Coupling metrics to diagnose land-atmosphere interactions

Mutual Information

- References:
  - Cover, T. M., and J. A. Thomas, 1991: *Elements of Information Theory*. Wiley, 542 pp.
  - DelSole, T., and M. K. Tippett, 2007: Predictability: Recent insights from information theory. *Rev. Geophys.*, **45**, RG4002, doi: 10.1029/2006RG000202.
- Principle:
  - Mutual information builds on concepts of information theory such as entropy, a measure of dispersion in a distribution. For a continuous probability distribution $p(x)$ for which $\int p(x)dx = 1$, the entropy across x is: $-\int p(x) \log p(x)\, dx$.
  - Relative entropy, or Kullback-Leibler divergence, quantifies the difference between two probability distributions in x, $p(x)$ and $q(x)$ as: $\int p(x) \log[p(x)/q(x)]\, dx$, which is 0 if the two distributions are identical.
  - Mutual information is a nonparametric measure of the dependence between two variables, similar to a correlation but without the assumption of linearity of the relationship. Stated another way, $I(x,y)$ is a measure of the reduction in the uncertainty of y due to knowledge of x.
  - Consider x as values of a variable in a probability distribution – the distribution could be across space or time or both. Consider another variable y with its own distribution. Plotting y versus x yields a scatter diagram, and the Pearson's correlation coefficient is a measure of the fit of that scatter to a straight line. Mutual information is a goodness-of-fit independent of the assumption of a specific functional relationship, quantified as:

    $$I(x,y) = \iint p(x,y) \log\left[\frac{p(x,y)}{p(x)p(y)}\right] dxdy$$

    If x and y are independent, the value is 0.
- Data needs:
  - Can be applied in the same situations as correlation – to relate two time series, two spatial distributions or space-time distributions.
  - Suitable for non-continuous or incomplete data as well as complete data sets. However, the joint probability distribution $p(x,y)$ must have exactly the same data as the marginal probability distributions $p(x)$ and $p(y)$, i.e., no cells or times when one has missing data and the other does not.
- Observational data sources:
  - Well suited to data where multiple components of the LoCo process chain are measured at the same location, such as flux tower data.
- Caveats:
  - For practical application to Earth system data, which is typically discrete in space (gridded or stations) and time (intervals or averages over intervals), integrals are replaced with summations across a finite number of bins; I is sensitive to bin choices.
  - As with correlations, causality between variables is not assured, but more subtle and complex relationships than possible with correlation can be revealed.
  - Rarely for a randomly distributed finite sample will $I(x,y) = 0$. Only when every bin in $p(x,y)$ contains exactly the same number of points does $I(x,y) = 0$
  - For any bins where the distribution $= 0$, the term inside the summation should be set to 0 (i.e., avoid dividing by 0).

Supported by NASA/MAP Grant: NNX13AQ21G