Nonlin. Processes Geophys., 17, 651–661, 2010
www.nonlin-processes-geophys.net/17/651/2010/
doi:10.5194/npg-17-651-2010

Nonlinear Processes in Geophysics

Markov chain analysis of regional climates

S. Mieruch[1], **S. Noël**[1], **H. Bovensmann**[1], **J. P. Burrows**[1], **and J. A. Freund**[2]

[1]Institute of Environmental Physics (IUP), University of Bremen, Otto-Hahn Allee 1, 28359 Bremen, Germany
[2]Institute for Chemistry and Biology of the Marine Environment (ICBM), University of Oldenburg, Carl-von-Ossietzky Str. 9-11, 26111 Oldenburg, Germany

Received: 4 August 2010 – Revised: 1 November 2010 – Accepted: 10 November 2010 – Published: 19 November 2010

Abstract. We present a novel method for regional climate classification that is based on coarse-grained categorical representations of multivariate climate anomalies and a subsequent Markov chain analysis. From the estimated transition matrix several descriptors, such as *persistence*, *recurrence time* and *entropy*, are derived. These descriptors characterise dynamic properties of regional climate anomalies and are connected with fundamental concepts from nonlinear physics like residence times, relaxation process and predictability. Such characteristics are useful for a comparative analysis of different climate regions and, in the context of global climate change, for a regime shift analysis.

We apply the method to the bivariate set of water vapour and temperature anomalies of two regional climates, the Iberian Peninsula and the islands of Hawaii in the central Pacific Ocean. Through the Markov chain analysis and via the derived descriptors we find significant differences between the two climate regions. Since anomalies are departures from seasonal and long term components, these differences relate to differences in the short term stability of both regional climates.

1 Introduction

Climate classifications represent the complex interaction of climate elements and climate factors as well as their impact on the Earth's surface in form of climate types, as stated by Lauer and Bendix (1993). Generally two types of climate classification exist, which are the *genetic classification* and the *empirical classification*. Genetic classification describes climate with respect to the climate genesis, e.g. continental and oceanic climate. In the 1950s the famous German climatologist Hermann Flohn developed a genetic climate classification based on global circulation systems (Flohn, 1957). In contrast, empirical climate classification is based on climate appearance, in the form of vegetation, temperature etc. The combination of genetic and empirical classification is called *integrative climate classification* and was developed by Lauer and Frankenberg (1988). In the context of the topology of life-forms Konrad Lorenz stated: "Without the essential principle of classification in the sense of abstract types, it would be impossible for our awareness to bring order and clarity into the overwhelming manifold of the forms around us…" (Lorenz, 1983), which is also true for climate.

The earliest attempt of climate classification probably goes back to *Parmenides* of *Elea* (500 BCE), who tried to differentiate climate zones of the at that time known world (Blüthgen and Weischelt, 1980). One of the most common empirical climate classifications, which is still used today, was developed by Wladimir Köppen in the year 1900, which is based on empirical observations of vegetation, temperature and precipitation. The quite useful Köppen classification was updated recently by Kottek et al. (2006) and Peel et al. (2007). Another approach was performed by Holdridge (1947) who divided the Earth in live zones based on evapotranspiration (evaporation from plants), precipitation and humidity. Climate classification is used on all scales, for instance, Gerstengarbe and Werner (1999) have updated the well-known classification of the European North Atlantic region in 29 *Großwetterlagen*. Nicolis et al. (1997) analysed more local weather patterns from Switzerland and parts of Austria. They employed coarsened descriptions of weather regimes via three main clusters (convective, advective and mixed weather) and transitions between them to model climate dynamics in the framework of Markov chains.

As stated by Nicolis et al. (1997), the main motivation for mapping meteorological fields onto a small number of symbols is the chance to make "predictions beyond the predictability time" of fine scale weather. Furthermore, in the context of climate change, an important application of

Correspondence to: S. Mieruch
(mieruch@iup.physik.uni-bremen.de)

climate classifications is the detection of gradual changes of climate types or abrupt transitions between prototypic states, so-called regime shifts. Such regime shifts are also analysed in the framework of bifurcation analysis (Scheffer and Carpenter, 2003). For instance, Beck et al. (2005) detected changes of the Köppen climate zones over time, which could be attributed to global warming.

The purpose of our paper is to open up the perspective for a new global climate classification scheme based on Markov chain analyses. Since our method evaluates aspects of climate dynamics that usually are neglected, we propose to use it supplementary to the existing climate classification methods and not as an alternative to these. Whereas existing climate classification schemes rely on absolute environmental quantities, our new approach investigates the temporal covariation and interaction of climate anomalies. We thus coin a *dynamic climate classification* and add it to the afore-mentioned existing climate classifications (genetic, empirical, integrative). In contrast to these our novel scheme classifies regional climates through a statistical analysis of departures from seasonal and trend behaviour. Consequently, quantities like persistence, recurrence time and entropy (to be defined below) have to be interpreted as dynamical characteristics of fluctuations. A relaxation of these fluctuations (towards the seasonal and long term components, in particular the annual cycle) is governed by the in general nonlinear dynamics and coupling of various climatic processes. Changes in the relaxation characteristics can therefore be interpreted in the context of nonlinear dynamics phenomena as, for instance, bifurcation scenarios.

Methodologically we transfer the ideas of Hill et al. (2004) and Freund et al. (2006), who analysed ecological communities through Markov chains, to water vapour and temperature measurements, an approach which is new in atmospheric research. As other methods of multivariate data analysis, e.g. a reduction to the first principal component (of a PCA), our method maps multivariate data series, such as water vapour and temperature series, to a univariate sequence of symbols in a rather transparent way. This method of data representation is also known as symbolic dynamics (e.g. Daw et al., 2003) and is widely used in the statistical analysis of chaotic time series resulting from nonlinear dynamics. Estimating one-step transition probabilities from the symbol series means to describe this sequence as a Markov chain. The efficiency of a first-order Markov description relies on the fact that the sampling interval Δt roughly matches the typical auto- and cross-correlation time τ of the multivariate series. The case of a mismatch (i.e. $\Delta t < \tau$) might be cured by resampling the data. The scale of resolution is determined by the number of symbols introduced. As elaborated in Sect. 2, the aim of high resolution has to be traded off against tolerable estimation errors of transition probabilities. The method is free from any assumptions on distributions (e.g. normal distribution) and can easily be extended to data with gaps.

The document is structured as follows: Sect. 2 represents a short overview on symbolic dynamics, Markov chains and derived descriptors. Section 3 introduces the data and presents the exemplary application of the methods to two climate regions. Finally, we discuss the results and give the conclusions and an outlook in Sect. 4.

