

AKM²D: An adaptive framework for online sensing and anomaly quantification

Hao Yan, Kamran Paynabar & Jianjun Shi

To cite this article: Hao Yan, Kamran Paynabar & Jianjun Shi (2020) AKM²D: An adaptive framework for online sensing and anomaly quantification, IISE Transactions, 52:9, 1032-1046, DOI: [10.1080/24725854.2019.1681606](https://doi.org/10.1080/24725854.2019.1681606)

To link to this article: <https://doi.org/10.1080/24725854.2019.1681606>



View supplementary material [↗](#)



Accepted author version posted online: 22 Oct 2019.
Published online: 21 Nov 2019.



Submit your article to this journal [↗](#)



Article views: 149



View related articles [↗](#)



View Crossmark data [↗](#)



AKM²D: An adaptive framework for online sensing and anomaly quantification

Hao Yan^a , Kamran Paynabar^b, and Jianjun Shi^b

^aSchool of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA; ^bSchool of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA

ABSTRACT

In point-based sensing systems such as coordinate measuring machines and laser ultrasonics where complete sensing is impractical due to the high sensing time and cost, adaptive sensing through a systematic exploration is vital for online inspection and anomaly quantification. Most of the existing sequential sampling methodologies focus on reducing the overall fitting error for the entire sampling space. However, in many anomaly quantification applications, the main goal is to estimate sparse anomalous regions at pixel-level accurately. In this article, we develop a novel framework named Adaptive Kernelized Maximum-Minimum Distance (AKM²D) to speed up the inspection and anomaly detection process through an intelligent sequential sampling scheme integrated with fast estimation and detection. The proposed method balances the sampling efforts between the space-filling sampling (exploration) and focused sampling near the anomalous region (exploitation). The proposed methodology is validated by conducting simulations and a case study of anomaly detection in composite sheets using a guided wave test.

ARTICLE HISTORY

Received 20 July 2018
Accepted 8 October 2019

KEYWORDS

Adaptive sampling; space-filling design; exploration and exploitation; kernel methods; anomaly quantification

1. Introduction

Systematic exploration of large areas for anomaly quantification is of particular importance in various applications including quality inspection, sensor network design, and structural health monitoring, etc. (e.g., in airplane maintenance (Wu *et al.*, 1996), wafer manufacturing (Jin *et al.*, 2012) and additive manufacturing (Gibson *et al.*, 2014)). For example, in metrology and Non-Destructive Evaluation (NDE), various point-based sensing systems are used for quality inspection and anomaly quantification. Examples include touch-probe coordinate measuring machines (CMMs) used for measuring the dimensional accuracy (Simpson, 1992), and non-destructive methods such as Guided Wave-field Tests (GWT) (Mesnil *et al.*, 2014) and laser ultrasonics (Aussel and Monchalain, 1989), utilized for defect quantification in composite sheets.

Most point-based sensing systems are only capable of measuring one point at a time. The algorithm provides a binary map to show those pixels that are anomalies. Given this binary map, the number, location, shape, or other features of anomalous regions can be easily derived by simply using the morphological operations in image processing. This binary map can be used to: (i) determine the number of defective areas (damages or imperfections); (ii) locate each defective area; and (iii) identify the shape of each defective area. This is subsequently used to determine if a part needs to be repaired or discarded; and to identify potential root causes in the part fabrication process. However, to achieve the foregoing goals, a point-based sensing method requires the sequential measurement of a large


number of points, which results in a time-consuming procedure that is not scalable to the online inspection of large areas. For example, using a touch-probe CMM, it may take more than 8 hours to measure one typical batch of wafers that includes 400 wafers of 11" diameter (Jin *et al.*, 2012). Also, using WWF test, the high-resolution inspection of a composite laminate of size 0.4 m² may take up to 4 hours (Mesnil and Ruzzene, 2016) (the experimental setup will be shown in Figure 1). However, due to the fact that anomalies are often clustered and sparse, most of the sensing points are actually irrelevant and the quick location of the important regions is important. Therefore, one can use a sequential and adaptive sampling strategy to reduce the measurement time or energy consumption by reducing the number of sampled points.

Existing adaptive sampling/sensing strategies in the literature can be classified into three groups: the multi-resolution grid strategy, Sequential Design Of Experiments (SDOE), and representative points selection.

1. Multi-resolution grid sensing has been widely used in practice. It begins with sensing over a coarse (low-resolution) grid to estimate the underlying functional mean (e.g., the image background in two-dimensional (2D) measurements) and find the rough locations of the anomalies. Then, sensing is continued over a finer (high-resolution) grid around the identified anomalies to estimate the shape and size of the anomaly. The performance of this method depends on the predefined size

CONTACT Hao Yan  haoyan@asu.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uiie.

 Supplemental data for this article can be accessed on the [publisher's website](#).

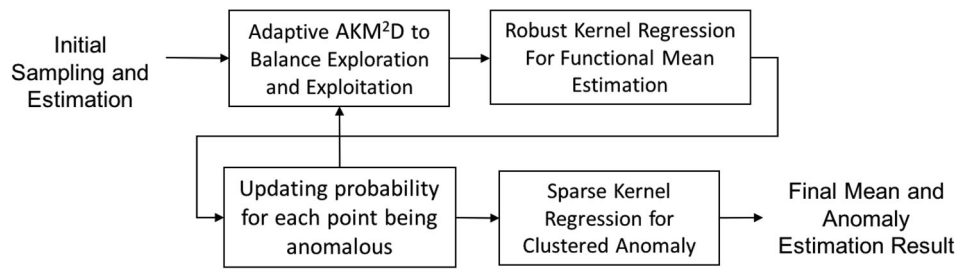


Figure 1. Procedure of the proposed sampling algorithm.

of both coarse and fine grids, which should be specified based on the size and shape of the expected anomalies. Since such information may not be available in advance, this method may result in either over-sampling or poor anomaly quantification caused by under-sampling.

2. Many sampling strategies have been developed in the area of the design of computer experiments for the modeling of spatial profiles. For example, some existing research focuses on selecting design points on the unit hypercube. Space-filling designs such as Latin Hypercube Design (LHD) (McKay *et al.*, 1979) and its many variations (e.g. Owen, 1994; Ye, 1998; Joseph and Hung, 2008) are widely used and have useful space-filling properties (Stein, 1987). Other popular designs such as the Halton sequence (Halton, 1964) and Sobol sequence (Sobol, 1967), are mainly applied in the Quasi Monte Carlo approach to evaluate numerical integrals; they have desirable properties, such as uniformity and low discrepancy (Sobol, 1976). The main problem is that these methods are not sequential and cannot direct the sampling direction toward the region of interest (anomaly region). To address this issue, a large amount of work has been reported in the literature on various (SDOE) methods to improve the model fitting of spatial profiles (Bernardo *et al.*, 1992; Welch *et al.*, 1992; Jones *et al.*, 1998; Ranjan *et al.*, 2008; Jin *et al.*, 2012). Loepky *et al.* (2010) classified the existing SDOE methods into model-based and distance-based (space-filling) methods. Model-based methods include maximizing the expected improvement criterion (Jones *et al.*, 1998; Ranjan *et al.*, 2008), minimizing the prediction error, minimizing the variance of the parameter estimates, e.g., D-optimal design (De Aguiar *et al.*, 1995), and optimizing a composite index (Jin *et al.*, 2012). Among the distance-based models, sequential LHD design (Kyriakidis, 2005; Xiong *et al.*, 2009) and sequential maximin design (Stinstra *et al.*, 2003; Loepky *et al.*, 2009) are popular. Loepky *et al.* (2009) and Loepky *et al.* (2010) showed the latter methods perform well, not only in selecting the initial sampling points, but also in determining the follow-up runs for SDOE, since one can place an upper bound on the Mean Squared Error (MSE) based on the distance measure of the design. The main problem of SDOE methods is that they only focus on improving the estimation of the functional mean over the entire sampling space without considering potential anomalies and non-smooth features.

3. Joseph *et al.* (2015) proposed the minimum energy design that selects representative points based on a known distribution over the design space, and sequentially chooses the next design points based on a criterion minimizing the total potential energy. However, the main problem of applying this approach for online anomaly quantification is that the anomaly distribution is often unknown *a priori*. Therefore, it lacks the ability of focused sampling near anomalous regions.

The second relevant body of literature deals with function estimation in the presence of anomalies. Robust kernel regression (Zhu *et al.*, 2008) and robust spline estimation (De Boor, 1972) are among these methods. However, their main focus is the estimation of the functional mean, not the anomaly, and hence, they do not consider the spatial structure of anomalies. To address this issue, Yan *et al.* (2015) and Yan *et al.* (2018) proposed Smooth-Sparse Decomposition (SSD) for anomaly detection in spatial profiles and temporal profiles, respectively. SSD can separate anomalies from the functional mean by utilizing the spatial structure of both the functional mean and anomalies. SSD, however, can only work when measurements are dense, hence, it is not applicable in point-based sensing and inspection systems.

The third relevant body of literature deals with the adaptive sampling problem for change point detection or Statistical Process Control (SPC). For example, Li and Qiu (2014) proposed a dynamic sampling scheme for SPC based on the p-value of the conventional CUSUM control chart. Liu *et al.* (2015) and Wang *et al.* (2018) proposed an adaptive sampling method to dynamically update the sampling location based on the current CUSUM statistics. However, the major focus of SPC is to detect out-of-control samples as soon as possible. The accurate quantification of the anomaly is not the primary focus for the SPC applications. Even though many SPC methods (Zou and Qiu, 2009; Liu *et al.*, 2015; Wang *et al.*, 2018) have the ability to identify the location of the anomaly, it is typically not possible to fully quantify the anomaly when it is first detected. In comparison, the proposed method focuses on online anomaly quantification with the least number of samples to achieve simultaneous anomaly quantification when it is first detected. In this case, we first need to detect the location of the anomaly and then to use more samples to provide accurate quantification. Online anomaly quantification is typically performed after SPC, to provide accurate anomaly quantification. Furthermore, SPC focuses on timely detection of a change in a dynamic setting. We assume the sample background and anomaly do not change over time.

The main objective of this article is to propose a new adaptive sampling framework along with estimation procedures for online anomaly quantification. The immediate benefit of the proposed framework is to help scale up point-based sensing methods so that they can be used for *in situ* inspection. It can also be used in sensing point selection for online anomaly detection in the sensor network (Wang *et al.*, 2004). An effective adaptive sensing strategy should consist of two major elements: first, it should randomly search the entire space (exploration) to spot anomalous regions and recover the functional mean; and second, it should perform focused sampling on the areas near the anomalous regions (exploitation) to determine the size and the shape of anomalies. To achieve this, the following two challenges should be addressed: (i) how to intelligently decide on the location of the next sampled point; and (ii) how to estimate anomalous regions as well as the functional mean online based on the sparsely sampled points. In this article, we will address the first challenge by proposing a new sensing strategy named Adaptive Kernelized Maximum-Minimum-Distance (AKM²D) combining the computer design of experiment approach for the random exploration of the entire space and the Hilbert Kernel approach (Devroye and Krzyżak, 1999) for the focused sampling in anomalous regions (exploitation). We also show the relationship of the proposed method to the existing SDOE methods. To address the second challenge concerning anomaly estimation, we propose a modeling framework based on robust kernel regression to estimate the background (profile mean) and sparse kernel regression to estimate and separate anomalies. In order to perform both estimation and adaptive measurement in real-time, we also propose efficient optimization algorithms.

