International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:4, No:12, 2010

Combining Fuzzy Logic and Data Mining to Predict the Result of an EIA Review

Kevin Fong-Rey Liu, Jia-Shen Chen, Han-Hsi Liang, Cheng-Wu Chen and Yung-Shuen Shen

***Abstract*—**The purpose of determining impact significance is to place value on impacts. Environmental impact assessment review is a process that judges whether impact significance is acceptable or not in accordance with the scientific facts regarding environmental, ecological and socio-economical impacts described in environmental impact statements (EIS) or environmental impact assessment reports (EIAR). The first aim of this paper is to summarize the criteria of significance evaluation from the past review results and accordingly utilize fuzzy logic to incorporate these criteria into scientific facts. The second aim is to employ data mining technique to construct an EIS or EIAR prediction model for reviewing results which can assist developers to prepare and revise better environmental management plans in advance. The validity of the previous prediction model proposed by authors in 2009 is 92.7%. The enhanced validity in this study can attain 100.0%.

***Keywords*—**Environmental impact assessment review, impact significance, fuzzy logic, data mining, classification tree.

I. Introduction

ENVIRONMENTAL impact assessment (EIA) stands for an environmental management plan based on scientific, objective and comprehensive surveys, forecasting, analyses and evaluations conducted prior to project implementation in order to determine the degree and scope of the potential impact of development activity or government policy on the environment, society, economy, culture and ecology, and the public explanation and review of such a plan. In Taiwan, development projects for which there is concern of adverse impact on the environment should prepare an environmental impact statement (EIS) for the phase-I EIA, and then transfer the EIS to the competent authority for review. The developer should edit an environmental impact assessment report (EIAR) for the phase-II EIA for those circumstances in which the review result of the EIS is concerned with a significant impact on the environment. The review results of EISs or EIARs can be classified into three categories: conditional approval, phase-II EIA, or disapproval on the development project. A major concern for developers is the possible review results of projects.

The prediction of a review results is feasible if there are sufficient previous review cases. From 1981 to present, the Taiwan EPA has been collecting EISs and EIARs and their associated review results and opinions, and disclosing these documents to the public on an online basis. Based on these past cases, the authors and their colleagues have proposed an integrated prediction model consisting of case-based reasoning and fuzzy logic to qualitatively forecast possible review results, which presents an overall prediction validity of 92.7% [5]. Our previous attempt was unable to obtain higher validity because the model is based on past EISs and EIARs. In fact, the environmental information provided in EISs and EIARs is primarily related to scientific facts (magnitudes of impacts induced by a development project); however, the review process can be viewed as highly subjective judgment because it has to ruminate over the scientific facts and subjective values (judgment, preference, value and concern). More specifically, review work determines the significance of impacts [1][2][3][4][5][6][7][8].

This paper intends to integrate two AI techniques to predict EIA review results with higher validity: (1) fuzzy logic for evaluating impact significance, (2) significance transformation for incorporating significance thresholds, and (3) data mining for predicting EIA review results.

II. Environmental Factors and Their Significance Criteria

The indicators considered in EIA for road construction projects, as listed in TABLE I **(**Liu et al., 2009b**)**, contains air (I_1), water (I_2), soil (I_3), solid waste (I_4), noise (I_5), terrestrial (I_6), aquatic (I_7), economics (I_8), society (I_9) and culture (I_{10}).

Sensitivity, spatial extent, mitigation measure reliability and information integrity are four major significance criteria used by the EIA review committee for reviewing road construction projects. These criteria are extracted from the EIA review opinions of 63 real cases from 1994 to 2009. The four criteria are discussed 28, 3, 43 and 84 times respectively within the period and their distributions over the ten indicators and their standard values (SV) are also shown in TABLE II.

III. Prediction Model

The concept of impact significance underlies an integrated prediction model of EIA review results of road construction

Department of Safety, Health and Environmental Engineering, Ming Chi University of Technology, Taipei, Taiwan 24301, ROC (tel.: +886-2 29089899 4697 ; fax: +886-2 29041914; e-mail: kevinliu@mail.mcut.edu.tw).
Department of Environmental Engineering, Da-Yeh University, Changhua, Taiwan 51542, ROC
Department of Architecture, National United University, Miaoli, Taiwan 36003, ROC
Institue of Maritime Information and Technology, National Kaohsiung Marine University, Kaohsiung 80543, Taiwan, ROC
General Education Center, Mackay Medical College, Taipei, Taiwan 25245, ROC

