*International Environmental Modelling and Software Society (iEMSs)*
*7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA,*
*Daniel P. Ames, Nigel W. T. Quinn, Andrea E. Rizzoli (Eds.)*
*http://www.iemss.org/society/index.php/iemss-2014-proceedings*

# Probabilistic Mapping With Bayesian Belief Networks: An Application On Ecosystem Service Delivery In Flanders, Belgium

**Dries Landuyt** [a b], **Steven Broekx**[b], **Katrien Van der Biest** [c] **and Peter Goethals**[a]

[a] *Laboratory of Environmental Toxicology and Aquatic Ecology, Ghent University, Jozef Plateaustraat 22, B-9000 Ghent, Belgium (dries.landuyt@ugent.be)*

[b] *Unit Environmental Modelling (RMA), Flemish Institute For Technological Research (VITO), Boeretang 200, B-2240 Mol, Belgium*

[c] *Ecosystem Management Research Group, University of Antwerp, Universiteitsplein 1C, B-2610 Wilrijk, Belgium*

**Abstract:** Ecosystem services are gaining more and more attention in decision support applications. Nevertheless, modelling and mapping ecosystem services to support landscape planning decisions remains challenging. Recently, Bayesian belief networks (BBNs), a probabilistic modelling technique, has been introduced in ecosystem service modelling. Major advantages of this modelling approach include high model transparency which enables stakeholder involvement in model development and evaluation, the ability to incorporate expert knowledge on top of data and the possibility to take into account uncertainties. To combine the advantages of BBNs and spatially explicit modelling in the context of ecosystem service modelling, we developed a Quantum GIS plug-in. The plug-in enables pixel-based application of BBN models to map ecosystem service delivery and associated uncertainties. The obtained probabilistic maps can be used for stakeholder involvement, decision support and probabilistic, regional ecosystem service accounting.

***Keywords:*** BBN Toolbox; decision support system; risk; uncertainty; GIS

## 1 INTRODUCTION

Goods and services delivered by natural and semi-natural ecosystems are generally referred to as ecosystem services [Daily, 1997]. A term that stresses the importance of ecosystems in socio-economic terms. Due to the attempt of the Millenium Ecosystem Assessment to mainstream the ecosystem services (ES) concept in 2005 [MEA, 2005], ecosystems are nowadays increasingly recognised for the benefits they deliver. Although the concept has the potential to support decisions related to evaluation of land use and management practices [Broekx et al., 2013], existing applications of the concept still face some challenges [De Groot et al., 2010]. Mapping of ES delivery for landscape planning is one of them.

Among the growing amount of literature on ES assessment [Seppelt et al., 2011], mapping is gaining more and more attention [Maes et al., 2012; Burkhard et al., 2012]. Current mapping approaches vary from simple indicator-based methods [Haines-Young et al., 2012] to complex mechanistic models [Kareiva et al., 2011; Nemec and Raudsepp-Hearne, 2013]. Although mechanistic models generally deliver more accurate results, they are only applicable to small scale case studies. For ES assessments at larger scales, indicator-based approaches may be the only possibility due to poor availability of data. An intermediary approach that was recently introduced in environmental modelling, and more specifically in ES modelling, is Bayesian belief network (BBN) modelling. Although the complexity of these probabilistic, graphical models is comparable to that of conventional indicator-based approaches, BBNs offer two important advantages: BBNs are able to complement limited available data with expert knowledge and can account explicitly for uncertainties [Aguilera et al., 2011]. The added value of these BBN characteristics to model ES delivery has been discussed by Landuyt et al. [2013]. However, as described in their paper, few of the BBN applications in ES modelling use BBNs for mapping purposes [Van der Biest et al., 2014; Dlamini, 2010; Smith et al., 2007]. A missing link between user-friendly and transparent BBN software and geographical information software (GIS) is probably one of the major reasons for this lack.

This paper describes a software plug-in which was developed to link BBN and GIS software. To enhance expert involvement in model development, as well as in result analysis, both the modelling and the mapping step were kept in a user-friendly and intuitive software environment, Netica [Norsys Software Corporation, 1998] and Quantum GIS (QGIS) [QGIS Development Team, 2012], respectively. The QGIS plug-in is implemented in Python and can be used to run BBNs on spatial input data. Although the tool can be of use in a broad range of applications, we focus on its applicability in ES modelling and mapping. The paper discusses the components of the software framework and demonstrates the tool for a small case study located in the north-eastern part of Belgium. The paper concludes with a general discussion on the applicability of the tool and the added value of uncertainty maps for decision support.

