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

Application of Bayesian Networks for agricultural land suitability classification: a case study of biosolids amendment

Ana Passuello[1], Oda Cadiach[1], Vikas Kumar[1,2], Marta Schuhmacher[1]
[1]*Environmental Analysis and Management Group, Departament Enginyeria Quimica, "Rovira I Virgili" University*
(ana.passuello/oda.cadiach/vikas.kumar/marta.schuhmacher@urv.cat)
[2]*Catchment Science Centre, Kroto Research Institute, University of Sheffield, Broad Lane, Sheffield S3 7HQ, UK.*

Abstract: Land suitability classification tools have been largely applied by environmental managers to support decision making. These tools are needed in the case of the management of biosolids on agricultural soil, to minimise environmental contamination and human exposure. To define the suitability of the agricultural areas is a complex task that involves integrating parameters that are incomparable and sometimes incommensurate, such as soil classification, crop type, etc. In addition, the lack of knowledge on the representation of natural systems is of concern and high uncertainty is associated to model development. In this study, we propose the application of Bayesian networks (BNs), a recognised modelling solution to deal with uncertain and complex problems, to classify the suitability of agricultural land to receive biosolids as an organic amendment. A case study of sewage sludge application in agricultural land in North-Eastern part of Spain (Lleida - Catalonia) was used to describe model's application. The developed Bayesian network represents causal relationships between the terrain characteristics and the identified issues related to the amending practice (environmental contamination and human exposure). The causal relationships were defined by local stakeholders and environmental experts in 3 workshops, where they were able to identify the main problems related to this practice and investigate the impacts according to different soil and landscape characteristics. As a final step, model outputs were represented in GIS, given the general trends for each agricultural area. In addition, uncertainty was represented in maps that give the probability of reaching each suitability class. These suitability maps represent an innovative way of evaluating the results of the land suitability classification.

Keywords*:* Bayesian network; Decision support; Sewage sludge; Waste management; Land suitability.

1 INTRODUCTION

Land suitability tools have been extensively applied to identify better management practices in agricultural areas. These tools evaluate the suitability of an agricultural land to a specific practice or land use. Soil and landscape properties are essential in this type of evaluation, fact that makes especially interesting the coupling of this type of model with Geographic Information Systems (GIS).

Suitable models and data are limited in diffuse contamination of sewage sludge (SS) in agricultural fields. However, it may be possible to build a decision model that considers the best combination of soil and landscape parameters, to prevent environmental contamination and human exposure to the contamination present in SS. In a previous study [Passuello et al. 2012], a traditional Multi-Criteria Decision Analysis (MCDA) model was applied to classify agricultural parcels. The model gave insights regarding soils suitability, but presented an important drawback, as the MCDA tool did not incorporate the uncertainty related to model development.

A method that is able to incorporate the uncertainty related to model development, integrating different sources and types of knowledge is needed to determine the suitability of agricultural areas. Bayesian networks (BNs) have been successfully applied to tackle this type of problem. This method is able to integrate knowledge from different disciplines, aggregating system complexity to the level that is appropriate to represent and communicate uncertainties [Lerner et al. 2010; Marcot et al. 2006].

BNs have been increasingly applied to land management and land use policy. For example, Cain et al. [2003] investigated whether Bayesian networks could provide the generic framework to develop a DSS for agricultural system management. Ticehurst et al. [2007] used BNs for assessing the sustainability of social, economic and environmental values within coastal lake catchments. In addition, Bacon et al. [2002] performed a two stage model of land use change. BNs were applied to explore the relations between personal satisfaction and the costs for the landholders to change their land use. Recently, Holzkaemper et al. [2012] used BN as decision-support tool for integrated catchment management. All the authors agree that BNs help the planning process, allowing the development of complex systems from a multi-disciplinary perspective, presenting key findings in a format that is easily interpreted and communicated, and bridging the gap between stakeholder debate and the experts.

In this work, we propose the application of BNs to define the suitability of agricultural areas to receive sewage sludge amendment. A case study of agricultural soils in North-Eastern part of Spain (Lleida) has been used to describe the model's application. This analysis, coupled with GIS, gives insights for policy makers and farmers on the best management options for the studied area.

