FIG 2018 ISTANBUL

Presented at the FIG Congress 2018, May 6-11, 2018 in Istanbul, Turkey

XXVI FIG Congress 2018

6–11 May 2018

ISTANBUL

Calculating Land Values by Using Advanced Statistical Approaches in Pendik

Prof. Dr. Arif Cagdas AYDINOGLU

Ress. Asst. Rabia BOVKIR and Dr. Ismail COLKESEN

Dept. of Geomatics Engineering Department, Gebze Technical University

Overview

- Introduction to land valuation
- Valuation methods
- Purpose of the study
- Determining factors affecting the land value
- Case study in Pendik, Istanbul
- Fuzzy logic approach for determining thematic value trends
- Random forest algorithm for determining land values
- Conclusion

Land Valuation

- In sustainable land development concept, **management of the land related information efficiently is highly important.** Because definitive and reliable information about land and real property promote the geographical enablement which leads to achieve a sustainable development.
- In addition, **governments need reliable value information of the land and real property** for implementing many legal practices such as taxation, expropriation, market capitalization, rural-urban transformation and land consolidation.
- Land valuation is **the process of determining the current market value of a real property** such as land, residence, cropland, work place, vineyard, garden for a certain time. There are many methods for valuation, yet no certain and objective method exists. The most well-known and used methods are sales comparison, cost and income methods which are called traditional approaches.
- In addition, with the developing technology **advanced statistics and machnine learning algorithms can be a good tool** for obtaining better results with the spatial a

- **Sales comparison method** is based on to analyse and **evaluate the different characteristics of recently sold properties** and to specify how these characteristics (location, landscaping, use, quality of buildings and services etc.) influence the prices.
- **Cost method** is based on the principle where the value is equal to the substitution cost, minus accrued depreciation. In this approach, the value is obtained by **adding the estimated value of the land to the construction cost** with all reproduction or replacement costs.
- In **the income method**, the present value of the future benefits and net income of property is estimated. The main criterion in determining the value by means of income is **the net income to be obtained in the future**.
- **The basis of statistical valuation is to create a mathematical model with numerical or proportional relations between land value and land related factors**. According to the literature statistical approaches such as multiple regression, artificial neural networks, decision trees, support vector machines and fuzzy logic have been used for land and real property valuations in many studies.

Purpose of the Study

- A wide variety of different parameters are affecting the property value and it makes objective evaluations quite difficult. In order to perform objective and accurate evaluations, **all parameters related to land and property** should be considered and taken into consideration.
- Because information alone is not enough for decision making, it should be analysed and interpreted logically and statistically for obtaining better results. Therefore, advanced statistics methods and machine learning algorithms can be a good tool for land valuation.
- In this study, random forest approach and **random forest algorithm were used for land valuation process in Pendik** with the help of GIS analysing tools. Resulting maps and performance of the algorithms were compared and analysed in detail for both methods.

Methodology:

Determining factors affecting land value

- According to the literature study including academic researches, nationally and internationally accepted standard documents, factors that significantly affects the value of the real estate are determined.

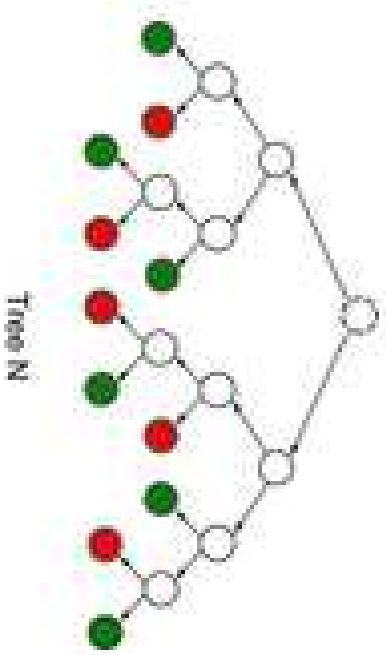
Value Factor	Explanation
Distance to Public Services	This factor expresses the importantance of administrative and official buildings for property valuation process.
Distance to Educational Facilities	Educational facilities like kindergarten, primary education, secondary education, high school and university effect the value of the property.
Distance to Health Services	The presence and proximity of health services like family health centers, polyclinics, hospitals effect the value of the property.
Distance to Religious Centers	Although having a lower effect, proximity of the religious centers to the property affects its value positively.
Distance to Cultural Centers	Presence of the cultural centers such as theaters, cinemas and libraries at the location where the property is located effect the value of the property positively.
Distance to Shopping Centers	Big shopping malls provides a lot of services about food, entertainment and shopping which make them important attraction centers for property.
Exit to the Main Roads	Closeness to the main roads increases the value of the property.
Distance to the Main Roads	The proximity to the main roads and motorways affected the value positively.
Distance to Bus Stations	The distance to the bus stations is an important element for property and it increases the value.
Distance to Metro and Tram Stations	The proximity of the property to the metro and tram stations affects the value positively.
Distance to the Green Areas	The presence and closeness of green areas such as parks, playgrounds, picnic areas etc. affects the value positively.
Parking Area	The proximity of the property to parking areas affects the value positively.
Distance to the Airport	Closeness to airway transportation is considerably influential on the property value.
Distance to Industrial Areas	The proximity to industrial and production areas affects the value negatively.
Distance to Cemeteries	The closeness to the cemetery areas generally effect the value negatively.
Distance to Infrastructure Facilities	Infrastructure facilities such as treatment facility, transformers, power transmission stations etc. have a negative effect on the value.
Slope	Slope characteristics of the land are very important for proper settlement and effects the property value.
Aspect	Because Turkey is located in the north pole properties in the south are more valuable.
Population Density	Very dense and very low population has a negative effect on the property value.
Education Level	According to the literature research, it is determined that education and culture levels of the people living in the environment where the property is located are affecting the value in the socio-cultural sense.

