



HAL
open science

The court of miracles of hydrology: can failure stories contribute to hydrological science?

Vazken Andréassian, Charles Perrin, Éric Parent, Andras Bárdossy

► To cite this version:

Vazken Andréassian, Charles Perrin, Éric Parent, Andras Bárdossy. The court of miracles of hydrology: can failure stories contribute to hydrological science?. *Hydrological Sciences Journal*, 2010, 55 (6), pp.849-856. 10.1080/02626667.2010.506050 . hal-01197604

HAL Id: hal-01197604

<https://hal.science/hal-01197604v1>

Submitted on 31 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

This article was downloaded by: [AgroParisTech]

On: 18 February 2013, At: 03:28

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Hydrological Sciences Journal

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/thsj20>

The Court of Miracles of Hydrology: can failure stories contribute to hydrological science?

Vazken Andréassian^a, Charles Perrin^a, Eric Parent^b & Andras Bárdossy^c

^a Cemagref, Hydrosystems and Bioprocesses Research Unit, Parc de Tourvoie, BP 44, F-92163, Antony Cedex, France

^b AgroParisTech - ENGREF, 19 Avenue du Maine, F-75732 Paris Cedex 15, France

^c Universität Stuttgart, Institut für Wasserbau, Universitätsbereich Vaihingen, Pfaffenwaldring 61, D-70569, Stuttgart, Germany

Version of record first published: 20 Aug 2010.

To cite this article: Vazken Andréassian, Charles Perrin, Eric Parent & Andras Bárdossy (2010): The Court of Miracles of Hydrology: can failure stories contribute to hydrological science?, Hydrological Sciences Journal, 55:6, 849-856

To link to this article: <http://dx.doi.org/10.1080/02626667.2010.506050>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

EDITORIAL

The Court of Miracles of Hydrology: can failure stories contribute to hydrological science?

Vazken Andréassian¹, Charles Perrin¹, Eric Parent² & Andras Bárdossy³

¹Cemagref, Hydrosystems and Bioprocesses Research Unit, Parc de Tourvoie, BP 44, F-92163 Antony Cedex, France
vazken.andreassian@cemagref.fr

²AgroParisTech – ENGREF, 19 Avenue du Maine, F-75732 Paris Cedex 15, France

³Universität Stuttgart, Institut für Wasserbau, Universitätsbereich Vaihingen, Pfaffenwaldring 61, D-70569 Stuttgart, Germany

Citation Andréassian, V., Perrin, C., Parent, E. & Bárdossy, A. (2010) Editorial – The Court of Miracles of Hydrology: can failure stories contribute to hydrological science? *Hydrol. Sci. J.* 55(6), 849–856.

Abstract In this article, we suggest that giving greater prominence to the analysis of failures and errors would more fruitfully advance the hydrological sciences. As widely recognised by philosophers of science, we can all learn from our mistakes, and errors can lead to discovery if they are properly diagnosed. However, failure stories are very seldom communicated and published, even though they represent the bulk of the results obtained by researchers and modellers. This article is the result of passionate discussions held in a workshop called the Court of Miracles of Hydrology held in Paris in June 2008. The participants had been invited to present their unpublished experience with what could be called monsters, anomalies, outliers and failures in their everyday practice of hydrology. The review of these studies clearly shows that in-depth analysis of these observations and results that deviate from the expected norm blazes a trail that can only lead to progress.

Key words hydrology; anomaly; outlier; failure; error; negative result

La Cour des Miracles de l'Hydrologie: les histoires d'échecs peuvent-elles contribuer à la science hydrologique?

Résumé Dans cet article, nous suggérons que d'accorder une plus large place à l'analyse des échecs et erreurs dans les sciences hydrologiques fournirait une voix féconde pour progresser. Comme cela est largement reconnu par les philosophes des sciences, nous pouvons apprendre de nos fautes, et nos erreurs peuvent être une voix vers la découverte si elles sont convenablement diagnostiquées. Cependant, les histoires d'échecs sont rarement communiquées ou publiées, alors qu'elles représentent le gros des résultats obtenus par les chercheurs et les modélisateurs. Cet article est le résultat de discussions passionnées qui se sont tenues lors de l'atelier intitulé la Cour des Miracles de l'Hydrologie de organisé à Paris en juin 2008. Les participants avaient été invités à présenter leurs expériences non publiées avec ce qui pourrait être appelé monstres, horsains et échecs, rencontrés dans leur pratique quotidienne de l'hydrologie. L'examen de ces travaux montre qu'une analyse détaillée de ces observations et résultats qui dévient d'une norme attendue sont réellement une voie pour progresser.

