Clim Dyn (2014) 42:253–270
DOI 10.1007/s00382-013-1759-z

Intrinsic modulation of ENSO predictability viewed through a local Lyapunov lens

Christina Karamperidou · Mark A. Cane · Upmanu Lall · Andrew T. Wittenberg

Received: 14 August 2012 / Accepted: 30 March 2013 / Published online: 13 April 2013

Abstract The presence of rich ENSO variability in the long unforced simulation of GFDL's CM2.1 motivates the use of tools from dynamical systems theory to study variability in ENSO predictability, and its connections to ENSO magnitude, frequency, and physical evolution. Local Lyapunov exponents (LLEs) estimated from the monthly NINO3 SSTa model output are used to characterize periods of increased or decreased predictability. The LLEs describe the growth of infinitesimal perturbations due to internal variability, and are a measure of the immediate predictive uncertainty at any given point in the system phase-space. The LLE-derived predictability estimates are compared with those obtained from the error growth in a set of re-forecast experiments with CM2.1. It is shown that the LLEs underestimate the error growth for short forecast lead times (less than 8 months), while they overestimate it for longer lead times. The departure of LLE-derived error growth rates from the re-forecast rates is a linear function of forecast lead time, and is also sensitive to the length of the time series used for the LLE calculation. The LLE-derived error growth rate is closer to that estimated from the re-forecasts for a lead time of 4 months. In the 2,000-year long simulation, the LLE-derived predictability at the 4-month lead time varies (multi)decadally only by 9–18 %. Active ENSO periods are more predictable than inactive ones, while epochs with regular periodicity and moderate magnitude are classified as the most predictable by the LLEs. Events with a deeper thermocline in the west Pacific up to five years prior to their peak, along with an earlier deepening of the thermocline in the east Pacific in the months preceding the peak, are classified as more predictable. Also, the GCM is found to be less predictable than nature under this measure of predictability.

Keywords ENSO · Predictability · Local Lyapunov exponents

C. Karamperidou (✉) · U. Lall
Earth and Environmental Engineering Department, Columbia University, 500 West 120th Street, Mudd Bldg, New York, NY 10027, USA
e-mail: ck2424@columbia.edu

C. Karamperidou
Department of Meteorology, School of Ocean and Earth Sciences and Technology, University of Hawaii, Honolulu, USA

M. A. Cane
Lamont-Doherty Earth Observatory, 61 Route 9W, PO Box 1000, Palisades, NY 10964, USA

A. T. Wittenberg
US DOC/NOAA/GFDL, Forrestal Campus, US Route 1, Princeton, NJ 08542, USA

1 Introduction

The limits of predictability of the state of the tropical Pacific are still not known, and the accuracy and range of ENSO predictions in dynamical, statistical, or hybrid models have not improved substantially since the first dynamical forecast by Cane et al. (1986) (Barnston et al. 1999, 2011). The skill of models used at present for operational ENSO forecasting varies with forecast lead-times (Landsea and Knaff 2000): depending on the event, models have provided skillful short-range (0–3 months lead), medium-range (6–9 months lead) and long-range (12–22 months lead) forecasts (e.g. Ruiz et al. 2005; Drosdowsky 2006; Lima et al. 2009). It has been shown that model skill depends on the amplitude of interannual ENSO variability, with active ENSO periods tending to be

better predicted than weaker ones (Kirtman and Schopf 1998). In addition, skill varies decadally (Chen et al. 2004) with decadal variations in tropical Pacific climate, and the limitations in our understanding of Pacific decadal variability impose limits on the skill of ENSO predictions (see Hazeleger et al. 2001; Wittenberg 2002; Karspeck et al. 2004; Power et al. 2006; Penland and Matrosova 2006, and references therein).

The limits of predictability depend on the mechanisms responsible for ENSO irregularity and equilibration at finite amplitude (Sarachik and Cane 2010). The former have been linked to either chaos (Munnich et al. 1991; Jin et al. 1994; Tziperman et al. 1994; Timmermann and Jin 2002; Ghil et al. 2008), or noise (Wittenberg 2002; Vecchi et al. 2006; Gebbie et al. 2007; Zavala-Garay et al. 2008; Kleeman 2008). Both depend on the stability—or lack of it—of the atmosphere–ocean interactions. It is possible that real ENSO behavior emerges from the dialectics of chaos and noise, with each dominating in any given decade.

Decadal variability of ENSO is present in historical and paleoclimate records, and has been simulated by a hierarchy of dynamical and statistical models (Vecchi and Wittenberg 2010; Collins et al. 2010; DiNezio et al. 2012; Emile-Geay et al. 2013a, b). As discussed in Lin (2007), the representation of the inter-decadal variability of ENSO in the Coupled GCMs participating in the IPCC AR4 ranges from constant periodicity or amplitude to significant inter-decadal variability in both period and amplitude. While long runs of intermediate dynamical models, such as the ZC model (Zebiak and Cane 1987), that exhibit inter-decadal and inter-centennial variability have been a subject of numerous studies, only recently have long runs of coupled GCMs become available. Wittenberg (2009) discusses the strong inter-decadal and inter-centennial ENSO variability in the 2,000-year run of the GFDL CM2.1 coupled GCM with solar irradiance, land cover and atmospheric composition held constant at 1,860 values (Wittenberg et al. 2006). The presence of such rich variability in the absence of variations in solar or volcanic forcing that could induce persistent regimes (Mann et al. 2005; Emile-Geay et al. 2007; Guilyardi et al. 2009), along with the length of the simulation, provides new ground for investigation of the causes of long-term modulation of ENSO behavior, and the implications for predictability at time-scales from the short-range to the decadal. Wittenberg (2009) shows that beyond 10 years the CM2.1 wait times between moderate-to-strong warm event peaks are indistinguishable from those of a Poisson process, thus indicating that ENSO modulation at inter-decadal and inter-centennial time scales need not require multi-decadal memory in the system.

Here, we address questions surrounding the variability in ENSO predictability in the context of dynamical systems theory. First, we briefly discuss concepts from the ergodic theory of dynamical systems used in this study, and some considerations in applying them to a high-dimensional model (Sect. 2). We then calculate estimates of local Lyapunov exponents (LLEs) from the monthly NINO3 SSTa time series of a 2,000-year long simulation from GFDL's CM2.1 model (Wittenberg et al. 2006; Wittenberg 2009) as a measure of ENSO predictability (Sect. 3). The LLEs are particularly useful in characterizing predictability locally in the attractor of a system that likely passes through phases of increased or decreased predictability. They describe the growth of infinitesimal perturbations due to internal variability over a finite time, and are a measure of the immediate evolution of predictive uncertainty at any given point in the system phase-space. However, practical ENSO forecasting is concerned with the evolution of finite, and not infinitesimal, uncertainties. In order to assess the relevance of the LLE-derived estimates of predictability to actual predictability, we compare error growth rates calculated from the LLEs with that calculated from a set of re-forecast experiments initialized from the same 2,000-year long CM2.1 simulation (Sect. 4). We show that the relationship between LLE-derived and actual predictability is dependent on the forecast lead time, and is sensitive to the length of simulation. We test a set of hypotheses to explain discrepancies between LLE-derived and actual predictability at different lead times, and show that the LLE-derived error growth rate is closer to actual growth rate at a lead time of 4-months.

This study then uses the LLE estimates at a lead time of 4 months as a relative measure to study fluctuations in predictability of this kind in the long 2,000-year simulation (Sect. 5). We also compare model predictability to predictability in nature quantified by the same methods (Sect. 6). We find that the loss of information in the GCM is faster than that inferred from observations, more so in active periods (post-1960). Finally, we explore the relationship between predictability and the evolution of SST and upper-ocean heat content anomalies (Sect. 7). We show that events with more heat pile-up in the Western Pacific five years prior to their peak, and with a deepening of the thermocline 3 months before the onset of the SST anomalies are more predictable.

To our knowledge, this is the first attempt to use LLEs to characterize predictability based on the output of a comprehensive climate model. Such measures have been primarily used within the lower levels of the hierarchy of models, and caution is required when one attempts to generalize these methods to models of higher complexity. Thus, the present paper serves as an initial assessment of the applicability of such methods to GCM output. In addition, the use of a computationally cheap relative measure of predictability in GCM simulations is interesting

for intra- and inter-model comparisons: characterizing relative predictability in a long climate model experiment could prove useful for identifying periods of the simulation that warrant further investigation of their dominant dynamics. Furthermore, a measure of relative predictability derived from model output can be useful for guiding classical predictability studies, e.g. for selecting initial conditions for running re-forecast experiments. Lastly, local predictability measures offer the possibility for more detailed comparisons between models and observations compared to global measures, such as the global Lyapunov exponents, since the models and the observations are required to agree at a large number of values rather than just one estimate of general predictability (Abarbanel et al. 1992).

2 Theory

2.1 Key concepts

We offer here a brief discussion of the key concepts of the ergodic theory of dynamical systems. The reader is referred to Abarbanel (1995) for details of the concepts that follow.

A commonly used method in studies of dynamical systems is phase-space reconstruction by time-delaying of a single state variable X (Takens 1981; Sauer et al. 1991). The key concept behind this method is that the time history of a single variable may act as a proxy for any of the variables of the system, since all variables are connected in a dynamical process. The measurement $X(t)$ is a result of the evolution of the dynamical system from $X(t-\tau)$ over a period τ during which all dynamical variables affect the variable X. Thus, the d-dimensional vector $\{X(t), X(t-\tau), X(t-2\tau), \ldots, X(t-d\cdot\tau)\}$ of the time delays of $X(t)$ stands as a proxy for observing d variables of the system. The dynamical system evolves over time towards subsets of the phase space, known as *attractors*.

To reconstruct the phase space one needs to determine the appropriate *embedding dimension* d and *time lag* τ.

The False Nearest Neighbor (FNN) method (Kennel et al. 1992) determines the sufficient embedding dimension for unfolding of the attractor in the following way: A vector $\mathbf{x}(t)$ in the d-dimensional phase space has neighboring vectors, denoted by $\mathbf{x}^{NN}(t)$. If a given $\mathbf{x}^{NN}(t)$ is a true neighbor of $\mathbf{x}(t)$, then it came to the neighborhood of $\mathbf{x(t)}$ through the evolution of the system dynamics. If, on the other hand, $\mathbf{x}^{NN}(t)$ is a false neighbor, then it has arrived in the neighborhood of $\mathbf{x}(t)$ not due to the dynamics, but by projection from a higher dimension: The dimension d does not unfold the attractor, and by moving to the next dimension $d+1$, the said false neighbor will be removed from the neighborhood of $\mathbf{x}(t)$. Consequently, the sufficient embedding dimension to unfold the attractor may be defined as the dimension above which no more false neighbors can be found.

Abarbanel et al (1993) tested the robustness of the FNN method by examining the effect of adding noise to a signal from the Lorenz attractor. They showed that until a ratio $N/R = 0.5$, where R is the RMS of the signal, and noise is uniform in $[-N, N]$, the FNN technique is able to definitively detect low-dimensional signals, and the residual percentage of false neighbors gives an indication of the noise level (see their figure 17).

In order to determine the appropriate *time lag* for the embedding, it is best to use a nonlinear measure, such as the Average Mutual Information (Shannon 1948). The Average Mutual Information (AMI) is defined as the information contained in $X(t+\tau)$ about $X(t)$ (Moon et al. 1995):

$$AMI(X(t);X(t+\tau)) = \sum_{X(t),X(t+\tau)} P(X(t),X(t+\tau))log\frac{P(X(t),X(t+\tau))}{P(X(t))P(X(t+\tau))} \quad (1)$$

The lag τ that corresponds to the first minimum of the AMI can be chosen as the optimal lag for the embedding. An additional, but linear, criterion is the first zero crossing of the autocorrelation function, which is usually at the same order of the first minimum of AMI (Abarbanel et al. 1993).

