Lecture 12:

Beyond Fisher distributions

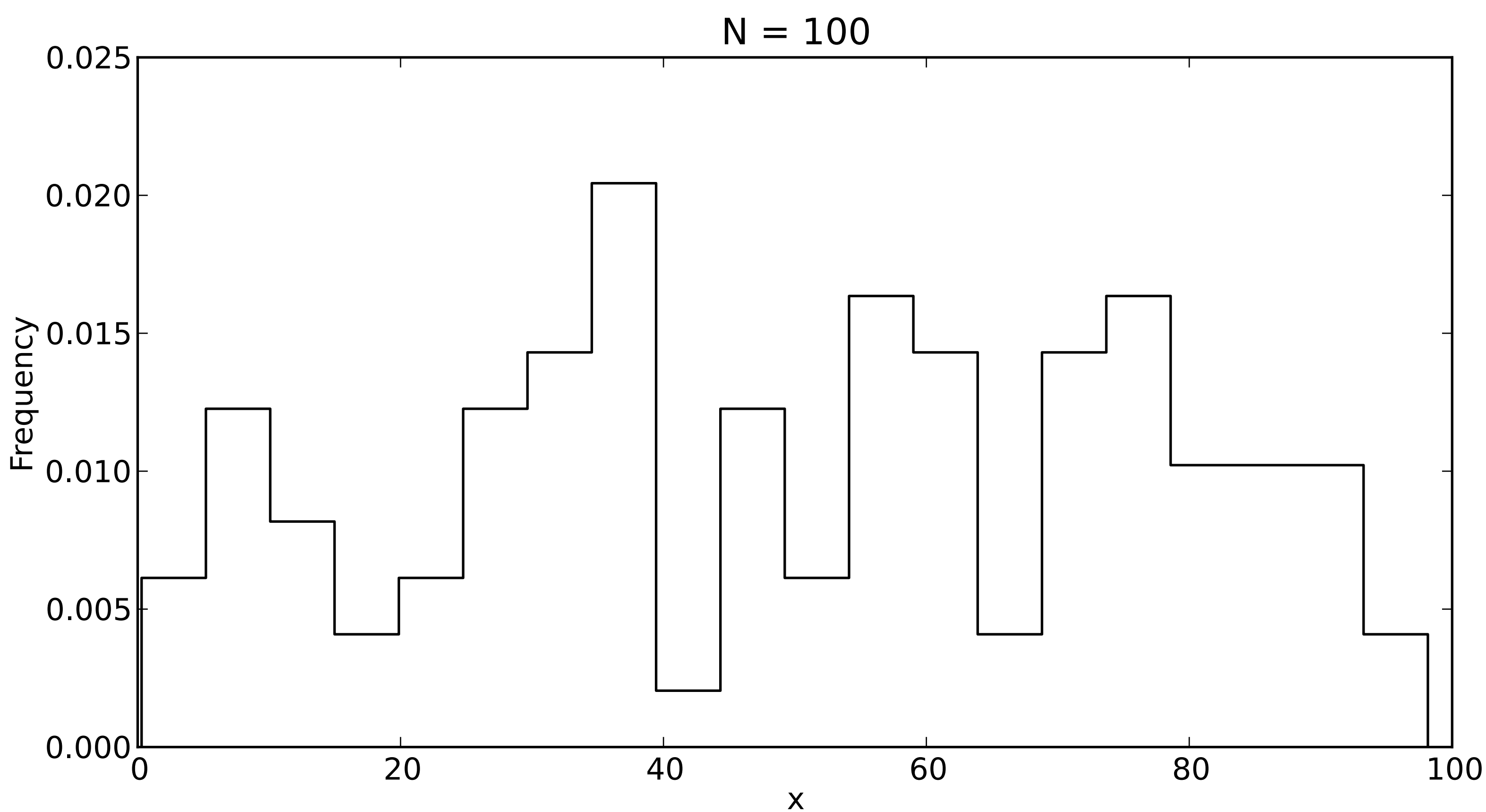
- How to tell if data are Fisher distributed
  - The magic of quantile-quantile plots
  - application to Fisher distributions
- Alternative statistical approaches for non-Fisherian data
- Applications to paleomagnetism
  - test for common direction
  - fold test

The magic of quantile-quantile plots
(Appendix B.1.5)

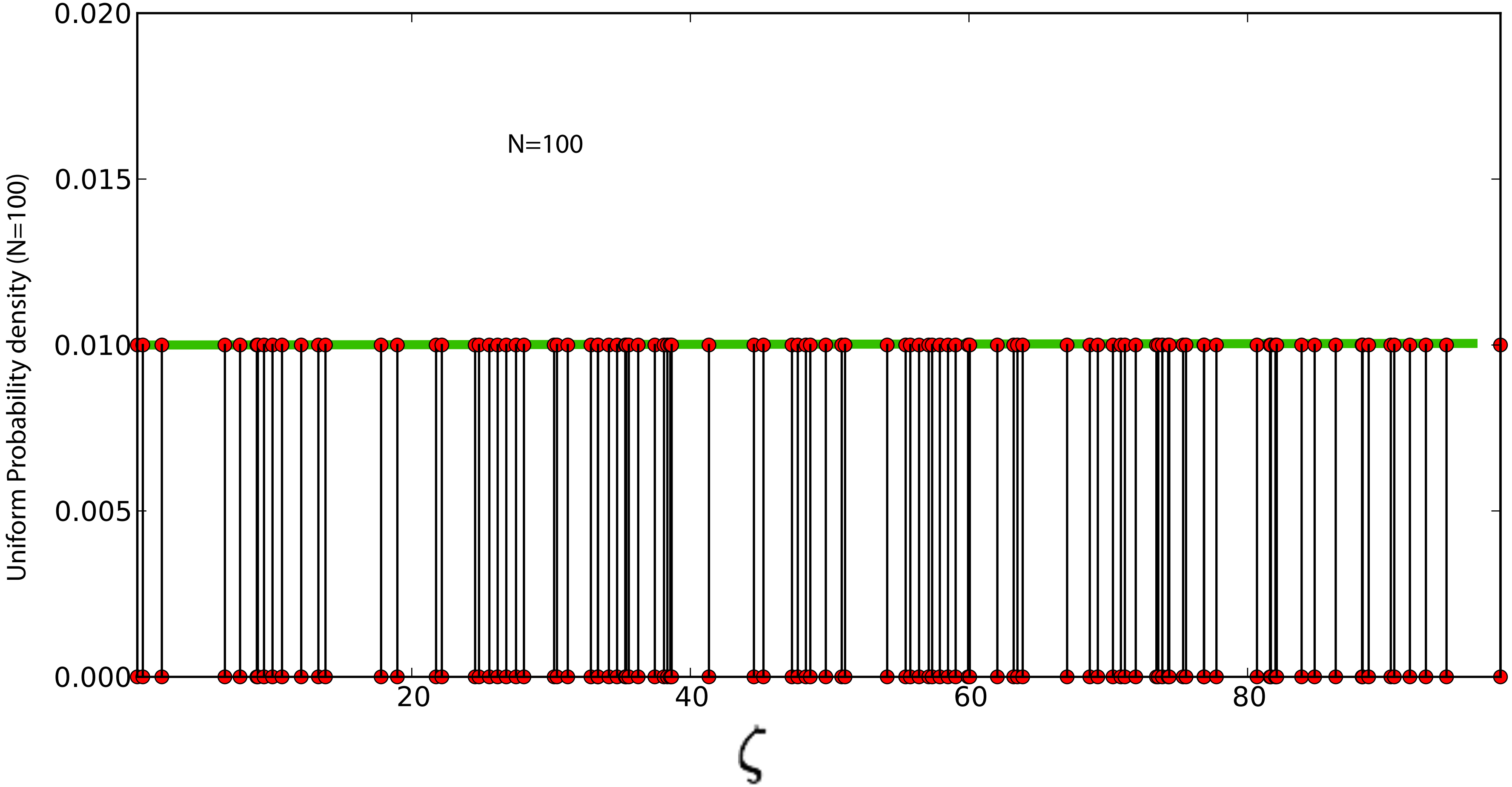
- In Q-Q plots, data are graphed against the value expected from a particular distribution
- If the data distribution is compatible with the chosen distribution, the data plot along a line
- Can quantify the degree of misfit and reject assumed distribution at 95% confidence

