Nonlinear Processes in Geophysics (2003) 10: 197–210

Nonlinear Processes in Geophysics

Cyclic Markov chains with an application to an intermediate ENSO model

R. A. Pasmanter[1] **and A. Timmermann**[2]

[1]KNMI, Postbus 201, 3730 AE De Bilt, the Netherlands
[2]Institut für Meereskunde, Düsternbrooker Weg 20, D-24105 Kiel, Germany

Received: 4 September 2001 – Revised: 3 June 2002 – Accepted: 24 June 2002

Abstract. We develop the theory of cyclic Markov chains and apply it to the El Niño-Southern Oscillation (ENSO) predictability problem. At the core of Markov chain modelling is a partition of the state space such that the transition rates between different state space cells can be computed and used most efficiently. We apply a partition technique, which divides the state space into multidimensional cells containing an equal number of data points. This partition leads to mathematical properties of the transition matrices which can be exploited further such as to establish connections with the dynamical theory of unstable periodic orbits. We introduce the concept of most and least predictable states. The data basis of our analysis consists of a multicentury-long data set obtained from an intermediate coupled atmosphere-ocean model of the tropical Pacific. This cyclostationary Markov chain approach captures the spring barrier in ENSO predictability and gives insight also into the dependence of ENSO predictability on the climatic state.

1 Introduction

One of the main challenges being addressed by climate research is that of climate prediction on time scales ranging from several months to years and up to decades. In this context, modern statistical techniques (e.g. Barnston and Ropelewski, 1992; Penland and Magorian, 1993; Xue et al., 1994; van den Dool and Barnston, 1995; Tangang et al., 1997) as well as climate models of different complexity (e.g. Cane et al., 1986; Zebiak and Cane, 1987; Blumenthal, 1991; Goswami and Shukla, 1991a, b; Balmaseda et al., 1994; Oberhuber et al., 1998; Stockdale et al., 1998; Ji et al., 1998; Mason et al., 1999; Grötzner et al., 1998) are being used in order to predict climate fluctuations sufficiently in advance for societies to take precautionary measures. Probably the most prominent example of this is the El Niño-Southern Oscillation (ENSO) phenomenon which, sometimes, can be predicted with an anticipation longer than six months.

Despite recent success in predicting ENSO using statistical and physical models (Latif et al., 1994) it turned out that, partly due to model-reality mismatches and partly due to initial state errors, the models perform very well in some years whereas they fail in others. Similar to the predictability of low-dimensional nonlinear dynamical systems (Smith et al., 1999), ENSO predictability is state dependent. To account for the state dependence of predictability in ENSO models, singular vector techniques have been employed (Chen et al., 1995; Eckert, 1997; Moore and Kleeman, 1997a, b). The singular vectors represent those states of the system which are associated with the strongest error-growth characteristics. However, by construction, they only capture the linear error growth along piecewise linearized trajectories. Furthermore, the computation of singular vectors in ENSO models requires a linearized version of the model code as well as of its adjoint. For fully coupled general circulation models (CGCMs) such an approach is not feasible due partly to limited computer resources and partly to the fundamental problem of formulating adjoint operators of coupled systems with different time scales. In our paper we describe an alternative approach, based on the statistical investigation of model generated time series rather than on manipulating computer code (Eckert, 1997; Chen et al., 1995; Moore and Kleeman, 1997a, b). The statistical technique employed here is based on the estimation of a dynamical equation for probability densities. In contraposition to the singular vector approach, see, e.g. Palmer et al. (1998), this technique is fully nonlinear and is also applicable to CGCM simulations. The essential part of the dynamical equation is a transition matrix which describes the probability of observing, in one time step, transitions between different states. This so-called Markov chain approach was already applied to the ENSO prediction problem by Fraedrich (1988)[1]. Fraedrich's results are based on rather short ENSO

Correspondence to: A. Timmermann
(atimmermann@fm.uni-kiel.de)

[1]This technique has been successfully applied also in other meteorological (Spekat et al., 1983; de Swart and Grasman, 1987;

time series; his analysis is univariate and it does not take into account the cyclostationarity of the ENSO system, see Flügel et al. (1999). The present work improves on these aspects by applying the theory of cyclic Markov chains to bivariate data series characterizing ENSO. We point out the analogies between cyclic Markov chains and the Floquet theory of ordinary differential equations with periodic coefficients. In this way, e.g. one is able to detect the so-called spring-barrier in ENSO's predictability.

Moreover, we partition the system's state space in such a way that each multidimensional cell contains the same number of observations, i.e. the system spends an equal amount of time in each cell. This turns out to be not only an efficient way of using the data and the state space, it also leads to a direct and transparent expression of the dynamics in terms of permutations matrices acting on the discretized state space, i.e. in terms of cycles. As the discretization becomes finer, one expects this cycles expansion to converge into the unstable periodic orbits (UPOs) expansion of chaotic attractors, see, e.g. P. Cvitanovic (1991) and the references therein. Hence, here one could use the term coarse-grained unstable periodic orbits in order to stress the fact that our discretization of the state space is not arbitrarily fine. A nice example of UPOs in the ENSO context can be found in Tziperman et al. (1997). In practice, one can construct reliable transition matrices for just a small number, for example three, dynamical variables. Therefore, it is of utmost importance to choose the right (three) variables, otherwise the results may be practically irrelevant.

The paper is organized as follows: Sect. 2 gives a brief introduction to the theory of cyclic Markov chains with a special emphasis on the connections with dynamical systems and unstable periodic orbits. Also, the connection between the concept of predictability and a suitably defined entropy production, or information loss rate, is brought to the fore. In Sect. 3 we study the state dependence of ENSO as simulated by an intermediate ENSO model (Zebiak and Cane, 1987). We apply univariate and bivariate Markov chains in order to determine both the least and the most predictable states in this model. Also, the predictability spring barrier is recovered. We conclude in Sect. 4 with a summary and discussion of the technique we applied, of the limitations that may show up in practice and of our main results.

2 Theoretical foundation of cyclic Markov chains

Before venturing into ENSO predictability in coupled models, we briefly describe the theory of Markov chains and discuss how to estimate master equations from data also in the case of cyclostationary systems.

Consider a dynamical system with V degrees of freedom denoted by $x^\alpha, 1 \leq \alpha \leq V$. The Markov chain description of the system is obtained by discretizing time as well as the system's configuration space. After discretizing the time variable t, the dynamical variables at time $t+1$ are related to their values at the previous time step t by

$$x^\alpha(t+1) = f^\alpha(\boldsymbol{x}(t), \xi(t), t),$$

where ξ stands for any possible random factors, while the explicit time dependence stands for the non-random external factors. For the sake of illustration and of simplicity, let us assume that

$$x^\alpha(t+1) = f^\alpha(x,t) + \xi^\alpha(t) \tag{1}$$

where the random terms $\xi^\alpha(t)$ are Gaussian, uncorrelated and their variances are

$$\left\langle \xi^\alpha(t)\xi^\beta(t')\right\rangle = \Delta_\alpha^2(t)\delta^{\alpha\beta}\delta(t-t'). \tag{2}$$

The case $\Delta_\alpha^2(t) = 0$ describes a deterministic system. Autonomous systems correspond to $f^\alpha(x(t),t) = f^\alpha(x(t))$ and $\Delta_\alpha^2(t) = \Delta_\alpha^2$ for all α.

Even in the purely deterministic case, it makes sense to consider the probability of observing, at time t, the dynamical variables having values between say $\boldsymbol{x}$ and $(\boldsymbol{x} + d\boldsymbol{x})$. We denote this probability density by $p(\boldsymbol{x},t)$ with $p(\boldsymbol{x},t) \geq 0$ and $\int p(\boldsymbol{x},t)\, d^V x = 1$. The time evolution of this probability density is expressed then in terms of the so-called Frobenius-Perron (FP) integral operator

$$p(\boldsymbol{y}, t+1) = \int d^V x\, \mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x}, t) p(\boldsymbol{x}, t).$$

One always has that $\mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x}, t) \geq 0$ and $\int d^V y\, \mathcal{L}(\boldsymbol{y}, \boldsymbol{x}, t) = 1$. In the case of additive Gaussian noise characterized by Eq. (1) and Eq. (2), the explicit form of the Frobenius-Perron operator is

$$\mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x}, t) := \left(\sqrt{2\pi}\right)^{-V} \Delta^{-1} \cdot \exp\left(-\sum_{\alpha=1}^{V} \frac{(y^\alpha - f^\alpha(\boldsymbol{x},t))^2}{2\Delta_\alpha^2(t)}\right), \tag{3}$$

with $\Delta(t) = \prod_{\alpha=1}^{V} \Delta_\alpha(t)$. For autonomous systems, i.e. $\mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x}, t) = \mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x})$, one has a stationary probability density[2], call it $p_o(\boldsymbol{x})$, which is invariant under the time evolution, i.e.

$$p_o(\boldsymbol{y}) = \int d^V x\, \mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x}) p_o(\boldsymbol{x}).$$

If the system is non-autonomous but its dynamics is periodic in time, i.e. $\mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x}, t+T) = \mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x}, t)$, where T is the corresponding period, then the analogue of the stationary distribution will be time-dependent: besides the oscillations with period T it may also contain (sub)harmonics. The system we shall consider, ENSO, is non-autonomous since its dynamics is modulated by the yearly seasonal cycle .

Vautard et al., 1990; Nicolis et al., 1997; Egger, 2001) and oceanographic (Cencini et al., 1999) contexts.

[2] For the sake of simplicity, we assume there is only one such invariant measure. In general, more than one invariant measure can exist if bounded noise is assumed.

