Smart Structures and Systems, *Vol. 32, No. 4 (2023) 267-279*
https://doi.org/10.12989/sss.2023.32.4.267

Towards high-accuracy data modelling, uncertainty quantification and correlation analysis for SHM measurements during typhoon events using an improved most likely heteroscedastic Gaussian process

Qi-Ang Wang*[1], Hao-Bo Wang [1b], Zhan-Guo Ma [1a], Yi-Qing Ni [2a], Zhi-Jun Liu**[1], Jian Jiang [1a], Rui Sun [1b] and Hao-Wei Zhu [1b]

[1] *State Key Laboratory of Intelligent Construction and Healthy Operation & Maintenance of Deep Underground Engineering & School of Mechanics and Civil Engineering, China University of Mining and Technology, Xuzhou 221008, China*
[2] *National Rail Transit Electrification and Automation Engineering Technology Research Center (Hong Kong Branch) and Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong*

(*Received December 1, 2022, Revised August 5, 2023, Accepted October 12, 2023*)

Abstract. Data modelling and interpretation for structural health monitoring (SHM) field data are critical for evaluating structural performance and quantifying the vulnerability of infrastructure systems. In order to improve the data modelling accuracy, and extend the application range from data regression analysis to out-of-sample forecasting analysis, an improved most likely heteroscedastic Gaussian process (iMLHGP) methodology is proposed in this study by the incorporation of the out-of-sample forecasting algorithm. The proposed iMLHGP method overcomes this limitation of constant variance of Gaussian process (GP), and can be used for estimating non-stationary typhoon-induced response statistics with high volatility. The first attempt at performing data regression and forecasting analysis on structural responses using the proposed iMLHGP method has been presented by applying it to real-world filed SHM data from an instrumented cable-stay bridge during typhoon events. Uncertainty quantification and correlation analysis were also carried out to investigate the influence of typhoons on bridge strain data. Results show that the iMLHGP method has high accuracy in both regression and out-of-sample forecasting. The iMLHGP framework takes both data heteroscedasticity and accurate analytical processing of noise variance (replace with a point estimation on the most likely value) into account to avoid the intensive computational effort. According to uncertainty quantification and correlation analysis results, the uncertainties of strain measurements are affected by both traffic and wind speed. The overall change of bridge strain is affected by temperature, and the local fluctuation is greatly affected by wind speed in typhoon conditions.

Keywords: data modelling; improved most likely heteroscedastic Gaussian process; structural health monitoring; typhoons; uncertainties

1. Introduction

In both communities of industry and academia, structural health monitoring (SHM) has raised growing concerns in the interest of guaranteeing the reliability of infrastructure systems and providing early warnings on structural damage or deterioration to avoid a catastrophic accident. The past two decades witnessed increasing study dedicated to structural health monitoring and structural health monitoring techniques has evolved from diagnosis to prognosis on the basis of monitoring data (Porter *et al*. 2002, Ju *et al*. 2011, Lee *et al*. 2014, Parida *et al*. 2020, Papadimitriou 2004, Kopsaftopoulos and Fassois 2011, Brownjohn 2007, Wang *et al*. 2022a, 2018a). The measured strain responses are increasingly used in structural state forecasting and diagnosis because they can directly manifest the safety state of the structural components or derive information on the bearing capacity of the whole structure (Ko and Ni 2005, Ni *et al*. 2009, Wan and Ni 2019a, b). Therefore, accurate data modelling for structural strain responses has become a critical step for the dependable diagnosis and prognosis of structural conditions.

Many cases show that structural health conditions can be estimated by analyzing their strain response (Wang *et al*. 2023, Herbko *et al*. 2022, Braunfelds *et al*. 2022, Glisic 2022, Theiler *et al*. 2014, Khan *et al*. 2021, Wang *et al*. 2022b). The strain evolution is a typical nonlinear dynamic process for large and complex structures subjected to live load and environmental load (Locke *et al*. 2020, Mohammadi *et al*. 2007), especially during extreme events (such as typhoons, floods, and geologic hazards). Assessing the strain responses of bridges during those extreme events is critical to the design, repair, maintenance, and other decisions. The typhoon event is one of the important causes

*Corresponding author, Ph.D., Professor,
E-mail: qawang@mail.nwpu.edu.cn
**Co-corresponding author, Ph.D., Professor,
E-mail: liuzhijun0331@cumt.edu.cn
[a] Ph.D.
[b] Graduate Student

ISSN: 1738-1584 (Print), 1738-1991 (Online)

of bridge damage. Monitoring the strain of bridges during typhoon events and processing the recorded strain data for structural performance evaluation has become a research hotspot. And, there is a strong need to develop an efficient and accurate data modelling and analysis method for structural strain responses.

Gaussian process (GP) has become one of the most practical tools to solve the all-kind problem in the field of machine learning (Agou *et al*. 2022, Iba and Akaho 2010), the main reason is that for nonlinear problems the Bayesian framework can be used to solve the problems such as data training, model selection, forecasting. And another reason is that the GP problem can be resolved through a relatively simple linear algebra method (Talaei and Ma 2007). The limitation in the GP model is the assumption that the noise is constant throughout the input space. However, in actual problems, the noise is always changing (Skolidis and Sanguinetti 2011, Urban *et al*. 2015), that is, the noise variance is different, so the heteroscedastic Gaussian process (HGP) appears (Mousavi and Gandomi 2021, Bludszuweit *et al*. 2009) to solve this kind of problem. Traditional HGP algorithms employ another GP to define the noise variance, but this process can only obtain an approximate solution through Markov Chain Monte Carlo (MCMC) sampling or analysis rather than an exact solution through learning or analysis (Munoz-Gonzalez *et al*. 2016, Kersting 2007). While most likely heteroscedastic Gaussian process (MLHGP) replaces the full posterior distribution of the changing noise by a point estimation at the most likely value, and in this manner, the predictive posterior over the test output can be evaluated analytically. Thus, the MLHGP framework is computationally attractive in dealing with regression issues in the situation of monitoring data with input-dependent noises compared with other traditional HGP algorithms (Binois and Gramacy 2021). In addition, data regression and forecasting methods also include Bayesian algorithms (Wright 2008, Baesens *et al*. 2002, Crawford *et al*. 2019), support vector machines (SVM) (Subasi 2013, Yoon and Kim 2010), Neural Network (Majkovic *et al*. 2016, Partal 2017), and other machine learning methods (Sierra-Garcia and Santos 2021, Von Krannichfeldt *et al*. 2007).

Despite the rapid advances in data science at SHM including the aforementioned machine learning algorithms, the specific goals of data modelling and analysis still face two typical challenges (Wan *et al*. 2022, Wan and Ni 2019b, Wang *et al*. 2022c, Wang *et al*. 2018b). (1) Firstly, relatively few studies have considered structural response forecasting under extreme events such as typhoon events. The structural response data under such extreme events fluctuates greatly and has strong randomness, which makes regression and forecasting more complex and difficult to deal with. (2) In addition, the accuracy and robustness of the research using the Gaussian process for structural response forecasting are insufficient, and the data processing method of heteroscedasticity needs to be further investigated.

In order to resolve these two problems, this study proposes an improved Most Likely Heteroscedastic Gaussian Process (iMLHGP) methodology, which calculates the maximum probability in the posterior distribution of variance and incorporates an out-of-sample forecasting algorithm. Compared with HGP algorithms, the proposed iMLHGP method can improve the data modelling accuracy for estimating non-stationary typhoon-induced response statistics with high volatility, and extend the application range from data regression analysis to out-of-sample forecasting analysis. The typhoon-induced strain responses of bridges are typical non-stationary data, and the proposed iMLHGP algorithm is especially suitable for data analysis for this kind of structural response. The main contributions of this article are as follows: Firstly, to the best of our knowledge, the proposed iMLHGP is initially used to perform regression analysis on the strain response of the bridge during typhoon events. Secondly, in this paper, the MLHGP algorithm is improved by incorporating an out-of-sample forecasting algorithm, which would extend its application range from regression analysis to out-of-sample forecasting analysis. In addition, the regression and out-of-sample forecasting performance of GP and iMLHGP were analyzed and compared in detail. Finally, by quantifying uncertainty and analyzing correlation, the influence of typhoon action on the bridge strain response is fully discussed and investigated.

The rest of the paper is organized as follows: Section 2 describes the formula derivation and improvement of GP, HGP, and iMLHGP; In the third section, the experimental verification results of the iMLHGP method are introduced in detail. In sections 3.2 and 3.3, GP and iMLHGP are used to perform regression and forecasting of structural strain response in typhoon and non-typhoon conditions, and the differences between the two methods are analyzed. In sections 4.1 and 4.2, the strain data uncertainties were quantified through the confidence interval width obtained by iMLHGP, and correlations in SHM data were investigated. Section 5 concludes the current works and the prospect of future directions.