Methodology:

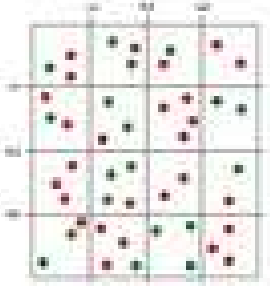
Fuzzy Logic
as Unsupervised Valuation

Random Forest
as Supervised Valuation

Value Factor
Distance to Public Services
Distance to Educational Facilities
Distance to Health Services
Distance to Religious Centers
Distance to Cultural Centers
Distance to Shopping Centers
Exit to the Main Roads
Distance to the Main Roads
Distance to Bus Stations
Distance to Metro and Tram Stations
Distance to the Green Areas
Parking Area
Distance to the Airport
Distance to Industrial Areas
Distance to Cemeteries
Distance to Infrastructure Facilities
Slope
Aspect
Population Density
Education Level

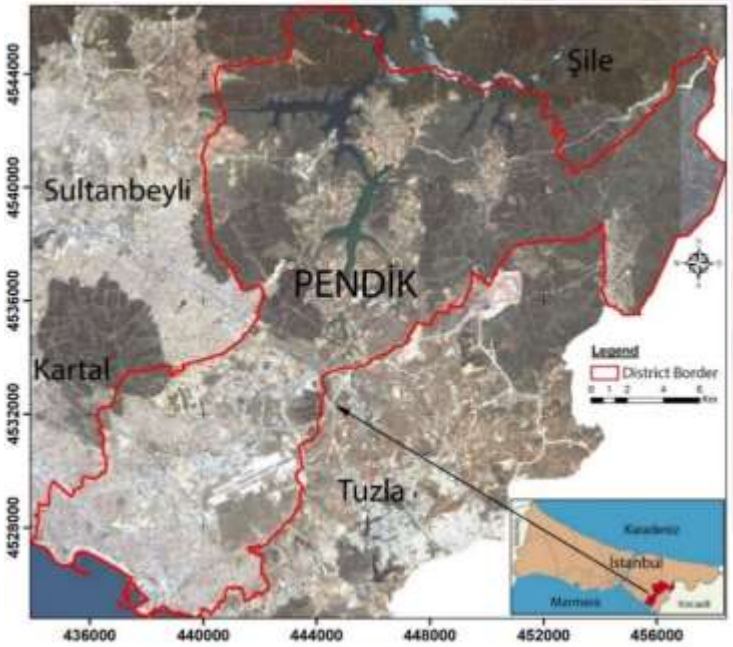


Training data set:
Ground truth value

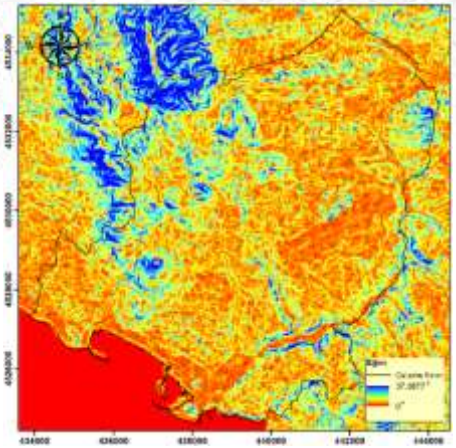


Case Study in Pendik

- Pendik is chosen is the case study area because of its large surface area and having urban-rural differences.
- Database having these criteria were produced in GIS environment for Pendik. Obtained geographic data sets were optimized with the help of the geographical analysis tools in ArcGIS software.
- Optimization should be performed for the successful analysis and process of the existing data. Raster surfaces should be produced which express the geographical characteristics of the study area. Analysis processes were performed in ArcGIS software environment.

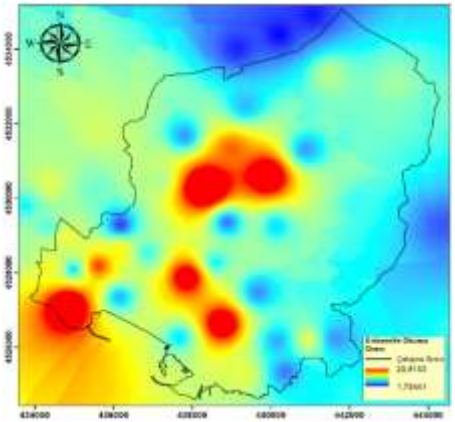


- Pixel-based thematic surfaces were created for determined thematic factors for the study area. Slope and Elevation analysis were performed by using Digital Elevation Model (DEM); Density Analysis for the creation of the surfaces reflecting population and education level characteristics and Euclidean Distance for calculating the thematic distances for other thematic factors in ArcGIS software.