Next, one partitions the state space into a finite number C of cells. We will identify each cell by a Latin subscript. The occupation probability of each cell is,

$$p^i(t) := \int_i d^V x \; p(\boldsymbol{x}, t) \geq 0, \tag{4}$$

and satisfies

$$\sum_{i=1}^{C} p^i(t) = 1.$$

Our partial knowledge about the state of the system at time t is contained in $\boldsymbol{p}(t)$, we call $\boldsymbol{p}(t)$ either the probability distribution or the state vector.
Finally, one defines the cell-to-cell transition rates $m_t(j, i)$ by

$$m_t(j, i) = \Omega_i^{-1} \int_j d^V y \; \int_i d^V x \; \mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x}, t),$$

with $\Omega_i = \int_i d^V x$.

The dimensionless numbers $m_t(j, i)$ are the fraction of times that one observes the system in the j-th cell at time $t + 1$ when the system was in the i-th cell at time t. The partition of initial and final states need not be identical. In the following, we deal always with initial and final partitions that are not (necessarily) identical but have the same number of cells C. The transition rates $m_t(j, i)$ can be seen then as elements of a nonnegative $C \times C$ matrix that we shall denote by $\mathbf{m}_t$. The conservation of the total probability takes now the form,

$$\sum_{i=1}^{C} m_t(i, j) = 1, \text{ for all } j. \tag{5}$$

This means also that the vector $(1, 1, \ldots, 1, 1)$ is a left eigenvector of the matrices m_t with eigenvalue 1. Nonnegative matrices satisfying (5) are called stochastic matrices; they form a semigroup, i.e. the product of two stochastic matrices is a stochastic matrix; the inverse of a stochastic matrix, if it exists, is not necessarily a stochastic matrix. The occupation probability of each cell will evolve according to

$$p^i(t + 1) = \sum_{j=1}^{C} m_t(i, j) p^j(t),$$

whenever the probability density $p(\boldsymbol{x}, t)$ and/or the kernel $\mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x}, t)$ is sufficiently constant over the x-cells. Moreover, by taking finer partitions with larger C one expects to approximate the real dynamics with higher precision.

Nonnegative matrices and, more specifically, stochastic matrices have been extensively studied, see, e.g. Berman and Plemmons (1979). We recall some of the important properties that they share. The Perron-Frobenius theorem tells us that for any stochastic matrix $\mathbf{m}$, 1) there is at least one positive eigenvector with eigenvalue 1, i.e. $\mathbf{m} \cdot \boldsymbol{v} = \boldsymbol{v}$ and $v^i > 0$ for all $1 \leq i \leq C$, such an eigenvector is called a Perron vector, 2) all other eigenvalues have absolute values smaller than or equal to 1 and 3) if $\mathbf{m}$ is irreducible[3], then the positive right eigenvector is unique. The Perron vector, is then the (unique) stationary distribution, $\boldsymbol{p}_o = \boldsymbol{v}$. The stochastic matrices we shall encounter are irreducible and, moreover, all their other eigenvalues have absolute values smaller than 1; such matrices are called primitive. Irreducibility means that there is only one attractor in state space.

As just stated, irreducible stochastic matrices have only one positive eigenvector. This means that all other eigenvectors have positive and negative components, moreover, some components may be complex. On the other hand, the state vectors $\boldsymbol{p}(t)$ describe probability distributions, i.e. their components are nonnegative at all times, $p^j(t) \geq 0$, for all j and for all $t \geq 0$, as it is ensured by the property (5) satisfied by the transition matrices $\mathbf{m}_t$. These two observations make evident that the non-Perron eigenvectors are not probability distributions but deviations away from the stationary distribution, i.e. deviations with respect to the Perron vector. Similarly, complex eigenvectors will always be accompanied by their complex conjugates; the presence of pairs of complex conjugate eigenvalues tells us that the corresponding eigenvectors do not decay monotonically but that they oscillate in time with a period that equals $(2\pi i) / \left[\ln(\arg \lambda)\right]$, when it is measured in units of the time step. In the general case, stochastic matrices are neither symmetric nor normal; on the physical relevance of nonnormal matrices see, e.g. Lorenz (1965) and Trefethen et al. (1993).

In order to avoid unnecessary complications, in the next Subsections we limit the discussion to autonomous systems. The generalization to non-autonomous systems is done in Sect. 2.4.

The concept of Markov chains is illustrated and summarized in Fig. 1. This figure also shows the main analogies with the expansion of deterministic dynamical systems in terms of eigenvectors of the Frobenius Perron operator.

2.1 Data (equi)partition

Suppose that we are given a long data series consisting of D measurements with $D \gg 1$. We assume that transient phenomena (stataistically transient), if present, play a negligible role in the data series. These measurements allow us to define the system state space and subsequently to partition this space into a number of cells. The construction of these cells can be done in an infinite number of different ways. The computation of some important dynamical quantities is sensitive to the way in which the cells are constructed. Consequently, one is interested in finding those partitions that make reliable and efficient computations possible.

The partitions we have been using are constructed as follows. Consider first the case when one measures only one dynamical variable. Then the D measurements are spread between a minimum and a maximum value; these extreme

[3] A nonnegative matrix $\mathbf{m}$ is irreducible if and only if there is no permutation $\mathbf{P}$ such that $\mathbf{PmP}^\dagger$ is of the form $\begin{pmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{C} & \mathbf{S}_2 \end{pmatrix}$, where $\mathbf{S}_1$ and $\mathbf{S}_2$ are square matrices

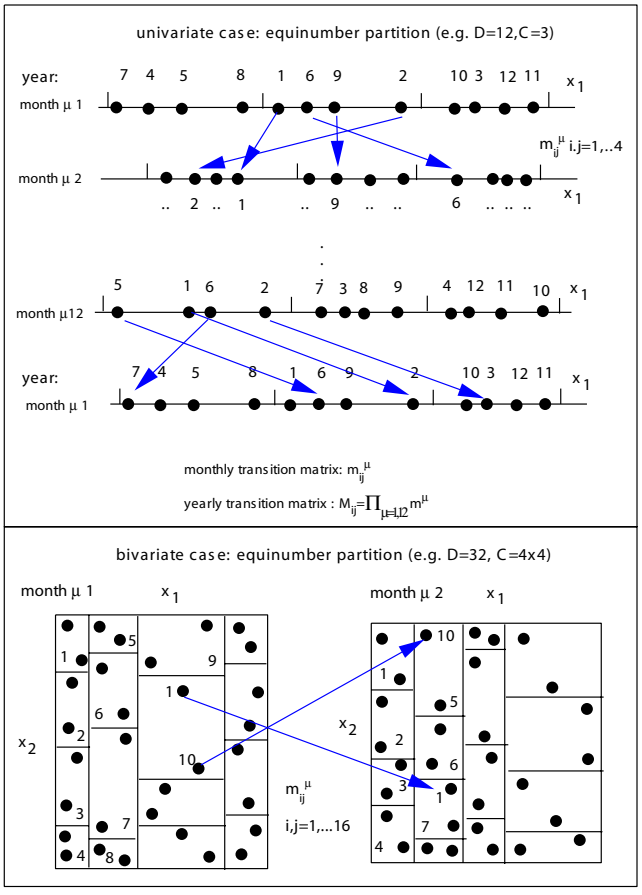


Fig. 1. Data partition for the univariate and bivariate cases.

values define the variable's range. Partition this range into C intervals such that each interval contains an equal number of data points. The number of data points in each interval is $d = D/C$. Ideally, we would like to have enough points in each cell so that, potentially, every cell can be accessed from every cell, i.e. one would like to have $d \geq C$, i.e. $D \geq C^2$.

Consider next the two dimensional case, i.e. when one measures two simultaneous dynamical variables. Firstly, partition the D data points into $\sqrt{C}$ cells along the coordinate corresponding to one of the dynamical variables, each cell containing $D/\sqrt{C}$ points[4]. Then partition each of these cells into $\sqrt{C}$ cells along the second coordinate corresponding to the other measured dynamical variable. In this way one generates C cells, each one containing $d = D/C$ data points. The procedure is illustrated in Fig. 1.

This partitioning can be applied to higher-dimensional measurements. In this way we can always partition the state space into 'cells of equal weight', i.e. the time series spends the same amount of time in each of the cells. It is intuitively clear that this is a very efficient way of defining a partition. This partition does not automatically ensure that the system's description satisfies the Markovian property. The Markovian property is assumed in our approach and has to be verified a posteriori. Like every sorting procedure, the manipulations required for this task are very time consuming and special algorithms have been developed in order to do this in an efficient way, see, e.g. Chap. 14 of *The nature of mathematical modelling* by Gershenfeld, Cambridge University Press, (1999) and the references therein.

[4] For the sake of simplicity, we are assuming that $\sqrt{C}$ and $D/\sqrt{C}$ are integers.

2.2 Double stochastic matrices and unstable periodic orbits

Since we constructed our partition in such a way that each cell contains the same number of occurrences, it follows that the probability distribution corresponding to the complete data set is represented by the state vector

$$\boldsymbol{p}_{data} = \begin{pmatrix} 1/C \\ \vdots \\ 1/C \end{pmatrix}. \tag{6}$$

Then one must have

$$\mathbf{m} \cdot \boldsymbol{p}_{data} = \boldsymbol{p}_{data},$$

i.e. $\boldsymbol{p}_{data}$ as given by Eq. (6) is the stationary distribution or Perron vector $\boldsymbol{p}_o$ with eigenvalue 1. This means that, besides the probability conservation constraint (5), the transition matrices satisfy also

$$\sum_{k=1}^{C} m(i,k) = 1. \tag{7}$$

Non-negative matrices satisfying both constraints (5) and (7) are called *doubly stochastic matrices*. Also, these matrices form a semigroup.

