# The Development of an Integrated Computing Platform for Measuring, Predicting and Analyzing Profile-specific Fixity of Railway Tracks

Fu, Qian; Easton, John; Burrow, Michael; Sweeney, James

Link to publication on Research at Birmingham portal

1  **The Development of an Integrated Computing Platform for Measuring, Predicting**
2  **and Analyzing Profile-specific Fixity of Railway Tracks**
3
4
5  **Qian Fu**
6  Research Fellow
7  Birmingham Centre for Railway Research and Education
8  Department of Electronic, Electrical and Systems Engineering, School of Engineering
9  University of Birmingham
10 Edgbaston, Birmingham B15 2TT, UK
11 Email: Q.Fu@bham.ac.uk
12 ORCID: 0000-0002-6502-9934
13
14 **John M. Easton**
15 Associate Professor
16 Birmingham Centre for Railway Research and Education
17 Department of Electronic, Electrical and Systems Engineering, School of Engineering
18 University of Birmingham
19 Edgbaston, Birmingham B15 2TT, UK
20 Email: J.M.Easton@bham.ac.uk
21 ORCID: 0000-0001-8745-6753
22
23 **Michael P. N. Burrow**
24 Reader
25 Birmingham Centre for Railway Research and Education
26 Department of Civil Engineering, School of Engineering
27 University of Birmingham
28 Edgbaston, Birmingham B15 2TT, UK
29 Email: M.P.Burrow@bham.ac.uk
30 ORCID: 0000-0001-5884-7763
31
32 **James Sweeney**
33 Senior Engineer
34 Network Rail
35 The Quadrant, Milton Keynes MK9 1EN, UK
36 Email: James.Sweeney@networkrail.co.uk
37

1  **ABSTRACT**

2  The current measures for railway track fixity in the UK's railway remain at a relatively low level
3  of granularity. This paper presents a pilot proof-of-concept study on the development of an
4  integrated computing framework for improving the measurement, prediction, and analysis of
5  profile-specific track fixity in the context of the UK's rail network. The framework is aimed to
6  produce a data integration and mining tool, which can determine track fixity parameters for any
7  given section of track. In this study, we propose to measure track movement based on point cloud
8  data and assess the track fixity by a set of parameters such as the direction and rate of the track
9  movement relative to the plane of rail within a certain period. We seek to integrate a data mining
10  algorithm into the framework to predict these parameters, given vast amounts of disparate and
11  heterogeneous data of potential influencing factors in the area. From the study, we have developed
12  a prototype framework, which allows the rapid implementation of data workflows with the
13  necessary functionality. The feasibility of the prototype was demonstrated by training a random
14  forest model on real data from an approximately 80-kilometer section of the East Coast Main Line,
15  southeast of Edinburgh, Scotland. The modelling results indicate that curvature, cant, and
16  maximum speed of trains are among the key factors that impact on, and are critical for predicting
17  and analyzing, the profile-specific track fixity.
18
19  **Keywords:** Railway track fixity, Track movement, Data integration, LiDAR point cloud, Random
20  forest

1    **INTRODUCTION**

2    Track fixity, which refers to the degree to which the position of a railway track remains unchanged
3    over time, is one of the key measures used to calculate clearances between rolling stock and
4    structures. In the UK's rail industry, the track fixity is typically presented as a simple value of
5    "low," "medium," or "high," with unrestrained ballast being the lowest and slab track being the
6    highest. The measurement of the fixity, however, remains at a low level of granularity; and there
7    is a lack of predictive tools that can provide more detailed, constantly updated information about
8    the movement of tracks through an automated process. As Network Rail – the major railway
9    infrastructure manager in the UK – seeks to improve the accuracy and reliability of its gauging
10   assessments, there is an immediate requirement for improved resolution of the track fixity across
11   its railway network. To achieve this would require explicit computation and prediction of the rate
12   and direction of the track movement within a given period, depending on the speed and use of the
13   track (e.g., six months or even a longer term). With such short-, medium- and long-term
14   calculations, track engineers would be able to better assess the risk of foul clearance developing
15   in a foreseeable future based on changes in the track alignment and historic surveys, thus making
16   better informed decisions about proactive actions against the risk. The assessment results would
17   in turn help identify and verify key factors influencing the track fixity in the area and thus allow
18   more intelligent prioritization of survey campaigns and preventative maintenance activities in the
19   event of resource challenges.
20        In essence, the possibility of some movement (i.e., shift or displacement) of a railway track
21   directly describes the track fixity. In the study discussed in this paper, accordingly, the track fixity
22   is measured in terms of both the rate and direction of the track movement, given a certain
23   confidence level. More specifically, the parameters of the track fixity examined in this study
24   include the displacements of any rail head in both the horizontal and vertical planes (relative to the
25   plane of the rail) within a certain period.
26        The track movement can be caused by numerous factors. Apart from known factors, such
27   as track geometry (*1*), track subgrade (*2*, *3*), track conditions (*4*), and train dynamics (*5*, *6*, *7*) there
28   can be an interplay of various factors, which can potentially have some direct or indirect impact
29   on track movement. To better understand the relative influence of these factors on track movement,
30   and identify any unidentified risk requires mining a large amount (e.g., terabytes worth of data)
31   and a wide spectrum of data, allied to the elicitation of information from experts. Usually, these
32   data are made available from disparate sources across different rail subsystems, such as
33   engineering structures and rolling stock; many of them, such as LiDAR data, are likely to be
34   clustered with redundant or irrelevant information and some may be unstructured. The diversity
35   and unobserved heterogeneity of the data resources often poses a major obstacle to data integration
36   for meeting the requirements of examining the profile-specific track fixity, given a higher degree
37   of granularity. To overcome these difficulties, there is clearly a critical need to develop an
38   integrated computing framework for facilitating a congruent workflow, involving: a) effective
39   integration of the heterogeneous data into a unified view, followed by b) implementation of
40   appropriate tools for timely calculation of track fixity and prediction of fixities of new structures.
41   Despite the growing salience of such kind of models (e.g., *8*, *9*, *10*, *11*), few studies in the literature
42   have investigated this subject.
43        This paper presents a pilot proof-of-concept study that seeks to: a) design a data pre-
44   processing workflow, which enables the smooth integration and management of a structured
45   corpus of data that is relevant to track fixity; and b) create a data mining tool as a prototype, which
46   can assist track engineers in measuring, predicting, and analyzing the track fixity parameters for

1   any given section of a railway track. In this phase of the framework development, we compile a
2   comprehensive database for a representative section of the railway track in the UK. It is used as
3   the fundamental building block of the computing framework for generating track fixity values,
4   which would serve as a basis for all reference curves used across the whole rail network. On this
5   basis, we develop data-driven models to investigate the relationships between potential influencing
6   factors and track movement. Through this analysis, we identify and verify some key parameters
7   that can serve as predictors of future track movements. Accordingly, the present paper is dedicated
8   to:
9      1)  proposing an effective metric and method of calculating track movements using LiDAR
10         (Light Detection and Ranging) point cloud data;
11     2)  developing an integrated data model with a machine learning model (e.g., a random forest
12         model), which is trained on data of calculated track movements and a set of empirically
13         selected relevant factors that potentially influence the movements; and
14     3)  verifying the identified influencing factors on the track movement.
15  The remainder of the paper consists of three sections. In the following section, we provide detailed
16  information on the data resources used in the study and propose a method for calculating the rate
17  and direction of track movement. Next, we demonstrate the method with a case-study example on
18  a selected section of the railway track of the East Coast Main Line between Edinburgh and the
19  Scottish border. We present a prototype computing framework for track fixity, which considers a
20  selected set of key factors influencing the track movement in the UK context. In the concluding
21  section, we summarize the outcomes of the work, discuss limitations of our study, and suggest
22  possible avenues for further research.
23
24
25  **METHODOLOGY**

26  To address the challenges outlined above, we gathered a wide spectrum of survey data, including
27  track geometry, tonnage, line speed, ballast age, geology, and many other potential factors (e.g.,
28  type of ballast materials, track quality measurements and lineside vegetation) that may influence
29  track fixity. These factors were empirically identified following an interview with Network Rail's
30  engineering staff. The disparate data resources are of various data formats, and they could be
31  extracted and made available from different subsystems (e.g., Civil Asset Register and Reporting
32  System) and data models (e.g., Corporate Network Model) of Network Rail. They, together with
33  historical track positions and movements, were processed on a uniformed computing environment
34  (i.e., using Python programming language). By doing so, a comprehensive database was created,
35  allowing the analysis, restructuring, integration and cross-referencing of the data associated with
36  the influencing factors, thus enabling the extraction of spatial-temporal information that is most
37  relevant to track movement. On this basis, we can further construct a new data model specifically
38  for track movement, which incorporates suitable machine learning algorithms to predict gauging
39  issues with a quantifiable confidence level of their risk of occurrence. However, it needs to be
40  pointed out that in this study, not all the above-named data resources were thoroughly evaluated
41  for their compatibility and potentiality of being integrated into the same data set. Therefore, only
42  a selected set of factors and their data were considered for the development of a protype model.
43
44  **Data Integration**
45  The data utilized in the study were mainly from the following four different resources.

