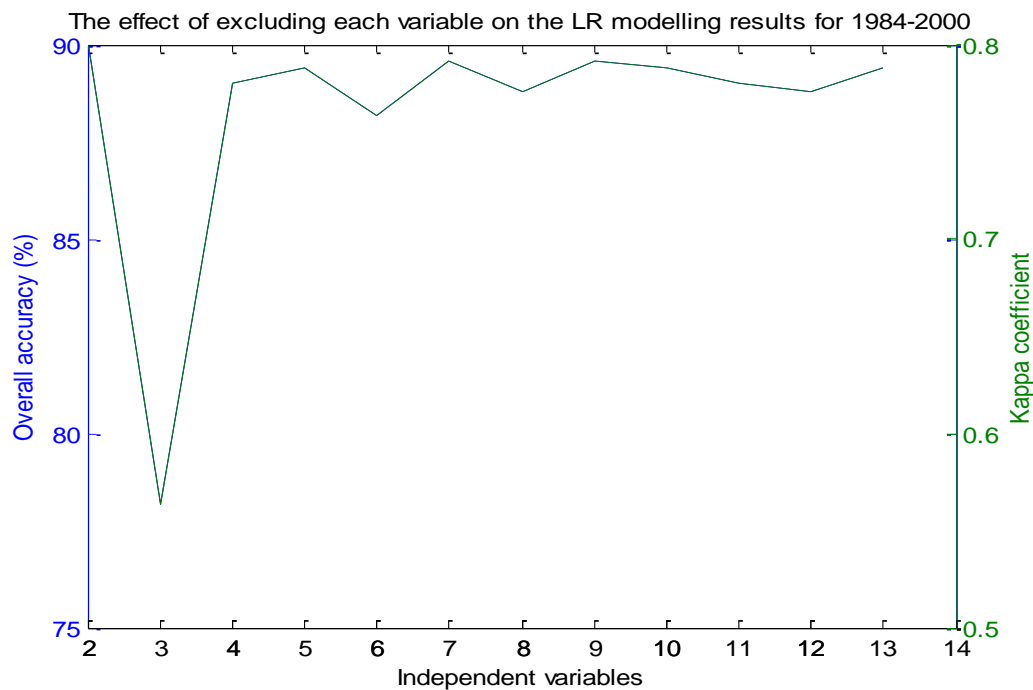


Figure 6.5 Showing the effect of excluding each variable from the model for 1984-2000

From Figure 6.3, 1963-1978; 1=distance to water; 2=distance to residential; 3=distance to industrial and commercial; 4=distance to major roads; 5=distance to railway; 6=distance to Lagos Island; 7=distance to international seaport; 8=distance to University of Lagos; 9=income potential; and 10=population potential. The calculated kappa coefficient and overall accuracy were lowest in the model when distance to industrial and commercial was excluded from the model; this implies that distance to industrial and commercial had the highest influence in the prediction. Figure 6.3 shows that distance to international seaport had the lowest influence on the model.

From Figure 6.4, 1978-1984; the independent variables 1 to 10 are the same as in Figure 6.3. The calculated kappa coefficient and overall accuracy were lowest in the model when distance to residential structures was excluded from the model; this implies that distance to residential structures had the highest influence in the prediction. Population potential had the lowest

influence on the model, because the model yielded its highest kappa coefficient and overall accuracy estimates when the population potential variable was excluded from the model.

From Figure 6.5, 1984-2000; 2=distance to water; 3=distance to residential; 4=distance to industrial and commercial; 5=distance to major roads; 6=distance to railway; 7=distance to Lagos Island; 8= distance to international airport; 9=distance to international seaport; 10=distance to University of Lagos; 11= distance to Lagos State University; 12=income potential; and 13=population potential. The calculated kappa coefficient and overall accuracy were lowest in the model when distance to residential structures was excluded from the model; this implies that distance to residential structures had the highest influence in the prediction. The highest kappa coefficient and overall accuracy estimates were obtained when distance to water was excluded from the model; which suggests that distance to water may have one of the lowest influences on the prediction. The findings in this experiment are not totally consistent with the LR test of hypothesis results presented in Tables 6.1-6.3.

From Tables 6.1-6.3, the variables with higher  $|t|$  values are more significant in the LR model than the variables with lower  $|t|$  values. The conclusion from the leave-one-pair-out holdout evaluation experiment results presented in Figures 6.3-6.5 is that the independent variables that were insignificant at the 95% CL are important in the LR model; this is because, distance to University of Lagos yielded the lowest  $|t|$  value in test of hypothesis experiment for period 1984-2000 given in Table 6.3; but the result presented in Figure 6.5 for period 1984-2000 indicated that distance to water had the lowest influence on the LR model. A reason for obtaining a relatively inconsistent result with the holdout experiment could be because all the calculated  $|t|$  values given in Table 6.3 are close, apart from the  $|t|$  value for distance to residential and distance to industrial and commercial that yielded high estimates of  $|t|$ .

All the calculated  $p$ -values presented in Tables 6.1-6.3 were all far from the maximum probability of 1.0000, therefore it can be deduced that all the explanatory variables contributed to the model, even though few of the explanatory variables were not significant at 95% CL. For the purpose of comparing the LR results with the GIS-based, ANN, SVM, and FSVM models all the independent variables were used for the GWR, LR, ANN, SVM, and FSVM models.

The non-CA modelling experiment is a classification experiment. The outcome of the classification is either developed or undeveloped. Cells with probabilities  $>0.5$  are classified as developed while cells with probabilities  $<0.5$  are classified as undeveloped. To predict the non-CA maps for periods 1963-1978, 1978-1984, and 1984-2000, the stratified random sampling approach was used to extract 1000 training points from each combination of periods 1963-1978, 1978-1984, and 1984-2000.

The two objectives of this modelling are to: (i) predict land use change between the periods, and (ii) evaluate the accuracy of the LR model in predicting land use change for periods 1963-1978, 1978-1984, and 1984-2000. The selected 1000 points were split into 10 equal datasets. Each dataset had a sample size of 100. Each dataset consists of 50 developed cells/points and 50 undeveloped cells/points. As usual, the developed cells were labelled +1 while the undeveloped cells were labelled 0. Note that, all the data for periods 1963-1978, 1978-1984, and 1984-2000 were scaled to  $[0, 1]$  (see section 5.7). The  $k$ -fold cross-validation technique (see section 3.5.3) (where  $k=10$ ) was used to evaluate the accuracy of the LR model in the prediction. The model was trained by putting together 9 subsets ( $k-1$ ) out of the 10 datasets while 1 subset was used to test the accuracy of the prediction. The experiment was repeated in 10 folds by eventually using all the 10 datasets for both training and testing. The kappa coefficient and overall accuracy were computed for each  $k$  dataset (see Figures 6.6-6.8).

The LR model invokes the land use change between periods 1963 and 1978, 1978 and 1984, and 1984 and 2000, based on training samples only selected from the regions/points common to 1963 and 1978, 1978 and 1984, and 1984 and 2000. The training data must not be extracted from the regions where change occurred between the three periods. This is the reason why the two land use maps between each of the three periods must be overlaid in order to detect the changed regions.

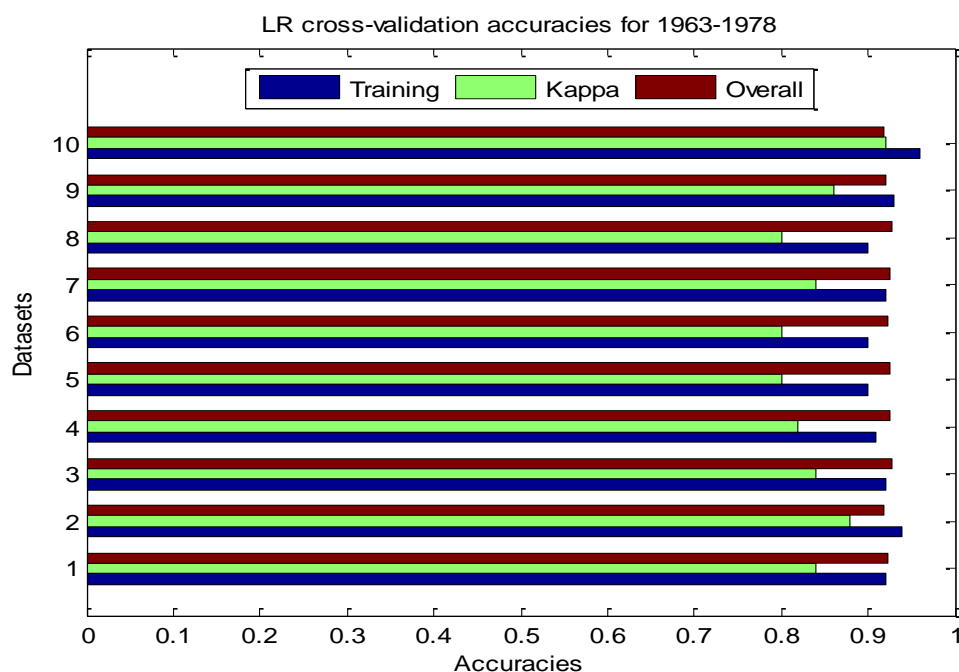


Figure 6.6 *LR modelling results for 1963-1978*

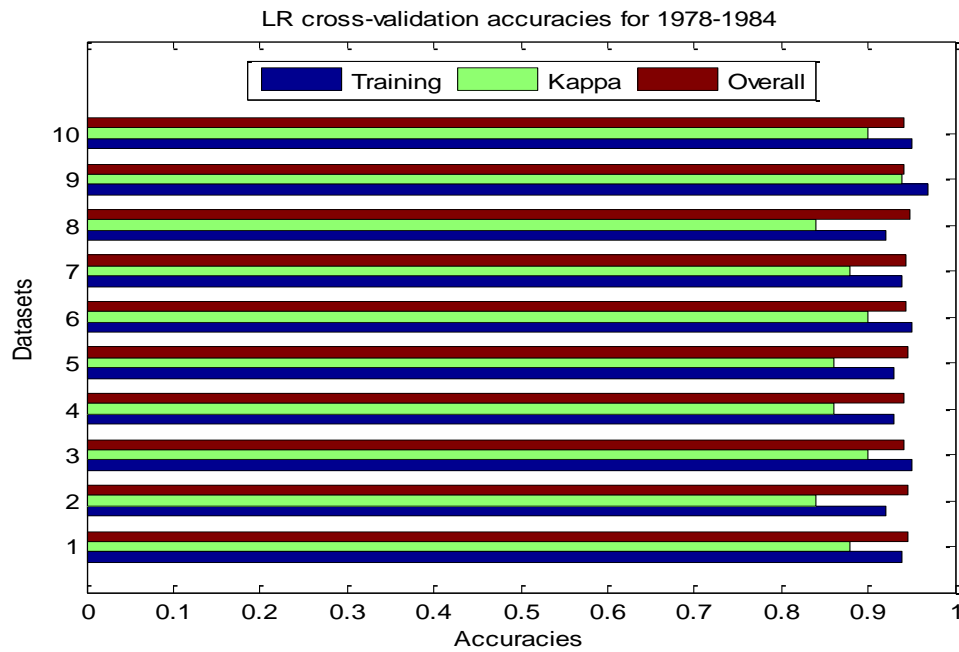


Figure 6.7 *LR modelling results for 1978-1984*

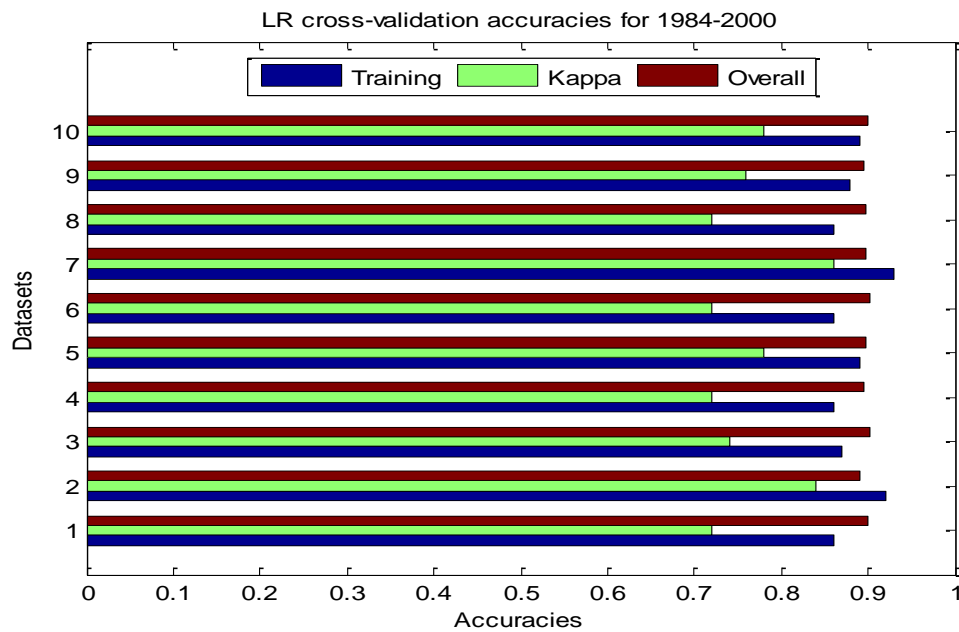


Figure 6.8 *LR modelling results for 1984-2000*

Using equation 6.3, development probabilities  $>0.5$  were classified as developed cells, while probabilities  $<0.5$  were classified as undeveloped cells. The kappa and overall accuracy results



shown in Figures 6.6-6.8 can help evaluate the strength of the model; it can perhaps be likened to the  $R^2$  value in the OLS modelling described in chapter 5.

The LR model accuracy was calculated by computing the mean accuracies from the kappa coefficients of all the datasets presented in Figures 6.6-6.8. The calculated model accuracies for periods 1963-1978, 1978-1984, and 1984-2000 were 0.8944, 0.9213, and 0.8479 respectively. The calculated accuracies for the three periods show that the highest kappa coefficient estimate was obtained with period 1978-1984 while the lowest kappa coefficient estimate was obtained with period 1984-2000.

The best maps yielded by the prediction for periods 1963-1978, 1978-1984, and 1984-2000 are given in Figure 6.9; and their computed confusion matrices are presented in Tables 6.4-6.6. The calculated kappa coefficients for the predicted maps given in Figure 6.9 were determined by comparing/validating the predicted maps with the actual maps in 1978, 1984, and 2000. The calculated kappa coefficients were 0.5057, 0.5525, and 0.5080 for periods 1963-1978, 1978-1984, and 1984-2000 respectively. The next subsection examines the bottom-up simulation results from the LR-based CA simulation.

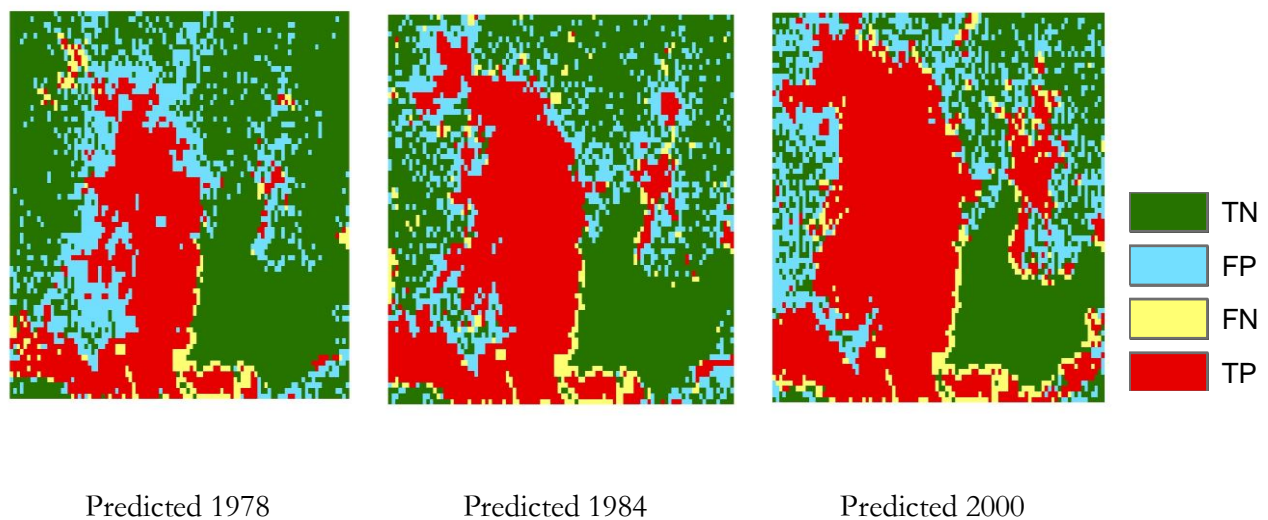


Figure 6.9 *LR predicted maps for periods 1963-1978, 1978-1984, and 1984-2000*

Table 6.4 *LR-based non-CA: confusion matrix for period 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1578	1387
Undeveloped	214	3821

Table 6.5 *LR-based non-CA: Confusion matrix for period 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	2224	1283
Undeveloped	284	3209

Table 6.6 *LR-based non-CA: Confusion matrix for period 1984-2000*

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	3019	1317
Undeveloped	413	2251

#### 6.2.4 CA modelling.

The basic difference between the CA model and the non-CA model is simply the incorporation of the Moore neighbourhood function  $\Omega_{3 \times 3}$  into the LR model (see equation 6.4). CA models simply use the neighbourhood influence of the initial state of the observed object and the independent variables to predict the future state of the target object. The bottom up approach of CA models makes their predictions highly accurate (Torrens & O'Sullivan, 2001).

There is no training of data in the CA modelling, the LR model has already been trained in the previous section. Only the neighbourhood function is updated to determine the conversion of undeveloped cells to develop. CA are iterative systems; for example, to predict from 1963-1978, the starting point of the simulation will be the 1963 data/map (that is iteration=0 is 1963), while the target map 1978 is used to validate the prediction. At iteration>0, the neighbourhood function  $\Omega_{3 \times 3}^{t-1}$  calculates the number of developed cells surrounding each undeveloped cell in the 1963 map. This process is repeated in each iteration.

Two hundred iterations were run to simulate the maps from periods 1963-1978, 1978-1984, and 1984-2000 (see Figure 6.11). There are no rules that guide the number of iterations needed for a simulation, but 100 to 200 iterations are common in most applications (Wu, 2002; Li & Yeh, 2004). Figure 6.10 depicts the mean kappa coefficients and standard deviations for 10,20,30,...,200 designated iteration thresholds, calculated by running each threshold ten times and comparing the simulated maps with the actual maps for periods 1963-1978, 1978-1984, and 1984-2000.

From Figure 6.10, at iteration=10 the mean kappa coefficients for 1963-1978, 1978-1984, and 1984-2000 were low, but gradually increased as the number of iterations increased. The highest mean kappa coefficients were found from 90 to 140 iterations. The mean kappa coefficients decreased beyond 140 iterations until the 200<sup>th</sup> iteration. The final maps for periods 1963-1978, 1978-1984, and 1984-2000 shown in Figure 6.11 were obtained from 90th to 170th iterations. The calculated kappa coefficients from the confusion matrices given in Tables 6.7-6.9 for the predicted maps for periods 1963-1978, 1978-1984, and 1984-2000 were 0.5847, 0.7543, and 0.7101. The highest kappa statistic estimate was obtained with 1978-1984 while the lowest was obtained with 1963-1978.

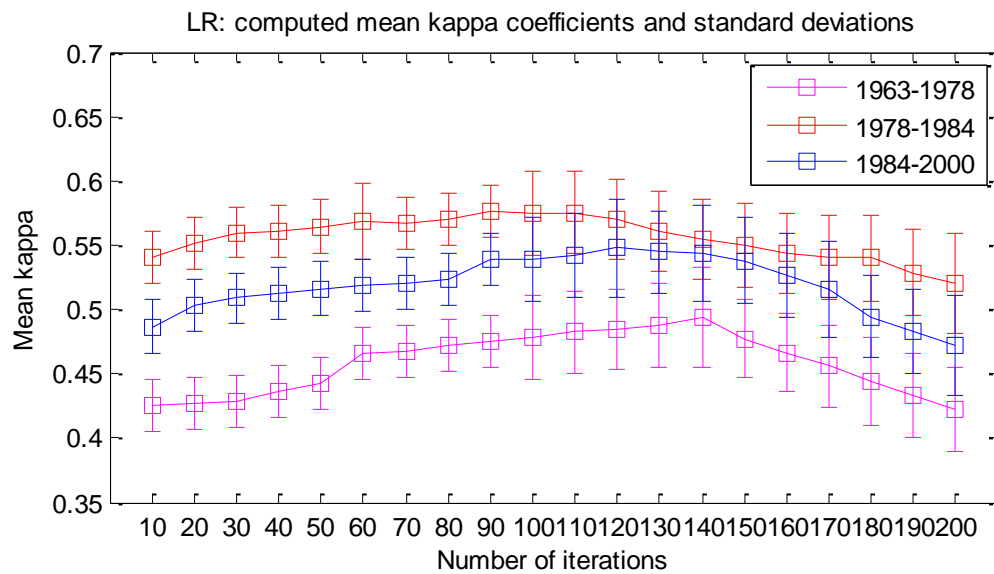


Figure 6.10 *LR modelling result: computed overall mean kappa and standard deviations for 200 designated iteration thresholds*

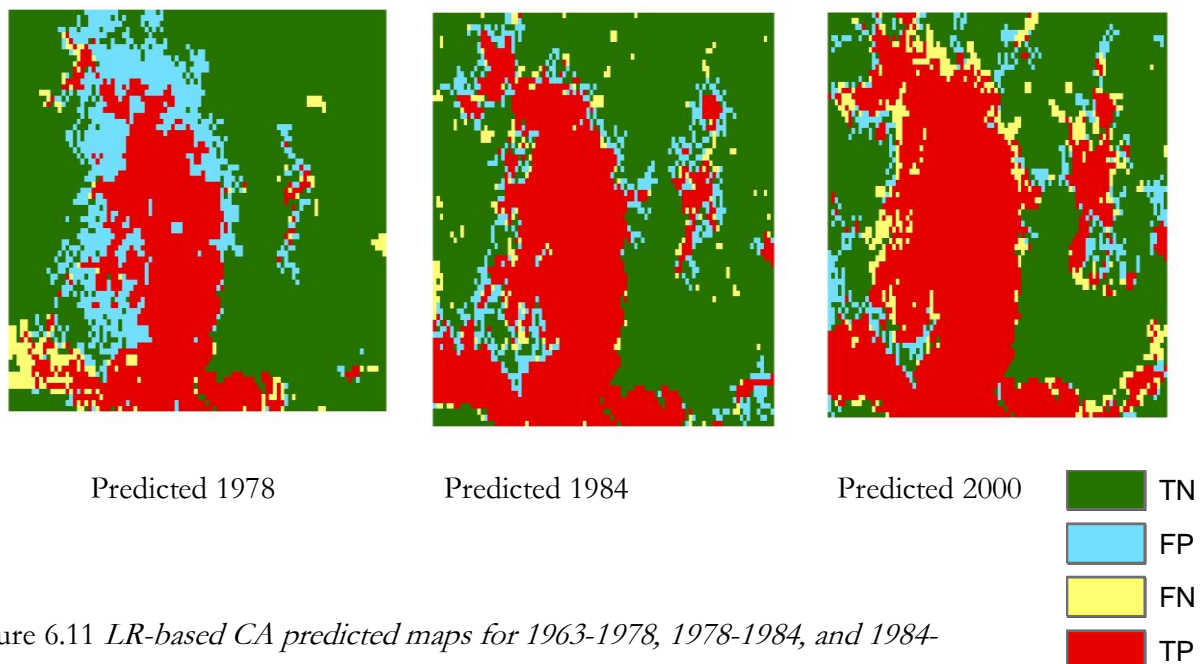


Figure 6.11 *LR-based CA predicted maps for 1963-1978, 1978-1984, and 1984-2000*

Table 6.7 *LR-based CA: confusion matrix for period 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1566	1048
Undeveloped	226	4160

Table 6.8 *LR-based CA: confusion matrix for period 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	2280	588
Undeveloped	228	3904

Table 6.9 *LR-based CA: confusion matrix for period 1984-2000*

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	2889	471
Undeveloped	543	3091

Evaluating the performance of the CA model can be intricate. The Receiver Operating Characteristics (ROC) was used to assess the performance of the LR-based CA model. The ROC is the plot of sensitivity against 1-specificity. Sensitivity is calculated by dividing the number of the true positive matches by the sum of the true positive and false negative matches; while specificity is calculated by dividing the number of the true negative matches by the sum of the true negative and false positive matches. The Area Under Curve (AUC) determines the result of the plot. Experiments that yield AUC indices  $<0.5$  are usually regarded as worthless. Figure 6.12 depicts the plot of mean sensitivity against mean 1-specificity, and their respective standard

deviations calculated from 10 ROC curves sampled at fixed 1-specificity points: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 (see Fawcett, 2004). The mean sensitivity and mean 1-specificity was calculated by comparing the simulated maps with the actual maps. The calculated AUC for 1963-1978, 1978-1984, and 1984-2000 were  $0.7354 \pm 0.0295$ ,  $0.7549 \pm 0.0267$ ,  $0.7451 \pm 0.0298$ . The calculated ROC results corroborate the results from the CA predicted maps because the order of best fit of the target maps were still: 1978-1984, 1984-2000, and 1963-1978 respectively.

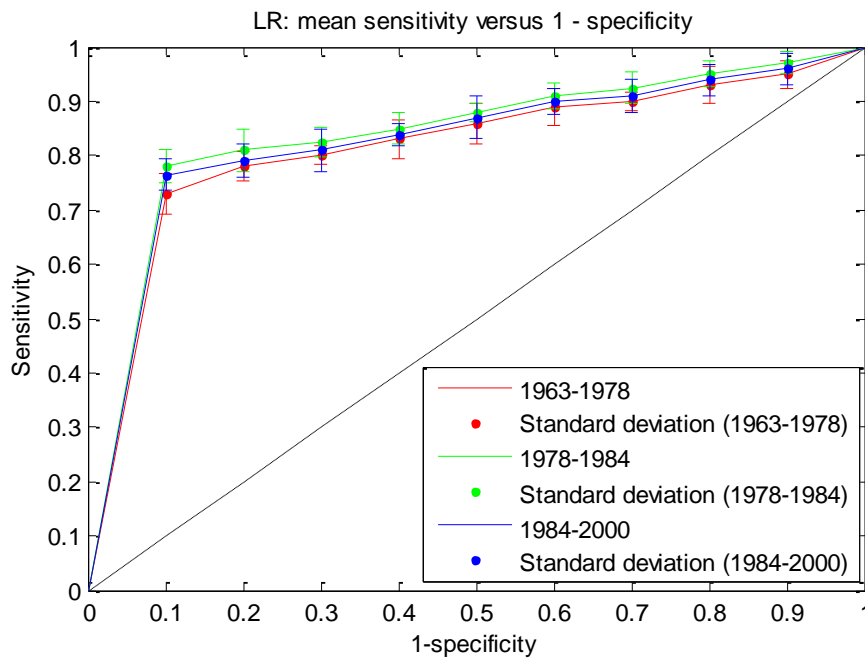


Figure 6.12 LR: plotted mean sensitivity versus 1-specificity and standard deviations for periods 1963-1978, 1978-1984, and 1984-2000

## 6.3 ANN-based CA modelling

### 6.3.1 ANN-based CA calibration.

A brief introduction on the ANN architecture is presented in section 2.3.2.6. A signal from neuron  $i$  of the first input layer of a cell  $x$ , at time  $t$  received by a neuron  $j$  of the hidden layer can be expressed as,

$$net_j(x,t) = \sum_j W_{i,j} S'_i(x,t) \quad (6.6)$$

where  $S'_i(x,t)$  denotes the site attributes given by variable (neuron)  $i$ ;  $W_{i,j}$  is the weight of the input from neuron  $i$  to neuron  $j$ ;  $net_j(x,t)$  is the signal received for neuron  $j$  of cell  $x$  at time  $t$ . The activation of the hidden layer of the signal is:

$$\frac{1}{1 + e^{-net_j(x,t)}} \quad (6.7)$$

The probability can be expressed as:

$$P(x,t) = \sum_j W_j \frac{1}{1 + e^{-net_j(x,t)}} \quad (6.8)$$

where,  $P$  is the development probability; as in case of the LR model given in equation 6.3. The final development probability can be written as,

$$P_{ij}^t = Q * \left( \sum_{i=1} W_j \frac{1}{1 + e^{-net_j(x,t)}} \right) * \left( 1 + (-\ln \gamma)^\alpha \right) * \Omega_{3 \times 3}^{t-1} * \prod_{i=1}^m cons_{ij} \quad (\text{Li \& Yeh, 2002}) \quad (6.9)$$

Equation 6.9 is the ANN-based CA model. The formula for land use transition from undeveloped to developed cells given in equation 6.5 (in the case of the LR-based CA model) is the same for the ANN-based CA model.

### 6.3.2 Data and methodology.

The procedure adopted in this section is virtually the same as that of the LR model described in section 6.2.2. The ANN experiment was implemented in MATLAB with the method of backpropagation (Rumelhart et al., 1986) using a two-layer feed-forward neural network.

The backpropagation algorithm is one of the most commonly used ANN methods (Asmadi et al., 2009). A backpropagation network has an input layer, an output layer, and one or two hidden layers; however, there is no limit to the number of hidden layers (Anderson, 1995; Chauvin & Rumelhart, 1995). It is very important to first choose the *random seed number* and the required *number of neurons* in the hidden layer (see Figure 6.13).

A neural network is initialised with initial weights; hence different results are always obtained every time the ANN model is run. To ensure the results remain the same at every run of the neural network the random seed number must be set. The random seed number is an arbitrary constant chosen by trial-and-error. After the random seed number is set, the ANN neurons then remain the only parameter that can be adjusted to vary the simulation results of the ANN. There seems to be no-hard-and-fast rule for determining the required number of neurons in the hidden layer to train a neural network (Almeida et al, 2008). A  $2n+1$  neurons may be required to produce a good fit of the neural network, where  $n$  is the number of neurons in the input layer (Li & Yeh, 2002);  $2n/3$  may produce a similar fit as the  $2n+1$ , with lesser computer processing time ( $n$  represents the number of independent variables that are input into the network through the input layer). The training of the neural network is simply the adjustment of the number of neurons in the hidden layer in order to minimise the training error. The training error is the discrepancy between the predicted and the actual value. The adjustment of the number of neurons is sustained until a training error that falls below a pre-determined threshold is found (Wang, 1994).



