A Methodology for the Characterization of Intelligent Environmental Decision Support Systems

Miquel Sànchez-Marrè

Karina Gibert

A. Cabello

F. Sem

International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

A Methodology for the Characterization of Intelligent Environmental Decision Support Systems

M. Sànchez-Marrè[1], K. Gibert[2], A. Cabello[3], F. Sem[3]
[1]*Knowledge Engineering and Machine Learning Group (KEMLG), Software Dept, Universitat Politècnica de Catalunya-BarcelonaTech,Barcelona,Catalonia*
[2]*Knowledge Engineering and Machine Learning Group (KEMLG), Statistics and Operations Research Dept. Universitat Politècnica de Catalunya-BarcelonaTech,Barcelona,Catalonia*
[3]*Barcelona School of Informatics, Universitat Politècnica de Catalunya-BarcelonaTech,Barcelona,Catalonia*
(miquel@lsi.upc.edu)

Abstract: The issue of the paper is to propose a methodology to characterize and describe any Intelligent Environmental Decision Support System (IEDSS). This characterization is necessary to analyse the IEDSSs built until now, to improve their abilities and their current output information. The methodology proposes a selection of a set of features to characterize all IEDSS, and in general, IDSS. The main characteristics of Intelligent Decision Support Systems have been identified through an accurate analysis of cognitive tasks, and several layers involved in this kind of systems. The resulting analysis has been implemented in a form. This form takes into account general features about the IDSS, about the decisions involved, or about the models used, to write only a few ones. The form has been created to describe those characteristics in real environmental applications (IEDSS). To test the viability of the methodology proposed an experimental evaluation has been carried out. This preliminary assessment is a test for the usability and appropriateness of the methodology. A broad sample of Scopus database articles regarding Intelligent Decision Support Systems applications has been analysed. Each one of the authors have analysed a set of papers, and filled the corresponding form. The methodology has been tested for general Intelligent Decision Support Systems, but it will be strengthen to focus on IEDSS. These papers were carefully read and analysed and main features of the proposed IDSS characteristics were identified according to the proposed form. Results of a preliminary analysis with 18 papers about IDSS, including 6 about IEDSS are presented to illustrate the applicability of the methodology. In the near future, a collaborative analysis through a web site is proposed to enlarge the IEDSS applications browsed.

***Keywords**:* Intelligent Environmental Decision Support Systems; Identification of IEDSS Characteristics; Collaborative Methodology.

1 INTRODUCTION

In the last decades, mathematical/statistical models, numerical algorithms and computer simulations have been used as the appropriate means to gain insight into environmental management problems and provide useful information to decision makers. To this end, a wide set of scientific techniques have been applied to environmental management problems for a long time and with good results. Most of these efforts were focused on problems with *relatively simple levels of complexity*. This kind of systems could be named as Decision Support Systems (DSSs). Consequently, many *complex environmental problems* have not been effectively addressed by the scientific community. However, the effort to integrate new tools to deal with more complex systems has led to the development of the so-called *Intelligent Environmental Decision Support Systems* (IEDSSs) [Guariso and Werthner, 1989], [Rizzoli and Young, 1997] [Cortés *et al*., 2000] [Poch *et al*. 2004] [Sànchez-Marrè *et al*., 2008b].

An IEDSS could be defined as done by R. Sojda [Sojda, 2002] who defines them as systems using a combination of models, analytical techniques, and information retrieval to help develop and evaluate appropriate alternatives [Adelman 1992; Sprague and Carlson 1982]; and such systems focus on strategic decisions and not operational ones. More specifically, decision support systems should contribute to reducing the uncertainty faced by managers when they need to make decisions regarding future options [Graham and Jones 1988]. Distributed decision making suits problems where the complexity prevents an individual decision maker from conceptualizing, or otherwise dealing with the entire problem [Boland et al. 1992; Brehmer 1991]. Others authors define an IEDSS them as an intelligent information system that reduces the time in which decisions are made in an environmental domain, and improves the consistency and quality of those decisions [Haagsma and Johanns, 1994], [Cortés et al., 2001]. Others definitions could be found in [D'Erchia et al, 2001].

IEDSSs have generated high expectations as a tool *to tackle problems with higher levels of complexity*. Thus, in a past review of the relevant literature in the topic, more than 600 references were found (including journal articles, conference papers, and technical reports) during the 90s, with only 10 references in 1992 and more than 150 references per year towards the end of the decade [Sànchez-Marrè *et al.*, 2002]. The range of environmental problems to which IEDSSs have been applied is wide and varied, with water management at the top (25% of references), followed by aspects of risk assessment (11.5%) and forest management (11.0%). Equally varied are the tasks to which IEDSSs have been applied, ranging from monitoring and data storage to prediction, decision analysis, control planning, remediation, management, and communication with society.

