



Combining point and distributed strain sensor for complementary data-fusion: A multi-fidelity approach

Seung-Seop Jin ^{a,*}, Sung Tae Kim ^b, Young-Hwan Park ^b

^a Sustainable Infrastructure Research Center, KICT, Goyang, Republic of Korea

^b Department of Infrastructure Safety Research, KICT, Goyang, Republic of Korea



ARTICLE INFO

Article history:

Received 31 July 2020

Received in revised form 2 January 2021

Accepted 3 February 2021

Keywords:

Complementary data-fusion

Structural monitoring

Point/distributed strain sensor

Gaussian process regression

Gaussian process mapping

ABSTRACT

The aim of this paper is to develop a complementary data-fusion algorithm for monitoring of the strain over whole structural system. Point strain sensors measure accurate strains (high-accuracy) at discrete measurement positions (low-spatial resolution), while a distributed strain sensor enables a quasi-continuous distributed measurement (high-spatial resolution) with less accurate strains (low-accuracy). This study firstly investigates the complementary data-fusion for the point and distributed strain sensor to combine their advantages in order to obtain the accurate strain distribution (i.e., high-accuracy and high-spatial resolution). For this purpose, a traditional multi-fidelity data-fusion framework can be applied based on Gaussian process regression in the form of auto-regressive scheme. However, the traditional method is not efficient to learn complex cross-correlations. To enable flexible learning for the complex cross-correlations, this study introduces a multi-fidelity data-fusion framework using the input-connected Gaussian process mapping. The numerical and experimental studies demonstrate that the proposed method outperforms the traditional methods with more accurate predictions by using fewer samples for the high-fidelity data. In this context, the proposed method has potential for monitoring the strain distribution over whole structural systems under limited budgets.

© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rapid advances in sensor and information technologies have led to structural monitoring in civil and mechanical structural systems. The structural monitoring system plays a key role in public safety to provide reliable information regarding the structural integrity [1]. Various sensors (e.g., strain gauge, accelerometer, and optical fiber) are embedded in structural systems like bridges to enable condition-based monitoring [2–4]. Since strain is sensitive to a local behavior of the structural systems, it has been widely investigated for monitoring the structural integrity [5–12]. Through accurate interpretation of the measured strain, condition-based monitoring can provide an early warning of an unsafe condition and enable proactive maintenance planning. Therefore, it can greatly reduce the inspection burden as well as maintenance costs.

The prerequisite for the structural monitoring is to obtain accurate measurement through the whole structural system. A traditional strain sensors only measure strains at discrete measurement positions (i.e., point sensors) [8,9,13], so that the number of sensors and their positions are critical considerations in the design of the overall system and total costs. In addition, the measurement positions for the critical locations are not previously known for detection of the warning events (e.g.,

* Corresponding author.

E-mail address: seungsab@kict.re.kr (S.-S. Jin).

crack formulation or severe damage after earthquake). If strain sensors are incorrectly positioned, a structural monitoring system results in completely useless measurements for structural integrity [5]. Therefore, the point strain sensors are carefully positioned as many as possible. Under insufficient budgets, it is challenging to install the point sensor densely over whole structural system.

A recent development of the optical fiber sensor technology enables a quasi-continuous distributed measurement for the strain along the optical fibers [14–17]. This type of sensor is referred to as a distributed strain sensor, and it is considered as a promising sensing technology to overcome the challenging issues of the point strain sensors [6,7]. Although the distributed strain sensor provides a quasi-continuous distributed measurement, the measured strains are relatively inaccurate than those from point strain sensors (e.g., strain gauge or Fiber Bragg Grating (FBG) sensor) due to a scattering-based sensing mechanism. If the measured strains along the optical fibers are inaccurate, structural monitoring systems may become unreliable and lead to misleading decision-making for structural integrity and maintenance planning.

This study firstly investigates the complementary data-fusion to fuse the advantages of the point and distributed strain sensor. Point strain sensors measure accurate high-fidelity strains at discrete measurement positions (a few samples of accurate high-fidelity data), while a distributed strain sensor measures distributed strains along the optical fiber with less accurate low-fidelity strains (a large number of inaccurate low-fidelity data). The goal of this complementary data-fusion is to obtain the distributed strains with high accuracy by combining their advantages.

Numerous studies related to data-fusion have investigated for structural health monitoring (SHM) in both temporal and spatial sense [18–29]. The recent overview of the data-fusion in SHM can be found in Wu and Jahanshahi [27]. Table 1 chronologically shows the recent studies on data-fusion for structural response estimation. The data-fusion for temporal responses estimation are prevalent with state-estimation methods (Kalman-filter or weighting method) using heterogeneous data (different response such as accelerometer and displacement) [18–20,22,23,28].

A few studies have investigated the data-fusion for spatial response estimation [21,24–26,29]. Some works have integrated the homogeneous data (strain) from the different sources (image and optical fiber sensor) by visual comparison to evaluate the cracking behavior of the concrete specimen and structure [25,29]. He, Zhang and Xu [26] have improved the multiscale response reconstruction based on Kalman-filter with revised observation equation. This method fuses the heterogeneous data (e.g., strain, velocity and displacement) to estimate unmeasured structural responses. Although it is successful to verify their method numerically, there are critical limitations for the practicability (e.g., collocated response measurements at unknown input locations). Recent works have developed unidirectional strain maps from the additive strain signal from soft elastomeric capacitor (SEC) based on electrical resistance strain gauge (ERS) [21,24]. The iterative signal fusion method have been developed to extract unidirectional strain maps from the additive strain (i.e., summation of the orthogonal strains, $\varepsilon_x + \varepsilon_y$) from the SEC based on unidirectional strains from the ERS. The iterative signal fusion method is an iterative correction method to minimize the discrepancy between predictions from three Krigings of the SEC and ERS. The rationale behind this method is that different sources are equally significant. In these studies, the measured responses from different sources are of equal importance (i.e., not hierarchical importance as in multi-fidelity data). In this context, the multi-fidelity data-fusion is somewhat new for the structural response estimation in a complementary manner.

Table 1
Summary of literatures on data-fusion for structural response estimation.

Year	Literature	Purpose	Methodology	Data type (Response)
2007	Smyth and Wu [18]	Temporal fusion	Kalman filter	Heterogeneous data(acceleration and displacement)
2016	Park, Lee, Sim, Jung and Spencer Jr. [20]	Temporal fusion	Weighting approach	Heterogeneous data(acceleration and strain)
2015	Chatzi and Fuggini [19]	Temporal fusion	Kalman filter	Heterogeneous data(acceleration and displacement)
2018	Ravizza, Ferrari, Rizzi and Chatzi [23]	Temporal fusion	Weighting approach	Heterogeneous data(acceleration and displacement)
2018	Park, Moon, Yoon, Gomez, Spencer Jr. and Kim [22]	Temporal fusion	Weighting approach	Heterogeneous data(acceleration and displacement)
2018	Downey, Sadoughi, Laflamme and Hu [21]	Spatial fusion	Additive strain map adjustment algorithm via Kriging (Gaussian process)	Homogeneous data (strain map via SEC ¹⁾)
2018	Sadoughi, Downey, Yan, Hu and Laflamme [24]			Homogeneous data (strain map via SEC + strain from ERS ²⁾)
2019	He, Zhang and Xu [26]	Spatial fusion	Kalman filter	Heterogeneous data(strain, velocity and displacement)
2019	Bado, Kaklauskas and Casas [25]	Spatial fusion	Visual comparison	Homogeneous data (internal and external strains via DOFS ³⁾ and DIC ⁴⁾)
2020	Bado, Casas, Dey and Berrocal [29]			Homogeneous data (internal and external strains via DOFS ³⁾ and DIC ⁴⁾)
2020	Zhu, Gao, Xia, Gao, Weng, Sun and Hu [28]	Temporal fusion	Kalman filter	Heterogeneous data(acceleration and strain)

¹⁾ SEC: soft elastomeric capacitor.

²⁾ ERS: electrical resistance strain gauge.

³⁾ DOFS: distributed optical fiber sensing.

⁴⁾ DIC: digital image correlation.

In optical sensing community, there were several studies to obtain the information from point (i.e., FBG) and distributed sensor (i.e., BOTDA/R) on a single fiber. Li, Xu and Kishida [30] proposed the FBG-indicator to precisely determine the position of the strain events. The hybrid sensing systems of BOTDA/R and FBG were proposed to measure the distributed strain by BOTDA/R and dynamic strain/temperature by FBG respectively [10,31]. However, these studies have not investigated the complimentary data-fusion for combining the point and distributed strains.

For the complementary data-fusion, a multi-fidelity data-fusion framework is utilized based on Gaussian process (GP) regression [32] in this study. The strain distribution with high-accuracy is obtained through the following multi-fidelity framework: (1) the GP regression extracts a less accurate strain distribution from low-fidelity strain data (distributed sensor); and (2) the multi-fidelity method improves the accuracy of the extracted strain distribution using a few samples of high-fidelity strain data (point sensor). For this purpose, a traditional multi-fidelity data-fusion method can be applied in the form of the linear auto-regressive scheme. However, the traditional multi-fidelity data-fusion method is not efficient to learn complex cross-correlations due to its simple linear auto-regressive scheme.

To address the limitation of the traditional multi-fidelity data-fusion method, this study introduces a novel multi-fidelity data-fusion method. The proposed method trains correlated surrogates on measured strains using GP regression [32,33] with input-connected GP mapping [34,35] in the form of auto-regressive scheme [36–39]. In order to learn complex cross-correlations between low-fidelity and high-fidelity strain data, the proposed method introduces a mapping function as a GP regression in terms of the low-fidelity prediction and corresponding input. By doing that, the proposed method can deal with various cross-correlations from simple to complex ones. To demonstrate the feasibility of the proposed methods, the proposed method was validated with the traditional multi-fidelity data-fusion method through numerical and experimental studies. The results show that the proposed method outperforms the traditional multi-fidelity method with more accurate predictions using fewer high-fidelity data.

The main contributions of this study are summarized as: (1) this study firstly proposes the multi-fidelity data-fusion method of combining the information from the point and distributed strain sensor; (2) the input-connected GP mapping is introduced for the complementary data-fusion to learn complex cross-correlations between the point and distributed strain sensors; and (3) through Monte Carlo simulation for the numerical and experimental study, the proposed method is validated by investigating the influence on the number of the high-fidelity data and their locations.

This study is organized as follows: Section 2 provides backgrounds related to our work. In Section 3, the proposed method is described. To demonstrate the flexible learning capability of the proposed method (MF_{GP}), four methods were evaluated numerically and experimentally: (1) GP regression from the low-fidelity data (LF); (2) GP regression from the high-fidelity data (HF); (3) multi-fidelity method using Hierarchical Kriging (MF_{HKRG}); and (4) multi-fidelity method using the simple linear auto-regressive model (MF_{KO}). In Section 4, numerical experiments were investigated to validate the feasibility of the MF_{GP} . Section 5 presents the experimental validation of the complementary data-fusion using the point and distributed strain sensor. In Section 6, we provide discussions on the MF_{GP} and practical issues for massive infrastructures. Lastly, we conclude with summary and remarks in Section 7.

2. Methodology

In this section, the background and materials are introduced. Firstly, the point and distributed strain sensor are presented with their pros and cons. Then, the idea of fusing their advantages by the complementary data-fusion is proposed. Since GP regression is used for the complementary data-fusion, the standard approach to modeling with GP regression is given. Lastly, the multi-fidelity data-fusion with simple linear auto-regression scheme is presented as the traditional method for the complementary data-fusion.

2.1. Point and distributed strain sensor and their potential information fusion

Among various responses for the structural monitoring, strain has been widely used for detecting local damage and fatigue [5–7,9–13]. A traditional point strain sensor is a strain gauge with a metallic foil pattern. The metallic foil pattern is supported by insulating flexible backing. Once the structural system is deformed, the metallic foil pattern changes and this results in change of the electrical resistance. This change is measured using a Wheatstone bridge and converted to strain by gauge factor.

Optical fiber-based measurement systems for strain uses characteristics of the light traveling within the optical fiber [40]. The measurement system transmits light into an optical fiber from a tunable frequency laser. Similarly to strain gauge, the deformation of the structural system changes the characteristics of the light. This changes are detected by reflected or back-scattered light, and they are converted into strain. Fiber Bragg grating (FBG) sensor is a single-point strain sensor of the optical fiber. FBG is made by writing Bragg grating with a specific grating space into the optical fiber. This Bragg grating generates a reflection centered on a wavelength that depends on the strain and temperature of the optical fiber. Once the optical fiber is deformed, it changes the grating space (change of the reflective index). Then this changes the characteristics of the reflected lights (i.e., Bragg wavelength shift), and strain is converted from the Bragg wavelength shift.

Strain gauge and FBG sensors provide accurate strains with a high signal-to-noise ratio (SNR). However, they only measure strains at discrete measurement positions (point strain sensor). Some important challenges regarding point strain

sensors remain as: (1) the number of sensors and their positions should be carefully chosen for the design of overall monitoring system and total costs; (2) they still lack the scalability for sensing new points rapidly (i.e., discrete sensing mechanism); and (3) it may fail to capture the important local changes (e.g., crack formulation after earthquake), if point sensors are incorrectly positioned. Consequently, it is challenging to install the point sensor in a dense array under insufficient budgets.

The challenging issues of the point sensors can be addressed using a distributed optical fiber sensor (DOFS). The distributed strain sensor enables a quasi-continuous distributed strains along the optical fibers from back-scattering of the light [41]. When an incident light is launched, this input light interacts with the optical fiber (i.e., optical medium). This interaction generates some phonons that return toward the light source (back-scattering). The back-scattering lights are generated along the optical fiber, and they contain information about the fiber properties (strain and temperature along the optical fiber). Three different scattering processes exist: Rayleigh, Raman and Brillouin scattering. While Raman scattering is only dependent to the temperature of the optical fiber, Rayleigh and Brillouin scattering relates with strain and temperature [42].

Table 2 summarized developments of the DOFS based on the recent literatures [42–45]. The Rayleigh-based DOFS can be categorized into the R-OTDR (Optical Frequency Domain Reflectometry) and R-OFDR (Optical Frequency Domain Reflectometry). The R-OTDR measures the attenuation of the Rayleigh backscattering of the time after an incident pulse. Based on the change in the attenuation, the strain and temperature are measured. The R-OTDR measures the strain and temperature over the long sensing distance (kilometer-metric) with the coarse spatial resolution (meter-metric). Various sensing techniques for the R-OTDR have been developed to improve measurement performances (i.e., spatial resolution).

The R-OFDR measures Rayleigh backscattering as distributed reflectivity pattern along the optical fiber [45]. The R-OFDR measures two signals from the reference and operating condition in the optical frequency domain. Then, it employs a Fast Fourier transform (FFT) for two signals, and divides the transformed into small segments using a sliding window. Based on the small segments, an inverse FFT (iFFT) was performed with zero-padding. The optical frequency shifts between the two signals via iFFT are used to estimate the strain or temperature variation. Contrary to the R-OTDR, the R-OFDR systems (also known as Rayleigh-OBR) provides the high-spatial resolution up to 1 mm. Due to the nonlinear phase noise in OFDR, the maximal sensing distance of the R-OFDR is limited to the short distance (hundred meter-metric). The advanced sensing techniques enable the longer sensing distance (up to kilometer-metric in laboratory) [45]. However, the longer sensing distance based on advanced sensing techniques (i.e., Fractional Fourier transform technique [46]) requires the cost of spatial resolution, accuracy and sensitivity [42,47,48].

Brillouin-based DOFS systems have been introduced for the monitoring of the strain over whole structural systems [6,7,17,41,42,44,57]. Brillouin scattering is generated by acoustic wave stimulated in the optical fiber. This acoustic wave produces Brillouin scattering wave by interacting with the input light (i.e., inelastic interaction). As a result, a frequency shift between the input light and Brillouin scattering wave occurs. This frequency shift is related to strain and temperature at the local region of the optical fiber (related to the spatial resolution), so that it is possible to measure strain and temperature distribution based on Brillouin scattering wave. In addition, any single-mode optical fibers of silica glass can be used as sensing unit and these optical fibers are very cheap, so that it is possible to embed the optical fiber over whole structural system. Soga and Luo [44] discussed the opportunities of the Brillouin-based DOFS for infrastructure monitoring (i.e., long sensing range for massive structure) to monitoring strains inside concrete and steel structure, tunnel deformation, railway monitoring and landslide and slope monitoring. Barrias, Casas and Villalba [42] mentioned that the Brillouin-based DOFS systems are very useful and potential for applications on massive infrastructures due to their extended sensing ranges and advancement of the sensing techniques for the better spatial resolution.

To improve spatial resolution of the quasi-continuous measurement, different techniques have been investigated, such as Brillouin optical time domain analysis/reflectometer (BOTDA/R) [6,7,10,42] and Brillouin optical correlation domain analysis (BOCDA) [15–17,41]. The sensing mechanism of the distributed sensors is inherently based on the back-scattering of the light, so that their SNR is relatively lower than those of the point strain sensors (e.g., strain gauge or Fiber Bragg Grating sen-

Table 2
Performance summary on distributed optical fiber sensor for strain.