As in the case of the LR, two categories of land use development will be used for the modelling: (i) simulation without using CA, and (ii) simulation using CA. The transition

module (see Figure 6.13) determines the final land use state of undeveloped cells. All the final simulated maps were visualised in the GIS. The next sub-section describes the ANN modelling without the use of CA.

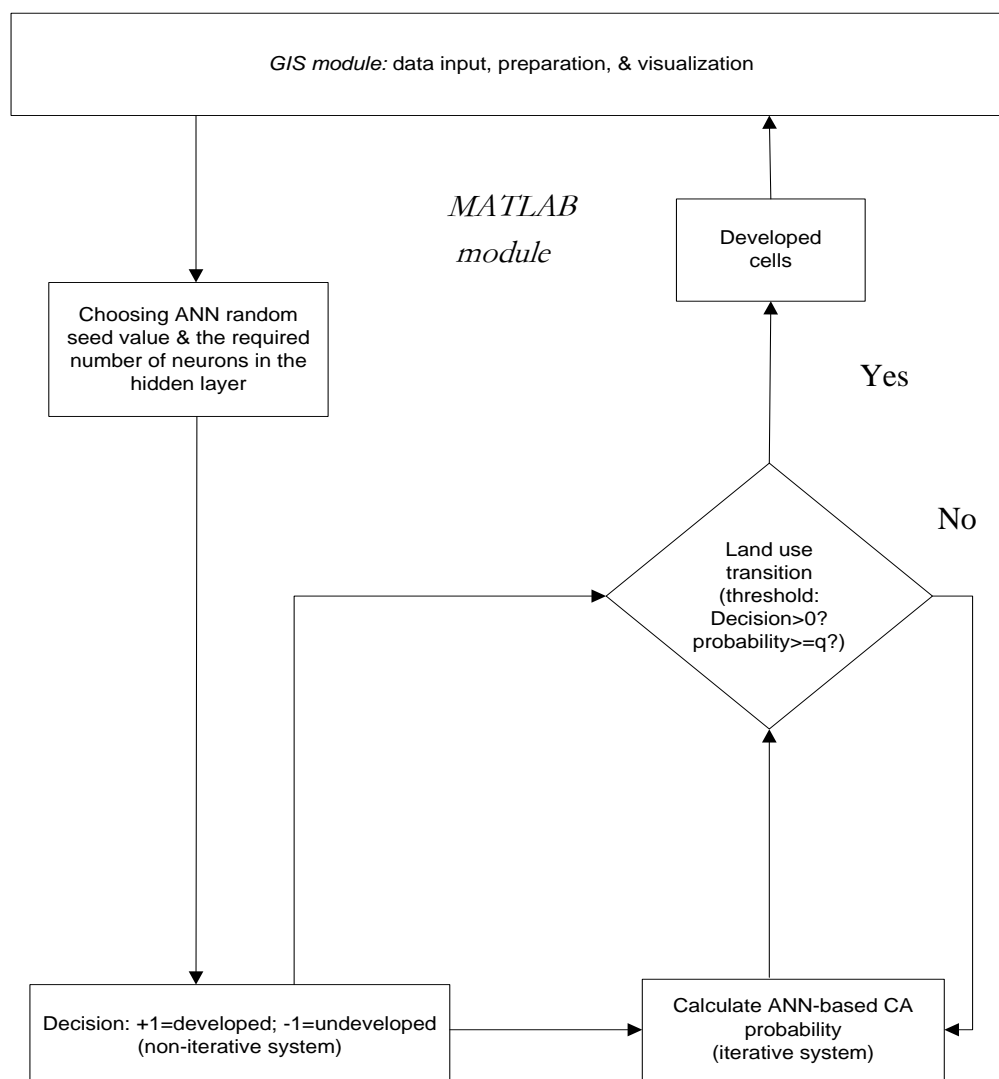


Figure 6.13 *Flowchart for ANN modelling*

### 6.3.3 CA modelling.

The same 1000 training points selected for each of periods 1963-1978, 1978-1984, and 1984-2000, were used in the ANN experiment. Just like the case of the LR modelling the experiment was implemented with the  $k$ -fold cross-validation procedure. Since there is no precise technique for selecting an optimal number of neuron required for training the network (or optimising the number of neurons in the hidden layer), 14 designated number of neurons: 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50 were used for optimising the number of neurons in the hidden layer. The dependent variables were represented with developed= +1, and undeveloped= -1.

In this experiment the 'random seed number' was set at 3558583436. Each of the designated neurons was used to compute the kappa coefficient for the ANN model with respect to periods 1963-1978, 1978-1984, and 1984-2000 (see Figures 6.14-6.16). The mean kappa coefficient computed for each designated neuron computed from Figures 6.14-6.16 is given in Figure 6.17.

From Figure 6.14, period 1963-1978, 3 neurons yielded the highest overall mean kappa coefficient while 50 neurons yielded the lowest. For 1978-1984, 2 neurons yielded the highest overall mean kappa coefficient while 50 neurons yielded the lowest. For 1984-2000, 1 neuron yielded the highest overall mean kappa coefficient while 50 neurons yielded the lowest.

The ANN model accuracy was determined by calculating the average of the mean kappa coefficients for periods 1963-1978, 1978-1984 from the mean kappa coefficients for each designated number of neurons given in Figure 6.18. The calculated model accuracies for periods 1963-1978, 1978-1984, and 1984-2000 were 0.8186, 0.8611, and 0.7354 respectively.

The number of neurons with the highest mean kappa coefficient (resulting from Figure 6.17) was used to predict the land use change for periods 1963-1978, 1978-1984, and 1984-2000 (see Figure 6.18). The computed confusion matrices for the predicted maps are given in Tables

6.10-6.12. The calculated kappa coefficients for the predicted maps given in Figure 6.18 were 0.5165, 0.5701, and 0.5149 for periods 1963-1978, 1978-1984, and 1984-2000 respectively.

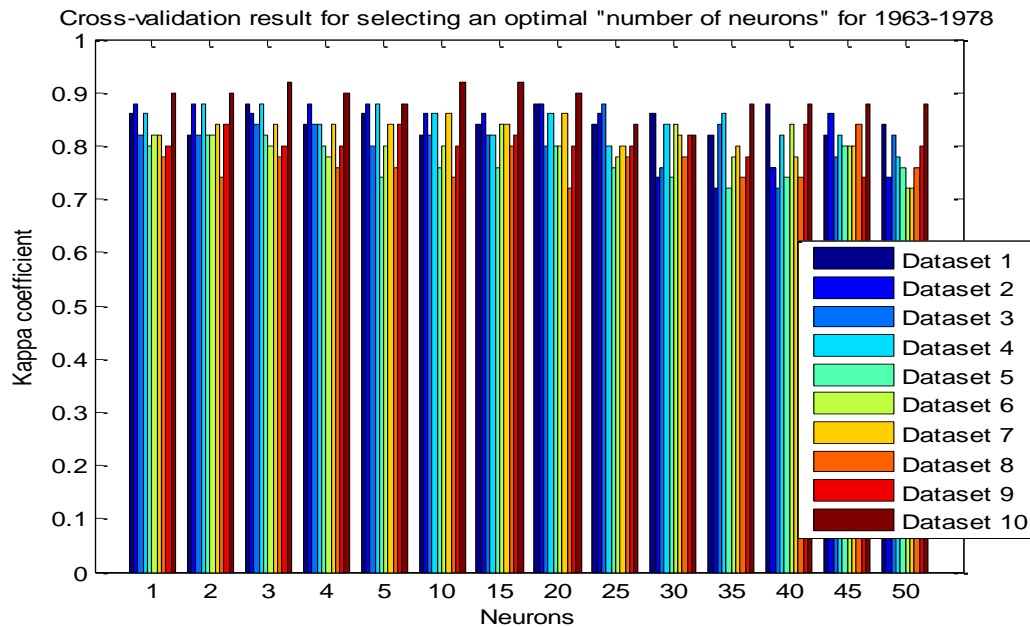


Figure 6.14 *Computed mean kappa for each dataset based on designated ANN neurons (1963-1978)*

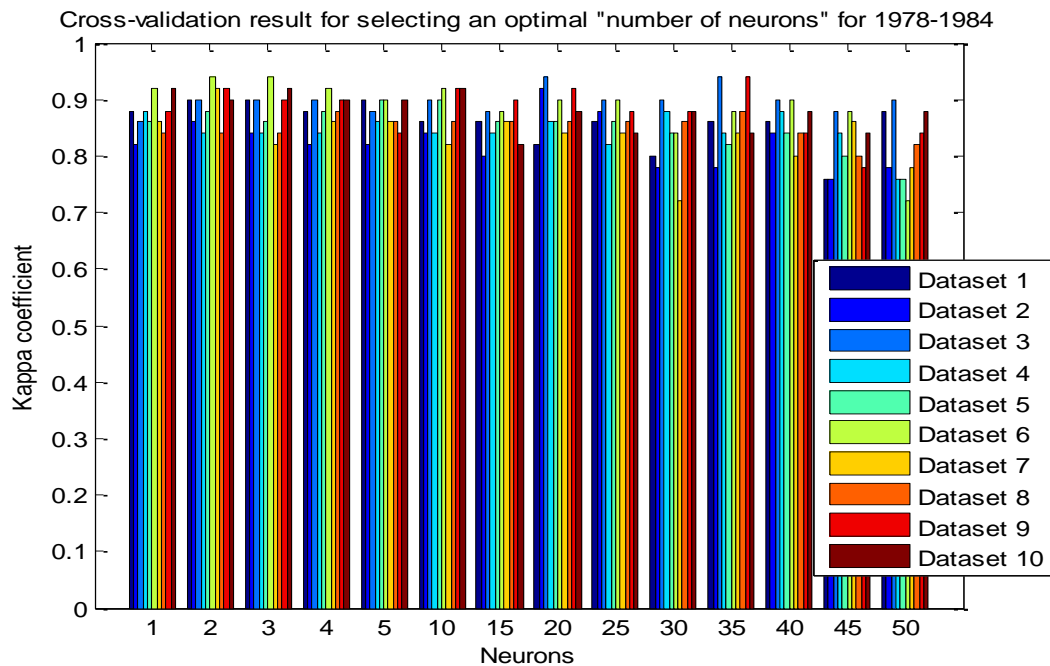


Figure 6.15 Computed mean kappa for each dataset based on designated ANN neurons (1978-1984)

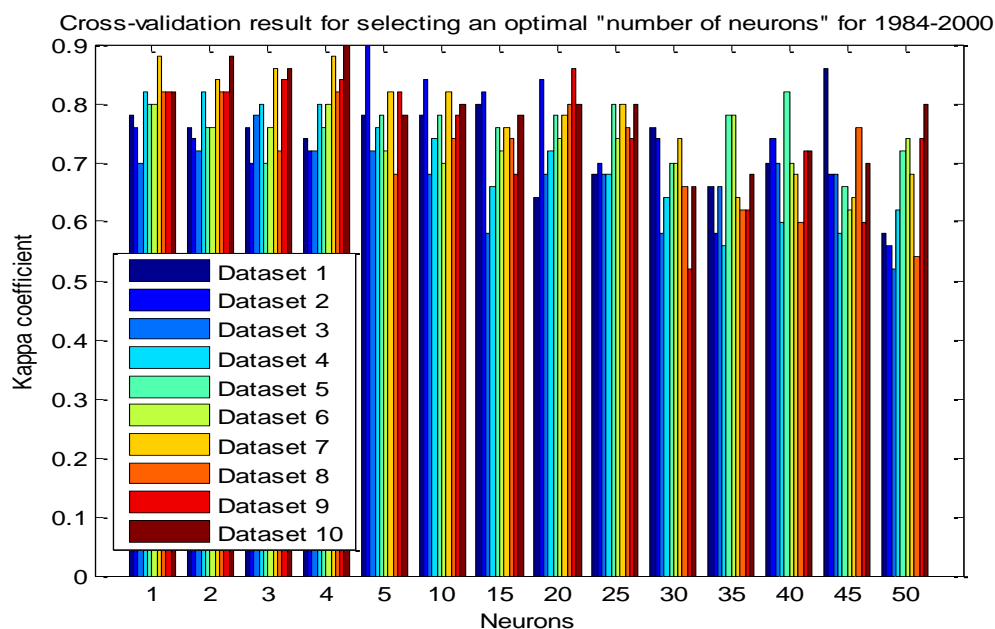


Figure 6.16 Computed mean kappa for each dataset based on designated ANN neurons (1984-2000)

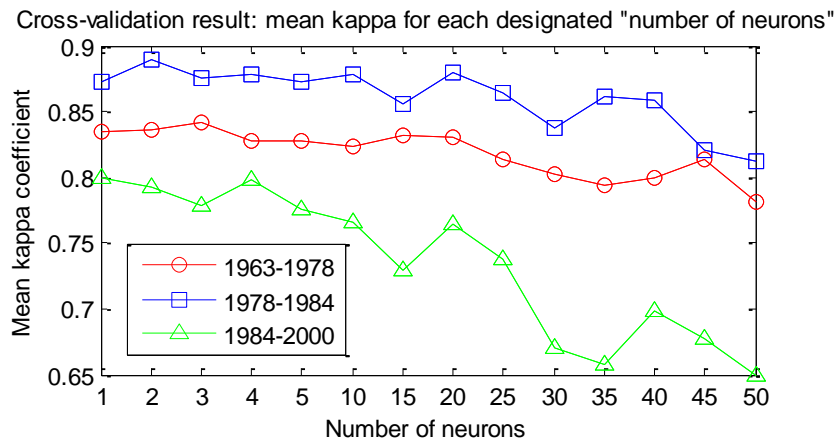


Figure 6.17 *The ANN overall mean kappa for 1963-1978, 1978-1984, and 1984-2000 based on the designated neurons*

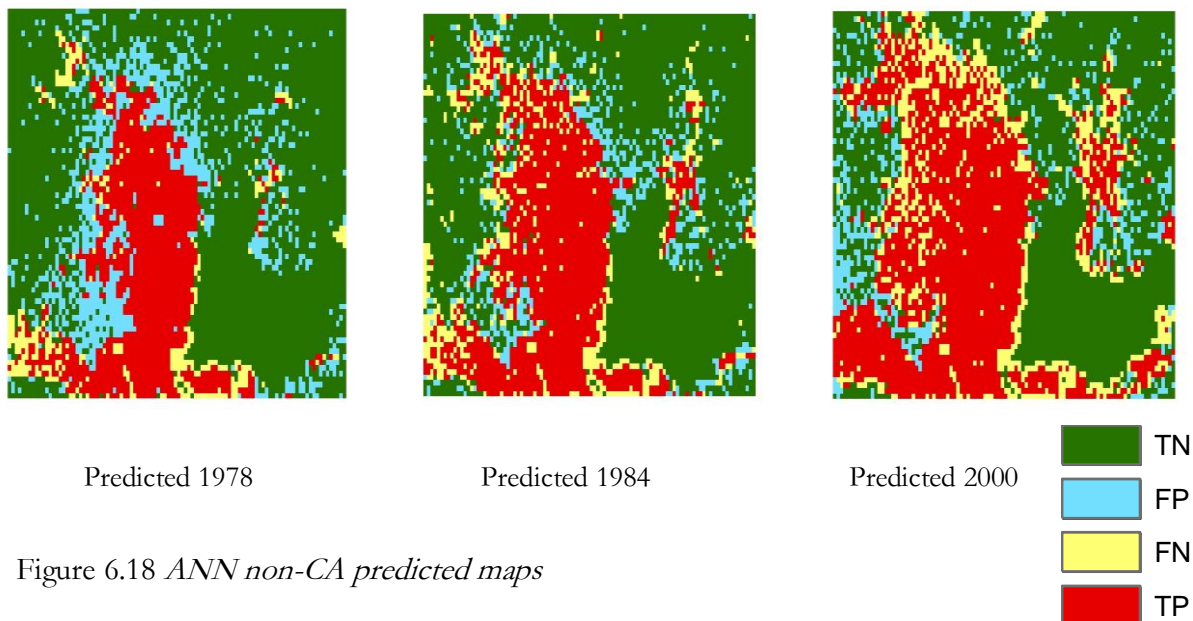


Figure 6.18 *ANN non-CA predicted maps*

Table 6.10 *ANN-based non-CA: confusion matrix for period 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1429	1100
Undeveloped	363	4108

Table 6.11 *ANN-based non-CA: confusion matrix for period 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	1806	679
Undeveloped	702	3813

Table 6.12 *ANN-based non-CA: confusion matrix for period 1984-2000*

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	2407	669
Undeveloped	1025	2899

In another experiment, the results of the input weights from a one neuron ANN (see Table 6.13) was compared with the derived LR coefficients given in Tables 6.1-6.3. Note that, a one neuron ANN network is not a black-box because its input weights can be equated with the LR coefficients. A comparison of the one neuron ANN input weights (see Table 6.13) and the LR coefficients (see Tables 6.1-6.3) shows that both results are basically the same. The LR coefficients are about 4 times higher than those of the one neuron ANN.

Unlike the LR, the one neuron ANN did not yield the expected negative values for the proximity variables and positive values for the weighted variables for period 1984-2000. However, the result of the one neuron ANN is correct because the designated values for the output layer for the experiment was [-1, +1]. This is the reason why the opposite signs were outputted rather than the usual signs.

Table 6.13 *Input weights from a one neuron neural network for periods 1963-1978, 1978-1984, and 1984-2000*

	1963-1978	1978-1984	1984-2000
	Input weights	Input weights	Input weights
Distance to water	-0.9023	-0.4545	0.13579
Distance to residential	-10.5806	-35.6294	20.3025
Distance to industrial and commercial	-2.5625	-0.8232	3.2947
Distance to major roads	-4.0632	-0.6593	1.04639
Distance to railway	-4.8828	-0.2645	0.6693
Distance to Lagos Island	-0.5946	-0.0060	0.1587
Distance to international airport	Not applicable	Not applicable	0.0745
Distance to international seaport	-0.2302	-0.1045	0.1877
Distance to University of Lagos	-0.4228	0.0577	0.0308
Distance to Lagos State University	Not applicable	Not applicable	0.5797
Income potential	0.2682	0.2062	-0.0791
Population potential	0.1141	0.0034	-0.0206
Input weight bias	-21.8324	-36.3103	25.0515
Model bias	0.0396	0.0765	0.1394

In another experiment using the LR and the one neuron ANN, the results from the  $k$ -fold cross-validation (computed from Figures 6.6-6.8 and Figure 6.17) given in Figure 6.19 indicate that but for period 1984-2000, that the computed mean kappa coefficients for the LR was higher than the those of the one neuron ANN for periods 1963-1978 and 1978-1984. Nonetheless, the

difference in the results from both methods is marginal. Therefore, it can be deduced that a one neuron ANN is an equivalent of the LR.

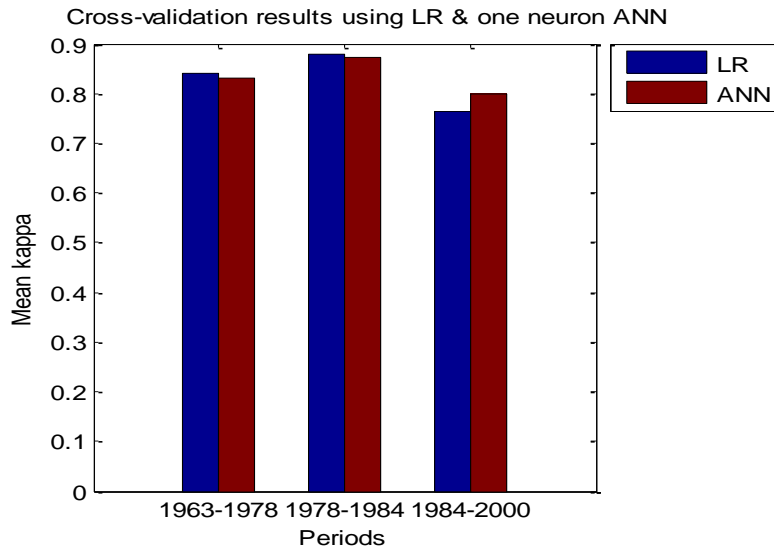


Figure 6.19 Computed mean kappa coefficients from the cross-validation results from the LR and a one neuron ANN

#### 6.3.4 CA modelling.

Both the predicted non-CA and CA maps were predicted with the optimal number of neurons selected with the  $k$ -fold cross-validation procedure. CA modelling is therefore implemented by introducing the Moore neighbourhood function  $\Omega_{3 \times 3}^{t-1}$  into the neural network.

Two hundred iterations were run to simulate the maps from periods 1963-1978, 1978-1984, and 1984-2000 (see Figure 6.21). Figure 6.20 depicts the mean kappa coefficients and standard deviations for 10,20,30,...,200 designated iteration thresholds, calculated by running each iteration threshold ten times and comparing the simulated maps with the actual maps for periods 1963-1978, 1978-1984, and 1984-2000. For 1978-1984 and 1984-2000, the highest mean kappa coefficients were yielded at the 130<sup>th</sup> iterations, while for 1963-1978, the highest mean kappa coefficient was yielded at the 120<sup>th</sup> iterations.



The predicted maps given in Figure 6.21 were obtained near thresholds with the highest mean kappa coefficient. The calculated kappa coefficients computed from the confusion matrices given in Tables 6.14-6.16 for the predicted maps for periods 1963-1978, 1978-1984, and 1984-2000 were: 0.5900, 0.7825, and 0.7161 respectively.

Figure 6.22 presents the ROC plot for periods 1963-1978, 1978-1984, and 1984-2000. The calculated AUC and their respective standard deviations were  $0.7574 \pm 0.0304$ ,  $0.7850 \pm 0.0308$ , and  $0.7613 \pm 0.0288$ , for 1963-1978, 1978-1984, and 1984-2000 respectively. The calculated AUC estimates indicate that the order of best fit of the observed data was: periods 1978-1984, 1984-2000, and 1963-1978. The computed ROC results corroborate the kappa coefficient results of the CA-based predicted maps.

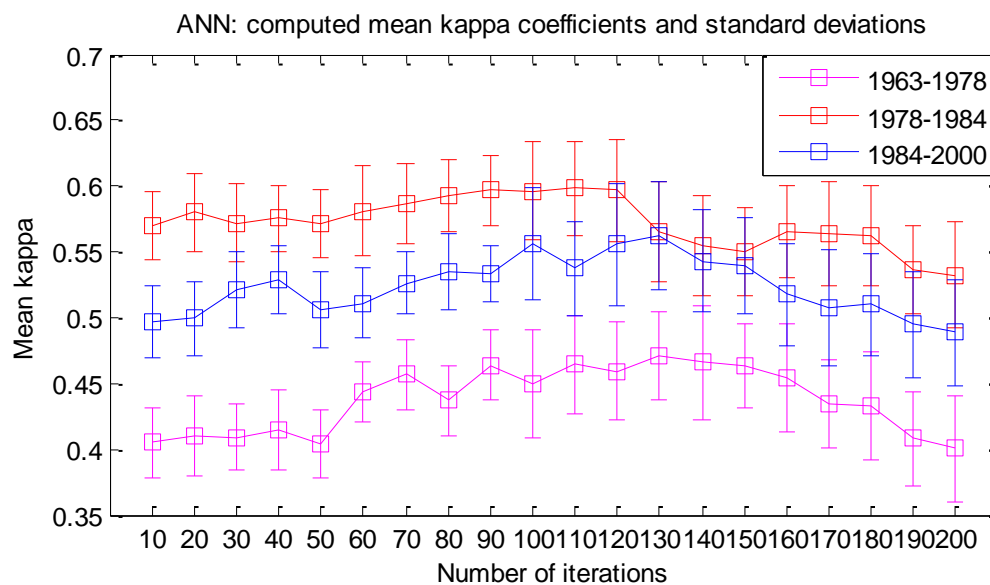


Figure 6.20 ANN result: overall mean kappa and standard deviations for 200 designated iteration thresholds

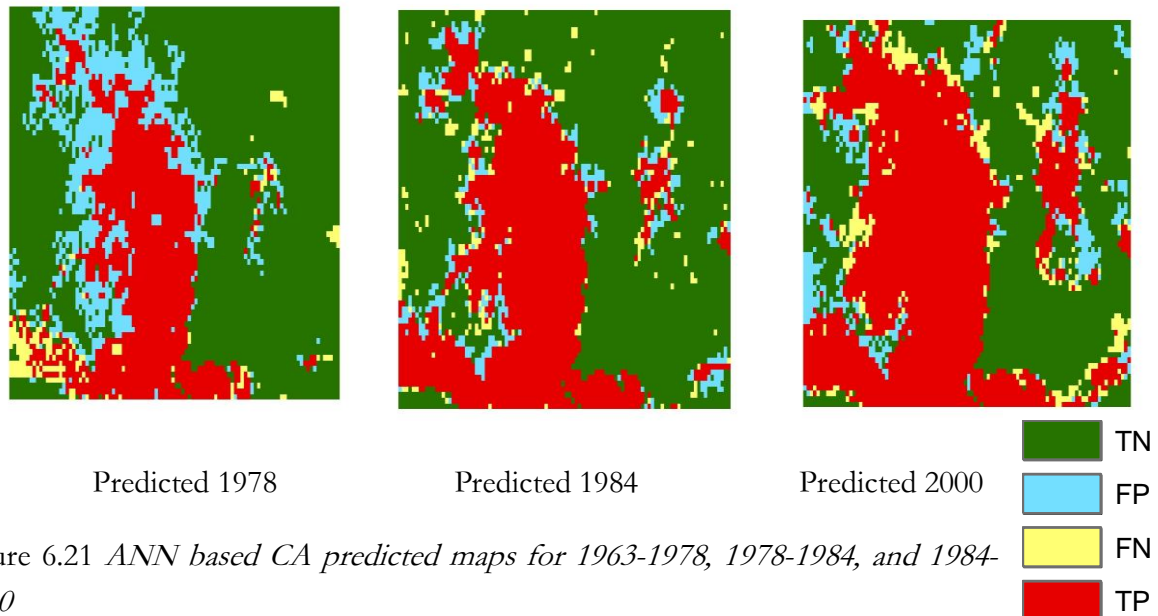


Figure 6.21 ANN based CA predicted maps for 1963-1978, 1978-1984, and 1984-2000

Table 6.14 ANN-based CA: confusion matrix for period 1963-1978

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1609	1092
Undeveloped	183	4116

Table 6.15 ANN-based CA: confusion matrix for period 1978-1984

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	2232	434
Undeveloped	276	4058

Table 6.16 ANN-based CA: confusion matrix for period 1984-2000

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	2996	558
Undeveloped	436	3010

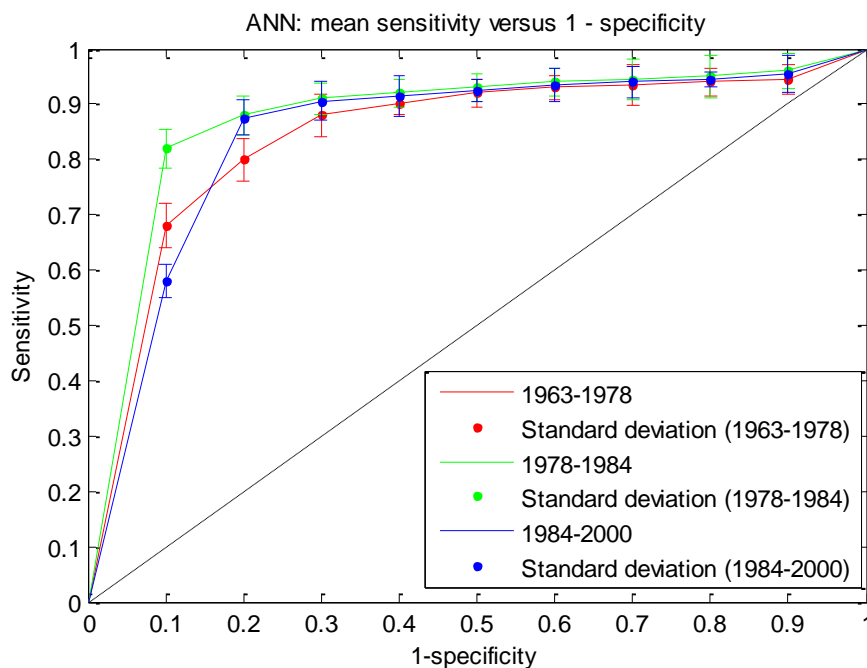


Figure 6.22 ANN-based CA ROC analysis: plotted mean sensitivity versus 1-specificity and standard deviations for 1963-1978, 1978-1984, and 1984-2000

## 6.4 Summary

The ANN kappa coefficients for both the non-CA and CA modelling were slightly higher than those of the LR model. The calculated AUC estimates for the ANN model were also moderately higher than those of the LR model. These results imply that the ANN model was fairly more robust to the Lagos data than the LR model.

The non-CA modelling accuracies from the LR and ANN models may not be an unbiased standard for assessing the accuracies of the LR and ANN models. This is because for the case of the ANN model, fourteen designated neurons were used to optimise the number of neurons in the hidden layer of the ANN model. This process may reduce the model accuracy of the ANN model since all the calculated kappa coefficients from all the designated neurons were averaged to determine the model accuracy. The calculated kappa coefficients of some of the designated neurons were relatively low and therefore reduced the overall model accuracy of the ANN model since the kappa coefficients from those neurons were used to calculate the model accuracy of the ANN model. The LR model does not have any parameter that must be optimised. Due to these reasons, the model accuracies from both the LR and ANN models were considered of no consequence in evaluating their accuracy.

## CHAPTER 7

### PROPOSED UNCONVENTIONAL METHODS: APPLICATION OF SUPPORT VECTOR MACHINE AND FUZZY SUPPORT VECTOR MACHINE BASED CELLULAR AUTOMATA MODELS TO LAND USE CHANGE MODELLING OF LAGOS

#### 7.1 Preamble

This chapter presents the use of the SVM and FSVM based CA models for modelling land use change in Lagos. As was in the case of the LR and ANN, the modelling is implemented in two stages: the non-CA and the CA based.

