EARTH SCIENCES: ISSUES AND QUESTIONS CONCERNING THE MULTIVARIATE APPROACH

Gianfausto SALVADORI

Dipartimento di Matematica e Fisica
Università del Salento
Lecce (Italy)

gianfausto.salvadori@unisalento.it

http://www.unisalento.it/people/gianfausto.salvadori

Columbia University Statistics Department
New York, October 11–12, 2013

http://dependence2013.wikischolars.columbia.edu/Copula+Conference

(1) CHOICE OF THE COPULA MODEL

A frequent request of practitioners is:

WHICH COPULA MODEL SHOULD I USE?

Apparently, at present, the Theory of Copulas is only able to suggest which models should NOT be used: e.g., do not use a Gumbel copula for discordant variables (with Kendall's $\tau < 0$).

The question is important and legitimate. The empirical answer is:

USE ANY COPULA THAT "REASONABLY FITS" THE DATA

but see point (2) later...

QUESTION: would it be possible to construct copulas from "first principles", possibly taking into account the physics of the phenomenon to be modeled? viz.: "this is the phenomenon, give me the right copula!"

(2) MULTIVARIATE GOODNESS-OF-FIT TESTS

Several multivariate Goodness-of-Fit tests exist to check whether a given copula model should be rejected for the phenomenon of interest.

The question is:

HOW MUCH POWERFUL ARE THESE TESTS
IF ONLY "SMALL" SAMPLE SIZES ARE AVAILABLE?

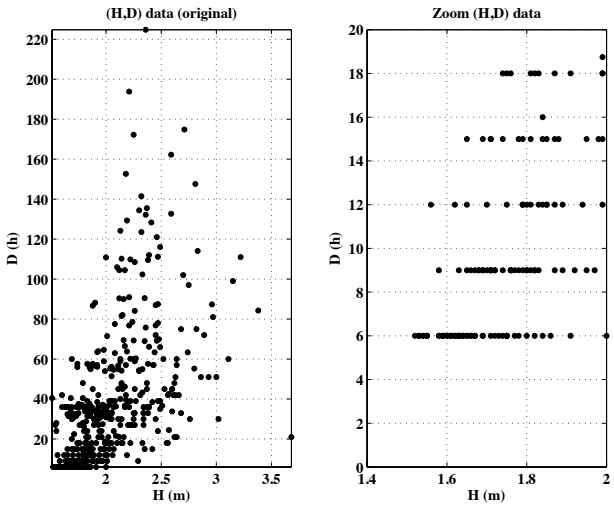
Earth scientists are not "as rich as" the colleagues working in Finance. For instance, in hydrology, usually time series of annual maxima are only 50 years long, or less. In coastal engineering, usually buoy observations are only 10-20 years long, or less. Similarly in other fields...

Typically, multivariate Goodness-of-Fit tests have only asymptotic justifications. Unfortunately, the sample size of environmental data bases will never attain such limit conditions. CAN THE TESTS BE IMPROVED?

(3) RANDOMIZATION

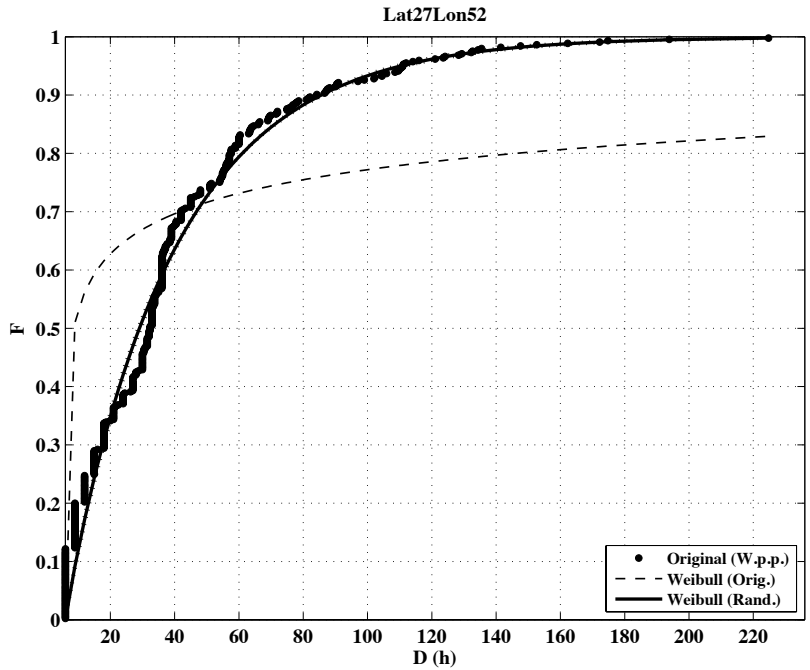
Quite often, due to limited instrumental resolutions or sampling procedures, environmental data bases report "repeated observations" (Ties). Ties are a serious complication for univariate and multivariate analyses.

Ties affect any non-parametric statistics based on the Ranks.



ANY WISE SOLUTION TO COPE WITH TIES?

Sea storm Duration data



Randomization may help!

(4) FURTHER OPEN (MULTIVARIATE) PROBLEMS

1. DEFINITION OF A RISK NOTION

   Introducing a coherent notion of Risk in a multivariate setting is quite troublesome, especially if the variables at play are not homogeneous, as it happens in environmental sciences.

   In Earth Sciences the Risk is usually a function of the Hazard, the Impact, and the Vulnerability. The only possible approach looks via the use of STRUCTURAL FAILURE FUNCTIONS (an empirical strategy). ANY CLEVER IDEA?

2. CLIMATE CHANGE (?) AND NON-STATIONARITY

   Assuming that a climate change is active (but data are contradictory! — Change-of-Distribution tests may help to check this), the assumption of i.i.d. samples fails. However, the available SAMPLE SIZES LOOK INSUFFICIENT TO PROPERLY WORK on a short time horizon. HOW TO COPE WITH IT?