2. An improved MLHGP methodology

An improved Most Likely Heteroscedastic Gaussian Process (iMLHGP) methodology is proposed in this study for high-accuracy data modelling and analysis of strain data with high volatility. The proposed iMLHGP is initially used to perform regression, forecasting and correlation analysis on the strain response of the bridge during typhoon events. The specific data modelling procedure using the proposed iMLHGP method is shown in Fig. 1.

2.1 Heteroscedastic Gaussian Process

Since GP framework employs Bayesian method to solve model learning, model selection, density estimation, etc, it can effectively solve the nonlinear regression problem, and also can deal with the over-fitting problem well (Zhang and Ni 2020). The purpose of using GP to perform regression on monitoring data is to obtain a function F, which corresponds to the data in each monitoring data set S to the data in time series T. Many uncertain factors in the actual monitoring process will cause the error of monitoring data. The

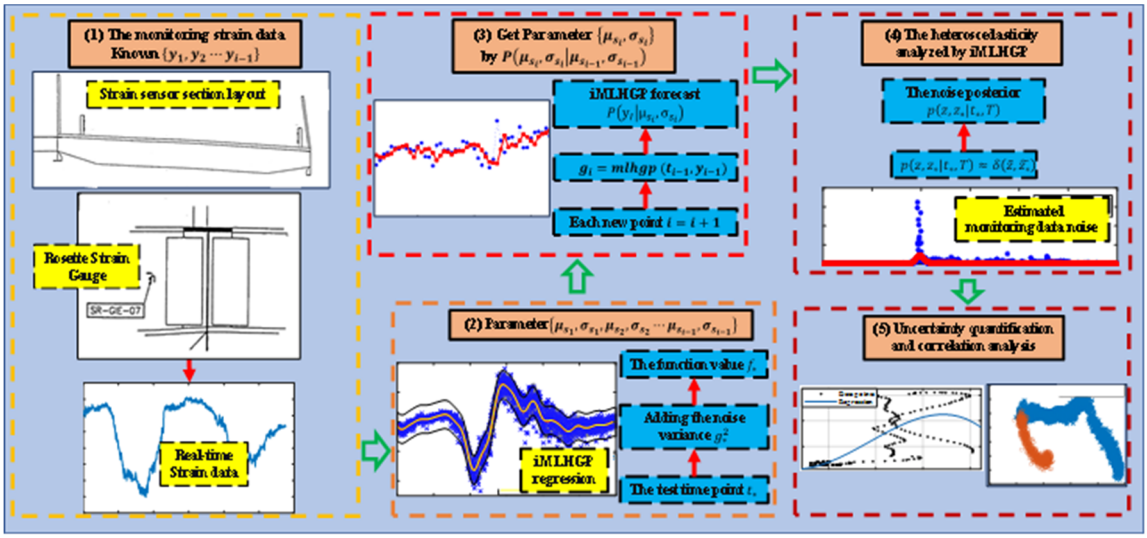

Fig. 1 Technical overview of SHM data modelling enabled by the iMLHGP model

monitoring error can be expressed by ε. The basic formula of the GP process can be expressed as follows

$$s_i = f(t_i) + \varepsilon_i \tag{1}$$

where the observation error ε_i is usually assumed to be an independent normal distribution in the GP with the zero mean value and the variance is g_i. The set of the variances g_i is G.

The Gaussian process algorithm belongs to a kind of nonparametric Bayesian modeling method for an unknown function, which can be defined through a mean function $m(t)$ and a covariance function $k(t, t')$. A simplified premise is placing prior information with a zero-mean GP on the function value, given as follows

$$p(f|t) = N(0, K) \tag{2}$$

where K is the covariance matrix, and k_{ij} is calculated at t_i and t_j by the covariance function $k(t_i, t_j)$. Generally, the squared exponential kernel is usually used as the prior form of matrix $k(t_i, t_j)$, expressed by

$$k(t_i, t_j) = \eta^2 exp\left[\left\|t_i - t_j\right\|^2 / (2l^2)\right] \tag{3}$$

where $\|\cdot\|$ is the Euclidean distance between s_i and s_j, η is the signal amplitude, and l is the characteristic length scale. η and l are measures of the similarity between the two monitoring values, which can be expressed by the parameter $\theta_f = \{\eta, l\}$. The joint Gaussian distribution of the monitoring value s and the function f_* at the time point t_* can be expressed as

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K + Q & k_* \\ k_*^T & k_{**} \end{bmatrix}\right) \tag{4}$$

where Q is the noise variance diagonal matrix $[Q]_{ii} = g_*^2$, which is assumed to be the independent normal distribution. k_* is the covariance vector between the training time point t_i and the test time point t_* obtained from the covariance function and k_{**} is the prior variance calculated from the covariance function at t_*. The posterior distribution for f_* at t_* would be calculated through the conditional identity of multivariate Gaussian distribution.

$$p(f_*|t_*, \theta_f, g, T, S) = N(\mu_{f_*}, \sigma_{f_*}^2), where \tag{5}$$

$$\mu_{f_*} = k_*^T (K + Q)^{-1} s \tag{6}$$

$$\sigma_{f_*}^2 = k_{**} - k_*^T (K + Q)^{-1} k_* \tag{7}$$

A posteriori distribution of the test output s_* can be obtained by adding the noise variance g_*^2 to the function value f_* at the test time point t_* as

$$p(s_*|t_*, \theta_f, g, T, S) = N(\mu_{s_*}, \sigma_{s_*}^2), where \tag{8}$$

$$\sigma_{s_*}^2 = \sigma_{f_*}^2 + g_*^2 = k_{**} - k_*^T (K + Q)^{-1} k_* + g_*^2 \tag{9}$$

Both the kernel parameter θ_f and the noise level $g(t)$ need to be learned from the data. In the GP, the noise level is constant throughout the input time series, so the noise $g^2(t) \equiv \sigma_n^2$ and the noise matrix $Q \equiv \sigma_n^2 I$ can be derived. The unknown kernel parameter θ_f and the noise matrix σ_n^2 are the hyperparameter $\theta_s = \{\theta_f, \sigma_n^2\}$ of the GP model and can be resolved by maximizing the logarithmic marginal likelihood.

$$\begin{aligned} log\, p(s|T, \theta_s) = &-\frac{1}{2} s^T (K + \sigma_n^2 I)^{-1} \\ &-\frac{1}{2} log|K + \sigma_n^2 I| - \frac{n}{2} log(2\pi) \end{aligned} \tag{10}$$

This process is known as the type II maximum likelihood (ML-II) estimate of the hyperparameters θ_t, which can be obtained by optimizing algorithms (Rasmussen and Nickisch, 2010) that pursue a local

maximum with an acceptable range or a possible global optimum.

If the heteroscedastic data is applied to the homoscedasticity model, the model will select an average noise level as the noise matrix σ_n^2. GP regression in the data with obvious heteroscedasticity will lead to a decrease in the accuracy of regression. Some improvement measures for the GP method can improve the accuracy of regression.

The HGP requires the estimation of noise at each point. An independent Gaussian process is placed for the logarithms of the data noises (Goldberg 1998), denoted as $z(t) = \log\sigma(t)$. This z-process employs a different set of hyperparameters θ_z and a covariance function k_z. In this manner, the noise levels σ and σ^* are expressed through a probabilistic model. Therefore, the HGP predictions would be evaluated through marginalization

$$p(s_*|t_*,S,T) = \iint p(s_*|t_*,z,z_*,S,T)\,p(z,z_*|t_*,T)dzdz_* \tag{11}$$

The analytic solution of this integral is difficult to obtain. Generally, the Monte Carlo sampling method is used to extract samples $\{(z_1,z_{1*}),\dots,(z_k,z_{2*})\}$ from $p(z,z_*|t_*,T)$ and approximate the integral by $\frac{1}{k}\sum_{j=1}^{k} p\left(s_*|x_*,z_j,z_{j*},T,S\right)$. This sampling method has the disadvantages of the huge amount of calculation and time-consuming. By comparison, the MLHGP framework is simple and computationally attractive to handle data modelling problems with varying noise. In the MLHGP framework, a point estimation on the most likely value is adopted to replace the posterior distribution of the input-dependent noises, and thus the predictive posterior would be analytically solved. The specific process of MLHGP will be introduced in the following chapters.

2.2 An improved most likely heteroscedastic Gaussian Process regression

The MLHGP algorithm replaces the posterior distribution of the input-dependent noises with the point estimation on the most likely value, which greatly improves the computational ability to handle data modelling problems with varying noise for computation. In order to further improve the accuracy, and extend its application range from regression analysis to out-of-sample forecasting analysis. This study proposed an improved MLHGP framework by incorporating an out-of-sample forecasting method.

