



This is a repository copy of *Sampling-based methods for uncertainty propagation in flood modeling under multiple uncertain inputs: finding out the most efficient choice*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/201631/>

Version: Published Version

---

**Article:**

Hajihassanpour, M. [orcid.org/0000-0002-9708-2393](https://orcid.org/0000-0002-9708-2393), Kesserwani, G. [orcid.org/0000-0003-1125-8384](https://orcid.org/0000-0003-1125-8384), Pettersson, P. et al. (1 more author) (2023) Sampling-based methods for uncertainty propagation in flood modeling under multiple uncertain inputs: finding out the most efficient choice. *Water Resources Research*, 59 (7). e2022WR034011. ISSN 0043-1397

<https://doi.org/10.1029/2022wr034011>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Water Resources Research<sup>®</sup>



## RESEARCH ARTICLE

10.1029/2022WR034011

# Sampling-Based Methods for Uncertainty Propagation in Flood Modeling Under Multiple Uncertain Inputs: Finding Out the Most Efficient Choice

Mahya Hajihassanpour<sup>1</sup> , Georges Kesserwani<sup>1</sup> , Per Pettersson<sup>2</sup>, and Vasilis Bellos<sup>3</sup>

<sup>1</sup>Department of Civil and Structural Engineering, University of Sheffield, Sheffield, UK, <sup>2</sup>NORCE Norwegian Research Centre, Bergen, Norway, <sup>3</sup>Laboratory of Ecological Engineering and Technology, Department of Environmental Engineering, Democritus University of Thrace, Xanthi, Greece

### Key Points:

- Compared five uncertainty propagation methods to reduce sample size over Standard Monte Carlo method for reproducing detailed histograms
- Detailed histograms of Flood Extent and average/maximum hazard rating were compared instead of means and variances of those quantities
- The Quasi-Monte Carlo method outperforms the other methods for probabilistic flood modeling with two to five random input variables

### Correspondence to:

G. Kesserwani,  
[g.kesserwani@sheffield.ac.uk](mailto:g.kesserwani@sheffield.ac.uk)

### Citation:

Hajihassanpour, M., Kesserwani, G., Pettersson, P., & Bellos, V. (2023). Sampling-based methods for uncertainty propagation in flood modeling under multiple uncertain inputs: Finding out the most efficient choice. *Water Resources Research*, 59, e2022WR034011. <https://doi.org/10.1029/2022WR034011>

Received 1 NOV 2022

Accepted 4 JUL 2023

### Author Contributions:

**Conceptualization:** Mahya Hajihassanpour, Georges Kesserwani, Per Pettersson  
**Formal analysis:** Mahya Hajihassanpour, Georges Kesserwani, Per Pettersson  
**Funding acquisition:** Georges Kesserwani  
**Investigation:** Mahya Hajihassanpour  
**Methodology:** Mahya Hajihassanpour, Georges Kesserwani, Per Pettersson, Vasilis Bellos  
**Project Administration:** Georges Kesserwani  
**Software:** Mahya Hajihassanpour, Per Pettersson  
**Supervision:** Georges Kesserwani, Per Pettersson  
**Visualization:** Mahya Hajihassanpour

© 2023. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

**Abstract** In probabilistic flood modeling, uncertainty manifests in frequency of occurrence, or histograms, for quantities of interest, including the Flood Extent and hazard rating (HR). Such modeling at the field-scale requires the identification of a more efficient alternative to the Standard Monte Carlo (SMC) method that can reproduce comparable output probability distributions with a relatively reduced sample size, including detailed histograms of quantities of interest. Latin hypercube sampling (LHS) is the most evaluated alternative for fluvial floods but yields no considerable sample size reduction. Potentially better alternatives include adaptive stratified sampling (ASS), Quasi Monte Carlo (QMC) and Haar-wavelet expansion (HWE), which are yet unevaluated for probabilistic flood modeling. To fulfill this gap, LHS, ASS, QMC, and HWE are compared to quantify sample size reduction to reproduce output detailed histograms—for Flood Extent, and average and maximum HR—while keeping the difference below 10% to the reference SMC prediction. The comparison is done for two test cases with two (i.e., inflow discharge and Manning's coefficient) and three (i.e., further including the ground elevation) input random variables, and a real case with five input random variables. With two input random variables, all four alternatives yield sample size reductions, with QMC and HWE considerably outperforming the others; with three and more input random variables, HWE becomes inflexible and LHS underperforms. Still, QMC is a better choice than ASS to boost sample size reduction for the real case and shall be preferred in probabilistic flood modeling. Accompanying research codes are openly available online.

## 1. Introduction

Flood risks are predicted to increase under global warming, affecting people's health, environment, and socio-economic activities (Arnell & Gosling, 2016; Gu et al., 2022). Flood prediction uses a flood physical solver to estimate two-dimensional maps of flood-related quantities of interest, including the Flood Extent and/or the flood Hazard Rating (HR) that can be further used to assess impacts from the velocity magnitude (Kreibich et al., 2009; Nkwunonwo et al., 2020; Shirvani et al., 2021; Shirvani & Kesserwani, 2021). A deterministic flood physical solver does not account for the variability inherent in the input variables, due to measurement error uncertainty, which might lead to suboptimal decisions in flood management and mitigation strategies (Di Baldassarre et al., 2010). Probabilistic flood modeling has therefore become standard to propagate the variability in the input variables into the output quantities of interest. Probabilistic flood modeling can involve variability in many input variables, non-smooth responses in any of the simulated quantities of interest, which could lead to statistical distributions that are hard to capture such as discrete distributions with multiple modes (Aitken et al., 2022; Beevers et al., 2020; Shaw et al., 2021).

Ideally, an uncertainty quantification (UQ) framework involves experimental observations to first model the uncertain data (e.g., using the Generalized Likelihood Uncertainty Estimation (GLUE) (Beven & Binley, 1992) and Bluecat (Koutsoyiannis & Montanari, 2022)). The present UQ analysis framework assumes forward propagation of Type-B uncertainty, for which the information of the parameters is extracted from published resources, that is often defaulted to a uniform distribution (Apel et al., 2004). Then, forward propagation from uncertain input variables into the probabilistically simulated quantities of interest is often achieved using the SMC method (referred herein to as SMC) (Aronica et al., 1998; Dimitriadis et al., 2016; Huang & Qin, 2014; Jung & Merwade, 2012; Jung & Merwade, 2015; Karamouz & Fereshtehpour, 2019; McMillan & Brasington, 2008; Rahman et al., 2002; Stephens & Bledsoe, 2020).

**Writing – original draft:** Mahya Hajihassanpour, Georges Kesserwani  
**Writing – review & editing:** Mahya Hajihassanpour, Georges Kesserwani, Per Pettersson, Vasilis Bellos

Brute-force SMC requires a large sample size, referred hereafter to  $N_s$ , to accurately propagate the uncertainty in the input parameters, but this is not ideal for field-scale applications where the framework would remain computationally intensive even with a fast flood physical solver such as LISFLOOD-FP (Aitken et al., 2022; Beavers et al., 2020). This study aims to identify alternative-to-SMC UQ methods that use reduced  $N_s$  to reproduce Flood Extent and HR-related histograms within a maximum allowed relative difference of 10%. Existing studies are often limited to analyze the reduction in the  $N_s$  for one alternative to SMC based on the conventional statistics like the mean or the variance (for example, Fan et al., 2015; Smemoe et al., 2007). However, the use of conventional metrics may not be adequate to suggest reliable  $N_s$  reductions, because such metrics permit to overlook key information that could be present in complex output histograms (e.g., with more-than-one statistical peaks or multimodalities (Aitken et al., 2022; Beavers et al., 2020; Shaw & Kesserwani, 2020)). Hence, the quantification of the reduction in the  $N_s$  for probabilistic flood modeling would be more reliable when considering the metric of the relative histogram difference. Among the few papers that explored this aspect, Beavers et al. (2020) found that the Latin hypercube sampling (LHS) method (McKay et al., 1979) offers no reduction in  $N_s$  over SMC to reproduce the Flood Extent histograms for modeling of fluvial floods with a  $D$ -dimensional uncertainty space, with more-than-one input variables ( $D > 1$ ). The present UQ analysis framework is aimed to further assess other alternatives to SMC, which could better reduce  $N_s$  when still accurately reproducing Flood Extent and HR-related histograms, including for rapid floods.

A wide range of UQ methods have been developed in many research fields and can be grouped into non-sampling and sampling methods (Zhang, 2021). Non-sampling methods based on spectral expansions in random spaces, also called intrusive stochastic Galerkin methods, and entail a redesign of the deterministic physical solver (Ghanem & Spanos, 2003; Pettersson et al., 2015; Shaw & Kesserwani, 2020), introducing solver-specific complexity that is hardly applicable in practice (Ge et al., 2008; Shaw et al., 2020). In contrast, sampling methods are readily useable with any deterministic physical solver and can either be based on random sampling or on deterministic realizations (Hickernell, 2018; Zhang, 2021). Random sampling uses generators of pseudo-random sequences to sample the uncertainty space (Matsumoto & Nishimura, 1998). The resulting estimator is then itself random, and accurately estimating the convergence for a statistical metric may require replications, leading to unaffordable costs for field-scale simulations (Rubinstein & Kroese, 2016; Beavers et al., 2020). Methods using deterministic realizations do not need replications since they generate non-random sequences to sample the uncertainty space. In this work, four alternatives to SMC are investigated, two based on random sampling and two on deterministic realizations.

Accelerated Monte Carlo methods with random sampling rely on variance reduction techniques to reduce  $N_s$  over SMC (James, 1985; Owen, 2013). From this category, LHS is a popular alternative (McKay et al., 1979; Pharr et al., 2016; Zhu et al., 2017; Zokagoa et al., 2021), which has been widely studied in many research areas (Kalagnanam & Diwekar, 1997; Yu et al., 2001). Kucherenko et al. (2015) analyzed the performance of LHS to SMC to estimate three types of synthetic response functions, concluding that LHS can only reduce  $N_s$ , by 10 times, for one function type featured by low sensitivity to the variations in the input variables, retrieving the findings in Yu et al. (2001). However, LHS did not offer reduction in  $N_s$  for any other type of synthetic response functions, including test cases of Flood Extent histograms of fluvial floods (Aitken et al., 2022; Beavers et al., 2020; Kucherenko et al., 2015). Other random sampling techniques include the stratified sampling (Botev & Ridder, 2017; Giunta et al., 2006; Pharr et al., 2016) and control variate approaches (Giles, 2015; Glasserman, 2004; Pasupathy et al., 2012). Stratified sampling and its adaptive versions, so-called adaptive stratified sampling (ASS) (Eto et al., 2011; Pettersson & Krumscheid, 2022; Shields et al., 2015) was shown to offer promising reductions in  $N_s$  over SMC for a conventional statistical metric: For synthetic fluid dynamics problems, ASS yielded 100-to-1,000 times, 10-to-40 times, and 5-to-10 times smaller  $N_s$  with a two-, three- and four-dimensional uncertainty space ( $D = 2, 3$  and 4), respectively (Pettersson & Krumscheid, 2022). Still, the ability of ASS is unassessed for probabilistic flood modeling. Alternatively, control variate approaches include Multi-Level Monte Carlo (Giles, 2015; Heinrich, 1998; Müller et al., 2013) and Multi-Fidelity Monte Carlo (Pasupathy et al., 2012; Ng & Willcox, 2014; Peherstorfer et al., 2016). For an academic flow reservoir problem with one-dimensional uncertainty space ( $D = 1$ ), Multi-Level Monte Carlo yielded 6-to-18 times smaller  $N_s$  against SMC for the conventional statistical metrics (Müller et al., 2013). However, the Multi-Level Monte Carlo approach deploys many physical grids of different resolution supported by a sensitivity analysis, which is test-specific, making this approach burdensome for probabilistic flood modeling including high dimensional uncertainty spaces ( $D > 1$ ) (Aitken et al., 2022). In contrast, the Multi-Fidelity Monte Carlo approach fixes the

grid resolution but varies the complexity in the physical solver(s), leading to 3-to-10 times smaller  $N_s$  against SMC (Peherstorfer et al., 2016). As this work aims to identify efficient alternatives to SMC for the same (fixed) physical solver and grid resolution, the control variate approaches are not considered.

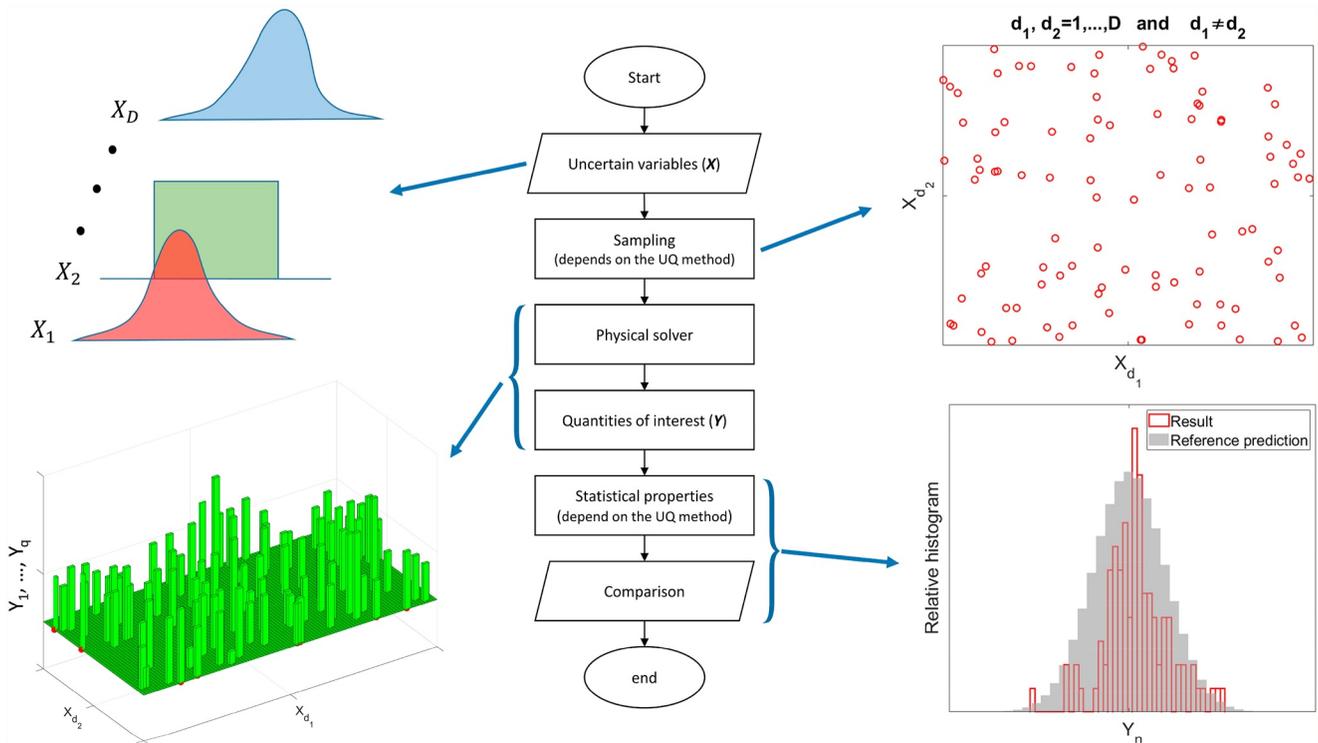
Deterministic realization methods include the Quasi Monte Carlo (QMC) and non-intrusive stochastic collocation approaches (Zhang, 2021). QMC (Morokoff & Caflisch, 1995; Wong et al., 1997) can be applied to generate low-discrepancy sequences, or quasi-random samples, for sampling the uncertainty space, which can be based on Sobol (Sobol, 1979), Halton (Halton, 1960) or Hammersley (Hammersley, 1960) sequences. For financial modeling, Chatzivasileiadis (2018) showed that QMC with Halton- and Sobol-based sequences yield 8 times and 2-to-5 times smaller  $N_s$  against SMC for the conventional statistical metrics, respectively. However, QMC with Hammersley-based sequences could lead to up to 40 times smaller  $N_s$  for high-dimensional uncertainty spaces ( $2 \leq D \leq 100$ ) (Kalagnanam & Diwekar, 1997; Kucherenko et al., 2015; Wang et al., 2004). Therefore, only the QMC alternative with the Hammersley-based sequences is considered for probabilistic flood modeling. In the non-intrusive stochastic collocation approach (Avasarala & Subramani, 2021; Eldred, 2009; Zio & Rochinha, 2012), global orthogonal polynomials can be used as continuous basis functions to span the uncertainty space (Shaw & Kesserwani, 2020; Xiu & Karniadakis, 2002). However, this choice is not suited for probabilistic flood modeling that needs an accurate representation of non-smooth responses in the quantities of interest that manifest in discrete histograms with multimodalities (Abgrall & Mishra, 2017; Shaw et al., 2020). Using discrete basis functions of the Haar-wavelet expansion (HWE) (Le Maître et al., 2004) remedies this issue, as exemplified by capturing critical physics of wetting and drying in the histograms using 4 times smaller  $N_s$  against SMC for a three-dimensional uncertainty space ( $D = 3$ ) (Shaw et al., 2020). So far, HWE was only evaluated for diagnostic river-channel flow problems; and, like ASS and QMC, its performance in the context of probabilistic flood modeling is yet to be assessed.

In this paper, we examine the capability of LHS, ASS, QMC, and HWE to efficiently capture relative histograms for Flood Extent, average and maximum HR ( $HR_{ave}$  and  $HR_{max}$ ) with reduced  $N_s$  against SMC. Section 2 describes the UQ analysis framework including details on: the generation of input random variables (Section 2.1); the choice of the physical solver used for the probabilistic runs and definitions of Flood Extent,  $HR_{ave}$  and  $HR_{max}$  (Section 2.2); the technical structure of alternative-to-SMC UQ methods (Section 2.3); the comparison approach based on the relative histogram difference against the reference SMC prediction (Section 2.4); and the definition, purpose and characteristics of validation test cases, including uncertainty spaces with  $2 \leq D \leq 5$ , considering torrential and fluvial flood types (Section 2.5). In Section 3, the comparative analysis is presented to identify the reduction in  $N_s^K$ , per alternative-to-SMC UQ method  $K$  ( $K =$  QMC, HWE, LHS, and ASS), to keep a difference below the target threshold of  $7.5\% \pm 2.5\%$  based on the relative histogram metric. Section 4 discusses the limitations and general applicability of the present study, and Section 5 summarizes the key findings and concludes on the most efficient UQ method for probabilistic flood modeling. In Appendix A, the comparative analysis is reproduced using the difference based on the standard errors of the mean and the variance, showing how the use of these conventional metrics lead to an over-optimistic reduction in  $N_s^K$ , compared to the relative histogram metric. Appendix B gives instructions for running the open research codes that can be used to reproduce this comparative study (Hajihassanpour et al., 2022).

## 2. UQ Analysis Framework

The present uncertainty propagation framework is described in Figure 1. The uncertainty in the values of the input parameters is represented by a vector of random variables,  $X = [X_1, \dots, X_D]$ , in which each random variable is assigned a probability distribution that should reflect the information available about the corresponding input parameter (Figure 1, upper left). According to Haan et al. (1998), the choice of which probability distributions to adopt to model the uncertain hydrological parameters is less important than acquiring good estimates of their means and standard deviations. In this work, therefore, the input random variables are generated based on given means and standard deviations from surveyed uncertainty ranges in published resources (Section 2.1), and using uniform, or rectangular, probability distributions are recommended (<https://www.isobudgets.com/type-a-and-type-b-uncertainty/#type-b-uncertainty-definition>; <https://physics.nist.gov/cuu/Uncertainty/typeb.html>) to conservatively model such Type-B uncertainty (i.e., with equal probabilities) (Apel et al., 2004).

According to the prescribed distributions, the input random variables are then sampled, with either a random sampling (e.g., Figure 1, upper-right) or deterministic sampling, depending on the UQ method (Section 2.3). A



**Figure 1.** The uncertainty quantification analysis framework. Upper-left: pre-defined uncertain variables, via probability distribution functions, spanning a  $D$ -dimensional uncertainty space; upper-right: a typical (random) sampling in the uncertainty space; lower-left: the output of a probabilistic run, after applying a physical solver for all samples, for the quantities of interest; lower-right: a relative histogram for a quantity of interest for evaluation against a reference prediction (brute-force standard Monte Carlo runs).

probabilistic run is achieved by applying a physical solver to all the samples ( $N_s$  simulations), leading to output samples for selected quantities of interest  $Y = [Y_1, \dots, Y_q]$  (Figure 1, lower-left). Histogram(s) for  $Y$  are generated for each UQ method, normalized to  $N_s$  to become relative histogram(s), to assess  $N_s$  reduction such that to keep the difference against the reference prediction (Figure 1, lower-right) below the target threshold of  $7.5\% \pm 2.5\%$  (Section 2.4).

### 2.1. Generation of the Input Random Variables

In probabilistic flood modeling, the three most significant uncertain variables are the inflow discharge (forcing driver), Manning coefficient (physical driver), and ground elevation (geometric driver) (Alipour et al., 2022; Bellos & Tsihrintzis, 2021; Dimitriadis et al., 2016; Jung & Merwade, 2012; Jung & Merwade, 2015; Savage et al., 2016; Tscheikner-Gratl et al., 2019). Here, each uncertain (i.e., random input) variable is assumed to result from the same measurement error at any point in time for the inflow discharge(s) and at any point in space for the Manning coefficient and the ground elevation. With this assumption, the dimensionality of the uncertainty space reduces to the number of random input variables to make the uncertainty propagation framework computationally feasible. Otherwise, non-correlated, cell-wise variations in the ground elevation would require a distinct random variable for every cell, leading to an unfeasible increase in the dimensionality of the uncertainty space (Stefanescu et al., 2012; Wechsler, 2007).

