Structural Management and Value of Information Analysis Accounting for Sensor Data Quality

Pier Francesco Giordano
Assistant professor, Dept. of Architecture, Built Environment and Construction Engineering, Politecnico di Milano, Milan, Italy

Said Quqa
Assistant professor, Dept. of Civil, Chemical, Environmental, and Materials Engineering, University of Bologna, Bologna, Italy

Maria Pina Limongelli
Associate Professor, Dept. of Architecture, Built Environment and Construction Engineering, Politecnico di Milano, Milan, Italy; Lisa Meitner Guest professor, Lund Technical University, Sweden

ABSTRACT: Structural health monitoring (SHM) can be used to assess the state of health of civil structures and infrastructures and acquire information that can support maintenance-related activities and post-disaster emergency management. Nevertheless, SHM outcomes may be susceptible to errors due to malfunctioning of the sensing system. The long-term benefit of SHM systems against the initial investment in sensing instrumentation is often quantified without considering the eventuality of faulty sensors. Inaccurate or missing sensor data, not accounted for when information from the SHM system is used to support decisions, can lead to the choice of sub-optimal maintenance actions, and associated economic losses. In the last two decades, Sensor Validation Tools (SVTs) have been proposed, which assess data quality before the SHM information is extracted to isolate and discard abnormal measurements. Nevertheless, automatic SVTs are still rarely implemented in real applications. Recently, a framework based on Bayesian decision theory has been proposed to quantify the benefit of using an SVT before it is implemented. The novel approach extends the traditional VoI to consider multiple "functioning" states of the SHM system with the final goal of quantifying the additional benefit obtained from SVTs. In this paper, this framework is demonstrated using a general example representative of different real situations. Uncertainties in the SVT results are accounted for to show that the adoption of an SVT enhances the overall benefit provided by an SHM system.

1. INTRODUCTION

Environmental factors, repeated loads, and exceptional events lead to the deterioration of civil structures and infrastructures over time. Informed decisions can be beneficial to optimizing expenses, prioritizing interventions, and managing structures in the aftermath of an exceptional event. Structural health monitoring (SHM) has become a mature discipline over the last decades, which allows for identifying and tracking the evolution of structural parameters over time. Once operational and environmental effects are accounted for, variations of these parameters are typically interpreted as modifications in the structural state and can be related to a change in its performance. Nevertheless, SHM systems are imperfect, as the assumptions of the identification algorithms may not always be respected. Also, the inherent noise of collected data and computational errors undermine the SHM outcomes.

Environmental factors and damaging events that jeopardize structural safety also affect the monitoring systems. In particular, sensors and cables are typically the most exposed elements of an SHM system. Their degradation inevitably affects data quality and, in turn, the quality of the SHM outcome. Inaccurate or missing data in the structural assessment and related decision-making can lead to significant economic loss (Smarsly & Law, 2014). Indeed, distorted data may generate false alarms, resulting in unnecessary operational interruptions, while missed detection might increase the probability of catastrophic accidents (Yi et al., 2017). A preliminary check on data quality, consisting of identifying and isolating inconsistent recording channels, is of the utmost importance to limit errors in the SHM process. However, data quality assessment is an additional service that requires specific algorithms and, therefore, involves a cost. While regular quality checks, e.g., detecting missing data, are typically embedded in standard monitoring systems, identifying accuracy and precision loss and the related isolation algorithms, need further efforts.

In the field of data quality assessment, sensor validation tools (SVTs) were first investigated by Dunia et al. (1996) to detect sensor malfunctions in the area of chemical process monitoring. Since then, several researchers have proposed SVTs employing different techniques, among which the ones based on one-class classifiers and multivariate statistical analysis were the most popular (Yi et al., 2017). While the former class studies the sensors individually to understand whether the sensing apparatus is functioning or faulty (Mertikas & Damianidis, 2007), the latter is based on the correlations among the data collected by different sensors in the network (Dunia et al., 1996). Recently, machine learning has gained particular interest in this field and is leading to the rapid development of fault identification algorithms (Mao et al., 2021).

