"Hurst-Kolomogorov Dynamics and Uncertainty"

Demetris Koutsoyiannis, National Technical University, Athens

There are several links with the previous presentation. To start with, I would say that I fully adopt the "non-equation" that Harry Lins presented, and, because I want to be honest, I would say that I have some problem with the very title of this workshop, and I would be happier if instead of "non-stationarity" was the word "non-static". Non-static frequency analysis and water management. I'll explain in a minute my thoughts about this. At the same time, I wish to say that I feel proud and honored that I am part of this very important workshop, and I thank the organizers for inviting me.

Feeling proud, I tried to make something good, but unfortunately, my presentation, what I prepared, is very long; I will upload it to the address that you see here, so if anybody wants to see what I was prepared to present, it's there. Anyway. I'll run, but my English doesn't help me too much to tell all the things that I wanted to tell.

Okay. My motivation is a real world problem. This real world problem is Athens. You see Greece here. So Athens is a very small part of Greece. Yet it has 40 percent of the population. And at the same time, it's a very dry place. 400 millimeters per year of rainfall.

Through history, we had to transfer water from other places. Now the modern system transfers water from four rivers. I'll focus on this river, the Boeoticos Kephisos River, with a 2,000 square kilometers basin, for several reasons; for instance, there exist several hydraulic works, and these works have started to be constructed much earlier than '50s or Al Gore. The history of hydraulic works here starts 3.5 thousand years ago. And the other important reason is that we have a record of a hundred years of measurements in this river.

So going back to the '90s, when we started to get involved in this problem of the water supply of Athens, we observed that there was a falling trend in runoff, downward trend, which is very significant. Not from a statistical point of view, but it's substantial. Almost 50% of the initial amount here in the '80s. There is also a trend in rainfall which explains the trend in runoff.

And above all this, in the end of the '80s and the beginning of the '90s, we had a very severe drought. It lasted seven years. You see the average up to the '80s, and this is the average of these seven years. It's less than half.

So I would like to add something to what Bob Hirsch said before. When we talk about droughts, it's not a matter of the distribution tail. It's a matter of the dependence and the clustering in time, which determine the duration of drought. Of course, the tail is important, but the duration is more important.

Okay. On top of this, we had to prepare the Olympic games, so we applied the knowledge that I'll explain later. And you see in this slide some evidence that there was water in the Olympics games, so our methodology is not so bad, perhaps. This is part of the opening ceremony, you must remember.

Okay. So here are the mottos motivating my presentation. The first one is, "stationarity is dead". The title of this article from the Science Magazine includes a question about water management. And for this question, I have a second motto that I seized from a blog, which says, "Hydrologists' work is used by engineers to plan large-scale projects designed to last many decades." Well, we know it, but this is from a non-engineering and non-hydrological blog. So I think it's important to know how people think about us. "They can't play with models, especially models that so plainly diverge from reality."

And the last motto, is a Greek one, which translates, "The start of wisdom is the visit of names," by Antisthenes, who was founder of the Cynic philosophy. I think the motto applies also to the Cynic philosophy, but this is another story.

Let me visit the names of stationarity and non-stationarity. And first, I would like to say that in science, it's important, when we see change and motion, to find some invariant properties. And of course, Newton's laws are about invariant properties; for instance, the first law says that velocity is constant. The second law is that acceleration is constant, and even the law of gravitation says -- or implies some constant things, for instance, angular momentum in the planetary motion. But this is good for one or two bodies. If we have a complex system with three or more bodies, then things are more complex.

Here, I have an example of a synthetic time series that I have produced in some way that I'll explain later. And it's a very convenient assumption to make that there is a constant mean here. It's clear that this mean is not a material thing. We cannot measure the mean. We cannot because it's just a mental construction. Okay. But we say, or assume, that this mean could be constant.

Okay. The idea behind this is that we don't model the system in its details. We use macroscopic descriptions, and this is very important, and it's the only way we have, to use macroscopic descriptions.

