Probabilistic Models for the Uncertain Hydrologist

Benjamin Renard

Mémoire présenté par

Benjamin Renard

Pour l'obtention de l'

Habilitation à Diriger des Recherches

Ecole Doctorale Sciences de l'Environnement / ED 251

Aix-Marseille Université

Probabilistic Models for the Uncertain Hydrologist

Composition du jury

Juliette Blanchet	**Rapporteure**
Directrice de Recherche, CNRS-IGE, Grenoble, France	
Alain Mailhot	**Rapporteur**
Professeur, INRS, Québec, Canada	
Louise Slater	**Rapporteure**
Professeure, University of Oxford, UK	
Benoit Hingray	**Examinateur**
Directeur de Recherche, CNRS-IGE, Grenoble, France	
Philippe Naveau	**Examinateur**
Directeur de Recherche, CNRS-LSCE, Gif-sur-Yvette, France	
Maria-Helena Ramos	**Examinatrice**
Directrice de Recherche, INRAE Antony, France	

Contents

General introduction

Modified from: `https://en.wikipedia.org/wiki/File:Diagram_of_the_Water_Cycle.jpg`

Consider the following practical questions, with a possible underlying motivation being provided in brackets:

1. What is the discharge flowing in this river right now? (Should action be taken against an upstream user because discharge is below the environmental threshold?)

2. What was the peak discharge reached during last week's flood? (Was it large enough to issue a disaster declaration?)

3. Did the great flood of 1910 in Paris exceed the 100-year threshold? (Should this event be taken as the reference for flood planning?)

4. How large will discharge be during the upcoming weeks? (Should we start emptying this reservoir?)

5. How much water will flow in French rivers during the summers of the next 50 years, and how warm will it be? (Will cooling of nuclear plants built along rivers be reliable?)

6. How would floods be affected if this patch of forest was turned into an urban area? (Should this area be urbanized, and if so with what kind of counter-measures?)

A few common threads emerge from these questions and allow setting the scene for the content of this document. First, all questions are related to the quantity of water flowing in rivers, so this manuscript will obviously be about **Hydrology**. Second, all questions request **estimating** unknown quantities. Third, it is quite intuitive that none of these estimates can be perfect, and that they are all affected, to varying degrees, by **uncertainty**. Quantifying this uncertainty is important because it strongly affects the decision to be made (i.e. the answer to the questions in brackets). My research work is structured around the development of **probabilistic models** to make such uncertain hydrologic estimations.

The work presented in this document is therefore essentially of methodological nature: the aim is not to answer the questions above (although I may occasionally try), but rather to build the tools that help others answering them. Moreover, while the methodological motto '*building probabilistic models for the uncertain hydrologist*' is a strong guiding principle, my work is not similarly structured around a few well-identified guiding hydrologic questions. In fact, the example questions above highlight a large diversity of situations, that will also be reflected in the applications presented in this document: analyses may focus on one hydrometric station or cover thousands of catchments all over the country or the world, use a couple of short time series or entire databases of long records, range from droughts to floods, study one given variable or several variables together, etc. The quantity to be estimated may also be in the present (question 1), the past (2-3), the future (4-5) or even in an hypothetical counterfactual world (6). Depending on the context but also on domain-specific usages, the estimation may be called an estimate, a reconstruction, a forecast, a projection, etc. Throughout this report, I will use the word **prediction** in a fairly liberal way to refer to any of these situations.

In order to address the questions above, Hydrologists have several strategies at their disposal. **Measuring streamflow** may be sufficient to answer questions 1-2. Understanding the dominant hydrologic processes that govern streamflow generation may be required to answer questions 4, 5 or 6, and such understanding is typically embedded into **hydrologic models**. Questions 3 and 5 require probabilistic models to describe **hydrologic variability** in time, space and even between variables. These three strategies define the structure of this manuscript:

1. Chapter 1 *Uncertainty in streamflow data* focuses on the production of streamflow time series, and in particular on the uncertainties affecting the rating curve used for this purpose. It describes the development of the generic BaRatin method (Bayesian Rating curve), along with other tools addressing hydrometric challenges such as rating shifts or complex rating curves for stations influenced by e.g. vegetation or backwater effects.

2. Chapter 2 *Uncertainty in and around Hydrologic Models* turns the focus to hydrologic models, and in particular the input, response and structural uncertainties that affect them. It revolves around the development of the BATEA framework (Bayesian Total

Error Analysis) to decompose the total predictive uncertainty into its constitutive sources, and the many challenges that accompanied this development.

3. Chapter 3 *Hydrologic variability* describes the development of probabilistic models of increasing flexibility to describe hydrologic variables. These models may vary in time, vary conditionally on some predictor, vary in space, be multi-variable or a combination of these properties. An underlying motivation behind these developments is to make the best possible use of available data to better understand the natural hydro-climatic variability.

Each chapter is structured in the same way. It starts with an introduction section containing a short and non-exhaustive review of the relevant literature (excluding the papers I contributed to). The aim of this section is to outline the scientific context into which my own work took place. The next few sections then describe the works I contributed to, ending with a specific section describing how this research work was transferred as operational tools. Finally, a last section concludes the chapter by outlining the most important research perspectives in my eyes. These three chapters follow what I considered as a logical order. In particular, streamflow data form the basic material used throughout this manuscript (and arguably in most hydrological studies), so it makes a natural starting point. Interestingly though, the chapters' order does not reflect the chronological order of the work. In particular, most of the work described in Chapter 2 *Uncertainty in and around Hydrologic Models* was made before the one described in Chapter 1 *Uncertainty in streamflow data* - which, in retrospect, was either optimistic or naive.

The work presented in this manuscript covers a 20-year period, since the beginning of my PhD in 2003. It was performed for the most part at the Riverly research unit (formerly known as Hydrology-Hydraulics), in the Lyon-Grenoble Auvergne-Rhône-Alpes regional center (formerly known as Lyon-Villeurbanne) of the INRAE research institute (a merger between INRA and Irstea, formerly known as Cemagref). Part of the work also took place during an 18-month postdoctoral contract at the University of Newcastle, Australia (2007-2008), a 6-month visit at Columbia University in the City of New York, USA (2013-2014) and a 2-year visit at the University of Adelaide, Australia (2019-2021).

Chapter 1

Uncertainty in streamflow data

1.1 Introduction: the production of streamflow time series

Streamflow time series are the cornerstone of hydrologic analyses. They are used to estimate freshwater resources, perform low-flow or flood frequency analysis, detect changes in hydrologic regimes, calibrate and evaluate hydrologic models, etc. Yet, unlike many environmental variables, streamflow time series are not direct measurements: they result from the transformation of continuously-measured water stages, by means of a model called the rating curve. This process induces specific challenges to derive, calibrate and maintain rating curves [WMO, 2010] and to quantify the surrounding uncertainties [Kiang et al., 2018]. Figure 1.1 provides an overview of the main steps needed to produce streamflow time series using rating curves and is commented in detail in the following sections.

1.1.1 Formulating the rating curve

In its simplest form, a rating curve is a relation between stage h [m] and discharge Q [$\mathrm{m^3\,s^{-1}}$] depending on a few parameters that can be grouped into a vector $\boldsymbol{\theta}$.
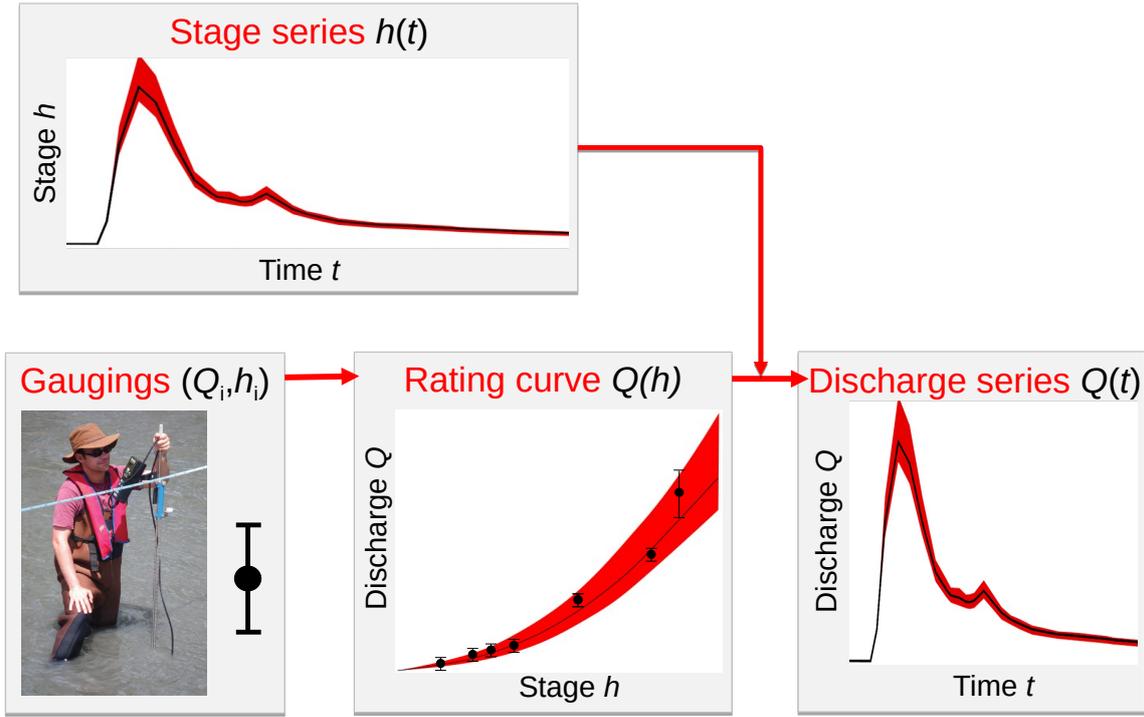
Figure 1.1: Schematic overview of the production of streamflow time series using rating curves. Note that the data shown in this figure are fake, as is the field hydrologist in the bottom left picture ;-)

$$Q = f_{RC}(h; \boldsymbol{\theta}) \tag{1.1}$$

The formulation of rating curve equation f_{RC} is site-specific and depends on the hydraulic controls governing the stage-discharge relation at the hydrometric station. Textbook hydraulics formula suggest that a power equation can be used for many standard hydraulic controls.

$$Q = f_{RC}(h; \boldsymbol{\theta}) = \begin{cases} a(h-b)^c \text{ if } h \geq b \\ 0 \text{ otherwise} \end{cases} \tag{1.2}$$

In this equation, the parameter vector $\boldsymbol{\theta} = (a, b, c)$ includes the cease-to-flow level b [m, also called offset], the exponent c [$-$] and the coefficient a [$\mathrm{m}^{3-c}\,\mathrm{s}^{-1}$]. This simple power formulation is quite versatile and holds for many standard hydraulic controls, as illustrated by the few formulas shown in Table 1.1.

In most cases, a single hydraulic control is not sufficient to describe the rating curve for the whole stage range. Instead, different hydraulic controls may be active at different stage ranges. As an illustration, the following situation is quite frequent: at low flows, the stage-discharge relationship is controlled by a weir (natural or artificial). When the stage increases, the weir gets drowned and the stage-discharge relationship is then controlled by the average geometry and roughness of the main channel. For an even larger stage, part of the water may flow in a floodway or over the banks of the main channel: the stage-discharge relationship is then controlled by two channels: the main channel and the floodway. This leads to a piecewise

Hydraulic control	Control properties	$Q = f_{RC}(h)$
Rectangular weir	Width B [m] Crest elevation b [m] Discharge coef. $C_r \approx 0.4$ [−]	$Q = \underbrace{C_r\sqrt{2g}B}_{a}(h-b)^{3/2}$
Triangular weir	Opening angle ν [−] Vertex elevation b [m] Discharge coef. $C_t \approx 0.31$ [−]	$Q = \underbrace{C_t\sqrt{2g}\tan(\nu/2)}_{a}(h-b)^{5/2}$
Orifice	Area A [m^2] Orifice elevation b [m] Discharge coef. $C_o \approx 0.6$ [−]	$Q = \underbrace{C_o\sqrt{2g}A}_{a}(h-b)^{1/2}$
Wide rectangular channel	Width B [m] Bed elevation b [m] Bed slope S [m] Strickler coef. K_S [m$^{1/3}$s^{-1}]	$Q = \underbrace{K_S\sqrt{S}B}_{a}(h-b)^{5/3}$

Table 1.1: Textbook hydraulics formula for a few standard controls. $g \approx 9.81$ [m s^{-2}] is the gravitational acceleration. Apologies for the notation b vs. B or a vs. A that will certainly confuse the statistically-inclined reader: Hydraulics has its reasons which Statistics does not comprehend.

formulation of the rating curve, which in turn requires defining the activation stages of each control.

1.1.2   Estimating the rating curve

Despite the formulas in Table 1.1 being based on measurable quantities, it is in general not possible to define a rating curve by just analyzing and measuring the physical characteristics of the river at the hydrometric station. This is because natural rivers do not strictly conform to the hypothetical controls for which the formulas of Table 1.1 apply. For instance natural weirs or channels are never strictly rectangular. For a natural channel, the width and slope may vary along the controlling reach, and roughness (Strickler coefficient) cannot be precisely measured. In the case of a multi-control rating curve, which is the rule rather than the exception, activation stages can also be difficult to identify (think about the drowning of a weir for instance).

    For these reasons, the rating curve must be estimated at each hydrometric station from occasional stage-discharge measurements (or "gaugings"). At first sight this may sound like a simple curve-fitting exercise, but a specific estimation strategy is needed to account for the following important points:

1. Gaugings are uncertain, especially the gauged discharge. Moreover, this uncertainty may strongly vary between gaugings, depending on the measurement technique, the gauging procedure, the flow condition, etc. [Herschy, 1999]. For instance, the standard uncertainty for a gauged discharge may be may be as low as 1.5% with the tracer dilution technique in good flow conditions, but as high 20% with surface-only measurements (e.g. floats, video, radar). Ignoring this varying uncertainty may result in some gaugings exerting an unduly large leverage on the inference.

2. While the rating curve parameters are not perfectly measurable quantities, they are still related to physical attributes, and it is important to respect at least the order of magnitude that these attributes imply. Failing to do so may result in hugely uncertain predictions when the rating curve is extrapolated beyond the gauged range - even if the fit within this range is good.

3. Like any model, the rating curve in an approximation of the reality and the discharge values it predicts are therefore affected by errors. These 'structural errors' exist irrespective of the quantity and quality of gaugings, and it is hence important to identify their properties and to quantify the resulting uncertainty.

1.1.3   Using the rating curve to estimate streamflow time series

The streamflow time series can be derived by applying the estimated rating curve to the measured stage series (Figure 1.1). The resulting uncertainty originates from the various sources of uncertainties affecting the rating curve as discussed above, but also from the uncertainties affecting the stage series itself [e.g. Petersen-Øverleir and Reitan, 2005]. There exist many potential sources of stage uncertainty, as discussed by Sauer and Turnipseed [2010]: sensor errors, waves, calibration drifts over time, etc. In order to deliver streamflow time series accompanied by a quantified uncertainty, the dominant sources of stage uncertainty at the hydrometric station need to be identified, quantified and propagated together with rating curve uncertainty.

The production of streamflow time series is further complicated by the fact that the stage-discharge relation may change in a variety of ways and for a variety of reasons [Herschy and Herschy, 2002, Morlot, 2014]. For instance, a rating shift corresponds to the stage-discharge relation suddenly changing, typically after a morphogenic flood inducing a modification of the channel's geometry. The stage-discharge relation may also change more continuously due to progressive sand deposit, seasonal vegetation growth or tidal influence, among others. These situations are encountered quite frequently in the operational practice and pose unresolved challenges to estimate such changing rating curves and to quantify the resulting uncertainties in streamflow series.

1.2   Rating curve uncertainties: the BaRatin framework

The BaRatin (Bayesian Rating curve) framework has been developed in a collegial way with colleagues coming from hydrometry, hydraulics and hydrology backgrounds [Le Coz et al., 2013, 2014b]. Its overarching objective is to address the main challenges described in section 1.1 for formulating, estimating and using rating curves.

1.2.1   BaRatin formulation of the rating curve

Given that most hydraulic controls can be approximated with a power relation between stage and discharge (equation (1.2)), the formulation of the rating curve mostly boils down to iden-
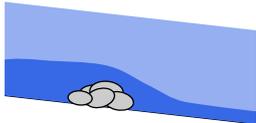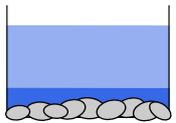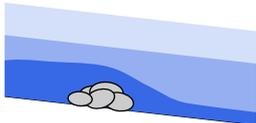
Description	Side view	Front view	Matrix

Natural riffle, then main channel

$$\begin{array}{c} \text{low stage} \\ \text{medium stage} \end{array} \begin{pmatrix} \overset{riffle}{1} & \overset{channel}{0} \\ 0 & 1 \end{pmatrix}$$

Natural riffle, then main channel, then main channel + floodway

$$\begin{array}{c} \text{low stage} \\ \text{medium stage} \\ \text{high stage} \end{array} \begin{pmatrix} \overset{riffle}{1} & \overset{channel}{0} & \overset{floodway}{0} \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

Artificial weir
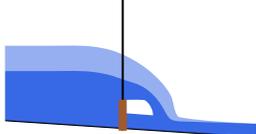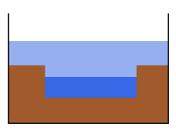
$$\begin{array}{c} \text{low stage} \\ \text{medium stage} \end{array} \begin{pmatrix} \overset{low\ crest}{1} & \overset{high\ crest}{0} \\ 1 & 1 \end{pmatrix}$$

Figure 1.2: A few typical examples of control matrices.

tifying the number, type and succession of hydraulic controls at the hydrometric station. In BaRatin, this specification is performed by means of the control matrix $M(r, j)$ (also called the Bonnifait matrix in honor of its creator), whose columns represent hydraulic controls, rows represent stage segments and the 0/1 value at cell (r, j) indicates whether the jth control is active for the rth stage range. Figure 1.2 provides examples of control matrices for fairly typical configurations.

The control matrix is of interest because it closely follows the line of thinking of the field hydrologist analyzing the hydraulic configuration of a site to formulate the rating curve. Moreover, it allows expressing most rating curves encountered in the operational practice using the unique but versatile rating curve equation below:

$$Q = f_{RC}(h; \boldsymbol{\theta}) = \sum_{r=1}^{N_{segment}} \left(\mathbf{1}_{[\kappa_r; \kappa_{r+1}]}(h) \times \sum_{j=1}^{N_{control}} M(r, j) \times a_j (h - b_j)^{c_j} \right) \tag{1.3}$$

In this equation, the parameter vector $\boldsymbol{\theta}$ that needs to be estimated comprises the activation stage, the coefficient and the exponent of each control ($\boldsymbol{\theta} = (\kappa_j, a_j, c_j)_{j=1,\ldots,N_{control}}$). It does not include the offsets b_j because they can be calculated from the other parameters due to the continuity of the rating curve [see Le Coz et al., 2014b, for details].

1.2.2 Bayesian estimation of the rating curve

As discussed in section 1.1.2, the parameter vector $\boldsymbol{\theta}$ is made of physically interpretable quantities for which information or constraints may exist. For instance, the coefficient a is a function of the geometric properties of the control (see Table 1.1), for which at least orders of magnitude

can be given, and possibly more precise values in the case of artificial controls (low-flow weirs typically). Alternatively, the exponent c only depends on the type of control and should remain close to the theoretical values shown in Table 1.1. In this context, the Bayesian framework is a logical choice to estimate the parameter vector $\boldsymbol{\theta}$ because the prior distribution $p(\boldsymbol{\theta})$ provides a natural way to specify the constraining information discussed above. From an operational perspective, the most natural way to proceed is for the user to specify a prior distribution for each physically interpretable quantity, and to deduce the resulting prior distributions in terms of (κ, a, c) parameterization. For instance, consider the case of a rectangular weir (first row in Table 1.1): the user specifies individual prior distributions for the width B, the discharge coefficient C_r and the gravitational acceleration g, and the resulting prior for parameter $a = C_r\sqrt{2g}B$ can be computed analytically if possible (e.g. if lognormal priors are used), approximately otherwise (e.g. Gaussian approximation or Monte Carlo propagation).

The prior distribution quantifies the information on the rating curve that can be gained from an hydraulic analysis of the controls prevailing at the station. The second main source of information is the gaugings dataset containing occasional stage-discharge measurements $(\tilde{h}_i, \widetilde{Q}_i)_{i=1,\dots,N}$. The use of this information to estimate the rating curve is based on the following error model:

$$\begin{cases} \widetilde{Q}_i = \underbrace{f_{RC}(\tilde{h}_i; \boldsymbol{\theta})}_{\widehat{Q}_i} + \delta_i + \varepsilon_i \\ \delta_i \sim \mathcal{N}(0, u_i) \\ \varepsilon_i \sim \mathcal{N}(0, \sigma_i); \sigma_i = \gamma_1 + \gamma_2 \widehat{Q}_i \end{cases} \tag{1.4}$$

Equation (1.4) assumes that the discrepancy between the gauged discharge $\widetilde{Q}_i$ and the rating-curve-estimated discharge $\widehat{Q}_i$ is due to two errors:

1. δ_i is the measurement error affecting the gauged discharge. In BaRatin this error is assumed to be a realization from a Gaussian distribution with mean zero and *known* standard deviation u_i. The latter represents the uncertainty affecting the gauged discharge: it depends on the measurement procedure and conditions and it can and should be estimated before rating curve estimation by means of an uncertainty analysis.

2. ε_i is thought of as the structural error affecting the rating curve. Like the measurement error δ_i, it is assumed to be a realization from a zero-mean Gaussian distribution, but unlike it, its standard deviation σ_i is assumed *unknown* and therefore needs to be estimated. This standard deviation is further parameterized as an affine function of the rating curve discharge to reflect the empirical observation that rating curve uncertainty tends to increase with discharge, without necessarily collapsing to zero for near-zero discharges.

The different treatment of measurement and structural errors is due to their fundamentally different nature: the former exists independently of the rating curve, while the latter is by definition related to the rating curve and its properties can hence hardly be known before

rating curve estimation. Moreover, it has to be acknowledged that the estimated σ_i does not exclusively represents structural uncertainty in most cases, since its estimation may compensate for any misspecified or ignored source of uncertainty (typically, underestimated measurement uncertainty or ignored rating shifts). The structural error is therefore also termed the remnant error, or the uncertainty garbage collector.

Further assuming independence within and between both types of errors, the error model of equation (1.4) induces the following likelihood:

$$p\left(\widetilde{\boldsymbol{Q}}|\boldsymbol{\theta},\boldsymbol{\gamma}\right) = \prod_{i=1}^{N} f_{\mathcal{N}}\left(\widetilde{Q}_i; \widehat{Q}_i(\boldsymbol{\theta}), \sqrt{u_i^2 + \left(\gamma_1 + \gamma_2\widehat{Q}_i(\boldsymbol{\theta})\right)^2}\right) \tag{1.5}$$

where $f_{\mathcal{N}}(x; m, s)$ is the probability density function (pdf) of a $\mathcal{N}(m, s)$ normal distribution with mean m and standard deviation s evaluated at value x. The likelihood encapsulates the information brought by the gaugings to estimate the rating curve parameters $\boldsymbol{\theta}$ and the structural error parameters $\boldsymbol{\gamma}$. In a Bayesian context, it can be combined with the hydraulic information encapsulated in the prior distribution to derive the posterior distribution:

$$p\left(\boldsymbol{\theta},\boldsymbol{\gamma}|\widetilde{\boldsymbol{Q}}\right) \propto p\left(\widetilde{\boldsymbol{Q}}|\boldsymbol{\theta},\boldsymbol{\gamma}\right) p\left(\boldsymbol{\theta}\right) p\left(\boldsymbol{\gamma}\right) \tag{1.6}$$

The posterior distribution quantifies the uncertainty in estimated parameters $(\boldsymbol{\theta}, \boldsymbol{\gamma})$. In practice, it is explored by means of a Markov Chain Monte Carlo (MCMC) sampler described in details in Renard et al. [2006a], leading to a large number of parameter vectors $\left(\boldsymbol{\theta}_j, \boldsymbol{\gamma}_j\right)_{j=1,...,N_{sim}}$ that are realizations from the posterior distribution. In addition, the parameter vector $\left(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\gamma}}\right)$ maximizing the posterior pdf can be used as a point-estimate and is called the maxpost parameter vector. This allows defining the following quantities, illustrated in Figure 1.3:

1. the **maxpost rating curve** $\widehat{Q}(h) = f_{RC}(h, \widehat{\boldsymbol{\theta}})$ represents the 'best' estimate of the rating curve.

2. the many rating curves $(f_{RC}(h, \boldsymbol{\theta}_j))_{j=1,...,N_{sim}}$ represent the **parametric uncertainty** around the rating curve, due to the imperfect estimation of parameters $\boldsymbol{\theta}$. For a given stage h, a 95% uncertainty interval can be obtained by computing the 2.5% and 97.5% empirical quantiles from the N_{sim} realizations.

3. the many rating curves $\left(f_{RC}(h, \boldsymbol{\theta}_j) + \varepsilon_j(\boldsymbol{\gamma}_j)\right)_{j=1,...,N_{sim}}$, where ε_j is a structural error sampled from the Gaussian distribution in equation (1.4), represent the **total uncertainty** around the rating curve, combining parametric and structural uncertainties.

The representation of uncertainties using many rating curves in points 2-3 is referred to as the 'spaghetti approach' for reasons that should be apparent from Figure 1.3.

1.2.3 Uncertain streamflow time series

The derivation of uncertain streamflow time series follows the same spaghetti-based approach as the one explained above for the rating curve: given a time series $h(t)$, which is typically
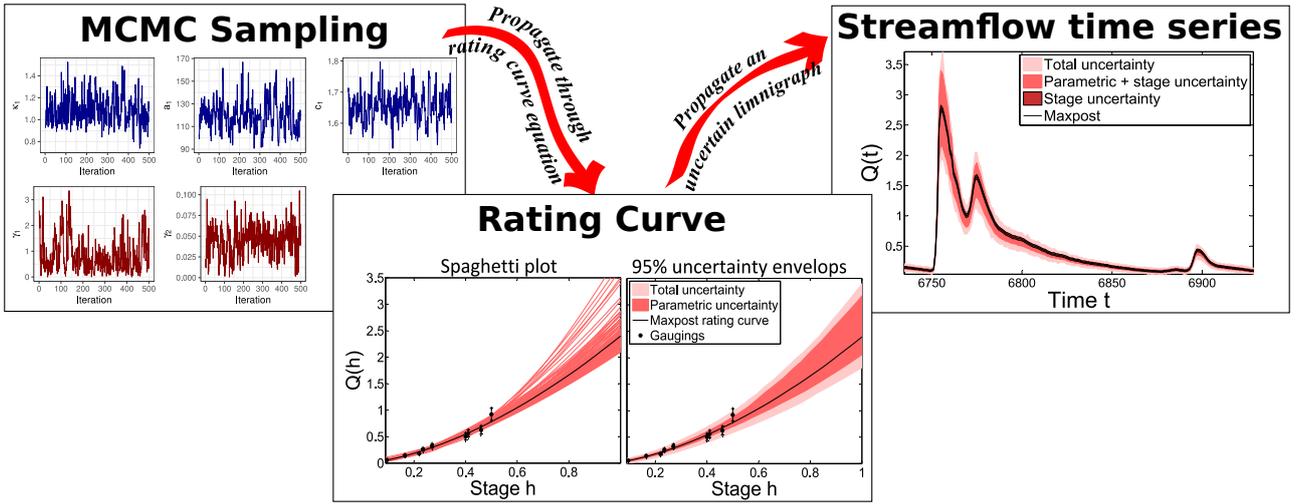
Figure 1.3: Use of MCMC samples from the posterior distribution to estimate the rating curve and the streamflow time series, with uncertainties.

measured at a fine and variable time step, a large number of streamflow time series can be computed as follows:

$$Q_j(t) = f_{RC}(h(t), \boldsymbol{\theta}_j) + \varepsilon_j(t, \boldsymbol{\gamma}_j) \text{ for } j = 1, \dots, N_{sim} \tag{1.7}$$

The representation of streamflow uncertainty using spaghetti has one key advantage: it allows computing the uncertainty of time-averaged streamflow series (e.g. daily or monthly series) in a straightforward way. Indeed, this can simply be achieved by computing e.g. the monthly-averaged series for each individual strand of spaghetti, leading to N_{sim} monthly series that can be further processed to derive an uncertainty interval for instance. By contrast, it would not be possible to deduce the uncertainty interval around a monthly series from the sole knowledge of the uncertainty interval around a finer-time-step series.

Moreover, two important remarks can be made on equation (1.7):

1. The structural error $\varepsilon_j(t, \boldsymbol{\gamma}_j)$ needs to be defined for each time step. In section 1.2.2, the assumption that structural errors are independent between successive gaugings was made to derive the likelihood. In order to remain consistent with this assumption, there is no choice but to resample $\varepsilon_j(t, \boldsymbol{\gamma}_j)$ from its Gaussian distribution in equation (1.4) independently at each time step. Admittedly, this is not a satisfying treatment of structural errors, and this point will be further discussed in section 1.5.

2. Using a single time series $h(t)$ implies assuming that there is no uncertainty in the measured stage time series. This is unrealistic and is discussed next.

In the spirit of the spaghetti approach, the propagation of stage uncertainty poses no technical difficulty. Indeed, if N_{sim} stage series $h_j(t)$ are available, they can all be passed through equation (1.7) to propagate uncertainty from stage to streamflow. The key question, however, is how to generate these N_{sim} stage series to realistically represent uncertainty. This question

was addressed by Horner et al. [2018b], who proposed a stage error model composed of two components: a non-systematic term representing errors independent from one time step to the next (waves, instrumental noise) and a systematic term typically representing the unknown bias induced by the occasional (e.g. monthly) re-calibration of the stage sensor against the reference staff gauge. Results highlighted that while non-systematic errors have in general a negligible effect, systematic stage errors may have a much more pronounced effect and may even be one of the main contributors to streamflow uncertainty. More precisely, the contribution of systematic stage errors to the total uncertainty was found to strongly depend on the flow range and the averaging interval. For instance, Figure 1.4 shows a case where stage uncertainty due to systematic errors is the main contributor to the uncertainty of weekly to monthly streamflow series during low flow periods. This was further related to the sensitivity of hydraulic controls in subsequent work [Horner et al., 2018a, 2022].

1.3   Rating curve and other hydrometric challenges

1.3.1   Complex rating curves

In some hydraulic configurations, discharge cannot be computed from stage alone. The rating curve does not take the form of a simple stage-discharge relation $Q = f(h)$ but becomes $Q = f(h, x)$ where x denotes another variable (or possibly several variables) needed to compute discharge. Such cases pose significant operational challenges and were studied in the PhD work of Mansanarez [2016]. The case of hydraulic hysteresis was addressed first: for a given stage value, discharge may be greater during the rising limb than during the falling limb of a flood event. Hysteresis can have noticeable effects when floods with large stage gradients propagate in a flat channel. Unduly ignoring hysteresis may result in biased streamflow time series, including an underestimated flood peak and a delayed flood hydrograph (Figure 1.5). Mansanarez [2016] studied several equations $Q = f(h, \partial h/\partial t)$ including the stage gradient $\partial h/\partial t$ as secondary variable needed to compute discharge and hence called Stage-Gradient-Discharge (SGD) models. These equations were embedded in the same Bayesian framework as described in section 1.2 for BaRatin to perform parameter estimation. It was shown that providing prior information is necessary to overcome the strong interactions affecting several parameters and make the problem well-posed, but that order-of-magnitude priors are sufficient for this purpose. Several gauging strategies were also compared in order to suggest an optimal strategy to estimate rating curves at stations where hysteresis occurs or is suspected to occur.

Some stations are affected by variable backwater effects due to an unsteady downstream boundary condition, corresponding to stage fluctuations typically induced by a reservoir, a lake or the tide. Computing discharge in such a case requires measuring the fall, i.e. the water surface slope. This is generally achieved by installing a second auxiliary gauge (measuring h_2) usually located along the same river reach as the main gauge (measuring h_1) but further downstream. The water surface slope can then be estimated as $(h_1 - h_2)/L$, where L is the distance between the two gauges. This leads to Stage-Fall-Discharge (SFD) models of the form
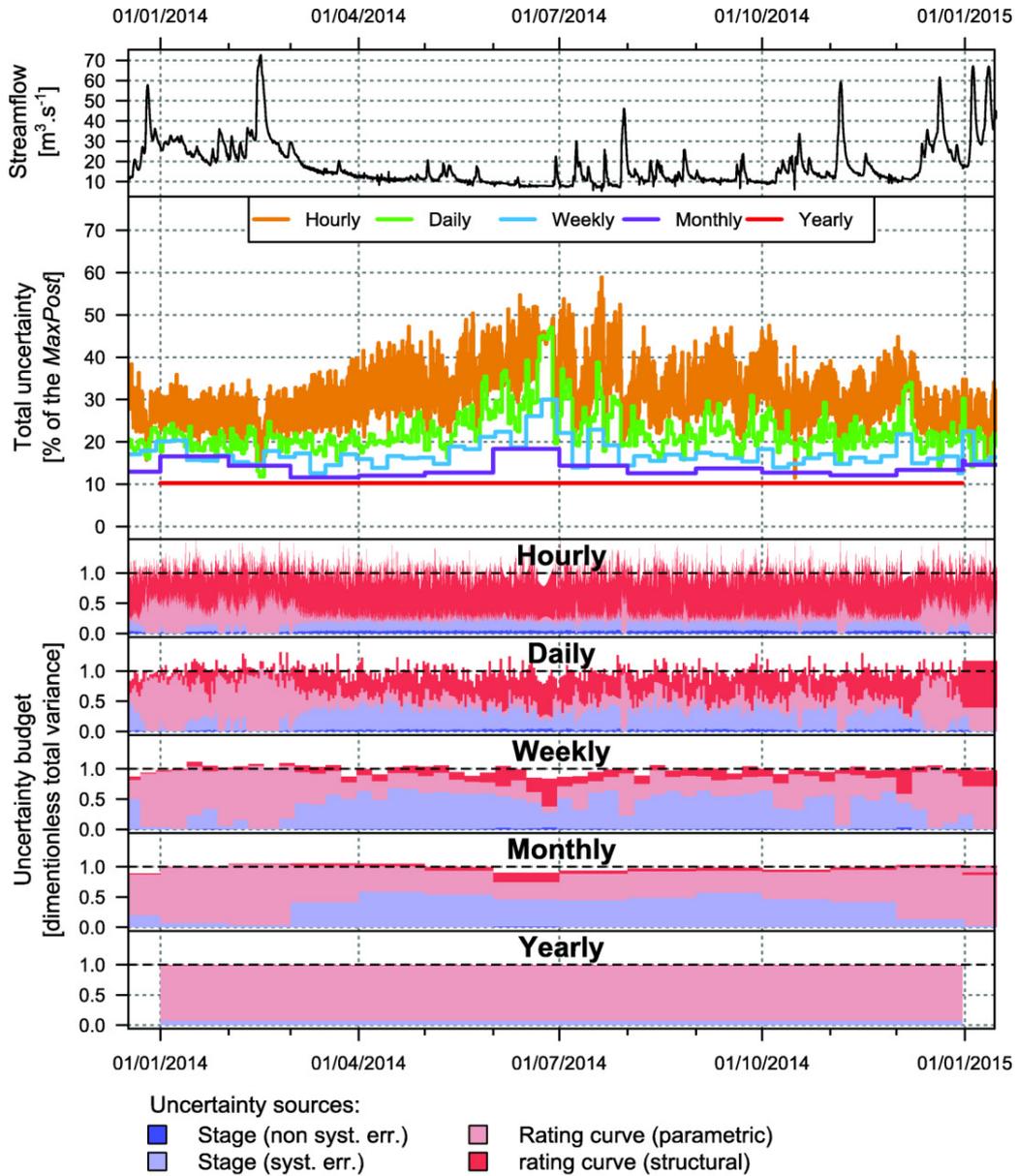
Figure 1.4: Example of uncertainty decomposition into different sources arising from the rating curve or the stage measurement (Blies River at Bliesbruck station). From top to bottom, the hourly MaxPost streamflow time series, the total uncertainty (in %) for various time averaging intervals, and the uncertainty decomposition for the same time averaging intervals. Reproduced from Horner et al. [2018b].
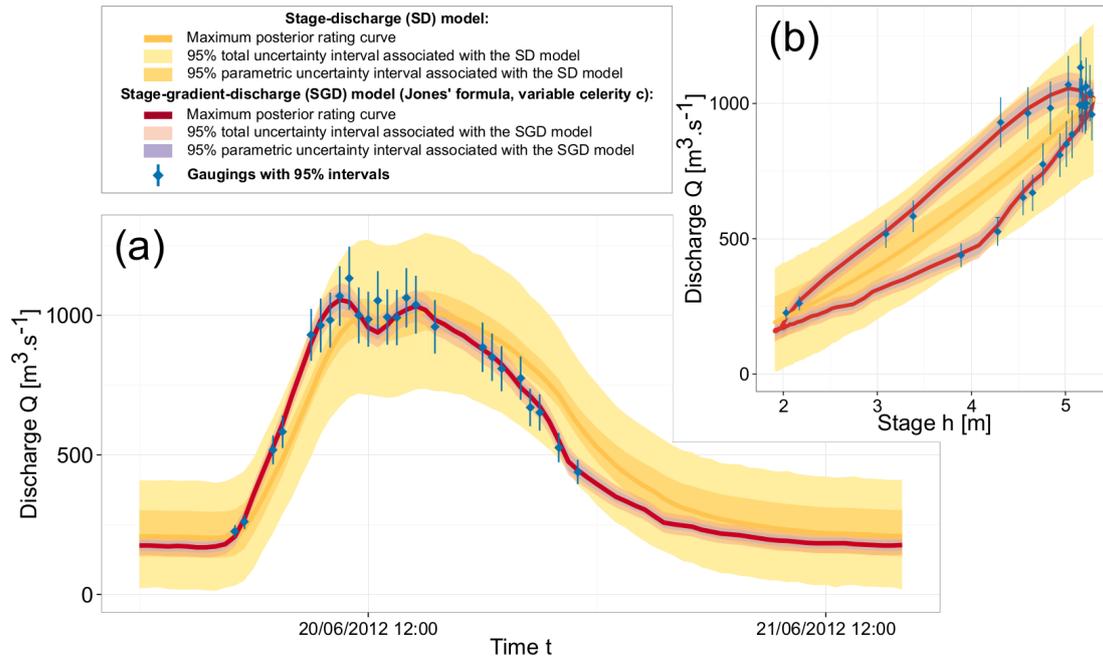
Figure 1.5: Stage-Gradient-Discharge (SGD) model applied to a flood event for the Ebro River at Ascó. The figure compares the streamflow computed from the SGD model to the one computed from a simple stage-discharge (SD) model that disregards stage gradients. (a) Streamflow time series; (b) rating curve. Reproduced from Mansanarez [2016].

$Q = f(h_1, h_2)$ that were studied in details by Mansanarez et al. [2016, 2017] under various hydraulic configurations. As with SGD models, it was found that the precise estimation of SFD rating curves is possible in an operational context (Figure 1.6) but requires informative priors and/or specifically adapted gauging strategies.