2 MODEL DEVELOPMENT

2.1 Workshops

Several workshops were organized with the objective of identifying the main issues related to the practice of amending agricultural soils with sewage sludge. These workshops were attended by stakeholders and environmental experts. In the first workshop, attended by experts, legislation accomplishment was pointed as one of the key issues to be solved by the decision support tool. Another identified point of concern was human exposure to contaminated crops and water. In general, experts identified two main groups of factors: the first regards human exposure to potential contaminations and the second identifies the factors that control environmental contamination (soil, groundwater and superficial water).

After the first workshop, the model structure was defined and the model applied to the area. Then, a second workshop was performed to present the results to the experts and model was calibrated. After extensive calibration, model was presented to different stakeholders (SS managers, environmental agency) at the third workshop. Model was positively accepted, especially by the managers. The need for improvement was indicated in some cases. Due to this fact, as a final step, model was validated through the application of questionnaires. In these questionnaires, stakeholders and environmental experts were asked to classify some real agricultural parcels, without knowing their location, only their characteristics. Results show a good agreement between model results and stakeholders expectations. The steps performed to develop the model are briefly described in the following sections.

2.2 Identification of model variables and structure

The BN model was developed to classify the suitability of different agricultural areas to receive sewage sludge as an organic amendment. The model was intended to represent causal relationships between the terrain characteristics and the identified problems related to the management practice.

The structure of the land classification BN is shown on Figure 1. Parent nodes description is provided at Table 1.

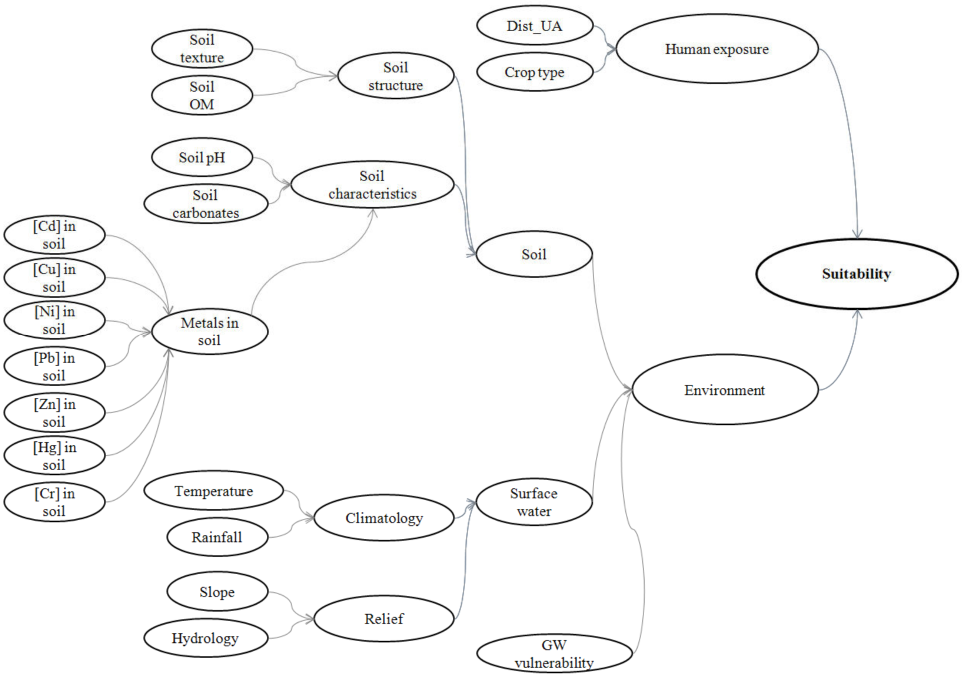

Figure 1. Structure of land classification BN

Two main groups of variables were defined. The first group regards the human exposure due to SS application in agricultural areas, through the assessment of the distance of the field to the urban areas, and the transference of contaminants to food, which is related to crop type. The second group of variables concerns the protection of environmental matrices (soil, surface water and groundwater). To assure environmental protection, soil and landscape properties were considered.