Besides the usefulness of SVTs, the imperfect nature of their outcomes should also be considered. Indeed, as for all the monitoring systems, the probability of false alarms of SVTs is never zero, and their probability of detection is never 100%. Therefore, in real applications, SVTs provide imperfect information, which reduces the uncertainties related to the functioning state of the SHM system but does not eliminate them. Due to the aforementioned cost and imperfect nature of SVTs and, more in general, of data quality analyses, it might be worth quantifying the economic benefit of providing an SHM apparatus with a system dedicated to data quality assessment.

Based on the Bayesian decision theory (Raiffa & Schlaifer, 1961), the concept of Value of Information (VoI) has been employed by several scholars to evaluate the long-term economic benefit provided by an SHM system before it is adopted. In these studies, the VoI was calculated as the expected reduction in management costs associated with the acquisition of new information on the monitored structure through the SHM system (Faber & Thöns, 2013; Pozzi & Der Kiureghian, 2011). This concept was then exploited for several applications, including the optimization of SHM system design (Zhang et al., 2022), the optimization of maintenance strategies (Vereecken et al., 2020), and emergency management (Giordano et al., 2022).

Few studies on the effect of data quality on the VoI exist. For instance, Ali et al. (2022) and Zhang et al. (2023) investigated the effect of biased measurements of the SHM system on the VoI. Very recently, the authors of this paper proposed a general framework to estimate the additional benefit of adopting an SVT coupled with an SHM system (Giordano et al., 2023). The present paper is aimed to extend the mentioned study by investigating the coupled effect of two main data quality issues, namely drift and noise, on the VoI.

The paper is organized as follows. Section 2 introduces the framework presented in (Giordano et al., 2023). Section 3 describes the case study and discusses the results of the VoI analysis accounting for both SHM and SVT information. Section 4 ends the paper with conclusions and general remarks, as well as future works.

2. THEORETICAL FRAMEWORK

The classical Bayesian decision analysis is based on the selection of optimal actions when the state of a system is not known with certainty. It is based on utility considerations and the Bayesian definition of probability. Accordingly, the decision maker selects the action associated with the maximum utility (or minimum cost in the engineering context). Also, the probability associated with unknown parameters reflects the knowledge of the decision maker or personal belief ranging from perfect confidence (probability equal to 1) to absolute no confidence (probability equal to 0).

There are different types of Bayesian decision analysis, depending on the degree of knowledge on the system, namely the Prior, the Posterior and the Pre-Posterior analyses. The Prior analysis is based on the available knowledge of the decision maker, without collecting further information.

The decision-maker computes the expected cost $E[u(A_n)]$ of each action A_n, $n = 1, \dots, N$, considering the prior probabilities $P(s_l)$ of the different states of the system s_l, $l = 1, \dots, L$, and the utility associated with different combinations of actions and states $u(A_n, s_l)$, as follows:

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

The optimal action $\hat{A}$ is the one which brings the maximum utility u_1:

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

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

The Posterior and Pre-Posterior analyses consider new information. Namely, the Posterior analysis is carried out when the outcome o_j is available. The Pre-Posterior analysis considers all the possible outcomes before they are available (for each outcome, a Posterior analysis is performed). Herein, it is assumed that the outcome o_j is provided by an SHM system. In both cases, the prior probabilities of the states of the system are updated through the Bayes' theorem, which reads:

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

with

$$P(o_j) = \sum_{l=1}^{L} P(o_j|s_l) P(s_l) \tag{5}$$

where $P(o_j|s_l)$ is the probability that the outcome o_j is observed when the state of the system is s_l and $P(o_j)$ is the total probability of o_j.

