Cyprus University of Technology
Faculty of Engineering and Technology

Doctoral Dissertation

Critical Investigation of Novel Computational Techniques for Automated Valuations of Real Estate Properties in Cyprus

Ph.D. Thesis: Thomas Dimopoulos

Limassol, April 2020

CYPRUS UNIVERSITY OF TECHNOLOGY

FACULTY OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF CIVIL ENGINEERING AND GEOMATICS

Doctoral Dissertation

Critical Investigation of Novel Computational Techniques for Automated Valuations of Real Estate Properties in Cyprus

Ph.D. Thesis: Thomas Dimopoulos

Limassol, April, 2020

Approval Form

Doctoral Dissertation

Critical Investigation of Novel Computational Techniques for Automated Valuations of Real Estate Properties in Cyprus

Presented by Thomas Dimopoulos

Examination Committee:

Research Supervisor: **Diofantos Hadjimitsis**, Professor, Department of Civil Engineering and Geomatics, Faculty of Engineering and Technology, Cyprus University of Technology.

Member of the examination committee: **Phaedon C. Kyriakidis**, Professor, Dean of Faculty Engineering and Technology, Department of Civil Engineering and Geomatics, Faculty of Engineering and Technology, Cyprus University of Technology.

Member of the examination committee: **Efi Dimopoulou**, Professor, School of rural and Surveying Engineering, National Technical University of Athens.

Doctoral Dissertation

Critical Investigation of Novel Computational Techniques for Automated Valuations of Real Estate Properties in Cyprus

Presented by Thomas Dimopoulos

Supervising Committee:

Research Supervisor: **Diofantos Hadjimitsis**, Professor, Department of Civil Engineering and Geomatics, Faculty of Engineering and Technology, Cyprus University of Technology.

Member of the Supervising committee: **Evangelos Akylas**, Associate Professor, Department of Civil Engineering and Geomatics, Faculty of Engineering and Technology, Cyprus University of Technology.

Member of the Supervising committee: **Nikolaos Karanikolas**, Associate Professor, School of Spatial Planning and Development, Aristotle University.

Cyprus University of Technology

Limassol, April, 2020

Copyrights

Acknowledgments.

Firstly, I would like to express my sincere gratitude to my advisor Prof. Diofantos Hadjimitsis for the continuous support throughout my Ph.D. study and its related research. His patience, motivation, his immense knowledge and most importantly, his straightforward honesty, has guided and helped me throughout the research and writing processes of this Ph.D. thesis. I could not imagine having a better advisor and mentor for my Ph.D. study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Phaedon Kyriakidis and Prof. Efi Dimopoulou, for their insightful comments and encouragement, and also for always being available. Assoc. Prof. Nikolaos Karanikolas has had a significant role, as he was the first one who introduced me to the subject of real estate valuation during my studies at Aristotle University, and together with Assoc. Prof. Akylas, have contributed as being members of the supervising committee.

I will forever be thankful to Dr. Nikolaos Bakas, not only for co-authoring research papers but also for mentoring me in the fields of AI and Machine Learning, for supporting and believing in me and for insisting, more than anyone else, that I complete this thesis. Mr. Varnavas Pashoulis has also been very supportive and probably the one who inspired me to choose the specific study area.

My sincere thanks also go to Mr. Yorkas, Mr. Moulas, Dr. Tyralis, Ms. Vertsoni, Mr. Antoniou and Mr. Protopapas for co-authoring papers, for their contributions and for the many inspiring discussions. I have a deep appreciation for Dr. Tassos Labropoulos for also being very supportive and for the stimulating discussions and the sleepless nights of research throughout the years.

I consider Mr. Nick Nunnington and Dr. Petros Sivitanides as key people in my academic career as they both broadened my horizons and inspired me. I know that I can always ask them for advice and opinions on all property-related issues. I also thank, Prof. Grover who taught me during my MSc year at Oxford Brookes University, and who offered his assistance and advice, and my friend Dr. Tsakonas for his bibliography related assistance.

I am also grateful to Ms. Cleovoulou and Ms. Chatzimina whose help was invaluable during the last critical month. Mr. Mamas, Ms. Kyprianou and the rest of the AXIA team (Mr. Kyriacou, Mr. Panagiotou and Ms. Antoni) have been very supportive throughout the years and have worked harder when necessary so that I could have the extra time I needed to complete this thesis.

I would also like to thank my students, especially those who have been eager to learn and who have provided me with motivation to become better (there are too many of you to list here, but you know who you are!).

I owe a lot to my family. My parents and my sister for supporting me spiritually throughout the writing process of this thesis, and in my life in general. My wife's family for also being very supportive and encouraging both personally and professionally. Last but not least, the key person though who deserves the most thanks is my soul-mate since 2009, and wife since 2014, Natasa. She is continuously by my side, supporting my decisions, and always tranquil. She has given me my two wonderful boys, Michael and Harry and she is currently pregnant with my third child (presumably boy as well).

The author would like to thank the support of the Remote Sensing and Geo-Environment Lab (ERC group) at the Department of Civil Engineering and Geomatics of the Cyprus University of Technology and the 'EXCELSIOR' H2020 Teaming Project. The Remote Sensing and Geo-Environment Lab (ERC group) at the Department of Civil Engineering and Geomatics is on the way to be upgraded to ERATOSTHENES Centre of Excellence (ECoE) through 'EXCELSIOR' H 2020 Widespread Teaming project (www.excelsior2020.eu). Indeed, this thesis is under the auspices of all the activities of the 'ERATOSTHENES: Excellence Research Centre for Earth Surveillance and Space-Based Monitoring of the Environment'- 'EXCELSIOR' project that has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 857510 and from the Government of the Republic of Cyprus through the Directorate General for the European Programmes, Coordination and Development.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857510.

This project has received funding from the Government of the Republic of Cyprus through the Directorate General of the European's Programmes, Coordination and Development.

ABSTRACT

Keywords: CAMA, AVM, Mass Appraisals, Valuation, Property taxation

Mass appraisals for valuation purposes using automated systems have gained a lot of traction in recent years, a fact which is highlighted when viewing the large amount of corresponding literature that has become available over the past decade. The main valuation governing bodies (IAAO, RICS, IVS, TEGOVA, national authorities, etc.) have all produced papers and articles referring to the models and systems that are used for mass valuations (Computer Assisted Mass Appraisals - CAMA, Automated Valuation Models - AVM, etc.) and how their application could be revolutionary within the sector. While automated systems are already being used in many countries and jurisdictions for taxation purposes, the demand for mass appraisals is growing as a result of the financial globalization process. Issues regarding ethics, licensing and responsibility of the valuations produced by automated systems remain pending but are being addressed constantly, as well as their importance and impact on the broader environment of valuation practice and the Real Estate industry.

The aim of this PhD thesis is to provide a rigorous and accurate analysis of the mass appraisal procedure, to highlight the relevant techniques and methodologies, and to propose innovative methods to advance the currently used mass appraisal system in Cyprus through worked case studies based also from the literature review findings. A global history of mass appraisals, as well as definitions, methodologies and models' specifications, calibration and adjustments are presented and the most common applications of mass appraisals are discussed. The models implemented by the Cyprus Department of Lands and Surveys (DLS) for taxation purposes are analyzed and the strengths and weaknesses of current systems are presented and assessed. The author uses an enhanced apartments' database to analyze the dependence on their deviation on the other parameters influencing a property's value (covered area, location, etc.). The results of the case studies that were carried out in Nicosia (Cyprus) and Thessaloniki (Greece) using Geographically Weighted Regression, Ordinary Least Squares, Random Forests as well as other mathematical techniques are presented, scrutinized and interpreted. The author provides novel recommendations for the improvement of the models and how their application could be implemented in the wider market. Finally, he provides a critical

judgment of the models' accuracy, by utilizing both his significant professional experience (with more than 15,000 valuations conducted throughout a 15-year career) on specific test cases and real valuation practice, with a focus on outliers and observations with high errors.

The main outcome of this Ph.D. is its contribution to the appropriateness of utilization of automated systems in the valuation procedure and in the broader property valuation environment based on the critical evaluation of the existing techniques and their implementation within the Cyprus region. Although the AMVs have many advantages and can be used in several sectors, they also present limitations on the real-world application. However, the use of AVMs can improve the quality of the valuation precision and lead to a higher achieved accuracy ratio per valuation, which could, in turn, create higher profits for any valuer, stakeholder and to the broader industry as well. In conclusion, mass appraisals are cost and time effective and a positive contribution to the sustainability of the broader economic and financial environment.

ΠΕΡΙΛΗΨΗ

Λέξεις-κλειδιά: CAMA, AVM, Μαζικές Εκτιμήσεις, Εκτίμηση, Φορολογία Ακινήτων

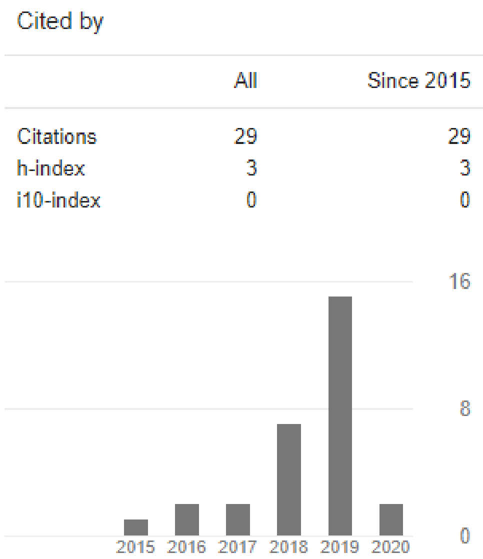
Οι μαζικές εκτιμήσεις με τη χρήση αυτοματοποιημένων μοντέλων, αποτελούν πόλο ερευνητικού ενδιαφέροντος τα τελευταία έτη, γεγονός το οποίο επιβεβαιώνεται μέσα από το μεγάλο όγκο σχετικής βιβλιογραφίας και των εργασιών που έχουν δημοσιευτεί τα τελευταία χρόνια. Οι κύριοι εκτιμητικοί φορείς (IAAO, RICS, IVS, TEGOVA, Εθνικές Υπηρεσίες, κ.λπ.) έχουν διεκπεραιώσει μελέτες και άρθρα σχετικά με τα μοντέλα και τα συστήματα τα οποία χρησιμοποιούνται στις μαζικές εκτιμήσεις (Computer Assisted Mass Appraisals - CAMA, Automated Valuation Models - AVM, κ.λπ.) και τον τρόπο με τον οποίο η εφαρμογή τους θα μπορούσε να αποφέρει σημαντικές στις ως σήμερα γνωστές διαδικασίες. Ενώ τα αυτοματοποιημένα συστήματα χρησιμοποιούνται ήδη σε πολλές χώρες και για φορολογικούς σκοπούς, η ζήτηση για μαζικές εκτιμήσεις αυξάνεται ως αποτέλεσμα της διαδικασίας της οικονομικής παγκοσμιοποίησης. Τα ζητήματα που αφορούν την ηθική, την αδειοδότηση και την ευθύνη των εκτιμήσεων που παράγονται από τα αυτοματοποιημένα συστήματα, καθώς επίσης και η σημασία και το αντίκτυπο που έχουν στο ευρύτερο εκτιμητικό περιβάλλον και στη βιομηχανία των ακινήτων, προβληματίζουν την εκτιμητική κοινότητα. Τα αυτοματοποιημένα μοντέλα, ωστόσο, προσαρμόζονται και βελτιώνονται συνεχώς με μεγάλη επιτυχία.

Σκοπός της παρούσας διδακτορικής διατριβής είναι να παρέχει μια αντικειμενική ανάλυση της διαδικασίας των μαζικών εκτιμήσεων, να επισημάνει τις σχετικές τεχνικές και μεθοδολογίες και να προτείνει καινοτόμες μεθόδους με σκοπό τη βελτίωση του συστήματος μαζικής εκτίμησης που χρησιμοποιείται σήμερα στην Κύπρο, μέσω μελετών που βασίζονται επίσης στα ευρήματα από τη βιβλιογραφική ανασκόπηση. Παρουσιάζεται η ιστορία των μαζικών εκτιμήσεων, καθώς επίσης και ορισμοί, μεθοδολογίες, προδιαγραφές των μοντέλων, η βαθμονόμηση και οι προσαρμογές τους, ενώ παρατίθενται και οι πιο γνωστές εφαρμογές των μαζικών εκτιμήσεων σε παγκόσμια κλίμακα. Αναλύονται τα μοντέλα που εφαρμόστηκαν από το Τμήμα Κτηματολογίου και Χωρομετρίας της Κύπρου για σκοπούς φορολογίας έως σήμερα, ενώ εντοπίζονται και αξιολογούνται παράλληλα τα πλεονεκτήματα και τις αδυναμίες τους. Ο συγγραφέας χρησιμοποιεί μια βάση δεδομένων για διαμερίσματα και αναλύσει τις παραμέτρους που

επηρεάζουν την αξία ενός ακινήτου (εμβαδό κλειστών χώρων, τοποθεσία κ.λπ.) καθώς και τις μεταξύ τους συσχετίσεις. Παρουσιάζονται, αναλύονται και ερμηνεύονται τα αποτελέσματα τα οποία προέκυψαν μέσα από μελέτες που πραγματοποιήθηκαν για την περιοχή της Λευκωσίας (Κύπρος) και της Θεσσαλονίκης (Ελλάδα) με τη χρήση της Γεωγραφικά Σταθμισμένης Παλινδρόμησης (GWR), της Μεθόδου των Ελαχίστων Τετραγώνων (OLS), της τεχνικής μηχανικής μάθησης Random Forest, καθώς επίσης και άλλων υπολογιστικών τεχνικών και μοντέλων. Ο συγγραφέας παρέχει καινοτόμες προτάσεις για τη βελτίωση των μοντέλων και τον τρόπο εφαρμογής τους στην ευρύτερη αγορά. Τέλος, ο συγγραφέας πραγματοποιεί κριτική σχετικά με την ακρίβεια των μοντέλων, αξιοποιώντας τα με βάση τη σημαντική επαγγελματική του πείρα (η οποία περιλαμβάνει περισσότερες από 15,000 εκτιμήσεις που διεξήχθησαν σε μια 15-ετή καριέρα) δίνοντας έμφαση στις ακραίες τιμές και τα μεγάλα σφάλματα.

Το σημαντικότερο αποτέλεσμα της παρούσας διδακτορικής διατριβής είναι η συμβολή της στην εφαρμογή και τη χρήση των αυτοματοποιημένων συστημάτων εντός της εκτιμητικής διαδικασίας αλλά και στο ευρύτερο οικονομικό περιβάλλον και βασίζεται στην κριτική αξιολόγηση των υφιστάμενων τεχνικών και την εφαρμογή τους στην Κύπρο. Παρόλο που τα Αυτοματοποιημένα Εκτιμητικά Μοντέλα έχουν σημαντικά πλεονεκτήματα και μπορούν να χρησιμοποιηθούν από διάφορους τομείς, παρουσιάζουν ωστόσο περιορισμούς στη δια ταύτα εφαρμογή τους. Παρ' όλα αυτά, η χρήση των AVMs μπορεί να βελτιώσει την ποιότητα της ακρίβειας των εκτιμήσεων και να οδηγήσει σε υψηλότερη αναλογία ακρίβειας ανά εκτίμηση, γεγονός που θα μπορούσε, με τη σειρά του, να επιφέρει σημαντικά οφέλη στους εκτιμητές, τους άμεσα ενδιαφερόμενους, αλλά και γενικότερα στην ευρύτερη βιομηχανία των ακινήτων. Συμπερασματικά, οι μαζικές εκτιμήσεις πλεονεκτούν ως προς το κόστος και τον χρόνο που απαιτείται, και συνεισφέρουν θετικά στη βιωσιμότητα του ευρύτερου οικονομικού και χρηματοοικονομικού περιβάλλοντος.

Pre- Ph.D. Publications



Journal publications: [1]–[6]

Chapter in Books: [7]

Peer Reviewed Confer. Proceedings: [8]–[14]

Peer Reviewed Confer. Presentations: [15]- [16]

Others: [17]-[18]

Invited Speaker: [19]-[30]

<u>Peer reviewed Journals</u>

1. Dimopoulos, T., & Bakas, N. (2019). Sensitivity analysis of machine learning models for the mass appraisal of real estate. Case study of residential units in Nicosia, Cyprus. *Remote Sensing*, *11*(24). https://doi.org/10.3390/rs11243047
2. Dimopoulos, T., Tyralis, H., Bakas, N. P., & Hadjimitsis, D. (2018). Accuracy measurement of Random Forests and Linear Regression for mass appraisal models that estimate the prices of residential apartments in Nicosia, Cyprus. *Advances in Geosciences*, *45*, 377–382. https://doi.org/10.5194/adgeo-45-377-2018
3. Dimopoulos, T., & Moulas, A. (2016). A Proposal of a Mass Appraisal System in Greece with CAMA System: Evaluating GWR and MRA techniques in Thessaloniki Municipality. *Open Geosciences*, *8*(1), 675–693. https://doi.org/10.1515/geo-2016-0064
4. Dimopoulos, T., & Bakas, N. (2019). An artificial intelligence algorithm analyzing 30 years of research in mass appraisals. *RELAND: International Journal of Real Estate & Land Planning*, *2*, 10–27.
5. Antoniou, D., & Dimopoulos, T. (2018). Southern Orientation and Natural Cross Ventilation: Mind the Gap (s) What Clients, Valuers, Realtors and Architects Believe. *RELAND: International Journal of Real Estate & Land Planning*, *1*, 426–439.
6. Dimopoulos, T. (2015). Theories and philosophy of property taxation. *Pandemos.Panteion*, *79*. Library.panteion.gr/pandemos

<u>Chapters in Books</u>

7. Baum, A., Mackmin, D., Nunnington, N., (2018). *The Income Approach to Property Valuation.*, 7th edition, pages 345-350, Routledge, https://doi.org/10.4324/9781315637099-8

Peer Reviewed Conference Proceedings:

8. Dimopoulos, T., & Bakas, N. (2019). *Artificial intelligence for mass appraisals of residential properties in Nicosia: mathematical modelling and algorithmic implementation*. *June 2019*, 69. https://doi.org/10.1117/12.2538430
9. Dimopoulos, T., & Yiorkas, C. (2017). Implementing GIS in real estate price prediction and mass valuation: the case study of Nicosia District, *Fifth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2017)* (Issue September 2017, p. 72). SPIE. https://doi.org/10.1117/12.2280255
10. Θ. Δημόπουλος. (2016). Εισήγηση για αποτελεσματικότερη φορολόγηση ακινήτων στην Κύπρο μέσω της Νέας Γενικής Εκτίμησης, Μαζικές Εκτιμήσεις & η χωρική διάσταση της ακίνητης περιουσίας. *Ελληνική Εταιρεία Γεωγραφικών Συστημάτων Πληροφοριών - Ελληνικό Ινστιτούτο Εκτιμητικής*.
11. Dimopoulos, T., & Varnavas, P. (2015). The General Valuation 2013 in Cyprus: Experiences and Good Practices. *FIG-OICRF*.
12. Dimopoulos, T., Labropoulos, T., & Hadjimitsis, D. G. (2015). Introducing local property tax for fiscal decentralization and local authority autonomy. *Third International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2015)*, *9535*, 953509.
13. Πασιουλής Βαρνάβας, & Δημόπουλος Θωμάς. (2016). Μαζικές Εκτιμήσεις και Νέα Γενική Εκτίμηση στην Κύπρο σε τιμές 01.01.2013. *Ελληνική Εταιρεία Γεωγραφικών Συστημάτων Πληροφοριών - Ελληνικό Ινστιτούτο Εκτιμητικής*.
14. Dimopoulos, T., Labropoulos, T., & Hadjimitsis, D. G. (2014). Comparative analysis of property taxation policies within Greece and Cyprus evaluating the use of GIS, CAMA, and remote sensing techniques. *Proceedings of SPIE - The International Society for Optical Engineering*, *9229*. https://doi.org/10.1117/12.2070457

Peer Reviewed Conference Presentations

15. Dimopoulos, T., Bakas, N. P., & Hadjimitsis, D. G. (2017). Assessment of the comparable evidence from RE transactions in Cyprus; Numerical investigation of the deviation between Declared and Accepted price. *Fifth International Conference on Remote Sensing and Geoinformation*.
16. Dimopoulos, T., & Bakas, N. P. (2016). Introducing Neural Network Techniques on property mass appraisal. Case study of 3,500 apartments comparable evidence in Nicosia. *Fourth International Conference on Remote Sensing and Geoinformation*.

Others

17. Dimopoulos, T. (2015). Overview of property taxation systems in Europe. *Working Paper - Neapolis University Pafos*. http://hdl.handle.net/11728/7097
18. Bank Lending Valuations and Mortgage Lending Value, RICS guidance note, 1st edition, 2018, (RICS working group), ISBN 978 1 78321 229 3

Invited Speaker

19. T. Dimopoulos (2019) (invited speaker), ‘27th ERES Industry Seminar: Greek Real Estate Market: REICs – Innovative Technologies, Friday 22nd March 2019 – Athens, Greece.
20. T. Dimopoulos (2012) (invited speaker), The impact of Real Estate in Cyprus’ Economy, RICS annual conference, Nicosia, September 2012.
21. T. Dimopoulos (2017) (invited speaker), Cyprus Real Estate Market & Expectations, RICS Conference on REO & NPL Management, March 8, 2017, Hilton, Nicosia, Cyprus
22. T. Dimopoulos (2017) (invited speaker), ‘Analysis of Cyprus’ Property Market 2012-2016’, 11ο συνέδριο ανάπτυξης γης και οικοδομών, June 8, 2017, Hitlon, Nicosia
23. T. Dimopoulos (2017) (invited speaker), ‘Πλεονεκτήματα των ATM σε θέματα εκτιμήσεων ακινήτων και παρουσίαση εφαρμογών’, Ο ΑΤΜ στην Εκτίμηση, Αξιοποίηση και Διαχείριση της Ακίνητης Περιουσίας, June 23-24, 2017, NTUA, Athens.
24. T. Dimopoulos (2018) (invited speaker), ‘Analysis of Cyprus’ Property Market 2012-2017’, 12ο συνέδριο ανάπτυξης γης και οικοδομών, May 17, 2018, Hilton Park, Nicosia.
25. T. Dimopoulos (2018) (invited speaker), ‘The Future of Valuations’, October 2, Ajax Hotel, Limassol, RICS CPD event.
26. T. Dimopoulos (2014) (invited speaker), Criticism on the New General Valuation, Neapolis University Pafos, November 2014.
27. T. Dimopoulos, V. Pashoulis (2014), Analytical presentation of the property taxation system in Cyprus, Workshop: ‘Market Value-Based Taxation of Real Property: Lessons from International Experience’. Mar 17–21, 2014, Ljubljana, Slovenia.
28. T. Dimopoulos (2014), GIS applications in Real Estate, 1st Annual Conference on Geographical Information Systems, Nicosia, 8-9 October, 2014.
29. P. Sivitanides, T. Dimopoulos (2012), Analysis of Cyprus Real Estate Market from 2006 to 2012, Pafos & Limassol, April 2012.
30. T. Dimopoulos, V. Pashoulis (2015), The new General Valuation in Cyprus at 1.1.2013, Workshop: ‘Market Value-Based Taxation of Real Property: Lessons from International Experience’. Mar 23 –27, 2015, Ljubljana, Slovenia, Ljubljana, Slovenia.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CUT:	Cyprus University of Technology
VAT:	Value Added Tax
CAMA:	Computer Assisted Mass Appraisal
AVM:	Automated Valuation Model
NUP:	Neapolis University Paphos
MRA:	Multiple Regression Analysis
GWR:	Geographic Weighted Regression
ANN:	Artificial Neural Networks
AI:	Artificial Intelligence
DLS	Department of Lands and Surveys
IMF:	International Monetary Fund
ECB:	European Central Bank
CBC:	Central Bank of Cyprus
GIS:	Geographical Information System
SMARP:	Standard on Mass Appraisal of Real Property
VIC:	Value Influence Centers
GIM:	Gross Income Multiplier
OAR:	Overall Rate Models
COV:	Coefficient of Variation
SEE:	Standard Error of the Estimate
NDEV:	Normalization of the Deviation
AEP:	Adaptive Estimation Procedure
GV:	General Valuation
VIF:	Variance Influential Factor

CBD:	Commercial and Business District
VAT:	Value Added Tax
AIC:	Akaike Information Criterion
CV:	Cross-Validation
UN:	United Nations
RF:	Random Forests
ITZA:	In Terms of Zone A
LTV:	Loan to Value
NGVMC:	New General Valuation Model of Cyprus
DP:	Declared Price
AP:	Accepted Price
PC:	Property Characteristics
NDEV:	Normalisation of the Deviation

1 Introduction

The greatest proportion of a person's or a county's wealth, in most cases, consists of real estate. However, real estate values fluctuate over time and follow trends and cycles. Considering that significant events, such as the current COVID-19 pandemic or the global financial crisis of 2008, can affect property values significantly, and have the potential to completely change their previous status, the need for the revaluation of all real estate assets is generated. As a task, this would be very difficult, if not almost impossible bearing in mind time, cost and other constraints. Therefore, the development of accurate Mass Appraisal systems is very important and crucial, and their need is especially apparent in current circumstances.

Half of the real estate valuation community believes that property valuation is a science, while the other half is of the belief that it can be best described as an art. Generally, it is widely accepted that it is actually a combination of both, and that property valuation is fundamental to all decisions regarding real estate. Consequently, valuers have a leading role in the property market as their verdicts (valuations and appraisals) affect not only their clients but also a much broader group of stakeholders. Georgia Warren-Myers (2015) would add that world economies depend on the reporting and international definition of market value. She is of the opinion that valuers have a significant role in providing advice and assessments of the market value for various purposes to primary, secondary, and tertiary stakeholders. The latter rely on the outcomes of a valuation. Valuations influence sales, investment decisions, lease agreements, rentals, and lending decisions not only for the local but also for the international markets as well. Valuers are decision-makers in a very complex and unstructured market, as is that of real estate, and their influence can be far-reaching. Their responsibility is such that they are held legally accountable for their assessment and valuation of a property, and it could be considered that valuers, in their role, are the ultimate decision influencers (Warren-Myers 2015).

This thesis studies mass appraisals as a practice of valuing multiple properties simultaneously, as of a given date, by applying a systematic and uniform application of valuation methods and techniques that allow for a statistical review and analysis of the results.

The role of these tools has become crucial in modern economies in several aspects. Property is considered to be a heterogeneous asset as it is not frequently traded or easily transferred, and together with the constraints of the imperfect property market and the unpredictability of human behavior, an exact valuation of a property is difficult if not impossible to determine.

A more in-depth analysis of the factors that contribute to the determination of the market value of a property is required in order to better understand the determination of property value. As stated above, property valuation is considered to be both an art and a science. The science aspect applies economic theory, standardized and governed mathematical models and frameworks, international standards, national and state professional institutes, texts and journals, as the basis of valuation. While there is a large bibliography relating to the methods, approaches and market analysis techniques that comprise the "science" component of valuation (Warren-Myers, 2015), there is little research pertaining to the behavioral elements, the "art", that influence the decision making in valuation praxis.

Through the application of experience, market knowledge and the ability to exercise judgement, based experience within the market, valuers are able to compute a rational value for a property. Consequently, their assessment of value is not a purely mathematical calculation. A valuer needs to know, understand, and comprehend the workings of the marketplace they are operating within. Valuers are obliged to examine and use comparable market transactions, their knowledge, understanding, and perceptions as a basis for their interpretations in valuation practice. The development of this market knowledge is used in conjunction with known factors, principles, ethics, rules, and practice guidance to create heuristics that values apply on a daily basis, allowing short cuts to be taken.

"The reliance on expert intuition developed in the past results in limited adoption of new knowledge and comprehension of change, requiring new strategic knowledge development by valuers to create new heuristics that identify appropriate anchors and make adjustments. Consequently, with so many relying on the valuer to provide the "right" answer, a change within the market can have widespread effects on stakeholders and markets" (Warren-Myers, 2015).

An Automated Valuation Model can assist the public and the public interest, valuers, organizations, and institutions, and significantly improve the valuation outcome, not necessarily in terms of accuracy, but in equal importance in terms of consistency, speed, and uniformity.

Extensive research on relevant literature shows that there is no further research on purely valuation methodologies. As per the American literature and the Appraisal Institute, there are three valuation approaches: the comparative method, the income approach, and the cost method. According to the British bibliography and the RICS, there are five methods of valuation: the comparative method, the income approach, the residual method, the profits method, and the cost approach. In the latter case, the profits method is just a subcategory of the income approach, and the residual method is a variation of the cost method.

On the other hand, significant research is conducted every year in the field of Mass Appraisals, and the automation in the valuation procedure is gaining more traction every day. The author aims to contribute with this thesis on the promotion of those approaches to the broader property valuation environment.

1.1 Aims, Objectives, and structure of the thesis

The work described in this thesis aims to provide a rigorous and accurate analysis of the mass appraisal procedure, to highlight the relevant techniques and methodologies, and to propose methods to advance the currently used mass appraisal system in Cyprus through worked case studies.

The study objectives are summarized below:

- Analysis of a wide range of corresponding literature.
- Discussion over the appropriateness, applicability, and necessity of complex mathematical models for taxation and other purposes, as well as their importance and impact on the broader environment of valuation practice and the Real Estate industry.
- Investigation of the Cyprus Mass Appraisal System (General Valuation).
- Scrutinisation of the provided database.
- Comparison of different mass appraising techniques and identification of the optimum computational approaches for the available data.

- Integration of Artificial Intelligence algorithms for the mass appraisal of properties in Cyprus.
- Critical judgment of the models' accuracy, by utilizing professional experience on specific test cases, with a focus on outliers and observations with high errors.

The valuation of a property is a task of high professional responsibility, and any mistakes made by human valuers, or errors within a computer model could lead to implications for the owner as well as any involved stakeholders. This Ph.D. marks the first time that a Ph.D. thesis has been prepared regarding the development of an AVM system in Cyprus as highlighted by the corresponding analysis of literature. Best practices and norms from abroad are identified, analysed and recommended, while some state of the art AI algorithms are successfully introduced.

Chapter 1: Introduction

This chapter provides a review of the art and science components of valuation. Additionally, it highlights the importance of valuation to international, regional and local economies. Aims, objectives, structure and novelty of the thesis are included.

Chapter 2: Literature Review

This chapter includes the history of mass appraisals, as well as definitions, methodologies and models' specifications, calibration and adjustments. Additionally, the most common applications of mass appraisals are highlighted, and their use for property taxation purposes is discussed more analytically due to their vast application in this field. A state of the art machine learning-based algorithm was applied on a database retrieved from the Scopus database (*Scopus*, 2019), where the authors, their collaborations, frequency and association of keywords, as well as a time-series of their evolution over the last 30 years, are automatically retrieved and presented.

Chapter 3: Mass Appraisal in Cyprus (General Valuation)

In Cyprus, four mass appraisals took place for taxation purposes, all conducted by the Department of Lands and Surveys (DLS), the corresponding Governmental Agency. This part of the thesis describes the implementation of the CAMA model that is currently applied by the DLS.

Chapter 4: The database

For the purposes of this thesis, an enhanced database was provided to the author by the Department of Lands and Surveys. A significant vagueness was found among the declared and accepted prices. The author analysed the dependence of their deviation on the other parameters influencing a property’s value (covered area, location, etc.), by applying regression analysis techniques and the Relieff method.

Chapter 5: Case studies: MRA, GWR, Random Forests, AI and machine learning

This chapter contains the main computational results of the Thesis. It starts with an application of Geographically Weighted Regression (GWR) for apartments in Nicosia and compares the results with the Ordinary Least Squares (OLS) where the GWR method was found to provide significantly better results. Similar are the results for Thessaloniki Municipality where the use GWR advantages are apparent, when compared to the results from a regression analysis, when a local database was examined. The third case study is the application of Random Forests, a machine learning algorithm that is based on decision trees. The aforementioned method is also the introduction of Artificial Intelligence in Mass Appraisals in the thesis. The author recognises that these complicated algorithms are widely considered as a “black-box”, in contrast with the government requirement for a successful implementation of a Mass Appraisal to be transparent. Having in mind the data complexities, he continues a step forward, and proposes a form of sensitivity analysis, in order to highlight how complex models work, as well as the underlying relationships amongst the involved predictors and their response (property value).

Chapter 6: Conclusions, discussion and future work

This section includes the synopsis of the main findings of the thesis. The author highlights the strengths of complicated models via literature references and evidence-based simulations. Some ethical issues are also raised and discussed.

Conducting research on the topic of mass appraisals is a complex task, involving multiple disciplines as well as their inter-relationships. GIS, data management and handling, building surveying, legal framework and taxation knowledge is required on the one hand, while statistics, mathematical models and modern AI algorithms are also required. Accordingly, the purpose of the thesis was accomplished by investigating a variety of multidisciplinary literature, interdisciplinary collaborations with colleagues from other fields, while the final outcomes were critically analyzed, based on the author's critical judgement, stemming from some thousands of 'real-world' valuations of properties. The results were published in high impact journals and the thesis constitutes a synthesis of all the above-mentioned stages. The future of the author includes concluding the thesis.

1.2 Research Methodology strategy and steps

The research approach implemented in the thesis, is the inductive approach, according to which known models are applied to test different databases. The steps that were followed are synopsized below:

Step 1: Set the problem.

The problem was the suitability of the use of known Mass Appraisal models, in Cyprus.

Step 2: Research the relevant literature.

An extensive literature review was conducted. Best practices, norms, models, functions etc. were identified and a machine learning algorithm was applied for the more profound understanding of critical issues and research trends.

Step 3: Data Analysis.

The database provided, was scrutinized. MRA and machine learning techniques such as the Relieff method were applied in order to better understand the features of the database.

Step 4: Case studies.

Regression, GWR, Random Forest and Gradient Boost were some of the algorithms that were tested.

Step 5: Results.

It was concluded that overall, the GWR method provided higher accuracy than linear and non-linear regression models, as it better incorporated the location influence as a component of value. Random Forest, a Machine Learning Algorithm, was successfully implemented and showed perfect adaptation in real estate data. Finally, the author introduced the Sensitivity Analysis as an appropriate tool to decode, and somehow explain, the "black box" issue associated with such methods.

2 Literature Review

2.1 Mass Appraisals: History, Definitions, and Methodologies

Mass Appraisals were originally developed for property taxation purposes. Property tax is an essential source of fiscal revenue in many developed economies. The introduction of property tax can expand local tax sources, regulate the real estate market and facilitate income redistribution to mediate wealth polarization (Wang & Li, 2019). From the perspective of tax administration, there is a direct tax and an indirect tax. Property tax is a type of asset tax which is categorized under the direct tax category within the comprehensive tax system. However, the establishment of a property tax system is complicated, and involves not only relevant policies and laws, but also requires a valuation mechanism and methods. Generally speaking, in order for a property taxation system to be established, a large number of tax base assessments of real estate need to be carried out in a relatively short period of time. This assessment should conform to the laws of property valuation. In practice, it is essential to adopt a mass appraisal model which fits a specific country's real estate market structure, and that is also adaptable to changes in this structure over time. The technological advancements, the development of statistical models and the use of AI allows for a variety of other applications of such models, other than only for taxation purposes.

2.1.1 Definitions of Mass Appraisal

There are many definitions of Mass Appraisal. Most of the definitions vary and depend on the purpose of the system being introduced. After an extensive literature review, the most popular and accurate definitions found are:

According to Eckert et al. (1990), "Mass Appraisal is the systematic appraisal of groups of properties as of a given date, using standardized procedures and statistical testing".

The International Association of Assessing Officers (2017), defines "Mass Appraisal or Mass Valuation as the process of valuing a group of properties as of a given date and using common data, standardized methods, and statistical testing".

The newly formed association, the European AVM Alliance, defines "Mass Valuation as the practice of valuing large numbers of properties as of a given effective fate by the systematic and uniform application of valuation methods and techniques that allow for statistical review and analysis of the results”.

Labropoulos, (2013) adds that the purpose of mass appraisal is to provide an equitable and efficient appraisal of all the properties that fall within the specific administrative boundaries for ad valorem tax purposes. In the process of mass appraisal, a valuation model capable of replicating the force of supply and demand over a wider area is created, and appraisal judgements relate to groups of properties rather than to single properties. The quality, accuracy and consistency control of these assessments are verified by statistical methods. Since mass appraisals are also used by the owners of a significant number of properties, they need to demonstrate the equitability of the taxation system.

Wang & Li, (2019), in their systematic literature review, they use the definition that comes from the SMARP (Standard of Mass Appraisal of Real Property) [1] which is as follows; “*Mass appraisal as the process of valuing a group of properties as of a given date and using common data, standardized methods, and statistical testing*.”, and it can be considered as the “Mass Appraisal 1.0”. Before the 21st century, relevant institutions and scholars produced abundant work on the theoretical construction and standard setting of real estate mass appraisals, which is summarized in Table 1.

Table 1.Main standards and institutions for mass appraisal.

Standard [1]	**Institution** [2]	**Year (1st version)**	**Year (latest version)**
SMARP [1]	IAAO	1976	2017
RICS Red Book [2].	RICS	1983	2020
IVS [3].	IVSC	1990’s	2020
USPAP [4].	AF	1987	2018

Many institutions, organizations and scholars have contributed significant research regarding mass appraisal theory and the mass appraisal criteria foundations. Not only

[1] *Full names of standards: SMARP (Standard on Mass Appraisal of Real Property), RICS Red Book (RICS Valuation – Global Standards, IVS (International Standards), and USPAP (Uniform Standards of Professional Appraisal Practice).*

[2] *Full names of institutions: IAAO (International Association of Assessing Officers), RICS (the Royal Institution of Chartered Surveyors), IVSC (the International Valuation Standards Council) and AF (the Appraisal Foundation).*

have they provided criteria regarding the process and methods of mass appraisal in the practical application process, but also they have provided their own detailed suggestions that include analytical guidance on complaint treatment following the assessment.

The most popular forms of Mass Appraisals (or Mass Valuations) are the CAMAs and the AVMs.

A.V.M. is the acronym for Automated Valuation Model. AVMs are mathematical based computer software programs that market analysts use to produce an estimate of the market value of a property. The distinguishing feature of an AVM is that it is a market appraisal produced through mathematical modelling. They are based on a variety of datasets that are collected separately, such as market analysis of location, market conditions, and real estate characteristics. The credibility of an AVM is dependent on the data used, and the abilities of the modeler who created the AVM.

On the other hand, C.A.M.A. is the acronym for Computer Assisted (or aided) Mass Appraisals. It also refers to a software package that offers a holistic mass valuation solution but for tax purposes.

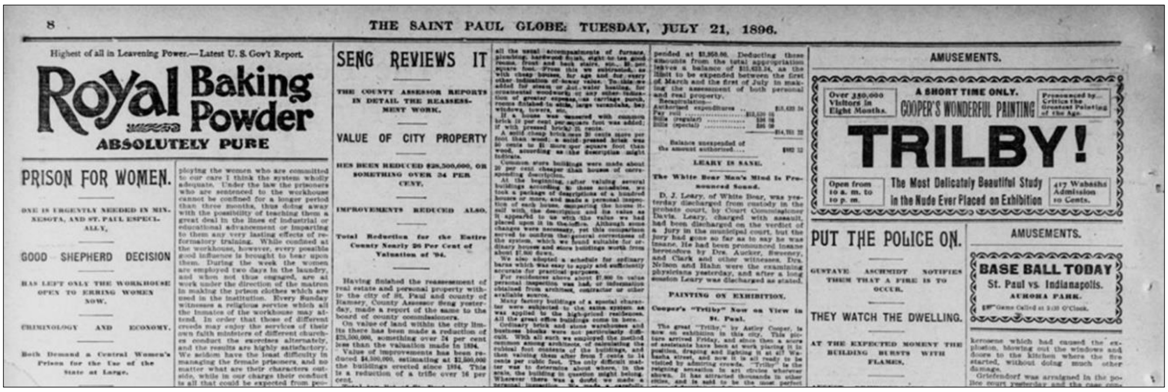
2.1.2 Mass Appraisal Systems

A more complete definition of a Mass Appraisal System, is provided by the author as: *"The independent automated and computer assisted system that collects and maintains property characteristics (including geographical data), process and cleans the data and estimates properties value with uniformity and consistency by providing also an error estimate. Such a system includes (normally) more than one mathematical model, and in the ideal scenario the transactions (sales) as well as the property attributes shall be updated automatically. Uniformity and accuracy will be ensured from known statistical measurements such as the COD or the MAPE".*

According to McCluskey et al. (1997), in mass appraisal modelling, "the aim is to construct a representative mathematical model that replicates the market within which the real estate is traded and to be clearly defined in accordance with the principles of microeconomics and at the same time being adaptable to groups of different types of properties". Eckert (1990), adds that "mass appraisal builds on the same principles as single-property appraisal". Mass appraisal techniques, however, emphasize valuation models, standardized practices, and statistical quality control. For the execution of mass appraisal, given the scale of valuations, statistical methods are used for quality control, in order to measure accuracy variations in the assessed value from actual sale prices. For most of the appraisal models if the average deviation from sales prices falls within a predetermined range, the model and quality is considered to be good. On the other hand, in single property valuation, this procedure is simpler, as the quality can be measured by direct comparison with specific comparable sales.

Renshow (1958), says that it may not be possible to identify and take into account all the factors that affect the selection and purchase of a property. However, it is possible to select some of them and examine their correlation with the values of these properties. Although this selection may seem arbitrary, it nevertheless seems to be sufficient to specify the value of the under-study properties with the required statistical accuracy. His view is actually the main idea behind all modern mass appraisal systems.

Eckert (1990), discusses that mass appraisals evolved out of the need for uniformity and consistency in ad valorem appraisals. The first attempts of mass appraisals took place in 1896 in the city of St. Paul - Minnesota, by William A. Somers. The necessity for the creation of such a system arose, as described in the relevant article at the time Clow (1896), when an economic recession had affected land values greatly, but not uniformly. Meanwhile, as reported in “The Saint Paul Globe” in 21.07.1896, issued by William A. Somers, the changes in existing land and building uses in combination with many overpriced commercial properties that remained vacant, led to the imperative need for a mass revaluation of the area's approximately 100,000 properties (Labropoulos, 2013).


8 THE SAINT PAUL GLOBE: TUESDAY, JULY 21, 1896.

Highest of all in Leavening Power.—Latest U. S. Gov't Report.

Royal Baking Powder

ABSOLUTELY PURE

PRISON FOR WOMEN.

ONE IS URGENTLY NEEDED IN MINNESOTA, AND ST. PAUL ESPECIALLY.

GOOD SHEPHERD DECISION

HAS LEFT ONLY THE WORKHOUSE OPEN TO ERRING WOMEN NOW.

CRIMINOLOGY AND ECONOMY.

Both Demand a Central Women's Prison for the Use of the State at Large.

SENG REVIEWS IT

THE COUNTY ASSESSOR REPORTS IN DETAIL THE REASSESSMENT WORK.

VALUE OF CITY PROPERTY

HAS BEEN REDUCED $28,500,000, OR SOMETHING OVER 24 PER CENT.

IMPROVEMENTS REDUCED ALSO.

Total Reduction for the Entire County Nearly 26 Per Cent of Valuation of '94.

LEARY IS SANE.

The White Bear Man's Mind Is Pronounced Sound.

PAINTING ON EXHIBITION.

Cooper's "Trilby" Now on View in St. Paul.

AMUSEMENTS.

A SHORT TIME ONLY.

COOPER'S WONDERFUL PAINTING

TRILBY!

The Most Delicately Beautiful Study in the Nude Ever Placed on Exhibition

PUT THE POLICE ON.

GUSTAVE ASCHMIDT NOTIFIES THEM THAT A FIRE IS TO OCCUR.

THEY WATCH THE DWELLING.

AT THE EXPECTED MOMENT THE BUILDING BURSTS WITH FLAMES.

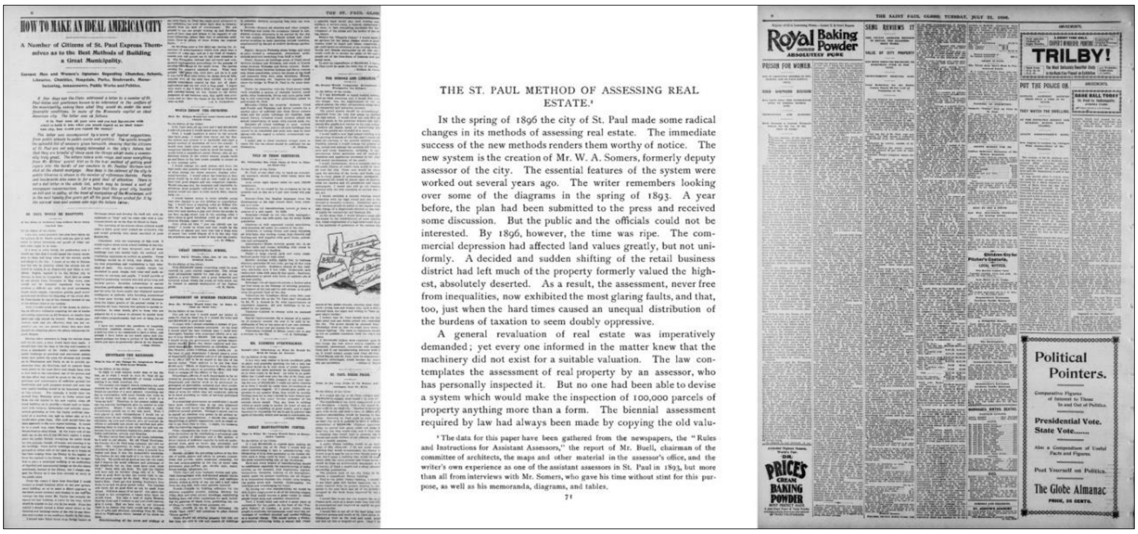
AMUSEMENTS.

BASE BALL TODAY

St. Paul vs. Indianapolis.

AURORA PARK.

Figure 1: Publication of the first Mass Appraisal attempt (Labropoulos, 2013)



HOW TO MAKE AN IDEAL AMERICAN CITY

THE ST. PAUL METHOD OF ASSESSING REAL ESTATE.[1]

In the spring of 1896 the city of St. Paul made some radical changes in its methods of assessing real estate. The immediate success of the new methods renders them worthy of notice. The new system is the creation of Mr. W. A. Somers, formerly deputy assessor of the city. The essential features of the system were worked out several years ago. The writer remembers looking over some of the diagrams in the spring of 1893. A year before, the plan had been submitted to the press and received some discussion. But the public and the officials could not be interested. By 1896, however, the time was ripe. The commercial depression had affected land values greatly, but not uniformly. A decided and sudden shifting of the retail business district had left much of the property formerly valued the highest, absolutely deserted. As a result, the assessment, never free from inequalities, now exhibited the most glaring faults, and that, too, just when the hard times caused an unequal distribution of the burdens of taxation to seem doubly oppressive.

A general revaluation of real estate was imperatively demanded; yet every one informed in the matter knew that the machinery did not exist for a suitable valuation. The law contemplates the assessment of real property by an assessor, who has personally inspected it. But no one had been able to devise a system which would make the inspection of 100,000 parcels of property anything more than a form. The biennial assessment required by law had always been made by copying the old valu-

[1] The data for this paper have been gathered from the newspapers, the "Rules and Instructions for Assistant Assessors," the report of Mr. Buell, chairman of the committee of architects, the maps and other material in the assessor's office, and the writer's own experience as one of the assistant assessors in St. Paul in 1893, but more than all from interviews with Mr. Somers, who gave his time without stint for this purpose, as well as his memoranda, diagrams, and tables.

71

Figure 2: Publication and explanation of the first Mass Appraisal attempt (Labropoulos, 2013)

2.1.3 Components, functions, and principles of Mass appraisal Systems

Eckert et al., (1990) divide a mass appraisal system into four key interdependent subsystems, the ***data management***, the ***sales analysis***, the ***valuation,*** and the ***administrative*** subsystem. Each of the four components may be further divided into smaller sub-categories and are briefly outlined below:

The ***Data Management System***: It has components for collection, entry, editing, organization, conversion, and storage of property and ownership data. This data management system is the heart of the mass appraisal system and should carefully be planned and designed, but also be subjected to quality controls in order to assure the accuracy of the values. The author would add that the integration with a GIS system is more than necessary as it assures all the above. The integration with GIS and Remote Sensing are analyzed further in the conference proceedings "*Comparative analysis of property taxation policies within Greece and Cyprus evaluating the use of GIS, CAMA, and remote sensing techniques*" (T. Dimopoulos et al, 2016).

The ***Sales Analysis System***: It has mechanisms for sales data collection, sales screening and processing, ratio studies and sales reporting. The main component of this system is the ratio studies component, that has a main objective to provide the best available measures of appraisal performance and is a valuable tool for monitoring appraisal results, identifying reappraisal priorities, adjusting valuations to the market and assisting management in planning and scheduling. Unfortunately, the existing procedure in Cyprus

DLS is not using their in-house CAMA model, and their assessment is based on desktop manual assessments. The above creates the paradox of a triple valuation figure; the declared price, the accepted price, and the value for taxation purposes. It is also recommended that the sales analysis should provide an internal property index for each property type for the necessary time adjustments.

The ***Valuation System***: It consists of mass appraisal applications for the comparison of sales, costs, and income approaches to valuation. Sales comparison applications include multiple regression analysis, adaptive estimation procedure (AEP), or feedback, and automated comparable sales analysis. The cost approach requires maintenance of computerized cost schedules and equations, derivation of depreciation schedules from market data, and reconciliation of cost-generated values with the market. Mass appraisal applications of the income approach include the development and use of income multipliers, and overall rates such as market yields, discount rates for different property types (hotels, businesses etc.). Values produced by these three approaches should be reviewed and reconciled to select a final value for assessment purposes.

The ***Administrative System***: It is composed of a variety of functions and activities subject to varying degrees of automation. One of the first functions to be automated was the preparation of assessment notices and tax bills for property taxation purposes. Many jurisdictions have also computerized other administrative functions. Finally, the administration system manages the appeals process. It should be added that the support of the appropriate legal framework is a prerequisite, especially for tax purposes. On the other hand, if the purpose is for portfolio management or loan management, the regulations from the Central Bank or the institutional internal procedures are required.