Type	Performance Specification	Advanced sensing technique
Rayleigh-OTDR (R-OTDR)	<ul style="list-style-type: none"> • - Sensing Distance: 10–50 km • - Spatial Resolution: 2–10 m • - Open loop control system 	<ul style="list-style-type: none"> - Polarization-OTDR [49] - Phase-sensitive OTDR [50,51]
Rayleigh-OFDR (R-OFDR)	<ul style="list-style-type: none"> • - Sensing Distance: 50–200 m • - Spatial Resolution: ≈ 1 mm • - Open loop control system 	<ul style="list-style-type: none"> - OFDR using narrow linewidth laser [52] - Fractional FT-OFDR [46]
Brillouin-OTDR (BOTDR)	<ul style="list-style-type: none"> • - Sensing Distance: 20–50 km • - Spatial Resolution: 1 m • - Open loop control system 	<ul style="list-style-type: none"> - STFT-BOTDR [53] - FFT-BOTDR [54]
Brillouin-OTDA (BOTDA)	<ul style="list-style-type: none"> • - Sensing Distance: 100–200 km • - Spatial Resolution: 0.02–5 m • - Closed loop control system 	<ul style="list-style-type: none"> - PPP-BOTDA [55] - DPP-BOTDA [56]
Brillouin-OCDA (BOCDA)	<ul style="list-style-type: none"> • - Sensing Distance: 2 km • - Spatial Resolution: 0.05–1 m • - Closed loop control system 	<ul style="list-style-type: none"> - BOCDA using differential measurement [15] - Phase-modulated BOCDA [16]

sor). In addition, the measured strain is averaged over the spatial resolution of the Brillouin-scattering based distributed sensors (i.e., 2 cm–100 cm). The definition of the spatial resolution is illustrated in Fig. 1.

For any DOFS systems, there are several measurement parameters to determine the measurement performance. These measurement parameters include spatial resolution, sensitivity, dynamic range, system bandwidth and so on [58]. It is worth noting that the optimal measurement performances in all DOFS systems cannot be achieved simultaneously without losing one of the measurement performances. Stated differently, there is a strong trade-off among these parameters for all DOFS systems [14,44,58–60]. One of the prevailing trade-offs among these parameters is “spatial resolution versus sensitivity” [58–60]. For example, with the DOFS based on Rayleigh backscattering systems (i.e., Luna Technologies OBR 4600), it can measure strain over very high-spatial resolution (1 mm) up to 70 m of the optical fiber; but, the measured strain is less accurate than over the low-spatial resolution (100 mm) [60]. In the DOFS based on Brillouin scattering, there is a trade-off between spatial resolution and SNR of the Brillouin gain spectrum (sensitivity) [59]. Since the DOFS systems (based on linear/non-linear backscattering or non-linear forward scattering) have their own advantages and disadvantages, the parameters in DOFS systems should be optimized based on the trade-offs (e.g., “spatial resolution vs sensitivity” and “spatial resolution vs maximal sensing distance”) to match the conditions for application areas. For all DOFS systems, the measurements are performed in a repeatable manner to obtain one profile of the strains. It is obvious that more average number improves the accuracy of strain. However, the larger average number results in longer measurement time. For massive infrastructures, the average number may be limited to afford measurement time for capturing the event of interest (i.e., loading test in bridge). Therefore, the compromise between accuracy and average number should be made carefully under allowable time. In this context, the trade-off between the measurement performances induces potential errors in the measurement for all DOFS systems. However, it is not easy to set optimal measurement parameters for application areas due to the difference conditions from laboratory (i.e., operating temperatures for devices and reference optical fiber).

When using the DOFS systems on infrastructures (i.e., concrete structures), the optical fibers are embedded to the structural system and they are typically exposed to harsh conditions. For installation and operation, the optical fibers are easily exposed to the risk of being damaged due to fragile nature of the optical fiber. When the optical fiber is damaged, it is necessary to replace the entire optical fiber. It is very inconvenient and costly, if the optical fiber was installed for very long distance (over the massive infrastructures). In this context, the optical fiber is required to be reinforced by protective materials (e.g., fiber reinforced polymer (FRP)) [61–63]). When fabricating bare optical fibers with fiber reinforcements (e.g., carbon FRP), the perfect composition of both components should be achieved via tightly bonding conditions to guarantee the strain transfer from the infrastructure to the fiber core. The strain transfer mechanism varies by the gauge length and material for the reinforcement as well as surface-treatment (e.g., adhesive or sand coating), so that a proper calibration test is needed to measure the actual strain within infrastructures [44]. Stated differently, the nominal strain coefficient differs from that of the bare optical fiber [63]. Otherwise, the actual strains can be over- or underestimated.

Although the distributed strain sensor provides less-accurate strains along the optical fiber than the points strain sensors, the distributed strain sensor can provide the key information about spatial distributions over whole structural systems. On the other hand, point strain sensors measure accurate strains at discrete measurement positions. The advantage and disadvantage of these sensors are schematically represented in Fig. 2. The less-accurate strain distribution can be improved using a few samples of the point strain sensors, so that the high-spatial information of the distributed strain sensor can be fused with high-accuracy from the point strain sensors. Although their complementary data-fusion is potentially promising and powerful, this has not been investigated yet for the structural response estimation.

2.2. Gaussian process regression model

A Gaussian process (GP) regression is a non-parametric Bayesian approach towards regression problem by using a finite sum over all data to make prediction [33]. More formally, a Gaussian process is a generalization of a multivariate Gaussian

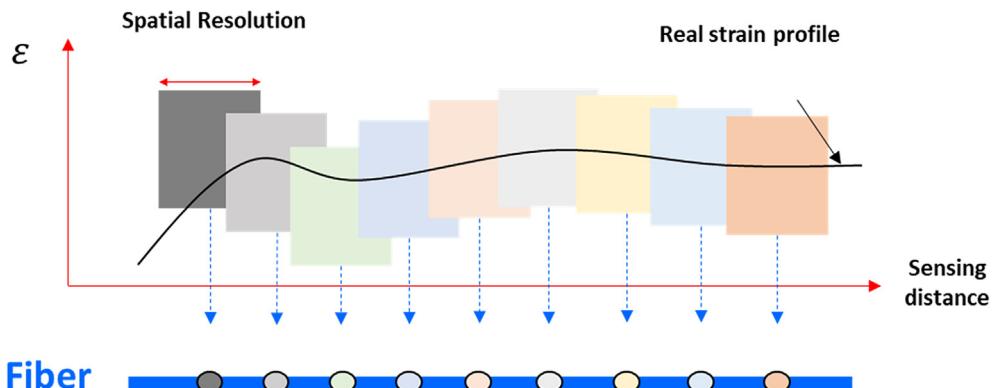


Fig. 1. Illustration of spatial resolution in distributed optical fiber sensing.

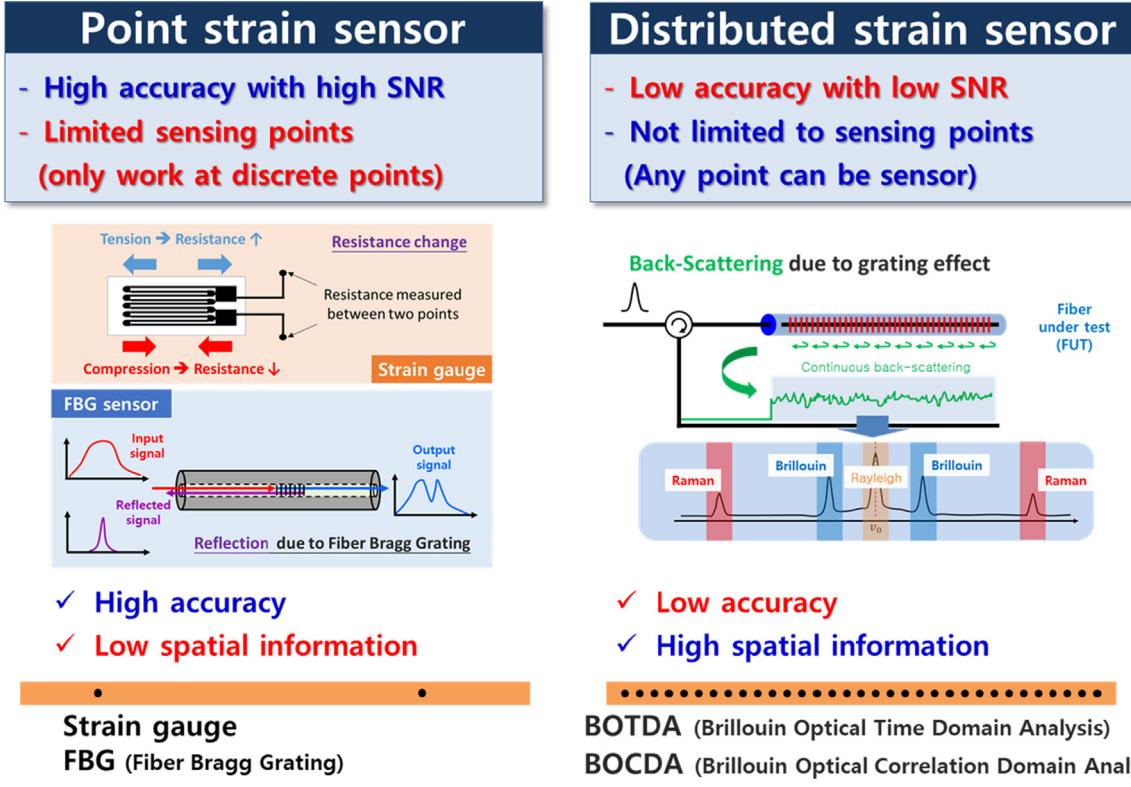


Fig. 2. Strength and weakness of point/distributed strain sensor.

distribution to infinitely dimensional object (i.e., output function of being fitted). In the GP regression, a noisy data y is assumed to output ($f(\mathbf{x})$) at input (\mathbf{x}) with noise (ϵ) as written in Eq. (1).

$$\mathbf{y} = f(\mathbf{x}) + \epsilon \quad (1)$$

where $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. Despite its similarity to the assumption of linear regression, the output ($f(\mathbf{x})$) in the GP regression is assumed to a random variable that follows a particular Gaussian distribution. The noise (ϵ) represents inherent randomness in the noisy data. In the GP regression, the output ($f(\mathbf{x})$) is distributed as a Gaussian process ($GP(\cdot, \cdot)$) as

$$f(\mathbf{x}) \sim GP(\mu, k(\mathbf{x}, \mathbf{x}')) \quad (2)$$

where μ and $k(\mathbf{x}, \mathbf{x}')$ denote mean and covariance function at input (\mathbf{x}), respectively. The mean function (μ) describes the expected values in the Gaussian distribution at input (\mathbf{x}), while covariance function ($k(\mathbf{x}, \mathbf{x}')$) represents the dependence between the function values at different inputs (\mathbf{x} and \mathbf{x}'). Setting prior mean function to zero ($\mu = 0$) has the merits of avoiding expensive posterior computation by doing the inference with only covariance function. In this regard, the prior mean function is often set to zero for numerical efficiency and simplicity. Setting the mean function to zero is equivalent to subtracting the mean from all data ($\mathbf{Y} - \mu$). The covariance function is commonly called the kernel of the Gaussian process [32]. The kernel function should be chosen to represent the patterns of the outputs in the data. One very popular assumption of the output behavior is that correlation between two inputs decreases with their distances. A squared exponential (SE) kernel is the most popular kernel to fulfill this assumption, and this is defined as

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{l}\right) \quad (3)$$

where the signal variance (σ_f^2) determines an amplitude of the output scale; and the length-scale (l) controls the smoothness between two inputs. There are unknown hyper-parameters in the SE kernel such as the signal variance (σ_f^2) and length-scale (l), and they need to be inferred with the noise variance (σ_ϵ^2) from the training data. In this study, the SE kernel is used for the GP regression to represent smooth patterns in the data.

Although Gaussian process are continuous (i.e., a distribution over continuous function), Gaussian process should be represented using a finite number of the samples in reality (i.e., training samples) based on the marginalization property of

Gaussians. For the target output $f(\mathbf{x})$, the training samples are given as a set of training inputs ($\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$) in the d -dimensional input space (\mathbf{X}) and their corresponding noisy data ($\mathbf{Y} = [y(\mathbf{x}_1) - \hat{\mu}, y(\mathbf{x}_2) - \hat{\mu}, \dots, y(\mathbf{x}_n) - \hat{\mu}]^T$; where $\hat{\mu}$ is the average value of all noisy data ($\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y(\mathbf{x}_i)$) and n denotes the number of the training samples. The joint prior distribution of the training set ($\mathbf{D} = [\mathbf{X}, \mathbf{Y}]$) can be represented by augmenting with new inputs (\mathbf{X}_*) and their outputs (\mathbf{f}_*), and a prior GP regression can be defined as

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_\epsilon^2 \mathbf{I} & \mathbf{K}(\mathbf{X}, \mathbf{X}_*) \\ \mathbf{K}(\mathbf{X}_*, \mathbf{X}) & \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix}\right) \quad (4)$$

where $\mathbf{K}(\mathbf{X}, \mathbf{X})$ is the symmetric and positive semi-definite matrix between all training data (\mathbf{X}) using kernel function of Eq. (3). This can be written as

$$\mathbf{K}(\mathbf{X}, \mathbf{X}) = \begin{bmatrix} \mathbf{k}(\mathbf{x}_1, \mathbf{x}_1) & \cdots & \mathbf{k}(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ \mathbf{k}(\mathbf{x}_n, \mathbf{x}_1) & \cdots & \mathbf{k}(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix} \quad (5)$$

The other components of covariance matrix in Eq. (4) (e.g., $\mathbf{K}(\mathbf{X}, \mathbf{X}_*)$) are described in a similar manner, as given in Eq. (5). \mathbf{I} is an $(n \times n)$ identity matrix and σ_ϵ^2 represents the noise level of noisy data (i.e., variance of the noise (ϵ) in Eq. (1)).

Before making predictions for new inputs (\mathbf{X}_*), the joint Gaussian prior distribution should be conditioned on the training data (\mathbf{X}) (i.e., posterior distribution). Stated differently, the prior GP regression learns the characteristics of the output from the training data (\mathbf{D}) by inferring hyper-parameters in GP regression ($\Theta = [\sigma_f^2, \mathbf{I}, \sigma_\epsilon^2]$). The optimal hyper-parameters ($\hat{\Theta}$) are estimated by minimizing the negative log-marginal likelihood (NLML) in Eq. (6).

$$\hat{\Theta} = \arg \min_{\Theta} \left(-\frac{1}{2} \mathbf{Y}^T [\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{Y} - \frac{1}{2} \log |\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_\epsilon^2 \mathbf{I}| - \frac{n}{2} \log 2\pi \right) \quad (6)$$

It is worth noting that the NLML can be viewed as a penalized log-likelihood function: the first term of Eq. (6) measures the goodness-of-fit; the second term penalizes the model complexity; and the last term is a normalizing constant. The minimization problem of Eq. (6) is generally performed using a multi-start gradient based optimization [64] by making use of the partial derivative of Eq. (6) with respect to the hyper-parameters (Θ), as given in Eq. (7).

$$\frac{\partial \text{NLML}}{\partial \Theta_i} = \frac{1}{2} \text{tr} \left((\boldsymbol{\alpha} \boldsymbol{\alpha}^T - [\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_\epsilon^2 \mathbf{I}]^{-1}) \frac{\partial [\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_\epsilon^2 \mathbf{I}]}{\partial \Theta_i} \right) \quad (7)$$

where $\boldsymbol{\alpha} = [\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{Y}$.

By updating our prior distribution with the optimal hyper-parameters ($\hat{\Theta}$), the posterior distribution of the \mathbf{X}_* is obtained as the conditional distribution ($p(\mathbf{f}_* | \mathbf{X}, \mathbf{Y}, \mathbf{X}_*, \hat{\Theta})$). This is defined as

$$\text{GPR} = p(\mathbf{f}_* | \mathbf{X}, \mathbf{Y}, \mathbf{X}_*, \hat{\Theta}) \sim \mathcal{N}(\hat{\mathbf{f}}(\mathbf{X}_*), \mathbf{s}^2(\mathbf{X}_*)) \quad (8)$$

The posterior mean ($\hat{\mathbf{f}}(\mathbf{X}_*)$) and the posterior variance ($\mathbf{s}^2(\mathbf{X}_*)$) are obtained using Eqs. (9) and (10), respectively.

$$\hat{\mathbf{f}}(\mathbf{X}_*) = \hat{\mu} + \mathbf{K}(\mathbf{X}_*, \mathbf{X}) [\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{Y} \quad (9)$$

$$\mathbf{s}^2(\mathbf{X}_*) = \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) - \mathbf{K}(\mathbf{X}_*, \mathbf{X}) [\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{K}(\mathbf{X}, \mathbf{X}_*) \quad (10)$$

where $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y(\mathbf{x}_i)$. The prediction ($\hat{y}(\mathbf{x})$) at any input (\mathbf{x}) is computed using the posterior mean ($\hat{\mathbf{f}}(\mathbf{X}_*)$) and the uncertainties associated with the prediction can be quantified by the posterior variance ($\mathbf{s}^2(\mathbf{X}_*)$).

2.3. Multi-fidelity data-fusion with recursive Gaussian process models

The fidelity of the data (or model) can vary from low-fidelity (LF) to high-fidelity (HF). The LF-data is relatively inexpensive to acquire, so it is possible to produce large amounts of samples for distribution information (output trend). However, the LF-data cannot be directly used for predicting high-fidelity performance due to its inaccuracy. Although the HF-data provides high accuracy, these data are expensive to acquire in terms of both time and money. Therefore, the limited amount of samples is available and this can significantly impair the ability to produce valid predictions.

The multi-fidelity data-fusion (MF) has been steadily gaining popularity in simulation based modeling, uncertainty quantification and optimization [36–39,65–70]. The MF (also known as variable-fidelity modeling) utilizes the cross-correlation between a LF- and HF-model to build a fast surrogate of the HF model [65,71]. The basic concept of the MF is to transfer knowledge of the cheap but potentially inaccurate LF-model to enhance the surrogate modeling of the HF-model [68].

