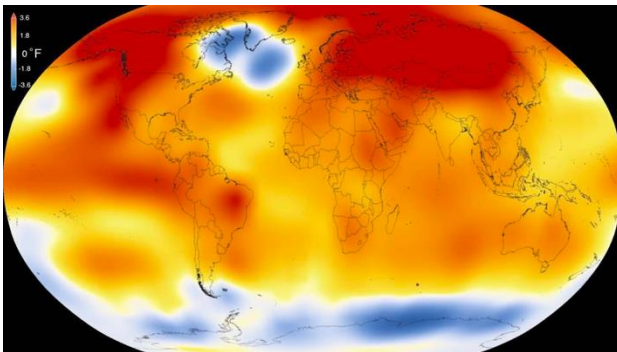# Variogram & Correlogram

- Both require data that have **stationarity:**
  - the pattern is roughly similar – has the same mean – throughout the dataset. In *non-stationary* data there is a systematic change in the local mean along the dataset. This can indicate an underlying gradient in the data.

Examples of non-stationarity in space and time

2015 temperature departure from normal

The "hockey stick" pattern in global mean temperature.
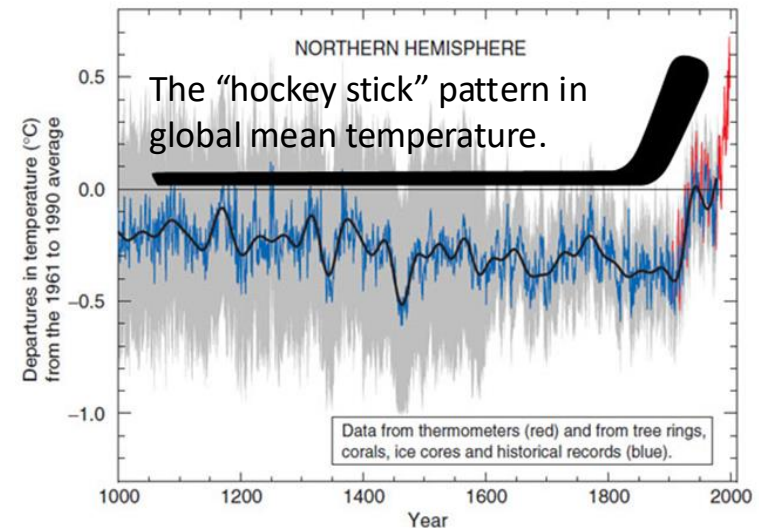
- Both assume **isotropy** *
  - Pattern being examined is about the same in all directions. *Anisotropy* means that the pattern varies according to the direction from which it is examined.*Anisotropy can be detected in a directional variogram. The 2015 map global temperature map also indicates some anisotropy in the southern ocean because the pattern differs horizontally vs. vertically.

Images: NASA, IPCC

Variogram vs. Correlogram

- Both X- axis = lag distance between points (pairwise distances are binned and are either represented as classes or actual values on the axis)

- The **null hypothesis** for semivariograms and correlograms is no pattern - i.e., a flat line.

- Y-axis:

  - **Variogram**: semivariance (<u>dissimilarity</u>): half the variance of the differences between all possible points spaced a constant distance apart.