The rate at which the nearby trajectories separate (diverge) in the phase space is described by the *Lyapunov exponents*, introduced by Oseledec (1968). Lyapunov exponents are metric invariants, in that they are insensitive to initial conditions or small perturbations of an orbit in the phase space. For a system to possess chaos, positive Lyapunov exponents have to exist, and their sum is equal to the Kolmogorov-Sinai entropy (Pesin 1977). In a dynamical system with positive entropy $h(X)$, two points that are unresolvable at $t = 0$ will follow after some time τ separate trajectories. The possible number of separable trajectories generated by the system after time τ is measured by $2^{h(X)\tau}$ (Gallager 1968; Rabinovich 1978). For $\tau_P \approx h(X)^{-1}$, this number approaches the total number of trajectories available for the system, so that all knowledge of the evolution of a specific orbit is lost; i.e. predictability is lost after time τ_P, although statistical information about the system is retained (Abarbanel et al. 1993). The Kolmogorov-Sinai entropy $h(X)$ is approximately equal to the largest global Lyapunov exponent $\lambda_1(X)$, therefore $\tau_P \approx \lambda_1(X)^{-1}$. This limit has been considered as a measure of predictability in the context of weather forecasting, in the sense that no new information to the climatology is added by the forecast.

The average prediction error in the system at time t, $E(t)$ grows exponentially from the initial error $E(0)$ as follows:

$$E(t) = E(0) \cdot e^{\lambda_1 t} \quad (2)$$

While global Lyapunov exponents characterize the average predictability of the attractor, the local behavior of instabilities is also of interest, in particular in the neighborhood of stable and unstable manifolds, where predictability varies dramatically relative to the global behavior. Such changes in predictability are of great interest with regard to the regime-like behavior of ENSO since they speak to the character of the underlying manifold in phase space. Note that the evolution of an initial uncertainty need not be uniform even in a uniform linear system due to the role of non-orthogonal eigen-bases on uncertainty growth (Smith et al. 1999). *Local Lyapunov exponents* $\lambda(\mathbf{x}, L)$ measure the growth or decay over L time steps of a perturbation made around a specific point $\mathbf{x}$ of the attractor (Kennel et al. 1994).

As $L \to \infty$, the local LLE estimates tend toward asymptotic values corresponding to the global exponent, i.e. $\lambda(\mathbf{x}, L) \to \lambda$. Abarbanel et al. (1992) argue that the largest average LLE will approach the global exponent from above, i.e. that finite-time predictability will, on average, be worse than the global Lyapunov exponent indicates. The local Lyapunov exponents, and thus local predictability, may vary significantly on the attractor indicating times of enhanced or reduced predictability (Abarbanel et al. 1991; Legras and Ghil 1985; Nese 1989; Nese and Dutton 1989; Siqueira and Kirtman 2012).

In general, the methods used in nonlinear time series analysis are burdened by the finite size of the dataset, the presence of stochastic noise, and the fractal nature of the attractor (Bryant et al. 1990). The sensitivity of the LLE estimates to the size of the dataset is discussed in detail in Sects. 4 and 6, where we present results from sub-sampling of the dataset. Here, we treat the model output in a perfect-model and noise-free sense, applying no smoothing.

2.2 Interpretation for ENSO

We next apply these methods to 2,000 years of monthly-mean NINO3 SSTs, simulated by the GFDL CM2.1 coupled GCM with its solar irradiance, land cover and atmospheric composition held constant at 1860 values (Wittenberg 2009). Numerous studies have shown the CM2.1 model to produce a reasonable simulation of tropical climate and variability, as one of the best models in the CMIP3 intercomparison (Delworth et al. 2006; Gnanadesikan et al. 2006; Wittenberg et al. 2006; Capotondi et al. 2006; Kug et al. 2010). The model is also highly relevant to ENSO predictions, since it is used routinely at GFDL for ocean-atmosphere data assimilation and seasonal-to-decadal forecasts (Zhang et al. 2004, 2005, 2007; Sun et al. 2007; Yang et al. 2013). In addition, CM2.1 remains close to the state-of-the-art in climate modeling, as the parent of all of the new models developed at GFDL for the CMIP5 intercomparison (Donner et al. 2011; Griffies et al. 2011; Galbraith et al. 2011; Delworth et al. 2012; Dunne et al. 2012, 2013).

The GCM is a high-dimensional dynamical system. In principle, at any time it has a set of LLEs (i.e the eigenvalues of the linear tangent model at that point in phase space) that are a function of the model equations and the system state. These exponents are not calculated directly; rather, they are estimated from a long time series, which in our case is the model output (monthly NINO3 SSTa, upper-ocean heat content etc.) over a restricted domain. Hence, the LLEs reflect the dominant dynamics of processes in this restricted domain as they evolve in the global domain. In reconstructing the phase-space of the model we find that the optimal embedding dimension is $d = 5$ (see Sect. 3 for details): therefore, we are reducing the dimensionality of the system from $O(10^8)$ dimensions to five dimensions. Much of this reduction comes from considering only the ENSO subsystem in the tropical Pacific. A final reduction to five dimensions is consistent with Tziperman et al. (1994), who found that the Zebiak-Cane model, which has $O(10^5)$ variables and simulates only the ENSO subsystem, could be reduced to a dimension less than 8.

The LLEs describe the growth of infinitesimal perturbations due to internal variability. The local e-folding time of small perturbations in the system is shorter when the LLEs are large and the system less predictable. If one interprets the NINO3-derived LLEs in the context of ENSO being a weakly damped oscillator sustained by wind perturbations (e.g. Neelin et al. 1998; Kirtman and Schopf 1998; Thompson and Battisti 2000, 2001; Fedorov and Philander 2001), then the derived LLEs would describe the capacity of wind perturbations, which slightly alter the initial conditions, to grow and thus hinder predictability. As was noted in Fedorov (2002) and Fedorov et al. (2003), the initial conditions are important for the influence of the westerly wind bursts: during the initiation of a warm event a westerly wind burst can accelerate the development of the event, while one after the peak of El Niño will simply prolong its duration. Therefore, it could be reasonable to treat them as possible slight perturbations of the monthly NINO3 index, on which our LLE calculations are performed. Note that these wind perturbations need not be external to the system (Eisenman et al. 2005; Vecchi et al. 2006; Tziperman and Yu 2007; Gebbie et al. 2007). In any case, whether these perturbations be external weather noise, or modulated by the ENSO state, the SST-derived LLEs that we examine here could describe the initial rate of error growth following such perturbations, as long as the main assumption of the embedding theorem, that the single variable, NINO3 SST, is sufficient to reflect underlying

system variables, holds. It is our expectation that this is true for modes that grow to be the substantial errors that matter for ENSO forecasting, though it is unlikely to be true for all the fast-growing but short-lived perturbations that the GCM allows (as for example in Lengaigne et al. 2004).

3 Reconstruction of the phase space

Figure 1 shows the NINO3 SST anomalies in the 2,000-year GFDL CM2.1 pre-industrial simulation. Color coding indicates terciles of predictability as inferred by the LLEs: green periods have small LLEs and are the most predictable, orange periods are of intermediate predictability, and red periods are the least predictable. These terciles of predictability will be discussed in detail in Sect. 5 of the paper.

Applying the False Nearest Neighbor (FNN) method described by Kennel et al. (1992) to the unsmoothed monthly time series of NINO3 SSTa, we estimate the optimal global embedding dimension as $d = 5$. Both the Average Mutual Information and the autocorrelation function, suggest a time lag τ of 11 months for the embedding. Consequently, the 5-dimensional reconstructed phase space spans approximately 5 years (55 months). To avoid the influence of seasonality in our analyses, we exclude points within 12 months of a target point in the search for nearest neighbors. We thus avoid considering sequentially-adjacent, serially-correlated vectors as neighbors.

At each point of the reconstructed phase-space, six nearest neighbors are found, and the average rate of divergence in the phase space of their trajectories from the reference trajectory over a window of length L is calculated, where L is the time scale of integration (hereafter referred to as lead time L). The LLEs are approximated via the local Jacobian matrix derived from a Taylor expansion in the small deviations from the center of the neighborhood. Thus, the local Lyapunov spectrum consists of one exponent for each local dimension and for each lead time L. The local dimension must be equal to or less than the global dimension ($d = 5$). While we have chosen the local dimension to be three, we have also computed the LLEs with local dimension equal to the global dimension $d = 5$ and the results are not substantially different. The number of neighbors found for each reference point in the embedding is chosen to be twice the order of the polynomial fit used to form the local Jacobians (equal to 3) for numerical stability. The average distance in the 5-D phase space of the closest nearest neighbor from the reference

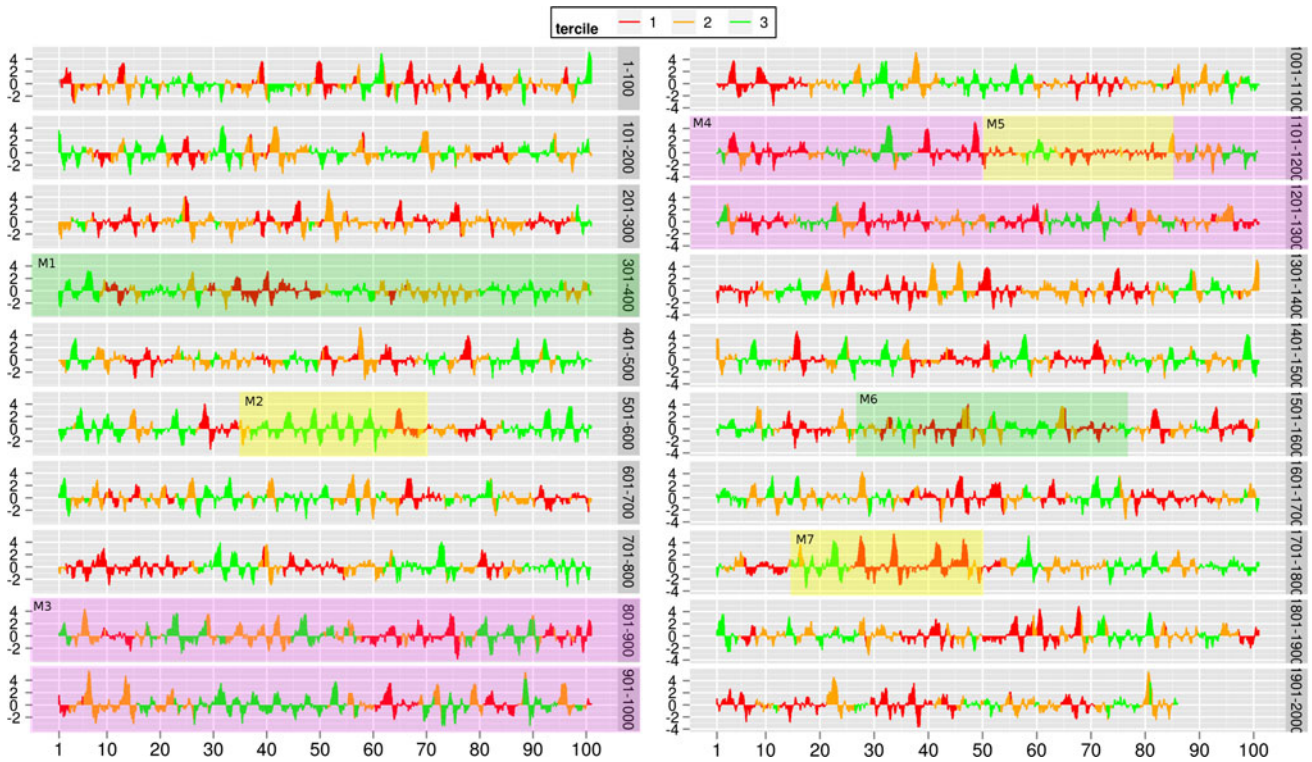


Fig. 1 The NINO3 Index times series in the 2,000-year unforced simulation of GFDL's CM2.1. *Colors* indicate terciles of predictability, as determined by the LLEs at lead time L = 4 months, with *red* being the least predictable tercile, *orange* intermediate, and *green* the most predictable. Predictability decreases by about 9 % on average from tercile to tercile. The *shaded regions* indicate epochs of distinct ENSO variability, as per Wittenberg (2009). The lack of LLE data at the end of the time series, is due to the integration for the calculation of LLEs being done forward in time

point is 0.75 °C, while the average difference in NINO3 SSTa between the closest neighbor and the reference point is 0.15 °C.

