Hurst, Joseph, colours and noises: The importance of names in an important natural behaviour

Demetris Koutsoyiannis

Department of Water Resources, Faculty of Civil Engineering, National Technical University of Athens, dk@itia.ntua.gr - http://www.itia.ntua.gr/dk

It is maintained that the names and terminology used in the study of a scientific concept are closely related to its understanding and explanation, at least at the initial stages of the study, and that bad names reflect (or even result in) difficulties in understanding. Furthermore, the basic terminology used in the study of the scaling behaviour or the Hurst phenomenon are reviewed and it is maintained that most of them are unfortunate.

Posted in http://landshape.org/enm/?p=25, March 2006

What's in a name? That which we call a rose
By any other name would smell as sweet.

William Shakespeare, "Romeo and Juliet, Act 2 scene 2

Introduction

Is the name given to a physical phenomenon or in a scientific concept (e.g. a mathematical object) really unimportant? Let us start with a characteristic example, the term "regression". The term was coined by Frances Galton who studied biological data and noticed that the offspring population were closer to the overall mean size than the parent population. For example, sons of unusually short fathers have heights typically closer to the mean height than their fathers. Today we know that this does not manifest a peculiar biological phenomenon but a normal and global statistical behaviour. The slope of the least squares straight line of two variables x and y is $r_{xy}\, s_y / s_x$, where s_x and s_y are the standard deviations of the variables and r_{xy} is the correlation coefficient. In the example of the height of fathers and sons, $s_x = s_y$, so the slope is precisely r_{xy}, which (by definition) is not greater than one; hence the "regression" towards the mean. Today no one has any problem with this generally accepted term, even though clearly it is not a good name. No one has problem to understand the statistical (rather than biological or physical) origin of the "regression" and its irrelevance with time: For example the fathers of exceptionally short people also tend to be closer to the mean than their sons. Just interchange y and x (and the axes in the graph) and you will have again another line whose slope (in the new graph) will be again r_{xy}, that is, not greater than unity. However, until people understood these simple truths, the improper term must have caused several fallacies (see Regression fallacies in the Wikipedia article “Regression toward the mean”, http://en.wikipedia.org/wiki/Regression_toward_the_mean).

Thus, it could be maintained that, at least at the initial stages of the study of a scientific concept and before its establishment and wide dissemination, the names used are closely related to understanding and explanation. Vit Klemes (1974), in a pioneering and famous paper notes how important the explanation and understanding is and how a model can hinder them: “Indeed it is the very success of an operational model that by diverting further attention from the problem, often delays satisfactory explanation and understanding. I think that a bad name can, too, hinder explanation and understanding, and that poor explanation and understanding puts bounds to progress in modelling. Klemes uses the example of the Ptolemaic planetary model saying “It was exactly because it ‘worked’ so well (its predictions of position of stars were more than accurate enough for the contemporary needs) that it hampered progress in astronomy for centuries. Here I could add that the geocentric model of Ptolemy (90-168 AD), despite its successful predictions was in fact a regression (with the literal rather than the statistical meaning of the word “regression”): four centuries earlier Aristarchus of Samos (310-230 BC) had formulated the heliocentric model of the solar system (1800 years before Copernicus, who admits this in a note), and figured out how to measure the distances to and sizes of the Sun and the Moon. And at about the same time, Eratosthenes (276-194 BC) measured with an error of only 3% the circumference of the earth, based on the angle of the sun’s rays at different places at noon; this happened 1700 years before Columbus, who must have used Ptolemy’s estimate (underestimated by about 30%) of the circumference of the earth, thus giving the incorrect name “Indians” to the people of the new continent.