Mots clefs hydrologie; anomalie; horsain; échec; erreur; résultat négatif

CONTEXT AND OBJECTIVES OF THE SPECIAL ISSUE ON THE COURT OF MIRACLES OF HYDROLOGY

Can we learn from our mistakes?

In the preface of his book *Conjectures and Refutations*, in which he expounds his theory of knowledge, Popper (1963) claims that “we can learn from our mistakes” and that “all our knowledge grows only through the correcting of our mistakes.” Popper believes we learn from our mistakes by proposing bold conjectures or guesses as

solutions to the problems raised and, by testing them and eventually refuting them, we learn where we went wrong. In his own words:

I much prefer an attempt to solve an interesting problem by a bold conjecture, *even (and especially) if it soon turns out to be false*, to any recital of a sequence of irrelevant truisms. We prefer this because we believe that in this way we learn from our mistakes; and that in finding that our conjecture was false, we shall have learnt much about the truth, and shall have got nearer to the truth.

This humble attitude is the very basis of modern scientific methodology. It has been discussed by other philosophers of science, including Chalmers (1973), who did not contest the approach itself but rather the definition of what a bold conjecture was and the fact that mistakes were presented as the *only* way for knowledge to grow. For Chalmers, learning from mistakes remains a core scientific value.

As stated in an old Roman saying – *errare humanum est* (it is human to make failures and errors) – taking lessons from failure is at the very heart of the feedback scheme of the learning process. After failure we try hard to understand the source of problems, so that next time we are more likely to succeed. If everybody agrees on this, then why must we dedicate a special issue to mistakes in hydrological sciences? First and foremost, we do so because of the famous truism captured by André Gide: “everything has been said before, but since no one listens, one must always start again.” Second, even if everybody agrees on the diagnosis and seems ready to make public the errors of others, we are much more reluctant to communicate our own errors (this tendency seems rather common with humans; see Matthew’s words in Matt. 7.3).

Passing from conviction to action has never been easy, and hydrology is no exception to this general rule. Failures are seldom communicated in scientific journal publications. Scientific articles focus almost exclusively on *success stories*, i.e. case studies where hydrologists have been able to produce so-called successful model runs. Failure stories, which could be just as instructive as (or even more instructive than) successes, meet much more resistance in the world of scientific publishing.

In everyday scientific practice, model failures are the rule rather than the exception. It takes any number of dead-end explorations before one hits on a suitable way to move forward. We are sometimes tempted to hide model failures, hoping to save face and thus artificially increase the confidence of model users. We either claim that these failures are *anomalous*, or that the data points are *outliers*; sometimes we merely consider the catchment in question to be a hydrological *monster*, isolated because of its strangeness. This is unfortunate because these very anomalies and outliers can help us identify new ways to improve the predictions of our models. By hiding our failures, we miss the opportunity to learn where we went wrong.

What is the “Court of Miracles of Hydrology”?

The workshop that led to this special issue was held at AgroParisTech – ENGREF in Paris on 18–20 June

2008, and was called the Court of Miracles of Hydrology. The name refers to Victor Hugo’s famous novel *The Hunchback of Notre Dame* published in 1831. Originally, under the Ancient Regime (17–18th century), the court of miracles referred to places in Paris where the beggars and thieves of the city lived in hiding. During the day, they went to the prosperous areas of the city to beg and rob, feigning one disability or another to arouse the compassion of the bourgeois. Returning at night, they dropped their beggar costumes, and suddenly the blind could see, the crippled could walk and the lame could run, as if by miracle. Hence, the court of miracles was the place where monsters rejected by society were able to metamorphose into ordinary human beings.

We postulate that such a place may well exist in the domain of hydrology, hiding most of what could be called hydrological monsters or failures that are not revealed and not published, and that this place may turn monsters and failures into new leads for the advancement of hydrology.

The workshop held in Paris (though not in the former court of miracles!) gathered a large panel of hydrologists, experimentalists, modellers, statisticians, researchers and operational end users willing to explore together the dark and narrow streets of our own court of miracles. The objective of this special issue is to present those papers that attempted to uncover, interpret and ultimately fix some of these failures within a sort of court of miracles of hydrological science.

ANOMALIES, MONSTERS, OUTLIERS, ETC.

A few definitions

“Anomalies”, “monsters”, “outliers”: these terms are often associated with errors, mistakes, or unexplained findings arising in the course of research or the modelling process. Allchin (2001) defines four types of error, namely material (e.g. violated experimental protocol), observational (e.g. sampling error), conceptual (inappropriate model or theory) and discursive (i.e. in communicating results). Errors can also be understood in a statistical sense while testing hypothesis H : a Type I error would be to reject H when it is actually true, and a Type II error would be to accept H when it is actually false. Scientists should obviously try to prevent these errors, or at least reduce the frequency of their occurrence.

