Physics-informed machine learning for the structural health monitoring and early warning of a long highway viaduct with displacement transducers

Mechanical Systems and Signal Processing 242 (2026) 113659

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Full Length Article

Physics-informed machine learning for the structural health monitoring and early warning of a long highway viaduct with displacement transducers

Enrico Cianci, Marco Civera *, Valerio De Biagi, Bernardino Chiaia

Department of Structural, Geotechnical and Building Engineering (DISEG), Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy

ARTICLE INFO

Communicated by Wei-Xin Ren

Keywords:
Bridge monitoring
Static monitoring
Displacement transducers
Physics-informed machine learning
Grey-box modelling
Structural health monitoring

ABSTRACT

In Bridge Structural Health Monitoring (SHM), damage detection is often hindered by environmental and operational variability. Confounding influences such as traffic, wind, and especially temperature can significantly affect measurements, making it difficult to distinguish true damage-related anomalies. This challenge is critical in static monitoring of long steel viaducts, where thermal effects dominate displacements, making small damage-induced perturbations difficult to detect. To address this, the study introduces a Physics-Informed Machine Learning (PIML) model that establishes a reliable baseline for the 'normal conditions' of the infrastructure. This baseline isolates anomalies attributable to structural damage while accounting for temperature effects. The proposed grey-box approach combines data-driven modelling with physical knowledge of thermal behaviour, enhancing both accuracy and interpretability. A real-world application is presented on a long-span highway viaduct, where longitudinal displacements are monitored using temperature and displacement sensors. By using only temperature and time as inputs, the model captures nonlinear daily and seasonal thermal cycles without additional instrumentation. To assess reliability, an Early Warning System (EWS) is developed based on displacement anomaly thresholds and Kernel Density Estimation (KDE). The PIML model is evaluated against black-box (purely data-driven) and white-box (purely physics-based) alternatives. Damage scenarios are simulated by introducing anomalies into experimental data to test each model's capability to detect abnormal behaviour while filtering out environmental effects. Results show that the grey-box PIML consistently outperforms black- and white-box models in accuracy, robustness, and anomaly discrimination. These findings demonstrate the potential of PIML to advance SHM practices and enable reliable automated EWSs for bridge monitoring.

1. Introduction

Structural Health Monitoring (SHM) has become a fundamental tool for ensuring the safety and longevity of civil infrastructures through continuous condition assessment. Bridges, in particular, represent a critical application domain where SHM systems enable condition-based maintenance, early damage detection, and risk management. As the number of sensor-equipped bridges increases, the volume of data available for analysis has grown exponentially, fostering the development of data-driven models capable of extracting

* Corresponding author.
E-mail addresses: enrico.cianci@polito.it (E. Cianci), marco.civera@polito.it (M. Civera), valerio.debiagi@polito.it (V. De Biagi), bernardino.chiaia@polito.it (B. Chiaia).

https://doi.org/10.1016/j.ymssp.2025.113659
Received 6 August 2025; Received in revised form 11 October 2025; Accepted 15 November 2025
Available online 21 November 2025

meaningful patterns from large and complex datasets while accounting for environmental and operational influences [1].

However, the practical deployment of SHM systems still faces significant challenges. One of the most pressing issues is the Environmental and Operational Variability (EOV) [2], which refers to the inherent fluctuations in environmental conditions (e.g., temperature, humidity) and operational loads (e.g., traffic) that can affect the structural response, often making it difficult to differentiate between these effects and damage [3,4]. A variety of approaches, including Machine Learning (ML) and statistical methods, have been proposed to mitigate these effects. Still, many rely on idealised or linear assumptions that fail to capture the nonlinear interactions between variables. This makes it difficult to compensate for EOV without oversimplifying the data or masking subtle damage indicators.

Another significant limitation in current SHM methodologies is the lack of interpretability in many black-box machine learning models. Techniques like Gaussian Processes (GPs), Support Vector Machines (SVMs), and Neural Networks (NNs) have shown promising results in terms of prediction accuracy [5], however, the inherent opacity of data-driven models makes it difficult for engineers to derive actionable insights or confirm the validity of the results. It is essential to develop models that support transparent decision-making, as trust and explainability are crucial for safety-critical applications. This leads to a research gap where the trade-off between model accuracy and interpretability remains unresolved, especially in real-world, complex SHM systems. Furthermore, although physics-based models are well-established and widely used for structural analysis, their limited flexibility in real-time monitoring environments limits their usefulness for SHM. These models are often constrained by assumptions about material properties, boundary conditions, and loading scenarios, making them difficult to apply in the field without significant adjustments.

As a result, recent research has increasingly explored hybrid strategies that combine the strengths of physics-based and data-driven approaches, leading to more robust and interpretable models that leverage both empirical observations and domain knowledge [6]. Several recent studies have proposed hybrid frameworks using GPs or NNs to capture complex structural responses. For instance, some works have applied physics-informed GP for vibration-based [7,8] or fatigue-related monitoring [9], or integrated neural-based models with physical constraints to improve the reliability of damage detection [10]. These approaches highlight the growing interest in embedding physics priors into machine learning models to enhance generalisation and interpretability. In contrast, the present study focuses on the static monitoring of bridge bearings, where dominant thermal effects are explicitly modelled, and residual analysis is leveraged to separate environmental variations from potential anomalies. This residual-based grey-box strategy enhances interpretability and transferability, complementing existing hybrid SHM frameworks while specifically addressing the challenges of long-term static monitoring.

Many existing SHM models are tailored to specific structures or environments, limiting their generalisability when applied to different types of bridges or regions with varying environmental conditions. To address this limitation, transfer learning has been proposed as a potential solution [11,12], in particular when training data from one bridge is limited or unavailable for a new structure [13]. However, developing robust domain-adaptation strategies that effectively handle differences in scale, material properties, and environmental conditions across bridges remains a significant challenge [14].

In parallel, some research efforts have focused on monitoring frameworks that explicitly treat temperature as the dominant factor influencing structural response. This has led to the emergence of Temperature-Based SHM (TBSHM), which leverages temperature-induced effects, typically characterised by high signal-to-noise ratios and reduced sensitivity to unpredictable operational conditions, to detect long-term trends and subtle structural anomalies [15].

The validation of SHM models under actual operational conditions remains a major obstacle. Most studies are validated using controlled, laboratory-like data, which may not accurately capture the noise and variability encountered in practice. Applications of SHM systems require models capable of handling noisy, sparse, and irregular data, a challenge that has yet to be fully addressed in current literature. This lack of validation under realistic conditions limits the practical deployment of SHM systems, making it difficult to assess their true robustness and reliability in operational environments.

In the context of bridge monitoring, Early Warning Systems (EWS) aim to promptly detect abnormal structural responses or damage progression, enabling timely interventions that can prevent catastrophic failures. Their implementation is crucial for infrastructures subject to ageing, environmental hazards, or heavy usage. Effective EWS can significantly reduce the risk of sudden collapses by transforming continuous monitoring data into real-time alerts [16]. Some recent studies have explored probabilistic models to support operational warnings, for instance, Bayesian Neural Networks applied to estimate cumulative bridge girder displacements [17]. Other recent frameworks have explored EWS applications in static monitoring setups, including systems based on the observation of bearing reaction forces [18], which are directly relevant to the approach presented in this study. Other strategies leverage correlations between temperature and longitudinal displacements to define alert conditions, demonstrating the potential of EWS to detect deviations from normal thermo-mechanical behaviour [19]. These systems rely on robust, damage-sensitive features and alert thresholds that account for environmental variability, a major challenge that remains in their design. Nevertheless, despite the increasing deployment of sensors, many SHM frameworks still lack well-calibrated early warning capabilities, particularly under non-ideal conditions such as nonlinear and hysteretic structural behaviours, as well as the confounding influence of EOVs. In this context, interpretable and adaptive decision-making frameworks play a key role in supporting real-time alerts with minimal false positives and missed detections.

Beyond classical EWS, advanced machine learning frameworks have been investigated to enhance operational and economic decision-making, as well as to improve understanding of structural behaviour. For example, deep reinforcement learning has been applied to optimise infrastructure lifecycle management, integrating inspection planning with machine learning for more efficient operational and economic management [20]. Similarly, interpretable ensemble models have been used to predict critical displacements, as in the case of cable-supported structures, providing insights into the factors influencing bridge behaviour [21]. These approaches illustrate how modern predictive and interpretable methods can complement EWS, improving both real-time alerts and long-term infrastructure management.

These research gaps underscore the need to further develop SHM methodologies, particularly in model interpretability, data normalisation under environmental variability, and validation in noisy environments. This research aims to address these gaps by developing a hybrid, interpretable, and adaptable model for SHM systems, focusing specifically on static bridge monitoring applications and on the principles of TBSHM.

The remainder of this article is organised as follows. Section 2 details the methodology of the hybrid (i.e. grey-box) model applied to define the structure's normal condition. Then, Section 3 presents the field case study. Section 4 discusses the results of applying the PIML model to this dataset. Departing from this baseline model, Section 5 examines damage detection by introducing artificial anomalies and testing a Kernel Density Estimation (KDE)-based EWS. Section 7 concludes this work, summarising the key findings of both PIML and EWS.

2. Methodology: Grey-Box modelling

2.1. From Black-Boxes to hybrid models

Machine Learning has revolutionised the modelling of complex nonlinear systems by leveraging large volumes of observational data to generate accurate predictive models. With the growing availability of SHM data, machine learning methods, such as GPs, SVMs, and NNs, have shown strong potential for tasks such as anomaly detection and system identification. Their ability to extract complex patterns from empirical observations makes them well-suited for modelling structural behaviour [5]. Such purely data-driven models (i.e., black-box models) are built solely on empirical observations and can deliver accurate predictions even without a physical description of the system. This advantage is particularly pronounced in complex and dynamic systems, such as wind turbines, bridges, or aircraft, where operational uncertainties and structural nonlinearities hinder accurate physical modelling [22].

However, in many engineering domains, especially those related to SHM and infrastructure management, data are often scarce, noisy, and incomplete. This inherent limitation restricts the effectiveness and reliability of purely data-driven ML models, which critically depend on the quality and representativeness of training datasets [23]. Moreover, despite their predictive power, black-box models often suffer from poor generalisability outside the training domain, limited interpretability, and a lack of transparency in safety–critical contexts [24]. Purely data-driven models, commonly referred to as black-box approaches, exhibit several fundamental weaknesses: lack of generalisation to unseen or extrapolative conditions, limited interpretability, and sensitivity to noise and environmental variations. These factors undermine their suitability for high-stakes decision-making where safety, reliability, and regulatory compliance are essential [25,26]. Moreover, black-box models are prone to overfitting, where learned relationships capture noise or artefacts in the training data rather than the underlying physical phenomena, leading to degraded predictive performance on new data [27]. This aspect is worsened in SHM applications, where structural behaviour is influenced by time-varying environmental and operational conditions, such as temperature fluctuations and load variations, as well as progressive degradation that may be poorly represented in historical measurements [6].

In recent years, the demand for model interpretability has grown significantly, especially within engineering domains where transparent and explainable predictions are essential to foster trust and support informed decision-making processes [28]. This has also motivated research into physics-guided data-driven methods, such as autoencoder-based approaches, that operate effectively even under sparse and noisy sensing conditions [29]. Hence, the current push is to move beyond black-box methods with PIML [30], not only limited to SHM but also in related fields such as nonlinear dynamics [31].

Traditional ML models generally do not provide insights into the physical mechanisms underlying their predictions, which limits their acceptance in safety–critical scenarios. Conversely, white-box models, such as those based on Finite Element Methods (FEMs) or differential equations, are founded on established physical laws and therefore offer high interpretability. Yet they, too, face limitations in practice: incomplete or uncertain boundary conditions, simplifying modelling assumptions, and difficulty in capturing real-time

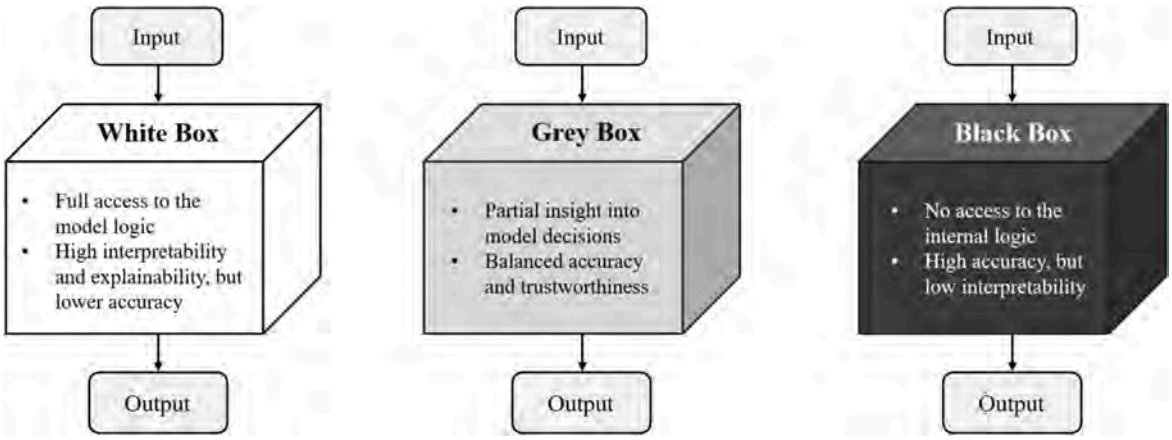

Fig. 1. Comparison between white-, grey-, and black-box models.

variability often reduce their predictive accuracy.

2.2. Physics-Informed Machine learning.

In response to the shortcomings of purely black-box and white-box approaches, PIML has emerged as a promising paradigm that combines the strengths of both. In particular, PIML focuses on the development of grey-box models, hybrid schemes that integrate physical knowledge within machine learning frameworks (Fig. 1). This integration leads to models that retain physical consistency and interpretability while gaining adaptability and predictive power through data adaptation. The advantages of PIML are numerous and significantly enhance the robustness and applicability of predictive models in engineering contexts. One key benefit lies in the improved generalisation capabilities. By incorporating physical constraints and enforcing known invariances, these models are less prone to overfitting and tend to provide reliable predictions even under diverse operational conditions and previously unseen scenarios [32]. Furthermore, PIML enhances the interpretability of machine learning outputs. The integration of physical principles ensures that the model's behaviour aligns with established engineering understanding, thus making the results more transparent and easier to validate by experts [33]. Another significant advantage is the reduction in data requirements. Since PIML leverages prior physical knowledge, it can attain accurate performance even when only limited or noisy data are available, which is a common situation in the context of infrastructure monitoring [25]. Finally, the inclusion of physical priors increases the model's resilience to environmental variability. This feature enables a more effective separation between structural responses and environmental influences, such as temperature or humidity, ultimately enhancing the sensitivity and reliability of anomaly detection in SHM applications [6].

Several empirical studies validate the efficacy of PIML frameworks in bridge monitoring contexts [34]. A common methodology involves residual modelling, where ML models (e.g., GPs and NNs) are trained on the residuals between observed sensor data and predictions from physics-based simulations, thereby capturing discrepancies attributable to damage or degradation. Physics-Informed Neural Networks (PINNs) have been successfully applied to embed governing partial differential equations of structural dynamics directly into deep learning architectures, ensuring physically consistent and interpretable predictions [35]. Additionally, the integration of ML corrections into digital twin frameworks has enabled real-time monitoring and predictive maintenance strategies that adapt to evolving system states and environmental conditions [36].