Coming back to the geophysical behaviour that Hurst discovered, it is interesting to quote again Vit Klemes: “Fortunately the success of fractional noises does not seem to be so universal that it could pose a similar danger to progress in hydrology and related sciences. My interpretation of this sentence is that Klemes (a) makes a clear distinction of the natural Hurst phenomenon and its mathematical modelling; (b) he disapproves the “fractional noise” as a model for this behaviour; but (c) simultaneously accepts the natural behaviour and seeks for an explanation of it (as seen from the entire context). I concur with Vit Klemes that it is important to make the distinction of the natural behaviour and the mathematical model; the fact that some fail to do this distinction always creates confusion. Besides, any mathematical model is only an approximation of reality; so one has the right not to like even a successful model and seek for a better one. In my opinion, “fractional Gaussian noise” is a good model (if demystified somewhat) but its name is not good. Nevertheless, this is not the only name associated with the Hurst phenomenon; both the natural behaviour and the models devised have been given a plethora of names, which alone creates confusion. Besides, several of these names are not good enough. In my opinion, the inappropriate names is one of the reasons (obviously there are additional ones that are not discussed here) that this natural behaviour has been regarded as a puzzle or a mystery, perhaps metaphysical, and that its consequences were not understood or were neglected, more than half a century after its discovery by Hurst.

Here are the lists of names (perhaps not complete) separated into two categories, names for the natural behaviour and names for models, along with my comments on the names.

A. Natural behaviour

A1. Hurst phenomenon: This is the best in my opinion; it respectfully attributes the behaviour to the engineer E. H. Hurst who discovered and studied it in geophysics. Here I wish to point out that the some people have identified the behaviour with the properties of a statistic called ‘rescaled range’ that Hurst used to report the behaviour. This, in my opinion, is not an ideal statistic (see Koutsoyiannis, 2002, 2003, 2006) and there is no reason to continue using it today and to identify the phenomenon with properties of this statistic.

A2. Joseph effect: This was coined by Benoit Mandelbrot (1977) who associated it to the biblical story of the seven fat and the seven thin cows. I have used it in talks addressed to general audience and I found that it helps people to approach the concept. However, the periodicity it implies and its association with the “magical” number seven (which some in the audiences have tried to point out) do not make it a good scientific term.

A3. Long memory: This is the worst name in my opinion. It stimulates people to imagine a mechanism inducing long memory (e.g. hundreds of years) and of course it is difficult to conceptualize such a mechanism. On the contrary, the mechanism dominating in this behaviour could be better characterized as absence of memory, as I tried to explain elsewhere (Koutsoyiannis, 2002).

A4. Long-range dependence: It is better than ‘long memory’ as it is free of the metaphorical meaning of ‘memory’. It is mathematically precise, so it is good to be used to describe a property of a model (that is, a stochastic process). However, it may be misleading in describing a natural behaviour and it does not point to any physical mechanism.

A5. Long-term persistence: ‘Persistence’ is a term more understandable, in physical terms, than ‘dependence’, whose conceptualization lies in stochastic processes. The term however implies some mystery, as in ‘Joseph effect’.

A6: Scaling behaviour: This is a concise and fashionable term, expressing the equivalence of (time) scales in this behaviour. I have used it a lot. The problem is that scaling is not a physical mechanism but a result of one or more other physical mechanisms or principles (perhaps the maximum entropy, as I tried to show in Koutsoyiannis, 2005a,b). Thus, it does not help understanding the physical concept.

A7. Multi-scale fluctuation: I have coined this term and I believe it demystifies the behaviour and makes it easily understandable. We are familiar with daily, seasonal and annual irregular fluctuations of weather and hydrologic quantities. If we expand these fluctuations at larger scales, say tens and hundreds of years (and there is no reason why we should not), then we obtain the Hurst behaviour (Koutsoyiannis, 2002).

B. Mathematical models

Firstly I should clarify that the names of models listed below do not refer to a single stochastic process but to two closely related processes; the fist (with names B1-B4) is a cumulative, continuous time, non-stationary process (such as in the cumulated rainfall depth at a site, which increases in time ever). The second (with names B5-B10) is the discrete-time stationary process that is obtained by taking the differences of the first process at equidistant times (such as in the annual rainfall depth); it could be also a continuous time process if derivative is used instead of difference. The second process more directly corresponds to what we study in geophysics and therefore is the most commonly used in branches of geophysics such as in climatology and hydrology.