The uncertainty in a given mean inflow discharge,  $\bar{Q}(t)$  at a time instant  $t$ , results from measurement errors in the estimation of rainfall-runoff, hydrological data and/or hydrometric data (Sharafati et al., 2020). This uncertainty is often assigned to a 16% range (Bates et al., 2011). Therefore, a uniform inflow discharge random variable,  $Q(t, \xi_Q)$ , follows:

$$Q(t, \xi_Q) = \bar{Q}(t) + \xi_Q \sigma_Q(t) \quad (1)$$

where  $\xi_Q$  is a random variable taking values in  $[-1, +1]$  and  $\sigma_Q(t) = 0.08\bar{Q}(t)$  is the range of variation with respect to the mean  $\bar{Q}(t)$ . In flood modeling, a number of  $P$  mean inflow discharges,  $\bar{Q}_1(t), \dots, \bar{Q}_P(t)$  ( $P > 1$ ), can be given (Kesserwani & Sharifian, 2023; Sharifian et al., 2023). In this case, the variation in each of  $Q_1(t, \xi_{Q_1}), \dots, Q_P(t, \xi_{Q_P})$  follows Equation 1, assuming that  $\xi_{Q_1}, \dots, \xi_{Q_P}$  are not intercorrelated (Neal et al., 2013), leading to an uncertainty space dimension of  $D = P$ .

Although measurement error uncertainty for the discharge has a strong impact on the outputs (Bellos & Tsihrintzis, 2021), the uncertainty in the spatial Manning coefficient can also be significant, such as for fluvial floods (Alipour et al., 2022; Beevers et al., 2020; Dimitriadis et al., 2016). The Manning coefficient's uncertainty depends on friction elements (e.g., vegetation, land, or soil type) and is often assigned a range of 10%, from bibliographical values and calibration to match observed flood-related data (Bellos et al., 2018; Neal et al., 2009). Therefore, the uniform Manning coefficient random variable follows:

$$n(x, y, \xi_n) = \bar{n}(x, y) + \xi_n \sigma_n(x, y) \quad (2)$$

where  $\xi_n$  is a random variable taking values in  $[-1, +1]$ ,  $\bar{n}(x, y)$  is a given mean Manning coefficient—that can be spatially varying to take a different value at each spatial point  $(x, y)$ —and  $\sigma_n(x, y) = 0.05 \bar{n}(x, y)$  is the range of variation with respect to  $\bar{n}(x, y)$ . For the considered test cases (Section 2.5), a constant mean Manning coefficient value is used, thus further assuming  $\bar{n}(x, y) = \bar{n}$  to reduce  $n(x, y, \xi_n)$  to  $n(\xi_n)$ . Note that, Equation 2 can be further used considering spatially varied mean Manning coefficient data,  $\bar{n}(x, y)$ , incorporating many land use types and channels (Sharifian et al., 2023), without loss of generality. However, using spatially varied roughness would lead to similar performance as with a uniform constant roughness (Alipour et al., 2022).

It is sometimes necessary to also include uncertainty in the mean ground elevation variable,  $\bar{z}(x, y)$ , with respect to which the range of variation  $\sigma_z(x, y)$  can be generated from the analysis of the Digital Elevation Model (DEM) data. The range of variation  $\sigma_z(x, y)$  is often assigned a measurement error uncertainty as high as 10% depending on the quality of the DEM data (Hu et al., 2009; Liu et al., 2015; Nalbantis et al., 2017; Shaw et al., 2020; West et al., 2018). Therefore, the uncertainty in  $\bar{z}(x, y)$  may be significant in some locations, informed by local estimates of  $\sigma_z(x, y)$ . Hu et al. (2009) and Liu et al. (2015) estimated  $\sigma_z(x, y)$  for LiDAR-based DEMs as a function of the diagonal length of the DEM's cell size, denoted by  $l_d$ , and the local curvature for the DEM's value at this cell, denoted by  $M(x, y)$ . Their estimates suggest significant  $\sigma_z(x, y)$  either when the DEM resolution is coarse or when the DEM's cells represent a topographic area with large curvatures (e.g., riverbanks or urban buildings). Furthermore, Hu et al. (2009) estimated  $\sigma_z(x, y)$  to  $|\frac{1}{8} M(x, y) l_d^2|$  using a linear interpolation of point cloud acquired by LiDAR sensors, whereas Liu et al. (2015) estimated it to  $|\frac{3}{8} M(x, y) l_d^2|$  using an interpolation based on Triangulated Irregular Network. As these estimates for  $\sigma_z(x, y)$  have different weights for  $M(x, y) l_d^2$  to analyze distinct DEM data types,  $\sigma_z(x, y)$  is here generally estimated to  $|c M(x, y) l_d^2|$ , where  $c$  is a user-defined weight identified by a DEM-specific sensitivity analysis. For the test cases considered (Section 2.5), the sensitivity analyses (not reported) suggest a  $c$  around 0.04 to 0.125 for realistic, uneven DEM data, and a higher value of 5 for idealistic, smooth DEM data. Note that, the selected choices for  $c$  were based on retrieving the target Root Mean Square Error of 0.3 m for smooth, semi-flat regions and the upper bound of 10% uncertainty for highly curved regions (Liu, 2011; Liu et al., 2015; Shaw et al., 2020; West et al., 2018). After estimating  $\sigma_z(x, y)$ , the uniform ground elevation random variable, for  $\xi_z$  varying between  $[-1, +1]$ , follows:

$$z(x, y, \xi_z) = \bar{z}(x, y) + \xi_z \sigma_z(x, y) \quad (3)$$

## 2.2. Physical Solver and Quantities of Interest

For a selected UQ method at a fixed  $N_s$ , a physical solver should be employed to make a probabilistic run, or ensemble of  $N_s$  simulations, to propagate the variations in the input variables (Equations 1–3) into output probability distributions. The first-order finite volume hydrodynamic solver of the LISFLOOD-FP suite was employed, using the version parallelized on Graphical Processing Units (GPU), so-called GPU-FV1 (Shaw et al., 2021). At the time, GPU-FV1 was the fastest physical solver available to minimize runtime cost per simulation for large spatial domains at fine DEM resolutions [that is, with more than half a million computational cells such as with two of the test cases investigated (Sections 2.5.2 and 2.5.3)]. For such simulations, GPU-FV1 is 30 times faster than the CPU version (CPU-FV1) on 16 threads (Shaw et al., 2021). Using the less mathematically complex

solver—neglecting convective acceleration terms—though a popular choice for UQ analysis (Beevers et al., 2020), had only a CPU version i.e., slower to run than GPU-FV1 for a simulation on grids with more than half a million cells (Shaw et al., 2021). Hence, GPU-FV1 is a mathematically complete and fast physical solver that can reproduce all types of hydrodynamic flow regimes and responses as accurately as second-order finite volume solvers for realistic floodplain flow simulations (Ayog et al., 2021; Zhao & Liang, 2022).

The probabilistic run leads to output samples for the water depth,  $h(x,y,t)$  and velocity field magnitude,  $V(x,y,t)$ , that are post-processed into the following quantities of interest.

- *Flood Extent.* It is the sum of the area of the computational cells with non-zero water depth  $h$ , that is, total wet area. The Flood Extent has often been used in probabilistic flood modeling to analyze Flood Extent frequencies (Bermúdez et al., 2017; Beevers et al., 2020; Aitken et al., 2022).
- *Average Hazard Rate ( $HR_{ave}$ ) and Maximum Hazard Rate ( $HR_{max}$ ).* The flood HR is defined as:  $HR = h \times (V + 0.5)$ ; and, it has been used to provide more information on velocity impacts to assets including structural damage to residential buildings, damages to road infrastructures, and risks to people's life and injury (Kreibich et al., 2009; Lumbroso & Davison, 2018; Maranzoni et al., 2022; Ramsbottom et al., 2006; Shirvani & Kesserwani, 2021; Shirvani et al., 2021). In this study,  $HR_{ave}$  and  $HR_{max}$  quantify the average and maximum flood HR over the computational area.
- *Floodtiming.* The quantities Flood Extent,  $HR_{ave}$  and  $HR_{max}$  are evaluated at the output floodtiming,  $t$ , at which the maximum Flood Extent is reached. At this time, the variation in the input random variables should have propagated complex responses into the output samples that in turn manifest in complex histograms representing the statistics of these quantities of interest.

A computational load of 5.5 million runs resulted in this study (that is, to compare five different UQ methods (Section 2.3) with convergence analyses for a range of  $N_s$  (Section 2.4) and considering three test cases (Section 2.5). This load was distributed across four NVIDIA Tesla V100 GPU cards (maximum) available on the Bessemer HPC cluster of the University of Sheffield ([https://docs.hpc.shef.ac.uk/en/latest/bessemer/cluster\\_specs.html#bessemer-specs](https://docs.hpc.shef.ac.uk/en/latest/bessemer/cluster_specs.html#bessemer-specs)). On such a GPU card, the lengthiest runtime for a run using GPU-FV1 did not exceed 7 min, allowing to make a probabilistic run for a batch of simulations, per UQ method and per  $N_s$ , in the allowable 7-day window of non-interrupted access to each GPU card. Meeting this limit was not possible with CPU-FV1 for grids involving more than half a million cells, where the elapsed runtime for some batches can exceed 7 days. Note that, it is possible to use CPU-FV1 to distribute the runs over up to 100 CPUs on the cluster, but distributing the runs of GPU-FV1 on the four GPU cards results in a 20% faster overall runtime (since CPU-FV1 is 30 times slower per run).

### 2.3. Alternative-To-SMC UQ Methods

In SMC the order of convergence is inversely proportional to  $N_s^{1/2}$ , requiring a large  $N_s$  to reproduce a true reference prediction (e.g., known analytical functions (Kucherenko et al., 2015)). Four selected alternative-to-SMC UQ methods are compared to potentially reduce  $N_s$  while keeping a relative histogram difference below 10% against the reference SMC prediction for the most relevant quantity of interest. The alternatives are HWE and QMC, from the deterministic realization methods, and ASS and LHS from the random sampling methods.

- *HWE.* This method describes the quantities of interest  $Y$  by expansion coefficients  $\hat{Y}$ , projected onto the discrete Haar-Wavelet basis  $\phi$ , as  $Y = \sum_{j=1}^{N_s} \hat{Y}_j \phi_j(X)$ . In the current implementation of HWE, the realizations in the uncertainty space are equidistantly, but sparsely distributed with a spacing of  $2^{DL}$  where  $L$  is the refinement level characteristic of the Haar-Wavelet basis (Shaw et al., 2020). This implies that for the same  $N_s$ , HWE leads to coarser sampling in the uncertainty space as  $D$  is increased. Shaw et al. (2020) assessed HWE for diagnostic river-channel flow problems. Here, the potential of HWE is assessed for realistic floodplain flow test cases.
- *Quasi Monte Carlo (QMC).* The QMC method uses Hammersley's space-filling sequences to evenly sample the uncertainty space. Since its order of convergence is  $(\log N_s)^{D-1} N_s^{-1}$ , it has a potential to significantly reduce  $N_s$  over SMC for uncertainty spaces with moderate dimensionality (Chatzivasileiadis, 2018; Kalagnanam & Diwekar, 1997; Wong et al., 1997; Zhang, 2021). This potential for QMC is here specifically investigated for probabilistic flood modeling.

- **ASS.** ASS is a quite recent UQ method (Pettersson & Krumscheid, 2022), which is here explored for probabilistic flood modeling. ASS is designed to adaptively improve the performance of SMC by sensing the local behavior in the variance of a scalar output, which is here selected to be any of the quantities of interest. According to the local variance estimates, the uncertainty space is split into disjoint subsets, or strata, where each stratum is sub-sampled independently. Among all strata, stratification is adaptively applied to maximize variance reduction. Stratification consists of splitting one existing stratum into two in the regions of high variance or leaving it intact otherwise. After an iteration of stratification,  $N_{\text{new}}$  sub-samples are added to be distributed between strata. With ASS,  $N_{\text{new}}$  and a parameter  $\alpha$  must be specified by the user (i.e., to  $N_{\text{new}} = 30$  and  $\alpha = 0.9$  following (Pettersson & Krumscheid, 2022)). The performance of ASS can only be analyzed for a single scalar output and, therefore, may need to be explored with each of the quantities of interest.
- **LHS.** In LHS, each dimension in the uncertainty space is partitioned into  $N_s$  intervals of equal length, and one-dimensional pseudo-random samples are generated such that there is exactly one sample in every interval. These marginal samples are then randomly combined, one from each dimension, into multivariate samples that preserve the space-filling property of the marginal distribution (McKay et al., 1979). LHS is the most explored alternative-to-SMC (Pharr et al., 2016; Zhu et al., 2017; Zokagoa et al., 2021) for probabilistic modeling of fluvial floods (Aitken et al., 2022; Beevers et al., 2020). Here, LHS is also explored for rapid floods to identify whether any of HWE, QMC or ASS can be a better alternative-to-SMC.

#### 2.4. Comparison Approach Using Relative Histograms

Compared to the conventional statistics like the mean or the variance, a histogram can inform on key aspects contributing to the overall statistics, such as multimodalities in a frequency of occurrence, which can otherwise be missed (i.e., as shown in (Shaw & Kesserwani, 2020)).

For a quantity of interest, among Flood Extent,  $HR_{\text{ave}}$ , and  $HR_{\text{max}}$ , the relative histogram predicted by the UQ method  $K$  for a fixed  $N_s^K$  ( $K = \text{SMC}, \text{QMC}, \text{HWE}, \text{LHS}, \text{and ASS}$ ) is compared to the relative histogram of the reference prediction—produced using SMC with a  $N_s$  that is 2.5-to-320 times larger than  $N_s^K$ , depending on the affordability of SMC per test case to make the probabilistic run (i.e., to produce the reference prediction). Comparing the difference between these two relative histograms may not be straightforward. On the one hand, a relative histogram is sensitive to the number of bins,  $N_{\text{bins}}$ , as too wide bins can overlook multimodalities. As a rule of thumb,  $N_{\text{bins}}$  depends on the  $N_s^K$ , which here ranges between 125 and 4,096 (Section 2.5.1). For this range of  $N_s^K$ , although Sturge's rule (<https://medium.datadriveninvestor.com/how-to-decide-on-the-number-of-bins-of-a-histogram-3c36dc5b1cd8>) calculates  $8 \leq N_{\text{bins}} \leq 13$ , a  $N_{\text{bins}}$  as large as 40 was reported to analyze Flood Extent frequency occurrence (Beevers et al., 2020). Therefore, the three values for  $N_{\text{bins}} = \{10, 20, 40\}$  are considered when comparing the difference between these two relative histograms. On the other hand, for a fixed  $N_{\text{bins}}$ , comparing two relative histograms, can either be bin-wise or via a cross-bin approach (Rubner et al., 2000). The bin-wise approach measures the difference bin-by-bin, and the cross-bin approach also incorporates the correlations from the differences at the neighboring bins. Although less sensitive to  $N_{\text{bins}}$ , the cross-bin approach tends to predict zero differences in regions of uniform frequency distributions (Rubner et al., 2000). Hence, the bin-wise approach is used to compare the two histograms, as follows (Stricker & Orengo, 1995):

$$\text{Relative histogram difference (\%)} = 100 \times \sum_{j=1}^{N_{\text{bins}}} |f_{\text{ref}_j} - f_j| \quad (4)$$

In Equation 4,  $f_j$  and  $f_{\text{ref}_j}$  denote the relative frequency (normalized by  $N_s^K$ ) inside the  $j$ th bin for the histogram predicted by the selected UQ method and for the histogram of the reference prediction, respectively.

The convergence analysis (Section 3.1) for the first test case (Section 2.5.1) uses Equation 4 with varying  $N_s^K$  to identify the  $N_s^K$  per UQ method that keeps relative histogram differences below the target threshold. The identified  $N_s^K$  per alternative-to-SMC UQ method  $K$  ( $K = \text{QMC}, \text{HWE}, \text{LHS}$  and  $\text{ASS}$ ) are then used to quantify the reduction in terms of the following relative-to-SMC speedup ratio:

$$\text{Speedup ratio} = N_s^{\text{SMC}} / N_s^K \quad (K = \text{QMC}, \text{HWE}, \text{LHS} \text{ and } \text{ASS}) \quad (5)$$

Note that, the quantified speedup ratios (Equation 5) from the convergence analysis (Section 3.1) are based on the relative histogram difference (Equation 4); in other words, not based on the standard errors of the mean and the

variance (for example, as in Yu et al. (2001), Smemoe et al. (2007), Fan et al. (2015), Kucherenko et al. (2015), Pettersson and Krumscheid (2022)), which lead to greatly larger speedup ratios (i.e., demonstrated in Appendix A vs. the results presented in Section 3.1). Worth also noting that a reliable convergence analysis required 60 times more runs, or replications, per random sampling method (i.e., SMC, ASS and LHS) to estimate converged differences for both the results presented in Section 3.1 and Appendix A. This number should be appropriately informed by Beevers et al. (2020) and Shirvani and Kesserwani (2021), where the authors reported making 10 and 20 replications to average acceptable differences with SMC and LHS, respectively.

## 2.5. Purpose, Definition, and Characteristics of the Test Cases

Three probabilistic flood modeling test cases are designed to assess the alternative-to-SMC UQ methods ( $K = \text{QMC, HWE, LHS and ASS}$ ) to potentially reduce  $N_s^K$  over  $N_s^{\text{SMC}}$  to keep a maximum threshold of 10% for the relative histogram difference (Equation 4), while also analyzing the effect of  $N_{\text{bins}} = \{10, 20, 40\}$ .

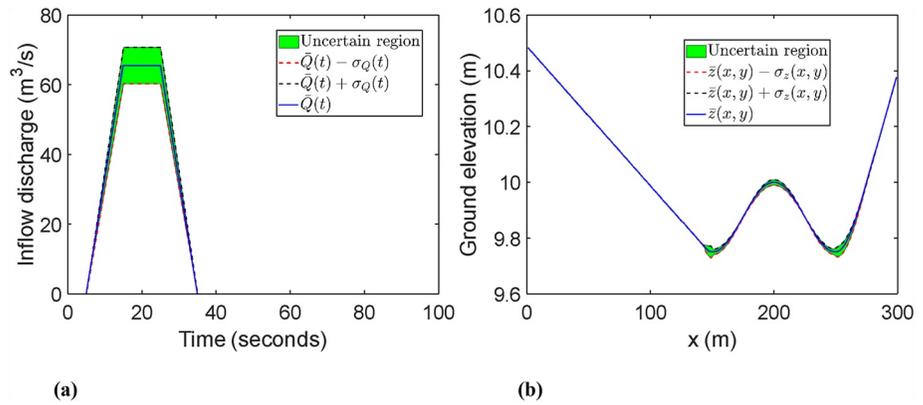
The first test case is a synthetic “Rapidly propagating flood over a smooth terrain” (Section 2.5.1). It is employed to conduct a detailed convergence analysis for each of the quantities Flood Extent,  $\text{HR}_{\text{ave}}$  and  $\text{HR}_{\text{max}}$ , for two sub-cases:  $D = 2$  for the two input random variables:  $Q(t, \xi_Q)$  and  $n(\xi_n)$ ; and,  $D = 3$  for the three input random variables:  $Q(t, \xi_Q)$ ,  $n(\xi_n)$  and  $z(x, y, \xi_z)$ . For computational feasibility, the convergence analysis is only carried out for this test case to identify the  $N_s^K$  that keeps the relative histogram difference (Equation 4) below the target threshold for the most uncertain quantity, then accordingly quantify the speedup ratios (Equation 5). Note that, making the replications required 4,830,862 probabilistic runs for the range of  $N_s^K$  considered in the first test case (Section 2.5.1), for which a run costs 3 s, leading to a total of runtime of 168 days. Making these replications was unfeasible for the second (Section 2.5.2) and third (Section 2.5.3) test cases, costing 100 and 400 s per run, which would require a runtime cost of 5,600 and 22,400 days, respectively.

The second test case is a more realistic “Torrential flooding over a rough river valley” (Section 2.5.2), also explored for the two sub-cases with  $D = 2$  and 3. It is aimed to further validate the identified choices for  $N_s^K$  to keep the average target threshold of 7.5% (Section 3.1). The validation re-measures the relative histogram difference (Equation 4) against the test-specific reference prediction, to consider a UQ method valid if it keeps this difference below the maximum target threshold of 10%. Accordingly, the best alternative-to-SMC from the deterministic realization methods (QMC and HWE) and from the random sampling methods (LHS and ASS) is selected for further validation over a real case study of a fluvial flooding. The third test case assesses the validity of the two selected candidates (QMC and ASS) for probabilistic modeling of the “Carlisle 2005 flooding” (Section 2.5.3). This test case is featured by a  $D = 5$  including three input random discharge variables (Section 2.1),  $Q_p(t, \xi_{Q_p})_{p=1,2,3}$ , as well as the input random variables for the Manning coefficient and the ground elevation,  $n(\xi_n)$  and  $z(x, y, \xi_z)$ . The assessment considers two choices for  $N_s^K$  (i.e., justified in Sections 3.1 and 3.2) to further validate QMC and ASS against the test-specific reference prediction.

### 2.5.1. Rapidly Propagating Flood Over a Smooth Terrain

This synthetic test case (Neelz & Pender, 2013) has been used to benchmark physical solvers for a rapidly propagating flow occurring over a short runtime (Ayog et al., 2021; Cea Gómez et al., 2020; De Luna-Cruz et al., 2019). The mean inflow discharge,  $\bar{Q}(t)$ , is  $65.5 \text{ m}^3 \text{ s}^{-1}$  and its variation  $Q(t, \xi_Q)$  is shown in Figure 2a. The inflow discharge variable  $Q(t, \xi_Q)$  enters from the left boundary into a small domain area of  $0.3 \text{ km} \times 0.1 \text{ km}$  that has a mean Manning coefficient  $\bar{n} = 0.01$  with a variation of  $n(\xi_n)$ . The mean ground elevation variable  $\bar{z}(x, y)$  is taken from a pseudo-two-dimensional DEM at 2 m resolution (finest available) and its variation  $z(x, y, \xi_z)$  is based on estimates for  $\sigma_z(x, y)$  using  $c = 5$  (Section 2.1). In Figure 2b, the variation in  $z(x, y, \xi_z)$  is detectable at the points of strong topographic curvature. The flood timing is 193 s, when a simulation stops for analysis of Flood Extent,  $\text{HR}_{\text{ave}}$  and  $\text{HR}_{\text{max}}$ . Using 7,500 computational cells, a single simulation took 3 s and the probabilistic SMC run used  $N_s = 40,000$  to produce the reference prediction.

