

A Hybrid Neural Model Approach for Health Assessment of Railway Transition Zones with Multiple Data Sources

Phusakulkajorn, Wassamon; Unsiwilai, Siwarak; Chang, Ling; Li, Zili; Nunez, Alfredo

DOI

[10.1109/TIM.2025.3575986](https://doi.org/10.1109/TIM.2025.3575986)

Publication date

2025

Document Version

Final published version

Published in

IEEE Transactions on Instrumentation and Measurement

Citation (APA)

Phusakulkajorn, W., Unsiwilai, S., Chang, L., Li, Z., & Nunez, A. (2025). A Hybrid Neural Model Approach for Health Assessment of Railway Transition Zones with Multiple Data Sources. *IEEE Transactions on Instrumentation and Measurement*, 74, Article 3544017. <https://doi.org/10.1109/TIM.2025.3575986>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

A Hybrid Neural Model Approach for Health Assessment of Railway Transition Zones With Multiple Data Sources

Wassamon Phusakulkajorn¹, Siwarak Unsiwilai², *Member, IEEE*, Ling Chang³, Zili Li⁴,
and Alfredo Núñez⁵, *Senior Member, IEEE*

Abstract—Transition zones in railway tracks often degrade faster than other locations, yet traditional health assessments rely on infrequent track geometry measurements, limiting early detection of dynamic changes. This research presents an approach for more frequent evaluation of transition zone health by integrating data sources from multiple monitoring technologies: track geometry, interferometric synthetic aperture radar (InSAR), and axle box acceleration (ABA). Missing InSAR data are addressed through spatiotemporal interpolation, and track longitudinal levels are predicted using a hybrid neural model that includes a hybrid convolutional neural network (CNN) with gated recurrent unit (GRU) network and a hybrid CNN with a long short-term memory (LSTM) network. The models fuse historical and interpolated data from InSAR and ABA, enabling high-frequency insights. A novel key performance index (KPI) based on predicted longitudinal levels is proposed to quantify track condition. The framework is validated on a transition zone at a railway bridge between Dordrecht and Lage Zwaluwe in The Netherlands. The results show that the hybrid model outperforms standalone methods and offers a good balance between accuracy and computational efficiency. The proposed approach enables earlier detection of irregularities, supporting prescriptive maintenance decisions.

Index Terms—Acceleration measurements, data fusion, hybrid neural networks, interferometric synthetic aperture radar (InSAR), railway infrastructure, track geometry.

NOMENCLATURE

ABA	Axle box acceleration.
AL	Alert limit.
CNN	Convolutional neural network.
CoSD	Combined standard deviation.

Received 20 February 2025; revised 2 May 2025; accepted 22 May 2025. Date of publication 2 June 2025; date of current version 23 June 2025. This work was supported in part by the ProRail and Europe's Rail Flagship Project IAM4RAIL—Holistic and Integrated Asset Management for Europe's RAIL System of European Union and in part by European Union's Horizon Europe Research and Innovation Program under Grant 101101966. The work of Wassamon Phusakulkajorn and Siwarak Unsiwilai was supported by the Royal Thai Government. The Associate Editor coordinating the review process was Dr. Siliang Lu. (*Corresponding author: Alfredo Núñez.*)

Wassamon Phusakulkajorn, Siwarak Unsiwilai, Zili Li, and Alfredo Núñez are with the Section of Railway Engineering, Delft University of Technology, 2628 CN Delft, The Netherlands (e-mail: W.Phusakulkajorn@tudelft.nl; S.Unsiwilai@tudelft.nl; Z.Li@tudelft.nl; A.A.Nunezvicencio@tudelft.nl).

Ling Chang is with the Faculty of Geo-Information Science and Earth Observation, University of Twente, 7522 NB Enschede, The Netherlands (e-mail: ling.chang@utwente.nl).

Digital Object Identifier 10.1109/TIM.2025.3575986

DTW	Dynamic time warping.
GRU	Gated recurrent unit.
IAL	Immediate action limit.
IL	Intervention limit.
InSAR	Interferometric synthetic aperture radar.
LSTM	Long short-term memory.
MAE	Mean absolute error.
RMSE	Root-mean-square error.
RNN	Recurrent neural network.
SAR	Synthetic aperture radar.
SAWP	Scale average wavelet power.
SSE	Sum squared error.
TQI	Track quality index.
TSS	Total sum of squared error.
WPS	Wavelet power spectrum.

I. INTRODUCTION

IN RAILWAYS, transition zones denote areas marked by structural discontinuities, leading to abrupt variations in track stiffness, damping, and track geometry. They are often observed at critical points such as bridges, culverts, tunnels, transitions between ballasted and slab tracks, and road crossings. The abrupt changes in train dynamic responses significantly accelerate the degradation of transition zones, resulting in substantial maintenance costs to sustain smooth operation and safety. Hence, it is crucial to assess the structural health condition of transition zones more frequently to detect changes at an early stage. Correct health assessment allows prescriptive interventions to maintain operational efficiency and mitigate potential risks.

The health assessment of a railway transition zone includes various aspects. Examples are track geometry inspection to measure the track deviations from design specifications, condition monitoring of rail and other track components in areas of transition to examine wear, defects, and damage, dynamic response analysis to identify any irregularities or excessive forces that could affect safety and health, and embankment stability inspection where the transition zone is situated to prevent derailment, among others. Information on the health of transition zones can be collected from three primary sources: trackside measurements, onboard systems, and remote sensing technologies. Each of these has specific limitations that can be addressed by combining insights from the other sources.

Trackside measurements provide information at specific positions using technologies such as borehole inclinometers, geophones, linear variable differential transformers, deflectometers, strain gauges, accelerometers, and advanced fiber-optic methods such as Rayleigh backscattering [1] and fiber Bragg grating [2]. Measurement techniques can be applied individually to meet specific objectives or combined to improve monitoring effectiveness by integrating data from different types of sensors [3], [4]. While these methods allow detailed monitoring of specific points, they are typically installed only in high-priority locations, offering the local coverage of responses [5], [6], [7]. The need for a broader perspective of track responses due to train loads necessitates the installation of a greater number of sensors. This involves costs from not only devices but also expenses related to labor and power supplies. Thus, trackside measurements are limited by their narrow spatial coverage and high resource demands.

Onboard monitoring systems collect data from moving vehicles. They offer continuous monitoring that improves trackside monitoring by covering extended track sections, including transition zones. For instance, onboard systems such as inertial measurement units, digital image correlation systems, and accelerometers are used to identify track irregularities and assess ballast conditions [8], [9], [10], [11], [12], [13], [14]. However, some challenges that onboard monitoring techniques face include their integration with existing systems, the need for efficient data processing and analysis due to vast amounts of generated data, and their adaptations to make them robust and implementable in, for instance, passenger trains.

Remote sensing technologies, including satellite-based measurements, provide a broader view of railway infrastructure and transition zones, detecting ground movements such as subsidence, uplift, and landslides with subcentimeter precision [15], [16]. Satellite data, such as that from the Sentinel-1 radar mission, deliver radar images on a biweekly basis, with (tens of) meter-level spatial resolution. They can be exploited to detect railway irregularities [17], [18] and characterize the deformation process resulting from subway-induced subsidence in the construction and operation periods [19]. Despite its ability to provide valuable data, some railway applications require higher spatial resolution and accuracy that are not currently possible to obtain from medium-resolution satellite data.

Assessing transition zone conditions relies significantly on measurement frequency and data density across the area. Frequent measurements are crucial for improving forecasts, gaining insights into long-term deterioration, and identifying severe events. This, in turn, supports more effective maintenance strategies. At the same time, high-density data collection along transition zones is essential for detecting track irregularities. While individual measurement approaches provide valuable insights, each has its strengths and limitations. To leverage their complementary advantages, this work considers integrating multiple data sources to predict track conditions at transition zones.

Various machine learning models, including principal component analysis [20], support vector regression [21], artificial neural networks [21], [22], [23], deep neural networks [24],

and convolutional encoder–decoder networks [25], have been employed to predict railway track conditions. The development of these models reflects variations in input data sources, prediction time horizons, and their applicability to specific railway section types. Studies such as [20], [21], [22], [24] and [23] rely on historical track geometry and maintenance data for degradation prediction. Among them, the works in [22], [24] and [25] adopt complex neural network architectures and achieve high predictive performance, with reported R^2 values ranging from 0.85 to over 0.95. However, such performance depends on the experiment design and local conditions. While models in [20], [21], [23], and [25] are limited to short-term predictions, those in [22] and [24] demonstrate capabilities for long-term forecasting. Although most studies [20], [21], [22], [23], [24] depend on historical track geometry measurements, the study in [25] explores an alternative by using dynamic inputs such as bogie, wheelset, and car-body accelerations, along with vehicle speed. This method infers track irregularities through acceleration-based inversion, reducing dependence on track geometry data but does not provide long-term prediction.

Transition zone condition assessment presents additional challenges due to the presence of both spatial and temporal dependencies. While CNNs are effective at capturing spatial patterns, as demonstrated in [25], RNN and its variants have shown success in learning temporal relationships [23]. Nevertheless, these models typically treat spatial and temporal aspects separately. To overcome this, hybrid models combining CNNs with RNN variants have been proposed, such as the CNN-LSTM approach in [26], which enables the joint learning of spatial–temporal dependencies. LSTM, though effective in capturing long-term dependencies due to their gating mechanisms and memory cells, is computationally intensive [23], [26]. Alternatively, GRUs, which use fewer parameters, offer a more efficient solution while still addressing the vanishing gradient problem [23]. Therefore, this work considers CNN-LSTM and CNN-GRU models to capture the spatiotemporal dependencies in track geometry changes.

To address the challenges of long-term prediction and reduce dependence on historical track geometry data, this article presents a data fusion framework that integrates track geometry, InSAR, and ABA measurements to assess railway transition zone conditions. Specifically, track geometry measurements are standardized and valued for accuracy and resolution. However, track geometry measurements may miss areas with poor dynamic train–track interactions. ABA measurements, in contrast, capture local dynamic responses, while InSAR offers broader spatial coverage and long-term deformation trends. By leveraging these complementary data sources, the proposed framework enables more frequent evaluations, addressing the limitation of current practices where track geometry is typically measured a few times per year. Our key contributions include the following.

- 1) A hybrid neural model approach to assess the health condition of transition zones that integrates three distinct types of signals for railway infrastructure geometry and dynamics: track geometry, InSAR, and ABA measurements. The developed key performance index (KPI)

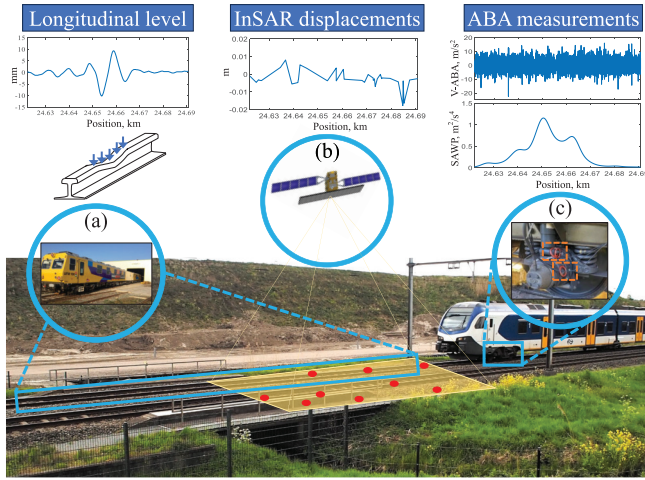


Fig. 1. (a) Track geometry measurement system, (b) InSAR, and (c) ABA system. SAWP of V-ABA is considered. Red points represent InSAR data points.

enables more frequent evaluations of transition zone health conditions and provides railway infrastructure managers with more information to plan and execute maintenance activities more efficiently.

- 2) Real-world field measurements at transition zones of a railway bridge in The Netherlands are used to demonstrate the applicability and performance of the framework.

II. MEASUREMENT TECHNOLOGY

Fig. 1 illustrates an overview of three measurement technologies used in this work. The technologies include a track geometry measurement system, InSAR, and ABA measurement system.