So with this introduction, let me show you the definition of stationarity. I have copied it from Papoulis. It's my favorite book on stochastic processes, because it is a book oriented to engineers -- Papoulis himself was an engineer -- and is also very rigorous from a scientific point of view.

First, the definition of stationarity applies to stochastic processes. This should be clear. And second, when we speak of non-stationarity, we must expect that some property, some statistical property, is a deterministic function of time; not just a change in time but a deterministic change.

So to clarify this, we depict in this graphic the real world, and we have a natural system that we study, which has a unique evolution, only one trajectory in time; and if we observe this evolution, we have a time series.

Now in the other part of the graphic we have the abstract world, and we have the models. Of course, we can build many models for the natural system. And any one of the models can give us an ensemble. The ensemble contains mental copies of the real-world system. The idea of mental copies is due to Gibbs, known from thermodynamics, statistical thermodynamics. And then we have the stochastic process, from which we can generate synthetic samples and we can have different ensembles. Then, stationarity and non-stationarity apply here, to the abstract objects. Not to the real objects. So please, let's not kill ideas. Especially useful ideas like stationarity. And even non-stationarity.

This slide depicts the example I showed earlier but now extended up to time 100. Clearly, this red line which represents the mean is 1.8 for times smaller than 70, and then it becomes 3.5. Is here the stationarity dead?

Let me reformulate the question. Does this red line reflect a deterministic function? If it is a deterministic function, then we have non-stationarity. If it is not deterministic but a realization of a stationary stochastic process, then we have stationarity.

And now I will reveal the model that I used to produce the time series, and the model is a stochastic model. The shifts in the red line, which represent the changes of the local mean, occur at random. They are random fluctuations, with random amplitude. So everything in the model is stationary, and the entire model is a stationary model.

But there is a big difference -- it's not a matter of semantics -- to call something stationary or non-stationary. The big difference is this: If we have stationarity, if the "red line" is a deterministic function, then every mental copy would have the same "red line", and now the variance is this, 0.04. By construction. But if the red line is a realization of a stochastic process, then the mental copies should be different also in their "red lines", and then the variance is 0.38 -- that is, 10 times greater.

So stationary is not synonymous to static. Non-stationary is not synonymous to changing. In a non-stationary process, the change is described by a deterministic function, and the deterministic description should be constructed by deduction, the Aristoteleian apodeixis, not by the data.

So to claim non-stationarity, we must establish a causative relationship, construct a model describing the change as a deterministic function of time, and ensure applicability of the deterministic model in the future. Non-stationarity, as I explained, reduces uncertainty; and unjustified claim of non-stationarity results in underestimation of variability, uncertainty, and risk.

So what makes us think that we can work with non-stationarity? Do we have credible models that predict the future? Do these models provide credible predictions?

I would like to make a small, brief parenthesis to remind you of this Aesop's fable about this man who asserted that, when he went to Rhodes, he made a jump, an extraordinary jump, but someone from the people told him -- "Rhodes is right here. Make the jump." It's more famous in its Latin version, which is, "Hic Rhodus, hic saltus." So "hic Rhodus" is the 20th Century in our case, and "hic saltus" is the skill of the models to reproduce reality.

We have studied, several now, cases of how climate model outputs compare to reality, and you can see here, in this part of the presentation, that the efficiencies are, almost in hundred percent of the cases, zero or smaller than zero.

Perhaps for the U.S., this diagram is more representative, because we have, on the graphic, the results of three climate models for the entire U.S. The blue line is the reality, the actual average data. The actual annual average for the U.S. rainfall is 700 millimeters. One model predicts 800 millimeters. The other model predicts 900 millimeters. The other, almost 1,000 millimeters. And a model that failed by so much for the past, I don't think that is credible for the future.

