

University of Groningen

## Extreme value laws and mean squared error growth in dynamical systems

Sterk, A.E.; Holland, M.P.

*Published in:*  
Dynamics and Statistics of the Climate System: An Interdisciplinary Journal

*DOI:*  
[10.1093/climsys/dzy007](https://doi.org/10.1093/climsys/dzy007)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2018

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Sterk, A. E., & Holland, M. P. (2018). Extreme value laws and mean squared error growth in dynamical systems. *Dynamics and Statistics of the Climate System: An Interdisciplinary Journal*, 3(1), 1-25.  
<https://doi.org/10.1093/climsys/dzy007>

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

Research Article

# Extreme value laws and mean squared error growth in dynamical systems

Alef E. Sterk<sup>a,\*</sup> and Mark P. Holland<sup>b</sup>

<sup>a</sup>Bernoulli Institute, University of Groningen, PO Box 407, 9700 AK Groningen, The Netherlands and

<sup>b</sup>College of Engineering, Mathematics and Physical Sciences, University of Exeter, Harrison Building, Streatham Campus, North Park Road, Exeter EX4 4QF, UK.

\*Correspondence Alef E. Sterk, Bernoulli Institute, University of Groningen, PO Box 407, 9700 AK Groningen, The Netherlands; E-mail: [a.e.sterk@rug.nl](mailto:a.e.sterk@rug.nl)

Received 29 January 2018; Accepted 1 October 2018

## Abstract

**Background.** Extreme value theory for chaotic deterministic dynamical systems is a rapidly expanding area of research. Given a system and a scalar observable defined on its state space, extreme value theory studies the asymptotic probability distributions of large values attained by the observable along evolutions of the system.

**Objective.** The aim of this paper is to study the relation between the statistics and predictability of extremes.

**Methods.** Predictability is measured by the mean squared error (MSE), which is estimated from the difference of pairs of forecasts conditional on one of the forecasts exceeding a threshold.

**Results.** Under the assumption that pairs of forecast variables satisfy a linear regression model, we show that the MSE can be decomposed into the sum of three terms: a threshold-independent constant, a mean term that always increases with threshold, and a variance term that can either increase, decrease, or stay constant with threshold. Using the generalised Pareto distribution to model excesses over a threshold, we show that the MSE always increases with threshold at sufficiently high threshold. However, when the forecasts have a negative tail index the MSE can be a decreasing function of threshold at lower thresholds.

**Conclusions.** Our method is illustrated by means of four examples: the tent map, the cusp map, and two low-order models for atmospheric regime transitions and the El Niño–Southern Oscillation phenomenon. These examples clearly show that predictability depends on the observable and the invariant measure of the system.

**Keywords:** Dynamical systems, extreme value laws, mean squared error, predictability.

## 1. Introduction

An important problem in meteorology and climatology is understanding the statistics and predictability of extreme events such as storms, floods or heat waves. The statistical theory of extremes in stochastic processes, both with and without serial dependence, is well developed (Galambos, 1978; Leadbetter *et al.*, 1983; Resnick, 1987; Castillo, 1988; Embrechts *et al.*, 1997; Coles, 2001; Beirlant *et al.*, 2004). Typically, the aim is to estimate the probability of the occurrence of future large values of a quantity, given a sample of past measurements of that quantity. To this end, the block maximum (BM) and peak-over-threshold (POT) methods are often used. The BM method amounts to dividing a time series in sufficiently long blocks and recording the largest observation of each block. The resulting block maxima are then fitted to the generalized extreme value distribution by methods such as maximum likelihood or L-moments. The POT method fits excesses over a selected threshold to a generalized Pareto distribution (GPD). Estimates of extreme value distributions can then be used to derive other quantities of interest such as return periods.

The extension of extreme value theory to the setting of chaotic, deterministic systems has become a rapidly expanding field in the last two decades (Collet, 2001; Freitas and Freitas, 2008; Freitas, 2009; Freitas *et al.*, 2010, 2011, 2012; Gupta, 2011; Holland *et al.*, 2012a, 2016; Chazottes and Collet, 2013). A comprehensive overview of the most important developments up to 2016 can be found in Lucarini *et al.* (2016). The main idea is to study time series obtained by evaluating a scalar observable along an evolution of the system. If the system is equipped with an invariant measure, then this time series can be interpreted as a stochastic process with serial dependence. Under suitable mixing conditions, the extreme value laws for dynamical systems are the same as for i.i.d. stochastic processes. Geophysical applications, in which dynamical systems arise as prognostic models and observables are physical quantities such as wind speed or temperature, form an important motivation for the development of the theory.

The predictability of extreme events is of great importance, and this topic has been studied by various authors. Hallerberg *et al.* (2007) studied the predictability of extreme *increments* in first-order autoregressive processes, wind speed recordings and long-range correlated autoregressive moving averages. In all their examples, extremes become better predictable with increasing event magnitude. The results in (Hallerberg and Kantz, 2008b) showed that for i.i.d. stochastic processes large increments are better predictable if the process is Gaussian, whereas large increments become less predictable if the underlying distribution has a power law tail. However, in the follow-up study (Hallerberg and Kantz, 2008a), which is concerned with *threshold crossings* instead of *increments*, it was found again that extremes are always better predictable. The work of Franzke (2012) and Franzke (2017) shows that also in the context of dynamic-stochastic models, extremes are better predictable than non-extremes. Bodai (2015) argues that in dynamical systems stronger predictability of extremes may be typical but not universal. Examples given by Sterk *et al.* (2012) and Sterk and Van Kekem (2017), however, show that the predictability of extreme values in deterministic dynamical systems can strongly depend on the observable, the attractor of the system and the prediction lead time.

The aim of this article is to explore the relation between statistics and predictability of extremes in dynamical systems by building on ideas presented in (Sterk *et al.*, 2016). Predictability is measured by the mean squared error (MSE), which is estimated from the difference of a pair of forecasts conditional on one of the forecasts exceeding a threshold. This approach has the advantage that we can consider finite-size errors in the initial condition, rather than infinitesimally small errors that are implicitly assumed when using predictability indicators such as finite-time Lyapunov exponents (Sterk *et al.*, 2012). Using a linear regression model, we show that the MSE can be decomposed into the sum of three terms: a threshold-independent constant, a mean term that always increases with threshold and a variance term that can either increase, decrease, or stay constant with threshold. Using the GPD to model excesses over a threshold, we show that the MSE always increases with threshold at sufficiently high threshold. However, the MSE can be a decreasing function of threshold at lower thresholds when the forecasts have finite upper bounds. The aim of this article is to illustrate the key ideas by means of engaging examples. A more rigorous mathematical approach including formal proofs will be presented elsewhere. Our examples clearly show that the predictability of extremes depends on the regularity properties of the invariant measure and the observable.

This article is structured as follows. In Section 2, we first summarize the theory of extreme value statistics in deterministic systems. Next, we explain how the MSE can be decomposed as a sum of three terms, and we use the GPD to study the behaviour of these terms on asymptotically high thresholds. In Section 3.1, we apply our methodology to the tent map, which is a basic model in dynamical systems exhibiting sensitive dependence on initial conditions and which has an explicit invariant measure; this example illustrates all possible behaviours of the MSE decomposition terms. In Section 3.2, we study the cusp map, which exhibits intermittent dynamics and also has an explicit invariant measure; this example shows how intermittency can lead to clustered extremes for which the methodology of Section 2.2 may no longer be applicable. As geophysical examples we consider wind speeds in a low-order barotropic vorticity

equation in Section 3.3 and sea surface temperatures in a low-order model for El Niño-Southern Oscillation (ENSO) in Section 3.4. Both models have a Shilnikov-like strange attractor that leads to irregular alternations between different physical regimes and slow decay of correlations. The article is concluded in Section 4 with a discussion.

## 2. Theoretical background

In this article, we consider a dynamical system, where  $\mathcal{X}$  is a  $d$ -dimensional Riemannian manifold, a measurable transformation, and is an  $f$ -invariant probability measure. In applications, the map  $f$  can also be a time-shift map associated with the flow of a differential equation. Assume that there is a compact invariant set  $\mathcal{A} \subset \mathcal{X}$  that supports the measure. Given an observable  $\phi: \mathcal{X} \rightarrow \mathbb{R}$ , we consider the stationary process defined by

$$X_\tau = \phi \circ f^\tau, \quad \tau \in \mathbb{N} \cup \{0\}. \quad (1)$$

The aim of this article is to explore the interplay between the statistics and the predictability of large values of the random variables  $X_\tau$ .

### 2.1. Statistics of extremes

Classical extreme value theory concerns the probability distribution of unlikely large events (Galambos, 1978; Leadbetter *et al.*, 1983; Resnick, 1987; Castillo, 1988; Embrechts *et al.*, 1997; Coles, 2001; Beirlant *et al.*, 2004). Consider a time series  $(X_i)_{i=0}^\infty$  of independent identically distributed random variables and define the partial maximum over the first  $n$  occurrences:

$$M_n = \max(X_0, \dots, X_{n-1}).$$

This variable has a degenerate limit as  $n \rightarrow \infty$ , and therefore, it is necessary to consider a rescaling. Assume that there exist sequences  $a_n > 0$  and  $b_n \in \mathbb{R}$  such that the rescaled variable  $a_n(M_n - b_n)$  converges to a non-degenerate distribution:

$$\lim_{n \rightarrow \infty} P(a_n(M_n - b_n) \leq z) = G(z). \quad (2)$$

Then, the extremal types theorem asserts that  $G$  is the generalized extreme value (GEV) distribution given by

$$G(z) = \exp\left(-\left[1 + \xi\left(\frac{z - \lambda}{\sigma}\right)\right]^{-1/\xi}\right), \quad (3)$$

which is supported on the set  $\{z \in \mathbb{R} : 1 + \xi(z - \lambda)/\sigma > 0\}$  (Leadbetter *et al.*, 1983; Coles, 2001). The parameters  $\lambda \in (-\infty, \infty)$  and  $\sigma \in (0, \infty)$  are called *location* and *scale*. The *tail index*  $\xi \in (-\infty, \infty)$  is the most important parameter because it determines the tail behaviour of the distribution and therefore the likelihood of extreme values. In the cases  $\xi = 0$ ,  $\xi > 0$  or  $\xi < 0$ , the distribution (3) is also referred to as Gumbel (type I), Fréchet (type II) or Weibull (type III), respectively. The parameters  $(\lambda, \sigma, \xi)$  can be estimated from a time series using block-maximum or peak-over-threshold methods (Embrechts *et al.*, 1997; Coles, 2001).

Recent work has extended extreme value theory to the setting of chaotic deterministic dynamical systems (see Lucarini *et al.*, 2016 for a comprehensive overview). Under suitable mixing conditions, the process  $M_n$  obtained from (1) satisfies (2). Most theoretical work focuses on observables of the form

$$\phi(x) = g(\text{dist}(x, \zeta)), \quad x \in \mathcal{X}, \quad (4)$$

where  $\zeta \in \mathcal{X}$  is a density point of the invariant measure  $\nu$  and  $g: [0, +\infty) \rightarrow \mathbb{R} \cup \{+\infty\}$  is a strictly decreasing bijection in a neighbourhood of 0 and has a global maximum at 0. However, typical observables used in applications are not of this form. For example, in physical applications one might be interested in energy, wind speed or vorticity. Such observables are typically of the form

$$\phi_E(x) = x^T E x, \quad \phi_W(x) = \|Wx\|, \quad \phi_V(x) = Vx, \quad (5)$$

respectively, where  $x$  is a point in the phase space  $\mathcal{X} = \mathbb{R}^d$ ,  $\|\cdot\|$  denotes the Euclidean norm and  $E \in \mathbb{R}^{d \times d}$ ,  $W \in \mathbb{R}^{2 \times d}$ ,  $V \in \mathbb{R}^{1 \times d}$  are matrices.

The parameters of the GEV distribution (3) strongly depend on the form of the observable  $\phi$ . For example, observables of the form (4) with

$$g_1(y) = -\log(y), \quad g_2(y) = y^{-1/\alpha}, \quad g_3(y) = -y^{1/\alpha}, \quad y \in \mathbb{R}, \quad (6)$$

where  $\alpha > 0$ , give a tail index of  $\xi = 0$ ,  $\xi = 1/\alpha d(\zeta)$  and  $\xi = -1/\alpha d(\zeta)$ , respectively, where

$$d(\zeta) := \lim_{r \rightarrow 0} \frac{\log v(\{x \in \mathcal{X} : \text{dist}(x, \zeta) < r\})}{\log r} \quad (7)$$

is the *local dimension* of the measure  $\nu$  at  $\zeta$  (see Lucarini *et al.*, 2016, Section 4.2.1). Note that  $\xi \rightarrow 0$  as  $d(\zeta) \rightarrow \infty$ , which means that for systems with high-dimensional attractors, the distribution of extremes is approximately a Gumbel law. For systems with measures that are not absolutely continuous, equations (2) and (3) do not hold for *linear* scaling sequences, as outlined in Holland *et al.* (2016). However, numerical fitting schemes might still give good estimates on the tail behaviour (see also Faranda *et al.*, 2012).

For observables of the form (5), however, determining the GEV distribution becomes a much more delicate problem. In this case, the GEV distribution is no longer determined by the local dimension of the measure  $\nu$ , as stated in equation (7). The asymptotic behaviour of  $\nu\{\phi(x) > \phi_M - u\}$  as  $u \rightarrow 0$ , where  $\phi_M$  is the maximum of  $\phi$ , still plays a crucial role. In fact, this critically depends on the geometry of the attractor  $\mathcal{A} \subset \mathcal{X}$  and the level sets of  $\phi$ . Problems arise, for instance, when the level set is a strip or a fractal set (see Holland *et al.*, 2012 for an extensive discussion). Note that the observables (5) are continuous functions, and hence, they are bounded when the system has an attractor which is compact. This implies that the time series (1) has an upper bound, and in this case, we expect a priori that  $\xi < 0$  for observables of the form (5).

