



HAL
open science

Impact of suspicious streamflow data on the efficiency and parameter estimates of rainfall–runoff models

Cyril Thébault, Charles Perrin, Vazken Andréassian, Guillaume Thirel, Sébastien Legrand, Olivier Delaigue

► To cite this version:

Cyril Thébault, Charles Perrin, Vazken Andréassian, Guillaume Thirel, Sébastien Legrand, et al.. Impact of suspicious streamflow data on the efficiency and parameter estimates of rainfall–runoff models. *Hydrological Sciences Journal*, 2023, 68 (12), pp.1627-1647. 10.1080/02626667.2023.2234893 . hal-04206286

HAL Id: hal-04206286

<https://hal.inrae.fr/hal-04206286>

Submitted on 13 Sep 2023

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.

Copyright



Impact of suspicious streamflow data on the efficiency and parameter estimates of rainfall-runoff models

Cyril Thébault, Charles Perrin, Vazken Andréassian, Guillaume Thirel, Sébastien Legrand & Olivier Delaigue

To cite this article: Cyril Thébault, Charles Perrin, Vazken Andréassian, Guillaume Thirel, Sébastien Legrand & Olivier Delaigue (2023) Impact of suspicious streamflow data on the efficiency and parameter estimates of rainfall-runoff models, Hydrological Sciences Journal, 68:12, 1627-1647, DOI: [10.1080/02626667.2023.2234893](https://doi.org/10.1080/02626667.2023.2234893)

To link to this article: <https://doi.org/10.1080/02626667.2023.2234893>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 14 Aug 2023.



Submit your article to this journal [↗](#)



Article views: 437








View related articles [↗](#)



View Crossmark data [↗](#)

Impact of suspicious streamflow data on the efficiency and parameter estimates of rainfall–runoff models

Cyril Thébault ^a, Charles Perrin ^a, Vazken Andréassian ^a, Guillaume Thirel ^a, Sébastien Legrand^b
and Olivier Delaigue ^a

^aUniversité Paris-Saclay, INRAE, HYCAR, Antony, France; ^bCompagnie nationale du Rhône, Lyon, France

ABSTRACT

Many sources of error in hydroclimatic data can affect hydrological modelling, yet the impact of streamflow data quality is poorly quantified. This work aims to investigate whether inconsistencies found in streamflow time series commonly available for hydrological studies (typically in national streamflow archives) have an impact on the efficiency and the parameter estimates of rainfall–runoff models. Hydroclimatic data were gathered at the hourly time step over the period 1998–2018 for a set of 30 catchments in France. Hydrological modelling was carried out with the lumped conceptual GR5H (standing for modèle du Génie Rural à 5 paramètres Horaire, i.e. Hourly 5-parameter rural engineering model) model. A typology of “realistic” suspicious streamflow was established to set up several error models in order to corrupt the data. Our results suggest that common suspicious streamflow data do not have a strong impact on model efficiency and parameter estimates overall, but may be an important source of instability and lack of robustness when working on a single catchment.

ARTICLE HISTORY

Received 22 July 2022
Accepted 6 June 2023

EDITOR

S. Archfield

GUEST EDITOR

P. Brigode

KEYWORDS

rainfall–runoff modelling;
sensitivity analysis;
streamflow data; random
and systematic errors

1 Introduction

1.1 Errors in streamflow estimates

There are many sources of errors in hydrological modelling. These stem from the choices made in the model structure, the techniques used to estimate model parameters, or the quality of observations used to run or calibrate the model (e.g. meteorological forcing and streamflow time series). Various studies have highlighted the impacts of errors in precipitation and potential evapotranspiration data on modelling, either in terms of model performance or directly on model parameter values (Ibbitt 1972, Troutman 1985, Patrel *et al.* 1995, Andréassian *et al.* 2001, 2004, Oudin *et al.* 2006, McMillan *et al.* 2010b, Singh and Dutta 2017). Recently, a few studies focused on the uncertainties associated with air temperature data, which can be used for the calculation of potential evapotranspiration, snow accumulation or snowmelt (Ruelland 2020, SchreinerMcGraw and Ajami 2022). Comparatively, fewer studies have focused on the impacts of errors in streamflow observations used to calibrate and evaluate rainfall–runoff models (Section 1.3). However, the sources of uncertainty in streamflow estimates are numerous.

By definition, a streamflow is the volume of water that a river produces per unit of time at a given cross-section. To calculate it, hydrometrologists measure the flow velocity profile in the river over the wetted area, i.e. the cross-sectional area of the bed where the water flows. However, streamflow remains difficult to evaluate, as it is not homogeneous over the whole section (the width, depth and shape of the bed vary

naturally). This is all the more true in high-flow conditions where there may be overbank flows, or in low-flow conditions where the heterogeneity of the riverbed may have strong impacts on the flow conditions (due to rocks, vegetation, etc.). The extreme flows are therefore also the most uncertain. Although hydrometric gauging techniques are becoming increasingly efficient, variability in measurements can be observed depending on the choice of the cross-section within a river reach or on the experience of the gauging crew (Despax *et al.* 2019). Direct measurements of streamflow are sparse and manually operated, hindering continuous measurement. Instead, water levels are automatically recorded and a rating curve based on simultaneous measurements of water depth and streamflow is used to create a streamflow record. Therefore, continuous streamflow observations should instead be called “streamflow estimates” since there is generally no direct measurement.

The water stage in the river is easier to obtain in a quasi-continuous and automatic way, typically using a sensor (e.g. float, pressure or electronic gauges). Like any measuring instrument, sensors are subject to errors linked to their calibration, performance and precision, to the measurement conditions (e.g. sensitivity to water temperature) or to the conversion into electrical signals for remote data transmission (Horner *et al.* 2018). A well-known issue in measured data is noise, which is due to random influences and inaccuracies that cannot be eliminated completely (Schouten *et al.* 1994). Some studies aimed to reduce noise in series with methods based on filters. They mostly concluded that this process is effective in

smoothing the data but that it can remove or distort a significant part of the original signal (Elshorbagy 2001, Peters *et al.* 2014, Karunasingha and Liong 2018, Wang *et al.* 2019a, 2019b).

To estimate streamflow time series, a rating curve giving the stage–discharge relationship in each monitored section is defined. The more information available on the river close to the measurement section (e.g. high-frequency measurements or presence of nearby hydraulic controls), the smaller the uncertainty in the rating curve (Le Coz *et al.* 2014). However, this stage–discharge relationship is likely to be modified with time since natural river beds evolve continuously, e.g. with vegetation growth or sediment transport, and especially during strong flood events or because of human activities (Mansanarez *et al.* 2019, Darienzo *et al.* 2021, Perret *et al.* 2021). In addition, the extremities of the rating curve remain very uncertain, because gauging is difficult under extreme climatic conditions (i.e. drought or floods), and is thus a source of error in streamflow records. This relationship is all the more difficult to estimate in contexts where temperature can play a key role, typically in high-mountain or high-latitude regions where river ice and ice jams modify the stage–discharge relationship. During such periods, the streamflow data may be created by gap-filling approaches involving subjective choices that might introduce biases and non-homogeneities. Streamflow reconstruction (e.g. post-flood) is therefore also a source of uncertainty.

However, the inconsistencies of streamflow time series are not limited to measurement errors and methodology. Indeed, human influences often result in non-natural streamflow, which is difficult to model. But the issue of naturalized streamflow records (see e.g. Terrier *et al.* 2021, for a review) or the inclusion of anthropogenic influences in hydrological models (Montanari *et al.* 2013) is out of the scope of our study.

When dealing with a large sample of catchments there can be numerous streamflow inconsistencies of varied or unknown causes. Wilby *et al.* (2017) proposed to summarize these error sources with 12 carefully chosen cases (the “dirty dozen”). They suggested several methods for detecting unrepresentative, poorly collected or erroneously archived data (e.g. visual inspection of raw data, comparison between sites, trend detection or metadata check). However, error detection and attribution are more difficult to conduct for large datasets, where observation networks are automated, or when various information sources have been combined.

Regardless of their sources, the various errors can be classified into two categories:

- Systematic errors: These most often result in a bias over a long period and are typically caused by an error in the rating curve or an incorrect initialization of the water depth measurement sensor.
- Random errors: These are sporadic and unpredictable; they can be directly inherited from metrological errors (e.g. the accuracy of measuring instruments), human errors (e.g. poor handling during measurement or data evaluation) or digital errors (e.g. data banking or remote transmission).

With the increased availability of large databases, Wagener *et al.* (2021) recently put forward some issues

in this context: where (regions) and when (periods) are available datasets informative or simply poor approximations of likely system properties? How can we best identify and acknowledge these gaps to better understand and reduce the uncertainty in characterizing hydrological systems? These questions are directly inherited from the works undertaken in Great Britain on the information (or disinformation) conveyed by various input data of hydrological models (Beven *et al.* 2011, Beven and Westerberg 2011, Beven and Smith 2015, Westerberg *et al.* 2020). Seibert and Westerberg (2022) have recently taken up this work to categorize disinformative data and their causes. However, it is difficult to reach a conclusion regarding when data are counterproductive for modelling, especially in a large sample of catchments. In addition, the causes of these inconsistent data are sometimes difficult, if not impossible, to find without studying the data precisely, catchment by catchment. The detection and treatment of disinformative data and the understanding of their impact on hydrological modelling remain a major challenge today.

Since streamflow data form the basis of many hydrological studies, it is important to ensure their quality. In France, for example, these data are quality-checked and corrected by data producers through predefined quality control procedures before being archived and made available for various studies (Puechberty *et al.* 2017). This limits systematic and random errors in time series available for hydrological studies (Leleu *et al.* 2014). Although greatly reduced, suspicious streamflow data nevertheless remain in time series archived in national databases.

1.2 Quantifying uncertainty in streamflow data

Due to the various sources of errors detailed above, streamflow estimates are intrinsically uncertain. The level of uncertainty associated with flow estimates varies greatly, whether in space or time. Based on an extensive literature review, Pelletier (1988) estimated that the uncertainty in streamflow estimates varies between 8% and 20% for a 95% confidence interval according to the method used. Many studies focused on the role of the stage–discharge relationship, which is considered a major source of uncertainty. Di Baldassarre and Montanari (2009) focused on the uncertainties related to the rating curve, especially when used in extrapolation (i.e. below the lowest or above the highest flow measurement). They found that, under non-stationary flow conditions, or following seasonal changes in the river roughness, the overall uncertainty in streamflow varies between 6.2% and 42.8% (25.6% on average), for a 95% confidence level, over the river reach studied, which is far from negligible.

Various methods exist to estimate the uncertainty associated with the rating curve. The Bayesian framework is well adapted to this, since new flow measurements can be considered as additional information useful for constraining the possible range of the rating curve (see e.g. Reitan and Petersen-Øverleir 2009, Le Coz *et al.* 2014). When applying the BaRatin method, which takes into account

the uncertainty of streamflow measurements but also the a priori knowledge of the hydraulic controls, Horner *et al.* (2018) highlighted the impact of water level errors on streamflow records and showed that random errors during gauging or hydraulic head measurements generate little uncertainty in the time series of six catchments. On the other hand, systematic errors in water level have a strong impact on streamflow time series.

Uncertainty can also be quantified by ensemble approaches, as suggested by McMillan *et al.* (2017). Indeed, given the uncertainties in rating curves, it is possible to generate multiple flow records. The authors address this issue by showing the economic, social and environmental benefits of quantifying uncertainty in streamflow data, using case studies from Norway and New Zealand. At such national scales, there may also be issues in the heterogeneity of data quality, which should be accounted for. Hamilton and Moore (2012) highlighted that data can come from a wide panel of providers, who use different instruments, data collection practices and methods to quantify uncertainty. This information is rarely available to data users, and the authors make recommendations to unify practices for quantifying uncertainties in streamflow data.

The issues involved in quantifying uncertainty are manifold and there is consensus on the need to use good-quality streamflow data. For this purpose, the largest uncertainties, i.e. related to the rating curve or to systematic errors in the water level records, must be minimized.

1.3 Sensitivity of rainfall–runoff modelling to the quality of streamflow data

Since streamflow data are uncertain, it is important to perform a sensitivity analysis to determine whether the various errors in streamflow estimates affect the performance and parameter estimates of rainfall–runoff models. Saltelli *et al.* (2000) define sensitivity analysis as the study of how the uncertainty in a model output can be attributed, qualitatively or quantitatively, to different sources of variation. In this work, sensitivity analysis focuses on how the model depends on the quality of the streamflow time series used for its calibration. We will therefore limit this review to issues related to data quality. Other studies have analysed the impact of the quantity of flow data on the performance and parameters of hydrological models (see e.g. Perrin *et al.* 2007, for a review).

Among the first sensitivity analysis studies, Ibbitt (1972) generated synthetic daily input data (precipitation, potential evapotranspiration and streamflow) and then corrupted them with random errors drawn from a normal distribution. The author concluded that errors usually encountered in hydrological records have a negligible effect on model parameter estimates since limited variations appear when the intensity of the errors is exaggerated. Borah and Haan (1991) used the same methodology in a real case study of a basin in Oklahoma, USA. In this study, corruption of precipitation

data appeared to introduce more uncertainty in the parameter estimates than errors in the flow records did. The authors also show that correlated errors can introduce significant errors in the estimated parameters. Montanari and Di Baldassarre (2013) analysed the sensitivity of models to flow errors according to their structure. Results obtained on a tributary of the Po River (Italy) indicate that the lower the model complexity, the larger the impacts of flow errors. The effect of uncertainty can also depend on the type of errors, i.e. systematic or random. Tillaart *et al.* (2013) show for two tributaries of the Meuse River that systematic errors introduced in flow data strongly deteriorate the performance and parameter stability of rainfall–runoff models, particularly those controlling the water balance. Random errors with autocorrelation also seem to influence the hydrograph shape, although to a lesser extent.

