



HAL
open science

Detection of Stage-Discharge Rating Shifts Using Gaugings: A Recursive Segmentation Procedure Accounting for Observational and Model Uncertainties

M. Darienzo, Benjamin Renard, Jérôme Le Coz, Michel Lang

► **To cite this version:**

M. Darienzo, Benjamin Renard, Jérôme Le Coz, Michel Lang. Detection of Stage-Discharge Rating Shifts Using Gaugings: A Recursive Segmentation Procedure Accounting for Observational and Model Uncertainties. *Water Resources Research*, 2021, 57 (4), 10.1029/2020WR028607 . hal-03318630

HAL Id: hal-03318630

<https://hal.science/hal-03318630v1>

Submitted on 10 Aug 2021

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.

Water Resources Research

RESEARCH ARTICLE

10.1029/2020WR028607

Key Points:

- We propose a method for detecting rating shifts through the segmentation of residuals between the gaugings and a reference rating curve
- The method accounts for observational and rating curve uncertainties and expresses change points in terms of time (rather than position)
- The method recursively re-estimates the reference rating curve to improve the detection of small shifts

Correspondence to:

M. Darienzo,
matteo.darienzo@inrae.fr

Citation:

Darienzo, M., Renard, B., Le Coz, J., & Lang, M. (2021). Detection of stage-discharge rating shifts using gaugings: A recursive segmentation procedure accounting for observational and model uncertainties. *Water Resources Research*, 57, e2020WR028607. <https://doi.org/10.1029/2020WR028607>

Received 14 AUG 2020
Accepted 18 MAR 2021

Detection of Stage-Discharge Rating Shifts Using Gaugings: A Recursive Segmentation Procedure Accounting for Observational and Model Uncertainties

M. Darienzo¹ , B. Renard¹ , J. Le Coz¹ , and M. Lang¹ 

¹INRAE, UR RiverLy, Villeurbanne, France

Abstract The stage-discharge rating curve is subject at many hydrometric stations to sudden changes (shifts) typically caused by morphogenic floods. We propose an original method for estimating shift times using the stage-discharge observations, also known as gaugings. This method is based on a recursive segmentation procedure that accounts for both gaugings and rating curve uncertainties through a Bayesian framework. It starts with the estimation of a baseline rating curve using all available gaugings. Then it computes the residuals between the gaugings and this rating curve with uncertainties. It proceeds with the segmentation of the time series of residuals through a multi-change point Bayesian estimation accounting for residuals uncertainties. Once the first set of shift times is identified, the same procedure is recursively applied to each sub-period through a “top-down” approach searching for all effective shifts. The proposed method is illustrated using the Ardèche River at Meyras in France (a typical hydrometric site subject to river bed degradation) and evaluated using several synthetic data sets for which the true shift times are known. The applications confirm the added value of the recursive segmentation compared with a “single-pass” approach and highlight the importance of properly accounting for uncertainties in the segmented data. The recursive procedure effectively disentangles rating changes from observational and rating curve uncertainties.

Plain Language Summary For many hydrological and hydraulic issues, such as flood forecasting, a reliable river discharge estimate is needed. In general, discharge is derived from the recorded water level (stage) through a stage-discharge relation (rating curve). This relation is calibrated using direct observations (gaugings). Unfortunately, the rating curve is not only uncertain but it can also be subject to sudden changes or shifts due for example to intense floods that modify the river bed geometry. One solution to identify periods of rating curve stability is to apply a segmentation procedure to the gaugings. We propose in this paper an original recursive segmentation procedure that accounts for both gaugings and rating curve uncertainties.

1. Introduction

1.1. Rating Curves

River discharge, or streamflow, is one of the most important variables in hydrology and hydraulics. Hydrometric data are essential for the calibration of hydrological models, flood forecasting and warning, engineering design (of dam and bridges for example), and policy decisions related to water resources and environmental management. However, streamflow time series are not direct observations as streamflow cannot generally be continuously measured in natural rivers. Instead, the water level (also called “stage”) is continuously monitored. Streamflow time series is hence derived from rating curves (Rantz, 1982; WMO, 2010), hereafter called “RCs”, which are models transforming an input stage into an output discharge. These models are estimated using occasional stage-discharge measurements (also known as gaugings) and hydraulic constraints. The physical relation between discharge and stage is determined by hydraulic controls, that is, physical characteristics of the river section or channel influencing the flow: geometry, friction, longitudinal slope, head losses, etc. A hydraulic analysis of the site, through field expertise or modeling, allows identifying the succession or addition of hydraulic controls as flow increases (Le Coz et al., 2014).

RCs are affected by many sources of uncertainty, including structural uncertainty (imperfection of the RC model), gaugings measurement uncertainty, and parametric uncertainty (estimation of RC parameters).

Several methods for quantifying RC uncertainty have been developed as recently reviewed by Kiang et al. (2018). Unfortunately, the stage-discharge relation can be unstable and affected by rating changes. When ignored, these changes may be the main source of RC uncertainty (Ibbitt & Pearson, 1987). Concerning the causes of this instability, it is useful to distinguish between transient and sudden rating changes. Transient changes are caused by progressive phenomena (Hersch, 1998) such as hysteresis in unsteady flow during floods (rising limb and falling limb have different discharge values for the same stage), variable backwater due to unsteady downstream boundary conditions (e.g., stage controlled by a reservoir, sea tidal effect), growth and decay of aquatic vegetation, the evolution of ice sheets covering cold-water rivers, river bed evolution due to sediment dynamics with progressive scour and fill. On the other hand, sudden changes (shifts) are related to specific and occasional events such as morphogenic floods, dike breaks, dams built by swimmers or beavers, etc.

The next sections review the methods proposed in the literature to estimate the magnitude of rating changes and therefore to track the variability of RCs in time. The aforementioned transient and sudden changes require different approaches.

1.2. Detecting and Modeling Transient Changes

Modeling transient changes requires dynamic approaches. Such methods have actually existed in the operational practice since the beginning of the 20th century at least, in the form of time-varying RCs accounting for gradual rating changes. Schmidt (2002) described Stout's method (circa 1900) based on gradual daily shifts estimated from the gaugings. The same author also described the similar Bolster's method (circa 1910) which interpolates gaugings every day and develops parallel rating curves. Recently, Westerberg et al. (2011) and Guerrero et al. (2012) proposed to estimate RCs on moving temporal windows containing 30 successive gaugings. Morlot et al. (2014) proposed to compute as many RCs as there are gaugings and introduced the concept of RC aging: following an RC update, uncertainty increases with time according to a variographic analysis (Jalbert et al., 2011). Reitan and Petersen-Øverleir (2011) developed a dynamic model based on time-varying RC parameters within a hierarchical Bayesian framework. Finally, in the specific context of sites affected by aquatic vegetation, Puechberty et al. (2017) proposed time-varying stage corrections and Perret et al. (2021) introduced the Bayesian estimation of a time-dependent rating curve model accounting for vegetation growth and decay, with year-specific parameters.

1.3. Detecting Sudden Changes

As rating changes often result from episodic morphogenic floods, models assuming sudden rating changes between stability periods are more widespread than dynamic models in the operational practice. This approach requires solving two main issues: detecting changes (which includes estimating shift times), and estimating the successive stable RCs with their associated uncertainties. In this paper, we will focus on the first issue only. We refer the reader to Mansanarez et al. (2019) for a review on the second issue.

Several methods exist for sudden rating change detection. The most common approach is arguably to use gaugings. For instance, an empirical rule (WMO, 2010) prescribes to start a new period (and hence a new RC) when a gauged discharge departs from the current RC by more than some predefined threshold, or when successive gauged discharges are systematically above or below the current RC. This rule varies across agencies and site conditions, for instance: $\pm 5\%$ of discharge or ± 0.6 cm (± 0.02 ft) in stage for low-flow controls for the USGS (Rantz, 1982); $\pm 5\%$ of discharge for the Water Survey of Canada (Rainville et al., 2016); $\pm 10\%$ of discharge and/or $\pm 2 - 3$ cm in the stage in France (Puechberty et al., 2017). While easy to apply, this approach is based on empirical rules and ignores both gaugings and RC uncertainties.

In addition, operational services monitor the evolution of the river bed elevation to detect a change in the corresponding RC parameter. They use field observations such as information about river works, gravel mining operations, and bathymetry surveys. It is also possible to install submersible pressure transducers at various locations along the reach (Phillips & Eaton, 2009). An observed drastic drop or raise in the stage record may indicate a sudden river bed change. Moreover, Łapuszek and Lenar-Matyas (2015) evaluated whether changes in the annual minimum stages may indicate changes in the river bed level. This method is useful to provide a trend of the river bed evolution, but due to its annual resolution, it cannot precisely

identify the dates of rating shifts. Alternatively, McMillan et al. (2010) proposed to arbitrarily select the 0.5-years return period discharge as a threshold triggering a new RC period.

Furthermore, some operational services perform correlation analyses with reference stable discharge time series (e.g., the output of a hydrological model, or a discharge time series from a stable and well gauged neighboring station). Changes in the correlation structure may be indicative of rating shifts.

Finally, a formal way to detect changes by using gaugings is to apply a segmentation procedure to the time series of residuals between the gaugings and a time-invariant RC in order to identify homogeneous families of gaugings, as done for instance by Morlot et al. (2014). This paper focuses on this approach.

1.4. Change Point Detection Methods

The problem of finding abrupt changes in a time series is known in the literature as Change Point Detection (CPD). Reviews of the most representative methods have been proposed by many authors, such as Aminikhanghahi and Cook (2017), Basseville and Nikiforov (1993), Ducré-Robitaille et al. (2003), Jandhyala et al. (2013), and Keogh et al. (2003).

A distinction can be made between single (sCPD) and multiple (mCPD) change point detection methods. sCPD methods are based on statistical tests questioning the existence of one single change affecting typically the mean or the median of the series, sometimes higher moments as well. The literature proposes several methods, e.g., likelihood-ratio tests (Chen et al., 2006; Chen & Gupta, 2012; Hinkley, 1970), nonparametric tests (Kruskal & Wallis, 1952; McGilchrist & Woodyer, 1975; Pettitt, 1979; Wilcoxon, 1945), “Standard Normal Homogeneity” tests (Alexandersson, 1986; Hawkins, 1977), and Bayesian procedures (Booth & Smith, 1982; Chernoff & Zacks, 1964; Lee & Heghinian, 1977; Perreault et al., 1999, 2000a, 2000b).

Alternatively, mCPD methods look for multiple change points in the series. Unlike sCPD, mCPD is affected by a combinatorial challenge induced by the large number of possible change point positions. The Binary Segmentation or BS (Scott & Knott, 1974) recursively performs sCPD until no more changes are detected in any of the obtained segments. However, this approach is prone to known issues such as premature termination (schematized in Figure 1a) and mislocated splits (Figure 1b). To overcome these issues, several options are available: for example, the Circular BS method proposed by Olshen et al. (2004), sequential methods (Hubert & Carbonnel, 1987; Hubert et al., 1989; Page, 1954), Dynamic Programming (Auger & Lawrence, 1989; Killick et al., 2012), Bayesian inference (Chib, 1998; Green, 1995; Lavielle & Lebarbier, 2001) or Hidden Markov Models (Cappé et al., 2005; Luong et al., 2012).

