Hydrol. Earth Syst. Sci. Discuss., 6, C2733–C2745, 2009
www.hydrol-earth-syst-sci-discuss.net/6/C2733/2009/

Hydrology and Earth System Sciences Discussions

Interactive comment on "HESS Opinions "A random walk on water"" by D. Koutsoyiannis

S. Weijs

s.v.weijs@tudelft.nl

Received and published: 25 November 2009

Steven V. Weijs

25 November 2009

The paper “A random walk on water” by D. Koutsoyiannis discusses determinism and randomness as two coexisting components of natural processes in time, identifying stochasticity with unpredictability. This view is said to contradict the traditional notion of determinism and randomness as two mutually exclusive components, associated with understanding, science, cause-effect on the one hand and noise and evil on the other hand. The points are demonstrated using a toy-model of a two dimensional (2D) caricature hydrological system with chaotic dynamics. I think the paper by Koutsoyiannis raises some important points regarding predictability at long time horizons and the absence of a clear distinction between deterministic and stochastic phenomena. I enjoyed reading it a lot.

Reading the paper crystallized some of my thoughts about the way we make predictions in hydrology. To my opinion, the example given is not describing the main cause for uncertainty in natural systems we as hydrologists usually deal with. For this kind of systems, the high dimensionality is a more natural explanation of uncertainty than chaotic interaction in low-dimensional systems. I have tried to clarify these views and contrast them to the notions brought forward by Koutsoyiannis. Apart from that, I have a few points I would like to add to the discussion, regarding the real-number argument, the principle of maximum entropy and the definition of understanding. This comment concludes with an extra argument for using probabilistic forecasts, based on a decom-

HESSD

6, C2733–C2745, 2009

position of an information-theoretical measure of forecast quality.

1 randomness vs. determinism

I Agree that what we define as randomness is mainly epistemological and can be seen as unpredictability. In other words, uncertainty is not a physical property of the system but just reflects lack of knowledge. However, current views on quantum mechanics actually say there also is fundamental uncertainty. The way classical mechanics emerges as a special case of quantum mechanics can thus be seen as "Emergence of determinism from randomness", in contrast to "Emergence of randomness from determinism (6616:3). From this apparent determinism, apparent randomness can emerge again (cf. the roll of dice, a coin toss). Maybe we just have to conclude that randomness and determinism can emerge from eachother and which one dominates depends on scale in general and not just on time, as is stated in (6612:14-15).

Actually, it does not matter whether the world it fundamentally deterministic or stochastic (fundamental uncertainty). In practical applications, we deal with more complex systems of interactions and enormous state-dimension than we could ever calculate, so epistemological uncertainty dominates the fundamental uncertainty.

2 Real numbers vs. quantized variables

It is true that it is fundamentally impossible to calculate real numbers with infinite precision (6619:12), but I doubt if the example states given in the caricature system are really real numbers, as is stated in (6619:13). The amount of soil water x_i can be quantized by the number of molecules present in the discrete volume and the flux v_i is also quantized by the number of molecules leaving the volume in one discrete timestep.

HESSD

6, C2733–C2745, 2009

Therefore, I think that the premise of incomplete precision is not something fundamental, but more of practical nature in this case. This is actually the case in many systems, as fundamental building blocks are often quantized variables. I think it is not necessary to bring in the real-number argument for showing that we can never precisely know the initial state of a system.

3 Chaotic dynamics or large state dimension?

The paper focuses on chaotic dynamics and incomplete precision in the knowledge of the initial state as main causes for unpredictability. However, I think it is more natural to see the large state dimension of natural systems as a fundamental reason for unpredictability and randomness. The paper presents it as somewhat surprising that randomness emerges from determinism, even with precisely formulated system dynamics (6618:20-26) and without the introduction of an explicit "agent" of randomness (6618:2-4). In contrast to this, I would rather see hydrological systems as high-dimensional complex systems, with surprising predictable macroscopic behaviour. The nature of this "emergence of predictability" is elaborated below.