Among the various PIML strategies, this research investigates a residual learning framework, in which ML is used to model the discrepancy between a physics-based model and observed data. This approach, known as residual modelling, allows the ML algorithm to capture unmodeled phenomena, compensate for simplifications in the physical model, and correct for bias induced by unknown boundary or loading conditions [37]. Within this framework, Gaussian Process Regression (GPR) provides a robust and versatile tool for SHM, owing to its probabilistic and non-parametric nature, as well as its ability to quantify prediction uncertainty. GPR has been widely adopted in SHM in general [38] as well as in previous PIML studies (in particular [7]) mainly due to its strong performance in modelling residuals. GPs are well-established and extensively studied in the literature; therefore, their fundamental equations are not explicitly derived here. For a comprehensive reference, see Rasmussen & Williams [39]. Its Bayesian formulation offers a key advantage in safety-critical systems, where understanding the confidence of a forecast is as crucial as the forecast itself. Moreover, GPs perform effectively even with sparse or noisy training data, making them ideal for real-world SHM applications where data quality and

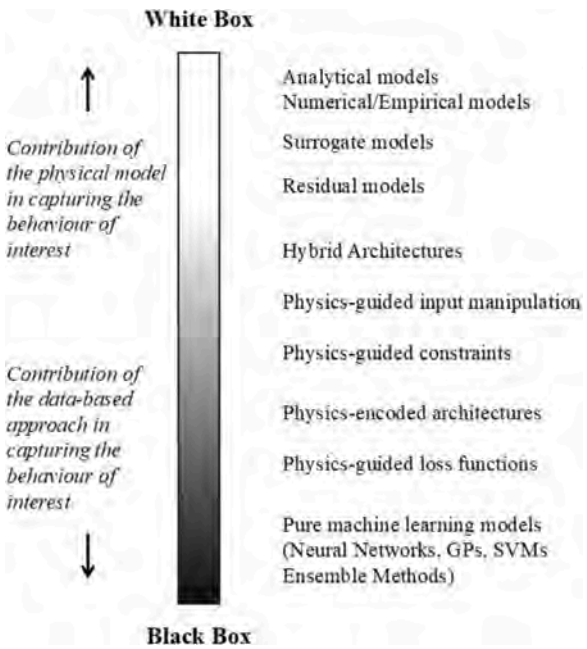

Fig. 2. Overview of the spectrum of Physics-Informed Machine Learning models (inspired by and adapted from Cross et al. [37]).

quantity may be limited.

'Physics-informed' and 'grey-box' are very vague terms that can be applied in a variety of specific implementation methods (Fig. 2), such as the one presented in this study, which are 'whiter' than 'black', as it is deemed better for maximum interpretability and explainability [37].

This research adopts a residual GPR framework to develop a grey-box SHM model. Rather than replacing the physics-based model, GPR is used to model its residuals, thereby enhancing the overall predictive accuracy and reliability of the model. This hybrid strategy leverages physical understanding while exploiting the flexibility of machine learning to adapt to unmodelled real-world complexity.

In summary, the growing complexity of engineering systems and the critical nature of their operation necessitate a paradigm shift beyond pure data-driven ML towards hybrid physics-informed approaches. PIML represents a crucial step forward, combining the strengths of physics-based understanding and data-driven adaptability to deliver reliable, interpretable, and efficient models that support advanced SHM and informed decision-making in engineering practice.

2.2.1. Integration of prior mean functions and residual learning

In system identification and modelling, engineers often possess detailed insights into the physical mechanisms governing the system's behaviour. However, conventional data-driven approaches, such as GPR, typically neglect this knowledge by employing uninformative priors, thus failing to exploit potentially valuable structural information. The grey-box modelling paradigm addresses this limitation by combining physics-based (white-box) formulations with data-driven (black-box) models, thereby enhancing both interpretability and predictive performance [35].

GPR requires specifying both a prior mean function and a covariance kernel, which jointly define the Gaussian process prior. Observed data are then used to update this prior, yielding a posterior distribution over functions. In standard practice, the prior mean is assumed to be zero, and the kernel is chosen from a class of flexible functions, such as the squared exponential or Matérn kernel, capable of representing a wide range of behaviours [39]. However, this zero-mean assumption neglects the possibility of embedding known system behaviour directly into the model structure.

A grey-box GPR framework overcomes this limitation by adopting a physically derived model, typically a simplified, mechanistic representation, as the prior mean function. The GP is then used to learn the residuals, i.e., the differences between the observed data and the predictions made by the white-box model. This decomposition enables the model to focus its learning capacity on patterns not captured by the physical formulation, thus efficiently combining theoretical knowledge with empirical flexibility [40]. In a grey-box, the white-box component serves as a deterministic baseline that encodes the expected trend. Importantly, since the GP is applied to the residuals, the covariance structure remains mathematically identical to that of a standard zero-mean GP trained on the discrepancies alone.

This modelling strategy offers several advantages, making it particularly suitable for applications in structural engineering. By incorporating prior knowledge through a deterministic mean function, the model embeds interpretable and physically explainable behaviour, an important feature in domains where relationships such as those between displacement and temperature are well established. This integration not only enhances the model's interpretability but also improves predictive accuracy. By isolating and learning only the residual and unexplained variability, the GP component can capture complex, nonlinear patterns that go beyond the descriptive power of traditional white-box models. Furthermore, the inclusion of prior physical knowledge reduces the learning burden of the data-driven component, enabling better generalisation even when training data are limited.

2.2.2. Analytical formulation

The proposed grey-box hybrid model adopts a residual-based approach, combining both physically informed and data-driven components, similarly to the methodology introduced by Zhang et al. [41], which serves as a foundation for this work. Specifically, it integrates a white-box term derived from analytical knowledge and a black-box term modelled through GP.

For the particular application of interest, the model follows the formulation proposed in [37]:

$$y(\Delta T, t) = f(\Delta T) + \delta(\Delta T, t) + \sigma \tag{1}$$

- $y(\Delta T, t)$ represents the total measured displacement, as a function of the temperature variation ΔT and time t.
- $f(\Delta T)$ denotes the white-box component derived from thermoelastic principles, calculated as:

$$f(\Delta T) = \delta_T = \alpha \cdot \Delta T \cdot L \tag{2}$$

i.e. the classic equation of linear thermal dilation. In this expression δ_T is the thermal displacement, which is equal to α, the coefficient of thermal expansion (which can be experimentally determined, or known for common materials such as structural steel), multiplied by ΔT, multiplied by L, i.e., the effective unrestrained length. This linear dependency correlates displacements with temperature variations, providing a straightforward and physically interpretable baseline model for the system. Importantly, this common formulation is not only linear but also does not account for time-dilated effects.

- $\delta(\Delta T, t) + \sigma$ represents the residual (or discrepancy) between the observed behaviour and the analytical model. Specifically, σ is a random (white Gaussian noise) factor that identifies the process noise, thus independent of time and temperature, while $\delta(\Delta T, t)$ is here assumed as nonlinear and learned based on GP regression. In this context, these two terms form the black-box component of the hybrid model.

Thus, as said, the primary input variable corresponds to the temperature variations recorded by the monitoring system. To enhance the model's ability to learn from temporal trends, additional time-related inputs were included during the training process. More in detail, a total of two vectors were employed: one encoding the time of day (expressed in hours with minutes converted to a decimal fraction), and another indicating the sequential day of the year. After adopting this approach, the model effectively accounted for both diurnal and seasonal patterns, thereby improving its ability to differentiate between short-term fluctuations and long-term displacement trends.

Summarising, the whole methodology can be defined in four consecutive steps:

1. Analytical evaluation of the expected displacements according to the white-box model.
2. Evaluation of the white-box model residuals.
3. Training on residuals of the black-box component, including time as an input to account for time-dependent phenomena
4. Prediction of the corrected displacements.

3. Case study

In this Section, the anonymised case study, the specific data subset, and the modelling approach will be briefly described.

The dataset analysed originates from a continuous monitoring system installed on a major Italian highway viaduct built in the mid-1970s and located within an urban area (Fig. 3). The structure consists of two parallel decks, each serving a single traffic direction, extending over 1,360 m. Each deck comprises 18 spans, each ranging from 52 to 92 m in length. The steel box girders for each carriageway vary in height between 2.75 and 3.55 m and have a base width of 6.00 m. These girders are topped by a 30 cm-thick, transversely prestressed reinforced concrete slab. For steel box girder bridges like the one examined, temperature fluctuations predominantly influence daily and seasonal displacement patterns, making accurate modelling of the relationship between environmental factors and structural response as challenging as crucial for effective SHM.

A critical structural component considered here is the bearing system located at the bridge abutments. Each abutment is equipped with four unidirectional sliding pot bearings (Fig. 4) designed to accommodate longitudinal displacements and reduce stress transfer between the deck and substructure. The bearings' sliding mechanism consists of two flat surfaces (a highly polished stainless-steel plate and a Polytetrafluoroethylene (PTFE) layer) that allow smooth longitudinal movement up to $\pm$ 200 mm, effectively accommodating thermal expansion and contraction.

The monitoring system gathers displacement data from eight inductive transducers (PZ34 rectilinear displacement transducers with cylindrical cases), four per abutment, each associated with a specific bearing. These Linear Variable Differential Transformer (LVDT)-based sensors, designed for industrial environments, feature a stainless-steel body for enhanced durability and articulated spherical joints for flexible mounting. They have a mechanical range of $\pm$ 150 mm, suitable for all expected displacement configurations. The 4–20 mA output signal minimises electrical interference and enables long-distance transmission. Thermal sensors installed along the deck provide ambient temperature measurements; for this study, only the temperature probes near the abutments, which represent local conditions, were selected. Data were initially acquired at a high frequency (1 sample per second) for 10 min, resulting in a very large dataset. To enable manageable analysis focused on the quasi-static structural response, the data were downsampled to one reading per minute. This yields approximately 1,440 records per sensor per day, about 44,500 records per month, and over 500,000 data points spanning the period from September 2023 (when the monitoring system was initiated) to September 2024. These were the only recorded measurements available at the time of this research.

To avoid biases related to temporal ordering and reduce computational demands, a random subset representing 2 % of the full dataset was extracted after thorough shuffling. This procedure is essential to prevent autocorrelation effects and temporal dependencies that could bias the training and evaluation of the models [42]. Proper shuffling also ensures a representative sampling of

Fig. 3. Satellite view of the viaduct in the densely populated area surrounding it.

Fig. 4. Unidirectional longitudinal bearing at abutment 1.

the data distribution, which improves the generalisation capability of machine learning models [43]. The reason behind the choice of the 2 % sampling rate will be further elaborated in subsection 4.6. This subset was then split into 70 % training data (training ratio) and 30 % testing data. For model training, this corresponds to approximately 7200 samples. The training process required roughly 5 min to complete, providing an efficient balance between model accuracy and computational cost. To comprehensively assess the model's predictive accuracy and robustness, a 5-fold cross-validation was performed. The shuffled dataset was divided into five equally sized folds, with each fold used once as a validation set while the others served for training. This K-fold approach ensures that each data subset contributes to both model training and evaluation, thereby minimising overfitting and enhancing reliability across different data segments. The results obtained from this procedure confirmed good predictive performance and model stability across folds.

Given the significant influence of seasonal and diurnal environmental variations on bridge response, the training was performed on a full year of data. This ensures the model is exposed to a comprehensive range of operational and environmental conditions, which is critical for normalisation and damage detection. Using an entire year's dataset helps capture the system's natural variability and prevents misinterpreting uncommon environmental states as structural anomalies, as shorter training durations have been shown to lead to such misclassifications [44].

In Fig. 5, the layout of the sensor arrangement for the monitoring system is shown. The displacement transducers (indicated in green) are positioned at the longitudinal disconnection axes between the various sections of the deck. Sensors are identified by a combination of a number and a letter: the number (1 for west abutment, 2 for east abutment) indicates the abutment, while the letter (A, B, C, D) specifies the individual bearing sensor. Although data from all sensors were analysed, results reported here focus primarily on sensors 1A and 2A (marked in red), which are representative of their respective abutments due to consistent sensor behaviour. It is essential to note that the sign convention differs between the two abutments because the sensors were installed in opposing orientations. Specifically, for sensors installed on abutment 2 (2A–2D), a positive displacement value corresponds to an elongation of the bridge deck, whereas for sensors on abutment 1 (1A–1D), a positive displacement value denotes a contraction of the deck.

For benchmarking purposes, linear regression models were also fitted, and their performance was assessed using the Pearson correlation coefficient (ρ).

Model evaluation employed standard regression metrics, including:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Normalised Mean Squared Error (NMSE)

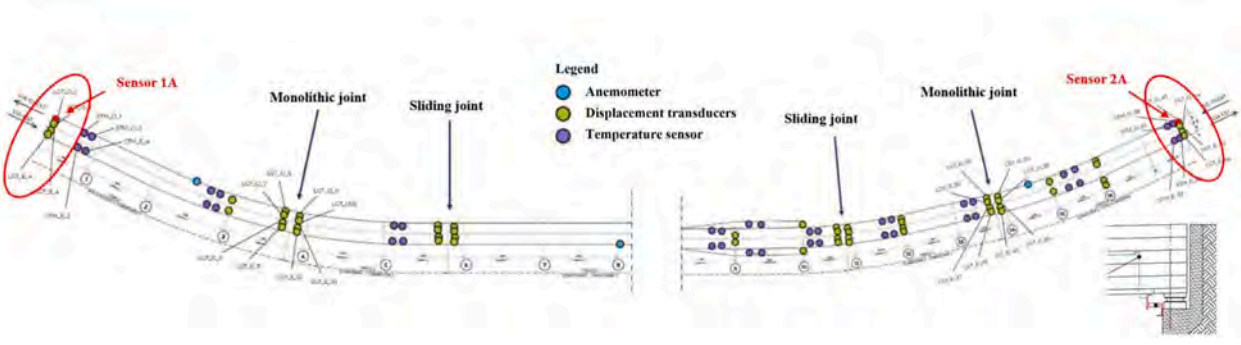

Fig. 5. Plan view of the sensor layout. Sensors 1A and 2A, located on the two abutments, are highlighted in red. The bottom right corner shows a detailed view of the sensor installed on the bearing at the abutment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

- Coefficient of Determination (R^2)

Results are presented as time series plots comparing observed and predicted displacements, along with error time series showing prediction residuals. Temperature-displacement graphs illustrate each model's ability to reproduce the cyclic thermal influence on the structure's response, offering clear insights into the GP model's effectiveness in capturing temperature-dependent displacement patterns.

In addition to GPR, SVM regression was also tested as an alternative data-driven approach, as it has been successfully applied in similar contexts in previous studies [45]. SVM is a kernel-based supervised learning algorithm that constructs a regression function by minimising a loss function while ensuring model flatness and robustness to outliers [46]. The goal is to find a function that deviates from the actual targets by no more than a predefined margin (ε) while keeping the model as simple as possible through regularisation [47]. Various kernel functions were evaluated, including the Radial Basis Function (RBF), but in all cases, the SVM model exhibited higher prediction errors and lower correlation with observed displacements than the GPR model. Given its inferior performance, the subsequent hybrid modelling and Early Warning analyses were conducted exclusively using GPR.

