VALUE OF INFORMATION ANALYSIS ACCOUNTING FOR SENSOR DATA QUALITY: FOCUS ON DRIFT

PIER FRANCESCO GIORDANO* AND MARIA PINA LIMONGELLI †

*Department of Architecture, Built environment and Construction engineering (DABC)
Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milan, Italy
e-mail: pierfrancesco.giordano@polimi.it

†Department of Architecture, Built environment and Construction engineering (DABC)
Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milan, Italy; Lisa Meitner Guest professor at Lund University Faculty of Engineering, Division of Structural Engineering. John Ericssons väg 1. SE-221 00 Lund, Sweden
e-mail: mariagiuseppina.limongelli@polimi.it

Key words: Structural Health Monitoring, Value of Information, Pre-Posterior analysis, Bayesian decision making, Data quality, Sensor fault, Drift.

Abstract. Structural health monitoring plays a crucial role in assessing the condition of civil structures, providing information for regular maintenance and post-disaster emergency management. However, the reliability of structural health monitoring outcomes can be compromised by sensor malfunctions. Over the past two decades, sensor validation tools have been proposed to identify and discard abnormal measurements before extracting information from the structural health monitoring system. The long-term benefits of structural health monitoring systems are commonly evaluated without considering the possibility of faulty sensors. This can lead to suboptimal maintenance decisions. Recently, a Bayesian decision theory-based framework has been introduced to account for different data quality issues and quantify the benefit of implementing a sensor validation tool. This novel approach expands the traditional Value of Information concept to encompass multiple "functioning" states of the structural health monitoring system. This paper mainly focused on a specific data quality issue, i.e., bias or drift in the monitoring outcome. Previous applications of this framework regard simplified decision scenarios, where the monitoring system was either “damaged” or “undamaged”, considering a fixed drift value. In this paper, the impact of uncertain drift levels on the Value of Information in structural health monitoring is investigated, addressing real-world complexities. A numerical case study is considered to illustrate the practical implications of the VoI framework.

1 INTRODUCTION

Structural Health Monitoring (SHM) has made significant advances in recent years, emerging as a critical tool for assessing the integrity and performance of civil structures, aerospace vehicles, and industrial machinery [1]. SHM systems are designed to track the evolution of structural parameters over time, offering timely evaluations of health conditions and security warnings. These systems rely on the concept of "damage-sensitive features," parameters that encapsulate the structural state and exhibit variations indicative of damage progression [2]. Accurate damage identification, localization, and quantification are pivotal for informed decision-making in structural maintenance, repair prioritization, and resource allocation [3].

However, despite the potential of SHM, the reliability of its outcomes can be compromised by various factors, including sensor data quality issues. Environmental factors, repeated loads, and exceptional events contribute to structural deterioration over time. These very factors can also undermine the performance of SHM systems, particularly when operating in challenging environments. Sensor apparatus exposed to such conditions may experience faults, leading to anomalies in collected data and impacting data quality. Inaccurate or missing data can result to sub-optimal decisions entailing economic losses, generating false alarms and missed detections, potentially leading to costly operational interruptions or catastrophic accidents [4].

Sensor data quality issues in vibration-based SHM systems are diverse, encompassing various types of sensor faults, ranging from drift to noise interference. These faults can be categorized as "soft" or "hard," depending on their impact on the usability of the collected data [5].

To address these challenges and bolster the efficacy of SHM systems, researchers have turned to Sensor Validation Tools (SVTs), designed to assess the quality of collected data. SVTs employ various techniques, including one-class classifiers, multivariate statistical analysis, and machine learning algorithms, to detect and isolate inconsistent data channels. While SVTs offer valuable insights into data quality, their outcomes are inherently imperfect [6].

The concept of the Value of Information (VoI) from Bayesian decision theory has played a crucial role in quantifying the long-term economic benefits of SHM systems [7]. Traditionally, VoI analyses have focussed on the expected reduction in management costs associated with acquiring new information through an SHM system [8]. However, recent research has begun to explore the impact of data quality issues on VoI [9].

This paper investigates a novel approach proposed by Giordano et al. [10] that extends the VoI framework to encompass the assessment of sensor data quality, with a particular emphasis on mitigating challenges related to drift. In prior applications of this framework, the SHM system was characterized by binary states: either “damaged” or “undamaged” [11]. Notably, in cases involving drift, the “damaged” classification hinged on a predefined drift value. This paper delves into a scenario that reflects real-world complexity, where multiple states of the SHM system coexist. Specifically, the precise value of drift, when present, remains uncertain and is not predetermined.