In the proposed iMLHGP framework, the posterior of noise $p(z,z_*|t_*,T)$ is approximated as

$$\begin{aligned} &p(z,z_*|t_*,T) \approx \delta(\tilde{z},\widetilde{z_*}) \\ &\delta(\tilde{z},\widetilde{z_*}) = 1, when\ \tilde{z} = \widetilde{z_*}: \\ &\delta(\tilde{z},\widetilde{z_*}) = 0, \text{otherwise} \end{aligned} \tag{12}$$

where $(\tilde{z},\widetilde{z_*})$ is the most likely log-noise level, and δ is the Dirac delta function. Thus, Eq. (11) can be approximated as

$$\begin{aligned} p(s_*|t_*,S,T) &\approx \iint p(s_*|t_*,z,z_*,S,T)\,\delta(\tilde{z},\widetilde{z_*})dzdz_* \\ &\approx p(s_*|t_*,\tilde{z},\widetilde{z_*}S,T) \end{aligned} \tag{13}$$

The most likely noise level is given by a noise posteriori

$$(\tilde{z},\widetilde{z_*}) = arg\max_{(\tilde{z},\widetilde{z_*})} log\,p(\tilde{z},\widetilde{z_*}|t_*,\theta_z,S,T) \tag{14}$$

The input noise is also given by GP, and the posteriori is normally distributed. The most likely noise level can be expressed as

$$(\tilde{z},\widetilde{z_*}) = \left(\mu_z,\mu_{z_*}\right) \tag{15}$$

where μ_z and μ_{z_*} represent the posterior mean of noise at training time point t and test time point t_*, respectively. The integral in Eq. (11) can be rewritten as

$$\begin{aligned} p\left(s_*|t_*,\theta_f,\theta_z,S,T\right) &\approx p\left(s_*|x_*,\theta_f,\tilde{z},\widetilde{z_*},S,T\right) \\ &= p\left(s_*|x_*,\theta_f,\mu_z,\mu_{z_*},S,T\right) \end{aligned} \tag{16}$$

Thus, the mean and variance of the predictive posterior distribution of the test output are given by Eqs. (8) and (9), which fit the Gaussian distribution.

The key point of the iMLHGP algorithm is the evaluation of the most likely noise level. The main methods used to estimate the noise level in the present MLHGP framework are illustrated in the following equation

$$g_i^2 = \frac{1}{q}\sum_{j=1}^{q} 0.5\left(s_i - s_i^j\right)^2 \tag{17}$$

where q is the size of the sample, and s_i^j are samples from

Algorithm 1 Improved MLHGP Method with Forecasting Function Based on SHM Data

input : Initial training data S= $\{\boldsymbol{T},\boldsymbol{Y}\}$ with $\boldsymbol{T} = \{\boldsymbol{t_i}\}|_{\boldsymbol{i=1}}^{\boldsymbol{n-1}}$ and Y= $\{\boldsymbol{y_i}\}|_{\boldsymbol{i=1}}^{\boldsymbol{n-1}}$, $\boldsymbol{\mu_0}$, $\boldsymbol{\sigma_0}$, $\boldsymbol{g_0}$, $\boldsymbol{q}$. *i* represents the index number of data point.

output : Forecasts of $\boldsymbol{y_{n-1},\cdots,y_N}$ at the times $\boldsymbol{t_{n-1},\cdots,t_N}$ in terms of mean vector $\boldsymbol{\mu}$ and variance vector $\boldsymbol{\sigma}$.

function mlhgp (t, y)
  for Each new point observed **do**
$\boldsymbol{i = i+1}$
$\boldsymbol{g_i = mlhgp\ (t_{i-1}, y_{i-1})}$
  update $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ such as $\boldsymbol{\mu = \mu \cup \mu_i, \sigma = \sigma \cup \sigma_i}$.
  update S= $\{\boldsymbol{T},\boldsymbol{Y}\}$ with $\boldsymbol{T = [T, t_i]\ and\ Y = [Y, y_i]}$
end for
end function

from the posterior predictive distribution of the training output s_i, which is Gaussian distribution with mean μ_{s_*} and variance $\sigma_{s_*}^2$ given in Eqs. (8) and (9), respectively.

The forecasting algorithm of iMLHGP is developed in this study to extend the application range for out-of-sample SHM data forecasting. The iMLHGP time series forecasting approach is detailed in Algorithm 1. This paper adopts the iMLHGP algorithm to process the real-time monitoring data obtained by the SHM system.

3. iMLHGP data modelling for strain measurements in typhoon conditions

In this section, in order to demonstrate the performance and accuracy of the iMLHGP data modeling method in typhoon conditions, the iMLHGP algorithm was applied to the real-world field monitoring data from the Ting Kau Bridge in Hong Kong. Specifically, iMLHGP is employed for strain monitoring data regression and forecasting in different external excitation conditions, including typhoon load and non-typhoon conditions.

Fig. 2 Ting Kau Bridge

3.1 Ting Kau Bridge with the sophisticated SHM system

Ting Kau Bridge is a three-tower cable-stayed bridge with a total span of 1,177 m from the northwest of Tsing Yi Island and Tuen Mun Road. Completed in 1998, the bridge serves as a primary connector between the Hong Kong International Airport and the rest of Hong Kong. Ting Kau Bridge features three bridge towers with heights of 170 m, 194 m, and 158 m, located on the Ting Kau headland, on a reclaimed island in Rambler Channel, and on the northwest of Tsing Yi shoreline, respectively. The three towers are provided with sufficient strength and stiffness by steel beams and transverse stability cables located under the deck of the bridge. Meanwhile, the main tower is equipped with longitudinal stability cables to connect the deck of the two pylons to improve the stability of the whole bridge structure. The decks of the Ting Kau Bridge are located on both sides of the towers, which can improve the structural

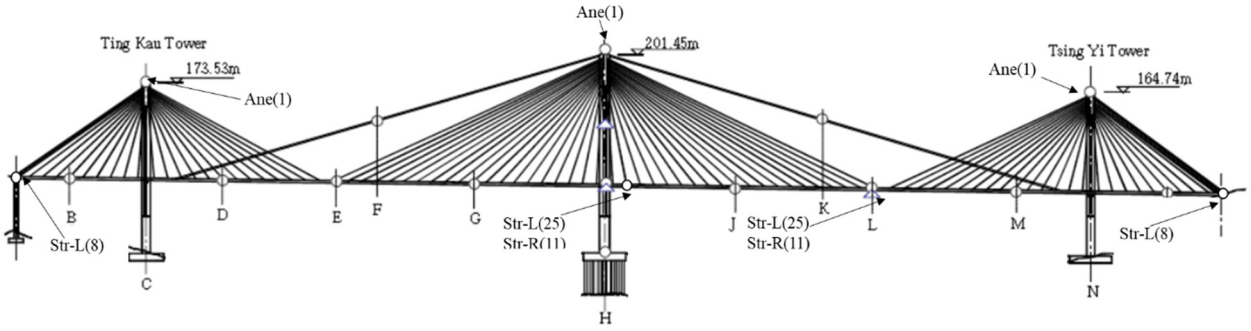

Fig. 3 Layout of anemometers and strain sensors

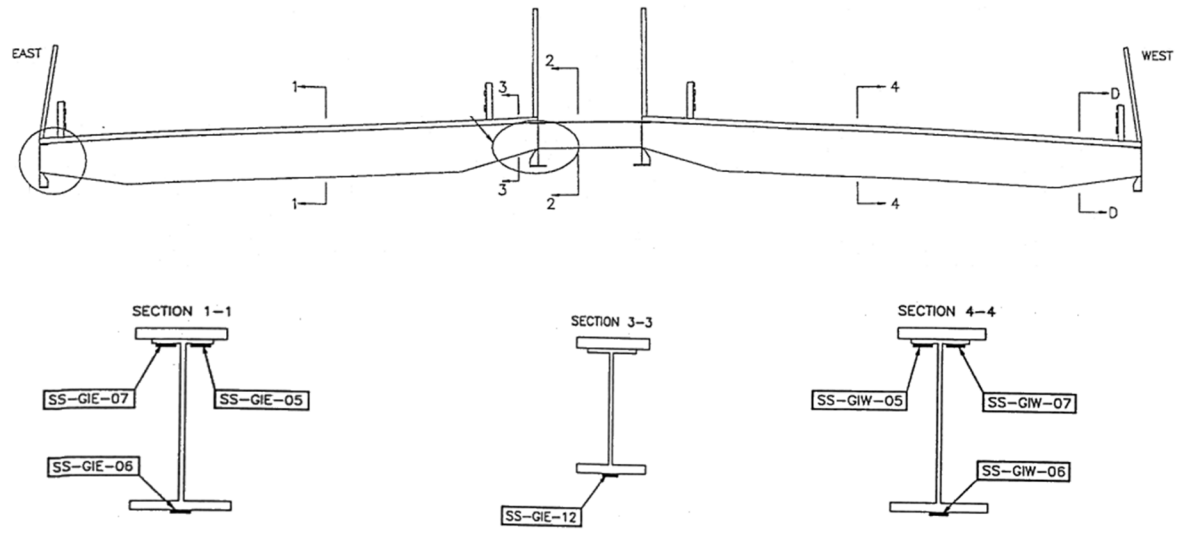

Fig. 4 Strain gauge arrangement of the L-L cross section

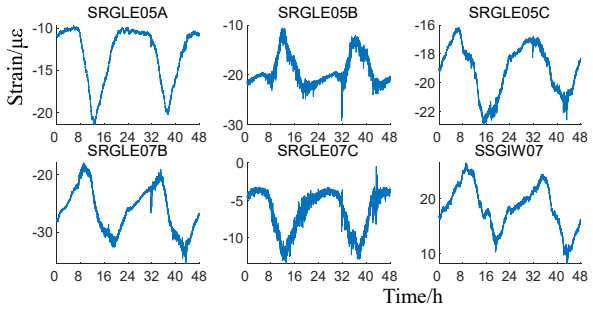

Fig. 5 48-hour SHM strain field data on 13-14 Oct. (non-typhoon)