However, these methods lack flexibility in the treatment of uncertainties affecting the segmented data. Indeed, the total uncertainty affecting RC residuals is induced by both gaugings and RC uncertainties. This uncertainty is not only potentially large, but it may also strongly vary from point to point. This should be accounted for in the segmentation procedure, as illustrated in the conceptual example of Figure 1c: ignoring data uncertainty clearly suggests three periods, while recognizing that two points have a much larger uncertainty than the others suggests a single period may suffice. An option is to consider rescaled residuals instead of absolute residuals (e.g., dividing the absolute residuals by the standard deviation representing their uncertainty). However, in some cases, this rescaling might create spurious periods as illustrated in Figure 1d.

Furthermore, mCPD methods attempt to detect all changes in a given data set. In the context of RC shifts, this data set is derived using a baseline RC fitted to all gaugings, which may be a very poor representation of the stage-discharge relation. The large scatter and uncertainty of residuals may hide smaller changes that may be missed by such a “single-pass” procedure. A recursive procedure, re-estimating the RC on each sub-period and deriving updated data sets of RC residuals, may hence be of interest.

Finally, the estimated change points provided by mCPD methods are not well-adapted to the context of RC shifts for the following reasons: (i) they are expressed in terms of position (i.e., observation index) rather than time, which is not ideal for irregular data such as gaugings; (ii) the uncertainty around the change point positions is rarely quantified. Nam et al. (2012) underlined the importance of accounting for the uncertainty of change point estimates. Estimating change points in terms of uncertain shift times would be

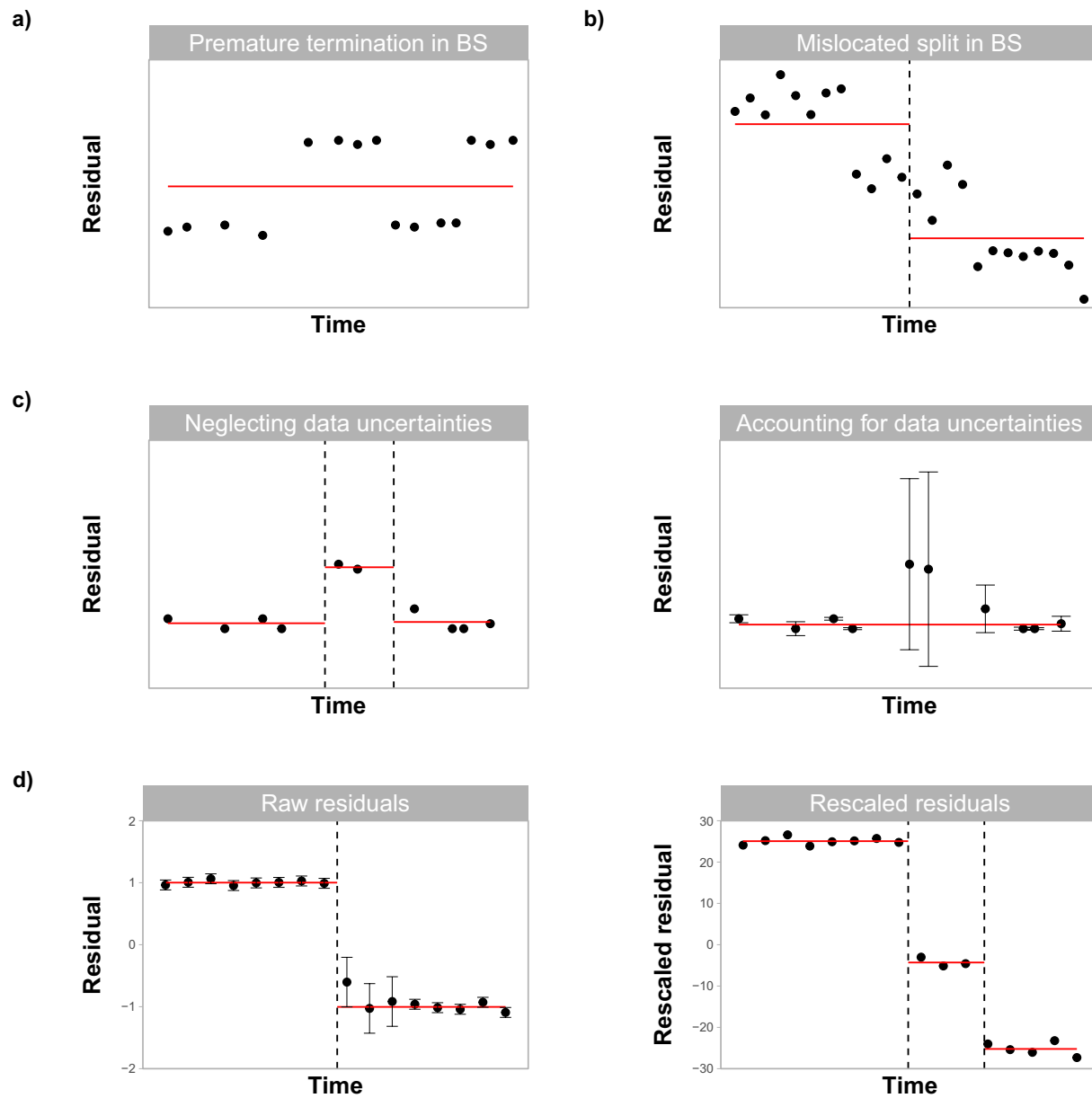


Figure 1. Schematic illustration of typical segmentation pitfalls: (a) premature termination in Binary Segmentation: no single change point model provides a reasonable fit to the data and BS, therefore, stops at the first iteration, with no change detected; (b) mislocated split in Binary Segmentation: optimal single change point is in the middle of a stable period; (c) neglecting versus accounting for data uncertainties; (d) artifact induced by residual rescaling: while raw residuals with standard deviation close to zero lead to rescaled residuals with high absolute values, highly uncertain residuals are sent to zero by the rescaling, thus creating a spurious period.

useful to look for specific events that may have caused the change—e.g., a large flood that would typically occur *in between* gaugings.

1.5. Objectives of the Paper

The main objective of this paper is to propose an original method for the detection of rating shifts through the segmentation of residuals between the gaugings and a baseline RC. The method must:

1. account for uncertainties in segmented data
2. recursively re-estimate the baseline RC and apply the segmentation on each sub-period to reveal minor changes
3. express change points in terms of time (rather than position), along with their uncertainty

The structure of the paper can be summarized as follows. Section 2 describes the proposed method. In Section 3, this method is applied to a typical hydrometric station with the aim of illustrating the main properties of the method and introducing several possible variants. Section 4 then describes a more thorough evaluation of the method's performance based on synthetic data sets where change points are known. Section 5 discusses results and proposes future perspectives. Finally, Section 6 summarizes the main findings.

2. The Proposed Method for Rating Shift Detection

2.1. Overview

Figure 2 illustrates in a schematic way the algorithm of the proposed method. The main steps are listed below and detailed in the next subsections.

1. Estimation of the baseline RC and its uncertainty using all available gaugings (Section 2.2)
2. Computation of the residuals between the gaugings and the baseline RC, and their uncertainties (Section 2.3)
3. Multiple change-point detection applied to the residuals time series (Section 2.4) and choice of the optimal number of change points (Section 2.5)
4. Shift times adjustment (Section 2.6)
5. "Top-down" recursion: re-apply steps 1–4 to each period until no more changes are detected (Section 2.7)

2.2. Estimation of the Baseline Rating Curve

The first step of the proposed method is to estimate the baseline RC and its uncertainty using all gaugings. Since one of the basic objectives of this study is to account for both gaugings and RC uncertainties, it is necessary to select an RC estimation method that provides quantitative uncertainties (see Kiang et al. (2018) for a review of such methods). In this study, the BaRatin method (Le Coz et al., 2014) is used for convenience. We refer to the aforementioned paper for a more detailed description and we stress that any other method could be used, as long as it provides the uncertainty around the RC.

2.3. Computation of Residuals and Their Uncertainty

The residuals between the gaugings and the RC are defined as follows:

$$r_i = \tilde{Q}_i - \hat{Q}_i \quad i = 1, \dots, N \quad (1)$$

where \tilde{Q}_i is the gauged discharge, \hat{Q}_i is the RC-estimated discharge and N is the number of gaugings. Each residual is affected by two sources of uncertainty. The first one is the measurement uncertainty affecting the observed discharge \tilde{Q}_i , which does not depend on the RC method and should ideally be determined by an uncertainty analysis of the measurement process. Assuming zero-mean Gaussian measurement errors, this uncertainty can be quantified by the standard deviation $u_{\tilde{Q}_i}$. By contrast, the uncertainty affecting the RC discharge \hat{Q}_i is obviously dependent on the RC method being used. The BaRatin method used in this paper assumes zero-mean Gaussian RC errors with standard deviation $u_{\hat{Q}_i} = \gamma_1 + \gamma_2 \hat{Q}_i$, where γ_1 and γ_2 are estimated as part of the RC estimation process. We reiterate that any other RC method could be used as long as it provides an uncertainty $u_{\hat{Q}_i}$.

Further assuming that measurement and RC errors are independent, the combined standard uncertainty affecting residuals r_i is equal to:

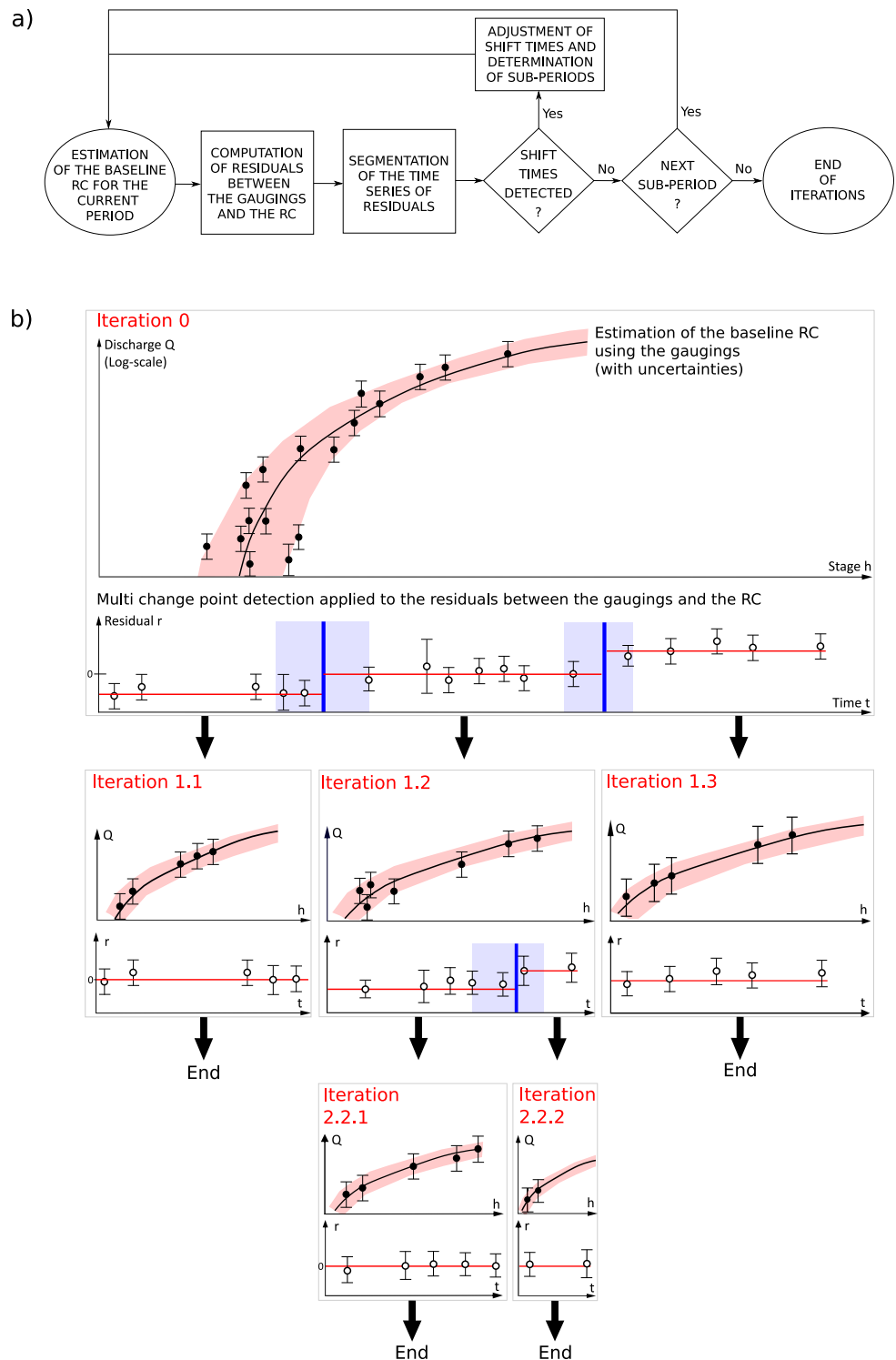


Figure 2. (a) Conceptual flowchart of the proposed algorithm. (b) Schematic representation of the iterative procedure. Each iteration consists of the succession of Steps 1–4 described in Section 2. Colored ribbons and error bars represent 95% uncertainty intervals for RCs (pink), shift times (blue), gaugings (black dots), and residuals (empty dots).