In the subsequent sections, the paper delves into the methodology, presents the case study, and discusses the obtained findings.

2 THEORETICAL FRAMEWORK

The theoretical framework of this paper builds upon classical Bayesian decision analysis, offering a systematic approach for making optimal decisions in situations where the true state of a system is uncertain [12]. Rooted in utility theory and the Bayesian definition of probability, this framework empowers decision-makers to select actions that maximize utility, considering various levels of knowledge and belief [13].

Three distinct types of Bayesian decision analysis are considered, each adapted to different levels of information.

The Prior Analysis relies solely on the decision maker's knowledge, without collecting additional data. For each action A_n, the decision-maker computes the expected cost $E[u(A_n)]$ by weighing the prior probabilities $P(s_l)$ of different system states s_l against the utility values associated with different combinations of actions and states $u(A_n, s_l)$:

$$E[u(A_n)] = \sum_{l=1}^{L} u(A_n, s_l)P(s_l) \tag{1}$$

The optimal action $\hat{A}$ is chosen to maximize utility u_1:

$$\hat{A} = \arg\max_n E[u(A_n)] \tag{2}$$

$$u_1 = E\left[u\left(\hat{A}\right)\right] = \sum_{l=1}^{L} u\left(\hat{A}, s_l\right)P(s_l) \tag{3}$$

The Posterior and Pre-Posterior Analyses incorporate new information. Posterior analysis is conducted when the outcome o_j becomes available, while Pre-Posterior analysis anticipates all possible outcomes before they are observed. In this paper, the outcomes originate from an SHM system. In both cases, the prior probabilities of system states are updated using Bayes' theorem:

$$P\left(s_l \middle| o_j\right) = \frac{P\left(o_j \middle| s_l\right)P(s_l)}{P\left(o_j\right)} \tag{4}$$

where $P\left(o_j | s_l\right)$ is the probability of observing outcome o_j when the system is in state s_l, and $P\left(o_j\right)$ is the total probability of outcome o_j.

The framework proposed in [10] extends the traditional Bayesian approach by considering the states of the SHM system and outcomes of a Sensor Validation Tools (SVTs). Specifically, the states of the SHM system are represented by a random variable m_k, which can assume K different values (e.g., $m_1 = damaged$ and $m_2 = undamaged$). The joint probability distribution of system states s_l and SHM system states m_k, denoted as $P(s_l, m_k)$, is updated using Bayes' theorem:

$$P\left(s_l, m_k \middle| o_j\right) = \frac{P\left(o_j \middle| s_l, m_k\right)P(s_l, m_k)}{P\left(o_j\right)} \tag{5}$$

The state of the SHM system is assumed to remain independent on the state of the structure. Consequently, the joint probability, $P(s_l, m_k)$, can be represented as the product of the individual probabilities, $P(s_l)$ and $P(m_k)$. This assumption holds valid when there are no partial or global collapses. In this context, when the SHM outcome o_j is known, the expected utility of an action A_n is calculated as:

$$E\left[u(A_n)\middle|o_j\right] = \sum_{l=1}^{L}\sum_{k=1}^{K} u(A_n, s_l)\,\frac{P\left(o_j\middle|s_l, m_k\right)P(s_l)P(m_k)}{P\left(o_j\right)} \tag{6}$$

Antecedent to the observation of the SHM system's outcome, decision-makers can compute the expected utility associated with informed decision-making, denoted as $u_{0,M}$. This computation encompasses the comprehensive spectrum of potential SHM outcomes and their respective probabilities, as delineated below:

$$u_{0,M} = \sum_{j=1}^{J} E\left[u\left(\breve{A}_{o_j}\right)\middle|o_j\right]P\left(o_j\right) \tag{7}$$

where $\breve{A}_{o_j}$ is the optimal action in case the SHM outcome is o_j.

If available, the SVT provides information about the state of SHM system. The SVT outcomes are modeled using a random variable c_h, whose realizations are typically equal to the number of SHM system states, i.e., $H = K$. Bayesian updating is employed to revise the prior probabilities of the SHM system states $P(m_k)$ based on SVT outcomes:

$$P(m_k|c_h) = \frac{P(c_h|m_k)P(m_k)}{P(c_h)} \tag{8}$$

where $P(c_h|m_k)$ is the probability of observing SVT outcome c_h when the SHM system is in state m_k, and $P(c_h)$ is the total probability of observing outcome c_h. Generally, the likelihood functions for the SVT outcome have the aspect of a confusion matrix [10].