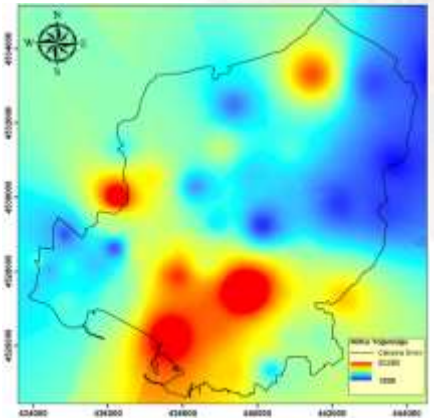


Slope

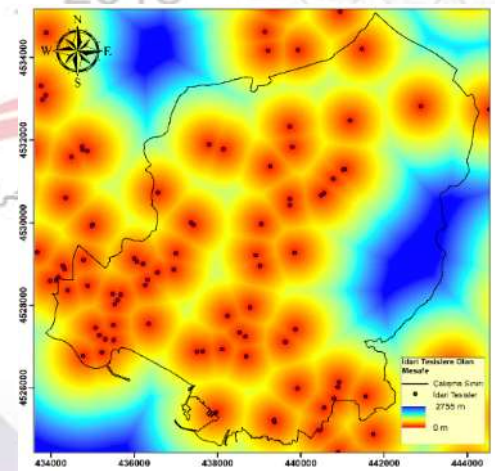
Aspect



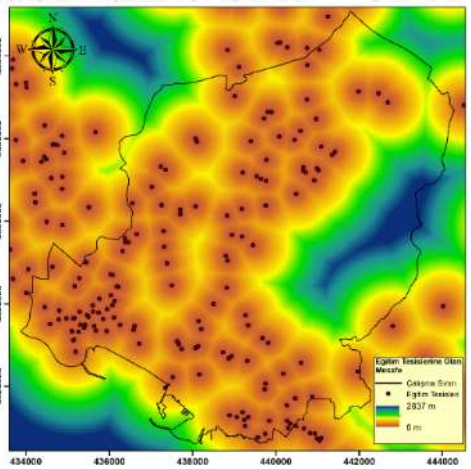
Education level



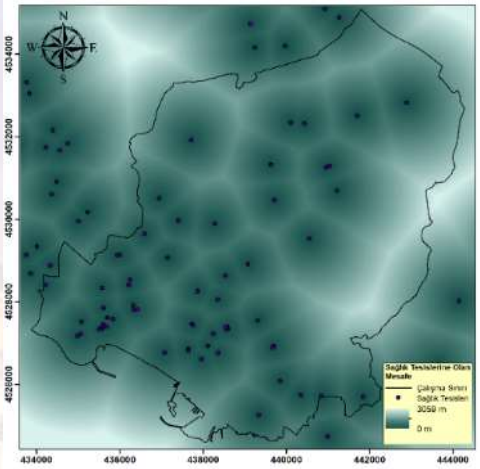
Population Density



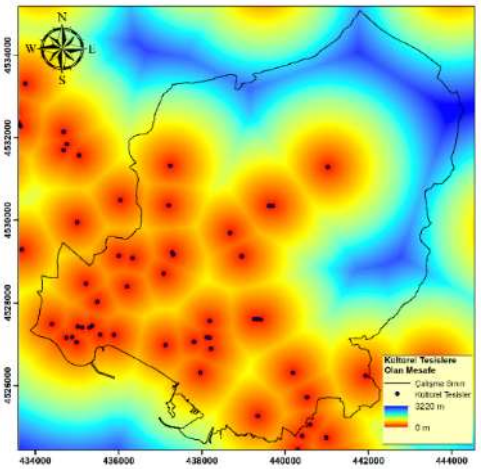
(a)



(b)

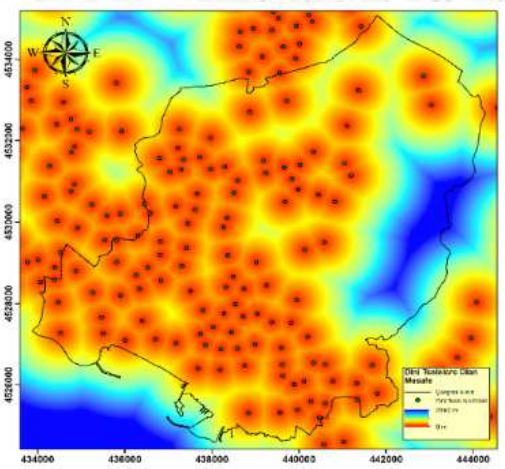


(c)

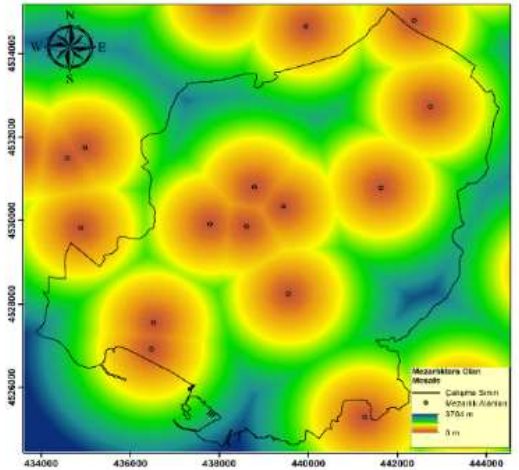


(d)

Thematic distance surfaces of Public (a), Education (b), Health (c), Cultural (d) Facilities for study area