The remainder of this article is organized as follows. Section 2 provides an overview of the proposed methodology. In Section 3, we propose the new adaptive sampling/sensing framework AKM²D. Section 4 elaborates mean estimation and anomaly detection algorithms. In Sections 5 and 6, simulated data and a case study of anomaly detection in composite laminates are used to evaluate the performance of the proposed methodology. Finally, we conclude this article with a short discussion and directions for future work in Section 7.

2. Methodology overview

We first briefly review the overall methodology proposed in this article. We assume that the quality measure can be modeled by a true background function $\mu(r)$ and the anomaly function $a(r)$, and the measurement at location r is given as

$$y(r) = \mu(r) + a(r) + \epsilon(r).$$

The goal is to estimate the background function $\mu(r)$ and anomaly function $a(r)$. This is different from SDOE methods, where the main goal is to estimate the mean function. Here we also assume that the functions $\mu(r)$ and $a(r)$ are static and thus do not change over time. Therefore, to achieve this goal, our methodology should include an adaptive sampling framework to identify the best location of the next point as well as an estimation procedure for recovering the background function and quantifying anomalies. The adaptive sampling

framework is to decide the sampling location r_{n+1} based on observations $y(r)$ at r_1, \dots, r_n . The estimation procedure should be able to update the functional estimation on $\mu(r)$ and $a(r)$ given a new observation r_{n+1} . In this article, we assume that the function background $\mu(r)$ is smooth and the anomaly $a(r)$ is sparse and clustered. Finally, we assume that the measurement noises $\epsilon(r)$ in different locations are independent and follow the normal distribution $N(0, \sigma^2)$.

For illustration purposes and simplicity, we use the 2D sampling space $[0, 1]^2$ in this article and further constrain the samples to be on a 2D fine grid defined as

$$\mathcal{G}_m = \left\{ \left(\frac{i}{m}, \frac{j}{m} \right) \mid i, j = 1, \dots, m \right\},$$

where m can be specified by the resolution capability of the sensing device. In the application of spatial sensing in composite part inspection, typically the dimension does not exceed two (e.g., x, y axes). However, the proposed method can be extended to a high-dimensional sampling space. More discussion is added in the methodology section.

The proposed methodology, illustrated in Figure 1, is summarized as follows: First, n_{init} initial points are sampled using a space-filling design (e.g., max-min distance, (Johnson *et al.*, 1990)) to explore the entire sampling space. Then, based on the outcome of the initial points, subsequent points are chosen by using AKM²D to balance between the space-filling sampling (exploration) and the focused sampling near the anomalous region (exploitation). After AKM²D chooses the location of a new sample, the functional mean is estimated (updated) via robust kernel regression. After a certain number of sampled points, if the functional mean estimate does not significantly deviate from the estimate obtained in the previous iteration, the functional mean estimation step can be skipped to reduce the computational time. The probability that a point in the sample space is anomalous is then updated after the estimation step of the functional mean, which is used as an input for AKM²D in the next iteration. Next, clustered anomalous regions are estimated (updated) via the proposed sparse kernel regression. Finally, this procedure is repeated until the desired sampling resolution is reached. The sampling resolution can be defined as the maximin distance, which can be defined as the maximum distance of points in the entire sampling space to the nearest sampled point.

In developing the proposed sampling methodology, we make the following assumptions: we assume that sparse anomalies are in the form of clusters. Also, for estimating the functional mean and anomalous regions, it is assumed that the functional mean is smooth and anomalies have different intensity values from the functional mean. It should be noted that the proposed AKM²D framework is general and does not require the smoothness assumption.

3. The AKM²D sensing algorithm

3.1. Formulation and algorithm

In this section, we present our new adaptive sensing framework, AKM²D, that helps sequentially choose the location

of samples. Denote \mathcal{M}_n as the observed samples in the n th iteration, defined as $\{r_k = (x_k, y_k) \in \mathcal{G}_m | k = 1, \dots, n\}$ are observed. Let $p_a(r_k)$ denote the known probability that the point r_k in this set is anomalous (the detailed procedure for estimating $p_a(r_k)$ will be discussed in Section 4.) To find the next sampled point r_{n+1} , we propose the following criterion:

$$r_{n+1} = \arg \max_r g_n(r) = \arg \max_r \psi_n(r) (f_n(r))^\lambda, \quad (1)$$

where $\psi_n(r)$ is the estimated distribution of anomalies. Therefore, maximizing $\psi_n(r)$ can encourage the focused sampling (exploitation), meaning that the next sampled point r_{n+1} continues searching in anomalous regions. $f_n(r)$ can be understood as the regularization term to prevent sampled points being too close to each other. In other words, $f_n(r)$ encourages the exploration of the entire sampling space for undiscovered anomalies (space-filling property). In this article, we define $\psi_n(r)$ as a mixture distribution of Gaussian distributions centered at each anomalous point observed, and a uniform distribution for the entire sampling space to account for unobserved anomalies. That is, $\psi_n(r) = (\sum_{k=1}^n p_a(r_k) K_h(r, r_k) + u)$ where

$$K_h(r, r_k) = \frac{1}{(\sqrt{2\pi}h)^2} \exp\left(-\frac{\|r - r_k\|^2}{2h^2}\right)$$

is the 2D-Gaussian kernel centered at point r_k used to model the clustered structure of the anomalies. $p_a(r_k)$ and u are, respectively, the mixture weights for the Gaussian distribution $K_h(r, r_k)$ and the uniform distribution. Note that $p_a(r_k)$ is also the probability that the sampled point r_k is anomalous. Even though the normalization weight

$$\frac{1}{\sum_{k=1}^n p_a(r_k) + u}$$

changes over different iterations, it can still be neglected, as it is independent of r and does not affect the optimization result in Equation (1). Furthermore, we define $f_n(r)$ by $f_n(r) := \min_{r_k \in \mathcal{M}_n} \|r - r_k\|$ to encourage the space-filling property. For a special case $\psi_n(r) = 1$, Equation (1) becomes $r_{n+1} = \arg \max_r \min_{k=1, \dots, n} \|r - r_k\|$, which is equivalent to a greedy approach to solve the maximum minimum-distance design proposed by Johnson *et al.* (1990). By plugging in $\psi_n(r)$ and $f_n(r)$, the sampling criterion given in Equation (1) can be rewritten as

$$r_{n+1} = \arg \max_r \left\{ \left(\sum_{k=1}^n p_a(r_k) K_h(r, r_k) + u \right) \min_{k=1, \dots, n} \|r - r_k\|^\lambda \right\}. \quad (2)$$

Algorithm 1: AKMMD

initialize

Initial n_{init} sampling based on max-min distance design

end

for $n = n_{init}, \dots, n_{max}$ **do**

Update $\psi_n(r)$ based on $\Psi_n = K_x^T P_A K_y + u \mathbf{1}_{m \times m}$

Update $f_n(r) = \min(f_{n-1}(r), \|r - r_n\|)$ for $r \in \mathcal{G}_m$

$r_{n+1} = \arg \max_{r \in \mathcal{G}_m} \psi_n(r) (f_n(r))^\lambda$

end

To efficiently solve Equation (2) on $r \in \mathcal{G}_m$, we compute $\psi_n(r)$ by the tensor product of two 1D-Gaussian kernels. That is, $\Psi_n = K_x^T P_A K_y + u \mathbf{1}_{m \times m}$, where

$$K_{x,ij} = K_{y,ij} = \frac{1}{(\sqrt{2\pi}h)^2} \exp\left(-\frac{\|i - j\|^2}{2h^2 m^2}\right),$$

and

$$P_{A,ij} = p_a\left(\frac{i}{m}, \frac{j}{m}\right) \mathbf{1}\left(\left(\frac{i}{m}, \frac{j}{m}\right) \in \mathcal{M}_n\right)$$

are the (i, j) component of the matrix K_x , K_y , and P_A , respectively. $\mathbf{1}(x)$ is an indicator function defined as

$$\mathbf{1}(x) = \begin{cases} 1 & x \text{ is true} \\ 0 & x \text{ is false} \end{cases}$$

and $\mathbf{1}_{m \times m}$ is an m by m matrix of ones. It is straightforward to show that $f_n(r), r \in \mathcal{G}_m$ can be updated recursively by $f_n(r) = \min(f_{n-1}(r), \|r - r_n\|), r \in \mathcal{G}_m$. Both the space and time complexity of this recursive update is $O(m^2)$, where m is the grid size in each dimension. Therefore, Equation (2) can be efficiently and recursively solved by Algorithm 1. Here, we would like to emphasize that the current optimization algorithm is based on a grid search approach, in which the function value of the grid \mathcal{G}_m is updated in each iteration. To optimize, the largest numerical value of the function $f(r)$ on the grid \mathcal{G}_m can be computed. However, we found that this grid searching algorithm is actually more efficient than the global solver in the 2D sampling space, as it is a recursive algorithm, and hence, it requires updating the function value for a small portion of the points. However, this approach is only feasible if the dimension of the sampling space is not too large. To optimize the design in a higher-dimensional sampling space, a global optimization solver, such as a particle swarm algorithm or genetic algorithm, can be used. For more details about using these solvers to find the near-optimal design, please refer to Mak and Joseph (2018).

3.2. AKM²D sampling properties

In this section, we study the properties of the proposed AKM²D. Let \mathcal{R}_i denote the neighborhood of a point r_i defined by $\mathcal{R}_i = \{r | \|r - r_i\| \leq \|r - r_k\|, \forall k = 1, \dots, n\}$. We first investigate the behavior of the sampling criterion $g(r)$ in the neighborhood of an anomalous point r_a , i.e., \mathcal{R}_a . (see Figure 2). It is easy to show that Equation (2) for $r \in \mathcal{R}_a$ can be decomposed into two terms: $g(r) = g_a(r) + g_{-a}(r)$, where $g_a(r) = (K_h(r, r_a) + u) \|r - r_a\|^\lambda$, $g_{-a}(r) = \left(\sum_{k \neq a} p(r_k) K_h(r, r_k)\right) \|r - r_a\|^\lambda$. The second term, $g_{-a}(r)$, can often be neglected in the neighborhood of r_a , especially when $\|r_k - r_a\| \gg h, \forall k \neq a$. For simplicity, in this subsection, we assume r_a is the only detected anomalous point with $p_a > 0$.