#### 7.2 SVM-based modelling

##### 7.2.1 SVM-based CA calibration.

From chapter 3, SVM output  $f(x) = \text{sign} \left[ \sum_{i=1}^n y_i \alpha_i^0 K(x_i, x) + b^0 \right]$  given in equation 3.16 of section 3.3 can be mapped into probabilities using a sigmoid function (Platt, 1999). Therefore, SVM-based land use development probability can be expressed as,

$$P = \frac{1}{1 + e^{-(\text{sign}[\sum_{i=1}^n y_i \alpha_i^0 K(x_i, x) + b^0])}} \quad (7.1)$$

By introducing the Moore neighbourhood function  $\Omega_{3 \times 3}$  (Wu, 2002), a coefficient  $Q$ , constraints contributions  $cons_{ij}$ , and a stochastic function  $1 + (-\ln \gamma)^\alpha$  (White & Engelen,

1993), equation 7.1 can be revised as the final development probability (Okwuashi et al., 2008; Okwuashi et al., 2009a; Okwuashi et al., 2009b),

$$P_{ij}^t = Q * \left( \frac{1}{1 + e^{-(\text{sign}[\sum_{i=1}^n y_i \alpha_i^0 K(x_i, x) + b])}} \right) * (1 + (-\ln \gamma)^\alpha) * \Omega_{3 \times 3}^{t-1} * \prod_{i=1}^m \text{cons}_{ij} \quad (7.2)$$

The equation for the conversion of cells from undeveloped to developed is the same as equation 6.5. Equation 3.16 is the SVM-based non-CA model, while equation 7.2 is the SVM-based CA model.

### 7.2.2 Data and methodology.

The first stage of the experiment is the preparation of data. The second stage is modelling in MATLAB; and the third stage is the visualisation of results in the GIS. The flowchart given in Figure 7.1 illustrates the processes of the modelling. The extracted land use variables for 1963-1978, 1978-1984, and 1984-2000 were described in section 5.6. The land use variables were prepared in the GIS.

Similar to the ANN, the SVM model is a black-box (Cristianini & Shawe-Taylor, 2000). Unlike the LR and OLS models, SVM models do not furnish information about the significance or contributions of respective independent variables used in the model.

The training data consist of developed and undeveloped cells. Developed cells were labelled +1 while undeveloped cells were labelled -1. The polynomial, RBF, and linear kernels were used for all the experiments. The SVM parameters (regularisation parameter  $C$ , polynomial kernel degree  $d$ , and RBF kernel gamma  $\gamma$ ) were first optimised (or selected) using a  $k$ -fold cross-validation procedure (where  $k=10$ ). The training data were classified, by determining the solutions of alpha  $\alpha_i$  using the optimisation equations 3.12 and 3.15. The simulation was

enhanced by varying the kernel parameters of the RBF and polynomial kernels. The optimisation equations were solved using Quadratic Programming (QP) (Gunn, 1998; Vapnik, 2000).

The exitflag value informs whether the result of the training is acceptable or not. Exitflag values other than 1 are unacceptable. The non-CA and the CA models were evaluated upon satisfactory training accuracy based on the support vectors furnished by  $\alpha_i$ . In cases where unacceptable accuracies were obtained, the process was repeated by readjusting the model parameters until desirable results were yielded. All the predicted maps were visualised in the GIS.

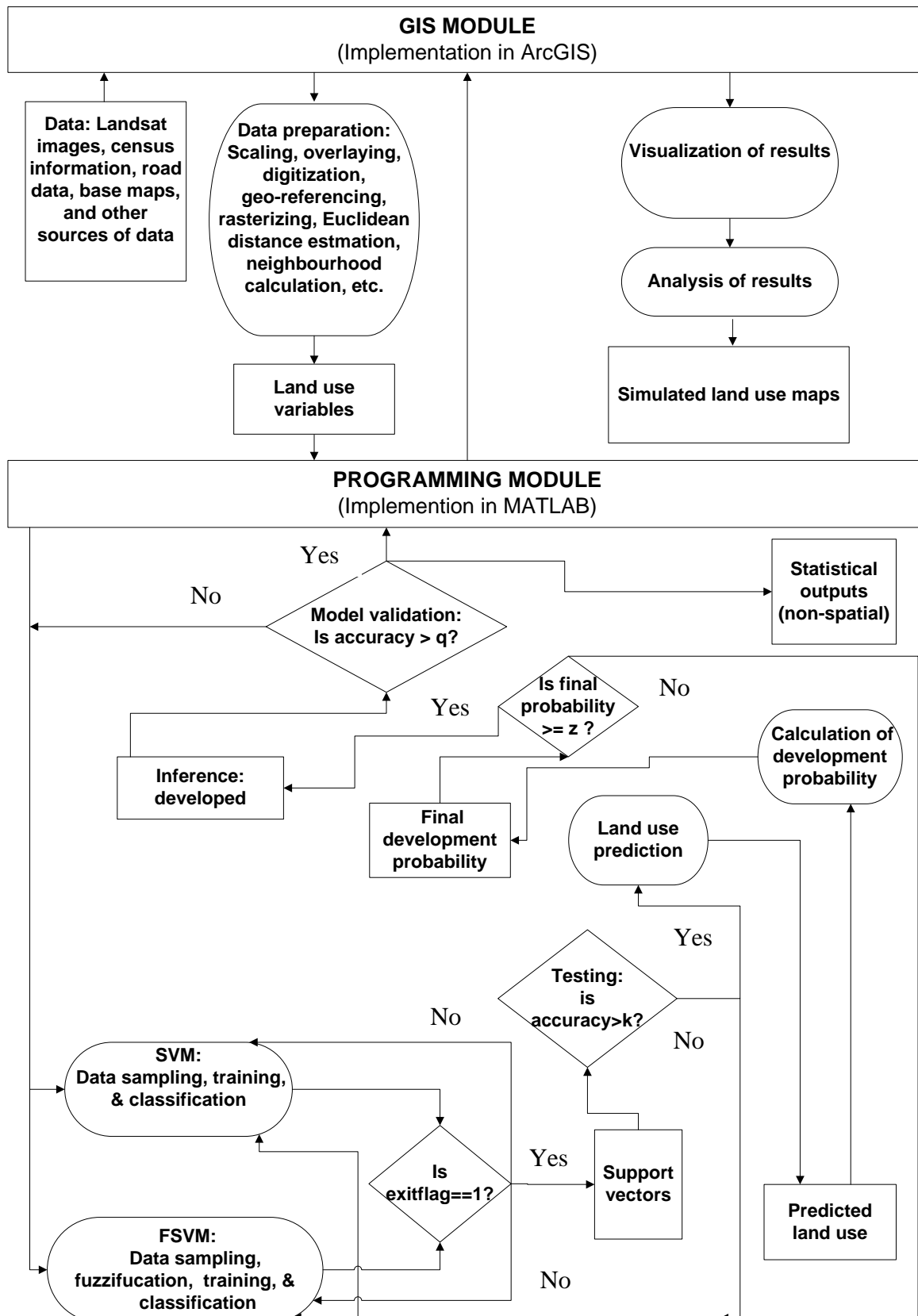


Figure 7.1 Flowchart for SVM and FSVM models



### 7.2.3 Non-CA modelling.

The SVM-based land use change modelling work by Xie (2006) was a classification experiment or non-CA experiment. It is important to implement the non-CA technique in this chapter, so that its results can be compared with that of the CA technique.

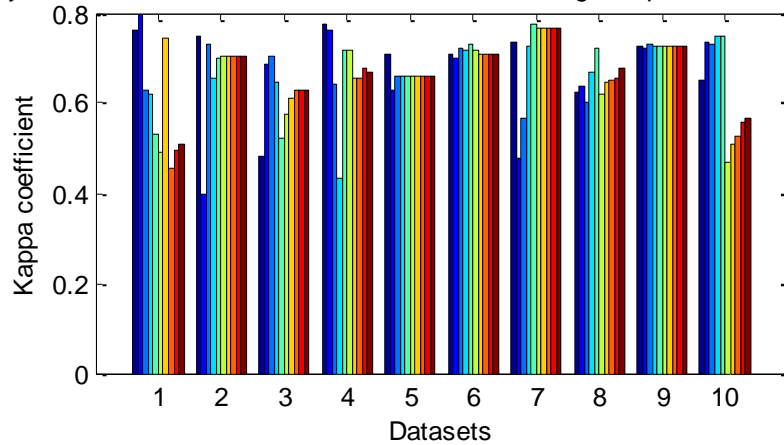
Numerical examples using SVM models are given in section 3.5. Apart from the polynomial and RBF kernels, the linear kernel does not have parameter/s that can be optimised. The results of the polynomial and RBF can be varied with the polynomial degree  $d$  and gamma  $\gamma$  respectively.

The first task in this modelling was to optimise the penalty value  $C$ , polynomial degree  $d$ , and gamma  $\gamma$ . Optimal SVM parameter values could vary for different datasets. These values are usually determined through trial-and-error (Vapnik, 2000; Ivanciuc, 2007). Ten designated  $C$  values,  $\log_{10}(10e0)$ ,  $\log_{10}(10e1)$ ,  $\log_{10}(10e2)$ ,  $\log_{10}(10e3)$ ,  $\log_{10}(10e4)$ ,  $\log_{10}(10e5)$ ,  $\log_{10}(10e6)$ ,  $\log_{10}(10e7)$ ,  $\log_{10}(10e8)$ , and  $\log_{10}(10e9)$ , were used to perform a  $k$ -fold cross-validation (where  $k=10$ ) just as in the case of the LR and ANN models. The training data were split into 10 equal datasets. Nine datasets out of the 10 datasets were put together to train the model, while the remaining one dataset was used to validate the model. The process was repeated until all the 10 datasets were used as both training and test sets.

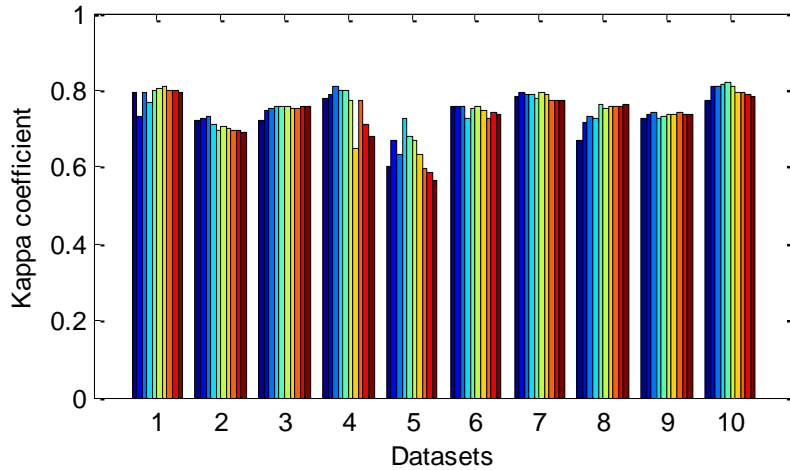
The designated values for the determination of an optimal values for  $\gamma$  and  $d$  were 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10. Before these designated values were adopted, trial experiments were run to determine the most probable range of values of  $C$ ,  $\gamma$ , and  $d$  that can be optimised.

The cross-validation results for  $C$  with respect to the RBF, polynomial, and linear kernels are given in Figures 7.2-7.5. The cross-validation results for determining optimal values for  $\gamma$  and  $d$  are depicted in Figure 7.6.

Polynomial kernel: cross-validation result for selecting an optimal C value for 1963-1978



RBF kernel: cross-validation result for selecting an optimal C value for 1963-1978



Linear kernel: cross-validation result for selecting an optimal C value for 1963-1978

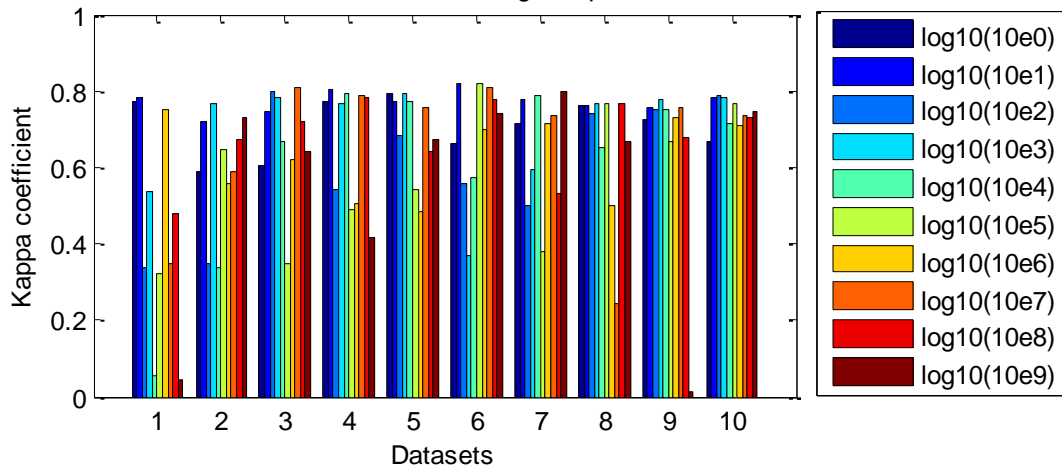
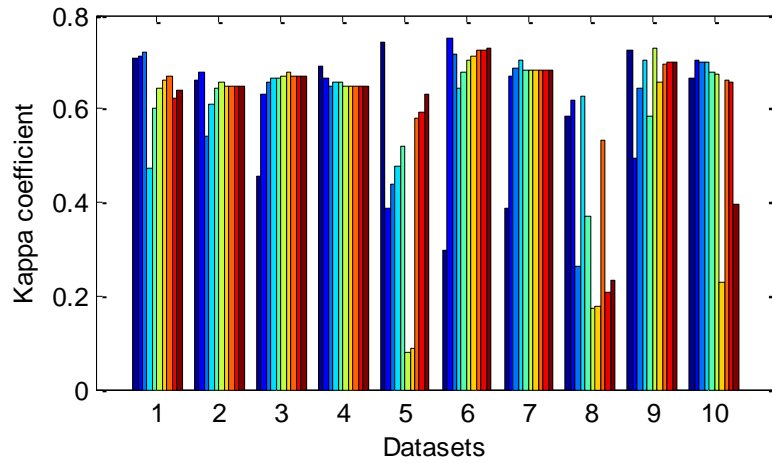
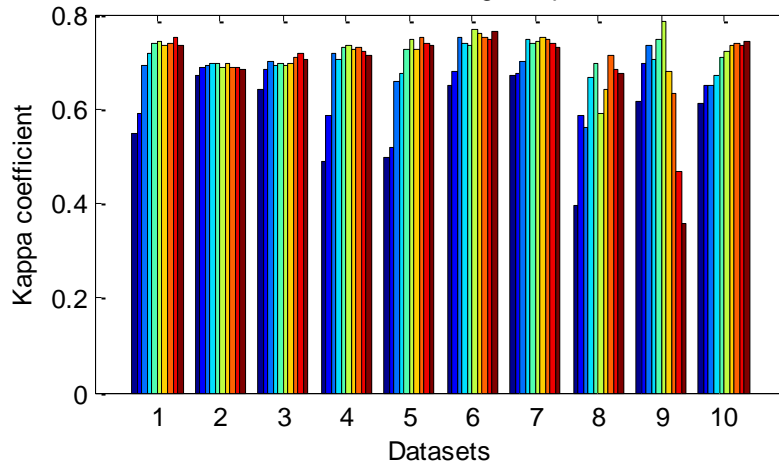


Figure 7.2 SVM-based non-CA: polynomial, RBF, and linear kernels cross-validation results for 1963-1978

Polynomial kernel: cross-validation result for selecting an optimal C value for 1978-1984



RBF kernel: cross-validation result for selecting an optimal C value for 1978-1984



Linear kernel: cross-validation result for selecting an optimal C value for 1978-1984

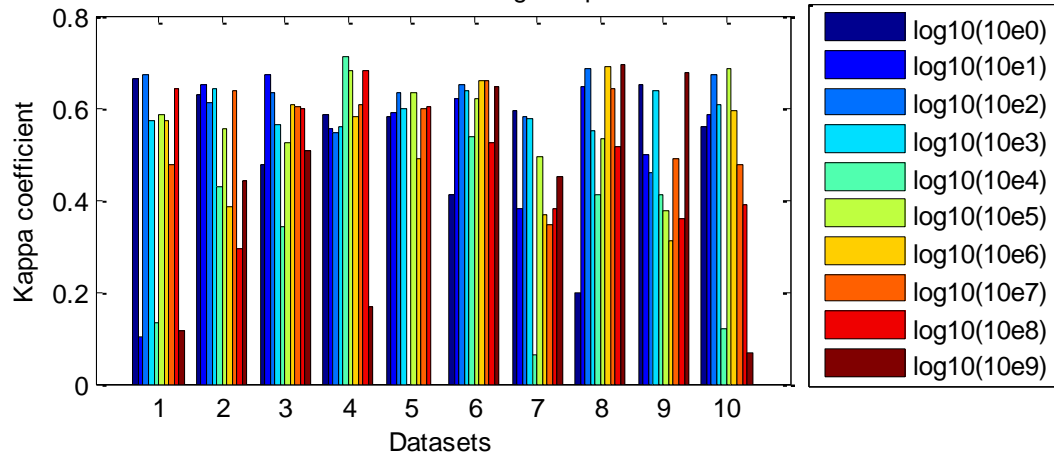
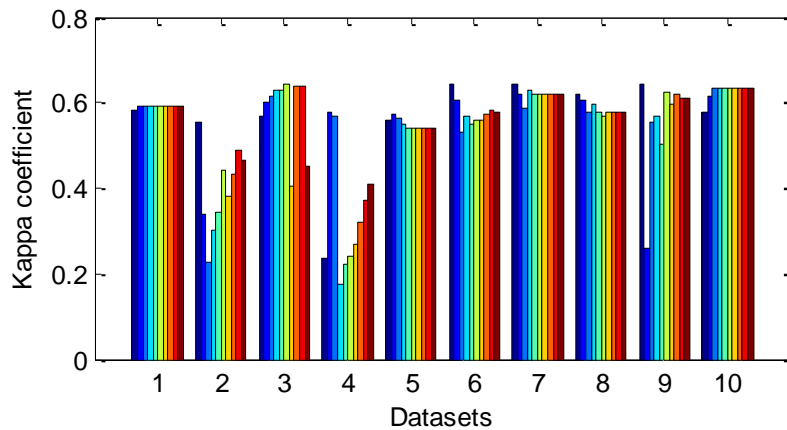
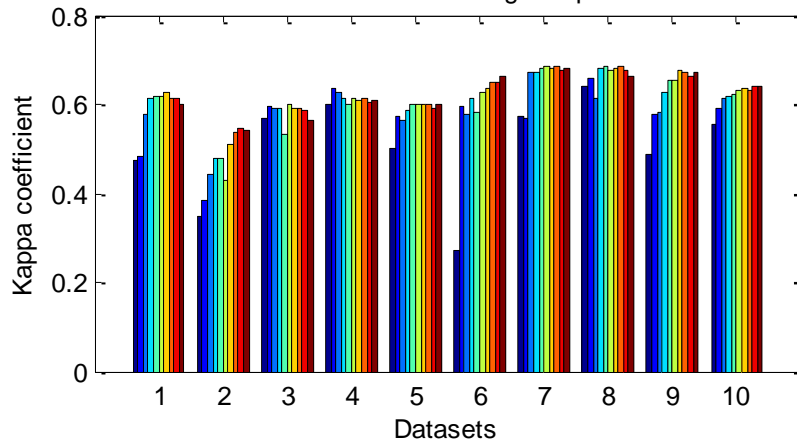


Figure 7.3 *SVM-based non-CA: polynomial, RBF, and linear kernels cross-validation results for 1978-1984*

Polynomial kernel: cross-validation result for selecting an optimal C value for 1984-2000



RBF kernel: cross-validation result for selecting an optimal C value for 1984-2000



Linear kernel: cross-validation result for selecting an optimal C value for 1984-2000

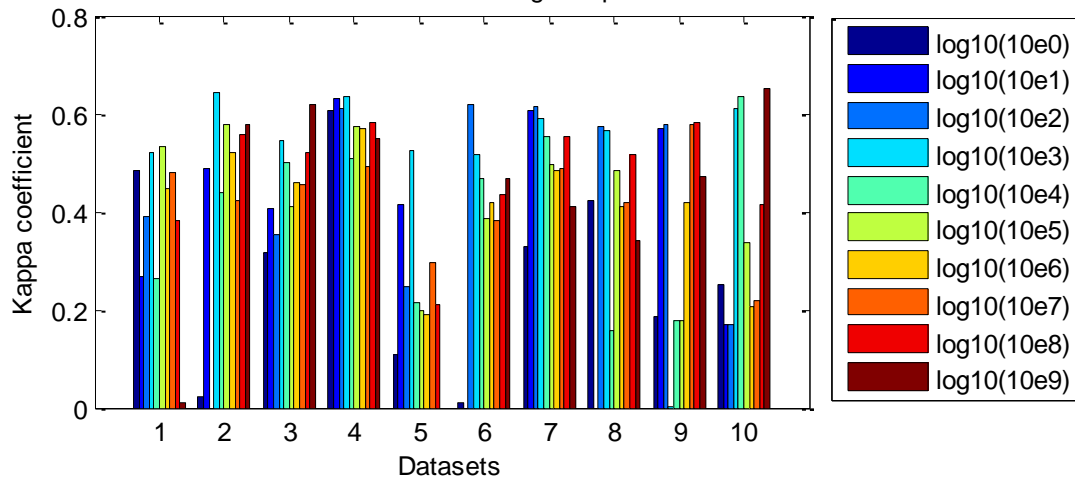


Figure 7.4 *SVM-based non-CA: polynomial, RBF, and linear kernels cross-validation results for 1984-2000*

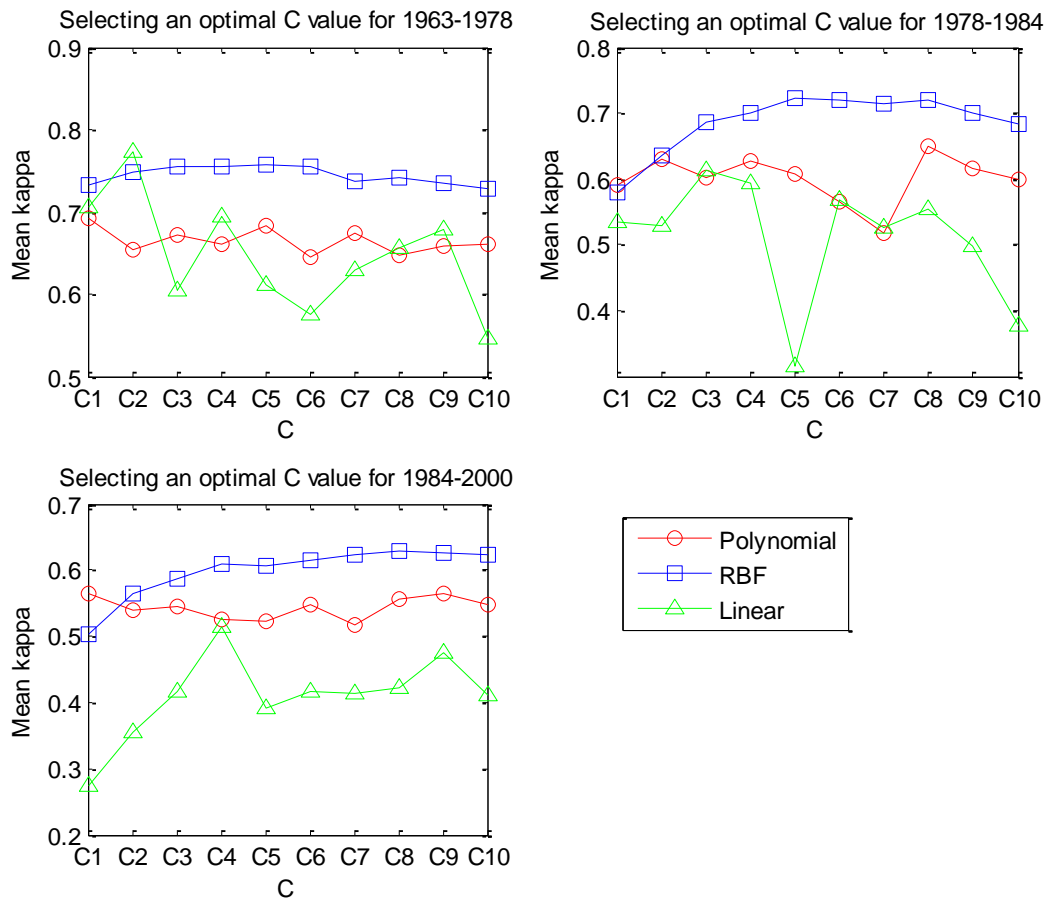


Figure 7.5 *SVM-based non-CA: selecting an optimal C value for periods 1963-1978, 1978-1984, and 1984-2000* ( $C1=\log_{10}(10e0)$ ,  $C2=\log_{10}(10e1)$ ,  $C3=\log_{10}(10e2)$ ,  $C4=\log_{10}(10e3)$ ,  $C5=\log_{10}(10e4)$ ,  $C6=\log_{10}(10e5)$ ,  $C7=\log_{10}(10e6)$ ,  $C8=\log_{10}(10e7)$ ,  $C9=\log_{10}(10e8)$ , and  $C10=\log_{10}(10e9)$ )

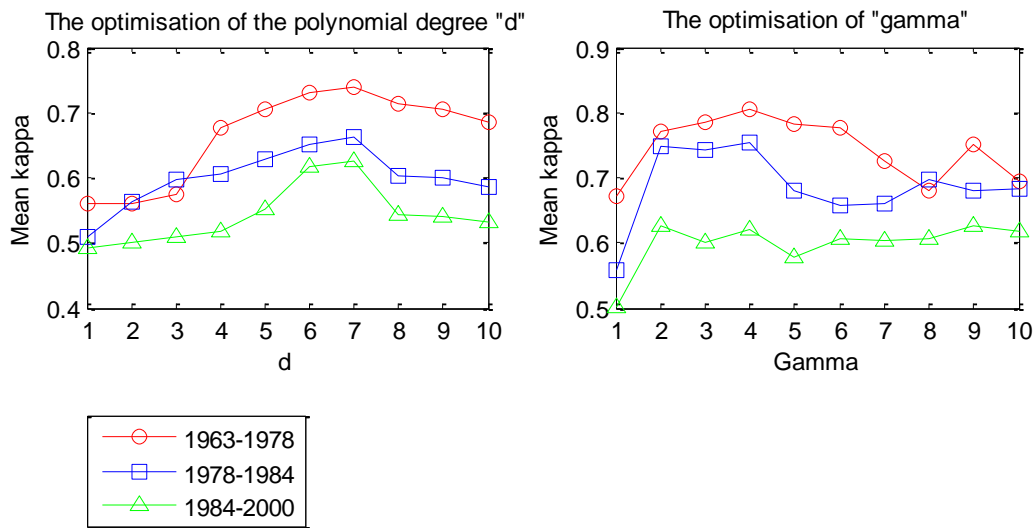


Figure 7.6 *SVM-based non-CA: cross-validation results for obtaining optimal values for  $d$  and  $\gamma$*

From Figure 7.5, 1963-1978, the optimal  $C$  values using polynomial, RBF, and linear kernels were  $\log_{10}(10e0)$ ,  $\log_{10}(10e5)$ , and  $\log_{10}(10e1)$  respectively. For 1978-1984, the optimal  $C$  values for the polynomial, RBF, and linear kernels were  $\log_{10}(10e7)$ ,  $\log_{10}(10e4)$ , and  $\log_{10}(10e2)$  respectively. For 1984-2000, the optimal  $C$  values for polynomial, RBF, and linear kernels were  $\log_{10}(10e0)$ ,  $\log_{10}(10e7)$ , and  $\log_{10}(10e3)$  respectively. From Figure 7.6, the optimal values of  $d$  was found to be 7 for the three periods 1963-1978, 1978-1984, and 1984-2000; while the optimal value for  $\gamma$  was found to be 4 for periods 1963-1978 and 1978-1984. The optimal gamma  $\gamma$  value for period 1984-2000 was 2.

The computed mean kappa statistic from Figure 7.5, given in Figure 7.7, shows the RBF yielded the highest overall mean kappa statistic from all the three predicted periods while the linear kernel yielded the lowest.

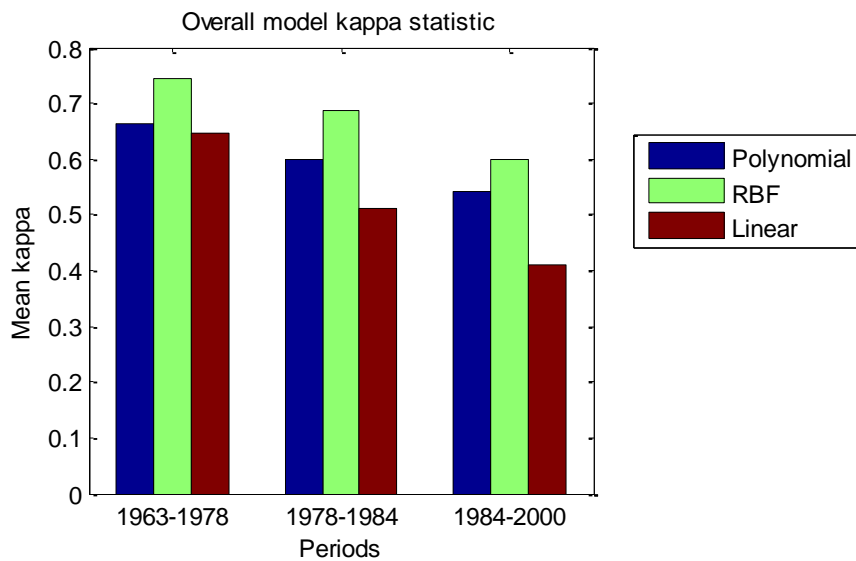


Figure 7.7 Overall model accuracy for periods 1963-1978, 1978-1984, and 1984-2000

The optimal values of  $C$ ,  $\gamma$ , and  $d$  that were obtained using the  $k$ -fold cross-validation were used to predict the land use maps for 1978, 1984, and 2000 given in Figure 7.8. The calculated kappa coefficients from the confusion matrices given in Tables 7.1-7.9 for the predicted maps using the polynomial, RBF, and the linear kernels for periods 1963-1978, 1978-1984, and 1984-2000 were: 0.5002, 0.5326, and 0.4756; 0.5436, 0.5789, and 0.5209; and, 0.4802, 0.5171, and 0.3545, respectively.

