



**HAL**  
open science

## Long-term predictability of mean daily temperature data

W. von Bloh, M. C. Romano, M. Thiel

► **To cite this version:**

W. von Bloh, M. C. Romano, M. Thiel. Long-term predictability of mean daily temperature data. *Nonlinear Processes in Geophysics*, 2005, 12 (4), pp.471-479. hal-00302597

**HAL Id: hal-00302597**

**<https://hal.science/hal-00302597>**

Submitted on 18 Jun 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Long-term predictability of mean daily temperature data

W. von Bloh<sup>1</sup>, M. C. Romano<sup>2</sup>, and M. Thiel<sup>3</sup>

<sup>1</sup>Potsdam Institute for Climate Impact Research, PO Box 60 12 03, 14412 Potsdam, Germany

<sup>2,3</sup>Institut für Physik, Universität Potsdam, Am Neuen Palais 10, 14469 Potsdam

Received: 30 September 2004 – Revised: 8 April 2005 – Accepted: 11 April 2005 – Published: 13 May 2005

Part of Special Issue “Nonlinear analysis of multivariate geoscientific data – advanced methods, theory and application”

**Abstract.** We quantify the long-term predictability of global mean daily temperature data by means of the Rényi entropy of second order  $K_2$ . We are interested in the yearly amplitude fluctuations of the temperature. Hence, the data are low-pass filtered. The obtained oscillatory signal has a more or less constant frequency, depending on the geographical coordinates, but its amplitude fluctuates irregularly. Our estimate of  $K_2$  quantifies the complexity of these amplitude fluctuations. We compare the results obtained for the CRU data set (interpolated measured temperature in the years 1901–2003 with  $0.5^\circ$  resolution, Mitchell et al., 2005<sup>1</sup>) with the ones obtained for the temperature data from a coupled ocean-atmosphere global circulation model (AOGCM, calculated at DKRZ). Furthermore, we compare the results obtained by means of  $K_2$  with the linear variance of the temperature data.

## 1 Introduction

Since the discovery of chaos in a conceptual climate model by Lorenz (1963) the predictability of weather (and climate) is still an open and not fully understood problem. Several similar conceptual models have been proposed so far (e.g. Lorenz, 1990). These models are usually given by systems of nonlinear ordinary differential equations, which can generate chaotic oscillations.

In this paper we consider daily mean temperature data from observed as well as gridded model data on a global scale. For the measured data the Climate Research Unit (CRU) dataset (Mitchell et al., 2005<sup>1</sup>) has been used. This monthly data on a  $0.5^\circ$  resolution has been interpolated from available observational data from all over the world (only

continental regions are considered). We apply a moving average filter to the simulated AOGCM data and then sample it down using each 15th data point. The resulting rather smooth temperature time series exhibit fluctuations on the time scale of one year (Fig. 3). Usually the annual fluctuations are considered to be a trivial (periodic) signal and are therefore filtered out (Govindan et al., 2002). However, the climate system is highly nonlinear and complex. If such a system is forced periodically, the resulting fluctuations may be highly complex and even chaotic (Lorenz, 1990) (see Fig. 4, where no clear annual cycle is observed). In order to quantify the complexity of the fluctuations of the signal's amplitude, we compute an estimate of the Rényi entropy of second order of each temperature time series, i.e. at any coordinate of the earth's surface. This measure quantifies how predictable the annual fluctuations are. This is an alternative approach to the ones presented in Fraedrich (1987); Latif and Barnett (1996); Chen et al. (1997).

We then compare the results obtained for the CRU data with the results for the data simulated based on an AOGCM, which is given on a coarser resolution of  $3.75^\circ$ , including the oceans, and find a good agreement. However, the estimated predictability depends on the geographical coordinates (for the CRU as well as for the AOGCM data), which is an interesting difference from the results obtained with the method of the detrending fluctuation analysis (DFA) applied in Govindan et al. (2002), which are rather independent on the geographical coordinates. Furthermore, we compare the results obtained with the estimate of the Rényi entropy with the variance of the data sets. A key result of this comparison is that over the continents the predictability correlates with the variance in the northern hemisphere and anticorrelates in the southern hemisphere.

The paper is structured as follows: after introducing the concept of the Rényi entropy of second order  $K_2$ , we outline the method that we use to estimate  $K_2$ . Then we describe the data we analyze in detail. After outlining the preprocessing, we discuss the results obtained with the CRU and AOGCM data and then conclude.

*Correspondence to:* W. von Bloh  
(bloh@pik-potsdam.de)

<sup>1</sup>Mitchell, T., Carter, T., Jones, P., Hulme, M., and New, N.: A comprehensive set of climate scenarios for Europe and the globe: the observed record (1901–2000) and the 16 scenarios (2001–2100), *J. Clim.*, submitted, 2005.

## 2 Rényi entropy of second order $K_2$

In this section, we first recall the definition of the Rényi entropy of second order. Let us therefore consider a trajectory  $\mathbf{x}(t)$  in a bounded  $d$ -dimensional phase space and suppose that the state of the system is measured at time intervals  $\tau$ . Let  $\{1, 2, \dots, M(\varepsilon)\}$  be a partition of the phase space in boxes of size  $\varepsilon$ . Then  $p(i_1, \dots, i_l)$  denotes the joint probability that  $\mathbf{x}(t=\tau)$  is in the box  $i_1$ ,  $\mathbf{x}(t=2\tau)$  is in the box  $i_2$ , ..., and  $\mathbf{x}(t=l\tau)$  is in the box  $i_l$ . The Rényi entropy of second order (Rényi, 1970; Grassberger, 1983) is then defined as

$$K_2 = - \lim_{\tau \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{l \rightarrow \infty} \frac{1}{l\tau} \ln \sum_{i_1, \dots, i_l} p^2(i_1, \dots, i_l). \quad (1)$$

This measure quantifies how fast the number of possible future evolutions increases as time goes by. If the system is perfectly deterministic in the classical sense, there is only one possibility for the trajectory to evolve. Hence, there exists only one series  $\{i_1, \dots, i_l\}$  of future steps with probability 1 and therefore,  $K_2=0$ . In contrast, one can easily show that for purely stochastic systems, the number of possible future trajectories increases to infinity so fast, that  $K_2 \rightarrow \infty$  (Thiel et al., 2003). Chaotic systems are characterized by a finite value of  $K_2$ , as they belong to an intermediate category. They are less predictable than purely periodic but more predictable than purely stochastic systems. Also in the chaotic case the number of possible trajectories diverges but not as fast as in the stochastic case. The inverse of  $K_2$  has units of time and can be interpreted as the mean prediction horizon/time of the system.