A general observation is that several of the names of models contain the term “noise”. I do not find this a good term for geophysics. Generally, “noise” is used (e.g. in electronics, information and communication) in contrast to “signal” and the distinction implies that there is some signal that contains information, which is contaminated by a (random) noise. Noise should be identified and removed from the signal to recover the maximum of information. Such a distinction may not have a meaning in geophysics. The evolution in time of temperature or rainfall, as we measure it at a site, has the characteristics of ‘noise’ rather than those of a typical signal of anthropic origin, i.e. it is irregular or random. Yet it has some structure and certainly it is the “signal” of nature, so I do not think we could classify it as “noise”. In recent studies, some attempted to find the signal-to-noise ratio in hydrological time series. They successfully applied algorithms (e.g. from the chaotic literature) to obtain a certain value of signal-to-noise ratio, but they failed to explain what signal and what noise represents. This failure is expected in my opinion, because nature’s signs are “signals” in their entirety even though look like “noise”.

B1. Self-similar process: This name is the most widely used in the mathematical literature today – but not so much in geophysics. It is a precise and concise name.

B2. Wiener spiral (or Wiener helix): This name (honouring the mathematician Norbert Wiener) was given to the process by the Russian mathematician Andrei Nikolaevich Kolmogorov (1940), who introduced and was the first to study it, to model turbulence. It is amazing that Kolmogorov introduced the process ten years before Hurst's celebrated paper and simultaneously that this contribution is so very little known to geophysical community (including myself). Thus, the name given by Kolmogorov is not at all used today, even though we still use the name "Wiener process" for the limiting form of the random walk process, which is a special (non-interesting, i.e. without Hurst behaviour) case of a self-similar process.

B3. Semi-stable process: This name was given to the process by the American mathematician John Lamperti (1962). Again this name (and perhaps Lamperti's significant contribution to its study) has been forgotten today.

B4. Fractional brownian noise. This is due to Benoit Mandelbrot (1965) and it is the most widespread. It may be a name mathematically rigorous, but I do not like it as a whole and each of the three terms separately. The first term, fractional, is not easily understandable, unless combined with fractals, which is not necessary. The second term, brownian, points to brownian motion (the movement of a particle in a liquid subjected to collisions and other forces), which again is not necessary as there not direct connection of the process with the brownian motion. The third term, noise, is unsuccessful as I described above.

B5. Stationary intervals of a self-similar process. This is a mathematically rigorous name of the discrete time stationary process but I think that it is too wordy and it is difficult to understand the meaning it communicates.

B6. Fractional gaussian noise. This corresponds to fractional brownian noise and again is due to Benoit Mandelbrot (1965). I do not like this name too for the reasons explained in B4 and the additional reason that it restricts our view to processes that are gaussian. The gaussian distribution may be not the case for several geophysical processes that are asymmetric (non-gaussian).

B7. *Fractional ARIMA process (abbreviated as FARIMA or ARFIMA)*. This is based on Hosking's (1984) work on fractional differencing (in fact meaning taking a weighted sum of infinite terms) of a Box-Jenkins ARMA process, which results in a process with long-range dependence. I do not think that this term communicates any information that helps in understanding.

B8. Red noise. I do not have enough information about the prevailing of this term and the message it contains (for example, the articles in Wikipedia “Colors of noise” – http://en.wikipedia.org/wiki/Colors_of_noise and “Red noise” – http://en.wikipedia.org/wiki/Red_noise are not very informative). Perhaps the “red” colour points to the fact that the multi-scale fluctuation, when studied in the frequency domain, results in high values of power spectrum for low frequencies. Besides, in the visible light, red is the colour with the lowest frequency. However, the scaling behaviour certainly could be not modelled as “monochromatic red”, because its power spectrum extends to the entire frequency domain. Besides, I do not like the term “noise” as I explained above. Therefore, I do not think that this name is successful.