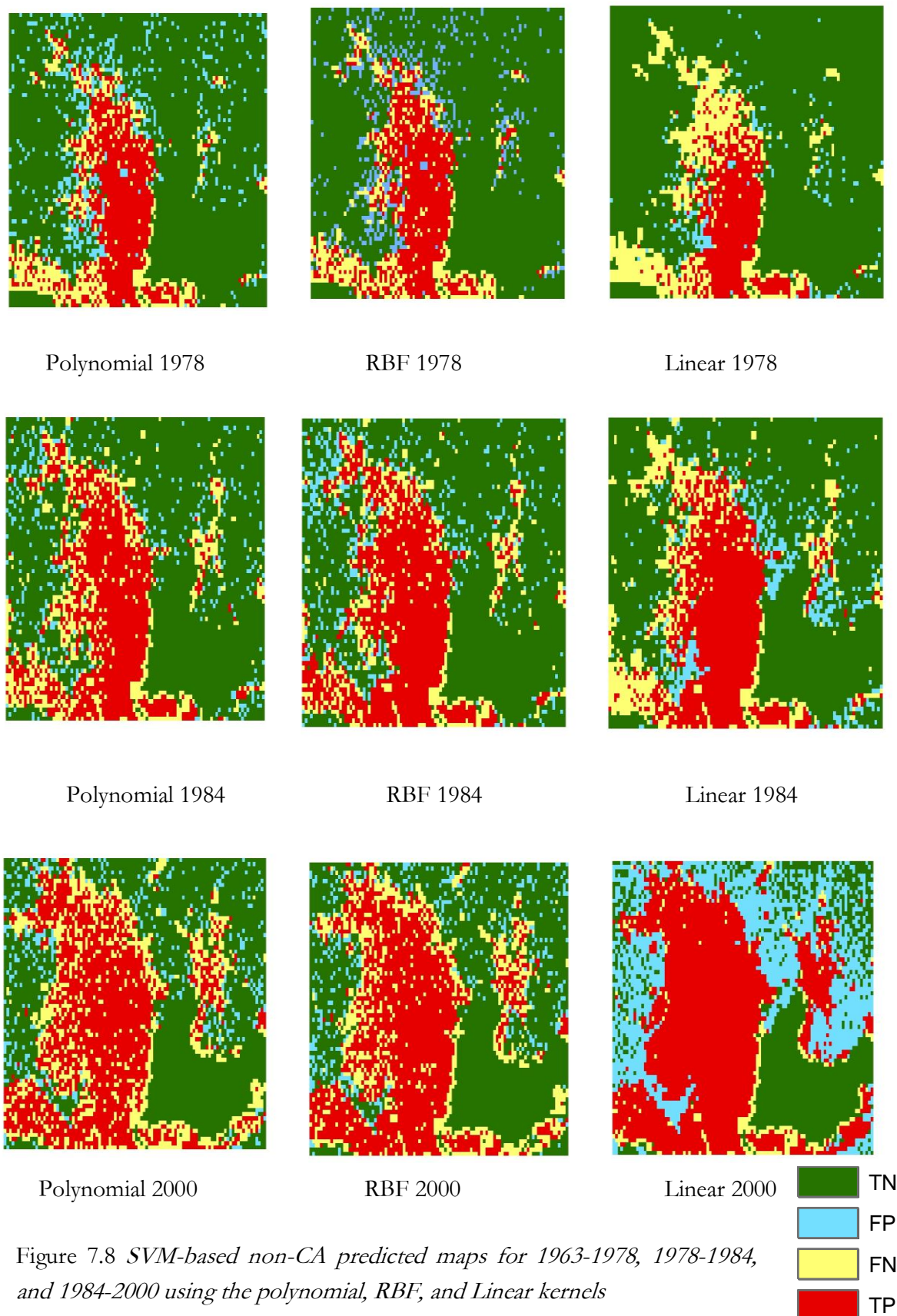


Figure 7.8 *SVM-based non-CA predicted maps for 1963-1978, 1978-1984, and 1984-2000 using the polynomial, RBF, and Linear kernels*



Table 7.1 *SVM-based non-CA: confusion matrix for polynomial 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1095	616
Undeveloped	697	4592

Table 7.2 *SVM-based non-CA: confusion matrix for RBF 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1114	536
Undeveloped	678	4672

Table 7.3 *SVM-based non-CA: confusion matrix for linear 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	809	220
Undeveloped	983	4988

Table 7.4 *SVM-based non-CA: confusion matrix for polynomial 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	1531	420
Undeveloped	977	4072

Table 7.5 *SVM-based non-CA: confusion matrix for RBF 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	1721	539
Undeveloped	787	3953

Table 7.6 *SVM-based non-CA: confusion matrix for linear 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	1542	515
Undeveloped	966	3977

Table 7.7 *SVM-based non-CA: confusion matrix for polynomial 1984-2000*

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	2106	484
Undeveloped	1326	3084

Table 7.8 *SVM-based non-CA: confusion matrix for RBF 1984-2000*

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	2237	488
Undeveloped	1195	3080

Table 7.9 SVM-based non-CA: confusion matrix for linear 1984-2000

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	3076	1922
Undeveloped	356	1646

## 7.2.4 CA modelling.

As mentioned in chapter 6, no training of data is done in this section. In the pervious section, the selected  $C$ ,  $d$ , and  $\gamma$  were used to train the SVM model. After training the model (already done in the previous sub-section), the neighbourhood function is used in conjunction with the independent variables for the CA modelling.

Two hundred iterations were run to determine the best predictions for periods 1963-1978, 1978-1984, and 1984-2000. Figures 7.9 shows the plotted mean kappa coefficients for periods 1963-1978, 1978-1984, and 1984-2000; obtained by running the CA model 10 times at each 20 designated iteration thresholds. For 1963-1978, the highest mean kappa coefficients for the polynomial, RBF, and linear kernels were obtained at the 100<sup>th</sup>, 120<sup>th</sup>, and 140<sup>th</sup> iterations respectively. For 1978-1984, the highest mean kappa coefficients for the polynomial, RBF, and linear kernels were obtained at the 90<sup>th</sup>, 70<sup>th</sup>, and 120<sup>th</sup> iterations respectively. For 1984-2000, the highest mean kappa coefficients for the polynomial, RBF, and linear kernels were obtained at the 90<sup>th</sup>, 140<sup>th</sup>, and 120<sup>th</sup> iterations respectively.

The derived optimal  $C$  values were used to predict the target maps in 1978, 1984, and 2000 (see Figure 7.10). The calculated kappa coefficients computed from the confusion matrices given in Tables 7.10-7.18 for the predicted maps using the polynomial, RBF, and the linear kernels for periods 1963-1978, 1978-1984, and 1984-2000 were: 0.5870, 0.6000, and 0.5467; 0.7848, 0.7919, and 0.7527; and, 0.7102, 0.7170, and 0.7054, respectively.

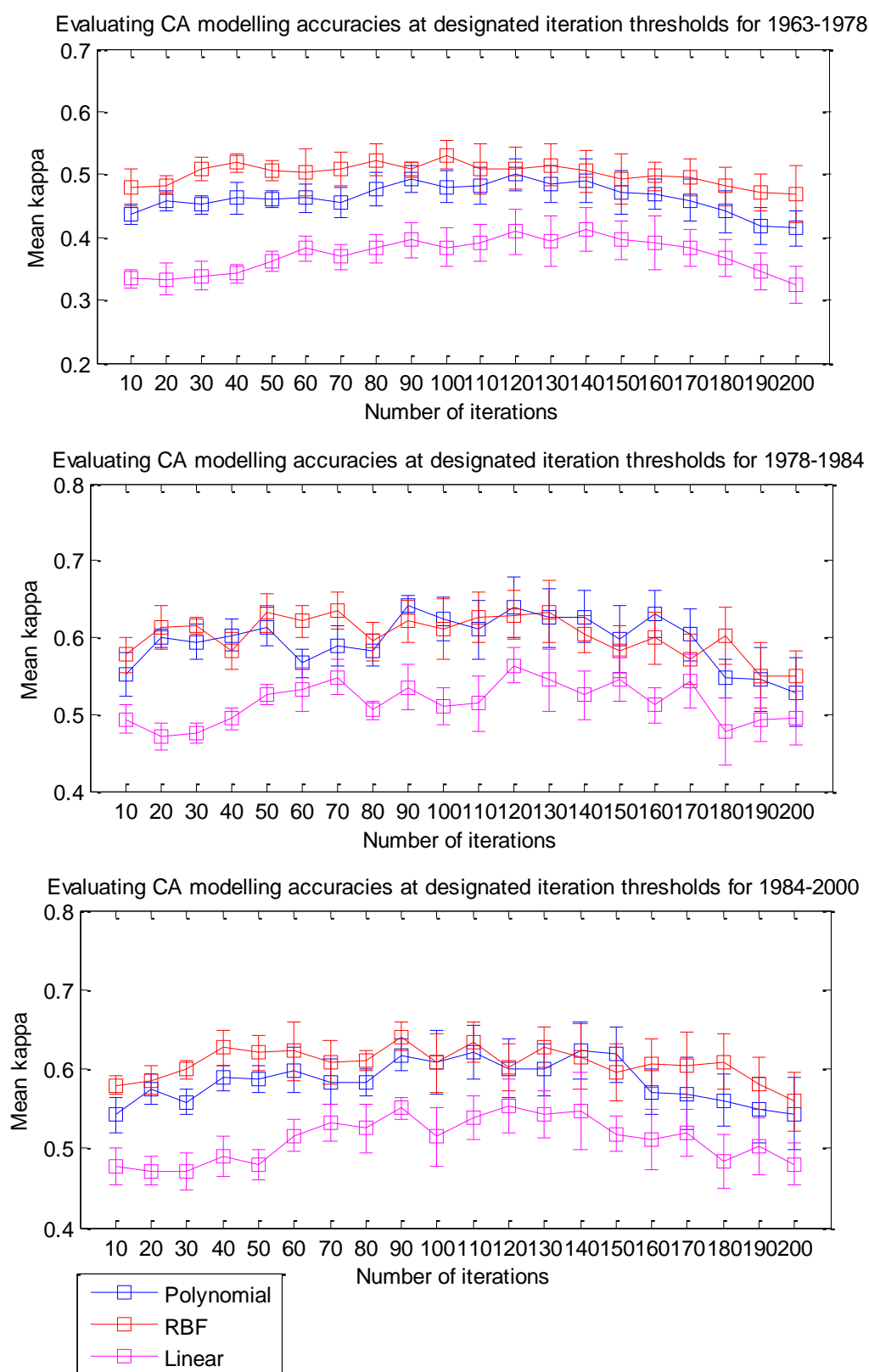


Figure 7.9 SVM-based non-CA: computed overall mean kappa and standard deviations for 200 designated iteration thresholds

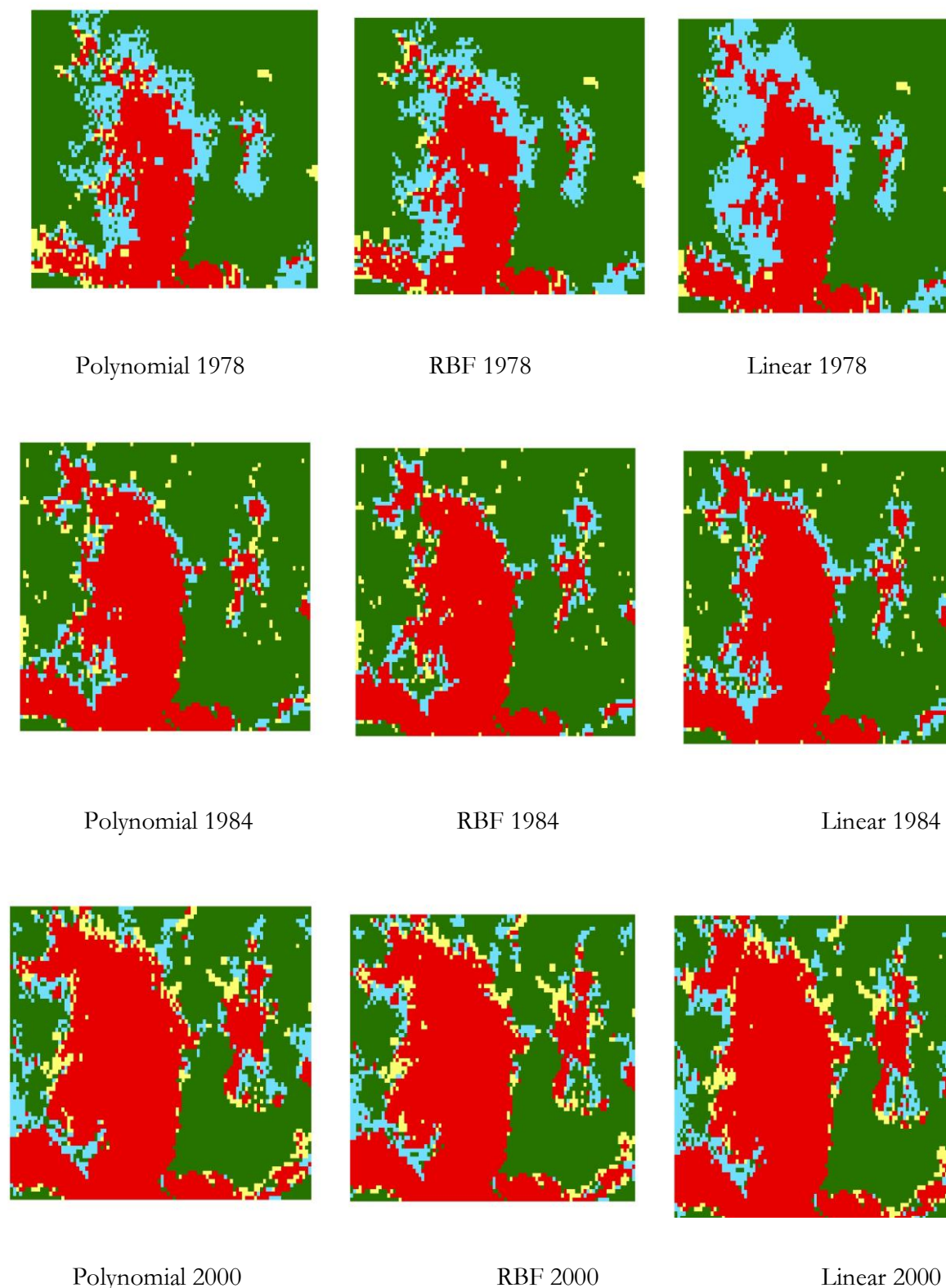


Figure 7.10 SVM-based CA predicted maps for periods 1963-1978, 1978-1984, and 1984-2000

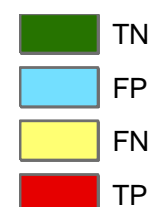


Table 7.10 *SVM-based CA: confusion matrix for polynomial 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1556	1024
Undeveloped	236	4184

Table 7.11 *SVM-based CA: confusion matrix for RBF 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1598	1037
Undeveloped	194	4171

Table 7.12 *SVM-based CA: confusion matrix for linear 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1690	1392
Undeveloped	102	3816

Table 7.13 *SVM-based CA: confusion matrix for polynomial 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	2250	446
Undeveloped	258	4046

Table 7.14 *SVM-based CA: confusion matrix for RBF 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	2263	436
Undeveloped	245	4056

Table 7.15 *SVM-based CA: confusion matrix for linear 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	2316	635
Undeveloped	192	3857

Table 7.16 *SVM-based CA: confusion matrix for polynomial 1984-2000*

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	3020	603
Undeveloped	412	2965

Table 7.17 *SVM-based CA: confusion matrix for RBF 1984-2000*

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	3004	563
Undeveloped	428	3005

Table 7.18 SVM-based CA: confusion matrix for linear 1984-2000

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	3019	619
Undeveloped	413	2949

In order to assess the simulation, the model accuracy was computed by repeatedly running two hundred iterations ten times. The model accuracy was computed by finding the average of all the resulting kappa coefficients. The computed model accuracies for the polynomial, RBF, and linear kernels are presented in Figure 7.11. The RBF kernel yielded the highest computed mean kappa while the linear kernel yielded the lowest.

The ROC plot was used to assess the strength of the SVM-based CA model. The ROC plots for periods 1963-1978, 1978-1984, and 1984-2000 given in Figures 7.12-7.14 were derived by running 200 iterations. The calculated AUC, 1963-1978, for polynomial, RBF, and linear kernels were:  $0.7629 \pm 0.0270$ ,  $0.7863 \pm 0.0273$ ,  $0.7464 \pm 0.0287$  respectively; for 1978-1984, were  $0.8003 \pm 0.0300$ ,  $0.8139 \pm 0.0248$ ,  $0.7673 \pm 0.0214$  respectively; and for periods 1984-2000,  $0.7804 \pm 0.0316$ ,  $0.7939 \pm 0.0290$ ,  $0.7714 \pm 0.0255$  respectively. All the results indicate that the RBF kernel yielded the highest AUC index for the three periods 1963-1978, 1978-1984, and 1984-2000, while the linear kernel yielded the lowest.



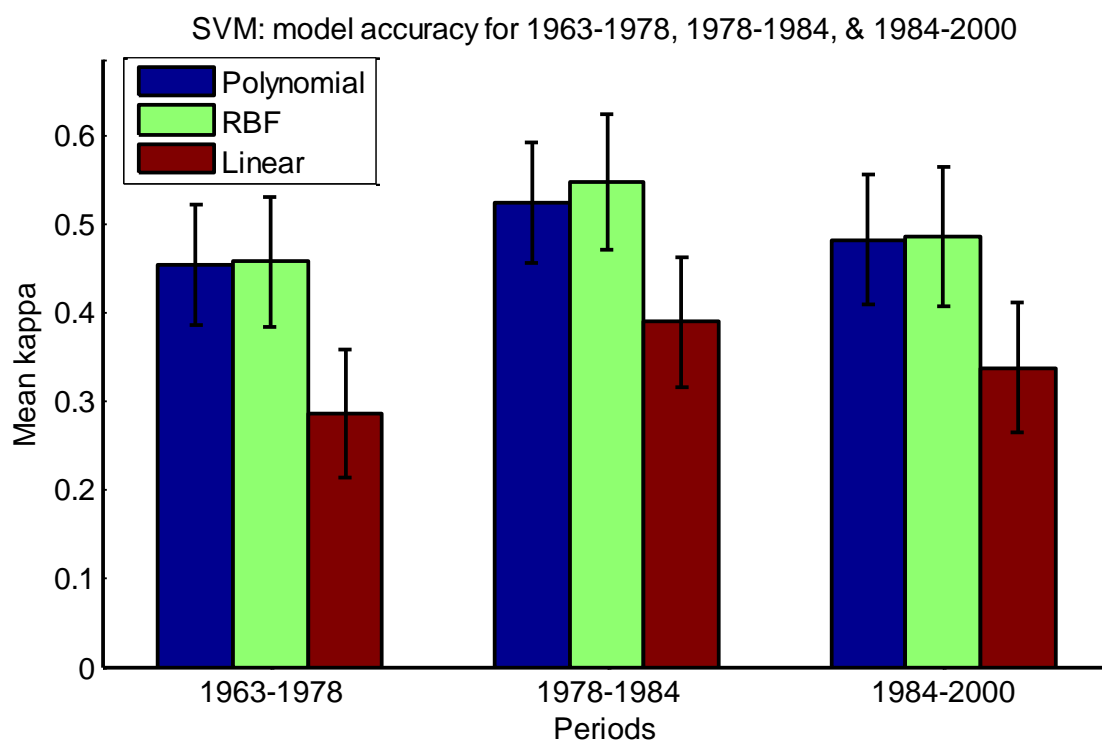


Figure 7.11 SVM based CA: model accuracy for periods 1963-1978, 1978-1984, and 1984-2000

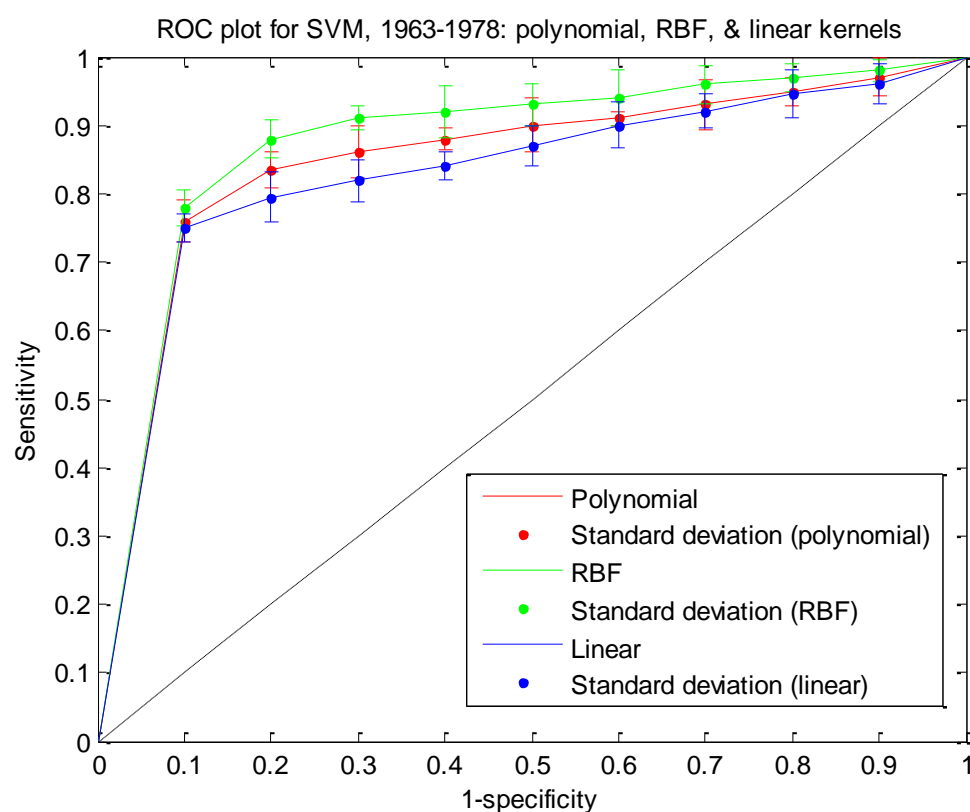


Figure 7.12 SVM based CA: ROC plot for 1963-1978

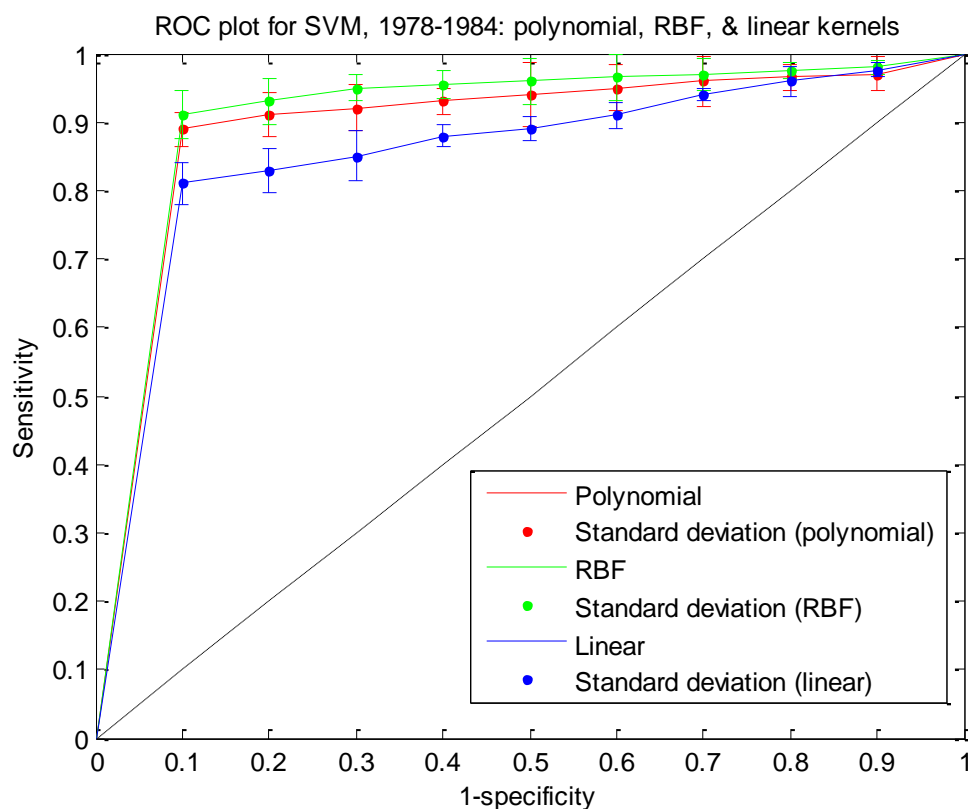


Figure 7.13 SVM based CA: ROC plot for 1978-1984

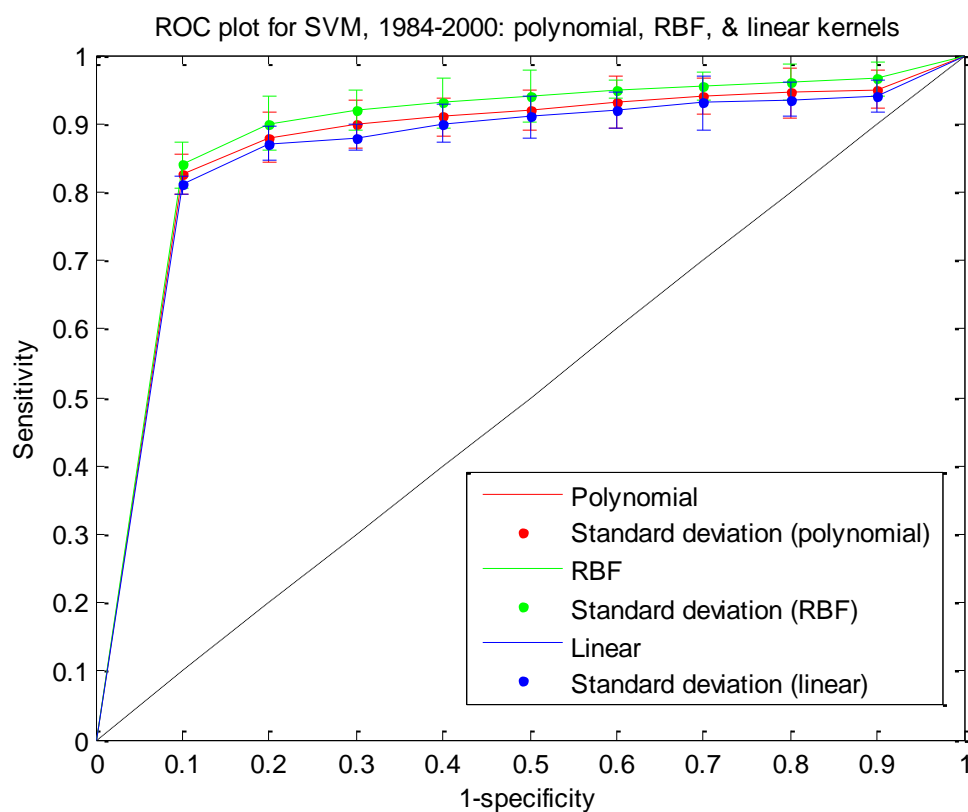


Figure 7.14 SVM based CA: ROC plot for 1984-2000

### 7.3 FSVM-based modelling

#### 7.3.1 FSVM-based CA calibration.

From section 3.4, unlike in the standard SVM model case explained in section 3.3, the optimisation problems given in equations 3.24 and 3.25 are subject to:  $\sum_{i=1}^n \alpha_i y_i = 0$ , and,

$$0 \leq \alpha_i \leq s_i C, \quad \text{for } i = 1, \dots, n, \quad \text{in order to obtain the solutions of}$$

$$f(x) = \text{sign} \left[ \sum_{i=1}^n y_i \alpha_i^0 (x_i \cdot x) + b^0 \right] \text{ and } f(x) = \text{sign} \left[ \sum_{i=1}^n y_i \alpha_i^0 K(x_i, x) + b^0 \right] \text{ (Lin \& Wang, 2002).}$$

In the same manner described in section 7.2.1 for the SVM-based CA case, FSVM outputs furnished by  $f(x)$  given in equations 3.24 and 3.25 can be mapped into probabilities using the sigmoid function (Platt, 1999). Note that, the remaining procedures are same as the SVM (see section 7.2.1).

The development probability is furnished by equation 7.1 and the final development probability by equation 7.2. The land use transition of undeveloped cells to develop is the same as equation 6.5.

#### 7.3.2 Data and methodology.

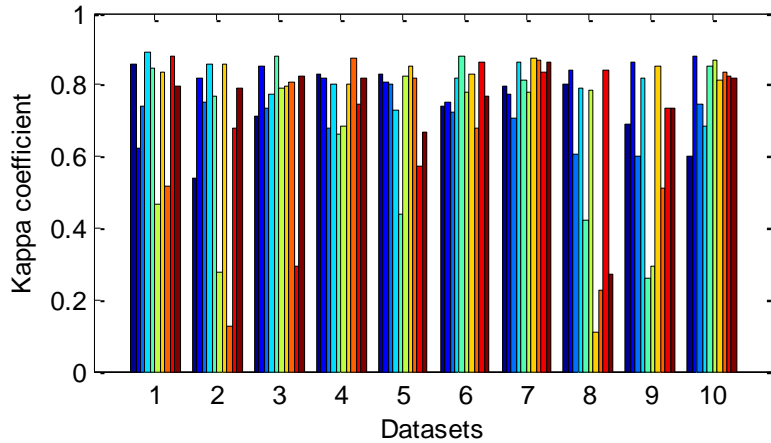
The procedures for the SVM model case described in section 7.2.2 and the FSVM model are basically the same (see Figure 7.1). The FSVM model only differs from the SVM-based model by the incorporation of the fuzzy membership function  $s_i$  into the optimisation equations given in equations 3.4 and 3.5. The FSVM model accuracy can be enhanced by adjusting the constant  $\delta$  (see equation 3.18). The remaining procedures are the same as the SVM-based model illustrated in section 7.2.2.

