

**Balancing Costs and Benefits in Selecting New Information  
Efficient Monitoring Using Deterministic Hydro-economic Models**

Raso, L.; Weijs, S. V.; Werner, M.

**DOI**

[10.1007/s11269-017-1813-4](https://doi.org/10.1007/s11269-017-1813-4)

**Publication date**

2018

**Document Version**

Final published version

**Published in**

Water Resources Management

**Citation (APA)**

Raso, L., Weijs, S. V., & Werner, M. (2018). Balancing Costs and Benefits in Selecting New Information: Efficient Monitoring Using Deterministic Hydro-economic Models. *Water Resources Management*, 32(1), 1-19. <https://doi.org/10.1007/s11269-017-1813-4>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

# Balancing Costs and Benefits in Selecting New Information: Efficient Monitoring Using Deterministic Hydro-economic Models

L. Raso<sup>1</sup>  · S. V. Weijs<sup>2</sup> · M. Werner<sup>3</sup>

Received: 5 May 2017 / Accepted: 13 September 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** Uncertainty reduces reliability and performance of water system operations. A decision-maker can take action, accepting the present uncertainty and facing its risks, or reduce uncertainty by first obtaining additional information. Information, however, comes at a cost. The decision-maker must therefore efficiently use his/her resources, selecting the information that is most valuable. This can be done by solving the Optimal Design (OD) problem. The OD problem balances the cost of obtaining new information against its benefits, estimated as the added value of better informed actions. Despite its usefulness, the OD problem is not widely applied. Solving the OD problem requires a stochastic hydro-economic model, whereas most of hydro-economic models presently used are deterministic. In this paper we introduce an innovative methodology that makes deterministic hydro-economic models usable in the OD problem. In deterministic models, Least Squares Estimation (LSE) is often used to identify unknown parameters from data, making them, and hence the whole model, de facto stochastic. We advocate the explicit recognition of this uncertainty and propose to use it in risk-based decisions and formulation of the OD problem. The proposed methodology is illustrated on a flood warning scheme in the White Cart River, in Scotland, where a rating curve uncertainty can be reduced by extra gaugings. Gauging at higher flows is more informative but also more costly, and the LSE - OD formulation results in the economically optimal balance. We show the clear advantage of monitoring based on OD over monitoring uniquely based on uncertainty reduction.

---

✉ L. Raso  
l.raso@tudelft.nl

<sup>1</sup> Department of Policy Analysis, Delft University of Technology, Delft, The Netherlands

<sup>2</sup> Department of Civil Engineering, University of British Columbia, Vancouver, BC, Canada

<sup>3</sup> UNESCO-IHE, Delft, The Netherlands

**Keywords** Monitoring · Information · Efficiency · Optimal design · Least squares · Rating curve · Decision · Uncertainty · Hydro-economic model · Robustness · Flood protection · White cart river

## 1 Introduction

Water management requires decisions to be taken under uncertainty. Under uncertainty, the outcome of decisions can be only partially predicted as the future is never completely predictable, nor can uncertainty be eliminated. Uncertainty reduces efficiency and jeopardize robustness and reliability. Risk-based decisions, where uncertainty is explicitly taken into account, reduce the cost due to uncertainty (Weijs 2011; Verkade and Werner 2011), augmenting the system resilience against possible negative outcomes. Nonetheless, uncertainty always poses a risk.

In the face of uncertainty, a decision maker has two possible options: either accepting the current level of uncertainty, or reducing it by obtaining more information. Information is an intangible but valuable good, because it leads to better decisions. Obtaining information, however, comes at a cost. The secondary problem of i) deciding whether to get additional information and ii) selecting the most valuable new observation, is referred to as the Optimal Design (OD) problem (DeGroot 1962; Raiffa 1974).

Uncertainty can be measured by selected “indicators”, related to the variance or to the entropy. Uncertainty reduction is measured as an improvement of the indicator value, such as a reduction in squared error or Kullback-Leibler divergence (Weijs et al. 2010; Nearing and Gupta 2015). If these criteria are employed to select new information in an OD problem, the solution would indicate which observation is the most informative.

The most informative observation, however, is not necessarily the most valuable one (Alfonso et al. 2016; Bode et al. 2016). In the context of a defined decision problem, OD can be employed to quantify the expected usefulness of information, in terms of performance improvement due to better actions, balancing it with the expected costs to gather that new information. In this case, solving the OD problem will suggest i) which, among all possible observations, is the most valuable, and ii) whether to keep gathering information, or to stop and take action, accepting the present level of uncertainty. OD has been extensively applied in various domains, including water management (Yokota and Thompson 2004; Davis 1971; Maddock 1973; Bhattacharjya et al. 2010; Reed and Kollat 2012), where applications focus mostly on remediation problems in groundwater systems (Ben-Zvi et al. 1988; James and Gorelick 1994; Nowak et al. 2012; Bierkens 2006; Kim and Lee 2007; Mantoglou and Kourakos 2007; Chadalavada and Datta 2008). In case the decision problem is a model selection problem, information-based cost functions are most appropriate (Nowak and Guthke 2016; Weijs et al. 2010), while in economically important decisions, where actions are taken, a value of information approach can optimize benefits through the OD problem (Eidsvik and Ellefmo 2013; Trainor-Guitton et al. 2014).

The application of OD is limited by two factors: i) OD problem requires a stochastic hydro-economic model, which is the explicit representation of the most relevant uncertainties. Currently, however, most of the existing models in this domain are deterministic, and therefore unusable for solving the OD problem; ii) Solving the OD Problem requires a large computational effort, limiting its application to small systems. In this paper we introduce

an innovative methodology that makes deterministic hydro-economic models with parameters identified using Least Squares Estimation (LSE) directly usable in OD problem. The methodology uses some properties of LSE to get a leaner approximation that can be solved in a shorter time and, potentially, applied to larger systems.

The proposed technique is applied on a test-case in the White Cart river, in Scotland. In this test-case, the primary decision is whether to issue a warning in case predicted water levels exceed a defined threshold. The uncertainty of the rating curve can be reduced by new gaugings, which can be selected on a scale between observations at low-flow, i.e. low-cost but also less informative, and observation at high flow, i.e. costly and more informative.

This paper is structured as follows: In Section 2 we introduce the decision problem, the OD problem, and the innovative solution to solve the OD problem using deterministic model based on LSE; in Section 3 the proposed solution is applied to the test-case; the conclusions are discussed in Section 4.

## 2 Methodology

We start from the primary decision problem that the decision maker is faced with. In the utility theory framework (Neumann and Morgenstern 1947; Raiffa 1974), a decision problem under uncertainty can be defined as in Eq. 1a–d.

$$\text{find } \mathbf{u}^* \text{ such that} \quad (1a)$$

$$\mathcal{J}(\mathbf{u}^*, f(\boldsymbol{\lambda})) = \min_{\mathbf{u}} \mathcal{J}(\mathbf{u}, f(\boldsymbol{\lambda})) \quad (1b)$$

where

$$\mathcal{J}(\mathbf{u}, f(\boldsymbol{\lambda})) = \mathbb{E}_{\mathbf{y}_{\mathcal{M}} \sim f(\mathbf{y}_{\mathcal{M}})} [J(\mathbf{y}_{\mathcal{M}})] \quad (1c)$$

$$\mathbf{y}_{\mathcal{M}} = \mathcal{M}(\mathbf{u}, \mathbf{x}_{\mathcal{M}}, \boldsymbol{\lambda}) \quad (1d)$$

In Eq. 1b,  $\mathbf{u}^*$  is the optimal decision, selected from the bounded set of alternatives  $\mathbf{u} \in \mathbb{U}$ , where  $\mathbb{U} \subseteq \mathbb{R}^{N_u}$ , and  $N_u$  is the number of decision variables.