Within the same body of work that, is considered as the cornerstone of Mass Appraisals, Eckert et al, (1990) distinguish three basic functions in a typical Mass Appraisal system the ***reappraisal***, the ***data maintenance*** and the ***general principle***s, all of which are analyzed below:

Reappraisal: A mass appraisal system that must be constantly kept up-to-date and reflect the market conditions as much as possible. For this reason, a periodic reappraisal of the values of the properties' under-study is required, with a proposed revaluation frequency of once per annum or a different periodicity. However, the application of complex AI algorithms may allow for instant updates. Even one new transaction may change the values of the under-appraisal properties instantly. The aforementioned is one of the main advantages that AI presents in comparison with traditional approaches that are based in MRA. The required activities for the reappraisal were gathered from several web sources from local authorities and Eckert et al., (1990) and are outlined in the paragraphs below:

- Reappraisal decisions are necessary as laws or administrative rules sometimes impose reappraisal requirements, which require a new valuation to be carried out. Some jurisdictions use a cyclical schedule, in which a portion of the properties under study is physically reviewed and revalued annually. Other jurisdictions revalue all properties under study in a mass appraisal at periodic intervals or in response to ratio study results or external factors. In any case, and due to the time-consuming process and its high cost, meticulous planning and a major commitment of resources is required.
- Planning, organization, the identification of the target completion date and performance objectives are also required. A specific action plan and schedule is developed in order to achieve the time objective. The plan defines critical activities and their completion dates, assigns responsibilities and sets production standards for data collection and fieldwork. In some cases, in order for the requirements listed above to be met, new employees may have to be hired.
- System development, which produces the procedures, methods, manuals, and software for each of the sub-systems of a mass appraisal system. The first products should include forms, manuals, and procedures for collecting and processing property characteristics, sales and income data.
- Data collection. The data is thoroughly edited and tested before being used for valuation. The quality of the available data will determine valuation accuracy. The author believes that this step is the most critical for the successful and strategic implementation of a reappraisal.
- Analysis of available resources, such as staff, budget, existing systems and practices, data processing support, and existing data and maps, is required. The resources available determine the type and capacity of the system that can be supported, and the time required in order for it to function.
- Values equitability and consistency with the current market conditions should be examined. The tool for such analysis is the ratio study. Ratio studies provide a set

of statistics describing the distribution of the ratios (such as central tendency and spread), as well as summaries of uniformity (horizontal and vertical equity). If performance is poor, a reappraisal is necessary, particularly if one has not been conducted for a long period of time.

- A pilot study, which tests new procedures in one or two areas of the jurisdiction, should be considered whenever major changes are carried out. A pilot study will show if the new system produces accurate and reliable values and also suggests modifications in procedures.
- Production of values, which begins with market analysis, model development, model calibration, and calculation of preliminary values. A ratio study then evaluated the accuracy and consistency of values between property types and areas. When models produce acceptable results, they can be used to produce values, which are then subjected to office and field reviews.
- Preparation of the assessment roll, which includes the final values in a form satisfying legal requirements. Assessors should have prepared for informal and formal appeals. Although processing appeals consumes much staff time, it provides an opportunity to review individual values in detail and make necessary corrections. After the appeals process is completed, tax billing can begin. At that point many researchers argue that consistency is preferred rather than a close estimation to the actual market value when such system is used for taxation purposes.
- Final performance analysis, where the assessor conducts a final ratio study to measure and evaluate the accuracy and uniformity of the new values. This study plays a key role in summarizing the achievements of the new system or reappraisal and in preparing for the next reappraisal.
- Data Maintenance: It is the process of capturing and valuing new construction and other changes to the property base. The process includes:

1. New records adding, that can arise due to some modifications to the land or even to the building.
2. Verification of existing information, implementing routinely inspections, during which the property and property record are compared.
3. Value updates that are annually based adjustments, applied to properties between reappraisals. A mass appraisal system can use ratio studies or other market

analyses to derive trending factors based on property type, location, size, age, and other similarities.

Principles of Mass appraisal: Although Mass appraisal builds on the same basic principles as single-property appraisal, mass appraisal techniques emphasize more in equations, tables and schedules, collectively called ***models***. Constructing such models can be viewed as a two-step process:

- **Specifying the general model**, that provides a framework to simulate supply and demand forces operating in a real estate market and is adaptable to many uses. The model builder specifies the variables, or property characteristics, to be used in the model and their relationships. A simple general model can be expanded to reflect the complexities of the market by expansion of each variable. Model specification is the first step in the development of any mass appraisal model. Mass appraisers must understand the models they use and be sure they reflect the way property is valued in the local market. When the model structure is imposed, as it often is in a commercial mass appraisal software, the user may not be able to select the property characteristics and their relationships in the model. In contrast, other software, particularly general-purpose statistical software, affords the user complete freedom in determining model structures. The choice of such model is not a standardized but is completely different for a different dataset and shall be adjusted as per the provisions of the regulatory jurisdiction.
- **Model calibration**, which is the process of adjusting mass appraisal formulas, tables and schedules to the current market. Although the structure of a mass appraisal model may be valid for many years, the model is usually calibrated or updated every year. To update for short periods, trend factors may suffice. Over longer periods, complete market analyses are required. At this point it is stated again the importance of an in-house index that time adjusts the sales transactions.

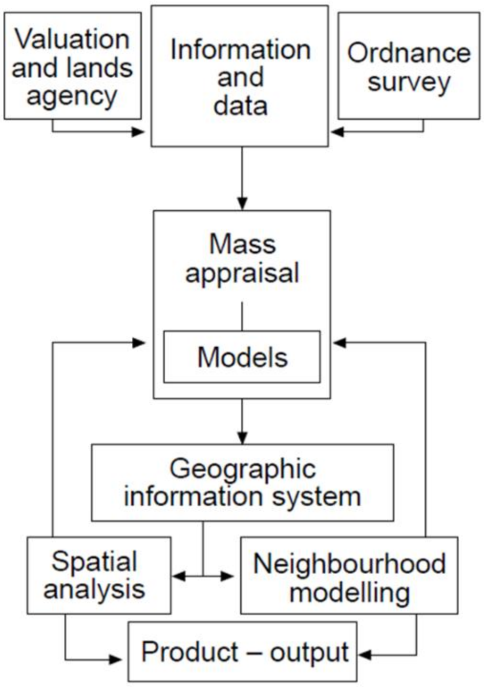


Figure 3: Graphical representation of a Mass Appraisal Model (McCluskey, 1997)

2.1.4 The three basic Mass Appraisal Approaches

Eckert et al., (1990), Labropoulos (2013) and other researchers divide the mass appraisals models into three major categories, following the three basic valuation approaches as per the American literature: The cost, the income and the comparative approach.

- **Cost Approach**
  - Cost tables: include base rates, per square foot adjustments and lump sum adjustments that are used to calculate the replacement cost as new.
  - Depreciation schedules: they are developed for each type of property and then tested to ensure that they reflect the local market, taking into account the condition and age of the property.
  - Time and location modifiers: they are used to adjust cost data for local variations and changes in costs over time.
  - Market adjustment factors: they are used in order to adjust the values obtained from the cost approach in order to reflect market conditions. These adjustments should be applied by type of property and area based on sales ratio studies or other types of market analysis. Accurate cost schedules,

condition ratings and depreciation schedules will minimize the need for the market adjustment factor.

- **Income approach**
    - Market and existing / passing rents: can be developed from income data obtained from the owners or in some cases local third-party sources.
    - Vacancy and expense ratios: they are used to adjust potential gross income to typical net income. These ratios must reflect the local market and differentiate according to the types of commercial properties and the age groups.
    - Income rates and multipliers: they are used to convert income to market value.
- **Comparable Sales Method**
    - Land valuation tables: they contain land values per unit, along with standard adjustments for topography, depth, water supply, frontage and other locational features.
    - Sales comparison adjustments: they are derived from local sales analysis and can be based on statistical techniques such as multiple regression analysis and feedback.
    - Multiple regression analysis equations: separate equations should be developed for each market area in a jurisdiction, or a general model should be adjusted for locational variations.
    - Feedback equations: it is another automated version of the sales comparison approach, while the ‘prices’ assigned to each property characteristic are based on sales analysis.
    - Base home table: they repackage a statistical equation, derived through multiple regression analysis equations.

However, in many cases the models are hybrid, where more than one approaches are utilized.

2.1.5 Mass appraisal model building and calibration

According to Eckert et al. (1990), “a model is a representation of how something works. Model building requires good theory, data analysis and research methods. The best valuation models will be accurate, rationale and explainable. Model specification and calibration are distinct steps in modeling. Model specification is the beginning of models based on economic and appraisal theory and market analysis”.

In practice, specification and calibration are performed in an iterative process, which includes the following steps (International Association of Assessing Officers, 2018):

1. Model specification
2. Model Calibration
3. Test the model
4. Make adjustments to model specifications
5. Recalibrate the model
6. Test the model
7. Repeat the process until the model quality assurance tests are met

Eckert et al., (1990) adds that there is a wide selection of data, variables and model structures available in specifying mass appraisal models:

- **Types of Data:**
  Data used in mass appraisal models can be either qualitative (e.g. neighborhoods, cooling systems, roof types) or quantitative (e.g. building age).
- **Data transformations:**
  The variables used in model building derive from property data. The variables can be simple – or untransformed – or transformed using mathematical transformations. Such a transformation is the conversion of qualitative data to:
    - binary variables, or
    - scalar variables.

  Mathematical transformations are useful in transforming quantitative data too. The most common mathematical transformations used in mass appraisal modeling include:
    - reciprocal transformation,
    - exponential transformation,
    - logarithmic transformation,
    - multiplicative transformation, and
    - quotient transformation.

- **Types of models:**

A model is composed of one dependent variable and one or more independent variables. The dependent variable is what is being estimated, usually the property value, while an independent variable is an item used to predict or explain the dependent variable (Eckert et al., 1990). The most common model structures are the following (International Association of Assessing Officers, 2018):

  - **Additive Models**, with the following form:

$$\mathbf{MV = B_0 + B_1 \times X_1 + B_2 \times X_2 + ...} \qquad \textbf{[2.1]}$$

where: MV = the dependent variable;
B_0 = a constant;
X_1 represents the independent variables in the model; and
B_1 = corresponding rates or coefficients.

  - **Multiplicative Models**, in which with the variables are either raised to powers or are themselves powers to which coefficients in the model are raised; the results are then multiplied rather than added. An example follows:

$$\mathbf{MV = B_0 \times X_1^{\ B1} \times X_2^{B2} \times \ldots} \qquad \textbf{[2.2]}$$

where: MV = the dependent variable;
B_0 = a constant;
$X_1^{\ B1}$ where X represents the independent variables in the model; and
$X_1^{\ B1}$ where B1 represents the corresponding rates or coefficients.

Multiplicative models consist of a constant B_0 and percentage adjustments. They have several advantages, including the ability to capture curvilinear relationships more effectively and the ability to proportionately adjust the value of the property being appraised. Multiplicative models are usually calibrated using linear regression packages. This requires some of the variables to be mathematically transformed for calibration. In this case the market value would require a transformation (International Association of Assessing Officers, 2018).

  - **Hybrid (Nonlinear) models**, which are the combination of additive and multiplicative models. The following example of a hybrid model is specified to the same as a cost model and demonstrates the flexibility of the hybrid model. A hybrid (nonlinear) model specification that separates value into building, land, and "other" components (e.g., outbuildings) is:

$$\mathbf{MV = \pi GQ \times [(\pi BQ \times \Sigma BA) + (\pi LQ \times \Sigma LA) + \Sigma OA]} \qquad \textbf{[2.3]}$$

where: πGQ = the product of general qualitative components

πBQ = the product of building qualitative components

πLQ = the product of land qualitative components

ΣBA = the sum of building additive components

ΣLA = the sum of land additive components

ΣOA = the sum of other additions additive components

Determining model structures from data analysis:

Mass appraisal models must reflect appraisal theory and market behavior. Successful modelling begins with market analysis, including a profile of the properties being modelled. Principal tools used in data analysis are data profiles and two – and three – way analyses (Eckert et al., 1990):

- Data profiles: consists of arrays and associated statistics such as the median, minimum, maximum, range and quartile. Qualitative data should be analyzed using frequency distributions, histograms and polygons.
- Two-way analyses: uses statistical tools to evaluate the relationship between two variables. Cross-tabulations, scatter diagrams and correlation analysis, breakdowns by strata are some of the most common and effective tools for analyzing relationships between variables.
- Three-way analyses: uses several techniques to evaluate the relationship between three variables. Contingency tables show how a response variable varies with changes in two control variables.

Data profiles and graphic analyses are limited to three dimensions or variables. The real world is more complex, and the calibration of mass appraisal models, therefore, requires multivariate statistical tools, such as MRA and feedback, that can handle many variables simultaneously. Real estate market is very complex by its nature, therefore, the use of techniques from data sciences and machine learning techniques are presenting better mathematical results.

2.1.6 Model specification

Model specification is used in order to determine which valuation models yield the best possible results by reviewing the data used. Model specification is based on data analysis and appraisal theory. Sample model specifications are provided below:

- **Cost approach specification:**

"The cost approach is an indirect method of arriving at market value, based on specification of replacement cost new less depreciation plus market derived land value. The cost approach is calibrated by review of construction costs and sales. This approach generally produces more acceptable results for newer properties, specialized properties and properties with insufficient sales. Model specification for the cost approach requires the estimation of separate land and building values" (International Association of Assessing Officers, 2018).

Two cost approach formulas for model specification are:

$$\mathbf{MV = \pi GQ \times [(1 - BQ_D) \times RCN + LV]} \qquad \textbf{[2.4]}$$

where: MV = the market value estimate;

πGQ represents the general qualitative variables such as location, economic adjustments, and time of sale;

BQ_D = a building qualitative variable representing depreciation;

RCN = the replacement/reproduction cost new;

LV = the land value.

and

$$\mathbf{MV = \pi GQ \times [(\pi BQ \times \Sigma BA) + (\pi LQ \times \Sigma LA) + \Sigma OA]} \qquad \textbf{[2.5]}$$

where: MV = the market value estimate;

πGQ = the product of general qualitative variables;

πBQ = the product of building qualitative variables;

ΣBA = the sum of building additive variables;

πLQ = the product of land qualitative variables;

ΣLA = the sum of land additive variables; and

ΣOA = the sum of other additive variables.

If a third party provides the cost tables, it is the responsibility of the market analyst to calibrate the cost tables to the local market in order to provide a valid indicator of value by the cost approach "It is also the responsibility of the analyst to fully understand the assumptions made by the third party in constructing the cost tables, the original source of the construction costs used, and the way the unit-in-place costs were aggregated by the third party to arrive at the published square foot rates" (IAAO, 2018).

The specification of the cost approach includes the following steps (Eckert et al., 1990):

1. Stratification improvements into homogenous groups. There is usually a different market for each group. For example, for residential properties, properties are usually stratified by occupancy, story height, and construction grade or wall type, while for non-residential properties, properties are stratified based on structure type and number of floors.
2. Specification of the typical characteristics, or base specifications of each model. Using the unit-in-place or quantity survey method, reproduction or replacement cost as new is then estimated. In mass appraisal, these costs are converted to comparative units of cost. For residential structures, costs can usually be divided into horizontal, vertical, building addition, constant building component, qualitative, and addition costs, while the non-residential cost models contain the six basic cost components mentioned, but with some important differences in the format of the horizontal and vertical components.

- **Sales comparison models specification:**

According to the International Association of Assessing Officers (2018), the Sales Comparison Approach can be implemented as a comparable sales model or as a direct market model.

- The Comparable sales model routine includes two steps. The first step involves the selection of the sales most comparable to the subject properties and a series of data filters. The second step adjusts the sale prices of the comparable sale properties to the subject properties based on differences in data characteristics. The adjusted sale prices are then used to determine an estimation of the value for the subject properties. Both steps in the comparable sales routine may be supported by statistics or a statistically-based direct market model. Model specification for the comparable sales method can be summarized as follows:

$$\mathbf{MV = SP_C + ADJ_C} \qquad \textbf{[2.6],}$$

Where: MV represents the market value estimate;
SP_C represents the selling price of comparable sale properties and
ADJ_C represents adjustments to the comparable sales.

- The Direct Market Model may use one of the three model structures mentioned above – the additive, the multiplicative or the hybrid model.

According to (Eckert et al., 1990), the specification of the sales comparison approach includes the following steps:

1. Stratification and location analysis. Three basic ways to analyze location in mass appraisal model building are: multiple models based on geographic stratification, multiple models based on cluster analysis, and a single model with location adjustments.
   - Geographic stratification: boundaries are drawn along rivers and other natural barriers, major streets and subdivision lines to reflect major differences in location. Separate models are then specified for each such area.
   - Stratification using cluster analysis: it combines properties into relatively homogenous strata-based categories such as location and physical characteristics, such as age, size, style and construction quality.
   - A single model with location adjustments: in this method, a single model with locational variables is used for the entire jurisdiction. A more sophisticated technique is used to identify value influence centers (VICs), a technique that adjusts location smoothly, virtually eliminating boundary problems.
2. Selection of the appropriate model considering the property type. Model specification varies by property type and can be additive, multiplicative or hybrid. The properties can be divided into three categories: residential properties, income generating properties and land. Each category can be further subdivided (e.g. the residential properties can be subdivided into

single-family residential, condominiums/townhouses etc.), using different types of models.

- **Income approach specification:**

The income approach is an indirect method for determining market value. The appraiser evaluates income for quantity, quality, direction, duration, and the expense incurred to earn this income, and then converts it by means of a capitalization rate into an estimate of market value (International Association of Assessing Officers, 2018).

In an income model, the dependent variable may be income, income per unit (net rentable area), expenses, and expenses per unit or capitalization rate. The steps in this approach are (Eckert et al., 1990):

1. Modelling Gross income,
2. Modelling Net Income,
3. Modelling a Gross Income Multiplier (GIM) - or Gross Rental Multipliers (GRM) (IAAO, 2018) -which express the relationship between gross income and rent,
4. Modelling an Overall Rate Models (OAR), which expresses the relationship between NOI and property value.

All the above can be estimated in one of two ways: stratification and typical units of comparison, often using spreadsheet software, or statistical models using MRA or alternative techniques.

2.1.7 Model calibration

Eckert, (1990) suggests that the calibration of each model should be as:

- **Cost models calibration:** As the cost approach is applied in four steps: estimate RCN, estimate and apply depreciation, estimate and add land value, and apply general qualitative adjustments. The calibration of the model includes the calibration of RCN and the development of formula-driven cost models, cost-trend factors, depreciation schedules and market adjustment factors.
- **Sales Comparison calibration:** The two more common techniques available for calibrating sales comparison models are MRA and feedback. MRA is older and more common. Feedback was first used in property appraisal in the late 1970s. MRA models can be additive, multiplicative or hybrid. Their objective is to model

the relationship between property characteristics and value, so that the latter can be estimated from the former. The results are evaluated using eight statistics: the coefficient of determination (R^2), the standard error of the estimate (SEE), the coefficient of variation (COV), the average percent error, the coefficient of correlation, the t-statistic, the F-statistic and the beta coefficients.

- **Income models calibration:** In mass appraisals, income and expense analysis is conducted mainly by microcomputer spreadsheet software. Entering available data, the appraiser establishes typical income figures and expense ratios by property type and estimates unreported figures. In mass appraisal, direct capitalization models are used in the form of GIMs and OARs. These models are developed in two ways, stratification and MRA. In general, income models should include data collected prior to property taxes being subtracted as an expense, with an effective property tax rate added to OARs and GIMs after calibration.

However, these suggestions nowadays seem very generic and outdated. Model calibration is the second task in building models, in which the values for the coefficients in a model are estimated. According to the International Association of Assessing Officers (2018), many Mass Appraisal Systems in use today rely on statistical models as the method of calibration. Some example techniques may be based on regression, geographic-weighted regression, or neural networks. All regression programs are based on statistics, while neural networks draw on the analogies of adaptive neurons learning. Therefore, the following calibrations are suggested:

2.1.7.1 Calibration Using Statistically Based Methods

Multiple regression analysis (MRA) is a statistically based analysis that evaluates the linear and / or curvilinear relationship between a dependent (response) variable and several independent variables(predictors). Models produced using MRA come with a rich set of diagnostic statistics that provide evaluation tools for the market analyst to compare results between and among specified models. These include “goodness of fit” statistics such as R^2, adjusted R^2, SEE, COV, and “measures of variable significance”, such as R, t, F, etc. (International Association of Assessing Officers, 2018). According to (Kauko, 2008), the importance of mass appraisal may also be seen from the possibility of linking the relationship between the property value, the property characteristics, and urban social and economic problems. The literature available on mass appraisal modelling tools (including automated valuation methods (AVMs)) is rich and ever evolving. While the standard multiple regression analysis (MRA)-based hedonic price models may not be the most suitable tools for capturing all the necessary information involved in the formation of value, MRA currently remains the most important theoretical framework regarding mass appraisals.

Tabales et al., (2013) adds that “their aim is to estimate the price of a complex good, such as a dwelling, as a function of its characteristics. These characteristics cannot be separated and should be treated as a whole. They include attributes such as the size of the residence, or its internal structure, and variables related to its environment, such as the building properties, its surroundings, and so on”.

Mooya, (2017) adds that the basic principle of hedonic models is that the value of a property is a based on the sum of its constituent characteristics. This suggests that a property being valued can be broken down into its fundamental components, and the market values of those components. For example, the market value of a specific residential property is regarded as a summation of the values of its size, location, age, and so on, all of which hold a particular value in the market. Typically, hedonic models use regression techniques to estimate the contribution of each characteristic of the property to its overall value. Using the same methodology, hedonic AVMs use regression analysis for the or estimation of property value. The term regression analysis is used to describe the various mathematical methods (including econometrics) that aggregate observations into a form in which a dependent variable is a mathematical function of independent variables ($y = f(x1, x2, \ldots, xn)$), often in a way that allows a statistical inference regarding the parameters of the function outside the specific sample. (Mooya, 2017)

The popularity of hedonic price modelling for mass appraisal purposes stems from several factors. First, it rests on multiple regression analysis (MRA). (Thériault et al., 2003) suggest that this conceptually sound and very powerful analytical tool combines probability theory with calculus, thereby allowing the separation of crossed influences that affect property values. Second, it perfectly fits the very definition of market value, expressed as ‘the most probable price’ that should be paid in a competitive and transparent market setting. A market value, being a probability distribution, calls for a statistical treatment that conveys objectivity through direct market reading, which is easily reproducible and offers adequate testing of result reliability; MRA provides for such qualities. Third, the hedonic approach is not confined to producing value estimates, as it adds very useful insights into the causal dimensions of property value determination. Thus, it is viewed as a decision-making tool that brings about market intelligence, even more so when used in combination with geographic information systems (GIS), a ‘natural’, and increasingly imperative, complement for adequately handling spatial issues.

“The Automated Valuation Method (AVM) is a computer software program that analyzes data using an automated process. It is related to the process of appraising a large set of properties, using shared data and standard appraisal methodologies. Generally, the AVM is based on quantitative models (statistical, mathematical, econometric, etc.), related to the valuation of the properties gathered in homogeneous groups (by use and location) for which samples of market data are collected” (Ciuna et al., 2017).

2.1.7.2 Artificial Intelligence and Artificial Neural Networks

Another way for calibrating real estate valuation models are the Artificial Neural Networks (ANNs). The history of the ANN technique as it is presented in Abidoye & Chan, (2016) work titled “Research Trend of the Application of Artificial Neural Network in Property Valuation” and can be categorized into four stages namely beginning of neural networks – the 1940s; the first golden age – the 1950s to 1960s; the quiet years – the 1970s; and the renewed enthusiasm period of the 1980s – now. The final stage is the present-day neural network research area where the application of the technique exploded and received more attention by scholars. The study of McCulloch & Pitts (1943) was the first to employ the ANN technique to demonstrate the ‘threshold logic’ in the field of mathematics.

In the real estate sector, the seminal study of Borst, (1991) was the first to apply the ANN technique to property appraisal. The author found that ANN is reliable and accurate for property value estimation, but recommended that more research efforts shall be invested in the application of ANN. Since it has been established that the ANN technique handles the shortcomings of most of other appraisal techniques (Do & Grudnitski, 1992), researchers in different real estate markets around the world have investigated its application in their domains and have mostly reported a positive result (Limsombunchai, 2004).

"Neural networks can calibrate models that consist of both linear and nonlinear terms simultaneously. The user inputs each variable with assigned weights (coefficients). The software exposes the data using an algorithm in a hidden layer where the weights are adjusted (calibrated) in a manner that reduces the squared error. This is an iterative process much like those found with hybrid (nonlinear) regression. The final output results in a single estimate of value with the exact algorithm remaining hidden from the market analyst" (International Association of Assessing Officers, 2018).

An ANN is like a non-linear regression or a multivariate regression model, with non-observable linking variables. Once the topology and the parameters of the network are specified, it can be presented as an ordinary statistical or econometric model. Neural networks are used with different purposes, such as the estimation of models, classification, forecasting, and so on. Here it is used as a modelling tool and as a practical alternative to the well-known econometric hedonic models. Also, it aims to forecast the value of properties, so it can be used to measure the downturn that occurred over the last few years within the real estate market (Tabales et al., 2013b).

Tabales et al. (2013b), adds that the marginal prices obtained with ANN are more realistic than the classical hedonic prices. The price to be paid, using this more complex methodology, is linked both to the mathematical difficulties in obtaining the marginal prices, even with simple network topologies, and to the need of a fairly large sample required to order obtain stable parameter estimates

In general, the ANN approach presents clear advantages over the classical hedonic methodology, despite its complexity and the large samples required to order obtain reasonable forecasting capabilities.

Classical hedonic models have been used for several year in real estate valuation in order to estimate prices of complex properties. The availability of neural network software packages has allowed for the use of this alternative methodology, with significantly improved results. The use of ANN is more flexible than the use of classical econometric models, when there is suffice sample data available. Some common problems in hedonic models, such as non-linearities in the extreme range of prices of real estate market, are resolved by using ANN, and even adapt to properties that could be labelled as outliers, although, some authors criticize this 'black box' approach. Neural networks allow for a more appropriate fit with the usual measures, such as the determination coefficient, or with error related statistics (RMSE, MAE, and so on). The 'dark side' of this methodology is the sample of data required. This can be realistically overcome, only with the collaboration of realtors with a broad presence in the market, with large internal databases that contain real data regarding properties, where a random sample can be selected, and the uncertainty about the exact identification of each dwelling is eliminated, allowing the inclusion of particular data belonging to each building and its location (Tabales et al., 2013b).

Due to the improved efficiency of AVMs over traditional methods, it is inevitable that comparisons between the two are made. Overall, AVMs have the advantage of speed, cost-effectiveness, consistency and objectivity over the traditional methods. Despite these advantages, AVMs also present a number of weaknesses. As no on-site inspection is carried out, they lack insights that are obtained by visiting a property. They also lack human judgment and intuition, attributes that could prove to be indispensable for the accurate interpretation of market conditions. Also, AVMs assume average conditions (for both land and buildings), which could produce ambiguous estimates in some situations (Mooya, 2017).

2.1.8 Model adjustments

According to International Association of Assessing Officers (2018), "*in the process of specification and calibration of AVMs, adjustments should be considered for location and market trends. Additionally, the data must be reviewed in order to ensure representativeness, and variables should be selected using advanced statistical analysis. These requirements can be met through the following processes*".

- **Time Series Analysis:** Time series analyses are techniques that can be used to measure the cyclical movements, random variations, seasonal variations, and cyclical trends observed over time. In property valuation, these analyses can be used to develop a multiplier or index factor to update existing appraised values or to adjust sale prices for individual properties to a valuation date. Since values can change at different rates in different markets, separate factors should be tested for each property type and market area. Methods used to develop time trend factors for property valuation include:

- Value per-unit analysis;
- Re-sales analysis;
- Sales/assessment or sales/AVM value ratio trend analysis;
- Median/mean value per period in the form of a moving average/median sale price;
- Inclusion of time sale variables in MRA/AVM models.

Once a time trend is established, it can be used to adjust values to any point within the sales period. Trend factors can be extrapolated for a brief period beyond the sales period, but this should be done with caution and grows increasingly unreliable as the period is lengthened.

- **Independent Variable Selection for Models:** Modelers should include statistically significant and reasonable variables. Variables (e.g., property characteristics) that are highly correlated with other variables or have a statistically insignificant effect should be excluded or used with caution in the model. Developing a comprehensive criterion for variable selection requires in-depth knowledge in appraisal of various property types, advanced statistical analysis, and mathematics.
- **Location Adjustments:** Variables to express the influence of location are critical in any model. The effect of factors external to the property should be determined. There are several methods of accounting for location. Two common methods for developing location adjustments are the creation or use of existing geo-economic areas and use of geocoordinates. Geo-economic areas are the traditional and most usual form of location analysis. In AVMs, geo-economic areas may be based upon streets and natural boundaries, and government designated areas.

  Geographic coordinate techniques relate common elements such as price to each property's unique location. Software uses a variety of smoothing techniques to compute a unique location adjustment for each property. This variable is then included, along with other variables, in a multiple regression or other model to capture location influences. Geographically Weighted Regression, another method of accounting for location, works by allowing model coefficients to vary by location in order to pick up spatially heterogeneous effects.
  - Model frameworks that incorporate these methods for accounting for location include the following: Spatial Regression Model (i.e. Spatial Lag Models, Spatial Error Models);

- Inclusion of Locational Variable in the Model Specification (i.e. - Regression with Locational Binaries, Random Forests with a Locational Variable as a Feature);
- Multi-Stage Locational Adjustment Models (i.e. - Regression Model Paired with Locational Value Response Surface, Market-Adjusted Cost by Geo-economic Area);
- Spatial Interpolation Models (i.e. Kriging);
- Locally-Weighted Regression Model (i.e. - Geographically-Weighted Regression);
- Segmented Models (i.e. - Separate Regression Models by Market Area);
- Cluster Analysis.

2.2 The most popular applications of Mass Appraisals

According to the RICS' Information Paper (2013), which refers to the use of Mass Appraisals as AVMs, it seems that they can be used in both the public and private sector (Amato, 2017) as described below.

- By banks (lenders) for granting a new loan or subsequent revaluation for adjustment decisions regarding an existing loan. Lenders can obtain an AVM that can sometimes be used prior to processing a case, to see if the proposed figures being discussed are likely to be adequate, without incurring the cost of a full valuation by a valuer. Alternatively, an AVM can be used mid-way through a mortgage term to check how a property value may be fluctuating.

- In arrears assessment and planning. Banks may require checking the value of a mortgaged property provided as security for a loan in order to establish if the property value still covers the amount of the mortgage, and what scope there is in arrears planning. An AVM can provide a cost-effective and quick indication.

- In an audit of valuations. Banks, and those who audit valuations, sometimes require a second valuation, in order to audit the original valuation for a property. An AVM can provide this second valuation for a range of properties, as well as in individual cases.

- For the identification of fraudulent activity. An AVM can be applied to a range of property valuations in order to determine whether there is any activity that does not follow normal market trends.

- For estimating potential compensation payments that need to be to owners of properties that have been affected by public construction work, such as the creation of a new road network that affects the living conditions of a neighborhood for a certain period of time. Public authorities can use AVMs as a quick and cost-effective way to estimating the likely cost of compensation as part of a total 'project' cost. An appropriate model can provide a pre-project value estimate of affected properties and keep this up to date as required.

- For lending (capital to risk-weighted assets). AVMs can be applied to a portfolio of properties, that are mortgaged, in order to obtain an indication of how well they are performing. This can also assist the bank and their accountants in determining capital adequacy ratios.

- In order to provide a valuation, estimate for large-scale assets. This could include large portfolios of properties or 'packages' of properties that were obtained by a bank through bad debt and are now being sold off. In these cases, an AVM can be used to provide cost-effective sample valuations for portfolio purposes or for the valuation of a whole portfolio.

- In order to provide a valuation for individual capital tax purposes. An AVM can be used in these situations for individual properties or for a portfolio of properties, in order to establish a quick estimate of likely tax implications. This can aid with tax planning.
- In order to analyze the cost and benefits that are associated with potential public expenditure. When estimating the cost or benefit of public works, an AVM can provide a total cost estimate of the residential value of properties likely to be affected, in a quick and cost-effective manner.

- For tax purposes. Mass appraisals can be used for local taxation purposes. A model called CAMA is a sophisticated AVM that has the ability to provide a cost-effective valuation estimate for thousands or millions of properties.

The only official published application noted in Cyprus so far is the General Valuation conducted by the Department of Lands and Surveys for taxation purposes. Therefore, property taxation and its importance will be further analyzed in the following chapter, 2.3.

2.3 Property Taxation

2.3.1 A philosophical approach

The main part of this chapter has been already published at the Review of Decentralization, Journal of Local Government and Regional Development. (Dimopoulos, 2015).

According to Aristotle, property forms part of the household; consequently, the art of acquiring it, forms the basis of managing the household. He asserted that a person who owns property has psychological satisfaction, for it fulfills the human instinct for ownership and possession, while he also considers property as the focal point for ensuring the prevalence of both democracy and the rule of law in any state. Furthermore, Aristotle views property as a significant factor that has a lot of importance in the development of a state. This is also true in modern societies and economies when trying to develop a sophisticated economic system.

The Greek philosopher believed that taxes on private property have a lot of significance, as they contribute to the pool of a country's resources. It is highlighted that taxation is a means for the replenishment of a country's wealth and should be applied to all who have a property. The best model of taxation should be based on wealth and not property. However, there should also be some sort of sanction, in the form of a tax too, for those who fail to utilize the property which is capable of generating wealth, in order to minimize the waste of resources and the taxation returns benefit of the citizens.

Aristotle and Plato acknowledged social fulfillment to be a product of political and economic equality. For instance, in today's economy, most leaders who do not exercise the aspects of democracy advocate for constitutions that would eliminate them from paying taxes. As a result, most citizens in the aforementioned countries meet the burden associated with high taxation. According to philosopher Mariana, if those in authority do not pay taxes, there is a risk of imposition of new taxes for the citizens on a daily basis.

The day-to-day increase of taxation benefits the leaders as they develop the ability to serve their interests and intervene in warfare using the taxpayers' money (Feinschreiber & Kent, 2009). The argument here is that if the system makes it mandatory for leaders to pay taxes, there will be no over-wasting of public money by those in government. Moreover, if one party pays taxes while the other does not, this will aggravate the existence of corruption in the public service. For example, in the modern world, if leaders and their subjects, both pay taxes, development will increase, and there will be no conflicts between the two parties. There will be harmonious living, which will lead to a virtuous life. In the words of Aristotle, man is a political animal who requires a partnership in the right and just things to create a stable household and city (Ellis et al., 2006).

Property is not wealth. Wealth is an accumulation of property - a byproduct of having and utilizing property - and a constituent of household (Ellis et al., 2006). Taxation on wealth would be demoralizing as many people think of paying taxes as a punishment than a duty to nation building. Moreover, taxation on wealth would amount to double taxation, as property is usually taxed before accumulation. In addition, the government may lack a means of keeping account of an individual's wealth, bearing in mind that wealth is volatile. On the other hand, taxes on land and property have been considered to be especially appropriate as a revenue source because real property is immovable, the tax cannot be evaded, it is easy to administer and lastly the tax base is visible to everyone. In any case, according to Aristotle's theory of distributive justice, equality is achieved when the inferior is not living under insufficiency because of the taxation regime and consequently, the best taxation structure benefits all the citizens (Ross et al., 1996).

Concluding, according to (Brunori, 2007), property tax remains an important source of revenue for local governments, essentially because it meets all the characteristics of a sound local tax system.

It has long provided a stable and reliable source of revenue for local governments. Stability and reliability are considered as the two requirements for creating a sound tax system. Only a few taxes are more reliable than those on property. Unlike all other local taxes, the property tax base cannot be moved. The revenue from such a "captive" tax base is more reliable than revenue from either wage or sales taxes-both of which have highly mobile tax bases. Property tax has consistently grown over the years. Real property values appreciate over time; because of that, property tax revenues continuously grow. The tax allows local governments to collect consistent amounts of revenue during economic downturns, recessions, and periods of high inflation. Thus, property tax is uniquely positioned to meet future public service needs. Property tax is an important source of revenue because it meets the general requirements of a sound local tax system: it offers an efficient, stable, and fair method of collecting revenue.

2.3.2 Attributes of Real Property Tax

Property taxes often are underutilized sources of revenue. Their unpopularity, coupled with opposition from taxpayers who benefit from entrenched inequities encourages "legislative neglect"(Almy, 2013). The property tax is the largest source of tax revenue for local governments especially in developing countries (Almy, 2013; Dillinger, 1992). In order to fully understand the real property tax framework, the attributes of property tax should first be presented. Following, the classifications of property tax according to (EUROSTAT, 2013) and the advantages and disadvantages of the real property tax are analyzed. In addition, the necessity for the appropriate legal framework in property taxation and the indicators which lead to a successful tax legislation are being underlined. Furthermore, the rationale of the differences among EU property taxation systems is highlighted. This subsection leads to the next section regarding the EU property taxation regimes overview.

2.3.3 Classifications of property tax

'Property taxes' are usually associated with the notion of recurrent taxes on immovable property, but in practice a variety of levies on the use, transfer, and ownership of property are included (J. Norregaard, 2013).

In order to investigate the different tax regimes, a categorization of the property tax should first be conducted. Property tax is classified into six basic types: recurrent taxes on immovable property; recurrent taxes on net wealth; estate, inheritance and gift taxes; taxes on financial and capital transactions; other non-recurrent taxes on property and other recurrent taxes on property. Some of the above six types are subdivided into further categories (EUROSTAT, 2013).

2.3.4 The Advantages and Disadvantages of the Real Property Tax

McCluskey & Plimmer, (2011) consider that real property tax has been characterized as the "perfect" tax. It provides a predictable and reliable source of revenue for the local economy and it allows for decentralization and local autonomy. Indeed, property tax is most important for local administration (communities, municipalities etc.). It is a tax that is characterized by immovability which allows for clear directive as to which government is entitled to the tax revenue. As property is a tangible object it also provides a clear indicator of a form of wealth.

Bahl, (2009) identifies some additional advantages of real estate property tax, such as the stability and the high taxpayer compliance. Real property tax collections are relatively stable as property values are not as quick to react to general economic fluctuations such as taxpayers' income (tax base for income tax) or regular spending habits (tax base for sales tax). Another significant advantage of property tax is the high compliance rate and corresponding ease of enforcement. Failure to pay property taxes may result in a lien on the property and even a foreclosure sale. As a result, compared to other taxes, collection rates for the property tax are high, ranging often from ninety-two to ninety-eight percent.

Recently a robust and revived interest in property taxation has been observed globally. This is becoming obvious in varied reform initiatives recently adopted or being under consideration in several countries and in a recent influx in literature the focuses on developing and transition economies. The revived interest may derive from completely different motivation sources in several country groupings. For instance, one source of motivation could be to strengthen local democracy propulsion in some transition economies, whereas in several developing countries the prime motives are revenue mobilization and providing incentives for the better use of land. Finally, the property tax transparency causes political accountability that may improve the quality of the overall public financial system (Bahl, 2009).

Additionally, according to Norregaard, (2013) property taxation could help reduce the dependency of local governments in regards to transfers, enhancing economic efficiency through the strengthening of local responsibility. However, there are doubts regarding the results of the strengthening of local finances—for instance, through broadening of property taxation—on the improvement of the overall fiscal balance.

Unfortunately, property tax also has its disadvantages. For example, some critics view property tax as inequitable. One of the reasons for this inequity is that property tax is a regressive tax. A tax is considered regressive when low-income taxpayers pay relatively more of their income in property tax than wealthier taxpayers. The regressive nature is present because property tax, in its most basic form, is calculated on the value of real property without regard to taxpayers' income. Because property taxes are not based on the ability to pay, a taxpayer's property tax bill may increase even in the face of declining, or zero, income. Property taxes usually yield a comparatively modest revenue, particularly in developing and rising economies (Norregaard, 2013).

Furthermore, when property taxes are applied to businesses, explicit issues arise and attention is required. When a major factor of production is being taxed, it may raise costs disproportionately for businesses which use relatively more property as factor input. This is the reason why many countries use special reliefs for agriculture, applying lower tax rates or full or partial exemptions in some cases. It is possible, especially when the tax concerns property belonging to a business, that harmful tax competition among native governments is created. This is reason why various countries set narrow bands in the tax rates fluctuations (Bahl, 2009).

Melnik and Cenedella, (2008) believe that despite the inequity issues discussed above, the relative stability of the tax base coupled with the high taxpayer compliance has made property tax a preferred source of revenue for local governments.

Grover et al., (2017) discuss about the obstacles and the barriers the countries that have successfully introduced value-based recurrent property taxes have had to overcome. First, there are technical obstacles concerned with how to levy an efficient value-based property tax. These include the comprehensiveness of property registration, the quality and availability of transaction price data, whether internationally recognised valuation standards are followed and the quality of tax collection systems. Second, there are political or governance barriers which stand in the way of satisfactory technical solutions being implemented. Property taxes are widely seen as being unpopular, and work has to be undertaken to convince the public that they are fair and necessary. Moreover, they tend to lack champions in government. They require different skills from other taxes levied. Their roles in reducing intergovernmental fiscal transfers and in countering the impact of globalisation on revenue yields are not widely understood.

2.3.5 The necessity for the appropriate legal framework in property taxation

A question that arises frequently is, why it is so important to set an appropriate legal framework? A country that adopts competitive taxation policies manages to attract productive factors, funds and investments from other countries. The tax system applied in a country, has a serious impact on cross-country competitiveness. The differences and imbalances between EU countries reflect the different tax regime structures applied and this problem seems to also have a spatial characteristic that imposes a significant regional problem for the EU, and especially EMU countries, that already have a common currency and monetary policy. On the other hand, the mobility of productive factors is directly related to the country's tax-regime differences, government budget funding from tax revenues and rates, which are the main fiscal policy tools (Thalassinos et al., 2014).

The tax system in most of the EU countries is the ad valorem system. Norregaard (2013) highlights that for any recurrent ad valorem tax, assessment problems arise. These problems can be solved only with the appropriate and adequate legal framework for each country regarding mass appraisals. This kind of tax brings into focus political and legal issues concerning functional elements of the law. Each government, choosing a political current and the appropriate legislation to support it may use the asset taxation, through a recurrent ad valorem tax or in another form of tax, such as an estate tax, can mitigate the concentration of wealth that typically accompanies a regressive tax structure.

2.3.6 Bases for Property Taxation

There is no uniform property tax base or method of assessment that applies in every country. In some countries, the tax is only applied to land, while in a few countries only buildings constitute the tax base. In most countries, though, both land and buildings are taxed (Mccluskey et al., 2013). There are many bases for property valuation and taxation. These can be divided into two broad categories, value-based and non-value-based.

"Value-based systems include market or capital value, rental value, business value, and use or productivity value".

"**Non-value-based systems** can be based on a fixed amount per property or on area or a fixed formula. Some countries have both kinds of bases. A market value system provides estimates of value that maximize uniformity, fairness, transparency, and understandability. Alternatively, an area-based system can provide a reasonable substitute when a market-based system is not practical. Market value can be determined only if there is a mature property marketplace and valuers have access to sales data" (International Association of Assessing Officers, 2014).

2.3.6.1 Non-Value-Based Systems

The most common non-value property tax systems are those based on land area, building area, or both. Under an area-based property tax system, taxes are determined simply by multiplying a measurement of area by a rate. In general, area-based systems are suitable only as long as revenues are negligible (International Association of Assessing Officers, 2014).

Area-based systems have the advantage of being simpler to administer, as only property classification information and area measurements are required. They are easier to implement, as market data does not have to be collected or analyzed, and there is no need for general reassessments. They are also more objective than value-based systems, because area measurements are less contestable than value estimates. On the other hand, area-based property tax systems are quickly perceived to be less fair. Highly desirable properties may pay the same taxes as undesirable properties. Individual assessments bear little relationship to either the ability to pay or benefits received, and this reduces public acceptance. These disadvantages of an area-based system can be lessened by the introduction of adjustment coefficients that reflect market factors. However, doing so reduces the system's simplicity and objectivity. Although taxpayers might see this as an advantage, area-based property taxes are less buoyant than value-based systems, unless frequent adjustments are made to the rates (International Association of Assessing Officers, 2014).

2.3.6.1 Value-Based Systems

An ad valorem tax system is based on the principle that the amount of tax paid should depend on the value of the property owned (Eckert et al., 1990). When a measure of value is the basis for a property tax, there are several options: market value, restricted value, or some notional (or normative) value. Moreover, value can be on a capital-value or an annual-value basis (Almy, 2001). When the standard of value is market value, capital value is the price that would be expected in an open-market, arm's length sale. Annual rental value is the expected annual rent (or income). Annual rental value can be expressed on a gross or net basis. On a gross basis, the owner is assumed to be paying all operating expenses; on a net basis, the occupier is assumed to be paying (specified) operating expenses (such as repairs and insurance). Under either basis, actual rentals can be on a different basis, requiring valuers to make adjustments (International Association of Assessing Officers, 2014).

A standard of value other than market value can be employed. Such a standard can be current-use value, insurance value, or acquisition price. In practice, when market value is not the basis, tax values are usually described as only accidentally reflecting market-value patterns. That is, they simply result from the application of rules, base rates, and adjustment coefficients. Countries with advanced economies usually have systems that are at least nominally based on market values. Each basis has advantages and disadvantages of a theoretical and practical nature depending on the nature of land tenure patterns and on other features of the property tax system, while a country's property tax system can use more than one basis (International Association of Assessing Officers, 2014).

When market value is the basis for taxation, important issues include the rights to be valued, the valuation (or assessment) date, the basis of valuation (the legal standard of value) and the revaluation frequency (Almy, 2001).

2.3.7 Property tax as a tool for fiscal decentralization and local authority autonomy

The main part of this chapter has been already published at the Proceedings of the Third International Conference on Remote Sensing and Geoinformation of the Environment (T. Dimopoulos et al., 2015).

Dimopoulos et al, (2015) expand Charles Tiebout's work "A Pure Theory of Local Expenditures" (1956). While Tiebout states in his theory that a tax system would enable the local government to replenish the funds expended in the provision of public goods, the theory does not go deep enough to identify the nature of the tax regime that would be effective in replenishing the expended funds. The debate surrounding the effectiveness of taxation as an avenue for raising public funds is something that has led to divisions among economists. On the one hand, there are economists who argue that local authorities' taxation on the local residents' income or property will only result in mass capital exodus and occasion a decline in the local authorities' tax base. These economists point to states like Boston in order to illustrate the dangers associated with the use of taxation as an avenue for raising funds. Boston imposed high property taxes on high income neighborhoods and watched in horror as dwellers in those neighborhoods sold their homes and moved to other states with lower property taxes. In the meantime, the individuals who purchased these properties in Boston circumvented the high property taxes by developing apartments that would house a large number of low-income earners. The outcome was a dip in the area's property value and a subsequent dip in the local authority's tax base. Other economists also point out that it is easy for wealthy residents to migrate to other states when a local authority attempts to impose taxes on their income. In such a scenario, the high-income residents sell their properties and move to other areas, leading to the dwindling in the revenues that the local authority draws from taxes.

On the other hand, there is a group of economists that argues that it is possible for local authorities to use high taxes as a sustainable source of public funds. These economists argue that such an outcome is possible when the local authorities use zoning laws to increase the wealthy owners' switching costs in the aftermath of the introduction of a high property tax regime. Fischel, (2013), one of the chief supporters of the use of taxation as an avenue for revenue generation, contends that it is possible for local authorities to derive sustainable sources of funds from high property tax. He also argues that the main factor that led to the failure of the Boston property tax regime is the fact that the local authority did not introduce a policy that made it expensive for the wealthy residents to move to other states and that it is only through the imposition of such restrictions that the local authorities can prevent the deadweight taxation loss that arises after the imposition of high property taxes. According to Fischel, (2013), the deadweight taxation loss denotes the economic loss that the local authority suffers either as a consequence of the use of the available land in an inefficient way or as a consequence of the movement of economic activities to regions with low tax regimes, as well as zoning laws are the most effective means of eliminating deadweight taxation loss because they make it uneconomical for the targets of the high property taxes to move to a new state. If a state imposes high property taxes on neighborhoods with wealthy homeowners and imposes zoning laws that restrict conversion of those properties into apartments that would house the poor, the wealthy landowners will be forced to either look for other wealthy buyers to purchase their property or significantly reduce the cost of their property so as to enable property buyers to offset the high taxation regime. Similarly, an individual purchasing the property will do so with the knowledge that the zoning regime prevents them from using the property in a less efficient way (Fischel, 2013). Thus, a zoning law would be instrumental in ensuring that the local authority maintains the income levels in the affected neighborhoods even after raising the property taxes and that the tax-owing landowners do not view the law as mere ink on paper.

The effectiveness of zoning laws in preventing a dip in the value of properties as a consequence of taxation has also been affirmed by studies on how land policies can act as an incentive for influencing land-use patterns. Lewis and Plantinga (2007), in their study, conducted an analysis on how land policies can become an effective tool for influencing landowners to spatially configure their properties so as to achieve positive environmental outcomes, since private landowners are generally reluctant to coordinate their land-use decisions to achieve a greater good, in terms of the overall value of their properties or the general quality of the environment within which they live. According to Lewis & Plantinga, (2007), land-use policies can provide an incentive for private landowners to develop a system for effectively coordinating their decisions so as to achieve certain predetermined outcomes that will improve the overall value of their properties. When one relates this argument to the issue of local property taxes, it becomes clear that zoning laws can become an avenue that local authorities can exploit for ensuring that the value of their tax base does not dip. In fact, zoning laws can be the vehicle to improve the value of properties within a local authority. The local authorities can design these zoning laws in a manner that encourages spatial configuration of private landowners' decision-making and ensure that they use their properties in a manner that will improve their value in the long-term. Thus, the local authority will benefit immensely from the general improvement in the value of the properties subject to the high taxation regime.

2.4 Bibliometric analysis of 30 years of research in mass appraisals

The main part of this chapter has been already published at the International Journal of Real Estate & Land Planning (Thomas Dimopoulos & Bakas, 2019a).

The research papers issued in scientific journals, for a variety of thematic areas, are not only increasing, but also exhibit an exponential growth over the last years. Accordingly, the researchers, struggled to retrieve information apropos of novel knowledge and become informed in their field, while the rigor and the extensive composition of surveys, reviews, and overviews of research works, has become difficult or even impossible, as the number of the available research studies is enormous. However, such reviews, contain vital information regarding the evolution of a scientific subject, the trends of the literature, the most significant concepts, and the concealed associations among research papers, their references, as well as authors' clusters. In this work, a scientometric study of the relevant to Mass Appraisals literature is accomplished for a first time, regarding the numerical models, computational procedures, and automated methods, utilized in the Mass Appraisals and Property Valuations literature. The study is based on an adequate pool of papers, constituted in Scopus database, utilizing a machine learning algorithm developed by one of the authors, for multidimensional scaling and clustering of the keywords found in the papers' database, the authors and their cooperation and the co-occurrences of the references in the papers studied. The time-series of the most frequent keywords are also computed, demonstrating the evolution of the mass appraisals research and identifying future trends.

This work aims to review the mass appraisals related literature, by revealing the major computational methods used, the inter-papers associations of the technical concepts, the most influential authors, the major journals as well as their associations and evolution over time. This task demanded vast and complicated work - as will be clarified in the next paragraphs - since the research papers available in scientific journals has increased exponentially in recent years. Bornmann & Mutz, (2015) investigated the growth rates of scientific publishing and found it to be 1% up to the middle of the 18th century, to 2-3% up to the period between the two world wars, and 8-9% up to 2010. Accordingly, the development of international scientific output doubles every nine years on average, as demonstrated by Van Noorden, (2014). Hence, novel methods have been produced in the field of bibliometric analysis, called bibliometrics, scientometrics, scientific mapping etc., that utilize computer processes to analyze a massive amount of research papers. In this work, a relevant procedure developed by Plevris et al., (2017) has been used for the bibliometric mapping, as described in the following Table 2. The procedure is based on the contingency table which indicates the simultaneous appearances of pairs of keywords in a paper. This appearance indicates a thematic association of the two concepts, as expressed by the pair of keywords. By doing this for all the pairs of the keywords in the studied database, all the inter items associations are revealed. However, in order to accomplish a visual representation of these associations, the bibliometric map is constituted in Figure 8, where the distances of each two items correspond to their co-occurrence in the database. In particular, the lesser the distance (the closer the items on the map) the higher their simultaneous appearance in the database. Conclusively, the developed bibliometric map offers a generic image of the studied papers' database, as well as their thematic relationship.