The seasonal growth and decay of aquatic vegetation affects rating curves for many rivers in France and in the World. Vegetation induces a continuous variation of the stage-discharge relation that is difficult to track because, unlike in the SGD and SFD cases above, there exists no bio-physical model that can be applied with the information typically available at hydrometric stations. The operational practice in France (and possibly elsewhere) is to keep the rating curve unmodified but to correct the stage time series to estimate the stage that would be observed if there were no vegetation. This approach has many limitations. First, it is manual and hence strongly relies on the expertise of the station manager that is in general not formalized and documented, making corrections hardly reproducible. It also heavily relies on the availability of many gaugings to track the evolution of the stage correction in time. Finally, and most importantly, it is conceptually unsatisfying: vegetation affects the *stage-discharge relation*, and it is therefore more logical to act on the rating curve rather than on the measured stage. To address these issues, Perret et al. [2020b] proposed a time-varying rating curve model $Q = f(h, t)$ based on the following premises:

1. aquatic vegetation primarily affects friction, i.e. the flow resistance of a channel;

2. the effect of vegetation on friction can be expressed as a function of the vegetation density and its ability to reconfigure (i.e. to bend and align in the flow direction);
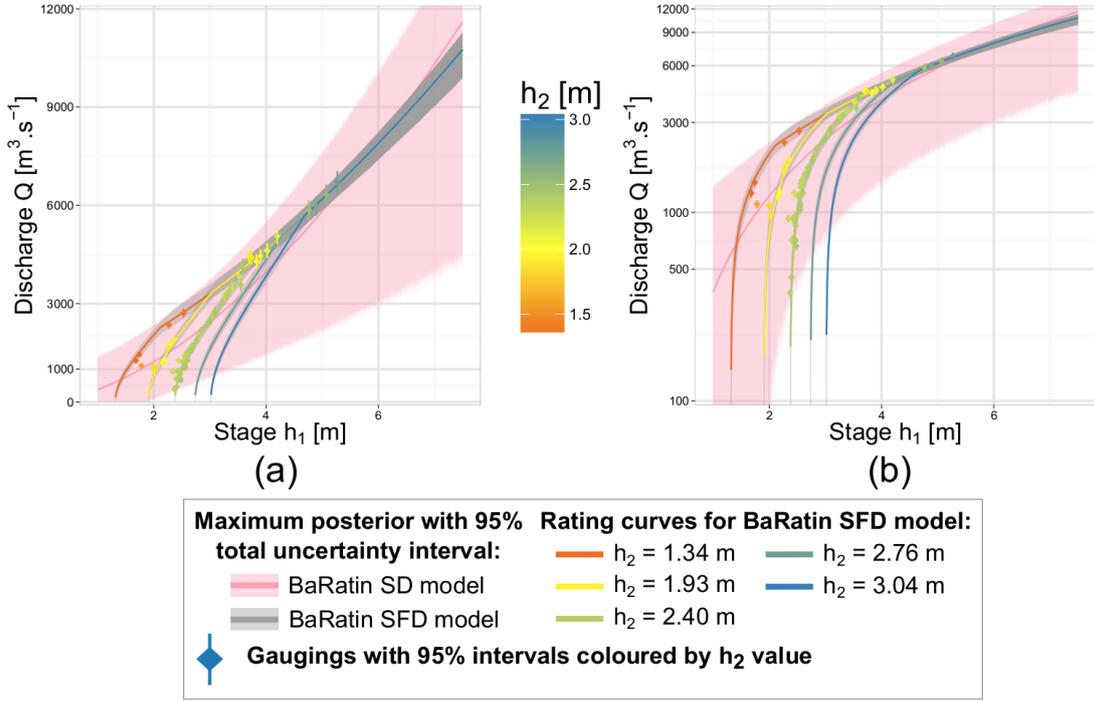
Figure 1.6: Stage-Fall-Discharge (SFD) rating curve for the Rhône River at Valence, compared with a simple stage-discharge (SD) rating curve that only uses h_1 and hence disregards variable backwater effects. (a) Natural scale; (b) logarithmic scale. The SFD rating curves are plotted for 5 values of h_2 corresponding to available gaugings ($h_2 = 1.34$, 1.93, 2.40 and 2.76 m) and to the maximum recorded value (3.04 m). Reproduced from Mansanarez [2016].

3. The growth and decay of vegetation can be parameterized as a deterministic function of time $d(t)$.

Perret et al. [2020b] demonstrated the applicability of this model in an operational context, provided that the parameters of the function $d(t)$ are re-estimated every year to account for the fact that the vegetation cycle, while always following a similar pattern, strongly varies in intensity and duration between years (Figure 1.7). An original aspect of this work was the incorporation of qualitative observations on the plant development state (e.g., no plant, growing phase, decaying phase), in addition to the usual stage-discharge gaugings. This additional information was found to be helpful for the identification of parameters varying every year. Subsequent work attempted to move from a time-varying to a dynamical model incorporating a biological component that describes the evolution of plant biomass as a function of environmental forcings such as irradiance and water temperature [Perret et al., 2022]. While promising, this model still requires adaptations and simplification to make it usable in the operational practice.

1.3.2   Rating shifts

Natural rivers are not static environments and their morphology may change for a variety of reasons, natural (e.g. large flood) or man-made (e.g. channelization). The resulting rating shifts constitute one of the most widespread challenges facing hydrometric data producers.
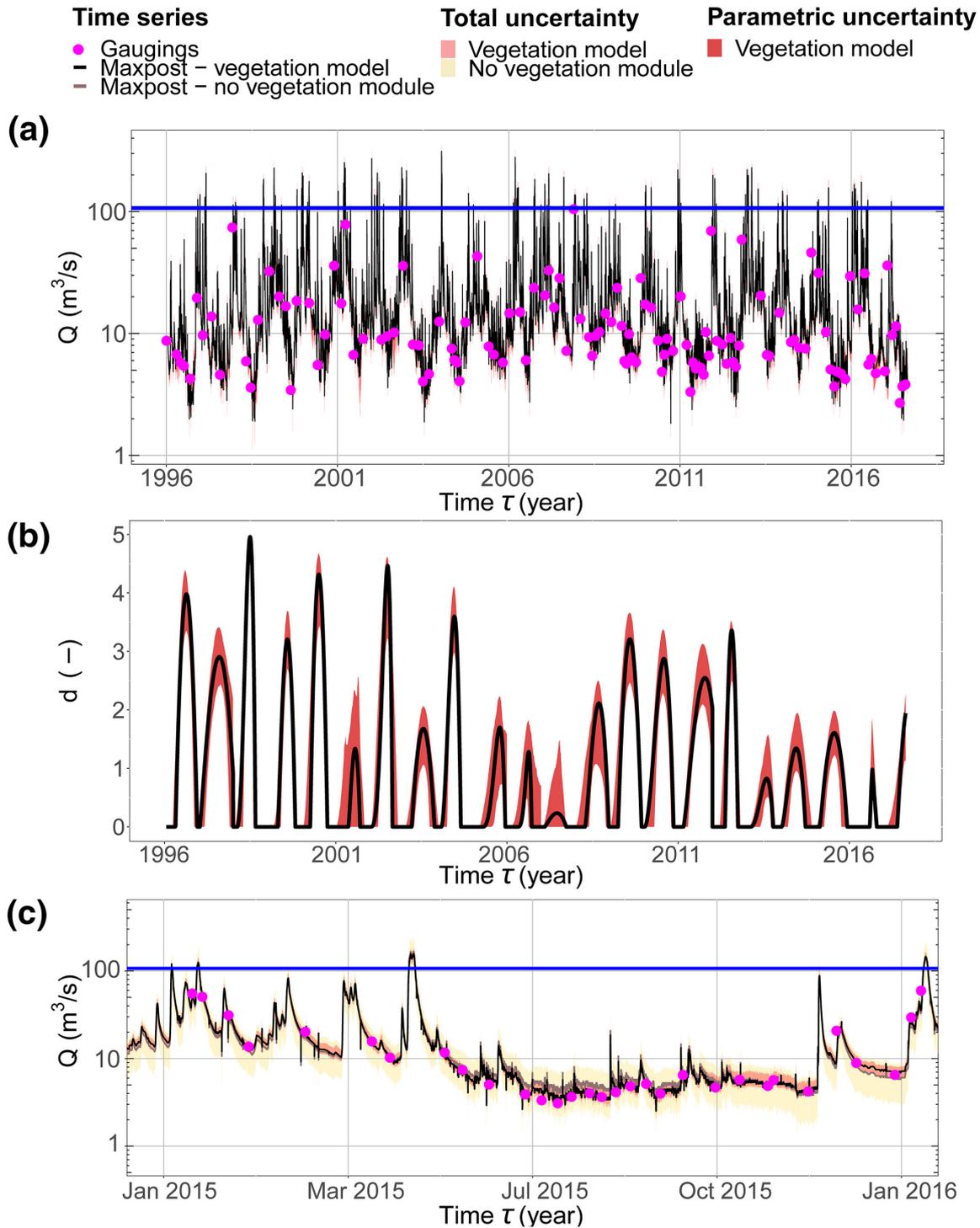
Figure 1.7: Vegetation model applied to the Ill River at Colmar-Ladhof station over a 22-year period. (a) Streamflow time series $Q(t)$ with uncertainties, and available gaugings; (b) estimated vegetation cycle function $d(t)$ with its parametric uncertainty, showing large inter-annual variability; (c) streamflow time series $Q(t)$ for year 2015 obtained with the temporal vegetation model and with a standard model with no vegetation module. Blue lines in (a) and (c) represent a change in hydraulic control occurring for $h > 3\ m$: the vegetation model is not valid above this line. Reproduced from Perret et al. [2020b].
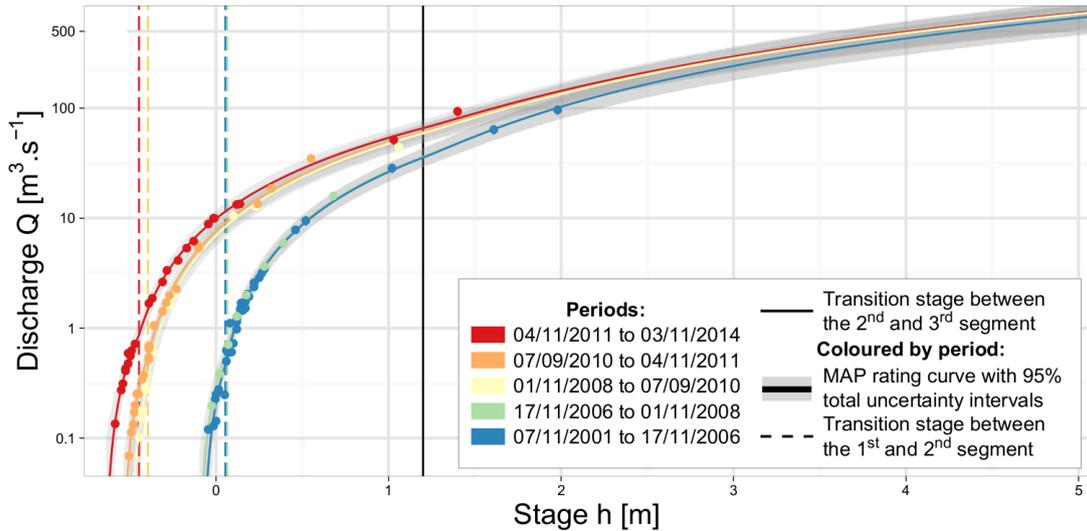
Figure 1.8: Stage-Period-Discharge (SPD) rating curve applied to the Ardèche River at Meyras station. Reproduced from Mansanarez [2016].

Two aspects need to be considered: how to estimate shifting rating curves, without restarting a new curve from scratch at the beginning of each stability period? And how to detect rating shifts?

The first question was addressed in the PhD thesis of Mansanarez [2016], who proposed a Stage-Period-Discharge (SPD) model $Q = f(h, k)$ to estimate shifting rating curves, assuming the stability periods k (or equivalently, the shift times) are known [Figure 1.8, Mansanarez et al., 2019]. The model is based on an hydraulic analysis of the changes affecting the river, which leads to a distinction between two types of rating curve parameters: 'static' parameters remain constant throughout all stability periods, while the value of 'varying' parameters change after each shift. For instance, if shifts only affect the elevation of the riverbed in the main channel, then the corresponding offset parameter b (see equation (1.2)) should be declared as 'varying', while the coefficient a (related to the channel width, slope and friction) and the exponent c (related to the channel shape) may remain 'static'. This distinction is important because it allows using a single model for all periods, and all rating curves can therefore be estimated in one go, using all available gaugings. This is to be compared with the awkward operational practice of repeatedly using gaugings for several periods. Another original aspect of the method is that prior specification is not directly performed on rating curve parameters, but rather on incremental changes, i.e. changes between two successive periods: the station manager is in general able to make at least an order-of-magnitude statement about the possible magnitude of a shift after a morphogenic flood. A change-of-variable then allows deducing the joint prior distribution of 'varying' rating curve parameters, which tend to be highly correlated: this favors the transfer of information between highly-gauged periods and poorly-gauged (or even possibly ungauged) ones.

The detection of rating shifts was studied as part of the PhD work of Darienzo [2020], who explored the use of several sources of information for this purpose. The most widely

used approach to detect rating shifts is based on gaugings through an analysis of the residuals between the gauged and the rating-curve discharges: rating shifts materialize as step changes in this residual series. Darienzo et al. [2021] proposed the recursive segmentation procedure schematized in Figure 1.9 to detect these step changes. It improves on existing segmentation methods by explicitly accounting for the uncertainties affecting both the gaugings and the rating curve. Its Bayesian nature allows accessing to the posterior distributions of shift times: this is useful because it allows searching for a large flood event within the time window covered by this posterior distribution, and allocating the precise shift time to the date of this flood. Finally, its recursive nature allows progressively refining the estimation of rating curves on each sub-period, which enables the detection of small shifts that may have been hidden by a large rating curve uncertainty at the preceding recursion step. Note that the previously-described SPD approach of Mansanarez et al. [2019] is used for multi-period rating curve estimation, which allows considering subperiods with a relatively small number of gaugings.

Using gaugings to detect rating shifts arguably constitutes the most natural approach, but its usefulness is limited by the frequency of gaugings: for instance, it is not possible to detect several shifts occurring in between distant gaugings. The measured stage series constitute an alternative, and possibly overlooked, source of information to detect rating shifts. Darienzo [2020] proposed to use the stage series in two different ways. First, some shifts may be visible in the lowest part of the stage series. Indeed, when streamflow tends to zero, stage tends to the offset b of the lowest hydraulic control, so that a change affecting this offset (strongly related to the riverbed) should materialize in the stage recessions. This idea was formalized in the procedure illustrated in Figure 1.10. It is based on the estimation of a decreasing curve for each recession event, with one parameter of this curve representing the asymptotic recession level, and the application of the segmentation procedure to this parameter. The second approach rather focuses on the highest part of the stage series and borrows ideas from sediment transport modeling. Its basic principle is that a minimum water depth is needed to trigger sediment transport and hence a potential morphological change in the riverbed. Moreover, the potential amplitude of the change is expected to be proportional to the duration and the peak of this triggering threshold excess. A retrospective analysis of the changes detected with other approaches allows estimating this triggering threshold and relating it to physical quantities used in sediment transport models. The main interest of this approach is that it can in principle be used in near-real-time, enabling the early detection of a potential rating shift without waiting for new gaugings or long stage recessions. The main difficulty is that it only allows detecting *potential*, not *actual* changes, because significant scour and fill processes may occur successively during the flood but compensate each other, potentially resulting in a small difference before and after the flood.

1.3.3 Uncertainties in gauging methods

While most of the work described in this chapter is directly related to the estimation of the rating curve, I also contributed to two studies aimed at quantifying the uncertainties affecting
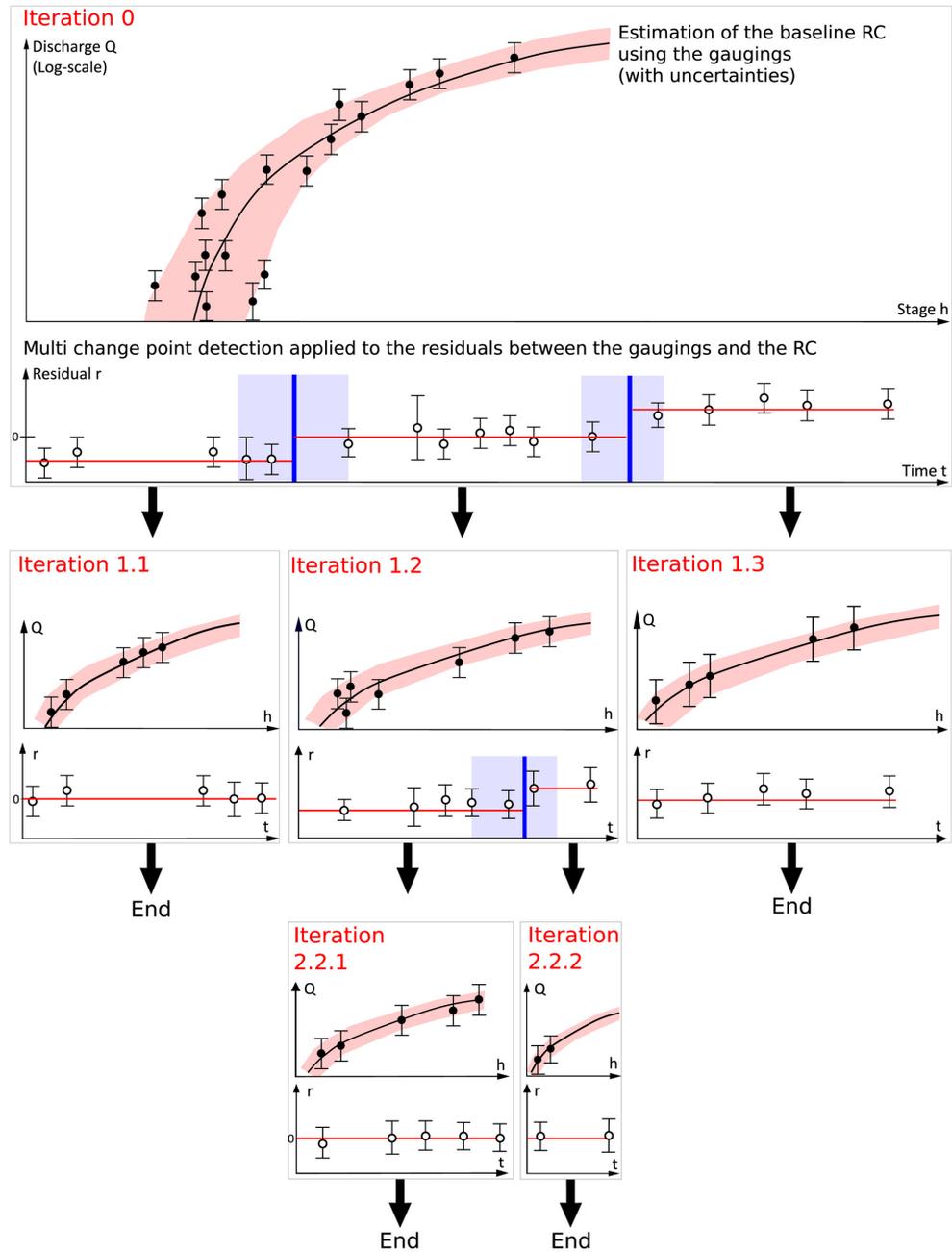
Figure 1.9: Schematic representation of the recursive segmentation procedure. Colored ribbons and error bars represent 95% uncertainty intervals for rating curves (pink), shift times (blue), gaugings (black dots), and residuals (empty dots). Reproduced from Darienzo [2020].
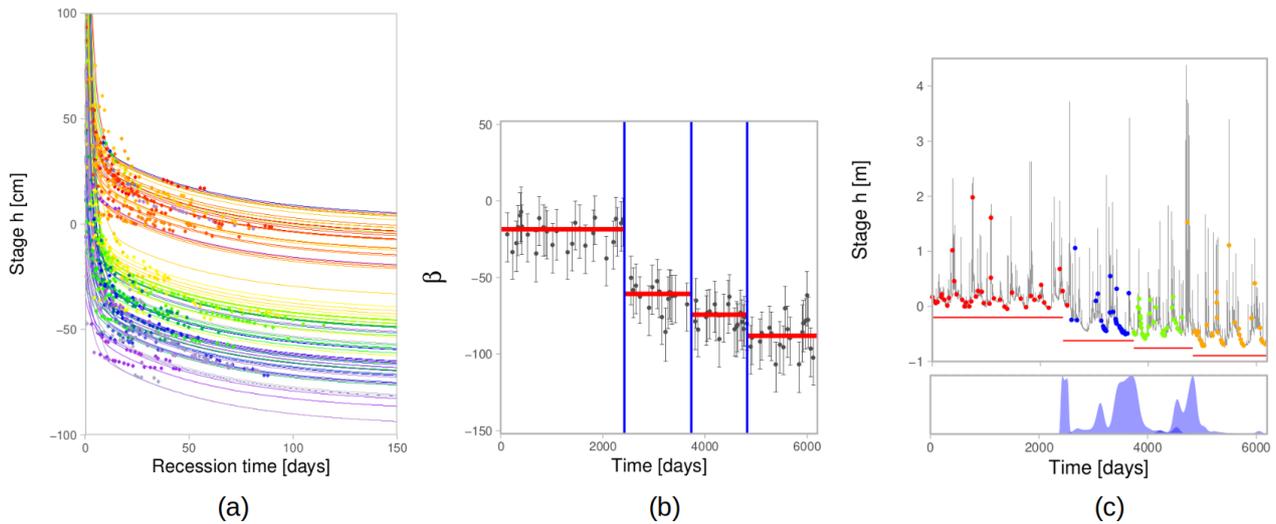
Figure 1.10: Stage recession analysis applied to the Ardèche River at Meyras station. (a) Estimation of recession curves; (b) segmentation applied to the time series of recession parameter β representing the asymptotic recession level; (c) detected shifts represented on the stage time series. Horizontal red lines denote the values of β estimated in (b), blue areas represent the posterior pdfs of shift times. Modified from Darienzo [2020].

specific gauging methods. The first study aimed at decomposing the sources of uncertainty affecting Acoustic Doppler Current Profiler (ADCP) streamflow measurements by analyzing the results of an 'ADCP regatta' - that is, repeated measurements of the same discharge by several operators using different instruments at several cross-sections [Despax et al., 2019]. My contribution restricted to helping in the interpretation of the two-way analysis of variance (ANOVA) that was applied to quantify the cross-section effect, the team (operator + instrument) effect, and their interaction. This type of analysis is interesting because it suggests practical strategies to reduce measurement uncertainties. Here for instance, it was found that for a given number of transects (i.e. measurement effort), exploring several cross sections is more efficient to reduce uncertainty than increasing the number of transects at the same cross section. Allowing several teams to explore these different cross sections is even more efficient, although it does represent a larger measurement effort.

I also contributed to the development of a method to estimate the uncertainty of video-based flow velocity and discharge measurements [Le Coz et al., 2021]. Video-based hydrometry is an area in rapid development since it allows performing contactless discharge measurements, which is particularly interesting in flood conditions. It also allows taking advantage of crowd-sourced videos to reconstruct flood discharges at ungauged locations. An important component of the process used to estimate discharge from videos is the orthorectification step, i.e. the transformation of the images so that each pixel has the same known physical size (Figure 1.11). Two approaches are typically used to estimate the parameters of this transformation: (i) calculate them from intrinsic (focal length, sensor size) and extrinsic (position, angles) parameters of the camera, which of course requires knowing these parameters quite precisely; (ii) calibrate them
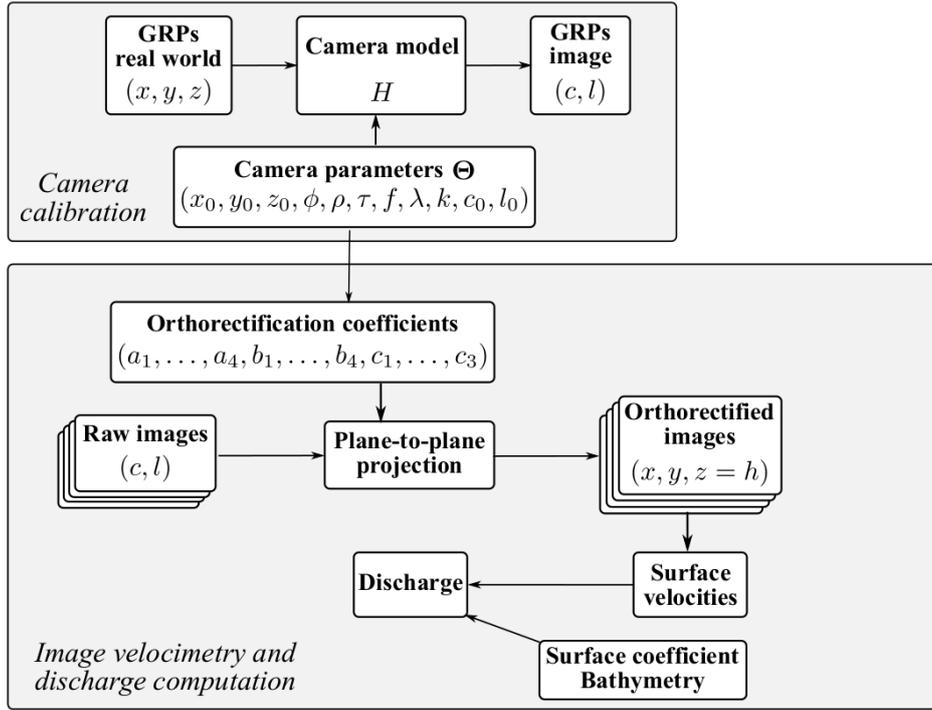
Figure 1.11: Schematic workflow from Bayesian camera calibration (top) to orthorectification, image velocimetry and discharge computation using Fudaa-LSPIV solvers [Le Coz et al., 2014a]. Reproduced from Le Coz et al. [2021].

from ground reference points (GRPs) with known image and real-world coordinates, which completely disregards the physical and optical meaning of the orthorectification parameters. Bayesian estimation offers a natural middle ground, taking advantage of both the availability of GRPs and the existence of prior knowledge on the parameters. Results show that combining observational and prior information is generally beneficial to get precise estimates, especially in edge cases where GRPs are not numerous and camera parameters are not all precisely known. An uncertainty propagation experiment also suggested that orthorectification is an important contributor to the uncertainty affecting video-based streamflow measurements, although not as important as the surface velocity coefficient (i.e. the ratio between the velocity measured at the water surface and the depth-averaged velocity).

1.4   Operational tools and applications

Given the practical significance of streamflow data, turning the methods developed in this chapter into operational tools has been - and still is - a permanent objective (Figure 1.12). BaRatin tools[1] are therefore released as open-source codes and software. BaRatin computational engine[2] takes the form of a FORTRAN-based cross-platform executable that performs all the Bayesian computations, in particular MCMC simulations exploring the posterior distribu-

---

[1]https://github.com/BaRatin-tools
[2]https://github.com/BaRatin-tools/BaRatin

tion. The graphical user interface BaRatinAGE[3] (BaRatin Advanced Graphical Environment) makes the method usable in the operational practice. BaRatinAGE v1 was developed as a Tcl/Tk tool by Laurent Bonnifait in 2013, and given the interest expressed by early users, it was re-implemented in Java and further developed in 2016 by Sylvain Vigneau and myself (BaRatinAGE v2.0). The current version BaRatinAGE v2.2.1 is a cross-platform and multi-language software that has reached maturity and is widely used by hydrometry services in France. This adoption is in part due to the existence of the BaRatinAGE software, but also to the many informal training sessions (a.k.a 'BaRatinades') organized over the years by Laurent Bonnifait, and the regular training now included in the continuous education plan of the Ministry of Environment [Renard et al., 2017]. BaRatin tools have also been tested by several hydrometry services abroad, such as the US Geological Survey [Mason et al., 2016] and the Instituto Nacional del Agua in Argentina [Kazimierski et al., 2021]. They are also used by other research groups for releasing datasets from experimental catchments [e.g. Tolsa et al., 2013, Lundquist et al., 2016, Francke et al., 2018, Gouy et al., 2021] or for addressing various hydrological questions [e.g. Zeroual et al., 2016, Ocio et al., 2017, Garcia et al., 2020, Qiu et al., 2021, Kastali et al., 2021, 2022, Ahrendt et al., 2022]. Finally, BaRatin was included into an international comparison of existing approaches to estimate rating curve uncertainties [Kiang et al., 2018].

While developing models other than the simple $Q = f(h)$ rating curve (see section 1.3), it became apparent that changing the model does not require fundamentally changing the statistical framework used in BaRatin. This led to the development of BaM (Bayesian Modeling), a generalization of BaRatin where the rating curve equation can be replaced by 'any' model [Renard, 2017a], including models for variables other than streamflow [e.g. constituent loads and sediment, Moatar et al., 2020, Perret et al., 2023]. Like BaRatin, BaM tools[4] are released as open-source codes and software. Compared with BaRatin, BaM's computational engine[5] has been generalized so that any model could be plugged in the same statistical framework. The R package RBaM[6] is available to pilot BaM engine, and an online interface[7] was developed by Ivan Horner to handle simple models that can be typed as simple formulas in a text box. Moreover, the v3 of BaRatinAGE is currently under development (by Ivan Horner[8]) and will use BaM as computational engine: this paves the way for the integration of complex rating curve models (SGD, SFD, SPD, vegetation) and tools for detecting rating shifts. Finally, BaM computational engine is already used in an operational context by the Compagnie Nationale du Rhône (CNR[9]) for managing twin-gauge stations with the SFD model, and by the US National Ecological Observatory Network[10] [Cawley, 2018, Harrison et al., 2018] for producing streamflow series.

---

[3]https://github.com/BaRatin-tools/BaRatinAGE
[4]https://github.com/BaM-tools
[5]https://github.com/BaM-tools/BaM
[6]https://github.com/BaM-tools/RBaM
[7]https://hydroapps.recover.inrae.fr/BaMit/
[8]https://www.ihdev.fr
[9]https://www.cnr.tm.fr/
[10]https://data.neonscience.org/
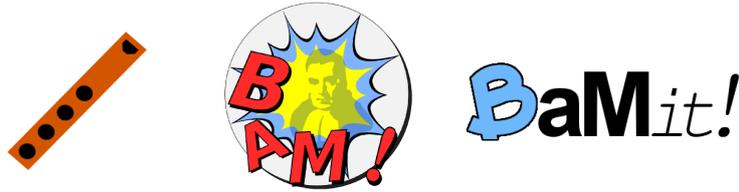
Figure 1.12: Logos for some operational tools: BaRatin, BaM! and BaMit!

Lastly, Matteo Darienzo implemented an R tool for the BAYesian Detection and Estimation of Rating Shifts (BayDERS[11]) based on the methods he developed during his PhD thesis (see section 1.3.2). It has been used in particular in the context of analyzing long hydrometric series [Lang et al., 2022, Lucas et al., 2023]. Work aimed at further developing operational tools for managing rating shifts is still ongoing.

1.5   Conclusion and perspectives

River streamflow is the most fundamental variable for surface hydrology, and streamflow time series are the starting point of many hydrological studies. The stakes are many: managing water resources and risks, evaluating water quality, assessing the impact of climate change and human disturbances on hydrologic regimes, or more fundamentally understanding the water balance and the associated fluxes in catchments. The need for reliable streamflow time series has therefore long been recognized. Moreover, there is a growing demand, arising from both data producers and downstream users, for a quantitative evaluation of the uncertainties involved in the process of producing streamflow series.

In this context, my main contribution has been the development of methods to estimate rating curves and to quantify the associated uncertainties. This includes the general framework BaRatin, built on simple principles from hydraulics and Bayesian statistics, and more specialized developments to address specific issues such as complex rating curves or rating shifts. This methodological work has been accompanied by operationally-oriented developments in the form of open-source codes and software, along with continuous education training modules. I hope this work contributed to demystifying uncertainty quantification and encouraging its routine application in hydrometric services. This may contribute to a virtuous circle: quantitative uncertainties enable data users to propagate them in subsequent analyses or decisions, which in turn further encourages data producers to communicate uncertainties.

The time-varying vegetation model described in section 1.3.1 constitutes a practical solution to the management of rating curves affected by vegetation, especially when completed with the ability to use qualitative observations of the vegetation development state. However this model is not used operationally yet. I believe this is not related to the model itself, but rather to the lack of a user-friendly interface to specify and use it, which can certainly be improved. Attempts at building dynamical models, where plant evolution would be predicted from environmental

---
[11]`https://github.com/MatteoDarienzo/BayDERS`

forcings based on biological principles, have been more frustrating so far, but remain an area worth further investigations [Perret et al., 2022].

Stage-Fall-Discharge (SFD) models proved to be efficient in operational cases where the variable backwater influence is induced by a reservoir. SFD models were also trialed for tidal rivers by Camenen et al. [2017], with mixed results. In particular, SFD models struggled to reproduce the discharge dynamics with strong tidal gradients, or when complex hydraulics phenomena such as floodplain overflow occur. The most promising research direction to overcome these issues is to replace the SFD model by a 1D hydrodynamic model such as MAGE [Perret et al., 2020a]. This represents a challenging but interesting test for BaM, which is supposed to be able to host 'any' model. The integration of MAGE into BaM is still under active development but a first successful attempt was performed recently [Mendez-Rios, 2022]. This opens an important research avenue for the analysis of uncertainties affecting complex and distributed hydrodynamics models, well beyond the specific case of tidal rivers.

From a methodological standpoint, two specific issues need to be further analyzed. The first one is the treatment of systematic errors affecting some gauging methods. In particular, video-based methods are able to provide high-frequency estimates of discharge during floods, but all these estimates rely on the same coefficient representing the ratio between surface and depth-averaged velocities. If this coefficient is wrong by e.g. 10%, all discharge estimates are affected by the same 10% error. Dedicated analyses suggested that ignoring the systematic nature of these errors was not an acceptable option, while a Monte-Carlo propagation approach provided a viable alternative [Renard, 2018]. The latter approach remains to be integrated in operational tools, with a RBaM-based workflow being currently developed by colleagues from Tenevia[12].

The second methodological challenge to be addressed is related to the assumption that structural errors are independent. This is probably not a major issue for rating curve estimation given that gaugings are generally far apart in time, but this is much more problematic when propagating structural uncertainty to streamflow time series having a high temporal resolution. In this case, structural errors are certainly not independent from one time step to the next. Since the uncertainty induced by independent errors quickly decreases with temporal averaging, unduly making an independence assumption may lead to an underestimation of the structural uncertainty affecting e.g. daily or monthly streamflow series. This is particularly problematic when structural uncertainty is large and dominates parametric uncertainty. Our practical advice to users in this case is to try reducing structural uncertainty, e.g. by investigating if rating shifts may have been missed or by trying alternative formulations of the rating curve. More satisfying methodological solutions to this problem remain to be found. In theory, it would be perfectly possible to use a time series model (e.g. of the ARMA family) to avoid this independence assumption. In practice, however, identifying such a temporal dependence structure is difficult because gaugings are performed too sporadically. Alternatives based on the modeling of conditional bias (i.e. a non-zero mean expressed as a function of stage) could

---

[12]https://www.tenevia.com/

be explored.

More generally, given the vitality of modern hydrometry and the regular introduction of new (or sometimes newly-rediscovered) technologies to measure streamflow [Le Coz, 2017], I trust my imaginative colleagues to never fall short of new challenges to tackle. The recent launch of the SWOT altimetry satellite[13] in December 2022 is worth a particular note. SWOT provides snapshots of water levels in lakes and large rivers, which is of immediate interest to estimate streamflow in large, poorly-gauged or trans-boundary catchments [Biancamaria et al., 2016, Durand et al., 2016, 2023]. In places where a dense hydrometric network exists, SWOT is not going to replace it [Fekete et al., 2015], but it might be able to bring interesting complementary information [Le Coz et al., 2018]. In particular, SWOT snapshots provide water surface slopes at a spatial resolution than can not be obtained with hydrometric stations, and which could be exploited to calibrate hydrodynamic models which, in turn, could be used for hydrometric purposes.

---

[13]`https://swot.cnes.fr`

Chapter 2

Uncertainty in and around Hydrologic Models

FIG. 2: PREDICTED CURVE
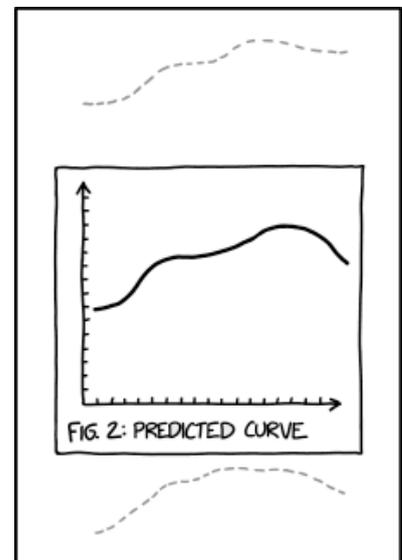
SCIENCE TIP: IF YOUR MODEL IS BAD ENOUGH, THE CONFIDENCE INTERVALS WILL FALL OUTSIDE THE PRINTABLE AREA.

Taken from: `https://www.xkcd.com/2311/`

2.1   Introduction: uncertainty in hydrologic modeling

Hydrologic models translate our understanding of the dominant processes and fluxes in the catchment into a system of mathematical equations, generally accompanied by a computer code that implements them. They can be viewed as scientific tools to formalize and evaluate competing hypotheses on the catchment functioning [Clark et al., 2011]. They are also highly

practice-oriented tools and are used in a myriad of applications: flood and drought forecasting [Cloke et al., 2013], flood frequency analysis [Paquet et al., 2013], quantifying the response of river streamflow to changes in the catchment or the climate [Vidal et al., 2016], etc. However, hydrologic models are, like any model, simplified and imperfect representations of the real world, and they are hence uncertain. Hydrologic models also use uncertain forcing and calibration data. Understanding and quantifying these various sources of uncertainty is necessary to properly evaluate hydrological functioning hypotheses and to make well-informed decisions.

2.1.1   Sources of uncertainty

Generally speaking, hydrologic modeling is affected by three main sources of uncertainty [Kuczera, 1982], as illustrated in Figure 2.1:

1. input uncertainty, typically due to sampling and measurement errors in estimates of spatial or catchment-averaged precipitation;

2. response uncertainty, in particular rating curve errors affecting streamflow time series;

3. structural uncertainty, arising from the simplified representation of hydrological processes in hydrologic models.

Moreover, most hydrologic models have parameters that are not directly measurable and must hence be estimated from the observed data: this constitutes a fourth source of uncertainty called parametric uncertainty [Kuczera, 1983, Kuczera and Parent, 1998]. Note that this decomposition of uncertainty sources is in fact quite generic and may apply beyond the case of hydrologic models [e.g. Nagel and Sudret, 2016].

Since streamflow (response) uncertainty has been treated in the previous chapter, the following paragraphs review the other three uncertainty sources. Input uncertainty arises from errors affecting the input variables of the model, typically precipitation, potential evapotranspiration and temperature. Precipitation uncertainty may be the most important input uncertainty to consider since precipitation is the main driver of streamflow response in many catchments. Precipitation over the catchment (either averaged or spatialized) is often estimated using rain gauges [Moulin et al., 2009] which are subject to various types of measurement errors, including mechanical limitations, wind effects, evaporation losses, etc. [Molini et al., 2005]. Moreover, precipitation is highly variable in both space and time: a small set of gauges may hence be poorly representative of the entire areal rain field, inducing a large sampling uncertainty [Villarini et al., 2008]. Weather radar provide a better representation of the spatial variability of rainfall and hence can reduce sampling uncertainty, but this comes at the cost of introducing new sources of uncertainty related to the radar data themselves and to their merging with rain gauges data [Severino and Alpuim, 2005]. Standard errors in estimating an areal precipitation may be of the order of 25% in unfavorable situations [Linsley and Kohler, 1988].