Although the MF has been mainly investigated in the deterministic computer experiments [37–39,65–68], a few researchers have been studied to leverage an inaccurate simulator and noisy experimental data in order to maximize the accuracy for quantity of interest [36,69,70].

Since GP regression has many merits such as Bayesian non-parametric nature and their analytical tractability properties, many researches using MF are based on the linear auto-regressive (AR) mapping scheme in combination with the GP regression (also known as the approach of Kennedy & O'Hagan [36]). Kennedy & O'Hagan [36] have introduced a Bayesian discrepancy-based multi-fidelity modeling framework, which is an extension to the co-Kriging model [41]. In their framework, a linear AR co-Kriging model with linear mapping (\mathbf{MF}_{KO}) is proposed by expressing the HF-data as the sum of the scaled LF-data and Gaussian process that associates with the discrepancy between the scaled LF- and HF-data. The \mathbf{MF}_{KO} is defined as

$$\mathbf{f}_{\text{HF}}(\mathbf{x}) = \rho \cdot \mathbf{f}_{\text{LF}}(\mathbf{x}) + \delta_d(\mathbf{x}) \quad (11)$$

where $\mathbf{f}_{\text{LF}}(\mathbf{x})$ and $\mathbf{f}_{\text{HF}}(\mathbf{x})$ denote LF- and HF-data, respectively; ρ is a linear mapping factor that reflects the cross-correlation between $\mathbf{f}_{\text{LF}}(\mathbf{x})$ and $\mathbf{f}_{\text{HF}}(\mathbf{x})$; and $\delta_d(\mathbf{x})$ is distributed to an independent GP regression ($\delta_d(\mathbf{x}) \sim \mathcal{N}(\mu_d, \mathbf{K}(\mathbf{x}, \mathbf{x}') + \sigma_{\epsilon_{\text{HF}}}^2 \mathbf{I})$) associated with discrepancies that cannot be captured by the linear mapping ($\delta_d(\mathbf{x}) = \mathbf{f}_{\text{HF}}(\mathbf{x}) - \rho \cdot \mathbf{f}_{\text{LF}}(\mathbf{x})$).

A recursive GP regression has been developed by Le Gratiet and Garnier [37], and this method has been widely adopted in the recent studies [38,39,65,69,70] due to its simplicity and numerical efficiency. The recursive GP regression replaces $\mathbf{f}_{\text{LF}}(\mathbf{x})$ in Eq. (8) with the posterior mean $\hat{\mathbf{f}}_{\text{LF}}(\mathbf{x})$ of the GP regression from the LF-data, while the training sets ($\{\mathbf{D}_{\text{LF}}, \mathbf{D}_{\text{HF}}\}$) have a nested structure ($\mathbf{D}_{\text{LF}} \supseteq \mathbf{D}_{\text{HF}}$). This allows one to decouple the AR co-kriging model into independent GP regressions with an identical posterior mean and variance of the AR co-kriging model [37,65,69].

The workflow of the \mathbf{MF}_{KO} is graphically shown in the left of Fig. 3. The \mathbf{MF}_{KO} first performs building the GP regression of the LF-data (\mathbf{GPR}_{LF}) using the training set of the LF-data ($\mathbf{D}_{\text{LF}} = [\mathbf{X}_{\text{LF}}, \mathbf{Y}_{\text{LF}}]$). As mentioned in Section 2.2, the optimal hyper-parameters of the \mathbf{GPR}_{LF} ($\hat{\Theta}_{\text{LF}} = [\hat{\sigma}_{\text{LF}}^2, \hat{\mathbf{L}}_{\text{LF}}, \hat{\sigma}_{\epsilon_{\text{LF}}}^2]$) can be achieved by minimizing Eq. (6). Then the \mathbf{GPR}_{LF} is represented as

$$\mathbf{GPR}_{\text{LF}} = p\left(\mathbf{f}_{*,\text{LF}} | \mathbf{X}_{\text{LF}}, \mathbf{Y}_{\text{LF}}, \mathbf{X}_*, \hat{\Theta}_{\text{LF}}\right) \sim \mathcal{N}\left(\hat{\mathbf{f}}_{\text{LF}}(\mathbf{X}_*), \mathbf{s}_{\text{LF}}^2(\mathbf{X}_*)\right) \quad (12)$$

where $\hat{\mathbf{f}}_{\text{LF}}(\mathbf{X}_*)$ and $\mathbf{s}_{\text{LF}}^2(\mathbf{X}_*)$ can be computed using Eqs. (9) and (10), respectively. In order to build the AR co-Kriging model, the hyper-parameters ($\hat{\Theta}_{\text{MF}_{\text{KO}}} = [\hat{\rho}, \hat{\mu}_d, \hat{\sigma}_{\text{HF}}^2, \hat{\mathbf{L}}_{\text{HF}}, \hat{\sigma}_{\epsilon_{\text{HF}}}^2]$) should be inferred. Their optimal values are achieved by minimizing the following NLML in Eq. (13).

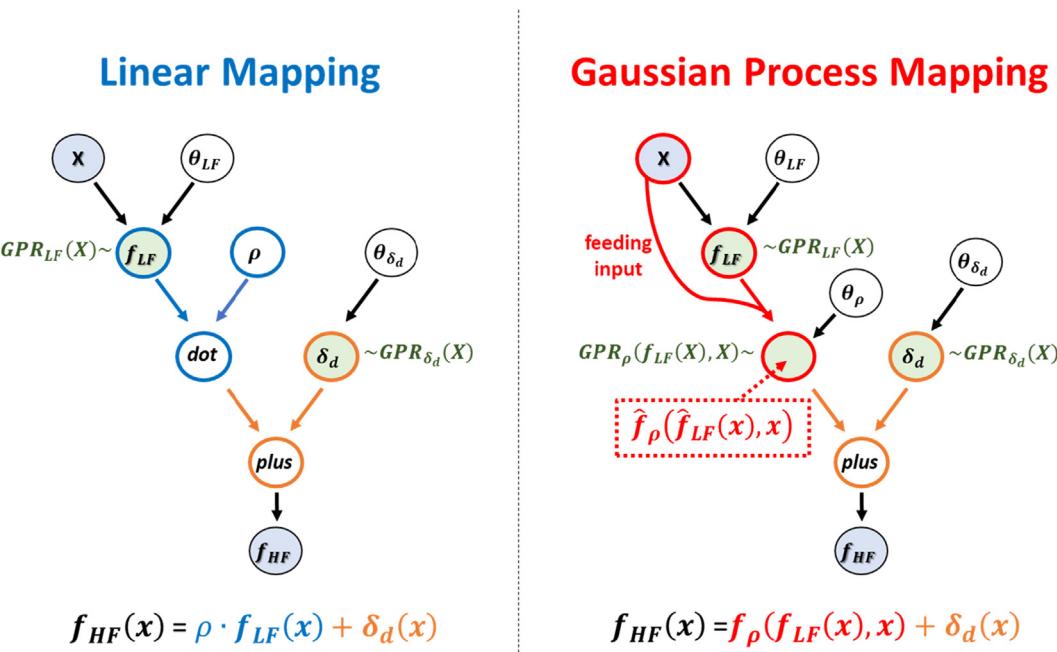


Fig. 3. Graphical representations of linear and Gaussian process mapping.

$$\hat{\Theta}_{MF_{KO}} = \arg \min_{\Theta} \left(\frac{n}{2} \log(\hat{\sigma}_{HF}^2) + \frac{1}{2} \log|R| + \frac{1}{2\hat{\sigma}_{HF}^2} \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \right) \quad (13)$$

where $\mathbf{R} = \mathbf{K}(\mathbf{X}_{LF}, \mathbf{X}_{HF}) + \sigma_{\epsilon_{HF}}^2 \mathbf{I}_m$ and $\mathbf{r} = \mathbf{Y}_{HF} - \mathbf{1}_m \mu_d - \rho \hat{\mathbf{f}}_{LF}(\mathbf{X}_{HF})$. $\mathbf{1}_m$ and \mathbf{I}_m are an $(m \times 1)$ vector and $(m \times m)$ matrix, respectively. After obtaining the optimal hyper-parameters ($\hat{\Theta}_{MF_{KO}}$), the posterior mean ($\hat{\mathbf{f}}_{MF_{KO}}(\mathbf{X}_*)$) of the multi-fidelity posterior distribution is calculated as

$$\hat{\mathbf{f}}_{MF_{KO}}(\mathbf{X}_*) = \hat{\mu}_d + \hat{\rho} \hat{\mathbf{f}}_{LF}(\mathbf{X}_*) + \mathbf{K}(\mathbf{X}_{HF}, \mathbf{X}_*)^T [\mathbf{K}(\mathbf{X}_{HF}, \mathbf{X}_{HF}) + \sigma_{\epsilon_{HF}}^2 \mathbf{I}_m]^{-1} (\mathbf{Y}_{HF} - \mathbf{1}_m \hat{\mu}_d - \hat{\rho} \hat{\mathbf{f}}_{LF}(\mathbf{X}_{HF})) \quad (14)$$

3. Proposed multi-fidelity data-fusion using Gaussian process mapping

To learn more complex cross-correlations between LF- and HF-data, the proposed method introduces an input-connected Gaussian process (GP) mapping in terms of the LF-predictions and inputs (herein, location). The input-connected Gaussian process (GP) mapping is firstly presented to allow flexible mapping (non-linear and input-dependent cross-correlation). Then, the workflow of the proposed method (MF_{GP}) is presented.

3.1. Gaussian process (GP) mapping

The linear mapping in Eq. (8) cannot learn complex scale variations from the LF- to HF-data (cross-correlation). In reality, the cross-correlation between the LF- and HF-data may be represented by the linear or non-linear change of the scale variation. Additionally, the input-dependency can exist locally or globally. Empirically, Deep learning models (e.g., deep neural net) can improve the quality of learning the complicated patterns using non-linear mapping recursively. Deep Gaussian processes (GPs) has been proposed by Damianou and Lawrence [35] as a deep belief network based on recursive GP mapping. Deep GPs are viewed as a neural network with a finite number of non-parametric, GP-distributed basis function at each layer [34]. In this context, the output of the GP at the previous layer is used as the input of the GP at the next layer (i.e., auto-associate mapping). This mapping is recursively performed each layer to represent the complicated patterns.

Duvenaud, Rippel, Adams and Ghahramani [34] have mentioned that deep GPs based on a squared-exponential kernel tends to make a degenerate problem due to pathology, as the number of layers increases. To address this problem, they have simply made each layer depend not only on the output of the previous GP, but also on the original input (i.e., input-connected networks). Motivated from these works regarding to the deep GPs, the proposed method (MF_{GP}) replaces the simple linear mapping with the input-connected GP mapping to make its learning more flexible. In this context, the mapping function is designed as a GP regression using the original input (\mathbf{x}) and LF-prediction from the GP regression ($\hat{\mathbf{f}}_{LF}(\mathbf{x})$). This is defined as given in Eq. (15).

$$\hat{\mathbf{f}}_{HF}(\mathbf{x}) = \mathbf{h}_\rho(\hat{\mathbf{f}}_{LF}(\mathbf{x}), \mathbf{x}) + \delta_d(\mathbf{x}) = \mathbf{f}_\rho(\hat{\mathbf{f}}_{LF}(\mathbf{x}), \mathbf{x}) + \mathbf{f}_{\delta_d}(\mathbf{x}) \quad (15)$$

where $\mathbf{h}_\rho(\hat{\mathbf{f}}_{LF}(\mathbf{x}), \mathbf{x})$ denotes GP mapping from low-fidelity to high-fidelity by the GP regression ($GPR_\rho(\hat{\mathbf{f}}_{LF}(\mathbf{x}), \mathbf{x})$); $\mathbf{f}_\rho(\hat{\mathbf{f}}_{LF}(\mathbf{x}), \mathbf{x})$ is the posterior mean of the $GPR_\rho(\hat{\mathbf{f}}_{LF}(\mathbf{x}), \mathbf{x})$; and $\mathbf{f}_{\delta_d}(\mathbf{x})$ is distributed to an independent GP regression associated with discrepancies that cannot be captured by the input-connected GP mapping ($\mathbf{f}_{HF}(\mathbf{x}) - \mathbf{f}_\rho(\hat{\mathbf{f}}_{LF}(\mathbf{x}), \mathbf{x})$). Since Gaussian process can be understood as one-hidden-layer perceptron with infinitely many hidden units [34,35], the MF_{GP} (Eq. (15)) is equivalent to deep GPs with two-layers connected to the original input as shown in Fig. 4. This can allow for flexible modeling of the complex cross-correlation and the input-dependency.

In order to provide the flexible learning capability for multi-fidelity data-fusion, the MF_{GP} introduces the squared exponential (SE) kernel for the covariance kernels and GP mapping. It should be noted that using the SE kernel is not optimal to construct GPs in each fidelity and mapping function for transferring the knowledge from low-fidelity to high-fidelity. Stated differently, using other covariance kernel can improve performances than using the SE kernel. Therefore, choosing a proper covariance kernel is crucial for maximizing the learning capability. However, the optimal covariance kernels in the MF_{GP} is problem-dependent (i.e., their correlation, number and location of data). Based on our previous studies [72,73], the compositional kernel learning (CKL) algorithm is one of the promising methods to construct the optimal covariance kernel. The CKL algorithm automatically discovers a compositional kernel for a richly structured kernel to maximize the learning capability. Elaborating the optimal kernel design (including CKL or other methods) is our on-going study.

3.2. Workflow

The proposed MF_{GP} is defined to the MF_{KO} in Section 2.3, except for the mapping operator. The workflow of the proposed method is graphically shown in the right of Fig. 3. The AR co-kriging model in Eq. (8) requires two times of building the GP

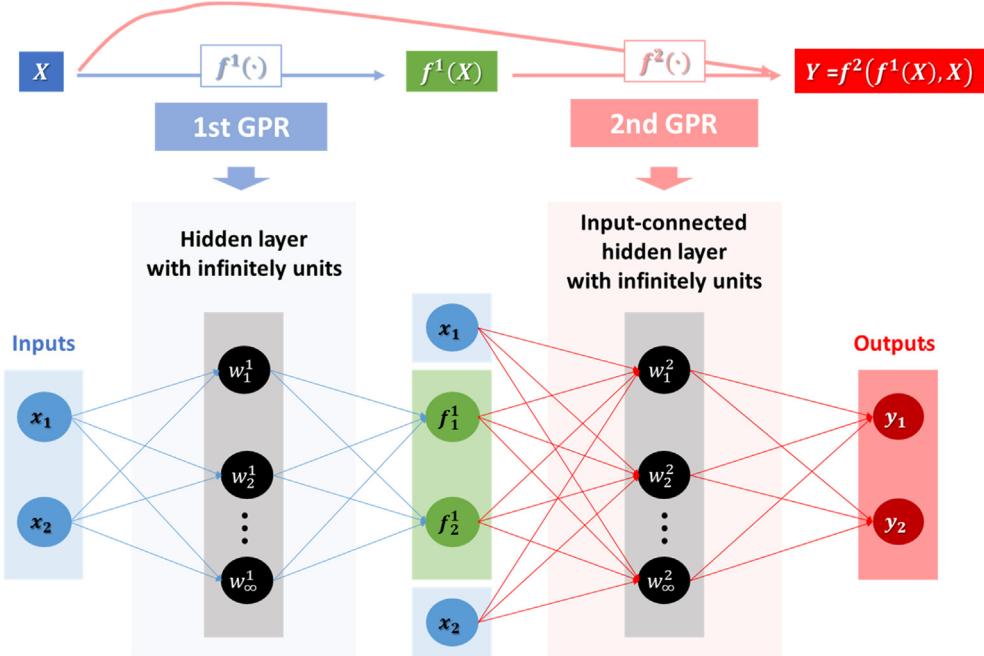


Fig. 4. Two-layer GP with input-connected architecture.

regression, while the proposed method needs three times to train the GP regression. By virtue of the recursive framework, both methods perform building the GP regression independently and sequentially according to their workflow.

The training set of the LF-data ($\mathbf{D}_{LF} = [\mathbf{X}_{LF}, \mathbf{Y}_{LF}]$) is used to build the GP regression of the LF-data (\mathbf{GPR}_{LF}). Based on the \mathbf{GPR}_{LF} , the low-fidelity predictions for inputs of the HF-data ($\hat{\mathbf{f}}_{LF}(\mathbf{X}_{HF})$) are estimated using $\hat{\Theta}_{LF} = [\hat{\sigma}_{LF}^2, \hat{\mathbf{l}}_{LF}, \hat{\sigma}_{\epsilon_{LF}}^2]$. The GP regression for mapping function ($\mathbf{h}_\rho(\hat{\mathbf{f}}_{LF}(\mathbf{x}), \mathbf{x})$) is constructed using the original input (\mathbf{x}) and low-fidelity prediction ($\hat{\mathbf{f}}_{LF}(\mathbf{X}_{HF})$) as \mathbf{GPR}_ρ .

$$\mathbf{GPR}_\rho = \mathbf{p}(\mathbf{f}_{*,\rho} | \mathbf{X}_{HF}, \hat{\mathbf{f}}_{LF}(\mathbf{X}_{HF}), \mathbf{Y}_{HF}, \hat{\Theta}_\rho) \mathcal{N}(\hat{\mathbf{f}}_\rho(\hat{\mathbf{f}}_{LF}(\mathbf{X}_*), \mathbf{X}_*), \mathbf{s}_\rho^2(\hat{\mathbf{f}}_{LF}(\mathbf{X}_*), \mathbf{X}_*)) \quad (16)$$

where $\hat{\Theta}_\rho = [\hat{\sigma}_\rho^2, \hat{\mathbf{l}}_\rho, \hat{\sigma}_{\epsilon_\rho}^2]$. The discrepancy between the HF-data (\mathbf{X}_{HF}) and the corresponding prediction of the GP mapping function ($\hat{\mathbf{f}}_\rho(\hat{\mathbf{f}}_{LF}(\mathbf{X}_{HF}), \mathbf{X}_{HF})$) is computed as

$$\delta_d(\mathbf{X}_{HF}) = \mathbf{Y}_{HF}(\mathbf{X}_{HF}) - \hat{\mathbf{f}}_\rho(\hat{\mathbf{f}}_{LF}(\mathbf{X}_{HF}), \mathbf{X}_{HF}) \quad (17)$$

To account for the discrepancy ($\delta_d(\mathbf{X}_{HF})$) in Eq. (17), an independent GP regression associated with $\delta_d(\mathbf{X}_{HF})$ is modelled using the optimal hyper-parameters ($\hat{\Theta}_{\delta_d} = [\hat{\mu}_d, \hat{\sigma}_d^2, \hat{\mathbf{l}}_d, \hat{\sigma}_{\epsilon_{HF}}^2]$).