Table 2: Bibliometric Mapping Algorithm

1. $c_{ij} := contingency\ table$ (co-occurrence of objects)

2. s_{ij}:= similarity

3. $ds_{ij} := \frac{1}{s_{ij}}$ (dis-similarity)

4. $d_{ij} = \|x_i - x_j\|$ (distance on map)

Optimization Algorithm

5. $f_{ij} := |ds_{ij} - d_{ij}|$ (objective function)

6. Optimality criteria satisfied? NO

YES

7. End => drawing of the bibliometric map

2.4.1 Papers obtained by Scopus query

An initial search in the Scopus database was performed, utilizing the two basic keywords *Mass Appraisal* and *Property Valuation* with 'OR' condition and simultaneously targeting more specific secondary keywords (MSK): *CAMA, Neural Networks, Geographically Weighted Regression, Regression Analysis, Automated Valuation,* *Spatial Analysis, Computer Aided, Computer Assisted and Machine Learning*. The secondary keywords (Figure 4) were identified manually in a literature examination by the authors and they were separated with 'OR' condition as well. At that point is should be stated that this is an initial investigation, future research could scrutinize in detail all the obtained papers not only from the Scopus database, but within other sources such as web of science etc.

Figure 4: Basic search and initial database

The resulting number of papers in Scopus were 123 relevant documents that were manually checked by the authors and approved as relevant. Still, the number of documents was not satisfactory enough for a proper bibliometric analysis. Accordingly, to expand the database, further research accomplished based on the main authors. Because in Scopus and any scientific database, some authors appear with small differences in their names, this further investigation was crucial for the construction of a larger and more complete database. These searches were executed in the title, abstract, and keywords in the Scopus fields (*TITLE-ABS-KEY*).

Hence, the second step was to use the keywords *CAMA, Geographically Weighted Regression, Automated Valuation, Spatial Analysis, Mass Appraisal, and Property Valuation*, together with each one of the authors' names. We excluded the generic keywords such as *neural networks* and *regression* because Scopus yielded irrelevant results from other thematic areas. Specifically, the authors' names were not used with their first name, but only the last name, which together with the keywords, should give correct results. The authors with the most document were: Lasota, Trawiński, Telec, Kauko, D'Amato, Davis (which results with this procedure were irrelevant and thus excluded), Haran (which results were selected manually), Dimopoulos (manually), Kempa (manually), McCluskey (with a lot of results which all were included), Borst (all relevant).

Consequently, a second search was performed in Scopus, utilizing the keyword *Real Estate* into the first part of the query, together with *Mass Appraisal* and *Property Valuation*, with OR statement among them. This way, Scopus yielded 935 research documents, with the top ten keywords, *Regression Analysis, Real Estate, Spatial Analysis, Neural Networks, Housing, Costs, Investments, Commerce, Housing Market* and *Forecasting*, which seemed to be relevant to the field under study. However, a more careful investigation of the titles and abstracts of the most cited papers of the resulting database, exhibited an approximate 30% of irrelevant papers and another 10% in the general thematic area of Real Estate, however dealing with Economics, Commercial etc, which are not directly relevant to the mass appraisals literature. A variety of transformations of the query was scrutinized, in order to accomplish a formulation of the query with consistent papers. Accordingly, when the word *Hous** (with * indicating any other character, such as *ing*, *e*, etc) was incorporated into the *OR Real Estate* part of the query, then the resulting (402) documents were relevant to the studied topic. Finally, the keyword Female was also excluded from the database, as it obviously referred to medical or social sciences papers. The structure of the query is demonstrated in Figure 5.

Figure 5: Refined search in Scopus database

Hence, the above-mentioned procedure produced a database with 486 research papers. The papers are relevant to the studied topic and at the same time adequate in terms of size, in order to produce generative conclusions. In Figure 7, the frequencies of the top twenty keywords, occurred in the Scopus database are demonstrated. The most frequent, were the keywords *property valuation*, *real estate*, and *mass appraisal*, indicating the consistency of the database and the accurate selection of the papers. Accordingly, the top twenty keywords can be classified into two categories, the methodological ones (property valuation, real estate, mass appraisal, housing etc.) and some technical / data analysis keywords, such as *GIS, spatial analysis, geographically weighted regression, bagging, artificial neural networks,* as classified in Figure 5.

property valuation, real estate, mass appraisal, housing, housing prices, real estate market, real estate appraisal, housing market, valuation	gis, spatial analysis, geographically weighted regression, bagging, artificial neural networks, neural networks, genetic fuzzy systems, data stream, trend functions, regression analysis, machine learning

Figure 6: Methodological and technical keywords

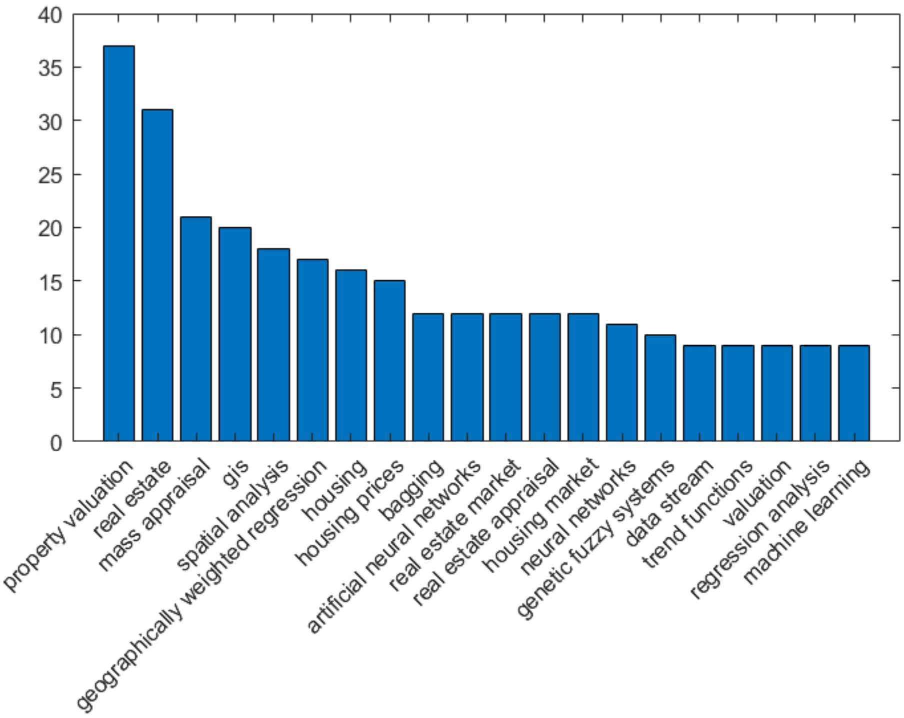


Figure 7: Frequencies of top twenty keywords in the Scopus database

2.4.2 Bibliometric results

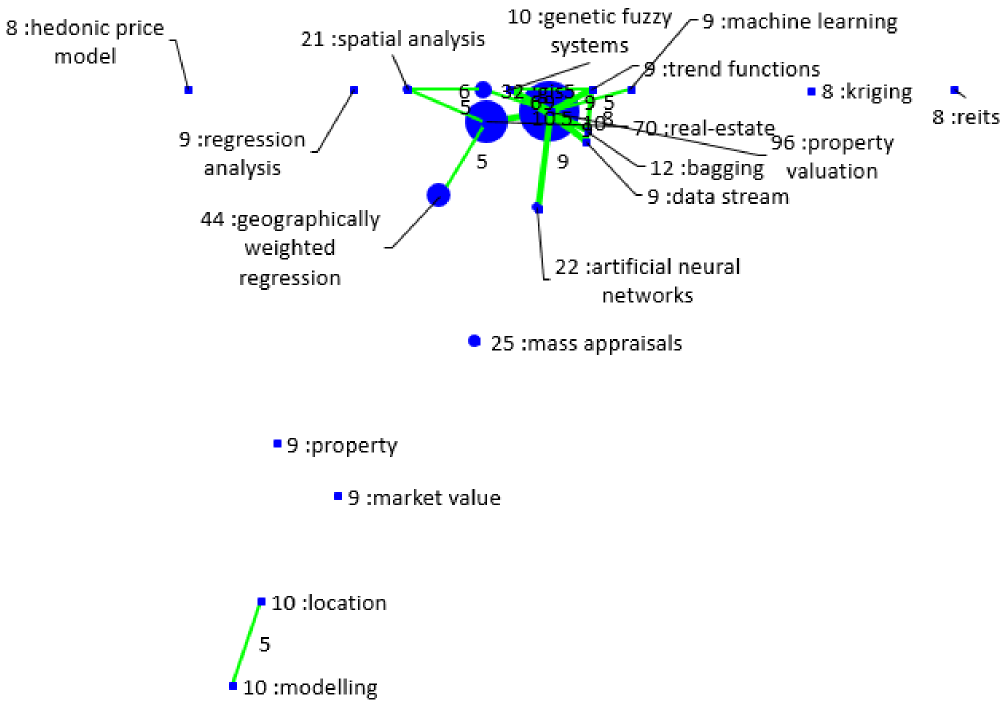


Figure 8: Bibliometric map of the top twenty keywords

In Figure 8, the twenty most frequent keywords are depicted. The distances on the keywords map represent their co-occurrences, that is to say the number of times they exist at the same time in a research paper. The extent of the circle (representing a keyword) is proportional, and corresponds to, the number of times this keyword exists in a research paper. From Figure 7, it is derived that the most frequent keyword is “property valuation”, with 96 occurrences (in the top right side of the map), followed by the keyword “real estate”, with 70 occurrences. Accordingly, the closest (hence the more related keywords) are “spatial analysis”, “geographically weighted regression”, “machine learning”, “artificial neural networks”, “trend functions”, concerning mainly data analysis methods, numerical modelling, spatial analysis and predictions. The exact numbers of the keywords occurrences and co-occurrences are demonstrated in the following Table 3. The research papers issued in scientific journals, for a variety of thematic areas, are not only increasing, but also exhibit an exponential growth over the last years. Accordingly, the researchers, struggled to retrieve information apropos of novel knowledge and become informed in their field, while the rigor and the extensive composition of surveys, reviews, and overviews of research works, has become difficult or even impossible, as the number of the available research studies is enormous. However, such reviews, contain vital information regarding the evolution of a scientific subject, the trends of the literature, the most significant concepts, and the concealed associations among research papers, their references, as well as authors' clusters. In this work, a scientometric study of the relevant to Mass Appraisals literature is accomplished for a first time, regarding the numerical models, computational procedures, and automated methods, utilized in the Mass Appraisals and Property Valuations literature. The study is based on an adequate pool of papers, constituted in Scopus database, utilizing a machine learning algorithm developed by one of the authors, for multidimensional scaling and clustering of the keywords found in the papers' database, the authors and their cooperation and the co-occurrences of the references in the papers studied. The time-series of the most frequent keywords are also computed, demonstrating the evolution of the mass appraisals research and identifying future trends.

This work aims to review the mass appraisals related literature, by revealing the major computational methods used, the inter-papers associations of the technical concepts, the most influential authors, the major journals as well as their associations and evolution over time. This task demanded vast and complicated work - as will be clarified in the next paragraphs - since the research papers available in scientific journals has increased exponentially in recent years. Bornmann & Mutz, (2015) investigated the growth rates of scientific publishing and found it to be 1% up to the middle of the 18th century, to 2-3% up to the period between the two world wars, and 8-9% up to 2010. Accordingly, the development of international scientific output doubles every nine years on average, as demonstrated by Van Noorden, (2014). Hence, novel methods have been produced in the field of bibliometric analysis, called bibliometrics, scientometrics, scientific mapping etc., that utilize computer processes to analyze a massive amount of research papers. In this work, a relevant procedure developed by Plevris et al., (2017) has been used for the bibliometric mapping, as described in the following Table 3. The procedure is based on the contingency table which indicates the simultaneous appearances of pairs of keywords in a paper. This appearance indicates a thematic association of the two concepts, as expressed by the pair of keywords. By doing this for all the pairs of the keywords in the studied database, all the inter items associations are revealed. However, in order to accomplish a visual representation of these associations, the bibliometric map is constituted in Figure 10, where the distances of each two items correspond to their co-occurrence in the database. In particular, the lesser the distance (the closer the items on the map) the higher their simultaneous appearance in the database. Conclusively, the developed bibliometric map offers a generic image of the studied papers' database, as well as their thematic relationship.

Table 3: Co-Occurrences of keywords

Contingency Table	property valuation	real-estate	geographically weighted regression	gis	mass appraisals	artificial neural networks	spatial analysis	bagging	location	modelling	genetic fuzzy systems	market value	property	machine learning	regression analysis	data stream	trend functions	reits	kriging	hedonic price model
property valuation	96	10	3	6	2	9	3	10	0	0	9	1	1	5	2	9	9	1	0	0
real-estate	10	70	5	2	1	1	5	0	0	1	0	0	2	0	3	0	0	1	0	2
geographically weighted regression	3	5	44	2	3	2	3	0	1	0	0	0	0	0	0	0	0	0	2	0
gis	6	2	2	32	1	0	6	0	1	1	0	0	1	0	0	0	0	0	0	1
mass appraisals	2	1	3	1	25	3	1	0	3	0	0	1	0	1	0	0	0	0	1	1
artificial neural networks	9	1	2	0	3	22	0	0	0	0	0	2	0	0	0	3	3	0	0	1
spatial analysis	3	5	3	6	1	0	21	0	0	0	0	0	0	0	0	0	0	0	1	0
bagging	10	0	0	0	0	0	0	12	0	0	4	0	0	0	0	0	0	0	0	0
location	0	0	1	1	3	0	0	0	10	5	0	0	2	0	0	0	0	0	0	0
modelling	0	1	0	1	0	0	0	0	5	10	0	2	2	0	0	0	0	0	0	0
genetic fuzzy systems	9	0	0	0	0	0	0	4	0	0	10	0	0	0	0	5	5	0	0	0
market value	1	0	0	0	1	2	0	0	0	2	0	9	0	0	0	0	0	0	0	1
property	1	2	0	1	0	0	0	0	2	2	0	0	9	0	0	0	0	0	0	0
machine learning	5	0	0	0	1	0	0	0	0	0	0	0	0	9	0	0	1	0	0	0
regression analysis	2	3	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0
data stream	9	0	0	0	0	3	0	0	0	0	5	0	0	0	0	9	8	0	0	0
trend functions	9	0	0	0	0	3	0	0	0	0	5	0	0	1	0	8	9	0	0	0
reits	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0
kriging	0	0	2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	8	0
hedonic price model	0	2	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	8

Accordingly, the timeseries of the keywords are demonstrated in the following Figure 9. The vertical axis represents the frequencies of the keywords per year, and the horizontal axis represents the years studied and in the legend the keywords with various colors. A high increase for the keywords "property valuation", as well as "real estate" is exhibited after the year 2010. The final years (2015, 2016 and 2017) a strong decrease was observed, however it cannot be considered as reliable, as the relevant articles might not be included in Scopus yet.

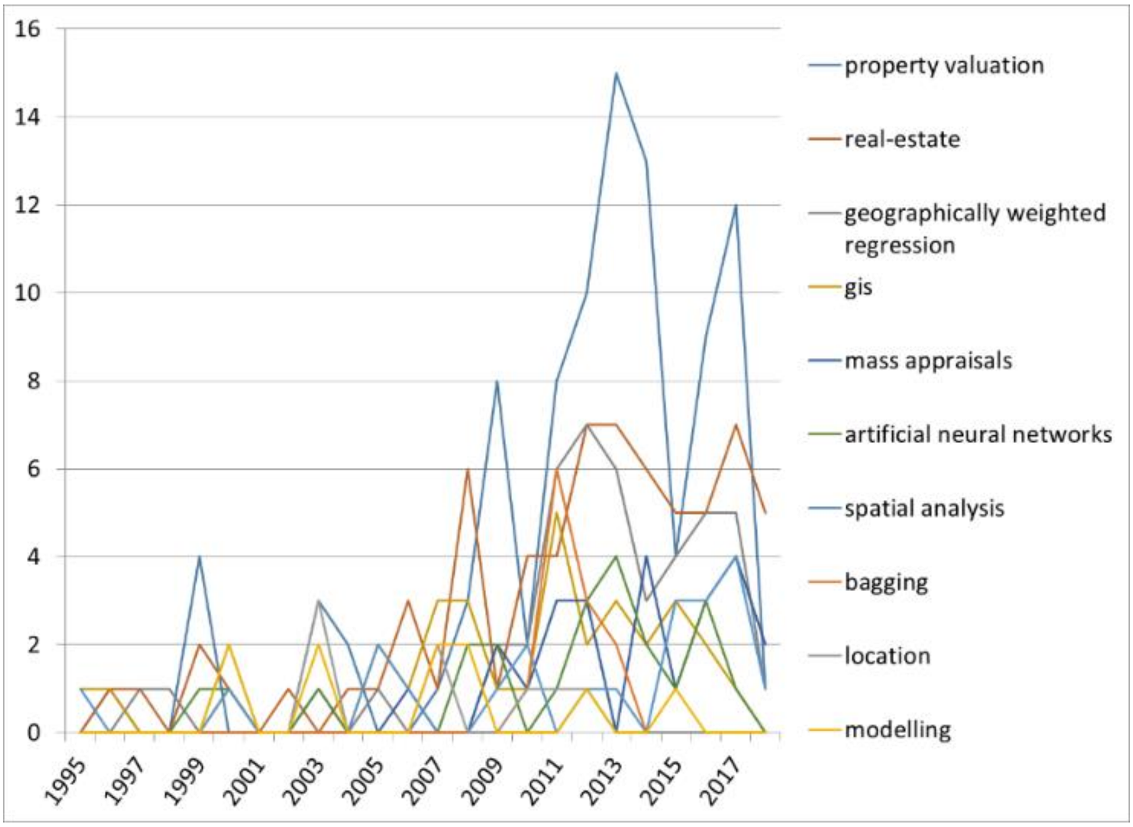


Figure 9: Keywords Occurrences Time-series

In Figure 10, the Bibliometric map of the twenty highest publishing authors is demonstrated. Each author is represented with a circle, while the radius of the circle is proportional to the number of articles published. The exact number of the published articles for each author is written with a numeric value before the name of the author. Some of the circles are linked with a green line, indicating that these authors have cooperated in scientific papers. The number of common papers, is indicated with a number in the middle of the line. The closer two circles are, more shared papers these authors have written.

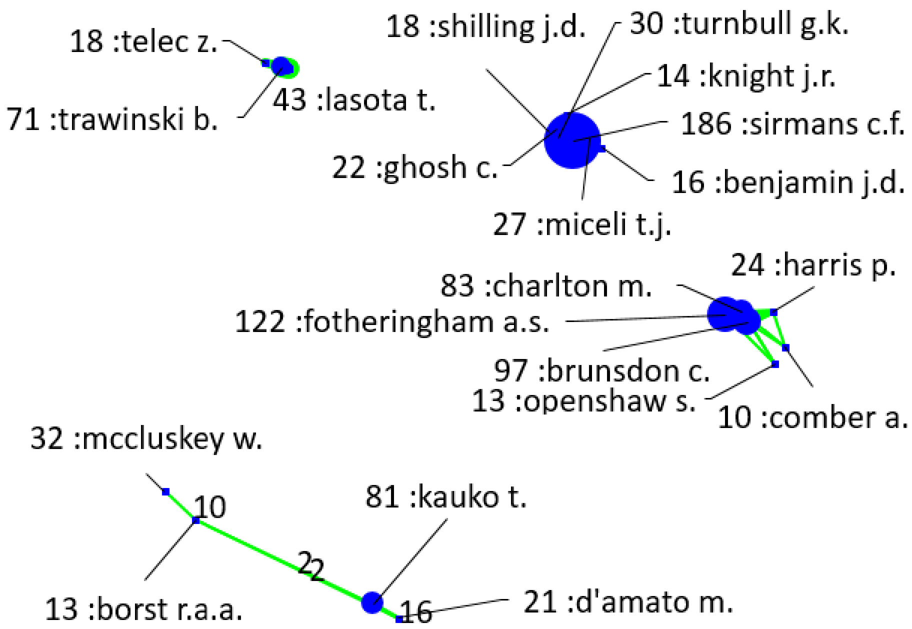


Figure 10: Bibliometric map of the top twenty authors in the Scopus database

Accordingly, in Figure 9, four clusters are constructed, one with authors McCluskey, Borst, Kauko and D' Amato, One with Fotherigham, Brunsdon, Charlton et. al., one with Trawinski, Lasota and Telec, and finally one with Sirmans, Turnbull, Ghosh et. al. The exact number of each author's papers as well as the co-authorship papers, are demonstrated in the following Table 4.

Table 4: Co-Authorship Matrix

Co-Occurances	sirmans c.f.	fotheringham a.s.	brunsdon c.	charlton m.	kauko t.	trawinski b.	lasota t.	mccluskey w.	turnbull g.k.	miceli t.j.	harris p.	ghosh c.	d'amato m.	shilling j.d.	telec z.	benjamin j.d.	knight j.r.	borst r.a.a.	openshaw s.	comber a.
sirmans c.f.	186	0	0	0	0	0	0	0	30	27	0	22	0	18	0	16	14	0	0	0
fotheringham a.s.	0	122	19	28	0	0	0	0	0	0	9	0	0	0	0	0	0	0	2	0
brunsdon c.	0	19	97	33	0	0	0	0	0	0	17	0	0	0	0	0	0	0	2	10
charlton m.	0	28	33	83	0	0	0	0	0	0	18	0	0	0	0	0	0	0	9	1
kauko t.	0	0	0	0	81	0	0	0	0	0	0	0	16	0	0	0	0	2	0	0
trawinski b.	0	0	0	0	0	121	68	0	0	0	0	0	0	0	28	0	0	0	0	0
lasota t.	0	0	0	0	0	68	43	0	0	0	0	0	0	0	18	0	0	0	0	0
mccluskey w.	0	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0	0	10	0	0
turnbull g.k.	30	0	0	0	0	0	0	0	30	7	0	0	0	2	0	4	1	0	0	0
miceli t.j.	27	0	0	0	0	0	0	0	7	27	0	0	0	0	0	0	0	0	0	0
harris p.	0	9	17	18	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	1
ghosh c.	22	0	0	0	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	0
d'amato m.	0	0	0	0	16	0	0	0	0	0	0	0	21	0	0	0	0	2	0	0
shilling j.d.	18	0	0	0	0	0	0	0	2	0	0	0	0	18	0	6	0	0	0	0
telec z.	0	0	0	0	0	28	18	0	0	0	0	0	0	0	18	0	0	0	0	0
benjamin j.d.	16	0	0	0	0	0	0	0	4	0	0	0	0	6	0	16	0	0	0	0
knight j.r.	14	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	14	0	0	0
borst r.a.a.	0	0	0	0	2	0	0	10	0	0	0	0	2	0	0	0	0	13	0	0
openshaw s.	0	2	2	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0
comber a.	0	0	10	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	10

Finally, in Figure 11, the timeseries of the papers published by the highest number of publications by one author is demonstrated. Two peaks were found, Sirmans with twenty papers in 2010 and Trawinski with 18 in 2012.

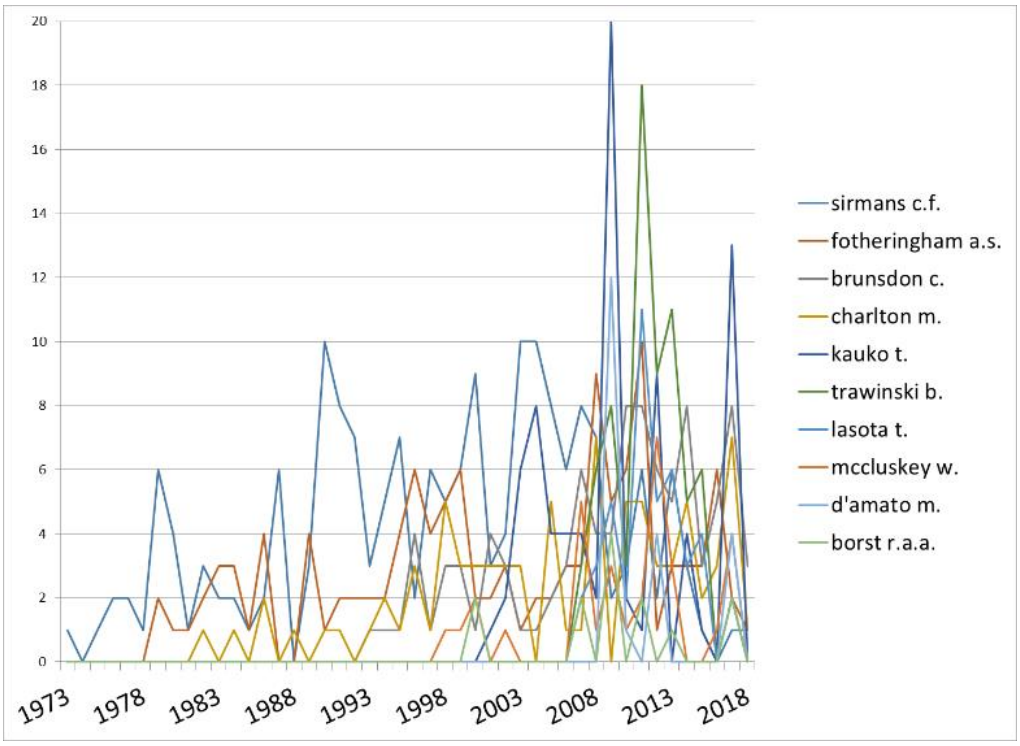


Figure 11: Authors time-series

2.4.3 Contribution of the thesis

Another more general search as it appears in the figure below (Figure 12) was made with some minor keyword adjustments in order to enhance the database of papers. The query resulted 735 papers.

Figure 12: Enhanced search in Scopus database

The papers were listed according to the author's affiliation country. The results are presented in the Table below (Table 5). Cyprus is ranked 18th with 10 publications. It should be stated that the six out of ten publications were authored or co-authored by Thomas Dimopoulos (Dimopoulos et al., 2014; Dimopoulos & Moulas, 2016; Dimopoulos et al., 2018a; Dimopoulos & Bakas, 2019; Dimopoulos et al., 2015; Dimopoulos & Bakas, 2019b). The above-mentioned investigation that refers to the globally available literature on Mass Appraisals highlights the scientific challenge, as well as the contribution of this thesis. Dr Demetris Demetriou has also contributed significantly to the field with three publications (Demetriou, 2017, 2018, 2015) that specialise in land consolidation, and one other publication dated in 2000 (Panayiotou et al., 2000) that was co-authored from Panayiotis Panayiotou, Prof. Frances Plimmer and others and tests ANN and MRA for the mass appraisal of properties in Strovolos Municipality for taxation purposes. 25 of the publications could not be categorised.

Table 5: Authors affiliation origin

#	Country	Papers	#	Country	Papers	#	Country	Papers
1	China	131	23	France	7	45	Jordan	2
2	United States	97	24	Turkey	7	46	Albania	1
3	United Kingdom	46	25	Austria	6	47	Croatia	1
4	Italy	45	26	Colombia	6	48	Denmark	1
5	Poland	42	27	Sweden	6	49	Ghana	1
6	Spain	26	28	Chile	5	50	Iraq	1
7	Australia	24	29	Singapore	5	51	Ireland	1
8	Hong Kong	19	30	Czech Republic	4	52	Kazakhstan	1
9	Germany	18	31	Israel	4	53	Latvia	1
10	Canada	16	32	Lithuania	4	54	Malta	1
11	Netherlands	15	33	Nigeria	4	55	Mexico	1
12	Taiwan	15	34	Romania	4	56	Pakistan	1
13	South Africa	14	35	Slovenia	4	57	Peru	1
14	Greece	12	36	Ecuador	3	58	Saudi Arabia	1
15	India	12	37	Finland	3	59	Slovakia	1
16	Japan	12	38	Portugal	3	60	Somalia	1
17	Russian Federation	11	39	Serbia	3	61	Sri Lanka	1
18	Cyprus	10	40	Switzerland	3	62	Tanzania	1
19	Brazil	9	41	Belgium	2	63	Thailand	1
20	Norway	9	42	Hungary	2	64	Ukraine	1
21	Malaysia	8	43	Indonesia	2	65	Viet Nam	1
22	South Korea	8	44	Iran	2	66	Undefined	25

It appears that the leading country is the China and the USA follows. The leading European countries are the United Kingdom, Italy and Poland. The results of Table 5 are graphically presented in Figure 13 below.

Figure 13: A world map that represents the authors origin.

2.5 Advantages, Criticism, and Constraints of AVMs

2.5.1 Advantages of Mass Appraisals

Tretton and Walt (2016) mentioned and described the different advantages that AVMs have:

1. Full transparency and public access facilitated. There has been an increase in the demand from the pubic for more information and for immediate access to date. Increasing the amount of information available "online" not only has its own justification but also increases public confidence in the integrity of the assessing body and its valuation.
2. Low cost. Revaluations will be significantly cheaper when done using an AVM model compared to when it is done as new valuation using traditional valuation process.

3. Consistency. In property taxation there has long been a conflict or potential conflict between accuracy and uniformity. Uniformity of assessment is highly desirable with a property tax as the principle of fairness is very much bound up with the acceptability of the tax to the public. AVMs deliver a high degree of consistency.
4. Speed. The more automated an assessment process becomes, the quicker a revaluation can be done.
5. Annual revisions are possible. The period between revisions varies greatly between jurisdictions. There is however a trend that revision periods become shorter.

2.5.2 Criticism on Mass Appraisals

Considerable criticism in the USA can be found of AVMs used by commercial companies for loans rather than those used for property taxation:

1. Concern that the public does not appreciate the difference between an automated valuation and one which involves an inspection, an appreciation of condition and a careful examination of comparable evidence close to the property. The Appraisal Institute 2002 Position Paper (Tretton, 2007) saw a danger of automated valuation being misrepresented as being equivalent to a professional appraisal. It wants the public to be property informed of the differences and calls for the two types of valuation to be clearly distinguished.
2. The use of outdated or very limited data. A recurring theme being the lack of data available outside the public sector. The number of data is crucial for the successful implementation.
3. Failure to take account of all the variables affecting value- a lack of individual inspection of the property. It appears that there are some AVMs in use seeking to achieve more than they are realistically capable of and using less than satisfactory or sufficient data.
4. There is great re-assurance in having a "real person" undertake the valuation. The public will be content with the use of AVMs providing they have the re-assurance that an expert person controls the program and the outputs are reviewed and refined by the expert. There is concern at flexibility in automated

systems. Where there is a large market for fairly homogenous properties such as is often the case with residential, then it is possible to set up a process to value such properties. This is more difficult when there are few properties or where the market is limited.

2.5.1 Main constraints of AVMs

Tretton (2007) concluded that: "Automated valuation programs assist in the production of valuation but its quality and accuracy are data and valuer led. One size does not fill all and there is no automated replacement for the subjective professional judgement of the valuer".

Gilbertson & Preston (2011) commented: "The fact that a valuer has little if any input is seen as a double-edged sword. It eliminates human error and bias but takes out of the equation not only the physical property inspection but also the skill, judgement and experience of the valuer".

Although AMVs have a lot of advantages and they can be used in several areas, however, there are some limitations. Robson and Downie (2008) referred to specific constraints on AVM use.

1. The need to inspect property: AMVs can only comprehend information about a the internal and external condition and the state of repair of a property, if it is inspected and the data is fed into the model. Value determinants such as orientation and aspect can be improved by including photographs and mapping information.
2. Data limitations: AVMs require accurate and comprehensive data that is collected consistently over a period. They can only provide accurate results if an adequate amount of sales or value data is available. They provide the most accurate results when they are used in conventional situations, such as in stable neighborhoods at where property prices near the median for the locality. The least accurate results occur when there are incomplete data records, when a low number of sales is recorded in the locality under study and when the properties under study are located in unique local markets. Problems also arise when trying to model purchasers' preferences for nonphysical property

characteristics such as views, gardens and sunshine, which are mentioned in literature.

3. Risk acceptance: The main obstruction to the further use of AVMs is attention over inaccuracy. When dealing with new loans that that have a high loan to value ratio, it is unlikely that an AVM will be used as a high level of accuracy is required. But, when dealing with a solid potential customer that has the ability to repay a loan, and where the property has been inspected already, as for second mortgages, accuracy is less critical, therefore AVMs may be judged acceptable despite the concern. The traditional valuation approaches are considered to be slow and expensive and are being replaced with time and cost-effective AVMs. Although the advantages to AVMs are appealing, banks are still required to provide wise loan decisions that they are liable for. Therefore, a combination of AVM confidence scores with credit and capacity assessment and Loan to Value (LTV) ratios is required. Banks are required to create own securitization guidelines, while also complying with rules and standards, in an attempt evolve constantly and follow market changes.

Thompson wrote that the phrase “garbage in /garbage out” captures the key message that the quality of the values produced is directly impacted by the quality of the data which are analysed to produce the value estimates (Kauko & Amato, 2008).

3 Mass Appraisal in Cyprus (General Valuation)

The author would like to acknowledge the assistance given for the authorship of this Chapter by Mr. Varnalas Pashoulis, Senior Valuation Officer at Department of Lands and Surveys and person in charge for the last two General Valuations at 1.1.2013 and 1.1.2018.

3.1 General Valuation 01.01.1920

The first coherent General Valuation in Cyprus dates back to 1920, when the first legislation was introduced under the Immovable Property (Registration and Valuation) Law, No. 12/1907. This law was introduced in order to carry out a compulsory registration and valuation of all properties on the island for the purpose of imposing property tax as well as to raise revenue through transfer fees. The General Registration on the island was completed in 1929 and it was followed by a Cadastral Survey. By law, every property had to be locally inspected. The G.V. was based on the capital value of each property (land and buildings).

After the completion of the new cadastral survey work on the island, each property was registered, along with the name its corresponding owner in a record known as the property register. At the same time, a valuation of each property was carried out and the value was recorded on every registration of title (Title Deed). The G.V. dated 01.01.1920 was carried out within, approximately, a 20-year period (1909-1929).

3.2 General valuation 01.01.1980

A new revolutionary legislation was introduced on 01.09.1946 in an attempt to resolve the so called defective ‘land tenure system’ namely, the Immovable Property (Tenure, Registration and Valuation) Law, Cap. 224. Under this law, special provisions were also enacted to improve the overall effectiveness and efficiency of the General Valuations. A new general valuation dated 01.01.1980 was ordered by the Council of Ministers in 1982 based on the new provisions of the aforementioned law and covered only government controlled areas. The valuation was based on the market value of the properties and this covered both land and buildings. It was carried out manually with no assistance of any computerized mass appraisal system, although statistical methods were applied. All types of properties were valued but different models were applied for different properties. For example, for residential properties the following formulae was applied:

$$V = D.F * (L + C + S) * area + P \qquad [3.1]$$

Where V = Value of property

DF = Development Factor (Developer's Remuneration or Profit and Risk)

L = Land Value per sq.m or density value per sq.m for horizontally divided units

C = Cost of Construction excluding heating and cooling (value per sq.m)

S = Services comprising of heat and cooling only (value per sq.m)

P = Parking Space

Regarding the building value of houses, the value per room was applied as well as the age (new, middle age, old, obsolete), condition and other deficiencies to adjust the value accordingly. Also, a similar model was applied for hotel valuations. Furthermore, for shops, the valuation was based on ITZA (In terms of Zone A) and for agricultural land, a multiplicative model was applied. As a general principle, the valuations were based on the analysis of comparable sales of similar properties within each geographical area that were completed close to the date of the G.V., 01.01.1980.

In terms of human resource engagement, about 50 employees were required in order to accomplish this task and it was completed in 12 years. Every property was inspected and the property characteristics and valuation calculations were recorded in the specified forms (N314).

3.3 General valuation 01.01.2013

The Cyprus Integrated Land Information System (CILIS) has been in operation since 2000 and it consists of three main systems, specifically, the legal, the fiscal and the GIS. Regarding the fiscal system, among other valuation systems, a computer assisted mass valuation system (CAMAS) was developed to assist the Department of Lands and Surveys to implement general valuations on a frequent basis, as and when they were ordered by the Council of Ministers. Due to the economic crisis and the bail in, the Cyprus government had to sign a memorandum of understanding (MOU) with EU/ECB/IMF in order to be able to get financial assistance, under measure 3.8, and a new

G.V. was ordered by the Council of Ministers as one of the many measures that had to be met. At that point in time the Department had about a year to accomplish this extremely difficult project. The major obstacle in accomplishing this task was that there was a lack of property characteristics in CILIS (only 20% of the total was available). The total number of parcels of land that needed to be valued was about 1.3 million, as well as 0.5 million buildings (433,000 residential and 67,000 commercial units). The total G.V. of all the properties was estimated at €202,618,669,610. The project was organized by breaking down the tasks into two subprojects namely the Data Capture Project and the G.V. Project. Each subproject was staffed with specific number of employees at district level and at central level two valuation officers were responsible, one to coordinate and supervise the Data capture Project and the other to coordinate and supervise the Valuation Analysis and Mass Valuation. The whole project was under the responsibility of two Senior Officers (General Valuation and Valuation Sections).

In order to accomplish the data capture project within a time span of a year, the Department had to also mobilize the District Town Planning Offices as well as the Municipal Authorities and District Administrations who are responsible for issuing building permits (District Administration is responsible for issuing all building permits for the countryside). All information made available by these authorities, was tabulated into the CAMAS in order to be able to perform a new general valuation dated 01.01.2013. Information, which was not possible to collect from these authorities, was undertaken by employees of the Department. Extensive satellite imagery and google earth observations were used in combination with existing data available in the legal/fiscal database by superimposing data and detecting unrecorded units for specific parcels. Furthermore, GIS tools were applied to capture land characteristics and thereafter were mass updated into the CAMAS. Also, the VISAT software in combination with GIS was also used providing at the street level of the property for information in 21 villages, a similar product to Google street map.

The 2013 G.V. was based on the same law as described in the 1980 G.V. but a major amendment was made by introducing a definition regarding the value of general valuation rather than the market value. Under S2 of Cap.224, “Value of G.V.” in relation to immovable property means “the amount which results from performing a G.V or revaluation or revising a G.V., which is as close as possible to the value" (value means market value).

The following property characteristics were used for the 2013 mass appraisal:

Land Characteristics	Unit Characteristics
• Type of Property • Area (sq.m) • Planning Zone • Location • Kind of access • Relation to the road • Shape	• Type of Unit • Area of Unit (sq.m) • Year of construction (Depr. Factor) • Year of substantial renovation • Category (luxury, A, B, C,D) • Condition (V. good, good, fair, bad) • View

Regarding the valuation methodology, valuation parameters were determined by geographical area (600), planning zone (2.226), location and property type of a predefined standard property. The basic rule was to use at least three comparable sales in determining the valuation parameters. These parameters were imported into the CAMAS and the base model was applied for all properties in order to perform general/mass valuations. The base model is based on the equation of MV= Lvalue + Bvalue*Development Factor. Separate parameters were estimated for land and buildings by analyzing comparable sales. Special properties such as golf courses, marinas/ports, airports, wind generators, photovoltaic parks, shopping centers, thematic parks, petroleum tanks and exceptional buildings were valued as single valuations, manually. Quality control (ratio studies-Median, COD, PRD) tests were performed regarding the uniformity of values as well as to ensure that horizontal and vertical equity is achieved to meet the compliance of IAAO standards, before the publication of the results. Following publication of the results, the Department applied a specific communication and public relations strategy in order to inform the public as well as to deal with public enquiries and objections. After the publications of values, the Department processed about 40.000 applications for correction of errors regarding property characteristics and about 1.700 applications for valuation objections. Correction of error applications were free but for valuation objections there was an administrative progressive fee and each application had to be annexed with a valuation report.

3.4 General Valuation 01.01.2018

There was a radical change in the methodology of Data Capture of property characteristics for updating the valuation base due to the development of the DLS Portal (Web application) that electronically interconnects all building authorities issuing building permits (Municipal Authorities, District Administrations). This application enables automatic updating of the Department's Database, both for new developments as well as for amendments to existing ones. Regarding the valuation approach, the same base models were applied to value all types of properties as in 2013, which are summarized below:

Land Models	Building Models
• Base Residential/Commercial building site • Base Industrial building site • Base Undeveloped field • Base Agricultural/Livestock fields	• Base Residential • Base Commercial • Base Industrial • Base Hotels / Tourist Establishments • Base Schools/Hospitals/Clinics • Base Livestock Units

The 2018 G.V. was refined and improved compared to the 2013 G.V. by adding more property characteristics such as the "Slope", "view", "subsoil", "High Voltage Tension Lines", "Foreshore Protection Zone". Furthermore, it became possible to perform automated valuations for parcels that lie within more than one planning zone. In addition, the valuation models were upgraded to be able to perform quantity adjustments for land extent. Also, it became possible to create more planning zone categories under the main ones, thus achieving a much higher degree of accuracy as these parameters were used to adjust the base value (constant) of the properties, according the property characteristics of each.

Another profound change of the 2018 G.V. was the participation of private valuers in the process of determining the land valuation parameters for all municipal administrations on the island. This was made possible through public tendering and the external valuers with their market knowledge and expertise worked together with the respective valuers employed in the Department. The private sector valuers also assisted the Department in tackling specific issues and market inconsistencies that do not follow the usual patterns of the property market.

It was estimated that the total number of properties valued for the 2018 G.V. were about 1.5 million land parcels and 525,00 buildings (458k residential, 67k commercial). The value of all the properties for 2018 was estimated at €178,779,262,000. Quality control (ratio studies-Median, COD, PRD) tests were performed regarding the uniformity of values as well as in order to ensure that horizontal and vertical equity is achieved to meet the compliance of IAAO standards, before the publication of results. A short-cut summary of quality results (ratio study) is available on the website of the Department.

It is worth noting that on 22 valuation objections were lodged over the period of six months period under the current legislation. Regarding the correction of error applications, these were also very low, approximately 200. The low number was due to the majority of property information being corrected during the process of the previous G.V. It is important to note that the lodging of both applications was possible only electronically through the DLS Portal. The burden of cost (time and administration) was minimized for both citizens and the Department. All the values of the general valuation are freely available to be viewed on the DLS Portal of the Department as well as their respective property characteristics.

4 Analysis of the apartments database

4.1 Limitations to the availability of data and data format

The database used in this thesis was provided by the DLS. The database was an enhanced database from what is normally provided to registered valuers that includes some additional information regarding the sales transactions, including condition, age, view etc. as it is analytically explained in Appendix I. The main problem of the database was that for every sale, there was not only one value, but a declared amount and an accepted amount. Therefore, the author tried to examine any significant factors that would find any correlations. The analysis that follows in paragraph 4.2, unfortunately did not prove any relation or influential factors.

4.2 Numerical investigation of the deviation between Declared and Accepted price in Real Estate transactions in Cyprus

Real Estate professionals suggest that the most important component for any property market operation is transparency. The transaction data in Cyprus is not publicly available; however, it is provided by the Department of Land and Surveys (DLS) to registered valuers. This data records the Declared Price (DP) that the two involved parties (buyer and seller) declare at the DLS and the Accepted Price (AP) by the DLS. In this work, a variety of statistical and machine learning algorithms were implemented in order to explain numerically the inconsistency between the DP and AP. The New General Valuation Model of Cyprus (NGVMC) database was utilized and a significant difference was observed between DP and AP, in the range of zero (0.00) to eighty three percent (+0.829) with a mean of approximately four percent (+0.0384). The AP was always higher than DP, indicating that the policy of the DLS was to identify properties with DP less than the actual transaction price, regarding the intention of owners to pay less taxes. However, the procedure for the determination of the AP was not explicitly described by the DLS and no correlation between the difference in prices and particular attributes of the recorded properties was provided. Thus, in this paper, statistical inference (ANOVA, effect size) as well as several regression models (linear, nonlinear and machine learning) and relevant techniques for feature extraction (Relieff, p-values, root mean square error of ensemble models) were implemented in order to compute the statistical significance, that is to say, the probability of correlation of the properties' attributes (such as covered area, region, construction year etc.), as well as their importance (weight factors or model derivatives) to the difference between AP and DP. The two parties declare the transaction price on the Dept. of Lands and Surveys. DLS' officer can either accept the declared price or 'correct' it (accepted price). The officer does not perform an on-site visit of the property. The AP is based on the outcome of an oral interview with the two involved parties and previous transactions recorded. Both parties are motivated to under declare the real price. The buyer, in an attempt to reduce the transaction fees (the higher the price the higher the fee they pay), and the seller in order to decrease the amount payable for capital gains tax.

4.2.1 Distributions of the Deviation

In this section, the NGVMC database was utilized to quantify the effect of the Properties Characteristics (PCs) to the deviation between DP and AP. The aim of the research paper was to examine if this deviation is random or if it is caused from some of the parameters (PCs). This deviation of DP minus AP vary within the range of zero (0.00) to eighty three percent (+0.829) with a mean of approximately four percent (+0.0384). The normalization of the deviation (hereafter NDEV) in terms of price per square meters was used, as a more meaningful quantity for the properties, rather than the deviation (DP-AP). The PCs were divided into two categories, depending on their input values: continuous and categorical.

A) <u>Categorical Variables</u>

The selected categorical variables of the NGVMC database are Town-Village, Road Code, Planning Zone Category, Unit description, Unit class code, Unit condition code and Unit view code

- Town Village: Eight Municipalities within the Local Town of Nicosia were selected. The selection was not random.
- Road Code: The Department of Land and Surveys categorized the road that every property (flat) abut. Categories are 1 and 2. Those in category 1 are considered as 'better'

- Planning Zone Category: From 1 to 5. 1 is residential, 2 is Residential / Commercial, 3 is Residential continuous density, 4 is urban centre and 5 is commercial.
- Unit Description: Flats or Offices
- Unit Class Code: Are classified from 1 to 4, 1 is better
- Unit Condition Code: Are classified from 1 to 4, 1 is better
- Unit View code: Are classified from 1 to 3, 1 is better

In the following Figure 14, the distribution of NDEV for the four classes of **Unit condition code** are demonstrated. The classes 1 to 3 exhibited similar mean values (63.12€/m^2 to 63.64€/m^2) of a normal distribution shape, while the distribution of class 4 was of a bimodal shape, with a mean value more than 400% over the other classes (266.15€/m^2), however the properties within this category were only a few (five) and the results are not sound.

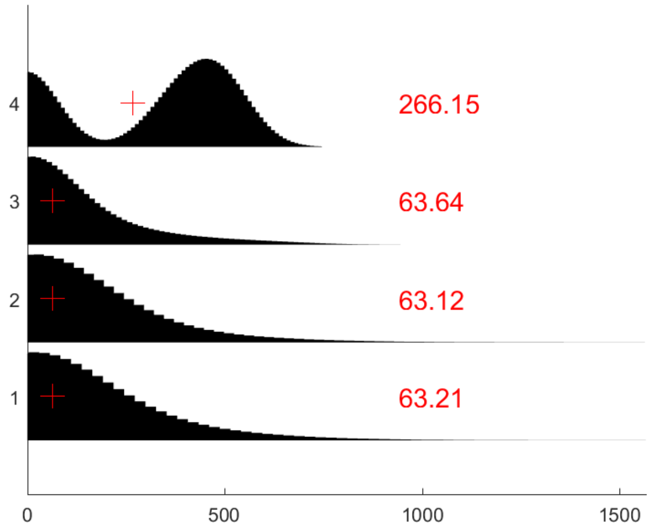


Figure 14: Distribution of NDEV for Unit condition code

Based on the above, it appeared that the condition of the property was not a factor that influenced the difference between declared and accepted prices. Accordingly, in Figure 15, the distributions of the NDEV were depicted within the eight municipalities (those that are regulated from the Local Town Plan of Nicosia) of Nicosia District studied (Strovolos, Nicosia, Latsia, Lakatamia, Egkomi, Geri, Aglantzia, Agios Dometios). For seven out of eight municipalities, the mean NDEV varied within the range of 42.66€/m^2

to 68.21€/m², while for municipality of Yeri, the mean NDEV was equal to 116.07€/m². The reason for this is either because the DLS had a wrong feeling and impression about the property values in Yeri or because the buyers there had an additional motivation to under declare the value of the property.

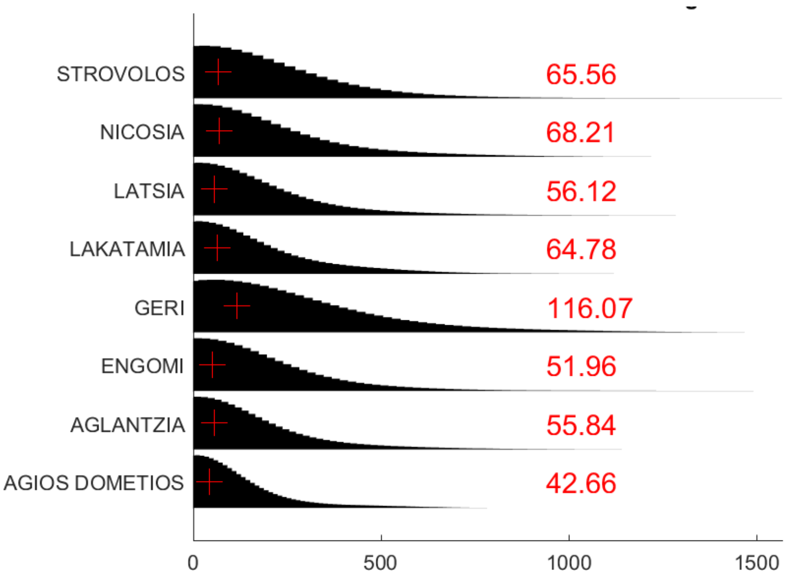


Figure 15: Distribution of NDEV per Municipality

The mean NDEV for Planning Zone Category 4 was found to be the higher (92.53€/m²), while for Zone 3 the mean NDEV was equal to 46.70€/m² (Figure 16). It seemed that the Planning Zone Category 4 had the biggest difference which is something normal. Accordingly, the planning zone did not seem to be an influential factor.

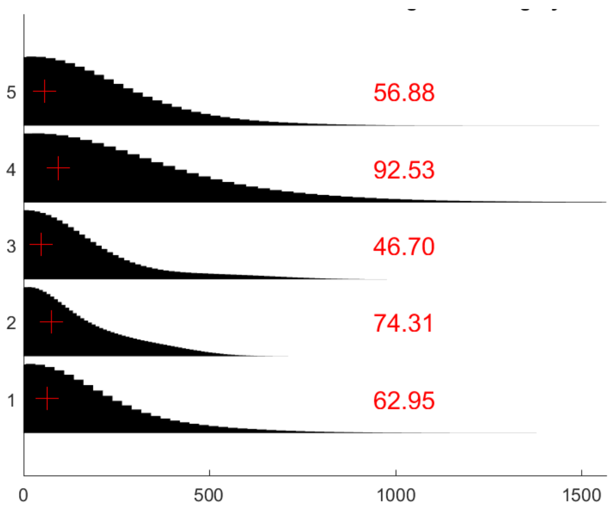


Figure 16: Distribution of NDEV for Planning Zone Category

For the for Unit class code, the highest mean of the NDEV appeared in class 1 (164.49€/m^2) as demonstrated in Figure 17. It appeared that luxury apartments (category 1) presented a significantly higher deviation and this ***<u>had to be further examined</u>***.

