Statistical Pattern Recognition for Optimal Sensor Placement in Damage Detection Applications

Theodora Liangou, Anastasios Katsoudas, Nicholas Silionis, and Konstantinos Anyfantis(✉)

School of Naval Architecture and Marine Engineering, National Technical University of Athens, Athens, Greece
kanyf@naval.ntua.gr

Abstract. A Structural Health Monitoring (SHM) architecture involves the processing of sensor measurements and their translation to decisions about the structure's condition. Given a specific SHM approach, the sensor topology, the feature selection, and the employed detector are the main elements that control the detection performance. This work provides an exploratory analysis of the statistical response patterns that govern a structure subjected to variable loads and methodically arrives at an optimal sensor topology, that maximizes the detection performance. For demonstration purposes, a thin square plate subjected to probabilistically described loads is considered. The damage of interest corresponds to a uniform thickness loss, the detection of which is evaluated at different damage levels (from 1% to a 90% unrealistic upper bound). The damage is to be identified indirectly, through strain sensing. The problem is numerically approached (Finite Elements and Monte Carlo Simulations). The generalized Gaussian likelihood ratio test is employed for setting up the detector. The effect of the feature vector arrangement to the detection performance is assessed through estimations of the probability of detection and false alarm, under the Neyman-Pearson framework. The optimal feature vector has been derived through case-based informal (selective process) or formal (Genetic Algorithms) optimization.

Keywords: Optimal sensor placement · Statistical pattern recognition · Detection theory · Genetic Algorithm · Corrosion

1 Introduction

Modern society relies heavily on structural and mechanical systems. Many of them are presently close to the furthest limit of their design life. Because they cannot be economically replaced, techniques for damage detection are being developed and implemented so that they can continue to be safely used if or when their operation is extended beyond their design service life [1]. These conditions require the early detection of damage to increase structure's reliability and reduce important economic consequences. Structural health monitoring (SHM) is a growing sector related with damaged detection. It has

P. Rizzo and A. Milazzo (Eds.): EWSHM 2022, LNCE 254, pp. 217–226, 2023.
https://doi.org/10.1007/978-3-031-07258-1_23

piqued the interest of the scientific community over the last decade. According to Ostachowicz et al. [2] the performance of the SHM methods is based on the quality of the information collected by the sensors. Even though there is a significant development in sensor technology over the past few years, there is a limit to their performance, so the need for research in other areas becomes significant. Optimal sensor placement is a solution that could lead to a reduction in cost of the SHM system without compromising on the quality of the monitoring approach [2] and is the main challenge of this study. Colombo et al. [3] provide a combined approach based on the Neyman-Pearson likelihood ratio test and Bayesian cost used in an optimization framework. According to the authors a multi-objective optimization scheme is vital to achieve a conjunct saving in costs and improvement in detection performances. Mallardo et al. [4] highlight the importance of impact location identification on composite structures and suggest an optimal sensor positioning on a composite-stiffened aircraft panel aiming to maximize the probability to detect the impact or minimize the error associated with a pre-assigned probability of detection by using Genetic Algorithms (GA). Another interesting work focused on marine applications is proposed in Silionis and Anyfantis [5], where the guiding principle of the damage identification strategy is based upon measuring, through a limited and constant number of sensors, the static strain redistribution caused by an extensive damage. The authors treat the problem as one of statistical pattern recognition, and apply methods derived from machine learning to tackle it. Other interesting works may be found in [6–8].

The present work addresses the important problem of optimizing sensor topology on a given structure. It provides an association of the sensor network arrangement that optimizes the detection performance based on the statistical patterns that govern the feature space, explored in a manual or automated fashion. More specifically, this paper focuses on a decision process for arriving at the topology for sensor placement over a domain of interest, under a variable operational environment. For research purposes, a fictitious SHM problem has been considered, which is described in the second section. In the third section, a description of the probabilistic numerical simulations is provided, followed by that of the employed detector in the fourth section. In the fifth section a presentation of the statistical response is provided. Detection results are given in the sixth section. Emphasis is given at the description of the optimal solution that has been derived by a GA. In the seventh section the conclusions drawn are presented along with some suggestions for future work.