Figure 2 shows an example of events and their nearest neighbors at a lead time $L = 4$ months. In Fig. 2a, we show the NINO3 SSTa of the nearest neighbors found in phase-space for the event of year 1722 (in thick black line). Since the lead time we use is L = 4 months, the neighbors are identified by their proximity in the reconstructed-phase space 4 months before the peak of the event, i.e. by the proximity to the reference point $NINO3(t - L, t - L + \tau, t - L + 2\tau, t - L + 3\tau, t - L + 4\tau, t - L + 5\tau)$, where t is peak month, L is the lead time, and τ is the embedding lag (11 months). This proximity is also reflected in the actual timeline as seen in the figure (note the convergence of trajectories at months 4, 7, 18). As mentioned above, in the search for nearest neighbors we exclude points in the phase space that are within 12 months from the reference point; thus, we avoid selecting the same—but slightly lagged—event as a neighbor of the reference event. The LLE at the peak of the event characterizes, within the reconstructed phase-space, the divergence of the trajectories of the neighbors from their reference trajectory in the 4-month window prior its peak. In this case, the LLE is small, i.e. the event is within our 'green tercile' of more-predictable events (see Sect. 5). The trajectories in phase-space do not diverge very fast in this 4-month window, which is reflected in the actual timeline as well. In contrast, Fig. 2b shows the evolution of the NINO3 SSTa for the nearest neighbors of the event of year 1733, which is characterized by its LLE as a lower predictability event (it belongs to our 'red tercile'). In the 4 months prior to the peak, the divergence of trajectories in the phase-space is higher, as is also reflected in the evolution of NINO3 SSTa in time.

For the main analyses presented in Sects. 5, 6, and 7 we chose a lead time of $L = 4$ months. It appears to be a generic property of dynamical systems, including low dimensional examples such as the Lorenz equations or the Ikeda map, to have rapid initial growth in the rate of divergence (i.e. the size of the LLE), which then slows with time and ultimately approaches an asymptotic limit set by the size of the entire attractor (e.g. Kennel et al. 1994). Very short lead times L put us in the rapid growth regime so the $LLE(x, L)$ overestimates the dispersion growth rate relevant to predictability. At long lead times the exponential assumption (Eq. 2) made in defining the LLE no longer holds, and the estimation procedure is no longer valid. Hence, we choose an intermediate lead time L = 4 months. This choice was also supported by the comparison between the LLE-derived error growth rates and those derived from a set of re-forecast experiments using CM2.1 (Sect. 4).

4 LLEs and actual predictability

In this section we explore the relationship of the LLEs calculated via reconstruction of the phase-space by time-delaying of the monthly NINO3 SSTa with the predictability estimates from a series of re-forecasts experiments performed with CM2.1.

The CM2.1 re-forecast experiments are described in detail in *Wittenberg, in prep*. Briefly, 11 different 1-January initial conditions were selected from the CM2.1 control run, each at least a decade apart and sampling epochs with a diverse range of ENSO amplitudes and ENSO regularity/irregularity. For each initial condition, the model trajectory is perturbed by adding an effectively infinitesimal (order of 10^{-4} °C) increment to the ocean surface temperature at a single gridpoint in the central equatorial Pacific. Forty such perturbations are made for each initial condition. These forty ensemble members are then integrated forward in time for 10 years, using the same CM2.1 model that generated the control run itself. The ensemble root-mean-square (RMS) difference of the monthly-mean NINO3-averaged SST from the ensemble mean is then computed at each re-forecast lead time, as a measure of the ensemble spread. Denoting this RMS difference by D(t), we then compute the ratio D(t)/D(0) as a measure of the dispersion

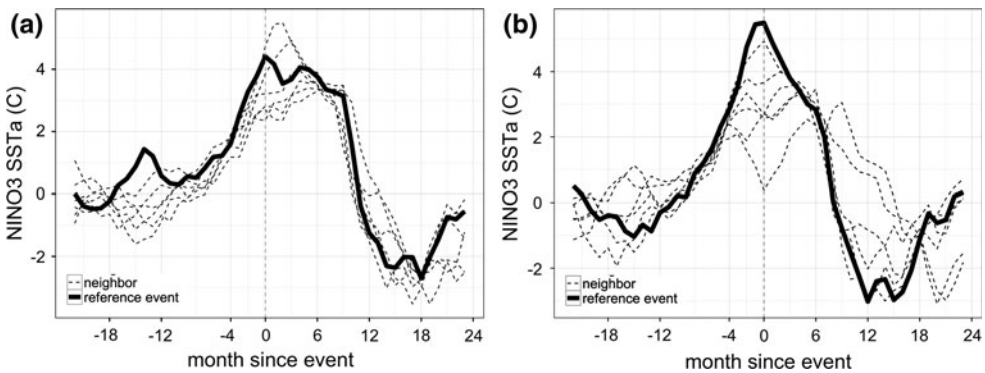


Fig. 2 **a** An event characterized as more predictable (*green* in Fig. 1, year 1722) and its nearest neighbors. **b** A least predictable event (*red*, year 1733) and its nearest neighbors. The nearest neighbors are found based on their distance from the reference event in the phase space, 4 months before its peak (see Sect. 3)

rate of the ensemble, where t = 0 corresponds to the central (16-January) time of the first monthly-mean.

To compute the e-folding time, we consider an exponential growth model, where:

$$D(t) = D(0) \cdot e^{\frac{t}{\tau_e}} \quad (3)$$

where $D(t)$ is RMS ensemble spread at time t, and τ_e is the e-folding time. It follows that the error-growth rate, or Re-forecast LLE Equivalent (RLE) at lead time t is:

$$RLE(t) = \frac{1}{\tau_e} = \frac{ln\frac{D(t)}{D(0)}}{t} \quad (4)$$

Figure 3a shows boxplots of LLEs versus lead time; as the lead time increases, the LLE estimates, and therefore the error-growth rate, decreases. Saturation occurs after approximately 64 months. At that time, the local exponents asymptotically approach the global estimate. Figure 3b shows the RLEs, defined as in Eq. 4; as the ensemble disperses, the growth rate slows. Saturation in the re-forecasts occurs at approximately 64 months also. At short lead times the spread among experiments is large.

Figure 4a presents the LLE-derived growth rate (LLE) versus the one calculated from the 11 re-forecast experiments (RLE). Colors indicate the forecast lead time, or scale of LLE integration, and the gray line denotes a one-to-one relationship. Over short lead times, i.e. 4 months and less, the LLEs underestimate the error growth rates, i.e.

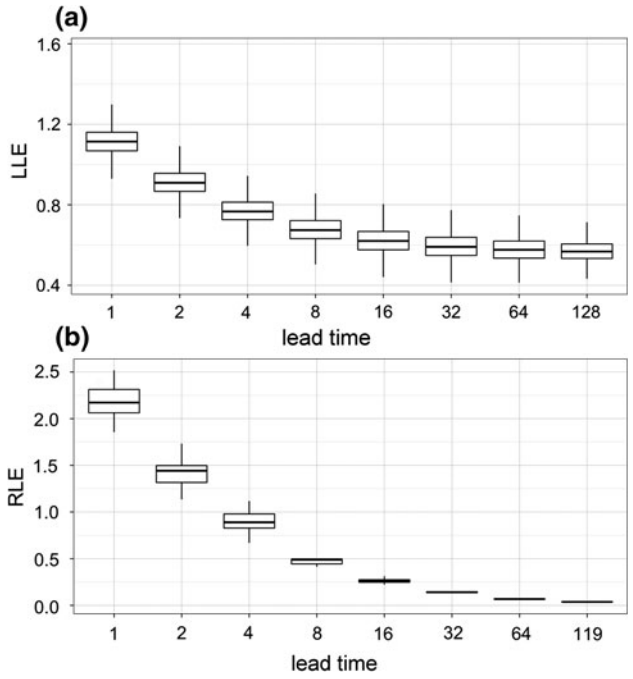


Fig. 3 **a** Boxplots of **a** LLEs versus lead time; saturation occurs after 64 months. **b** Re-forecast LLE equivalents (RLEs) versus lead time, calculated from 11 re-forecast experiments. The RLEs are defined by Eq. 4

overestimate the e-folding time τ_e. On the contrary, over longer lead times, the LLEs overestimate the true rates, i.e. underestimate τ_e. Interestingly, the relationship between LLEs and RLEs is a linear function across lead times:

$$LLE(t) = 0.56 + 0.25 \cdot RLE(t) \quad (5)$$

4.1 Short lead times

One possible explanation for the underestimation of error dispersion over short lead times is that the LLE estimation is based on neighbors in the phase space that are farther apart than in the case of the re-forecasts, where the initial perturbation is truly infinitesimal. The exponential growth model for errors is only a linear tangent approximation, and if the initial separation is large this approximation is compromised by the difference-based derivative approximations. This hypothesis would imply that, given a long-enough CM2.1 simulation, the neighbors based on which the LLE approximation is done would be closer because they are drawn from a larger sample. Hence the approximation to true LLEs would be better. Indeed, when the LLE calculation is performed on half the time series, for short lead times the approximation is worse, as seen in the diamond-shaped points in Fig. 4.

To further test this hypothesis, we consider the average difference in NINO3 SSTa between the closest neighbor and the reference point, which was found to be 0.15 °C (see Sect. 3). Figure 5 (left panel) shows the growth of the re-forecast spread versus lead time in the first 10 months of the re-forecast. As was also seen in Fig. 3, the growth rate slows as the ensemble disperses. The 0.15 °C RMS of the nearest-neighbor distance is analogous to the spread of the second monthly-mean of the re-forecasts, as shown by the red line in Fig. 5. The subsequent growth rate for this spread, evaluated via a forward difference between the second and third-month mean, is approximately one e-folding per month, as seen in the right panel of Fig. 5.

In order to test whether the initial distance of the nearest neighbors is the culprit for the discrepancy between LLE estimates and RLEs, we replace in Eq. 4 the RMS spread of the first monthly-mean $D(0)$ with that of the second monthly-mean, $D(1)$. Figure 4b shows the LLEs versus RLE for the second monthly-mean growth rate. The LLEs now overestimate the second monthly-mean growth rate at all lead times. This could be expected, since the new estimate of actual error growth is aligned in the direction of the modes that grow fastest between $t = 0$ and $t = 1$ (and were originally underestimated by the LLEs). These modes do not grow as fast from $t = 1$ to the next time steps as the ones that grew from random seeds at $t = 0$. But, the LLE estimate still assumes a random seed, so it should now be larger than the RLE of the second montly-mean spread.