The Posterior and the Pre-Posterior analyses assume that the SHM system used to gain new knowledge is working correctly. In reality, it could provide altered information due to malfunctioning. In turn, knowledge on the state of the SHM system could be collected through SVTs. In (Giordano et al., 2023), the Posterior and the Pre-Posterior decision analysis are extended to account for different states of the SHM system and SVT information. In this section, this framework is recalled to make the paper self-contained. The different states of the SHM system are modelled through the random variable m_k that can assume K different values, $k = 1, \dots, K$. A simple example is $m_1 = damaged$ and $m_2 = undamaged$. The state of the system (that now includes both structure and SHM system) can be described through the joint probability distribution of s_l and m_k. The prior probability of s_l and m_k, $P(s_l, m_k)$, is updated through the Bayes theorem, where the likelihood function is conditioned not only on the state of the structure but also on the state of the SHM system itself, $P(o_j|s_l, m_k)$, as follows:

$$P(s_l, m_k|o_j) = \frac{P(o_j|s_l, m_k) P(s_l, m_k)}{P(o_j)} \tag{6}$$

In this context, if the SHM outcome is available, the expected utility of an action A_n reads:

$$E\big[u(A_n)|o_j\big] = \sum_{l=1}^{L} \sum_{k=1}^{K} u(A_n, s_l) \frac{P(o_j|s_l, m_k) P(s_l) P(m_k)}{P(o_j)} \tag{7}$$

In Eq. (7) as well in the remainder of this paper, it is assumed that the state of the SHM system does not depend on the state of the structure, i.e., $P(s_l, m_k) = P(s_l)P(m_k)$. This is reasonable in absence of partial or global collapses. Before observing the outcome of the SHM system, the decision-makers can compute the expected utility of the informed decision making $u_{0,M}$ by considering all the possible SHM outcomes and their probability of occurrence, as follows:

$$\begin{aligned} u_{0,M} &= \sum_{j=1}^{J} E\left[u\left(\breve{A}_{o_j}\right)|o_j\right] P(o_j) \\ &= \sum_{j=1}^{J}\sum_{l=1}^{L}\sum_{k=1}^{K} u\left(\breve{A}_{o_j}, s_l\right) P(o_j|s_l, m_k) P(s_l) P(m_k) \end{aligned} \tag{8}$$

where $\breve{A}_{o_j}$ is the optimal action when the SHM outcome is o_j. The VoI associated with the SHM information, $\mathrm{VoI_M}$, is computed as the difference between the expected utilities from Eq. (8) and Eq. (3), namely:

$$\mathrm{VoI_M} = u_{0,M} - u_1 \tag{9}$$

When an SVT is adopted, it can give insights into the state of the SHM system. The SVT outcome is modelled through the random variable c_h, $h = 1, \dots, H$. Commonly, the number of SVT outcomes is equal to the number of SHM system states, i.e., $H = K$. The SVT outcome can be used to update the prior probabilities of the state of the SHM system $P(m_k)$, as follows:

$$P(m_k|c_h) = \frac{P(c_h|m_k)P(m_k)}{P(c_h)} \tag{10}$$

where $P(c_h|m_k)$ is the probability of observing the SVT outcome c_h when the state of the SHM system is m_k and $P(c_h)$ is the probability of observing (c_h). The denominator of Eq. (10) can be obtained as:

$$P(c_h) = \sum_{k=1}^{K} P(c_h|m_k)P(m_k) \tag{11}$$

Before observing the outcome of the SHM system and the SVT, the decision-maker can compute the expected utility of the informed decision making u_{0,M^2} by considering all the possible outcomes of both outcomes and their probability of occurrence, as follows:

$$u_{0,M^2} = \sum_{h=1}^{H}\sum_{j=1}^{J} E\left[u\left(\breve{A}_{o_j c_h}\right)|o_j, c_h\right] P(o_j)P(c_h) \tag{12}$$

where $E\left[u\left(\breve{A}_{o_j c_h}\right)|o_j, c_h\right]$ is the optimal action when the outcomes o_j and c_h are observed.