a) LiDAR data, which contains point cloud data showing the spatial position of every profile-specific railway track, represented as three-dimensional (3D) geographic coordinates, of rail heads – This data was available in LAS / LAZ file format (*12*). With data preprocessing, position data aligning with the rail head was made available for every roughly one-meter length. The positions are described by geographic coordinates measured in the OSGB36 (Ordnance Survey Great Britain 1936) reference system (*13*). Note that all the geographic coordinates utilized in this study are OSGB36 based.

b) Survey data that is related to the infrastructure that supports the rails – This data provides basic information such as curvature, cant, maximum allowable train speed and axle load, as well as the types of embankments, cuttings, rails, and sleepers used, for every varying length of meters of a railway track. It was available in CSV file format (*14*).

c) Data of structures, including overline and underline bridges, retaining walls, tunnels, and stations – To further clarify, an "overline bridge" refers to a bridge structure that spans over a railway line, while an "underline bridge" refers to a railway bridge that passes over a road. This data encompasses information about their locations, materials, structural forms, and construction details. It was available in DGN file format (*15*, *16*).

d) Data of a range of parameters associated with track geometry – This data comprises information on the layout and geographic locations of the railway tracks within the UK's rail network, as well as reference data that associates the tracks with different infrastructure. It was available in shapefile format (*17*).

Noticeably, these resources present distinctly different data and file formats. To integrate the data, we transform them into a unified format and visualize the preprocessed data through the use of the Python programming language (hereafter referred to as Python). All the preprocessed data were stored in a database managed by a PostgreSQL server. On this basis, an application programming interface (API) is established, as a prototype, for further data processing in a fully Python-supported computing environment. Utilizing this approach not only facilitates efficient storage, retrieval, and extraction of the most relevant information among all the available data, but also offers greater flexibility and extensibility in terms of software engineering for modelling and future development of an integrated computing framework, as compared to using commercial tools. With the prototype API in this pilot study, we shall be able to:

1) calculate the displacement of rail heads in terms of both rate and direction;
2) cross-reference the track fixity measures with data of any identified influencing factors (given their availability);
3) integrate these data in both spatial and temporal contexts to create a comprehensive data set; and
4) develop a prototype track fixity prediction model using an appropriate machine learning model and the data set.

To be more specific, the data set can be created at a specified level of resolution, such as 1 m, 10 m, or 100 m intervals, by matching the location and time across the different data resources. In this study, it relies entirely on open-source tools, including PyHelpers (*18*), PyRCS (*19*), PyDriosm (*20*), LAStools (*21*), Laspy (*22*), folium (*23*), and Open3D (*24*), all of which are under free licenses. The method of how to calculate the displacement is detailed and illustrated in the case-study example in the next section of the paper.

**Calculation of Track Movement**

Traditionally, the calculation of the rail displacements would necessarily entail data about on-site in-situ measurement; some researchers have also used digital images (*25, 26*). Inspired by (*27*)'s work, we propose in our study to use the LiDAR data to calculate the track movement. The LiDAR data provides comprehensive information on all visible objects within a certain area that scanners can reach. It generates a 'point cloud,' namely, a set of 3D points, which could be used to describe and represent the shape and relative spatial position of an object, given a reference system such as the OSGB36.

In this pilot study, the point cloud data extracted from the raw LiDAR data is a dense, ordered set of 3D Cartesian coordinates, representing spatial position of the external surface of the rail heads. Each point is in the format of (X, Y, Z), referring to Easting, Northing, and elevation, respectively. In view of the wheel-rail contact where the most significant impact on track movement is, the focus in this study is on the top surface and running edge of the rail heads. More specifically, with any two sets of the point cloud data – collected from scanning the same railway track at two different times, we shall be able to reproduce two trajectories – also referred to as curves or 'polylines' – of the mobile scanner moving along the track. Calculating the track movement is therefore essentially equivalent to measuring the relative displacement (or shift) of one polyline to another formed by the point clouds.

One way to approach this problem is to measure the similarity between the two polylines. In this regard, there are several methods for conducting the comparisons of two given polylines. One possible option is to compute Fréchet distance (*28, 29, 30*), which, however, usually describes the smallest of the maximum pairwise distances. Alternatively, one may consider the Pompeiu-Hausdorff distance (or Hausdorff distance) which, as suggested by (*31*), may not be adequate for measuring the polylines' similarities. Besides the notion of mathematical distance, one may also utilize statistical theory and consider Kolmogorov-Smirnov statistics (*32*), and weighted least squares (i.e. the sum of the absolute differences between observed values and expectations divided by the observed values). The investigation of such approaches is beyond the scope of this pilot study but may usefully be explored in the next phase of the study. In this study, we propose a more straightforward method, which can have two options:

    a) to calculate the distance between a point (of a later observed data set) and its nearest line segment formed of two adjacent points (of an earlier observed data set); or

    b) to calculate the distance between a line segment formed of two adjacent points (of a later observed data set) and its corresponding one (of an earlier observed data set).

The line segment can be of arbitrary length (e.g., 1 m, 10 m, or 100 m) given the granularity of the point cloud data provided. To minimize computing errors, we consider the minimum distance between any two points as a unit line segment. On this basis, we shall be able to calculate the average and/or moving average of track movement at equal distances, or that at any greater distances by aggregating the displacements of their unit line segments. In this study, the unit line segment is approximately 1 m. After an interview with experienced railway track engineers at Network Rail, we adopt the method b) in this pilot study, as illustrated in **Figure 1**.
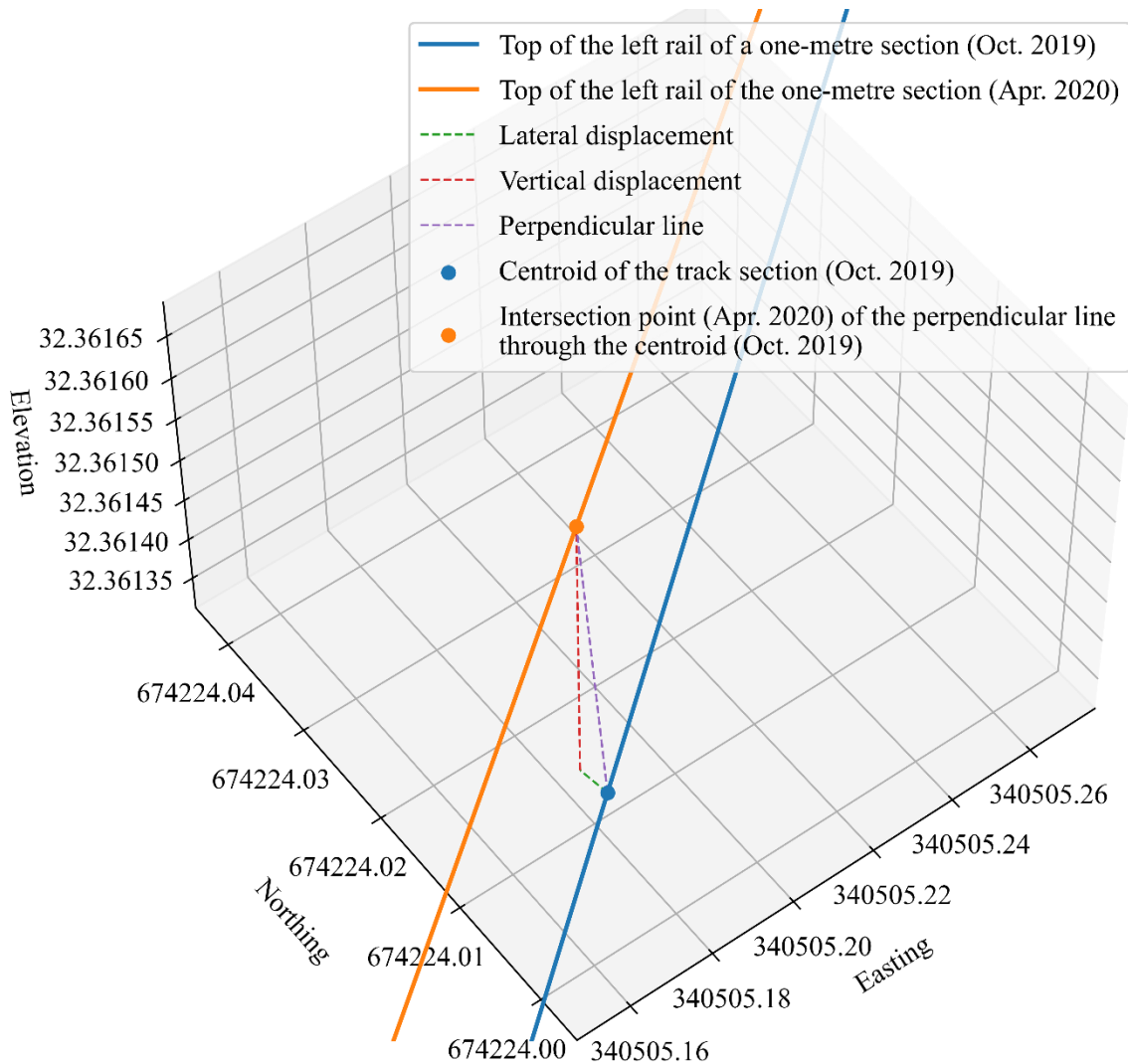
1
2 **FIGURE 1: Illustration of the displacement (movement) of the top surface of a one-meter**
3 **length of rail head.**
4
5 **Figure 1** shows an example of calculating the movement of the top surface of a rail head, with the
6 blue line representing a 1 m rail head measured in October 2019 and the orange, nearest line
7 segment, measured in April 2020. Taking a point on the blue line – empirically, the centroid – we
8 draw and compute the length of a perpendicular line segment (in the dotted purple in **Figure 1**)
9 from that point to the orange line, and hence obtain the displacement of the blue line from October
10 2019 to April 2020. On this basis, the lateral displacement (dotted green) and vertical displacement
11 (dotted red) of the top surface can be easily computed.
12      It must be noted that a) the line segments of the same rail head observed/measured at
13 different times would not be necessarily parallel with each other; b) the two ends of any line
14 segment would not necessarily correspond to the same ones in the different observation times; and
15 c) the two line segments (e.g., the blue and orange lines in **Figure 1**) may not be of exactly the
16 same length, due to measuring errors. For these reasons, using the above proposed method would
17 entail offsetting the errors by cutting out a section of the same length for each line segment in
18 either the horizontal or the vertical plane. That way we could directly connect the two centroids