From the reference prediction,  $D$ -dimensional response surfaces for each of Flood Extent,  $\text{HR}_{\text{ave}}$  and  $\text{HR}_{\text{max}}$  can be produced to analyze how the variations in the input variables (from their mean values) propagate into each of the quantities of interest. Note that, a response surface is a metamodel created via the “response surface methodology” using numerical experiments (i.e., the SMC simulations) to understand the non-smoothness level in the response behavior to the random variables within the uncertainty space. Qualitatively, the response surface projections inform on which are the uncertain variables that the quantities of interest are more sensitive to (Sections 2.5),



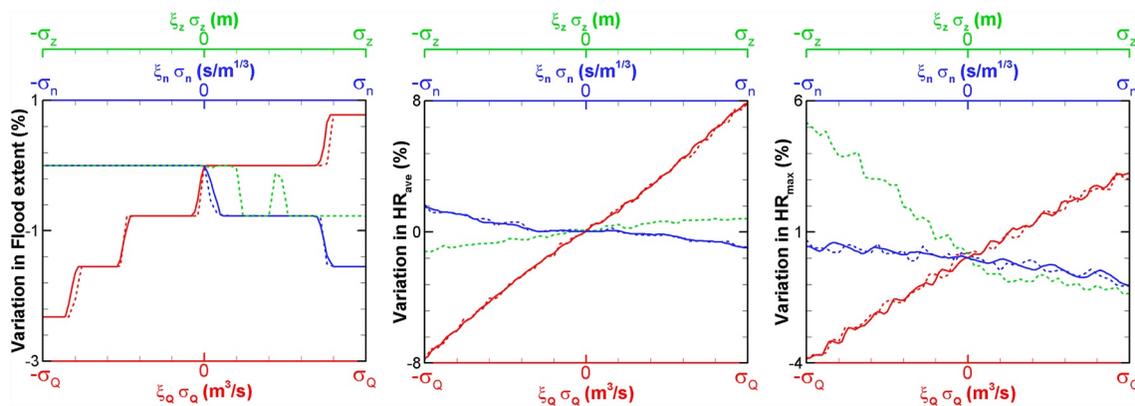
**Figure 2.** Rapidly propagating flood over a smooth terrain (Section 2.5.1). Variation in (a) the mean inflow discharge variable and (b) the mean ground elevation variable.

and therefore should retrieve the conclusions from other studies that quantitatively assess the sensitivity of flood model outputs including Sobol indices (for example, Alipour et al., 2022; Savage et al., 2016).

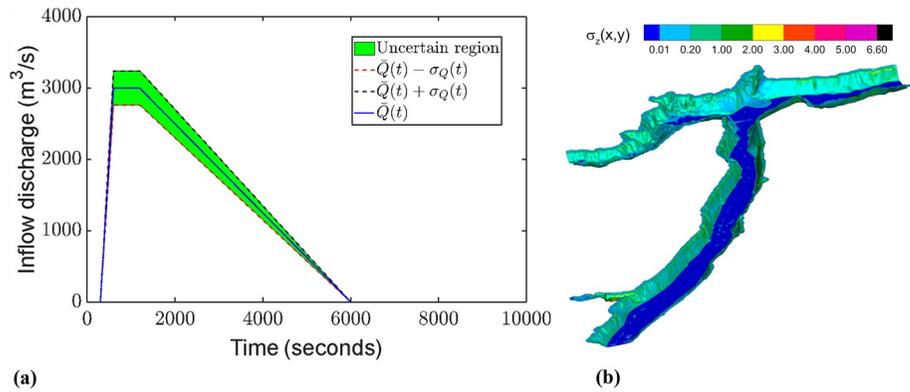
Figure 3 shows the plots of the centerline projections of the response surfaces of each random variable for the sub-cases with  $D = 2$  (solid lines) and with  $D = 3$  (dotted lines). With  $D = 2$ , larger variations are seen due to the inflow discharge variable compared to variations due to the Manning coefficient variable. For Flood Extent, the variations due to these two variables are few, flat states connected by steep jumps, that are more frequent for the inflow discharge variable. For  $HR_{ave}$ , the variations are quite linear; but the variation due to the inflow discharge variable is steeper and far more deviates from the close-to-flat variation due to the Manning coefficient variable. For  $HR_{max}$ , though the variations have a lower range of variability, they display non-smooth waviness compared to  $HR_{ave}$ . Its lower range of variability arise from the fact that Floodtiming (Section 2.2) is closer to the time when the maximum of  $HR_{ave}$  is reached, than to the time when  $HR_{max}$  reaches its maximum; this explains why  $HR_{max}$  is less sensitive than  $HR_{ave}$  to the variations in the input random variables.

With  $D = 3$ , the variations in both dimensions of the inflow discharge variable and the Manning coefficient remain quite unchanged for all the quantities of interest. In the ground elevation variable, the variations for Flood Extent and  $HR_{ave}$  are close to zero, thus relatively insignificant, which is in contrast with the variation for  $HR_{max}$  that is noticeably large, significantly adding up to the overall  $HR_{max}$  variation. Overall, the analysis of Figure 3 indicates that  $HR_{max}$  is the most uncertain quantity of interest in this test, in which the responses for Flood Extent and  $HR_{ave}$  are similar for both sub-cases (with  $D = 2$  and 3).

A range of choices for  $N_s^K$  per UQ method is needed to conduct the convergence analysis, including 500 samples per uncertain input variable as recommended in Beavers et al. (2020). For HWE, the range of  $N_s^{HWE} = \{64,$



**Figure 3.** Rapidly propagating flood over a smooth terrain (Section 2.5.1). Centerline plots of the projections from the  $D$ -dimensional response surfaces for Flood Extent,  $HR_{ave}$  and  $HR_{max}$ , showing their variations in each uncertainty dimension for the case with  $D = 2$  (solid lines) and the case with  $D = 3$  (dotted lines).



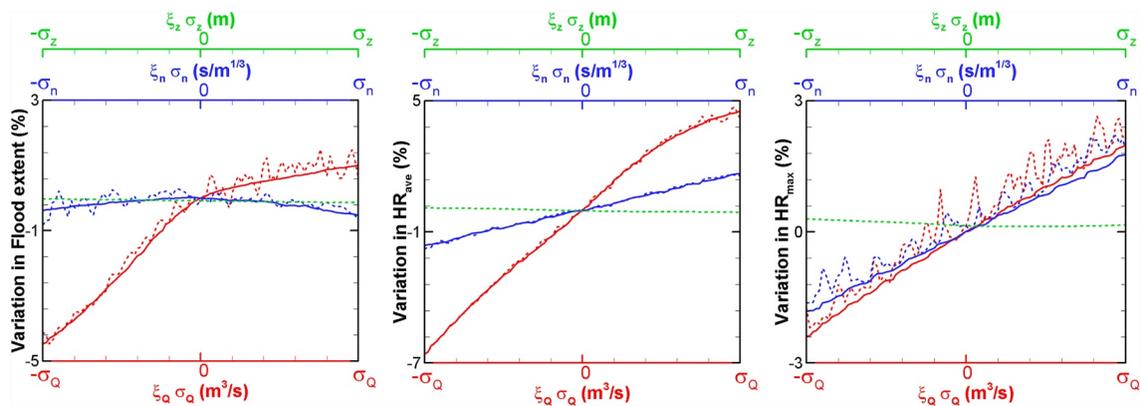
**Figure 4.** Torrential flooding over a rough river valley (Section 2.5.2). Variation in (a) the mean inflow discharge variable and (b) the mean ground elevation variable.

256,1,024, 4,096} is used with  $D = 2$  (i.e., refinement levels of  $L = \{3, 4, 5, 6\}$ ) and of  $N_s^{HWE} = \{64, 512, 4,096\}$  with  $D = 3$  (i.e., refinement levels of  $L = \{2, 3, 4\}$ ) (Shaw et al., 2020). The other UQ methods (SMC, LHS, ASS and QMC) can be given the same range of  $N_s^K = \{125, 250, 500, 1,000, 2,000, 4,000\}$ . In Section 3.1, the identified  $N_s^K$  that keep the relative histogram difference below the average threshold of 7.5% are discussed, together with quantifications of their relative-to-SMC speedup ratios and analysis of their sampling patterns in the uncertainty space and ability to reproduce different shapes of relative histograms.

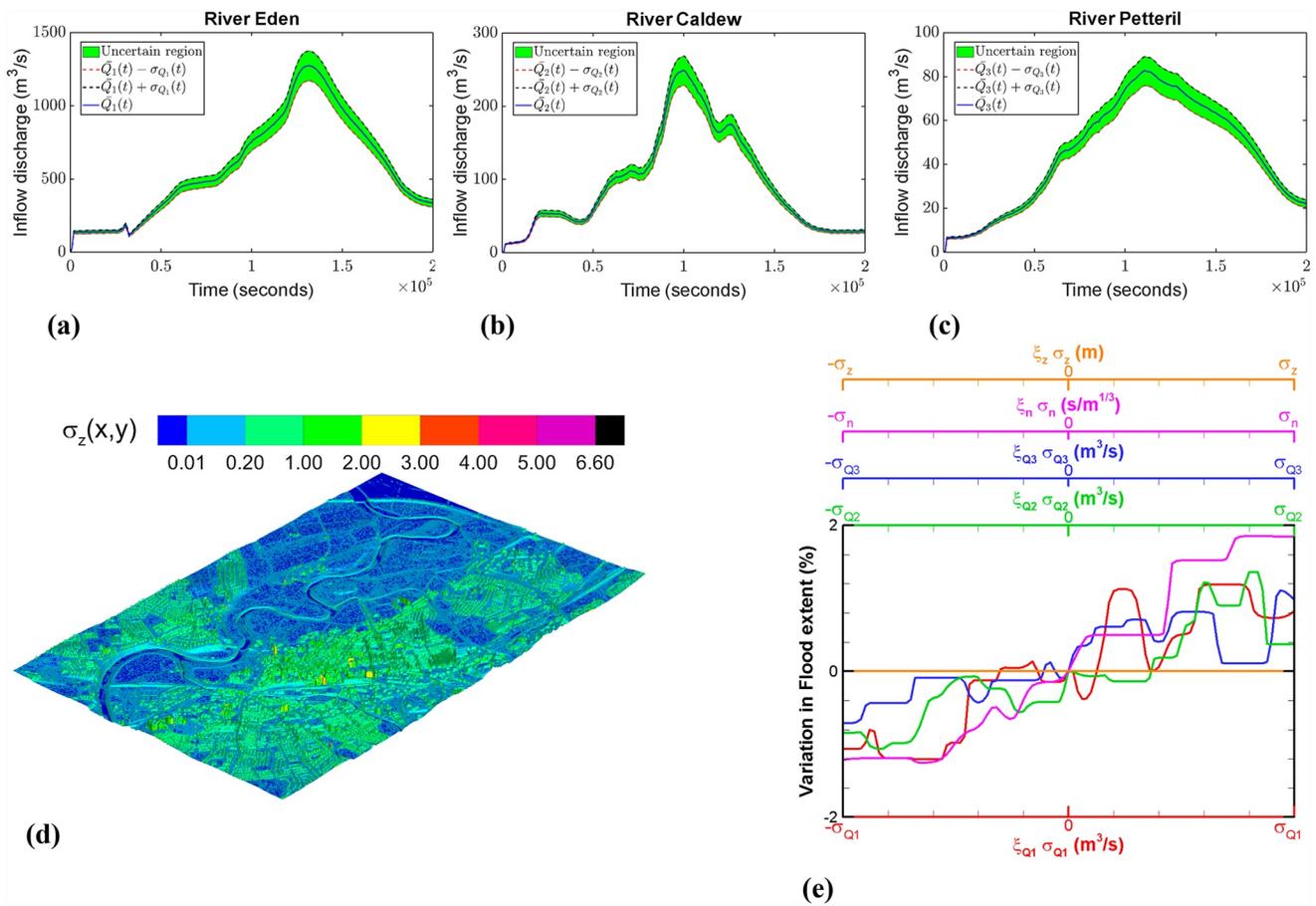
### 2.5.2. Torrential Flooding Over a Rough River Valley

This is another rapidly propagating flood test case (Neelz & Pender, 2013), often used to assess physical solvers but for a more realistic mean ground elevation variable  $\bar{z}(x, y)$  and a rougher mean Manning coefficient of  $\bar{n} = 0.04$  (Ayog et al., 2021; Shaw et al., 2021). From a breach located southwest of a down-sloping 17.0 km  $\times$  0.8 km area, floodwater flows from a mean inflow discharge variable  $\bar{Q}(t) = 3,000 \text{ m}^3 \text{ s}^{-1}$  and its variation  $\sigma_z(x, y)$  is shown in Figure 4a. The mean ground elevation variable,  $\bar{z}(x, y)$ , is generated from the DEM data at 10 m resolution (finest available), with a variation  $z(x, y, \xi_z)$  based on estimates for  $\sigma_z(x, y)$  with  $c = 0.125$ . The flood timing is 4.35 hr, when a simulation stops to analyze Flood Extent,  $HR_{ave}$ , and  $HR_{max}$ . A simulation with 1.7 million cells costs 100 s to run and the reference prediction is produced via a probabilistic SMC run with  $N_s = 40,000$ .

Again,  $D$ -dimensional response surfaces for Flood Extent,  $HR_{ave}$ , and  $HR_{max}$  are produced from the reference prediction, to analyze how the variations in the input random variables (from their mean values) propagate into each of these quantities of interest. Figure 5 includes the plots of the centerline projections of the response surfaces of each random variable for the sub-cases with  $D = 2$  (solid lines) and with  $D = 3$  (dotted lines). With  $D = 2$ , larger variations are observed for all the quantities of interest due to the inflow discharge variable



**Figure 5.** Torrential flooding over a rough river valley (Section 2.5.2). Centerline plots of the projections from the  $D$ -dimensional response surfaces for Flood Extent,  $HR_{ave}$ , and  $HR_{max}$ , showing their variations in each uncertainty dimension for the case with  $D = 2$  (solid lines) and the case with  $D = 3$  (dotted lines).



**Figure 6.** Carlisle 2005 flooding (Section 2.5.3). (a–c) show the variation in inflow discharge originating from rivers Eden, Caldew, and Petteril, respectively; (d) shows the variation from the mean ground elevation variable; and subfigure (e) shows the centerline plots of the projections from the 5-dimensional response surface into each uncertainty dimension.

compared to the Manning coefficient. For Flood Extent, the variation is nonlinear due to the inflow discharge variable but is quite flat for the Manning coefficient variable. Comparatively, the variations for  $HR_{ave}$  are larger and quite linear; but the variation due to the inflow discharge variable is steeper compared to the variation due to the Manning coefficient variable. Compared to the variations for Flood Extent and  $HR_{ave}$ , the variations for  $HR_{max}$  due to both random variables have a lower range of variability, remaining close to each other, and show smooth linear patterns.

With  $D = 3$ , the variations for any quantity of interest due to the ground elevation variable remain somewhat flat, thus relatively insignificant compared to the variations due to the other two variables: no difference in the variations for  $HR_{ave}$  is detected (compare with  $D = 2$ ) but the variations for Flood Extent and  $HR_{max}$  become non-smooth, showing noisy patterns (not seen with  $D = 2$ ). Overall, the analysis of Figure 5 indicates that Flood Extent is the most uncertain quantity of interest in this test, together with  $HR_{ave}$  for the sub-case with  $D = 2$  or alternatively with  $HR_{max}$  for the sub-case with  $D = 3$ . In Section 3.2, the UQ method  $K$  ( $K = SMC, QMC, HWE, LHS, \text{ and } ASS$ ) is validated for the identified  $N_s^K$  for reproducing different shapes of relative histograms for the quantities of interest.

### 2.5.3. Carlisle 2005 Flooding

This test case (Fewtrell et al., 2011; Horritt et al., 2010; Kabir et al., 2020; Kesserwani & Sharifian, 2023), is also adjusted to become probabilistically impacted by five input random variables. The fluvial flooding is driven by three given mean inflow discharge variables,  $\bar{Q}_p(t)_{p=1,2,3}$ , originating from rivers Eden, Caldew and Petteril, with the variations of their random variables,  $Q_p(t, \xi_{Q_p})_{p=1,2,3}$ , shown in Figures 6a–6c. The inflow drivers lead to a flood propagation over a 14.5 km<sup>2</sup> area in the city of Carlisle, where the mean Manning coefficient is often

considered to be a constant of  $\bar{n} = 0.055$ , based on which the random input variable  $n(\xi_n)$  is introduced. The mean ground elevation variable,  $\bar{z}(x, y)$ , is based on a 5 m resolution DEM (finest available), and its random input variable  $z(x, y, \xi_z)$  was based on estimates for  $\sigma_z(x, y)$  using  $c = 0.04$ , taking the values shown in Figure 6d. A single simulation used 581,061 cells and was stopped at the flood time of 40 hr, costing a runtime of 400 s. This results in a large, elapsed, runtime to make a probabilistic SMC run, limiting the reference prediction to  $N_s = 10,000$ .

Informed by the analysis of the response surfaces projections in Section 2.5.2, Flood Extent is the most uncertain quantity of interest to investigate, as is usually the case with such a slowly propagating flood over rough and realistic terrains (Aitken et al., 2022; Beevers et al., 2020; Bermúdez et al., 2017). Therefore, only the 5-dimensional response surface for Flood Extent is produced from the reference prediction to analyze its variation due to each of the five random variables—again via the plots of centerline projections included in Figure 6e. As opposed to the previous test cases (Sections 2.5.1 and 2.5.2), the strongest, largest and widest range of variation in the response surface arises due to the Manning coefficient variable. The range of variation due to each of three inflow discharge variables is comparatively weaker, but their cumulative variations is expected to lead to an overall stronger variation than that due to the Manning coefficient variable. For any of these four random variables, the type of variation for Flood Extent is quite similar, exhibiting semi-flat states in some regions or non-smooth patterns in other regions, which are connected by steep jumps. Comparatively, a flat variation is observed due to the ground elevation variable, suggesting that this variation would insignificantly influence the predictions compared to any other variation. This finding is in line with that in Section 2.5.2, seeming to suggest giving less importance to the variation in the ground elevation variable for probabilistic modeling over real and highly rough topologies. In Section 3.3, only ASS and QMC are compared using  $N_s^K = 2000$  and 4,000 (i.e., justified in Sections 3.1 and 3.2), with validation to reproducing the relative histogram for the Flood Extent quantity.

### 3. Results and Discussion

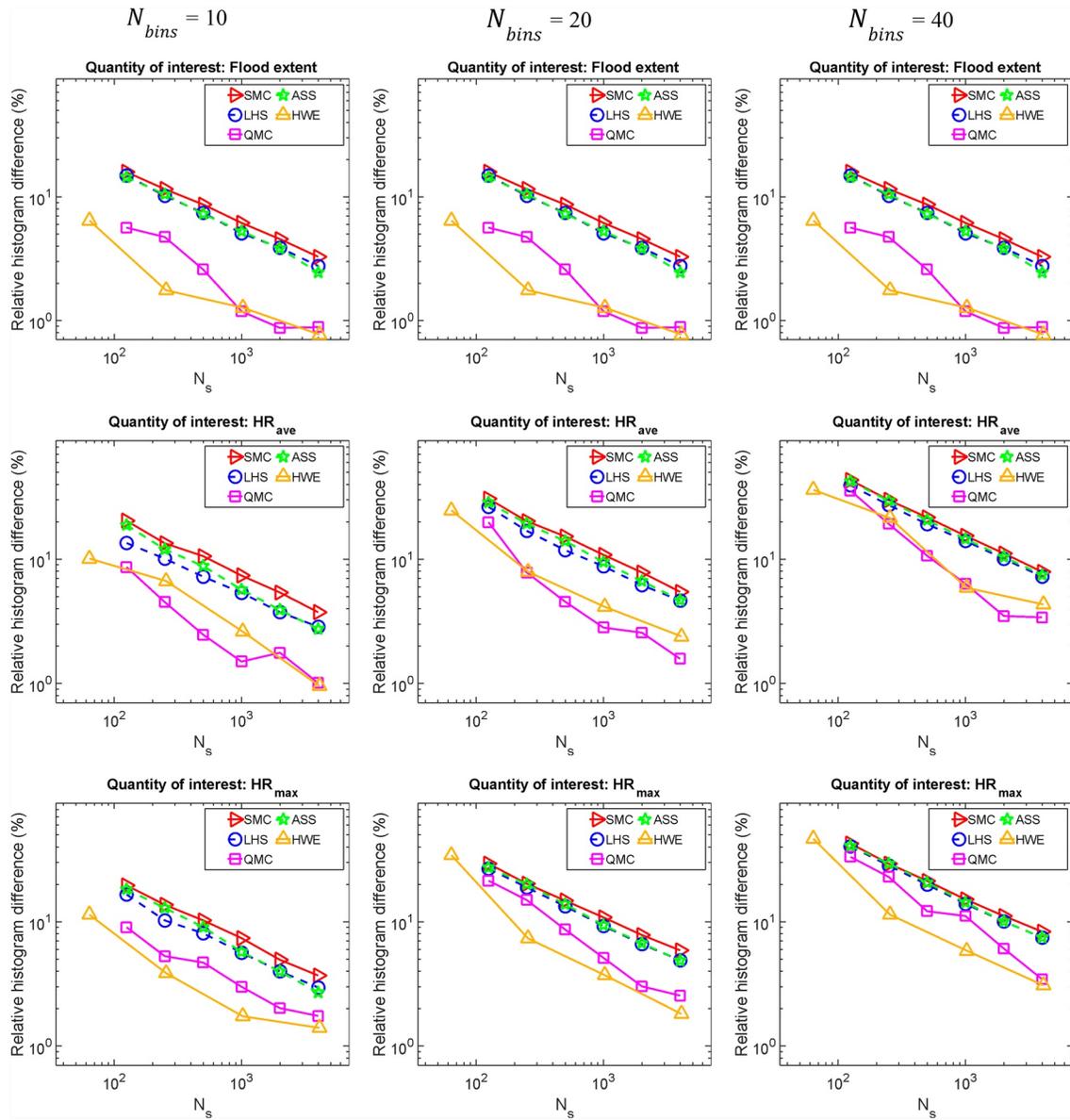
#### 3.1. Rapidly Propagating Flood Over a Smooth Terrain (Section 2.5.1)

##### 3.1.1. Sub-Case With $D = 2$

Figure 7 displays the order of convergence per UQ method based on the relative histogram difference (Equation 4) calculated for each of the quantities of interest (shown in the upper, medium and lower panels for Flood Extent,  $HR_{ave}$ , and  $HR_{max}$ , respectively), for  $N_{bins} = \{10, 20, 40\}$  (from left to right). Each UQ method preserves its order of convergence, that is, the decay rate for the relative histogram difference with increased  $N_s$ . The deterministic realization methods (QMC and HWE) show generally faster decay rates than the random sampling methods (LHS, ASS and SMC). QMC and HWE yield lower order of error magnitudes that tend to become close to the orders of error magnitudes obtained for LHS, ASS, and SMC with increased complexity in the responses for the quantity of interest (Figure 3) and with increased  $N_{bins}$ —thus becoming the closest to each other for  $HR_{max}$  and using  $N_{bins} = 40$  (Figure 7, lower-right panel).

