Chapter 3

Symbol sequence analysis of climatic time signals

3.1 Introduction

The question of whether or not there are climatic attractors has evoked considerable interest and debate in the past two decades [45, 96]. Time-series analysis [1, 5, 23, 46, 65] have, in recent years, greatly helped in understanding complex temporal phenomena underlying climate phenomena, providing insight into issues such as whether, for instance, the data is indicative of underlying chaotic dynamics, whether the variables arise from a low dimensional attractor etc. There are host of unresolved issues in this analysis, some of which, for instance, whether the attractor has a fractal dimension which can be accurately inferred from the time-series data, what the utility of such an inference is, namely can this be used in modeling climate, etc. Studies in this regard have exploited several robust methods to estimate quantities such as fractal dimensions, Lyapunov exponents or the Lyapunov dimension from time-series data [33, 44, 100, 133]. The reconstruction of chaotic attractor or its crucial properties from time signals has been the subject of intensive study [59, 65, 100, 122].

In ecology and climatology, it is also relevant to address the following related issue: given two or more time signals obtained experimentally, can one reliably ascertain whether or not they arise from the same dynamical system? Recent studies of chaotic time signals which arise from model sys-
tems have attempted to devise techniques to detect the similarity or the inherent dynamic coupling even though the individual signals may be unpredictable and/or chaotic. Further, for dynamically coupled signals, can delay information be extracted? Such information crucially guides the process of model-building.

As described in chapter 2, Lehrman et al. [68] provided a partial answer to these questions. The time signals obtained from different variables in the Lorenz system are converted to symbol sequences and application of information-theoretic concepts is made to show that they come from the same dynamical system even though cursory examination does not reveal any dynamical coupling among them.

Our study is also similarly motivated. Here we make use of the symbolic analysis [68, 112, 113] directed toward establishing whether different climatic time-signals have their origin in the same underlying dynamics, and the extent to which they are mutually coupled. Studies have suggested that there may be low dimensional attractors that underlie the climate system [45, 96]. We extend the utility of the entropic method in analyzing the relative coupling strength of two variables. For dynamically coupled variables, say, $X$ and $Z$, the dependence of $X$ on $Z$ may not be the same as dependence of $Z$ on $X$. This difference can be seen by calculating the conditional entropies $E(X/Z)$ and $E(Z/X)$ for relative shifts of $X$ & $Z$ with respect to each other. This application is also found to be of help in deducing delays intrinsically present in such coupled systems.

Coarse-graining the dynamics of the phase space, in conjunction with information-entropic analysis can allow for inference of some of the crucial issues described above. This method has potential applications in the analysis of real data since it is frequently not known if there is indeed an underlying deterministic dynamical system. Dynamical correlations, determined by computing the correlation coefficients, or the cross-correlation function $C_{XZ}(\tau) = \langle X(t)Z(t+\tau) \rangle$ of signals $X(t)$ and $Z(t)$ are not always clear indicator with regard to issues raised above, and the application to climate data substantiates this fact. Our analysis shows that whereas the conventional correlation methods do indicate the dynamical correlations, it is often not possible to know the dynamical coupling of aperiodic or chaotic variables from such methods. Moreover, conventional methods give no indication of the relative coupling strength, in contrast to the present coarse-graining method. Thus the approach adopted here seems robust and reveals the extent to
which different variables are dynamically coupled. The present method is a ‘minimalist’ approach in extracting useful information from time signals.

3.2 Methodology

In the study of aperiodic or chaotic time signals arising from deterministic dynamical systems, we have observed that the Pearson correlation coefficient or the cross-correlation function is not an adequately sensitive detector of dynamical correlations. Application to climate data presents the similar picture. In contrast, the coarse-graining analysis is able to extract some of the crucial properties of such systems. In this section we briefly review the methods that we adopt for subsequent analysis.

Consider the time-series, $X(i), Z(i)$, with $i = 1, 2, \ldots N$ being a discrete time index. The Pearson correlation coefficient is

$$r_{XZ} = \frac{1}{(N-1)\sigma_X\sigma_Z} \sum_{i=1}^{N} (X(i) - \overline{X})(Z(i) - \overline{Z}),$$  \hspace{1cm} (3.1)

where $\overline{X}$ and $\overline{Z}$ are their respective means and $\sigma_X, \sigma_Z$ are the standard deviations. When $r = 1$, the quantities are perfectly positively correlated, for $r = -1$, the quantities are perfectly negatively correlated, while if $r = 0$, the two variables are not related.