The VoI associated with both SHM and SVT information, $\mathrm{VoI_{M^2}}$, is computed as the difference between the expected utilities from Eq. (12) and Eq. (3), namely:

$$\mathrm{VoI_{M^2}} = u_{0,M^2} - u_1 \tag{13}$$

The supplementary benefit ΔVoI supplied by the SVT is:

$$\Delta\mathrm{VoI} = \mathrm{VoI_{M^2}} - \mathrm{VoI_M} = u_{0,M^2} - u_{0,M} \tag{14}$$

3. APPLICATION

3.1. Description of the case study

The case study analyzed in this section is a generic structure that can be in two states, i.e., s_1 (healthy) or s_2 (damaged). Two actions can be carried out, namely A_1 "Do nothing", or A_2 "Shut the structure down". The (unitless) utilities associated with different combinations of actions and structural states are displayed in Table 1.

Table 1: Utility table.

	s_1	s_2
A_1	0	−1
A_2	−0.5	−0.5

The decision maker is planning to install an SHM system which can be in two states, namely m_1 – good conditions – and m_2 – faulty conditions. It is assumed that the decision-maker does not have any prior knowledge on the state of the SHM system, accordingly $P(m_1) = P(m_2) =$

0.5. If the SHM system is in good condition, it provides a continuous outcome which is modelled through Normal distributions $N(\mu, \sigma)$, where μ and σ are the mean value and the standard deviation, respectively. These parameters depend on the state of the system, which includes both the structure and the SHM system. When the SHM system is working correctly (state m_1), the distributions are $N(1, 0.1)$ and $N(0.7, 0.1)$ for the structure in s_1 and s_2, respectively. These distributions describe the information about the structure provided by the SHM system, thereby they can be interpreted as the distributions of a generic damage index. The corresponding Probability Density Functions (PDFs), used as likelihood functions, are shown in Figure 1.

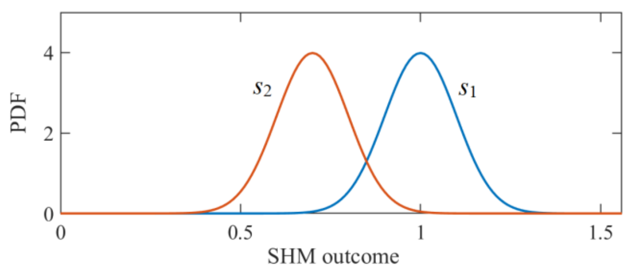

Figure 1: Likelihood functions $P(o|s_l, m_1)$ when the SHM system is working correctly (state m_1).

To exemplify an SHM system in a "faulty" condition it will be assumed that the outcome of the SHM system is affected by an additional source of noise (excessive noise) and is affected by a systematic error. The latter is modelled herein as a positive drift δ of the mean value while the excessive noise is modelled as an increase ε of the standard deviation (Rogers, 2003). Depending on the physical phenomena that generate the fault, the two data quality issues can also occur simultaneously, i.e., $N(\mu + \delta, \sigma + \varepsilon)$. It is assumed that the SHM outcomes in the two possible structure states s_1 and s_2 are affected by the same values of drift and excessive noise.

In addition to the SHM system, the decision maker is considering the adoption of an SVT to identify possible faults in the SHM system. The SVT can provide two outcomes, namely c_1 and c_2. The SVT is not perfect; therefore, its outcomes are affected by uncertainty. The probabilities of observing a given SVT outcome in the different states of the SHM system are shown in Table 2.

Table 2: Likelihood of the SVT.

	m_1	m_2
c_1	0.8	0.2
c_2	0.2	0.8

3.2. Results

The VoI analysis is performed for several values of the prior probability $P(s_2)$ i.e., $P(s_2) = 0.2$, 0.5, and 0.8. The values 0.2 and 0.8 corresponds to low and high, respectively, prior degree of belief on behalf of the decision maker that the structure is in the damaged state. The value 0.5 corresponds to an equal degree of belief in the two states (healthy and damaged) of the structure. To investigate the impact of the severity of the SHM fault on the VoI the sensitivity to values of drift and excessive noise in the interval 0-1 is studied. Figure 2 shows the results for $P(s_2) = 0.2$.