## 2.2. Predictability of extremes

In this article, we study predictability in the context of a perfect model scenario: instead of comparing forecasts with observations we compare predictions with each other. Given two initial conditions  $x_0, y_0 \in \mathcal{A}$ , where  $\text{dist}(x_0, y_0)$  is sufficiently small, we can make the twin predictions

$$X_\tau = \phi(f^\tau(x_0)) \quad \text{and} \quad Y_\tau = \phi(f^\tau(y_0)). \quad (8)$$

We will interpret  $X_\tau$  as an ‘observation’ and  $Y_\tau$  as a ‘forecast’. We can now measure the predictability of extremes by the MSE between the twin predictions given that the observation exceeds a threshold:

$$\text{MSE}(\tau, u) = E([X_\tau - Y_\tau]^2 | X_\tau > u). \quad (9)$$

The aim of this article is to understand how the MSE depends on the threshold  $u$ . The following result provides a decomposition of the MSE for fixed  $\tau$  as a function of  $u$  under the assumption that the ‘forecast’ and ‘observation’ are related through a linear model. In Sterk *et al.* (2016), a similar methodology was applied to a data set of wind speed forecasts produced by an operational weather forecasting model. The aim of this article is to apply the methodology to simple dynamical systems examples to investigate the relation between the predictability of extremes and certain recurrence properties of the system such as intermittency.

**Lemma 1.** *Assume that the ‘observation’  $X_\tau$  and ‘forecast’  $Y_\tau$  are identically distributed at  $\tau = 0$  and that for  $\tau \geq 0$  they are related by the linear regression model*

$$Y_\tau = \mu + \rho_\tau(X_\tau - \mu) + \varepsilon, \quad (10)$$

where the error term  $\varepsilon$  is independent of  $X_\tau$  and  $E(\varepsilon) = 0$ . Then  $X_\tau$  and  $Y_\tau$  are identically distributed for all  $\tau \geq 0$ , and the coefficients of the linear model are given by

$$\mu = E(X_\tau) = E(Y_\tau) \quad \text{and} \quad \rho_\tau = \frac{\text{Cov}(X_\tau, Y_\tau)}{\text{Var}(X_\tau)}. \quad (11)$$

Moreover, we have the following decomposition:

$$E([Y_\tau - X_\tau]^2 | X_\tau > u) = (1 - \rho_\tau^2) \text{Var}(X_\tau) + (\rho_\tau - 1)^2 \text{Var}(X_\tau | X_\tau > u) + (\rho_\tau - 1)^2 E(X_\tau - \mu | X_\tau > u)^2. \tag{12}$$

*Proof.* If  $X_\tau$  and  $Y_\tau$  are identically distributed at  $\tau = 0$ , then they are identically distributed for  $\tau \geq 0$  by invariance of the measure  $\nu$ . Equation (11) follows from a straightforward computation. Note that  $\mu$  is independent of  $\tau$  since  $\nu$  is invariant under the map  $f$ .

Combining the standard formula for conditional variance (Ross, 1997) with the linear model of equation (10) gives

$$\begin{aligned} E([Y_\tau - X_\tau]^2 | X_\tau > u) &= \text{Var}(Y_\tau - X_\tau | X_\tau > u) + E(Y_\tau - X_\tau | X_\tau > u)^2 \\ &= \text{Var}((\rho_\tau - 1)(X_\tau - \mu) + \varepsilon | X_\tau > u) \\ &\quad + E((\rho_\tau - 1)(X_\tau - \mu) + \varepsilon | X_\tau > u)^2 \\ &= \text{Var}((\rho_\tau - 1)(X_\tau - \mu) + \varepsilon | X_\tau > u) \\ &\quad + [(\rho_\tau - 1)E(X_\tau - \mu | X_\tau > u) + E(\varepsilon | X_\tau > u)]^2. \end{aligned}$$

The assumption that the error  $\varepsilon$  is independent of  $X_\tau$  gives

$$\begin{aligned} \text{Var}((\rho_\tau - 1)(X_\tau - \mu) + \varepsilon | X_\tau > u) &= (\rho_\tau - 1)^2 \text{Var}(X_\tau - \mu | X_\tau > u) + \text{Var}(\varepsilon | X_\tau > u) \\ &= (\rho_\tau - 1)^2 \text{Var}(X_\tau | X_\tau > u) + \text{Var}(\varepsilon). \end{aligned}$$

Note that equation (10) implies that  $\text{Var}(Y_\tau) = \rho_\tau^2 \text{Var}(X_\tau) + \text{Var}(\varepsilon)$ . Since  $\text{Var}(Y_\tau) = \text{Var}(X_\tau)$ , we have

$$\text{Var}(\varepsilon) = (1 - \rho_\tau^2) \text{Var}(Y_\tau) = (1 - \rho_\tau^2) \text{Var}(X_\tau).$$

Finally, noting that  $E(\varepsilon | X_\tau > u) = E(\varepsilon) = 0$  completes the proof. □

Recall the GPD, which is given by

$$F(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp(-x/\beta) & \text{if } \xi = 0, \end{cases} \tag{13}$$

where  $\beta > 0$  is the scale parameter and  $\xi \in \mathbb{R}$  is the tail index (which is also referred to as the shape parameter). If  $X$  is any random variable, then under widely applicable conditions (Balkema and de Haan, 1974; Pickands, 1975), the distribution of the excess  $X - u$  conditional on  $X > u$  for sufficiently large  $u$  is well approximated by the GPD in which the scale parameter becomes a linear function of threshold, namely

$$\beta(u) = \sigma + \xi(u - \lambda), \tag{14}$$

where  $(\lambda, \sigma, \xi)$  are the parameters of the GEV distribution (3) (see Coles, 2001). The next result will be useful to analyse the MSE at high thresholds  $u$ .

**Lemma 2.** Assume that the random variable  $X$  has a GPD with scale parameter  $\beta > 0$  and tail index  $\xi \in \mathbb{R}$ . If  $\xi < 1/2$ , then the first and second moment of  $X$  are finite and

$$E(X - \mu | X > u)^2 = \left( \frac{\beta + (1 - \xi)(u - \mu)}{1 - \xi} \right)^2, \tag{15}$$

and

$$\text{Var}(X | X > u) = \frac{(\beta + \xi u)^2}{(1 - \xi)^2 (1 - 2\xi)}. \tag{16}$$

*Proof.* Let  $x_F = \sup\{x \geq 0 : F(x) < 1\}$  denote the right endpoint of the distribution function  $F$  in (13). If  $\xi < 0$ , then  $x_F = -\beta/\xi$ ; otherwise,  $x_F = \infty$ . For  $k \in \mathbb{N}$ , we have

$$E((X-u)^k | X > u) = \frac{1}{1-F(u)} \int_u^{x_F} (x-u)^k F'(x) dx.$$

The first two conditional moments are given by

$$E(X-u | X > u) = \frac{\beta + \xi u}{1 - \xi} \quad \text{and} \quad E((X-u)^2 | X > u) = \frac{2(\beta + \xi u)^2}{(1 - \xi)(1 - 2\xi)}.$$

This gives

$$E(X-\mu | X > u)^2 = (E(X-u | X > u) + u - \mu)^2 = \left( \frac{\beta + (1 - \xi)(u - \mu)}{1 - \xi} \right)^2$$

and

$$\text{Var}(X-u | X > u) = E((X-u)^2 | X > u) - E(X-u | X > u)^2 = \frac{(\beta + \xi u)^2}{(1 - \xi)^2(1 - 2\xi)}.$$

Noting that  $\text{Var}(X | X > u) = \text{Var}(X-u | X > u)$  completes the proof.  $\square$

According to the discussion above, for sufficiently large  $u$  the distribution of  $X_\tau - u$  conditional on  $X_\tau > u$  can be approximated by a GPD for which the scale parameter is given by (14). Combining this with Lemma 1 gives

$$\text{MSE}(\tau, u) = (1 - \rho_\tau^2) \text{Var}(X_\tau) + (1 - \rho_\tau^2) \{E(u) + V(u)\},$$

where

$$E(u) = \left( \frac{\beta(u) + (1 - \xi)(u - \mu)}{1 - \xi} \right)^2 \quad \text{and} \quad V(u) = \frac{(\beta(u) + \xi u)^2}{(1 - \xi)^2(1 - 2\xi)}.$$

The mean component  $E(u)$  increases monotonically for  $u > \mu + \xi(\lambda - \mu) - \sigma$ . The variance component  $V(u)$  either increases, stays constant or decreases with  $u$  depending upon whether  $\xi > 0$ ,  $\xi = 0$  or  $\xi < 0$ , respectively.

For  $\xi \geq 0$ , the sum  $S(u) = E(u) + V(u)$  is a strictly monotonically increasing function of threshold. For  $\xi < 0$ , a decreasing  $V(u)$  may compensate for an increasing  $E(u)$  and create a minimum  $S(u)$  at

$$u_{\min} = \frac{(1 - 2\xi)(1 - \xi)\mu + \lambda\xi - \sigma}{1 - 2\xi + 4\xi^2}. \quad (17)$$

In summary, the MSE always increases with threshold at sufficiently high threshold, although it can be a decreasing function of threshold at lower thresholds but only if the ‘observations’  $X_\tau$  have a negative tail index. Computing the minimum  $u_{\min}$  requires an accurate estimate of the parameters  $(\mu, \lambda, \sigma, \xi)$ . However, in practical applications, it may be difficult to obtain a precise estimate for the tail index  $\xi$  (Holland *et al.*, 2012).

In concrete applications, the expectation in (9) is replaced by its empirical counterpart that is computed from a sufficiently large sample. This amounts to drawing samples from the invariant measure for the pairs of initial conditions in (8). Given an arbitrary point in  $\mathcal{X}$ , we first iterate the map  $f$  sufficiently many times to get an initial point  $x_0$  on the attractor. Next, we construct a perturbation  $y_0$  of  $x_0$  for the twin prediction in (8). Note that by simply adding a random perturbation to  $x_0$  we would get a point that is almost surely *not* on the attractor. In this case, the points  $x_0$  and  $y_0$  would not have the same distribution. Instead, we construct perturbations by finding a so-called *analogue* of the point  $x_0$ . For a specified distance  $\delta > 0$ , we find the first  $j \geq 100$  such that  $y_0 = f^j(x_0)$  satisfies  $\text{dist}(x_0, y_0) < \delta$ . By requiring that  $j \geq 100$  we ensure that the short-time evolutions starting at  $x_0$  and  $y_0$  do not overlap. We then replace the point  $x_0$  by  $f(x_0)$  and repeat this procedure until a chosen sample size  $N$  is reached. In the experiments below, we have used the parameters  $N = 10^5$ ,  $\delta = 0.1$  (tent map),  $\delta = 0.05$  (cusp map),  $\delta = 0.025$  (low-order barotropic model) and  $\delta = 0.1$  (low-order ENSO model).

### 3. Examples

We first present two examples of one-dimensional dynamical systems. These examples have the advantage that the invariant measure is known. In particular, the extreme value law is also known when we restrict to observables of the form (4). This allows us to check to what extent the conclusions of Lemma (2) hold. Next, we present two geophysical models; for these examples, it is unknown whether the invariant measure has a smooth density or whether the extreme value law is of the form (3).

#### 3.1. The tent map

Our first example is a uniformly expanding map. We consider the tent map given by  $f(x) = \pi - 2|x|$  for  $x \in \mathcal{X} = [-\pi, \pi]$ . The invariant measure  $\nu$  is the Lebesgue measure on  $\mathcal{X}$  rescaled by  $1/2\pi$ . We consider the observables in equation (4) in which the distance between two points is given by  $\text{dist}(x, y) = |x - y|$ . It is easy to verify that the local dimension in equation (7) is given by  $d(\zeta) = 1$  for all  $\zeta \in \mathcal{X}$ . In particular, this implies that the tail index for the process (1) with observables of the form (4) is given by  $\zeta = 0$  when  $g = g_1$ ,  $\zeta = 1/\alpha$  when  $g = g_2$  and  $\zeta = -1/\alpha$  when  $g = g_3$ . In the following, we will take  $\zeta = 1$ .