The cross-correlation function, on the other hand, is

$$C_{XZ}(n_0) = \langle X(n)Z(n+n_0) \rangle,$$  \hspace{1cm} (3.2)

where $n_0$ is relative shift, and $\langle \rangle$ denotes an average over the sample. If the variables are (anti)correlated, then a (dip) peak appears in $C_{XZ}(n_0)$ as a function of $n_0$, with the location of the peak or dip giving an indication of whether there is any time–delay in the correlation or anticorrelation.

We apply the coarse-graining method as explained in Chapter 2. We present some new results taking cue from the very basic understanding of the conditional information entropy Eq. (2.4), which can not only be used to understand the dynamical correlation but also the conditional dependence of one variable on the other. This is accomplished by observing both $E(Z/X)$ and $E(X/Z)$. For making such a comparison, we do a minor change in this formula by noting that $N_t$ now stands for total number of all possible $\ell_X$ or $\ell_Z$ values.
$E(Z/X)$ thus represents the information about the values that $Z$ is constrained to take if $X$ takes a particular value and is not necessarily the same as $E(X/Z)$; the conditional probabilities $P(\ell_Z/\ell_X)$ and $P(\ell_X/\ell_Z)$ can differ, depending upon the relative dependences of two variables on each other. This is in turn determined by the coupling parameters. A detailed study of the Lorenz and other systems [13] examines the manner in which the change of parameter values can lead to asymmetric relative dependences.

As described in detail in chapter 2, the coarse-graining method can be successfully used to extract delay information from time signals and another useful context here can be to examine delayed coupling of two variables, say when the present value of $X$ is determined by a previous value of $Z$, the conditional entropy is minimised when the lag matches the delay.

### 3.3 Application & Results

Lehrman et al. [68] show that the conditional entropy $E(Z/X)$ for the Lorenz attractor has a sharp minimum for relative shift $n_0 = 0$ (see Fig. 2.2). We find such variation to depend on the order: $E(X/Z) \neq E(Z/X)$ (Fig. 3.1-3.2).

As discussed above, such differences arise because of differential coupling strength of parameters in the model. To explore further, we generated $X$ and $Z$ signals by setting different values of the parameters and searching for the parameters which shows three distinct possibilities of such a dependence: Fig. 3.1 shows a greater dependence of $X$ on $Z$ than $Z$ on $X$ for parameter values $a = 10, b = \frac{8}{3}$ and $r = 28$; reverse trend can be seen in Fig. 3.2 for parameter values $a = 10, b = 1$ and $r = 28$ and nearly a symmetric dependence for $a = 6, b = 1$ and $r = 24$. Obviously such a study can help in developing a model where the degree of mutual dependence can be well anticipated. This result seems interesting in the context of climate data analysis where the knowledge of relative dependence of variables would help immensely in understanding the climate dynamics and would be of great use in the process of model building.

Further we made a comparison vis-a-vis the conventional correlation methods. We show in Fig. 3.3 the plot of cross-correlation function $C_{XZ}(n_0) = \langle X(t)Z(t+n_0) \rangle$ for the $X$ and $Z$ time signals of Rössler and Lorenz equations. While $X$ and $Z$ signals show correlation for the Rössler system, the same do not appear to be correlated for the Lorenz system. The Pearson
Figure 3.1: The conditional entropies (a) $E(Z/X)$ and (b) $E(X/Z)$ as a function of shift parameter $n_0$ of $X$ and $Z$ time signals of Lorenz model Eqs. (1.2)-(1.4).

Figure 3.2: The conditional entropies (a) $E(Z/X)$ and (b) $E(X/Z)$ of $X$ and $Z$ time signals of Lorenz equations (1.2)-(1.4) with parameters $a = 10, b = 1$ and $r = 28$. 
Figure 3.3: Cross-correlation function $C_{XZ}$ as a function of shift parameter $n_0$ corresponding to $X$ and $Z$ signals of (a) Rossler model and (b) Lorenz model.

correlation coefficient (see chapter 3) for $X$ & $Z$ signals of Rossler system is $\approx 0.49$ and for Lorenz it is $\approx 0.02$.

