**Hydrology and
Earth System
Sciences**

HESS Opinions
"A random walk on water"

D. Koutsoyiannis

Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Greece

Abstract. According to the traditional notion of randomness and uncertainty, natural phenomena are separated into two mutually exclusive components, random (or stochastic) and deterministic. Within this dichotomous logic, the deterministic part supposedly represents cause-effect relationships and, thus, is physics and science (the "good"), whereas randomness has little relationship with science and no relationship with understanding (the "evil"). Here I argue that such views should be reconsidered by admitting that uncertainty is an intrinsic property of nature, that causality implies dependence of natural processes in time, thus suggesting predictability, but even the tiniest uncertainty (e.g. in initial conditions) may result in unpredictability after a certain time horizon. On these premises it is possible to shape a consistent stochastic representation of natural processes, in which predictability (suggested by deterministic laws) and unpredictability (randomness) coexist and are not separable or additive components. Deciding which of the two dominates is simply a matter of specifying the time horizon and scale of the prediction. Long horizons of prediction are inevitably associated with high uncertainty, whose quantification relies on the long-term stochastic properties of the processes.

Αἰὼν παῖς ἐστι παίζων πεσσεύων. Παιδὸς ἡ βασιληίη.
(*Time is a child playing, throwing dice. The ruling power is a child's*; Heraclitus; ca. 540–480 BC; Fragment 52)

I am convinced that He does not throw dice. (Albert Einstein, in a letter to Max Born in 1926)

1 What is randomness?

In his foundation of the modern axiomatic theory of probability, A. N. Kolmogorov (1933) avoided defining randomness. He used the notions of random events and random variables in a mathematical sense but without explaining what randomness is. Later, in about 1965, A. N. Kolmogorov and G. J. Chaitin independently proposed a definition of randomness based on complexity or absence of regularities or patterns (which could be reproduced by an algorithm). Specifically, a series of numbers is random if the smallest algorithm capable of specifying it to a computer has about the same number of bits of information as the series itself (Chaitin, 1975; Kolmogorov, 1963, 1965; Kolmogorov and Uspenskii, 1987; from Shiryaev, 1989). Interestingly, Chaitin proved that, although randomness can be precisely defined in this manner and can even be measured, there cannot be a proof that a given real number (regarded as a series of its digits) is random.

The move from this mathematical abstraction of a real number to the realm of real physical phenomena is not straightforward. Here, commonly, randomness is contrasted to determinism. The movement of planets is a typical example of a deterministic phenomenon, whereas that of dice is thought to be random. This reflects a dichotomous logic, according to which there exist two mutually exclusive types of events or processes – deterministic and random (or stochastic). Such dichotomy is perceived either on ontological or on epistemological grounds. In the former perception the natural events are thought to belong, in their essence, to these two different types, whereas in the latter it is regarded convenient to separate them into these types, where processes that we do not understand or explain are considered random. When a

classification of a specific process into one of these two types fails – and it usually does, except in a few cases such as the above examples of planets and dice – then a separation of the process into two different, usually additive, parts is typically devised. This perception has been dominant in geosciences, including hydrology. This thinking proceeds so as to form a reductionist hierarchy. Thus, each of the parts may be further subdivided into subparts (e.g. deterministic subparts such as periodic and aperiodic or trends). This dichotomous logic is typically combined with a manichean perception, in which the deterministic part supposedly represents cause-effect relationships and reason and thus is physics and science (the "good"), whereas randomness has little relationship with science and no relationship with understanding (the "evil"). The random part is also characterized as "noise", in contrast to the deterministic "signal". "Noise" is a contaminant that causes uncertainty, a kind of illness that should be remedied or eliminated.

Probability theory and statistics, which traditionally provided the tools for dealing with randomness and uncertainty, have been regarded by some as the "necessary evil", but not as an essential part of physical sciences. This view has also affected hydrology and geophysics, particularly in the last couple of decades. Some tried to banish probability from hydrology, replacing it with deterministic sensitivity analysis and fuzzy-logic representations. Others attempted to demonstrate that irregular fluctuations observed in natural processes are au fond manifestations of underlying deterministic dynamics with low dimensionality, thus rendering probabilistic descriptions unnecessary. Some of the above views and recent developments are simply flawed because they make erroneous use of probability and statistics, which, remarkably, provide the tools for such analyses (Koutsoyiannis, 2006).

The entire logic of contrasting determinism with randomness is just a false dichotomy. To see this, it suffices to recall that P.-S. Laplace, perhaps the most famous proponent of determinism in the history of philosophy of science (cf. Laplace's demon), is, at the same time, one of the founders of probability theory. According to Laplace (1812), "probability theory is, au fond, nothing but common sense reduced to calculus[1]". This harmonizes with J. C. Maxwell's view that "the true logic for this world is the calculus of Probabilities" (Maxwell, 1850). Recently, the same view was epigrammatically formulated in the title of E. T. Jaynes's (2003) book "Probability Theory: The Logic of Science".

Here I argue that the dominant dichotomous logic reflects a naïve and inconsistent view of randomness. It cannot help us see the unity of Nature. Are the movement of planets and that of dice qualitatively different natural phenomena? Do they not obey the same physical laws? Abandoning this logic and seeking a more consistent view, I propose to identify randomness with unpredictability. Randomness exists in

processes that we may understand, we may explain, but we cannot predict[2]. In other words, randomness and determinism (which, in turn, could be identified with predictability) coexist in the same process, but are not separable or additive components. It is a matter of specifying the time horizon and scale of prediction to decide which of the two dominates. This view, which will be illustrated in the sequel, is consistent with Kolmogorov's and Chaitin's views of mathematical randomness, as well as with K. Popper's (1992) indeterministic world view.

In identifying randomness with unpredictability, the latter is meant in deterministic terms: any specific prediction algorithm is not able to accurately calculate the future state of a system and thus there is uncertainty. If, in our attempt to predict the future, we are not able to know the precise state, we could lower our target and try to know how high or low the uncertainty is. Quantification of uncertainty is a useful target and a feasible one. Naturally, it is achieved by means of probability. The classical definition of probability, i.e. the ratio of the favourable outcomes of a specified event to the number of possible outcomes, has several logical problems and does not help to study natural processes. Rather, we should follow the Kolmogorov (1933) system, in which probability is a *normalized measure*, i.e. a function that maps sets (areas where the unknown quantities lie) to real numbers. An event A is just a set (of elements called elementary events) to which a probability, i.e. a number $P(A)$ in the interval $[0, 1]$, is assigned. The notion of a *random variable*, i.e. a single-valued function $\underline{x}$ of the set of all elementary events (so that to each elementary event ξ it maps a real number $x(\xi)$), is central in this system, and is associated with a *probability distribution function* ($F(x):=P\{\underline{x}\le x\}$)[3] and a *probability density function* ($f(x):=\mathrm{d}F(x)/\mathrm{d}x$). The notion of a random variable allows *probabilization of uncertainty*, typical in Bayesian statistics (not to be confused with the lately abused term of "Bayesian beliefs").

2   Emergence of randomness from determinism

To illustrate that randomness coexists with determinism and that the two do not imply different types of mechanisms, or different parts or components in the time evolution, we will study a toy model of a caricature hydrological system. The system, shown in Fig. 1, and its toy model are designed

---

[1] Original in French: "la théorie des probabilités n'est, au fond, que le bon sens réduit au calcul".

[2] According to Niels Bohr, "Prediction is difficult, especially of the future".

[3] Some texts distinguish random variables, which are functions, from their values, which are numbers, by denoting them with upper case and lower case letters, respectively. Since this convention has several problems (e.g. the Latin x and the Greek χ, if put in upper case, are the same symbol X), other texts do not distinguish the two at all, thus creating other type of ambiguity. Here we follow another convention, in which random variables are underscored and their values are not.
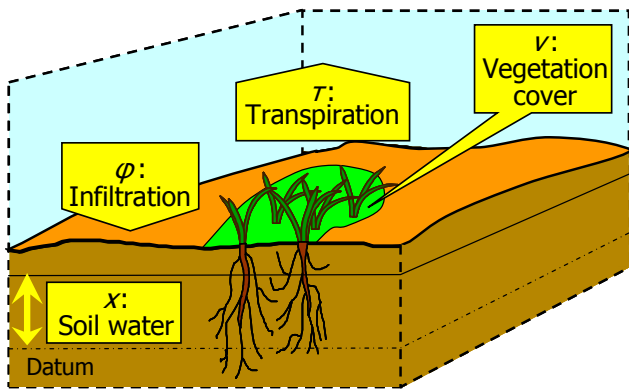
Fig. 1. A caricature hydrological system for which the toy model is constructed.

intentionally to be simple. A large piece of land is considered, on which water infiltrates and is stored in the soil (without distinction from groundwater), from where it can transpire though vegetation. Except infiltration, transpiration and water storage in the soil, no other hydrological processes are considered. To simplify the system, no change is imposed to its external "forcings". That is, the rates of infiltration ϕ' and potential transpiration τ'_p are assumed constant in time. The toy model is constructed assuming discrete time, denoted as $i = 1, 2, ...$, and that in each time unit Δt (say, "year"), the input is $\phi := \phi' \Delta t = 250$ mm and the potential output $\tau_p := \tau'_p \Delta t = 1000$ mm. The internal state variables, which are allowed to vary in time, are two, thus shaping a two-dimensional (2-D) dynamical system: the fraction of the land that is covered by vegetation, v_i ($0 \leq v_i \leq 1$) and the soil water storage x_i. The latter is measured above a certain datum, so that it can take positive values up to some upper bound α (assuming that water above α spills as runoff) or negative values without a bound (i.e. $-\infty \leq x_i \leq \alpha$). The constant α is assumed to be 750 mm. If the vegetation at time i is v_i, the actual output through transpiration will be $\tau_i = v_i \tau_p$. Thus, the water balance equation is

$$x_i = \min(x_{i-1} + \phi - v_i \tau_p, \alpha) \tag{1}$$

We can observe that, if at some time i, $v_i = \phi / \tau_p = 250/1000 = 0.25$, then the water balance results in $x_i = x_{i-1} + \phi - v_i \tau_p = x_{i-1}$. Assuming that the system dynamics is fully deterministic, continuity demands that there should be some specific value of x_{i-1} for which $v_i = v_{i-1}$. Without loss of generality, we set this value $x=0$; that is, we define the datum in such a way that the vegetation remains unchanged if water is stored up to the datum. Thus, the state ($v=0.25$, $x=0$) represents an equilibrium state: if at some time the system happens to be at this state, it will remain there for ever. In other words, once the system reaches its equilibrium state, it becomes a "dead" system, exhibiting no change.