A theorem by Birkhoff (1946) states a remarkable property of these matrices: every doubly stochastic $C \times C$ matrix can be written as a convex combination of permutation matrices, i.e.

$$\text{doubly stochastic } \mathbf{M} \rightarrow \mathbf{M} = c_1\mathbf{P}_1 + \cdots + c_N\mathbf{P}_N \tag{8}$$

$$\text{with } c_i > 0,\ \sum_{i=1}^{N} c_i = 1 \ \text{ and } \ N \leq C^2 - 2C + 2.$$

Recall that a permutation matrix is a doubly stochastic matrix whose elements are either 0 or 1. It is easy to see that such a matrix only exchanges the contents between cells and that by rising it to an appropriate power one gets the identity matrix. In other words, a permutation matrix describes a (combination) of cyclic, periodic evolution. When the number of cells C goes to infinity, one expects this Birkhoff expansion to converge into the cycles expansion that has been developed for studying dynamical systems, see, e.g. Cvitanović (1991) and the references therein. It describes the dynamics in terms of coarse-grained unstable periodic orbits. They are unstable to the extent that $c_i < 1$ for all i; in fact, one can associate a lifetime $1/|\ln c_i|$ to the permutation $\mathbf{P}_i$. Just like

the cycles expansion, the Birkhoff expansion is not unique. For example, the Markov chain described by the following doubly stochastic transition matrix,

$$\begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

can be decomposed in terms of permutations in more than one way. One possible decomposition is

$$\frac{1}{2}\begin{pmatrix} 0&0&0&1 \\ 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0&0&1&0 \\ 0&1&0&0 \\ 1&0&0&0 \\ 0&0&0&1 \end{pmatrix}.$$

Another possible decomposition is:

$$\frac{1}{2}\begin{pmatrix} 0&0&0&1 \\ 0&1&0&0 \\ 1&0&0&0 \\ 0&0&1&0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0&0&1&0 \\ 1&0&0&0 \\ 0&1&0&0 \\ 0&0&0&1 \end{pmatrix}.$$

Generally, one is interested in the expansion (8) with the smallest possible number of terms N and, consequently, with the largest amplitudes c_i (or longest lifetimes).

In closing this Subsection, it is worthwhile recalling that, for a large class of irreducible stochastic matrices $\mathbf{m}$, it is possible to find positive diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ such that $\mathbf{D}_1 \cdot \mathbf{m} \cdot \mathbf{D}_2$ is a doubly stochastic matrix, see Brualdi et al. (1966) and Sinkhorn and Knopp (1967).

2.3 Information loss

The transition matrices $\mathbf{m}$ may be singular, i.e. one or more eigenvalues may be zero. Evidently, in such a case, some of the information contained in $\boldsymbol{p}(t)$ is irretrievably lost when passing to $t+1$. More generally, the decay of all modes[5] but for the Perron vector, means that information about the departure of the state vector $\boldsymbol{p}(t)$ from the stationary state $\boldsymbol{p}_o$ is lost. How can we quantify this information loss? A convenient and often used measure of the 'distance' between, say, a state $\boldsymbol{p}(t)$ and $\boldsymbol{p}_o$ is[6]

$$I(t) = +\sum_{i=1}^{C} p^i(t) \ln \frac{p^i(t)}{p_o^i} \geq 0,$$

This quantity is also bounded from above, $\ln C \geq I(t)$, the maximum value of $I(t)$ is achieved when only one component of $\boldsymbol{p}(t)$ is different from zero. The decay of the modes implies that, on the average, the information content $I(t)$ diminishes[7] until it reaches zero, i.e. when $p^i(t) = p_o^i$.

[5] We use 'mode' and 'eigenvector' indistinguishably.

[6] There are other possible definitions; this one is the only extensive one, i.e. extensive over uncorrelated degrees of freedom.

[7] The decay of $I(t)$ is not necessarily monotonic: transiently it may happen that $I(t)$ increases.

Suppose that at $t = 0$ we know with high precision the state of our system, i.e. at $t = 0$ the probability distribution is totally localized in one cell, say it is in cell k,

$$p^i(0) = \delta_k^i$$

so that the information attains its maximum possible value $I_k(0) = \ln C$. One time-step later, the state will be $p^i(1) = m(i,k)$ and the corresponding information content will be

$$I_k(1) = \sum_{i=1}^{C} m(i,k) \ln m(i,k) + \ln C.$$

Therefore, the information lost in the first time step of the evolution is

$$\Delta I_k := I_k(0) - I_k(1) = -\sum_{i=1}^{C} m(i,k) \ln m(i,k) \geq 0.$$

By construction, all cells have equal weight (or probability), therefore the probability of starting from cell k is just $1/C$. Averaging over all possible initial conditions, i.e. over all initial cells, we get the average loss of information in one time-step

$$\langle \Delta I \rangle = -\frac{1}{C} \sum_{k=1}^{C} \sum_{i=1}^{C} m(i,k) \ln m(i,k) \geq 0. \tag{9}$$

Since permutation matrices are characterized by $m(i,k) = \delta^i_{f(k)}$ for some invertible $f(k)$ we see that permutation matrices achieve the lower bound in information loss. In fact, permutation matrices are the only doubly stochastic matrices with $\langle \Delta I \rangle = 0$. This agrees with our expectations since a permutation corresponds to a totally reversible dynamics.

Notice that the information loss is strongly dependent upon the chosen partition. For example, one can construct the cells in such a contrived way that all the transition rates equal $1/C$, i.e. $m(i,k) = 1/C$ for all (i,j) and the information loss achieves its largest possible value $\ln C$. On the other hand, the minimum information loss, *taken over all possible partitions*, is an intrinsic property of the system.

2.4 Cyclic Markov chains and Floquet theory

The effects of the yearly cycle are clearly present in the ENSO phenomenon. In this Subsection we explain how to include this fact in a Markov-chain description.

The data on which the analysis is based are monthly averages of some relevant variables. The partition described in the previous Subsections is applied to each month, i.e. we create 12 equipartitions, each with a number C of cells, each cell containing $D/12C$ number of measurements. This is illustrated in Fig. 1. The precise position in state space of the cells differs from month to month. Next, we compute 12 transition-rate matrices, e.g. $m_1(i,j)$ denotes the transition rate from the j-th January cell to the i-th February cell and so successively up to the December to January transition-rates matrix $\mathbf{m}_{12}$. Therefore, $\mathbf{M}_1$ the transition matrix after a

yearly cycle starting from month 1 and ending in the same month one year later, is given by

$$\mathbf{M}_1 := \mathbf{m}_{12} \cdot \mathbf{m}_{11} \cdot \cdots \cdot \mathbf{m}_2 \cdot \mathbf{m}_1$$

Notice that it is not necessary to compute the twelve intermediary month-to-month transition matrices $\mathbf{m}_\mu$ in order to compute $\mathbf{M}_1$, it can be computed directly from the yearly transition rates, i.e. $\mathbf{M}_1(i, j)$ is the fraction of month-1 data points that were in the j-th cell and reached the i-th cell of the same month one year later. In this way, it is possible to construct twelve year-to-year transition matrices $\mathbf{M}_\mu$, $\mu = 1, \ldots, 12$. Since the $\mathbf{m}_\mu$ are doubly stochastic and irreducible, so are the yearly transition rate matrices $\mathbf{M}_\mu$.

In general, the twelve yearly transition matrices $\mathbf{M}_\mu$ will be different, however, it is easy to show that they have the same eigenvalues $\{\rho_n \,|\, 0 \le n \le (C-1)\}$ and that their eigenvectors are simply related. This can be seen as follows: multiply the eigenvalue equation

$$\mathbf{M}_1 \cdot \boldsymbol{p}_n = \rho_n \boldsymbol{p}_n$$

on the left by the transition matrix m_1 and obtain

$$\mathbf{M}_2 \cdot \mathbf{m}_1 \cdot \boldsymbol{p}_n = \rho_n \mathbf{m}_1 \cdot \boldsymbol{p}_n$$
$$\text{with} \qquad \mathbf{M}_2 = \mathbf{m}_1 \cdot \mathbf{m}_{12} \cdot \cdots \cdot \mathbf{m}_2.$$

In other words, $\mathbf{m}_1 \cdot \boldsymbol{p}_n$ is an eigenvector of $\mathbf{M}_2$ corresponding to the same eigenvalue ρ_n. Similarly, $\mathbf{m}_2 \cdot \mathbf{m}_1 \cdot \boldsymbol{p}_n$ is the $\mathbf{M}_3$-eigenvector corresponding to ρ_n.

Being doubly stochastic, all the yearly transition matrices $\mathbf{M}_\mu$ have the same Perron vector $[1, 1, \ldots, 1]$ with eigenvalue $\rho_0 = 1$. However, recall that the partitions associated with each month are different; consequently, the Perron vector in, e.g. January describes a different probability distribution than the, e.g. February Perron vector. In fact, these twelve Perron vectors describe the stationary pdf's of each month.

The analogy with the classical Floquet theory of ordinary differential equations (1883), see, e.g. Cronin (1980), becomes evident by introducing a set of matrices $\mathbf{Y}_\mu$, the analogue of Floquet's fundamental matrix solution, as

$$\mathbf{Y}_\mu := [\boldsymbol{p}_0^\mu, \boldsymbol{p}_1^\mu, \ldots, \boldsymbol{p}_{C-1}^\mu], \ \mu = 1, \ldots, 12,$$

where the columns $\boldsymbol{p}_n^\mu$ are the eigenvectors of the transition matrix from month μ to the same month one year later, i.e.

$$\mathbf{M}_\mu \cdot \boldsymbol{p}_n^\mu = \rho_n \boldsymbol{p}_n^\mu \text{ with } n = 0, 1, \ldots, C-1,$$

and the e.g. eigenvalues are, e.g. ordered according to their absolute values, $|\rho_n| \ge |\rho_{n+1}|$ with $\rho_0 = 1$. Our index μ corresponds in the Floquet approach to the time modulo the period, in our case, modulo twelve months. Then one has that

$$\mathbf{Y}_{\mu+12} = \mathbf{Y}_\mu \cdot \Phi$$

where Φ, the analogue of Floquet's monodromy matrix, is the diagonal matrix $[\rho_0, \rho_1, \ldots, \rho_{C-1}]$ and the time-index μ is shifted by twelve months, i.e. by one complete cycle.

