OXFORD

NONLINEAR TIME SERIES ANALYSIS WITH R

Ray Huffaker · Marco Bittelli · Rodolfo Rosa

Nonlinear Time Series Analysis with R provides a practical guide to emerging empirical techniques allowing practitioners to diagnose whether highly fluctuating and randomly appearing data are most likely driven by random or deterministic dynamic forces. It joins the chorus of voices recommending 'getting to know your data' as an essential preliminary evidentiary step in modelling. Time series are often highly fluctuating with a random appearance. Observed volatility is commonly attributed to exogenous random shocks to stable real-world systems. However, breakthroughs in nonlinear dynamics raise another possibility: highly complex dynamics can emerge endogenously from astoundingly parsimonious deterministic nonlinear models. Nonlinear time series analysis (NLTS) is a collection of empirical tools designed to aid practitioners detect whether stochastic or deterministic dynamics most likely drive observed complexity. Practitioners become 'data detectives' accumulating hard empirical evidence supporting their modelling approach.

This book is targeted at professionals and graduate students in engineering and the biophysical and social sciences. Its major objectives are to help non-mathematicians with limited knowledge of nonlinear dynamics to become competent in NLTS; and in this way to pave the way for NLTS to be adopted in the conventional empirical toolbox and core coursework of the targeted disciplines. Consistent with modern trends in university instruction, the book makes readers active learners with hands-on computer experiments in R code, directing them through NLTS methods and helping them understand the underlying logic. The computer code is explained in detail so that readers can adjust it for use in their own work. The book also provides readers with an explicit framework condensed from sound empirical practices recommended in the literature—that details a step-by-step procedure for applying NLTS in real-world data diagnostics.

To view and download the computer code please visit: http://www.dista.unibo.it/~bittelli/

Ray Huffaker is a professor at the Department of Agricultural and Biological Engineering, University of Florida, USA.

Marco Bittelli is a professor at the Department of Agricultural Sciences, University of Bologna, Italy.

Rodolfo Rosa is a professor at the National Research Council (CNR), Institute for Microelectronics and Microsystems, Section of Bologna, Italy.

Cover image: Shutterstock.com

ISBN 978-0-19-880825-1

9 780198 808251

OXFORD
UNIVERSITY PRESS

www.oup.com

Huffaker · Bittelli · Rosa

NONLINEAR TIME SERIES ANALYSIS WITH R

OXFORD

Nonlinear Time Series Analysis with R

Ray Huffaker

Department of Agricultural and Biological Engineering,
University of Florida, USA

Marco Bittelli

Department of Agricultural Sciences, University of Bologna, Italy

Rodolfo Rosa

National Research Council (CNR), Institute for Microelectronics and Microsystems,
Section of Bologna, Italy

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2017

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2017953063

ISBN 978–0–19–878293–3 (hbk.)
ISBN 978–0–19–880825–1 (pbk.)

DOI 10.1093/oso/9780198782933.001

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

NONLINEAR TIME SERIES ANALYSIS WITH R

Preface

Nonlinear Time Series Analysis with R joins the chorus of voices recommending 'getting to know your data' as an essential preliminary evidentiary step in scientific inquiry. Time series are often highly fluctuating, with a random appearance. Observed volatility is commonly attributed to exogenous random shocks to stable real-world systems. Consequently, investigators are driven to reproduce volatility with a variety of linear-stochastic and probabilistic methods. However, breakthroughs in nonlinear dynamics raise another possibility: highly complex dynamics can emerge endogenously from astoundingly parsimonious deterministic models.

Nonlinear time series analysis (NLTS) is a collection of empirical tools that allow practitioners to diagnose whether observed data are most likely generated by stochastic or deterministic dynamics. In particular, practitioners can use NLTS in an attempt to reconstruct, characterize and model real-world dynamics from a single time series or multiple causally interactive time series. This information can be used, along with scientific principles and other expert information, to guide the specification of mechanistic models used to build theory or to support high-stakes public policy. Models used for public policy are increasingly subjected to formal government audit to ascertain how well they correspond to reality. The compatibility of audited models with NLTS-detected dynamics offers evidence of proper model specification.