As discussed above, the rating curve is often considered a major source of uncertainty. McMillan *et al.* (2010a) analysed the impact of uncertainties associated with the rating curve on hydrological modelling at the hourly time step. By applying a method for quantifying the uncertainty in the stage–discharge relationship, they showed, for the Wairau River in New Zealand, that explicit consideration of streamflow uncertainty increases the variability in parameters. By analysing the effect of multiple sources of uncertainties (input data, parameters, and model structure) on streamflow simulation at the hourly time step in the Orsali basin in Norway, Engeland *et al.* (2016) confirmed that the use of different stage–discharge relationships, representing various levels of systematic errors, led to variability in the water balance parameters. However, it was difficult to determine a trend in the model performance while evaluating it within an independent period, as both positive and negative variations were shown. In Australia, Peña-Arancibia *et al.* (2015) concluded that rating curve uncertainties can have important consequences on the calibration and efficiency of hydrological models by showing a variance heterogeneity (heteroscedasticity) in streamflow estimates and the existence of streamflow extrapolation beyond the highest or lowest stage measurement in many operational rating curves. Recently, Westerberg *et al.* (2020) studied the impact of uncertain streamflow data (due to the rating curve) on the calibration of hydrological models. The authors developed an objective function that considers streamflow uncertainties, which provides more reliable results.

This short literature review shows the evolution of research from studies that corrupted the time series with random errors at the daily time steps, with a limited impact on model results, towards more detailed studies at the hourly time step, considering various types of uncertainties affecting streamflow observations. Systematic errors linked to bias in the water level time series or generated by the uncertainty of the rating curves were shown to be detrimental to model performance and/or parameter estimates.

However, to our knowledge, only a few studies have focused on random errors, and most such studies were based on the hypothesis of errors drawn from a normal distribution without any real expertise on their typology (i.e.

without considering the types of errors most likely to affect real time series). In addition, the majority of the studies reviewed above focus on a single case study or on synthetic data.

1.4 Objectives

To overcome the limitations of existing studies, we propose to investigate the sensitivity of the performance and parameter estimates of rainfall–runoff models when the information on streamflow is corrupted in a random and realistic way. By “realistic” we mean comparable to the level of errors commonly found in streamflow time series available for hydrological studies, i.e. typically available in national flow archives. Here we will not focus on raw flow records directly retrieved from measurement devices, but rather on time series that have undergone an initial quality control based on the expertise of the data producer before being stored in a database accessible to end users. More specifically, we address the issue from a large-sample hydrology (LSH) perspective (Andréassian *et al.* 2006, Gupta *et al.* 2014), in which several catchments are used and where the detailed critical analysis of streamflow time series is complex. We wish to answer the following question: Do errors in streamflow time series have a strong impact on model performance and parameter estimates when working with several catchments?

To this end, we first tried to characterize the errors commonly found in hourly time series available for hydrological studies, by expert analysis. We then carried out a sensitivity analysis using flow series corrupted by simple error models representing the various types of errors identified.

The remainder of the paper is organized as follows. First, the hydroclimatic data, catchment sets, and hydrological model used for these tests are presented. A typology of suspicious data found in the time series is proposed before addressing the methodological approach used for this work. We then present, analyse and discuss the results. Last, we summarize the main conclusions of this work and introduce different research perspectives.

2 Material and methods

2.1 Hydroclimatic data

This study was conducted at an hourly time step using precipitation, potential evapotranspiration and streamflow time series over the 1998–2018 period (Delaigue *et al.* 2020).

Precipitation data were extracted from the COMEPHORE (standing for COMbinaison en vue de la Meilleure Estimation de la Précipitation HORaiRE, i.e. Combination for best hourly precipitation estimate) re-analysis produced by Météo-France (Tabary *et al.* 2012), which provides information at a 1 km² resolution and which has already been extensively used in hydrological studies (Artigue *et al.* 2012, van Esse *et al.* 2013, Bourgin *et al.* 2014, Lobligeois *et al.* 2014, Saadi *et al.* 2021).

Potential evapotranspiration is calculated with the formula proposed by Oudin *et al.* (2005). This equation was chosen for its simplicity, as the only input required is daily air temperature (from the SAFRAN (standing for Système d’Analyse

Fournissant des Renseignements À la Neige, i.e. Reanalysis system providing information on snow) reanalysis of Météo-France, see Vidal *et al.* 2010) and extra-terrestrial radiation (which depends only on latitude). Once calculated, the daily potential evapotranspiration was disaggregated to the hourly time step using a simple parabola (Lobligeois 2014).

Streamflow time series were extracted from the national flow archive Banque Hydro (Leleu *et al.* 2014). This database gathers data produced by hydrometric services in regional environmental agencies in charge of measuring flows in France as well as by other data producers (e.g. hydropower companies, dam managers, etc.). Before being archived, streamflow data undergo quality control strategies applied by data producers, with corrections and gap-filling procedures applied when necessary (Puechberty *et al.* 2017). Quality codes are also available, although this information is not uniformly provided for all stations. These data are freely available on the Banque Hydro website (<https://www.hydro.eaufrance.fr>) and are widely used in France for hydraulic and hydrological studies. They can therefore be considered as standard streamflow data.

2.2 Catchments

A total of 30 catchments spread over mainland France within the hydroclimatic database set up by Delaigue *et al.* (2020) were selected for this study, 15 because of the common suspicious data they display in streamflow records, and 15 for the apparent quality of their streamflow time series. They were also chosen to represent a variety of hydrological behaviours in terms of physical characteristics or hydroclimatic conditions and to provide complete streamflow time series over the period 1998–2018 (with a maximum of 5% missing data per year). Nevertheless, this set may not be representative of all conditions over France (e.g. dominant snow or glacial regimes are missing here due to incomplete time series). Although a larger catchment set could have been used, the identification of these catchments within the large Banque Hydro (> 4000 gauging stations) was time consuming and we found that this number was a good compromise between the time necessary to set up

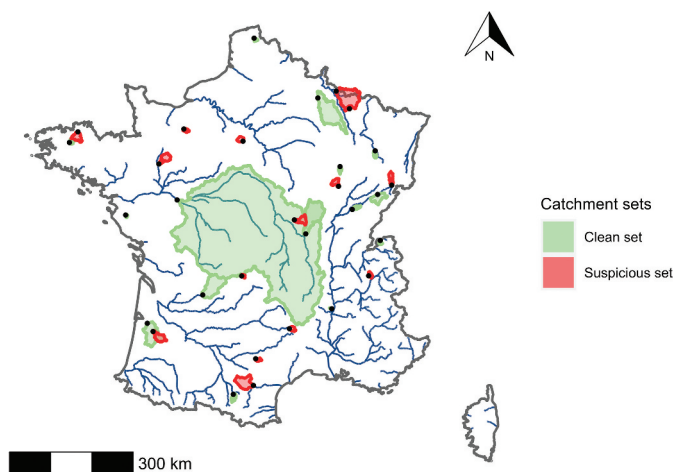


Figure 1. Catchment boundaries of the clean (green) and suspicious (red) French catchment sets selected for this study. Black dots are catchment outlets.

Table 1. Minimum, median and maximum values of the main characteristics of the catchment sets.

	Suspicious set			Clean set		
	Min.	Med.	Max.	Min.	Med.	Max.
Catchment area [km ²]	42.0	230.0	1978.0	8.0	119.0	81 314.0
Mean annual precipitation [mm y ⁻¹]	680.0	958.0	1506.0	755.0	992.0	1310.0
Mean annual potential evapotranspiration [mm y ⁻¹]	614.0	668.0	829.0	648.0	687.0	801.0
Mean annual flow [mm y ⁻¹]	116.0	391.0	985.0	227.0	355.0	569.0
Humidity index (P/PE) [-]	1.0	1.4	2.4	1.1	1.4	2.0
Runoff coefficient (Q/P) [-]	0.2	0.4	0.7	0.2	0.3	0.5

P: precipitation; PE: potential evapotranspiration; and Q: flow.

the study and perform the test, on the one hand, and the generalizability of our results on the other.

The first subset of 15 catchments, called the “suspicious set”, was used to identify a variety of errors in streamflow data and to build a simple typology of suspicious data in streamflow records (Section 2.3). The second subset of 15 catchments, called the “clean set”, was used to apply data corruption (Section 2.4) and to perform a sensitivity analysis (Section 3). The quality of the record is defined according to the occurrence of errors listed in the typology (Section 2.3). The catchment locations are illustrated in Fig. 1 and the main characteristics are summarized in Table 1 (details are given in Appendices A and B). Both subsets are rather uniformly spread over France, and their characteristics follow similar ranges (except for the maximum catchment area).

2.3 Typology of suspicious data

As a reminder, the actual streamflow is never known and suspicious data may arise from recording errors or from unusual – but real – natural or human-induced processes. Here, by “suspicious data” we mean data that appear hydrologically inconsistent compared to neighbouring data in the time series or within the record itself. The automatic detection of such data is not trivial because it can be tricky to differentiate between hydrologically consistent and inconsistent streamflow without expertise. For example, a very slow recession can be detected as a linear interpolation issue while a short and intense rainfall can lead to a curious spike. The objective is not to find the origin of this inconsistency (upstream influence, problem in the measurement device, etc.) but to identify data that do not seem to be related to common natural catchment responses to climate inputs. Some detection algorithms (Vallis *et al.* 2014, Dancho and Vaughan 2020) from the domain of signal processing have been tested without success since they did not enable the differentiation between natural and suspicious behaviours of the streamflow time series (many false positives were detected). Therefore, the suspicious set of catchments was submitted to expert visual inspection in order to detect hydrologically inconsistent data in the records. These time series were plotted in linear and logarithmic scales to better identify errors in high and low flows, respectively. Each streamflow record was analysed by two hydrologists (among the authors of this article) to enhance consistency. Due to the variety of the natural behaviours of rivers or to possible human influences, it is difficult to categorize all the suspicious data. However, four main patterns (Fig. 2) seem to stand out from our analysis:

- (A) Noise: It is characterized by an additive random signal (both positive or negative) to the initial signal carrying the information. It appears most often during low-flow periods (sometimes over several weeks). It may originate from the measuring device, from issues of sensitivity or from local anthropogenic influences.
- (B) Curious spikes: These are sharp peaks over a short period (a few hours), which differ from the expected trend. It is the most common pattern detected in the hourly time series. Curious spikes are typically the effect of random error (e.g. measurement problems) but can also be explained by human interference, such as river pumping or effluent discharge.
- (C) Drops: This pattern corresponds to a sudden change, lowering or raising the hydrograph, before returning to the initial signal. It is an uncommon error, appearing as a systematic error in a moderate part of the time series (e.g. a few days or weeks). The pattern may have various origins. For example, a sudden decrease may appear when obstructions disrupt the water level station. The influence of water management operations (typically outputs of reservoirs for low-flow augmentation in summer) may also be responsible for these behaviours.
- (D) Linear interpolations: This error is directly visible when a segment is drawn between two points far apart in the time series (some weeks). It stems from a problem with the coding of missing data. Instead of creating a gap, the data are filled by linear interpolation, which leads to a source of disinformation.

Note that the term “suspicious” was intentionally preferred to the term “erroneous” in the article to avoid reference to the origin of the data considered inconsistent. Suspicious data may originate from problems in measurements (i.e. actual errors), but also from unusual natural behaviour or artificial influences (i.e. not actual errors). A more comprehensive analysis would be to find the actual source of inconsistency in each case, but this is out of the scope of this article.

On average, over the 15 catchments analysed, almost 3% of the total record was assessed as suspicious, which represents about 7 months out of 20 years of data. These errors are mostly found during low-flow periods, since 87% of them are detected below the median flow in each time series. Although numerous, the curious spikes do not have a strong temporal impact since they are point errors over a few hours. Noise is the pattern that affects the most time steps. Note that several different patterns can be found in the same record (Fig. 3).

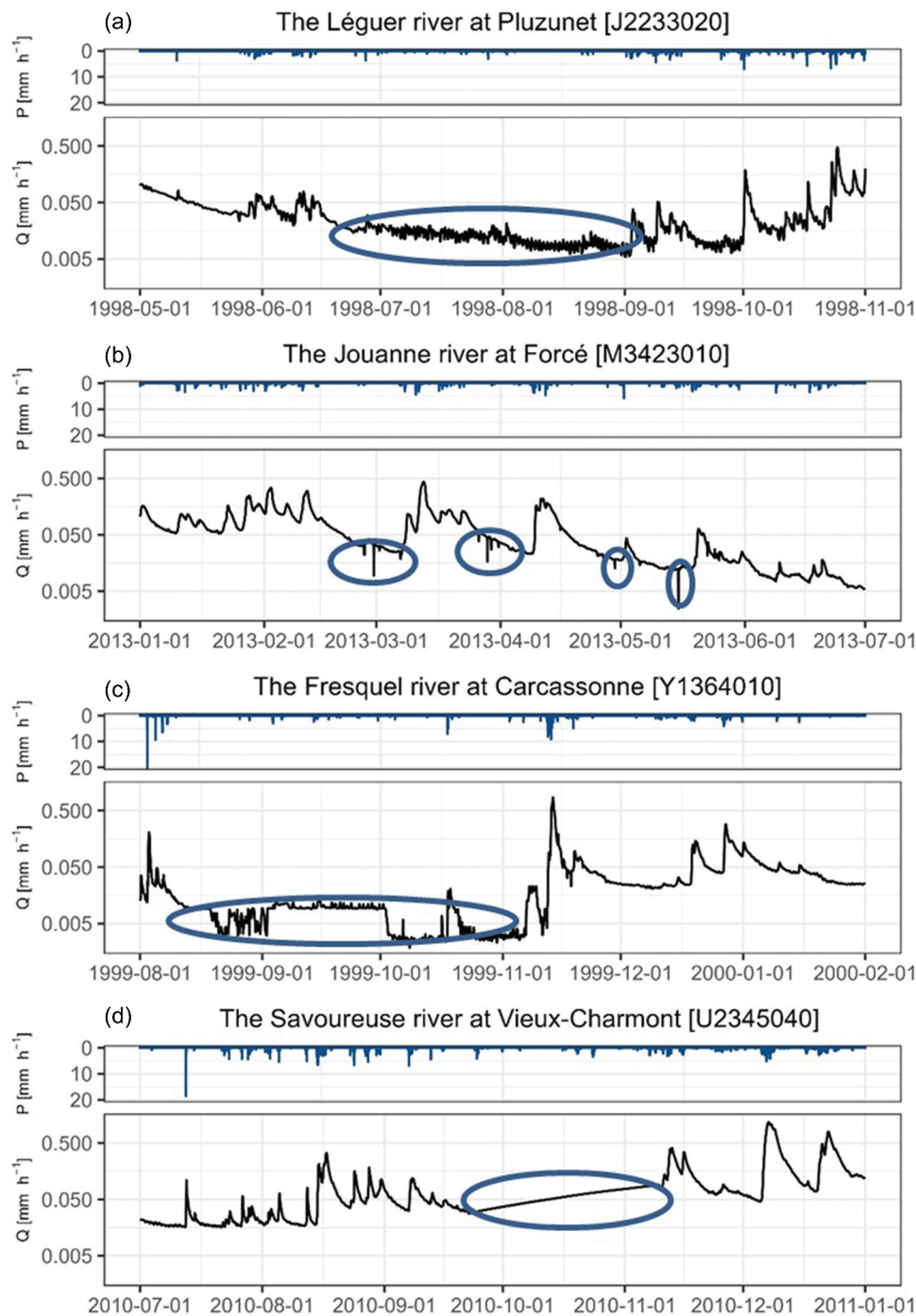


Figure 2. Illustration of the main suspicious patterns – (a) noise, (b) curious spikes, (c) drops and (d) linear interpolations – found in the hourly streamflow data of four catchments in the database.