2 Methods

2.1 Symbolic dynamics – coarse grained data representations

Quite often processes of interest correspond to a dynamics evolving in an m-dimensional continuous space. Recordings of such processes lead to multivariate time series with values $\boldsymbol{x}(t_n) = \{x_1(t_n), \ldots, x_m(t_n)\}$ measured at discrete (equidistant) sampling times $t_n (= n\Delta t)$. An analysis and interpretation of the data series is traditionally done against the backdrop of linear stochastic processes. More recently the comprehension of nonlinear dynamics and the paradigms of deterministic chaos have launched diverse methods of nonlinear time series analysis (Kantz and Schreiber, 2004). Many of these methods work on a coarse-grained representation of data series which is effected by partitioning state space or the *attractor* X, i.e. the subset supporting the asymptotic dynamics, into cells $\mathcal{C}_i$ uniquely labeled with symbols c_i $(i = 1, \ldots, \lambda)$ (e.g. Ebeling and Nicolis, 1992; Daw et al., 2003). The shape of these cells can in principle be chosen arbitrarily. The collection of cells $\{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_\lambda\}$ constitutes a *partition* (Lind and Marcus, 1996) if its union covers the attractor (or state space) completely and if cells are pairwise mutually disjoint, i.e.

$$X = \bigcup_i \mathcal{C}_i \quad \text{and} \quad \mathcal{C}_i \bigcap \mathcal{C}_j = \emptyset, \quad \forall i \neq j. \tag{1}$$

The collection of related symbols $\{c_1, c_2, \ldots, c_\lambda\}$ may be called an alphabet $\mathcal{A}$ of size $|\mathcal{A}| = \lambda$. The finiteness of the partition is more a practical than a rigorous demand. Likewise, for reasons of better interpretation cells are frequently chosen as simply connected subsets. The maximum of all cell diameters is defined as *refinement* of the partition and can also be viewed as the scale of resolution. The strive for high resolution would favour partitions with quite many small cells. However, since the essence of symbolic dynamics is to evaluate the statistics of symbol subsequences the demand for large cell numbers must be weighed against expected estimation errors. The latter-mentioned are typically tied to the sample size, which is equal or less than the length of the time series.

Lets suppose we have chosen a partition with λ cells and we want to reconstruct the probabilities for subsequences $\sigma_1, \ldots, \sigma_n$ (where each $\sigma_i \in \mathcal{A}$) of length n, also called n-words. The number of possible combinations is given by λ^n. The combinatorial explosion poses a considerable problem for statistical estimation because the sample size, i.e. the

length N of the sequence, definitely should be larger or at least not less than the number of states. As a rule of thumb we might demand that $N \geq 2\lambda^n$ which, for an equidistribution, means to allow each state to occur twice in the sample. Since N is fixed by the length of the time series this usually poses severe constraints on λ and n. This coarse grained representation of time series will be described in the following as a stochastic process, particularly a first order Markov chain.

2.2 Markov chains – a brief review

A Markov chain is a time and state discrete stochastic process (Norris, 1998), which obeys the Markov property: given the present state σ_t, the future state σ_{t+1} is independent from past states σ_{t-k} (for all positive k). Mathematically, the above statement is expressed via conditional probabilities as

$$P(\sigma_{t+1}|\ldots,\sigma_{t-2},\sigma_{t-1},\sigma_t) = P(\sigma_{t+1}|\sigma_t) \tag{2}$$

and has the consequence that joint probabilities can be decomposed in the following way

$$\begin{aligned} P(\sigma_1,\ldots,\sigma_t) \;=\; & P(\sigma_t|\sigma_{t-1}) \cdot P(\sigma_{t-1}|\sigma_{t-2}) \cdot \\ & \ldots \cdot P(\sigma_2|\sigma_1) \cdot P(\sigma_1)\,. \end{aligned} \tag{3}$$

The conditional probabilities $P(\sigma_{t+1}|\sigma_t)$ can be viewed as transition probabilities and be organised as entries of a $\lambda \times \lambda$ transition matrix $\mathbf{P}_t$ where the subscript t indicates the fact that the matrix is in general time variant. Time independent transition probabilities describe so-called homogeneous Markov chains which are fully characterised by *the* transition matrix $\mathbf{P}$. We note that entries of each column must sum to one (i.e. $\sum_i \mathbf{P}_{ij} = 1$) since leaving state j the system must arrive at any of the states $i \in \mathcal{A}$.

The probability distribution of the states of a homogeneous Markov chain at time t denoted as $\boldsymbol{\pi}_t$ is recursively related to the transition matrix (Norris, 1998) via

$$\boldsymbol{\pi}_t = \mathbf{P}\boldsymbol{\pi}_{t-1}. \tag{4}$$

A stationary distribution $\boldsymbol{\pi}$ obeys

$$\boldsymbol{\pi} = \mathbf{P}\boldsymbol{\pi}\,, \tag{5}$$

which means $\boldsymbol{\pi}$ is an eigenvector of the transition matrix with eigenvalue one. A sufficient condition for the existence of a unique stationary distribution exists: it is enough for the non-negative matrix $\mathbf{P}$ to be primitive, i.e. $\mathbf{P}^m > 0$ for some $m \geq 1$ (see e.g. Horn and Johnson, 1990).

We mention that generalisations of the above introduced Markov chains are possible. Replacing Eq. (2) by

$$P(\sigma_{t+1}|\ldots,\sigma_{t-2},\sigma_{t-1},\sigma_t) = P(\sigma_{t+1}|\sigma_{t-k+1},\ldots,\sigma_{t-1},\sigma_t) \tag{6}$$

defines a so-called Markov chain of order k and comprises standard Markov chains as of order one. The order k determines the correlation range which can be shown by applying a specific information-theoretic measure: given the probabilities of n-words $P(\sigma_1,\ldots,\sigma_n)$ we can define n-block entropies (Ebeling and Nicolis, 1992)

$$H_n := -\sum_{(\sigma_1,\ldots,\sigma_n)\in\mathcal{A}^n} P(\sigma_1,\ldots,\sigma_n)\log_\lambda P(\sigma_1,\ldots,\sigma_n)\,. \tag{7}$$

as generalisations of Shannon's famous information measure (Shannon, 1948). Due to choosing the alphabet size λ as base of logarithms it is clear that $0 \leq H_n \leq n$. The quantity H_n represents the average amount of information necessary for a prediction (or gained after observation) of an n-word $(\sigma_1,\ldots,\sigma_n)$. From this it is obvious that the conditional entropy (Freund et al., 1996)

$$h_0 := H_1 \quad \text{and} \quad h_n := H_{n+1} - H_n \quad (n = 1,2,\ldots), \tag{8}$$

is nothing but the average information necessary when trying to predict the symbol σ_{n+1} given perfect knowledge of the prehistory $(\sigma_1,\ldots,\sigma_n)$. For a Markov chain of order k it can rigorously be shown (Gatlin, 1972) that

$$h_{n\geq k} = H_{k+1} - H_k = h \quad \text{(entropy of the source)} \tag{9}$$

which means that the profile of conditional entropies stagnates exactly at the Markov order k shaping a plateau at height h (cf. dotted line in Fig. 5 in Sect. 3.3). As seen in Fig. 5 when computing conditional entropies from finite length time series this ideal is affected by the estimation problem (Schürmann, 2004), which is actually an underestimation of the entropy caused by biased probabilities of n-words. The reason for these underestimated probabilities is the finite sample size and the fact that some n-words thus are too rare and too frequent. Nevertheless, the profile of h_n can be used to assess the order of a Markov process and how useful is the standard (i.e. first order) Markov approximation.