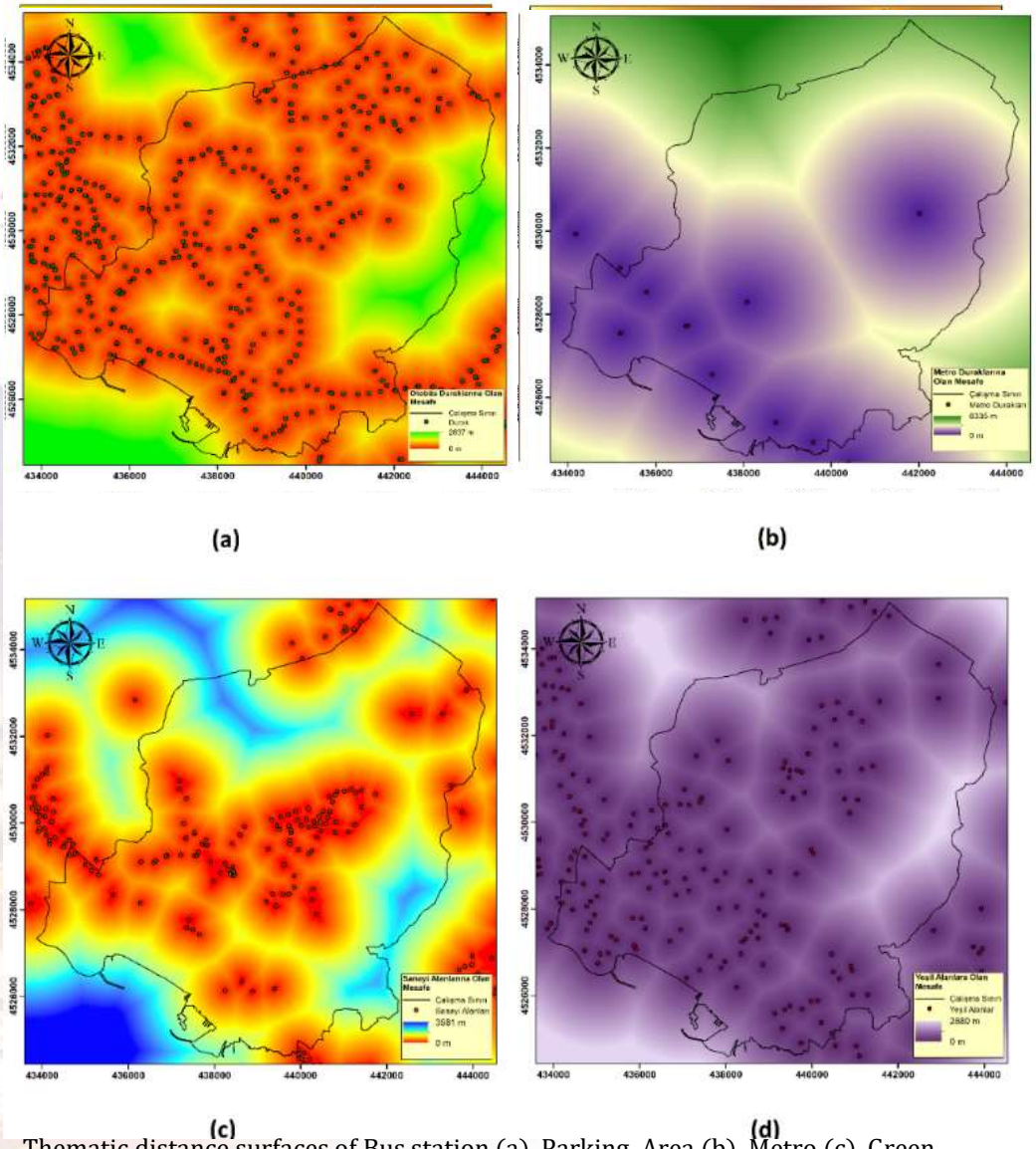


(a)

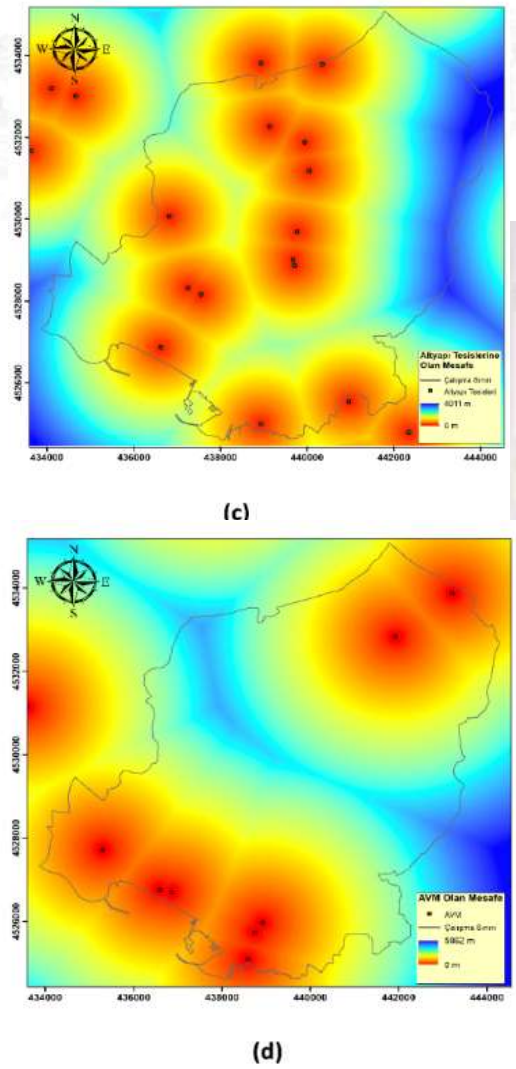


(b)

Thematic distance surfaces of Religious Centers(a), Cemetries (b) for study area



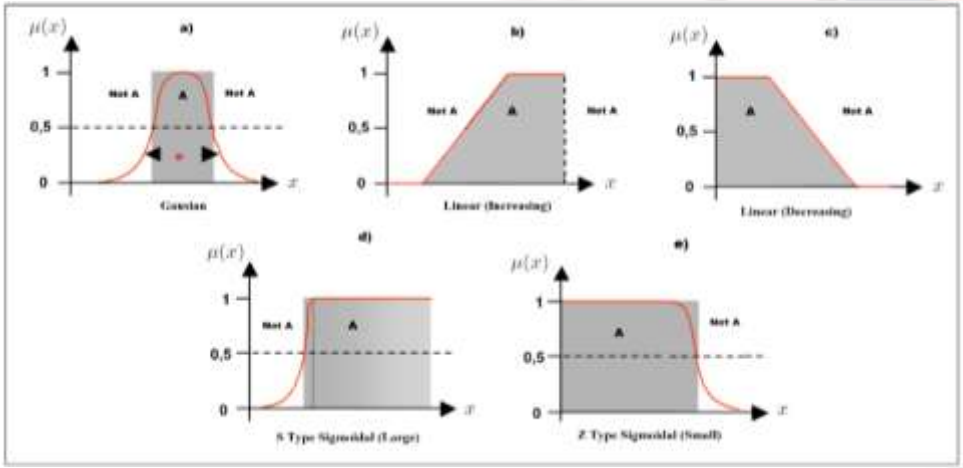
Thematic distance surfaces of Bus station (a), Parking Area (b), Metro (c), Green Space (d) for study area



Thematic distance surfaces of Infrastructure Services (c), Shopping (d) Areas for study area