In the Floquet theory one introduces the monodromy matrix in order to show that it is time-independent; we have already proved that, i.e. that the Φ as defined above is, indeed, independent of μ.

2.5 Reversible and irreversible dynamics

Let us split a doubly stochastic transition matrix $\mathbf{M}$ into its symmetric and antisymmetric parts,

$$M_{ik} = S_{ik} + A_{ik} \tag{10}$$

$$S_{ik} = \frac{1}{2}(M_{ik} + M_{ki}) \quad A_{ik} = \frac{1}{2}(M_{ik} - M_{ki}). \tag{11}$$

Due to the doubly stochastic character of $\mathbf{M}$, one has

$$0 \le S_{ik} = S_{ki} \le 1, \tag{12}$$
$$\sum_i S_{ik} = 1,$$
$$\sum_k S_{ik} = 1.$$

In words: the symmetric part of a doubly stochastic matrix is also doubly stochastic. The dynamics generated purely by such a doubly stochastic, symmetric $\mathbf{S}$ would consist of decaying, non-oscillating modes, we say that such an $\mathbf{S}$ is purely dissipative or diffusive. In mechanics, such a dynamics is called irreversible. The matrix elements of the antisymmetric part $\mathbf{A}$ satisfy

$$-\frac{1}{2} \le A_{ik} \le \frac{1}{2}, \tag{13}$$
$$\sum_i A_{ik} = 0, \tag{14}$$
$$\sum_k A_{ik} = 0, \tag{15}$$
$$|A_{ik}| \le S_{ik}. \tag{16}$$

The dynamics generated purely by the anti-symmetric part $\mathbf{A}$ consists of non-decaying, oscillating modes, i.e. $\mathbf{A}$ alone would generate a non-dissipative (or conservative) time evolution. In mechanics, such a dynamics is called reversible. In the next paragraphs we will discover more reasons for associating $\mathbf{A}$ with the conservative part of the dynamics and $\mathbf{S}$ with the diffusive part.

The complete dynamics, i.e. the one generated by $\mathbf{M} = \mathbf{S} + \mathbf{A}$, cannot be obtained by a simple juxtaposition or factorization of the dynamics generated by these two components separately because, in general, $\mathbf{SA} \neq \mathbf{AS}$, consequently, the eigenvectors of these matrices will differ.

Let us express the average information loss in terms of the symmetric and anti-symmetric components,

$$\langle \Delta I \rangle = -\frac{1}{C} \sum_{k=1}^{C} \sum_{i=1}^{C} [S_{ik} + A_{ik}] \ln [S_{ik} + A_{ik}]$$
$$= -\frac{1}{C} \sum_{k=1}^{C} \sum_{i=1}^{C} S_{ik} \ln S_{ik} \left[1 + \frac{A_{ik}}{S_{ik}}\right]$$

$$
\begin{aligned}
&-\frac{1}{C}\sum_{k=1}^{C}\sum_{i=1}^{C} A_{ik}\ln\left[S_{ik}+A_{ik}\right] \\
&=-\frac{1}{C}\sum_{k=1}^{C}\sum_{i=1}^{C} S_{ik}\ln S_{ik} \\
&\quad-\frac{1}{C}\sum_{k=1}^{C}\sum_{i=1}^{C} S_{ik}\ln\left[1+\frac{A_{ik}}{S_{ik}}\right] \\
&\quad-\frac{1}{C}\sum_{k=1}^{C}\sum_{i=1}^{C} A_{ik}\ln\left[S_{ik}+A_{ik}\right]. \qquad (17)
\end{aligned}
$$

Therefore, we can write $\langle\Delta I\rangle = \langle\Delta I\rangle_S + \langle\Delta I\rangle_{A/S}$ where $\langle\Delta I\rangle_S$ is the information loss that would be generated by the doubly stochastic matrix S exclusively, namely

$$
\langle\Delta I\rangle_S = -\frac{1}{C}\sum_{k=1}^{C}\sum_{i=1}^{C} S_{ik}\ln S_{ik} \geq 0,
$$

while $\langle\Delta I\rangle_{A/S}$ depends both on $\mathbf{S}$ and $\mathbf{A}$,

$$
\begin{aligned}
\langle\Delta I\rangle_{A/S} &= -\frac{1}{C}\sum_{k=1}^{C}\sum_{i=1}^{C} S_{ik}\ln\left[1+\frac{A_{ik}}{S_{ik}}\right] \\
&\quad-\frac{1}{C}\sum_{k=1}^{C}\sum_{i=1}^{C} A_{ik}\ln\left[S_{ik}+A_{ik}\right], \qquad (18)
\end{aligned}
$$

and it vanishes for $\mathbf{A}=0$. Taking into account the antisymmetry of $\mathbf{A}$, one has

$$
\begin{aligned}
\langle\Delta I\rangle_{A/S} &= -\frac{1}{C}\sum_{k=1}^{C}\sum_{i<k}^{C} S_{ik}\ln\left[1-\left(\frac{A_{ik}}{S_{ik}}\right)^2\right] \\
&\quad-\frac{1}{C}\sum_{k=1}^{C}\sum_{i<k}^{C} A_{ik}\ln\frac{S_{ik}+A_{ik}}{S_{ik}-A_{ik}}. \qquad (19)
\end{aligned}
$$

Since $|A_{ik}| \leq S_{ik}$, one sees that each term under the first summation sign is negative or zero and that each term under the second summation sign is positive or zero. Moreover, one can check that $a\ln\left[(s+a)/(s-a)\right] \geq -s\ln\left[1-(a/s)^2\right]$ for all $0 \leq s \leq 1$ and $|a| \leq s$. Therefore, $\langle\Delta I\rangle_{A/S} \leq 0$, i.e. a non-vanishing anti-symmetric component $\mathbf{A}$ implies that the entropy production of the doubly stochastic matrix $(\mathbf{S}+\mathbf{A})$ is smaller than that of the doubly stochastic matrix $\mathbf{S}$. The most extreme example of this reduction is provided by the permutations matrices: since $\langle\Delta I\rangle^{\text{permutation}} = 0$ one has that $\langle\Delta I\rangle^{\text{permutation}}_{A/S} = -\langle\Delta I\rangle^{\text{permutation}}_{S} < 0$.

For all these reasons it is natural to associate the anti-symmetric part $\mathbf{A}$ with the conservative part of the dynamics. On the other hand, the symmetric part $\mathbf{S}$ stems not only from the diffusive part but also from the conservative part of the dynamics, as can be clearly seen in the case of a permutation matrix: a permutation matrix is a purely conservative dynamics and yet it has a nonvanishing symmetric part $\mathbf{S}$. One is led to define the purely diffusive components of $\mathbf{M}$ as $S_{kl}-|A_{kl}|$ and to measure the overall purely diffusive character of $\mathbf{M}$ by $\left(1-C^{-1}\sum_{k,l}|A_{kl}|\right)$.

3 Predictability of an intermediate ENSO model

In this section we apply the Markov chain concepts sketched above to data generated by the Zebiak and Cane (ZC) ENSO model (Zebiak and Cane, 1987). The ZC model is a coupled atmosphere-ocean model for the tropical region. The atmospheric component consists of a Gill-type, steady-state, linear shallow water model (Gill, 1980) which is formulated on an equatorial beta plane. Dissipation is parameterized in terms of linear Newtonian cooling and Rayleigh friction. Furthermore, a surface-wind parameterization of low-level moisture convergence is used. This model simulates reasonably well the steady state atmospheric response to typical sea surface temperature anomalies (SSTA) in the tropics. The ocean model is formulated for a rectangular tropical ocean basin. It is based on a linear, reduced gravity model, including a 50 m deep frictional layer, which accounts for surface intensification of wind-driven currents. The thermodynamic core of this ocean model takes into account three-dimensional temperature advection by mean and anomalous ocean currents, a linear dependence between surface heat flux anomalies and SSTA and the asymmetric effect of vertical advection on temperature. Subsurface temperature anomalies are diagnosed from the variations of the model's upper layer thickness. The seasonal background fields of surface winds and wind divergence, as well as of sea surface temperature are prescribed.

The Zebiak and Cane model generates chaotic ENSO oscillations in the standard parameter set. As to whether the observations are more adequately described in terms of a stochastically excited damped oscillator (Penland and Sardeshmukh, 1995) or of a chaotic oscillation is an interesting issue that has not been solved yet. We do not intend to dwell into the details of this interesting controversy. Small changes in the standard parameters of the ZC model can lead to stable ENSO oscillations, for which noise becomes crucial., This sensitivity should be kept in mind when interpreting the results of the ZC model. Very often, however, the possibility is ignored that ENSO can be self-sustained during some decades, whereas it might be stable during other decades. In fact there is observational evidence (An and Jin, 2000) that interdecadal background changes in the tropical Pacific can trigger changes in the growth rate of ENSO, strong enough to cross the Hopf bifurcation point.

We performed a 640 year long simulation with the Zebiak and Cane ENSO model using the standard parameter configuration; we chose to limit the length of the data set to 640 years because it would be very difficult to generate longer time series from more complex GCMs . Our analysis makes use of the simulated sea surface temperature anomalies as well as of the thermoclince depth anomalies. In particular, we compute the popular Niño 3 SSTA index (Zebiak and Cane, 1987) which characterizes ENSO effectively. This index is defined as the SSTA averaged over the region 5° S–5° N, 150° W–90° W. For our bivariate analysis we also use the thermocline depth anomalies averaged over the region 10° S–10° N, 120° W–180° W.