$$u_{r_i} = \sqrt{u_{Q_i}^2 + u_{Q_i}^2} \quad i = 1, \dots, N \quad (2)$$

2.4. Segmentation Model and Bayesian Inference

The third step of the proposed method is the mCPD of the time series of residuals $(t_i, r_i)_{i=1..N}$ through the Bayesian estimation of a segmentation model.

2.4.1. General Segmentation Model

Generally speaking, a segmentation model can be viewed as a piece-wise constant model of the form:

$$r_i = \hat{r}_i + \varepsilon_i \quad (3)$$

$$\hat{r}_i = \begin{cases} \mu_1, t_1 \leq t_i < \tau_1 \\ \mu_2, \tau_1 \leq t_i < \tau_2 \\ \vdots \\ \mu_K, \tau_{K-1} \leq t_i \leq t_N \end{cases} \quad (4)$$

In Equation 4, K is the known number of segments; it will be selected based on a model-selection procedure described in the following Section 2.5. The means μ_j of each segment j and the change point τ_j that separate segment j from segment $j + 1$ are unknown and are grouped into the vector of inferred parameters $\theta = (\mu_1, \dots, \mu_K, \tau_1, \dots, \tau_{K-1})$. The treatment of segmentation errors ε_i depends on how these errors are interpreted: it is a key focus of this study, and two distinct approaches will be presented in the next section. Finally, it is noted that many segmentation models in the literature use the observation index i rather than the time t_i . We favor the latter option because it will allow expressing uncertainties on the change point in terms of time rather than position, which is particularly useful in the RC context where gaugings are performed irregularly.

2.4.2. Two Approaches for Describing Segmentation Errors

Segmentation errors ε_i are generally assumed to be realizations from a zero-mean Gaussian distribution. The two approaches considered here differ in the way they treat their standard deviation σ_i :

1. Type-1 approach: σ_i is assumed to be unknown but identical for all segmented data, i.e., $\sigma_i = \sigma$
2. Type-2 approach: σ_i is assumed to be known but to vary between segmented data, i.e., $\sigma_i = u_{t_i}$

Type-2 approach is particularly suitable for cases where the segmented data r_i are RC residuals (or more generally, residuals between a model and observations): indeed, Equation 2 provides the known standard deviation u_{t_i} to be used in this case. Type-1 approach is arguably the most standard procedure, since it corresponds to the assumption made in standard regression models with homoscedastic errors. However, it ignores the uncertainty affecting the segmented data, despite the fact that it is known before applying the segmentation procedure.

2.4.3. Bayesian estimation

Assuming that segmentation errors ε_i are independent, the likelihood associated with the segmentation model can be written as follows:

$$\begin{aligned} \text{Type - 1 approach: } p(\mathbf{r} | \theta, \sigma) &= \prod_{i=1}^N \phi(r_i | \hat{r}_i(\theta), \sigma) \\ \text{Type - 2 approach: } p(\mathbf{r} | \theta) &= \prod_{i=1}^N \phi(r_i | \hat{r}_i(\theta), u_{t_i}) \end{aligned} \quad (5)$$

where $\phi(z | m, s)$ is the probability density function (pdf) of a Gaussian distribution with mean m and standard deviation s , evaluated at value z .

Bayesian inference requires specifying the prior distribution of parameters (θ, σ) . Independent priors are specified for each inferred parameter. By default a uniform prior distribution is specified for each change point, $\tau_j \sim \mathcal{U}(t_1, t_N)$. Note that on top of this prior distribution, change points are also constrained by the relation $\tau_1 < \dots < \tau_{K-1}$. An order-of-magnitude Gaussian distribution is specified for each segment mean,

$\mu_j \sim \mathcal{N}(0, 10^m)$; the value of m is case-specific and should reflect the expected order of magnitude of RC residuals, which in turn is specific to the studied catchment.

Bayes' theorem allows combining the information brought by the data through the likelihood with the prior information on the inferred parameters into a posterior distribution of the parameters, whose pdf is defined by:

$$\begin{aligned} \text{Type - 1 approach: } p(\boldsymbol{\theta}, \sigma | \mathbf{r}) &\propto p(\mathbf{r} | \boldsymbol{\theta}, \sigma) p(\boldsymbol{\theta}, \sigma) \\ \text{Type - 2 approach: } p(\boldsymbol{\theta} | \mathbf{r}) &\propto p(\mathbf{r} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) \end{aligned} \quad (6)$$

An MCMC approach based on a multi-block Metropolis algorithm is used to explore this multidimensional posterior distribution. The variance of each parameter jump distribution is also adapted during iterations in order to reach a user-defined acceptance rate. The implemented algorithm is detailed in Renard et al. (2006). By default, the first half of the simulations is ignored ("burned") and only the second half of simulated values is used for inference. MCMC convergence is assessed by visually inspecting trace plots and density plots (except in the synthetic case studies of Section 4 where this is not feasible given a large number of replications). In addition, the Potential Scale Reduction Factor (Brooks & Gelman, 1998) is computed for each parameter, verifying that it is smaller than 1.2.

Finally, to avoid short segments containing no observations, which may lead to an ill-posed inference, additional constraints can be enforced: a minimum number of points $N_{\min} \geq 1$ for each segment and a minimum duration $d_{\min} \geq 0$ between two consecutive change points. These constraints are case-specific and user-defined.

2.5. Choice of the Optimal Number of Segments

The number of segments K is selected by minimizing the Deviance Information Criterion (DIC, Spiegelhalter et al., 2002), in the formulation suggested by Pooley and Marion (2018). The DIC is selected because it is adapted to the Bayesian estimation described in Section 2.4.3. We note that a maximum-likelihood estimation could also be used and possibly favored by some users. In this case, the DIC should be replaced by an alternative model-selection criterion such as the Akaike Information Criterion (AIC, Akaike, 1974), the Bayesian Information Criterion (BIC, Schwarz, 1978), or the Hannan-Quinn information Criterion (HQC, Hannan & Quinn, 1979).

Let $\boldsymbol{\Theta}$ denote the vector of all inferred parameters, i.e., $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \sigma)$ for the type-1 approach ($N_{\text{par}} = 2K$) and $\boldsymbol{\Theta} = \boldsymbol{\theta}$ for the type-2 approach ($N_{\text{par}} = 2K - 1$). Moreover, let $D(\boldsymbol{\Theta})$ denote the deviance defined as $D(\boldsymbol{\Theta}) = -2 \ln(p(\mathbf{r} | \boldsymbol{\Theta}))$. The four criteria discussed above are computed as follows:

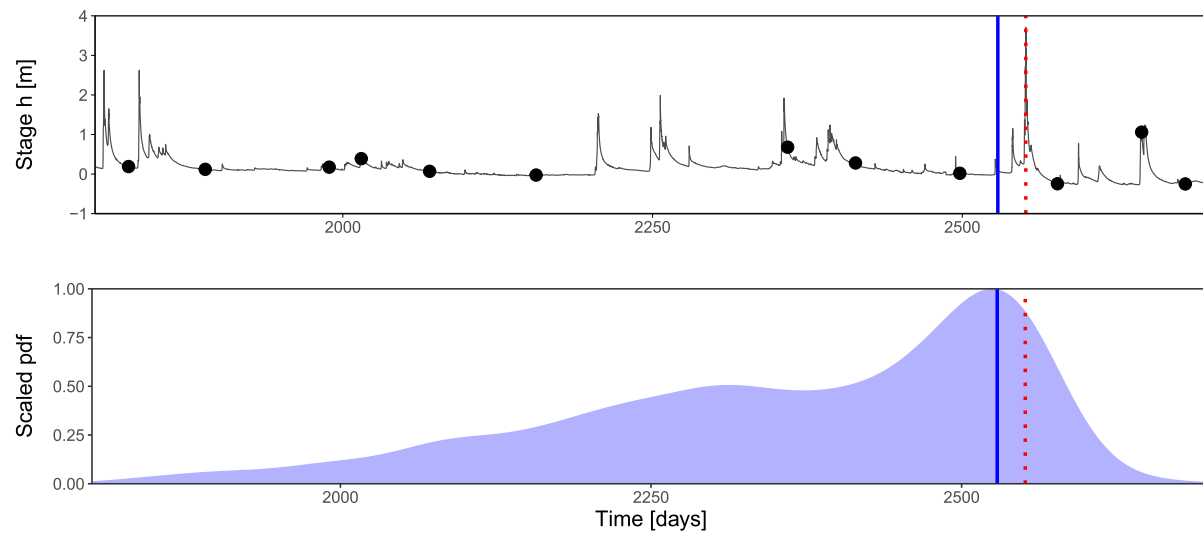
$$\begin{aligned} AIC &= D(\hat{\boldsymbol{\Theta}}) + 2N_{\text{par}} \\ HQC &= D(\hat{\boldsymbol{\Theta}}) + 2N_{\text{par}} \ln(\ln(N)) \\ BIC &= D(\hat{\boldsymbol{\Theta}}) + N_{\text{par}} \ln(N) \\ DIC &= E[D(\boldsymbol{\Theta})] + \frac{1}{2} \text{Var}[D(\boldsymbol{\Theta})] \end{aligned} \quad (7)$$

In the first three criteria, $\hat{\boldsymbol{\Theta}}$ is the maximum-likelihood parameter estimate. In the fourth criterion DIC, $E[\cdot]$ and $\text{Var}[\cdot]$ represent the posterior mean and variance, and the corresponding quantities can easily be computed using MCMC samples.

2.6. Adjustment of Shift Times

Bayesian estimation results provide the marginal posterior distribution of each inferred parameter. Rating shift times $\mathbf{s} = (s_1, \dots, s_{K-1})$ can be obtained from the posterior distributions of parameters $\boldsymbol{\tau}$.