A. Track Geometry Measurement

Track geometry measurement is a standardized measurement used in various railway companies. It is often performed using measurement cars equipped with sensors and instruments. These cars move along the track and collect data on various parameters. Typically, the track geometry measurement includes the longitudinal level of both rails, alignment of both rails, gauge, and twist. The spatial resolution of track geometry measurements can range from centimeters to decimetres. The frequency of track geometry measurements depends on several factors, including budget and the availability of measurement cars. The level of train traffic and maintenance activities also affect the frequency of track geometry measurements due to track possession. Generally, track geometry measurements are performed regularly. In The Netherlands, for example, the measurement campaign is conducted once or twice per year. This results in less frequent measurements with a high density of track geometry data.

B. InSAR Radar

InSAR is a remote sensing technique that uses SAR to measure and monitor surface changes of the Earth over time.

SAR is an active sensing technique that emits electromagnetic pulses (in the microwave range) toward the surface of the Earth and records the backscattered radar signal. The radar pulses are emitted along the satellite line-of-sight direction. Radar signals are composed of two main components: amplitude and phase. To derive the ground displacement that occurred between the two acquisitions, InSAR involves comparing the phase difference of radar signals acquired over the same area at different times, particularly for repeat-pass SAR satellites. A stack of radar signals over different times is used to get a time series of ground target movement. InSAR data, namely InSAR displacements for our study, are collected every 6–12 days, and the spatial density of InSAR data varies depending on the studied area and processing strategy, which can be in the range of meters. This results in frequent measurements with less density of InSAR data over an area. However, some areas may be monitored by multiple SAR satellites, each providing independent observations at different time intervals and viewing angles. Combining data from multiple SAR sources, therefore, offers a more comprehensive assessment of changes in the study area.

C. ABA Measurement System

ABA measurement system is an onboard measurement technique. Its fundamental concept is to use a train as a moving load to excite the infrastructure and to detect defects and irregularities through an analysis of time–frequency characteristics. This analysis is derived from the dynamic response of the train–track interaction captured by accelerometers installed on the axle boxes. Accelerometers can be installed on the axle boxes in three directions: vertical, longitudinal, and lateral. These produce ABA signals that capture dynamic responses in each corresponding direction. Fig. 1(c) shows an example of signals obtained from the vertical ABA (V-ABA). To analyze frequency characteristics associated with specific track irregularities, we apply a continuous wavelet transform to convert the time-domain ABA signals into the time–frequency domain. SAWP is then used to extract features within the spatial frequency range of $0.04\text{--}0.33\text{ m}^{-1}$, corresponding to track longitudinal irregularity wavelength between 3 and 25 m. This wavelength range has been reported to indicate changes in track substructure conditions [12].

In comparison to other measurement techniques that use dedicated measurement vehicles, ABA technology comes with the advantages of lower cost and easier maintenance. The ABA measurement system achieves high spatial resolution, typically within the millimeter range due to a high sampling frequency. While its technology readiness level is increasing [27], the ABA system still needs to evaluate its robustness and generalization. This involves conducting extensive measurement and validation campaigns in different locations and under different measurement conditions. Once the technology advances and matures, more frequent ABA measurements are anticipated. This involves installing ABA systems on existing passenger trains or considering these systems already in the design of new generation trains with embedded smart sensors, offering real-time or near-real-time monitoring of track conditions. However, in this case study, ABA measurements are obtained

yearly. This results in less frequent measurements with a higher density than track geometry data.

III. PROBLEM FORMULATION

This work assumes the availability of datasets $\mathcal{D}_j^r = \{\mathcal{G}_j^r, \mathcal{S}_j^r, \mathcal{A}_j^r\}$ containing three different measurement technologies for a given railway track $j \in \mathbb{N}$ and rail $r \in \{\text{inner, outer}\}$. The measurements include: first, a set of track geometry measurements, $\mathcal{G}_j^r$; second, a set of displacements derived from InSAR data $\mathcal{S}_j^r$; and finally, a set of ABA measurements $\mathcal{A}_j^r$. The measurements are collected from a transition zone at a position x within the kilometer range from x_b to x_e and during the month t within time frame t_b and t_e . They are mathematically expressed as follows:

$$\begin{aligned}\mathcal{G}_j^r &= \{l_j^r(x, t) | x = x_1^G, \dots, x_{\gamma_1}^G \text{ and } t = t_1^G, \dots, t_{\beta_1}^G\} \\ \mathcal{S}_j^r &= \{d_j^r(x, t) | x = x_1^S, \dots, x_{\gamma_2}^S \text{ and } t = t_1^S, \dots, t_{\beta_2}^S\} \\ \mathcal{A}_j^r &= \{a_j^r(x, t) | x = x_1^A, \dots, x_{\gamma_3}^A \text{ and } t = t_1^A, \dots, t_{\beta_3}^A\}\end{aligned}$$

where $l_j^r(x, t)$, $d_j^r(x, t)$, and $a_j^r(x, t)$ denote, respectively, the track longitudinal level, the displacement, and the ABA data collected at a position x and in month t . Variables γ_1 – γ_3 represent the number of measurements in the spatial domain corresponding to track geometry, displacements, and ABA signals, respectively. Similarly, β_1 – β_3 represent the number of measurements in the temporal domain for each respective dataset. For all the measurements, missing data can happen in which $l_j^r(x, t)$, $d_j^r(x, t)$, and $a_j^r(x, t)$ are represented by the not-a-number value.

The average spatial resolutions of the track geometry, ABA measurements, and InSAR displacements are defined as $\Delta x^G = 1/(\gamma_1 - 1)(x_{\gamma_1}^G - x_1^G)$, $\Delta x^S = 1/(\gamma_2 - 1)(x_{\gamma_2}^S - x_1^S)$, and $\Delta x^A = 1/(\gamma_3 - 1)(x_{\gamma_3}^A - x_1^A)$, respectively. Similarly, the average temporal resolutions of the track geometry, ABA measurements, and InSAR displacements are defined as $\Delta t^G = 1/(\beta_1 - 1)(t_{\beta_1}^G - t_1^G)$, $\Delta t^S = 1/(\beta_2 - 1)(t_{\beta_2}^S - t_1^S)$, and $\Delta t^A = 1/(\beta_3 - 1)(t_{\beta_3}^A - t_1^A)$, respectively. In this work, we assume that Δt^G and Δt^A are similar, but both are much less than Δt^S . Conversely, Δx^A is significantly higher compared to Δx^G and Δx^S , with Δx^G being denser than Δx^S . Fig. 2 shows which measurements were obtained across track positions and time. Track geometry measurements are in blue solid lines, InSAR measurements obtained from Sentinel-1 satellites with four different orbital trajectories represented by dotted lines in other colors, and the ABA measurements with red solid lines. The spatial and temporal resolutions of the different data sources can be estimated from such a diagram.

For track j and rail r , let $\mathcal{H}_j^r$ denote a 2-D spatiotemporal function that provides the unobserved value of ABA data $\hat{a}_j^r(x, t^A)$ at position $x \in [x_b, x_e]$ and in the ABA measurement month $t^A = t_1^A, \dots, t_{\beta_3}^A$ such that

$$\hat{a}_j^r(x, t^A) = \mathcal{H}_j^r(x, t^A). \quad (1)$$

Likewise, let $\mathcal{K}_j^r$ denote a 2-D spatiotemporal function that provides the unobserved value of the displacements $\hat{d}_j^r(x, t)$ at position $x \in [x_b, x_e]$ and in month $t \in [t_b, t_e]$ such that

$$\hat{d}_j^r(x, t) = \mathcal{K}_j^r(x, t). \quad (2)$$

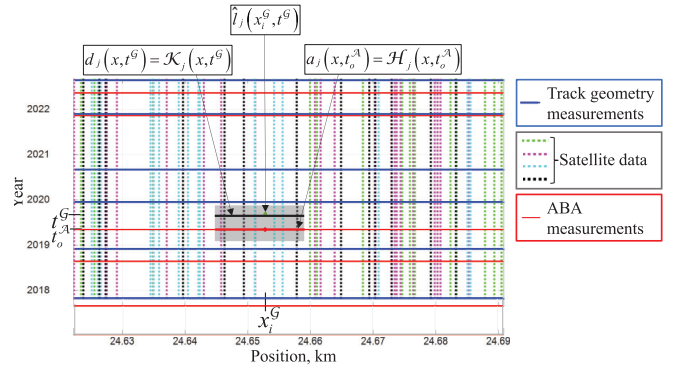


Fig. 2. Availability of the data considered in this work across different positions and times, obtained from track geometry measurements (blue solid lines), InSAR measurements from four different orbital trajectories (dotted lines in four distinct colors), and ABA measurements (red solid lines). White spaces represent missing information, while the gray area defines the neighboring measurements used to construct the interpolation functions.

To predict the track longitudinal level $\hat{l}_j^r(x^G, t^G)$ at the observed position $x^G = x_1^G, \dots, x_{\gamma_1}^G$ and in the unobserved month t^G , $t^G \neq t_1^G, \dots, t_{\beta_1}^G$, we exploit the interpolated values of InSAR and ABA data to construct a hybrid neural model $\mathcal{M}_j$ defined for all rails r such that

$$\begin{aligned}\hat{l}_j^r(x^G, t^G) &= \mathcal{M}_j(\mathcal{H}_j^r(x^G, t_o^A), \mathcal{K}_j^r(x^G, t^G)) \\ &= \mathcal{M}_j(\hat{a}_j^r(x^G, t_o^A), \hat{d}_j^r(x^G, t^G))\end{aligned} \quad (3)$$

where $\mathcal{H}_j^r$ is derived from the closest available ABA measurement month t_o^A , $o \in \{1, \dots, \beta_3\}$, occurring prior to the measurement month t^G of the predicted track geometry. $\mathcal{K}_j^r$ is defined at the position x^G and the same month t^G of the predicted track longitudinal level. Then, a KPI is evaluated across unobserved months spanning the entire transition zone utilizing the predicted track longitudinal level $\hat{l}_j^r(x^G, t^G)$.

This article proposes a methodology that involves the identification of (1)–(3) for an estimation of a KPI that aims at facilitating a more frequent assessment of transition zone health. The proposed methodology offers the capability to evaluate transition zone health prior to the next measurement of track geometry. With the new available InSAR data that comes frequently, the KPI is updated to provide early warnings to support decision-making regarding the time and locations for maintenance interventions.

IV. METHODOLOGY

Fig. 3 illustrates the framework of this article. It involves the development of two distinct 2-D spatiotemporal interpolation functions to estimate values from InSAR and ABA data. Then, a hybrid neural model is developed to predict the track longitudinal levels using these interpolated values. Finally, utilizing the predicted track longitudinal levels, the KPI is evaluated across unobserved months spanning the entire transition zone.

A. Spatiotemporal Interpolation

The railway track is a distributed system characterized by dynamic variations over time and locations. Temporal variation is due to continuous usage, degradation, and maintenance,

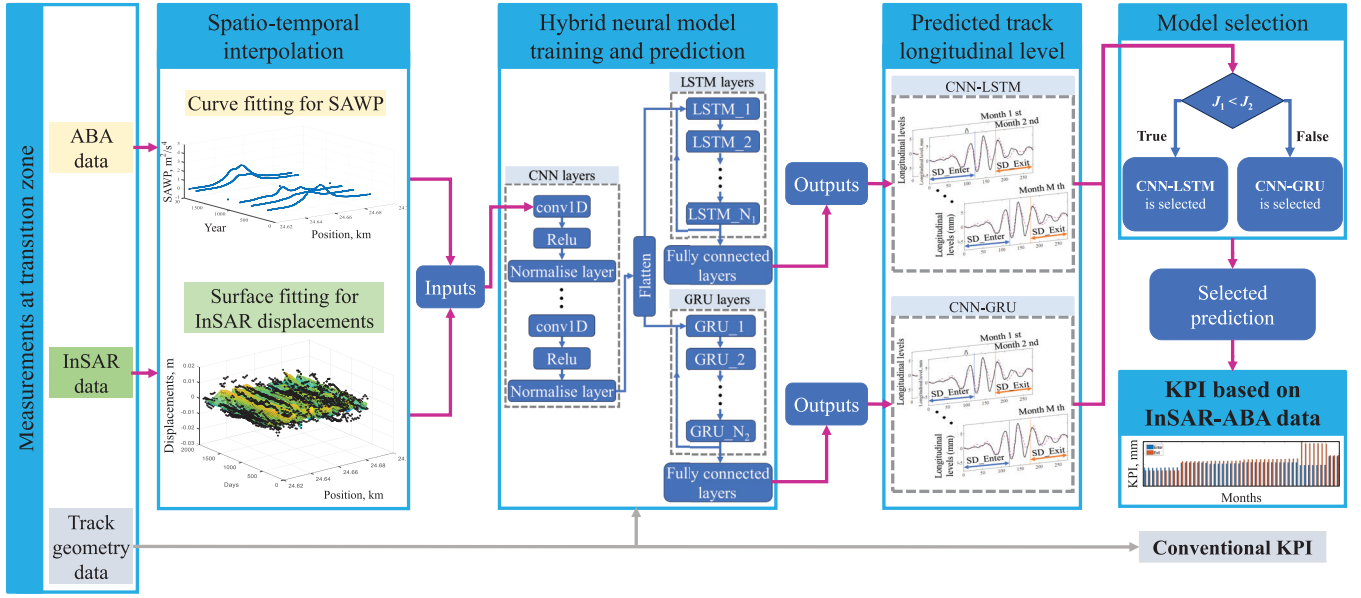


Fig. 3. Proposed framework.

while spatial variations arise from distinct dynamics at different locations, including transition zones. As performing an interpolation within a smaller region allows for addressing local variations over different locations and times, this work considers the development of spatiotemporal interpolation within smaller regions. The workflow of the spatiotemporal interpolation involves domain discretization, surface fitting for the displacements, and curve fitting for the ABA data within the subdomains.