2.4 Corruption of streamflow time series using simple error models of suspicious data

In this study, we want to evaluate the impact of suspicious streamflow data on modelling results, all else being equal. Since the process of correcting suspicious data to produce natural data is not straightforward, we chose to corrupt the streamflow data in non-erroneous time series (i.e. starting from the clean dataset). Therefore we obtained non-erroneous and erroneous time series differing only in the corrupted time steps. Oudin *et al.* (2006) proposed to corrupt

precipitation and potential evapotranspiration time series with biased and random errors, and then to analyse their impact on the parameters and performance of hydrological models. This approach was repeated here by attempting to reproduce the four patterns found in streamflow records. The objective was to implement realistic suspicious data in the streamflow time series of the clean set according to the typology defined previously (Section 2.3).

Noise most often corresponds to the appearance of values close to the initial signal, in terms of intensity and spaced out by short time intervals, corrupting the main

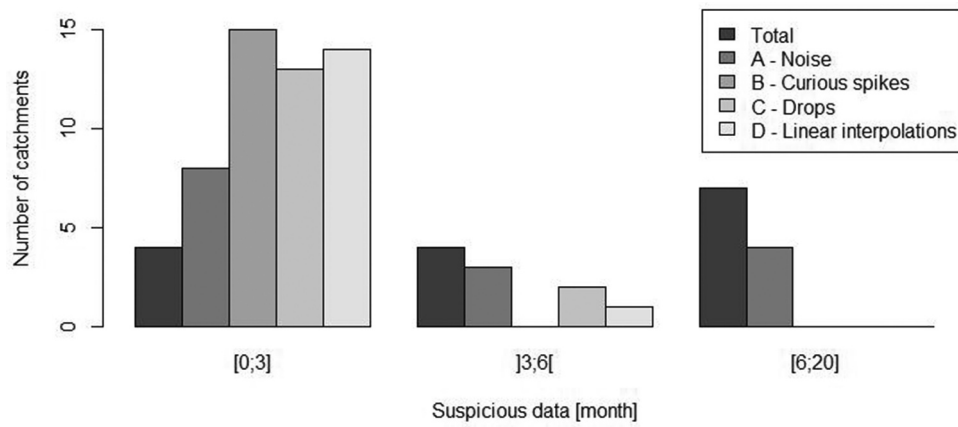


Figure 3. Distribution of total duration of suspicious data (in months) over the suspicious set (15 catchments over the 20-year period), according to their type.

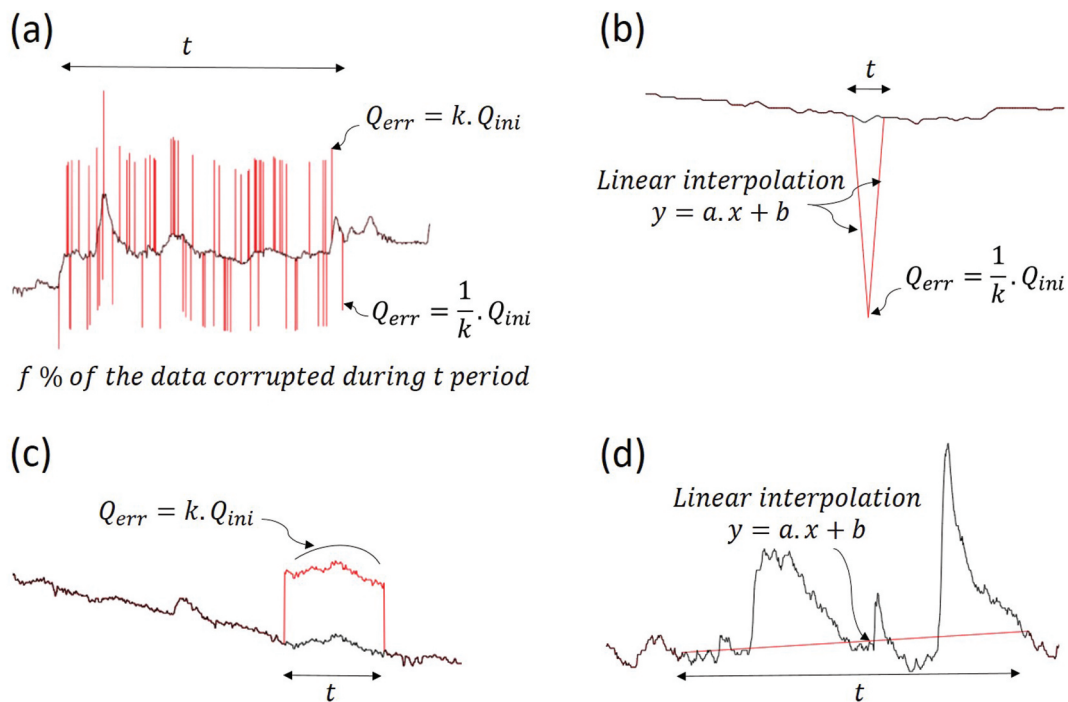


Figure 4. Illustration of (a) numerical implementation of the noise, (b) curious spike, (c) drop and (d) linear interpolation patterns within the time series of the clean catchment set. Original series are in black and corrupted parts are in red.

information up to several weeks. For this purpose, a simple positive multiplier coefficient k , greater than 1, was applied to a percentage f of the time window t on which the pattern is applied. The ascending or descending factor of the spikes was determined randomly over the pattern period (Fig. 4(a)).

Curious spikes are random errors that can be visible at the hourly time step. This pattern often shows a triangular shape. Here, we proposed to implement this error by defining the peak by the intensity coefficient k for an increase (or its inverse for a decrease) and then to link this erroneous flow value to the surrounding parts of the observed hydrograph by simple linear

Table 2. Ranges for the parameters of the error models used for streamflow corruption according to the typology of suspicious data adopted. n is the number of patterns in 10 years, k the multiplicative coefficient of the suspicious pattern, t the length of the time window of the suspicious pattern, and f the percentage of time steps to be corrupted by the suspicious pattern over the time window.

	n [-]	k [-]	t [h]	f [%]
Noise	[1; 7]	[1.3; 2.1]	[360; 1800]	5
Curious spikes	[1; 4]	[1.3; 2.1]	[1; 12]	n/a
Drops	[5; 30]	[1.3; 2.1]	[120; 600]	n/a
Linear interpolation	[1; 4]	n/a	[288; 1440]	n/a

interpolation over the time window t representing the duration of the pattern (Fig. 4(b)).

The behaviour of drops seems to resemble a bias affecting the time series up to several weeks and appearing randomly within the record. This pattern is implemented here by shifting the time series upward with an intensity coefficient k (or downward with its inverse) applied on the whole pattern over the time window t (Fig. 4(c)).

Finally, the linear interpolation pattern was implemented by linearly linking flow values chosen at both ends of a time window t (Fig. 4(d)).

Table 2 summarizes the set of parameters that can be modified to represent artificial, yet realistic, suspicious streamflow patterns. For this study, the parameter f was set to 5% because it does not seem to vary strongly in the suspicious set analysed. A parameter n was used representing the number of identical suspicious patterns introduced over 10 years in the streamflow record. The corruption parameters of the records were chosen to represent the actual dynamics of the suspicious data found in streamflow time series. For this purpose, a gamma distribution, based on the real suspicious data encountered in the suspicious dataset (Section 2.3), was assigned to each corruption parameter of each pattern (Fig. 5).

The analysis of the impact of suspicious data was carried out on the clean set by alternatively considering the original and corrupted flow time series. Each error pattern could be studied separately and implemented in various conditions (e.g. calibration/evaluation, high flows/low flows).

In this work, we focused on analysing the sensitivity of hydrological models (performance and parameter estimates) to the suspicious patterns commonly found in streamflow time series.

2.5 Rainfall-runoff model

We used the continuous GR5H (standing for modèle du Génie Rural à 5 paramètres Horaire, i.e. Hourly 5-parameter rural engineering model) hydrological model, which is lumped and runs at the hourly time step (Le Moine 2008, Ficchi 2017, Ficchi *et al.* 2019). The model estimates river streamflow from previous meteorological conditions, using three stores (interception, production, and routing), a unit hydrograph, and a function to account for inter-catchment groundwater exchanges (Fig. 6).

The model seeks to represent in a synthetic way the catchment-scale dynamics of processes at play in the catchment (e.g. infiltration, evapotranspiration, runoff, etc.). All the calculations in the model are mathematical relationships, which depend on different parameter values that need to be determined during the model calibration phase (Table 3). The model was implemented using the airGR package (Coron *et al.* 2017, 2021), which includes a calibration algorithm well adapted to the model.

Although other models could have been used, the GR5H model has two main advantages for our study. First, it has already been tested extensively in France on several hundreds of catchments and has shown a good level of performance (Ficchi *et al.*

2019). Second, its parsimony (only five free parameters to estimate) makes it easy to calibrate and facilitates the interpretation of parameter variations in terms of sensitivity analysis.

Here the model was used without any snowmelt module. The influence of snow remains limited on most catchments, and we preferred to limit model complexity and avoid introducing additional parameters. Given the comparative framework of our methodology, we think that this choice has no significant impact on our conclusions.

2.6 Hydrological model testing methodology

For this study, the common calibration–evaluation testing framework (Klemeš 1986) was applied for the 1999–2018 period with the hydrological model. The first decade, 1999–2008, was used to estimate the parameters of the hydrological model and the second, 2009–2018, was used to evaluate the simulation. The 1998 data were used for model initialization. However, one year of warm-up may not be enough for some basins. Therefore, the pre-1998 decade was added by uniformly disaggregating the SAFRAN daily precipitation at the hourly time step.

Hydrological model testing was performed for the 15 catchments of the clean dataset only (the catchments of the suspicious set were only used to set up the error models). The hydrological model was first applied (i.e. calibrated and evaluated) using the original flow time series (considered without suspicious data) and the corresponding simulations were considered as the reference. Then, the hydrological model was applied to each catchment using the flow series corrupted with the four types of suspicious data and various levels of corruption by modifying the parameters of the error models. This resulted in 8820 corrupted time series over the 15 catchments, i.e. 588 per catchment (1680 tests for the noise type, 3360 for the curious spikes type, 3360 for the drops type and 420 for the linear interpolations type). Although the actual series frequently displays a mix of suspicious patterns (see Fig. 3), we excluded the case of mixing different error types in our tests because it would have resulted in too many combinations, rendering the results difficult to interpret. We believe the results that had a single suspicious pattern in each case already provide useful insights.

The chosen criterion for calibration and evaluation is the Kling-Gupta efficiency (KGE) (Gupta *et al.* 2009) calculated on the square root of streamflow in order to obtain a compromise between low and high flows:

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (1)$$

where r is the correlation, α is the ratio of standard deviations and β is the ratio of means (bias) of the observed and simulated streamflow.