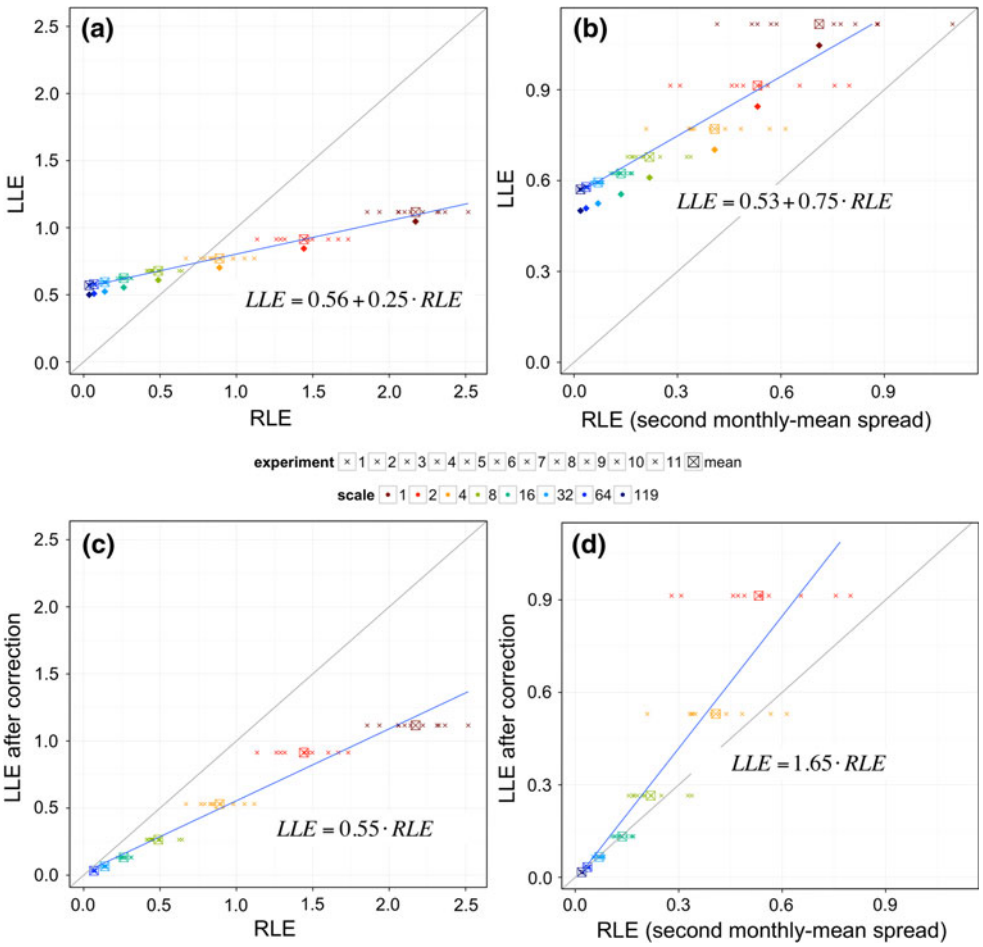


Fig. 4 LLE- versus RLE-derived growth rate $1/\tau_e$ **a** before any correction, **b** after correcting for the initial distance of nearest neighbors (Sect. 4.1), **c** after introducing a threshold for error growth at long lead times (Sect. 4.2), and **d** after both corrections **b** and **c** are applied. *Colors* indicate the forecast lead time, and the *gray line* denotes a one-to-one relationship. The *blue line* is the best fit from a linear regression model (the equation is shown). *Diamond-shaped points* denote the average values for calculations performed on the first half and second half of the time series

The relation between LLEs and the RLEs of the second montly-mean spread is now:

$$LLE(t) = 0.53 + 0.75RLE(t) \quad (6)$$

In conclusion, the underestimation of error-growth by the LLEs for short lead times is attributed to the fact that in a finite time series the nearest neighbors based on which the error-growth is calculated are not close enough for the approximation to capture the more rapid initial growth characteristic of dynamical systems.

4.2 Long lead times

If the nearest-neighbor growth rate of approximately 'one e-folding per month' were to continue, the spread would saturate at the climatological value in just over 3 months (green line in Fig. 5). But for the actual re-forecasts, this growth rate is not sustained, and it takes more than 7 months to first reach that climatological level. As the scale increases, the actual growth rate slows down (also see Fig. 3b); the RLE and the LLE-approximated growth rates approach each other at lead time $L = 4$, as seen in Fig. 4a. For lead times over 8 months, the LLEs greatly overestimate error growth. This could be a consequence of the saturation of error growth near the size of the attractor (Kennel et al. 1994; Boffetta et al. 1998), and of the assumption that error growth is exponential until a finite limit (i.e. the global LE). It is clear, however, that exponential growth of errors does not hold for long lead times, and that finite-size perturbations and prediction errors cannot grow indefinitely. If the actual error has stopped growing and the LLE still assumes exponential growth, then the discrepancy in growth rates will increase linearly with lead time L, as found in Fig. 4a.

To constrain error growth at long lead times, Kennel et al. (1994) introduce a limiting factor $\rho = lnR$, where R is the ratio of the geometric mean over uncorrelated pairs of attractor points to the geometric mean of the initial perturbation magnitude. To introduce an analogous correction, we consider the limiting factor to be the ratio of the final RMS (standard deviation of NINO3 for the 2,000-year simulation, i.e. 1.25 °C) to the initial RMS, i.e. the initial distance of the neighbors (0.15 °C). We constrain the LLE estimate as follows:

$$LLE'(x,L) = \min\left(LLE(x,L), \frac{\rho}{L}\right), \quad \text{where } \rho = ln\left(\frac{sd(NINO3)}{D(0)}\right) \tag{7}$$

Figure 4c shows the LLE versus RLEs after the correction of Eq. 7. The estimates are now very close to the one-to-one line, especially at long lead times. Note that this correction has no effect at short lead times. Figure 4d shows the same LLE's as Fig. 4c, but as in Fig. 4b the re-forecast equivalents are calculated using the second monthly-mean spread as an initial value.

To conclude, error growth is characterized by an initial rapid growth, followed by a regime of exponential growth, leading to a saturation for long lead times. The LLE estimates computed here cannot capture the initial rapid growth possibly because they are approximated based on an initial distance of neighbors that is much larger than infinitesimal and is analogous to the spread of the second monthly-mean of actual re-forecasts. For the longer lead time of 4 months, where the exponential error growth is more relevant, the LLE estimate and actual growth rate are very close to each other. For lead times beyond 8 months, the LLEs greatly overestimate the error growth, since they cannot distinguish between the fast modes and the slow modes that extent predictability beyond these lead times (e.g. as in Goswami and Shukla 1991; Blumenthal 1991; Goswami et al. 1991).

For the remainder of the paper, we choose to focus on LLEs at the intermediate scale of 4 months.

5 Variability in predictability in the GFDL CM2.1-1860 simulation

At a lead time of $L = 4$ months, we calculate two positive and one negative LLE with [mean, standard deviation] equal to [0.77, 0.068], [0.233, 0.068], and [−0.728, 0.109]. To test whether the calculated LLEs as significantly different from the ones calculated from an AR(2) process, we fitted an AR(2) model to the 2,000-year time series, generated samples from this process, and calculated their LLEs. A Kolmogorov-Smirnov test confirms that the LLEs calculated from CM2.1 could not have come from such an autoregressive process at the 95 % significance level.

In order to quantify variability in predictability over the long simulation, we define terciles of predictability based on the 33rd and 67th percentile of the largest positive LLE. The mean and standard deviation of the LLEs for each level is equal to [0.85, 0.048], [0.77, 0.015] and [0.70, 0.032], respectively. On average, predictability increases by approximately 9 % from one tercile to the next lower one.

The variations in predictability in the 2,000-year run are shown in Fig. 1. Red color (first tercile) indicates periods

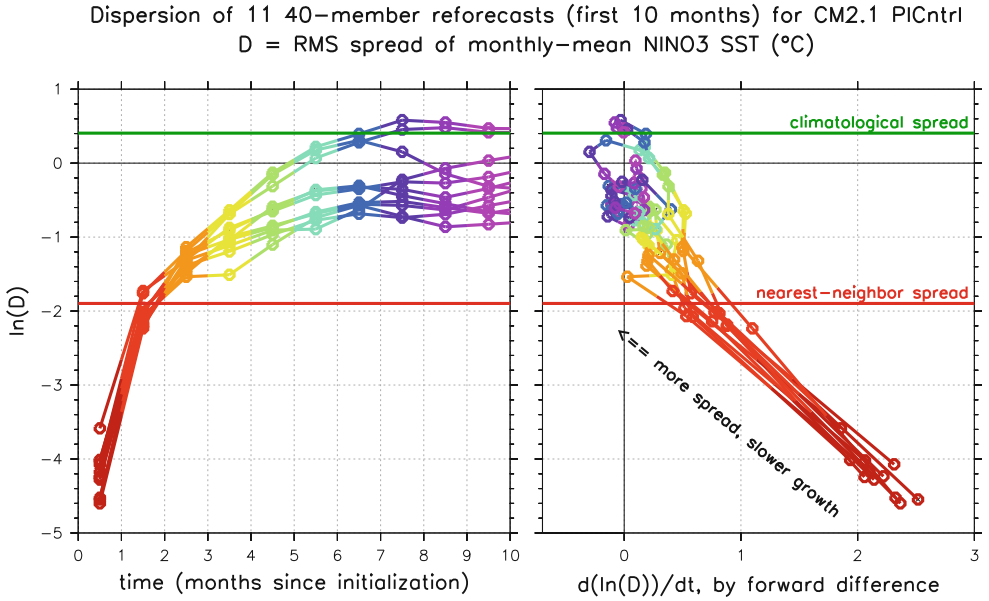
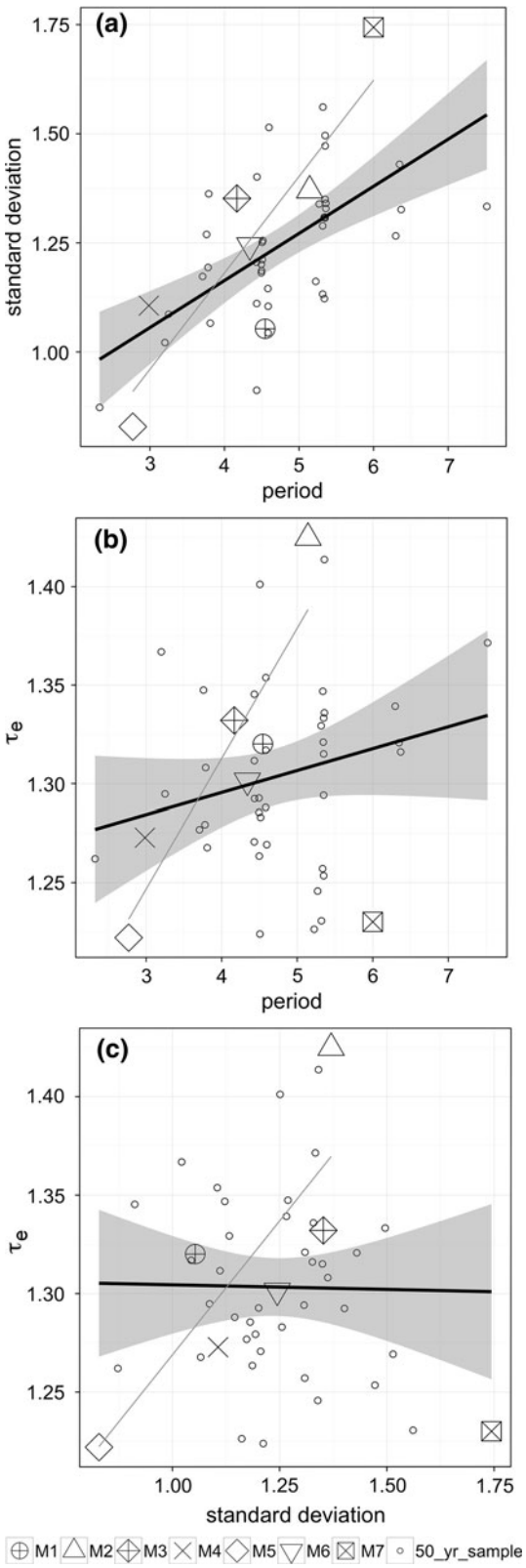


Fig. 5 Growth of the monthly-mean re-forecast spread versus lead time for the first 10 months in 11 re-forecast experiments (*left panel*). The *right panel* shares the same ordinate but shows the growth rate on the abscissa, which is calculated via a forward difference between month t and month $t + 1$. The *red line* indicates the logarithm of the initial distance of nearest neighbors for the LLE calculation (0.15 °C). The *green line* shows the climatological spread of randomly selected NINO3 SSTa monthly means (1.5 °C)



◀ **Fig. 6** Scatterplots of ENSO standard deviation (in degrees C), period (in years), and the e-folding time (in months) for each epoch of distinct ENSO behavior. Open circles indicate consecutive 50-year periods. The *shaded area* indicates the 95 % confidence intervals for the linear regression fit. The relationship is statistically significant at the 5 % level only for standard deviation versus period (**a**). *Gray lines* indicate the best linear regression fit for epochs M1–M6. The e-folding time is calculated from LLEs at L = 4 months

with the least predictability; orange periods (second tercile) are intermediate, and green periods (third tercile) are in the most predictable tercile. The difference in predictability between the most (green) and least (red) periods is approximately 18 %.