2 Problem Definition

2.1 SHM Detection Basis

The considered SHM problem is focused on detecting corrosion-induced thickness loss (CITL), in a thin-walled structural component. This problem is of interest in marine structures, as these operate in a predominantly corrosive environment. Added thickness is commonly used in plates across vulnerable regions as a preventive measure. The SHM strategy proposed herein is oriented towards predictive, condition-based maintenance. The structure's plate thickness is inferred, indirectly, from capturing load redistributions

caused by stiffness reduction. A classifier, based on detection theory is introduced, which translates the strain-based feature vector into inferences about the structure's condition.

2.2 Case Study Description

A thin plate subjected to probabilistically described static loads is considered. Hence, their magnitude and location are considered as random variables (RVs) that obey a given statistical structure. Strain measurements are extracted and used as the informative features. Damage is introduced as general uniform corrosion by simulating eleven damage levels, ranging from 1% to 90% thickness loss.

A Level I SHM system [9] is developed to discriminate between two states, one corresponding to the intact plate (healthy state), and the other to the corroded plate (damaged state). This binary classification task is treated using a detector based on the maximum likelihood ratio test. The optimal feature vector, with respect to detector performance, is then chosen via an optimization strategy that optimizes both the sensor topology and the measured strain component.

The problem is numerically studied on a clamped square plate ($a = 1000$ mm and $t_0 = 10$ mm) as shown in Fig. 1a. A non-uniform pressure profile, $q(x, y)$, is considered with the mathematical description provided in Eq. (1).

$$q(x, y; Q_0, X_0, Y_0) = Q_0 \exp\left[-\frac{1}{2}\left(\frac{x - X_0}{\sigma_{x_0}}\right)^2\right] \exp\left[-\frac{1}{2}\left(\frac{y - Y_0}{\sigma_{y_0}}\right)^2\right] \tag{1}$$

Its variability (load uncertainty) is controlled through the amplitude Q_0 at the controlled location (X_0, Y_0). These quantities are considered as normal RVs, with $Q_0 \sim N(q_0; \mu_{Q_0}, \sigma^2_{Q_0})$, $X_0 \sim N(x_0; \mu_{x_0}, \sigma^2_{x_0})$ and $Y_0 \sim N(y_0; \mu_{y_0}, \sigma^2_{y_0})$. Parameters μ_{Q_0}, σ_{Q_0} are selected such that the plate remains within its elastic range. To constrain the non-zero magnitude area within the extreme locations of the plate, parameters μ_{x_0} and μ_{y_0} are taken equal to zero and parameters σ_{x_0} and σ_{y_0} are taken equal to $a/12$ (statistically symmetric loading). Some indicative realizations of X_0, Y_0 pairs are plotted in Fig. 1b.

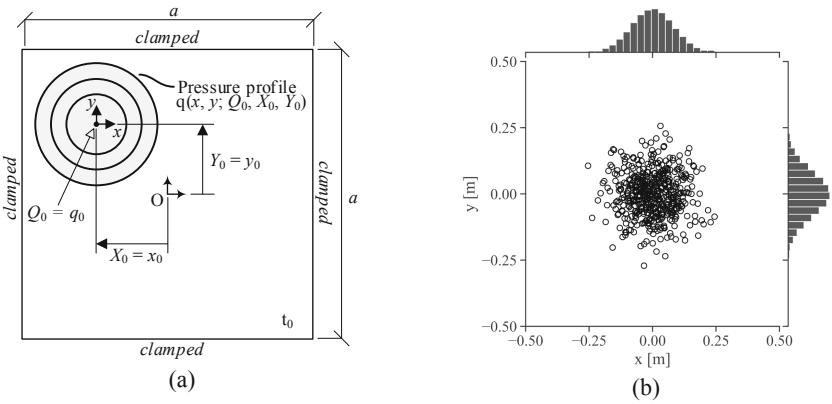

Fig. 1. Considered problem (a) and realizations of the pressure's profile peak (X_0, Y_0) (b).