### 7.3.3 Non-CA modelling.

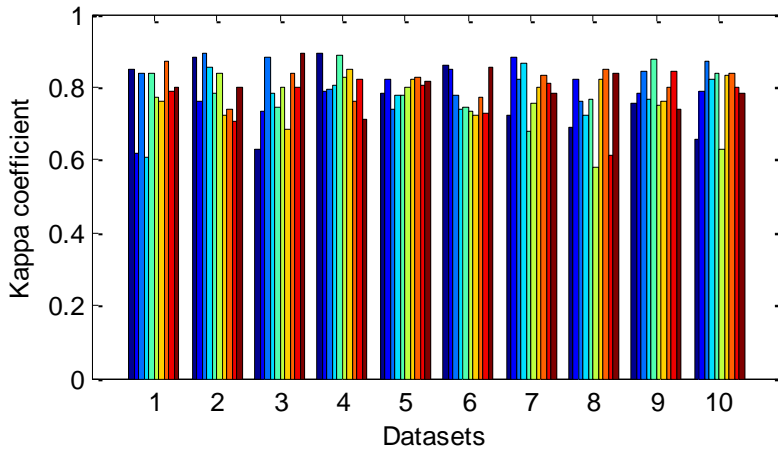
The same data and  $k$ -fold cross-validation procedure used in the case of the SVM was used in this section. Same as the SVM modelling, the designated values for the determination of an optimal value for  $C$  were  $\log_{10}(10e0)$ ,  $\log_{10}(10e1)$ ,  $\log_{10}(10e2)$ ,  $\log_{10}(10e3)$ ,  $\log_{10}(10e4)$ ,  $\log_{10}(10e5)$ ,  $\log_{10}(10e6)$ ,  $\log_{10}(10e7)$ ,  $\log_{10}(10e8)$ , and  $\log_{10}(10e9)$ ; and  $\gamma$  and  $d$  were 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

The cross-validation results for  $C$  with respect to the RBF, polynomial, and linear kernels are given in Figures 7.15-7.18. The cross-validation results for the determining optimal values for  $\gamma$  and  $d$  are depicted in Figure 7.19.

Polynomial kernel: cross-validation result for selecting an optimal C value for 1963-1978



RBF kernel: cross-validation result for selecting an optimal C value for 1963-1978



Linear kernel: cross-validation result for selecting an optimal C value for 1963-1978

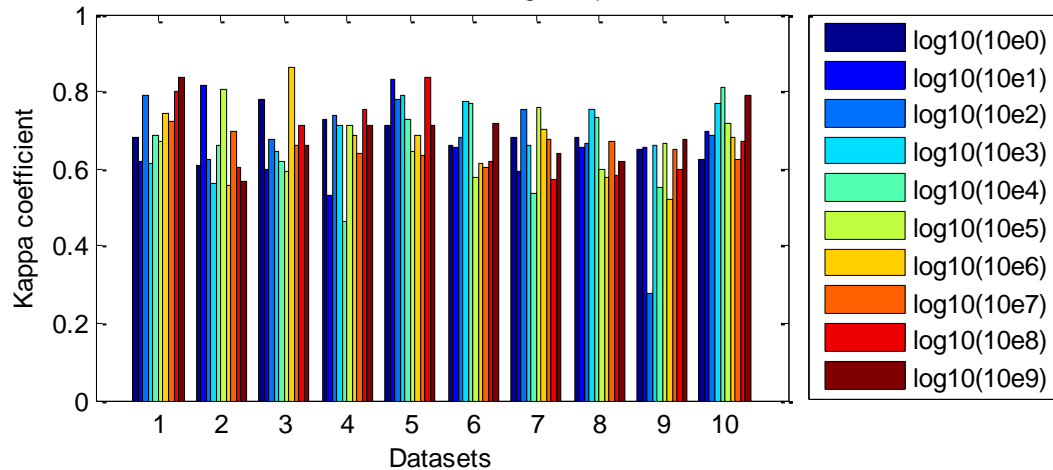
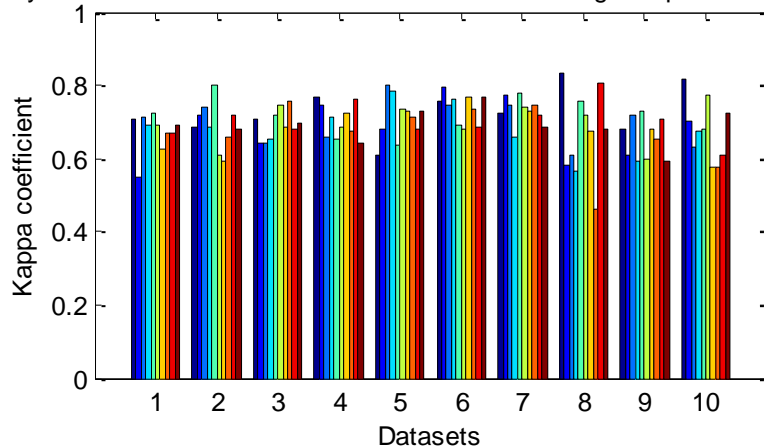
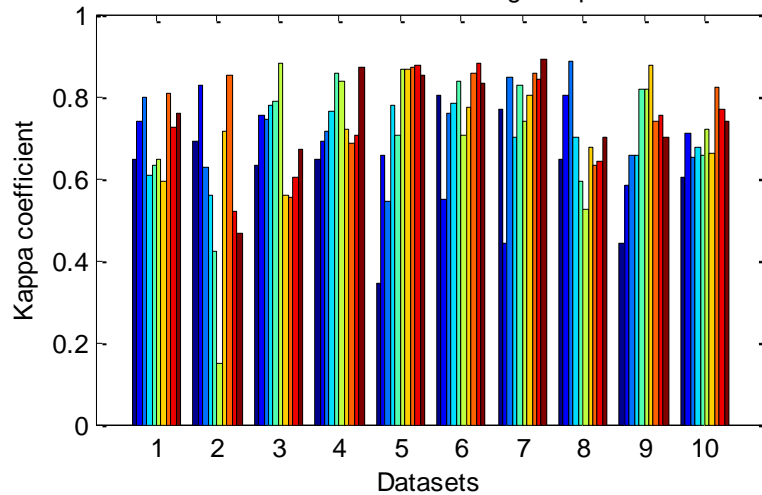


Figure 7.15 *FSVM-based non-CA polynomial, RBF, and linear kernels cross-validation results for 1963-1978*

Polynomial kernel: cross-validation result for selecting an optimal C value for 1978-1984



RBF kernel: cross-validation result for selecting an optimal C value for 1978-1984



Linear kernel: cross-validation result for selecting an optimal C value for 1978-1984

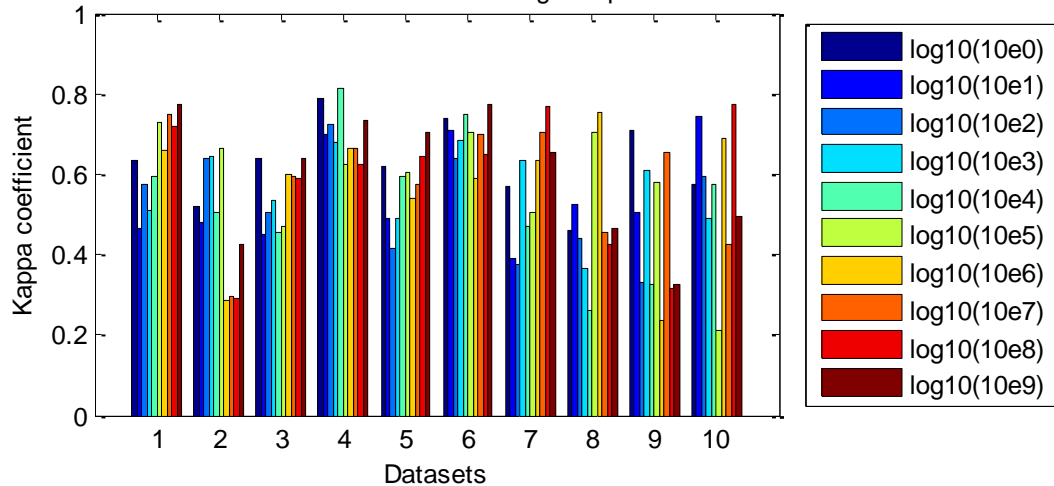
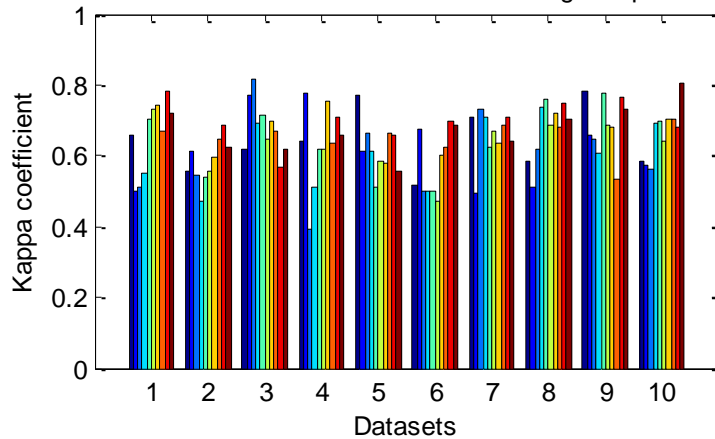
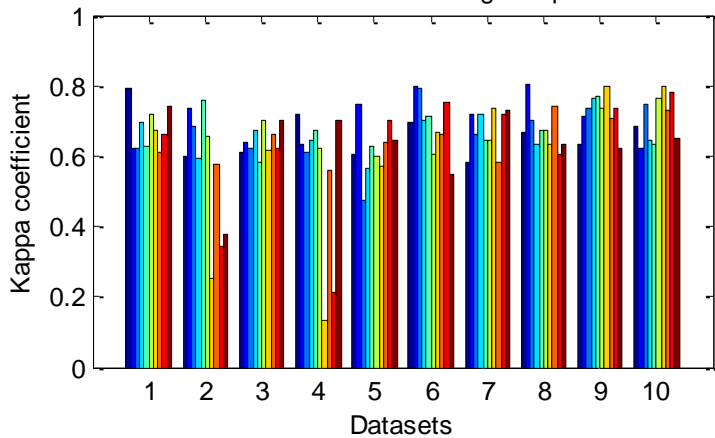


Figure 7.16 *FSVM-based non-CA polynomial, RBF, and linear kernels cross-validation results for 1978-1984*

Polynomial kernel: cross-validation result for selecting an optimal C value for 1984-2000



RBF kernel: cross-validation result for selecting an optimal C value for 1984-2000



Linear kernel: cross-validation result for selecting an optimal C value for 1984-2000

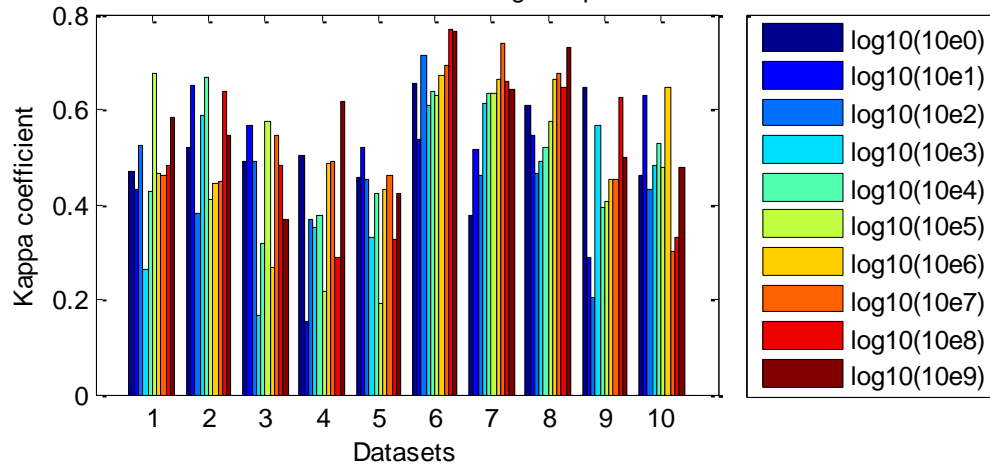


Figure 7.17 *FSVM-based non-CA polynomial, RBF, and linear kernels cross-validation results for 1984-2000*

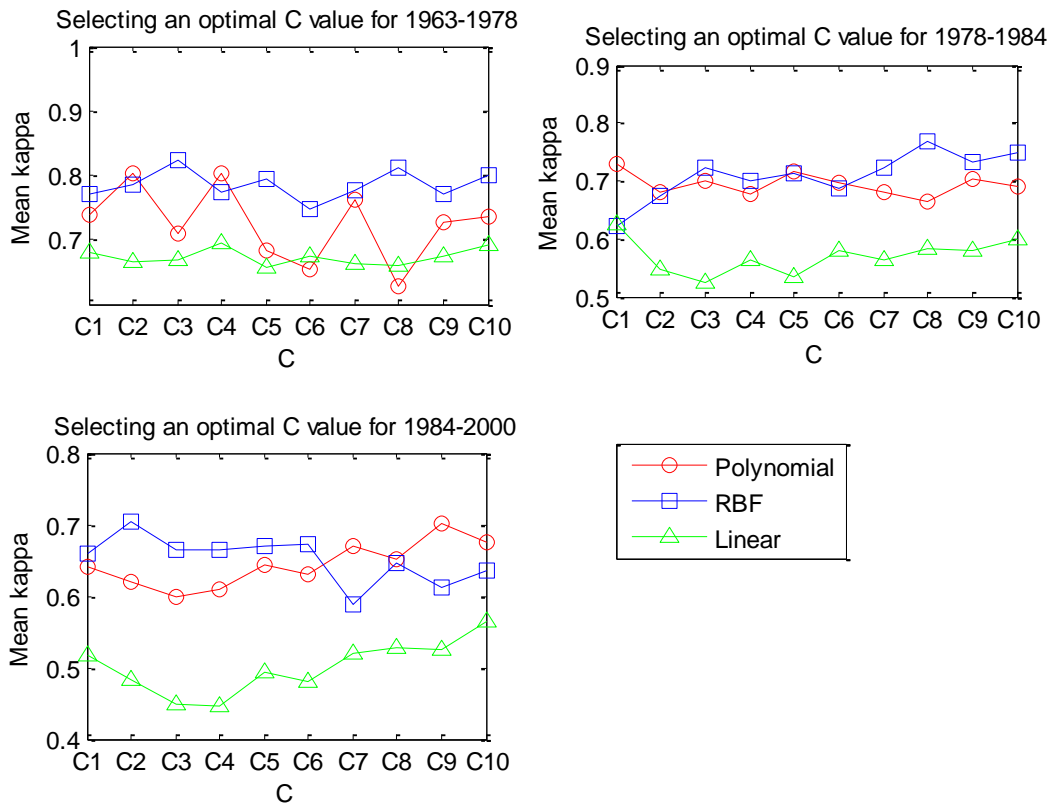


Figure 7.18 *FSVM-based non-CA: selecting an optimal C value for periods 1963-1978, 1978-1984, and 1984-2000*

From Figure 7.18, 1963-1978, the optimal  $C$  values using polynomial, RBF, and linear kernels were  $\log_{10}(10e1)$ ,  $\log_{10}(10e2)$ , and  $\log_{10}(10e9)$  respectively. For 1978-1984, the optimal  $C$  values for the polynomial, RBF, and linear kernels were  $\log_{10}(10e0)$ ,  $\log_{10}(10e7)$ , and  $\log_{10}(10e0)$  respectively. For 1984-2000, the optimal  $C$  values for polynomial, RBF, and linear kernels were  $\log_{10}(10e8)$ ,  $\log_{10}(10e1)$ , and  $\log_{10}(10e9)$  respectively. From Figure 7.19, the optimal value of  $d$  was 7 for periods 1963-1978 and 1978-1984; while period 1984-2000 was 8. The optimal value for  $\gamma$  for periods 1963-1978 and 1978-1984 was 4; while 1984-2000 was 2.



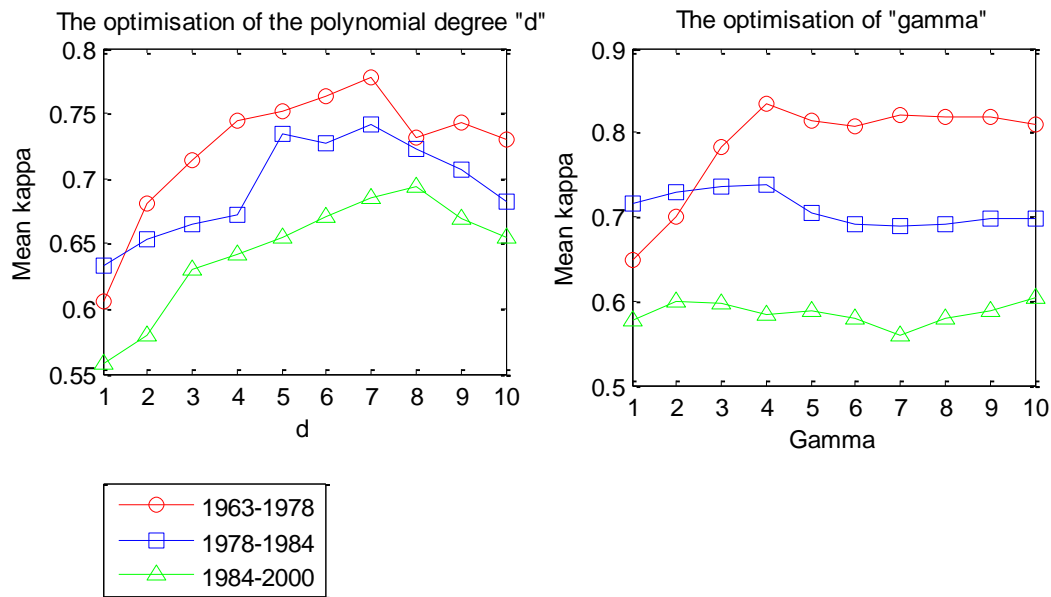


Figure 7.19 *FSVM-based non-CA: cross-validation results for obtaining an optimal value for d and gamma*

From Figure 7.20, the computed overall model accuracies for periods 1963-1978, 1978-1984, and 1984-2000 indicate that the RBF yielded the highest overall mean kappa statistic for all the three predicted periods while the linear kernel yielded the lowest.

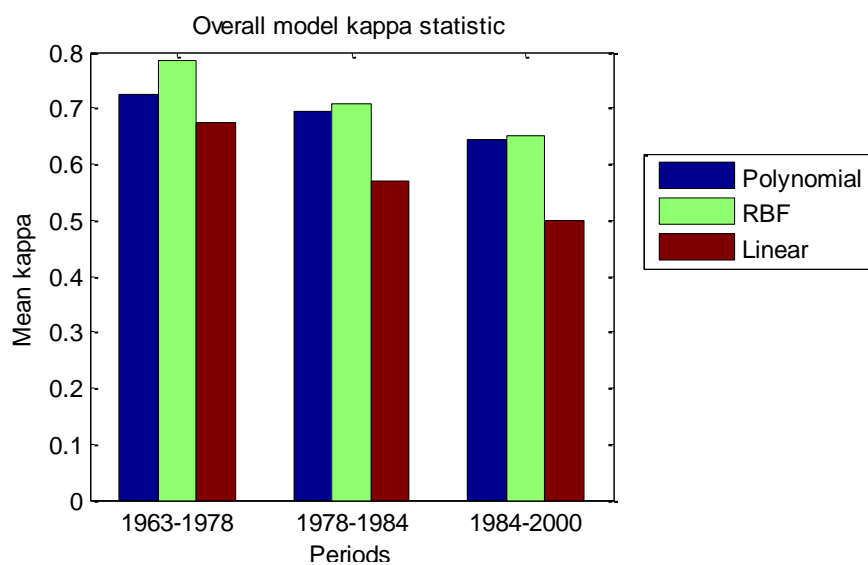


Figure 7.20 *FSVM-based non-CA: overall model accuracies for periods 1963-1978, 1978-1984, and 1984-2000*

The optimal values of  $C$ ,  $\gamma$ , and  $d$  were used to predict the land use maps for 1978, 1984, and 2000 given in Figure 7.21. The calculated kappa coefficients computed from the confusion matrices given in Tables 7.19-7.27 using the polynomial, RBF, and the linear kernels for periods 1963-1978, 1978-1984, and 1984-2000 were: 0.5339, 0.5559, and 0.5142; 0.5641, 0.6660, and 0.5534; and 0.5293, 0.5307, and 0.4404, respectively.

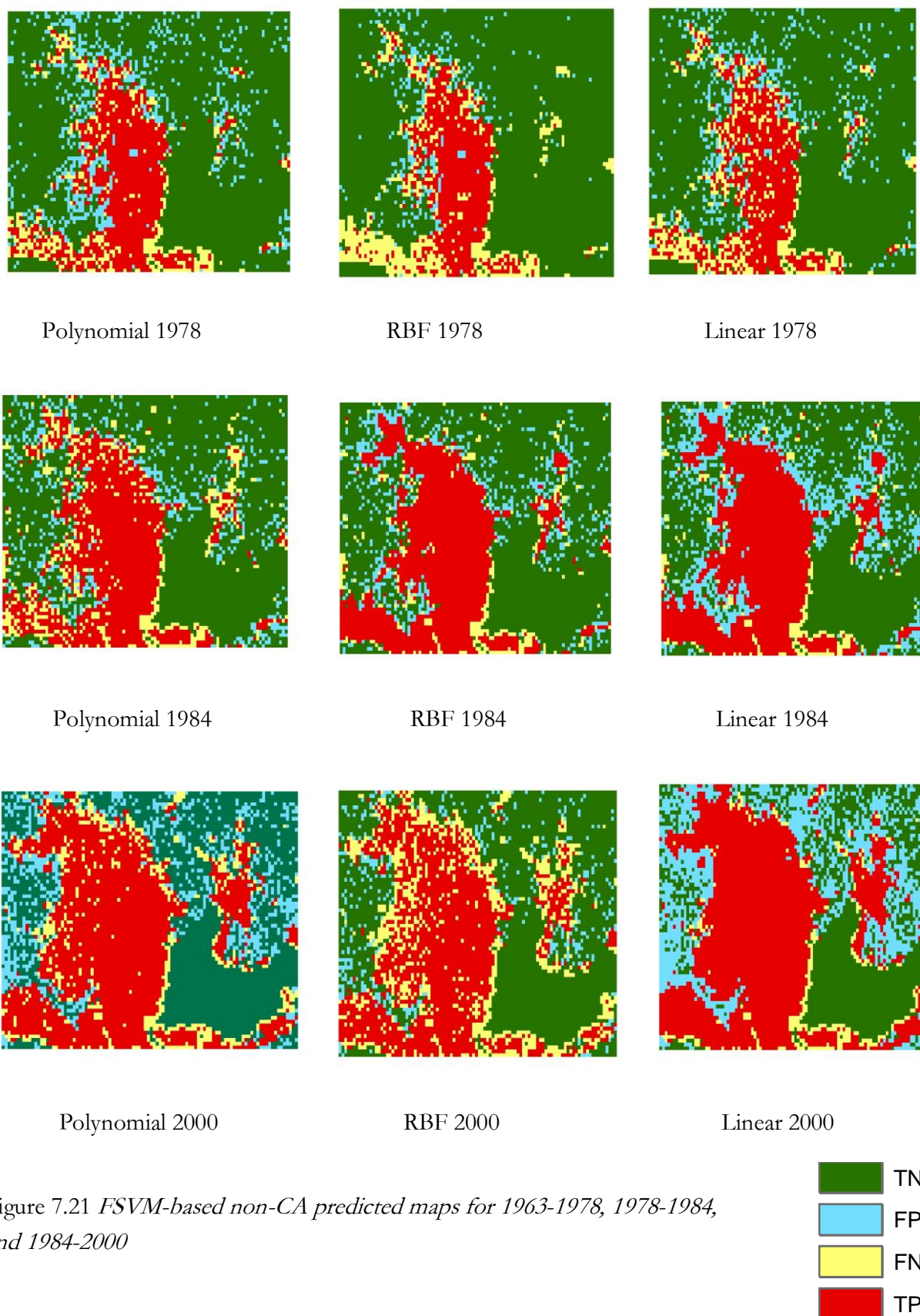


Figure 7.21 FSVM-based non-CA predicted maps for 1963-1978, 1978-1984, and 1984-2000

Table 7.19 *FSVM-based non-CA: confusion matrix for polynomial 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1259	762
Undeveloped	533	4446

Table 7.20 *FSVM-based non-CA: confusion matrix for RBF 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1028	325
Undeveloped	764	4883

Table 7.21 *FSVM-based non-CA: confusion matrix for linear 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1115	600
Undeveloped	677	4608

Table 7.22 *FSVM-based non-CA: confusion matrix for polynomial 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	1785	674
Undeveloped	723	3818

Table 7.23 *FSVM-based non-CA: confusion matrix for RBF 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	2138	740
Undeveloped	370	3752

Table 7.24 *FSVM-based non-CA: confusion matrix for linear 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	2226	1282
Undeveloped	282	3210

Table 7.25 *FSVM-based non-CA: confusion matrix for polynomial 1984-2000*

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	2728	945
Undeveloped	704	2623

Table 7.26 *FSVM-based non-CA: confusion matrix for RBF 1984-2000*

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	2234	437
Undeveloped	1198	3131

Table 7.27 *FSVM-based non-CA: confusion matrix for linear 1984-2000*

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	3093	1633
Undeveloped	339	1935

### 7.3.4 CA modelling.

The FSVM modelling in this section is the same as the SVM already described in section 7.2.4. Same as the SVM two hundred iterations were run to determine the best predictions for periods 1963-1978, 1978-1984, and 1984-2000. Figure 7.22 shows the plotted overall mean kappa coefficients for periods 1963-1978, 1978-1984, and 1984-2000; obtained by running the CA model 10 times at each 20 designated iteration thresholds. For 1963-1978, the highest mean kappa coefficients for the polynomial, RBF, and the linear kernels were obtained at the 130<sup>th</sup>, 110<sup>th</sup>, and 100<sup>th</sup> iterations, respectively. For 1978-1984, the highest mean kappa coefficients for the polynomial, RBF, and the linear kernels were obtained at the 120<sup>th</sup>, 110<sup>th</sup>, and 100<sup>th</sup> iterations, respectively. For 1984-2000, the highest mean kappa coefficients for the polynomial, RBF, and the linear kernels were obtained at the 120<sup>th</sup>, 110<sup>th</sup>, and 130<sup>th</sup> iterations, respectively.

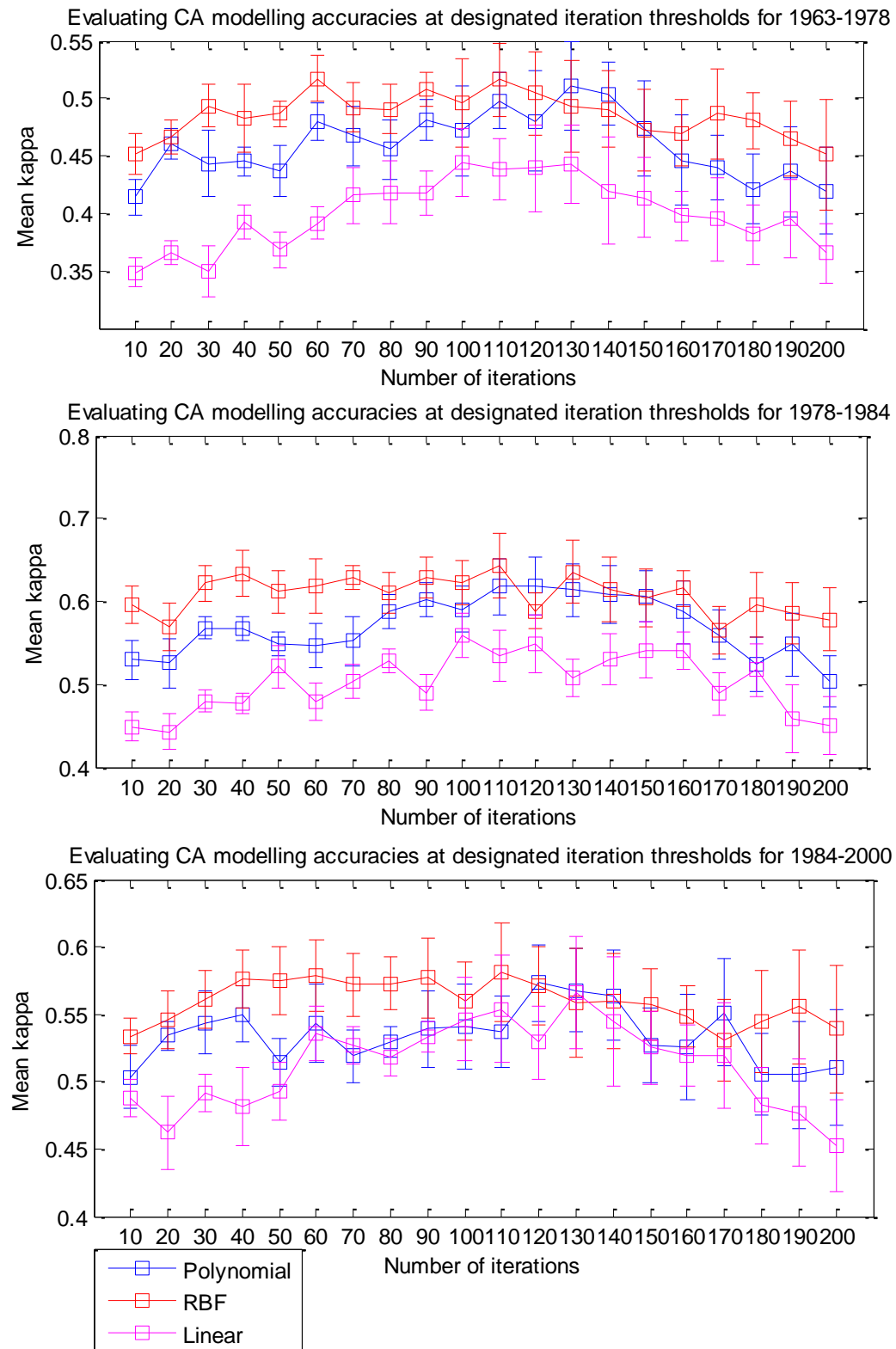


Figure 7.22 *FSVM-based CA: computed overall mean kappa and standard deviations for 200 designated iteration thresholds*

The predicted maps in 1978, 1984, and 2000 (see Figure 7.23) were based on the optimised model parameters ( $C$ ,  $d$ , and  $\gamma$ ) which was used to train the model. The calculated kappa coefficients computed from the confusion matrices given in Tables 7.28-7.36 using the polynomial, RBF, and the linear kernels for periods 1963-1978, 1978-1984, and 1984-2000 were: 0.6062, 0.6257, and 0.5911; 0.7896, 0.8031, and 0.7543; and, 0.7197, 0.7266, and 0.7190, respectively.