The caricature system in the example is one of few states and complex nonlinear interactions. This leads to chaotic system behaviour, which in it's turn leads to unpredictability. It is then stated that "All alive natural systems behave in more or less this way" (6621:1). In contrast to this, I think most natural systems, including the one that the caricature system represents, are most accurately discribed as systems of a high dimensional state and relatively simple interactions (far simpler than in eq. 2). The 2D system can be seen as an approximation of the emergent behaviour of such a complex system. The large number of interactions and feedbacks in the complex system make it even more chaotic and unpredictable than the 2D complex interactions system.

However, the macro-states, which are for example sums or averages such as the water

HESSD

6, C2733–C2745, 2009

storage and vegetation cover, are far more predictable than the micro-states, such as the position of all water molecules and the activity of the individual stomata in the vegetation leaves. In this case one could speak of “emergence of (some degree of) determinism from randomness” or the “emergence of predictability from randomness”. I think this very fundamental mechanism of emergence, both visible in evolutionary processes and the movement of systems towards maximum entropy, is the reason why we can make hydrological predictions in the first place.

Even though it is impossible to predict the paths of individual water molecules with accuracy, the macro-states of a large number of molecules interacting is surpisingly predictable. Sometimes this predictability simply follows from the calculus of probabilities. Given a large number of equally probable microstates, probability in a complex system often concentrates in a small number of possible macrostates. An example is the sum of the outcomes of a large number of dice. The uncertainty about the precise microstate is equal to the sum of uncertainties of the individual dice, but the uncertainty about the sum is far less.

If we recognize that the states used in the caricature system are in fact macro-states of a more complex system, then we do not need the incomplete precision in describing real numbers as an argument for not knowing the initial state precisely. The incomplete precision simply follows from random fluctuations of the sum of “a very large number of dice”. Or, as Grandy Jr (2008) puts it: “Effects of the microscopic dynamical laws can only be studied at the macroscopic level by means of probability theory”.

4 The principle of maximum entropy

In this section of the commentary, the connection between maximum entropy thermodynamics and the more general principle of maximum entropy as a tool for inference of a distribution is explored. The key observation is that principle of maximum entropy

HESSD

6, C2733–C2745, 2009

as used by Koutsoyiannis (6625:12-6626:2) is not in itself a justification for the normal distribution of the probabilistic forecast at long time horizon. I am curious if there would be some explanation for the constraints on mean and variance, which reflect the information available about the state at long lead time.

Because of the large number of interacting particles in a full microscopic representation of a hydrological system, it is also practically impossible to make predictions about the precise full state even one timestep ahead. The only reason we can make predictions about future behaviour of the world is the fact that patterns emerge from these complex interactions (the most simple pattern being effects that average out, such as the sum of dice). The patterns are far more predictable than the microstates themselves. This is the result of constraints on the macro-states that follow from conservation laws and quantities that are measurable on macro-scale. The most likely probability distribution of the micro-states is the one that maximally spreads out probability, given these constraints on macro-states, which are usually expectation values of functions of microstates. This principle is related to the fact that the most likely distribution of micro-states is the one that can be realized in the largest number of ways, satisfying constraints on macroscopic quantities, like the total energy in the system. A well-known example is the pressure of gas in the atmosphere, which follows a exponential distribution. This matches the maximum entropy distribution for a given mean and positive values. The mean can be interpreted as a fixed potential height energy and the restriction to positive values as the boundary of the earth's surface. Whether the distribution is a good estimate depends on whether the known macrostates are sufficient to characterize the system. The distribution thus reflects a state of knowledge contained in the constraints on the macrostates.