each located on a cut-out section (slightly shorter than 1 m) to obtain a perpendicular line between the two segments.

Besides the method illustrated in **Figure 1**, there can be different alternatives. For example, one could also project either an end or the centroid of a line segment onto the other line in the space to obtain a perpendicular line, whereby calculations of the lateral and vertical displacements could also be obtained. However, this method is likely to result in more errors when many pairs of line segments are highly mismatched in terms of the spatial position, in which case the cut-out section can be shorter than 0.5 m. In this paper, we consider only the method illustrated in **Figure 1**.

**Predicting Track Movements**

The data integration and further data processing generates a comprehensive data set that contains both:

1) calculated results of average displacements for every 1 m track section and for a given period (i.e., six months), and
2) specific information of the associated factors affecting the track fixity, translated into numerical forms.

With the data of calculated track movements for every 1 m, the track fixity is measured as the direction and average rate of movement of the rail head for any track section length 1 m and above. As part of the computing framework discussed in this study, we propose to further integrate it with an appropriate machine learning model (e.g., a random forest model) to establish a functional link between the track fixity and those influencing factors for analyzing the movement of the rail heads, which should be capable of assisting track engineers in the analysis of track movement over time.

Given the immediate availability of data in this pilot study, we consider the following variables as some major influencing factors:

a) curvature and cant of a railway track;
b) presence of structures, including overline bridges, underline bridges, tunnels, stations and retaining walls; and
c) axle load (*33*) and speed of trains (*34*).
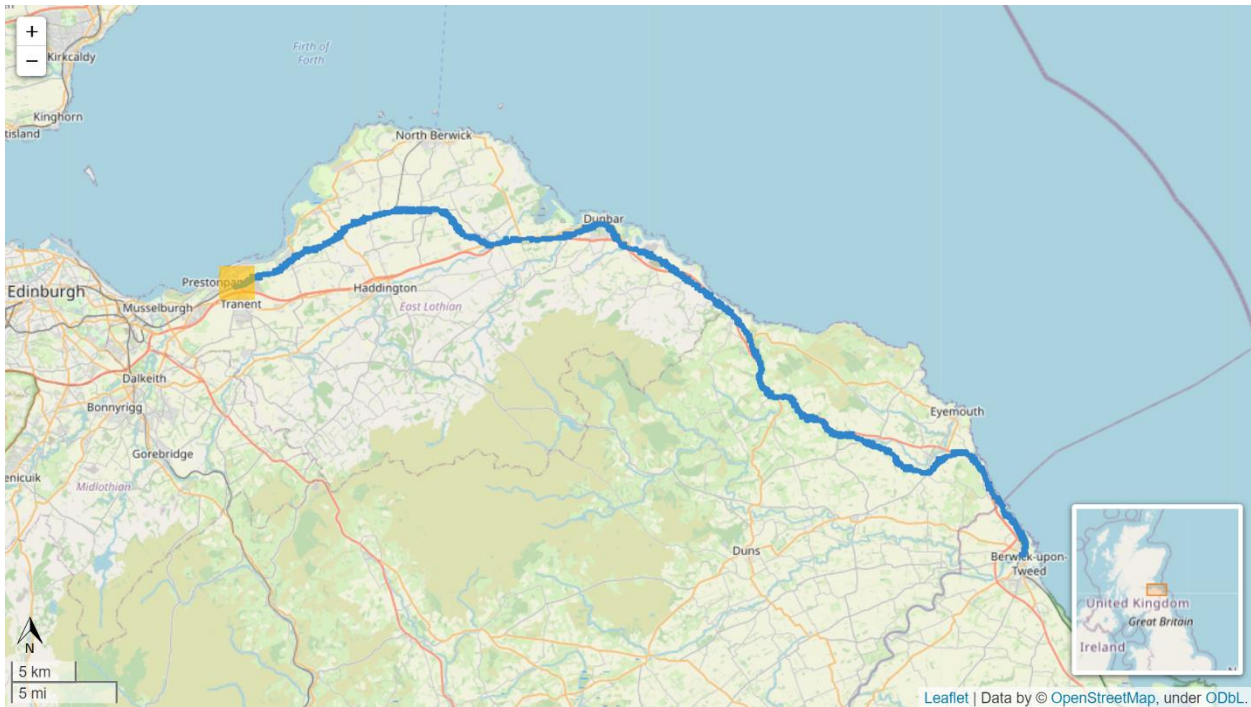
**A CASE-STUDY EXAMPLE**

This section demonstrates the methodology proposed in the previous section with a case-study example in the context of the UK's railway system.

**Case-study Region**

To develop and demonstrate our proposed computing framework, we selected an approximately 80-km section of railway track along the East Coast Main Line in the UK. This section of track, as highlighted in blue in **Figure 2(a)**, passes through four stations located between Prestonpans and Berwick-upon-Tweed, southeast of Edinburgh in Scotland, and is primarily used for passenger rail, though there may also be some freight train traffic. In this study, we refer to this section of track as the "example rail line" (or "example track"), which we treat as a representative of the entire network. The area highlighted in yellow in **Figure 2(a)** is magnified in **Figure 2(b)**.
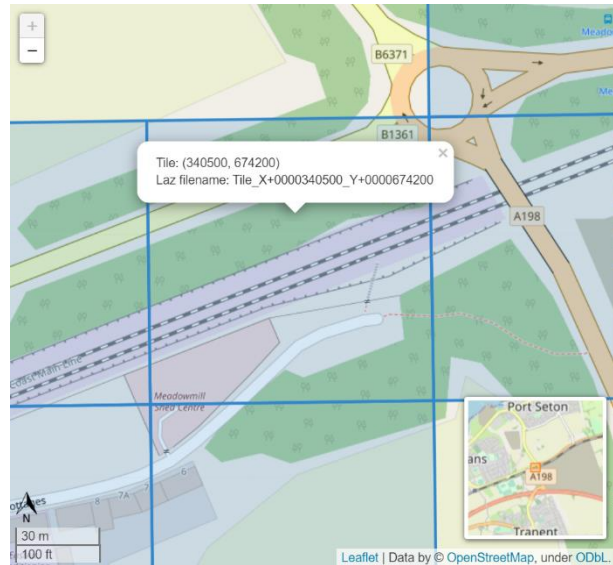
1



**(a)** An 80-km section of the East Coast Main Line for case study
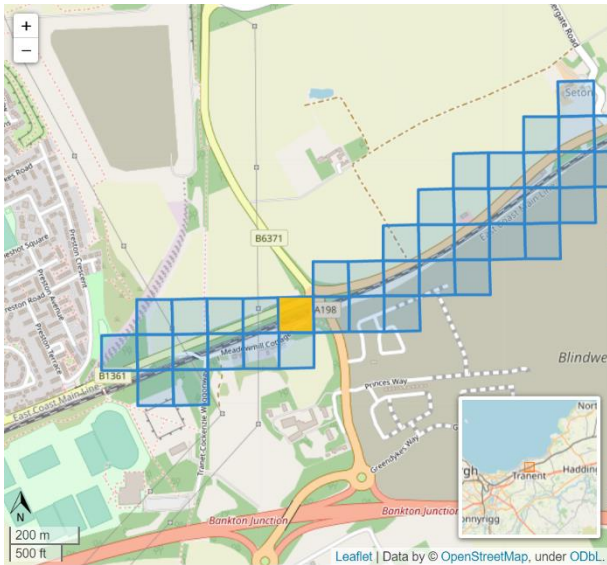


**(b)** The west end of the track section



**(c)** An example 100m×100m survey grid

2 **FIGURE 2: Illustration of an 80-km section of the East Coast Main Line and its survey**
3 **grids for case study.**
4
5 The point cloud data for the example rail line was available for two survey periods: October 2019
6 and April 2020. The data was collected from about 2,000 survey grids, each of which was a
7 100m×100m area measured in the OSGB36 reference system, as illustrated in **Figure 2(b)**. In this
8 study, we randomly selected a survey grid highlighted in yellow in **Figure 2(b)** and magnified it
9 in **Figure 2(c)**. The example grid was originally labelled "Tile_X+0000340500_Y+0000674200",