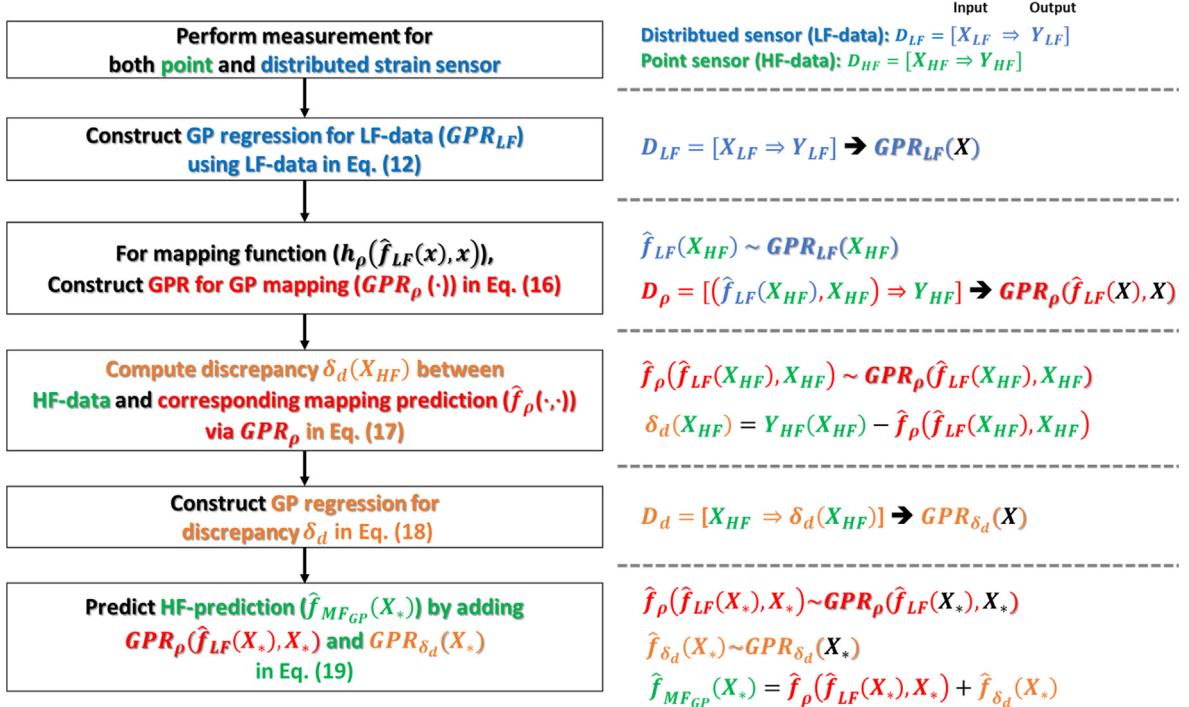
$$\mathbf{GPR}_{\delta_d} = \mathbf{p}(\mathbf{f}_{*,\delta_d} | \mathbf{X}_{HF}, \hat{\mathbf{f}}_{LF}(\mathbf{X}_{HF}), \mathbf{Y}_{HF}, \mathbf{X}_*, \hat{\Theta}_{\delta_d}) \mathcal{N}(\hat{\mathbf{f}}_{\delta_d}(\mathbf{X}_*), \mathbf{s}_{\delta_d}^2(\mathbf{X}_*)) \quad (18)$$

Since two GP regressions are independent each other, the posterior mean $\hat{\mathbf{f}}_{MF_{GP}}(\mathbf{X}_*)$ can obtain by the additive form as

$$\hat{\mathbf{f}}_{MF_{GP}}(\mathbf{X}_*) = \hat{\mathbf{f}}_\rho(\hat{\mathbf{f}}_{LF}(\mathbf{X}_*), \mathbf{X}_*) + \hat{\mathbf{f}}_{\delta_d}(\mathbf{X}_*) \quad (19)$$

where $\hat{\mathbf{f}}_\rho(\hat{\mathbf{f}}_{LF}(\mathbf{X}_*), \mathbf{X}_*)$ and $\hat{\mathbf{f}}_{\delta_d}(\mathbf{X}_*)$ is the posterior mean from \mathbf{GPR}_ρ and \mathbf{GPR}_{δ_d} , respectively. The flowchart of the proposed method is shown in Fig. 5.

The hierarchical implementation of the \mathbf{MF}_{GP} has several advantages in computational efficiency and stability. Firstly, the hierarchical implementation of the standard GP is computationally efficient over the \mathbf{MF}_{KO} . The \mathbf{MF}_{KO} requires the construction of the large cross-covariance matrix (\mathbf{R} matrix in Eq. (13)) using all LF- and HF-data, while the \mathbf{MF}_{GP} sequentially constructs the cross-covariance matrix in each fidelity. By virtue of the sequential implementation, the \mathbf{MF}_{GP} is less expensive than the \mathbf{MF}_{KO} by using smaller cross-covariance matrix. Secondly, the number of the hyper-parameters for the GP modeling can be reduced, so that it minimizes errors in the hyper-parameter estimation. In addition, various correlation kernel

**Fig. 5.** Flowchart of proposed method.

can be introduced separately in each fidelity. In numerical stability, the MF_{GP} is potentially less instable than the MF_{KO} for matrix manipulation (i.e., matrix inversion).

4. Numerical study

In this section, the numerical studies are presented to highlight the flexible learning capability of the proposed method. The goals of these studies are as follows: (1) this study demonstrate the performance of the proposed method by comparing with other methods; (2) since the number of HF-data is limited in the real problems, the influence on the number of the HF-data was evaluated by considering the variability in their locations; and (3) two numerical problems with different cross-correlations were investigated to show the flexibility of the proposed method. The five different methods were compared: (1) the GP regression of the LF-data (**LF**), (2) the GP regression of the HF-data (**HF**), (3) the MF using Hierarchical Kriging (**MF_{HKRG}**), (4) the MF using Kennedy & O'Hagan method (**MF_{KO}**), and (5) the MF using GP mapping (**MF_{GP}**). As the existing method for multi-fidelity data-fusion, the **MF_{HKRG}** was introduced for comparison [74]. It utilizes the Kriging model of the LF-data as a trend of the Kriging model for the HF-data. This method does not assume any linear cross-correlation between LF and HF-data. For the **MF_{HKRG}**, UQLab [75] was utilized based on the implementation of the **MF_{HKRG}** in Lataniotis, Marelli and Sudret [76].

Prediction performances of these methods are evaluated using two performance measures: (1) root mean square error (RMSE) and (2) maximum absolute error (MAE). The RMSE is widely used to aggregate the squared root of the mean square error between observed values (y_i) and predicted values (\hat{y}_i) into a single measure of the prediction performance. The RMSE is defined as

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}} \quad (20)$$

where n represents the total number of the data. The MAE measures the maximum value among the absolute difference between y_i and \hat{y}_i , and this is defined as

$$MAE = \max(|y_1 - \hat{y}_1|, |y_2 - \hat{y}_2|, \dots, |y_n - \hat{y}_n|) \quad (21)$$

where $|\cdot|$ denotes the absolute value. The RMSE provides measure for the global accuracy, while the MAE measure the local accuracy.

To evaluate the influence on the number of the HF-data, a Monte Carlo (MC) simulation was performed by varying the number of the HF-data and their locations. For all MC simulations, the number of the LF-data is fixed. Based on the results of the MC simulation, the proposed method (\mathbf{MF}_{GP}) was validated by comparing with the four methods (\mathbf{LF} , \mathbf{HF} , \mathbf{MF}_{HKRG} and \mathbf{MF}_{KO}).

4.1. Numerical study #1: local variation in cross-correlation problem

We considered a one-dimensional test function involving two levels of fidelity under the presence of the local variation in cross-correlation. The noiseless LF-function (f_1) and HF-function (f_2) are expressed respectively as

$$f_1(\mathbf{x}) = 0.5(6\mathbf{x} - 2)^2 \sin(12\mathbf{x} - 4) - 50(\mathbf{x} - 0.5) \quad (22)$$

and

$$f_2(\mathbf{x}) = (6\mathbf{x} - 2)^2 \sin(12\mathbf{x} - 4) \quad (23)$$

where $x \in [0, 1]$. To generate reference sets for the LF- and HF-data ($\mathbf{D}_{LF} = [\mathbf{X}_1, \mathbf{Y}_1]$ and $\mathbf{D}_{HF} = [\mathbf{X}_2, \mathbf{Y}_2]$), grid sampling was used by taking $\mathbf{m}_{LF} = \mathbf{m}_{HF} = 50$; where \mathbf{m}_{LF} and \mathbf{m}_{HF} denote the number for the LF- and HF-data in the reference sets, respectively. In this context, the LF- and HF-data in the reference set can be nested ($\mathbf{D}_{LF} \supseteq \mathbf{D}_{HF}$). These reference sets were used to generate sub-sets randomly for MC simulation. In order to consider measurement noise, the reference sets of the LF- and HF-data were perturbed by Gaussian noise (ϵ_i).

$$\mathbf{Y}_i = \mathbf{f}_i(\mathbf{X}_i) + \epsilon_i, \quad i = 1, 2 \quad (24)$$

where $\epsilon_1 \sim \mathcal{N}(0, 0.3)$ and $\epsilon_2 \sim \mathcal{N}(0, 0.1)$. The validation data was generated using grid sampling with $\mathbf{m}_{VALID} = 100$ from HF-function (f_2) to evaluate the prediction performance. The noiseless functions of the LF- and HF-data (denoted by solid line and dotted line) and the noisy LF- and HF-data (denoted by red-cross marker and dark dot marker) are shown in Fig. 6(a). Fig. 6(b) shows the noises injected to the reference set for LF- and HF-data (ϵ_i). The cross-correlation between the noisy LF- and HF-data is shown in Fig. 6(c). The non-linear variation of the cross-correlation exists at the local regions of the input (i.e., $Y_1 \in [-10, -3]$), while the other region ($Y_1 \in [-4, 8]$) exhibits the linear cross-correlation.

To evaluate the four different methods, the MC simulation was performed by varying the number of the HF-data. To consider variability in the position of the HF-data, the range of the input space was divided into N_{HF} equally sub-intervals (N_{HF} subset) like Latin hypercube design [77]. For each repetition of the MC simulation, one HF-data was selected randomly from each subset of the reference set. The seven case studies (with different number of the HF-data) were investigated with 100 repetitions of the MC simulation, as tabulated in Table 3.

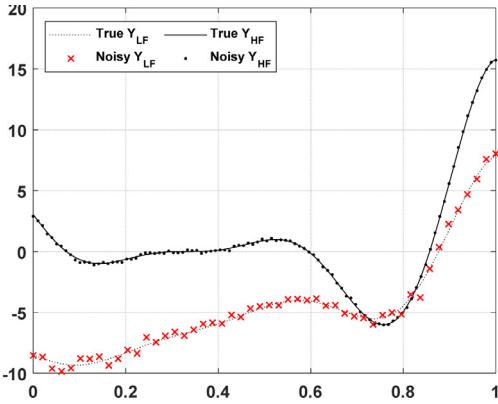
The RMSE and MAE values were calculated using the validation data to evaluate the performance. Since the performance measures of the \mathbf{LF} are constant RMSE and MAE values, they are used as a normalized value to compare the global and local accuracy of the remaining four methods (\mathbf{HF} , \mathbf{MF}_{HKRG} , \mathbf{MF}_{KO} and \mathbf{MF}_{GP}). Fig. 7 shows boxplots of the normalized RMSE and MAE values (NRMSE and NMAE) over 100 MC runs, while Fig. 8 illustrates best realizations of the HF-predictions from the \mathbf{HF} , \mathbf{MF}_{HKRG} , \mathbf{MF}_{KO} and \mathbf{MF}_{GP} with that of the \mathbf{LF} .

Based on Fig. 7, the \mathbf{MF}_{GP} generally provides better prediction performance than those of the \mathbf{MF}_{HKRG} , \mathbf{MF}_{KO} and \mathbf{HF} . It is seen clearly that the \mathbf{MF}_{GP} provides lower NRMSE and NMAE values than those of the \mathbf{HF} , \mathbf{MF}_{HKRG} , and \mathbf{MF}_{KO} after using more than the seven HF-data ($N_{HF} > 7$). It is observed that three methods based on the multi-fidelity data-fusion (\mathbf{MF}_{HKRG} , \mathbf{MF}_{KO} and \mathbf{MF}_{GP}) shows the improvement of the prediction performances in accordance with the increase of the HF-data. Since the locations of the HF-data were different for each Monte Carlo run, the \mathbf{MF}_{GP} results in the large variability. Some realizations of the HF-data are not informative to learn the cross-correlation, while some realizations are informative for multi-fidelity data-fusion. However, it is observed that the prediction performances of the \mathbf{MF}_{GP} are better than or similar to those from other methods ($N_{HF} = 5, 6, 7$). This indicates that the \mathbf{MF}_{GP} generally outperforms or performs similarly to the best one among existing methods. Stated differently, the \mathbf{MF}_{GP} can produce the HF-predictions accurately and efficiently with fewer number of the HF-data than other methods.

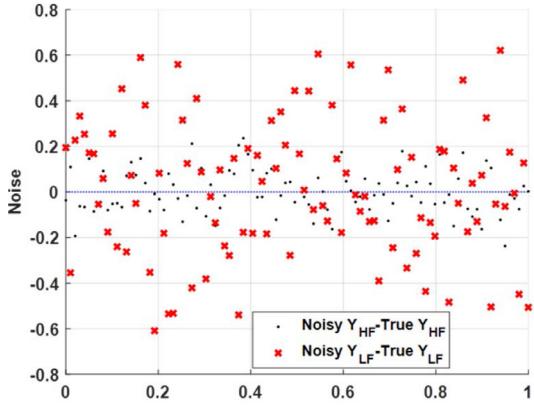
Based on Fig. 8(a), the HF-predictions from the \mathbf{MF}_{HKRG} , \mathbf{MF}_{KO} and \mathbf{HF} produce large deviations in some regions, while that of the \mathbf{MF}_{GP} represents the true values of the HF-function (f_2) very accurately. When the sufficient number of the HF-data is used ($N_{HF} = 9$), all methods (\mathbf{HF} , \mathbf{MF}_{HKRG} , \mathbf{MF}_{KO} and \mathbf{MF}_{GP}) provide the accurate HF-predictions for all regions. It is worth noting that the \mathbf{MF}_{GP} always provides the best HF-prediction in terms of both global and local accuracy. Based on these results, the \mathbf{MF}_{GP} can transfer useful information (i.e., functional trend) from the LF-data to enhance the modeling

Table 3
MC simulation for Numerical study #1 by varying N_{HF} .

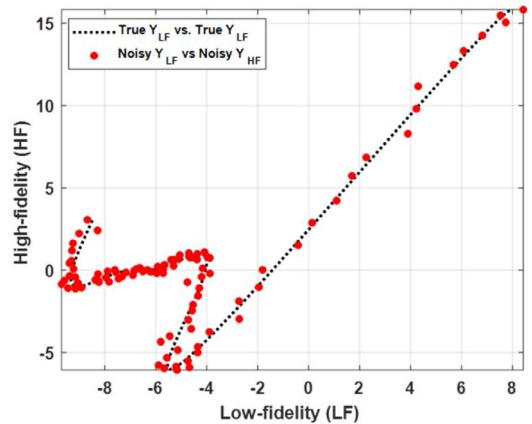
Numerical study #1	#1	#2	#3	#4	#5	#6	#7
# of LF-data (N_{LF})	50	50	50	50	50	50	50
# of HF-data (N_{HF})	3	4	5	6	7	8	9



(a) LF- and HF-fidelity functions and data



(b) Noises in LF- and HF-fidelity data



(c) Cross-correlation

Fig. 6. Mild non-linear cross-correlation problem ($d = 1$).

of the HF-predictions under the non-linear variation of the cross-correlation at the sub-region of the input space ($x \in [-10, -3]$ in Fig. 6(b)).

4.2. Numerical study #2: highly non-linear cross-correlation

To demonstrate the flexible learning capability of the proposed method, one-dimensional test functions were considered. The noiseless LF-function (f_1) and HF-function (f_2) are defined respectively as

$$f_1(\mathbf{x}) = \sin(8\pi\mathbf{x}) \quad (25)$$

and

$$f_2(\mathbf{x}) = (\mathbf{x} - \sqrt{2})^2 f_1^2(\mathbf{x}) \quad (26)$$

where $\mathbf{x} \in [0, 1]$. Similarly in Numerical study #1, the reference sets for the LF- and HF-data (\mathbf{D}_{LF} and \mathbf{D}_{HF}) were generated using grid sampling by taking $\mathbf{m}_{LF} = \mathbf{m}_{HF} = 70$. The noise components for the LF- and HF-data (ϵ_i) were generated by the following Gaussian distributions: $\epsilon_1 \sim \mathcal{N}(0, 0.3)$ and $\epsilon_2 \sim \mathcal{N}(0, 0.1)$, and they were injected to the reference sets using Eq.