Prior to observing the results of both the SHM system and the SVT, the decision-maker may calculate the anticipated utility associated with informed decision-making, u_{0,M^2}. This calculation takes into account all potential outcomes for both systems, along with their respective probabilities, as illustrated below:

$$u_{0,M^2} = \sum_{h=1}^{H}\sum_{j=1}^{J} E\left[u\left(\breve{A}_{o_jc_h}\right)\middle|o_j, c_h\right]P\left(o_j\right)P(c_h) \tag{9}$$

In this equation, $E\left[u\left(\breve{A}_{o_jc_h}\right)\middle|o_j, c_h\right]$ represents the optimal action when the outcomes o_j and c_h have been observed.

The comprehensive framework proposed in [10] allows for calculating the VoI associated with both SHM and SVT information. Namely, the $\mathrm{VoI_M}$ measures the difference in expected utility between Pre-Posterior analysis (with SHM information) and Prior analysis (without SHM information):

$$\mathrm{VoI_M} = u_{0,M} - u_1 \tag{10}$$

The term $\mathrm{VoI_{M^2}}$ quantifies the value gained by incorporating additional SVT information alongside SHM data:

$$\mathrm{VoI_{M^2}} = u_{0,M^2} - u_1 \tag{11}$$

The additional gain provided by the SVT, ΔVoI, considering both SHM and SVT information, reads:

$$\Delta\mathrm{VoI} = \mathrm{VoI_{M^2}} - \mathrm{VoI_M} = u_{0,M^2} - u_{0,M} \tag{12}$$

This theoretical framework equips decision-makers with a robust methodology for informed decision-making in scenarios where structures and sensors are subject to uncertainties and potential malfunctions. It enables the integration of multiple information sources and quantification of their impact on utility and decision outcomes.

3 APPLICATION

In this section, the analyzed case study features a generic structure, e.g., a bridge, that can be in two states: the healthy state s_1 or the damage state s_2. There are two possible actions available: A_1, which involves doing nothing, and A_2, which entails interrupting the functionality of the structure. Table 1 presents the utilities associated with different combinations of actions and structural states, all of which are expressed in unitless terms.

Table 1: Utility table

	s_1	s_2
A_1	0	−1
A_2	−0.5	−0.5

The decision maker is considering the installation of an SHM system that provides continuous outcomes, modeled as Normal distributions $N(\mu,\sigma)$, where μ represents the mean value, and σ represents the standard deviation. The SHM system can be in a healthy condition, denoted as m_1 (good condition), or in one of multiple damage conditions, denoted as $m_{(k>1)}$, characterized by the presence of a drift in the SHM outcome. When the SHM system is functioning correctly (in state m_1), the distributions follow $N(1,0.1)$ for s_1 and $N(0.7,0.1)$ for s_2. To illustrate a "faulty" SHM system, it is assumed that a systematic error affects its outcome, modeled as a positive drift δ in the mean value, i.e., $N(\mu+\delta,\sigma)$. Both structure states, s_1 and s_2, are affected by the same values of the drift.

Figure 1 displays the Probability Density Functions (PDFs) of the SHM outcome for different states of both the structure and the SHM system, which are used as likelihood functions. Specifically, the drift δ is assumed to range from 0 (corresponding to m_1) to 1 with a resolution of 0.01, for a total of $K = 101$ states (with 100 states affected by drift presence).

