New Developments and Challenges in Remote Sensing, Z. Bochenek (ed.)

Investigation of water quality parameters by using multiple regression and fuzzy logic in the Istanbul Strait, Turkey

B. Ustun
Istanbul Technical University, Remote Sensing Division, Maslak-34469, Istanbul, Turkey; ustunbe@itu.edu.tr

S. Ekercin
Aksaray University, Engineering Faculty, Department of Geodesy and Photo., Aksaray, Turkey; ekercin@itu.edu.tr

A. Oztopal
Istanbul Technical University, Meteorology Department, Maslak-34469, Istanbul, Turkey;oztopal@itu.edu.tr

Keywords: remote sensing, fuzzy logic, water quality, regression analysis

ABSTRACT: This study focuses mainly on the water quality modelling through remotely sensed imagery and ground data around Istanbul, Turkey. Water quality models are firstly developed by using fuzzy logic for the assessment of water quality parameters together with spectral reflectance. The proposed methodology is compared with the already available classical multiple regression (MR) technique model that is mostly used in many water quality studies. A Landsat-7 ETM satellite image collected on July 2, 2000 is selected as remotely sensed data source for water quality modelling in addition to almost simultaneously performed *in-situ* measurements. Water quality parameters are total suspended sediment (TSS) and chlorophyll-a (Chl-a). It is concluded that the fuzzy logic model yields more refined results than the multiple regression approach. As a result of this study, it is suggested that future water quality measurements should be supported by current satellite images. On the other hand, the number of hydraulic stations located in the Istanbul Strait and its vicinity should be increased with careful positioning of new stations, especially along the Black Sea coasts of Istanbul.

1 INTRODUCTION

Water quality is the general term that describes whether or not the surrounding environment may be endangered by pollutants in the water. From a historical perspective, poor water quality has not been a perceived problem before the industrial revolution and the rapid growth of cities. Since about the turn of the century,attempts at water quality control have concentrated on sewage and industrial discharges. These attempts require a number of water sampling from a set of conveniently distributed locations which are time consuming and require a lot of field and laboratory efforts in addition to budget restriction problems. At this point, it can be mentioned that remote sensing has an important and effective role in water quality management, processing time reduction and requirement of several ground measurements (GM).

Remote sensing satellites measure the amount of solar radiation at various wavelengths reflected by surface water. In this procedure, the optical properties (i.e. reflectance) of water depend on the concentration and character of water quality parameters. In water quality studies, measured solar radiation and *in-situ* measurements are correlated to develop water quality models by using different methods, which are almost invariably multiple regression (MR) procedure with a set of restrictive assumptions and problems in model structure and model parameter estimations. Provided that the assumptions are satisfied in practice and the time is available, it is an effective and economic approach for estimation of water quality over large areas, especially when the spatial coverage of satellite image is considered. Hence, it is possible to reduce the number of *in-situ* measurements needed and increases the spatial and temporal coverage of the estimates (V). If any one of the assumptions is not valid for the problem at hand then the use of MR model yields to over or under estimations.

The main purpose of this paper is to investigate and present the feasibility of MR analysis in addition to the fuzzy logic (FL) for water quality modelling in and around the Istanbul Strait. The consistency of these two established models are interpreted comparatively for water quality parameters, namely, total suspended sediment and chlorophyll-a.

2 METHODOLOGY

2.1 *Study Area*

The study area is located in the north western part of Turkey and is known as the area where the two continents Asia and Europe meet each other. This area is called Bosphorus. This is a Thracian word for the Istanbul Strait. One of the world's most strategic waterways, Bosphorus is the strait between the Black Sea and the Sea of Marmara; it is an inundated valley that follows an irregular northeast-southwest course 30 km long, 700–3700 m wide, 30–120 m deep. It is a former river valley which was drowned by the sea at the end of the Tertiary period. This is a very busy strait with many ships and oil tankers, as well as local fishing and passenger boats. The shores of the strait are heavily populated as the city of Istanbul (in excess of 11 million inhabitants) straddles it.

