

Negligent killing of scientific concepts: the stationarity case

Demetris Koutsoyiannis¹ and Alberto Montanari²

¹Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Greece (dk@itia.ntua.gr – <http://itia.ntua.gr/dk>)

²Department DICAM, University of Bologna

Abstract In the scientific vocabulary, the term “process” is used to denote change in time. Even a stationary process describes a system changing in time, rather than a static one which keeps a constant state all the time. However, this is often missed, which has led to misusing the term “nonstationarity” as a synonym of “change”. A simple rule to avoid such misuse is to answer the question: can the change be predicted in deterministic terms? Only if the answer is positive it is legitimate to invoke nonstationarity. In addition, we should have in mind that models are made to simulate the future rather than to describe the past; the past is rather characterized by observations (data). Usually future changes are not deterministically predictable and thus the models should, on the one hand, be stationary and, on the other hand, describe in stochastic terms the full variability, originating from all agents of change. Even if the past evolution of the process of interest contains changes explainable in deterministic terms (e.g. urbanization), again it is better to describe the future conditions in stationary terms, after “stationarizing” the past observations, i.e. adapting them to represent the future conditions.

Introduction

While a phrase like “there is a lot of heat in the kitchen” is perfectly understandable in everyday language, the same expression is unacceptable for a scientific text, or even in oral communications among scientists. One would expect from a scientist to say “there is a lot of thermal energy in the kitchen” or reformulate it as “the temperature in the kitchen is high”. Interestingly, despite such scientific concepts like *heat*, (*thermal*) *energy* and *temperature* are centuries old, misconceptions and misuses are still common, even among university students of thermal physics (Georgiou and Sharma, 2012) and may hinder learning and understanding. Perhaps the difficulties stem from the fact that thermal physics concepts are au fond statistical concepts, while our education systems favour deterministic approaches and give much less importance to probability, statistics and stochastic processes, areas that are collectively described by the term *stochastics*.

While stochastics offers a much more powerful approach to model physical systems (Koutsoyiannis 2010) including hydrological systems (Montanari and Koutsoyiannis 2012), it also requires a different type of understanding for the concepts it introduces. Lack of such understanding may result in misuse of the terms and the related concepts.

Take for instance the terms *stationarity* and *nonstationarity*. Interestingly, while the adjective *stationary* is contained in English dictionaries with the meaning *not moving* or *not changing*, the nouns *stationarity* and *nonstationarity* do not appear in common dictionaries like the Cambridge (dictionary.cambridge.org), the Dictionary.com (dictionary.reference.com/), the Merriam-Webster (www.merriam-webster.com/dictionary/) and the Oxford (www.oxforddictionaries.com/). The reason is that they are only scientific terms, not words used in everyday language. Thus, in contrast to *heat* and *temperature* which may be used colloquially in a loose manner, misuse of the *stationarity* and

nonstationarity can be more easily detected and hopefully avoided as there is no ambiguity owing to difference in everyday language and scientific meaning.

However, for a wide variety of reasons, including its stochastic origin, ideologico-political influences and a clustering effect, misuse of the terms *stationarity* and *nonstationarity* has been common. Illustration of the clustering effect in the term's misuse is offered by the paper by Milly et al. (2008) with the bold title "Stationarity is dead". A Google Scholar search reveals that up to now (March 2014) 1350 papers contain the phrase "Stationarity is dead". This can be interpreted as an indication that more or less those papers have adopted the misuse of the term in the original paper. Interestingly, only four papers query "Is stationarity dead?", not a single paper says "Stationarity is not dead" and only one paper says "Stationarity is alive". Other searches, again using Google Scholar, reveal that (as of March 2014) 660 papers speak about "Nonstationary world", 551 about "Nonstationary climate", 15 about "Nonstationary catchment(s)" and 14 700 papers about "Nonstationary data". It can be conjectured that most of them misuse the term *nonstationary*, as will be justified below.

In an attempt to combat such misuse, the World Meteorological Organization (Commission for Hydrology 2013) has issued a statement with clarifications about the terms. This note has exactly the same purpose.

Semantics and historical review

In this section we will try to reconstruct a possible history of the concept of stationarity based on some key references including historical ones. The term *stationary* is originally a classical mathematical term: given a function $x(t)$, a *stationary point* thereof is a point in which the derivative is zero, i.e. $x'(t) = 0$. In dynamical systems, the notion of a *stationary point* has a richer meaning. A dynamical system described by its state $x(t)$ is typically characterized by a differential equation of the form $x'(t) = g(x(t))$. On the other hand, it is common to express the dynamical system in terms of a transformation S_t of its initial state $x(0)$ (at time 0) to its current state $x(t)$ (at time t), that is, $x(t) = S_t(x(0))$ (Lasota and Mackey 1994). In this case, if x is a stationary point, i.e. a solution of $x'(t) = 0$, it also satisfies $x = S_t(x)$, that is, it remains invariant under the transformation S_t . In classical algebra a point x satisfying $x = S_t(x)$ is called a *fixed point* or an *invariant point* and is distinguished from the *stationary point*. However in differential equations books (e.g. Conrad, 2003) and dynamical systems books (e.g. Wiggins, 2003) all these terms (and other ones, e.g., "equilibrium point", "rest point") are used interchangeably.