Proposition 1. *The local maximum of $g_a(r), r \in \mathcal{R}_a$ is attained at*

$$\|r - r_a\| = d_a^* = h \sqrt{\lambda - 2W\left(-\frac{\pi h^2 \lambda u}{p_a} \exp\left(\frac{\lambda}{2}\right)\right)}$$

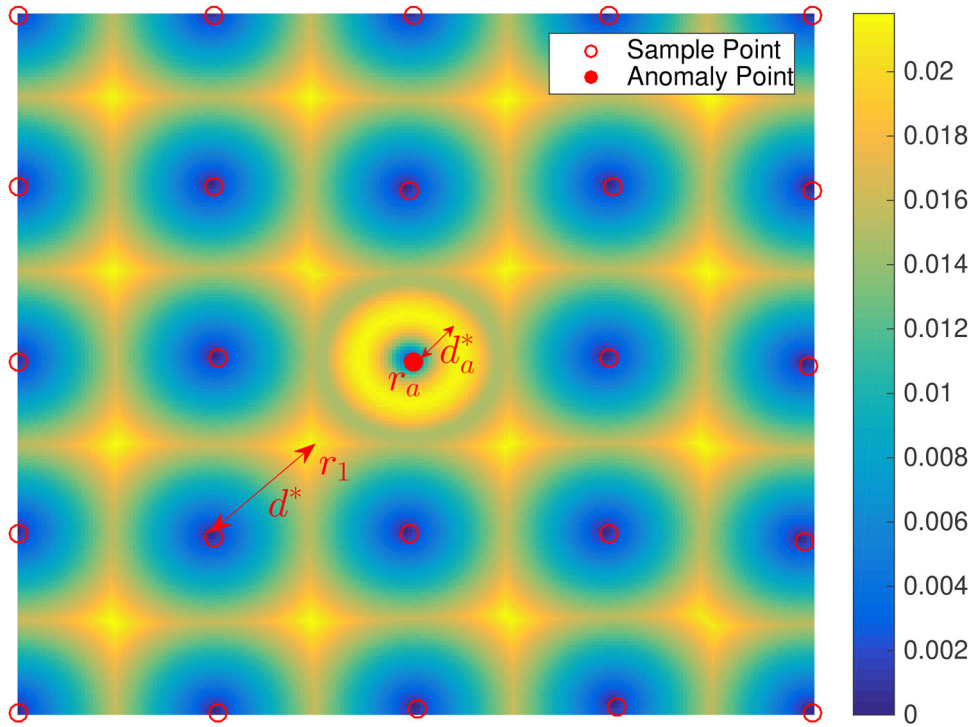


Figure 2. The behavior of $g(r)$ with the center point as the anomaly point.

if $\{r : \|r - r_a\| = d_a^*\} \in \mathcal{R}_a$. W is the Lambert W -function defined as $W(z) = \{w|z = w \exp(w)\}$.

The proof is given in Appendix A.

Proposition 1 guarantees that $g_a(r)$ in the neighborhood of r_a will generate a local maximum ring with radius d_a^* (as shown in Figure 2), which encourages the next sampled point to be chosen near the potential anomalous point r_a (exploitation), but with the distance of d_a^* to avoid over-exploitation. Proposition 1 only guarantees the local optimality. However, the next sampled point is selected on the local maximum ring only if it is the global maximum of $g(r)$. To study this and show how criterion (2) is able to balance sampled points between exploration and exploitation, we give the following necessary condition under which the algorithm selects r_a^* .

Proposition 2. Let d^* denote the current sampling Max-Min Distance (MMD) defined as the maximum distance of each point in the entire sampling space with its closest sampled point, i.e., $d^* := \max_j \min_{r_k \in \mathcal{M}_n} \|r - r_k\|$ and suppose r_a is the only sampled point with $p_a > 0$. Then

$$\|r - r_a\| = d_a^* = h \sqrt{\lambda - 2W\left(-\frac{\pi h^2 \lambda u}{p_a} \exp\left(\frac{\lambda}{2}\right)\right)}$$

is the global maximum of Equation (2) if

$$d^* < \tilde{d}^* := \left(\frac{1}{(1 + \exp(-c^2))} \times \frac{(d_a^*)^2}{2((d_a^*)^2 - \lambda h^2)}\right)^{\frac{1}{2}} d_a^*, \quad (3)$$

where c is a constant satisfies

$$c < \min_{r_k \neq r_a, r_k \in \mathcal{M}_n} \|r_k - r_a\| / 2 \sqrt{2h^2 \ln\left(\frac{p_a}{2\pi h^2 u}\right)}.$$

The proof is given in Appendix B.

Proposition 2 shows that the proposed algorithm first samples the entire sampling space up to a certain resolution \tilde{d}^* and then, starts focused sampling. This ensures that the proposed method does not miss any anomaly with a radius greater than \tilde{d}^* . Furthermore, this proposition can be used to select the tuning parameters, which will be discussed in the next section.

To illustrate the implication of this proposition, we plot the behavior of $g(r)$ in Figure 2. The center point in this figure is an anomalous point (the point indicated by r_a), which generates a local optimal ring with a radius d_a^* . It will be global optimum if this optimal value is larger than the other local maximum in the center of the potential sampled points (the point indicated by r_1) as shown in Figure 2. Proposition 2 shows that if Equation (3) holds, the algorithm will select a point on the local maximum ring centered at r_a as the global optimum, and hence, as the next sampled point.

Proposition 1 and Proposition 2 study the transient behavior of the algorithm and how it balances the exploration and exploitation when the anomaly is first discovered. We also study the limiting behavior of the proposed sampling algorithm in Proposition 3 and Remark 4.

Proposition 3. If $\psi_n(r) = 1$, the sampling points distribution will converge to the uniform distribution. Furthermore, the MMD of the sampling points decrease at the rate of $O\left(\frac{1}{n^{1/p}}\right)$ for a p -dimensional sampling space.

The proof of Proposition 3 follows directly by Theorem 1 in the Minimum Energy Design (Joseph et al., 2015). Proposition 3 demonstrates the limiting distribution of the proposed algorithm when the anomaly is not present (i.e.,

$\psi_n(r)$ is a constant) is actually the uniform distribution. Furthermore, even though this rate $O\left(\frac{1}{n^{1/p}}\right)$ can only be proved when $\psi_n(r) = 1$, it is proved in Joseph *et al.* (2015), that $O\left(\frac{1}{n^{1/p}}\right)$ is actually the upper bound of the convergence rate for any design. We found in our simulation study that the MMD also decreases at the rate of $O\left(\frac{1}{n^{1/p}}\right)$. This result can guide practitioners on deciding the number of points needed for the sampling algorithm.

Remark 4. If the $\psi_n(r)$ converges to $\psi(r)$, with appropriate initial sampling points, the limiting distribution of the sampling algorithm $r_{n+1} = \arg \max_r \left\{ \psi(r) \min \|r - r_k\|^\lambda \right\}$ will converge to $\psi(r)^{p/\lambda}$, where p is the dimension of the sampling space.

Remark 4 is a conjecture and depends on whether the limiting behavior of Minimum Energy Design (Joseph *et al.*, 2015) can be proved. The assumptions of Remark 4 are discussed in Appendix C.

Remark 4 demonstrates the limiting behavior of the proposed algorithm. It demonstrates that the limiting distribution is related to the function $\psi_n(r)$. Recall that $\psi_n(r) = \left(\sum_{k=1}^n p_a(r_k) K_h(r, r_k) + u \right)$, which is a combination of the background and anomaly. Therefore, the proposed algorithm is able to balance exploration and exploitation. Finally, as the number of sampling points $n \rightarrow \infty$, the kernel density estimation $1/n \sum_{k=1}^n p_a(r_k) K_h(r, r_k)$ will approach the true anomaly distribution. In the limiting behavior, when $n \rightarrow \infty$, more samples will be put on the anomaly regions, due to the increasing weight according to the $\sum_{k=1}^n p_a(r_k) K_h(r, r_k)$ that results from the sample size.

3.3. Tuning parameter selection

In this section, we will discuss how these propositions can help us select the tuning parameters λ, h , and u . First, based on our numerical experiments in the simulation study, we suggest the kernel bandwidth h is selected approximately at the scale of the desired anomaly sampling resolution (e.g., Anomaly MMD).

In the simulation, we find out λ needs to have a value of five for the algorithm to demonstrate the behavior of both exploration and exploitation. Correspondingly, u is normally set as small as $u < 10^{-7}$. To decide the exact value of λ and u , we can use the transient behavior of the algorithm demonstrated in Propositions 1 and 2 to select the tuning parameter. Since

$$d_a^* = h \sqrt{\lambda - 2W \left(-\frac{\pi h^2 \lambda u}{p_a} \exp \left(\frac{\lambda}{2} \right) \right)} \approx h \sqrt{\lambda}$$

when u is small, $h \sqrt{\lambda}$ should be roughly the desired AMMD for exploiting the anomaly. Furthermore, according to Proposition 2, since

$$d_a^* - \lambda h^2 \approx 2 \frac{\pi h^2 \lambda u}{p_a} \exp \left(\frac{\lambda}{2} \right),$$

we know

$$d^* = \left(4\pi u \exp \left(\frac{\lambda}{2} \right) \right)^{-1/\lambda} h \sqrt{\lambda}.$$

Note that when computing \tilde{d}^* , we ignore

$$\left(\frac{1}{1 + \exp(-c^2)} \right)^{\frac{1}{2}}$$

since it is close to one when $c > 3$ and $\lambda > 5$. For example, $c = 3$, $\lambda = 5$,

$$\left(\frac{1}{1 + \exp(-c^2)} \right)^{\frac{1}{2}} = 0.99998.$$

If d^* is larger than the size of sampling space (i.e., $d^* > 1$), the algorithm may be trapped in the anomalous region, as it may never start exploration. Therefore, we can set

$$\left(4\pi u \exp \left(\frac{\lambda}{2} \right) \right)^{-1/\lambda} h \sqrt{\lambda} < 1.$$

for example, if the desired $h = 0.02$ and we set $\lambda = 5$, $u = 1e - 9$, this inequality implies λ should have a value of at least five.

3.4. Relationship with model-based criterion in SDOE

In this section, we aim to link the proposed criterion in Equation (1) with the existing SDOE methods and show how it can be derived from the SDOE perspective. We can use a Gaussian process on the entire sampling space to represent the mean response $\mu \sim GP(0, K(\cdot))$, with the K matrix defined as

$$K_{ij} = K(r_i - r_j) = \exp \left(-\frac{\|r_i - r_j\|^2}{h^2} \right).$$

The anomaly can be defined as the location where the measurement y deviates from the mean response μ (i.e., $y - \mu$ is large). Equivalently, the anomaly should be located in locations where the MSE of the prediction $MSE(r)$ is large. It is straightforward to prove that for a Gaussian process $MSE(r) \leq c_0 \min_{k=1, \dots, n} \|r - r_k\|^2$ (Loeppky *et al.*, 2010), where c_0 is a constant. Since we only care about the anomalous regions, we define

$$IMSE = \left[\int \left\{ \sqrt{MSE(r, D, \theta)} \psi_n(r) \right\}^\beta dr \right]^{1/\beta},$$

as the Integrated MSE (IMSE) over the anomalous density, where $\psi_n(r)$ is the estimated distribution of anomaly r . IMSE can be also explained as the integrated confidence intervals of the prediction.