International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:4, No:12, 2010

TABLE I
FACTORS FOR EIA AND INFORMATION OF STUDY CASE

Indicator	Subindicator	Unit	SV	Study Case		
				BC	PIWOM	PIWM
Air (I_1)	CO (I_{11})	ppm	35.0	0.7	0.9	0.8
	SO_2 (I_{12})	ppb	250.0	6.8	39.5	16.6
	NO_2 (I_{13})	ppb	250.0	24.5	94.6	66.5
	TSP (I_{14})	μg/m^3	250.0	71.8	136.6	97.7
Water (I_2)	DO (I_{21})	mg/L	6.5	7.1	7.1	7.1
	BOD (I_{22})	mg/L	3.0	4.1	4.1	4.1
	SS (I_{23})	mg/L	20.0	5.3	5.3	5.3
	NH_3-N (I_{24})	mg/L	0.5	4.7	4.7	4.7
Soil (I_3)	Heavy metal (I_{31})	0-100	100.0	66.9	66.9	66.9
Solid waste (I_4)	Rubbish (I_{41})	0-100	20.0	4.0	9.0	5.0
	Construction waste (I_{42})	0-100	30.0	0.0	33.0	3.0
Noise (I_5)	Noise (I_{51})	dB	60.0	59.8	71.5	68.0
	Vibration (I_{52})	dB	60.0	33.2	60.1	52.0
Terrestrial (I_6)	Threatened terrestrial animals (I_{61})	%	20.0	3.0	20.0	10.0
	Threatened terrestrial plants (I_{62})	%	20.0	3.0	20.0	10.0
	Threatened endangered terrestrial species (I_{63})	%	10.0	1.0	3.0	2.0
Aquatic (I_7)	Threatened terrestrial animals (I_{71})	%	20.0	15.0	20.0	18.0
	Threatened terrestrial plants (I_{72})	%	20.0	10.0	15.0	12.0
	Threatened endangered terrestrial species (I_{73})	%	10.0	0.0	0.0	0.0
Economics (I_8)	Land-use and development obstacle (I_{81})	0-100	20.0	2.0	3.0	3.0
	Life-quality decline (I_{82})	0-100	10.0	2.0	3.0	3.0
	Economic activity disturbance (I_{83})	0-100	10.0	0.0	1.0	1.0
Society (I_9)	Public facility inaccessibility (I_{91})	0-100	20.0	0.0	0.0	0.0
	Transportation inaccessibility (I_{92})	0-100	50.0	10.0	50.0	30.0
	Community disconnection (I_{93})	0-100	20.0	5.0	10.0	6.0
Culture (I_{10})	Cultural heritage destruction (I_{101})	0-100	20.0	0.0	0.0	0.0
	Landscape demolition (I_{102})	0-100	20.0	3.0	10.0	8.0

Note: SV: standard value; BC: baseline condition;
PIWOM: predicted pollution increment without mitigation measures;
PIWM: predicted pollution increment with mitigation measures.

TABLE II
SIGNIFICANCE CRITERIA AND THEIR COUNTS IN 63 ROAD CONSTRUCTION EIA REVIEW CONCLUSIONS FROM 1994 TO 2009

Indicator	Sensitivity		Spatial extent		Mitigation measure liability		Information integrity	
	Count	SV	Count	SV	Count	SV	Count	SV
Air (I_1)	2	30	1	20	3	70	6	70
Water (I_2)	9	30	0	20	8	70	9	70
Soil (I_3)	1	NA	0	0	3	70	3	70
Solid waste (I_4)	1	NA	0	NA	9	70	17	70
Noise (I_5)	5	25	1	20	5	70	8	70
Terrestrial (I_6)	3	0	0	10	3	70	9	70
Aquatic (I_7)	4	0	0	10	2	70	9	70
Economics (I_8)	1	0	0	NA	2	70	3	70
Society (I_9)	1	NA	0	NA	3	70	14	70
Culture (I_{10})	1	0	1	NA	5	70	6	70
Total count	28		3		43		84	