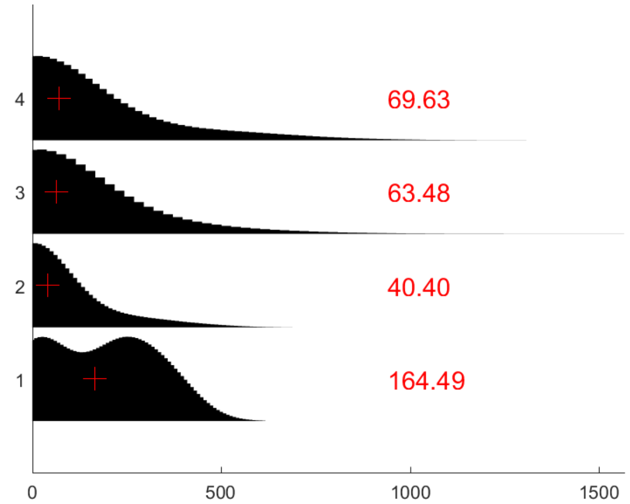


Figure 17: Distribution of NDEV for Unit class code

In Figure 18, a high difference of the mean NDEV for offices and apartments was exhibited. In particular, the mean NDEV for offices equated to 92.52€/m^2, while for apartments it was 62.47€/m^2. The aforementioned is something that the author expected for two reasons. Firstly, due to the number of transactions concerning offices being significantly lower and secondly as the offices' market had some additional variables that were not included in the NGVMC database (facilities such as raised floors, server rooms, additional fittings etc.)

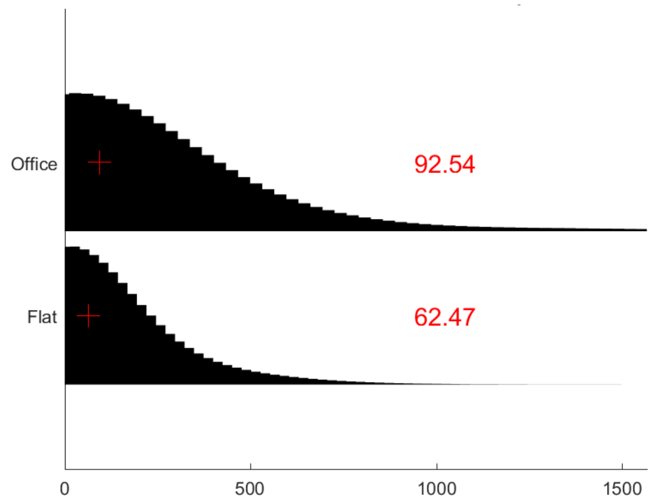


Figure 18: Distribution of NDEV for Unit Description

Finally, the mean NDEV did not exhibit a significant difference, varying for *Unit view code* within 51.30€/m^2 to 64.53€/m^2, and for *Road code* 61.41€/m^2 -65.69€/m^2. This was reasonable, as Nicosia did not present significant variations on this parameter. The author would expect higher NDEV for coastal cities.

B) <u>Continuous PCs</u>

Although the categorical variables exhibited significant differences regarding the NDEV for the particular aforementioned cases, the six continuous variables studied (Parcel Extent, DLO File Year, Unit built year, Unit enclosed extent, Unit covered extent, Unit uncovered extent) were found to be numerically uncorrelated with the NDEV as demonstrated in the following Figures 19 to 24. In the same figures, the distributions of each variable is demonstrated, while the distribution of the NDEV was highly skewed, with the majority of the values being equal to zero. This is taken into account later for the regression analysis.

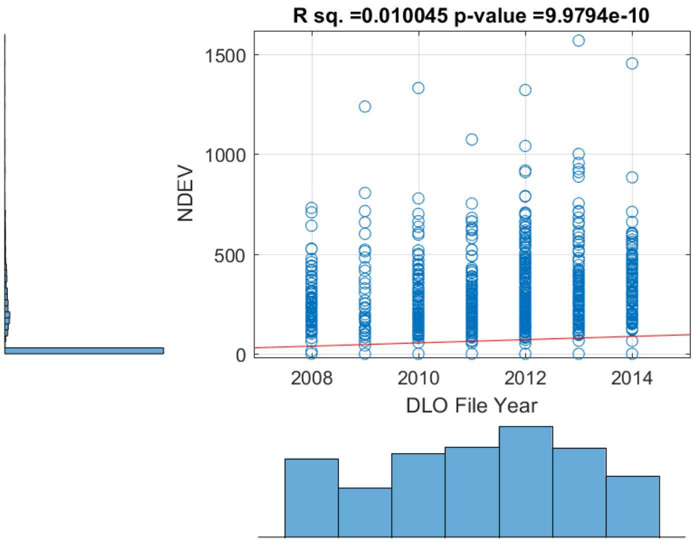


Figure 19: Distributions and correlation of NDEV with DLO File Year

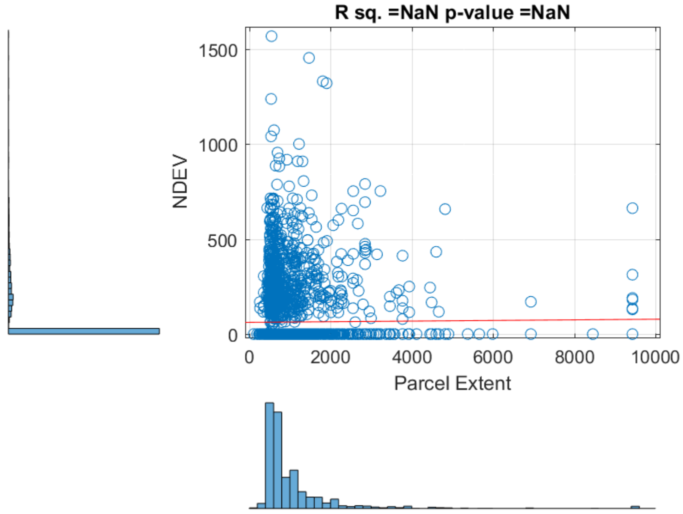


Figure 20: Distributions and correlation of NDEV with Parcel Extent

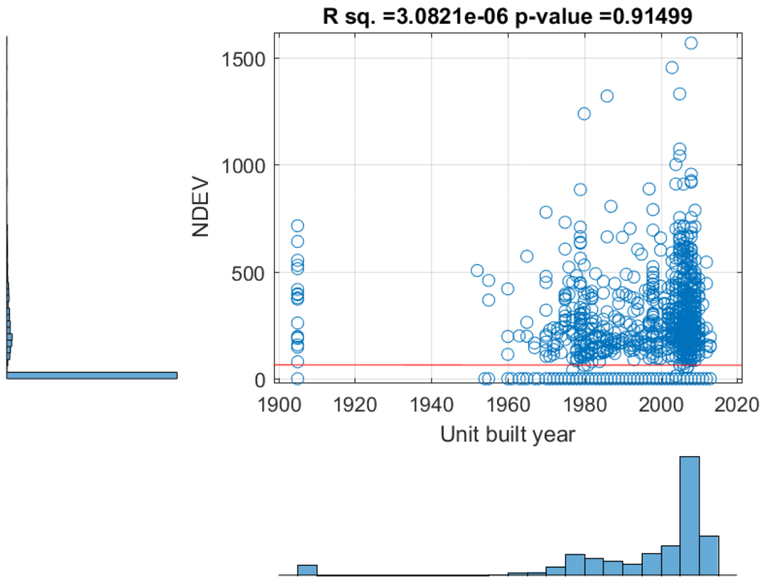


Figure 21: Distributions and correlation of NDEV with Unit Built Year

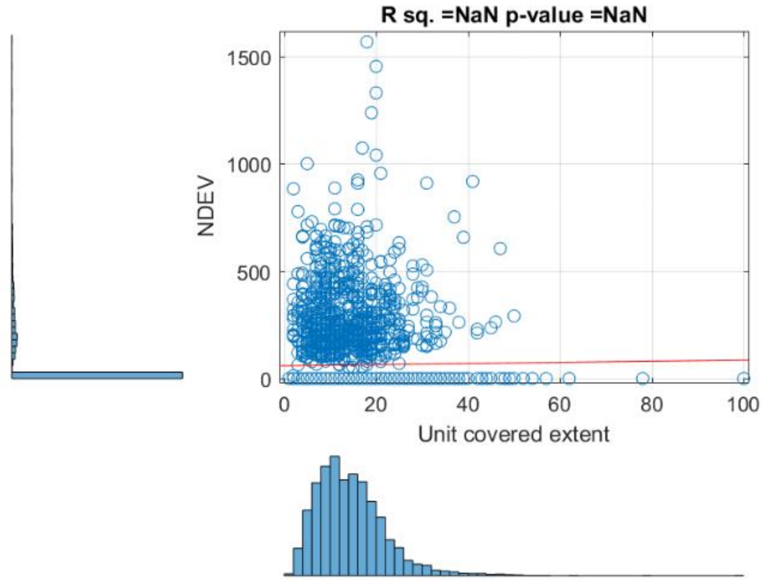


Figure 22: Distributions and correlation of NDEV with Unit covered extent

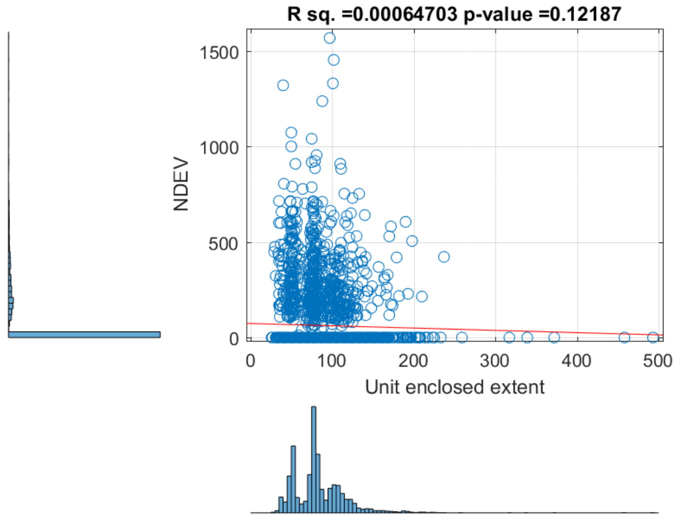


Figure 23: Distributions and correlation of NDEV with Unit enclosed extent

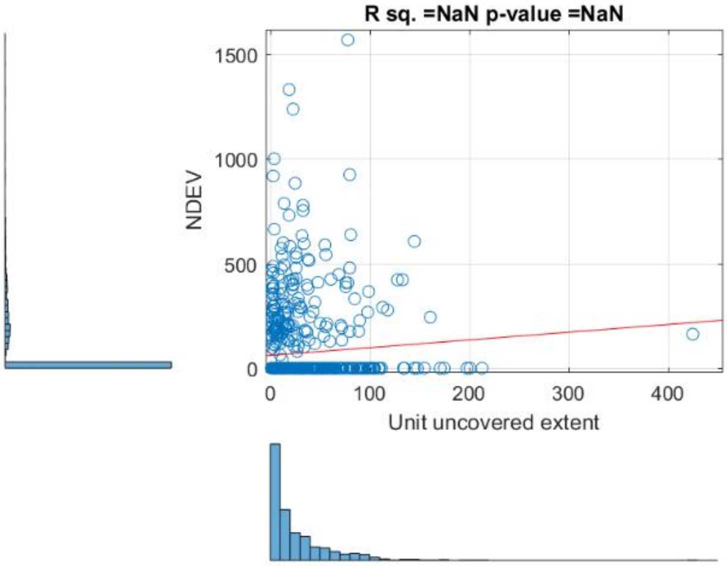


Figure 24: Distributions and correlation of NDEV with Unit uncovered extent

4.2.2 Regression Analysis

In order to investigate the further potential associations of the properties' characteristics with the deviation between Declared and Accepted price, linear and nonlinear regression analyses were implemented. In particular, the dependent variable assumed the raw Deviation (DEV), while the independent variables assumed the previously analyzed PCs. The regression was stepwise for both cases (linear and nonlinear), so as to keep only the statistically significant variables in the model. The selection criterion was the p-value for the continuous variables while for the categorical the chi-squared test, with a threshold of 5% for both cases. The nonlinear model, examined squared, cubic, quadric, logarithmic, exponential and sigmoid functions of the independent variables as predictors. The continuous variables (denoted by "_1" in the following Tables 6 to 9) were normalized to the domain between zero and unit [0, 1] so as to avoid numerical instabilities as well as to be able to compare the regression weight as importance measurements of the independent variables to the response. Additionally, a test set of 15% was utilized in order to estimate the behavior of the model for the out of sample data. The r-squared for the test was close or even greater than the train set for the linear regression, while the nonlinear exhibited a higher r-squared.

Table 6: Linear Regression for all the values of DEV

Fitted	Estimate	C.I. 95%(-)	C.I. 95%(+)	SE	tStat	pValue
(Intercept)	0.0024	-0.0060	0.0109	0.0043	0.5622	0.5740
Town-Village-AGIOS DOMETIOS_1	-0.0162	-0.0322	-2.8265e-04	0.0081	-1.9955	0.0461
Planning Zone Category-4_1	0.0369	0.0213	0.0525	0.0080	4.6393	3.6385e-06
DLO File Year	0.0218	0.0122	0.0314	0.0049	4.4609	8.4484e-06
Unit condition code-4_1	0.2139	0.1315	0.2964	0.0420	5.0882	3.8282e-07
Unit enclosed extent	0.0849	0.0377	0.1320	0.0240	3.5314	4.1935e-04
Unit covered extent	0.0483	0.0085	0.0880	0.0203	2.3827	0.0172
Unit uncovered extent	0.1388	0.0732	0.2045	0.0335	4.1483	3.4384e-05
Adjusted Rsquared =0.037349						
Test Set Rsquared =0.049462						

Table 7: Linear Regression for the non-zero values of DEV

Fitted	Estimate	C.I. 95%(-)	C.I. 95%(+)	SE	tStat	pValue
(Intercept)	0.0228	-0.0028	0.0485	0.0131	1.7478	0.0810
Town-Village-NICOSIA_1	0.0195	0.0014	0.0376	0.0092	2.1209	0.0343
Planning Zone Category-4_1	0.1348	0.0860	0.1836	0.0249	5.4245	8.1697e-08
DLO File Year	0.0532	0.0257	0.0807	0.0140	3.8020	1.5688e-04
Unit condition code-4_1	0.1770	0.0273	0.3266	0.0762	2.3218	0.0205
Unit enclosed extent	0.2917	0.2267	0.3566	0.0331	8.8228	9.9509e-18
Unit covered extent	0.0647	0.0083	0.1212	0.0287	2.2525	0.0246
Unit uncovered extent	0.2281	0.0888	0.3675	0.0710	3.2150	0.0014
Adjusted Rsquared =0.25288						
Test Set Rsquared =0.35992						

Table 8: Nonlinear Regression for all the values of DEV

Fitted	Estimate	C.I. 95%(-)	C.I. 95%(+)	SE	tStat	pValue
(Intercept)	-0.0055	-0.0273	0.0163	0.0111	-0.4973	0.6190
Planning Zone Category-4_1	0.0263	0.0107	0.0418	0.0079	3.3069	9.5392e-04
DLO File Year	0.0582	0.0469	0.0695	0.0058	10.0846	1.4727e-23
Unit class code-2_1	-0.0386	-0.0620	-0.0152	0.0119	-3.2374	0.0012
Unit class code-3_1	-0.0173	-0.0320	-0.0026	0.0075	-2.3036	0.0213
Unit condition code-4_1	0.1707	0.0874	0.2539	0.0424	4.0208	5.9380e-05
Unit uncovered extent	-0.7058	-1.3583	-0.0534	0.3328	-2.1211	0.0340
log(Unit enclosed extent)	0.1053	0.0638	0.1469	0.0212	4.9720	6.9827e-07
log(Unit uncovered extent)	0.6326	0.1133	1.1519	0.2649	2.3885	0.0170
Unit built year^3	1.4407	1.0813	1.8002	0.1833	7.8587	5.2923e-15
Unit built year^4	-1.3142	-1.6327	-0.9958	0.1624	-8.0921	8.3019e-16
sigmoid(Unit built year)	-0.1658	-0.2194	-0.1123	0.0273	-6.0764	1.3776e-09
sigmoid(Unit enclosed extent)	-0.1505	-0.2379	-0.0630	0.0446	-3.3742	7.4934e-04
Stepwise Regression for:Deviation						
Adjusted Rsquared =0.066866						
Test Set Rsquared =0.048401						

Table 9: Nonlinear Regression for the non-zero values of DEV

Fitted	Estimate	C.I. 95%(-)	C.I. 95%(+)	SE	tStat	pValue
(Intercept)	0.1087	0.0568	0.1605	0.0264	4.1165	4.3384e-05
Town-Village-NICOSIA_1	0.0421	0.0216	0.0625	0.0104	4.0430	5.9033e-05
Town-Village-STROVOLOS_1	0.0301	0.0099	0.0503	0.0103	2.9284	0.0035
Planning Zone Category-4_1	0.0982	0.0533	0.1431	0.0229	4.2939	2.0213e-05
Unit built year	0.1259	0.0298	0.2219	0.0489	2.5732	0.0103
Unit covered extent	0.0971	0.0407	0.1536	0.0287	3.3810	7.6509e-04
Unit uncovered extent	-1.7817	-2.7957	-0.7677	0.5164	-3.4503	5.9592e-04
DLO File Year^2	0.0497	0.0244	0.0750	0.0129	3.8586	1.2532e-04
Unit enclosed extent^2	0.4240	0.3286	0.5194	0.0486	8.7290	2.1180e-17
log(Unit uncovered extent)	1.6814	0.8196	2.5432	0.4389	3.8312	1.3976e-04
sigmoid(Unit built year)	-0.1773	-0.2743	-0.0803	0.0494	-3.5882	3.5782e-04
Stepwise Regression for:Deviation						
Adjusted Rsquared =0.3009						
Test Set Rsquared =0.28107						

4.2.3 Machine Learning Investigation

In order to confirm numerically the hypothesis that the NGVSC did not follow a specific method for the unacceptance of the declared price, a machine learning algorithm named Relieff was utilized. In particular the aim of this study, was to detect conditional dependencies between the PCs and DEV, while the Relieff was selected as a feature subset selection method, since it split the data to k-nearest neighborhoods, by means of their normalized n-dimensional Euclidean distance. This procedure made it possible for an unsupervised classification of the properties with similar characteristics (PCs). Accordingly, the importance of each particular variable to the response could be computed, depending on if a change of this variable caused a change of the class or not. The dataset was split into 20,40,60,80 and 100 classes, in order to examine if the results depended on the volume. Finally, the Relieff method was applied for all and for the non-zero values of the deviation as well. The calculated weights were lower than 0.025 if all the values (Figure 25) and lower than 0.04 (Figure 26) if only the non-zero values of the DEV were examined. As a value of zero indicates inexistence of association and values close to unit, high importance of the predictors, the calculated values denoted essentially no association.

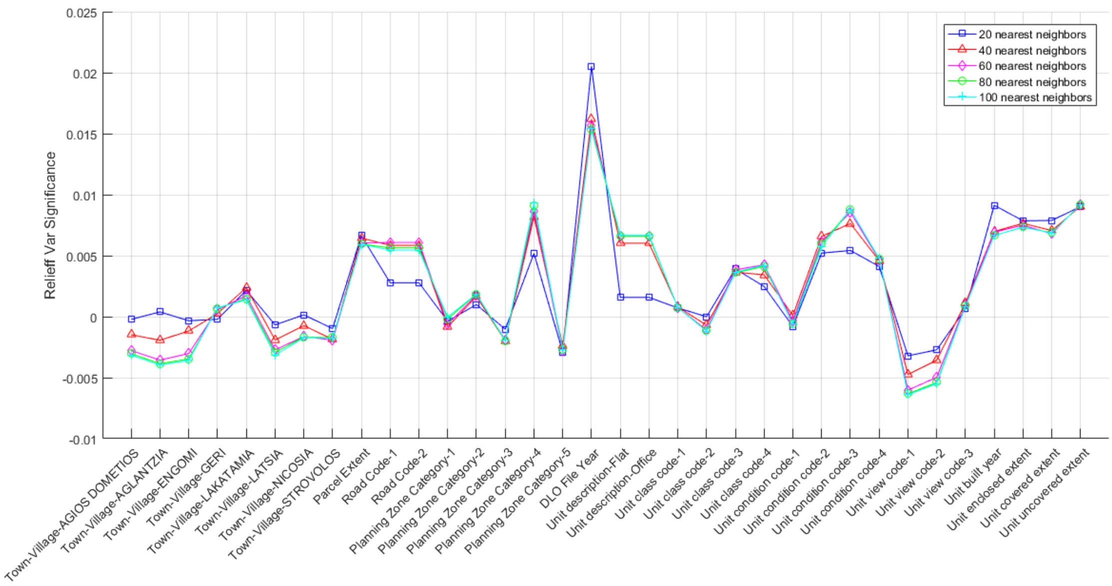


Figure 25: Relieff weights for all the values of DEV

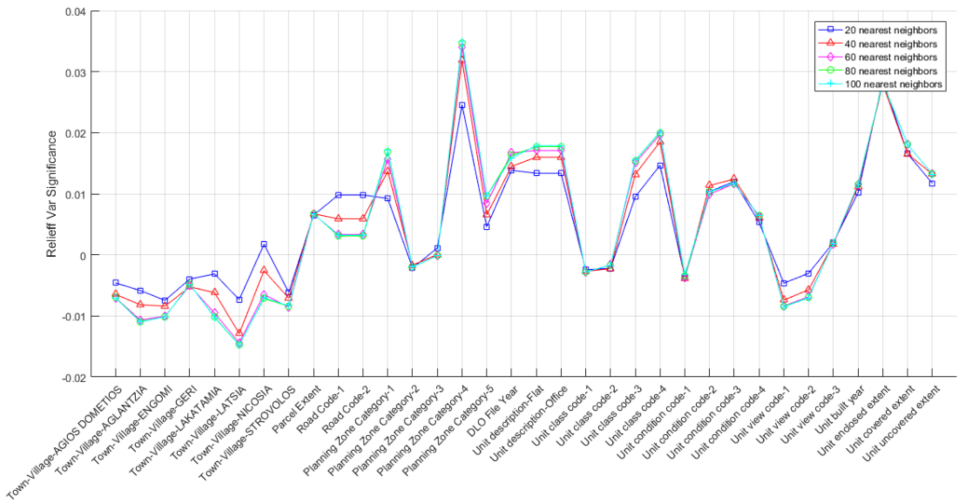


Figure 26: Relieff weights for the non-zero values of DEV.

4.2.4 Discussion about the data appropriateness and limitations

McCluskey et al., (1997) states that having access to information in itself is not enough. There are problems involved in unlocking significant potential. Given the imperfection and inherent inefficiency of the property market (applicable in most countries), greater access to more structured information is a desirable "in principle" objective. Consider the role currently played by data in the work of valuers. The main problems fall within the following areas:

1. Utilization. Valuers and assessors who will be involved should be able to handle and process digital databases. Basic knowledge of GIS is a prerequisite.
2. Availability. The data should be available and appropriate legal framework is required. The latest GDPR law in Europe appears as an obstacle. Data collection is always very expensive.
3. Quality. Even if enough data is available, they must present consistency and be realistic, and the transaction declared or accepted prices the closest possible to the Market Value. Other qualitative and quantitative data are also expected to be realistic.
4. Technology. A cloud – based GIS software is a basic requirement. The use of drones or oblique photography is also of great assistance on the data verification and also in the appeal process.
5. Targets. The goal of the Mass Appraisal shall be set a-priori. The aim determines the model the cost and the data requirements.
6. Legality. The legal barriers that exist shall be overcome before such a system is implemented.
7. Impacts. An open access system would impact the valuation profession significantly and great consideration shall be taken into account.

4.2.5 The lack of standardized and harmonized data

Quality depends not only on good quality data; it also is related to different perceptions and interpretations of information and data. In this respect the particular skills of the competent valuer can utilize data in the most effective manner.

Research demonstrates that recording information in different ways, without a recognized industry standard for collection and recording of data, militates against quality and perhaps more critically against data sharing. Arguably standardization, particularly the need for a recognized unique property identifier, is a prerequisite for a systematic data set built on information sharing. Users need compatible formats of data for transfer and downloading of data from one system to another. Common forms of deficiency in datasets are:

1. Limitations – structural deficiencies in a dataset which become clear when it is used for purposes other than intended.

2. Uncertainties – introduced when variables are measured against a non-objective standard, for example when an area is classified as belonging to a particular neighbourhood grouping which itself may be poorly defined.

5 Case studies: MRA, GWR, Random Forests, AI and machine learning

5.1 Multiple Regression and Geographically Weighted Regression

The relationship between real estate prices and its explanatory variables is highly complex. Residential property values are traditionally simulated and explained based on the pioneering work of Rosen (1974), following the hedonic model framework. This regression technique has been widely used to estimate property values in the real estate market, as well as to investigate interactions between real estate prices and their related attributes (Páez, 2009; Páez et al., 2011), including the impacts of amenities or disseminates on the property sale price (Paterson & Boyle, 2002). Within this model, properties are considered as the entities that provide utilities to their owners, and housing prices are calculated based on the amount of utility-created amenities that are present within the associated properties (Liu, 2013).

In the advancement of MRA, Goodman & Thibodeau illustrated through linear modelling that price is a consequence of a choice involving the tiered interaction among property characteristics, fundamentals of neighbourhood and housing markets (Goodman, 1998). For many years, the MRA method was the standard method used and is still applied for real estate mass appraisals. However, this method originated through non-spatial disciplines, and thus, it does not account for peculiarities of spatial data (Jahanshiri et al, 2011). Several studies have integrated spatial effects in such models to consider location in the hedonic framework (Anselin, 1998). This approach incorporates spatial independence and structural instability, as well as spatial drift and spatial lags that eliminate the error term (Gao et al., 2006). As a result, spatial studies utilized the spatial nature of property datasets to examine spatial causation in the regression models (Thériault et al., 2003).

Noteworthy is the rapid increase of interest in absorbing the temporal dimension in the hedonic price modeling, either in global models (Worth, 2004) or in local models (Huang et al., 2010). This is because the residential valuation process evolves not only over space but also over time. Academic studies have explored regression models using distance coefficients and specific dummy variables in a geographical manner. Although this approach has improved Ordinary Least Squares (OLS)-based regression techniques, some biased coefficients still exist (Bidanset & Lombard, 2014a).

Despite the effort for accurately modelling locational effects in the OLS context, housing price residuals customarily exhibit spatial dependence which creates problems in the construction of house price index models (Mae, 1997) and mass appraisals (Bourassa et al., 2003). Basu & Thibodeau, (1998) argue that nearby properties often have similar structural attributes and share locational features as a result, spatial dependence is created. For that purpose, a plethora of studies have integrated the hedonic framework of pricing models with GIS (Geographic Information Systems) technologies to develop environmental amenity variables that might affect transaction prices (Geoghegan et al., 1997). GIS-derived locational characteristics have been increasingly used in many mass appraisal studies, deriving that property values are provided by the value of the attributes collected (Lake et al., 2000). Based on GIS, such studies have explored more advanced statistical techniques, known as Geographically Weighted Regression (GWR), indicating significant improvements on the model performance, since spatial weighting function is performed and coefficients are fluctuating within a geographic space (McMillen, 2004).

In recent years, the GWR method has presumed greater prominence for price estimates, since it combines and isolates spatial heterogeneity and dependency, considering locational consequences and market segmentation (Páez, 2005). Bitter, Mulligan & Dall'erba analysed the prediction accuracy of the GWR approach through examining spatial heterogeneity within a residential price dataset (Bitter et al., 2007). The result of their study depicts GWR as a superior method for predictive accuracy and explanatory power. Furthermore, the local variation of GWR allows the analysis of the locational relationships between properties, while it also addresses a function that models spatial dependency and spatial heterogeneity (James et al., 2005). However, Dimopoulos (2016) recommends that prior to a GWR model, it is important to identify the significance of attributes in pricing models based on OLS regression techniques.

GWR is a non-stationary method that has been applied to capture spatial varying relationships by calibrating local models, in which a local neighbourhood is determined using kernel functions with fixed or varying bandwidth (Fotheringham & Crespo, 2015). This technique provides a point wise calibration in a local form of regression, concerning the influence of nearer observations around each regression entity in order to determine a local set of coefficients estimated through weighted least squares (Fotheringham et al., 1998). It is a linear regression technique that creates an equation for each feature in the dataset in order to define any potential spatial variation regarding the relationships amongst dependent and explanatory variables (Esri, n.d.). Lu, Chartlon, Harris & Fotheringham provided the general equation of the GWR model as (Lu et al., 2014):

$$\boldsymbol{yi = \beta i0 + \sum_{k=1}^{m} \beta ikxik + \varepsilon i} \qquad \textbf{[5.1]}$$

where yi represents the dependent variable at location i; xik depicts the explanatory variable at location i; the number of the explanatory variables is provided with m; βi0 is the intercept parameter at location i; βik provides the local regression coefficient for explanatory variable kth at location i; and εi represents the standard error at location i.

5.1.1 Spatial determinants and Geographic Information System

A plethora of real estate studies has long identified the significance of location on real estate values. In general, real estate spatial dimension is a principal factor that has contributed a new field of study (Rodriguez et al., 1995). Prior the development of GIS, the impact of location in prediction models was unlikely to be measured. GIS was developed to facilitate an analytic spatial reasoning for the scientific geographers, performing a series of spatial processes. GIS consists of spatial functionalities that allow for data to be analysed across geographic space. Dimopoulos & Moulas provided GIS as a substantial platform that could maximize efficiency and accuracy within the real estate sector, since it has the capabilities to provide large databases and efficient spatial analysis tools, as well as to process and visualise geographical data in high quality (Dimopoulos & Moulas, 2015). Particularly, GIS plays a vital role in spatial diffusion models or spatial interaction models which might be used to determine an optimal site location or create forecasts about population movements, absorption rates or neighbourhoods' growth. Within GIS, property data can be easily disaggregated towards spatial market segments, while market area could be examined through spatially joining available datasets. GIS technology can produce a set of variables that might be applied to analyse the real estate sector, providing technologies for spatial support and geostatistical analysis.

Computerized mass appraisal systems have been developed by government authorities in order to generate an automated property valuation process for real estate taxation purposes (Dimopoulos & Moulas, 2015). For that reason, it is a government responsibility to provide fairness and equality in terms of property taxes and thus, property valuation systems must ensure accuracy, reliability and uniformity in regards to property values. McCluskey, Deddis, Mannis, McBurney & Borst argue that integrating GIS with such valuation systems is a major development in the recent years in property valuation sector, as it enables spatial data to be adopted in tax valuation models (McCluskey et al., 1997). GIS has the capabilities to visualise data values on a map so as to provide direct communication of data with the user, while it provides a more transparent and accountable system in indicating spatial data distribution. Integrating GIS with valuation systems enables a mechanism that manages, queries and models spatial property information.

Tian & Yang have explored the combination of the hedonic framework with GIS in order to spatially analyse the effects of several factors on real estate values and thus, to create a powerful computerised mass appraisal system in China (Tian & Yang, 2015). Using a sample of 270 residential sale transactions of three districts, a series of multi-criteria decision was developed relating to the proximity of the properties to some desirable localities. A fitting analysis of four different hedonic models was carried out, ending up with the most optimal model of being the logarithmic form of hedonic regression. In this study, researchers indicated that the proximity of such selected properties with regards to business services, the district and the age of each of the properties have a negative correlation with real estate appraisal valuation, whilst the proximity towards fundamental facilities like schools and bus stops is positively correlated with the appraisal valuation, and thus strongly influences real estate values. Mulaku & Kamau follow a similar statistical procedure as Tian & Yang to derive the significance of location on property values regarding residential land in Nairobi, Kenya, using geostatistical analysis and Multiple Regression Analysis (MRA) (Mulaku & Kamau, 2010). In this study, a sample of 97 properties were obtained from three middle-class residential neighbourhoods in Nairobi in order to analyze the variation of residential land values concerning the proximity to major neighbourhood amenities (such as schools, hospitals and churches), as well as the distance from the major road junctions and the proximity to the Central Business District (CBD). As Tiag & Yang demonstrated, the researchers of this study also indicated that public amenities and the CBD have a strong positive impact on property prices in urban areas (Tian & Yang, 2015). Furthermore, the author also concluded with illustrating the significance of GIS applications in analyzing locational characteristics and spatial distributions to generate desirable automation for mass appraisal valuations.

Kuburic, Tomic & Ivic noted that mass valuation models must be consistent, functional and adaptable to real estate conditions and market trends of each nation. At the same time, the spatial effects on property values and a sufficient number of spatial attributes are crucial tasks in modeling real estate mass appraisal systems (Kuburić et al., 2012). In their paper, they use multi-criteria valuations of spatial attributes to create an efficient mass appraisal system in Serbia. Apart from the social and economic characteristics provided in their data sample, they further analyze the contribution of the spatial characteristics of real estate prices. Specifically, they determined a series of spatial multi-criteria valuations, including the distance from the capital, various traffic corridors, such as highways, airports, rivers, main roads, and railway lines, as well as amenity spots and distance from natural resources, like energy potential, mining resources, and waterways. Spatial analysis was also carried out by Koramaz & Dokmeci based on GIS technologies to identify the influence of spatial characteristics on residential values. At the same time, they integrated regression and interpolation techniques regarding the spatial prediction of housing prices (Koramaz & Dokmeci, 2012). In addition to the structural and neighbourhood characteristics provided within the model, the author also included distance from the CBD, coasts and transportation arteries as the locational attributes of such predictive model. The results of their analysis revealed that spatial determinants in terms of distance variables significantly affect housing prices, while the CBD was illustrated as the major attribute, which is positively proportional to real estate values in Istanbul.

These previous studies shed light on the validity of spatial characteristics as important determinants for real estate price predictions. It seems that some variables are more significant than others. For example, the location might well comfort the ideal sense of prediction models in more accurate results than other factors do. However, spatial data must be adaptable to the current market condition of each nation (Kuburić et al., 2012), to provide functional and consistent results. Prediction models are treated as accurate if not only structural attributes are considered in the statistical process for real estate mass appraisals. Based on that statement, this study aimed to analyze, through implementing GIS techniques, the significance of various spatial attributes on residential property prices in Nicosia, providing a more accurate technique than the one already used by the Department of Land and Surveys in Cyprus. Real-estate mass appraisals in Cyprus are generated based on MRA that processes only structural characteristics, completely ignoring the spatial term that affects real estate values.

5.1.2 Applying GWR for residential apartments in Nicosia.

The main part of this chapter has been already published in the Proceedings Volume 10444, Fifth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2017).

Dimopoulos & Yiorkas, (2017) implemented the use of GIS and GWR techniques to improve the results of the New General Valuation (1.1.2013). On a sample of comparative evidence for flats in Nicosia District, GIS was used to measure the impact of spatial attributes on real estate prices and to construct a prediction model in terms of spatially estimating apartment values. In addition to the structural property characteristics, some spatial attributes (landmarks) were also analysed to assess their contribution on the prices of the apartments, including the Central Business District (CBD), schools and universities, as well as the major city roads and the restricted zone that divides the country into two parts; the occupied by Turkish area and the Greek area. The values of the spatial attributes, or locational characteristics, were determined by employing GIS, considering an established model of multi-criteria analysis. The price prediction model was analyzed using the OLS method and calibrated based on the GWR method. The results of the statistic process indicated an accuracy of 81.34%, showing better performance than the mass valuation system applied by the Department of Land and Surveys in Cyprus with an accuracy of 66.76%. This approach suggested that GIS systems are fundamentally important in mass valuation procedures in order to identify the spatial pattern of the attributes, provided that the database comprises of a sufficient amount of comparable information, and is continuously updated.

This study proved that the property mass appraisal system applied by the Department of Land and Surveys in Cyprus to impose property taxes was not so accurate since the spatial attribute that describes the location of each property was not taken into consideration.

The most frequent question arises as to what factors affect property values. According to Dimopoulos & Moulas, all properties are unique, and their values are affected by endless qualitative and quantitative factors (Dimopoulos & Moulas, 2015). Real estate prices depend on various social and economic attributes that reflect the physical needs of the property owners in terms of income or services provided, such as location, size, quality, regulations etc. In terms of residential price modeling, Basu & Thibodeau argue that all these factors that affect property values could be configured or classified into three domains (Basu & Thibodeau, 1998). The first describes the structural characteristics that households are willing to pay for the purchase of a housing unit in terms of their shelter preferences (i.e. dwelling area, age, and number of bedrooms). The second domain includes the socio-economic attributes describing the characteristics of the neighborhood (i.e. population, unemployment rate). In contrast, the third domain reflects the locational characteristics regarding accessibility and centrality (i.e. proximity to amenities, services and pleasant landscapes) (Kockelman, 1997).

5.1.2.1 Data Analysis & Methodology

The empirical research conducted in this study focused on Nicosia, the capital of Cyprus, which comprises of ten Municipalities. However, eight of these Municipalities were selected, as the area of interest in this study which is presented in Figure 33 below, while the other two were rejected due to the lack of information and the limited number of sales in the real estate sector in the years of the study. The reference unit selected in this study was residential apartments, while the data obtained represented the social and physical environment of this property type. Primarily, the data were obtained from secondary sources, while some data were calculated or determined through primary sources. Specifically, a sample dataset of 1,341 apartment transactions of the study area from 2008Q1 to 2014Q3 were retrieved from the Department of Land and Surveys in Cyprus, as well as the Map Layer that presents the parcels geospatially within the area of interest. The sample dataset provided some structural characteristics for each transacted property (apartment) including the declared and accepted transacted prices, the date of transaction, the size (providing the enclosed area, the covered area and the uncovered area individually), the age, the administrative area and the physical condition with values varying from 1 to 3 (the value of 1 depicts apartments in bad condition, 2 is for average condition and 3 for excellent condition). Some data that referred to spatial features/landmarks were obtained from the Geospatial Information Portal of Cyprus in a GML format. Since GML format is incompatible with ArcGIS 10.4.1, the data were firstly visualized in QGIS 2.14.2 and were exported as new features in a Shapefile format. These spatial features described the boundaries of the study area. They also included some spatial landmarks such as public roads, historical monuments and museums, private and public universities and schools, health units, emergency stations, units of entertainment, the UN buffer zone, Nicosia's city centre, and the Central Business District (CBD). All spatial features were transformed into the local projected coordinate system CGRS 1993_LTM.

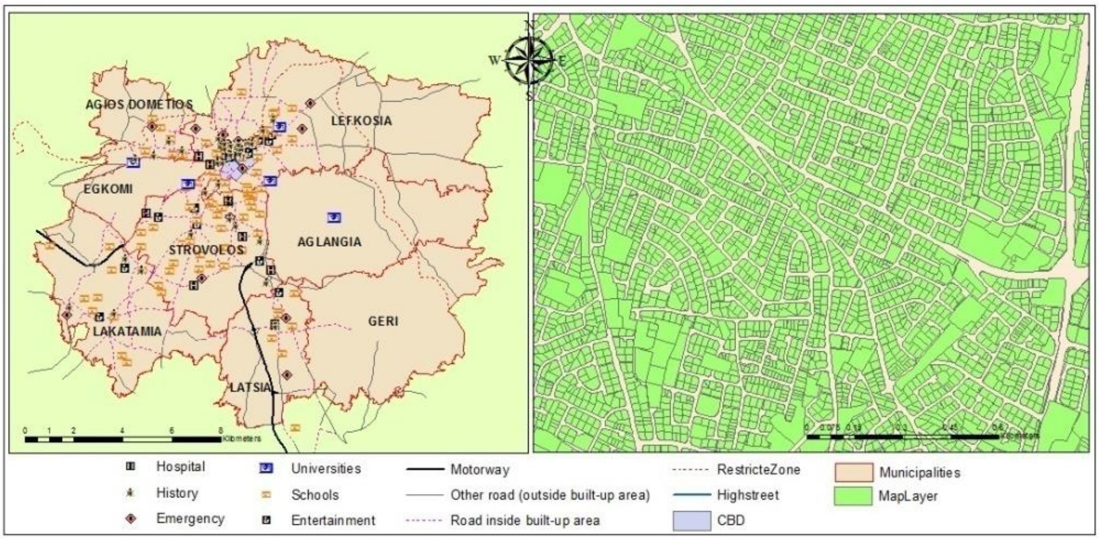


Figure 27: Map Layer of Nicosia District and spatial features (landmarks) located within the study area. Source: Geospatial Information Portal of Cyprus and the Department of Land and Surveys in Cyprus.

5.1.2.2 Data analysis and adjustments

Performing any statistical method (using the hedonic price modelling), required the adjustment of the declared and accepted property prices accordingly in order to provide a single price for each entity (apartment) in the dataset. More specifically, the market value of each property was determined by using the weighted average, adopting a weight of 35% for the declared prices and 65% for the accepted prices. However, the value of each apartment provided in the database represented the transacted sale price that corresponded to the share of the property sold and thus, the market value of each property was adjusted to represent the whole property share. At this stage it was also essential to consider that the calendar year 2008 was highlighted as the peak year for the Cyprus economy with property prices being at their highest value. However, the effects of the financial crisis that struck the economy of the island were firstly observed early in 2010, with the real estate industry suffering a significant loss, and property prices dropped. Even though the sale prices were correctly presented in each quarter, they were provided at different market periods. Therefore, to eliminate the significant discrepancy of their time variation, property prices were adjusted and transformed in order to be presented within the same quarter of the same calendar year, and thus within the same market period (Kuburić et al., 2012). Similarly to the paper of Fotheringham, Crespo & Yao, the apartment prices were adjusted to the Q3 2014 equivalent price using two different price indices as tools in order to allow the results to be comparable over time (Fotheringham & Crespo, 2015). The transformation of the values was established in relation to the RICS Cyprus index and the index produced from Central Bank of Cyprus (CBC). The trends of apartment prices in Nicosia that were determined by each property index are illustrated in Figure 2. For the purpose of this study, it is believed that RICS Cyprus index was more accurate and realistic than the CBC price index and thus, a weight of 65% was assigned for RICS index and 35% for CBC index in order to assess the quarterly percentage change with the benchmark being the Q3 2014.

QUARTER	RICS	CBC	Adjusted
2008Q1 - 2009Q4			-35.0%
2010Q1 - 2014Q3	-35.5%	-19.4%	-29.9%
2010Q2 - 2014Q3	-35.9%	-18.4%	-29.8%
2010Q3 - 2014Q3	-35.5%	-18.7%	-29.6%
2010Q4 - 2014Q3	-34.4%	-18.9%	-29.0%
2011Q1 - 2014Q3	-34.4%	-19.0%	-29.0%
2011Q2 - 2014Q3	-33.6%	-18.9%	-28.5%
2011Q3 - 2014Q3	-29.9%	-18.7%	-26.0%
2011Q4 - 2014Q3	-29.3%	-17.7%	-25.2%
2012Q1 - 2014Q3	-28.9%	-16.4%	-24.5%
2012Q2 - 2014Q3	-28.3%	-15.4%	-23.8%
2012Q3 - 2014Q3	-27.4%	-15.5%	-23.2%
2012Q4 - 2014Q3	-25.5%	-13.8%	-21.4%
2013Q1 - 2014Q3	-18.9%	-12.5%	-16.6%
2013Q2 - 2014Q3	-14.5%	-10.5%	-13.1%
2013Q3 - 2014Q3	-10.7%	-8.4%	-9.9%
2013Q4 - 2014Q3	-4.1%	-6.2%	-4.8%
2014Q1 - 2014Q3	-2.8%	-3.5%	-3.0%
2014Q2 - 2014Q3	-1.3%	-2.2%	-1.6%
2014Q3 - 2014Q3	0%	0%	0.0%

Figure 28: RICS and CBC price indices for apartment values during the period 2010-2014 on a quarterly basis.

5.1.2.3 Visualization of the transactions within the map of the study area

In the paper of Rosiers & Thériault, the buildings under study were geographically visualized through the geocoding process, using street addresses provided in the related attribute of each entity (Des Rosiers & Thériault, 2006). Lu, Charlton, Harris, &

Fotheringham visualised the residential properties in London by geocoding data using the postcode of each house (Lu et al., 2014). Each transaction in the dataset had a parcel ID number that corresponds to a particular geographic location in the city (Morrow-Jones et al., 2005). Thus, due to the absence of addresses or postcodes in the dataset of this study, the sample comprised of the property transactions was imported in ArcGIS and spatially joined with the Map Layer (parcels in a shapefile format) based on that ID number. In that way, all property attributes were integrated within a spatial dataset. With spatial reference the centroid of each parcel, all related properties were visualized as point features on the map.

5.1.2.4 Spatial data selection and analysis

Based on useful tools provided by ArcGIS, a series of criteria were analyzed and integrated into the assessment process of the spatial dataset. Previous studies depict that Euclidean Distances to various services/amenities and proximity to CBD are useful and objective measures that can be easily assessed to identify their significance in the property market. For that purpose and similarly to Tian & Yang, (2015), spatial analysis functions were applied to quantize several feature variables (landmarks), especially buffer analysis, while OLS and GWR models were calibrated with Euclidean Distances. The criteria selected in this paper in order to describe or evaluate real estate aimed to represent the social and physical environment of each property. The factors/chosen criteria for being included in the dataset are presented below:

- Euclidean Distance from each property to the high street of Nicosia (Archbishop Makariou III Avenue).
- Proximity of each property to the nearest historical monument or museum in a straight line.
- Number of properties that fall within a range of 0-500m from CBD.
- Number of properties that fall within a range of 500m from the motorway axis.
- Number of properties within a range of 150m from major road axis (roads within build-up area).
- Number of properties that fall within a buffer zone of 500m from the restricted line.
- Proximity of each property based on Euclidean Distance to the nearest school.
- Proximity of each property based on Euclidean Distance to the nearest Emergency station.
- Number of properties that fall within a range of 300m from hospitals or clinics.
- Number of properties that fall within a buffer zone of 1km from University of Cyprus and 500m from the other universities in the city.

- Proximity of each property to the nearest theatre, cinema or library using a straight-line distance.

Following the regression model developed by Fotherinhgam, Crespo & Yao (2015), the explanatory variables used in this study to predict residential prices included structural, socio-economic, and locational attributes. Absorbing a hedonic price modelling method, the adjusted transaction price of each property was considered as the dependent variable. At the same time, an OLS-based linear regression technique was processed to assess the significance of individual explanatory variables, presented in Table 11, on property market values in Nicosia. The prediction model of the real estate prices was locally calibrated based on the GWR method.

Table 10: Structural, Socio-economic, and spatial variables included within the prediction model to assess their significance on apartment prices of Nicosia District.

	Variable Name	Type	Definition
Dependent Variable:	Adj_Prty_P	Numerical	Price in Euros of each property
Independent Variable:			
• **Structural attributes**	Prty_Age	Numerical	Age of each property
	Prty_Area	Numerical	Adjusted area of each property considering enclosed area + covered area/3 + uncovered area/5
	Prty_Cond	Numerical	Property Condition giving a value of 1 for bad condition, 2 for average and 3 for excellent.
	Admin_Area	Numerical	Administration area within which each property is located. The municipal districts of Nicosia, Strovolos, Lakatamia, Aglatzia, Latsia, Engomi, Agios Dometios and Geri assigned by 1,2,3,4,5,6,7,8 respectively.
• **Socio-economic attributes**	Population	Numerical	Population of each municipal district within which is property is located is assigned.
• **Locational (spatial) attributes**	BCentre500	Dummy	Value of 1 for properties within a range of 0-500m from the Central Business Centre, otherwise value of 0.
	Mtrway500	Dummy	Value of 1 if properties fall within a buffer zone of 500m from motorway, otherwise value of 0.
	MjRoad100	Dummy	Value of 1 if properties fall within a range of 100m from Major road, otherwise value of 0.
	RZone500	Dummy	Value of 1 if properties fall within the Restricted Zone of 500m, otherwise value of 0.
	Prox_Scls	Numerical	Proximity to the nearest school in metres.
	Prox_Emerg	Numerical	Proximity to the nearest emergency station in metres.
	Hsptl300	Dummy	Value of 1 if properties fall within a range of 300m from hospitals or clinics, otherwise value of 0.
	Uni500_1km	Dummy	Value of 1 if properties fall within a range of 1km from university of Nicosia and 500m from other Universities. Otherwise value of 0.
	Prox_HSt	Numerical	Euclidean Distance in meters from each property to the High street in Nicosia.
	Hstry100	Dummy	Value of 1 if properties fall within a range of 100m from historical monuments and museums, otherwise value of 0.
	Prox_Entrt	Numerical	Proximity to the nearest theatre, cinema or library in meters (entertainment).

5.1.2.5 Analysis of explanatory variables

Calibrating the dataset using the global linear regression technique (OLS), it was noticed that the Adjusted R^2 value was fairly low, indicating that the explanatory variables were explaining or interpreting the model by 55.37%. The values provided as Multiple R^2 and Adjusted R^2 measured the performance of the model and vary within a range of 0 (for 0%) to 1.0 (for 100%). As per the Esri's model, the value of the Adjusted R^2 reflects the complexity of the model and consequently explains more accurately the performance of the model than the Multiple R^2. To ensure credibility and accuracy of the results, bias estimates (outliers) were identified and removed from the model. Hoen et al., (2014) assumed that observations above and below 1% and 99% of percentile are problematic outliers, while Bidanset & Lombard, (2014b) established the outliers based on an IQRx3 approach. However, in the paper of O'Connor (2013), outliers were defined and excluded from the dataset based on the residuals determined with standard deviation falling outside ±3. Similarly, to this approach, outliers of this study were established and omitted from the model, removing the entities with standardized residuals falling outside ±1.5. In total, 237 individual objects were detected and excluded from the dataset as influential points since they could have had a disproportionate impact on the prediction model and regression coefficients. As a result, the sales transactions comprised within the dataset were reduced from 1,341 properties to 1,104 properties. Based on the sales transactions remaining within the dataset, 54 properties were randomly collected and excluded from the database called 'Active model', in order to form a new database, called 'Passive model'. Within the 'Passive model' database, the corresponding market value of each entity was estimated through processing, based on linear regression, the 'Active model'. The result was compared with respect to the actual transaction prices that correspond to each property in order to verify the price prediction performance of the 'Active model'.

The dataset was processed, outliers were excluded from the sample, and the final database (Active model) of 1,050 transactions was assembled and utilized for the analysis. Table 2 illustrates the significant improvement of the regression coefficients (and the model performance in general) from Adjusted R^2 = 55.37% to Adjusted R^2 = 78.98%. Furthermore, the descriptive statistics of each variable in the sample dataset is also provided in Table 2, illustrating the variability within the model. The average property price of the sample was approximately €100,000, while the mean of property age was 15.27 years old, and the mean of the adjusted property size was 88.04sq.m. Based on this statistical analysis, it could be observed that 16% of the sampled properties fell within the range of 0-500m from CBD, 2% within the restricted zone (UN buffer zone), 7% within the predefined zone from the universities of the city and 1% within the range of 100m from historical monuments and museums. The average proximity to schools and emergency stations was 571m and 1,367m, respectively. The standard deviation of each variable demonstrated the data dispersion, in which the higher the value states greater the spread in the dataset. On the other hand, the range of each attribute indicated the difference between the maximum and the minimum values of specific attributes, representing the interval that includes all values in the dataset (Görög, 1994).

Table 11: Diagnosis of the 'Active' model based on OLS and the descriptive statistics of each variable in the dataset.