In practical applications one must estimate the entries of the transition matrix from anomalies, i.e. from deseasonalised and detrended time series. Denoting the observed frequency of transitions from states j at time t to state i at time $t+1$ by n_{ij} a standard estimator for the transition probabilities, when working with anomalies, is given by

$$\widehat{p_{ij}} = \frac{n_{ij}}{\sum_i n_{ij}}\,. \tag{10}$$

Notice that the presumed stationarity of anomalies corresponds to the assumption of a homogeneous Markov chain. In case the resulting empirical matrix $\widehat{\mathbf{P}}$ turns out to be primitive one can compute the related stationary distribution $\boldsymbol{\pi}$. Likewise, an empirical distribution $\widehat{\boldsymbol{\pi}}$ can be estimated from the data via

$$\widehat{\pi}_j = \frac{n_j}{\sum_j n_j} \tag{11}$$

where n_j denotes the observed frequency of state j. The observation that $\widehat{\boldsymbol{\pi}}$ is not significantly different from $\boldsymbol{\pi}$ provides additional support for the assumption of stationarity.

2.3 Markovian descriptors

The aim of our Markov chain analysis is to characterise the regional climate system through several descriptors derived from the estimated transition matrix $\widehat{\mathbf{P}}$. Following Hill et al. (2004) these descriptors can be formulated at the level of single states or the system as a whole. Below we will consider only the latter-mentioned option and further only the most important descriptors *persistence*, *recurrence time* and *entropy*:

2.3.1 Persistence

Persistence gives the probability that the system residing in state j at time t remains there at time $t+1$

$$P(\text{persistence}) = \sum_{j=1}^{\lambda} \widehat{p}_{jj}\widehat{\pi}_j . \tag{12}$$

For instance, a state that is dryer and warmer than the average seasonal condition will have the general tendency to return to the more humid and more temperate state of the seasonal average. The persistence is a measure for the fraction of deviations that do not leave the related cell in the following time step. It can also be interpreted in terms of the average residence time within an arbitrary cell, e.g. drier and warmer.

2.3.2 Recurrence time

The Smoluchowski recurrence time describes the average time elapsing between leaving a state j and then returning to it again. The recurrence time for state j is given by the ratio of the number of states $i \neq j$ and the number of unbroken blocks of states $i \neq j$. Kac (1947) elucidates this connection with an example. Suppose just having two states $X(t) \in [0,1]$ and, for instance, the following sequence of length 14:

$$10101010101010, \tag{13}$$

here the recurrence time for state 1 is the number of 0's divided by the number of unbroken blocks of 0's, which is $7/7 = 1$. Suppose another sequence is observed

$$100100100100100, \tag{14}$$

here we find 10 zeros and five unbroken blocks of zeros, thus computing a recurrence time for state 1 as $10/5 = 2$.

In accordance with the definition by Hill et al. (2004), the recurrence time for a single state is given by

$$\phi_j = \frac{1-\widehat{\pi}_j}{(1-\widehat{p}_{jj})\widehat{\pi}_j} . \tag{15}$$

Thus, for the whole system we can define

$$\langle\phi\rangle = \sum_{j=1}^{\lambda} \phi_j \widehat{\pi}_j \tag{16}$$

$$= \sum_{j=1}^{\lambda} \frac{1-\widehat{\pi}_j}{1-\widehat{p}_{jj}} . \tag{17}$$

The recurrence time identifies the time it takes between leaving the drier and warmer state before returning to it again. Just as the persistence it is related to the relaxation time, i.e. the time to return to the average seasonal state, but not identical with it, since the return to the average state may be followed by several deviations into quite different directions.

2.3.3 Entropy

As already mentioned above the Shannon information (Shannon, 1948) or entropy is a measure of unpredictability. The entropy of the Markov chain is defined by the expression:

$$H(\mathbf{P}) = \sum_{j=1}^{\lambda} \left\{ (-) \sum_{i=1}^{\lambda} \widehat{p}_{ij} \log \widehat{p}_{ij} \right\} \widehat{\pi}_j . \tag{18}$$

From the following identities

$$H(\mathbf{P}) = \left\langle \sum_{i=1}^{\lambda} \widehat{p}_{ij} \log \frac{1}{\widehat{p}_{ij}} \right\rangle_j \tag{19}$$

$$= \left[\sum_{i,j=1}^{\lambda} \widehat{\Pi}_{ij} \log \widehat{\Pi}_{ij}^{-1} \right] - \left[\sum_{j=1}^{\lambda} \widehat{\pi}_j \log \widehat{\pi}_j^{-1} \right] \tag{20}$$

$$= H_2 - H_1 = h_1 , \tag{21}$$

where $\Pi_{ij} = p_{ij}\pi_j$ denotes the joint probability (to observe state j at time t *and* state i at time $t+1$), we see that $H(\mathbf{P})$ is, in fact, the conditional entropy h_1 (cf. Eq. 8). Division by the maximum entropy $H_{\max}(\mathbf{P}) = \log\lambda$ (which is attained for the equidistribution $1/\lambda$) corresponds to computing logarithms to base λ (since $\log_\lambda x = \log x / \log\lambda$) and normalises the entropy, i.e.

$$0 \leq H_{\mathrm{r}}(\mathbf{P}) = \frac{H(\mathbf{P})}{H_{\max}(\mathbf{P})} \leq 1 . \tag{22}$$

Finally, the entropy is a measure for the lack of average predictability of the next anomalous deviation that follows the present one. Typically, the stagnation of an anomaly expressed by high persistence is a situation that is easy to predict. Therefore, in our comparison of the two climate regions we found a decrease of entropy, i.e. an increase of average predictability, with increasing persistence (cf. Figs. 7 and 9).

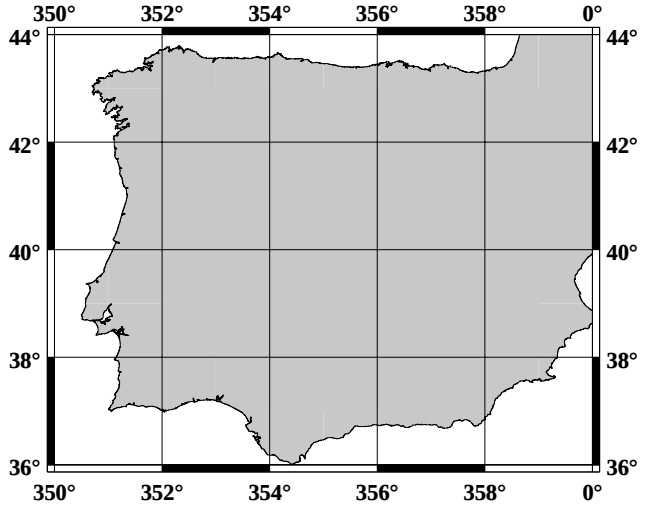

Fig. 1. Iberian Peninsula subdivided into 20 $2^\circ \times 2^\circ$ grid pixels.