In Eq. 1c, d,  $J(\cdot)$  is the loss function,  $\mathcal{M}(\mathbf{u}, \mathbf{x}_{\mathcal{M}}, \boldsymbol{\lambda})$  is the system model. The loss function is defined here as a cost to be minimized. Alternatively, Expression (1c) can be written as a benefit to be maximized.  $\mathbf{y}_{\mathcal{M}}$  is the model output,  $\mathbf{x}_{\mathcal{M}}$  is the model input, and  $\boldsymbol{\lambda}$  the model parameters. Model  $\mathcal{M}(\cdot)$  produces an output  $\mathbf{y}_{\mathcal{M}}$  given inputs  $\mathbf{x}_{\mathcal{M}}$ ,  $\mathbf{u}$  and parameters  $\boldsymbol{\lambda}$ . Both model inputs and parameters can be uncertain. In this case inputs  $\mathbf{x}_{\mathcal{M}}$ , parameters  $\boldsymbol{\lambda}$  and outputs  $\mathbf{y}_{\mathcal{M}}$  are stochastic variables, such that  $\mathbf{x}_{\mathcal{M}} \sim f(\mathbf{x}_{\mathcal{M}})$ ,  $\boldsymbol{\lambda} \sim f(\boldsymbol{\lambda})$ , and  $\mathbf{y}_{\mathcal{M}} \sim f(\mathbf{y}_{\mathcal{M}})$ , where  $f(\cdot)$  is the probability density function (pdf). Output uncertainty can be estimated by integrating input and parameter uncertainty.  $\mathbb{E}(\cdot)$  is the average operator, and  $\mathcal{J}(\mathbf{u}, f(\boldsymbol{\lambda}))$  is the expected cost for actions  $\mathbf{u}$ , given the present state of knowledge  $f(\boldsymbol{\lambda})$ ;  $f(\boldsymbol{\lambda})$  contains the present state of information about  $\boldsymbol{\lambda}$  after all available information has been assimilated.

The problem defined in Eq. 1a–d provides a firm theoretical framework that has been used in water management applications for risk based decision making (van Overloop et al. 2008; Verkade and Werner 2011; Raso et al. 2014; Vogel 2017).

## 2.1 Optimal Design Problem

The OD problem, defined in Eq. 2a–c, identifies the next observation that has the highest expected marginal value.

$$\text{find } \chi^* = \arg \max [\mathcal{B}(\chi) - \mathcal{C}(\chi)] \quad (2a)$$

$$\text{in which} \quad (2b)$$

$$\mathcal{B}(\chi) - \mathcal{C}(\chi) = \mathbb{E}_{\mathbf{y}_\chi \sim f(\mathbf{y})} [B(\mathbf{y}_\chi) - C(\mathbf{y}_\chi)] \quad (2c)$$

In Problem (2a–c), metadata  $\chi$  and data  $\mathbf{y}_\chi$  make up the new observation.  $B$  and  $C$  are the conditional benefit and cost of data  $\mathbf{y}_\chi$ , and  $\mathcal{B}$  and  $\mathcal{C}$  are the average benefit and cost of getting information from  $\chi$ . Metadata  $\chi$  allows framing the new data within the available model to give meaning to it. Meta-data specifies where and when an observation has been realized, the observational uncertainty of the measuring instrument, etcetera. Metadata  $\chi$  must identify univocally the variable, defined on the bounded set  $\chi \in \mathcal{X}$ , to which data  $\mathbf{y}_\chi$  is related.  $\mathcal{X}$  is the space of observable new information. Data  $\mathbf{y}_\chi \in \mathbb{R}^{N_{\mathbf{y}_\chi}}$  is the numerical value registered after that the observation has been obtained;  $N_{\mathbf{y}_\chi}$  is the dimension of new data.

The conditional benefit  $B$  of data depends on the present available information  $f(\boldsymbol{\lambda})$  and the new information  $\mathbf{y}$ .  $B$  quantifies the benefit of new information conditioned to the new data, as described in Eq. 3.

$$B(\mathbf{y}) = \mathcal{J}(\mathbf{u}_0^*, f(\boldsymbol{\lambda}|\mathbf{y})) - \mathcal{J}(\mathbf{u}^*, f(\boldsymbol{\lambda}|\mathbf{y})) \quad (3)$$

In Eq. 3,  $f(\boldsymbol{\lambda}|\mathbf{y})$  is the present information after the new information  $\mathbf{y}$  has been taken into account,  $\mathbf{u}_0^*$  is the optimal action conditional to information before the obtaining of new data, such that  $\mathbf{u}_0^* = \arg \min \mathcal{J}(\mathbf{u}, f(\boldsymbol{\lambda}))$ ;  $\mathcal{J}(\mathbf{u}_0^*, f(\boldsymbol{\lambda}|\mathbf{y}))$  is the expected cost if action did not adapt, calculated a posteriori, once  $\mathbf{y}$  is given;  $\mathcal{J}(\mathbf{u}^*, f(\boldsymbol{\lambda}|\mathbf{y}))$  the costs after that new information has been assimilated and actions adapt consequently;  $\mathbf{u}^*$  is the optimal decision after new information has been assimilated. The benefit of information is proportional to the performance improvement, which ameliorates thanks to the better actions due to better information.

The cost  $\mathcal{C}(\chi)$  depends on how the new piece of information is collected, and includes only the marginal costs of getting the new information. Costs that do not depend directly on the specific decision about getting additional data are “sunk costs” (Menger 1981), and must be excluded. Costs can include “generalized” costs, such as the cost related to the time to wait before getting the data. Equation 4 defines the conditional cost.

$$C(\mathbf{y}) = C_D(\chi) + C_S(\mathbf{y}, f(\boldsymbol{\lambda})) \quad (4)$$

In Eq. 4, costs are decomposed in a deterministic  $C_D$  and stochastic  $C_S$ . The stochastic component depends on the data outcome. Additional information reduces uncertainty in the stochastic component of costs. According to the certainty equivalence principle (Philbrick and Kitanidis 1999; Van de Water and Willems 1981), any other randomness independent of  $\mathbf{y}$  or other problem components can be treated as deterministic, using its expected value only.

In Problem (2a–c), however, the variable  $\mathbf{y}$  is yet to be observed. Nonetheless, the pdf of  $\mathbf{y}$  can be estimated using the best present knowledge, as in Eq. 5.

$$f(\mathbf{y}) = \mathbb{E}_{\boldsymbol{\lambda}} [f(\mathbf{y}|\boldsymbol{\lambda})] = \int_{\Lambda} f(\mathbf{y}|\boldsymbol{\lambda}) \cdot f(\boldsymbol{\lambda}) \cdot d\boldsymbol{\lambda} \quad (5)$$

In Eq. 5,  $f(\mathbf{y})$  is the pdf of  $\mathbf{y}$ ;  $f(\mathbf{y})$  depends on both present information  $f(\boldsymbol{\lambda})$  and observational uncertainty  $p(\mathbf{y}|\boldsymbol{\lambda})$ . Equation 5 can then be used to estimate the first term in Eq. 2c.

When, on average,  $\mathcal{J}(\mathbf{u}_0^*, f(\boldsymbol{\lambda}|\mathbf{y})) < \mathcal{J}(\mathbf{u}^*, f(\boldsymbol{\lambda}))$ , new information is expected to improve the actions and get better results.  $\mathcal{B}$  is always non-negative (Parmigiani and Inoue 2009). However, if  $\mathbf{u}$  does not change for all possible values of  $\mathbf{y}$ , then both terms of the right side of Eq. 3 are equal, and  $\mathcal{B}=0$ . In such a case, additional information has no added benefit.

OD can also be used as stopping criterion: when the expected benefit is lower than the expected cost for all possible observations, the next piece of information is expected to be not worth its cost, hence a level of “rational ignorance” (Simon 1990) is reached: reducing uncertainty can still be reduced, but it is not effective from a cost-benefit point of view.

The dependency of  $f(\mathbf{y})$  on  $f(\boldsymbol{\lambda})$  implies that OD is an iterative problem. Every time a new piece of information is obtained, this is processed and assimilated within present information, and the uncertainty represented by  $f(\boldsymbol{\lambda})$  is reduced. Then, the problem in Eq. 2a–c is reformulated using the new state of information.

## 2.2 OD Using LSE

The OD Problem requires: 1)  $f(\mathbf{y}_\lambda|\boldsymbol{\lambda})$  for all observable variables  $\lambda$ , 2)  $f(\boldsymbol{\lambda})$ , 3)  $f(\mathbf{y})$ , and 4)  $f(\boldsymbol{\lambda}|\mathbf{y})$ . In the following we will show how to define these distributions if data are assimilated using Least Squares Estimation (LSE).

LSE can be used to identify the model parameters given the data and a deterministic model structure. LSE are extensively used in many fields (Kariya and Kurata 2004), including hydrology (Weijs et al. 2013). Equation 6a, b presents the rule for parameter estimation using LSE.

$$\min_{\boldsymbol{\lambda}} \sum_{i=1}^k \mathbf{e}_i^2 \quad (6a)$$

where

$$\mathbf{e}_i = \mathbf{y}_i - \mathcal{R}(\mathbf{x}_i, \boldsymbol{\lambda}) \quad (6b)$$

In Eq. 6a, b,  $k$  is the number of data points,  $\mathbf{e}_i \in \mathbb{R}^{N_y}$  are residuals, defined in Eq. 6b as the difference between the observed data and the results from relation  $\mathcal{R}(\cdot)$ ;  $\mathcal{R}$  is a sub-component of the whole system model  $\mathcal{M}$ ;  $\mathcal{R}(\cdot)$  is intended to predict  $\mathbf{y}$ ; in  $\mathcal{R}(\cdot)$ ,  $\mathbf{x}$  is the input and  $\boldsymbol{\lambda}$  the parameters. Uncertainty on effects of actions can be included considering actions  $\mathbf{u}$  as part of the inputs of model  $\mathcal{R}(\cdot)$ , such that  $\mathbf{u} \subseteq \mathbf{x}$ .