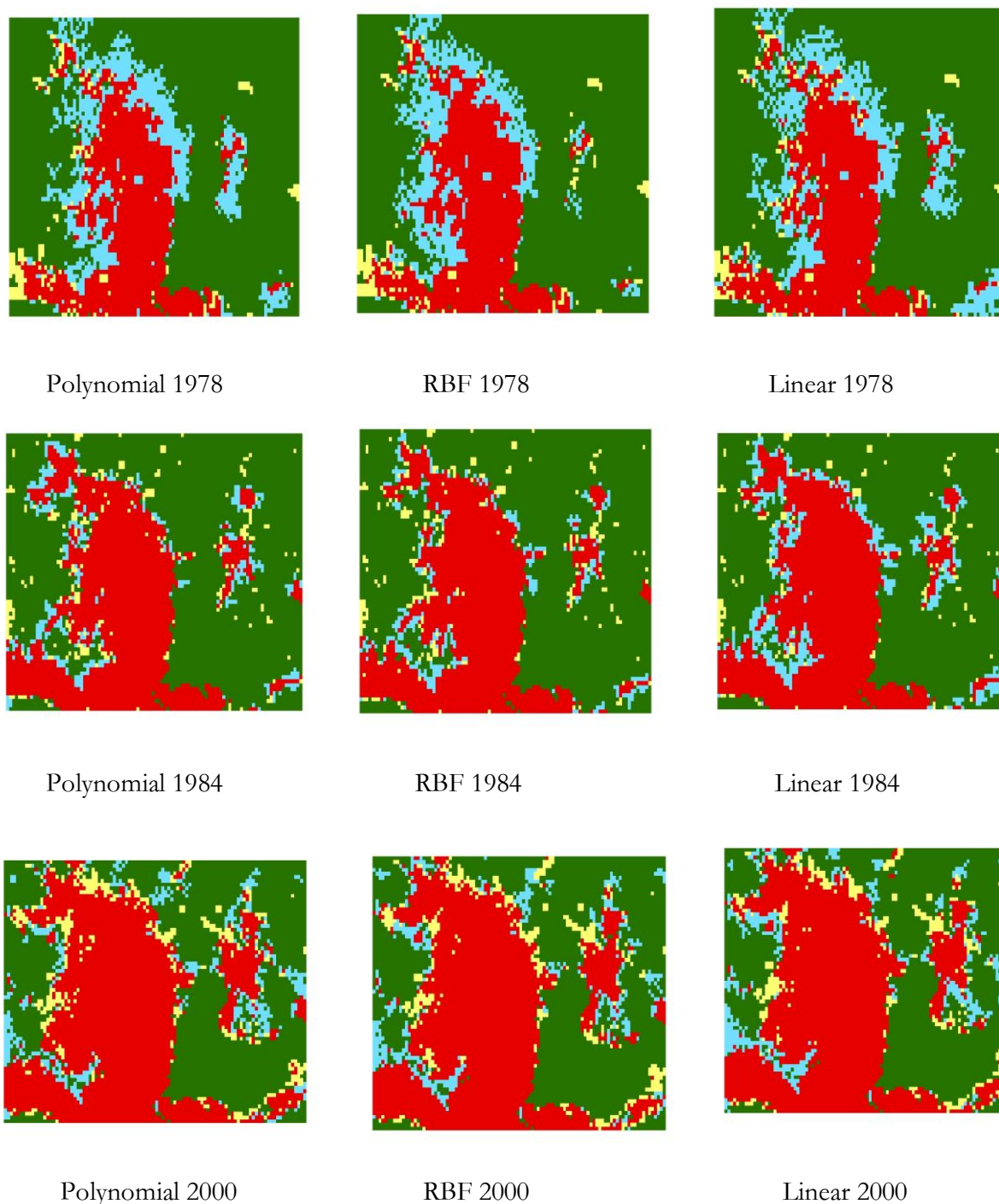


Figure 7.23 FSVN-based CA predicted maps for periods 1963-1978, 1978-1984, and 1984-2000



Table 7.28 *FSVM-based CA: confusion matrix for polynomial 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1580	987
Undeveloped	212	4221

Table 7.29 *FSVM-based CA: confusion matrix for RBF 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1590	929
Undeveloped	202	4279

Table 7.30 *FSVM-based CA: confusion matrix for linear 1963-1978*

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1571	1031
Undeveloped	221	4177

Table 7.31 *FSVM-based CA: confusion matrix for polynomial 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	2247	426
Undeveloped	261	4066

Table 7.32 *FSVM-based CA: confusion matrix for RBF 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	2236	367
Undeveloped	272	4125

Table 7.33 *FSVM-based CA: Confusion matrix for linear 1978-1984*

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	2320	634
Undeveloped	188	3858

Table 7.34 *FSVM-based CA: confusion matrix for polynomial 1984-2000*

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	2974	523
Undeveloped	458	3045

Table 7.35 *FSVM-based CA: confusion matrix for RBF 1984-2000*

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	2982	507
Undeveloped	450	3061

Table 7.36 *FSVM-based CA: confusion matrix for linear 1984-2000*

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	3014	566
Undeveloped	418	3002

Just like the SVM, the FSVM-based CA model accuracy was obtained by repeatedly running two hundred iterations ten times. The computed model accuracies for the polynomial, RBF, and linear kernels are given in Figure 7.24. The RBF kernel yielded the highest mean kappa coefficient while the linear kernel yielded the lowest.

The ROC results derived by running 200 iterations for periods 1963-1978, 1978-1984, and 1984-2000 are depicted in Figures 7.25-7.27. The calculated AUC, 1963-1978, for polynomial, RBF, and linear kernels were:  $0.7856 \pm 0.0290$ ,  $0.7934 \pm 0.0278$ ,  $0.7736 \pm 0.0247$  respectively; for 1978-1984, were  $0.8139 \pm 0.0153$ ,  $0.8219 \pm 0.0094$ ,  $0.8049 \pm 0.0153$  respectively; and for periods 1984-2000,  $0.7879 \pm 0.0191$ ,  $0.8037 \pm 0.0205$ ,  $0.7739 \pm 0.0262$  respectively. The RBF kernel yielded the highest AUC estimate, followed by the polynomial and linear kernels respectively. The computed ROC results corroborate the kappa coefficient results of the CA-based predicted maps. The order of high performance of the three kernel functions was: RBF, polynomial, and linear.

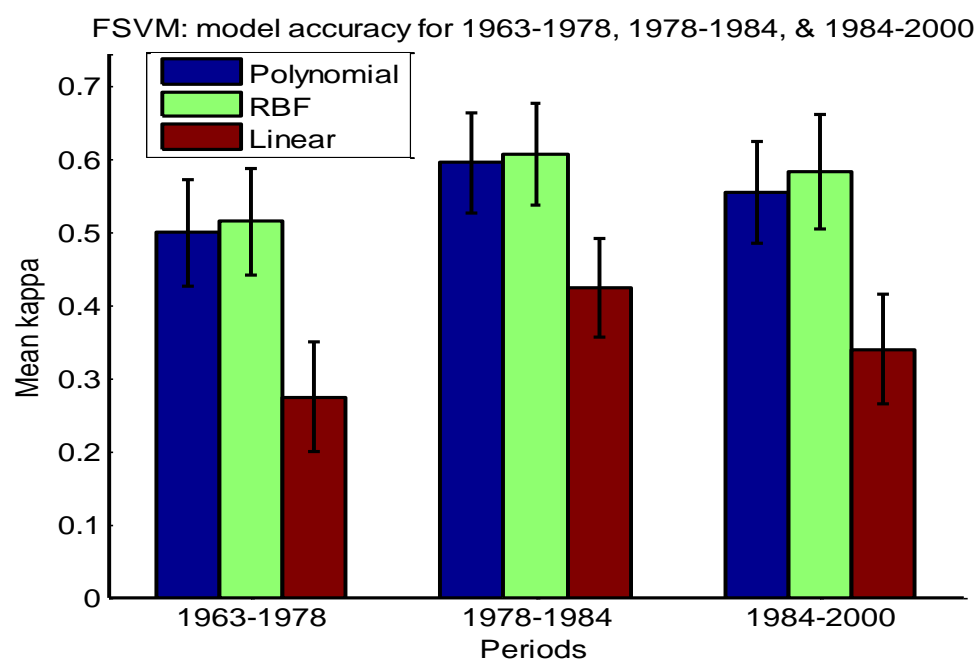


Figure 7.24 FSVM-based CA model accuracy

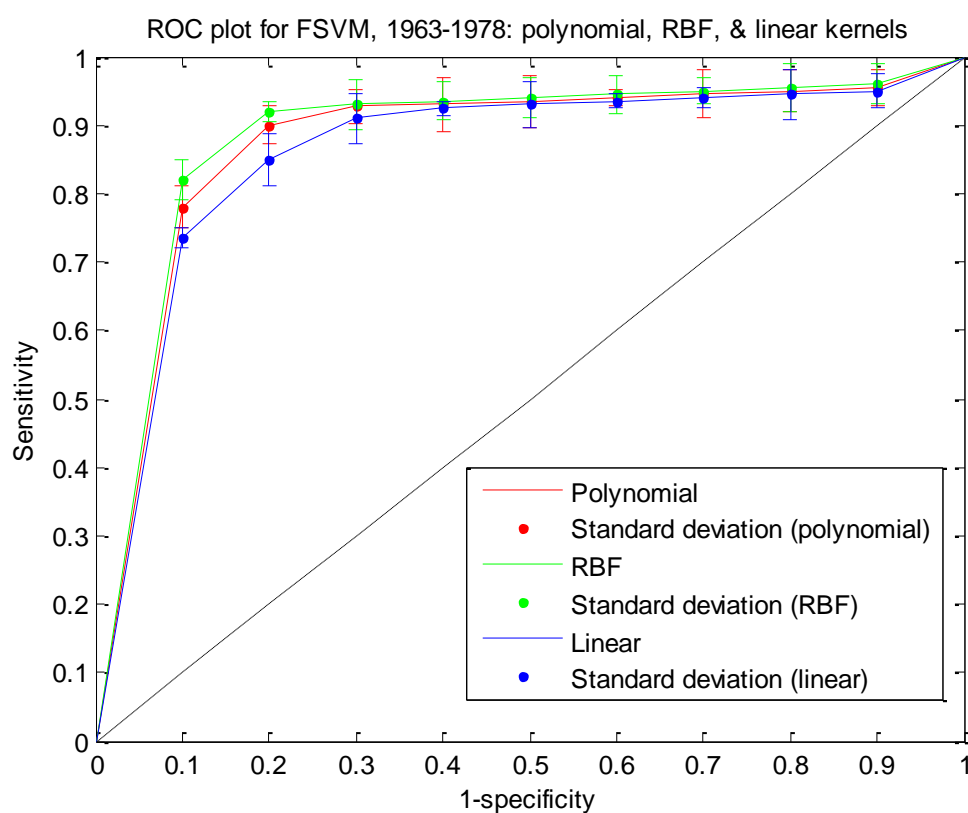


Figure 7.25 FSVM-based CA: ROC plot for period 1963-1978

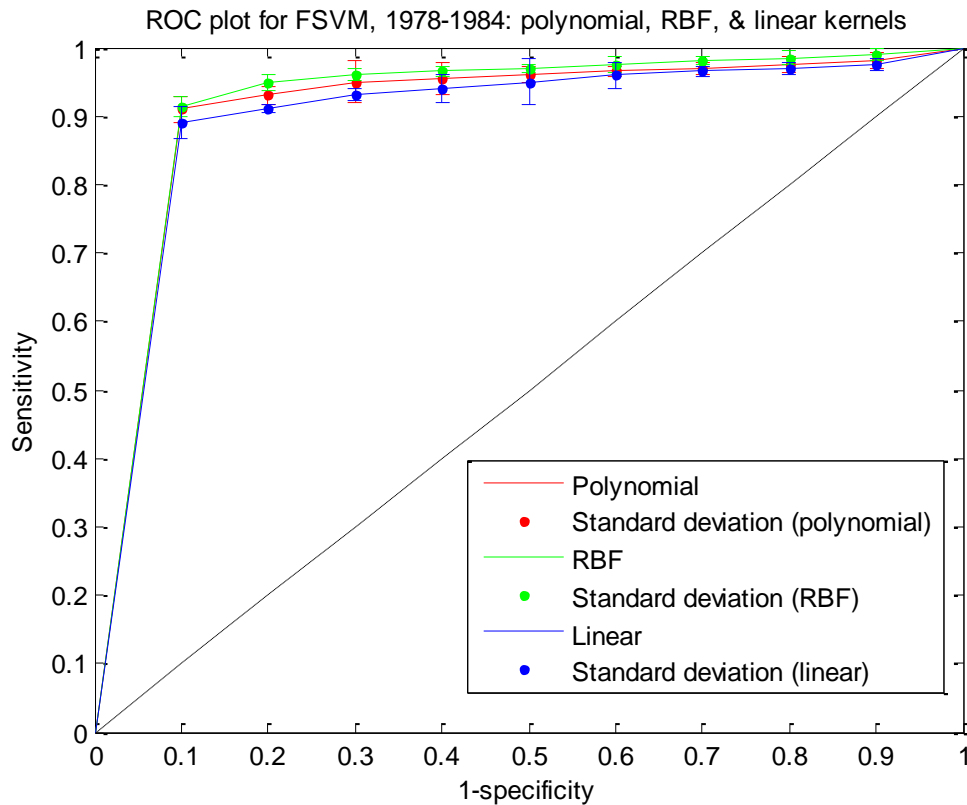


Figure 7.26 *FSVM-based CA: ROC plot for period 1978-1984*

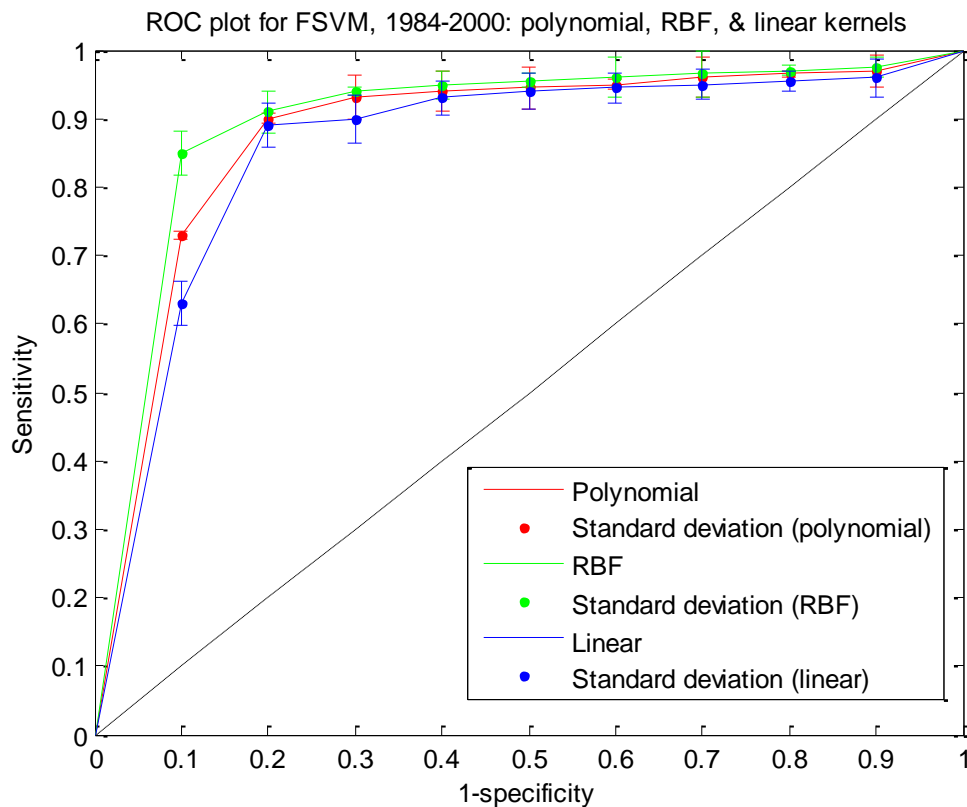


Figure 7.27 *FSVM-based CA: ROC plot for period 1984-2000*

#### **7.4 The effect of the neighbourhood function on SVM and FSVM based CA modelling**

The von Neumann neighbourhood function is based on the influence of the four neighbours within a cell; while the Moore neighbourhood is evaluated using eight surrounding neighbours within a cell (see Figures 2.2 and 2.3). The SVM and FSVM simulation results depicted in Figure 7.28 show that the Moore neighbourhood function performed better than the von Neumann neighbourhood function. Most land use change experiments are based on the Moore neighbourhood function. The FSVM mean kappa coefficients were slightly higher than those of the SVM. Of the three kernel functions used in this experiment, the RBF kernel yielded the highest mean kappa coefficient while the linear kernel yielded the lowest.

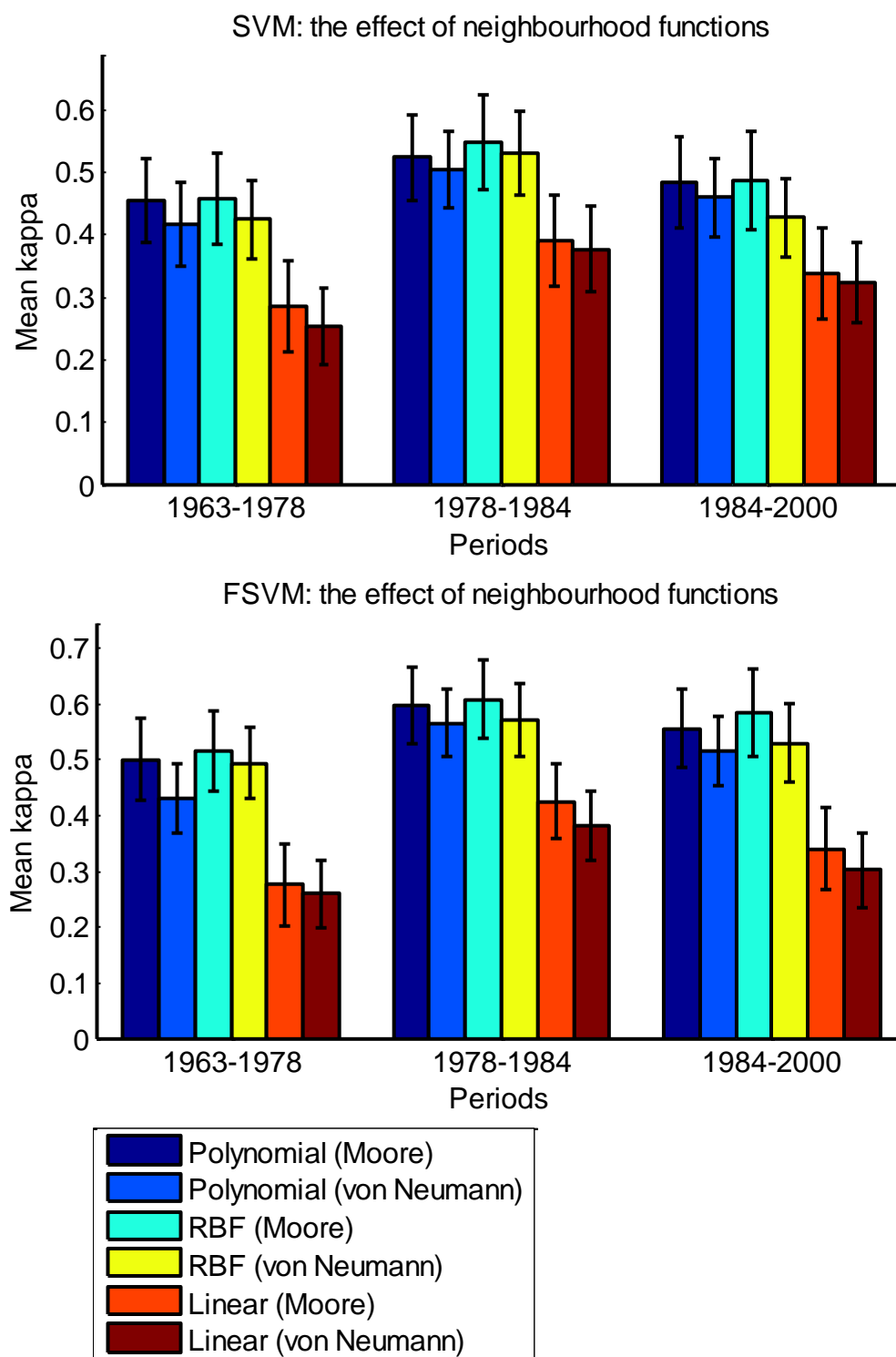


Figure 7.28 *The effect of the Moore and von Neumann neighbourhood functions on the SVM and FSVM based CA modelling*



## 7.5 Predicting future land use using SVM and FSVM based CA

The land use historical trend from 1963-1978, 1978-1984, and 1984-2000 was evaluated and used to forecast the most probable land use maps in 2015 and 2030. Shapefiles for 2015 and 2030 were created in the GIS, based on the evaluation of the most probable land use forms in 2015 and 2030. The GIS shapefiles were converted from vector to raster format and imported into MATLAB for modelling. The resulting shapefiles for 2015 and 2030 were both binary land use maps (that is the maps only consisted of developed and undeveloped pixels/cells; developed cells = +1, and undeveloped cells = -1).

The 2015 and 2030 shapefiles were overlaid with a known land use map in 2000, to determine the changed regions from 2000-2015 and 2000-2030. The same land use variables used for 1984-2000 subsist for the 2000-2015 and 2000-2030 predictions. Note that, empirical measurements were only done in 2000, since it is the known data. Training points must not be taken from the change regions from periods 2000-2015 and 2000-2030 since those change regions are not common to both overlaid maps. The same CA bottom-up technique described in 6.2.4 applies to all the CA experiments implemented in this research.

Future land use in 2015 and 2030 were derived by running the SVM and FSVM based CA models iteratively. The predicted maps in 2015 and 2030 using the SVM and FSVM with the polynomial, RBF, and linear kernels are depicted in Figure 7.29. The forecasted maps in 2015 and 2030 cannot be validated since their actual land use forms can only be known in the future.



## 7.6 Summary

The FSVM performed slightly better than the SVM. The next chapter is the concluding chapter, where the results from all the models will be analysed.

## CHAPTER 8

### CONCLUSION

#### 8.1 Appraisal of modelling results

Section 8.1.1 compares the results from the GIS-based GWR, the LR and ANN based non-CA and CA modelling, with those of the proposed SVM and FSVM based non-CA and CA modelling. Section 8.1.2 compares the results from the LR, ANN, SVM and FSVM based non-CA modelling with those from the LR, ANN, SVM and FSVM based CA modelling.

##### 8.1.1 GIS-based GWR, LR, and ANN versus the proposed SVM and FSVM.

###### *8.1.1.1 GIS-based GWR versus SVM and FSVM.*

Since the SVM and FSVM do not furnish  $R^2$  values, the comparison of the GWR and the proposed models (SVM and FSVM) will be based on the calculated kappa coefficients of the simulated maps.

Table 8.1 contains kappa coefficients for the GWR, SVM, and FSVM predicted maps. From Table 8.1 for periods 1963-1978 and 1984-2000, the GIS-based GWR kappa coefficient for 1963-1978 and 1984-2000 were higher than all kernel results from the SVM and FSVM. For 1978-1984, the GWR kappa statistic was lower than the SVM and FSVM polynomial and RBF results, except for the linear kernels.

Table 8.1 *Calculated kappa coefficients for the GWR, SVM, and FSVM simulated maps*

	GWR	SVM			FSVM		
		Polynomial	RBF	Linear	Polynomial	RBF	Linear
1963-1978	0.8858	0.5870	0.6000	0.5467	0.6062	0.6257	0.5911
1978-1984	0.8366	0.7848	0.7919	0.7527	0.7896	0.8031	0.7543
1984-2000	0.8812	0.7102	0.7170	0.7054	0.7197	0.7266	0.7190

The significance of the difference in the kappa coefficients of the GWR and SVM/FSVM presented in Table 8.1 can be assessed using the McNemar's test (Bradley, 1968; Agresti, 1996). The McNemar's test is used to assess the statistical significance of two related samples (Bradley, 1968; Agresti, 1996; Foody, 2004; Huang et al., 2010). The McNemar's test is based on the elements in a confusion matrix. Therefore, McNemar's test in this test is based on the confusion matrices that yielded the kappa coefficients given in Table 8.1. According to Foody (2004), the McNemar's test evaluates the  $z$ -score from a standardised normal test statistic,

$$z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \quad (8.1)$$

where  $f_{12}$  and  $f_{21}$  are the sum of incorrectly classified pixels resulting from a 2-class problem (see Table 8.2).

Table 8.2 *A typical 2-class confusion matrix*

	Map 1	
	Correct	Incorrect
Map 2		
Correct	$f_{11}$	$f_{12}$
Incorrect	$f_{21}$	$f_{22}$

In this test, let the GWR predicted maps be map 1 while SVM/FSVM predicted maps be map 2. The analysis was done by comparing the GWR and SVM/FSVM predicted maps. The respective values of  $f_{12}$  and  $f_{21}$  were used to compute the respective  $z$ -scores using equation 8.1. The statement of hypothesis can be written as,

$H_0$ : there is no significant difference between the two predicted maps at 95% CL

$H_1$ : there is a significant difference between the two predicted maps (reject  $H_0$  if  $p\text{-value} < 0.05$ )

Using a two-tailed test, the null hypothesis is rejected at 5% significant level that is when  $|z| > 1.96$ . The computed confusion matrices, calculated  $z$ -scores, and  $p$ -values are respectively presented in Table 8.3.

Table 8.3 *Computed confusion matrices, z-scores, and p-values respectively for the GWR versus SVM/FSVM simulated maps (\*significant at  $p < 0.05$  or  $|z| > 1.96$ )*

	GWR versus polynomial SVM	GWR versus RBF SVM	GWR versus linear SVM	GWR versus polynomial FSVM	GWR versus RBF FSVM	GWR versus linear FSVM
1963 - 1978	$\begin{bmatrix} 1652 & 202 \\ 915 & 4231 \end{bmatrix}$ -21.3335 (0.0000*)	$\begin{bmatrix} 1676 & 178 \\ 843 & 4303 \end{bmatrix}$ -20.8118 (0.0000*)	$\begin{bmatrix} 1634 & 220 \\ 968 & 4178 \end{bmatrix}$ -21.7017 (0.0000*)	$\begin{bmatrix} 1669 & 185 \\ 966 & 4180 \end{bmatrix}$ -23.0204 (0.0000*)	$\begin{bmatrix} 1627 & 227 \\ 953 & 4193 \end{bmatrix}$ -21.1347 (0.0000*)	$\begin{bmatrix} 1758 & 96 \\ 1324 & 3822 \end{bmatrix}$ -32.5877 (0.0000*)
1978 - 1984	$\begin{bmatrix} 2439 & 329 \\ 257 & 3975 \end{bmatrix}$ 2.9743 (0.0029*)	$\begin{bmatrix} 2438 & 330 \\ 261 & 3971 \end{bmatrix}$ 2.8383 (0.0045*)	$\begin{bmatrix} 2545 & 223 \\ 406 & 3826 \end{bmatrix}$ -7.2967 (0.0000*)	$\begin{bmatrix} 2396 & 372 \\ 207 & 4025 \end{bmatrix}$ 6.8572 (0.0000*)	$\begin{bmatrix} 2428 & 340 \\ 245 & 3987 \end{bmatrix}$ 3.9278 (8.4000e-05*)	$\begin{bmatrix} 2555 & 213 \\ 399 & 3833 \end{bmatrix}$ -7.5186 (0.0000*)
1984 - 2000	$\begin{bmatrix} 3069 & 479 \\ 428 & 3034 \end{bmatrix}$ 1.6934 (0.0904)	$\begin{bmatrix} 3098 & 450 \\ 482 & 2970 \end{bmatrix}$ -1.0482 (0.2945)	$\begin{bmatrix} 3071 & 477 \\ 418 & 3034 \end{bmatrix}$ 1.9722 (0.0486*)	$\begin{bmatrix} 3101 & 447 \\ 522 & 2930 \end{bmatrix}$ -2.4093 (0.0160*)	$\begin{bmatrix} 3103 & 445 \\ 464 & 2988 \end{bmatrix}$ -0.6302 (0.5286)	$\begin{bmatrix} 3121 & 427 \\ 517 & 2935 \end{bmatrix}$ -2.9293 (0.0034*)

The results from the McNemar's test indicate that the accuracy of the GWR predicted maps for 1963-1978 and 1978-1984 were significantly better than those of all the SVM and FSVM predicted maps using the polynomial, RBF, and the linear kernels. For period 1984-2000, but for the polynomial SVM, RBF SVM, and RBF FSVM the McNemar's test shows that the accuracy of the GWR predicted maps were significantly better than those of the linear SVM, polynomial FSVM, and linear FSVM.

Even though the accuracies of the GIS-based OLS and GWR models were satisfactory for modelling land use change, they do not satisfy basic statistical assumptions required for all linear regression models when used for land use change modelling. The statistical validity of GIS-based OLS and GWR models when applied to land use change modelling has been the main rationale why researchers have continued to crave alternative methods to the conventional GIS-based OLS and GWR models.

### 8.1.1.2 LR versus SVM and FSVM.

Table 8.4 presents the non-CA kappa coefficients of the LR, SVM, and FSVM predicted maps for 1963-1978, 1978-1984, and 1984-2000. From Table 8.4, for periods 1963-1978 and 1978-1984, but for the polynomial SVM and linear SVM, the LR kappa coefficient was lower than those of the RBF SVM, polynomial FSVM, RBF FSVM, and linear FSVM. For 1984-2000, but for the polynomial SVM, linear SVM, and linear FSVM, the LR kappa coefficient was lower than those of the RBF SVM, polynomial FSVM, and RBF FSVM.

Table 8.4 *Non-CA modelling results: calculated kappa coefficients for the LR, SVM, and FSVM simulated maps*

	LR	SVM			FSVM		
		Polynomial	RBF	Linear	Polynomial	RBF	Linear
1963-1978	0.5057	0.5002	0.5326	0.4756	0.5339	0.5559	0.5142
1978-1984	0.5525	0.5436	0.5789	0.5209	0.5641	0.6660	0.5534
1984-2000	0.5080	0.4802	0.5171	0.3545	0.5293	0.5307	0.4404

Table 8.5 presents the CA kappa coefficients of the LR, SVM, and FSVM predicted maps for 1963-1978, 1978-1984, and 1984-2000. From Table 8.5, for all the three periods, 1963-1978, 1978-1984, and 1984-2000 but for the linear SVM, the LR kappa coefficient was lower than the kappa coefficients of all the SVM and FSVM kernel functions.