An example with uniform distribution

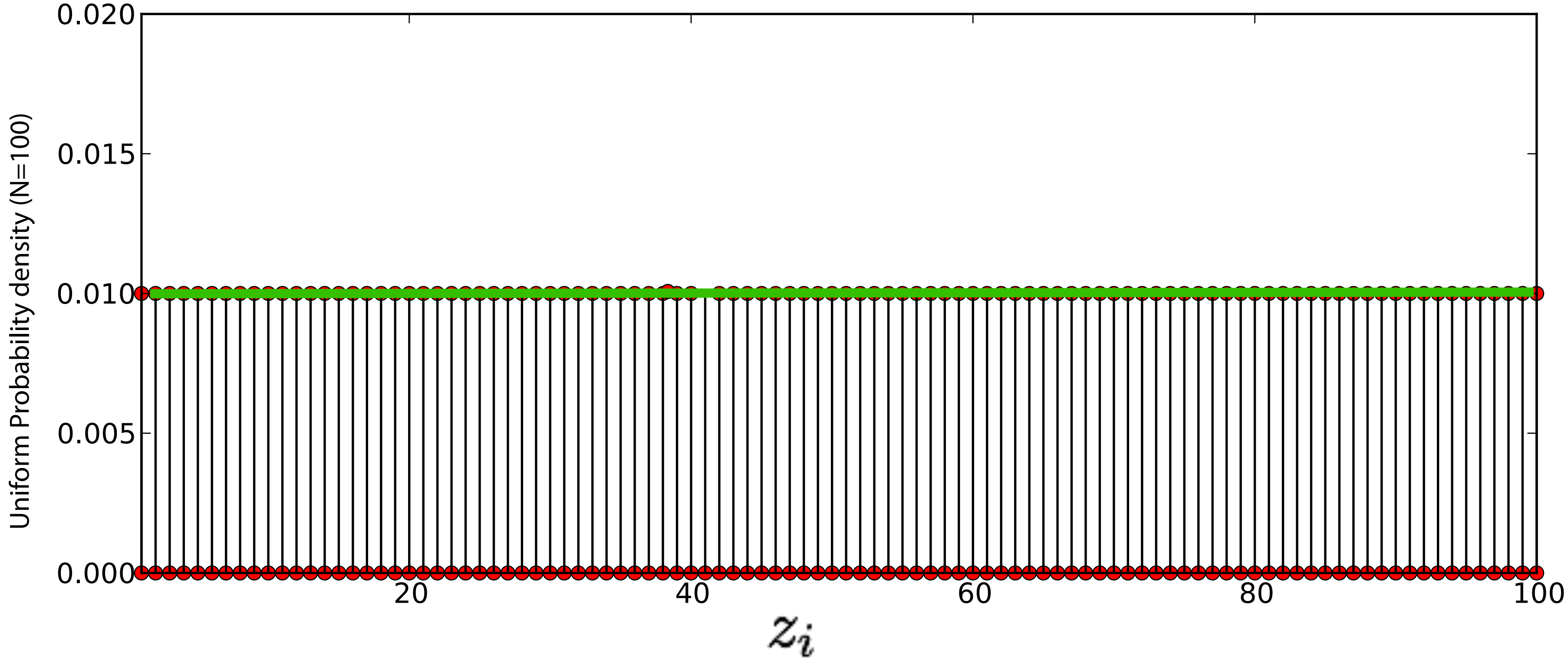
Generate a list of numbers from a uniform distribution (e.g., with command random.uniform(100))



Sort the data into an increasing list and locate each data point on the assumed distribution (in this case: uniform - green line)

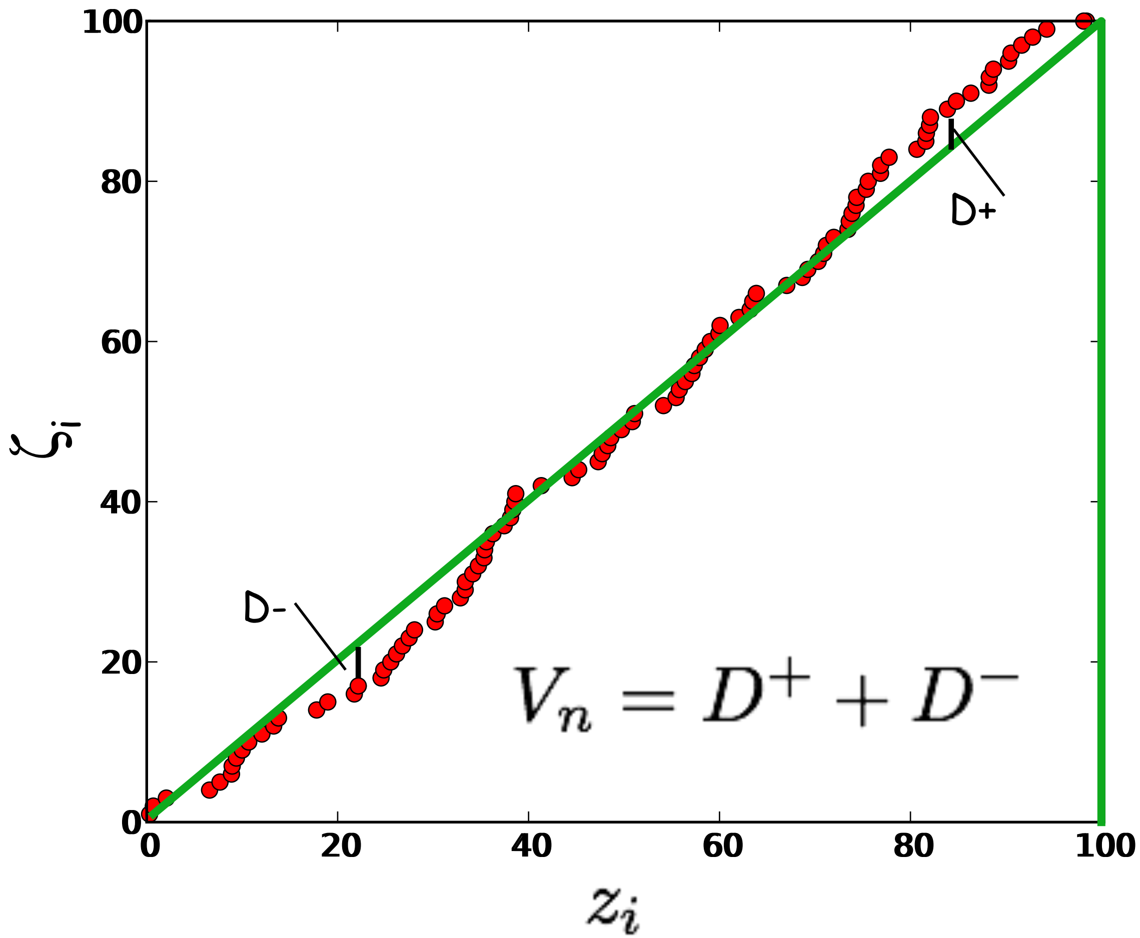


Break the expected distribution (green line) into N regions of equal area



This gives you a second list of z's - one for every original data point

Plot the each data point (ζ_i) against the corresponding values from the assumed distribution (z_i)