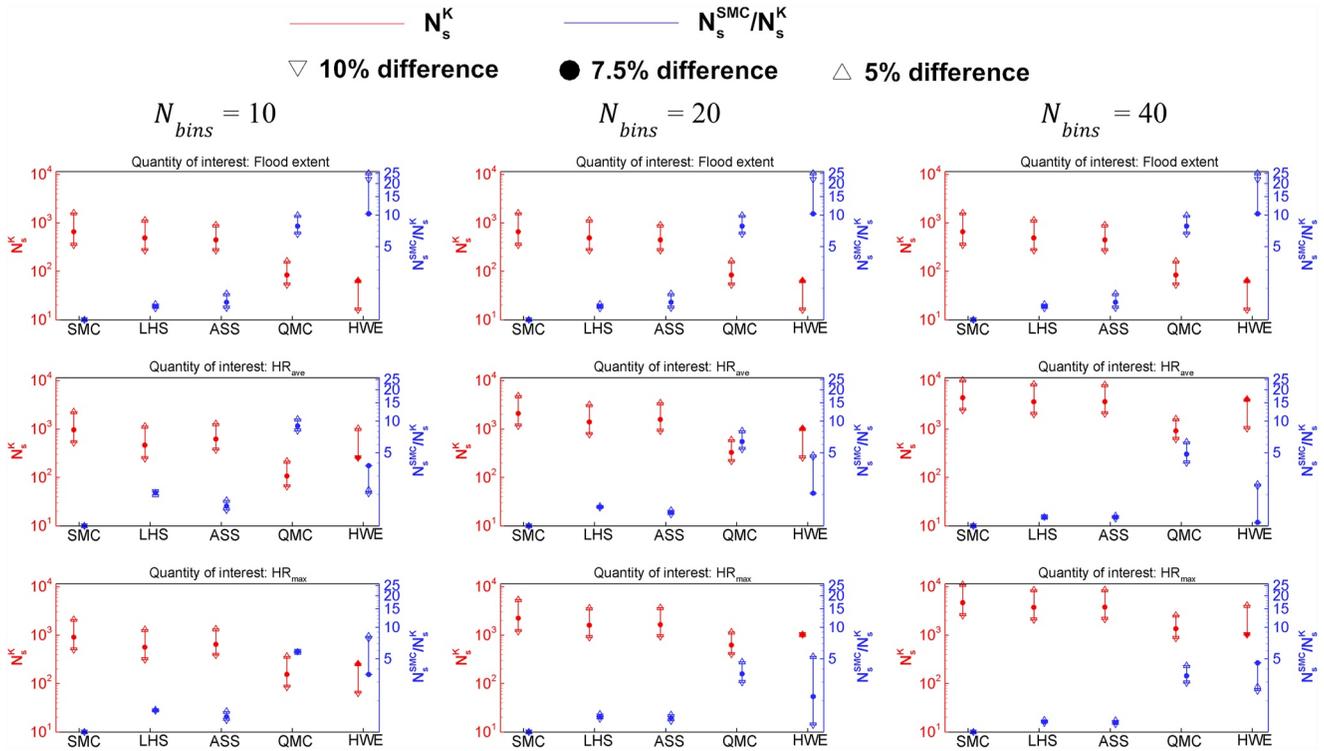
For Flood Extent, the relative histogram difference is insensitive to  $N_{bins}$  and remains one order of error magnitude smaller compared to the relative histogram difference for  $HR_{ave}$  and  $HR_{max}$ . This can be expected as the responses for Flood Extent consist of a few flat states (Figure 3), that are smaller in number than the smallest  $N_{bins} = 10$ . In contrast, the responses for  $HR_{ave}$  are non-flat states arising from larger continuous variations, further exhibiting waviness for  $HR_{max}$  (Figure 3). Therefore, the representation of the  $HR_{ave}$  and  $HR_{max}$  states are expected to become more detailed with increasingly finer bins, leading to increasingly larger order of error magnitudes (Figure 7, medium and lower panels: left-to-right). Still, the order of error magnitudes reached by the deterministic realization methods (QMC and HWE) remain lower than those reached by the random sampling methods (LHS and ASS). This indicates that QMC and HWE offer more reduction in  $N_s$  compared to LHS and ASS, but this reduction tends to decrease with increased non-flat variations in the responses of the quantity of interest and with the refinement of  $N_{bins}$ .

Consequently, for the target threshold of relative histogram difference, the  $N_s^K$  identified would be larger for  $HR_{ave}$  and  $HR_{max}$  as compared to the  $N_s^K$  identified for Flood Extent, and as  $N_{bins}$  is increased. Moreover, the  $N_s^K$  identified for the random sampling methods (LHS and ASS) would be larger compared to the  $N_s^K$  identified for the deterministic realization methods (QMC and HWE). Figure 8 shows the identified  $N_s^K$  and relative-to-SMC speedup ratios  $N_s^{SMC}/N_s^K$  to keep the target threshold for  $N_{bins} = \{10, 20, 40\}$ .



**Figure 7.** Rapidly propagating flood over a smooth terrain (Section 2.5.1). Order of convergence per uncertainty quantification method  $K$  ( $K$  = standard Monte Carlo, QMC, Haar-wavelet expansion, Latin hypercube sampling and adaptive stratified sampling) for the relative histogram difference (Equation 4) calculated for: Flood Extent (upper panel),  $HR_{ave}$  (middle panel), and  $HR_{max}$  (lower panel), using  $N_{bins} = \{10, 20, 40\}$  (left to right).

All the UQ methods achieve speedup ratios greater than one, thus lead to identifying  $N_s^K$  that offer a reduction over  $N_s^{SMC}$ . However, the reduction differs depending on the quantity of interest. Flood Extent, since it has few flat variations in its responses, leads to the smallest  $N_s^K$  alongside the highest speedup ratios and irrespective of  $N_{bins}$ . The identified  $N_s^{LHS}$  and  $N_s^{ASS}$  are at least one order of magnitude larger than  $N_s^{QMC}$  and  $N_s^{HWE}$ , yielding speedup ratios that range between 1.1 and 1.8 for LHS and ASS and between 5 and 25 for QMC and HWE. For  $HR_{ave}$  and  $HR_{max}$ , the identified  $N_s^K$  are expectedly larger, up to one order of magnitude with the largest  $N_{bins} = 40$ . This leads to a drop in the lower bound of the speedup ratios for QMC and HWE from around 5 to around 1.1. However, this drop occurs due to the overly large  $N_s^{HWE} = 4096$  identified for  $HR_{ave}$ . With the lower  $N_s^{HWE} = 1,024$ , the relative histogram difference for Flood Extent and  $HR_{max}$  are 1.3% and 6.4%, respectively; however, the difference for  $HR_{ave}$  becomes 8.1%, which is not below the average target difference of 7.5%. The only next possible choice for an  $N_s^{HWE}$  that is larger than 1,024, is  $N_s^{HWE} = 4096$ , given the inflexibility of HWE in the selection of  $N_s^{HWE}$ , which is four-times larger leading to a difference of 3.8% for  $HR_{ave}$  that is below the target average

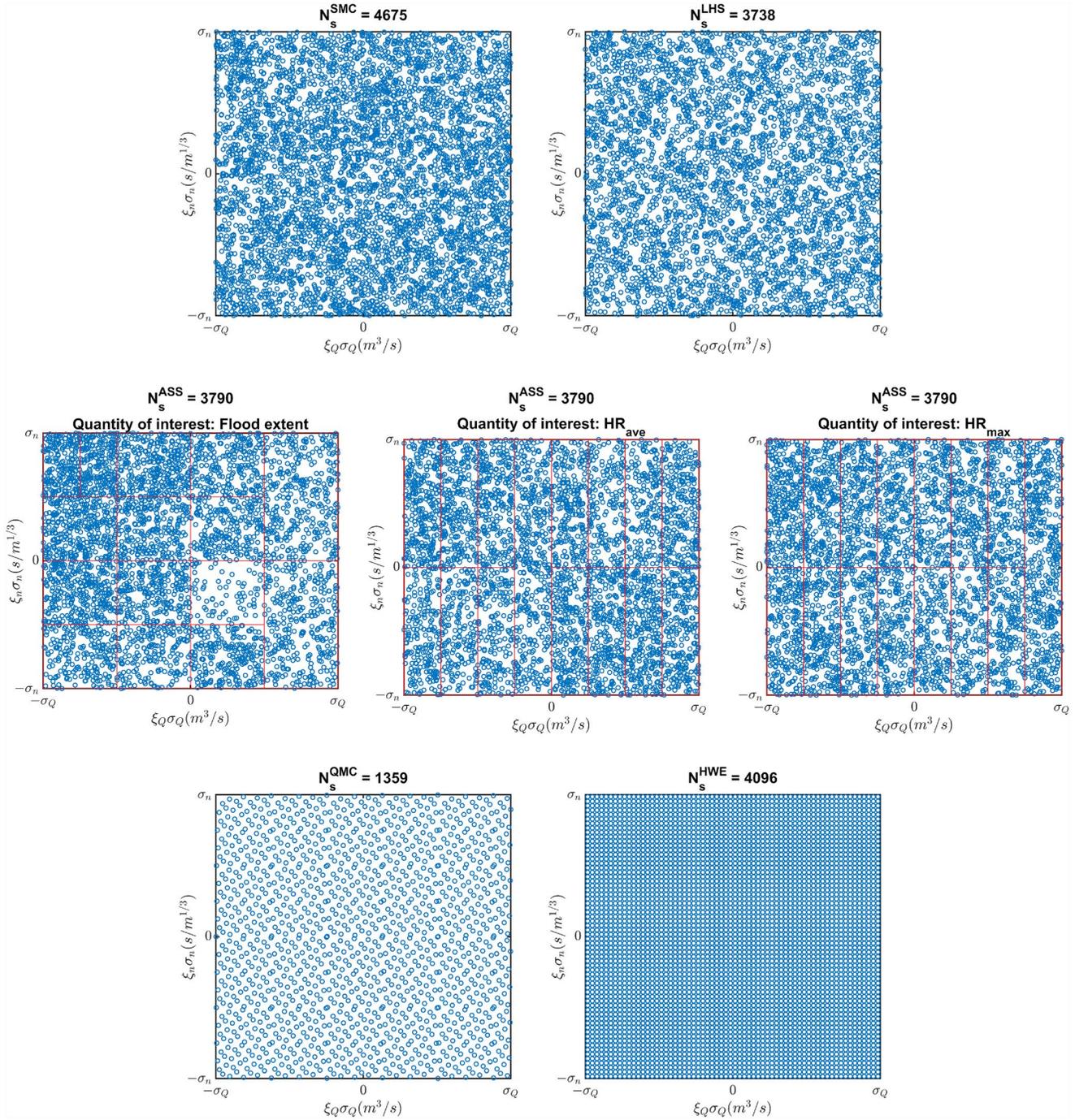


**Figure 8.** Rapidly propagating flood over a smooth terrain (Section 2.5.1). The  $N_s^K$  and relative-to-SMC speedup ratio (Equation 5) to meet the target difference of  $7.5\% \pm 2.5\%$  for Flood Extent (upper panel),  $HR_{ave}$  (middle panel), and  $HR_{max}$  (lower panel), using  $N_{bins} = \{10, 20, 40\}$  (left to right).

difference of 7.5%. Therefore, QMC is a better alternative since its lower bound remains around four-times higher than 1.1. The speedup ratios for LHS and ASS remain similar to the ratios identified for Flood Extent, subject to a slightly lower upper bound, of 1.30.

Figure 9 illustrates the sampling patterns in the uncertainty space with  $D = 2$  for each UQ method for the largest  $N_s^K$  identified to meet the average threshold difference of 7.5% for all the quantities of interest with  $N_{bins} = 40$ . The sampling with SMC shows disadvantages caused by the presence of clumps, resulting in the largest  $N_s$  ( $N_s^{SMC} = 4,675$ ). A more efficient sampling is achieved by LHS, with 937 fewer samples ( $N_s^{LHS} = 3,738$ ). ASS performs similar to LHS, needing slightly more samples ( $N_s^{ASS} = 3,790$ ), though its sampling is different per quantity of interest. More strata and samples emerge for Flood Extent in the regions showing strong jumps in its responses (Figure 3), but no strata refinement is seen for  $HR_{ave}$  and  $HR_{max}$  that have non-flat variations (Figure 3). This suggests that ASS may be a more suited alternative-to-SMC to reproduce the relative histogram for a quantity of interest with responses exhibiting close-to-flat states connected by jumps (for example, for the “Carlisle 2005 flooding” (Section 2.5.3) as can be seen in Figure 6e). HWE leads to overtly refined sampling with  $D = 2$  ( $N_s^{HWE} = 4,096$ ), caused by its aforementioned inflexibility to allow a choice for  $N_s^{HWE}$  somewhere in-between 1,024 and 4,096. QMC achieves the highest reduction ( $N_s^{QMC} = 1,359$ ) compared to any other alternative-to-SMC, which suggests that it provides the most efficient sampling of the uncertainty space with  $D = 2$ .

Figure 10 shows the plots of the relative histograms for Flood Extent,  $HR_{ave}$ , and  $HR_{max}$ , extracted per UQ method for the identified  $N_s^K$  based on  $N_{bins} = 40$  (shown in Figure 9), and in which the coarser relative histograms using  $N_{bins} = \{10, 20\}$  are based on the same  $N_s^K$ . For Flood Extent, the relative histograms exhibit a discrete distribution for  $N_{bins} = \{20, 40\}$ , which tends to look like a left-skewed unimodal distribution with the coarsest  $N_{bins} = 10$ , suggesting that larger  $N_{bins}$  is still needed despite the few discrete states in this quantity of interest. For  $HR_{ave}$ , the relative histograms follow an almost symmetric distribution that becomes more complex with larger  $N_{bins}$ . The relative histograms seen for  $HR_{max}$  are the most complex overall, in particular as  $N_{bins}$  is enlarged to 40 leading to a quite irregular distribution. Among the histogram distributions, the discrete one is the most accurately captured by all the UQ methods. The symmetric distribution seems to be more challenging to capture, though it is better reproduced with the deterministic realization methods (QMC and HWE) than with the random sampling methods

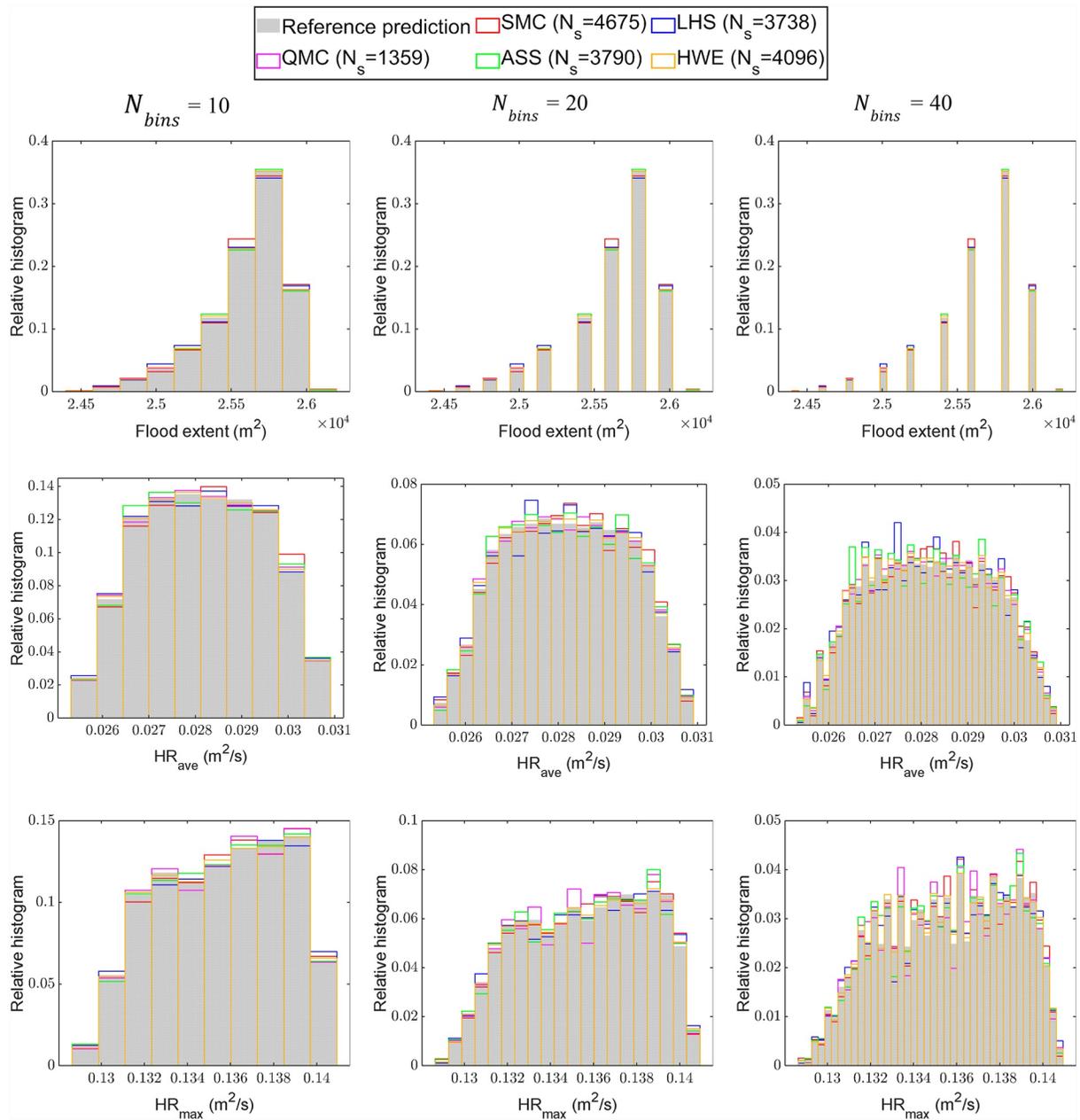


**Figure 9.** Rapidly propagating flood over a smooth terrain (Section 2.5.1). Sampling patterns for each uncertainty quantification method for the largest  $N_s^K$  identified (Figure 8) to meet the target average difference of 7.5% for all the quantities of interest.

(SMC, LHS and ASS). The distribution for  $HR_{max}$ , involving multimodalities, is the most challenging to capture for which all the UQ methods reach the average threshold difference of 7.5%.

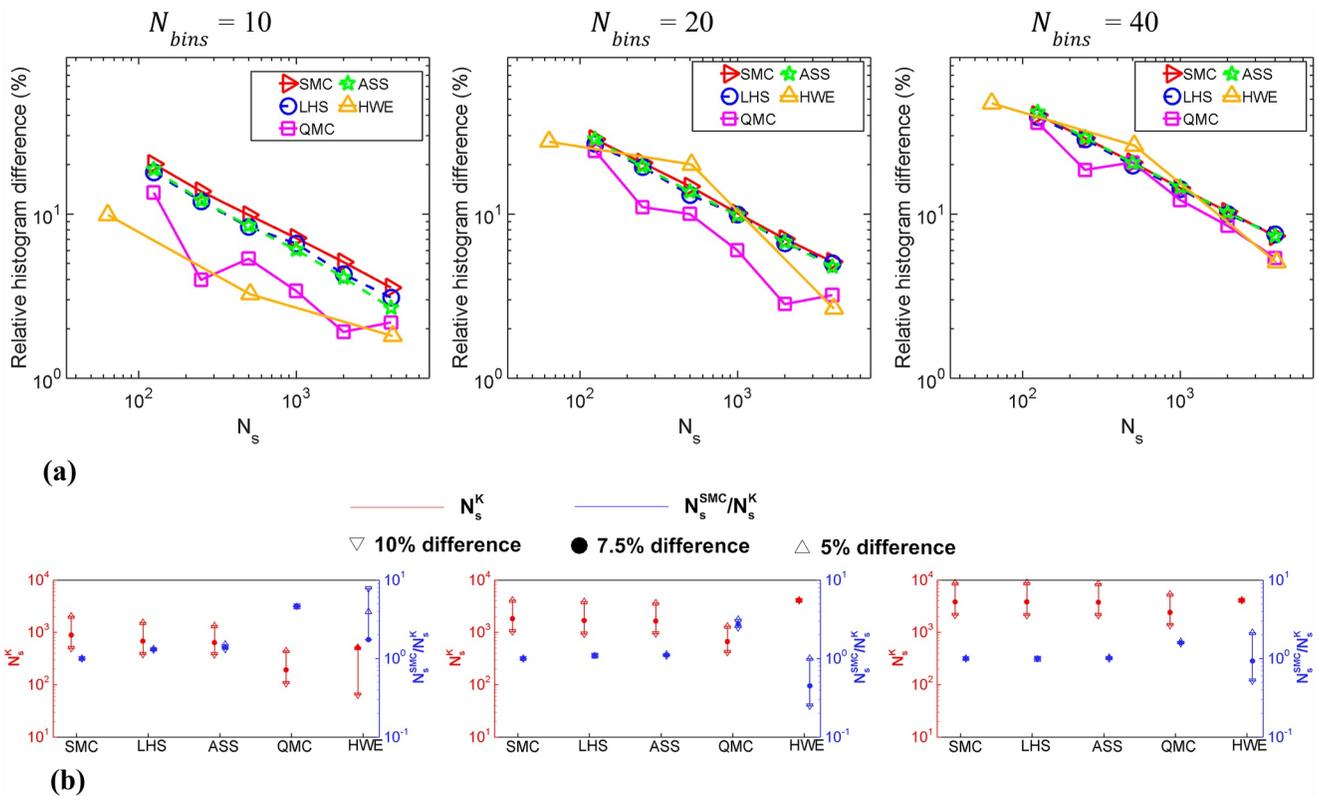
### 3.1.2. Sub-Case With $D = 3$

Here, the analysis is restricted to  $HR_{max}$  as it is the only quantity of interest that changes drastically as  $D$  is increased to 3—compared to Flood Extent and  $HR_{ave}$  that retain similar responses to as with  $D = 2$  (Figure 3). Figure 11a displays the order of convergence per UQ method based on the relative histogram difference



**Figure 10.** Rapidly propagating flood over a smooth terrain (Section 2.5.1). Relative histogram plot per uncertainty quantification method for the  $N_s^K$  shown in Figure 9 based on  $N_{bins} = 40$ . The plots for the coarser relative histograms using  $N_{bins} = \{10, 20\}$  are based on the same  $N_s^K$ , thus are downsampled from the relative histograms using  $N_{bins} = 40$ .