3 Numerical Simulations

Strain measurements are numerically obtained using Finite Element Analysis (FEA). A mesh consisting of 2500 linear (4-node) shell elements was used and a linear elastic and isotropic material (E = 200 GPa, $\nu = 0.3$) was assigned. Every load realization ($X_0 = x_0$, $Y_0 = y_0$ and $Q_0 = q_0$) represents a virtual statistical experiment, as indicatively presented in Fig. 2.

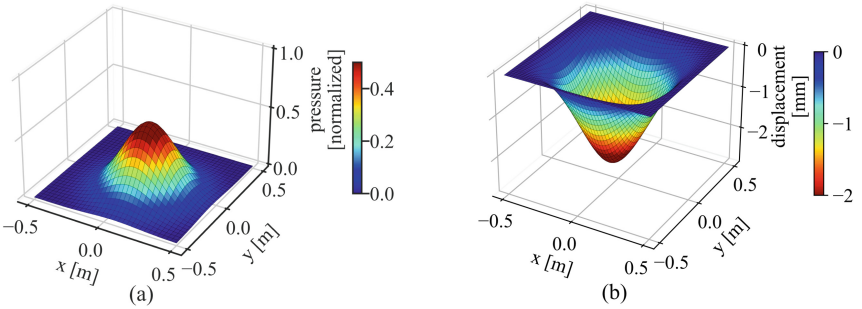

Fig. 2. Indicative pressure distribution (a) and corresponding deflection of the plate (b).

Figure 3 presents an indicative strain distribution of the plate for the healthy state. The form of the response is similar for the damaged state due to the nature of the load. Eleven different damage states are investigated and simulated. Each damaged condition results from the simulation of a FE model with a different thickness from 1% thickness wastage to 90% unrealistic upper bound.

Monte Carlo Simulations (MCS) were employed as the means for forward uncertainty quantification. Latin hypercube sampling is employed to generate independent realizations of Q_0, X_0 and Y_0. Each set corresponds to a numerical statistical test (deterministic FEA). Convergence was achieved at ~10^4 samples. The three in-plane strain components (axial ε_{xx}, transverse ε_{yy} and shear ε_{xy}) of the top surface and at the element center (gauss point location) were stored for each sample.

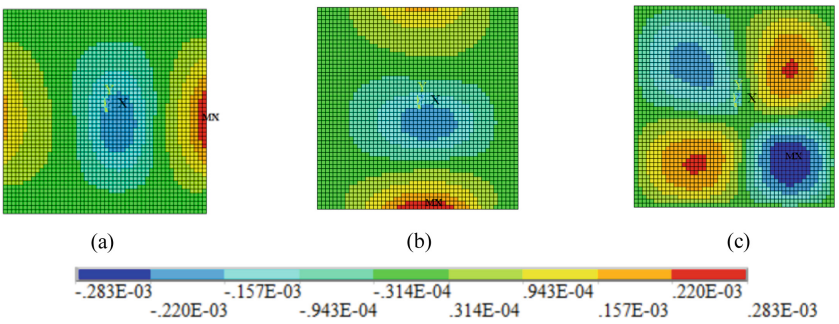

Fig. 3. Contour plot of strain ε_{xx} (a), ε_{yy} (b) and ε_{xy} (c) for the healthy state for the realized case shown in Fig. 2.