Table 1. Description of the model root variable nodes

Variable	Definition
Distance to urban areas	Considered an important factor for reducing human health risk and bad odours.
Crop type	Different crop types have different potential of accumulating contaminants.
Soil texture	Soil texture was classified according to the European Soil Database (European Communities, 2006). In this regard, a soil with high clay content was considered to have a better structure than a sandy soil.
Soil organic matter (OM) content	Organic matter provides a better soil structure. OM is related to soil organic carbon by the Van Bemmelen factor.
Soil pH	Between soil parameters, pH is the most important factor in metals mobility and bioavailability in soil.
Carbonates	Carbonates decrease metals mobility in soil.
Metals in soils	Soils with lower metals content are preferred for SS amendment. These levels are controlled by the Spanish Royal Decree 1310 (1990).
Temperature	Considers the mean annual temperature. Zones with higher temperatures present a higher rate of organic contaminants degradation.
Rainfall	Regards the mean annual precipitation values. Higher precipitation rates are responsible for contaminants movement through lixiviate and surface runoff.
Slope	Higher slope values lead to higher rates of surface runoff. The classification of the European Digital Archive on Soil Maps of the World was considered.
Hydrology	Fields far from water bodies are considered more suitable for sewage sludge amendment. The Euclidean distance (m) from the field to the open waters was measured.
GW vulnerability	In accordance with the Nitrate Directive (European Commission, 2010), the vulnerability map elaborated by the local environmental agency was employed in this study.

2.3 Categorization of the variables

Having defined the structure, the next step was specifying the states of the variables. In this case, 4 suitability ranges were defined: poor (unsuitable), marginally, moderately and highly suitable. Four states were also defined for all the nodes, except for the inputs that were classified by default (crop type, soil texture, and groundwater). For the latter, the same classes already assigned were considered. Specifically for the inputs, the states were defined as the ranges where important effects are observed, based on expert knowledge and literature review.

2.4 Simulating Conditional Probability Table

In this step, conditional probability tables (CPTs) are used to describe the probability of each value of the child node, conditioned on every possible combination of values of its parent nodes. This step was performed based on the previous MCDA model [Passuello et al., 2012]. The MCDA model was built considering the 12 variables (criteria) described on Table 1. These criteria were aggregated through the Logic Scoring of Preference (LSP) method, that is based on mathematical models that use Generalized Conjunction/Disjunction (GCD) and other continuous preference logic functions (Dujmovic, 2007), taking into consideration the different levels in the hierarchy of criteria, weights and constraints, over those criteria. As the aggregation of partial preferences is based on GCD logic, the simultaneity and replaceability of the criteria was assessed (Dujmovic and Larsen, 2007). This is a key point of MCDA method, as the stakeholders had previously described some criteria as replaceable and others as complementary. Other major points that were considered in method selection are: flexibility in the number of criteria to be inserted; versatility in model's design; and model's capacity to generate correct logic results in all points of the attribute space. To understand criteria complementarity, let's consider the soil structure node. Appropriate soil structure is characterized by high organic matter content or appropriate texture. Soil with poor texture (sandy, sandy loam soil), that has been amended for several years, may have a high OM content and, in consequence, an appropriate structure (Figure 1). In this case, soil structure node is likely to be considered highly suitable.

To translate MCDA inputs into CPTs, three steps were followed:

(i) Classify each input value according to the previously defined ranges;
(ii) Run the MCDA method for all the combinations of these ranges;
(iii) Calculate the outputs probabilities, which are used to populate CPTs.

All the input ranges of the parent nodes were considered as uniform distributions, and the combinations of these input ranges were used to simulate the probabilities outputs for the child nodes. The probability was obtained by classifying the results of the child node in four predefined categories [Passuello, 2011].

3 CASE STUDY

Lleida is a province of north-eastern Spain, in the western part of Catalonia (Figure 2). The region has a population of 442,308 inhabitants [IDESCAT, 2011] and is the most significant agricultural region of Catalonia, with an important fruit industry. The region is flat at the centre and the south, and mountainous at the North (Pyrenees). Agricultural areas are mainly situated in the flat area.