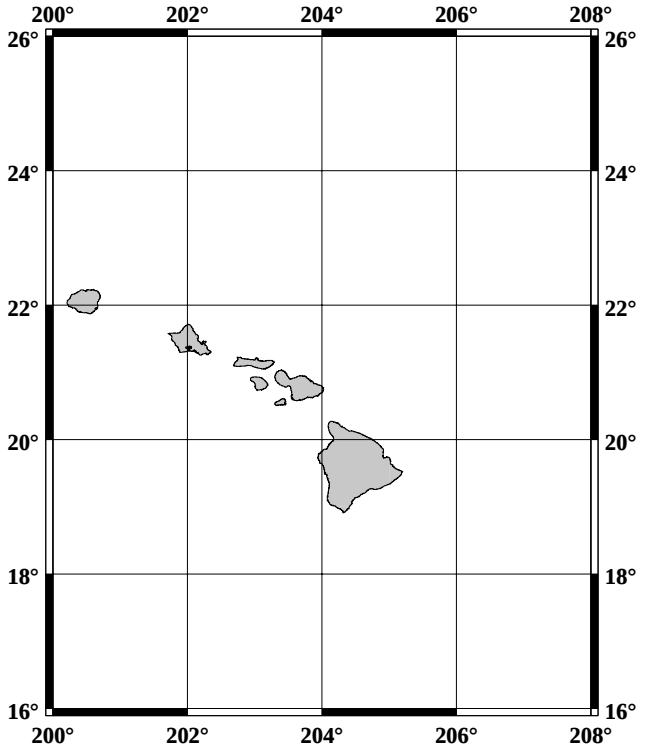

Fig. 2. A region centred over Hawaii subdivided into 20 $2^\circ \times 2^\circ$ grid pixels.

The above derived descriptors, which are new in atmospheric science, can be viewed as characteristics of a regional climate system. As we will explain in Sect. 3.5, their absolute values are strongly dependent on the choice of a partition. Therefore, we propose to use these values in a comparative analysis with the aim to differentiate climatic regions or to assess trends or sudden shifts across time. Of course, conclusive statements require an analysis of statistical significance. Since this is standard practice and can be achieved by surrogate simulations or resampling techniques as, for instance, jackknife and bootstrap methods, we do not detail it here but exemplify it in the next section.

Methodologically, the computation of descriptors means to condense the wealth of information contained in the λ^2 matrix entries $\widehat{p}_{ij}$ in comparatively few and well interpretable numbers. This is useful for subsequent classification and, quite generally, for gaining a better overview. For a more detailed analysis one can compute similar transition specific descriptors (Hill et al., 2004) or investigate statistically significant state or transition changes across a set of matrices.

3 Results – application to a climatological example

Regarding the Köppen classification, which is mainly based on temperature and humidity, we apply the above concepts to two temperature and water vapour data sets with the aim to classify two distinct climatic regions: the Iberian Peninsula (Fig. 1) and the Hawaiian Islands (Fig. 2). In future studies it is also planned to incorporate more climate variables and expand the analysis globally.

3.1 Data base

One pillar of the Markov chain analysis is constituted by measurements of the satellite instruments GOME (Global Ozone Monitoring Experiment) (Burrows et al., 1999) and SCIAMACHY (SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY) (Burrows et al., 1990, 1995; Bovensmann et al., 1999; Gottwald et al., 2006), which provide the total atmospheric water vapour column (unit: $\mathrm{g\,cm^{-2}}$) since 1995. The data have been retrieved by the AMC-DOAS method (Noël et al., 2004). The main advantages of the GOME/SCIAMACHY data are the independence from external information and the ability to retrieve water vapour also over land, which is e.g. not possible with microwave sensors like SSM/I (Special Sensor Microwave Imager) (e.g. Andersson et al. (2010)). The GOME/SCIAMACHY data are gridded on a global $0.5^\circ \times 0.5^\circ$ lattice and aggregated to monthly means. Based on this data a global trend study for the time span from 1996 to 2006 was performed by Mieruch et al. (2008).

Several global temperature products are available, e.g. the HadCRUT3 (on a $5^\circ \times 5^\circ$ grid) data from the University of East Anglia or the GISS (Goddard Institute of Space Studies) surface temperatures. In the present study the GISS data set is used and builds the second pillar of the analysis,

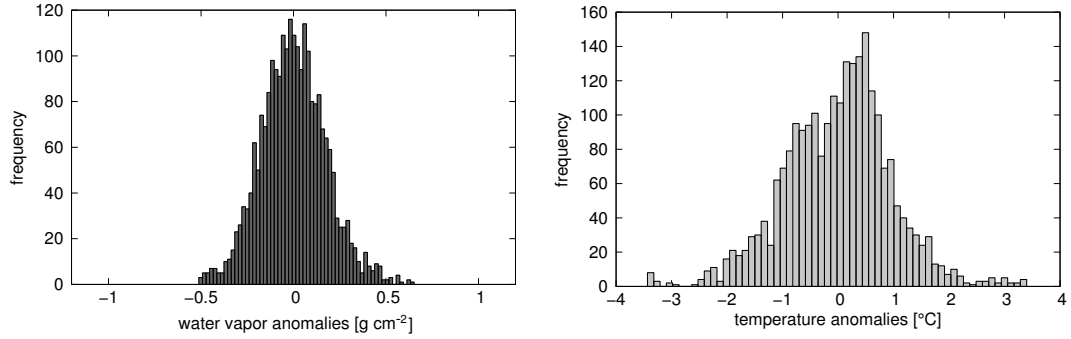

Fig. 3. By removing the trend and the seasonal components from monthly mean temperature and water vapour data we derive the water vapour and temperature anomalies. The frequency distributions of these anomalies for the time span from 1996 to 2005 for the region of the Iberian Peninsula are shown in the above figure. Altogether we gather for each climate parameter 2400 monthly anomalies for the considered region and time span.