  - **Correlogram**: correlation (or Moran's I); measure of <u>similarity</u>

Covariance, Correlation, Semivariance, Variance

- Relationships are close to inverse
- Variance here is mean variance across study area
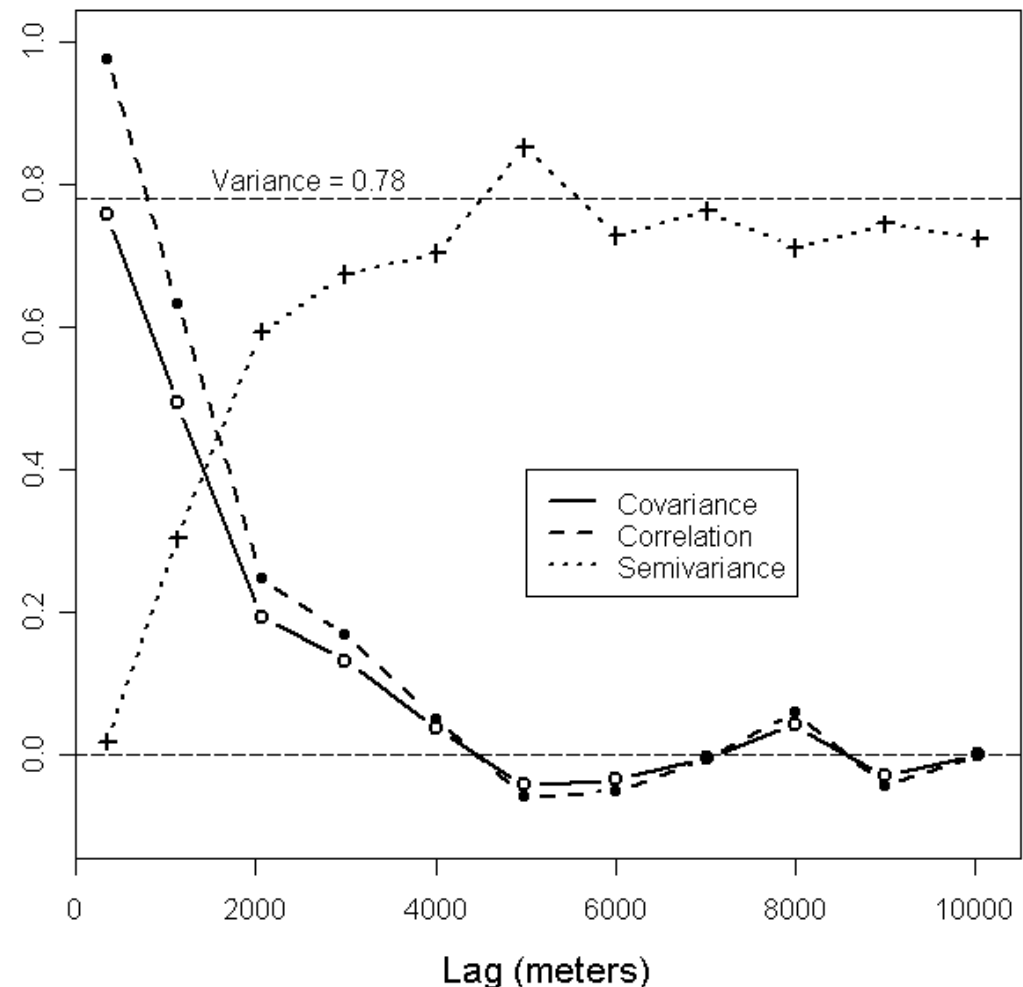
Variogram

- A measure of how much two samples taken from an area will vary depending on the distance between those samples.
- X-axis: distance between points (lag)
- Y-axis: Semivariance
- Dotted line= sill
- Y-intercept = nugget
- Asymptote (~ 100 km) = Range

Asymptote

Geographic variation along shoreline in Sea urchin gonad weight
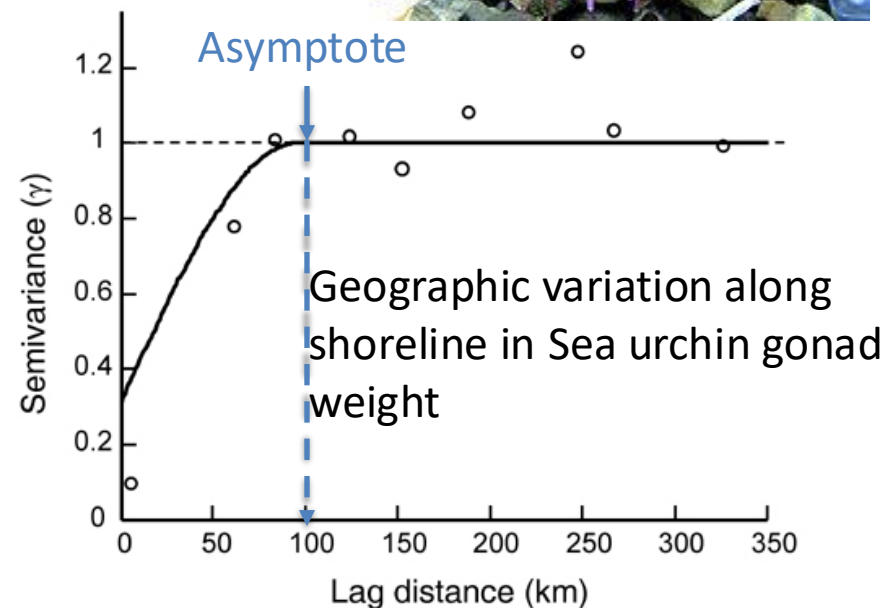
Sarah E. Lester, Steven D. Gaines, and Brian P. Kinlan. 2007. Reproduction on the edge: large-scale patterns of individual performance in a marine invertebrate. Ecology 88:2229–2239.