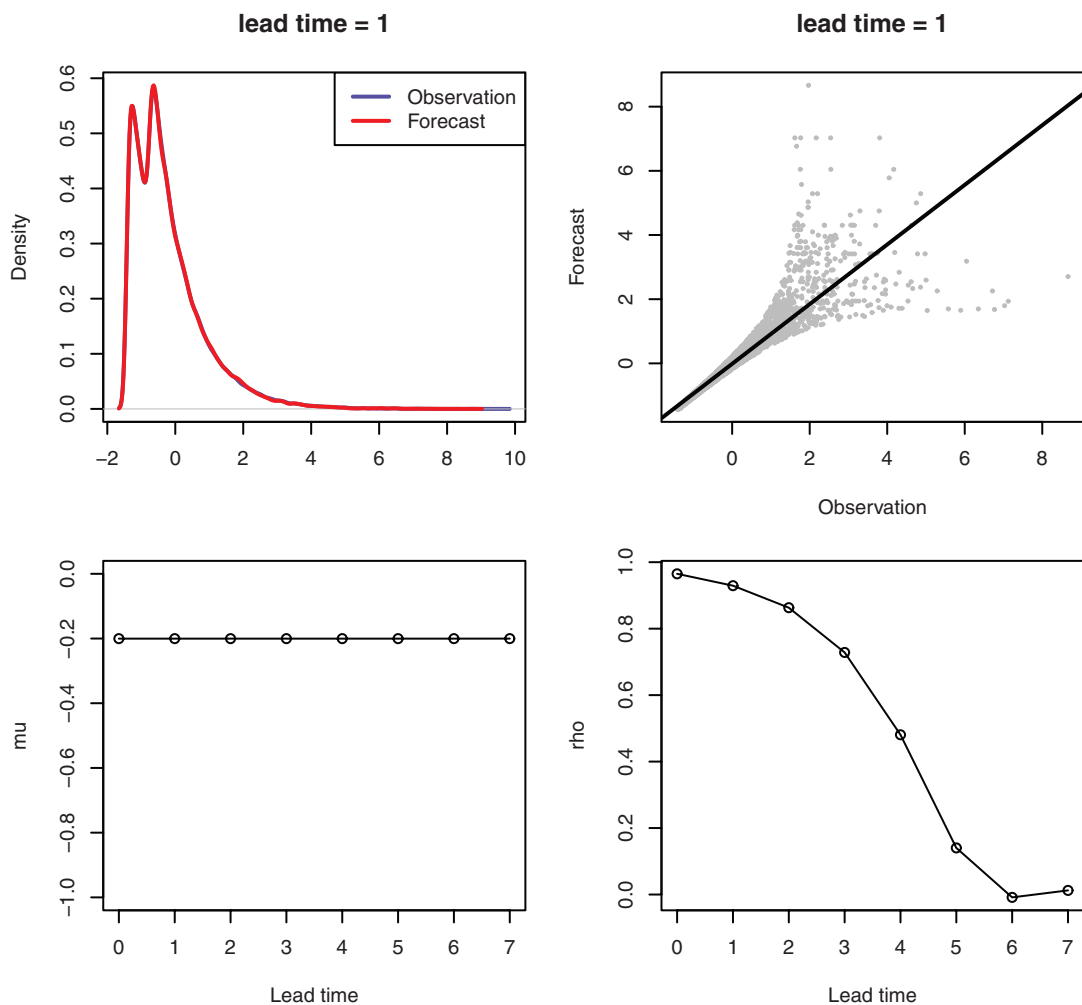


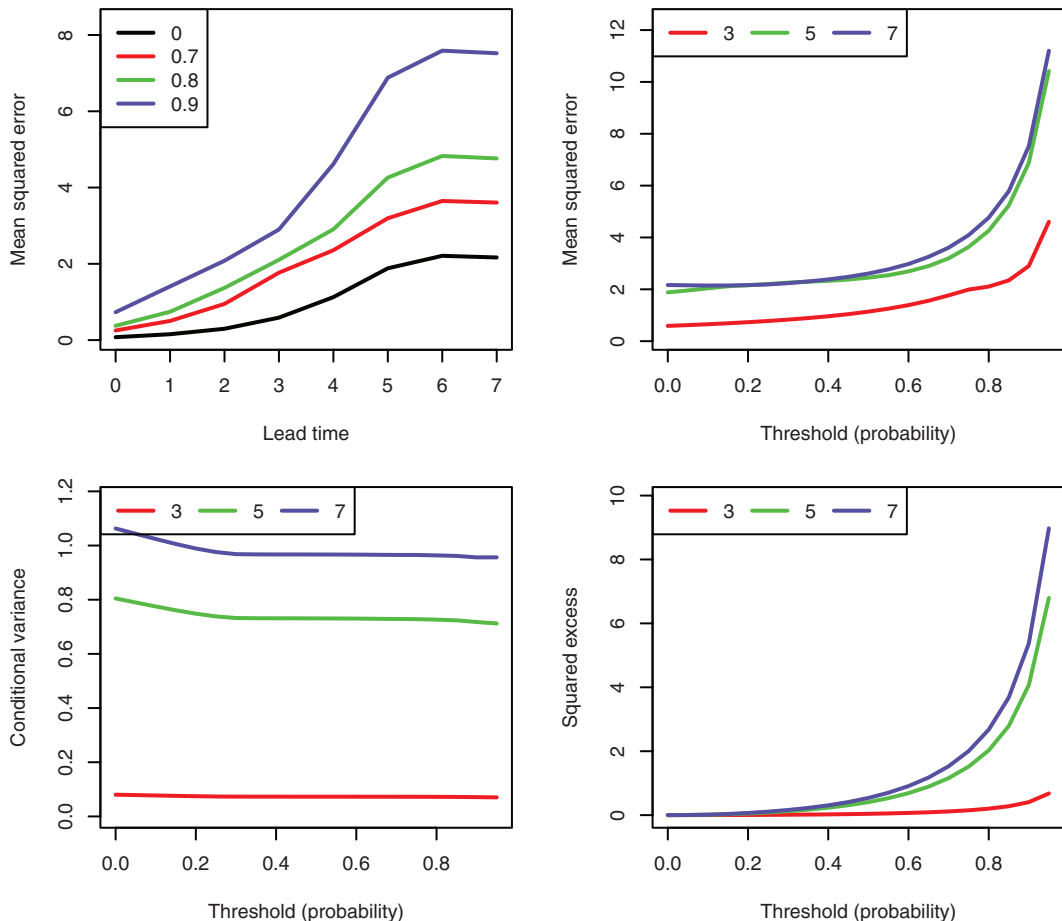
Figure 1. Statistics of tent map and the observable given by equation (4) with  $g = g_1$  and  $\zeta = 1$ .



We first discuss the case  $g = g_1$ . Figure 1 shows the invariant density and a scatter plot of  $X_\tau$  versus  $Y_\tau$  for the lead time  $\tau = 1$ . Clearly, the forecast and observation have the same distribution. Points in the scatter plot for which  $X_\tau > 2$  or  $Y_\tau > 2$  deviate from the linear model, but these points form less than 5% of the total number of points in the scatter plot. In addition, for  $\tau = 1$ , the correlation coefficient is  $\rho_\tau \approx 0.95$ . Figure 1 also shows that the coefficient  $\mu$  is a constant function of the prediction lead time  $\tau$ , which can be explained by the fact that the measure is invariant under the map  $f$ . The correlation coefficient  $\rho_\tau$  decays very fast with lead time and we have  $\rho_\tau \approx 0$  for  $\tau \geq 6$ . This can be explained by the fact that the doubling map has exponential decay of correlations for a large class of observables (Boyarsky and Góra, 1997).

Figure 2 shows that the MSE increases as a function of the prediction lead time  $\tau$  and saturates for  $\tau \geq 5$ . For fixed lead times, the MSE monotonically increases as a function of threshold. Note that the MSE is plotted against the exceedance probability rather than the actual threshold. The decomposition terms of the MSE in equation (12) are also shown in Figure 2. Clearly, the conditional variance is constant at high thresholds whereas the squared excess term increases with threshold. This behaviour is explained by the asymptotic expressions derived in Lemma 2 since the tail index  $\xi = 0$ .

For the observable (4) with  $g = g_2$  and  $\alpha = 10$ , the coefficients  $\mu$  and  $\rho$  of the linear model behave in qualitatively the same way, and hence, their graphs are omitted. The MSE and its decomposition terms are shown in Figure 3. In this case,  $\xi = 1/10 > 0$ , and hence, Lemma 2 explains why the conditional variance term is increasing. For the observable (4)



**Figure 2.** Analysis of the mean squared error in equation (9) as functions of lead time and threshold for the tent map and the observable given by equation (4) with  $g = g_1$  and  $\zeta = 1$ . The conditional variance and square excess refer to the decomposition terms defined in equation (12).

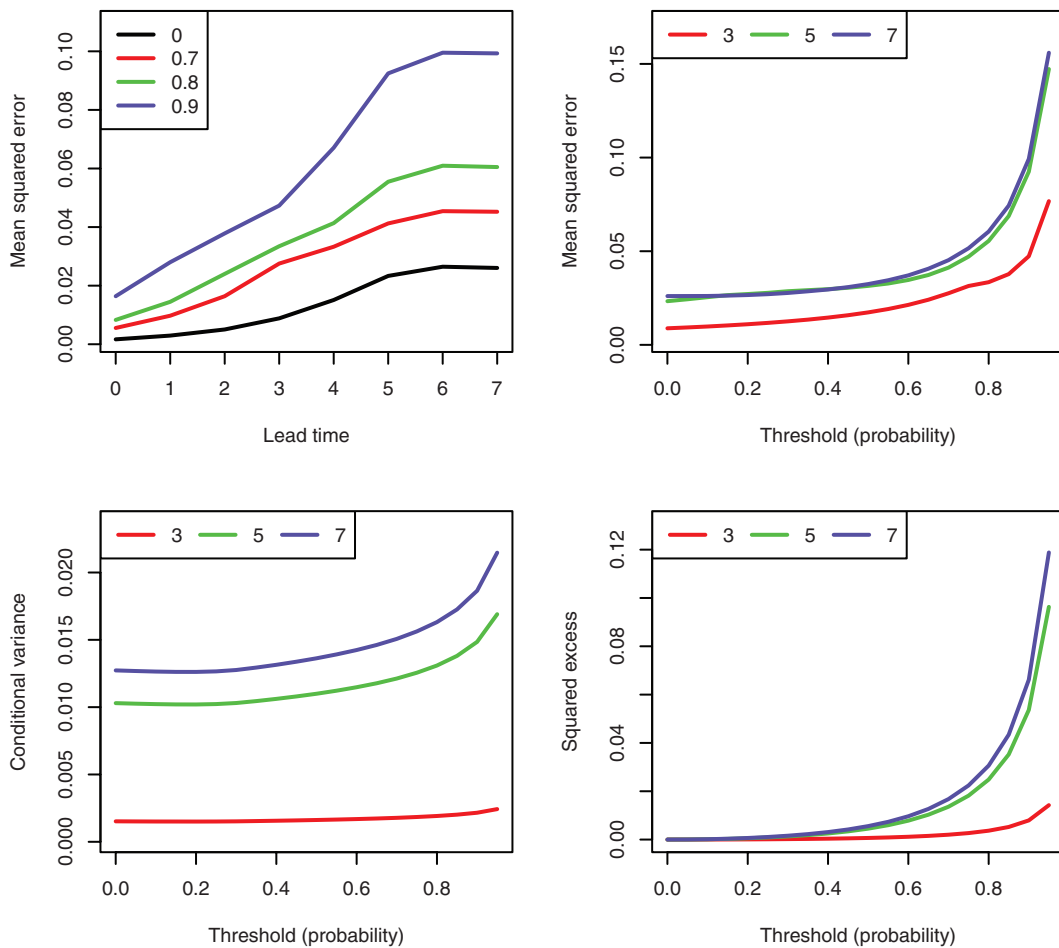


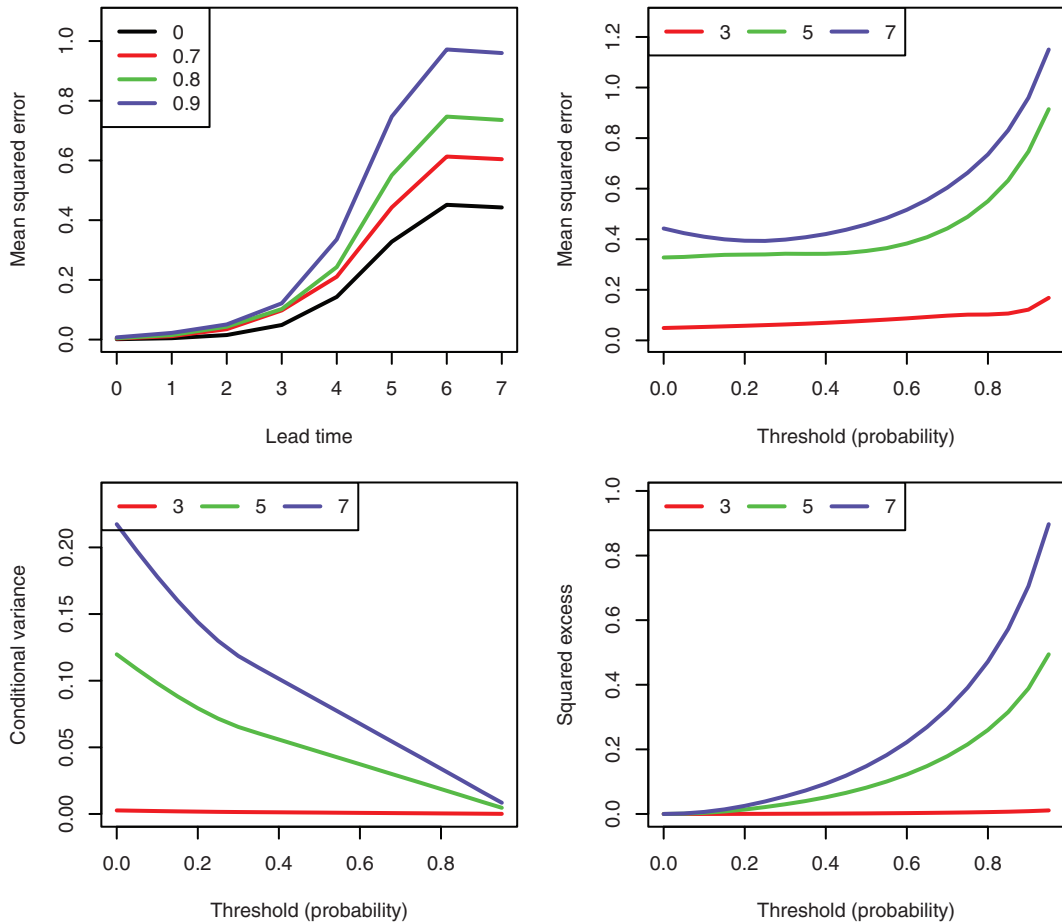
Figure 3. As Figure 2, but for the observable in equation (4) with  $g = g_2$ ,  $\alpha = 10$  and  $\zeta = 1$ .

with  $g = g_3$  and  $\alpha = 2$ , the MSE and its decomposition terms are shown in Figure 4. In this case, we have  $\xi = -1/2 < 0$ , and the conditional variance term is decreasing. Note that a minimum in the MSE as a function of threshold is not clearly visible. Fitting  $X_0 - u$  conditional on  $X_0 > u$  with  $u = -0.8$  (which corresponds to the 80th percentile) to a GPD with the peaks over threshold method (Coles, 2001) gives the estimates  $(\hat{\lambda}, \hat{\sigma}, \hat{\xi}) = (0.031, 0.012, -0.452)$ . Note that the estimated tail index is close to the true value  $\xi = -1/2$ . Together with the estimate  $\hat{\mu} = 1.21$ , formula (17) gives  $\hat{u}_{min} \approx -1.239$ . The estimated value of  $u_{min}$  lies far below thresholds for which the GPD is applicable, and this may explain why no clear minimum is observed in Figure 4.

The tent map clearly illustrates that all types of behaviour of the MSE as discussed in Section 2.2 can occur, depending on the choice of the observable. However, the tent map is a rather simple example. In more complicated examples, the local dimension varies across the attractor of the dynamical system. In this case, it is expected that the results also depend on the point  $\zeta \in \mathcal{X}$  where the observable attains its maximum. In the next section, we discuss an example for which this is indeed the case.