Input Features:	Model
Number of Observations:	1050
Multiple R-Squared [d]:	0.792996
Joint F-Statistic [e]:	247.327775
Joint Wald Statistic [e]:	3560.983415
Koenker (BP) Statistic [f]:	32.119668
Jarque-Bera Statistic [g]:	26.268313
Dependent Variable:	ADJ_PRTY_P
Akaike's Information Criterion (AICc) [d]:	23332.397078
Adjusted R-Squared [d]:	0.789790
Prob(>F), (16,1033) degrees of freedom:	0.000000*
Prob(>chi-squared), (16) degrees of freedom:	0.000000*
Prob(>chi-squared), (16) degrees of freedom:	0.009647*
Prob(>chi-squared), (2) degrees of freedom:	0.000002*

Descriptive Statistics

Variables	Mean	Median	Standard Deviation	Range	Minimum	Maximum	Valid Cases
Adj_Prty_P	100092.49	97863.75	34970.26	228148.31	21619.50	249767.81	1050
Prty_Age	15.27	10.00	12.53	53.00	1.00	54.00	1050
Prty_Area	88.04	84.33	29.49	210.40	29.00	239.40	1050
Admn_Area	2.59	2.00	1.77	7.00	1.00	8.00	1050
BCentre500	0.16	0.00	0.37	1.00	0.00	1.00	1050
RZone500	0.02	0.00	0.15	1.00	0.00	1.00	1050
Prox_Scls	571.76	384.79	594.01	3791.31	40.27	3831.58	1050
Prox_Emerg	1367.65	1181.73	809.30	4133.43	62.32	4195.75	1050
Hsptl300	0.05	0.00	0.21	1.00	0.00	1.00	1050
Prty_Cond	2.37	2.00	0.60	2.00	1.00	3.00	1050
Mtrway500	0.02	0.00	0.12	1.00	0.00	1.00	1050
MjRoad100	0.34	0.00	0.47	1.00	0.00	1.00	1050
Uni500_1km	0.07	0.00	0.26	1.00	0.00	1.00	1050
Prox_HSt	2960.48	2338.13	2179.30	9409.22	25.98	9435.20	1050
Hstry100	0.01	0.00	0.08	1.00	0.00	1.00	1050
Prox_Entrt	1305.20	1179.15	742.88	4445.24	67.20	4512.44	1050
Population	48529.25	55014.00	19630.37	59669.00	8235.00	67904.00	1050

A hedonic price model was created in order to investigate the relationships between apartment values and related determinants. The corresponding coefficient of each explanatory variable in the model reflected the type and strength of their relationship to the dependent variable (property price). If the sign of the coefficient was positive, the relationship between the dependent variable and the associated independent variable was also positive and vice versa. All the coefficients were provided in the same units of the corresponding explanatory variables, specifying the expected change of the dependent variable (property price) in each unit change of the relevant explanatory variable, while holding the other attributes constant. Statistically significant variables were provided with an asterisk in the probability (or Robust Probability) column, indicating that it was important to consider statistically significant coefficients in the regression model. Variables with no statistically significant coefficients do not help a model and must be excluded unless theory describes them as critical.

Eliminating all the variables from the model with no statistically significant coefficients as determined through the OLS technique, it was observed that property age, restricted zone, proximity to schools and major roads had a negative impact on property prices, whereas property area, property condition and proximity to universities are positively influenced apartment values in Nicosia. Even if the CBD was not highlighted with statistically significant coefficient, there was a lot of background history that indicated the importance of that 'landmark' on real estate values. Esri also stated that the results provided according to a global regression model might be unreliable if explanatory variables exhibit multicollinearity or redundancy. Redundancy between variables is measured through the values of the Variance Inflation Factor (VIF) in which, according to the rule of thumb, variables associated with high VIF values (greater than 7.5) should be excluded from the model. Although all explanatory variables used in the 'Active model' provided satisfactory VIF values, most of them were excluded from the model due to the lack of significance either statistically or theoretically indicated. Table 12 portrays that correlations between the explanatory variables used within the 'Active model' were more than satisfactory, since the variation was very low. However, property condition and age exhibited some level of collinearity, as Pearson Correlation factor was determined at approximately 60%. On the other hand, the correlation among the other variables was calculated below the factor of 40%, with the majority varying within a range of 0% to 20%. In other words, this meant that multicollinearity or heterogeneity did not exist.

Table 12: Correlations between the explanatory variables of the prediction model.

	Prty_Age	*Prty_Area*	*BCentre500*	*RZone500*	*Prox_Scls*	*Prty_Cond*	*MjRoad100*	*Uni500_1km*
Prty_Age	1							
Prty_Area	0.254989	1						
BCentre50(	0.351649	0.097979	1					
RZone500	0.042476	0.005425	0.03561796	1				
Prox_Scls	-0.21428	-0.18449	-0.200546539	-0.073708708	1			
Prty_Cond	-0.60019	-0.21441	-0.301671311	-0.052533724	0.152793529	1		
MjRoad10(	0.152145	0.062944	0.07532625	0.119700445	-0.25052328	-0.110963539	1	
Uni500_1k	-0.03442	-0.09678	-0.104767804	0.054759091	0.224028058	-0.009805625	-0.077486524	1

Table 13: Summary statistics of the model variables with the attributes derived as statistically significant.

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	8941.330168	3594.648411	2.487400	0.013012*	3660.109699	2.442913	0.014721*	--------
PRTY_AGE	-1243.582120	51.939437	-23.942926	0.000000*	54.102499	-22.985669	0.000000*	1.714711
PRTY_AREA	1009.625272	17.689363	57.075276	0.000000*	18.696355	54.001182	0.000000*	1.102204
BCENTRE500	280.080154	1462.424716	0.191518	0.848152	1501.201405	0.186571	0.852027	1.187550
RZONE500	-8684.216710	3362.943418	-2.582326	0.009941*	3513.976354	-2.471336	0.013608*	1.023952
PROX_SCLS	-1.654064	0.913800	-1.810095	0.070573	0.841443	-1.965746	0.049587*	1.193278
PRTY_COND	9665.693667	1055.126892	9.160693	0.000000*	1067.122737	9.057715	0.000000*	1.603821
MJROAD100	-2592.715923	1097.055306	-2.363341	0.018279*	1079.285561	-2.402252	0.016455*	1.091823
UNI500_1KM	4215.307536	1973.880291	2.135544	0.032936*	1890.901769	2.229258	0.025994*	1.073315

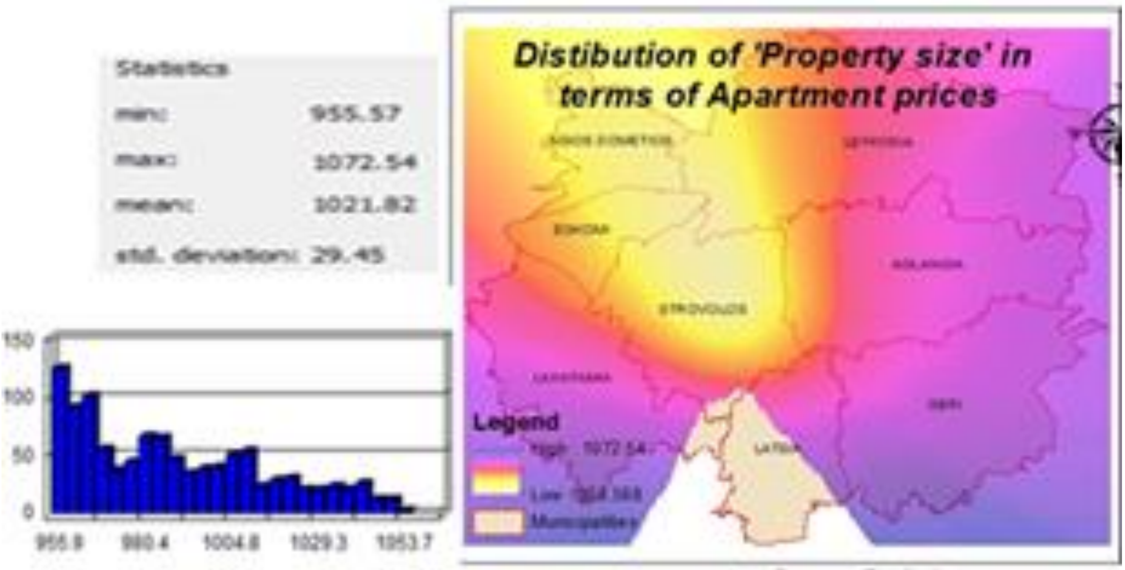
Table 13 presents the OLS results of the new 'Active' model that only included the explanatory variables associated with significant coefficients. The result of the regression model verified that the observations about Universities positively affecting real estate prices. Furthermore, it was also indicated that the restricted zone (UN buffer zone) parameter was inversely proportional to apartment values whereas, property condition had a greater influence on prices than age and size. Surprisingly, the schools had a negative impact on property prices with a narrow coefficient value of 1.65. This could be simplified to say that, apartments in remote areas were negatively affected by the absence of schools in a more significant proportion than apartments located at a very close distance where the presence of schools might have adversely affected apartment values due to the noise or crowding created. For this reason, the coefficient determined was very narrow. Using the notation specified in Table 13, the hedonic pricing model for Nicosia apartments could be expressed as depicted in the Equation:

$$\mathbf{Apart.Value} = \mathbf{8,941} - \mathbf{1,244} \times [\boldsymbol{Property\ Age}] + \mathbf{1,010} \times [\boldsymbol{Property\ Area}] + \mathbf{280} \times [\boldsymbol{fall\ within\ CBD}] - \mathbf{8,684} \times [\boldsymbol{fall\ within\ Restricted\ Zone}] - \mathbf{2} \times [\boldsymbol{distance\ to\ the\ nearest\ School}] + \mathbf{9,666} \times [\boldsymbol{Property\ Condition}] - \mathbf{2,593} \times [\boldsymbol{fall\ within\ Major\ Road\ Zone}] + \mathbf{4,215} \times [\boldsymbol{fall\ within\ University\ zone}] \quad \textbf{[5.2]}$$

5.1.2.6 Model calibration based on GWR

The hedonic pricing model presented in Equation 2 was calibrated using a local regression technique, known as GWR, which expands the global framework by enabling parameters to differ over space (Fotheringham & Crespo, 2015). GWR provides a pointwise calibration of hedonic pricing models in a linear regression form, which creates an equation for each entity in the dataset based on nearby entities, and thus, it determines the influence of all spatial features on property values. Perhaps, the most significant parameter for GWR is the number of neighbours or bandwidth used for local estimations, which controls the degree of the model smoothness. Typically, a neighbour value or bandwidth is applied by selecting either the cross-validation (CV) method or the Akaike Information Criterion (AICc) method as the bandwidth parameter. Both ways try to determine an optimal adaptive number or fixed distance of neighbours, with the only difference being on the criteria used for the expression of the 'optimal' (Bidanset & Lombard, 2014b). The units of the bandwidth depending on the kernel type specified for the model, in which an adaptive type allows distance to change towards the spatial density of the features used. On the other hand, a fixed Gaussian Kernel bandwidth indicates a range that has the same unit with the associated function in the input feature class. Stone stated that both bandwidth methods (Cross-Validation and AICc) are asymptotically equivalent (Criterion, 1977).

In this case study, a GWR calibration of the price model in Equation 2 was carried out based on an adaptive Gaussian kernel bandwidth with an AICc method. A series of continuous surfaces in a raster form were also generated with an output cell size of 60 pixels in order to define the spatial variations in apartment values over time. Values of the parameters were accommodated in each pixel in order to determine the equations of the regression, namely the regression coefficients and the intercept (Patriche & Vasiliniuc, 2009). In each pixel, a distinctive regression model was formulated. The quality of the models applied varied remarkably from one place to another, declaring that there were places where property values dramatically depended on age and size, while in other areas, these spatial patterns and relations faded considerably. In that way, spatial differentiation could be identified, enabling effective decision-making (Dimopoulos, 2015). For instance, as illustrated in Figure 29: Spatial distribution of property size and age parameters determined based on GWR., property values along the southern part of the study area were greatly affected by the property size attribute, whereas property age had a greater extent on apartment prices along the northern part of the study area. As it was expected, age was negatively correlated to real estate values due to the building depreciation and obsolescence that might occur. Thus, coefficient variations were negative in all places within the study area. Regarding the calculated spatial distribution of ‘age’, it was indicated there was lower demand for apartments in the suburbs than in more centralised areas when the age parameter was constant. On the other hand, moving away from the center, the need for larger spaces increased, as the minimum coefficient value was illustrated for apartments within the centralised places and the maximum values along with the decentralized areas. Accurately, the highest ‘Area’ coefficient value was recorded in Lakatamia municipal district (1,070) and the lowest in Lefkosia and Agios Dometios (950). On the other hand, the lowest and highest values for ‘age’ were recorded in Nicosia (-1,725) and Lakatamia (-1,090) respectively.



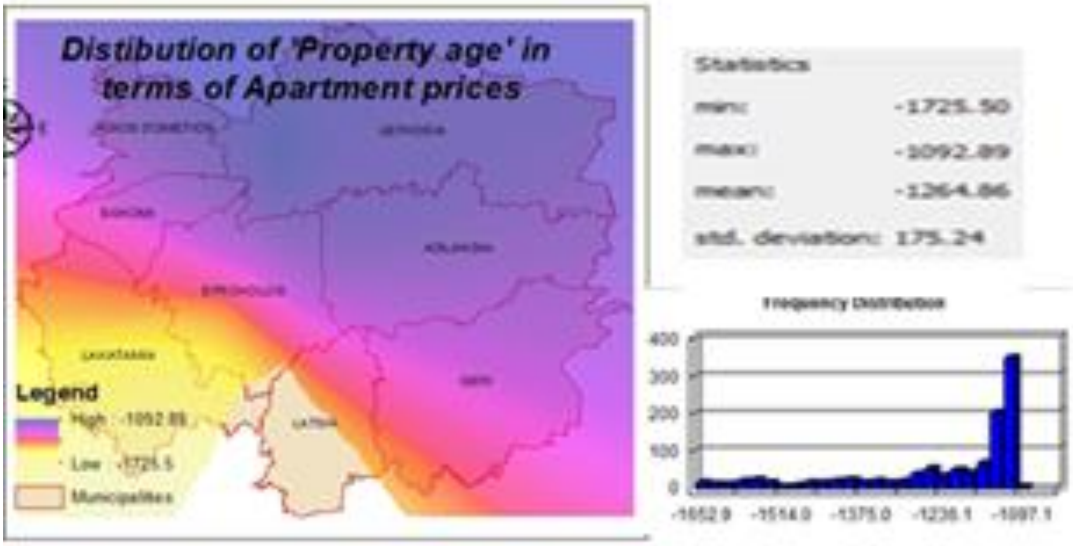


Figure 29: Spatial distribution of property size and age parameters determined based on GWR.

5.1.2.7 *Model performance and spatial autocorrelation*

As a classical linear regression model, R^2 was calculated to describe the model performance, interpreting how well it fit the data. Similarly, to OLS, the adjusted R^2 is a more representative indicator to portray the model, since it normalizes the fraction concerning degrees of freedom. Under the circumstances of this study, the statistical results of the model based on GWR illustrated that data was interpreted at very satisfactory percentage level of 79.38%. Examining the pattern of the model residuals that are spatially visualized in Figure 30, it was concluded that the model was performing well since the presence of red and dark blue areas was limited. Red circles indicate the properties under prediction values (where actual prices are higher than predicted values), whereas blue circles indicate the overpredicted properties (where actual prices are lower than predicted values).

For correctly specified regression models, residuals must be randomly distributed instead of clustering since statistically, significant clustering indicates that the model might be misspecified. For that reason, a spatial autocorrelation based on Moran's Index was processed to ensure that model regression residuals were spatially random. The statistic is the standard procedure to utilize spatial autocorrelation. Moran's Index calculates the similarity of the attributes between the value of each variable at one location and value of the same variable at another location. A positive value exists when nearby places are similar in attributes, whereas a negative value is obtained when nearby areas are dissimilar (Yomralioglu & Nisanci, 2004). In this paper, the Moran's Index estimate of nearly 0.00 (z-value: -0.76, p-value: 0.45) statistically confirms that no spatial autocorrelation exists. Thus, the null hypothesis is accepted to complete spatial randomness (Figure 30). Esri suggests that the null hypothesis always indicates that features are randomly distributed within the study area.

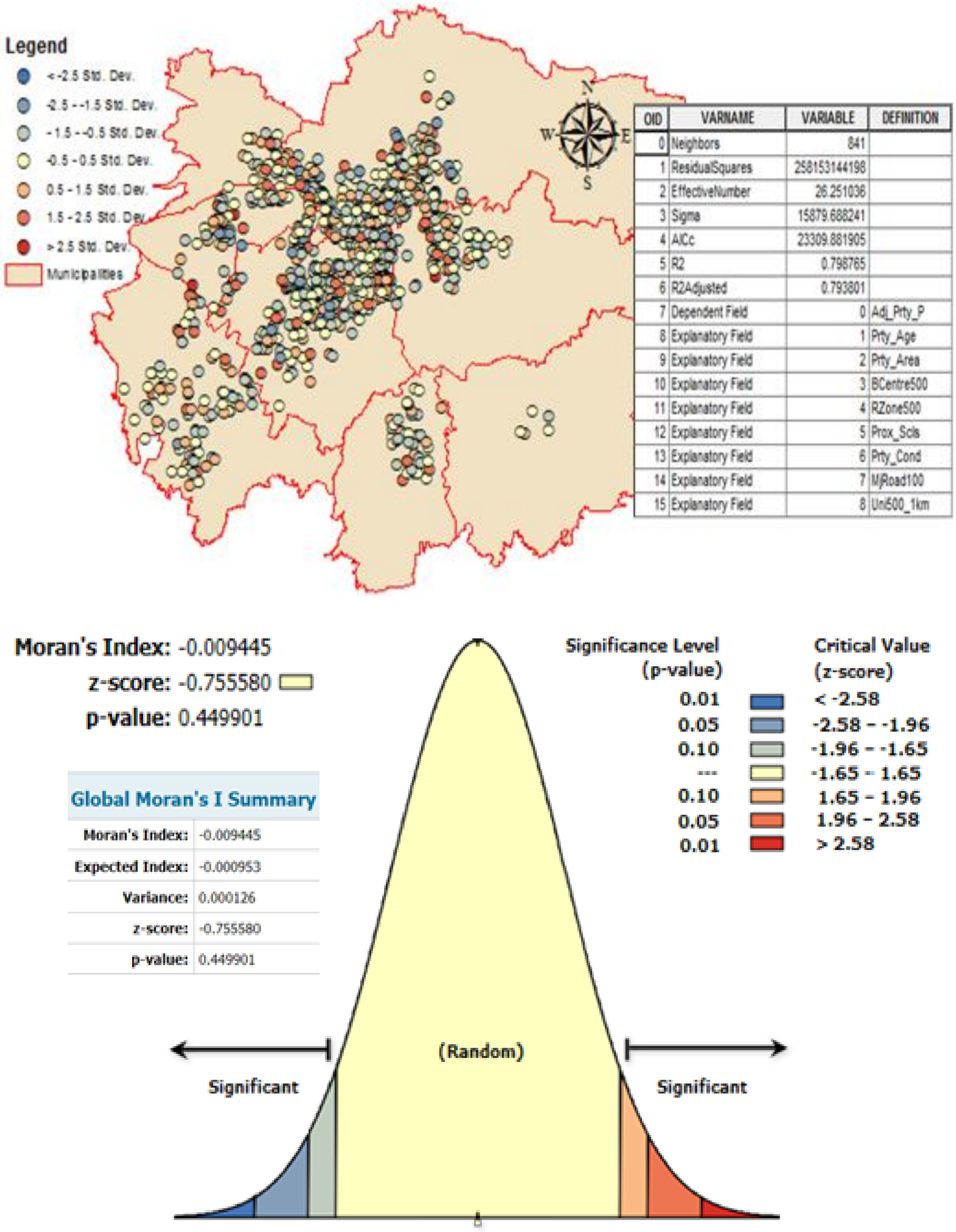


Figure 30: GWR results and spatial autocorrelation results of the prediction model.

5.1.2.8 *Spatial distribution of coefficients intercept*

Figure 31 illustrates the intercept coefficients of the model on property prices, clearly indicating that the lowest values are concentrated across the centralised areas, while moving outward from the Nicosia centre, the coefficients of variable interception tended to increase. Adopting a graduated colour pattern of the model coefficients, it was specified that the lowest values of interception were concentrated in the north-west part of the study area, specifically in Agios Dometios. In contrast, the highest values were indicated across the south-western area (in Lakatamia). Noteworthy, Figure 31 demonstrates the significant negative influence of ‘age’ and ‘restricted zone’ on apartment prices, since some properties in Agios Dometios depicts negative values of coefficient intercept. Agios Dometios is relatively small in size compared to the Municipality of Nicosia, which is mainly consists of old buildings and is located at a very close distance from the UN Buffer zone. On the other hand, Lakatamia is a suburb of Nicosia district, in which the majority of buildings are brand new properties, and thus, values of coefficient intercept are high. This spatial distribution could also be indicated by analyzing the Municipal district of Strovolos, which is divided into the old area and the newly built area, where the intercept values increase as moving towards the modern part of Strovolos.

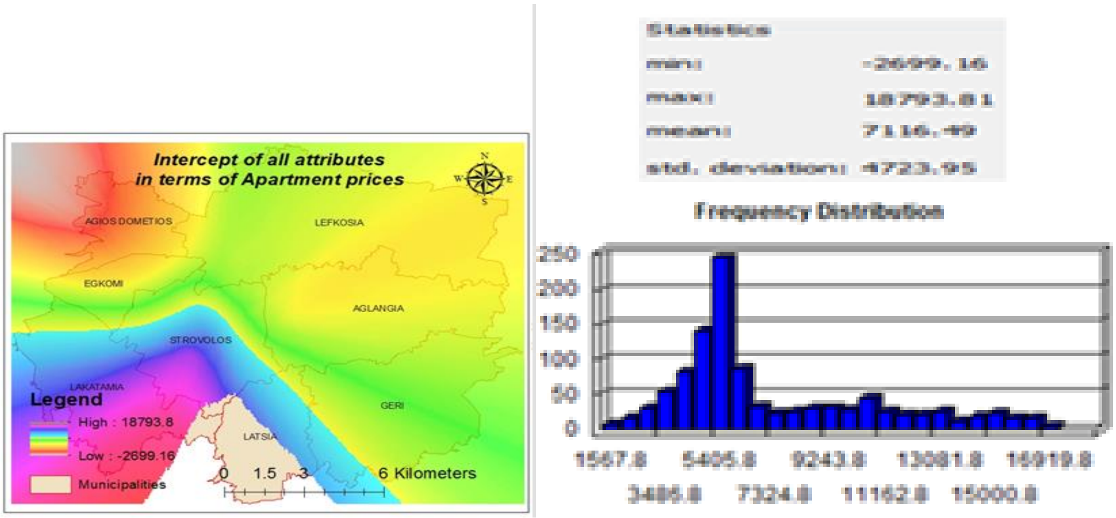


Figure 31: Spatial distribution of the intercept coefficients calculated by GWR.

5.1.2.9 *Model predictions and verifications*

Having created, through GWR, a well-specified prediction model, the prices of the 54 apartments in the ‘Passive model’ were automatically estimated. The accuracy of the prediction results was performed and assessed by calculating the Pearson Correlation factor between the actual values of the apartments and the values predicted by the model. Additionally, the correlation between the exact prices and the associated taxable prices (valuations obtained by the Department of Land and Surveys in Cyprus for taxation purposes) was also assessed and compared. The correlation between the actual prices and the predicted prices determined by the GWR model was calculated at 81.34%. This indicated that the GWR model provided much more accurate property estimates since the correlation factor between the actual apartment prices and the values determined by the Cypriot Authorities was much lower with an accuracy rate of 66.76%. The results also verified the hypothesis of the term study, indicating that the spatial attribute was fundamentally significant for ensuring the accuracy of price-prediction models in regards to real estate. The better accuracy of the model is demonstrated in Figure 32. At this point, it is crucial to notice that prices of the 'Active model' used to undertake predictions of the apartment values within the 'Passive model', were adjusted to represent the same calendar year, and thus the same market period. Price adjustments were achieved based on two reliable price index tools (RICS index and CBC index) that provided a general view of price fluctuations. Thus, the apartment prices used in the model as actual values are based on approximations. Therefore, the GWR model has the possibility to be further improved in accuracy by providing property transactions achieved within the same market period.

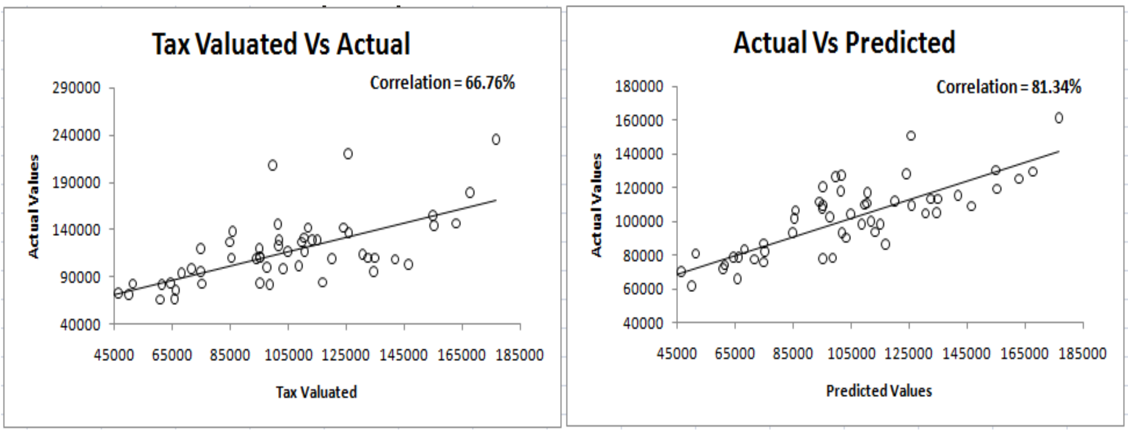


Figure 32: Accuracy of General Valuation (Left Figure) and accuracy of prediction model of this study (Right Figure)

5.1.2.10 Conclusions

Mccluskey et al., (2013) argues that, from an industry perspective, accuracy is fundamentally essential in predicting model approaches, without rejecting transparency and defensibility of the model requirements. However, (Jahanshiri et al., 2011) provided that speed, consistency, uniformity, and accuracy of mass real estate appraisals are the key factors that tax authorities are challenged with. This paper aimed to examine the significance of spatial variables in apartment prices in Nicosia, Cyprus, on the basis of constructing an accurate price prediction model. According to Yomralioglu and Nisanci, the basis of any property prediction model is significantly formatted based on geospatial data, by the fact that property prices are significantly determined by their spatial characteristics (Yomralioglu & Nisanci, 2004).

Gathering a sample dataset of 1,341 apartment sales in Nicosia, each property was spatially coordinated in ArcGIS 10.4.1 by joining each entity with the associated parcel based on a common ID field. The transaction prices comprised within the dataset were adjusted based on two remarkable price indices in order to represent the same calendar year. Removing any outlier from the dataset, the sample was divided into two models; the 'Active model' used for the creation of the prediction model and the 'Passive model' used for verification. This study analyzed whether a spatial statistical prediction model constructed in ArcGIS performed better than the traditional system used by the Department of Land and Surveys in Cyprus that ignores the spatial influence. This was achieved by considering a set of spatial attributes within the model that were applied based on the multi-criteria analysis. In addition to the structural characteristics included in the sample used, five spatial variables were also determined, while the others were rejected due to the lack of their statistical significance. The coefficients of each variable were analyzed and assessed using an OLS-based regression technique, while the prediction model was calibrated based on the GWR method. The statistical results of the GWR model illustrated very satisfactory levels, indicating that it fit the data at 79.38%. The correlation between the actual prices and the prices estimated from the model was calculated at 81.34%, showing more accurate results than the appraisal model already performed by the local Tax Authority with an accuracy of 66.76%.

To sum up, the integration of mass appraisal models with GIS processes could ensure a complete visualization and intercommunication of the valuation process, while enabling a mechanism that allows managing, querying, and modeling spatial information. Creating an Automated Valuation Model (AVM) based on GIS applications, government appraisers, mortgage insurers, real estate agencies, and financial institutions could provide better-informed insurance, lending, and property allocation decisions. Appraisers can also take advantage of their knowledge of the property market to perform market analysis through statistical and GIS systems. This study could be further improved by considering a sample composing of more property entities with updated transaction prices, so as to eliminate any possibility of biases that might occur through approximations. Additional to that, there are many more fundamental spatial features (parks, green areas, sidewalks etc.) that could be analysed to define their contribution to the property market. Further studies could also be conducted based on comparing different prediction approaches to determine the one that describes the model more accurate. Finally, it might be interesting to apply the model of this study to a coastal city where sea view substantially contributes to property values.

5.1.3 Applying GWR for residential apartments in Thessaloniki, Greece.

The main part of this chapter has been already published in the Open Geosciences Journal (T Dimopoulos & Moulas, 2016).

The Municipality of Thessaloniki, located in North Greece, was also selected as a case study in this thesis. The municipality comprises of six Municipal Districts (MD) which are shown in the following figure (a base map of the wider area of Thessaloniki was used as background):

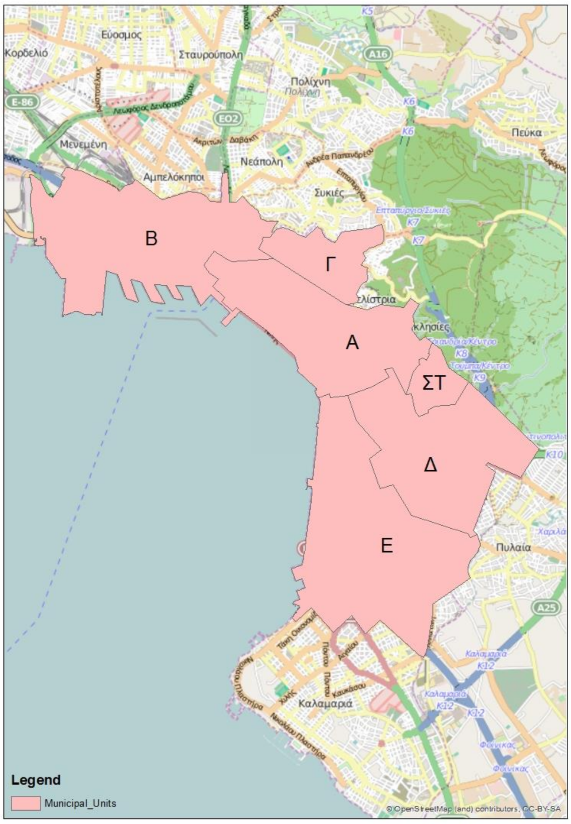


Figure 33: Municipal Districts of the Municipality of Thessaloniki

The Municipality of Thessaloniki was selected as a case study as it was considered suitable and representative for the following reasons:

- It disposed of one of the most well-organised GIS portals in Greece (http://gis.thessaloniki.gr/gis2014/), which offered invaluable spatial information used in this study, such as public spaces (parks, squares etc.), parking facilities, health facilities, public transport means lines/stations and many more. Most of this information was considered as extremely useful when assessing property values since values are directly or indirectly affected by numerous spatial factors.
- It is the second-largest municipality in Greece in terms of population (c.325,000 inhabitants) and the fifth largest in terms of population density (c.17,000 / sq km). This implies a dynamic property market where property values are being determined by demand and supply levels.
- It disposes of features the impact of which on property values are worth to be studied. For example, it has a large frontage on the sea which usually results in

higher values for those properties that are closer to it. Its city centre is historic and commercially active which might also command higher values for properties that are close/in it, while it also has low and high-income areas with different features and residents' profile.

- It was possible to collect a considerable amount of comparable data (c.4,450 properties) from the Bank of Greece (Real Estate Market Analysis Section), which allowed for the application of the selected data-hungry statistical methods (OLS & GWR).
- It is the home-town of the author which facilitated the proper interpretation of the results. Given that the spatial features of an area have a key role in the valuation process, good knowledge of the various features and particularities of the subareas was deemed crucial.
- It consists of typical urban subareas. Thus, the results of the study could be applied to and/or compared with other areas of the country that dispose of similar features.

For the needs of this study, the selected reference unit was **residential apartments that comprised horizontal ownerships in the urban environment** (Municipality of Thessaloniki). The main reasons for this were: (a) residential properties correspond to 80% of the total building stock in Greece; thus the benefits of this sector from such a study are self-explanatory, (b) Greece has amongst the highest rates of self-occupied residential units in Europe, which demonstrate the importance of this property sector to Greeks, (c) urban areas are considered more dynamic property markets compared to rural areas; thus property values are being determined by demand and supply levels, (d) apartments (horizontal ownerships) is a type of property with more common and homogenous features compared to other types of properties (e.g. residential villas, offices, retail shops etc). Thus, the features affecting the values could be much better determined (e.g. floor, age, size, views etc).

5.1.3.1 Data Collection & Processing

The data used in this study were collected primarily through secondary sources, but also, where deemed necessary, primary data were collected or calculated. The data collected referred to map layers (shapefiles) in ArcGIS, comparable properties database (sample), but also spatial data for each property as calculated in ArcGIS. More analytically, the data sources were as follows:

- **Map layers (shapefiles) in ArcGIS**: The map layers were provided by the Urban Planning Department (GIS dept.) of the Municipality of Thessaloniki and are as follows:
    - Public transport means (bus) lines and stops
    - Public areas, such as parks, sports facilities and squares
    - Educational units, such as nursery schools, kindergartens, primary schools, secondary schools, high schools, universities and others
    - Hotel units of all categories
    - Health units, such as hospitals, pharmacies and others
    - Parking facilities, including public spaces, charge-free spaces, short-term spaces and bicycle spaces
    - Urban blocks with their respective planning regulations (e.g. minimum frontage and surface, building coefficient, site coverage ration etc)
    - Municipal districts of the Municipality of Thessaloniki
    - Post code areas

All the above layers are illustrated in the following figures (the whole municipality of Thessaloniki and part of it):

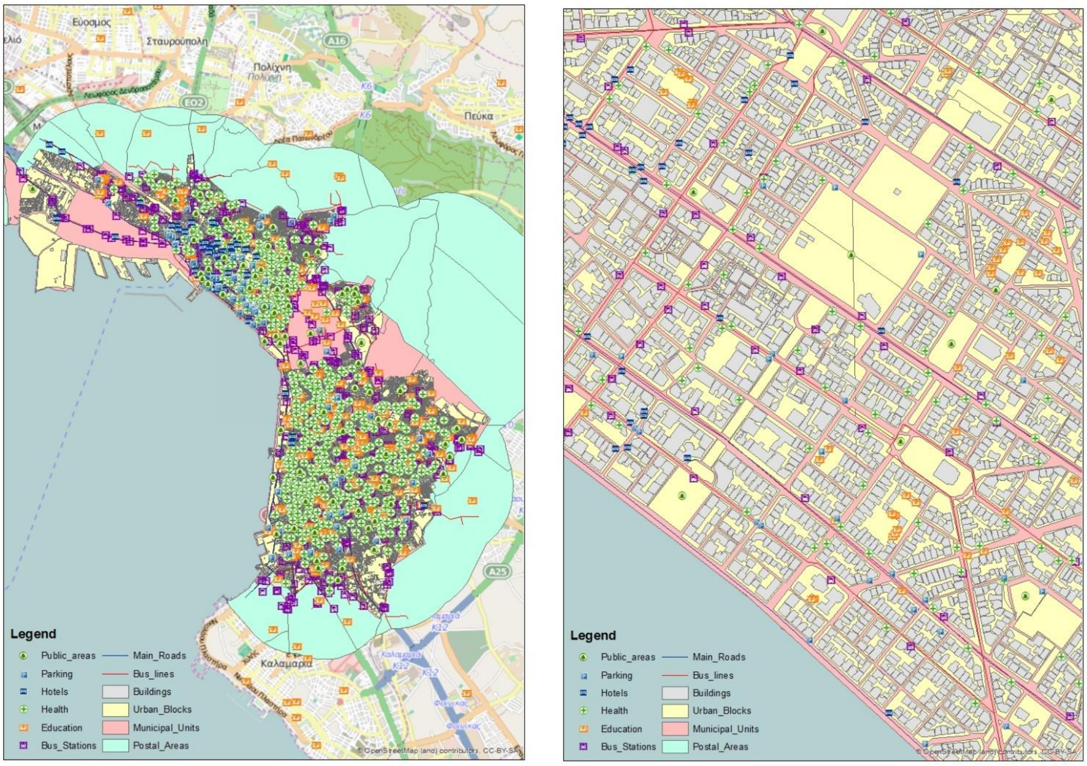


Figure 34: Map layers of the Municipality of Thessaloniki (left figure) and at part of it (right figure)

- **Comparable properties database (sample)**: The property sample is perhaps the most important component for the implementation of an effective prediction model based on the selected statistical methods. Data were provided by the Bank of Greece (Real Estate Market Analysis Section), though it should be noted that the values of the properties collected referred to estimated market values (by qualified valuers) and not actual transacted prices. However, for the needs of this study these values were deemed very satisfactory. The data that were requested refers to residential apartments within the boundaries of the municipality of Thessaloniki and for the period between Jan. ’10 and Jun. ’14. Despite the fact that the street number was not provided (for confidentiality purposes), the street name and its postcode were provided based on which it was made possible to georeference as accurately as possible each property on the map. The sample consisted of 4,435 entities (comparable properties) each with the following data (Bank of Greece, 2015β):

- Type of property
- Address (street, postcode, district)
- Valuation date
- Age
- Floor number
- Surface (main use areas)
- Number and surface of storage and parking space(s)
- Other features (excellent construction quality, position/view/environment, recently renovated, depreciated district)
- Market Value
- Other

- **Other secondary data**: It was deemed necessary to collect additional secondary data that would facilitate the study. More specifically, the price index of residential properties was collected by Bank of Greece (Real Estate Market Analysis Section), which allowed the econometric analysis of values and their transformation into present values for comparison purposes.

5.1.3.2 Initial data processing

It should be noted that the sample provided by the Central Bank of Greece had numerous errors and omissions that had to be corrected. After those transformations, the initial sample of 4,435 came down to 2,583 entities. Also, for consistency and comparability purposes, the values of all entities were referenced into Q3’14 terms as they were referring to the period Q1’10-Q2’14. This was achieved by the price index calculated by the Bank of Greece. Furthermore, it was deemed necessary to add further data in the database.

5.1.3.3 Geocoding

All properties were imported in ArcGIS at their exact geographical position through the process of geocoding (by using MS Office Excel '07 add-on application called 'Terra Excel Geocoder'). The results of this process are demonstrated in the following figures:

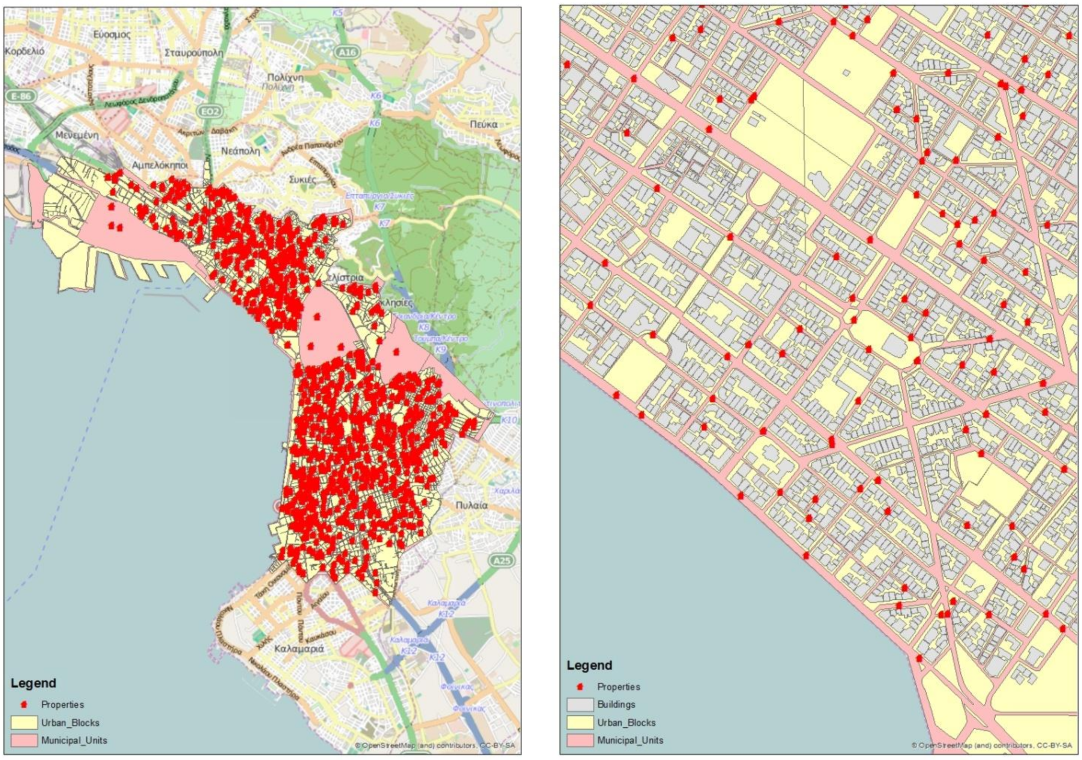


Figure 35: Geocoded properties in the Municipality of Thessaloniki (left figure) and in part of it (right figure). Source: Municipality of Thessaloniki. data processing by author.

5.1.3.4 Spatial data calculation

According to the literature review, but also based on working experience, a number of spatial factors that might affect property values were selected. Each of those factors was treated as an independent variable in the regression models in an attempt to determine and measure their impact on property values. These factors were divided into two broad categories; (a) Points of Interest (POIs), and (b) Spatial Features (SF), as shown in the table below:

Table 14: Points of Interest (POIs) and Spatial Features (SF)

Category	Code	Description	Spatial relationship
POIs	Educ200	Number of educational units	within a 200m zone from the property
POIs	BusStat200	Number of bus stations	within a 200m zone from the property
POIs	BusLine50	Number of bus lines	within a 50m zone from the property
POIs	Health100	Number of health units	within a 100m zone from the property
POIs	Parking	Distance from nearest parking facility	in meters (from the property)
POIs	Parks200	Public spaces (parks, squares, sports facilities)	within a 200m zone from public space 0=falls outside, 1=falls within
SF	Seafront	Sea frontage	0=without frontage, 1=with frontage
POIs	MRoad50_01	Main road axis	within a 50m zone from main road axis 0=falls outside, 1=falls within
SF	CBD_01	Central Business District (CBD)	0=falls outside, 1=falls within
POIs	CityCentre	Distance from city centre	in meters (from the property)

As regards spatial features, they are polygons where every property that falls within it takes the value 1 and all others the value 0. The two SF factors are the 'Seafront' and 'CBD_01'; properties with a sea frontage usually show high values due to the unobstructed sea views and, similarly, properties located in the CBD show higher values due to their convenient location, easy access to shops, public authorities etc. and proximity to the commercial activity of the city. The following figures demonstrate the respective polygons as well as the properties that fall within each polygon (green color).

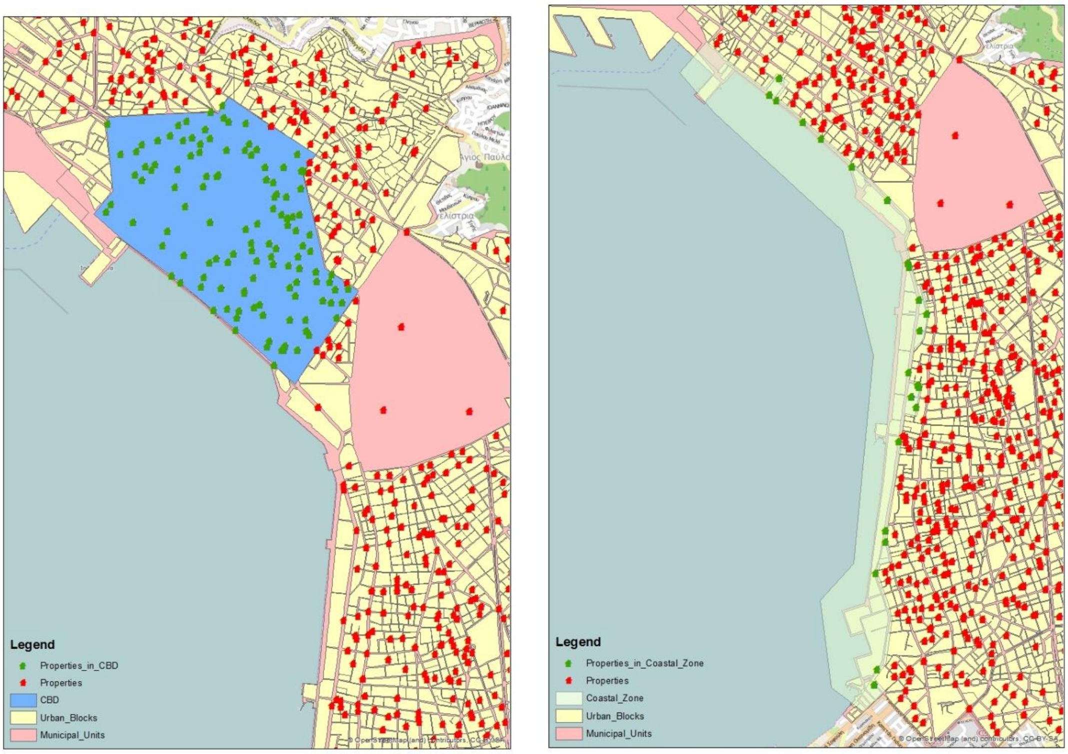


Figure 36: Properties that fall within the CBD (left figure) and the Seafront polygons (right figure)

As regards points of interest, these are point features, and various tools from ArcToolbox were used in order to measure distances, buffer zones etc. from the sample properties. Despite the fact that no drivetimes polygons were calculated (through network analysis) due to lack of necessary layers, measurements were still deemed satisfactory. The following factors/variables were calculated through ArcGIS:

- The number of educational units within a range of 200m from each property. Proximity to schools or universities usually had a positive impact on property values due to easy access to students.
- The number of public transport (bus) stops within a range of 200m from each property. Again, the higher this number, the higher the value, as it implied very convenient access.
- Number of public transport (bus) lines within a range of 50m from each property. It was expected that the higher this number, the lower the value, since it could have had a negative impact due to increased traffic, noise, environmental pollution, and parking difficulties.
- A number of health units (pharmacies, hospitals etc) within a range of 100m from each property. It was expected that the higher this number, the lower the price, since it would imply high population density (in order to serve their needs), thus the low quality of the built environment and life as well as parking difficulties.
- The distance of each property from the nearest parking space. Proximity to such a space might have had a positive impact on property values since it could save valuable time from on-street parking search.
- Properties that fell within a range of 200m from a public space (e.g., park or square) got the value of 1, while properties outside this zone got the value of 0. It is expected that proximity to a park or square might have had a positive impact on property values.
- Properties that fell within a range of 50m from main road axis. It was expected that the values of those properties might either have been negatively affected due to increased traffic, noise, environmental pollution, and parking difficulties, or positively affected due to easy accessibility.

- The distance of each property from the city centre. Proximity to the city centre might have had a positive impact on property values, due to longer drive times to the commercial area, public authorities etc., but also the fact that the use of cars seemed inevitable.

The following figures demonstrate extracts from ArcGIS during the calculation process of the aforementioned factors/variables.

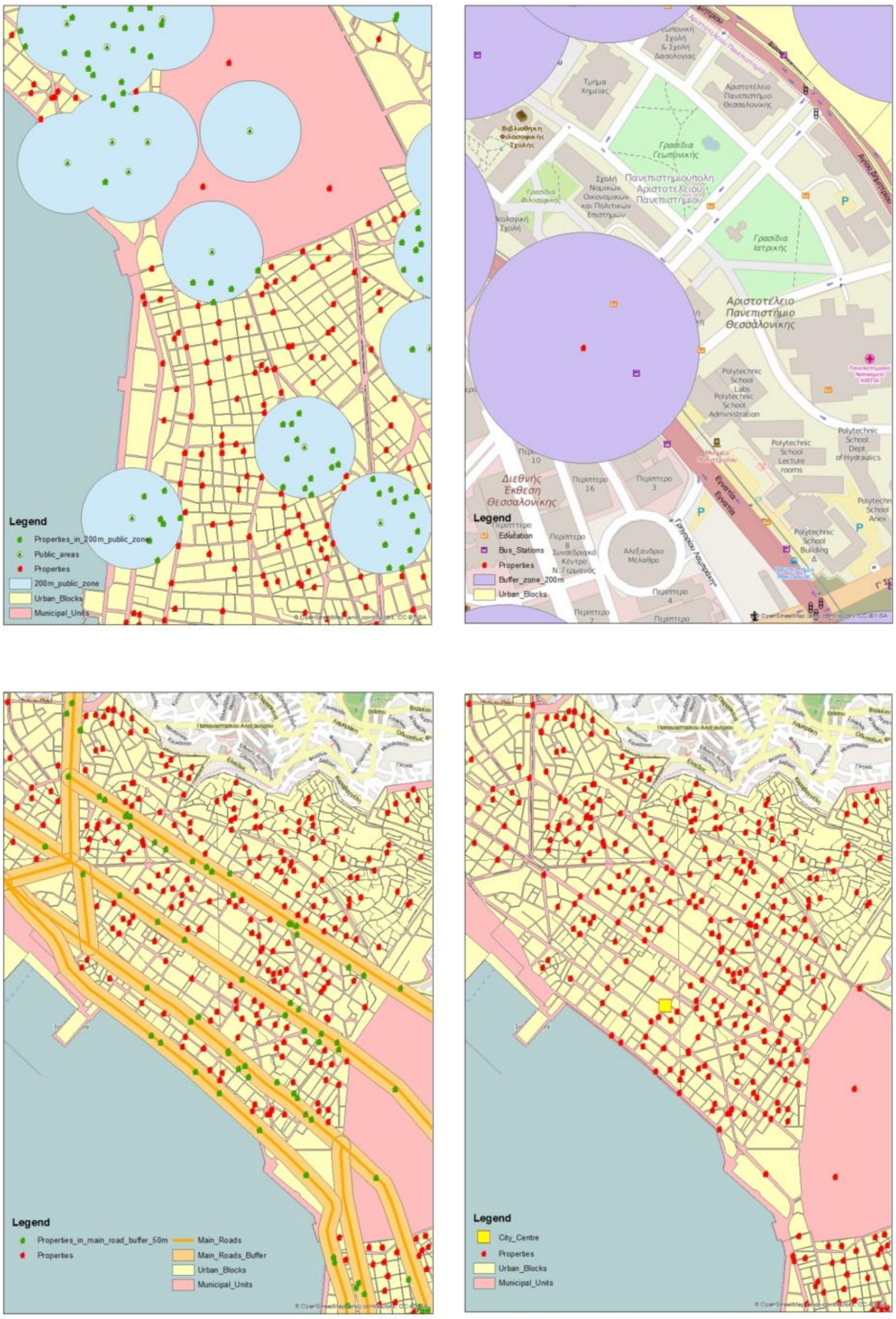


Figure 37: Extracts from the calculation process of variables in POIs category

Note: Properties (green color) that lay within a 200m range from public spaces (up left figure), Number of educational units and bus stops within a 200m range from each property (upright figure), Properties (green color) that were within a 50m range from main road axis (down left figure), Distance of each property from the city centre (yellow color) (downright figure)

The final table of properties included the following variables:

Table 15: Variables that are included in the final table of properties

Variable	**Description**
S_N	Serial Number
ValTotWeig	Market Value
Age	Years since it was built
Floor	Level that the property is located
Area_Main	Surface of property (sq m)
Storages	Number of storages that belong to the property
ParkSpaces	Number of parking spaces that belong to the property
Poor_Qual	Poor quality in terms of location, area, construction quality, state of repair (Y/N)
Good_Point	Number of superior features (e.g. excellent views, renovated and privileged location)
Educ200	Number of educational units within a 200m range from each property
BusStat200	Number of bus stops within a 200m range from each property
BusLine50	Number of bus lines within a 50m range from each property
Health100	Number of health units within a 100m range from each property
Parking	Distance from nearest parking facility from each property
Parks200	Properties that lie within a 200m range from public space (e.g. park, square etc) (0/1)
Seafront	Properties that lie in the seafront zone (0/1)
MRoad50_01	Properties that lie within a 50m range from main road axis (0/1)
CBD_01	Properties that lie within the City Business District (CBD) (0/1)
CityCentre	Distance of each property to the city centre

5.1.3.5 Results & Commentary

As discussed, data were initially analysed in SPSS in order to determine the variables that are significant for the prediction of property values. The following table presents the descriptive statistics for all variables.