Note that when $\beta \rightarrow \infty$, $IMSE \leq \max_r \min_{k=1, \dots, n} \|r - r_k\| \psi_n(r)$, and therefore, the solution of $r_{n+1} = \arg \max_r \min_{k=1, \dots, n} \|r - r_k\| \psi_n(r)$ gives the point which contributes most to reducing the integrated confidence intervals of the prediction. This provides an SDOE-based justification for the proposed sampling criterion and strategy in Equation (1). The only difference between the IMSE-based and AKM²D criteria is the parameter λ that is used to adjust

the relative importance of the $\psi_n(r)$ compared with the exploration part.

4. Mean and anomaly estimation using sparse samples

In the previous section, we proposed a general adaptive sampling strategy and discussed its properties. Here, we propose methods for estimating the mean function, as well as anomalous regions, using the sparse measurements obtained by AKM²D. Specifically, we present a robust kernel regression algorithm for functional mean estimation and a sparse kernel regression algorithm for anomaly estimation.

4.1. Robust kernel regression for functional mean estimation

Let z_k denote the recorded measurement at point $r_k = (x_k, y_k)$ and $z = (z_1, \dots, z_k, \dots, z_n)$ be the vector of measurements for all n sampled points. To model the smooth functional mean μ in the presence of anomalies, A Reproducing Kernel Hilbert Space (RKHS) is utilized. From the representer theorem (Schölkopf *et al.*, 2001), it is known that every function in an RKHS can be written as a linear combination of kernel functions evaluated at sampled points. If anomalies do not exist, then kernel regression can be used to estimate the functional mean. However, since anomalies have a different functional structure from the mean, they behave as outliers when estimating the functional mean. Therefore, we utilize robust kernel regression to alleviate the effect of anomalies on mean estimation. To estimate the functional mean μ , we minimize:

$$\sum_{k=1}^n \rho(z_k - \mu_k) + \lambda \|\mu\|_H, \quad (4)$$

in which $\rho(x)$ is the Huber loss function, defined by

$$\rho(x) = \begin{cases} x^2 & |x| \leq \frac{\gamma}{2} \\ \gamma|x| - \frac{\gamma^2}{4} & |x| > \frac{\gamma}{2} \end{cases}$$

and $\lambda \|\mu\|_H$ is the Hilbert norm penalty, which controls the smoothness of the functional mean. The Robust kernel regression can be solved efficiently via an iterative soft-thresholding function (Mateos and Giannakis, 2012). See Appendix C for a detailed derivation and optimization algorithm. The functional mean μ is almost constant after sensing enough sampled points. Therefore, to speed up the algorithm, we stop updating μ when the estimation difference after adding a new sampled point is smaller than a certain threshold. After estimating the functional mean μ_k , the residuals can be computed by $\hat{e} = [\hat{e}_k] = [z_k - \hat{\mu}_k]$.

4.2. Updating probability $p_a(r_k)$

We conduct a hypothesis test on the residual \hat{e}_k to test whether there exist anomalies in the specimen at the location r_k . The null hypothesis is $H_0 : \mu_{e_k} = 0$, implying no anomalies

exist. The p-value of this test can be used to update the probability of the sampled point r_k being anomalous. That is,

$$\begin{aligned} p_a(r_k) &= P(|e_k| > |\hat{e}_k| | e_k \sim N(0, \hat{s}^2)) = 1 - 2P(e_k > \hat{e}_k) \\ &= 2\Phi(|\hat{e}_k|/\hat{s}) - 1, \end{aligned}$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution, \hat{s} is the standard deviation of the noise e , which can be estimated by the median absolute deviation under the normality assumption as $\hat{s} = \text{median}\{|\hat{e}|\}/0.6745$. $p_a(r_k)$ is used as an input to AKM²D as discussed earlier. Moreover, the selection of γ can be determined based on a specified false positive rate, α_0 , associated with the hypothesis test. If no anomalies exist (H_0 is true), the false positive rate can be computed by

$$P\left(|e_k| > \frac{\gamma}{2} | e_k \sim N(0, \hat{s}^2)\right) = 2\left(1 - \Phi\left(\frac{\gamma}{2\hat{s}}\right)\right) = \alpha_0.$$

Consequently, γ can be selected by $\hat{\gamma} = 2\hat{s}\Phi^{-1}(1 - \frac{\alpha_0}{2})$. See Appendix C for the reason as to why $\frac{\gamma}{2}$ is a good threshold to determine whether a point is anomalous.

4.3. Sparse kernel regression for clustered anomaly estimation

In this subsection, we estimate the size, shape, and boundary of anomalous regions. Specifically, we model the spatial structure of clustered anomalies by a Gaussian kernel K_a through optimizing

$$\underset{\theta_a}{\text{argmin}} \|\hat{e} - K_a \theta_a\|^2 + \gamma_a |\theta_a|_1. \quad (5)$$

Problem (5) can be solved efficiently by existing L1 solvers such as the Accelerated Proximal Gradient (APG) algorithm. The APG algorithm for solving Problem (5) is given in Algorithm 2. For the tuning parameter γ_a , as has been pointed out by Yan *et al.* (2015), the Generalized Cross Validation (GCV) approach usually tends to select more points, leading to a larger false positive rate. Therefore, instead of using GCV, we choose γ_a based on a specified false positive rate α . Since there is no closed-form solution for Problem (5) with general K_a , Monte Carlo simulations can be used to select γ_a as follows: We first generate white noise from $e \sim NID(0, \hat{s}^2)$, where \hat{s} is the standard deviation of the noise e . We then select γ_a such that $\alpha \times 100\%$ of $\hat{a} = K_a \hat{\theta}_a$ are non-zero. Note that since K_a changes over time, γ_a should be recomputed whenever a new point is measured, which is time-consuming. Therefore, an approximate procedure for tuning parameter selection is proposed. When K_a is orthogonal, θ_a has a closed-form solution computed by

$$\hat{\theta}_a = S_{\frac{1}{\alpha}}(K_a^T \hat{e}),$$

or equivalently,

$$\hat{\theta}_{ai} = S_{\frac{1}{\alpha}} \left(\sum_j K_a(r_j, r_i) \hat{e}_j \right).$$

When K_a is close to orthogonal, the soft-thresholding function gives a reasonable approximate solution. The false

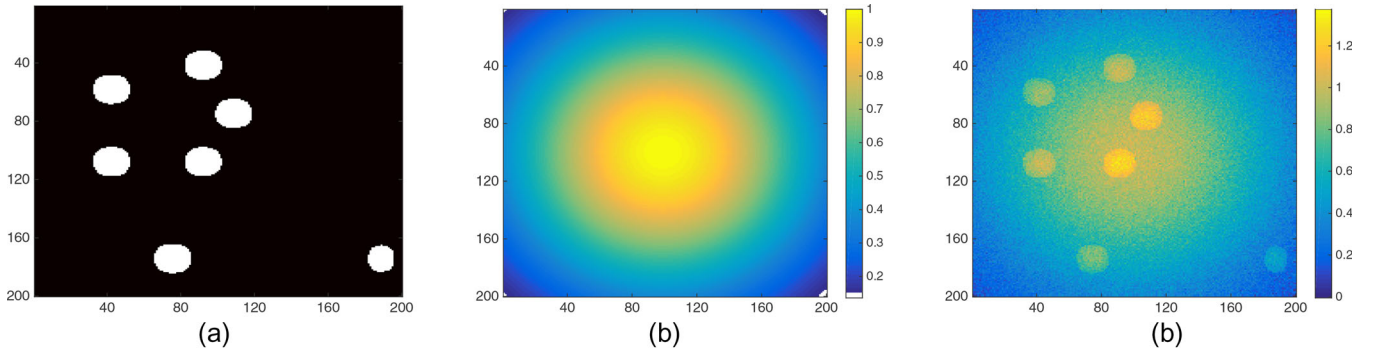


Figure 3 Simulated images with both functional mean and anomalies: (a) simulated anomalies; (b) simulated functional mean; and (c) simulated images coupled with noise and anomalies.

positive rate can then be computed by

$$\begin{aligned}\alpha &= P(\hat{\theta}_{ai} \neq 0) \\ &= 2P\left(|z| > \frac{\gamma_a}{2} \mid z \sim N(0, l^2 \hat{s}^2)\right) = 2\Phi\left(1 - \frac{\gamma_a}{2l\hat{s}}\right),\end{aligned}$$

where $l^2 = \sum_j K_a(r_j, r_i)^2$. Therefore, γ_a can be approximated by $\gamma_a = 2l\hat{s}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$.

To determine the anomalous regions, since the Gaussian kernel is not localized, we threshold the solution to Problem (5) with a small threshold w to ensure noises are not detected. Consequently, anomalous regions are estimated by $1(\hat{a} > w)$, where $1(x)$ is the indicator function. In our study, we select $w = 0.005\hat{s}$. Furthermore, as the number of points in anomalous regions increases, the corresponding kernel size should decrease accordingly. Therefore, we update the bandwidth of kernel K_a (i.e., h_a) proportionally to the sampling resolution in anomalous regions. That is, $h_a = c_h \max_{r \in \hat{a}} \min_{r_k} \|r - r_k\|$. From the simulation study, we found $c_h = 0.2$ works reasonably well.

Algorithm 2: APG algorithm for sparse kernel estimation of anomalies

initialize

Choose a basis for the background as B

$$\theta_a^{(0)} = 0$$

end

while $|\theta_a^{(k-1)} - \theta_a^{(k)}| > \epsilon$ **do**

$$\text{Update } \theta_a^{(k+1)} \text{ by } \theta_a^{(k+1)} = S_{\frac{\gamma}{2}}\left(x^{(k)} + K_a^T(e - K_a x^{(k)})\right)$$

$$\text{Update } t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$

$$\text{Update } x^{(k+1)} = \theta_a^{(k)} + \frac{t_k - 1}{t_{k+1}}\left(\theta_a^{(k)} - \theta_a^{(k-1)}\right)$$

end

5. Simulation study

To evaluate the performance of the proposed methodology, we simulate 200×200 images with a smooth functional mean denoted by matrix M whose elements are obtained by evaluating

$$M(x, y) = \exp\left(-\frac{(x^2 + y^2)}{4}\right)$$

at points

$$x = \frac{i}{201}, y = \frac{j}{201}; i, j = 1, \dots, 200.$$

In this study, seven anomaly clusters are generated by $A = B_s A_s B_s^T$, in which B_s is a cubic B-spline basis with 13 knots, and A_s is a 13 by 13 sparse matrix with seven randomly selected non-zero entries denoted by S_A . The elements of A_s are defined by $A_s(i, j) = \delta \cdot 1(a_{ij} \in S_A)$, where $\delta = 0.3$ characterizes the intensity difference between anomalies and the functional mean. Random noises E are generated from $E \sim NID(0, \sigma^2)$ with $\sigma = 0.05$. Finally, the set of 200×200 simulated images, Y , is generated by adding the anomalies and random noises to the functional mean, i.e., $Y = M + A + E$. A sample of simulated functional mean, anomalies, and a noisy image with anomalies are shown in [Figure 3](#). The goal of this simulation study is to accurately estimate anomalous regions with the least number of sampled points.