1     indicating that its lower left corner is located at the OSGB36 coordinates (340500, 674200). We
2     used this example grid to further demonstrate the framework and methodology for calculating and
3     predicting track movement.
4

5     **Track Movement**
6     The raw LiDAR data from the example grid, Tile_X+0000340500_Y+0000674200, for the two
7     survey periods, October 2019 and April 2020, are shown in **Figure 3(a)** and **Figure 3(b)**,
8     respectively. The colors in the figures represent different elevations of objects, where warmer
9     colors (e.g., red, orange, and yellow) indicate relatively higher elevations, while cooler colors (e.g.,
10     green, blue, and purple) indicate relatively lower elevations. According to **Figure 2(c)**, it can be
11     inferred that the blue lines in both **Figure 3(a)** and **Figure 3(b)** represent the railway tracks; the
12     warmer colors surrounding the tracks represent mostly lineside vegetation, which notably had
13     grown considerably higher from since October 2019.
14     Further, **Figure 3(c)** and **Figure 3(d)** illustrate the point cloud data of the rail heads in
15     **Figure 3(a)** and **Figure 3(b)**, respectively. We linked every two adjacent points in sequential order
16     to create a polyline for each of the elements, including top surfaces, running edges and the center
17     of a track, as illustrated in **Figure 3(e)** and **Figure 3(f)**. On this basis, we could use the method
18     described in the methodology section to calculate the displacements of each of the polylines.
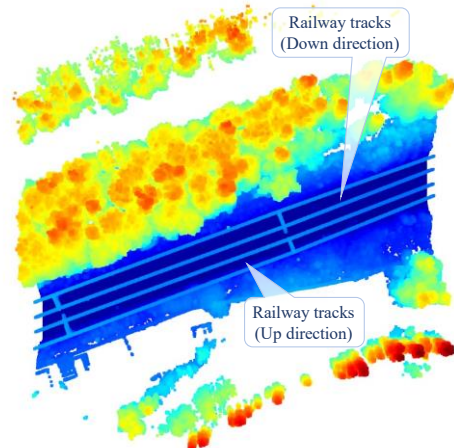19     For example, consider the top surface of the rail head of the left rail of the track (hereafter
20     referred to as "left rail top") in the Up direction where all trains run towards Edinburgh. Note that
21     in the UK, the terms "Up" and "Down" are conventionally used to indicate, respectively, the
22     directions of trains running towards and away from a major destination, such as Edinburgh and
23     London. This predefined reference system is used across the entire UK rail network and provides
24     a convenient and straightforward way for on-site railway staff to conduct inspections and
25     maintenance work. In our case-study example, Edinburgh is identified as a major destination, and
26     the direction in which trains run towards it is thus referred to as the Up direction.
27     **Figure 4** shows violin plots for the calculated movements of the left rail top of the example
28     80-km track in the Up direction. The plots depict the probability density and boxplot information
29     of average displacements in both lateral and vertical planes for every 10-m track section between
30     October 2019 and April 2020. The direction of track movement is indicated by a positive (+) or
31     negative (−) sign, where a positive sign (+) denotes that the track had moved towards the left in
32     the Up direction since October 2019, and a negative sign (−) denotes movement to the right.
33     As illustrated in **Figure 4**, the lateral track movement exhibited an average displacement
34     of less than 5 mm per 10-m track section, with the maximum displacement of nearly 20 mm
35     observed in the rightward movement (i.e., movement of the left rail top towards the center of the
36     track). Notably, there was a greater range of values for leftward movement (i.e., movement of the
37     left rail top away from the center of the track), which may indicate a need for further investigation
38     into the underlying factors contributing to this variation. On the vertical track movement, however,
39     the 10-m average displacement was mostly around 2 mm or less; and a few extreme values ranging
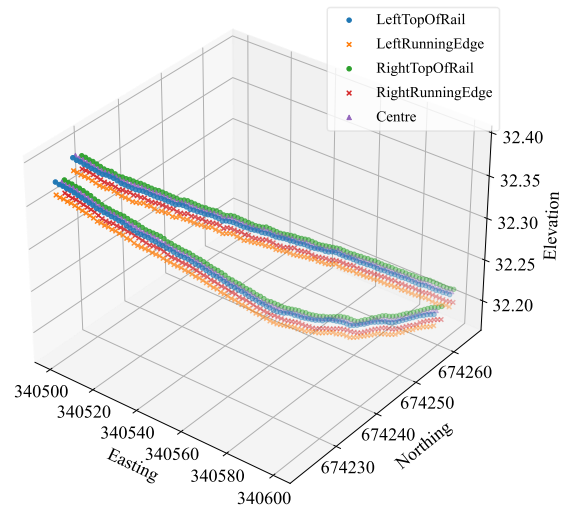40     from 5 to 15 mm were also observed.
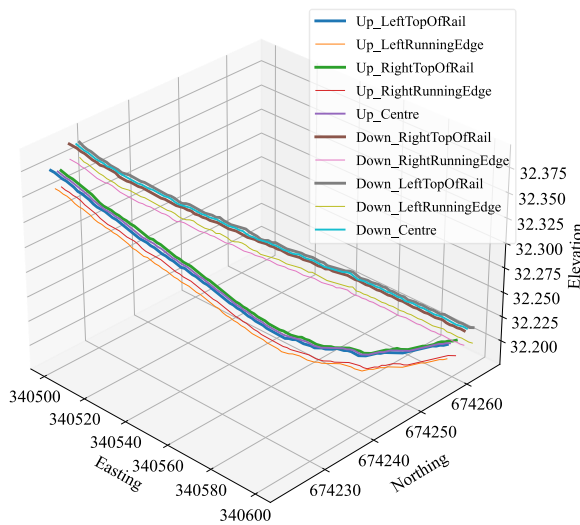41
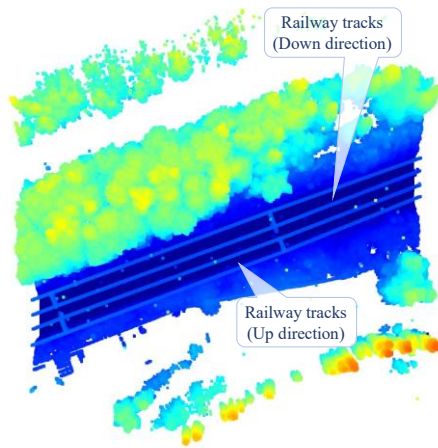
**(a)** Raw LiDAR data (October 2019)　　　**(b)** Raw LiDAR data (April 2020)
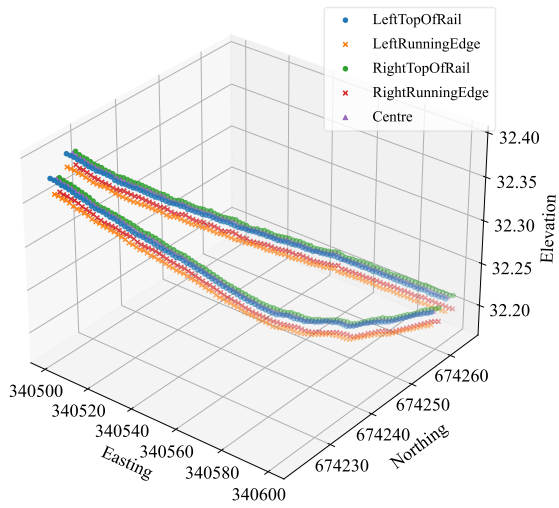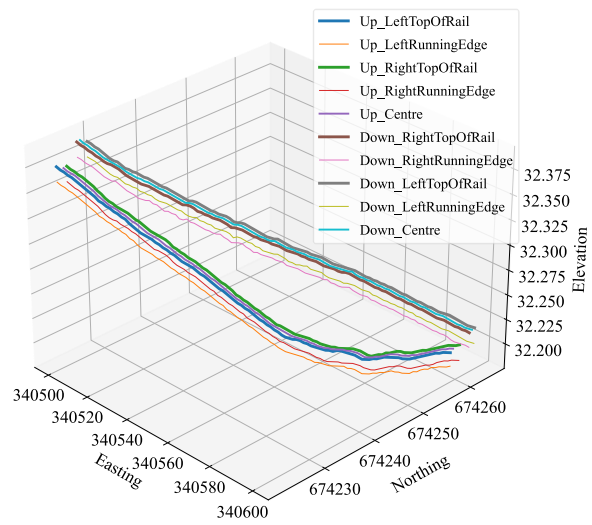
**(c)** Point cloud of the rail heads (October 2019)　　**(d)** Point cloud of the rail heads (April 2020)
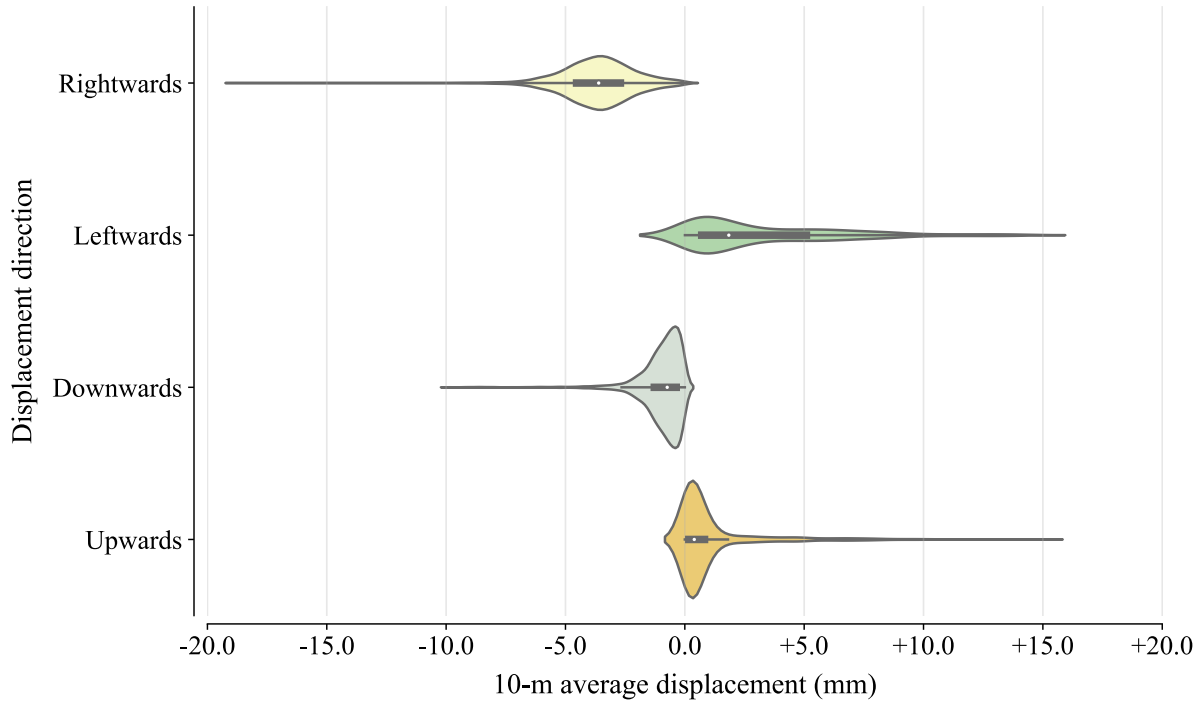
**(e)** Polyline based on the point cloud (October 2019)　**(f)** Polyline based on the point cloud (April 2020)