The algorithm to estimate  $K_2$  is linked to a weather prediction scheme introduced by Lorenz (1963). He proposed to use naturally occurring analogues for prediction. His idea is to record long series of data describing the state of the atmosphere (or any other system under consideration). To predict the weather one then has to compare the actual state with all the states in the database and to identify a former state which is (extremely) similar/close to the current one, with respect to some metric, so that the mismatch could also be attributed to a measurement error. Such a state is called ‘‘analogue’’. The prediction of the future is then given by the time evolution of the former state. In this paper we do not focus on prediction but rather on quantifying the predictability of a system. Therefore, we identify close analogues of a given state and consider all their respective evolutions. Then we quantify the time that these possible evolutions remain similar. This time is directly linked to  $K_2$ .

Next, we introduce a tool to visualize recurrences of trajectories of (dynamical) systems in phase space, which is called Recurrence Plot (RP) (Eckmann et al., 1987). These plots have proved to be rather useful for the analysis of time series, as they give a first impression of the behavior of the system under study. But the most interesting aspect of RPs for our purposes is that they allow to estimate  $K_2$  from time series in a very robust way.

## 3 Estimation of $K_2$

Let us consider a dynamical system represented by the trajectory  $\{\mathbf{x}_i\}$  for  $i=1, \dots, N$  in a  $d$ -dimensional phase space (i. e.  $\mathbf{x}_i \in \mathcal{R}^d$ ). Then we compute the recurrence matrix

$$R_{i,j} = \Theta(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|), \quad i, j = 1 \dots N, \quad (2)$$

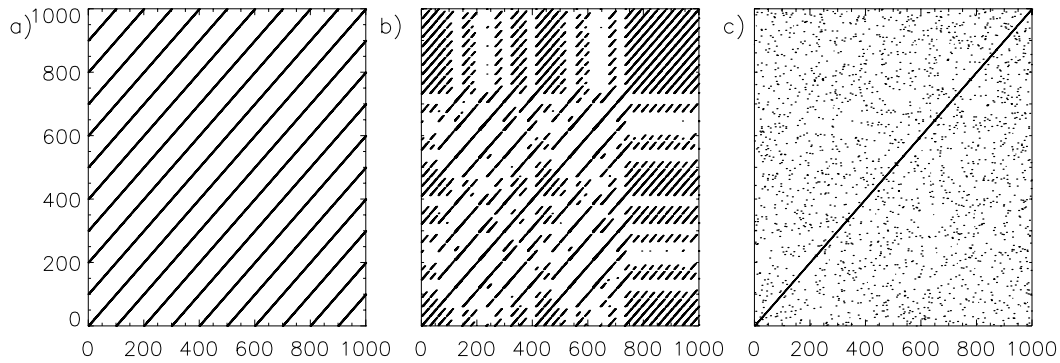
where  $\varepsilon$  is a predefined threshold and  $\Theta(\cdot)$  is the Heaviside function. The norm used in Eq. (2) is in principle arbitrary, but for theoretical reasons (Thiel et al., 2004), it is advantageous to use the maximum norm. The graphical representation of  $R_{i,j}$ , called Recurrence Plot (RP), is obtained encoding the value one as ‘‘black’’ and zero as ‘‘white’’ point. The analogues of a point  $\mathbf{x}_i$  are visualized in the  $i$ -th column of the RP as black points, i.e. if  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are analogues, then  $R_{i,j}=1$ . If the evolution of  $\mathbf{x}_i$  and  $\mathbf{x}_j$  is similar for  $n$  time steps we have  $R_{i+1,j+1}=1, \dots, R_{i+n,j+n}=1$ , i.e. we have a diagonal line in the RP. Figure 1a shows the RP of a sine function, i.e. a circle in phase space. Its RP is then characterized by non-interrupted diagonal lines. Figure 1b represents the RP of the Rössler system in a chaotic regime (Rössler, 1976). In this case, the predominant structure are diagonal lines, which are interrupted. Figure 1c represents the RP of white noise. It is homogeneous with mainly single points, which is typical for independent stochastic systems, as the state at time  $i+1$  is unrelated to the one at time  $i$ . From these plots, we can easily make out that there is a certain connection between the length of diagonal lines and the ratio of determinism or predictability inherent to the system. The better the predictability of a system is, the longer are these diagonals. Now we show how these diagonals can be used to estimate  $K_2$ .

Thiel et al. (2004) have shown that the following relationship holds

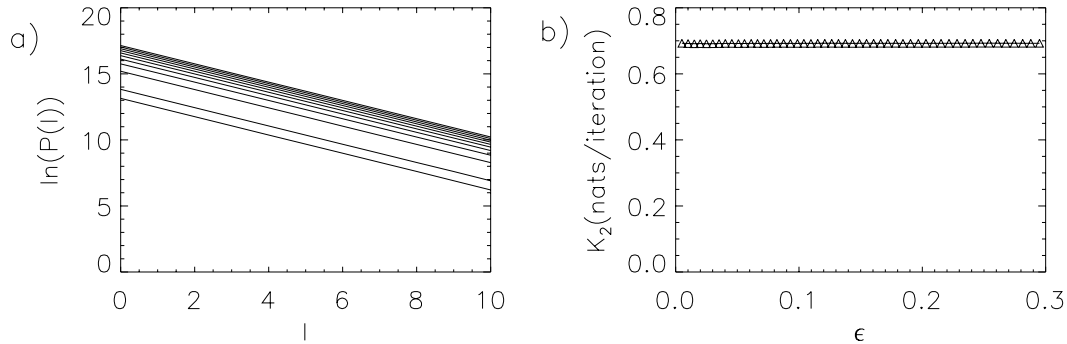
$$P_\varepsilon^c(l) \sim \varepsilon^{D_2} \exp\left(-\hat{K}_2(\varepsilon)\tau l\right), \quad (3)$$