As an example consider the logistic differential equation

$$x'(t) = g(x(t)) = a x(t) (1 - x(t)/b) \quad (1)$$

Solving the differential equation for $x(t) = S_t(x(0))$ with initial condition $x(0) \equiv x$ we find that the transformation $S_t(x)$ is

$$S_t(x) = \frac{b}{1 + e^{-at}(b/x - 1)} \quad (2)$$

It can be readily seen that there are two stationary points, $x = 0$ and $x = b$, which satisfy both $x' = g(x) = 0$ and $x = S_t(x)$.

Now, a more effective description of dynamical systems is achieved when, instead of studying trajectories of points (i.e. the $S_t(x)$ for varying t), we study the evolution in time t of probability densities $f(x)$ (Lasota and Mackey, 1994). In this case, the description through $S_t(x)$ is replaced by that based on $K^t f(x)$, where K^t is a transformation operator. Accordingly, a stationary density is defined to be a probability density function f which satisfies $f = K^t f$ thus remaining unchanged in the course of time. For a simple account about why evolution of probability densities provide a better description of

a system than trajectories of points and for more details about transformation operators, the reader is referred to Koutsoyiannis (2010).

In this context, Kolmogorov (1931) used the term *stationary* to describe a probability density function that is unchanged in time. In the same work, Kolmogorov introduced the term *stochastic process* although he cited Bachelier (1900) as having already used stochastic processes in continuous time; however, it appears that Bachelier did not use the term *stochastic process*. A more formal definition of a *stochastic process* and *stationarity* was given a few years later by Khinchin (Khinchine 1934; notice the different transliteration in the English and the German literature of author's name, originally appearing in the Cyrillic alphabet). A concise presentation thereof has been given by Kolmogorov (1938) as follows:

[...] a stationary stochastic process in the sense of Khinchin [...] is a set of random variables x_t depending on the parameter t , $-\infty < t < +\infty$, such that the distributions of the systems

$$(x_{t_1}, x_{t_2}, \dots, x_{t_n}) \text{ and } (x_{t_1+\tau}, x_{t_2+\tau}, \dots, x_{t_n+\tau}) \quad (3)$$

coincide for any n, t_1, t_2, \dots, t_n , and τ .

Kolmogorov (1947) has also defined *wide-sense stationarity* in terms of independence on time t of the expectation $E[x_{t+\tau} x_t]$.

These authors were Russian (then Soviet) and their papers were published mostly in German. The concepts were transplanted in the English literature, unfortunately with some loss of clarity. Perhaps the first who introduced them into the English literature was Doob (1934), from whom we quote the following definition:

A stochastic process is defined by Khinchine [(1934)] to be a one parameter set of chance variables: $x(t)$, $-\infty < t < \infty$. It is supposed that if t_1, \dots, t_n is any finite set of values of t , and $a_j < x < b_j, j = 1, \dots, n$ any set of intervals, the probability that

$$a_j < x(t_j) < b_j, j = 1, \dots, n \quad (4)$$

is defined. If the probability that [(4)] is true is independent of translations of the t -axis, the process is called stationary.

By comparing with the original paper (in German) or with Kolmogorov's definition in (3) it can be seen that Doob distorts somewhat Khinchin's definition. Furthermore, the famous book by Kendall and Stuart (1966, p. 404), while essentially keeping the Khinchin-Kolmogorov definition, speaks about a *stationary series* instead of a *stationary process*. It further states that *stationary time-series* are a particular case of the *theory of stochastic processes*. However, the definition of a time series given in the book does not support the latter statement, while a definition of a stochastic process is missing. Specifically the definition of a time series given by Kendall and Stuart (1966, p. 342) is the following:

Observations on a phenomenon which is moving through time generate an ordered set known as a time series. The values assumed by a variable at time t may or may not embody an element of random variation, but in the majority of cases with which we shall be concerned some such element is present, if only as an error of observation.

This seems to recognize a time series as a series of observations, which could be a series of realizations of the random variables that constitute the stochastic process (else known as a sample function of the stochastic process). Also, it could be a series of values not necessarily associated with a stochastic process. However, in the definition of the stationarity, the concept of a time series looks to be treated as identical to (or subcase of) that of a stochastic process.

Indeed, in the English literature the concept of a time series is ambiguous, sometimes denoting a realization of a stochastic process and other times denoting the stochastic process *per se* (with the

specification that the index set defining the stochastic process denotes time; e.g. Parzen 1957). Nevertheless, there are books in the English literature characterized by perfect clarity, of which we mention Papoulis (1991; first edition 1965). Such books follow the Khinchin-Kolmogorov definition of stochastic processes and stationarity.

Some preliminary conclusions that we can draw from this historical review are: (a) stationarity refers to stochastic processes; (b) stochastic processes are families of random variables usually (but not necessarily) indexed by time; and (c) random variables are variables associated with a probability distribution or density function. It follows that any attempt to conceptualize stationarity without reference to the notion of a stochastic process will be inconsistent with the theory.