B9. Brown noise. This has been used as synonymous to “red noise”. Perhaps it is better than “red noise”, as brown is not one of the colours of the visible light spectrum (it is a mixture of colours) and also reminds “Brownian”, as discussed above.

B10. Simple scaling signal or *Simple scaling stochastic process* (abbreviated as an *SSS process*). I have proposed these terms and the abbreviation (Koutsoyiannis, 2003) thinking that they are less misleading than other terms described above, more understandable and more helpful in understanding of the process described by this name. In fact, the definition of the process is a simple scaling relationship involving a power-law of time scale. The abbreviation SSS could be derived in other ways, too (observe that most of the terms above start with an ’s’); for example “stationarized self-smimilar” with “stationarized” standing either for taking the “stationary intervals” (or taking the difference at equidistant times as explained above) or for taking the “stationary derivative” (at any time instant) in the continuous time version.

References

Hosking, J. R. M. (1984), Modeling persistence in hydrological time series using fractional differencing, *Water Resources Research*, 20(12) 1898-1908.

H. E. Hurst (1950), Long-Term Storage Capacity of Reservoirs, *Proceedings of the American Society of Civil Engineers*, 76(11).

Klemes, V. (1974), The Hurst phenomenon: A puzzle?, *Water Resources Research*, 10(4) 675-688.

Kolmogorov, A. N. (1940), Wienersche Spiralen und einige andere interessante Kurven in Hilbertschen Raum, *Comptes Rendus (Doklady) Acad. Sci. USSR* (N.S.) 26, 115–118.

Koutsoyiannis, D. (2002), The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrological Sciences Journal*, 47(4), 573-595.

Koutsoyiannis, D. (2003), Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrological Sciences Journal*, 48(1), 3-24.

Koutsoyiannis, D. (2005a), Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling, *Hydrological Sciences Journal*, 50(3), 381-404.

Koutsoyiannis, D. (2005b), Uncertainty, entropy, scaling and hydrological stochastics, 2, Time dependence of hydrological processes and time scaling, *Hydrological Sciences Journal*, 50(3), 405-426.

Koutsoyiannis, D. (2006), Nonstationarity versus scaling in hydrology, *Journal of Hydrology*, (in press).

Lamperti, J. W. (1962), Semi-stable stochastic processes, *Transactions of the American Mathematical Society*, 104, 62-78.

Mandelbrot, B. B. (1965), Une classe de processus stochastiques homothetiques a soi: Application a la loi climatologique de H. E. Hurst, *Compte Rendus Academie Science*, 260, 3284-3277.

Mandelbrot, B. B. (1977), *The Fractal Geometry of Nature*, Freeman, New York.

Comment 1[*]

Regarding a question about removing periodic signals in analysis but not removing recent trends, my answer is this:

1. It is difficult, if not impossible, to remove anything from a time series of observations. We sometimes think that we can remove "periodic signals" by standardizing a time series on a sub-annual scale. For example, in a monthly time series we think we can remove periodicity by subtracting the mean and dividing by standard deviation of each month separately. This may be a widespread simple approximate technique but it cannot remove periodicity. Take, for instance the skewness coefficient for each moth of the standardized time series. This will be exactly the same as in the original time series, before standardization. Take the lag-one autocorrelation. Again, this will be exactly the same as in the original time series. (The proof is simple for both). As both skewness and autocorrelation vary with month in the original time series, they will vary in the standardized time series, too. Thus, we have not removed periodicity.

2. Things are even worse with trends. As I have tried to show in Koutsoyiannis (2006) a "trend" is an essential part of the time series and if removed it distorts the time series largely and gives a false impression for the natural behaviour. In addition, trend-looking patterns of natural time series are not trends but parts of large scale fluctuations. Besides, I do not think that anyone could define a trend in a time series objectively.