where  $P_\varepsilon^c(l)$  is the cumulative distribution of diagonal lines in the RP, i. e. it represents the probability of finding a diagonal in the RP of at least length  $l$ , and  $\hat{K}_2(\varepsilon)$  is an estimator of  $K_2$ .  $D_2$  is the correlation dimension of the system under consideration (Grassberger, 1983). Therefore, the representation of  $P_\varepsilon^c(l)$  on a logarithmic scale versus  $l$  yields a straight line with slope  $-\hat{K}_2(\varepsilon)\tau$  for large  $l$ . For chaotic systems this slope is independent of  $\varepsilon$  in a rather large range of  $\varepsilon$ . This is shown in Fig. 2a for the chaotic Bernoulli map  $x_{n+1}=2x_n \bmod 1$ . Finally, one represents the slope of the curves for large  $l$  in dependence on  $\varepsilon$ . Then, for chaotic systems a plateau is found. The value of the ordinate of this plateau determines  $\hat{K}_2$  (Fig. 2b). In the case of the Bernoulli map, one obtains  $\hat{K}_2=0.6917$ , in good accordance with the theoretical values of  $\ln(2)$ .

In the representation of  $\ln(P_\varepsilon^c(l))$  vs.  $l$  we usually find two different slopes: one slope for short diagonal lines and another one for long diagonals (Thiel et al., 2003). Since  $K_2$  is defined for large  $l$  (Eq. 1), we have to detect the second slope. This can be done automatically (see Appendix). An



**Fig. 1.** Prototypical examples of RPs: **(a)** RP of a sine function, **(b)** RP of the Rössler system in chaotic regime, **(c)** RP of white noise.



**Fig. 2.** **(a)** Number of diagonal lines of at least length  $l$  versus  $l$  in the RP of the Bernoulli map for different values of the threshold  $\epsilon$ . The mean slope of the curves is equal to 0.6917. **(b)** Estimator of the Rényi entropy of second order  $\hat{K}_2$  vs.  $\epsilon$  for the Bernoulli map.

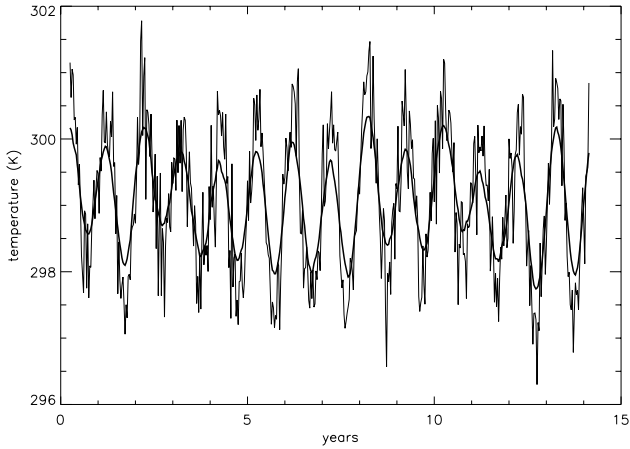
advantage of this algorithm to estimate  $K_2$  is that the result is independent of the embedding parameters used (see Thiel et al., 2003), i.e. even without embedding the time series at all, we obtain reasonable results. However, using embedding allows the algorithm to recognize the scaling region more easily. Hence, we will embed the time series in our analysis. Furthermore, our studies have shown that the algorithm yields robust results even if a time series is corrupted by noise or if one adds a nonstationarity.

#### 4 Description of the data and preprocessing

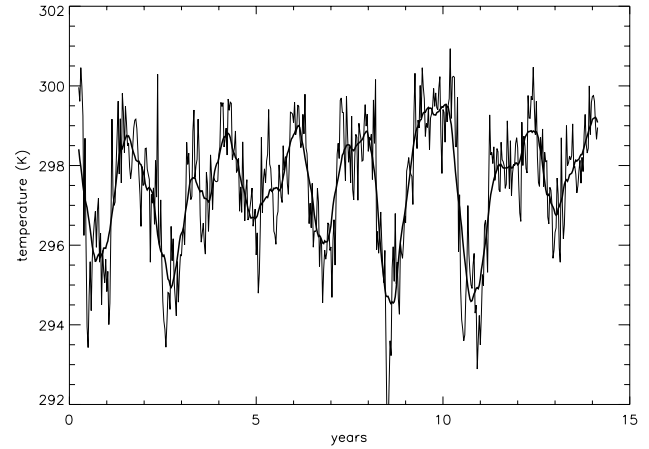
As an example for (partially) measured data the Climate Research Unit (CRU) dataset (Mitchell et al., 2005<sup>1</sup>) has been used. This monthly data on a spatial resolution of  $0.5^\circ$  has been constructed from global available observed data. It is based on interpolation from all stations available at any moment in time. The primary purpose was to create an input data set for environmental modeling. The CRU data set has been widely used as e.g. a driving force for models of the global vegetation dynamics (Sitch et al., 2003). The monthly data has been embedded in a two-dimensional space with a delay of one month.

The global climate model (AOGCM) consists of the spectral atmospheric model ECHAM4 Roeckner et al. (1996) and the ocean model HOPE-G Wolff et al. (1997), both devel-

oped at the Max-Planck-Institute of Meteorology in Hamburg. In our case the model ECHAM4 has a horizontal resolution of T30 (approximately  $3.75^\circ \times 3.75^\circ$ ) and 19 vertical levels, five of them located above 200 hPa. The horizontal resolution of the ocean model HOPE-G is about  $2.8^\circ \times 2.8^\circ$  with a grid refinement in the tropical regions, where the meridional grid-point separation decreases progressively to the equator, reaching a value of  $0.5^\circ$ . This increased resolution allows for instance for a more realistic representation of El Niño southern oscillation (ENSO) events. The ocean model has 20 vertical levels. In this simulation, the model has been driven by estimations of three past external forcing factors: solar variability, greenhouse gas concentrations in the atmosphere and an estimation of the radiative effects of stratospheric volcanic aerosols. No changes in the anthropogenic atmospheric aerosol concentrations have been considered. Changes in vegetation cover or land-use have also been neglected. The atmospheric concentrations of two greenhouse gases, carbon dioxide and methane, have been estimated from analysis of air bubbles trapped in Antarctica ice cores (Etheridge et al., 1996; Blunier et al., 1995). The past variations of solar output have been derived from the values used in Crowley (2000). For the period after 1610 A. D., past solar variations are empirically estimated from observations of sun spots (Lean et al., 1995) and between 1500 and 1610 A. D. they are based on concentrations of the cosmogenic isotope  $^{10}\text{Be}$ . The third external factor is the



**Fig. 3.** Temperature time series from AOGCM at the coordinates (86.25° W, 11.25° N) (solid line) and its corresponding filtered signal (bold line). The annual period is predominant in this time series.