Two types of tests were performed:

- (A) First, only the calibration period was corrupted and the evaluation period was left unchanged (see Section 3.1). This would correspond to the practical case where a modeller uses raw flow data as retrieved from a database for model calibration, without further

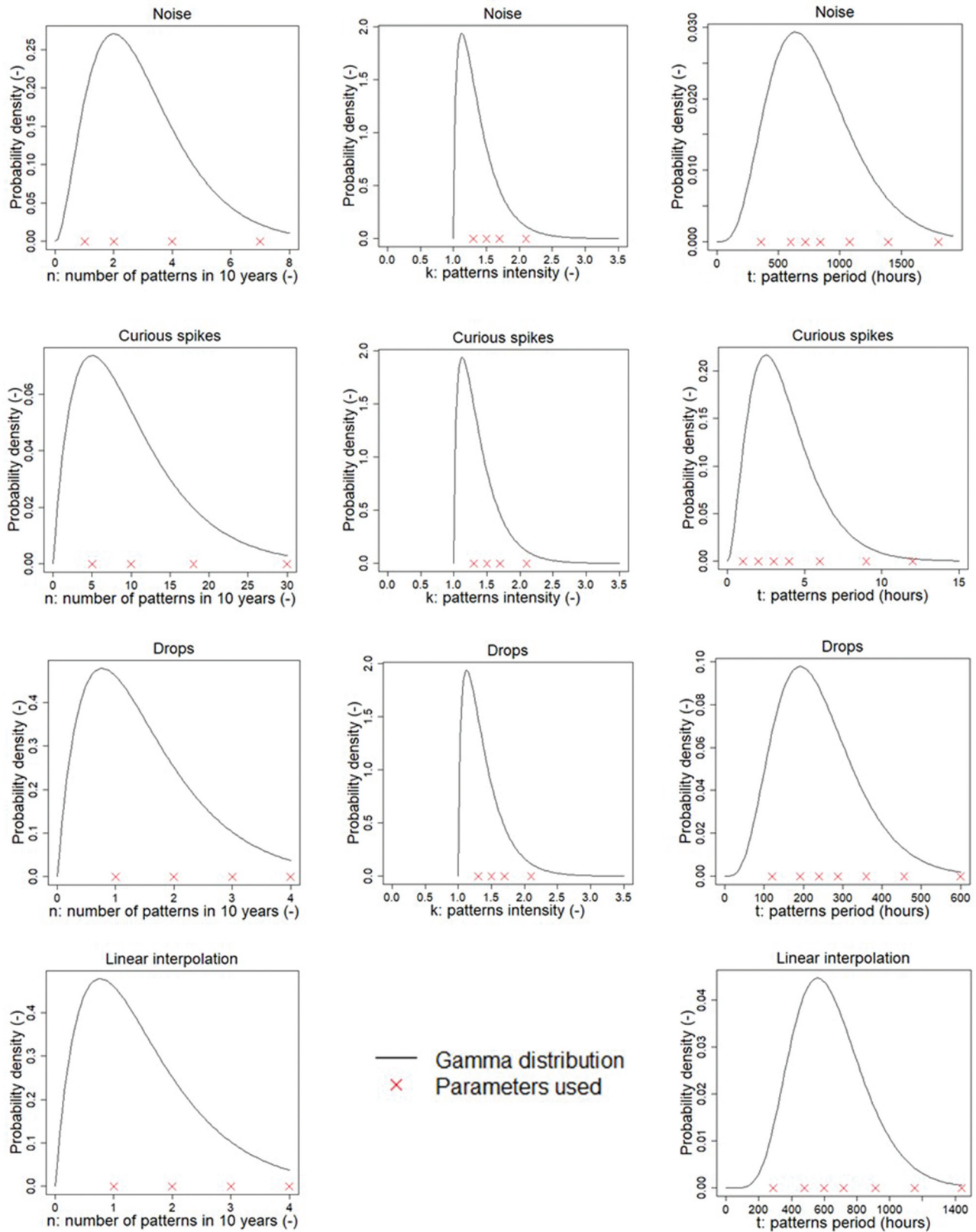


Figure 5. Chosen pattern parameters and gamma distributions defined to mimic suspicious data for the four types of errors.

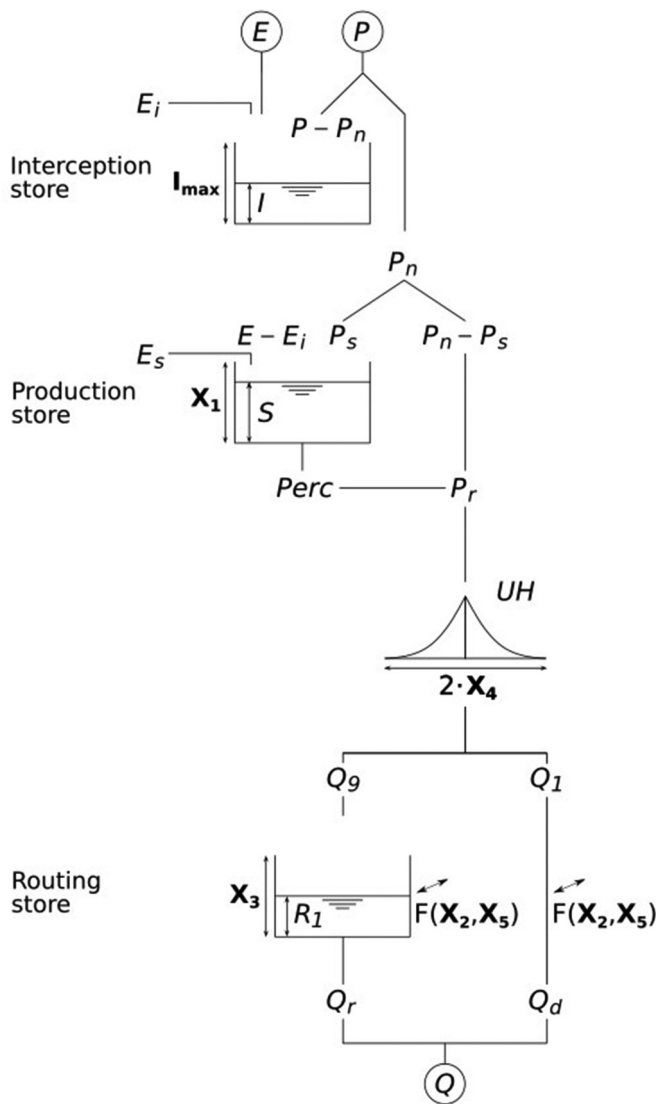


Figure 6. Schematic diagram of the GR5H model (X1 to X5 represent model parameters; see Table 3); P, E and Q stand for precipitation, potential evapotranspiration and streamflow, respectively; other letters are internal state variables) (Coron *et al.* 2021).

quality checks. The model was calibrated using either original or corrupted data, which leads to several sets of parameters. In order to assess the impact on the parameter values of the model, a variation can be calculated between the parameter values obtained by model calibration with the corrupted streamflow time series and with the original record. In this case, the variation is defined as follows:

$$var_param = \frac{X_{cor} - X_{ori}}{X_{ori}} \times 100 \quad (2)$$

where X_{ori} and X_{cor} are the model parameters obtained with the original and corrupted flow series, respectively. A positive

(negative) value of var_param indicates an increase (a decrease) in the parameter value when moving from the original to the corrupted series.

The model performance in the calibration and evaluation periods is calculated using the original observed streamflow record as the reference, and thus the performance is strictly comparable between the various tests. The value of a relative approach in this context is to bear in mind that the closer the KGE value is to 1, the more difficult it is to improve the performance. Lerat *et al.* (2012) proposed to calculate the variation in performance between two time series by:

$$var_perf1 = \frac{m(\sqrt{Q_{simori}} \sim \sqrt{Q_{obsori}}) - m(\sqrt{Q_{simcor}} \sim \sqrt{Q_{obsori}})}{m(\sqrt{Q_{simori}} \sim \sqrt{Q_{obsori}}) + m(\sqrt{Q_{simcor}} \sim \sqrt{Q_{obsori}})} \quad (3)$$

where $m(X \sim Y)$ is $1 - KGE(X \sim Y)$. $KGE(X \sim Y)$ is the KGE criterion calculated for the variable X as an estimate of the reference variable Y. Q_{simcor} and Q_{simori} are, respectively, the corrupted and original simulated streamflow time series, while Q_{obsori} is the original observed streamflow record. A positive (negative) value of var_perf1 indicates an increase (a decrease) in performance value when moving from the original to the corrupted series. A more detailed interpretation of this relative performance is available in Table 4.

(A) Then, the calibration period was left unchanged (i.e. using the original uncorrupted data) and the evaluation period was corrupted (see Section 3.2): The model was calibrated on the original data, and the impact was assessed only on model performance in the evaluation. This would correspond to the case where a modeller uses flow data that are as good as possible for model calibration, but then applies the model to a period with suspicious streamflow data, which may be the case typically for real-time applications. This test therefore has no impact on the parameter values or the calibration performance. However, a sensitivity analysis on the performance during the evaluation period can be carried out, using the single simulated streamflow record as the reference and the observed corrupted time series as the variable. To this end, a variation can be calculated between the performance values obtained with the corrupted streamflow time series and with the original record. The variation is also based on Lerat *et al.* (2012) and is here defined by:

$$var_perf2 = \frac{m(\sqrt{Q_{simori}} \sim \sqrt{Q_{obsori}}) - m(\sqrt{Q_{simori}} \sim \sqrt{Q_{obsor}})}{m(\sqrt{Q_{simori}} \sim \sqrt{Q_{obsori}}) + m(\sqrt{Q_{simori}} \sim \sqrt{Q_{obsor}})} \quad (4)$$

where Q_{obsor} is the corrupted observed streamflow time series. A positive (negative) value of var_perf2 indicates an increase (a decrease) in performance value when moving from the original to the corrupted series. A more detailed interpretation of this relative performance is available in Table 4.

Table 3. Parameters of the GR5H model.

Parameter	Meaning	Unit	Typical range in France
X ₁	Production store capacity	mm	100 to 400
X ₂	Inter-catchment exchange coefficient	mm h ⁻¹	-0.4 to 0.0
X ₃	Routing store capacity	mm	40 to 390
X ₄	Unit hydrograph time constant	h	2 to 27
X ₅	Inter-catchment exchange threshold	-	0.1 to 0.3

Table 4. Interpretation of the relative performance (*var_perf*) comparing the performance of a benchmark simulation (*ori*) with an alternative (*cor*) using a metric $m = 1 - KGE$ (adapted from Lerat *et al.* 2012).

<i>var_perf</i>	Interpretation
1	<i>cor</i> is perfect according to the metric m : $m(\sqrt{Q_{sim\ ori}} \sim \sqrt{Q_{obs\ ori}}) = 0$
1/2	<i>cor</i> is three times smaller (better) than <i>ori</i> according to the metric m : $3 \times m(\sqrt{Q_{sim\ cor}} \sim \sqrt{Q_{obs\ ori}}) = m(\sqrt{Q_{sim\ ori}} \sim \sqrt{Q_{obs\ ori}})$
0	<i>cor</i> and <i>ori</i> are equal according to the metric m : $m(\sqrt{Q_{sim\ cor}} \sim \sqrt{Q_{obs\ ori}}) = m(\sqrt{Q_{sim\ ori}} \sim \sqrt{Q_{obs\ ori}})$
-1/2	<i>cor</i> is three times larger (worse) than <i>ori</i> according to the metric m : $m(\sqrt{Q_{sim\ cor}} \sim \sqrt{Q_{obs\ ori}}) = 3 \times m(\sqrt{Q_{sim\ ori}} \sim \sqrt{Q_{obs\ ori}})$
-1	<i>ori</i> is perfect according to the metric m : $m(\sqrt{Q_{sim\ ori}} \sim \sqrt{Q_{obs\ ori}}) = 0$

The choice to calculate variations rather than absolute values was made in order to have an equivalent comparison for each basin, whatever the range of parameters or performances.

The bias (β component of the KGE, Equation 1) will also be used as evaluation criterion, to show the impacts of suspicious data on total volume. Due to the dissymmetry of β around the optimal value of 1 for under- and overestimation, we chose the symmetrical formulation proposed by Perrin (2000), which varies between $-\infty; 1]$ (1 being the optimal value):

$$\beta' = 1 - \left| \sqrt{\beta} - \frac{1}{\sqrt{\beta}} \right| \tag{5}$$

It is possible to compute meaningful β' values over the set of catchments but at the cost of a loss of information about over- and underestimation of the water balance.

The modelling tests were applied to two different cases:

- (1) when data were corrupted without a preconceived notion of streamflow range;
- (2) when only data in low-flow periods, i.e. under the median, were corrupted.

3 Results

Here we investigate the impact of the suspicious patterns usually encountered in streamflow records on hydrological modelling, i.e. on model parameter values and model performance (in calibration and evaluation). All the results are presented in terms of relative variation from the initial value (i.e. before the corruption). However, a table summarising the distribution of absolute parameter and performance values over the whole clean sample, whatever the suspicious pattern implemented, is available in Appendices C and D.

3.1 Data corruption over the calibration period

3.1.1 Impact on parameter estimates

Here we focus on the impact of suspicious streamflow data on model parameters when the corruption was made only during the calibration period. To this end, the variation used for this section was *var_param* (Equation 2).

Figure 7 represents the distribution of parameter variations obtained from the 15 catchments under this analysis framework. This corruption was initially carried out without any preconceived notion of streamflow range (shown in blue in Fig. 7). The distribution of the variations in the five model

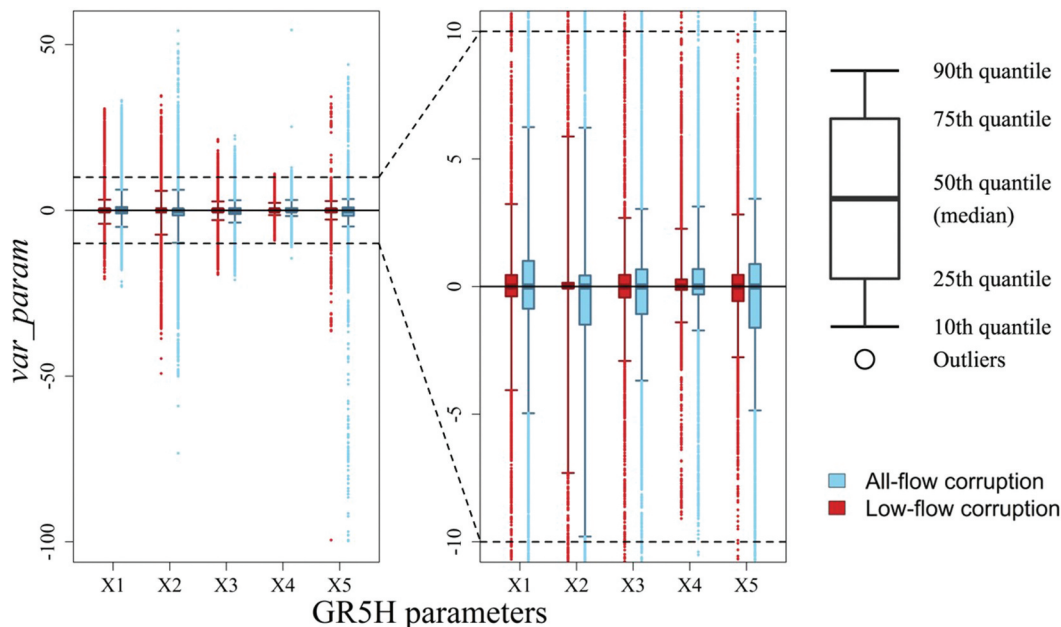


Figure 7. Distribution of parameter estimate variations obtained from calibration of the 15 catchments. The box plots represent the 10th, 25th, 50th, 75th and 90th quantiles and contain 8820 values each, including outliers. Red corresponds to corruption in the low-flow periods and blue to corruption without restriction in the flow range. Data corruption was performed over the calibration period only.

parameter estimates over the entire set of 8820 calibration tests shows that the variations in parameter values are very limited. Indeed, whichever parameter is considered, its variation is lower than 10% for at least 80% of the cases. However, it should be noted that, occasionally, some parameter values can increase by up to 50% and decrease by up to 100%. The variations in the production store capacity, X_1 , remain small (between -4% and $+6\%$ for the 10th and 90th quantiles). Streamflow data corruption has little impact on the routing part of the model since parameters X_3 and X_4 remain mostly unaffected. X_2 is the most sensitive parameter to data corruption when considering the 10–90 inter-quantile range of the distribution. X_5 shows strong variations in some cases (between -100% and $+44\%$ when considering all the tests). The X_2 and X_5 parameters control the water exchange function with aquifers but also act to close the catchment water balance.