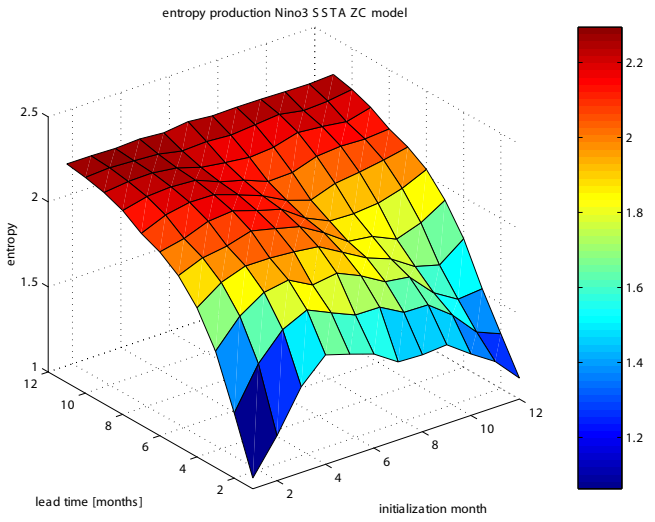


Fig. 2. Seasonal dependence of the entropy production: Entropy as a function of lead time and of the initialization month.

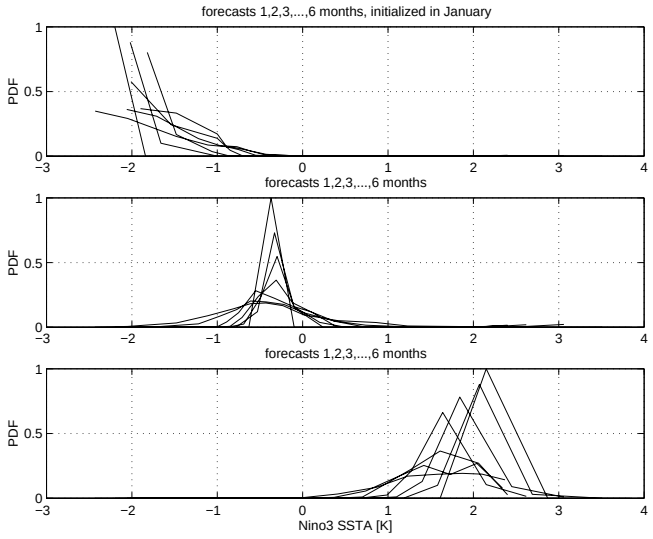


Fig. 3. Initial-state dependence of dispersion in probability space; the initialization month is January.

It is well-known (Balmaseda et al., 1995, Kumar and Hoerling, 1998) that ENSO forecast skills depend on the initial state and in particular, on the season of the year. The so-called predictability barrier characterizes the fact that ENSO predictions initialized in the boreal fall have a significantly better skill than those initialized in the boreal spring[8]. This effect might be due to the seasonal changes of ENSO instability related to changing ocean stratification and atmosphere-ocean interactions. As to whether the Zebiak Cane model is a "realistic" model with respect to simulating the seasonality of the atmosphere-ocean interactions shall not be discussed here[9].

We studied this phenomenon by looking at the information loss as a function of lead time and initialization month[10].

[8]Recently Torrence and Webster (2000) introduced the term persistence barrier in order account for the fact that the drop of the autocorrelation observed during spring (Balmaseda et al., 1995) is not necessarily associated with a complete lack of predictability. Dynamical ENSO prediction models, however, show that ENSO forecast skills are strongly affected by this barrier. In the following, we will thus, use the old term predictability barrier.

[9]As for the role of the ZC-model, the ZC-model was chosen not because we consider it to be a perfect model for ENSO but just as an illustrative example. The cyclic Markov chain concept can be applied to any kind of ENSO model that yields either stochastically excited linear oscillations or self-sustained chaotic oscillations.

[10]The transition matrices for a ν-month lead forecast starting from month μ can be computed as $\mathbf{m}(\nu+\mu) = \mathbf{m}(\nu+\mu-1)\cdot\cdot\mathbf{m}(\mu)$. However, such a matrix multiplication might increase the rounding errors and this should be avoided in order to ensure that Eq. (5) is satisfied. Hence, we compute the corresponding transition matrix elements, called them m_{ab}, directly by counting how many data points in month μ located in cell a end up in cell b ν months later.

3.1 Univariate analysis

Let us first consider an univariate analysis using only the El Niño 3 SSTA index. The computed information loss based on the 640-year long time series for different initialization months and different lead prediction times is shown in Fig. 2. The partition used here is $D = 640$, $C = 16$, $d = 40$ which ensures the accessibility conditions $d \geq C$ and $D \geq C^2$.

One observes that the entropy production quantifying the rate of spread of the probability density is modulated by the annual cycle. More precisely: there is a larger entropy production from March and April and a lower entropy production for late summer to autumn initialization months; this is the so-called ENSO predictability barrier. One observes furthermore, that for forecasts initialized in, e.g. August (month 8) the entropy production levels off at a lead time of about six to seven months whereas it increases again for longer lead times. These results are not new but they do illustrate the utility of cyclic Markov chains in order to quantify predictability.

Notice that in this approach information about the predictability of the system is extracted from one, long, time series and not from a number of simulations initialized on different points of the attractor, as is done in ensemble-forecast studies.

In addition to the seasonal cycle effect, the predictability of ENSO might depend also on the state of the tropical Pacific itself. This implies that an El Niño forecast starting, e.g. from an El Niño state might have a different quality than a forecast initialized during an intermediate state.

Figure 3 shows three Markov chain probability forecasts, all of them starting in January but from different initial conditions. The initial probability density functions (pdf) were chosen to be $p_0^i = \delta_j^i$, with j=1, 7 and 14, respectively. The first one corresponds to strong La Niña conditions, the last

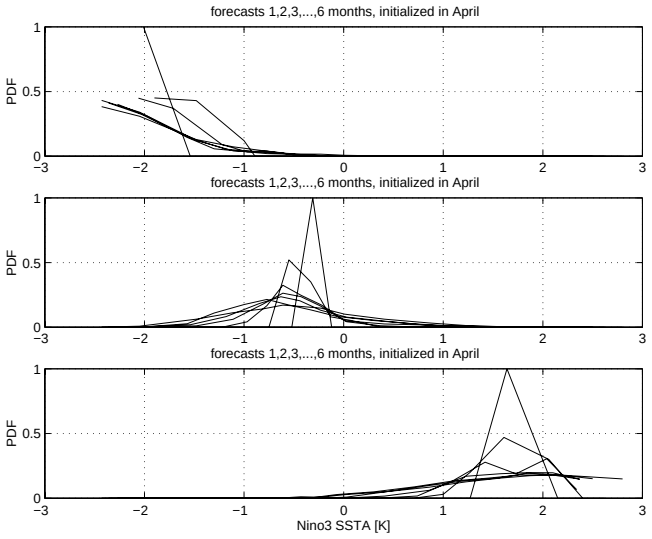


Fig. 4. Initial-state dependence of dispersion in probability space; the initialization month is April.

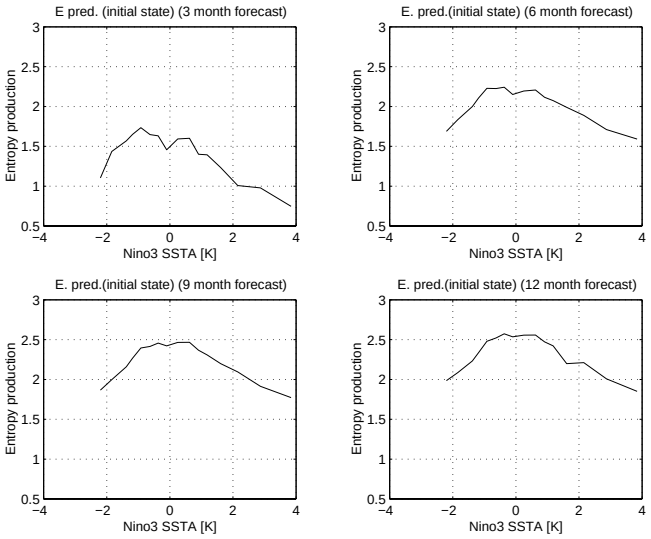


Fig. 5. Entropy as a function of the January initial state for forecasting lengths of 3, 6, 9 and 12 months.

one to strong El Niño conditions. The time evolution for these initial pdfs is shown for lead times from one to six months.

One observes that, starting from strong La Niña conditions, upper panel of Fig. 3 the system remains in the La Niña state for about six months. The same holds for strong El Niño conditions, see lower panel of Fig. 3. In contrast to these two large anomaly cases, neutral conditions are much less predictable as can be seen from the fast spread of the pdfs in the middle panel of this Figure. Probably this is a manifestation of the fact that under neutral conditions an oscillation exhibits a velocity maximum in state space. This could imply that neutral ENSO conditions in January are much more unstable than extreme ENSO conditions.

In Fig. 4 we show the results obtained from the same type of analysis but done now for highly localized initial conditions starting in April. We observe qualitatively the same features as for the January initialization case. However, the dispersion rate of the pdfs is much stronger during the boreal spring than during the boreal winter. The main quantitative differences with respect to Fig. 3 can be seen for leads ranging from one to three months.

Had we chosen a partition of the state-space into equidistant intervals of the index, then the transition rates from and to the extreme-value cells would have been estimated with large errors. Our partition into equal-weight cells has the advantage that the transition rates for the extreme values are estimated as precisely as those for the neutral conditions. On the other hand, a drawback of this partition may be the reduced physical resolution for extreme ENSO events.