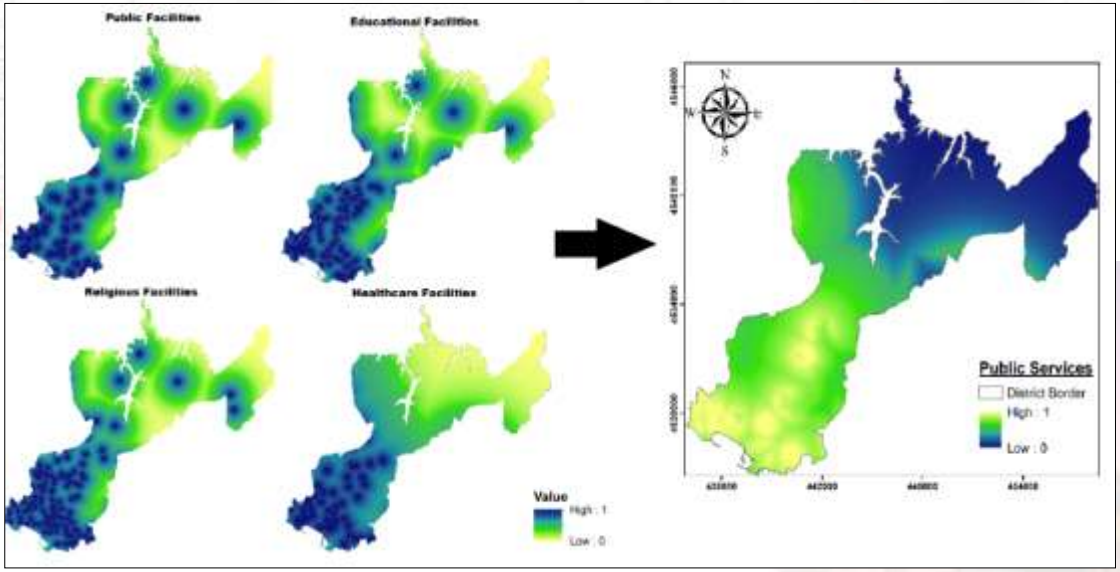
Fuzzy logic approach for determining land values

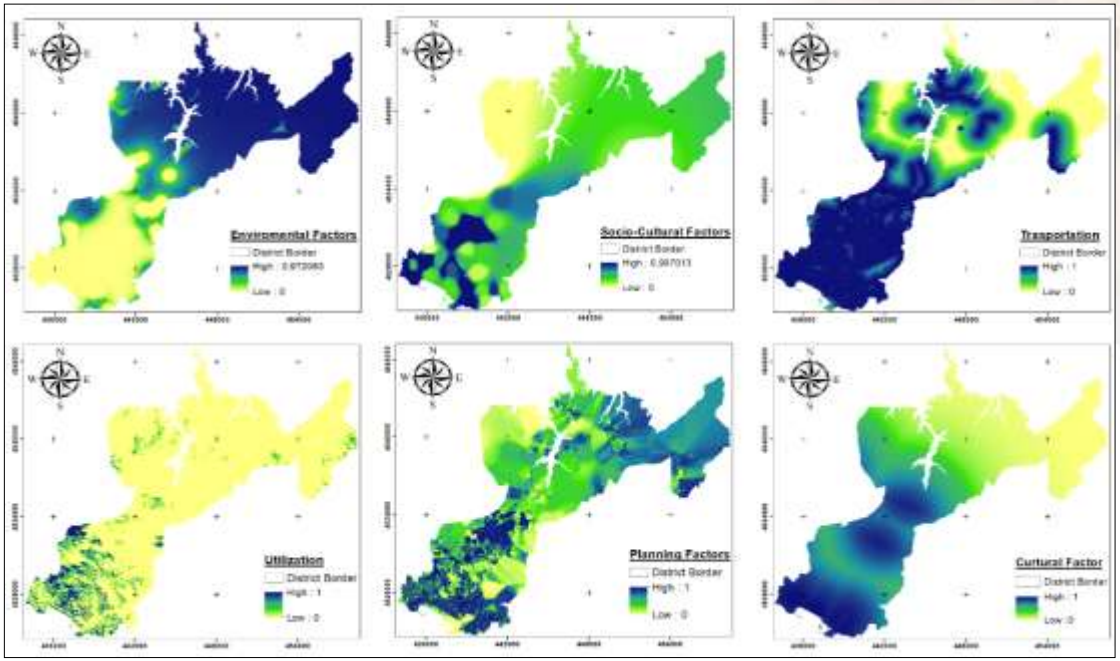
- The factors affecting the land value are highly complex and depending on various parameters. Because information can be the form of verbal expressions such as large, small and little in fuzzy logic, **it is an effective method for processing oral variables, which is very important for valuation process.**
- An element of a fuzzy set can be full member (100% membership) or a partial member (between 0% and 100% membership). That means **an assigned membership value of an element is not restricted with two values, but can be 0, 1 or any value in-between.** The shape of the membership function varies according to the application area requirements.



- Produced surfaces were gathered in seven thematic groups according to their characteristics as public services, transportation, planning, socio-cultural factors, utilization, environmental factors and cultural factors.

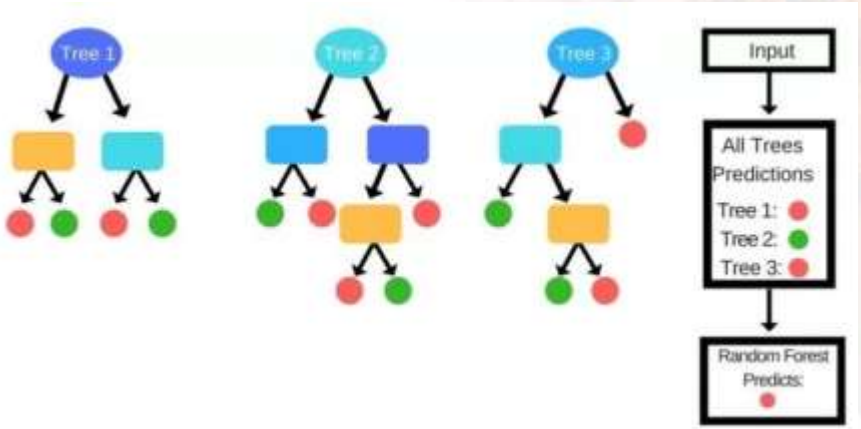
- Fuzzy membership functions were defined and applied for each thematic raster surfaces according to the characteristics; then fuzzy overlay maps were created for each thematic group.