This book targets students and professionals in physics, engineering, biology, agriculture, and economics and other social sciences. Our major objectives are to put key concepts of NLTS – developed in the mathematical physics literature – within the operational reach of non-mathematicians with limited knowledge of nonlinear dynamics, and in this way to pave the way for NLTS to be adopted in the conventional empirical toolbox and core coursework of other disciplines. Consistent with modern trends in university instruction, the book makes readers active learners with hands-on computer experiments in `R` code directing them through NLTS methods. The computer code is explained in detail so that readers can adjust it for use in their own work. The book also provides readers with an explicit framework – condensed from sound empirical practices recommended in the literature – that proposes a strategy for implementing NLTS in real-world data diagnostics. Practitioners become 'data detectives', accumulating hard empirical evidence directing scientific inquiry.

We used `R 3.3.1` and the following packages to construct the code in this book:

`animation 2.5; boot 1.3-18; crqa 1.0.6; deSolve 1.12; extRemes 2.0-7; fields 8.4-1; fractal 2.0-1; glmnet 2.0-5; gplots 3.0.1; graphics 3.3.1; igraph 1.0.1; MESS 0.4-3; mpoly 1.0.3; multispatialCCM 1.0; nonlinearTseries 0.2.3; pdc 1.0.3; pdist 1.2; phaseR 1.3; plotrix 3.6-3; ppls 1.6-1; psych 1.7.3.21; rgl; Rssa 0.13-1; scatterplot3d 0.3-37; stats 3.5.0; tseriesChaos 0.1-13; tseriesEntropy 0.6-0`

Acknowledgements

We would like to thank the following institutions for institutional and financial support: the University of Florida (USA), the University of Bologna (Italy) and the National Research Council (Italy). MB thanks Roberto Olmi and RR thanks Simone Giannerini for collaborative and fruitful research over many years of friendship and collaboration. We also wish to thank our students, whose feedback helped us to improve the material presented in this book. RGH gratefully acknowledges Dr Gerhard Schiefer for providing numerous opportunities to present this material and develop productive international collaborations at the annual IGLS-FORUM 'System Dynamics and Innovation in Food Networks', and Maurizio Canavari, Ernst Berg, Rafael Muñoz-Carpena, Klaus Frohberg and Miles Medina for valuable input and collaborations. We thank Sonke Adlung, Ania Wronski and especially Mac Clarke for expert editorial support. And thanks for their endearing support to Ann Huffaker, Andrea Vogt and AnnaMaria Bononcini.

Contents

1 Why Study Nonlinear Time Series Analysis?

```
'"Data!data!data!" he cried impatiently. "I can't make bricks without clay."'
                     Arthur Conan Doyle, The Adventure of the Copper Beeches
```

1.1 Introduction

Nonlinear time series analysis (NLTS) requires time series data on only a single variable to diagnose, reconstruct, characterize and model the dynamics of the real-world system generating the data. This is possible because any single variable in an interdependent dynamic system encodes the history of its interactions with other system variables. The famous naturalist John Muir (1911) intuited this result in the early twentieth century, observing that: 'When we try to pick something up by itself, we find it hitched to everything else in the universe.'

Perhaps the most compelling reason for studying NLTS is it that facilitates scientific inquiry. However, demonstrating how NLTS data diagnostics fit into the scientific method is complicated by the lack of consensus regarding what that method is in the first place. The *Stanford Encyclopedia of Philosophy* concludes that 'there is not any unique, easily described scientific method' because 'scientific activity varies so much across disciplines, times, places, and scientists that any account which manages to unify it all will either consist of overwhelming descriptive detail, or trivial generalizations' (Andersen and Hepburn, 2016).

A conventional view is that scientific inquiry cycles between inductive reasoning that converts detected regularities in data into testable hypotheses, and deductive reasoning that elevates those hypotheses to theory whose predicted consequences are corroborated by further observation or experimentation (Andersen and Hepburn, 2016). This view is roundly criticized as an overly simplistic representation of how science converts experience into knowledge. For one thing, there has been unending disagreement over epistemic issues, including the proper balance between empirical observation and deductive reasoning on the one hand and the requirements of corroboration on the other. Moreover, the view excludes important human, social and political aspects of science, including the talent, imagination and objectivity of scientists; the benefits of transdisciplinary collaboration; and disincentives created by science and political communities that discourage researchers from undertaking truly novel research projects

Nonlinear Time Series Analysis with R. Ray Huffaker, Marco Bittelli and Rodolfo Rosa, Oxford University Press (2017).
 DOI: 10.1093/oso/9780198782933.001.0001

(Haack, 1999; Geman and Geman, 2016). Also, it does not account for how science progresses cataclysmically as new paradigms overthrow old ones in scientific revolutions (Kuhn, 1962). The lack of consensus leaves us between two extremes: the narrow epistemic framework of conventional views and the nihilistic positions that science is merely politics with no epistemic authority.