“Anomaly” is probably the term that generally best describes the apparently unsolvable or inexplicable problems we find in our everyday hydrological practice. An anomaly characterises something that is

different from what is expected, or not in agreement with something else and therefore not satisfactory. Darden (1991) proposed to distinguish *monster* anomalies from *model* anomalies. The first are simply ordinary outliers which do not allow one to seriously modify the structure of the model, while the second suggest that changes are needed at some level in the model's structure. Although here we are interested in the second category, we are often tempted to relegate anomalies to the first category because we lack a satisfactory explanation of them.

The term "monster" defines any frightening creature or person, especially one which is extraordinarily large, strange, hideous, or that deviates grotesquely from a natural or normal form. The adjective "monstrous" has many synonyms, such as bizarre, weird, freakish. These adjectives may be used to define something incongruous, unexpected, deviating from the norm or unpredictable. In absolute terms, they describe something unpredictable, in relative terms, something unexpected, i.e. which deviates from an expected standard (Mathevet & Garçon, 2010).

An outlier is something or part of something situated apart from a main or related body. In statistics, an outlier is an observation that is numerically distant from the rest of the data, that deviates markedly from other members of the sample.

These terms are often used in hydrology when observations or events deviate from an expected value or when the model fails; they carry an exclusively negative connotation, and we are often prone to reject them, since we are reluctant or unable to diagnose the associated errors. However, this diagnostic phase may be an early stage of scientific discovery. Elliott (2004) advocates error probing and proposes seven strategies that could help scientists to characterise and localise anomalies in the ultimate process of eliminating them, advocating that errors should be seen as part of the learning process.

Hydrological "monstrosity": an illustration of how we account for anomalies

To illustrate how anomalies are treated in the hydrological literature, we use the example of regionalisation studies (see Boldetti *et al.*, 2010, for a more complete discussion). In these studies, one tries to transfer information (e.g. model parameters) from gauged (donor) catchments to ungauged (target) catchments to estimate the model parameters on the latter. The quality of the information provided by donors

may play an important role in the outcome. Therefore, when considering the whole set of donor catchments, donors with poor model performance are untrustworthy, as these outliers may provide overly uncertain information. Several strategies can be adopted to handle these outliers:

- (a) donor catchments with model performance under some predefined threshold can be excluded;
- (b) donor catchments can be excluded using a threshold of acceptability on model performance based on sensitivity analysis;
- (c) a weighted approach can weight the better-modelled donors more heavily; and
- (d) potentially undesirable donors can be identified on an objective basis (Boldetti *et al.*, 2010, proposed to define them as catchments that do not fit in their neighbourhood).

Boldetti *et al.* (2010) clearly show that these various treatments of outliers yield significant differences in results. These different ways of treating the outliers are in fact different hypotheses on the reasons for the anomalies. The anomalies may result from actual problems in observations or data, but also the erroneous conceptions upon which we base our models. Discriminating between the various hypotheses requires appropriate tests, and enables us to learn about the nature of information.

HOW COULD WE BETTER LEARN FROM ANOMALIES?

Given the past and present generations of brilliant hydrologists, it is quite difficult here to advise original initiatives to advance the progress of hydrological sciences in terms of scientific questions and technical options. For example, Beven (2002), Kirchner (2006), McDonnell *et al.* (2007), Todini (2007) and, recently, Wagener *et al.* (2010) provided a list of promising leads for observational networks and model developments. Here we suggest that further progress could be made by paying more attention to anomalies. The following aspects deserve more attention:

Implementing severe tests

More demanding and severe tests should be designed to evaluate the reliability of models with the aim of improving them (Andréassian *et al.*, 2009). We follow Mayo (2010) when she argues that:

the growth of knowledge has not to do with replacing or confirming or probabilifying or “rationally accepting” large-scale theories, but with testing specific hypotheses in such a way that there is a good chance of learning something – whatever theory it winds up as part of. This learning, in the particular experimental account we favor, proceeds by testing experimental hypotheses and inferring those which pass probative or severe tests – tests that would have unearthed some error in, or discrepancy from, a hypothesis H, were H false.

We need to implement thorough tests to assess models as hypotheses, as advocated by Beven (2010). This will require defining limits of acceptability or benchmarks that can help reject or accept these hypotheses. Following good modelling practice guidelines for model verification, testing and validation could also help increase the confidence in the results and models we develop and then transfer to end-users (Refsgaard & Henriksen, 2004; Refsgaard *et al.*, 2005).