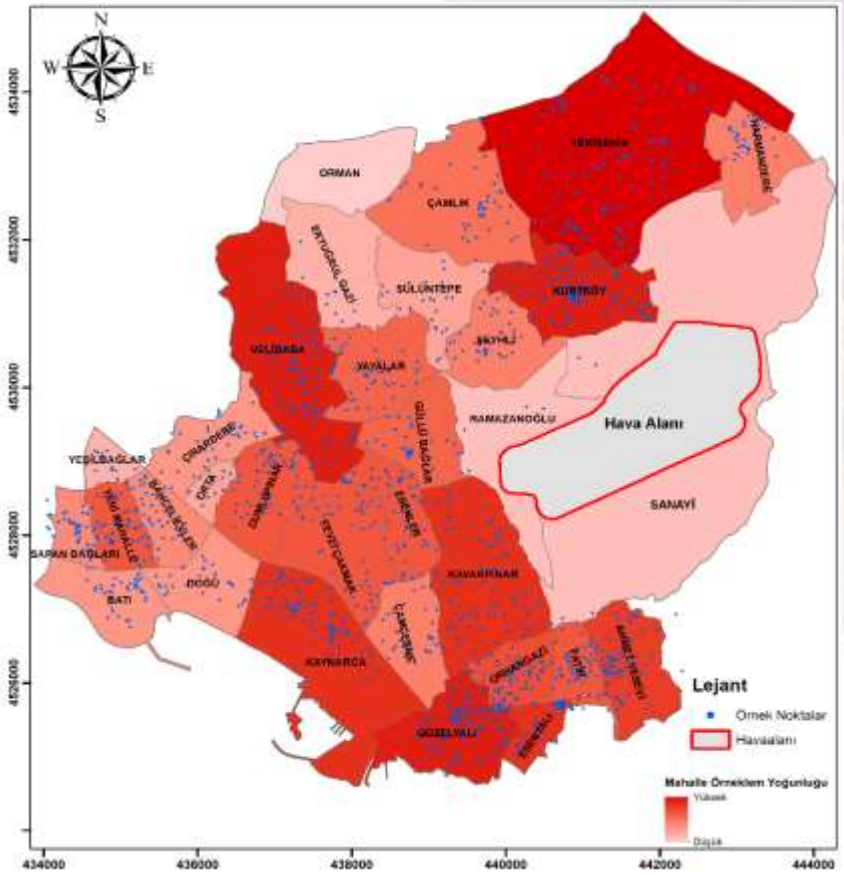


Random forest algorithm for determining land values

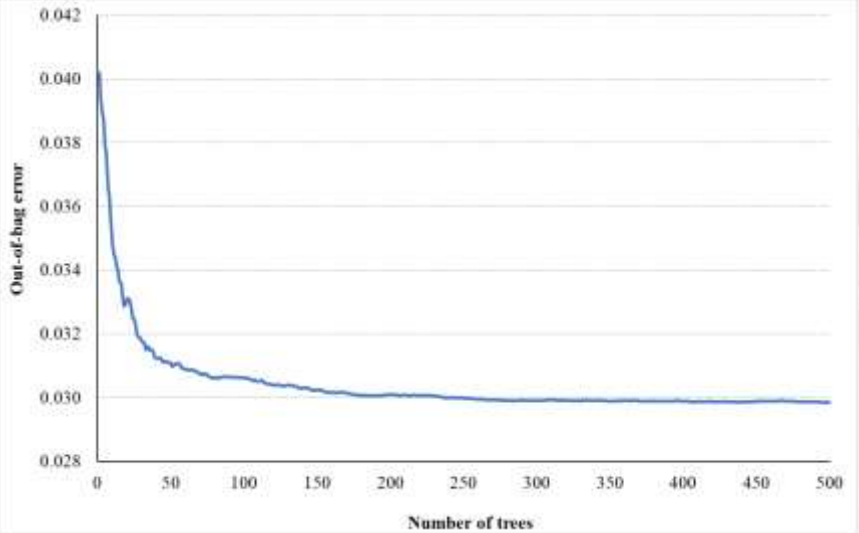
- Random forest (RF) is one of **the popular machine learning algorithms** applied for various classification and regression problems. RF, widely known as **tree based ensemble learning algorithm**, is based on the idea that training a set of decision tree (DT) learners, using randomly selected samples through bootstrap aggregating (Bagging) strategy to make final a prediction.
- For the implementation of RF algorithm, two parameters (the number of trees and the number of variables) have to be set by the analyst. In order to construct a RF prediction model, two randomization processes are employed. First, training samples for each individual tree are randomly selected by applying bootstrap sampling strategy. Second, rather than choosing the best split among all attributes, the tree inducer randomly samples a subset of the attributes and chooses the best one.