3. These comments do not mean that we cannot do anything to model natural phenomena with stochastic processes. Modelling does not presuppose decomposing of time series into parts such as "trendy", periodic and random components. This is a bad modelling practice, in my opinion. There exist consistent "holistic" modelling techniques.

4. Specifically, at annual and overannual time scales, we could use an SSS process as a model, which will reproduce the trend-looking behaviour, as I have tried to show in Koutsoyiannis (2006). If one wishes to reproduce exactly an observed "trend" in the past, there are two solutions: (a) One can use an explicit conditional simulation technique as described in Koutsoyiannis (2000). (b) One can use a Monte Carlo conditional simulation technique as in David Stockwell, "A new temperature reconstruction" by (http://landshape.org/enm/?p=15).

[*] Comment #2 in http://landshape.org/enm/?p=25.

5. At a sub-annual scale, the answer is to use a cyclostationary, rather than a stationary model, without any removal of periodicity. A simple model of this category is the periodic autoregressive model of order 1 (PAR(1)). This, however, will not reproduce the Hurst behaviour and the trend-looking patterns at overannual scales. Here there are at least two solutions: (a) One can couple PAR(1) or another simple model of this type that describes short scale cyclostationary properties, with an SSS process that takes account of large scale properties. This is described in Koutsoyiannis (2001). (b) One can use a cyclostationary SSS modelling strategy as described in Langousis and Koutsoyiannis (2006). Both cases are more difficult that typical stochastic modelling. However, the first model in Langousis and Koutsoyiannis (2006) is simple enough and can be implemented conveniently; also the single-site case of Koutsoyiannis (2001) associated with a PAR(1) model is simple enough and could be made even simpler (Koutsoyiannis and Manetas, 1996).

References

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. (2001), Coupling stochastic models of different time scales, Water Resources Research, 37(2), 379-392.

Koutsoyiannis, D. (2006), Nonstationarity versus scaling in hydrology, Journal of Hydrology (in press).

Koutsoyiannis, D., and A. Manetas (1996), Simple disaggregation by accurate adjusting procedures, Water Resources Research, 32(7), 2105-2117.

Langousis, A., and D. Koutsoyiannis (2006), A stochastic methodology for generation of seasonal time series reproducing overyear scaling, Journal of Hydrology (in press).

Comment 2[†]

Regarding the discussion on the definition of 'trends' and the suggestion that the word "trend" really just means average slope, my answer is this:

If you identify tend with slope, I can agree. In this case it is correct to say that now the temperature has a negative trend (because at this very moment that I am writing this note, here in Bologna it is afternoon so the temperature decreases with time). Also, we can speak of a "zero trend" – if the derivative is zero.

I can also agree if you identify trend with "average slope" but I would ask two questions: "average at which time scale?" and "slope at what time(s) exactly?" So, if you reply these questions, I think that there is not any problem, but on the other hand we do not add any information about the process at hand, introducing the term "trend".

The problem starts when we try to "dress up" the term "trend" with either of two properties:

(a) A property of "globality", implying that we have a phenomenon lasting for a long time or even forever, so that we feel obliged to describe it as a "nonstationarity".

(b) A property of "statistical significance", implying that this "trend" is something extraordinary, so that it really is a "nonstationarity". But statistical significance depends on the model we use in the null hypothesis; usually this is IID, which is good when we play dice but not good when we study natural or human-related phenomena.

In this respect, I do not favour the use of the term "trend" thinking that it creates problems rather than helps in understanding. In addition, I think that we should always distinguish models from reality and in case that we use models we should always do it consciously and admit it. For example, if we had determined a "statistically significant trend" using, say, Kendall's test, we must be conscious that we used a model, specifically the IID model, and that the result is specific to this model; perhaps it would be opposite if we had used another model, say the SSS model.

Having said that, I also think that it is not accurate to say "a series may have infinite variance". The infinite variance is for the process (i.e. model), not for the series (i.e. reality). A series of observations with finite length has always finite variance.

[†] Comment #18 in http://landshape.org/enm/?p=25.