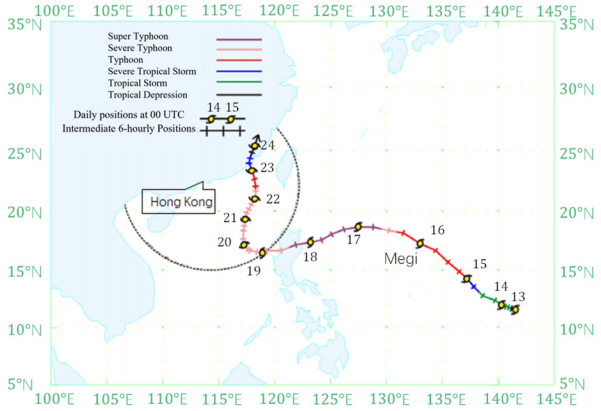

Fig. 6 Track of Megi (1013) on 13-24 Oct. 2010

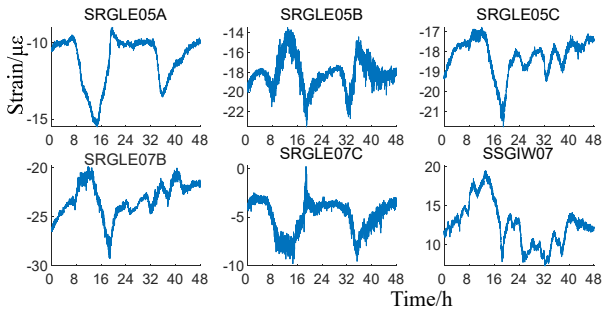

Fig. 7 48-hour SHM strain field data on 21-22 Oct. (typhoon)

wind resistance performance. The stay cables are distributed on four cable planes, which are respectively anchored to the fixed ends of the main girders of the combined bridge deck, and the tensioned ends are located in the steel anchor box at the top of the bridge tower.

The SHM system of Ting Kau Bridge mainly includes accelerometer, anemometer, displacement transducer, strain gauge, temperature sensor, GPS, etc. The monitoring field data analyzed in this study include the strain measurements, wind speeds and temperature data. Totally, there are 66 one-direction dynamic strain gauges and 22 three-direction dynamic strain gauges at Ting Kau Bridge. The data used in this section mainly comes from 12 sensor channels, as show in Figs. 5 and 7. The distribution of the strain sensors is shown in Fig. 3. The sensors data used in this paper are from the strain sensors located on the L-L cross section. L-L section is roughly located in the middle of one span, where the structural responses are relatively large, and the structural responses are more significantly affected by external effects such as vehicle load and wind load, so the sensor data on this cross section is selected for data modelling and analysis. The specific sensor layout for this cross section is shown in Fig. 4. For example, “SECTION 3-3” in Fig. 4 shows the position of the sensor arrangement with channel number SS-GLE-12, which is arranged at the bottom of “SECTION” 3-3 I-beam at the mid-span L-L cross section. The serial prefix SS is the abbreviation of “Strain Single”, and the serial prefix SR is the abbreviation of “Strain Rosette”. “DETAIL 1” shows the location of the sensor layout with channel number SR-GLE-05, which is located at the west edge of “DETAIL 1” at the L-L cross section. Fig. 5 shows the strain response data recorded by the sensor system at the L-L cross section on 13-14 October in non-typhoon condition. The subtitles in Fig.5 represent the sensor channel number. For example, for SRGLE05A, the serial prefix SS is the abbreviation of “Strain Single”, and the serial prefix SR is the abbreviation of “Strain Rosette”. The abbreviation “GLE” in the middle represents the sensor located on the eastern side of the L-L section. Similarly, “GIE” in the middle represents the sensor located on the eastern side of the I-I section. “GIW” in the middle represents the sensor located on the western side of the I-I section. “05” is the serial number of the sensor, and “A” represents one of the directions of the sensor.

In October 2010, the typhoon named Megi hit Hong Kong. According to the statistics by the Hong Kong Observatory, Megi was the only super tropical cyclone over the northwestern Pacific in the year 2010. When Typhoon Megi was approximately 570 km south-southeast away from Hong Kong, “Standby Signal No. 1” was issued on 20 October. “Standby Signal No. 1” is the lowest warning signal for tropical cyclones in Hong Kong. In the Hong Kong area, winds were fresh north to northeasterly, sporadically strong in offshore waters. With Megi moving closer on 21 October at a distance of approximately 480 km southeast of Hong Kong, “Strong Wind Signal No. 3” was released. “Strong Wind Signal No. 3” is a tropical cyclone warning signal with one level higher than “Standby Signal No. 1”. On 22 October, Megi reached 430 km to the east-southeast at 2 a.m. Local winds also diminished gradually and the Strong Wind Signal No. 3 was superseded by the Standby Signal No. 1 at 6:05 p.m. As winds weakened further, all signals were countermanded at 8:40 p.m.

In this study, recorded strain data in two time periods, i.e., 13-14 October (non-typhoon) and 21-22 October (typhoon), with different external load conditions were selected. The strain sensor has a sampling frequency of 51.2 Hz, collecting 184,320 data points per hour. Fig. 7 shows the strain response data recorded by the sensor system at the L-L cross section on 21-22 October when typhoon Megi had the greatest impact on the bridge. By comparing Fig. 5 and Fig. 7, it can be found that the regularity of strain response is stronger under non-typhoon action. Under typhoon, the randomness and uncertainty of strain monitoring data are greater, and the heteroscedastic

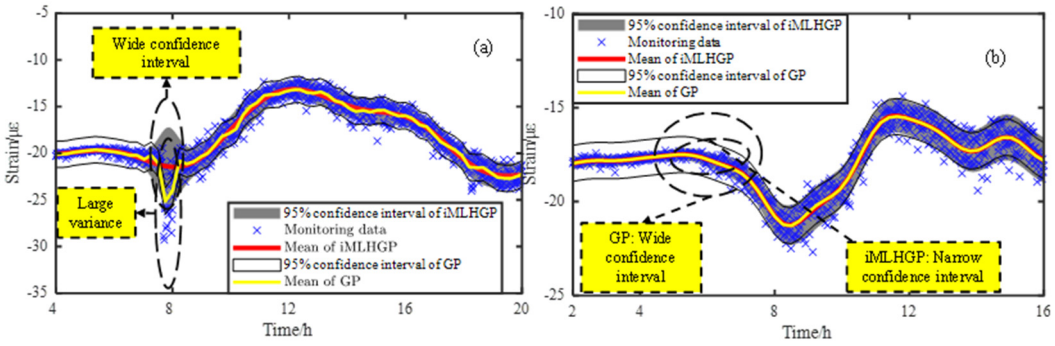

Fig. 8 Regression results for strain monitoring data: (a): non-typhoon conditions (14 Oct.); (b): typhoon conditions (22 Oct.)

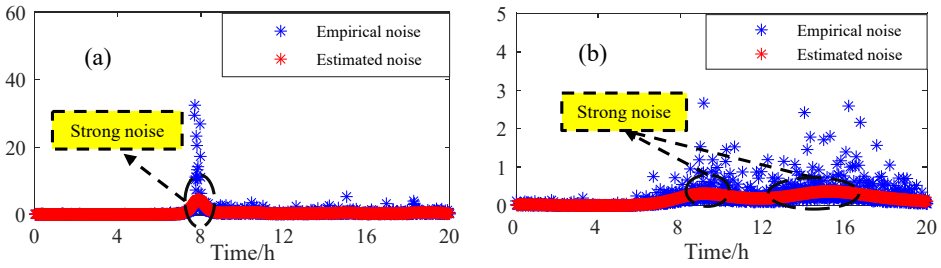

Fig. 9 Data noise and regression error: (a) estimated monitoring data noise (non-typhoon); (b) estimated monitoring data noise (typhoon)

characteristic is more obvious.