The concept of nonstationarity

Inevitably, the negation of stationarity, i.e. *nonstationarity*, needs to be defined and conceived again within the notion of stochastic processes. The negation could be conceived with reference to Eqn. (3), meaning that in a nonstationary process the probability density $f(x_{t_1+\tau}, x_{t_2+\tau}, \dots, x_{t_n+\tau})$ for some (or all) τ should not equal $f(x_{t_1}, x_{t_2}, \dots, x_{t_n})$. In turn, this means that the mathematical expression of $f(x_{t_1+\tau}, x_{t_2+\tau}, \dots, x_{t_n+\tau})$ should explicitly contain the time shift τ , or else that it should be a deterministic function of τ .

Often, in looser terms, a stationary process is thought of as a process whose statistical properties do not change with time, while, by negation, a nonstationary process is thought of as a process whose statistical properties do change with time. The latter statement, if combined with a loose perception of what statistical properties are, and perhaps dismissal of “statistical” from “properties” has led to very widespread misconceptions. Two common examples of such misconceptions are (a) that random changes in some properties indicate nonstationarity, and (b) that it is not necessary to have deterministic functions of time in order to claim nonstationarity.

We can clarify these misconceptions, by considering Eqn. (3) with just one variable, x_{t_1} , also assuming $t_1 = 0$, so that we have to compare statistical properties of \underline{x}_τ , with those of \underline{x}_0 (notice that underlined symbols denote random variables and stochastic processes—the Dutch convention). For misconception (a), let us assume that \underline{x}_τ can be decomposed in two parts as follows:

$$\underline{x}_\tau = d_\tau + \underline{v}_\tau \quad (5)$$

where \underline{v}_τ is a stationary stochastic process. If the component d_τ is a constant ($d_\tau \equiv d$, where d is a number), then obviously \underline{x}_τ is also a stationary stochastic process. If d_τ is a deterministic function of time ($d_\tau \equiv d(\tau)$), then \underline{x}_τ will be a nonstationary stochastic process as its mean will be $E[\underline{x}_\tau] = d(\tau) + E[\underline{v}_\tau]$, i.e., a function of time. Now if d_τ changes in time, but not according to a deterministic function, then it should be modelled as a stochastic process itself ($d_\tau \equiv \underline{d}_\tau$). If the latter process is a stationary one (see further clarification about this in section “Stationarity under change”) then the process \underline{x}_τ is stationary (because $E[\underline{x}_\tau] = E[\underline{d}_\tau] + E[\underline{v}_\tau] = \text{constant}$)—even though many would regard it as nonstationary.

As an example to analyse misconception (b), let us consider as statistical properties of \underline{x}_τ its moments $E[\underline{x}_\tau^q]$ for varying $q = 0, 1, \dots$. Such moments are not random variables, and thus they can be either constants ($E[\underline{x}_\tau^q] \equiv m_q$) or deterministic functions of τ ($E[\underline{x}_\tau^q] \equiv m_q(\tau)$). If we assume that all moments $E[\underline{x}_\tau^q]$ for all q are not functions of τ then, by virtue of the moment theorem (Papoulis 1991, p. 116) the probability density function $f(x_\tau)$ is uniquely determined by the moments and will not be a function of τ *per se*. Therefore, it is necessary that at least one moment $E[\underline{x}_\tau^q]$ for a certain q be a deterministic function $m_q(\tau)$ in order to have a time dependent $f(x_\tau)$.

The above example is for first-order stationarity, because we used only one \underline{x}_τ . A process can be first-order stationary yet being nonstationary, if joint statistical properties of variables corresponding

to different times are deterministic functions of time (e.g., if the joint moment $E[x_\tau x_{\tau+t}]$ is a function of both t and τ , the process is nonstationary). However, the above result can be generalized for two or more variables as the moment theorem holds for more than one variable (e.g., Papoulis 1991, p. 160). Hence, if a stochastic process is nonstationary, then at least one of its moments, marginal or joint, should be a deterministic function of time.

In conclusion, as the meaning of *change* in everyday language is quite general and certainly includes random change, statements like “a nonstationary process is a process whose statistical properties *change* with time” may be misunderstood. In order to be strictly valid, such statements should be accompanied by the clarification that *change* here is meant as a deterministic function of time.

Philosophical reflections

While misconceptions about stationarity and nonstationarity can be relieved because the concept is only 80 years old, the discourse in the Introduction (in particular, the typical phrases that involve *nonstationarity*) reveals that there may be a more fundamental misconception about a notion that is 2.5 thousand years old. Specifically, this is related to confusing the material world with the world of models, which usually are abstract (mathematical) entities. The distinction between the two and the relevant discussion are at least as old as Plato’s metaphysical theory.