Repeating experiments

As advocated by Underwood (1999), we should avoid single unrepeated experiments as they increase the risks of Type I and Type II errors in hypothesis testing. Underwood adds that a vast amount of additional information can be provided by repeated experiments and that this helps interpreting failures. In the context of hydrological modelling, avoiding the one-case-study application should become the rule rather than the exception, as also argued by Andréassian *et al.* (2006). This is one convincing way to make our results more generalisable and our models more reliable (Takle *et al.*, 2007).

Developing a culture of publishing failure stories

The publication of failure stories (as stand-alone papers or in papers reporting on both successes and failures) should be seriously considered, possibly even encouraged and better recognised as driving progress, as also argued Bergström (1991) who criticised the “cult of success”. Obviously, such publications should follow the established rules for any publication in terms of scientific rigour. The entire publishing process should be prepared to deal with these contributions: authors need to be willing to show their unsuccessful tests; reviewers and editors need to be willing to accept such articles for publication, provided that they follow the standards of scientific research; and, lastly, authors need

to recognise the utility of citing these articles, which would make them all the more valuable for publication. It is probably unrealistic to think that there could be journals that specialise in failure publishing, as whimsically suggested by Palmer (2007) with his *Journal of Close, But No Cigar*. However, some initiatives have already been undertaken by peer-reviewed journals. For example, in 1999 the *Marine Ecology Progress Series* created a Theme Section called Negative Results. The coordinator recounts the hurdles that had to be crossed in convincing colleagues to contribute to this section, but argues that negative results may be important for several reasons: they can provide more balance in a subject area, indicate that a subject is not as mature as previously suspected, or show that a particular line of research is not worthy of further effort (Browman, 1999). Publication of negative results could also help keep a memory of past mistakes. Allchin (2010) claims that memory of errors is critical to progress in science, and advocates the building of error repertoires at institutional levels. Lastly, Gould (1993) points out that underreporting negative results strongly biases scientific literature towards an impression of efficacy and achieved understanding, whereas “the great bulk of daily scientific work never sees the light of a published day (and who would wish for changes here, as the ever-increasing glut of journals makes keeping up in one’s own field impossible and exploration of others inconceivable?)”. For example, Beven (2001), discussing the case of distributed modelling, states that there was little reporting of cases where models have failed in their spatial predictions over the past 30 years, whereas the progress made in the meantime may be found limited to some respect. This underreporting also unconsciously limits researchers’ initiatives to try to publish results and reviewers to accept them.

Developing post-audit and post-evaluation

In most cases, modelling studies are never post-audited or post-evaluated although it would be very useful. In operational conditions, evaluating the performance of models after a series of events or after a few years can help to point out model deficiencies or mistakes, and possibly show that the model does not meet the expected level of performance shown during the calibration study. Making such an exercise, Kerr (2007) suggests that little progress was made in operational flow forecasting over the past two decades. Post-evaluation should be an explicit stage of the modelling process, as detailed by Scholten *et al.*

(2007). It should help reinforce the confidence of end users. Although it is likely to be more difficult to do in a research context, post-evaluation could be very beneficial. It could help to show which directions truly produced advances and which proved to be dead ends. In this respect, the post-evaluation of his own work made by Refsgaard (2007), or the critical evaluation of the TOPMODEL developments made by Beven (1997), should serve as examples for hydrologists. Comparative analysis and the use of reference data sets (though not very easy to implement in practice as new types of data emerge) could help in measuring more objectively the actual improvements made in modelling over several decades, as advocated by Andréassian *et al.* (2009).

OVERVIEW OF THIS SPECIAL ISSUE

The 2008 Court of Miracles of Hydrology Workshop was divided into four poster sessions, each introduced by an invited keynote speech. The four sessions discussed: the definition of “monsters” in hydrology; the interpretation of model failures; the treatment of various sources of uncertainty; and decision-making problems. A total of 46 posters were presented, spread over the four sessions. The 20 papers subsequently submitted to *Hydrological Sciences Journal* consisted of extended and revised versions of the workshop presentations. All manuscripts were peer reviewed and revised; three were either rejected, or temporarily withdrawn as reviewers had suggested further work before they could be accepted for publication. This special issue comprises 17 papers, resulting from the four introductory keynote speeches (Krzysztofowicz, 2010; Kuczera *et al.*, 2010; Mathevet & Garçon, 2010; Refsgaard & Hansen, 2010) and 13 posters. In the following, we summarise some of the main outcomes of these studies.

How bizarre is bizarre in fact?