4. Grey-box modelling: Results and discussion

4.1. Observed daily trends and hysteretic cyclic behaviour

To explore the influence of intraday thermal variations on bridge displacement, a representative case study was selected, focusing on the behaviour recorded on August 8, 2024. On this day, as on many others, diurnal temperature fluctuations prompted observable cycles of expansion and contraction of the bridge deck. During the morning and afternoon, rising temperatures induce thermal expansion, producing longitudinal displacements in the deck bearings. Conversely, during the evening and night, cooling causes contraction and a reversal of displacement direction. These daily cycles tend to be more pronounced on days characterised by large temperature gradients between daytime highs and nighttime lows.

In case of a purely linear, thermoelastic behaviour, the longitudinal displacement δ_T of a bridge deck subjected to a temperature change ΔT is governed by the well-known relation, already presented in Eq. (2). In this case, α is the coefficient of thermal expansion for steel, and L is the unrestrained length of the continuous deck between the abutment and the first sliding joint. This formulation assumes an idealised system governed exclusively by linear elasticity and instantaneous thermal response.

However, empirical evidence suggests a deviation from this idealised linear behaviour. As illustrated in Fig. 6, the temperature–displacement relationship exhibits clear inelastic and hysteretic behaviour, deviating from the expected linear trend, with a distinct heating path clearly differentiated from the cooling one. This hysteresis effect, commonly observed in real structures, lacks a universally accepted mechanistic explanation, and remains an open topic in bridge engineering research [48]. The observed behaviour implies that, although temperature-induced displacements are a well-established phenomenon, the cyclic nonlinear dynamics associated with them are more complex than previously assumed.

One possible hypothesis attributes the hysteresis to frictional resistance within the sliding bearings. Although such bearings are engineered to accommodate thermal movements, in practice, they exhibit a threshold resistance due to static friction. Displacement occurs only when the thermal-induced force exceeds this threshold, resulting in delayed movement. This behaviour is commonly referred to as stick–slip behaviour [49]. During the heating phase, this leads to a lag in expansion: the bearing remains stationary until

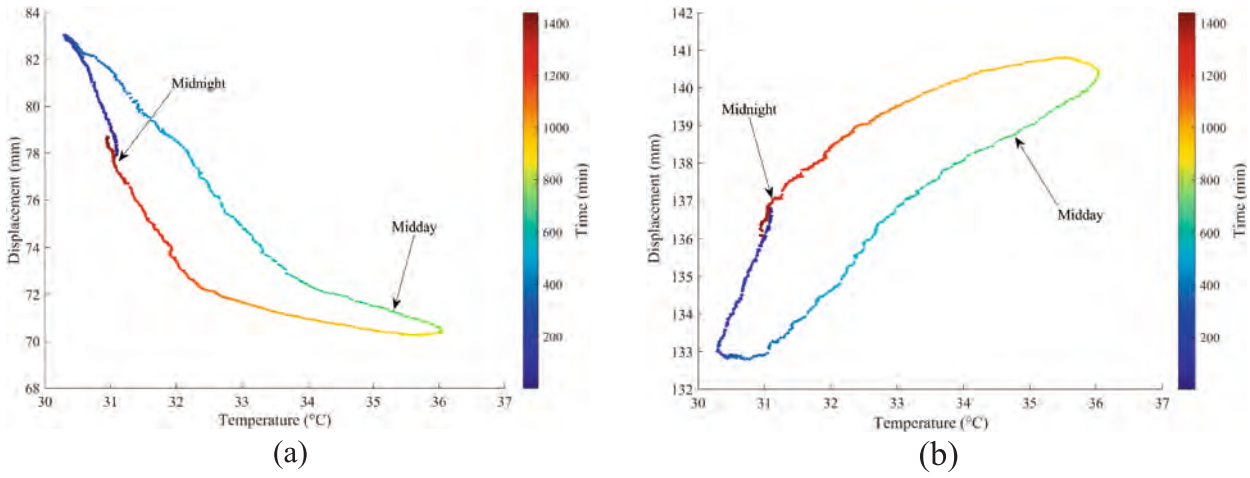

Fig. 6. Daily cyclic behaviour of the temperature-displacement correlation for sensor 1A (a) and sensor 2A (b). The colour scale starts at midnight (dark blue) and moves to the same hour of the next day (dark red). Midnight and midday are indicated as reference timesteps, but do not coincide with the minimum and maximum recorded temperatures due to thermal inertia. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sufficient thermal force accumulates to overcome friction, after which displacement occurs. A similar pattern is observed during cooling, wherein contraction is delayed until the opposing frictional threshold is exceeded.

This friction-induced hysteresis produces a looped temperature–displacement curve throughout the day, characteristic of energy dissipation and mechanical nonlinearity. The magnitude of the frictional resistance, and thus the extent of the hysteresis, may increase with bearing wear, corrosion, or inadequate maintenance [50]. As such, the state of the bearings is a critical parameter in the thermo-mechanical response of bridges.

This cyclic and nonlinear relationship between temperature and structural displacement highlights the need to account for temporal dynamic effects, such as bearing friction and hysteretic behaviour, when developing predictive models.

4.2. Seasonal trends

Regarding long-term effects, seasonal variations have a more significant and lasting impact on the bridge's longitudinal displacements. In the summer, the overall rise in temperatures causes a pronounced expansion of the bridge deck, while in the winter months, cold conditions lead to contraction. The annual temperature trend (Fig. 7) illustrates how summer temperatures, particularly those in August 2024, are consistently higher than the rest of the year. As expected, in the same period of the year, the bearing displacement also reaches its maximum.

Once again, when plotting the displacement trend over more days in the temperature domain, a cyclic behaviour is observed, as these cycles are the superposition of daily ones. The values observed in the hottest month (August 2024) are shown in Fig. 8.

4.3. Temperature-Displacement linear correlation

A preliminary assessment of the temperature–displacement relationship is carried out using a basic white-box model, illustrated in Fig. 9 as the purple line. This model relies on a linear regression applied across the complete dataset, with the parameters estimated by minimising the sum of squared residuals (i.e., the MSE). This procedure was deemed more correct than other linearisation options. The quality of the linear fit was evaluated using the Pearson correlation coefficient, although the regression parameters were obtained by minimising the residual sum of squares.

The resulting fit demonstrates a strong linear association between temperature and longitudinal displacement; for instance, sensor 2A exhibits a Pearson correlation coefficient of $\rho = 0.954$, indicating a high degree of alignment between the two variables. The predicted displacement values were then generated and compared against the actual measurements. The corresponding time series plots of the predictions and associated residuals are presented in Fig. 10, offering a visual assessment of the model's performance over time.

Table 1 reports the performance metrics of the white-box model. The magnitude of the estimation errors suggests that a linear

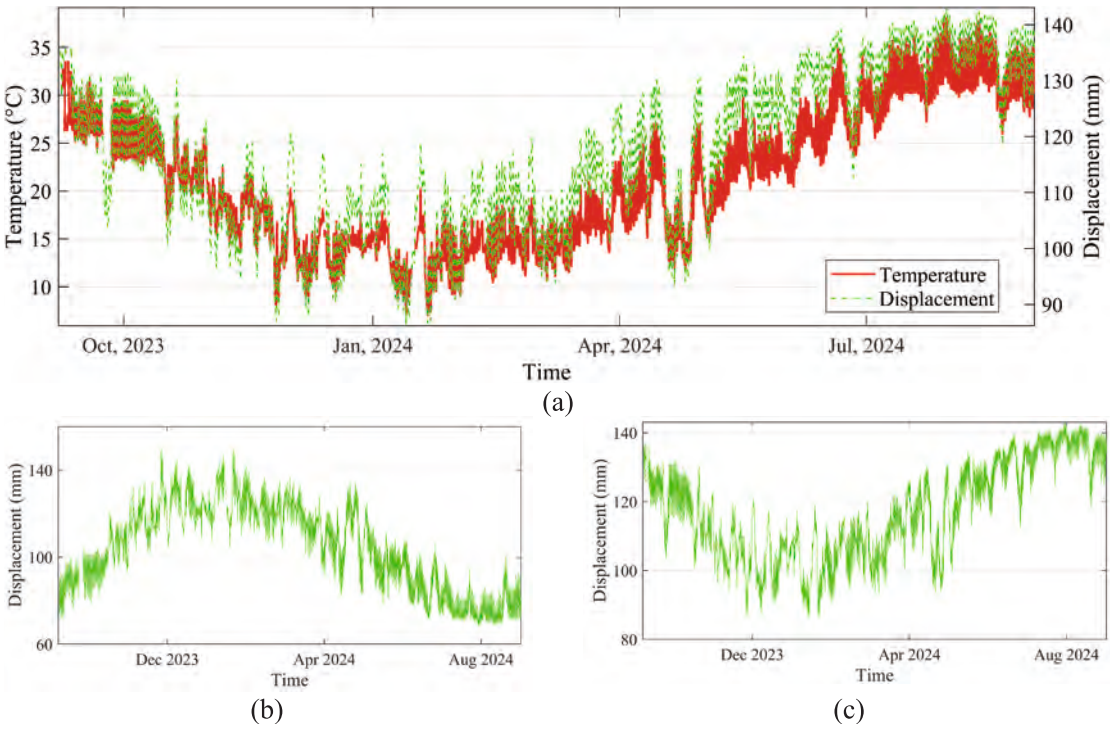

Fig. 7. Annual temperature trend and sensor 2A displacement (a). Annual displacement trend for sensor 1A (b) and sensor 2A (c).

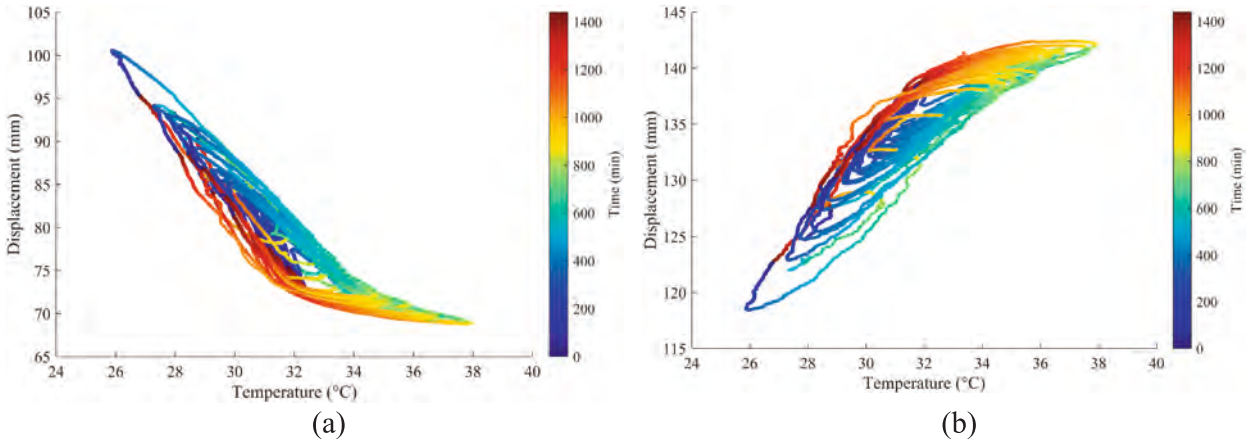

Fig. 8. Cyclic behaviour of the temperature-displacement correlation observed in August 2024 for sensor 1A (a) and sensor 2A (b), showing both the intraday and interday variability of measured data.

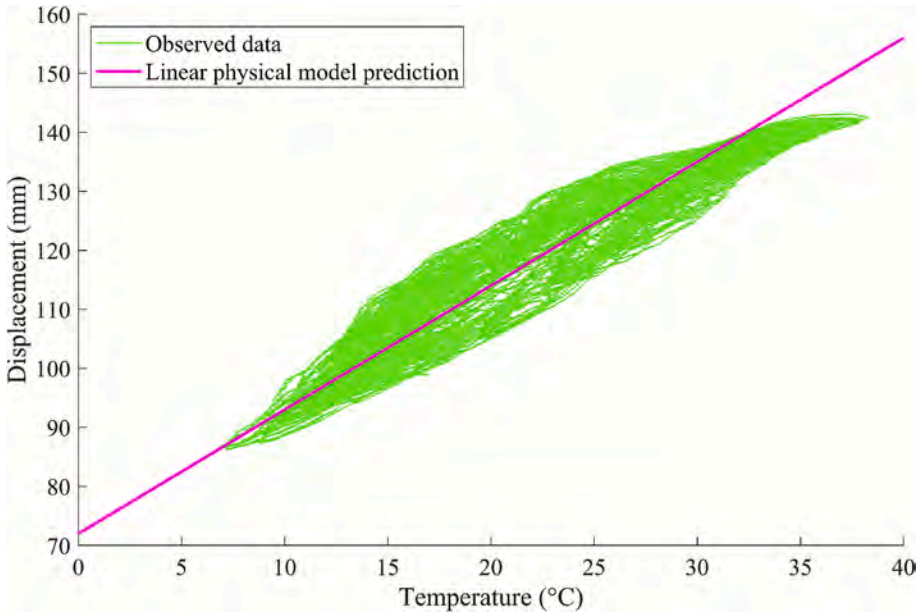

Fig. 9. Linear correlation between temperature and abutment bearing displacements (sensor 2A).

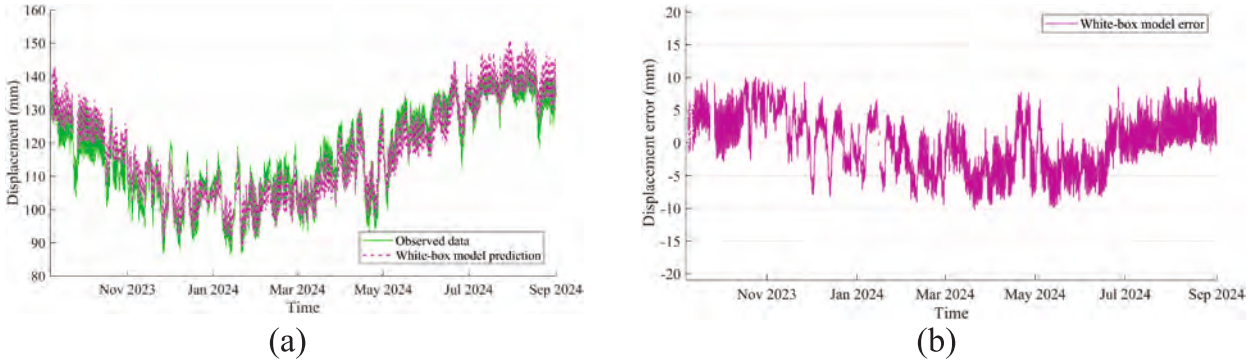

Fig. 10. Observed and predicted displacements (a) and displacement prediction error (b) of the white-box model (sensor 2A).

temperature–displacement correlation captures the dominant structural response of the bearings, as temperature is identified as the primary driver of the measured displacements. This high explanatory power provides additional support for the legitimacy of using the linear prior within the grey-box framework, as it confirms that the primary physical mechanism is already well represented. Although effects such as bearing deterioration or time-varying friction exist, their impact is relatively minor compared to the temperature-driven response and is effectively accounted for by the black-box residual. Values of MAE below 0.5 mm and RMSE below 1 mm are considered satisfactory for this application, as these thresholds ensure that the model's typical prediction error remains well below the magnitude of the displacements to be detected. In the grey-box PIML formulation, the data-driven residual refines the physical prior, thus improving accuracy. Preliminary investigations indicated that moderate variations in bearing friction or gradual stiffness degradation produce residuals within the magnitude range already captured, supporting the adequacy of the linear prior for first-order predictions. For real-world deployment, periodic recalibration of the model is recommended to account for potential long-term changes in structural behaviour, ensuring sustained prediction performance over time.

