Environmental Modelling & Software 83 (2016) 74–87

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Demonstrating correspondence between decision-support models and dynamics of real-world environmental systems

Ray Huffaker [a], Rafael Muñoz-Carpena [a,*], Miguel A. Campo-Bescós [b], Jane Southworth [c]

[a] Department of Agricultural and Biological Engineering, University of Florida, 281 Frazier Rogers Hall, Gainesville, FL, 32611-0570, USA
[b] Projects and Rural Engineering Department, Public University of Navarre, Ed. Los Olivos, Pamplona, Spain
[c] Geography Department, University of Florida, Gainesville, FL, USA

ARTICLE INFO

Article history:
Received 11 March 2015
Received in revised form
2 February 2016
Accepted 25 April 2016

Keywords:
Model evaluation
Nonlinear dynamics
Phase space reconstruction
Extreme value statistics

ABSTRACT

There are increasing calls to audit decision-support models used for environmental policy to ensure that they correspond with the reality facing policy makers. Modelers can establish correspondence by providing empirical evidence of real-world behavior that their models skillfully simulate. Since real-world behavior—especially in environmental systems—is often complex, credibly modeling underlying dynamics is essential. We present a pre-modeling diagnostic framework based on *Nonlinear Time Series* (*NLTS*) methods for reconstructing real-world environmental dynamics from observed data. The framework is illustrated with a case study of saltwater intrusion into coastal wetlands in Everglades National Park, Florida, USA. We propose that environmental modelers test for systematic dynamic behavior in observed data before resorting to conventional stochastic exploratory approaches unable to detect this valuable information. Reconstructed data dynamics can be used, along with other expert information, as a rigorous benchmark to guide specification and testing of environmental decision-support models corresponding with real-world behavior.

1. Introduction

Real-world environmental systems are complex, ever-changing, and beyond anyone's capacity to model closely. Despite this, policymakers rely on decision-support models to regulate real-world environmental systems. Faulty model representations lead to ineffective and wasteful policies (Saltelli and Funtowitz, 2014). Consequently, there are increasing calls to review decision-support models used for public policy (Saltelli and Funtowitz, 2014; Oreskes et al., 1994). For example, the European Commission's Joint Research Centre formally audits models used in impact assessments of EU initiatives, legislation, and policy (Joint Research Centre).

If modelers are required to demonstrate correspondence of their models with real-world behavior—as recommended by Oreskes et al. (1994) and Saltelli and Funtowitz (2014)—what is a reasonable burden of proof? The literature is clear that modelers cannot be reasonably required to verify their models as accurate representations of reality since verification is a logical impossibility. Moreover, demonstrating a 'good fit' between model output and observed data does not constitute validation since the possibility remains that other models with very different structures and representations of reality can be parameterized to provide good fits (Oreskes et al., 1994; Rykiel, 1996; Hornberger and Spear, 1981). Because real-world environmental systems are in a constant state of flux, we propose that a reasonable burden of proof would require auditing modelers to present persuasive empirical evidence of real-world dynamic behavior that their models skillfully simulate.

What constitutes persuasive evidence of real-world dynamic behavior? Most would agree that observed data provide an essential portal to real-world environmental systems to which there is only limited access. Evidence of real-world dynamic behavior must be drawn somehow from data that often exhibit a highly volatile, irregular, and random appearance.

Past work in the environmental and resource management literature presumes that irregular data reflect two major sources of uncertainty: (1) 'Real' uncertainty due to the inherent randomness and natural variation of real-world biophysical processes, and (2) 'Perceived' uncertainty due to decision-maker's limited perceptions of reality (Uusitalo et al., 2015; Feder, 1979; Dixon and Howitt, 1980; Johnson and Pasour, 1981). Typically, stochastic decision-support

* Corresponding author.
E-mail address: carpena@ufl.edu (R. Muñoz-Carpena).

http://dx.doi.org/10.1016/j.envsoft.2016.04.024
1364-8152/

models are believed to be required to capture this management uncertainty (Uusitalo et al., 2015), while deterministic methods are believed to ignore or assume away this uncertainty, and thus incapable of representing risk-responsive management behavior (Karp, 1987).

Contrary to this common presumption, principles of randomness—set out in the philosophy of science literature—reveal that random-appearing data are not evidence that stochastic decision-support models are required to model uncertainty (Horan, 1994; Huffaker, 1998). Stochastic modeling is required only when real-world processes are 'physically random' (indeterministic). Since indeterministic processes are irreducibly probabilistic, causal relationships do not exist to support deterministic formulations. Alternatively, deterministic modeling is feasible when real-world processes are physically nonrandom (deterministic). Deterministic processes are governed by laws such that, given them, nothing else can happen. Importantly, apparent 'mathematically random' process output can be generated by both indeterministic and deterministic processes. Indeed, breakthroughs in nonlinear dynamics demonstrate that irregular and apparently-random dynamic behavior can emerge endogenously from deterministic, low-dimensional and nonlinear interactions among system variables (Glendinning, 1994; Medio, 1993). Equally importantly, one cannot work backward from observing mathematically random output to prove whether it was generated by an indeterministic or deterministic process.