A major source of anomaly in hydrology comes from the observation of hydrological phenomena that can be attributed to various problems related to measurement and spatial representation. Giving a series of examples, Mathevet & Garçon (2010) show that several assumptions made about data are false. They mention that the data are often considered to be a reliable representation of reality, of natural processes, or of catchment spatial fields, whereas many examples show that actually they are not. One good illustration

is the case of precipitation observation in snow-affected areas. Using catchments in Québec, Turcotte *et al.* (2010) describe the problem of simulating spring floods caused by snowmelt. One major problem stems from the estimation of the initial ground-based snow–water equivalent before melting starts. These measurements and the representativeness of sampling could partly explain why some water seems to be missing at the end of the simulated floods. Another good example of observation problems is given by Lang *et al.* (2010), who raise the issue of uncertainty associated with the extrapolation of the rating curve and its impact on the estimation of high discharge values. They show how the uncertainty of the estimation of flood quantiles can be limited with a mixed approach, using hydraulic modelling and a proper statistical description of hydrological data to better define the rating curve.

A priori good-looking and weird catchments: why do they turn into a modeller’s nightmare?

Beyond the problems of observation, many of the limitations of our current knowledge in hydrology come from our mistakes or misconceptions in the modelling processes. Refsgaard & Hansen (2010) introduced this topic with the example of a real-life modelling study that aimed to identify measures to reduce diffuse nitrate pollution in Denmark. This study turned out to be far from an ideal modelling practice protocol because of a number of deficiencies (insufficient funding, overly optimistic expectations, incomplete conceptual understanding) that led to what could be termed a failure. However, the study turned out to be food for thought, and much was learnt from a back analysis. Hence, the authors conclude that such studies should be seen as part of a learning process. Goswami & O’Connor (2010) raise the issue of model verification. They give the example of a study that successfully modelled the rainfall–runoff transformation of the Fergus catchment in Ireland using the SMAR model, but this success later proved to have been achieved “for the wrong reasons”. This highly non-conservative karstic catchment had been well modelled based on a mistake in the model’s structure. Revision of the model made it work even better, this time for the right reasons.

Two articles raise the issue of catchments that are considered outliers because they do not properly fit within the modelling mould, i.e. they are difficult to model. In the case of snow-affected mountainous catchments in France and Sweden, Valéry *et al.* (2010) present a non-dimensional representation of catchment water balance, which can identify potential outliers.

They then show how the identification of elevation–precipitation relationships can help reduce the occurrence of outliers (although it does not suffice to eliminate them all). Boldetti *et al.* (2010) challenge the practice of data cleansing in regionalisation studies, removing catchments where models do not perform well. They propose another approach for discriminating catchments that look weird from those that can actually contribute information in the regionalisation process.

Theoretical considerations may also provide clues to improve our understanding of systems that are apparently prone to generating errors in the modelling process. Romanowicz (2010) proposed a method to estimate baseflow based on a stochastic transfer function approach applied to log-transformed flows. Applied to several “monstrous” catchments in Poland and the UK, the approach is shown to provide more consistent results than other existing approaches and has a theoretically stronger formulation. Schertzer *et al.* (2010) analyse the way precipitation fields are considered. They argue that many problems in the scaling treatment of precipitation fields originate from their being erroneously considered regular although they are most likely singular measures.

There are no hydrological monsters, only models with huge uncertainties

Hydrological modelling is still far from providing truly satisfactory solutions, and great expectations have now arisen for better quantification of the associated uncertainties. Kuczera *et al.* (2010) indicate that when errors (whatever they may be) are treated too simplistically, they can severely limit our ability to model catchments. Using the Bayesian framework (BATEA) on an apparently anomalous catchment, they show that the information contained in rainfall–runoff records may be insufficient to give insight into our prior knowledge of uncertainty sources.

Leviandier (2010) discusses the issue of estimating extreme events and the uncertainties associated with a rainfall–runoff model’s transformation of extreme variables. Based on conditional probabilities, he proposes a more general formulation of the extreme value index which can allow an increase of the index value from rainfall to streamflow that may sometimes be observed. This more general formulation may therefore limit what could be called exceptional events.

Hingray *et al.* (2010) propose a framework to handle the limited information available for the calibration of a semi-distributed model on the upper Rhône

River in the Swiss Alps. They apply a sequential calibration approach based on the hydrological signature. In spite of this cautious approach, the authors conclude that considerable uncertainties remain, partly due to the problem of input estimation in mountainous areas, as mentioned above, as well as existing regulations on this catchment which are difficult to model.

In a similar concern to find adequate balance between several objectives for model parameter calibration, Booij & Krol (2010) evaluate three strategies to account for four different fit measures. They apply the approach to an eight-parameter version of the HBV semi-distributed model on nine sub-catchments of the Meuse basin. They conclude that differences between methods can be considerable, and differences could also be observed between basins.