We compare our proposed adaptive sampling framework, AKM²D, with four benchmark methods, the random sampling method (designated by “Random”) and multi-resolution grid sampling (designated by “Grid”), adaptive maximum-minimum design (designated by “DOE”), and adaptive Gaussian process criterion (designated by “Variance”). In the Random sampling method, the sampled points are selected purely at random. In Grid sampling, the sampled points are first selected on a 15×15 coarse grid. If $p_a > 0.5$, a finer grid with a five-times-higher resolution is then used to sample within the coarse grid containing anomalous points. For the adaptive maximum-minimum design, the point is selected by maximizing the minimum distance of the entire sampling space. For adaptive Gaussian process criterion, we first fit a Gaussian process model to the entire space. Second, the point is selected at the point with the largest confidence interval of the fitted value. We apply the proposed estimation method to the sampled points obtained by both AKM²D and the benchmarks to estimate the anomalous regions. In this way, the difference in anomaly detection performance can only be attributed to the sampling strategy.

To evaluate the anomaly quantification accuracy, we propose to use the precision, recall, and F1-score to evaluate the pixel-level image segmentation accuracy. The average

Table 1. Anomaly detection result with 250 sampled points.

Methods	250 sampled points				
	AKM ² D	Random	Grid	Variance	DOE
Precision	0.69(0.0011)	0.80(0.0023)	0.80(0.0013)	0.74(0.006)	0.74(0.0033)
Recall	0.78(0.0021)	0.19(0.0021)	0.27(0.0015)	0.05(0.0008)	0.26(0.0017)
F	0.72(0.0011)	0.30(0.0027)	0.39(0.0016)	0.098(0.0014)	0.38(0.0021)
ER (%)	18	5.4(0.03)	5.6(0.03)	1.9(0.03)	4.2(0.03)
AMMD	0.036(0.002)	0.073(0.00)	0.049(0.00)	0.07(0.00)	0.057(0.0002)
MMD	0.068(0.0001)	0.119(0.00)	0.070(0.00)	0.122(0.0002)	0.060(0.00)
Time (seconds)	0.0046	0.0026	0.0025	0.473	0.003

Table 2. Anomaly detection result with 400 sampled points.

Methods	400 sampled points				
	AKM ² D	Random	Grid	Variance	DOE
Precision	0.75(0.009)	0.80(0.002)	0.66(0.001)	0.68(0.005)	0.76(0.0025)
Recall	0.88(0.0015)	0.23(0.0019)	0.65(0.0014)	0.05(0.0008)	0.28(0.0011)
F	0.80(0.0003)	0.35(0.0023)	0.65(0.0005)	0.101(0.0014)	0.41(0.0015)
ER (%)	18(0.03)	5.4(0.03)	14.17(0.05)	2.1(0.04)	4.7(0.03)
AMMD	0.027(0.001)	0.063(0.0002)	0.049(0.00)	0.07(0.00)	0.04(0.00)
MMD	0.056(0.00)	0.095(0.0003)	0.070(0.00)	0.11(0.0002)	0.04(0.00)
Time (seconds)	0.0046	0.0026	0.0025	0.473	0.003

value and standard deviations of the following criteria are computed over 5000 simulation replications: Precision, defined as the percentage of detected anomalies by the algorithm that are indeed anomalous; Recall, defined as the percentage of the true anomalous regions detected by the algorithm; F-measure, defined as the harmonic mean of precision and recall; Exploitation Ratio (ER), defined as the percentage ratio of sampled points in the true anomalous regions to the total number of sampled points; Anomaly Max-Min Distance (AMMD), defined as the maximum distance of points in the true anomalous region to the nearest sampled point; MMD, defined as the maximum distance of points in the entire sampling space to the nearest sampled point; and the computational time of the sampling procedure for each sampled point. Here, Precision, Recall, and F-measure are related to the accuracy of anomaly estimation, which evaluate how the algorithms locate all anomalies (e.g., exploration) and how well it quantifies each anomalous region (e.g., exploitation). AMMD, ER, and MMD are direct quantification on the performance of the sampling algorithm in terms of exploration and exploitation. For example, AMMD and ER are related to the exploitation performance of the proposed AKM²D algorithm. MMD is related to the exploration of the proposed AKM²D algorithm.

These average values and standard deviations of 250 points and 400 points are reported in Table 1 and Table 2, respectively. From these tables, it is clear that the proposed AKM²D overall outperforms other benchmark methods. For example, with 250 sampled points, the recall of AKM²D is 0.78(0.0021) indicating that 78% (with the standard deviation 0.21%) of the anomalous regions have been detected by AKM²D with only 250 points. This is much higher than the recall of benchmarks that is at most about 27%. Although benchmark methods have slightly higher precision, the overall classification accuracy, measured by F is in favor of AKM²D. The F-measure of AKM²D is around 0.72 (with standard deviation

0.0011), whereas it is at most 0.39 for benchmark methods. The MMD value of the AKM²D is only larger than the pure DOE method and smaller than all other methods. This is expected since DOE only focus on the exploration of the entire sampling space without paying attention to any focus sampling. Therefore, the AMMD values of the AKM²D are also much smaller than the benchmark methods, which indicates the proposed AKM²D achieves better-focused sampling near the anomalous regions.

Similarly, the ER of AKM²D with 250 sampled points method is around 18% (with standard deviation 3%), 3.6 times larger than that of Random, Grid, DOE (around 5%) and Variance (around 2%). This implies that the proposed method is able to quickly locate anomalous regions and sample about 3.6 times more points in those regions than benchmark methods. Note that the area of anomalous regions covers about 5.4% of the entire sampling space. However, AKM²D with around 0.6% of the full sampled points (250 sampled points out of 200 × 200), is able to detect at least 78% of the true anomalous regions. If we increase the number of sampled points to 400, this number increases to 88%, whereas for other benchmark methods it is at most 65%. The main reason for the poor performance of Random, DOE, and Variance is that they lack any abilities to focus on the discovered anomalous regions. Grid has some power of focusing on the discovered anomalous regions. The reason for the bad performance of the Variance method lies in the stability and boundary issue. We observe that the boundary is not correctly estimated with the Gaussian process and the algorithm may tend to put more points in the boundary. However, the fine sampling grid is rigid, and hence, it is not flexible enough to detect arbitrarily shaped anomalies. Although AKM²D is slightly slower than the benchmarks, all methods except Variance satisfy the real-time speed requirement for online sensing. Finally, the standard deviation of the proposed method in all these criteria is also quite small, which implies the proposed

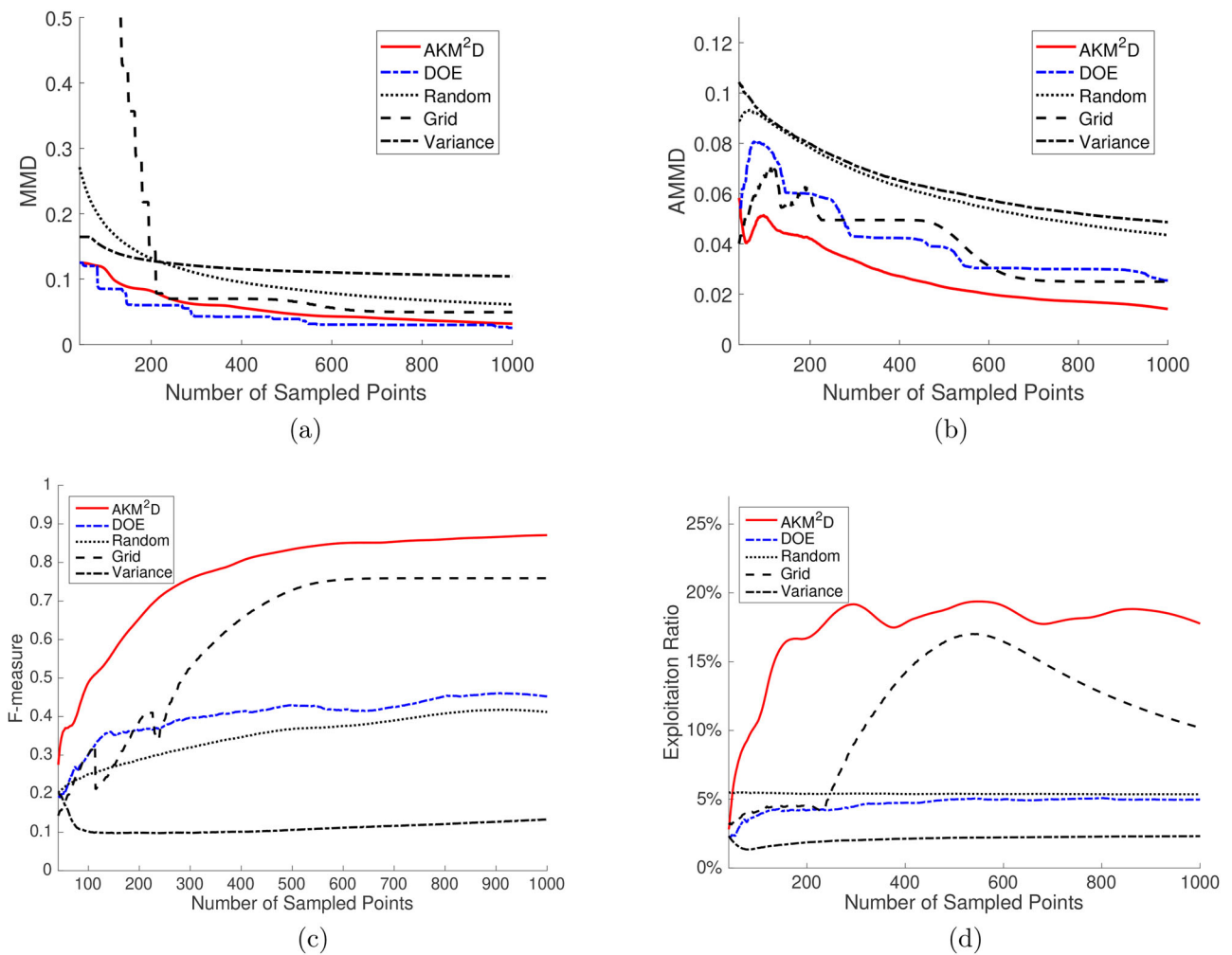


Figure 4. (a) MMD, (b) AMMD, (c) F-measure and (d) ER.