projects. Furthermore, the example of air pollution is depicted in Fig. 1. As discussed in previous section, significance is a complex concept that relates to not only impact magnitude but also other considerations and its evaluation may be viewed as a highly subjective judgment. Therefore, in part (a) of Fig. 1, fuzzy logic is employed to infer significances because it can imitate experts' thinking process. It should be noted that the level of impact significance is expressed as a percentage of thresholds, ranging from 0% (insignificant) to 200% (completely significant), for the sake of incorporating legal requirements, ecological and social-economic tolerance standards. The determination of impact significances is developed into three tiers. The first tier produces first-order significance which mainly evaluates the impact magnitudes of an indicator according to the pollution levels of its subindicators. For example, as shown in Fig. 1, air pollution evaluation refers to the appraisal of emission of carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2) and total suspended particulates (TSP). Second-order significance is an outcome of simultaneous consideration of baseline condition (BC), predicted pollution increment without and with mitigation measures (PIWOM and PIWM). The first and second-order significance are evaluated based on the objective measurements or predictions; however, the last tier, termed third-order significance, merges significance criteria which are extracted from the opinions given in EIA reviews. The second part of the model predicts EIA review results according to impact significance. A data mining technique is adopted for the task of prediction. First of all, the third-order significances of all EIA cases from 1992 to 2009 (i.e. 50 cases gained conditional approval, 4 cases were phase-II and another 9 cases gained disapproval) are used to derive a classification tree. This classification tree is then exploited to predict the possible review result for a new case, as shown in part (b) of Fig. 1.

The software used to construct the integrated prediction model included the MATLAB Fuzzy Logic Toolbox (The Mathworks Inc. USA) and the Data Mining Tools See5 (RuleQuest Research Pty Ltd, Australia). The graphical user interfaces editors and viewers in the MATLAB Fuzzy Logic Toolbox make users easy to build the rules, define the membership functions, and analyze the behavior of a fuzzy

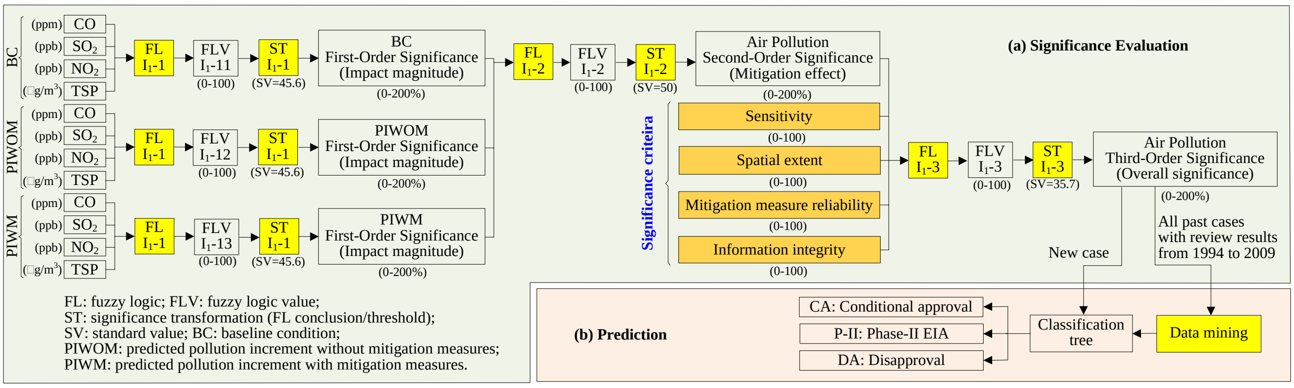

Fig. 1 Integrated prediction model for EIA review results of road construction projects (air pollution).

inference systems. It was used to develop the 21 rulebases containing 2,302 fuzzy rules and their corresponding fuzzy inference systems for the purpose of evaluating impact significance. The 21 fuzzy inference systems is available free of cost from the corresponding author. See5 is also easy to use and assist authors to establish the 30 classification trees for predicting review results.

IV. Case Study

A. Case Description

Taiwan High-Speed Rail (THSR), commencing on January 5, 2007, is capable of running at up to 300 kilometers per hour and travels from Taipei City to Kaohsiung City in about 90 minutes, compared the 4.5 hours spent by trains on the conventional western trunk line of the Taiwan Railway Administration. It passes 13 major cities and runs 344.68 kilometers, including about 252 kilometers of overpasses. The spaces under the viaducts were designed as part of the transportation network and were successively constructed as connections between cities and THSR stations. As shown in Fig. 6, the case study is a 12.147-kilometer extension of provincial highway no.31, which runs beneath the THSR bridges and is divided by piers as north-south bounds. This project intends to minimize the cost of land purchase and the demolition of existing buildings, and it also intends to achieve the goal of easing the traffic burden between Taoyuan and Hsinchu counties.