**FIGURE 3: Representation of the rail heads based on their point cloud data within the example 100m×100m grid (340500, 674200).**

**FIGURE 4: Violin plot of the average lateral and vertical track movements for every 10-m section of the example track.**

Based on the calculation results, **Figure 5(a)** and **Figure** 5**(b)** show the hotspots of significant lateral and vertical displacements of the left rail top, respectively. The color scale ranges from cooler colors like green, indicating relatively lower rates of movement, to warmer colors like yellow for moderate rates and red for higher rates. It needs to be noted that these heatmaps are based only on the absolute values of the calculated track movements.

Further to the calculation of the track fixity parameters, we proceeded to integrate all the available data of the several selected factors influencing the track fixity to create a comprehensive data set for developing a prototype machine learning model capable of analyzing the movement of the rail heads.
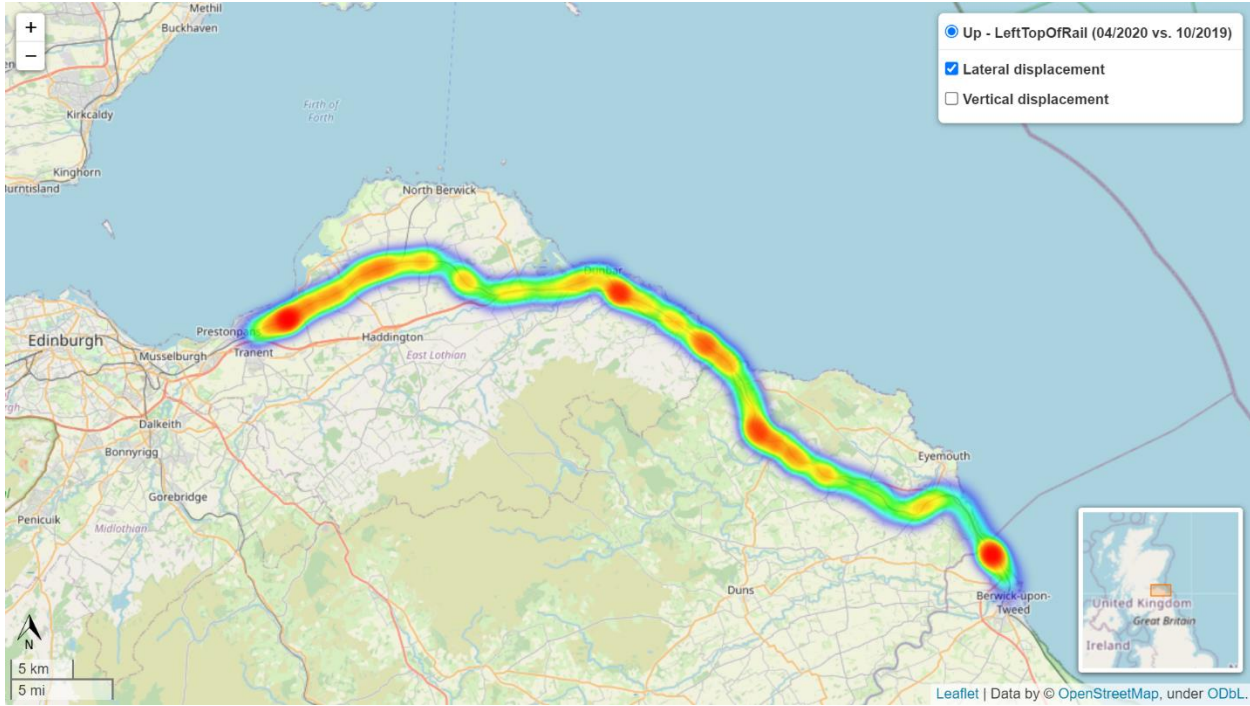
**(a)** Lateral track movement
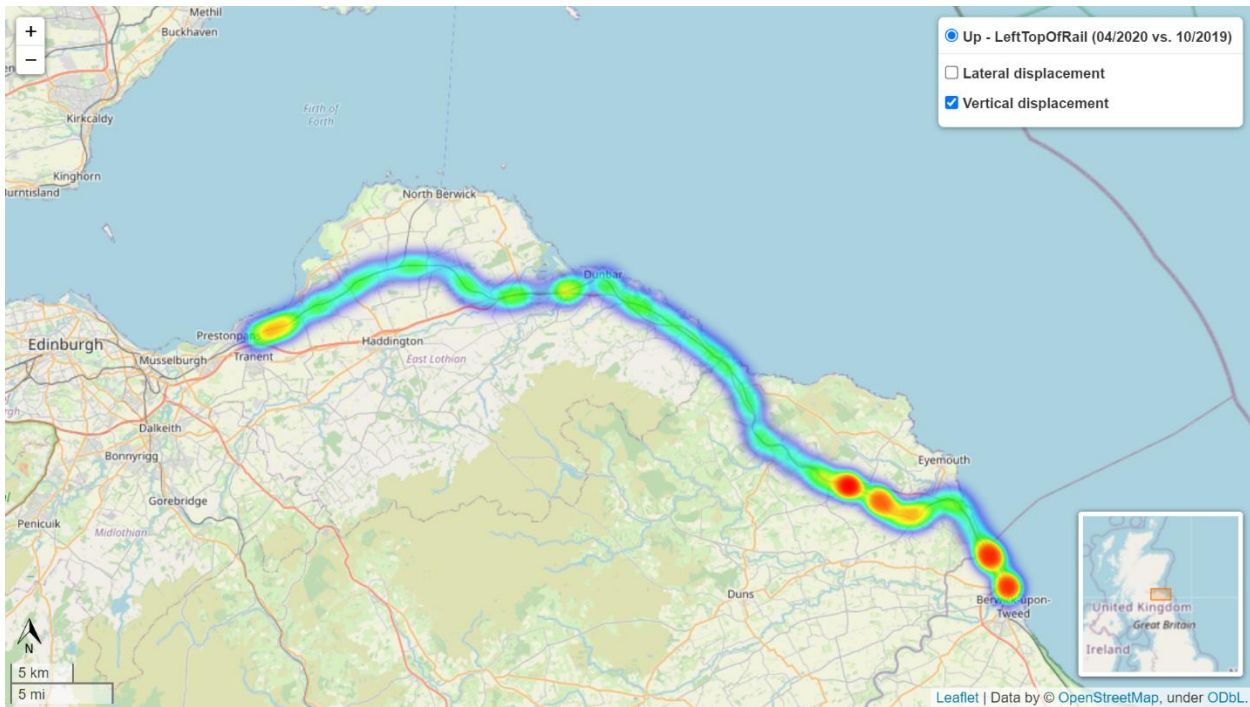