Table 4
MC simulation for Numerical study #2 by varying N_{HF} .

Numerical study #2	#1	#2	#3	#4	#5	#6	#7
# of LF-data (N_{LF})	70	70	70	70	70	70	70
# of HF-data (N_{HF})	4	7	10	13	16	19	22

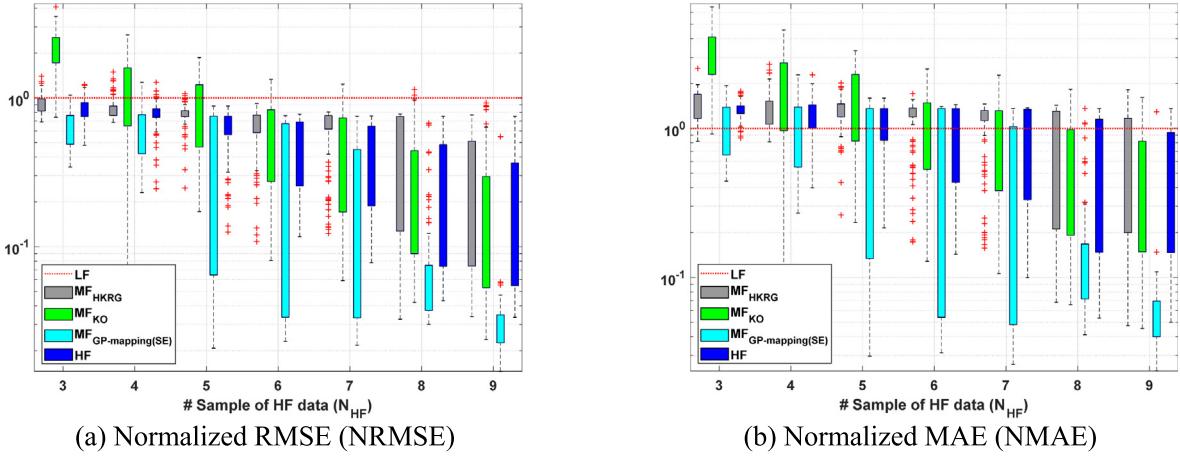
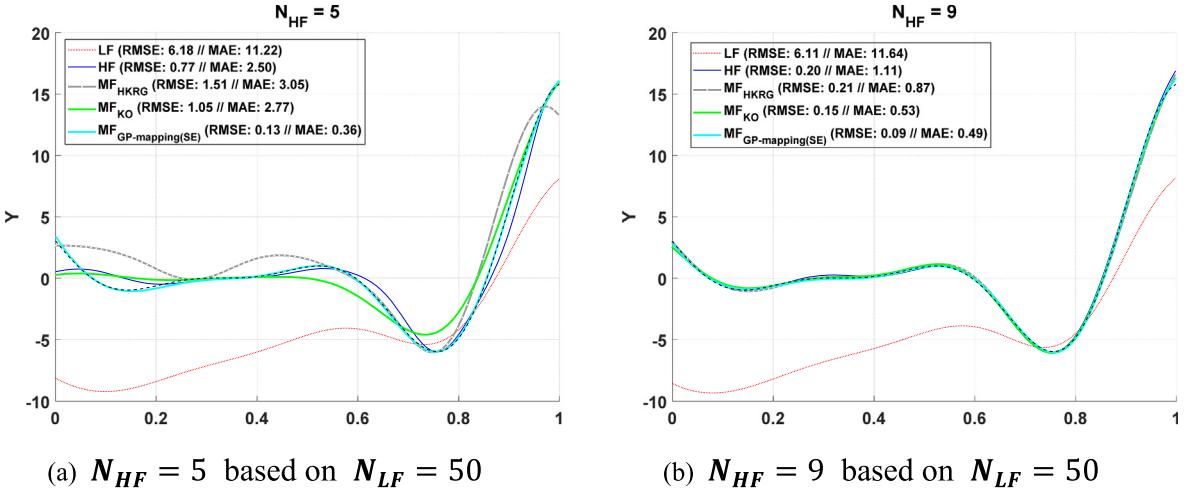


Fig. 7. Boxplots for performance measures over 100 repetitions.

Fig. 8. Predictions from the four methods with $N_{HF} = 5, 9$ based on $N_{LF} = 50$

(24). To evaluate the prediction performance, the validation data was generated using grid sampling with $m_{VALID} = 100$ from HF-function (f_2). Fig. 9(a) shows the true functions with the LF- and HF-data, while the noise levels for LF- and HF-data are shown in Fig. 9(b). As shown in Fig. 9(c), the cross-correlation between true LF- and HF-data exhibits highly non-linearity with input-dependency. The injected noise distills information on the cross-correlated between true LF- and HF-data, and this makes difficult to combine the information from the LF- and HF-data.

Similarly to the MC simulation in Numerical study #1, the range of the input space was divided into N_{HF} equally sub-intervals to evaluate the influence for variability in the position of the HF-data. The summary of the MC simulation was tabulated in Table 4. All case studies were evaluated with 100 repetitions of the MC simulation with different random seeds to select the subset randomly.

Fig. 10 shows boxplots of the NRMSE and NMAE over 100 MC runs. The **HF** cannot improve the global and local accuracy, as the the HF-data increases. The prediction performances of the **MF_{HKG}** are very similar to those of The **HF**. As the HF-data increases, the **MF_{GP}** and **MF_{KO}** improve the global and local predictions. It is worth noting that the **MF_{GP}** generally produces more accurate HF-predictions than those of the **MF_{KO}** with fewer HF-data. When the number of HF-data is equal or larger than ten ($N_{HF} \geq 10$), the values of RMSE and MAE from the **MF_{GP}** are slightly larger than those of the **HF**. At each realization of the MC runs, the locations of the HF-data are different under the given number of the HF-data. As a result, some realizations were slightly insufficient to learn the cross-correlations between LF- and HF-data even with the **MF_{GP}**. However, it should be noted that the values of the **MF_{GP}** are better than or similar to those from other methods.

Fig. 11 shows best realizations of the HF-predictions from **HF**, **MF_{HKG}**, **MF_{KO}** and **MF_{GP}** with that of the **LF**. In both $N_{HF} = 7$ and $N_{HF} = 22$, the **HF** only captures the quadratic trend in the HF-function (f_2) by smoothing out the periodic component due to insufficient number of the training data and noise effect. Stated differently, the periodic component was con-

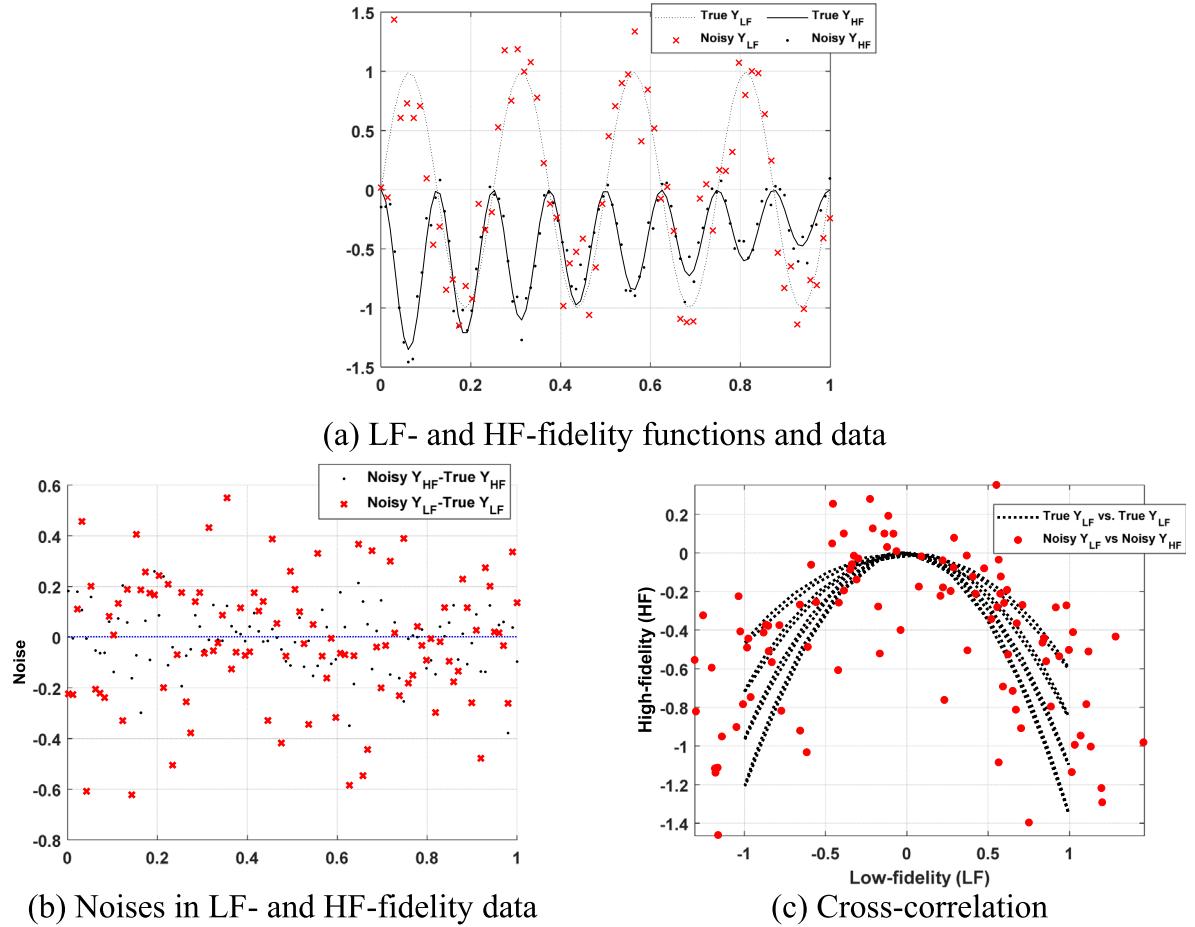


Fig. 9. Highly non-linear cross-correlation problem ($d = 1$).

sidered as the noise by estimating large value of σ_e^2 in Eq. (9). In order to capture both trend and periodicity, more HF-data is required. The \mathbf{MF}_{HKRG} produced similar results of the \mathbf{HF} . Although the \mathbf{MF}_{HKRG} utilized the infomration from the LF-data (i.e., periodicity) as the trend of the HF-predictions, it smoothed out the trend and only captured the quadratic trend. In $N_{HF} = 7$, the \mathbf{MF}_{KO} cannot represent the trend and periodic component correctly. The \mathbf{MF}_{KO} with $N_{HF} = 22$ can represent the HF-function (f_2), but it has the relatively large error at the certain region ($x = 0.15$). The \mathbf{MF}_{GP} can produce quadratic trend with the periodic component in both $N_{HF} = 7$ and $N_{HF} = 22$. It is worth noting that the \mathbf{MF}_{GP} is only successful to generate the accurate HF-prediction by capturing quadratic trend with periodicity in the HF-function through the proposed multi-fidelity method.

5. Experimental study

This section presents experimental studies to validate the feasibility and effectiveness for the proposed method using a lab-scaled test specimen under two loading cases. As the LF-data, a distributed optical fiber sensor system based on the Brillouin optical correlation domain analysis (BOCDA) technique was used for the distributed strain sensor to measure strains with the spatial resolution of 5 cm [15]. As the HF-data, an electrical resistance strain gauge (ERS) was used for the point strain sensor with the gauge factor of 2.13. Fig. 12 shows the experimental set-up for the point and distributed strain sensor.

A test specimen are shown with the sensor configuration at the top of Fig. 13. Since the spatial resolution of the BOCDA is 5 cm, the measured strains are the averaged values for 5 cm along the optical fiber. As a result, the sensing point of the BOCDA was equally spaced by 5 cm. Strain gauges were attached to these sensing points of the BOCDA in close proximity. Total number of the LF- and HF-data is fifteen ($N_{LF} = N_{HF} = 15$), and the strain gauge and polyimide-coated optical fiber were bonded by strain gauge adhesives at the bottom of the test specimen. The test specimen was simply supported by

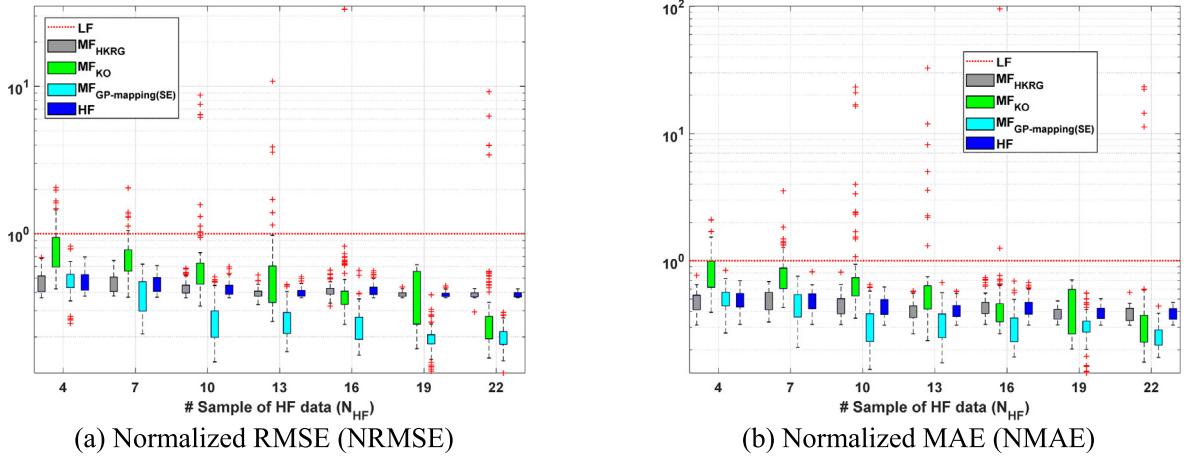
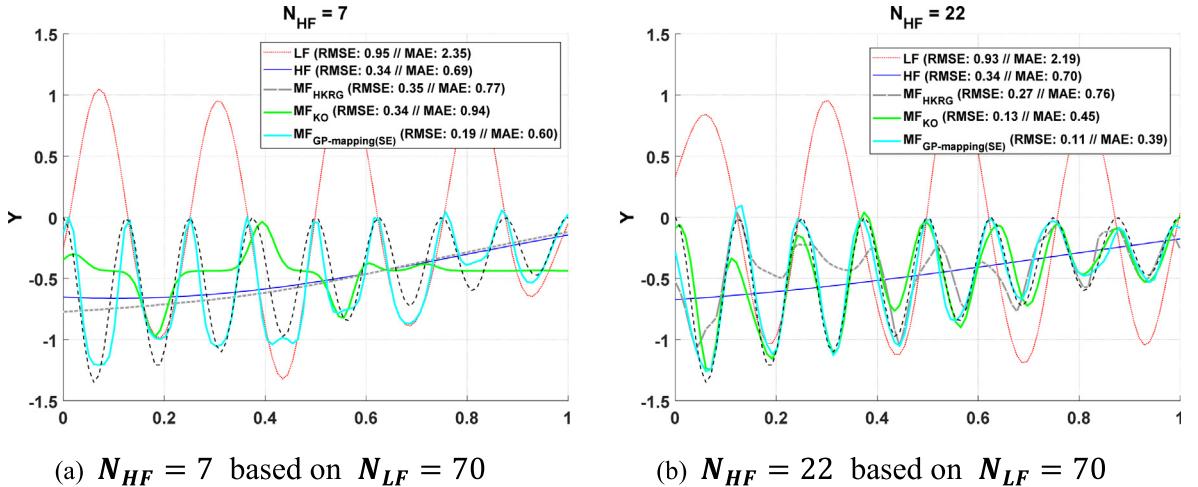


Fig. 10. Boxplots for performance measures over 100 repetitions.

Fig. 11. Predictions from the four methods with N_{HF} = 7, 22 based on N_{LF} = 70

hinges. As shown in the bottom of Fig. 13, two loading cases were considered based on the location of the loading as (1) asymmetric distribution (Experimental study #1) and (2) symmetric distribution (Experimental study #2).

In order to investigate the influence by both number and location of the HF-data, the MC simulation was performed by varying the number of the HF-data. The static loading tests were carefully performed with three times to put the masses to the identical location. As a result, two sets of the LF- and HF-data were measured (15 points, respectively). By averaging them, averaged values of the LF- and HF-data were obtained. Based on the averaged values, the MC repetitions were performed according to the given number of the HF-data. According to the number of the HF-data (N_{HF}) as tabulated in Table 5, the range of the input space was divided into N_{HF} sub-regions (N_{HF} subset). The HF-data was chosen randomly from each subset. All LF-data was used to extract the strain distribution, so that the LF- and HF-data inherently have the nest structure. The seven case studies were investigated depending on the number of the HF-data.

Table 5
Experimental study with varying N_{HF}.