There are no hydrological monsters, only decision-making issues

Another step that may generate monsters is the way models are evaluated and their outputs communicated and used in decision-making processes. Krzysztofowicz (2010) argues that many of the problems associated with cases that are difficult to model can be addressed systematically and coherently within the mathematical framework of Bayesian forecast–decision theory. After presenting five examples that can be supported by this type of framework, the author argues that hydrological predictions would be more effectively used by decision makers if modellers adopted the point of view of these users to some extent. Modellers often provide too confident a picture of the model’s capabilities. This may come from the way models are evaluated.

Berthet *et al.* (2010a, 2010b) show that the way root-squared criteria are used can be misleading, since, in most cases, it is necessary to have time series that are longer than a few decades to obtain criterion values close to their statistical expectation. Therefore calculating them on a few years of data, as is usually done, gives a partial evaluation of model performance. Moreover, in the context of flood forecasting, they show that most of the total error amount stems from a few time steps where the flow variations are the greatest, which also restricts how informative these criteria are to those few time steps where the model errors are the greatest.

Moussa (2010) also questions the interpretation of the commonly-used Nash-Sutcliffe criterion in the case of event-based modelling. Using the case study of the flashy Gardon d’Anduze catchment in southern France, he shows that a monster catchment may simply originate

from a temporal translation or homothetic transformation of the hydrograph. He also shows that performance interpretation is strongly related to the benchmark used, and seemingly good model performance may turn poor if the benchmark model is changed. Therefore, the author proposes a multi-criteria analysis to prevent misinterpretation of performance criterion values.

Lastly, Mathevet & Garçon (2010) discuss the problems of interpreting model results in an operational flood forecasting context where uncertainties are sometimes difficult for the forecaster to interpret.

CONCLUSION

We hope that this special issue will generate interest and reactions among our colleagues, and that it will stimulate a renewed interest in collectively learning from our mistakes. There are already independent initiatives in the same direction (such as the session convened by L. Pfister, P. Matgen and F. Fenicia at the EGU conference in 2009, entitled “Progress in hydrological sciences: what do we learn from our mistakes?”).

The *Hydrological Sciences Journal* provided this tribune to highlight this important question of publishing so-called failure stories as an impetus for the advancement of science. Quoting James Joyce, we could say that “mistakes are the portals of discovery”. This means that hydrologists should be encouraged to dedicate part of their publication efforts to reporting their mistakes or what can be called negative results. Of course, this requires a humble attitude, but hydrology is definitely a domain of science that makes people humble! As experts of water sciences, hydrologists may recognise themselves in the words of Niels Bohr (Nobel laureate in Physics in 1922), who wrote that “an expert is a person who has made all the mistakes that can be made in a very narrow field.” So it should be feasible to accept our scientific failures and expose them to the scientific community, as was done quite enthusiastically by the participants in the Court of Miracles of Hydrology Workshop.

Should we think that the Court of Miracles of Hydrology is currently depopulated, as argued by Schertzer *et al.* (2010)? On the contrary, we think that there is a wealth of hydrological monsters out there! Of course, there would be no point in replacing the cult of success by the cult of failure: all failures are not equally instructive. All negative results are not relevant, significant and constructive for the research community and they cannot all be turned into positive knowledge (Allchin, 1999). But we do believe that the most

valuable lessons lie in catchments, hydrometeorological situations and extreme events (flood and low flows) that somehow caused unexpected or apparently unsolvable problems in terms of understanding and modelling hydrological behaviour. Too often, model failures have been conveniently attributed to “data errors”. We reject the systematic nature of this approach, for the very reason that it is often an excuse to avoid challenging the structure of models and their underlying hypotheses.

We hope that paying attention to both “right” and “wrong” outcomes will increase the confidence in results, be they positive or negative. What should be contested are not negative results in themselves, but those that are too uncertain or too inconclusive, as is often the case for seemingly “right” results. Therefore, we agree with Allchin (1999), who states that:

thinking must shift from an exclusively right/wrong distinction to include a certain/uncertain distinction. “Positive” knowledge is defined by being certain, not by being either right or wrong. The fundamental aim is reliability. The lesson is key – and one worth instilling in students.