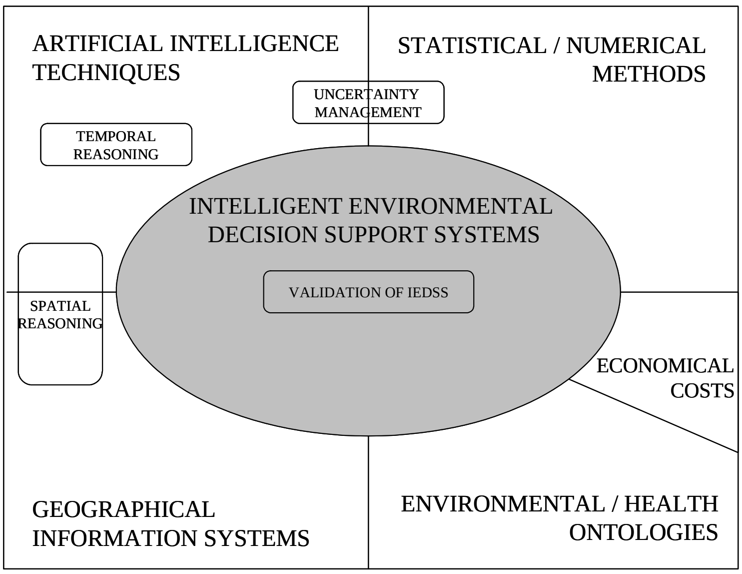

Figure 1. IEDSS conceptual components

Environmental issues belong to a set of critical domains where wrong management decisions may have disastrous social, economic and ecological consequences. Decision-making performed by IEDSSs should be collaborative, not adversarial, and decision makers must inform and involve those who must live with the decisions. What an IEDSS contributes is not only an efficient mechanism to find an optimal or sub-optimal solution, given any set of whimsical preferences, but also a mechanism to make the entire process more open and transparent. In this context, IEDSSs can play a key role in the interaction of humans and ecosystems, as they are tools designed to cope with the multidisciplinary nature and high complexity of environmental problems. Decisions are made when a deviation from an expected, desired state of a system is observed or predicted. This implies a problem awareness that in turn must be based on information, experience and knowledge about the process. Those systems are built by integrating several artificial intelligence methods, geographical information system components, mathematical or statistical techniques, and environmental/health ontologies, and some minor economical components (see figure 1).

Therefore, IEDSS are highly complex systems. Many IEDSS have been developed to solve several environmental problems in different environmental areas. The success of those systems has been quite different, but all the researchers in the area share the opinion that this kind of systems can be improved.
To improve the operation and abilities of Intelligent Environmental Decision Support Systems, it is mandatory to find out what is the current state of the developed IEDSS in the literature to identify their advantages and drawbacks. Notwithstanding, it is not an easy task to analyse, to characterise, and to understand the deployed IEDSS, because most of them are described with a different detail level and usually using a different terminology to refer to the same items.
Thus, the aim of this paper is to propose a methodology, based on the selection of a common set of features to characterize an IEDSS. This set of features will be a common ontology to describe, to understand, and to share knowledge about the Intelligent Environmental Decision Support Systems. In the next section, this methodology will be explained. In section 3, a survey on a sample of 18 papers about IDSS will be detailed as a preliminary test of viability of the methodology. Also the results of the survey are discussed in section 4. Finally, some conclusions and future work are presented in section 5.

2 CHARACTERIZATION OF AN IEDSS

A cognitive analysis of the tasks undertaken in those systems has been carried out to propose the right features to characterize an IEDSS. Taking into account the cognitive approach architecture for the development of IEDSS proposed by [Sànchez-Marrè et al., 2008a] and taking into account the main cognitive tasks of human beings within the decision-making environment, the common underlying tasks in most of IEDSS can be extracted. The general architecture proposed for IEDSS outlined three kinds of tasks involved in the operation of an IEDSS:

- Analysis tasks: In this layer is where most of the analytical processes are run. At this layer, the data gathering processes as well as the knowledge discovery process by means of some data mining techniques are undertaken to get diagnostic models. These models will provide the IEDSS with hard analytical power to get an insight of the environmental system/ process being supervised in real-time or managed in an off-line basis. Most of these techniques come from the data mining field.
- Synthesis tasks: This layer wraps all the work necessary to synthesize possible alternative solutions for the different diagnostics found in the previous step. This synthetisation task can be done through several solution-generation methods based on Statistical techniques (a regression

model, etc.), on Artificial Intelligence techniques (a knowledge-based sytem, etc.) or on Mathematical/Numerical techniques (an optimization technique, etc.). The integration of different methods could enhance the problem solving ability of the IEDSS.