Evidently, it is more interesting to study our system when it is "alive", that is, out of the equilibrium. To this aim, as the system is 2-D, we need one equation additional to Eq. (1) to model it, which we seek in conceptualizing the dynamics of vegetation. As $x=0$ represents the state where the vegetation does not change, we may assume that soil water in excess, $x > 0$, will result in increase of vegetation and soil water in deficit, $x < 0$, will result in decrease of the vegetation cover. The graph in Fig. 2 was constructed heuristically, according to this logic, and is described by the following equation:

$$v_i = \frac{\max(1 + (x_{i-1}/\beta)^3, 1) v_{i-1}}{\max(1 - (x_{i-1}/\beta)^3, 1) + (x_{i-1}/\beta)^3 v_{i-1}} \tag{2}$$

where $\beta = 100$ mm is a standardizing constant to make the equation dimensionally consistent.

Summarizing, we have a 2-D toy model in discrete time i whose state is described by the state variables $(x_i, v_i) =: \boldsymbol{x}_i$ (with $\boldsymbol{x}_i$ in bold denoting a vector) and whose dynamics is represented by Eqs. (1) (water balance) and (2) (vegetation cover dynamics). To avoid approximation errors, the dynamics is expressed in discrete time by construction of the model (rather than derived as an approximation of continuous time differential equations) and thus is fully accurate as far as the toy model is concerned. The system parameters are four and are assumed to be known precisely: $\phi = 250$ mm, $\tau_p = 1000$ mm, $\alpha = 750$ mm and $\beta = 100$ mm. The model is easy to program on a hand calculator or a spreadsheet[4]. The system dynamics is graphically demonstrated in Fig. 3, where interesting geometrical surfaces appear, showing that the transformation $\boldsymbol{x}_i = f(\boldsymbol{x}_{i-1})$ is not invertible. It should be stressed that no explicit "agent" of randomness (e.g. perturbation by a random number generator) has been introduced into the system.

Moreover, there is no external forcing imposing change onto the system. If any change occurs, it is caused by internal reasons, that is, by an "imbalance" of the vegetation cover and water stored. To see this, let us assume that at time $i=0$ the system state is somewhat different from the equilibrium state, setting initial conditions $x_0 = 100$ mm ($\neq 0$) and $v_0 = 0.30$ ($\neq 0.25$). Using Eqs. (1) and (2) we can calculate the system state (x_i, v_i) at times $i > 0$. The trajectories of x and v for time $i=1$ to 100 are shown in Fig. 4. Apparently, the system remains "alive", i.e. it exhibits change all the time, and its state does not converge to the equilibrium. The trajectories, albeit produced by simple deterministic dynamics, are interesting and seem periodic; we will discuss their properties in Sect. 5.

As the dynamics is fully deterministic, one may be tempted to cast predictions for arbitrarily long time horizons. For example, for time 100, iterative application of the simple dynamics allows to calculate the prediction $x_{100} = -244.55$ mm, $v_{100} = 0.7423$ (as plotted in the right end

---

[4]The interested reader can find a spreadsheet with the toy model in http://www.itia.ntua.gr/en/docinfo/923/.
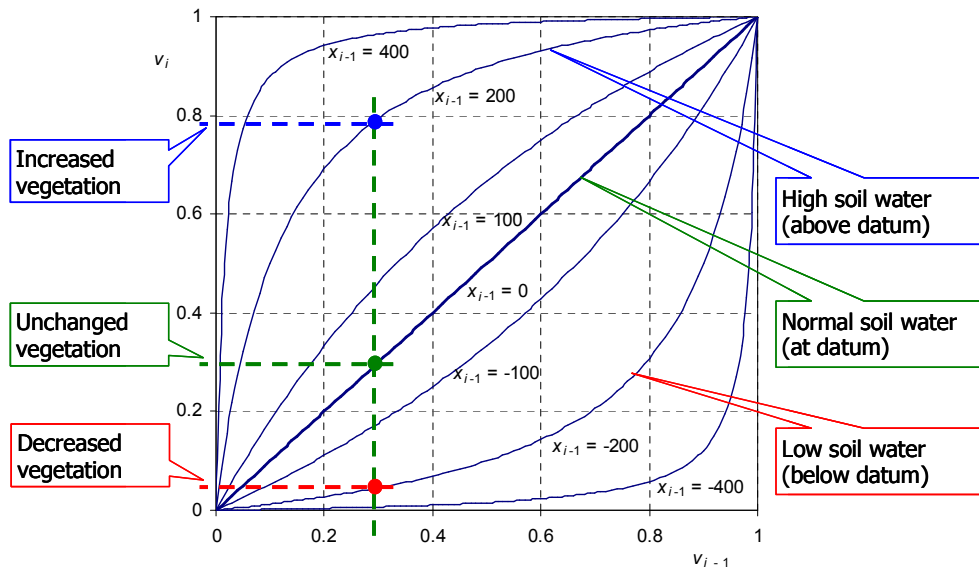
Fig. 2. Conceptual dynamics of vegetation for the caricature hydrological system.
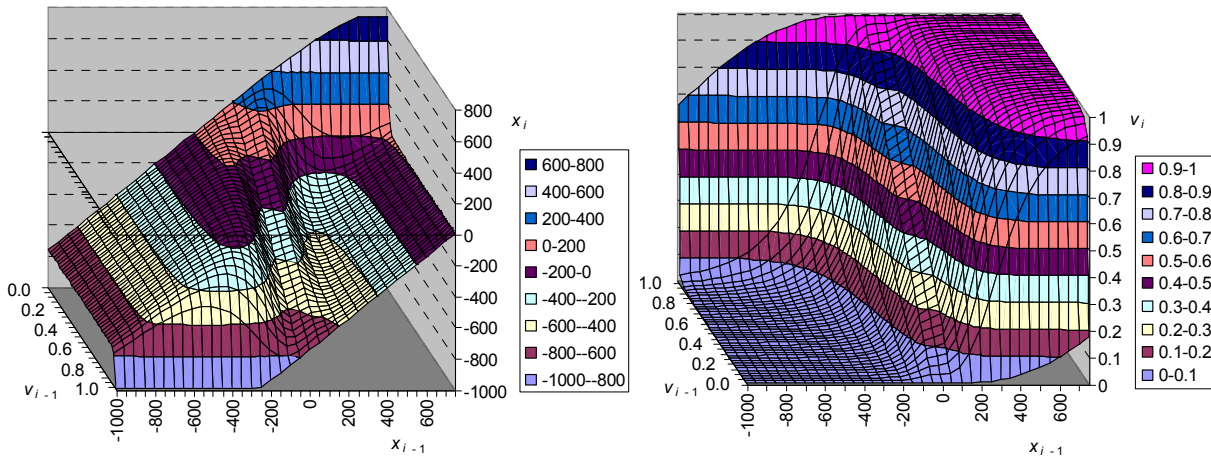

Fig. 3. Graphical depiction of the toy model dynamics.

of Fig. 4). Furthermore, one may be tempted to think that the "primitive science" that this system represents, if any, has come to an end with the above discourse. We have already achieved an understanding of the system, its driving mechanisms and the causative relationships: (a) there is water balance (conservation of mass); (b) excessive soil water causes increase of vegetation; (c) deficient soil water causes decrease of vegetation; (d) excessive vegetation causes decrease of soil water; and (e) deficient vegetation causes increase of soil water. And we have completely and precisely formulated the system dynamics, which is fully consistent with this understanding, very simple, fully deterministic, nonlinear and chaotic.

However, science is not identical to understanding. As R. Feynman (1965) stated, "I think I can safely say that nobody understands quantum mechanics" – and this does not preclude quantum mechanics from being science. Literally,

the name *science* points to *overstanding*[5], and understanding is not identical, nor a prerequisite, to overstanding. Perhaps an imbalance of understanding, which pertains to observing the details of a system, and overstanding, which aims at an overall image of the system, the "forest" rather than the "tree", keeps science alive.

We can thus hope that, despite achieving a good understanding of the system mechanisms and a precise formulation of its dynamics, science may have not come to an end, as far as our toy model is concerned. Let us now focus on predictions, especially of the future, which is a crucial target of science – with even higher importance in engineering. Does, really, deterministic dynamics allow a reliable prediction at

---

[5]Science < Latin Scientia < translation of Greek Episteme (Επιστήμη) < Epistasthai (Επίστασθαι) = to know how to do < [epi (επί) = over] + [histasthai (ίστασθαι) = to stand] = to overstand.
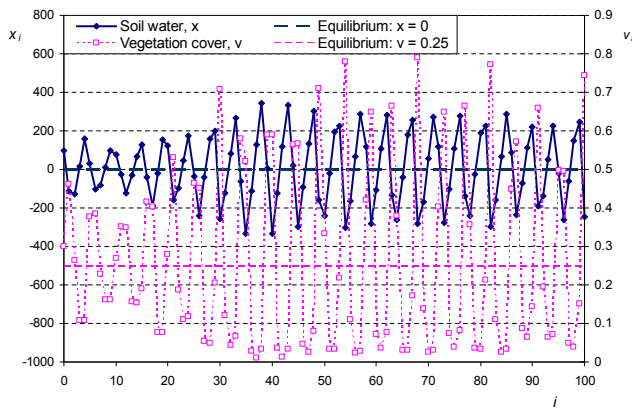
Fig. 4. System trajectory for 100 time steps assuming initial conditions x_0=100 mm and v_0=0.30.

Fig. 5. Evolution of soil water x as in Fig. 4 but with uncertainty in initial conditions: bold blue line corresponds to initial conditions s_0=100 mm, v_0=0.30 and the other five lines represent initial conditions slightly (<1%) different.

an arbitrarily long time horizon, as in our above example? In the previous section, in constructing our prediction for time 100 (x_{100}=−244.55 mm, v_{100}=0.7423), we, explicitly or implicitly, assumed that we know the parameters and initial state with full precision. However, these are real numbers. It is now well known that not only can real numbers not be known with full precision, but (with probability 1), they are not computable (Chaitin, 2004). Therefore, our further investigations will incorporate the premise that a continuous (real) variable cannot ever be described with full (infinite) precision, particularly if it varies in time. This premise, which will be called the premise of incomplete precision, is consistent with mathematics (cf. Chatin's results), as well as with physics (cf. W. Heisenberg's, 1927, uncertainty principle).