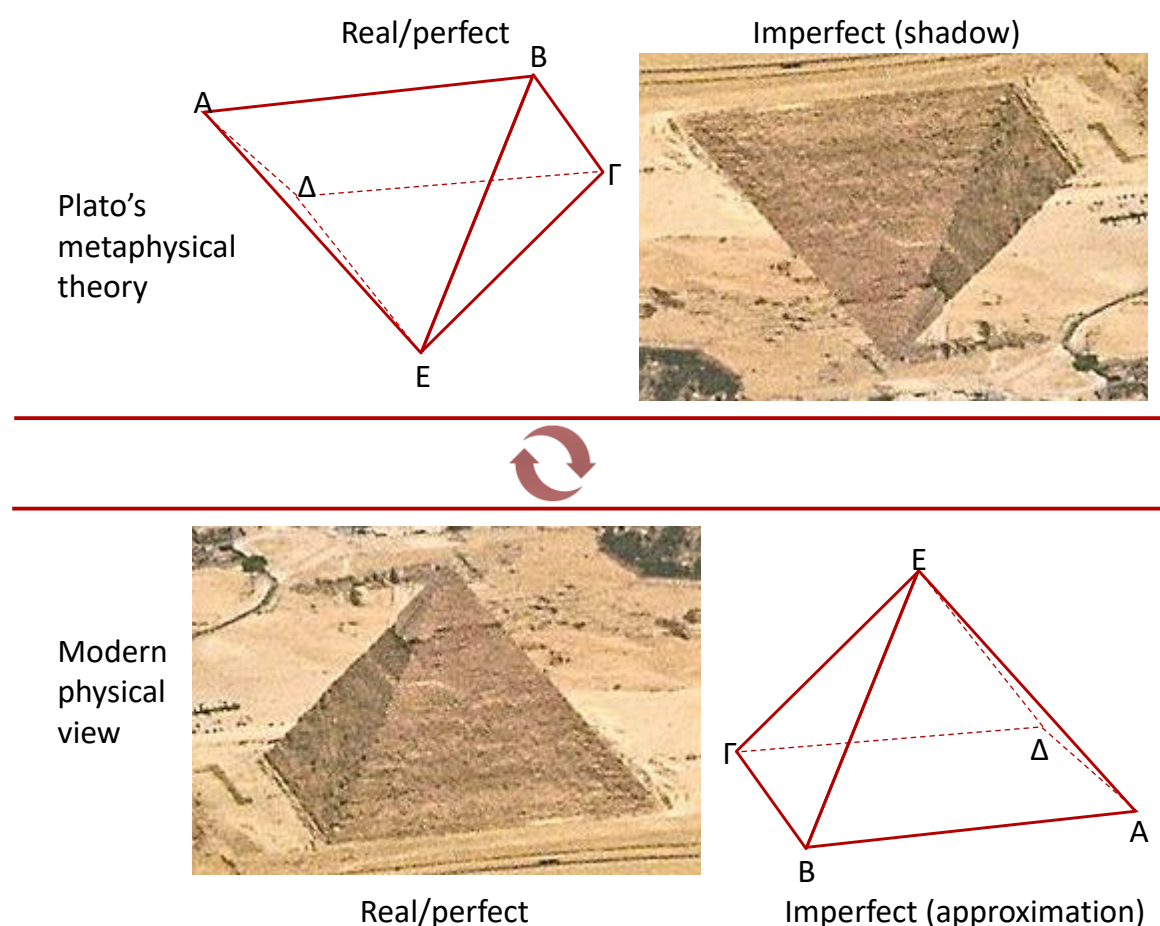


Figure 1 A schematic to illustrate Plato’s metaphysical theory (upper row) in comparison to a view more compatible to modern science (lower row). The left column depicts the real world, which in Plato’s theory is the world of the archetypes, while the material world is composed by imperfect “shadows” of the archetypes. By rotating the schematic corresponding to Plato’s theory by 180° we identify the material world with the real world and we replace the ideal archetypes with imperfect models of reality. In both cases, however, the intellectual structures are distinguished from the material ones (image of pyramid from en.wikipedia.org/wiki/File:Giza-pyramids.JPG).

According to Plato's theory, the real world is a world of ideal or perfect forms (*αρχέτυπα*, *archetypes*). It is unchanging and unseen and it can only be perceived by reason (*νοούμενα*, *nooumena*). The physical world is an imperfect image of the world of archetypes. Physical objects and events are “shadows” of their ideal forms, are subject to change and can be perceived by senses (*φαινόμενα*, *phenomena*).

By turning upside-down Plato's theory (or making a 180° turn, as illustrated in Fig. 1) we obtain a view that is more consistent with modern science. According to the latter, the physical world is the real world. It is perfect and it is perpetually changing. Abstract representations or *models* of the real world are imperfect but can be useful to describe the real world. It should be noted though that, while Plato's archetypes and modern models have in common the fact that are abstract concepts perceived by reason, the two notions are not identical.

Whether one accepts the original or the upside-down version of Plato's theory, it is important that both make the distinctions: physical world \neq models (or forms) and phenomena \neq noumena. Also, both recognize that in the physical world change is the rule, a fact that had been earlier and aptly expressed by Heraclitus as “Πάντα ῥεῖ” (*Panta rhei*, Everything flows). The latter aphorism has now become the emblem of the IAHS initiative for the current decade (Montanari et al. 2013).

With respect to the discussion of the earlier section, it is quite easy to understand that probability and all concepts based on it are abstract concepts, and not properties of the physical world. Attempts to define probability in terms of experience from the physical world resulted in logical problems (e.g. circular logic), which were resolved only when Kolmogorov (1933) introduced the axiomatic foundation of probability. The abstract character of probability is more evident in the Bayesian interpretation thereof. Likewise, the stochastic processes, as well as the random variables they consist of and their probability density functions, are models and not physical objects. This means that stationarity and nonstationarity are properties belonging to the world of models. In contrast, the objects and processes of the real world are neither stationary nor nonstationary; they are just perpetually changing.