- Prognosis tasks: At this upper layer relays the inherent ability of IEDSS to decision support tasks. At this level, the several predictive models, which can be numerical (mostly simulations) or rather qualitative (qualitative reasoning or qualitative simulations), are used to estimate the consequences of several actions proposed in the previous step by the solution-generation methods. These "what if" models let the final user/decisor to make a decision based on the evaluation of several possible alternatives. At this stage, the temporal and spatial features could be very important for a good environmental modelling practise.

Starting from these cognitive levels, an accurate identification of main features involved in each one of the layers within an IEDSS was done, in order to be able to characterise and analyse them in a right way.

1. Publication
Authors   Journal   Volume & Number   Pages   Year

2. IDSS
To which domain is applied the IDSS?
Static / dynamic IDSS?
Uses feedback to improve its decisions?
Uses on-line information?

3. Decisions
How many decisions are involved?
How many decisors are involved?
Type of decisions (operational / tactic / strategic)?

4. Data / Knowledge
Use of GIS?
Use of domain ontologies?
Use of temporal data?
Use of spatial data?
Use of preprocessing techniques?
Which preprocessing techniques were used? (Data Transformation / New data generation / Data Cleaning / Outliers' management / Missing value management / Data Selection / Feature Selection / Feature Weighting)

5. Models used (AI or Statistical or Hybrid)
Descriptive Models
Associative Models
Discriminant Models
Predictive Models
Use of Hybrid methods?

6. Decision Support
Model Integration?
Predictive skills?
Explanation capabilities?
Alternatives' evaluation ?
Real-time actuation?

7. Validation/Postprocessing
Use of any validation technique?
Use of any postprocessing technique?

Figure 2. Summary of the form to describe an IDSS

The main features identified were related to the following items:

- Target field or domain

- Kind of IDSS
- IDSS information managed
- Kind and number of decisions and decisors involved
- Kind of data types in the dataset
- Data pre-processing techniques
- Type of domain knowledge used
- Models used in the IDSS
- Combination/hybrid methods
- Decision Support level capabilities
- Data post-processing techniques
- Validation methods
- Software tools

All this features coming from the cognitive analysis were implemented in a form. This form is summarized in the figure 2. This form is the actual implementation of the methodology for a practical use.

3 SURVEY ON A SAMPLE OF IDSSs

In order to confirm that the proposed features were the right ones to characterise an IDSS or IEDSS, a survey on several publications from a sample of Scopus database articles regarding Intelligent Decision Support Systems applications was planned. As a preliminary survey, the search was broadened to all fields (IDSS), and not exclusively to environmental domains (IEDSS). The reason behind that is to check whether the features identified for characterising the IDSS were reliable enough, before being applied to a deeper analysis in IEDSS.

Scopus database was searched for articles or reviews containing the words "Intelligent Decision Support System" in the title. One hundred and fifty six articles were obtained. The search of the words was restricted only to the article title, because there were more than 2000 articles with these words on whatever search field (title, abstract, keywords, references, authors, etc.). From those 156 publications, only those ones with the full text available were considered, and from those ones, 18 representative papers from several application domains (Environment, Medicine, Manufacturing, etc.) were finally selected. These papers were carefully read and analysed. For each paper, a row of a data matrix was filled according to the questionnaire in the form described in figure 2. The contents of the data matrix were given as input to a statistical package, and some descriptive statistics, like minimum and maximum values, mean value, standard deviation, 1st and 3rd quartiles and median for continuous variables, and frequency tables, mode value for qualitative variables were performed to get an insight of main features of real IDSSs developed.
Among the 18 papers referring to IDSS, there were 6 of them which were within the environmental fields, i.e., they were IEDSS.

4 RESULTS OF THE STUDY

The papers analysed were published between years 1996 and 2009. All were published in well-known and respected international journals indexed in the SCI database. All were journal articles or journal reviews, because conference articles were explicitly excluded. Of course, some interesting systems are never published in journals, but the reason to constrain the test to journals is because normally in journal articles all the systems/products are thoroughly described, and all the features to characterize them would be found in the text. In other articles published in conference papers or other publications, probably not all the information would be available.