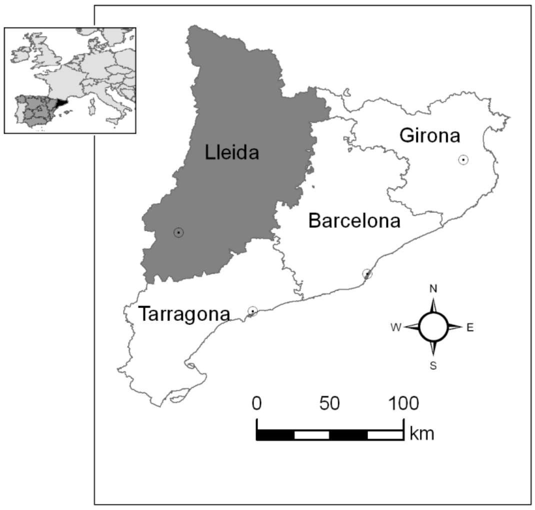

Figure 2. Situation map of Lleida.

Organic amendment is a common practice at the zone, fact that leads to groundwater contamination by nitrates especially in the region close to the provincial capital.

To evaluate the suitability of agricultural areas, the model coupled with GIS was applied. This procedure was selected due to the easiness of data gathering in public geo-referenced databases. Moreover, results representation on GIS is a highly innovative point of this study, which allowed the assessment of a large amount of data. The evaluated agricultural area has more than 457,000 ha, and the pixel size is 200 x 200m. In total, more than 114,000 pixels were evaluated.

The procedure is shown in Figure 3. The first step involved data gathering from GIS. These data were stored on text files and inserted in the software Netica [Norsys, 2010]. Then, the model was run, according to the method described in section 2. Model results were exported to GIS through text files. In this case, 5 series of data were generated: (i) the most probable (MP) class, showing the suitability class with the higher probability for each pixel; and (ii) the probabilities of reaching each of the 4 states (classes). Data gathering and results representation were performed on ArcGIS [Esri, 2011]. All model inputs are described by Passuello et al. [2012].

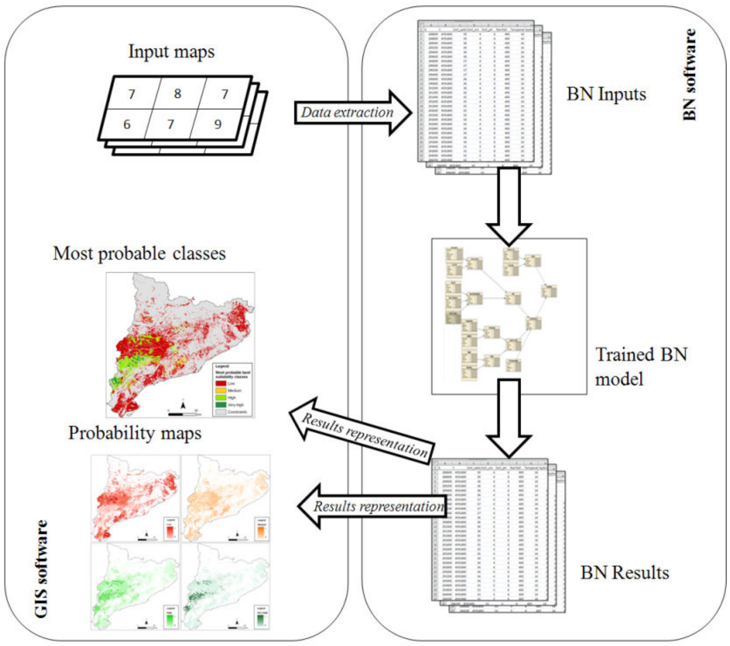

Figure 3. Framework for model application to the case study.

4 RESULTS AND DISCUSSION

Figure 4 shows the results for suitability node, in two different ways. Figure 4 (a) to (d) represent the probability of reaching each suitability class, while Figure 4 (e) gives the general trends regarding soil suitability for receiving sewage sludge amendment.

The majority of the areas (69%) present “unsuitable” as the most probable class. These areas are represented by the dark points in Figure 4(a) and have an historical problem of groundwater contamination. Also, moderately (25%) and highly suitable (3%) classes are observed in the south part of the region (Figure 4 (c)-(e)). These high values are related to low groundwater vulnerability and the presence of fruit fields, leading to a low human exposure.

