Information Theoretic Approaches to Rapid Discovery of Relationships in Large Climate Data Sets

Kevin H. Knuth¹, William B. Rossow²

1. NASA Ames Research Center, Moffett Field CA, 94035-1000
2. NASA Goddard Institute for Space Studies, New York NY, 10025

Mutual information as the asymptotic Bayesian measure of independence is an excellent starting point for investigating the existence of possible relationships among climate-relevant variables in large data sets. As mutual information is a nonlinear function of its arguments, it is not beholden to the assumption of a linear relationship between the variables in question and can reveal features missed in linear correlation analyses. However, as mutual information is symmetric in its arguments, it only has the ability to reveal the probability that two variables are related. It provides no information as to how they are related; specifically, causal interactions or a relation based on a common cause cannot be detected. For this reason we also investigate the utility of a related quantity called the transfer entropy. The transfer entropy can be written as a difference between mutual informations and has the capability to reveal whether and how the variables are causally related. The application of these information theoretic measures is tested on some familiar examples using data from the International Satellite Cloud Climatology Project (ISCCP) to identify relations between global cloud cover and other variables, including equatorial Pacific sea surface temperature (SST), over seasonal and El Nino Southern Oscillation (ENSO) cycles.
Mutual information can be seen as the asymptotic Bayesian measure of independence. It is an excellent starting point for investigating the existence of linear or non-linear relationships between climate-related variables in large data sets. As mutual information is defined as the amount of information that can be gained about one variable by observing another variable, it is not biased towards the assumption of a linear relationship between the variables in question and can reveal features missed in linear correlation analyses. However, as mutual information is symmetric in its arguments, it only has the ability to reveal the probability that two variables are related. It provides no information about how they are related, specifically, causal interactions or a relation based on a common cause cannot be detected. For these reasons, we also investigate the utility of a related quantity called the transfer entropy. The transfer entropy can be written as a difference between mutual informations and has the capability to reveal causalities in data for which the variables are usually redundant. The application of these information-theoretic measures is tested on some familiar examples using data from the International Satellite Cloud Climatology Project (ISCCP) to identify relationships between global cloud cover and other variables, including equatorial Pacific sea surface temperature (SST) over seasonal and El Nino Southern Oscillation (ENSO) cycles.

In studying a complex dynamical system, one of the first approaches is to break it down into a set of dynamically coupled subsystems. In recent years, this has been possible, and in fact, it has been possible to define a set of strongly coupled subsystems or Markovian systems. The Markovian systems are important because they can be used to describe the behavior of the system at a time series. This is done by considering the transition matrix and the transition probabilities.

At this stage it is important to step back from the problem and consider the assumptions. Rather than focusing on driving systems and parameters, we can examine how our understanding of the state of the system at a future time is affected by our knowledge about the present or past states of the system. We consider the subsystems to be Markovian systems. This will provide much information regarding the interactions among the subsystems as well as giving us an idea as to the predictability of the various parts of the system.

We can characterize the behavior of a system X by looking at the probability distribution over the set of states of the system at a time t. We can also express the system in terms of an unknown stochastic process using a Markov process. We can then apply the Markov process over a series of states x(t) to find the transition matrix as

\[ P(x(t+1)|x(t)) = \log \left( \frac{P(x(t+1))}{P(x(t))} \right) \]

If we average this over all states, this gives us a measure of our uncertainty. This quantity is called the Shannon entropy [4].

\[ H(X) = \sum_{x} p(x) \log \frac{1}{p(x)} \]

If the system states are described with multiple parameters, the entropy can still be computed by averaging over all possible states (here we have assumed for states described by X and Y)

\[ H(X,Y) = \sum_{x,y} p(x,y) \log \frac{1}{p(x,y)} \]

Now, if we consider two subsystems X and Y, which together make up a larger system, we can compute what is called the Mutual Information (MI) by

\[ MI(X,Y) = H(X,Y) - H(X) - H(Y) \]

Notice that this describes the difference between the uncertainty when we are treated separately and when they are treated jointly. If these two subsystems are independent of one another, the MI will be zero. However, if there is any interaction between these subsystems, the MI will be positive. This can perhaps be estimated by writing it as

\[ MI(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \]

If X and Y are independent then the probability is the same factor and the MI is zero.

However, as the MI is symmetric with respect to interchange of X and Y it cannot distinguish the direction of any causal influence as well as the effect of any common influence on both subsystems. A third system called the Transfer Entropy can be written as a function of MI

\[ T(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x|y)}{p(x)} \]

Where X and Y are the systems defined by the set of states from, respectively. This is represented by the previous states from X. This is a measure of the knowledge obtained from the current state of Y. For a 1-D this can clearly be written in terms of Shannon entropy (although not as clearly interpretive) as

\[ T(X,Y) = \sum_{x} p(x) \log \frac{p(x|y)}{p(x)} \]

In contrast the cloud cover over Paris France has a low MI with respect to the cloud cover in New York City. This is reflected by the probability density of the joint probability density in the right. Note that there is little dependence of cloud cover on sea surface temperature.

The transfer entropies were found to be difficult to estimate with precision because they expected to improve in the future. Using a Gaussian kernel density estimator, we found that the transfer entropy above the cloud cover was large, suggesting we know that the seasonality is an autonomous variable.