If distribution fits - plot will be linear

Calculate Vn

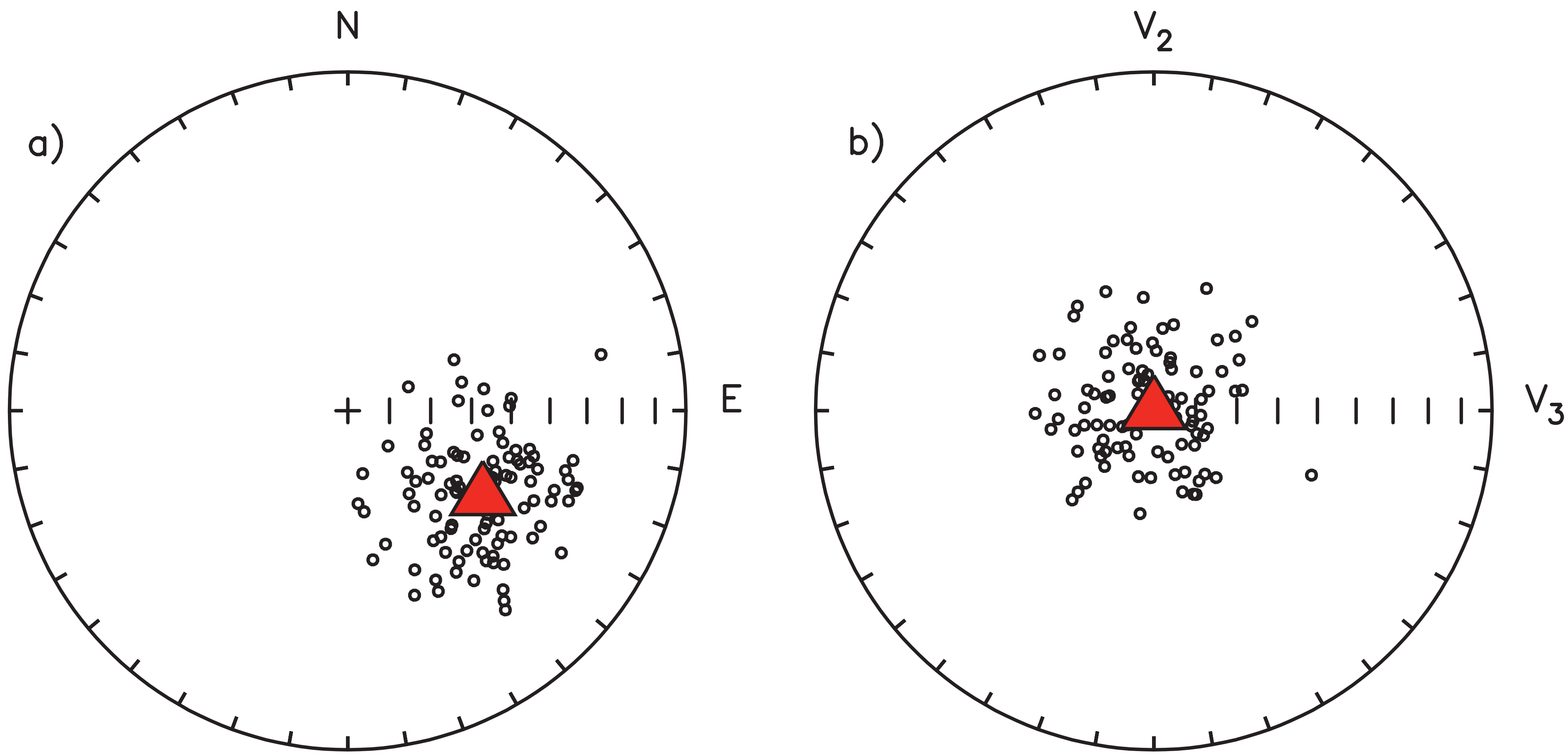
Compare Vn with maximum number allowed for 95% confidence:

$$M_u = V_n \left(\sqrt{N} - 0.567 + \frac{1.623}{\sqrt{N}} \right)$$

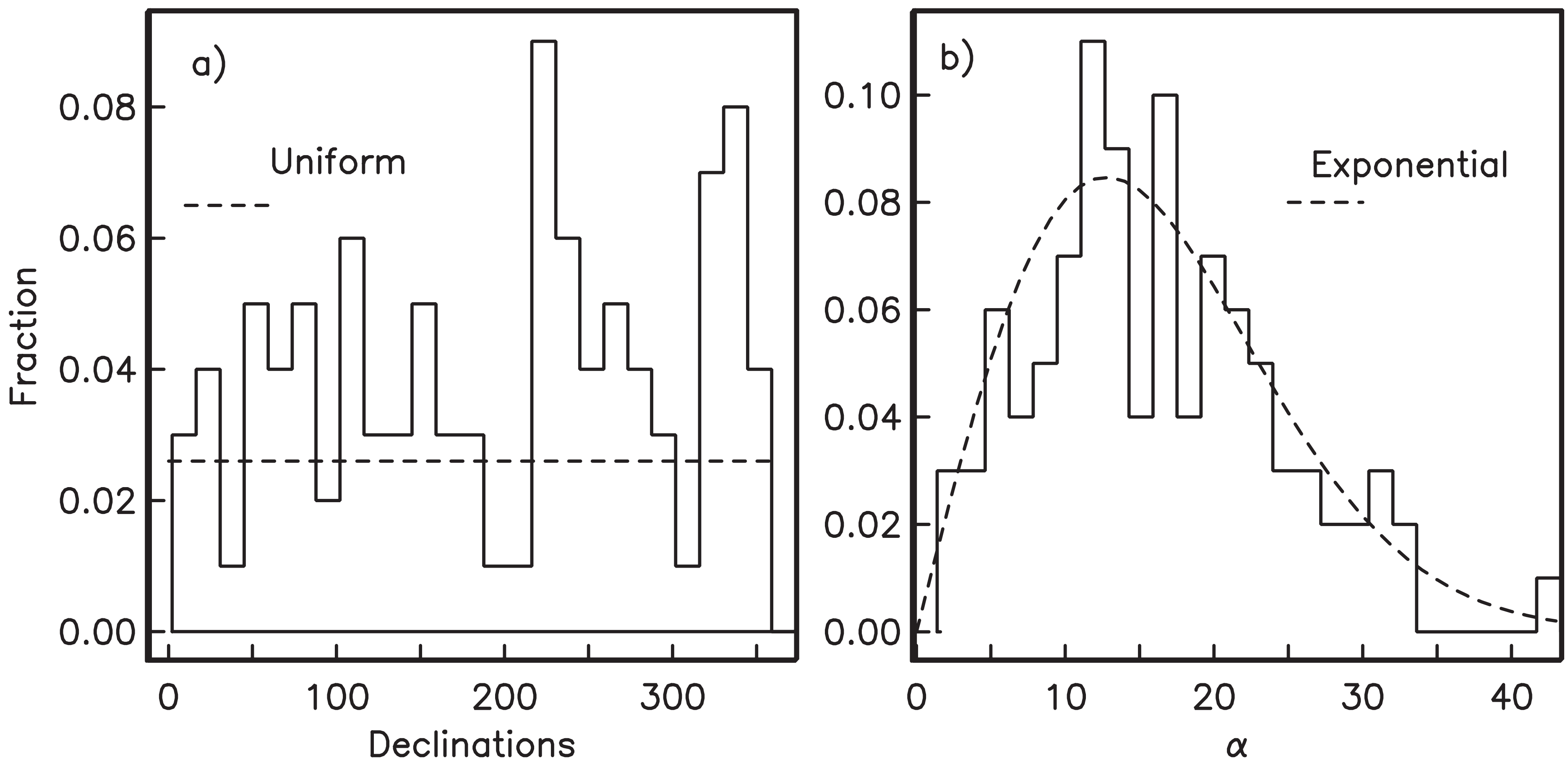
If $M_u > 1.207$, not uniform

Applied to paleomagnetic data

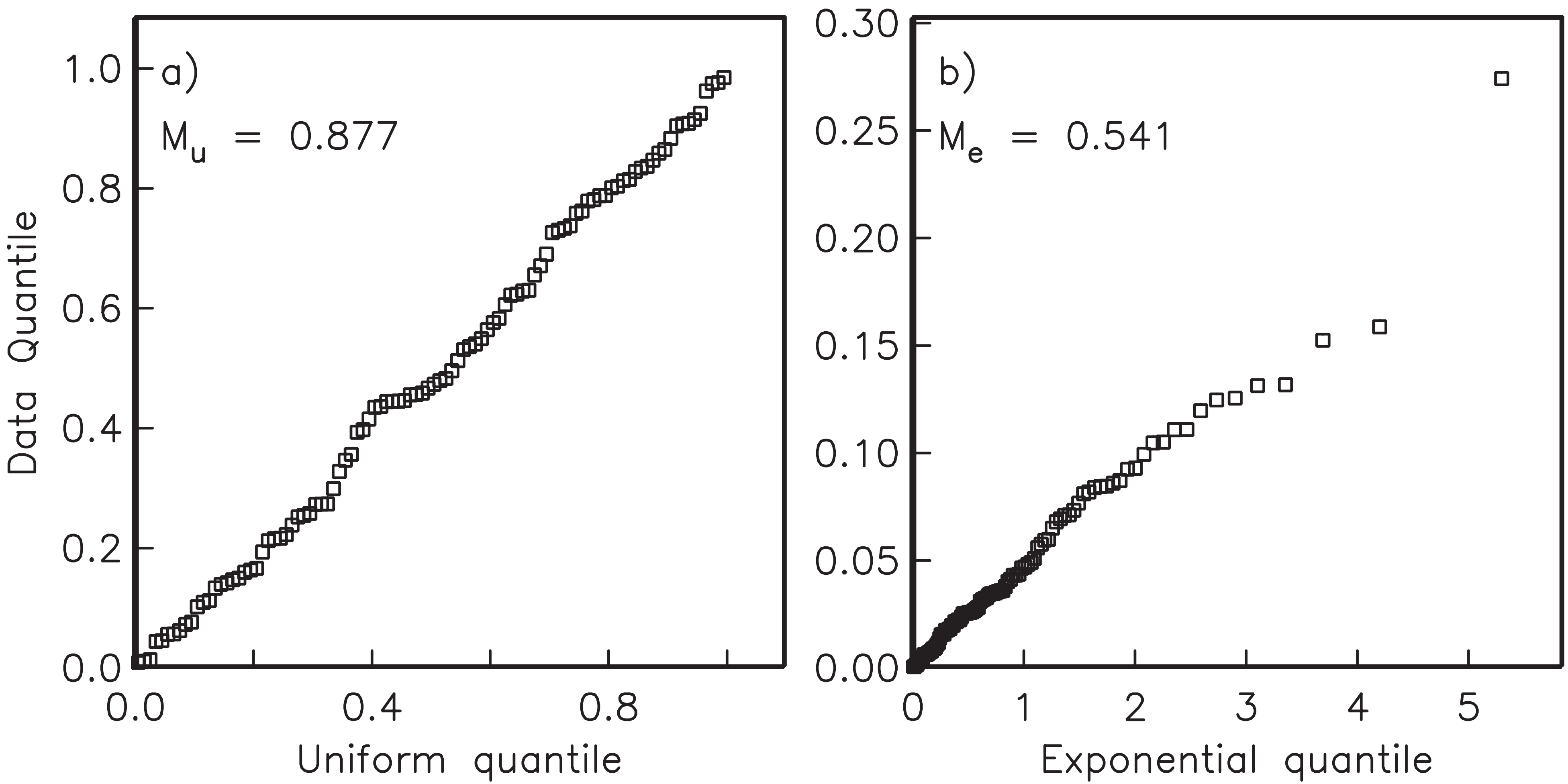
- First - transform the data set to the mean direction (see Chapter 11 for details)



Remember Fisher declinations are uniform and inclinations are exponential



If either Mu or Me exceed critical values, not a Fisher distribution

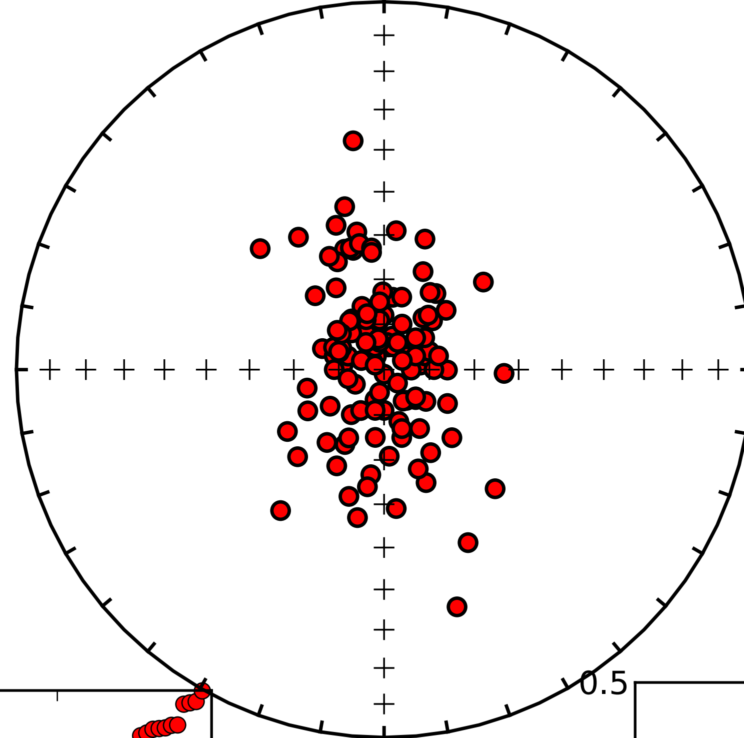


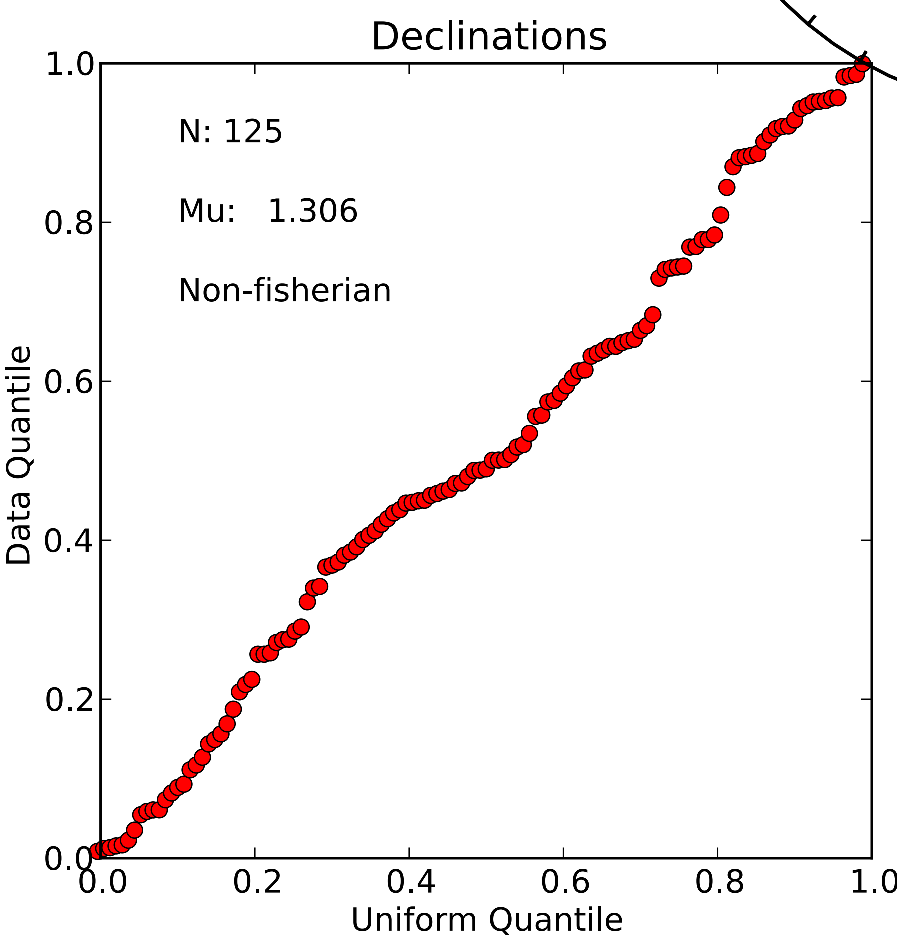
can do this with fishqq.py in PmagPy distribution but only good for large data sets (N~100)

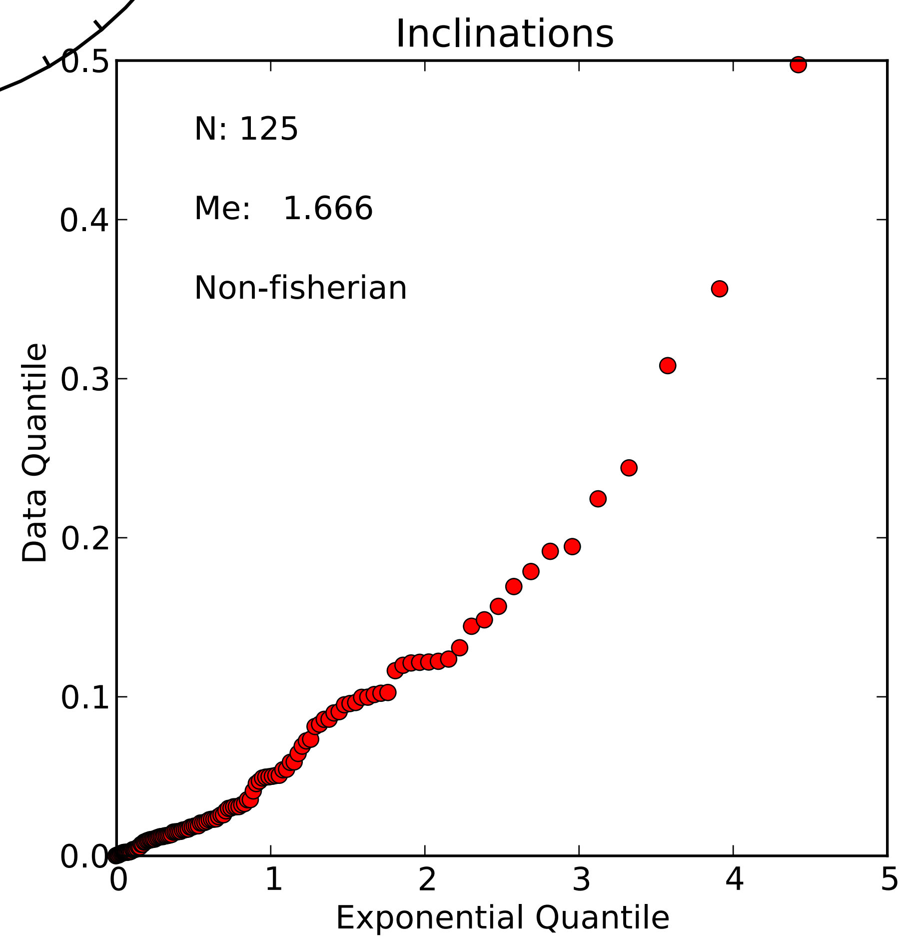
What to do if your data aren't Fisherian