4.4. Accuracy Comparison of hybrid and fully Data-Driven models

To effectively capture the nonlinear behaviour of the bridge bearing displacements, data-driven components are added, employing both black-box and grey-box methodologies. The predictive models leverage a univariate temperature input, along with two auxiliary variables: time of day and day of the year. These additional temporal inputs were included to reflect the inherent diurnal and seasonal cyclicity observed in the structural response. Preliminary investigations into other environmental variables, such as wind speed and relative humidity, revealed that their influence on longitudinal bearing movements was negligible, in line with findings reported in SHM literature [3,6]. As their inclusion did not yield a measurable improvement in model performance, these variables were excluded to maintain parsimony without compromising predictive accuracy.

For the GPR models, an exponential kernel (or covariance) function was selected. Due to its robustness and flexibility, it is well-suited for modelling smooth nonlinear dependencies. Model hyperparameters, including the length scale and signal variance, were automatically estimated by maximising the Negative Log Marginal Likelihood (NLML) of the training data, following standard Bayesian inference procedures [39]. This estimation was carried out by MATLAB's "fitrgp" function using internal numerical optimisation routines [51].

For illustration, the GP in the black-box models estimated a length scale of approximately 21,000, a signal variance of 136, and an NLML of –3,665. In contrast, the grey-box models yielded a length scale of 760, a signal variance of 23, and an NLML of –3,433. These values are provided for illustrative purposes only, and they may vary depending on the specific dataset or model configuration, reflecting the grey-box approach's ability to capture the main trends through the physical model while leaving the GP to model smaller residual variations. The grey-box approach, thanks to data-driven refinement, provided superior results in replicating the bearings' actual displacements (Table 2).

In particular, the NMSE for this model was exceptionally low, indicating high accuracy in the predicted deck movements. These results underscore the capability of hybrid modelling approaches to more effectively track both daily and annual variations in structural response, outperforming purely empirical methods. This is also supported by the excellent correspondence between the predicted and measured displacements, as illustrated in Fig. 11 and Fig. 12.

The incorporation of time-related variables enables the grey-box model to more effectively capture both diurnal and seasonal trends, thereby markedly improving the replication of the observed displacement behaviour. This enhanced capability is evident in the model's ability to reproduce the cyclic patterns present in the data, as illustrated in Fig. 13. While temperature as a single input captures the nonlinear yet predominantly elastic response, characterised by a smooth, mean-path behaviour during both heating and cooling phases, it does not fully explain the observed displacement dynamics. By including time-related variables, the model can capture not only this nonlinearity but also the inelastic, hysteretic nature of the response, effectively representing the cycles and energy dissipation, leading to a more faithful reproduction of the complex displacement behaviour exhibited by the bearings. Importantly, this temporal input allows the model to implicitly account for the system's memory, capturing hysteresis effects that arise from the path-dependent accumulation of thermal and mechanical influences over time.

To provide a comprehensive view, Table 3 also presents the results obtained from sensor 1A, located on the opposite abutment. As expected, the outcomes show patterns consistent with those observed for abutment 2, confirming the uniformity of behaviour on both sides of the structure. This alignment supports the robustness of the modelling strategy and strengthens confidence in the predictive framework's applicability to different parts of the viaduct.

Table 1
Performance metrics of the white-box model predictions.

White-box Model		
Performance Metrics	*sensor 1A*	*sensor 2A*
MAE	2.699	3.582
Std. dev.	3.368	4.283
MSE	11.346	18.365
RMSE	3.369	4.285
NMSE	2.911	9.350
R^2	0.9709	0.9065

Table 2
Performance metrics comparison of the GP models (sensor 2A).

Metrics Comparison (sensor 2A)			
Performance Metrics	*White-box*	*Black-box*	*Grey-box*
MAE	3.582	0.459	0.143
Std. dev.	4.283	2.116	0.226
MSE	18.365	4.519	0.051
RMSE	4.285	2.126	0.226
NMSE	9.350	2.316	0.026
R^2	0.9065	0.9766	0.9997

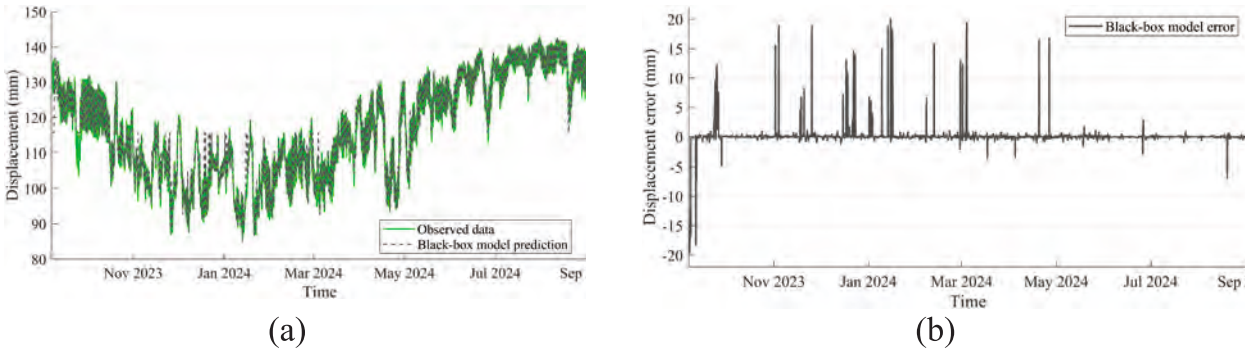

Fig. 11. Observed and predicted displacements (a) and displacement prediction error (b) of the black-box model (sensor 2A).

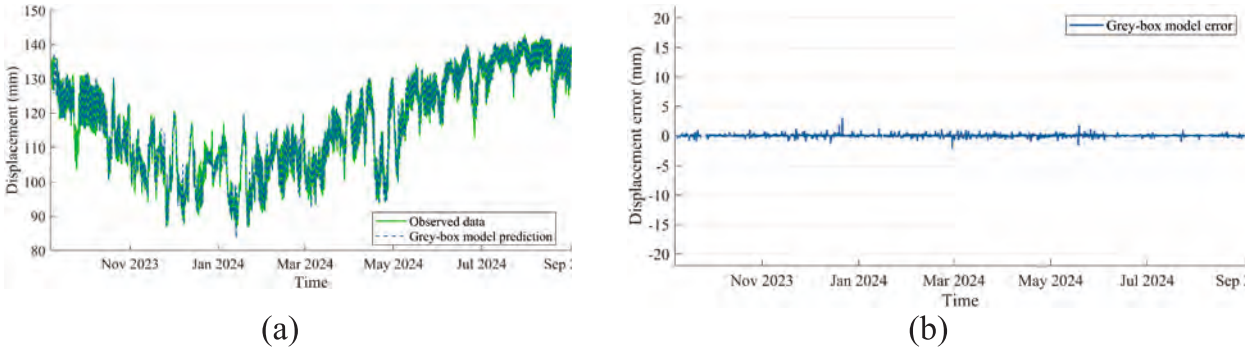

Fig. 12. Observed and predicted displacements (a) and displacement prediction error (b) of the grey-box model (sensor 2A).

Overall, the analysis highlights the significant nonlinearity exhibited by the bearings, as evidenced by the displacement trends across temperature variations. The presence of clear cyclic patterns underscores the limitations of simple linear physical laws in fully capturing the structural response. To better represent this complexity, temperature inputs were augmented with temporal information at multiple levels. Given that cyclic phenomena are intrinsically time-dependent, including time as a predictive feature proved essential for capturing the full behavioural dynamics of the structure. This enhancement significantly improved model performance, particularly in resolving nonlinear dependencies. The hybrid grey-box approach yielded robust results. The approach consistently demonstrated strong generalisability across all the monitored bearings. Notably, the grey-box model outperformed both the standalone white-box and black-box models, demonstrating that residual-based correction strategies lead to higher predictive accuracy.

4.5. Analysis of residuals and path-dependent effects

To further investigate subtle behaviours not captured by the white-box component of the grey-box model, we performed a dedicated analysis of the trained GP residual function $\delta(\Delta T,t)$. Residuals were separated into heating and cooling phases based on the sign of the temperature rate and grouped into temperature bins (Fig. 14). The comparison of binned mean residuals, shown together with their standard errors (SE), highlights a small but consistent offset between the two phases, with cooling residuals slightly higher than heating residuals across a wide temperature range. The SE provides an indication of the variability of the mean residuals within each temperature bin. Its inclusion allows distinguishing genuine trends from statistical fluctuations: the systematic offset observed between

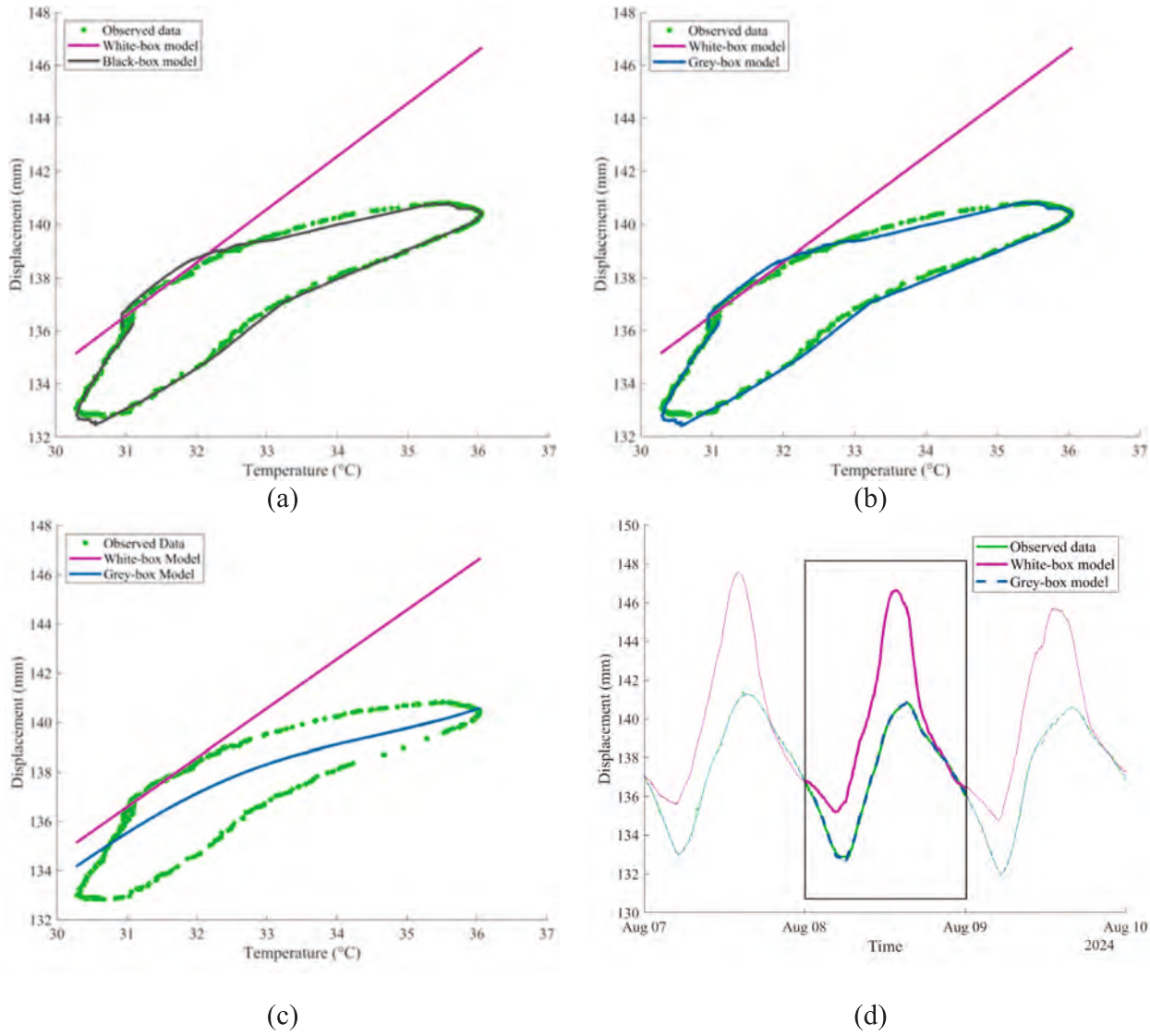

Fig. 13. Prediction of a daily cycle using the black-box (a) and the grey-box (b) model. The same prediction is also performed with the grey-box model, excluding time as a training input (c). Time series of the observed and the predicted displacements, with white- and grey-box models in the corresponding period, i.e., August 8 (d) (sensor 2A).

Table 3
Performance metrics comparison of the models (sensor 1A).

Metrics Comparison (sensor 1A)			
Performance Metrics	*White-box*	*Black-box*	*Grey-box*
MAE	2.699	0.389	0.149
Std. dev.	3.368	2.240	0.235
MSE	11.346	5.017	0.055
RMSE	3.369	2.240	0.235
NMSE	2.911	1.295	0.014
R^2	0.9709	0.9871	0.9999

heating and cooling lies beyond the SE envelope, confirming that the effect is not due to random noise but rather to an underlying path-dependent mechanism. While its magnitude is limited and the direct causal link to mechanisms such as variations in bearing friction cannot be firmly established, the observed difference between heating and cooling phases clearly illustrates the presence of minor hysteresis-like effects. These findings demonstrate that the residual component of the grey-box model can serve as a sensitive indicator

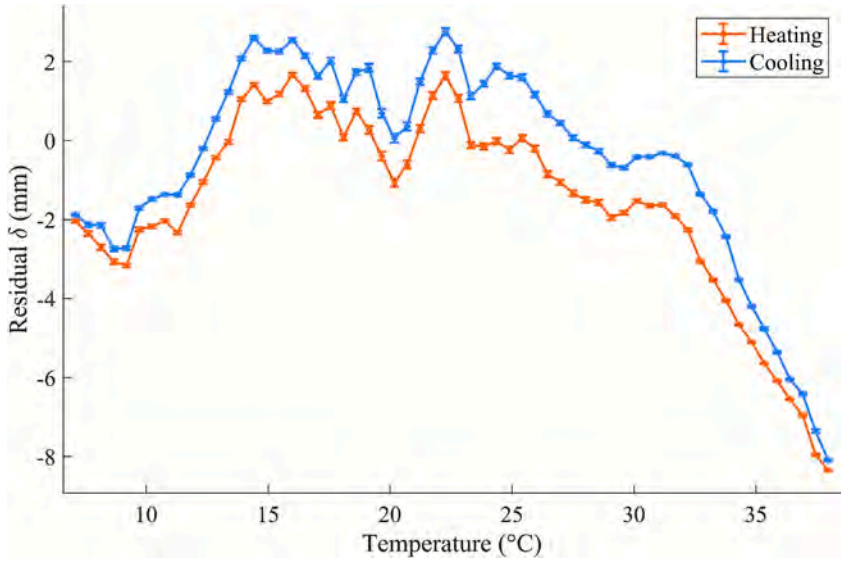

Fig. 14. Binned mean residuals (δ) for heating and cooling phases.

of unmodelled physical behaviours, offering additional interpretative value beyond predictive performance.