To study the dynamical coupling of climate variables at a given location, meteorological data recorded at three-hourly intervals in the period January, 1988 to December, 1992 in Delhi are analyzed. These are respectively the temperature($T, ^\circ C$), Fig. 3.4(a), the relative humidity($Rh, \%$), Fig. 3.4(b), the rainfall($Rn, mm$), Fig. 3.4(c), the cloud cover($C, oktas$), Fig. 3.4(d), and the wind speed ($W, km/h$), Fig. 3.4(e) (data shown for 312 days, beginning January, 1988). Although one may anticipate some dynamical correlations among such variables, nothing definitive can be inferred by looking at the grossly obvious features of such time signals.

To understand the influence of climate at one location on the climate of a neighbouring station or a far away station, the coarse-graining methodology is applied to meteorological data recorded at 3 hour interval from January, 1988 to December, 1992 of four different stations in the Indian subcontinent: Delhi, Jaipur, Sundernagar and Chennai. Jaipur is located $\approx 300$ km south
Figure 3.4: Time records of (a) temperature (°C), (b) relative humidity (%), (c) rainfall (in mm), (d) cloud cover (oktas) and (e) wind speed (km/h) of Delhi.
west, Sundernagar ≈ 400 km north and Chennai ≈ 2000 km south of Delhi. Two of the stations are close to Delhi and one far away from it and they have been selectively chosen to see how the coupling of climate systems are affected with distance. The three variables taken for analysis are temperature(°C), relative humidity(%) and rainfall (mm).

The time series of the variables are discretized as described in Chapter 2 with \( \tau' = 15 \).

The correlation coefficient \( r_{XZ} \), where \( X \) and \( Z \) are pairs of variables taken out of the above climate variables of Delhi, is calculated. A significant correlation is seen for the pairs like \( T, Rh \), and \( W, Rh \), with \( r_{TRh} = -0.56, r_{WRh} = -0.41 \), while for some pairs the correlation coefficients are very small, \( r_{CW} = 0.12, r_{TC} = 0.18, r_{RnRh} = 0.15 \) and \( r_{RnC} = 0.18 \).

We also computed the cross-correlation function \( C_{XX}(n_0) \) for all above pairs. \( C_{TRh}, C_{RWh}, C_{CW} \) and \( C_{TC} \) show oscillations with shift \( n_0 \), Fig. 3.5(a) shows variation in \( C_{TC} \) with relative shift and similarly for others. \( C_{RnRh} \) (Fig. 3.5(b)) and \( C_{RnC} \) show a maximum at \( n_0 = 0 \).

In the coarse-graining analysis, we adopt the binary symbolic language \( (m = 2 \) and \( L = 5 \)) and determine the critical point for temperature as \( \{T_0 = T_{min}, T_1 = 28.5, T_2 = T_{max}\} \), \( T_{max} \) and \( T_{min} \) being the maximum
and minimum values of \{T\}. Similarly, the critical points for relative humidity, rainfall, cloud cover and wind speed are \(Rh_1 = 66, Rn_1 = 0.1, C_1 = 1\) and \(W_1 = 3\) respectively, the two other critical points for each are their respective maximum and minimum values. The conditional entropy \(E(Z/X)\), where \{Z\} and \{X\} are variables taken out of the set \{T, Rh, Rn, C, W\} is calculated for time shifts \(n_0\).

Some of the results of the above analysis are shown in Figs. 3.6(a-d), for the change in conditional entropy with respect to time shift \(n_0\) for \(E(Rh/T)\), \(E(W/Rh)\), \(E(C/T)\) and \(E(W/C)\) respectively. The presence of a clear minimum around \(n_0 = 0\) shows the dynamical coupling of the pairs of variables, in contrast to the oscillations typically seen for the cross-correlation functions.

Fig. 3.7(a) & (b) shows the variation in conditional entropy for \(E(Rh/Rn)\) and \(E(C/Rn)\) respectively. As can be expected, the relative humidity & rainfall or cloud cover & rainfall are significantly coupled variables and a prominent dip in conditional entropy at \(n_0 = 0\) validates this fact. As seen
in Fig. 3.5(b) above, the cross-correlation function $C_{Rn/Rh}$ shows a maximum peak at shift $n_0 = 0$ and similar result is observed for $C_{Rn/C}$.