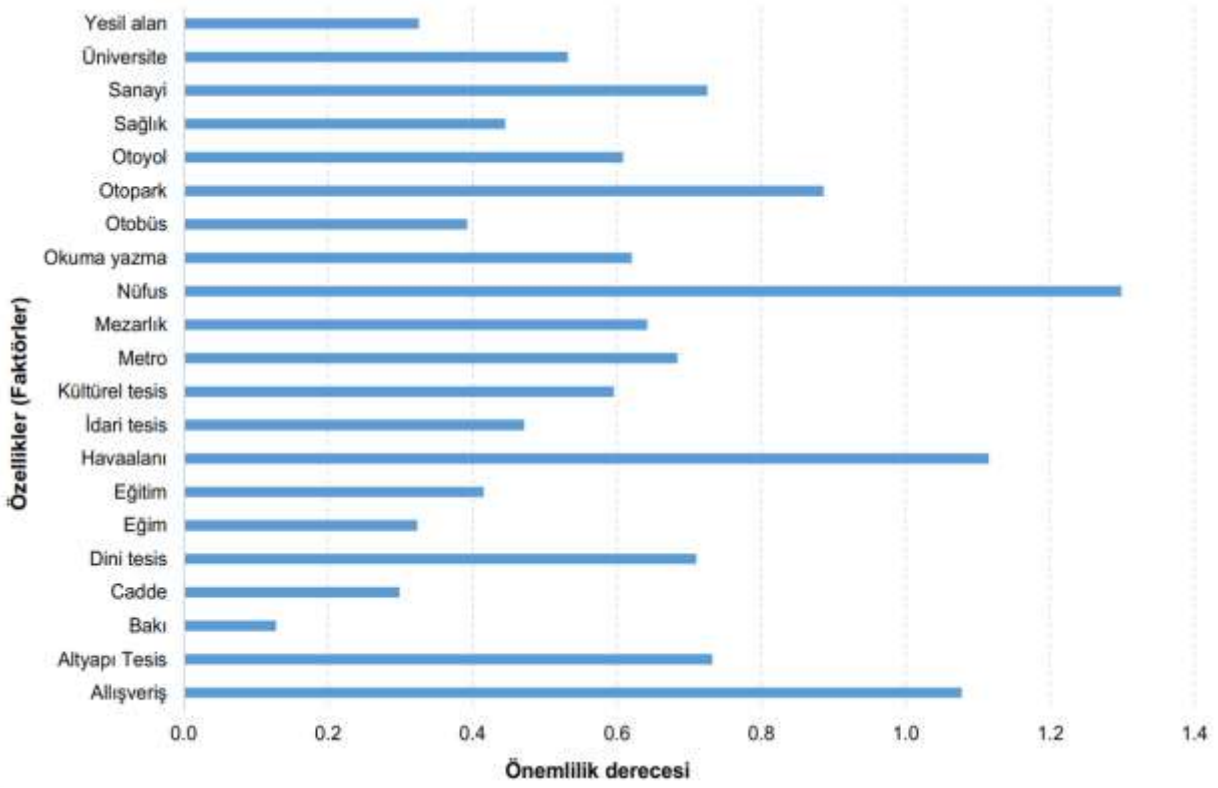
- Since the **RF algorithm is a supervised learning algorithm**, **up-to-date market value for 1.789 real-estate** located in Pendik district of Istanbul province were determined as ground-truth data.
- Then, the ground-truth dataset was randomly split into **70:30 ratio for construction (training) and validation of the model** respectively to evaluate the prediction performance of RF algorithm. As a result, 1.250 points were selected as a training set and 537 points were selected as a validation set.



- The input data set consisted of 20 factors, hence the **number of input variables (m) was set to be 5** (i.e. $k = \sqrt{m}$ variables at each split).
- **Out-of-bag (OOB) error results** of RF ensemble model was used **to determine the number of trees parameter.** For this purpose, input data set was firstly classified using **a large number of trees (i.e. 500 trees)** to estimate changes in OOB error with increasing number of trees. It was observed that there was a sharp decline in OOB error from 0.040 to less than 0.030 as number of tree increased from 1 to 150.
- After that, OOB error continued to decrease slightly until the number of trees takes value 250. From this critical point to larger tree sizes, OOB error stays stable. For this reason, **the number of trees (n) was set to be 250** for the current study.



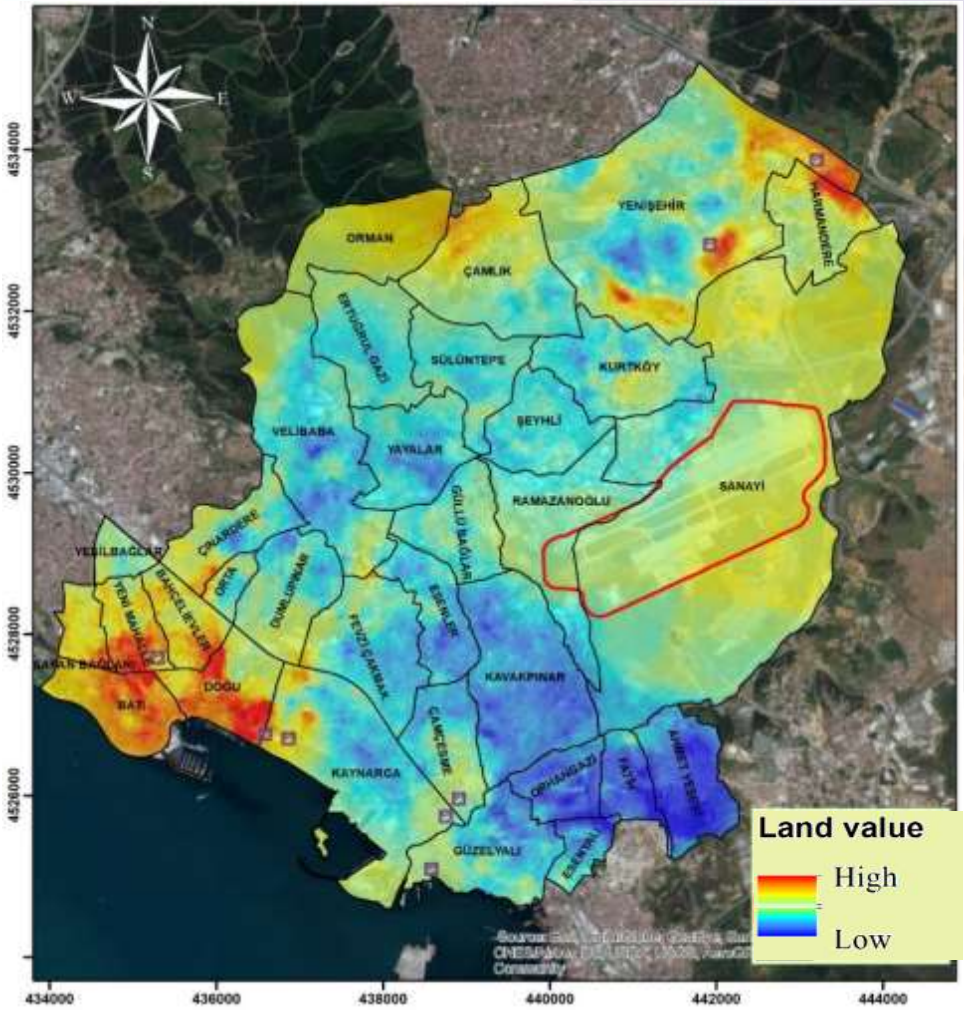
- RO represents **the combination of many decision tree classifiers**. In such structure, each decision tree is trained with randomly selected sample sets from the training data set. After the training of decision trees, some data are extracted from training data sets and the degree of importance of different properties is determined. Finally, **all properties in the data set are ordered by importance according to the Z value**.