4.6. *Influence of sampling proportion on model accuracy*

Additional tests were conducted to evaluate the GP grey-box model's sensitivity to the sampling proportion of the available dataset. To reduce computational cost and avoid biases due to temporal dependencies, random subsets of the full dataset were extracted and used for both training and testing. three different proportions were considered: 2 %, 5 %, and 20 % of the full dataset. Each subset was obtained after thorough shuffling to ensure representative sampling and to mitigate autocorrelation.

The results, summarised in Table 4, show that, as expected, reducing the dataset size leads to a slight decrease in predictive accuracy for the grey-box model. However, even with the smallest subset (2 %), the GP models maintain high accuracy and consistently outperform the white-box model. Increasing the sampling proportion to 5 % or 20 % results in only marginal improvements in accuracy, while substantially increasing computational time and resource requirements. These findings indicate that a 2 % random subset of the dataset is sufficient to achieve reliable predictions. This strategy strikes an effective balance between computational efficiency and model performance, demonstrating that the GP approach remains robust even with a very limited training dataset.

5. Early warning system

The implementation and testing of an early warning system for detecting anomalous bearing displacements will be presented hereinafter. To evaluate the effectiveness of the proposed method, it was necessary to have data representing a damaged state of the viaduct. However, as no such data were available, a simulated damage scenario was created by introducing an artificial abnormal displacement into the existing dataset. This approach enabled testing the system's capacity to detect and respond to abnormal conditions, providing valuable insights into its practical application and effectiveness.

5.1. *Introduction to early warning systems*

SHM is crucial for early detection and prevention of structural failures or degradation that could lead to sudden collapse or severe

Table 4
Performance metrics comparison for the grey-box model (sensor 2A) using different sampling proportions).

Metrics Comparison (sensor 2A) using different Sampling Proportions			
Performance Metrics	*2 %*	*5 %*	*20 %*
MAE	0.143	0.122	0.105
Std. dev.	0.226	0.201	0.177
MSE	0.051	0.047	0.041
RMSE	0.226	0.217	0.202
NMSE	0.026	0.022	0.018
R^2	0.9997	0.9998	0.9999

damage. Within this framework, EWS are essential tools for promptly identifying signs of anomalies or deterioration. In civil engineering, such systems are fundamental for the effective management and maintenance of critical infrastructure, including bridges, dams, tunnels, and high-risk buildings.

Civil structures, particularly those subjected to cyclic loading and environmental exposure, undergo progressive degradation over time. This deterioration can lead to unforeseen failures, posing serious threats to public safety and generating substantial economic burdens. In this context, EWS play a crucial role by enabling the timely and accurate detection of anomalous variations in structural parameters. Such systems, with continuous real-time monitoring indicators, facilitate interventions at the onset of degradation, offering a level of precision that surpasses that of traditional visual inspections. By allowing for early detection and targeted maintenance, EWS support a shift from reactive to preventive strategies, thereby mitigating the likelihood of irreversible damage and reducing maintenance costs. Beyond safeguarding structural integrity, these systems contribute to more sustainable infrastructure management. Through optimised resource allocation and the minimisation of emergency interventions, EWS enhance operational efficiency, thereby extending the service life of structures and improving economic and environmental sustainability. In the static monitoring of bridges, EWS rely on a network of strategically positioned sensors installed on key structural components such as decks, piers, abutments, bearings, and joints. These systems often include temperature sensors to account for seasonal variations influencing thermal expansion and contraction. The data collected are transmitted to a central monitoring unit, where they are processed and compared in real time against a reference baseline. This baseline must be sensitive to meaningful changes in the structural system, and insensitive to normal operational changes [15].

A key application of EWS lies in the implementation of hierarchical alert frameworks that enable calibrated responses based on the severity of detected anomalies. This multilevel structure allows for a nuanced interpretation of monitoring data: low-priority alerts may correspond to minor deviations within expected operational ranges, whereas intermediate alerts can prompt further investigation through visual inspections or targeted non-destructive testing. When predefined safety thresholds are exceeded, high-priority alerts are generated, potentially necessitating immediate structural assessments or even the temporary closure of the infrastructure to ensure public safety. Such a stratified alert system is crucial for tailoring responses to the nature and magnitude of observed changes, e.g., distinguishing between normal operational variations, such as seasonal displacements, and anomalies that may signal the onset of structural degradation.

5.2. *Early warning method for bearing displacements*

Bridge bearings are critical components in bridge engineering, serving as the interface between the superstructure and the substructure. Their proper functioning is essential to ensure the bridge's structural integrity and serviceability. Consequently, monitoring their displacements is fundamental for assessing overall structural health. Longitudinal displacements can arise from various factors, including temperature fluctuations, traffic loads, and wind actions. Early detection of abnormal displacements can provide valuable indications of potential structural issues, such as excessive movements, which may result from structural changes, including bridge expansion joint failures [52,53]. Another possible cause of abnormal bearing displacement could be the partial restriction of movable bearings [54].

To accurately identify deviations from expected behaviour, it is essential to define a robust baseline representing the expected displacements under normal operational conditions. The core principle of predictive modelling in this context is to replicate the structural response under typical environmental and loading scenarios. If the model can reliably forecast the expected structural response, any significant deviation between the predicted and observed displacements, referred to as residuals, may serve as a reliable indicator of potential anomalies or early-stage damage. Under healthy conditions, these residuals are typically small and often assumed to follow a Gaussian distribution, indicating that the model effectively captures the system's actual behaviour [55]. However, deviations from this expected distribution may suggest that the structure is no longer operating within its normal regime, possibly

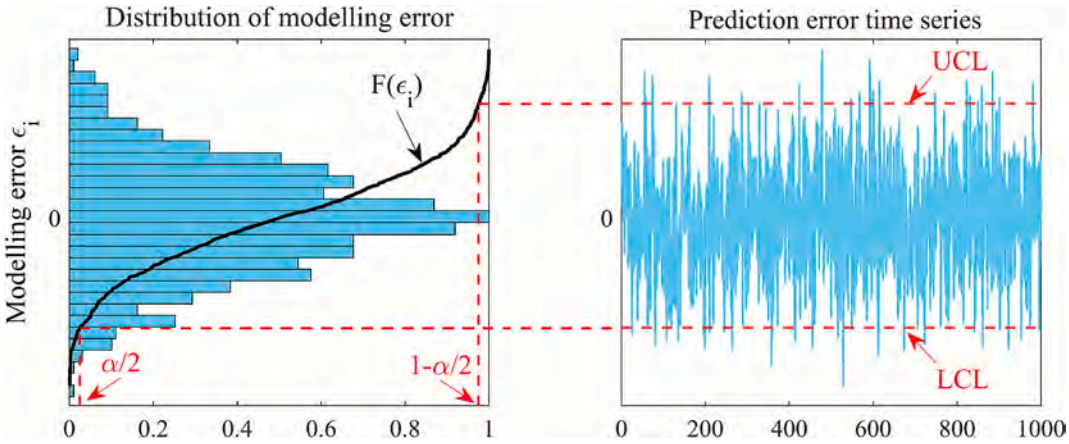

Fig. 15. Diagram of abnormality warning on a random dataset.

indicating degradation or other irregularities. The key concept is that while the model accounts for normal operational variability, significant changes in the distribution of residual errors may reveal novel structural conditions.

Once the predictive model based on temperature variation is trained over a sufficiently long period, it is used to establish the baseline of expected displacements under normal operational conditions, assuming the structure is in a healthy state and no abnormal displacements are present. This baseline model helps evaluate modelling error, defined as the difference between the measured and model-predicted displacements. When analysing new data, the prediction error is calculated as the difference between the observed displacement and the predicted displacement:

$$\varepsilon = d - \widehat{d} \tag{3}$$

where ε is the estimation error, d is the observed displacement, and $\widehat{d}$ is the predicted displacement.

Under normal operating conditions, the statistical properties of the prediction error are expected to resemble those of the modelling error [19]. However, when the bearing displacement is abnormal, the statistical characteristics of the prediction error will shift. As illustrated in Fig. 15, comparing the prediction error with the modelling error provides a diagnostic criterion for determining whether the bearing displacement is abnormal or exceeds the design limits. Specifically, when the prediction error exceeds the control limits established from the modelling error, it may indicate the occurrence of potential anomalies or displacements that exceed the design specifications. The control limits are derived from the distribution of the modelling error associated with the adopted predictive model and serve as reference thresholds for anomaly detection.

Assuming the data follow a normal distribution, there is a 99.73 % probability that prediction errors fall within a $\pm 3\sigma$ range around the mean, which is zero. This condition is typically met under normal circumstances. However, when the errors do not follow a normal distribution, control limits are determined by fitting the error distribution using the KDE method and evaluating the required quantiles. KDE is a non-parametric technique used to estimate the probability density function of a dataset without assuming any specific underlying distribution. It smooths the observed data points to obtain a continuous density curve, providing a flexible representation of the data distribution, even in the presence of skewness, multimodality, or outliers. For a more detailed mathematical treatment of KDE and bandwidth selection, see Parzen [56] and Silverman [57].

The control limits, referred to as the Upper Control Limit (UCL) and Lower Control Limit (LCL), mark the boundaries beyond which displacement deviations are considered abnormal and trigger an early warning. The UCL and LCL can be expressed as:

$$\frac{\alpha}{2} \leq F(LCL \leq \varepsilon \leq UCL) \leq 1 - \frac{\alpha}{2} \tag{4}$$

where α is the desired confidence level parameter, and $F(\varepsilon)$ is the cumulative distribution function. The control limits are calculated as:

$$LCL = F^{-1}\left(\frac{\alpha}{2}\right) \quad \text{and} \quad UCL = F^{-1}\left(1 - \frac{\alpha}{2}\right) \tag{5}$$

In this study, a multi-tiered warning system is proposed, based on control limits derived from confidence intervals. Two initial warning levels are established based on statistical thresholds associated with the predictive model's uncertainty.

An additional warning level can be introduced by monitoring exceedances of the design displacement limits, which account for expected loads, environmental conditions, and a safety margin. Whenever the measured displacement exceed this threshold, it would signal a potential risk to the bridge's structural integrity. This third-level warning involves stricter control limits based on design displacement. Unlike the model-based thresholds, this highest warning level is tied to the bridge design specifications. It can be defined using either nominal design values or empirical displacement ranges observed during normal operation. The definition of warning levels may be tailored to various factors, including the bridge's structural typology, its functional role, environmental exposure, and historical performance data.

When formal design methods or code-specified displacement margins are unavailable, the difference between the annual maximum and minimum displacements can serve as a preliminary reference for setting control limits. In our proof-of-concept, a 10 % threshold of this yearly displacement range was adopted to illustrate the procedure. This choice is consistent with recommendations in the literature [19], where it is suggested that, to prevent the bearing displacement from exceeding design limits, the prediction error should not exceed 0.1 times the basic displacement. The basic displacement is defined as the total displacement resulting from factors such as structural temperature variations, concrete shrinkage and creep, and traffic loads. Accordingly, the prediction error must remain within this margin to ensure that the bearing displacement does not exceed its design limits. It is important to note that this value has no universal engineering justification and is provided solely as an example; in practice, thresholds should be defined on a case-by-case basis according to bridge-specific characteristics, engineering judgment, historical monitoring data, or reliability-based calibration. Accordingly, illustrative control limits can be expressed as:

$$LCL = -\beta|max[d] - min[d]| \tag{6}$$

$$UCL = +\beta|max[d] - min[d]| \tag{7}$$

where β is a reference fraction of the annual displacement range (e.g., 0.1 in our example) and should be adapted to the specific structural context.

5.3. Case study: Predictive models' accuracy for EWS

To evaluate the proposed methodology, long-term monitoring data from the highway viaduct were utilised. Temperature and displacement data were collected over a one-year period, from September 2023 to August 2024, with January and August identified as the months with the lowest and highest temperatures. Structural temperatures ranged from 6.84 °C to 38.22 °C, capturing the full thermal variation. Extreme-temperature months tend to produce maximum and minimum displacement values, increasing the risk of exceeding designed limits. High temperatures during this month can cause significant thermal expansion, making it essential for assessing abnormal bearing displacements. For this reason, this case study will focus on August, as it is reasonable to consider this period the most critical with respect to abnormal bearing displacements. The high temperatures observed during this month can induce substantial thermal expansion, potentially leading to displacement values exceeding or approaching the structural design thresholds.

To develop an EWS capable of detecting anomalous displacements, it is essential to establish what the expected values are under normal conditions. Baseline data are commonly used to characterise the normal structural response, serving as a reference against which new measurements can be compared to assess whether the current structural behaviour remains within normal conditions [55]. The baseline is generated by utilising the predictive models discussed previously. For a comprehensive evaluation, GP-based models are implemented utilising both black- and grey-box approaches. Additionally, white-box results are included as a reference benchmark. For the GP models, the exponential kernel function, previously identified as the most effective, is used, with hyperparameters automatically optimised during training. Again, a 70 % training ratio is applied.

The results (Table 5) exhibit a strong consistency with those obtained when evaluating the models over the entire year of available data. However, a notable distinction emerges in the white-box model's performance, which shows a marked tendency to overestimate displacements during this focused time window (Fig. 16).

Conversely, both the black-box and grey-box models demonstrate improved accuracy (Table 5). Results are also reported in Fig. 17 and Fig. 18. This enhancement is likely due to their ability to capture recent patterns under relatively stable environmental conditions. Over more extended periods, data distribution shifts, evolving external influences, or variable drift can hinder model generalisation. Thus, when applied to short and stable time frames, data-driven models offer superior precision.

For completeness, the results for abutment 1 are also presented in Table 6. Here, a notable difference is observed (Fig. 19): the white-box model tends to underestimate displacements, whereas it overestimates in abutment 2. This discrepancy can be attributed to the opposite sign convention applied to abutment 1. In this configuration, the bridge deck's thermal expansion results in a negative displacement.

5.4. Early warning analysis for bearing displacement

In line with established practices in the SHM literature [19,58], the EWS is evaluated using test data that simulate artificially altered displacements; an anomaly was introduced to emulate the effects of a damaged bridge bearing. Since real damage events are often unavailable [3], this approach allows for a controlled evaluation of model performance. Since no actual damage was observed in the monitored bridge, the simulated anomalies constitute the sole viable option, even with all their limitations, as they cannot fully reproduce the progressive and nonlinear development of real structural damage, such as gradual crack propagation, corrosion, or stiffness degradation. Furthermore, in real-world applications, thresholds must be calibrated dynamically to account for changing environmental and operational conditions, including seasonal temperature variations and evolving traffic loads.