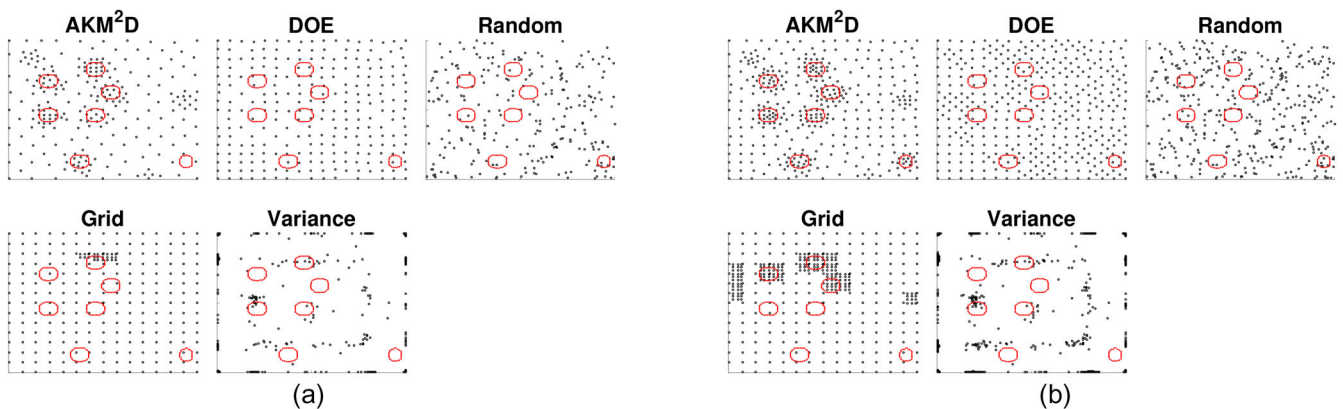


Figure 5. Sampled point pattern for all methods for (a) 250 and (b) 400 points.

methods are robust to random anomaly locations and random noises.

The average values of the MMD, AMMD, F-measure, and the ER against the iteration number (number of sampled points) are plotted in Figure 4. From this figure, we can conclude that the F-measure of AKM²D is strictly better than other benchmark methods for any number of sampled points. Furthermore, the ER of AKM²D increases to 18% with only 200 points and then oscillates around 18%, showing its superiority to quickly locate and sample the

anomalous regions. The MMD of the proposed AKM²D is better than Random, Grid, Variance, and only second to DOE. However, the AMMD of the proposed AKM²D is much better than all other benchmark methods, which demonstrates its excellent overall sampling performance. The ER of Grid stays at 4% during the coarse grid sampling and only begin to increase up to 16% when performing the fine-grid sampling (after 225 points). Finally, the ER of Random, DOE, and Variance stays lower than 6%, which is the percentage of true anomalous regions.

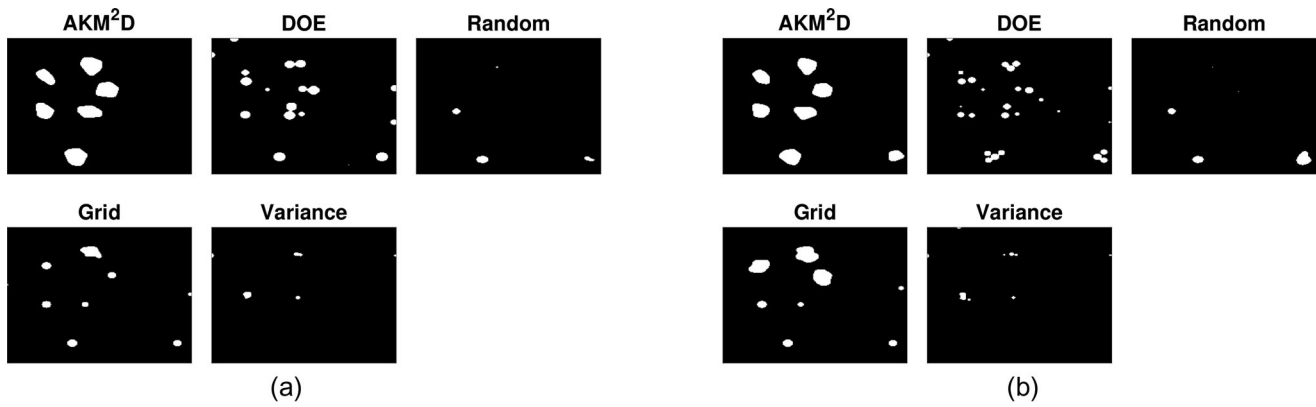


Figure 6. Anomaly estimation result for all methods for (a) 250 and (b) 400 points.

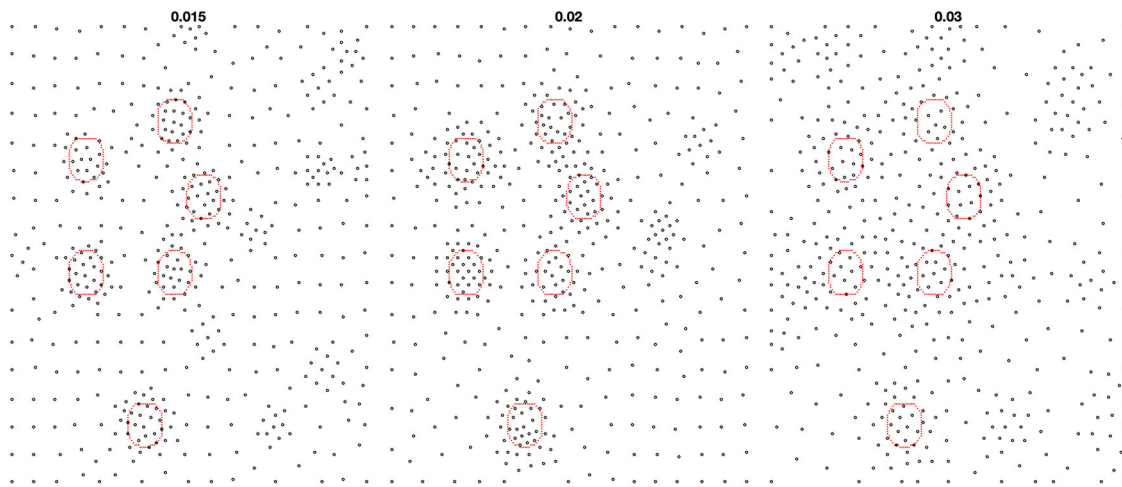


Figure 7. Effect of h (e.g., $h = 0.015, 0.02, 0.03$ from left to right).

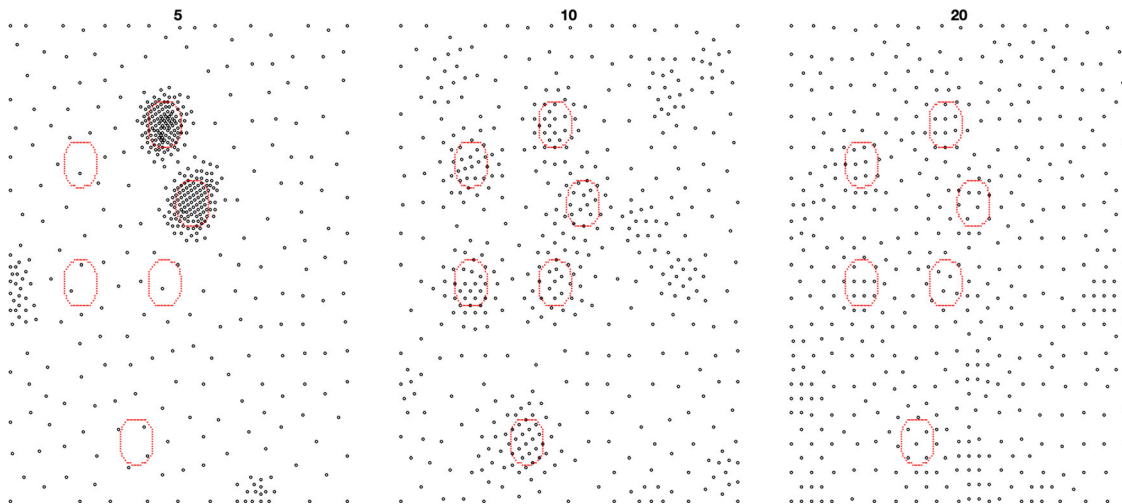


Figure 8. Effect of λ (e.g., $\lambda = 5, 10, 20$ from left to right).

Furthermore, we investigate the pattern of sampled points (with 250 and 400 points) in Figure 5. From the figure, we can observe that with only 250 sampled points, AKM²D discovers all anomalous regions one, with a better space-filling point distribution. However, Random, Variance, and DOE only put a few points on the anomalous regions, which fail to detect any of the anomalous regions and Grid can only detect one. On the other hand,

400 sampled points are enough for AKM²D to detect all seven anomalous regions. However, again Random, Variance, and DOE fail to discover any anomalous regions and Grid finishes with the fine-grid sampling of only three regions. We plot the detected anomalies corresponding to 250 and 400 sampled points in Figure 6, which again indicates the superior performance of AKM²D in anomaly detection.

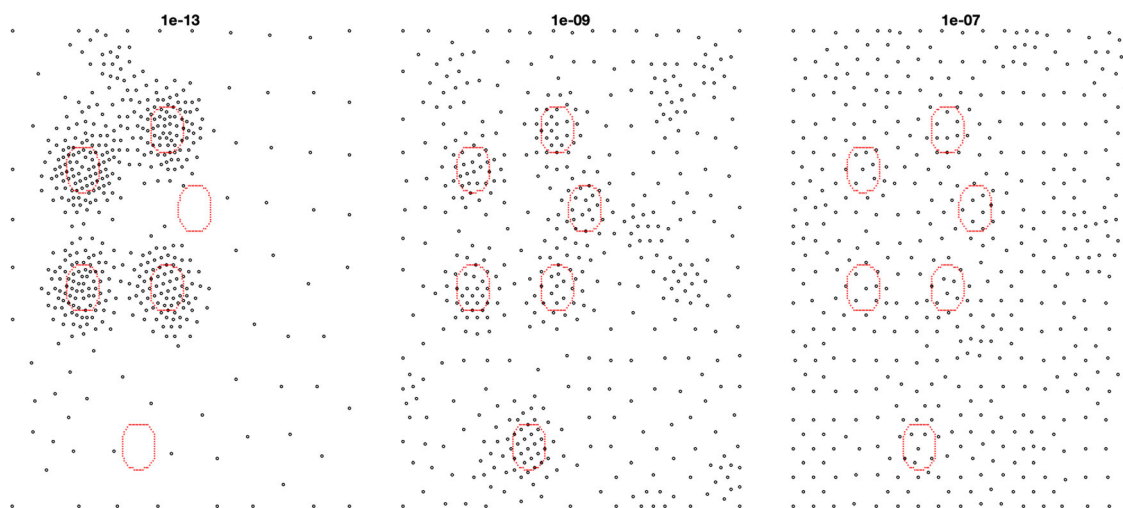


Figure 9. Effect of u (e.g., $u = 10^{-13}, 10^{-9}, 10^{-7}$, from left to right).