Haack (1999) provides a middle ground:

> 'An adequate account of scientific knowledge and scientific inquiry must acknowledge a subtle interplay of logical, personal, and social aspects. The interplay begins at the beginning, of course, with talented individuals coming up with imaginative conjectures on which others build and which are subject to the scrutiny of the whole relevant community, and it is present at every stage. The warrant of any empirical proposition depends in part on experimental evidence, i.e., on what some individual observe(s) see(s) or hear(s), etc, and so, on how justified others are in thinking the observer(s) reliable.'

The chief epistemic value of this middle ground is a 'respect for evidence' that meets general standards of good inquiry, namely 'good, strong, supportive evidence and ... well-conducted, honest, thorough, imaginative inquiry' (Haack, 1999).

Data provide an evidentiary portal to the real world to which there is only limited access. For example, Charles Darwin consolidated several months of studying numerous marine and terrestrial samples collected in the Galapagos into his famous sketch of the evolutionary tree (Berra, 2009). In another example, Leonardo da Vinci compiled a lifetime of observing nature into notebooks explaining diverse behaviours, including water and sediment movement in river systems, waves in ponds, sound waves in air, and even whether spirits can speak (da Vinci, 1519). The contributions of both scientists are especially renowned because they detected patterns that were not obvious in the data, and processed that information into astounding knowledge of systematic natural behaviour that has withstood the test of time.

Haack (1999) sets 'realism' as the goal of scientific inquiry. The results of scientific inquiry must agree with reality if they are to be put to reliable use. For example, policymakers reasonably expect that theory supporting high-stakes public policy adequately represents the real world that they are charged with regulating. Otherwise, policies 'leave the real problem unaddressed, waste resources, and impeded learning' (Saltelli and Funtowitz, 2014). Recently, a 2010 US Congressional special hearing, *Building a Science of Economics for the Real World* (US Congress Subcommittee on Science and Technology, 2010), criticized the performance of macroeconomic models for failing to reproduce temporal patterns of booms and busts observed in the 2008 financial crisis. Congress concluded that 'if major crises are a recurrent feature of the economy then our models should incorporate this possibility', and expressed frustration that 'because our experts' way of looking at the economy left them blind to the crisis that was building, we were unprepared to deal with the crisis'. Most ominously, Congress broadly questioned why 'we continue to rely upon [theoretical models] for so many critical decisions, so much practical policy advice'.

An economist empanelled at the 2010 hearing recommended that policy models be formally audited by the National Science Foundation. An earlier recommendation by Oreskes *et al.* (1994) would place the burden on the modeller 'to demonstrate the degree of correspondence between the model and the material world it seeks to represent' when 'public policy and public safety are at stake'. The European Commission's Joint Research Centre conducts a formal audit of models used to assess the impacts of EU initiatives, legislation and policy.

NLTS facilitates well-conducted evidentiary scientific inquiry by providing a collection of mathematically rigorous procedures that help practitioners to extract information on real-world dynamics from observed data that often have a complex, highly variable and random appearance. Applied science disciplines conventionally presume that apparent randomness of volatile data must result from exogenous shocks to inherently stable dynamic systems (Feder, 1979; Uusitalo *et al.*, 2015), and turn to stochastic methods without further justification. Alternatively, the theory of randomness teaches us that mathematically random output can be generated by both physically random (indeterministic) and physically nonrandom (deterministic) processes (Horan, 1994), and breakthroughs in nonlinear dynamics demonstrate that parsimonious deterministic models can produce surprisingly irregular and complex behaviour (Glendinning, 1994).