The mean number of pages by each article was 17.56. The minimum number was 9 pages and the maximum number was 33 pages. Of course, the number of pages it is not an important feature per se, but it is worth mentioning that most of the papers were long, and that extracting the main characteristics of the IDSSs described in the papers was not an easy task. This was an additional reason to limit the preliminary study to only 18 papers.

The target domains to which they were applied ranged from air pollution, nuclear emergency, petroleum-contaminated sites, urban development, chemical processes among the IEDSS, and from manufacturing, stock markets, national defence budgets, financial decision strategies, automated systems, medical systems, spaceflight operations, digital preservation, or product development among the general IDSS.

Table 1.Kind of IDSS in the sample

Kind of IDSS	Frecuency
Static IDSS	7
Dynamic IDSS	6
Both Static and Dynamic IDSS	5

The kind of IDSS was equally divided among the static IDSS (7), the dynamic IDSS (6) or both kinds of IDSS (5) as it can be seen in the table 1. The number of decisions involved in the IDSS range from 2 to 9, and the average value was 3.76. Thus, the most common value was 2. Thus, normally only 2 decisions were involved within the IDSS.
Regarding the kind of decisions involved (operational, tactic or strategic) was discovered that the most common decisions are operational (13), and the tactic (8) or strategic (9) decisions were almost equally distributed among the IDSSs, as it is shown in the statistics analysis depicted in table 2.

Table 2. Operational, tactic and strategic decisions in the IDSS in the sample

Kind of Decisions involved	Yes	No	Total
Involving Operational decisions	13	5	18
Involving Tactic Decisions	8	10	18
Involving Strategic Decisions	9	9	18

Regarding the kind of data managed by the IEDSS in their datasets, most IDSS used numerical data, qualitative data and temporal data. Spatial 2D data or spatial 3D data or the use of GIS is not a common feature within the IDSS analysed.

Table 3. Kind of pre-processing techniques used in the sample

Pre-processing Techniques used	Yes	No	Total
Data Transformation	7	11	18
New data Generation	8	10	18
Missing Data Treatment	2	16	18
Ouliers Management	1	17	18
Data Selection	7	11	18
Feature Selection	5	13	18
Feature weighting	4	14	18

The use of pre-processing techniques was employed in 13 systems, and the other 5 IDSS did not use these techniques previously to the data mining step to obtain some data-driven model. The techniques used in the development of these systems are described in the table 2, were the most used methods were data transformations, feature selection and/or weighting the second most used, and finally the least used were missing and outlier management. The average number of techniques used in each IDSS is around 2.

Table 4a. Most common methods used in the sample

Most Common Methods Used	Yes	No	Total
Clustering	2	16	18
Association Rules	8	10	18
Multivariate Analysis Methods (CPA, etc.)	0	18	18
Bayesian Networks	2	16	18
Case-Based Reasoning	5	13	18
Classification Rules	9	9	18
Decision Trees	4	14	18
Discriminant Analysis	2	16	18
Regression Trees	0	18	18
Support Vector Machines	0	18	18
Artificial Neural Networks	8	10	18
Linear Regression	4	14	18
ANOVA	0	18	18
Time Series Analysis	2	16	18

There is a huge variety of available models to be used in an IDSS. An exhaustive list of all possible methods in the questionnaire could not be provided, because the questionnaire would have been too large and probably some of them would have been forgotten. Therefore, according to the authors experience in IDSS topic, the most common methods were directly put in the questionnaire. Most common used techniques were classification rules, association rules and artificial neural networks. After those techniques, case-based reasoning, decision trees and statistical regression were also used. Clustering techniques, Bayesian networks, discriminant analysis and time series analysis were used only in 2 systems. Finally, multivariate analysis methods, regression trees, support vector machines and ANOVA analysis were not used at all. The detailed statistics are depicted in table 4a.

Additionally, there was one item in the questionnaire named as “Others methods used?” to let specify the other methods used in the IDSS which were not in the list of common methods. The other methods used, which were not in the list of common methods initially considered in the questionnaire, were fuzzy logic, genetic algorithms, rule-based reasoning and multi-agent systems, which were also used (see table 4b).