The state dependence of ENSO predictability is summarized in Fig. 5. It displays the information loss associated with the El Niño 3 SSTA index for different forecasting lengths. The January initial states are chosen as $p_0^i = \delta_j^i$, j=1,..., 16.

In agreement with the previous findings, one observes saturation of the entropy production at around a six-month lead and that predictions started from neutral ENSO conditions are less predictable than those initialized during large ENSO events. For a given lead time, we shall call Most Unpredictable Modes (MUMs) those initial states that generate the largest information loss and Least Unpredictable Modes (LUMs) those initial states that generate the smallest information loss. In our analysis we obtain MUMs and LUMS for each month of the year, separately.

In closing this Subsection, let us analyze the eigenmodes of the year-to-year transition matrices $\mathbf{M}_\mu = \Pi_{i=0}^{11}\mathbf{m}_{\mu+i}$. As shown in previous sections, all the $\mathbf{M}_\mu$ have a Perron eigenvector $\boldsymbol{p}_0^\mu = (1/C, \ldots, 1/C)$, such that $\mathbf{M}_\mu \cdot \boldsymbol{p}_0^\mu = \boldsymbol{p}_0^\mu$. These twelve Perron vectors are nothing else but the average, stationary distributions corresponding to each month. Recalling that each month has its own partition of state space, one realizes that, indeed, the twelve Perron vectors describe different distributions in state space. It is only after a complete twelve-month cycle that we recover the same state space distribution. The distributions corresponding to all twelve Perron vectors have the maximum possible entropy for a partition with C cells, namely, their entropy is $\ln C$. Accordingly, there is no information loss when one starts from this distribution. Another way of stating this is that the Perron eigenvector has an infinite lifetime. The time evolution of the Perron vectors starting from the initial month of January is shown in Fig. 6.

We observe that in the first stages of the time evolution the eigenvector of $\mathbf{M}$ is rotated away from the El Niño state, whereas the pdf does not change significantly for SSTA values smaller than 2 K. The eigenvector turns back to its original position after about 7 months.

The other eigenvectors have eigenvalues whose absolute value is smaller than 1, i.e. $|\rho_n| < 1$ for $n > 0$. This means

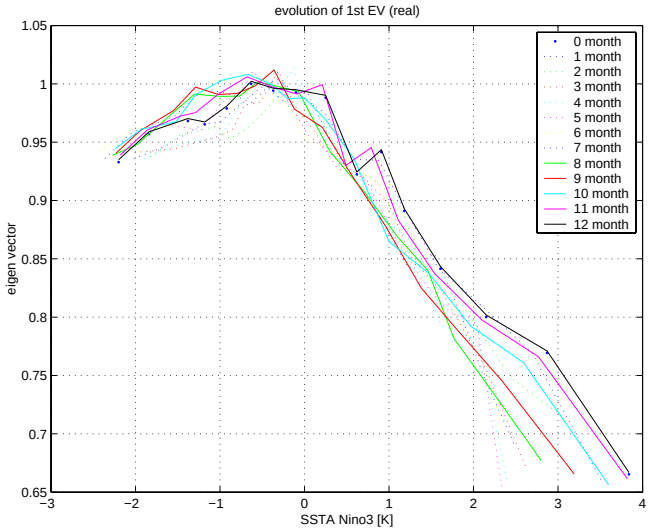


Fig. 6. Time evolution of the leading eigenmode of the yearly transition matrix M during the course of a year.

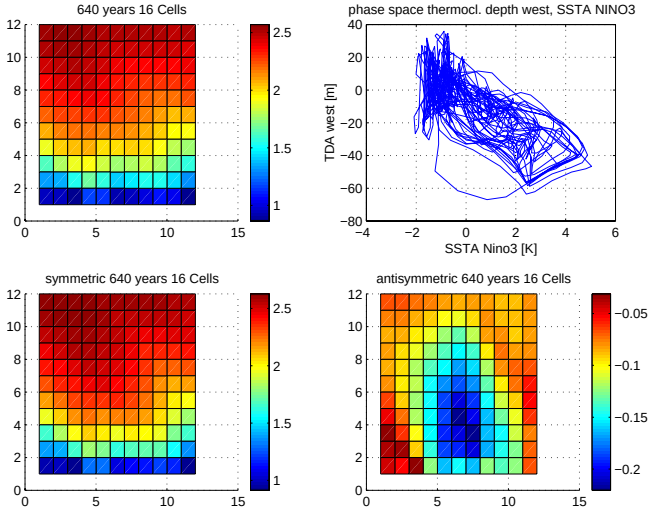


Fig. 7. Entropy as a function of lead time and initialization month as computed for the total (upper left panel), symmetric (lower left panel) and non-dissipative dynamics (lower right panel). A state space view of the SSTA-TDA- trajectory is displayed in the upper right panel.

that their lifetimes are finite, i.e. that they decay in time. There is another essential difference with respect to the Perron vectors, namely, the other eigenvectors are not positive. This reflects the fact that these eigenvectors are not probability distributions but only departures of probability distributions with respect to the steady distribution, i.e. from the corresponding Perron vector. In meteorological parlance, they are anomalies. Since the matrices $\mathbf{M}_\mu$ are real, complex eigenvalues and their corresponding complex eigenvectors appear always in conjugate pairs; in such a case, the corresponding departures from the Perron vector do not decay monotically but decay as damped oscillations. There is a connection between these eigenvalues and the periods and lifetimes of the coarse-grained unstable periodic orbits, see, e.g. Cvitanović et al. (1991).

3.2 Bivariate analysis

Next, we present a bivariate study, i.e. Markov chains which are derived from two ENSO-characterizing variables. In the first example, we compute the information loss based on the El Niño 3 SSTA index as generated by the ZC model and the west equatorial thermocline depth anomalies, averaged over the 10° S–10° N, 120° W–180° W for different initialization months and different lead prediction times. We use Eq. (9) and the cyclostationary transition matrices computed according to the scheme shown in the lower panel of Fig. 1. The partition we use is characterized by $D = 640, C = 16, d = 40$ which ensures the accessibility conditions $d \geq C$ and $D \geq C^2$. The results are displayed in Fig. 7 where the entropy production has been split into three parts.

– The entropy production $\langle \Delta I \rangle$ based on the full dynamics is shown in Fig. 7 upper left panel. Qualitatively and quantitatively we observe very similar features as in the univariate case displayed in Fig. 2. The spring barrier is retrieved as well as the slowing down of entropy production after about six months.

– The entropy production $\langle \Delta I \rangle_S$ due only to the symmetric part of the transition rate matrix is shown in Fig. 7 lower left panel. It can be seen that the total entropy production is mainly dominated by the symmetric part.

– The non-dissipative part $\langle \Delta I \rangle_{A/S}$ which is related to the anti-symmetric part of the transition rate matrix explains only a small part of the information loss. However, a strongly pronounced seasonal modulation of information gain becomes apparent with strongest information gain to be seen in the boreal summer season.

In Fig. 8 we show the one-month lead transition matrix starting in June, its eigenvectors, eigenvalues, and its symmetric and anti-symmetric parts. The eigenvectors are shown in cell-number space. As one can see, also in this case the total transition matrix is dominated by the symmetric part. Hence, even for a one month forecast the dynamics is largely dissipative. The eigenvalue spectrum (Fig. 8f) shows several oscillating modes with frequencies comparable to their decay rates. How these oscillatory modes can be interpreted physically is a question that requires further work. A first step would consist in studying whether they remain unchanged or not when one increases the number of state space cells C. As it is evident, when $C = 16$ the longest period that one can detect is 16 months. Similarly, more research should lead to a better understanding of the information contained, e.g. in the asymmetric part of the transition matrices.

In Fig. 9 we present the same results for a three month lead forecast. If this and the previous matrix would contain

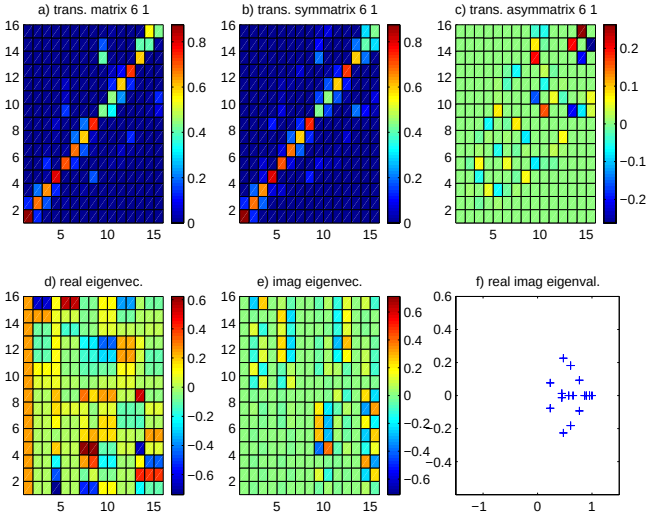


Fig. 8. (a) one month forecast transition matrix starting in June, **(b)** its symmetric part, **(c)** its anti-symmetric part, **(d)** real part of its eigenvectors, **(e)** imaginary part of its eigenvectors, **(f)** its spectrum.