And with this thought, I would shift to stationary descriptions of the future. Okay. If we adopt a stationary method, then we could use statistics of stationary processes. For instance, the autocorrelogram. If we had white noise, there would be a vertical line here in the autocorrelogram. If it was Markovian, it would be this line with the magenta color. The process that I demonstrated earlier has this autocorrelogram, the blue line. So what moved the autocorrelogram from the Markovian line or from the vertical line there to the blue line? What shifted it to the right? The change.

In this case, we have the change with a characteristic time of about 50 years, and this change produces this autocorrelogram, the blue line. So the autocorrelogram, which reflects the dependence, does not indicate memory but change; and long-term change is equivalent to long-term dependence.

And this is the power spectrum of the same process, but this was presented earlier, so I skip to the climacogram, which is a much simpler object than the power spectrum; it shows the standard deviation versus the scale. We aggregate the process onto several scales, and we take the standard deviation at each scale.

We have this shape here for our example. If it were white noise, the slope would be -0.5, but now it's -0.75. The slope plus 1 gives us what we call the Hurst coefficient, and remember that in that example, we have only two scales of fluctuation. If we had had several scales of fluctuation, we could obtain a simpler process than that, in which the standard deviation is a power function of the scale. Very simple. And also the autocorrelogram is a power function and the power spectrum is a power function.

The cascade of scales is very common in nature, and I have this visual example here, which is the hydraulic jump, a phenomenon familiar to all of us. In this case here, we have molecular motion or change, and we also have microturbulence, because we have a high Reynolds number, but then we have the hydraulic jump, and we have also macroturbulence.

Kolmorogov, who was the first to study this stochastic nature of turbulence, was able, in 1940, to establish the stochastic model that today we call fractional Brownian motion; and 10 years after, Hurst studied several geophysical time series, and he observed that although in random events groups of high or low values do occur, their tendency to occur in natural events is greater. He said, this is the main difference between natural and random events. And essentially, what these two said are not so different.

Here, I show perhaps the longest available instrumental record, which is the record of the Nile minimum water level, between 600 and 1500 AD. You see the instrument that was used in measurements -- much more solid a device than we use today. So if we make the climacogram of this process here, we see that the Hurst-Kolmogorov model is a very good representation of reality.

And if you compare this figure of the Nilometer data with random noise that I simulated here, you can see the difference on the 30-year scale. The fluctuations in the real-world process are much more intense and frequent than the stable curve of the 30-year average in the "roulette" process.

Okay. This example was already presented by Tim Cohn and Harry Lins. It is a time series of the lower tropospheric temperature. The red line represents the Hurst-Kolmogorov model. These are the theoretical standard deviations. And if we subtract the bias from the theoretical model, which is also calculated by the model, then we have a perfect agreement between model and reality.

As I said before, the model produces that there is some bias, in statistical metrics related to dispersion, and also, there is an increase of uncertainty from the classical statistics, which is very substantial. For a Hurst coefficient 0.9, we have six times larger variability of the sample mean in comparison to the classical statistics.

Okay. Just to show a few more examples, this here is the Atlantic multi-decadal oscillation index. You can see, again, the fitting of the Hurst model to the reality. And here in the Greenland temperature example we have scales of 20 years, of 500 years, the pink curve, and the red curve is 2,000 years averages. The same example, where we have 2,000 and 8,000 years averages. In those scales, you can see the change, and you can see the model, which fits fairly well.

So what are the implications in engineering design? In particular, in Athens? This is a repetition of the figure that I showed you before. Now we have this drought here, and we try to estimate its return period. Because we had a record of about 100 years, we expected that the return period of the most critical drought would be about this hundred years.

But look what happens. In one-year scale, it was indeed close to hundred years, but if we aggregate the scale to seven, eight, nine, ten years, then the return period would become a hundred thousand years, and we couldn't believe that we were so lucky hydrologists to experience a very infrequent and important drought event in our lifetime. So we said, something goes wrong here, so that's motivation to see other parts of this problem.