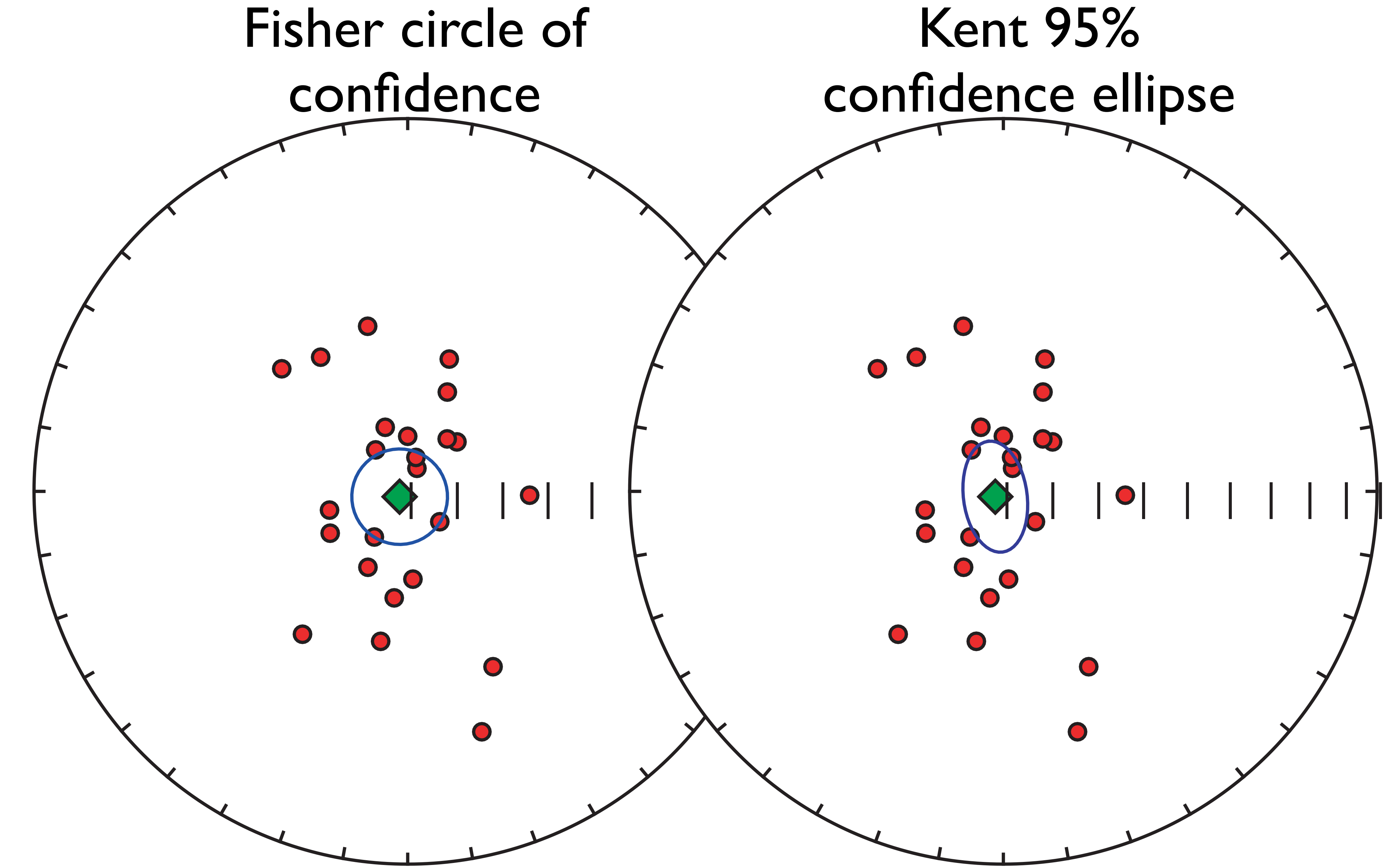
- Parametric confidence ellipses (see Chapter 12)
  - Kent distribution
  - Bingham distributions
  - These don't have nifty tests for common mean, etc.
- Non-parametric (bootstrap)

Non-fisherian data set









$$F = c(\kappa, \beta)^{-1} \exp(\kappa \cos \alpha + \beta \sin^2 \alpha \cos 2\phi).$$

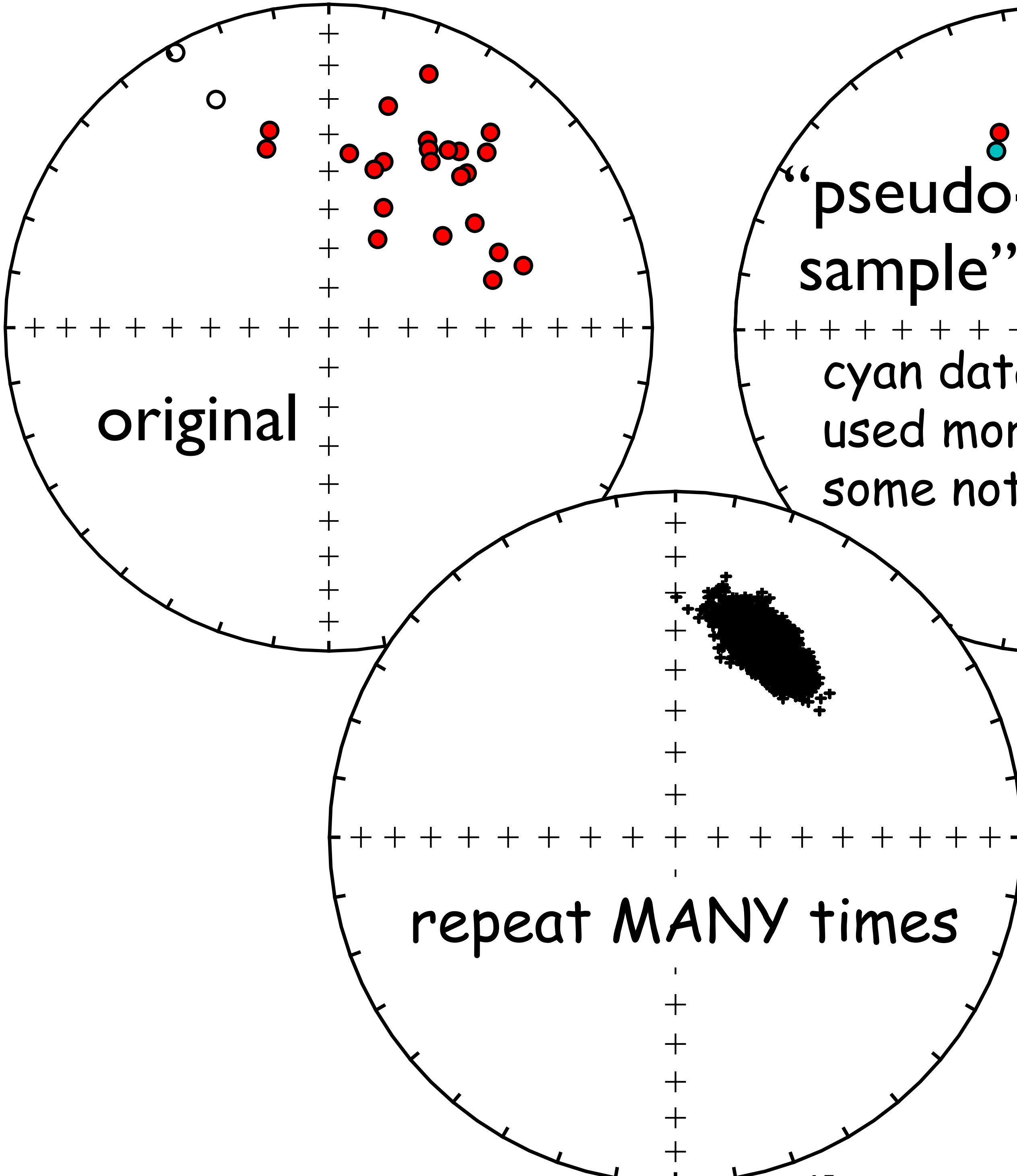
Kent like Fisher, but with "ovalness" parameter, β

- Kent is nice (allows elliptical data distributions)
- But one major cause for non-Fisherian data is reversals!
- Bingham distribution (based on eigenvector of orientation tensor and not vector mean) allows for bi-polar data - see Chapter 12
- BUT. does not allow a test for whether normal and reverse data are antipodal.
- AND - none of these has the handy tests available for Fisherian data (Vw, etc.)