Acknowledgements The authors warmly thank and acknowledge the support, encouragements and feedback of the *HSJ* Co-editors Demetris Koutsoyiannis and Zbigniew W. Kundzewicz, who accepted to edit and publish this special issue, and the *HSJ* Editorial office, especially Cate Gardner for the fruitful discussions in the early stages of this special issue, and Frances Watkins for her constant support in easing the review and editorial process and guiding us in the newly implemented online submission and review system. We also wish to thank all the participants to the workshop for their stimulating discussions and inputs, all the authors who accepted to present and discuss their hydrological monsters in this special issue and all the reviewers for their work on the submitted manuscripts. Finally, our thanks go to AgroParisTech – ENGREF for hosting the workshop and Cemagref and EDF-DTG for their financial support.

REFERENCES

- Allchin, D. (1999) Negative results as positive knowledge, and zeroing in on significant problems. *Marine Ecology Progress Series* **191**, 303–305.
- Allchin, D. (2001) Error types. *Perspectives on Science* **9**, 38–59.
- Allchin, D. (2010) Error repertoires. Submitted manuscript, <http://www.tc.umn.edu/~allch001/papers/err-rep.pdf>.
- Andréassian, V., Hall, A., Chahinian, N. & Schaake, J. (2006) Why should hydrologists work on a large number of basin data sets? In: *Large Sample Basin Experiments for Hydrological Model*