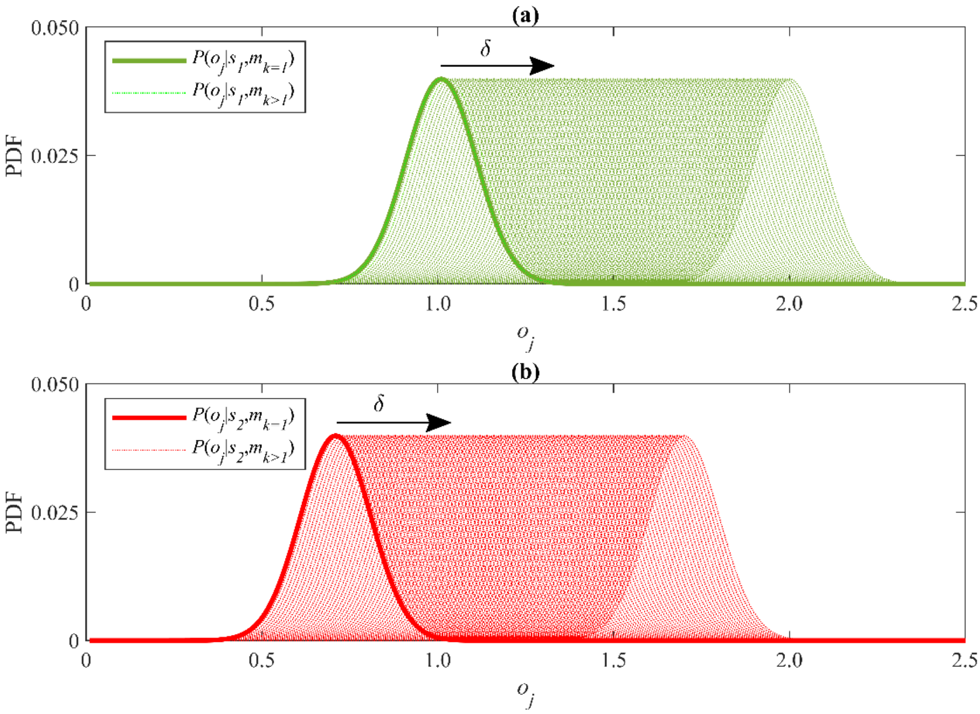

Figure 1: Likelihood functions $P(o|s_l, m_k)$ for (a) structural states s_1 and (b) s_2 for different states of the monitoring system.

Additionally, besides the SHM system, the decision-maker is considering adopting a sensor validation tool (SVT) to identify potential faults in the SHM system. The number of outcomes of the SVT matches the number of states in the SHM system. The SVT provides perfect outcomes. Therefore, the confusion matrix related to the SVT's results consists of a 101x101 square matrix with ones along the diagonal and zeros in other positions.

4 RESULTS

In the previous section, two parameters were not specified, namely the prior probabilities of the states of the structures and the prior probabilities of the states of the SHM system. In turn, the results are expressed as a function of the prior probability $P(s_2)$ that the structure is damaged (state s_2) and the prior probability $P(m_1)$ that the SHM system is working correctly (state m_1). The prior probability $P(s_1)$ that the structure is undamaged (state s_1) can be expressed as $P(s_1) = 1 - P(s_2)$. Regarding the prior probability of the various states of the SHM system, it is assumed that the damage states of the SHM system $m_{(k>1)}$ have uniform probability distributions, i.e., $P(m_{k>1}) = \big(1 - P(m_1)\big)/(K - 1)$.

Figure 2 shows the value of the information provided by the SHM system, without considering the SVT outcome, $\mathrm{VoI_M}$. In general terms, the benefit increases when $P(s_2)$ approaches 0.5 and for increasing values of $P(m_1)$. When both $P(s_1)$ and $P(s_2)$ equal 0.5, the expected costs of the actions A_1 and A_2 are the same: $E[u(A_1)] = 0 \cdot 0.5 + (-1) \cdot 0.5 = -0.5$; $E[u(A_2)] = (-0.5) \cdot 0.5 + (-0.5) \cdot 0.5 = -0.5$. In this scenario, the decision-maker lacks clarity on the optimal course of action, underscoring the particular usefulness of information derived from the SHM system. Regarding the condition of the SHM system, a high probability that the SHM system is in state m_1 corresponds to a high probability that the information it provides is of good quality, leading to an increase in the VoI.

Figure 3 displays the value of the information provided by both the SHM system and the SVT, $\mathrm{VoI}_{\mathrm{M}^2}$. In this instance, the VoI remains constant regardless of the value of m_1. This indicates that the presence of drift does not affect decision-making when the decision-maker is informed about its existence. Drift, being a systematic error, can be rectified when its presence is known.

Figure 4 shows the additional VoI provided by the SVT.