(Equation 4) calculated for  $HR_{max}$  (left-to-right); while Figure 11b shows the  $N_s^K$  identified per UQ method and their relative-to-SMC speedup ratios  $N_s^{SMC}/N_s^K$  to keep the target threshold difference with  $N_{bin} = \{10, 20, 40\}$ . With larger  $N_{bins}$ , the orders of convergence among the UQ methods become similar, reaching maximum similarity with the slightly higher error magnitudes with  $N_s < 2,000$  for  $N_{bins} = 40$  (Figure 11a). This suggests that all UQ methods tend to have a similar decay rate for  $N_s^K < 2,000$ ; and, as  $N_{bins}$  is refined, the choice among them is not as significant with  $D = 3$  compared to with  $D = 2$ . This observation also arises from the fact that  $HR_{max}$  is the most relevant quantity of interest under consideration (Figure 3), which would require larger  $N_s^K$  (with all the UQ methods  $K$ ) than any other quantity of interest, to reproduce refined relative histograms using  $N_{bins} = 40$ . As  $N_s$  exceeds 4,000, the deterministic realization methods (HWE and QMC) show lower relative histogram differences



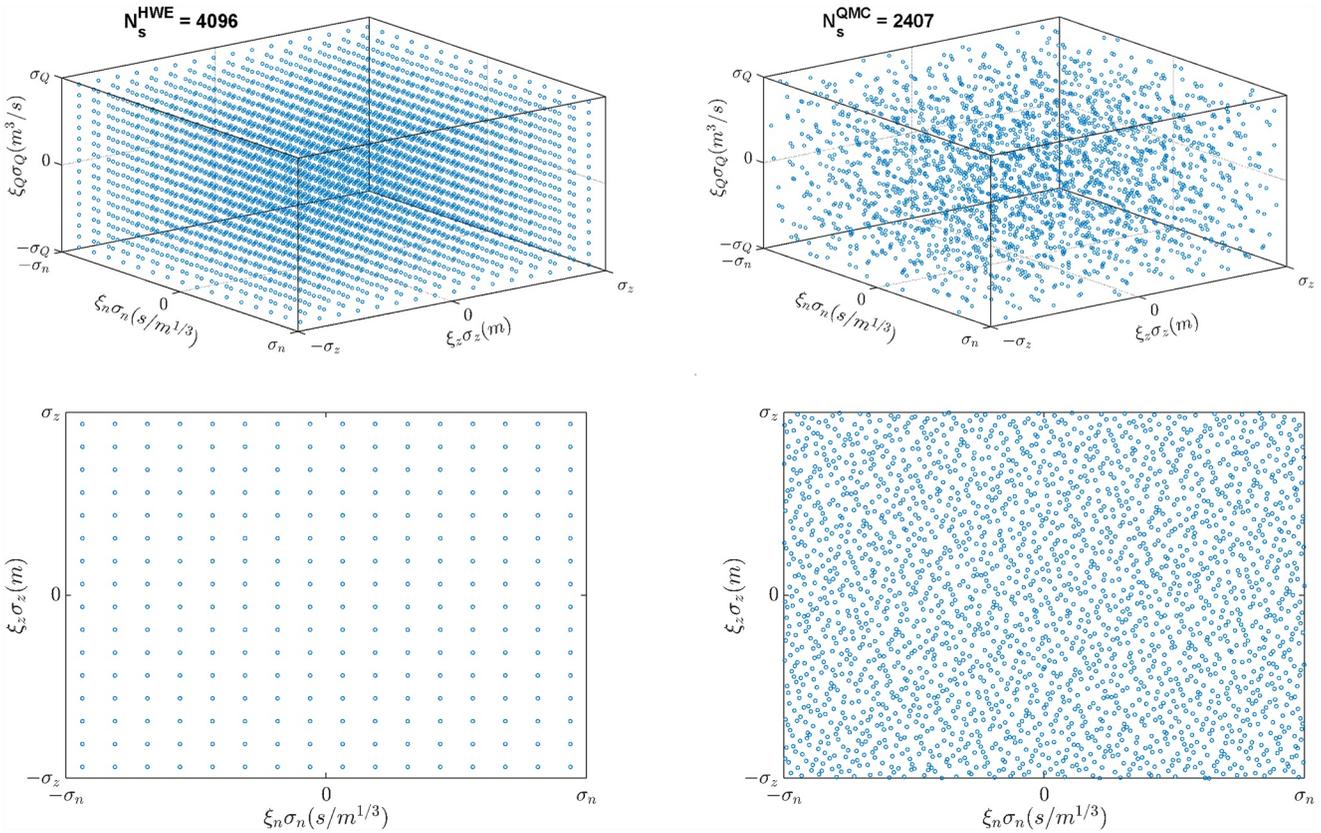
**Figure 11.** Rapidly propagating flood over a smooth terrain (Section 2.5.1). Panel (a) displays the order of convergence per uncertainty quantification (UQ) method  $K$  ( $K =$  standard Monte Carlo, QMC, Haar-wavelet expansion, Latin hypercube sampling and adaptive stratified sampling) based on the relative histogram difference (Equation 4) calculated for  $HR_{max}$ , and using  $N_{bin} = \{10, 20, 40\}$  (left to right); and panel (b) shows the  $N_s^K$  identified with each UQ method  $K$  to keep the target relative histogram difference of  $7.5\% \pm 2.5\%$  alongside the relative-to-SMC speedup ratios  $N_s^{SMC}/N_s^K$ .

than the random sampling methods (SMC, LHS and ASS), but these differences become less significant as  $N_{bins}$  is refined to 40 compared to with  $N_{bins} = \{10, 20\}$ .

Consequently, with  $D = 3$ , the identified  $N_s^K$  would be larger than with  $D = 2$  along with larger  $N_{bins}$ , and larger  $N_s^K$  should be expected with the random sampling methods (LHS and ASS) compared with the deterministic realization methods (QMC and HWE). As can be observed from Figure 11b, there is no notable reduction in the identified  $N_s^{LHS}$  and  $N_s^{ASS}$  since they lead to a speedup ratio range close to 1. The identified  $N_s^{QMC}$  and  $N_s^{HWE}$  yield a speedup ratio range of around 0.5-to-2.0. These upper and lower bounds are, however, fluctuations reached by HWE, which leads to an average speedup ratio that is lower than the average speedup ratio of 1.6, which is consistently preserved by QMC. This suggests that QMC is a better choice to meet the average threshold difference of 7.5% while having  $N_s$  somewhere between 2,000 and 4,000.

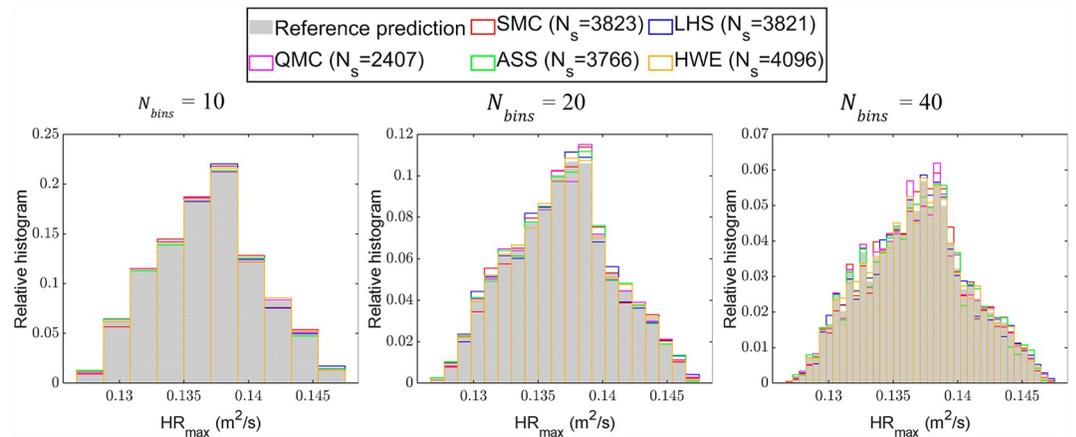
Figure 12 illustrates the sampling patterns in the three-dimensional uncertainty space for QMC and HWE, using the  $N_s^{QMC}$  and  $N_s^{HWE}$  identified to keep the average threshold difference of 7.5% for the finest  $N_{bins} = 40$  (Figure 11). The sample distributions for the random sampling methods ( $K =$  SMC, LHS and ASS) are not included, to save space, as they lead to more or less similar  $N_s^K$  with similar distribution patterns. Cross-sections of a two-dimensional projection are included to distinguish between the sampling abilities more clearly, among HWE and QMC. Contrasting the sampling patterns between the cross-sections, QMC achieves more efficient filling of samples ( $N_s^{QMC} = 2407$ ) compared to HWE that achieves a relatively sparse sampling even with about twice the number of samples ( $N_s^{HWE} = 4,096$ ). Note that, this  $N_s^{HWE}$  was identified for HWE with  $D = 2$ , where the sampling patterns were not as sparse (Figure 9), thus demonstrating how HWE becomes less practical in its sampling with increased dimensionality,  $D$ , of the uncertainty space.

Figure 13 includes the plots of the relative histograms for  $HR_{max}$ , shown for each UQ method for the identified  $N_s^K$  to meet the average threshold difference of 7.5%, based on  $N_{bins} = 40$  (Figure 11). The relative histograms

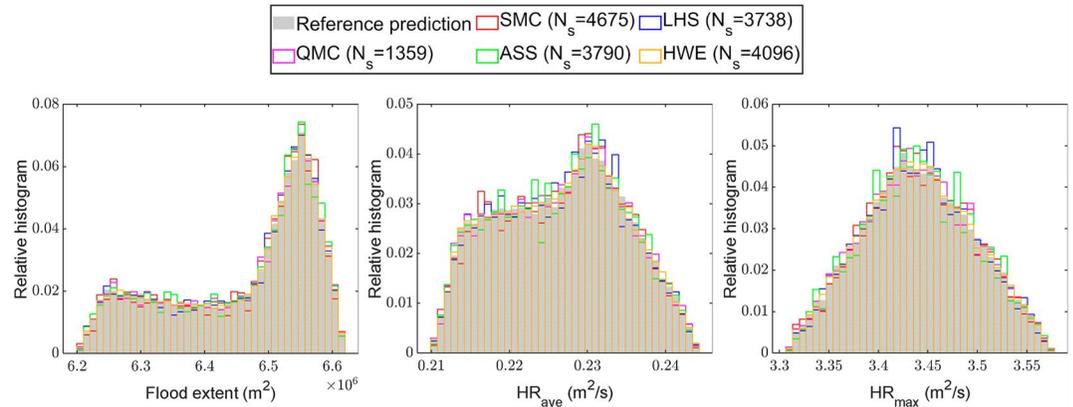


**Figure 12.** Rapidly propagating flood over a smooth terrain (Section 2.5.1). The upper panel shows the sampling inside the three-dimensional uncertainty space using the largest  $N_s^K$  identified (for QMC and Haar-wavelet expansion) to keep the target average difference of 7.5%. The lower panel displays a two-dimensional projection to distinguish between the sampling abilities more clearly.

follow a triangular distribution that displays sharper details as  $N_{bins}$  is enlarged, making it even more challenging to capture by the UQ methods  $K$  with a  $N_s^K$  between around 2,000 and 4,000 ( $K = SMC, LHS, ASS, \text{ and } QMC$ ). QMC again outperforms the other methods with a relatively smaller  $N_s^{QMC} = 2,400$ , that is closer to the lower bound of 2,000, compared to any other UQ method  $K$  that yield a  $N_s^K$  that is closer to the upper bound of 4,000.



**Figure 13.** Rapidly propagating flood over a smooth terrain (Section 2.5.1). Relative histogram plot per uncertainty quantification method for the  $N_s^K$  identified to keep the target average difference of 7.5% based on  $N_{bins} = 40$ . The plots for the coarser relative histograms using  $N_{bins} = \{10, 20\}$  are based on the same  $N_s^K$ , thus are downsampled from the relative histograms using  $N_{bins} = 40$ .



**Figure 14.** Torrential flooding over a rough river valley (Section 2.5.2). Relative histograms with  $N_{bins} = 40$  for Flood Extent (left),  $HR_{ave}$  (middle) and  $HR_{max}$  (right) per uncertainty quantification method  $K$  ( $K =$  standard Monte Carlo, QMC, Haar-wavelet expansion, Latin hypercube sampling, and adaptive stratified sampling) for the  $N_s^K$  identified for the sub-case with  $D = 2$  (Section 3.1.1).

These analyses have quantified the range of the speedup ratio for the UQ methods targeting a relative histogram difference below the (average) threshold to 7.5%. Remarkably, targeting a difference based on the relative histogram metric yields a range with 4 and 175 times smaller lower and upper bounds, respectively, as compared to targeting the difference based on the standard errors of the mean and the variance (Appendix A). Consequently, the metric of the relative histogram difference provides more informative estimates of the relative-to-SMC speedup ratios for the UQ methods to capture the frequency of occurrence for the flood-related quantities of interest.

### 3.2. Torrential Flooding Over a Rough River Valley (Section 2.5.2)

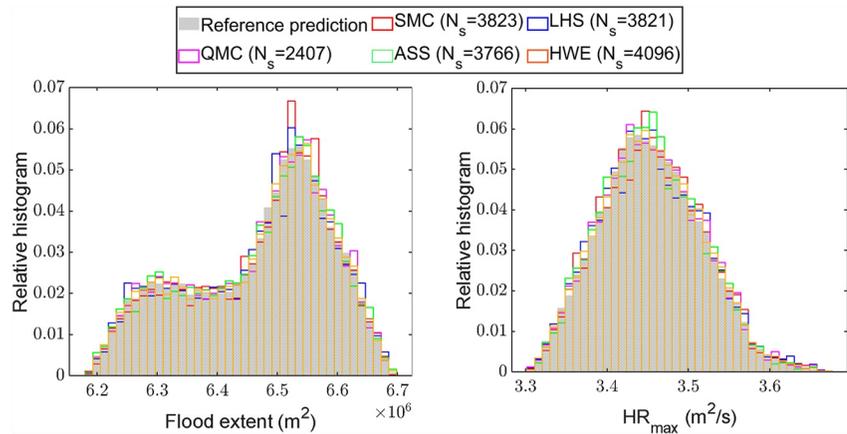
#### 3.2.1. Sub-Case With $D = 2$

Figure 14 includes the plots of the relative histograms for Flood Extent,  $HR_{ave}$ , and  $HR_{max}$  per UQ method  $K$  for the identified  $N_s^K$  with  $N_{bins} = 40$  (Section 3.1.1); note that the coarser relative histograms with  $N_{bins} = \{10, 20\}$  are not shown hereafter, but their relative differences are quantified (Table 1). As may be expected from the behavior of the response surfaces (Figure 5), the relative histograms for Flood Extent and  $HR_{ave}$  follow almost bimodal distributions, but the distribution for  $HR_{ave}$  is weaker and more akin to have an inflexion; whereas the

**Table 1**  
Torrential Flooding Over a Rough River Valley (Section 2.5.2)

$N_{bins}$	Quantity of interest	Relative histogram difference (%)				
		$N_s^{SMC} = 4675$	$N_s^{LHS} = 3738$	$N_s^{ASS} = 3790$	$N_s^{QMC} = 1359$	$N_s^{HWE} = 4096$
10	Flood Extent	3.7	3.5	2.4	1.6	0.8
	$HR_{ave}$	3.6	3.5	3.1	1.6	1.4
	$HR_{max}$	3.8	3.5	3.8	1.5	1.2
20	Flood Extent	5.1	3.7	5.0	3.6	1.7
	$HR_{ave}$	5.2	5.1	4.6	2.7	2.1
	$HR_{max}$	4.8	4.6	6.0	4.3	2.0
40	Flood Extent	8.8	5.3	8.6	5.7	2.6
	$HR_{ave}$	7.1	7.4	7.8	5.3	3.1
	$HR_{max}$	6.9	7.7	9.3	6.2	3.3

*Note.* Relative histogram differences (via Equation 4) against the test-specific reference prediction (Section 2.5.2) for  $N_{bins} = \{10, 20, 40\}$  for the UQ methods  $K$  ( $K =$  SMC, QMC, HWE, LHS, and ASS) considering the  $N_s^K$  identified for the sub-case with  $D = 2$  (Section 3.1.1).



**Figure 15.** Torrential flooding over a rough river valley (Section 2.5.2). Relative histograms with  $N_{\text{bins}} = 40$  for Flood Extent (left) and  $HR_{\text{max}}$  (right) per uncertainty quantification method  $K$  ( $K =$  Standard Monte Carlo, QMC, Haar-wavelet expansion, Latin hypercube sampling, and adaptive stratified sampling) for the  $N_s^K$  identified for the sub-case with  $D = 3$  (Section 3.1.2).

relative histogram for  $HR_{\text{max}}$  follows a triangular distribution. These distributions are captured by the UQ methods with a maximum difference below 10%. This indicates that the  $N_s^K$  quantified previously (Section 3.1.1) are still valid for the more realistic setting of the second test case (Section 2.5.2).

However, the capture of the relative histogram become harder for the UQ methods with larger  $N_{\text{bins}}$ . Table 1 includes the relative histogram differences against the test-specific reference prediction (Section 2.5.2) for all the UQ methods and the quantities of interests for  $N_{\text{bins}} = \{10, 20, 40\}$ . As expected, the larger the  $N_{\text{bins}}$ , the larger the difference, which averages for all the UQ methods and the quantities of interests, to 6.3% for  $N_{\text{bins}} = 40$ , to 4% for  $N_{\text{bins}} = 20$  and to 2.6% for  $N_{\text{bins}} = 10$ . Moreover, this difference for the random sampling methods (SMC, LHS and ASS) tends to have a faster growth rate compared to the differences for the deterministic realization methods (QMC and HWE). The largest differences occur with  $N_{\text{bins}} = 40$ , varying between 5.3% and 9.3% for SMC, LHS and ASS and between 2.6% and 6.2% for QMC and HWE, considering all the quantities of interest. LHS and ASS lead to a difference up to 7.7% and 9.3%, respectively, which is very close to that of SMC, being around 7.0% but using about 900 fewer samples. HWE attains a difference of 3.3% that is lower than the 6.2% difference attained by QMC with three times fewer samples. This suggests that QMC is the most efficient alternative-to-SMC to reduce  $N_s$  and keep the threshold difference of 7.5%.

### 3.2.2. Sub-Case With $D = 3$

Figure 15 shows the plots of the relative histograms for Flood Extent and  $HR_{\text{max}}$  per UQ method for the identified  $N_s^K$  with  $N_{\text{bins}} = 40$  (Section 3.1.2); the relative histogram for  $HR_{\text{ave}}$  is not included as it is very similar to the one shown in Figure 14, also informed by the behavior of the response surfaces (Figure 5). Now, the relative histogram predicted for Flood Extent exhibits a weaker bimodality within its distribution and that predicted for  $HR_{\text{max}}$  becomes a unimodal distribution with an upper tail. The predicted reference distributions are well-captured by all the UQ methods leading to a maximum difference below 10%, thus indicating that the  $N_s^K$  quantified previously with  $D = 3$  (Section 3.1.2) remain valid.