Wittenberg (2009) noted distinct periods of ENSO behavior, shown in the shaded regions in Fig. 1. Some of these epochs correspond to distinct periods of ENSO predictability. Epoch M1, whose ENSO variability mimics pre-1960 observations is marked by decadal variations in predictability. The decades with stronger variability and persistence of events, such as years 330–350 belong to the first tercile (red), indicating decreased predictability. Epoch M6 agrees well with the post-1960 observed variability, characterized by weak, biennial oscillations, followed by a large warm event, then several smaller events, another large warm event, and then a long quiet period. This irregularity seems reflected in predictability, as shown in Fig. 1.

On the other hand, epoch M2 with moderate ENSO events which exhibit regular periodicity, is marked as a period with constantly enhanced predictability. Epoch M3, a period of consistently strong variability, has long periods of enhanced predictability. The quiet epoch M5 is characterized by more La Niña events; in this epoch, predictability is decreased, as indicated by the prominence of red periods. Epoch M7 with strong warm events is also classified as less predictable by the LLEs, while the irregularity of ENSO in epoch M6 is accompanied by irregularity in predictability.

Is there correspondence between the decadal variability in ENSO magnitude and frequency and the decadal variability in ENSO predictability? Figure 6 shows scatterplots of NINO3 standard deviation, period and mean e-folding time for epochs M1 to M7, and for consecutive 50-year periods (open circles). The mean period for each 50-year period is defined from the peak of the wavelet power spectrum. In this model, higher variance is associated with larger period, i.e. the stronger the events the longer it takes for the Pacific to 'recharge' to give another event so the longer the wait-time between events (Fig. 6a). The relationships of the e-folding time with period and standard deviation are not statistically significant (see best linear fit line in Fig. 6b, c). Only if one excludes the consecutive 50-year periods and the mega-ENSO period M7 is there a correlation of e-folding time with period and standard deviation (see gray best linear fit lines for M1–M6 in Fig. 6). Note that the 'inactive' period M4 has shorter

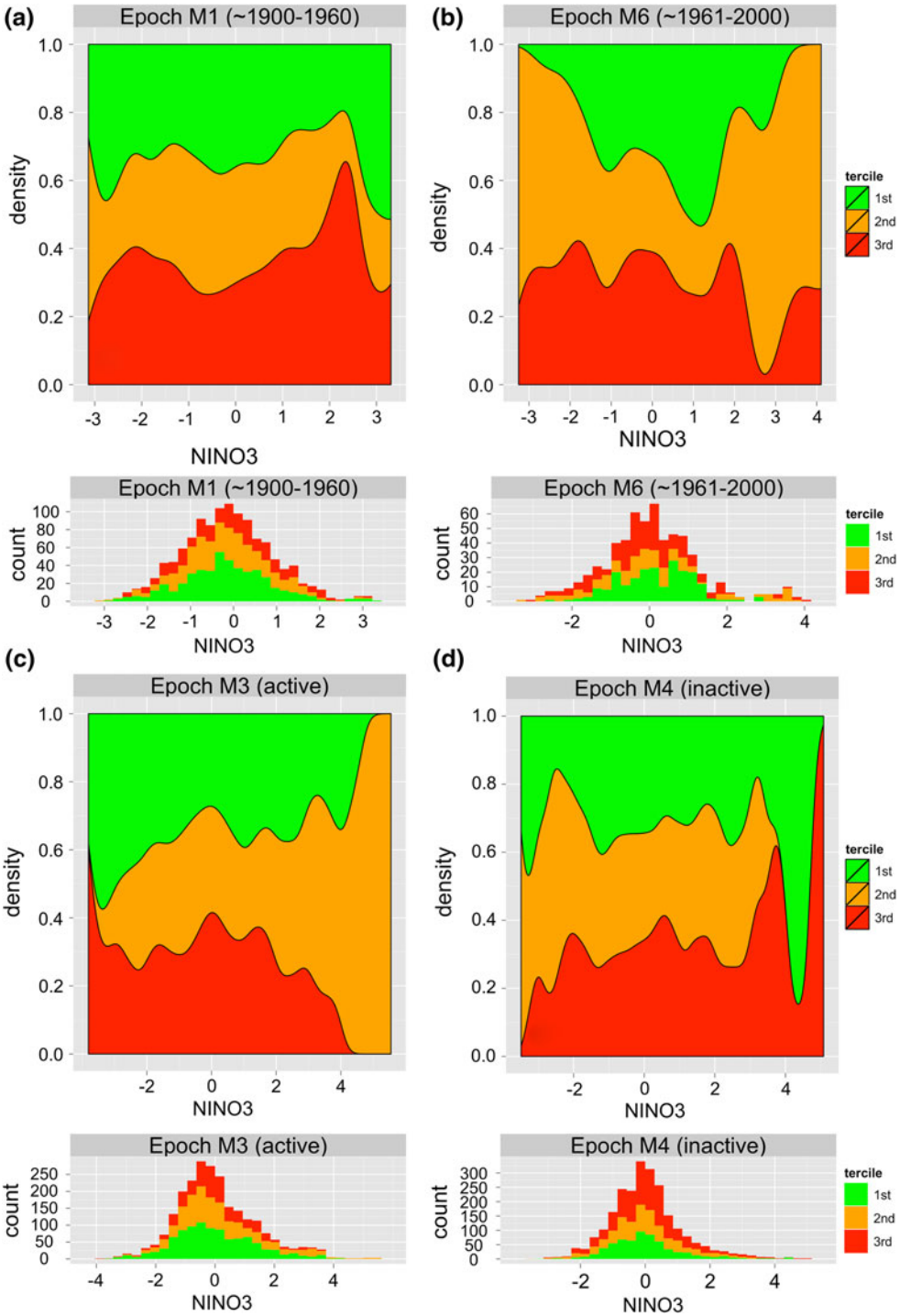


Fig. 7 The probability that an event of certain magnitude be associated with a certain tercile of predictability, in epoch **a** M1, which mimics the pre-1960 observations, **b** M6, which agrees well with the post-1960 observations, **c** 'active' period M3, and **d** 'inactive' period M4. To show the number of events that belong to each tercile, the *bottom panel* of *each subfigure* shows stacked histograms of the events. Strong El Niño events are more predictable in epoch M6. El Niño predictability is enhanced during the 'active' period M3 compared to the 'inactive' one (M4). Terciles of predictability are based on LLEs at L = 4 months, as in Fig. 1

e-folding time than the 'active' period M3; the latter is characterized by longer periodicity and higher variance. Epoch M2, which is the most predictable has the longer period and highest variance.

Is there is a relationship between the magnitude of individual events and their classification in terms of predictability? There seems to be no such relationship over the whole 2,000-year run. However, such relationships arise when one looks within the epochs whose spectral characteristics resemble observed ENSO periods. Figure 7 shows the probability density function that an event of certain magnitude be associated with each of the three levels of predictability. In epoch M1, which mimics the pre-1960 observed record, strong warm events are deemed less predictable, in contrast to strong cold events (Fig. 7a). For example, strong warm events of magnitude 2.5 °C have a

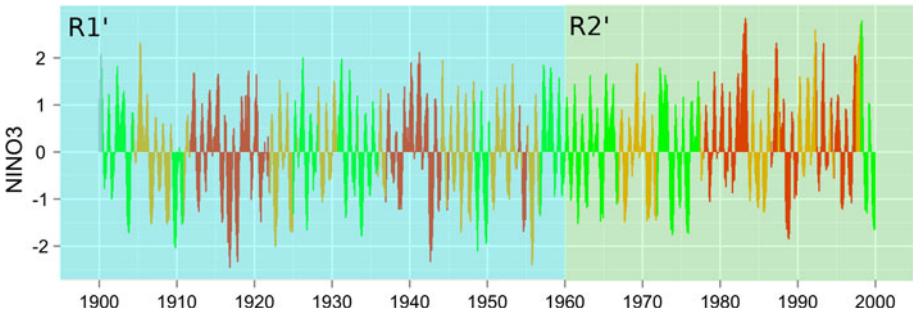


Fig. 8 The NINO3 Index from the ERSST.v3 dataset. *Colors* indicate terciles of predictability based on LLEs at L = 4 months, with *red* being the least predictable, *orange* the immediately more predictable (by 27 %), and *green* the most predictable (by 44 % compared to *red*). The *shaded regions* approximate epochs R1 and R2 (pre- and post-1960, respectively), as per Wittenberg (2009)

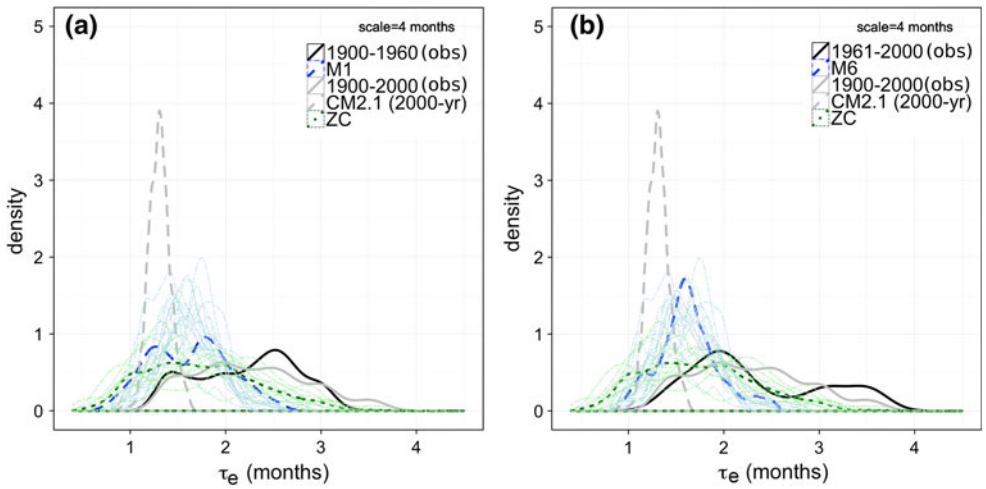


Fig. 9 The pdf's of the mean e-folding time for **a** pre-1960 observations and their corresponding model period M1, and **b** post-1960 observations and the corresponding epoch M6. In general, the model seems less predictable than nature. The *gray lines* refer to the full 2,000-year long CM2.1 simulation (*dashed*) and the full 100-year record (*solid*). The *dark green dotted line* refers to randomly selected 100-year samples from an unforced 150,000-year long ZC simulation. *Thin lines* refer to consecutive 100-year periods from the GCM (*blue*) and the ZC model (*green*)

63 % probability of being less predictable (level 1). Strong cold events of magnitude −2 have 40 % probability of belonging to level 1, and 60 % probability of belonging to one of the other two categories of enhanced predictability. The picture is quite different during active epoch M6, which agrees with the post-1960 observed ENSO variability, as seen in Fig. 7b. This epoch has stronger warm ENSO events (note the abscissa). The majority of the 2.5 °C events here are deemed more predictable (levels 2 and 3). This result suggests that caution should be exercised when one infers relationships between predictability and magnitude of ENSO events based on the short record: conclusions are epoch-dependent, and no such relationships would be inferred from our examination of the full 2,000-year long simulation.