In sum, random-appearing environmental data are not evidence of indeterministic biophysical processes requiring stochastic modeling. The interesting possibility remains that deterministic decision-support models might generate observed uncertainty endogenously. In this paper, we propose a pre-modeling diagnostic framework to empirically test observed data for this possibility. The framework is based on *Nonlinear Time Series* (*NLTS*) methods—developed in mathematical physics—that reconstruct system dynamics from time-series data on a single variable (Schreiber, 1999). The *NLTS* framework extends Larsen et al. (2014) who proposed an exploratory approach in which modelers experiment with simple nonlinear deterministic structures to find those capable of generating complex observed environmental dynamics (Larsen et al., 2014). We address the open question of how to diagnose whether real-world environmental dynamics are nonlinear and deterministic in the first place. *NLTS* data diagnostics provide mathematically and statistically rigorous evidence of real-world dynamic behavior, and can guide specification and testing of deterministic environmental decision-support models corresponding with diagnosed behavior.

We present an intuitive description of each component of the proposed *NLTS* data diagnostics framework, and illustrate it with a case study of saltwater intrusion into coastal wetlands in Everglades National Park, Florida, USA.

2. Framework for *NTLS* pre-modeling data diagnostics and data-informed modeling

The proposed *NLTS* diagnostic framework is summarized in Fig. 1. In a nutshell, we first apply *Singular Spectrum Analysis* (*SSA*) (Elsner and Tsonsis, 2010; Ghil et al., 2002; Golyandina et al., 2001)—a signal processing technique—to separate an observed time series into signal (structured variation) and noise (unstructured variation). *SSA* is a data-adaptive signal processing approach that can accommodate highly irregular (anharmonic and potentially non-sinusoidal) oscillations in signals (Elsner and Tsonsis, 2010; Ghil et al., 2002; Golyandina et al., 2001; Vautard, 1999). When the time series is converted to anomalies from the mean, and the *Toeplitz* method of *SSA* is applied, signal strength is measured as the fraction of variation explained in the observed time series from its mean (Ghil et al., 2002; Golyandina et al., 2001). We test a sufficiently strong signal for low-dimensional nonlinear dynamic structure with *Phase Space Reconstruction* (Schreiber, 1999; Kantz and Schreiber, 1997) and *Surrogate Data* methods (Theiler et al., 1992). A low dimensional nonlinear dynamic is defined as driven by a small number (low-dimension) of nonlinear components. We propose a novel application of *Extreme Value Statistics* (Katz, 2010) to model noise separated from an observed time series

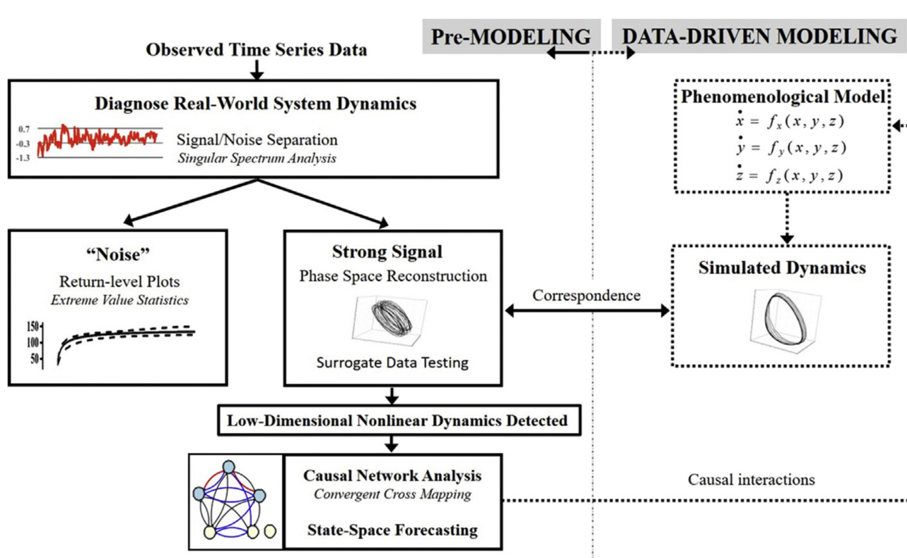

Fig. 1. Framework for NLTS pre-modeling data diagnostics and data-driven modeling. We first separate an observed time series into signal (structured variation) and noise (unstructured variation) components, and use the separated signal to reconstruct the dynamics of the real-world system generating it. We test whether empirically-detected dynamics are deterministic, low-dimensional, and nonlinear, and then whether attractors reconstructed from multiple observed signals are causally interrelated. We apply this information to simulate empirically-detected dynamics with a phenomenological (data-driven) model composed of polynomial ordinary differential equations. We complete the diagnostics by modeling unstructured noise separated from the observed time series probabilistically with Return-level Plots. These data diagnostics inform the structure of mechanistic models capable of simulating real-world system dynamics.