Variogram components

- **Sill**: where there is maximum semivariance in the data (max dissimilarity between points)
- **Nugget**: where there is minimum semivariance in the data (min. similarity between points; often variability below sampling scale)
- **Range**: lag distance at which the semivariance levels off and there is no longer correlation between points (max distance at which there is a relationship between points; beyond this range point pairs are dissimilar)
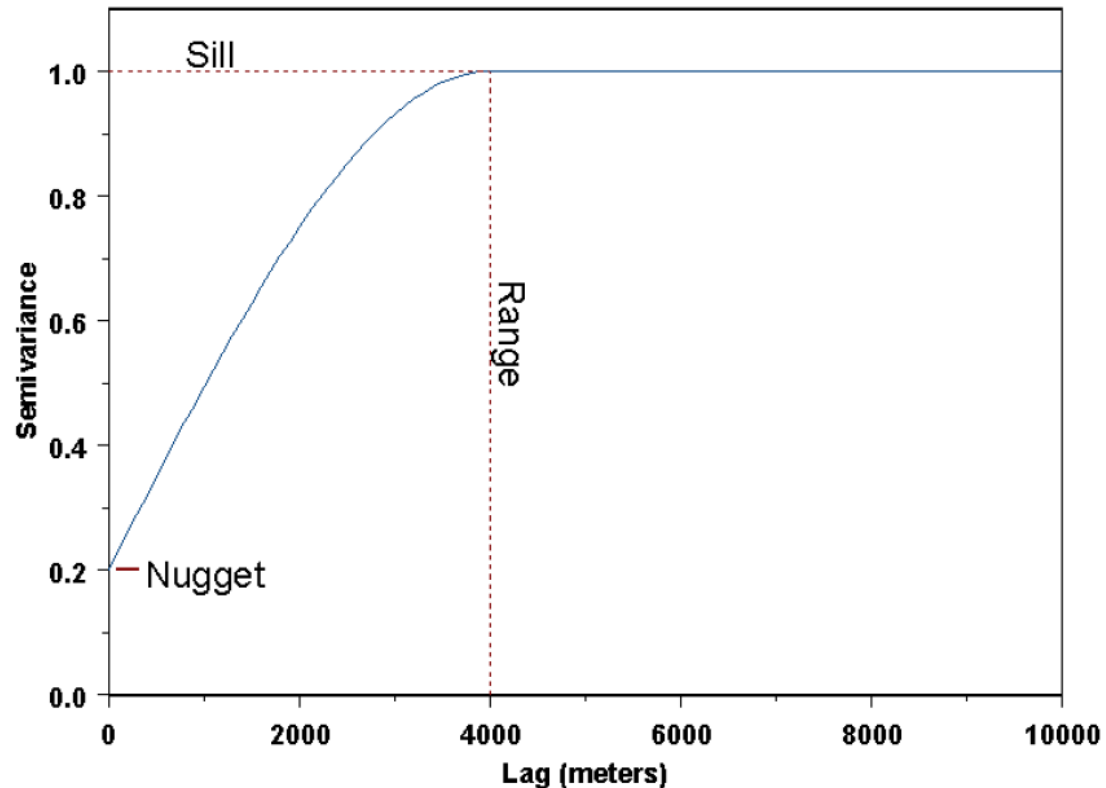
Creating a Variogram

1.