The current flows north to south; however, a strong subsurface countercurrent with numerous points and coves sets up swirls and eddies that make navigation dangerous to the inexperienced.

The location of the study area can be given with the geographic coordinates shown in Figure 1.

3 MATERIALS

In this study, Landsat-7 ETM image acquired on July 2, 2000 is used as the satellite image data for water quality modelling. The Satellite image is acquired

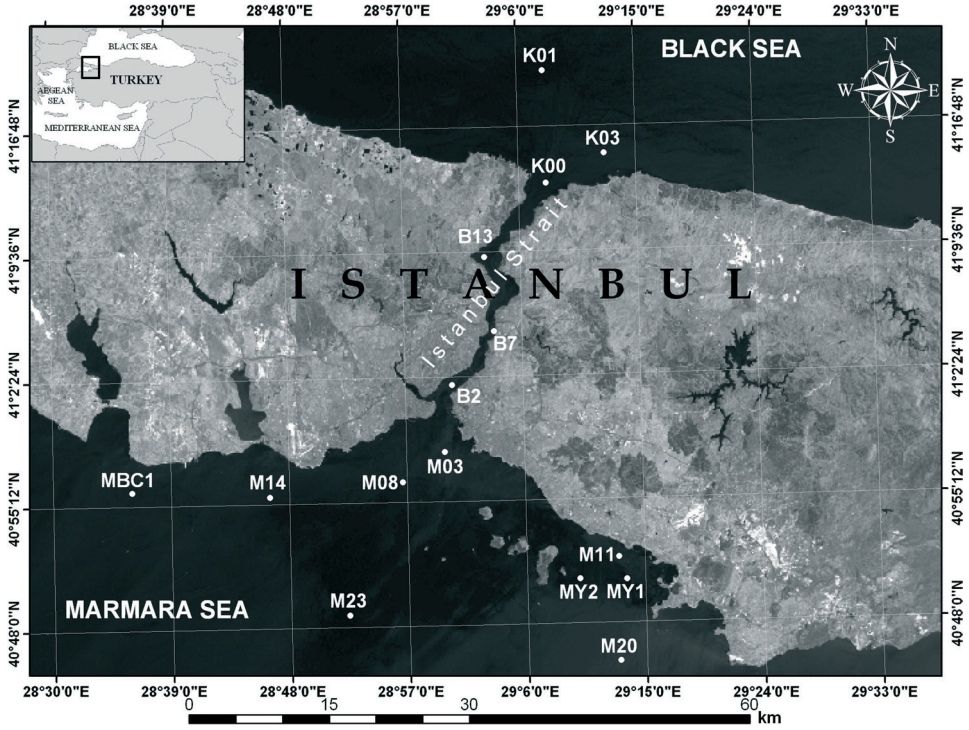

Figure 1. Location of the study area.

contemporaneously with the field data (X). Table 1 presents some basic information on the Landsat-7 ETM image selected for the study.

Additionally, topographic maps in the scale of 1:25000 produced by General Command of Mapping are used to get ground control points (GCPs) for the geometric correction process. These maps are compiled with NATO level A standards (III, IV).

On the other hand, the ground data are provided by Istanbul Water and Sewerage Administration Authority (İSKİ). *In-situ* measurements for water quality monitoring are carried out by Institute of Marine Science and Management, University of Istanbul, in

Table 1. Satellite data selected for this study.

Sensor	Date	Band (μm)	Spatial Resolution (m)	Path/Row
Landsat-7 ETM	July, 2000	1: 0.45–0.52	30	180/31
		2: 0.52–0.60	30	
		3: 0.63–0.69	30	
		4: 0.76–0.90	30	
		5: 1.55–1.70	30	
		6: 10.4–12.5	30	
		7: 2.08–2.35	30	
		Pan: 0.52–0.90	15	

the first week of July, 2000 in the name of İSKİ (VI). The total number of water sampling sites is 15 and they are located in an area of 60×60 km 2 from the southern part of the Black Sea to the northern part of the Marmara Sea through the Istanbul Strait (Figure 1).

