International Precipitation Conference (IPC10)
Coimbra, Portugal, 23 - 25 June 2010
Topic 4 Instrumentation and novel precipitation observing systems

A note of caution for consistency checking and correcting methods of point precipitation records

D. Koutsoyiannis
Department of Water Resources and Environmental Engineering
National Technical University of Athens
(http://www.itia.ntua.gr/en/docinfo/1002/)

1. Abstract

Point precipitation data are routinely subject to consistency checking and adjustment of emerging inconsistencies, a process also known as data homogenization. The double mass curve is the most popular method of this type. While this is a graphical and empirical method with a high degree of subjectivity, there exist more objective and statistically sound versions. However, all versions tacitly rely on the assumption that precipitation is independent in time over long (e.g. annual) time scales. On the other hand, long precipitation time series reveal that they may exhibit long-range dependence, also known as the Hurst-Kolmogorov (HK) behaviour. A simulation study is performed, which shows that under HK behaviour different slopes appearing in the double mass curve are regular and do not necessarily indicate data inconsistency or inhomogeneity. Thus, application of the routine method to correct the data in fact modifies correct measurements, which are rendered inconsistent. Thus, if we hypothesize that the HK behaviour is common in precipitation, application of such methods may enormously distort correct data, based on a vicious circle logic: (a) we assume time independence of the rainfall process; (b) we interpret manifestation of dependence (the HK behaviour in particular) as incorrectness of data; (c) we modify the data so as to remove the influence of dependence; (d) we obtain series much closer to the faulty assumption of independence. The caution derived from the simulation study is that such methods should never be applied blindly. Unless information on local conditions and station archive justify that inconsistencies or errors exist, corrections of data should be avoided.

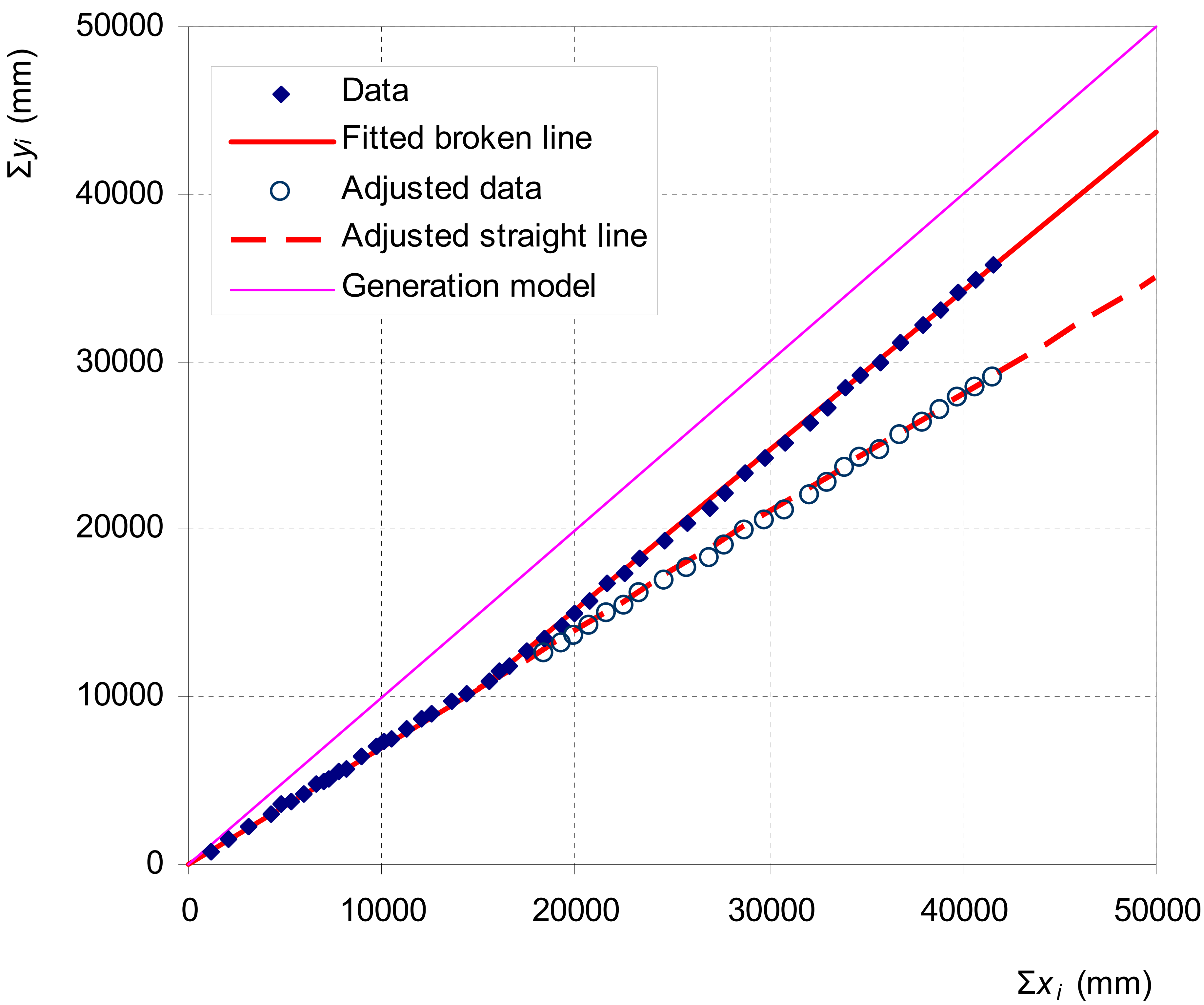
2. The example of the double mass curve in consistency checking and correcting of rainfall data