Experimental study	#1	#2	#3	#4	#5	#6	#7
# of LF data (N _{LF})	15	15	15	15	15	15	15
# of HF data (n _{HF})	4	5	6	7	8	9	10

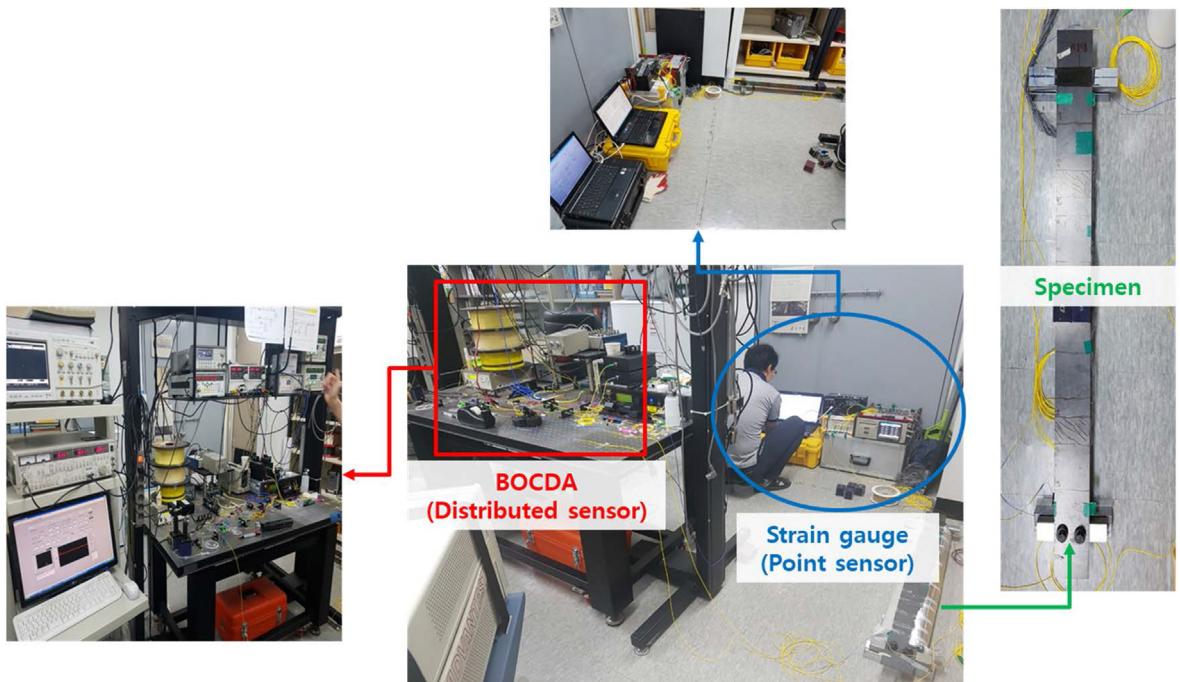


Fig. 12. Experimental set-up using point/distributed strain sensor.

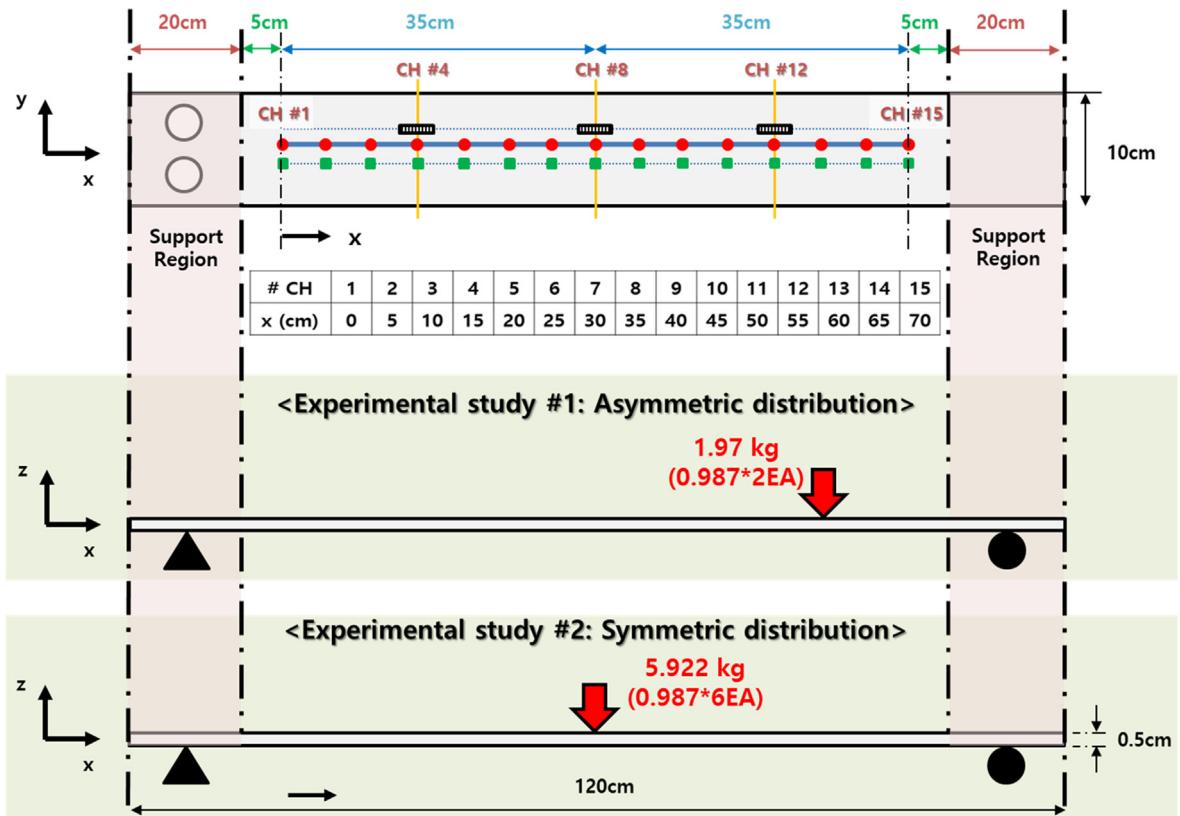


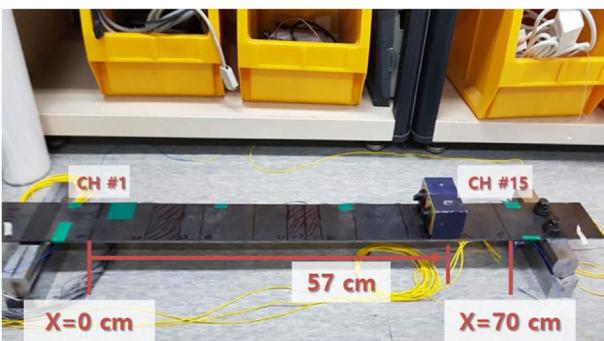
Fig. 13. Test specimen with sensor configuration.

5.1. Experimental study #1: asymmetric strain distribution

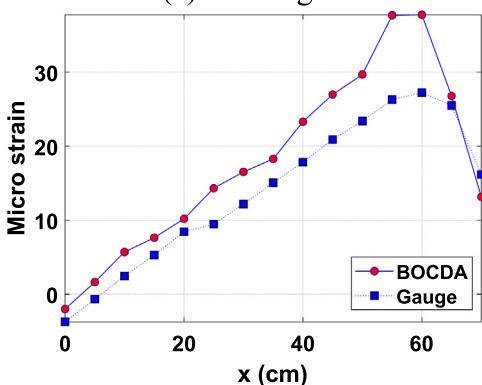
The static loading test was performed using the mass of 1.97 kg as shown in Fig. 14(a). The test was repeated three times under the identical condition in order to show the accuracy and reproducibility of the measured strains. For the three repetitions, the masses were carefully located to the identical location. Fig. 14(b) shows the measured strains from the three repetitions. The measured strains from the BOCDA has the variation of their values with the range of $\pm 10\mu\epsilon$ (LF-data), while those from the strain gauges are less than $\pm 4\mu\epsilon$ (HF-data).

By averaging the measured strains across the three repetitions, the LF-data (Y_{BOCDA}) and HF-data (Y_{gauge}) were obtained as shown in Fig. 14(c). It should be noted that the strains from the BOCDA is expected to be lower than those of the ERC due to spatial resolution of the BOCDA (i.e., near the loading locations at $x = 58\text{cm}$ as shown in Fig. 14(a)). However, the measured strains from the BOCDA are higher than those from the ERC. The measured Brillouin frequency shift contains temperature effects with mechanical strains. The experiments were performed in summer, so that the temperature in the room slightly increased. During this experiment, there was the mistake to forget turning on the air conditioner in order to maintain temperature in the room constantly. As a result, the temperature slightly increased (up to 0.7°C) from the reference condition (i.e., unloading condition) to the loading condition. Considering that the coefficients for strain and temperature are $1\text{MHz}/20\mu\epsilon$ and $1\text{MHz}/1^\circ\text{C}$ for Brillouin frequency shift, $10\mu\epsilon$ is equivalently converted by the increase of 0.5°C . The Brillouin frequency shift increased due to temperature variation during this experiment, so that the measured strains from the BOCDA were higher than those of the ERC. Although the temperature variation results in inaccurate strain distribution from the BOCDA, the trend of the strain distribution from the BOCDA is well matched to those of the ERC. In this context, the multi-fidelity data-fusion methods can be evaluated whether they are successful to ensure the accurate strain distributions. The cross-correlation between LF- and HF-data is shown in Fig. 14(d). It is observed that Y_{BOCDA} seems to be linearly correlated with Y_{gauge} ($Y_{BOCDA} > Y_{gauge}$); however, this trend ($Y_{BOCDA} > Y_{gauge}$) abruptly changes to the adverse trend ($Y_{BOCDA} < Y_{gauge}$) at the sub-region of $x > 60\text{cm}$. This indicates that the cross-correlation between LF- and HF-data are input-dependent (i.e., position) due to the location of the loading.

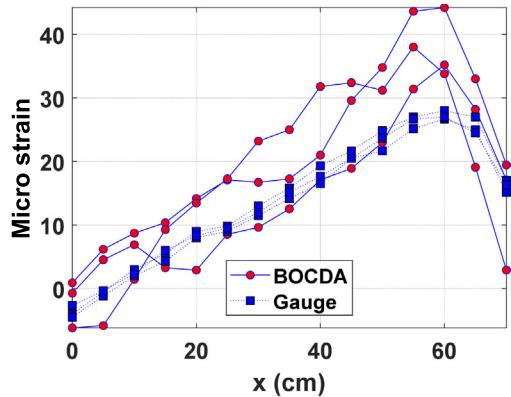
As tabulated in Table 6, the seven case studies with different number of the HF-data were performed. The NRMSE and NMAE were calculated using the whole HF-data ($N_{HF} = 15$) and the predictions. Fig. 15 shows the boxplots of the NRMSE and NMAE over 100 MC runs. It is observed that all methods provide better prediction performances as the number of



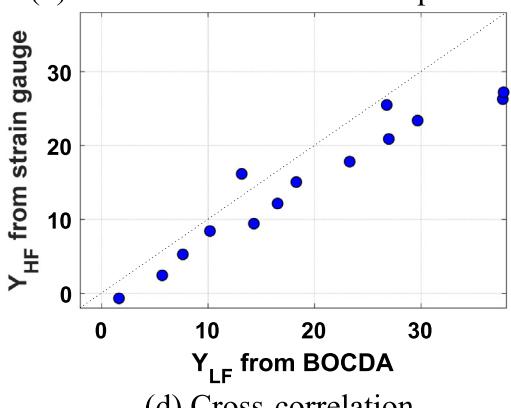
(a) Loading test



(c) Mean strain



(b) Measured strains with 3 repetitions



(d) Cross-correlation

Fig. 14. Loading test and measured strains: asymmetric strain distribution.

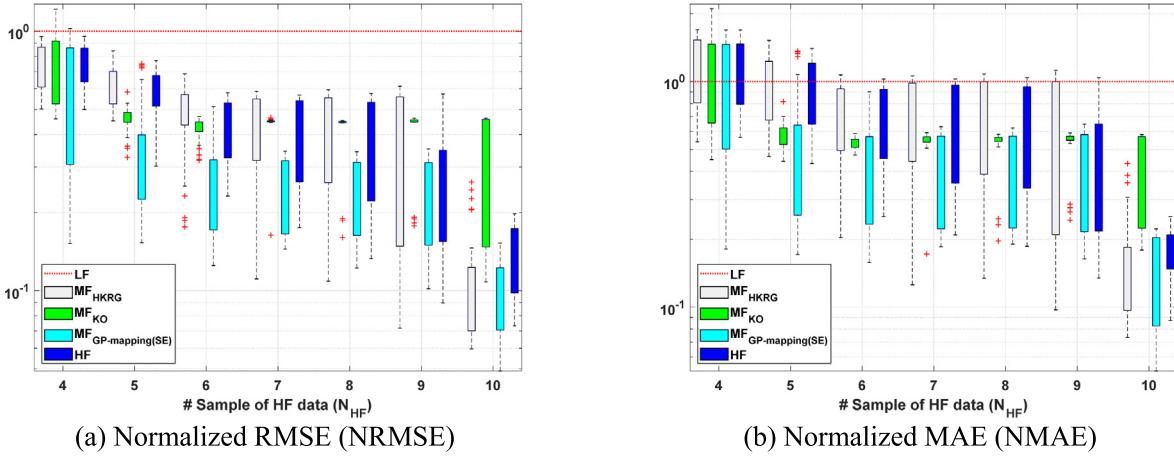


Fig. 15. Boxplots for performance measures over 100 repetitions.

the HF-data increases. It is worth noting that the prediction performances of the **HF** are generally superior to those of the **MF_{HKRG}** and **MF_{KO}** after $N_{HF} = 7$. Although the cross-correlation between Y_{BOCDA} and Y_{gauge} are linearly correlated over the most regions ($x < 60cm$), the **MF_{KO}** cannot perform well due to the local change of the input-dependency at $x = 60cm$. The **MF_{HKRG}** are slightly better than the **MF_{KO}** after $N_{HF} = 7$, despite the larger variabilities in the prediction performances. On the other hand, for all case studies, the prediction performances from the **MF_{GP}** are better than other methods in terms of global and local accuracy.

The best realizations of the HF-predictions are illustrated with that of the **LF** in Fig. 16. Based on Fig. 16(a), the **MF_{GP}** provides the most accurate HF-predictions among the methods with only five number of the HF-data. Both the **MF_{HKRG}** and **MF_{KO}** produce relatively inaccurate HF-predictions over the sub-region of $x > 50cm$ by failing to learn the abrupt change of the cross-correlation, while the **HF** cannot capture such local change of the HF-data due to the insufficient data ($N_{HF} = 5$). In Fig. 16(b), the three methods (**HF**, **MF_{KO}** and **MF_{GP}**) provide the similar prediction performance under $N_{HF} = 7$. As the number of HF-data increases ($N_{HF} = 9$ in Fig. 16(c)), the prediction performances from the **HF**, **MF_{HKRG}** and **MF_{GP}** are very similar. However, the prediction performance of the **MF_{KO}** is worse than that of using $N_{HF} = 7$. Although the number of the HF-data is sufficient to construct the accurate prediction, the **MF_{KO}** is not successful to produce the similar or better predictions than those of the HF-data. Stated differently, the linear mapping may fail to fuse the information from the LF- and HF-data properly. Although the **MF_{HKRG}** exhibits the large variabilities as shown in Fig. 15, it shows the performance improvements with more HF-data. On the other hand, the input-connected GP mapping in the **MF_{GP}** can capture the local changes of the input-dependency and provides better predictions using the less HF-data (herein, $N_{HF} = 5$).

5.2. Experimental study #2: symmetric strain distribution

To produce a symmetric strain distribution, the masses of 5.922 kg was located on the center of the test specimen as shown in Fig. 17. The loading test was performed three times, and the measured strains from the BOCDA and strain gauge are shown in Fig. 17(b). In order to minimize the temperature variation during this experiment, the air conditioner was turned on to maintain the room temperature constantly. Similarly to Experimental study #1, the measured strains from the BOCDA has relatively larger variability than those of the strain gauge due to the low SNR and averaging effect (5cm of the spatial resolution in BOCDA). As shown in Fig. 17(c), the LF-data (Y_{BOCDA}) and HF-data (Y_{gauge}) were obtained by averaging the measured strains. The cross-correlation between the LF- and HF-data is shown in Fig. 17(d). The mean values of the strain from the BOCDA for $x < 40cm$ are larger than those of the strain gauges. For $x \geq 40cm$, the averaged strains from BOCDA and strain gauge are well matched. Therefore, the input dependency in the cross-correlation was changed based on the location of the loading.

Similarly to Experimental study #1, the influence on the number of the HF-data and their locations was also investigated through the seven case studies in Table 6. The boxplots of the NRMSE and NMAE over 100 MC runs are illustrated in Fig. 18. The **MF_{GP}** outperforms the other methods for both NRMSE and NMAE. It is observed that the prediction performances from the **MF_{KO}** and **HF** are slightly improved with more HF-data. Contrary to Experimental study #1, the prediction performances of the **MF_{HKRG}** are worse than those of the **HF**.