4 IMAGE PROCESSING

Image processing and interpretation processes are performed in the laboratory of Remote Sensing Division at the Istanbul Technical University by using ERDAS Imagine software package. A regression adjustment technique for atmospheric correction (haze reduction) is used to correct the atmospheric errors in the satellite image. The atmospheric haze correction is based on linear regression of each band from TM band 1 to 7 using brightness values from an area of homogeneous depth and non-turbid water (II).

During the geometric correction process, the radio-metrically corrected Landsat TM scene is geo-referenced to a common topographic base in a UTM coordinate system (Zone 35) by using 80 ground control points. As the study area includes statistical analysis of water surface reflectance, a first-order polynomial transformation and nearest neighbour re-sampling methods are used to create the output images with 30 m ground resolution. The root mean square error of the polynomial transformation is less than a half pixel (XI). Check points are measured on the geometrically corrected image set and compared with the topographic sheets to evaluate the quality of the geometric correction (X).

5 WATER QUALITY MODELLING

5.1 *Multiple Regression (MR)*

In this study, the average digital number of pixels (a 3×3 window) surrounding the sample pixel is used to develop water quality models. Following multiple regression equation is widely used in water quality researches (I, *V)*.

$$WQP = a + b * ETM1 + c * ETM2 + d * ETM3 + e * ETM4 \quad (1)$$

Where; WQP is the water quality parameter as dependent variable with four independent variables ETM1, ETM2, ETM3 and ETM4, which are the digital numbers in Landsat-7 ETM image data from band 1 to band 4 as presented in Table 2. Finally, there are five model parameters as a, b, c, d and e which are unknowns. They must be predicted from a given set of data. The set of restrictive assumptions concerning the regression models is already presented by Şen (XIII). Fuzzy Logic (FL)

The concept of fuzzy logic was conceived by Lotfi Zadeh (XIV), a professor at the University of California. The real world simply does not map well to binary distinctions, and numerical precision is often unhelpful in making qualitative statements. Fuzzy logic gives us a way to deal with such situations, and is closer to natural language; therefore,

Table 2. Digital numbers and averages (3 by 3 windows) at the stations obtained from Landsat-7 ETM, 2000 image data.

Station Name	Geographic Location	ETM1 0.45-0.52 μm		ETM2 0.52-0.60 μm		ETM3 0.63-0.69 μm		ETM4 0.76-0.90 μm	
		DN	Average	DN	Average	DN	Average	DN	Average
K01	Black Sea	75	75,84	42	43,32	26	28,88	12	11,04
K03	Black Sea	90	88,12	57	57,00	43	42,92	18	18,16
K00	Black Sea	81	81,88	52	50,12	35	34,20	14	14,00
B13	Istanbul Strait	85	86,44	56	56,40	40	42,16	17	17,88
B07	Istanbul Strait	90	91,76	60	61,72	48	50,32	20	21,48
B02	Istanbul Strait	89	86,72	60	57,56	46	44,44	20	18,72
M03	Marmara Sea	93	89,76	60	59,76	46	48,84	19	20,24
M08	Marmara Sea	85	85,72	55	55,64	44	43,52	17	17,84
M14	Marmara Sea	81	83,20	52	50,78	38	39,04	17	15,96
M23	Marmara Sea	78	73,84	45	43,76	29	29,32	12	11,04
M20	Marmara Sea	93	94,32	66	66,12	60	57,20	25	24,92
M11	Marmara Sea	79	79,84	46	49,68	42	36,52	15	15,04
MY2	Marmara Sea	77	76,56	47	46,88	34	33,48	14	13,20
MY1	Marmara Sea	73	73,56	43	43,04	31	29,60	12	11,80
MBC1	Marmara Sea	85	82,04	52	53,76	42	41,04	16	16,16