And of course, history is the key to the past. To the present. And to the future. So we started to compare with other cases, and of course, the Nilometer time series was a very nice example, and we located a very similar trend in about 700 AD. And then we saw, of course, that the evolution was not just as a projection of the same trend.

We applied the Hurst-Kolmogorov framework, and we found that the reality seems to comply with, to be consistent with, the model, and we did the calculations about the return period again, and then we found that the reality was consistent with the Hurst-Kolmogorov logic. And indeed we're not so lucky. So this Hurst-Kolmogorov framework implies a very different perspective, particularly in aggregate scales.

Here is the distribution function, the distribution quantile versus normal variate, for the one-year mean discharge in the river. If we move from the classical statistics to Hurst-Kolmogorov statistics, the quantile would be the same, but the uncertainty is increased.

The classical statistics produced this gray zone here. The Hurst-Kolmogorov produces this wider uncertainty. But if we aggregate at the 30-year scales, then there is a huge difference, because not only the range changes, but also the quantile is different.

So the uncertainty at the 30-year scale within Hurst-Kolmogorov happens to be exactly the same, at 200 percent, as the uncertainty in the annual scale of the classical model. So according to the classical model, climate -- actually, this is part of the definition -- climate is what we expect, weather is what we get, and according to the Hurst-Kolmogorov approach, weather is what we get immediately, and climate is what we get if you keep expecting for a long time.

Now, of course, we implemented this methodology, and of course, the first step is a stochastic generator, and we devised one for using the generalized Hurst-Kolmogorov process. Of course, it produces changes in long time scales. That's immediate.

And we implemented, then, the generated series, according to a methodology that we have developed, called parameterization-simulation-optimization. We have several papers about this methodology, as well this methodology here. Some of them are in WRR. I will not go into detail about how this methodology works. But I will present two alternatives.

Of course, the one obvious and very well-known alternative is to project the trend. But, as you see in the graphic, in this case, flow would disappear by 2050. Also, the uncertainty would decrease for the reasons I've explained. So the conclusion is that it is absurd to use simplistic methods as prediction.

The obvious second alternative is to use climate models. Of course, the climate models did not have any relation with reality, so we did kind of adaptations, also known as downscaling, to bring their outputs closer to reality, and this that you see here is after the adjustment, of course. Look what all these climate models predict: a stable future, very stable. We cannot promise a stable future. Rather, we estimated the uncertainty and the reliability on the climatic scale, based on the Hurst-Kolmogorov approach, and we found these red lines, and we plan and manage the system based on this uncertainty.

Okay. Of course, this Hurst-Kolmogorov framework applies also to floods, and I have quoted a phrase here from a recent paper by Blöschl and Montanari which says, "five of the six largest floods have occurred in the last two decades". This is a trick, of course, because the person who would say that, lived in 1900. That's the quote -- and that's the entire record of floods including the 20th century.

And change has also implications to the distribution tails and if you start with an exponential type of tail, you will end up with a power-type tail, and there is a simple mathematical proof about this.

Okay. Final remarks and conclusions. The advantages of the new Hurst-Kolmogorov approach. Perhaps some people think that these are disadvantages. Anyway, it is old-fashioned and not trendy, because it starts in 1940. It is stationary, and not non-stationary, and demonstrates how stationarity can coexist with change at all time scales.

It is linear, and not nonlinear; the deterministic dynamics need to be nonlinear to produce realistic trajectories, but stochastic models need not. It's simple, parsimonious, inexpensive; not complicated, inflationary, and expensive. And it's honest and not deceitful; does not hide uncertainty, does not pretend to predict the distant future.

So, conclusions. Change is nature's style. Change occurs at all time scales. Change is not non-stationarity. Hurst-Kolmogorov dynamics is the key to perceive multi-scale change and to model the implied uncertainty and risk. That's my opinion, of course.

In general, the Hurst-Kolmogorov approach can incorporate the deterministic descriptions if they are available and credible. In absence of credible predictions of the future, Hurst-Kolmogorov dynamics admits stationarity. So, long live stationarity.