As illustrated in Figure 3, each posterior distribution provides a point estimate $\hat{\tau}_j$ and a credibility interval CI_j . Typically, $\hat{\tau}_j$ is the Maximum A Posteriori (MAP) estimate maximizing the posterior density. The



Legend

— Stage record ● Gaugings | $\hat{\tau}_1$: $t_{\text{flood},1}$ ■ Posterior pdf of change point time τ_1

Figure 3. Example of shift time adjustment options. Instead of setting the shift time to the MAP estimate $\hat{\tau}_1$, a better option may be to choose the time of the maximum stage $t_{\text{flood},1}$ within CI_1 .

interval CI_j can be explored to find a physically justified shift time. The following three options can be considered for instance:

1. $s_j = \hat{\tau}_j$ can be used as a default option
2. If the stage record is available, s_j can be set to the time of the largest stage value within CI_j (cf. Figure 3), reasoning that a large flood is a likely cause of the shift
3. s_j can also be set manually within CI_j using any other information on possible causes of the shift (e.g., gravel mining operations, beavers/swimmers dams, works in the river bed, earthquake)

2.7. Recursive Segmentation

Once the first set of shift times has been identified, a recursive “top-down” procedure is performed (Figure 2). The segmentation procedure described in the previous sections (2.2–2.6) is recursively performed within each sub-period. At each iteration, only the gaugings of the current sub-period are considered. The iterations stop when within all periods no more shift times are detected.

It may happen that at some iteration of the recursive segmentation only few gaugings are available. This may lead to challenging RC estimation and, if the BaRatin method is used, to challenging quantification of $u_{\hat{Q}_i}$. When moving from one period to its sub-period, a better fit and hence a smaller standard deviation $u_{\hat{Q}_i}$ is expected; thus a prior uniform distribution between zero and the posterior mean of the parent period is specified for $u_{\hat{Q}_i}$.

3. Application to a Real Case Study: The Ardèche River at Meyras, France

3.1. Presentation of the Station

The Ardèche River at Meyras station is located in Mediterranean France, with a catchment area of 98 km². This station is characterized by a gravel bed degrading during each important flood. It has been already studied by Le Coz et al. (2014), Mansanarez et al. (2019), and Sikorska and Renard (2017). They all proposed a three-control hydraulic configuration: a section control governed by a natural riffle for low flows, the main channel control for medium flows and one floodway channel control added to the main channel for very high flows. The stage record is available for the period between November 07, 2001 and October 29, 2018.

Gaugings and comments about shift times have been provided by the hydrometric Service in charge of the station (UHPC Grand Delta).

3.2. Segmentation Strategies

Several segmentation strategies are used in order to compare “single-pass” versus recursive procedures and type-1 versus type-2 treatment of segmentation errors:

- A) “Single-pass” mCPD method from the literature. The R function *cpt.mean* of the *changept* package (Killick & Eckley, 2014) is used with the following options: (i) a maximum number of segments $K = 30$; (ii) change in the mean only; (iii) Binary Segmentation method (Scott & Knott, 1974); (iv) Normal statistical test; (v) BIC selection of K ; (vi) a minimum number of data in a segment $N_{min} = 1$. Note that in this method, the segmentation model of Equation 4 is expressed in terms of observation index i rather than time t_i , and does not provide uncertainty on the change point. Consequently, each change point τ_k is assumed to be uniformly distributed between times t_{i_k} (the time associated with the k -th detected change point) and $t_{i_{k-1}}$. The shift times are then adjusted (Section 2.6) on the largest stage value within this interval
- B) “Single-pass” mCPD method proposed in this paper with a type-1 treatment of segmentation errors (i.e., unknown but constant uncertainty). The following options are chosen: (i) maximum number of segments $K = 30$; (ii) DIC selection of K ; (iii) a minimum number of data in a segment $N_{min} = 1$; (iv) minimum duration of a segment $d_{min} = 0$ days. This approach is very similar to the previous Strategy A, except that the segmentation model of Equation 4 is expressed in terms of time t_i . Shift time adjustment is therefore applied by looking for the largest flood in the 95% credibility interval of the change point, as described in Section 2.6
- C) Recursive mCPD method proposed in this paper with a type-1 treatment of segmentation errors (i.e., unknown but constant uncertainty). The maximum number of segments is now set to 5 because by using a recursive procedure there is no requirement to find all changes during the first pass. All other options are identical to approach B
- D) Recursive mCPD method proposed in this paper with a type-2 treatment of segmentation errors (i.e., known residuals uncertainties). All options are identical to approach C

3.3. Results with Strategy D

Figure 4 shows the results of some significant steps of Strategy D. The structure and the enumeration of the iterations are schematized in (a). Seven shift times are detected in 13 iterations. Panels (b–d) zoom to the intermediate results of iterations 0, 1.2, 1.3.

For each iteration Figure 4b shows the RC estimated using gaugings from the current period. At iteration 0 the baseline RC has a large uncertainty, confirming the presence of multiple stage-discharge relations for this data set. This uncertainty decreases in subsequent iterations, reflecting the fact that the RC is estimated using more homogeneous gaugings.

Figure 4c shows the evolution of four criteria (AIC, BIC, HQC, and DIC) for the choice of the optimal segmentation (see Section 2.5). Similar behavior is observed between BIC, DIC, and HQC, in particular between BIC and DIC. On the contrary, AIC tends to favor a higher number of periods for the iterations shown in the figure and for other iterations (not shown).

Figure 4d shows the segmented residuals. These three iterations illustrate the added value of the “top-down” recursion: the large RC uncertainty at iteration 0 leads to the detection of two major shifts only; then iterations 1.2 and 1.3 lead to the detection of other minor shifts based on refined RCs. Iteration 1.3 also illustrates that, because of the uncertainty in the estimated change point location, the adjusted shift time (based on flood occurrence) may be far from the optimal time identified using gaugings only.

Finally, by the last iterations, segmentation errors (ϵ_i in Equation 3) do not generally exhibit any significant autocorrelation (not shown).

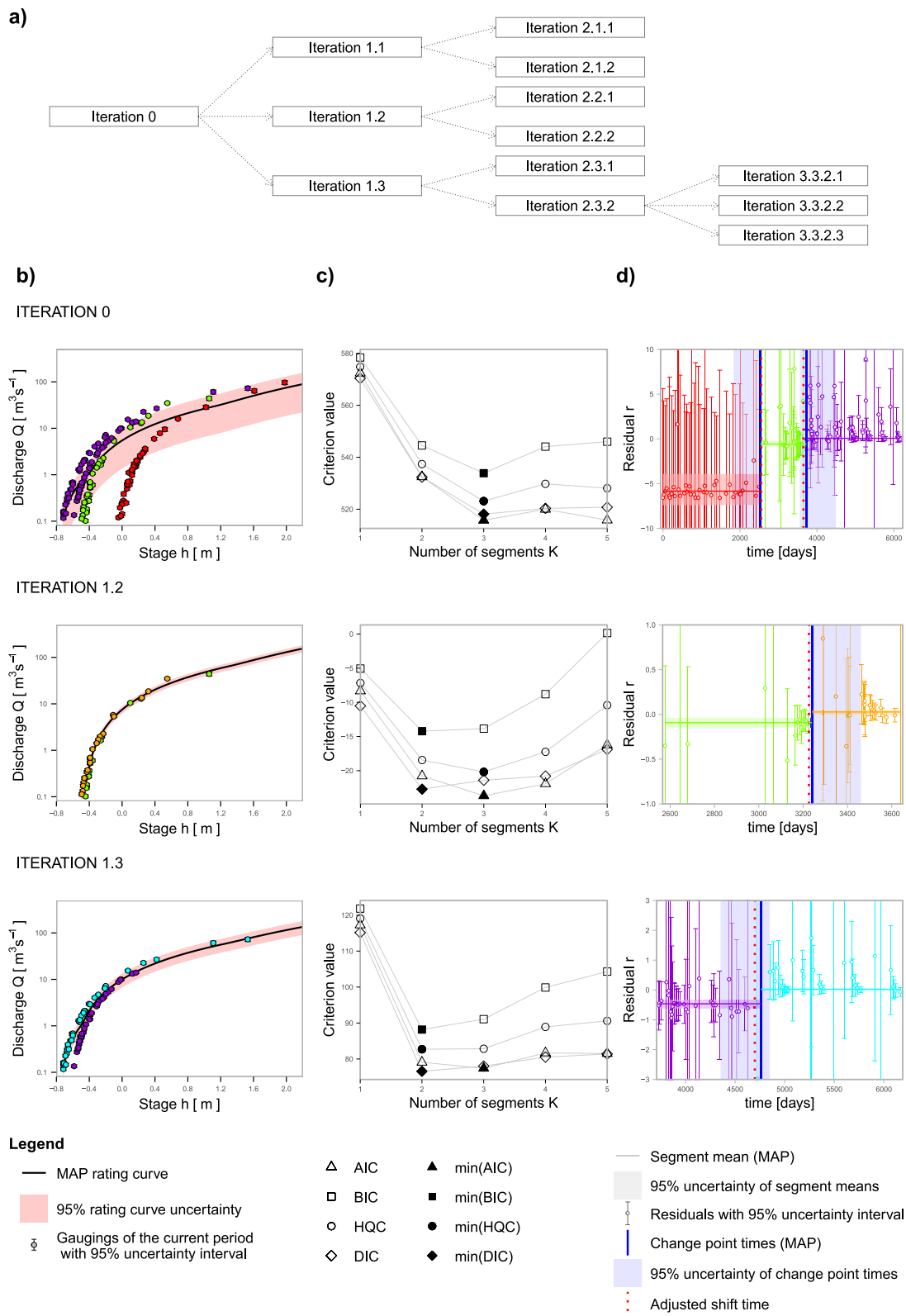


Figure 4. Strategy D applied to the gaugings of the Ardèche River at Meyras. Subfigure (a) structure of the recursion. Subfigures (b) baseline RC using the gaugings of the current period. Subfigures (c) evolution of four criteria for the selection of the optimal segmentation applied to the residuals. Subfigures (d) results of the segmentation considering the lowest DIC.