If relation  $\mathcal{R}$  is in the form as Eq. 7, then a LSE solution exists in closed form.

$$g_y(\mathbf{y}) = \sum_j^{N_\lambda} \lambda_j \cdot g_j(\mathbf{x}) \quad (7)$$

In Eq. 7,  $g_j(\cdot)$  and  $g_y(\cdot)$  are functions, also non-linear, of  $\mathbf{x}$  and  $\mathbf{y}$ .

In LSE, when residuals  $\mathbf{e}$  are normal, independent, and identically distributed, i.e.  $f(\mathbf{y}|\boldsymbol{\lambda}) \sim N(0, \boldsymbol{\Sigma})$ , then parameters  $\boldsymbol{\lambda}$  are also normal (Leon 1980), defined as in Eq. 8a–c.

$$f(\boldsymbol{\lambda}) \sim \mathcal{N}(\boldsymbol{\mu}_\lambda, \boldsymbol{\Sigma}_\lambda) \quad (8a)$$

$$\boldsymbol{\mu}_\lambda = (X^T X)^{-1} X^T Y \quad (8b)$$

$$\boldsymbol{\Sigma}_\lambda = \boldsymbol{\Sigma} (X^T X)^{-1} \quad (8c)$$

In Eqs. 8b, c,  $X = [\underline{g}(\mathbf{x}_1), \dots, \underline{g}(\mathbf{x}_k)]$  is a matrix of dimension  $k \times N_\lambda$ , made of  $k$  data-points of  $N_\lambda$ -dimension; regressors  $\underline{g}(\mathbf{x}_i) = [g_1(\mathbf{x}_i), \dots, g_{N_\lambda}(\mathbf{x}_i)]$ ;  $Y = [g_y(\mathbf{y}_1), \dots, g_y(\mathbf{y}_k)]$  is a matrix of dimensions  $k \times N_y$ , made of the observed outputs.  $\Sigma$  is estimated as in Eq. 9.

Under the same conditions on the residuals, LSE parameter estimation is equivalent to Maximum Likelihood Estimation (Charnes et al. 1976). Maximum Likelihood estimation is equivalent to using Bayes rule starting from a condition of no a-priori information and optimal learning from data (Jaynes and Bretthorst 2003)).  $f(\mathbf{y}|\lambda)$  is in fact the conjugate prior in the Bayes Equation for the conjugate distributions  $f(\lambda)$  and  $f(\lambda|\mathbf{y})$ , that are both normal. When  $\Sigma$  is a known parameter, then the conjugate prior is also normal (Diaconis et al. 1979). When observational uncertainty variance is estimated from data, then the conjugate prior is Inverse-Gamma distributed, or Inverse-Wishart distributed in the multidimensional case (Diaconis et al. 1979). Normal and Inverse-Gamma distributions converge when  $k$  is sufficiently large (Leemis and McQueston 2008).

$f(\mathbf{y}|\lambda)$  is the distribution of data after its observation. The uncertainty of this stochastic variable is considered irreducible. In  $f(\mathbf{y}|\lambda)$ , the average is zero and the covariance matrix is  $\Sigma$ . When the average is not zero, any knowledge about the presence of a bias can be used to correct its estimation and bring it to zero average (Sorooshian and Dracup 1980).  $\Sigma$  is a measure of the observational uncertainty: it can be either a given parameter, when, for example, the observational uncertainty is known from the characteristics of the observation technique, or estimated from data. When observational uncertainty is to be estimated from data, the parameter set is extended to include variance parameters, such that  $\Sigma \in \lambda$ . In LSE, covariance  $\Sigma$  can be estimated as in Eq. 9.

$$\Sigma = \frac{E^*}{k - N_\lambda} \tag{9}$$

In Eq. 9,  $E^*$  is the minimum of sum of squares errors, as defined in Eq. 6a. For models as in Eq. 7,  $E^* = (Y - X(X^T X)^{-1} X^T)^T (Y - X(X^T X)^{-1} X^T)$ , and  $k - N_\lambda$  is the degree of freedom.

The distribution of the next data-point can be derived from Eq. 5. Given observable  $\chi = \mathbf{x}$ , and related data  $\mathbf{y}_\chi$  Eq. 10a–c defines the distribution of next data-point.

$$f(\mathbf{y}_\chi|\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}_y, \Sigma_y) \tag{10a}$$

$$\boldsymbol{\mu}_y = \boldsymbol{\mu}_\lambda \mathbf{x} \tag{10b}$$

$$\Sigma_y = \mathbf{x} \Sigma_\lambda \mathbf{x}^T \tag{10c}$$

Once a new data-point is available, this is assimilated in the set of available data, such that  $X \leftarrow [X, \underline{g}(\mathbf{x}_{k+1})]$  and  $Y \leftarrow [Y, g_y(\mathbf{y}_{k+1})]$ , where  $[\mathbf{x}_{k+1}, \mathbf{y}_{k+1}]$  is the new data-point.

Once  $f(\mathbf{y}|\lambda)$  are defined for all observable variables  $\chi$ ,  $f(\lambda)$ , and  $f(\lambda|\mathbf{y})$  are defined using LSE properties, these can be used within the OD problem.

### 3 Application

The methodology outlined in the previous section is applied to a test case in the White Cart River in Scotland. This case has been selected as it offers a simple non-trivial case of a monitoring problem related to a well defined decision problem. Monitoring based on OD is compared to monitoring uniquely based on uncertainty reduction. In the latter, new data is

selected according to an information criterium, as in Alfonso et al. (2010) and Mogheir and Singh (2002).

In this river a flood warning scheme has been established. A hydrological model of the upstream catchment is used to predict the discharge at the gauging station at Overlee. The rating curve available at Overlee is used to transform the predicted discharge to a predicted water level, which is then compared to the flood warning threshold. If the threshold is exceeded, a warning may be issued and response to mitigate flood losses initiated. A more complete description of the system can be found in Werner et al. (2009) and in Verkade and Werner (2011). The flood warning scheme considered here is simplified for the purpose of this study, when compared to the actual operational scheme.

In predicting the levels, we consider two sources of uncertainty: i) discharge prediction and ii) rating curve. In this study, we focus only on the rating curve uncertainty and its reduction by new data. The parameters of the rating curve are estimated using the gauging data. This data includes a set of observed water level and discharge pairs, referred to as gaugings. These are collected through an in-situ discharge measurement by a team of hydrometrists. A complete rating curve estimation requires gaugings at a range of levels/discharges. The rating curve model used at Overlee is a power law equation, with the parameters of this power law established through regression using the available set of gaugings. The uncertainty in the rating curve affects the uncertainty of water level predictions, especially for flood situations (Tomkins 2014; Sikorska et al. 2013; Domeneghetti et al. 2012; Di Baldassarre and Montanari 2009; Pappenberger et al. 2004; Le Coz et al. 2014; McMillan et al. 2012). The uncertainty in the rating curve can be reduced by adding data points in the regression, which are obtained through extra gaugings. These gaugings are laborious and when carried out manually, such as at Overlee, labour intensive, particularly at higher flows, when gauging may even become too dangerous to be done. The timing of the new gauging must be selected to coincide with either low-flow conditions, that are low-cost but also less informative, or high flow, that are costly and more informative, particularly for flood events.

In the primary decision problem the decision maker must decide whether to issue a flood warning. In the secondary monitoring problem the decision maker must efficiently select the additional gaugings to be carried out, and at what flow conditions.