**Fig. 4.** Temperature time series from AOGCM at the coordinates (63.75° E, 3.75° S) (solid line) and its corresponding filtered signal (bold line). In this time series no annual period is observed.

stratospheric loading of volcanic aerosols. After a volcanic eruption, these are washed out by precipitation, that influence the acidity of the ice layers in ice cores. Changes in optical densities of the stratosphere can be thus estimated from ice acidity through a semi-empirical model (Crowley, 2000). We only used the first 200 years (1500–1700), not the full record.

In contrast to the CRU dataset we have a daily resolution of the AOGCM data, but greenhouse gas emissions are not present in this period. As we are interested in the annual fluctuations of the amplitude of the temperature signals, we apply a moving average filter to the data and then downsample them, considering only each 15th data point. The resulting temperature time series shows oscillations with more or less constant frequency, depending on the geographical position, but with varying amplitude. This preprocessing of the data was applied to the AOGCM data. The data was embedded in a three-dimensional space with a delay of 5 days. These embedding parameters were fixed applying the conventional methods of the autocorrelation function and the false nearest neighbors method (Kantz and Schreiber, 1997), although it is not necessary to embed the time series to estimate  $K_2$  from RPs, as we have mentioned above (Thiel et al., 2003). We embed to facilitate the automatized estimation by our algorithm. In Figs. 3 and 4 two examples are represented. The time series in Fig. 3 shows a predominant annual cycle, whereas the fluctuations displayed in Fig. 4 do not. This last figure shows that there is no strict annual cycle at all geographical positions.

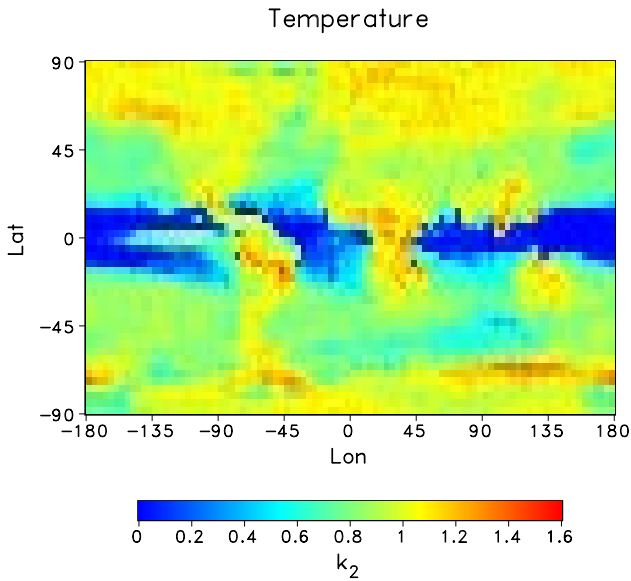
## 5 Results

In this section, we estimate  $K_2$  for the CRU data and for the AOGCM model data by means of the method of RPs. Due to the high dimensionality of the models and/or the random elements involved, the slope of  $P_\varepsilon^c(l)$  depends, as expected, on  $\varepsilon$  (see Eq. 3 and Fig. 13 Thiel et al., 2002). In Sect. 7 we discuss this feature in more detail. The entropy  $K_2$  is also

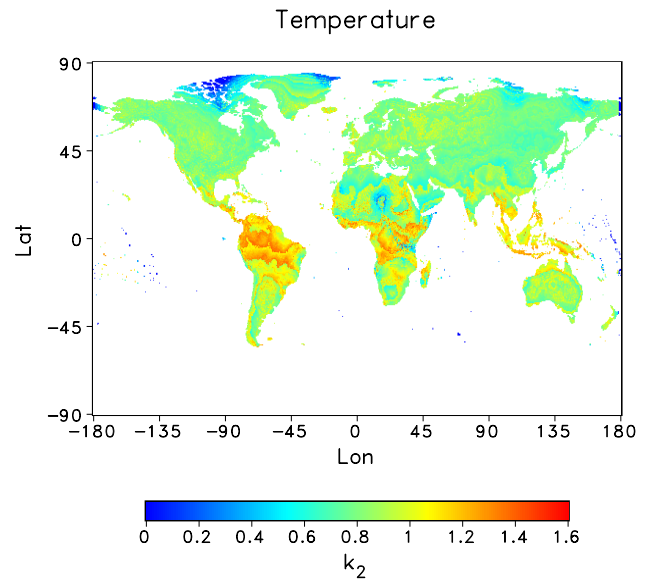
well defined in high dimensional and/or stochastic systems but according to its definition one has to estimate the limit  $\varepsilon \rightarrow 0$ . However, due to the finite length of the time series, this is not possible. Several approaches have been developed to solve this problem (Grassberger, 1983; Urbanowicz and Holyst, 2003). The RP-based method has shown to be very suitable for the estimation of  $K_2$  (Thiel et al., 2004). It also yields very reliable estimates if e.g. nonstationarities are involved (Romano, 2004). Moreover, the RP based method has been already applied to experimental turbulent flow data and has confirmed results which were reported in the literature (Thiel et al., 2004; Read et al., 1992). Various filters did not have to be applied as the RP based method can deal with a rather high degree of nonstationarity and noise. Furthermore, for turbulent systems, such as the coupled ocean-atmosphere one, the predictability depends on the scale considered. Hence, we set  $\varepsilon=0.1$ , which corresponds to tenth a Kelvin prediction error, and determine  $\hat{K}_2$  by the slope of  $P_\varepsilon^c(l)$  vs.  $l$ .