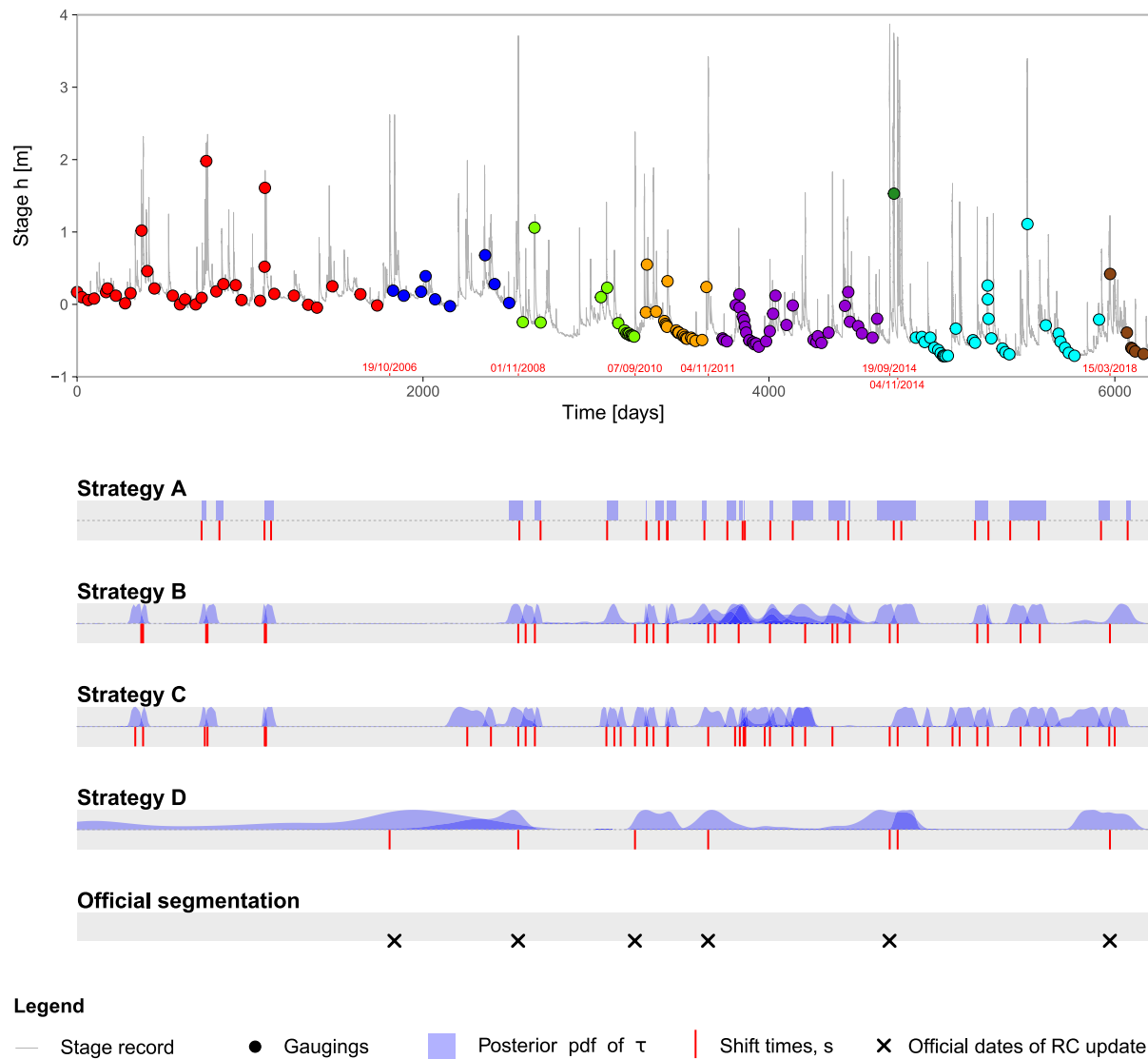


Figure 5. Results of the segmentation procedure applied to gaugings of the Ardèche River at Meyras, France. Gaugings are plotted against the stage record with a different color for each period of RC stability identified by Strategy D. For each strategy, results are presented as posterior pdf of τ_j (blue ribbons) and adjusted shift times s_j (red segments).

3.4. Comparison of Strategies A-D

The results of the four strategies are then summarized in Figure 5 in terms of detected shift times against the stage record. The official dates of RC updates are provided by the hydrometric Service UHPC Grand Delta and are confirmed by the analysis proposed by Mansanarez et al. (2019). Only Strategy D shows results similar to the official segmentation. However, the official dates cannot be considered as a “truth” against which the performance of competing methods can be judged. Comparing the results of Strategies A-B-C-D may still be insightful.

The first striking observation is that the four Strategies lead to markedly different segmentations, in particular between Strategy D and Strategies A, B, C.

While both Strategies C and D are recursive, they lead to very different numbers of shifts (42 detected in 56 iterations vs. 7 detected in 13 iterations). This indicates that the treatment of segmentation errors (type-1 vs. type-2) is of prime importance.

Strategies A and B are both nonrecursive, and differ in the following two aspects: Strategy A is index-based while B is time-based; they use a different criterion for selecting the number of shifts. Both strategies lead to very similar results in terms of a number of shifts (27 vs. 29) and their location. The slight differences may be due to the stronger penalty term of the BIC or to Binary Segmentation issues as illustrated in Figures 1a and 1b.

Strategies A, B, and C lead to many more change points than Strategy D. During the first period (red gaugings) the detected shifts correspond to the largest gaugings. This might suggest that the shifts result from the much larger uncertainty affecting these residuals (which cannot be accommodated with the constant-sigma type-1 approach), rather than from a genuine change in the underlying RC. This evidences the problem illustrated in Figures 1c and 1d.

However, the objective choice of the most efficient segmentation strategy is challenging without knowing the true shift times. In the next section, the model selection criteria and the segmentation strategies are compared based on synthetic data with known shift times and magnitudes, thus enabling a more objective evaluation of their performance.

4. Performance Evaluation from Simulated Rating Shifts

4.1. Generation of Synthetic Data

The generation of synthetic gauging data is based on the following steps (the corresponding R code is also provided as online material):

1. Select the length of the studied period $[0; T]$ (in years)
2. Shift times: generate inter-shift durations from an exponential distribution with rate λ_s (e.g., $\lambda_s = 1/5$ corresponds to 1 shift every 5 years on average); shift times are then derived as the cumulative sum of the inter-shift durations. The generation stops when the shift time exceeds T , leading to N_s shifts
3. Shift magnitudes: it is assumed that RC shifts only affect the offset of the lower control (i.e., the b in equation $Q = a(h - b)^c$). Each shift magnitude $\delta b^{(i)}$ is generated from a Gaussian distribution with mean 0 and standard deviation σ_b . If there exists at least one shift, for each stable period j ($j \geq 2$), the offset parameter $b^{(j)}$ is hence equal to $b^{(j)} = b^{(1)} + \sum_{i=2}^j \delta b^{(i)}$
4. Gauging times: use the same approach as for shift times, using a rate λ_g leading to N_g gaugings
5. Gauging true discharge: for each gauging, the true discharge Q_i is generated by first sampling a non-exceedance probability p between 0 and 1, then transforming it into discharge using the quantile function of a LogNormal $\mathcal{LN}(\ln(50), 0.5)$ distribution. Probability p is sampled from a beta distribution $\mathcal{B}(0.1, 0.9)$ which is strongly skewed toward zero, mimicking the typical situation where gaugings are mostly performed during low flows
6. Gauged stage: for each gauging, the stage h_i is computed by applying the inverse RC function to the true discharge Q_i
7. Gauged discharge: for each gauging, the gauged discharge is obtained by adding a measurement error ξ_i to the true discharge Q_i . Error ξ_i is sampled from a Gaussian distribution with mean 0 and standard deviation $\rho_i \times Q_i$

4.2. Design of Experiments

In order to assess how the properties of the data set impact the performances of the segmentation approaches, several classes of simulation are defined as described in Table 1. Each class is characterized by fixed values of the parameters described in Section 4.1. Comparing these classes allows assessing the impact of the following Properties:

- P1: number of hydraulic controls N_c (classes 8 and 10)
- P2: mean number of gaugings per period, which is equal to the ratio between the gaugings frequency λ_g and the shifts frequency λ_s (classes 1, 6, 2, 7, 3, 8), as suggested by Ibbitt and Pearson (1987)
- P3: uncertainty in gauged discharges as controlled by ρ_{LF} and ρ_{HF} (classes 8 and 9)
- P4: shift magnitude is controlled by σ_b (classes 4, 5, and 8)

Table 1
Classes of Simulation for the Performance Evaluation

Class	Frequency of gaugings λ_g (year ⁻¹)	Frequency of shifts λ_s (year ⁻¹)	Mean number of gaugings/period λ_g/λ_s	Shift st.dev. σ_b (m)	Gauging error		Number of controls N_c
					Low flows $\rho_{LF}(\%)$	High flows $\rho_{HF}(\%)$	
1	2	1/5	10	0.5	2.5	5	1
2	4	1/5	20	0.5	2.5	5	1
3	7	1/5	35	0.5	2.5	5	1
4	10	1/5	50	0.1	2.5	5	1
5	10	1/5	50	0.3	2.5	5	1
6	10	1	10	0.5	2.5	5	1
7	10	1/2	20	0.5	2.5	5	1
8	10	1/5	50	0.5	2.5	5	1
9	10	1/5	50	0.5	10	15	1
10	10	1/5	50	0.5	2.5	5	3

For each class, 10 replications are generated, for a total of 100 simulations (some data sets are reported in Figure 6). To minimize computational cost all data sets are generated with a maximum number of 150 gaugings and a maximum number of 15 true shift times within a period of 15 years. As an order of magnitude, the CPU time to apply approach C to one data set with 54 gaugings is around 50 min. This is to be multiplied by the number of data sets (100) times the number of approaches (4) or the number of criteria (4), which amounts to several days of effective running time.

The first experiment aims at comparing criteria AIC, BIC, DIC, and HQC (see Section 2.5). To this aim, strategy D is applied to all classes above.

The second experiment aims at comparing Strategies A, B, C, and D. To this aim, all four Strategies are applied to the same 100 data sets of the first experiment. The stage record is not available for synthetic data sets, thus in Strategies B-C-D, the shift times are taken as the estimated parameters $\hat{\tau}_j$ (i.e., option 1 in Section 2.6). Since Strategy A provides the estimated change point as an observation index k (rather than a time), the shift time is taken as the middle of the interval $[t_{k-1}; t_k]$.

4.3. Metrics for Performance Evaluation

The performance evaluation uses some of the metrics proposed by Aminikhanghahi and Cook (2017). At the end of the segmentation procedure each gauging is classified into one of TP , FN , FP , TN (see example in Figure 7), where:

- A gauging is classified TP (true positive) if it is the nearest neighbor of a true shift and this true shift is within the 95% CI of an estimated shift
- A gauging is classified FN (false negative) if it is the nearest neighbor of a true shift but this true shift is outside all 95% CI of estimated shifts
- A gauging is classified FP (false positive) if it is the nearest neighbor of an estimated shift but the 95% CI of this estimated shift does not contain any true shift
- Otherwise a gauging is classified TN (true negative)
- $N_g = n_{TP} + n_{FN} + n_{FP} + n_{TN}$

The Accuracy A is defined as the rate of correctly classified gaugings:

$$A = \frac{n_{TP} + n_{TN}}{N_g} \quad (8)$$

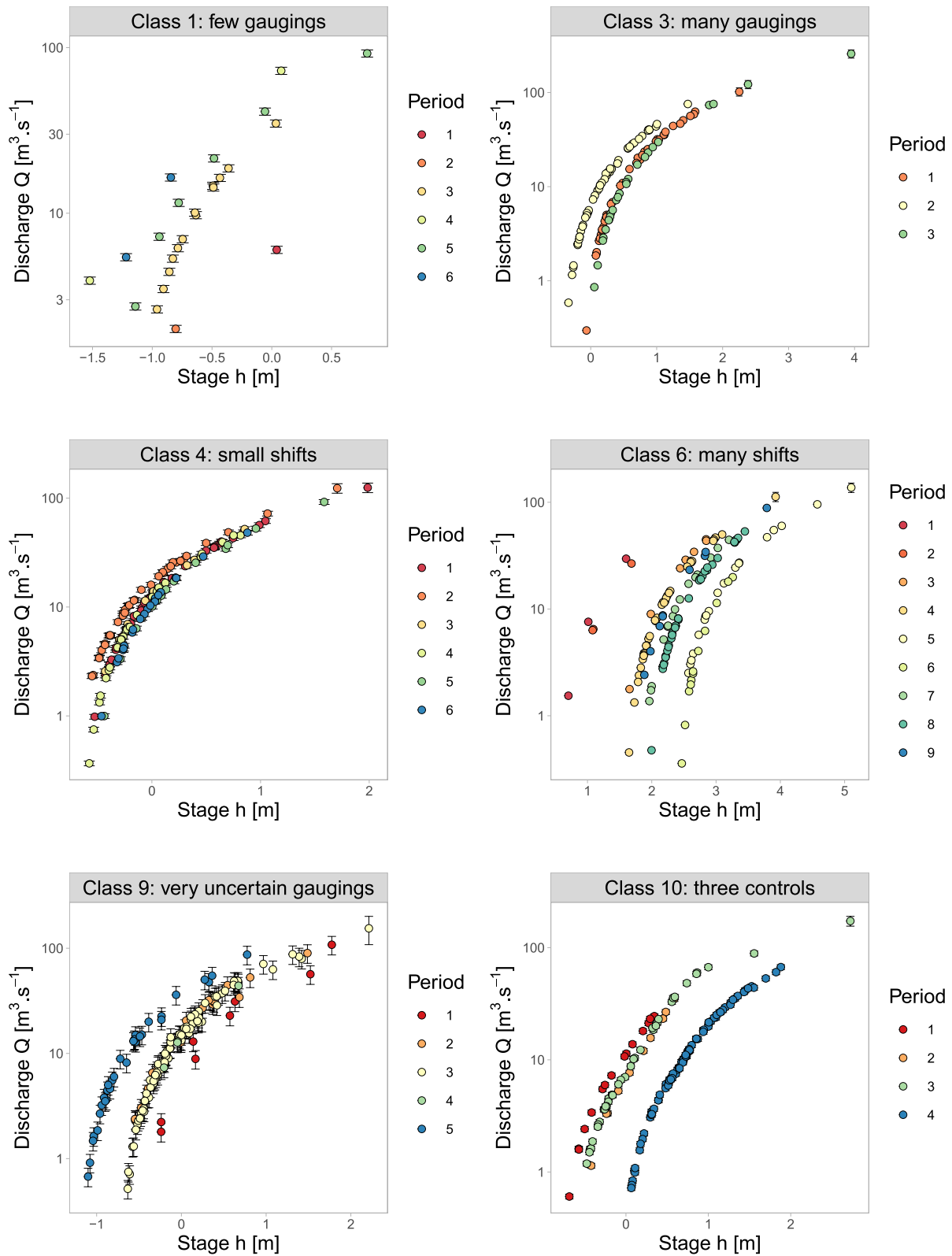


Figure 6. Examples of synthetic data sets characterized by known rating shift times using parameters of Classes 1-3-4-6-9-10 defined in Table 1.

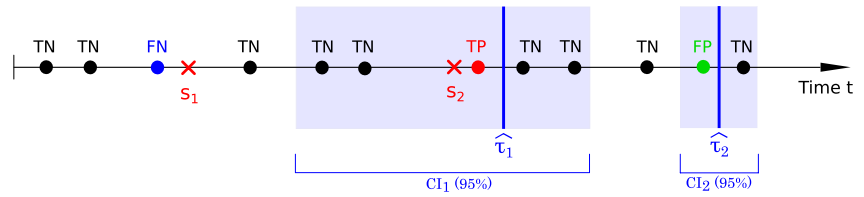


Figure 7. Schematic example of the gaugings classification into true positive (TP), false negative (FN), true negative (TN), false positive (FP) for the performance evaluation of the segmentation results. s_i represent the known shift times, while $\hat{\tau}_i$ are the change point estimates.

The Sensitivity S is maximal when no shift has been missed; low values hence correspond to under-segmentation:

$$S = \frac{n_{TP}}{n_{TP} + n_{FN}} \quad (9)$$

The Precision P is maximal when all detected shifts are real; low values hence correspond to over-segmentation:

$$P = \frac{n_{TP}}{n_{TP} + n_{FP}} \quad (10)$$

The $RMSE$ between the times of correctly detected shifts s_i^{TP} and the times of corresponding true shifts $s_{k_i}^{true}$ is also computed:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n_{TP}} (s_i^{TP} - s_{k_i}^{true})^2}{n_{TP}}} \quad (11)$$

4.4. Results of the Experiments

4.4.1. Comparison of Criteria for Choosing the Number of Change Points

Figure 8 summarizes the results of the first experiment. Results reveal that:

- BIC and DIC on the one hand and AIC and HQC on the other hand lead to similar performance metrics
- BIC has the highest Accuracy/Precision, closely followed by DIC, while AIC and HQC have the lowest. The same ranking holds for most simulation classes (not shown). AIC and HQC lead to a higher degree of over-segmentation with, on average, 125% and 67%, respectively, more detected shift times than there really are. On the contrary, BIC and DIC over-estimate the number of shifts by only 13% and 40%, respectively
- BIC and DIC sometimes miss a few shifts leading to lower values of sensitivity than AIC and HQC

These results indicate that AIC and HQC have a marked tendency to over-segmentation and should therefore be avoided. BIC and DIC have similar performances, and it is, therefore, sensible to select the one that is conceptually more adapted to the chosen inference paradigm. More specifically, BIC is solely based on the likelihood and is hence more adapted to maximum-likelihood estimation (despite what its name confusingly suggests). By contrast, DIC makes use of the whole posterior distribution and should therefore be favored in a Bayesian context such as the one adopted in this paper.

4.4.2. Comparison of Segmentation Strategies

Figure 9 summarizes the results of the second experiment. Strategies A, B, C, and D are compared considering all simulations.

Strategy D is quite markedly the best-performing approach in terms of Accuracy, Precision, and $RMSE$. On the contrary Strategy A is quite markedly the worst one, due to a strong tendency to over-detection. Its

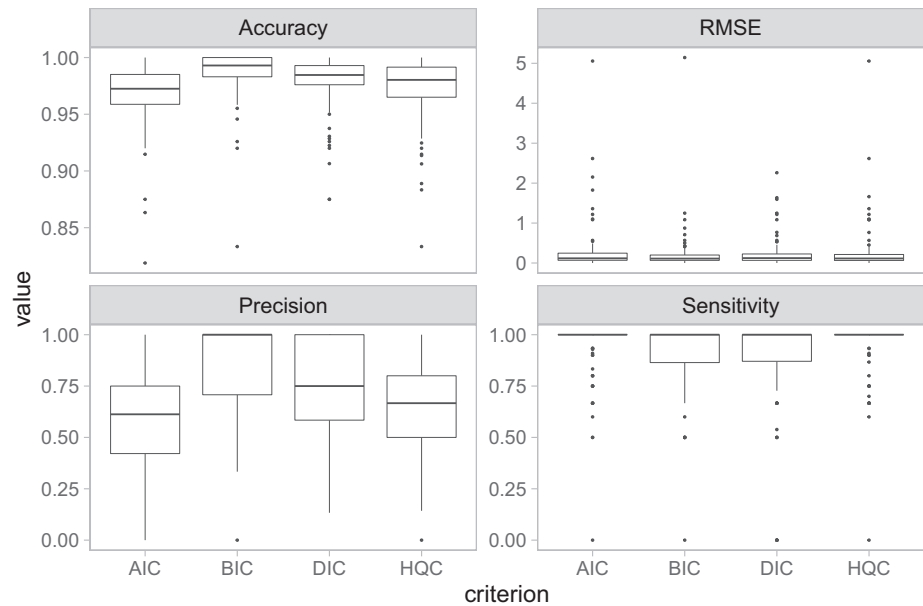


Figure 8. Results of the first experiment with Strategy D, comparing four criteria for the optimal choice of K and using all 100 simulated data sets.

good Sensitivity means that it does not miss many changes, but this comes at the cost of detecting too many spurious ones.

Strategies C and D are both recursive procedures and both yield better results in terms of Accuracy, Precision, and RMSE compared with the “single-pass” Strategies A and B, demonstrating the added value of the recursive approach. However, Strategy D yields better results than Strategy C in terms of all metrics, emphasizing the added value of type-2 errors treatment.

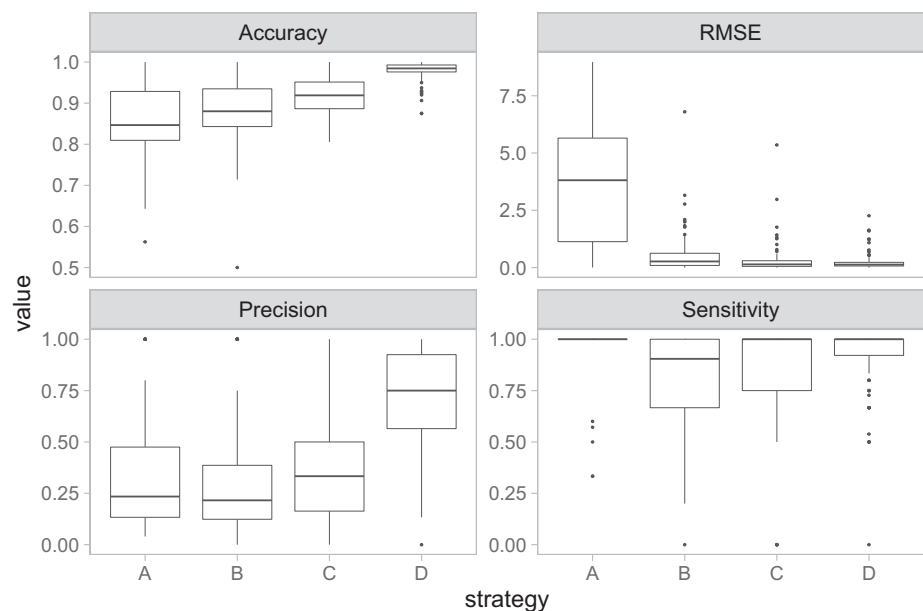


Figure 9. Results of the second experiment: comparison between the four strategies for the segmentation of residuals.

Finally, Strategies A and B are conceptually similar but, surprisingly, yield quite different results. Strategy B yields better results in terms of Accuracy and RMSE. In particular, the lower RMSE may be due to the added value of expressing change points in terms of time rather than position.

4.4.3. Factors Influencing the Method Performance

Figure 10 focuses on the results of Strategy D and evaluates its performance for data sets with varied properties as described in Section 4.2.

The main factor affecting the performance of the segmentation is the mean number of gaugings per period. As expected, performance increases with this number, especially in terms of Accuracy and RMSE. Results suggest that 20 gaugings per period are sufficient to achieve a good-quality segmentation. For higher values, the Accuracy seems to stabilize, while the RMSE continues decreasing, mainly because of the smaller inter-gauging interval. Performance markedly deteriorates with only 10 gaugings per period. This confirms that shifty curves require a high frequency of gaugings. Sensitivity sharply increases when moving from 10 to 20/35 gaugings per period. Then, for 50 gaugings per period, few particular realizations have Sensitivity = 0 (since characterized by only one single small change which unfortunately has been missed). Instead, Precision is weakly influenced by the number of gaugings per period. However, with an average of 10 gaugings per period, the segmentation detects some false shifts and Precision slightly decreases.

Other factors have a lower impact on performance. Different hydraulic configurations slightly influence the performance of the segmentation. More complex rating curves (with multiple controls) lead to a slight decrease in Accuracy and Precision and a slightly better Sensitivity.

Surprisingly, increasing the gaugings uncertainty does not markedly impact performance. It leads to slightly lower but still very high values of Accuracy. It also leads to slightly lower Precision, and, surprisingly, to higher Sensitivity. The RMSE values remain very similar.

Finally, increasing the shift magnitude does not markedly influence the performance. However, few simulations characterized by very small shifts show lower Accuracy.