1. For all pairs of points, (*i,j*), obtain all pairwise distances and corresponding values *v* from the variable of interest at all positions just like correlogram

2. Bin pairwise distances (binning) to obtain the distance class, *h.*
   - group the pairs so that they have a common distance and direction.
   - In the landscape scene of 12 locations, you can see the pairing of all the locations with one location, the red point. Similar colors for the links between pairs indicate similar bin distances.
   - You do this for all pairs of points, such that all distances are accounted for (all blue lines to the right)
     - The numbers pointing to certain blue line distances reflect common distance-direction combinations that are used in the binning process (1&2 bin together; 3&4 bin together) See page 63 in the link on the next page for more details about this process.

2.

Legendre 1989, ArcGIS.com
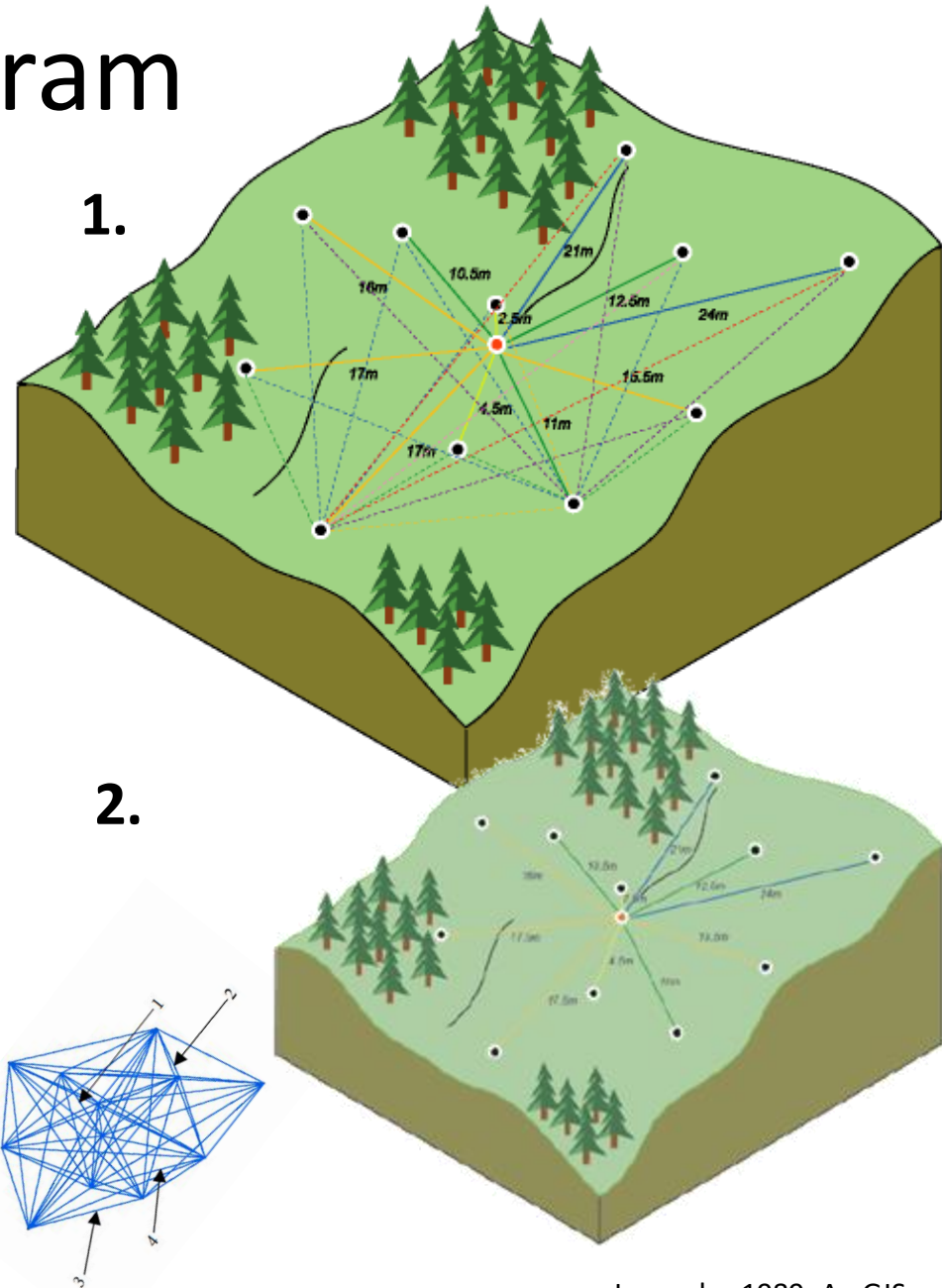
Creating a Variogram
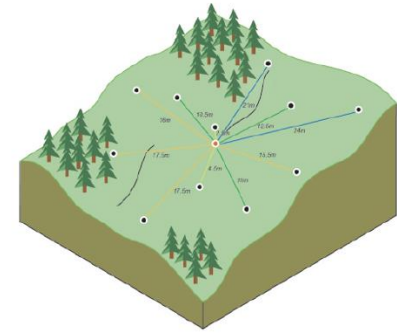
3. Compute Semivariance