- The example used is the most common in routine data quality checking and correcting.
- It is rather empirical and graphical, but its logic is followed even in more statistically sound methods (e.g. Worsley, 1983).
- The double mass curve is a plot of the successive cumulative annual precipitation Σy_i at the gauge that is checked versus the successive cumulative annual precipitation Σx_i for the same period of a control gauge (or the average of several gauges in the same region).
- Provided that the stations are close to each other and lie in a climatically homogeneous region, the annual values should be correlated to each other.
- A fortiori, provided that the two series are consistent with each other, the cumulated values Σy_i and Σx_i are expected to follow a proportionality relationship.
- A departure from this proportionality can be interpreted as a systematic error or inconsistency, which should be corrected. Such a departure is usually reflected in a change in the slope of the trend of the plotted points.
- The ratio λ of the values of slopes is used to correct the data so that the broken line becomes a straight line.
- The method is typically applied for correction of as short as 5-year trends (e.g. Dingman, 1994)

Worsley, K. J. (1983) Testing a two-phase multiple regression, *Technometrics*, 25(1), 35-41.
Dingman, S. L. (1994), *Physical Hydrology*, Prentice Hall, Englewood Cliffs, New Jersey.

3. Illustration of the method of double mass curve

- Typical double mass curve for 50 pairs of values representing annual precipitation at two points, whose cross-correlation (between x_i and y_i) is 0.82.
- The first 25 (newest, plotted on the left) and the last 25 (oldest, on the right) form slopes $m = 0.7$ and $m' = 0.95$, respectively.
- The adjusted points, also shown, have been calculated using a multiplicative factor $\lambda = m/m' = 0.737$.



