Wiley Series in Operations Research and Management Science

BIG DATA AND DIFFERENTIAL PRIVACY

ANALYSIS STRATEGIES FOR RAILWAY TRACK ENGINEERING

Nii O. Attoh-Okine

WILEY

Big Data and Differential Privacy

Wiley Series in Operations Research and Management Science

A complete list of the titles in this series appears at the end of this volume.

Big Data and Differential Privacy

Analysis Strategies for Railway Track Engineering

Nii O. Attoh-Okine

WILEY

This edition first published 2017

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloguing-in-Publication Data

Names: Attoh-Okine, Nii O., author.
Title: Big data and differential privacy : analysis strategies for railway track engineering / Nii O. Attoh-Okine.
Other titles: Wiley series in operations research and management science.
Description: Hoboken, NJ : John Wiley & Sons, 2017. | Series: Wiley series in operations research and management science | Includes bibliographical references and index.
Identifiers: LCCN 2017005398 (print) | LCCN 2017010092 (ebook) | ISBN 9781119229049 (cloth) | ISBN 9781119229056 (pdf) | ISBN 9781119229063 (epub)
Subjects: LCSH: Railroad tracks–Mathematical models. | Data protection–Mathematics. | Big data. | Differential equations.
Classification: LCC TF241 .A88 2017 (print) | LCC TF241 (ebook) | DDC 625.1/4028557–dc23
LC record available at https://lccn.loc.gov/2017005398

Cover design: Wiley
Cover image: (Top Image) © Jaap Hart/iStockphoto; (Bottom Image) © mbbirdy/Gettyimages

Set in 10/12pt WarnockPro by SPi Global, Chennai, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Contents

Preface *xi*
Acknowledgments *xiii*

1 Introduction *1*
1.1 General *1*
1.2 Track Components *2*
1.3 Characteristics of Railway Track Data *4*
1.4 Railway Track Engineering Problems *6*
1.5 Wheel–Rail Interface Data *11*
1.5.1 Switches and Crossings *14*
1.6 Geometry Data *15*
1.7 Track Geometry Degradation Models *20*
1.7.1 Deterministic Models *21*
1.7.1.1 Linear Models *21*
1.7.1.2 Nonlinear Models *22*
1.7.2 Stochastic Models *22*
1.7.3 Discussion *25*
1.8 Rail Defect Data *25*
1.9 Inspection and Detection Systems *33*
1.10 Rail Grinding *37*
1.11 Traditional Data Analysis Techniques *40*
1.11.1 Emerging Data Analysis *41*
1.12 Remarks *41*
References *42*

2 Data Analysis – Basic Overview *49*
2.1 Introduction *49*
2.2 Exploratory Data Analysis (EDA) *49*
2.3 Symbolic Data Analysis *53*
2.3.1 Building Symbolic Data *54*
2.3.2 Advantages of Symbolic Data *54*

2.4 Imputation *54*
2.5 Bayesian Methods and Big Data Analysis *56*
2.6 Remarks *57*
References *57*

3 Machine Learning: A Basic Overview *59*
3.1 Introduction *59*
3.2 Supervised Learning *60*
3.3 Unsupervised Learning *61*
3.4 Semi-Supervised Learning *61*
3.5 Reinforcement Learning *61*
3.6 Data Integration *63*
3.7 Data Science Ontology *63*
3.7.1 Kernels *64*
3.7.1.1 General *64*
3.7.1.2 Learning Process *64*
3.7.2 Basic Operations with Kernels *65*
3.7.3 Different Kernel Types *65*
3.7.4 Intuitive Example *65*
3.7.5 Kernel Methods *66*
3.7.5.1 Support Vector Machines *66*
3.8 Imbalanced Classification *69*
3.9 Model Validation *70*
3.9.1 Receiver Operating Characteristic (ROC) Curves *70*
3.9.1.1 ROC Curves *71*
3.10 Ensemble Methods *71*
3.10.1 General *71*
3.10.2 Bagging *72*
3.10.3 Boosting *74*
3.11 Big P and Small N ($P \gg N$) *74*
3.11.1 Bias and Variances *75*
3.11.2 Multivariate Adaptive Regression Splines (MARS) *75*
3.12 Deep Learning *79*
3.12.1 General *79*
3.12.2 Deep Belief Networks *81*
3.12.2.1 Restricted Boltzmann Machines (RBM) *81*
3.12.2.2 Deep Belief Nets (DBN) *82*
3.12.3 Convolutional Neural Networks (CNN) *83*
3.12.4 Granular Computing (Rough Set Theory) *83*
3.12.5 Clustering *89*
3.12.5.1 Measures of Similarity or Dissimilarity *89*
3.12.5.2 Hierarchical Methods *90*
3.12.5.3 Non-Hierarchical Clustering *91*

3.12.5.4 k-Means Algorithm *92*
3.12.5.5 Expectation–Maximization (EM) Algorithms *93*
3.13 Data Stream Processing *95*
3.13.1 Methods and Analysis *95*
3.13.2 LogLog Counting *96*
3.13.3 Count–Min Sketch *97*
3.13.3.1 Online Support Regression *104*
3.14 Remarks *105*
References *105*

4 Basic Foundations of Big Data *113*
4.1 Introduction *113*
4.2 Query *116*
4.3 Taxonomy of Big Data Analytics in Railway Track Engineering *123*
4.4 Data Engineering *124*
4.5 Remarks *130*
References *130*