A sharp fall in conditional entropy at or around shift $n_0 = 0$ implies that the variables are dynamically coupled and are governed by the same underlying dynamics. Further, it is worth noting that the variation in conditional entropy around shift $n_0 = 0$ is greater for some pairs of variables; for example, the minimum of $E(Rh/Rn)$ at shift $n_0 = 0$ is greater than that of $E(Rh/T)$, implying thus, the rainfall and relative humidity are more strongly coupled than are temperature and relative humidity. In some instances it is possible that that variation in conditional entropy is not very prominent, for example, $E(T/Rn)$ (not shown here).

We calculated the conditional entropy for the above pairs now with the order reversed. We observe a distinct minimum conditional entropy for $E(T/Rh), E(Rh/W), E(T/C), E(C/W)$ and $E(Rn/C)$. We don’t observe a clear minimum for $E(Rn/Rh)$ implying that every occurrence of high relative humidity does not result in rainfall. The present analysis can therefore be used as a tool in building up models with some idea of the relative coupling.
Figure 3.8: Gaussian noise of $SNR=2$ is added to time signals of rainfall and relative humidity to calculate (a) correlation function and (b) conditional entropy as a function of the shift $n_0$. Further the noise level is increased to $SNR=1$ to calculate the (c) correlation function and the (d) conditional entropy.

between different variables.

The robustness of this method in comparison to standard cross-correlation analysis can be seen by adding noise to the data. Gaussian noise of strength characterized by the signal to noise ratio, $SNR=\sqrt{\sigma_s^2/\sigma_n^2}$, where $\sigma_s$ is the standard deviation of the noise free time series and $\sigma_n$ that of noise, is added to the rainfall and relative humidity data and the variation in cross-correlation function and the conditional entropy at shift $n_0 = 0$ observed as a function of $SNR$. Fig. 3.8 shows the results for $SNR=2$ and $SNR=1$. At $SNR=1$ (Figs. 3.8(c) & (d)) ($\sigma_{Rh} \sim 22.69, \sigma_{Rhnoise} = \sigma_{Rh}, \sigma_{Rn} \sim 12.69$ (average taken over nonzero values), $\sigma_{Rnnoise} = \sigma_{Rn}$), while the maximum of $C_{RnRh}$ at $n_0 = 0$ vanishes (see Fig. 3.8(c)), the minimum of $E(Rh/Rn)$ still persists (Fig. 3.8(d)).

The next course of our analysis examines the coupling of the variables of a station with that of the other selected stations as mentioned above. Fig. 3.9
Figure 3.9: Time records of (a) rainfall (in mm) at Delhi, (b) relative humidity (%) of Jaipur, (c) temperature (°C) of Sundernagar and (d) temperature (°C) of Chennai.

(a), (b), (c) and (d) are the time records of data of rainfall of Delhi, relative humidity of Jaipur, temperature of Sundernagar and temperature of Chennai for 312 days, beginning January, 1988. With respect to dynamical coupling of some of such variables, again nothing much can be inferred from the gross features of the signals.

For simplicity Delhi, Jaipur, Sundernagar and Chennai are denoted D, J, S and C. The critical points $T_1$ for Jaipur, Sundernagar and Chennai are determined as 26.8, 23 and 28 respectively, the two other critical points for each are their respective maximum and minimum values. The critical points $Rh_1$ for relative humidity of Jaipur, Sundernagar, and Chennai are 45, 77 and 79 respectively. The critical points $Rn_1$ for rainfall of all four stations are almost same, $Rn_1 \approx 0.1$.

The results that we obtained are described as follows:
Temperature: The variation in conditional entropy of temperature for J, S and C with respect to D, for relative shift $n_0$ is shown in Fig. 3.10. We observe that $E(J/D)$ and $E(S/D)$ show almost similar trends with minimum at shift $n_0 = 0$. This implies that the temperature of Jaipur and Sundernagar are affected by a change in temperature of Delhi. We also calculated $E(D/J)$ and $E(D/S)$ and the results are similar, i.e. the temperature in Delhi is correlated with the temperatures of Jaipur and Sundernagar. Results for $E(S/J)$ and $E(J/S)$ also show minimum around $n_0 = 0$. Thus the temperature of three stations Delhi, Jaipur and Sundernagar are coupled and the relative dependence is seen to be mutually directional. In contrast, $E(C/D)$ (see Fig. 3.10) and $E(D/C)$ show no such variation: the temperatures of Delhi and Chennai do not seem to be dynamically coupled.