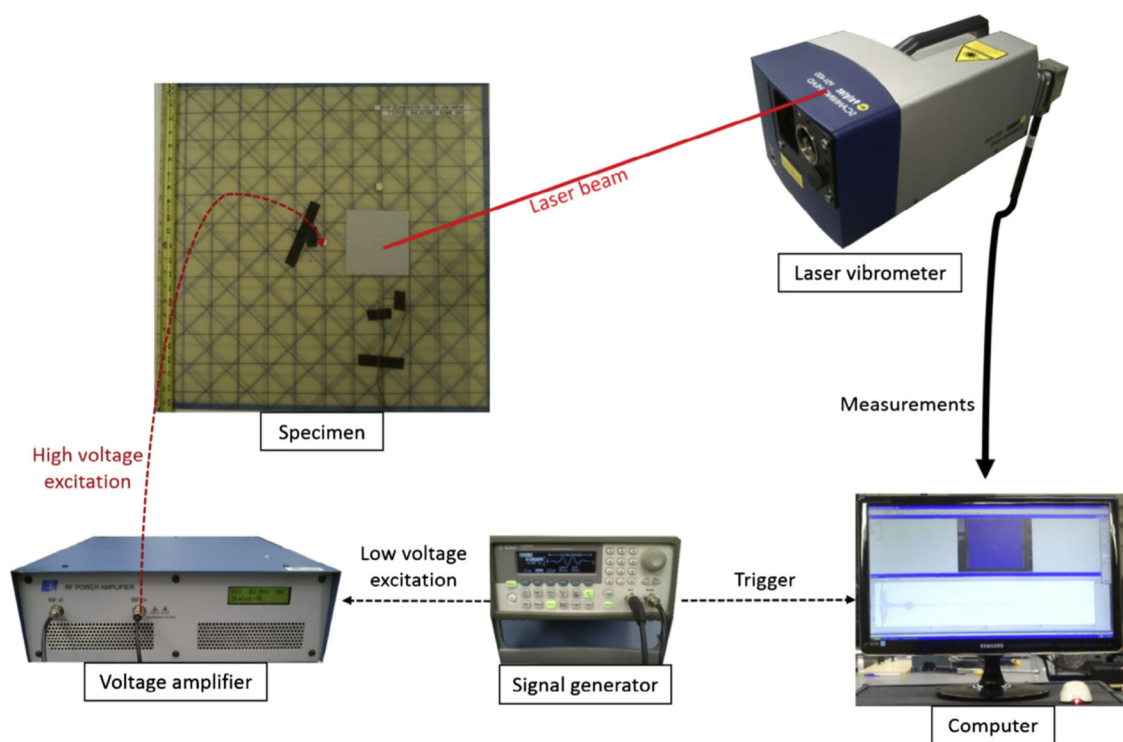


Figure 10. Guided wavefield experiment setup (Mesnil and Ruzzene, 2016).

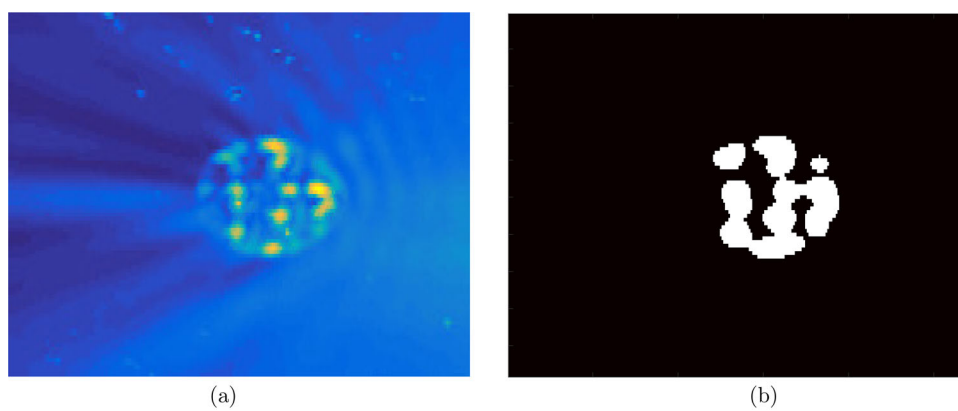


Figure 11. (a) Energy map of the entire wavefield and (b) detected anomaly.

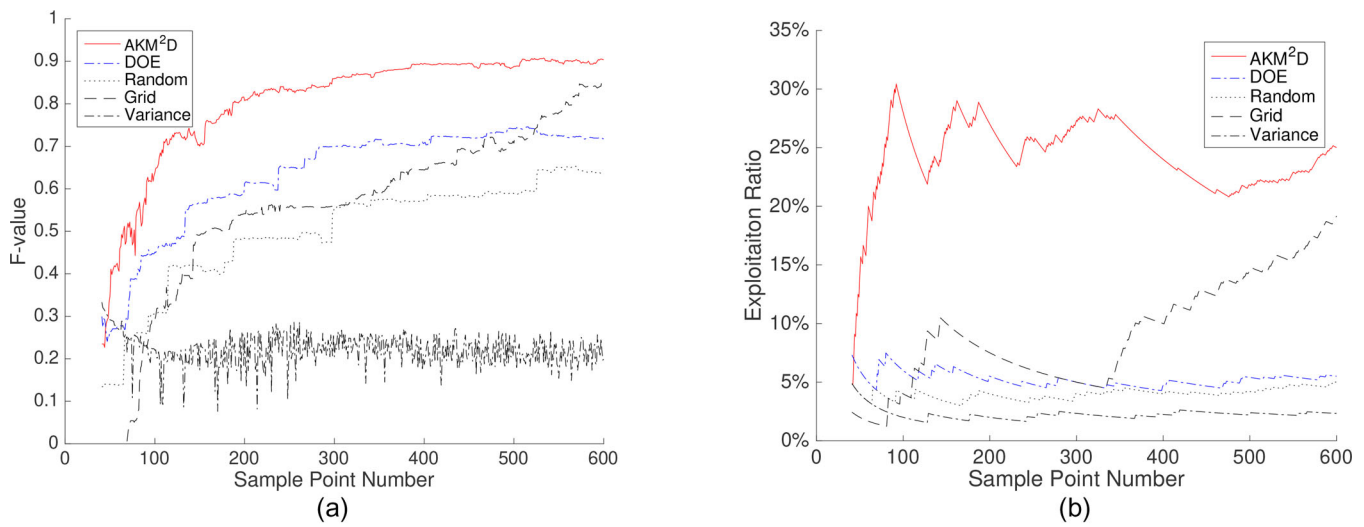


Figure 12. F-measure and (b) ER.

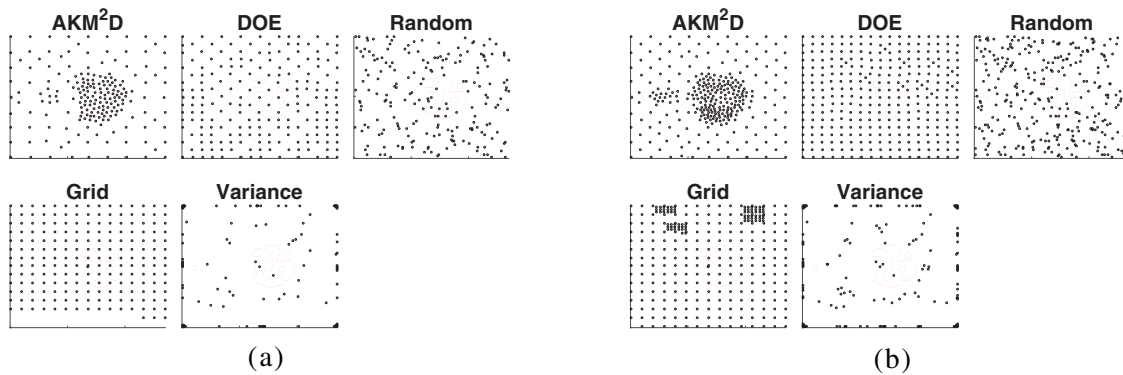


Figure 13. Sampled point pattern for all methods for (a) 200 and (b) 300 points.

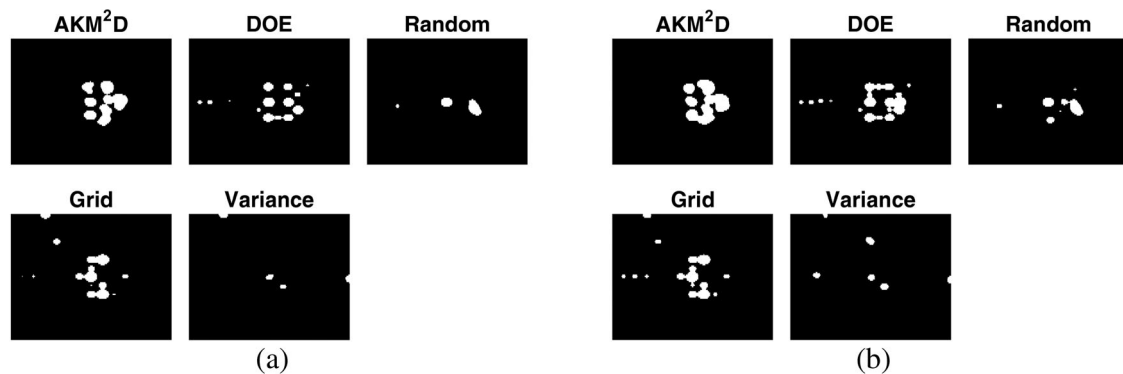


Figure 14. Anomaly estimation result for all methods for (a) 200 and (b) 300 points.

We also plot different sampling point patterns with different tuning parameters λ, u, h in Figures 7–9. We can conclude that smaller λ and u or larger h tends to lead to a better exploration of the entire background. Larger λ and u or smaller h tends to lead to better exploitation of the anomaly. Therefore, the balance of exploration and exploitation for different tasks will decide the tuning parameter that will be used in the algorithm. Furthermore, to help practitioners to understand what best tuning parameter combination is suitable for their practical need, we also perform a complete tuning sensitivity analysis in Appendix E.

6. Case study

In this section, the proposed adaptive sampling and estimation framework is applied to a real dataset in the NDE area. The case study pertains to anomaly detection in composite laminates using a GWT inspection system. Lamb wave-based inspection is one of the popular methods in NDE and structural health monitoring, due to its high sensitivity to detecting anomalies invisible to the naked eye (Mesnil and Ruzzene, 2016). However, existing GWT techniques are point-based and require the whole-field inspection of a specimen. A whole-field inspection is typically a time-consuming

process, as it requires sensing of a large number of points to avoid spatial aliasing and to achieve the desired resolution (Mesnil and Ruzzene, 2016). Therefore, it is vital to reduce the data acquisition time by reducing the number of sampled points using an adaptive sampling strategy.

The setup of our GW experiment is shown in Figure 10. A scanning laser Doppler vibrometer is employed for wavefield measurement over a grid of points with the resolution of 270×100 . It takes around 4 hours to inspect a $600 \times 600 \times 1.6$ mm composite laminate with eight layers. The specimen contains several artificial delaminations in the center, as shown in Figure 11(a), which is the energy map of the entire wavefield based on complete sampling. To speed up the GW test so that it can be used for online inspection, we reduce the number of sensing points by using adaptive sampling strategies. For comparison purposes, we show detected anomalies using a complete sampling strategy (i.e., Figure 11(a)) in Figure 11(b). The objective is to achieve a similar detection accuracy with the least number of sampled points.