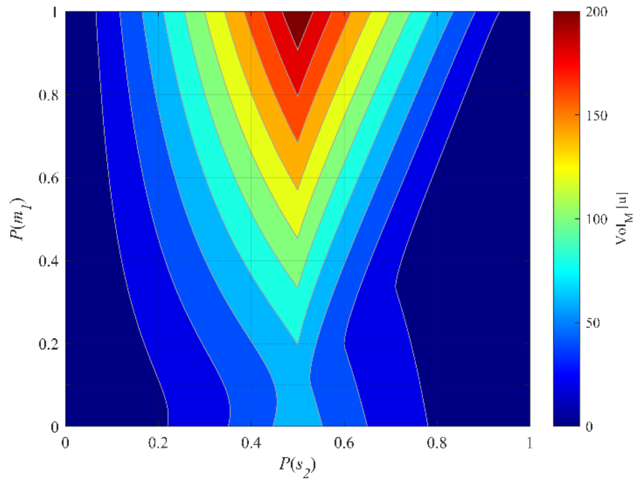

Figure 2: VoI analysis results: The value of SHM information, $\mathrm{VoI}_{\mathrm{M}}$

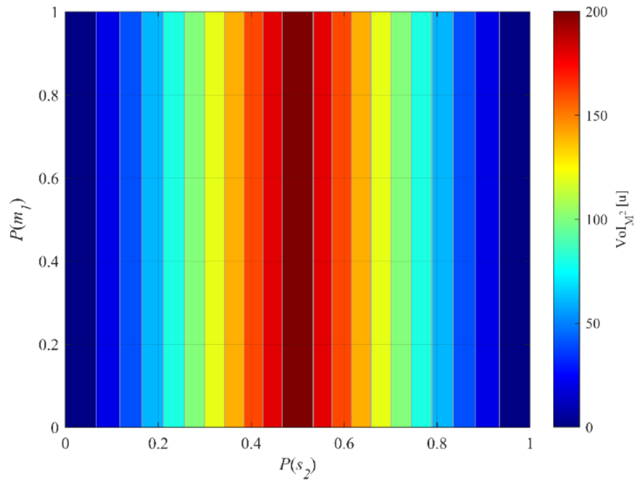

Figure 3: VoI analysis results: The value of SHM and SVT information, $\mathrm{VoI}_{\mathrm{M}^2}$

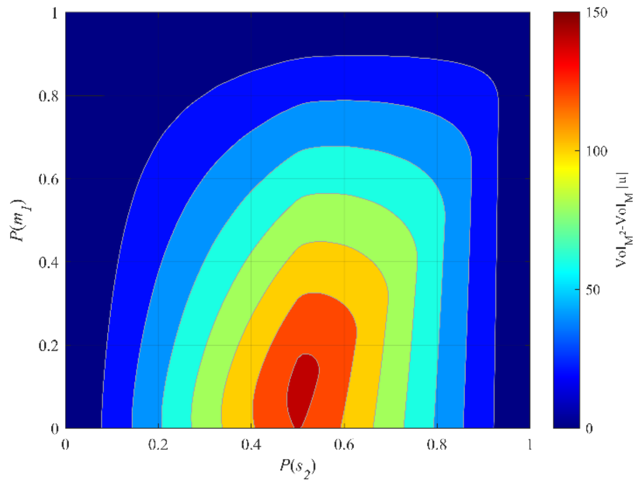

Figure 4: Additional VoI provided by the SVT information, $\mathrm{VoI}_{\mathrm{M}^2} - \mathrm{VoI}_{\mathrm{M}}$

5. CONCLUSION

Issues such as sensor drift and noise interference have the potential to compromise the reliability of SHM outcomes, hampering effective decision-making. Within this context, SVTs have been introduced to identify inconsistencies within collected data, offering valuable insights into data quality. Despite their potential, the practical application of SVTs remains limited in real-world scenarios. Recently, a novel approach that broadens the scope of VoI analysis to encompass the assessment of sensor data quality has been proposed. This framework empowers decision-makers to quantify the additional benefits gained from SVTs and make informed decisions in the face of data quality uncertainties.

In previous applications of this framework, the SHM system was simplified into binary states: either "damaged" or "undamaged." Notably, in cases involving sensor drift, the determination of "damaged" relied on a predefined drift threshold. This paper focuses on a more complex real-world scenario where multiple states of the SHM system coexist. In this context, the exact magnitude of sensor drift, when it occurs, remains uncertain and is not predetermined.