5. Discussion

5.1. Contributions to the Operational Practice and the Scientific Literature

The proposed method represents a more formal way to detect rating shifts using gaugings, compared to empirical rules commonly used in the operational practice (e.g., WMO, 2010). A similar formalization objective was pursued by Morlot et al. (2014) who applied the Hubert segmentation method (Hubert et al., 1989) for segmenting the residuals between the gaugings and a baseline RC. Their method yielded satisfying results, even though it neglected the residuals' uncertainties. However, the synthetic simulations and the real case study proposed in this paper suggest that neglecting residuals uncertainties may lead to over-segmentation.

The proposed method differs from the mCPD literature in its handling of shift times in two aspects: considering time rather than position and providing shift time uncertainty. Existing performance evaluation metrics proposed in the literature (Aminikhanghahi & Cook, 2017) were also adapted to account for shift times uncertainty while comparing the true shift times and the detected ones.

In addition to the segmentation procedure, this paper proposes a protocol for the generation of synthetic data sets of gaugings and shift times. This protocol is very useful to evaluate the performance of the segmentation method.

The case studies proposed in this paper indicate that the mean number of gaugings per period remarkably affects the performance of the segmentation: this is consistent with the observation of Ibbitt and Pearson (1987). It is therefore important to consider this indicator when planning gauging campaigns or deriving gauging strategies. According to the simulations, the availability of 20 gaugings per period on average leads to an acceptable identification of rating shifts. On the contrary, less than 10 gaugings per period may lead to poor segmentation. However, these numbers should be considered as rough orders of magnitudes

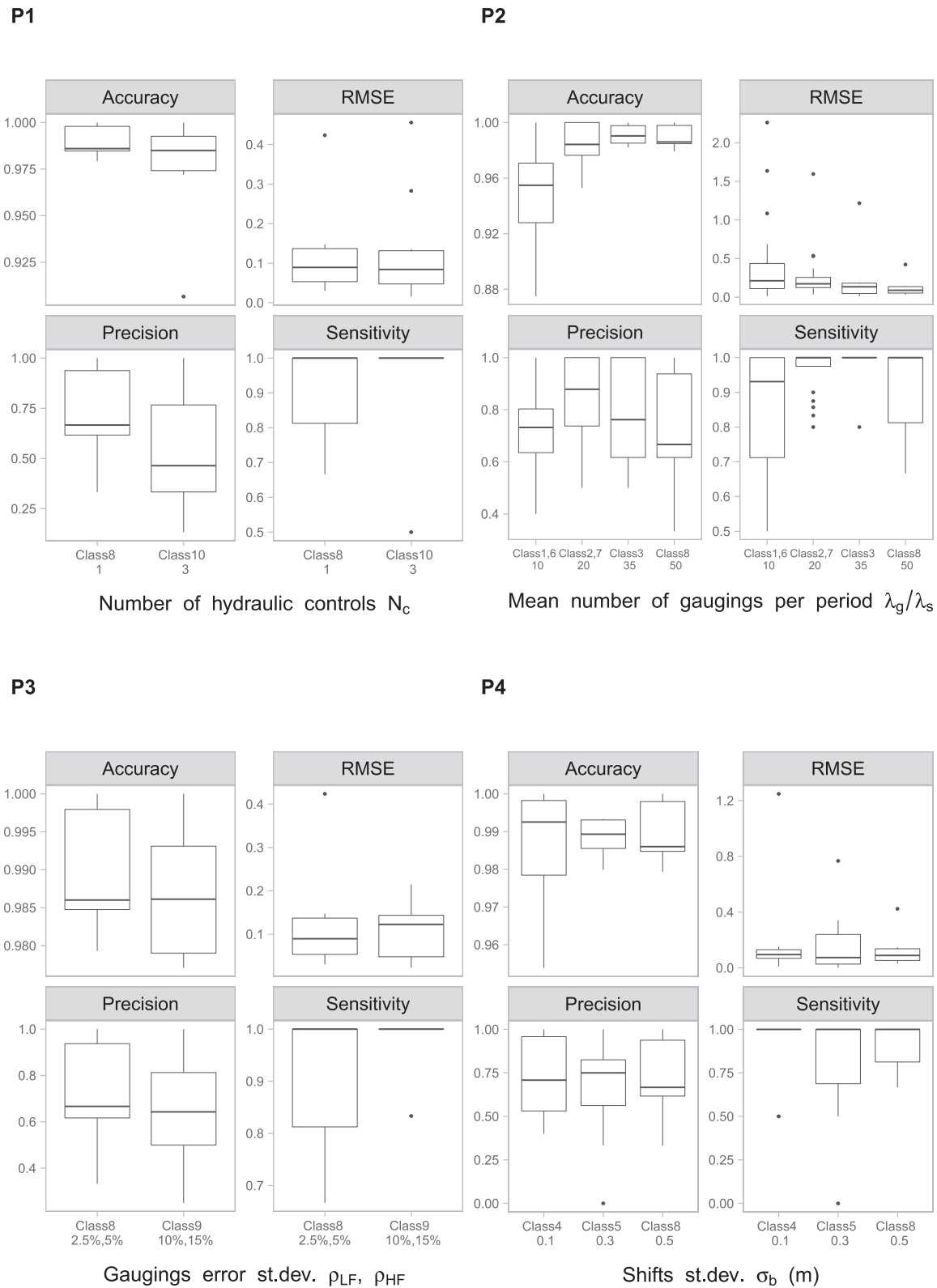


Figure 10. Results of the second experiment: impact of the four properties P1 to P4 on the performance metrics (Strategy D).

rather than precise figures since not all existing hydraulic configurations, shift magnitudes and gaugings uncertainties have been tested.

5.2. Current Limitations

The method proposed in this paper is built on the main assumption that changes correspond to sudden shifts (as opposed to slower transient changes), which may be inadequate for phenomena such as vegetation growth and decay. The segmentation model proposed here is not designed to analyze a trend in the residuals. Dynamic approaches such as those discussed in the introduction Section 1.2 should instead be favored.

The case studies have illustrated the added value of expressing change points in terms of time (rather than position) and of adjusting the shift time by looking for some causative events within the uncertainty bound of the estimated change point, as described in Section 2.6. Nevertheless, this adjustment must be done with precaution. Determining the cause of a shift is a complex decision since several potential processes might be suspected (e.g., flood, gravel mining). This necessarily comes with a degree of expertise and subjectivity. For instance, how to separate the sudden shift created by a morphogenic flood from apparent shifts induced by transient phenomena such as vegetation, backwater, etc.? How to choose when a single shift may be attributed to several floods? Introducing some degree of expertise and subjectivity is not problematic in our opinion. It may even well improve an otherwise fully automated procedure.

5.3. Avenues for Future Work

The method proposed in this paper can, in principle, be used with any RC method that provides RC uncertainty. Results from various RC methods could be compared in future work to assess the extent to which the detected changes are robust with respect to the RC method.

The method is also based on the analysis of residuals computed with respect to discharge. They may be computed with respect to the stage too, as suggested by Morlot et al. (2014). It would therefore be of interest to modify the method to enable the use of stage residuals, and to evaluate whether it has an important impact on the detected changes. We note however that the treatment of stage residuals, and in particular of their uncertainty, is not straightforward. This is because the equivalent of Equation 2 for the stage is not immediately available and may require some additional error propagation.

More work could also be carried out to refine operational gauging strategies in the presence of shifty RCs. Many factors may affect the efficiency of the rating shift detection, in particular, the number of gaugings, their uncertainty, and their location along the RC. A single gauging may be sufficient if it is precise and far away from the base RC. But in the case of minor shifts or very uncertain gaugings, several gaugings may be required. Notice also that the gaugings for Meyras station (Section 3) have been performed for mostly all flow conditions. This leads to the estimation of relatively precise RCs. More uncertain RCs are likely to require more gaugings to detect a shift of a given magnitude.

The case studies have also evidenced that, among the proposed criteria for the choice of the optimal number of segments, the BIC and the DIC obtained the highest performance. However, the segmentation appears sensitive to these criteria. Future work includes a more exhaustive comparison, as discussed by Buckland et al. (1997) and Burnham and Anderson (2004), by analyzing the weight of each penalty term related to data fit and number of parameters.

The proposed method is also inherently limited by gaugings availability - no change can be detected in their absence. However, we stress that gaugings are not the only information available at hydrometric stations. The proposed tool may be complemented with other sources of information such as the stage record (e.g., by detecting a change in recession shapes or by deriving sediment transport estimates), other independent hydrologic data (e.g., correlation analysis with neighboring stations or with the output of a rainfall-runoff model) and direct observations (e.g., bathymetric surveys).

Finally, from a general perspective, the proposed method could probably be extended to other fields where a relationship between two variables, calibrated with uncertain data, is subject to sudden changes (e.g., relation turbidity vs. total suspended sediment concentration).

6. Conclusions

We propose a method for the detection of rating shifts using gaugings. The method applies a recursive segmentation procedure to the time series of residuals between the gaugings and a time-invariant baseline RC. Unlike other classical methods for the segmentation of residuals, the proposed method formally accounts for both gaugings and RC uncertainties through a Bayesian approach. It also expresses change points and their uncertainties in terms of time rather than position, which is of interest to search for specific events that may have caused the shift. It performs a “top-down” recursive procedure, progressively refining the RC estimation on homogeneous sub-periods and leading to the detection of minor shifts.

The method yielded encouraging results for the Ardèche River at Meyras, France, with the detection of effective rating shifts, in good agreement with the official dates of the RC update. Accounting for the uncertainty in the change point times allowed identifying flood events as likely causes of the shifts. Furthermore, a performance evaluation procedure based on synthetic gauging data sets for which the true shift times are known highlighted the added value of the recursive segmentation procedure and the importance of accounting for both gaugings and RC uncertainties. This approach yielded more accurate results than a “single-pass” strategy or a strategy assuming homoscedastic residuals.

Data Availability Statement

The data and the codes used in this study can be downloaded from INRAE's forge platform (<https://forge.irstea.fr/projects/bam/files>). The codes use the DMSL Fortran library developed by Dmitri Kavetski (University of Adelaide, Australia).