Table 2 includes the relative histogram differences against the test-specific reference prediction (Section 2.5.2) for all the UQ methods for  $N_{\text{bins}} = \{10, 20, 40\}$ , considering the  $N_s^K$  identified for the sub-case with  $D = 3$  (Section 3.1.2). Again, the larger the  $N_{\text{bins}}$  the bigger the relative histogram differences for all the UQ methods, suggesting that the capturing of the relative histogram is more demanding with larger  $N_{\text{bins}} = 40$ . In this case, the difference for Flood Extent varies between 4.6% and 9.3% and that for  $HR_{\text{max}}$  between 4.0% and 8.8%, suggesting that the bimodal Flood Extent distribution is slightly more difficult to capture compared to the almost unimodal  $HR_{\text{max}}$  distribution. In doing so, the deterministic realization methods (QMC and HWE) outperform, leading to a difference between 4.6% and 7.0%, compared to the difference between 7.5% and 9.3% achieved by the random sampling methods. For Flood Extent, ASS achieves a difference of 7.7% that is lower than the 9.3% difference obtained by LHS, suggesting that ASS may become a better option as  $D$  is increased from 2 to 3 and as the quantity of interest exhibits non-smooth variations in the response surfaces (Figure 5). HWE achieves a difference of

**Table 2**  
Torrential Flooding Over a Rough River Valley (Section 2.5.2)

$N_{bins}$	Quantity of interest	Relative histogram difference (%)				
		$N_S^{SMC} = 3,823$	$N_S^{LHS} = 3,821$	$N_S^{ASS} = 3,766$	$N_S^{QMC} = 2,407$	$N_S^{HWE} = 4,096$
10	Flood Extent	4.3	2.6	2.7	2.3	1.7
	$HR_{max}$	2.7	4.4	3.4	1.5	1.5
20	Flood Extent	6.6	5.5	4.7	4.0	2.4
	$HR_{max}$	4.6	5.7	6.5	2.9	1.9
40	Flood Extent	7.5	9.3	7.7	7.0	4.6
	$HR_{max}$	7.5	8.2	8.8	5.4	4.0

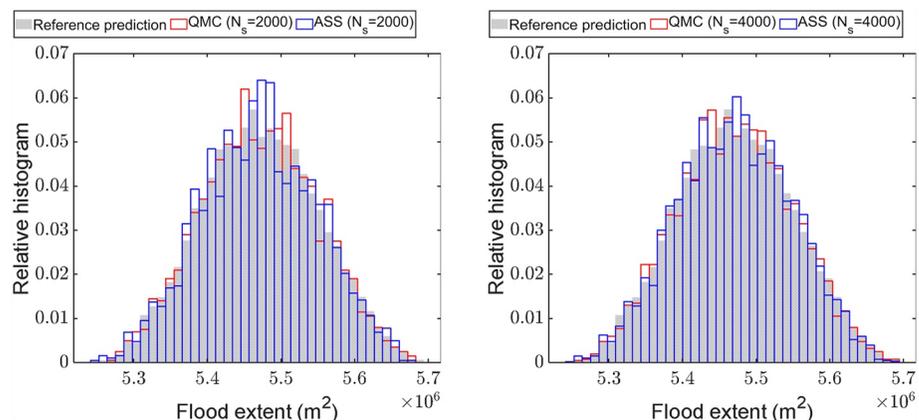
Note. Relative histogram differences (via Equation 4) against the test-specific reference prediction (Section 2.5.2) for  $N_{bins} = \{10, 20, 40\}$  for the UQ methods  $K$  ( $K = SMC, QMC, HWE, LHS, \text{ and } ASS$ ) considering the  $N_s^K$  identified for the sub-case with  $D = 3$  (Section 3.1.2).

4.6% while QMC achieves a difference of 7.0% by employing 1,689 less samples. However, as  $D$  is increased from 2 to 3, the difference for HWE increased by a higher rate, of 1.3%, compared to that with QMC, of 0.8%, which introduced a further reduction of 789 samples compared to HWE. For  $HR_{max}$ , the differences are slightly lower than for Flood Extent, except for ASS that is designed to be better suited to capture responses with local, non-flat variations (Section 3.1).

Overall, the alternative-to-SMC UQ methods with the identified  $N_s^K$  (Section 3.1) leads to predictions that are close to the average threshold difference of 7.5% without exceeding the maximum threshold of 10%. Among these methods, QMC and ASS keep the relative histogram prediction for Flood Extent close to 7.5% and tend to perform better as  $D$  is increased, with  $N_s^{ASS}$  and  $N_s^{QMC}$  remaining within 2,000 and 4,000. Therefore, ASS and SMC are investigated further for a real case study involving five input random variables ( $D = 5$ ) to analyze their Flood Extent histograms (Section 3.3), predicted using  $N_s^K = \{2,000, 4,000\}$ .

### 3.3. Carlisle 2005 Flooding (Section 2.5.3)

In Figure 16, the relative Flood Extent histograms predicted by QMC and ASS are compared to the reference prediction for  $N_s^K = 2,000$  (Figure 16, left) and 4,000 (Figure 16, right). The histograms follow a symmetric distribution that seems to be better captured by QMC compared to ASS. Table 3 includes the relative histogram differences against the test-specific reference prediction (Section 2.5.3) for QMC and ASS considering  $N_{bins} = \{10, 20, 40\}$  and  $N_s^K = \{2,000, 4,000\}$ . For  $N_s^K = 2000$ , the relative histogram difference increases with both QMC and ASS with larger  $N_{bins}$ , but the difference with ASS increases at faster rate than the difference



**Figure 16.** Carlisle 2005 flooding (Section 2.5.3). Relative Flood Extent histograms (using  $N_{bins} = 40$ ) per uncertainty quantification method  $K$  ( $K = QMC$  and adaptive stratified sampling) for  $N_s^K = 2,000$  (left) and 4,000 (right).

**Table 3**  
Carlisle 2005 Flooding (Section 2.5.3)

$N_{\text{bins}}$	Relative histogram difference (%)			
	$N_s = 2,000$		$N_s = 4,000$	
	ASS	QMC	ASS	QMC
10	4.1	2.6	2.4	4.4
20	8.7	5.1	6.1	6.2
40	12.5	8.6	8.5	7.7

Note. Relative histogram differences (via Equation 4) for Flood Extent against the test-specific reference prediction (Section 2.5.3) considering  $N_{\text{bins}} = \{10, 20, 40\}$  for the UQ methods  $K$  ( $K = \text{QMC}$  and  $\text{ASS}$ ) each considered with  $N_s = 2,000$  and  $4,000$ .

with QMC, leading to differences around 8.6% and 12.5% for the largest  $N_{\text{bins}} = 40$ , respectively. This suggests favoring QMC over ASS when using a  $N_s^K$  that is close to 2000. For  $N_s^K = 4,000$ , the relative histogram difference with ASS is lower than that with QMC for  $N_{\text{bins}} = 10$ , almost identical for  $N_{\text{bins}} = 20$  but slightly higher for  $N_{\text{bins}} = 40$ , suggesting that the increase in  $N_{\text{bins}}$  may be of less influence as  $N_s^K$  is close to 4,000. Overall, QMC leads to a difference that is closer to the average threshold difference of 7.5%, with both  $N_s^{\text{QMC}} = 2,000$  and  $4,000$ , and irrespective of  $N_{\text{bins}}$ ; whereas ASS only meets this criterion with  $N_s^{\text{ASS}} = 4,000$ . This means that to keep the relative histogram difference below the maximum threshold of 10%, using  $N_s^{\text{QMC}} = 2,000$  is feasible to achieve a relative-to-SMC speedup ratio of 5 with QMC, for this test case, whereas only a speedup ratio of 2.5 can be achieved with ASS by using  $N_s^{\text{ASS}} = 4,000$ .

Among the two selected UQ candidates, QMC seems able to maximize both the predictive accuracy in the capture of relative histograms and the speedup ratio to gain efficiency over SMC. Given its simplicity compared to ASS and its outperformance on efficiency in the other two test cases involving rapid floods over smooth and rough terrains (Sections 3.1 and 3.2), it can be concluded that QMC is a compelling alternative-to-SMC UQ method for complex, probabilistic modeling of real-world floods with multiple uncertain inputs.

#### 4. Limitations and General Applicability

The performance analysis of the alternative-to-SMC UQ methods for probabilistic flood modeling is limited to: using the GPU-FV1 physical solver of LISFLOOD-FP, making simulation runs for moderately sized spatial domains ( $\leq 15 \text{ km}^2$ ), having an input uncertainty space with a dimension  $2 \leq D \leq 5$  in which the variations in the input random variables are all represented by uniform probability distribution, and splitting the load of parallel probabilistic runs on four NVIDIA Tesla V100 GPU cards. The applicability of the analysis beyond any of these limitations is discussed in the following.

##### 4.1. Non-Uniform Distributions for the Uncertain Input Variables

The present UQ analysis framework can be reproduced using a Gaussian or any other probability distribution per input random variable. In fact, such non-uniform distributions can be generated from the cumulative distribution functions of the uniform distributions (Equations 1–3) by means of the Inverse Transform Sampling (Steinbrecher & Shaw, 2008) or the Box-Muller algorithm (Box & Muller, 1958). However, these non-uniform distributions would have the same data means and standard deviations, and thus may not significantly change the findings of this study (Haan et al., 1998). Rather, there is an impending need for producing well-characterized uncertainty with statistically rigorous estimation of the potential underlying correlations in their domain of variations, followed by model-based reduction of the uncertainty space (Stefanescu et al., 2012).

##### 4.2. Choice of the Physical Solver

The present GPU-FV1 physical solver on LISFLOOD-FP is best suited for probabilistic modeling of rapid flood types. At present, it runs on a single GPU card and is thus limited to case studies with spatial domain size to DEM resolution leading to a number of cells that fits within the GPU card's memory capacity. Simulations over large-size domains require deploying another GPU-FV1 solver that can effectively utilize hundreds of GPUs across thousands of CPUs in a heterogeneous computing architecture (for example, Morales-Hernández et al., 2021), to also perform the batches of probabilistic runs more efficiently.

For probabilistic modeling of slowly propagating fluvial/pluvial floods over large catchments, a less complex (mathematically and numerically) physical solver may be more appropriate to expedite runtimes, such as the reduced acceleration solver on LISFLOOD-FP (Beevers et al., 2020), which has a version for the GPU with grid-resolution adaptivity to maximize runtime efficiency (Sharifian et al., 2023). Alternatively, physical solvers based on distributed hydrological modeling can be used for multi-physics modeling such as to incorporate the

feedbacks between groundwater and land surface processes (for example, Baroni et al., 2019). In the latter case, a relatively large input uncertainty space would be expected, with  $D > 5$ , (for example, Merz et al., 2020), requiring further measures to address the curse of dimensionality that will affect all the UQ methods.

#### 4.3. Measures to Address the Curse of Dimensionality ( $D > 5$ )

As identified in this study, the higher the  $D$  the higher the  $N_s$  required by any of the UQ methods, making them not affordable with physical solvers for probabilistic hydraulic and/or hydrological simulations with  $D \gg 5$  at the catchment scale and beyond. In such a case, one measure would be to use a data-driven emulator to reduce the number of evaluations of the physical solver to just train a sub-sample of  $N_s$  to produce the data set from which the emulator produces the predictions. This type of modeling has been explored for flood modeling using physical solvers, including LISFLOOD-FP, to train and validate different types of emulators (for example, Chu et al., 2020; Donnelly et al., 2022; Kabir et al., 2020). Still, this study provides strong evidence to prefer using QMC with any type of emulators since it offers a higher reduction in  $N_s$  compared to the other UQ methods as  $D$  is increased.

#### 4.4. Assumption of Fixed and Correct Input Uncertainty

This work has assumed fixed and correct input uncertainty assumptions with no attempt to gain knowledge about the input uncertainty from observation such as using Bayesian analysis, like in GLUE (Beven & Binley, 1992). The work can be considered usefully applicable for real-world flood modeling problems where detailed data sets (i.e., water depth and/or velocity time series) are unavailable to acquire a characterization of the input uncertainty as is too often the case (Boelee et al., 2019). Still, it highlights a potential for deploying QMC in general uncertainty analysis framework for flood modeling to estimate Bayesian inference more efficiently (Buchholz & Chopin, 2019), such as within GLUE and Bluecat (Koutsosoyannis & Montanari, 2022).

### 5. Summary and Conclusions

Four UQ methods were assessed to find alternatives to the SMC method for reproducing flood-related histograms, efficiently at a reduced sample size: two based on random sampling, which are LHS and ASS, and two based on deterministic realization, which are Quasi Monte Carlo (QMC) and HWE. The reproduced flood-related histograms were evaluated for three quantities of interest, the Flood Extent and the average and maximum HR ( $HR_{ave}$  and  $HR_{max}$ ). These quantities stemmed from the probabilistic modeling of torrential and fluvial floodplain flows, impacted by uncertainty from at least two input random variables amongst the inflow discharge(s), the Manning coefficient and the ground elevation. The relative histograms predicted by each of the four alternative-to-SMC methods were validated against the reference SMC prediction achieved by brute-force probabilistic runs using a much larger sample size.

First, the four methods were exhaustively compared for a synthetic rapidly propagating flood over a smooth terrain to include diagnostic analyses of their orders of converge for two sub-cases: one with two input random variables for the inflow discharge and the Manning coefficient; and, the other with three input random variables, further incorporating the input random variable for the ground elevation. The analyses identified a sample size between 2,000 and 4,000 for the four methods to keep the relative histogram difference below an average threshold around 7.5% with respect to the reference prediction. The identified sample sizes were mostly based on  $HR_{max}$ , since it is the most uncertain quantity of interest, exhibiting the highest non-smoothness level in the responses, and on the largest number of bins, of 40, since using fewer bins led to better sample size reductions. With two input random variables, the sample size for LHS was slightly smaller than that for ASS, both yielding about 900 samples less than the sample size predicted for SMC, or a relative-to-SMC speedup ratio in the range of 1.2-to-1.8. However, LHS and ASS yielded no considerable speedup ratios with three input random variables; despite this, the sample size for ASS became smaller than that of LHS, suggesting a tendency for ASS to potentially outperform if the number of input random variables is increased beyond three. QMC and HWE yield higher speedup ratios, in range of 1.1-to-25, with two input random variables; the lower bound of 1.1 was caused by the inflexibility of HWE to use any sample size between 1,024 and 4,096, as opposed to QMC that had a four-times higher lower bound. With three input random variables, the speedup ratios for QMC and HWE dropped, to a range of 0.5-to-2.0, though QMC preserved an average speed up ratio of 1.6; again, the lower and upper bounds

were fluctuations arising from the (aforementioned) inflexibility of HWE. Common to both sub-cases, QMC entailed a sample size closer to 2000, compared to the sample size for LHS, ASS and HWE that were closer to 4,000.

Second, the two sub-cases were reconsidered for a torrential flooding over a rough river valley to validate the identified sample sizes against the test-specific reference prediction. Flood Extent was the most uncertain quantity of interest, exhibiting the highest non-smooth responses level. ASS captured the bimodality in the Flood Extent histogram with a difference of 7.7%, as the number of input random variables increased to three, which is in contrast to LHS that achieved a higher difference, of 9.3%. In both sub-cases, QMC outperformed, predicting the histogram with a difference of 7%, while using a sample size that is three times smaller than that of HWE.

Finally, therefore, ASS and QMC were validated using sample sizes of 2,000 and 4,000 to reproduce the reference prediction for the Flood Extent histogram of a real-world fluvial flooding scenario with five input random variables. For the sample size of 2,000, both ASS and QMC capture the reference Flood Extent histogram, with a difference below the maximum threshold of 10%, with low number of bins, of 10 and 20; however, ASS failed to meet this threshold when the number of bins is enlarged to 40. For the larger sample size of 4,000, ASS and QMC predicted relative histogram differences that meet the maximum threshold, irrespective of the number of bins, suggesting that they are both valid choices for sample sizes as large as 4,000 to get a speedup ratio of 2.5. However, only QMC could meet this threshold to further reduce the sample size somewhere close to 2,000 and boost the speedup ratio to 5.

Overall, the comparative analyses in this study identify QMC to be the simplest and most efficient alternative-to-SMC for probabilistic flood modeling applications, including rapid and slow flows, driven by more than one input random variable but not exceeding five. The speedup ratio for QMC is in the range of 1.6-to-5. However, this ratio has been quantified for suboptimal conditions (i.e., for the relative histogram metric in order to capture the full details of the probability distributions, for the most uncertain flood-related quantity of interest and the largest number of bins); thus should be larger when using less sensitive metrics (e.g., the standard errors of the mean and variance) or when targeting a flood-related quantity with low and/or smooth variations in its responses. Despite the limitations of this study, its findings still provide useful insights into the potential utility of QMC to speedup probabilistic modeling of more sophisticated water resource problems including more than five random variables such as, for example, to use QMC with a multi-physics solver to support a data-driven model to minimize the number of sub-samples of the training data set.

### Appendix A

The convergence analyses, presented in Section 3.1, are re-explored using a threshold difference, with respect to the reference prediction, that is based on the standard errors of the mean,  $\bar{Y}$ , and the variance,  $S^2$ . Table A1 summarizes how  $\bar{Y}$  and  $S^2$  have been calculated per UQ method  $K$  ( $K = \text{SMC, LHS, ASS, QMC, HWE}$ ), noting that with ASS the weighted sum over all strata should be performed first, wherein the weight of the  $k$ th stratum is its relative volume  $p_k$  (Owen, 2013).

The convergence analyses are conducted for the same range of  $N_s^K$  (Section 2.5.1), to identify the  $N_s^K$  per UQ method  $K$  ( $K = \text{SMC, QMC, HWE, LHS, and ASS}$ ) that keep a threshold difference of  $0.0075\% \pm 0.0025\%$  and

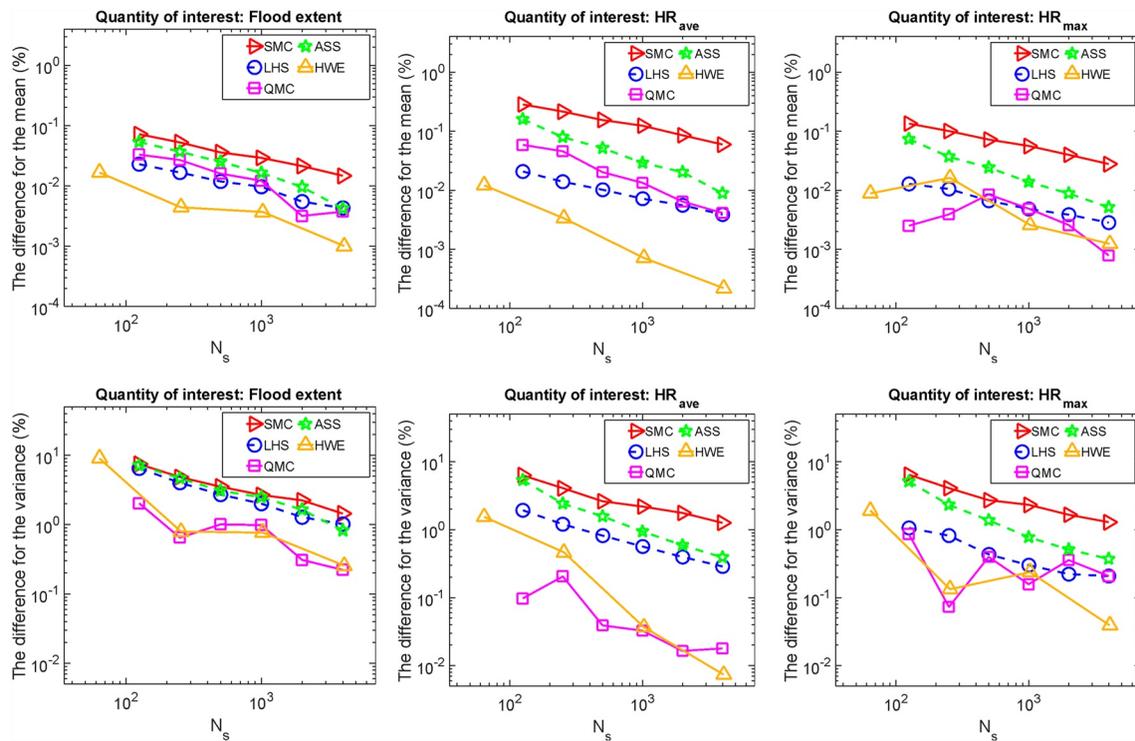
UQ method	$\bar{Y}$	$S^2$
SMC	$\frac{1}{N_s} \sum_{j=1}^{N_s} Y_j$	$\frac{1}{N_s} \sum_{j=1}^{N_s} (Y_j - \bar{Y})^2$
LHS		
QMC		
ASS	Number of strata $\sum_{k=1} p_k \bar{Y}_k$	Number of strata $\sum_{k=1} p_k S_k^2 + p_k (\bar{Y}_k - \bar{Y})^2$
HWE	$\hat{Y}_0$	$\frac{1}{2DL} \sum_{l=2}^{N_s} \sum_{m=1}^{N_s} \hat{Y}_j^2 \phi_j^2(X_m)$

of  $0.3\% \pm 0.1\%$  for  $\bar{Y}$  and  $S^2$ , respectively. From the identified  $N_s^K$  per UQ method  $K$  ( $K = \text{QMC, HWE, LHS, and ASS}$ ), the relative-to-SMC speedup ratios (Equation 5) are quantified.

