Physica D 313 (2015) 37–50

Contents lists available at ScienceDirect

Physica D

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

Predictability of threshold exceedances in dynamical systems

Tamás Bódai *

Meteorological Institute, University of Hamburg, Grindelberg 5, 20144 Hamburg, Germany

HIGHLIGHTS

- Predictability of extremes in a chaotic dynamical system is examined.
- Stronger extremes are more predictable assuming no model errors.
- The skills of model and data-driven predictions are the same under certain conditions.
- The stability of trajectories determines the prediction skill in a nontrivial way.

ARTICLE INFO

Article history:
Received 6 August 2014
Received in revised form
19 August 2015
Accepted 22 August 2015
Available online 10 September 2015
Communicated by H.A. Dijkstra

Keywords:
Extreme event
Data-driven prediction
Precursory structure
Prediction skill
ROC curve
Finite-time Lyapunov exponent

ABSTRACT

In a low-order model of the general circulation of the atmosphere we examine the predictability of threshold exceedance events of certain observables. The likelihood of such binary events – the cornerstone also for the *categoric* (as opposed to probabilistic) prediction of threshold exceedances – is established from long time series of one or more observables of the same system. The prediction skill is measured by a summary index of the ROC curve that relates the hit- and false alarm rates. Our results for the examined systems suggest that *exceedances of higher thresholds are more predictable*; or in other words: rare large magnitude, i.e., extreme, events are more predictable than frequent typical events. We find this to hold provided that the bin size for binning time series data is optimized, but not necessarily otherwise. This can be viewed as a confirmation of a counterintuitive (and seemingly contrafactual) statement that was previously formulated for more simple autoregressive stochastic processes. However, we argue that for dynamical systems in general it may be typical only, but not universally true. We argue that when there is a sufficient amount of data depending on the precision of observation, the skill of a class of data-driven categoric predictions of threshold exceedances approximates the skill of the analogous model-driven prediction, assuming strictly no model errors. Therefore, stronger extremes in terms of higher threshold levels are more predictable both in case of data- and model-driven prediction. Furthermore, we show that a quantity commonly regarded as a measure of predictability, the finite-time maximal Lyapunov exponent, does not correspond directly to the ROC-based measure of prediction skill when they are viewed as functions of the prediction lead time and the threshold level. This points to the fact that even if the Lyapunov exponent as an intrinsic property of the system, measuring the instability of trajectories, determines predictability, it does that in a nontrivial manner.

1. Introduction

Extreme events have fundamental importance to life, as they are often associated with survival and losses. Extreme events to do with gains or amusement receive far less attention in general, dissociated from individual events. Rare and large magnitude events of interest arise in physical, technological, social, and other systems [1]. The classical theory of extremes in uncorrelated sequences (or, in sequences in which the auto-correlation is decaying sufficiently fast) [2–4] has a statistical orientation; it is not- and cannot be concerned with *prediction* or with uncovering mechanisms that can produce extremes; but it is rather concerned with e.g. expected return times, which can be useful in designing structures of a certain required life time, such as sea walls [5].

Since Newton revolutionized science, it has become a paradigm that predictions should be based on validated models. These models describing fluctuating phenomena often take the form of a system of differential equations, also referred to as a *dynamical system*. Since the work of Lorenz it has become clear that even though some phenomena can be modeled quite accurately,

* Tel.: +49 40 42838 9205.
E-mail address: tamas.bodai@uni-hamburg.de.

http://dx.doi.org/10.1016/j.physd.2015.08.007
0167-2789/

they can be inherently unpredictable because of their extreme sensitivity to initial conditions [6]. Such systems are called chaotic, characterized by positive Lyapunov exponents. This imposes a time horizon on predictions; beyond that only statistical properties can be robustly estimated, which is what classical extreme value theory is concerned with. In contrast with that, in our analysis we consider prediction lead times shorter than the decorrelation time in a time series.

In the context of model-based or *model-driven* predictions (MDP) equivalent with initial value problems for deterministic differential equations, like e.g. a weather forecast, one can often read that extremes are much harder to predict. Unfortunately a systematic study of the dependence of some appropriate prediction skill score – or a measure of predictability in a more general sense – of any model on the magnitude of events is still lacking. Inaccuracy of the model may be an important factor leading to such a dependence of its prediction skill on the event magnitude, beside details of its chaotic nature. In contrast, in pure *data-driven* prediction (DDP), model errors are not present, as the basis of the prediction (of any kind) does not involve a model in the form of equations or an algorithm, only observational/measurement data. Instead, beyond errors in measuring the present conditions (as with initial conditions for MDP), the prediction is compromised by the finite size of the data set. That is, the said virtue of DDP can be exploited – when employing it in its pure form – only if enough and good quality data (with a high precision of observation and high signal-to-noise ratio) is available [7].

