Connectivity Analysis

- References:
  - Western, A. W., G. Bloeschl, and R. B. Grayson, 2001: Toward capturing hydrologically significant connectivity in spatial patterns, *Water Resour. Res.*, **37,** 83–97, doi: 10.1029/2000WR900241.
  - Koch, J., A. Siemann, S. Stisen, and J. Sheffield, 2016: Spatial validation of large-scale land surface models against monthly land surface temperature patterns using innovative performance metrics. *J. Geophys. Res.*, **121,** 5430–5452, doi: 10.1002/2015JD024482.

- Principle:
  - Connectivity is a measure of heterogeneity or smoothness in a distribution, typically of a 2-D spatial field but it can be applied to any number of dimensions including time. It is often used in hydrogeology with regard to characterizing percolation, but the method is broadly applicable to characterize patterning, similar to a variogram but providing more information that just decorrelation scales.
  - A field $F(x)$, where *x* may be one or more dimensions, is decomposed into a binary distribution $B_\theta(x)$ by choosing a threshold θ setting all values of B_θ to 1 where $F > \theta$ and setting B_θ to 0 where $F \leq \theta$. A range of thresholds θ_i are chosen, ideally along equal percentiles of the distribution of *F* to produce an unbiased result. Each choice of θ_i produces in phase-space x a set of clusters C_i wherein $B_\theta = 1$. The probability of connection is:

$$\Gamma(\theta_i) = \frac{\Sigma_{C_i} n_{C_i}^2}{\left[\Sigma_x B_{\theta i}\right]^2},$$

    where the numerator is a sum of the square of the number of grid cells in each cluster over all the clusters; the denominator is the square of the total number of cells where $B_\theta = 1$.
  - The distributions of Γ for all i can be compared between two sources of data for the variable in field F; e.g., a model versus observations, or two models, or between two variables expected to have related clustering properties. RMSE of Γ is a measure of goodness of match.
  - This can also be applied to the other half of the distribution, clusters of $B_\theta = 0$, revealing potential asymmetries in the connectivity.
  - Clustering may also be defined in different ways, for instance on a rectilinear grid as only cells that touch along edges, or cells that also touch on vertices.

- Data needs and caveats:
  - Discontinuities in data will introduce artifacts in the calculation – complete fields are ideal.

Supported by NASA/MAP Grant: NNX13AQ21G