Table 16: Descriptive stats for all variable and all property sample of 2,583 entities

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
ValTotWeig	2583	1,247,266	4,633	1,251,899	99,094.7	1,697.0	86,245	7438175130
EUR/sqm	2583	5,637	159	5,796	1,114.7	10.2	521	271,089
Age	2583	92	0	92	31.8	.3	17	298
Floor	2583	10	0	10	2.8	.0	2	4
Area_Main	2583	589	10	599	84.7	.8	39	1,519
Storages	2583	3	0	3	.1	.0	0	0
ParkSpaces	2583	3	0	3	.0	.0	0	0
Poor_Qual	2583	1	0	1	.0	.0	0	0
Good_Point	2583	3	0	3	.3	.0	0	0
Educ200	2583	15	0	15	2.7	.1	3	8
BusStat200	2583	10	0	10	2.2	.0	2	2
BusLine50	2583	1	0	1	.4	.0	0	0
Health100	2583	6	0	6	.9	.0	1	1
Parking	2583	1,349	5	1,354	421.3	4.9	250	62,455
Parks200	2583	1	0	1	.5	.0	0	0
Seafront	2583	1	0	1	.0	.0	0	0
MRoad50_01	2583	1	0	1	.1	.0	0	0
CBD_01	2583	1	0	1	.1	.0	0	0
CityCentre	2583	7,042	95	7,137	3,322.2	34.7	1,761	3,101,608
Valid N (listwise)	2583							

It should be noted that the sample was divided into two parts: (a) 90% of the properties (2,337 entities), also called 'Input Group', that was used for the creation of the models, and (b) the rest 10% of the properties (246 entities), called 'Control Group', was not used in the model formation but in its verification and statistical check.

5.1.3.6 OLS method

After a thorough statistical analysis on SPSS by using the MRA method, a prediction model that satisfied all statistical checks (significance of independent variables, independency of entities, multicollinearity, Homoscedasticity, normality, and linearity) was created. It should be noted that a number of outliers and influential points (116 entities) were removed from the sample in order to improve its credibility and prediction power (indeed, the coefficient of determination was significantly improved from R^2=61.2% to R^2=75.9%). Also, the results were exactly the same by using the OLS method in ArcGIS. The following table demonstrates the position of the 'Control Group' entities (246 properties), those entities that were removed from the sample as outliers or influential points (116 properties), and the 'Input Group' entities (2,221 properties).

The final prediction model is presented below:

Value = **– 8.503 + 1.221 x** [Surface] **– 679 x** [Age] **+ 36.606 x** [falls within CBD] + **138.934 x** [falls within seafront zone] **+ 2.923 x** [Floor] **+ 2 x** [distance from city centre] + **5.138 x** [number of superior features] **+ 5.853 x** [number of storages] **– 1.614 x [number of health units within 100m] – 11.489 x** [poor quality] **[5.3]**

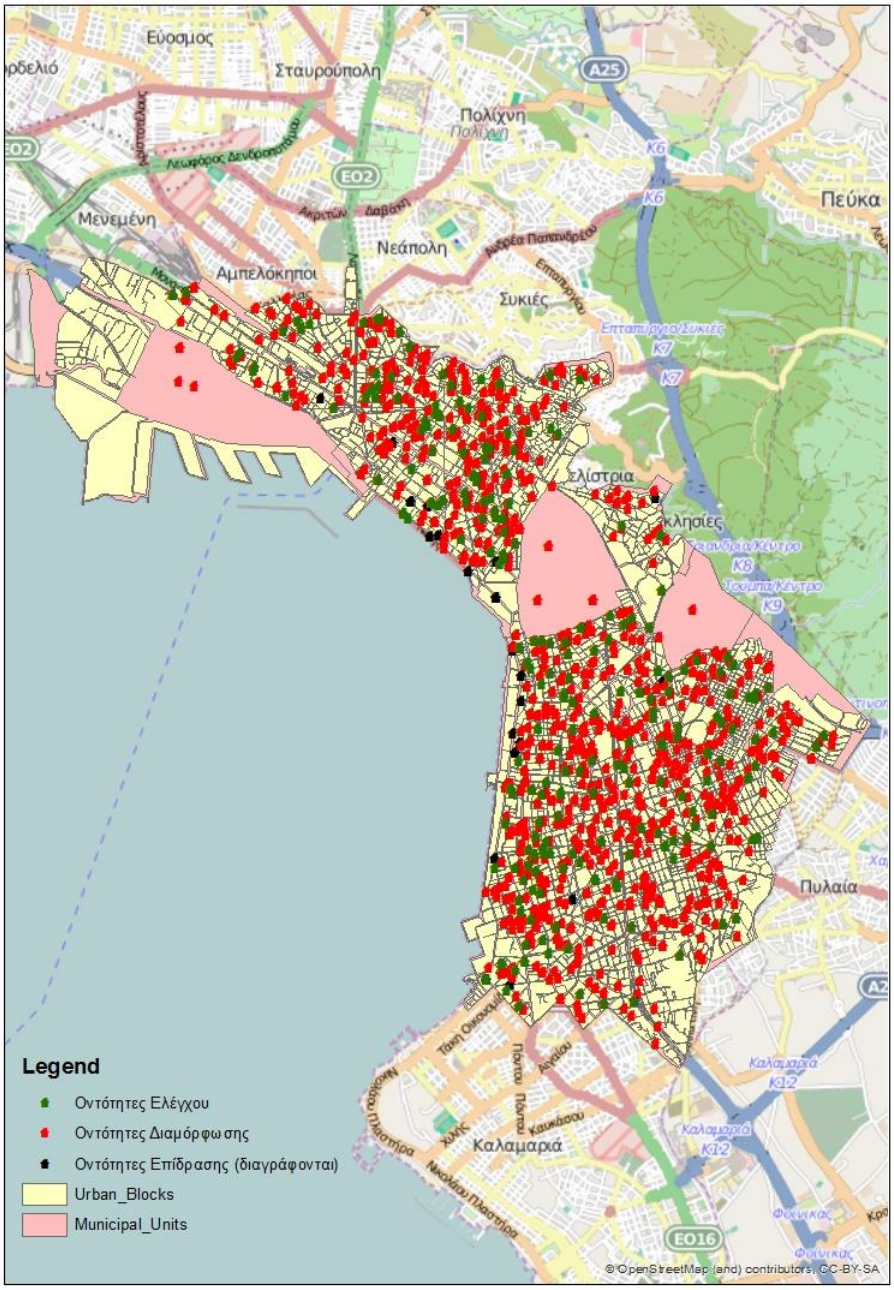


Figure 38: Illustration of entities groups (control group in green color, input group in red color and outliers/influential points in black color).

The model above showed a very satisfactory coefficient of determination R^2=75.9% and the Akaike's Information Criterion (AICc), which was used for comparison with the GWR model, is 51,450 (models with low AICc value are preferred). The results of the OLS method are summarized in the following report table:

Table 17: Report table of statistical results of OLS method in ArcGIS

Summary of OLS Results

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	-8502,586074	3037,983124	-2,798760	0,005177*	2817,219968	-3,018077	0,002584*	--------
AGE	-678,547506	35,219753	-19,266106	0,000000*	40,117377	-16,914055	0,000000*	1,221853
FLOOR	2923,477759	301,052982	9,710841	0,000000*	314,256923	9,302827	0,000000*	1,115536
AREA_MAIN	1220,738300	18,442812	66,190464	0,000000*	23,783462	51,327191	0,000000*	1,103468
STORAGES	5852,740656	1698,730642	3,445361	0,000597*	1869,219763	3,131114	0,001779*	1,154343
PARKING	2507,443203	3020,500277	0,830142	0,406535	3602,452309	0,696038	0,486473	1,178317
POOR_QUAL	-11489,355520	4634,898712	-2,478880	0,013242*	3504,620043	-3,278346	0,001077*	1,015658
GOOD_POINT	5137,902256	1158,679668	4,434273	0,000013*	1231,145637	4,173269	0,000037*	1,035063
EDUC200	-380,992376	210,385346	-1,810926	0,070292	208,418006	-1,828020	0,067684	1,117862
BUSSTAT200	-659,704730	395,802190	-1,666754	0,095717	374,330830	-1,762357	0,078154	1,226604
BUSLINE50	1982,751994	1422,967749	1,393392	0,163655	1485,005865	1,335181	0,181965	1,578619
HEALTH100	-1613,896187	599,280899	-2,693055	0,007131*	574,068095	-2,811332	0,004981*	1,173104
PARKING1	-0,186148	2,468103	-0,075421	0,939870	2,430932	-0,076575	0,938952	1,227427
PARKS200	483,695429	1136,920238	0,425444	0,670570	1153,258183	0,419416	0,674968	1,068100
MROAD50_01	-190,703551	1859,453547	-0,102559	0,918306	1895,146435	-0,100627	0,919839	1,378759
CBD_01	36605,843435	2338,645619	15,652582	0,000000*	2621,851108	13,961832	0,000000*	2,115415
CITYCENTRE	2,054066	0,420924	4,879898	0,000002*	0,407615	5,039235	0,000001*	1,802058
SEAFRONT	138933,542377	8302,377742	16,734187	0,000000*	8283,026527	16,773282	0,000000*	1,028644

OLS Diagnostics

Input Features:	Οντότητες Διαμόρφωσης	Dependent Variable:	VALTOTWEIG
Number of Observations:	2221	Akaike's Information Criterion (AICc) [d]:	51450,420330
Multiple R-Squared [d]:	0,759280	Adjusted R-Squared [d]:	0,757423
Joint F-Statistic [e]:	408,748024	Prob(>F), (17,2203) degrees of freedom:	0,000000*
Joint Wald Statistic [e]:	4568,275970	Prob(>chi-squared), (17) degrees of freedom:	0,000000*
Koenker (BP) Statistic [f]:	289,385674	Prob(>chi-squared), (17) degrees of freedom:	0,000000*
Jarque-Bera Statistic [g]:	840,437971	Prob(>chi-squared), (2) degrees of freedom:	0,000000*

Having estimated a satisfactory prediction model, its prediction accuracy should have been controlled. To achieve this, the model was applied to the 'Control group' of entities which corresponded to 10% of the whole sample. It is reminded that those entities were not included in the model formation in order to not affect it. This control took place in SPSS by calculating the Pearson Correlation factor between the actual property values and the predicted ones. This factor equaled 0.87, which was considered as very satisfactory as the value of 1.0 means 100% prediction accuracy of the model. The very good fit of the model is depicted in the following figure of the correlation between the actual and predicted values.

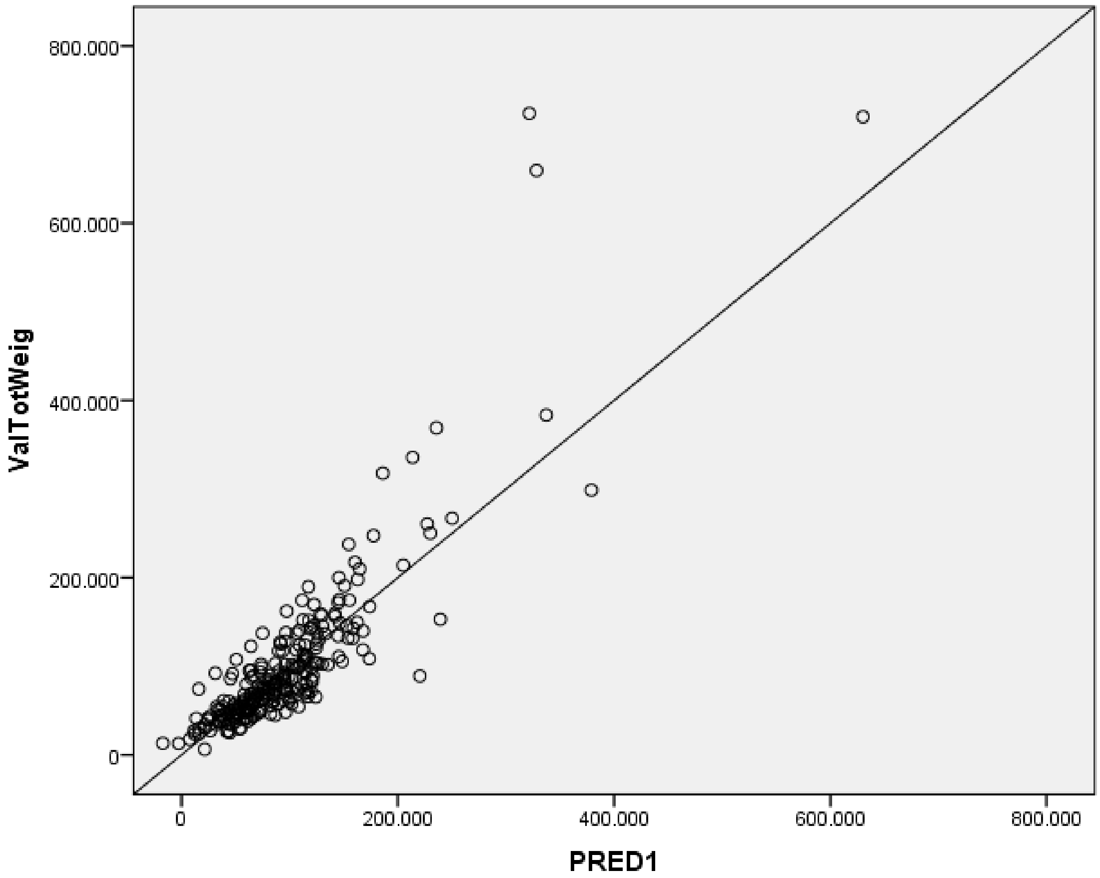


Figure 39: Correlation diagram between actual and predicted values (10% of the sample size - 'Control Group' of entities) (OLS method)

5.1.3.7 GWR method

As mentioned, GWR is a local regression method that highlights spatial differences instead of spatial similarities between entities in a given area. This means that independent variables that have spatial concentration should not be included in the model (e.g., properties that fall within the CBD or seafront zone), as they create a multicollinearity problem. After many trial-and-error efforts, the four variables Health100, Seafront, CBD_01, CityCentre, and Poor_Qual were excluded from the model, which remains with five independent variables (Age, Floor, Area/Surface, Number of storages, Number of superior features). Given that GWR created one equation for each property based on nearby properties, any spatial features affecting property values were taken into consideration.

Again, the created GWR model satisfied all statistical checks. The coefficient of determination (R^2=83.3%) was much higher than that of OLS method, and its Akaike's Information Criterion (AICc) (51,224) was lower than that of the OLS method (lower AICc values are preferred). Those two parameters demonstrated that the GWR model was superior compared to the OLS one. The report table of statistical results of GWR method is presented below:

Table 18: Report table of statistical results of GWR method in ArcGIS

Table

GeographicallyWeightedRegression103_supp

OBJECTID *	VARNAME	VARIABLE	DEFINITION
1	Neighbors	125	
2	ResidualSquares	1021277140936,1146	
3	EffectiveNumber	355,119705	
4	Sigma	23395,36992	
5	AICc	51223,681374	
6	R2	0,832713	
7	R2Adjusted	0,800965	
8	Dependent Field	0	ValTotWeig
9	Explanatory Field	1	Age
10	Explanatory Field	2	Floor
11	Explanatory Field	3	Area_Main
12	Explanatory Field	4	Storages
13	Explanatory Field	5	Good_Point

The significant advantage of the GWR method, and ArcGIS accordingly, was the creation of a series of thematic maps (raster layers) of the variables' coefficients and the constant term. This allowed for the identification of spatial differentiations within the study area, which could assist in effective decision-making. Through these maps, it was possible to obtain excellent insight into the key parameters that affect property value in a specific area. For example, the age of a property may have a significant negative impact on its value in a newly developed region where the majority of properties are brand new. On the other hand, it may have a positive effect on the city centre where older buildings dispose of extreme architectural features and historic significance.

To shed more light on this, the results of 'Age' coefficient are presented. As seen in the following figure, its values show a relative variation with of most of them being negative while there are also some positive values. As someone would expect, in the majority of cases, age was inversely proportional to property value, since the older the property, the lower its value due to obsolescence, deterioration, and depreciation. However, in some parts of the study area, property values were directly proportional to age for the reasons mentioned above. More specifically, coefficient values ranged between c. -€1,800 and +€280, the average was c.-€730 and the standard deviation €370.

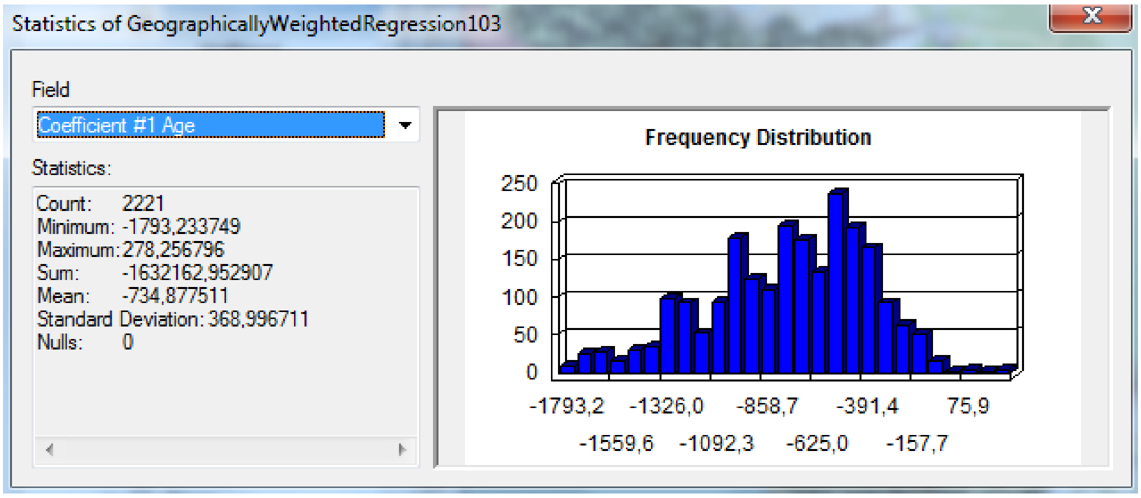


Figure 40: Statistics and frequency distribution of the variable 'Age' coefficient (GWR method)

Interestingly, as noticed in the following figure, positive values of the coefficient were concentrated in the historic city centre of Thessaloniki, where there are many neoclassic listed buildings. For those buildings, age is not usually a negative factor due to their rarity and character. On the other hand, the highest negative values of the variable’s coefficient were concentrated in the affluent south-east suburbs of the city, where brand new properties are sought after by buyers; demand is low for old properties. Thus, age had a negative impact on property values.

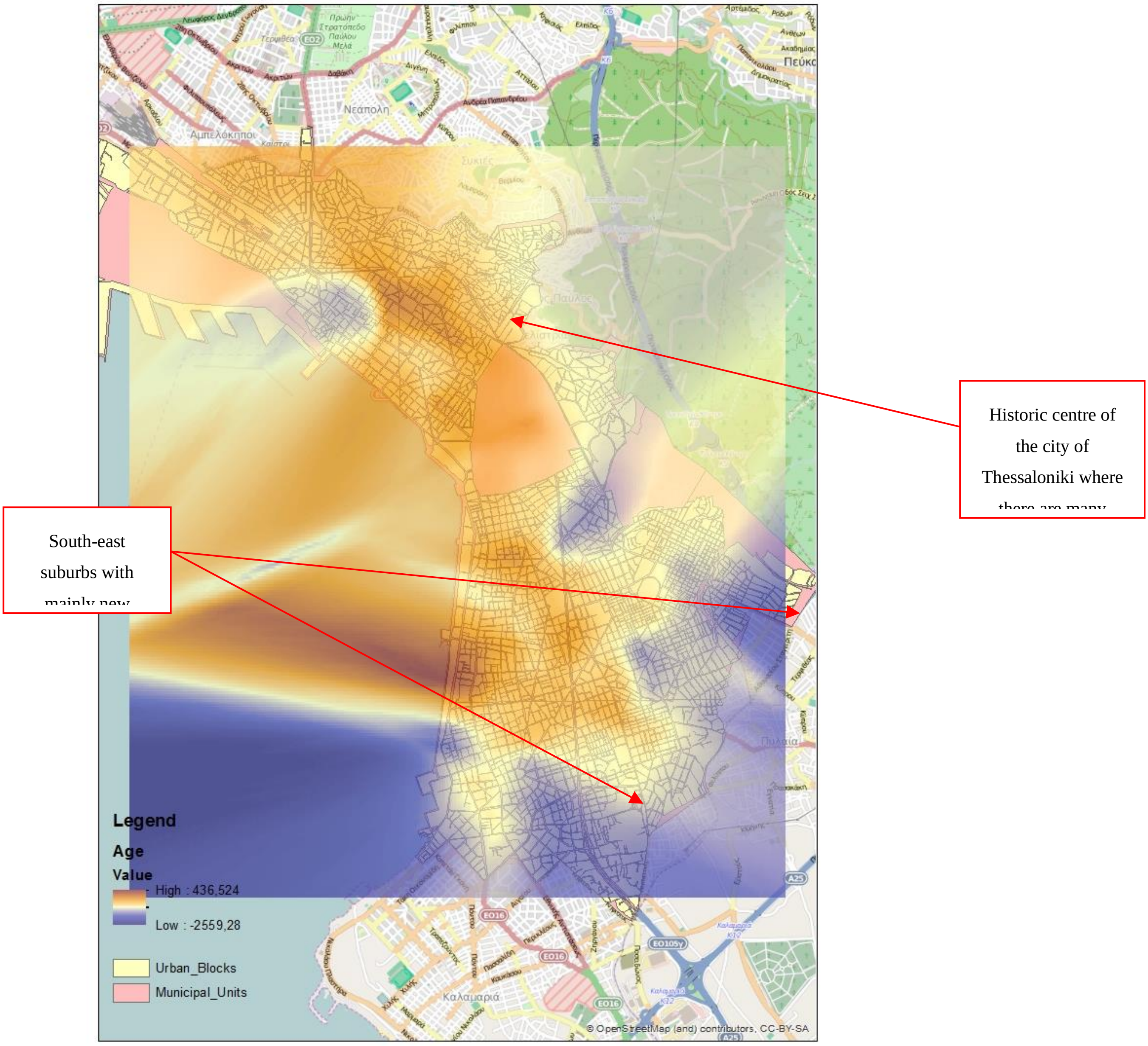


Figure 41: Thematic map of coefficient values of 'Age' variable (GWR method)

Having estimated a satisfactory prediction model through GWR method, its prediction accuracy should have been controlled. ArcGIS automatically calculated predicted values for 10% of entities ('Control Group'). The following table presents the results of this process, which shows that each property (or groups of properties) got different local R^2, predicted values, and coefficients' values.

Table 19: Results table of the GWR method with predicted values per property, coefficients' values and other useful information

Table

feature_class_GWR103

FID	Shape *	Cond	LocalR2	Predicted	Intercept	C1_Age	C2_Floor	C3_Area_Ma
0	Point	13,220509	0,797704	99245,80696	5480,57095	-1286,788753	6717,901622	1434,104618
1	Point	8,550179	0,901939	124051,734006	29924,784632	-1651,336873	3945,78239	1249,169797
2	Point	8,095496	0,87343	250291,616124	29226,953258	-1597,354448	3485,756538	1245,721779
3	Point	8,095496	0,87343	246805,859585	29226,953258	-1597,354448	3485,756538	1245,721779
4	Point	8,9915	0,897877	121205,291673	17626,813071	-1489,658933	5284,595245	1283,985234
5	Point	13,601658	0,787308	166291,965441	10402,698994	-1417,137639	6146,639429	1476,521208
6	Point	13,518971	0,774137	309303,097855	10183,85163	-1475,529999	5603,826712	1519,318817
7	Point	12,84878	0,796966	186158,743866	-6256,67374	-1170,274574	6254,541757	1516,090949
8	Point	13,293507	0,785178	120956,625027	-2719,759799	-1264,682571	5884,030357	1537,807758
9	Point	13,293507	0,785178	210445,641417	-2719,759799	-1264,682571	5884,030357	1537,807758
10	Point	12,099448	0,820896	138610,944005	-18021,543847	-980,322489	6446,024899	1520,120594
11	Point	8,956141	0,732007	127047,688228	18226,033011	-1267,502438	4382,598511	1140,719442
12	Point	9,918665	0,865971	184533,621283	-12365,472065	-1068,842112	5469,378372	1424,196558
13	Point	8,979279	0,848079	54278,162109	3680,128721	-1307,644034	4399,91178	1354,340825
14	Point	11,148874	0,840378	138088,435269	-19093,536635	-903,7261	5600,584311	1488,34594
15	Point	9,567733	0,847495	95963,124776	-7837,724025	-962,261664	4503,837203	1378,355499
16	Point	8,858201	0,817385	42652,562005	7817,208996	-1075,161488	3328,290324	1278,335849
17	Point	10,560994	0,665609	118098,643755	21036,142741	-1103,675332	3051,228946	1076,527848
18	Point	9,313071	0,826146	64685,214507	11499,500859	-1189,231648	4919,8531	1144,222586
19	Point	10,411156	0,866027	65142,269924	-4065,888955	-913,929397	3642,758939	1356,298103
20	Point	9,45095	0,839872	150512,880775	10563,232751	-1235,18457	5270,419218	1160,20578
21	Point	10,989521	0,770554	69651,423992	27270,378159	-1253,419749	1548,729095	1136,024785
22	Point	12,285384	0,840559	60197,715432	5747,57327	-1047,565443	3657,942406	1310,092423
23	Point	13,371067	0,789849	152951,812949	18699,252528	-1158,444962	3045,446821	1237,181528
24	Point	12,383367	0,837285	115604,942899	1857,928501	-808,403882	2885,421809	1243,04703
25	Point	13,569503	0,835995	112179,339496	8554,909107	-872,68521	2963,679057	1189,479115
26	Point	11,262354	0,77544	80943,970522	11049,008491	-781,398743	3039,652407	1087,271272
27	Point	11,484462	0,789108	49141,111122	7548,801399	-746,265204	3247,599434	1107,377709
28	Point	14,497051	0,759494	46146,131036	22007,677388	-961,292169	3079,350729	1082,358124
29	Point	11,905294	0,835634	39567,992089	-5607,245775	-566,063118	2563,766272	1219,840814
30	Point	11,905294	0,835634	137994,600272	-5607,245775	-566,063118	2563,766272	1219,840814
31	Point	13,729433	0,784677	115491,47725	8532,497129	-779,157146	2474,988756	1151,720878
32	Point	13,729433	0,784677	127498,656574	8532,497129	-779,157146	2474,988756	1151,720878
33	Point	13,470774	0,636109	117172,518375	19058,776294	-855,512451	3183,343025	1097,830888
34	Point	12,374185	0,785064	139864,556473	-8145,549593	-575,306917	1458,401144	1289,786411
35	Point	13,434516	0,522682	86382,73105	24385,355008	-1057,026933	2164,732628	1246,828413
36	Point	13,580956	0,740606	69346,431125	12306,301056	-758,918964	2870,87362	1098,854138
37	Point	12,181636	0,799441	110116,002876	-8194,537181	-500,354602	2075,067617	1235,54994
38	Point	12,581631	0,634007	58235,032228	11283,869822	-593,013779	2722,375305	1053,348592
39	Point	13,293222	0,80006	93078,966823	-14726,778735	-492,419669	3332,489497	1252,719231
40	Point	13,094688	0,751763	88463,166475	4485,807614	-588,477145	1984,389338	1130,812709
41	Point	12,771795	0,796361	48164,695678	1565,170938	-629,586775	1699,065858	1177,550221
42	Point	12,555625	0,785067	64783,258713	-1623,589019	-603,278973	1889,246983	1203,962412
43	Point	12,555625	0,785067	122663,246709	-1623,589019	-603,278973	1889,246983	1203,962412

As a next step, the prediction accuracy of the created GWR model was performed in SPSS by calculating the Pearson Correlation factor between the actual property values and the predicted ones. This factor equaled 0.84, which was considered very satisfactory despite the fact that it was slightly lower than that of OLS method. The very good fit of the model is depicted in the following figure of the correlation between the actual and predicted values.

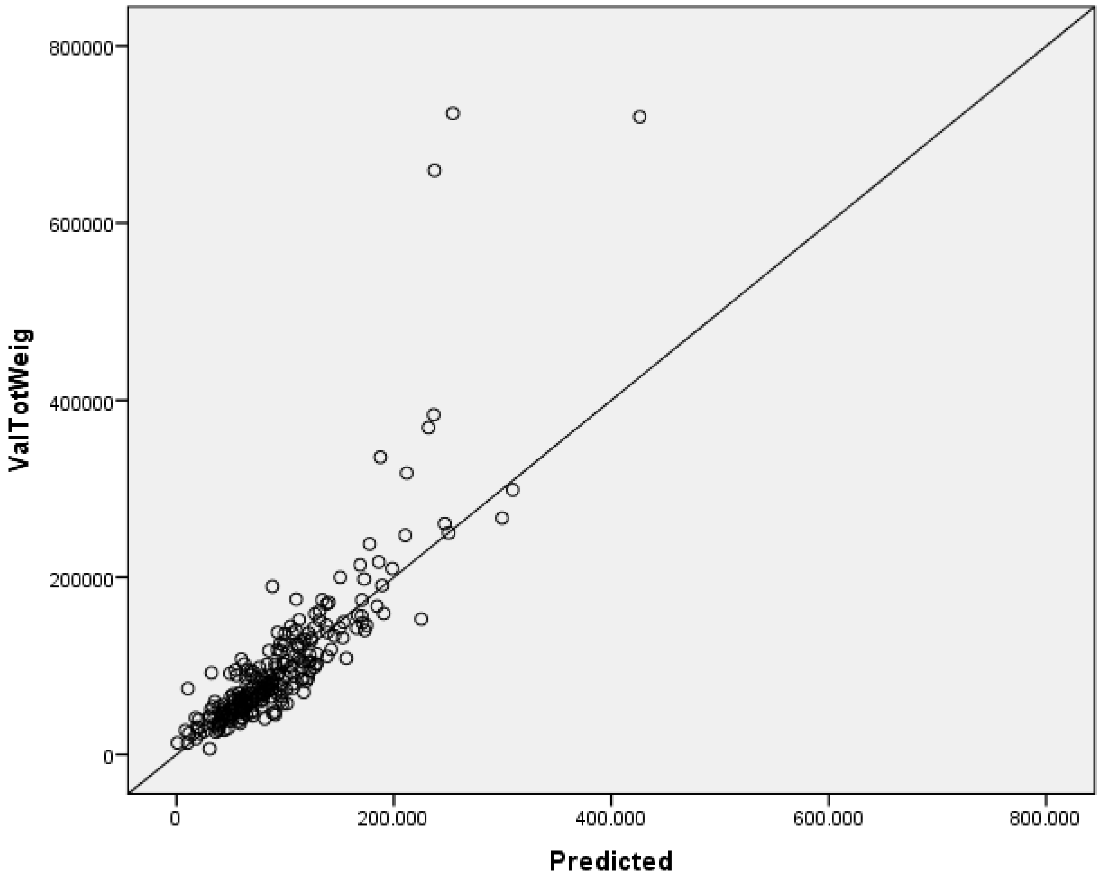


Figure 42: Correlation diagram between actual and predicted values (10% of the sample size – ‘Control Group’ of entities) (GWR method)

5.1.3.8 Conclusions

Two prediction models were created by applying the OLS and the GWR methods in ESRI ArcGIS 10.1. Both models showed very good fit for the input data, passed all statistical checks, and had excellent prediction accuracy. However, as shown through the analysis, GWR method had several advantages compared to the OLS one. The following summary table shows an immediate comparison between the two methods to extract useful conclusions.

Table 20:Summary table of comparison between the OLS and GWR methods

	OLS	GWR
Statistical checks	✔	✔
Coefficient of determination R2	75.8%	83.3%
Akaike Information Criterion (corrected)	51,450	51,224
Prediction accuracy (Pearson correl.)	0.870	0.840
Number of independent variables	10	5
Type of input data	Property & Spatial data	Property data
Coefficient per variable	One	Many

More specifically, the GWR method had a much higher coefficient of determination R2 which means that the created model(s) fit much better in the data. The higher percentage showed that the dependent variable (property value) was better explained (by 83.3%) from the selected independent variables. In comparison, this percentage was lower in the OLS method (75.8%), though still high. The AICc value of GWR model was lower compared to that of OLS (51,224 versus 51,450 respectively). According to theory, the absolute value of AICc does not mean anything, though models with lower AICc are preferred. As regards the predictive accuracy, both models had a very high Pearson correlation factor (OLS: 0.87, GWR: 0.84) which demonstrated their very good prediction accuracy.

Perhaps the most significant advantages of the GWR method are the three last points, which relate to the ability of this method to create different local equation(s) for each entity or a set of entities in the nearby area. This ability allows the removal of any variable with spatial concentration, as it creates a multicollinearity problem. In this study, GWR method used only 5 independent variables compared to OLS method, which used 10 variables. It was not only the fact that more variables do not necessarily lead to better results but also more practical difficulty. The calculation of each of the 5 variables' values that were removed from GWR model were both a time-consuming process and required good knowledge of GIS. Additionally, the type of input data that the GWR method required is solely property data (i.e., information like the floor, age, size etc.). In contrast, the OLS method required spatial data as well (i.e., information like distance from sea/city centre/public transport etc.). The calculation of spatial data was again a time-consuming process. Still, most importantly, it required a number of layers that are not always available (e.g., layers of bus stops/lines, educational & health units etc.). Last but not least, the ability of GWR to create thematic maps with the variables' coefficients was a key advantage that greatly assisted decision-making and extraction of invaluable conclusions.

For all the reasons above, it should was concluded that the GWR method can lead to superior prediction models compared to the traditional OLS method. At the same time, it became clear that the use of GIS can have a great positive impact on mass appraisals field through the application of advanced statistical and spatial analysis techniques.

5.2 Accuracy measurement of Random Forests and Linear Regression for mass appraisal models that estimate the prices of residential apartments in Nicosia, Cyprus

The main part of this chapter has been already published at the Journal Advances in Geosciences (T Dimopoulos et al., 2018b).

The purpose of this section was to examine the prediction accuracy of the Random Forests, a machine learning method, when it was applied for residential mass appraisals in the city of Nicosia, Cyprus. The analysis was performed using transaction sales data from the Cyprus Department of Lands and Surveys, the Consumer Price Index of Cyprus from the Cyprus Statistical Service, and the Central Bank of Cyprus' Residential Index (Price index for apartments). The Consumer Price Index and the price index for apartments record quarterly price changes. At the same time, the dependent variables for the computational models were the Declared and the Accepted Prices that were conditional on observed values of a variety of independent variables. The Random Forests method exhibited enhanced prediction accuracy, especially for the models that comprised of a sufficient number of independent variables, indicating the method as prominent. However, it has not yet been utilized adequately for mass appraisals.

The Random Forests (RF) method is broadly used for predictive modeling as well as data analysis. It has been deemed significant in a wide variety of scientific thematic areas, such as Computer Science (Data Mining), Engineering, Medicine, Business, etc. The method is based on the so-called decision trees, which is a machine learning method for classification and regression. Specifically, decision trees are algorithmic structures that consist of nodes and branches. At each node, a decision is made, whether a variable is higher or less than a value. The nodes lead to branches, which in turn lead to another node, with a sequential decision established each time. Through an iterative procedure, the model is trained to predict the desired output (dependent variable) with the minimum possible errors. The method exhibits high accuracy; however, many times, it overfits the training data, making it unable to generalize the predictions. In 2001, the Random Forests was developed by Breiman, as indicated by the name of the method, to utilize a vast group of different decision trees, creating a final result that is an average of the particular trees. This procedure results in high prediction performance (accuracy), as well as the significance of each input variable, which is estimated automatically, regarding its contribution to the prediction errors.

A search in Scopus database of papers, for the terms “random forests” and “mass appraisal” existing in the Title, Abstract or Keywords of a paper, returned only two results: Antipov and Pokryshevskaya (2012), and Pokryshevskaya and Antipov (2011). In the first article, that was cited most (23 times), the authors declared that they investigated the performance of a variety of methods such as multiple regression analysis, artificial neural networks, and others, for the mass appraisal of residential apartments, but that the Random Forests method exhibited highest prediction accuracy. However, a search query in Scopus for the term “random forests” only, existing in the Title, Abstract, or Keywords, resulted in 16,655 papers with the first one (Breiman, 2001) being cited 22,498 times. Furthermore, an “inversed” search in Scopus database, only for the term “regression” in the Title, Abstract or Keywords of a paper, Scopus yields 1,158,892 results. In contrast, the most frequently used one, in terms of citations, is again the RF method by Breiman (2001).

5.2.1 The test case and data processing

In this paper, the accuracy of linear regression was compared with the accuracy of the Random forest method, in order to investigate the latter’s relative performance as a popular regression method, however with limited publications in the mass appraisals research field. The studied database regarded 3527 transactions of residential apartments in Nicosia, obtained from the Department of Lands and Surveys of Cyprus (DLS). The period studied, including transactions from 2008 to 2014, which regarded properties in Nicosia district. The database obtained by the DLS included a vast number of sales features.

The following variables were excluded: Town village name, Planning zone name, Block, Unit built year, declared price over the enclosed extent, declared price over enclosed and covered extent, declared price over total extent, declared price over the adjusted extent, accepted price over the enclosed extent, accepted price over enclosed and covered extent, accepted price over the total extent, accepted price over the adjusted extent, total value 2013 over the enclosed extent, total value 2013 over enclosed and covered extent, total value 2013 over the total extent, total value 2013 over adjusted extent. The database, contains the sales data, the Consumer Price Index of Cyprus as well as the Central Bank of Cyprus residential index (Flats Index). In Table S5 in the Supplement, the analytical description of the fields in the studied Database is presented.

The following variables were excluded because most of them contained too many missing values: District, Quarter, Sheet, Parcel no, Access code, Planning zone code, Secondary planning zone code, DLO file, DLO file year, Remark, Main sbp cat, Status, Building code, Building sbpi id no, Unit code, Field46. For the purposes of the study, some new variables were created, to improve the forecasting:

– Dali Municipality and Latsia Municipality concatenated to Dali-Latsia Municipality, due to a low number of observations (transactions).

– the Unit_condition_mod_code merged class 3 & 4 into one variable (3), as the 4 class had only few observations.

– The time data were converted to count the difference of days between the built date and the age at the year 2013 and then converted to years, by dividing with 365.25

– The sale acceptance quarter categorical variable, was converted to numeric, with regard to the month of the year (< = 3 - > “Q1”, < = 6 - > “Q2”, < = 9 - >“Q3”, > 10 - > “Q4”).

– Furthermore, a combined variable was created, regarding every apartment’s “adjusted” area (AE), utilizing the Unit Enclosed Extent (UEE), the Unit Covered Extent (UCE) and the Unit Uncovered Extent (UUE): $AE = UEE + 0{,}5 * UCE + 0.2 * UUE$ (1)

– Additionally, another variable was created, equal to the accepted price divided by the apartment’s adjusted area (AE).

– The Accepted minus the Declared price was also investigated as the dependent variable. This price is always equal to or bigger than 0, for the above-mentioned reasons.

Last but not least, the following variables, were excluded, as they were not of interest in the study: Fiscal property type, inflation increase sale acceptance date, Share numerator, Share denominator, Unit desc, Sale acceptance date, Unit built date, Sale acceptance month, sale acceptance quarter, Unit built month. The location (municipality) variables were transformed to nominal variables, and the database was rearranged accordingly. After the above-mentioned data preparation and cleaning, an initial exploration of the data was performed by creating frequency tables, scatter plots, and distribution graphs. In

Figure 43, the percentage of condition code per municipality is demonstrated, as well as each variation. In Figure 44, the Histogram of the Declared Price is demonstrated, exhibiting a significant variation, as well as a skewed distribution. In Figure 45, the Accepted Price vs. Unit Enclosed and Covered Extent is demonstrated, where the Price variation exhibits similar shapes among the studied years. as the majority of the observations were before the 2012–2013 Cypriot financial crisis. Finally, as it appears in Figure 46 and Figure 47, (Accepted Price vs. Municipality and Accepted Price vs. Planning Zone) the region of Engomi and Zones Ea3, Ka7 depicts the highest apartment prices.

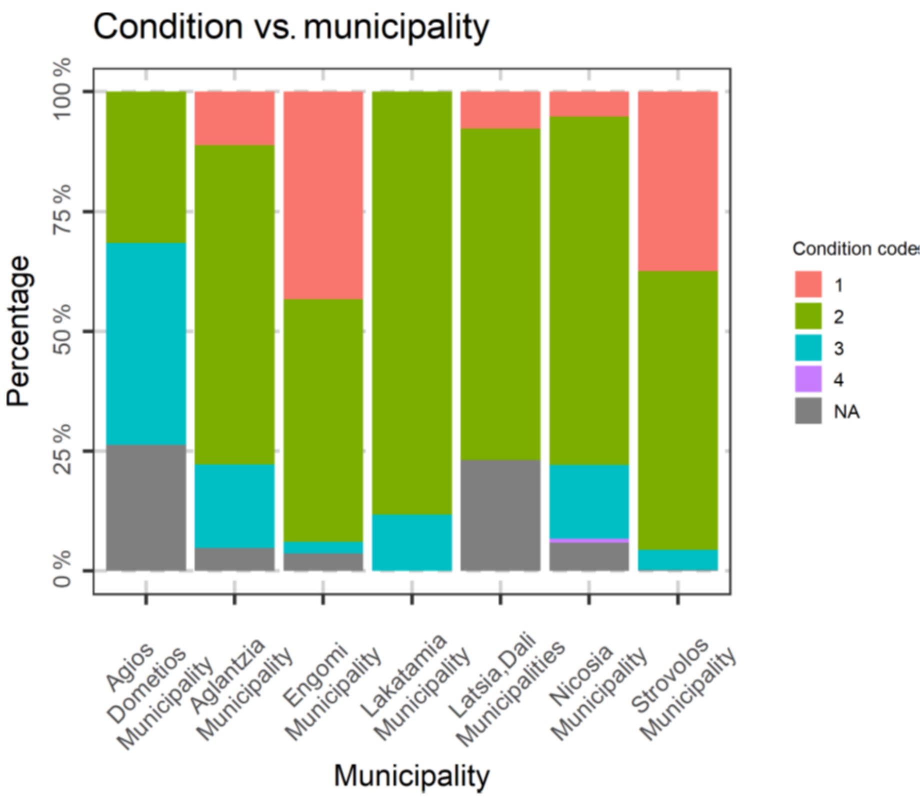


Figure 43: Percentage of condition code per municipality

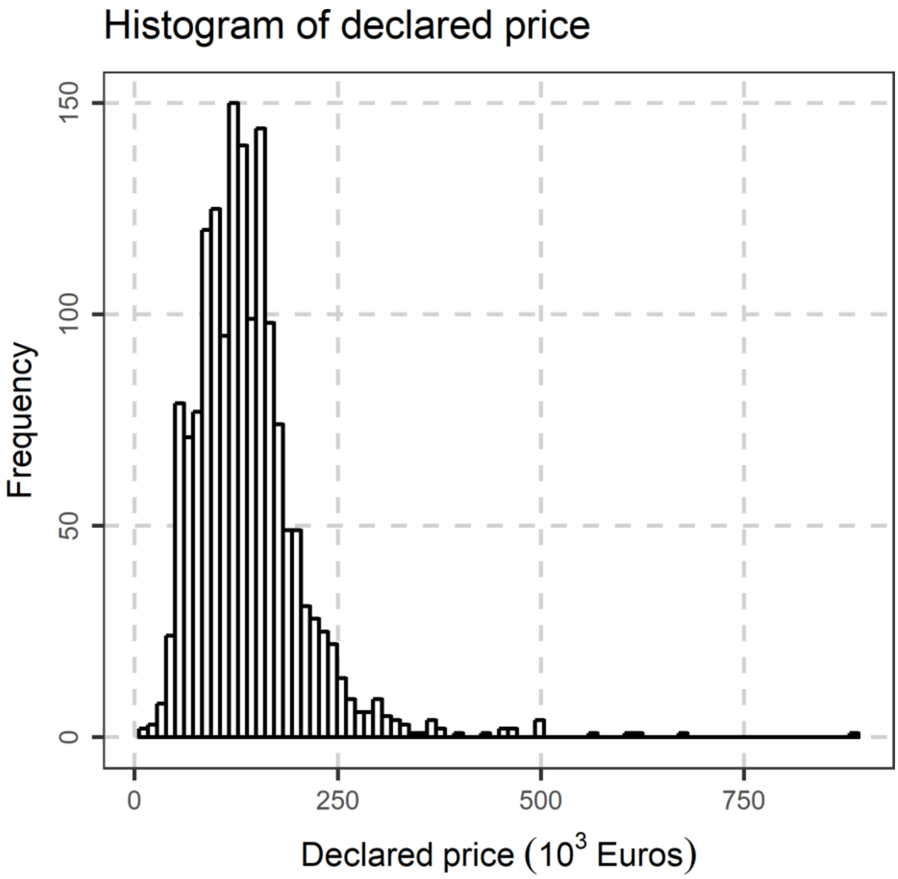


Figure 44: Histogram of Declared Price.

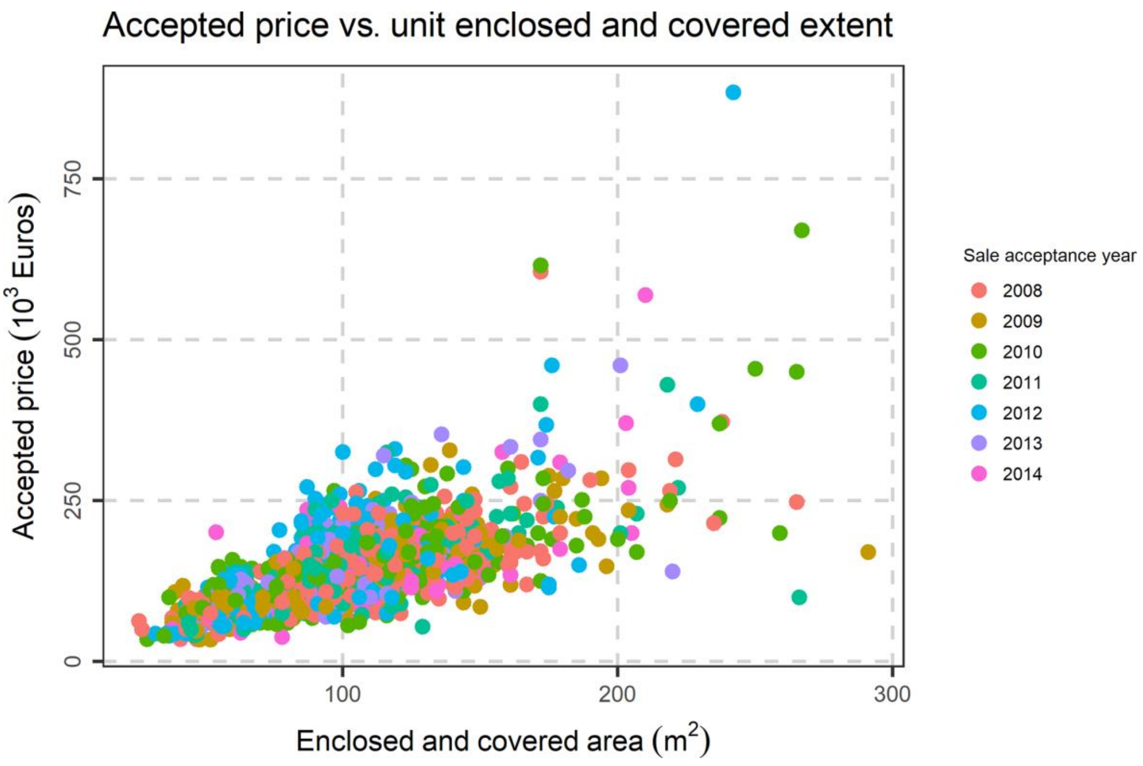


Figure 45: Accepted Price vs. Unit Enclosed and Covered Extend.