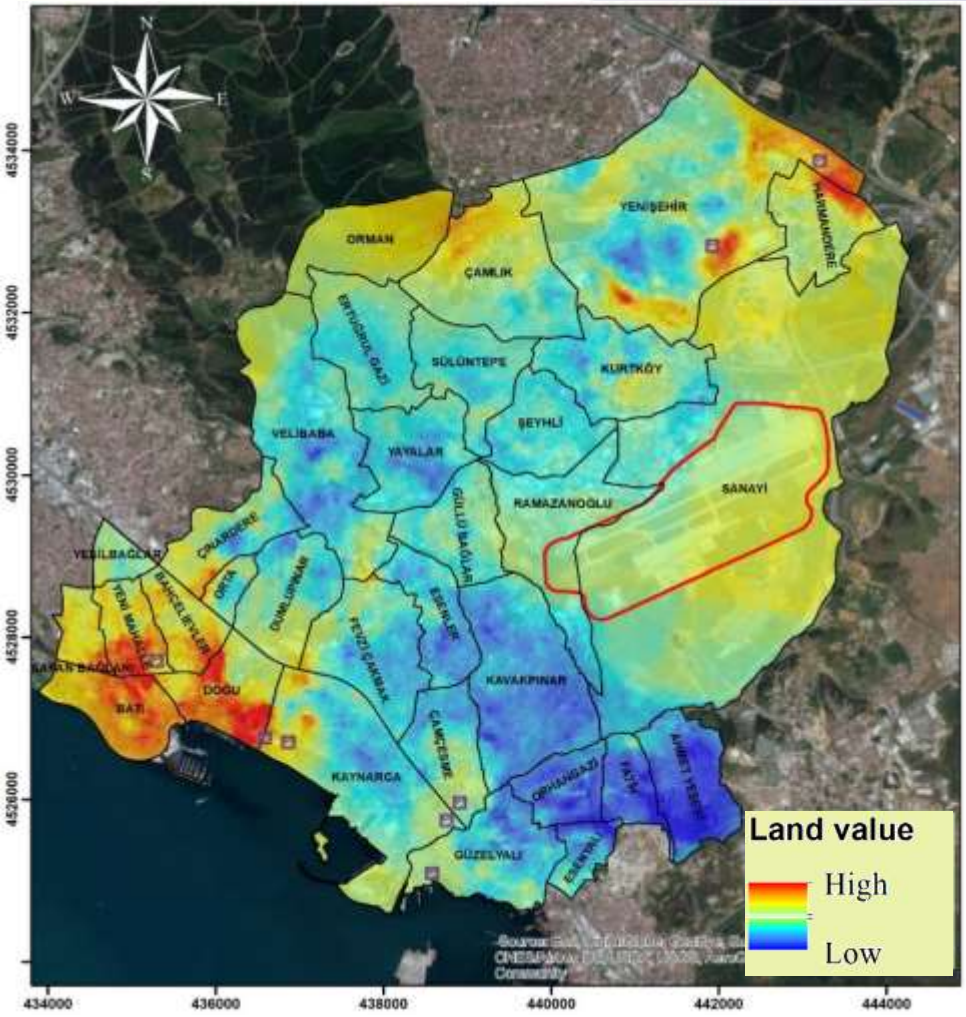
- The predictive accuracy of RF model constructed with the user-defined parameters was tested using;

Root-mean-square error (RMSE)=0.103
Mean-absolute-error (MAE)= 0.0758 and
Correlation coefficient **(CorC)= 0.799**

- The RF model constructed with optimal parameter setting applied to whole dataset to produce land valuation map of the study area. The resulting thematic map was produced for Pendik.



- When the produced thematic map was analysed, it was observed that the south-west part of the Pendik district especially the lands situated along the sea side were predicted as high land value.
- Furthermore, areas close to the shopping centre located to the north-east part of the study area were also predicted as zones with high value.



Conclusion

- The resulting thematic map produced with **RF algorithm demonstrates the model predictions (supervised)**, while thematic maps produced with **fuzzy logic approach demonstrate value trends in each thematic factor group (unsupervised)**.
- if you have ground truth value or real market value with geographic location, random forest can be used to predict real value of each pixel.
- According to the results, it can be said that **fuzzy logic and random forest algorithms with the help of GIS analysis** capabilities gives successful results and can be used **for mass real estate valuation.**
- Both of them can be programmed easily, it can be used together and seperately.
- You can produe valuation map more accurate with less parameter selections by users.

Thank you for your attention…

Prof. Dr. Arif Cagdas AYDINOGLU
Gebze Technical University
Dept. of Geomatics Engineering 41400
Gebze/Kocaeli Turkey
Email: aydinoglu@gtu.edu.tr
Web: arifcagdas@com
Twitter: https://www.twitter.com/arifcagdas