Transfer Entropy

- References:
  - Schreiber, T., 2000: Measuring information transfer, *Phys. Rev. Lett.*, **85**, 461–464, doi: 10.1103/PhysRevLett.85.461.
  - Ruddell, B. L., and P. Kumar, 2009a: Ecohydrologic process networks: 1. Identification. *Water Resour. Res.*, **45**, W03419, doi: 10.1029/2008WR007279.
- Principle:
  - Transfer entropy (TE) is closely related to Mutual Information; it is the reduction in the uncertainty of the *current state* of variable (e.g. time series) y gained from information in variable (e.g., time series) x that is not already present in y. Thus it can identify coupling between variables.
  - TE is defined by considering blocks or spans of information in x and y; the spans need not be the same. Generally:

$$T(X_t \rightarrow Y_t) = \sum_{y_t, y_t^{[k]}, x_t^{[l]}} p\left(y_t, y_t^{[k]}, x_t^{[l]}\right) \ln \frac{p\left(y_t \middle| \left(y_t^{[k]}, x_t^{[l]}\right)\right)}{p\left(y_t \middle| y_t^{[k]}\right)}$$

    where p denotes a probability distribution (marginal or joint as the case may be), $[k]$ and $[l]$ denote different blocks of data of lengths k and l – histories preceding the current time for time series, bars denote conditional probabilities.
  - Lags may be introduced so the blocks $[k]$ and $[l]$ do not immediately precede the time t. For simplicity, when lags are introduced blocks may be reduced to size 1, lest the number of possible combinations grow exponentially.
  - TE can be expressed in terms of the Shannon entropy $H(x) = -\sum p(x) \ln p(x)$, as:

$$T(X_t \rightarrow Y_t, \tau) = H(x_{t-\tau}, y_{t-1}) + H(y_t, y_{t-1}) - H(y_{t-1}) - H(x_{t-\tau}, y_t, y_{t-1})$$

    for x as a factor at lag τ, assuming the strongest determinant of y within its own history is its value at the previous time step.
- Data needs:
  - Typically applied to time series locally, although X could be a remote forcing.
  - Ruddell and Kumar (2009a) find 10-20 bins to determine p, and 500-1000 data points are adequate for robust estimation of TE.
  - Lags may be viewed as a measure of time scale, especially when applied across a range of τ, behaviors become apparent. Note that T is not symmetric, so

$$T(X_t \rightarrow Y_t, \tau) \neq T(Y_t \rightarrow X_t, \tau)$$

    even though there may be transfers in both directions at different lags.
- Observational data sources:
  - Because of the considerations listed above, this approach is well suited to the vagaries of observational data
- Caveats:
  - As a nonparametric statistic, TE significance can generally be established only by Monte Carlo methods, such as shuffling (shifting) X versus Y randomly in time, or bootstrapping (sampling with replacement), depending on the nature of the time series.
  - Periodic "noise" (diurnal cycle, seasonal cycle) need not be removed in advance as the lag process can pull out any signals buried among known periodicities, as long as the signal to noise ratio for the coupled signal >1.

Supported by NASA/MAP Grant: NNX13AQ21G