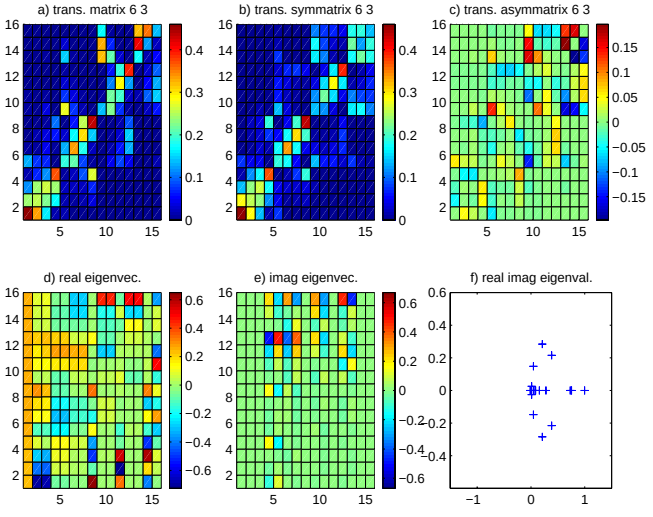
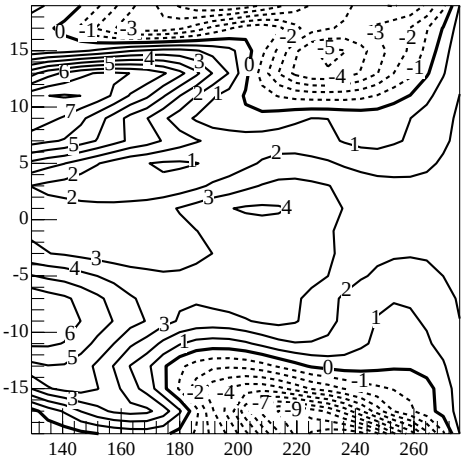


Fig. 9. (a) transition matrix for a three month forecast initialized in June, **(b)** its symmetric part, **(c)** its anti-symmetric part, **(d)** real part of its eigenvectors , **(e)** imaginary part of its eigenvectors, (f) its spectrum.

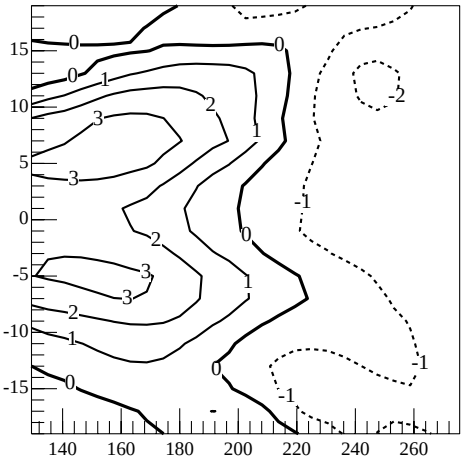
exactly the same modes, then their spectra would be simply related to each other: the damping factor and the phase of the eigenvalues of the second matrix would be the cubic power and three times those of the one month lead matrix, respectively. The associated eigenvectors would show the evolution in state space of the corresponding anomaly. As one can see, the spectra of the two matrices are not so simply related. This is not surprising since there may be more than sixteen modes that dominate the evolution at different phases of the yearly cycle. By increasing the number of cells[11] one may detect more dynamically relevant modes, for example, those

[11]This will require a larger set of data.

MUM ZC model, 7 month forecast, initialized in January



MUM ZC model, 7 month forecast, initialized in May



MUM ZC model, 7 month forecast, initialized in September

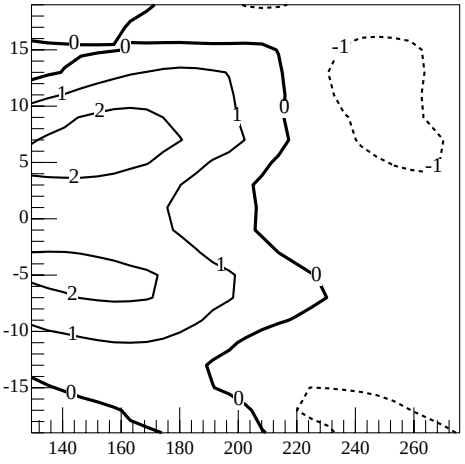


Fig. 10. Most unpredictable thermocline patterns (y-axis represent the latitude ranging from 20° S to 20° N, x-axis represents longitude ranging from 130° E to 85° W) for a 7 month forecast initialized in January, May and September, respectively.

with periods longer than 16×3 months $= 48$ months. Moreover, as discussed in Sect. 2.2, the transition matrices can be expanded in terms of cyclic permutations that correspond to the coarse-grained unstable periodic orbits of dynamical systems' theory, Cvitanović et al. (1991). Such a decomposition in terms of UPOs may, in fact, be more revealing than a decomposition in terms of eigenvectors. Again, by enlarging the number of cells, it should be possible to detect more UPOs and to disentangle them better. In this respect, it should be noticed that probably the best way to detect and to disentangle UPOs is to work with three or more dynamical variables, as it is done in Tziperman et al. (1997).

3.3 Most Unpredictable Modes (MUM's)

It has become quite popular to study similar aspects of error growth in ENSO forecasts (Chen et al., 1995b, Xue et al., 1994, Moore and Kleeman, 1997; Eckert, 1997) due to initial state errors within the linear framework of singular vectors (Lorenz, 1965). These vectors are associated with the fastest linear growth of the initial perturbations due to the non-normality of the tangent linear propagator (Trefethen et al., 1993). They can be computed from the integral propagator of the linearized model. In the case of reduced complexity, coupled, atmosphere-ocean models (Chen et al., 1995b; Xue et al., 1994; Moore and Kleeman, 1997; Eckert, 1997) these linearized models and their adjoint can be obtained without facing fundamental difficulties. However, the determination of singular vectors associated with ENSO using a comprehensive CGCM is highly non-trivial and has not been achieved yet. The reason is not only computer power but also the fundamental difference in atmospheric and oceanic time scales. As shown above, the Markov chain approach allows for the extraction of those initial states (or patterns) which are associated with the highest (or the lowest) predictability without any linearization whatsoever. In order to illustrate the type of possible application we have in mind, we have decomposed the simulated thermocline depth field of the ZC model into Empirical Orthogonal functions (EOFs). The two leading EOFs explain about 80% of the variance. These large values of the explained variance are typical for 1 1/2 layer models. We constructed then a Markov chain based on these two leading principal components and searched for those localized initial states $p_0^j = \delta_i^j$ which are characterized by the largest entropy production for a given forecast length. This state is then transformed from cell space back into physical space giving a pair of values for the two principal components of thermocline depth anomaly which is associated with the largest rate of information loss. Finally, these values of the two principal components are multiplied with their respective EOF patterns such as to give an impression of what the initial state which leads to the lowest predictability looks like.

We apply this to three seven-month forecasts, initialized in January, May and September, respectively. The resulting MUM's are displayed in Fig. 10. We see that the most unpredictable mode changes during the course of the year. The MUM computed for a seven-month forecast initialized in January characterizes the discharging of the warm pool. Warm thermocline waters swash the equatorial eastern Pacific. This situation is very similar to a situation shortly before a major El Niño event is developed and downwelling Kelvin waves propagate from east to west. The most unpredictable mode for a September forecast has a similar pattern to the one initialized during May. MUMs for May and September are characterized by a deep warm pool thermocline and a shallow eastern equatorial Pacific. This pattern is dominated mainly by the leading EOF. The MUM for May corresponds to the situation approximately 7 months before a major El Niño event. The warm pool is anomalously deep and short perturbations can lead to the initiation of downwelling Kelvin waves which need about two to three months to cross the Pacific. It is worthwhile noticing that the MUMs computed for different initialization months have very similar structures to the singular vectors computed for the ZC model, see, e.g. Chen et al. (1995b).

4 Summary and discussion

Our main objective has been to illustrate the capacity and utility of the Markov chain approach in a geophysical context, more specifically, in characterizing ENSO and its predictability. To this end, we presented the formalism for cyclostationary Markov chains and introduced an efficient way of partitioning the state space that also leads directly to an interpretation of the dynamics in terms of coarse-grained unstable periodic orbits. One of the most attractive aspects of this approach is that it does not require a linear approximation of the dynamics; this stands in contraposition to the singular vector approach. Another attractive aspect is that the analysis is based on one long, experimental or numerically generated, data series; this should be compared with ensemble forecasting which requires numerous numerical simulations. Among the straightforward applications presented in Sect. 3, is the identification of the physical configuration that leads to the largest (alternatively, the smallest) uncertainty in predictions with a predefined forecast lead; these states are identified separately for each month. A more standard application is the detection of the spring predictability barrier and, less standard, its quantification. It should be noted that the partition into equal weight cells implies that the cells corresponding to the most unprobable observations[12] occupy a relatively large interval of the variable's physical range. Sometimes, this may be undesirable.

Needless to say, in order to be able to predict, a good knowledge of the system's past behaviour is required; this is an inescapable fact that all practical forecasting techniques have to face. In the context of the present article, this means that 1) sufficiently long data series are required in order to be able to make accurate and interesting predictions and 2) the

[12] These are usually the most extreme values of the dynamical variables under consideration.

records should be of more than one relevant dynamical variable. Moreover, the choice of these variables is crucial, e.g. if two variables are strongly correlated then one of them is nearly redundant and should be ignored. These two requirements have been satisfied in, e.g. Tziperman et al. (1997).

Another important question that must be addressed is the so-called *Markovian assumption* that lays behind the Markov chain approach. The Markovian assumption or Markovian approximation consists in assuming that $p(\boldsymbol{x}, t+1)$, the probability at time step $t+1$, is completely determined by $p(\boldsymbol{x}, t)$, the probability at the previous time step t. Notice that this is more general than a Markov chain, i.e. the 'Markovian models' often found in the literature are not necessarily Markov chains. In practice, the Markovian assumption is often violated because the set of variables $\boldsymbol{x}$ that one uses is usually (much) smaller than the total number of relevant variables. Whether the Markovian assumption holds or not can be checked a posteriori, e.g. by computing correlation functions and comparing them with the corresponding correlation functions in the original data. In many cases one will find that the real system shows stronger, longer lived correlation functions than the Markov chain model does. In principle, the violations of the Markovian approximation can be eliminated by refining the partition and/or by enlarging the set $\boldsymbol{x}$ of dynamical variables, e.g. by including past values of the dynamical variables, etc. In practice, this is hardly feasible.