We now compare the results described above with conclusions that are obtained from finite-time Lyapunov exponents (FTLEs). The FTLE is the finite-time counterpart of the traditional infinite-time Lyapunov exponent, which measures the average exponential growth rate of infinitesimally small errors in the initial condition. For an interval map  $f: I \rightarrow I$ , the time- $\tau$  FTLE is defined as

$$\lambda_\tau(x_0) = \frac{1}{\tau} \sum_{k=0}^{\tau-1} \log|f'(x_k)| \quad \text{where} \quad x_{k+1} = f(x_k). \tag{18}$$



**Figure 4.** As [Figure 2](#), but for the observable in equation (4) with  $g = g_3$ ,  $\alpha = 2$  and  $\zeta = 1$ .

Typically,  $\lambda_\tau$  strongly depends on both the initial condition  $x_0$  and the prediction lead time  $\tau$ . However, for the tent map we have  $f'(x_0) = 2$  for almost every  $x_0 \in \mathcal{X}$  and this implies that  $\lambda_\tau(x_0) = \log(2)$  is independent of both the initial condition and the prediction lead time. Hence, measuring predictability with FTLEs implies that for the tent map all events are equally unpredictable and that predictability does not depend on lead time.

An alternative finite-time predictability indicator can be defined by

$$\sigma_\tau(x_0) = \prod_{k=0}^{\tau-1} |f'(x_k)|,$$

which has been used in ([Harle et al., 2006](#); [Sterk and Van Kekem, 2017](#)). Clearly, this indicator is related to the FTLE by the relation  $\lambda_\tau(x_0) = \log(\sigma_\tau(x_0))/\tau$ . Indeed, for a fixed value of  $\tau$  one can interpret  $\sigma_\tau$  and  $\lambda_\tau$  as the same predictability indicators expressed in different units. For the tent map we have  $\sigma_\tau(x_0) = 2^\tau$ . Unlike  $\lambda_\tau$ , the indicator  $\sigma_\tau$  does show that predictability decreases with increasing lead time  $\tau$  as one would expect. However,  $\sigma_\tau$  suggests that prediction errors keep growing without bound, whereas in practice prediction errors often saturate for  $\tau$  sufficiently large (see, e.g., [Figs 2–4](#)). Note that for the tent map,  $\sigma_\tau$  is still independent of the initial condition and therefore not suitable as a predictability indicator for specific events such as extreme events.

### 3.2. The cusp map

Our next example is a non-uniformly expanding map. We consider the cusp map  $f(x) = 1 - 2\sqrt{|x|}$  on  $\mathcal{X} = [-1, 1]$ . This map is a qualitative model of a Poincaré return map of the Lorenz–63 system ([Lorenz, 1963](#)). The map has an

invariant measure that is absolutely continuous with respect to the Lebesgue measure with density  $dv/dx = (1-x)/2$  (Hemmer, 1984). The point  $x = -1$  is a neutral fixed point, and time series of the cusp map show intermittent behaviour near  $x = -1$ . This is also reflected by the invariant measure: at  $x = -1$ , the density  $dv/dx$  is maximal, so that Birkhoff's ergodic theorem implies that iterates of the map spend longer times near  $x = -1$ .

We consider the observable (4) with  $g = g_3$  and  $\alpha = 2$  and two choices for  $\zeta \in \mathcal{X}$ . Due to the intermittent dynamics of the map  $f$ , the correlation coefficient of the linear model decays much slower than in the case of the tent map (see Fig. 5 for the case  $\zeta = 1$ ). In fact, the cusp map only has polynomial decay of correlations (Bhansali and Holland, 2007). A straightforward computation shows that the local dimension of the measure  $\nu$  is given by  $d(\zeta) = 1$  for all  $\zeta \in [-1, 1)$  and  $d(\zeta) = 2$  for  $\zeta = 1$ . In particular, we expect that the tail index is given by

$$\xi = \begin{cases} -1/2 & \text{for } \zeta \in [-1, 1), \\ -1/4 & \text{for } \zeta = 1. \end{cases}$$

The behaviour of the MSE and its decomposition terms strongly depend on the choice of  $\zeta$ . For  $\zeta = 1$ , the MSE increases with threshold as the calculations in Section 2.2 suggest (see Fig. 6). However, for  $\zeta = -1$ , the MSE decreases very fast with high thresholds (see Fig. 7). This cannot be explained by the results in Section 2.2. Note that for  $\zeta = -1$ , the observable is maximized at a fixed point of  $f$ . Figure 8 shows time series of the observable (4) with  $g = g_3$  and  $\alpha = 2$  for  $\zeta = -1$  and  $\zeta = -0.6$ . In the former case, the extremes form clusters: once the observable exceeds a high threshold, the next values in the time series also exceed that threshold.

The statistics of clustered extremes can still be described by equation (3) when the location and scale parameters  $(\lambda, \sigma)$  are replaced with the modified parameters (see Coles, 2001)

$$(\lambda^*, \sigma^*) = \begin{cases} (\lambda - \frac{\sigma}{\xi}(1 - \theta^{-\xi}), \theta^{\xi}\sigma) & \text{if } \xi \neq 0, \\ (\mu + \sigma \log \theta, \sigma) & \text{if } \xi = 0. \end{cases} \tag{19}$$

The new parameter  $\theta \in (0, 1]$  is referred to as the *extremal index* that can be interpreted as the reciprocal of the limiting mean cluster size, where the limiting is in the sense of clusters of exceedances of increasingly high threshold. Clustering in time series of random variables is discussed in Embrechts et al. (1997 and Coles (2001). A comprehensive discussion on recurrence properties of dynamical systems and clustering of extremes can be found in Lucarini et al. (2016).

If  $f(\zeta) = \zeta$  and  $1 < |f'(\zeta)| < \infty$ , then  $\zeta$  is an expanding fixed point of  $f$  and under suitable conditions the extremal index is given by

$$\theta = 1 - \frac{1}{|f'(\zeta)|}, \tag{20}$$

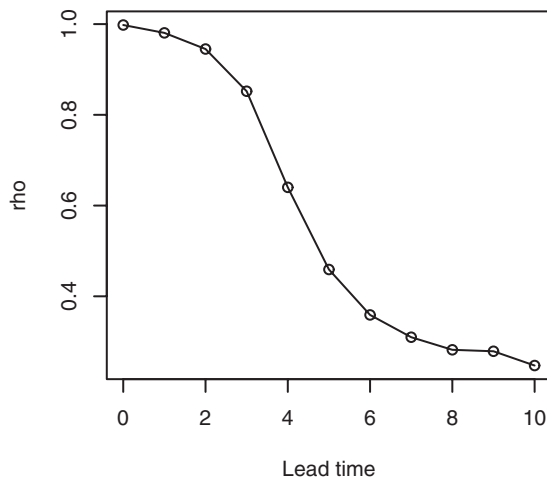
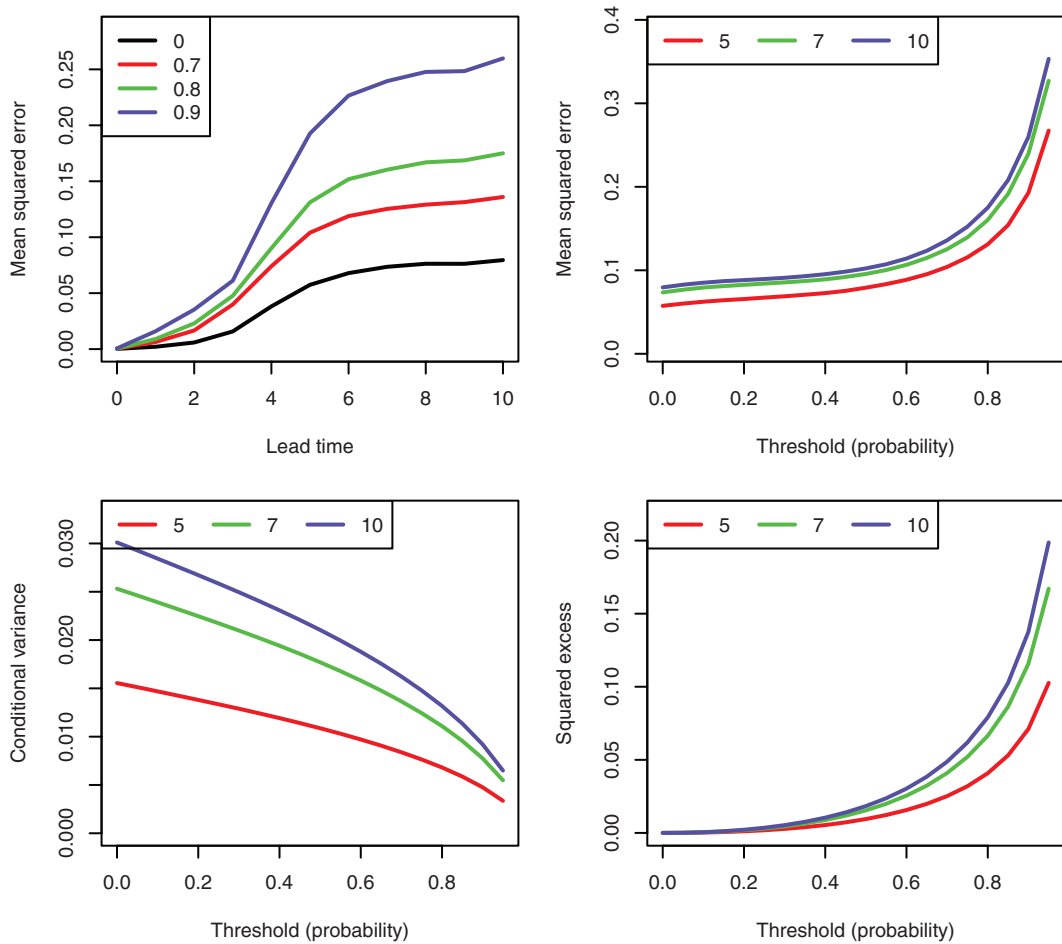


Figure 5. The correlation coefficient  $\rho$  of the linear model (10) for the cusp map with observable given by equation (4) with  $\alpha = 2$  and  $\zeta = 1$ . Note that the decay of correlations is much slower than in the case of the tent map.



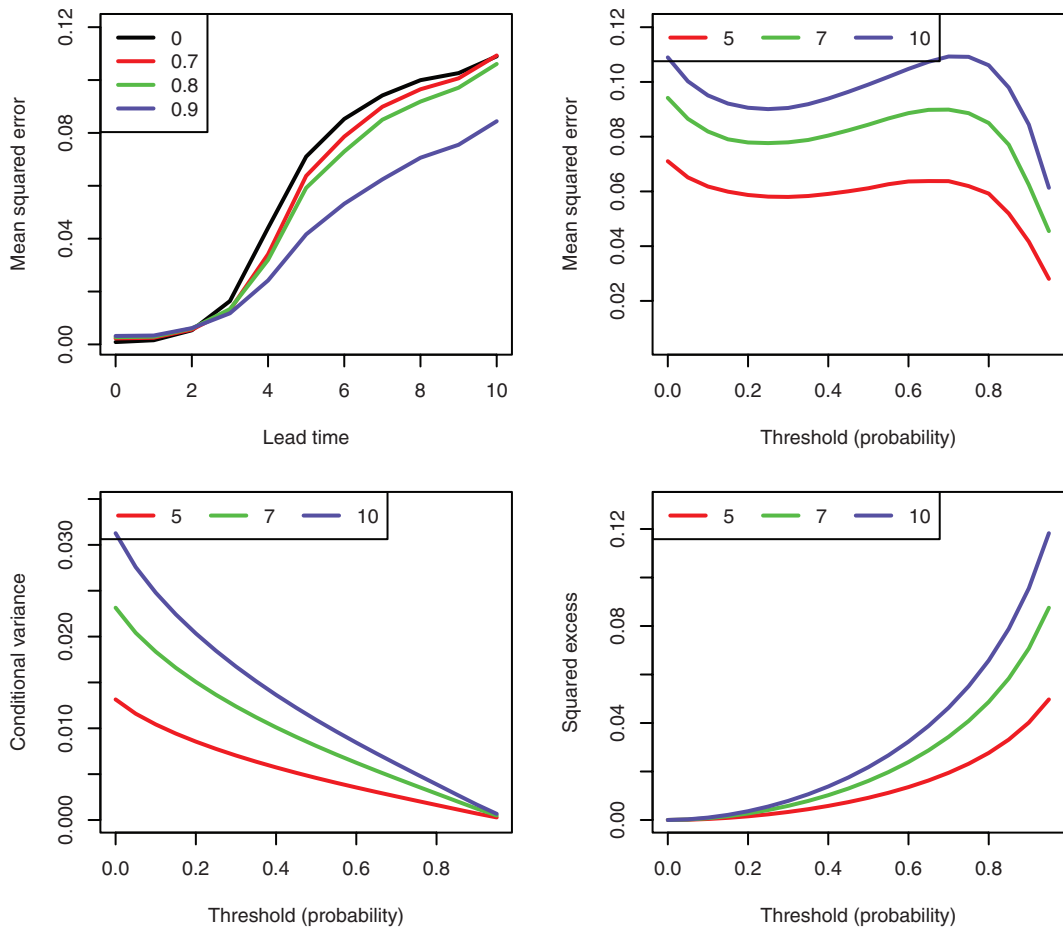
**Figure 6.** As Figure 2, but for the cusp map with the observable in equation (4) with  $g = g_3$ ,  $\alpha = 2$  and  $\zeta = 1$ .

which implies that  $\theta \in (0, 1)$ . In particular, this relation also shows that when the fixed point becomes less repelling, the extremal index decreases and hence the limiting mean cluster size of extremes becomes larger. However, for the cusp map, the point  $\zeta = -1$  is an *indifferent* fixed point, which means that  $|f'(\zeta)| = 1$ . In Freitas *et al.* (2016), it is proven for the Manneville–Pomeau map with the observable  $\phi(x) = -\log|x - \zeta|$  that the extremal index is zero when  $\zeta$  is an indifferent fixed point. In this case, the scaling sequence in equation (2) has to be modified. However, the proof techniques in Freitas *et al.* (2016) only work for a suitable range of parameter values and are not directly applicable to the cusp map.