Finally, we compare one active and one inactive period. During active epoch M3 very strong warm events are more predictable in contrast to strong cold events (Fig. 7c). During the inactive period M4 strong warm events are less predictable, and, on average, the whole inactive period M4 is less predictable than the active period M3 (compare Fig. 7c, d). The same conclusion can be drawn from Fig. 6. This result is consistent with the finding of Kirtman and Schopf (1998) that forecast skill is higher in periods of high amplitude interannual variability.

6 Predictability in CM2.1 versus observations

In this section, we compare the LLE characteristics in epochs M1 and M6 to the LLEs computed from the observed record.

We compute the LLEs at lead time $L = 4$ months from the NOAA Extended Reconstructed SST ERSST.v3 record; Fig. 8 shows the NINO3 time series, with colors indicating terciles of predictability, as in previous sections.

The shaded regions R1′ and R2′ are the two periods discussed in Wittenberg (2009), corresponding qualitatively to model epochs M1 and M6, respectively.

Figure 9 shows the probability density function of the e-folding time for (a) epoch M1 and its corresponding period 1900–1960 in the observed record, and (b) epoch M6 and its corresponding period 1961–2000. Note that for a consistent comparison with the 100-year long observational record, the LLEs are computed here using 100-year periods (M1 and the period 1501–1600 containing epoch M6). The gray dashed line shows the probability density function estimated for the full 2,000-year long run, while the gray solid line shows the pdf estimated from the 100-year long observational record. The green dotted line delineates the pdf of the e-folding time computed from randomly selected 100-year long periods from an unforced 150,000-year long simulation from the Zebiak–Cane (ZC) model (Zebiak and Cane 1987).

In general, the mean e-folding time in the model is smaller than in the observed record, indicating that the rate of information loss in the model is faster than in nature. The median e-folding time for epochs M1 and M6 is 1.67 and 1.61 months, respectively, compared to 2.3 and 2.1 months in the pre- and post-1960 observed periods. The lower variance in the model pdf, especially in epoch M6, indicates less variability in predictability compared to observations. The model also has on average less predictability in the more active epoch M6 (Fig. 9b). The probability mass of the e-folding time from observations is shifted to lower values in the active post-1960 period (the median value shifts from 2.3 to 2.1), and median predictability is reduced by 8.5 %. However, the tails of the distribution are heavier and extend up to 4 months in the active 1961–2000 period. In CM2.1, the shift in median values is of the order of 4 %. The e-folding time in the ZC model is similar to that of CM2.1 (with median value 1.7), but has larger variance and heavier tails in better accordance with the observed record.

It could be argued from Fig. 9 that real-world predictability is an upper bound for model predictability, since the probability mass of the e-folding time from observations lies above that of the model. In an informal poll we found that most people expected nature to be less predictable than any model, contrary to our finding here. Some reflection may persuade the reader that there is no necessary ordering. For example, the pdf of e-folding time for the ZC model has heavier tails, so its predictability could rank higher or lower than CM2.1 and observations, depending on the region of the attractor (or the time of initialization of a re-forecast). In addition, changing a parameter (e.g. to increase the strength of ocean-atmosphere coupling and introduce more noise in the model) could make it less predictable and could result in its pdf being more similar to that of CM2.1. Our findings are consistent with the results of Schneider et al. (2003), who found that tropical SST anomalies in coupled models tend to be less persistent than in observations, and argued that initial shocks in the coupled models can be transmitted rapidly to remote locations by oceanic wave propagation and lead to degradation of the forecasts through coupled interactions. Finally, the discrepancy between between real-world and GCM predictability could be arising from the external radiative forcings present in the historical reconstruction, which could lend apparent predictability to the record.

The reduction in variance between the 2,000-year sample and the 100-year sample, seen in Fig. 9, gives a sense of the uncertainty due to sampling when considering only a short record period or model run. This suggests that phase-space reconstruction based on only a century of data may not suffice to capture the system's dynamics; this concern is similar to the concern of Wittenberg (2009) about whether the available ENSO record suffices to constrain ENSO simulations. As we showed in Sect. 4, the finite nature of the time series used for the LLE estimation has an important effect on the accuracy of the derived error growth rates, especially at short lead times.

7 Relationship of NINO3 predictability and underlying ENSO variables

As explained in Sect. 2, the fundamental idea of phase space reconstruction by time-delay embedding of a single variable is that the delay coordinates constituting the embedding contain information about state variables that are not explicitly sampled, and are thus able to capture the dynamics of the underlying high-dimensional system. The LLEs measure the rate of divergence of nearby trajectories in this reconstructed phase space. Typically, the number of delay coordinates needed is less than or equal to $2d + 1$, where d is the "true" dimension of the underlying phase space if all relevant variables were sampled and available to form the phase space. It is of interest, then, to examine the capacity of the phase-space reconstruction based solely on the NINO3 Index to capture underlying ENSO dynamics.

Subsequent analyses are based on the phase space reconstruction from the first 500 years of simulations, the period for which we had access to upper-ocean heat content data. We composite the 120-month Hovmöller diagrams of SST and upper-ocean heat content anomalies (0–300 m, 10°S–10°N) for the warm events classified as more (green) and less (red) predictable. An event is considered when the NINO3 SSTa exceeds 1 °C, and the classification in green and red terciles is based on the LLE estimates at lead time 4 months (shown in Fig. 1). Figure 10 shows the

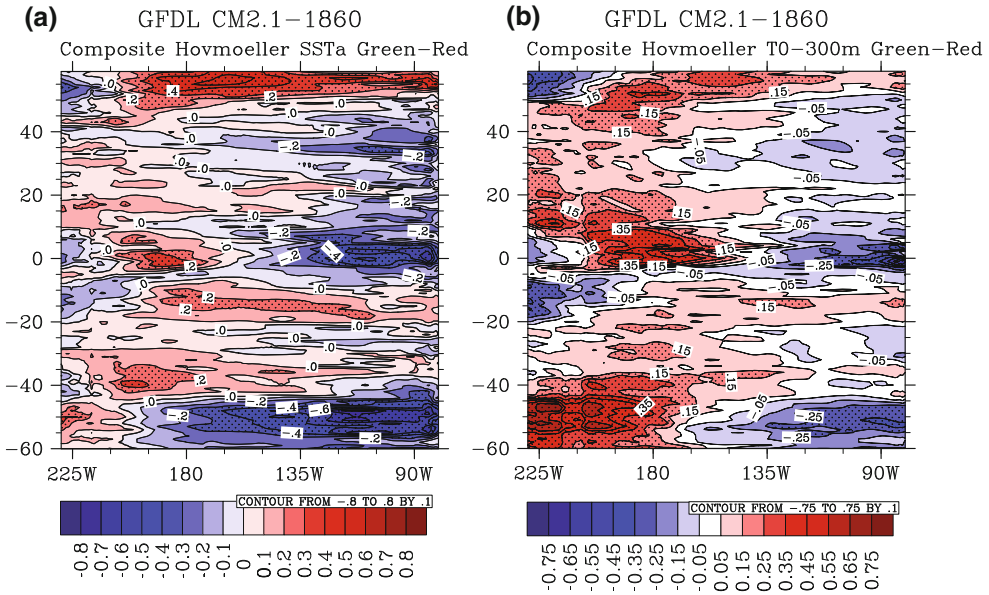


Fig. 10 The difference of the composite 120-month Hovmöller diagrams of SST and upper ocean heat content anomalies for the events classified as more (*green*) or less (*red*) predictable. *Stippled areas* denote statistical significance based on bootstrapping. The SST difference plot is also field-significant at the 75 % level, while the heat-content difference plot is field-significant at the 97 % level. Note the heat pile-up in the west Pacific five years prior to the event, and the associated cold anomalies in the east Pacific

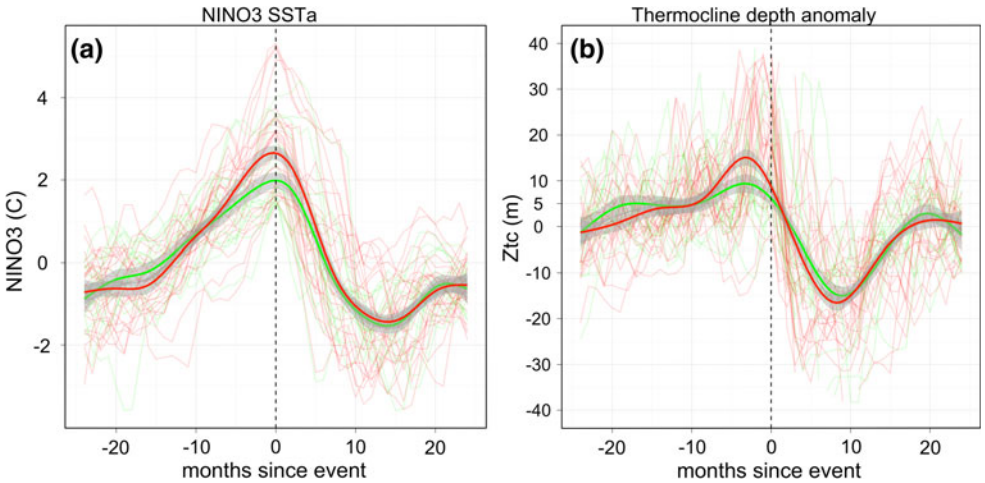


Fig. 11 **a** SST and **b** thermocline depth anomaly within 24 months of the peak of the *green* (most predictable) and *red* (least predicable) events. *Light green* and *red lines* show individual events, while *thick lines* are their composite

difference of the composite diagrams (green minus red). Stippled areas denote statistical significance at the 95 % level based on bootstrapping. The SST difference plot is also field-significant at the 75 % level, while the heat-content difference plot is field-significant at the 97 % level.[1] The most notable difference is the heat pile-up in the west Pacific five years prior to the event, and the associated cold anomalies in the east Pacific. The least predictable events are stronger, which accounts for the larger SST anomalies in the central Pacific at the peak of the event (t = 0), so a claim cannot be made on the basis of these results that central Pacific El Niños are more predictable. Also, the predictability classification is done here on the basis of the NINO3 Index; Kim et al. (2009) show that the NINO4 Index is more predictable than the NINO3 Index and conclude that central Pacific El Niños are more predictable than eastern Pacific ones.

Figure 11 shows the SST and thermocline depth anomaly within 48 months of the peak of the events. Light green and red lines show individual events, while thick lines are their composite. While the onset of the SST

[1] In order to test field-significance we used bootstrapping to create 1,000 pseudo-difference maps; these maps exhibit contiguous areas of significant differences, due to spatio-temporal correlations between gridpoints. The percentage of grid-points that has significant values in the original map is at the 75th and 97th percentile of the percentage in the 1,000 pseudo-difference maps for SST and heat content, respectively.

anomalies is almost simultaneous for both groups, the thermocline starts deepening approximately 4 months earlier in the most predictable events. Thus, the LLEs seem to be reflecting a reduced uncertainty regarding the onset of an event given a thermocline that starts depending early on. The difference in thermocline between the most and least predictable events might also be reflecting a succession of wind events, that could alter subsequent predictability. The amplitude of the earlier thermocline anomaly in the most predictable events -almost five meters- can be captured by satellites. The results hint towards the notion that the thermocline depth contains more information than the SST anomaly, i.e. that the slower variable of the system has more predictability than the faster one (also see Lima et al. 2009).