One might expect that the slogan that 'extremes in comparison with more moderate events are harder to predict' extends to DDP. In fact, just the opposite has been reported by Hallerberg and Kantz [8] for simple autoregressive processes at least, indifferently to whether the probability distribution is exponentially decaying or according to a power-law, and also for some observational data [9]: *stronger events are easier to predict*. This counter-intuitive statement is based on a measure of prediction skill that derives from the so-called *receiver operating characteristic* (ROC) *curve* [10,11] that takes into account the true positives – meaning that an event is correctly predicted to happen – as well as the false negatives. Concerning rare events, such a measure of prediction skill is regarded [12] more meaningful than other *proper* [13] so-called *skill scores* for probabilistic predictions like the Brier or Ignorance scores. This is so, because the ROC statistics has been viewed to not depend on the relative frequency of events (only that the accurate evaluation of the statistics requires a sufficient number of events). The latter characteristic is thought to allow for the comparison of the ROC-predictability of events between two situations where the events have different frequency [12].

Whether the above statement [8] can be *maintained* in case of more complex processes has been an open question so far—addressed but not settled with a consensus. Recently two studies [14,15] have been published concerning the predictability of extreme events in dynamical systems with seemingly contradictory results as to whether stronger events are more predictable. Franzke [14] applied the method set out in [10] to predict extreme threshold exceedances in a systematically derived stochastic dynamical system representing climate variability by the resolved (slow) variable(s) and weather variability by noise in place of the unresolved (fast) variable(s) [16]. He maintained the earlier statement [8] in this case, measuring the prediction skill by the ROC statistics, but on the basis of considering only two high threshold values. On the other hand, Sterk et al. [15] considered a number of dynamical systems of various complexity, and various physical observables. They evaluated finite-time maximal Lyapunov exponents (FTMLE) of trajectories that lead to extremes, and concluded that no generally applicable statements can be made, but the predictability of extremes depends on the system (and so the attractor geometry) and on the observable in question, as well as the prediction lead time. We emphasize that in their study the authors did not take model errors into account.

To summarize the essence of the above review, we can list three different views encountered in the literature regarding the predictability of extremes:

(1) Stronger extremes are better predictable.
(2) Stronger extremes are less predictable.
(3) Stronger extremes can be better or less predictable depending on various factors.

Without giving details, e.g. assumptions of these statements, they seem to be contradictory to each other. On this basis we set out the following objectives for the present paper:

(i) Keeping to the assumption of (1), we evaluate the predictability of *peak-over-threshold* events measured by a ROC-based quantity, using time series data of finite length produced by the Lorenz-84 model [17]. With an attention to (3), we evaluate (i.a) the dependence of predictability *itself* on various factors, and also what is more relevant to the question: (i.b) the magnitude-dependence of predictability – whether increasing/decreasing or nonmonotonic – depending on some of those same factors.
(ii) We argue for an analogy between a certain class of DDPs and MDP, and that the latter is usually understood as something that below we will refer to as an *on-demand* MDP, in which case any input data belongs to a single time instant. We believe this is an assumption of (2). This objective (ii) is to reconcile (1) and (2), suggesting that (2) can be true when model errors are present, even if the predictability is measured by the same ROC-based quantity as that assumed by (1).
(iii) For on-demand MDP and the analogous DDP where the time of input is arbitrary, we will be able to carry out an assessment of the lead time-dependence of the predictability of what we will call *threshold-exceedance-in-an-interval* events in a straightforward manner. This will turn out to have a bearing on the magnitude-dependence of predictability. This objective (iii) together with (i.b) are to revisit point (1), possibly extending that point from stochastic processes to dynamical systems.
(iv) However, to show that (3) does not necessarily contradict (1), we recall that (3) was stated on the basis of measuring predictability by FTMLEs. This also assumes an arbitrary input time on-demand MDP or analogous DDP. Accordingly, in the autonomous L84 we calculate the FTMLEs of trajectories that lead to extremes, and compare their average to the ROC-based measure of prediction skill—looking for any qualitative mismatch.

To motivate our top objective (i) and (iii) we remark that DDP is gaining increasing prominence nowadays given that data is relatively much more easily accessible than models. This is certainly the case with geophysical phenomena that we are primarily interested in, such as meteorology. Furthermore, performing predictions based on data can be far less costly than those based on simulating complex models, while the skill may not be much worse [18].

Next we recapitulate the methodology of the applied prediction scheme and the used ROC-based measure of prediction skill. Lorenz's 1984 model of global atmospheric circulation, simulated to produce time series data for the purpose of assessing the predictability of threshold exceedance events, is also briefly described. Subsequently, in Section 3, we present our results on the dependence of predictability on several factors, such as: the makeup of the so-called precursory structure – made use for a prediction – in terms of the observables involved, the prediction