### 3.1 Decision Problem

Equations 11–12 define the optimal decision problem for this system.

$$\mathcal{J}^* = \min_{\{u_t\}_t^H} \sum_{t=1}^H \mathcal{J}_t \quad (11)$$

where

$$\begin{aligned} \mathcal{J}_t &= u_t \times L \times y_{F,t} \\ &= \begin{bmatrix} u_t \\ 1 - u_t \end{bmatrix}^T \times \begin{bmatrix} C_w + L_u & C_w \\ L_a + L_u & 0 \end{bmatrix} \times \begin{bmatrix} p_t \\ 1 - p_t \end{bmatrix} \\ &= \begin{bmatrix} u_t \\ 1 - u_t \end{bmatrix}^T \times \begin{bmatrix} p_t \cdot (C_w + L_u) + (1 - p_t) \cdot C_w \\ p_t \cdot (L_a + L_u) \end{bmatrix} \\ &= \begin{bmatrix} u_t \\ 1 - u_t \end{bmatrix}^T \times \begin{bmatrix} C_w \\ p_t \cdot L_a \end{bmatrix} \end{aligned} \quad (12)$$



In Eq. 11,  $\mathcal{J}_t$  is the time-step cost function;  $u_t \in \{0, 1\}$  is the daily decision whether a flood warning is to be issued. This decision must be taken at all time-steps  $t \in \{1, \dots, H\}$ , where  $H$  is the length of the problem horizon.

In Eq. 12,  $y_{F,t} \in \{0, 1\}$  is a stochastic variable defining the occurrence of a flood event at day  $t$ . A flood event occurs when water level exceeds the flood threshold, as defined in Expression (14).  $p_t$  represents the flood probability, i.e.  $P(y_{F,t} = 1) = p_t$ .  $L$  is a  $2 \times 2$  matrix representing the cost structure for all possible combinations of events and decisions. In matrix  $L$ ,  $C_w$  is the cost of warning and the cost of response to the warning,  $L_u$  is the unavoidable loss in case of flood, and  $L_a$  is the loss that can be avoided due to the response to a flood warning, while  $L_a + L_u$  is the total loss in case of a flood occurring without warning. Costs are all positive values, and such that  $L_a > C_w$ . Maintenance costs are not considered, being irrelevant in this decision problem.

The no-flood no-warning situation, in the lower-right quadrant of  $L$ , is the “business as usual” case. The flood warning situation, upper-left quadrant, is the “hit” case, where the warning is correctly issued before a flood event. The no-flood warning situation, upper-right quadrant, is the “false alarm” case, where the warning issues turned out to be unnecessary. The flood no-warning situation, lower-left quadrant, is the “missen case. The warning system failed to anticipate the event. Costs due to uncertainty are due to the occurrence of the last two cases. In this experiment,  $C_w = 1000$  \$,  $L_a = 4000$  \$, and  $L_u = 1000$  \$.

The action is selected according to Rule (13).

$$u_t = \begin{cases} 1 & \text{if } p_t > C_w/L_a \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

This rule is derived from Expression (12), simplifying its application: action  $u_t$  can be selected according to this simple rule rather than solving the correspondent optimization problem.

The flood event is defined by Condition (14).

$$y_{F,t} = \begin{cases} 1 & \text{if } h_t > \bar{h} \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

Condition (14) states that flood occurs when water level  $h_t$  exceeds a given threshold  $\bar{h}$ , or equivalently, when  $y_t > \bar{y}$ , where  $y_t = \log h_t$  and  $\bar{y} = \log \bar{h}$ . In the numerical experiment, the flood warning threshold is set at  $\bar{h} = 1.1$  m, corresponding to 1/2 year return period. This threshold is lower than the true warning threshold at Overlee, but it ensures a sufficient number of flood events in the experiment. The discharge uncertainty is included by adding an artificial noise to include the upstream rating curve uncertainty and the presence of an unknown lateral discharge. The forecasted discharge is considered log-normally distributed with sigma parameter equal to 0.05.

Water level can be estimated from river discharge using the rating curve. Equation 15 defines the rating curve, relating water level  $h$  and discharge  $q$ , with the parameters  $a$ ,  $b$  and  $c$  determined through fitting the curve to the available set of gauged discharge-level pairs.

$$q = a \cdot (h + b)^c \quad (15)$$

The rating curve in Eq. 15 is a deterministic power law. Rating curve estimations generally use multiple parameter sets for different river stages (Reitan and Petersen-Øverleir 2009; Petersen-Øverleir and Reitan 2005). Two parameters sets are used at Overlee, for  $h$  below or above 0.5. We are interested in floods, and therefore we will consider only the upper part of the rating curve.

Rating curve parameters identification can be made easier by transforming Eq. 15 in a linear regression. Equation 16 is obtained by log transform of Eq. 15.

$$x = \lambda_0 + \lambda_1 \cdot y \tag{16}$$

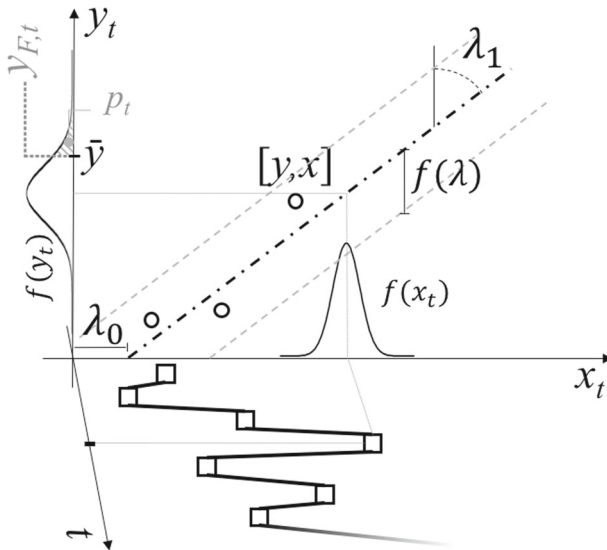
In Eq. 16,  $x = \log q$ ,  $y = \log(h + b)$ ,  $\lambda_0 = \log a$ , and  $\lambda_1 = c$ ;  $b$  has little influence on the high-flow part of the rating curve, therefore it will be considered known (or zero);  $\lambda = [\lambda_0, \lambda_1]$  are parameters of Relation (16).

The system model is made of i) Relation (15), ii) Condition (14), iii) Decision Rule (13), and iv) Loss matrix  $L$ . The causal chain of model  $\mathcal{M}$  is  $x_t \rightarrow y_t \rightarrow y_{F,t} \rightarrow p_t \rightarrow u_t^* \rightarrow \mathcal{J}_t^*$  (Fig. 1).

### 3.2 Optimal Design problem

The uncertainty of the rating curve, in Eq. 16, can be reduced by additional data. Setting up the OD Problem requires defining the space of possible observations, relative data, cost and observational uncertainty for all possible observations.

Each observation  $\chi$  is defined by the water level at which the gauging is realized,  $\chi = [y]$ . The relative data is the couple water level-discharge,  $[y, x]$ . The space of possible observations is  $\chi \in \mathcal{X} \subseteq \mathbb{R}^+$ . In this experiment, however, the space of possible observation is limited by the available data-points, made of 173 pairs. Observational uncertainty is  $f(y|\lambda) \sim N(0, \sigma_{RC})$ , where  $\sigma_{RC}$  is either a given parameter, derived from instrument characteristics, or estimated from data. We consider the case when  $\sigma_{RC}$  is estimated from data, as in Eq. 9. Observation costs  $C(\chi)$  are made up of the direct gauging costs and indirect



**Fig. 1** Test case model schema. At each timestep  $t$ , discharge  $\hat{x}_t$  (squares) is forecasted with uncertainty  $f(x)$ . Rating curves parameters  $\lambda = [\lambda_0, \lambda_1]$  are estimated from data-points  $[y, x]$  (circles). Rating curve (bold dash-dot line) is also uncertain  $f(\lambda)$  (grey dash lines). Water level uncertainty  $f(y)$  is the result of discharge and rating curve uncertainty. Flood event  $y_{F,t}$  happens if  $y_t > \bar{x}$  (grey dot line), with probability  $p_t$  (dashed area)

costs, or the “cost-to-wait”. The gauging cost has a fixed component, required for all gaugings, and a variable one, which increases with increasing flows. The cost-to-wait is related to lower frequency of higher water levels. Waiting for the occurrence of higher water levels has a cost, because, if the waiting time increases, costly flood events are more likely to happen. Eq. 17a–c defines the cost of gauging. Discharge dynamics and its uncertainty (Angulo et al. 2000) are not considered.

$$C(\mathbf{y}) = C_g(y) + C_w(x) \quad (17a)$$

$$= c_g + c_q \cdot \exp(x) \quad (17b)$$

$$+ \mathcal{J}^*/H \cdot \frac{1}{P(Y \geq y)} \quad (17c)$$

In Eq. 17a–c, Line (17b) defines the direct cost of gauging:  $c_g$  is the fixed cost component, and  $c_q$  the part proportional to discharge, such that  $C_g(q) = c_g + c_q \cdot q$ , written in Eq. 17b as function of  $x$ . In the experiment,  $c_g=500$  \$ and  $c_q=7$  \$/m<sup>3</sup>/s. Line (17c) defines the indirect cost of gauging, linked to the cost to wait. Cost-to-wait is proportional to the expected cost per day due to possible floods,  $\mathcal{J}^*/H$ , where  $\mathcal{J}^*$  is the expected cost for the entire horizon  $H$  as defined in Expression (11), multiplied by the waiting time in days,  $1/P(Y \geq y)$ . This implies that additional information reduces observation cost. Time is inversely proportional to the frequency of events larger than  $y$ ,  $P(Y \geq y)$ .  $P(Y \geq y) = 1 - P(Y \leq y) = 1 - F(Y)$ , where  $F(\cdot)$  is the cumulative density function of  $y$ , estimated here using the empirical distribution on available data  $y$ .