For preventing a lateral impact on the adjacent environment along the area of the case study within the construction and operation stages, an environmental impact statement (EIS) concerning the natural, biological, social and economical impacts was submitted to the review committee in July 2009 and ultimately conditionally approved in December of the same year. According to the information provided in this EIS, as summarized in the last three columns of TABLE I, the prediction model of the review result proposed in this paper demonstrates its use. In TABLE I, three conditions are discussed: the baseline condition (BC) before construction, prediction of the impact without mitigation measures (PIWOM) and prediction of the impact with mitigation measures (PIWM). Indicators I_1, I_2, I_3 and I_5 take the average of all measurement points and the others are subjectively estimated by experts as follows.

The case study is located in rural areas which are made up of mostly agricultural land, ponds and a few residences. Thereby, the rubbish load (I_{41}) is minor for BC and is rated 4.0; for PIWOM, about 200 construction workers generate rubbish in the amount of 0.2 ton per day and thus, the I_{41} score is 9.0; the mitigation measure deals with the waste by government cleaning units and enables I_{41} to decrease to 5.0 in PIWM. As for construction waste (I_{42}), BC does not produce industrial waste, therefore, it is rated as 0; in PIWOM, a large number of excavations are generated, so I_{42} waste is estimated at 33.0. After the mitigation measures for PIWM, almost all the excavations are reused on site, causing I_{42} to decrease to 3.0. As the terrestrial animals (I_{61}) in these areas include common native species and the terrestrial plants (I_{62}) are mostly crops along the projected line, the impact of existing THSR on the surrounding terrestrial is minor, with about 3% of them being affected in BC. The construction will have a significant impact on them, about 20% in PIWOM. After taking some mitigation measures such as reducing air pollution to not cover surface of the plan leaves or reducing noise and vibration as to avoid disturbance of the nearby wildlife habitats, the affected percentage can be brought down to 10% in PIWM. The surrounding endangered species (I_{63}) are myna, falco tinnunculus, serpent eagle and brown shrikes. They make up about 9.3% of all species. The impact of the existing THSR on these species is minor, about 1.0% in BC. The construction will not influence them much because it is certain distance away, about 3.0% in PIWOM. Through the mitigation measures, for

TABLE III
SIGNIFICANCE EVALUATION AND REVIEW RESULT PREDICTION FOR STUDY CASE

Indicator	First-order significance			Second-order significance	Significance criteria				Third-order significance	Review result prediction
	BC	PIWOM	PIWM		Sensitivity	Spatial extent	Mitigation measure reliability	Information integrity		
Air (I_1)	25.3%	35.6%	29.1%	45.6%	60.0	0	95.0	100.0	30.8%	CA (< 33.0%)
Water (I_2)	115.2%	115.2%	115.2%	106.2%	30.0	100.0	85.0	85.0	80.9%	CA (< 86.0%)
Soil (I_3)	73.9%	73.9%	73.9%	87.4%	NA	0	100.0	100.0	31.3%	CA (< 49.4%)
Solid waste (I_4)	18.8%	97.6%	35.6%	66.6%	NA	NA	90.0	95.0	51.5%	CA (< 65.1%)
Noise (I_5)	83.7%	105.2%	99.1%	101.6%	46.2	50.0	90.0	85.0	69.4%	CA (< 71.5%)
Terrestrial (I_6)	41.1%	74.6%	57.1%	77.4%	10	10.0	80.0	80.0	67.4%	CA (< 75.8%)
Aquatic (I_7)	54.5%	61.2%	58.5%	75.6%	5	10.0	80.0	85.0	66.7%	CA (< 75.8%)
Economics (I_8)	42.2%	45.1%	45.1%	60.4%	0	NA	100.0	100.0	27.8%	CA (< 52.1%)
Society (I_9)	47.4%	74.4%	57.8%	77.8%	NA	NA	100.0	95.0	47.8%	CA (< 49.6%)
Culture (I_{10})	38.6%	55.8%	51.6%	69.4%	0	NA	100.0	95.0	33.1%	CA (< 54.0%)

Note: BC: baseline condition; PIWOM: predicted pollution increment without mitigation measures;
PIWM: predicted pollution increment with mitigation measures NA: Not applicable; CA: Conditional approval..