this allows us to generate fuzzy statements through mathematical calculations. Zadeh introduced linguistic variables as variables whose values are sentences in a natural or artificial language. In both contexts, fuzzy propositions, i.e. IF- THEN statements, are used to characterize the state of a system and the truth value of the proposition is a measure for how well the description matches the state of the system. Unfortunately, the fuzzy concept was not welcome in the literature since many uncertainty techniques such as probability theory, statistics and stochastic processes were commonly employed at that time. But fuzzy logic has been developing since then, and is now being used especially in Japan for automatic control of commercial products such as washing machines, cameras and robotics. Many textbooks provide basic information on the concepts and operational fuzzy algorithms (VII, VIII, IX, XII).

In this study, the program named MATLAB was used for fuzzy model. As it is seen in Figure 2, every input has three membership functions which are triangle. Backpropagation in ANFIS is selected as the training method and number of epochs is 100.

6 RESULTS AND DISCUSSION

In the applications both the multiple regression and fuzzy logic models are adopted as having four input (independent variables) and one output (dependent, prediction) parameter, WQP. Table 3 shows constant coefficients (a, b, c, d, e) and determination coefficients of multiple regression models for each parameter. In Table 4, there are outputs of multiple regression and fuzzy logic models with ground measurements. It is obvious that the fuzzy logic model predictions of WQP (calculated values) are

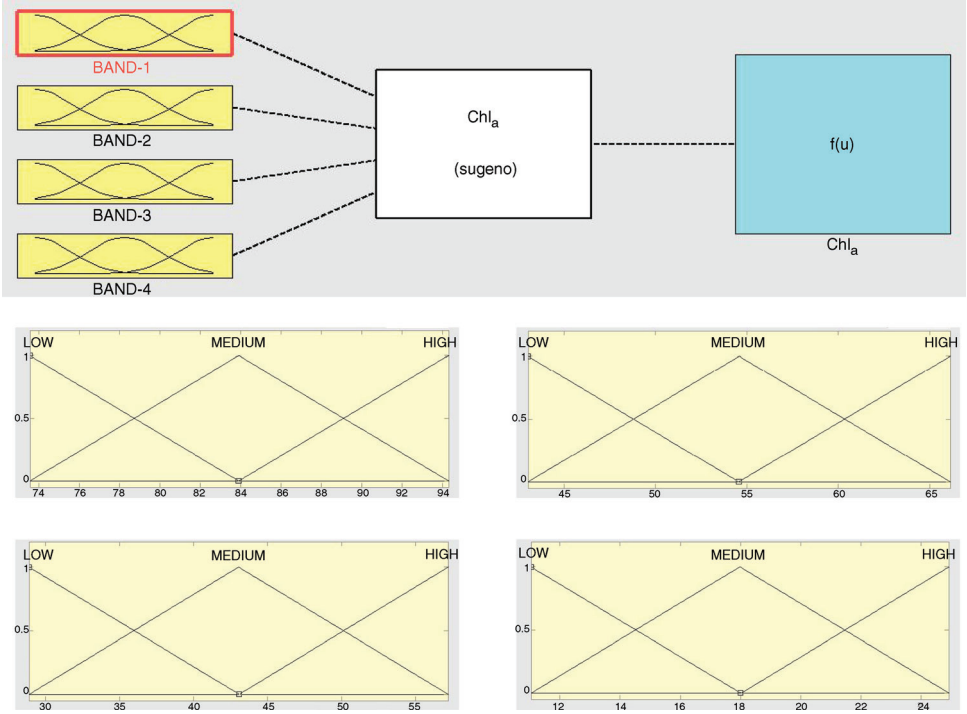

Figure 2. Membership functions for inputs (spectral bands).

comparatively better than the multiple regression results (Figure 3). This point is clear from the closeness of the scatter points around the straight-line; where in the case of fuzzy logic model the deviations from the straight-line are smaller.