1) *Domain Discretization*: This work predicts track longitudinal levels by utilizing interpolated values derived from InSAR and ABA measurements. Subsequently, the domain is discretized to ensure that each subdomain contains displacements from at least two distinct locations and ABA values from at least one measurement. Based on these assumptions, a distance Δx can be defined for a spatial domain discretization such that $\Delta x \geq \max\{(x_i^S - x_{i-1}^S), i = 2, \dots, \gamma_2\}$. Likewise, Δt for a temporal domain discretization can be defined such that $\Delta t \geq \max\{(t_i^A - t_{i-1}^A), i = 2, \dots, \beta_3\}$.

Given Δx and Δt and a spatiotemporal domain $\mathcal{T} = \{(x, t) | x \in [x_b, x_e] \text{ and } t \in [t_b, t_e]\}$, we have that the spatial domain is uniformly discretized into M subspatial domains and the temporal domain is uniformly discretized into N subtemporal domains. This results in a total of $M \times N$ subdomains. Mathematically, a subdomain $T_{m,n}$ is defined as

$$T_{m,n} = \{(x, t) | x \in [x_b + (m-1) \cdot \Delta x, x_b + m \cdot \Delta x], t \in [t_b + (n-1) \cdot \Delta t, t_b + n \cdot \Delta t]\} \quad (4)$$

where $m = 1, \dots, M$ and $n = 1, \dots, N$.

The shaded region depicted in Fig. 2 provides an example of a subdomain where the interpolation functions are formulated based on the available InSAR and ABA measurements within that specific subdomain. Within this shaded region, ABA data from the month t_o^A and InSAR data collected from positions and months confined by the spatial and temporal boundaries of

the shaded area will be utilized to construct the interpolation functions.

2) *Surface Fitting for Displacements*: Given a subdomain $T_{m,n}$ and the displacements $d_j^{r,m,n}(x^S, t^S)$ collected from rail r of track j at position $x^S \in [x_b + (m-1) \cdot \Delta x, x_b + m \cdot \Delta x]$ and in month $t^S \in [t_b + (n-1) \cdot \Delta t, t_b + n \cdot \Delta t]$ that are available within $T_{m,n}$, this article considers 2-D polynomial interpolation to estimate values of displacements, $\hat{d}_j^{r,m,n}(x^G, t^G)$ at the position, x^G , and in the month, t^G , of the predicted track longitudinal level within a subdomain $T_{m,n}$. The 2-D polynomial interpolation $\mathcal{K}_j^{r,m,n}$ defined for rail r of track j within a subdomain $T_{m,n}$ is mathematically expressed as

$$\mathcal{K}_j^{r,m,n}(x, t) = \sum_{p=0}^{P_j^{r,m,n}} \sum_{q=0}^{Q_j^{r,m,n}} c_{j,pq}^{r,m,n} x^p t^q \quad (5)$$

where $P_j^{r,m,n}$ and $Q_j^{r,m,n}$ denote the degree of the polynomial used for interpolation in the spatial and temporal subdomain, respectively. The coefficients $c_{j,pq}^{r,m,n}$, $p = 0, \dots, P_j^{r,m,n}$ and $q = 0, \dots, Q_j^{r,m,n}$ are determined using the given data points in each corresponding subdomain, aiming at minimizing the sum of the squared differences between the observed and predicted values. The degree of the polynomial defined in each subdomain is optimized to attain the lowest sum square error.

3) *Curve Fitting for the ABA Signal*: Given a subdomain $T_{m,n}$ and the ABA data $d_j^{r,m,n}(x^A, t^A)$ obtained from rail r of track j at position $x^A \in [x_b + (m-1) \cdot \Delta x, x_b + m \cdot \Delta x]$ and in month $t^A \in [t_b + (n-1) \cdot \Delta t, t_b + n \cdot \Delta t]$ that are available within $T_{m,n}$, even though ABA data exhibit temporal variation, this article assumes a constant variation of the ABA values between two consecutive measurement months. This allows the formulation of 2-D spatiotemporal interpolation to be simplified into 1-D interpolation. To estimate the values of the ABA data $\hat{d}_j^{r,m,n}(x^G, t_o^A)$ at the position x^G from the closest ABA measurement month, t_o^A , $o \in \{1, \dots, \beta_3\}$, occurring prior

to the measurement month t^G , this article considers 1-D polynomial interpolation $\mathcal{H}_{j,t^G}^{r,m,n}(x)$ defined for rail r of track j within a subdomain $T_{m,n}$, which is mathematically expressed as

$$\mathcal{H}_{j,t^G}^{r,m,n}(x) = \sum_{\xi=0}^{R_j^{r,m,n}} b_{j,\xi}^{r,m,n} x^\xi \quad (6)$$

where $R_j^{r,m,n}$ denotes the degree of the polynomial used for interpolation in the spatial subdomain $T_{m,n}$. The coefficients $b_{j,\xi}^{r,m,n}, \xi = 0, \dots, R_j^{r,m,n}$, are determined using the given data points in each corresponding subdomain, aiming at minimizing the sum of the squared differences between the observed and predicted values of the ABA signals. The degree of the polynomial defined in each subdomain is optimized to attain the lowest sum square error.

B. Hybrid Neural Model Approach

Each transition zone is affected by different structural, environmental, and operational factors, leading to unique temporal-spatial signatures in both track geometry and dynamic responses. These variations can limit the generalization capability of a single-model architecture.

To address this, we propose a hybrid approach that uses CNN-LSTM and CNN-GRU models with a selection mechanism. Both models are designed to have CNN layers on top. Then, the LSTM or GRU structure is designed to follow the CNN layers, as seen in Fig. 3. The inputs to the CNN layers are the interpolated values of the displacements derived from the InSAR data and the ABA measurements. After the CNN layers, their output is fed into the LSTM and GRU layers and the fully connected layers.

Both CNN-LSTM and CNN-GRU models are effective in capturing spatial and temporal dependencies. The LSTM is particularly effective at capturing long-term dependencies, while the GRU, with its simpler architecture and fewer parameters, tends to offer better training efficiency and reduced risk of overfitting, especially when working with limited data. Rather than assuming a universally applicable model, we empirically evaluate the suitability of both architectures for each transition zone. This strategy allows us to take advantage of their complementary strengths, thereby improving adaptability across diverse scenarios. Prior studies indicate that CNN-GRU architectures exhibit greater robustness compared to CNN-LSTM, attributed to their simpler structure and faster convergence. This work addresses robustness and generalization through cross validation and applying early stopping to limit overfitting.

Given railway track j , rail r , and the interpolated values of the displacements $\hat{d}_j^{r,m,n}(x^G, t^G)$ and the ABA signals $\hat{a}_j^{r,m,n}(x^G, t^A)$, this article considers the input vector to a hybrid neural model as a concatenated vector denoted as $X = [\hat{d}_j^{r,m,n}(x^G, t^G), \hat{a}_j^{r,m,n}(x^G, t^A)]$. For the f th convolutional layer, let denote its input as X_f , its filters as W_f , the bias term as b_f , and the activation function as ϕ . The output z_f of the f th convolutional layer can be expressed mathematically as follows:

$$z_f = \phi(W_f * X_f + b_f) \quad (7)$$

where $*$ denotes the convolution operation, which involves sliding the filters W_f over the input X_f and computing the dot product at each position. The result is then passed through the activation function ϕ . The obtained output z_f represents the feature maps passed on to the next layers.

After the CNN layers, LSTM layers or GRU layers are considered. For the GRU layers, the core building blocks include hidden state (h), update gate (v), reset gate (r), candidate hidden state ($\tilde{h}$), and hidden state update rule. For a time step t , the hidden state h_t is the memory of the GRU that captures information from previous time steps. The update gate v_t determines how much of the previous hidden state h_{t-1} to retain and how much of the new candidate hidden state $\tilde{h}_t$ to incorporate. The update gate is computed using a sigmoid activation function (σ) as seen in (8). The reset gate determines how much of the previous hidden state h_{t-1} to forget. It is also computed using a sigmoid activation function (σ) as seen in (9). The candidate hidden state at the time step t represents the new information that could be added to the memory. It is computed by applying the hyperbolic tangent ($\tanh$) activation function to the weighted sum of the reset-gated previous hidden state and the current input, as seen in (10). Then, the new hidden state h_t is computed by combining the previous hidden state h_{t-1} and the candidate hidden state $\tilde{h}_t$ using (11)

$$v_t = \sigma(W_v \cdot [h_{t-1}, z_t]) \quad (8)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, z_t]) \quad (9)$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \odot h_{t-1}, z_t]) \quad (10)$$

$$h_t = \tanh((1 - v_t) \odot h_{t-1} + v_t \odot \tilde{h}_t) \quad (11)$$

where z_t is an input to the GRU layers obtained from the last convolution layer of the CNN layers, W_v is the weight matrix for the update gate, W_h is the weight matrix for the candidate hidden state, and $\odot$ represents the element-wise multiplication.

Similar to the GRU layers, the LSTM cell has three gates: the input gate, the forget gate, and the output gate. The forget gate determines what information from the previous cell state should be discarded. The input gate decides what new information to store, and the output gate controls how much of the current cell state should be revealed as the output. Additionally, there are a cell state update, $\tilde{c}_t$, and the actual cell state c_t designed for LSTM. Mathematically, they are expressed as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, z_t] + b_f) \quad (12)$$

$$I_t = \sigma(W_I \cdot [h_{t-1}, z_t] + b_I) \quad (13)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, z_t] + b_c) \quad (14)$$

$$c_t = f_t \cdot c_{t-1} + I_t \cdot \tilde{c}_t \quad (15)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, z_t] + b_o) \quad (16)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (17)$$

where z_t is an input to the LSTM layers obtained from the last convolution layer of the CNN layers. f_t denotes the forget gate, I_t is the input gate, $\tilde{c}_t$ is the cell state, c_t is the update cell state, o_t is the output gate, and h_t is the hidden state update. The weight matrices are denoted as W_f, W_I, W_c , and W_o , and the bias vectors are represented by b_f, b_I, b_c , and b_o .

The final layer after the hybrid neural model consists of a fully connected layer. The output y_α^λ from the α th node of the λ th layer is mathematically expressed as

$$y_\alpha^\lambda = \sum_j \omega_{\alpha\epsilon}^\lambda (\phi(h_{t,\alpha}^\lambda) + b_\alpha^\lambda) \quad (18)$$

where $\omega_{\alpha\epsilon}^\lambda$ represents the weight of the α th and ϵ th node for the layer λ and $\lambda - 1$, respectively; ϕ is an activation function; and b_α^λ is its bias.

Then, a selection mechanism is implemented to choose the better-performing output from the CNN-LSTM or CNN-GRU. The selection mechanism is guided by an empirical strategy based on a tradeoff between two main criteria: predictive accuracy (accounting for spatial and temporal errors) and computational efficiency (considering time per epoch). While additional aspects may be considered in future extensions, this work focuses on these two factors as they most directly impact model performance and deployment feasibility. We formulate the model selection as follows:

$$i = \arg \min (J_1, J_2) \quad (19)$$

$$J_i = \alpha \cdot Y_{1i} + (1 - \alpha) \cdot Y_{2i}, \quad i \in \{1, 2\} \quad (20)$$

where index i represents the model we will select, with $i = 1$ representing CNN-LSTM and $i = 2$ representing CNN-GRU; J_i is the cost function related to model i , J is the minimum value between J_1 and J_2 ; Y_{1i} is the MAE of the predicted longitudinal level at a transition zone and represents performance of model i that considers spatial (about 30 m from each end of the bridge) and temporal behavior (about a year ahead prediction); Y_{2i} represents the complexity of the training via the average time per epoch for model i (with the same stopping criteria for both models); and $\alpha \in [0, 1]$ is a tuning parameter that controls the relative importance of predictive accuracy versus computational efficiency.