**(b)** Vertical track movement

**FIGURE 5: Average track movement (mm per 10-meter length) of the top surface of rail head of the left rail in the Up direction (October 2019 vs. April 2020).**

1  **Prototype Predictive Model for the Region**
2  This subsection describes how we could develop a prototype model, which is capable of predicting
3  the track movement under specific conditions, drawing on a comprehensive data set of recorded
4  asset data, local operational factors, and a knowledge of the track design.
5
6  *Influencing Factors and Modelling*
7  For the data integration of influencing factors, we created a series of contiguous circular "buffer
8  zones" overlaying the example track, as illustrated in **Figure 6**. Each of the buffer zones represents
9  a virtual area surrounding a specific section of the track, where we calculate the track movement
10  and collect data of any factors that may influence the track fixity. The diameter of the buffer zone
11  is equal to the length of the track section (or rather, the rail head), for which average track
12  movement is calculated. This approach allows us to gather information on influencing factors and
13  associate them with the track movement within the same designated buffer zone. For instance, in
14  **Figure 6**, the green dots indicate the presence of overline bridges within each of the buffer zones.
15  It should be noted that the buffer zones in **Figure 6** have a diameter of 1 km, which is used for
16  demonstration purposes only.
17
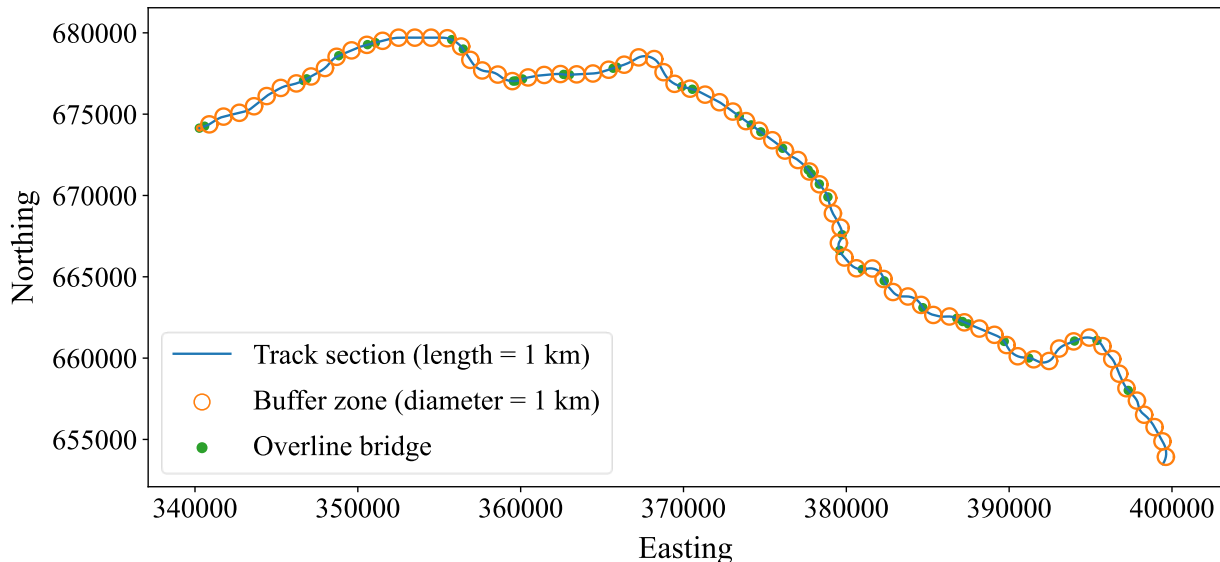


18
19  **FIGURE 6: Illustration of buffer zones and the presence of overline bridges on the example**
20  **rail line.**
21
22  As mentioned in the methodology, our modelling trial in this case-study example was limited to
23  nine factors, including curvature, cant, maximum allowable train axle load and speed, as well as
24  the presence of overline bridges, underline bridges, retaining walls, tunnels, and stations. To create
25  a prototype predicative tool for the track movement in this study, we trialed, a random forest (RF)
26  model (*35*), considering only the nine factors.
27      The RF model is one of the most popular machine learning methods used in many
28  applications (*36*). We chose RF as a starting point from among various machine learning models
29  due to its robustness and interpretability. In comparison to other models, the RF is less susceptible
30  to overfitting and is capable of handling irrelevant factors in the data. Moreover, the output of an
31  RF model can offer valuable insights into which factors are most important for making predictions.
32  In essence, an RF model is an ensemble learning method that combines a set of decision tree

1    models (hereafter referred to as "trees"), each of which may have insufficient individual
2    competence in using the same data for making predictions (*37*). More specifically, an RF model
3    evaluates all the predictions made independently by its component trees and provides a
4    comprehensive prediction result (*38*).
5         Take for example the lateral displacement of the left rail top. Following consultation with
6    Network Rail engineers, we categorized the displacement into five ranges, including "$\leq -4.45$
7    mm", "($-4.45$ mm, $-3.5$ mm]", "($-3.5$ mm, $-2.5$ mm]", "($-2.5$ mm, $0.0$ mm]" and "$> 0.0$ mm".
8    Note again that negative sign ($-$) indicates that the left rail top moves rightwards in the Up direction,
9    and leftwards otherwise. The predicted class for a given case should correspond to the class with
10   the highest probability across all the decision trees in the RF model. In this case-study example, a
11   total of 6,792 valid cases of 10-m track movements were obtained; we shuffled and divided the
12   data set into a training set of 5,433 cases and a test set of 1,359 cases. We trained a commonly
13   used RF classifier on the training set using a Python package, scikit-learn (*39*), considering
14   different numbers of individual trees (i.e., 50, 150, 200, 250, 300, and 350) and different maximum
15   allowed depths of the trees (i.e., 5, 10, 15, 20, 25, and 30). Through an exhaustive search from
16   different combinations of the two sets of values, we identified the best RF model on the training
17   set, which was formed of 300 decision trees, each with a maximum allowed depth of 15. This
18   model was based on 5-fold cross validation.
19
20   *Results*
21   The results from the trained RF model presented valuable evidence of the relative importance of
22   the identified factors on predicting track fixity, which could be used to better understand how much
23   impact these factors would have on the track fixity and hence the prediction of future track
24   movement.
25        To get the overall performance of the RF model in predicting the lateral track movement,
26   **Figure 7** shows a confusion matrix on a set of test data. It illustrates a comparison between the
27   predicted values of the model with the actual values of the test data. For instance, the model
28   predicted that the lateral displacement of 255 10-m left-rail-top section fell within the range of
29   ($-4.45$ mm, $-3.5$ mm], which was consistent with their corresponding calculated values. Due to
30   limited data availability, the model's absolute accuracy is only around 50% overall. However, the
31   confusion matrix shows that most cases are centered around the diagonal, indicating that the
32   model's predictions of lateral displacements are largely consistent with the calculated track
33   movements. The output of the trained RF model demonstrates good predictive capability and
34   shows that the proposed computing framework has enormous potential as a tool for predicting and
35   further exploring the sensitivity of the factors influencing track fixity.