4 The Employed Detector

A detector can be regarded of as the transition from a feature vector to a decision [3]. Only binary decisions will be investigated in this research, where one must normally decide whether features received from a test case belong to the reference condition or to the corroded plate. The likelihood ratio test is the detector that is associated with a hypothesis test where a null hypothesis H_0 is defined for the healthy state and the alternative hypothesis H_1 is defined for a modeled damaged state, as follows:

$$H_0 : \mathbf{x} \sim f_{\mathrm{H}}(\mathbf{x}; \boldsymbol{\theta}_{\mathrm{H}}) = N(\mathbf{x}; \boldsymbol{\mu}_{\mathrm{H}}, \boldsymbol{\Sigma}_{\mathrm{H}}) \tag{2}$$

$$H_1 : \mathbf{x} \sim f_{\mathrm{D}}(\mathbf{x}; \boldsymbol{\theta}_{\mathrm{D}}) = N(\mathbf{x}; \boldsymbol{\mu}_{\mathrm{D}}, \boldsymbol{\Sigma}_{\mathrm{D}}) \tag{3}$$

where $\mathbf{x}$ is the n-dimensional feature vector $\mathbf{x} = \{x_1, x_2, .., x_n\}^T$ that holds the strain components registered from a specific sensor network design. The random vector is considered to follow the multinormal distribution with a parameter vector $\boldsymbol{\theta}_{\mathrm{H\,or\,D}}$, that incorporates the mean vector $\boldsymbol{\mu}_{\mathrm{H\,or\,D}}$ and the covariance matrix $\boldsymbol{\Sigma}_{\mathrm{H\,or\,D}}$. A test feature vector, $\mathbf{x}$, is associated with the damaged state (alternative hypothesis) if the likelihood ratio is larger than a threshold γ, as stated below:

$$L(\mathbf{x}) = \frac{N(\mathbf{x}; \boldsymbol{\mu}_{\mathrm{D}}, \boldsymbol{\Sigma}_{\mathrm{D}})}{N(\mathbf{x}; \boldsymbol{\mu}_{\mathrm{H}}, \boldsymbol{\Sigma}_{\mathrm{H}})} > \gamma \tag{4}$$

The likelihood ratio test is written as:

$$L(\mathbf{x}) = \frac{\sqrt{\det \boldsymbol{\Sigma}_{\mathbf{H}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{D}})^T \boldsymbol{\Sigma}_{\mathbf{D}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{D}})\right]}{\sqrt{\det \boldsymbol{\Sigma}_{\mathbf{D}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{H}})^T \boldsymbol{\Sigma}_{\mathbf{H}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{H}})\right]} \tag{5}$$

By taking the logarithm of both sides and rearranging the terms, the detector (hypothesis test) is expressed in terms of a test statistic, $T(\mathbf{x})$, that is in turn associated with a new threshold γ'. The decision rule is:

$$\begin{cases} \text{decide } H_1, \text{ if } T(\mathbf{x}) > \log \gamma + \frac{1}{2} \log \frac{\det \boldsymbol{\Sigma}_{\mathrm{H}}}{\det \boldsymbol{\Sigma}_{\mathbf{D}}} \equiv \gamma' \\ \text{decide } H_0, \text{ if } T(\mathbf{x}) < \log \gamma + \frac{1}{2} \log \frac{\det \boldsymbol{\Sigma}_{\mathrm{H}}}{\det \boldsymbol{\Sigma}_{\mathbf{D}}} \equiv \gamma' \end{cases} \tag{6}$$

Magnitude $T(\mathbf{x})$ is a univariate random variable defined as follows:

$$T(\mathbf{x}) \equiv \frac{(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{H}})^T \boldsymbol{\Sigma}_{\mathbf{H}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{H}})}{2} - \frac{(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{D}})^T \boldsymbol{\Sigma}_{\mathbf{D}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{D}})}{2} \tag{7}$$