C. Evaluation of KPI

Let the proposed KPI based on InSAR and ABA data denote as $\text{KPI}_{\text{InSAR+ABA}}$. To assess the health of transition zones, this work focuses on computing $\text{KPI}_{\text{InSAR+ABA}}$ for two track segments $k \in \{\text{Entrance, Exit}\}$. These segments cover a distance of 30 m each, extending from the entrance and exit sides of the transition zones [12]. Specifically, the spatial boundaries are defined by x_b and $x_b + 30$ m for the entrance side, while for the exit side, the spatial boundaries are defined by $x_e - 30$ m and x_e . With a single value assigned to each segment k , two values of $\text{KPI}_{\text{InSAR+ABA}}^{k,j}$ are determined for a given track j . Their calculation involves determining the standard deviation of the predicted longitudinal levels, $\widehat{\text{SD}}_r$, at rail $r \in \{\text{inner, outer}\}$. The proposed $\text{KPI}_{\text{InSAR+ABA}}^{k,j}$ for a given segment k of track j is obtained as follows:

$$\text{KPI}_{\text{InSAR+ABA}}^{k,j} = \frac{(\widehat{\text{SD}}_{\text{inner}}^{k,j} + \widehat{\text{SD}}_{\text{outer}}^{k,j})}{2} \quad (21)$$

where

$$\widehat{\text{SD}}_r^{k,j} = \sqrt{\frac{\sum_{i=1}^{\gamma_1} (\hat{r}_j(x_i^G, t^G) - \bar{r}_j)^2}{\gamma_1}} \quad (22)$$

$\hat{r}_j(x_i^G, t^G)$ is the predicted track longitudinal levels at a position x_i^G contained within the spatial boundaries of segment k and in a month t^G , $\bar{r}_j$ is the mean of the predicted track longitudinal levels from rail r of track j , and γ_1 is the number of track geometry measurements in the spatial domain.

D. Evaluation Metrics

The CNN-LSTM and CNN-GRU models are trained to minimize the difference between the observed and the predicted track longitudinal level. In regression tasks, RMSE [defined in (24)] is typically used as a loss function for model training. Given two sequences of the observed track longitudinal levels $Y = [y_1, \dots, y_i, \dots, y_N]$ and the predicted track longitudinal levels $\hat{Y} = [\hat{y}_1, \dots, \hat{y}_i, \dots, \hat{y}_N]$ and N being the length of the sequences, the evaluation metrics for model performance in this article include the following.

- 1) MAE calculates the absolute differences between the actual and predicted values generated by the model and then averages these absolute differences. MAE gives equal weight to all errors without considering overestimation or underestimation. MAE is expressed as [28]

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N \{|y_i - \hat{y}_i|\}. \quad (23)$$

- 2) RMSE is the square root of the average of the squared differences between the actual and predicted values generated by the model. RMSE is expressed as [26], [28], [30]

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}. \quad (24)$$

- 3) R^2 measures the proportion of variance explained by the model and ranges from 0 to 1, with 1 indicating a perfect fit. However, a negative R^2 can occur when the SSE exceeds the TSS. This indicates that the model fails to capture the underlying patterns in the data, resulting in predictions that deviate substantially from the actual values. R^2 is expressed as [29]

$$\begin{aligned} R^2 &= 1 - \frac{\text{SSE}}{\text{TSS}} \\ &= 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \end{aligned} \quad (25)$$

where $\bar{y}$ denotes the mean of the observed values.

- 4) DTW distance is a metric derived from the DTW algorithm. It measures the similarity between two temporal sequences that may be misaligned. DTW accounts for variations in time between the predicted and actual sequences. The larger the DTW distance, the more dissimilarity and temporal misalignment. In this article, Euclidean distance is considered. Thus, the DTW distance between Y and $\hat{Y}$ is expressed as follows:

$$\text{DTW}(y_i, \hat{y}_i) = \min_w \sum_{k=1}^K \sqrt{(y_{i_k} - \hat{y}_{j_k})^2} \quad (26)$$

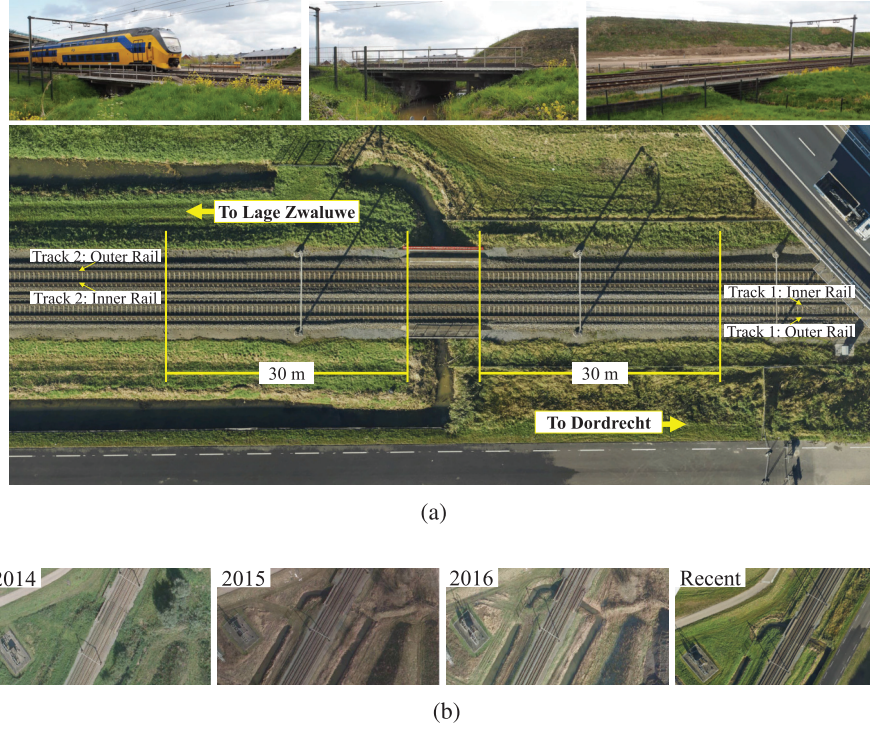


Fig. 4. Our study area of transition zones. (a) Photographs at the double-track railway bridge across a water channel from the side and top view. (b) Examples of aerial photographs showing the variability of the area near the railway track from 2014, 2015, 2016, and the most recent year (source: BBMS, ProRail).

where $w = [(i_1, j_1), \dots, (i_k, j_k), \dots, (i_K, j_K)]$ is the warping path with the length of K and k is the index along the warping path, from 1 to K .

V. CASE STUDY

This article selects transition zones at a railway bridge between Dordrecht and Lage Zwaluwe station as our case study. It is a 9-m single-span bridge that crosses over a water channel. The bridge supports two tracks of a railway line with fixed travel directions. Track 1 is for trains traveling from Lage Zwaluwe to Dordrecht, and Track 2 is for trains from Dordrecht to Lage Zwaluwe. This transition zone features the entrance and exit sides on each of the two tracks, as well as the inner and outer rails. The entrance side marks the beginning of the transition in which trains move from the conventional track to the bridge. The exit side of a railway transition zone is the segment where the trains return back to the standard configuration of tracks. Fig. 4(a) shows the photographs of the double-track railway bridge across a water channel from the side and top views. In addition to ground-level photographs, aerial photographs are used to provide a broader perspective on the surrounding environment and its evolution over time. They offer several advantages for railway infrastructure management, such as supporting asset inventory, detecting missing objects or obstructions, monitoring changes in land use, and documenting the infrastructure state before and after maintenance actions. Fig. 4(b) shows aerial photographs from different years illustrating changes surrounding the transition zones due to seasonal variations and human activities. These provide valuable contextual information, for example, nearby stable civil structures visible in the

aerial photographs are used to select reliable InSAR reference points, since these structures remain nondeformed over time compared to the railway track and transition zones. Moreover, broader environmental changes can indirectly influence transition zone conditions through mechanisms such as soil instability, surface runoff, or changes in drainage patterns. While these changes are not included in the approach proposed in this article, they could be embedded in future research. This article focuses on developing a KPI for frequently monitoring track geometry at transition zones, using ABA, InSAR, and track geometry measurements. Each measurement method has unique sensitivities and captures different aspects of track conditions. As a result, relying on a single data source can lead to gaps or misinterpretations when assessing transition zone conditions. Therefore, integrating multisource data is essential to overcome these limitations, providing a more comprehensive understanding of transition zone conditions.

A. Description of the Measurements

All the measurement technologies are analyzed at the track sections at both ends of the bridge, from 30 m before the railway bridge on the entrance side to 30 m after the bridge on the exit side and 30 m away from both tracks. Therefore, the boundary area of our case study is defined to include measurements collected at positions x within the kilometer range from $x_b = \text{edge1} - 0.03 \text{ km}$ to $x_e = \text{edge2} + 0.03 \text{ km}$. The transition zone of the bridge at the entrance side of Track 1, denoted as edge1 , is marked at 24.652 km. The transition zone of the bridge at the exit side of Track 1, denoted as edge2 , is

marked at 24.661 km. The distance from x_b to x_e accounts for a total distance of 69 m, including the bridge length of 9 m.

1) *Track Geometry Measurements*: This article considers only track longitudinal levels from track geometry measurements conducted from a railway track $j \in \{1, 2\}$ at rail $r \in \{\text{inner}, \text{outer}\}$. The measurements used are available yearly between November 2018 and August 2022. This accounts for five measurements in which t_1^G corresponds to November 2018, t_2^G corresponds to December 2019, t_3^G corresponds to August 2020, t_4^G corresponds to November 2021, and t_5^G corresponds to August 2022. The measurements are processed with a bandpass filter within a wavelength range of 3–25 m, and the measurements are reported in the spatial domain with a resolution $\Delta x^G = 0.25$ m [30]. This accounts for 276 datapoints within the boundary area. Hence, a set of track longitudinal levels collected from a rail r of track j at position x and month t can be defined as

$$G_j^r = \{l_j^r(x, t) | x = x_1^G, x_2^G, \dots, x_{276}^G \text{ and } t = t_1^G, \dots, t_5^G\}.$$

2) *InSAR Data*: We use InSAR data that SkyGeo processed, and data were obtained from four orbital trajectories: west-1a, west-2a, midden-1, and midden-2. In this article, the data were collected by Sentinel-1 SAR satellites, with multitrack images, every 6–12 days from September 2018 to August 2022. Due to the temporal variations in InSAR data collection, this article considers representing InSAR data on a monthly basis to ensure a uniform measurement time across different satellites. We also perform a moving average technique to mitigate noise inherent in InSAR data, which can be due to temporal changes on the ground. Hence, the measurement month t_1^S aligns with September 2018. With subsequent measurement months determined at a monthly resolution, culminating in the final measurement month t_{48}^S , which corresponds to August 2022. Within the boundary area, only 67 InSAR points were available. These InSAR points are spatially aligned by projecting their positions perpendicularly onto tracks 1 and 2. The projection yields InSAR data points spaced at an average interval of 1 m, i.e., $\Delta x^S = 1$ m, with a maximum discrepancy between individual data points equal to 5.4 m. For this article, we assume that the InSAR data for the inner rail are the same as those of the outer rail of the same track. Hence, a set of line-of-sight displacements collected from track j at position x and month t can be defined as

$$S_j = \{d_j(x, t) | x = x_1^S, x_2^S, \dots, x_{67}^S \text{ and } t = t_1^S, \dots, t_{48}^S\}.$$

3) *ABA Measurements*: The ABA signals are collected from four wheelsets at a rail $r \in \{\text{inner}, \text{outer}\}$ and from a track $j \in \{1, 2\}$. The measurements are available yearly between 2018 and 2019 and 2021 and 2022. This accounts for four measurements in which t_1^A corresponds to September 2018, t_2^A corresponds to June 2019, t_3^A corresponds to November 2021, and t_4^A corresponds to May 2022. In this article, the measurements come from the high sampling frequency of 25.6 kHz, e.g., local dynamic responses are measured at every millimeter approximately for a measuring speed of 100 km/h. Only the V-ABA signals corresponding to the vertical

train-track dynamic are considered in this work. A low-pass filter at 100 Hz is applied to filter out high-frequency contents within the V-ABA signals unrelated to transition zone conditions.