Table 1 shows the main variables and parameters of the decision and design problem.

### 3.3 Results

Starting from a rating curve with 3 data-points taken at low-flow conditions, new observations are selected by iteratively solving the OD problem. The OD problem is solved

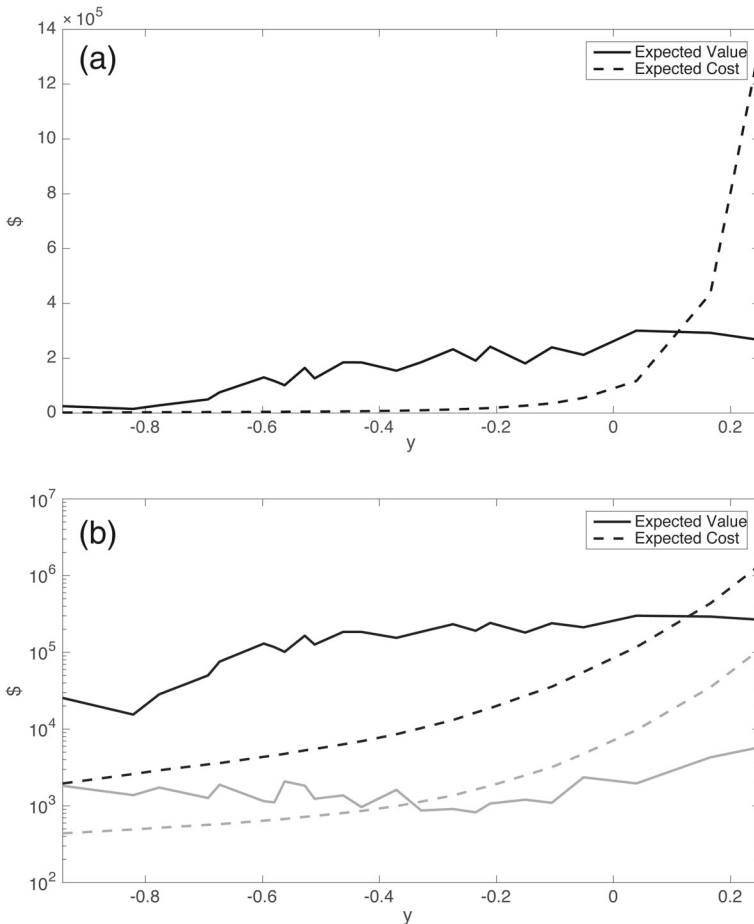
**Table 1** List of main variables and parameters

Description	Symbol	Type	Domain/Value
Discharge	$q_t$	Input	$\mathbb{R}^+$
Water level	$h_t$	Output	$\mathbb{R}^+$
Rating curve parameters	$a, b, c$	Parameters	$\mathbb{R}^3$
Discharge, log scale	$x_t$	Model input	$\mathbb{R}$
water level, log scale	$y_t$	Internal variable	$\mathbb{R}$
Rating curve param., log scale	$\lambda_0, \lambda_1$	Unknown parameters	$\mathbb{R}^2$
Flood event	$y_{F,t}$	Model output	[0, 1]
Issue warning	$u_t$	Action	[0, 1]
Warning costs	$C_w$	Parameter	1000 \$
Unavoidable costs	$L_u$	Parameter	1000 \$
Avoidable costs	$L_a$	Parameter	4000 \$
Gauging costs	$c_g, c_q$	Parameters	500 \$, 7 \$/m <sup>3</sup> /s
Gauged water level	$y$	Selected observation	$\mathbb{R}$
Gauged discharge	$x$	Observed data	$\mathbb{R}$

numerically using a sample average approximation. The samples dimensions are: 60 from  $f(\lambda)$ , 20 from  $f(x_t)$ ,  $\forall t$ , 30 from  $f(x|y)$ .

Figure 2, plot *a*, shows expected observation benefit and cost at different water levels, for the selection of the first additional gauging. Observations are more informative and more costly as the water level increases. Cost starts rising rapidly from about  $y > 0$ , exceeding the benefit of information in the proximity of  $y > 0.1$ , corresponding to water level  $h = 1.1$  m. The most valuable observation is where the difference between benefit and cost is the largest. In this case the OD will select gauging at  $y = -0.21$ , corresponding to water level  $h = 0.81$  m.

Figure 2, plot *b*, shows the effect of different initial states of knowledge on observation benefit and cost. The black lines refer to an initial condition where the three available gauging pairs are sampled at a low water level, all in the neighbourhood  $y = -0.52763$ . The grey lines refer to an initial condition where the three data-points are evenly distributed



**Fig. 2** Above: Plot *a*. Expected Benefit and Cost for observations at different water levels, at first gauging. Below: Plot *b*. Expected Benefit and Cost for observations at different water levels, at first gauging, with different initial information: low-flow data-points only (black lines) and evenly distributed data-points (grey lines)

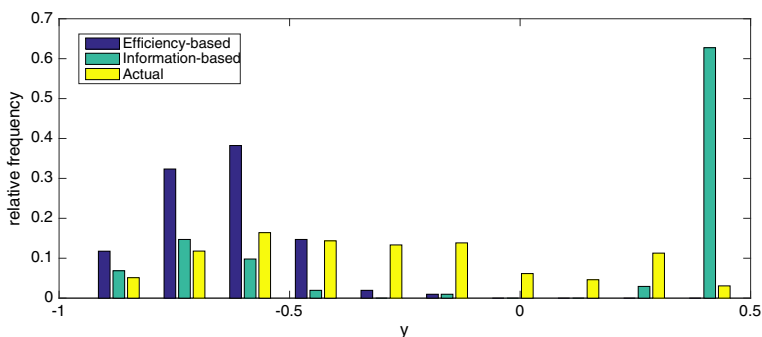
on the water level distribution, namely  $y = [-0.53, -0.49, -0.30]$ . The vertical scale is logarithmic.

In the case with evenly distributed data-points, the initial condition contains more information than in the case with low-flow data-points. Despite having the same number of data-points the distributed data allow a better estimation of the parameters. When starting from a more informative situation, the added value of one additional data-point is lower. Consequently, in Fig. 2, plot *b*, the continuous grey line is below the continuous black line for all water levels. More information allows a better prediction of flood events, reducing type I and type II errors, consequently reducing average cost. Observation cost depends on expected cost,  $\mathcal{J}^*(\cdot)$ , on which cost-to-wait depends. New information reduces the expected cost-to-wait, hence the cost of observation, on average. In Fig. 2*b*, in fact, the dashed grey line is below the dashed black line. Data-points taken at high water level will increase the relative benefit of observation at lower water levels. In the evenly distributed case the OD suggests an optimal trade-off at  $y = -0.56$ , corresponding to a water level  $h = 0.57$  m, lower than the optimal trade-off in the low-flow data case.

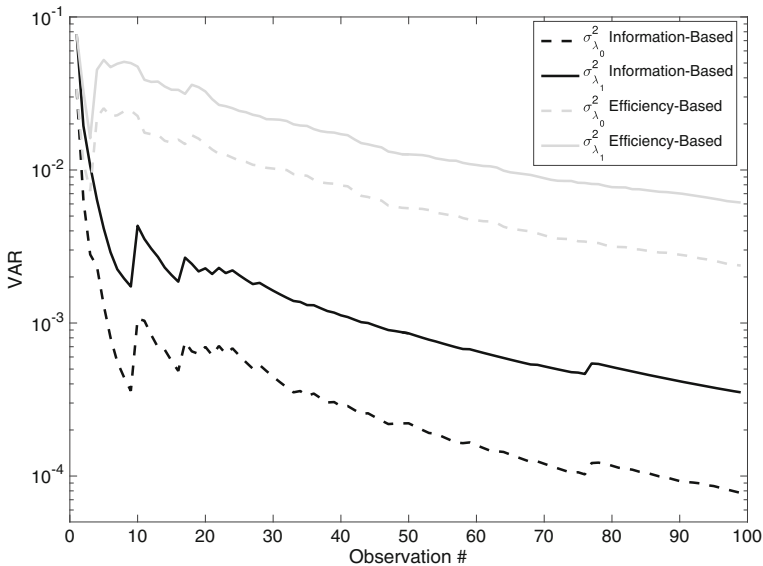
Monitoring based on OD is compared to monitoring uniquely based on uncertainty reduction. In the latter case, the criterium is selecting an observation that reduces the uncertainty of the water level estimation the most. Uncertainty is quantified by the total variance. i.e.  $\min \text{VAR}(y)$ , as defined in Appendix I, Eq. 20a, b.