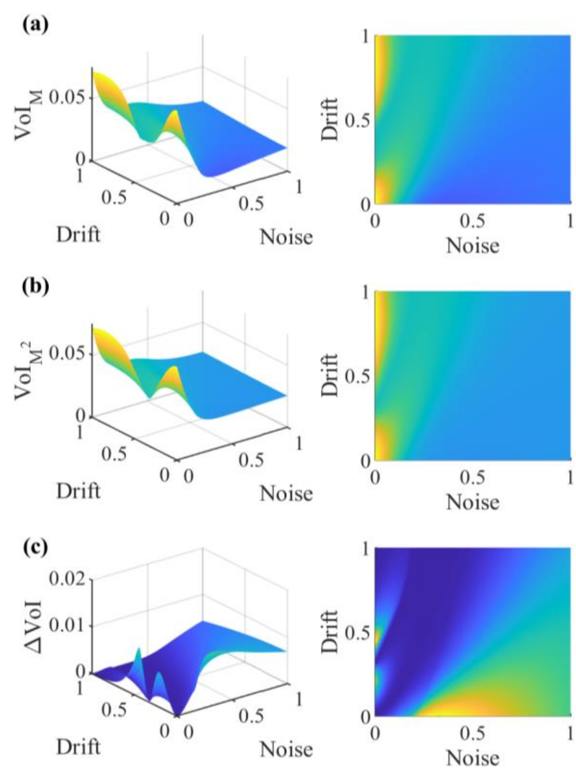

Figure 2: Results of the VoI analysis for $P(s_2) = 0.2$.

In particular, Figure 2(a) displays the VoI associated with the sole SHM information, $\mathrm{VoI_M}$, Figure 2(b) presents the VoI associated with both the SHM and the SVT information, $\mathrm{VoI_{M^2}}$, and Figure 2(c) illustrates $\Delta\mathrm{VoI}$, the supplementary benefit supplied by the SVT. The value of the sole SHM information, $\mathrm{VoI_M}$, is maximum for null values of drift and excessive noise. This benefit generally decreases for increasing values of noise. As for the drift, the $\mathrm{VoI_M}$ drops in the proximity of $\delta = 0.3$. In this situation, the likelihood function $\mathrm{P(o|s_2, m_2)}$ associated with the damaged state of the structure $\mathrm{s_2}$ and the faulty state of the SHM system overlaps with the likelihood $\mathrm{P(o|s_1, m_1)}$ associated with the healthy state of the structure s_1 and the good state of the SHM system m_1. In this situation, the decision maker is not able to distinguish between the two structural states (s_1 and s_2) and the effectiveness of the SHM system decreases.

The VoI associated with both the SHM and the SVT information, i.e., $\mathrm{VoI_{M^2}}$, is generally higher than $\mathrm{VoI_M}$. This happens especially in proximity of ($\delta = 0$, $\varepsilon = 0.3$) and two areas close to ($\delta = 0.3$, $\varepsilon = 0$), that is for $\delta = 0.2$ and $\delta = 0.5$, see Figure 2(c).

Figure 3 illustrates the results of the VoI analysis for $P(s_2) = 0.5$. The VoI values are significantly higher than the corresponding values in Figure 2. Note that a different scale of the z-axis has been used. In the Bayesian context, when $P(s_1) = P(s_2) = 0.5$, the decision maker does not have any prior knowledge on the state of the structure. Thus, new information is particularly valuable. The additional VoI provided by the SVT in Figure 3(c) is particularly high for ($\delta = 0.3$, $\varepsilon = 0$), that is where the plots of $\mathrm{VoI_M}$ and $\mathrm{VoI_{M^2}}$ drop in Figure 3(a) and Figure 3(b), respectively. In this situation, the overlap of the likelihoods relevant to the damaged and healthy state of the structure makes the SHM information unclear, and thereby the contribution of the SVT is particularly valuable.