Then, the ABA signals in the time domain are transformed into the time-frequency domain using the Morlet wavelet. After the WPS of the ABA signal is obtained, their corresponding SAWP is calculated to investigate characteristic frequency responses at the transition zones. Only the spatial frequency range between 0.04 and 0.33 m^{-1} , corresponding to track irregularities in the wavelength from 3 to 25 m, is selected for our analysis as this is related to the condition of the substructure layer reported in [12]. This article considers the average SAWP from four ABA signals corresponding to the same rail to reduce uncertainty from the measurements, in which the SAWP is reported with a spatial resolution of 1 cm, i.e., $\Delta x^A = 0.01$ m. This accounts for 6900 data points within the boundaries of the spatial domain. Hence, a set of the average SAWP derived from the ABA measurements collected from rail r of track j at position x and month t can be defined as

$$A_j^r = \{a_j^r(x, t) | x = x_1^A, x_2^A, \dots, x_{6900}^A \text{ and } t = t_1^A, \dots, t_4^A\}.$$

B. Health Assessment of Railway Transition Zones

Maintaining railway tracks at transition zones depends on their condition, which is assessed through track geometry profiles. Two track geometry-based indicators can be considered for this task: an indicator for isolated defects and an indicator for overall TQI. For an indicator for isolated defects, five track geometry parameters defined in EN 13848-1 can be used: track gauge, longitudinal level, cross level, alignment, and twist [30]. According to EN 13848-5 [31], three severity levels are established for each geometry parameter to guide maintenance actions: IAL, IL, and AL [32]. For an indicator for overall TQI, it is defined by EN 13848-6 [32] as the CoSD, which is calculated by combining the weighted standard deviations of individual geometric parameters as

$$\text{CoSD} = \sqrt{w_{\overline{\text{AL}}} \text{SD}_{\overline{\text{AL}}}^2 + w_G \text{SD}_G^2 + w_C \text{SD}_C^2 + w_{\overline{\text{LL}}} \text{SD}_{\overline{\text{LL}}}^2} \quad (27)$$

where, for the individual geometry parameter i , SD_i is its standard deviation, w_i is its weighting factor, $\overline{\text{AL}}$ is the average alignment of the inner and outer rails, G is the track gauge, C is the cant, and $\overline{\text{LL}}$ is the average longitudinal level between two rails. This CoSD is calculated considering a track segment of the same distance.

In this work, the conventional TQI, referred to as CoSD, and the performance index based on the average between the standard deviation of the track longitudinal levels at the inner rail and that at the outer rail, referred to as avgSD, are considered baseline comparisons against our proposed $\text{KPI}_{\text{InSAR+ABA}}$. Note that the avgSD can be obtained by considering the observed value of track longitudinal levels in (21) and (22).

VI. RESULTS

A. Implementation Details

Due to limitations in the available measurement history, we discretize the temporal domain into five fixed subdomains.

Then, we vary the number of subdomains in the spatial domain to assess their impact on the overall analysis. For a given subdomain, the degrees of polynomials used in the curve fitting for the SAWP and the surface fitting for displacements are experimented with. The polynomials up to the fifth degree are trials. The hybrid neural model is developed separately for Tracks 1 and 2, in which the same network architecture is used to predict the longitudinal levels from both the inner and outer rails of the respective track. We experiment with the number of hidden layers used within the CNN, GRU, and LSTM structures. For the CNN structure, up to three convolutional layers are considered, and four different numbers of filters are trial: 64, 32, 16, and 8. We also experiment with different filter sizes: 3×3 , 5×5 , 7×7 , and 11×11 . The rectifier or ReLU activation function is used in CNN. For the GRU and LSTM structure, up to two layers are considered, and six different numbers of hidden units are trial: 256, 128, 64, 32, 16, and 8. This is followed by a dropout layer to prevent overfitting. Next, a dense layer and the L_2 regularization are used to maintain generalizability. The hyperbolic tangent and sigmoid functions are exploited as the state and gate activation functions, respectively, in this dense layer.

For the model training, the following parameter setting is specified: the number of data in each batch (small mini-batch size) = 10, the maximum number of epochs = 300, and the learning rate for parameter updates = 0.01. The dataset is divided into training and test sets with a ratio of 60:40, in which 90% of the training set is used to train the model and the other 10% is used for validating the trained model. The trained model is then tested with the holdout test set to evaluate its performance for the regression task. The models are trained with fivefold cross validation considering inner and outer rail data. Furthermore, the early stopping technique is also used to prevent overfitting during model training.

B. Results of Different Interpolation Functions

This section investigates the impact of using different interpolation functions on predictive performance. In this analysis, the CNN-GRU model is used to showcase the analysis ($\alpha = 0.0$) and we consider uniformly discretizing the domain into 35 subdomains. This is obtained by dividing the spatial and temporal domains uniformly into seven and five subdomains, respectively. The tested interpolation functions include local linear regression, polynomial function, nearest interpolation, biharmonic interpolation, cubic spline, Kriging, and Gaussian process regression. We assess the performance in two aspects: interpolation performance and predictive performance. To assess the interpolation and predictive performance, the deviations of the estimates of the missing values from the actual values are measured using RMSE and R^2 .

Table I shows that the polynomial function yields the best predictive performance, although the nearest, biharmonic, cubic spline interpolation, Kriging, and Gaussian process regression methods exhibit superior fitting performance. The experiments reveal that the optimal polynomial degree varies across different subdomains, reflecting the diverse distribution of displacement data in each area. By tailoring the polynomial degree to each subdomain, the polynomial interpolation

TABLE I
IMPACT ON THE PREDICTIVE PERFORMANCE WHEN USING DIFFERENT INTERPOLATION FUNCTIONS. THE RESULTS OBTAINED FROM THE HYBRID CNN-GRU OF TRACK 1 ARE SHOWN

Function	Performance			
	Interpolation		Predictive	
	RMSE	R^2	RMSE	R^2
Local linear regression	0.0030	0.5547	0.0489	0.5061
Polynomial	0.0027	0.7731	0.0403	0.8048
Nearest interpolation	0.0	1.0	0.0472	0.7244
Biharmonic	0.0	1.0	0.0458	0.1959
Cubic spline	0.0	1.0	0.0862	-2.7106
Kriging	0.0007	0.9743	0.0361	0.7571
Gaussian process	0.0011	0.9998	0.0886	0.5796

method has reasonable consistency between the predicted and observed values. The surface fitting of the displacements yields the average RMSE values of 0.0027 m and $R^2 = 0.7731$. This subdomain-specific approach ensures that the polynomial degree remains appropriate for the local data characteristics, avoiding the oscillatory behavior associated with overfitting. As a result, the polynomial interpolation achieves a balance between capturing local variations and maintaining global trends, both of which are critical for predictive performance.

Furthermore, Table I reveals that the success of interpolation methods in terms of RMSE and R^2 does not necessarily guarantee its suitability for predictive tasks. These methods might focus on minimizing local interpolation error, which can produce smooth interpolated surfaces that lack the distinctive features necessary for capturing spatial and temporal dependencies required by the CNN-GRU model.

While our work focuses on interpolation, it is worth mentioning that the DTW algorithm can also be considered as an alternative approach to address the challenges posed by different temporal resolutions among multiple data sources. Preliminary tests with DTW in this study indicated comparable performance to our interpolation method in terms of accuracy, but it did not yield significant improvements. Nevertheless, DTW may enhance temporal alignment in datasets with different characteristics and is worth exploring further. Additionally, time-series resampling strategies may also be employed to achieve better temporal synchronization and data quality.

C. Results of Different Numbers of Subdomains

Considering the polynomial functions obtained from Section VI-B to estimate values of displacements, this section investigates the impact of discretizing the domain into different numbers of subdomains. The CNN-GRU model is also used to showcase the analysis ($\alpha = 0.0$). It is observed in Table II that employing large subdomains for interpolation fails to capture local variations along the track. This leads to poor predictive performance, as indicated by a high error and a large R^2 value when discretizing the domain into five subdomains. However, performing an interpolation with subdomains that are too small can result in overfitting, where the interpolation passes through all the data points. This is evident in the well-fit performance with a small error and an R^2 value close to 1 when the domain is discretized into 50 subdomains. Analyzing the

TABLE II

IMPACT ON THE PREDICTIVE PERFORMANCE WHEN DISCRETIZING DOMAIN INTO DIFFERENT SIZES OF SUBDOMAINS. THE RESULTS ARE SHOWN FROM USING A POLYNOMIAL FUNCTION IN THE INTERPOLATION AND THE HYBRID CNN-GRU OF TRACK 1 IN THE PREDICTION

Numbers of subdomains	Performance			
	Interpolation		Predictive	
	RMSE	R^2	RMSE	R^2
5	0.0042	0.1197	0.0649	-6.4755
15	0.0035	0.3690	0.1127	0.3451
35	0.0027	0.7731	0.0403	0.8048
45	0.0021	0.7996	0.0782	0.0563
50	0.0018	0.8466	0.0703	0.0344
55	0.0020	0.8020	0.0989	-37.2659

results in Table II reveals that the best predictive performance is achieved through uniform discretization into 35 subdomains.

D. Comparative Study for Track Longitudinal Level Prediction

Following the results of Sections VI-B and VI-C, we discretize the domain into 35 subdomains using a polynomial for interpolation of displacements. For the SAWP, the degrees of polynomials used in the interpolation are also experimented with. The results suggest the utilization of a fifth-degree polynomial for the interpolation, in which the obtained interpolation performance is given with average RMSE values of $0.0012 \text{ m}^2/\text{s}^4$ and $R^2 = 0.9999$.

To evaluate the effectiveness of the hybrid neural approach, this article compares the results with CNN, GRU, LSTM, CNN-LSTM, CNN-GRU, and transformer-based models for CNN-LSTM and CNN-GRU trained with an attention mechanism. Each model employs the same parameter settings for the model training and undergoes the same experiments with different numbers of filters and sizes as described in Section VI-A. In order to assess the impact of the attention mechanism on accuracy, the network architectures of the CNN-GRU-based transformer and CNN-LSTM-based transformer are designed to align with their respective hybrid CNN-GRU and CNN-LSTM models. Then, we conduct experiments to optimize the number of attention heads and output sizes for the attention layer. Note that the notation 64 A ($h = 8$) denotes an attention layer with an output size of 64 and eight attention heads.

Table III compares performance obtained from different predictive models developed using data obtained from railway Tracks 1 and 2. The reported performance is an average value between the inner and outer rail. Different network architectures are designed for different railway tracks, with those outlined in Table III exhibiting the lowest RMSE values. Notably, the GRU and LSTM models perform better than the CNN model. This improvement is due to their ability to handle temporal characteristics, as demonstrated by a 38.32% reduction in DTW distance, indicating better temporal alignment between prediction and observation. On average, their RMSE improves from that of the CNN model by 24.15%, and their R^2 improves by 12.19%. Within the comparison between the GRU and LSTM models, the GRU model, designed according

to the network architecture specified in Table III, exhibits higher performance than the LSTM model for the dataset from Tracks 1 and 2. Employing a hybrid model approach leads to enhanced performance. This is evident in both CNN-LSTM and CNN-GRU models. The CNN-GRU model outperforms all the individual models for Tracks 1 and 2, improving with an average RMSE of 32.88%, R^2 of 40.40%, and DTW distance of 47.55%. The transformer-based models outperform the hybrid CNN-LSTM and CNN-GRU models, achieving lower RMSE and higher R^2 values on both tracks. On average, their RMSE improves by 12.26%, and their R^2 improves by 6.24%.

Notably, the CNN-GRU model yields inferior performance to the CNN-LSTM model for Track 1, which is impaired by 7.75% for RMSE, 1.24% for R^2 , and 7.05% for DTW distance. Conversely, the CNN-GRU model yields superior performance for Track 2, surpassing the CNN-LSTM model with an improved RMSE of 7.96%, R^2 of 7.45%, and DTW distance of 1.53%. This suggests that neither CNN-LSTM nor CNN-GRU consistently outperforms the other; rather, their effectiveness depends on the specific context and data characteristics. This observation aligns with findings reported in the literature (see [34], [35], [36]). To address these, we propose a unified framework that can adapt to these variations through an automated model selection mechanism, as described in (19) and illustrated in Fig. 3. This mechanism enables us to select between CNN-LSTM and CNN-GRU models based on the specific objectives of each case, for example, balancing predictive accuracy and training efficiency. Additionally, these accuracy variations can be attributed to the distinct spatiotemporal characteristics and operational conditions unique to each transition zone. Future work could incorporate dynamic modeling techniques using time-varying parameters, such as state-space models, to capture changes in system behavior over time. Additionally, online learning algorithms that update model parameters as new data becomes available can be particularly beneficial for adapting to evolving conditions without requiring complete retraining [36], [37].