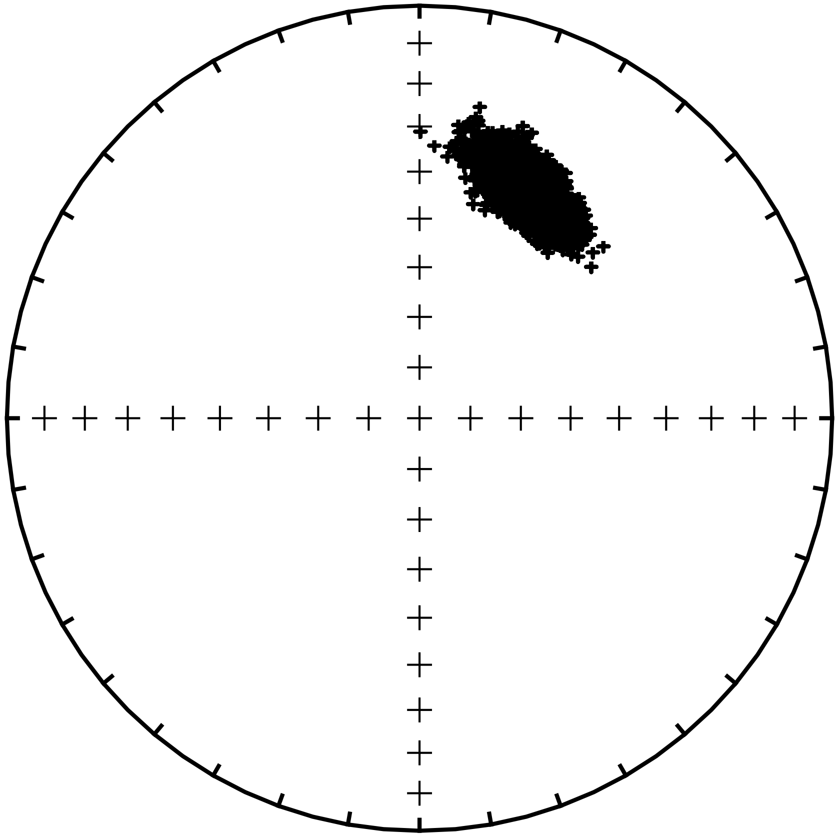
Non-parametric approach (the bootstrap)

- The bootstrap is like the Monte Carlo test we encountered earlier.
- Calculate parameter of interest (e.g., the mean direction) for random samples of the original data many many times (>1000)
- Each resampled data set (called a "pseudo-sample") has the same N

calculate vector mean



Maps out probability distribution of means

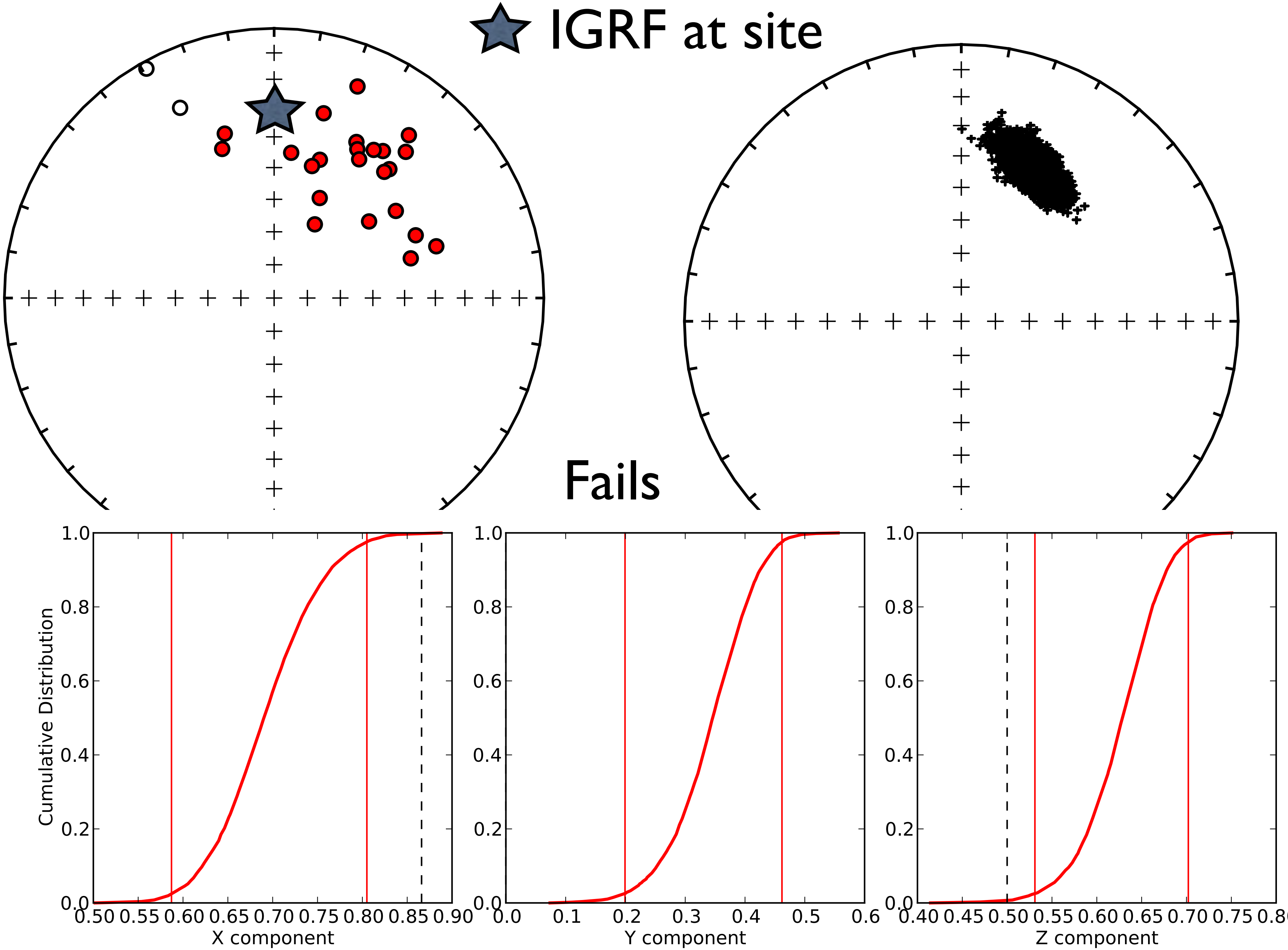


Now you have some options

- If you want ellipses, you can assume a distribution for the MEANS (e.g., Kent)
- You can test your hypothesis with the components of the bootstrapped mean vectors directly (preferred)