Table 4b. Other methods used in the sample

Other Methods used	Yes	No	Total
Multi-Agents Systems	6	12	18
Fuzzy Logic / Fuzzy Systems	5	13	18
Rule-Based Reasoning	2	16	18
Genetic Algorithms	1	17	18
Constraint Programming	1	17	18
Multiple Category Classification	1	17	18
ARIMA Model	1	17	18
Natural Language Generation	1	17	18

The different number of models/techniques used in the same IDSS ranged from 1 technique up to 8. The average value of techniques used per IDSS was 3.28. The first quartile (Q1) is 2 and the Q3 is 4, being 3 the median value. A bar plot showing the frequency distribution is depicted in figure 3. It is worth to mention that the plot is skewed to the left, meaning than most of the IDSS analysed have used between 2 and 4 models.

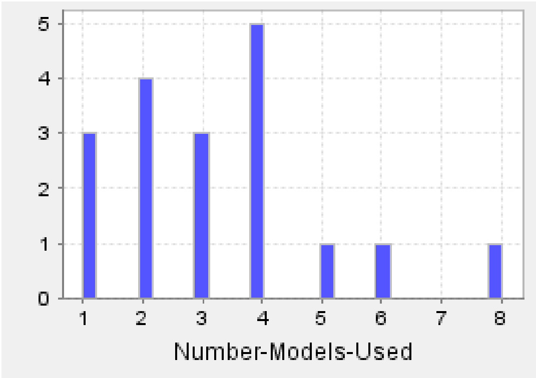

Figure 3. Bar plot of the number of models used in the sample

Post-processing techniques were generally not used in the IDSS deployment. Only 6 systems used them. Regarding validation techniques, 6 systems did not use any technique, or at least, they were not reported. Most commonly used techniques were visualization techniques (5) and expert validation (5). One system used a benchmark procedure, and another one used a cross-validation strategy.

Almost all the IDSS used only one software tool, and four of them used a couple of tools. Common used data mining tools like Weka or R software were not used. Most software tools were self-developed tools or some basic tools like Microsoft Excel or Access, or some programming languages like Prolog, C++, or some other specific software tools like JADE or G2.

5 CONCLUSIONS AND FUTURE WORK

A methodology to characterize Intelligent Environmental Decision Support Systems has been presented. The methodology proposes a set of concepts and features that can characterize different IEDSSs. These features have been derived from a cognitive analysis of IEDSS tasks. Moreover, these proposed features constitute a common ontology about IEDSS knowledge allowing sharing a common terminology for all IEDSSs.

This is a necessary step to analyse, to compare and to assess the different IEDSS deployed in real applications. After that analysis is done, the current limitations of IEDSS will be outlined. This is a mandatory step to overcome these limitations of these systems in the future. Deploying improved IEDSS will be possible due to this characterization methodology.

The reliability of the proposed methodology has been preliminary tested through the survey of a sample of 18 papers describing each one an IDSS. The methodology has shown to be a promising one, because all IDSSs have been able to be analysed and characterised according to the several features and concepts proposed.

Even though not definitive significant conclusions can be extracted from this small sample, the survey has provided both a positive feasibility and reliability test of the methodology, and also a preliminary overview of how are being deployed Intelligent Decision Support Systems. Thus, it has showed some main trends in the deployment of IDSSs. These trends must be confirmed with broader studies, which will be undertaken in the near future.

Some trends extracted from the survey can be outlined, but they must be confirmed in next broader samples:

- Static IDSS, Dynamic IDSS, and even both kind of IDSS have been equally deployed in analysed systems

- The most common number of decisions supported by IDSS analysed is two.
- Most common decisions are operational ones (72%), and static and strategic decisions are almost equally distributed (14% each one).
- Spatial 2D or 3D data or the use of GIS was not a common feature in the IDSSs analysed.
- Near 33% of IDSSs did not use any pre-processing technique in their datasets.
- Most common models/techniques/methods used in the IDSS analysed were, in decreasing order of use: classification rules, associative rules, artificial neural networks, multi-agent systems, case-based reasoning, fuzzy logics/fuzzy systems, decision trees, linear and regression models.
- The 75% of IDSSs used between 1 and 4 different models, being 3 the median value. This fact probably implies that the integration, cooperation and interoperability of methods should be a working direction in the improvement of IDSSs.
- Post-processing techniques were generally not used in IDSSs.
- Software tools used were self-developed tools, or basic tools or specific tools. This fact outlines that there is not available a general tool for the deployment of IDSSs.

In a next step, more papers will be analysed under the same methodology, and especially the aim will be focused on IEDSS. With bigger set of data deeper information could be extracted from data, going further than basic descriptive statistics. Then, significant inferences could be made.
Notwithstanding, the analysis undertaken have been very revealing to test the characterisation methodology proposed in this paper.