The estimates of  $K_2$  for the AOGCM data set are shown in Fig. 5. We find that the entropy  $\hat{K}_2$  is maximal in the tropical region. Rather high values of  $\hat{K}_2$  can also be found in the northern hemisphere (Alaska, central and northern Europe). ENSO can clearly be identified over the ocean as an elevated area of  $\hat{K}_2$ . Therefore the predictability of temperature fluctuations is reduced in that area (the prediction horizon is proportional to the inverse of  $K_2$ ).

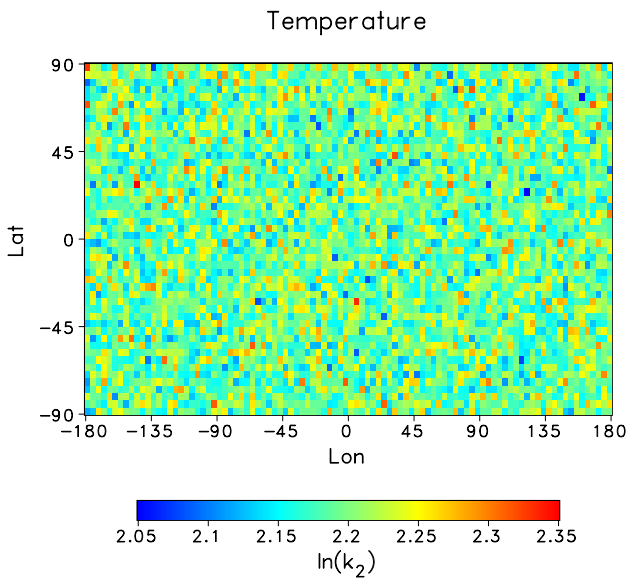
To test the reliability of the estimates of  $K_2$  for the AOGCM data, a bootstrap resampling has been performed. We used the shuffled surrogates because they destroy the correlations and also reduce the predictability of the time series. The results show that the structures obtained by mean of  $\hat{K}_2$  are significant (see Fig. 6). The picture lacks all coherent structure. No continents can be made out. Also the values of  $\hat{K}_2$  are much higher for the shuffled data. In order to validate these results we compute  $\hat{K}_2$  also for the CRU data. The respective estimates are represented in Fig. 7. We observe,



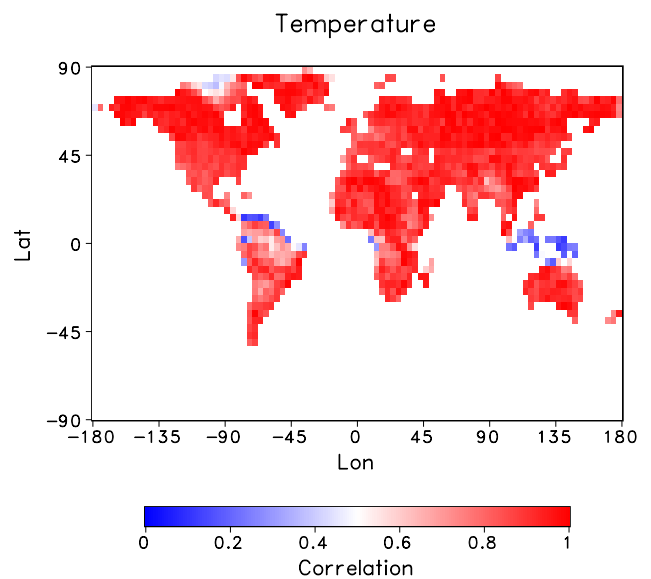
**Fig. 5.**  $K_2$  estimates for the AOGCM data set for a fixed  $\varepsilon=0.1$  K.



**Fig. 7.**  $K_2$  estimates for the CRU data set for a fixed  $\varepsilon=0.1$  K.



**Fig. 6.**  $K_2$  estimates for the randomly shuffled AOGCM data set for a fixed  $\varepsilon=0.1$  K.



**Fig. 8.** The correlation index  $r$  between the estimates of  $K_2$  for the CRU and AOGCM data set.

that the structures are similar in both CRU and AOGCM data (e. g. the values of  $\hat{K}_2$  are high in the tropical regions in both cases).

In order to compare quantitatively the estimates of  $K_2$  for the AOGCM and the CRU data, we compute the following index

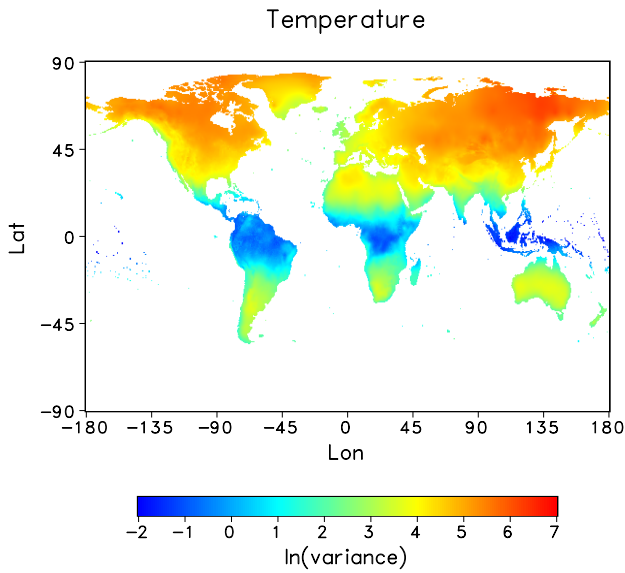
$$r := 1 - |\tilde{K}_{2\text{CRU}} - \tilde{K}_{2\text{AOGCM}}|, \quad (4)$$

where  $\tilde{K}_2$  denotes the values rescaled to the interval  $[0, 1]$ . The index  $r$  is of the order of 1 if the estimates of  $K_2$  for the AOGCM and CRU data are correlated and of the order of 0 if they are anticorrelated. Figure 8 displays the index