The detector's performance is measured in terms of the probability of detection P_{D}:

$$P_{\mathrm{D}} = \Pr\{T(\mathbf{x}) > \gamma'; H_1\} = \int_{\gamma}^{+\infty} f_{\mathrm{D}}(\mathbf{x}; H_1) d\mathrm{x} \tag{8}$$

and the probability of false alarm, P_{FA}, given by the following equation:

$$P_{\mathrm{FA}} = \Pr\{T(\mathbf{x}) > \gamma'; H_0\} = \int_{\gamma}^{+\infty} f_{\mathrm{H}}(\mathbf{x}; H_0)\mathrm{d}\mathbf{x} \tag{9}$$

Under the Neyman-Pearson (NP) framework, given a specific value for P_{FA}, the threshold for a selected sensor topology, i.e. given $\boldsymbol{\theta}_{\mathrm{H}}$ and $\boldsymbol{\theta}_{\mathrm{D}}$, can be obtained by minimizing an error functional defined by comparing the constant P_{FA} with the numerically computed version of the same, $P_{\mathrm{FA_{MC}}}$, through MCS [3]:

$$\gamma'_{MC} = \arg\min_{\gamma}(\ln((P_{FA} - P_{FA_{MC}}(\gamma))^2)) \tag{10}$$

5 Statistical Analysis of the Strain Response

The present chapter presents a statistical exploration of the strain pattern over the plate. The strain data set is extracted from the healthy state (H_0) and eleven strain datasets from each corresponding damaged state (H_1).

Every damaged state is considered as a distinct problem. A 80/20 sample ratio was considered for training/testing purposes. Some important statistics are computed based on a thickness loss of 20% of the nominal thickness of the plate, as a major corrosion scenario that reflects to a loss corresponding to the thickness margin introduced to marine plates for corrosion wastage purposes over the structure's design life (20–25 years). Figure 4 presents contour maps of the mean shift between the undamaged and damaged states, over the plate domain for each strain component. The difference between mean values is proportional to the corrosion wastage level. The shift between sample standard deviations follows a similar pattern. The ratio of the standard deviation of the healthy state and the standard deviation of the damaged state is almost constant over the plate. The ratio increases with higher levels of corrosion and reduces with lower levels.

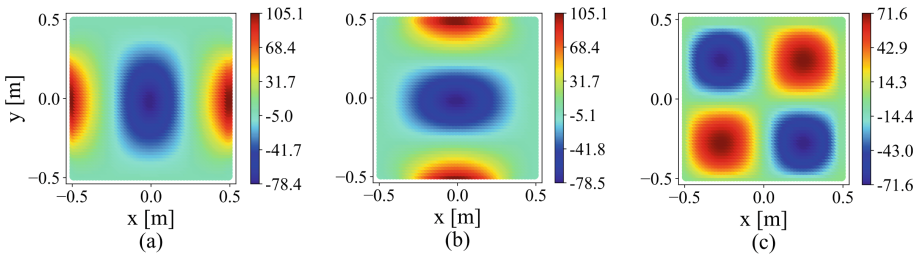

Fig. 4. Difference between the MC estimates of the expected values of $\varepsilon_{\mathrm{xx}}$ (a), $\varepsilon_{\mathrm{yy}}$ (b) and $\varepsilon_{\mathrm{xy}}$ (c) in [με] for 20% thickness wastage (damaged) and the reference (healthy) state.

6 Detector Results

6.1 Detector Performance Assessment

The detector's performance is assessed for the 20% thickness wastage case (strain pattern in Fig. 4). A single strain component ($\varepsilon_{xx}/\varepsilon_{yy}/\varepsilon_{xy}$) is first selected as the unique informative feature (univariate problem). The element-wise evaluation of the performance measures (P_D and P_{FA}) is presented in Fig. 5 for a threshold $\gamma = 1$ used in Eq. (4). These plots reveal information regarding the effect of the location for sensor placement to the detector's performance. There is a clear association between the regions where a high mean shift is recognized in Fig. 4 and the zones that exhibit high P_D in Fig. 5. It is evident though, that P_D and P_{FA} maximize (or minimize) at the same zones within the plate, regardless the selected feature. It is noteworthy that we are aiming at a sensor design that maximizes P_D and minimizes P_{FA}, in the Neyman-Pearson setting. Nevertheless, an optimally placed single sensor yields poor performance metrics ($P_D \approx 65\%$ and $P_{FA} \approx 21\%$), indicating the statistical complexity of the problem at hand.