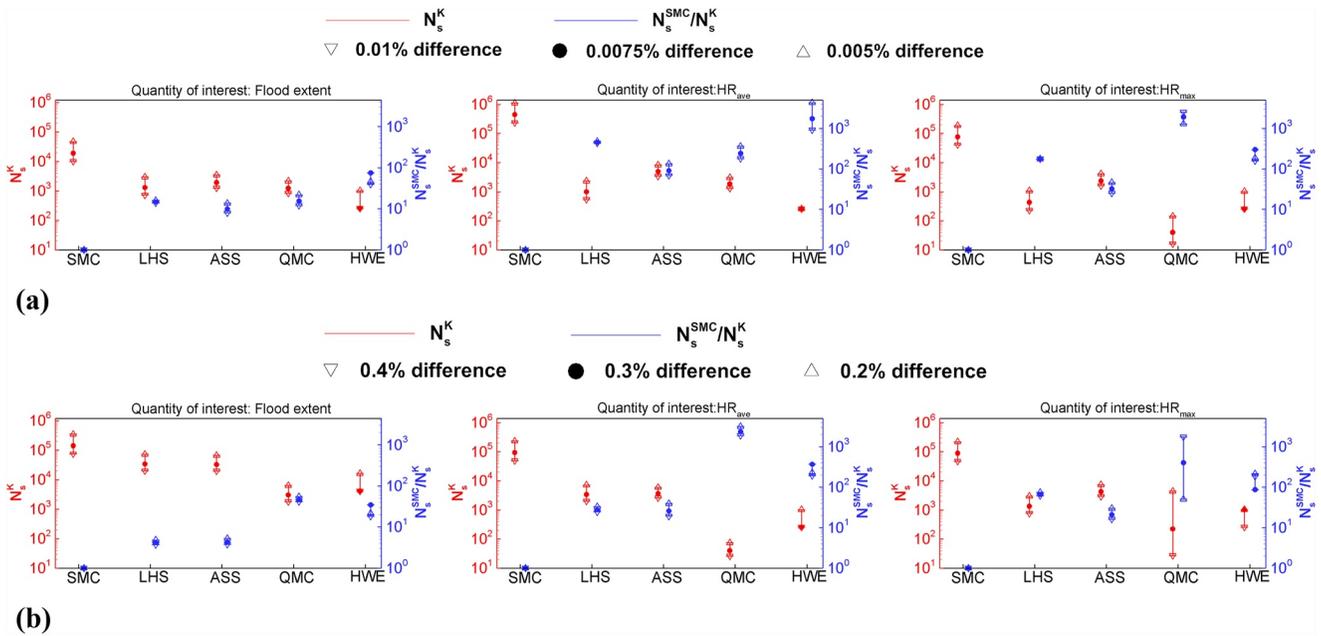
**A1. Sub-Case With  $D = 2$**

Figure A1 shows the order of convergence using the difference for the mean (upper panel) and variance (lower panel) with increased  $N_s$ , for each of the three quantities of interest with all the UQ methods. Based on the difference for the mean (Figure A1, upper panel), the order of error magnitudes are about three levels lower compared to the order of error magnitudes based on the relative histogram difference (compare with Figure 7). This shows that, for any selected  $N_s$ , the UQ methods better preserve the threshold difference for the mean as compared to the threshold relative histogram difference. Even with the smallest  $N_s$ , the differences for the mean are less than 1%, irrespective of the quantity of interest, which suggests that any UQ method is useable to accurately preserve the threshold difference for the mean for a small  $N_s$ . With larger  $N_s$ , any alternative-to-SMC UQ method has increasingly lower order of error magnitudes than that of SMC, thus become a better choice. The deterministic realization methods (HWE and QMC) outperform the random sampling methods (LHS and ASS) since they lead to lower order of error magnitudes. Among the former and latter methods, HWE and LHS have lower order-of-magnitudes than QMC and LHS, respectively, suggesting that they are slightly better options when considering the threshold mean difference.

Based on the difference for the variance (Figure A1, lower panel), the order of error magnitudes are about two levels bigger than the order of error magnitudes seen with the threshold mean difference (compare with Figure A1, upper panel), but still one level lower than the order of error magnitude seen with the threshold relative histogram difference (compare with Figure 7). Therefore, the UQ methods can still offer better reduction in  $N_s$  compared to the relative histogram difference even when targeting a difference for the variance—that is more sensitive in variability than the mean. For the smallest  $N_s$ , the order of error magnitudes are very close to each other, far below 10%, indicating any UQ method is useable to accurately to keep the target difference for the variance for a small  $N_s$ . As  $N_s$  is increased, the alternative-to-SMC outperform: the deterministic realization methods (HWE and



**Figure A1.** Rapidly propagating flood over a smooth terrain (Section 2.5.1). Order of convergence per uncertainty quantification method  $K$  ( $K = \text{standard Monte Carlo, QMC, Haar-wavelet expansion, Latin hypercube sampling, and adaptive stratified sampling}$ ) calculated for Flood Extent,  $HR_{ave}$ , and  $HR_{max}$  based on the difference in the mean (upper panel) and variance (lower panel).



**Figure A2.** Rapidly propagating flood over a smooth terrain (Section 2.5.1). The  $N_s^K$  and relative-to-SMC speedup ratio for Flood Extent,  $HR_{ave}$ , and  $HR_{max}$  to meet the target difference: (a) of  $0.0075\% \pm 0.0025\%$  based on the standard error of the mean (upper panel), and (b) of  $0.3\% \pm 0.1\%$  based on the standard error of the variance (lower panel).

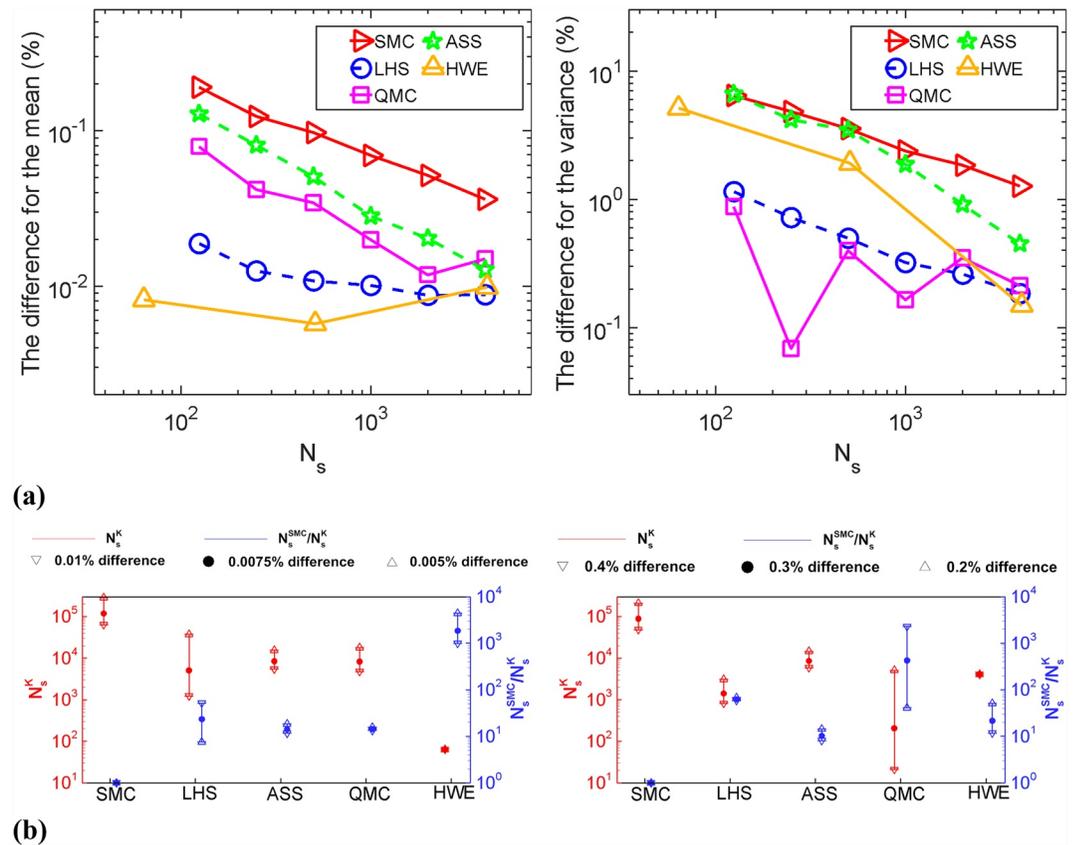
QMC) show lower order of error magnitudes compared the random sampling methods (SMC, LHS, and ASS), thus can deliver a lower difference for the variance for a fixed  $N_s$  leading to more reductions. Among QMC and HWE there is no clear winner, and LHS outperforms ASS given its lower order-of-magnitudes.

Figure A2 shows the  $N_s^K$  identified for each uncertainty quantification (UQ) method  $K$  and the relative-to-SMC speedup ratios  $N_s^{SMC}/N_s^K$  to keep the threshold difference of  $0.0075\% \pm 0.0025\%$  for the mean  $\bar{Y}$  (Figure A2, upper panel), and of  $0.3\% \pm 0.1\%$  for the variance  $S^2$  (Figure A2, lower panel). For the threshold mean difference, a speedup ratio range of 6-to-4334 is achieved by the UQ methods. This is much higher than the range of 1.1-to-25 identified using the relative histogram difference (compare with Figure 8). The deterministic realization methods (QMC and Haar-wavelet expansion (HWE)) achieve a range of 12-to-4334 whereas the random sampling methods (Latin hypercube sampling (LHS) and adaptive stratified sampling) have a comparatively lower range of 6-to-474. Among the former methods, HWE outperform leading to higher lower and upper bounds, of 39 and 4,334, compared to those reached by QMC, of 12 and 2,538, respectively. Whereas LHS outperforms among the latter methods leading to higher upper and lower bounds, of 14 and 474, compared to those reached by ASS, of 6 and 116, respectively.

For the threshold variance difference, the speedup ratios achieved by the UQ methods are a bit lower than the ratio identified for the threshold mean difference (compare Figure A1, lower vs. upper panels), and has a range of 4-to-3174. This range is much higher than that the range identified for the relative histogram difference (compared with Figure 8). This shows that the variance is a slightly more demanding on the  $N_s$  for a UQ method compared to the mean, which may be expected, but none of these conventional statistical metrics are as challenging compared to when considering the relative histogram metric. Here, the deterministic realization methods (QMC and HWE) again outperform the random sampling methods (LHS and ASS) with a speedup ratio range of 18-to-3174 for the former methods that is higher than that of 4-to-74 for the latter methods. The highest speedup ratio of 3,174 is obtained by QMC, suggesting that it is likely to be the best performer overall to reduce  $N_s$  when capturing sensitive responses such as the variance for  $HR_{max}$ .

A2. Sub-Case With  $D = 3$

Figure A3(a) shows the order of convergence for the threshold mean difference (left) and for the threshold variance difference (right) for  $HR_{max}$ . Increasing the uncertainty space to  $D = 3$  makes the choice among the UQ methods become more competitive, especially among LHS, QMC, and HWE that now have smaller order of error magnitudes than SMC and ASS. The order of error magnitudes of LHS lies between HWE and QMC. For the threshold mean difference, HWE leads to the lowest order of error magnitudes, while the lowest order of error magnitudes for the variance are attained by QMC. Therefore, the deterministic realization methods (QMC and HWE) are a better choice than the sampling methods (LHS and ASS). Figure A3(b) shows the  $N_s^K$  identified per UQ method  $K$  and the relative-to-SMC speedup ratios  $N_s^{SMC}/N_s^K$  to keep the threshold difference of  $0.0075\% \pm 0.0025\%$  for  $\bar{Y}$  (left), and of  $0.3\% \pm 0.1\%$  for  $S^2$  (right). The speedup ratio has a range of 12-to-4,431, which is much greater than the range of 1-to-2 seen for the relative histogram difference (compared with Figure 11b). HWE and QMC deliver the highest range for the speedup ratio, with higher upper and lower bounds identified with HWE based on the threshold mean difference and with QMC based on the threshold variance difference. Hence, QMC is better suited to attain the most reduction in  $N_s$  when analyzing a complex quantity of interest using a sensitive statistical metric.



**Figure A3.** Rapidly propagating flood over a smooth terrain (Section 2.5.1). Panel (a) shows the order of convergence per UQ method  $K$  ( $K = SMC, QMC, HWE, LHS,$  and  $ASS$ ) calculated for  $HR_{max}$  based on the difference in the standard errors of the mean (left) and the variance (right); and, panel (b) shows the  $N_s^K$  and relative-to-SMC speedup ratio for  $HR_{max}$  to meet the target difference: of  $0.0075\% \pm 0.0025\%$  based on the standard error of the mean (left), and of  $0.3\% \pm 0.1\%$  based on the standard error of the variance (right).

## Appendix B

Matlab source codes are openly available (Hajihassanpour et al., 2022) to run the UQ analysis framework. The codes are separated into two parts: (i) *sampling and realization* and (ii) statistics and difference (error) study. Part (i) contains the codes to generate samples for a selected UQ method and to then calls the GPU-FV1 physical solver of LISFLOOD-FP (Shaw et al., 2021) to perform realizations at each sample, to finally post-process the realizations into the three quantities of interest. The parameters listed in Table B1 need to be defined in the script file `main.m`. The files needed to run LISFLOOD-FP (i.e., with extensions of `.par`, `.bci`, `.dby`, and `.dem`) should be placed in the same directory as `main.m`. Running `main.m` will produce the following output files in `.mat` format.

<code>rand.mat</code>	Samples in the uncertainty space generated by a UQ method
<code>UI.mat</code>	The information of the uncertain input variables including their original values, perturbation parameters, perturbed values, etc.
<code>stratum.mat</code>	Only needs to be activated with ASS to record the stratification information such as the location of each stratum in the uncertainty space, the position of samples in each stratum, the mean value, the variance value, etc.
<code>FloodExtent.mat</code>	Flood Extent quantity evaluated at the samples
<code>HR_ave.mat</code>	$HR_{ave}$ quantity evaluated at the samples
<code>HR_max.mat</code>	$HR_{max}$ quantity evaluated at the samples

**Table B1**

*Part (i) Input Parameters to Generate Samples and to Perform Simulations Using LISFLOOD-FP 8.0*

Parameters	Description
SamplingMethod	Set to "Pseudorandom" to run SMC Set to "LHS" to run LHS Set to "Hammersley" to run QMC Set to "HaarQuadrature" to run HWE
RefinementLevel	Only used with HWE, and is the refinement level. It takes a positive integer value
NumSamples	Number of samples with LHS, QMC and SMC
UncertaintyTypes	The type of uncertain input variables for the three test cases in Section 3: Rapidly propagating flood over a smooth terrain (Sections 2.5.1 and 3.1) ["Manning," "Topography"] for the case with $D = 2$ ["Discharges," "Manning," "Topography"] for the case with $D = 3$ Torrential flooding over a rough river valley (Sections 2.5.2 and 3.2) ["Manning," "Topography"] for the case with $D = 2$ ["Discharges," "Manning," "Topography"] for the case with $D = 3$ Carlisle 2005 flooding (Sections 2.5.3 and 3.3) ["Discharges," "Discharges," "Discharges," "Manning," "Topography"] for the case with $D = 5$

**Table B1**  
*Continued*

Parameters	Description
UncertaintyTypesValues	The total amount of uncertainty for each uncertain variable defined in <code>UncertaintyTypes</code> . For example, the inflow discharge variable has $\pm 8\%$ uncertainty, which means 16% in total Rapidly propagating flood over a smooth terrain (Sections 2.5.1 and 3.1) [16.0, 10.0] for the case with $D = 2$ [16.0, 10.0, 10.0] for the case with $D = 3$ Torrential flooding over a rough river valley (Sections 2.5.2 and 3.2) [16.0, 10.0] for the case with $D = 2$ [16.0, 10.0, 0.25] for the case with $D = 3$ Carlisle 2005 flooding (Sections 2.5.3 and 3.3) [16.0, 16.0, 16.0, 10.0, 0.08] for the case with $D = 5$
TimesToEvalFloodAndHR	The time when Flood Extent, $HR_{ave}$ or $HR_{max}$ evaluate. 193 for the Rapidly propagating flood over a smooth terrain test case (Sections 2.5.1) 15,660 for the Torrential flooding over a rough river valley test case (Sections 2.5.2) 144,000 for the Carlisle 2005 flooding test case (Sections 2.5.3)
FileExeAddress	The address of the location of the physical solver (LISFLOOD-FP 8.0) executable file
ParameterFileName	The file name with the extension <code>.par</code> needed for running LISFLOOD-FP 8.0
QuantOfInter	Only for ASS, and indicates the quantity of interest that can be any of "Flood Extent," " $HR_{ave}$ " or " $HR_{max}$ "

Part (ii) utilizes outputs of part (i) to evaluate the difference against the reference prediction and the relative-to-SMC speedup ratio considering: the relative histogram difference as a metric (Sections 3.1–3.3), or the statistical metrics of the mean and the variance (Appendix A). Part (ii) includes three Matlab source codes.

- `ErrorStudyWithReplications.m` to evaluate the difference against the reference prediction (Sections 3.1 and Appendix A);
- `ComparingMethodsHistogram.m` to qualitatively and quantitatively compare a relative histogram per UQ method against the reference relative histogram (Sections 3.1–3.3);
- `PostProcessing.m` to plot the uncertain input variables along with associated response surfaces, and the sampling patterns in the uncertainty space (Sections 2.5 and 3.1).

### Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

### Data Availability Statement

Some or all data, models, or code generated or used during the study are available in a repository or online in accordance with funder data retention policies. The GPU-FV1 code is openly available on LISFLOOD-FP 8.0 with <https://doi.org/10.5281/zenodo.4073011>, with instructions on how to download, set-up and run the code available on: <https://www.seamlesswave.com/LISFLOOD8.0>. The codes to reproduce the comparative analysis for the five UQ methods are openly available on: <https://doi.org/10.5281/zenodo.7050213>, with the instructions for running them detailed in Appendix B.

**Acknowledgments**

We wish to thank Onno Bokhove (University of Leeds) and Domenico Bau (University of Sheffield) for their useful feedback on Sections 2 and 3. Mahya Hajihassanpour and Georges Kesserwani were supported by the UK Engineering and Physical Sciences Research Council Grant EP/R007349/1. This work is part of the SEAMLESS-WAVE project (SoftwareE infrAstructure for Multi-purpose Flood modELLing at variouS scaleS based on WAVElets (<https://www.seamless-wave.com>)). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising.

**References**

Abgrall, R., & Mishra, S. (2017). Uncertainty quantification for hyperbolic systems of conservation laws. In *Handbook of numerical analysis* (pp. 507–544).

Aitken, G., Beevers, L., & Christie, A. (2022). Multi-level Monte Carlo models for flood inundation uncertainty quantification. *Water Resources Research*, 58(11), 1–25. <https://doi.org/10.1029/2022wr032599>

Alipour, A., Jafarzadegan, K., & Moradkhani, H. (2022). Global sensitivity analysis in hydrodynamic modeling and flood inundation mapping. *Environmental Modelling and Software*, 152, 105398. <https://doi.org/10.1016/j.envsoft.2022.105398>

Apel, H., Thieken, A. H., Merz, B., & Blöschl, G. (2004). Flood risk assessment and associated uncertainty. *Natural Hazards and Earth System Sciences*, 4(2), 295–308. <https://doi.org/10.5194/nhess-4-295-2004>

Arnell, N. W., & Gosling, S. N. (2016). The impacts of climate change on river flood risk at the global scale. *Climatic Change*, 134(3), 387–401. <https://doi.org/10.1007/s10584-014-1084-5>

Aronica, G., Hankin, B., & Beven, K. (1998). Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data. *Advances in Water Resources*, 22(4), 349–365. [https://doi.org/10.1016/s0309-1708\(98\)00017-7](https://doi.org/10.1016/s0309-1708(98)00017-7)

Avasarala, S., & Subramani, D. (2021). A non-Gaussian Bayesian filter for sequential data assimilation with non-intrusive polynomial chaos expansion. *International Journal for Numerical Methods in Engineering*, 122(23), 7156–7181. <https://doi.org/10.1002/nme.6827>

Ayog, J. L., Kesserwani, G., Shaw, J., Sharifian, M. K., & Bau, D. (2021). Second-order discontinuous Galerkin flood model: Comparison with industry-standard finite volume models. *Journal of Hydrology*, 594, 125924. <https://doi.org/10.1016/j.jhydrol.2020.125924>

Baroni, G., Schalge, B., Rakovec, O., Kumar, R., Schüler, L., Samaniego, L., et al. (2019). A comprehensive distributed hydrological modeling intercomparison to support process representation and data collection strategies. *Water Resources Research*, 55(2), 990–1010. <https://doi.org/10.1029/2018wr023941>

Bates, P. D., Pappenberger, F., & Romanowicz, R. J. (2011). Uncertainty in flood inundation modelling. In *Applied uncertainty analysis for flood risk management* (pp. 232–269). Imperial College Press.

Beevers, L., Collet, L., Aitken, G., Maravat, C., & Visser, A. (2020). The influence of climate model uncertainty on fluvial flood hazard estimation. *Natural Hazards*, 104(3), 2489–2510. <https://doi.org/10.1007/s11069-020-04282-4>

Bellos, V., Nalbantis, I., & Tsakiris, G. (2018). Friction modeling of flood flow simulations. *Journal of Hydraulic Engineering*, 144(12), 04018073. [https://doi.org/10.1061/\(asce\)hy.1943-7900.0001540](https://doi.org/10.1061/(asce)hy.1943-7900.0001540)

Bellos, V., & Tsihrintzis, V. (2021). Uncertainty aspects of 2D flood modelling in a benchmark case study. In *17th International Conference on Environmental Science and Technology, Athens, Greece*.

Bermúdez, M., Neal, J. C., Bates, P. D., Coxon, G., Freer, J. E., Cea, L., & Puertas, J. (2017). Quantifying local rainfall dynamics and uncertain boundary conditions into a nested regional-local flood modeling system. *Water Resources Research*, 53(4), 2770–2785. <https://doi.org/10.1002/2016wr019903>

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, 6(3), 279–298. <https://doi.org/10.1002/hyp.3360060305>

Boelee, L., Lumbroso, D. M., Samuels, P. G., & Cloke, H. L. (2019). Estimation of uncertainty in flood forecasts—A comparison of methods. *Journal of Flood Risk Management*, 12(Suppl. 1), e12516. <https://doi.org/10.1111/jfr3.12516>

Botev, Z., & Ridder, A. (2017). *Variance reduction* (pp. 1–6). Wiley StatsRef: Statistics Reference Online.

Box, G. E., & Muller, M. E. (1958). A note on the generation of random normal deviates. *The Annals of Mathematical Statistics*, 29(2), 610–611. <https://doi.org/10.1214/aoms/1177706645>

Buchholz, A., & Chopin, N. (2019). Improving approximate Bayesian computation via quasi-Monte Carlo. *Journal of Computational & Graphical Statistics*, 28(1), 205–219. <https://doi.org/10.1080/10618600.2018.1497511>

Cea Gómez, L., Bladé i Castellet, E., Sanz Ramos, M., Fraga, I., Sañudo Costoya, E., García-Leal, O., et al. (2020). *Benchmarking of the Iber capabilities for 2D free surface flow modelling* (pp. 1–52). Universidade da Coruna.

Chatzivasileiadis, T. (2018). Quasi-Monte Carlo application in CGE systematic sensitivity analysis. *Applied Economics Letters*, 25(21), 1521–1526. <https://doi.org/10.1080/13504851.2018.1430322>

Chu, H., Wu, W., Wang, Q. J., Nathan, R., & Wei, J. (2020). An ANN-based emulation modelling framework for flood inundation modelling: Application, challenges and future directions. *Environmental Modelling and Software*, 124, 104587. <https://doi.org/10.1016/j.envsoft.2019.104587>

De Luna-Cruz, F., Ramos-Hernández, J. G., Fuentes-Mariles, O. A., & Gracia-Sánchez, J. (2019). FluBiDi—A model to estimate flood based on runoff: Validation using extreme and natural basin conditions. *Hydrological Sciences Journal*, 64(3), 297–317.