The possibility of deterministic volatility should not be surprising. Many essential biophysical processes exhibit strong patterns of dynamic behaviour to align with environmental regularities. Eating and sleeping in animals (including humans), as well as leaf movements and photosynthetic reactions in plants, exhibit built-in circadian (roughly 24-hour) rhythms that sunlight adjusts to the local environment. Tidal transitions exhibit a tidal (12.4-hour) rhythm, and tidal amplitudes a lunar (29.5-day) rhythm. Climate exhibits regular (diurnal and seasonal) cycles, quasi-periodic cycles (e.g. El Nino) and highly irregular cycles (e.g. volcanic winters).

NLTS allows us to replace presumption of stochasticity with rigorous empirical evidence. The data themselves can serve as our initial guide for ascertaining whether observed volatility is driven by stochastic or deterministic nonlinear real-world dynamics. This distinction matters critically both for theory and its practical application. For example, the *efficient-markets hypothesis* in economics is based on the presumption that market instability results from exogenous shocks to otherwise stable markets. The hypothesis holds that markets tend to equilibrate in response to these shocks as economic agents process all available information in adjusting supply to demand conditions (Fama, 1970). Observed volatility reflects corrective supply and demand adjustments. The hypothesis supports *laissez faire* market policies that do not interfere with corrective adjustments.

The failure of the efficient-markets hypothesis to predict the 2008 financial crisis awoke the profession to another possibility: real-world markets do not naturally equilibrate but may be inherently unstable. Consequently, *The Economist* recommended that 'like physicists, [economists] should study instability instead of assuming that economies naturally self-correct' (Economist, 2016*a*). In striving to understand the economics of market instability, economists returned to the earlier work of Minsky (1992), who developed the *financial instability hypothesis* explaining how financial booms systematically breed their own busts (see also (Economist, 2016*b*)). According to this hypothesis, markets do not provide a natural corrective mechanism, and public intervention should be geared to smoothing systematic boom and bust cycles, and buffering their negative impacts on consumers and producers (Huffaker *et al.*, 2016*a*).

The 2008 crisis led *The Economist* to endorse a more evidentiary approach for the profession in an article entitled *If Economists Reformed Themselves* (Economist, 2016*c*):

'Economists are good at reducing a complicated world to a few assumptions, then adding bells and whistles to make their models more realistic. But problems arise when they mistake the map for the territory ... In future, big data and new "machine learning" techniques could help test the relative power of competing theories.'

1.2 Nonlinear Dynamics and a Strategy for Applying NLTS

Sophisticated application of NLTS requires a firm theoretical foundation in basic nonlinear dynamics that we provide in Part 1 of this book (Chapters 2–5). We investigate how complex behaviour can arise from a simple deterministic nonlinear specification (Chapter 2), how system dynamics can be reconstructed from a single solution variable with ***phase space reconstruction*** techniques (Chapter 3) and how tools from nonlinear dynamics can be used to measure characteristics of reconstructed dynamics (Chapters 4 and 5).

Phase space reconstruction is the centrepiece of NLTS (Kantz and Schreiber, 1997). Phase space records the level of system (*state*) variables at each point in time. For example, assume that a dynamic system is fully expressed with two state variables x_t and y_t that change with time t. Phase space plots y_t against x_t, so that one point could be their values in the year 2000: (x_{2000}, y_{2000}). Running a line through this point and past and future points results in a phase space *trajectory* depicting the evolution of the system through time. In *dissipative* dynamic systems, trajectories converge to a subset of phase space where they oscillate aperiodically along an ***attractor*** – a geometric structure with noticeable regularity – forever after (Brown, 1996) . Consequently, an attractor encapsulates the long-term dynamic behaviour of the system. Prior to the 1980s, researchers assumed that data on all system variables were needed to construct phase space representations of system dynamics. This is problematic in practice because we cannot reasonably identify all of the variables at work in real-world dynamic systems, and we might not be able to adequately measure some variables that we can identify. Researchers then discovered that phase space dynamics could be reconstructed from time series data on a single variable using delay coordinates (Breeden and Hubler, 1990). As a result, we can potentially reconstruct real-world system dynamics from time series data on a single observed variable.