Fig. 19 illustrates best realizations of the HF-predictions from **HF**, **MF_{HKRG}**, **MF_{KO}** and **MF_{GP}** with that of the **LF**. The **MF_{GP}** provides slightly better HF-predictions than the **MF_{HKRG}** and **MF_{KO}** for all cases. In Fig. 19(a), the **HF** has the relatively large errors over $20 \leq x \leq 55cm$ due to the insufficient HF-data ($N_{HF} = 5$). The **MF_{HKRG}** provides similar results of the **MF_{KO}**. The only **MF_{GP}** can produce very accurate HF-predictions even with the insufficient HF-data. Fig. 20(b) shows that the **MF_{GP}** and **MF_{KO}** provide similar prediction performances. However, the **MF_{HKRG}** slightly improves the performance than its results in

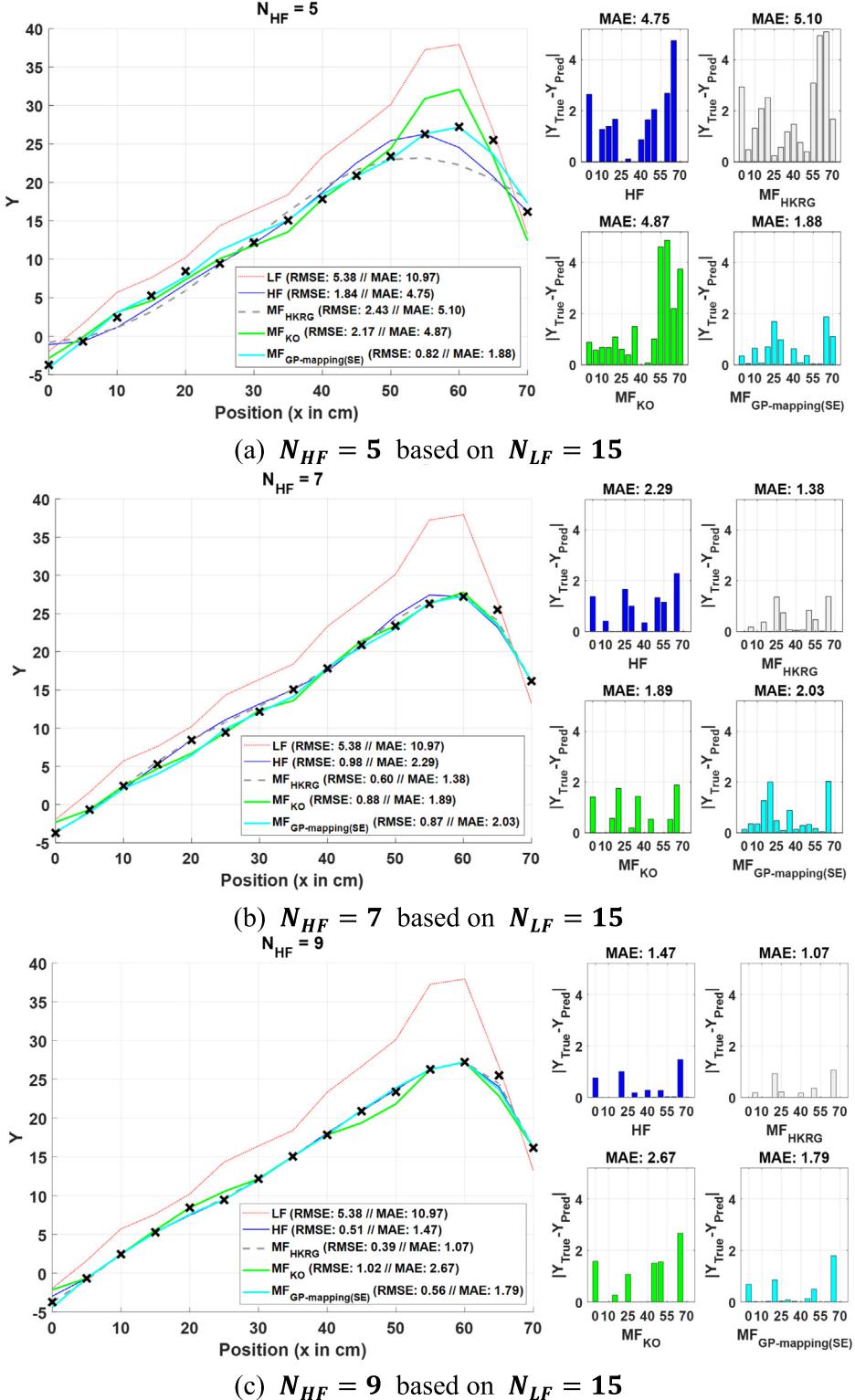
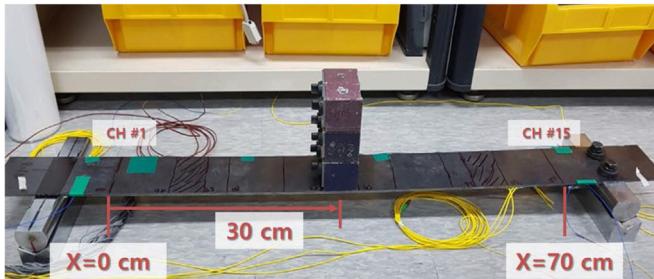
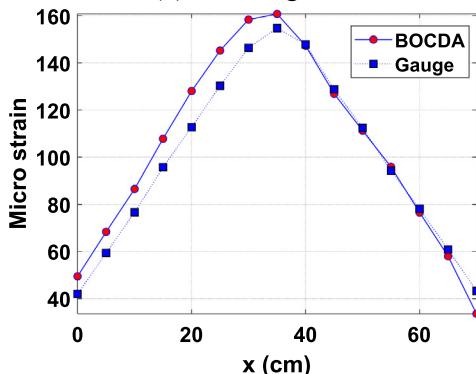


Fig. 16. Predictions from the four methods with $N_{HF} = 5, 7, 9$ based on $N_{LF} = 15$

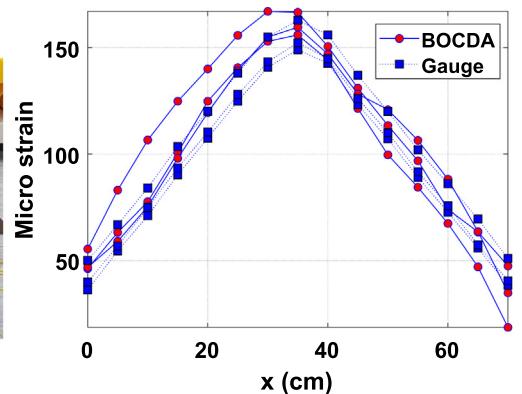
Fig. 20(a). On the other hand, the relatively large errors in the **HF** are observed over $35 \leq x \leq 50$ cm. When using $N_{HF} = 9$ (Fig. 20(c)), the HF-predictions from the three methods (**HF**, MF_{KO} and MF_{GP}) are sufficiently accurate for the true HF-data, except for the MF_{HKRG} .



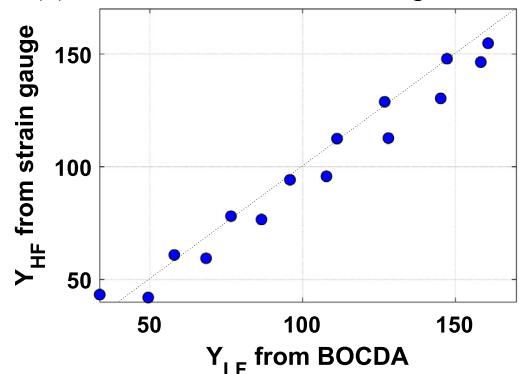
(a) Loading test



(c) Mean strain

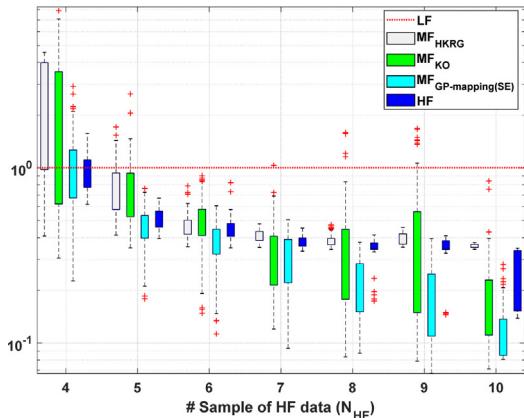


(b) Measured strains with 3 repetitions

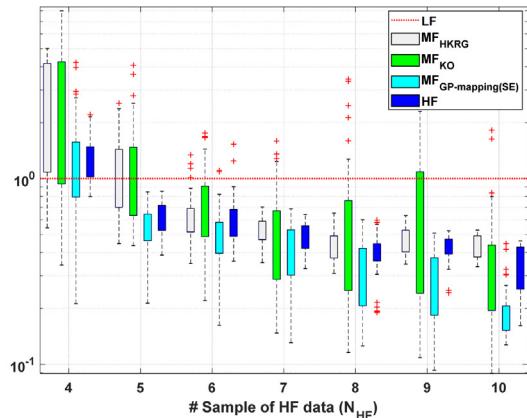


(d) Cross-correlation

Fig. 17. Static test and measured strains: symmetric strain distribution.



(a) Normalized RMSE (NRMSE)



(b) Normalized MAE (NMAE)

Fig. 18. Boxplots for performance measures over 100 repetitions.

The **MF_{KO}** can predict the true HF-data accurately with more HF-data, since the linear mapping in the **MF_{KO}** is still admissible to transfer the information from the LF-data in this experiment. However, the **MF_{HKRG}** does not show reasonable performances in this experiment (contrary to Experiment #1), although it utilized the LF-predictions as trend of the HF-predictions. On the other hand, the prediction performance of the **MF_{GP}** still outperforms the **HF**, **MF_{HKRG}** and **MF_{KO}**. This indicates that the **MF_{GP}** can learn various cross-correlations between the LF- and HF-data flexibly with less HF-data.

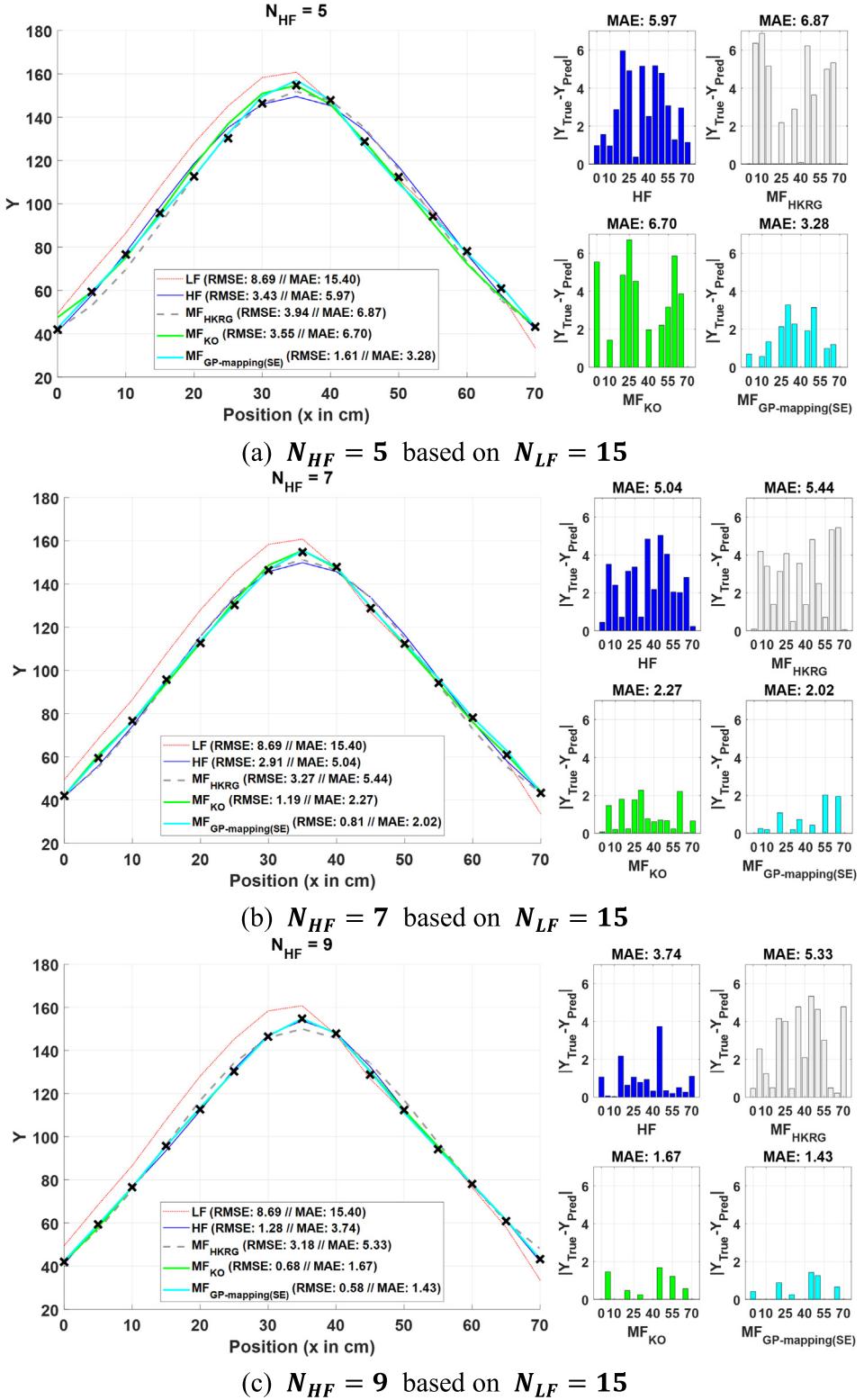


Fig. 19. Predictions from the four methods with $N_{HF} = 5, 7, 9$ based on $N_{LF} = 15$.

6. Discussion on proposed method

Based on the numerical and experimental studies, it can be concluded that the proposed method (MF_{GP}) is successful to learn the various types of the cross-correlation for multi-fidelity data-fusion. As a result, the MF_{GP} produces more accurate HF-predictions than other methods (HF , MF_{HKRG} and MF_{KO}) with less HF-data. It is observed that the MF_{HKRG} performed reasonably well in Sections 5.1; however, it performed worse than the HF in Section 5.2. Similar results were observed in the MF_{KO} . The MF_{KO} performed reasonably well in Section 5.2, while it did not perform well in Section 5.1. This indicates that its learning capability is limited and problem-dependent. Considering that the structural systems are exposed to various conditions, the different types of the strain distributions can be realized (as seen in Section 5). As a result, different cross-correlations can be generated even for the identical structural systems. In this context, the MF_{GP} is more promising to deal with various cross-correlations flexibly and efficiently for the complementary data-fusion. We provide the discussion on the MF_{GP} in terms of potential improvement and computational issues. Then, practical issues on field applications are summarized.

6.1. Potential improvement of proposed method

As seen in Section 4 and 5, it is observed that the MF_{GP} sometimes provides larger variabilities of the prediction performance over other methods. However, the MF_{GP} provides much lower values of both RMSE and MAE than the other methods. These lower values imply that the MF_{GP} is successful to transfer the information from the LF-data to the HF-prediction. Stated differently, the MF_{GP} has more chances to provide HF-predictions that are more accurate than other methods. The larger variabilities in the MF_{GP} can be alleviated by selecting the optimal location of the HF-data. In this context, pursuing optimal sensor placement is required as further study.

6.2. Computational issues on proposed method

One challenging issue in computational demands is to model the GP under large number of the LF-data. Because standard GP (also Kriging) requires solving a large matrix of linear equations (i.e., inversion and determinant of the covariance matrix), its computational complexity exponentially increases according to the sample size ($O(n^3)$). In this context, the standard GP limits its scalability and becomes unaffordable for large-scale datasets [78]. In the proposed data-fusion framework, the distributed optical fiber sensor (DOFS) can generate numerous LF-data. Some DOFS (e.g., BOCDA) have the high spatial resolution (e.g., cm) for the long distance of fiber under test (e.g., km). For example, the BOCDA generates 20,000 samples with the spatial resolution of 5 cm when measuring the optical fiber of 1 km (e.g., massive infrastructures such as long-span bridges). As a result, it is unaffordable to implement the standard GP for large LF-data. To address the computational complexity under large-scale datasets, scalable GP algorithms [79,80] can be adopted for the MF_{GP} . The adoption of the scalable GP algorithm is our on-going work for the proposed data-fusion framework.

6.3. Practical issues for field implementation

Since the DOFS systems using Rayleigh and Brillouin scattering are dependent to both mechanical strain and temperature variation, the temperature effect cannot be negligible. If the temperature varies during the monitoring, the temperature effect on measurements is a potential problem for field applications. The simple way of removing the temperature effect on strain is to use the dummy optical fiber that is not bonded to the infrastructures. Since the dummy optical fiber measures the strain caused by the temperature variation, the mechanical strain can be extracted by subtracting the measurement of the dummy optical fiber from the measurement of the optical fiber bonded with the infrastructures.

7. Conclusions

Over the past decades, various researches have been conducted to develop an efficient and reliable monitoring system. The structural monitoring system utilizes sensors to measure the response of the structural system, and the measured response is used to evaluate the structural capability (structural integrity) to retain its intended purpose through continuous monitoring. Especially in civil structures (infrastructure), some important challenges exist as follows: (1) the target structures such as bridge and dam are massive, so that dense deployment of the sensors results in the complexity of overall system and total costs; and (2) prior knowledge for the critical points is not known in advance, so that the sensors are deployed densely under allowable resources. The starting point for the successful structural monitoring is to obtain accurate responses over the structural system.

In this context, the distributed strain sensor is the promising sensing technique to obtain very dense strains using the quasi-continuous sensing. However, due to its scattering-based sensing mechanism, the distributed strain sensor typically lacks the accuracy and reproducibility of the measured strain. On the other hand, point strain sensors such as strain gauge and FBG can provide high accurate and reproducible strains at discrete measurement positions. The point strain sensor provides a few samples of accurate high-fidelity data, while the distributed strain sensor produce a large number of cheap and

inaccurate low-fidelity data. Although they can be complemented each other, their complementary data-fusion using point and distributed strain sensors has not been studied yet for the structural monitoring system.

The complementary data-fusion can be achieved by the traditional multi-fidelity method using simple linear auto-regressive mapping. However, the traditional multi-fidelity method is not efficient to fuse the point and distributed strain sensors under complex cross-correlations. To overcome the limitation of the traditional method, this study developed an efficient complementary data-fusion method based on the input-connected GP mapping for the structural monitoring system.