It is reasonable, then, to assume that there is some small uncertainty, at least in the initial conditions (initial values of state variables). Perhaps it would be reasonable to assume that there is uncertainty also in the parameters and in model Eq. (1) (but not in Eq. (1), which represents preservation of mass). However, to keep our study simple, we will restrict our investigation to the uncertainty of initial conditions. Figure 5 shows the trajectory of the soil water x for the already examined initial conditions (x_0=100 mm, v_0=0.30) and for five more sets of initial conditions only slightly (<1%) different from the basic set. At short times, the differences in the trajectories are not visible in Fig. 5. At about time 20, the differences become visible and slightly later (time ~30) they become large. Soon thereafter, the different trajectories become unrelated to each other. This shows that a tiny uncertainty in initial conditions gets amplified after some time, a fact well known in chaotic systems since H. Poincaré's discovery of chaos (e.g. Poincaré, 1908). As a result, the deterministic dynamics can produce good predictions only for short time horizons. For longer time horizons, the deterministic predictions become extremely inaccurate and useless. In other words, at long times the system behaviour is unpredictable, that is, random, whereas at short times it is very well described by the deterministic dynamics. This could
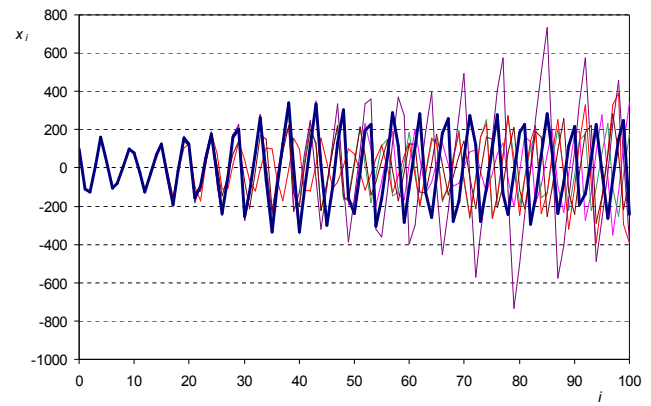
be thought of as the emergence of randomness from determinism. Later, in Sect. 5, we will study the inverse case, i.e. whether randomness could give rise to determinism (at a macroscopic level).

We can easily imagine that if the system dynamics were different, so as to drive it to its "dead" equilibrium state, there would not be uncertainty in the future. The nonlinear type of dynamics we used is the agent that made the system "alive", i.e. changing and not dying. Apparently, what makes the system alive is the same agent that creates the uncertainty. Only dead systems are certain – and this might be useful to recall when thinking about eliminating uncertainty.

The type of uncertainty we observe here could be hardly classified in categories typically used in hydrology. It is not model uncertainty, i.e. incomplete representation of reality, because our system is artificial. It is not parameter uncertainty, because we assumed that the parameters are completely known. It is not even data uncertainty, as our inputs and outputs are assumed fully known and constant, and in fact we have not assumed any measurement error. The uncertainty in the initial conditions should be thought of as a consequence of the premise of incomplete precision, rather than as a measurement error. One may think that the assumed uncertainty 1% is too high to represent this premise. But we used this number just for better illustration. One may easily experiment with lower uncertainties in initial conditions to see that the behaviour does not change. Only the time span of predictability changes. For example, reducing uncertainty from 1% to 10^{-6} will extend the predictability time span, but not more than double it.

All alive natural systems behave more or less this way, and only the predictability time span changes. This view unifies phenomena as diverse as the movement of dice and planets, although in the former the time span of predictability is less than a second, whereas in the latter it is several millions of years. As strange as it seems, even the solar system is chaotic

and unpredictable in such long horizons (Laskar, 1989). For example, it has been shown that it may never be possible to accurately calculate the location of the Earth in its orbit 100 million years in the past or into the future (Duncan, 1994; Lissauer, 1999), whereas an accurate solution for the orbital and precessional motion of the Earth is limited between 35 to 50 million years (Laskar, 1999).

3   From determinism to stochastics

As simple and obvious as the premise of incomplete precision may seem, it implies a radically different perception and study of physical phenomena. First of all, the proper visualization of the trajectory of a system's evolution can no longer be a line or a thread. Rather it should be a stream tube (a notion familiar to hydrologists) of nonzero size (distance between its imaginary walls). The path this tube follows is important to know, but the size of the tube is equally important. This size is not constant, but varies in time. When an observation is made at a time i_o, the size of the tube at this time becomes tiny. In our hypothetical system this small size represents an error related to the premise of incomplete precision but in real-world systems it represents a (usually much higher) observation error. However, for future times of prediction ($i > i_o$), as well as past times ($i < i_o$) at which no observation had been made, the size becomes much larger.

Initially, we can try to define the imaginary walls of this stream tube as the envelope curves of several model runs with perturbed initial conditions within the assumed uncertainty bounds. The tube visualization of this type for x_i is shown in Fig. 6 in two cases, when only x_0 is observed and when x_0 to x_{30} are observed, and with only 5 model runs (as in Fig. 5) as well as with 1000 model runs. At small times the tube has a size too tiny to be seen and a rough shape. The latter is typical in systems with discrete-time dynamics (in continuous time it would be smooth, but the dynamics and the calculations of a continuous-time model would be too complicated to serve our purpose). At long times, the size gets much larger and increases with the number of model runs that are used to construct the envelope curves. It can be expected that, as this number tends to infinity, the stream tube (the zone between the envelopes) will tend to cover all available space; in the case of x, this is the interval between $-\infty$ and $\alpha = 750$ mm. Apparently, the dependence of the size on the number of model runs and the large or infinite size of tube as this number increases are deficiencies of the envelope method. Because of these deficiencies, a deterministic approach of this type, i.e. based on lower and upper physical bounds (cf. the probable maximum precipitation concept), cannot help to effectively describe the stream tube size and the uncertainty.
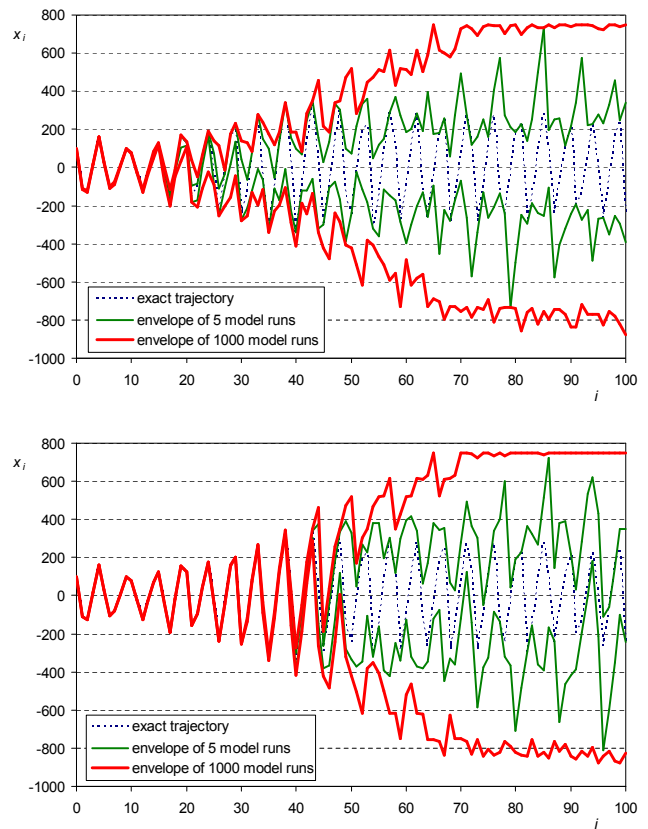

Fig. 6. Stream tube representation of the evolution of soil water x (with initial conditions as in Fig. 4 and 1% uncertainty) in a deterministic setting using envelope curves constructed by 5 or 1000 simulation runs: (upper) when only x_0 is observed; (lower) when x_0 to x_{30} are observed.

Here comes probability and stochastics, which will give us a good description of the tube size (independent of the number of model runs) as well as a profile of the likelihood that the system state is at any specified position, a profile reminding of the profile of longitudinal velocity across a stream tube in real flows. One may wonder: Is it permissible to use probability in a system that is purely deterministic, as the system we investigate here is? The answer I propose is a categorical "yes". This answer is consistent with the unified notion of randomness discussed above, as well as with the concept of *probabilization of uncertainty*, that is, the axiomatic reduction from the notion of an uncertain quantity to the notion of a random variable (Robert, 2007). To my perception, nothing in the Kolmogorov (1933) axiomatic system prohibits this reduction. In a probability-theoretic context, an unknown value x_i is a realization of a random variable $\underline{x}_i$ and is associated with a probability density function $f(x_i)$. A family of random variables $\underline{x}_i$, (arbitrarily, usually infinitely, large) is a stochastic process whereas a realization of the stochastic process, i.e. a series of numbers x_i is a *time series*[6].

---

[6]In some texts the two terms are used as synonymous, but distinction helps avoid ambiguity and misunderstanding.

Probability along with the related fields of *statistics* and *stochastic processes* are currently described by the collective term *stochastics*[7]. The current meaning of this scientific term is no different from that first given by Jakob Bernoulli (1713 – Ars Conjectandi, written 1684–1689). Specifically, Bernoulli defined stochastics as the *science of prediction*, or the science of measuring as exactly as possible the probabilities of events. (In this respect, stochastics should not be identified with the very common ARMA or similar types of models.)

To make up a stochastic formulation of the evolution of the system, first, we fully utilize the known deterministic dynamics $x_i = S(x_{i-1})$, where S is the vector function representing the deterministic dynamics. In our system, the state x_i is the vector (x_i, v_i) and the transformation S is described by Eqs. (1) and (2) and is graphically depicted in Fig. 3. Second, we assume that the probability density at time 0, $f(x_0)$, is known. Third, we use the following concept from the theory of dynamical systems: given the probability density function at time $i-1$, $f(x_{i-1})$, that of next time i, $f(x_i)$, is given by the *Frobenius-Perron operator* FP, i.e. $f(x_i) = FP f(x_{i-1})$, uniquely defined by an integral equation (e.g. Lasota and Mackey, 1991), which in our case takes the following simplified form,

$$FP f(x) = \frac{\partial^2}{\partial x \, \partial v} \int_{S^{-1}(A)} f(u) du \qquad (3)$$

where $A := \{x; \, x \le (x, v)\}$ and $S^{-1}(A)$ is the *counterimage* of A, i.e. the set containing all points x whose mappings $S(x)$ belong to A. Iterative application of the equation can determine the density $f(x_i)$ for any time i.

This shows that the stochastic representation has an analytical expression, as has the deterministic. However, the stochastic representation refers to the evolution in time of admissible sets and densities (the stream tube visualization), rather than to trajectories of points (the thread visualization). From the deterministic, "exact" but inaccurate, thread-like trajectory $x_i = S(x_{i-1})$, we have moved to the tube-like trajectory:

$$f_i(x) = \frac{\partial^2}{\partial x \, \partial v} \int_{S^{-1}(A)} f_{i-1}(u) du \qquad (4)$$

which has been written in a manner slightly different from Eq. (3). Clearly, the stochastic formulation does not disregard the deterministic dynamics: it is included in the counterimage $S^{-1}(A)$. However, it can be easily extended to describe non-deterministic dynamics by generalizing the FP operator (Lasota and Mackey, 1991).