Let us recall that 87% of the suspicious streamflow data were in low-flow periods, i.e. below the median streamflow of each record (Section 2.3). Therefore, corruption was re-applied only to the low flows in order to assess the impact of the suspicious patterns under these conditions (shown in red in Fig. 7). Results show that parameter values are less impacted

when suspicious data occur only during low-flow periods. It should be recalled that the calibration criterion is calculated over the entire range of streamflow.

Figure 8 extends this analysis by separating the types of error. It indicates that although noise and spike patterns are the most frequent, they seem to have little impact on the calibration of the parameters since they show variations lower than 5% in 80% of the cases. On the other hand, drops and linear interpolations lead to larger variations, especially in X_1 , X_2 , and X_5 values whose fluctuations can reach between 15% and 26% for the 10th or 90th quantiles.

3.1.2 Impact on model performance

Here we focus on the impact of suspicious streamflow data on model performance (in both calibration and evaluation) when the corruption was applied only during the calibration period. To this end, the variation calculation used for this section was var_perf1 (Equation 3).

Figure 9 represents the distribution of performance variations in the 15 catchments under this analysis framework. It highlights that performance variations are small in both calibration and evaluation periods, after corrupting either low

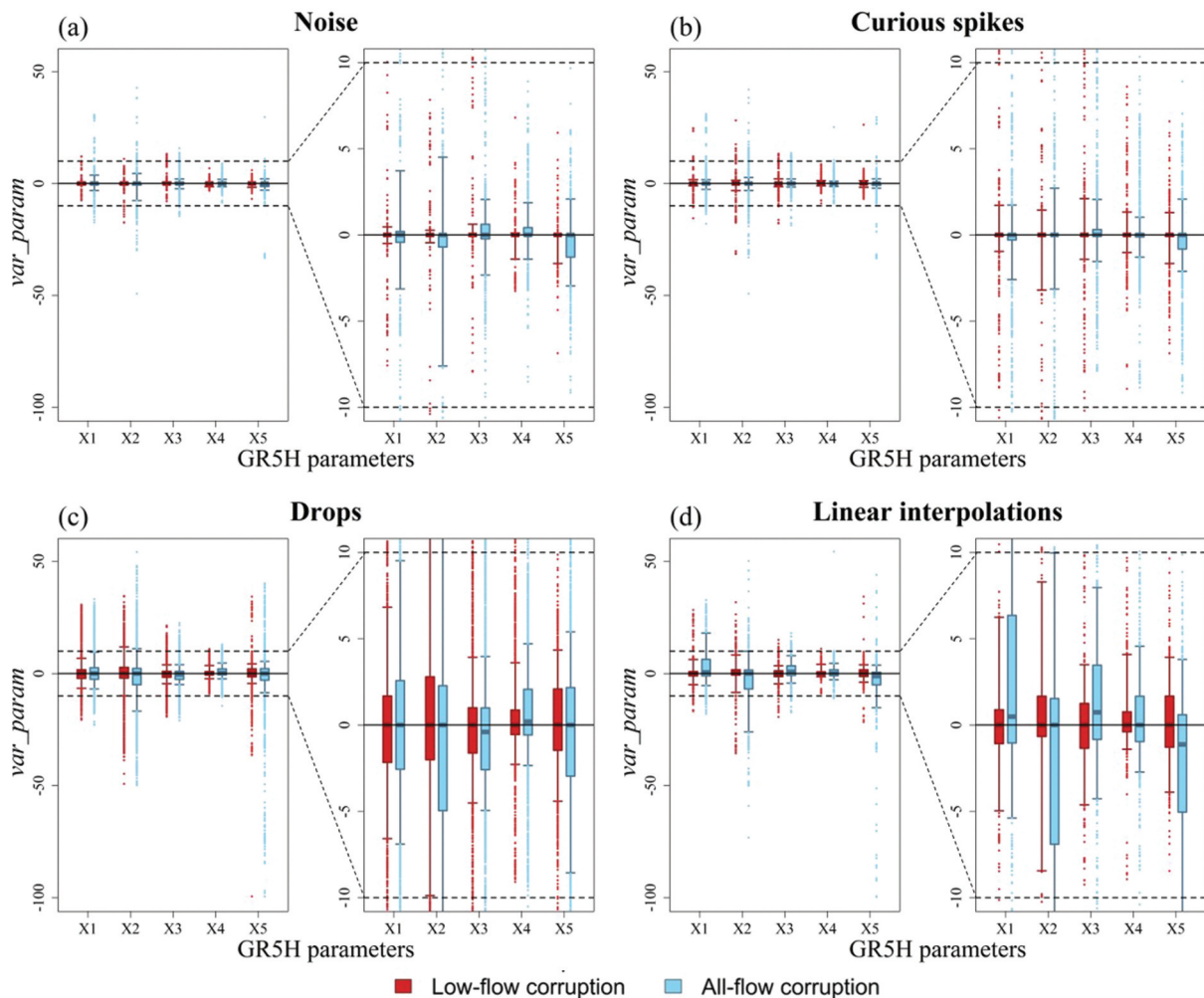


Figure 8. Distributions of parameter estimate variations according to the suspicious pattern. The box plots represent the 10th, 25th, 50th, 75th and 90th quantiles. They contain 1680 (a: noise), 3360 (b: spikes, c: drops) and 420 (d: linear interpolations) values each, including outliers. Red corresponds to corruption in the low-flow periods and blue to corruption without restriction in the flow range. Data corruption was performed over the calibration period only.

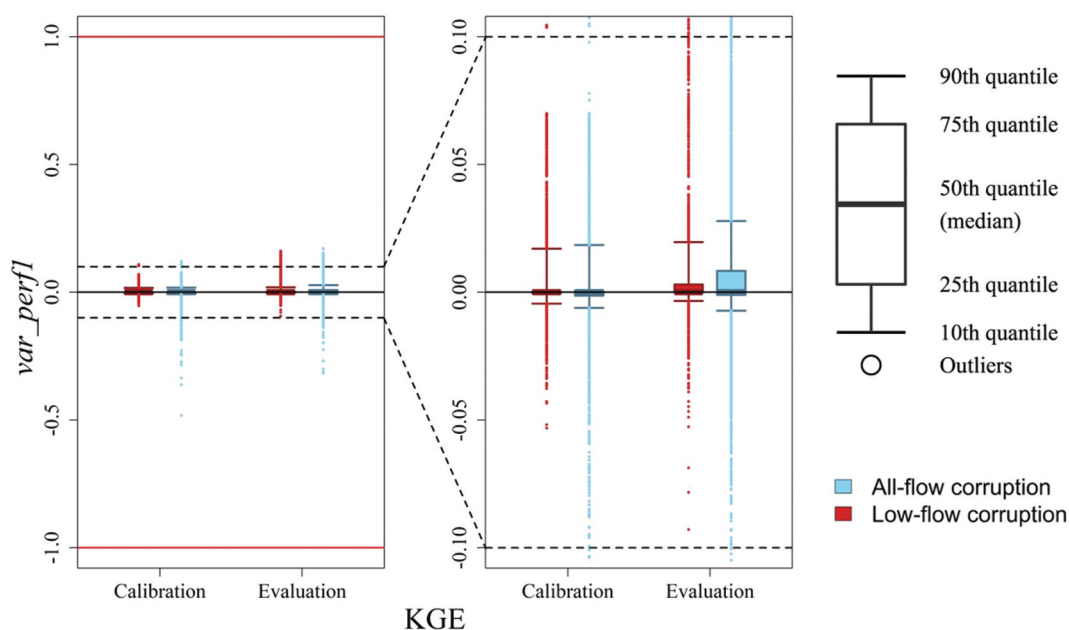


Figure 9. Distribution of the variation in KGE values (calculated on the square root of flows), in the calibration period (left) and the evaluation period (right). The box plots represent the 10th, 25th, 50th, 75th and 90th quantiles and contain 8820 values each, including outliers. Red corresponds to corruption in the low-flow periods and blue corresponds to corruption without restriction in the flow range. Data corruption was performed over the calibration period only. A positive value of the variation represents an improvement in performance after the corruption of streamflow records.

flows or the whole flow range. Evaluation performance seems to be more impacted than calibration performance. This result may be due to the specificities of the calibration and evaluation periods used in this study but may also show an impact of suspicious streamflow data on the model robustness. Note that, in a certain number of cases, the KGE tends to improve in calibration and evaluation following the implementation of the errors. However, this increase is negligible because the variation of the 90th quantile is very close to zero.

Figure 10 aims to differentiate the impact on the model performance of each suspicious pattern implemented in the streamflow time series. Noise and curious spikes have almost no impact on model performance. Once again, drops and linear interpolations seem to be responsible for the performance variations, although these remain very low.

Predicting total water availability is a frequently desired output of hydrological modelling. Figure 11 shows the difference with the initial value of the bias reformulated into β^2 in order to quantify the impact of the various suspicious patterns generated on the water balance. As for the KGE, the largest variations come from drops and linear interpolations. However, differences with the initial value remain limited, whether beneficial or detrimental.

3.2 Data corruption over the evaluation period

Here we focus on the impact of suspicious streamflow data on model performance (in evaluation) when the corruption was made only during the evaluation period. To this end, the variation used for this section was var_perf2 (Equation 4).

Figure 12 represents the distribution of performance variations obtained in the evaluation of the 15 catchments when the corruption is made only over the evaluation period. It shows

that, in general, suspicious streamflow data in the evaluation period have little (if any) influence on the performance criterion, although occasionally they can lead to a sharp decrease in performance when the corruption is applied without a preconceived notion on streamflow range. Note that the biases were also analysed but did not show significant variations (less than 0.04% for 90% of cases and reaching a maximum of 6%).

Figure 13 shows the previous results separated by type of error. It highlights that this loss of performance is mostly driven by linear interpolations (and, to a lesser extent, by drops), while noise and curious spikes seem to have no effect. It should be noted that for linear interpolations, the difference in performance impact between all-flow and low-flow corruption is much higher than for the other patterns.

4 Discussion

In this work, we decided to use actual measurements and not theoretical synthetic time series. Consequently, the actual streamflow (i.e. the streamflow that would have been observed if no error or influence had occurred) is not known and we thus needed to make the following hypotheses:

- (1) Each implemented suspicious pattern approximated the observed reality by a simple and parsimonious error model, constrained by one, two or three realistically varying parameters.
- (2) Each original streamflow record was assumed to represent the actual streamflow.

The parameters used to represent the suspicious patterns were carefully generated from a gamma distribution representative

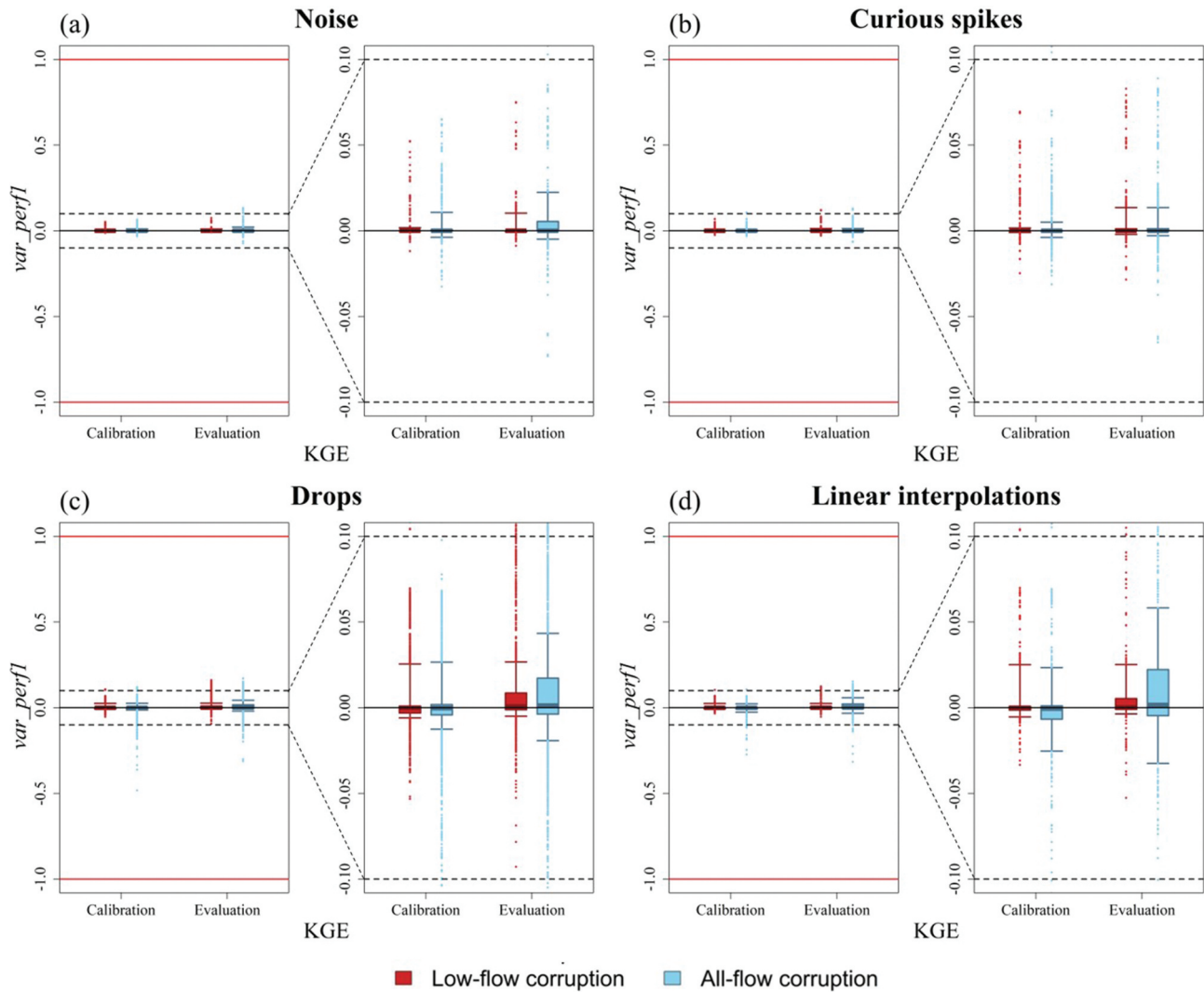


Figure 10. Distribution of the variation in KGE values (calculated on the square root of flows) according to the suspicious pattern, in the calibration period (left) and the evaluation period (right). The box plots represent the 10th, 25th, 50th, 75th and 90th quantiles. They contain 1680 (a: noise), 3360 (b: spikes, c: drops) and 420 (d: linear interpolations) values each, including outliers. Red corresponds to corruption in the low-flow periods and blue corresponds to corruption without restriction in the flow range. Data corruption was performed over the calibration period only. A positive value of the variation represents an improvement in performance after the corruption of streamflow records.

of the reality observed in the 15 catchments of the suspicious set (Section 2.3). A more detailed study with a larger number of catchments would allow us to define a more suitable distribution of these parameters, but this would be an extremely time-consuming task.

