

Modeling Ecological Integrity with Bayesian Belief Networks

**J. M. Barrios, R. Sierra-Alcocer, C. González,
F. E. Mora, M. Munguía, O. M. Pérez, I. Trejo**

General Coordination of Information and Analysis,
National Commission for Knowledge and Use of Biodiversity *
Liga Periférico – Insurgentes Sur, No. 4903.
Col. Parques del Pedregal, Tlalpan, 14010, México, D.F.

Abstract

Although the concept of ecological integrity is referred in many country legislations there is no consensus on how to formalize and implement it. One possible definition is as the capacity of an ecosystem to support and maintain a balanced, integrated, and adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of a natural habitat of the region. Our objective is to model this interpretation of ecological integrity from a set of ecological measures that can be estimated from ecological inventory data.

Introduction

The goal of this project is to develop a system that helps CONABIO and other institutions to assess the status of ecosystems across the Mexican territory and provide insights about the risks that pose future developments to the ecosystems. For this purpose we define ecological integrity as “the capability of supporting and maintaining a balanced, integrated, adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of the natural habitat of the region” (Karr and Dudley 1981). Our main challenge is to develop a measure that captures this concept. The most common values considered to evaluate an ecosystem integrity are measures like the number of species, selection of physical and chemical parameters, or analysis of risk factors. These are important measures to characterize parts of an ecosystem but fail to capture the ecosystem as a whole.

Bayesian belief networks (BBN) are useful tools for modeling ecological predictions and identifying potential conservation risks in development (Marcot et al. 2006; McCloskey, Lilieholm, and Cronan 2011). We propose a way to estimate the concept of ecological integrity, or ecosystem

health, and its relation with human activities with a BBN model built from expert opinion.

One first problem is the complexity of the concept of ecological integrity so it has been important to discuss the idea among people from different areas of expertise. We have to balance this knowledge with data insights in order to get a model that captures the hidden components that keep an ecosystem in optimal conditions.

This paper reports preliminary results, work in progress and open questions toward the goals of this project.

The model

As a foundation to model the concept of ecological integrity (EI) we use three concepts called: naturalness (N), self-organization (SO), and stability (S). We see these concepts as emergent properties of an ecosystem and are modeled as latent variables.

Naturalness measures how close is an ecosystem to its pristine condition, i.e. how free from human intervention. Self-organization is the ability of an ecosystem to maintain some sort of macroscopic order. Finally, stability is the ability of an ecosystem to maintain its natural structure and operation even after it is altered.

These three concepts emerge from the combination of a set of variables that are derived from ecological inventory data. The relations are as follows: naturalness is given by co-existence (CE), evenness (E), compensation (C), functional diversity (FD), connectivity (CN), and complexity (CX); self-organization by regularity (R), functional diversity, connectivity, and complexity; while stability is given by resistance (RS), occupancy (O), prevalence (P), and compensation.

One problem is that these variables are not comparable among regions because their range of values is closely tied to the type of ecosystem. For this purpose we added Holdridge life zones (LZ) (Holdridge 1947; Díaz-Maeda 2012) to the model as an extra variable that works like a standardization factor.

Another important factor to evaluate ecological integrity is the presence/absence of human alterations, this is represented by the latent variable human impact (HI) which groups the following measures: population density (PD), accessibility (A), landscape transformation (LT), and power infrastructure (PI).

*The National Commission for Knowledge and Use of Biodiversity (CONABIO) is a permanent interdepartmental commission in Mexico. The mission of CONABIO is to promote, coordinate, support and carry out activities aimed at the conservation of biodiversity and its sustainable use. This work is the combined effort of two areas at CONABIO: the Coordination of the Decision Support System on Impacts to the Biodiversity (SIESDIB) and the Coordination of Eco-informatics.

Copyright © 2015, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

The variables connectivity loss (CL), biotic homogenization (BH), trophic downgrading (TD), and extinction debt (ED) is a third set of variables that try to capture some of the degradations processes that show damaged ecosystems. We use these variables as an aid to evaluate the ecological integrity of a region.

Although, in the ecological literature there is still an ongoing discussion about how to quantify the concepts mentioned above, the SIESDIB team believes the current definitions are good enough for our purposes. A discussion of all this concepts is outside the scope of this paper. From the modeling perspective our challenge is to find out if these measurements are good proxies to characterize the ecological integrity of an ecosystem and the adverse effects of human development.

In summary, we are framing our goal as that of modeling the joint distribution for all the aforementioned variables and the value of ecological integrity. Our BBN is modeled after expert knowledge in three ways. First, the definition of all the input nodes comes from ecological theory; second, the topology of the network comes from a combination of ecosystems theory and data considerations; and finally, the evidence of Ecological integrity levels is taken from experts' beliefs on the conditions of some areas.