## 2 METHODS

### 2.1 Bayesian belief network models

BBNs are multivariate statistical models that comprise two structural components: a causal network, represented as a directed acyclic graph (DAG), and conditional probability tables (CPTs) that probabilistically quantify the causal relations in the graph. The DAG consists of a set of nodes, representing the system's variables of interest, and a set of arrows, indicating the causal relationships among the system's variables. Each node or variable in the model has a finite set of states it can manifest and the extent to what its different states are realized is expressed through a probability distribution over these states. The probability distribution of a node X is determined by the realized states of its preceding or parent nodes, using the conditional probability P(X|parents(X)) described in Bayes' theorem (1). These conditional probabilities are tabled in each node's CPT. For a detailed model description and statistical background we refer to Jensen [2001].

$$P(X|parents(X)) = \frac{P(parents(X)|X) * P(X)}{P(parents(X))} \tag{1}$$

### 2.2 Software framework

The software framework of the plug-in comprises three main components: model development, carried out in Netica, data and model preprocessing, carried out in Python and model application and output visualisation, performed in QGIS. A schematic representation of the software framework is shown in Figure 1.

As most experts, such as economists, ecologists and sociologists, are not familiar with abstract modelling programs, we have chosen for Netica as model development environment. The user-friendliness and transparency of this software package facilitates the inclusion of knowledge of a broad range of experts into the model. Although some freely available BBN packages exist, such as Hugin [Hugin, 2011] and SMILE [Druzdzel, 1999], Netica outperforms most of them in terms of graphical user interface and user support. For a detailed description of model development in Netica we refer to Netica's user manual [Norsys Software Corporation, 1998]. The in Netica developed model, usually saved as a .neta file, can be used directly as one of the input files of the plug-in.

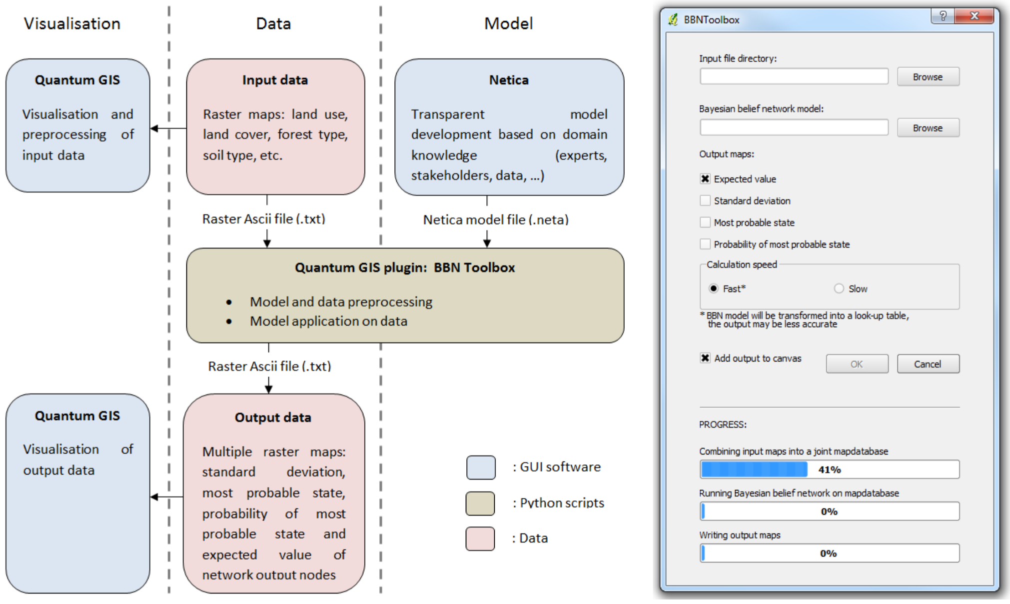


**Figure 1.** A schematic representation of the software framework (left) and a screen-shot of the plug-in dialog screen (right)

The required format for raster input data is ascii, a universal raster data type which can be opened in almost all GIS packages as well as in simple text formatting programs. Each raster input file needs to correspond with one of the input nodes of the developed BBN. As ascii-files only support numerical data, each raster input file should be accompanied with a legend file (excel-format) which assigns to each mapped, numerical code a state name of the corresponding input node. The ascii files are subsequently preprocessed in Python to exclude non-overlapping areas and to eliminate raster offset. After reshaping, all data files are merged into a joint map database (.csv format). Subsequently, all numerical codes are translated into state names based on the provided legend files.