The results of the streamflow data corruption in the model (Section 3) highlight that suspicious patterns, especially during low-flow periods, generally have a limited impact on performance (KGE and bias) or parameter values. This finding must be put into perspective against the quality of the input data. Indeed, in France, although many suspicious data remain in the records, the data producers carry out a streamflow validation procedure to remove aberrations before making them available, which may explain the low impact on the efficiency and parameter estimates of rainfall–runoff models. However, the quality-checking methodology varies around the world and is sometimes lacking, resulting in a wide disparity

in the quality of the streamflow records in the different national archives. Extending the database to catchments located all over the world and not only in France could lead to different results, potentially widening the range of parameters of the error models and perhaps identifying new suspicious patterns. In addition, this finding also needs to be put into the context of the modelling choices. The criterion used for calibration or evaluation is the KGE on the square root of the flows. Although this criterion takes into account low flows, greater weight is still given to high flows (Santos *et al.* 2018). Given that 87% of the detected patterns are present during low-flow periods, it is likely that suspicious streamflow data would be more impactful for a study focusing on low-flow periods only. Moreover, all patterns except linear interpolations are implemented with a multiplicative coefficient k to represent the intensity of the suspicious pattern. This choice is intended to produce consistent relative variations

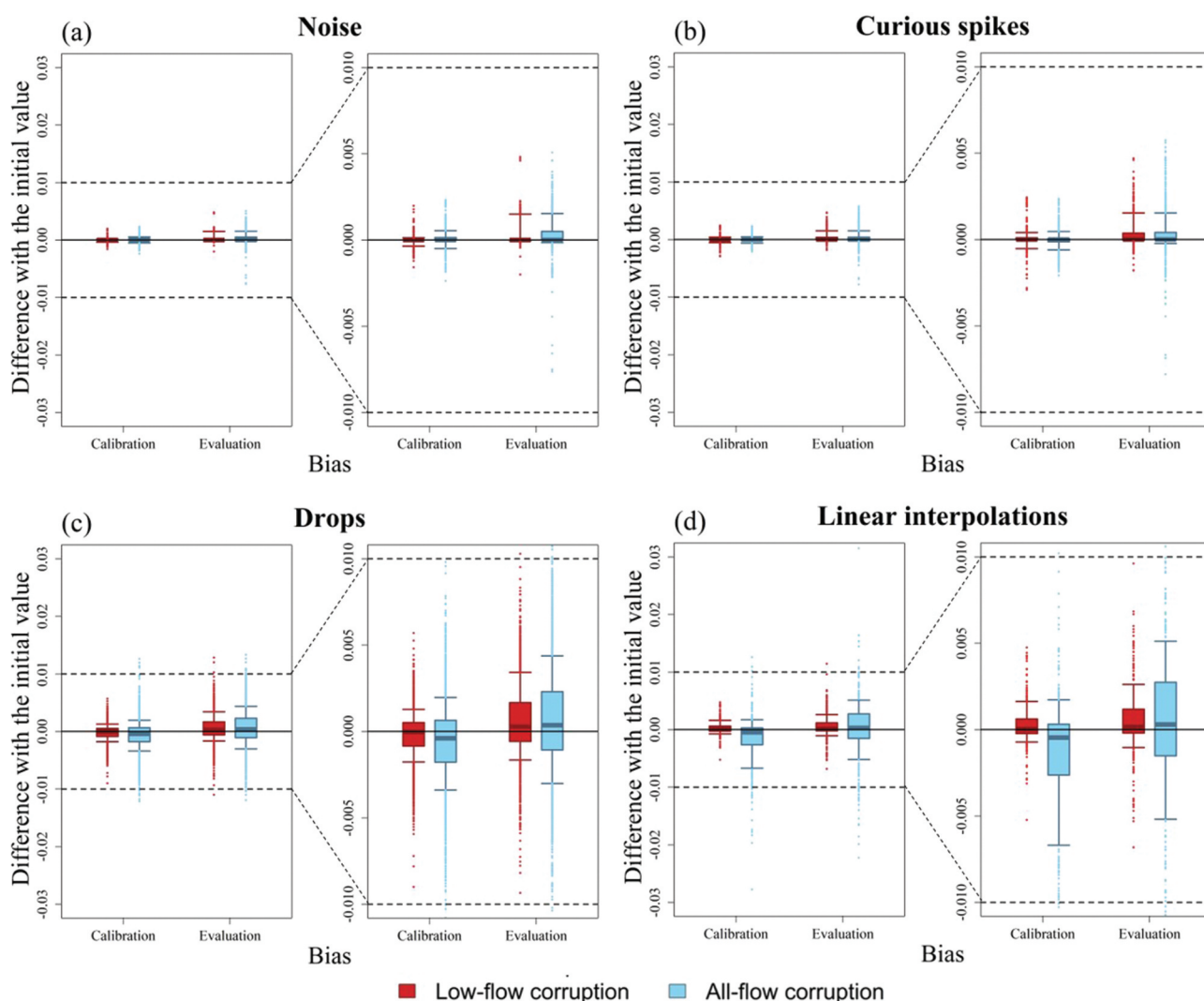


Figure 11. Distribution of the difference with the initial value in reformulated bias β' according to the suspicious pattern, in the calibration period (left) and the evaluation period (right). The box plots represent the 10th, 25th, 50th, 75th and 90th quantiles. They contain 1680 (a: noise), 3360 (b: spikes, c: drops) and 420 (d: linear interpolations) values each, including outliers. Red corresponds to corruption in the low-flow periods and blue corresponds to corruption without restriction in the flow range. Data corruption was performed over the calibration period. A positive value of the variation represents an improvement in performance after the corruption of streamflow records.

regardless of the streamflow range but also has the effect of amplifying the absolute error during floods. The linear interpolations appear to be much more sensitive to the range of streamflow compared with the other implemented patterns. The natural dynamics of the river can partly explain this. Indeed, during low-flow periods the river can have a stable behaviour over several weeks when it does not rain, which can be approximated by a straight line. On the other hand, this hypothesis is not suitable when phenomena fluctuating rapidly in time, such as floods, are taken into account. The results of the streamflow data corruption in the model also highlight that suspicious patterns might occasionally have a strong impact on performance or parameter values. However, a remaining issue is that there is no clear trend between error characteristics (intensity, time window affected or number of patterns existing in the time series) and their impact on the efficiency and the parameter estimates of rainfall-runoff models. Of course, the more pessimistic the scenario,

the more likely the drop in performance. However, it is difficult to anticipate whether a suspicious pattern will have an impact or not because it is also largely dependent on its location in the corrupted time series. Figure 14 highlights this phenomenon: counterintuitively, the KGE is here much more affected by the less pessimistic scenario (-20% vs. -1%).

Data corruption over the calibration period mainly leads to a variation of the X_1 , X_2 and X_5 parameters (Section 3.1.1). While it is possible that suspicious streamflow data can lead to a variation in the parameter X_1 , i.e. the production part of the model, the fluctuations of X_2 and X_5 are quite surprising because implemented patterns should not have such an impact on the underground flows. Here, these parameters seem to deviate from their original function and instead aim to close the water balance of the system modified by the data corruption. It should also be noted that these parameter estimates are usually close to 0 (Table 3). The relative variation in regard to the initial

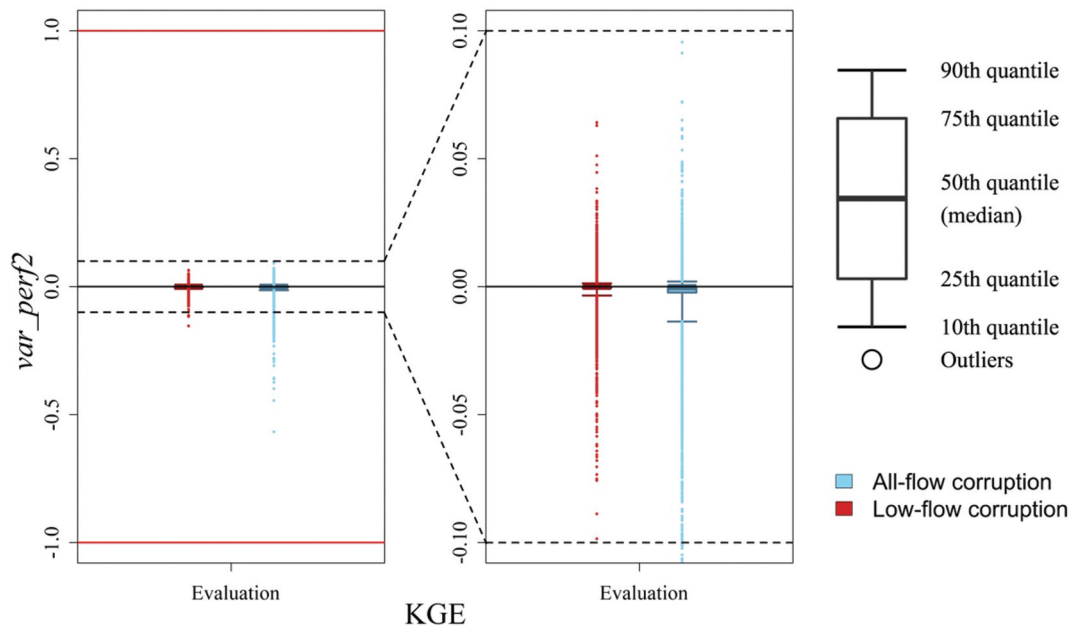


Figure 12. Distribution of KGE on the square root of the flow variations, in the evaluation period. The box plots represent the 10th, 25th, 50th, 75th and 90th quantiles and contain 8820 values each, including outliers. Red corresponds to corruption in the low-flow periods and blue to corruption without restriction in the flow range. Data corruption was performed over the evaluation period only. A positive value of the variation represents an improvement in performance after the corruption of streamflow records.

value, therefore, tends to be more sensitive than for the other parameters.

Finally, suspicious data lead to a stronger variation in performance, although they remain small, when they appear during the evaluation period (Section 3.2). A large variation in performance values between the calibration and evaluation period is often attributed to the model robustness, while this phenomenon can be directly inherited from the suspicious streamflow data.

5 Conclusion

The main conclusions of this work are as follows:

- (1) On average, over several catchments, the errors (noise, curious spikes, drops and linear interpolations) affecting commonly available streamflow time series for hydrological studies have a limited impact on model results. However, they may still be an important source of model instability and lack of robustness when working on a single catchment.
- (2) Amongst the four types of model error tested, drops and linear interpolations have a higher impact on model results in specific catchments, in terms of either performance or parameter estimates. However, there is no clear trend between the characteristics of the errors (intensity, time window affected or number of patterns existing in the time series) and the impact on hydrological modelling.
- (3) The suspicious patterns seem to more greatly affect the production function of the model (i.e. the part of the model responsible for adjusting the catchment water balance) than the routing part. The

parameters initially intended to control underground water exchanges can be utilized to compensate for the loss or gain of volume to close the water balance.

- (4) Model outputs seem to be more impacted by suspicious streamflow data during high-flow periods. However, this result is dependent on the modelling framework. Here, the objective function and the evaluation criterion used is the KGE on the square root of the streamflow, which makes it possible to take into account the whole range of flows but gives a greater weight to the high-flow values.

These results suggest that, when working on several catchments, using time series directly retrieved from an existing database without further quality checks may have only a limited impact on the overall modelling results. Therefore, the time-consuming task of checking the streamflow series a second time may not be necessary. However, note that we say “a second time”, because we assume that this work has already been done by the hydro-metric services in charge of streamflow data collection.

The results were obtained here with a single hydrological model and using data from the French national flow archive, which has its own quality check rules. Further tests may be necessary to generalize these results to other models or datasets and under other conditions (e.g. typically different objective functions, the presence of gaps in data or the time window of calibration–evaluation periods). It would also be useful to evaluate the relative role of errors in flows in comparison with other variables used in the model, typically precipitation and potential evapotranspiration (see the study by Oudin *et al.* 2006). This would help in offering recommendations regarding prioritizing where efforts should be made when checking datasets for large-sample hydrology studies.