---

[7] Stochastics < Greek Stochasticos (Στοχαστικός)< Stochazesthai (Στοχάζεσθαι) < Stochos (Στόχος); Stochos = target; Stochazesthai = (a) to aim, point, or shoot (an arrow) at a target > (b) to guess or conjecture (the target) > (c) to imagine, think deeply, bethink, contemplate, cogitate, meditate.
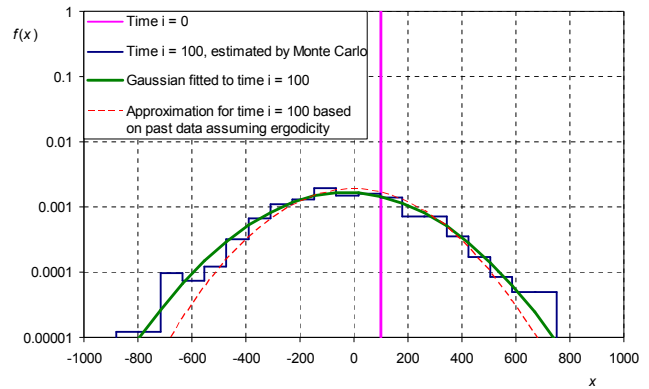

Fig. 7. Probability density functions $f_i(x)$ of the system state x (soil water) for times $i=0$ and 100.

In the iterative application of the stochastic description of the system evolution we encounter two difficulties. First, despite being simple, the function S is not invertible and the integral over the counterimage $S^{-1}(A)$ needs to be evaluated numerically. Second, as the deterministic formulation is quite satisfactory for short time horizons, the stochastic formulation gets more meaningful for long ones. Iterative application of Eq. (4) over time will result in multiple integrations, so that eventually, for long time horizons, we need to perform a *high dimensional numerical integration*. This is difficult, unless a stochastic integration method is used. Specifically, it is easily shown (e.g. Metropolis and Ulam, 1949; Niederreiter, 1992) that for a number of dimensions $d > 4$, a *stochastic (Monte Carlo) integration* method (in which the function evaluation points are taken at random) is more accurate (for the same total number of evaluation points) than classical numerical integration, based on a grid representation of the integration space.

In our example, the Monte Carlo method does not involve other calculations than those we did to construct the envelopes above. It is so very simple that it even bypasses the calculation of $S^{-1}(A)$. Results for the density function $f_i(x)$ of the system state x (soil water) for time $i=100$, in comparison to that for time $i=0$, are shown in Fig. 7. The Monte Carlo integration was performed assuming $f(x_0)$ to be uniformly extending 1% around the value $x_0=(100$ mm, $0.30)$ and using 1000 simulations. The Monte Carlo method is very powerful, yet so easy that we may fail to notice that we are doing numerical integration. Technically, its application to construct Fig. 7 would be done even intuitively, without knowing that there is some concrete mathematical background (Eq. 3) behind our simulations. However, being conscious of the theoretical background and the associated strengths and limitations, results in most efficient use of the method, in acceleration of scientific progress, and in avoidance of confusions and potential errors.

It is observed in Fig. 7 that moving from time $i=0$ to $i=100$, the density changes from concentrated to broad and from uniform to Gaussian; the theoretical Gaussian curve is also

plotted. Knowing a priori, for theoretical reasons, that the probability distribution, after a long time, will be Gaussian is very important and substantially simplifies the solution of problem. But are there theoretical reasons implying Gaussian distribution? Jaynes (2003) lists a number of them. The most widely known is the Central Limit Theorem. In its most common formulation, which involves sums of random variables, it seems inapplicable here as there are no such sums. However, if interpreted as a statement about the properties of density functions under convolution (multiple integrals of a number of density functions tends to the Gaussian density) it may give the required explanation. However, it is more convenient to express our reasoning in terms of the *principle of maximum entropy*: for fixed mean and variance, the distribution that maximizes entropy is the normal distribution (or the truncated normal, if the domain of the variable is an interval in the real line; Papoulis, 1991). Entropy[8] is a probabilistic concept, which for a continuous random variable $\underline{x}$ is defined as

$$\phi[\underline{x}] := E[-\ln f(\underline{x})] \tag{5}$$

where $E[g(\underline{x})]$ denotes the expectation of any function $g(x)$, i.e.

$$E[g(\underline{x})] := \int_{-\infty}^{\infty} g(x) f(x) \mathrm{d}x \tag{6}$$

Entropy is a typical measure of uncertainty, so its maximization indicates that the uncertainty becomes as high as possible. Entropy could be used to quantify the notions of randomness (high entropy) and determinism (lowest entropy). Given that information and entropy are more or less the same quantity, this quantification agrees with Kolmogorov's and Chaitin's view of mathematical randomness.

The principle of maximum entropy was introduced by Jaynes (1957) as a method to infer unknown probabilities from known information. However, much earlier, in statistical thermophysics, entropy maximization had constituted the theoretical basis of the *Second Law of thermodynamics*, where the optimization constraints of fixed mean and variance correspond to the conservation of momentum and energy, respectively (assuming that the microscopic description is in terms of momenta). In this case, it is used to determine macroscopic thermodynamic states of physical systems. Some regard the application of the principle of maximum entropy in statistical thermophysics fundamentally different from its application in inference, as used here. In my opinion, though, the two must be the same. Observing that Nature spontaneously maximizes entropy, i.e. uncertainty, why not use just the same principle in logic for inference about natural systems? (For a profound discussion of the concept of entropy and its controversial views in thermophysics, the interested reader is referred to Robertson, 1993).

___

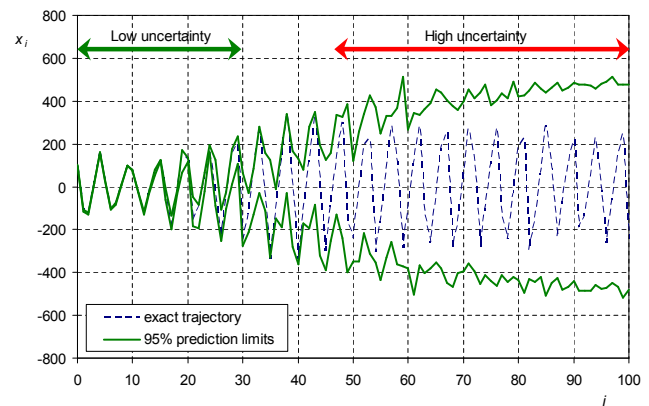[8] Entropy < Greek εντροπία < entrepesthai (εντρέπεσθαι) = to turn into.

Fig. 8. Stream tube representation of the evolution of soil water x (with initial conditions as in Fig. 4 and 1% uncertainty) in a stochastic setting using Monte Carlo prediction limits for 95% probability of bracketing the true state between the stream tube "walls" (1000 simulation runs).

Coming back to our example, we can use the same simulation method – in fact the same simulation runs – to trace the probability densities for all times. The propagation of uncertainty in time is typically visualized through prediction intervals, whose limits define the stream tube "walls" according to the probabilistic approach. An example is shown in Fig. 8, for a certain probability, say 95%, of the true state being contained in the stream tube. Now the stream tube size does not depend on the number of simulations and does not diverge to infinity (except if the probability chosen to visualize the stream tube is 100%). Thus, the probabilistic stream tube visualization is more meaningful, convenient and informative than the envelope representation of Fig. 6. It is observed that for long time horizons the stream tube becomes less rough and its size, i.e. the uncertainty, tends to stabilize to a maximum value. This defines another type of equilibrium, a statistical thermodynamic equilibrium of maximum entropy. To distinguish it from the static or "dead" equilibrium of Sect. 2, we can call it the "alive" equilibrium.

4 The power of data

As our prediction horizon increases and we approach the "alive" equilibrium, we may find it natural to raise this question: Do we really need the deterministic dynamics to make a long-term prediction? Intuitively, from Fig. 8, the answer seems to be negative. But, then, what can replace the dynamics? Recalling that the form of the density function may be known a priori, as discussed above, what we need to completely express the density function are the two parameters of the Gaussian curve, namely its mean and standard deviation. But these could be estimated from data. Hence, for long horizons past data render knowledge of dynamics unnecessary. For illustration, Fig. 9 shows a record of 100 past values of x_i corresponding to times $i = -100$ to -1. Here,

because our caricature system is imaginary, the past data are synthetic, generated by the same model, but in a real system with really unknown dynamics these would be past observations of the system state. To generate the data here we assume initial conditions: $\boldsymbol{x}_{-100}$=(73.99 mm, 0.904), for which the resulting state at time i=0 is $\boldsymbol{x}_0$=(99.5034≈100 mm, 0.3019≈0.30). This state is compatible (within precision 1%) with the rounded off initial state $\boldsymbol{x}_0$=(100, 0.30) that we used in earlier investigations. Interpreting past data as a statistical sample, we estimate a sample mean μ=−2.52 mm and a sample standard deviation σ=209.13 mm. With these values we can obtain the complete density function for time i=100, which is plotted in Fig. 7 along with the results obtained by the Monte Carlo simulation, in which the deterministic dynamics was explicitly taken into account. It can be seen that the empirical result without considering the dynamics is a good approximation.

Despite being empirical, this result and, more generally, the use of past data in prediction, can find a theoretical justification in the concept of *ergodicity*[9], an important concept in dynamical systems and stochastics. By definition (e.g. Lasota and Mackey, 1994, p. 59), a transformation is ergodic if all its invariant sets are trivial (have zero probability). In other words, in an ergodic transformation starting from any point, a trajectory will visit all other points, without being trapped to a certain subset. (In contrast, in non-ergodic transformations there are invariant subsets, such that a trajectory starting from within a subset will never depart from it). An important theorem by G. D. Birkhoff (1931) says that for an ergodic transformation $\boldsymbol{S}$ and for any integrable function g the following property holds true:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} g\left(\boldsymbol{S}^i(\boldsymbol{x})\right) = \int_{-\infty}^{\infty} g(\boldsymbol{x}) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \quad (7)$$

with the right-hand side representing the expectation $E[g(\underline{\boldsymbol{x}})]$. For example, for $g(\boldsymbol{x})=\boldsymbol{x}$, setting $\boldsymbol{x}_0$ the initial system state, observing that the sequence $\boldsymbol{x}_0$, $\boldsymbol{x}_1$=$\boldsymbol{S}(\boldsymbol{x}_0)$, $\boldsymbol{x}_2$=$\boldsymbol{S}^2(\boldsymbol{x}_0)$, ..., represents a trajectory of the system, and taking the equality in the limit as an approximation with finite (n) terms, we obtain that the time average equals the true (ensemble) average $E[\underline{x}]$:

$$\frac{1}{n} \sum_{i=0}^{n-1} x_i \approx \int_{-\infty}^{\infty} x f(x) \mathrm{d}x \quad (8)$$