Relative humidity: Fig. 3.11 shows the result for $E(J/D)$, $E(S/D)$ and $E(C/D)$. A fall in $E(J/D)$ & $E(S/D)$ at $n_0 = 0$ and not so for $E(C/D)$ can be clearly seen. Similar results can be seen for $E(D/J)$, $E(D/S)$ and $E(D/C)$, though variation in $E(D/J)$ at $n_0 = 0$ is not so pronounced. $E(S/J)$ and $E(J/S)$ show weak minima around $n_0 = 0$. As with temper-
nature, we found the relative humidity of Delhi to be coupled with that of Jaipur and Sundernagar and vice versa. There appears to be no coupling in the relative humidity of Delhi and Chennai.

**Rainfall:** No significant variation in conditional entropy is observed. The rainfall at one of the stations does not appear to have significant effect on rainfall at other stations.

**Relative humidity & Rainfall:** We show in Fig. 3.12 the results for $E(J/D)$, $E(S/D)$ and $E(C/D)$, where, for example, in $E(J/D)$, the two time signals under consideration are relative humidity of Jaipur and the rainfall in Delhi. Similarly for others. While $E(J/D)$ and $E(S/D)$ show sharp dip at $n_0 = 0$, $E(C/D)$ goes almost like flat line. It shows that the rainfall in Delhi effectively increases the relative humidity in Jaipur and Sundernagar. While $E(D/S)$ shows similar result, the minimum of $E(D/J)$ appears at $n_0 = 4$ as shown in Fig. 3.13, meaning thus, the rainfall in Jaipur leads to an increase in relative humidity in Delhi, almost 2 days later. $E(S/J)$ and $E(J/S)$ also shows minimum at $n_0 = 0$, but they are not so prominent. For any station, $E(Rh/Rn)$ shows a very sharp minimum peak at null shift. If we now flip the variables, i.e., in $E(J/D)$, the variables of interest are rainfall in Jaipur and relative humidity of Delhi, the results show no remarkable variation in

![Figure 3.11: Same as in Fig. 3.10, for relative humidity.](image-url)
Figure 3.12: As in Fig. 3.10, for the relative humidity and rainfall. The first argument in conditional entropy indicates the relative humidity and the second the rainfall of the respective stations.

conditional entropy with relative shift. The result can be expected since an increase in relative humidity does not result in rainfall most of the times at other stations or even at same station.

**Relative humidity & Temperature:** The relative humidity and temperature at any location are clearly coupled, but no significant variation in conditional entropy is observed for inter-station analysis. The results are almost same when the two variables are interchanged, i.e., $E(T/Rh)$ and $E(Rh/T)$ show minimum around $n_0 = 0$ for all stations, but this is not so when the analyses is repeated for two different stations. The results show that change in relative humidity at a particular location will affect the temperature there and vice versa. Further, the change in one variable at a station does not seem to affect significantly the other variable at any other location.

**Temperature and Rainfall:** $E(T/Rn)$ shows minimum at zero relative shift for Delhi and at $n_0 = -3$ (see Fig. 3.14(a)) for Sundernagar and not so for Jaipur (possibly due to scanty rainfall there and relatively high temperature). No such variation is observed for inter-station analysis. $E(Rn/T)$ shows a minimum at shift $n_0 = 3$ for Sundernagar (Fig. 3.14(b)); no such variation is observed for Delhi and Jaipur as well for inter-station analy-
Figure 3.13: As in Fig. 3.12, for $E(D/J)$ vs $n_0$.