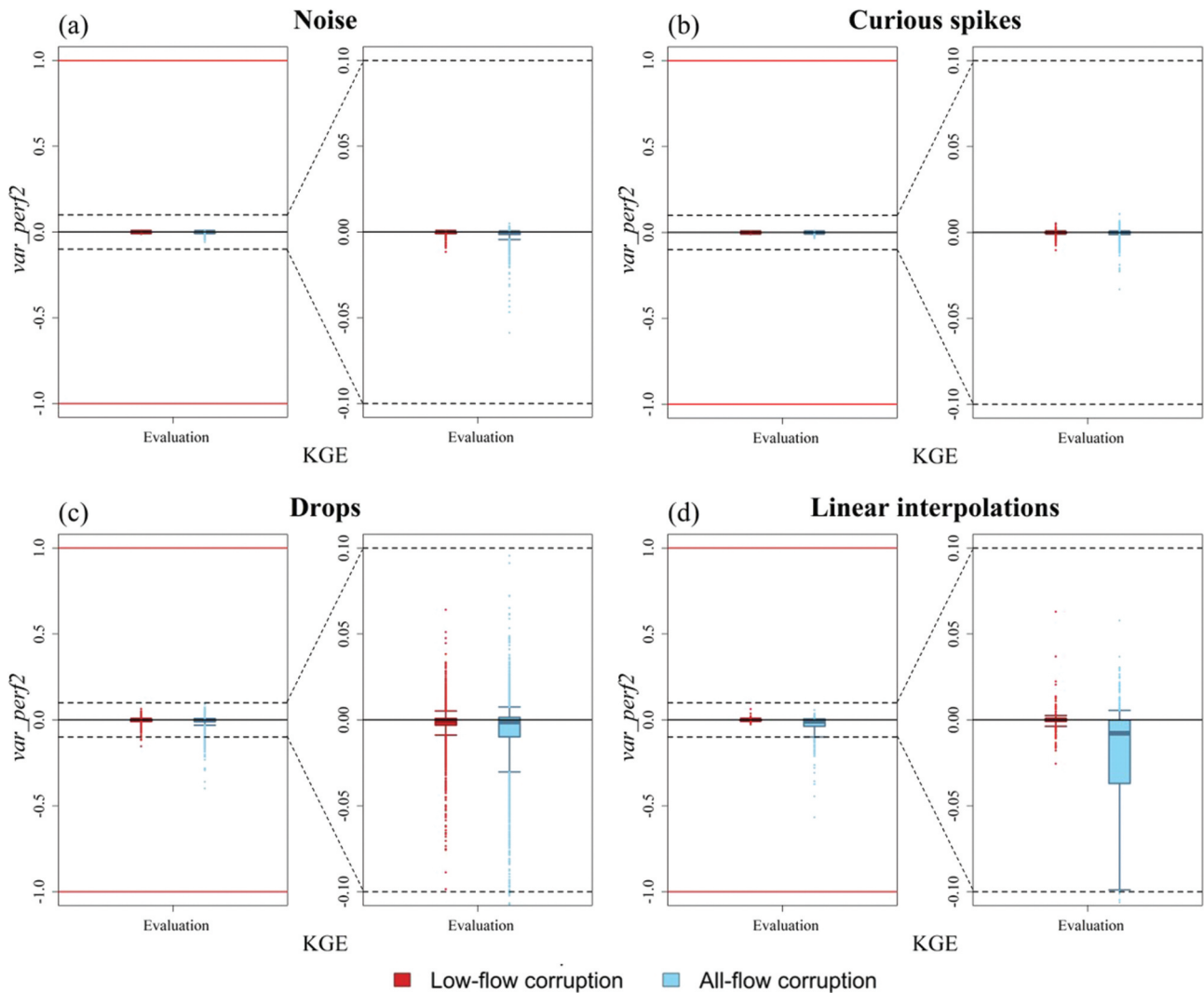


Figure 13. Distribution of the variation in KGE values (calculated on the square root of flows) according to the suspicious pattern, in the evaluation period. The box plots represent the 10th, 25th, 50th, 75th and 90th quantiles. They contain 1680 (a: noise), 3360 (b: spikes, c: drops) and 420 (d: linear interpolations) values each, including outliers. Red corresponds to corruption in the low-flow periods and blue to corruption without restriction in the flow range. Data corruption was performed over the evaluation period only. A positive value of the variation represents an improvement in performance after the corruption of streamflow records.

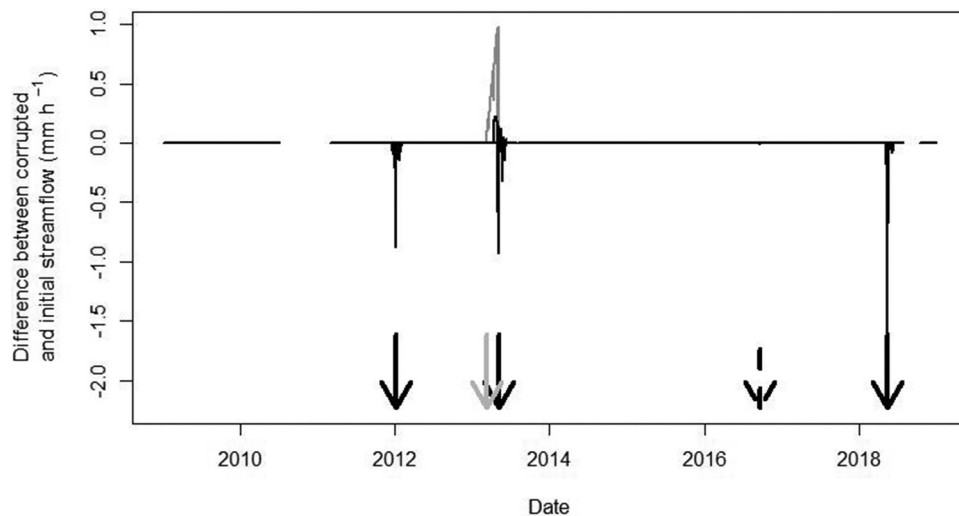


Figure 14. Difference between corrupted and initial streamflow record for the River La Clauge at La Loye during the evaluation period for a linear interpolation over 2 months with n (number of patterns) equal to one (grey) and four (black). Arrows represent the corrupted time steps and dotted lines mean that only low flow is affected.

Acknowledgements

The authors thank Météo-France and the SCHAPI (managing the Banque Hydro website), for making available the climate and hydrological data used in this study. Thibault Hallouin and Laurent Strohmenger are also thanked, for their advice, and Isabella Athanassiou for copyediting the manuscript. The authors thank the Editor Stacey Archfield, the Associate Editor Pierre Brigode, David Post and an anonymous reviewer for their feedback and comments on the manuscript, which helped to improve its overall quality.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE) and Compagnie Nationale du Rhône (CNR), which provided financial support to the PhD grant of the first author.

ORCID

Cyril Thébault  <http://orcid.org/0000-0002-4259-0497>
 Charles Perrin  <http://orcid.org/0000-0001-8552-1881>
 Vazken Andréassian  <http://orcid.org/0000-0001-7124-9303>
 Guillaume Thirel  <http://orcid.org/0000-0002-1444-1830>
 Olivier Delaigue  <http://orcid.org/0000-0002-7668-8468>

References

- Andréassian, V., *et al.*, 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *Journal of Hydrology*, 250 (1), 206–223. doi:10.1016/S0022-1694(01)00437-1
- Andréassian, V., *et al.*, 2006. Introduction and synthesis: why should hydrologists work on a large number of basin data sets? In: *Large sample basin experiments for hydrological parametrization: results of the models parameter experiment- MOPEX*. IAHS Red Books Series, IAHS, Wallingford, UK, Vol. 307. 1–5. Available from: https://iahs.info/uploads/dms/13599_02-1-6-INTRODUCTION.pdf [Accessed 22 Jul 2022].
- Andréassian, V., Perrin, C., and Michel, C., 2004. Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models. *Journal of Hydrology*, 286 (1), 19–35. doi:10.1016/j.jhydrol.2003.09.030
- Artigue, G., *et al.*, 2012. Flash flood forecasting in poorly gauged basins using neural networks: case study of the Gardon de Mialet basin (southern France). *Natural Hazards and Earth System Sciences*, 12 (11), 3307–3324. doi:10.5194/nhess-12-3307-2012
- Beven, K. and Smith, P., 2015. Concepts of information content and likelihood in parameter calibration for hydrological simulation models. *Journal of Hydrologic Engineering*, 20 (1), A4014010. doi:10.1061/(ASCE)HE.1943-5584.0000991
- Beven, K., Smith, P.J., and Wood, A., 2011. On the colour and spin of epistemic error (and what we might do about it). *Hydrology and Earth System Sciences*, 15 (10), 3123–3133. doi:10.5194/hess-15-3123-2011
- Beven, K. and Westerberg, I.K., 2011. On red herrings and real herrings: disinformation and information in hydrological inference. *Hydrological Processes*, 25 (10), 1676–1680. doi:10.1002/hyp.7963
- Borah, A. and Haan, C.T., 1991. Impact of random data errors on parameter estimation in hydrologic modeling. *Transactions of the ASAE (American Society of Agricultural Engineering)*, 34 (3), 0857–0864. doi:10.13031/2013.31740
- Bourgin, F., *et al.*, 2014. Investigating the interactions between data assimilation and post-processing in hydrological ensemble forecasting. *Journal of Hydrology*, 519, 2775–2784. doi:10.1016/j.jhydrol.2014.07.054
- Coron, L., *et al.*, 2017. The suite of lumped GR hydrological models in an R package. *Environmental Modelling & Software*, 94, 166–171. doi:10.1016/j.envsoft.2017.05.002
- Coron, L., *et al.*, 2021. *airGR: suite of GR Hydrological Models for Precipitation-Runoff Modelling*. R package version 1.6.12. Antony, France: Portail Data INRAE, INRAE. doi:10.15454/EX11NA
- Dancho, M. and Vaughan, D., 2020. *Anomalize: tidy Anomaly Detection*. R package version 0.2.2. Available from: <https://CRAN.R-project.org/package=anomalize> [Accessed 22 Jul 2022].
- Darienzo, M., *et al.*, 2021. Detection of stage-discharge rating shifts using gaugings: a recursive segmentation procedure accounting for observational and model uncertainties. *Water Resources Research*, 57 (4), e2020WR028607. doi:10.1029/2020WR028607
- Delaigue, O., *et al.*, 2020. *Database of watershed-scale hydroclimatic observations in France*. Antony: INRAE, HYCAR Research Unit, Hydrology group. Available from: <https://webgr.inrae.fr/activites/base-de-donnees/> [Accessed 22 Jul 2022].
- Despax, A., *et al.*, 2019. Decomposition of uncertainty sources in acoustic doppler current profiler streamflow measurements using repeated measures experiments. *Water Resources Research*, 55 (9), 7520. doi:10.1029/2019WR025296
- Di Baldassarre, G. and Montanari, A., 2009. Uncertainty in river discharge observations: a quantitative analysis. *Hydrology and Earth System Sciences*, 13 (6), 913–921. doi:10.5194/hess-13-913-2009
- Elshorbagy, A., 2001. Noise reduction approach in chaotic hydrologic time series revisited. *Canadian Water Resources Journal/Revue Canadienne Des Ressources Hydriques*, 26 (4), 537–550. doi:10.4296/cwrj2604537
- Engeland, K., *et al.*, 2016. Effects of uncertainties in hydrological modelling. A case study of a mountainous catchment in Southern Norway. *Journal of Hydrology*, 536, 147–160. doi:10.1016/j.jhydrol.2016.02.036
- Ficchi, A., 2017. *An adaptive hydrological model for multiple time-steps: diagnostics and improvements based on fluxes consistency*. Thesis (PhD). Université Pierre et Marie Curie - Paris VI, 27 February. Available from: <https://tel.archives-ouvertes.fr/tel-01619102>
- Ficchi, A., Perrin, C., and Andréassian, V., 2019. Hydrological modelling at multiple sub-daily time steps: model improvement via flux-matching. *Journal of Hydrology*, 575, 1308–1327. doi:10.1016/j.jhydrol.2019.05.084
- Gupta, H.V., *et al.*, 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology*, 377 (1), 80–91. doi:10.1016/j.jhydrol.2009.08.003
- Gupta, H.V., *et al.*, 2014. Large-sample hydrology: a need to balance depth with breadth. *Hydrology and Earth System Sciences*, 18 (2), 463–477. doi:10.5194/hess-18-463-2014
- Hamilton, A.S. and Moore, R.D., 2012. Quantifying uncertainty in streamflow records. *Canadian Water Resources Journal/Revue Canadienne Des Ressources Hydriques*, 37 (1), 3–21. doi:10.4296/cwrj3701865
- Horner, I., *et al.*, 2018. Impact of stage measurement errors on streamflow uncertainty. *Water Resources Research*, 54 (3), 1952–1976. doi:10.1002/2017WR022039
- Ibbitt, R.P., 1972. Effects of random data errors on the parameter values for a conceptual model. *Water Resources Research*, 8 (1), 70–78. doi:10.1029/WR008i001p00070
- Karunasingha, D.S.K. and Liong, S.-Y., 2018. Enhancement of chaotic hydrological time series prediction with real-time noise reduction using Extended Kalman Filter. *Journal of Hydrology*, 565, 737–746. doi:10.1016/j.jhydrol.2018.08.044
- Klemeš, V., 1986. Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31 (1), 13–24. doi:10.1080/02626668609491024
- Le Coz, J., *et al.*, 2014. Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: a Bayesian approach. *Journal of Hydrology*, 509, 573–587. doi:10.1016/j.jhydrol.2013.11.016
- Leleu, I., *et al.*, 2014. La refonte du système d'information national pour la gestion et la mise à disposition des données hydrométriques. *La Houille Blanche*, 1 (1), 25–32. doi:10.1051/lhb/2014004
- Le Moine, N., 2008. *Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performances et du réalisme des modèles pluie-débit?* These de doctorat. Paris 6. Available from: <https://www.theses.fr/2008PA066468> [Accessed 22 Jul 2022].