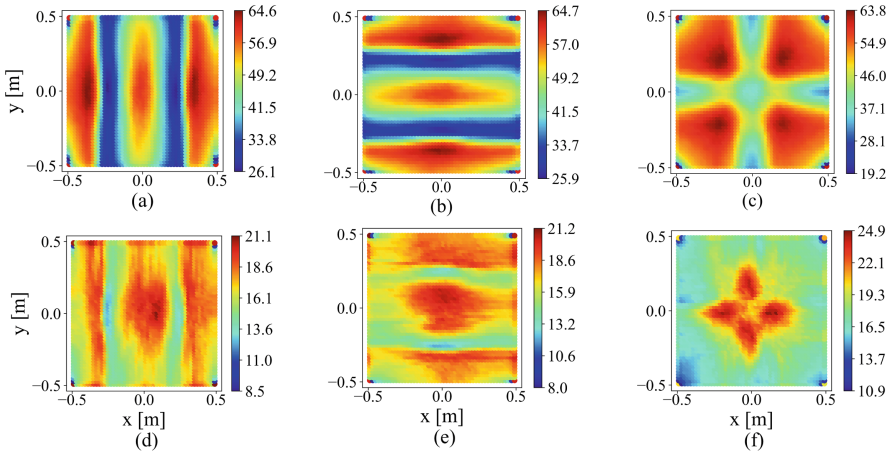

Fig. 5. Distribution of detector's performance over the plate for a univariate (1d) feature vector. Subplots ε_{xx} (a), ε_{yy} (b) and ε_{xy} (c) present measure P_D and subplots ε_{xx} (d), ε_{yy} (e) and ε_{xy} (f) present the corresponding measure P_{FA}. Paired subplots (a)–(d), (b)–(e) and (c)–(f) are to be considered for examination.

On an effort to increase the information content, all three strain components at each element are considered to construct the feature vector (tri-variate problem). The corresponding distribution of the performance metrics over the domain is presented in Fig. 6. A correlation between the distribution of the statistics obtained from single features (Fig. 4) and the detector's performance distribution in Fig. 6 is intractable. Although, an increased performance is observed in some limited elements located at the boundary of the plate, this region is considered singular and is preferred to be disregarded. The remaining smooth pattern within the plate is the pool of candidate locations for sensor placement. However, the information gain from all strain components is marginal ($P_D \approx 69.8\%$ and $P_{FA} \approx 18.7\%$).

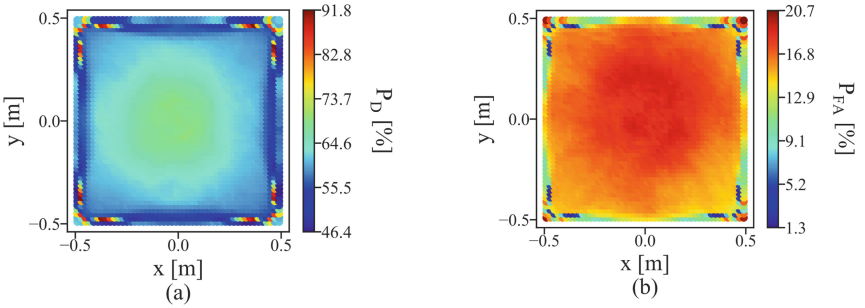

Fig. 6. Distribution of detector's performance over the plate for a tri-variate element-wise (3d) feature vector: P_{D} (a) and P_{FA} (b).