3.2 iMLHGP regression performance for typhoon-induced strain responses

In this section, iMLHGP was employed for strain monitoring data regression analysis in both typhoon and non-typhoon conditions. The results are shown in Fig. 8, and the employed strain monitoring data are recorded from SR-GLE-05A on 14 October (non-typhoon) and 22 October (typhoon). In order to reduce the running time of the proposed algorithm, the original data is preprocessed by moving average process.

As shown in Fig. 8, the regression means and 95% confidence interval were obtained by the iMLHGP method. To illustrate the accuracy of the iMLHGP method, the results obtained by the GP method are also provided as a comparison. In non-typhoon conditions, the mean squared error (MSE) of the iMLHGP and the GP are 0.0798 and 0.4865, respectively. And in typhoon conditions, the MSE of the iMLHGP and the GP are 0.1481 and 0.2316, respectively. This indicates that the iMLHGP regression accuracy obtained is higher both for typhoon and non-typhoon conditions. As shown in Figs. 8(a) and (b), the strain data points from 0:00 to 7:00 are more concentrated, and the confidence interval obtained by the iMLHGP algorithm is narrow, while the confidence interval obtained by GP is wider. In Fig. 8(a), at 8:00 for the rush hour, the noise in the data becomes larger and the confidence interval obtained by iMLHGP becomes wider, while the confidence interval obtained by GP does not change. Generally, the iMLHGP confidence interval width varies as the data characteristic changes both for typhoon and non-typhoon conditions, which indicates that iMLHGP can better capture the heteroscedasticity of strain data compared with the GP method.

In addition, the variation of data noise with time is obtained by the iMLHGP algorithm, and the variation of regression error with regard to time is calculated, both for non-typhoon and typhoon conditions. The results are shown in Fig. 9.

Regarding to non-typhoon conditions shown in Fig. 9(a), the data noise around 8:00 is relatively higher and the estimated confidence interval obtained by iMLHGP shown in Fig. 8(a) also gets wider. However, the confidence interval of GP evaluation does not change much, as shown in Fig. 8(a). Thus, the confidence interval of the iMLHGP method can be changed with the variation characteristics of the noise, so it can better reflect the characteristics of the data by the iMLHGP method. As shown in Fig. 9(a) that the main trend of data noise change is affected by the vehicle load. During the period from 0:00 to 7:00 when there is less vehicle traffic on the bridge, the noise in the data is small. During the daytime, the vehicle traffic increases and the noise becomes larger. Especially in the morning around 8:00 for the rush hour, the traffic is heavy, and thus the noise in the data also reached a peak.

For typhoon conditions, as shown in Fig. 9(b), the data noises around 9:00 and 16:00 are relatively higher, and the estimated confidence intervals by iMLHGP shown in Fig. 8(b) get wider. However, the confidence intervals of GP evaluation do not change much, as shown in Fig. 8(b). Typhoon Megi had the greatest impact on Hong Kong on 22 October. Affected by this typhoon, there was almost no vehicle passage on the bridge during this day, so the strain data noise on that day was mainly caused by the typhoon. In

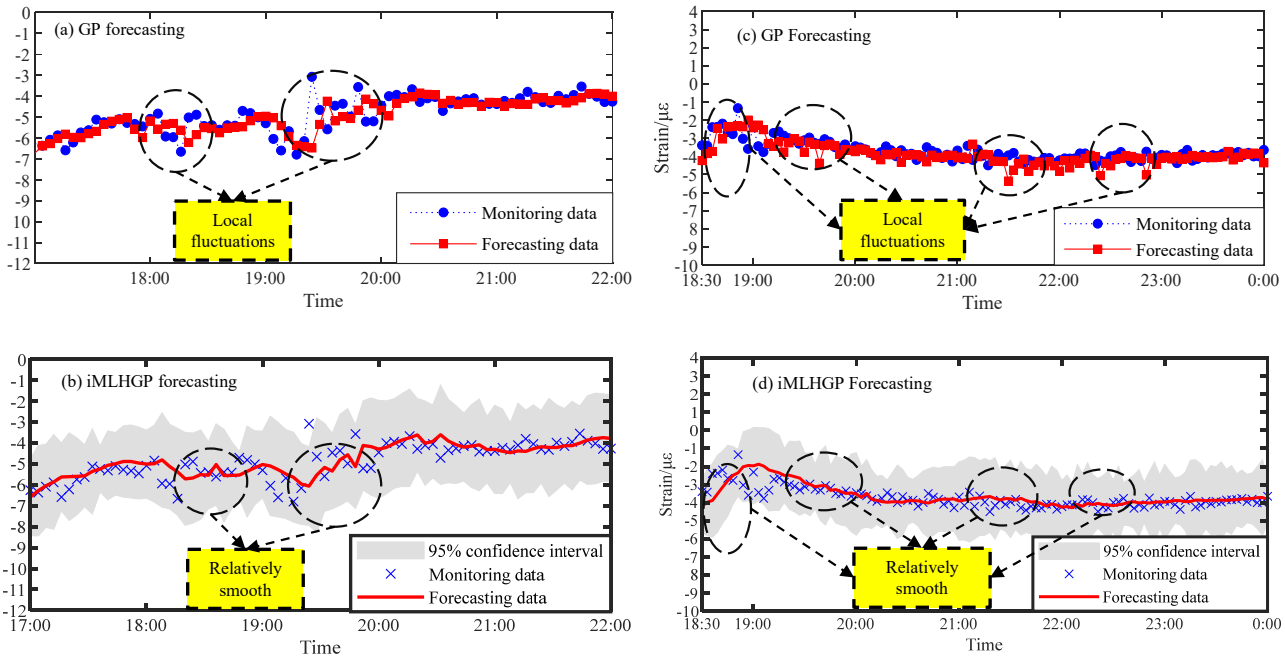

Fig. 10 Forecasting results on the mean for strain monitoring data: (a) GP forecasting (non-typhoon); (b) iMLHGP forecasting (non-typhoon); (c) GP forecasting (typhoon); (d) iMLHGP forecasting (typhoon)

Fig. 9(b), empirical noise represents the actual noise in strain monitoring data and the estimated noise is evaluated through iMLHGP. There are two peaks for the noise around 9:00 and 16:00, respectively. At this time, the confidence interval of iMLHGP regression in Fig. 8(b) is smaller than that of GP regression, which is more consistent with the rule of the noise distribution. After 16:00, the noise gradually decreases, and the confidence interval of iMLHGP regression in Fig. 8(b) also gradually decreases.

3.3. iMLHGP forecasting performance for typhoon-induced strain responses

By incorporating an out-of-sample forecasting algorithm, the iMLHGP would extend its application range from regression analysis to out-of-sample forecasting analysis. In order to verify the forecasting performance of the proposed iMLHGP, the validation tests are carried out based on the strain monitoring data from SR-GLE-07 in both typhoon and non-typhoon conditions, compared with the GP method. And the corresponding results are shown in Fig. 10.

Regarding to non-typhoon conditions shown in Fig. 10(a) and (b), the forecasting curve obtained by the GP algorithm has obvious local fluctuations near 18:00 and 20:00, while the forecasting curve obtained by the iMLHGP algorithm is relatively smooth and is more consistent with the characteristics of the original data. As shown in Table 1, for non-typhoon conditions, the MSE of the iMLHGP algorithm is 0.2976, with a 3.25% improvement in accuracy compared to the GP algorithm (0.3076). Thus, the forecasting accuracy of the iMLHGP method is higher.

For typhoon conditions, as shown in Fig. 10(c) and (d), there are more local fluctuations in the forecasting curve obtained by the GP method, and the iMLHGP forecasting curve remains relatively smooth. The curve fluctuates greatly because the forecasting results are greatly affected by noise. Thus, the iMLHGP forecasting results are less affected by noise. In terms of strain monitoring data, more attention is paid to its overall change trend. The iMLHGP forecasting curve has good smoothness, so it can better reflect the overall trend change. As shown in Table 1, for typhoon conditions, the MSE of the iMLHGP algorithm is 0.1698, which improves the accuracy by 30.61% compared with the GP algorithm (0.2447). Therefore, the percentage improvement in the accuracy of strain response forecasting using the proposed iMLHGP in typhoon conditions is much higher than that using the iMLHGP in non-typhoon conditions. Compared with Fig. 10(b), the confidence interval obtained in Fig. 10(d) has relatively large variations, indicating that the data has heteroscedasticity in typhoon conditions. The structural strain noise under the action of the typhoon is more complex with higher randomness, so the iMLHGP algorithm is more suitable for the data-driven forecasting and analysis of typhoon-induced strain responses.