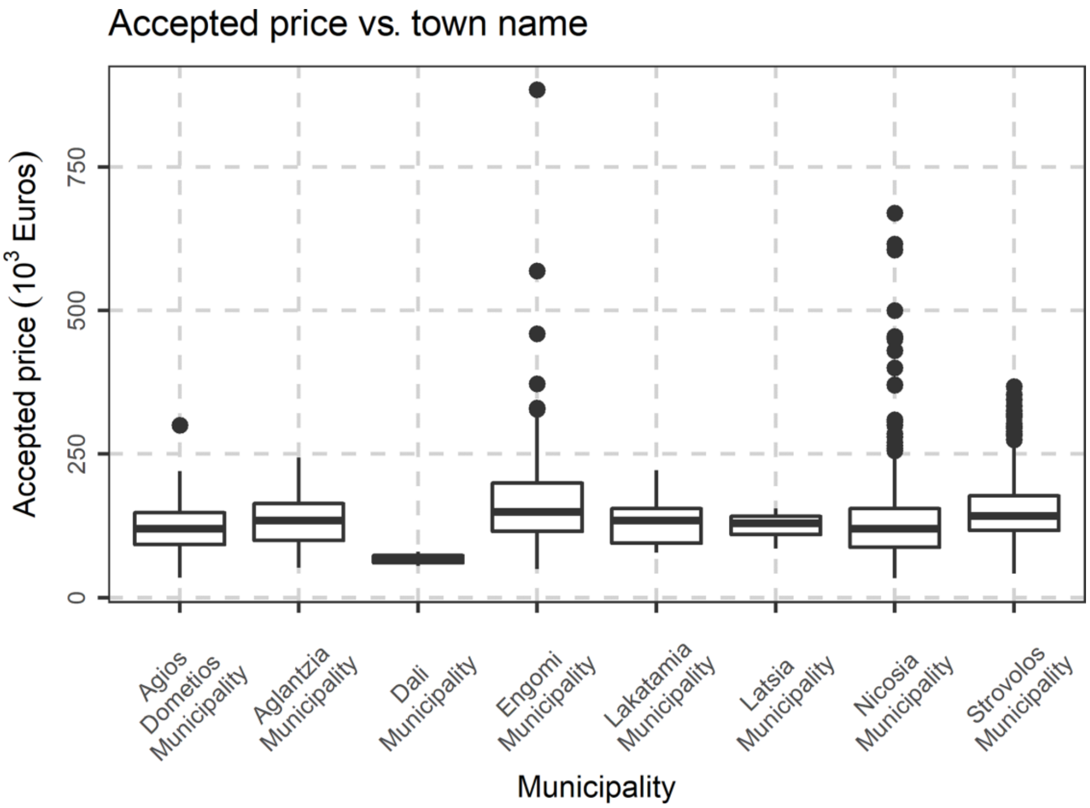


Figure 46: Accepted Price vs. Municipality

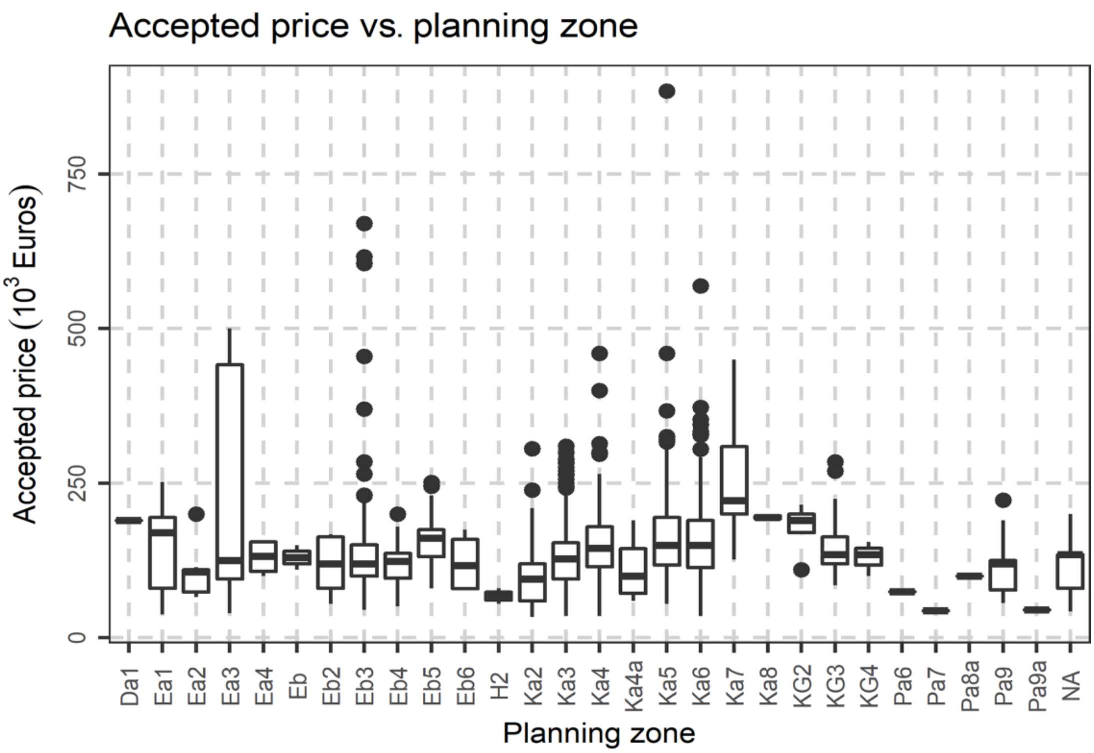


Figure 47: Accepted Price vs. Planning Zone

5.2.2 The Dependent Variable

The data referred to records of actual sale transactions. For a sale to be registered at the DLS, the two parties (buyer and seller) declare an amount of money to the DLS. However, the DLS performs a desktop appraisal, and in some cases the value accepted, might be different from the declared one (Declared vs. Accepted Price). At this point, it should be highlighted that buyer and seller have a special interest in declaring a lower amount of money than the amount agreed. The buyer benefits by paying lower transfer fees, and the seller benefits by paying lower capital gains tax. On the other hand, the DLS, serving the government's interests, aims to collect more taxes. Thus, the quality of data collected is debatable, and further analysis, as is discussed in this thesis, is necessary. This deviation among the Declared Price (DP) minus the Accepted Price (AP) varied within the range of zero (0.00) to eighty-three percent (+0.829) with a mean of approximately four percent (+0.0384). Hence, we investigated this difference further using the normalization of the deviation (hereafter NDEV) in terms of price per square meter (of the adjusted area) as a more significant quantity for the properties, rather than the deviation (DP-AP). The PCs were divided into two categories, depending on their input values: continuous and categorical. Accordingly, in Fig. A, the distributions of the NDEV are depicted within the eight municipalities (those that are regulated from the Local Town Plan of Nicosia) of Nicosia District studied (Strovolos, Nicosia, Latsia, Lakatamia, Egkomi, Geri, Aglantzia, Agios Dometios). For seven out of eight municipalities, the mean NDEV varied within the range of 42.66€/m2 to 68.21€/m2, while for the Municipality of Yeri, the mean NDEV was equal to 116.07€/m2. The reason for this was either because the DLS had an inaccurate understanding and impression of the property values in Yeri or because the buyers there had an additional motivation to under-declare the value of the property.

In order to confirm numerically the hypothesis that the DLS did not follow a specific method for the unacceptance of the declared price, a machine learning algorithm named Relieff (Robnik-Šikonja & Kononenko, 2003) was utilized. The aim was to detect dependencies among the PCs and DEV, while the Relieff was selected as a feature subset selection method, since it split the data to k-nearest neighborhoods, by means of their normalized n-dimensional Euclidean distance. This procedure provided an unsupervised classification of properties with similar characteristics. Accordingly, the importance of each particular variable to the response could be computed, depending on if a change of this variable causes or does not cause a change of the class. The dataset was split into 20,40,60,80 and 100 classes, in order to examine whether the results depended on the volume. Finally, the Relieff method was applied for all and for the non-zero values of the deviation as well. The calculated weights were lower than 0.025 for all the values (Figure 25) and lower than 0.04 (Figure 26) if only the non-zero values of the DEV were examined. As a value of zero indicates the inexistence of association and values close to the unit, the high importance of the predictors, the calculated values denoted essentially no association.

Table 21: Accuracy metrics and comparison

Metric	α	RMSE	MAE	MAPE	SR	Method
Mean	0.504	39 849.124	30 408.391	0.254	1.115	Linear Regression
St. Dev	0.294	8974.123	7968.162	0.085	0.070	Linear Regression
Min.	0.000	30 495.690	22 887.060	0.180	1.050	Linear Regression
Max.	0.790	55 198.710	44 376.310	0.410	1.240	Linear Regression
Mean	0.444	40 664.314	30 272.058	0.252	1.123	Random Forests
St. Dev	0.246	10 358.919	8982.956	0.091	0.062	Random Forests
Min.	0.000	27 603.670	19 723.790	0.150	1.060	Random Forests
Max.	0.710	62 465.260	47 559.540	0.430	1.250	Random Forests
Mean	9.731	−1.272	1.444	2.073	−0.734	Comparison (%diff)
St. Dev	15.431	6.303	6.430	7.807	1.685	Comparison (%diff)
Min.	−42.105	−23.533	−13.712	−10.526	−4.651	Comparison (%diff)
Max.	36.220	11.168	15.272	23.529	3.419	Comparison (%diff)

5.2.3 Models' description and diagnostics

Accordingly, a training and a testing set were created, with the training set to be utilized to fit the studied models and the testing set, to test the out-of-sample accuracy of the models. The training set corresponded to the 80% of the observations and the test set to the remaining 20% (randomly permuted), while the accuracy measures were calculated for the test set. The investigated regression formulas are described in Table A (Appendix II). The error metrics utilized in each model, either Linear regression or Random forest, with respect to the Predicted Values (PV), the Dependent Variable (DV), and the number of observations N, were: the linear coefficient a, the Root Mean Squared Error (R.M.S.E), the Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MAPE), and the Average Sales Ratio (SR).

$$\boldsymbol{PV = Alpha * DV + Beta} \qquad \textbf{[5.4]}$$

$$\mathbf{RMSE} = \sqrt{\frac{\sum(\mathbf{PV}-\mathbf{DV})^2}{\mathbf{N}}} \qquad \textbf{[5.5]}$$

$$\boldsymbol{MAE} = \frac{\sum|\boldsymbol{PV}-\boldsymbol{DV}|}{\boldsymbol{N}} \qquad \textbf{[5.6]}$$

$$\boldsymbol{MAPE} = \frac{\mathbf{1}}{\boldsymbol{N}}\sum\frac{|\boldsymbol{PV}-\boldsymbol{DV}|}{\boldsymbol{DV}} \qquad \textbf{[5.7]}$$

$$\boldsymbol{SR} = \frac{\mathbf{1}}{\boldsymbol{N}}\sum\frac{\boldsymbol{PV}}{\boldsymbol{DV}} \qquad \textbf{[5.8]}$$

The detailed results are demonstrated in Table B (Appendix II). For the comparison, the percentage difference (*%diff*) was utilized as:

$$\boldsymbol{diff} = \mathbf{100} * \frac{\boldsymbol{LR}-\boldsymbol{RF}}{\frac{\boldsymbol{LR}+\boldsymbol{RF}}{\mathbf{2}}} \qquad \textbf{[5.9]}$$

With *LR* indicating each Linear Regression Measure and *RF* the corresponding measure of the Random Forests method. The Accuracy metrics of the equations 5.4 to5.8 for each method, as well as their comparison (equation 7) (diff), are presented in Tables B, C and D (Appendix II), and accordingly, the Mean, the Standard Deviation, the Minimum and the Maximum for the models utilized in the investigation (Table A, Appendix II) are summarized in Table 21: Accuracy metrics and comparison. The means of the differences (Equation 5.6) were 9.73% for the Linear coefficient (Eq. 5.4), with the plus sign indicating positive values, and hence enhanced prediction of the Random Forests, -1.27% for the RMSE, 1.44% and 2.07% for the MAE and MAPE and -0.73% for the SR (Table 21). However, for the final five models (41-45), the differences for the RMSE were 7.87% to 11.17% lesser for the Random Forests Method than the equivalent of the regression (Table D, AppendixII), signifying an important prediction accuracy of the RF method, and in particular when adequate variables were incorporated in the model.

5.2.4 Conclusions

Through this work, a comparative study of the prediction performance of the Random Forests method was accomplished, with respect to the corresponding results of the Linear Multivariate Regression. A variety of regression models were scrutinized, apropos the independent variables involved, for both methods. The database regarded actual transactions of the Cyprus Department of Lands and Surveys. A significant part of this work involved the data preparation and cleaning, the identification of the significant database features, as well as the handling of the missing values and the skewed input distributions. Furthermore, the author made suggestions for the mass appraisal system in Nicosia, to examine the dependent variable (Accepted Price), its differences from the Declared Price (by the buyer and seller), and its dependency on the property's characteristics (independent variables). The overall performance of the Linear models and RF was in the range of approximately 53.000 € to 27.000 € for the RMSE, for an average value of properties price 150.000 €. The Random Forests outperformed the linear models with an RMSE difference up to 3.000 € (7.87% to 11.17%). The most important predictor variables were found to be the Enclosed Extend while other important variables were the Planning Zone Density and the Property's Class & Condition Code. However, although Random Forests exhibited low prediction errors and generalization ability (test set), an extensive analysis of the literature of some hundreds of papers, applying machine learning algorithms, demonstrated a lack of utilization of the real estate mass appraisals problem, highlighting further this work's contribution to the importance of machine learning and more specifically the Random Forests in mass appraisals.

5.3 Sensitivity Analysis of Machine Learning Models for the Mass Appraisal of Real Estate. Case study of residential units in Nicosia, Cyprus

5.3.1 State of the art

It appears that machine intelligence up until today can successfully replace humans only on the execution of specific tasks (Brynjolfsson et al., 2018; Pagano, 2018), which are often repetitive, dull, and time-consuming (Bryson, 2010). Typical examples of Machine Intelligence for the case of Real Estate Valuations could be the collection of comparable evidence, automated exclusion of incorrect registrations in a database (anomaly detection) (Agrawal & Agrawal, 2015), calculation of the uncertainty of prediction in each particular chosen region through the computation of the local outliers and calibration of the prediction according to spatial parameters (T. Dimopoulos & Moulas, 2016).

Contradictory, Artificial Intelligence cannot be considered to accurately understand specific property characteristics regarding the quality of construction, aesthetic characteristics, design, internal materials and appliances, the view to sea or nature, the deterioration of specific structural elements, local price peaks where comparable evidence is not available, property ownership issues (shares of ownership, rights of use, etc.) and tax, legal or governmental special cases, because such models are complex and incomprehensible (Chan & Abidoye, 2019).

Artificial intelligence and machine learning methods have been widely utilized in Real-Estate, and a variety of studies have been performed. In Arribas et al., (2016), the Hierarchical Linear Model is utilized in Mass appraisals of residential properties, to overcome the limitations of traditional econometric models such as Ordinary Least Squares. The absence of data of comparable properties is a major issue, while the consideration of micro- & macro-level characteristics of the properties should be considered (Ciuna et al., 2017). In Chica-Olmo et al., (2019), the spatial and temporal variation of properties is investigated, by a regression-kriging method. However, it appears that no study exists on the interpretation of the black-box machine learning models, regarding Real Estate Mass Appraisals. The purpose of this case study was to investigate how the complex machine learning models work, regarding Real Estate price predictions, and present the various models and the corresponding results. In Section 5.3.2, the analyzed dataset is explained, as well as its variables, followed by the Machine Learning Methods, utilized for the target task, as well as the generic algorithm to obtain the closed-form formula for the Higher-Order Regression Model, via an automated, step-wise method. In Section 3, the sensitivity analysis results of the predictors, regarding Real Estate prices is presented. In Section 5.3.3, the influence of the dataset volume is also investigated, by a parametric study, for a variety of partitions of the given dataset. In Section 5.3.4 highlights the obtained formulas, utilizing five (5), and ten (10), and in A.1, for one hundred (100) nonlinear terms. The equation is presented in Appendix III.

5.3.2 Comparable evidence and Methods

5.3.2.1 Database, pre-processing, methods and performance metrics

The studied database was obtained from the Department of Lands and Surveys. The data were used for the purposes of the Cyprus new General Valuation (1.1.2013) and referred to transactions between 2008 and 2014, out of which only transactions for apartments in Nicosia District were studied. Although it does not contain important socioeconomic variables (Lelo et al., 2018), it was considered as vastly useful amongst professional valuers, as it contained comparable evidence in certain property types. Hence, the level of information available for the valuer could be greatly enhanced; however, the reliable exploitation of the contained information remained vague. A significant effort was spent in order to prepare the database in a predictors-output format. At this point, the author highlights that the data would be significantly enhanced if remote sensing was integrated in order to enrich the database provided that was completed by on-site or drive-by observations.

In particular, 4261 observations of apartment/office sales in Nicosia existed, nevertheless from column Unit_desc, only values “APPARTMENT” & “2-FLOOR APARTMENT” were kept, resulting in 3786 remaining observations. Furthermore, only Municipalities that are regulated by the Nicosia Local Town Plan were selected, the Quarters with less than 20 observations were deleted, and finally, 3561 sales data were used for the analysis and predictions. In order to enhance the prediction accuracy of the models, Urban Planning data was added for each Planning Zone, and in particular, the maximum building density, the number of stories, height and coverage of the allowed building, the minimum sq.m. per resident and the expected sq.m. per resident. Due to multicollinearity among urban planning variables, only the maximum building density was finally kept. The transaction dates were converted to reflect the date 30/09/2018 as floating numbers constituting a continuous variable, and the prices were adjusted to 1-Jan-2013 utilizing the Central Bank of Cyprus Index.

The utilized variables were the following, with their abbreviations in parentheses, for each Unit (Apartment)

- Unit Enclosed extent, which is the Internal Area in m2 (IntArea).
- The Unit covered extent, which is the Area of covered verandahs in m2 (CovVer).
- The Unit uncovered extent, which is the Area of uncovered verandahs in m2 (UnCovVer).
- Parcel extent is the Area of the parcel (or plot) in m2 (ParcExt).

- The Built Years, calculated as the difference between the date the transaction happened and the date the building was constructed, in years (BuiltYrs).
- The Unit condition code (Cond), that denotes the condition of the building, and takes values from 1 (best condition) to 4 (worst condition).
- The Unit's view code (View), which denotes the view of the unit, with values from 1 (best view) to 4 (worst view).
- The Unit's class code (Class), denoting the class of the building. It takes Values from 1 (best class) to 4 (worst class).
- Density (Dens), as the maximum allowed density (built m_2, over plot's m_2) of the specific district.

The dependent variable was the apartment's price as Accepted by the Cyprus Department of Lands and Surveys. This price was adjusted by utilizing the Central Bank of Cyprus Index, and the dates were transferred to 30/09/2018. The abbreviation for the dependent variable is (Adj. Accepted Price).

5.3.2.2 Error Metrics

Machine learning methods exhibit diverse performance on a studied dataset, with respect to the error metrics each time utilized. Hence, four methods were investigated, Univariate (UR), Multivariate Linear (MLR), and Higher Order Regression (HOR), as well as Artificial Neural Networks (ANN). The Coefficient Of Dispersion (COD) was used as defined by Appraisal Ratio Studies (NCSS Statistical Software, n.d.), as a standard metric utilized in Real Estate Mass Appraisals. It is based on the Predicted Values (PV), the Dependent Variable (DV), and the number of observations N. COD is defined by

$$\boldsymbol{COD = 100 * \frac{\frac{1}{N}\sum(\left|\frac{PV}{DV}\right| - \frac{1}{N}\sum\frac{PV}{DV})}{\frac{1}{N}\sum\frac{PV}{DV}}} \qquad \textbf{[5.10]}$$

Furthermore, the utilized error metrics were the Root Mean Squared Error (R.M.S.E.), the Mean Absolute Error (M.A.E), and the Mean and Maximum Absolute Percentage Error (M.A.P.E) as well as the Pearson Correlation Coefficient which are defined as below:

$$RMSE = \sqrt{\frac{\sum(PV-DV)^2}{N}} \quad \textbf{[5.5]}$$

$$MAE = \frac{\sum|PV-DV|}{N} \quad \textbf{[5.6],}$$

$$MAPE = \frac{1}{N}\sum\frac{|PV-DV|}{DV} \quad \textbf{[5.7],}$$

$$\boldsymbol{MAXAPE = max}\left(\frac{|\boldsymbol{PV-DV}|}{\boldsymbol{DV}}\right) \quad \textbf{[5.8],}$$

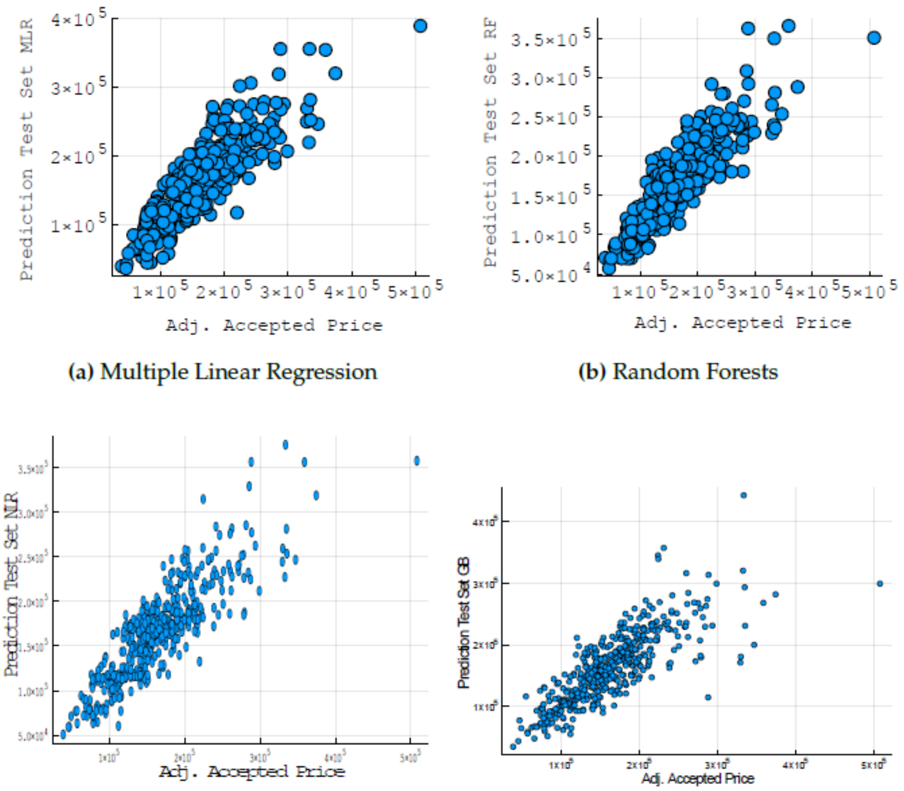
$$\rho_{X,Y} = \frac{cov(PV,DV)}{\sigma_{PV}\sigma_{DV}} \quad \textbf{[5.9],}$$

$$cov(X,Y) = E[PV - E(PV))(DV - E(DV))].$$

5.3.2.3 Anomaly Detection

Although the observations in the studied database regarded official registration in the DLS, some extremely unreasonable records occurred. For example, property in Nicosia Municipality, Ag. Andreas Quarter, built-in 1965, with 66 sq.m covered areas, and a price of 3.524e, Latsia / Ag. Georgios (1977), 68 sq.m, with a price of 17.781 e, Nicosia / Ag. Omologites (1982), 44 sq.m, 15.724 e, Nicosia /Ag. Antonios (1973), 35 sq.m, 22.562 e, and. Strovolos / Chryseleousa (1986), 76 sq.m, 17.283

Accordingly, an iterative procedure was implemented in order to identify the outliers and eliminate at each step the observation, which violated a specified threshold. The corresponding results were highly enhanced, as even for the Linear Regression (LR) (Figure 48) the R squared was increased from 0.611 to 0.744, while the shape of the scattered observations is closer to a straight line after the removal of the outliers.



(a) Multiple Linear Regression

(b) Random Forests

(c) Higher-Order Nonlinear Regression

(d) Gradient Boosting

Figure 48: Accepted Price vs Simulated for the test-set data.

The Algorithm 1 was selected, in order to exclude observations with high prediction errors, as they represented apartments which were under- or overpriced by the DLS, for some particular reason.

The algorithm developed by the author was selected amongst others because it presented better results in terms of percentage errors that are easier understood by property professionals.

Algorithm 1: Anomaly Detection

Data: Predictors($IntArea, CovVer, UnCovVer, ParcExt, BuiltYrs, Cond, View, Class, Dens$), and Responce ($Adj.AcceptedPrice$) for the entire Dataset.

Result: New, Decreased Dataset

Do Linear Regression
Compute $MAXAPE$
while $MAXAPE \geq 50\%$ **do**
  Linear Regression;
  Compute $MAXAPE$;
  Find index i of observation with $APE_i = MAXAPE$;
  Delete observation with index i;
end
return Decreased Dataset

5.3.2.1 *Analysis* of *Machine Learning Methods*

In order to evaluate more complex models, apart from Multiple Linear Regression (MLR), a Higher Order, Nonlinear Regression (NLR) was implemented. In particular, all combination of the variables were created, up to third-order

$$x_i * x_j * x_k,$$

with i, j, k ϵ [1, 9] for all the nine independent variables. Afterward, a forward step-wise algorithm was implemented, in order to sequentially add to the model, the combined variable with $x_i * x_j * x_k$, which corresponded to the model with the lowest APE. Algorithm 2 that was also developed for the purposes of the study, represented the applied procedure.

Algorithm 2: Step-wise, Higher Order Regression

Data: Predictors($IntArea, CovVer, UnCovVer, ParcExt, BuiltYrs, Cond, View, Class, Dens$), Responce ($Adj.AcceptedPrice$) for the Decreased Dataset, and desired number of features n_f

Result: NLR Model

Create Nonlinear Features $x_i * x_j * x_k$
Compute all APE with Uni-variate Regression
Store i, j, k combination corresponding to the minimum APE
for $i_1 = 2 : n_f$ **do**
    for $i_2 = 1$: *all non-Stored features* **do**
        Add i_2 to the Model
        Compute all APE with Multi-variate Regression
    end
    Store i, j, k combination corresponding to the i_2, with minimum APE
end
return all combinations of i, j, k for the NLR Model

Furthermore, Random Forests were used (RF) (Breiman, 2001) as implemented in Sadeghi, B. DecisionTree.jl. (2013) Available online: https://github.com/bensadeghi/DecisionTree.jl (accessed on 1 June 2019). and Gradient Boosting (GB) Xu, B.; Chen, T. XGBoost.jl. (2014) Available online: https://arxiv.org/abs/1603.02754 (accessed on 1 June 2019). All analyses were run on Julia programming language by utilizing the mentioned packages, as well as code written by the author, as described in Algorithms 1 and 2.

5.3.2.2 Regression Analysis

The regression results are presented in Table 22 for the four methods studied and the corresponding error metrics.

Table 22: Regression Results for the four methods studied, and error metrics.

Methods	ρ	MAE	RMSE	MAPE	MAXAPE	SR	α	COD
				Train Set				
Random Forests	0.914	17931.100	28854.237	0.111	1.307	1.031	0.739	10.778
Gradient Boosting	0.992	2630.784	8923.668	0.016	0.441	1.002	0.983	1.753
Linear Regression	0.863	24546.300	34745.422	0.151	0.550	1.027	0.746	14.703
Non-Linear Regression	0.880	23520.570	32700.793	0.146	1.100	1.032	0.775	14.197
				Test Set				
Random Forests	0.877	20817.165	27950.722	0.134	0.802	1.040	0.753	12.950
Gradient Boosting	0.803	24485.519	35946.437	0.151	1.092	1.009	0.776	15.017
Linear Regression	0.858	22977.825	30047.707	0.146	0.506	1.025	0.789	14.279
Non-Linear Regression	0.862	22525.779	29500.974	0.144	0.552	1.032	0.761	13.984

5.3.2.3 Sensitivity Analysis

A modified version of the Profile method (Gevrey et al., 2003; Olden & Jackson, 2002b) was utilized, in order to investigate the contribution of each independent variable to the dependent variable. In particular, each input variable varies within its given (raw) range, while all the other input variables were kept constant for a specific value. This constant takes three discrete values: 25% Percentile, Median, 75% Percentile. Through Sensitivity Analysis, the comparison of the black-box models could be illuminated, as we compared the effect of a predictor (i.e. Unite Enclosed Extent in Figure 49) to the studied variable (Adj. Accepted Price), indicating a decreasing pattern for IntArea higher than 180 m_2, which could not be identified with the Linear Model. Accordingly, in Figure 50, the Adj. Accepted Price was decreased with respect to the built years, however, the Machine Learning models concurred that this effect was weakening for built years more than 30. However, although all models exhibited similar patterns, different sensitivity curves were obtained for each model. This effect indicated the complexity of such models, which should be utilized critically or as ensembles (Sagi & Rokach, 2018).

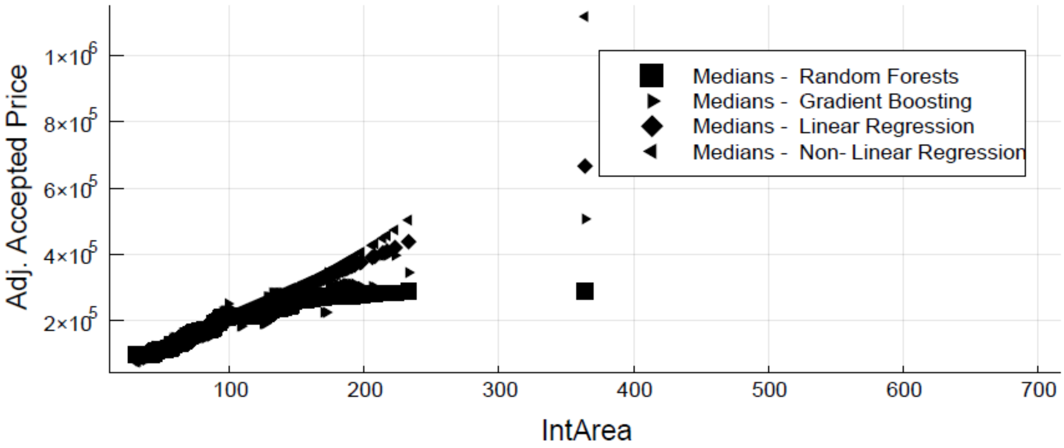


Figure 49: Sensitivity Analysis for Unit Enclosed Extend

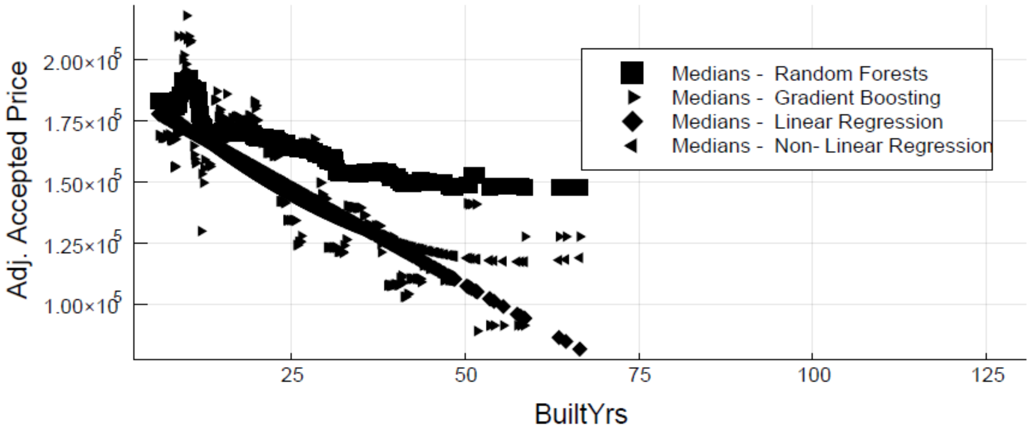


Figure 50: Sensitivity Analysis for Unit Built Years

In most applications of machine learning models to real-world problems, it is immensely challenging to identify the influence of each of the input variables on the final calculated outcome. For example, in the case of Real Estate Automated Valuations, it is useful to know, to what extent the covered area or the quality of construction of an apartment affects a specific property's value, as well as with what kind of pattern, linear or not. That is why the application of sensitivity analysis is a necessary tool when using the so-called *black-box,* AI models.

A prevalent topic in recent years is the interpretability of complex computational models. The creation of a model with excellent performance is not the only thing required; the explanation of certain aspects of the model is also necessary. The understanding of the model is crucial for validation purposes, in order to detect biases or outliers, and sometimes even for the identification of new patterns. The complication of such a model is due to the state-of-the-art performance that is required and this usually comes at the price of interpretability.

The interpretability can be further categorized into two main subcategories:

- The overall interpretability, which gives explanations about the behavior of the model over the entire set of observations,
- and the local interpretability, which gives an explanation regarding a specific prediction.

The understanding of a machine learning model is attained by conducting a type of sensitivity analysis to identify the impact of each feature on the model's prediction. To calculate a feature's sensitivity, its value needs to be altered somehow, while keeping all the other features constant. By examining the output produced by the model after this alteration, the impact of the selected feature can be concluded.

Feature sensitivity is an insightful technique that helps in the understanding of each feature's influence on the response (dependent variable), by utilizing the trained model. For example, Figure 51 illustrates the generic process of how an AI model predicts an output. Any AI model is actually a complex mechanism which, by utilizing algorithms and mathematical processes, yields a value. Now assuming that the test model is in the form of:

$$f(x, y) = x^3 + x^2 - xy + y^4,$$

where x,y are some normalized predictors and $f(x,y)$ *is* the response. Although this formula is simple, when it is compared with the *black-box* models, as it contains only two predictors, and specifically in polynomial form, its interpretation is almost impossible. For example, if the variable x grows, then the terms x^3 and x^2 grow as well, however, the variable xy is being subtracted from the response due to the negative sign, and it is not known which one will be the dominant term, as it depends on the value of y. Hence, a clear image of such an association of x versus $f(x,y)$, cannot be illustrated even in the case where the formula is given.

Figure 51: Black-box model

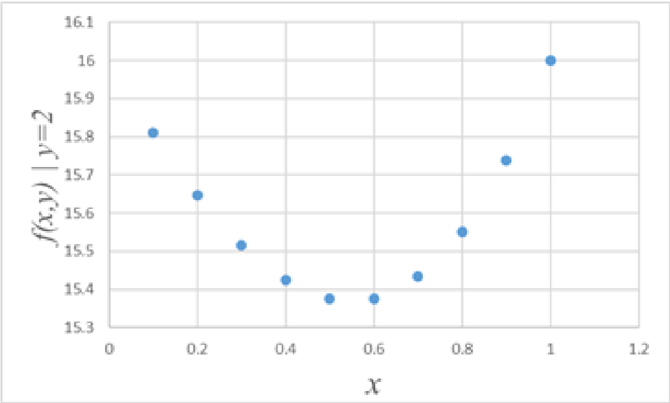


Figure 52 Sensitivity Analysis for the importance of *x* to *f(x,y)*

This is where the sensitivity analysis is introduced into the process. Particularly, if the variable *y is constrained* to its mean value (in this case, *y=2*), and vary the *x* variable within the values {*0.1, 0.2, ..., 1*}, then the image portrayed in Figure 52 is obtained. Interestingly, it now becomes clear that *f* attains its minimum value for intermediate values of *x* and not its extrema, lower of maximum. Furthermore, the visual representation of how the model behaves in terms of association of each one of the inputs is provided, with respect to its response. In this thesis, a modified version of the profile method, as presented in the work "*Review and comparison of methods to study the contribution of variables in artificial neural network models*" (Gevrey et al., 2003; Olden & Jackson, 2002a), obtained the sensitivity curves for all the predictors, with four different methods, in order to get robust conclusions for the sensitivity curves, and so forth the impact of the predictors on the response.

5.3.2.4 *How much data is Big enough?*

A common problem in simulation with machine learning methods is the amount of data. To investigate the importance of the data volume to the accuracy of prediction, we utilized random portions of the dataset, and each time we fitted a Random Forests model to the partition of the data. In Figure 53, we present the corresponding Mean Absolute Percentage Errors concerning the number of observations. Afterward, we fit a logarithmic curve:

$$y = a\log(x) + b, (1) \qquad [5.11]$$

To the obtained results, and extended the curve up to 5000 observations from the results it is seen that the number of data is an important factor influencing the prediction accuracy, with a clear decreasing pattern.

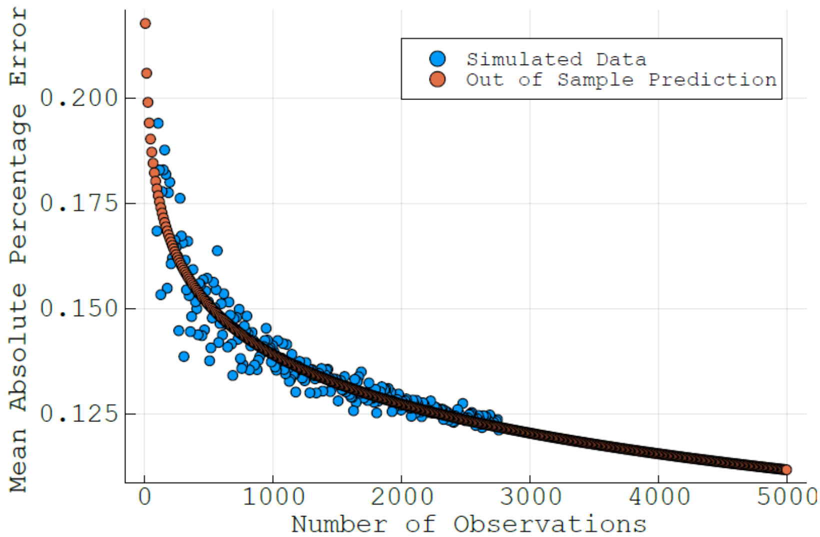


Figure 53: Number of data importance

5.3.2.5 *Prediction Formula*

With nonlinear Regression, we obtained the following Equations for the prediction:

A) With five terms (MAE=23993€)

Adj.AcceptedPrice = 2.13785E + 03 * IntArea

- 2.44629E + 01 * BuiltYrs * IntArea

$$
\begin{aligned}
&+ 4.67313E - 01 * BuiltYrs * CovVer * IntArea \\
&+ 3.52720E + 02 * UnCovVer \\
&+ 2.43798E + 02 * Dens * View * CovVer + 1.09116E + 04 \qquad [5.11]
\end{aligned}
$$

B) With ten terms (MAE=23748€, also used in sensitivity)

$$
\begin{aligned}
Adj.AcceptedPrice \quad &= 2.87808E + 03 \_ IntArea \\
&- 3.52523E + 01 * BuiltYrs * IntArea \\
&+ 1.35281E - 02 * BuiltYrs * CovVer * IntArea \\
&+ 3.30431E + 02 * UnCovVer + 5.16573E \\
&+ 02 * Dens * View * CovVer + 3.01148E \\
&- 01 * BuiltYrs * BuiltYrs * IntArea \\
&+ 2.32119E - 02 * IntArea * IntArea * IntArea \\
&- 6.87503E + 00 * IntArea * IntArea \\
&- 1.57789E + 00 * View * Cond * ParcExt \\
&+ 3.23269E - 04 * ParcExt * ParcExt * Dens - 7.24500E + 03 \qquad [5.12]
\end{aligned}
$$

5.3.2.6 Conclusions

Sensitivity analysis for features' importance to the dependent variable (Adj. accepted Price), demonstrated similar patterns, for all the four methods used. However, certain differences were also depicted, which highlighted the need for such analyses on the trained machine learning models. The accurate modelling of a studied system is challenging, and its predictive value is controversial (Oreskes et al., 1994) and (T Dimopoulos et al., 2018a), while the hopeful prospects that computers and refined models, will accomplish high prediction accuracy, were repeatedly defeated (Makridakis & Bakas, 2016). The utilization of a more accurate model instead of empirical rules exhibited enhanced prediction accuracy in property valuations. However, mathematical models without error estimation could jeopardize valuations hence author recommends that one obtains an initial estimation +/- a prediction error, as well as comprehensively investigating the errors' extrema and distributions. Machine learning algorithms can be used to validate professional valuations and not to replace human judgment, in order to avoid the impact of the highly improbable (Lybeck, 2017).

5.4 Discussion on the methods & results

The GWR approach was tested in two different datasets: Apartments in Nicosia and apartments in Thessaloniki. The main difference between the two cases was the source of data. In Nicosia, sales transactions were retrieved from the DLS, while in Thessaloniki, the data were not actual sales but valuations from practitioners from practitioners that were provided from the Central Bank of Greece. While in the first dataset, no significant geographical transformations were practiced, for Thessaloniki geocoding was deemed as necessary as the number of the street was removed for GDPR purposes. On the other hand, Thessaloniki had more location influential factors such as bus stops and other points of density also because the city is more densely populated, and the GWR model could be applied better. Both case studies indicated that GWR advantaged over the OLS. The GWR is an enhanced form of regression that batter takes into account the location influence on properties value. GWR is considered a transparent method and is recommended from the traditional researchers.

On the other hand, alternative and more sophisticated algorithms from Artificial Intelligence and Machine Learning are preferred by a vast majority of researchers. Therefore, the author examined the application of Random Forests as the third case study. It is an algorithm that comes from the machine learning scientific area. It constructs decision trees and calculates the mean prediction (regression) of the individual trees. It is based on the idea of bagging and the random selection of features to construct the collection of random decision trees, but with controlled variance. The method was tested in Nicosia's database, proved significant robustness, and is highly recommended for similar datasets.

The main issues of such models are the transparency and the explainability of the outcome. The author, having recognized the fact that these complicated algorithms are widely considered as a "black-box", continued a step forward and proposes sensitivity analysis, in order to highlight how complex models, work, as well as the underlying relationships among the involved predictors and the response (predicted property value). It is the first time in the literature that the sensitivity analysis has been applied in a real estate dataset.

Machine Intelligence imitates human perception, by utilizing mathematical models that compete against humans to deliver certain tasks such as the assessment and analysis of a studied system and predictions of out-of-sample observations. The accomplished tasks can be highly complex, based on mathematical models which simulate a physical, social, financial and so forth, system of study (Makridakis & Bakas, 2016). Machine learning algorithms belong to the wider thematic area of Artificial Intelligence, with applications in Healthcare (Jiang et al., 2017), Automotive (self-driving cars) (Pozna & Antonya, 2016), Finance and Economics (predictions, assets management) (Gupta et al., 2018), Military (drones capable of autonomous action) (de Swarte et al., 2019), Advertising (predict/quantify the behavior of customers) (Olson & Levy, 2018), Image Recognition (Vyas et al., 2018), and so forth. The idea that machines could exhibit intelligence is not a new concept, rather it stems from ancient times (Cave & Dihal, 2018), for example, the robot Talos made by Hephaestus in Ancient Greek Mythology (Lleo, 2019). At this point, it should be noted that remote sensing, aerial or oblique photos can be used in order to obtain this information automatically (this thesis though, only focuses on the mathematical modelling of price prediction. It should also be noted, however, that professional valuers hesitate to utilize such algorithms (Dimopoulos et al., 2018), and certain questions arise concerning the application of mathematical models for the improvement of the valuer's work. Questions such as, whether Machine Intelligence can replace Human Intelligence, if the mathematical models can replace the judgment of the individual valuer, and who would sign the outcome of an Automated Valuation Model are commonly debated topics. Yann LeCun quoted: "*Our intelligence is what makes us human, Artificial intelligence is an extension of that quality. Many discussions have been had over recent years about whether there shall be a limit to restrict artificial intelligence and which level of artificial intelligence is optimal*". AI can perform tasks that a human is unable to perform either at the same pace, quality, at the same cost or at all. The question arises whether artificial intelligence can replace the human valuer, taking into consideration computer-assisted mass appraisal and automated valuation models. It needs to be stated that, within the environment of appraisals, CAMA and AVM have been used since as early as the 1950s (Worzala, E.; Lenk, 1995), and were further developed in the 1960s.

6 Conclusions, discussion and future work

6.1 Conclusions

Mass appraisals have already been used for taxation purposes, and the initial part of this thesis regards the analysis of the corresponding best practices. Nevertheless, the demand for mass appraisals is growing, as a result of the financial globalization process, which promotes capitals to move investing in property markets around the world and, accordingly, the urgency to know similar rules, risks and guarantees associated to the investment process.

Cauko and D'Amatao (2018) state that normalization and harmonization of appraisal methods are increasing issues for property markets within developed countries. In some areas like the European Union, appraisal methods have tended to converge in recent years as a result of the coordination of European regulations, following the need to establish such policy practices, in order to determine the 'single and reliable' value for any type of property. This will guarantee the value of the collateral to an investor with the appropriate transparency.

The author used traditional approaches of mass appraisals, such as linear and nonlinear regression on a test database, where the results exhibited poor performance, when compared with the results of the GWR. The GWR approach was used for a database of residential apartments in Nicosia, Cyprus, and another database for residential flats in Thessaloniki, Greece. It should be noted that the author lived and professionally practiced in both areas, and the computational results were critically analyzed by a real-world point of view. In both cases, the results were robust and transparent. Both cases proved that the geographically weighted regression is more accurate than MRA, and highlights that if the location influence - which the most dominant attribute of a property's value - is captured by a statistical method, the accuracy increases. Regression-based methods are more transparent compared with the AI applications, as the mathematical model regards an explicit equation.

Random forests are a machine learning algorithm from data science that are based on random decision trees. A large number of relatively uncorrelated models (trees) operating as a committee outperformed any of the individual constituent models. The low correlation between models is the key to obtaining a model that generalizes the results. It was the second time, noted in literature, that this algorithm was successfully implemented for mass appraising properties. Even though such approaches appear to provide higher accuracy, they lack transparency, and the results cannot be easily explained by humans.

The author highlighted the strengths and limitations of Machine Learning and AI in mass appraisal modelling. The number and the quality of the data provided was a major implication for the case of Cyprus. However, it appears that AI models offer higher accuracy than the MRA for the training as well as, out-of-sample, test sets. This is in accordance with the fact that properties are heterogenous commodities and can be appraised not only by applying a predefined mathematical law, but with the talent (heuristics) of the valuer that is called the “art”, and is captured better by AI models. The author goes a step forward and introduces sensitivity analysis in the application of AI in order to highlight the importance of the various predictors (covered extent, location, etc.) to the estimated property value.

6.2 Discussion

In the decision making process to whether a Mass Appraisal System shall be used, the outermost important factors that the author recommends to be examined are:

- Time
- Money
- Quality
- Accuracy
- Bureaucracy
- Responsibility
- Regulations
- Licenses
- Initial cost
- Neutrality
- Available data.

Every single property valuation is a unique project and has a clear start and end date. Manual valuations are usually both time and money resource intensive and often deliver results in crucial revaluations later or sometimes never (Bird et al., 2013). In a project, there is always a trade-off between Time, Money and Quality. Increasing one of the factors almost automatically decreases the remaining two. For example, a valuer who tries to complete more valuations within a given period, either must decrease the quality of each valuation to be faster per valuation or must hire more staff to deliver more valuations. AI does not have any of these constraints. It can work 24/7 and with the correct data and can produce a theoretically infinite amount of valuations. Practically, the amount is limited to the available data as well as the input of this data by a human source. Data has been mentioned as an important component. CAMA and AVM can only exhibit high computational efficiency if the database contains adequate data. Theoretically, one could state that if no data is available, AI cannot be used. On the other hand, without precise data, any human-based valuation would not be very precise either. It takes years of studying and obtaining practical experience as well as local market knowledge for a valuer to be able to deliver accurate valuations and appraisals. This process of learning is time-consuming and rather expensive. AI can do so within a short period of time and can improve its performance based on past observations. Due to this, human valuers are considered to be expensive. AI can offer a much less expensive rate for any valuation since costs such as travel time and travel expenses to the property can be saved. However, AI has a higher initial cost, as it is expensive to set up a model. The maintenance of the database and feeding the AI model with more data are usually the highest running expenses. Any invention that may replace workers with machines in a particular field can have a positive effect on society by "reducing the price of goods, increasing real income" (Autor, 2015). Research conducted in this context suggests that the methods, currently used extensively, have inherent errors regarding how they derive their value estimates (Ismail & Buyong, 1998).

Many scientists stated that feelings and sympathy are what make us human. These are unarguably great assets of every human; however, in valuations, they can create inaccuracies due to the loss of neutrality. Humans can only control their doings up to a certain level. AI does not lose neutrality and hence accuracy, due to sympathy, therefore, in this aspect it can create more accurate valuations. Carrying out an official valuation requires, in almost every country, a license. These licenses are often provided by human-based associations. Often political reasons block any technological process as some humans fear losing their job to AI. This political lobbying reduces progress considerably and by doing so the human valuer is heavily favored. Human valuers often argue about the responsibility and legal pursuit of AI. A valuation carried out by a human valuer can always be challenged and one can sue the person who completed the valuation. But, the questions to be answered are who do you sue when an automated Mass Appraisal valuation is in question, and who signs an automated Mass Appraisal valuation. The above two questions can unfortunately not be answered easily. Looking for the responsible party of a Mass Appraisal valuation is a tricky process, which is one of the major drawbacks of AI. However, if we feed the AI model with enough data and constantly maintain and update the database, the possible margin of error shall be small enough to be negligible, and costly legal processes could be avoided or minimized.

As well as the above, we must understand in which situations we value properties and if all valuations need to be legally appropriate in terms of responsibility and suitability. Nowadays, countless valuations are done daily; mostly valuations for courts or banks giving out mortgages or attempting to repossess distressed/mortgaged assets, but there are also so many more valuations conducted for many other reasons.

All the explanations described in the above paragraph could be ideal situations for the use of AI, in order to provide cheaper and faster valuations. Having this kind of valuation completed by AI models would, of course, reduce the total number of valuations completed by human valuers. However, it has to be stated that the effect of artificial intelligence on the level of human employment will be dramatically reduced (Frey & Osborne, 2017). This, however, does not necessarily mean that any human valuer should lose their job. It could mean the opposite. Human valuers could focus more on each valuation, and automatically increase the quality of every valuation completed by a human valuer. Special reference must be made to complex valuations where a valuer needs a lot of time to fully understand and adjust the influencing factors. By giving human valuers more time to focus on these complex valuations and valuations for bank lending or repossessing purposes, the quality increases significantly. The improvement of quality will automatically lead to a higher achieved price per valuation which could, in the end, create higher profits for any valuer, stakeholder and for the wider financial and economic environment.

6.3 Future work

The author plans to continue his research in the area. His immediate plans are synopsized in the following five axes:

1. Continue his role as an active member of the academic community. He teaches at the only dedicated Real Estate programmes in Cyprus – both at Neapolis University (the BSc in Real Estate Valuation and Development and the MSc in Real Estate – accredited from RICS). Since 2018, he has been elected as Lecturer. He has also been teaching the relevant Valuation class at the Dept. of Civil Engineers and Geomatics of Cyprus University of Technology where he is appointed as an Expert Scientist since 2012, and has conducted many seminars.
2. Research is challenging and demanding area. He plans to enrich his knowledge in data science and participate or lead a research programme with stakeholders such as the Central Bank of Cyprus, the Dept. of Lands and Surveys and a Research Centre that will contribute to the development of an AVM in Cyprus. He also plans to progress the models he has built and apply them to real-life situations, where data is available
3. Keep carrying out real-life valuations. In 2012 he founded a company that services all the major stakeholders and financial institutions in Cyprus. The firm currently employs eight people. Five of them have a degree in surveying engineering and a Masters in real estate.
4. He is also co-founder and director of another company that was established in 2018 and developed a GIS- based software that is currently used by two major asset management companies in Cyprus. He plans to implement the algorithms he introduced in this thesis to this software package and offer a Mass Appraisal solution to the market.
5. Keep contributing serving the profession. The author has been the Chairman of RICS in Cyprus since 2017. On January 2020 he was appointed as a member of the European Board of the IVSC (International Valuation Standards Council). He is also an elected board member of the local Association of Property Valuers since 2016. He participated in several working groups that produced Guidance Notes (Bank lending valuations and mortgage lending value, RICS 2019). He also co-authored the 2nd edition of the “Application of the RICS Valuation – Professional

Standards in Cyprus” that is expected to be published soon. He is also participating in several international thinking tanks including RICS’ Virtual Valuation Network. Last but not in least, he is member of RICS’ working group for developing standards for AVMs.

REFERENCES

Abidoye, R. B., & C, C. A. P. (2016). Research Trend of the Application of Artificial Neural Network in Property Valuation. *33rd CIB W78 Conference 2016, November*, 1–9.

Agrawal, S., & Agrawal, J. (2015). Survey on anomaly detection using data mining techniques. *Procedia Computer Science*, *60*(1), 708–713. https://doi.org/10.1016/j.procs.2015.08.220

Almy, R. (2001). A Survey of Property Tax System in Europe. In *Almy, Gloudemans, Jacobs & Denne: Property Taxation and Assessment Consultants* (Issue March). http://www.agjd.com/EuropeanPropertyTaxSystems.pdf

Almy, R. (2013). *A Global Compendium and Meta-Analysis of Property Tax Systems*.