Even if there were no errors in the observed forcings and responses, the hydrologic model would still fail to exactly reproduce the response due to the structural errors resulting from
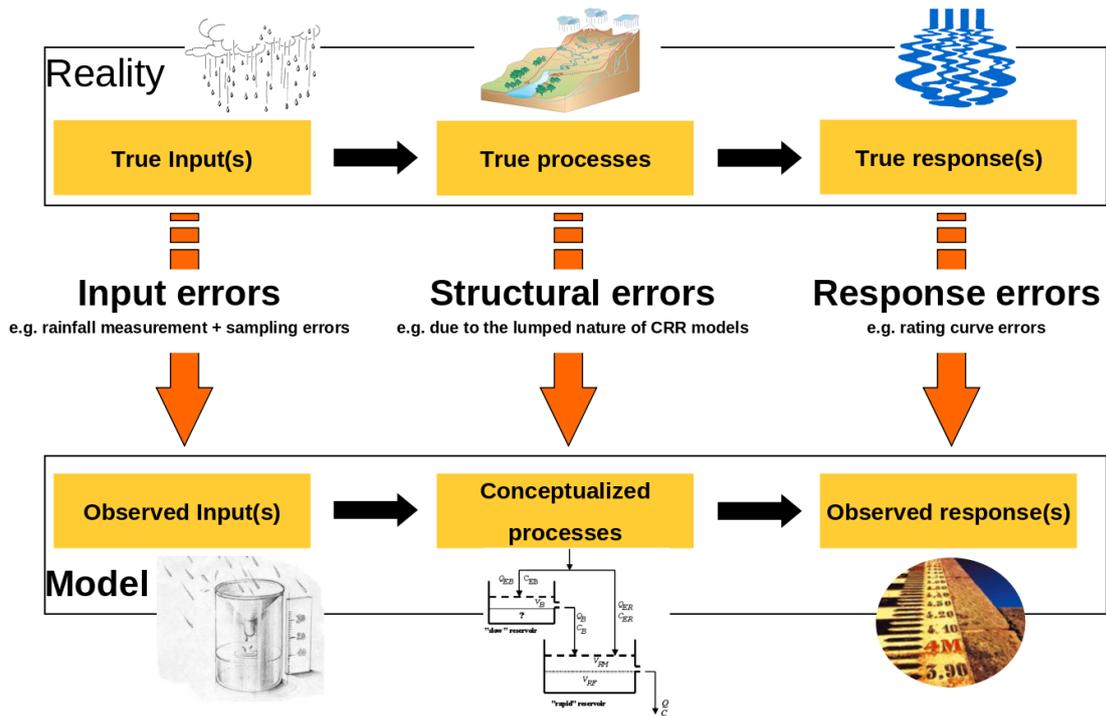
Figure 2.1: Model vs. reality: schematic representation of the three main error sources in hydrologic modeling.

modeling approximations [Beven and Binley, 1992, Beven, 2005, Kuczera et al., 2006]. Such approximations may arise from the use of conceptual stores representing complex 2-D or 3-D features of the catchment in a highly simplified way, from the flux equations that are often empirical in nature, or by ignored sub-grid variability [Andréassian et al., 2004, Ching et al., 2006]. There latter problem is particularly pronounced for lumped or semi-lumped models defined at the scale of catchments or sub-catchments: for given initial conditions and areal rainfall, these models will always produce the same response, whereas in reality streamflow may strongly depend on *where* the rain falls (saturated or unsaturated part of the catchment).

Parametric uncertainty refers to the inability to specify exact values for the model parameters. The first cause of this uncertainty is the finite length of the series used to estimate parameters (a.k.a. 'sampling uncertainty' in statistical terminology). In general, sampling uncertainty decreases with the length of calibration data [Mantovan and Todini, 2006], and may hence tend to zero when long series are available. In practice, the rate of this uncertainty decrease depends on the statistical model used for calibration, and may even be null under specific assumptions (e.g. presence of unknown biases). Some authors have argued that the declining contribution of parametric uncertainty to the total uncertainty is an undesirable property [Beven et al., 2008]: the principle of equifinality under data and structural uncertainties [Beven, 2006] suggests that many parameter sets should lead to indistinguishable model performances, even with hypothetical infinite calibration series.

Note that there is a fundamental difference between data uncertainties (input and response) and structural uncertainty. Data errors arise and exist independently from the hydrologic

model, and their properties (e.g., means and variances of rainfall and streamflow errors) can hence be estimated prior to model calibration. By contrast, structural uncertainty is an inherent feature of the hydrologic model: it depends on the model formulation, on the specific catchment it is applied to, on the spatial and temporal scale of the analysis, etc. It can therefore hardly be estimated before model calibration.

2.1.2   Methods for uncertainty quantification

Characterizing the uncertainty in hydrologic models has attracted the attention of hydrologists over many years. Numerous methods have been proposed for this purpose, including the Generalized Likelihood Uncertainty Estimation (GLUE) [Beven and Binley, 1992], instrumental-variable methods [Young, 1998], frequentist approaches [Montanari and Brath, 2004], standard Bayesian approaches [Kuczera and Parent, 1998, Krzysztofowicz, 2002, Feyen et al., 2007], Bayesian Recursive Estimation [Thiemann et al., 2001], Bayesian hierarchical models [Kavetski et al., 2006a,b, Kuczera et al., 2006, Huard and Mailhot, 2006, 2008], Bayesian model averaging [Duan et al., 2007, Marshall et al., 2007] and others. In spite of this methodological diversity, the implementation of uncertainty quantification methods still faces many challenges, both practical and theoretical [Wagener and Gupta, 2005].

It is useful to distinguish between the quantification of predictive uncertainty and its decomposition into elementary sources (input, response, structural). The former essentially entails deriving a realistic statistical model to describe the residuals between simulated and observed streamflow. Such a model typically needs to include components representing the correlated, heteroscedastic, and non-Gaussian nature of residuals [Schoups and Vrugt, 2010, McInerney et al., 2017]. Alternatively, the GLUE approach relies on performance metrics rather than explicit probabilistic models to quantify parametric uncertainty and transform it into predictive uncertainty [Beven and Binley, 1992]. However, this lumped residuals-based approach is questionable because errors affecting model calibration and prediction are often different. For instance, in a forecasting context, errors affecting the observed rainfalls used for calibration are different from errors in forecasted rainfall. Likewise, the rating curve errors affecting residuals should not be propagated in the forecast because the aim is to predict the actual future streamflow, not the observed one. The following drawbacks can also be mentioned:

1. parameters may be biased by unrecognized data errors, confounding regionalization attempts;

2. the best strategy to reduce predictive uncertainty is difficult to identify without knowing its main causes (improve the model or get better data?);

3. discriminating between competing model hypotheses is difficult because the precise causes of poor model performance are unclear.

Uncertainty decomposition methods aim at quantifying the individual contributions of input, response and structural uncertainties to the total predictive uncertainty [Moradkhani et al.,

2005a, Kuczera et al., 2006, Huard and Mailhot, 2008]. Achieving this decomposition poses significant challenges and requires developing dedicated statistical techniques. Most importantly, decomposition approaches require specifying error models for each source of uncertainty. Developing realistic error models and identifying the prior knowledge necessary to achieve a well-posed inference are significant challenges, as reviewed next.

Regarding input uncertainty, Kavetski et al. [2006a,b] proposed a simple model based on rainfall multipliers to account for the effect of rainfall errors on model calibration. Since the spatialized or catchment-averaged precipitation used as input of the hydrologic model is often based on some geostatistical analysis of rain gauge and/or radar data, prior information on the amount of rainfall uncertainty can be derived from geostatistical approaches such as kriging [Kuczera and Williams, 1992, Goovaerts, 2000, Leblois and Creutin, 2013, Delrieu et al., 2014] and conditional simulation [Onibon et al., 2004, Clark and Slater, 2006, Gotzinger and Bardossy, 2008, Vischel et al., 2009], or by analyzing dense rain gauge networks [Willems, 2001, Villarini et al., 2008]. Regarding response uncertainty, the streamflow data used to calibrate the hydrologic model generally comes from a hydrometric station and is therefore affected by rating curves errors: models for describing such errors have been extensively discussed and described in Chapter 1 and are therefore not repeated here.

Models for structural errors are arguably the most challenging: there is currently no agreement on how best to describe them and several options are available. They can be roughly classified as follows:

1. Exogenous treatment: structural errors are modeled as an (usually additive) streamflow error term [Huard and Mailhot, 2008]. The use of independent and identically distributed (*iid*) Gaussian errors represents the simplest case, but more complex models are needed in general to account for autocorrelation, heteroscedasticity etc. This error term can be combined with a second error term representing the uncertainty in streamflow data (this is identical to the approach used in BaRatin, see Chapter 1 section 1.2.2)

2. Endogenous treatment: structural errors are represented by a stochastic component within the hydrologic model. This can be achieved by perturbating internal model states [Bras and Rodriguez-Iturbe, 1984, Moradkhani et al., 2005b], by making some model parameters stochastic [Kuczera et al., 2006, Reichert and Mieleitner, 2009] or by formulating internal relationships within the model as a joint probability density function [Bulygina and Gupta, 2009].

In addition to the approaches discussed above, alternatives that are not based on formal probabilistic models have also been proposed [Beven and Binley, 1992, Jacquin and Shamseldin, 2007]. Lastly, Bayesian Model Averaging has also been used to generate structural uncertainty from a collection of models [Duan et al., 2007, Marshall et al., 2007, Schöniger et al., 2014, Valdez et al., 2022]. This complements, but does not replace, the characterization of the structural uncertainty affecting one given model.
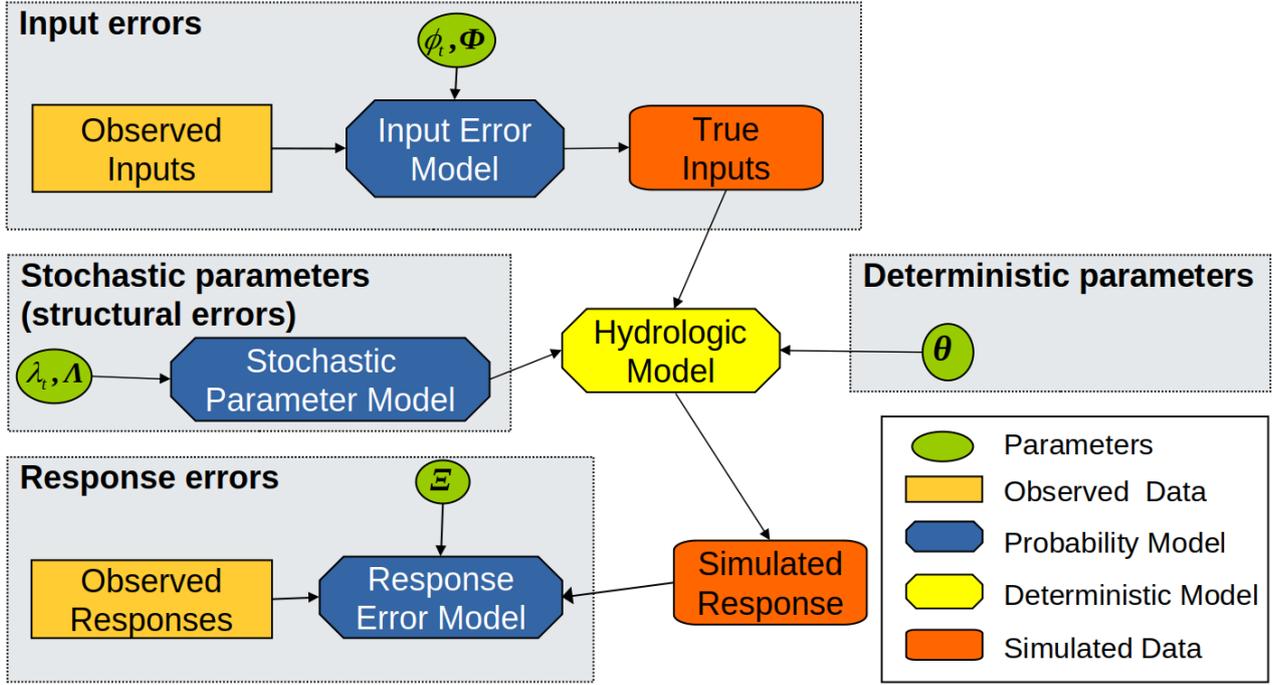
Figure 2.2: Schematic of the Bayesian Total Error Analysis (BATEA) framework.

2.2   Developing the Bayesian Total Error Analysis framework

My work revolving around uncertainties in hydrologic models started during an 18-month post-doctoral position at the University of Newcastle (Australia) in the research team of George Kuczera, who developed the Bayesian Total Error Analysis (BATEA) framework.

2.2.1   Overview of the Bayesian Total Error Analysis framework

The BATEA framework was developed to explicitly represent each source of uncertainty affecting calibration and prediction of hydrologic models [Kavetski et al., 2006a,b, Kuczera et al., 2006]. In a nutshell, it transforms the schematic of Figure 2.1 into a probabilistic hierarchical model, with submodels for describing each type of error (Figure 2.2). BATEA therefore allows, even requires, modelers to explicitly hypothesize, infer and evaluate assumptions regarding each source of uncertainty, and generates model predictions accounting for all uncertainties included in the analysis.

More formally, BATEA assembles the components described below. The hydrologic model can be viewed as a function $\mathcal{H}$ that simulates the streamflow response $\widehat{Q}_t$ given parameters $\boldsymbol{\theta}$ and precipitation input $\boldsymbol{R}_{1:t}$:

$$\widehat{Q}_t = \mathcal{H}\left(\boldsymbol{R}_{1:t}, \boldsymbol{\theta}\right) \tag{2.1}$$

The time subscript reminds that for simulating responses at the tth time step, the whole history of precipitation up to time t is needed, reflecting the dynamical nature of an hydrologic model. Moreover, the dependence on initial state conditions S_0 and on input variables other than precipitation (potential evaporation, temperature) is made implicit for simplifying purposes. Finally, the notation in equation (2.1) is adapted to a lumped model rather than a distributed one.

The input error component is a probabilistic model $p(R_t | \widetilde{R}_t, \boldsymbol{\Phi})$ relating true (R_t) and observed $(\widetilde{R}_t)$ precipitation, with $\boldsymbol{\Phi}$ denoting the parameters of this model. Kavetski et al. [2006a,b] proposed using rainfall multipliers to derive a possible rainfall error model as follows:

$$R_t = \widetilde{R}_t \exp(\phi_t); \; \phi_t \sim \mathcal{N}(\mu_R, \sigma_R)$$
$$\Longleftrightarrow R_t \sim \mathcal{LN}(\mu_R + \log(\widetilde{R}_t), \sigma_R) \tag{2.2}$$

Likewise, the response error component is a probabilistic model $p(\widetilde{Q}_t | \widehat{Q}_t, \boldsymbol{\Xi})$ relating observed $(\widetilde{Q}_t)$ and simulated $(\widehat{Q}_t)$ streamflow. The following model is the simplest possibility:

$$\widetilde{Q}_t = \widehat{Q}_t + \varepsilon_t; \; \varepsilon_t \sim \mathcal{N}(0, \sigma_Q)$$
$$\Longleftrightarrow \widetilde{Q}_t \sim \mathcal{N}(\widehat{Q}_t, \sigma_Q) \tag{2.3}$$

An original aspect of the BATEA framework is that it allows describing structural errors by means of stochastic parameters. To achieve this, at least one component of the parameter vector $\boldsymbol{\theta}$ is allowed to vary in time according to some distribution [Kuczera et al., 2006]. For instance, if a parameter λ corresponds to the maximum capacity of a model's store, one may assume that it changes in time according to:

$$\lambda_t \sim \mathcal{LN}(\mu_\Lambda, \sigma_\Lambda) \tag{2.4}$$

Note that equations (2.2) to (2.4) are just examples of possible error models. In particular, the response error model (2.3) might be further developed to make σ_Q increase with streamflow $\widehat{Q}_t$, to include autocorrelation or variable transformation [Engeland et al., 2010], or to add a second error term representing streamflow data errors (as done in BaRatin, see Chapter 1 section 1.2.2). Likewise, the stochastic parameter model (2.4) might describe an autocorrelated stochastic process as done in Reichert and Mieleitner [2009].

Once these error models are defined, unknown quantities such as the parameters $\boldsymbol{\theta}$ of the hydrologic model and the parameters $(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Xi})$ of the error models need to be inferred. Since running the hydrologic model requires knowing the values taken by the stochastic parameter λ_t and the true rainfall R_t (or equivalently, the rainfall multiplier ϕ_t, see Figure 2.2), these two time series are treated as latent variables that also need to be inferred. This results in a Bayesian hierarchical model which itself leads to the following posterior distribution, whose
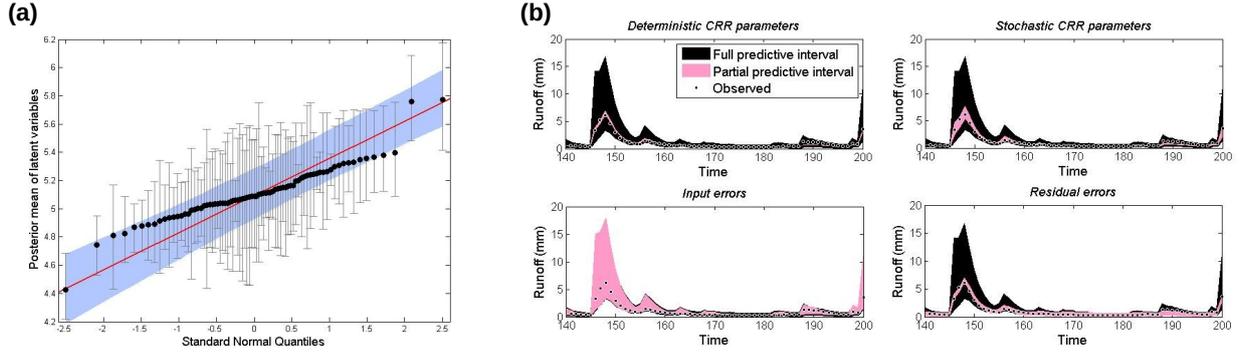
Figure 2.3: Examples of post-processing tools. (a) Diagnostic qq-plot to test the normality of a latent variable; (b) contributions of individual sources of uncertainty to the total predictive uncertainty.

factorization closely follows the schematic shown in Figure 2.2:

$$p\left(\boldsymbol{R}, \boldsymbol{\Phi}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\Lambda}, \boldsymbol{\Xi} | \widetilde{\boldsymbol{Q}}, \widetilde{\boldsymbol{R}}\right) \propto$$

$$\underbrace{p\left(\widetilde{\boldsymbol{Q}} | \underbrace{\boldsymbol{R}, \boldsymbol{\theta}, \boldsymbol{\lambda}}_{\text{allows computing } \widehat{\boldsymbol{Q}}}, \boldsymbol{\Xi}\right)}_{\text{Response error model}} \times \underbrace{p\left(\boldsymbol{R} | \widetilde{\boldsymbol{R}}, \boldsymbol{\Phi}\right)}_{\text{Input error model}} \times \underbrace{p\left(\boldsymbol{\lambda} | \boldsymbol{\Lambda}\right)}_{\text{Stochastic parameter model}} \times \underbrace{p\left(\boldsymbol{\Phi}, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\Xi}\right)}_{\text{Prior}} \quad (2.5)$$

2.2.2   Specific developments

My first contribution to the development of BATEA was its implementation in the form of a flexible and efficient code. More specific objectives were as follows [Renard et al., 2007]:

1. allowing rainfall errors and stochastic parameters to vary at the scale of storm epochs rather than at each time step;

2. implementing pre-processing sensitivity analyses to determine which parameter(s) of the hydrologic model should be made stochastic;

3. designing a code architecture that would allow building the BATEA model by simply selecting error models and distributions from existing plug-in catalogs;

4. designing and implementing an efficient and robust MCMC strategy;

5. implementing post-processing tools to visualize posterior estimates, implement diagnostics, perform predictions etc. (Figure 2.3).

A particular difficulty associated with BATEA is that, like many hierarchical models, it leads to a high-dimensional inference. In principle, it is possible to drastically reduce dimensionality by integrating out the latent variables, rather than conditioning on them. However, this merely

replaces a high-dimensional sampling problem by a high-dimensional integration problem which is, in general, not easier to solve. We evaluated a third alternative proposed by Ajami et al. [2007], based on resampling a new realization of the latent variables at each model evaluation, but we concluded that this approach did not allow solving the dimensionality issue [Renard et al., 2009].

A dedicated MCMC algorithm was then implemented to explore the posterior distribution of equation (2.5) [Kuczera et al., 2010a]. The standard Metropolis algorithm is not well suited to such a high-dimensional posterior because it attempts to update the whole parameter vector at once, which makes it fast but very difficult to tune. Instead, 'Metropolis-within-Gibbs' samplers are commonly used in Bayesian hierarchical inference: they update the parameter vector one component at a time and are hence much easier to tune because the jump distributions are uni-dimensional. Moreover, the increase in computational cost is limited because in most hierarchical models, a latent variable only affects a small subset of the data, so that updating a single latent variable does not involve recomputing the whole likelihood - only a small part of it. Unfortunately, this does not apply to dynamical models such as hydrologic models: modifying, for instance, a rainfall multiplier ϕ_t will affect all simulated streamflow located in the future with respect to t. Consequently, a Metropolis-within-Gibbs sampler accounting for this memory involves many long re-runs of the hydrologic model, leading to a computational cost that grows quadratically, and hence prohibitively, with the length N of calibration series. We proposed a workaround that exploits the fact that this memory is decaying, so that only re-running the model on a few future time steps provides an acceptable approximation, which can be controlled by a user-defined tolerance. This 'limited-memory' trick is able to bring the computational cost down to being almost linear with N, with negligible differences in terms of posterior estimates compared to a 'full-memory' sampler (Figure 2.4).

Case studies were also performed in order to better understand how the model used to describe data errors could influence parameter estimates. In Thyer et al. [2009], we compared three competing hypotheses on the temporal structure of input errors: (1) daily rainfall multipliers; (2) identical rainfall multipliers for time steps within the same storm event; (3) a single rainfall multiplier for the whole period. The second is more parsimonious and less dependent on the chosen time step than the first. The third corresponds to a constant bias and is probably too simplistic - the amount of rainfall error strongly depends on *where* the rain falls, and hence varies from storm to storm. In addition, a response error model arising from a rating curve analysis was also specified. Results showed that calibration schemes ignoring data uncertainty or treating it too simplistically led to precise but very unstable parameter estimates (Figure 2.5): very different values were obtained when changing the rainfall calibration data, suggesting that the parameters were 'fitting to rainfall errors' to some extent. By contrast, BATEA schemes that recognized rainfall uncertainty provided consistent, albeit more uncertain, parameter estimate. This can roughly be described as being 'vaguely correct rather than precisely wrong'. Such uncertain but stable parameter estimates can be used more confidently in regionalization studies or when the model is extrapolated or forced with forecasted or pro-
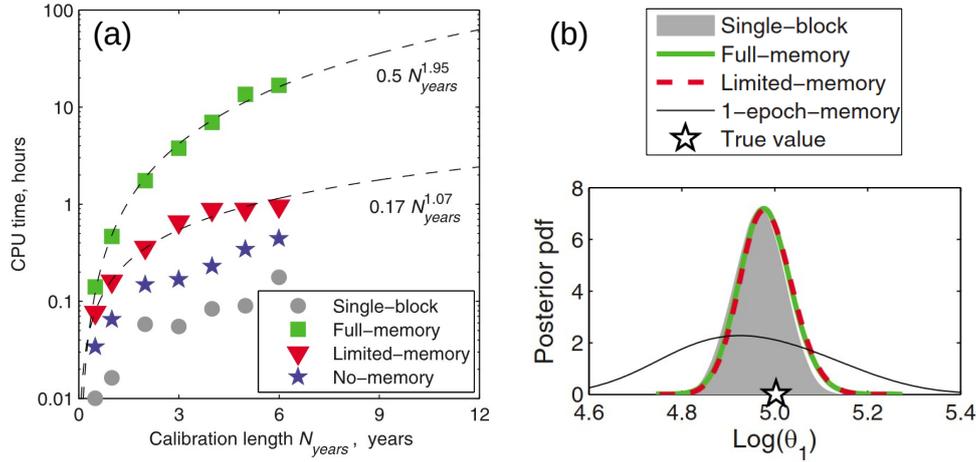
Figure 2.4: Evaluation of the 'limited memory' trick to speed up MCMC sampling. (a) CPU time to generate 10,000 MCMC samples from BATEA posterior, as a function of the calibration data length (2.0 GHz laptop CPU with 1 GB of RAM); (b) posterior distributions of the first hydrologic model parameter estimated using different MCMC samplers(100,000 samples). Note that differences between the limited-memory and the full-memory samplers are barely noticeable, while a 1-epoch memory leads to a markedly different posterior distribution. Modified from Kuczera et al. [2010a].

jected rainfall. Despite these encouraging results, this paper ended with a note of caution: we found that the estimated amount of rainfall errors was unrealistically large, suggesting that the inferred rainfall errors were compensating for the simplistic treatment of structural errors. Understanding such interactions between data and structural errors is important if a meaningful uncertainty decomposition is to be achieved.

2.3   Interactions between data and structural uncertainties

2.3.1   Diagnosing the non-identifiability of input and structural errors

The first applications of the BATEA approach [Kavetski et al., 2006a,b, Thyer et al., 2009] focused on the treatment of rainfall errors and lacked a separate characterization of structural errors. The work by Kuczera et al. [2006] was the first to implement the full BATEA model illustrated in Figure 2.2, including the use of stochastic parameters to characterize structural errors. However, this study was based on maximum-posterior estimates of the parameters and did not implement a full Bayesian treatment of the posterior distribution, including its exploration through MCMC sampling. Attempts to achieve this latter approach proved to be challenging: MCMC sampling systematically failed to converge when both input errors and stochastic parameters were included. We investigated the reasons for this behavior through several synthetic and real-life case studies [Renard et al., 2010, Kuczera et al., 2010b], and
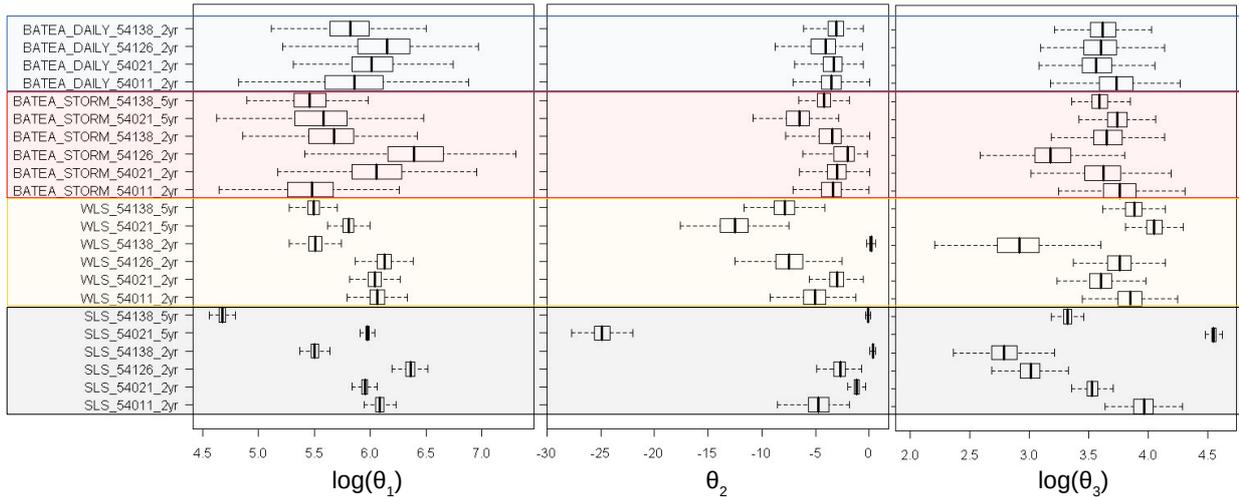
Figure 2.5: Boxplots of marginal posterior distributions of three GR4J parameters, depending on the calibration setup. Labels on the y-axis have the form (calibration scheme)_(rain gauge used as input)_(length of calibration period). SLS = 'Standard Least Square', *iid* Gaussian response error model as given in equation (2.3); WLS = 'Weighted Least Square', heteroscedastic Gaussian response error model arising from a rating curve analysis; 'BATEA_STORM': WLS + stochastic rainfall errors acting at the scale of storm epochs; 'BATEA_DAILY': WLS + stochastic daily rainfall errors.

concluded that convergence failures had nothing to do with the efficiency of the MCMC samplers we used. Instead, they were due to the intrinsic non-identifiability of input and structural errors, which led to a ill-posed problem.

A very simple example is useful to better understand non-identifiability and its symptoms. Consider the model $Y \sim \mathcal{N}(\theta_1 + \theta_2, 1)$, and suppose that a sample of 20 values is observed, with a mean approximately equal to 1. This sample provides information on the mean assumed by the model, i.e. on $\theta_1 + \theta_2$, but it does not say anything about individual θ_1 and θ_2 values: $(\theta_1 = 0, \theta_2 = 1)$ is indistinguishable from $(\theta_1 = -99, \theta_2 = 100)$. Figure 2.6(a) shows the likelihood function, which is proportional to the posterior pdf when flat priors are used: it is characterized by an infinite ridge elongated along the line $\theta_1 + \theta_2 = 1$. An attempt at applying a MCMC sampler to this (improper) pdf leads to non-convergence (Figure 2.6(b)): the chains for θ_1 and θ_2 diverge as they may walk indefinitely along the ridge with no information to favor any particular region. This situation corresponds to a model that is non-identifiable from the data, leading to a ill-posed problem [see Renard et al., 2010, for formal definitions].

Although extremely simple, the example above is of interest because it closely corresponds to the symptoms we encountered in both synthetic and real-life case studies [Renard et al., 2010, Kuczera et al., 2010b]. As an illustration, Figure 2.7 highlights near-perfect negative correlations between latent variables for rainfall multipliers and stochastic parameters: each correlated pair has a bivariate posterior pdf similar to the improper one shown in Figure 2.6(a). MCMC sampling in such a case never converges. So in a nutshell, a key conclusion of these studies was that decomposing the total uncertainty into its components is impossible based on
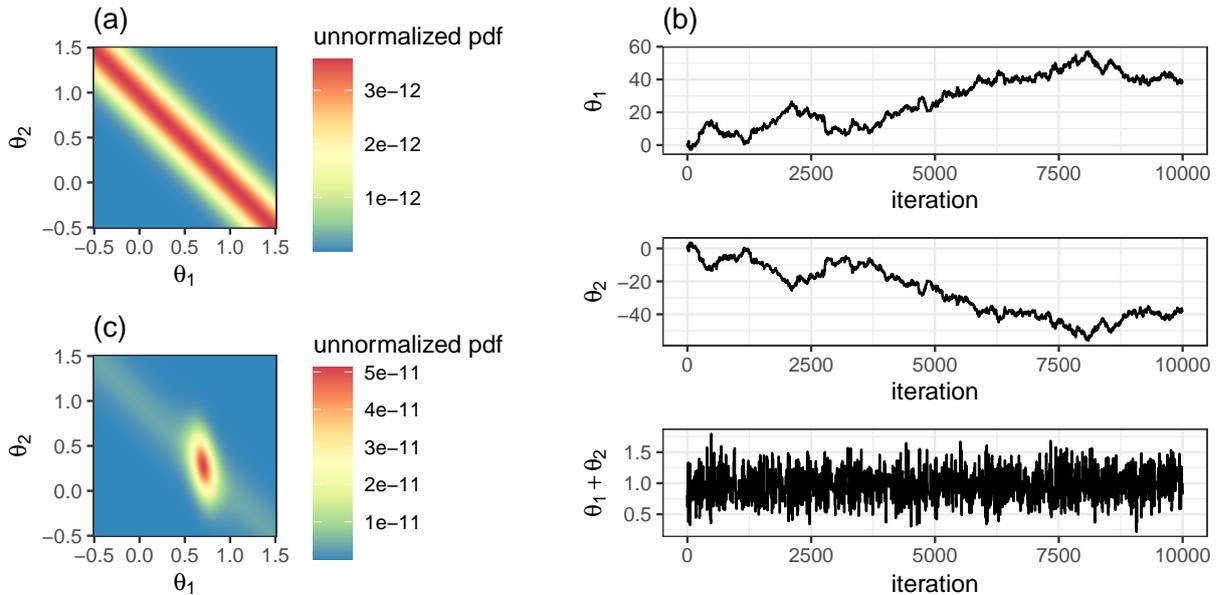
Figure 2.6: Symptoms and cure for non-identifiability. (a) Posterior pdf associated with the model $Y \sim \mathcal{N}(\theta_1 + \theta_2, 1)$, using $n = 20$ observations generated from a $\mathcal{N}(1, 1)$ distribution and flat priors. (b) MCMC sampling from the posterior in (a); (c) same as (a), but using an informative prior distribution $\theta_1 \sim \mathcal{N}(0.7, 0.15)$.

observed streamflow data alone: the discrepancy between observed and simulated streamflow only provides information about total errors.

What are the available options to overcome non-identifiability and ill-posedness? Using longer series does not solve the problem: the 'infinite ridge' aspect of Figure 2.6(a) remains irrespective of sample size. Improving the MCMC sampler, although a legitimate first reaction to non-convergence, is pointless: no sampler will be able to sample properly from a posterior like the one shown in Figure 2.6(a). In fact, we even found that MCMC non-convergence was a systematic symptom so that it should rather be used as a clear warning to go back to the model and understand the reason for ill-posedness. This leaves the following options:

1. Reparameterizing: in the simple non-identifiability example, summing the two non-convergent chains for θ_1 and θ_2 leads to a perfectly convergent chain for $\theta_1 + \theta_2$ (Figure 2.6(b)). This means that using a model parameterized in terms of $\eta = \theta_1 + \theta_2$ solves non-identifiability, but at the cost of abandoning the decomposition objective, which is not a satisfying solution.

2. Using prior information: in the simple non-identifiability example, using a prior information on only one component leads to a well-posed posterior (Figure 2.6(c)). We similarly found in case studies [Renard et al., 2010, Kuczera et al., 2010b] that providing precise priors on the properties of rainfall errors (typically their mean and variance) led to a much better-behaved posterior. Note, however, that this does not make rainfall uncertainty identifiable *from the data*: the information rather comes from the prior, and its
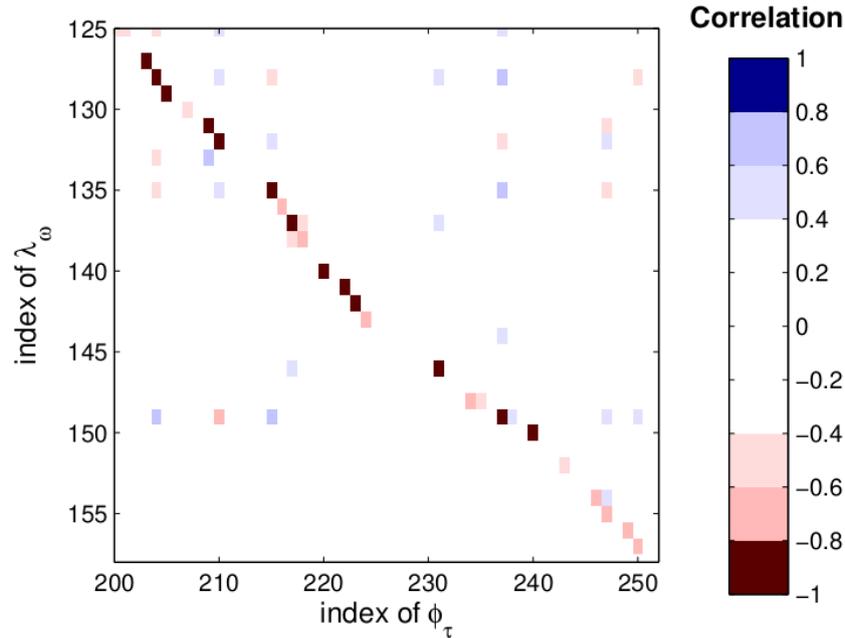
Figure 2.7: Illustration of the reason for the non-identifiability of input and structural errors together: the figure shows posterior correlations between latent variables for rainfall multipliers and stochastic parameters in a full BATEA model as schematized in Figure 2.2. Dark cells near the diagonal highlight near-perfect correlations between rainfall multipliers and stochastic parameters acting on the same time steps.

specification is hence of paramount importance.

3. Changing the treatment of structural errors: we tried to represent them through a dedicated error term in the response error model (in addition to the streamflow data error), rather than with stochastic parameters [Renard et al., 2010]. The resulting posterior was indeed well-posed, but we found that rainfall uncertainty was strongly overestimated, unless constrained by precise priors. In other words, what was labeled as 'rainfall uncertainty' was in fact a mixture of rainfall and structural uncertainties, as already suspected by Thyer et al. [2009].

Overall, the results presented in this section lead to a conclusion that '*there is no free lunch in hydrology*' [Kuczera et al., 2010b]: providing precise priors on data uncertainty is an absolute necessity if a meaningful decomposition of the total predictive uncertainty is to be achieved.

2.3.2 Ways forward

The uncertainty affecting the spatialized or catchment-averaged precipitation used as input of the hydrologic model can be quantified by means of conditional simulations, as explored in Renard et al. [2011] thanks to the rainfall simulator SAMPO developed by Etienne Leblois [Leblois and Creutin, 2013]. Figure 2.8(a) illustrates the principle of this technique: after a geostatistical model has been specified and estimated, several realizations of the rainfall fields are generated. These realizations have the same values at rain gauges, but different values
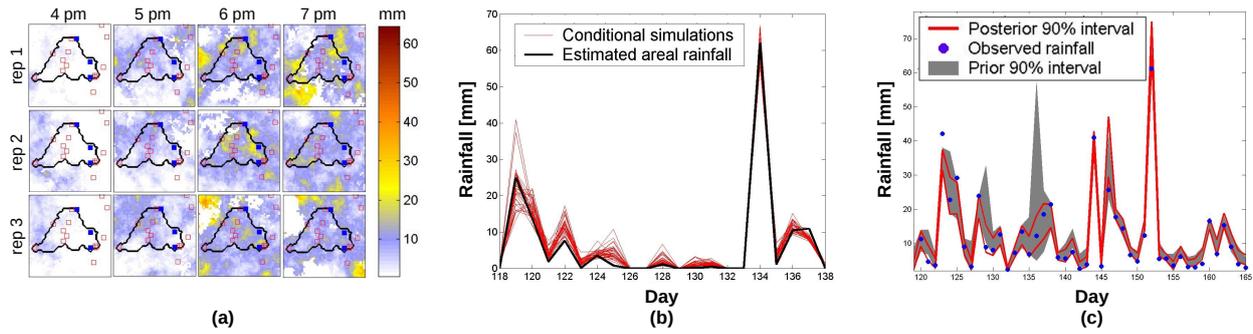
Figure 2.8: Use of rainfall conditional simulations in hydrologic modeling. (a) Three conditional simulations on four consecutive hourly time steps, with identical values at rain gauges but different values in between. (b) Several realizations of the catchment-averaged daily precipitation time series. (c) Comparison between prior rainfall uncertainty (estimated from conditional simulations before model calibration) and posterior rainfall uncertainty (estimated as part of model calibration with BATEA). Note that for most time steps, the prior and posterior intervals are similar.

in between rain gauges, thus representing the sampling uncertainty affecting spatial rainfall. Each realization can be aggregated in space and time to provide one possible realization of the rainfall series used as input of the hydrologic model (Figure 2.8(b)). The multiple realizations can finally be used to specify a prior distribution of rainfall errors. Note that conditional simulations hence provide information at each time step, not merely on the mean and variance of rainfall errors.

A case study based on the Yzeron catchment (France) showed that the inclusion of this information was sufficient to overcome the ill-posedness problems described in the previous section: simultaneous estimation of rainfall multipliers and stochastic parameters (representing structural errors) in BATEA was achieved with no MCMC convergence issue. This indicates that the prior information provided by conditional simulations exerts a strong influence on the inference (Figure 2.8(c)). By contrast, estimating the same 'full' BATEA model without prior rainfall information led to non-convergent MCMC sampling. The simultaneous estimation of rainfall multipliers and stochastic parameters allowed isolating the contributions of input and structural errors to the total predictive uncertainty (Figure 2.9): in this particular case study, it appeared to be dominated by structural errors. We also found that the 'full' BATEA model yielded more reliable streamflow predictions than approaches that ignore or lump different sources of uncertainty.