When comparing the computational time measured as the average time taken to complete one epoch, Table III indicates that, on average, the CNN-GRU and CNN-LSTM models require 56.62% more time per epoch than the individual models to achieve these improvements. This is due to the more complex structure employed in their models. Compared to the transformer-based models, while they achieve higher prediction accuracy, these improvements come with an additional computational cost. The transformer-based models require approximately 27%–31% more time per epoch compared to their hybrid counterparts. This highlights a tradeoff between enhanced model performance and computational efficiency. While the transformer-based models enhance accuracy, their higher computational demand may limit their practicality in resource-constrained environments.

Additionally, Table III highlights a tradeoff between accuracy and computational efficiency when comparing the CNN-LSTM and CNN-GRU models. Considering the average performance across Tracks 1 and 2, CNN-GRU achieves slightly better accuracy, with an average RMSE reduction of 1.62% and an average R^2 values improvement of 2.78%.

TABLE III

PERFORMANCE COMPARISON AMONG DIFFERENT PREDICTIVE MODELS. THE REPORTED PERFORMANCE IS AN AVERAGE VALUE BETWEEN THE INNER AND OUTER RAIL AND COMPUTED ALONG THE TRACK SEGMENT OVER FIVE PREDICTED YEARS. THE HYBRID EXECUTES CNN-LSTM AND CNN-GRU MODELS IN PARALLEL, AND THE TRAINING TIME PER EPOCH CORRESPONDS TO THE LONGER OF THE TWO MODELS

Track	Model	Architecture	RMSE	MAE	R ²	DTW distance	Time per epoch (sec)
1	CNN	128-64	0.0828	0.0579	0.5203	239.23	0.1200
	GRU	32-16-8	0.0477	0.0327	0.5845	120.95	0.1620
	LSTM	128-64	0.0779	0.0584	0.5829	170.53	0.7200
	CNN-LSTM transformer	32-16-8-32-16-32A(h=4)	0.0339	0.0014	0.8437	72.99	0.6000
	CNN-GRU transformer	32-16-8-128-64-64A(h=4)	0.0381	0.0297	0.8299	73.97	0.2900
	CNN-LSTM	32-16-8-32-16	0.0374	0.0015	0.8149	80.14	0.4680
	CNN-GRU	32-16-8-128-64	0.0403	0.0275	0.0848	85.79	0.2150
	Hybrid approach ($\alpha = 0.0$)	32-16-8-32-16 and 32-16-8-128-64	0.0403	0.0275	0.0848	85.79	0.4680
	Hybrid approach ($\alpha = 1.0$)	32-16-8-32-16 and 32-16-8-128-64	0.0374	0.0015	0.8149	80.14	0.4680
2	CNN	64-64	0.1045	0.0744	0.5222	293.56	0.1454
	GRU	32-16-8	0.0558	0.0396	0.5848	132.89	0.1646
	LSTM	64-32	0.0868	0.0629	0.5436	233.70	0.6320
	CNN-LSTM transformer	32-16-8-32-16-128A(h=8)	0.0619	0.0435	0.7278	113.21	0.6714
	CNN-GRU transformer	32-16-8-128-64-32A(h=4)	0.0501	0.0343	0.8194	105.40	0.2674
	CNN-LSTM	32-16-8-32-16	0.0553	0.0395	0.7016	124.31	0.4800
	CNN-GRU	32-16-8-128-64	0.0509	0.0355	0.7539	122.41	0.2236
	Hybrid approach ($\alpha = 0.0$)	32-16-8-32-16 and 32-16-8-128-64	0.0509	0.0355	0.7539	122.41	0.4800
	Hybrid approach ($\alpha = 1.0$)	32-16-8-32-16 and 32-16-8-128-64	0.0509	0.0355	0.7539	122.41	0.4800

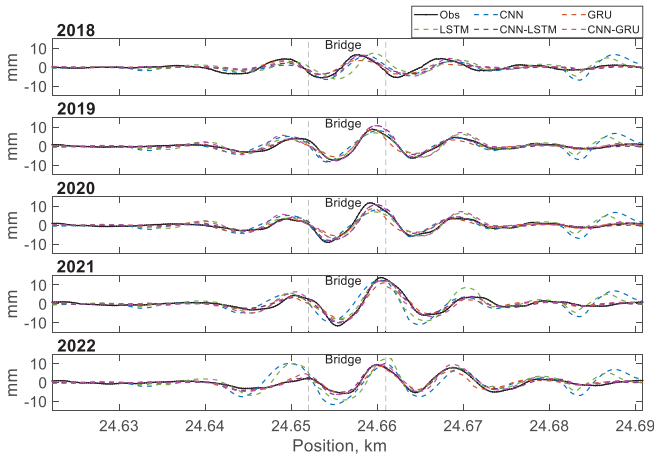


Fig. 5. Predicted track longitudinal levels obtained from different models. The inner rail of Track 2 is shown.

However, CNN-GRU demonstrates a notable advantage in computational efficiency, being approximately 53.72% more efficient on average. For applications where computational efficiency is critical, CNN-GRU thus offers a viable solution with competitive accuracy while being significantly faster. Conversely, for use cases where accuracy holds higher priority and sufficient computational resources are available, CNN-LSTM becomes a favorable choice. Future optimization efforts could focus on reducing the number of model parameters. This involves incorporating pruning techniques to enhance computational efficiency further, enabling these models to adapt more effectively to diverse application requirements.

Fig. 5 illustrates a comparison between the predicted track longitudinal levels obtained from different models in the years 2018–2022. These are the results obtained from the inner rail of Track 2. It can be seen that different models demonstrate different capabilities in capturing the spatial and temporal

characteristics presented in the track longitudinal level changes over the years. This capability can be investigated through the spatial and temporal relationship between the predictions and measurements. In other words, a consistent trend between the predictions and measurements along the track positions and across the years should be observed. The predicted values obtained from the individual models, i.e., CNN, GRU, and LSTM, exhibit inconsistent trends with the real measurements. The inconsistency trend is more pronounced in 2022. This is evident by their smaller R^2 values, indicative of a less consistent trend with the observed track longitudinal levels. In contrast, the CNN-LSTM and CNN-GRU models exhibit more consistent trends, as evidenced by their higher R^2 values compared to the individual models. This suggests a superior ability to capture spatial characteristics. As time evolves, the increased errors in the predicted values from CNN, GRU, and LSTM models are observed. Also, the trend in their predictions shifts, failing to accurately capture peaks. While a slight drop in capturing peaks is observed in the CNN-GRU model, it demonstrates an ability to adapt to the trends of the track longitudinal changing over time. This emphasizes the effectiveness of the CNN-LSTM and CNN-GRU models in predicting track longitudinal levels. Regarding robustness and generalization, the use of fivefold cross validation, early stop, dense layer, and L2 regularization appears to be sufficient in view of the prediction results. The model was evaluated across five years, including a nearly two-year prediction horizon during COVID-19 (when ABA data could not be collected with our measurement train), demonstrating consistent performance under varying operational conditions.

E. KPI Based on InSAR and ABA Data for the Health Assessment of Railway Transition Zones

This section employs the results obtained from the CNN-GRU to showcase the evaluation of the proposed

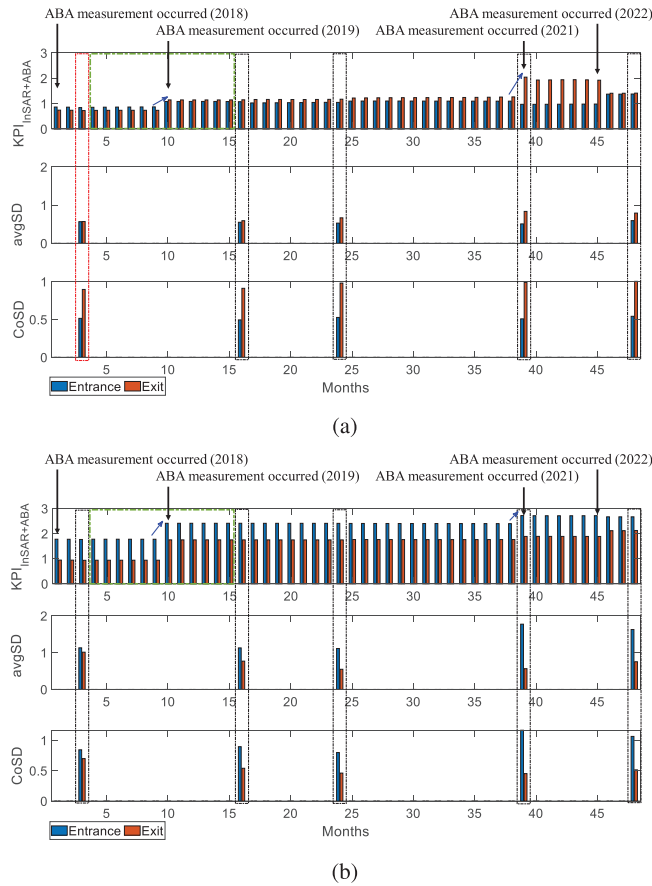


Fig. 6. Continuous evolution of the KPI at the entrance and exit side of the transition zones from September 2018, marked by month 1, to August 2022, marked by month 48. (a) Track 1. (b) Track 2.

$KPI_{InSAR+ABA}$ ($\alpha = 0.0$). Fig. 6 shows a continuous evolution of the proposed $KPI_{InSAR+ABA}$, at the entrance and exit side of the transition zones from September 2018, marked by month 1, to August 2022, marked by month 48. There were only five measurements of track geometry parameters within this period. Notably, the $KPI_{InSAR+ABA}$ provides consistent health conditions with the CoSD and the avgSD in the months highlighted in the black boxes for Tracks 1 and 2. Specifically, from June 2019, marked by month 10, the $KPI_{InSAR+ABA}$ is similar to the trend of the CoSD and avgSD derived from the track geometry measurements. The transition zone at the exit side was more degraded than the entrance for Track 1, whereas the entrance side was more degraded than the exit for Track 2. Likewise, the CoSD, avgSD, and $KPI_{InSAR+ABA}$ consistently revealed that Track 2 was more degraded than Track 1, evident by their higher value.

In month 3, a contrasting trend is observed in Track 1, as highlighted by the red boxes. The $KPI_{InSAR+ABA}$ analysis for Track 1 uncovered that the entrance side of the transition zone exhibited greater degradation compared to the exit side. This is consistent with the information obtained from the ABA measurements depicted in Fig. 7(a), where the average values of SAWP on the entrance side of the transition zone surpass those on the exit side for Track 1. The inconsistency with the CoSD in Track 1 can be attributed to the inherent integration

of ABA measurements in predicting track longitudinal levels. This integration enables the inclusion of local changes within the substructure layers in the proposed KPI, a capability that traditional track geometry measurements may lack. This results from the consideration of the spatial frequency range of $0.04\text{--}0.33\text{ m}^{-1}$ (substructure related) in the SAWP calculation.

Furthermore, the track longitudinal levels at unobserved months can be obtained using our hybrid model. This results in a more frequent estimation of the KPI that can be used to facilitate regular assessments of transition zone health, owing to a high-frequency measurement of InSAR data. This approach can enhance the ability to detect changes in track irregularities early, even before obtaining the next measurement of track geometry profiles, as illustrated in the green box. In Fig. 6(a) and (b), abrupt changes of the $KPI_{InSAR+ABA}$ value are observed between months 9 and 10 and 38 and 39. This was when the track longitudinal levels were predicted from the ABA measurements of the most recent year. As seen in Fig. 7, the SAWP values for Tracks 1 and 2 in 2019 increased from those of 2018, resulting in more degraded conditions of the transition zones in 2019 compared to 2018. This pattern is followed between months 38 and 39 as the increased SAWP values are observed in Fig. 7(c) and (f) for Tracks 1 and 2 compared to those in 2019. Consequently, the $KPI_{InSAR+ABA}$ in month 10 is higher than in month 9, and similarly, the $KPI_{InSAR+ABA}$ in month 39 exceeds that of month 38. This consistent trend is observed for both tracks. Evaluating from the $KPI_{InSAR+ABA}$ values, in cases where the severity is classified as high, this proactive information allows for effective planning of maintenance actions.