For the cusp map, we do not have an analytical expression for the extremal index, but we can estimate  $\theta$  from the time series shown in Figure 8 using the algorithm described in Ferro and Segers (2003). For  $\zeta = -1$ , we have  $\theta \approx 0.002$  that lends support to the conjecture that  $\theta = 0$ . This numerical estimate is consistent with the result proved in Freitas *et al.* (2016) for the Manneville–Pomeau map. Also observe that in Figure 8, the clusters are very large. The predictability of extremes in the presence of clusters will be studied in more detail in forthcoming work. In particular, we will study how the results presented in Section 2.2 need to be modified when  $0 < \theta < 1$  and also in the extreme case  $\theta = 0$  for which formula (19) clearly no longer holds.

### 3.3. A low-order barotropic model

In this section, we consider a low-order model for atmospheric regime transitions. The existence and properties of an invariant measure are unknown. The model exhibits slow decay of correlations due to Shilnikov-like intermittency, and this implies that linear model in Section 2.2 remains valid for long lead times.



**Figure 7.** As Figure 6, but for  $\zeta = -1$ . In this case, the behaviour of the MSE with threshold is very different from the theoretical results presented in Section 2.2.

The barotropic vorticity equation on a  $\beta$ -plane channel for atmospheric flow over orography is given by

$$\frac{\partial}{\partial t} \Delta \psi = -J(\psi, \Delta \psi + \beta y + \gamma h) - C \Delta(\psi - \psi^*), \quad (21)$$

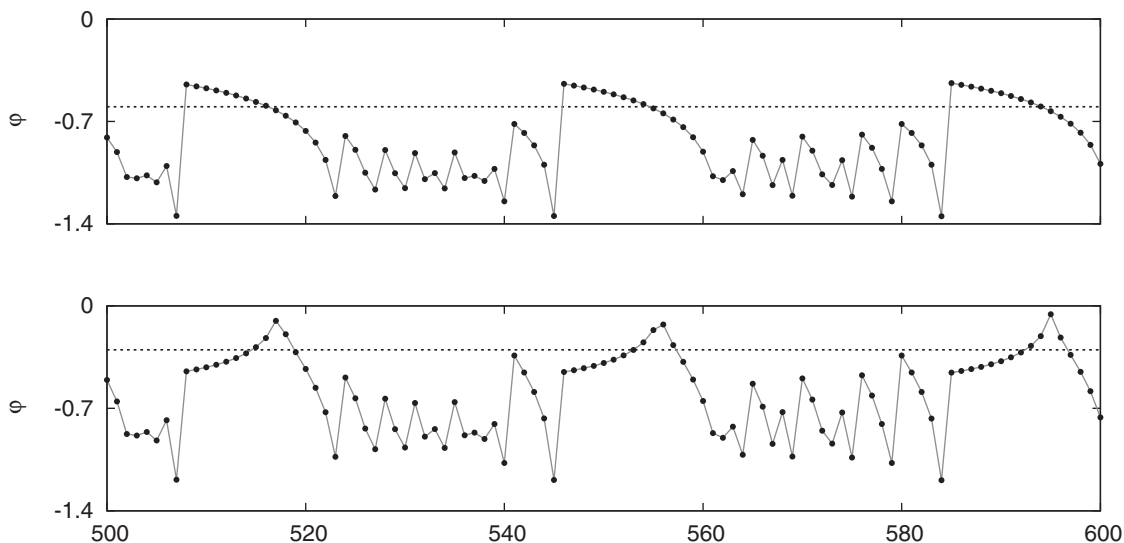
where  $\psi$  is the stream function,  $\psi^*$  is the forcing,  $\beta y$  is the Coriolis force,  $h$  is the orography and  $\Delta$  is the Laplace operator. In addition,  $J$  is the Jacobian operator which is defined as

$$J(f, g) = \frac{\partial f}{\partial x} \frac{\partial g}{\partial y} - \frac{\partial f}{\partial y} \frac{\partial g}{\partial x}.$$

The domain is  $[0, 2\pi] \times [0, \pi b]$  so that  $b/2$  is the aspect ratio of the  $\beta$ -plane channel. We take periodic boundary conditions in  $x$  and for  $y = 0, \pi b$  we require

$$\frac{\partial \psi}{\partial x} = 0 \quad \text{and} \quad \int_0^{2\pi} \frac{\partial \psi}{\partial y} dx = 0,$$

which, respectively, mean that on the lateral boundaries the northward component of the velocity vanishes and that the net zonal flow is zero.



**Figure 8.** Time series obtained by evaluating the observable  $\phi(x) = -|x - \zeta|^{1/2}$  along the iterates of the cusp map for  $\zeta = -1$  (top panel) and  $\zeta = -0.6$  (bottom panel). For  $\zeta = -1$ , exceedances over a threshold form large clusters.

Equation (21) can be spatially discretized by expanding the stream function as a Fourier series:

$$\psi(x, y, t) = \sum_{n,m} \widehat{\psi}_{n,m}(t) \phi_{n,m}(x, y), \quad \phi_{n,m}(x, y) = \begin{cases} \sqrt{2} \cos(my/b) & \text{if } n = 0, \\ \sqrt{2} e^{inx} \cos(my/b) & \text{if } n \neq 0, \end{cases} \quad (22)$$

and then applying a Galerkin projection onto a finite set of Fourier modes. Charney and DeVore (1979) studied a three-dimensional spectral truncation of (21) and discovered coexisting equilibria representing blocked and zonal flow patterns. De Swart (1989) studied higher order truncations, including the six-dimensional truncation

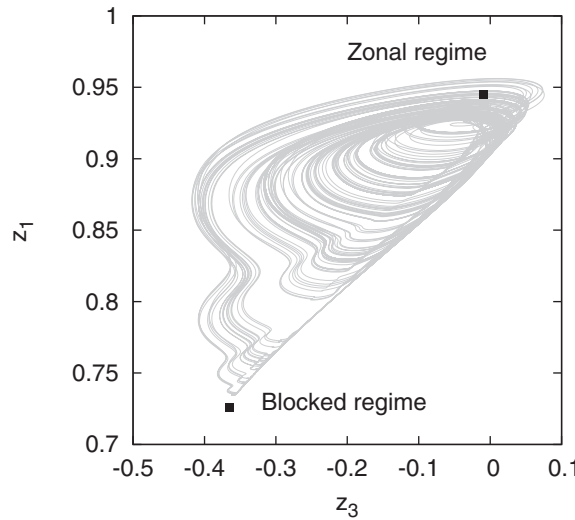
$$\begin{aligned} \dot{z}_1 &= \tilde{\gamma}_1 z_3 - C(z_1 - z_1^*), \\ \dot{z}_2 &= -(\alpha_1 z_1 - \beta_1) z_3 - C z_2 - \delta_1 z_4 z_6, \\ \dot{z}_3 &= (\alpha_1 z_1 - \beta_1) z_2 - \gamma_1 z_1 - C z_3 + \delta_1 z_4 z_5, \\ \dot{z}_4 &= \tilde{\gamma}_2 z_6 - C(z_4 - z_4^*) + \varepsilon(z_2 z_6 - z_3 z_5), \\ \dot{z}_5 &= -(\alpha_2 z_1 - \beta_2) z_6 - C z_5 - \delta_2 z_4 z_3, \\ \dot{z}_6 &= (\alpha_2 z_1 - \beta_2) z_5 - \gamma_2 z_4 - C z_6 + \delta_2 z_4 z_2, \end{aligned} \quad (23)$$

where the state space variables  $z_1$  and the Fourier coefficients  $\widehat{\psi}_{n,m}$  are related as follows:

$$\begin{aligned} z_1 &= \frac{1}{b} \widehat{\psi}_{0,1}, \\ z_2 &= \frac{1}{b\sqrt{2}} (\widehat{\psi}_{1,1} + \widehat{\psi}_{-1,1}), \\ z_3 &= \frac{i}{b\sqrt{2}} (\widehat{\psi}_{1,1} - \widehat{\psi}_{-1,1}), \\ z_4 &= \frac{1}{b} \widehat{\psi}_{0,2}, \\ z_5 &= \frac{1}{b\sqrt{2}} (\widehat{\psi}_{1,2} + \widehat{\psi}_{-1,2}), \\ z_6 &= \frac{i}{b\sqrt{2}} (\widehat{\psi}_{1,2} - \widehat{\psi}_{-1,2}). \end{aligned}$$

**Table 1.** Parameters and coefficients of the model (23) as used in [Crommelin et al. \(2004\)](#)

$C = 0.1$	$\alpha_2 = 32\sqrt{2}(b^2 + 3)/15\pi(b^2 + 4)$	$\gamma = 0.2$
$b = 0.5$	$\beta = 5/4$	$\gamma_1 = 4\sqrt{2}b\gamma/3\pi(b^2 + 1)$
$z_1^* = 0.95$	$\beta_1 = \beta b^2/(b^2 + 1)$	$\gamma_2 = 32\sqrt{2}b\gamma/15\pi(b^2 + 4)$
$r = -0.801$	$\beta_2 = \beta b^2/(b^2 + 4)$	$\tilde{\gamma}_1 = 4\sqrt{2}b\gamma/3\pi$
$z_4^* = rz_1^*$	$\delta_1 = 64\sqrt{2}b^2/15\pi(b^2 + 1)$	$\tilde{\gamma}_2 = 8\sqrt{2}b\gamma/15\pi$
$\alpha_1 = 8\sqrt{2}b^2/3\pi(b^2 + 1)$	$\delta_2 = 64\sqrt{2}(b^2 - 3)/15\pi(b^2 + 4)$	$\varepsilon = 16\sqrt{2}/15\pi$



**Figure 9.** Projection of a Shilnikov-like strange attractor of the low-order model (23) onto the  $(z_3, z_1)$ -plane. Evolutions intermittently visit two locations in state space that correspond to blocked and zonal flow patterns.

The parameters are chosen as in [Crommelin et al. \(2004\)](#) and are listed in [Table 1](#). With this choice, the  $\beta$ -plane channel  $[0, 2\pi] \times [0, \pi b]$  has dimensions  $5000 \times 1250$  km, the orography  $h(x, y) = \cos(x) \sin(y/b)$  has a dimensional amplitude of 200 m, and one unit of time corresponds to 1 day.

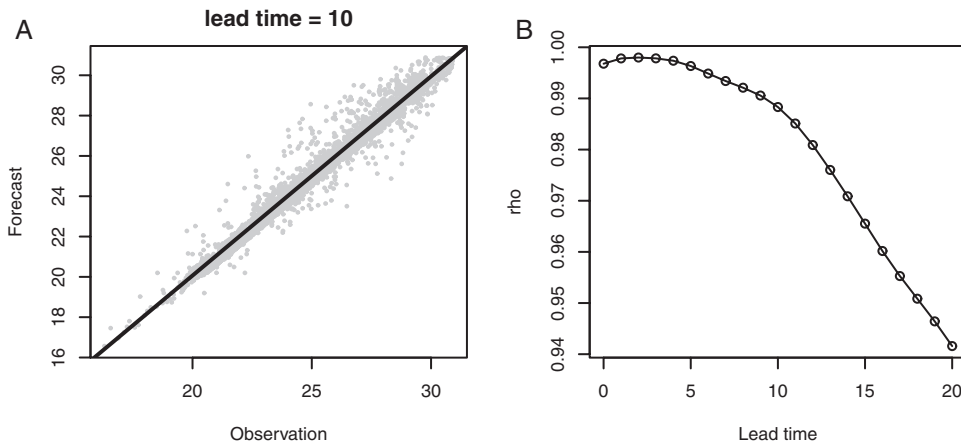
The low-order model (23) was used in [Crommelin et al. \(2004\)](#) to explain intermittent transitions between two regions in state space corresponding to zonal and blocked flow regimes (see [Fig. 9](#)). The dynamics consists of three recurrent episodes: (i) transitions from zonal to blocked flows, typically taking 30 days, (ii) transitions from blocked to zonal flows, typically taking 40–80 days, and (iii) spiralling behaviour around the zonal regime, typically lasting  $>200$  days. This intermittent behaviour can be explained by the presence of Shilnikov-like strange attractors due to homoclinic bifurcations taking place near a Hopf-saddle-node bifurcation ([Broer and Vegter, 1984](#); [Broer and Vitolo, 2008](#)).

Sampling the flow of (23) at multiples of a time step  $t_s$  gives a map  $f: \mathcal{X} \rightarrow \mathcal{X}$  with  $\mathcal{X} = \mathbb{R}^6$ . In the following, we will take  $t_s = 1$ , which corresponds to one day. We will consider observables that measure wind speed at particular points in the  $\beta$ -plane channel. For a given state vector  $(z_1, \dots, z_6) \in \mathcal{X}$  we can reconstruct the stream function  $\Psi$  from equation (22). The velocity field and the wind speed are then given by

$$(u, v) = \left(-\frac{\partial \Psi}{\partial y}, \frac{\partial \Psi}{\partial x}\right) \quad \text{and} \quad w = \sqrt{u^2 + v^2}$$

respectively. Evaluating the wind speed at a geographical location  $(x, y)$  in the  $\beta$ -plane channel gives observables  $\phi: \mathbb{R}^6 \rightarrow \mathbb{R}$ , which are indeed of the form  $\phi_W$  in equation (5), but their explicit expressions will be omitted. We will study wind speeds at three locations in the  $\beta$ -plane channel: the side of the orography  $(x, y) = (0, 3\pi b/4)$ , the top  $(x, y) = (0, \pi b/2)$  and the valley





**Figure 10.** Statistics of the time-1 map of the low-order model (23) and the observable  $\phi_{\text{sid}}$ . The scatter plot and the graph of the correlation coefficient  $\rho$  versus lead time lend support to the assumption that the forecasts and observations are related through the linear model (10). For the observables  $\phi_{\text{top}}$  and  $\phi_{\text{val}}$ , the results are similar (not shown).

$(x, y) = (\pi, \pi b/2)$ . The corresponding observables are denoted by  $\phi_{\text{sid}}$ ,  $\phi_{\text{top}}$  and  $\phi_{\text{val}}$ , respectively. The intermittent dynamics of (23) causes the correlation coefficient  $\rho_\tau$  to decrease slowly with prediction lead time  $\tau$ . This also implies that the linear model between the forecasts and observations remains valid for long lead times. For the observable  $\phi_{\text{sid}}$ , this is illustrated in Figure 10. The decay of  $\rho_\tau$  for the observables  $\phi_{\text{top}}$  and  $\phi_{\text{val}}$  is similar (not shown).