The work described in the previous paragraph represented an important step forward in the inclusion of rainfall uncertainty in model calibration. However, the inclusion of streamflow uncertainty still relied on an overly simplistic response error model that did not recognize the partly systematic nature of rating curve errors. This shortcoming was an important motivation to the development of the BaRatin method described in Chapter 1. Moreover, a collaboration with Anna Sikorska, from the University of Zurich, allowed exploring the impact of parametric and structural rating curve errors on the calibration of hydrologic models [Sikorska and Renard,
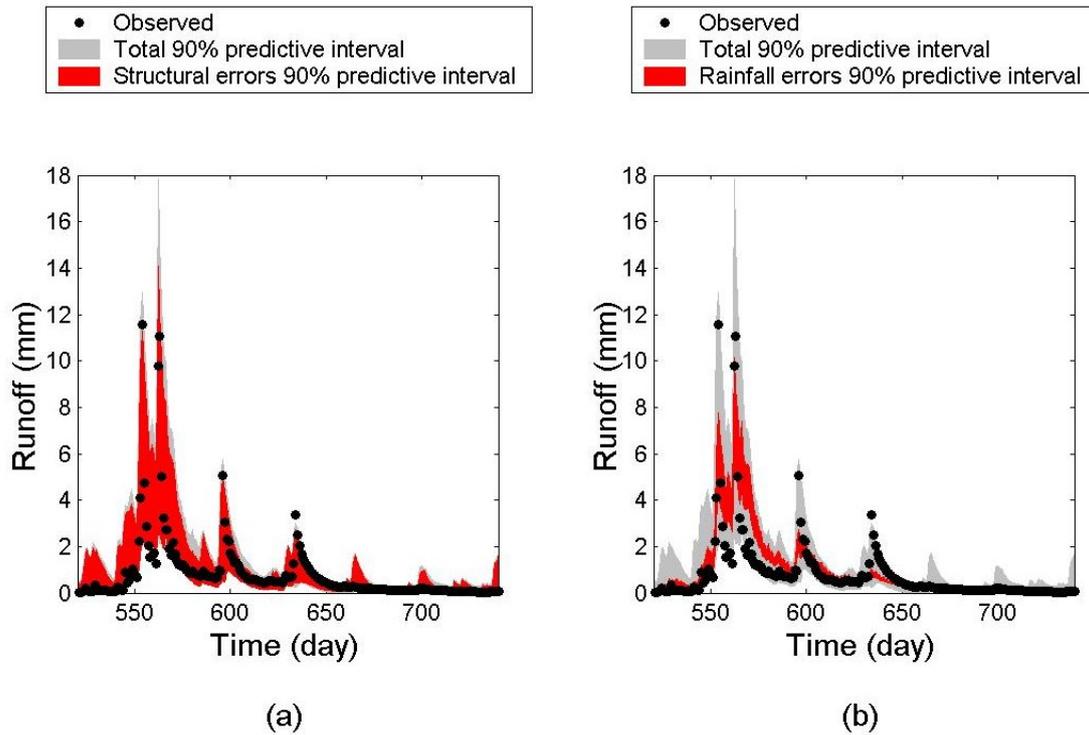
Figure 2.9: Contributions of structural (a) and rainfall (b) uncertainties to the total predictive uncertainty.

2017]. The basic idea behind this work was to couple the hydrologic model with the inverse rating curve, leading to a 'rainfall-runoff-stage' model (Figure 2.10(a)). This coupled model is parameterized with both hydrologic and rating curve parameters, and accounts for both hydrologic and rating curve structural errors: this allows acknowledging the various sources of uncertainty (coming from hydrologic and rating curve models) and assessing their contribution to the total predictive uncertainty. Calibration is made in two steps: the rating curve is first calibrated on gaugings with the BaRatin approach of Chapter 1. The rainfall-runoff-stage model is then calibrated on the observed stage series, with the prior on rating curve parameters ($\boldsymbol{\theta}_{RC}$ and $\boldsymbol{\gamma}$ in Figure 2.10(a)) being defined as the BaRatin posterior of step 1.

A case study in the upper Ardèche catchment in France indicated that the structural uncertainty of the hydrologic model largely dominated other uncertainty sources (Figure 2.10(b)). It was also found that the inclusion of rating curve parametric uncertainty led to marked changes in some parameters of the hydrologic model compared with calibration schemes ignoring it. Some rating curve parameters were also markedly modified by the calibration of the rainfall-runoff-stage model to the observed stage series (step 2). The extent to which the rating curve was modified is hardly defensible (see inset in Figure 2.10(b)) and does not correspond to a meaningful improvement of the rating curve in our eyes. Instead, it is a sign that the error models we used do not convincingly weight the information brought by the gaugings and the stage time series. A first direction to improve this is to avoid re-estimating the rating curve during calibration stage 2: this can be achieved by means of a Monte-Carlo propagation approach, where the hydrologic model is estimated many times for many possible realizations of
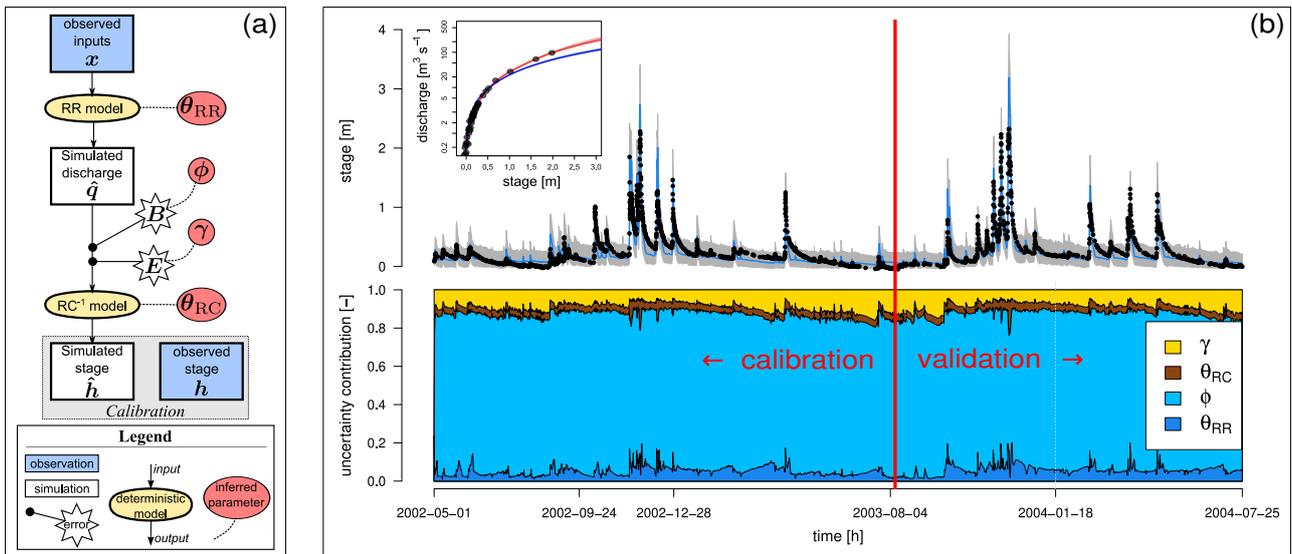
Figure 2.10: Uncertainty decomposition in a rainfall-runoff-stage model. (a) Coupling between an hydrologic rainfall-runoff (RR) model and an inverse rating curve (RC) model. B represent the structural errors of the RR model, E those of the RC model. (b) Total uncertainty in the predicted stage, and decomposition into contributing sources. The top left inset shows the rating curve estimated with gaugings (red), and the rating curve after calibration to the observed stage time series (blue). Modified from Sikorska and Renard [2017].

the streamflow series (i.e. the BaRatin 'spaghetti' of section 1.2.3). The second approach is to improve the error models. In particular, the structural error models of both the hydrologic model and the rating curve assume independant Gaussian errors, which is probably too simplistic (see discussion in section 2.5). Moreover, rainfall errors were completely ignored in this case study, and this may have played a role as well.

Finally, a collaboration during the PhD work of David Wright [2017], from the University of Adelaide, offered the opportunity to study and develop influence metrics that allow identifying data points that have a disproportionate impact on model parameters, performance and/or predictions. Such metrics are widely used in regression, and we implemented some modifications to make them applicable to nonlinear and dynamical hydrologic models. Moreover, further modifications were also required to make them applicable to general objective functions derived from posterior distributions, which are characterized by possible heteroscedasticity, autocorrelation, non-normality and the use of prior information. The resulting Generalised Cook's Distance [Wright et al., 2019] was found to be a computationally cheap and reliable tool to identify influential data in a wide variety of hydrological and environmental modeling problems. This identification is useful not only to question the reliability of the data, but also to critically evaluate some key assumptions made in the hydrologic and probabilistic models.

2.4 Operational tools and applications

Given that the decomposition of predictive uncertainty still faces several unresolved challenges, operational applications of the work described in this chapter are very few. The Australian Bureau of Meteorology uses an operational version of BATEA for seasonal streamflow forecast[1] [Tuteja et al., 2011], but it is a version distinct from the code described in section 2.2.2. Moreover, the operational BaM tool described in the previous chapter (see section 1.4) incorporates code libraries and general ideas that were developed as part of the BATEA work described in this chapter. BaM can hence be applied to hydrologic models and already includes the GR4J model [Perrin et al., 2003]. Finally, we released as part of the BATEA development the RFortran library[2] to call R from FORTRAN code [Thyer et al., 2011], but development and support have been stopped due to issues with the underlying COM[3] technology.

2.5 Conclusion and perspectives

Hydrologic models are widely used for a variety of purposes, both scientific (understanding the dominant processes at play in the catchment) and operational (flood and low flow forecasting, water resources management, etc.). However, several sources of uncertainty affect their estimation, testing and predictions: quantifying and accounting for these uncertainties is an important research endeavor. This chapter described my contribution to the development of a Bayesian Total Error Analysis approach. An important conclusion from this work is that the rainfall–runoff record on its own is insufficient to decompose the total predictive uncertainty into its constitutive sources. Stated in a more optimistic way, this means that the following ingredients need to be gathered to achieve a meaningful decomposition of uncertainty:

1. An estimation of data uncertainties prior to model calibration (they exist independently of the hydrologic model!);

2. Realistic error models describing data errors, both input (precipitation) and output (streamflow) data;

3. Realistic error models describing structural errors.

The work presented in this chapter allowed to move forward on each of these three points, but without ticking all three boxes together so far. Regarding the first point, we explored specific tools to quantify data uncertainties: conditional simulations for rainfall uncertainty and BaRatin for streamflow uncertainty. Both are mature approaches and provide viable solutions to the estimation of data uncertainty, but they are certainly not the only ones and alternative approaches exist and could be further evaluated. Regarding the second point, the description

---

[1] http://www.bom.gov.au/water/about/publications/document/dynamic_seasonal_streamflow_forecasting.pdf

[2] https://code.google.com/archive/p/rfortran/

[3] https://en.wikipedia.org/wiki/Component_Object_Model

of rainfall errors by means of a latent variable representing rainfall multipliers was found to be a viable approach as long as it is associated with informative priors (typically derived from conditional simulations). The search for an adequate streamflow error model in BATEA has been less conclusive so far. Early applications of BATEA made an unrealistic assumption of independent streamflow errors, which disregards the systematic errors induced by the rating curve. The approach of calibrating the hydrologic model in stage space described in section 2.3.2 precisely aimed at addressing this shortcoming, and it certainly holds some promises. However, the case study we performed also raised issues that need to be addressed, in particular regarding the weighting of information brought by a few dozens of gaugings vs. a stage time series that may comprise hundreds or thousands of values.

The biggest challenge to be addressed in my opinion is the derivation of realistic models for structural errors. With BATEA we mostly focused on an internal representation by means of stochastic parameters. We demonstrated that this is a viable approach as long as informative priors on rainfall errors are available - otherwise the problem is ill-posed. Alternative internal representations of structural errors could be evaluated, in particular the use of state (rather than parameter) perturbations, as done in some data assimilation approaches [Moradkhani et al., 2005a].

The external representation of structural errors, by means of a dedicated error term in the response error model, should also be further appraised. The challenge in this case is to derive a probabilistic model that properly describes the complex nature of structural errors, in particular their heteroscedasticity, autocorrelation, non-normality [Schoups and Vrugt, 2010] and partly systematic nature. The latter point is due to the fact that structural errors arise from model deficits that will systematically manifest themselves for given inputs and initial conditions: this leads structural errors to have a non-zero mean (a.k.a. a conditional bias) for given inputs and initial conditions. Developing error models that explicitly account for such a conditional bias is an important avenue for improvement in my opinion.

Finally, an important research perspective is the adaptation of the BATEA approach described here for distributed or multi-scale hydrologic models [Braud et al., 2010, 2014], or even other types of distributed models such as hydraulic models [Roux and Dartus, 2005, 2008]. This raises many questions: how to spatialize model parameters? if a parameter is to be treated as stochastic, how to induce stochasticity in a whole spatial field? how to regionalize the properties of structural errors in order to quantify uncertainty everywhere, not just where streamflow calibration series are available? Beyond streamflow series, what other types of data can be used to constrain the estimation problem? How to weight these different types of data that may strongly differ in terms of sparsity, space-time resolution, quality and even nature?

Chapter 3

Hydrologic variability

HOW TO DETECT A CHANGE IN THE SLOPE OF YOUR DATA

NOVICE METHOD:

$$Y = \begin{cases} m_1 x + b_1 & : x < n \\ m_2 x + b_2 & : x \geq n \end{cases}$$

$r = N.M$

$y_r = y - mx + b$

$\sigma =$

DO A BUNCH OF STATISTICS

EXPERT METHOD:

HEY LOOK, IT BENDS HERE

TIP THE GRAPH SIDEWAYS

Taken from: `https://xkcd.com/2701`
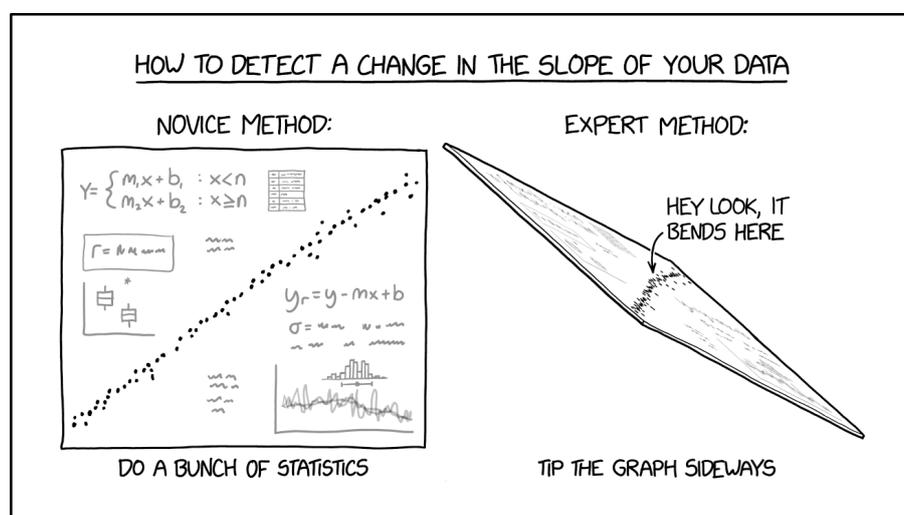
3.1 Introduction: hydrologic variability and probabilistic models

The management of water resources and risks requires a good knowledge of the temporal variability of hydrologic regimes, especially in the extreme domain. As an illustration, a dam spillway is designed so that it can pass a T-year flood event without overtopping: this requires estimating a quantile from the interannual distribution of streamflow extremes. In a climate change context, the capacity of the spillway may need to be modified if future flood projections suggest significant changes [Hirabayashi et al., 2013, Arnell and Gosling, 2014, Dankers et al., 2014, Li et al., 2022]: this involves evaluating the time-varying distribution of floods. Moreover, dam operation can benefit from seasonal streamflow forecasts (e.g. keeping the dam below its level setpoint if higher-than-usual flows are forecasted): such forecasts arise from the conditional

distribution of streamflow given some predictor [e.g. Sea Surface Temperature (SST), Lima and Lall, 2010a,c]. Many applications also require understanding the spatial variability of hydrologic extremes. For instance, large hydroelectricity companies, national water agencies, insurance companies or international humanitarian institutions are interested in the spatial aspects of flood and drought hazards [Wilhelmi and Wilhite, 2002, Uhlemann et al., 2010, Braman et al., 2013, Ward et al., 2015, Coughlan de Perez et al., 2016]. Finally, some applications require the joint description of several variables. For instance, renewable energy production depends on the joint variability of wind, solar radiation and river streamflow [Engeland et al., 2017], while the wildfire hazard results from drought, heat wave and high wind hazards [Barbero et al., 2014, Sharples et al., 2016].

The situations described above are quite diversified but they share one common aspect: they all require a probabilistic model to describe the distribution of the target hydrologic variable(s). This distribution may be as simple as the interannual distribution at one given site, but it may also vary in time, vary conditionally on some covariate, vary in space, be multi-variate or a combination of those. The following sections provide an overview of the development of such distributions in the literature.

3.1.1   Frequency analysis: estimating a marginal distribution

Frequency analysis (FA) refers to the set of methods used to estimate the nonexceedance probability (generally expressed as a return period) associated with a given streamflow value, or inversely, the streamflow value associated with a given return period. It has many important applications: inundation mapping for urban planning, natural disaster declaration, water usage restrictions, design of at-risk structures such as dams, levees, power plants, and many more [Botto et al., 2017]. The most basic FA approach assumes that at a given site, data are independent and identically distributed (*iid*) realizations from a distribution with unknown parameters. Such a model therefore focuses on the marginal (a.k.a. interannual) distribution of data. The choice of the distribution depends on the target hydrologic variable. It may be suggested by official guidelines [Interagency Advisory Committee on Water Data, 1982], but it is also sometimes guided by local usage or empirical adequacy. For extreme variables characterizing floods and droughts, extreme value theory offers a strong theoretical foundation [Coles, 2001, Katz et al., 2002, Naveau et al., 2005]. The Generalized Extreme Value (GEV) distribution is hence often used for floods characterized with annual maxima variables [Martins and Stedinger, 2000]. Likewise, the Generalized Pareto Distribution (GPD) is the natural distribution for flood peaks above a high threshold [Rosbjerg, 1985, Lang et al., 1999, Mailhot et al., 2013]. Parameter estimation and uncertainty quantification can be achieved in many well-documented ways [e.g. Interagency Advisory Committee on Water Data, 1982, Commonwealth of Australia, 2019, Ramachandra Rao and Hamed, 2019].

Put simply, FA hence corresponds to estimating the marginal distribution of the target hydrologic variable in order to use its cumulative distribution function (cdf) or its quantile function. But this apparent simplicity hides a genuine practical challenge: the sampling un-

certainty affecting FA estimates is in general large, or even huge when high return periods ($T \geq 100$ years) are sought, which is the case for many applications. It follows that uncertainty quantification is an integral part of the FA exercise [Coles et al., 2003]. Moreover, developing strategies to reduce this uncertainty has been an important focus in the research literature. Three main directions can be considered for this purpose: temporal, causal or spatial expansion of information [Merz and Bloschl, 2008a,b, Viglione et al., 2013]:

1. Temporal expansion is mostly based on the use of historical or paleoflood information that may date back several centuries [e.g. Benito et al., 2004, Brázdil et al., 2006]. The typical workflow involves reconstructing flood peaks using available information, and using them in combination with systematic data from hydrometric stations. Specialized statistical techniques are required to account for the censored nature of historical series [Stedinger and Cohn, 1986, Naulet et al., 2005, Payrastre et al., 2011] and the large uncertainty affecting flood peak reconstructions [Reis and Stedinger, 2005].

2. Causal expansion refers to deducing the distribution of floods from the distribution of its main forcings [precipitation in particular, Eagleson, 1972, Fiorentino and Iacobellis, 2001, Sivapalan et al., 2005]. While earlier approaches attempted to achieve this in the form of analytical formulas, more recent developments tend to use a hydrologic model coupled to a rainfall generator to achieve this transformation by means of a large number of simulations. In France, several methods widely used in the engineering practice are based on this approach, in particular the GRADEX [Guillot and Duband, 1967] and its successor the SCHADEX [Paquet et al., 2013] methods, or the SHYREG method [Arnaud and Lavabre, 2002, Arnaud et al., 2016].

3. Spatial expansion refers to the joint use of data from several sites and will be discussed in section 3.1.3.

3.1.2 Modeling variability in time

The assumption of identical distribution underlying the approaches described in the previous section may be unrealistic because of some trend affecting the data [e.g. Nogaj et al., 2006, Slater et al., 2021]. Alternatively, this assumption does not allow using external information from climate covariates, for instance. To address these issues, a time-varying conditional distribution can be obtained by assuming that the parameters vary as a function of some temporal covariates. Time itself can be used as a covariate, resulting in a non-stationary distribution [Perreault et al., 2000a,b]. Other typical covariates include large-scale climate information such as global temperature [Westra et al., 2012] or climate indices [Steirou et al., 2019], synoptic-scale information such as weather type [Garavaglia et al., 2010] or airflow descriptors [Maraun et al., 2010], paleoclimate information [Devineni et al., 2013, Ho et al., 2015] or even non-climatic information [Prosdocimi et al., 2015, Slater et al., 2019]. Note that these distributions may or may not be stationary, depending on the stationarity of the covariates themselves.

The use of covariate-based models to describe time variability is now a well established approach, with generic tools available [e.g. Stasinopoulos and Rigby, 2007, Carpenter et al., 2017], but it still faces several methodological challenges, as reviewed by Slater et al. [2020]. In particular, the limited length of many station series induces a large sampling uncertainty and hence limits the power to detect trends or covariate effects [Bertola et al., 2020]. To overcome this limitation, the use of historical data in time-varying models has been explored [e.g. Machado et al., 2015, Xiong et al., 2020], along with regionalization approaches, as described next.

3.1.3   Spatial models

It is sometimes necessary to analyze the variable of interest at multiple sites. In particular, the spatial expansion of information mentioned in section 3.1.1 is based on the joint use of data from several sites in an attempt to reduce estimation uncertainty. This is known as Regional Frequency Analysis (RFA) in hydrology [Dalrymple, 1960, Cunnane, 1988, Bobée and Rasmussen, 1995, Hosking and Wallis, 1997]. The basic principle is to assume that within an homogeneous region, some parameters are the same for all sites (typically the shape parameter of a GEV distribution or its coefficient of variation) while other are site-specific. The estimation of regionally-constant parameters is hence informed by data from several sites, which should reduce estimation uncertainty. Research on RFA methods has initially focused on the definition of homogeneous regions [e.g. Burn, 1990] and on methods to estimate site-specific parameters at ungauged locations [Stedinger and Tasker, 1985]. Covariate-based models have also been used to describe the spatial variability of parameters by means of a regression with spatially-varying covariates [Panthou et al., 2012]. Finally, hierarchical models have been introduced as a general and flexible way to implement RFA schemes. The principle is to use a 2-level model, where the first level describes the distribution of data and the second level describes the spatial hyperdistribution of parameters, typically with a Gaussian spatial process which may also include a regression with spatial covariates [Cooley et al., 2007, Dyrrdal et al., 2015]. Moreover, approaches to explicitly model inter-site data dependence have been proposed, based on spatial copulas [Sang and Gelfand, 2010, Ghosh and Mallick, 2011, Bracken et al., 2016a] or on max-stable processes for extremes [Padoan et al., 2010, Westra and Sisson, 2011, Blanchet and Davison, 2011, Ribatet et al., 2012, Le et al., 2018].

Note that spatial models can also be combined with the approaches discussed in preceding sections. In particular, when historical and systematic data are available at several nearby stations, it is possible to use historical information within a RFA procedure [e.g. Jin and Stedinger, 1989, Gaume et al., 2010, Nguyen et al., 2014], hence combining the strength of spatial and temporal expansion of information. Alternatively, it is possible to apply the basic regionalization idea to time-varying models, by making assumptions on the spatial variability of parameters controlling the effect of temporal covariates. For instance, one may hypothesize that a trend is identical for all stations within a region, or varies smoothly in space. The resulting model produces distributions that vary in both space and time [Cunderlik and Burn, 2003, Aryal et al., 2009, Lima and Lall, 2010b, Gregersen et al., 2013, Panthou et al., 2013,

Sun et al., 2015a, Steinschneider and Lall, 2015, Ossandon et al., 2021, Le Roux et al., 2021].

3.1.4 Modeling of several variables

Multivariate models can result from multiple types of variables [Zscheischler et al., 2018], rather than multiple sites. Multi-variable models have also become a well-established approach, with inter-variable dependence being typically described using copulas [Favre et al., 2004, Salvadori and De Michele, 2004] or extreme-specific models [De Haan and De Ronde, 1998, Heffernan and Tawn, 2004]. They have also been combined with other approaches: multi-variable time-varying models have been proposed for drought management [Sarhadi et al., 2016] or dam safety [Bracken et al., 2018]; multi-variable spatial models have been used to build stochastic weather generators [Ailliot et al., 2015, Ahn, 2020]; 'full' multi-variable space-time models have even been derived for studying air pollution [De Iaco, 2011] or merging rainfall and radar data [Sideris et al., 2014], albeit within a Gaussian geostatistical framework that may be too restrictive for extreme or discrete variables, for example.

3.2 Frequency analysis of hydrologic extremes

My contribution on the topic of hydrologic variability has revolved around the development of probabilistic models to describe the space-time variability of one or several hydrologic variables, with a particular focus on flood frequency analysis.

3.2.1 Local and regional frequency analysis models

It is useful to start this section by introducing the notation that will consistently be used throughout the chapter. Let $Y(s, t)$ denote the random variable of interest - for instance, annual maximum streamflow - at time t and site s, and $y(s, t)$ denote the corresponding observed value. A probabilistic model can be viewed as an assumption on the distribution that generated these observations. For instance, focusing on a single site, the basic frequency analysis model assumes that $y(s, t)$ are independent and identically distributed (*iid*) realizations from a distribution $\mathcal{D}$ parameterized by a vector with C components $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_C)$:

$$Y(s, t) \sim \mathcal{D}(\boldsymbol{\theta}) \tag{3.1}$$

Typically, for annual maxima:

$$Y(s, t) \sim GEV(\mu, \sigma, \xi) \tag{3.2}$$

For most contributions described in this chapter, Bayesian inference was used for parameter estimation and uncertainty analysis [Renard et al., 2013b]. Two contributions were made in the field or regional frequency analysis (RFA). The first one addressed the question of accounting for spatial inter-site dependence and illustrated the feasibility of using a Gaussian copula for
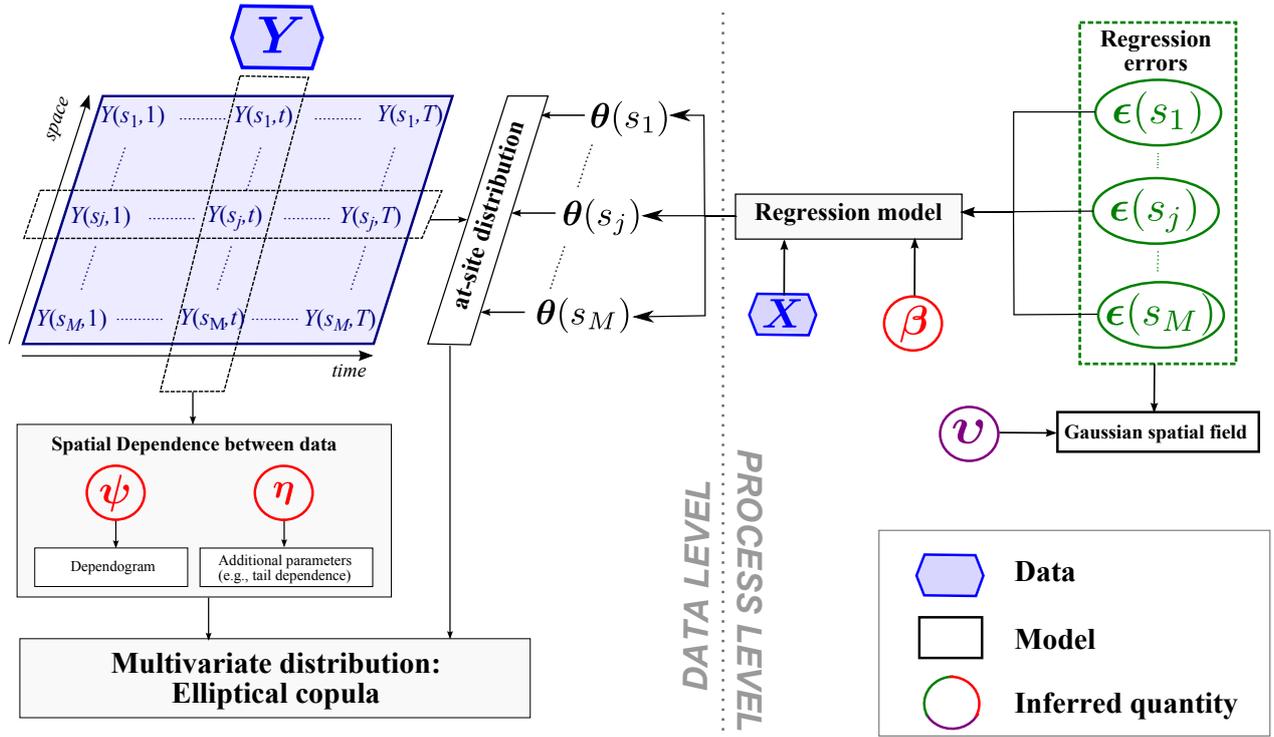
Figure 3.1: Schematic of the hierarchical model proposed for regional frequency analysis. Reproduced from Renard [2011].

this purpose [Renard and Lang, 2007]. The second one proposed a general framework for RFA based on hierarchical modeling [Renard, 2011]. The originality of this work at the time was that it underlined a common modeling framework for several works it was inspired from [Cooley et al., 2007, Aryal et al., 2009, Lima and Lall, 2009, 2010b, among others], and it included an explicit treatment of spatial dependence based on elliptical copulas. The description of spatial variability in this framework can be described as follows (see also Figure 3.1):

$$
\begin{cases}
Y(s,t) \sim \mathcal{D}\left(\boldsymbol{\theta}(s)\right) \\
g_c\left(\theta_c(s)\right) = h_c\left(\boldsymbol{x}(s), \boldsymbol{\beta}_c\right) + \varepsilon_c(s), \ \forall c = 1 \ldots C \\
(\varepsilon_c(s_1), \ldots, \varepsilon_c(s_M)) \sim GP(\boldsymbol{0}, \boldsymbol{\Sigma}_c(\boldsymbol{v}_c))
\end{cases}
\tag{3.3}
$$

The first line of equation (3.3) states the assumed distribution of the data, with parameters varying in space (but not in time). The second line is the spatial regression part of the model: for each component of the parameter vector, the parameter value is derived from a regression with spatially-varying covariates $\boldsymbol{x}(s)$ describing site or catchment properties (e.g., elevation, distance to sea, catchment size). The regression function $h_c(.,.)$ may be linear or more complex. Also note that similar to generalized linear models [McCullagh and Nelder, 1989], the link function g_c is used to map the range of parameter θ_c to $(-\infty, +\infty)$. For instance the logarithm function can be used if $\theta_c > 0$, or the logit function if $\theta_c \in (0,1)$. Finally, the third line of equation (3.3) states that the regression errors $\varepsilon_c(s)$ are realizations from a zero-mean spatial Gaussian process, with a covariance function parameterized by $\boldsymbol{v}_c$. Lines 2 and 3 essentially
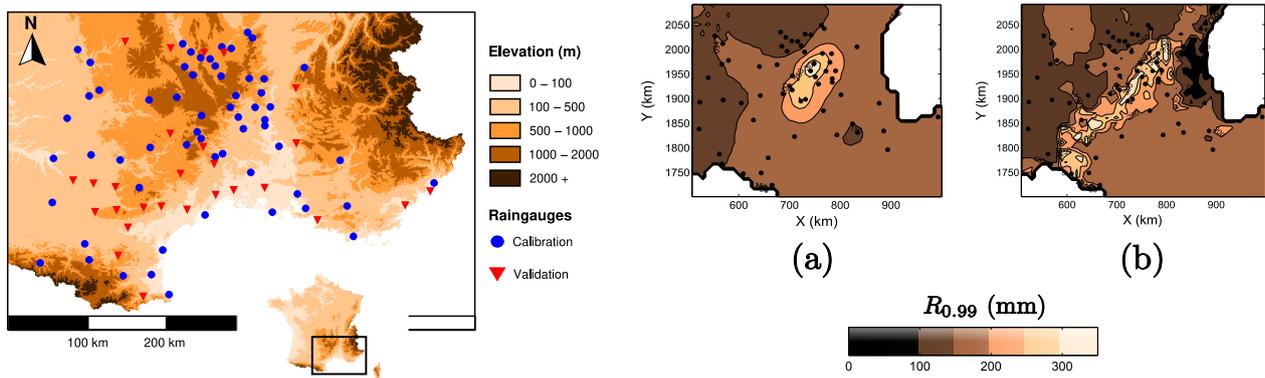
Figure 3.2: Estimation of the 100-year daily rainfall using a hierarchical model. The left panel shows the raingauges location in Southern France, the right panel shows the map of estimated quantiles (posterior median) based on a constant-mean (a) and an elevation-dependent (b) Gaussian spatial process. Modified from Renard [2011].

correspond to the 'external drift kriging' geostatistical model embedded into the second level of the hierarchical model. This allows interpolating between calibration sites and hence deriving maps of estimates, as illustrated in Figure 3.2. Moreover, such a 2-level hierarchical model avoids performing RFA estimation in two steps (local estimates first, then regression between local estimates and site properties), which enables a seamless propagation of sampling and regression uncertainties.

The ExtraFlo[1] research project (2009-2013), led by Michel Lang, aimed at comparing the main FA methods used in France for estimating extreme rainfall and floods - essentially, it was an FA competition. This required defining the spirit and the rules of the game, which in turned required addressing the following questions [Renard et al., 2013a]:

1. What is a good FA method? FA is a predictive exercise, not a curve-fitting one (is a dam designed so that it can withstand upcoming floods, or so that it would have withstood past floods?). Consequently, FA methods need to be judged on their predictive ability, not their goodness-of-fit.

2. What strategies can be used to compare methods fairly? Strategies based on simulated data are useful and handy since FA results can be compared against a known truth. However, conclusions entirely depend on the simulation setup: how realistic is it? Is it fair or might it favor specific types of methods? To avoid these questions, we opted for a data-based comparison, which is by essence more realistic but creates challenges due to the fact that the underlying truth is unknown.

3. How to deal with the limited length of existing series? Unless it is blatantly wrong, a 100-year flood estimate cannot be falsified at a single site. However, repeating the evaluation on many sites might be able to reveal shortcomings of the underlying estimation method.

---

[1]https://extraflo.inrae.fr

4. How to accommodate a wide variety of methods? Split-sample schemes adapted to local, regional or model-based FA methods were defined, so that methods could all be compared based on exactly the same validation data.

5. What criteria can be used? We defined criteria to judge the reliability of estimated quantiles based on validation data from many sites.

6. How to compare not only estimates but also their uncertainty? We proposed to integrate uncertainties into a predictive distribution, that can then be compared using the previously-defined strategy.

The project's results yielded interesting insights into the performance of methods used operationally [Kochanek et al., 2014, Lang et al., 2014, Neppel et al., 2014]. For floods in particular, two methods dominated their competitors in terms of predictive performances, namely the local version of the SHYPRE continuous simulation method and the mixed local+regional estimation of a GEV distribution. The Gumbel distribution is still widely used in the operational practice but the results demonstrated that it underestimates flood quantiles in Mediterranean catchments. On the other hand, using a locally-estimated GEV distribution and ignoring its uncertainty is not recommended either, because the difficulty in estimating the shape parameter results in frequent predictive failures. Interestingly, the predictive distribution is able to solve most of these failures, and it has been used in subsequent flood studies [Alliau et al., 2015]. Finally, all the purely regional methods displayed a quite poor reliability, suggesting that prediction in completely ungauged catchments remains a challenge.

Finally, all analyses discussed so far in this section were based on streamflow time series, but the associated uncertainty has been ignored. This might appear surprising considering that an entire chapter was devoted to this topic (Chapter 1). The impact of data uncertainty on low-, medium- or high-flow frequency analyses was investigated by Horner et al. [2016] by means of a Monte-Carlo propagation of data and sampling uncertainties. In a nutshell, the results of this study suggested that sampling uncertainty is by far the main contributor to the uncertainty around flow quantiles for floods and medium flows, at least with the typical sample sizes currently available ($\sim$30-60 years). By contrast, the contribution of data uncertainty is much more noticeable for low flow frequency analyses.

3.2.2 Using historical data

The use of historical data requires a specific statistical treatment to account for the censored nature of the historical flood sample and the large uncertainty around the reconstructed flood peak discharges. The Bayesian method we developed in Neppel et al. [2010] contributes to a more realistic characterization of the latter by distinguishing between independent errors (typically resulting from imperfectly-known water levels) and systematic errors (resulting from the imperfect rating curves used to reconstruct peak flows). Figure 3.3(b) illustrates a historical flood dataset consisting of a mixture of points $y(t)$ and intervals $[l(t); u(t)]$, with the latter

corresponding to floods for which the water level reached at the peak is poorly known. Importantly, Figure 3.3(b) disregards the fact that all discharge values (points, intervals but also the perception threshold q_0) have been obtained by means of rating curves. For ancient data, these rating curves have not been established at hydrometric stations as described in Chapter 1, but rather with a hydraulic model affected by large and non-negligible uncertainties (in particular in terms of bathymetry and friction). To account for this, Neppel et al. [2010] introduced systematic multiplicative errors γ_k: all discharges related to the kth rating curve may be affected by the same multiplicative error γ_k. If $\widetilde{Y}(t)$ denotes the unknown true discharge, this can be formalized as:

$$\widetilde{Y}(t) = \gamma_k Y(t) \tag{3.4}$$

A value $\gamma_k = 1.2$ hence means that the true discharge values are systematically 20% higher that the 'observed' ones shown in Figure 3.3(b) during the whole period of validity of the kth rating curve. Prior distributions can be specified for each term γ_k to represent the possible range of under/overestimation, typically based on a sensitivity analysis of the hydraulic model. Priors for recent γ_k will hence tend to be concentrated around 1, while priors for ancient γ_k will be less precise.

The objective is to estimate the distribution of the unknown true discharge. Under a GEV assumption $\widetilde{Y}(t) \sim GEV(\mu, \sigma, \xi)$, the likelihood associated with the observed dataset (composed of the perception threshold, point and interval data) is as follows:

$$
\begin{aligned}
p(q_0, \boldsymbol{y}, \boldsymbol{l}, \boldsymbol{u} | \mu, \sigma, \xi, \boldsymbol{\gamma}) = \prod_{k=1}^{K} & \left[F_{GEV}(q_0 | \mu/\gamma_k, \sigma/\gamma_k, \xi) \right]^{n_k} \times \\
& \prod_{t \in P_k} f_{GEV}(y(t) | \mu/\gamma_k, \sigma/\gamma_k, \xi) \times \\
& \prod_{t \in I_k} \frac{F_{GEV}(u(t) | \mu/\gamma_k, \sigma/\gamma_k, \xi) - F_{GEV}(l(t) | \mu/\gamma_k, \sigma/\gamma_k, \xi)}{u(t) - l(t)}
\end{aligned}
\tag{3.5}
$$

The first line of equation 3.5 corresponds to the n_k below-threshold values within the period of validity of the kth rating curve. The second and third lines correspond to point and interval data, respectively, within the period of validity of the kth rating curve. Note that systematic errors γ_k affect all three terms.

Figure 3.3(c) illustrates the quantile curve resulting from the application of this Bayesian approach to the data shown in Figure 3.3(b), and assesses how estimates vary with the analyzed period. A more thorough analysis suggested that when systematic errors are taken into account, the final uncertainty in estimated quantiles does not necessarily decrease when the analysis period increases [Lang et al., 2010]. The method was also subsequently applied to a case study in New Zealand by Griffiths et al. [2017], and also to other case studies performed by operational services in France (see section 3.5).