- **Y axis on semivariogram (with binning) is computed for each distance bin *h*.**

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i,j=1}^{N(h)} (v_i - v_j)^2$$

γ(h) = Semivariance for the distance (or lag) *h (=bin)*

v_i, v_j = Values (of the variable of interest) at position *i* and *j*

N(h) = Number of pairs within distance *h*

For each bin, form the squared difference from the values for all pairs of locations that are linked, and these are then averaged and multiplied by 0.5 to give one empirical semivariogram value per bin. These are plotted on the next page to form the semivariogram plot.
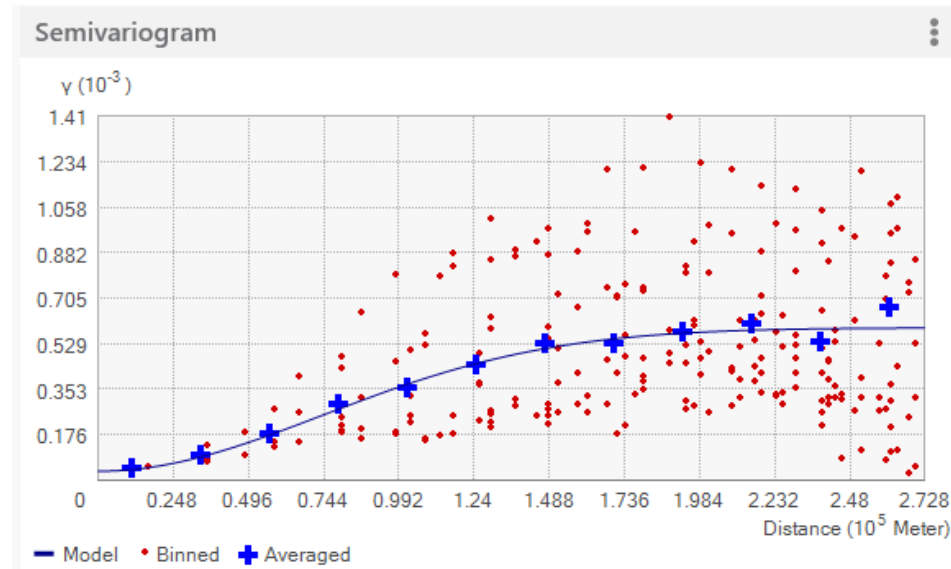
Creating a Variogram

4. Create the empirical variogram

- Plot the binned distances, *h* (x-axis) = distance from the center of the bin to the origin (center of semivariogram surface), and semivariances (y-axis)

- = red points at right

5. Plot the average semivariance per bin (blue crosses)

6. Fit a theoretical model to the empirical variogram (blue curve)

*note that the x-axis here is distance (m) instead of a distance class.