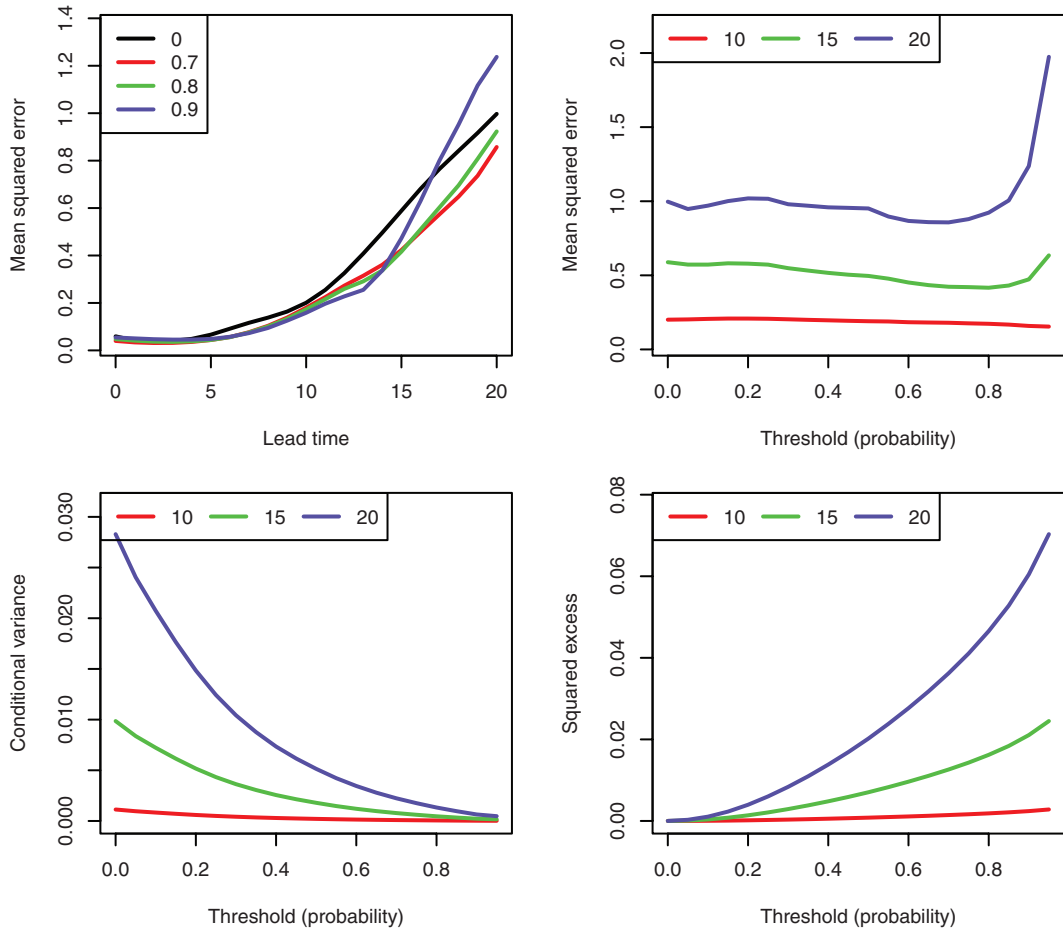
The low-order model (23) belongs to a class of forced dissipative systems in which the quadratic nonlinearities preserve the total energy. It is well known that such systems have a trapping region (Lorenz, 1980), which means that the orbits remain in some ball around the origin for all  $t \geq 0$ . Since the wind speed observables are continuous functions, their evolutions along orbits of the system have an upper bound. Therefore, we expect a priori that the process (1) has a negative tail index. In particular, we expect that the conditional variance term of the MSE always decreases as a function of threshold, but the rate of decrease might be different for the three observables.

Figure 11 shows the MSE and its decomposition for the time-1 map of (23) with the observable  $\phi_{\text{sid}}$ . For lead time  $\tau = 20$ , the MSE slowly decreases with threshold exceedance probability  $p$  but shows a steep increase after  $p = 0.8$ . The behaviour of the MSE decomposition terms is consistent with the fact that  $\xi < 0$ : the conditional variance term decreases and the squared excess term increases. A clear minimum in the MSE is not visible in this case. For the observable  $\phi_{\text{top}}$  (see Fig. 12), the MSE always seems to increase with  $p$ . The conditional variance term stays relatively constant up to  $p = 0.8$  and then suddenly decreases. The squared excess term increases fast after  $p = 0.8$ . For the observable  $\phi_{\text{val}}$  (see Fig. 13), the MSE is nearly constant as a function of threshold up to  $p = 0.7$  and increases fast thereafter. The conditional variance term decreases fast for low thresholds, whereas the squared excess term increases fast for high thresholds. This might explain why no minimum in the MSE is visible.

Figure 14 shows the attractor of the low-order model (23) in which colours indicate the error between the forecast and observation for the observables  $\phi_{\text{sid}}$  and  $\phi_{\text{val}}$  and lead time  $\tau = 10$ . These figures show that the prediction error is large for only a small fraction of points on the attractor and that most of these points are located near the zonal regime. The initial conditions leading to an extreme event at a specific lead time are precisely located in the unpredictable region of the attractor. This qualitatively explains why the MSE increases for large thresholds. The results presented in this article are consistent with results presented in our earlier article (Sterk et al., 2012), which were obtained by means of FTLEs. However, in the latter case, the unpredictable part of the attractor was significantly larger. This might be a consequence of the fact that FTLEs do not take into account that prediction errors saturate with lead time and hence overestimate the growth of prediction errors.

### 3.4. A low-order model for ENSO

As a final example we consider a low-order model for the ENSO phenomenon that was presented in Timmermann and Jin (2002). The model consists of the three prognostic equations



**Figure 11.** As Figure 2, but for the time-1 map of the low-order model (23) with the observable  $\phi_{\text{sid}}$ .

$$\begin{aligned}
 \dot{T}_1 &= -\alpha(T_1 - T_r) - \frac{u(T_2 - T_1)}{L/2}, \\
 \dot{T}_2 &= -\alpha(T_2 - T_r) - \frac{w(T_2 - T_{\text{sub}})}{H_m}, \\
 \dot{h}_1 &= -r\left(b_1 + \frac{bL}{2}\tau\right),
 \end{aligned} \tag{24}$$

where  $T_1$  and  $T_2$ , respectively, denote the western and eastern equatorial sea surface temperatures and  $b_1$  denotes the anomalous thermocline depth.

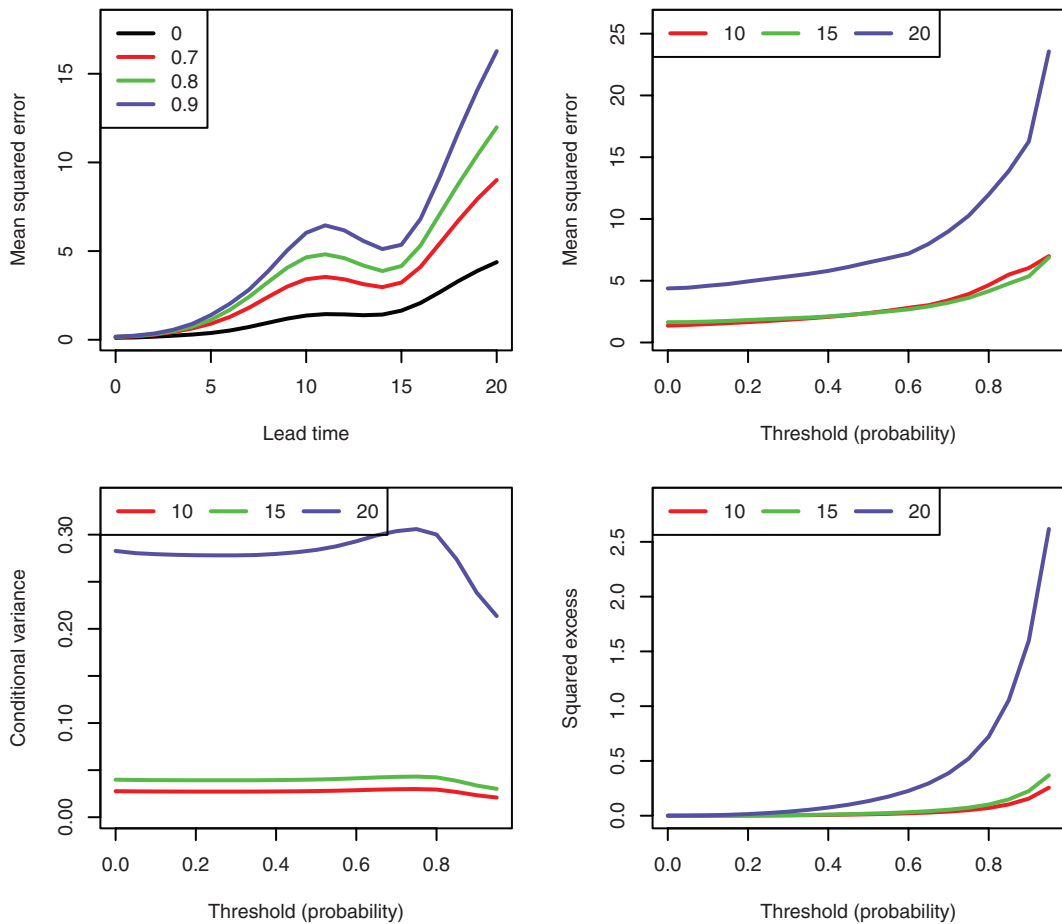
Wind stress anomalies  $\tau = -\mu(T_1 - T_2)/\beta$ , generated by the zonal temperature difference, drive both equatorial upwelling  $w$  in the eastern equatorial Pacific and zonal advection  $u$  via the diagnostic equations

$$\frac{u}{L/2} = \varepsilon\beta\tau \quad \text{and} \quad \frac{w}{H_m} = -\beta\tau.$$

The relation between the subsurface temperature and the thermocline depth is given by

$$T_{\text{sub}} = T_r - \frac{T_r - T_{r,0}}{2} \left[ 1 - \tanh\left(\frac{H + b_2 - z_0}{h^*}\right) \right],$$

where  $b_2 = b_1 + bL\tau$  is the depth deviation of the eastern equatorial thermocline from the reference depth  $H$  and  $z_0$  is the depth at which  $w$  takes its characteristic value. The parameter  $h^*$  controls the sharpness of the thermocline. For



**Figure 12.** As Figure 2, but for the time-1 map of the low-order model (23) with the observable  $\phi_{\text{top}}$ .

more details and physical justifications, see [Timmermann and Jin \(2002\)](#) and references therein. Together with these diagnostic relations, the model (24) is a closed system. The values of the parameters are listed in [Table 2](#).

For the parameters listed in [Table 2](#), the low-order model (24) has a Shilnikov-like chaotic attractor in which orbits spiral around an equilibrium of saddle-focus type (see [Fig. 15](#)). The equilibrium represents the tropical climate mean state of the model, and the spiralling motion corresponds to irregular temperature oscillations which lead to an ENSO event when  $T_2$  reaches a maximum. This regime behaviour is also reflected in the probability density of  $T_2$  shown in [Figure 16](#) which is clearly bimodal. The two modes can be interpreted as different regimes: the dominant mode  $T_2 = 21$  corresponds to the tropical climate mean, whereas the much weaker mode  $T_2 = 28.5$  corresponds to an ENSO event. As for the low-order atmosphere model presented in the previous section, [Figure 17](#) shows that correlations decay slowly in a non-monotonic manner.

The flow of (24) is sampled at intervals of 0.25 years and as observable we use the eastern equatorial sea surface temperature  $T_2$ . [Figure 18](#) shows the MSE and its decomposition terms as functions of lead time and threshold. The MSE is a non-monotonic function of  $\tau$ , and this behaviour becomes more pronounced for larger thresholds  $\mu$ . Note that the MSE has a local maximum at lead times for which the correlation coefficient  $\rho_\tau$  has a local minimum and vice versa. Non-monotonic behaviour of prediction errors as a function of lead time has been observed earlier in the context of an operational weather forecasting model ([Sterk et al., 2016](#)) and toy-models ([Sterk and Van Kekem, 2017](#)). This phenomenon might be a consequence of the intermittent dynamics of the system. Perhaps the simplest possible explanation can be given in terms of the cusp map in Section 3.2. [Figure 19](#) shows two predictions produced by the cusp map starting at two nearby initial conditions. The evolution starting at the initial condition  $x_0 = -0.9$  drifts away from

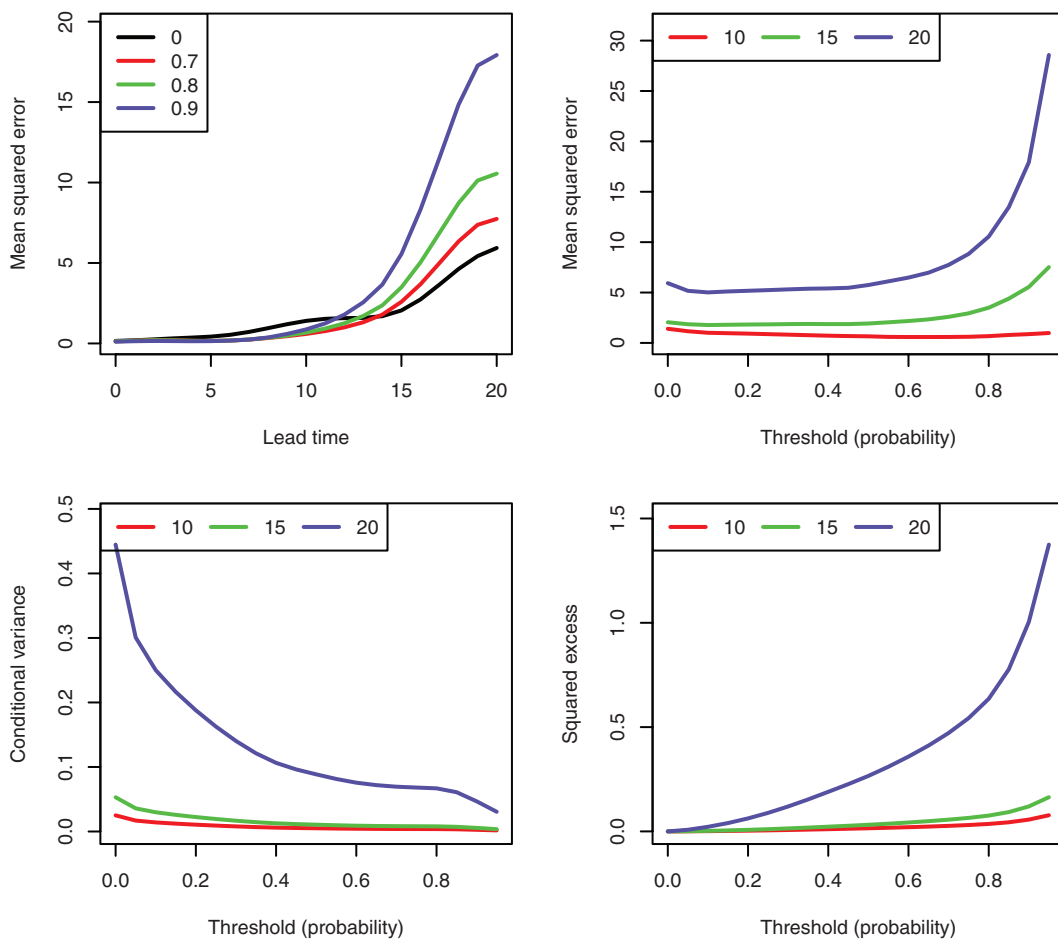


Figure 13. As Figure 2, but for the time-1 map of the low-order model (23) with the observable  $\phi_{val}$ .