Stationarity under change

Stationarity does not contradict change. Rather it offers a powerful way to model change. Without change, we would not need the concept of a stationary process, not even that of a process. Note that a process is defined in common English dictionaries, such as those mentioned above, as a series of changes—or actions. In addition, in the scientific literature the term *process* has been introduced as synonymous to change, as evident in Kolmogorov's (1931) pioneering paper, which starts as “A *physical process* [is] a change of a certain physical system”.

It is very common in science to try to identify invariant properties within change (Koutsoyiannis 2011). For example, in the absence of an external force, the position of a body in motion changes in time but the velocity is unchanged (Newton's first law). If a constant force is present, then the velocity changes but the acceleration is constant (Newton's second law). If the force changes, e.g. the gravitational force with changing distance in planetary motion, the acceleration is no longer constant, but other invariant properties emerge, e.g. the angular momentum (Newton's law of gravitation; see also Koutsoyiannis 2011).

Also, in motion of fluids we speak about steady flow. This is not a contradiction. Obviously, there is change because of the flow (*change* and *flow* are tightly connected), but in a steady flow the velocity does not change in time.

Likewise, the theory of dynamical systems is a theory describing change. Amazingly, most of dynamical systems used to model natural processes are autonomous systems, that is, they are expressed by autonomous differential equations, e.g., $x'(t) = g(x(t))$, which do not explicitly depend on

time. In the more general case of non-autonomous systems, i.e. those expressed as $x'(t) = g(x(t); t)$, the dynamics is not the same on the intervals $[0, t']$ and $[t, t + t']$. However, this is rarely the case because the laws of nature which hold now are identical to those holding for any time in the past or future. Even if change in the laws happens to be the case (e.g. in a macroscopic description of a complex system) again the non-autonomous system is converted into an autonomous one. This is a rather easy task as it only needs the definition of new dependent variables (Mackey 1992).

Now, coming to the notion of a stochastic process, we note that it was invented to describe the irregular changes of complex natural systems, which are impossible to model deterministically in full detail or predict their future evolution in detail and with precision. Here, the great scientific achievement is the invention of macroscopic descriptions instead of modelling the details. This is essentially done using stochastics. Here lies the essence and usefulness of the stationarity concept, which seeks invariant properties in complex systems (Koutsoyiannis, 2011). As in the case of converting non-autonomous to autonomous systems, again even if a process is nonstationary, it should be converted into a stationary one, which can be effectively studied and used in predictions.

As the stochastic model is a mathematical construction, the use of a stationary or a nonstationary model is a modelling option rather than a property of the observational data of the real world. Of course any model, whether deterministic or stochastic, stationary or nonstationary, should be consistent with the data. But the data alone do not suffice to make a correct modelling choice.

This is illustrated in Figure 2, which depicts 1000 terms of a synthetic time series; the present time, as designated in the figure, divides the time axis into two parts, the past (with 150 observations) and the future. Considering the 150 observations of the past, one sees two step changes: a large rise at time 70 and a smaller drop at time 120. According to common statistical practice, one would describe such changes as nonstationarities and would construct a nonstationary model with different statistical properties for each of the three periods designated by the two step changes. But such a description is totally useless if prediction of the future is of interest and, at the same time, it contradicts the fact that the stochastic process which generated the time series is in fact a stationary process, which is described in detail in Koutsoyiannis (2011).