Thus, ergodicity allows estimation of the system properties using past data only. The question then arises: If the dynamics is known, should a long-term prediction be better based on the data or on the dynamics? To explore this question, let us compare two different types of predictions: (a) a typical deterministic prediction, based on applying the dynamics as

---

[9] Ergodicity < Greek εργοδικός< [ergon (έργον) = work] + [odos (oδός) = path].
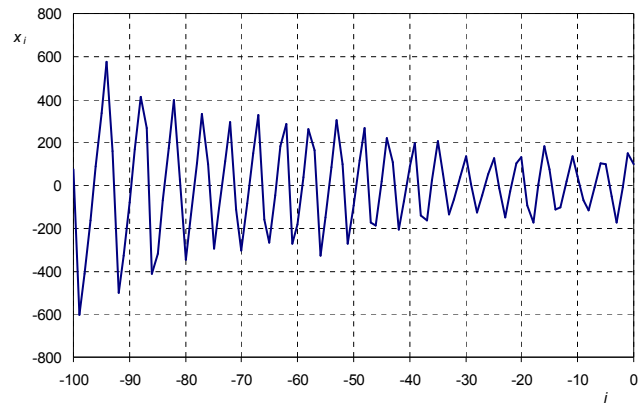

Fig. 9. A sample of past data of soil water x_i for times i=−100 to −1.

done initially in Fig. 4; (b) a naïve statistical prediction, according to which the future equals the average of past data. Stochastics provides the tool to compare the two predictions, which is the standard (root mean square – RMS) error,

$$e_i := \sqrt{E\left[(\underline{x}_i - \hat{x}_i)^2\right]} \quad (9)$$

where $\underline{x}_i$ denotes the random variable representing the state and $\hat{x}_i$ denotes the specific prediction for time i provided by either method (a) or (b). It is easily seen that for method (b), in which $\hat{x}_i$=μ (mean, estimated at −2.52 mm), the standard error equals the standard deviation σ (estimated at 209.13 mm), and is constant for all i. In method (a), e_i is different for different times i and can be evaluated again by Monte Carlo integration, now of Eq. (9). The results are shown in Fig. 10. Clearly, in short lead times (<∼30) the deterministic forecast is better, but in long lead times (>∼45) the naïve statistical forecast is superior.

However, this is not a surprise. Actually, stochastics can give us an a priori estimate of the deterministic model error, applicable near the "alive" equilibrium, where the uncertainty has stabilized. Thus, from Eq. (9) we obtain

$$e_i := \sqrt{E\left[(\underline{x}_i - \hat{x}_i)^2\right]} = \sqrt{E\left[\left\{(\underline{x}_i - \mu) + (\mu - \hat{x}_i)\right\}^2\right]} \quad (10)$$

which after typical manipulations results in

$$e_i = \sqrt{\sigma^2 + (\mu - \hat{x}_i)^2} \quad (11)$$

When it happens that $\hat{x}_i$=μ, then e_i=σ, as in the statistical prediction; otherwise the error in the deterministic prediction is obviously greater than σ. The a priori error estimates are also plotted in Fig. 10 and agree well with those obtained by Monte Carlo simulation. By treating $\hat{x}_i$ also as a random variable we easily obtain that the average error of the deterministic forecast over all times (above ∼45) will be e_i=$\sigma\sqrt{2}$ (also plotted in Fig. 10), which shows that, on the average, the naïve statistical prediction outperforms the deterministic prediction by a factor of $\sqrt{2}$.
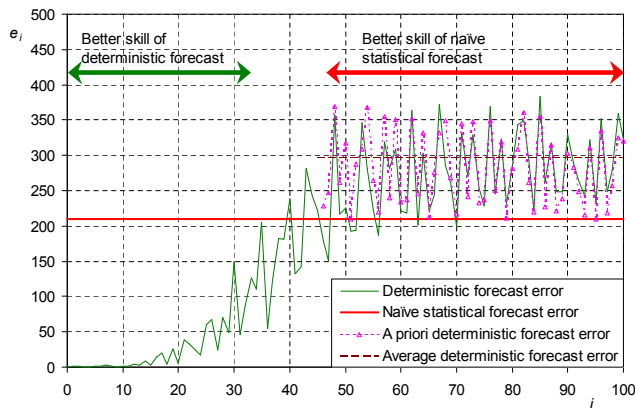
Fig. 10. Comparison of the RMS prediction error of soil water x for the deterministic prediction and the naïve statistical prediction.

This example shows that for long horizons the use of deterministic dynamics gives misleading results and a dangerous illusion of exactness. Unless a stochastic framework is used, neglecting deterministic dynamics in long-term prediction is preferable. In very complex systems, the same behaviour could emerge also in the smallest prediction horizons. This justifies, for example, the so-called ensemble forecasting in precipitation and flood prediction. In essence, it does not differ from this stochastic framework discussed, and is much more effective and reliable than a single deterministic forecast.

In seeking a more informative prediction than the naïve prediction, a natural question is: Is reduction of uncertainty possible for long time horizons? In our simple example the answer is categorical: No way! For there is no margin for better knowledge of dynamics (we have assumed full knowledge already). And there is indifference of potentially improved knowledge of initial conditions. As mentioned above, reduction of initial uncertainty from 1% to 10^{-6} results in no reduction of final uncertainty at $i=100$. Therefore, a more informative prediction cannot be a prediction with reduced uncertainty. Rather, it must be a point prediction accompanied by quantified uncertainty. This has already been done in Fig. 7.

In summary, for long time horizons, the stochastic inference using (a) past data, (b) ergodicity, and (c) maximum entropy, provides an informative prediction. Knowledge of dynamics does not improve this prediction. For short time horizons, the stochastic framework also incorporates the deterministic dynamics and uses it in a Monte Carlo framework. Thus, the stochastic representation is an all-times solution, good for both short and long horizons, and helps figure out when the deterministic dynamics should be considered or neglected. It also highlights the usefulness and importance of data (which may have been missed today as indicated by the ongoing dismantling of hydrological monitoring networks).

In theory, a good data set allows even the recovery of dynamics, if it is unknown, employing the Whitney (1936) and

Takens (1981) embedding theorems. The recovery is based on time-delay vectors $\boldsymbol{x}_i^m := (x_i, x_{i-1}, \ldots, x_{i-m+1})$ of a single observable x_i, with the required vector size m being no more than $2d+1$, where d is the system dimensionality, which can also be estimated from the time series. Forming time-delay vectors $\boldsymbol{x}_i^m$ with trial size m, we are able to calculate the multidimensional entropy $\phi_m(\varepsilon)=E[-\ln p]$, where ε is a scale length (side length of hypercube) related to a grid covering the m-dimensional space, on which the empirical probability p that a data point $\boldsymbol{x}_i^m$ belongs to each hypercube is calculated (notice the difference from definition Eq. (5), i.e. the replacement of the probability density f with the probability mass p). The entropy $\phi_m(\varepsilon)$ is a decreasing function of ε and tends to infinity as ε tends to zero. The limit of $\phi_m(\varepsilon)/(-\ln\varepsilon)$ as ε tends to zero, which (according to de l'Hôpital's rule) is also equal to the limit of the slope $d_m(\varepsilon):=-\Delta\phi_m(\varepsilon)/\Delta\ln\varepsilon$, gives the dimension of the subspace of the m-dimensional space where the set of $\boldsymbol{x}_i^m$ lies. For small ε, $d_m(\varepsilon)$ cannot exceed m nor d. Application of a standard algorithm that implements this idea for increasing trial values of m (Grassberger and Procaccia, 1983; Koutsoyiannis, 2006) is demonstrated in Fig. 11, where it can be seen that d_m does not exceed $d=2$, thus capturing the system dimensionality, which is 2. Note, however, that a large data set is required for the application of this technique. Our toy model can easily give us arbitrarily long time series (here we used a time series of 10 000 points rather than 100 used before). But shortness of data or poor attentiveness in the statistical properties of the data may result in erroneous conclusions (Koutsoyiannis, 2006). Other stochastic tools that can recover deterministic controls from data are discussed in the next section.

5   Exploration of the long term stochastic properties of the system

Arguably, when we are interested in a prediction for a long time horizon, we wish to know not the exact system state at a specified time but an average behaviour around that time and perhaps a measure of the dispersion of the extremes around it. This implies a different perspective, macroscopic rather than detailed, of long-term prediction and predictability. In this we can disregard the "instantaneous" system state, which in atmospheric sciences is referred to as the *weather*, and try to predict the long-term average for a future period, commonly referred to as the *climate*. According to a common definition, climate is "the long-term average of conditions in the atmosphere, ocean, and ice sheets and sea ice described by statistics, such as means and extremes (US Global Change Research Program, 2009). The usefulness of the notion of climate as a long-term average extends also to hydrological processes. Actually, in studies of "climate change" and its impacts, including those in hydrology and water resources, long-term predictions always refer to long-term average conditions.
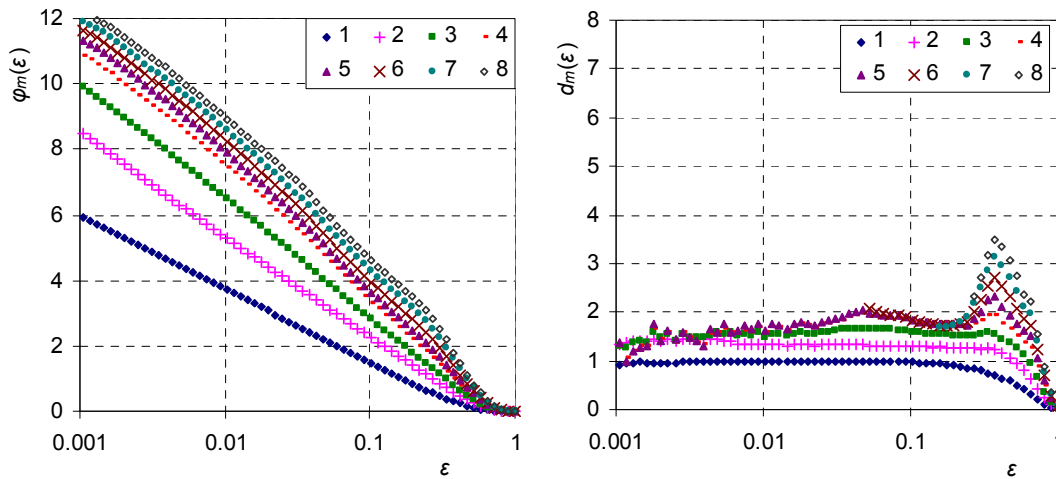
Fig. 11. Graphical depiction of the application of the algorithm to recover the system dynamics (with emphasis of its dimensionality) from a time series of 10 000 points; numbers in the legend box are trial embedding dimensions m.