The principle of maximum (Shannon-)entropy, formulated by Jaynes (1957), generalizes this to a method of statistical inference. It states that the best estimate for a distribution of some variable is the one that has maximum entropy (uncertainty), constrained by what we do know. The principle of maximum entropy distributions is very

HESSD

6, C2733–C2745, 2009

applicable to the case of forecasting behaviour of large numbers of interacting particles or processes and probably also for forecasting the probability distribution of the state after a large number of timesteps.

However, a maximum entropy distribution should always be accompanied by constraints. In other words, simply maximizing uncertainty makes no sense, if we do not constrain it by the information we do have. If we do not have any information and maximize entropy, we end up with the a uniform distribution on $(-\infty, \infty)$.

In the example given (6625:12-22), the principle of maximum entropy is used as a theoretical reason for the distribution becoming Gaussian after some time. However, in itself, the principle is no justification for the Gaussian distribution. Also the fact that the mean and variance are fixed should be justified theoretically (cf. the constraints on the lower bound and mean for the exponential distribution for gas in the atmosphere). In other words, it must be shown that the mean and standard deviation sufficiently characterize the system to make a correct inference about distribution of micro-states.

Another point regarding the maximum entropy principle is the difference between the tendency to thermodynamic equilibrium and the principle of maximum entropy as a method of inference based on limited information. Although links exist, which are not always straightforward, they are different concepts. I think the forecast distribution of the state at long time-horizons is a typical example of inference with incomplete information and I do not see an obvious connection with the concept of a thermodynamic equilibrium of maximum entropy (as suggested in 6626:10). The thermodynamic equilibrium is this system is only reached when it is in the dead state. The reason why the system remains alive is the fact that it is not isolated and low-entropy (free) energy flows in (sunlight). Although the energy is not explictly modeled, this is probably the only way to explain the chaotic dynamics in equation 2 (page 6617).

HESSD

6, C2733–C2745, 2009

5 Prediction vs understanding

On (6618:27-29) the following statements are made: “Science is not identical to understanding” and “Nobody understands quantum mechanics”. I fully agree with the two statements, but for different reasons, because I would like to propose another view on defining understanding. I think this definition of understanding is closer to what Feynman was arguing in the second quote. At this point, I do think the distinction between understanding and overstanding, as brought forward by Koutsoyiannis (6618:15 - 6619:9), is another issue here. That distinction refers to making inferences based on observed emergence behaviour versus deduction from first principles. However, the first principles are also inferred from observed behaviour. The difference is just a matter of scale and uncertainty and both are valuable tools for improving predictions. I therefore think that even the statement that “prediction ... is a crucial target in science-with even higher importance in engineering” (6619:7-9) does understate the importance of predictions in science. I would say that prediction is the fundamental goal and understanding is just a means. This is clarified in the following section, along with some discussion of the relation between prediction and data-compression.

I think science is nothing more and nothing less than trying to make good predictions. The way we can make predictions is to see patterns in what we observe and assume these patterns are the result of some physical laws that are generally applicable. If we describe the pattern in a way longer than necessary, including more parameters to be estimated, we miss part of the pattern, because the description allows for a wider range of patterns to be represented and extra information is necessary to specify which applies to the future.

Therefore, the best way to make predictions is to find some form of minimum description length (MDL) for all observations (model + data). Formal approaches to the MDL principle can be found in Rissanen (2007). In other words, we try to codify observed behaviour in laws as much as possible (leave little noise) and a shorter code is preferred

HESSD

6, C2733–C2745, 2009

over a larger one. Laws should therefore be as general as possible and re-use of laws for various problems is encouraged. This is view is strongly related to data compression, information theory and model complexity control (see for example Schoups et al. (2008)). In science, a good set of models is the set of models that best describes all observations so far and with the smallest total complexity (there is a trade-off between describing data and limiting complexity). This set of models coincides with the one that makes the best predictions, given our state of knowledge. Whenever it is possible to unify two parts of physics into one and it yields a shorter MDL, progress has been made. Sometimes this progress is visible in terms of prediction of thus far unobserved phenomena (this progress is cashed at the moment when it is observed, for example if the Higgs Boson is found). Another way to advance science is to observe phenomena that cannot be explained by the current set of models, forcing the models to become more complex in order to make good predictions of the phenomena causing these new observations.