The plug-in offers two possible ways to run the model: a fast, approximate or a slow, accurate run. The slow run mode is based on a built-in function of Netica and runs the BBN model for each pixel independently. The fast method first converts the BBN model into a look-up table that list all possible combinations of input states with their corresponding probabilistic model output. This table, implemented as a dictionary structure in Python, is then used to assign a model output to each pixel of the study area. In slow mode, computational time increases exponentially with the amount of pixels in the study area. The slow method can, therefore, only be used for small study areas (upto 1 million pixels). For larger study areas, unacceptably long run times and memory errors may occur. A major disadvantage of the fast, table-based approach is its inability to deal with missing data in one of the input variables of a particular pixel. The built-in Netica function, used in slow run mode, tackles this frequently occurring problem by assigning uniform distributions to the input nodes for which data is not available. The maximum length of a dictionary structure in Python imposes an additional technical limitation of the fast method. This implies that the table-based approach cannot be used for BBN models with a large number of input variables and a large number of states per input variable. For these models, the slow method is preferable. Alternative data structures in Python or alternative implementations in more performant programming languages will probably resolve this technical issue in the future. The dialog screen of the plug-in (Figure 1) offers the possibility to select among both approaches.

The spatial output of the model can be visualised in four different ways. For each pixel, the output of the BBN model can be mapped as the expected value, the standard deviation, the most-probable state or the probability of the most-probable state. Users can select one or more of these output formats in the dialog screen of the plug-in.

## 3 CASE STUDY APPLICATION

### 3.1 Study area

To illustrate its functionalities, we applied the plug-in on a small scale study area located in the north-eastern part of Belgium, the Belgian landdune region (Figure 2). The study area approximately covers an area of 160 km$^2$ and its main land uses are urban land use (26 %), grassland (25 %), cropland (21 %) and forestry (16 %). Although the region delivers a broad range of ES, we only consider one to illustrate the plug-in. Soil organic carbon (SOC) storage for climate regulation was selected as one of the most important services in the study area. The study area and especially the wet grasslands and forests located along the rivers have a high potential to store organic carbon. Taking into account SOC storage to evaluate alternative land use scenarios will be particularly important for these wet valley bottoms. As almost all land uses in the study area deliver this service, SOC storage is also the most convenient service to map. Other important services in the study area such as food production and wood production are, in contrast, only delivered by a limited number of land uses in the study area.