Amato, M. (2017). *Advances in Automated Valuation Modeling* (Vol. 86). https://doi.org/10.1007/978-3-319-49746-4

Anselin, L. (1998). GIS Research Infrastructure for Spatial Analysis of Real Estate Markets. *Journal of Housing Research*, *9*(1), 113–133. https://doi.org/10.5555/jhor.9.1.e523670p713076p1

Antipov, E. A., & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics. *Expert Systems with Applications*, *39*(2), 1772–1778. https://doi.org/10.1016/j.eswa.2011.08.077

Antoniou, D., & Dimopoulos, T. (2018). Southern Orientation and Natural Cross Ventilation: Mind the Gap (s) What Clients, Valuers, Realtors and Architects Believe. *RELAND: International Journal of Real Estate & Land Planning*, *1*, 426–439.

Arribas, I., García, F., Guijarro, F., Oliver, J., & Tamošiūnienė, R. (2016). Mass appraisal of residential real estate using multilevel modelling. *International Journal of Strategic Property Management*, *20*(1), 77–87. https://doi.org/10.3846/1648715X.2015.1134702

Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of

Workplace Automation. *The Journal of Economic Perspective*, *29*(3), 3–30. https://doi.org/10.1057/9780230226203.0037

Bahl, R. (2009). Property tax reform in developing and transition countries. *Fiscal Reform and Economic Governance Project, United States Agency for International Development*.

Basu, S., & Thibodeau, T. G. (1998). Analysis of Spatial Autocorrelation in House Prices. *Journal of Real Estate Finance and Economics*, *17*(1), 61–85. https://doi.org/10.1023/A:1007703229507

Baum, A., Baum, C. M., Nunnington, N., & Mackmin, D. (2017). *The income approach to property valuation* (Seventh, pp. 345–350). Routledge.

Bidanset, P. E., & Lombard, J. R. (2014a). Evaluating Spatial Model Accuracy in Mass Real Estate Appraisal A Comparison of Geographically Weighted Regression and the Spatial Lag Model. : *: A Journal of Policy Development and Research*, *16*(3), 169–182. https://www.jstor.org/stable/pdf/26326913.pdf

Bidanset, P. E., & Lombard, J. R. (2014b). The effect of kernel and bandwidth specification in geographically weighted regression models on the accuracy and uniformity of mass real estate appraisal. *Journal of Property Tax Assessment & Administration*, *11*(3), 5–14.

Bird, R., Slack, E., & Guevara, M. (2013). Real Property Taxation in the Philippines. *International Handbook of Land and Property Taxation; Edward Elgar Publishing: Cheltenham, UK*.

Bitter, C., Mulligan, G. F., & Dall'erba, S. (2007). Incorporating spatial variation in housing attribute prices : a comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems*, *9*(1), 7–27. https://doi.org/10.1007/s10109-006-0028-7

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*(11), 2215–2222.

Borst, R. A. (1991). Artificial neural networks: The next modelling/calibration technology for the assessment community. *Property Tax Journal*, *10*(1), 69–94.

Bourassa, S. C., Hoesli, M., & Peng, V. S. (2003). Do housing submarkets really matter ? *Journal of Housing Economics*, *12*(1), 12–28. https://doi.org/10.1016/S1051-1377(03)00003-2

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Brunori, D. (2007). *Local tax policy: A federalist perspective*. The Urban Insitute.

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What Can Machines Learn, and What Does It Mean for Occupations and the Economy? *AEA Papers and Proceedings*, *108*, 43–47. https://doi.org/10.1257/pandp.20181019

Bryson, J. J. (2010). Robots should be slaves. In Y. Wilks (Ed.), *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues* (pp. 63–74). https://doi.org/10.1075/nlp.8.11bry

Cave, S., & Dihal, K. (2018). Ancient dreams of intelligent machines: 3,000 years of robots. In *Nature*. https://doi.org/10.1038/d41586-018-05773-y

Chan, A. P. C., & Abidoye, R. B. (2019). Advanced Property Valuation Techniques and Valuation Accuracy: Deciphering the Artificial Neural Network Technique. *International Journal of Real Estate and Land Planning*, *2*, 2623–4807.

Charalambous, M., & Hanlon, N. O. (2015). *The Study on Refining the Parameters of the CAMA Model*. *January*. https://portal.dls.moi.gov.cy/el-gr/thedepartment/Documents/GV2013- Final study on Refining Parameters of the CAMA model.pdf

Chica-Olmo, J., Cano-Guervos, R., & Chica-Rivas, M. (2019). Estimation of housing price variations using spatio-temporal data. *Sustainability*, *11*(6), 1551.

Ciuna, M., Milazzo, L., & Salvo, F. (2017). A mass appraisal model based on market segment parameters. *Buildings*, *7*(2), 34.

Clow Frederick R. (1896). The St. Paul Method of Assessing Real Estate. *Journal of Political Economy*, *5*(1), 71–80.

Criterion, A. (1977). Equivalence of Choice of Model by Cross-validation An Asymptotic Akaike ' s Criterion. *Journal of the Royal Statistical Society*, *39*(1), 44–47.

de Swarte, T., Boufous, O., & Escalle, P. (2019). Artificial intelligence, ethics and human values: the cases of military drones and companion robots. *Artificial Life and Robotics*. https://doi.org/10.1007/s10015-019-00525-1

Demetriou, D. (2017). A spatially based artificial neural network mass valuation model for land consolidation. *Environment and Planning B: Urban Analytics and City Science*, *44*(5), 864–883. https://doi.org/10.1177/0265813516652115

Demetriou, D. (2018). Automating the land valuation process carried out in land consolidation schemes. *Land Use Policy*, *75*, 21–32. https://doi.org/10.1016/j.landusepol.2018.02.049

Demetriou, D. (2015). A GIS-based hedonic price model for agricultural land. In T. K. P. G. Michaelides S. Hadjimitsis D.G. (Ed.), *Proceedings of SPIE - The International Society for Optical Engineering* (Vol. 9535). SPIE. https://doi.org/10.1117/12.2192498

Des Rosiers, F., & Thériault, M. (2006). Mass Appraisal, Hedonic Price Modelling and Urban Externalities: Understanding Property Value Shaping Processes. *Advances in Mass Appraisal Methods Seminar*.

Dillinger, W. R. (1992). *Urban property tax reform: Guidelines and recommendations* (Vol. 1). World Bank Publications.

Dimopoulos, T., Labropoulos, T., & Hadjimitsis, D. G. (2014). Comparative analysis of property taxation policies within Greece and Cyprus evaluating the use of GIS, CAMA, and remote sensing techniques. *Proceedings of SPIE - The International Society for Optical Engineering*, *9229*. https://doi.org/10.1117/12.2070457

Dimopoulos, T., Labropoulos, T., & Hadjimitsis, D. G. (2015). Introducing local property tax for fiscal decentralization and local authority autonomy. *Proceedings of SPIE - The International Society for Optical Engineering*, *9535*. https://doi.org/10.1117/12.2193545

Dimopoulos, T., & Moulas, A. (2016). A Proposal of a Mass Appraisal System in Greece with CAMA System: Evaluating GWR and MRA techniques in Thessaloniki Municipality. *Open Geosciences*, *8*(1), 675–693. https://doi.org/10.1515/geo-2016-0064

Dimopoulos, T, & Bakas, N. (2019). Artificial intelligence for mass appraisals of residential properties in Nicosia: Mathematical modelling and algorithmic implementation. In M. S. A. V. H. D. G. Themistocleous K. Papadavid G. (Ed.), *Proceedings of SPIE - The International Society for Optical Engineering* (Vol. 11174). SPIE. https://doi.org/10.1117/12.2538430

Dimopoulos, T, & Moulas, A. (2016). A Proposal of a Mass Appraisal System in Greece with CAMA System: Evaluating GWR and MRA techniques in Thessaloniki Municipality. *Open Geosciences*, *8*(1), 675–693. https://doi.org/10.1515/geo-2016-0064

Dimopoulos, T, Tyralis, H., Bakas, N. P., & Hadjimitsis, D. (2018a). Accuracy measurement of Random Forests and Linear Regression for mass appraisal models that estimate the prices of residential apartments in Nicosia, Cyprus. *Advances in Geosciences*, *45*, 377–382. https://doi.org/10.5194/adgeo-45-377-2018

Dimopoulos, T, Tyralis, H., Bakas, N. P., & Hadjimitsis, D. (2018b). Accuracy measurement of Random Forests and Linear Regression for mass appraisal models that estimate the prices of residential apartments in Nicosia, Cyprus. *Advances in Geosciences*, *45*, 377–382. https://doi.org/10.5194/adgeo-45-377-2018

Dimopoulos, Thomas. (2015). Theories and philosophy of property taxation. *Pandemos.Panteion*, *79*. Library.panteion.gr/pandemos

Dimopoulos, Thomas, & Bakas, N. (2019a). An artificial intelligence algorithm analyzing 30 years of research in mass appraisals. *RELAND: International Journal of Real Estate & Land Planning*, *2*, 10–27.

Dimopoulos, Thomas, & Bakas, N. (2019b). Sensitivity analysis of machine learning models for the mass appraisal of real estate. Case study of residential units in Nicosia, Cyprus. *Remote Sensing*, *11*(24). https://doi.org/10.3390/rs11243047

Dimopoulos, Thomas, & Bakas, N. P. (2016). Introducing Neural Network Techniques on property mass appraisal. Case study of 3,500 apartments comparable evidence in Nicosia. *Fourth International Conference on Remote Sensing and Geoinformation*.

Dimopoulos, Thomas, Bakas, N. P., & Hadjimitsis, D. G. (2017). Assessment of the comparable evidence from RE transactions in Cyprus; Numerical investigation of the deviation between Declared and Accepted price. *Fifth International Conference on Remote Sensing and Geoinformation*.

Dimopoulos, Thomas, Labropoulos, T., & Hadjimitsis, D. G. (2015). Introducing local property tax for fiscal decentralization and local authority autonomy. *Third International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2015)*, *9535*, 953509.

Dimopoulos, Thomas, Tyralis, H., Bakas, N. P., & Hadjimitsis, D. (2018). Accuracy measurement of Random Forests and Linear Regression for mass appraisal models that estimate the prices of residential apartments in Nicosia, Cyprus. *Advances in Geosciences*, *45*, 377–382. https://doi.org/10.5194/adgeo-45-377-2018

Dimopoulos, Thomas, & Varnavas, P. (2015). The General Valuation 2013 in Cyprus: Experiences and Good Practices. *FIG-OICRF*.

Dimopoulos, Thomas, & Yiorkas, C. (2017). Implementing GIS in real estate price prediction and mass valuation: the case study of Nicosia District. In G. Papadavid, D. G. Hadjimitsis, S. Michaelides, V. Ambrosia, K. Themistocleous, & G. Schreier (Eds.), *Fifth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2017)* (Issue September 2017, p. 72). SPIE. https://doi.org/10.1117/12.2280255

Do, A. Q., & Grudnitski, G. (1992). A neural network approach to residential property appraisal. *The Real Estate Appraiser*, *58*(3), 38–45.

Eckert, J. K., Gloudemans, R. J., & Almy, R. R. (Eds.). (1990). *Property Appraisal and Assessment Administration*. International Association of Assessing Officers.

Ellis, R. D., Fischer, N., & Sauer, J. B. (2006). *Foundations of civic engagement: Rethinking social and political philosophy*.

Esri. (n.d.). *Geographically Weighted Regression (GWR)*. Retrieved August 15, 2016, from https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/geographically-weighted-regression.htm

European AVM Alliance. (n.d.). *Glossary of Terms & Definitions*.

européenne. Commission européenne, U., & EUROSTAT. (2013). *Taxation Trends in* *the European Union: Data for the EU Member States, Iceland and Norway*. Office for official publications of the European Communities.

Feinschreiber, R., & Kent, M. (2009). *Multistate Taxation*.

Fischel, W. A. (2013). *Fiscal Zoning and Economists' Views of the Property Tax*.

Fotheringham, A. S., Charlton, M., & Brunsdon, C. (1998). Geographically weighted regression : a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, *30*, 1905–1927.

Fotheringham, A. S., & Crespo, R. (2015). Exploring , modelling and predicting spatiotemporal variations in house prices. *The Annals of Regional Science*, *54*(2), 417–436. https://doi.org/10.1007/s00168-015-0660-6

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Gao, X., Asami, Y., & Chung, C. F. (2006). *An empirical evaluation of spatial regression models*. *32*, 1040–1051. https://doi.org/10.1016/j.cageo.2006.02.010

Geoghegan, J., Wainger, L. A., & Bockstael, N. E. (1997). Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. *Ecological Economics*, *23*, 251–264.

Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, *160*(3), 249–264. https://doi.org/10.1016/S0304-3800(02)00257-0

Gilbertson, B., & Preston, D. (2011). *Practice briefing A vision for valuation*.

Goodman, A. C. (1998). *Housing Market Segmentation*. *143*, 121–143.

Görög, G. (1994). An excel program for calculating and plotting receiver-operator characteristic (ROC) curves, histograms and descriptive statistics. *Computers in Biology and Medicine*, *24*(3), 167–169.

Grover, R., Törhönen, M.-P., Munro-Faure, P., & Anand, A. (2017). Achieving successful implementation of value-based property tax reforms in emerging European economies. *Journal of European Real Estate Research*.

Gupta, S., Sharma, A., & Abubakar, A. (2018). Artificial Intelligence–Driven Asset Optimizer. In *SPE Annual Technical Conference and Exhibition* (p. 7). Society of Petroleum Engineers. https://doi.org/10.2118/191551-MS

Hoen, B., Brown, J. P., Jackson, T., Thayer, M. A., Wiser, R., & Cappers, P. (2014). Spatial Hedonic Analysis of the Effects of US Wind Energy Facilities on Surrounding Property Values. *Journal of Real Estate Financial Economics*, *51*(1), 22–51. https://doi.org/10.1007/s11146-014-9477-9

Huang, B., Wu, B., Barry, M., & Security, L. T. (2010). *Geographically and Temporally Weighted Regression for Modeling Spatio- Temporal Variation in House Prices*. *September 2015*. https://doi.org/10.1080/13658810802672469

International Association of Assessing Officers. (2014). Guidance on International Mass Appraisal and Related Tax Policy. *Journal of Property Tax Assessment & Administration*, *11*(1), 5–33. https://search.proquest.com/docview/1524709595?accountid=14775%0Ahttp://openurl.ac.uk/athens:uls/?genre=article&issn=13571419&volume=11&issue=1&title=Journal+of+Property+Tax+Assessment+%26+Administration&spage=5&date=2014-01-01&atitle=Guidance+on+Internat

International Association of Assessing Officers. (2017). *Standard on Mass Appraisal of Real Property International Association of Assessing Officers*. www.iaao.org

International Association of Assessing Officers. (2018). *Standard on Automated Valuation Models*. https://www.iaao.org/media/standards/AVM_STANDARD_2018.pdf

Ismail, S., & Buyong, T. (1998). Residential property valuation using geographic information system. *Buletin Geoinformasi*, *2*(2), 249–266.

Jahanshiri, E., Buyong, T., Rashid, A., & Shariff, M. (2011). *A Review of Property Mass Valuation Models*. *19*(August 2010), 23–30.

James, V., Wu, S., Gelfand, A., & Sirmans, C. (2005). Apartment rent prediction using spatial modeling. *Journal of Real Estate Research*, *27*(1), 105–136.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. In

Stroke and Vascular Neurology. https://doi.org/10.1136/svn-2017-000101

Kauko, T., & Amato, M. (2008). *Mass appraisal methods: An international perspective for property valuers* (T. Kauko & D. Maurizio (Eds.); First). Wiiley-Blackwell.

Kockelman, K. M. (1997). Effects of location elements on home purchase prices and rents in San Francisco bay area. *Transportation Research Record*, *1606*, 40–50. https://doi.org/10.3141/1606-06

Koramaz, T. K., & Dokmeci, V. (2012). Spatial Determinants of Housing Price Values in Istanbul Spatial Determinants of Housing Price Values in Istanbul. *European Planning Studies*, *20*(7), 1221–1237. https://doi.org/10.1080/09654313.2012.673569

Kuburić, M., Tomić, H., & Ivić, S. M. (2012). Use of Multicriteria Valuation of Spatial Units in a System of Mass Real Estate Valuation. *Cartography and Geoinformation*, *11*(17), 58–74.

Labropoulos, T. (2013). *Αυτοματοποίηση μαζικών εκτιμήσεων αξιών ακινήτων για τον Ελληνικό χώρο, Διδακτορική Διατριβή*. National Technical University of Athens.

Lake, I. R., Lovett, A. A., Ian, J., & Day, B. (2000). Using GIS and large-scale digital data to implement hedonic pricing studies. *International Journal of Geographical Information Science*, *14*(6), 521–541. https://doi.org/10.1080/136588100415729

Lelo, K., Monni, S., & Tomassi, F. (2018). Urban inequalities in Italy: A comparison between rome, Milan and Naples. *Entrepreneurship and Sustainability Issues*, *6*(2), 939–957. https://doi.org/10.9770/jesi.2018.6.2(31)

Lewis, D. J., & Plantinga, A. J. (2007). Policies for Habitat Fragmentation: Combining Econometrics with GIS-Based Landscape Simulations. *Land Economics, 83*(2), 109–127. https://doi.org/10.3368/le.83.2.109

Limsombunchai, V. (2004). House price prediction: hedonic price model vs. artificial neural network. *New Zealand Agricultural and Resource Economics Society Conference*, 25–26.

Liu, X. (2013). Spatial and Temporal Dependence in House Price Prediction. *Journal of Real Estate Finance and Economics*, *47*(2), 341–369. https://doi.org/10.1007/s11146-011-9359-3

Lleo, S. (2019). Gods and Robots: Myths, Machines, and Ancient Dreams of Technology. *Quantitative Finance*. https://doi.org/10.1080/14697688.2019.1591632

Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S. (2014). Geographically weighted regression with a non-Euclidean distance metric: A case study using hedonic house price data. *International Journal of Geographical Information Science*, *28*(5), 660–681. https://doi.org/10.1080/13658816.2013.865739

Lybeck, E. (2017). *The Black Swan: the Impact of the Highly Improbable*. Taylor & Francis.

Mae, F. (1997). *Spatial Dependence and House Price Index Construction*. *222*, 203–222.

Makridakis, S., & Bakas, N. (2016). Forecasting and uncertainty: A survey. *Risk and Decision Analysis*, *6*(1), 37–64. https://doi.org/10.3233/RDA-150114

Mccluskey, W. J., Cornia, G. C., Walters, L. C., Kitchen, H., & Slack, E. (2013). *A Primer on Property Tax Administration and Policy Edited by* (W. J. Mccluskey, G. C. Cornia, L. C. Walters, H. Kitchen, & E. Slack (Eds.)). Blackwell Publishing Ltd.

McCluskey, W. J., Deddis, W., Mannis, A., McBurney, D., & Borst, R. (1997). Interactive application of computer assisted mass appraisal and geographic information systems. *Journal of Property Valuation & Investment*, *15*(5), 448–465.

McCluskey, W. J., & Plimmer, F. (2011). The creation of fiscal space for the property tax: The case of Central and Eastern Europe. *International Journal of Strategic Property Management*, *15*(2), 123–138.

McMillen, D. P. (2004). Geographically weighted regression: the analysis of spatially varying relationships. *American Journal of Agricultural Economics*, *86*(2), 554–556.

Melnik, S. V, & Cenedella, D. S. (2008). Real property taxation and assessment processes: A case for a better model. *NYUJ Legis. & Pub. Pol'y*, *12*, 259.

Mooya, M. M. (2017). Automated Valuation Models and Economic Theory. In *Advances in Automated Valuation Modeling* (pp. 33–57). Springer International

Publishing. https://doi.org/10.1007/978-3-319-49746-4_3

Morrow-jones, H. A., Anacker, K. B., Chen, W., Homer, M. W., & Malecki, E. (2005). Rapid Property Transactions and Price Changes: An Explorato. *Focus on Geography, 48*(3), 1–25.

Moulas, A., & Dimopoulos, T. (2015). *A proposal of a mass appraisal system in Greece with CAMA system. Evaluating GWR and MRA techniques. The case study of Thessaloniki Municipality.*

Mulaku, G. C., & Kamau, J. (2010). Computer-Assisted Analysis of the Impact of location on residential property value: A case study of Nairobi, Kenya. *Appraisal Journal*, *78*(3), 270–282.

NCSS Statistical Software. (n.d.). *Appraisal Ratio Studies*.

Norregaard, J. (2013). Taxing Immovable Property Revenue Potential and Implementation Challenges. *IMF Working Papers*, *13*(129), 1. https://doi.org/10.5089/9781484369050.001

Norregaard, M. J. (2013). *Taxing immovable property revenue potential and implementation challenges* (Issues 13–129). International Monetary Fund.

O’Connor, P. M. (2013). Use of statistical models and GIS in residential market analysis. *Journal of Property Tax Assessment & Administration*, *10*(3), 37–60.

Olden, J. D., & Jackson, D. A. (2002a). Illuminating the “black box”: A randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*. https://doi.org/10.1016/S0304-3800(02)00064-9

Olden, J. D., & Jackson, D. A. (2002b). Illuminating the “black box”: Understanding variable contributions in artificial neural networks. *Ecological Modelling*, *154*, 135–150.

Olson, C., & Levy, J. (2018). Transforming marketing with artificial intelligence. *Applied Marketing Analytics*.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science*, *263*(5147), 641 LP – 646. https://doi.org/10.1126/science.263.5147.641

Páez, A. (2005). Local Analysis of Spatial Relationships : A Comparison of GWR and the Expansion Method. *In Computational Science and Its Applications*, 162–172.

Páez, A. (2009). Recent research in spatial real estate hedonic analysis. *Journal of Geographical Systems*, *11*(4), 311–316. https://doi.org/10.1007/s10109-009-0103-y

Páez, A., Farber, S., & Wheeler, D. (2011). A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A*, *43*(12), 2992–3010. https://doi.org/10.1068/a44111

Pagano, D. (2018). *Machine Learning will replace tasks, not jobs, say MIT researchers*. MIT News.

Panayiotou, P. A., Pattichis, C., Jenkins, D., & Plimmer, F. (2000). Modular artificial neural network valuation system. *Proceedings of the Mediterranean Electrotechnical Conference - MELECON*, *2*, 457–460.

Paterson, R. W., & Boyle, K. J. (2002). Out of sight, out of mind? Using GIS to incorporate visibility in hedonic property value models. *Land Economics*, *78*(3), 417–425. https://doi.org/10.2307/3146899

Patriche, C. V., & Vasiliniuc, I. (2009). Aspects regarding the usefulness of Geographically Weighted Regression (GWR) for digital mapping of soil parameters ASPECTS REGARDING THE USEFULNESS OF GEOGRAPHICALLY WEIGHTED REGRESSION (GWR). *Lucrări Ştiinţifice*, *52*(seria Agronomie), 415–420.

Plevris, V., Bakas, N., Markeset, G., & Bellos, J. (2017). Literature review of masonry structures under earthquake excitation utilizing machine learning algorithms. *COMPDYN 2017 - Proceedings of the 6th International Conference on Computational Methods in Structural Dynamics and Earthquake Engineeringtructural Dynamics and Earthquake Engineering*, *1*, 2685–2694. https://doi.org/10.7712/120117.5598.18688

Pokryshevskaya, E. B., & Antipov, E. A. (2011). Applying a CART-based approach for the diagnostics of mass appraisal models. *Economics Bulletin*, *31*, 2521–2528. https://doi.org/10.2139/ssrn.1729627

Pozna, C., & Antonya, C. (2016). Issues about autonomous cars. *SACI 2016 - 11th IEEE International Symposium on Applied Computational Intelligence and Informatics, Proceedings*. https://doi.org/10.1109/SACI.2016.7507360

Renshow, E. F. (1958). Scientific appraisal. *National Tax Journal*, *11*(4), 314–322.

RICs. (2019). *Bank Lending Valuations and Mortgage Lending Value, 1st edition*. *March*.

RICS. (2013). Automated Valuation Models (1st Ed.). *RICS Information Paper*.

Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and Empirical Analysis of ReliefF and RReliefF. *Machine Learning*. https://doi.org/10.1023/A:1025667309714

Robson, G., & Downie, M. L. (2008). *Automated Valuation Models: an international perspective*. *November*.

Rodriguez, M., Sirmans, C., & Marks, A. (1995). Using geographic information systems to improve real estate analysis. *Journal of Real Estate Research, 10*(2), 163–173.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55. https://doi.org/10.1086/260169

Ross, S. M., Kelly, J. J., Sullivan, R. J., Perry, W. J., Mercer, D., Davis, R. M., Washburn, T. D., Sager, E. V, Boyce, J. B., & Bristow, V. L. (1996). *Stochastic processes* (Vol. 2). Wiley New York.

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1249. https://doi.org/10.1002/widm.1249

Scopus. (2019).

Tabales, J. N., Carmona, F. R., & y Ocerin, J. M. (2013a). Implicit Prices in Urban Real Estate Valuation. *Revista de La Construcción*, *12*(2), 116–126. https://doi.org/10.4067/s0718-915x2013000200009

Tabales, J. N., Carmona, F. R., & y Ocerin, J. M. (2013b). Implicit Prices in Urban Real Estate Valuation. *Revista de La Construcción*, *12*(2), 116–126.

https://doi.org/10.4067/s0718-915x2013000200009

Thalassinos, E., Liapis, K., & Thalassinos, J. (2014). The role of the rating companies in the recent financial crisis in the Balkan and black sea area. In *Economic Crisis in Europe and the Balkans* (pp. 79–115). Springer.

Thériault, M., Des Rosiers, F., Villeneuve, P., & Kestens, Y. (2003). Modelling interactions of location with specific value of housing attributes. *Property Management*.

Tian, Y., & Yang, J. (2015). Application of Geographic Information System on Urban Residential Real Estate Mass Appraisal. *Applied Mechanics and Materials*, *744–746*, 1665–1668. https://doi.org/10.4028/www.scientific.net/AMM.744-746.1665

Tiebout, C. (1956). A pure theory of local expenditures. *The Journal of Political Economy*, *64*(5), 416–424. https://doi.org/10.1086/257839

Tretton, D. (2007). Where is the world of property valuation for taxation purposes going? *Journal of Property Investment and Finance*, 25(5): 482-514.

Van Noorden, R. (2014). Elsevier opens its papers to text-mining. *Nature*, *506*(7486), 17.

Vyas, A., Yu, S., & Paik, J. (2018). Fundamentals of digital image processing. In *Signals and Communication Technology*. https://doi.org/10.1007/978-981-10-7272-7_1

Walt, J. van der. (2016). *AN ANALYSIS OF THE USE OF MASS APPRAISAL METHOS FOR AGRICULTURAL PROPERTIES*. *October*.

Wang, D., & Li, V. J. (2019). Mass appraisal models of real estate in the 21st century: A systematic literature review. *Sustainability (Switzerland)*, *11*(24), 1–14. https://doi.org/10.3390/su11247006

Warren-Myers, G. (2015). Editorial. *Journal of Property Investment & Finance*, *33*(2), 1–3. https://doi.org/10.1108/jpif-01-2015-0001

Worth, F. (2004). *Modeling Spatial and Temporal House Price Patterns : A Comparison of Four Models*. 167–191.

Worzala, E.; Lenk, M. . S. (1995). An Exploration of Neural Networks and Its

Application to Real Estate Valuation. *Journal of Real Estate Research*.

Yomralioglu, T., & Nisanci, R. (2004). Nominal Asset Land Valuation Technique by GIS Nominal Asset Land Valuation Technique by GIS. *Proceedings of the FIG Working Week 2004*, 1–9.

Θ. Δημόπουλος. (2016). Εισήγηση για αποτελεσματικότερη φορολόγηση ακινήτων στην Κύπρο μέσω της Νέας Γενικής Εκτίμησης, Μαζικές Εκτιμήσεις & η χωρική διάσταση της ακίνητης περιουσίας. *Ελληνική Εταιρεία Γεωγραφικών Συστημάτων Πληροφοριών - Ελληνικό Ινστιτούτο Εκτιμητικής*.

Πασιουλής Βαρνάβας, & Δημόπουλος Θωμάς. (2016). Μαζικές Εκτιμήσεις και Νέα Γενική Εκτίμηση στην Κύπρο σε τιμές 01.01.2013. *Ελληνική Εταιρεία Γεωγραφικών Συστημάτων Πληροφοριών - Ελληνικό Ινστιτούτο Εκτιμητικής*.

APPENDIX I

Data provided from the Cyprus Department of Lands and Surveys

This section is part of the The Study on Refining the Parameters of the CAMA Model (Charalambous & Hanlon, 2015)**.**

The DLS maintains a register of all real property transactions in Cyprus. Each transaction is processed and recorded through the Land Information System (LIS). The following characteristics are recorded:

- Sale file and date;
- Declared sale price;
- Accepted sale price;
- Sales analysis indicator (code 1: D=C without local inspection, code 2 D≠C without local inspection, code 3 D≠C with, code 4 D≠C with local inspection and court decision);
- Property identifier;
- Contract of sale file and date;
- Seller and buyer;
- Planning zone and category;
- Property type;
- Remarks;

The seller and the buyer attend the District Lands Office at the date of the transfer. Under existing laws, both the seller and buyer declare the agreed value of the property and the Department charges the relevant transfer fees based on that amount. An authorized valuer evaluates the sale transaction, at the date of the transfer, to determine if the declared sales price corresponds to the market value. If the declared value does not equal to the market value at the date of the transfer, the authorized valuer proceeds with a new valuation that is considered to be the true market value of the property based on comparable sales in the neighbourhood. The buyer has to pay transfer fees on the estimated market value, otherwise he can object within 45 days by providing a valuation report that he believes is the true market value. The District Lands Office (DLO) Valuer will assess the evidence of the private valuation and proceed with a local inspection of the property, if necessary.

The resultant decision is communicated to the buyer within 3 months. If the buyer disagrees with the decision of the DLO Valuer, then he can apply to the Supreme Court for the determination of the market value under Section 80 of the Immovable Property Law (tenure, registration and valuation) Cap. 224.

When the declared value equals the accepted sale price in the LIS, the sales analysis indicator automatically shows the code 1, which means "genuine sale". In cases where the unit is under construction and it does not yet have a title, parties agree to sign a contract of sale instead of directly proceeding with the transfer. There are also some cases where the parties agree to sign a contract of sale, rather than complete the transfer, even with the existence of a title. In such cases, the contract of sale date, rather than the transfer date, is used for valuation purposes.

The existing process for registration of transactions does not record property characteristics beyond those listed above. On the other hand, the valuer gathers additional data on property characteristics from the buyer/seller during the transfer process.

For the purpose of this study, the sales register has been supplemented by information on property characteristics that have been collected from owners during the registration of transactions as well as during the data capture process and from the external inspection of properties for the implementation of the new GV. This database was utilised in the study to test the potential for applying MRA analysis in the application of the sales comparison approach to mass appraisal of property.

For Flats, the following property characteristics were available for analysis for the period Q1 2008 – Q2 2014:

- Enclosed area in square meters;
- Covered area in square meters;
- Uncovered area in square meters;
- Year built;
- Class (luxury, A- very good, B- standard, C below standard, D- very poor);
- Condition (very good, good, fair, bad);
- View (restricted, standard, premium, sea view);
- District (5 administrative district of Cyprus);
- Town/village identifier;

- Quarter of town/village;
- Planning zone type (17 different types);
- Planning zone density.

For houses, the following property characteristics were available for analysis for the period Q1 2008 – Q3 2013:

- Enclosed area in square meters;
- Covered area in square meters;
- Uncovered area in square meters;
- Year Built;
- Class (luxury, A- very good, B- standard, C below standard, D- very poor);
- Condition (very good, good, fair, bad);
- View (restricted, standard, premium, sea view);
- District (5 administrative district of Cyprus);
- Town/village identifier;
- Quarter of town/village;
- Planning zone type (17 different types);
- Planning zone density.

For undeveloped land, the following property characteristics were available for analysis for the period Q1 2008 – Q3 2014:

- Area in square meters;
- District (5 administrative district of Cyprus);
- Town/village identifier;
- Quarter of town/village;
- Planning zone type;
- Planning zone density;
- Accessibility (no access, access, right of way);
- Road parcel relation type (none, side access, corner, two sides, three sides, four sides);
- Shape (regular, irregular, highly irregular).

APPENDIX II

Table A: Formulas utilized in the models

Number	**Investigated Formula**
1	Accepted_price ~ Town_village_name_mod
2	Accepted_price ~ Road_code
3	Accepted_price ~ Shape_code
4	Accepted_price ~ Unit_class_code
5	Accepted_price ~ Unit_condition_code
6	Accepted_price ~ Unit_condition_mod_code
7	Accepted_price ~ Unit_view_code
8	Accepted_price ~ Unit_enclosed_extent
9	Accepted_price ~ Unit_covered_extent
10	Accepted_price ~ Unit_uncovered_extent
11	Accepted_price ~ Unit_encl_and_covered_extent
12	Accepted_price ~ Unit_total_extent
13	Accepted_price ~ Unit_adjusted_extent
14	Accepted_price ~ price_index_sale_acceptance_date
15	Accepted_price ~ flats_index_sale_acceptance_date
16	Accepted_price ~ Sale_acceptance_year

17	Accepted_price ~ age_at_sale
18	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent
19	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + Unit_uncovered_extent
20	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + Town_village_name_mod
21	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + Unit_class_code
22	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + Unit_condition_code
23	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + Unit_condition_mod_code
24	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + price_index_sale_acceptance_date
25	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + flats_index_sale_acceptance_date
26	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + Sale_acceptance_year
27	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale
28	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Town_village_name_mod
29	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Unit_class_code
30	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Unit_condition_code

31	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Unit_condition_mod_code
32	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + price_index_sale_acceptance_date
33	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + flats_index_sale_acceptance_date
34	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Sale_acceptance_year
35	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Unit_condition_mod_code + Town_village_name_mod
36	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Unit_condition_mod_code + Unit_class_code
37	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Unit_condition_mod_code + price_index_sale_acceptance_date
38	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Unit_condition_mod_code + flats_index_sale_acceptance_date
39	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Unit_condition_mod_code + Sale_acceptance_year
40	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Unit_condition_mod_code + flats_index_sale_acceptance_date + Town_village_name_mod
41	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Unit_condition_mod_code + flats_index_sale_acceptance_date + Unit_class_code
42	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Unit_condition_mod_code + flats_index_sale_acceptance_date + Unit_class_code + Town_village_name_mod

43	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Unit_condition_mod_code + flats_index_sale_acceptance_date + Unit_class_code + Town_village_name_mod + Unit_uncovered_extent
44	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Unit_condition_mod_code + flats_index_sale_acceptance_date + Unit_class_code + Town_village_name_mod + Unit_uncovered_extent + price_index_sale_acceptance_date
45	Accepted_price ~ Unit_enclosed_extent + Unit_covered_extent + age_at_sale + Unit_condition_mod_code + flats_index_sale_acceptance_date + Unit_class_code + Town_village_name_mod + Unit_uncovered_extent + price_index_sale_acceptance_date + Unit_view_code

Table B: Accuracy Measures for Linear Regression

model	lin_coef	RMSE	MAE	MAPE	sr	method
1	0.05	53216.5	42414.9	0.39	1.23	Linear Regression
2	0	55032	44374.4	0.4	1.24	Linear Regression
3	0	54683.4	44076.5	0.4	1.23	Linear Regression
4	0.04	52935.8	41734.4	0.38	1.21	Linear Regression
5	0.12	51147.5	40382.8	0.37	1.21	Linear Regression
6	0.12	51213	40508.9	0.37	1.21	Linear Regression
7	0.05	52198.7	42011.9	0.38	1.21	Linear Regression
8	0.45	43077.7	33357	0.27	1.12	Linear Regression
9	0.33	45443.4	34504.8	0.3	1.16	Linear Regression
10	0.07	52615.6	42011.3	0.38	1.22	Linear Regression
11	0.53	40859.7	30740.6	0.24	1.1	Linear Regression
12	0.51	41193.3	31209.6	0.25	1.12	Linear Regression
13	0.52	41146.1	31448.4	0.25	1.11	Linear Regression
14	0	54683	44027.9	0.4	1.24	Linear Regression
15	0	54793.9	44097.9	0.41	1.24	Linear Regression
16	0	55198.7	44376.3	0.41	1.24	Linear Regression
17	0.15	49313	39069.6	0.35	1.19	Linear Regression
18	0.57	39678.8	28811.1	0.23	1.1	Linear Regression
19	0.6	38436.2	28169.8	0.22	1.1	Linear Regression
20	0.65	37996.5	28112	0.22	1.08	Linear Regression
21	0.55	38447.9	28313.1	0.22	1.09	Linear Regression
22	0.65	35698.7	26481.4	0.2	1.08	Linear Regression
23	0.65	35799.8	26539.2	0.21	1.08	Linear Regression
24	0.59	39641.8	29273.4	0.23	1.09	Linear Regression
25	0.58	39875	29248.4	0.23	1.1	Linear Regression
26	0.59	39614.4	29373.6	0.23	1.1	Linear Regression

27	0.73	31621.4	23221.8	0.18	1.06	Linear Regression
28	0.74	31232.6	22887.1	0.18	1.06	Linear Regression
29	0.73	31702	23502.1	0.19	1.06	Linear Regression
30	0.76	31519	23399.4	0.19	1.05	Linear Regression
31	0.75	31774	23472.9	0.19	1.06	Linear Regression
32	0.74	31920.5	23607.3	0.19	1.06	Linear Regression
33	0.73	31708.8	23312.6	0.18	1.06	Linear Regression
34	0.74	31603.9	23266.5	0.18	1.05	Linear Regression
35	0.76	31547.7	23209.7	0.18	1.06	Linear Regression
36	0.75	31927.8	23676.7	0.19	1.06	Linear Regression
37	0.75	31974.5	23742.5	0.19	1.06	Linear Regression
38	0.75	31767.5	23465.3	0.19	1.06	Linear Regression
39	0.75	31776.2	23583.4	0.18	1.05	Linear Regression
40	0.76	31512.1	23176.3	0.18	1.06	Linear Regression
41	0.75	31918.3	23665.7	0.19	1.06	Linear Regression
42	0.76	31667.2	23411.6	0.19	1.06	Linear Regression
43	0.77	30868.9	23027.4	0.18	1.06	Linear Regression
44	0.78	30732	23106.5	0.18	1.05	Linear Regression
45	0.79	30495.7	22973.8	0.18	1.05	Linear Regression

Table C: Accuracy Measures for the Random Forests Method

model	lin_coef	RMSE	MAE	MAPE	sr	method
1	0.05	53211.2	42398.7	0.39	1.23	Random Forests
2	0	55001.3	44342.1	0.4	1.24	Random Forests
3	0	54648.9	44019.9	0.4	1.23	Random Forests
4	0.04	52939.6	41736.9	0.38	1.21	Random Forests
5	0.12	51135.3	40365.5	0.36	1.21	Random Forests
6	0.12	51232.2	40543.9	0.37	1.21	Random Forests
7	0.05	52202.5	42018.6	0.38	1.21	Random Forests
8	0.46	46422.5	34488.6	0.27	1.11	Random Forests
9	0.41	48011.5	36344.7	0.31	1.16	Random Forests
10	0.09	54920.8	43351.6	0.38	1.2	Random Forests
11	0.5	45033	31782.6	0.24	1.09	Random Forests
12	0.59	45231.8	31994.7	0.25	1.1	Random Forests
13	0.5	44896.3	33133.9	0.26	1.1	Random Forests
14	0.01	59853.3	47559.5	0.43	1.24	Random Forests
15	0.01	56177.9	45280	0.42	1.25	Random Forests
16	0	55192.8	44356.3	0.41	1.24	Random Forests
17	0.23	62465.3	44821	0.36	1.15	Random Forests
18	0.6	44138.9	29343	0.22	1.09	Random Forests
19	0.58	38781.9	27819.4	0.22	1.1	Random Forests
20	0.56	40920.1	27676.9	0.22	1.1	Random Forests
21	0.48	37838.1	28108.4	0.23	1.11	Random Forests
22	0.54	37359.4	26894.2	0.22	1.1	Random Forests
23	0.52	37173	27065.2	0.22	1.11	Random Forests
24	0.55	41474.8	28768.3	0.22	1.09	Random Forests
25	0.55	40426.3	28831	0.22	1.09	Random Forests
26	0.52	41408.3	29358.4	0.23	1.1	Random Forests

27	0.69	30659.6	21574.3	0.17	1.06	Random Forests
28	0.61	31166.7	21881.3	0.17	1.07	Random Forests
29	0.57	31182.6	23234	0.19	1.09	Random Forests
30	0.57	32217.1	23674.4	0.19	1.09	Random Forests
31	0.57	32244.9	23779	0.19	1.09	Random Forests
32	0.65	30730.7	21856	0.17	1.07	Random Forests
33	0.66	29845.2	21583.6	0.17	1.07	Random Forests
34	0.62	30843.9	22166.5	0.18	1.08	Random Forests
35	0.55	32365.3	23636.7	0.19	1.09	Random Forests
36	0.52	32518	24520.5	0.2	1.1	Random Forests
37	0.55	32605.8	24089.8	0.2	1.1	Random Forests
38	0.57	31605.6	23487.6	0.19	1.09	Random Forests
39	0.54	32362.1	24101.8	0.2	1.1	Random Forests
40	0.71	29418.4	20132.1	0.15	1.06	Random Forests
41	0.7	29499.5	20647.8	0.16	1.06	Random Forests
42	0.69	29050	20133.3	0.15	1.06	Random Forests
43	0.7	27603.7	19760.1	0.15	1.06	Random Forests
44	0.71	28122.1	19856.6	0.15	1.06	Random Forests
45	0.71	27755.9	19723.8	0.15	1.07	Random Forests

Table D: Accuracy Measures Comparison

model	lin_coef	RMSE	MAE	MAPE	sr	method
1	0.000	0.010	0.038	0.000	0.000	comparison
2	-	0.056	0.073	0.000	0.000	comparison
3	-	0.063	0.128	0.000	0.000	comparison
4	0.000	-0.007	-0.006	0.000	0.000	comparison
5	0.000	0.024	0.043	2.740	0.000	comparison
6	0.000	-0.037	-0.086	0.000	0.000	comparison
7	0.000	-0.007	-0.016	0.000	0.000	comparison
8	-2.198	-7.474	-3.336	0.000	0.897	comparison
9	-21.622	-5.496	-5.194	-3.279	0.000	comparison
10	-	-4.287	-3.140	0.000	1.653	comparison
11	5.825	-9.717	-3.333	0.000	0.913	comparison
12	-14.545	-9.346	-2.484	0.000	1.802	comparison
13	3.922	-8.717	-5.220	-3.922	0.905	comparison
14	-	-9.028	-7.712	-7.229	0.000	comparison
15	-	-2.494	-2.645	-2.410	-0.803	comparison
16	-	0.011	0.045	0.000	0.000	comparison
17	-42.105	-23.533	-13.712	-2.817	3.419	comparison
18	-5.128	-10.642	-1.829	4.444	0.913	comparison
19	3.390	-0.895	1.251	0.000	0.000	comparison
20	14.876	-7.409	1.560	0.000	-1.835	comparison
21	13.592	1.599	0.725	-4.444	-1.818	comparison
22	18.487	-4.546	-1.547	-9.524	-1.835	comparison
23	22.222	-3.764	-1.963	-4.651	-2.740	comparison
24	7.018	-4.519	1.740	4.444	0.000	comparison
25	5.310	-1.373	1.437	4.444	0.913	comparison
26	12.613	-4.428	0.052	0.000	0.000	comparison

27	5.634	3.088	7.355	5.714	0.000	comparison
28	19.259	0.211	4.493	5.714	-0.939	comparison
29	24.615	1.652	1.147	0.000	-2.791	comparison
30	28.571	-2.191	-1.169	0.000	-3.738	comparison
31	27.273	-1.471	-1.296	0.000	-2.791	comparison
32	12.950	3.798	7.704	11.111	-0.939	comparison
33	10.072	6.055	7.702	5.714	-0.939	comparison
34	17.647	2.434	4.842	0.000	-2.817	comparison
35	32.061	-2.559	-1.823	-5.405	-2.791	comparison
36	36.220	-1.832	-3.501	-5.128	-3.704	comparison
37	30.769	-1.955	-1.452	-5.128	-3.704	comparison
38	27.273	0.511	-0.095	0.000	-2.791	comparison
39	32.558	-1.827	-2.174	-10.526	-4.651	comparison
40	6.803	6.873	14.058	18.182	0.000	comparison
41	6.897	7.876	13.621	17.143	0.000	comparison
42	9.655	8.621	15.057	23.529	0.000	comparison
43	9.524	11.168	15.272	18.182	0.000	comparison
44	9.396	8.869	15.129	18.182	-0.948	comparison
45	10.667	9.407	15.224	18.182	-1.887	comparison
%diff	**9.730765**	**-1.27179**	**1.443673**	**2.072518**	**-0.73431**	

APPENDIX III

Prediction Formula with 100 terms (MAE=19694 €)

$$
\begin{aligned}
&Adj.\,Accepted\ Price \\
&= 3.31113E+03 * IntArea - 4.68398E+01 * BuiltYrs * IntArea \\
&- 5.65708E-01 * BuiltYrs * CovVer * IntArea + 1.86782E+03 \\
&* UnCovVer - 2.78991E+02 * Dens * View * CovVer + 4.88587E-01 \\
&* BuiltYrs * BuiltYrs * IntArea + 6.60039E-02 * IntArea * IntArea \\
&* IntArea - 1.70363E+01 * IntArea * IntArea + 3.24869E+00 \\
&* View * Cond * ParcExt + 1.70068E-03 * ParcExt * ParcExt * Dens \\
&+ 1.30015E+01 * Cond * BuiltYrs * IntArea - 5.17967E-07 \\
&* ParcExt * ParcExt * ParcExt - 1.00466E+01 * Dens * BuiltYrs \\
&* IntArea - 5.65825E-02 * View * BuiltYrs * ParcExt - 8.15277E \\
&+ 00 * Class * ParcExt + 4.49178E-02 * BuiltYrs * UnCovVer \\
&* UnCovVer + 6.90546E+01 * CovVer * IntArea - 1.04803E+02 \\
&* CovVer * CovVer - 1.55104E-03 * ParcExt * UnCovVer * UnCovVer \\
&- 1.27722E+01 * Dens * BuiltYrs * BuiltYrs + 3.00378E+03 \\
&* BuiltYrs - 1.23882E+01 * Cond * CovVer * BuiltYrs - 5.87984E \\
&+ 01 * BuiltYrs * BuiltYrs + 5.66044E-01 * BuiltYrs * BuiltYrs \\
&* BuiltYrs - 1.16139E+02 * Class * Class * UnCovVer - 1.00749E \\
&+ 03 * Class - 1.54169E+02 * Class * Cond * BuiltYrs - 5.23747E \\
&- 02 * CovVer * CovVer * CovVer - 1.69265E-01 * CovVer * IntArea \\
&* IntArea - 3.02085E+03 * CovVer + 1.27557E+01 * Dens \\
&* ParcExt * Dens + 8.11517E+04 * Dens + 2.19071E+00 * Cond \\
&* UnCovVer * UnCovVer - 2.34493E+02 * Class * UnCovVer \\
&- 7.94913E-01 * BuiltYrs * CovVer * CovVer + 3.26919E-01 * Cond \\
&* ParcExt * CovVer - 7.49578E-06 * ParcExt * ParcExt * IntArea \\
&+ 1.38552E+02 * Dens * View * IntArea - 2.01899E+03 * View \\
&* View - 4.53539E-01 * Class * ParcExt * CovVer - 2.15116E+00 \\
&* Cond * IntArea * IntArea - 3.77756E-01 * Cond * BuiltYrs \\
&* ParcExt + 9.21765E-01 * Dens * BuiltYrs * ParcExt + 1.84117E \\
&- 03 * ParcExt * IntArea * IntArea - 3.30262E-01 * ParcExt \\
&* IntArea + 3.04830E+00 * Cond * Cond * ParcExt - 1.46508E+01 \\
&* Dens * Cond * ParcExt - 3.86245E-02 * UnCovVer * IntArea \\
&* IntArea + 5.24053E-01 * UnCovVer * CovVer * CovVer - 2.75663E \\
&+ 01 * Cond * UnCovVer * CovVer + 7.11439E+01 * View * BuiltYrs \\
&* UnCovVer + 1.02919E-01 * UnCovVer * UnCovVer * IntArea \\
&- 3.28336E+00 * Class * UnCovVer * UnCovVer - 7.87081E+02 \\
&* Dens * Cond * CovVer - 2.26100E+03 * View * View * Cond \\
&+ 5.63615E-01 * UnCovVer * CovVer * BuiltYrs + 4.64562E+01 \\
&* View * Cond * Cond + 2.81664E+01 * CovVer * CovVer * View \\
&- 1.93202E-01 * View * CovVer * ParcExt - 4.70719E+00 * Dens \\
&* BuiltYrs * UnCovVer + 6.68377E+03 * Dens * View * Class \\
&+ 3.48819E+01 * View * UnCovVer * CovVer + 5.50960E+01 * Dens \\
&* UnCovVer * CovVer + 6.22286E+01 * Dens * UnCovVer + 8.19344E \\
&+ 02 * Dens * Class * UnCovVer - 4.44616E+04 * Dens * Dens \\
&- 1.24587E+01 * Dens * Class * ParcExt + 5.88687E+00 * Class \\
&* Class * ParcExt - 1.66661E-01 * CovVer * CovVer * IntArea \\
&+ 6.65160E+03 * Dens * Dens * Cond + 3.01051E+01 * Cond
\end{aligned}
$$

$$* CovVer * CovVer + 1.08769E + 02 * BuiltYrs * CovVer - 2.07236E + 01 * View * BuiltYrs * CovVer - 5.19018E + 00 * Cond * BuiltYrs * UnCovVer - 1.55653E - 01 * UnCovVer * IntArea * BuiltYrs + 1.09671E + 00 * Dens * ParcExt * CovVer - 1.29912E + 02 * View * BuiltYrs * Dens - 7.65347E + 00 * BuiltYrs * UnCovVer * Class + 9.07541E + 01 * Dens * Cond * IntArea - 4.31216E - 04 * ParcExt * UnCovVer * CovVer - 1.37379E + 01 * View * UnCovVer * UnCovVer + 5.77048E - 03 * ParcExt * ParcExt - 5.29492E - 03 * BuiltYrs * BuiltYrs * ParcExt - 1.10686E + 03 * Class * Class * View + 6.79802E - 01 * Dens * UnCovVer * IntArea - 2.49660E + 00 * Dens * View * ParcExt + 1.30759E + 03 * Class * CovVer - 1.71103E + 02 * Class * CovVer * Class + 9.43679E + 01 * Cond * CovVer * Cond + 2.89773E + 01 * Dens * Dens * BuiltYrs - 1.90894E + 02 * UnCovVer * CovVer - 1.54163E + 03 * Dens * View * UnCovVer - 2.77202E + 03 * Dens * Dens * Dens + 5.71547E + 00 * Dens * BuiltYrs * CovVer + 2.71628E + 02 * View * UnCovVer * View + 2.80269E + 01 * UnCovVer * UnCovVer - 7.70039E + 01 * BuiltYrs * UnCovVer - 2.44312E + 00 * Dens * UnCovVer * UnCovVer + 2.38252E + 01 * Class * UnCovVer * CovVer - 4.06621E - 05 * ParcExt * ParcExt * CovVer - 7.10543E + 04$$