Table 1 MSEs for GP and iMLHGP forecasting results

	Non-typhoon conditions (14 Oct.)	Typhoon conditions (22 Oct.)
MSE of the iMLHGP	0.3076	0.2447
	0.2976	0.1698

In order to show the data smoothing effect of the iMLHGP forecasting analysis, the training data and forecasting data in typhoon conditions are shown in Fig. 11. As shown in Fig. 11, the first two-thirds of the data are

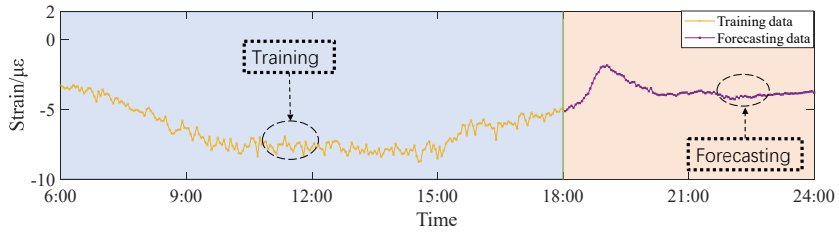

Fig. 11 Illustration of training data and forecasting data

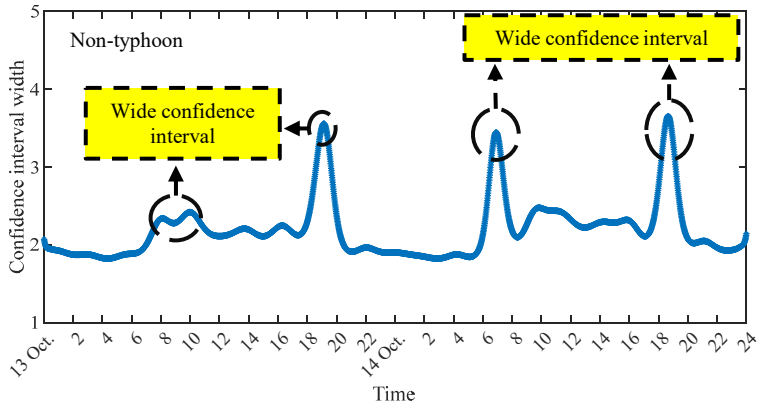

Fig. 12 Strain monitoring data uncertainty quantified by confidence interval width under non-typhoon conditions (13-14 Oct.)

monitoring raw data which are used as training data, and the last third is the iMLHGP forecasting data. It can be seen that there is more fluctuation in the training data curve because of the typhoon-induced noise. The forecasting data curve from the iMLHGP method is smoother, which can better reflect the overall trend of strain changes.

4. Uncertainty quantification and correlation analysis for strain measurements

4.1 Uncertainty quantification through the iMLHGP framework

In the process of regression and forecasting modelling, the data in different periods have different volatility and noise, and thus data uncertainties in different periods vary. In this study, the confidence interval width of the iMLHGP model is used to quantify the data uncertainty. A wide confidence interval indicates strong uncertainty in the monitoring data, and the accuracy of the data modelling will deteriorate. In order to investigate the variation of data uncertainty, this study carried out the regression analysis for strain monitoring data. Specifically, regression analysis on strain monitoring data obtained by the SR-GLE-07 sensor in non-typhoon conditions was carried out. The confidence interval width obtained by this regression can reflect the uncertainty degree of data. The confidence interval width can be obtained by the confidence interval in the iMLHGP regression results obtained in Section 3.2. Specifically, the confidence interval width can be calculated by solving the difference between the upper and lower confidence intervals at the sample point.

Fig. 12 shows the variation of confidence interval width with regard to time for non-typhoon conditions, respectively. As shown in Fig. 12, the width of the confidence interval changes obviously over time. Interesting pattern changes in the confidence interval width can be disclosed when taking a closer look at this figure. Around 8:00 and 19:00 on 13 October and 14 October in non-typhoon conditions, there are some obvious peaks with wide confidence intervals. These periods also coincide with peak traffic hours. While the width of the confidence interval is small from 21:00 on 13 October to 6:00 on 14 October when there are few vehicles during the night and early morning hours. In summary, in the cases of non-typhoon, the uncertainty of data is mainly affected by the vehicle load. However, in typhoon conditions, due to the significant reduction of vehicles, the peak values of the confidence interval in the peak traffic hours are not obvious and prominent, which is smaller than that in the non-typhoon vehicle peak period.

4.2 Correlation Analysis for Strain Measurements

In order to confirm the correlation between the strain data uncertainty and the wind speed in typhoon conditions, the proposed iMLHGP framework is employed to investigate this correlation. The uncertainty of strain monitoring data can be quantified by the confidence interval, so the relationship between confidence interval and wind speed can be used to verify the relationship between the strain data uncertainty and wind speed. The wind speed data and strain data on 21 October in typhoon conditions

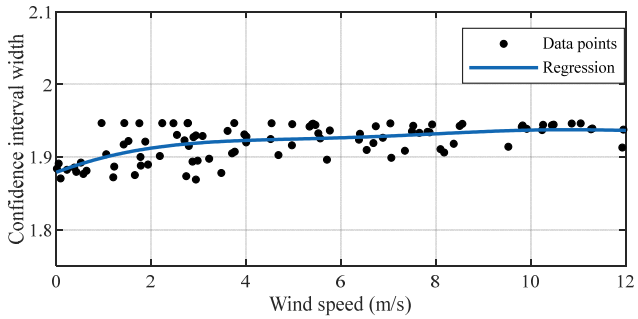

Fig. 13 Correlation between confidence interval and wind speed: data points and regression curves on 21 Oct. (typhoon)

were adopted for correlation analysis. And the corresponding results are shown in Fig. 13. In Fig. 13, the regression curve is obtained through the least square polynomial fitting algorithm on the basis of data points.

Fig. 13 shows the correlation between the wind speed and the confidence interval width in typhoon conditions on 21 October. According to the regression curve in Fig. 13, the width of the confidence interval gradually increases with the increase of the wind speed. And the wind speed was relatively small at that time, which made a weak contribution to the uncertainty of the strain data of the bridge in non-typhoon conditions. It shows that when wind speed is high, the correlation between wind speed and confidence interval width is high. But in non-typhoon conditions, there is no such change rule. The correlation between confidence interval width and wind speed is weak in non-typhoon conditions.

Due to the thermal expansion and cold contraction of building materials, it is critical to investigate the effect of temperature on strain monitoring data. In order to explore the relationship between temperature variation and strain monitoring data, temperature and strain monitoring data points in both non-typhoon and typhoon conditions are investigated for correlation analysis, and the corresponding results are shown in Fig. 14

Fig. 14(a) shows the correlation between the strain measurements and the temperature in non-typhoon conditions from 13:30 on 13 October to 13:30 on 14 October. As the temperature rises, the absolute value of strain increases gradually. As the temperature drops, the absolute value of strain decreases gradually. And in the process of temperature rise and temperature drop, a closed annular curve is formed, indicating that the strain of the structure is in the elastic stage and there is no residual strain under the action of temperature.

During typhoon events, Fig. 14(b) shows the correlation between the strain measurements and the temperature from 15:00 on 21 October to 15:00 on 22 October. As marked in Fig. 14(b), in the stage of “temperature rise”, the absolute value of strain data increases gradually with the temperature. In the stage of “temperature drop”, the absolute value of strain data decreases gradually with the temperature. However, the marked parts of Fig. 14(b) show some local changes in typhoon conditions compared with that in non-typhoon conditions. In these “local change” areas, the correlation between the strain and temperature monitoring data is ambiguous. The main reason maybe that: The typhoon has an influence on the strain responses, resulting in the variation form of strain data in some local change parts, and thus the correlation trend between the strain and temperature monitoring data does not form a closed annular curve compared with that in non-typhoon conditions. In typhoon conditions, the overall change of bridge strain data is mainly affected by the temperature, while the local fluctuation is affected by the typhoon event.