This is the reason why in all the cases of fuzzy logic the coefficient of determination, R^2, is invariably larger than the corresponding multiple regression counterparts. According to the selected two water quality parameters determination coefficients for multiple regression are 0.2889 for total suspended sediment and 0.7391 for chlorophyll-a, respectively, whereas these values are 0.7036 and 0.9570 for fuzzy logic model.

Table 3. Multi regression analysis of parameters and spectral reflectance

Parameters of water quality	Regression equations	Determination Coefficients (R^2)
Total Suspended Sediment (TSS)	TSS = 281,7741 − 5,2418*ETM1 + 1,7266*ETM2 + 2,0578*ETM3 + 0,5478*ETM4	0,2889
Chlorophyll-a (Chl-a)	Chl-a = 117,8677 − 1,9259*ETM1 − 0,0298*ETM2 −0,7081*ETM3 + 4,5710*ETM4	0,7391

Table 4. The ground measurements and calculated values from Landsat ETM data by using multi regression and fuzzy logic methods.

	Total Suspended Sediment (TSS) (mg/L)			Chlorophyll-a (Chl-a) (μg/L)		
Station	GM	MR	FL	GM	MR	FL
K01	15,00	24,51	19,2	0,30	0,53	−0,17
K03	18,00	16,55	21,2	0,40	−0,93	0,03
K00	8,60	17,16	19,6	0,40	−1,54	0,32
B13	22,00	22,60	33,9	0,50	1,59	0,46
B07	14,20	22,66	12,4	0,70	1,86	0,74
B02	45,70	28,29	29,6	0,80	3,24	0,52
M03	9,00	26,04	13,5	0,70	1,15	0,53
M08	44,90	27,84	42,8	1,70	1,85	1,02
M14	40,00	22,41	38,9	1,20	1,43	1,51
M23	28,60	36,66	29,3	1,20	4,06	2,76
M20	23,10	32,88	23,2	8,60	7,65	8,57
M11	44,30	32,43	30,1	5,80	5,51	4,83
MY2	28,80	37,53	26,1	4,40	5,65	5,15
MY1	36,10	37,87	32,4	10,10	7,89	8,68
MBC1	45,00	37,86	46,7	6,20	3,07	5,92

(GM: Ground Measurement, MR: Multi Regression, FL: Fuzzy Logic.)

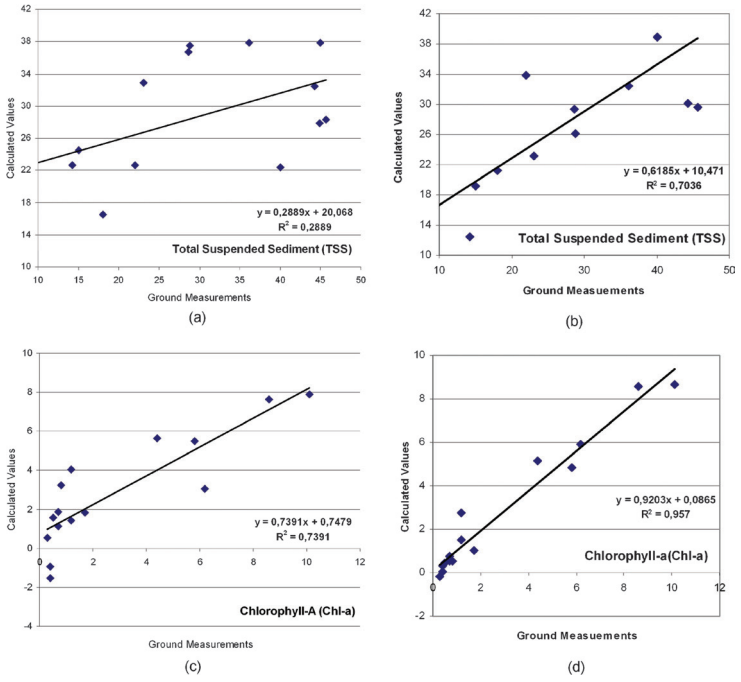

Figure 3. Resulting algorithms developed by using multiple regression (MR) and fuzzy logic (FL) methods.