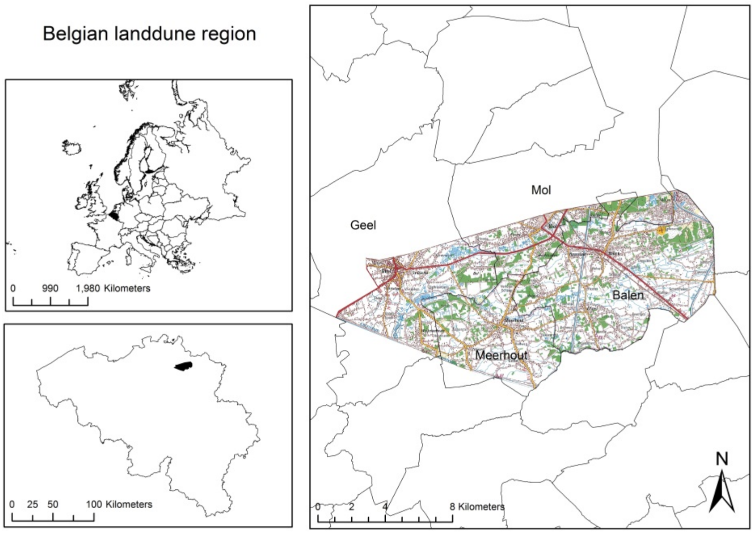


**Figure 2.** Location and topographic map of the study area

### 3.2 Model development

To developed a BBN model for SOC sequestration, we based ourselves on the available literature. Although mechanistic SOC models have been developed in the past [Skjemstad et al., 2004; Byun and Schere, 2006], their extensive data requirements often restrict their use to small field scale studies. Therefore, we developed a BBN model based on an empirical study on SOC storage conducted in Flanders [Meersmans et al., 2008]. In this study, Meersmans et al. [2008] developed a regression model which predicts SOC storage based on soil texture, soil moisture content and land use (grassland, heathland, cropland and forest). The graphical network of the BBN was structured accordingly (Figure 3). The regression model's SOC estimates for each combination of soil texture, soil moisture content and land use were used to populate the CPTs of the BBN. As confidence intervals were available for each estimate, uncertainties could be taken into account in the model's CPTs. To expand the land use classes with a wetland class, additional literature data on SOC storage in wetlands was consulted [Post et al., 1982; Adhikari et al., 2009]. According to the guidelines of Kirschbaum et al. [2001] to model SOC stocks, a correction factor was included to account for five different types of management intensity [Ogle et al., 2005]. The required input raster layers to run this model comprise a land use map, a soil texture map as well as a map representing soil moisture content and a map representing management type. Data on soil characteristics were obtained from the Belgian national soil classification system [AGIV, 2012] with a spatial resolution of 5 m.

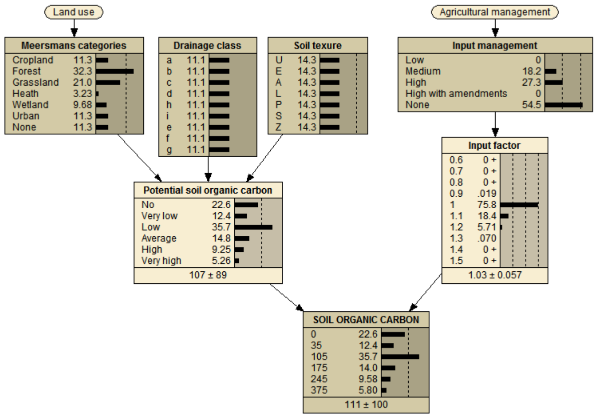


**Figure 3.** The developed Bayesian belief network to estimate SOC storage. For the 'land use' and 'agricultural management' nodes, states are not represented as there were too many to visualise. Note that the represented network does not contain prior knowledge regarding the study area as all input nodes are distributed uniformly.

### 3.3 Results

After running the plug-in with the developed network and the required input maps, four probabilistic output maps were obtained. As can be seen in Figure 4, a considerable amount of similarities can be observed between map a and map c and between map b and map d. Although the mapped indicator differs for each map, map a and map c both represent the amount of carbon stocked in each pixel, while map b and map d both represent the uncertainties associated with the estimates in map a and map c.

As can be seen in map a and map b, the wet grasslands and forests are clearly visible as dark zones around the rivers, denoting high potential SOC storage. However, map b and map d assign high uncertainty values to most of these pixels. Although these wetlands have a high potential, the model results suggest that the realized SOC stock varies considerably in these parcels. The urbanized regions in the study area are visible as white star-shaped spots in map a and c which denotes a very low SOC storage in these regions. Nevertheless, urbanized regions do have the potential to store SOC, for example, in parks, gardens, green roofs, etc. As no data was available for these urban land uses, they were not taken into account as potential SOC sinks.

## 4 Discussion

Although all steps of the software framework (model development, model run and data visualisation) are relatively straightforward for the end-users of the tool, interpreting probabilistic output maps remains an important challenge. We suggest two possible ways to visualise uncertainties: via the expected value (Figure 4c) and the standard deviation (Figure 4b) or via the most-probable state (Figure 4a) and the probability associated with this state (Figure 4d). While the latter approach is more intuitive and thus probably preferred by most of the end-users, statisticians are generally more confident with the first approach. Additional advantages of mapping the most probable state and an associated probability layer include the ease of mapping a fixed amount of classes and the ability to compare