The four sub-catchments shown in Figure 3.3(a) were also used as part of the PhD work
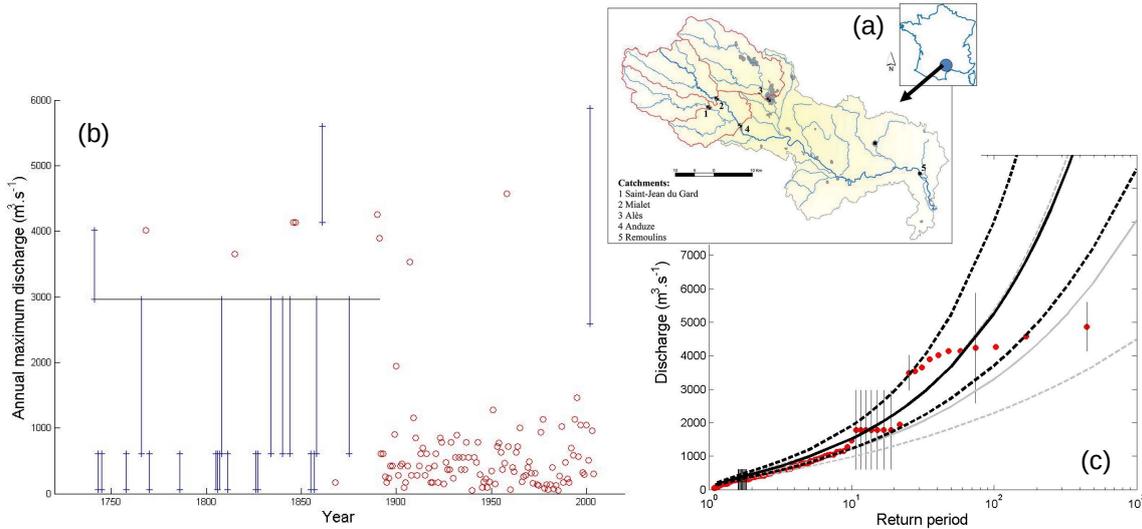
Figure 3.3: Historical case studies in the Gardons catchments. (a) Location of four sub-catchments of the Gard River where historical data are available; (b) historical dataset for the Anduze sub-catchment (the horizontal line is the perception threshold); (c) corresponding quantile curves and 90% uncertainty intervals obtained using the whole period 1741–2005 (thick black lines) or the systematic period 1892–2005 only (thin gray lines). Modified from Neppel et al. [2010].

of Anne Sabourin [2013] to develop a multivariate joint model for analyzing floods at all four sites. The objective was twofold: mixing regional estimation and the use of historical data, and studying how inter-site dependence evolves for very large floods. Results first indicated that approaches that either use historical floods or perform regional estimation or both lead to similar estimates. By contrast, a local estimation ignoring historical floods leads to markedly different estimates: this illustrates the benefit of extending the at-site sample using either historical or regional information or both to improve the robustness of estimated quantiles. Moreover, the availability of exceptionally large historical floods in several nearby catchments allowed demonstrating the existence of significant asymptotic dependence, calling for adapted multivariate statistical models [Sabourin and Renard, 2015].

More recently, we started to investigate the use of long stage records for frequency analysis, in particular during the PhD work of Mathieu Lucas [2023]. The motivation behind this work is the existence of several long, multi-century daily stage series in hydrometric archives, that are currently not used. The historical context is quite different from the one discussed up to now: historical information does not take the form of a censored sample of flood peak discharge, but rather of a systematic daily stage series. The challenge to address is therefore to build a chain of uncertainty estimation methods to move from the stage series to a discharge series that can be used for frequency analysis. This chain should account for many sources of uncertainties: uncertainty in stage values that were recorded once a day at a fixed time by observers, hence potentially missing flood peaks; uncertainty in reconstructing a whole history

of rating curves based on limited information and gaugings; uncertainty in correctly detecting rating shifts; uncertainty in estimating the frequency analysis distribution. Figure 3.4(a) illustrates the reconstruction of a more-than-200-year-long streamflow series for the Rhône River at Beaucaire, and highlights the highly variable uncertainty around this series [from about 30% during the XIXth Century to 5% in the last decades, Lucas et al., 2023]. Figure 3.4(b) shows the resulting quantile curve and its uncertainty. In this particular case, streamflow and sampling uncertainties contribute roughly equally to the final uncertainty affecting high flood quantiles. A similar analysis has been initiated for the Rhine River [Lang et al., 2022], and others could follow (e.g. the Garonne River). This work contributes to giving a second life to this rich limnimetric heritage.

3.3 Modeling variability in time

3.3.1 Trend detection

Determining the impact of climate change on hydrologic regimes is an important endeavor that attracted a lot of attention in the scientific literature. It may be addressed in a variety of ways, but one of the facets of this question it to determine if changes can be detected in observed streamflow series. This was the main topic of my PhD thesis [Renard, 2006]. More specifically, my PhD work addressed the following questions: how to detect changes in hydroclimatic series? What are the observed changes affecting hydrologic regimes, and in particular floods and droughts, in France?

The first question regarding the methodology for detecting trends was addressed at two spatial scales. At the local scale, many existing at-site tests for change were compared by simulations, leading to a selection diagram identifying the most suitable test given the properties of the target variable [Renard et al., 2006c]. At the regional scale, two aspects were studied: field significance and regional consistency [Renard et al., 2008]. The former aspect arises from the fact that when a test is repeated on many sites, some amount of false detection is bound to occur. Field significance tests therefore aim at evaluating the H_0 hypothesis: 'data from ALL sites are stationary', given the number of sites and the local significance level. Several existing methods were compared, and the False Discovery Rate approach of Benjamini and Hochberg [1995] was found to be both easily applicable and efficient. Regional consistency, on the other hand, aims at assessing whether trends are consistent within a homogeneous hydro-climatic region. A specific method was developed to test the hypothesis that, after normalization of all local series, an identical trend is affecting all sites. This method accounts for the existence of spatial dependence.

These local and regional testing procedures were then applied to a number of hydrologic variables describing several facets of floods, droughts and mean flows for nearly 200 hydrometric stations in France. These stations were selected on the basis of several requirements including the absence of a significant direct human influence (dam, water withdrawal, etc.). This dataset of 'near-natural' stations was updated several times in subsequent years [Giuntoli et al., 2012,
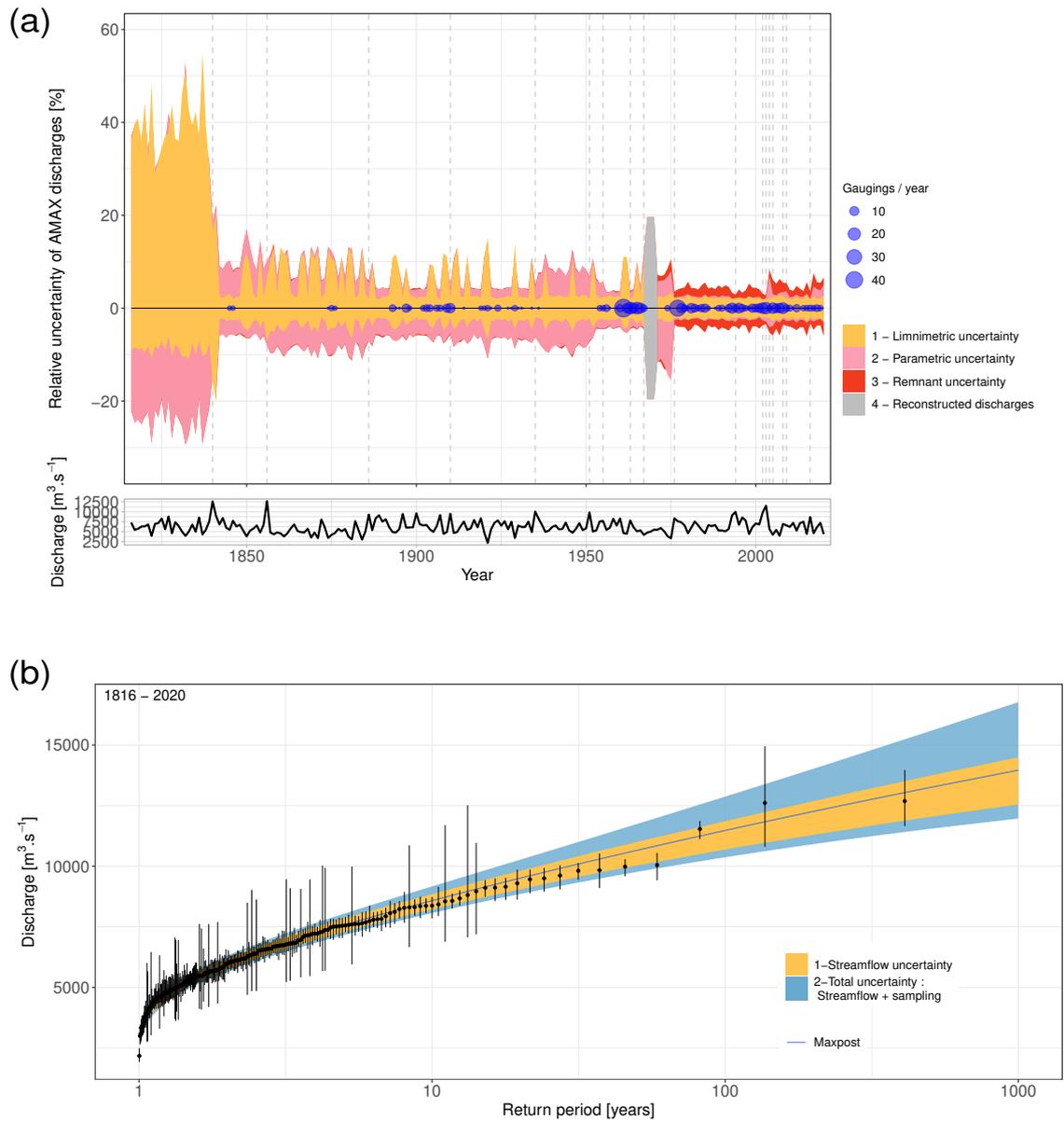
Figure 3.4: The Rhône River at Beaucaire, 1816-2020. (a) Reconstructed series of annual maximum discharge (bottom) and associated relative uncertainties associated with three sources of error (top); (b) quantile curve and associated uncertainties. Modified from Lucas et al. [2023].
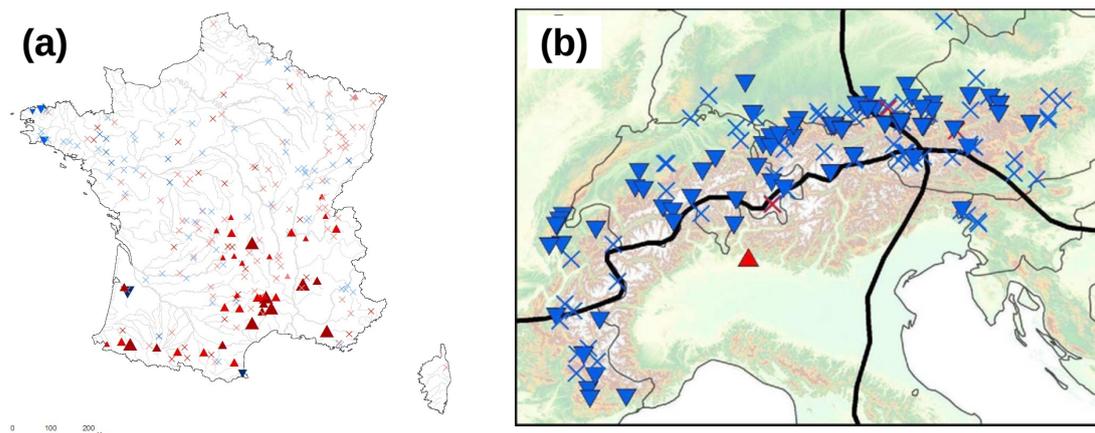
Figure 3.5: Two examples of analyses where significant trends were detected. (a) Drought duration in France, 1968-2008 [Giuntoli et al., 2013]; (b) Start of snowmelt flows in the Alps, 1961-2006 [Bard et al., 2015]. Color indicates the change direction (blue = decrease, red = increase), crosses denote non-significant changes.

2013] and has been made freely available through international databases [Hall et al., 2015] such as the GRDC[2]. Overall, no generalized change was found at the national scale on the basis of at-site tests and field significance assessment. Homogeneous hydro-climatic regions were also defined, and for most of them the regional consistency test did not reveal consistent trends. There were, however, three exceptions: in the northeast flood peaks were found to increase; in the Pyrenees high and low flows showed decreasing trends; in the Alps, earlier snowmelt-related floods were detected, along with less severe drought and increasing runoff due to glacier melting.

Throughout the years I contributed to several trend analyses as summarized in Table 3.1. It is difficult to shortly summarize the results given that these studies differed in the target variables, the spatial scale of analysis and the underlying objectives. However, a few common threads emerged. Significant changes were generally detected in catchments strongly influenced by snow (Figure 3.5(b) for instance). This can be explained by the fact that temperature is a major driver for such catchments, and temperature changes are strong and widespread. At the opposite, flood changes were very few, which is consistent with the literature and may be explained by the diversity and complexity of flood-generating mechanisms [e.g. Sharma et al., 2018]. Clear patterns of change were also sometimes found for low flow variables (Figure 3.5(a) for instance), but in many cases these patterns turned out to be highly sensitive to the analyzed period, suggesting that multidecadal variability might play a role.

3.3.2 Conditional modeling

Modeling time-varying distributions is a natural follow-up to the analyses described in the previous section: it may be used to describe a trend or the effect of some climate covariate

---

[2]https://www.bafg.de/GRDC/EN/Home/homepage_node.html

References	Flow Component	Spatial Extent	Peculiarities and key findings
Galéa et al. [2005]	Low, medium and high flows	2 catchments in Southwest France	Influenced catchments: trends consistent with expected effect of water withdrawal for irrigation
Renard et al. [2006c], Lang and Renard [2007], Renard et al. [2008]	Low, medium and high flows	France	No change in most regions, with 3 exceptions: increasing floods in the Northeast, decreasing flows in the Pyrenees, snowmelt-related changes in the Alps
Giuntoli et al. [2012, 2013]	Low, medium and high flows	France	Update of the studies above. Split between a wetter north and a drier south. In particular, a consistent increase of drought severity is found in southern France
Bard et al. [2012a,b, 2015]	Low, medium and high flows	Alps	Clear trends in glacial- and snowmelt-dominated regimes: less severe winter droughts, earlier start and increased duration of the snowmelt season, increasing volume and peak of snowmelt flows for glacial regimes
Dudley et al. [2017]	Snowmelt flows	US	Widespread trends toward earlier snowmelt flows
Hodgkins et al. [2017a]	Groundwater levels	Glacial aquifer system, Northern US	Increases in groundwater levels dominate where human influence is low
Hodgkins et al. [2017b]	Major floods	North America and Europe	No compelling evidence for changes in the occurrence of major floods. Influence of modes of multidecadal variability such as the AMO in some regions.
Dudley et al. [2019]	Low flows	US	Comparison reference vs. human-impacted stations. Agricultural: consistent with reference stations; Dam-regulated: many upward trends; Urban. many trends, both upward and downward
Hodgkins et al. [2023]	Extreme low flows	World	Decreasing occurrence rate for cold-season low flow regimes. No clear trend for warm-season low flow regimes

Table 3.1: Summary of the trend analyses I contributed to.

within a framework that allows making probabilistic statements and hence goes beyond trend detection. The most basic approach to build a time-varying probability model is to assume that data are realizations from a given distribution $\mathcal{D}$ (e.g. Gumbel, GEV, etc.) whose parameters vary in time [Renard et al., 2006b, 2013b]:

$$Y(t) \sim \mathcal{D}\left(\boldsymbol{\theta}(t)\right) \tag{3.6}$$

The temporal variations of parameter $\boldsymbol{\theta}(t)$ can be described by setting up, for each component c of the parameter vector, a regression with time-varying covariates $\boldsymbol{x}(t)$ (Figure 3.6(a)):

$$g_c\left(\theta_c(t)\right) = h_c\left(\boldsymbol{x}(t), \boldsymbol{\beta}_c\right) \tag{3.7}$$

Notation is similar to the one used in section 3.2.1 for a spatial regression: g_c is the link function used to map the range of parameter θ_c to $(-\infty, +\infty)$. $h_c(., .)$ is the regression function and $\boldsymbol{\beta}_c$ its parameters that need to be estimated. Bayesian inference is performed which allows including some prior knowledge and using Bayesian model averaging of multiple candidate models [Renard et al., 2006b, 2013b].

While quite simple, the approach embodied in equation (3.7) is flexible and allows covering a variety of situations. The regression might just be a constant for a time-invariant parameter, or take the form of a linear trend or a step change or a more complex structure [Renard et al., 2006b]. The time-varying covariate $\boldsymbol{x}(t)$ may be time itself, or climate indices such as ENSO/IPO [Westra et al., 2015]. Finally, changes may affect any parameter of the distribution, which in particular allows evaluating changes in variability and not just in mean. This was investigated during the PhD work of Maleki Badjana [2018], who analyzed trends in streamflow and rainfall variables within the Kara River catchment in West Africa. A variety of distributions was used depending on the target variable, with trends affecting both location and scale parameters. For some variables, trends in variability were found to be more noticeable than trends in mean [Figure 3.7, Badjana et al., 2017].

In spite of this flexibility, the local approach presented here is limited by the strong sampling uncertainty that results from a single-site analysis, thus restricting the ability to identify time-varying components. As previously with frequency analysis (section 3.2.1) or trend detection (section 3.3.1), a natural approach to address this limitation is to move to a regional scale.

3.4 Modeling variability in space and time

3.4.1 Space-time probabilistic models

When multiple sites are analyzed together, the parameters of the parent distribution may vary in both space and time:

$$Y(s, t) \sim \mathcal{D}\left(\boldsymbol{\theta}(s, t)\right) \tag{3.8}$$
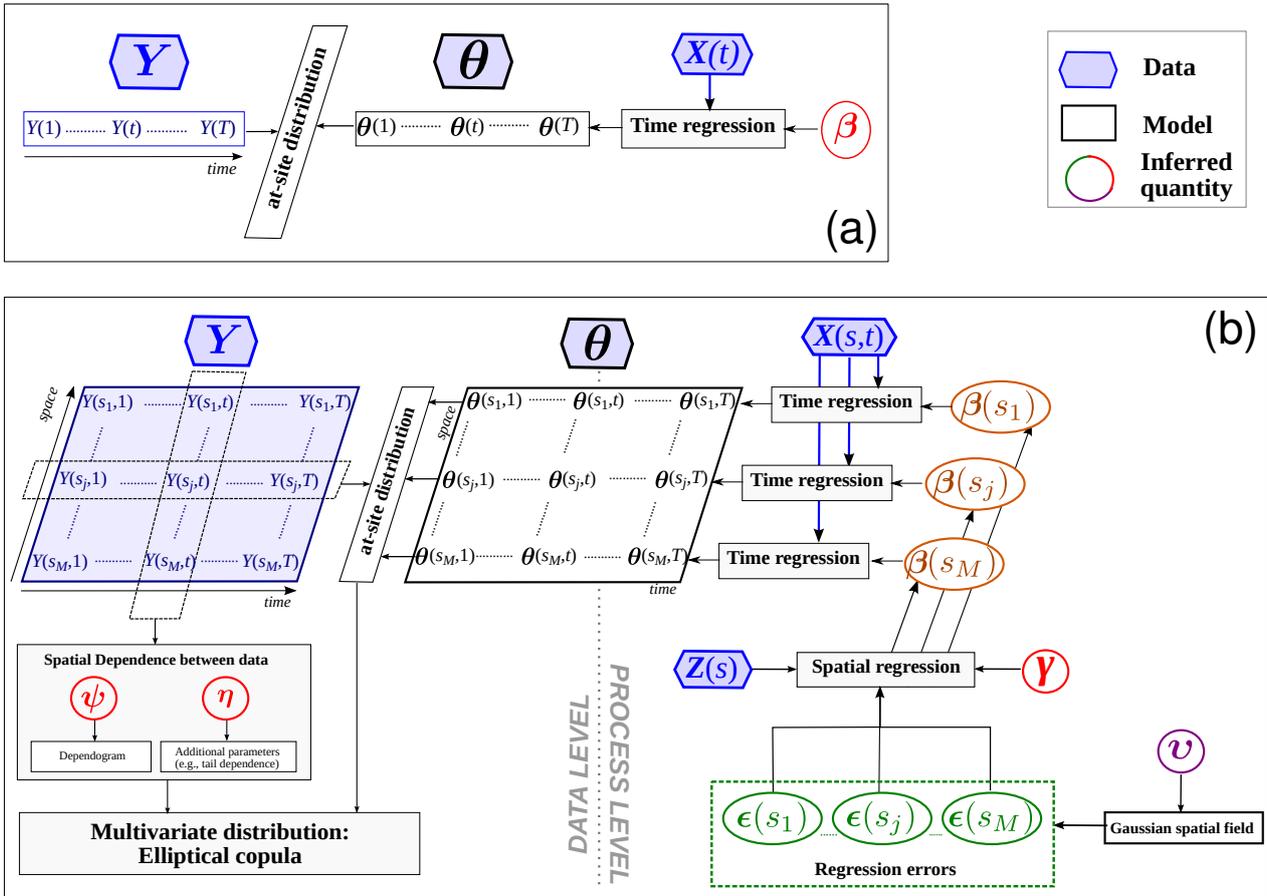
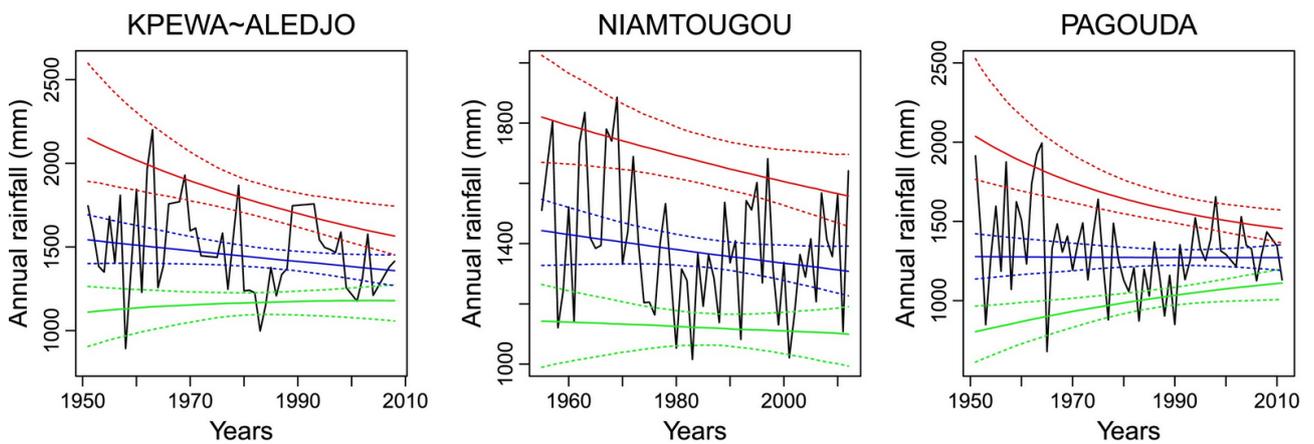Figure 3.6: Schematic of time-varying models. (a) local model; (b) regional model.

Figure 3.7: Annual rainfall time series at three stations and estimated trends expressed on rainfall quantiles. The green, blue and red solid lines represent 0.1, 0.5 and 0.9 quantiles respectively. Dotted lines represent 95% posterior intervals. Reproduced from Badjana et al. [2017].

A simple strategy to describe the space-time variability of parameters is to use the time-varying model of section 3.3.2 (Figure 3.6(a)) at each site, and assume that some parameters are constant across sites, while others are site-specific. This strategy was applied to a time-varying GEV model for annual maximum streamflow in Renard et al. [2006a], with the trend and the shape parameters μ_1 and ξ being treated as spatially constant:

$$Y(s,t) \sim GEV\left(\mu_0(s)(1 + \mu_1 t), \sigma(s), \xi\right) \qquad (3.9)$$

Such a model is of interest to obtain a precise estimation of parameters that are typically difficult to estimate locally, but it is affected by two limitations. First, the assumption of constant parameters is quite strong and may be unrealistic in large regions. Second, the existence of spatial dependence needs to be accounted for.

During his PhD thesis, Xun Sun [2013] therefore defined a more complete framework for building space-time models. The basic idea was to merge the time-varying model of section 3.3.2 (Figure 3.6(a)) and the regionalization approach of section 3.2.1 (Figure 3.1), leading to the space-time modeling framework schematized in Figure 3.6(b). This framework first describes temporal variability or trends by means of a regression with time-varying covariates, and then describes how the covariates effects vary in space with a spatial hierarchical model. Going back to the simple GEV model above, this framework would allow for instance that the trend varies as a function of the site elevation and/or stochastically in space. Moreover, the framework includes a description of spatial dependence based on elliptical copulas. This module was carefully implemented to allow for missing values, which are unavoidable in regional datasets where data availability varies between sites.

This framework was used to build a model describing the influence of ENSO on precipitation extremes. The model was first developed at a small regional scale, using 16 raingauges located in Southeast Queensland, Australia [Sun et al., 2014]. This case study allowed setting up the nonlinear regression model and demonstrated the interest of regionalizing, with uncertainties around quantile estimates being divided by a factor of two (Figure 3.8(a)). A similar study was then performed at the global scale, by repeatedly applying the regional model on 'regions' defined as 5×5 degrees cells [Sun et al., 2015b]. The effect of ENSO was found to vary in space and between seasons, with the strongest effect generally occuring in boreal winter (DJF, Figure 3.8(b)). Importantly, the effect of ENSO on extreme precipitation was found to be asymmetric, with most parts of the world experiencing a significant effect only for a single ENSO phase (El Niño or La Niña).

3.4.2 Hidden Climate Indices models

The approach described in the previous section allows modeling temporal variability by means of known time-varying covariates. Standard climate indices (SCIs) such as the Southern Oscillation Index (SOI), the North Atlantic Oscillation (NAO) index and many others[3] are typically

---

[3]`https://climatedataguide.ucar.edu/climate-data/overview-climate-indices`
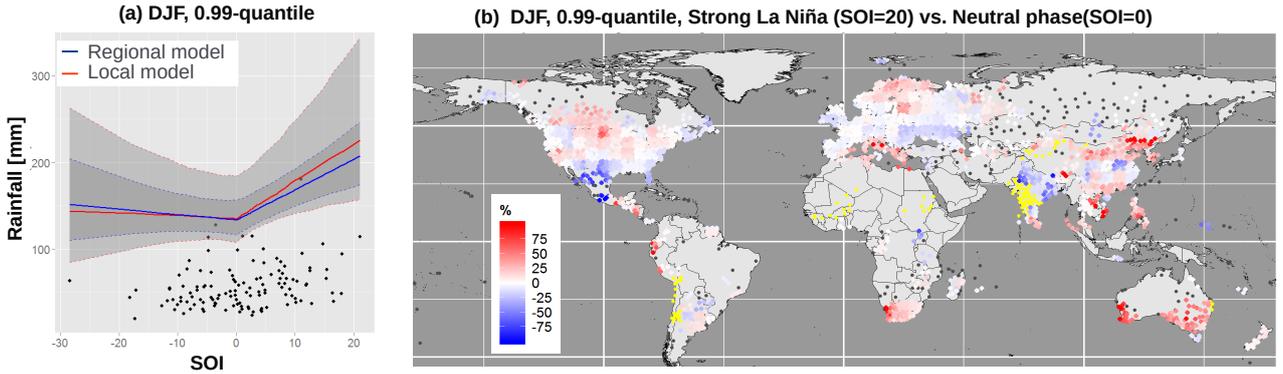
Figure 3.8: (a) Summer maximum daily rainfall at one site in Southeast Queensland plotted against the Southern Oscillation Index (SOI), and corresponding 0.99-quantile estimated with local and regional models. Gray areas represent 90% credibility intervals; (b) effect of La Niña on 0.99 precipitation quantile during the DJF season. Grey dots denote cells with too few stations to perform a regional analysis, yellow dots denote dry regions with frequent zero precipitations during DJF. Modified from Sun et al. [2014, 2015b].

chosen for this purpose. However, identifying relevant SCIs in France proved to be difficult: relationships with floods or rainfall extremes were for the most part non significant, while relationships with droughts were at best moderate [Giuntoli et al., 2013]. However, relationships with the frequency of weather patterns specifically derived for France were stronger [Garavaglia et al., 2011, Giuntoli et al., 2013]. This motivated the need for a different approach: instead of relying on standard climate indices, which are essentially predefined time series summarizing atmospheric or oceanic space-time fields, we proposed to treat the time-varying covariates as unknown temporal latent variables that need to be inferred from the target data. In other words, we proposed to uncover the hidden climate indices (HCIs) that govern the temporal variability of the target data.

The development of the HCI modeling framework was made in several steps. The first one was implemented during a 6-month research visit at the Water Center of Columbia University (US). This first model started with the space-and-time-varying distribution of equation (3.8), and further assumed that a single component of the parameter vector $\boldsymbol{\theta}$ varies in time as follows [Renard and Lall, 2014]:

$$g\left(\theta(s,t)\right) = \lambda(s)(1 + \tau(t)) \tag{3.10}$$

In this equation, $\lambda(s)$ can be interpreted as the 'normal' parameter value at site s, and $\tau(t)$ is the deviation from this value at time t: $\tau(t) = +0.2$, for instance, means that the parameter is 20% above normal. Note that this formulation hence implies that the relative effect of $\tau(t)$ is the same on all sites. Importantly, $\tau(t)$ is treated as an unknown time series and is therefore estimated from the data. This first HCI model was applied to 16 gauging stations in Mediterranean France (Figure 3.9(a)). The target variable was the number of flood events recorded at each station during the autumn season (OND), modeled with a Poisson
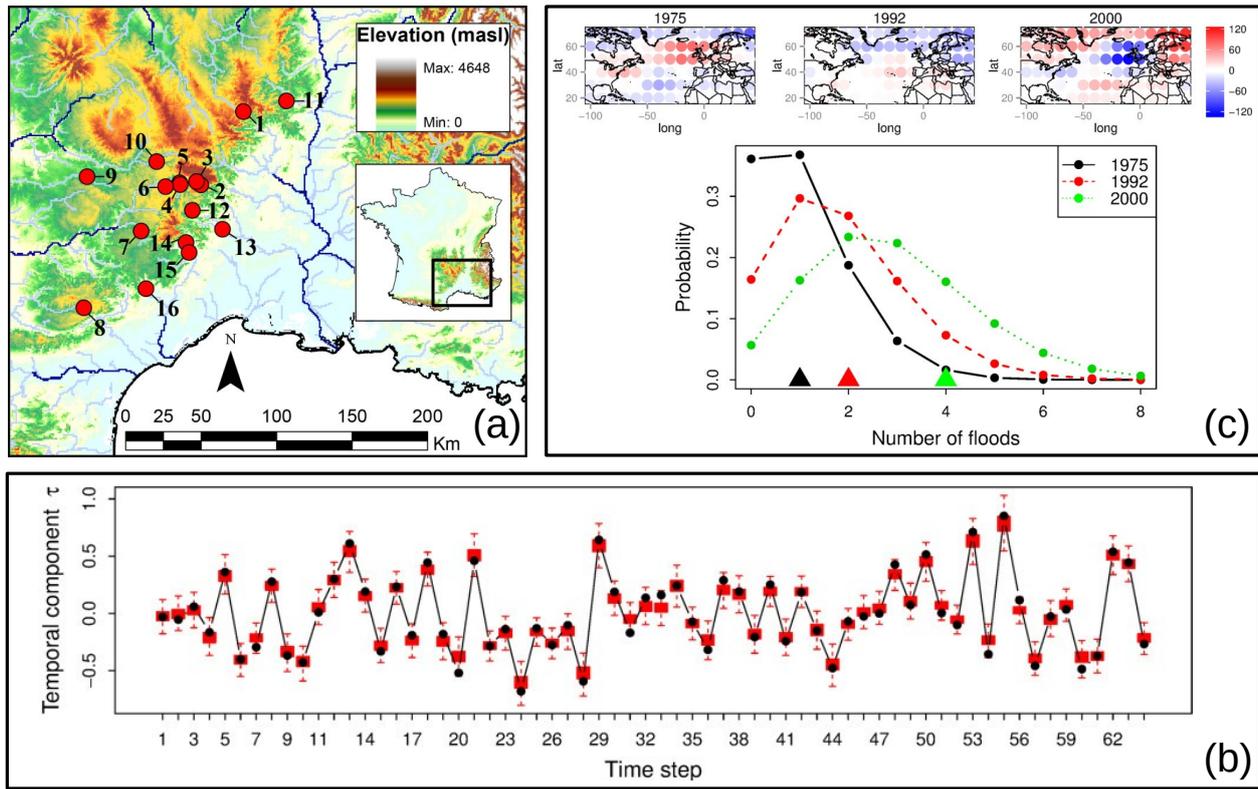
Figure 3.9: First application of a Hidden Climate Index (HCI) model. (a) Location of the gauging stations used in the case study; (b) estimated HCI (boxplots represent the posterior distribution, the line and points represent the modal estimate); (c) predictive distribution of the number of floods for site 5, conditional on the observed geopotential height anomalies for the years 1975, 1992, and 2000. Triangles represent the observed number of floods. Modified from Renard and Lall [2014].

distribution. Figure 3.9(b) shows that the HCI $\tau(t)$ could be estimated quite precisely. In addition, this HCI could be related to specific spatial patterns of atmospheric pressure in the North Atlantic, which could be used to predict the distribution of the number of floods from geopotential heights anomalies (Figure 3.9(c)).

This first model made strong assumptions (a single HCI leading to a spatially constant relative effect) which restricted its application to fairly small regions. This was addressed during a 2-year research visit at the University of Adelaide (Australia) as part of the HEGS project[4] (Hydrologic Extremes at the Global Scale). Renard and Thyer [2019] first enabled the use of several HCIs with spatially-varying effects, but within the restrictive context of occurrence 0/1 data modeled with a Bernoulli distribution:

$$\begin{cases} Y(s,t) \sim \mathcal{B}\left(\theta(s,t)\right) \\ logit\left(\theta(s,t)\right) = \lambda_0(s) + \lambda_1(s)\tau_1(t) + \cdots + \lambda_K(s)\tau_K(t) \end{cases} \tag{3.11}$$

Estimating parameters in equation (3.11) requires imposing identifiability constraints. An easy glimpse into this problem can be obtained by remarking that multiplying $\lambda_k(s)$ and dividing

---

[4]https://globxblog.github.io/hegs

$\tau_k(t)$ by the same number does not change anything to equation (3.11). Consequently, each HCI time series $\tau_k(t)$ is constrained to have mean 0 and standard deviation 1. In addition, a stepwise inference is adopted, i.e. the model is first estimated with a single component, then the second component is estimated conditionally on the previously-estimated component 1, etc. Finally, a spatial Gaussian process is used to model HCI spatial effects $\lambda_k(s)$, in the spirit of the hierarchical models described in the previous sections (Figure 3.1 and Figure 3.6(b)).

The formulation of the model is completed by assuming that all $Y(s, t)$ are independent, **conditionally** on the values taken by the temporal HCIs $\boldsymbol{\tau}$ and their spatial effects $\boldsymbol{\lambda}$. Note that this is not equivalent to assuming space-time independence: nearby sites similarly affected by the same HCIs will have similar temporal variations, i.e. they will show inter-site dependence. From this standpoint, the use of a common set of HCIs to describe temporal variability in a set of stations can even be viewed as an indirect approach to model spatial dependence. This indirect treatment is of great practical interest for the following reasons:

1. the treatment of missing values is straightforward, which is a major advantage for highly irregular station-based datasets.

2. discrete variables can easily be handled, which offers advantages over copula-based approaches for which specific difficulties arise with discrete variables [e.g. non-uniqueness, non-identifiability, see Genest and Nešlehová, 2007, Faugeras, 2017].

Several synthetic and real-life case studies were conducted to study the properties of this multi-HCI model [Renard and Thyer, 2019]. The first conclusion was that estimating several HCIs and their effects from a multi-site dataset of flood occurrences was feasible and not prone to overfitting when the number of sites is large. A case study in France (Figure 3.10) allowed identifying several HCIs that did not correspond to any preexisting climate index, but that were nonetheless linked with specific spatial patterns in atmospheric variables, making them interpretable in terms of climate variability and opening the way for predictive applications. A case study in Australia showed that the first estimated HCI was strongly correlated with the NINO4 index, indicating that the model correctly identified the ENSO-related climate drivers of floods in this area.

The full generalization of this model was finally described by Renard et al. [2021]. In particular, this general framework enables the use of any user-specified distribution. It can also be used in a multi-variable case, where several distinct distributions are needed to model several variables [François et al., 2014]. Formally, let $Y_v(s, t)$ denote the random variable representing variable v $(= 1, \ldots, V)$ at time t and site s. The assumptions made when building an HCI model are illustrated in Figure 3.11 and start by the data model:

$$Y_v(s, t) \sim \mathcal{D}_v\left(\boldsymbol{\theta}_v(s, t)\right) \tag{3.12}$$

The distribution $\mathcal{D}_v$ is variable-specific. For instance if Y_1 denotes an annual number of flood events and Y_2 denotes an annual average flow, then it would be sensible to use the Poisson distribution for $\mathcal{D}_1$ and the lognormal distribution for $\mathcal{D}_2$. The parameter vector $\boldsymbol{\theta}_v(s, t)$ is also
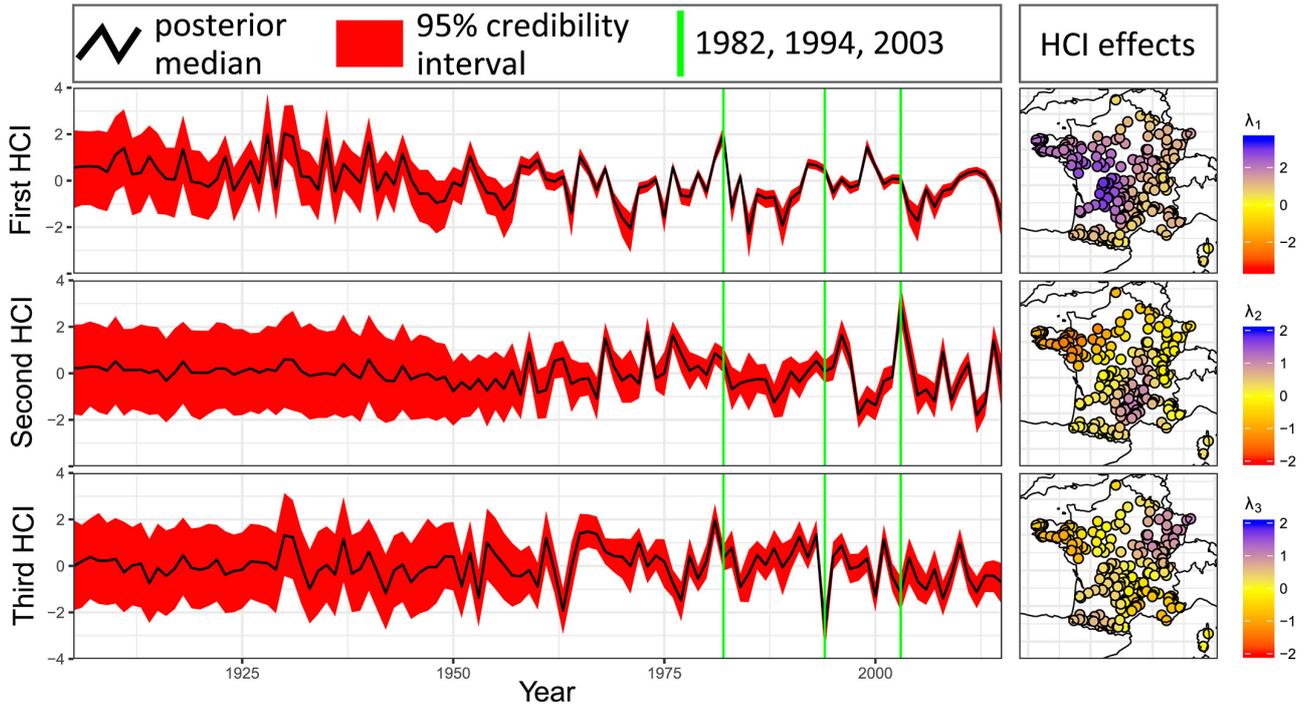
Figure 3.10: Three Hidden Climate Indices (HCIs; τ_1, τ_2 and τ_3) and their effects (λ_1, λ_2 and λ_3) identified from flood occurrence data in France. The increasing uncertainty as one moves back in time is due to the decreasing data availability. Reproduced from Renard and Thyer [2019].

variable-specific. Each component $\theta_{v,c}$ of the parameter vector is allowed to vary in space and time as follows (box 2 in Figure 3.11):

$$g_{v,c}\left(\theta_{v,c}(s,t)\right) = \lambda_{0,v,c}(s) + \lambda_{1,v,c}(s)\tau_1(t) + \cdots + \lambda_{K,v,c}(s)\tau_K(t) \qquad (3.13)$$

This is very similar to the previous equation (3.11), except that the link function $g_{v,c}$ and the spatial HCI effects $\lambda_{k,v,c}$ depend on the variable v. Note, however, that the temporal HCIs τ_k are assumed to be the same for all variables. The motivation is that using a common set of HCIs for all variables can induce inter-variable dependence, in the same way as it can induce spatial dependence as previously discussed.