Consider a simple preliminary example of phase space reconstruction using data on snowshoe hares (prey) and lynx (predator) collected by the Hudson Bay company in Canada from 1845 to 1935 (Odum, 1953). The time series for each population cycles through time (Figure 1.1a). System dynamics are portrayed in phase space by plotting lynx against hares at each point in time (Figure 1.1b). The populations co-evolve repeatedly along a predator–prey cycle constituting the attractor for the system. A large predator population over-consumes available prey and crashes for want of food. This allows the prey to recover until pressed again by recovering predators, and cycling continues. A shadow version of the predator–prey attractor is reconstructed in phase space from a single variable by plotting either the prey or predator population against its level a period later (Figure 1.1c). This is the ***time delay embedding*** method of phase

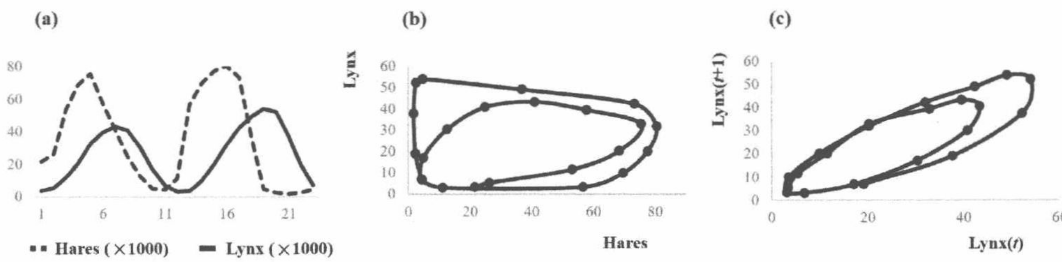

Fig. 1.1 Example of phase space reconstruction: (a) snowshoe hare (prey) and lynx (predator) population time series data; (b) phase space solution (from model to dynamics); (c) reconstructed 'shadow' phase space (from data to dynamics). The data were collected by the Hudson Bay Company in Canada from 1845 to 1935. The figure uses data from 1908 to 1930.

space reconstruction. Takens (1981) derived sufficient conditions guaranteeing that a *shadow* phase space preserves essential dynamic properties of the original phase space.

In this clear-cut example, the shadow dynamics reconstructed from the lynx population are already obvious in the cycling time series plots. The full potential of phase space reconstruction is appreciated when reconstructing deterministic dynamics concealed in a volatile and random-appearing time series. Consider an example from Kaplan and Glass (1995). There are two time series: x_t (plotted in Figure 1.2a) and y_t (plotted in Figure 1.2d). One of these time series is randomly generated, while the other is the solution to a deterministic difference equation. We apply three methods to distinguish between the two: casual observation, autocorrelation functions testing for linear correlations in the data and NLTS time-delay embedding plots. First, neither plot exhibits obvious behavioural patterns on casual observation – any dynamic structure is well concealed. Second, the autocorrelation functions do not exhibit patterns indicative of corresponding patterns in either time series (Figure 1.2b, e). Time-delay embedding plots succeed in distinguishing between random and deterministic dynamics. The time-delay plot using x_t is randomly distributed on the plane, correctly detecting that this series was randomly generated (Figure 1.2c). In contrast, the time-delay plot using y_t shows structure: a parabolic serial correlation missed by the autocorrelation functions (Figure 1.2f). Indeed, y_t was generated by the deterministic parabolic logistic map whose dynamics we investigate in Chapter 2:

$$x_{t+1} = 4x_t(1 - x_t), \qquad x_t \in [0, 1]$$

In Part 2 (Chapters 6–11), we focus on the application of NLTS to reconstruct real-world dynamic structure from observed time series data. We propose a strategy for implementing NLTS that is modified from previous versions in Huffaker (2015), Huffaker *et al.* (2016*b*) and (Huffaker *et al.*, 2016*a*) (Figure 1.3). We emphasize from the outset that NLTS is capable of reconstructing linear as well as nonlinear deterministic system dynamics, and of diagnosing the presence of linear stochastic dynamics. Our objective is not limited to finding evidence pointing to nonlinear deterministic structure, but extends to diagnosing the structure most closely corresponding to reality whether that be linear, nonlinear, deterministic or stochastic.