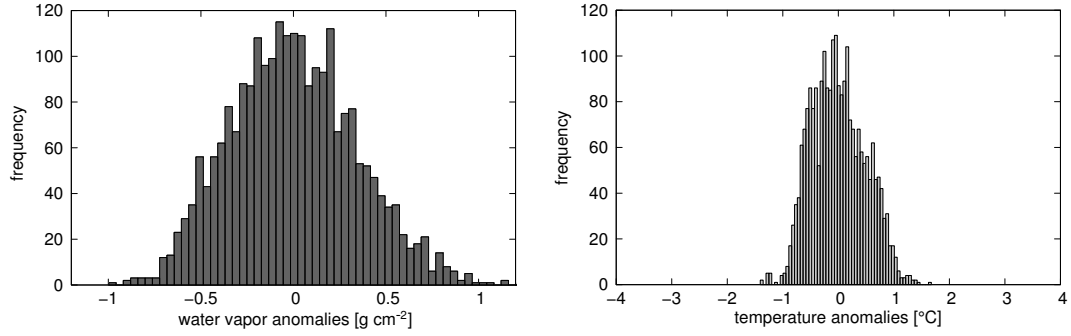

Fig. 4. Frequency distributions of 2400 water vapour and temperature anomalies from the islands of Hawaii. Similarly to Fig. 3 the anomalies have been derived from monthly mean data, where we removed the trend and the seasonal components.

because of its higher spatial resolution ($2^\circ \times 2^\circ$). The GISS data (Hansen and Lebedeff, 1987) are based on the Global Historical Climatology Network (GHCN), which comprises 7280 stations, the United States Historical Climatology Network (USHCN) with more than 1000 stations and the Scientific Committee on Antarctic Research (SCAR) with stations in Antarctica. The data sets are adjusted to the overlapping time span, which is from January 1996 to December 2005, and the GOME/SCIAMACHY data are gridded to a $2^\circ \times 2^\circ$ grid.

The Markov chain analysis is exemplary applied to the regions of the Iberian Peninsula and Hawaii, which are covered by 20 $2^\circ \times 2^\circ$ grid pixels (shown in Figs. 1 and 2), hence the analysis is based on 20 water vapour and temperature time series each consisting of 120 monthly mean measurements from 1996 to 2005. Thus, each region provides overall 2400 samples – the problem that these are not independent but spatially correlated will be dealt with by a resampling strategy. As mentioned in the introduction we are not interested in absolute values but rather in the interaction of the anomalies, which goes beyond systematic connections like seasonalities. As stated above, to obtain the anomalies we removed systematic terms, i.e. the linear trend plus the offset, and also the seasonal components, which were modelled as a Fourier series including annual and semiannual cycles.

3.2 Data processing and construction of the Markov chains

Employing the method of symbolic dynamics (cf. Sec. 2.1) the continuous-valued water vapour and temperature anomalies are mapped onto a sequence of discrete states labeled with symbols. The distributions of the water vapour ($v(t)$) and temperature ($\vartheta(t)$) anomalies from the Iberian Peninsula and Hawaii are shown in Figs. 3 and 4. Interestingly the water vapour anomalies from the Iberian Peninsula are smaller than those from Hawaii, whereas the temperature anomalies from the Iberian Peninsula are larger than the anomalies from Hawaii. This finding is important for a more detailed climatological interpretation of the results in the light of short term climate stability aspects, because the final partitioned sequences are independent of former absolute values.

As explained above (Sect. 2.1), the size of the partition λ (the number of cells in the state space spanned by all selected parameters) must be weighed against sample size N. Consider the case of P parameters. We might specify the partition by the vector $K = (k_1, \ldots, k_P)$ where k_i is the number of intervals used to segment the observed range of parameter i. This leads to a total of $\lambda = \prod_{i=1}^{P} k_i$ cells in the form of hypercubes. Of course, more sophisticated choices are thinkable, for instance, cell boundaries reflecting covariation of parameters, however, will not be considered here. In the present case we use the combination of temperature and water vapour ($P = 2$) and, moreover, will only consider $k_1 = k_2 = k$; therefore our $\lambda = k^2$.

From the consideration of our sample size $N = 2400$ it seems appropriate to map each data series onto two ($k = 2$) or three ($k = 3$) cells. It is recommended to choose adaptive partitions (Cellucci et al., 2005) which generates cells with equal probabilities (equipartition). However, we prefer to define cell borders for each parameter independently in order to maintain a more transparent physical interpretation. As a consequence we do not account for covariation of temperature and water vapour and, therefore, do not arrive at an exact equipartition (cf. Fig. 6).

For the twice binary partition we choose cell symbols:

$$v(t) \rightarrow \{l_v, h_v\}$$

$$\vartheta(t) \rightarrow \{l_\vartheta, h_\vartheta\}$$

and for the ternary partition we transform:

$$v(t) \rightarrow \{l_v, m_v, h_v\}$$

$$\vartheta(t) \rightarrow \{l_\vartheta, m_\vartheta, h_\vartheta\}.$$

Since we are interested in the interaction of both climate variables we merge the water vapour and temperature sequences to a single symbol sequence by combining both symbol sets, thus forming the alphabets $\mathcal{A}^2 = \{[l_v, l_\vartheta], [l_v, h_\vartheta], [h_v, l_\vartheta], [h_v, h_\vartheta]\}$ of size $\lambda = 4$ and $\mathcal{A}^3 = \{[l_v, l_\vartheta], [l_v, m_\vartheta], [l_v, h_\vartheta], [m_v, l_\vartheta], [m_v, m_\vartheta], [m_v, h_\vartheta], [h_v, l_\vartheta], [h_v, m_\vartheta], [h_v, h_\vartheta]\}$, which is of size $\lambda = 9$. Furthermore, we perform our calculations for finer partitions of size $k = 4, \ldots, 10$ to show the dependence on the coarse graining, although critical estimation problems arise for partition sizes larger than $\lambda = 9$.

3.3 The conditional entropy

As discussed in Sect. 2.2 the conditional entropy can be used to assess the order of a Markov chain. Consequently we have exemplarily estimated the h_n for the Hawaiian water vapour – temperature sequence for the four states alphabet $\mathcal{A}^2$ (sample size $N = 2400$) and show the result in Fig. 5 as filled circles connected with a line. In addition, the theoretical h_n for an independent sequence, a Bernoulli chain, (open squares connected with a line) and a first order Markov chain (open circles connected with a line) are depicted in Fig. 5. While, for infinite sample size, the Bernoulli chain gives rise to a constant profile the Markov chain would stagnate after $n = 1$ (the order of the Markov chain). Furthermore, we have simulated an ensemble of Bernoulli and first order Markov chains and plotted the ensemble averaged h_n as filled squares and filled triangles connected with lines, respectively. The thick grey curves represent the 2σ confidence intervals for the h_n of the simulated Bernoulli and Markov chains.

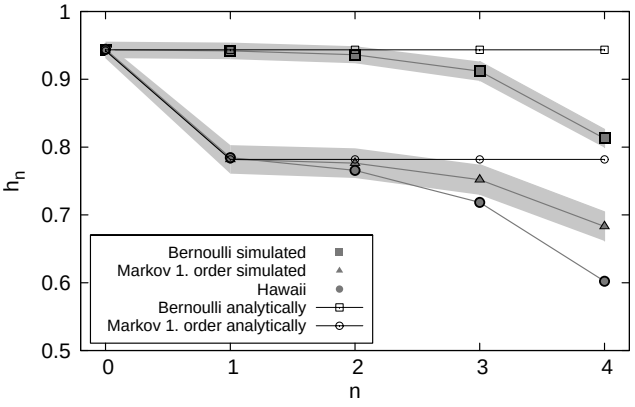

Fig. 5. Theoretical and estimated conditional entropies for several sequences.