4. What is the problem of the method?

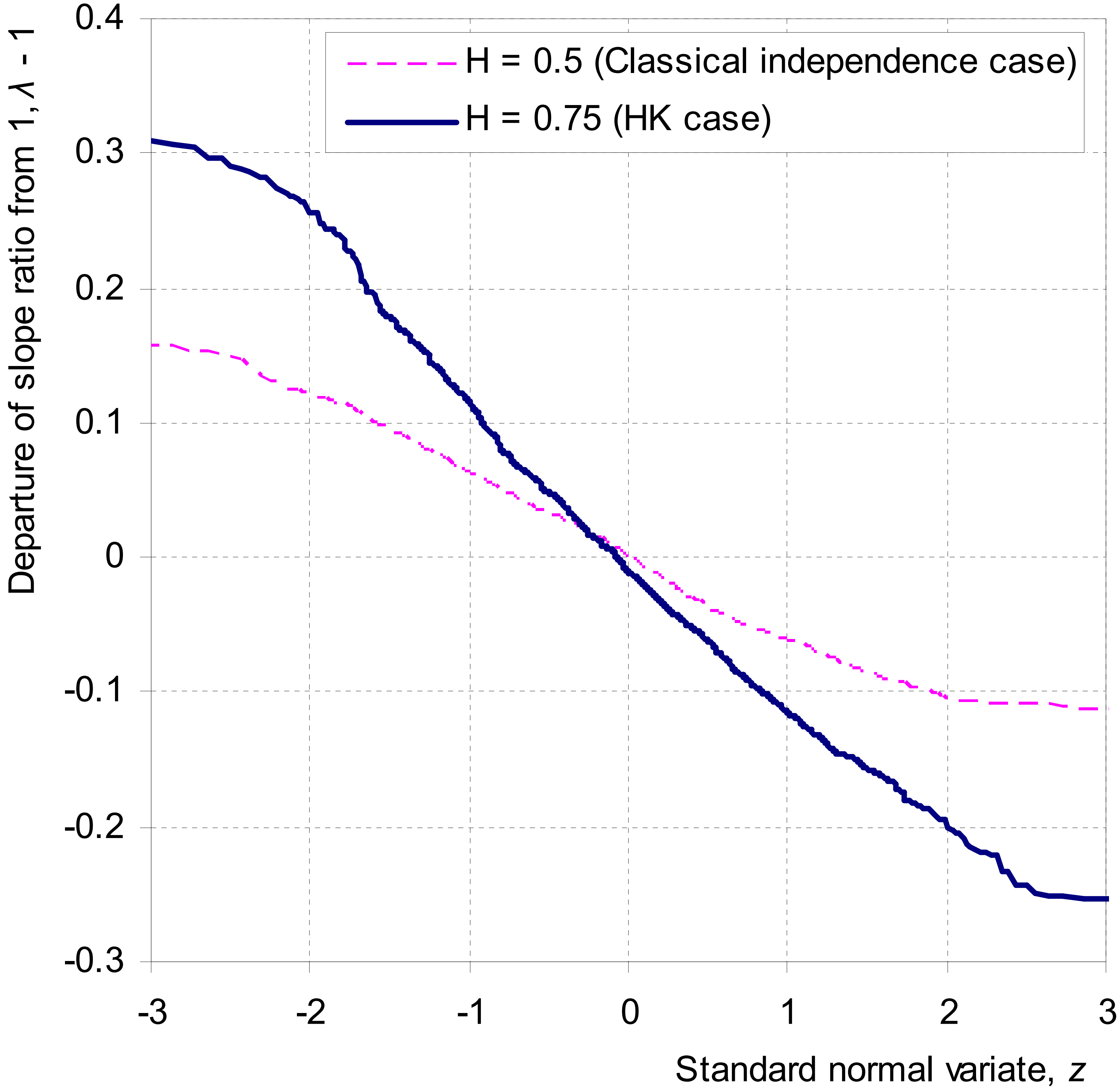
- The data values in the figure are generated so that both stations have equal mean and standard deviation (1000 and 250 mm, respectively), be correlated to each other (with correlation 0.71) and, most importantly, exhibit Hurst-Kolmogorov (HK) behaviour (with $H = 0.75$, compatible with the values found in the real world time series).
- Hence, evidently, all values are *correct*, *consistent*, and *homogeneous*, because they were produced by the same model assuming no change in its parameters.
- Thus, the example illustrates that the method can be dangerous, as it can modify measurements, seemingly inconsistent, which however are correct.
- While this risk inheres even in time independent series, it is largely magnified in the presence of HK behaviour (see figure below). For the same probability, the departure of λ from unity in the HK case is twice as high as in the classical independence case.
- For the HK case, departures of ±0.25 from unity appear to be quite normal for 25-year trends and even more so for finer time scales, i.e. ±0.35 to ±0.40 for 10-year to 5-year consecutive trends (not shown in figure).
- Application of the method most probably results in distortion rather than correction of hydrological time series.
- The "correction" of the series removes "trends" in one of the two time series.
- Removal of trends results in reduction of the estimated Hurst coefficient or even elimination of the exhibited HK behaviour (Koutsoyiannis, 2003).
- The widespread use of the double mass curve method in routine processing of precipitation time series may thus have caused enormous distortion of real history of precipitation at numerous stations worldwide, also hiding the HK behaviour.

A definition of the Hurst-Kolmogorov process on discrete time: The process whose standard deviation $\sigma^{(k)}$ at time scale k is given as $\sigma^{(k)} = k^{H-1} \sigma$, where H is the Hurst coefficient: $H = 0.5$ indicates a random process and $0.5 < H < 1$ indicates long term persistence.

Koutsoyiannis D (2000) A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resources Research*, 36 (6), 1519–1533.
Koutsoyiannis, D. (2002) The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrological Sciences Journal*, 47 (4), 573–595.

5. A Monte Carlo experiment

- Comparison of probability distributions (on normal probability plot) of the departure of the ratio of slopes λ from unity for series:
  - independent in time
  - with HK behaviour with $H = 0.75$
- The distributions were calculated using a Monte Carlo method based on synthetic series with a total size of 1000.



6. Discussion

- Most data pre-processing techniques are influenced by tacit assumptions inspired by classical (dice-throwing) statistics, such as light (exponential) distribution tails and time independence.
- Real world time series exhibit heavy tails and time dependence, which result, respectively, in intense and frequent extremes, and in grouping of similar states in time.
- Thus, routine methods perform procrustean operations on the time series:
  - They eliminate the extreme values that occurred in reality;
  - They eliminate the long-term variability implied by the HK behaviour.
- The effect of both these procrustean operations on data is a serious underestimation of design precipitation and flow in engineering constructions and management decisions.
- Thus, all methods of this type should never be applied blindly.
- An inspection of local conditions (environment of the gauging station, practices followed by the observer) as well as of the station's archive history is necessary before any action is taken towards altering the data.
- Unless information on local conditions and archive history justify that inconsistencies or errors exist, corrections of data should be avoided.