- Lerat, J., *et al.*, 2012. Do internal flow measurements improve the calibration of rainfall-runoff models? *Water Resources Research*, 48 (2). doi:10.1029/2010wr010179
- Lobligeois, F., 2014. *Mieux connaître la distribution spatiale des pluies améliore-t-il la modélisation des crues? Diagnostic sur 181 bassins versants français*. These de doctorat. AgroParisTech.
- Lobligeois, F., *et al.*, 2014. When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. *Hydrology and Earth System Sciences*, 18 (2), 575–594. doi:10.5194/hess-18-575-2014
- Mansanarez, V., *et al.*, 2019. Shift happens! Adjusting stage-discharge rating curves to morphological changes at known times. *Water Resources Research*, 55 (4), 2876–2899. doi:10.1029/2018WR023389
- McMillan, H.K., *et al.*, 2010a. Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. *Hydrological Processes*, 24 (10), 1270–1284. doi:10.1002/hyp.7587
- McMillan, H.K., *et al.*, 2010b. Input uncertainty in hydrological models: an evaluation of error models for rainfall. *Journal of Hydrology*, 400 (1–2), 83–94. doi:10.1016/j.jhydrol.2011.01.026
- McMillan, H.K., *et al.*, 2017. How uncertainty analysis of streamflow data can reduce costs and promote robust decisions in water management applications. *Water Resources Research*, 53 (7), 5220–5228. doi:10.1002/2016WR020328
- Montanari, A., *et al.*, 2013. Panta rhei—everything flows”: change in hydrology and society—the IAHS scientific decade 2013–2022. *Hydrological Sciences Journal*, 58 (6), 1256–1275. doi:10.1080/02626667.2013.809088
- Montanari, A. and Di Baldassarre, G., 2013. Data errors and hydrological modelling: the role of model structure to propagate observation uncertainty. *Advances in Water Resources*, 35th Year Anniversary Issue 51, 498–504. doi:10.1016/j.advwatres.2012.09.007
- Oudin, L., *et al.*, 2005. Which potential evapotranspiration input for a lumped rainfall-runoff model?: part 2—Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *Journal of Hydrology*, 303 (1–4), 290–306. doi:10.1016/j.jhydrol.2004.08.026
- Oudin, L., *et al.*, 2006. Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. *Journal of Hydrology the Model Parameter Estimation Experiment*, 320 (1), 62–83. doi:10.1016/j.jhydrol.2005.07.016
- Paturel, J.E., Servat, E., and Vassiliadis, A., 1995. Sensitivity of conceptual rainfall-runoff algorithms to errors in input data — case of the GR2M model. *Journal of Hydrology*, 168 (1), 111–125. doi:10.1016/0022-1694(94)02654-T
- Pelletier, P.M., 1988. Uncertainties in the single determination of river discharge: a literature review. *Canadian Journal of Civil Engineering*, 15 (5), 834–850. doi:10.1139/l88-109
- Peña-Arancibia, J.L., *et al.*, 2015. Streamflow rating uncertainty: characterisation and impacts on model calibration and performance. *Environmental Modelling & Software*, 63, 32–44. doi:10.1016/j.envsoft.2014.09.011
- Perret, E., Renard, B., and Le Coz, J., 2021. A rating curve model accounting for cyclic stage-discharge shifts due to seasonal aquatic vegetation. *Water Resources Research*, 57 (3), e2020WR027745. doi:10.1029/2020WR027745
- Perrin, C., 2000. *Vers une amélioration d'un modèle global pluie-débit*. Thesis (PhD). Institut National Polytechnique de Grenoble - INPG. Available from: <https://tel.archives-ouvertes.fr/tel-00006216> [Accessed 22 Jul 2022].
- Perrin, C., *et al.*, 2007. Impact of limited streamflow data on the efficiency and the parameters of rainfall—runoff models. *Hydrological Sciences Journal*, 52 (1), 131–151. doi:10.1623/hysj.52.1.131
- Peters, A., *et al.*, 2014. Separating precipitation and evapotranspiration from noise – a new filter routine for high-resolution lysimeter data. *Hydrology and Earth System Sciences*, 18 (3), 1189–1198. doi:10.5194/hess-18-1189-2014
- Puechbert, R., *et al.*, 2017. *Charte qualité de l'hydrométrie. Guide de bonnes pratiques*, 83. Ministère de l'environnement, de l'énergie et de la mer, France. Available from: <https://hal.inrae.fr/hal-02606950> [Accessed 22 Jul 2022].
- Reitan, T. and Petersen-Øverleir, A., 2009. Bayesian methods for estimating multi-segment discharge rating curves. *Stochastic Environmental Research and Risk Assessment*, 23 (5), 627–642. doi:10.1002/2016WR020328
- Ruelland, D., 2020. Should altitudinal gradients of temperature and precipitation inputs be inferred from key parameters in snow-hydrological models? *Hydrology and Earth System Sciences*, 24 (5), 2609–2632. doi:10.5194/hess-24-2609-2020
- Saadi, M., Oudin, L., and Ribstein, P., 2021. Physically consistent conceptual rainfall-runoff model for urbanized catchments. *Journal of Hydrology*, 599, 126394. doi:10.1016/j.jhydrol.2021.126394
- Saltelli, A., Tarantola, S., and Campolongo, F., 2000. Sensitivity analysis as an ingredient of modeling. *Statistical Science*, 15 (4), 377–395. doi:10.1214/ss/1009213004
- Santos, L., Thirel, G., and Perrin, C., 2018. Technical note: pitfalls in using log-transformed flows within the KGE criterion. *Hydrology and Earth System Sciences*, 22 (8), 4583–4591. doi:10.5194/hess-22-4583-2018
- Schouten, J.C., Takens, F., and van den Bleek, C.M., 1994. Estimation of the dimension of a noisy attractor. *Physical Review E*, 50 (3), 1851–1861. doi:10.1103/PhysRevE.50.1851
- Schreiner-McGraw, A.P. and Ajami, H., 2022. Combined impacts of uncertainty in precipitation and air temperature on simulated mountain system recharge from an integrated hydrologic model. *Hydrology and Earth System Sciences*, 26 (4), 1145–1164. doi:10.5194/hess-26-1145-2022
- Seibert, J. and Westerberg, I.K., 2022. Hunting after shoals of red herrings - cataloguing disinformative data and their causes. *Presented at the IAHS*. Montpellier.
- Singh, S.K. and Dutta, S., 2017. Observational uncertainty in hydrological modelling using data depth. *Global Nest Journal*, 19 (3), 489–497. doi:10.30955/gnj.002354
- Tabary, P., *et al.*, 2012. A 10-year (1997–2006) reanalysis of quantitative precipitation estimation over France: methodology and first results. *IAHS-AISH Publication*, 351, 255–260.
- Terrier, M., *et al.*, 2021. Streamflow naturalization methods: a review. *Hydrological Sciences Journal*, 66 (1), 12–36. doi:10.1080/02626667.2020.1839080
- Tillaart, S., Booij, M., and Krol, M., 2013. Impact of uncertainties in discharge determination on the parameter estimation and performance of a hydrological model. *Hydrology Research*, 44 (3), 454–466. doi:10.2166/nh.2012.147
- Troutman, B.M., 1985. Errors and parameter estimation in precipitation-runoff modeling: 2. Case Study. *Water Resources Research*, 21 (8), 1214–1222. doi:10.1029/WR021i008p01214
- Vallis, O.S., Hochenbaum, J., and Kejarawal, A., 2014. *AnomalyDetection: anomaly detection using seasonal hybrid extreme studentized deviate test. R package version 1.0*. Available from: <https://github.com/twitter/AnomalyDetection> [Accessed 22 Jul 2022].
- van Esse, W.R., *et al.*, 2013. The influence of conceptual model structure on model performance: a comparative study for 237 French catchments. *Hydrology and Earth System Sciences*, 17 (10), 4227–4239. doi:10.5194/hess-17-4227-2013
- Vidal, J.-P., *et al.*, 2010. A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *International Journal of Climatology*, 30 (11), 1627–1644. doi:10.1002/joc.2003
- Wagener, T., *et al.*, 2021. On doing hydrology with dragons: realizing the value of perceptual models and knowledge accumulation. *WIREs Water*, 8 (6), e1550. doi:10.1002/wat2.1550
- Wang, H.-W., *et al.*, 2019a. Filtering continuous river surface velocity radar data. *Water*, 11 (4), 764. doi:10.3390/w11040764
- Wang, X., Pan, M., and Xia, C., 2019b. Water level data preprocessing method based on savitzky-golay filter, 288–294. *Presented at the 2019 International Conference on Modeling, Simulation and Big Data Analysis* (MSBDA Conference, Wuhan, China, 23–24 June 2019). Atlantis Press. doi:10.2991/msbda-19.2019.44

Westerberg, I.K., *et al.*, 2020. Hydrological model calibration with uncertain discharge data. *Hydrological Sciences Journal*, 1–16. doi:10.1080/02626667.2020.1735638

Wilby, R.L., *et al.*, 2017. The ‘dirty dozen’ of freshwater science: detecting then reconciling hydrological data biases and errors. *WIREs Water*, 4 (3), e1209. doi:10.1002/wat2.1209

Appendices

Appendix A.

Table A1. Main characteristics of the 15 catchments included in the suspicious set.

Suspicious set					
Code	Name	Area [km ²]	Average annual precipitation [mm y ⁻¹]	Average annual potential evapotranspiration [mm y ⁻¹]	Average annual streamflow [mm y ⁻¹]
A8006210	Vaux stream at Morgemoulin	42	863	652	264
B4631010	Chiers at Carignan	1978	914	651	391
H4223110	Remarde River at Saint-Cyr-sous-Dourdan	152	680	690	116
I2001010	Dives River at Saint-Lambert-sur-Dive	88	783	667	144
J2233020	Léguer River at Pluzunet	362	1067	652	512
K1753110	Alène River at Cercy-la-Tour [Coueron]	333	1035	699	398
M3423010	Jouanne River at Forcé	404	796	705	239
O4704030	Dadou River at Paulinet [Saint-Jean-de-Jeanne]	71	1140	712	557
O7035010	Bramont River at Saint-Bauzile [Les Fonts]	117	1036	619	424
P3001010	Vézère River at Saint-Merd-les-Oussines [Maisonnial]	42	1384	614	985
S2134010	Petite Leyre River at Belhade	413	958	800	225
U1204010	Tille River at Crécey-sur-Tille	230	941	674	351
U2345040	Savoureuse River at Vieux-Charmont	235	1437	668	790
V1315050	Leyse River at la Ravoire	87	1506	631	836
Y1364010	Fresquel River at Carcassonne [Pont Rouge]	936	845	829	173

Appendix B.

Table B1. Main characteristics of the 15 catchments included in the clean set.

Clean set					
Code	Name	Area [km ²]	Average annual precipitation [mm y ⁻¹]	Average annual potential evapotranspiration [mm y ⁻¹]	Average annual streamflow [mm y ⁻¹]
A4632010	Avière River at Frizon	103	992	668	360
E4907005	Peene Becque River at Ochteezele	55	755	648	253
H5033310	Suize River at Villiers-sur-Suize	83	965	658	401
H6221010	Aisne River at Givry	2887	883	667	322
J3601810	Aulne River at Scrignac [Le Goask]	117	1196	655	569
K1341810	Arroux River at Rigny-sur-Arroux	2271	978	684	326
L8000020	Loire River at Saumur	81 315	844	707	227
N0113010	Falleron River at Falleron	33	817	748	313
O1584610	Douctouyre River at Dun	136	1056	763	354
P7001510	Isle River at Bassilac [Charrieras]	1859	1048	754	362
S2242510	Eyre River at Salles	1678	981	801	273
U2425260	Cusancin River at Baume-les-Dames	541	1310	664	396
U2565020	Clauge River at La Loye	119	1092	729	264
V0355010	Foron River at Sciez	51	1222	665	465
V4025010	Embroye River at Toulaud	8	1027	745	457

Appendix C.

Table C1. Distribution of the parameter values obtained before and after data corruption.

	X_1 [mm]		X_2 [mm h ⁻¹]		X_3 [mm]		X_4 [h]		X_5 [-]	
	Before	After (Cal)	Before	After (Cal)	Before	After (Cal)	Before	After (Cal)	Before	After (Cal)
min	71	66	-0.36	-0.40	20	19	1	1	0.05	0
Q10	101	93	-0.25	-0.27	50	44	4	3	0.07	0.06
Q25	148	151	-0.17	-0.21	64	62	6	6	0.09	0.09
Q50	180	180	-0.09	-0.09	116	116	11	11	0.12	0.12
Q75	221	232	-0.08	-0.08	173	186	18	19	0.17	0.16
Q90	342	391	-0.04	-0.04	216	226	47	48	0.19	0.19
max	628	664	-0.04	-0.02	349	427	68	77	0.24	0.24

Appendix D.

Table D1. Distribution of the performance values obtained before and after data corruption depending on the calibration and evaluation period.

	KGE [-]						Bias β [-]					
	Before		After (Cal)		After (Eval)		Before		After (Cal)		After (Eval)	
	Cal	Eval	Cal	Eval	Cal	Eval	Cal	Eval	Cal	Eval	Cal	Eval
min	0.86	0.83	0.86	0.77	0.86	0.65	0.99	0.91	0.97	0.89	0.99	0.88
Q10	0.88	0.85	0.87	0.84	0.87	0.84	0.99	0.93	0.99	0.93	0.99	0.93
Q25	0.91	0.88	0.91	0.88	0.91	0.88	0.99	0.95	0.99	0.95	0.99	0.95
Q50	0.92	0.91	0.92	0.91	0.92	0.91	1.00	1.00	1.00	1.01	1.00	1.01
Q75	0.96	0.93	0.97	0.95	0.97	0.95	1.00	1.02	1.00	1.03	1.00	1.03
Q90	0.97	0.96	0.97	0.97	0.97	0.97	1.00	1.03	1.01	1.03	1.00	1.03
max	0.98	0.97	0.98	0.98	0.98	0.97	1.00	1.04	1.04	1.04	1.00	1.05