sis. The minimum in $E(T/Rn)$ and not so in $E(Rn/T)$ for Delhi shows that temperature is substantially affected by rainfall there, but the change in temperature does not imply most of the time that there would be a rainfall. The shifted minimum ($n_0 = -3$) in $E(T/Rn)$ for Sundernagar shows that temperature change may result in rainfall almost two days later. The effect is reciprocative; $E(Rn/T)$ shows minimum at $n_0 = 3$. Sundernagar being a hilly region where the temperature is normally low, the rain may be expected if there is a temperature change, our result shows that rainfall may occur with a delay. We emphasize that since the minimum does not seem to be very robust, the result is an indicative of such a delay and other factors like local topography may also be crucial in such phenomena.
Figure 3.14: Conditional entropy (a) $E(T/Rn)$ and (b) $E(Rn/T)$ for Sundernagar. $T$ and $Rn$ stands for temperature and rainfall.

3.4 Discussion

We have shown here the utility of the information-theoretic method in detecting dynamical coupling between different climatic variables in real data. The coarse-graining method appears to be a robust procedure for determining whether different noisy time signals, taken from experimental observations, possibly arise from the same underlying dynamics. The analysis is based on obtaining a symbolic representation of the dynamics via a coarse-graining procedure, and subsequently computing the conditional probability. The method is relatively stable to errors in the dynamics or errors in the coding.

We show that an underlying dynamics governs the climate of a location; we observed varying degree of coupling strength between climate variables of Delhi. The local climates of Delhi, Jaipur and Sundernagar appear to be coupled and can perhaps be arising from the same underlying dynamics. The conditional entropy shows the conditional dependence of one variable on the other: $E(Z/X)$ and $E(X/Z)$ may not show the same variation with relative shift.

Conventional correlation methods are often inadequate when analyzing the dynamical coupling of complex data such as climate. This is of significance since conventional correlation methods such as the Pearson correlation
coefficient or cross-correlation function are not always reliable in finding dynamical coupling between aperiodic or chaotic variables.

It has been well substantiated by our analysis that correlation function is of little help in analyzing the dynamical coupling of chaotic time signals of numerical models as well as real experimental data. While the X & Z signals of the Rössler attractor can be seen to be coupled from correlation function analysis, those for Lorenz attractor show no such result.

Although one may anticipate variables like rainfall and relative humidity at a given location will originate from the same dynamics, it is often not clear which variables are dynamically coupled and further, which variables are strongly/weakly coupled to each other. The differential relative dependence of two variables on each other is another significant issue that has been addressed comprehensively through our study. In modeling complex systems, in particular in climatology, such information is of help, and we expect that the analysis presented here will find application in such a context.

The specific results pertaining to the climate data from the four cities studied here are as follows. Locations within \(\approx 400 \text{ km}\) of each other, \(D, J& S\), have a coupled temperature system and similarly for relative humidity. The rainfall at a location, seems not to depend on rainfall at other locations.

The relative humidity at one station is affected by the rainfall at the same location or nearby stations; the rainfall is not seen to be remarkably affected by change in relative humidity. Further, the relative humidity and temperature of a station shows coupling which is not conspicuous for inter-station analysis. In case of temperature and rainfall, \(E(T/Rn)\) shows minimum at \(n_0 = 0\) for Delhi, but no such variation for \(E(Rn/T)\). This indicates that the temperature of Delhi is significantly affected by rainfall there, which is not so the other way round. There is no notable variation as such for Jaipur and also for two different stations. Results for Sundernagar show that a change in temperature there may result in rainfall two days later. This result is important in the context of climate, since one may anticipate delayed occurrence of a phenomenon in response to a cause producing it. We also discovered such lagged coupling for relative humidity of Delhi and rainfall in Jaipur. The climate of Chennai, which is separated \(\approx 2000 \text{ km}\) from rest three stations appears to be uncoupled to the local climates there.

Time delay which is central theme of chapter 2 appears to be an important factor in regulating the dynamical coupling of climate variables. We observe
from the results above that the coarse-graining analysis present a novel tool in extracting delay information from climate data.

This study gathers information from climate time series data, which may be utilized in developing climate models. We have seen that time signals coming from same model may show differential dependence among themselves, depending on the coupling parameters, this fact has been amply deciphered using coarse-graining analysis. The present technique is likely to be of help in modeling complex systems, climate being a potential one. Further, in climatology, delay is an important factor in determining the dynamical coupling of different variables. The knowledge of time-delays implicitly present in such data would further help in this regard.