Figure 4 displays the VoI results for $P(s_2) = 0.8$.

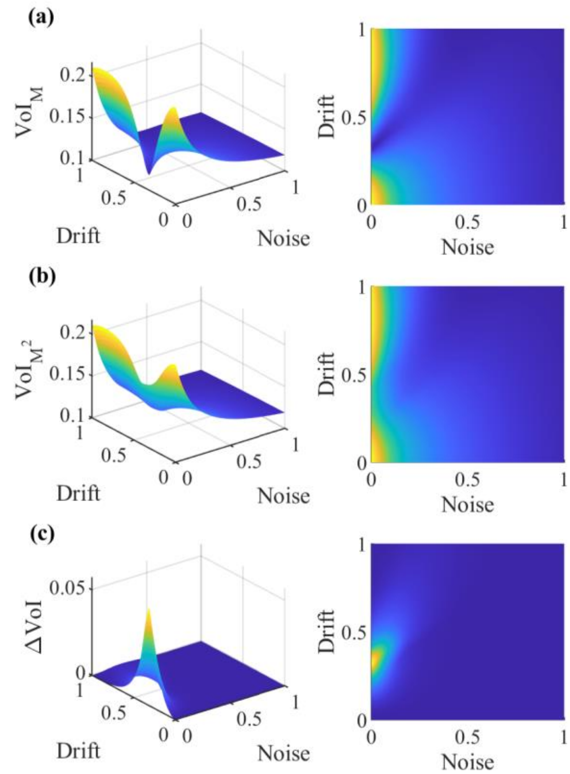

Figure 3: Results of the VoI analysis for $P(s_2) = 0.5$.

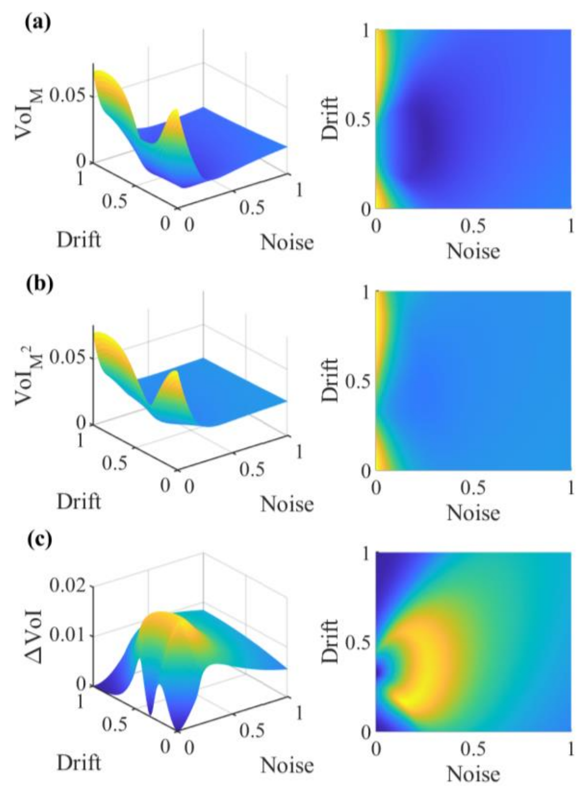

Figure 4: Results of the VoI analysis for $P(s_2) = 0.8$.

The VoI values in Figure 4(a) and (b) are lower than the corresponding ones shown in Figure 3. In turn, they are similar to those in Figure 2. The additional VoI provided by the SVT Figure 4(c) is particularly high in the area surrounding ($\delta = 0.3$, $\varepsilon = 0$) and for $\varepsilon = 0.3$.

To highlight and interpret the additional contribution ΔVoI of the SVT reported in Figure 2(c), Figure 3(c), and Figure 4(c), the absolute difference Δ between the posterior expected utilities of the two actions $\Delta = \left|E\left[u(A_2)|o_j\right] - E\left[u(A_1)|o_j\right]\right|$ is plotted in Figure 5. These expected utilities are computed accounting for the SHM outcome, according to Eq. 7.