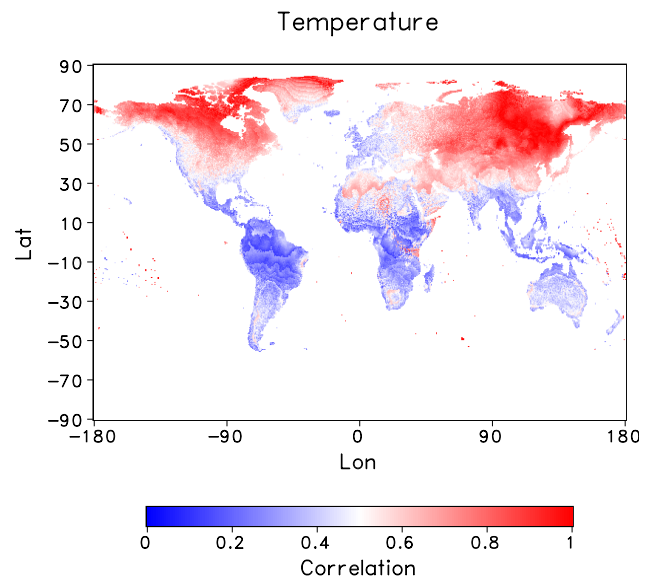
$r$ . We can confirm, that the structures seen in the  $K_2$  estimates are similar in both CRU and AOGCM data, indicated by the homogeneous red pattern<sup>2</sup>. Hence, with respect to the  $\hat{K}_2$  statistic the AOGCM data reproduced the structures obtained with the CRU data set. This analysis shows that with respect to the predictability the CRU and the AOGCM data are qualitatively (and almost quantitatively) equivalent.

Now, we address the question what can be learned from the  $K_2$  estimates beyond what can be found by linear techniques. Usually, the models are validated by means of tra-

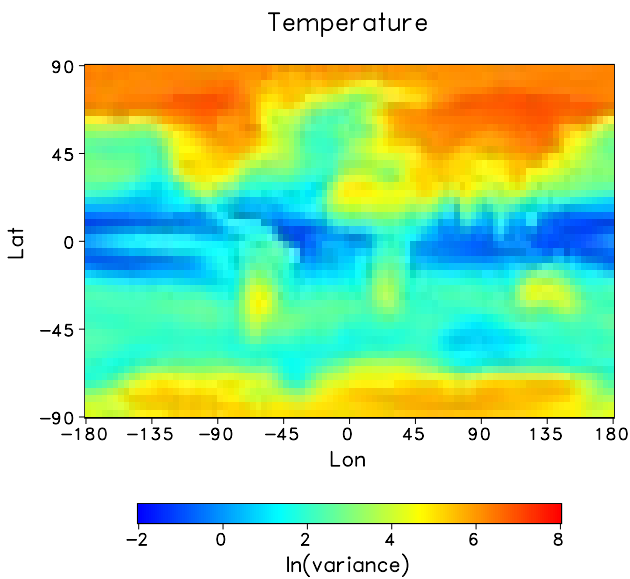
<sup>2</sup>The blue regions in Fig. 8 are due to the lower resolution of the AOGCM with respect to the CRU data.



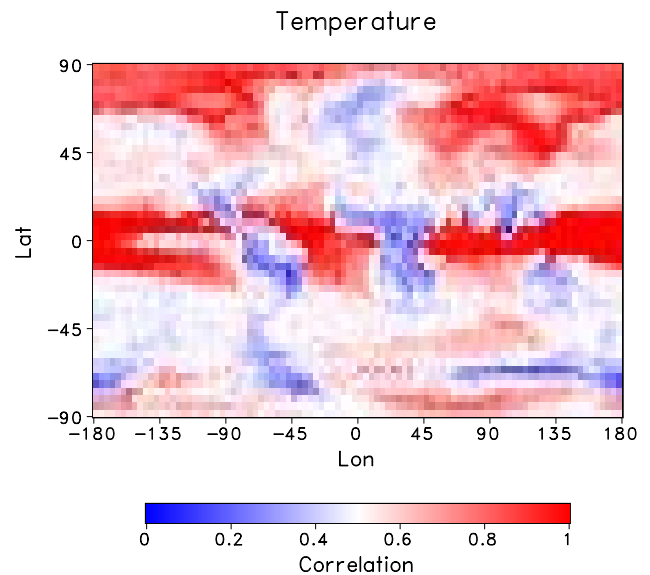
**Fig. 9.** Variance for the CRU data set.



**Fig. 11.** The correlation index  $r$  between  $\hat{K}_2$  and variance for the CRU data set.



**Fig. 10.** Variance for the AOGCM data set.



**Fig. 12.** The correlation index  $r$  between  $K_2$  and variance for the AOGCM data set.

ditional linear time series analysis (Govindan et al., 2002). Hence, we also compute the variance of the filtered time series for both CRU (Fig. 9) and AOGCM (Fig. 10) data. In order to compare quantitatively the results obtained with  $K_2$  and the variance, we compute the index  $r$  for the CRU and for the AOGCM data, analogously to Eq.(4),

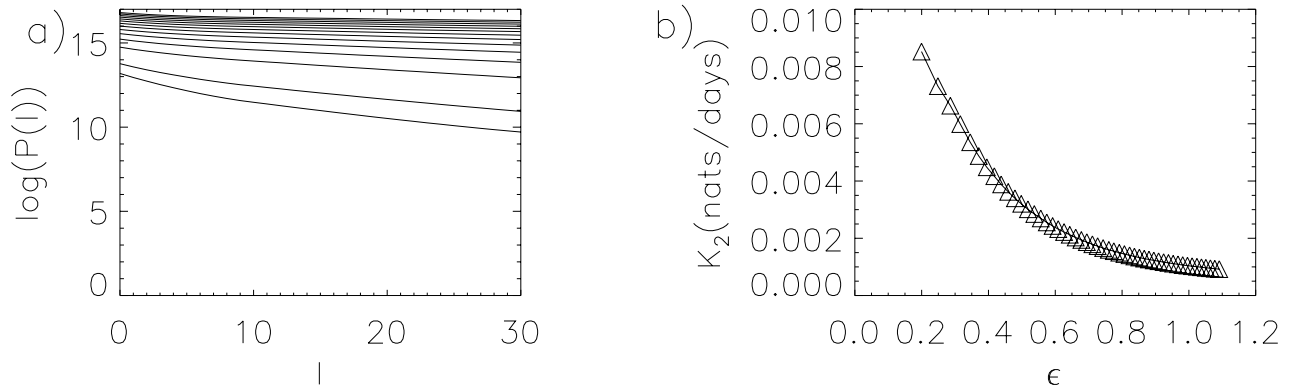
$$r := 1 - |\tilde{K}_2 - \tilde{Var}|, \quad (5)$$

where  $\tilde{Var}$  denotes the values rescaled to the interval  $[0, 1]$ . The results are plotted in Fig. 11 and Fig. 12, respectively.