As can be seen in Figure 4 (e), the most suitable regions are located close to the capital, where most of the sewage sludge is generated. However, most of the areas are considered “unsuitable”, proving that the model is highly restrictive.

In general, probability and most probable maps present the same trends. However, as can be seen in Figure 4, some of the unsuitable areas do not present a high probability of being unsuitable. In this case, a high uncertainty is associated to these results. For these results, the MP map gives incomplete information and the actual trends can only be understood by the evaluation of the probability maps.

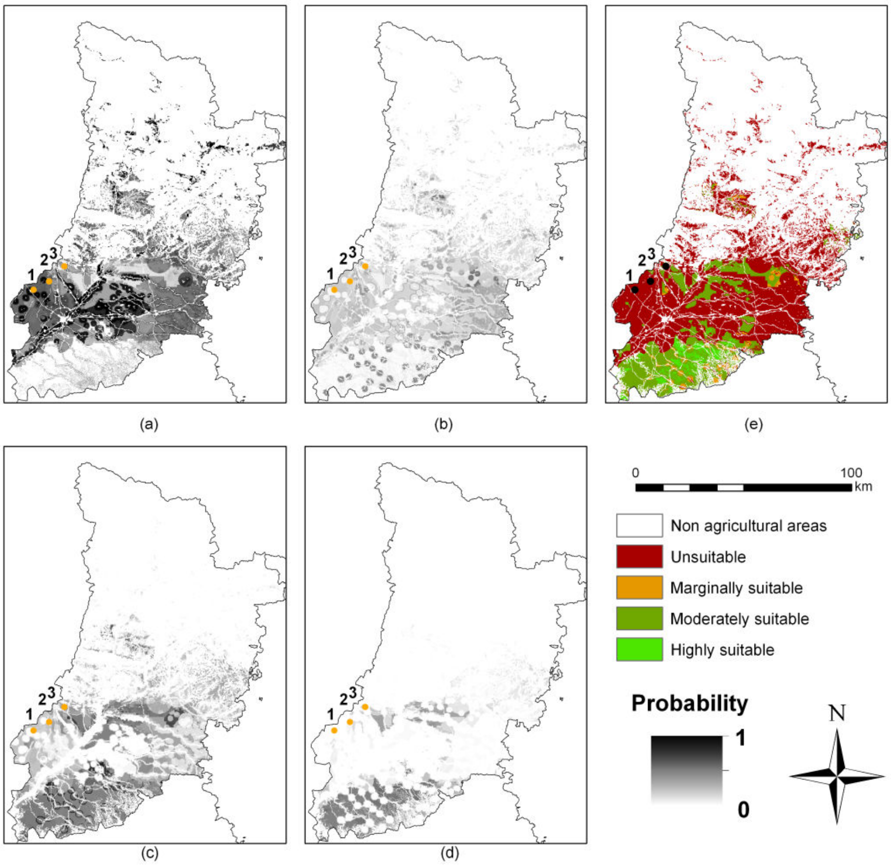

Figure 4. BN results for suitability node. Probability maps defining the likelihood of the different suitability ranges for (a) unsuitable (poor), (b) marginally, (c) moderately and (d) highly suitable classes, and most probable land suitability classes (e).

Uncertainty representation in the model is better understood if we consider the CPTs (Figure 5) for 3 different regions of the study (points in Figure 4). In this case, points 1 and 2 have the same most probable suitability class (Unsuitable), but a different uncertainty is related to those results. While point 1 is certainly unsuitable (P=0.96), point 2 is likely unsuitable (P=0.59). Point 3 is located quite close to points 1 and 2, however it presents a moderately suitable classification (P=0.38). In this case, a high uncertainty is related to the classification for this point, as the probabilities of reaching each class are very similar. This example shows a clear limitation of the model, and proves that the information given by the MP map (Figure 4 (e)) is incomplete. This type of incongruence is hard to define *a priori*, and an evaluation case by case should be considered, especially in those areas where abrupt changes are observed for the MP class.