In the yellow areas in Figure 5, the two expected utilities have approximately the same value. Hence, in these areas, the selection of the optimal action is not trivial (even if the SHM outcome is available).

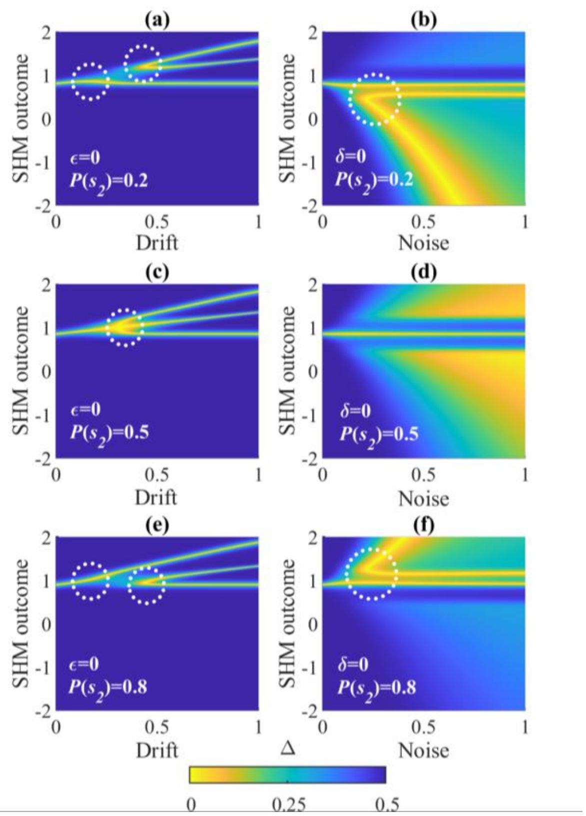

Figure 5: The absolute difference Δ between the posterior expected utilities of the two actions for different values of $P(s_2)$, $\varepsilon = 0$, and $\delta = 0$.

In these situations, the SVT has the highest impact on the selection of the optimal action and thereby the information it provides about the state of the SHM system reaches the highest value. More in detail, Figure 5(a), (c), and (e) relate to $\varepsilon = 0$. In Figure 5(a) and (e), the yellow areas concentrate in correspondence of $\delta = 0.2$ and $\delta = 0.5$ (dashed circles), that is where some ΔVoI maxima are localized in Figure 2(c) and Figure 4(c), respectively, Similarly, in Figure 5(c), the yellow area is concentrated close to $\delta = 0.3$, that is where the ΔVoI peak is located in Figure 3(c). To explain the other ΔVoI maxima, Figure 5(b), (d), and (f), which relate to $\delta = 0$, are analyzed. Namely, in Figure 5(b) and (f), the yellow areas concentrate in correspondence of $\varepsilon = 0.3$, that is where the other ΔVoI maxima are placed in Figure 2(c) and Figure 4(c). As for Figure 5(b), there are no specific areas in which the yellow areas concentrate. There is a single peak in Figure 3(c).

4. CONCLUSION

Altered SHM information can lead to suboptimal management of civil structures and infrastructures. This paper investigates the benefit of the information provided by SVTs on the state of SHM systems. This benefit is computed through a recently proposed framework based on the VoI from Bayesian decision theory. This framework accounts for the different states of SHM systems and the uncertainty in SVT information. A numerical case study is investigated in which two common data issues are considered simultaneously, namely drift and excessive noise in the SHM outcome. Specifically, it is assumed that the SHM system can be in two states: the healthy state and a faulty state corresponding to a combination of drift and excessive noise effects. Results indicate that the VoI from the SHM system decreases when, due to malfunctioning, it is not able to correctly distinguish between the different structural states. In general, the SVT information is valuable to decision-makers. The overall value of information from SHM and SVT is equal to or higher than the value of SHM information alone. The SVT provides significant additional value when the expected utilities (in other words, the expected costs) of management actions estimated accounting for the SHM outcome are similar.