The formulation of the model is completed by assuming that each spatial HCI effect λ_k and temporal HCI τ_k is a realization from a spatial or temporal Gaussian process (boxes 3 and 4 in Figure 3.11). Non-identifiability is handled as previously by assuming that each HCI time series $\tau_k(t)$ has mean 0 and standard deviation 1 and by performing stepwise inference. Finally, full conditional independence is assumed.

Three case studies were used to evaluate the HCI modeling framework. The first one focused on 'hot-and-dry', fire-prone summer conditions in Southeast Australia. In this study, three physical variables (temperature, precipitation and streamflow) were measured on three distinct station networks, with varying data availability and representing hundreds of sites in total (Figure 3.12(a)). Four statistical variables were defined from these times series: the
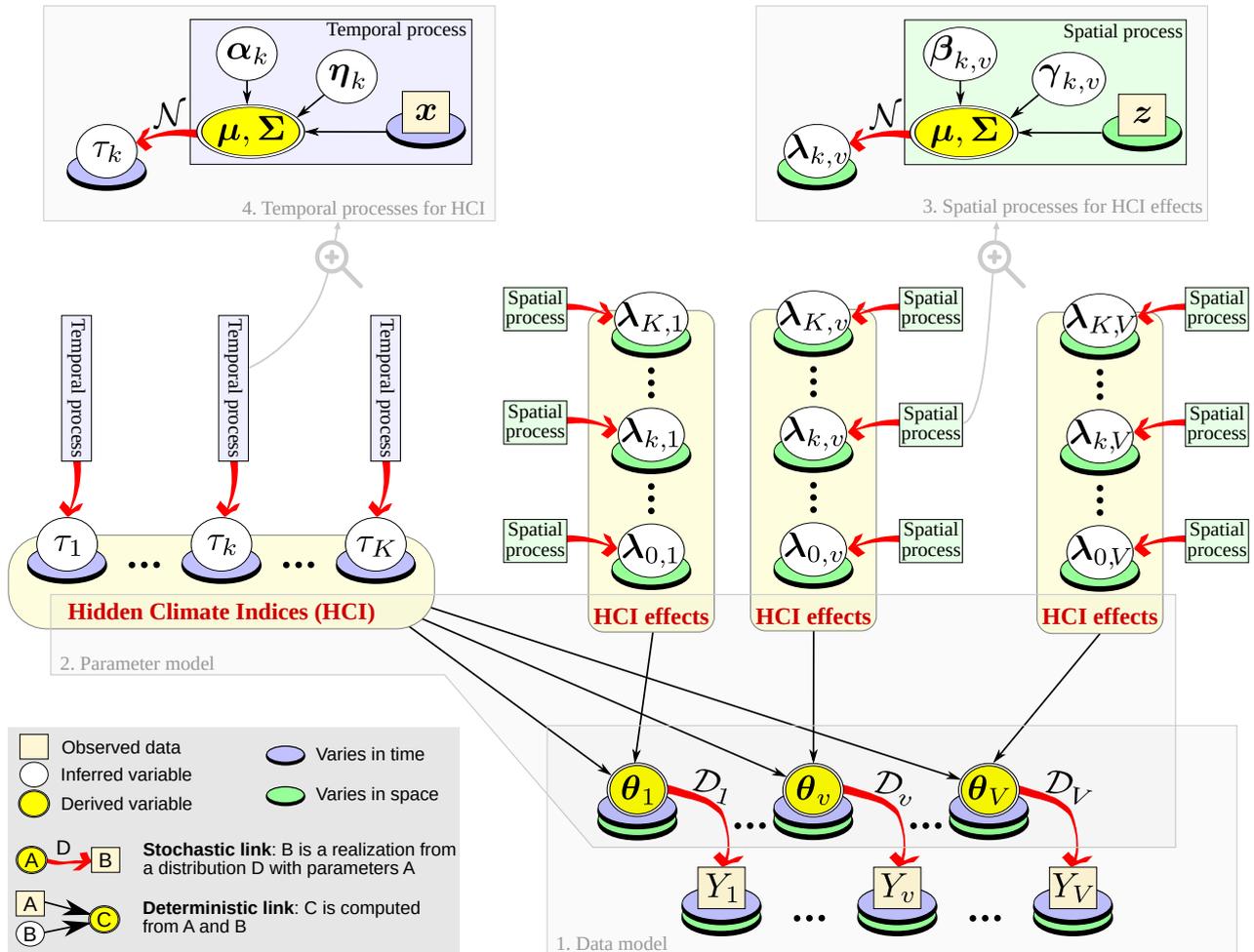
Figure 3.11: Schematic of the hidden climate indices (HCI) modeling framework. Reproduced from [Renard et al., 2021].

number (Tn) and maximum intensity (Tx) of heatwaves, the dry duration (Pd, % of days during the summer with near-zero precipitation) and the drought duration (Qd, % of days with streamflow being below a low-flow threshold). The joint modeling of these four variables is challenging because, apart from the highly variable data availability, it includes continuous (Tx), discrete (Tn) and even mixed continuous+discrete variables (Pd and Qd, which may both reach values of 0% and 100% with a non-zero probability). The HCI model set up for this case study used Generalized Pareto, Poisson and censored Gaussian distributions for each variable type. Figure 3.12(b) illustrates some of the HCIs that were identified. The first HCI has a positive effect on all four variables over the whole region, and hence describes the general 'hot-and-dryness' of each year. In contrast, the third HCI has a positive effect on temperature variables but a small effect on dry and drought durations: this can hence be interpreted as a 'temperature-only' HCI, and the increasing trend that can be discerned in the recent decades as a warming signal. Finally, Figure 3.12(c) illustrates the time-varying distributions arising from the HCI model, and how the co-variability in this distributions creates dependence, both inter-site and inter-variable in this case since the two variables Pd and Qd are measured at distinct locations. Finally, an interesting aspect of Figure 3.12(c) is that predictions can be made for the variable Qd even during no-data periods: this is because the HCIs have been estimated during the whole study period, and since they are common to all variables, equation (3.13) can be used to estimate Qd even before the availability of any streamflow data. This represents a transfer of information between variables that may have interesting applications.

The second case study aimed at taking advantage of this ability to transfer information between variables [Renard, 2023]. More precisely, it was based on the joint analysis of annual maxima at hydrometric stations and flood marks at sites (Figure 3.13), the latter being recorded in the recent national database 'repères de crues'[5]. The idea was to use flood marks, available over the period 1705–2015, to reconstruct flood probabilities at stations over this whole long period, i.e. well before the stations even existed. The usual historical approach based on using the flood marks to reconstruct peak discharge using hydraulic modeling (see section 3.2.2) is difficult to apply to such a large number of stations. We therefore opted for a simpler HCI-based approach, preserving the original localization of flood marks at sites and their original nature as time series of occurrence/non-occurrences. The following model was used:

$$
\begin{cases}
Q(s,t) \sim GEV\left(e^{\mu_0(s)} \times \left(1 + \sum_{k=1}^{K} \mu_k(s)\tau_k(t)\right), e^{\mu_0(s)} \times e^{\gamma(s)}, \xi(s)\right) \\
O(r,t) \sim \mathcal{B}\left(logit^{-1}\left(\lambda_0(r) + \sum_{k=1}^{K} \lambda_k(r)\tau_k(t)\right)\right)
\end{cases}
\tag{3.14}
$$

The idea behind this model is the same as in the previous case study: using the same set of HCIs $\tau_k(t)$ to model the temporal variability of flood peaks at stations $Q(s,t)$ and flood marks at sites $O(r,t)$ is an indirect way to model dependence between them, and can be used as a device to transfer information. Results demonstrated that flood marks allow estimating the time-varying probability of exceeding some high discharge threshold at stations during the

---

[5]https://www.reperesdecrues.developpement-durable.gouv.fr/
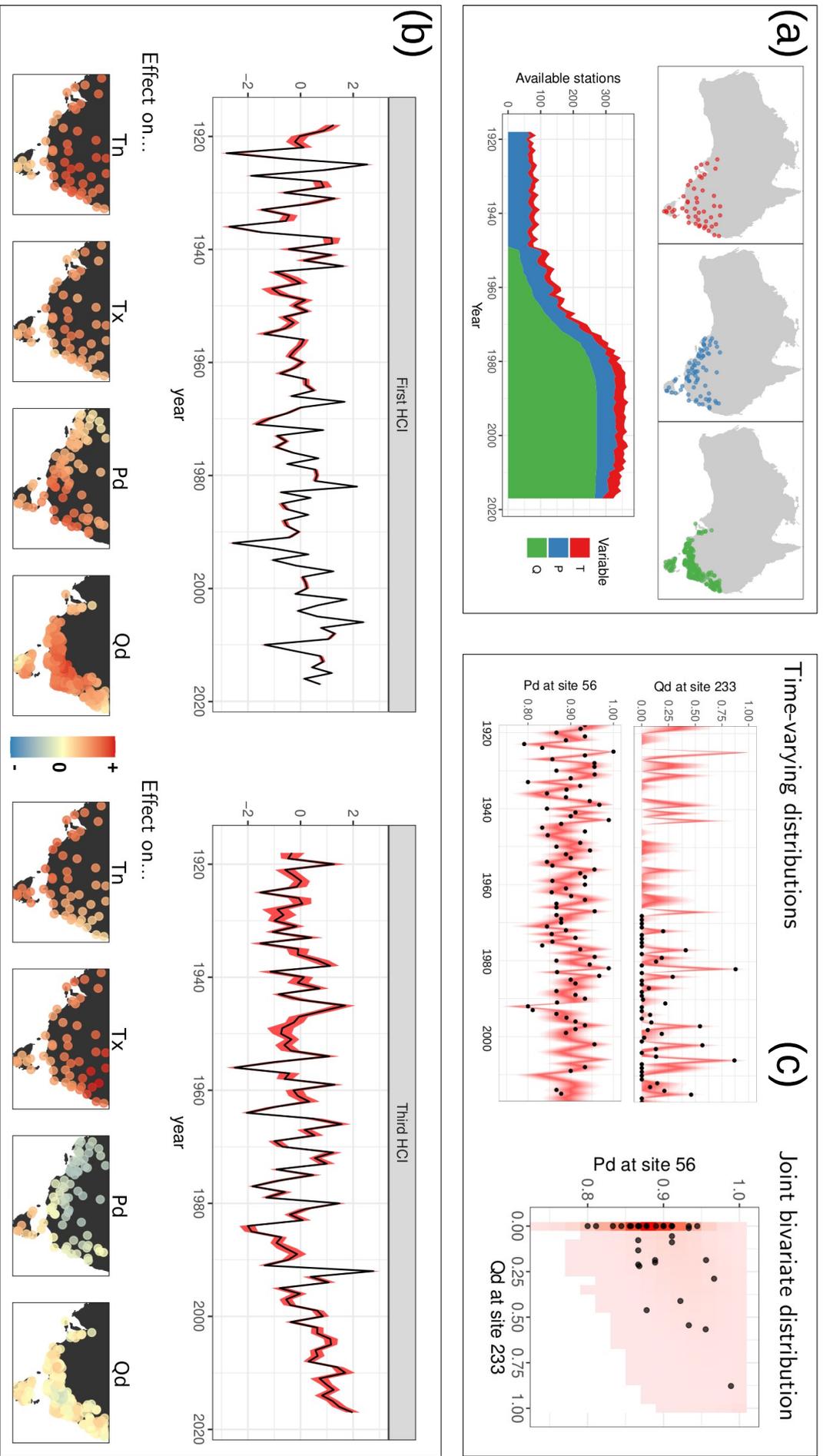
Figure 3.12: Hot-and-dry summers in Southeast Australia. (a) Location of Temperature (T), Precipitation (P) and streamflow (Q) stations and data availability; (b) Estimated first and third hidden climate indices (HCIs; posterior median and 90% credibility intervals) and their spatial effects on the four target variables; (c) Time-varying distributions for drought and precipitation durations at two sites, and resulting joint bivariate distribution. Colored bands on the left represent nested prediction intervals up to the 95% level, colored area on the right represents the bivariate pdf. Black dots are observed values. Modified from Renard et al. [2021]. For a sonified animation of this case study, see https://vimeo.com/60089870g.
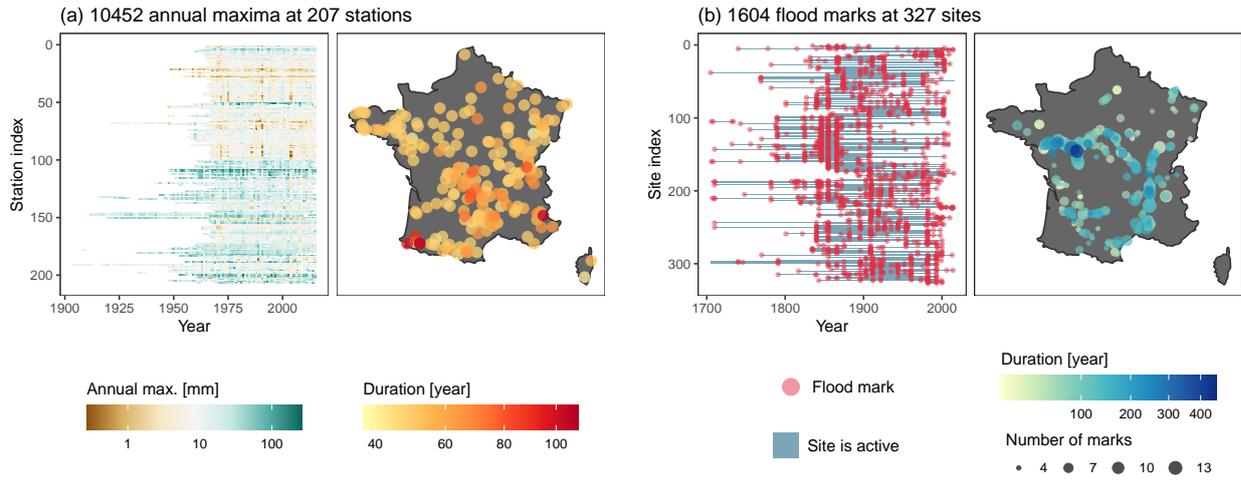
Figure 3.13: Two sources of data to characterize floods in France: (a) Annual maxima at hydrometric stations, 1904–2015; (b) Flood marks at flood sites, 1705–2015. Reproduced from Renard [2023]. For a sonified animation of this case study, see `https://vimeo.com/815008124`

whole period 1705–2015. The resulting probability maps, illustrated in Figure 3.14, provide quantitative information on the extent and spatial structure of ancient floods.

The third and last application of the HCI framework aimed at studying heavy precipitation and floods at the global scale [Renard et al., 2023], using HadEX [Donat et al., 2013, Dunn et al., 2020] and GSIM [Do et al., 2018, Gudmundsson et al., 2018] datasets (Figure 3.15). The analysis was based on the joint modeling of seasonal maxima of streamflow $Q(s,t)$ and precipitation $P(s,t)$ at more than 3,000 stations over a 100-year period. All series were transformed into nonexcedance probabilities using locally-estimated GEV distributions, making the Beta distribution a natural choice for modeling the resulting data belonging to the interval $(0;1)$:

$$
\begin{cases}
P\left(s,t\right) \sim Beta\left(\mu_P\left(s,t\right), \nu_P\left(s,t\right)\right); Q\left(s,t\right) \sim Beta\left(\mu_Q\left(s,t\right), \nu_Q\left(s,t\right)\right) \\
logit\left(\mu_P\left(s,t\right)\right) = \zeta_{\mu_P}(s) + \sum_{k=1}^{K} \lambda_{k,P}(s)\tau_k(t) + \sum_{k=1}^{K} \theta_{k,P}(s)\delta_k(t) \\
logit\left(\mu_Q\left(s,t\right)\right) = \zeta_{\mu_Q}(s) + \sum_{k=1}^{K} \lambda_{k,Q}(s)\tau_k(t) + \sum_{k=1}^{K} \theta_{k,Q}(s)\omega_k(t)
\end{cases}
\tag{3.15a}
$$

In this equation, the distribution $Beta(\mu,\nu)$ is parameterized so that μ is the mean and ν is a concentration parameter. As previously, a common set of HCIs $\tau_k(t)$ is used to describe the co-variability between P and Q. However, unlike in previous case studies, we also use specific HCIs $\delta_k(t)$ and $\omega_k(t)$ to describe the variability that is unique to P and Q. Estimation of this model allowed identifying strong and wide-ranging trends in precipitation-specific HCIs, while trends affecting flood-specific HCIs appeared weaker and with much more localized effects.

Another probabilistic model was also derived as part of this case study to link HCIs and large-scale atmospheric variables (pressure, wind, temperature) from the long 20CRv3 reanal-
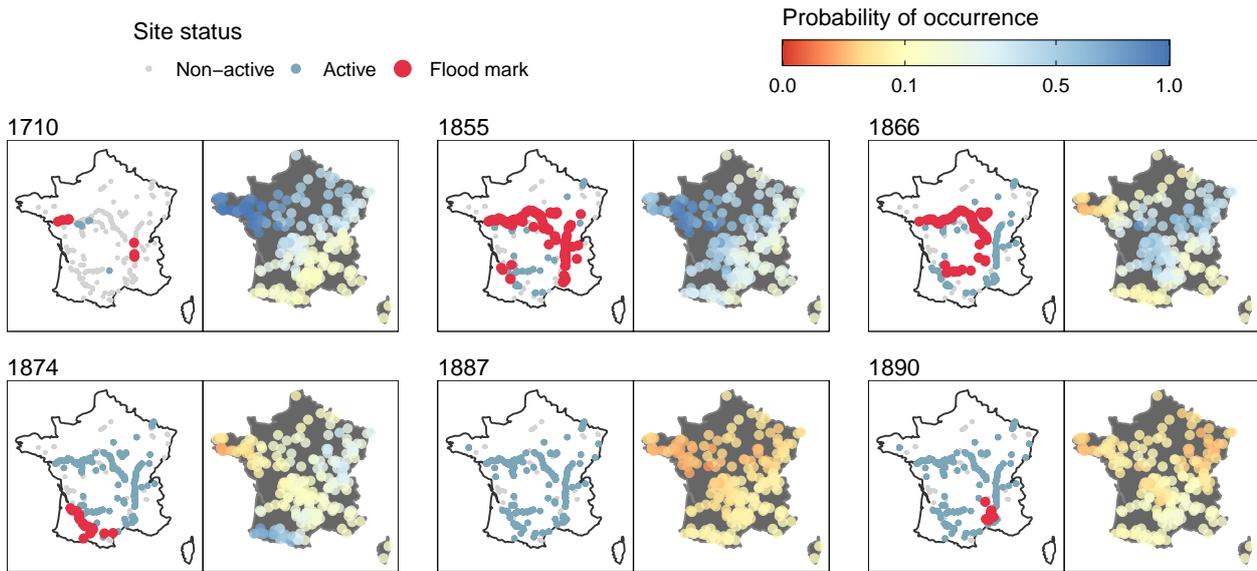
Figure 3.14: Probability maps estimated by transferring information from flood marks sites to hydrometric stations. For each selected year, the map on the left shows occurrences of flood marks and site status, and the map on the right shows the probability of exceeding a 10-year flood at stations. Reproduced from Renard [2023].
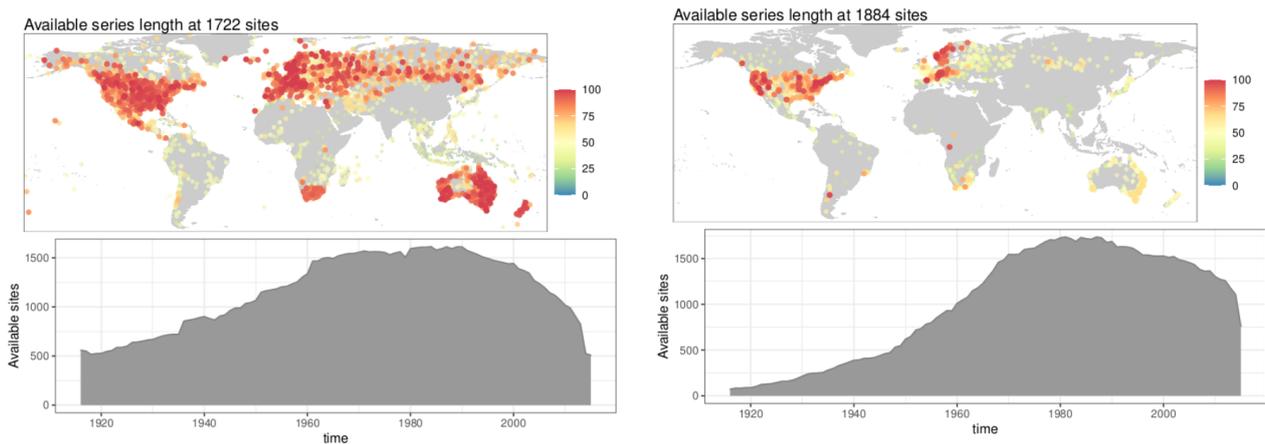
Figure 3.15: Location of stations used to study heavy precipitation (P, left) and floods (Q, right) at the global scale, and evolution of the number of available stations with time. For a sonified animation of this case study, see `https://vimeo.com/802751683`.
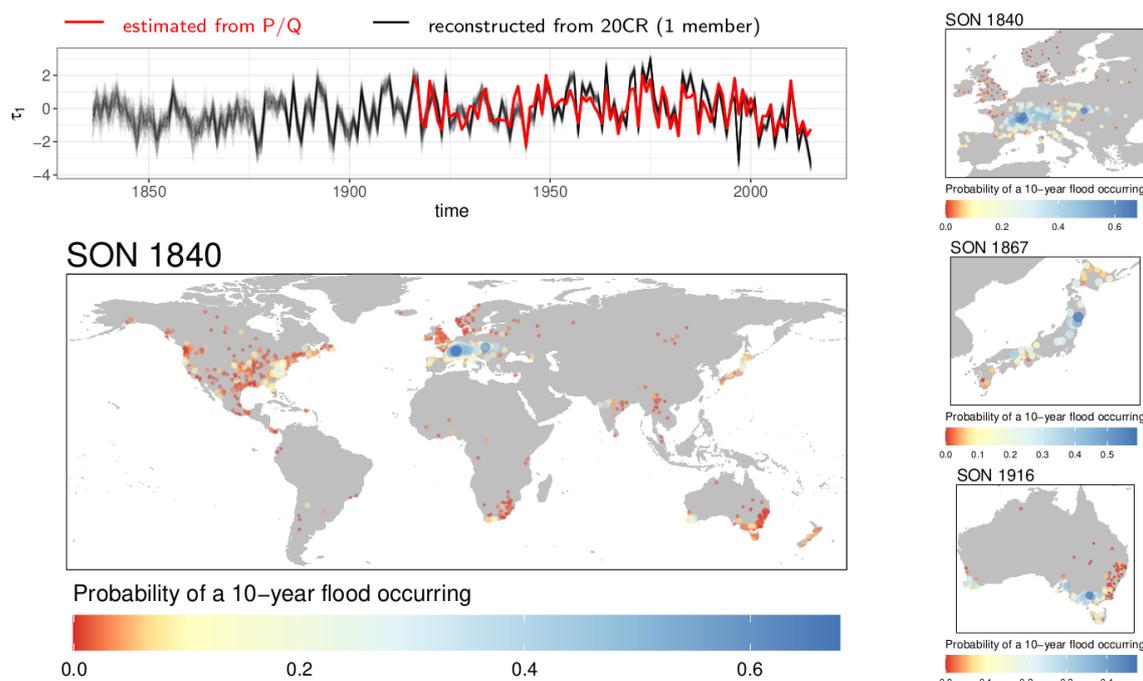
Figure 3.16: Reconstruction of flood probabilities since 1836. Top left: Hidden Climate Index (HCI) reconstructed from atmospheric data (pressure, wind, temperature) for the first component (SON season). The red line corresponds to the HCI estimated from floods and heavy precipitation data, each thin black line is a reconstruction based on one member of the 20CRv3 reanalysis. Maps show examples of reconstructed probability maps at the global scale or for specific regions.

ysis (1836-2015). This allowed reconstructing HCIs since 1836, and in turn estimating the probability of occurrence of floods and heavy precipitation at the global scale. This 180-year reconstruction is illustrated in Figure 3.16 with a few flood probability maps that highlight flood hot-spots and hot-moments in the distant past, well before the establishment of perennial monitoring networks.

3.5 Operational tools and applications

Several codes developed as part of the work described in this chapter have been released and are used either operationally or by research colleagues (Figure 3.17). The R package HydroPortail-Stats[6] [Renard, 2022] performs the statistical computations proposed by the HydroPortail[7], the French online portal to access data from the national hydrometric network. A 2-day training session is organized annually to explain and discuss these computations [Renard, 2017b]. The main task of this package is to estimate a user-selected distribution and the associated uncertainty using a data sample extracted from the HydroPortail database. A few other utilities are available, e.g. to implement basic trend or step-change tests. Several methods are available for parameter estimation (moments, L-moments, maximum likelihood, Bayesian) and uncer-

---

[6]`https://github.com/benRenard/HydroPortailStats`
[7]`https://hydro.eaufrance.fr/`

tainty quantification (bootstrap, parametric bootstrap, asymptotic Gaussian approximation, Bayesian), but preliminary analyses suggested that the combination L-moments + parametric boostrap was by far the most robust [Renard, 2016a,b]: its is hence used as the default choice in the HydroPortail. Note that the package can also be used off-line as a standard R package, which is in my eyes of great interest for hydrometry or hydrology services who wish to integrate it within their own workflow. As an illustration, the DREAL PACA[8] (regional environment agency in Southeastern France) is using the Bayesian estimation utility to constrain the GEV distribution for floods with some order-of-magnitude prior information for the shape parameter; the DREAL IdF[9] (Paris Region) is building a user interface around it; the DREAL Normandy[10] has integrated it in its statistical analyses. Finally, another R package called disTRIMbution[11] [Renard, 2021a] has been developed to estimate trimmed distributions, i.e. distributions having either reachable or unreachable bounds, but it is not yet used operationally as far as I know.

Before its integration into the HydroPortailStats R package, the Bayesian estimation of a distribution had been implemented in a tool called JBay[12], composed of a FORTRAN-based executable embedded in a Java user interface. This software was primarily used as part of a continuous education module on uncertainty quantification [Renard et al., 2017], but it has also been used for operational purposes, in particular for updating flood quantiles estimates in the Rhône catchment[13] [Bard and Lang, 2018]. A related code called HBay[14] was devoted to the integration of historical data, following the method developed in Neppel et al. [2010]. It was used for a New Zealand case study by Griffiths et al. [2017], and in a more operational context it has been used by colleagues from Cerema[15] in collaboration with DREAL ARA[16] (center-east France) to develop a systematic workflow for integrating historical information [Fromental and Bonnifait, 2018]. Note however that JBay and HBay will not be further developed in the future because they are superseded by the tool presented in the next paragraph.

STooDs[17] [probabilistic models for data varying in Space, Time or other Dimensions, Renard, 2021d] is a FORTRAN-based computational engine that can be considered as a synthesis of most developments described in this chapter. Indeed, STooDs allows building and estimating a wide range of models: models based on any distribution (continuous or discrete), multi-variable models, models varying in space, time or even in other dimensions (e.g. duration for rainfall), models using covariates, latent variable models, etc. It allows using censored historical data and enforcing identifiability constraints that are needed in Hidden Climate Indices models for instance. It comes with the R package RSTooDs[18] [Renard, 2021c] to help the user

---

[8]https://www.paca.developpement-durable.gouv.fr/

[9]https://www.drieat.ile-de-france.developpement-durable.gouv.fr/

[10]https://www.normandie.developpement-durable.gouv.fr/

[11]https://github.com/benRenard/disTRIMbution

[12]https://forge.irstea.fr/projects/thebay/news

[13]https://www.plan-rhone.fr/publications-131/actualisation-de-lhydrologie-des-crues-du-rhone-1865.html

[14]https://forge.irstea.fr/projects/thebay/files

[15]https://www.cerema.fr/

[16]https://www.auvergne-rhone-alpes.developpement-durable.gouv.fr/

[17]https://github.com/STooDs-tools/STooDs

[18]https://github.com/STooDs-tools/RSTooDs

Figure 3.17: Logos for some operational tools: HydroPortailStats, STooDs and JBay.

build the model and handle computations. The development of this tool is fairly recent and it has hence mostly been used in a research context so far: all HCI case studies described in section 3.4.2 rely on STooDs, and it has also been used by colleagues Michel Lang and Mathieu Lucas for analyzing long historical series [Lang et al., 2022, Lucas, 2023, Lucas et al., 2023].

3.6 Conclusions and perspectives

Describing the space-time variability of hydrologic variables with a probabilistic model is at the core of many important applications such as water resources management, flood and low flow frequency analysis, seasonal forecasting and more. My main contribution to this topic has been the development of models of increasing flexibility (and hence complexity) to move from the local to the regional scale, to include historical information, to enable the modeling of trends of other forms of temporal variability, and to enable the joint use of several variables. An underlying and continuous motivation behind these developments has been to make the best possible use of available data.

Several avenues for future works can be identified in terms of applications. First, a low-flow equivalent of the global floods and heavy precipitation analysis [Renard et al., 2023] could be implemented, building on a recent global low-flow trend analysis I participated to [Hodgkins et al., 2023]. This analysis would also be an opportunity to integrate data relevant to the wildfire hazard, which is highly related to droughts and for which new collaborations are emerging. More generally, the latest case studies I have performed point toward several directions I will follow in the future: the integration of several sources of information (multi-variable space-time station datasets, historical information, reanalyses or climate model outputs) and the move toward global-scale analyses. The latter direction will require participating to initiatives aimed at collating and documenting global streamflow datasets, and in particular to identify near-natural vs. heavily influenced catchments (see e.g. the ROBIN[19] initiative).

From a more methodological standpoint, I plan to explore in further depth the similarities existing between HCI models and several approaches from the statistical and machine learning literature. To start with, HCI models can be viewed as a continuous extension of Hidden

---

[19]https://www.ceh.ac.uk/our-science/projects/robin

Markov models: indeed, the temporal latent variable takes the form of a categorical variable describing the unknown climate state in the latter, with a hyperdistribution controlling the probability of transition between states [Thyer and Kuczera, 2003a,b, Bracken et al., 2016b]. Moreover, the potential of HCI models to indirectly represent dependence is highlighted by similarities with standard methods such as Principal Component Analysis (PCA) or Canonical Correlation Analysis (CCA), which have been re-interpreted as Gaussian models with unknown latent variables [Tipping and Bishop, 1999, Bach and Jordan, 2005, Klami et al., 2013]. In the case of a space-time dataset, these latent variables play a similar role to the HCIs (but within a restrictive Gaussian framework). Alternatively, many machine learning methods can be interpreted as latent variable models [Murphy, 2022, 2023]. Finally, the max-stable spatial model of Reich and Shaby [2012] also uses time-varying latent variables to describe spatial dependence between extreme data. Understanding all these approaches under the unifying view of probabilistic models may highlight important connections between seemingly unrelated methods. In turn, this may help identifying practical solutions found by others (e.g. the reparameterization proposed by Pourzanjani et al. [2020] to solve the non-identifiability of latent processes) or suggest new directions (e.g. using the HCI approach to derive spatial models for extremes).

Finally, several important challenges I shied away from are worth mentioning in this concluding section. The first one is the quantification, or even the modeling, of human influences. The work described in this chapter (with the exception of a small participation to the work of Dudley et al. [2019]) was indeed entirely based on selections of 'near-natural' stations where human influences could simply be discarded. I followed this admittedly not fully satisfying approach because I felt the modeling of human influences was too complex and case-specific to be applicable - for the moment at least - to the large-scale analyses I was heading toward. I also did not address the trend attribution problem, which is the logical step after trend detection. The first explanation is that in many cases trend analyses on hydrologic extremes revealed no strong trend, but it is a false excuse because attributing the absence of trends is actually a valid and interesting question[20]. A more fundamental reason is the complexity of the attribution process: it relies on a counterfactual world build by means of Global Climate Models (GCMs) and hydrologic models, and where anthropogenic greenhouse gas emissions have been removed [Hannart et al., 2016]. The conclusions of chapter 2 highlight many open questions on the quantification of uncertainties in hydrologic models that I feel should be addressed first. Moreover, the use of GCMs is also challenging due to the their poor ability to reproduce key variables for hydrology, especially extremes. Bias-correction is a possible approach to circumvent (but not solve) this difficulty. Another approach is to weight GCMs according to their ability to reproduce these key variables. We worked with Jean-Philippe Vidal on such an approach as part of the COMPLEX research project[21], but the papers were left unpublished [Renard and Vidal, 2017a,b]. For similar reasons, my participation to projects aimed at future projections of hydrologic variables has been very limited [Thirel et al., 2019].

---

[20]*"If precipitation extremes are increasing, why aren't floods?"* [Sharma et al., 2018]
[21]https://cordis.europa.eu/project/id/308601

General conclusion

Word cloud of this manuscript (bubble area is proportional to word frequency)

Summary of my research activity

My scientific activity revolves around the development of probabilistic models enabling the production of uncertain hydrologic predictions. This activity stretches from applied statistics to hydrology, but with a clear asymmetry: the former provides the toolbox while the latter is the objective. As a consequence, nearly all articles I contributed to are in hydro-climatology journals (with two exceptions in entomology [Sauvion et al., 2005] and civil engineering [Giacomini et al., 2009]), while I am still to co-author my first paper in a statistics journal. This

simply reflects the strategy I follow to build my contribution: using well-established statistical methods and adapting them to address hydrological questions.

My activity can be structured in three parts. The first one aims at developing methods to quantify the uncertainty affecting streamflow series, which form the raw material used in many hydrologic analyses (Chapter 1). The main outcome of this part is the development of the BaRatin method to estimate and propagate rating curve uncertainties. It is by now a mature tool that is largely used operationally and that provides, I hope, a satisfying solution to estimate and use simple stage-discharge rating curves and their uncertainties. The problem, however, is that such simple rating curves are the exception rather than the rule. To start with, virtually all stations on natural rivers are subject to rating shifts that need to be detected and managed. The PhD works of Darienzo [2020] and Mansanarez [2016] offered solutions to these challenges, but they still need to be streamlined and integrated into operational tools (this is work in progress). Moreover, complex rating curves are often required in situations where stage alone is not sufficient to determine discharge. Significant progresses have also been achieved in this area with the development of models accounting for hydraulic hysteresis, backwater or vegetation influences, while other cases like tide-influenced stations remain challenging. Work is still ongoing to either improve the operational transfer or further develop complex rating curve models. The BaM framework plays an important part in this work: it has indeed been developed with the double objective of being operational and of allowing other researchers to easily test new models. Indeed, the infrastructure for parameter estimation and uncertainty quantification built into BaM allows focusing on the model formulation rather than on technical difficulties.

The second part of my research activity is to develop a framework to quantify the different sources of uncertainty (input, response and structural) that affect hydrologic models (Chapter 2). This work is largely unfinished and constitutes the number one challenge I wish to tackle in the upcoming years. A key conclusion we achieved is that decomposing the total predictive uncertainty into its constitutive sources requires estimating data uncertainties prior to model calibration. Moreover, the accuracy of this decomposition entirely relies on the accuracy of the specified data uncertainties. This is one of the reasons why the quantification of streamflow uncertainty (Chapter 1) took such a prominent role in my research activity. Having achieved acceptable solutions for data uncertainties allows turning the attention to the remaining grand challenge, which is the treatment of structural uncertainty. Defining structural errors as the difference between true and simulated responses leads to characterizing structural uncertainty with an additive response error term. This is arguably the most natural approach but it is affected by the difficulty of deriving a realistic error model. More precisely, the partly systematic nature of structural errors needs to be recognized by means of a non-zero mean representing conditional bias. How to model this conditional bias, and with which conditioning variables (streamflow itself? inputs? state variables?), remain open questions. Model-internal approaches based on parameter or state stochastic perturbations represent a promising alternative because even perturbations arising from a very simple probabilistic model transit through the dynamic hy-

drologic model and may result in highly-structured perturbations in the streamflow response. However, many open questions also remain regarding the choice of the parameter or state to be made stochastic, the temporal scale at which this stochasticity operates, etc. Finally, the adaptation of the work made so far to the case of distributed models will represent an important part of my future research activity. This is in line with the activity of the research team RHAX[22] (hydro-meteorological risks) I joined in June 2022 and which develops the SMASH[23] distributed model. This will also allow considering other types of model such as hydraulic models, with possible applications in hydrometry or flood risk management, just to name a few. This new direction raises many questions in terms of modeling the various sources of uncertainty in a spatially distributed context, and may require resorting to specific techniques such as surrogate models [Sudret, 2008, Rouzies et al., 2023] to make computations feasible.

The third part of my research activity is the development of probabilistic models to describe the space-time variability of potentially several hydrologic variables (Chapter 3). This very general description hides a variety of more specific purposes or topics that provided the original motivation for many of the developed models: flood frequency analysis, regionalization, integration of historical data, detecting and modeling trends or climate effects in hydrologic variables, etc. A general modeling framework progressively emerged from these elementary developments. In particular, the PhD work of Sun [2013] allowed integrating techniques for time-varying models and for regionalization into a unique space-time modeling framework. The development of Hidden Climate Indices models and of the associated computing code STooDs went one step further by also enabling the joint modeling of several variables and the use of censored data. This provides a flexible platform to integrate several sources of information such as stations datasets, historical information, climate model outputs, etc. In the upcoming years, I wish to focus on applications of this platform rather than on new methodological developments. A first possible direction would be to pursue the analysis of multi-century daily stage series that exist in French hydrometric archives. These series cannot be transformed into a streamflow times series of homogeneous quality and therefore need to be analyzed with specific methods, along the lines of the PhD work of Lucas [2023]. Hopefully, such analyses may highlight the value of these long series when treated with adequate methods. In the longer term, the joint analysis of long series, flood marks and hydroclimatic reconstructions may shed more light on the historical variability of hydrologic regimes in France over the last few centuries. A second direction I wish to pursue is the analysis of low flows at the global scale, in relation with other variables that drive the wildfire hazard (humidity, temperature, etc.). The research unit RECOVER[24] (risks, ecosystems, vulnerability, environment, resilience) I now belong to includes a research team specialized on the wildfire risk[25], which offers promising collaboration opportunities.