These comments on the validity of the Markovian approximation have important implications for some of the results we have obtained. In particular, it should be clear that some results are no more than *lower bounds* to predictability. In its turn, this has implications for the development of sophisticated physical models: the effort needed in order to develop physical, accurate dynamical models can be justified only if the predictions obtained from such models have a lower uncertainty than the ones obtained from Markov chain models like the one presented in this article.

Our future work will be devoted to the application of cyclic Markov chains to ENSO data obtained from long CGCM simulations. It is our goal to identify those patterns which are associated with least and highest predictability. Furthermore, we plan to study the time evolution of CGCM ENSO predictability more in detail. This can be achieved by projecting the physical fields onto the MUMs and LUMs. The resulting timeseries quantify how simulated ENSO predictability is modulated in time. Maybe it is possible to attribute physical meaning to these predictability timeseries, such as establishing that, e.g. if the thermocline is anomalously deep for several decades then we might expect better intrinsic predictability of the ENSO system as compared to eras with relatively shallow thermocline. This will eventually lead to the generation of probabilistic physical ENSO models.

Acknowledgements. RAP thanks A. Knauf (Erlangen University) and H. Voss (Freiburg University) for very stimulating conversations. This work was partly sponsored by the German Science Foundation (DFG), partly by the EU project SINTEX ENV4-CT98–0714.

References

An, S.-I. and Jin, F.-F.: An eigenanalysis of the interdecadal changes in the structure and frequency of ENSO mode, Geophys. Res. Lett., 27, 2573–2576, 2000.

Balmaseda, M. A., Anderson, D. L. T., and Davey, M. K.: ENSO prediction using a dynamical ocean model coupled to statistical atmospheres, Tellus, 46 A, 497–511, 1994.

Balmaseda, M. A., Davey, M. K., and Anderson, D. L. T.: Decadal and seasonal dependence of ENSO prediction skill, J. Climate, 8, 2705–2715, 1995.

Barnston, A. G. and Ropelewski, C. F.: Prediction of ENSO episodes using canonical correlation analysis, J. Climate, 5, 1316–1345, 1992.

Berman, A. and Plemmons, R. J.: Nonnegative matrices in the mathematical sciences, Academic Press, 1979.

Birkhoff, G.: Tres observaciones sobre el algebra lineal, Rev. Univ. Nac. Tucumán, Ser. A, 5, 147–150, 1946 .

Brualdi, R. A., Parter, S. V., and Schneider, H.: The diagonal equivalence of a nonnegative matrix to a stochastic matrix., J. Math. Anal., Appl., 16, 31–50, 1966.

Cane, M. A., Zebiak, S. E., and Dolan, S. C.: Experimental forecasts of El Niño, Nature, 321, 827–832, 1986.

Cencini, M., Lacorata, G., Vulpiani, A., and Zambianchi, E.: Mixing in a meandering jet: A Markovian Approximation, J. Phys. Oceanogr., 29, 2578–2593, 1999.

Chen, D., Zebiak, S. E., Busalacchi, A. J., and Cane, M. A.: An improved procedure for El Niño forecasting, Science, 269, 1699–1702, 1995a.

Chen, Y-Q., Battisti, D. S., Palmer, T. N., Barsugli, J., and Sarachik, E. S.: A study of the predictability oftropical Pacific SST in a coupled atmosphere/ocean model using singular vector analysis: the role of the annual cycle and the ENSO cycle, Mon. Wea. Rev., 125, 831–845, 1995b.

Cronin, J.: Differential equations, Marcel Dekker, New York, 1980.

Cvitanović, P.: Periodic orbits as the skeleton of classical and quantum chaos, Physica D, 51, 138, 1991.

Eckert, C.: PhD Dissertation, University of Hamburg, 1997.

Egger, J.: Master equations for climatic parameter sets, submitted, 2000.

Flügel, M. and Chang, P.: Does the predictability of ENSO depend on the seasonal cycle? J. Atmosph. Sciences, 55, 3230–3243, 1998.

Fraedrich, K.: El Niño-Southern Oscillation predictability, Mon. Wea. Rev., 116, 1001–1012, 1988.

Gill. A.: Some simple solutions for heat induced tropical circulation, Q. J. R. Meteor. Soc., 106, 447–462, 1980.

Goswami, B. and Shukla, J.: Predictability of a coupled ocean-atmosphere model, J. Climate, 4, 3–22, 1991a.

Goswami, B. and Shukla, J.: Predictability and variability of a coupled ocean-atmosphere model, J. Marine. Sys., 1, 217–228, 1991b.

Grötzner, A., Latif, M., Timmermann, A., and Voss, R.: Interannual to decadal predictability in a coupled ocean-atmosphere general circulation model, J. Climate, 12, 2607–2624, 1998.

Ji, M., Behringer, D. W., and Leetmaa, A.: An improved coupled model for ENSO prediction and implications for ocean initialization, Part II: The coupled model, Mon. Wea. Rev., 126, 1022–1034, 1998.

Kumar, A. and Hoerling, M. P.: Annual cycle of Pacific-North American seasonal predictability associated with different phases of ENSO, J. Climate, 11, 3295–3308, 1998.

Latif, M., Barnett, T. P., Cane, M. A., Flügel, M., Graham, N. E., von Storch, H., Xu, J.-S., and Zebiak, S. E.: A review of ENSO prediction studies, Climate Dynamics, 9, 167–179, 1994.

Lorenz, E. N.: A study of the predictability of a 28-variable atmospheric model, Tellus, 17, 321–333, 1965.

Mason, S. J., Goddard, L., Graham, N. E., Yulaeva, E., Sun, L., and Arkin, P. A.: The IRI seasonal climate prediction system and the 1997/8 El Niño event, Bull. Am. Met. Soc., 80, 1853–1873, 1999.

Mo, R. and Straus, D. M.: Probability forecasts for seasonal average anomalies based on GCM ensemble means, COLA Technical Report No. 74, 33, 1999.

Moore, A. M. and Kleeman, R.: The singular vectors of a coupled atmosphere-ocean model of ENSO, I: Thermodynamics, energetics and error growth, Quart. J. Roy. Met. Soc., 123, 953–981, 1997a.

Moore, A. M. and Kleeman, R.: The singular vectors of a coupled atmosphere-ocean model of ENSO, II: Sensitivity studies and dynamical interpretation, Quart. J. Roy. Met. Soc., 123, 983–1006, 1997b.

Nicolis, C., Ebeling, W., and Baraldi, C.: Markov processes, dynamic entropies and statistical prediction of mesoscale weather regimes, Tellus, 49 A, 108–118, 1997.

Oberhuber, J. M., Roeckner, E., Christoph, M., Esch, M., and Latif, M.: Predicting the 1997 El Niño Even with a global climate model, Geophys. Res. Lett., 13, 2273–2276, 1998.

Palmer, T. N., Gelaro, R., Barkmeijer, J., and Buizza, R.: Singular vectors, metrics and adaptive observations, J. Atmos. Sci., 55, 633–653, 1998.

Penland, C. and Magorian, T.: Prediction of Niño 3 sea surface temperature using linear inverse modeling, J. Climate, 6, 1067–1076, 1993.

Penland, C. and Sardeshmukh, P. D.: The optimal growth of tropical sea surface temperature anomalies, J. Climate, 8, 1999–2024, 1995.

Sinkhorn, R. and Knopp, P.: Concerning nonnegative matrices and doubly stochastic matrices, Pacific J. Math., 21, 343–348, 1967.

Smith, L. A., Ziehmann, C., and Fraedrich, K.: Uncertainty dynamics and predictability in chaotic systems, Q. J. R. Meteor. Soc., 2855–2886, 1999.

Spekat, A., Heller-Schulze, B., and Lutz, M.: "Grosswetter" circulation analysed by means of Markov chains, Meteorol. Rdsch., 36, 243–248, 1983.

Stockdale, T. N., Anderson, D. L. T., Alves, J. O. S., and Balmaseda, M. A.: Global seasonal rainfall forecasts using a coupled ocean-atmosphere model, Nature, 392, 370–373, 1998.

de Swart, H. E. and Grasman, J.: Effect of stochastic perturbations on a low order spectral model of theatmospheric circulation, Tellus, 39 A, 10–24, 1987.

Tangang, F. T., Hsieh, W. W., and Tang, B.: Forecasting the equatorial Pacific sea surface temperatures byneural network models, Climate Dynamics, 13, 135–147, 1997.

Torrence, C. and Webster, P. J.: The annual cycle of persistence in the El Niño-Southern oscillation, Quart. J. Roy. Meteor. Soc., 124, 1985–2004. 2000.

Trefethen, L. N., Trefethen, A. E, Reddy, S. C., and Driscoll, T. A.: Hydrodynamic stability without eigenvalues, Science, 261, 578–584, 1993.

Tziperman, E., Scher, H., Zebiak, S. E., and Cane, M. A.: Controlling spatiotemporal chaos in a realistic El Niño prediction model, Phys. Rev. Lett., 79, 1034–1038, 1997.

van den Dool, H. M. and Barnston, A. G.: Forecasts of global sea surface temperature out to a year using the constructed analogue method, Proceedings of the 19th Annual Climate Diagnostics Workshop, Nov. 14–18, 1994, College Park, Maryland, 416–419, 1995.

Vautard, R., Mo, K., and Ghil, M.: Statistical significance test for transition matrices of atmospheric Markov chains, J. Atmosph. Sc., 47, 1926–1931, 1990.

Xue, Y., Cane, M. A., Zebiak, S. E., and Blumenthal, M. B.: On the prediction of ENSO: a study with a low-order Markov model, Tellus, 46 A, 512–528, 1994.

Zebiak, S. E. and Cane, M. A.: A model El Niño-Southern Oscillation, Monthly Weather Review, 115, 2262–2278, 1987.