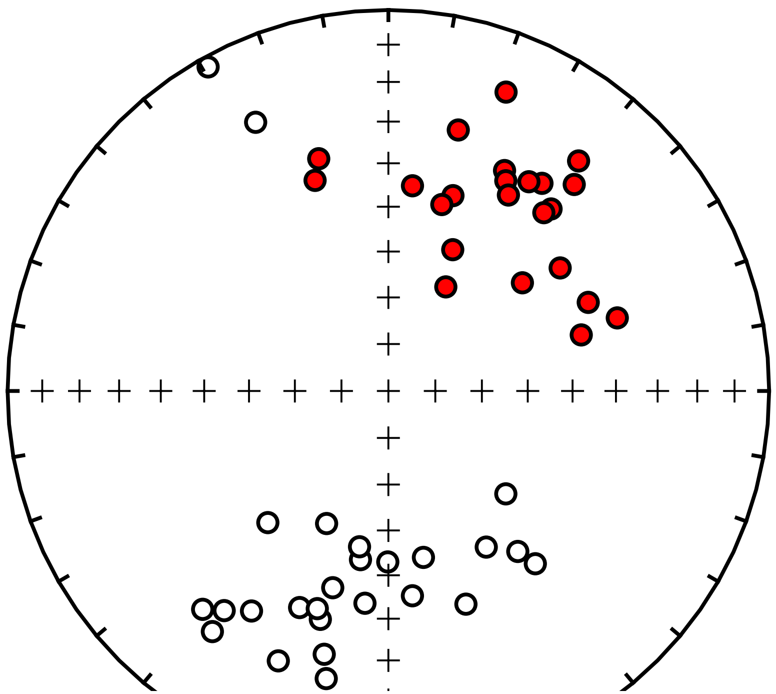
Test for common direction

- Comparison of paleomagnetic direction with known direction (IGRF value)
- Comparison of one paleomagnetic direction with another
  - normal and reverse data from the same study (the "reversals test")
  - data from different studies or locations
  - direction predicted from a reconstructed location or paleomagnetic pole



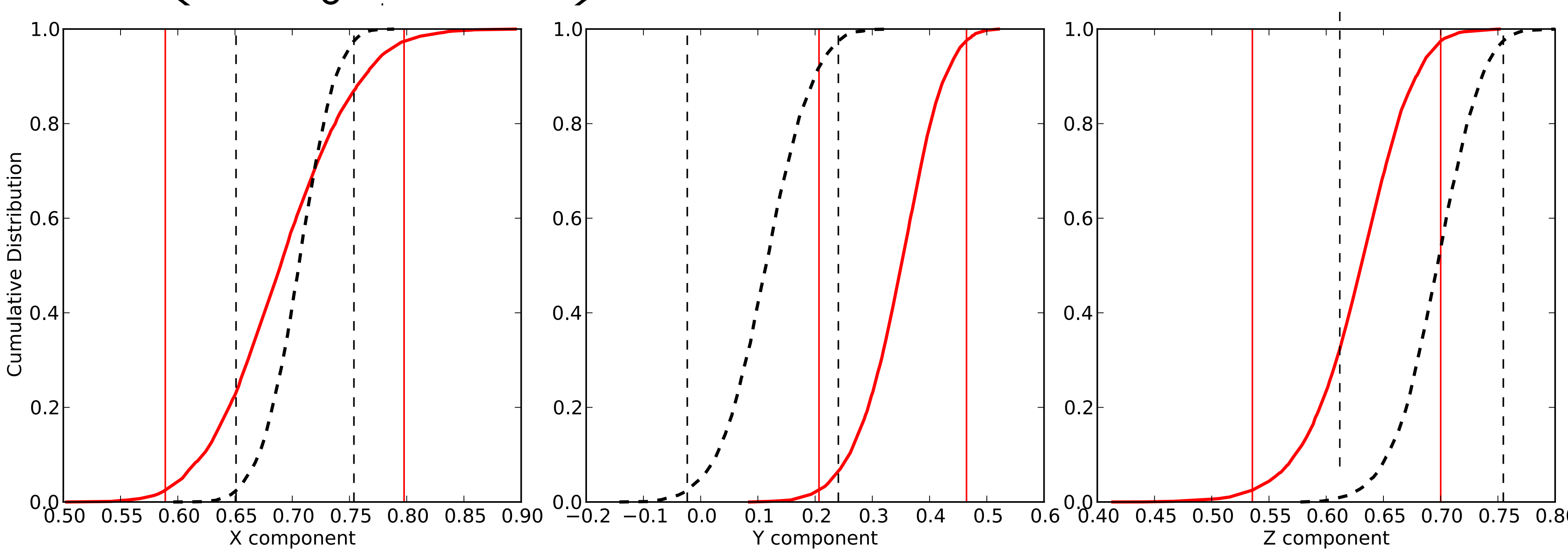
IGRF at site
Fails
Cumulative Distribution
X component
Y component
Z component
0.0
0.2
0.4
0.6
0.8
1.0
0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90
0.0 0.1 0.2 0.3 0.4 0.5 0.6
0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80

Two sets of directions



Reversals test
compare normal mode with antipodes of reverse mode

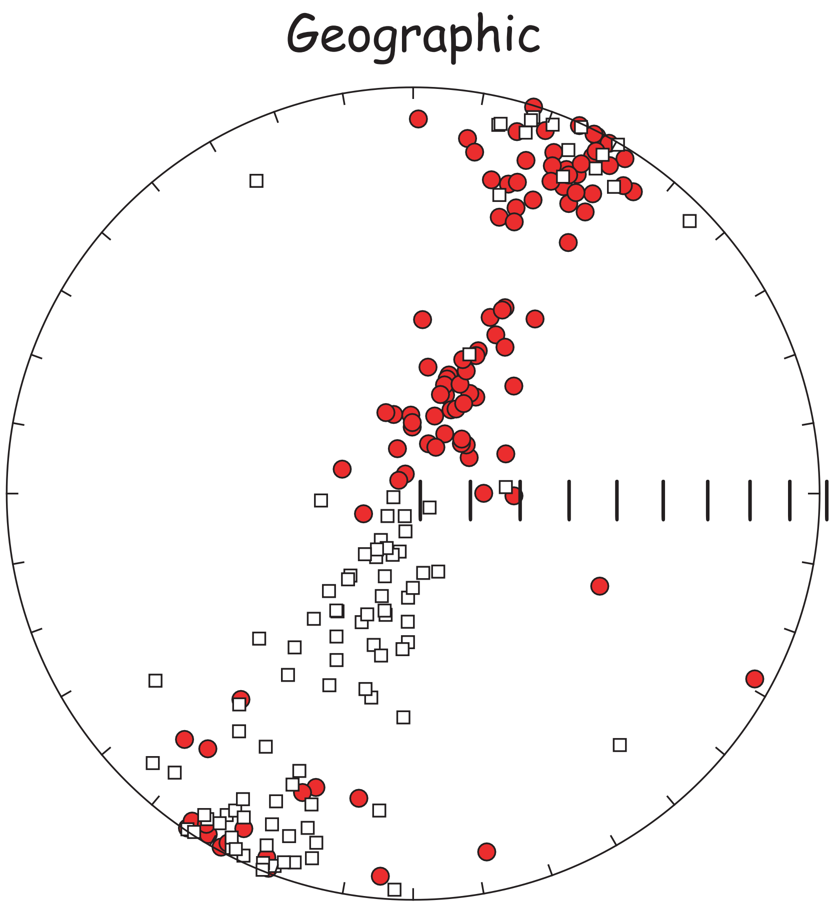
Passes!