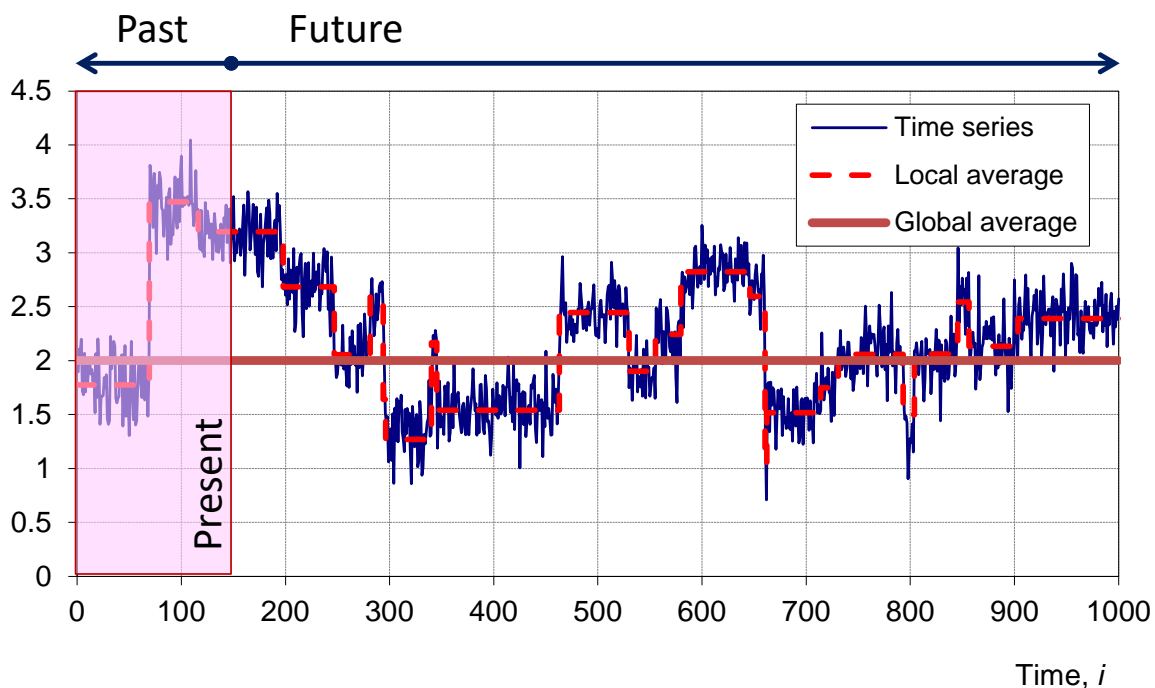


Figure 2 Schematic for a prediction problem based on a time series.

It is recalled that nonstationarity is determined based on ensemble means or, more generally, ensemble statistical properties of a stochastic process. The ensemble mean is defined as

$$E[\underline{x}] = \int_{-\infty}^{\infty} xf(x)dx \quad (6)$$

while the temporal mean is

$$\bar{\underline{x}}_{t,t'}(t) := \frac{1}{t'} \int_t^{t+t'} \underline{x}(t)dt \quad (7)$$

It is seen that the temporal mean is a stochastic process *per se*, that is, a random function of time. The ensemble mean is not a random quantity, but in a stationary process it is a single number. As explained in section “The concept of nonstationarity”, if the process is nonstationary, $E[\underline{x}]$ can vary in time, being a deterministic function of time (i.e., precisely known and perfectly predictable). But, as the graphical example shows, this cannot be inferred from data only, because the data can give us only the temporal mean (actually a realization thereof) and not the ensemble mean. Rather, the data can mislead us to adopt a nonstationary description, while the underlining process is stationary. To establish a deterministic function of time, as required in order to claim nonstationarity, we need at least both of the following conditions to hold: (a) deductive reasoning in order to establish the deterministic function of time; (b) validation of the deterministic function by data which were not used in the model construction.

It seems the frequent invocation of nonstationarity, as well as deterministic trends, shifts and cycles, is a remnant of deterministic thinking, while using stochastic descriptions. The actual stochastic concept that should replace all these, within a stationary stochastic setting is *dependence*, possibly a long-range one. The handling of dependence, and eventually its utilization in prediction, is made through conditioning on known information, e.g. observations of the past. For example, assuming that x_0, x_1, \dots, x_n , are observations at n past times t_0, t_1, \dots, t_{n-1} , and at the present time t_n , the conditional average of the process $\underline{x}(t)$ for any future time $t > t_n$ will be a function of time, $h(t)$:

$$E[\underline{x}(t)|x_0, x_1, \dots, x_n] = h(t) \neq E[\underline{x}(t)] \quad (8)$$

Likewise, the conditional variance is also a (non-decreasing) function of time and every statistical characteristic, conditional on observations, can be a function of time. All these functions of time are deterministic functions, derived by deductive reasoning (using stochastic processes algebra). Hence, conditioning results in statistical characteristics that vary in time, even though overall the model is stationary.

Figure 2 further illustrates the effect of conditioning on the past, within stationarity, in order to predict the future. While the past values can be known from observation, the evolution of the process in future times is unknown. Ignorance of the future does not concern only the precise value at each time. The local averages, shown as dashed lines in the figure are also unknown (had they been known, the model would be nonstationary). Let us suppose we are required to make a prediction for the future. If the prediction horizon is long, then we will use the global (i.e. the true or ensemble) average and the global variance for our prediction. However, if the prediction horizon is short, then we will use the local average at the present time and a reduced variance. More generally, by utilizing the dependence structure of the model, which can be inferred from the observations, assuming that their length is sufficient, we are able to make predictions conditioning on the observations, i.e. formally evaluating Eqn. (8) for the assumed stationary stochastic model and for the required prediction horizon t .

Justified use of nonstationary models

As clarified above, in a nonstationary process, to describe change we use both stochastic variations and deterministic functions. By definition, a deterministic function should be constructed by deduction

(the Aristotelian apodeixis), not by induction (the Aristotelian epagoge) which makes direct use of the data (see also next section). Because it explains in deterministic terms part of the variability, a nonstationary description is associated with reduced uncertainty. Hence unjustified or inappropriate claim of nonstationarity results in underestimation of variability, uncertainty and risk.

However, there are cases where use of a nonstationary description is justified. Sometimes we undertake modelling of the past, in order to explain observed behaviours. Changes in hydrological behaviours happen all the time, as a result of changes in quantifiable characteristics of catchments and conceptual parameters of models. If we know the evolution of these characteristics and parameters (e.g. in addition to hydrological observations, we have information about how the percent of urban area changed in time), then we can build a nonstationary model. The specific information has reduced uncertainty thus enabling a nonstationary description. In contrast, if we see a changing behaviour but we do not have this quantitative information, then we may treat the catchment characteristics and parameters as random variables. In this case we build stationary models entailing larger uncertainty.