From these figures, we can conclude the following points:

- The GCM and CRU data exhibit qualitatively (and quantitatively) the same long term predictability.

- On the continent the variance and  $\hat{K}_2$  of the annual temperature fluctuations are correlated in the northern hemisphere and anticorrelated in the southern hemisphere. Western Europe seems to be an exception from this thumb rule.
- Over the oceans the variance and  $\hat{K}_2$  are correlated in the polar regions and in the vicinity of the equator and uncorrelated in the mid-latitudes. The southwest region of America, where the ENSO takes place is the exception from this rule.



**Fig. 13.** (a) Number of diagonal lines of at least length  $l$  versus  $l$  in the RP of one temperature time series from AOGCM for different values of the threshold  $\epsilon$ . (b) Corresponding estimator of the Rényi entropy of second order  $\hat{K}_2$  vs.  $\epsilon$ .

## 6 Conclusions

The analysis shows that the predictability does not simply correlate with the variance of the mean daily temperature data. The degree of correlation depends on the latitude and on the fact whether one is on a continent or not. The annual fluctuations exhibit a complex pattern and in some cases it might not be suitable to filter them out.

## 7 Problems of the method

The method to estimate  $K_2$  which we have proposed (Sect. 3) allows to investigate automatically large amounts of data and yields rather robust results. However, the algorithm must be applied carefully in some cases.

For high dimensional or stochastic signals, the slope of  $P_\epsilon^c(l)$  depends in general on  $\epsilon$  (Thiel et al., 2003). Only for periodic and chaotic systems  $K_2$  is independent of  $\epsilon$ . For the AOGCM data  $P_\epsilon^c(l)$  vs.  $l$  is shown in Fig. 13a in dependence on  $\epsilon$ . Figure 13b displays the respective dependence of  $\hat{K}_2$  on  $\epsilon$ . One now has to compute the limit  $\epsilon \rightarrow 0$ . One possibility to overcome this problem is to fit a curve e.g. a polynomial to  $\hat{K}_2(\epsilon)$ , and then to extrapolate to  $\epsilon=0$  (Urbanowicz and Holyst, 2003). But there is some ambiguity choosing the type of function one might use for the fitting.

An even simpler approach is used in Sect. 5. We fix  $\epsilon = \epsilon_{\text{fix}}$  and then consider  $\hat{K}_2(\epsilon_{\text{fix}})$  as an estimate of the complexity of the signal.  $\hat{K}_2(\epsilon)$  is a measure for the time that two trajectories stay within an  $\epsilon$ -ball or -tube. Based on this approach we estimate the time, for which the “prediction” is correct up to an error less or equal to  $\epsilon$ . Hence,  $\hat{K}_2$  can be interpreted in a very straight forward and physical manner.

In this paper we have presented an algorithm to estimate the complexity and predictability of the earth surface temperature fluctuations. The algorithm is based on recurrences of the temperature signal in an abstract embedding space. We have analyzed two data sets (CRU and AOGCM) and found structurally the same results. Comparing the results with the variance of the temperature signal, one finds that the estimate

of the  $K_2$  entropy yields complementary information to the variance. The variance correlates with  $\hat{K}_2$  over the continents in the northern hemisphere and anticorrelates in the southern hemisphere. This correlation and anticorrelation still lacks a physical interpretation. To find an explanation of this dependence would give further insights into the behavior of the ocean/atmosphere system. However, the combination of linear and nonlinear data analysis techniques can be used to analyze and quantify data beyond linearity. The analysis suggests that the annual fluctuations of the surface temperature are not trivial but exhibit a complex behavior. This complexity depends on the geographical position in a nontrivial way. Hence, filtering the annual oscillations out may yield spurious results.

The presented algorithm resembles an idea introduced by Lorenz (1969), where a similar algorithm to estimate the predictability of different climate variables (such as temperature, pressure, etc.) is presented. The recurrence based algorithm to estimate the predictability has been applied for the first time to a high dimensional climate system. To validate the correctness of the application of the algorithm to such a system it would be necessary to apply it to model systems with many coupled temporal and spatial time scales (Lorenz, 1991). This will be the subject of a forthcoming paper. However, our algorithm can be supposed to yield rather reliable results in high dimensional systems for two reasons: 1) In lower dimensional cases it gives robust results with rather short time series and is often more efficient than standard techniques (Thiel et al., 2003). 2) Other measures which quantify predictability (DET, DIV) and are estimated from RPs have been successfully applied to climate data. The comparison of these measures with  $K_2$  in lower dimensional systems suggests that the latter has independently of the dimension of the underlying systems notable advantages. It is, e.g. numerically close to the largest Lyapunov exponent.



## Appendix: Automatization of the algorithm

For many applications, e.g. if spatio-temporal data has to be analyzed, it is desirable to automate the algorithm to estimate  $K_2$  based on RPs. Such an automated algorithm is also more objective, as otherwise the choice of the proper scaling regions of  $P_\varepsilon^c(l)$  depends to some extent on the data analyst.

The crucial step for the automatization is the estimation of the scaling region of  $\ln P_\varepsilon^c(l)$  vs.  $l$ . For many systems not only one slope is found in the representation of  $\ln(P_\varepsilon^c(l))$  vs.  $l$ . Usually, for small  $l$  the slope is greater than for large  $l$ . As  $K_2$  is defined for long diagonal lines, the slope for large  $l$  has to be identified. On the other hand, too long lines cannot be taken into account, due to the fact that the time series is finite and hence, the length of the longest diagonal line to be found is limited. For this reason, the decay of  $\ln(P_\varepsilon^c(l))$  for very long diagonals is not the correct one.