Figure 3 shows the relative frequency of observations at different levels of  $y$  for the OD problem and the information-based. The relative frequency at which the actual gaugings were taken is also shown. The OD criterion selects mostly low-flow values, due to their relative low cost. The information-based criterion, on the other hand, distributes observations approximately equally between very high flow and low flow conditions, in order to reduce the degree of freedom of the rating curve relation. In fact, if data-points are to be selected considering an information criterium only, it is preferable to explore the entire relation domain and, in the linear case, its extremes. Actual observations are evenly distributed. In the latter case, other criteria to select at what discharge/level the gauging is taken, more than just data for flood protection, are likely to be present. Moreover, the present OD problem setting considers average observation cost, without taking into account current information about river stage dynamics. In case of high-flow, in fact, one can take immediate advantage of the present condition to obtain a highly informative data-point, without having to wait for it.

In the efficiency-based case, for the present cost structure, the average benefit per observation is about 270 \$. If compared to the problem costs, this implies that 4 gaugings make



**Fig. 3** Observations frequency at different water levels



**Fig. 4** Rating curve parameters variance in the information-based (black line) and the efficiency-based (grey line) cases

up the cost of a false alarm, and about 15 gaugings the extra-cost for failing in forecasting a flood.

In the information-based the average observation value is  $-13 \times 10^3$  \$. The high negative value means that the costs of observation is much larger than its benefit, showing the inconvenience of selecting additional data using an information-based criterion only. In practice, however, if observations are costly, there will be a balance between informativeness and efficiency. In this case the weights to be assigned to information and cost are subjective. In the efficiency based approach this duality is solved by evaluating the benefit of information, which can be compared to the cost to obtain it.

The information based approach, despite its inefficiency, is highly effective in reducing uncertainty. Figure 4 represents how rating curve uncertainty is reduced as a function of the number of gauging data pairs when data is selected using efficiency and information based criteria, taking the variance of the parameters  $\lambda_0$  and  $\lambda_1$  as indicator of uncertainty.

Figure 4 shows the evolution of rating curve parameters variance with observations. In the long run, uncertainty is reduced in both cases, but in the information-based case the variance is smaller, resulting in better defined parameters.

## 4 Conclusions

In this paper we propose and test an innovative methodology to make deterministic hydro-economic models calibrated with Least Squares Estimation (LSE) usable for solution of the Optimal Design (OD) problem.

The application of the OD problem has been limited by its large computational burden and the need to use stochastic models, whereas most of existing hydro-economic models

are deterministic. The solution we present uses some of the properties of LSE to handle deterministic models and get a leaner approximation of the OD problem. Thanks to this innovation, LSE can be used to identify relation parameters given the data and a deterministic model structure. Under some mild assumptions of the distribution of residuals, the parameter distribution family can be derived and used in the OD problem, and the OD problem boils down to a parametric estimation. Using this approach, the OD can be solved in a shorter time and, potentially, applied to larger systems.

Despite the proven effectiveness of the methodology, OD remains a computationally intensive method: scaling it up to large hydro-economic models is not explored here, and requires further research. Moreover, the methodology demands some assumptions on residual distribution to be valid. The test case is well characterised, with residual sufficiently “well distributed. The validity of the method when these assumptions on residuals distribution do not hold is to be explored.

Solving the OD problems indicates the piece of information that has the higher net benefit, i.e. the difference between the added value of that piece of information in terms of decision improvements, and the cost to obtain it. We presented the procedure to set a OD problem. Required elements of an OD problem are: i) a defined decision problem, ii) the space of possible observations, and, for all possible observations: ii-a) data uncertainty and ii-b) cost structure.

The proposed methodology is applied to a simplified although non-trivial decision problem under uncertainty: the flood warning system in the White Cart River, Scotland. The decision-maker must decide whether to issue a flood warning. River discharge forecast and the rating curve are uncertain. The rating curve parameters estimation can be improved by new gaugings, which can be selected on a scale between low-flow observations, that are low-cost but also less informative, and high flow ones, that are costly and more informative. The OD finds a balance between these two suggesting the most efficient data collection. Results from OD problem using an efficiency based approach is compared to results from an information based approach, where the criterium is reducing uncertainty as such: the two criteria lead to completely different results: the information based criterium leads to effective, but very costly, data-points selection; an economic criteria leads to economic efficiency. We suggest including an economic evaluation in data selection, in order to maximize monitoring efficiency.

The methodology can be applied to a large set of monitoring problems related to a well defined decision problem. The monitoring problem regards the decision on which piece of information to get, when getting that information has some cost: the purchase of remote sensing data, the optimal number and position of rain-gauges, the trajectory of drones are a few examples of monitoring problems. Example of defined decision problems are, at the operational level, a flood warning systems, as seen in the test case, or reservoir operation with defined objectives, i.e. electricity production. At the planning level, some examples of possible applications are: levee dimensioning, reservoir location, or even non-structural policies, such as new rules, the entity of incentives or Pigouvian tax.

The methodology presented here can open up new potential application for efficient monitoring. Efficient monitoring, guided by the value of information, is closely related to the needs of the decision maker. Value of information could be used as a means to inform the decision maker on priorities, and, being closer to the cognitive framework of the decision maker, it is probably more convincing than just uncertainty reduction.

To become fully mature, the methodology presented must prove its applicability to larger systems, and future research will tackle the computational issues related to more complex applications.

**Acknowledgements** We would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper. We are also grateful to the AXA Research Fund ([www.axa-research.org](http://www.axa-research.org)).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix I: Output Variance

Equation 18 is the variance for the output variable, sum of  $\lambda_i x_i$ .

$$VAR(y) = VAR\left(\sum_{i=1}^{N_\lambda} \lambda_i x_i\right) = \sum_{i=1}^{N_\lambda} \sum_{j=1}^{N_\lambda} COV(\lambda_i x_i, \lambda_j x_j) \quad (18)$$

Equation 19a–d defines the covariance between products  $\lambda_i x_i$  in function of multipliers average, variance, and covariance.

$$COV(\lambda_i x_i, \lambda_j x_j) = \mathbb{E}[\lambda_i x_i \cdot \lambda_j x_j] - \mathbb{E}[\lambda_i x_i] \mathbb{E}[\lambda_j x_j] \quad (19a)$$

$$= \mathbb{E}[\lambda_i \cdot \lambda_j] \cdot \mathbb{E}[x_i \cdot x_j] - \mathbb{E}[\lambda_i x_i] \mathbb{E}[\lambda_j x_j] \quad (19b)$$

$$= (\sigma_{\lambda_i, \lambda_j} + \mu_{\lambda_i} \cdot \mu_{\lambda_j}) \cdot (\sigma_{x_i x_j} + \mu_{x_i} \cdot \mu_{x_j}) - \mu_{\lambda_i} \cdot \mu_{x_i} \cdot \mu_{\lambda_j} \cdot \mu_{x_j} \quad (19c)$$

$$= \sigma_{\lambda_i \lambda_j} \cdot \sigma_{\lambda_i \lambda_j} + \sigma_{\lambda_i \lambda_j} \cdot \mu_{x_i} \mu_{x_j} + \sigma_{x_i x_j} \cdot \mu_{\lambda_i} \mu_{\lambda_j} \quad (19d)$$

In Eq. 19a–d,  $\mu$  is the average,  $\sigma^2$  the variance and  $\sigma$  the covariance of generic variables written in the subscript. Line (19b) assumes  $\lambda$  independent on  $x$ .