To study this macroscopic notion of prediction and predictability, we need long simulations and data series that enable observation of long-term behaviours. In all following illustrations we use time series with lengths of 10 000. The first thousand terms of the time series of the soil water x_i, generated with the same initial conditions as before, are shown in Fig. 12 (upper panel). The plot shows high variability at the shortest time scale (i.e. 1), with peculiar variation patterns not visible in the plots of fewer data points presented earlier. Yet it shows relatively flat time average at scale 30 ("climate"). Despite high variability at scale 1, the trajectory of the system state does not resemble a purely irregular or random pattern. Rather the trajectory seems cyclical, but a more careful investigation reveals that the time series differs from that of a typical periodic deterministic system. Specifically, as shown in Fig. 13 (upper panel) there is no constant periodicity, but the time δ between successive peaks of the time series x_i varies between 4 and 10 time steps, with a period of 6 time steps being the most frequent.

The relatively flat average of the soil water x_i makes prediction of the long-term average rather trivial in this caricature system. However, other problems, in which variability plays a role (e.g. long term behaviour of extremes), are less trivial and more interesting, even in this very simple system. To study the peculiar variability of $\underline{x}_i$, we introduce the random variable $\underline{y}_i := |\underline{x}_i - \underline{x}_{i+6}|$, where the time lag 6 was chosen to be equal to the most frequent δ. We call $\underline{y}_i$ the *variability index* and we study its long-term behaviour in comparison to that of $\underline{x}_i$. It can be easily verified that $\underline{y}_i$ represents the sample standard deviation of the size-two sample $\underline{x}_i$ and $\underline{x}_{i+6}$. Apparently, the standard deviation of a number of consecutive x_i (e.g. the seven terms $\underline{x}_i$ to $\underline{x}_{i+6}$) would give a more representative variability index, but we choose this simpler definition to avoid artificial dependence between successive time series terms or else to avoid the need to change scale. Besides, the simple definition serves well our exploration

purpose. A plot of the first 1000 terms of the time series of y_i is shown in Fig. 12 (lower panel) and reveals a different behaviour than x_i (upper panel). Here the variability is high, not only at a short time scale but also at a long one. The long "excursions" of the moving average of 30 values ("the climate") from the global average (of 10 000 values) are quite characteristic.

To explore the long-term stochastic properties of our system, including periodicity and time dependence or persistence, we use three stochastic tools. The first is the periodogram, i.e. the square absolute value of the Discrete Fourier Transform of the time series. It is a real function $q(\omega)$, where ω is frequency. The quantity $q(\omega)d\omega$ is proportional to the fraction of variance explained by ω and thus excessive values of $q(\omega)$ indicate cyclicity with period $1/\omega$. The periodogram of 10 000 terms x_i is shown in Fig. 13 (lower panel).

The second is the empirical autocorrelation function (autocorrelogram), i.e. the Finite Fourier Transform of the periodogram (divided by variance). It is a sequence of values ρ_j, where j is a lag. It is alternatively defined and more easily determined as $\rho_j = \text{Cov}[\underline{x}_i, \underline{x}_{i-j}] / \text{Var}[\underline{x}_i]$, where $\text{Cov}[\underline{x}_i, \underline{x}_{i-j}] := E[(\underline{x}_i - \mu)(\underline{x}_{i-j} - \mu)]$, $\text{Var}[\underline{x}_i] := \text{Cov}[\underline{x}_i, \underline{x}_i] = E[(\underline{x}_i - \mu)^2]$, and $\mu = E[\underline{x}]$. The autocorrelograms of both x_i and y_i are shown in Fig. 14.

The third tool aims at a multi-scale stochastic representation. Based on the process x_i at scale 1, we define a process $\underline{x}_i^{(k)}$ at any scale $k \geq 1$ as:

$$\underline{x}_i^{(k)} := \frac{1}{k} \sum_{l=(i-1)k+1}^{ik} \underline{x}_l \qquad (12)$$

A key multi-scale characteristic is the standard deviation $\sigma^{(k)}$ of $\underline{x}_i^{(k)}$. The quantity $\sigma^{(k)}$ is a function of the scale $k \geq 1$, here
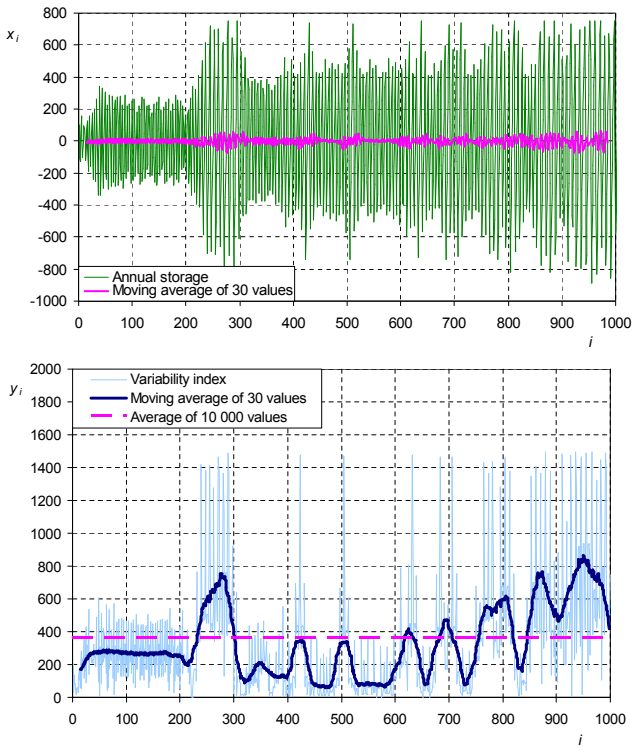
Fig. 12. Evolution of the soil water x_i, the variability index y_i, as well as their moving averages of length 30, for time i up to 1000.

referred to as the *climacogram*[10] and typically depicted on a double logarithmic plot. While the periodogram and the autocorrelogram are related to each other through a Fourier transform, the climacogram is related to the autocorrelogram by a simpler transformation, i.e.

$$\sigma^{(k)} = \frac{\sigma}{\sqrt{k}}\sqrt{\alpha_k}, \quad \alpha_k = 1 + 2\sum_{j=1}^{k-1}\left(1 - \frac{j}{k}\right)\rho_j \leftrightarrow$$

$$\rho_j = \frac{j+1}{2}\alpha_{j+1} - j\alpha_j + \frac{j-1}{2}\alpha_{j-1} \tag{13}$$

It is directly verified (actually this is the most classical statistical law) that in a purely random process

$$\sigma^{(k)} = \frac{\sigma}{\sqrt{k}} \tag{14}$$

However, this law may not be verified in natural systems. The simplest alternative is

$$\sigma^{(k)} = \frac{\sigma}{k^{1-H}} \tag{15}$$

where H is a constant ($0 < H < 1$) known as the Hurst coefficient, after Hurst (1951) who first analyzed statistically the long-term behaviour of geophysical time series. Earlier, Kolmogorov (1940), in studying turbulence, had proposed

---

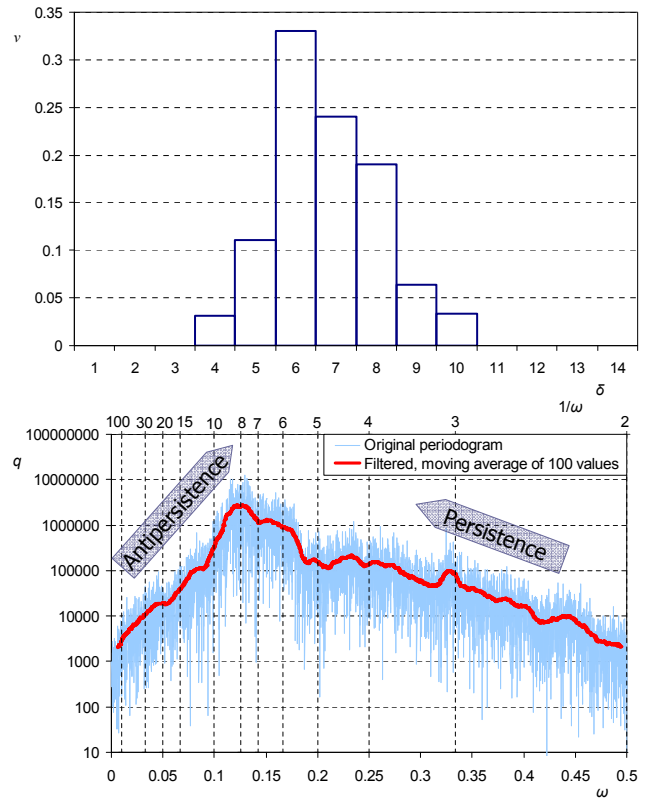[10]Climacogram < Greek Κλιμακόγραμμα < [climax (κλιμαξ) = scale] + [gramma (γράμμα = written].

Fig. 13. Graphical depiction of periodic properties of the system based on (upper): a histogram of the time between successive peaks, δ, of the soil water x_i, where ν is the fraction of occurrences of a certain δ over the total number of occurrences; (lower): the periodogram of the time series of x_i, where ω is frequency and q is the spectral density; 10 000 terms of time series were used for the construction of the plots.

a mathematical model to describe this behaviour. This behaviour has been known with several names including *Hurst phenomenon*, *long-term persistence*, *long range dependence*, and *Hurst-Kolmogorov (HK) behaviour* or *HK (stochastic) dynamics*. At the same time, Eq. (15) defines a simple stochastic model that reproduces this behaviour, known as a *simple scaling stochastic model* or *fractional Gaussian noise* (due to Mandelbrot and van Ness, 1968), or the *HK model*. Climacograms for the time series x_i and y_i are shown in Fig. 15, where the departure from the classical law (Eq. 14) of a purely random process is evident.

A purely random process would have a flat periodogram, but Fig. 13 (lower panel) indicates a different behaviour for x_i. Furthermore, fixed periodicities would be manifested in the periodogram as high impulses in the specific periods, but in the periodogram of Fig. 13 no impulses exist. Rather, the figure indicates relatively higher densities q at a broad band of periods $1/\omega$, between 5–12 time units (agreeing with the simpler representation in the upper panel of Fig. 13). The most noticeable characteristics in the periodogram are the
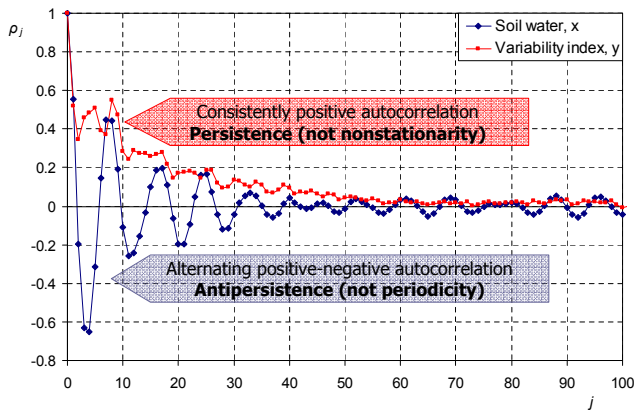
Fig. 14. Empirical autocorrelation functions of 10 000 terms x_i and y_i.