It is important to distinguish explanation of observed phenomena in the past from modelling that is made for the future. Except for trivial cases, the future is not easy to predict in deterministic terms. If changes in the recent past are foreseen to endure in the future (e.g. urbanization, hydraulic infrastructures), then the model of the future should be adapted to the most recent past. This may imply a stationary model of the future that is different from that of the distant past (prior to the change). It may also require “stationarizing” of the past observations, i.e. adapting them to represent the future conditions (e.g. the flow data prior to the construction of the dam could be adapted to determine what the flow would be if the dam existed).

In the case of planned and controllable future changes (e.g. catchment modification by hydraulic infrastructures, water abstractions), which indeed allow prediction in deterministic terms or at least condition our predictions, nonstationary models are justified.

Stationarity, ergodicity and inductive inference

It is important to note that stationarity is also related to ergodicity, which in turn is a prerequisite to make inference from data, that is, induction. In dynamical systems, by definition (e.g. Mackey, 1992, p. 48), ergodicity is the property of a system whose all invariant sets under the dynamic transformation are trivial (have zero probability). In other words, in an ergodic transformation starting from any point, the trajectory of the system state will visit all other points, without being trapped to a certain subset. The ergodic theorem (Birkhoff 1931; see also Mackey 1992 p. 54) allows redefining ergodicity within the stochastic processes domain (Papoulis 1991 p. 427; Koutsoyiannis 2010) in the following manner: A stochastic process \underline{x}_t is ergodic if the time average of any (integrable) function $g(\underline{x}_t)$, as time tends to infinity, equals the true (ensemble) expectation $E[g(\underline{x}_t)]$. Thus, with reference to Eqn. (7), in an ergodic system the time average, as $t' \rightarrow \infty$, will tend to the ensemble average given in (6). This allows the estimation (i.e. approximate calculation) of the true but unknown property $E[\underline{x}_t]$ from the $\bar{\underline{x}}_{t'}$, that is from the available data. Without ergodicity inference from data would not be possible.

Now, if the system that is modelled in a stochastic framework has deterministic dynamics (meaning that a system input will give a single system response, as happens for example in most hydrological models) then a theorem applies (Mackey 1992, theorem 4.5 p. 52), according to which a dynamical system with dynamics $S_t(x)$ has a stationary probability density *if and only if* it is ergodic. Therefore, a stationary system is also ergodic and vice versa, and a nonstationary system is also non-ergodic and vice versa. Here we note that even if a system has deterministic dynamics, again it is legitimate to use a stochastic description, replacing the study of the evolution of system states $S_t(x)$ with the evolution of probability densities of states $f(x)$ as already mentioned in the section “Semantics and historical review”; one reason to prefer the stochastic description over the pure deterministic

description is that the former includes quantification of uncertainty, whereas the deterministic dynamics does not eliminate uncertainty (Koutsoyiannis 2010). Furthermore, we clarify that the deterministic description through the transformation $S_t(x)$ is fully compatible with a stochastic description that is stationary and ergodic, according to theorem stated above: while the system state is changing in time t according to the transformation $S_t(x)$, its statistical properties (and the probability density $f(x)$) can be constant in time.

If the system dynamics is stochastic (a single input could result in multiple outputs), then ergodicity and stationarity do not necessarily coincide. However, recalling that a stochastic process is a model and not part of the real world, we can always conveniently device a stochastic process that is ergodic, provided that we have excluded nonstationarity. To clarify this idea we give the following example. A well-known stochastic process \underline{x}_t which is stationary and non-ergodic is (Papoulis 1991, p. 434)

$$\underline{x}_t := \underline{a} \underline{u}_t \quad (9)$$

where \underline{a} is a random variable and \underline{u}_t is a stationary and ergodic stochastic process. A single realization of \underline{x}_t does not enable the estimation of its true statistics because all its history corresponds to a single realization a of the random variable \underline{a} . However, the single realization allows the estimation of the statistics of the process $\underline{x}'_t := a \underline{u}_t \equiv \underline{x}_t | \underline{a} = a$, which is stationary and ergodic. In most practical problems corresponding to physical systems, which actually do not allow different realizations but rather have a unique evolution, the properties of the ergodic process \underline{x}'_t rather than those of the non-ergodic \underline{x}_t are actually of interest.

In conclusion, from a practical point of view ergodicity can always be assumed when there is stationarity, while this assumption is fully justified by the theory if the system dynamics is deterministic. Conversely, if nonstationarity is assumed, then ergodicity cannot hold, which forbids inference from data. This contradicts the basic premise in geosciences, where data are the only reliable information in building models and making inference and prediction.