Table 8.5 *CA modelling results: calculated kappa coefficients for the LR, SVM and FSVM simulated maps*

	LR	SVM			FSVM		
		Polynomial	RBF	Linear	Polynomial	RBF	Linear
1963-1978	0.5847	0.5870	0.6000	0.5467	0.6062	0.6257	0.5911
1978-1984	0.7543	0.7848	0.7919	0.7527	0.7896	0.8031	0.7543
1984-2000	0.7101	0.7102	0.7170	0.7054	0.7197	0.7266	0.7190

The McNemar's test (equation 8.1) was used to assess whether the kappa results from the non-CA and CA simulations presented in Tables 8.4 and 8.5 were significantly different at 5% significant level.

The calculated  $z$ -scores and  $p$ -values for the non-CA results (see Table 8.6) indicate that for LR versus linear FSVM 1978-1984 that there was a significant difference between the LR and the SVM and FSVM predictions for all the three periods.

Table 8.6 *Non-CA modelling: computed confusion matrices, z-scores, and p-values respectively for the LR versus SVM/FSVM simulated maps (\*significant at  $p < 0.05$  or  $|z| > 1.96$ )*

	LR versus polynomial SVM	LR versus RBF SVM	LR versus linear SVM	LR versus polynomial FSVM	LR versus RBF FSVM	LR versus linear FSVM
1963 - 1978	$\begin{bmatrix} 1569 & 1396 \\ 142 & 3893 \end{bmatrix}$ 31.9756 (0.0000*)	$\begin{bmatrix} 1532 & 1433 \\ 118 & 3917 \end{bmatrix}$ 33.3903 (0.0000*)	$\begin{bmatrix} 1015 & 1950 \\ 14 & 4021 \end{bmatrix}$ 43.6852 (0.0000*)	$\begin{bmatrix} 1842 & 1123 \\ 179 & 3856 \end{bmatrix}$ 26.1617 (0.0000*)	$\begin{bmatrix} 1594 & 1371 \\ 121 & 3914 \end{bmatrix}$ 32.3613 (0.0000*)	$\begin{bmatrix} 1344 & 1621 \\ 9 & 4026 \end{bmatrix}$ 39.9274 (0.0000*)
1978 - 1984	$\begin{bmatrix} 2211 & 1296 \\ 49 & 3444 \end{bmatrix}$ 34.0021 (0.0000*)	$\begin{bmatrix} 1893 & 1614 \\ 58 & 3435 \end{bmatrix}$ 38.0532 (0.0000*)	$\begin{bmatrix} 1886 & 1621 \\ 171 & 3322 \end{bmatrix}$ 34.2530 (0.0000*)	$\begin{bmatrix} 2878 & 629 \\ 0 & 3493 \end{bmatrix}$ 25.0799 (0.0000*)	$\begin{bmatrix} 2313 & 1194 \\ 146 & 3347 \end{bmatrix}$ 28.6292 (0.0000*)	$\begin{bmatrix} 3371 & 136 \\ 137 & 3356 \end{bmatrix}$ -0.0605 (0.9518)
1984 - 2000	$\begin{bmatrix} 2638 & 1698 \\ 87 & 2577 \end{bmatrix}$ 38.1308 (0.0000*)	$\begin{bmatrix} 4169 & 167 \\ 829 & 1835 \end{bmatrix}$ -20.9763 (0.0000*)	$\begin{bmatrix} 2516 & 1820 \\ 74 & 2590 \end{bmatrix}$ 40.1194 (0.0000*)	$\begin{bmatrix} 4322 & 14 \\ 404 & 2260 \end{bmatrix}$ -19.0755 (0.0000*)	$\begin{bmatrix} 3532 & 804 \\ 141 & 2523 \end{bmatrix}$ 21.5674 (0.0000*)	$\begin{bmatrix} 2607 & 1729 \\ 64 & 2600 \end{bmatrix}$ 39.3210 (0.0000*)

For the CA modelling (see Table 8.7) but for the LR versus polynomial SVM/RBF SVM/polynomial FSVM/linear FSVM the calculated  $z$ -scores and  $p$ -values indicate that there was a significant difference between the LR and the SVM and FSVM predictions for all the three periods.

Table 8.7 *CA modelling: computed confusion matrices, z-scores, and p-values respectively for the LR versus SVM/FSVM simulated maps (\*significant at  $p < 0.05$  or  $|z| > 1.96$ )*

	LR versus polynomial SVM	LR versus RBF SVM	LR versus linear SVM	LR versus polynomial FSVM	LR versus RBF FSVM	LR versus linear FSVM
1963 - 1978	$\begin{bmatrix} 2163 & 451 \\ 439 & 3947 \end{bmatrix}$ 0.4022 (0.6875)	$\begin{bmatrix} 2239 & 375 \\ 328 & 4058 \end{bmatrix}$ 1.7726 (0.0763)	$\begin{bmatrix} 2343 & 271 \\ 176 & 4210 \end{bmatrix}$ 4.4933 (8.0000e-06*)	$\begin{bmatrix} 2161 & 453 \\ 419 & 3967 \end{bmatrix}$ 1.1514 (0.2496)	$\begin{bmatrix} 2476 & 138 \\ 606 & 3780 \end{bmatrix}$ -17.1577 (0.0000*)	$\begin{bmatrix} 2212 & 402 \\ 423 & 3963 \end{bmatrix}$ -0.7311 (0.4647)
1978 - 1984	$\begin{bmatrix} 2529 & 339 \\ 167 & 3965 \end{bmatrix}$ 7.6463 (0.0000*)	$\begin{bmatrix} 2601 & 267 \\ 98 & 4034 \end{bmatrix}$ 8.8459 (0.0000*)	$\begin{bmatrix} 2643 & 225 \\ 308 & 3824 \end{bmatrix}$ -3.5951 (3.2400e-04*)	$\begin{bmatrix} 2493 & 375 \\ 180 & 3952 \end{bmatrix}$ 8.2773 (0.0000*)	$\begin{bmatrix} 2470 & 398 \\ 133 & 3999 \end{bmatrix}$ 11.5000 (0.0000*)	$\begin{bmatrix} 2644 & 224 \\ 310 & 3822 \end{bmatrix}$ -3.7216 (1.9800e-04*)
1984 - 2000	$\begin{bmatrix} 3229 & 131 \\ 268 & 3372 \end{bmatrix}$ -6.8586 (0.0000*)	$\begin{bmatrix} 3154 & 206 \\ 426 & 3214 \end{bmatrix}$ -8.7511 (0.0000*)	$\begin{bmatrix} 3135 & 225 \\ 354 & 3286 \end{bmatrix}$ -5.3611 (8.2573e-08*)	$\begin{bmatrix} 3175 & 185 \\ 448 & 3192 \end{bmatrix}$ -10.4533 (0.0000*)	$\begin{bmatrix} 3162 & 198 \\ 476 & 3164 \end{bmatrix}$ -10.7082 (0.0000*)	$\begin{bmatrix} 3167 & 193 \\ 400 & 3240 \end{bmatrix}$ -8.5005 (0.0000*)

A further validation of the CA modelling was done with the ROC analysis. Table 8.8 presents the calculated AUC and standard deviations for the LR, SVM, and FSVM models. The computed AUC given in Table 8.8 indicate that the AUC for the LR model was lower than those of all the SVM and FSVM kernel results for periods 1963-1978, 1978-1984, and 1984-2000 with the exception of the SVM linear kernel. Now, we can test statistically whether the calculated AUC for the LR and the SVMs (that is, SVM and FSVM) differ significantly at 95% CL.

Table 8.8 *ROC analysis: calculated mean AUC and their respective standard deviations for LR versus SVM/FSVM simulated maps*

	LR	SVM			FSVM			
		Polynomial	RBF	Linear	Polynomial	RBF	Linear	
1963-1978	0.7354±0.0295	0.7629±0.0270	0.7863±0.0273	0.7464±0.0287	0.7856 ±0.0292	0.7934±0.0278	0.7736±0.0247	Tsai and Chen (2004) present
1978-1984	0.7549±0.0267	0.8003±0.0300	0.8139±0.0248	0.7673±0.0214	0.8139±0.0153	0.8219±0.0094	0.8049±0.0153	
1984-2000	0.7451±0.0298	0.7804±0.0316	0.7939±0.0290	0.7714±0.0255	0.7879±0.0191	0.8037±0.0205	0.7739±0.0262	

ed a technique for testing the significance of the AUC estimates obtained from ROC plots. According to Tsai and Chen (2004), ROC indices from two samples with unequal variance can be assessed using the  $t$ -statistic,

$$T_i = \frac{\bar{y}_{it} - \bar{y}_{ic}}{\sqrt{S_{ic}^2/n_1 + S_{it}^2/n_2}} \quad (8.2)$$

where  $\bar{y}_{it}$  and  $\bar{y}_{ic}$  are the means of sample 1 and sample 2;  $S_{it}^2$  and  $S_{ic}^2$  are the variances of sample 1 and sample 2 ; and,  $n_1$  and  $n_2$  are the sizes of samples 1 and 2, all respectively. The hypothesis can be stated as,

$H_0$ : there is no significant difference between samples 1 and 2 at 95% CL

$H_1$ : there is a significant difference between samples 1 and 2 (reject  $H_0$  if  $p\text{-value} < 0.05$  )

The AUC and their respective standard deviations given in Table 8.8 were derived by computing the mean of running 200 iterations 10 times. Therefore,  $n_1 = n_2 = 10$  . Then, degrees of freedom (df) =  $n_1 + n_2 - 2 = 18$ . Using equation 8.2 based on a two-tailed test, the  $t$ -statistics

and their respective  $p$ -values were calculated using the calculated AUC in Table 8.8. The test results given in Table 8.9 show that the performance of the LR model was significantly less than the polynomial SVM, RBF SVM, polynomial FSVM, RBF FSVM, and linear FSVM in all the three periods. The test also indicated that there was no difference between the performances of the LR and the linear SVM in all the three periods.

Table 8.9 Calculated  $t$ -statistic and  $p$ -values (in parenthesis) for LR versus SVM/FSVM simulated maps (\*significant at  $p < 0.05$  or  $|t| > 1.96$ )

	SVM			FSVM		
	LR versus polynomial SVM	LR versus RBF SVM	LR versus linear SVM	LR versus polynomial FSVM	LR versus RBF FSVM	LR versus linear FSVM
1963-1978	2.17458 (0.04330*)	4.0046 (0.00083*)	0.84517 (0.40920)	3.8245024 (0.00124*)	4.52478 (0.00026*)	3.139668 (0.00567*)
1978-1984	3.57481 (0.00217*)	5.11993 (0.00007*)	1.14596 (0.26721)	6.06292 (0.00001*)	7.484988 (0.000001*)	5.13807 (0.00007*)
1984-2000	2.57001 (0.01928*)	3.71123 (0.00160*)	2.120498 (0.04816*)	3.82379 (0.00125*)	5.12325 (0.000071*)	2.29522 (0.03397*)

### 8.1.1.3 ANN versus SVM and FSVM.

Table 8.10 presents the non-CA kappa coefficients of the ANN, SVM, and FSVM predicted maps for 1963-1978, 1978-1984, and 1984-2000. From Table 8.10, for periods 1963-1978, but for the polynomial SVM, linear SVM, and linear FSVM the ANN kappa coefficient was lower than the RBF SVM, polynomial FSVM, and RBF FSVM. For 1978-1984, but for the RBF SVM and RBF FSVM the ANN kappa coefficient was higher than the polynomial SVM, linear SVM, polynomial FSVM, and linear FSVM. For 1984-2000, but for the polynomial SVM, linear SVM, and linear FSVM the ANN kappa coefficient was lower than the RBF SVM, polynomial FSVM, and RBF FSVM.

Table 8.10 *Non-CA modelling: calculated kappa coefficients for the ANN, SVM, and FSVM simulated maps*

	ANN	SVM			FSVM		
		Polynomial	RBF	Linear	Polynomial	RBF	Linear
1963-1978	0.5165	0.5002	0.5326	0.4756	0.5339	0.5559	0.5142
1978-1984	0.5701	0.5436	0.5789	0.5209	0.5641	0.6660	0.5534
1984-2000	0.5149	0.4802	0.5171	0.3545	0.5293	0.5307	0.4404

Table 8.11 presents the CA kappa coefficients of the ANN, SVM, and FSVM predicted maps for 1963-1978, 1978-1984, and 1984-2000. From Figure 8.11, 1963-1978, but for the polynomial SVM and linear SVM the ANN kappa coefficient was lower than those of the RBF SVM, polynomial FSVM, RBF FSVM, and linear FSVM. For 1978-1984, but for the linear SVM and linear FSVM the ANN kappa coefficient was lower than those of the polynomial SVM, RBF SVM, polynomial FSVM, and RBF FSVM. For 1984-2000, but for the polynomial SVM and linear SVM the ANN kappa coefficient was lower than those of the RBF SVM, polynomial FSVM, RBF FSVM, and linear FSVM.

Table 8.11 *CA modelling: calculated kappa coefficients for the ANN, SVM, and FSVM simulated maps*

	ANN	SVM			FSVM		
		Polynomial	RBF	Linear	Polynomial	RBF	Linear
1963-1978	0.5900	0.5870	0.6000	0.5467	0.6062	0.6257	0.5911
1978-1984	0.7825	0.7848	0.7919	0.7527	0.7896	0.8031	0.7543
1984-2000	0.7161	0.7102	0.7170	0.7054	0.7197	0.7266	0.7190

The McNemar's test statistic (equation 8.3) was used to test whether there was a significant difference between the ANN and SVMs kappa coefficients for the non-CA and CA modelling given in Tables 8.10 and 8.11 respectively. From Tables 8.12, the computed  $z$ -scores and  $p$ -values indicate that but for the polynomial FSVM in period 1978-1984, there was a significant difference between the ANN and all the SVM and FSVM predictions.

Table 8.12 *Non-CA modelling: computed confusion matrices,  $z$ -scores, and  $p$ -values respectively for the ANN versus SVM/FSVM simulated maps (\*significant at  $p < 0.05$  or  $|z| > 1.96$ )*

	ANN versus polynomial SVM	ANN versus RBF SVM	ANN versus linear SVM	ANN versus polynomial FSVM	ANN versus RBF FSVM	ANN versus linear FSVM
1963 - 1978	$\begin{bmatrix} 1418 & 1111 \\ 293 & 4178 \end{bmatrix}$ 21.8308 (0.0000*)	$\begin{bmatrix} 1343 & 1186 \\ 307 & 4164 \end{bmatrix}$ 22.7488 (0.0000*)	$\begin{bmatrix} 1025 & 1504 \\ 4 & 4467 \end{bmatrix}$ 38.6270 (0.0000*)	$\begin{bmatrix} 1425 & 1104 \\ 290 & 4181 \end{bmatrix}$ -21.8018 (0.0000*)	$\begin{bmatrix} 1266 & 1263 \\ 87 & 4384 \end{bmatrix}$ 32.0067 (0.0000*)	$\begin{bmatrix} 1569 & 960 \\ 452 & 4019 \end{bmatrix}$ 13.5191 (0.0000*)
1978 - 1984	$\begin{bmatrix} 1633 & 1033 \\ 424 & 3910 \end{bmatrix}$ 15.9547 (0.0000*)	$\begin{bmatrix} 1831 & 835 \\ 429 & 3905 \end{bmatrix}$ 11.4196 (0.0000*)	$\begin{bmatrix} 1613 & 1053 \\ 338 & 3996 \end{bmatrix}$ 19.1709 (0.0000*)	$\begin{bmatrix} 1948 & 537 \\ 511 & 4004 \end{bmatrix}$ 0.8031 (0.4219)	$\begin{bmatrix} 2438 & 47 \\ 1070 & 3445 \end{bmatrix}$ -30.6090 (0.0000*)	$\begin{bmatrix} 2197 & 288 \\ 681 & 3834 \end{bmatrix}$ -12.6250 (0.0000*)
1984 - 2000	$\begin{bmatrix} 2193 & 883 \\ 397 & 3527 \end{bmatrix}$ 31.5841 (0.0000*)	$\begin{bmatrix} 2252 & 824 \\ 473 & 3451 \end{bmatrix}$ 9.7462 (0.0000*)	$\begin{bmatrix} 2988 & 88 \\ 2010 & 1914 \end{bmatrix}$ -41.9615 (0.0000*)	$\begin{bmatrix} 2352 & 724 \\ 319 & 3605 \end{bmatrix}$ 12.5404 (0.0000*)	$\begin{bmatrix} 2846 & 230 \\ 827 & 3097 \end{bmatrix}$ -18.3627 (0.0000*)	$\begin{bmatrix} 3062 & 14 \\ 1664 & 2260 \end{bmatrix}$ -40.2799 (0.0000*)

From Table 8.13, the McNemar's test statistic indicated that for period 1963-1978, aside the polynomial SVM, RBF SVM, polynomial FSVM that there was a significant difference between the ANN result and those of the linear SVM, RBF FSVM, and linear FSVM. For 1978-1984, aside the polynomial SVM, RBF SVM, and the polynomial FSVM there was a significant difference between the ANN result and those of the linear SVM, RBF FSVM, and linear FSVM. For 1984-2000, aside the RBF SVM and linear FSVM there was a significant difference between

the ANN result and those of the polynomial SVM, linear SVM, polynomial FSVM, and RBF FSVM.

Table 8.13 *CA modelling: computed confusion matrices, z-scores, and p-values respectively for the ANN versus SVM/FSVM simulated maps (\*significant at  $p < 0.05$  or  $|z| > 1.96$ )*

	ANN versus polynomial SVM	ANN versus RBF SVM	ANN versus linear SVM	ANN versus polynomial FSVM	ANN versus RBF FSVM	ANN versus linear FSVM
1963 - 1978	$\begin{bmatrix} 2166 & 535 \\ 436 & 3863 \end{bmatrix}$ 3.1771 (0.0015*)	$\begin{bmatrix} 2215 & 486 \\ 352 & 3947 \end{bmatrix}$ 4.6290 (4.0000e-06*)	$\begin{bmatrix} 2293 & 408 \\ 226 & 4073 \end{bmatrix}$ 7.2281 (0.0000*)	$\begin{bmatrix} 2179 & 522 \\ 401 & 3898 \end{bmatrix}$ 3.9828 (6.8000e-05*)	$\begin{bmatrix} 2492 & 209 \\ 590 & 3709 \end{bmatrix}$ -13.4788 (0.0000*)	$\begin{bmatrix} 2188 & 513 \\ 447 & 3852 \end{bmatrix}$ 2.1301 (0.0332*)
1978 - 1984	$\begin{bmatrix} 2412 & 254 \\ 284 & 4050 \end{bmatrix}$ -1.2934 (0.1959)	$\begin{bmatrix} 2466 & 200 \\ 233 & 4101 \end{bmatrix}$ -1.5859 (0.1128)	$\begin{bmatrix} 2508 & 158 \\ 443 & 3891 \end{bmatrix}$ -11.6254 (0.0000*)	$\begin{bmatrix} 2434 & 232 \\ 239 & 4095 \end{bmatrix}$ -0.3225 (0.7471)	$\begin{bmatrix} 2390 & 276 \\ 213 & 4121 \end{bmatrix}$ 2.8490 (0.0044*)	$\begin{bmatrix} 2506 & 160 \\ 448 & 3886 \end{bmatrix}$ -11.6799 (0.0000*)
1984 - 2000	$\begin{bmatrix} 3140 & 414 \\ 357 & 3089 \end{bmatrix}$ 2.0528 (0.0401*)	$\begin{bmatrix} 3199 & 355 \\ 381 & 3065 \end{bmatrix}$ -0.9584 (0.3379)	$\begin{bmatrix} 3151 & 403 \\ 338 & 3108 \end{bmatrix}$ 2.3878 (0.0170*)	$\begin{bmatrix} 3211 & 343 \\ 412 & 3034 \end{bmatrix}$ -2.5112 (0.0120*)	$\begin{bmatrix} 3202 & 352 \\ 436 & 3010 \end{bmatrix}$ -2.9924 (0.0028*)	$\begin{bmatrix} 3190 & 364 \\ 377 & 3069 \end{bmatrix}$ -0.4776 (0.9552)

The AUC indices given in Table 8.14 show that the performance of the ANN was less than that of the SVM and FSVM for all three periods (1963-1978, 1978-1984, and 1984-2000). The  $t$ -statistic given in equation 8.4 was used to compare AUC values from the ANN, SVM, and FSVM simulations. The calculated  $t$ -statistic and  $p$ -values (see Table 8.15) indicate that there were significant differences between the performances of the ANN and the RBF SVM, polynomial FSVM, and RBF FSVM but there were no significant differences between the performances of the ANN and those of the polynomial SVM, linear SVM, and linear FSVM.



Table 8.14 ROC results for the ANN, SVM, and FSVM simulated maps

	ANN	SVM			FSVM		
		Polynomial	RBF	Linear	Polynomial	RBF	Linear
1963-1978	0.7574 ± 0.0304	0.7629 ± 0.0270	0.7863 ± 0.0273	0.7464 ± 0.0287	0.7856 ± 0.0292	0.7934 ± 0.0278	0.7736 ± 0.0247
1978-1984	0.7850 ± 0.0308	0.8003 ± 0.0300	0.8139 ± 0.0248	0.7673 ± 0.0214	0.8139 ± 0.0153	0.8219 ± 0.0094	0.8049 ± 0.0153
1984-2000	0.7613 ± 0.0288	0.7804 ± 0.0316	0.7939 ± 0.0290	0.7714 ± 0.0255	0.7879 ± 0.0191	0.8037 ± 0.0205	0.7739 ± 0.0262

Table 8.15 Calculated *t*-statistics and *p*-values (in parenthesis) for ANN versus SVM/FSVM simulated maps (\*significant at  $p < 0.05$  or  $|t| > 1.96$ )

	SVM			FSVM		
	ANN versus polynomial SVM	ANN versus RBF SVM	ANN versus linear SVM	ANN versus polynomial FSVM	ANN versus RBF FSVM	ANN versus linear FSVM
1963-1978	0.42776 (0.67394)	2.23672 (0.03820*)	0.832033 (0.41631)	2.11558 (0.04859*)	2.76351 (0.01280*)	1.30788 (0.20740)
1978-1984	1.12529 (0.27529)	2.31113 (0.03288*)	1.492409 (0.15291)	2.65739 (0.01604*)	3.62357 (0.00194*)	1.82983 (0.08389)
1984-2000	1.41268 (0.17483)	2.52233 (0.02130*)	0.8303 (0.41725)	2.43408 (0.02557*)	3.79284 (0.00133*)	1.02338 (0.31972)

### 8.1.2 A comparison of the non-CA and CA predicted maps for LR, ANN, SVM, and FSVM models.

The McNemar's test described in section 8.1.1.1 was also used to statistically assess the whether there are significant differences between the non-CA and CA predicted maps from LR, ANN, SVM, and FSVM models. The computed confusion matrices, calculated  $z$ -scores, and  $p$ -values by comparing the predicted LR, ANN, SVM, and FSVM non-CA and CA maps are presented in Tables 8.16-8.19.

The calculated  $z$ -scores and  $p$ -values indicate that the predicted non-CA and CA maps for all the three periods (1963-1978, 1978-1984, and 1984-2000) were significantly different at the 95% CL.

Table 8.16 *LR: computed confusion matrices,  $z$ -scores, and  $p$ -values respectively (\*significant at  $p < 0.05$  or  $|z| > 1.96$ )*

LR: non-CA versus CA		
1963-1978	$\begin{bmatrix} 2275 & 339 \\ 690 & 3696 \end{bmatrix}$	-10.9421 (0.0000*)
1978-1984	$\begin{bmatrix} 2724 & 144 \\ 783 & 3349 \end{bmatrix}$	-20.9875 (0.0000*)
1984-2000	$\begin{bmatrix} 3141 & 219 \\ 1195 & 2445 \end{bmatrix}$	-25.9552 (0.0000*)

Table 8.17 *ANN: computed confusion matrices, z-scores, and p-values respectively (\*significant at  $p < 0.05$  or  $|z| > 1.96$ )*

ANN: non-CA versus CA		
1963-1978	$\begin{bmatrix} 2428 & 101 \\ 1908 & 2563 \end{bmatrix}$	-40.3151 (0.0000*)
1978-1984	$\begin{bmatrix} 1949 & 717 \\ 536 & 3798 \end{bmatrix}$	5.1133 (3.1671e-07*)
1984-2000	$\begin{bmatrix} 2518 & 1036 \\ 558 & 2888 \end{bmatrix}$	11.9725 (0.0000*)

Table 8.18 *SVM: computed confusion matrices, z-scores, and p-values respectively (\*significant at  $p < 0.05$  or  $|z| > 1.96$ )*

	Polynomial SVM	RBF SVM	Linear SVM
1963-1978	$\begin{bmatrix} 1395 & 1172 \\ 255 & 4178 \end{bmatrix}$ 24.2749 (0.0000*)	$\begin{bmatrix} 1339 & 1263 \\ 372 & 4026 \end{bmatrix}$ 22.0353 (0.0000*)	$\begin{bmatrix} 960 & 1559 \\ 69 & 4412 \end{bmatrix}$ 36.9283 (0.0000*)
1978-1984	$\begin{bmatrix} 1909 & 1042 \\ 351 & 3698 \end{bmatrix}$ 18.5141 (0.0000*)	$\begin{bmatrix} 1628 & 1068 \\ 323 & 3981 \end{bmatrix}$ 19.9753 (0.0000*)	$\begin{bmatrix} 1656 & 1043 \\ 401 & 3900 \end{bmatrix}$ 16.8947 (0.0000*)
1984-2000	$\begin{bmatrix} 2424 & 1156 \\ 301 & 3119 \end{bmatrix}$ 22.3994 (0.0000*)	$\begin{bmatrix} 2268 & 1221 \\ 322 & 3189 \end{bmatrix}$ 22.8864 (0.0000*)	$\begin{bmatrix} 3261 & 236 \\ 1737 & 1766 \end{bmatrix}$ -33.7923 (0.0000*)

Table 8.19 *FSVM: computed confusion matrices, z-scores, and p-values respectively (\*significant at  $p < 0.05$  or  $|z| > 1.96$ )*

	Polynomial FSVM	RBF FSVM	Linear FSVM
1963-1978	$\begin{bmatrix} 1599 & 1036 \\ 422 & 3943 \end{bmatrix}$ 16.0801 (0.0000*)	$\begin{bmatrix} 1438 & 1142 \\ 277 & 4143 \end{bmatrix}$ 22.9628 (0.0000*)	$\begin{bmatrix} 1304 & 1778 \\ 49 & 3869 \end{bmatrix}$ 40.4507 (0.0000*)
1978-1984	$\begin{bmatrix} 2458 & 496 \\ 420 & 3626 \end{bmatrix}$ 2.5111 (0.0120*)	$\begin{bmatrix} 1912 & 691 \\ 547 & 3850 \end{bmatrix}$ 4.0926 (4.2000e-05*)	$\begin{bmatrix} 2483 & 190 \\ 1025 & 3302 \end{bmatrix}$ -23.9551 (0.0000*)
1984-2000	$\begin{bmatrix} 2422 & 1201 \\ 249 & 3128 \end{bmatrix}$ 25.0007 (0.0000*)	$\begin{bmatrix} 3029 & 538 \\ 644 & 2789 \end{bmatrix}$ -3.0832 (0.0020*)	$\begin{bmatrix} 3440 & 198 \\ 1286 & 2076 \end{bmatrix}$ -28.2431 (0.0000*)

## 8.2 Overview

### 8.2.1 Research questions and answers.

A recap of the research questions posed in section 1.3 of chapter one are:

- (i) Did the conventional GIS-based GWR model perform better than the proposed SVM and FSVM based CA models?
- (ii) Did the LR-based CA model yield better accuracies than the proposed SVM and FSVM based CA models?

- (iii) Did the ANN-based CA model outperform the proposed SVM and FSVM based CA models?
- (iv) Were the results of the CA based techniques better than those of the non-CA based techniques?

Answers to the above questions: From section 8.1.1.1, the calculated kappa coefficients from the predicted GWR, SVM, and FSVM maps indicate that the GWR performed better than the SVM and FSVM. The statistical test also validates that the GWR results were better than the SVM and FSVM. From sections 8.1.1.1 and 8.1.1.3, the tests show that the SVM and FSVM results were better than those of the LR and ANN. The tests done in section 8.1.2 indicate that the CA results were better than those of the non-CA.

### **8.2.2 The pros and cons of the proposed SVM and FSVM models.**

One of the weaknesses of the SVM and FSVM was that, unlike the GIS-based OLS/GWR, LR, and one neuron neural network models, they could not be used to assess the contribution of each independent variable. Another shortcoming that they had was the rigour of optimising their kernel parameters (that is, finding their optimal parameters) before implementation. The advantages achieved with the SVM and FSVM models include: (i) the use of kernel functions to overcoming the problem of curse of dimensionality, (ii) the realisation of various simulation outputs as a result of using different kernel functions, (iii) that, with the exception of the GIS-based model, the predicted SVM and FSVM results were better than those of the other unconventional methods.

### 8.3 The benefit of the thesis to Lagos and Nigeria

Lagos is the largest and the most congested city in Nigeria; and as a result, poses the greatest planning burden to the Nigerian government. As explained in section 1.1, the use of predictive models for planning is presently lacking in Lagos and Nigeria. Incorporating predictive models into planning will transform the present prevailing spontaneous planning to proactive and sustainable planning.

This CA-based research will assist urban planners in forecasting urban expansion in Lagos, as well as other sprawling cities in Nigeria. As an academic in one of Nigerian government owned universities, I will make this research freely available to the Nigerian planning authorities including the academia where I belong. The SVMs employ several kernel functions that can be used to derive multiple scenarios of urban disposition in a complex urban environment like Lagos.

### 8.4 The scientific benefit of the thesis

CA-based models have been found very attractive to researchers because of their simplicity and bottom-up approach based on very simple rules (O'Sullivan & Torrens, 2000). The LR model was first explored by Wu (2002) for CA model calibration while the ANN model was first explored by Li and Yeh (2002).

In order to enhance the level of accuracy in predicting urban systems researchers have continued to seek new mathematical models that can effectively model urban systems. This research explored for the first time the use of the SVM-based and FSVM-based models to calibrate a CA model. This research has found that the SVM and FSVM models are very satisfactory for calibrating CA models for land use change prediction. The proposed models can be applied to cities that share similar characteristics as Lagos. This research has been peer-

reviewed and presented in three significant international conferences<sup>4</sup>; and has also been accepted for publication in the Journal of Land Use Science<sup>5</sup>. The satisfactory results from the SVM and the FSVM models have showed that SVMs remain promising tools for modelling land use change.