ArcGIS has a good resource that covers this process of creating variograms more completely. In addition to the series of webpages linked in Lab5 (https://space-lab-msu.github.io/MSUGradSpatialEcology/lab5_variograms.html) starting with "Empirical semivariogram and covariance functions", check out **Chapter 3 The principles of geostatistical analysis**, in their Geostatistical Analysis PDF, starting on page 49:
http://downloads2.esri.com/support/documentation/ao_/Using_ArcGIS_Geostatistical_Analyst.pdf ArcGIS.com

Variograms: Notes

- Really "Semi-variogram" but called "variogram" for short

  - As the lag distance approaches infinity, variogram functions converge to twice the variance. Therefore, dividing by 2 means that the **sill** approximates the variance.

- Unlike correlogram with Moran's I (univariate; or multivariate: Mantel test), cannot really perform statistical hypothesis test with variograms, but you can test the fit of the theoretical variogram model to the empirical variogram.

Legendre 1989

Variograms: Types

- 2 types of variograms:
  - Experimental
    - Created from data using the semivariance formula
  - Theoretical
    - Fitted to experimental variogram
    - Useful for kriging (an interpolation method)
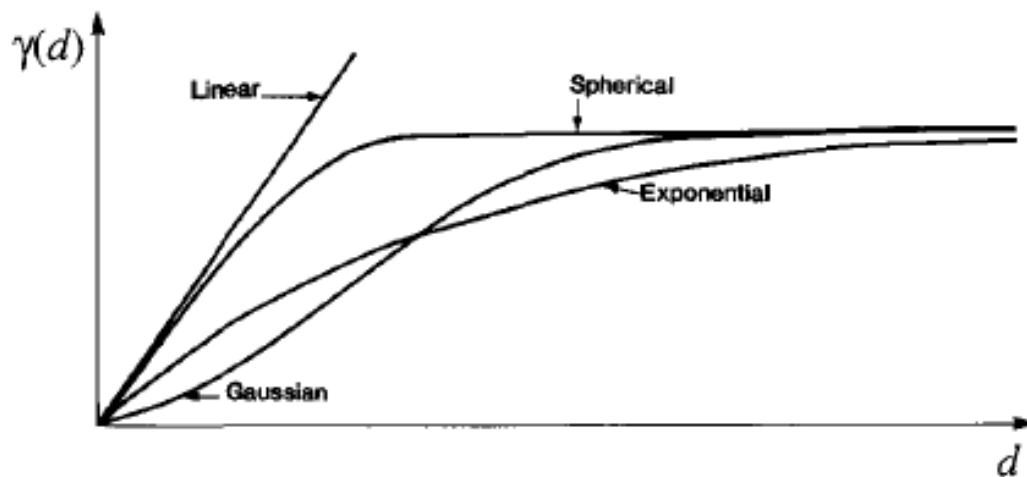
- Several types of **theoretical fits** are commonly used. Except for linear, they all describe spatial influence of the process of interest
  - Linear (linear spatial gradient; no sill)
  - Gaussian
  - Exponential (may not have a sill)
  - Spherical
  - Flat (no nugget; no spatial structure; pure nugget effect)
- **Isotropic?**
  - If variograms fit for all directions all show same shape (anisotropic if not)
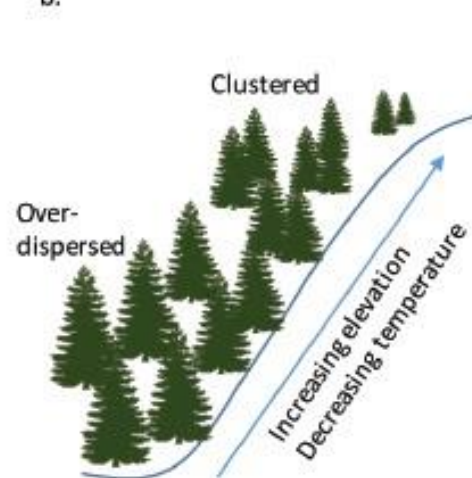

Legendre 1989

Variograms: Anisotropy

- **Isotropy**: variance depends only on the distance between two measured objects, not the direction from one to another