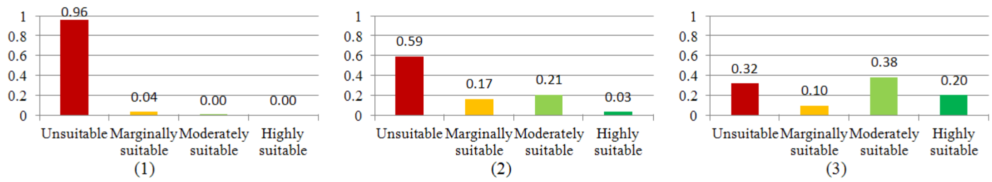

Figure 5. CPTs, represented as distributions, for the Suitability node. The numbers in parenthesis correspond to the points indicated on Fig. 4.

5 CONCLUSIONS

The use of BN techniques for land classification presents several advantages when compared to other integration techniques. The most important ones are the ability of dealing with uncertainty related to model development and the capacity of integrating incomparable data.

The probability maps gave an improved perception of systems uncertainty, clearly indicating the areas with more likelihood of reaching any of the defined classes.

Future improvements of the model should consider improving the evaluation scale, i.e., working in a larger and more detailed scale. Moreover, the development of an open tool to present the results to stakeholders could provide an improved understanding of the model.

ACKNOWLEDGEMENTS

Financial support was provided by the Ministry of Science and Innovation, Spain, through grant No. CTM2007-64490, the SOSTAQUA project, founded by CDTI in the framework of the Ingenio 2010 Program under the CENIT call, and the SCARCE project financed by the Spanish Ministry of Science and Innovation through the project Consolider-Ingenio 2010 CSD2009-00065. A. Passuello received a PhD scholarship from AGAUR (Commissioner for Universities and Research of the Department of Innovation, Universities and Enterprises of the "Generalitat de Catalunya" and the European Social Fund) and a mobility grant by the Ministry of Education through grant No. TME2009-00406 (Orden EDU/2933/2009).

REFERENCES

Bacon, P.J., J.D. Cain, and D.C. Howard, Belief network models of land manager decisions and land use change, *Journal of Environmental Management 65*, 1-23, 2002.

Cain, J.D., K. Jinapala, I.W. Makin, P.G. Somaratna, B.R. Ariyaratna, and L.R. Perera, Participatory decision support for agricultural management. A case study from Sri Lanka, *Agricultural Systems 76*, 457-482, 2003.

CEC, Council of the European Communities, Council Directive 91/676/EEC of 12 December 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources, 1991.

Esri, ArcGIS, v.10,2011. < http://www.esri.com/software/arcgis/index.html>

Holzkaemper, A., V. Kumar , B. Surridge, A. Paetzold, and D.N. Lerner, Bringing diverse knowledge sources together – A meta-model for supporting integrated catchment management, *Journal of Environmental Management 96(1)*,116-127, 2012.

IDESCAT, Padró municipal d'habitants. Xifres Oficials. Recomptes Lleida. Sèrie temporal. Institut d'Estadística de Catalunya, 2011.

<http://www.idescat.cat/territ/BasicTerr?TC=5&V0=5&V3=863&V4=435&ALLINFO=TRUE&PARENT=1&CTX=B&V1=25&VOK=Confirmar>

Lerner, D.N., V. Kumar, A. Holzkämper, B.W.J. Surridge, and B. Harris, Challenges in developing an integrated catchment management model, *Water and Environment Journal 25*, 345-354, 2010.

Marcot, B.G., J.D. Steventon, G.D. Sutherland, and R.K. McCann, Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation, *Canadian Journal of Forest Research 36*, 3063-3074, 2006.

Norsys, Netica, 4.16 ed., 2010. <www.norsys.com>

Passuello A. Development of environmental tools for the management of sewage sludge on agricultural soils. Ph.D. thesis. Universitat Rovira I Virgili, Tarragona, Spain, 2011.

Passuello A, O. Cadiach, Y. Perez, and M. Schuhmacher. A spatial multicriteria decision making tool to define the best agricultural areas for sewage sludge amendment. *Environment International 38,* 1-9, 2012.

Ticehurst, J.L., L.T.H. Newham, D. Rissik, R.A.Letcher, and A.J. Jakeman, A Bayesian network approach for assessing the sustainability of coastal lakes in New South Wales, Australia, *Environmental Modelling and Software 22*, 1129-1139, 2007.