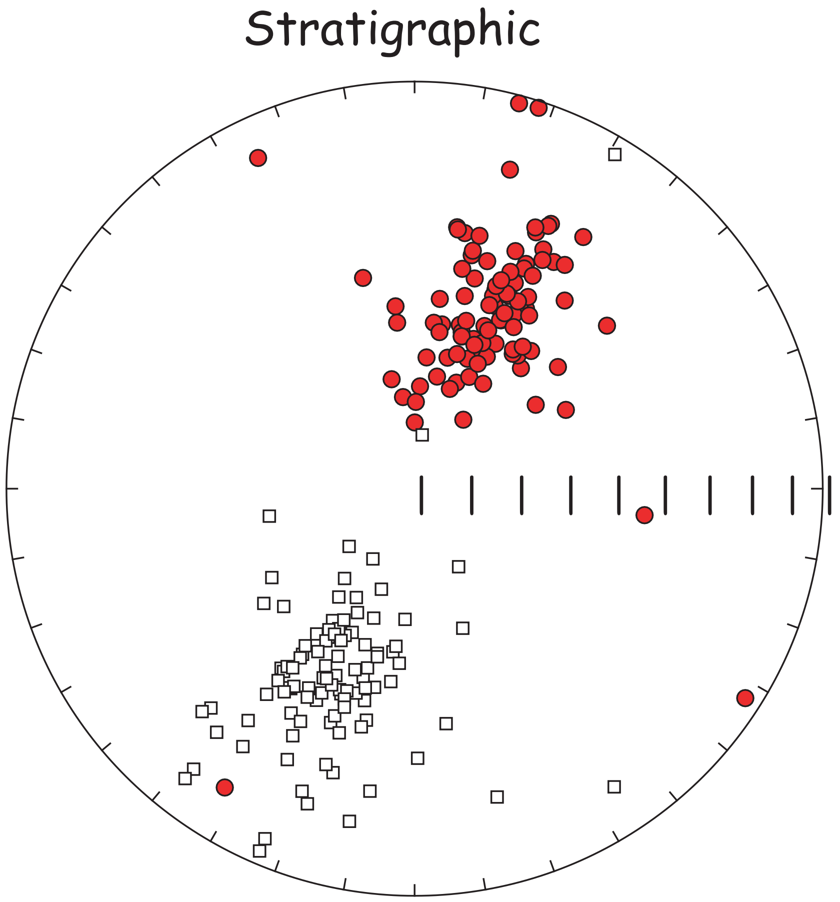


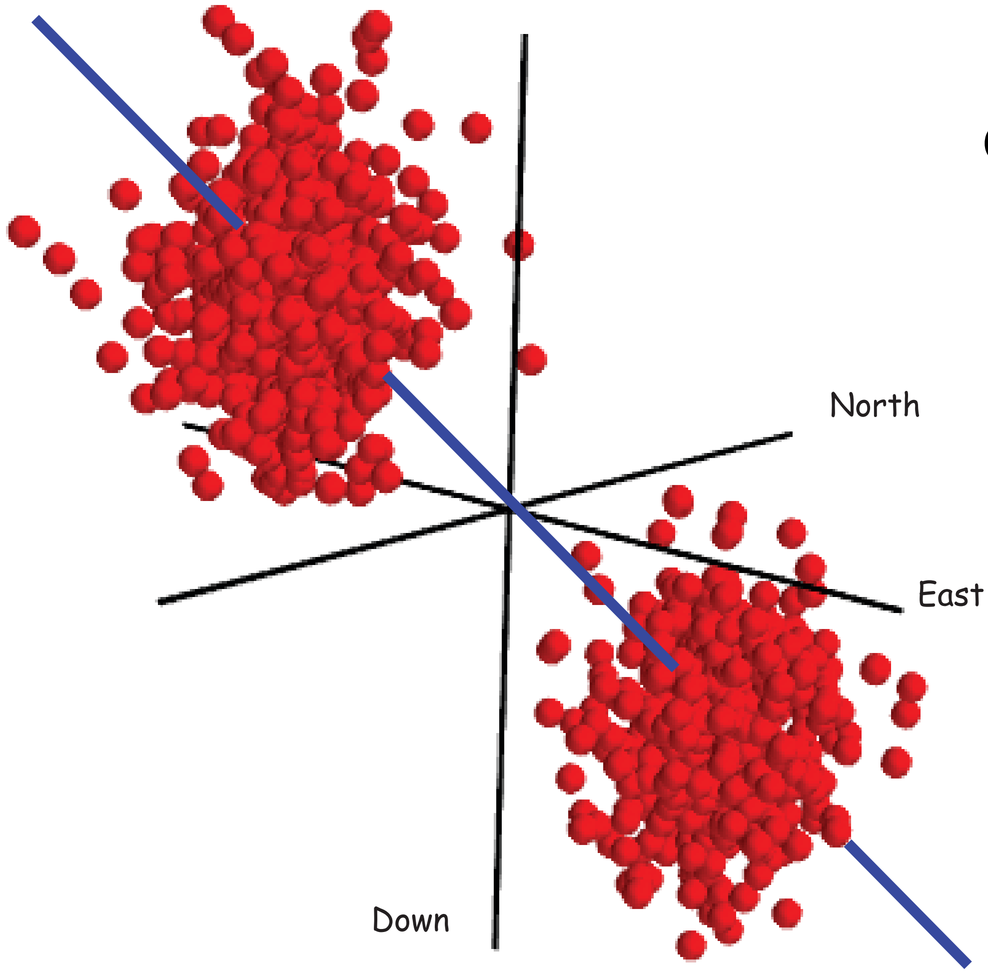
Foldtest

- Relies on testing whether directions are better grouped before or after correcting for tilt
- Many versions in the literature - they all give pretty much the same answer....
- The bootstrap approach does not require separation of data into normal and reverse modes or arbitrary groupings of data
- Simply calculate eigenparameters of orientation matrix as function of untilting
- Perform bootstrap to get bounds on tightest grouping

Example

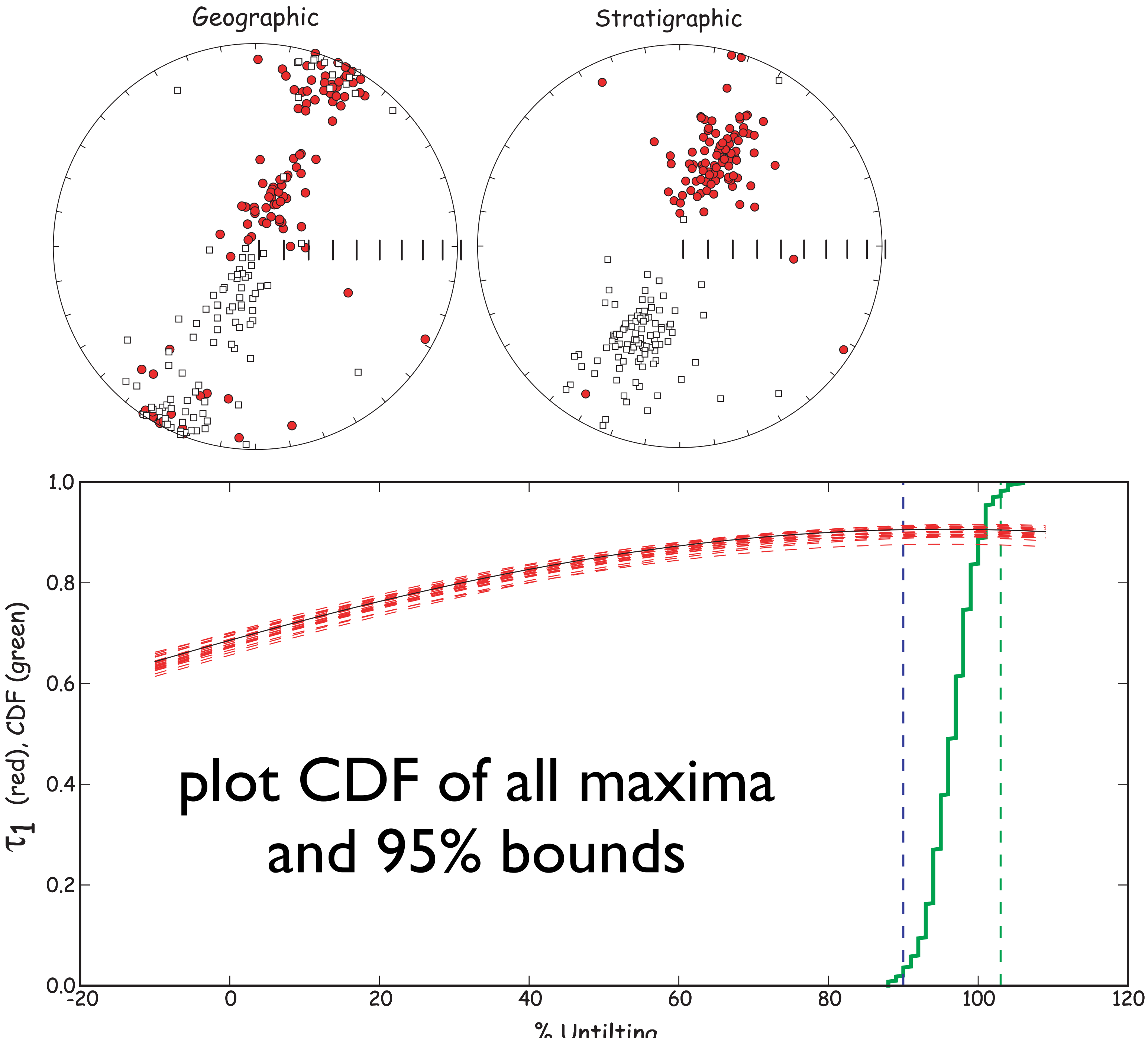






Reminder

convert all directions to unit vectors, then calculate eigenparameters. Blue line is "principal" ($\mathbf{V_1}$) corresponding to most variance (τ_1)



plot CDF of all maxima and 95% bounds

Let’s talk about possible project topics