For the given sample size 2400 the above mentioned length-effect can hardly be seen for $n = 1$. From $n = 2$ on the finite sample size progressively affects the entropy estimates. Nevertheless, Fig. 5 supports the idea, that the symbol sequence, constructed from water vapour and temperature measurements at a region around the islands of Hawaii, can quite adequately be described by a first order Markov chain.

3.4 Estimation of transition probabilities

In line with the homogeneity of the Markov chain, justified by considering anomalies, the transition probabilities are estimated via observed frequencies of symbol pairs across the complete time span from January 1996 to December 2005, which comprises 120 months. In addition to spatial and temporal auto-correlations of single variables, spatio-temporal cross-correlations between temperature and water vapour reflect their coupling through the underlying climatic processes. Estimating probabilities from the full space and time range of the Iberian Peninsula and Hawaii means to rely on stationarity and homogeneity of the interplay between these variables. We estimate the transition matrices $\widehat{\mathbf{P}}$ and the empirical distributions $\widehat{\boldsymbol{\pi}}$ across all states as detailed in Eqs. (10) and (11). Exemplarily, the 4×4 transition matrix $\widehat{\mathbf{P}}$ for the islands of Hawaii is given in Table 1.

Table 1. Transition matrix $\widehat{\mathbf{P}}$, estimated from 2380 transitions in 120 months for the Hawaiian islands. The entries $\widehat{p_{ij}}$ represent the transition probabilities of changing the state from column j to row i.

	l_v, l_ϑ	l_v, h_ϑ	h_v, l_ϑ	h_v, h_ϑ
l_v, l_ϑ	**0.617**	0.452	0.186	0.100
l_v, h_ϑ	0.200	**0.323**	0.051	0.070
h_v, l_ϑ	0.089	0.075	**0.279**	0.194
h_v, h_ϑ	0.094	0.151	0.485	**0.635**

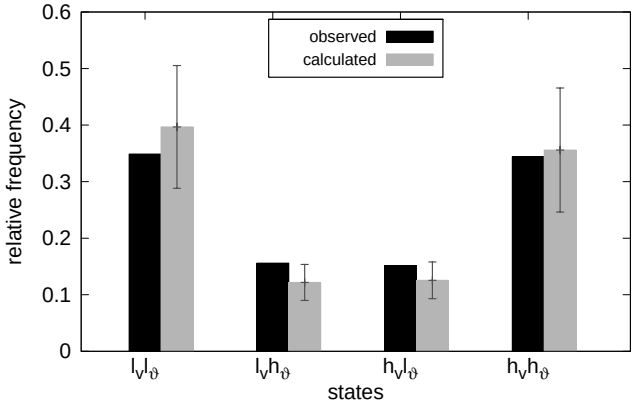

Fig. 6. Histogram of the observed ($\widehat{\pi}_j$) and calculated (π_j) relative frequencies of the states. The errorbars (2σ) have been estimated by a resampling technique.

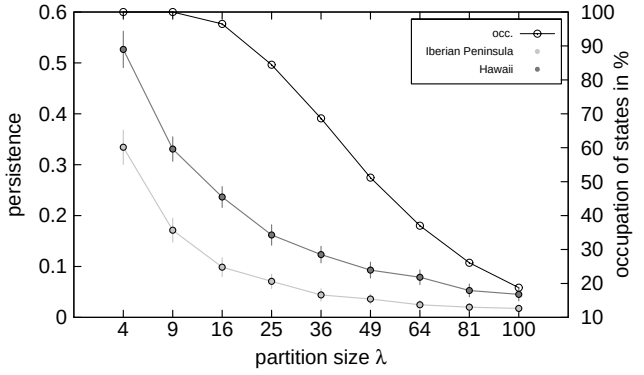

Fig. 7. Persistence of the regional climate systems of Hawaii (dark grey) and the Iberian Peninsula (light grey). From partition size $\lambda = 16$ full occupation of the states of the transition matrix cannot be fulfilled.

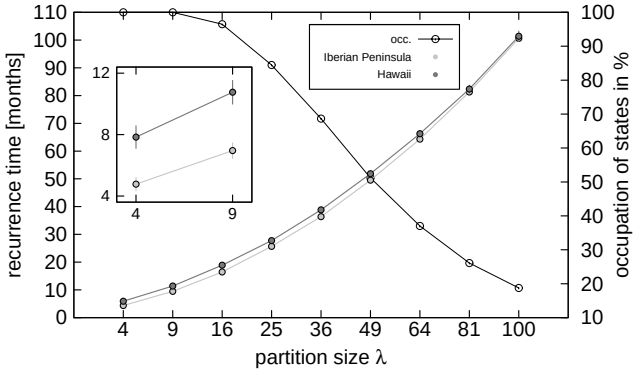

Fig. 8. Recurrence time of the regional climate systems of Hawaii (dark grey) and the Iberian Peninsula (light grey). From partition size $\lambda = 16$ full occupation of the states of the transition matrix cannot be fulfilled.

Finally, the observed relative frequencies $\widehat{\pi}_j$ of the states (cf. Eq. 11) can be compared with the calculated equilibrium distributions of the states π_j, which are derived via $\boldsymbol{\pi} = \widehat{\mathbf{P}}\boldsymbol{\pi}$. Figure 6 shows the relative frequencies (observed: black and calculated: grey) of the states as a histogram.

As can be seen, the differences between the observed and estimated frequencies are marginal, thus the time and space homogeneous system of water vapour and temperature of the islands of Hawaii has already reached the stationary distribution. In the sense of the Markov chain analysis this result is observed under the decisive assumption of temporal homogeneity, which implies a constant transition matrix over time.

3.5 Markovian descriptors

According to Sect. 2.3 the Markovian descriptors have been calculated for the two regional climate systems and the results are shown in Figs. 7, 8 and 9. We have calculated the descriptors for several partition sizes, where for instance $\lambda = 4$ results from combining binary water vapour and temperature sequences. Additionally, we show the percentage of occupied entries exemplarily for the Hawaiian transition matrix. 100% means that all λ^2 transitions were observed at least once, while for increasing $\lambda \geq 16$ the fraction of unobserved transitions increases.