Therefore, the automatic detection of the right slope is realized applying a cluster dissection algorithm (Späth, 1992). The algorithm divides the set of points into distinct clusters. In each cluster a linear regression is performed. The algorithm minimizes the sum of all square residuals in order to determine the scaling region. The following parameters are used:

- We consider only diagonal lines up to a fixed length  $l_{\max}=200$ . Longer lines are excluded because of finite size effects, as explained above. Reasonable values of  $l_{\max}$  are at about 10% of the length of the time series.
- We consider only values of  $P_\varepsilon^c(l)$  with  $P_\varepsilon^c(l)>500$  to obtain a reliable statistic.
- We further have to specify the number of clusters when applying the cluster dissection algorithm: two different clusters seem to be a rather good choice for many systems. Then, we use the slope of the largest cluster.

These choices have proven to be the most appropriate ones for the estimation of the scaling regions. All these parameters are defaults of a computer program.

*Acknowledgements.* We thank the SPP 1114 program supported by the DFG and the Promotionskolleg “Computational Neuroscience of Behavioral and Cognitive Dynamics” supported by the Ministerium für Forschung und Kultur Brandenburg (Germany). We also thank E. Zorita and the Deutsches Klimarechenzentrum (DKRZ) for providing the AOGCM data and scientific support.

Edited by: J. von Hardenberg

Reviewed by: two referees

## References

- Blunier, T., Chapellaz, J. A., Scwander, J., Stauffer, B., and Raynaud, D.: Variations in atmospheric methane concentration during the Holocene epoch, *Nature*, 374, 46–49, 1995.
- Y.-Q. Chen D.S. Battisti, T.N. Palmer, J.J. Barsugli and E.S. Sarachik, A study of the predictability of Tropical 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 (1997).
- Crowley, T. J.: Causes of climate change over the past 1000 years, *Science*, 289, 270–277, 2000.
- Eckmann, J.-P., Kamphorst, S. O., and Ruelle, B.: Recurrence Plots of Dynamical Systems, *Europhys. Lett.*, 4, 973–977, 1987.
- Etheridge, D., Steele, L. P., Langenfelds, R. L., Frandcey, R. J., Barnola, J. M., and Morgan, V. I.: Natural and anthropogenic changes in atmospheric CO<sub>2</sub> over the last 1000 years from air in Antarctic ice and firn, *J. Geophys. Res.*, 101, 4115–4128, 1996.
- Fraedrich, K.: Estimating weather and climate predictability on attractors. *J. Atmos. Sci.*, 44, 722–728, 1987.
- Grassberger, P.: Generalized Dimensions of Strange Attractors, *Phys. Lett. A*, 97, 6, 227–230, 1983.
- Govindan, R. B., Vyushin, D., Bunde, A., Brenner, S., Havlin, S., and Schellnhuber, H.-J.: Global Climate Models Violate Scaling of the Observed Atmospheric Variability, *Phys. Rev. Lett.*, 89, 2, 0285011–0285014, 2002.
- Kantz, H., and Schreiber, T.: *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, 1997.
- Latif, M. and Barnett, T. P.: Decadal climate variability over the North Pacific and North America: Dynamics and predictability, *J. Climate*, 9, 10, 2407–2423, 1996.
- Lean, J., Beer, J., and Bradley, R.: Reconstruction of solar irradiance since 1610: Implications for climate change, *Geophys. Res. Lett.*, 22, 23, 3195–3198, 1995.
- Lorenz, E. N.: Deterministic Nonperiodic Flow, *J. Atmos. Sci.*, 20, 130–141, 1963.
- Lorenz, E. N.: Can chaos and intransitivity lead to interannual variability?, *Tellus*, 42A, 378–389, 1990.
- Lorenz, E. N.: Dimension of weather and climate attractors, *Nature*, 353, 241–244, 1991.
- Rényi, A.: *Probability theory*, North-Holland, Appendix, 1970.
- Read, P. L., Bell, M. J., Johnson, D., and Small, R. M.: Quasi-periodic and chaotic flow regimes in a thermally driven, rotating fluid annulus, *J. Fluid Mech.*, 238, 599, 1992.
- Roeckner, E., Arpe, L., Bengtsson, L., Christoph, M., Claussen, M., Dümenil, L., Esch, M., Giorgetta, M., Schlese, U., and Schulzweida, U.: Atmospheric general circulation model ECHAM-4: Model description and simulation of present-day climate, MPI Report No. 218, Max-Planck-Institute for Meteorology, Hamburg, Germany, 1996.
- Romano, M. C.: Synchronization Analysis by Means of Recurrences in Phase Space, Doctoral thesis, <http://pub.ub.uni-potsdam.de/2004/0075/romano.pdf>, 2004.
- Rössler, O. E.: An equation for continuous chaos, *Phys. Lett.* 57A, 397–398, 1976.
- Sitch, S., Smith, B., Prentice, I. C., Arneeth, A., Bondeau, A., Cramer, W., Kaplan, J., Levis, S., Lucht, W., Sykes, M., Thonicke, K., and Venevski, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ Dynamic Vegetation Model, *Global Change Biology*, 9, 161–185, 2003.
- Späth, H.: *Cluster Dissection and Analysis*, Horwood, Chichester, 1992.
- Thiel, M., Romano, M. C., Kurths, J., Meucci, R., Allaria, E., and Arecchi, F. T.: Influence of observational noise on the recurrence quantification analysis, *Physica D*, 171, 3, 138, 2002.
- Thiel, M., Romano, M. C., and Kurths, J.: Analytical Description of Recurrence Plots of White Noise and Chaotic Processes, *Applied Nonlinear Dynamics*, 11, 3, 20–29, 2003.

Thiel, M., Romano, M. C., Read, P., and Kurths, J.: Estimation of Dynamical Invariants without Embedding by Recurrence Plots, *Chaos*, 14, 2, 234–243, 2004.

Urbanowicz, K. and Holyst, J. A.: Noise-level estimation of time series using coarse-grained entropy, *Phys. Rev. E*, 67, 046218, 2003.

Wolff, J.-O., Maier-Reimer, E., and Legutke, S.: The Hamburg Ocean Primitive Equation Model, Technical report, No. 13, German Climate Computer Center (DKRZ), Hamburg, 98, 1997.