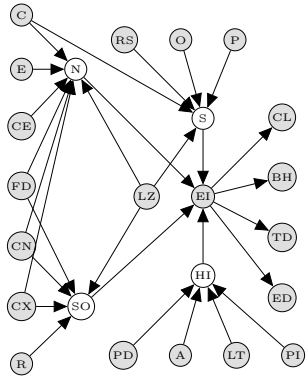


Figure 1: Bayes network proposed to evaluate ecological integrity. Shaded nodes represent measurable or known quantities.

Remark. All variables were discretized on three intervals. For our purpose we assume that the variable ecological integrity has three values: High, medium, and low; and to illustrate the difference between these three groups, expert opinions help to label 2,076 data points.

Results

As we mentioned above, there are 2,076 hand labeled points with a value for ecological integrity, 634 have low EI, 565 medium EI, and 877 high EI. Each point corresponds to a centroid of 25 Km² square, these points are a subset of a grid for all Mexico. For each point, we have at most 21 values for the measurable variables, some of the points contain missing data.

The variables C, E, CE, FD, CN, CX, R, RS, O, P, CL, BH, TD, and ED are indexes calculated from species distri-

Table 1: Description of train and test data used for fit the model.

	High	Medium	Low	Total
Train data	625	400	475	1500
Test data	252	165	159	576

Table 2: Confusion matrix for the 576 test points.

Actual	High	Medium	Low
High	246	3	3
Medium	8	148	9
Low	0	6	153

bution models (SDM) of mammals, estimated with MaxEnt and post-edited by experts.

The conditional probability tables for the BBN are estimated using the implementation of Expectation-maximization in Netica (Norsys Corporation 2014). We split the dataset of 2,076 labeled points into 1,500 training data points, and 576 data points to test the estimated model. In Table 1, the data distributions are described.

The confusion matrix from the test is presented in Table 2, there is a 5% of overall error. It can be noticed that there is no confusion between high and low values, this is important because it means that confusions are mild, that all the confusion is either between high and medium values or between low and medium values.

In contrast when a naïve network is considered, the overall error is about 78% for EI with the same dataset.

Discussion

We are trying to model and understand the impact of human activities in an ecosystem's health so that we can identify the potential ecological risks of, for example, infrastructure projects. So, from a conservation point-of-view an important question is what are the most important factors to keep an ecosystem functioning. As a possible answer to this question we have proposed to analyze the sensitivity of the variable ecological integrity with respect to the variables C, E, CE, FD, CN, CX, R, RS, O, and P, given the life zone; assuming that a life zone acts like an ecosystem unit. This could be a good estimation of what variable is the most important for each ecosystem and focus protection programs on preserving that factor.

One source of concern is that this model mostly uses prey-predator interactions, but there are many other sources of information, like vegetation, that should be considered. However, the larger the model the more difficult it is to understand it and we need an explanatory model.

Our current model may be useful to evaluate changes on ecological integrity but fails to address the relations between human alterations and particular ecosystem degradation processes. This is due to the network topology. In this context, our model has been developed to capture most of the expert knowledge while keeping it interpretable.

Going from theoretical concepts to measurable indexes is a difficult task, which involves not only computational

considerations, but inevitably involves subjectivity. The final system must generate confidence in the final users, and before that it must have the confidence of the biologists in the sense that all the necessary concepts are included. There is a trade off between including all the theory and building a usable and interpretable system. In order to find a balance between these two we need better ways to evaluate this model. Evaluation, however, is challenging because there is no ground truth. We would like to evaluate this model at a larger scale, but we still need to define what does it mean to do that, one way would be to ask more specialists to mark more sample areas with different labels of ecological integrity, that would help to evaluate the predictions on EI. We also need to evaluate the explanatory capabilities of the model, that is, how well the dynamics on the BBN are in tune with the dynamics in the real world.

References

- Díaz-Maeda, P. 2012. Zonas de vida de Holdridge para México. Resolución de 926 m (0.00833333338 grados decimales). Technical report, CONABIO.
- Holdridge, L. R. 1947. Determination of world plant formations from simple climatic data. *Science* 105(2727):367–368.
- Karr, J., and Dudley, D. 1981. Ecological perspective on water quality goals. *Environmental Management* 5(1):55–68.
- Marcot, B. G.; Steventon, J. D.; Sutherland, G. D.; and Mccann, R. K. 2006. Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Canadian Journal of Forest Research* (36):3063–3074.
- McCloskey, J. T.; Lilieholm, R. J.; and Cronan, C. 2011. Using Bayesian belief networks to identify potential compatibilities and conflicts between development and landscape conservation. *Landscape and Urban Planning* 101(2):190–203.
- Norsys Corporation. 2014. Netica application v.5.15.