1
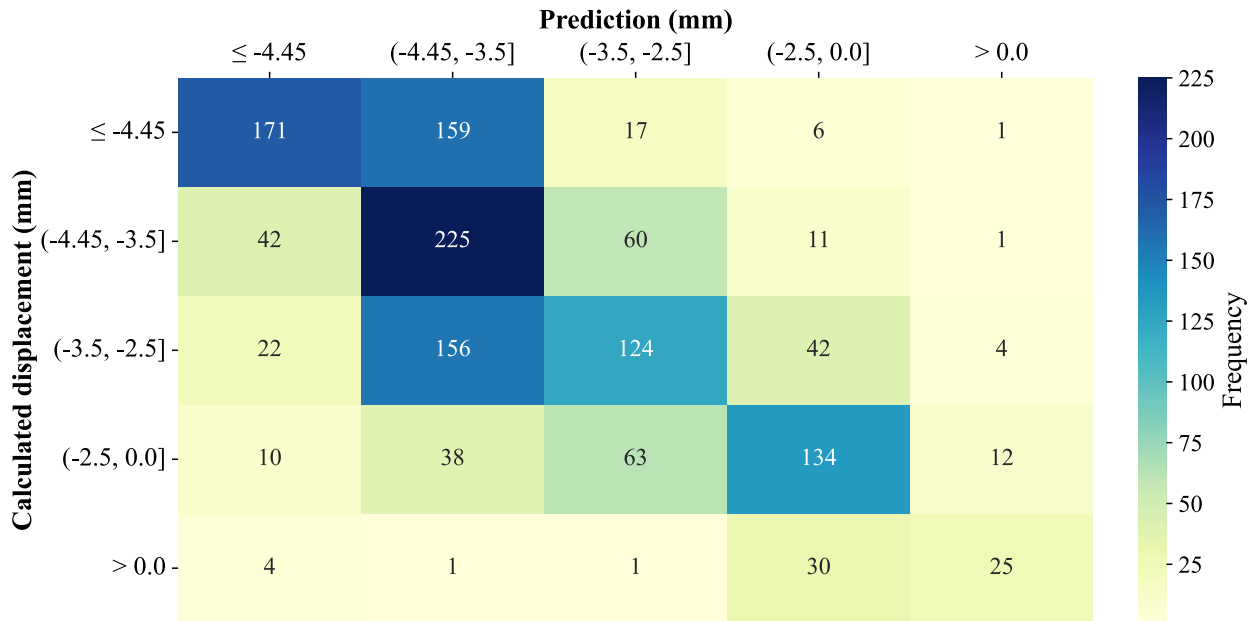


FIGURE 7: Confusion matrix on the test data set.

The importance of the different influencing factors (relative to all the others) being considered in the RF model is presented in **Table 1**, where the values associated with the different factors sum to one. The table is sorted in the descending order of the importance values; the higher the value, the more significant the impact that the corresponding factor can exert on the track fixity in the lateral plane.

TABLE 1: Relative importance of factors influencing the track fixity.

| Factor | Relative importance |
|---|---|
| Curvature | 0.3887 |
| Cant | 0.3725 |
| Maximum speed | 0.2016 |
| Presence of underline bridges | 0.0095 |
| Presence of overline bridges | 0.0073 |
| Maximum axle load | 0.0067 |
| Presence of retaining walls | 0.0060 |
| Presence of tunnels | 0.0058 |
| Presence of stations | 0.0019 |

As expected, curvature and cant proved to be the most important among all that were considered in the model. Besides the track geometry, axle load and train speed would also be expected to significantly impact track fixity (*33*). In terms of the presence of structures, track fixity of ballasted track can be more vulnerable to movement than fixed structures such as retaining walls and tunnels that may offer a greater degree of track bed stability. Track sections within station areas are much less likely to suffer from fixity issues given much slower train speeds and lack of track curvature.

16

1   In summary, the integrated computing framework in the current phase of this study allows for high
2   flexibility for further development with more, adequate data across the railway network. There
3   remains, however, a need for trialing and validating the data model in different areas of the railway
4   network to help further refine the methodology.
5
6
7   **CONCLUSION AND DISCUSSIONS**
8   This paper presents a pilot study that seeks to design the most comprehensive integrated computing
9   framework to date for track fixity in the context of the UK's railway system, allowing for a step
10  change in temporal and spatial resolution of understanding profile-specific track fixity. Although
11  it is currently in the prototype stage, the design of the data flow pipeline enables it to gather,
12  process and combine as much relevant information as possible to track fixity. With the established
13  workflow, we propose a new metric for assigning track fixity values to a given track profile in
14  terms of its movement relative to the plane of rail within a given period. A random forest model
15  was trialed to predict future movement of the track. The prototype framework has demonstrated
16  its capability of predicting track movement with an engineering acceptable confidence level, with
17  most movements classified either in the predefined ranges or within a single bin width of the true
18  value (based on calculation with point cloud data). There is thus a potential of using the framework
19  to explore the sensitivity of track fixity to the factors affecting it and calculate future track fixity
20  for new structures. Also, the prediction model developed from this research is highly adaptable in
21  different contexts, given the availability of similar data resources of the factors that the model
22  accounts for.
23          However, it is recognized there are three main limitations in this pilot study. Firstly, the
24  calculation of track movement was based only on 3D geographical coordinates of rail head position,
25  which were made available in the form of point cloud data. However, errors generated from data
26  collection, as well as that in the data collected at different times, are unavoidable, notwithstanding
27  the same geographical coordinate system. While the actual track movement is mostly within a few
28  millimeters, it can possibly be even smaller than the errors in the measurement at the same point.
29  Arguably, on the one hand, the errors can depend largely on the technology used for collecting the
30  data; on the other hand, it also needs to be noted that the development of the prediction model was
31  not intended for predicting the precise movement but the movement within a predefined set of
32  displacement ranges, into which the movement would be most likely to fall. Whether the errors in
33  the measurements could be fully contained within the predefined range would be worth further
34  investigation for a specific data collection technology. Secondly, the data integration for the model
35  development relies heavily on mapping heterogeneous information (about the factors influencing
36  track movement) onto a same geographical system. Due to a deficiency of accurate location
37  referencing data, it was not possible to consider all influencing factors in the prototype model.
38  Thirdly, data was only available from two time periods (October 2019 and April 2020). Therefore,
39  to further the development of the framework, the following issues should be addressed:
40      1)  Improving the quality of reference data across various location identifiers in different data
41          resources.
42      2)  The comprehensive data set should be extended to include additional line sections with
43          differing reference curves and speed profiles. In this way, it will become possible to test
44          the potential of the model in a more general context, and to begin to a) determine what and
45          how many prediction models would need to be trained to obtain national coverage, and b)

investigate the trade-off between the type and number of models and the individual discriminatory power of each instance.

3) Using data from additional measurement campaigns to enable the existing model to be further developed, leading to improved accuracy and greater confidence in the results produced by the model.

4) To look ahead in a longer term, would require a more harmonized and unified data codification system across the rail industry to accelerate the further development of the proposed framework and the implementation of a full-fledged, automated computing platform to be integrated into the railway track system.

## ACKNOWLEDGEMENT

## AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: Q. Fu, J. M. Easton, M. P. N. Burrow, and J. Sweeney; data collection: J. M. Easton, M. P. N. Burrow, and J. Sweeney; analysis and interpretation of results: Q. Fu; draft manuscript preparation: Q. Fu and J. M. Easton. All authors reviewed the results and approved the final version of the manuscript.

## REFERENCES

1. Weston, P., C. Roberts, G. Yeo, and E. Stewart, Perspectives on railway track geometry condition monitoring from in-service railway vehicles. Vehicle System Dynamics, Vol. 53, No. 7, 2015, pp. 1063-1091. DOI: 10.1080/00423114.2015.1034730.
2. Brough, M.J., G.S. Ghataora, A.B. Stirling, K.B. Madelin, C.D.F. Rogers, and D.N. Chapman, Investigation of railway track subgrade. I: In-situ assessment. Proceedings of the Institution of Civil Engineers - Transport, Vol. 156, No. 3, 2003, pp. 145-154. DOI: 10.1680/tran.2003.156.3.145.
3. Brough, M.J., G.S. Ghataora, A.B. Stirling, K.B. Madelin, C.D.F. Rogers, and D.N. Chapman, Investigation of railway track subgrade. Part 2: Case study. Proceedings of the Institution of Civil Engineers - Transport, Vol. 159, No. 2, 2006, pp. 83-92. DOI: 10.1680/tran.2006.159.2.83.
4. Burrow, M., P.F. Texeira, T. Dahlberg, and E. Berggren, Track stiffness considerations for high speed railway lines, in Railway Transportations: Policies, Technology and Perspectives, N.P. Scott, Editor. 2009, Nova Science Publishers: Hauppauge, NY, USA.
5. Kish, A., G. Samavedam, and D. Wormley, Fundamentals of track lateral shift for high-speed rail applications, Federal Railroad Administration, USA, 1998. Available from: https://rosap.ntl.bts.gov/view/dot/8515 (accessed: March 2023).

6.  Kish, A. and W. Mui, Track buckling research, John A. Volpe National Transportation Systems Center, USA, 9 July 2003. Available from: https://rosap.ntl.bts.gov/view/dot/11985 (accessed: March 2023).

7.  Mott MacDonald, RSSB 1386 (Revised) The effects of railway traffic on embankment stability - Final report, Rail Safety and Standards Board (RSSB), UK, March 2011. Available from: https://www.networkrail.co.uk/wp-content/uploads/2017/02/RSSB-report-The-effects-of-railway-traffic-on-embankment-stability-1.pdf (accessed: March 2023).

8.  Damme, P., M. Birkenbach, C. Bitsakos, M. Boehm, P. Bonnet, F. Ciorba, M. Dokter, P. Dowgiallo, A. Eleliemy, C. Faerber, G. Goumas, D. Habich, N. Hedam, M. Hofer, W. Huang, K. Innerebner, V. Karakostas, R. Kern, T. Kosar, A. Krause, D. Krems, A. Laber, W. Lehner, E. Mier, M. Paradies, B. Peischl, G. Poerwawinata, S. Psomadakis, T. Rabl, P. Ratuszniak, P. Silva, N. Skuppin, A. Starzacher, B. Steinwender, I. Tolovski, P. Tözün, W. Ulatowski, Y. Wang, I. Wrosz, A. Zamuda, C. Zhang, and X. Zhu, DAPHNE: An open and extensible system infrastructure for integrated data analysis pipelines, in Conference on Innovative Data Systems Research. 2022: Santa Cruz, California, USA.