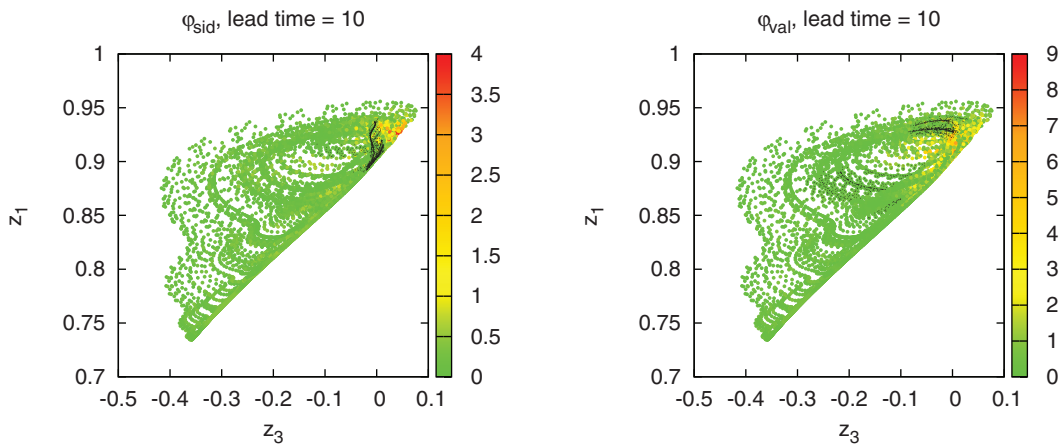
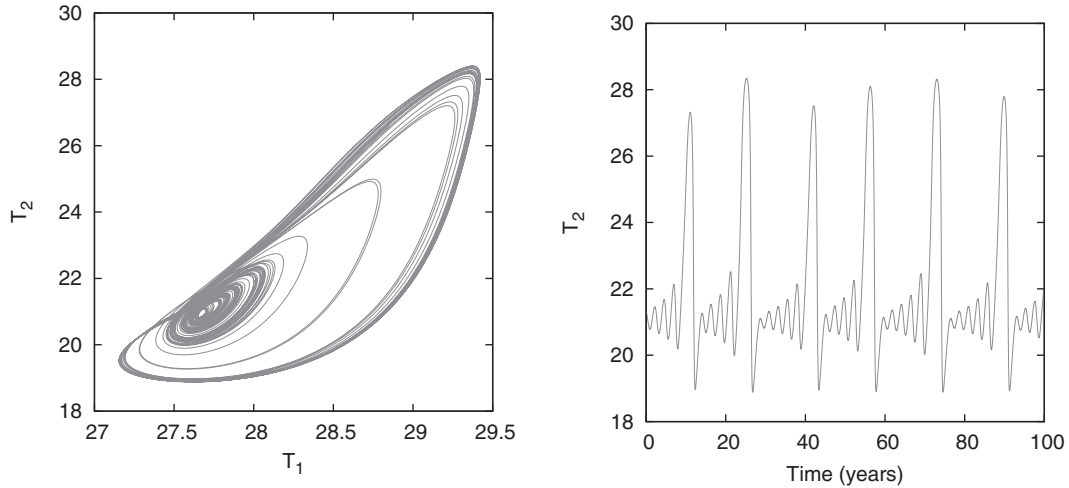


Figure 14. As Figure 9, but with colour indicating the error between the forecast and observation for the lead time  $\tau = 10$  days for the observables  $\phi_{sid}$  and  $\phi_{val}$ . The initial conditions that lead to an exceedance of the 90th percentile of the observable are marked with black dots. The result for the observable  $\phi_{top}$  is similar to that of  $\phi_{sid}$  (not shown).

**Table 2.** Parameter values of the low-order ENSO model (24) as used in Timmermann and Jin (2002)

$\alpha = 1/180 \text{ day}^{-1}$	$r = 1/400 \text{ day}^{-1}$
$T_{r,0} = 16^\circ\text{C}$	$T_r = 29.5^\circ\text{C}$
$H_m = 50 \text{ m}$	$H = 100 \text{ m}$
$z_0 = 75 \text{ m}$	$h^* = 62 \text{ m}$
$\mu = 0.0026 \text{ K}^{-1} \text{ day}^{-1}$	$\mu b L / \beta = 22 \text{ mK}^{-1}$
$L = 15 \times 10^6 \text{ m}$	$\varepsilon = 0.086$

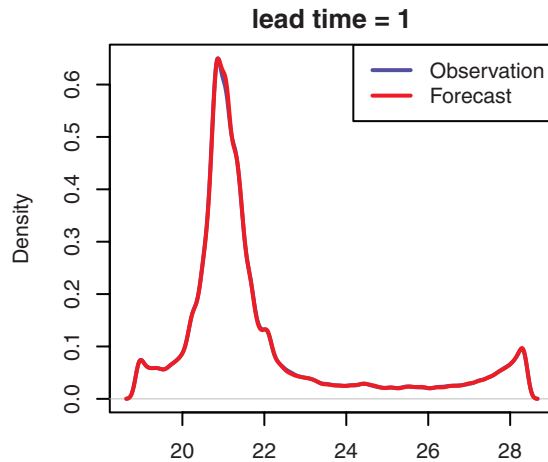
**Figure 15.** Projection of a Shilnikov-like strange attractor of the low-order model (24) onto the  $(T_1, T_2)$ -plane and a time series of the eastern equatorial sea surface temperature  $T_2$ .

the neutral fixed point at  $x = -1$ . The evolution starting at the initial condition  $y_0 = -0.8$  enters the chaotic regime at lead time  $\tau = 15$ . At lead time  $\tau = 22$ , the error suddenly decreases to almost zero again as the latter evolution is again reinjected in the laminar regime near the neutral fixed point.

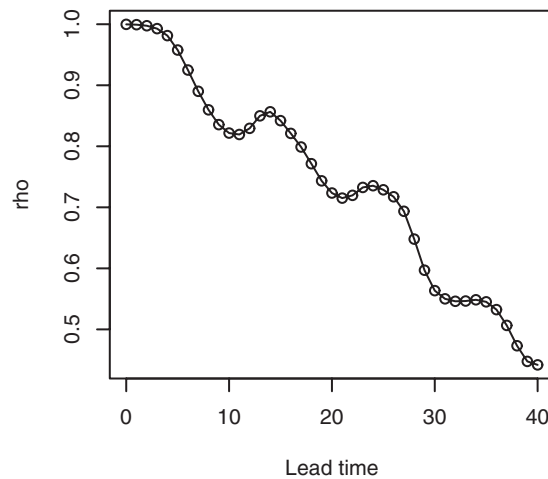
For fixed lead times, the MSE as a function of threshold is increasing. The conditional variance term shows a sharp decrease at large thresholds, whereas the squared excess term shows a fast increase at high thresholds. This explains why no minimum is observed. Figure 20 shows the attractor of the low-order model (24) in which colours indicate the error between the forecast and observation for the lead time  $\tau = 20$  (which corresponds to 5 years) and the observable  $T_2$ . As in Figure 14, the unpredictable part of the attractor is located near the saddle-focus equilibrium and initial conditions leading to an extreme event are located in the unpredictable region of the attractor.

#### 4. Discussion and conclusion

In this article, we have explored the relation between statistics and the predictability of extremes in dynamical systems. We have adopted a perfect model scenario by comparing forecasts with each other rather than comparing forecasts with observations. We measured predictability in terms of an MSE computed from the difference of a pair of forecasts conditional on one of the forecasts exceeding a threshold. We showed that the MSE can be written as a sum of three terms: a threshold-independent constant, a mean term that always increases with threshold and a variance term that can either increase, decrease or stay constant with threshold. Using the GPD to model excesses over a threshold, we showed that the MSE always increases with threshold at sufficiently high threshold. However, when the forecasts have a negative tail index, the MSE can be a decreasing function of threshold at lower thresholds. In this article, we have only considered twin forecasts, but the methodology can be extended to ensemble forecasts (Sterk *et al.*, 2016).



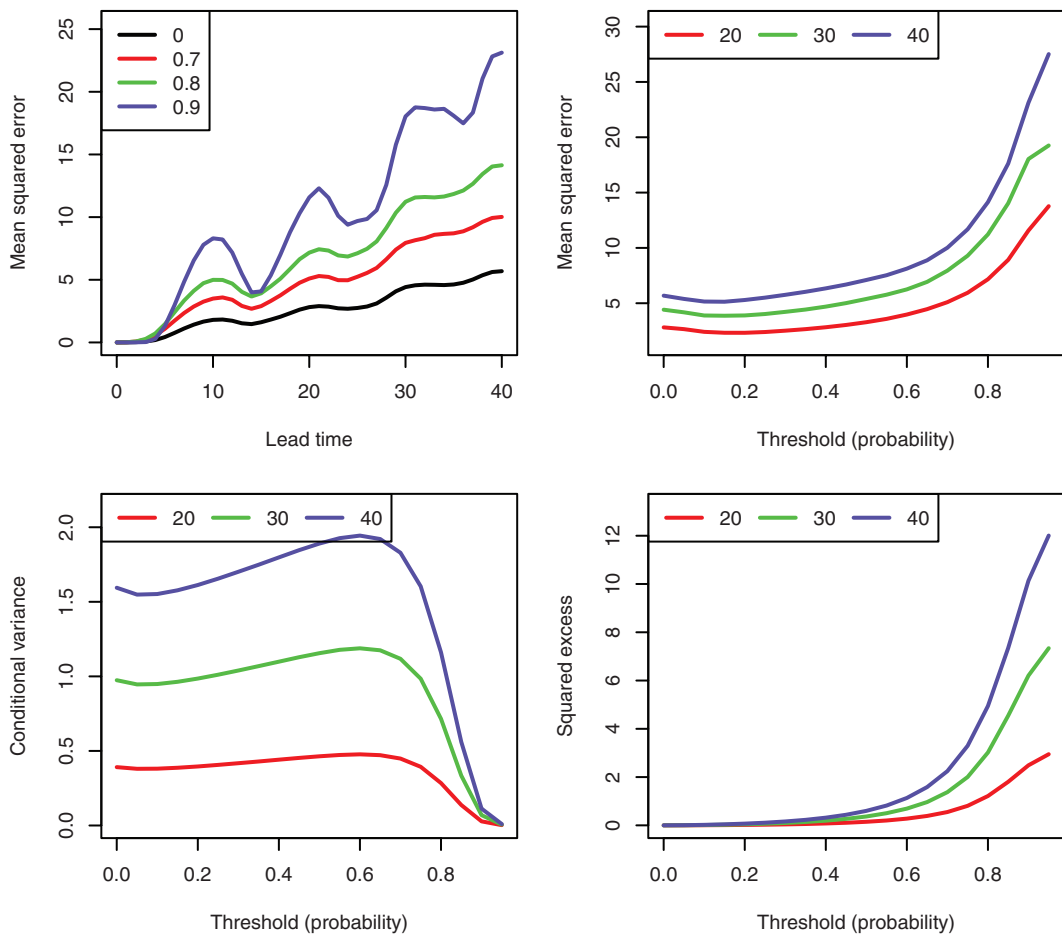
**Figure 16.** Empirical probability density functions for forecasts and observations produced by the low-order model (24) with the observable  $\phi(T_1, T_2, b_1) = T_2$ . Note that the densities are bimodal. The largest mode corresponds to the tropical climate mean of the model, whereas the weaker mode corresponds to an ENSO event.



**Figure 17.** Graph of the correlation coefficient of the linear model (10) as a function of the prediction lead time for the time-0.25 map of the low-order model (24) with the observable  $T_2$ . Note that as in the case of the low-order barotropic model (23), the correlations decay slowly with lead time.

In earlier articles (Sterk *et al.*, 2012; Sterk and Van Kekem, 2017), we have used FTLE-like predictability indicators to assess the predictability of extremes. However, these indicators have at least two disadvantages. First, these indicators are based on the assumption that errors in the initial condition are infinitesimally small, which is a very strong idealization. Second, their computation requires the explicit availability of a prediction model and its variational equations (also referred to as a tangent linear model). If the prediction model has a high-dimensional state space, then computationally expensive algorithms are needed to obtain FTLEs from singular value decompositions. With the method presented in this article, one can study the effect of finite size errors. More importantly, our method can be used in a data-driven framework in which a prediction model is not explicitly available. Indeed, only a data set of forecasts and observations is required.