5. Conclusions

The initial attempt at performing data modelling and analysis on the SHM measurements during typhoon events using the iMLHGP method has been presented in this study. Based on the original regression function, the iMLHGP method incorporates an out-of-sample forecasting method

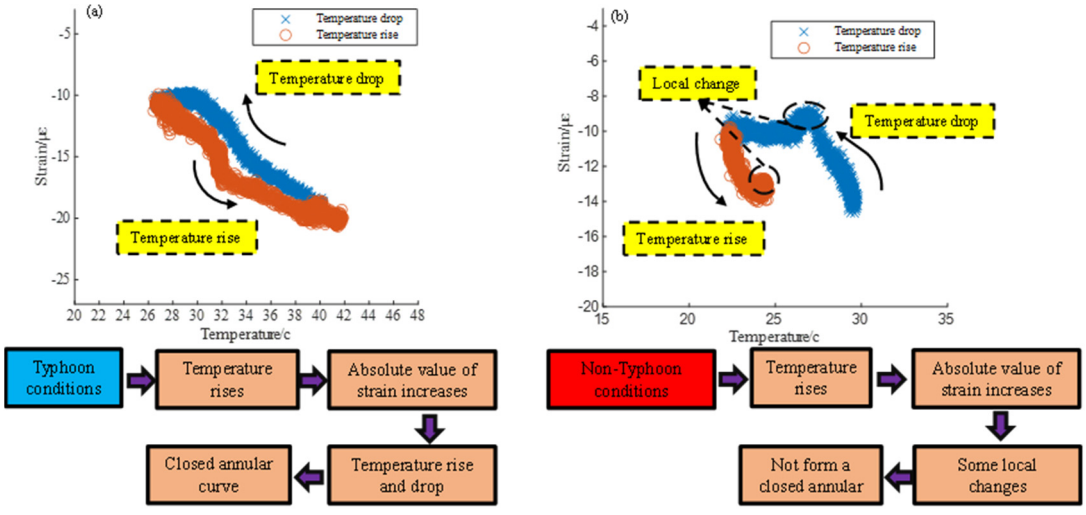

Fig. 14 Correlation between temperature and strain: (a) non-typhoon conditions (13-14 Oct.); (b) typhoon conditions (21-22 Oct.) (c) non-typhoon conditions (13-14 Oct.); (d) typhoon conditions (21-22 Oct.)

that can extend its application range from regression analysis to out-of-sample forecasting analysis. The proposed iMLHGP method overcomes this limitation of constant variance of Gaussian process (GP), and can be used for estimating non-stationary typhoon-induced response statistics with high volatility. To be specific, the iMLHGP method calculates the maximum probability in the posterior distribution of variance, to achieve the high-accuracy regression and forecasting results of heteroscedastic data with high volatility. Making accurate regression and forecasting for SHM data under extreme events with large volatility and strong uncertainty is an important problem in the SHM field. In this context, the iMLHGP method can be of high value because of its superiority in processing heteroscedastic data.

The accuracy of the iMLHGP for data regression and forecasting is demonstrated by dealing with the strain monitoring data of a real-world long-span cable-stayed bridge (i.e., TKB in Hong Kong) during typhoons. Some interesting observed conclusions specific to the problem solved above are as follows:

(1) Regarding to regression analysis: The results show that compared with GP regression, the iMLHGP algorithm has smaller MSE and higher regression accuracy regardless of the typhoon, iMLHGP regression is more consistent with the noise characteristics of the data itself. It can also be concluded that the change in the mean curve is mainly caused by temperature and is not greatly affected by the typhoon. Meanwhile, in non-typhoon conditions, the noise is mainly caused by vehicle load. In typhoon conditions, the noise is mainly caused by wind load due to the reduction of vehicle movement.

(2) With respect to forecasting analysis, the accuracy of the iMLHGP algorithm is higher in both typhoon and non-typhoon conditions than that of the GP algorithm, and the forecasting strain curve processed by the iMLHGP algorithm is smoother. Meanwhile, the degree of accuracy improvement using iMLHGP in typhoon conditions is much higher than that in non-typhoon conditions, which proves that the algorithm has better robustness performance, especially for typhoon-induced structural response forecasting.

(3) In terms of uncertainty quantification, the uncertainty in strain data is quantified by confidence interval width. In non-typhoon conditions, the uncertainty of data is mainly affected by the vehicle load. In typhoon conditions, the uncertainty of data is affected by typhoon events and the confidence interval width in the peak traffic hours is smaller than that in the non-typhoon vehicle peak period.

(4) In terms of the correlation analysis, the correlation between strain data uncertainty and wind speed, and the correlation between temperature and strain are analyzed. The following conclusions can be drawn. Firstly, the uncertainty in the strain data of the bridge was mainly caused by vehicles in non-typhoon conditions and was affected by both vehicles and wind speed in typhoon conditions. Secondly, in non-typhoon conditions, the variation trend of bridge strain is affected by temperature. In typhoon conditions, the overall change of bridge data is affected by temperature and the local fluctuation is affected by wind speed.

Through the iMLHGP regression analysis, the heteroscedasticity of monitoring data error over time can be quantitatively considered, and the monitoring data regression model can be established to obtain scientific and accurate response data. The iMLHGP forecasting model can accurately predict the future change trend of the bridge strain response based on the previous historical monitoring data, and carry out early warning to avoid the occurrence of catastrophic accidents. As for future works, the authors intend to automate the proposed methodology to realize the real-time forecasting of monitoring data embedded in the SHM system. In addition, the iMLHGP framework will be investigated to analyze the temporal and spatial correlation between sensor data to further improve the accuracy of the algorithm.

Acknowledgments

The study was supported by the National Key Research and Development Program of China under Award Number 2019YFE0118500, China Postdoctoral Science Foundation (2019M652006) and the National Natural Science Foundation of China (NSFC) under Award Numbers (51708545 and 52078478). The authors wish to express their gratitude to the staff and students in the Structural Engineering Laboratory for their extensive assistance. The data used to support the findings of this study are available from the corresponding author upon request.

References

Agou, V.D., Pavlides, A. and Hristopulos, D.T. (2022), "Spatial modeling of precipitation based on data-driven warping of Gaussian processes", *Entropy-Switz*, **24**(3), p. 321. https://doi.org/10.3390/e24030321

Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J. and Dedene, G. (2002), "Bayesian neural network learning for repeat purchase modelling in direct marketing", *Eur. J. Oper. Res.*, **138**(1), 191-211. https://doi.org/10.1016/S0377-2217(01)00129-1

Binois, M. and Gramacy, R.B. (2021), "hetGP: Heteroskedastic Gaussian Process Modeling and Sequential Design in R", *J. Stat. Softw.*, **98**(13), 1-44. https://doi.org/10.18637/jss.v098.i13

Bludszuweit, H., Dominguez-Navarro, J.A. and Llombart, A. (2008), "Statistical analysis of wind power forecast error", *IEEE Transact. Power Syst.*, **23**(3), 983-991. https://doi.org/10.1109/TPWRS.2008.922526

Braunfelds, J., Senkans, U., Skels, P., Janeliukstis, R., Porins, J., Spolitis, S. and Bobrovs, V. (2022), "Road Pavement Structural Health Monitoring by Embedded Fiber-Bragg-Grating-Based Optical Sensors", *Sensors*, **22**(12). https://doi.org/10.3390/s22124581

Brownjohn, J.M.W. (2007), "Structural health monitoring of civil infrastructure", *Philosoph. Transact. Royal Soc. A: Mathe. Phys.*

Eng. Sci., **365**(1851), 589-622. https://doi.org/10.1098/rsta.2006.1925

Crawford, L., Wood, K.C., Zhou, X. and Mukherjee, S. (2019), "Bayesian Approximate Kernel Regression With Variable Selection", *J. Am. Stat. Assoc.*, **113**(524), 1710-1721. https://doi.org/10.1080/01621459.2017.1361830

Glisic, B. (2022), "Concise historic overview of strain sensors used in the monitoring of civil structures: The first one hundred years", *Sensors*, **22**(6), p. 2397. https://doi.org/10.3390/s22062397

Goldberg, P.W. (1998), "Regression with input-dependent noise: A Gaussian process treatment", *Adv. Neural Inform. Process. Syst.*, **10**, 493-499.

Herbko, M., Lopato, P., Psuj, G. and Rajagopal, P. (2022), "Application of Selected Fractal Geometry Resonators in Microstrip Strain Sensors", *IEEE Sens. J.*, **22**(13), 12656-12663. https://doi.org/10.1109/JSEN.2022.3177932

Iba, Y. and Akaho, S. (2010), "Gaussian process regression with measurement error", *IEICE Transact. Inform. Syst.*, **93**(10), 589-622. https://doi.org/10.1587/transinf.E93.D.2680

Ju, M., Park, C. and Kim, G. (2011), "Structural Health Monitoring (SHM) for a cable stayed bridge under typhoon", *KSCE J. Civil Eng.*, **19**(4), 1058-1068. https://doi.org/10.1007/s12205-015-0039-3

Kersting, K. (2007), "Most Likely Heteroscedastic Gaussian Process Regression", *Proceedings of the 24th International Conference on Machine Learning*, pp. 393-400.