In the context of a bridge, any abnormalities in bearing displacement resulting from changes in structural performance are expected to increase gradually over time, as structural degradation is a long-term progressive process. In a short time, significant abnormalities in bearing displacement typically do not occur continuously, except in cases of unforeseen incidents. Therefore, introducing a slight variation in the displacement data during the test phase is a reasonable approach to simulate abnormalities in bearing displacements resulting from changes in structural performance [19]. This consideration enables the accurate reflection of the gradual abnormalities in bearing displacement induced by a slowly changing structural configuration. This leads to the expression:

$$d_{abn} = d + \delta_e \tag{8}$$

where d is the measured bearing displacement, d_{abn} is the abnormal displacement and δ_e is the artificially added displacement. The simulated displacement abnormality δ_e was varied from 0 to 10 mm in increments of 0.01 mm, creating a continuous damage distribution from 0 mm (healthy condition) to 10 mm (very high damage). This approach enabled assessment of the alert system's

Table 5
Performance metrics of the GP models tested on data from August (sensor 2A).

Metrics Comparison (sensor 2A)			
Performance Metrics	*White-box*	*Black-box*	*Grey-box*
MAE	3.039	0.142	0.092
Std. dev.	2.355	0.577	0.144
MSE	14.374	0.334	0.021
RMSE	3.791	0.578	0.143
NMSE	71.379	1.563	0.102
R^2	0.2862	0.9844	0.9990

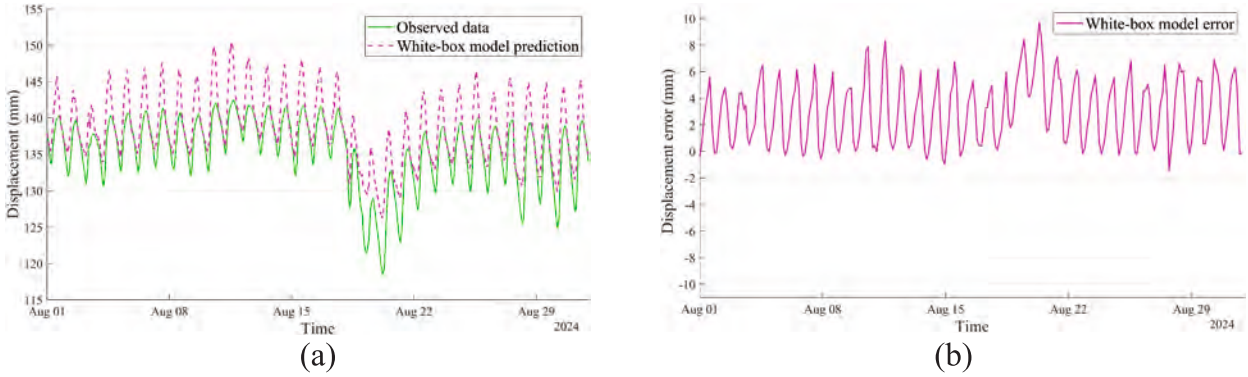

Fig. 16. Observed and predicted displacements (a) and displacement prediction error (b) of the white-box model (sensor 2A, tested on data from August 2024).

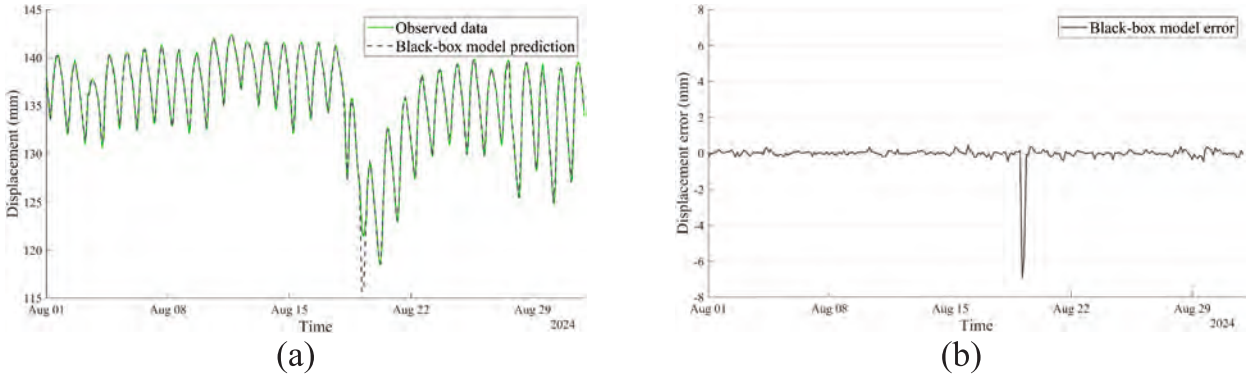

Fig. 17. Observed and predicted displacements (a) and displacement prediction error (b) of the black-box model (sensor 2A, tested on data from August 2024).

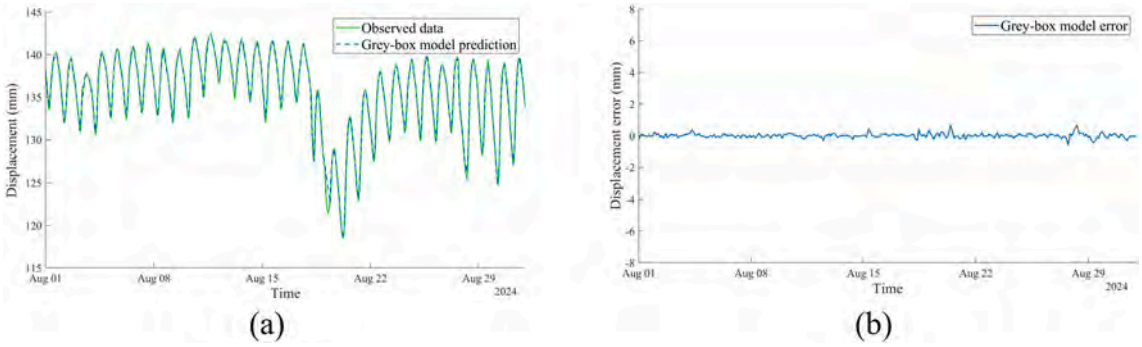

Fig. 18. Observed and predicted displacements (a) and displacement prediction error (b) of the grey-box model (sensor 2A, tested on data from August 2024).

performance across various stages of structural degradation and of the EWS's ability to detect and quantify damage accurately.

5.5. Warning thresholds

Three different warning thresholds have been defined to assess bearing displacement abnormalities:

1. The first threshold corresponds to a 95.4 % confidence interval, representing a range of $\mu \pm 2\sigma$ for a normal distribution.
2. The second threshold is set at a 99.7 % confidence interval, corresponding to $\mu \pm 3\sigma$, again assuming a normal distribution.

Table 6
Performance metrics of the GP models tested in August (sensor 1A).

Metrics Comparison (sensor 1A)			
Performance Metrics	*White-box*	*Black-box*	*Grey-box*
MAE	3.370	0.160	0.112
Std. dev.	2.784	0.713	0.160
MSE	16.294	0.507	0.026
RMSE	4.037	0.712	0.160
NMSE	37.390	1.119	0.058
R^2	0.6237	0.9888	0.9994

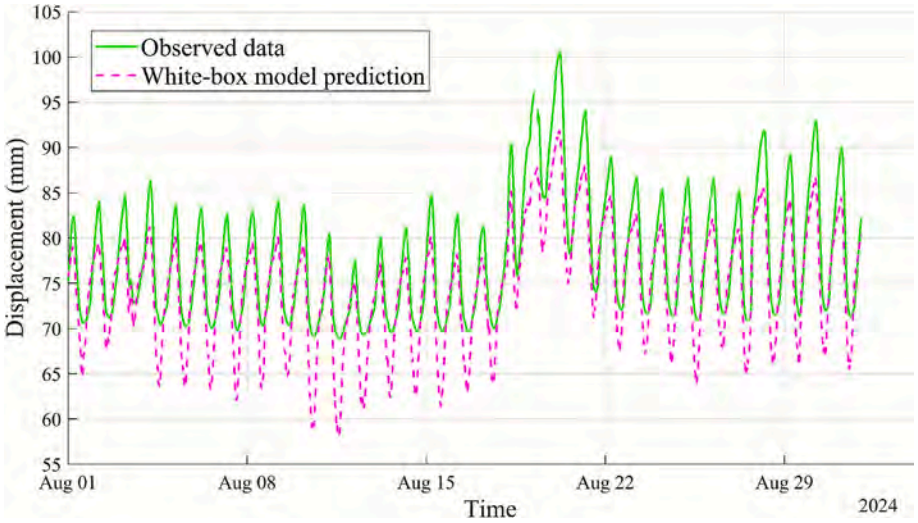

Fig. 19. Displacement prediction of the white-box model tested on data from August (sensor 1A).

3. The third threshold ("Critical Damage") is related to the structural safety margin, set at 10 % of the maximum displacement range between the observed maximum and minimum displacement values for each bearing over the entire dataset.

The prediction error for each bearing's displacement was analysed across different models. Notably, the error distributions for the grey-box and black-box models follow a normal distribution, confirming that they are Gaussian. In contrast, the white-box model's error did not follow a normal distribution, as shown in Fig. 20. In this case, the confidence interval was calculated by estimating the probability density function of the errors of the white-box model using the KDE method. The bandwidth h of the Gaussian kernel was chosen according to the implementation of Silverman's rule of thumb [57], computed as:

$$h = 1.06 \cdot \sigma \cdot n^{-1/5} \tag{9}$$

where σ is the standard deviation of the error data and n is the number of samples. For the white-box model, this resulted in h = 0.72933 for sensor 1A and h = 0.76014 for sensor 2A. Using this explicitly computed bandwidth ensures reproducibility of the estimated quantiles and the corresponding confidence intervals.

The quantiles corresponding to the probabilities assumed in the first two thresholds were determined, as shown in Fig. 21.

The different thresholds corresponding to each warning level are presented in Table 7, Table 8, and Table 9. For simplicity, the values are reported only for sensors 1A and 2A as references, as all other sensors exhibit very similar results.

A hierarchical structure of warning levels can be established to represent increasing degrees of alert severity. In this framework, the warning ratio is defined as the percentage of observations that meet or exceed a given threshold. When the first two warning levels consistently reach 100 %, this strongly suggests the presence of structural anomalies and should trigger a timely inspection.

An escalation in the critical damage ratio further reinforces the need for diagnostic evaluations and immediate maintenance interventions. Should all warning ratios simultaneously reach 100 %, this condition signals a critical exceedance of allowable bearing displacements, indicating a severe threat to structural safety. In such cases, the immediate suspension of bridge operation is warranted to allow for urgent maintenance. In August and January, as expected, bearing displacements tend to approach or reach their annual maximum and minimum values. If displacement abnormalities arise during these months, the likelihood of exceeding the designed displacement limit increases significantly.

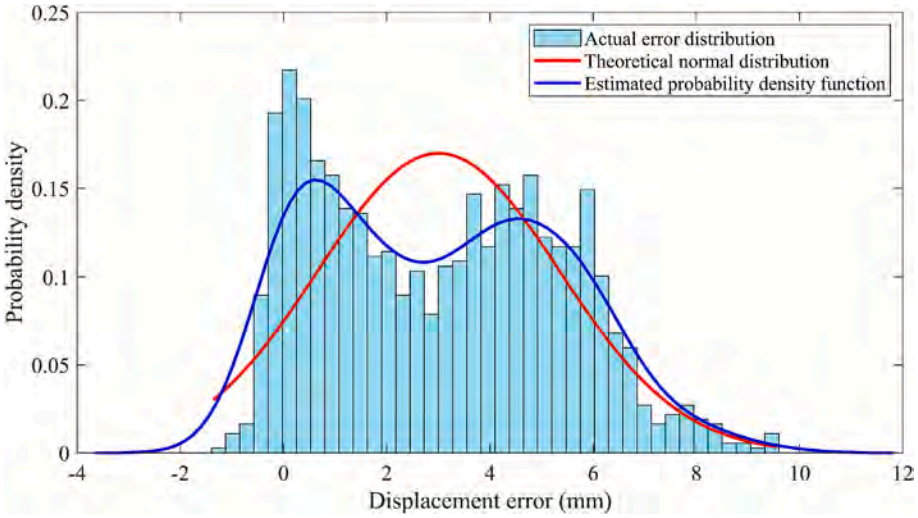

Fig. 20. Estimated Probability Density Functions of the white-box model errors (sensor 2A). The theoretical normal distribution is shown in red for comparison. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

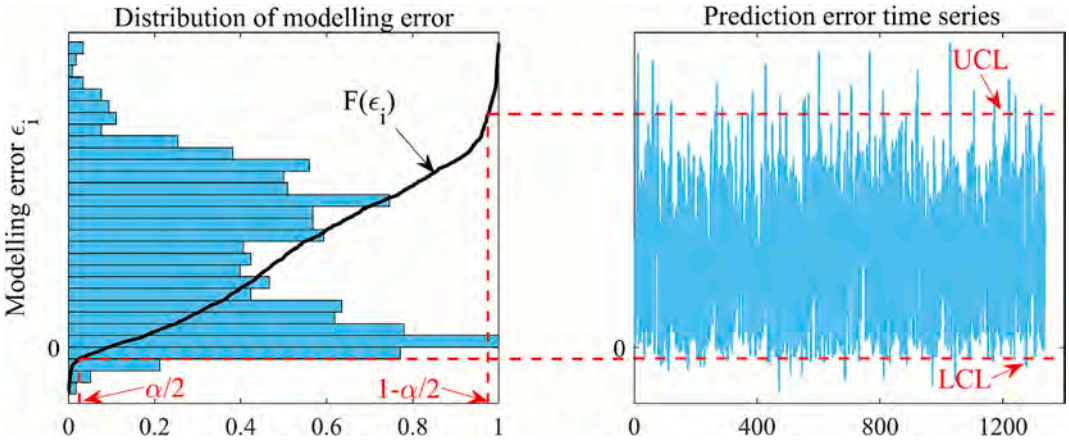

Fig. 21. Diagram of abnormality warning applied to the white-box model errors distribution (sensor 2A).

Table 7
95.4 % confidence interval warning thresholds for the GP model.

95.4 % Abnormality Thresholds		
sensor	1A	2A
White-box UCL (mm)	2.611	8.011
Black-box UCL (mm)	1.446	1.142
Grey-box UCL (mm)	0.345	0.296
White-box LCL (mm)	−8.548	−1.348
Black-box LCL (mm)	−1.446	−1.142
Grey-box LCL (mm)	−0.345	−0.296

5.6. Early warning results with the GP model and sensor 2A

The damage is simulated by adding an increasing positive value of artificial damage to the prediction errors obtained from the white-, black-, and grey-box GP models. The artificial abnormal displacement ranges from 0 mm to 10 mm in steps of 0.01 mm. This procedure allows a systematic assessment of the EWS across a broad spectrum of damage levels, providing a controlled benchmark for evaluating detection sensitivity and threshold performance. By applying these artificial increments to the prediction errors, we can effectively quantify the EWS's ability to distinguish subtle deviations from normal conditions and to calibrate alarm thresholds under

Table 8
99.7 % confidence interval warning thresholds for the GP model.