In the rating curve regression, Eq. 16,  $\sigma_{x_0}^2 = 0$ ,  $\sigma_{x_0, x_1} = 0$ , and  $\mu_{x_0} = 1$ . Equation 20a, b defines the output variance in the test-case.

$$VAR(y) = VAR(\lambda_0) + VAR(\lambda_1 x) + 2 \cdot COV(\lambda_0, \lambda_1 x) \quad (20a)$$

$$= \sigma_{\lambda_0}^2 + \sigma_{\lambda_1}^2 \sigma_{x_1}^2 + \sigma_{\lambda_1}^2 \mu_{x_1}^2 + \mu_{\lambda_1}^2 \sigma_{x_1}^2 + 2\sigma_{\lambda_0 \lambda_1} \mu_{x_1} \quad (20b)$$

## Appendix II: List of Main Symbols

Decision Problem

<b>u</b>	Decisions
<b>u*</b>	Optimal Decisions
<b>x</b>	Inputs
<b>y</b>	Outputs
<b><math>\lambda</math></b>	Parameters
<b><math>f(\cdot)</math></b>	Probability density function (p.d.f.)
<b><math>\mathcal{M}</math></b>	Model
<b><math>\mathcal{J}</math></b>	Objective
<b><math>\mathcal{J}^*</math></b>	Optimal value of objective function



## Design Problem

$\chi$	New data, metadata
$y_\chi$	New data, data
$\mathcal{B}(\chi)$	Benefit of new data $\chi$
$B(y_\chi)$	Conditional Benefit of new data ( $y_\chi$ )
$\mathcal{C}(\chi)$	Cost of new data $\chi$
$C(y_\chi)$	Conditional cost of new data ( $y_\chi$ )

## Least Squares Estimation

$\mathbf{e}$	Model error
$g_i, g_y$	Transformation on model input and output
$\boldsymbol{\mu}_\lambda$	Parameters expected value
$\Sigma_\lambda$	Parameters covariance matrix

## Appendix III: OD Procedure

The following pseudo-algorithm summarizes the procedure used in this paper.

- 1** Set Decision Problem: Define  $\mathbf{u} \in \mathbb{U}$ ,  $J$ ,  $\mathcal{M}$  and  $f(\boldsymbol{\lambda})$ , as in Problem (1)
- 2** Set Optimal Design Problem: Define  $\chi \in \mathcal{X}$ , and,  $\forall \chi$ , define  $f(y_\chi|\boldsymbol{\lambda})$  and  $C$ , as in Problem (2)
  - i** Solve Problem (1), find  $\mathbf{u}^*$
  - ii** Solve Problem (2), find  $\chi^*$
  - iii** If  $\mathcal{B}(\chi^*) < \mathcal{C}(\chi^*)$ , then take action  $\mathbf{u}^*$  and go to **3**, else:
  - iv** Select  $\chi^*$  and obtain  $\mathbf{y}$
  - v** Update  $f(\boldsymbol{\lambda})$  by  $\mathbf{y}$  by data assimilation, such that  $f(\boldsymbol{\lambda}) \leftarrow f(\boldsymbol{\lambda}|\mathbf{y})$
  - vi** Repeat from **i**
- 3** Stop

The OD procedure starts from a given primary Decision Problem (1) made of i) system model and objectives,  $\mathcal{M}$  and  $J$ , ii) set of possible actions  $\mathbf{u} \in \mathbb{U}$ , and iii) initial level of information  $f(\boldsymbol{\lambda})$ , as in Step 1. Then the Optimal Design Problem (2) is defined by: i) the complete set of observable variables  $\mathcal{X}$ , characterised by the level of uncertainty left after their observation  $f(y_\chi|\boldsymbol{\lambda})$ , and their cost structure  $C(\cdot)$ , as in Step 2. Problem (2) is solved and the most valuable information is selected, as in Step ii. The expected value of most valuable information is evaluated (Step iii). If the information is not worth its cost, then the procedure stops and the action is taken. Otherwise, the observation is realized (Step iv), the new data assimilated within existing information (Step v), repeating the procedure from Step i until satisfaction of condition at Step iii.

Solving OD Problem is selecting the observation with the highest net value. The OD Problem at Step ii must be solved numerically. For each observation, expected benefits and costs, which are the average of conditional benefit, as defined in Eq. 3 and cost, as defined in Eq. 4, are evaluated for all possible  $\mathbf{y}$ .

## Appendix IV: OD Numerical Solution

The following pseudo-algorithm shows a Montecarlo procedure to solve OD Problem numerically.

**A**  $f(\lambda)$  and  $\mathcal{X}$  are given. For all  $\chi_j \in \mathcal{X}$ :

**1** Sample  $n$  random  $\mathbf{y}_i$  from  $f(\mathbf{y})$ . For all  $i = 1, \dots, n$ :

**i** Evaluate  $f(\lambda|\mathbf{y}_i)$  as by data assimilation

**ii** Evaluate  $B(\mathbf{y}_i)$  as in Eq. 3

**iii** Evaluate  $C(\mathbf{y}_i)$  as in Eq. 4

**2** Evaluate  $\mathcal{B}(\chi_j) = \frac{1}{n} \sum_{i=1}^n B(\mathbf{y}_i)$

**3** Evaluate  $\mathcal{C}(\chi_j) = \frac{1}{n} \sum_{i=1}^n C(\mathbf{y}_i)$

**B** Find  $\chi^* = \arg \max[\mathcal{B}(\chi_j) - \mathcal{C}(\chi_j)]$

At Step **1**, select  $n$  randomly selected  $\mathbf{y}_i$  such that  $\mathbf{y} \sim f(\mathbf{y}) = f(\mathbf{y}|\lambda) \cdot f(\lambda)$ , as in Eq. 5. This can be obtained numerically by a simple Montecarlo method: sample  $n$  random  $\lambda_i \sim f(\lambda)$  and  $n$  random  $\mathbf{y}_i \sim f(\mathbf{y}|\lambda_i)$ . Step **i** assimilate  $\mathbf{y}_i$ , **ii** and **iii** evaluate conditional benefit and cost, for all  $\mathbf{y}_i$ . Steps **2–3** evaluate expected benefit and cost, average of conditional values, i.e.  $\mathcal{B}(x) = \mathbb{E}[B(\mathbf{y})]$ , and  $\mathcal{C}(x) = \mathbb{E}[C(\mathbf{y})]$ . Step **B** selects the  $\chi$  with the highest difference between expected benefit and cost.

This procedure presents different critical issues that must be carefully handled. At Steps **2–3**, expected benefit and cost are evaluated by average on a sample of size  $n$ , selected at Step **1**. This introduces a numerical approximation that depends on the sampling size  $n$ , but also on the problem structure and the problem dimensionality. When the cost function is related to a reliability problem, i.e. avoiding rare failure situations, the average of a purely random sampling may not approximate sufficiently its expected value. In this case importance sampling may be a more appropriate choice (?) to sample from  $f(\mathbf{y})$ .

At Step **ii**, Eq. 3 is evaluated  $n$ -times. This may require a significant computational burden which can be limiting in practical applications. The second term of Eq. 3,  $\mathcal{J}^*(f(\lambda|\mathbf{y}_i))$ , contains an optimization problem, the solution of which may be time-consuming. There are ways to reduce the computational time. For example, results for different  $\mathbf{y}$  are independent from each other, so Steps **i–iii** can be easily parallelized. Otherwise, relations among observable variables  $\chi$  can be used to avoid an exhaustive evaluation over all their domain  $\mathcal{X}$ .

At Step **i**, assimilating the new data changes the pdf of  $\lambda$ . This is a functional evaluation, i.e. the new data change the pdf for all  $\lambda$ . To be solved numerically, the entire  $\lambda$ -space must be discretized, and the Bayes rule applied to all interpolation points. Alternatively,  $f(\lambda)$  can be parametrised by selecting a distribution family, and the Bayes update will affect the parameters of the given distribution.

## Appendix V: Discharge Forecast Uncertainty

Discharge is observed upstream of Overlee at a station far enough to offer sufficient anticipation. In this experiment, however, the observed discharge at Overlee is considered as observed upstream. The discharge uncertainty is included by adding an artificial noise

to include the upstream rating curve uncertainty and the presence of an unknown lateral discharge. Discharge data, available from 01/01/1982 to 29/05/2011, at 15 minutes frequency, is aggregated at the daily time step. At each time instant, i.e. once a day, average daily discharge  $\hat{q}_{up,t}$  is observed. Upstream discharge is  $q_{up,t} = \hat{q}_{up,t} \cdot \xi_t$ , where  $\xi_t \sim LN(0, \sigma_{up})$ , i.e. lognormal distributed. Lateral discharge is proportional to  $q_{up,t}$ , such that  $q_{lat,t} = \kappa \cdot q_{up,t} \cdot \eta_t$ , where  $\eta_t \sim LN(0, \sigma_l)$ . Discharge at Overlee  $q_t$ , is  $q_t = q_{up,t} + q_{lat,t}$  which, after a log transform and for small lateral contribution, i.e. for  $\kappa \cdot \eta$  sufficiently small, one can obtain Eq. 21.

$$x_t \approx \hat{x}_{up,t} + \varepsilon_t + \kappa \cdot \delta_t \quad (21)$$

In Eq. 21,  $\hat{x}_{up,t} = \log \hat{q}_{up,t}$ ,  $\varepsilon_t \sim N(0, \sigma_{up})$ , and  $\delta_t \sim N(0, \sigma_l)$ . In the experiment,  $\sigma_{up}^2 + \sigma_l^2 = 0.05$ .