8 Summary and discussion

The presence of rich ENSO variability in the long unforced simulation of GFDL's CM2.1 motivated the use of tools from the ergodic theory of dynamical systems to study variability in ENSO predictability, and its possible connections to ENSO magnitude, frequency, and physical evolution at the epochal and individual-event levels. To that end, this paper uses local Lyapunov exponents computed from time-delaying of the monthly NINO3 SSTa model output. The local Lyapunov exponents computed here describe the growth of perturbations due to internal variability; as such they are a measure of the predictive uncertainty at any given point along the system trajectory.

We first studied the relevance of LLE-derived local predictability estimates to predictability estimates derived from a set of re-forecast experiments. To the authors' knowledge this paper is the first to discuss such relationships using long simulations and an extensive set of re-forecasts from a fully coupled GCM. The comparison of LLE-derived predictability to error growth from the set of re-forecast experiments yielded the following main results:

1. The LLEs underestimate the error growth rates for short forecast lead times (less than 8 months), while they overestimate it for longer lead times.
2. The underestimation of error growth at short lead times is attributed to the fact that in a finite time series the nearest neighbors in the phase space, based on which the error-growth is calculated, are not close enough for the approximation to be optimal.
3. Longer model simulations would provide a larger sample from which to find nearest neighbors. A better set of nearest neighbors should reduce the departure of LLE-derived predictability from 'actual predictability' (e.g. as derived from re-forecasts).
4. The overestimation of error-growth at long lead times is attributed to the fact that finite sized perturbations cannot grow exponentially and indefinitely. Adding a threshold for error growth brings the LLE-derived estimate closer to the actual error growth rates.
5. The departure of LLE-derived error growth rates (proportional to $1/\tau_e$) from the actual re-forecast rates is a linear function of forecast lead time.
6. The growth-rate saturation derived from LLEs occurs at approximately 5 years, which is broadly consistent with the re-forecast results.

As shown above, the relationship between LLE-derived predictability and 'actual predictability' as inferred from the re-forecast experiments is lead-time dependent, and sensitive to the length of simulation. Measures like finite-size (Boffetta et al. 1998) or scale-dependent (Gao et al. 2007) Lyapunov exponents could be better suited to characterize the multi-scale nature of ENSO predictability. Nevertheless, our study of the relationship between LLE estimates and predictability from re-forecast experiments in a high-complexity GCM contributes to the investigation whether exponential growth of errors (as in Eq. 2) is relevant for quantifying error scaling at both small and substantially larger levels of error in these systems (see Kennel et al. 1994; Boffetta et al. 1998; Smith et al. 1999, for discussion of these issues).

We found good agreement between the LLE-derived and the actual error growth at lead times near 4 months. This is beyond the lead times dominated by initial rapid error growth and before the exponential growth assumption underlying the LLE construct becomes inapplicable. Thus, in the remainder of the paper, we focused on the 4-month LLEs to characterize periods of increased or decreased seasonal predictability in the long CM2.1 simulation.

Our main findings can be summarized as follows:

1. Predictability as measured by local Lyapunov exponents varies (multi)decadally by 9–18 %.
2. 'Active' ENSO periods are slightly more predictable than 'inactive' ones. Also, epochs with regular periodicity and moderate ENSO magnitude are classified as the most predictable by the local Lyapunov exponents.
3. The e-folding time is linearly related to ENSO frequency and standard deviation during epochs of distinct ENSO variability. However, the linear relationship between predictability and standard deviation does not hold for all 50-year periods of the simulation.
4. The ERSST.v3 reconstruction appears to lose information less rapidly than the unforced CM2.1 GCM. This could be revealing a discrepancy between real-world and GCM predictability: The GCM could be more 'chaotic' than the real world, due to an

overactive thermocline feedback, and deficient damping from evaporation and cloud-shading, likely related to an equatorial cold tongue bias. Or, the noise level, associated with atmospheric weather, could be higher in the GCM than in nature. Finally, this discrepancy may arise from and external radiative forcings present in the historical reconstruction.

5. Events with more west Pacific heat pile-up 5 years prior to the El Niño events are found to be slightly more predictable. Also, the thermocline starts deepening approximately 4 months earlier than the onset of the SST anomalies in the most predictable events.

The variation in predictability reported here is not large (9–18 %). It is possible then that changes in atmospheric noise (in the tropics or in the extratropics) at decadal timescales could produce such (multi)decadal variations in ENSO short-term predictability. Changes in the coupling strength between the ocean and the atmosphere could also produce these variations. Preliminary results using the methods employed here suggest that the stochasticity of winds play a role in determining predictability variations. Further investigation using the intermediate ZC model, and using measures of noise and ENSO stability in the GCM simulation might provide some explanation for the variability in predictability that is measured here by the LLEs.

Our investigation also answered the following questions:

1. Are century-long data sufficient to draw conclusions in terms of variability in ENSO predictability? In a related context, are century-long simulations or observations sufficient to capture key aspects of the ENSO dynamics?
2. Is the NINO3 SSTa variable sufficient to capture underlying ENSO dynamics?

With regard to the first question, we have shown that the relationship between predictability and ENSO variance and magnitude varies with epoch. It is not obvious that such a relationship exists when one studies the full 2,000-year simulation. This result suggests that caution be exercised when interpreting the ENSO predictability-magnitude relationships based on limited records. Moreover, the reduction in the variance of the local Lyapunov exponents computed from the 100-year versus the 2,000-year sample gives a sense of the uncertainty of predictability estimates due to sampling, and suggests that long model simulations are useful for putting the predictability inferred from the 100-year long record into perspective.

The second question was approached by assessing the capacity of the reconstructed phase-space of the NINO3 Index to reflect underlying ENSO dynamics. It was shown that the classification in terms of predictability on the basis of this phase-space reconstruction is in reasonable agreement with expectations from our physical understanding of the ENSO system. Heat-pile up in the west Pacific five years before the event, as well as a deepening of the thermocline four months earlier than the onset of the SST anomalies, are precursors of strong El Niño events. It seems that the tools we have used here provide a usable reflection of some dynamical system characteristics related to predictability, the limitations of the Takens embedding theorem and the use of a single state variable to represent a complex system notwithstanding. However, further analysis is required to ascertain these initial conclusions. Second, the heat content anomaly (or the thermocline depth anomaly) may be a better variable for phase-space reconstruction than the NINO3 Index, which is contaminated by weather noise. Therefore, the real issue here is associated with the predictability gain in a 'slow' versus a 'fast' variable, and with the predictability gain associated with spatial and temporal averaging.

The present paper is an early assessment of the applicability to GCM output of methods from dynamical systems theory that use observed data. The use of a computationally cheap relative measure of predictability in GCM simulations is of interest for intra- and inter-model comparisons. Characterizing relative predictability in a long climate model experiment could prove useful for identifying periods of the simulation that warrant further investigation of their dominant dynamics, or for guiding classical predictability studies, e.g. for selecting initial conditions for running re-forecast experiments.

Acknowledgments We thank Dr. Ken Takahashi, and an anonymous reviewer for their constructive comments that significantly improved this manuscript. We acknowledge the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA for dissemination of the NOAA ERSST.v3 datasets. CK was partially supported by the Alexander S. Onassis Public Benefit Foundation Scholarship Program. MAC was supported by grants NOAA-CICAR NA08OAR4320912, DOE DE-SC0005108 and the Office of Naval Research MURI (N00014-12-1-0911).

References

Abarbanel H (1995) Analysis of observed chaotic data. Institute for Nonlinear Science, Springer, Berlin

Abarbanel H, Brown R, Kennel M (1991) Variation of Lyapunov exponents on a strange attractor. J Nonlinear Sci 1:175–199

Abarbanel H, Brown R, Kennel M (1992) Local Lyapunov exponents computed from observed data. J Nonlinear Sci 2:343–365

Abarbanel H, Brown R, Sidorowich J, Tsimring L (1993) The analysis of observed chaotic data in physical systems. Rev Mod Phys 65:1331–1392

Barnston A, Glantz M, He Y (1999) Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997–98 El Niño episode and the 1998 La Niña onset. Bull Am Meteorol Soc 80:217–243

Barnston A, Tippett M, L'Heureux M, Li S, Dewitt D (2011) Skill of real-time seasonal enso model predictions during 2002–2011-is our capability increasing? In: Science and technology infusion climate bulletin, NOAAs National Weather Service, Fort Worth, TX, 3–6 Oct 2011, [36th NOAA annual climate diagnostics and prediction workshop]

Blumenthal M (1991) Predictability of a coupled ocean-atmosphere model. J Clim 4:766–784

Boffetta G, Giuliani P, Paladin G, Vulpiani A (1998) An extension of the Lyapunov analysis for the predictability problem. J Atmos Sci 55(23):3409–3412

Bryant P, Brown R, Abarbanel H (1990) Lyapunov exponents from observed time series. Phys Rev Lett 65:1523–1526

Cane M, Zebiak S, Dolan S (1986) Experimental forecasts of El Niño. Nature 321:827–832

Capotondi A, Wittenberg A, Masina S (2006) Spatial and temporal structure of tropical Pacific interannual variability in 20th century coupled simulations. Ocean Model 15:274–298. doi:10.1016/j.ocemod.2006.02.004

Chen D, Cane M, Kaplan A, Zebiak S, Huang D (2004) Predictability of El Niño over the past 148 years. Nature 428:733–6

Collins M, An SI, Cai W, Ganachaud A, Guilyardi E, Jin FF, Jochum M, Lengaigne M, Power S, Timmermann A, Vecchi G, Wittenberg A (2010) The impact of global warming on the tropical Pacific and El Niño. Nature Geosci 3:391–397. doi:10.1038/ngeo868

Delworth TL et al (2006) GFDL's CM2 global coupled climate models. Part I: formulation and simulation characteristics. J Clim 19:643–674

Delworth TL, Rosati A, Anderson W, Adcroft AJ, Balaji V, Benson R, Dixon K, Griffies SM, Lee HC, Pacanowski RC, Vecchi GA, Wittenberg AT, Zeng F, Zhang R (2012) Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. J Clim 25:2755–2781. doi:10.1175/JCLI-D-11-00316.1

DiNezio PN, Kirtman BP, Clement AC, Lee SK, Vecchi GA, Wittenberg A (2012) Mean climate controls on the simulated response of ENSO to increasing greenhouse gases. J Clim 25:7399–7420. doi:10.1175/JCLI-D-11-00494.1

Donner LJ, Wyman BL, Hemler RS, Horowitz LW, Ming Y, Zhao M, Golaz JC, Ginoux P, Lin SJ, Schwarzkopf MD et al (2011) The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. J Clim 24:3484–3519. doi:10.1175/2011JCLI3955.1

Drosdowsky W (2006) Statistical prediction of ENSO (Niño 3) using sub-surface temperature data. Geophys Res Lett 33:L03710. doi:10.1029/2005GL024866

Dunne JP, John JG, Adcroft AJ, Griffies SM, Hallberg RW, Shevliakova E, Stouffer RJ, Cooke W, Dunne KA, Harrison MJ, Krasting JP, Malyshev SL, Milly PCD, Phillipps PJ, Sentman LT, Samuels BL, Spelman MJ, Winton M, Wittenberg AT, Zadeh N (2012) GFDL's ESM2 global coupled climate-carbon Earth System Models. Part I: physical formulation and baseline simulation characteristics. J Clim 25:6646

Dunne JP, John JG, Shevliakova E, Stouffer RJ, Krasting JP, Malyshev SL, Milly PCD, Sentman LT, Adcroft AJ, Cooke W, Dunne KA, Griffies SM, Hallberg RW, Harrison MJ, Levy H, Wittenberg AT, Phillips PJ, Zadeh N (2013) GFDL's ESM2 global coupled climate-carbon Earth System Models. Part II: carbon system formulation and baseline simulation characteristics. J Clim. doi:10.1175/JCLI-D-12-00150.1

Eisenman I, Yu L, Tziperman E (2005) Westerly wind bursts: ENSOs tail rather than the dog? J Clim 18:5224–5238