Figure 7 shows the aggregated detection performance metrics for each selected feature vector (univariate or tri-variate case) with respect to the examined corrosion levels. Each point corresponds to the maximum P_{D} obtained within the plate and the corresponding P_{FA}, which does not correspond to the Neyman-Pearson optimal detector, as prescribed by Eq. (10). It is evident that detection accuracy breaks the 50% threshold for corrosion levels higher than 10%. For corrosion wastage above 40%, the detection scheme becomes more powerful achieving a P_{D} of 87%.

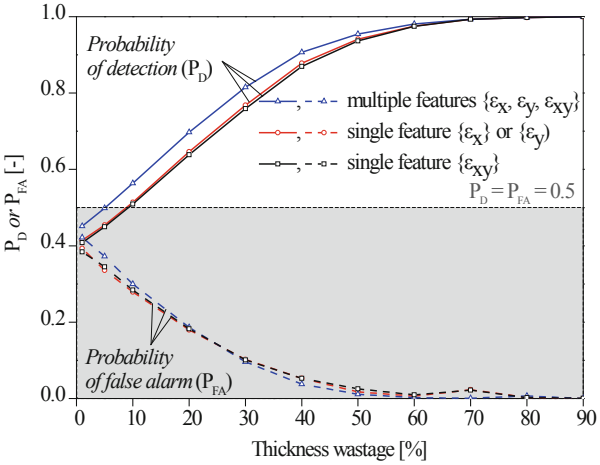

Fig. 7. Maximum probability of detection and corresponding false alarm rate per corrosion level.

6.2 Determining an Optimal Sensor Network

This section focuses on the determination of the optimal sensor topology and the corresponding strain component that maximizes the detector's performance. This is a constrained optimization problem and is tackled through the employment of Genetic Algorithms (GA). The aim is to arrive at a N-dimensional feature vector (holding the design variables), $\mathbf{x}$, corresponding to N sensors together with the specific strain component, k = xx, yy or xy measured at a location.

$$\mathbf{x} = \left\{ x_1^k, \ldots, x_j^k \ldots x_N^k \right\} \tag{11}$$

The features are selected from a database holding ($2500 \times 3 \times 10^4$) entries: (number of elements x number stain components x MC samples) for the healthy and the damaged state (20% corrosion). For each design vector, the parameters $\boldsymbol{\theta}_{\mathrm{H\ or\ D}}$ are calculated and used within Eq. (6) and Eq. (7) for detection performance evaluation. The fitness function was defined as:

$$\mathbf{x}_{\mathrm{opt}} = \arg\max_{\mathbf{x}}(P_{\mathrm{D}}(\mathbf{x})) \tag{12}$$

The results from four optimization cases are presented in Fig. 8. Initially a case were ten sensors measuring a single strain component each was examined. The corresponding feature vector is described by a 10-dimensional multinormal distribution. The problem is solved over the entire plate domain with a resulting max(P_{D}) equal to 99.8% and corresponding P_{FA} equal to ~0%. The optimal sensor topology is presented in Fig. 8a. However, running the optimization by excluding the singular boundary zones (Fig. 6), detection performance drops to 92.4% for max(P_{D}) and increases to ~2.2% for P_{FA}. The optimal position for each single sensor is located at high P_{D} zones over the plate according to Fig. 5.

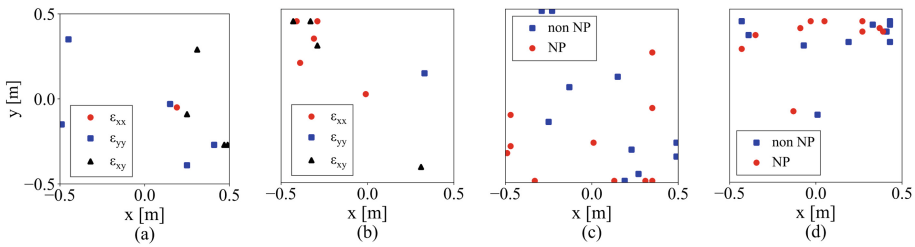

Fig. 8. Optimal positions of single sensors (a) with 3 zones and (b) without 3 zones and positions of rosettes for 20% corrosion (c) with 3 zones and (d) without 3 zones.

