Entropy train diagrams for information-based measures of statistical association

Jobst Heitzig and Jakob Runge

Potsdam Institute for Climate Impact Research, Transdisciplinary Concepts and Methods, Potsdam, Germany
(heitzig@pik-potsdam.de, +49 331 288-2640)

Nonlinear statistical associations between geophysical processes are increasingly often analyzed with concepts from information theory, i.e., by measuring certain amounts of *entropy,* and often taking into account possible time-lags. We offer an intuitive visual tool that allows for easy understanding and comparison of many information-theoretic statistics, and use it to discuss existing measures and to introduce two new measures of sensitivity and coupling strength.

Given two (possibly multivariate) processes X, Y, the most prominent information-theoretic association measures are the symmetric measure of *mutual information* $I(X; Y)$ and the directed measure of *transfer entropy* $T(X \to Y)$. Mutual information can be interpreted as that "part" of the entropy $H(X_t, Y_t)$ of the present joint state (X_t, Y_t) that "is part of" both the entropy of X_t and of that of Y_t. Transfer entropy can be interpreted as that amount of the entropy of Y_t that "first visited" the joint system (X, Y) at subsystem X at some past time $t - k$ and did not "visit" subsystem Y before the present time t (but might have visited X again between $t - k$ and t). In the special case of discrete-time Gaussian processes, such measures can typically be expressed in terms of certain (co-)variances between unlagged or lagged variables or residuals. Transfer entropy is then proportional to the linear measure of *Granger "causality"* and is thus influenced by possible auto-correlations in X but not by those in Y.

We introduce *entropy train diagrams* as an intuitive way to illustrate the above kind of entropy flow interpretation. Such a diagram depicts a possible path that a part of the entropy (or a "signal") can take through the joint system (X, Y) and its subsystems X and Y over time. It combines a time axis with a Venn diagram that corresponds to the decomposition of the entropy $H(X_t, Y_t)$ into the sum of mutual information and two conditional entropies, $H(X_t, Y_t) = H(X_t|Y_t) + I(X_t; Y_t) + H(Y_t|X_t)$. A simple colour-coding is used to mark mandatory, optional, and excluded places the signal "visits" at different time-points. It allows for an easy understanding and comparison of the often lengthy and formalistic definitions of information-based association measures.

We first use entropy train diagrams to illustrate different decompositions of transfer entropy of discrete-time processes into lag-specific parts, which can be used to determine the probable time lags at which two systems might be coupled, or to filter for relevant lags that are predicted by some theory. Finally, we use entropy train diagrams to motivate a modification of transfer entropy in two opposing directions, resulting in new aggregate measures of non-linear association and their symmetrized and linear counterparts: *Backdoor entropy* reflects the aggregate effect on the present value of Y of all signals which visited X before ever visiting Y. It is possibly mediated by auto-correlation in X or Y and can thus be interpreted as a measure of "sensitivity". *Leap entropy,* in contrast, reflects that part of the aggregate effect of these signals which was not mediated by auto-correlation in X or Y, and is thus a measure of "coupling strength".