## References

- Alfonso L, Lobbrecht A, Price R (2010) Information theory based approach for location of monitoring water level gauges in polders. *Water Resour Res* 46:W03528. <https://doi.org/10.1029/2009WR008101>
- Alfonso L, Mukolwe MM, Di Baldassarre G (2016) Probabilistic flood maps to support decision-making: mapping the value of information. *Water Resour Res* 52(2):1026–1043
- Angulo J, Bueso M, Alonso F (2000) A study on sampling design for optimal prediction of space–time stochastic processes. *Stoch Env Res Risk A* 14(6):412–427
- Ben-Zvi M, Berkowitz B, Kesler S (1988) Pre-posterior analysis as a tool for data evaluation: application to aquifer contamination. *Water Resour Manag* 2(1):11–20
- Bhattacharjya D, Eidsvik J, Mukerji T (2010) The value of information in spatial decision making. *Math Geosci* 42(2):141–163
- Bierkens M (2006) Designing a monitoring network for detecting groundwater pollution with stochastic simulation and a cost model. *Stoch Env Res Risk A* 20(5):335–351
- Bode F, Nowak W, Loschko M (2016) Optimization for early-warning monitoring networks in well catchments should be multi-objective, risk-prioritized and robust against uncertainty. *Transp Porous Media* 114(2):261–281
- Chadalavada S, Datta B (2008) Dynamic optimal monitoring network design for transient transport of pollutants in groundwater aquifers. *Water Resour Manag* 22(6):651–670
- Charnes A, Frome E, Yu P-L (1976) The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. *J Am Stat Assoc* 71(353):169–171
- Davis DR (1971) Decision making under uncertainty in systems hydrology. The University of Arizona
- DeGroot MH (1962) Uncertainty, information, and sequential experiments. *Ann Math Stat* 33(2):404–419
- Di Baldassarre G, Montanari A (2009) Uncertainty in river discharge observations: a quantitative analysis. *Hydrol Earth Syst Sci* 13(6):913–921
- Diaconis P, Ylvisaker D et al. (1979) Conjugate priors for exponential families. *The Annals of Statistics* 7(2):269–281
- Domeneghetti A, Castellarin A, Brath A (2012) Assessing rating-curve uncertainty and its effects on hydraulic model calibration. *Hydrol Earth Syst Sci* 16(4):1191–1202
- Eidsvik J, Ellefmo SL (2013) The value of information in mineral exploration within a multi-gaussian framework. *Math Geosci* 45(7):777–798
- James BR, Gorelick SM (1994) When enough is enough: the worth of monitoring data in aquifer remediation design. *Water Resour Res* 30(12):3499–3513
- Jaynes ET, Bretthorst GL (2003) Probability theory: the logic of science. Cambridge University Press
- Kariya T, Kurata H (2004) Generalized least squares. Wiley
- Kim K-H, Lee K-K (2007) Optimization of groundwater-monitoring networks for identification of the distribution of a contaminant plume. *Stoch Env Res Risk A* 21(6):785–794
- Le Coz J, Renard B, Bonnifait L, Branger F, Le Boursicaud R (2014) Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: a Bayesian approach. *J Hydrol* 509:573–587
- Leemis LM, McQueston JT (2008) Univariate distribution relationships. *Am Stat* 62(1):45–53
- Leon SJ (1980) Linear algebra with applications. Macmillan, New York

- Maddock T (1973) Management model as a tool for studying the worth of data. *Water Resour Res* 9(2):270–280
- Mantoglou A, Kourakos G (2007) Optimal groundwater remediation under uncertainty using multi-objective optimization. *Water Resour Manag* 21(5):835–847
- McMillan H, Krueger T, Freer J (2012) Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. *Hydrol Process* 26(26):4078–4111
- Menger C (1981) Principles of economics. Ludwig von Mises Institute
- Mogheir Y, Singh V (2002) Application of information theory to groundwater quality monitoring networks. *Water Resour Manag* 16(1):37–49
- Nearing GS, Gupta HV (2015) The quantity and quality of information in hydrologic models. *Water Resour Res* 51(1):524–538
- Neumann J. v., Morgenstern O (1947) Theory of games and economic behavior. Princeton University, Princeton
- Nowak W, Guthke A (2016) Entropy-based experimental design for optimal model discrimination in the geosciences. *Entropy* 18(11):409
- Nowak W, Rubin Y, de Barros FPJ (2012) A hypothesis-driven approach to optimize field campaigns. *Water Resour Res* 48:W06509. <https://doi.org/10.1029/2011WR011016>
- Pappenberger F, Matgen P, Beven KJ, Henry J, Pfister L, de Fraipont P (2004) The influence of rating curve uncertainty on flood inundation predictions. *Flood Risk Assessment*, Bath
- Parmigiani G, Inoue L (2009) Decision theory: principles and approaches, vol 812. Wiley
- Petersen-Øverleir A, Reitan T (2005) Objective segmentation in compound rating curves. *J Hydrol* 311(1):188–201
- Philbrick CR, Kitanidis PK (1999) Limitations of deterministic optimization applied to reservoir operations. *J Water Resour Plan Manag* 125(3):135–142
- Raiffa H (1974) Applied statistical decision theory. Division of Research, Graduate School of Business Administration, Harvard University
- Raso L, Schwanenberg D, van de Giesen NC, van Overloop PJ (2014) Short-term optimal operation of water systems using ensemble forecasts. *Adv Water Resour* 71:200–208
- Reed P, Kollat J (2012) Save now, pay later? multi-period many-objective groundwater monitoring design given systematic model errors and uncertainty. *Adv Water Resour* 35:55–68
- Reitan T, Petersen-Øverleir A (2009) Bayesian methods for estimating multi-segment discharge rating curves. *Stoch Env Res Risk A* 23(5):627–642
- Sikorska A, Scheidegger A, Banasik K, Rieckermann J (2013) Considering rating curve uncertainty in water level predictions. *Hydrol Earth Syst Sci* 17(11):4415–4427
- Simon H (1990) Reason in human affairs. Stanford University Press
- Sorooshian S, Dracup JA (1980) Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases. *Water Resour Res* 16(2):430–442
- Tomkins KM (2014) Uncertainty in streamflow rating curves: methods, controls and consequences. *Hydrol Process* 28(3):464–481
- Trainor-Guitton WJ, Hoversten GM, Ramirez A, Roberts J, Juliusson E, Key K, Mellors R (2014) The value of spatial information for determining well placement: a geothermal example. *Geophysics* 79(5):W27–W41
- Van de Water H, Willems J (1981) The certainty equivalence property in stochastic control theory. *IEEE Trans Autom Control* 26(5):1080–1087
- van Overloop P-J, Weijs S, Dijkstra S (2008) Multiple model predictive control on a drainage canal system. *Control Eng Pract* 16(5):531–540
- Verkade JS, Werner MGF (2011) Estimating the benefits of single value and probability forecasting for flood warning. *Hydrol Earth Syst Sci Discuss* 15:3751–3765. <https://doi.org/10.5194/hess-15-3751-2011>
- Vogel RM (2017) Stochastic watershed models for hydrologic risk management. *Water Security* 1:28–35. <https://doi.org/10.1016/j.wasec.2017.06.001>. ISSN 2468-3124
- Weijs S, Schoups G, Van De Giesen N (2010) Why hydrological predictions should be evaluated using information theory. *Hydrol Earth Syst Sci* 14(EPFL-ARTICLE-167375):2545–2558
- Weijs SV (2011) Information theory for risk-based water system operation. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands
- Weijs SV, Mutzner R, Parlange MB (2013) Could electrical conductivity replace water level in rating curves for alpine streams? *Water Resour Res* 49(1):343–351
- Werner M, Cranston M, Harrison T, Whitfield D, Schellekens J (2009) Recent developments in operational flood forecasting in England, Wales and Scotland. *Meteorol Appl* 16(1):13–22
- Yokota F, Thompson KM (2004) Value of information literature analysis: a review of applications in health risk management. *Med Dec Making* 24(3):287–298