To illustrate the practical implications of the novel VoI framework, it is applied to a case study involving a generic structure. Examined scenarios relate to the VoI provided by the SHM system alone, by the combined information from SHM and SVT sources, and the additional benefit contributed by the SVT information. Results reveal that the presence of sensor drift does not influence decision-making when the decision-maker is informed about its existence. Sensor drift, being a systematic error, can be effectively corrected when its presence is known in advance.

In conclusion, this research illuminates the critical relationship between data quality and decision-making within the context of SHM. By extending the VoI analysis to account for sensor data quality, particularly in scenarios marked by sensor drift, this framework equips decision-makers with a potent instrument to assess the benefit of SHM systems and SVTs. The potential outcomes extend to optimized resource allocation, ultimately culminating in improved infrastructure safety and economic savings.

ACKNOWLEDGMENT

This study was carried out within the RETURN Extended Partnership and received funding from the European Union Next-GenerationEU (National Recovery and Resilience Plan – NRRP, Mission 4, Component 2, Investment 1.3 – D.D. 1243 2/8/2022, PE0000005)

REFERENCES

1. Worden, K., Farrar, C.R., Manson, G., Park, G.: The fundamental axioms of structural health monitoring. Proc. R. Soc. A Math. Phys. Eng. Sci. 463, 1639–1664 (2007). https://doi.org/10.1098/rspa.2007.1834.
2. Magalhães, F., Cunha, A., Caetano, E.: Vibration based structural health monitoring of an arch bridge: From automated OMA to damage detection. Mech. Syst. Signal Process. 28, 212–228 (2012). https://doi.org/10.1016/j.ymssp.2011.06.011.
3. Tonelli, D., Verzobio, A., Cappello, C., Bolognani, D., Zonta, D., Salvatore Bursi, O., Costa, C.: Expected Utility Theory For Monitoring-Based Decision Support System. In: Structural Health Monitoring 2017. DEStech Publications, Inc., Lancaster, PA (2017). https://doi.org/10.12783/shm2017/14095.
4. Kullaa, J.: Distinguishing between sensor fault, structural damage, and environmental or operational effects in structural health monitoring. Mech. Syst. Signal Process. 25, 2976–2989 (2011). https://doi.org/10.1016/j.ymssp.2011.05.017.
5. Kullaa, J.: Detection, identification, and quantification of sensor fault in a sensor network. Mech. Syst. Signal Process. 40, 208–221 (2013). https://doi.org/10.1016/j.ymssp.2013.05.007.
6. Soleimani-Babakamali, M.H., Soleimani-Babakamali, R., Sarlo, R.: A general framework for supervised structural health monitoring and sensor output validation mitigating data imbalance with generative adversarial networks-generated high-dimensional features. Struct. Heal. Monit. 21, 1167–1182 (2022). https://doi.org/10.1177/14759217211025488.
7. Zonta, D., Glisic, B., Adriaenssens, S.: Value of information: impact of monitoring on decision-making. Struct. Control Heal. Monit. 21, 1043–1056 (2014). https://doi.org/10.1002/stc.1631.
8. Kamariotis, A., Chatzi, E., Straub, D.: Value of information from vibration-based structural health monitoring extracted via Bayesian model updating. Mech. Syst. Signal Process. 166, 108465 (2022). https://doi.org/10.1016/j.ymssp.2021.108465.
9. Zhang, W.-H., Qin, J., Lu, D.-G., Liu, M., Faber, M.H.: Quantifying the value of structural health monitoring information with measurement bias impacts in the framework of dynamic Bayesian Network. Mech. Syst. Signal Process. 187, 109916 (2023). https://doi.org/10.1016/j.ymssp.2022.109916.
10. Giordano, P.F., Quqa, S., Limongelli, M.P.: The value of monitoring a structural health monitoring system. Struct. Saf. 100, 102280 (2023). https://doi.org/10.1016/j.strusafe.2022.102280.
11. Giordano, P.F., Quqa, S., Limongelli, M.P.: Value of information analysis accounting for data quality. In: Su, Z., Limongelli, M.P., and Glisic, B. (eds.) Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2023. p. 59. SPIE (2023). https://doi.org/10.1117/12.2657933.
12. Raiffa, H., Schlaifer, R.: Applied Statistical Decision Theory. Division of Research, Graduate School of Business Administration, Harvard University, Boston (1961).
13. Benjamin, J.R., Cornell, C.A.: Probability, statistics, and decision for civil engineers. McGraw-Hill, New York (N.Y.) (1970).