example districted construction areas, the affected proportion may drop to 2.0% in PIWM. The case study crosses over four slightly polluted rivers and nine ponds which are already heavily or moderately polluted, therefore the aquatic animals (I_{71}) and plants (I_{72}) have been serious affected, about 15% and 10%, respectively, in BC. In PIWOM, the slight leakage of oil and sewage brings the scores up to 20% and 15%, respectively. Mitigation measures such as collecting sewage and controlling oil leaks appropriately can make the affected ratios fall to 18% and 12%, respectively, in PIWM. The scores are all 0% under any conditions for indicator I_{73} since there are no endangered species in these ponds. As the projected line runs through agricultural land with small population, land-use and development obstacle (I_{81}), life-quality decline (I_{82}) and economic activity disturbance (I_{83}) are only slight. Thus, these factors are rated as 2.0, 2.0 and 0, respectively, in BC. In PIWOM, these three indicators create a slight impact and are rated as 3.0, 3.0 and 1.0, respectively. Because of failure to take mitigation measures in PIWM, the scores are unchanged. There are no accessibility problems for public facilities around the project. Thus, indicator I_{91} is rated as 0 under any condition. In addition, the existing THSR disconnects transportation and communities slightly, so indicators I_{92} and I_{93} are assessed as 10.0 and 5.0, respectively, in BC. The construction will produce impacts on the above two indicators, and thus, the scores are 50.0 and 10.0, respectively, in PIWOM. Taking mitigation measures, such as traffic management, can make them decrease to 30.0 and 6.0, respectively, in PIWM. Indicator I_{101} is rated as 0 in any condition because the project is in an area with no cultural heritage. The existing THSR slightly damages the landscape and the indicator I_{102} is thereby assessed as 3.0 in BC. The construction will heavily change the landscape, making the score go up to 10.0 in PIWOM. Mitigation measures, such as beautification of construction fences and management of construction layout, can drop the score to 8.0 in PIWM.

B. Evaluation through Fuzzy Logic and Prediction by Classification Tree

The above-mentioned case information provides the basis for first-order significance assessment of the ten indicators in the three conditions (BC, PIWOM and PIWM), the results of which are listed in TABLE III. In this table, the expression of percentage, a ratio compared to standard values, is used to involve the concept of significance thresholds, as stated in Section 3.3. TABLE III shows, obviously, that first-order significances of water pollution (I_2) in all conditions are beyond 100%, indicating they are over-standard, as a result of the slightly polluted rivers and heavily polluted ponds. In addition, noise (I_5) in PIWOM is also over-standard due to excessive construction noise. The second-order significance, in the fifth column of TABLE III, merges the first-order significances of BC, PIWOM and PIWM to produce a percentage score which, similarly, expresses a proportion to standard values. As expected, water pollution (I_2) and noise (I_5) are also assessed to be over-standard (beyond 100%) according to second-order significance, which is represented by a single-underline in TABLE III.

The third-order significance mingles values with second-order significance. The values coming from the EIA review committee are referred to as significance criteria. These criteria include sensitivity, spatial extent, mitigation measure reliability and information integrity. On the basis of rating guidelines, the entire line of the study case is located in a class-III air pollution control region and therefore its air pollution sensitivity is 60.0, as shown in TABLE III. Water pollution has a sensitivity of 30.0 since it is located in a water pollution control zone. The case study obtains a sensitivity score of 46.2 for noise pollution, which is the average derived when crossing through two category-II and five category-III noise pollution control zones. Spatial extent refers to the proportion of measurement points which are over-standard. All rivers and ponds with at least slight pollution make the spatial extent of water pollution 100.0; noise pollution is over standard by 50.0% even in PIWM; terrestrial and aquatic are only over-standard by 10.0% of the spatial extent. As for the assessment of mitigation measure reliability and integrity, it requires professional and subjective judgments, the results of which are shown in Table 6. Compared to the predefined standard values of the four significance criteria, the over-standard scores are single-underlined in this table. The third-order significances, the penultimate second column of this table, are the final outcomes when combining these significance criteria with second-order significances; that is, they are 30.8%, 80.9%, 31.3%, 51.5%, 69.4%, 67.4%, 66.7%, 27.8%, 47.8% and 33.1%, respectively, for the ten indicators. Although several indicators such as air (I_1), water (I_2) and noise (I_5) are over-standard in second-order significance or significance criteria, however, all indicators are ultimately under-standard and are also below conditional approval splitting values in third-order significance. This is mainly because of the reliable mitigation measures and adequate information. Eventually, the review result for this case study is predicted to receive conditional approval (CA), as shown in the last column of TABLE III.