Fig. 15. Climacograms (plots of standard deviation $\sigma^{(k)}$ vs. averaging scale k) of the time series x_i and y_i (where standard deviations were estimated by the classical statistical estimator of time series averaged on each scale k).

increasing spectral densities for low frequencies ($1/\omega > 8$) and the decreasing ones for high frequencies ($1/\omega < 8$). The two different behaviours are indicators of *antipersistence* and *persistence*, respectively.

The autocorrelogram in Fig. 14 depicts the same behaviours in a different manner. It is observed that the autocorrelation for lag one is positive, which is expected because of physical consistence (states in neighbouring times should be positively correlated because the changes in small time should be small) and indicates short-term persistence. For higher lags the autocorrelation of x_i oscillates between negative and positive values. The existence of negative autocorrelations is an indication of antipersistence. In the simplest case, an antipersistent process should have all its autocorrelations negative: it can be verified that in an HK process defined by Eq. (15) with $H < 0.5$, the autocorrelation function is negative everywhere. But this cannot be the case in nature because short-term persistence demands that some autocorrelations should be positive. Therefore, here we broaden this more common but too simplistic and unrealistic notion of antipersistence so as to include alternating negative and positive autocorrelations without a constant period. A strictly periodic behaviour would also result in autocorrelation oscillating between positive and negative values, which creates the risk of misinterpretation of antipersistence as periodicity. However, the intervals between different peaks or troughs in Fig. 14 have not constant length, so here we have antipersistence (sometimes called "quasi-periodic" behaviour).

Figure 14 also shows the autocorrelation of the variability index y_i. In this case, the autocorrelation is always positive, indicating persistence both in short and long term. Processes with consistently positive autocorrelation functions lead to large and long "excursions" from the mean as shown in Fig. 12 (lower panel), which often tends to be interpreted as nonstationarity. The latter, however, would require that the system's dynamics changes in time in a deterministic manner, which does not happen here (and in most of the cases).
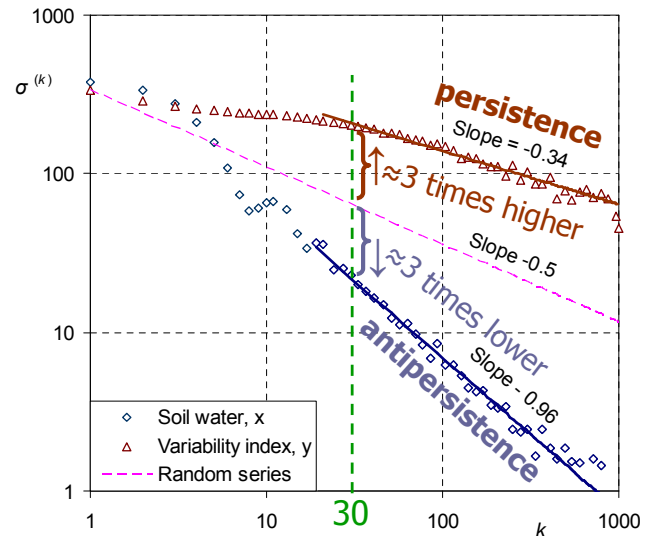
The most characteristic and useful plot of all three is the climacogram in Fig. 15. For the soil water series x_i and for scales 1–3, the slope formed by the empirical points is very low, reflecting short-term persistence. For large scales ($k > 10$) the empirical points are arranged in a straight line with steep slope, -0.98. From Eq. (15) we can see that this slope equals $H-1$, so that in this case $H=0.02$ asymptotically (for large scales). Likewise, the plot for the variability index y_i indicates a slope of -0.34 for large scales, which corresponds to $H=0.66$. The slope for a purely random process, also shown in figure, is -0.5, which corresponds to $H=0.5$. Generally, an H between 0.5 and 1 characterizes long-term persistence, whereas an H between 0 and 0.5 indicates antipersistence. Thus, this figure verifies the antipersistent and persistent behaviour already detected for x_i and y_i, respectively.

Most importantly, this figure provides insights on the predictability issue. It is well known that for a one-step-ahead prediction at scale 1, a purely random process x_i is the most unpredictable. Dependence enhances one-step-ahead predictability. For example, in a process with $\rho_1=0.5$ (comparable to that of our series x_i and y_i) the standard deviation, conditional on known present state, is a fraction $\sqrt{(1-\rho_1^2)}$ of the unconditional one, i.e. 13% smaller. However, in the climatic-type predictions, in which the average behaviour rather than exact values is studied, the situation is different. In our example, as clearly shown in Fig. 15, at the climatic scale of 30 time steps, predictability is deteriorated by a factor of 3 for the persistent process y_i (thus eliminating and largely exceeding the 13% reduction due to conditioning on the present state). On the other hand, for the antipersistence process x_i, the long-term predictability is improved by
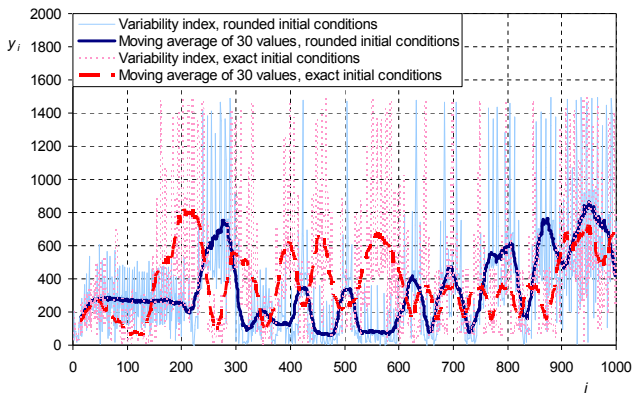
Fig. 16. Plot of 1000 terms of the time series y_i (variability index) at scales 1 and 30 and for two sets of initial conditions deferring less than 1%.

a factor of about 3. In summary (and perhaps contrary to what is believed), long-term persistence substantially deteriorates predictability over long time scales – but antipersistence improves it.

Figure 16 provides further demonstration of the unpredictability of persistent processes. The plot shows 1000 items of the time series y_i (variability index) at the annual and the climatic scale and for the two sets of initial conditions discussed above, the exact and rounded off, which differ by less than 1%. After about 30 time steps (one time unit of "climate"), the departures in the two cases are pronounced. Thus, even a *completely deterministic system* is *completely unpredictable* at a large (climatic) time scale, when there is persistence.

This may seem counterintuitive to many. The conventional argument is that averaging at a large temporal or spatial scale results in "elimination of local noise" or "cancellation of positive and negative tendencies" and gives rise to the *emergence of determinism from randomness*. While the latter is an indisputable truth and is the basis of macroscopic determinism in thermodynamic systems, extreme caution is required not to generalize unjustifiably. Theoretical justification of the emergence of determinism from randomness relies on probability theory and on application of its laws, already mentioned above. Practically, we can speak of emergence of determinism in a macroscopic quantity if the uncertainty of an averaged quantity is so low that we can neglect it. Two things are most important to derive the uncertainty of the macroscopic quantity: the scale of averaging k (number of microscopic elements) and the dependence among the microscopic elements. An example from thermodynamics of ideal gases is enlightening: in a mole of a gas the macroscopic quantities are averages of a number of $k = N_A = 6.022 \times 10^{23}$ molecules (N_A is the Avogadro number) and independence between different molecules can be assumed (e.g. Stowe, 2007). Simple calculations using the classical statistical law (Eq. 14) yield that the uncertainty associated to a single molecule, expressed by the

standard deviation σ, is multiplied by $1/\sqrt{N_A} = 1.3 \times 10^{-12}$ when macroscopic properties are calculated. This tiny number explains why we do not need an accurate representation at the molecular level to have the macroscopic quantity correct; in other words it justifies the emergence of determinism. However, even with that high averaging scale $k = N_A$, if there is substantial positive dependence between the different elements, the uncertainty may still be high. For example, application of the HK law (Eq. 15) with $H = 0.99$ yields that the ratio of macroscopic to microscopic uncertainty is $1/N_A^{-0.01} = 0.58$, i.e. more than 11 orders of magnitude higher than in classical statistics. Evidently, then, we cannot speak of determinism in the latter case. As an additional example, when $k = 30$ and $H = 0.95$ the ratio of macroscopic to microscopic uncertainty is $1/30^{-0.05} = 0.84$. Again, we cannot speak of emergence of determinism. In contrast, in processes with antipersistence the emergence of determinism is much more relevant (here, indeed, positive and negative tendencies cancel each other).

6 From the toy model to the real world

In comparison to our toy model, a natural system, such as the atmosphere, a river basin, etc.: (a) is much more complex; (b) has time-varying inputs and outputs; (c) has spatial extent, variability and dependence (in addition to temporal); (d) has greater dimensionality (virtually infinite); (e) has dynamics that to a large extent is unknown and difficult or impossible to express deterministically; and (f) has parameters that are unknown. Hence uncertainty and unpredictability are naturally even more pronounced in a natural system. The role of stochastics is then even more crucial: (a) to infer dynamics (laws) from past data; (b) to formulate the system equations; (c) to estimate the involved parameters; and (d) to test any hypothesis about the dynamics. Data offer the only solid ground for all these tasks, and failure to rely upon, and test against, evidence from data renders inferences about hypothesized dynamics worthless.

Despite the huge difference of the toy model and natural reality, we may hope that the above discourse can help address several questions, from philosophical to technical. Given the current dominant trend in hydrological (and other geophysical) sciences for *physically-based* modelling, relevant questions are: What is physically-based modelling? Is physics a synonym for determinism? Is physically-based synonymous to mechanistic? Are first principles mechanistic principles? Is not statistical physics part of physics? Is not entropy maximization a first principle? Is not stochastic modelling part of physical modelling? Will it ever be possible to achieve such a physically-based modelling of hydrological systems that will not depend on data or stochastic representations? Can detailed representations and reduction to first principles render hydrologic measurements unnecessary? What level of detail is needed in such reductionist

modelling for a catchment of, say, $1000\,\text{km}^2$? How far can the current research trend toward detailed models advance hydrology and water resources science and technology?

Hydrological uncertainty and its reduction is currently a core research issue and also very important from a water resources engineering and management perspective. Relevant questions are: To what extent can hydrological uncertainty be reduced? Can uncertainty be eliminated by uncovering the system's deterministic dynamics? Is uncertainty epistemic or inherent in nature? When there is potential for reduction of uncertainty, what is the most effective means for reduction? Is it better understanding, better deterministic modelling, more detailed discretization, or better data? When the limits of uncertainty reduction have been reached, what is the appropriate scientific and engineering attitude? Is it confession of failure and no action, or quantification of uncertainty and risk through stochastics? Are current stochastic methods consistent with observed natural behaviours? Is there potential to improve current stochastic methods in hydrology? Can deterministic methods provide solid scientific grounds for water resources engineering and management? Are there physical upper and lower limits (deterministic envelopes) in extreme hydrological phenomena, such as precipitation and flood, whose determination could constitute the basis of hydrological design? Can there be risk-free hydraulic engineering and water management?