F. Discussion

This section delves deeper into the analysis of predicted track longitudinal levels, providing insights into the physics. The investigation is showcased using data collected in 2020 from Track 1. While consistency is evident between the prediction and observation in the time domain, depicted in Fig. 8(a), a distinct variation in energy distribution within the frequency domain is noticeable. As seen in Fig. 8(b) and (c), differences are observed in the frequency content for wavelengths ranging from 3 to 25 m when comparing observations to predictions. This is more pronounced, particularly when the wavelength is less than 5 m for the entrance and exit sides of the inner and outer rails. The predicted track longitudinal levels at the railway bridge transition zones exhibit different characteristics, reflecting the limitation arising from the data-driven nature of the predictive models employed in this article. This comes from a lack of the incorporation of physical information from the track system.

Now, to evaluate the impact of wrong predictions obtained from our proposed methodology on the decision for maintaining tracks at transition zones, we consider the indicator for isolated defects, specifically focusing on the track longitudinal level. As described in Section V-B, IAL stands for immediate action limit. It refers to a threshold level of track condition severity. When IAL is exceeded, urgent maintenance actions are required. Maintenance activities or other countermeasures, such as closing the line or reducing operational speed, must

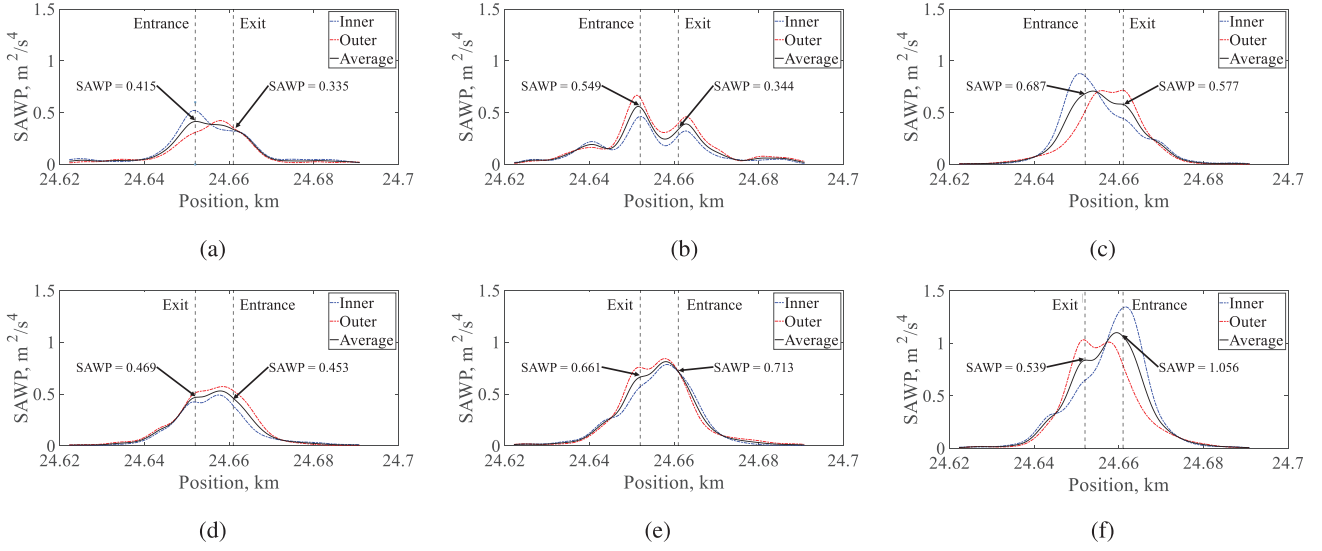


Fig. 7. Interpolated SAWP for predicting the track longitudinal levels. (a) Track 1, 2018. (b) Track 1, 2019. (c) Track 1, 2021. (d) Track 2, 2018. (e) Track 2, 2019. (f) Track 2, 2021.

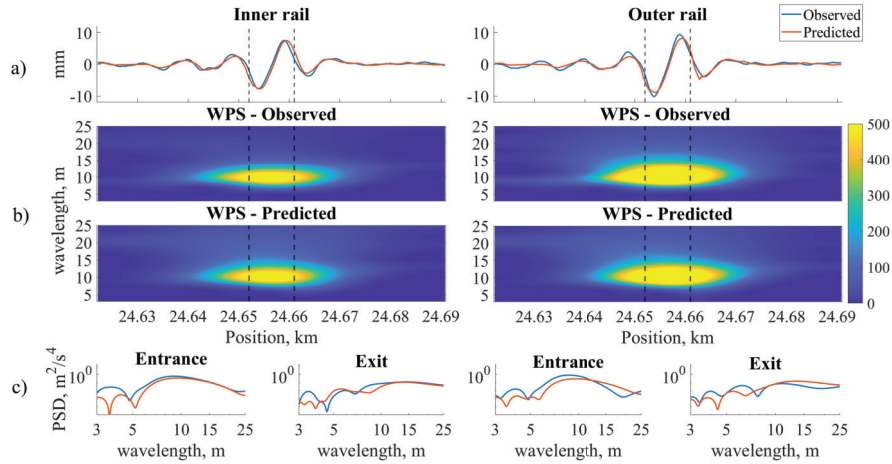


Fig. 8. Comparison results between the predicted and observed track longitudinal levels in (a) time domain, and frequency domain considering (b) WPS and (c) PSD at the transition zones from track 1.

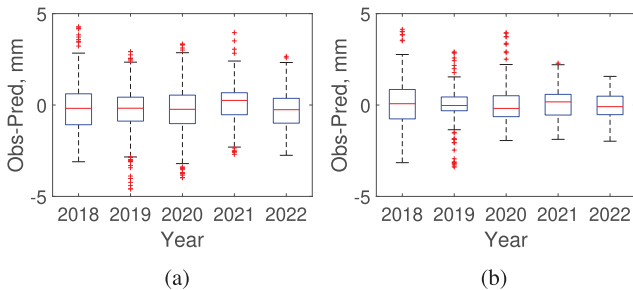


Fig. 9. Distribution of discrepancy between the observations and predictions along the distance over years. The results from the inner rail of Track 2 are shown. (a) CNN-GRU. (b) CNN-LSTM.

be conducted immediately to reduce the risk of derailment and ensure the safety of train operations. IAL is directly related to safety in train operations compared to AL and IL, which serve as additional indicators for ensuring good service quality

and supporting maintenance planning. Thus, we use IAL to evaluate the influence of wrong predictions. For our case study, the typical operational speed on the Dutch railway network is 140 km/h, so the applicable IAL is 23 mm (absolute value) according to EN 13848-5 [31]. It can be seen from Fig. 9 that the discrepancies between the predictions and observations are below the IAL for both hybrid models. This suggests that even with some prediction errors, the necessary safety measures and maintenance activities can still be identified and implemented without risk.

A key contribution of this work is the integration of three complementary measurement techniques (ABA, InSAR, and TG) into a unified KPI. Existing studies in the literature utilize heterogeneous data sources, such as bogie acceleration or maintenance logs, which are not registered for our current test case. Consequently, direct benchmarking against those approaches is not possible. Future research can consider the development of a benchmark with diverse multisource

measurements, at various transition zones, and even across different countries. Such a benchmark would facilitate a rigorous comparative evaluation of techniques.

Although AI is advancing rapidly, with newer models such as ResNet, EfficientNet, ConvNeXt, and Vision Transformers offering improved scalability and performance, our focus is on practical implementation with well-established spatiotemporal models for railway condition monitoring. Nonetheless, we acknowledge the importance of benchmarking against state-of-the-art methods in future work, as more comprehensive and diverse datasets become accessible. Emerging sequential and spatiotemporal modeling techniques, such as temporal convolutional networks, informer, spatiotemporal graph neural networks, or techniques that combine physics information and machine learning, offer promising alternative approaches that are interesting for further research.

VII. CONCLUSION

This work presents a data fusion framework that can be used to enable a more frequent evaluation of transition zone health by integrating multiple monitoring technologies, including track geometry, InSAR, and ABA measurements. To illustrate its effectiveness, a case study at a railway bridge between Dordrecht and Lage Zwaluwe station in The Netherlands is used. Compared to individual models, i.e., CNN, GRU, and LSTM, hybrid CNN-LSTM and CNN-GRU exhibit a superior ability to capture the spatial and temporal relationships between the track longitudinal levels, InSAR, and ABA measurements. Although the transformer-based models enhance prediction accuracy for both tracks, these improvements come at the cost of increased training time. This tradeoff highlights the balance between model performance and computational efficiency. By using a simple CNN and LSTM/GRU combination, our approach ensures computational efficiency, making it well suited for applications in the railway industry where computational efficiency is critical. The hybrid approach allows for more accurate predictions of track longitudinal levels during unobserved months, leading to a more frequent estimation of the KPI. This is due to the high-frequency measurement provided by InSAR data. As new measurements become available, the hybrid models facilitate the timely estimation of KPI values. The proposed KPI proves effective in detecting changes in track irregularities, thereby enabling regular assessments of transition zone health.

Future research lines include evaluating the generalization and robustness of the proposed methodology and the proposed KPI based on InSAR and ABA data for assessing the health of transition zones. This involves using a larger dataset containing diverse locations under varying transition zone conditions. This alleviates the risk of overfitting. Advanced data fusion techniques, such as Bayesian fusion or ensemble learning, could be explored to achieve a more sophisticated and nuanced integration of heterogeneous data sources. Then, probabilistic models, e.g., the fuzzy-interval method and Bayesian networks, can be considered as predictive models to incorporate and quantify uncertainty in the data arising from diverse locations and conditions. The inclusion of physical variables and maintenance information in the predictions can also be

considered via physics-informed machine learning to better capture the dynamic behavior of transition zones. Given that ABA measurements depend on speed, a method to mitigate the impact of speed variations in measurements is needed to facilitate the broader application of the proposed framework. Furthermore, railway systems are inherently dynamic and subject to sudden changes, such as maintenance activities and environmental factors. These changes may not be uniformly reflected across all data sources, potentially impacting model accuracy over time. Future work could explore the development of adaptive models or feedback mechanisms capable of capturing and responding to these dynamic shifts in data relationships. Future work should also focus on extending the robustness analysis through broader case studies, such as nationwide assessments of transition zones and cross-national comparisons to evaluate transfer learning capabilities. Additional evaluations under perturbed conditions, e.g., noisy inputs or missing data, along with regularization and data augmentation techniques, are recommended to enhance model generalization in data-limited scenarios.

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 article.

ACKNOWLEDGMENT

Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

REFERENCES

- [1] L. N. Wheeler, E. Pannese, N. A. Hoult, W. A. Take, and H. Le, "Measurement of distributed dynamic rail strains using a Rayleigh backscatter based fiber optic sensor: Lab and field evaluation," *Transp. Geotechnics*, vol. 14, pp. 70–80, Mar. 2018.
- [2] K. Nasrollahi, J. Dijkstra, and J. C. O. Nielsen, "Towards real-time condition monitoring of a transition zone in a railway structure using fibre Bragg grating sensors," *Transp. Geotechnics*, vol. 44, Jan. 2024, Art. no. 101166.
- [3] B. Akpinar and E. Gulal, "Multisensor railway track geometry surveying system," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 1, pp. 190–197, Jan. 2012.
- [4] Y. Li, Q. Zhao, J. Guo, R. Fang, and J. Zheng, "Multisensor fusion for railway irregularity inspection system: Integration of RTK GNSS, MEMS IMU, odometer, and laser," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–12, 2024.
- [5] C. Alves Ribeiro, R. Calçada, and R. Delgado, "Experimental assessment of the dynamic behaviour of the train-track system at a culvert transition zone," *Eng. Struct.*, vol. 138, pp. 215–228, May 2017.
- [6] D. Mishra, E. Tutumluer, H. Boler, J. P. Hyslip, and T. R. Sussmann, "Railroad track transitions with multidepth deflectometers and strain gauges," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2448, no. 1, pp. 105–114, Jan. 2014.
- [7] H. Wang, V. Markine, and X. Liu, "Experimental analysis of railway track settlement in transition zones," *Proc. Inst. Mech. Engineers, F, J. Rail Rapid Transit*, vol. 232, no. 6, pp. 1774–1789, Jul. 2018.
- [8] K. Nasrollahi, J. C. O. Nielsen, E. Aggestam, J. Dijkstra, and M. Ek, "Prediction of long-term differential track settlement in a transition zone using an iterative approach," *Eng. Struct.*, vol. 283, May 2023, Art. no. 115830.
- [9] D. F. Hesser, K. Altun, and B. Markert, "Monitoring and tracking of a suspension railway based on data-driven methods applied to inertial measurements," *Mech. Syst. Signal Process.*, vol. 164, Feb. 2022, Art. no. 108298.