Our methodology is applicable under the condition that forecast/observation pairs satisfy a linear relation for sufficiently large prediction lead times. This condition is met when the correlations decay sufficiently slow. For example, this is the case for the cusp map and the low-order atmosphere and ocean models discussed in Section 3. All these case studies exhibit a form of intermittent dynamics, i.e. alternating phases of regular and chaotic dynamics. Intermittency plays an important role in the dynamics of climate models and has been proposed as an explanation for atmospheric

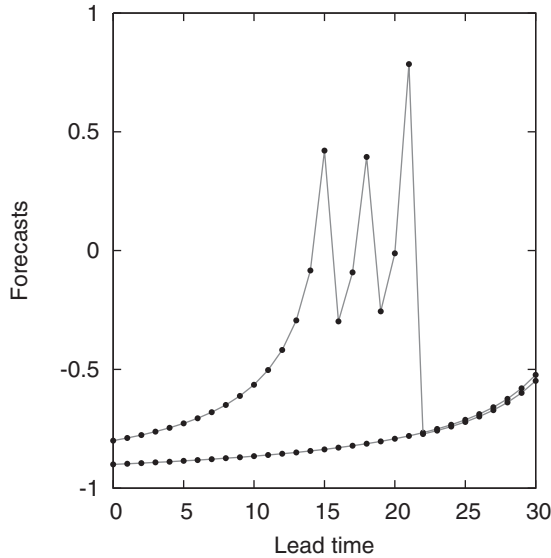


**Figure 18.** As Figure 2, but for the time-0.25 map of the low-order model (24) with the observable  $T_2$ .

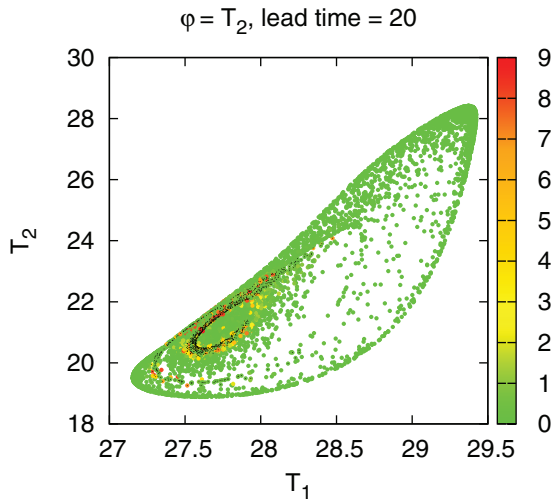
low-frequency variability due to homo- and heteroclinic connections (Crommelin, 2002, 2003) and bifurcations of periodic orbits (Sterk *et al.*, 2010), atmospheric regime dynamics (Crommelin *et al.*, 2004) and active and passive ocean regimes in coupled atmosphere–ocean models (Van Veen *et al.*, 2001). We expect that our methodology can be applied in a wide range of climate contexts in which correlations decay sufficiently slow.

The examples presented in this article show that predictability strongly depends on the invariant measure of the system and the regularity of the observable. For observables of the form (4), the forecasts can have a non-negative tail index in which case the MSE always increases with threshold. In climate models, however, observables are typically of the form (5), which are continuous functions on the state space  $\mathcal{X} = \mathbb{R}^d$ . Climate models often have dissipation terms, such as linear damping or diffusion, which implies that their evolutions are confined to attractors which are compact sets in state space. This means that observables of the form (5) attain a maximum value, which implies that the forecasts have a negative tail index. In this case, the MSE might decrease to a minimum before eventually increasing at high thresholds. Whether the minimum is observed or not, however, depends also on the location and scale parameters.

We now point out two directions for further research. In this article, we have studied predictability in the perfect model scenario that amounts to comparing forecasts with each other rather than comparing forecasts with real observations. An interesting direction topic for future research is the extension of our method to an imperfect model scenario. In that case, the forecast and observation do not have the same distribution, which implies that the assumptions of Lemma 1 are not satisfied and equation (12) is no longer applicable. Perhaps a more refined statistical model



**Figure 19.** Two forecasts produced by the cusp map  $f(x) = 1 - 2\sqrt{|x|}$  using the initial conditions  $x_0 = -0.9$  and  $y_0 = -0.8$ . Initially, the two forecasts drift apart, but at lead time  $\tau = 22$ , the error between the two forecasts suddenly decreases again and almost vanishes.



**Figure 20.** As Figure 14, but for the time-0.25 map of the low-order model (24) with the observable  $T_2$ . Note that the lead time  $\tau = 20$  corresponds to 5 years.

can be constructed to relate forecasts and observations to obtain a decomposition of the MSE. Such a statistical model should take into account the bias between the predictions and the observations.

A second important direction for future research is to adapt the methodology in Section 2.2 to the setting of clustered extremes. The cusp map example shows that when an observable is maximized at a fixed point, then the extremes form clusters. In this case, the MSE can *decrease* at high thresholds. When the observable is maximized at a point sufficiently close to this fixed point, the extremes may still show a moderate degree of clustering. To describe the statistics of clustered extremes, the location and scale parameters in extreme value law (3) have to be modified with an additional parameter known as the extremal index. The latter parameter can be interpreted as the reciprocal of the



mean cluster size. We believe that detailed studies of one-dimensional maps can help to obtain a better mathematical understanding of the predictability of clustered extremes. Of particular interest are the so-called intermittency maps that exhibit slow decay of correlations (Holland, 2005; Bhansali and Holland, 2007).

Clustering of extremes is not only relevant in the setting of low-dimensional dynamical systems, but also for meteorological phenomena. For example, a heat wave or a cold spell might persist for a couple of days, and one might be interested in predicting the duration of an extreme event. In this context, it is also interesting to look at observables different from those in equation (5). For persistent extremes, one might be interested in observables that exceed a threshold for a subsequent number of elements in a time series, and for such observables, the extreme value laws are probably different from equation (3). Another example is recent observational work that shows that large-scale flow patterns, such as the North Atlantic Oscillation, which itself might be a manifestation of intermittency, cause temporal clustering of storms (Mailier *et al.*, 2006; Vitolo *et al.*, 2009). A better understanding of clustered extremes in dynamical systems may therefore also lead to a better understanding of real-world extremes.

## Declaration

Funding: EPSRC (EP/P034489/1).

Ethical approval: none.

Conflict of interest: none.

## Acknowledgements

The anonymous referees are kindly thanked for their comments and suggestions that have helped to improve this article.

## References

- Balkema AA, de Haan L. Residual life time at great age. *Ann Probab* 1974; 2: 792–804.
- Beirlant J, Goegebeur Y, Teugels J, Segers J. *Statistics of Extremes: Theory and Applications*. Berlin, Germany: John Wiley and Sons, 2004.
- Bhansali RJ, Holland MP. Frequency analysis of chaotic intermittency maps with slowly decaying correlations. *Stat Sinica* 2007; 17: 15–41.
- Bodai T. Predictability of threshold exceedances in dynamical systems. *Phys D Nonlinear Phenom* 2015; 313: 37–50.
- Boyarsky A, Góra P. *Laws of Chaos. Invariant Measures and Dynamical Systems in One Dimension*. Boston, MA: Birkhäuser, 1997.
- Broer HW, Vegter G. Subordinate Šil'nikov bifurcations near some singularities of vector fields having low codimension. *Ergod Theory Dyn Syst* 1984; 4: 509–25.
- Broer HW, Vitolo R. Dynamical systems modelling of low-frequency variability in low-order atmospheric models. *Discrete Contin Dyn Syst B* 2008; 10: 401–19.
- Castillo E. *Extreme Value Theory in Engineering. Statistical Modeling and Decision Science*. Boston, MA: Academic Press, 1988.
- Charney JG, DeVore JG. Multiple flow equilibria in the atmosphere and blocking. *J Atmos Sci* 1979; 36: 1205–16.
- Chazottes JR, Collet P. Poisson approximation for the number of visits to balls in non-uniformly hyperbolic dynamical systems. *Ergod Theory Dyn Syst* 2013; 33: 49–80.
- Coles S. *An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics*. New York, NY: Springer, 2001.
- Collet P. Statistics of closest return for some non-uniformly hyperbolic systems. *Ergod Theory Dyn Syst* 2001; 21: 401–20.
- Crommelin DT. Homoclinic dynamics: a scenario for atmospheric ultralow-frequency variability. *J Atmos Sci* 2002; 59: 1533–49.
- Crommelin DT. Regime transitions and heteroclinic connections in a barotropic atmosphere. *J Atmos Sci* 2003; 60: 229–46.
- Crommelin DT, Opsteegh JD, Verhulst F. A mechanism for atmospheric regime behavior. *J Atmos Sci* 2004; 61: 1406–19.
- De Swart HE. Analysis of a six-component atmospheric spectral model: chaos, predictability and vacillation. *Phys D Nonlin Phenom* 1989; 36: 222–34.
- Embrechts P, Klüppelberg C, Mikosch T. *Modelling Extremal Events, Volume 33 of Applications of Mathematics: Stochastic Modelling and Applied Probability*. Berlin: Springer-Verlag, 1997.
- Faranda D, Lucarini V, Turchetti G, Vaienti S. Extreme value distributions for singular measures. *Chaos* 2012; 22: 023135.
- Ferro CAT, Segers J. Inference for clusters of extreme values. *J R Stat Soc Ser B Stat Methodol* 2003; 65: 545–56.
- Franzke C. Predictability of extreme events in a nonlinear stochastic-dynamic model. *Phys Rev E* 2012; 85: 031134.
- Franzke C. Extremes in dynamic-stochastic systems. *Chaos* 2017; 27: 012101.
- Freitas ACM. Statistics of the maximum for the tent map. *Chaos Solitons Fract* 2009; 42: 604–8.

- Freitas ACM, Freitas JM. On the link between dependence and independence in extreme value theory for dynamical systems. *Stat Probab Lett* 2008; 78: 1088–93.
- Freitas JM, Freitas ACM, Todd M. Hitting times and extreme value theory. *Probab Theory Relat Fields* 2010; 147: 675–710.
- Freitas JM, Freitas ACM, Todd M. Extreme value laws in dynamical systems for non-smooth observations. *J Stat Phys* 2011; 142: 108–26.
- Freitas ACM, Freitas JM, Todd M. Extremal index, hitting time statistics and periodicity. *Adv Math* 2012; 231: 2626–65.
- Freitas ACM, Freitas JM, Todd M, Vaienti S. Rare events for the Manneville–Pomeau map. *Stoch Process Appl* 2016; 126: 3463–79.
- Galambos J. *The Asymptotic Theory of Extreme Order Statistics*. New York, NY: John Wiley and Sons, 1978.
- Gupta C. Extreme value distributions for some classes of non-uniformly partially hyperbolic dynamical systems. *Ergod Theory Dyn Syst* 2011; 30: 757–71.
- Hallerberg S, Altmann EG, Holstein D, Kantz H. Precursors of extreme increments. *Phys Rev E* 2007; 75: 016706.
- Hallerberg S, Kantz H. How does the quality of a prediction depend on the magnitude of the events under study? *Nonlinear Process Geophys* 2008a; 15: 321–31.
- Hallerberg S, Kantz H. Influence of the event magnitude on the predictability of an extreme event. *Phys Rev E* 2008b; 77: 011108.
- Harle M, Kwasiok F, Feudel U. Growth of finite errors in ensemble prediction. *Nonlinear Process Geophys* 2006; 13: 167–76.
- Hemmer PC. The exact invariant density for a cusp-shaped return map. *J Phys A Math Gen* 1984; 17: L247.
- Holland MP. Slowly mixing systems and intermittency maps. *Ergod Theory Dyn Syst* 2005; 25: 133–59.
- Holland MP, Nicol M, Török A. Extreme value theory for non-uniformly expanding dynamical systems. *Trans Am Math Soc* 2012a; 364: 661–88.
- Holland MP, Rabassa P, Sterk AE. Quantitative recurrence statistics and convergence to an extreme value distribution for non-uniformly hyperbolic dynamical systems. *Nonlinearity* 2016; 29: 2355–94.
- Holland MP, Vitolo R, Rabassa P, Sterk AE, Broer HW. Extreme value laws in dynamical systems under physical observables. *Phys D* 2012b; 241: 497–513.
- Leadbetter MR, Lindgren G, Rootzén H. *Extremes and Related Properties of Random Sequences and Processes*. New York, NY: Springer-Verlag, 1983.
- Lorenz EN. Deterministic nonperiodic flow. *J Atmos Sci* 1963; 20: 130–41.
- Lorenz EN. Attractor sets and quasi-geostrophic equilibrium. *J Atmos Sci* 1980; 37: 1685–99.
- Lucarini V, Faranda D, Freitas ACM et al. *Extremes and Recurrence in Dynamical Systems*. Hoboken, NJ: Wiley, 2016.
- Mailier PJ, Stephenson DB, Ferro CAT, Hodges KI. Serial clustering of extratropical cyclones. *Mon Weather Rev* 2006; 134: 2224–40.
- Pickands J. Statistical inference using extreme order statistics. *Ann Stat* 1975; 3: 119–31.
- Resnick SI. *Extreme Values, Regular Variation, and Point Processes, Volume 4 of Applied Probability: A Series of the Applied Probability Trust*. New York, NY: Springer-Verlag, 1987.
- Ross S. *A First Course in Probability*, 5th edn. Upper Saddle River, NJ: Prentice Hall, 1997.
- Sterk AE, Holland MP, Rabassa MP, Broer HW, Vitolo R. Predictability of extreme values in geophysical models. *Nonlinear Process Geophys* 2012; 19: 529–39.
- Sterk AE, Stephenson DB, Holland MP, Mylne KR. On the predictability of extremes: does the butterfly effect ever decrease? *Q J R Meteorol Soc* 2016; 142: 58–64.
- Sterk AE, Van Kekem DL. Predictability of extreme waves in the Lorenz-96 model near intermittency and quasi-periodicity. *Complexity* 2017; 2017: 9419024.
- Sterk AE, Vitolo R, Broer HW, Simó C, Dijkstra HA. New nonlinear mechanisms of midlatitude atmospheric low-frequency variability. *Phys D Nonlinear Phenom* 2010; 239: 702–18.
- Timmermann A, Jin FF. A nonlinear mechanism for decadal El Niño amplitude changes. *Geophys Res Lett* 2002; 29: 3-1–3-4.
- Van Veen L, Opsteegh T, Verhulst F. Active and passive ocean regimes in a low-order climate model. *Tellus A* 2001; 53: 616–28.
- Vitolo R, Stephenson DB, Cook IM, Mitchell-Wallace K. Serial clustering of intense European storms. *Meteorol Z* 2009; 18: 411–24.