Of course, scientific progress is not limited to fundamental physics and explaining the unexplained. Sometimes the emergent behaviour of a complex system can be described in a much shorter way than the reductionist explanation, avoiding the need to specify the full micro-state. Especially because the full micro-state is impossible to observe anyway, it has no value trying to predict it. Modeling relations between macro-states is what we mainly try to do in hydrology (referred to by Koutsoyiannis as 'overstanding'), but in a way also in Newtonian mechanics (With the difference that it almost perfectly describes the emergent behaviour in many conditions).

If science is just about compressing observations to get good predictions, where is the understanding? I think that what is usually seen as understanding is nothing more than seeing analogies between the mathematical relations that give good predictions and our intuition based on observations in everyday life. We intuitively understand conservation of mass because we see it everyday. We understand the movement of molecules in a gas because it is analogous to bouncing marbles in some way. We

HESSD

6, C2733–C2745, 2009

understand the concept of waves by picturing the movement of ripples when we throw a stone in the water.

Understanding is thus nothing more than picturing the predictive model in terms of similar relations in our observable world. This is also the reason why "Nobody understands quantum mechanics" (Feynman, 1965). If the way the world behaves at a certain scale does not have analogous counterparts on a human-observable scale, it is impossible to understand, given this notion of understanding.

Furthermore, I think that understanding is often overrated. It is often stated as an objective as such. From an aesthetic point of view it is of course nice if a model is understandable in the sense that it has analogies to observable behaviour on a human scale. An example is to picture a catchment as a series of interconnected buckets. However, in many cases, the system we try to model simply does not behave in such a way. In those cases, the understandable model structure compromises prediction accuracy and is not closer to how the actual system works than a black-box model fitted to the data. The advantage of conceptual models is that knowledge that has been gained from past observations, like conservation of momentum and mass, is easily added to the model in the form of constraints on the structure. Data that has been previously transformed into knowledge of laws is helping predictions. The fact that we have an extra constraint on the structure helps the model to learn more from the information in the data and improves predictions. So again, the overall goal of good predictions already captures the benefits of physically understandable parameters. Also, by keeping the formulation of the relations for prediction restricted to the known physical laws, we do not unnecessarily extend the description length with extra relations, that are only usable in one specific hydrological system.

Another distinction that can be made is the prediction of simple systems of few states (like in particle physics) and prediction of the macro-states in far larger systems (like in hydrology). In the paper by Koutsoyiannis this difference is referred to as the difference between understanding and "overstanding". Indeed this is an important distinction,

HESSD

6, C2733–C2745, 2009

because in the former case we aim at perfect predictions, while in the latter, we know beforehand that our model gives approximate predictions of macroscopic behaviour of complex systems. In this case, the understanding of deterministic dynamics does not get us far if we are to predict future behaviour.

In these cases intuitive understanding (through "overstanding") can sometimes still be achieved if one realizes that there are analogies with many systems in nature where an "intention" of the system emerges from randomness. Examples are adaptations and optimality that emerge from evolution and maximum entropy distributions that emerge from microscopic randomness, deterministic dynamics and macroscopic constraints. Especially under idealized assumptions, the relations between the macro-quantities can sometimes be of the same form as some of the relations between micro-quantities, which even enhances the idea of understanding.

However, it is dangerous to generalize this kind of understanding. Heterogeneity, for example, can completely change the relation between the macro-states into a form that has no relation with similar processes on micro-scale. Fitting a model structure that still assumes that the form of the relation is "understandable" in terms of simple mechanics will yield bad predictions and thus is bad science.