## 8.5 Recommendations for further research

Mathematical models such as Gaussian Processes (GP), K nearest Neighbour (KNN), Gaussian Mixture Model (GMM), Maximum Likelihood Classifier (MLC), and Situation Theory (ST) are among notable machine learning algorithms that have not been explored for CA modelling. The aforementioned mathematical models are recommended for further research into CA modelling.

---

<sup>4</sup> Okwuashi, O., McConchie, J., Nwilo, P., & Eyo, E. (2008). *Fuzzy support machine constrained GIS-based cellular automata for urban change simulation*. Paper presented at the 24th New Zealand Geographical Conference: Inequality, Sustainability, Policy; Across the Divide, Wellington, New Zealand, July 2<sup>nd</sup> – 5<sup>th</sup>.

Okwuashi, O., McConchie, J., Nwilo, P., & Eyo, E. (2009a). Enhancing a GIS cellular automata model of land use change using support vector machine. *Proceedings of the 17th International Conference on Geoinformatics*, August 12<sup>th</sup> – 14<sup>th</sup>, Fairfax, VA: IEEE Xplore.

Okwuashi, O., McConchie, J., Nwilo, P., & Eyo, E. (2009b). Stochastic GIS cellular automata for land use change simulation: Application of a kernel based model. In B. G. Lees & S. W. Laffan (Eds.), *10<sup>th</sup> International Conference on Geocomputation* (pp. 203-209). UNSW, Sydney, 30<sup>th</sup> November – 2<sup>nd</sup> December.

<sup>5</sup> Accepted for publication on 19<sup>th</sup> August 2010

## REFERENCES

- Abiodun, J. O. (1974). Urban growth and problems in metropolitan Lagos. *Urban Studies*, 11 (33), 341-347.
- Abiodun, J. O. (1997). The challenges of growth and development in metropolitan Lagos. In C. Rakodi (Ed.), *The urban challenge in Africa: Growth and management of its large cities* (pp. 192-222). Tokyo, Japan: United Nations University Press.
- Agresti, A. (1996). *An introduction to categorical data analysis*. New York, NY: Wiley.
- Aguiar, A. P. D. (2006). *Modelling land use change in the Brazilian Amazon: Exploring the intra-regional heterogeneity* (Unpublished doctoral dissertation). INPE, São José dos Campos, SP, Brazil.
- Aizerman, M. A., Braverman, E. M., & Rozonoer, L. I. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25, 821 – 837.
- Al-Ahmadi, K., Heppenstall, A., Hogg, J., & See, L. (2009). A fuzzy cellular automata urban growth model (FCAUGM) for the city of Riyadh, Saudi Arabia. Part 2: Scenario testing. *Appl. Spatial Analysis*, 2, 85 – 105.
- Almeida, C. M., Monteiro, A. M. V., & Camara, G. (2002). *Modelling urban land use dynamics through Bayesian Probabilistic methods in a cellular automaton environment*. Paper presented at the 29th International Symposium on Remote Sensing of Environment, Buenos Aires, Argentina, April 8<sup>th</sup> – 12<sup>th</sup>.
- Almeida, C. M., Batty, M., Monteiro, A. M. V., Camara, G., Soares-Filho, B. S., Cerqueira, G. C., & Pennachin, C. L. (2003a). Stochastic cellular automata modelling of urban land use dynamics: Empirical development and estimation. *Computers, Environment and Urban Systems*, 27(5), 481 – 509.
- Almeida, C. M., Gleriani, J. M., Castejon, E. F., & Soares-Filho, B. S. (2008). Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *International Journal of Geographical Information Science*, 22 (9), 943-963.
- Almeida, C. M., Monteiro, A. M. V., Câmara, G., Soares-filho, B. S., Cerqueira, G. C., Araújo, W. L., & Pantuzzo, A. E. (2003b). *Simulating urban land use change through CA-based*



- modeling and logistic regression*. Paper presented at Anais XI SBSR, Belo Horizonte, Brasil, April 5<sup>th</sup> – 10<sup>th</sup>.
- Alsmadi, M. K. S., Omar, K. B., & Noah, S. A. (2009). Back propagation algorithm: The best algorithm among the multi-layer perceptron algorithm. *International Journal of Computer Science and Network Security*, 9(4), 378 – 383.
- Anderson, J. A. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press.
- Angotti, T. (1993). *Metropolis 2000: Planning, poverty and politics*. London, England: Routledge.
- Anselin, L., & Griffith, D.A. (1988). Do spatial effects really matter in regression analysis? *Pap. Reg. Sci. Assoc.* 65, 11 – 34.
- Antoni, J. (2001). Urban sprawl modelling : A methodological approach. In Cybergeog: European Journal of Geography [En ligne], Dossiers, *Quantitative and theoretical geography* (pp. 2-14). St-Valeryen-Caux, France: 12th European Colloquium.
- Baker, W. L. (1989). A review of models of landscape change. *Landscape Ecology* 2, 111 – 133.
- Barredo, J.I., Demicheli, L., Lavelle, C., Kasanko, M., & McCormick, N. (2004). Modelling future urban scenarios in developing countries: An application case study in Lagos, Nigeria. *Environment and Planning B: Planning and Design*, 32, 65-84.
- Batty, M., & Longley, P. (1994). *Fractal cities*. London, England: Academic Press.
- Batty, M., (1997). Cellular automata and urban form: A primer. *Journal of the American Planning Association*, 63(2), 266-274.
- Batty, M. (2005). *Cities and complexity: Understanding cities with cellular automata, agent-based models, and fractals*. Cambridge, MA: MIT Press.
- Bellman, R. E (1961). *Adaptive control processes*. Princeton, NJ: Princeton University Press.
- Bernstein, J. (1995). *The urban challenge in national environmental strategies* (Environmental Management Series Paper no 012). Washington, DC: Environment Department, The World Bank.

- Bhardwaj, N., Langlois, R., Zhao, G., & Lu, H. (2005). Kernel-based machine learning protocol for predicting DNA-binding proteins. *Nucleic Acids Research*, 33, 6486 – 6493.
- Bivand, R., & Lucas, A. (2000). Integrating models and geographical information systems. In S. Openshaw & R. J. Abrahart (Eds.), *Geocomputation* (pp. 331-364). London, England: Taylor & Francis.
- Bradley, J.V., (1968). *Distribution-free statistical tests*. Upper Saddle River, NJ: Prentice-Hall.
- Braimoh, A. K., & Onishi, T. (2007). Spatial determinants of urban land use change in Lagos, Nigeria. *Land Use Policy*, 24(2), 502-515.
- Brown, C. T., & Witschey, W. R. T. (2003). The fractal geometry of ancient Maya settlement. *Journal of Archaeological Science*, 30, 1619-1632.
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1998). geographically weighted regression - modelling spatial non-stationarity. *The Statistician*, 47(3), 431-443.
- Burrough, P. A. (1986). *Principles of geographic information systems for land resources assessment*. Oxford, England: Oxford University Press.
- Burrough, P. A. & McDonnell R. A. (1998). *Principles of geographic information systems*. Oxford, England: Oxford University Press.
- Camps-Valls, G. & Bruzzone, L. (2005). Kernel-based techniques for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6), 1351-1362.
- Caruso, G. (2005). *Integrating urban economics and cellular automata to model periurbanisation: Spatial dynamics of residential choice in the presence of neighbourhood externalities* (Unpublished doctoral dissertation). Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
- Chauvin, Y., & Rumelhart, D. E. (Eds.) (1995). *Backpropagation: Theory, architectures and applications*. Hillsdale, NJ: Erlbaum.
- Clarke, K. C., Gaydos, L., & Hoppen, S. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B*, 24, 247 – 261.

- Clarke, K. C., & Gaydos, L. (1998). Long term urban growth prediction using a cellular automaton model and GIS. *International Journal of Geographical Information Science*, 12(7), 699-714.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Colwell, R. N. (1997). History and place of photographic interpretation. In W. R. Philipson (Ed.), *Manual of photographic interpretation* (pp. 3-47). Bethesda, MD: American Society for Photogrammetry and Remote Sensing.
- Cortes, C., & Vapnik, V. (1995). Support vector networks. *Machine Learning* 20(3), 273-297.
- Couclelis H. (1988). Of mice and men: what rodent populations can teach us about complex spatial dynamics. *Environment and Planning A*, 20, 99-109.
- Couclelis, H. (1997). From cellular automata to urban models: New principles for model development and implementation. *Environment and Planning B*, 24, 165-174.
- Couclelis, H. (1998). GeoComputation in context. In P. A. Longley, S. M. Brooks, R. McDonnell, & B. Macmillan (Eds.). In *Geocomputation: A Primer* (pp. 17-29). Chichester, England: John Wiley & Sons Ltd.
- Couclelis, H. (2002). Modelling frameworks, paradigms and approaches. In K. C. Clarke, B. E. Parks, & M. P. Crane(Eds). *Geographic information systems and environmental modelling* (pp. 36-50). Upper saddle River, NJ: Prentice Hall.
- Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines*. Cambridge, England: Cambridge University Press.
- Demirel, H. & Cetin, M. (2010). Modelling urban dynamics via cellular automata. In ISPRS Archive vol. xxxviii, Part 4-8-2-W9, *Core spatial databases – updating, maintenance, and services – for theory to practice*, Haifa, Israel.
- Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., et al. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, 30, 609 – 628.

- Drucker, H., Wu, D. H., & Vapnik, V. N. (1999). Support vector machines for spam categorization. *IEEE Transactions on Neural Networks*, 10(5), 1048 – 1054.
- Duch, W. & Diercksen, G.H.F. (1995). Feature space mapping as a universal adaptive system. *Computer Physics Communications*, 87 (3), 341-371.
- Elhers, M. (1997). Rectification and registration. In J. L. Star, J. E. Estes, & K. C. McGwire (Eds.), *Integration of geographic information systems and remote sensing* (pp. 13-36). Cambridge, England: Cambridge University Press.
- Elhers, M., Edwards, G., & Bedard, Y. (1989). Integration of remote sensing with geographic information system: A necessary evolution. *Photogrammetric Engineering and Remote Sensing*, 55 (11), 1619-1627.
- Erener, A., Sebnem, H., & Düzgün, B. (2010). Improvement of statistical landslide susceptibility mapping by using spatial and global regression methods in the case of More and Romsdal (Norway). *Landslides*, 7, 55 – 68.
- Farber, S. & Páez, A. (2007). A systematic investigation of cross-validation in GWR model estimation: Empirical Analysis and Monte Carlo simulations. *Journal of Geographical Systems*, 9(4), 371-396.
- Fkirin, M. A., Badwai, S. M., & Mohamed, S. A. (2009). Change detection using neural network with improvement factor in satellite images. *American Journal of Environmental Sciences*, 5 (6), 706-713.
- Fawcett, T. (2004). *ROC graphs: Notes and practical considerations for data mining researchers* (Technical report HPL-2003-4). Palo Alto, CA: HP Laboratories.
- Fletcher, R. (Ed.) (1987). *Practical methods of optimization*. New York, NY: John Wiley & Sons.
- Foody, G. M. (2004). Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogramm. Eng. Remote Sensing*, 70, 627 – 633.
- Fotheringham, A. S., Brunsdon, C. & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester, England: John Wiley & Sons Ltd.

- Fox, J.C., Ades, P. K., & Bi, H. (2001). Stochastic structure and individual-tree growth models. *For. Ecol. Manage.*, 154, 261 – 276.
- Fung, T., & LeDrew, E. (1988). The Determination of optimal threshold levels for changes detection using various accuracy indices. *Photogrammetric Engineering and Remote Sensing*, 54(10), 1449-1454.
- Gandy, M. (2006). Planning, anti-planning and the infrastructure crisis facing metropolitan Lagos. *Urban Studies*, 43(2), 371 – 396.
- Gardner, M., (1970). The fantastic combinations of John Conway's new solitaire game 'life.' *Scientific American*, 223(4), 120- 123.
- Gong, P. (1996). Integrated analysis of spatial data from multiple sources: Using evidential reasoning and artificial neural network techniques for geological mapping. *Photogrammetric Engineering & Remote Sensing*, 62(5), 513 – 523.
- Guan, Q., Wang, L., & Clarke, K.C. (2005). Artificial neural network based constrained CA model for simulating urban growth. *Cartography and Geographic Information Science*, 32(4), 369-380.
- Gunn, S. (1998). *Support vector machines for classification and regression* (Technical Report, ISIS). Southampton, England: Department of Electronics and Computer Science, University of Southampton.
- Herbert, D. T. & Thomas, C. J. (Eds.) (1997). *Cities in space: City as place*, (3rd ed.). London, England: David Fulton Publishers.
- Holland, J. H. (1998). *Emergence: From chaos to order*. Reading, MA: Addison-Wesley.
- Huang, B., Xie, C., & Tay, R. (2010). Support vector machines for urban growth modelling. *Geoinformatica*, 14, 83 – 99.
- Ivanciuc, O. (2007). Applications of support vector machines in Chemistry. In K. Lipkowitz, & T. Cundari (Eds.), *Reviews in computational chemistry* (pp. 291-400). Texas, TX: Wiley-VCH, John Wiley & sons, Inc.
- Jensen, J. R. (1996). Issues involving the creation of digital elevation models and terrain correction orthoimagery using soft-copy photogrammetry. In C. Greve (Eds.), *Digital*

- photogrammetry: An addendum to the manual of photogrammetry* (pp. 167-179). Bethesda, MD: American Society of Photogrammetry and Remote Sensing.
- Jimenez, L.O. & Landgrebe, D.A. (1997). Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Transactions on Systems, Man and Cybernetics*, 28(1), 39 – 54.
- Joachims, T. (1999). Making large-scale SVM learning practical. In B. SchÖlkopf, C. J. C. Burges, & A. J. Smola (Eds.), *Advances in kernel methods - support vector learning* (pp. 169-184). Cambridge, MA: MIT Press.
- Kimsey, M. J., Moore, J., & McDaniel, P. (2008). A geographically weighted regression analysis of Douglas-fir site index in north central Idaho. *Forest Science*, 54, 356 – 366.
- Kocabas, V., & Dragicevic, S., (2007). Enhancing a GIS cellular automata model of land use change: Bayesian networks, influence diagrams and causality. *Transactions in GIS* 11(5), 681-702.
- Kupfer, J. A., & Farris, C. A. (2007). Incorporating spatial nonstationarity of regression coefficients into predictive vegetation models. *Landscape Ecology*, 22, 837 – 852.
- Leung, Y., Mei, C., & Zhang, W. (2000). Testing for spatial autocorrelation among the residuals of the geographically weighted regression. *Environment and Planning A*, 32(5), 871-890.
- Li, X. and Magill, W. (2001). Modelling fire spread under environmental influence using a cellular automaton approach. *Complexity International*, 8, 1-14.
- Li, X., & Yeh, A. G. (2002). Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, 16(4), 323-343.
- Li, X., & Yeh, A. G. (2004). Data mining of cellular automata transition rules. *International Journal of Geographical Information Science*, 18(8), 723-744.
- Lin, C. F. & Wang, S. D. (2002). Fuzzy support vector machines. *IEEE Transactions on Neural Networks*, 13, 464-471.

- Liu, Y. (2001). *Modelling urban development with geographical information systems and cellular Automata: A case study of Sydney since 1971* (Unpublished doctoral dissertation). University of Queensland, Queensland, Australia.
- Liu, X., & Li, X. (2006). Retrieving CA nonlinear transition rule from high-dimensional feature space. *Journal of Geographical Science*, 61(6), 663-672.
- Liu, X., Li, X., Ai, B., Wu, S., & Liu, T., (2007). *Kernel-based cellular automata for urban simulation*. Presented at the 'Third International Conference on Natural Computation, Haikou, China, August 24-27.
- Lo, C. P., & Hu, Z. (2007). Modelling urban growth in Atlanta using logistic regression. *Computers, Environment, and Urban Systems*, 31, 667-688.
- Lo, C. P., & Shipman, R. L. (1990). A GIS approach to land use change dynamics detection. *Photogrammetric Engineering and Remote Sensing*, 56 (11), 1483-1491.
- Lo, C.P. & Yeung, A.K.W. (Eds.) (2007). *Concepts and techniques of geographic information systems* (2<sup>nd</sup> ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Longley, P. & Batty, M. (Eds.). (2003). *Advanced spatial analysis: The CASA book of GIS*. Redlands, CA: ESRI Press.
- Lunetta, R.S., Congalton, R.G., Fenstermaker, L.K., Jensen, J.R., & Tinney, L.R. (1991). Remote sensing and geographic information system data integration: Error sources and research issues. *Photogrammetric Engineering and remote sensing*, 57(6), 677-687.
- Ma, Z. & Redmond, R. L. (1995). Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing* 61, 435-439.
- Macmillan, W. D. (1998). Epilogue. In P. A. Longley, S. M. Brooks, R. McDonnell, & W. Macmillan (Eds.), *Geocomputation: A Primer* (pp. 257-264). Chichester, England: John Wiley & Sons, Inc.
- Macy, M. W. & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modelling. *Annual Review of Sociology*, 28, 143-166.
- Maerivoet, S. & De Moor, B. (2005). Cellular automata models of road traffic. *Physics Reports*, 419 (1), 1-64.

- Makse, H. A., Andrade, J. S., & Batty, M. (1998). Modelling urban growth patterns with correlated percolation. *Physics Review E*, 58(6), 7054-7062.
- Mandelas, E., Hatzichristos, T., & Prastacos, P. (2007). *A fuzzy cellular automata based shell for modelling urban growth – A pilot application in Mesogia area*. Paper presented at the 10th AGILE International Conference on Geographic Information Science, Aalborg University, Denmark.
- Mandelas, E., Prastacos, P., & Hatzichristos, T. (2008). *Modelling urban growth using fuzzy cellular automata*. Paper presented at the 11th AGILE International Conference on Geographic Information Science, University of Girona, Girona, Spain.
- Mandelbrot, B. B. (1967). How long is the coast of Britain? Statistical self similarity and fractal dimension. *Science*, 156, 636-638.
- McMillen, D. P. (1989). An empirical model of urban fringe land use. *Land Economics*, 65, 138 – 145.
- Menard S. (1995). *Applied logistic regression analysis*. Sage University paper series on quantitative applications in the Social Sciences (Series no. 07-106). Thousand Oaks, CA: Sage.
- Mennis, J. (2006). Mapping the results of geographically weighted regression. *The Cartographic Journal*, 43(2), 171-179.
- Mercer, J. (1909). Functions of positive and negative type and their connection with the theory of integral equations. *Transactions of the London Philosophical Society A*, 209, 415-446.
- Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society Series B*, 10, 243-251.
- Noresah M. S. & Ruslan, R. (2009). *Modelling urban spatial structure using geographically weighted regression*. Paper presented at the 18th World IMACS / MODSIM Congress, Cairns, Australia, July 13<sup>th</sup> -17<sup>th</sup>.
- Oduwaye, L. (2009). Challenges of sustainable physical planning and development in metropolitan Lagos. *Journal of Sustainable Development*, 2(1), 159-171.



- Okwuashi, O., McConchie, J., Nwilo, P., & Eyo, E. (2008). *Fuzzy support machine constrained GIS-based cellular automata for urban change simulation*. Paper presented at the 24th New Zealand Geographical Conference: Inequality, Sustainability, Policy; Across the Divide, Wellington, New Zealand, July 2<sup>nd</sup> – 5<sup>th</sup>.
- Okwuashi, O., McConchie, J., Nwilo, P., & Eyo, E. (2009a). Enhancing a GIS cellular automata model of land use change using support vector machine. *Proceedings of the 17th International Conference on Geoinformatics*, August 12<sup>th</sup> – 14<sup>th</sup>, Fairfax, VA: IEEE Xplore.
- Okwuashi, O., McConchie, J., Nwilo, P., & Eyo, E. (2009b). Stochastic GIS cellular automata for land use change simulation: Application of a kernel based model. In B. G. Lees & S. W. Laffan (Eds.), *10<sup>th</sup> International Conference on Geocomputation* (pp. 203-209). UNSW, Sydney, November 30<sup>th</sup> – 2<sup>nd</sup> December 2<sup>nd</sup>.
- Olden, J.D., & Jackson, D.A. (2001). Fish-habitat relationships in lakes: Gaining predictive and explanatory insight by using artificial neural networks. *Trans. Am. Fish. Soc.* 130, 878 – 897.
- Osborne, P. E., Foody, G. M., & Suárez-Seoane, S. (2007). Non-stationarity and local approaches to modelling the distributions of wildlife. *Diversity and Distributions*, 13, 313 – 323.
- O'Sullivan, D. (2001). Exploring spatial process dynamics using irregular cellular automaton models. *Geographical Analysis*, 33 (1), 1-18.
- O'Sullivan, D. (2004). Complexity science and human geography. *Transactions of the Institute of British Geographers*, 29(3), 282-295.
- O'Sullivan, D. & Torrens, P. M. (2000). Cellular models of urban systems. In S. Bandini & T. Worsch (Eds.), *Theoretical issues on cellular automata* (pp. 108-117). London, England: Springer-Verlag.
- Overmars, K. P.; Koning, G. H., & Veldkamp, A. (2003). Spatial autocorrelation in multiscale land use models. *Ecological Modelling*, 164, 257-270.
- Ozah, A. P., Adesina, F.A., & Dami, A. (2010). A Deterministic Cellular Automata Model for Simulating Rural Land Use Dynamics: A Case Study of Lake Chad Basin. In ISPRS

- Archive vol. xxxviii, Part 4-8-2-W9, *Core spatial databases – updating, maintenance, and services – for theory to practice*, Haifa, Israel.
- Park, S., & Wagner, D. F. (1997). Incorporating cellular automata simulators as analytical engines in GIS. *Transactions in GIS*, 2(3), 213-231.
- Parker, D. C. (2005). Integration of geographic information systems and agent-based models of land use: Challenges and prospects. In D. J. Maguire, M. F. Goodchild, & M. Batty (Eds.), *GIS, spatial analysis and modelling* (pp. 403-422), Redlands, CA: ESRI Press.
- Platt, J. (1999). Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In A. J. Smola, P. Bartlett, B. Schölkopf, D. Schuurmans (Eds.), *Advances in large margin classifiers* (pp. 61 – 74). Cambridge, MA: MIT Press.
- Platt, R. V. (2004). Global and local analysis of fragmentation in a mountain region of Colorado. *Agriculture Ecosystems & Environment*, 101(2-3), 207-218.
- Pohlmann, J. T., & Dennis, W. L. (2003). A comparison of ordinary least squares and logistic regression. *Ohio Journal of Science*, 103 (5), 118-125.
- Polhill, J. G., Pignotti E., Gotts, N. M., Edwards, P., & Preece, A. (2007). A semantic grid service for experimentation with an agent-based model of land-use change. *Journal of Artificial Societies and Social Simulation*, 10(2), 2.
- Putman, S. H. (1992). *Integrated urban models 2: New research and applications of optimization and dynamics*. London, England: Pion Press.
- Putman, S. H. (1993). *Integrated urban models: Policy analysis of transportation and land use*. London, England: Pion Press.
- Putman, S. H. (1995). *EMPAL and DRAM Location and Land Use Models: an Overview*. Paper distributed at TMIP Land Use Modelling Conference, Dallas, TX.
- Rees, P., & Turton, I. (1998). *Investigation of the effects of input uncertainty on population forecasting*. Paper presented at the 3rd International Conference on GeoComputation, Bristol, UK, September 17<sup>th</sup> – 19<sup>th</sup>.

- Richards, J. F. (1990). Land transformation. In: B. L. Turner, W.C. Clark, R.W. Kates, J.F. Richards, J.T. Mathews, & W.B. Meyer (Eds.), *The earth as transformed by human action* (pp. 163-178). New York, NY: Cambridge Univ. Press.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: explorations in the microstructure of cognition* (pp. 318-362). Cambridge, MA: MIT Press.
- Shen, G. (1997). A fractal dimension analysis of urban transportation networks. *Geographical and Environmental Modelling*, 1(2), 221-236.
- Shi, H., Laurent, E. J., LeBouton, J., Racevskis, L., Hall, K. R., Donovan, M., Doepker, R. V., et al. (2006). Local spatial modelling of white-tailed deer distribution. *Ecological Modelling*, 190, 171 – 189.
- Silva, E. A., & Clarke, K. C. (2002). Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems*, 26(6), 525-552.
- Sietchiping, R. (2004). *A geographic information systems and cellular automata-based model of informal settlement growth* (Unpublished doctoral dissertation). University of Melbourne, Melbourne, Australia.
- Simonett, D. S. (1983). *The development and principles of remote sensing, in Manual of Remote Sensing*. In R. N. Colwell (Ed.), *Manual of remote sensing* (pp. 1-35). Falls Church, VA: American Society of Photogrammetry.
- Tang, B.-S., & Choy, L. H.-T. (2000). Modelling planning control decisions: a logistic regression analysis on office development applications in urban Kowloon, Hong Kong. *Cities*, 17(3), 219 – 225.
- Thapa, R. & Murayama, Y. (2009). Land use change factors in Kathmandu Valley: A GWR approach. In B. G. Lees & S. W. Laffan (Eds.), *10<sup>th</sup> International Conference on Geocomputation* (pp. 255-260). UNSW, Sydney, November 30<sup>th</sup> – December 2<sup>nd</sup>.
- Tobler W., (1970). A computer movie simulating urban growth in the Detroit Region. *Economic Geography*, 46(2), 234-240.

- Torrens, P. M., & Benenson, I. (2005). Geographic automata systems. *International Journal of Geographical Information Science*, 19(4), 385-412.
- Torrens, P.M. & O'Sullivan, D. (2001). Cellular automata and urban simulation: Where do we go from here? *Environment and Planning B*, 28, 163-168.
- Tsai, C. & Chen, J. (2004). Significance analysis of ROC indices for comparing diagnostic markers: Applications to gene microarray data. *Journal of Biopharmaceutical Statistics*, 14(4), 985-1003.
- Turner II, B. L., Clark, W. C., Kates, R. W., Richards, J. F., Mathews, J. T., & Meyer, W. B. (1990). *The earth as transformed by human action: Global and regional changes in the biosphere over the past 300 years*. Cambridge, England: Cambridge University Press.
- Ulam S. (1952). Random processes and transformations. *Proceedings of the International Congress on Mathematics*, (2), 264-275. Providence, RI: American Mathematical Society.
- United Nations (2002). *World urbanization prospects: The 2001 Revision*. New York, NY: United Nations.
- Vapnik, V. (1995). *The nature of statistical learning theory*. New York, NY: Springer-Verlag.
- Vapnik, V.N. (2000). *The nature of statistical learning theory*. New York, NY: Springer-Verlag.
- Veldkamp, A., & Fresco, L.O. (1996). CLUE-CR: An integrated multi-scale model to simulate land use change scenarios in Costa Rica. *Ecological Modelling*, 91, 231-248.
- Verburg, P. H., Schot, P., Dijst, M., & Veldkamp, A. (2004). Land use change modelling: current practice and research priorities. *Geojournal*, 61(4), 309 – 324.
- Verburg, P.H., Veldkamp, A., de Koning, G.H., Kok, K. & Bouma, J. (1999). A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. *Ecological Modelling*, 116, 45-61.
- von Neumann, J. (1966). *The theory of self reproducing automata*. Edited by A. W. Burks, Urbana, IL: Univ. of Illinois Press.
- Waddell, P. (2002). UrbanSim: Modelling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association*, 68(3), 297 – 314.

- Wagner, D. F. (1997). Cellular automata and geographic information systems. *Environment and Planning B*, 24, 219-234.
- Waldrop, M. M. (1992). *Complexity: The emerging science at the edge of order and chaos*. New York, NY: Touchstone.
- Wang, F., 1994, The use of artificial neural networks in a geographical information systems for agricultural land suitability assessment. *Environment and Planning A*, 26, 265 – 284.
- Webster, C., & Wu, F. (2001). Coarse, spatial pricing and self-organising cities. *Urban Studies*, 38, 2037- 2054.
- Wheeler, D., & Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7, 161 – 187.
- White, R., & Engelen, G. (1993). Fractal urban land use patterns: A cellular automata approach. *Environment and Planning A*. 25, 1175-1199.
- White, R., & Engelen, G. (2000). High resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems*, 24, 383-440.
- Windle, M. J. S., Rose, G. A., Devillers, R., & Fortin, M.J. (2010). Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): An example from the Northwest Atlantic. *ICES Journal of Marine Science*, 67, 145 – 154.
- Wolfram, S. (1983). Statistical mechanics of cellular automata. *Review of Modern Physics*, 55, 601 – 644.
- Wolfram, S. (2002). *A new kind of science*. Champaign, IL: Wolfram Media.
- Wu, F. (1996). A linguistic cellular automata simulation approach for sustainable land development in a fast growing region. *Computers, Environment, and Urban Systems*, 20, 367-387.
- Wu, F. (1998). Simulating urban encroachment on rural land with fuzzy-logic-controlled cellular automata in a geographical information system. *Journal of Environmental Management*, 53(4), 293-308.

- Wu, F. (2002). Calibration of stochastic cellular automata: The application to rural urban land conversions. *International Journal of Geographical Information Science*, 16(8), 795-818.
- Xie, C. (2006). *Support vector machines for land use change modelling* (Unpublished master's thesis). University of Calgary, Calgary, Canada.
- Yeh, A. G., & Li, X. (2001). A constrained CA model for the simulation and planning of sustainable urban forms by using GIS. *Environment and Planning B: Planning and Design*, 28, 733 - 753.
- Zadeh L.A. (1962). From circuit theory to system theory. *Proceedings of Institution of Radio Engineers*, 50, 856-865.
- Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, 8 (3), 338-353.
- Zadeh, L. A. (1971). Towards a theory of fuzzy systems. In R. E. Kalman & N. DeClaris (Eds.), *Aspects of network and systems theory* (pp. 209-245). New York, NY: Holt, Rinehart and Winston.
- Zadeh, L.A. (1984). Making computers think like people. *IEEE Spectrum*, 8, 26-32.
- Zhang, L., & Gove, J. H. (2005). Spatial assessment of model errors from four regression techniques. *Forest Science*, 51(4), 334 – 346.
- Zhou, K., & Kang, Y. (2005). *MATLAB neural network model and its simulation programming*. Beijing, China: Tsinghua University Press.