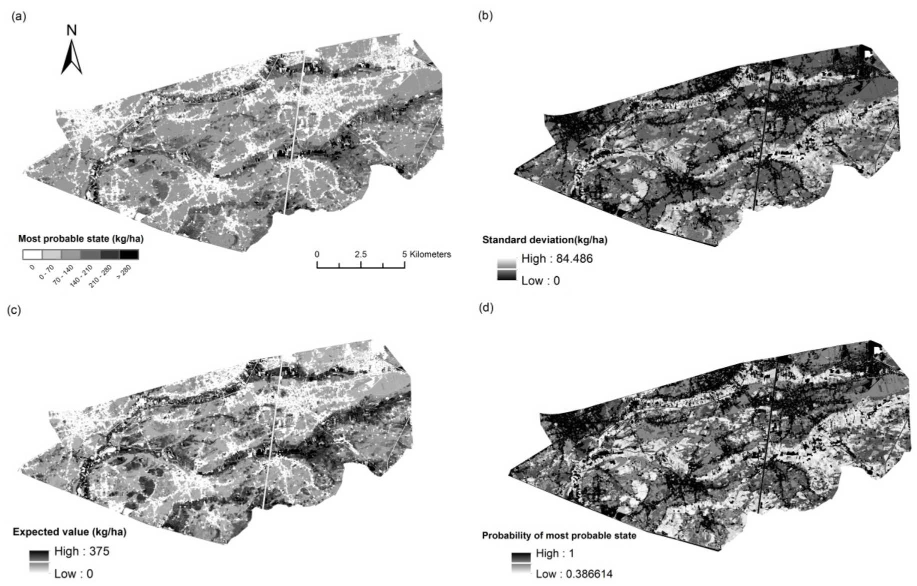


**Figure 4.** Probabilistic outputmaps of the ecosystem service soil organic carbon storage. The maps represent for each pixel the most probable state (a), the standard deviation (b), the expected value (c) and the probability of the most probable state (d).

probability percentages among pixels which is not possible with mapped standard deviations. A major disadvantage of this approach is that the probabilities depend on the number of states of the output variable. A large number of states in the output node of the network will increase the chance for low probabilities in the uncertainty layer. The standard deviation as a measure for uncertainty, on the other hand, is not dependent on the number of states in the output node.

Ideally, both the magnitude of ES delivery and the associated uncertainties are plotted on a single map. Working with a semi-transparent uncertainty layer on top of a base layer which represents the magnitude of delivery is one possibility. Also 3D visualisation techniques can offer a solution for this mapping problem. Nevertheless, maps of this kind will be hard to understand for users without GIS experience. Delivering understandable maps that represent both information types is an important challenge for future research.

For decision makers, uncertainties can be valuable, however, usually not at the pixel level. To answer relevant question for decision makers, such as, predicting the chance that ES delivery in a study area is higher under a particular alternative land use or management scenario, we need to be able to aggregate uncertainty values over the entire study area. On top, spatial interactions, such as, the relation between upstream and downstream management actions may be of interest for decision makers. Currently these spatial interactions cannot be modelled with the presented pixel-based approach.

## 5 Conclusion

The developed QGIS plug-in has the potential to bridge the gap between current BBN research and spatial analyses, not only in the ES research domain but also in a broad range of other domains where uncertainties may be of interest. Nevertheless, mapping uncertainties and interpreting these uncertainties remains a complex issue. Further research on visualisation techniques, reasoning with uncertain pixels and interpreting probabilistic maps is, therefore, needed. On top, the applicability of the tool should be tested to see whether end-users fully understand the delivered uncertainty maps and whether the provided information is useful in day-to day decision making.

## References


Adhikari, S., Bajracharaya, R. M., and Sitaula, B. K. (2009). A review of carbon dynamics and sequestration in wetlands. *Journal of Wetlands Ecology*, 2(1):42–46.

AGIV (2012). Vectorile versie van de bodemkaart. Technical report, AGIV,IWT, Laboratorium voor Bodemkunde van de Universiteit Gent.

Aguilera, P. A., Fernandez, A., Fernandez, R., Rumi, R., and Salmeron, A. (2011). Bayesian networks in environmental modelling. *Environmental Modelling and Software*, 26(12):1376–1388.

Broekx, S., Liekens, I., Peelaerts, W., De Nocker, L., Landuyt, D., Staes, J., Meire, P., Schaafsma, M., Van Reeth, W., Van den Kerckhove, O., and Cerulus, T. (2013). A web application to support the quantification and valuation of ecosystem services. *Environmental Impact Assessment Review*, 40(0):65–74.

Burkhard, B., Kroll, F., Nedkov, S., and Mller, F. (2012). Mapping ecosystem service supply, demand and budgets. *Ecological Indicators*, 21(0):17–29.

Byun, D. and Schere, K. L. (2006). Review of the governing equations, computational algorithms, and other components of the models-3 community multiscale air quality (cmaq) modeling system. *Applied Mechanics Reviews*, 59(1/6):51.

Daily, G. C. (1997). *Nature's services: societal dependence on natural ecosystems*. Island Press, Washington, D.C.

De Groot, R. S., Alkemade, R., Braat, L., Hein, L., and Willemen, L. (2010). Challenges in integrating the concept of ecosystem services and values in landscape planning, management and decision making. *Ecological Complexity*, 7(3):260–272.