When viewing Figs. 7, 8, 9, it becomes clear that all three descriptors show a pronounced dependence on the partition size. Increasing the partition size leads to choosing more cells and a smaller maximum cell diameter. From this it follows immediately that the persistence will decrease – the shrinking cell diameter is traversed in shorter time – the recurrence time will increase – returning to a smaller cell takes longer because nearby trajectory segments miss it – and the normalised entropy will decrease – the number of observed transitions grows slower than the number of all possible transitions due to refinement. These statements elucidate why absolute values of our descriptors cannot be easily interpreted. Persistence and the recurrence time are continuously decreasing/increasing, respectively, with more refined partitions. The entropy profile of the Hawaiian

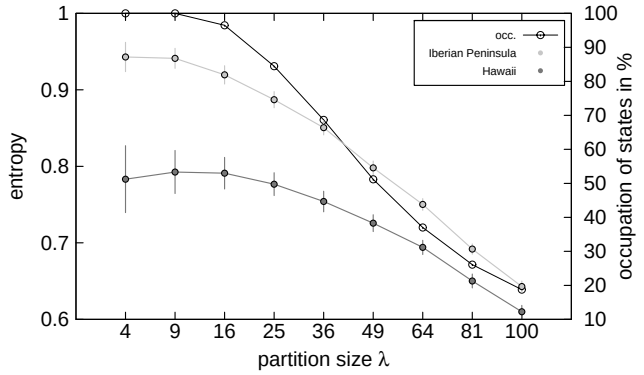

Fig. 9. Entropy of the regional climate systems of Hawaii (dark grey) and the Iberian Peninsula (light grey). From partition size $\lambda = 16$ full occupation of the states of the transition matrix cannot be fulfilled.

islands exhibits a very shallow maximum near partition size $\lambda = 9$, however, the significance is not guaranteed. Still it might be tempting to define an optimal partition through an entropy maximum criterion.

As already mentioned, our aim is to apply the Markov chain analysis in a comparative way, which means that differences between descriptors should have the potential to discriminate two or more regions. The persistence of the Markov chain of the Hawaiian system is shown in Fig. 7 as dark grey filled circles together with two times the standard error. The estimation of the errors of the descriptors is shown in the following Sect. 3.6. The persistence from Hawaii is systematically significantly larger than the persistence from the Iberian Peninsula (light grey). Regarding Fig. 8 the recurrence time of the Hawaiian system (dark grey) is larger then the recurrence time of the system of the Iberian Peninsula (light grey). The difference is significant, as can be seen in the blow up for partition sizes $\lambda = 4$ and $\lambda = 9$, where the errorbars (two times the standard errors) are significantly separated. Finally a significant difference between Hawaii and the Iberian Peninsula has been detected with the entropy descriptor in Fig. 9.

3.6 Significance of the descriptors

The usefulness of the descriptors for differentiating between regional climates relies on finding significant differences. We base our significance analysis of descriptors for the Iberian Peninsula and Hawaii on simulated surrogate data. To this end we use the algorithm MIAAFT (Multivariate Iterated Amplitude Adjusted Fourier Transform), which was developed by Schreiber and Schmitz (2000). In a comparative study (Venema et al., 2006) the univariate IAAFT turned out to be optimal. The MIAAFT algorithm shuffles the original data, thus exactly preserving the original distribution. In addition it preserves also the auto- and cross-correlation structure of our multivariate time series. In the following we simulated an ensemble of 100 multivariate data sets and estimated descriptors from these surrogate time series. The related standard deviations of this ensemble provide estimates for the standard errors of the descriptors, which are used in Figs. 7, 8 and 9, where we have plotted two times the standard error. As can be seen, the error bars are far separated, hence the differences between the Hawaiian descriptors and the descriptors from the Iberian Peninsula are significant.

4 Conclusions and outlook

The two climate regions are significantly different in all three descriptors. This is not at all self-evident as one might think when relating to obvious differences in seasonal behaviour. For an interpretation of the above differences it is important to keep in mind that anomalies are *departures* from seasonal behaviour and that our descriptors thus measure dynamical features of the relaxation process. Additionally it has to be kept in mind that we are analysing climate states and not the variables themselves. Furthermore, partitioned data are independent from absolute values, which are also important for an interpretation of the results regarding short term stability and susceptibility to e.g. external forcings. The fact that the Hawaiian region has larger persistence, larger recurrence time and smaller entropy than the Iberian Peninsula is a reflection of the fact that departures from normal behaviour return with a larger relaxation time.

Regarding the famous Köppen climate classification (Kottek et al., 2006) the two regions are assigned to different climate types, i.e. the Iberian Peninsula is classified as warm/steppe/hot summer (south)/warm summer (north), whereas Hawaii is warm/fully humid/hot summer. As can be seen from the Köppen classification the main difference between the Iberian Peninsula and Hawaii is the humidity and this is captured in our analysis including the water vapour data. Because the average climate of the Hawaiian Islands and the Iberian Peninsula are apparently different these differences may not be surprising. However, it is not so clear in advance in which direction these climatic regions differ with respect to dynamical aspects of anomalies.

Our findings offer the opportunity to develop a new supplementary global climate classification scheme, which we name *dynamic climate classification*, on the basis of real measurements in the form of condensed Markovian descriptors. The incorporation of more climate parameters such as clouds, precipitation and vegetation into the Markov chain analysis is, of course, desirable. However, in view of the finite sample size problem it might be advisable to perform the analyses of many parameters with just a few principle components (e.g. via a PCA). In addition to existing climate classifications, which use absolute climate values,

our *dynamic climate classification* describes the interplay or coupling of climate variables through climate states and is thus suitable for a short term climate stability discussion.

Interpreting the seasonal dynamics as a limit cycle this means that the Hawaiian short term climate is less stable than that of the Iberian Peninsula. This conclusion is bit surprising since naively one might expect a stabilising influence of the Pacific Ocean. The concept of local stability that is typically investigated in nonlinear dynamics might be or not be adequate to characterise the stability of climate regions. An extended duration of a deviation in climate state space that corresponds to an anomaly from the seasonal cycle will be no problem as long as the system (sooner or later) returns to the seasonal average. Of course, longer lifetimes of anomalies are typically accompanied by larger excursions which might well trigger catastrophic events in ecosystems (e.g. extinction, starvation) and which, in turn might feed back on climate dynamics. However, even more pronounced might be supercritical anomalies causing a climatic regime shift even without feedback of the ecosystem. In such a case the system does not return to the previous seasonal cycle but switches to another mode. Whether such tipping points exist is a prominent question in current climate research (Lenton et al., 2008). In this context, in a sliding-window analysis our proposed descriptors might gain importance as early-warning signals for critical transitions (Scheffer et al., 2009).

Acknowledgements. SCIAMACHY is a national contribution to the ESA ENVISAT project, funded by Germany, the Netherlands and Belgium. SCIAMACHY data are provided by ESA. This work is funded by DLR-Bonn and by the University of Bremen. Temperature data are provided by Goddard Institute of Space Studies. We thank two Referees for their constructive comments.

Edited by: W. Hsieh
Reviewed by: R. V. Donner and another anonymous referee

References

Andersson, A., Fennig, K., Klepp, C., Bakan, S., Graßl, H., and Schulz, J.: The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data - HOAPS-3, Earth Syst. Sci. Data Discuss., 3, 143–194, doi:10.5194/essdd-3-143-2010, 2010.

Beck, C., Grieser, J., Kottek, M., Rubel, F., and Rudolf, B.: Characterizing Global Climate Change by means of Köppen Climate Classification, Klimastatusbericht, 139–149, 2005.