ACKNOWLEDGMENT

PFG and MPL were partially funded by the Italian Civil Protection Department within the project Accordo CSLLPP e ReLUIS "WP3: Analisi, revisione e aggiornamento delle Linee Guida".

REFERENCES

Ali, K., Qin, J., & Faber, M. H. (2022). On information modeling in structural integrity management. *Structural Health Monitoring*, *21*(1), 59–71. https://doi.org/10.1177/1475921720968292

Dunia, R., Qin, S. J., Edgar, T. F., & McAvoy, T. J. (1996). Identification of faulty sensors using principal component analysis. *AIChE Journal*, *42*(10), 2797–2812. https://doi.org/10.1002/aic.690421011

Faber, M., & Thöns, S. (2013). On the value of structural health monitoring. In *Safety, Reliability and Risk Analysis: Beyond the Horizon - Proceedings of the European Safety and Reliability Conference, ESREL 2013* (pp. 2535–2544).

Giordano, P. F., Iacovino, C., Quqa, S., & Limongelli, M. P. (2022). The value of seismic structural health monitoring for post-earthquake building evacuation. *Bulletin of Earthquake Engineering*. https://doi.org/10.1007/s10518-022-01375-2

Giordano, P. F., Quqa, S., & Limongelli, M. P. (2023). The value of monitoring a structural health monitoring system. *Structural Safety*, *100*, 102280. https://doi.org/10.1016/j.strusafe.2022.102280

Mao, J., Wang, H., & Spencer, B. F. (2021). Toward data anomaly detection for automated structural health monitoring: Exploiting generative adversarial nets and autoencoders. *Structural Health Monitoring*, *20*(4), 1609–1626. https://doi.org/10.1177/1475921720924601

Mertikas, S. P., & Damianidis, K. I. (2007). Monitoring the quality of GPS station coordinates in real time. *GPS Solutions*, *11*(2), 119–128. https://doi.org/10.1007/s10291-006-0044-6

Pozzi, M., & Der Kiureghian, A. (2011). Assessing the value of information for long-term structural health monitoring. In T. Kundu (Ed.), *Health Monitoring of Structural and Biological Systems 2011* (p. 79842W). SPIE Press. https://doi.org/10.1117/12.881918

Raiffa, H., & Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Division of Research, Graduate School of Business Administration, Harvard University.

Rogers, S. (2003). Sensor noise fault detection. *Proceedings of the 2003 American Control Conference*, *5*, 4267–4268. https://doi.org/10.1109/ACC.2003.1240506

Smarsly, K., & Law, K. H. (2014). Decentralized fault detection and isolation in wireless structural health monitoring systems using analytical redundancy. *Advances in Engineering Software*, *73*, 1–10. https://doi.org/10.1016/j.advengsoft.2014.02.005

Vereecken, E., Botte, W., Lombaert, G., & Caspeele, R. (2020). Bayesian decision analysis for the optimization of inspection and repair of spatially degrading concrete structures. *Engineering Structures*, *220*, 111028. https://doi.org/10.1016/j.engstruct.2020.111028

Yi, T.-H., Huang, H.-B., & Li, H.-N. (2017). Development of sensor validation methodologies for structural health monitoring: A comprehensive review. *Measurement*, *109*, 200–214. https://doi.org/10.1016/j.measurement.2017.05.064

Zhang, W.-H., Qin, J., Lu, D.-G., Liu, M., & Faber, M. H. (2023). Quantifying the value of structural health monitoring information with measurement bias impacts in the framework of dynamic Bayesian Network. *Mechanical Systems and Signal Processing*, *187*, 109916. https://doi.org/10.1016/j.ymssp.2022.109916

Zhang, W.-H., Qin, J., Lu, D.-G., Thöns, S., & Faber, M. H. (2022). VoI-informed decision-making for SHM system arrangement. *Structural Health Monitoring*, *21*(1), 37–58. https://doi.org/10.1177/1475921720962736