Emile-Geay J, Cane M, Seager R, Kaplan A, Almasi P (2007) El Niño as a mediator of the solar influence on climate. Paleoceanography 22:PA3210. doi:10.1029/2006PA001304

Emile-Geay J, Cobb KM, Mann ME, Wittenberg AT (2013a) Estimating central equatorial Pacific SST variability over the past millennium. Part I: methodology and validation. J Clim. doi:10.1175/JCLI-D-11-00510.1

Emile-Geay J, Cobb KM, Mann ME, Wittenberg AT (2013b) Estimating central equatorial Pacific SST variability over the past millennium. Part II: reconstructions and implications. J Clim. doi:10.1175/JCLI-D-11-00511.1

Fedorov A (2002) The response of the coupled tropical ocean-atmosphere to westerly wind bursts. Q J Roy Meteorol Soc 128:1–23

Fedorov A, Philander S (2001) A stability analysis of the tropical ocean-atmosphere interactions (bridging the measurements of, and the theory for El Niño). J Clim 14:3086–3101

Fedorov A, Harper S, Philander S, Winter B, Wittenberg A (2003) How predictable is El Niño? Bull Am Meteorol Soc 84:911–919

Galbraith ED, Kwon EY, Gnanadesikan A, Rodgers KB, Griffies SM, Bianchi D, Sarmiento JL, Dunne JP, Simeon J, Slater RD, Wittenberg AT, Held IM (2011) Climate variability and radiocarbon in the CM2Mc Earth System Model. J Clim 24:4230–4254. doi:10.1175/2011JCLI3919.1

Gallager RG (1968) Information theory and reliable communication. Wiley, New York

Gao J, Cao Y, Tung WW, Hu J (2007) Multiscale analysis of complex time series: integration of chaos and random fractal theory, and beyond. Wiley, New York

Gebbie G, Eisenman I, Wittenberg A, Tziperman E (2007) Modulation of westerly wind bursts by sea surface temperature: a semistochastic feedback for ENSO. J Atmos Sci 64:3281–3295. doi:10.1175/JAS4029.1

Ghil M, Chekroun M, Simonnet E (2008) Climate dynamics and fluid mechanics: natural variability and related uncertainties. Physica D 237:2111–2126

Gnanadesikan A, Dixon KW, Griffies SM et al (2006) GFDL's CM2 global coupled climate models, part II: the baseline ocean simulation. J Clim 19:675–697. doi:10.1175/JCLI3630.1

Goswami B, Shukla J (1991) Predictability of a coupled ocean atmosphere model. J Clim 3:2–22

Goswami B, Rajendran K, Sengupta D (1991) Source of seasonality and scale dependence of predictability in a coupled ocean-atmosphere model. Mon Wea Rev 125:846–858

Griffies SM, Winton M, Donner LJ, Horowitz LW et al (2011) GFDL's CM3 coupled climate model: characteristics of the ocean and sea ice simulations. J Clim 24:3520–3544. doi:10.1175/2011JCLI3964.1

Guilyardi E, Wittenberg A, Fedorov A et al (2009) Understanding El Niño in ocean-atmosphere general circulation models: progress and challenges. Bull Am Meteorol Soc 90:325–340

Hazeleger W, Visbeck M, Cane M, Karspeck A, Naik N (2001) Decadal upper ocean temperature variability in the tropical Pacific. J Geophys Res 106:8971–8988. doi:10.1029/2000JC000536

Jin FF, Neelin JD, Ghil M (1994) El Niño on the devil's staircase: annual subharmonic steps to chaos. Science 264:70–72

Karspeck A, Seager R, Cane M (2004) Predictability of tropical pacific decadal variability in an intermediate model. J Clim 17(14):2842–2850

Kennel M, Brown R, Abarbanel H (1992) Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys Rev A 45(6):3403–3411

Kennel M, Abarbanel H, Sidorowich JJ (1994) Prediction errors and local Lyapunov exponents. Chaotic dynamics found at arXiv: chao-dyn/9403001v1

Kim HM, Webster P, Curry J (2009) Impact of shifting patterns of Pacific Ocean warming on North Atlantic tropical cyclones. Science 325(5936):77–80

Kirtman BP, Schopf P (1998) Decadal variability in enso predictability and prediction. J Clim 11:2804–2822
Kleeman R (2008) Stochastic theories for the irregularity of ENSO. Philos Trans R Soc A 366(1875):2509–2524
Kug JS, Choi J, An SI, Jin FF, Wittenberg AT (2010) Warm pool and cold tongue El Nino events as simulated by the GFDL CM2.1 coupled GCM. J Clim 23:1226–1239. doi:10.1175/2009JCLI43293.1
Landsea C, Knaff J (2000) How much skill was there in forecasting the very strong 1997–98 El Niño?. Bull Am Meteorol Soc 81(9):2107–2120
Legras B, Ghil M (1985) Persistent anomalies, blocking and variations in atmospheric predictability. J Atmos Sci 42:433–471
Lengaigne M, Guilyardi E, Boulanger JP, Menkes C, Delecluse P, Inness P, Cole J, Slingo J (2004) Triggering of El Niño by westerly wind events in a coupled general circulation model. Clim Dyn 23(6):601–620. doi:10.1007/s00382-004-0457-2
Lima C, Lall U, Jebara T, Barnston A (2009) Statistical prediction of enso from subsurface sea temperature using a nonlinear dimensionality reduction. J Clim 22:4501–4519
Lin JL (2007) Interdecadal variability of ENSO in 21 IPCC AR4 coupled GCMs. Geophys Res Lett 34:L12702. doi:10.1029/2006GL028937
Mann M, Cane M, Zebiak S, Clement A (2005) Volcanic and solar forcing of El Niño over the past 1000 years. J Clim 18:447–456
Moon YI, Rajagopalan B, Lall U (1995) Estimation of mutual information using kernel density estimators. Phys Rev E 52:2318–2321
Munnich M, Cane M, Zebiak S (1991) A study of self-excited oscillations of the tropical ocean-atmosphere system. Part II: nonlinear cases. J Atmos Sci 48:1238–1248
Neelin J, Battisti D, Hirst A, Jin FF, Wakata Y, Yamagata T, Zebiak SE (1998) Enso theory. J Geophys Res 14:261–290
Nese J (1989) Quantifying local predictability in phase space. Physica D 35(1–2):237–250
Nese J, Dutton J (1989) Quantifying predictability variations in a low-order occan-atmosphere model: a dynamical systems approach. J Climate 6:185–204
Oseledec V (1968) Multiplicative ergodic theorem: characteristic Lyapunov exponents of dynamical systems. Trudy MMO 19:179–210 (in Russian)
Penland C, Matrosova L (2006) Studies of El Niño and interdecadal variability in tropical sea surface temperatures using a nonnormal filter. J Clim 19:5796–5815
Pesin Y (1977) Characteristic Lyapunov exponents and smooth ergodic theory. Russ Math Surv 32(4):55–114
Power S, Haylock M, Wang RCX (2006) The predictability of interdecadal changes in enso activity and enso teleconnections. J Clim 19:4755–4771
Rabinovich M (1978) Stochastic self-oscillations and turbulence. Usp Fiz Nauk 125:123–168
Ruiz J, Cordery I, Sharma A (2005) Integrating ocean subsurface temperatures in statistical enso forecasts. J Clim 18:3571–3586
Sarachik E, Cane M (2010) The El Niño-southern oscillation phenomenon. Cambridge University Press, Cambridge
Sauer T, Yorke J, Casdagli M (1991) Embedeology. J Stat Phys 65
Schneider E, Kirtman B, DeWitt D, Rosati A, Ji L, Tribbia J (2003) Retrospective ENSO forecasts: sensitivity to atmospheric model and ocean resolution. Mon Wea Rev 131:3038–3060
Shannon C (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423
Siqueira L, Kirtman B (2012) Predictability of a low-order interactive ensemble. Nonlinear Process Geophys 19:273–282
Smith L, Ziehmann C, Fraedrich K (1999) Uncertainty dynamics and predictability in chaotic systems. QJR Meteorol Soc 125:2855–2886
Sun C, Rienecker MM, Rosati A, Harrison M, Wittenberg A, Keppenne CL, Jacob JP, Kovach RM (2007) Comparison and sensitivity of ODASI ocean analyses in the tropical Pacific. Mon Wea Rev 135:2242–2264. doi:10.1175/MWR3405.1
Takens F (1981) Detecting strange attractors in turbulence. In: Rand D, Young LS (eds)Dynamical systems and turbulence, Warwick 1980. Lecture notes in mathematics, vol 898. Springer Berlin, Heidelberg, pp 366–381. ISBN:978-3-540-11171-9
Thompson C, Battisti D (2000) A linear stochastic dynamical model of ENSO. Part I: model development. J Clim 13:2818–2832
Thompson C, Battisti D (2001) A linear stochastic dynamical model of ENSO. Part II: analysis. J Clim 14:445–466
Timmermann A, Jin FF (2002) A nonlinear mechanism for decadal El Niño amplitude changes. Geophys Res Lett 29(1):1003. doi:10.1029/2001GL013369
Tziperman E, Yu L (2007) Quantifying the dependence of westerly wind bursts on the large-scale tropical pacific SST. J Clim 20:2760–2768
Tziperman E, Stone L, Cane M, Jarosh H (1994) El Niño chaos: overlapping of resonances between the seasonal cycle and the pacific ocean-atmosphere oscillator. Science 264:72–74
Vecchi GA, Wittenberg AT (2010) El Niño and our future climate: where do we stand? Wiley Interdiscip Rev Clim Change 1:260–270. doi:10.1002/wcc.33
Vecchi GA, Wittenberg AT, Rosati A (2006) Reassessing the role of stochastic forcing in the 1997–1998 El Niño. Geophys Res Lett 33:L01706. doi:10.1029/2005GL024738
Wittenberg A (2009) Are historical records sufficient to constrain enso simulations?. Geophys Res Lett 36:L12702. doi:10.1029/2009GL038710
Wittenberg A, Rosati A, Lau NC, Ploshay JJ (2006) GFDL's CM2 global coupled climate models. Part III: tropical Pacific climate and ENSO. J Clim 19:698–722
Wittenberg AT (2002) ENSO response to altered climates. PhD thesis, Princeton University, 475 pp. Available at http://www.gfdl.noaa.gov/atw/research/thesis
Yang X, Rosati A, Zhang S, Delworth TL, Gudgel RG, Zhang R, Vecchi G, Anderson W, Chang YS, DelSole T, Dixon K, Msadek R, Stern WF, Wittenberg A, Zeng F (2013) A predictable AMO-like pattern in GFDL's fully-coupled ensemble initialization and decadal forecasting system. J Clim 26:650–661. doi:10.1175/JCLI-D-12-00231.1
Zavala-Garay J, Zhang C, Moore AM, Wittenberg AT, Harrison MJ, Rosati A, Vialard J, Kleeman R (2008) Sensitivity of hybrid ENSO models to unresolved atmospheric variability. J Clim 21:3704–3721. doi:10.1175/2007JCLI1188.1
Zebiak S, Cane M (1987) A model for El Niño-southern oscillation. Mon Wea Rev 115:2262–2278
Zhang S, Anderson JL, Rosati A, Harrison M, Khare SP, Wittenberg AT (2004) Multiple time level adjustment for data assimilation. Tellus 56A:2–15. doi:10.1111/j.1600-0870.2004.00040.x
Zhang S, Harrison MJ, Wittenberg AT, Rosati A, Anderson JL, Balaji V (2005) Initialization of an ENSO forecast system using a parallelized ensemble filter. Mon Wea Rev 133:3176–3201. doi:10.1175/MWR3024.1
Zhang S, Harrison MJ, Rosati A, Wittenberg A (2007) System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. Mon Wea Rev 135:3541–3564. doi:10.1175/MWR3466.1