Blüthgen, J. and Weischelt, W.: Allgemeine Klimageographie, Walter de Gruyter, 1980 (in German).

Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., and Goede, A. H. P.: SCIAMACHY-Mission objectives and measurement modes, J. Atmos. Sci., 56, 127–150, 1999.

Burrows, J. P., Schneider, W., Geary, J. C., Chance, K. V., Goede, A. P. H., Aarts, H. J. M., de Vries, J., Smorenburg, C., and Visser, H.: Atmospheric remote sensing with SCIAMACHY, Digest of Topical Meeting on Optical Remote Sensing of the Atmosphere, Optical Society of America, Washington, 4, 71–74, 1990.

Burrows, J. P., Hölzle, E., Goede, A. P. H., Visser, H., and Fricke, W.: SCIAMACHY – Scanning imaging absorption spectrometer for atmospheric chartography, Acta Astronaut., 35, 445–451, 1995.

Burrows, J. P., Weber, M., Buchwitz, M., Rozanov, V., Ladstätter-Weißenmayer, A., Richter, A., de Beek, R., Hoogen, R., Bramstedt, K., Eichmann, K.-U., Eisinger, M., and Perner, D.: The Global Ozone Monitoring Experiment (GOME): Mission Concept and First Scientific Results, J. Atmos. Sci., 56, 151–175, 1999.

Cellucci, C. J., Albano, A. M., and Rapp, P. E.: Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms, Phys. Rev. E, 71, 066208, doi:10.1103/PhysRevE.71.066208, 2005.

Daw, C. S., Finney, C. E. A., and Tracy, E. R.: A review of symbolic analysis of experimental data, Rev. Sci. Instrum., 74, 916–930, 2003.

Ebeling, W. and Nicolis, G.: Word frequency and entropy of symbolic sequences: a dynamical perspective, Chaos Soliton. Fract., 2, 635–650, doi:10.1016/0960-0779(92)90058-U, available at: http://www.sciencedirect.com/science/article/B6TJ4-4C7WKT1-5/2/5bc140e76179003ff45a6795e6ff14d5, 1992.

Flohn, H.: Zur Frage der Einteilung der Klimazonen, Erdkunde archive for scientific geography, 11, 161–175, 1957 (in German).

Freund, J., Ebeling, W., and Rateitschak, K.: Self-similar sequences and universal scaling of dynamical entropies, Phys. Rev. E, 54, 5561–5566, 1996.

Freund, J. A., Pöschel, T., and Wiltshire, K. H.: Markovsche Analyse nasser Gemeinschaften, Logos, Berlin, 99–110, 2006.

Gatlin, L. L.: Information Theory and the Living System, Columbia Univ. Press, New York, 1972.

Gerstengarbe, F.-W. and Werner, P. C.: Katalog der Großwetterlagen Europas (1881–1998) Nach Paul Hess und Helmuth Brezowsky, 5. verbesserte und ergänzte Auflage, 1999 (in German).

Gottwald, M., Bovensmann, H., Lichtenberg, G., Noël, S., von Bargen, A., Slijkhuis, S., Piters, A., Hoogeveen, R., von Savigny, C., Buchwitz, M., Kokhanovsky, A., Richter, A., Rozanov, A., Holzer-Popp, T., Bramstedt, K., Lambert, J.-C., Skupin, J., Wittrock, F., Schrijver, H., and Burrows, J.: SCIAMACHY, Monitoring the Changing Earth's Atmosphere, DLR, 2006.

Hansen, J. E. and Lebedeff, S.: Global trends of measured surface air temperature, J. Geophys. Res., 92(D11), 13345–13372, doi:10.1029/JD092iD11p13345, 1987.

Hill, M. F., Witman, J. D., and Caswell, H.: Markov Chain Analysis of Succession in a Rocky Subtidal Community, Am. Nat., 164, E46–E61, 2004.

Holdridge, L. R.: Determination of world plant formations from simple climatic data, Science, 105, 367–368, 1947.

Horn, R. A. and Johnson, C. R.: Matrix Analysis, Cambridge University Press, 1990.

Kac, M.: On the notion of recurrence in discrete stochastic processes, B. Am. Math. Soc., 53, 1002–1010, 1947.

Kantz, H. and Schreiber, T.: Nonlinear Time Series Analysis, Cambridge University Press, Cambridge, UK, 2004.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World Map of the Köppen-Geiger climate classification updated, Meteorol. Z., 15, 259–263, doi:10.1127/0941-2948/2006/0130, 2006.

Lauer, W. and Bendix, J.: Klimatologie, Westermann, Germany, 1993 (in German).

Lauer, W. and Frankenberg, P.: Klimaklassifikation der Erde, Geograpische Rundschau, Westermann Verlag, 40 pp., 1988.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., and Schellnhuber, H. J.: Tipping elements in the Earth's climate system, P. Natl. Acad. Sci. USA, 105, 1786–1793, doi:10.1073/pnas.0705414105, 2008.

Lind, D. and Marcus, B.: An Introduction to Symbolic Dynamics and Coding, Cambridge University Press, 1996.

Lorenz, K.: Das Wirkungsgefüge der Natur und das Schicksal des Menschen, Piper, München, 1983 (in German).

Mieruch, S., Noël, S., Bovensmann, H., and Burrows, J. P.: Analysis of global water vapour trends from satellite measurements in the visible spectral range, Atmos. Chem. Phys., 8, 491–504, doi:10.5194/acp-8-491-2008, 2008.

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

Noël, S., Buchwitz, M., and Burrows, J. P.: First retrieval of global water vapour column amounts from SCIAMACHY measurements, Atmos. Chem. Phys., 4, 111–125, doi:10.5194/acp-4-111-2004, 2004.

Norris, J. R.: Markov Chains, Cambridge University Press, 1998.

Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci., 11, 1633–1644, doi:10.5194/hess-11-1633-2007, 2007.

Scheffer, M. and Carpenter, S. R.: Catastrophic regime shifts in ecosystems: linking theory to observation, Trends Ecol. Evol., 18, 648–656, 2003.

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., van Nes, E. H., Rietkerk, M., and Sugihara, G.: Early-warning signals for critical transitions, Nature, 461, 53–59, doi:10.1038/nature08227, 2009.

Schreiber, T. and Schmitz, A.: Surrogate time series, Physica D, 142, 346–382, doi:10.1016/S0167-2789(00)00043-9, 2000.

Schürmann, T.: Bias Analysis in Entropy Estimation, J. Phys. A-Math. Gen., 37, L295–L301, 2004.

Shannon, C. E.: A Mathematical Theory of Communication, Bell Syst. Tech. J., 27, 623–656, 1948.

Venema, V., Bachner, S., Rust, H. W., and Simmer, C.: Statistical characteristics of surrogate data based on geophysical measurements, Nonlin. Processes Geophys., 13, 449–466, doi:10.5194/npg-13-449-2006, 2006.