- **Anisotropy**: variance depends on the distance between two measured objects. This is directional dependency [spatially-dependent variation]; different properties in different directions and often treated as random error. It is possible to know the underlying reason:
  - e.g., wood is easier to split along grain, but not across grain
  - e.g., Elevation gradient: environmental conditions change with elevation (temperature, stress gradient, degree of competition), so spatial relationships between multiple variables across elevation are likely to be **anisotropic.** Trees along elevation: over-dispersed at lower elevations (bc of competition), clumped at higher elevations (ameliorate harsh conditions). (Buckley et al. 2016 *Chance*)
  - In variogram: https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/accounting-for-directional-influences.htm
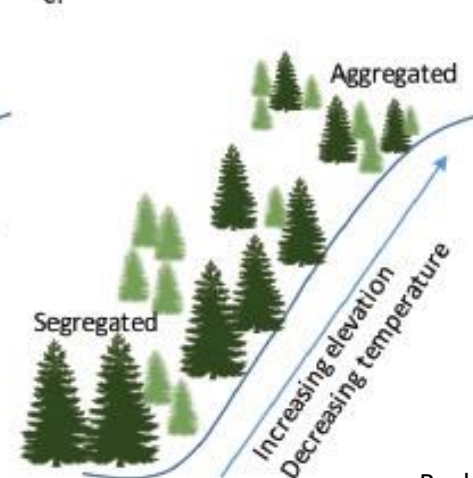

Buckley et al. 2016 Chance

Variograms: Anisotropy

- ## Sugar maple abundance

  – surveyed in same study locations as the hemlock data presented for correlograms

- ## Anisotropic

  – Anisotropic ratio=range in direction with longest range/range in direction with smallest range

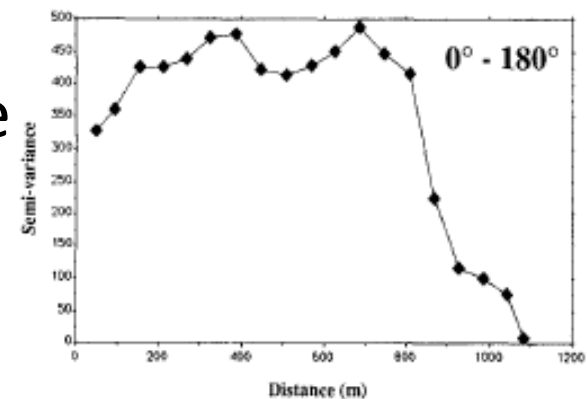   - ~1 is isotropic ; >1 is anisotropic
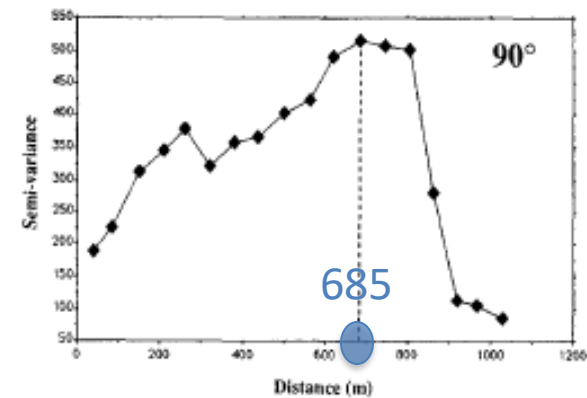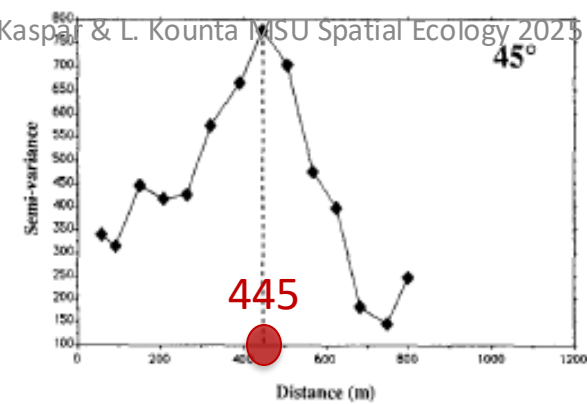
    - 685/445=1.5

Fig. 9. Three experimental variograms computed for the *Acer saccharum* data. See text. Abscissa: distance classes. Ordinate: values of the semi-variance function $\gamma(d)$. Dashed lines: ranges. Modified from Fortin (1985).