Environmental Modelling & Software 81 (2016) 1–11

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

System learning approach to assess sustainability and forecast trends in regional dynamics: The San Luis Basin study, Colorado, U.S.A

Alejandra M. González-Mejía [1], Tarsha N. Eason, Heriberto Cabezas*

Sustainable Technology Division, National Risk Management Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, 26 W. Martin Luther King Drive, Cincinnati, OH 45268, USA

ARTICLE INFO

Article history:
Received 18 May 2015
Received in revised form
31 October 2015
Accepted 2 March 2016
Available online 22 March 2016

Keywords:
Artificial neural network
Fisher information
Forecast
Prediction
Baseline scenario
Sustainability
Regional system

ABSTRACT

This paper presents a methodology that combines the power of an Artificial Neural Network and Information Theory to forecast variables describing the condition of a regional system. The novelty and strength of this approach is in the application of Fisher information, a key method in Information Theory, to preserve trends in the historical data and prevent over fitting projections. The methodology was applied to demographic, environmental, food and energy consumption, and agricultural production in the San Luis Basin regional system in Colorado, U.S.A. These variables are important for tracking conditions in human and natural systems. However, available data are often so far out of date that they limit the ability to manage these systems. Results indicate that the approaches developed provide viable tools for forecasting outcomes with the aim of assisting management toward sustainable trends. This methodology is also applicable for modeling different scenarios in other dynamic systems.

Published by Elsevier Ltd.

1. Introduction

Determining how to assess and manage aspects of a system towards a sustainable path is one of the most critical challenges in sustainability science (Kates, 2011). Accordingly, developing future scenarios is an important management need (Boyko et al., 2012). A key element in this effort is developing plausible projections given an assessment of historical data to identify trends representing typical system conditions. Artificial neural networks offer great promise for handling time series forecasting and form the basis of our approach.

Artificial neural networks (ANN) are powerful data-driven modeling techniques based on iterative algorithms that have the ability to estimate a function from a great array of dependent or independent inputs (Zaihong et al., 2012). They provide an advantage over many traditional statistical approaches because there is no need to find a causal relationships among variables. ANNs are useful in a variety of applications from data processing and regression analysis to adaptive learning and forecasting, yet like any estimation technique, ANNs are prone to over-fit forecasted data (Krogh, 2008). In response to this challenge, we developed a methodology to "bound" the forecast and ensure that ANN projections describe typical patterns found in the historical data. Here, we use Fisher information to characterize system condition as defined by patterns in variables.

Fisher information is a measure of the amount of information about system behavior that is present in data (Fisher, 1922). While, its roots are in statistical estimation theory, Fisher information has been adapted to provide a means of monitoring variables which characterize system behavior by collapsing them into an index that can be monitored to assess dynamic order (patterns in data), system regimes and regime shifts (Fath et al., 2003; Karunanithi et al., 2008). For this application, Fisher information was initially employed to assess the stability of the patterns (how much they are changing) in the historical data. The trends in the index also served as a constraint to ensure that level of stability found in the historical data is preserved during the time series projection. The central purpose of this research effort is to develop a data-driven forecasting method that ensures that the projections contain patterns

* Corresponding author.
E-mail addresses: gonzalez.alejandra@epa.gov (A.M. González-Mejía), eason.tarsha@epa.gov (T.N. Eason), cabezas.heriberto@epa.gov (H. Cabezas).

[1] Research Fellow, Oak Ridge Institute for Science and Education and Research Fellow presently at Bangor University, Wales, UK. Sêr Cymru National Research Network for Low Carbon, Energy and Environment, School of Environment, Natural Resources and Geography.

http://dx.doi.org/10.1016/j.envsoft.2016.03.002
1364-8152/Published by Elsevier Ltd.

consistent with the historical behavior of the system.

As a study case, we applied the method to data collected on a regional system in south-central Colorado. This work is a follow up to the San Luis Basin (SLB) Sustainability Metrics Project initiated in 2006 by the U.S. Environmental Protection Agency (USEPA), Office of Research and Development (USEPA, 2010). Using publically available data and four integrated metrics, researchers computed these metrics to assess trends in key aspects of sustainability: ecological footprint representing environmental burden (Hopton and White, 2012), green net regional product accounting for economic well-being (Heberling et al., 2012), emergy which measures energy resources and flow through the system (Campbell and Garmestani, 2012), and Fisher information captures the stability of the system condition (Eason and Cabezas, 2012). The research team found that the dynamics of the region were fairly steady yet slowly moving away from sustainability. The stability was due to such drivers as relatively low diversity in land use, high bioavailability and a relatively stable economic base (Hopton et al., 2010). One of the primary goals of the study was help inform decision making on land and environmental management issues. However, data availability presented a key limitation and resulted in the production of sustainability metrics with a lag of three or more years. Therefore, these calculations were not the optimal for stakeholders to make near-time informed management decisions (Heberling and Hopton, 2014). Accordingly, the team highlighted the need for forecasting methodologies to aid in assessing plausible scenarios and alternative futures. This research article presents a method of developing and applying a forecasting methodology to aid in producing scenarios for the SLB. The approach presented teams a dynamic autoregressive artificial neural network (DARX) with Fisher information in order to extrapolate time series from 2011 to 2025 while preserving the trends found from 1969 to 2010. This approach simulates a “business as usual” scenario that eliminates both unexpected patterns and model over fitting. Although this initial installment focuses on developing a baseline scenario, the methods can be adapted to assess alternate futures and other types of systems.