In order to enhance the information content of the feature vector, all three strain components were considered at each candidate location, resulting at a 30-dimensional multinormal statistical description. The number of locations for sensor placement remains the same ($N = 10$). This optimization problem is formulated as in the previous cases (we maximize P_{D} and evaluate the corresponding P_{FA} – non NP) and additionally under the NP setting, where we maximize P_{D} for a given P_{FA} (NP). Equation (10) has been

employed for implementing the NP optimal detector and the optimization problem was solved through the employment of GAs. Respective sensor topologies are presented in Fig. 8c–d. The performance for the non-NP cases is $\max(P_D) = 99.9\%/P_{FA} = \sim 0\%$ and $\max(P_D) = 98.4\%/P_{FA} = 0.4\%$, for the entire (Fig. 8c) and the reduced plate domain (Fig. 8d), respectively. On the other hand, the performance of the NP detector for $P_{FA} = 1\%$ has arrived at a $\max(P_D) = 99.9\%$ by optimizing over the entire domain and at a $\max(P_D) = 93.7\%$ by optimizing over the reduced domain. It is evident that the boundary zones that are considered as singular, include highly informative content, however, as aforementioned, these locations may be sensitive to error and inaccuracies arising from e.g. sensor misplacement. The sensor locations are distributed over the plate when the boundary zones are included compared to the sensor concentration in their absence (Fig. 8c–d).

7 Conclusions

This study investigated the statistical pattern recognition for optimal sensor placement in damage detection due to thickness loss. The behavior of a detector under NP framework was examined for all possible sensor placement positions and different levels of corrosion. It proved that the number, locations, and measuring features of the sensors could not be selected manually, as such a choice is not immediately understandable. Thus, the selection of an appropriate optimization technique is considered necessary. A GA is used to locate optimal – in terms of P_D – combinations of sensor locations and strain components measured for a constant number of sensors, in a short computational time, compared to the individual performance of the detector. The solved problem could be extended to investigate the minimum number of sensors that could be placed.

References

1. Farrar, C.R., Worden, K.: Structural Health Monitoring: A Machine Learning Perspective, 1st edn. Wiley, UK (2013)
2. Ostachowicz, W., Soman, R., Malinowski, P.: Optimization of sensor placement for structural health monitoring: a review. Struct. Health Monit. **18**(3), 963–988 (2019)
3. Colombo, L., Todd, M.D., Sbarufatti, C., Giglio, M.: On statistical multi-objective optimization of sensor networks and optimal detector derivation for structural health monitoring. Mech. Syst. Signal Process. **167**, 108528 (2022)
4. Mallardo, V., Aliabadi, M.H., Khodaei, Z.S.: Optimal sensor positioning for impact localization in smart composite panels. J. Intell. Mater. Syst. Struct. **24**(5), 559–573 (2013)
5. Silionis, N.E., Anyfantis, K.N.: Static strain-based identification of extensive damages in thin-walled structures. Struct. Health Monit. (2021)
6. Worden, K., Burrows, A.P.: Optimal sensor placement for fault detection. Eng. Struct. **23**(8), 885–901 (2001)
7. Barthorpe, R.J., Worden, K.: Emerging trends in optimal structural health monitoring system design: from sensor placement to system evaluation. J. Sens. Actuator Netw. **9**(3), 31 (2020)
8. Tan, Y., Zhang, L.: Computational methodologies for optimal sensor placement in structural health monitoring: a review. Struct. Health Monit. **19**(4), 1287–1308 (2020)
9. Rytter, A.: Vibration based inspection of civil engineering structures. PhD Thesis, University of Aalborg, Denmark (1993)