99.7 % Abnormality Thresholds		
Sensor	1A	2A
White-box UCL (mm)	4.176	10.299
Black-box UCL (mm)	2.169	1.714
Grey-box UCL (mm)	0.518	0.444
White-box LCL (mm)	−11.094	−2.677
Black-box LCL (mm)	−2.169	−1.714
Grey-box LCL (mm)	−0.518	−0.444

Table 9
Critical Damage warning thresholds.

Critical Damage Thresholds		
sensor	1A	2A
UCL (mm)	8.136	5.685
LCL (mm)	−8.136	−5.685

idealised conditions. An example of different levels of artificial damage is provided in Fig. 22, where the corresponding warning thresholds are reported.

For the white-box model, the critical damage threshold is notably lower than the other two thresholds, reflecting the model's relatively low accuracy, as evidenced by its high standard deviation. Since the accuracy is very high for both the black- and grey-box models, the first two warning thresholds are minimal, resulting in a significant gap between them and the critical damage level. This means that it is unlikely that the modelling error triggers the third warning due to a prediction error alone. Consequently, the critical level is reached only when the artificially added damage matches or exceeds the UCL. Once the damaged datasets were generated for varying levels of added displacement, the warning ratios for each model were evaluated as the artificial damage increased. Results are reported in Fig. 23. The warning ratio rises with increasing artificial displacement variation.

For the white-box model, the critical damage warning is triggered even under nominal conditions (i.e., 0 mm of added displacement), primarily due to its relatively high modelling errors. Consequently, the critical threshold is exceeded at displacement values lower than those corresponding to the 95.4 % and 99.7 % confidence intervals, causing the red curve (critical damage) to precede the yellow and orange curves (confidence-based thresholds). This indicates that, when the critical alarm is activated, the likelihood that it reflects actual structural damage (rather than a modelling error) is still low. This results in a high incidence of false positives. Hence, the limited predictive accuracy of the white-box model significantly compromises its reliability and sensitivity in detecting true critical displacements.

In the black-box model, the red curve is clearly positioned well to the right of the other two, indicating that the critical warning is triggered only when there is full confidence in the presence of abnormal displacement, i.e., after the first two warning levels have

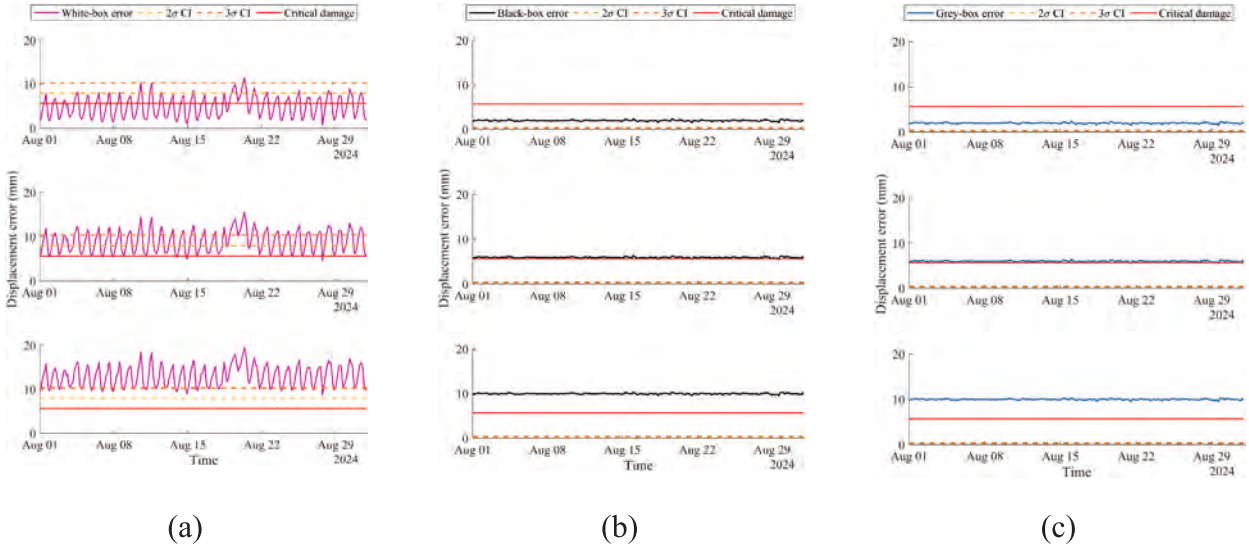

Fig. 22. White- (a), black- (b) and grey-box (c) models' prediction error with different levels of damage (sensor 2A).

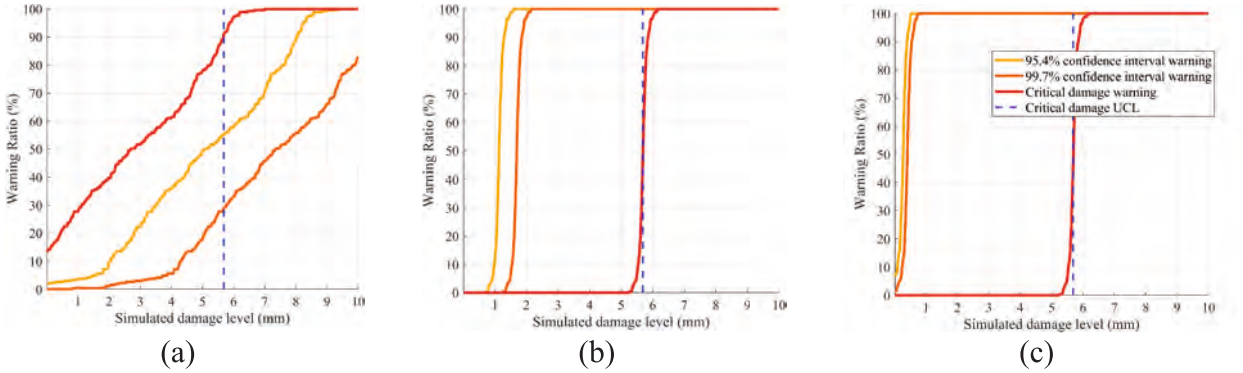

Fig. 23. Warning ratio for the white- (a), black- (b) and grey-box (c) models (sensor 2A).

already reached 100 %. This reduces the likelihood of false alarms. Moreover, the critical warning ratio increases significantly only as the added damage approaches the UCL, confirming that the alarm is raised only under truly critical situations. Notably, even a slight abnormal displacement of just 2 mm leads to the exceedance of the 99.7 % confidence threshold in nearly all cases, demonstrating the model's high sensitivity. This indicates that even minor deviations from normal behaviour are consistently recognised as signs of potential damage, underscoring the model's reliability in early-stage damage detection.

The grey-box model exhibits further improvement, achieving strong accuracy in detecting minimal displacements (remarkably, even submillimetric) with high confidence. This exceptional sensitivity significantly enhances the model's ability to detect degradation in the early stages. To highlight the different performances obtained by each model, their results are compared directly in Fig. 24 (for each different warning level).

The grey-box model consistently performs best across all warning levels, proving superior accuracy and high sensitivity to damage. The black-box model also performs well, though with slightly lower sensitivity than the grey-box model. In contrast, the white-box model shows a high false-positive rate at the critical damage level and fails to reliably detect significant damage, resulting in low confidence in its suitability for accurate early warning applications.

5.7. Early warning results with the GP model and sensor 1A

The same procedure was also repeated for the bearing displacement measured by sensor 1A. In this case, there is an evident tendency of the white-box model to underestimate the displacement (see also Fig. 19).

Results are reported in Fig. 25. Since the white-box model tends to underestimate displacements, the warning levels are very low, even in cases of significant displacement abnormalities, thus heavily increasing the risk of false negatives. This underestimation means the model may fail to flag potentially critical conditions. However, the higher warning ratios associated with the 2σ and 3σ confidence intervals at least ensure that the system is less likely to produce false alarms. For the other two models, the results are very similar to the previous case, showing that the grey-box consistently outperforms the black-box, triggering high warning ratios at the first two warning levels, even for very small added damage.

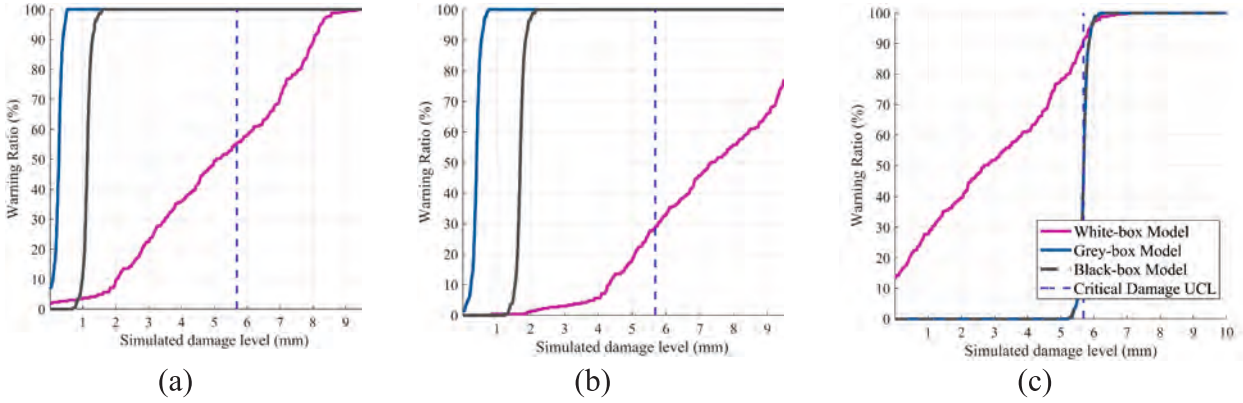

Fig. 24. Comparison of the models using different levels of warning: 95.4 % confidence interval (a), 99.7 % confidence interval (b), and critical damage (c) (sensor 2A).

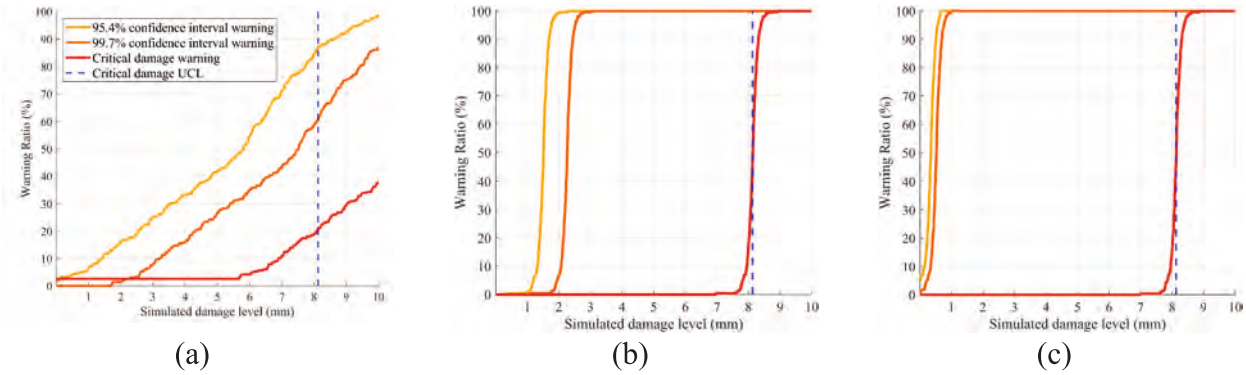

Fig. 25. Warning ratio for the white- (a), black- (b) and grey-box (c) models (sensor 1A).

5.8. Discussion of EWS findings

Monitoring of bearing displacement in bridges is crucial because changes in these displacements can indicate significant structural deterioration or damage to the bearing devices. This study establishes an early warning method for three abnormal levels of bearing displacement by utilising different models, including white-, black-, and grey-box models, with GPs used as regression tools. In particular, the grey-box model has demonstrated its ability to accurately predict displacements, suggesting that deviations from its predictions can serve as an effective anomaly indicator. It is observed that the smaller the modelling error in displacement predictions, the more accurately the predictive model assesses displacement abnormalities. Generally, a significant gap between the yellow (or the orange) and red curves indicates that models with low modelling error can effectively distinguish between an abnormal displacement level and a critical damage level. Conversely, when the modelling error is significant, the model's ability to accurately identify displacement abnormalities diminishes, resulting in a smaller interval between the yellow/orange and red curves, or, in the worst cases, a switch between them (see Fig. 23(a)). In such cases, it becomes crucial to pay closer attention to changes in warning information, as a high prediction modelling error may lead to unreliable assessments of the bridge's structural condition. This highlights the importance of minimising modelling errors to enhance the reliability of early warning systems in SHM. In practical applications, the hybrid grey-box modelling approach is particularly effective for achieving high predictive accuracy, as it offers enhanced sensitivity in detecting even minor structural displacements.

6. Conclusions

This study presented a grey-box modelling approach for predicting bridge bearing displacements. To the best of the Authors' knowledge, this is the first application of Physics-Informed Machine Learning principles to the static monitoring data of a real-life, real size highway viaduct. In more detail, this study falls within the context of residual modelling: a nonlinear, data-driven correction coefficient was inserted into the analytical formulation of thermoelasticity. This research aimed to address the limitations of purely physics-based (white-box) and purely data-driven (black-box) models. By integrating engineering knowledge with machine learning, the hybrid model effectively captured the complex, nonlinear, temperature-dependent, and time-dependent response of the structure to environmental variability, achieving both interpretability and predictive accuracy. In this sense, the grey-box model successfully addressed the shortcomings of white-box approaches, which often fail to account for unmodeled components (especially nonlinear terms), and black-box models, which lack transparency and generalisability.

Key findings include:

- Incorporating time alongside temperature as inputs significantly improved the model's ability to capture thermal effects on displacements.
- The residual-based grey-box model consistently outperformed both black- and white-box approaches. For example, on sensor 2A, it achieved a NMSE of 0.026 %, compared to 2.316 % for the black-box and 9.350 % for the white-box models.
- GPR demonstrated superior accuracy compared to SVM-based alternatives.
- While the grey-box model could extrapolate beyond the training range, its performance depended on the accuracy of the underlying physics-based prior, particularly visible in its tendency to overestimate displacements, inherited from the white-box component.

The second key contribution of this work was the integration of the aforementioned grey-box model with a KDE-based Early Warning System that detects subtle anomalies. In these regards:

- The grey-box model could reliably detect displacement deviations smaller than 1 mm.
- Conversely, all else being equal, the black-box model required deviations greater than 2 mm for consistent detection.
- The white-box model failed to detect even severe anomalies, showing its unreliability for early warning applications.

It should be noted that the study relies on simulated anomalies due to the absence of actual damage. While these provide a controlled benchmark to evaluate the EWS, they cannot fully reproduce the nonlinear progression of real structural damage. As a limitation of the scope, the present study focuses solely on the measurable effects of structural anomalies on bearing displacements. It does not, instead, investigate the underlying physical cause of the damage. Future work will extend the methodology to datasets with progressive damage, implement adaptive thresholding, and explore continuous model updating strategies, as proposed in Worden et al. [59], to maintain reliable detection under changing environmental and operational conditions.

These results also highlight the importance of minimising predictive errors to ensure the effectiveness of SHM alarm systems under environmental and operational variability.