In the last decades, financial support for research in hydrology and water resources engineering and management has been strongly linked to research on climate, a practice that does not favour hydrology as an autonomous scientific discipline. Therefore, questions related to climate research and the relationship between hydrology and climate are quite important: Are deterministic climate models a proper means to assessing future climate and are there alternatives to them? Is the current interface of climate and hydrology satisfactory? Should hydrology and water resources planning rely on climate model outputs? Are climate models properly validated and is the meaning of model validation and prediction the same in hydrology and in climate modelling? Is the evolution of climate and its impacts on water resources deterministically predictable? With respect to the last questions, we can observe that climate modellers do not hesitate to offer arbitrarily long predictions, with time horizons from 2100 AD (Battisti and Naylor, 2009), to 3000 AD (Solomon et al., 2009), to 100 000 AD (Shaffer et al., 2009) – to mention a few of the most recent publications in the most reputable journals. Given the definition of climate (detailed above) as a long-term average state, the behaviours illustrated by the toy model are quite relevant. The high Hurst exponents estimated in several instrumental and proxy climatic time series, especially for temperature ($H > 0.90$; Cohn and Lins, 2005; Koutsoyiannis and Montanari, 2007; Koutsoyiannis et al., 2009), support the view of most climate processes as persistent ones and, hence, far more unpredictable than a purely random process. This raises the question: Is there any

indication that climate is predictable in deterministic terms? For the latter question, we can observe that the last numerical example in Sect. 5 is relevant to climate ($k=30$ as typically assumed in climate studies and $H=0.95$, a value that is not uncommon for annual temperature). The resulting very little reduction of uncertainty as we move from the annual to the 30-year scale indicates that the standard perception of climate as less variable and less uncertain, in comparison to the annual fluctuation, may be flatly wrong.

A number of philosophical questions related to the exposed view of uncertainty and its implications on society, economy and life could be relevant to this study, but I refrain from presenting here. However, as scientific texts often emphasize the negative aspects of uncertainty or generally imply a negative meaning of it, it may be useful to refer to it as a positive quality. As demonstrated here, uncertainty is tightly related to change and, hence, to creativity, evolution, and progress, in life and in science. As far as science is concerned, recognizing and studying uncertainty, wherever there is uncertainty, should be interpreted not as a pessimistic conceding of defeat, but just as an important scientific task, consistent with the aim of science for the pursuit of truth.

7 Conclusions

The following summarizing questions can represent the conclusions of this article:

- Can natural processes be divided in deterministic and random components?

- Are probabilistic approaches unnecessary in systems with known deterministic dynamics?

- Is stochastics a collection of mathematical tools, unable to give physical explanations?

- Are deterministic systems deterministically predictable in all time horizons?

- Do stochastic predictions disregard deterministic dynamics in all time horizons?

- Can uncertainty be eliminated (or radically reduced) by discovering a system's deterministic dynamics?

- Does positive autocorrelation (i.e. dependence) improve long term predictions?

- Are deterministic predictions of climate possible?

- Are the popular climate "predictions" or "projections" trustworthy and able to support decisions on water management, hydraulic engineering, or even "geoengineering" to control Earth's climate?

The most common answer to all these questions is "yes". Hopefully, the above discourse explained why my answers to all of them are "no".

References

Battisti, D. S. and Naylor, R. L.: Historical warnings of future food insecurity with unprecedented seasonal heat, Science, 323, 240–244, 2009.

Bernoulli, J.: Ars Conjectandi, Thurnisii fratres, Basel, 306+35 pp., 1713.

Birkhoff, G. D.: Proof of the ergodic theorem, Proc. Nat. Acad. Sci., 17, 656–660, 1931.

Chaitin, G. J.: Randomness and mathematical proof, Sci. Am., 232(5), 47–52, 1975.

Chaitin, G. J.: How real are real numbers?, http://arxiv.org/abs/math.HO/0411418, last access: March 2010, 2004.

Cohn, T. A. and Lins, H. F.: Nature's style: Naturally trendy, Geophys. Res. Lett., 32, L23402, doi:10.1029/2005GL024476, 2005.

Duncan, M. J.: Orbital stability and the structure of the solar system, in 1994, in: Circumstellar Dust Disks and Planet Formation, Proceedings of the 10th IAP Astrophysics Meeting, Institut d'Astrophysique de Paris, edited by: Ferlet, R. and Vidal-Madjar, A., Editions Frontiers, Gif sur Yvette, France, 245–256, 1994.

Feynman, R.: The Character of Physical Law, MIT Press, Cambridge, MA, 1965.

Grassberger, P. and Procaccia, I.: Characterization of strange attractors, Phys. Rev. Lett., 50(5), 346–349, 1983.

Heisenberg, W.: Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik, Z. Phys., 43, 172–198, doi:10.1007/BF01397280, 1927, English translation by: Wheeler, J. A. and Zurek, H., Quantum Theory and Measurement, Princeton Univ. Press, 62–84, 1983.

Hurst, H. E.: Long term storage capacities of reservoirs, Trans. Am. Soc. Civil Engrs., 116, 776–808, 1951.

Jaynes, E. T.: Information theory and statistical mechanics. Phys. Rev., 106(4), 620–630, 1957.

Jaynes, E. T.: Probability Theory: The Logic of Science, Cambridge Univ. Press, 728 pp., 2003.

Kolmogorov, A. N.: Grundbegrijfe der Wahrscheinlichkeitsrechnung, Ergebnisse der Math. (2), Berlin, 1933; 2nd English Edition: Foundations of the Theory of Probability, 84 pp. Chelsea Publishing Company, New York, 1956.

Kolmogorov, A. N.: Wienersche Spiralen und einige andere interessante Kurven in Hilbertschen Raum, Dokl. Akad. Nauk URSS, 26, 115–118, 1940.

Kolmogorov, A. N.: On tables of random numbers, Sankhyā Ser. A, 25, 369–375, 1963.

Kolmogorov, A. N.: Three approaches to the definition of quantity of information, Probl. Inf. Transm., 1(1), 1–7, 1965.

Kolmogorov, A. N., and Uspenskii, V. A.: Algorithms and randomness, Theor. Probab. Appl., 32, 389–412, 1987.

Koutsoyiannis, D.: On the quest for chaotic attractors in hydrological processes, Hydrolog. Sci. J., 51(6), 1065–1091, 2006.

Koutsoyiannis, D. and Montanari, A.: Statistical analysis of hydroclimatic time series: Uncertainty and insights, Water Resour. Res., 43(5), W05429, doi:10.1029/2006WR005592, 2007.

Koutsoyiannis, D., Montanari, A., Lins, H. F., and Cohn, T. A.: Climate, hydrology and freshwater: towards an interactive incorporation of hydrological experience into climate research – DISCUSSION of "The implications of projected climate change for freshwater resources and their management", Hydrolog. Sci. J., 54(2), 394–405, 2009.

Laplace, P.-S.: Théorie Analytique des Probabilités / par M. le Comte Laplace. 1812. Oeuvres complètes de Laplace VII. 1-2. Publiées sous les auspices de l'Académie des sciences, par MM. les secrétaires perpétuels, Paris, Gauthier-Villars, 1878–1912.

Laskar, J.: A numerical experiment on the chaotic behaviour of the Solar System, Nature, 338, 237–238, 1989.

Laskar, J.: The limits of Earth orbital calculations for geological time-scale use, Philos. Tr. R. Soc. Lond. A, 357, 1735–1759, 1999.

Lasota, A. and Mackey, M. C.: Chaos, Fractals and Noise, Springer-Verlag, New York, 1994.

Lissauer, J. L.: Chaotic motion in the Solar System, Rev. Mod. Phys., 71(3), 835–845, 1999.

Mandelbrot, B. B. and van Ness J. W.: Fractional Brownian Motion, fractional noises and applications, SIAM Review, 10, 422–437, 1968.

Maxwell, J. C.: A letter to Lewis Campbell, 1850; quoted in: Campbell, L. and W. Garnett, The Life of James Clerk, Maxwell, London, 1882, Digitally reproduced by Sonnet Software, Inc., p. 80, 1997.

Metropolis, N. and Ulam, S.: The Monte Carlo method, J. Am. Stat. Assoc., 44(247), 335–341, 1949.

Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods, Society for Industrial and Applied Mathematics, Philadelphia, 1992.

Papoulis, A.: Probability, Random Variables, and Stochastic Processes, 3rd ed., McGraw-Hill, New York, 1991.

Poincaré, H.: Science et Méthode, 1908; reproduced in Chance, The World of Mathematics, Simon & Schuster, New York, p. 1382, 1956.

Popper, K.: Quantum Theory and the Schism in Physics, Routledge, London, 230 pp., 1992.

Robert, C. P.: The Bayesian Choice, From Decision-Theoretic Foundations to Computational Implementation, Springer, New York, 2 edn., 602 pp., 2007.

Robertson, H. S.: Statistical Thermophysics, Prentice Hall, Englewood Cliffs, NJ, USA, 1993.

Shaffer, G., Olsen, S. M., and Pedersen, J. O. P.: Long-term ocean oxygen depletion in response to carbon dioxide emissions from fossil fuels, Nat. Geosci., 2, 105–109 doi:10.1038/NGEO420, 2009.

Shiryaev, A. N.: Kolmogorov, Life and creative activities, The Annals of Probability, 17(3), 866–944, 1989.

Solomon, S., Plattner, G.-K., Knutti, R., and Friedlingstein, P.: Irreversible climate change due to carbon dioxide emissions, Proc. Natl. Acad. Sci., 106(6), 1704–1709, 2009.

Stowe, K.: Thermodynamics and Statistical Mechanics, 2nd ed., Cambridge Univ. Press, Cambridge, UK, 2007.

Takens, F.: Detecting strange attractors in turbulence, in: Dynamical Systems and Turbulence, edited by: Rand, D. A. and Young, L.-S., Lecture Notes in Mathematics no. 898, Springer-Verlag, New York, USA, 336–381, 1981.

US Global Change Research Program: Climate Literacy: The Essential Principles of Climate Sciences, Second Version, March 2009, www.climatescience.gov/Library/Literacy/default. php, last access: March 2010, 2009.

Whitney, H.: Differentiable manifolds, Ann. Math., 37, 645–680, 1936.

www.hydrol-earth-syst-sci.net/14/585/2010/

Hydrol. Earth Syst. Sci., 14, 585–601, 2010