---

[22] https://www6.paca.inrae.fr/recover/Qui-sommes-nous/Nos-equipes/RHAX
[23] https://smash.recover.inrae.fr/
[24] https://www6.paca.inrae.fr/recover
[25] https://www6.paca.inrae.fr/recover/Qui-sommes-nous/Nos-equipes/EMR

Miscellaneous methodological thoughts

The word cloud of this document shown at the beginning of the chapter is of interest because it offers an opportunity to reflect on some important aspects of the work described in this manuscript.

The most frequently used word is *model*, which should come at no surprise since the whole manuscript revolves around the search of models that could realistically have generated the data. Unlike rating curve or hydrologic models, probabilistic models are not deterministic, but despite this fundamental difference the underlying strategies are remarkably similar: (i) formulate hypotheses and formalize them as mathematical equations; (ii) estimate the model based on available data; (iii) assess the adequacy of the initial hypotheses, and if needed (iv) modify them by going back to point (i) to formulate alternative hypotheses. Once hypotheses are considered satisfying, the model can be used to make uncertain predictions.

During discussions with colleagues, I found it interesting that different people were more comfortable with either a model view or an algorithmic view of statistical methods. I am definitely of the first kind: I am lost without a model, and I struggle to develop intuitions on algorithms - but this seems to be nothing more than a personal preference. This is probably the reason why I have not made an heavy use of methods from the machine learning (ML) or artificial intelligence (AI) communities so far, as I generally came across algorithmic-oriented presentations of their functioning (and this is also the way they were taught at the University). However, the books by Murphy [2022, 2023] show that there is a probabilistic model (often a Gaussian one) behind virtually all ML/AI methods. This is quite enlightening to me, and it is also an important vector of innovation - for instance, knowing the Gaussian model behind Principal Component Analysis makes it easier to adapt it to 0/1 occurrence data by switching to a Bernoulli model. In fact I may have used ML/AI methods without even knowing it: for instance, the HCI approach of Chapter 3 can be interpreted as a non-Gaussian encoder-decoder in ML terms.

The word *data* only comes in tenth position, but combined with words like *streamflow* and *series* it goes back to the top rank. Again, this is not surprising for an hydrology manuscript heavily based on applied statistics - a different outcome would have been a disgrace! An important part of my work has been to quantify uncertainties in streamflow data, but in doing so my aim has been to serve them, not to raise suspicion. Quantifying uncertainty allows escaping this binary view that a data is either good enough to make the cut or has to be dropped. It also places responsibility to account for this uncertainty on the modeler or the decision-maker, not implicitly on the data producer. In my opinion, uncertainty is not a problem - ignoring it is.

Another motivation underpinning the probabilistic models developed in this manuscript is to adapt them to the properties of available data, and not the other way around. For instance, restricting trend analyses to a common period is a strategy I've never come to terms with (although I did implement it): it represents a huge waste of data mostly justified by the inadequacy of a linear trend model. Likewise, gridded datasets make the modeler's life

easier, but in some cases I feel that going back to station data and adapting the model to its idiosyncrasies (missing values, varying availability, irregular spatial sampling) is a better solution. This is the case for extremes in particular, since gridding induces some form of spatial smoothing that may strongly affect extremes' properties. Overall, I consider that the effort made by a modeler like me to adapt the model to the data is negligible compared with the efforts made by generations of data producers to get the data in the first place - it is hence a duty to make this effort if it allows making a better use of existing data.

The word cloud is also of interest for the words that do *not* appear as prominently as could have been expected. In particular, the word *Bayesian* is surprisingly hard to find given that Bayesian estimation was used throughout this manuscript. It is admittedly an approach I enjoy and enthusiastically advertise [Renard et al., 2013b, Kuczera et al., 2017], because when it comes to uncertainty quantification, I find it very practical and easier to understand than frequentist approaches. Typically, I find MCMC sampling from a posterior distribution easier to understand and manipulate than e.g. asymptotic approximations of the sampling distribution of estimates. Alternatively, properly defining the concept of a predictive distribution is straightforward in the Bayesian paradigm, but surprisingly tricky in the frequentist one [see discussion and references in Renard et al., 2013a]. However, this is again a personal preference, and in any case I see the Bayesian approach as a means (estimation) to an end (the probabilistic model), not an objective in its own right. In fact, I believe many of the analyses presented in this manuscript could have been performed with an alternative estimation approach - but almost none could have been performed without a probabilistic model.

Words attempting to define a taxonomy of uncertainty, such as '*epistemic*' or '*aleatoric*' (among many others), do not appear at all in the word cloud. This may seem surprising given that many texts treating with uncertainty start by defining some form of classification into various uncertainty types. Moreover, there exists a long-lasting debate in hydrology about the adequacy of probabilistic models to describe epistemic uncertainty [e.g. Di Baldassarre et al., 2016, Nearing et al., 2016, and references therein]. In all honesty, I never fully understood any of these general-purpose taxonomies[26], and the arguments used in the aforementioned debate are largely beyond my grasp. I am therefore following a pragmatic and technically-oriented road: I take it for granted that probabilistic modeling is an adequate way to represent uncertainty, whatever its type, and I am trying to address some of the technical challenges rising up on the way. At least this approach has not been proven wrong yet, and if one day it does, it will still have been an interesting and hopefully useful journey, and I will have made a living out of it!

Concluding remarks on doing research

This manuscript ends with a few thoughts on the way I conducted my research activities in the past and what this way might look like in the future. I am purely expressing personal views and preferences here, and I am in no way suggesting that this is an optimal approach everyone should

---

[26]It seems I am not the only one, see e.g. discussions in
`https://statmodeling.stat.columbia.edu/2022/02/03/epistemic-and-aleatoric-uncertainty`

follow - in fact, quite the contrary. Observing and discussing with colleagues made it clear that there exists a huge diversity of research paths, with different people putting different emphasis on various facets of the researcher's work (performing analyses and experiments, supervising, teaching, reviewing, managing, consulting, doing operational transfer, communicating, taking scientific or administrative leadership roles, etc.). This diversity of positioning is a necessity as it is impossible for a single individual to tick all those boxes, but it is also an asset in my opinion and should therefore be preserved and respected.

It is very important for me to keep doing technical work, and I cannot conceive a future where I would stop coding, analyzing datasets, etc. This implies a need to find a balance between supervising and doing analyses myself, and I am hence regularly making conscious (and maybe selfish) decisions to keep tasks for myself. This also implies the need to consciously manage time. I had a few difficulties in this respect at the beginning of my research career, mostly because I ended up being overwhelmed by too many supervising and project commitments. I have since learned to say no and to live with the frustration of giving up on interesting things - which I find much easier to manage than the frustration of having to do things superficially.

Coding is an important part of my activity, first because I really enjoy it, and second because I find it extremely useful and efficient in structuring ideas. In turn, this helps moving from case-specific approaches to more general modeling frameworks. In my experience, there is a surprisingly strong link between the framework's equations and the code that implements them. Every attempt I made to implement a general model without having properly worked out the equations resulted in complete failure. Alternatively, implementation difficulties often resulted in the realization that there was in fact something wrong in some equation. Equations and code are therefore two sides of the same coin, and I'm only comfortable when I have achieved both a coherent set of equations and a robust working code - which, in general, takes quite a few iterations.

Operational transfer is another important part of my activity, and it is based on two main pillars: software development and continuous education. The former is the natural end product of the codes discussed in the previous paragraph, but developing a computing code and developing a graphical user interface are two different things. I took part to the development of the BaRatinAGE software and, while it was a very interesting and rewarding experience, it was also very time-consuming. I will probably restrict to developing computing codes and non-graphical interfaces such as BaM, STooDs or HydroPortailStats in the future. Continuous education is in my eyes an efficient way to make operational practice closer to the current scientific state of the art, especially when accompanied with operational software. It is also a very enjoyable and rewarding activity given the enthusiasm of participants, and it allows identifying salient operational challenges and getting feedback on released tools. Finally, another 'non-research' activity I recently took up is scientific communication, with the creation of a blog[27] based on the visualization and sonification[28] of the hydro-climatic datasets I use as part of my activity [Renard, 2021b,e, Renard and Le Bescond, 2022].

---

[27]`https://globxblog.github.io/`
[28]`https://blogs.egu.eu/divisions/hs/2023/02/01/`

Finally, I think scientific research is a very rewarding and fulfilling activity, and I am trying my best to preserve the positive sides of it and to disregard the occasional annoyances and frustrations. Such positive aspects include the continuous opportunity to learn, the freedom to take directions out of curiosity and interest, the sheer fun of the problem-solving process and even the occasional 'eureka' moments. Most importantly, I highly value the collective dimension of research and the diversity that comes with it: while I am the sole author of this manuscript, it is the result of collaborations with many colleagues having different backgrounds, interests, cultures and personalities. It therefore seems fitting that I conclude this manuscript by thanking them all.

Bibliography

K.-H. Ahn. Coupled annual and daily multivariate and multisite stochastic weather generator to preserve low- and high-frequency variability to assess climate vulnerability. *Journal of Hydrology*, 2020. doi: 10.1016/j.jhydrol.2019.124443.

S. Ahrendt, A. R. Horner-Devine, B. D. Collins, J. A. Morgan, and E. Istanbulluoglu. Channel Conveyance Variability can Influence Flood Risk as Much as Streamflow Variability in Western Washington State. *Water Resources Research*, 2022. doi: 10.1029/2021WR031890.

P. Ailliot, D. Allard, V. Monbet, and P. Naveau. Stochastic weather generators: an overview of weather type models. *Journal de la société française de statistique*, 156(1):101–113, 2015. URL http://www.numdam.org/item/JSFS_2015__156_1_101_0/.

N. K. Ajami, Q. Y. Duan, and S. Sorooshian. An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resour. Res.*, 43(1), 2007.

D. Alliau, J. De Saint Seine, M. Lang, E. Sauquet, and B. Renard. Etude du risque d'inondation d'un site industriel par des crues extrêmes: de l'évaluation des valeurs extrêmes aux incertitudes hydrologiques et hydrauliques. *La houille blanche*, (2):70–77, 2015.

V. Andréassian, A. Oddos, C. Michel, F. Anctil, C. Perrin, and C. Loumagne. Impact of spatial aggregation of inputs and parameters on the efficiency of rainfall-runoff models: A theoretical study using chimera watersheds. *Water Resources Research*, 2004. doi: 10.1029/2003WR002854.

P. Arnaud and J. Lavabre. Coupled rainfall model and discharge model for flood frequency estimation. *Water Resour. Res.*, 2002.

P. Arnaud, P. Cantet, and Y. Aubert. Relevance of an at-site flood frequency analysis method for extreme events based on stochastic simulation of hourly rainfall. *Hydrological Sciences Journal*, 2016. doi: 10.1080/02626667.2014.965174.

N. Arnell and S. Gosling. The impacts of climate change on river flood risk at the global scale. *Climatic Change*, 2014. doi: 10.1007/s10584-014-1084-5.

S. K. Aryal, B. C. Bates, E. P. Campbell, Y. Li, M. J. Palmer, and N. R. Viney. Characterizing and modeling temporal and spatial trends in rainfall extremes. *Journal of Hydrometeorology*, 10, 2009. doi: 10.1175/2008JHM1007.1.

F. R. Bach and M. I. Jordan. A probabilistic interpretation of canonical correlation analysis, 2005. URL https://www.stat.berkeley.edu/~jordan/688.pdf.

H. M. Badjana. *River Basin's Assessment and Hydrologic Processes Modeling for Integrated Land and Water Resources Management in West Africa.* PhD thesis, University of Abomey-Calavi, West African Science Service Centre on Climate Change and Adapted Land Use, 2018.

H. M. Badjana, B. Renard, J. Helmschrot, K. S. Edjamé, A. Afouda, and K. Wala. Bayesian trend analysis in annual rainfall total, duration and maximum in the kara river basin (west africa). *Journal of Hydrology: Regional Studies*, 2017. doi: https://doi.org/10.1016/j.ejrh. 2017.08.009.

R. Barbero, J. T. Abatzoglou, E. A. Steel, and N. K Larkin. Modeling very large-fire occurrences over the continental united states from weather and climate forcing. *Environmental Research Letters*, 9, 2014. doi: 10.1088/1748-9326/9/12/124009.

A. Bard and M. Lang. Actualisation de l'hydrologie des crues du Rhône. Technical report, HY-DRO CONSULTANT, Irstea, 2018.

A. Bard, B. Renard, and M. Lang. Floods in the alpine areas of europe. In Z. W. Kundzewicz, editor, *Changes in flood risk in Europe.* IAHS Press, 2012a.

A. Bard, B. Renard, and M. Lang. Observed trends in the hydrologic regime of alpine catchments. *La houille blanche*, 2012b. doi: DOI10.1051/lhb/2012006.

A. Bard, B. Renard, M. Lang, I. Giuntoli, J. Korck, G. Koboltschnig, M. Janža, M. d'Amico, and D. Volken. Trends in the hydrologic regime of alpine rivers. *Journal of Hydrology*, 2015. doi: http://dx.doi.org/10.1016/j.jhydrol.2015.07.052.

G. Benito, M. Lang, M. Barriendos, M. C. Llasat, F. Francés, T. Ouarda, V. Thorndycraft, Y. Enzel, A. Bardossy, D. Coeur, and B. Bobée. Use of systematic, palaeoflood and historical data for the improvement of flood risk estimation. review of scientific methods. *Natural Hazards*, 2004. doi: 10.1023/B:NHAZ.0000024895.48463.eb.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological*, 1995.

M. Bertola, A. Viglione, D. Lun, J. Hall, and G. Blöschl. Flood trends in Europe: are changes in small and big floods different? *Hydrology and Earth System Sciences*, 2020. doi: 10.5194/ hess-24-1805-2020.

K. J. Beven. On the concept of model structural error. *Water Science and Technology*, 52(6), 2005.

K. J. Beven. A manifesto for the equifinality thesis. *Journal of Hydrology*, 320, 2006.

K. J. Beven and A. M. Binley. The future of distributed hydrological models: model calibration and uncertainty prediction. *Hydrological Processes*, 6, 1992.

K. J. Beven, P. J. Smith, and J. E. Freer. So just why would a modeller choose to be incoherent? *Journal of Hydrology*, 354(1-4), 2008.

S. Biancamaria, D. P. Lettenmaier, and T. M. Pavelsky. The SWOT Mission and Its Capabilities for Land Hydrology. *Surveys in Geophysics*, 2016. doi: 10.1007/s10712-015-9346-y.

J. Blanchet and A. C. Davison. Spatial Modeling of Extreme Snow Depth. *The Annals of Applied Statistics*, 2011. doi: 10.2307/23069351.

B. Bobée and P. F. Rasmussen. Recent advances in flood frequency analysis. *Reviews of Geophysics*, 1995. doi: 10.1029/95RG00287.

A. Botto, D. Ganora, P. Claps, and F. Laio. Technical note: Design flood under hydrological uncertainty. *Hydrology and Earth System Sciences*, 21, 2017. doi: 10.5194/hess-21-3353-2017.

C. Bracken, B. Rajagopalan, L. Cheng, W. Kleiber, and S. Gangopadhyay. Spatial bayesian hierarchical modeling of precipitation extremes over a large domain. *Water Resources Research*, 52, 2016a. doi: 10.1002/2016WR018768.

C. Bracken, B. Rajagopalan, and C. Woodhouse. A bayesian hierarchical nonhomogeneous hidden markov model for multisite streamflow reconstructions. *Water Resources Research*, 52, 2016b. doi: doi:10.1002/2016WR018887.

C. Bracken, K. D. Holman, B. Rajagopalan, and H. Moradkhani. A bayesian hierarchical approach to multivariate nonstationary hydrologic frequency analysis. *Water Resources Research*, 2018. doi: 10.1002/2017WR020403.

L. M. Braman, M. K. van Aalst, S. J. Mason, P. Suarez, Y. Ait-Chellouche, and A. Tall. Climate forecasts in disaster management: Red Cross flood operations in West Africa, 2008. *Disasters*, 2013. doi: 10.1111/j.1467-7717.2012.01297.x.

R. L. Bras and I. Rodriguez-Iturbe. *Random Functions and Hydrology*. Addison-Wesley, 1984.

I. Braud, H. Roux, S. Anquetin, M.-M. Maubourguet, C. Manus, P. Viallet, and D. Dartus. The use of distributed hydrological models for the Gard 2002 flash flood event: Analysis of associated hydrological processes. *Journal of Hydrology*, 2010. doi: 10.1016/j.jhydrol.2010.03.033.

I. Braud, P. A. Ayral, C. Bouvier, F. Branger, G. Delrieu, J. Le Coz, G. Nord, J. P. Vandervaere, S. Anquetin, M. Adamovic, J. Andrieu, C. Batiot, B. Boudevillain, P. Brunet, J. Carreau, A. Confoland, J. F. Didon-Lescot, J. M. Domergue, J. Douvinet, G. Dramais, R. Freydier, S. Gérard, J. Huza, E. Leblois, O. Le Bourgeois, R. Le Boursicaud, P. Marchand, P. Martin, L. Nottale, N. Patris, B. Renard, J. L. Seidel, J. D. Taupin, O. Vannier, B. Vincendon, and A. Wijbrans. Multi-scale hydrometeorological observation and modelling for flash flood understanding. *Hydrology and Earth System Sciences*, 2014. doi: 10.5194/hess-18-3733-2014.

R. Brázdil, Z. W. Kundzewicz, and G. Benito. Historical hydrology for studying flood risk in europe. *Hydrological Sciences Journal*, 51:739–764, 2006. doi: 10.1623/hysj.51.5.739.

N. Bulygina and H. Gupta. Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation. *Water Resour. Res.*, (12), 2009. doi: 10.1029/2007wr006749.

D. H. Burn. Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, 1990. doi: 10.1029/WR026i010p02257.

B. Camenen, G. Dramais, J. Le Coz, T. D. Ho, N. Gratiot, and S. Piney. Estimation d'une courbe de tarage hauteur-dénivelée-débit pour une rivière influencée par la marée. *La Houille Blanche*, 2017. doi: 10.1051/lhb/2017039.

B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76, 2017. doi: 10.18637/jss.v076.i01.

K. Cawley. NEON User Guide to Stream Discharge. Technical report, National Ecological Observatory Network, 2018.

J. Ching, J. Herwehe, and J. Swall. On joint deterministic grid modeling and sub-grid variability conceptual framework for model evaluation. *Atmospheric Environment*, 2006. doi: 10.1016/j.atmosenv.2006.01.021.

M. P. Clark and A. G. Slater. Probabilistic quantitative precipitation estimation in complex terrain. *Journal of Hydrometeorology*, 7(1), 2006.

M. P. Clark, D. Kavetski, and F. Fenicia. Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resour. Res.*, 2011. doi: 10.1029/2010wr009827.

H. L. Cloke, F. Pappenberger, S. J. van Andel, J. Schaake, J. Thielen, and M. Ramos. Hydrological ensemble prediction systems. *Hydrological Processes*, 2013. doi: 10.1002/hyp.9679.

S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer-Verlag, London, 2001.

S. Coles, L. R. Pericchi, and S. Sisson. A fully probabilistic approach to extreme rainfall modelling. *Journal of Hydrology*, 273(1-4), 2003.

Commonwealth of Australia. *Australian Rainfall and Runoff: A Guide to Flood Estimation.* Commonwealth of Australia, 2019.

D. Cooley, D. Nychka, and P. Naveau. Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840, 2007.

E. Coughlan de Perez, B. van den Hurk, M. K. van Aalst, I. Amuron, D. Bamanya, T. Hauser, B. Jongma, A. Lopez, S. Mason, J. Mendler de Suarez, F. Pappenberger, A. Rueth, E. Stephens, P. Suarez, J. Wagemaker, and E. Zsoter. Action-based flood forecasting for triggering humanitarian action. *Hydrology and Earth System Sciences*, 2016. doi: 10.5194/hess-20-3549-2016.

J. M. Cunderlik and D. H. Burn. Non-stationary pooled frequency analysis. *Journal of Hydrology*, 276, 2003.

C. Cunnane. Methods and merits of regional flood frequency analysis. *Journal of Hydrology*, 1988. doi: 10.1016/0022-1694(88)90188-6.

T. Dalrymple. Flood frequency analyses. *Water-supply paper 1543-A*, 1960.

R. Dankers, N. W. Arnell, D. B. Clark, P. D. Falloon, B. M. Fekete, S. N. Gosling, J. Heinke, H. Kim, Y. Masaki, Y. Satoh, T. Stacke, Y. Wada, and D. Wisser. First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble. *Proceedings of the National Academy of Sciences*, 2014. doi: 10.1073/pnas.1302078110.

M. Darienzo. *Real-time Bayesian estimation of dynamic stage-discharge models.* PhD thesis, Irstea, Université de Lyon, Université Grenoble Alpes, 2020.

M. Darienzo, B. Renard, J. Le Coz, and M. Lang. Detection of Stage-Discharge Rating Shifts Using Gaugings: A Recursive Segmentation Procedure Accounting for Observational and Model Uncertainties. *Water Resources Research*, 2021. doi: 10.1029/2020WR028607.

L. De Haan and J. De Ronde. Sea and wind: Multivariate extremes at work. *Extremes*, 1, 1998.

S. De Iaco. A new space–time multivariate approach for environmental data analysis. *Journal of Applied Statistics*, 38, 2011. doi: 10.1080/02664763.2011.559206.

G. Delrieu, A. Wijbrans, B. Boudevillain, D. Faure, L. Bonnifait, and P.-E. Kirstetter. Geostatistical radar–raingauge merging: A novel method for the quantification of rain estimation accuracy. *Advances in Water Resources*, 2014. doi: 10.1016/j.advwatres.2014.06.005.

A. Despax, J. Le Coz, A. Hauet, D. S. Mueller, F. L. Engel, B. Blanquart, B. Renard, and K. A. Oberg. Decomposition of Uncertainty Sources in Acoustic Doppler Current Profiler Streamflow Measurements Using Repeated Measures Experiments. *Water Resources Research*, 2019.

N. Devineni, U. Lall, N. Pederson, and E. Cook. A tree-ring-based reconstruction of delaware river basin streamflow using hierarchical bayesian regression. *Journal of Climate*, 26, 2013. doi: 10.1175/JCLI-D-11-00675.1.

G. Di Baldassarre, L. Brandimarte, and K. Beven. The seventh facet of uncertainty: wrong assumptions, unknowns and surprises in the dynamics of human–water systems. *Hydrological Sciences Journal*, 2016. doi: 10.1080/02626667.2015.1091460.

H. X. Do, L. Gudmundsson, M. Leonard, and S. Westra. The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata. *Earth Syst. Sci. Data*, 2018. doi: 10.5194/essd-10-765-2018.

M. G. Donat, L. V. Alexander, H. Yang, I. Durre, R. Vose, R. J. H. Dunn, K. M. Willett, E. Aguilar, M. Brunet, J. Caesar, B. Hewitson, C. Jack, A. M. G. Klein Tank, A. C. Kruger, J. Marengo, T. C. Peterson, M. Renom, C. Oria Rojas, M. Rusticucci, J. Salinger, A. S. Elrayah, S. S. Sekele, A. K. Srivastava, B. Trewin, C. Villarroel, L. A. Vincent, P. Zhai, X. Zhang, and S. Kitching. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *Journal of Geophysical Research: Atmospheres*, 2013. doi: 10.1002/jgrd.50150.

Q. Duan, N. K. Ajami, X. Gao, and S. Sorooshian. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources*, 30(5), 2007.

R. W. Dudley, G. A. Hodgkins, M. R. McHale, M. J. Kolian, and B. Renard. Trends in snowmelt-related streamflow timing in the conterminous united states. *Journal of Hydrology*, 2017. doi: http://dx.doi.org/10.1016/j.jhydrol.2017.01.051.

R. W. Dudley, R. M. Hirsch, S. A. Archfield, A. G. Blum, and B. Renard. Low streamflow trends at human-impacted and reference basins in the United States. *Journal of Hydrology*, 2019. doi: https://doi.org/10.1016/j.jhydrol.2019.124254.

R. J. H. Dunn, L. V. Alexander, M. G. Donat, X. Zhang, M. Bador, N. Herold, T. Lippmann, R. Allan, E. Aguilar, A. A. Barry, M. Brunet, J. Caesar, G. Chagnaud, V. Cheng, T. Cinco, I. Durre, R. Guzman, T. M. Htay, W. M. Wan Ibadullah, M. K. I. Bin Ibrahim, M. Khoshkam, A. Kruger, H. Kubota, T. W. Leng, G. Lim, L. Li-Sha, J. Marengo, S. Mbatha, S. McGree, M. Menne, M. Milagros Skansi, S. Ngwenya, F. Nkrumah, C. Oonariya, J. D. Pabon-Caicedo, G. Panthou, C. Pham, F. Rahimzadeh, A. Ramos, E. Salgado, J. Salinger, Y. Sané, A. Sopaheluwakan, A. Srivastava, Y. Sun, B. Timbal, N. Trachow, B. Trewin, G. Schrier, J. Vazquez-Aguirre, R. Vasquez, C. Villarroel, L. Vincent, T. Vischel, R. Vose, and M. N. Bin Hj Yussof. Development of an Updated Global Land In Situ-Based Data Set of Temperature and Precipitation Extremes: HadEX3. *Journal of Geophysical Research: Atmospheres*, 2020. doi: 10.1029/2019JD032263.

M. Durand, C. J. Gleason, P. A. Garambois, D. Bjerklie, L. C. Smith, H. Roux, E. Rodriguez, P. D. Bates, T. M. Pavelsky, J. Monnier, X. Chen, G. Di Baldassarre, J. Fiset, N. Flipo, R. P.

d. M. Frasson, J. Fulton, N. Goutal, F. Hossain, E. Humphries, J. T. Minear, M. M. Mukolwe, J. C. Neal, S. Ricci, B. F. Sanders, G. Schumann, J. E. Schubert, and L. Vilmin. An intercomparison of remote sensing river discharge estimation algorithms from measurements of river height, width, and slope. *Water Resources Research*, 2016. doi: 10.1002/2015WR018434.

M. Durand, C. J. Gleason, T. M. Pavelsky, R. Prata de Moraes Frasson, M. Turmon, C. H. David, E. H. Altenau, N. Tebaldi, K. Larnier, J. Monnier, P. O. Malaterre, H. Oubanas, G. H. Allen, B. Astifan, C. Brinkerhoff, P. D. Bates, D. Bjerklie, S. Coss, R. Dudley, L. Fenoglio, P. Garambois, A. Getirana, P. Lin, S. A. Margulis, P. Matte, J. T. Minear, A. Muhebwa, M. Pan, D. Peters, R. Riggs, M. S. Sikder, T. Simmons, C. Stuurman, J. Taneja, A. Tarpanelli, K. Schulze, M. J. Tourian, and J. Wang. A Framework for Estimating Global River Discharge From the Surface Water and Ocean Topography Satellite Mission. *Water Resources Research*, 2023. doi: 10.1029/2021WR031614.

A. V. Dyrrdal, A. Lenkoski, T. L. Thorarinsdottir, and F. Stordal. Bayesian hierarchical modeling of extreme hourly precipitation in norway. *Environmetrics*, 26, 2015. doi: 10.1002/env.2301.

P. S. Eagleson. Dynamics of flood frequency. *Water Resources Research*, 1972. doi: 10.1029/WR008i004p00878.

K. Engeland, B. Renard, I. Steinsland, and S. Kolberg. Evaluation of statistical models for forecast errors from the HBV model. *Journal of Hydrology*, 384:142–155, 2010.

K. Engeland, M. Borga, J.-D. Creutin, B. François, M.-H. Ramos, and J.-P. Vidal. Space-time variability of climate variables and intermittent renewable electricity production – a review. *Renewable and Sustainable Energy Reviews*, 79:600–617, 2017. doi: 10.1016/j.rser.2017.05.046.

O. P. Faugeras. Inference for copula modeling of discrete data: a cautionary tale and some facts. *Dependence Modeling*, 5, 2017. doi: 10.1515/demo-2017-0008.

A. C. Favre, S. El Adlouni, L. Perreault, N. Thiemonge, and B. Bobee. Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.*, 40, 2004.

B. M. Fekete, R. D. Robarts, M. Kumagai, H.-P. Nachtnebel, E. Odada, and A. V. Zhulidov. Time for in situ renaissance. *Science*, 2015. doi: 10.1126/science.aac7358.

L. Feyen, J. A. Vrugt, B. O. Nuallain, J. van der Knijff, and A. De Roo. Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model. *Journal of Hydrology*, 332(3-4), 2007.

M. Fiorentino and V. Iacobellis. New insights about the climatic and geologic control on the probability distribution of floods. *Water Resources Research*, 2001. doi: 10.1029/2000WR900315.

T. Francke, S. Foerster, A. Brosinsky, E. Sommerer, J. A. Lopez-Tarazon, A. Güntner, R. J. Batalla, and A. Bronstert. Water and sediment fluxes in Mediterranean mountainous regions: comprehensive dataset for hydro-sedimentological analyses and modelling in a mesoscale catchment (River Isábena, NE Spain). *Earth System Science Data*, 2018. doi: 10.5194/essd-10-1063-2018.

B. François, M. Borga, S. Anquetin, J. D. Creutin, K. Engeland, A. C. Favre, B. Hingray, M. H. Ramos, D. Raynaud, B. Renard, E. Sauquet, J. F. Sauterleute, J. P. Vidal, and G. Warland. Integrating hydropower and intermittent climate-related renewable energies: a call for hydrology. *Hydrological Processes*, 2014. doi: 10.1002/hyp.10274.

A.-M. Fromental and L. Bonnifait. Utilisation de données anciennes pour l'analyse fréquentielle des débits. Stations d'Anduze, Mialet, Saumane, Saint-Jean-du-Gard. Technical report, Cerema, 2018.

G. Galéa, S. Vasquez-Paulus, B. Renard, and P. Breil. L'impact des prélèvements d'eau pour l'irrigation sur les régimes hydrologiques des sous-bassins du tescou et de la séoune (bassin adour-garonne, france). *Revue des Sciences de l'Eau*, 18(3):273–305, 2005.

F. Garavaglia, J. Gailhard, E. Paquet, M. Lang, R. Garcon, and P. Bernardara. Introducing a rainfall compound distribution model based on weather patterns sub-sampling. *Hydrology and Earth System Sciences*, 14, 2010. doi: 10.5194/hess-14-951-2010.

F. Garavaglia, M. Lang, E. Paquet, J. Gailhard, R. Garcon, and B. Renard. Reliability and robustness of a rainfall compound distribution model based on weather pattern sub-sampling. *Hydrology and Earth System Sciences.*, 2011. doi: 10.5194/hess-15-519-2011.

R. Garcia, V. Costa, and F. Silva. Bayesian Rating Curve Modeling: Alternative Error Model to Improve Low-Flow Uncertainty Estimation. *Journal of Hydrologic Engineering*, 2020. doi: 10.1061/(ASCE)HE.1943-5584.0001903.

E. Gaume, L. Gaál, A. Viglione, J. Szolgay, S. Kohnová, and G. Blöschl. Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites. *Journal of Hydrology*, 2010. doi: 10.1016/j.jhydrol.2010.01.008.

C. Genest and J. Nešlehová. A primer on copulas for count data. *ASTIN Bulletin*, 37, 2007. doi: 10.1017/S0515036100014963.

S. Ghosh and B. K. Mallick. A hierarchical bayesian spatio-temporal model for extreme precipitation events. *Environmetrics*, 22, 2011. doi: 10.1002/env.1043.

A. Giacomini, O. Buzzi, B. Renard, and G. P. Giani. Experimental studies on fragmentation of rock falls on impact with rock surfaces. *International Journal of Rock Mechanics & Mining Sciences*, 46:708–715, 2009.

I. Giuntoli, B. Renard, and M. Lang. Floods in france. In Z. W. Kundzewicz, editor, *Changes in flood risk in Europe.* IAHS Press, 2012.

I. Giuntoli, B. Renard, J. P. Vidal, and A. Bard. Low flows in france and their relationship to large scale climate indices. *Journal of Hydrology*, 2013. doi: 10.1016/j.jhydrol.2012.12.038.

P. Goovaerts. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, 228(1-2), 2000.

J. Gotzinger and A. Bardossy. Generic error model for calibration and uncertainty estimation of hydrological models. *Water Resources Research*, 44, 2008.

V. Gouy, L. Liger, S. Ahrouch, C. Bonnineau, N. Carluer, A. Chaumot, M. Coquery, A. Dabrin, C. Margoum, and S. Pesce. Ardières-Morcille in the Beaujolais, France: A research catchment dedicated to study of the transport and impacts of diffuse agricultural pollution in rivers. *Hydrological Processes*, 2021. doi: 10.1002/hyp.14384.

I. B. Gregersen, H. Madsen, D. Rosbjerg, and K. Arnbjerg-Nielsen. A spatial and nonstationary model for the frequency of extreme rainfall events. *Water Resources Research*, 2013. doi: 10.1029/2012wr012570.

G. A. Griffiths, A. McKerchar, and B. Renard. Flood frequency analysis: combining a systematic record with historical, regional, model and analogue information. *Journal of Hydrology(NZ)*, 56(1):1–12, 2017.

L. Gudmundsson, H. X. Do, M. Leonard, and S. Westra. The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment. *Earth Syst. Sci. Data*, 2018. doi: 10.5194/essd-10-787-2018.

P. Guillot and D. Duband. La méthode du gradex pour le calcul de la probabilité des crues à partir des pluies. *AISH Publications*, 1967.

J. Hall, B. Arheimer, G. T. Aronica, A. Bilibashi, M. Boháč, O. Bonacci, M. Borga, P. Burlando, A. Castellarin, G. B. Chirico, P. Claps, K. Fiala, L. Gaál, L. Gorbachova, A. Gül, J. Hannaford, A. Kiss, T. Kjeldsen, S. Kohnová, J. J. Koskela, N. Macdonald, M. Mavrova-Guirguinova, O. Ledvinka, L. Mediero, B. Merz, R. Merz, P. Molnar, A. Montanari, M. Osuch, J. Parajka, R. A. P. Perdigão, I. Radevski, B. Renard, M. Rogger, J. L. Salinas, E. Sauquet, M. Šraj, J. Szolgay, A. Viglione, E. Volpi, D. Wilson, K. Zaimi, and G. Blöschl. A european flood database: facilitating comprehensive flood research beyond administrative boundaries. *Proc. IAHS*, 2015. doi: 10.5194/piahs-370-89-2015.

A. Hannart, J. Pearl, F. E. L. Otto, P. Naveau, and M. Ghil. Causal Counterfactual Theory for the Attribution of Weather and Climate-Related Events. *Bulletin of the American Meteorological Society*, 2016. doi: 10.1175/BAMS-D-14-00034.1.

N. M. Harrison, K. Cawley, B. Renard, and J. Le Coz. The National Ecological Observatory Network: Creating stage-discharge rating curves by combining physical channel inputs with probabilistic modeling techniques. In *AGU Fall Meeting*, 2018.

J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society*, 66, 2004.

R. W. Herschy, editor. *Hydrometry: principles and practices.* John Wiley, New York, 2nd ed edition, 1999. ISBN 978-0-471-97350-8.

R. W. Herschy and R. Herschy. *Streamflow Measurement.* CRC Press, 2002. ISBN 978-0-429-17972-3. doi: 10.1201/9781482271485. URL https://www.taylorfrancis.com/books/9781482271485.

Y. Hirabayashi, R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae. Global flood risk under climate change. *Nature Clim. Change*, 2013. doi: 10.1038/nclimate1911.

M. Ho, A. S. Kiem, and D. C. Verdon-Kidd. A paleoclimate rainfall reconstruction in the murray- darling basin (MDB), australia: 1. evaluation of different paleoclimate archives, rainfall networks, and reconstruction techniques. *Water Resources Research*, 51, 2015. doi: 10.1002/2015WR017058.

G. A. Hodgkins, R. W. Dudley, M. G. Nielsen, B. Renard, and S. L. Qi. Groundwater-level trends in the u.s. glacial aquifer system, 1964–2013. *Journal of Hydrology*, 2017a. doi: https://doi.org/10.1016/j.jhydrol.2017.07.055.

G. A. Hodgkins, P. H. Whitfield, D. H. Burn, J. Hannaford, B. Renard, K. Stahl, A. K. Fleig, H. Madsen, L. Mediero, J. Korhonen, C. Murphy, and D. Wilson. Climate-driven variability in the occurrence of major floods across north america and europe. *Journal of Hydrology*, 2017b. doi: https://doi.org/10.1016/j.jhydrol.2017.07.027.

G. A. Hodgkins, B. Renard, P. H. Whitfield, G. Laaha, K. Stahl, J. Hannaford, D. H. Burn, M. A. Thyer, S. Westra, A. Fleig, C. Murphy, J. Korhonen, O. Rossler, and L. Mediero. Climate driven trends and variability in extreme low flows on four continents. *Submitted to Water Resources Research*, 2023.

I. Horner, B. Renard, J. Le Coz, and M. Lang. Impacts des incertitudes hydrométriques dans l'estimation des débits caractéristiques. Technical report, Irstea, 2016.

I. Horner, J. Le Coz, B. Renard, and F. Branger. Impact de la sensibilité des contrôles hydrauliques sur les incertitudes hydrométriques. *La houille blanche*, (1):27–35, 2018a.

I. Horner, B. Renard, J. Le Coz, F. Branger, H. K. McMillan, and G. Pierrefeu. Impact of stage measurement errors on streamflow uncertainty. *Water Resources Research*, 2018b. doi: doi:10.1002/2017WR022039.

I. Horner, J. Le Coz, B. Renard, F. Branger, and M. Lagouy. Streamflow uncertainty due to the limited sensitivity of controls at hydrometric stations. *Hydrological Processes*, 2022. doi: 10.1002/hyp.14497.

J. R. M. Hosking and J. R. Wallis. *Regional Frequency Analysis: an approach based on L-Moments*. Cambridge University Press, Cambridge, UK, 1997.

D. Huard and A. Mailhot. A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model "abc". *Water Resour. Res.*, 42, 2006.

D. Huard and A. Mailhot. Calibration of hydrological model GR2M using Bayesian uncertainty analysis. *Water Resources Research*, 44, 2008.

Interagency Advisory Committee on Water Data. *Guidelines for determining flood-flow frequency: Bulletin 17B of the Hydrology Subcommittee*. U.S. Geological Survey, 1982.

A. Jacquin and A. Y. Shamseldin. Development of a possibilistic method for the evaluation of predictive uncertainty in rainfall-runoff modeling. *Water Resources Research*, 43, 2007.

M. Jin and J. R. Stedinger. Flood frequency analysis with regional and historical information. *Water Resources Research*, 1989. doi: 10.1029/WR025i005p00925.

A. Kastali, A. Zeroual, M. Remaoun, R. Serrano-Notivoli, and T. Moramarco. Design Flood and Flood-Prone Areas under Rating Curve Uncertainty: Area of Vieux-Ténès, Algeria. *Journal of Hydrologic Engineering*, 2021. doi: 10.1061/(ASCE)HE.1943-5584.0002049.

A. Kastali, A. Zeroual, S. Zeroual, and Y. Hamitouche. Auto-calibration of HEC-HMS Model for Historic Flood Event under Rating Curve Uncertainty. Case Study: Allala Watershed, Algeria. *KSCE Journal of Civil Engineering*, 2022. doi: 10.1007/s12205-021-1051-4.

R. W. Katz, M. B. Parlange, and P. Naveau. Statistics of extremes in hydrology. *Advances in Water Resources*, 2002. doi: 10.1016/s0309-1708(02)00056-8.

D. Kavetski, G. Kuczera, and S. W. Franks. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resour. Res.*, 42(3), 2006a.

D. Kavetski, G. Kuczera, and S. W. Franks. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. *Water Resour. Res.*, 42(3), 2006b.

L. Kazimierski, P. García, N. Ortiz, M. Morale, and M. Re. Aforos de ríos y arroyos en la Cuenca Matanza-Riachuelo. Elaboración de relaciones altura – caudal (curvas HQ). Technical Report Informe LHA 05-397-21, Instituto Nacional del Agua (INA) – ACUMAR, Ezeiza, Argentina, 2021.