- [10] A. Sabato and C. Niezrecki, "Feasibility of digital image correlation for railroad tie inspection and ballast support assessment," *Measurement*, vol. 103, pp. 93–105, Jun. 2017.
- [11] I. La Paglia, M. Carnevale, R. Corradi, E. Di Gialleonardo, A. Facchinetti, and S. Lisi, "Condition monitoring of vertical track alignment by bogie acceleration measurements on commercial high-speed vehicles," *Mech. Syst. Signal Process.*, vol. 186, Mar. 2023, Art. no. 109869.
- [12] S. Unsiwilai, L. Wang, A. Núñez, and Z. Li, "Multiple-axle box acceleration measurements at railway transition zones," *Measurement*, vol. 213, May 2023, Art. no. 112688.
- [13] M. Khosravi, A. Ahmadi, and A. Nissen, "A multi-objective approach for position alignment of track geometry measurements," *Eng. Failure Anal.*, vol. 149, Jul. 2023, Art. no. 107260.
- [14] H. A. Fernández-Bobadilla and U. Martin, "Modern tendencies in vehicle-based condition monitoring of the railway track," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–44, 2023.
- [15] Y. Zhu, M. Shuang, D. Sun, and H. Guo, "Algorithm and application of foundation displacement monitoring of railway cable bridges based on satellite observation data," *Appl. Sci.*, vol. 13, no. 5, p. 2868, Feb. 2023.
- [16] Y. Zhu et al., "Early detection of potential landslides along high-speed railway lines: A pressing issue," *Earth Surf. Processes Landforms*, vol. 48, no. 15, pp. 3302–3314, Dec. 2023.
- [17] L. Chang, N. P. Sakpal, S. O. Elberink, and H. Wang, "Railway infrastructure classification and instability identification using Sentinel-1 SAR and laser scanning data," *Sensors*, vol. 20, no. 24, p. 7108, Dec. 2020.
- [18] H. Wang, L. Chang, and V. Markine, "Structural health monitoring of railway transition zones using satellite radar data," *Sensors*, vol. 18, no. 2, p. 413, Jan. 2018.
- [19] X. Xu, D. Zhao, C. Ma, and D. Lian, "Monitoring subsidence deformation of Suzhou subway using InSAR timeseries analysis," *IEEE Access*, vol. 9, pp. 3400–3416, 2021.
- [20] A. Lasisi and N. Attoh-Okin, "Principal components analysis and track quality index: A machine learning approach," *Transp. Res. C, Emerg. Technol.*, vol. 91, pp. 230–248, Jun. 2018.
- [21] J. S. Lee, S. H. Hwang, I. Y. Choi, and I. K. Kim, "Prediction of track deterioration using maintenance data and machine learning schemes," *J. Transp. Eng., A, Syst.*, vol. 144, no. 9, Sep. 2018, Art. no. 04018045.
- [22] H. Khajehi, A. Ahmadi, I. Soleimanmeigouni, M. Haddadzade, A. Nissen, and M. J. Latifi Jebelli, "Prediction of track geometry degradation using artificial neural network: A case study," *Int. J. Rail Transp.*, vol. 10, no. 1, pp. 24–43, Jan. 2022.
- [23] M. Sedghi, *Enhancing the Effectiveness of Neural Networks in Predicting Railway Track Degradation*. Cham, Switzerland: Springer, 2024, pp. 651–664.
- [24] L. Han, Y. Liao, H. Wang, and H. Zhang, "Long-term prediction for railway track geometry based on an optimised DNN method," *Construction Building Mater.*, vol. 401, Oct. 2023, Art. no. 132687.
- [25] M. Chen, S. Zhu, W. Zhai, Y. Sun, and Q. Zhang, "Inversion and identification of vertical track irregularities considering the differential subgrade settlement based on fully convolutional encoder-decoder network," *Construct. Building Mater.*, vol. 367, Feb. 2023, Art. no. 130057.
- [26] X. Wang, Y. Bai, and X. Liu, "Prediction of railroad track geometry change using a hybrid CNN-LSTM spatial-temporal model," *Adv. Eng. Informat.*, vol. 58, Oct. 2023, Art. no. 102235.
- [27] W. Phusakulkajorn et al., "Detection of rail surface defects based on axle box acceleration measurements: A measurement campaign in Sweden," in *Proc. The 10th Transp. Res. Arena (TRA)*, Dublin, Ireland, 2024, pp. 1–6.
- [28] A. Islam, S. B. Belhaouari, A. U. Rehman, and H. Bensmail, "KNNOR: An oversampling technique for imbalanced datasets," *Appl. Soft Comput.*, vol. 115, Jan. 2022, Art. no. 108288.
- [29] J. Hao and R. Liu, "Prediction for geometric degradation of track on machine learning," *Proc. SPIE*, vol. 12790, Sep. 2023, Art. no. 127903U.
- [30] Railway Applications-Track-Track Geometry Quality—Part 1: Characterization of Track Geometry, Standard 13848-1:2019, CEN, NEN-EN, 2019. [Online]. Available: <https://www.nen.nl/en/nen-en-13848-1-2019-en-257518>
- [31] Railway Applications-Track-Track Geometry Quality—Part 5: Geometric Quality Levels-plain Line, Switches and Crossing, Standard 13848-5:2017, CEN, NEN-EN, 2017. [Online]. Available: <https://www.nen.nl/nen-en-13848-5-2017-en-237864>
- [32] Railway Applications-Track-Track Geometry Quality—Part 6: Characterization of Track Geometry Quality, Standard 13848-6:2014, CEN, NEN-EN, 2014. [Online]. Available: <https://www.nen.nl/nen-en-13848-6-2014-en-194073>
- [33] F.-S. Hsu et al., "Benchmarking of eight recurrent neural network variants for breath phase and adventitious sound detection on a self-developed open-access lung sound database—HF_Lung_V1," *PLoS ONE*, vol. 16, no. 7, Jul. 2021, Art. no. e0254134.
- [34] T. Wilaiprasitporn, A. Dittaporn, K. Matchaparn, T. Tongbuasirilai, N. Banluesombatkul, and E. Chuangsuwanich, "Affective EEG-based person identification using the deep learning approach," *IEEE Trans. Cogn. Devel. Syst.*, vol. 12, no. 3, pp. 486–496, Sep. 2020.
- [35] H. Naeem and A. A. Bin-Salem, "A CNN-LSTM network with multi-level feature extraction-based approach for automated detection of coronavirus from CT scan and X-ray images," *Appl. Soft Comput.*, vol. 113, Dec. 2021, Art. no. 107918.
- [36] D. Chen, R. Chen, Y. Li, and T. Tang, "Online learning algorithms for train automatic stop control using precise location data of balises," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1526–1535, Sep. 2013.
- [37] O. Marschall, K. Cho, and C. Savin, "A unified framework of online learning algorithms for training recurrent neural networks," *J. Mach. Learn. Res.*, vol. 21, pp. 1–34, Jan. 2019.



Wassamon Phusakulkajorn received the B.Sc. degree (Hons.) in mathematics from the Prince of Songkla University, Songkhla, Thailand, in 2007, and the Ph.D. degree in railway engineering from Delft University of Technology, Delft, The Netherlands, in 2024.

She was a Post-Doctoral Researcher at Delft University of Technology until April 2025 contributing to Europe's Rail Joint Undertaking (ERJU) Flagship Project IAM4RAIL. Currently, she is with the Rail and Modern Transports Research Group, National Metal and Materials Technology Center (MTEC), National Science and Technology Development Agency (NSTDA), Pathum Thani, Thailand. Her research interests lie in machine learning, big data analytics, data fusion, predictive analytics, and their applications in the condition monitoring and maintenance of railway infrastructure and machinery.

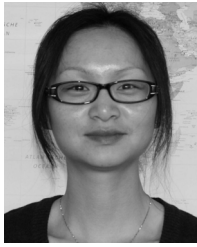
Dr. Phusakulkajorn received the Best Paper Award from Oxford University Press in 2023 for her comprehensive review of AI in railway infrastructure.



Siwarak Unsiwilai (Member, IEEE) received the B.Eng. and M.Eng. degrees in civil engineering from Chulalongkorn University, Bangkok, Thailand, in 2010 and 2013, respectively, and the Ph.D. degree in railway engineering from Delft University of Technology, Delft, The Netherlands, in 2024.

He is a Post-Doctoral Researcher with the Section of Railway Engineering, Delft University of Technology, for the term of 2024–2025, contributing to Europe's Rail Joint Undertaking (ERJU) Flagship Project IAM4RAIL. His research focuses on axle

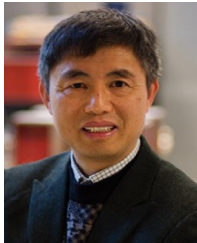
box acceleration (ABA)-based monitoring to assess track support conditions and applying data fusion techniques to support railway maintenance activities. In 2025, he will join the Department of Civil Engineering, Chulalongkorn University, to continue his academic career as a Lecturer.



Ling Chang received the M.S.E. degree in geodesy and survey engineering from Tongji University, Shanghai, China, in 2010, and the Ph.D. degree from Delft University of Technology, Delft, The Netherlands, in 2015.

She is currently an Associate Professor of microwave remote sensing with the Department of Earth Observation Science, Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, The Netherlands. Her research interests include statistical hypothesis test-

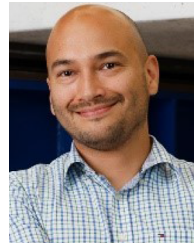
ing, time-series modeling, and change detection using satellite-borne remote sensing technology, particularly interferometric synthetic aperture radar techniques.



Zili Li received the B.Sc. and M.Sc. degrees in mechanical engineering from Southwest Jiaotong University, Chengdu, China, in 1988 and 1991, respectively, and the Ph.D. degree in computational mechanics on numerical solution of rolling contact from Delft University of Technology, Delft, The Netherlands, in 2002.

From 1999 to 2005, he was with the Institute of Road Transportation, TNO, Delft, where he was involved in research and software development of multibody dynamics and finite element method for crash safety. In 2005, he joined the Faculty of Civil Engineering and Geosciences, Delft University of Technology, where he taught and performed research on railway engineering. He is currently a Full Professor with the Section of Railway Engineering, Delft University of Technology. He is also the Head of the Section of Railway Engineering and the Scientific Director of TU Delft Rail Institute (DelftRail), Delft University of Technology. His current research interests include health monitoring and asset management of railway infrastructure; numerical solution of frictional rolling contact and its applications to analyses of wear, rolling contact fatigue, and vehicle dynamics; and train-track interaction, particularly in the high-frequency/short-wave range and at switches and crossings, and friction adhesion between wheel and rail.

Dr. Li is a member of the Editorial Board of *International Journal of Rail Transportation*, *International Journal of Railway Technology*, and *Advances in Mechanical Engineering*.



Alfredo Núñez (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Chile, Santiago, Chile, in 2010.

He was a Post-Doctoral Researcher with Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands; and a Visiting Scholar at universities in Slovenia, Italy, Spain, Chile, Colombia, China, and USA. He is an Associate Professor at Delft University of Technology, specializing in intelligent railway infrastructures. His work has been supported by the Dutch Research Council, ProRail, and European projects. He has contributed to major transportation conferences as a member of local organizations and as a speaker, delivering over 50 presentations. He has mentored award-winning Ph.D., M.Sc., and Eng.D. students and actively contributes to education, industry collaborations, and academic initiatives in railway infrastructure. His research focuses on intelligent transportation systems, railway engineering, and computational intelligence, with over 150 publications.

Dr. Núñez belongs to the Editorial Board of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, *Applied Soft Computing* (Elsevier), and *Intelligent Transportation Infrastructure* (Oxford Academic). He has been a Guest Editor of special issues in *Wear* (Elsevier) and *Control Engineering Practice*. He has held key roles in EU projects and served as an associate editor and a guest editor for leading journals.