9.  Young, M.A., S.P. Blainey, T. Gowland, and S. Nagella, An automated online tool to forecast demand for new railway stations and analyse potential abstraction effects. Transport Practitioners' Meeting, July 10–11, 2019. 2019, Oxford.

10. Hall, J., UK reveals new platform for infrastructure data analysis and simulation modelling. Proceedings of the Institution of Civil Engineers - Civil Engineering, Vol. 172, No. 3, 2019, pp. 102-102. DOI: 10.1680/jcien.2019.172.3.102.

11. Smarzaro, R., C.A. Davis, and J.A. Quintanilha, Creation of a multimodal urban transportation network through spatial data integration from authoritative and crowdsourced data. ISPRS International Journal of Geo-Information, Vol. 10, No. 7, 2021, p. 470. DOI: 10.3390/ijgi10070470.

12. ASPRS, LAS Specification Version 1.4 - R14, American Society for Photogrammetry and Remote Sensing (ASPRS), Bethesda, MD, USA, 26 March 2019. Available from: https://github.com/ASPRSorg/LAS (accessed: March 2023).

13. Ordnance Survey, A guide to coordinate systems in Great Britain, Ordnance Survey, UK, 2020. Available from: https://www.ordnancesurvey.co.uk/documents/resources/guide-coordinate-systems-great-britain.pdf (accessed: March 2023).

14. Shafranovich, Y., Common format and MIME type for comma-separated values (CSV) files. No. RFC 4180. Internet Engineering Task Force (IETF), 2005. Available from: https://datatracker.ietf.org/doc/html/rfc4180#page-1 (accessed: 12 March 2023).

15. Uchôa, H.N., M.C.M. De Paulo, L.C.T.C. Filho, and P.R. Ferreira, Evaluation of data conversion of vectorial geographic features in topographic maps using free software tools, in Workshop de Software Livre. 2006: Porto Alegre, Brasil.

16. Warmerdam, F., DGNLib: A Microstation DGN (ISFF) Reader. http://dgnlib.maptools.org/. Accessed March, 2023.

17. ESRI, Shapefile technical description: An ESRI white paper, Environmental Systems Research Institute (ESRI), Inc., Redlands, CA, USA, 1998. Available from: https://www.esri.com/content/dam/esrisites/sitecore-archive/Files/Pdfs/library/whitepapers/pdfs/shapefile.pdf (accessed: March 2023).

18. Fu, Q., PyHelpers: an open-source toolkit for facilitating Python users' data manipulation tasks (Version 1.4.6). 2023, Zenodo. DOI: 10.5281/zenodo.7679196. Available from: https://github.com/mikeqfu/pyhelpers.

19. Fu, Q., PyRCS: an open-source tool for collecting railway codes used in different UK rail industry systems (Version 0.3.5). 2023, Zenodo. DOI: 10.5281/zenodo.7679247. Available from: https://github.com/mikeqfu/pyrcs.

20. Fu, Q., PyDriosm: an open-source tool for downloading, reading and PostgreSQL-based I/O of OpenStreetMap data (Version 2.1.2). 2023, Zenodo. DOI: 10.5281/zenodo.7681653. Available from: https://github.com/mikeqfu/pydriosm.

21. Isenburg, M., LAStools - Efficient LiDAR processing software (Version 181001 academic). 2019, rapidlasso GmbH: Gilching, Germany. Available from: https://rapidlasso.com/lastools/.

22. Brown, G., T. Montaigu, and contributors, Laspy: Native Python ASPRS LAS read/write library (Version 2.4.1). 2023, Python Package Index (PyPI). Available from: https://pypi.org/project/laspy/2.4.1/.

23. Filipe, Frank, R. Story, J. Gardiner, M. Journois, H. Rump, A. Bird, A. Lima, J. Cano, dbf, M. Oefelein, J. Leonel, J. Baker, T. Sampson, J. Reades, B. Welsh, Q. Kong, O. Komarov, A. Crosby, G. Harris, R. Dumas, D. KRIEF, D. Kato, D. Mendoza, N. Wilson, T.P. Nogueira, kenmatsu4, L. Furtado, A. Patil, and J. Duke, python-visualization/folium: v0.13.0 (Version 0.13.0). 2022, Zenodo. DOI: 10.5281/zenodo.7158711. Available from: https://pypi.org/project/folium/0.13.0/.

24. Zhou, Q.-Y., J. Park, and V. Koltun, Open3D: A modern library for 3D data processing. arXiv, 2018. DOI: 10.48550/arXiv.1801.09847.

25. Iryani, L., H. Setiawan, T. Dirgantara, and I.S. Putra, Development of a railway track displacement monitoring by using digital image correlation technique. Applied Mechanics and Materials, Vol. 548-549, 2014, pp. 683-687. DOI: 10.4028/www.scientific.net/AMM.548-549.683.

26. Murray, C.A., W.A. Take, and N.A. Hoult, Measurement of vertical and longitudinal rail displacements using digital image correlation. Canadian Geotechnical Journal, Vol. 52, No. 2, 2015, pp. 141-155. DOI: 10.1139/cgj-2013-0403.

27. Ye, C., S. Acikgoz, S. Pendrigh, E. Riley, and M.J. DeJong, Mapping deformations and inferring movements of masonry arch bridges using point cloud data. Engineering Structures, Vol. 173, 2018, pp. 530-545. DOI: 10.1016/j.engstruct.2018.06.094.

28. Heiter, T. and H. Manila, Computing discrete Fréchet distance, Christian Doppler Laboratory for Expert Systems, Technical University Vienna, Vienna, Austria, Technical Report CD-TR 94/64, 25 April 1994. Available from: http://www.kr.tuwien.ac.at/staff/eiter/et-archive/cdtr9464.pdf.

29. Alt, H. and M. Godau, Computing the Fréchet distance between two polygonal curves. International Journal of Computational Geometry & Applications, Vol. 5, No. 01n02, 1995, pp. 75-91. DOI: 10.1142/S0218195995000064.

30. Buchin, K., M. Buchin, and C. Wenk, Computing the Fréchet distance between simple polygons. Computational Geometry, Vol. 41, No. 1-2, 2008, pp. 2-20. DOI: 10.1016/j.comgeo.2007.08.003.

31. Aronov, B., S. Har-Peled, C. Knauer, Y. Wang, and C. Wenk, Fréchet distance for curves, revisited, in European Symposium on Algorithms (ESA), Y. Azar and T. Erlebach, Editors. 2006, Springer: Berlin, Heidelberg. p. 52-63.

32. Fasano, G. and A. Franceschini, A multidimensional version of the Kolmogorov–Smirnov test. Monthly Notices of the Royal Astronomical Society, Vol. 225, No. 1, 1987, pp. 155-170. DOI: 10.1093/mnras/225.1.155.

33. Burrow, M.P.N., D. Bowness, and G.S. Ghataora, A comparison of railway track foundation design methods. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, Vol. 221, No. 1, 2007, pp. 1-12. DOI: 10.1243/09544097jrrt58.

34. Heelis, M.E., A.C. Collop, A.R. Dawson, D.R. Chapman, and V.V. Krylov, Predicting and measuring vertical track displacements on soft subgrades, in Proceedings of the 2nd International Conference 'Railway Engineering-99', M.C. Forde, Editor. 1999: London, UK.

35. Breiman, L., Random forests. Machine Learning, Vol. 45, No. 1, 2001, pp. 5-32. DOI: 10.1023/A:1010933404324.

36. Wang, G., M. Burrow, and G. Ghataora, Study of the factors affecting road roughness measurement using smartphones. Journal of Infrastructure Systems, Vol. 26, No. 3, 2020, p. 04020020. DOI: 10.1061/(ASCE)IS.1943-555X.0000558.

37. Krusche, A., S. Clarke, Q. Fu, L. Wilde, C. Warburton, J.M. Easton, and D. Chick, Using real-time prediction to influence passenger behaviour in order to improve the management of disruption on the railway, Rail Safety and Standards Board, UK, COF-DSP-02, 2021. Available from: https://www.rssb.co.uk/research-catalogue/CatalogueItem/COF-DSP-02 (accessed: 11 March 2023).

38. Tong, W., H. Hong, H. Fang, Q. Xie, and R. Perkins, Decision forest: Combining the predictions of multiple independent decision tree models. Journal of Chemical Information and Computer Sciences, Vol. 43, No. 2, 2003, pp. 525-531. DOI: 10.1021/ci020058s.

39. Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg, Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, Vol. 12, 2011, pp. 2825-2830. DOI: 10.5555/1953048.2078195.