Acknowledgments

The PhD fellowship of M. Darienzo is funded by INRAE, Electricité de France (EDF), the Compagnie Nationale du Rhône (CNR) and the French National Hydrological Services (SCHAPI) under Action 2.1 HYDROM of Program 181 action 10 (Prévention des risques/Prévention des risques naturels et hydrauliques). Data and expert knowledge on the Ardèche River at Meyras station (V5004030) were provided by the French National Hydrological Services (Unité d'Hydrométrie et de Prévision des Crues Grand Delta, Guillaume Fourquet). The authors finally acknowledge the constructive and insightful comments of the three anonymous reviewers and the Associate Editor of this study.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Alexandersson, H. (1986). A homogeneity test applied to precipitation data. *Journal of Climatology*, 6(6), 661–675. <https://doi.org/10.1002/joc.3370060607>
- Aminikhanghahi, S., & Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51(2), 339–367. <https://doi.org/10.1007/s10115-016-0987-z>
- Auger, I., & Lawrence, C. (1989). Algorithms for the optimal identification of segment neighborhoods. *Bulletin of Mathematical Biology*, 51(1), 39–54. <https://doi.org/10.1007/BF02458835>
- Basseville, M., & Nikiforov, I. (1993). Detection of Abrupt Change - Theory and Application (5). PTR Prentice-Hall.
- Booth, N. B., & Smith, A. F. M. (1982). A Bayesian approach to retrospective identification of change-points. *Journal of Econometrics*, 19(1), 7–22. [https://doi.org/10.1016/0304-4076\(82\)90048-3](https://doi.org/10.1016/0304-4076(82)90048-3)
- Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational & Graphical Statistics*, 7(4), 434–455. <https://doi.org/10.1080/10618600.1998.10474787>
- Buckland, S. T., Burnham, K. P., & Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics*, 53, 603–618. <https://doi.org/10.2307/2533961>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference. *Sociological Methods & Research*, 33, 261–304. <https://doi.org/10.1177/0049124104268644>
- Cappé, O., Moulines, E., & Ryden, T. (2005). *Inference in hidden Markov models (Springer series in statistics)*. Springer-Verlag. <https://doi.org/10.1007/0-387-28982-8>
- Chen, J., & Gupta, A. (2012). *Parametric statistical change point analysis: With applications to genetics, medicine, and finance*. Birkhäuser Boston. <https://doi.org/10.1007/978-0-8176-4801-5>
- Chen, J., Gupta, A., & Pan, J. (2006). Information criterion and change point problem for regular models. *Sankhyā: The Indian Journal of Statistics (2003-2007)*, 68(2), 252–282. Retrieved from <http://www.jstor.org/stable/25053496>
- Chernoff, H., & Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Annals of Mathematical Statistics*, 35(3), 999–1018. <https://doi.org/10.1214/aoms/1177700517>
- Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics*, 86(2), 221–241. [https://doi.org/10.1016/S0304-4076\(97\)00115-2](https://doi.org/10.1016/S0304-4076(97)00115-2)
- Ducré-Robitaille, J.-F., Vincent, L. A., & Boulet, G. (2003). Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology*, 23(9), 1087–1101. <https://doi.org/10.1002/joc.924>
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4), 711–732. <https://doi.org/10.1093/biomet/82.4.711>
- Guerrero, J.-L., Westerberg, I. K., Halldin, S., Xu, C.-Y., & Lundin, L.-C. (2012). Temporal variability in stage-discharge relationships. *Journal of Hydrology*, 446–447, 90–102. <https://doi.org/10.1016/j.jhydrol.2012.04.031>

- Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B*, 41(2), 190–195. Retrieved from <http://www.jstor.org/stable/2985032>
- Hawkins, D. M. (1977). Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association*, 72(357), 180–186. Retrieved from <http://www.jstor.org/stable/2286934>
- Herschy, R. W. (1998). Streamflow measurement. In *Encyclopedia of hydrology and lakes* (pp. 635–637). Springer Netherlands. https://doi.org/10.1007/1-4020-4497-6/textf/_j214
- Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, 57(1), 1–17. <https://doi.org/10.2307/2334932>
- Hubert, P., & Carbonnel, J.-P. (1987). Approche statistique de l'aridification de l'Afrique de l'Ouest. *Journal of Hydrology*, 95(1), 165–183. [https://doi.org/10.1016/0022-1694\(87\)90123-5](https://doi.org/10.1016/0022-1694(87)90123-5)
- Hubert, P., Carbonnel, J. P., & Chauouche, A. (1989). Segmentation des séries hydrométéorologiques - application à des séries de précipitations et de débits de l'Afrique de l'Ouest. *Journal of Hydrology*, 110(3), 349–367. [https://doi.org/10.1016/0022-1694\(89\)90197-2](https://doi.org/10.1016/0022-1694(89)90197-2)
- Ibbitt, R. P., & Pearson, C. P. (1987). Gauging frequency and detection of rating changes. *Hydrological Sciences Journal*, 32(1), 85–103. <https://doi.org/10.1080/02626668709491164>
- Jalbert, J., Mathevet, T., & Favre, A.-C. (2011). Temporal uncertainty estimation of discharges from rating curves using a variographic analysis. *Journal of Hydrology*, 397(1), 83–92. <https://doi.org/10.1016/j.jhydrol.2010.11.031>
- Jandhyala, V., Fotopoulos, S., MacNeill, I., & Liu, P. (2013). Inference for single and multiple change-points in time series. *The Journal of Time Series Analysis*, 34(4), 423–446. <https://doi.org/10.1111/jtsa.12035>
- Keogh, E., Chu, S., Hart, D., & Pazzani, M. (2003). Segmenting time series: A survey and novel approach. *Data Mining in Time Series Databases*, 57. https://doi.org/10.1142/9789812565402/textf/_j0001
- Kiang, J. E., Gazoorian, C., McMillan, H., Coxon, G., Le Coz, J., Westerberg, I. K., et al. (2018). A comparison of methods for streamflow uncertainty estimation. *Water Resources Research*, 54(10), 7149–7176. <https://doi.org/10.1029/2018WR022708>
- Killick, R., & Eckley, I. (2014). Changepoint: An R Package for Changepoint Analysis. *Journal of Statistical Software*, 58. <https://doi.org/10.18637/jss.v058.i03>
- Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500), 1590–1598. <https://doi.org/10.1080/01621459.2012.737745>
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583–621. <https://doi.org/10.2307/2280779>
- Lapuszek, M., & Lenar-Matyas, A. (2015). *Methods of analysis the riverbed evolution. A case study of two tributaries of the upper Vistula River. Infrastructure and Ecology of Rural Areas*, Nr IV/3/2015, Polska Akademia Nauk. <https://doi.org/10.14597/infraeco.2015.4.3.095>
- Lavielle, M., & Lebarbier, E. (2001). An application of mcmc methods for the multiple change-points problem. *Signal Processing*, 81(1), 39–53. [https://doi.org/10.1016/S0165-1684\(00\)00189-4](https://doi.org/10.1016/S0165-1684(00)00189-4)
- Le Coz, J., Renard, B., Bonnifait, L., Branger, F., & Le Boursicaud, R. (2014). Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach. *Journal of Hydrology*, 509, 573–587. <https://doi.org/10.1016/j.jhydrol.2013.11.016>
- Lee, A. F. S., & Heghinian, S. M. (1977). A shift of the mean level in a sequence of independent normal random variables—A Bayesian approach. *Technometrics*, 19(4), 503–506. <https://doi.org/10.2307/1267892>
- Luong, T., Perduca, V., & Nuel, G. (2012). *Hidden markov model applications in change-point analysis*. arXiv. Retrieved from <https://arxiv.org/abs/1212.1778>
- Mansanarez, V., Renard, B., Coz, J. L., Lang, M., & Darienzo, M. (2019). Shift happens! Adjusting stage-discharge rating curves to morphological changes at known times. *Water Resources Research*, 55(4), 2876–2899. <https://doi.org/10.1029/2018WR023389>
- McGilchrist, C. A., & Woodyer, K. D. (1975). Note on a distribution-free cusum technique. *Technometrics*, 17(3), 321–325. <https://doi.org/10.2307/1268068>
- McMillan, H., Freer, J., Pappenberger, F., Krueger, T., & Clark, M. (2010). Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. *Hydrological Processes*, 24(10), 1270–1284. <https://doi.org/10.1002/hyp.7587>
- Morlot, T., Perret, C., Favre, A.-C., & Jalbert, J. (2014). Dynamic rating curve assessment for hydrometric stations and computation of the associated uncertainties: Quality and station management indicators. *Journal of Hydrology*, 517, 173–186. <https://doi.org/10.1016/j.jhydrol.2014.05.007>
- Nam, C. F. H., Aston, J. A. D., & Johansen, A. M. (2012). Quantifying the uncertainty in change points. *Journal of Time Series Analysis*, 33(5), 807–823. <https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9892.2011.00777.x>
- Olshen, A. B., Venkatraman, E. S., Lucito, R., & Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*, 5(4), 557–572. <https://doi.org/10.1093/biostatistics/kxh008>
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2), 100–115. <https://doi.org/10.2307/2333009>
- Perreault, L., Bernier, J., Bobée, B., & Parent, E. (2000a). Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. *Journal of Hydrology*, 235(3), 221–241. [https://doi.org/10.1016/S0022-1694\(00\)00270-5](https://doi.org/10.1016/S0022-1694(00)00270-5)
- Perreault, L., Bernier, J., Bobée, B., & Parent, E. (2000b). Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting. *Journal of Hydrology*, 235(3), 242–263. [https://doi.org/10.1016/S0022-1694\(00\)00271-7](https://doi.org/10.1016/S0022-1694(00)00271-7)
- Perreault, L., Haché, M., Slivitzky, M., & Bobée, B. (1999). Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach. *Stochastic Environmental Research and Risk Assessment*, 13(3), 201–216. <https://doi.org/10.1007/s004770050039>
- Perret, E., Renard, B., & 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. <https://doi.org/10.1029/2020WR027745>
- Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Applied Statistics*, 28(2), 126–135. <https://doi.org/10.2307/2346729>
- Phillips, J. C., & Eaton, B. C. (2009). Detecting the timing of morphologic change using stage-discharge regressions: A case study at Fish-trap Creek, British Columbia, Canada. *Canadian Water Resources Journal*, 34, 285–300. <https://doi.org/10.4296/cwrj3403285>
- Pooley, C. M., & Marion, G. (2018). Bayesian model evidence as a practical alternative to deviance information criterion. *Royal Society Open Science*, 5(3), 171519. <https://doi.org/10.1098/rsos.171519>
- Puechberty, R., Perret, C., Poligot-Pitsch, S., Battaglia, P., Belleville, A., Bompard, P., & Rauzy, G. (2017). Charte qualité de l'hydrométrie: Guide de bonnes pratiques. Groupe Doppler Hydrométrie. Ministère de l'Environnement, de l'Énergie et de la Mer, France (83). Retrieved from <https://side.developpement-durable.gouv.fr/Default/doc/SYRACUSE/355845/charte-qualite-de-l-hydrometrie-guide-de-bonnes-pratiques>

- Rainville, F., Hutchinson, D., Stead, A., Moncur, D., & Elliott, D. (2016). Hydrometric manual: Data computations, stage-discharge model development and maintenance. *Water Survey of Canada, Environment and Climate Change Canada, Canada*. <https://doi.org/10.26419/res.00126.003>
- Rantz, S. E. (1982). *Measurement and computation of streamflow* (2). Computation of discharge. Water-Supply Paper ed. <https://doi.org/10.3133/wsp2175>
- Reitan, T., & Petersen-Overleir, A. (2011). Dynamic rating curve assessment in unstable rivers using Ornstein-Uhlenbeck processes. *Water Resources Research*, 47(2). <https://doi.org/10.1029/2010WR009504>
- Renard, B., Garreta, V., & Lang, M. (2006). An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima. *Water Resources Research*, 42(12). <https://doi.org/10.1029/2005WR004591>
- Schmidt, A. (2002). *Analysis of stage-discharge relations for open-channel flows and their associated uncertainties* (Doctoral dissertation, University of Illinois at Urbana-Champaign). Retrieved from <http://hdl.handle.net/2142/83191>
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464. Retrieved from <https://doi.org/10.1214/aos/1176344136>
- Scott, A. J., & Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 30(3), 507–512. <https://doi.org/10.2307/2529204>
- Sikorska, A. E., & Renard, B. (2017). Calibrating a hydrological model in stage space to account for rating curve uncertainties: general framework and key challenges. *Advances in Water Resources*, 105, 51–66. <https://doi.org/10.1016/j.advwatres.2017.04.011>
- Spiegelhalter, D. J., Best, N., Carlin, B. P., & Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society: Series B*, 64, 1–34. <https://doi.org/10.1111/1467-9868.00353>
- Westerberg, I., Guerrero, J.-L., Seibert, J., Beven, K. J., & Halldin, S. (2011). Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras. *Hydrological Processes*, 25, 603–613. <https://doi.org/10.1002/hyp.7848>
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80–83. <https://doi.org/10.2307/3001968>
- WMO. (2010). Manual on Stream Gauging (II). Computation of discharge (No. 1044 ed.).