Khan, M.S., Caprani, C., Ghosh, S. and Ghosh, J. (2021), "Value of strain-based structural health monitoring as decision support for heavy load access to bridges", *Struct. Infrastruct. Eng.*, **18**(4), 521-536. https://doi.org/10.1080/15732479.2021.1890140

Ko, J.M. and Ni, Y.Q. (2005), "Technology developments in structural health monitoring of large-scale bridges", *J. Eng. Struct.*, **27**(12), 1715-1725. https://doi.org/10.1016/j.engstruct.2005.02.021

Kopsaftopoulos. F.P. and Fassois, S.D. (2011), "Scalar and vector time series methods for vibration based damage diagnosis in a scale aircraft skeleton structure", *J. Theor. Appl. Mech.*, **49**(3), 727-756. https://doi.org/10.1088/1742-6596/305/1/012056

Lee, W.F., Cheng, T.T., Huang, C.K., Yen, C.I. and Mei, H.T. (2014), "Performance of a highway bridge under extreme natural hazards: Case study on bridge performance during the 2009 Typhoon Morakot", *J. Perform. Constr. Facil.*, **28**(1), 49-60. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000418

Locke, W., Sybrandt, J., Redmond, L., Safro, I. and Atamturktur, S. (2020), "Using drive-by health monitoring to detect bridge damage considering environmental and operational effects", *J. Sound Vib.*, **468**. https://doi.org/10.1016/j.jsv.2019.115088

Majkovic, D., O'Kiely, P., Kramberger, B., Vracko, M., Turk, J., Pazek, K. and Rozman, C. (2016), "Comparison of using regression modeling and an artificial neural network for herbage dry matter yield forecasting", *J. Chemometr.*, **30**(4), 203-209. https://doi.org/10.1002/cem.2770

Mohammadi, M., Al-Fuqaha, A., Sorour, S and Guizani, M. (2007), "Deep learning for IoT big data and streaming analytics: A survey", *IEEE Commun. Surv. Tutorials*, **20**(4), 2923-2960. https://doi.org/10.1109/COMST.2018.2844341

Mousavi, M. and Gandomi, A.H. (2021), "Structural health monitoring under environmental and operational variations using MCD prediction error", *J. Sound Vib.*, **512**. https://doi.org/10.1016/j.jsv.2021.116370

Munoz-Gonzalez, L., Lazaro-Gredilla, M. and Figueiras-Vidal, A.R. (2016), "Laplace approximation for divisive gaussian processes for nonstationary regression", *IEEE T. Pattern Anal.*, **38**(3), 618-624. https://doi.org/10.1109/TPAMI.2015.2452914

Ni, Y.Q., Xia, Y., Liao, W.Y. and Ko, J.M. (2009), "Technology innovation in developing the structural health monitoring system for Guangzhou new TV tower", *Struct. Control Health Monitor.*, **16**(1), 73-98. https://doi.org/10.1002/stc.303

Papadimitriou, C. (2004), "Optimal sensor placement methodology for parametric identification of structural systems", *J. Sound Vib.*, **278**(4-5), 923-947. https://doi.org/10.1016/j.jsv.2003.10.063

Parida, S.S., Nikellis, A., Sett, K. and Singla, P. (2020), "Model-data fusion for seismic performance evaluation of an instrumented highway bridge", *Earthq. Eng. Struct. Dyn.*, **49**(14), 1559-1578. https://doi.org/10.1002/eqe.3317

Partal, T. (2017), "Wavelet regression and wavelet neural network models for forecasting monthly streamflow", *J. Water Clim. Change*, **8**(1), 48-61. https://doi.org/10.2166/wcc.2016.091

Porter, K.A., Beck, J.L. and Shaikhutdinov, R.V. (2002), "Sensitivity of building loss estimates to major uncertain variables", *Earthq. Spectra*, **18**(4), 719-743. https://doi.org/10.1193/1.1516201

Rasmussen, C.E. and Nickisch, H. (2010), "Gaussian Processes for Machine Learning (GPML) Toolbox", *J. Mach. Learn Res.*, **11**, 3011-3015. https://doi.org/10.1115/1.4002474

Sierra-Garcia, J.E. and Santos, M. (2021), "Improving wind turbine pitch control by effective wind neuro-estimators", *IEEE Access.*, **9**, 10513-10425. https://doi.org/10.1109/ACCESS.2021.3051063

Skolidis, G. and Sanguinetti, G. (2011), "Bayesian Multitask Classification with Gaussian Process Priors", *IEEE T. Neural Networ.*, **22**(12), 2011-2021. https://doi.org/10.1109/TNN.2011.2168568

Subasi, A. (2013), "Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders", *Comput. Biol. Med.*, **43**(5), 576-586. https://doi.org/10.1016/j.compbiomed.2013.01.020

Talaei, S. and Ma, H.W. (2007), "Vibration-based Structural Damage Detection Using Twin Gaussian Process (TGP)", *Structures*, **16**, 10-19. https://doi.org/10.1016/j.istruc.2018.08.006

Theiler, M., Frangi, A. and Steiger, R. (2014), "Strain-based calculation model for centrically and eccentrically loaded timber columns", *Eng. Struct.*, **56**, 1103-1116. https://doi.org/10.1016/j.engstruct.2013.06.032

Urban, S., Ludersdorfer, M. and van der Smagt, P. (2015), "Sensor Calibration and Hysteresis Compensation with Heteroscedastic Gaussian Processes", *IEEE Sens. J.*, **15**(11), 6498-6506. https://doi.org/10.1109/JSEN.2015.2455814

Von Krannichfeldt, L., Wang, Y and Hug, G. (2007), "Online Ensemble Learning for Load Forecasting", *IEEE T. Power Syst.*, 36(1), 545-648. https://doi.org/10.1109/TPWRS.2020.3036230

Wan, H.P. and Ni, Y.Q. (2019a), "Binary Segmentation for Structural Condition Classification Using Structural Health Monitoring Data", *J. Aerospace Eng.*, **32**(1). https://doi.org/10.1061/(ASCE)AS.1943-5525.0000956

Wan, H.P. and Ni, Y.Q. (2019b), "Bayesian modeling approach for forecast of structural stress response using structural health monitoring data", *J. Struct. Eng.*, **144**(9). https://doi.org/10.1061/(ASCE)ST.1943-541X.0002085

Wan, H.P. and Ni, Y.Q. (2019c), "Bayesian multi-task learning methodology for reconstruction of structural health monitoring data", *Struct. Health Monit.*, **18**(4), 1282-1309. https://doi.org/10.1177/1475921718794953

Wan, H.P., Dong, G.S., Luo, Y.Z. and Ni, Y.Q. (2022), "An improved complex multi-task Bayesian compressive sensing approach for compression and reconstruction of SHM data", *Mech. Syst. Signal Pr.*, **167**. https://doi.org/10.1016/J.YMSSP.2021.108531

Wang, Q.A., Wu, Z. and Liu, S. (2018a), "Multivariate probabilistic seismic demand model for the bridge

multidimensional fragility analysis", *KSCE J. Civil Eng*., **22**(9), 3443-3451. https://doi.org/10.1007/s12205-018-1792-4

Wang, Q.A., Wu, Z. and Liu, S. (2018b), "Multivariate probabilistic seismic demand model for the bridge multidimensional fragility analysis", *KSCE J. Civil Eng*., **22**(9), 3443-3451. https://doi.org/10.1007/s12205-018-1792-4

Wang, Q.A., Wang, C.B., Ma, Z.G., Chen, W., Ni, Y.Q., Wang, C.F., Yan, B.G. and Guan, P.X. (2022a), "Bayesian dynamic linear model framework for SHM data forecasting and missing data imputation during typhoon events", *Struct. Health Monitor*., **21**(6), 2933-2950. https://doi.org/10.1177/14759217221079529

Wang, Q.A., Zhang, C., Ma, Z.G., Jiao, G.Y., Jiang, X.W., Ni, Y.Q., Wang, Y.C., Du, Y.T., Qu, G.B. and Huang, J. (2022b), "Towards long-transmission-distance and semi-active wireless strain sensing enabled by dual-interrogation-mode RFID technology", *Struct. Control Health*., **29**(11), e3069. https://doi.org/10.1002/stc.3069

Wang, Q.A., Zhang, C., Ma, Z.G. and Ni, Y.Q. (2022c), "Modelling and forecasting of SHM strain measurement for a large-scale suspension bridge during typhoon events using variational heteroscedasic Gaussian process", *Eng. Struct*., **251**, 113554. https://doi.org/10.1016/j.engstruct.2021.113554

Wang, Q.A., Dai, Y., Ma, Z.G., Wang, J.F., Lin, J.F., Ni, Y.Q., Ren, W.X., Jiang, J., Yang, X. and Yan, J.R. (2023), "Towards high-precision data modeling of SHM measurements using an improved sparse Bayesian learning scheme with strong generalization ability", *Struct. Health Monitor*. https://doi.org/10.1177/14759217231170316

Wright, J.H. (2008), "Bayesian Model Averaging and exchange rate forecasts", *J. Econometr*., **146**(2), 329-341. https://doi.org/10.1016/j.jeconom.2008.08.012

Yoon, S. and Kim, S. (2010), "k-Top Scoring Pair Algorithm for feature selection in SVM with applications to microarray data classification", *Soft Comput*., **14**(2), 151-159. https://doi.org/10.1007/s00500-009-0437-x

Zhang, Q.H. and Ni, Y.Q. (2020), "Improved most likely heteroscedastic Gaussian process regression via Bayesian residual moment estimator", *IEEE Transact. Signal Process*., **68**, 3450-3460. https://doi.org/10.1109/TSP.2020.2997940

FC