We apply AKM²D as well as four other benchmark methods (i.e., Random, Grid, Variance, and DOE) for adaptive sampling and use the proposed estimation methods for anomaly detection. We compare the detection results obtained from the adaptive sampling methods with those of the complete sampling, shown in Figure 11(b), (as the ground truth), and compute the F-measure and ER profiles depicted in Figure 12. We can observe that with only 300 points (1.1% of complete sensing) AKM²D is able to achieve an F-measure of 0.8, much higher than those of other benchmark methods, which is at most 0.5.

The pattern of sampled points and detected anomalous regions by using 200 and 300 points are also shown in Figures 13 and 14, respectively. From these figures, it is clear that the irregular anomalous regions can be fully explored by the proposed AKM²D with only 200 sampled points (0.7% of full sampling), which can reduce the measurement time from 4 hours to 2 minutes. However, using other methods, very few sampled points are selected in the anomalous regions.

7. Conclusion

Adaptive sampling for clustered anomaly detection is vital in scaling up point-based inspection systems. In this article, we proposed a novel methodology for real-time adaptive sampling and anomaly detection in large sampling spaces. In our methodology, we first developed an adaptive sampling framework, namely the AKM²D, by optimizing a composite index. We also studied the sampling properties and showed that the proposed method is able to balance sampling between the exploration of the entire space and the focused sampling near anomalies. We developed efficient and recursive algorithms to determine the location of the next sampled point by solving the optimization problem in real time. Then, we proposed robust kernel regression and sparse kernel regression to update the estimates of the functional mean and the anomalous regions after a new sample is collected. In the simulation study, we showed that the

proposed AKM²D outperformed existing adaptive sampling approaches, which fail to locate and focus on anomalous regions. Finally, the proposed method was applied to a real case study on the anomaly detection of composite laminates via a GWT. We showed that our method can achieve a similar detection accuracy to that of the complete sampling by sensing only 0.7% of the sampled points, and hence, it can significantly reduce the inspection time.

There are several potential research directions to be investigated. One possible direction is to extend this method to batch sampling, in which multiple sampled points can be selected simultaneously by the algorithm. The other direction is to extend this into a higher-dimensional sampling space (e.g., larger than 2D) by more efficient optimization techniques than the grid-based methods.

Notes on contributors

Hao Yan received a B.S. degree in physics from Peking University, Beijing, China, in 2011. He also received an M.S. degree in Statistics, an M.S. degree in computational science and engineering, and a Ph.D. degree in industrial engineering from Georgia Institute of Technology, Atlanta, in 2015, 2016, 2017, respectively. Currently, he is an assistant professor in the School of Computing, Informatics, & Decision Systems Engineering at Arizona State University. His research interests focus on developing scalable statistical learning algorithms for large-scale high-dimensional data with complex heterogeneous structures to extract useful information for the purpose of system performance assessment, anomaly detection, intelligent sampling and decision making. He is a member of INFORMS and IISE.

Kamran Paynabar received his B.Sc. and M.Sc. in industrial engineering from Iran University of Science and Technology and Azad University in 2002 and 2004, respectively, and his Ph.D. in industrial and operations engineering from The University of Michigan in 2012. He also holds an M.A. in statistics from The University of Michigan. Currently, he is a Fouts Family Career Professor and associate professor at the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta. His research interests include data fusion for multi-stream waveform signals and functional data, engineering-driven statistical modeling, sensor selection in distributed sensing networks, probabilistic graphical models, and statistical learning with applications in manufacturing and healthcare systems. Dr. Paynabar is a member of Institute of Industrial and Systems Engineers The Institute for Operations Research and the Management Sciences and IEEE Robotics and Automation Society.

Jianjun Shi received B.S. and M.S. degrees in electrical engineering from Beijing Institute of Technology, Beijing, China, in 1984 and 1987, respectively, and a Ph.D. degree in mechanical engineering from the University of Michigan, Ann Arbor, in 1992. He is currently the Carolyn J. Stewart Chair Professor with the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta. His research interests include the fusion of advanced statistical and domain knowledge to develop methodologies for modeling, monitoring, diagnosis, and control for manufacturing systems. Dr. Shi is a Fellow of the Institute of Industrial and Systems Engineering, a Fellow of American Society of Mechanical Engineering, a Fellow of the Institute of Operations Research and the Management Sciences, an Academician of the International Academy for Quality, and a member of National Academy of Engineering.

ORCID

Hao Yan  <http://orcid.org/0000-0002-4322-7323>

References

- Aussel, J.D. and Monchalin, J.-P. (1989) Precision laser-ultrasonic velocity measurement and elastic constant determination. *Ultrasonics*, **27**(3), 165–177.
- Bernardo, M.C., Buck, R., Liu, L., Nazaret, W., Sacks, J., Welch, W.J. et al. (1992) Integrated circuit design optimization using a sequential strategy. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, **11**(3), 361–372.
- De Aguiar, P.F., Bourguignon, B., Khots, M.S., Massart, D.L. and Phan-Thau-Luu, R. (1995) D-optimal designs. *Chemometrics and Intelligent Laboratory Systems*, **30**(2), 199–210.
- De Boor, C. (1972) On calculating with b-splines. *Journal of Approximation Theory*, **6**(1), 50–62.
- Devroye, L. and Krzyżak, A. (1999) On the Hilbert kernel density estimate. *Statistics & Probability Letters*, **44**(3), 299–308.
- Gibson, I., Rosen, D.W. and Stucker, B. (2014) *Additive Manufacturing Technologies*, (Vol. 17). Springer, New York.
- Halton, J.H. (1964) Algorithm 247: Radical-inverse quasi-random point sequence. *Communications of the ACM*, **7**(12), 701–702.
- Jin, R., Chang, C.J. and Shi, J. (2012) Sequential measurement strategy for wafer geometric profile estimation. *IIE Transactions*, **44**(1), 1–12.
- Johnson, M.E., Moore, L.M. and Ylvisaker, D. (1990) Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, **26**(2), 131–148.
- Jones, D.R., Schonlau, M. and Welch, W.J. (1998) Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, **13**(4), 455–492.
- Joseph, V.R. and Hung, Y. (2008) Orthogonal-maximin Latin hypercube designs. *Statistica Sinica*, **18**(1), 171.
- Joseph, V.R., Dasgupta, T., Tuo, R. and Wu, C.F.J. (2015) Sequential exploration of complex surfaces using minimum energy designs. *Technometrics*, **57**(1), 64–74.
- Kyriakidis, P.C. (2005) Sequential spatial simulation using Latin hypercube sampling, in *Proceedings of the Conference on Geostatistics*, pp. 65–74, Springer, Dordrecht.
- Li, Z. and Qiu, P. (2014) Statistical process control using a dynamic sampling scheme. *Technometrics*, **56**(3), 325–335.
- Liu, K., Mei, Y. and Shi, J. (2015) An adaptive sampling strategy for online high-dimensional process monitoring. *Technometrics*, **57**(3), 305–319.
- Loeppky, J.L., Sacks, J. and Welch, W.J. (2009) Choosing the sample size of a computer experiment: A practical guide. *Technometrics*, **51**(4), 366–376.
- Loeppky, J.L., Moore, L.M. and Williams, B.J. (2010) Batch sequential designs for computer experiments. *Journal of Statistical Planning and Inference*, **140**(6), 1452–1464.
- Mak, S. and Joseph, V.R. (2018) Minimax and minimax projection designs using clustering. *Journal of Computational and Graphical Statistics*, **27**(1), 166–178.
- Mateos, G. and Giannakis, G.B. (2012) Robust nonparametric regression via sparsity control with application to load curve data cleansing. *IEEE Transactions on Signal Processing*, **60**(4), 1571–1584.
- McKay, M.D., Beckman, R.J. and Conover, W.J. (1979) Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**(2), 239–245.
- Mesnil, O. and Ruzzene, M. (2016) Sparse wavefield reconstruction and source detection using compressed sensing. *Ultrasonics*, **67**(2016), 94–104.
- Mesnil, O., Yan, H., Ruzzene, M., Paynabar, K. and Shi, J. (2014) Frequency domain instantaneous wavenumber estimation for damage quantification in layered plate structures, in *Proceedings of the EWSHM-7th European Workshop on Structural Health Monitoring*. Université de Nantes, Jul 2014, IFFSTTAR, Inria, pp. 2338–2345.
- Owen, A.B. (1994) Controlling correlations in Latin hypercube samples. *Journal of the American Statistical Association*, **89**(428), 1517–1522.
- Ranjan, P., Bingham, D. and Michailidis, G. (2008) Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, **50**(4), 527–541.
- Schölkopf, B., Herbrich, R. and Smola, A.J. (2001) A generalized representer theorem, in *Proceedings of the International Conference on Computational Learning Theory*, Springer, Berlin, Heidelberg.
- Simpson, J.A. (1992) Mechanical measurement and manufacturing. *Control and Dynamic Systems: Advances in Theory and Applications*, **45**, 17–30.
- Sobol, I.M. (1976) Uniformly distributed sequences with an additional uniform property. *USSR Computational Mathematics and Mathematical Physics*, **16**(5), 236–242.
- Sobol, I.M. (1967) The distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, **7**(4), 784–802.
- Stein, M. (1987) Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, **29**(2), 143–151.
- Stinstra, E., den Hertog, D., Stehouwer, P. and Vestjens, A. (2003) Constrained maximin designs for computer experiments. *Technometrics*, **45**(4), 340–346.
- Wang, A., Xian, X., Tsung, F. and Liu, K. (2018) A spatial-adaptive sampling procedure for online monitoring of big data streams. *Journal of Quality Technology*, **50**(4), 329–343.
- Wang, H., Yao, K., Pottier, G. and Estrin, D. (2004) Entropy-based sensor selection heuristic for target localization, in *Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks*. ACM, Berkeley, CA, USA.
- Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P., Mitchell, T.J. and Morris, M.D. (1992) Screening, predicting, and computer experiments. *Technometrics*, **34**(1), 15–25.
- Wu, D., Salerno, A., Malter, U., Aoki, R., Kochendörfer, R., Kächele, P.K., Woiße, K., Pfister, K. and Busse, G. (1996) Inspection of aircraft structural components using lockin-thermography. *Quantitative Infrared Thermography*, **96**, 251–256.
- Xiong, F., Xiong, Y., Chen, W. and Yang, S. (2009) Optimizing latin hypercube design for sequential sampling of computer experiments. *Engineering Optimization*, **41**(8), 793–810.
- Yan, H., Paynabar, K. and Shi, J. (2015) Anomaly detection in images with smooth background via smooth-sparse decomposition. *Technometrics*, **59**(1), 102–114.
- Yan, H., Paynabar, K. and Shi, J. (2018) Real-time monitoring of high-dimensional functional data streams via spatio-temporal smooth sparse decomposition. *Technometrics*, **60**(2), 181–197.
- Ye, K.Q. (1998) Orthogonal column Latin hypercubes and their application in computer experiments. *Journal of the American Statistical Association*, **93**(444), 1430–1439.
- Zhu, J., Hoi, S.C.H. and Lyu, M.R.T. (2008) Robust regularized kernel regression. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, **38**(6), 1639–1644.
- Zou, C. and Qiu, P. (2009) Multivariate statistical process control using lasso. *Journal of the American Statistical Association*, **104**(488), 1586–1596.