Dlamini, W. M. (2010). A bayesian belief network analysis of factors influencing wildfire occurrence in swaziland. *Environmental Modelling and Software*, 25(2):199–208.

Druzdzel, M. J. (1999). Smile: Structural modeling, inference, and learning engine and genie: a development environment for graphical decision-theoretic models. In *Proceedings of the national conference on artificial intelligence*, pages 902–903. John Wiley and Sons LTD.

Haines-Young, R., Potschin, M., and Kienast, F. (2012). Indicators of ecosystem service potential at european scales: Mapping marginal changes and trade-offs. *Ecological Indicators*, 21:39–53.

Hugin (2011). The hugin graphical user interface. http://download.hugin.com/webdocs/manuals/Htmlhelp/.

Jensen, F. V. (2001). *Bayesian networks and decision graphs*. Springer, New York.

Kareiva, P., Tallis, H., Ricketts, T. H., Daily, G. C., and Polasky, S. (2011). *Natural capital: theory and practice of mapping ecosystem services*. Oxford University Press, New York.

Kirschbaum, M. U. F., Schlamadinger, B., Cannell, M. G. R., Hamburg, S. P., Karjalainen, T., Kurz, W. A., Prisley, S., Schulze, E. D., and Singh, T. P. (2001). A generalised approach of accounting for biospheric carbon stock changes under the kyoto protocol. *Environmental Science and Policy*, 4(2):73–85.

Landuyt, D., Broekx, S., D'Hondt, R., Engelen, G., Aertsens, J., and Goethals, P. L. M. (2013). A review of bayesian belief networks in ecosystem service modelling. *Environmental Modelling and Software*, 46(0):1–11.

Maes, J., Egoh, B., Willemen, L., Liquete, C., Vihervaara, P., Schgner, J. P., Grizzetti, B., Drakou, E. G., Notte, A. L., Zulian, G., Bouraoui, F., Luisa Paracchini, M., Braat, L., and Bidoglio, G. (2012). Mapping ecosystem services for policy support and decision making in the european union. *Ecosystem Services*, 1(1):31–39.

MEA (2005). Ecosystems and human well-being: synthesis. Technical report, Island press, Washington, D.C.

Meersmans, J., De Ridder, F., Canters, F., De Baets, S., and Van Molle, M. (2008). A multiple regression approach to assess the spatial distribution of soil organic carbon (soc) at the regional scale (flanders, belgium). *Geoderma*, 143(1):1–13.

Nemec, K. T. and Raudsepp-Hearne, C. (2013). The use of geographic information systems to map and assess ecosystem services. *Biodiversity and conservation*, 22(1):1–15.

Norsys Software Corporation (1998). *Netica Application User's Guide*. Norsys Software Corporation, Vancouver, BC, Canada.

Ogle, S. M., Breidt, F. J., and Paustian, K. (2005). Agricultural management impacts on soil organic carbon storage under moist and dry climatic conditions of temperate and tropical regions. *Biogeochemistry*, 72(1):87–121.

Post, W. M., Emanuel, W. R., Zinke, P. J., and Stangenberger, A. G. (1982). Soil carbon pools and world life zones. *Nature*, 298:156–159.

QGIS Development Team (2012). Qgis geographic information system. open source geospatial foundation project. Technical report, http://qgis.osgeo.org.

Seppelt, R., Dormann, C. F., Eppink, F. V., Lautenbach, S., and Schmidt, S. (2011). A quantitative review of ecosystem service studies: approaches, shortcomings and the road ahead. *Journal of Applied Ecology*, 48(3):630–636.

Skjemstad, J. O., Spouncer, L. R., Cowie, B., and Swift, R. S. (2004). Calibration of the rothamsted organic carbon turnover model (rothc ver. 26.3), using measurable soil organic carbon pools. *Soil Research*, 42(1):79–88.

Smith, C. S., Howes, A. L., Price, B., and McAlpine, C. A. (2007). Using a bayesian belief network to predict suitable habitat of an endangered mammal - the julia creek dunnart (sminthopsis douglasi). *Biological Conservation*, 139(3-4):333–347.

Van der Biest, K., D'hondt, R., Jacobs, S., Landuyt, D., Staes, J., Goethals, P., and Meire, P. (2014). Ebi: An index for delivery of ecosystem service bundles. *Ecological Indicators*, 37(Part A):252–265.