- Parameterization. Results of the Model Parameter Experiment - MOPEX* (V. Andréassian, A. Hall, N. Chahinian & J. Schaake, eds.), 1–5. Wallingford: IAHS Press, IAHS Publ. 307.
- Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M. H. & Valéry, A. (2009) Crash tests for a standardized evaluation of hydrological models. *Hydrol. Earth System Sci.* **13**, 1757–1764.
- Bergström, S. (1991) Principles and confidence in hydrological modelling. *Nordic Hydrology* **22**, 123–136.
- Berthet, L., Andréassian, V., Perrin, C. & Loumagne, C. (2010a) How significant are quadratic criteria? Part 1. How many years are necessary to ensure the data-independence of a quadratic criterion value? *Hydrol. Sci. J.* **55**(6), 1051–1062 (*this issue*).
- Berthet, L., Andréassian, V., Perrin, C. & Loumagne, C. (2010b) How significant are quadratic criteria? Part 2. On the relative contribution of large flood events to the value of a quadratic criterion. *Hydrol. Sci. J.* **55**(6), 1063–1073 (*this issue*).
- Beven, K. (1997) TOPMODEL: a critique. *Hydrol. Processes* **11**(9), 1069–1085.
- Beven, K. (2002) Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrol. Processes* **16**(2), 189–206.
- Beven, K. (2010) Preferential flows and travel time distributions: defining adequate hypothesis tests for hydrological process models. *Hydrol. Processes*, **24**(12), 1537–1547.
- Beven, K. J. (2001) How far can we go in distributed hydrological modelling? *Hydrol. Earth System Sci.* **5**(1), 1–12.
- Boldetti, G., Riffard, M., Andréassian, V. & Oudin, L. (2010) Dataset cleansing practices and hydrological regionalization: is there any valuable information among outliers? *Hydrol. Sci. J.* **55**(6), 941–951 (*this issue*).
- Booij, M. J. & Krol, M. S. (2010) The balance between calibration objectives in a conceptual hydrological model. *Hydrol. Sci. J.* **55**(6), 1017–1032 (*this issue*).
- Browman, H. I. (1999) The uncertain position, status and impact of negative results in marine ecology: philosophical and practical considerations. *Marine Ecology Progress Series* **191**, 301–302.
- Chalmers, A. F. (1973) On learning from our mistakes. *Brit. J. Philos. Sci.* **24**(2), 164–173.
- Darden, L. (1991) *Theory Change in Science*. New York: Oxford University Press.
- Elliott, K. (2004) Errors as means to discovery. *Philosophy of Science* **71**, 174–197.
- Goswami, M. & O'Connor, K. M. (2010) A “monster” that made the SMAR conceptual model “right for the wrong reasons”. *Hydrol. Sci. J.* **55**(6), 913–927 (*this issue*).
- Gould, S. J. (1993) Cordelia’s Dilemma. *Natural History* **102**, 10–18.
- Hingray, B., Schaeffli, B., Mezghani, A. & Hamdi, Y. (2010) Signature-based model calibration for hydrological prediction in mesoscale Alpine catchments. *Hydrol. Sci. J.* **55**(6), 1002–1016 (*this issue*).
- Kerr, R. A. (2007) River-level forecasting shows no detectable progress in 2 decades. *Science* **316**, 1555.
- Kirchner, J. (2006) Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res.* **42**, W03S04, doi: 10.1029/2005WR004362.
- Krzysztofowicz, R. (2010) Decision criteria, data fusion and prediction calibration: a Bayesian approach. *Hydrol. Sci. J.* **55**(6), 1033–1050 (*this issue*).
- Kuczera, G., Renard, B., Thyer, M. & Kavetski, D. (2010) There are no hydrological monsters, just models and observations with large uncertainties! *Hydrol. Sci. J.* **55**(6), 980–991 (*this issue*).
- Lang, M., Pobanz, K., Renard, B., Renouf, E. & Sauquet, E. (2010) Extrapolation of rating curves by hydraulic modelling, with application to flood frequency analysis. *Hydrol. Sci. J.* **55**(6), 883–898 (*this issue*).
- Leviandier, T. (2010) A model capable of changing the heaviness of the tail of a probability distribution function. *Hydrol. Sci. J.* **55**(6), 992–1001 (*this issue*).
- Mathevet, T. & Garçon, R. (2010) Tall tales from the hydrological crypt: are models monsters? *Hydrol. Sci. J.* **55**(6), 857–871 (*this issue*).
- Mayo, D. (2010) Severe testing, error statistics and the growth of theoretical knowledge. In: *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability and the Objectivity and Rationality of Science* (D. Mayo & A. Spanos, eds.), 28–57. Cambridge: Cambridge University Press.
- McDonnell, J. J., Sivapalan, M., Vache, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M. L., Selker, J. & Weiler, M. (2007) Moving beyond heterogeneity and process complexity: a new vision for watershed hydrology. *Water Resour. Res.* **43**(7), W07301, doi:10.1029/2006WR005467.
- Moussa, R. (2010) When monstrosity can be beautiful while normality can be ugly: assessing the performance of event-based flood models. *Hydrol. Sci. J.* **55**(6), 1074–1084 (*this issue*).
- Palmer, J. (2007) Learning from our mistakes. *Physics World* (1), 2.
- Popper, K. (1963) *Conjectures and Refutations*. London: Routledge and Keegan Paul.
- Refsgaard, J. C. (2007) *Hydrological modelling and river basin management*. Doctoral Thesis, Geological Survey of Denmark and Greenland.
- Refsgaard, J. C. & Hansen, J. R. (2010) A good-looking catchment can turn into a modeller’s nightmare. *Hydrol. Sci. J.* **55**(6), 899–912 (*this issue*).
- Refsgaard, J. C. & Henriksen, H. J. (2004) Modelling guidelines – terminology and guiding principles. *Adv. Water Resour.* **27**, 71–82.
- Refsgaard, J. C., Henriksen, H. J., Harrar, W. G., Scholten, H. & Kassahun, A. (2005) Quality assurance in model based water management – review of existing practice and outline of new approaches. *Environ. Model. Software* **20**(10), 1201–1215.
- Romanowicz, R. (2010) An application of a log-transformed low-flow (LTLF) model to baseflow separation. *Hydrol. Sci. J.* **55**(6), 952–964 (*this issue*).
- Schertzer, D., Tchiguirinskaia, I., Lovejoy, S. & Hubert, P. (2010) No monsters, no miracles: in nonlinear sciences hydrology is not an outlier! *Hydrol. Sci. J.* **55**(6), 965–979 (*this issue*).
- Scholten, H., Kassahun, A., Refsgaard, J. C., Kargas, T., Gavardinas, C. & Beulens, A. J. M. (2007) A methodology to support multi-disciplinary model-based water management. *Environ. Model. Software* **22**, 743–759.
- Takle, E. S., Roads, J., Rockel, B., Gutowski, W. J., Arritt, R. W., Meinke, I., Jones, C. G. & Zadra, A. (2007) Transferability intercomparison - An opportunity for new insight on the global water cycle and energy budget. *Bulletin of the American Meteorological Society* **88**(3), 375–384.
- Todini, E. (2007) Hydrological catchment modelling: past, present and future. *Hydrol. Earth System Sci.* **11**(1), 468–482.
- Turcotte, R., Fortier Fillion, T.-C., Lacombe, P., Fortin, V., Roy, A. & Royer, A. (2010) Simulation hydrologique des derniers jours de la crue de printemps: le problème de la neige manquante. *Hydrol. Sci. J.* **55**(6), 872–882 (*this issue*).
- Underwood, A. J. (1999) Publication of so-called “negative” results in marine ecology. *Marine Ecology Progress Series* **191**, 307–309.
- Valéry, A., Andréassian, V. & Perrin, C. (2010) Regionalisation of precipitation and air temperature over high-altitude catchments – learning from outliers. *Hydrol. Sci. J.* **55**(6), 928–940 (*this issue*).
- Wagener, T., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H. V., Kumar, P., Rao, P. S. C., Basu, N. B. & Wilson, J. S. (2010) The future of hydrology: an evolving science for a changing world. *Water Resour. Res.* **46**, W05301.