6 Why we should not make deterministic predictions

It is stated that quantification of uncertainty is a useful target (6615:14). I think it is important here to make a formal distinction here between quantification of uncertainty, which is putting a number to the amount of uncertainty, and "probabilization" of uncertainty, which is specifying the distribution. The quantification of uncertainty within the framework of probability theory amounts to calculating the entropy of the distribution that reflects both our knowledge and our uncertainty about the variable under question. I think the real target is finding that distribution, which is the same as making a proba-

HESSD
6, C2733–C2745, 2009

bilistic prediction. In fact there is a relation between specifying the forecast distribution as accurately as possible and minimizing the uncertainty about the truth. Below, this relation is given for the case of forecasts of an event with a discrete number of possible outcomes.

$$\frac{1}{N}\sum_{t=1}^{N}\mathrm{D}_{KL}(o_t||f_t) = \frac{1}{N}\sum_{k=1}^{K}n_k\mathrm{D}_{KL}(\bar{o}_k||f_k) - \frac{1}{N}\sum_{k=1}^{K}n_k\mathrm{D}_{KL}(\bar{o}_k||\bar{o}) + \mathrm{H}(\bar{o}) \qquad (1)$$

In which N is the number of forecast-observation pairs, K the number of unique forecasts, n_k the number of forecasts of one unique type and o_t, $\bar{o}$, $\bar{o}_k$ and f_k are probability mass functions of respectively one single observation, the average observation (climate), the conditional observation given one particular forecast and the forecast distribution.

The remaining uncertainty about the truth can be measured in hindsight by the Kullback-Leibler divergence (relative entropy) from the observations (assuming they are perfect) to the forecast distribution (the left hand side (LHS) of the equation). This remaining uncertainty can be mathematically decomposed into three components (Weijs et al., 2009). The first term on the RHS is measures the divergence (in terms of information) between the forecast distribution and the real conditional distribution of the variable, given the information on which the forecast was based. Although not formulated in information-theoretical terms before, this concept is known in meteorological forecast verification as reliability, although a better term would be unreliability. One can easily see that in order to minimize the LHS, the first divergence term should be as small as possible. The second term is expectation of the information distance from the conditional distributions given a forecast to the climate. This expectation is also known as the mutual information between the forecasts and the observations. The last term is the uncertainty (Shannon-entropy) of the climatological distribution, which measures the uncertainty knowing only past observations and assuming ergodicity.

HESSD

6, C2733–C2745, 2009

From the decomposition we can see that the uncertainty about the outcome is first determined by the average properties of the past data (3rd term). Then it can be reduced by using the mutual information between the actual state and the future state (2nd term) and some of this information is lost again by not properly converting it into a probability estimate. This information-loss (1st term) is measured as the average divergence of the forecasted distribution to the conditional distribution, given the information available. In deterministic forecasts, this reliability-term goes to infinity for a non-perfect forecast, indicating an infinite information loss. This can be seen as the penalty for pretending to be certain in an uncertain world. In other words, deterministic forecasts give an infinite information loss and therefore increase uncertainty about the truth in a information-theoretical sense. We are currently working on a paper further exploring this issue.

References

W.T. Grandy Jr. *Entropy and the time evolution of macroscopic systems*. Oxford University Press, USA, 2008.

E.T. Jaynes. Information theory and statistical mechanics. II. *Physical review*, 108(2):171–190, 1957.

J. Rissanen. *Information and complexity in statistical modeling*. Springer Verlag, 2007.

G. Schoups, NC Van de Giesen, and HHG Savenije. Model complexity control for hydrologic prediction. *Water Resources Research*, 44(null), 2008.

S. V. Weijs, R. R. P. van Nooyen, and N. C. van de Giesen. Kullback-leibler divergence as a forecast skill score with classical reliability–resolution–uncertainty-decomposition. *Monthly Weather Review*, ,submitted, 2009.

HESSD

6, C2733–C2745, 2009