2. Methods

The core method involves two main processes: (1) Characterizing the system given data series from 1969 to 2010 and (2) forecasting of a baseline scenario for the period 2011–2025. Process 1 involves using Fisher information to assess trends in the system variables. Process 2 consists in designing (simulation, calibration, cross-validation, test, and optimization) and applying the architecture of the constrained neural net to extrapolate trends found in process 1. Both processes incorporate techniques to handle data quality and quantity issues (e.g., sparseness) inherent in real systems by pre and post processing the variables describing an area of study.

2.1. Artificial neural nets

Inspired by biological neural networks, researchers from different disciplines design artificial neural nets in order to address a variety of problems, such as pattern classification, clustering, optimization, and prediction among others (Jain et al., 1996). The analogy between the artificial and the biological system is the high capacity of interconnection (neurons), learning (what is more probable), generalizing (rules), and making decisions (model). These characteristics provide a powerful tool for answering questions about the future, based on past behavior. In that sense, neural nets are an excellent forecasting choice that does not need prior knowledge of the relationship among data (not always evident in observations), and that infers from examples in spite of noise content (Darbellay and Slama, 2000). Further, neural nets have been applied successfully to model a wide variety of real-world applications, such as: the prediction of the stock market index (Guresen et al., 2011), forecasting energy consumption (Kankal et al., 2011) and estimating wind power output (Tu et al., 2012). Other examples include the forecast of environmental variables such as the net ecosystem metabolism (a proxy for system tropic state) within a freshwater wetland (Young II et al., 2011), water quality (Zaihong et al., 2012), and municipal waste generation (Antanasijević et al., 2013). As mentioned, Artificial Neural Networks (generally called neural nets) have demonstrated promise in the forecasting arena (Maier et al., 2010; Guresen et al., 2011; Zaihong et al., 2012; Antanasijević et al., 2013). However, when used for projections, they are sometimes prone to over fitting and produce results that are outside of the realm of realistic system behavior (Voyant et al., 2011). Hence, it is critical that mechanisms be developed to help establish appropriate constraints on projections.

2.2. Fisher Information

Integrated indicators or metrics have proven a very useful tool to present a scientific, straightforward, consistent, and multidimensional view of a system (Bond and Morrison-Saunders, 2010). One critically important characteristic is the patterns in variables describing the condition of the system. Fisher information (Fisher, 1922), denoted I in Equation (1), was selected to measure the dynamics and stability (Pawlowski et al., 2005) of the system under study.

$$I = \int \frac{1}{p(s)} \left[\frac{dp(s)}{ds}\right]^2 \mathrm{d}s \tag{1}$$

Derived from Information Theory, Fisher information is a statistical quantity that measures the degree to which a parameter (e.g., s: state of a system) can be estimated (Frieden, 2004; Mayer et al., 2007) from a given data set, where p(s) is the probability of observing a particular state (condition) of the system. Fisher information (I) has been adapted to assess dynamic order (patterns in data) and applied to various types of systems for evaluating such aspects as urban and regional sustainability (Eason and Cabezas, 2012; Gonzalez-Mejia et al., 2012; Gonzalez Mejia et al., 2014). Further details on the derivation and computation of the index may be found in Karunanithi et al., 2008 and the USEPA (2010). Simply stated, higher I implies greater predictability of the system's state at a point in time (Mayer, 2008) and higher information content in the data set.

According to the Sustainable Regimes Hypothesis, well-functioning systems are predictable. The patterns may fluctuate within a limited range exist in regimes but they remain relatively steady through time; accordingly, the time averaged Fisher Information is constant, ($dI/dt \approx 0$) (Karunanithi et al., 2008). Taking this a bit further, a system is considered to be in a stable regime when values are within two standard deviations ($2sd(I)$) from the mean value I computed for the historical data (Equation (2)). A regime shift exists only when the I drops by more than two standard deviations from I as shown in Equation (3).

$$\text{Stable period} \Rightarrow \langle I \rangle - 2 \cdot sd(I) \le I(t) \le I + 2 \cdot sd(I) \tag{2}$$

$$\text{Regime shift} \Rightarrow I(t) < I - 2 \cdot sd(I) \tag{3}$$

The choice of two standard deviations from the mean as a criteria is based on an application of a Chebyshev's Theorem, which