While promising, the model was validated only under typical environmental conditions. Future research will explore its performance under extreme or unseen scenarios through prolonged monitoring of real data, as well as the integration of long-term environmental data, and, potentially, further refinement of the physics-based component to improve generalisability. Finally, although this study focused on bridge bearings, the proposed residual-based grey-box strategy is not limited to this context. The methodology has the potential to be extended to other structural typologies (e.g., multi-storey buildings) by adapting the physical component. In principle, it could also be applied to vibration-based SHM by computing residuals relative to the time series of modal frequencies extracted from vibration measurements. In this case, an appropriate white-box component describing the expected frequency evolution would be required, as the residuals are defined with respect to the physical model. Such an extension, however, remains a scope for future work.

Nevertheless, this research work on a complex case study demonstrates the potential of grey-box models to enhance bridge monitoring systems in real-life conditions, by combining data-driven precision with physical insight. The approach offers a more reliable and interpretable framework for predictive assessment and early anomaly detection, key steps toward robust, automated SHM solutions, robust to EOVs, which are a common source of misleading sensor readings and false alarms in SHM.

CRediT authorship contribution statement

Enrico Cianci: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Marco Civera:** Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Data curation, Conceptualization. **Valerio De Biagi:** Writing – review & editing, Supervision, Resources, Project administration. **Bernardino Chiaia:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

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.

Acknowledgements

This work is part of the research activity developed by the authors within the framework of the “PNRR”: MOST – Sustainable Mobility National Research Centre - SPOKE 7 “Cooperative Connected and Automated Mobility and Smart Infrastructures” - WP4“ and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1033 17/06/2022).

Furthermore, the Authors would like to thank Movyon S.p.A., especially Benedetto Carambia, Davide Chiola, Dario Coletta, and Pierpaolo Dragonetti, as well as the whole Autostrade per l’Italia (ASPI) group, for providing the opportunity for this research.

Data availability

The data were shared from the industrial partner under a non-disclosure agreement, and only the agreed level of detail has been reported here.

References

[1] E. Figueiredo, J. Brownjohn, Three decades of statistical pattern recognition paradigm for SHM of bridges, Struct. Health Monit. 21 (2022) 3018–3054, https://doi.org/10.1177/14759217221075241.
[2] E. Cross, On Structural Health Monitoring in Changing Environmental and Operational Conditions, University of Sheffield, 2012. PhD thesis.
[3] H. Sohn, C.R. Farrar, F.M. Hemez, J.J.A. Czarnecki, Review of Structural Health Review of Structural Health Monitoring Literature, 1996–2001, 2002.
[4] B. Peeters, G. De Roeck, One-year monitoring of the Z24-Bridge: environmental effects versus damage events, Earthquake Engng. Struct. Dyn. 30 (2001) 149–171, https://doi.org/10.1002/1096-9845(200102)30:2<149::AID-EQE1>3.0.CO;2-Z.
[5] K. Worden, Farrar CR, Manson G, Park G. The fundamental axioms of structural health monitoring. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 2007;463:1639–64. 10.1098/rspa.2007.1834.
[6] C.R. Farrar, K. Worden, Structural Health Monitoring: A Machine Learning Perspective, John Wiley and Sons, 2012, 10.1002/9781118443118.
[7] E.J. Cross, T.J. Rogers, D.J. Pitchforth, S.J. Gibson, S. Zhang, M.R. Jones, A spectrum of physics-informed Gaussian processes for regression in engineering, Data-Centric Eng. (2024) 5, https://doi.org/10.1017/dce.2024.2.
[8] B.T. Svendsen, O. Øiseth, G.T. Frøseth, A. Rønnquist, A hybrid structural health monitoring approach for damage detection in steel bridges under simulated environmental conditions using numerical and experimental data, Struct. Health Monit. 22 (2023) 540–561, https://doi.org/10.1177/14759217221098998.
[9] S.J. Gibson, T.J. Rogers, E.J. Cross, Distributions of fatigue damage from data-driven strain prediction using Gaussian process regression, Struct. Health Monit. 22 (2023) 3065–3076, https://doi.org/10.1177/14759217221140080.

[10] Z. Sun, M. Sun, D.M. Siringoringo, Y. Dong, X. Lei, Predicting bridge longitudinal displacement from monitored operational loads with hierarchical CNN for condition assessment, Mech. Syst. Signal Process. 200 (2023) 110623, https://doi.org/10.1016/J.YMSSP.2023.110623.
[11] K. Worden, L.A. Bull, P. Gardner, J. Gosliga, T.J. Rogers, E.J. Cross, E. Papatheou, W. Lin, N. Dervilis, A brief introduction to recent developments in population-based structural health monitoring, Front. Built Environ. 6 (2020) 146, https://doi.org/10.3389/fbuil.2020.00146.
[12] V. Giglioni, J. Poole, R. Mills, I. Venanzi, F. Ubertini, K. Worden, Transfer learning in bridge monitoring: laboratory study on domain adaptation for population-based SHM of multispan continuous girder bridges, Mech. Syst. Signal Process. 224 (2025) 112151, https://doi.org/10.1016/J.YMSSP.2024.112151.
[13] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (2010) 1345–1359, https://doi.org/10.1109/TKDE.2009.191.
[14] K. Worden, E.J. Cross, On switching response surface models, with applications to the structural health monitoring of bridges, Mech. Syst. Signal Process. 98 (2018) 139–156, https://doi.org/10.1016/j.ymssp.2017.04.022.
[15] M.T. Yarnold, F.L. Moon, Temperature-based structural health monitoring baseline for long-span bridges, Eng. Struct. 86 (2015) 157–167, https://doi.org/10.1016/j.engstruct.2014.12.042.
[16] Z. Deng, M Huang, N Wan, J. Zhang, The Current Development of Structural Health Monitoring for Bridges: A Review, Buildings 13 (2023) 1360, https://doi.org/10.3390/BUILDINGS13061360.
[17] X. Lei, Z. Sun, A. Wang, T. Guo, T. Nagayama, Estimation of bridge girder cumulative displacement for component operational warning using bayesian neural networks, Struct. Control Health Monit. 2025 (2025), https://doi.org/10.1155/STC/9974584.
[18] Li S, Gan L, Zhao R, Wang S, Zhou Y. Research on Bridge Integrity Assessment and Early Warning Monitoring Methods Based on Bearing Reaction Force. Buildings 2024, Vol 14, Page 763 2024;14:763. 10.3390/BUILDINGS14030763.
[19] G.-M. Wu, T.-H. Yi, D.-H. Yang, H.-N. Li, H. Liu, Early warning method for bearing displacement of long-span bridges using a proposed time-varying temperature–displacement model, J. Bridg. Eng. (2021) 26, https://doi.org/10.1061/(asce)be.1943-5592.0001763.
[20] X. Lei, Y. Dong, D.M. Frangopol, Integration of inspection and monitoring data for RL-enhanced sustainable life-cycle management of infrastructure networks, Struct. Infrastruct. Eng. 21 (2025) 1288–1302, https://doi.org/10.1080/15732479.2025.2453484.
[21] X. Lei, D.M. Siringoringo, Y. Dong, Z. Sun, Interpretable machine learning methods for clarification of load-displacement effects on cable-stayed bridge, Measurement 220 (2023) 113390, https://doi.org/10.1016/J.MEASUREMENT.2023.113390.
[22] C. Lathourakis, A. Cicirello, Physics enhanced sparse identification of dynamical systems with discontinuous nonlinearities, Nonlinear Dyn. 112 (2024) 11237–11264, https://doi.org/10.1007/S11071-024-09652-2.
[23] S.L. Brunton, B.R. Noack, P. Koumoutsakos, Machine learning for fluid mechanics, Annu. Rev. Fluid Mech. 52 (2020) 477–508, https://doi.org/10.48550/arXiv.1905.11075.
[24] A. Cicirello, Physics-Enhanced Machine Learning: a Position Paper for Dynamical Systems Investigations, 2024, 10.1088/1742-6596/2909/1/012034.
[25] A. Karpatne, G. Atluri, J.H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, V. Kumar, Theory-guided data science: a new paradigm for scientific discovery from data, IEEE Trans. Knowl. Data Eng. 29 (2017) 2318–2331, https://doi.org/10.1109/TKDE.2017.2720168.
[26] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nat. Mach. Intell. 1 (2019) 206–215, https://doi.org/10.1038/s42256-019-0048-x.
[27] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.
[28] Doshi-Velez F, Kim B. Towards A Rigorous Science of Interpretable Machine Learning 2017.
[29] N.N. Kulkarni, A. Sabato, Full-field expansion and damage detection from sparse measurements using physics-informed variational autoencoders, Struct. Health Monit. (2024), https://doi.org/10.1177/14759217241289575.
[30] E. Maria Tronci, A.R.J. Downey, A. Mehrjoo, P. Chowdhury, D. Coble, Physics-Informed Machine Learning Part I: Different Strategies to Incorporate Physics into Engineering Problems. Conference Proceedings of the Society for Experimental Mechanics Series 2025:1–6. 10.1007/978-3-031-68142-4_1.
[31] K. Worden, C.X. Wong, U. Parlitz, A. Hornstein, D. Engster, T. Tjahjowidodo, F. Al-Bender, D.D. Rizos, S.D. Fassois, Identification of pre-sliding and sliding friction dynamics: Grey box and black-box models, Mech. Syst. Signal Process. 21 (2007) 514–534, https://doi.org/10.1016/J.YMSSP.2005.09.004.
[32] S. Wang, Y. Teng, P. Perdikaris, Understanding and mitigating gradient flow pathologies in physics-informed neural networks, SIAM J. Sci. Comput. 43 (2021) 3055–3081, https://doi.org/10.1137/20M1318043.
[33] J. Willard, X. Jia, M. Steinbach, V. Kumar, S. Xu, Integrating Physics-Based Modeling With Machine Learning: A Survey 2020;1:34. 10.1145/1122445.1122456.
[34] M. Haywood-Alexander, W. Liu, K. Bacsa, Z. Lai, E. Chatzi, Discussing the spectrum of physics-enhanced machine learning: a survey on structural mechanics applications, Data-Centric Eng. 5 (2024) e31.
[35] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707, https://doi.org/10.1016/J.JCP.2018.10.045.
[36] A. Fuller, Z. Fan, C. Day, C. Barlow, Digital twin: enabling technologies, challenges and open research, IEEE Access 8 (2020) 108952–108971, https://doi.org/10.1109/ACCESS.2020.2998358.
[37] E.J. Cross, Gibson SJ, Jones MR, Pitchforth DJ, Zhang S, Rogers TJ. Physics-Informed Machine Learning for Structural Health Monitoring. Structural Integrity, vol. 21, Springer Science and Business Media Deutschland GmbH; 2022, p. 347–67. 10.1007/978-3-030-81716-9_17.
[38] M. Dalmasso, M. Civera, V. De Biagi, B. Chiaia, Gaussian Process Regression (GPR)-based missing data imputation and its uses for bridge structural health monitoring, Adv. Bridge Eng. 6 (2025) 1–21, https://doi.org/10.1186/S43251-025-00169-1.
[39] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.
[40] M.A. Álvarez, L. Rosasco, N.D. Lawrence, Kernels for vector-valued functions: a review, Foundations and Trends® in Machine Learning 4 (2012) 195–266, https://doi.org/10.1561/2200000036.
[41] S. Zhang, T.J. Rogers, E.J. Cross, Gaussian process based grey-box modelling for shm of structures under fluctuating environmental conditions, Lecture Notes in Civil Engineering vol. 128 (2021) 55–66, 10.1007/978-3-030-64908-1_6.
[42] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, The MIT Press, 2016.
[43] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer, New York, NY, 2009, 10.1007/978-0-387-84858-7.
[44] E. Sousa Tomé, M. Pimentel, J. Figueiras, Damage detection under environmental and operational effects using cointegration analysis – application to experimental data from a cable-stayed bridge, Mech. Syst. Signal Process. 135 (2020) 106386, https://doi.org/10.1016/j.ymssp.2019.106386.
[45] B. Behkamal, A. Entezami, C. De Michele, A.N. Arslan, Elimination of thermal effects from limited structural displacements based on remote sensing by machine learning techniques, Remote Sens (Basel) 15 (2023), https://doi.org/10.3390/rs15123095.
[46] A.V. Joshi, Support Vector Machines, Mach. Learn. Artif. Intelligence (2023) 89–99, https://doi.org/10.1007/978-3-031-12282-8_8.
[47] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (2004) 199–222, https://doi.org/10.1023/B:STCO.0000035301.49549.88.
[48] X. Yang, Y. Zhou, X. Li, J. Zhang, Effect of temperature changes on bearing motion of long-span steel truss continuous girder bridge, Constr. Build. Mater. 430 (2024) 136440, https://doi.org/10.1016/J.CONBUILDMAT.2024.136440.
[49] J. Alexander, M. Yarnold, Quasi-static bearing evaluation and monitoring—a case study, Front. Built Environ. 6 (2020), https://doi.org/10.3389/fbuil.2020.00069.
[50] Y. Xi, Y. Zhao, R. Larsson, J. Mao, Time-dependent hysteresis friction behaviors of linear rolling bearings, Int. J. Adv. Manuf. Technol. 94 (2018) 3109–3116, https://doi.org/10.1007/S00170-016-9425-3.
[51] fitrgp - Fit a Gaussian process regression (GPR) model - MATLAB n.d. https://www.mathworks.com/help/stats/fitrgp.html (accessed July 24, 2025).
[52] D.-H. Yang, T.-H. Yi, H.-N. Li, Y.-F. Zhang, Correlation-based estimation method for cable-stayed bridge girder deflection variability under thermal action, J. Perform. Constr. Facil (2018) 32, https://doi.org/10.1061/(asce)cf.1943-5509.0001212.
[53] Q. Xia, Y. Xia, H.P. Wan, J. Zhang, W.X. Ren, Condition analysis of expansion joints of a long-span suspension bridge through metamodel-based model updating considering thermal effect, Struct. Control Health Monit. (2020) 27, https://doi.org/10.1002/stc.2521.
[54] M.E. Warren, N.C. Dubbs, Periodic temperature-based bearing assessment of a steel multigirder bridge, Transp. Res. Rec. 2642 (2017) 26–34, https://doi.org/10.3141/2642-04.

[55] E.J. Cross, K. Worden, K.Y. Koo, J.M.W. Brownjohn, Filtering environmental load effects to enhance novelty detection on cable-supported bridge performance. Bridge Maintenance, Safety, Management, Resilience and Sustainability - Proceedings of the Sixth International Conference on Bridge Maintenance, Safety and Management, Taylor and Francis - Balkema; 2012, p. 745–52. 10.1201/b12352-101.
[56] E. Parzen, On Estimation of a Probability Density Function and Mode. 1962;33:1065–76. 10.1214/AOMS/1177704472.
[57] B.W. Silverman, Density Estimation: For Statistics and Data Analysis, CRC Press, 1998, 10.1201/9781315140919.
[58] G. Marasco, F. Oldani, B. Chiaia, G. Ventura, F. Dominici, C. Rossi, F. Iacobini, A. Vecchi, Machine learning approach to the safety assessment of a prestressed concrete railway bridge, Struct. Infrastruct. Eng. 20 (2024) 566–580, https://doi.org/10.1080/15732479.2022.2119581.
[59] K. Worden, Baldacchino T, Rowson J, Cross EJ. Some recent developments in SHM based on nonstationary time series analysis n.d. 10.1109/JPROC.2016.2573596.