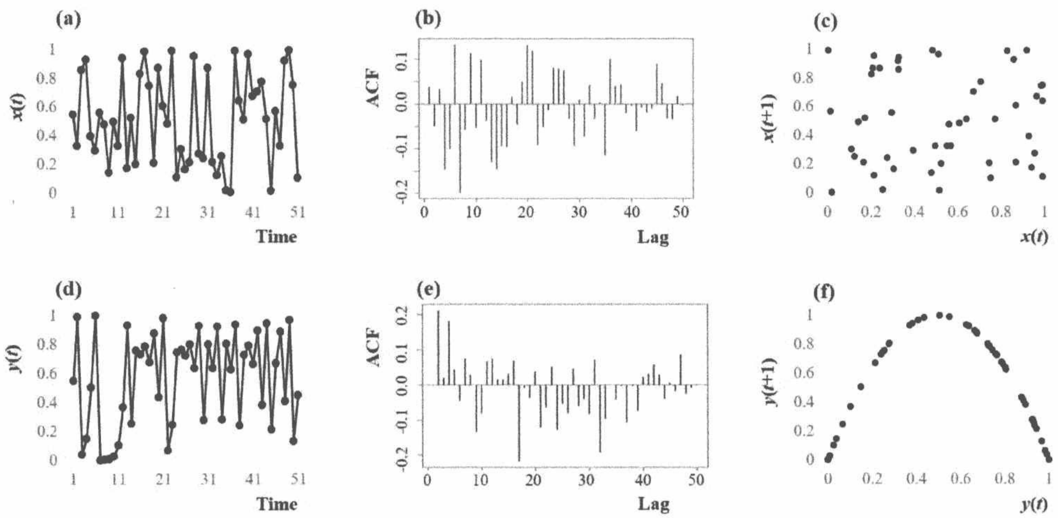

Fig. 1.2 Detection of nonlinear dynamics in data: (a) plot of fifty uniform random variates; (b) autocorrelation plot of random variates; (c) time-delay embedding plot of random variates; (d) plot of nonlinear logistic map; (e) autocorrelation plot for logistic map; (f) time-delay embedding plot for logistic map. Conventional linear methods fail to detect nonlinear deterministic structure in data.

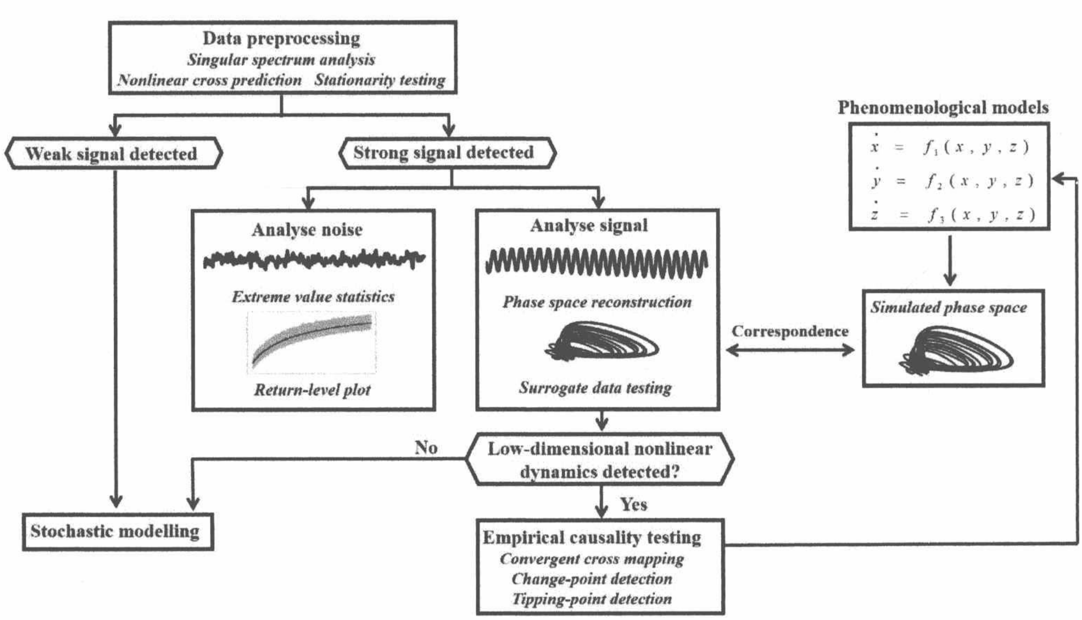

Fig. 1.3 NLTS diagnostics strategy.