Conclusions

Change is Nature's style and occurs at all times and all time scales (i.e. Πάντα ῥεῖ). The frequent use of the term *nonstationarity* lately indicates that it is confused with *change*. However, change is not synonymous to nonstationarity: While change is a general notion applicable everywhere, including to the real (material) world, stationarity and nonstationarity apply only to models, not to the real world, and are defined within stochastics. The confusion about these terms extends also to the world of models, as several properties related to stochastic dependence of processes are often interpreted as nonstationarity. Nonstationary descriptions are justified only if the future can be predicted in deterministic terms and are associated with reduction of uncertainty, while misuse of nonstationarity results in underestimation of variability, uncertainty and risk. In absence of credible predictions of the future, admitting stationarity (and larger uncertainty) provides a more consistent and more effective modelling option.

Acknowledgment The present work was developed within the framework of the Panta Rhei Research Initiative of the International Association of Hydrological Sciences (IAHS). We thank two anonymous reviewers, the eponymous reviewer Tim Cohn and the Guest Editor Guillaume Thirel for their comments which resulted in expansion and improvement of the paper.

References

- Bachelier, M. L., 1900. Théorie de la spéculation, *Ann. Ecole Norm. Super.* 17, 21-86.
- Birkhoff, G. D., 1931. Proof of the ergodic theorem, *Proc. Nat. Acad. Sci.*, 17, 656–660.
- Commission for Hydrology, 2013, *CHy Statement on the Terms Stationarity and Nonstationarity*, World Meteorological Organization, http://www.wmo.int/pages/prog/hwrp/publications/statements/Stationarity_CHy_Statement.pdf.
- Conrad, B. P., 2003, *Differential Equations, A Systems Approach*, Prentice Hall, Upper Saddle River, NJ, USA, 462 pp.
- Doob, J.L., 1934. Stochastic processes and statistics. *Proceedings of the National Academy of Sciences of the United States of America*, 20(6), p.376.
- Georgiou, H., and Sharma, M. D. 2012, University students' understanding of thermal physics in everyday contexts, *International Journal of Science and Mathematics Education*, 10, 1119-1142.
- Kendall, M. G. & Stuart, A. (1966) *The Advanced Theory of Statistics, vol. 3: Design and Analysis, and Time-series*. 552 pp., Griffin, London, UK.
- Khintchine, A., 1934. Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen*, 109 (1), 604–615.
- Kolmogorov, A. N., 1931. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung, *Math. Ann.* 104, 415-458. (English translation: On analytical methods in probability theory, In: Kolmogorov, A.N., 1992. *Selected Works of A. N. Kolmogorov - Volume 2, Probability Theory and Mathematical Statistics* A. N. Shiriyayev, ed., Kluwer, Dordrecht, The Netherlands, pp. 62-108).
- Kolmogorov, A. N., 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Ergebnisse der Mathematik und Ihrer Grenzgebiete), Berlin (2nd English Edition: *Foundations of the Theory of Probability*, 84 pp. Chelsea Publishing Company, New York, 1956).
- Kolmogorov, A.N., 1938, A simplified proof of the Birkhoff-Khinchin ergodic theorem, *Uspekhi Mat. Nauk* 5, 52-56. (English edition: Kolmogorov, A.N., 1991, *Selected Works of A. N. Kolmogorov - Volume 1, Mathematics and Mechanics*, Tikhomirov, V. M. ed., Kluwer, Dordrecht, The Netherlands, pp. 271-276).
- Kolmogorov, A.N., 1947, Statistical theory of oscillations with continuous spectrum, *Collected papers on the 30th anniversary of the Great October Socialist Revolution*, Vol. 1, Akad. Nauk SSSR, Moscow-Leningrad, pp. 242-252. (English edition: Kolmogorov, A.N., 1992. *Selected Works of A. N. Kolmogorov - Volume 2, Probability Theory and Mathematical Statistics* A. N. Shiriyayev ed., Kluwer, Dordrecht, The Netherlands, pp. 321-330).
- Koutsoyiannis, D., 2010, A random walk on water, *Hydrology and Earth System Sciences*, 14, 585–601.
- Koutsoyiannis, D., 2011, Hurst-Kolmogorov dynamics and uncertainty, *Journal of the American Water Resources Association*, 47 (3), 481–495.
- Lasota, A., and Mackey, M. C.: *Chaos, Fractals and Noise*, Springer-Verlag, 1994.
- Mackey, M. C. 1992, *Time's Arrow: The Origins of Thermodynamic Behavior*, Dover, Mineola, NY, USA, 175 pp.
- Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier and R. J. Stouffer, 2008, Stationarity is dead: whither water management?, *Science*, 319, 573-574.
- Montanari, A., and D. Koutsoyiannis, 2012, A blueprint for process-based modeling of uncertain hydrological systems, *Water Resources Research*, 48, W09555, doi:10.1029/2011WR011412.
- Montanari, A., et al. 2013, “Panta Rhei – Everything Flows”, *Change in Hydrology and Society – The IAHS Scientific Decade 2013-2022*, *Hydrological Sciences Journal*, 58 (6), 1256–1275.
- Papoulis, A., 1991, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, New York, USA.
- Parzen, E. 1957. On consistent estimates of the spectrum of a stationary time series, *The Annals of Mathematical Statistics* 28 (2), 329-348.
- Wiggins, S., 2003. *Introduction to applied nonlinear dynamical systems and chaos*, Ed. 2. Springer-Verlag, New York, USA, 843 pp.