J. E. Kiang, C. Gazoorian, H. McMillan, G. Coxon, J. Le Coz, I. K. Westerberg, A. Belleville, D. Sevrez, A. E. Sikorska, A. Petersen-Øverleir, T. Reitan, J. Freer, B. Renard, V. Mansanarez, and R. Mason. A Comparison of Methods for Streamflow Uncertainty Estimation. *Water Resources Research*, 2018. doi: doi:10.1029/2018WR022708.

A. Klami, S. Virtanen, and S. Kaski. Bayesian canonical correlation analysis. *Journal of Machine Learning Research*, 14:965–1003, 2013.

K. Kochanek, B. Renard, P. Arnaud, Y. Aubert, M. Lang, T. Cipriani, and E. Sauquet. A data-based comparison of flood frequency analysis methods used in france. *Nat. Hazards Earth Syst. Sci.*, 2014. doi: 10.5194/nhess-14-295-2014.

R. Krzysztofowicz. Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology*, 268(1-4), 2002.

G. Kuczera. On the relationship between the reliability of parameter estimates and hydrologic time series data used in calibration. *Water Resources Research*, 1982. doi: 10.1029/WR018i001p00146.

G. Kuczera. Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty. *Water Resources Research*, 1983. doi: 10.1029/WR019i005p01151.

G. Kuczera and E. Parent. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm. *Journal of Hydrology*, 1998.

G. Kuczera and B. J. Williams. Effect of Rainfall Errors on Accuracy of Design Flood Estimates. *Water Resour. Res.*, 28(4), 1992.

G. Kuczera, D. Kavetski, S. Franks, and M. Thyer. Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *Journal of Hydrology*, 2006.

G. Kuczera, D. Kavetski, B. Renard, and M. Thyer. A limited-memory acceleration strategy for MCMC sampling in hierarchical bayesian calibration of hydrological models. *Water Resources Research.*, 46, 2010a.

G. Kuczera, B. Renard, M. Thyer, and D. Kavetski. There are no hydrological monsters, just models and observations with large uncertainties! *Hydrological Sciences Journal*, 55(6), 2010b.

G. Kuczera, D. Kavetski, B. Renard, and M. Thyer. Bayesian methods. In V. P. Singh, editor, *Handbook of Applied Hydrology, Second Edition.* McGraw-Hill Education, 2017.

M. Lang and B. Renard. Analyse régionale sur les extrêmes hydrométriques en france : détection de changements cohérents et recherche de causalité hydrologique. *La houille blanche*, pages 83–89, 2007.

M. Lang, T. B. M. J. Ouarda, and B. Bobée. Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, 1999.

M. Lang, K. Pobanz, B. Renard, E. Renouf, and E. Sauquet. Extrapolation of rating curves by hydraulic modelling, with application to flood frequency analysis. *Hydrological sciences Journal.*, 55(6):883–898, 2010.

M. Lang, P. Arnaud, J. Carreau, N. Deaux, L. Dezileau, F. Garavaglia, A. Latapie, L. Neppel, E. Paquet, B. Renard, J.-M. Soubeyroux, B. Terrier, J.-M. Veysseire, Y. Aubert, A. Auffray, F. Borchi, P. Bernardara, J.-C. Carre, D. Chambon, T. Cipriani, J.-L. Delgado, H. Doumenc, R. Fantin, S. Jourdain, K. Kochanek, A. Paquier, E. Sauquet, and Y. Tramblay. Résultats du projet ExtraFlo (ANR 2009-2013) sur l'estimation des pluies et crues extrêmes. *La houille blanche*, (2):5–13, 2014.

M. Lang, M. Darienzo, J. Le Coz, and B. Renard. Evaluation des incertitudes et de l'homogénéité de longues séries de débits de crue sur le Rhin à Bâle (1225–2017) et Maxau (1815–2018). *La Houille Blanche - Revue internationale de l'eau*, 2022. doi: 10.1080/27678490.2022.2053313.

P. D. Le, M. Leonard, and S. Westra. Modeling spatial dependence of rainfall extremes across multiple durations. *Water Resources Research*, 54(3):2233–2248, 2018. doi: 10.1002/ 2017WR022231.

J. Le Coz. *Quantifying discharges and fluxes of matters in rivers*. Habilitation à Diriger des Recherches, Université Grenoble Alpes, 2017.

J. Le Coz, C. Chaléon, L. Bonnifait, R. Le Boursicaud, B. Renard, F. Branger, J. Diribarne, and M. Valente. Analyse bayésienne des courbes de tarage et de leurs incertitudes : la méthode BaRatin. *La houille blanche*, (6):31–41, 2013.

J. Le Coz, M. Jodeau, A. Hauet, B. Marchand, and R. Le Boursicaud. Image-based velocity and discharge measurements in field and laboratory river engineering studies using the free Fudaa-LSPIV software. In *Proceedings of the International Conference on Fluvial Hydraulics RIVER FLOW 2014*, 2014a.

J. Le Coz, B. Renard, L. Bonnifait, F. Branger, and R. Le Boursicaud. Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A bayesian approach. *Journal of Hydrology*, 2014b. doi: http://dx.doi.org/10.1016/j.jhydrol.2013.11.016.

J. Le Coz, H. Chen, B. Renard, and P. O. Malaterre. Bayesian estimation of streamflow using satellite data and ground discharge measurements. In AGU, editor, *AGU Fall Meeting*, 2018.

J. Le Coz, B. Renard, V. Vansuyt, M. Jodeau, and A. Hauet. Estimating the uncertainty of video-based flow velocity and discharge measurements due to the conversion of field to image coordinates. *Hydrological Processes*, 2021. doi: 10.1002/hyp.14169.

E. Le Roux, G. Evin, N. Eckert, J. Blanchet, and S. Morin. Elevation-dependent trends in extreme snowfall in the French Alps from 1959 to 2019. *The Cryosphere*, 2021. doi: 10.5194/tc-15-4335-2021.

E. Leblois and J.-D. Creutin. Space-time simulation of intermittent rainfall with prescribed advection field: Adaptation of the turning band method. *Water Resources Research*, 2013. doi: 10.1002/wrcr.20190.

Z. Li, S. Gao, M. Chen, J. J. Gourley, C. Liu, A. F. Prein, and Y. Hong. The conterminous United States are projected to become more prone to flash floods in a high-end emissions scenario. *Communications Earth & Environment*, 2022. doi: 10.1038/s43247-022-00409-6.

C. H. R. Lima and U. Lall. Hierarchical Bayesian modeling of multisite daily rainfall occurrence: Rainy season onset, peak, and end. *Water Resources Research*, 2009. doi: W0742210.1029/ 2008wr007485.

C. H. R. Lima and U. Lall. Climate informed long term seasonal forecasts of hydroenergy inflow for the Brazilian hydropower system. *Journal of Hydrology*, 2010a. doi: 10.1016/j.jhydrol. 2009.11.026.

C. H. R. Lima and U. Lall. Spatial scaling in a changing climate: A hierarchical bayesian model for non-stationary multi-site annual maximum and monthly streamflow. *Journal of Hydrology*, 383, 2010b. doi: 10.1016/j.jhydrol.2009.12.045.

C. H. R. Lima and U. Lall. Climate informed monthly streamflow forecasts for the Brazilian hydropower network using a periodic ridge regression model. *Journal of Hydrology*, 2010c. doi: 10.1016/j.jhydrol.2009.11.016.

R. K. Linsley and M. A. Kohler. *Hydrology for Engineers*. McGraw Hill, London, 1988.

M. Lucas. *Comment valoriser les données anciennes pour l'analyse fréquentielle des crues : application au Rhône à Beaucaire de 1500 à 2020*. PhD thesis, INRAE, Université de Lyon, 2023.

M. Lucas, B. Renard, J. Le Coz, M. Lang, A. Bard, and G. Pierrefeu. Are historical stage records useful to decrease the uncertainty of flood frequency analysis ? A 200-year long case study. *Journal of Hydrology*, 2023. doi: 10.1016/j.jhydrol.2023.129840.

J. D. Lundquist, J. W. Roche, H. Forrester, C. Moore, E. Keenan, G. Perry, N. Cristea, B. Henn, K. Lapo, B. McGurk, D. R. Cayan, and M. D. Dettinger. Yosemite Hydroclimate Network: Distributed stream and atmospheric data for the Tuolumne River watershed and surroundings. *Water Resources Research*, 2016. doi: doi:10.1002/2016WR019261.

M. J. Machado, B. A. Botero, J. López, F. Francés, A. Díez-Herrero, and G. Benito. Flood frequency analysis of historical flood data under stationary and non-stationary modelling. *Hydrology and Earth System Sciences*, 2015. doi: 10.5194/hess-19-2561-2015.

A. Mailhot, S. Lachance-Cloutier, G. Talbot, and A.-C. Favre. Regional estimates of intense rainfall based on the Peak-Over-Threshold (POT) approach. *Journal of Hydrology*, 2013. doi: 10.1016/j.jhydrol.2012.10.036.

V. Mansanarez. *Non-unique stage-discharge relations: Bayesian analysis of complex rating curves and their uncertainties*. PhD thesis, Irstea, Communauté Université Grenoble Alpes, 2016.

V. Mansanarez, J. Le Coz, B. Renard, M. Lang, G. Pierrefeu, and P. Vauchel. Bayesian analysis of stage-fall-discharge rating curves and their uncertainties. *Water Resources Research*, 2016. doi: 10.1002/2016WR018916.

V. Mansanarez, R. Le Boursicaud, J. Le Coz, B. Renard, M. Lang, I. Horner, G. Pierrefeu, and K. Pobanz. BaRatin-SFD, analyse bayésienne des courbes de tarage à double échelle et de leurs incertitudes. *La houille blanche*, (5):22–28, 2017.

V. Mansanarez, B. Renard, J. L. Coz, M. Lang, and M. Darienzo. Shift Happens! Adjusting Stage-Discharge Rating Curves to Morphological Changes at Known Times. *Water Resources Research*, 2019. doi: 10.1029/2018wr023389.

P. Mantovan and E. Todini. Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. *Journal of Hydrology*, 330(1-2), 2006.

D. Maraun, H. W. Rust, and T. J. Osborn. Synoptic airflow and UK daily precipitation extremes development and validation of a vector generalised linear model. *Extremes*, 13, 2010. doi: 10.1007/s10687-010-0102-x.

L. Marshall, D. Nott, and A. Sharma. Towards dynamic catchment modelling: a Bayesian hierarchical mixtures of experts framework. *Hydrological Processes*, 21(7), 2007.

E. S. Martins and J. R. Stedinger. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour. Res.*, 36(3), 2000.

J. R. R. Mason, J. E. Kiang, and T. A. Cohn. Rating curve uncertainty: A comparison of estimation methods. 2016. doi: 10.1201/9781315644479-115.

P. McCullagh and J. A. Nelder. *Generalized linear models (Second Edition)*. Monographs on Statistics & Applied Probability. Chapman and Hall/CRC, 1989. ISBN 0-471-66719-6.

D. McInerney, M. Thyer, D. Kavetski, J. Lerat, and G. Kuczera. Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*, 2017. doi: 10.1002/2016WR019168.

F. Mendez-Rios. Courbes de tarage dynamiques - Influence de la marée. Technical report, INRAE, 2022.

R. Merz and G. Bloschl. Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. *Water Resources Research*, 2008a. doi: 10.1029/2007WR006744.

R. Merz and G. Bloschl. Flood frequency hydrology: 2. Combining data evidence. *Water Resources Research*, 2008b.

F. Moatar, M. Floury, A. J. Gold, M. Meybeck, B. Renard, M. Ferréol, A. Chandesris, C. Minaudo, K. Addy, J. Piffady, and G. Pinay. Stream Solutes and Particulates Export Regimes: A New Framework to Optimize Their Monitoring. *Frontiers in Ecology and Evolution*, 2020. doi: 10.3389/fevo.2019.00516.

A. Molini, L. G. Lanza, and P. La Barbera. The impact of tipping-bucket raingauge measurement errors on design rainfall for urban-scale applications. *Hydrological Processes*, 19(5), 2005.

A. Montanari and A. Brath. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resour. Res.*, 40, 2004.

H. Moradkhani, K. L. Hsu, H. Gupta, and S. Sorooshian. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, 41(5), 2005a.

H. Moradkhani, S. Sorooshian, H. V. Gupta, and P. R. Houser. Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, 28, 2005b.

T. Morlot. *La gestion dynamique des relations hauteur-débit des stations d'hydrométrie et le calcul des incertitudes associées : un indicateur de gestion, de qualité et de suivi des points de mesure.* PhD thesis, Grenoble University, Laboratoire d'étude des transferts en hydrologie et environnement, Grenoble, France, 2014. URL `http://www.theses.fr/2014GRENU029`.

L. Moulin, E. Gaume, and C. Obled. Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations. *Hydrology and Earth System Sciences*, 13, 2009.

K. P. Murphy. *Probabilistic Machine Learning: An introduction.* MIT Press, 2022. URL `probml.ai`.

K. P. Murphy. *Probabilistic Machine Learning: Advanced Topics.* MIT Press, 2023. URL `http://probml.github.io/book2`.

J. B. Nagel and B. Sudret. A unified framework for multilevel uncertainty quantification in Bayesian inverse problems. *Probabilistic Engineering Mechanics*, 2016. doi: 10.1016/j.probengmech.2015.09.007.

R. Naulet, M. Lang, T. B. M. J. Ouarda, D. Coeur, B. Bobee, A. Recking, and D. Moussay. Flood frequency analysis on the ardeche river using french documentary sources from the last two centuries. *Journal of Hydrology*, 313, 2005. doi: 10.1016/j.jhydrol.2005.02.011.

P. Naveau, M. Nogaj, C. Ammann, P. Yiou, D. Cooley, and V. Jomelli. Statistical methods for the analysis of climate extremes. *Comptes Rendus Geoscience*, 337(10-11), 2005.

G. S. Nearing, Y. Tian, H. V. Gupta, M. P. Clark, K. W. Harrison, and S. V. Weijs. A philosophical basis for hydrological uncertainty. *Hydrological Sciences Journal*, 61, 2016. doi: 10.1080/02626667.2016.1183009.

L. Neppel, B. Renard, M. Lang, P. A. Ayral, D. Coeur, E. Gaume, N. Jacob, O. Payrastre, K. Pobanz, and F. Vinet. Flood frequency analysis using historical data: accounting for random and systematic errors. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 2010. doi: 10.1080/02626660903546092.

L. Neppel, P. Arnaud, F. Borchi, J. Carreau, F. Garavaglia, M. Lang, E. Paquet, B. Renard, J.-M. Soubeyroux, and J.-M. Veysseire. Résultats du projet extraflo sur la comparaison des méthodes d'estimation des pluies extrêmes en france. *La houille blanche*, (2):14–19, 2014.

C. Nguyen, E. Gaume, and O. Payrastre. Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations. *Journal of Hydrology*, 2014. doi: 10.1016/j.jhydrol.2013.09.058.

M. Nogaj, S. Yiou, S. Parey, F. Malek, and P. Naveau. Amplitude and frequency of temperature extremes over the north atlantic region. *Geophysical Research Letters*, 33(10), 2006.

D. Ocio, N. Le Vine, I. Westerberg, F. Pappenberger, and W. Buytaert. The role of rating curve uncertainty in real-time flood forecasting. *Water Resources Research*, 2017. doi: 10.1002/2016WR020225.

H. Onibon, T. Lebel, A. Afouda, and G. Guillot. Gibbs sampling for conditional spatial disaggregation of rain fields. *Water Resources Research*, 40(8), 2004.

A. Ossandon, B. Rajagopalan, and W. Kleiber. Spatial-temporal multivariate semi-bayesian hierarchical framework for extreme precipitation frequency analysis. *Journal of Hydrology*, 600:126499, 2021. ISSN 00221694. doi: 10.1016/j.jhydrol.2021.126499. URL https://linkinghub.elsevier.com/retrieve/pii/S0022169421005461.

S. A. Padoan, M. Ribatet, and S. A. Sisson. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105, 2010.

G. Panthou, T. Vischel, T. Lebel, J. Blanchet, G. Quantin, and A. Ali. Extreme rainfall in West Africa: A regional modeling. *Water Resources Research*, 2012. doi: 10.1029/2012WR012052.

G. Panthou, T. Vischel, T. Lebel, G. Quantin, A. F. Pugin, J. Blanchet, and A. Ali. From pointwise testing to a regional vision: An integrated statistical approach to detect nonstationarity in extreme daily rainfall. Application to the Sahelian region. *Journal of Geophysical Research: Atmospheres*, 2013. doi: 10.1002/jgrd.50340.

E. Paquet, F. Garavaglia, R. Garçon, and J. Gailhard. The SCHADEX method: a semi-continuous rainfall-runoff simulation for extreme flood estimation. *Journal of Hydrology*, 2013. doi: http://dx.doi.org/10.1016/j.jhydrol.2013.04.045.

O. Payrastre, E. Gaume, and H. Andrieu. Usefulness of historical information for flood frequency analyses: Developments based on a case study. *Water Resources Research*, 2011. doi: 10.1029/2010WR009812.

L. Perreault, J. Bernier, B. Bobee, and E. Parent. Bayesian change-point analysis in hydrometeorological time series. part 1. the normal model revisited. *Journal of Hydrology*, 235, 2000a. doi: 10.1016/S0022-1694(00)00270-5.

L. Perreault, J. Bernier, B. Bobee, and E. Parent. Bayesian change-point analysis in hydrometeorological time series. part 2. comparison of change-point models and forecasting. *Journal of Hydrology*, 235, 2000b. doi: 10.1016/S0022-1694(00)00271-7.

E. Perret, J. Le Coz, and B. Renard. Courbes de tarage dynamiques - végétation aquatique et influence de la marée. Technical report, INRAE, 2020a.

E. Perret, B. Renard, and J. Le Coz. A rating curve model accounting for cyclic stage-discharge shifts due to seasonal aquatic vegetation. *Water Resources Research*, 2020b. doi: 10.1029/2020WR027745.

E. Perret, J. Le Coz, and B. Renard. Courbes de tarage dynamiques pour la végétation aquatique saisonnière. *La Houille Blanche - Revue internationale de l'eau*, 2022. doi: 10.1080/27678490.2022.2082339.

E. Perret, B. Camenen, C. Berni, K. El kadi Abderrezzak, and B. Renard. Uncertainties in Models Predicting Critical Bed Shear Stress of Cohesionless Particles. *Journal of Hydraulic Engineering*, 2023. doi: 10.1061/JHEND8.HYENG-13101.

C. Perrin, C. Michel, and V. Andreassian. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1-4), 2003.

A. Petersen-Øverleir and T. Reitan. Uncertainty in flood discharges from urban and small rural catchments due to inaccurate head measurement. *Hydrology Research*, 2005. doi: 10.2166/nh.2005.0018.

A. A. Pourzanjani, R. M. Jiang, B. Mitchell, P. J. Atzberger, and L. R. Petzold. Bayesian Inference over the Stiefel Manifold via the Givens Representation. *Bayesian Analysis*, 2020. doi: 10.1214/20-BA1202.

I. Prosdocimi, T. R. Kjeldsen, and J. D. Miller. Detection and attribution of urbanization effect on flood extremes using nonstationary flood-frequency models. *Water Resources Research*, 51(6):4244–4262, 2015. doi: 10.1002/2015WR017065.

J. Qiu, B. Liu, X. Yu, and Z. Yang. Combining a segmentation procedure and the BaRatin stationary model to estimate nonstationary rating curves and the associated uncertainties. *Journal of Hydrology*, 2021. doi: 10.1016/j.jhydrol.2021.126168.

A. Ramachandra Rao and K. H. Hamed. *Flood Frequency Analysis.* CRC Press, 1 edition, 2019. ISBN 978-0-429-12881-3. doi: 10.1201/9780429128813.

B. J. Reich and B. A. Shaby. A hierarchical max-stable spatial model for extreme precipitation. *Annals of Applied Statistics*, 6, 2012. doi: Doi10.1214/12-Aoas591.

P. Reichert and J. Mieleitner. Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. *Water Resour. Res.*, 45, 2009.

D. S. Reis and J. R. Stedinger. Bayesian MCMC flood frequency analysis with historical information. *Journal of Hydrology*, 2005. doi: 10.1016/j.jhydrol.2005.02.028.

B. Renard. *Détection et prise en compte d'éventuels impacts du changement climatique sur les extrêmes hydrologiques en France.* Ph.D Thesis, INPG / Cemagref, Lyon, France, 2006.

B. Renard. Détection et prise en compte d'éventuels impacts du changement climatique sur les extrêmes hydrologiques en france. *La houille blanche*, (1):109–117, 2008.

B. Renard. A bayesian hierarchical approach to regional frequency analysis. *Water Resources Research*, 2011. doi: 10.1029/2010WR010089.

B. Renard. Procédures statistiques de la banque HYDRO 3: Evaluation et vérification des codes de calcul. Technical report, Irstea, 2016a.

B. Renard. Procédures statistiques de la banque HYDRO 3: documentation des codes de calcul, 2016b.

B. Renard. BaM ! (bayesian modeling): Un code de calcul pour l'estimation d'un modèle quelconque et son utilisation en prédiction, 2017a.

B. Renard. Probabilités et statistiques appliquées à l'hydrologie, 2017b.

B. Renard. Prise en compte des erreurs systématiques dans l'estimation des courbes de tarage et des modèles hydrologiques. Technical report, Irstea, 2018.

B. Renard. benRenard/disTRIMbution: DisTRIMbution v0.1.0 (beta version), 2021a. URL https://zenodo.org/record/4905580.

B. Renard. benRenard/musicXML: musicXML r package v0.1.0 (beta version), 2021b. URL https://zenodo.org/record/4924680.

B. Renard. STooDs-tools/RSTooDs: RSTooDs package v0.1.1, 2021c. URL https://zenodo.org/record/5075760.

B. Renard. STooDs-tools/STooDs: STooDs engine v0.1.0, 2021d. URL https://zenodo.org/record/5075586.

B. Renard. benRenard/sequenceR: sequenceR v0.1.0 (beta version), 2021e. URL `https://zenodo.org/record/4924597`.

B. Renard. benRenard/HydroPortailStats: HydroPortailStats r package v1.0.0, 2022. URL `https://zenodo.org/record/6769234`.

B. Renard. Use of a national flood mark database to estimate flood hazard in the distant past. *Hydrological Sciences Journal*, 2023. doi: 10.1080/02626667.2023.2212165.

B. Renard and U. Lall. Regional frequency analysis conditioned on large-scale atmospheric or oceanic fields. *Water Resources Research*, 2014. doi: 10.1002/2014WR016277.

B. Renard and M. Lang. Use of a gaussian copula for multivariate extreme value analysis: some case studies in hydrology. *Advances in Water Resources*, 30(4):897–912, 2007.

B. Renard and C. Le Bescond. The Music of Water. Vienna, Austria, 2022. doi: 10.5194/egusphere-egu22-9759.

B. Renard and M. Thyer. Revealing hidden climate indices from the occurrence of hydrologic extremes. *Water Resources Research*, 2019. doi: 10.1029/2019WR024951.

B. Renard and J.-P. Vidal. Performance weighting of GCMs. Part 1: A method based on explicit probabilistic models and accounting for observation uncertainty. *unpublished*, 2017a.

B. Renard and J.-P. Vidal. Performance weighting of GCMs. Part 2: Daily atmospheric variability across Europe. *unpublished*, 2017b.

B. Renard, V. Garreta, and M. Lang. An application of bayesian analysis and markov chain monte carlo methods to the estimation of a regional trend in annual maxima. *Water Resources Research*, 2006a. doi: 10.1029/2005WR004591.

B. Renard, M. Lang, and P. Bois. Statistical analysis of extreme events in a non-stationary context via a bayesian framework. *Stochastic Environmental Research and Risk Assessment*, 21:97–112, 2006b.

B. Renard, M. Lang, P. Bois, A. Dupeyrat, O. Mestre, H. Niel, J. Gailhard, C. Laurent, L. Neppel, and E. Sauquet. Evolution des extrêmes hydrométriques en france à partir de données observées. *La houille blanche*, (6):48–54, 2006c.

B. Renard, M. Thyer, G. Kuczera, and D. Kavetski. Bayesian total error analysis for hydrologic models: Sensitivity to error models. In *MODSIM Congress*, 2007.

B. Renard, M. Lang, P. Bois, A. Dupeyrat, O. Mestre, H. Niel, E. Sauquet, C. Prudhomme, S. Parey, E. Paquet, L. Neppel, and J. Gailhard. Regional methods for trend detection: Assessing field significance and regional consistency. *Water Resources Research*, 2008. doi: 10.1029/2007WR006268.

B. Renard, D. Kavetski, and G. Kuczera. Comment on "an integrated hydrologic bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction." by n. k. ajami, q. y. duan, and s. sorooshian. *Water Resources Research*, 45, 2009.

B. Renard, D. Kavetski, M. Thyer, G. Kuczera, and S. W. Franks. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, 46, 2010.

B. Renard, D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, and S. W. Franks. Towards a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation. *Water Resources Research*, 2011. doi: 10.1029/2011WR010643.

B. Renard, K. Kochanek, M. Lang, F. Garavaglia, E. Paquet, L. Neppel, K. Najib, J. Carreau, P. Arnaud, Y. Aubert, F. Borchi, J. M. Soubeyroux, S. Jourdain, J. M. Veysseire, E. Sauquet, T. Cipriani, and A. Auffray. Data-based comparison of frequency analysis methods: A general framework. *Water Resources Research*, 2013a. doi: 10.1002/wrcr.20087.

B. Renard, X. Sun, and M. Lang. Bayesian methods for non-stationary extreme value analysis. In S. Sorooshian, D. Easterling, A. AghaKouchak, S. Schubert, and K. Hsu, editors, *Hydrologic Extremes in a Changing Climate - Detection, Analysis & Uncertainty*. Springer, 2013b.

B. Renard, J. Le Coz, B. Blanquart, and L. Bonnifait. Calcul d'incertitudes en hydrologie : applications à la prédétermination et à l'hydrométrie., 2017.

B. Renard, M. Thyer, D. McInerney, D. Kavetski, M. Leonard, and S. Westra. A hidden climate indices modeling framework for multi-variable space-time data. *Water Resources Research*, 2021. doi: 10.1029/2021WR030007.

B. Renard, D. McInerney, S. Westra, M. Leonard, D. Kavetski, M. Thyer, and J. Vidal. Floods and Heavy Precipitation at the Global Scale: 100-Year Analysis and 180-Year Reconstruction. *Journal of Geophysical Research: Atmospheres*, 2023. doi: 10.1029/2022JD037908.

M. Ribatet, D. Cooley, and A. C. Davison. Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 2012. doi: 10.5705/Ss.2009.248.

D. Rosbjerg. Estimation in partial duration series with independent and dependent peak values. *Journal of Hydrology*, 1985. doi: 10.1016/0022-1694(85)90098-8.

H. Roux and D. Dartus. Parameter identification using optimization techniques in open-channel inverse problems. *Journal of Hydraulic Research*, 2005. doi: 10.1080/00221680509500125.

H. Roux and D. Dartus. Sensitivity Analysis and Predictive Uncertainty Using Inundation Observations for Parameter Estimation in Open-Channel Inverse Problem. *Journal of Hydraulic Engineering*, 2008. doi: 10.1061/(ASCE)0733-9429(2008)134:5(541).

E. Rouzies, C. Lauvernet, B. Sudret, and A. Vidard. How is a global sensitivity analysis of a catchment-scale, distributed pesticide transfer model performed? Application to the PESH-MELBA model. *Geoscientific Model Development*, 2023. doi: 10.5194/gmd-16-3137-2023.

A. Sabourin. *Mélanges bayésiens de modèles d'extrêmes multivariés : application à la prédétermination régionale des crues avec données incomplètes*. PhD thesis, Université Claude Bernard - Lyon I,, 2013.

A. Sabourin and B. Renard. Combining regional estimation and historical floods: A multivariate semiparametric peaks-over-threshold model with censored data. *Water Resources Research*, 2015. doi: 10.1002/2015WR017320.

G. Salvadori and C. De Michele. Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research*, 40, 2004.

H. Sang and A. E. Gelfand. Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics*, 15, 2010. doi: 10.1007/s13253-009-0010-1.

A. Sarhadi, D. H. Burn, M. Concepción Ausín, and M. P. Wiper. Time-varying nonstationary multivariate risk analysis using a dynamic bayesian copula. *Water Resources Research*, 52, 2016. doi: 10.1002/2015WR018525.

V. B. Sauer and D. P. Turnipseed. Stage measurement at gaging stations. In *U.S. Geological Survey Techniques and Methods book 3, chap. A7*, page 45. USGS, 2010. URL https://doi.org/10.3133/tm3A7.

N. Sauvion, V. Mauriello, B. Renard, and N. Boissot. Impact of melon accessions resistant to aphids on the demographic potential of silverleaf whitefly. *Journal of Economic Entomology*, 98(2):557–567, 2005.

G. Schoups and J. A. Vrugt. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resour. Res.*, 2010. doi: 10.1029/2009wr008933.

A. Schöniger, T. Wöhling, L. Samaniego, and W. Nowak. Model selection on solid ground: Rigorous comparison of nine ways to evaluate Bayesian model evidence. *Water Resources Research*, 2014. doi: 10.1002/2014WR016062.

E. Severino and T. Alpuim. Spatiotemporal models in the estimation of area precipitation. *Environmetrics*, 16, 2005.

A. Sharma, C. Wasko, and D. P. Lettenmaier. If Precipitation Extremes Are Increasing, Why Aren't Floods? *Water Resources Research*, 2018. doi: 10.1029/2018wr023749.

J. J. Sharples, G. J. Cary, P. Fox-Hughes, S. Mooney, J. P. Evans, M.-S. Fletcher, M. Fromm, P. F. Grierson, R. McRae, and P. Baker. Natural hazards in Australia: extreme bushfire. *Climatic Change*, 139, 2016. doi: 10.1007/s10584-016-1811-1.

I. V. Sideris, M. Gabella, R. Erdin, and U. Germann. Real-time radar-rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of switzerland: Real-time radar-rain-gauge merging. *Quarterly Journal of the Royal Meteorological Society*, 140, 2014. doi: 10.1002/qj.2188.

A. E. Sikorska and B. Renard. Calibrating a hydrological model in stage space to account for rating curve uncertainties: general framework and key challenges. *Advances in Water Resources*, 2017. doi: https://doi.org/10.1016/j.advwatres.2017.04.011.

M. Sivapalan, G. Blöschl, R. Merz, and D. Gutknecht. Linking flood frequency to long-term water balance: Incorporating effects of seasonality: FLOOD FREQUENCY AND HYDROLOGIC SEASONALITY. *Water Resources Research*, 2005. doi: 10.1029/2004WR003439.

L. Slater, G. Villarini, S. Archfield, D. Faulkner, R. Lamb, A. Khouakhi, and J. Yin. Global Changes in 20-Year, 50-Year, and 100-Year River Floods. *Geophysical Research Letters*, 2021. doi: 10.1029/2020GL091824.

L. J. Slater, G. Villarini, A. A. Bradley, and G. A. Vecchi. A dynamical statistical framework for seasonal streamflow forecasting in an agricultural watershed. *Climate Dynamics*, 2019. doi: 10.1007/s00382-017-3794-7.

L. J. Slater, B. Anderson, M. Buechel, S. Dadson, S. Han, S. Harrigan, T. Kelder, K. Kowal, T. Lees, T. Matthews, C. Murphy, and R. L. Wilby. Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management. *Hydrology and Earth System Sciences*, 2020. doi: 10.5194/hess-2020-576.

D. M. Stasinopoulos and R. A. Rigby. Generalized additive models for location scale and shape (GAMLSS) in r. *Journal of Statistical Software*, 23(7):46, 2007. ISSN 1548-7660. doi: 10.18637/jss.v023.i07.

J. R. Stedinger and T. A. Cohn. Flood frequency-analysis with historical and paleoflood information. *Water Resources Research*, 22(5):785–793, 1986. ISSN 0043-1397.

J. R. Stedinger and G. D. Tasker. Regional hydrologic analysis: 1. Ordinary, weighted and generalized least squares compared. *Water Resources Research*, 21(9), 1985.

S. Steinschneider and U. Lall. A hierarchical bayesian regional model for nonstationary precipitation extremes in northern california conditioned on tropical moisture exports. *Water Resources Research*, 2015. doi: 10.1002/2014WR016664.

E. Steirou, L. Gerlitz, H. Apel, X. Sun, and B. Merz. Climate influences on flood probabilities across europe. *Hydrology and Earth System Sciences*, 23, 2019. doi: 10.5194/hess-23-1305-2019.

B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 2008. doi: 10.1016/j.ress.2007.04.002.

X. Sun. *Regional frequency analysis of precipitation accounting for climate variability and change.* PhD thesis, Irstea Lyon, Grenoble University, 2013.

X. Sun, M. Thyer, B. Renard, and M. Lang. A general regional frequency analysis framework for quantifying local-scale climate effects: A case study of ENSO effects on southeast queensland rainfall. *Journal of Hydrology*, 2014. doi: 10.1016/j.jhydrol.2014.02.025.

X. Sun, U. Lall, B. Merz, and N. V. Dung. Hierarchical bayesian clustering for nonstationary flood frequency analysis: Application to trends of annual maximum flow in germany. *Water Resources Research*, 51, 2015a. doi: doi:10.1002/2015WR017117.

X. Sun, B. Renard, M. Thyer, S. Westra, and M. Lang. A global analysis of the asymmetric effect of ENSO on extreme precipitation. *Journal of Hydrology*, 2015b. doi: http://dx.doi.org/10.1016/j.jhydrol.2015.09.016.

M. Thiemann, H. Trosset, H. Gupta, and S. Sorooshian. Bayesian recursive parameter estimation for hydrologic models. *Water Resources Research*, 37(10), 2001.

G. Thirel, K. Gerlinger, C. Perrin, G. Drogue, B. Renard, and J.-P. Wagner. Quels futurs possibles pour les débits des affluents français du Rhin (Moselle, Sarre, Ill) ? *LHB*, (5-6): 140–149, 2019.

M. Thyer and G. Kuczera. A hidden markov model for modelling long-term persistence in multi-site rainfall time series. 2. real data analysis. *Journal of Hydrology*, 275, 2003a.

M. Thyer and G. Kuczera. A hidden markov model for modelling long-term persistence in multi-site rainfall time series 1. model calibration using a bayesian approach. *Journal of Hydrology*, 275, 2003b.

M. Thyer, B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modelling: a case study using bayesian total error analysis. *Water Resources Research*, 45, 2009.

M. Thyer, M. Leonard, D. Kavetski, S. Need, and B. Renard. The open source RFortran library for accessing r from fortran, with applications in environmental modelling. *Environmental Modelling & Software*, 26(2):219–234, 2011.

M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999. ISSN 1467-9868. doi: 10.1111/1467-9868.00196.

M. Tolsa, Y. Aubert, J. Le Coz, B. Renard, P. Arnaud, J.-A. Fine, and D. Organde. Méthode de consolidation des courbes de tarage pour les crues d'occurrence rare sur le bassin versant expérimental du real collobrier. *La houille blanche*, (6):16–23, 2013.

N. Tuteja, D. Shin, R. Laugesen, U. Khan, Q. Shao, E. Wang, M. Li, H. Zheng, G. Kuczera, D. Kavetski, G. Evin, M. Thyer, A. MacDonald, T. Chia, and B. Le. Experimental evaluation of the dynamic seasonal streamflow forecasting approach. Technical report, Bureau of Meteorology, Melbourne, Vic., Australia, 2011.

S. Uhlemann, A. H. Thieken, and B. Merz. A consistent set of trans-basin floods in Germany between 1952-2002. *Hydrology and Earth System Sciences*, 2010. doi: 10.5194/hess-14-1277-2010.

E. S. Valdez, F. Anctil, and M.-H. Ramos. Choosing between post-processing precipitation forecasts or chaining several uncertainty quantification tools in hydrological forecasting systems. *Hydrology and Earth System Sciences*, 2022. doi: 10.5194/hess-26-197-2022.

J. P. Vidal, B. Hingray, C. Magand, E. Sauquet, and A. Ducharne. Hierarchy of climate and hydrological uncertainties in transient low-flow projections. *Hydrology and Earth System Sciences*, 2016. doi: 10.5194/hess-20-3651-2016.

A. Viglione, R. Merz, J. L. Salinas, and G. Blöschl. Flood frequency hydrology: 3. A Bayesian analysis. *Water Resources Research*, 2013. doi: 10.1029/2011wr010782.

G. Villarini, P. V. Mandapaka, W. F. Krajewski, and R. J. Moore. Rainfall and sampling uncertainties: A rain gauge perspective. *Journal of Geophysical Research-Atmospheres*, 113 (D11), 2008.

T. Vischel, T. Lebel, S. Massuel, and B. Cappelaere. Conditional simulation schemes of rain fields and their application to rainfall-runoff modeling studies in the Sahel. *Journal of Hydrology*, 375(1-2), 2009.

T. Wagener and H. V. Gupta. Model identification for hydrological forecasting under uncertainty. *Stochastic Environmental Research and Risk Assessment*, 19(6), 2005.

P. J. Ward, B. Jongman, P. Salamon, A. Simpson, P. Bates, T. De Groeve, S. Muis, E. C. de Perez, R. Rudari, M. A. Trigg, and H. C. Winsemius. Usefulness and limitations of global flood risk models. *Nature Climate Change*, 2015. doi: 10.1038/nclimate2742.

S. Westra and S. A. Sisson. Detection of non-stationarity in precipitation extremes using a max-stable process model. *Journal of Hydrology*, 406, 2011. doi: http://dx.doi.org/10.1016/j.jhydrol.2011.06.014.

S. Westra, L. V. Alexander, and F. W. Zwiers. Global increasing trends in annual maximum daily precipitation. *Journal of Climate*, 2012. doi: 10.1175/jcli-d-12-00502.1.

S. Westra, B. Renard, and M. Thyer. The ENSO–precipitation teleconnection and its modulation by the interdecadal pacific oscillation. *Journal of Climate*, 2015. doi: 10.1175/JCLI-D-14-00722.1.

O. Wilhelmi and D. Wilhite. Assessing Vulnerability to Agricultural Drought: A Nebraska Case Study. *Natural Hazards*, 2002. doi: 10.1023/A:1013388814894.

P. Willems. Stochastic description of the rainfall input errors in lumped hydrological models. *Stochastic Environmental Research and Risk Assessment*, 15(2), 2001.

WMO. *Manual on Stream Gauging, Vol. II: Computation of discharge.* Manual on Stream Gauging. World Meteorological Organization (WMO), 2010.

D. P. Wright. *Influence diagnostics in hydrological modeling.* PhD thesis, The University of Adelaide, Australia, 2017.

D. P. Wright, M. Thyer, S. Westra, B. Renard, and D. McInerney. A generalised approach for identifying influential data in hydrological modelling. *Environmental Modelling & Software*, 2019. doi: https://doi.org/10.1016/j.envsoft.2018.03.004.

B. Xiong, L. Xiong, S. Guo, C. Xu, J. Xia, Y. Zhong, and H. Yang. Nonstationary frequency analysis of censored data: A case study of the floods in the yangtze river from 1470 to 2017. *Water Resources Research*, 2020. doi: 10.1029/2020WR027112.

P. Young. Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Modelling & Software*, 13(2), 1998.

A. Zeroual, M. Meddi, and A. A. Assani. Artificial Neural Network Rainfall-Discharge Model Assessment Under Rating Curve Uncertainty and Monthly Discharge Volume Predictions. *Water Resources Management*, 2016. doi: 10.1007/s11269-016-1340-8.

J. Zscheischler, S. Westra, B. J. J. M. van den Hurk, S. I. Seneviratne, P. J. Ward, A. Pitman, A. AghaKouchak, D. N. Bresch, M. Leonard, T. Wahl, and X. Zhang. Future climate risk from compound events. *Nature Climate Change*, 8(6):469–477, 2018. doi: 10.1038/s41558-018-0156-3.