7 CONCLUSIONS

One of the most important advantages of remote sensing is that it is a reliable and economical data source for research applications including large areas. At this junction, this study has become a good and explanatory example with a study area of approximately 3,700 km^2. The results show that water quality modelling via fuzzy logic yields more accurate results compared to the multiple regression method for the same case study.

It is suggested that present water quality measurements should be supported by current satellite images and the number of hydraulic stations located in the Istanbul Strait and especially in the Black Sea coasts of Istanbul and its surroundings should be increased.

ACKNOWLEDGEMENTS

The authors appreciate the support provided by the Istanbul Water and Sewerage Administration Authority (İSKİ) and the project team of the Institute of Marine Science and Management, University of Istanbul. Also, the authors would like to thank to Prof. Dr. Zekai Sen for his generous supports.

REFERENCES

Berka, C., Schreier, H. and Hall, K. 2001. Linking water quality with agricultural intensification in a rural watershed. Water, Air, and Soil Pollution, 127, 389–401.

Bilge, F., Yazici, B., Dogeroglu, T. and Ayday, C. 2003. Statistical evaluation of remotely sensed data for water quality monitoring. International Journal of Remote Sensing, 24, 5317–5326.

Coskun, G., Ekercin, S., Öztopal, A. 2001, ''Monitoring and water pollution modelling of the Bosphorus by regression analysis using Landsat-TM data'', EARSeL 21thSymposium Observing our Environment from Space/New Solutions for a New Millennium , ENSG, 14–16 May, Marne-la-Vallee (Paris), France.

Ekercin, S. 2006, Coastline Change Assessment at the Aegean Sea Coasts in Turkey Using Multitemporal Landsat Imagery. *Journal of Coastal Research (in press)*.

Hellweger, F.L., Schlosser, P., Lall, U., Weissel, J.K. 2004. Use of satellite imagery for water quality studies in New York Harbor. *Estuarine, Coastal and Shelf Science*, 61, 437–448.

İSKİ (General Directorate of Istanbul Water and Sewerage Administration, Turkey), 2001. Water Quality Monitoring, Third Progress Report. Project Team: *Institute of Marine Science and Management, University of Istanbul*. January-2001, Istanbul, Turkey, 150 p.

Klir, G.J., Fogel, T.A. 1988. ''*Fuzzy Sets*, Uncertainty and Information'', Prentice Hall, New York.

Kosko, B. 1992.''*Fuzzy Thinking: The new Science of Fuzzy Logic*'' Hyperion, New York

McNEİL F.M., Thro, E. 1994. ''*Fuzzy Logic: A practical Approach*'', AP Professional, Boston.

Metternicht, G.I. and Zinck, J.A. 1998. Evaluating the information content of JERS-1 SAR and Landsat TM data for discrimination of soil erosion features. *International Journal of Photogrammetry and Remote Sensing*, 53, 143–153.

Ormeci, C. and Ekercin, S. 2005. Water Pollution Assessment at the Salt Lake by using Remote Sensing Technique. 10th Map Meeting, 28 March-01 April, Ankara, Turkey (in press).
Şen, Z. 1998, ''Fuzzy Algorithm For Estimation Of Solar Irradiation From Sunshine Duration'', *Solar Energy*, 63, 39–49
Şen, Z. 2001, Angstrom equation parameter estimation by unrestricted method. Solar Energy 71 (2): 95–107.
Zadeh, L. 1967, ''*Fuzzy Sets*'', Information and Control, 8, 38–53.