In the near future, the authors want to propose the design and the set-up of a web site where Environmental and Computer Scientists could fill in the analysis of whatever paper he/she reads about IEDSS. That way, a significant corpus of analysed applications will be explored in a collaborative way in a relatively short time. After a minimum number of articles will be analysed, then the information from the website will be mined using basic statistical analysis and some specific treatments for multi-response variables, to get a first picture of what is currently being done in the Environmental applications of Intelligent Decision Support Systems. Some R scripts have been defined to replicate the analysis in the future with extended samples coming from the website or other means.

REFERENCES

Adelman, L. 1992. Evaluating Decision Support and Expert Systems. John Wiley and Sons, New York, NY.

Boland, R.J., Mahewshwari, A.K., Te'eni, D., Schwartz, D.G. and Tenkasi, R.V. 1992. Sharing perspectives in distributed decision making. In: Proceedings of the Conference on Computer-supported Cooperative Work, Association for Computing Machinery, New York, NY.

Brehmer, B. 1991. Distributed decision making: some notes on the literature. In: Rasmussen, J., Brehmer, B. and Leplat, J. (Eds.), Distributed Decision Making: Cognitive Models for Cooperative Work, John Wiley and Sons, Chichester, England.

Cortés, U., Sànchez-Marrè, M., Sangüesa, R., Comas, J., R-Roda, I., Poch, M. and Riaño, D. Knowledge Management in Environmental Decision Support Systems. AI Communications. 14(1), 3-12 2001.

Cortés, U., M. Sànchez-Marrè, L. Cecaronni, I. R.-Roda, and M. Poch, Artificial Intelligence and Environmental Decision Support Systems, Applied Intelligence, 13(1), 77–91, 2000.

D'Erchia, F., C, Korschgen, M. Nyquist, R. Root, R. Sojda, and P. Stine. 2001. A framework for ecological decision support systems: building the right systems and building the systems right. U.S. Geological Survey, Biological Resources Division, Information and Technology Report USGS/BRD/ITR-2001-0002.

Graham, I. and Jones, P.L. 1988. Expert Systems: Knowledge, Uncertainty, Decision, Chapman and Hall, New York, NY.

Guariso, G. and Werthner H., Environmental Decision Support Systems. Ellis Horwood-Wiley, New York, 1989.

Haagsma, I.G. and Johanns, R.D. Decision support systems: An integrated approach. In Environmental Systems Vol. II. Zannetti, P.(Ed.),pp.205-212, 1994.

Poch M., Comas J., Rodríguez-Roda I., Sànchez-Marrè M. & Cortés U. Designing and building real environmental decision support systems. Environmental Modelling and Software, 19(9), pp. 857-873 (2004).

Rizzoli, A.E. and Young, W.Y. Delivering Environmental Decision Support Systems: software tools and techniques. Environmental Modelling and Software 12(2-3):237-249, 1997.

Sànchez-Marrè, M., Cortés, U., Comas, J., Rodríguez-Roda, I. And Poch, M. Research report, UPC, 2002.

Sànchez-Marrè, M., J. Comas, I. Rodríguez-Roda, M. Poch and U. Cortés. (2008a). Towards A Framework for the Development of Intelligent Environmental Decision Support Systems. 4th International Congress on Environmental Modelling and Software (iEMSs'2008). iEMSs'2008 Procceedings, Vol. 1, pp. 398-406, 2008 ISBN: 978-84-7653-074-0.

Sànchez-Marrè M., Gibert K., Sojda R. S., Steyer J. P., Struss P., Rodríguez-Roda I., Comas J., Brilhante V. and Roehl E. A. (2008b). Intelligent Environmental Decision Support Systems. Chapter 8 in the book edited by Jakeman, A.J., Rizzoli, A., Voinov, A. and Chen, S., Environmental Modelling, Software and Decision Support. State of the Art and New Perspectives, pp. 119-144. In the series: Developments in Integrated Environmental Assessment, vol. 3. Elsevier Science, Amsterdam, 2008. ISBN: 978-0-08-056886-7.

Sojda, R. S. 2002. Artificial intelligence based decision support for trumpeter swan management. PhD Dissertation. Colorado State University. Fort Collins, Colorado. 183 pages.

Sprague Jr., R.H. and Carlson, E.D. 1982. Building Effective Decision Support Systems, Prentice-Hall, Englewood Cliffs New Jersey.