V. Conclusions

The purpose of predicting EIA review results is to help developers identify and deal with risky nuisances to the environment in their development projects, and then enable them to pass EIA reviews. The prediction can be achieved by a comprehensive use of several AI technologies, which is carried out through: (1) fuzzy logic for evaluating impact significance; (2) significance transformation for incorporating significance thresholds; and (3) data mining for predicting EIA review results. A case study of road construction in north Taiwan was used to illustrate the use of the integrated system. The major advantages of fuzzy logic are that it can model the human thought process in analyzing complex systems and decisions, and it allows a human expert to naturally express his knowledge. Therefore, it is easy to model experts' knowledge of significance evaluation. In this study, a total of 21 fuzzy logic systems containing 2,302 fuzzy rules were developed. For the case study, although several indicators such as air (I_1), water (I_2) and noise (I_5) are over-standard in second-order significance or significance criteria, the third-order significances of the ten indicators were under-standard and were estimated as 30.8% for air (I_1), 80.9% for water (I_2), 31.3% for soil (I_3), 51.5% for solid waste (I_4), 69.4% for noise (I_5), 67.4% for terrestrial (I_6), 66.7% for aquatic (I_7), 27.8% for economics (I_8), 47.8% for society (I_9) and 33.1% for culture (I_{10}), sequentially. Nevertheless, water (I_2), noise (I_5), terrestrial (I_6) and aquatic (I_7) are the four relatively significant indicators. The main benefits of classification trees are their abilities to explore the most relevant attributes and their corresponding splitting values at each node, and to easily and explicitly explain classification knowledge. Ten classification trees were established for the ten indicators, which present the CASVs for the ten indicators as 33.0%, 86.0%, 49.4%, 65.1%, 71.5%, 75.8%, 75.8%, 52.1%, 49.6% and 54.0%, respectively; the DASVs of the ten indicators are 108.7%, 152.1%, 116.6%, 149.0%, 141.9%, 170.3%, 159.7%, 130.8%, 116.6% and 134.3%, respectively. For the study case, all indicators were ultimately below CASVs in third-order significance and therefore the case was predicted to receive conditional approval (CA). The critical step of model development verifies feasibility and correctness. This study adopted six projects from the authors' previous work to test the validity of the prediction model. The validity of the previous prediction model is 92.7% and it can be enhanced to 100.0% in this paper. The genuine goal of assessment is to make improvement. If mitigation measures and information integrity is improved for this study case thereby the predicted significant indicators water (I_2), noise (I_5), terrestrial (I_6) and aquatic (I_7) can be largely improved 29.7%, 36.8%, 63.7 and 64.5%, respectively.

References

[1] R.B. Beattie, "Everything you already know about EIA (but don't often admit)," *Environmental Impact Assessment Review*, vol. 15(2), pp.109-14. 1995.

[2] B. Kontic, "Why are some experts more credible than others?" *Environmental Impact Assessment Review*, vol. 20(4), pp.427-34, 2000.

[3] D.P. Lawrence, "Impact significance determination-Back to basics," *Environmental Impact Assessment Review*, vol. 27(8), pp.755-769, 2007.

[4] K.F.R. Liu, J.H, Lai, "Decision-support for environmental impact assessment: A hybrid approach using fuzzy logic and fuzzy analytic network process," *Expert Systems with Applications*, vol. 36, pp.5119-5136, 2009a.

[5] K.F.R. Liu, C.W. Yu, "Integrating case-based and fuzzy reasoning to qualitatively predict risk in environmental impact assessment review," *Environmental Modelling & Software*, vol. 24(10), pp.1241-1251, 2009b.

[6] K.F.R. Liu, H.H. Liang, K. Yeh, C.W. Chen, "A qualitative decision support for environmental impact assessment using fuzzy logic. *Journal of Environmental Informatics*, vol. 13(2), pp.94-104, 2009c.

[7] H. Wilkins, "The need for subjectivity in EIA: discourse as a tool for sustainable development," *Environmental Impact Assessment Review*, vol. 23(4), pp.401-414, 2003.

[8] G. Wood, "Thresholds and criteria for evaluating and communicating impact significance in environmental statements: 'See no evil, hear no evil, speak no evil'?" *Environmental Impact Assessment Review*, vol. 28(1), 22-38, 2008.