Di Baldassarre, G., Schumann, G., Bates, P. D., Freer, J. E., & Beven, K. J. (2010). Flood-plain mapping: A critical discussion of deterministic and probabilistic approaches. *Hydrological Sciences Journal*, 55(3), 364–376. <https://doi.org/10.1080/02626661003683389>

Dimitriadis, P., Tegos, A., Oikonomou, A., Pagana, V., Koukouvinos, A., Mamassis, N., et al. (2016). Comparative evaluation of 1D and quasi-2D hydraulic models based on benchmark and real-world applications for uncertainty assessment in flood mapping. *Journal of Hydrology*, 534, 478–492. <https://doi.org/10.1016/j.jhydrol.2016.01.020>

Donnelly, J., Abolfathi, S., Pearson, J., Chatrabgoun, O., & Daneshkhah, A. (2022). Gaussian process emulation of spatio-temporal outputs of a 2D inland flood model. *Water Research*, 225, 119100. <https://doi.org/10.1016/j.watres.2022.119100>

Eldred, M. (2009). Recent advances in non-intrusive polynomial chaos and stochastic collocation methods for uncertainty analysis and design. In *50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. American Institute of Aeronautics and Astronautics.

Etore, P., Fort, G., Jourdain, B., & Moulines, E. (2011). On adaptive stratification. *Annals of Operations Research*, 189(1), 127–154. <https://doi.org/10.1007/s10479-009-0638-9>

Fan, Y. R., Huang, W. W., Li, Y. P., Huang, G. H., & Huang, K. (2015). A coupled ensemble filtering and probabilistic collocation approach for uncertainty quantification of hydrological models. *Journal of Hydrology*, 530, 255–272. <https://doi.org/10.1016/j.jhydrol.2015.09.035>

Fewtrell, T. J., Neal, J. C., Bates, P. D., & Harrison, P. J. (2011). Geometric and structural river channel complexity and the prediction of urban inundation. *Hydrological Processes*, 25(20), 3173–3186. <https://doi.org/10.1002/hyp.8035>

Ge, L., Cheung Kwok, F., & Kobayashi Marcelo, H. (2008). Stochastic solution for uncertainty propagation in nonlinear shallow-water equations. *Journal of Hydraulic Engineering*, 134(12), 1732–1743. [https://doi.org/10.1061/\(asce\)0733-9429\(2008\)134:12\(1732\)](https://doi.org/10.1061/(asce)0733-9429(2008)134:12(1732))

Ghanem, R. G., & Spanos, P. D. (2003). *Stochastic finite elements: A spectral approach*. Courier Corporation.

Giles, M. B. (2015). Multilevel Monte Carlo methods. *Acta Numerica*, 24, 259–328. <https://doi.org/10.1017/s096249291500001x>

Giunta, A. A., McFarland, J. M., Swiler, L. P., & Eldred, M. S. (2006). The promise and peril of uncertainty quantification using response surface approximations. *Structure and Infrastructure Engineering*, 2(3–4), 175–189. <https://doi.org/10.1080/15732470600590507>

Glasserman, P. (2004). *Monte Carlo methods in financial engineering*. Springer.

- Gu, L., Chen, J., Yin, J., Slater, L. J., Wang, H.-M., Guo, Q., et al. (2022). Global increases in compound flood-hot extreme hazards under climate warming. *Geophysical Research Letters*, *49*(8), e2022GL097726. <https://doi.org/10.1029/2022gl097726>
- Haan, C. T., Storm, D. E., Al-Issa, T., Prabhu, S., Sabbagh, G. J., & Edwards, D. R. (1998). Effect of parameter distributions on uncertainty analysis of hydrologic models. *Transactions of the ASAE*, *41*(1), 65–70. <https://doi.org/10.13031/2013.17158>
- Hajihassanpour, M., Kesserwani, G., Pettersson, P., & Bellos, V. (2022). Sampling-based uncertainty quantification methods. *Zenodo*. <https://doi.org/10.5281/zenodo.7050213>
- Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, *2*(1), 84–90. <https://doi.org/10.1007/bf01386213>
- Hammersley, J. M. (1960). Monte Carlo methods for solving multivariable problems. *Annals of the New York Academy of Sciences*, *86*(3), 844–874. <https://doi.org/10.1111/j.1749-6632.1960.tb42846.x>
- Heinrich, S. (1998). Monte Carlo complexity of global solution of integral equations. *Journal of Complexity*, *14*(2), 151–175. <https://doi.org/10.1006/jcom.1998.0471>
- Hickernell, F. J. (2018). The trio identity for quasi-Monte Carlo error. In *Paper presented at Monte Carlo and Quasi-Monte Carlo Methods*. Springer International Publishing.
- Horritt, M. S., Bates, P. D., Fewtrell, T. J., Mason, D. C., & Wilson, M. D. (2010). Modelling the hydraulics of the Carlisle 2005 flood event. *Proceedings of the Institution of Civil Engineers - Water Management*, *163*(6), 273–281. <https://doi.org/10.1680/wama.2010.163.6.273>
- Hu, P., Liu, X., & Hu, H. (2009). Accuracy assessment of digital elevation models based on approximation theory. *Photogrammetric Engineering and Remote Sensing*, *75*(1), 49–56. <https://doi.org/10.14358/pers.75.1.49>
- Huang, Y., & Qin, X. (2014). Uncertainty analysis for flood inundation modelling with a random floodplain roughness field. *Environmental Systems Research*, *3*(1), 9. <https://doi.org/10.1186/2193-2697-3-9>
- James, B. A. P. (1985). Variance reduction techniques. *Journal of the Operational Research Society*, *36*(6), 525–530. <https://doi.org/10.1057/jors.1985.88>
- Jung, Y., & Merwade, V. (2015). Estimation of uncertainty propagation in flood inundation mapping using a 1-D hydraulic model. *Hydrological Processes*, *29*(4), 624–640. <https://doi.org/10.1002/hyp.10185>
- Jung, Y., & Merwade, V. (2012). Uncertainty quantification in flood inundation mapping using generalized likelihood uncertainty estimate and sensitivity analysis. *Journal of Hydrologic Engineering*, *17*(4), 507–520. [https://doi.org/10.1061/\(asce\)he.1943-5584.0000476](https://doi.org/10.1061/(asce)he.1943-5584.0000476)
- Kabir, S., Patidar, S., Xia, X., Liang, Q., Neal, J., & Pender, G. (2020). A deep convolutional neural network model for rapid prediction of fluvial flood inundation. *Journal of Hydrology*, *590*, 125481. <https://doi.org/10.1016/j.jhydrol.2020.125481>
- Kalagnanam, J. R., & Diwekar, U. M. (1997). An efficient sampling technique for off-line quality control. *Technometrics*, *39*(3), 308–319. <https://doi.org/10.1080/00401706.1997.10485122>
- Karamouz, M., & Fereshtehpour, M. (2019). Modeling DEM errors in coastal flood inundation and damages: A spatial nonstationary approach. *Water Resources Research*, *55*(8), 6606–6624. <https://doi.org/10.1029/2018wr024562>
- Kesserwani, G., & Sharifian, M. K. (2023). (Multi)wavelet-based Godunov-type simulators of flood inundation: Static versus dynamic adaptivity. *Advances in Water Resources*, *171*, 104357. tentatively accepted. <https://doi.org/10.1016/j.advwatres.2022.104357>
- Koutsogiannis, D., & Montanari, A. (2022). Bluecat: A local uncertainty estimator for deterministic simulations and predictions. *Water Resources Research*, *58*(1), e2021WR031215. <https://doi.org/10.1029/2021wr031215>
- Kreibich, H., Piroth, K., Seifert, I., Maiwald, H., Kunert, U., Schwarz, J., et al. (2009). Is flow velocity a significant parameter in flood damage modelling? *Natural Hazards and Earth System Sciences*, *9*(5), 1679–1692. <https://doi.org/10.5194/nhess-9-1679-2009>
- Kucherenko, S., Albrecht, D., & Saltelli, A. (2015). Exploring multi-dimensional spaces: A comparison of Latin hypercube and quasi Monte Carlo sampling techniques. arXiv:1505.02350.
- Le Maitre, O. P., Knio, O. M., Najm, H. N., & Ghanem, R. G. (2004). Uncertainty propagation using Wiener–Haar expansions. *Journal of Computational Physics*, *197*(1), 28–57. <https://doi.org/10.1016/j.jcp.2003.11.033>
- Liu, H., & Hu, P. (2015). Accuracy assessment of LiDAR-derived digital elevation models based on approximation theory. *Remote Sensing*, *7*(6), 7062–7079. <https://doi.org/10.3390/rs70607062>
- Liu, X. (2011). Accuracy assessment of Lidar elevation data using survey marks. *Survey Review*, *43*(319), 80–93. <https://doi.org/10.1179/003962611x12894696204704>
- Lumbroso, D., & Davison, M. (2018). Use of an agent-based model and Monte Carlo analysis to estimate the effectiveness of emergency management interventions to reduce loss of life during extreme floods. *Journal of Flood Risk Management*, *11*(S1), S419–S433. <https://doi.org/10.1111/jfr3.12230>
- Maranzoni, A., D’Oria, M., & Rizzo, C. (2022). Quantitative flood hazard assessment methods: A review. *Journal of Flood Risk Management*, *16*(1), e12855. <https://doi.org/10.1111/jfr3.12855>
- Matsumoto, M., & Nishimura, T. (1998). Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, *8*(1), 3–30. <https://doi.org/10.1145/272991.272995>
- McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*(2), 239–245. <https://doi.org/10.1080/00401706.1979.10489755>
- McMillan, H. K., & Brasington, J. (2008). End-to-end flood risk assessment: A coupled model cascade with uncertainty estimation. *Water Resources Research*, *44*(3). <https://doi.org/10.1029/2007wr005995>
- Merz, R., Tarasova, L., & Basso, S. (2020). Parameter’s controls of distributed catchment models - How much information is in conventional catchment descriptors? *Water Resources Research*, *56*(2), e2019WR026008. <https://doi.org/10.1029/2019wr026008>
- Morales-Hernández, M., Sharif, M. B., Kalyanapu, A., Ghafoor, S. K., Dullo, T. T., Gangrade, S., et al. (2021). TRITON: A multi-GPU open source 2D hydrodynamic flood model. *Environmental Modelling and Software*, *141*, 105034. <https://doi.org/10.1016/j.envsoft.2021.105034>
- Morokoff, W. J., & Caflisch, R. E. (1995). Quasi-Monte Carlo integration. *Journal of Computational Physics*, *122*(2), 218–230. <https://doi.org/10.1006/jcph.1995.1209>
- Müller, F., Jenny, P., & Meyer, D. W. (2013). Multilevel Monte Carlo for two phase flow and Buckley–Leverett transport in random heterogeneous porous media. *Journal of Computational Physics*, *250*, 685–702. <https://doi.org/10.1016/j.jcp.2013.03.023>
- Nalbantis, I., Papageorgaki, I., Sioras, P., & Ioannidis, C. (2017). Effect of uncertainty in Digital Surface Models on the extent of inundated areas. *Hydrological Processes*, *31*(9), 1760–1775. <https://doi.org/10.1002/hyp.11148>
- Neal, J., Keef, C., Bates, P., Beven, K., & Leedal, D. (2013). Probabilistic flood risk mapping including spatial dependence. *Hydrological Processes*, *27*(9), 1349–1363. <https://doi.org/10.1002/hyp.9572>
- Neal, J. C., Bates, P. D., Fewtrell, T. J., Hunter, N. M., Wilson, M. D., & Horritt, M. S. (2009). Distributed whole city water level measurements from the Carlisle 2005 urban flood event and comparison with hydraulic model simulations. *Journal of Hydrology*, *368*(1), 42–55. <https://doi.org/10.1016/j.jhydrol.2009.01.026>

- Neelz, S., & Pender, G. (2013). *Benchmarking the latest generation of 2D hydraulic modelling packages: Report - SCI20002*. Environment Agency.
- Ng, L. W. T., & Willcox, K. E. (2014). Multifidelity approaches for optimization under uncertainty. *International Journal for Numerical Methods in Engineering*, *100*(10), 746–772. <https://doi.org/10.1002/nme.4761>
- Nkwunonwo, U. C., Whitworth, M., & Baily, B. (2020). A review of the current status of flood modelling for urban flood risk management in the developing countries. *Scientific African*, *7*, e00269. <https://doi.org/10.1016/j.sciaf.2020.e00269>
- Owen, A. B. (2013). *Monte Carlo theory, methods and examples*. Art Owen.
- Pasupathy, R., Schmeiser, B. W., Taaffe, M. R., & Wang, J. (2012). Control-variate estimation using estimated control means. *IIE Transactions*, *44*(5), 381–385. <https://doi.org/10.1080/0740817x.2011.610430>
- Peherstorfer, B., Willcox, K., & Gunzburger, M. (2016). Optimal model management for multifidelity Monte Carlo estimation. *SIAM Journal on Scientific Computing*, *38*(5), A3163–A3194. <https://doi.org/10.1137/15m1046472>
- Pettersson, M. P., Iaccarino, G., & Nordstrom, J. (2015). Polynomial chaos methods for hyperbolic partial differential equations. *Springer Math Eng*, *10*(1007), 978–973.
- Pettersson, P., & Krumscheid, S. (2022). Adaptive stratified sampling for nonsmooth problems. *International Journal for Uncertainty Quantification*, *12*(6), 71–99. <https://doi.org/10.1615/int.j.uncertaintyquantification.2022041034>
- Pharr, M., Jakob, W., & Humphreys, G. (2016). *Physically based rendering: From theory to implementation*. Morgan Kaufmann.
- Rahman, A., Weinmann, P. E., Hoang, T. M. T., & Laurensen, E. M. (2002). Monte Carlo simulation of flood frequency curves from rainfall. *Journal of Hydrology*, *256*(3), 196–210. [https://doi.org/10.1016/s0022-1694\(01\)00533-9](https://doi.org/10.1016/s0022-1694(01)00533-9)
- Ramsbottom, D., Wade, S., Bain, V., Hassan, M., Penning-Rowsell, E., Wilson, T., et al. (2006). *Flood risks to people methodology: Phase 2*. R&D Technical Report FD2321/TR2, DEFRA – Department for the Environment, Food and Rural Affairs, UK.
- Rubinstein, R. Y., & Kroese, D. P. (2016). *Simulation and the Monte Carlo method*. John Wiley and Sons.
- Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). The Earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, *40*(2), 99–121. <https://doi.org/10.1023/a:1026543900054>
- Savage, J. T. S., Pianosi, F., Bates, P., Freer, J., & Wagener, T. (2016). Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model. *Water Resources Research*, *52*(11), 9146–9163. <https://doi.org/10.1002/2015wr018198>
- Sharafati, A., Khazaei, M. R., Nashwan, M. S., Al-Ansari, N., Yaseen, Z. M., & Shahid, S. (2020). Assessing the uncertainty associated with flood features due to variability of rainfall and hydrological parameters. *Advances in Civil Engineering*, 7948902. <https://doi.org/10.1155/2020/7948902>
- Sharifian, M. K., Kesserwani, G., Chowdhury, A. A., Neal, J., & Bates, P. (2023). LISFLOOD-FP 8.1: New GPU accelerated solvers for faster fluvial/pluvial flood simulations. *Geoscientific Model Development*, *16*(9), 2391–2413. <https://doi.org/10.5194/gmd-16-2391-2023>
- Shaw, J., & Kesserwani, G. (2020). Stochastic Galerkin finite volume shallow flow model: Well-balanced treatment over uncertain topography. *Journal of Hydraulic Engineering*, *146*(3), 04020005. [https://doi.org/10.1061/\(asce\)hy.1943-7900.0001705](https://doi.org/10.1061/(asce)hy.1943-7900.0001705)
- Shaw, J., Kesserwani, G., Neal, J., Bates, P., & Sharifian, M. K. (2021). LISFLOOD-FP 8.0: The new discontinuous Galerkin shallow-water solver for multi-core CPUs and GPUs. *Geoscientific Model Development*, *14*(6), 3577–3602. <https://doi.org/10.5194/gmd-14-3577-2021>
- Shaw, J., Kesserwani, G., & Pettersson, P. (2020). Probabilistic Godunov-type hydrodynamic modelling under multiple uncertainties: Robust wavelet-based formulations. *Advances in Water Resources*, *137*, 103526. <https://doi.org/10.1016/j.advwatres.2020.103526>
- Shields, M. D., Teferra, K., Hapij, A., & Daddazio, R. P. (2015). Refined Stratified Sampling for efficient Monte Carlo based uncertainty quantification. *Reliability Engineering and System Safety*, *142*, 310–325. <https://doi.org/10.1016/j.res.2015.05.023>
- Shirvani, M., & Kesserwani, G. (2021). Flood-pedestrian simulator for modelling human response dynamics during flood-induced evacuation: Hillsborough stadium case study. *Natural Hazards and Earth System Sciences*, *21*(10), 3175–3198. <https://doi.org/10.5194/nhess-21-3175-2021>
- Shirvani, M., Kesserwani, G., & Richmond, P. (2021). Agent-based simulator of dynamic flood-people interactions. *Journal of Flood Risk Management*, *14*(2), e12695. <https://doi.org/10.1111/jfr3.12695>
- Smemoe, C. M., Nelson, E. J., Zundel, A. K., & Miller, A. W. (2007). Demonstrating floodplain uncertainty using flood probability Maps1. *JAWRA Journal of the American Water Resources Association*, *43*(2), 359–371. <https://doi.org/10.1111/j.1752-1688.2007.00028.x>
- Sobol, I. M. (1979). On the systematic search in a hypercube. *SIAM Journal on Numerical Analysis*, *16*(5), 790–793. <https://doi.org/10.1137/0716058>
- Stefanescu, E. R., Bursik, M., Cordoba, G., Dalbey, K., Jones, M. D., Patra, A. K., et al. (2012). Digital elevation model uncertainty and hazard analysis using a geophysical flow model. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *468*(2142), 1543–1563. <https://doi.org/10.1098/rspa.2011.0711>
- Steinbrecher, G., & Shaw, W. T. (2008). Quantile mechanics. *European Journal of Applied Mathematics*, *19*(2), 87–112. <https://doi.org/10.1017/s0956792508007341>
- Stephens, T. A., & Bledsoe, B. P. (2020). Probabilistic mapping of flood hazards: Depicting uncertainty in streamflow, land use, and geomorphic adjustment. *Anthropocene*, *29*, 100231. <https://doi.org/10.1016/j.ancene.2019.100231>
- Stricker, M. A., & Orengo, M. (1995). Similarity of color images. In *Proceedings of SPIE 2420, Storage and Retrieval for Image and Video Databases III*.
- Tscheikner-Gratl, F., Bellos, V., Schellart, A., Moreno-Rodenas, A., Muthusamy, M., Langeveld, J., et al. (2019). Recent insights on uncertainties present in integrated catchment water quality modelling. *Water Research*, *150*, 368–379. <https://doi.org/10.1016/j.watres.2018.11.079>
- Wang, R., Diwekar, U., & Grégoire Padró, C. E. (2004). Efficient sampling techniques for uncertainties in risk analysis. *Environmental Progress*, *23*(2), 141–157. <https://doi.org/10.1002/ep.10020>
- Wechsler, S. P. (2007). Uncertainties associated with digital elevation models for hydrologic applications: A review. *Hydrology and Earth System Sciences*, *11*(4), 1481–1500. <https://doi.org/10.5194/hess-11-1481-2007>
- West, H., Horswell, M., & Quinn, N. (2018). Exploring the sensitivity of coastal inundation modelling to DEM vertical error. *International Journal of Geographical Information Science*, *32*(6), 1172–1193. <https://doi.org/10.1080/13658816.2018.1444165>
- Wong, T.-T., Luk, W.-S., & Heng, P.-A. (1997). Sampling with Hammersley and Halton points. *The Journal of Graphics Tools*, *2*(2), 9–24. <https://doi.org/10.1080/10867651.1997.10487471>
- Xiu, D., & Karniadakis, G. E. (2002). The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, *24*(2), 619–644. <https://doi.org/10.1137/s1064827501387826>
- Yu, P.-S., Yang, T.-C., & Chen, S.-J. (2001). Comparison of uncertainty analysis methods for a distributed rainfall-runoff model. *Journal of Hydrology*, *244*(1), 43–59. [https://doi.org/10.1016/s0022-1694\(01\)00328-6](https://doi.org/10.1016/s0022-1694(01)00328-6)
- Zhang, J. (2021). Modern Monte Carlo methods for efficient uncertainty quantification and propagation: A survey. *WIREs Computational Statistics*, *13*(5), e1539. <https://doi.org/10.1002/wics.1539>
- Zhao, J., & Liang, Q. (2022). Novel variable reconstruction and friction term discretisation schemes for hydrodynamic modelling of overland flow and surface water flooding. *Advances in Water Resources*, *163*(1), 104187. <https://doi.org/10.1016/j.advwatres.2022.104187>

- Zhu, F., Zhong, P.-A., Sun, Y., & Yeh, W. W. G. (2017). Real-time optimal flood control decision making and risk propagation under multiple uncertainties. *Water Resources Research*, 53(12), 10635–10654. <https://doi.org/10.1002/2017wr021480>
- Zio, S., & Rochinha, F. A. (2012). A stochastic collocation approach for uncertainty quantification in hydraulic fracture numerical simulation. *International Journal for Uncertainty Quantification*, 2(2), 145–160. <https://doi.org/10.1615/int.j.uncertaintyquantification.v2.i2.50>
- Zokagoa, J.-M., Soulaïmani, A., & Dupuis, P. (2021). Flood risk mapping using uncertainty propagation analysis on a peak discharge: Case study of the Mille Îles River in Quebec. *Natural Hazards*, 107(1), 285–310. <https://doi.org/10.1007/s11069-021-04583-2>