To demonstrate the proposed method, the numerical and experimental studies were performed by considering the number of the HF-data and their locations. The prediction performance of the proposed method is consistently superior to other methods using less HF-data. In this context, the proposed method can allow for the flexible modeling of the cross-correlation, so that the proposed method can learn the various cross-correlations between the point and distributed strain sensor under different loading cases. The further studies and computational issues on the proposed method are discussed with the practical issues for field applications. The proposed method can be readily extended to other problems having multi-fidelity data (e.g., FBG sensor and Raman-based distributed temperature sensor for monitoring the temperature).

CRediT authorship contribution statement

Seung-Seop Jin: Conceptualization, Methodology, Experiment, Software, Validation, Supervision, Writing - original draft, Writing - review & editing, Funding acquisition. **Sung Tae Kim:** Methodology, Experiment. **Young-Hwan Park:** Conceptualization, Experiment.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1C1C1009236).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ymssp.2021.107725>.

References

- [1] K. Worden, C.R. Farrar, G. Manson, G. Park, The fundamental axioms of structural health monitoring, *P. R. Soc. A* 463 (2007) 1639–1664.
- [2] S.S. Jin, H.J. Jung, Vibration-based damage detection using online learning algorithm for output-only structural health monitoring, *Struct. Health Monit.* 17 (2018) 727–746.
- [3] S. Jang, H. Jo, S. Cho, K. Mechitov, J.A. Rice, S.H. Sim, H.J. Jung, C.B. Yun, B.F. Spencer, G. Agha, Structural health monitoring of a cable-stayed bridge using smart sensor technology: deployment and evaluation, *Smart Struct. Syst.* 6 (2010) 439–459.
- [4] M.S. Safizadeh, S.K. Latifi, Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell, *Inform. Fusion* 18 (2014) 1–8.
- [5] G. Luyckx, E. Voet, N. Lammens, J. Degrieck, Strain measurements of composite laminates with embedded fibre Bragg Gratings: criticism and opportunities for research, *Sensors-Basel* 11 (2011) 384–408.
- [6] S. Minakuchi, Y. Okabe, T. Mizutani, N. Takeda, Barely visible impact damage detection for composite sandwich structures by optical-fiber-based distributed strain measurement, *Smart Mater. Struct.* 18 (2009).
- [7] S. Minakuchi, H. Tsukamoto, N. Takeda, Detecting water accumulation in honeycomb sandwich structures by optical-fiber-based distributed temperature measurement, *J. Intel. Mat. Syst. Str.* 20 (2009) 2249–2255.
- [8] N. Tanaka, Y. Okabe, N. Takeda, Temperature-compensated strain measurement using fiber Bragg grating sensors embedded in composite laminates, *Smart Mater. Struct.* 12 (2003) 940–946.
- [9] M.D. Todd, J.M. Nichols, S.T. Trickey, M. Seaver, C.J. Nichols, L.N. Virgin, Bragg grating-based fibre optic sensors in structural health monitoring, *Philos. T. R. Soc. A* 365 (2007) 317–343.
- [10] F. Li, W.G. Zhao, H.B. Xu, S.P. Wang, Y.L. Du, A Highly integrated BOTDA/XFG sensor on a single fiber for simultaneous multi-parameter monitoring of slopes, *Sensors-Basel* 19 (2019).
- [11] J. Alvarez-Montoya, A. Carvajal-Castrillon, J. Sierra-Perez, In-flight and wireless damage detection in a UAV composite wing using fiber optic sensors and strain field pattern recognition, *Mech. Syst. Signal. Pr.* 136 (2020).
- [12] T.L. Li, M.Y. Liu, R.Y. Li, Y. Liu, Y.G. Tan, Z.D. Zhou, FBG-based online monitoring for uncertain loading-induced deformation of heavy-duty gantry machine tool base, *Mech. Syst. Signal. Pr.* 144 (2020).
- [13] J. Frieden, J. Cugnoni, J. Botsis, T. Gmur, D. Coric, High-speed internal strain measurements in composite structures under dynamic load using embedded FBG sensors, *Compos. Struct.* 92 (2010) 1905–1912.
- [14] A. Guemes, A. Fernandez-Lopez, B. Soller, Optical fiber distributed sensing - physical principles and applications, *Struct. Health Monit.* 9 (2010) 233–245.
- [15] J.H. Jeong, K. Lee, K.Y. Song, J.M. Jeong, S.B. Lee, Differential measurement scheme for Brillouin optical correlation domain analysis, *Opt. Express* 20 (2012) 27094–27101.
- [16] A. Zadok, Y. Antman, N. Primerov, A. Denisov, J. Sancho, L. Thevenaz, Random-access distributed fiber sensing, *Laser Photonics Rev.* 6 (2012) L1–L5.
- [17] K. Hotate, Fiber distributed Brillouin sensing with optical correlation domain techniques, *Opt. Fiber Technol.* 19 (2013) 700–719.

- [18] A. Smyth, M. Wu, Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring, *Mech. Syst. Signal. Pr.* 21 (2007) 706–723.
- [19] E. Chatzi, C. Fuggini, Online correction of drift in structural identification using artificial white noise observations and an unscented Kalman Filter, *Smart Struct. Syst.* 16 (2015) 295–328.
- [20] J.-W. Park, K.-C. Lee, S.-H. Sim, H.-J. Jung, B.F. Spencer Jr., Traffic safety evaluation for railway bridges using expanded multisensor data fusion, *Comput.-Aided Civ. Infrastruct. Eng.* 31 (2016) 749–760.
- [21] A. Downey, M. Sadoughi, S. Laflamme, C. Hu, Fusion of sensor geometry into additive strain fields measured with sensing skin, *Smart Mater. Struct.* 27 (2018) 075033.
- [22] J.-W. Park, D.-S. Moon, H. Yoon, F. Gomez, B.F. Spencer Jr., J.R. Kim, Visual–inertial displacement sensing using data fusion of vision-based displacement with acceleration, *Struct. Control Health Monit.* 25 (2018) e2122.
- [23] G. Ravizza, R. Ferrari, E. Rizzi, E. Chatzi, Effective Heterogeneous Data Fusion procedure via Kalman filtering, *Smart Struct. Syst.* 22 (2018) 631–641.
- [24] M. Sadoughi, A. Downey, J. Yan, C. Hu, S. Laflamme, Reconstruction of unidirectional strain maps via iterative signal fusion for mesoscale structures monitored by a sensing skin, *Mech. Syst. Signal. Pr.* 112 (2018) 401–416.
- [25] M.F. Bado, G. Kaklauskas, J.R. Casas, Performance of Distributed Optical Fiber Sensors (DOFS) and Digital Image Correlation (DIC) in the monitoring of RC structures, *IOP Conf. Ser.: Mater. Sci. Eng.* 615 (2019) 012101.
- [26] J. He, X. Zhang, B. Xu, XF-based multiscale response reconstruction under unknown inputs with data fusion of multitype observations, *J. Aerosp. Eng.* 32 (2019) 04019038.
- [27] R.-T. Wu, M.R. Jahanshahi, Data fusion approaches for structural health monitoring and system identification: past, present, and future, *Struct. Health Monit.* 19 (2020) 552–586.
- [28] H. Zhu, K. Gao, Y. Xia, F. Gao, S. Weng, Y. Sun, Q. Hu, Multi-rate data fusion for dynamic displacement measurement of beam-like supertall structures using acceleration and strain sensors, *Struct. Health Monit.* 19 (2020) 520–536.
- [29] M.F. Bado, J.R. Casas, A. Dey, C.G. Berrocal, Distributed optical fiber sensing bonding techniques performance for embedment inside reinforced concrete structures, *Sensors (Basel)* 20 (2020) 5788.
- [30] J.Z. Li, L.X. Xu, K. Kishida, FBG-based positioning method for BOTDA sensing, *IEEE Sens. J.* 16 (2016) 5236–5242.
- [31] J.P. He, Z. Zhou, J.P. Ou, Simultaneous measurement of strain and temperature using a hybrid local and distributed optical fiber sensing system, *Measurement* 47 (2014) 698–706.
- [32] C.E. Rasmussen, C.K.I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Mass., Cambridge, 2006.
- [33] E. Schulz, M. Speekenbrink, A. Krause, A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions, *J. Math. Psychol.* 85 (2018) 1–16.
- [34] D.K. Duvenaud, O. Rippel, R.P. Adams, Z. Ghahramani, Avoiding pathologies in very deep networks, *Proceedings of the 34th International Conference on Machine Learning*, JMLR, Reykjavik, Iceland, 2014, pp. 202–210.
- [35] A.C. Damianou, N.D. Lawrence, Deep Gaussian Processes, *International Conference on Artificial Intelligence and Statistics (AISTATS)*, Scottsdale, AZ, USA, 2013.
- [36] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, *J. Roy. Stat. Soc. B* 63 (2001) 425–450.
- [37] L. Le Gratiet, J. Garnier, Recursive co-kriging model for design of computer experiments with multiple levels of fidelity, *Int. J. Uncertain. Quan.* 4 (2014) 365–386.
- [38] P. Perdikaris, D. Venturi, G.E. Karniadakis, Multifidelity information fusion algorithms for high-dimensional systems and massive data sets, *SIAM J. Sci. Comput.* 38 (2016) B521–B538.
- [39] P. Perdikaris, M. Raissi, A. Damianou, N.D. Lawrence, G.E. Karniadakis, Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling, *P. R. Soc. A* 473 (2017).
- [40] D.A. Krohn, *Fiber Optic Sensors: Fundamentals and Applications*, 3rd ed., ISA, Research Triangle, NC, 2000.
- [41] A.H. Hartog, An introduction to distributed optical fibre sensors.
- [42] A. Barrias, J.R. Casas, S. Villalba, A review of distributed optical fiber sensors for civil engineering applications, *Sensors (Basel)* 16 (2016) 748.
- [43] C. Du, S. Dutta, P. Kurup, T. Yu, X. Wang, A review of railway infrastructure monitoring using fiber optic sensors, *Sens. Actuators, A* 303 (2020) 111728.
- [44] K. Soga, L. Luo, Distributed fiber optics sensors for civil engineering infrastructure sensing, *J. Struct. Integrity Maint.* 3 (2018) 1–21.
- [45] Z. Ding, C. Wang, K. Liu, J. Jiang, D. Yang, G. Pan, Z. Pu, T. Liu, Distributed optical fiber sensors based on optical frequency domain reflectometry: a review, *Sensors-Basel* 18 (2018) 1072.
- [46] L. Shihol, A. Eyal, Distributed acoustic and vibration sensing via optical fractional Fourier transform reflectometry, *Opt. Express* 23 (2015) 4296–4306.
- [47] B.J. Soller, D.K. Gifford, M.S. Wolfe, M.E. Froggatt, High resolution optical frequency domain reflectometry for characterization of components and assemblies, *Opt. Express* 13 (2005) 666–674.
- [48] M. Froggatt, D. Gifford, S. Kreger, M. Wolfe, B. Soller, Distributed strain and temperature discrimination in unaltered polarization maintaining fiber, *Optical Fiber Sensors*, Cancun (2006), pp. ThC5.
- [49] Z. Zhang, X. Bao, Distributed optical fiber vibration sensor based on spectrum analysis of Polarization-OTDR system, *Opt. Express* 16 (2008) 10240–10247.
- [50] H. He, L.-Y. Shao, B. Luo, Z. Li, X. Zou, Z. Zhang, W. Pan, L. Yan, Multiple vibrations measurement using phase-sensitive OTDR merged with Mach-Zehnder interferometer based on frequency division multiplexing, *Opt. Express* 24 (2016) 4842–4855.
- [51] A. Masoudi, M. Belal, T.P. Newson, A distributed optical fibre dynamic strain sensor based on phase-OTDR, *Meas. Sci. Technol.* 24 (2013) 085204.
- [52] Z. Ding, X.S. Yao, T. Liu, Y. Du, K. Liu, Q. Han, Z. Meng, J. Jiang, H. Chen, Long measurement range OFDR beyond laser coherence length, *IEEE Photonics Technol. Lett.* 25 (2013) 202–205.
- [53] L. Luo, B. Li, Y. Yu, X. Xu, K. Soga, J. Yan, Time and frequency localized pulse shape for resolution enhancement in STFT-BOTDR, *J. Sens.* 2016 (2016) 3204130.
- [54] F. Wang, C. Zhu, C. Cao, X. Zhang, Enhancing the performance of BOTDR based on the combination of FFT technique and complementary coding, *Opt. Express* 25 (2017) 3504–3513.
- [55] D. Meng, F. Ansari, X. Feng, Detection and monitoring of surface micro-cracks by PPP-BOTDA, *Appl. Opt.* 54 (2015) 4972–4978.
- [56] A. Minardo, A. Coscetta, L. Zeni, R. Bernini, High-spatial resolution DPP-BOTDA by real-time balanced detection, *IEEE Photonics Technol. Lett.* 26 (2014) 1251–1254.
- [57] V. Varma, Bhabha Atomic Research Centre., Health monitoring of civil structures using fiber optic sensors, Bhabha Atomic Research Centre, Mumbai, 2003.
- [58] A. Rogers, Distributed optical-fibre sensing, *Meas. Sci. Technol.* 10 (1999) R75–R99.
- [59] J. Kadum, C. Feng, T. Schneider, Characterization of the noise induced by stimulated Brillouin scattering in distributed sensing, *Sensors (Basel)* 20 (2020).
- [60] R. Regier, N.A. Hoult, Distributed strain behavior of a reinforced concrete bridge: case study, *J. Bridge Eng.* 19 (2014) 05014007.
- [61] Y. Bao, G. Chen, W. Meng, F. Tang, Y. Chen, Kilometer-Long Optical Fiber Sensor for Real-Time Railroad Infrastructure Monitoring to Ensure Safe Train Operation, 2015 Joint Rail Conference, 2015.
- [62] V.P. Matveenko, I.N. Shardakov, A.A. Voronkov, N.A. Kosheleva, D.S. Lobanov, G.S. Serovaev, E.M. Spaskova, G.S. Shipunov, Measurement of strains by optical fiber Bragg grating sensors embedded into polymer composite material, *Struct. Control Health Monit.* 25 (2018) e2118.
- [63] K. Cho, S.T. Kim, Y.-H. Park, J.-R. Cho, Measurement of mechanical and thermal strains by optical FBG sensors embedded in CFRP rod, *J. Sens.* 2019 (2019) 5345901.

- [64] Z. Ugray, L. Lasdon, J. Plummer, F. Glover, J. Kelly, R. Martí, Scatter search and local NLP solvers: a multistart framework for global optimization, *INFORMS J. Comput.* 19 (2007) 313–484.
- [65] P. Perdikaris, G.E. Karniadakis, Model inversion via multi-fidelity Bayesian optimization: a new paradigm for parameter estimation in haemodynamics, and beyond, *J. R. Soc. Interface* 13 (2016) 20151107.
- [66] A.I.J. Forrester, A. Sóbester, A.J. Keane, Multi-fidelity optimization via surrogate modelling, *Proc. R. Soc. A: Math., Phys. Eng. Sci.* 463 (2007) 3251–3269.
- [67] C. Park, R.T. Haftka, N.H. Kim, Remarks on multi-fidelity surrogates, *Struct. Multidiscip. Optim.* 55 (2017) 1029–1050.
- [68] H. Liu, Y.-S. Ong, J. Cai, Y. Wang, Cope with diverse data structures in multi-fidelity modeling: a Gaussian process method, *Eng. Appl. Artif. Intell.* 67 (2018) 211–225.
- [69] H. Babaee, P. Perdikaris, C. Chrysostomidis, G.E. Karniadakis, Multi-fidelity modelling of mixed convection based on experimental correlations and numerical simulations, *J. Fluid Mech.* 809 (2016) 895–917.
- [70] M. Raissi, P. Perdikaris, G. EmKarniadakis, Inferring solutions of differential equations using noisy multi-fidelity data, *J. Comput. Phys.* 335 (2017) 736–746.
- [71] B. Khaleghi, A. Khamis, F.O. Karay, S.N. Razavi, Multisensor data fusion: a review of the state-of-the-art, *Inform Fusion* 14 (2013) 28–44.
- [72] S.-S. Jin, Compositional kernel learning using tree-based genetic programming for Gaussian process regression, *Struct. Multidiscip. Optim.* 62 (2020) 1313–1351.
- [73] S.-S. Jin, Accelerating Gaussian Process surrogate modeling using Compositional Kernel Learning and multi-stage sampling framework, *Appl. Soft Comput.* 106909 (2020).
- [74] I. Abdallah, C. Lataniotis, B. Sudret, Parametric hierarchical kriging for multi-fidelity aero-servo-elastic simulators – application to extreme loads on wind turbines, *Probab. Eng. Mech.* 55 (2019) 67–77.
- [75] S. Marelli, B. Sudret, UQLab: A Framework for Uncertainty Quantification in Matlab, Vulnerability, Uncertainty, and Risk, 2014, pp. 2554–2563.
- [76] C. Lataniotis, S. Marelli, B. Sudret, The Gaussian process modelling module in UQLab, *J. Soft Comput. Civ. Eng.* 2 (2018) 91–116.
- [77] M.D. McKay, R.J. Beckman, W.J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* 42 (2000) 55–61.
- [78] H. Liu, Y. Ong, X. Shen, J. Cai, When Gaussian process meets big data: a review of scalable GPs, *IEEE Trans. Neural Networks Learn. Syst.* 31 (2020) 4405–4423.
- [79] E. Snelson, Z. Ghahramani, Sparse Gaussian processes using pseudo-inputs, in: Proceedings of the 18th International Conference on Neural Information Processing Systems, MIT Press, Vancouver, British Columbia, Canada, 2005, pp. 1257–1264.
- [80] J. Quiñonero-Candela, C.E. Rasmussen, A unifying view of sparse approximate Gaussian process regression, *J. Mach. Learn. Res.* 6 (2005) 1939–1959.