5 Hilbert–Huang Transform, Profile, Signal, and Image Analysis *133*
5.1 Hilbert–Huang Transform *133*
5.1.1 Traditional Empirical Mode Decomposition *134*
5.1.1.1 Side Effect (Boundary Effect) *138*
5.1.1.2 Example *139*
5.1.1.3 Stopping Criterion *139*
5.1.2 Ensemble Empirical Mode Decomposition (EEMD) *143*
5.1.2.1 Post-Processing EEMD *144*
5.1.3 Complex Empirical Mode Decomposition (CEMD) *144*
5.1.4 Spectral Analysis *145*
5.1.5 Bidimensional Empirical Mode Decomposition (BEMD) *146*
5.1.5.1 Example *147*
5.2 Axle Box Acceleration *150*
5.2.1 General *150*
5.3 Analysis *151*
5.4 Remarks *153*
References *153*

6 Tensors – Big Data in Multidimensional Settings *157*
6.1 Introduction *157*
6.2 Notations and Definitions *158*
6.3 Tensor Decomposition Models *161*
6.3.1 Nonnegative Tensor Factorization *162*
6.4 Application *164*

6.5 Remarks *170*
References *171*

7 Copula Models *175*
7.1 Introduction *175*
7.1.1 Archimedean Copulas *179*
7.1.1.1 Concordance Measures *180*
7.1.2 Multivariate Archimedean Copulas *183*
7.2 Pair Copula: Vines *184*
7.3 Computational Example *186*
7.3.1 Results *187*
7.4 Remarks *192*
References *193*

8 Topological Data Analysis *197*
8.1 Introduction *197*
8.2 Basic Ideas *197*
8.2.1 Topology *197*
8.2.2 Homology *198*
8.2.2.1 Simplicial Complex *199*
8.2.2.2 Cycles, Boundaries, and Homology *200*
8.2.3 Persistent Homology *201*
8.2.3.1 Filtration *201*
8.2.4 Persistence Visualizations *201*
8.2.4.1 Persistence Diagrams *201*
8.3 A Simple Railway Track Engineering Application *203*
8.3.1 Embedding Method *203*
8.4 Remarks *204*
References *204*

9 Bayesian Analysis *207*
9.1 Introduction *207*
9.1.1 Prior and Posterior Distributions *208*
9.2 Markov Chain Monte Carlo (MCMC) *210*
9.2.1 Gibbs Sampling *210*
9.2.2 Metropolis–Hastings *210*
9.3 Approximate Bayesian Computation *210*
9.3.1 ABC – Rejection algorithm *211*
9.3.2 ABC Steps *216*
9.4 Markov Chain Monte Carlo Application *216*
9.5 ABC Application *219*
9.6 Remarks *221*
References *222*

10 **Basic Bayesian Nonparametrics** *225*
10.1 General *225*
10.2 Dirichlet Family *226*
10.2.1 Moments *226*
10.2.1.1 Marginal Distribution *227*
10.3 Dirichlet Process *227*
10.3.1 Stick-Breaking Construction *228*
10.3.2 Chinese Restaurant Process *229*
10.3.3 Chinese Restaurant Process (CRP) for Infinite Mixture *229*
10.3.4 Nonparametric Clustering and Dirichlet Process *231*
10.4 Finite Mixture Modeling *231*
10.5 Bayesian Nonparametric Railway Track *232*
10.6 Remarks *233*
References *233*

11 **Basic Metaheuristics** *235*
11.1 Introduction *235*
11.1.1 Particle Swarm Optimization *235*
11.1.2 PSO Algorithm Parameters *237*
11.2 Remarks *237*
References *239*

12 **Differential Privacy** *241*
12.1 General *241*
12.2 Differential Privacy *242*
12.2.1 Differential Privacy: Hypothetical Track Application *243*
12.3 Remarks *247*
References *247*

Index *249*

Preface

The ability of railway track engineers to handle and process large and continuous streams of data will provide a considerable opportunity for railway agencies. This will help decision makers to make informed decisions about the maintenance, reliability, and safety of the railway tracks. Now a period is beginning in which the problem is collecting the railway track data and analyzing it in a defined period of time. Therefore, the tools and methods needed to achieve this analysis need to be addressed. Knowledge derived from big data analytics in railway track engineering will become one of the foundational elements of any railway organization and agency. Also, another key issue has been the protection of data by different railway organizations. Therefore, although the data are available, they are really shared among different agencies. This makes the issue of differential privacy of utmost importance in the railway industry. Also, it is not clear if the industry has developed a clear way of both protecting and accessing the data from third parties.

Data science is an emerging field that has all the characteristics needed by railway track engineers to address and handle the enormous amounts of data generated by various technology platforms currently in place. The major objective is for railway track engineers to have an understanding of big data. Using the right tools and methodologies, railway track big data will also uncover new directions for monitoring and collecting railway track data; this apart from the engineering side will also have a major business impact on railway agencies.

This book provides the fundamental concepts needed to work with big data applications for railway engineers. The concepts serve as a foundation, and it is assumed that the reader has some understanding of railway engineering. The book does not attempt to address railway track engineering as a subject, but it does address the use of data science and the big data paradigm in railway track applications. Colleagues in industry will find the book very handy, but it will also serve as a new direction for graduate students interested in data science and the big data paradigm in infrastructure systems. The work in this book is intended to be accessible to an audience broader than those in railway track engineering.

Furthermore, I hope to shed a bright light on the enormous potential and future development that the big data paradigm will bring to railway track engineering. The amount of data railway agencies already have and the amount they are planning to collect in the future make this book an important milestone. This book attempts to bring together new emerging topics in a coherent way that can address different methodologies that can be used in solving a variety of railway track problems in the analysis of large data from various inspection technologies. In preparing the book, I tried to achieve the following objectives: (a) to develop some data science ontologies, (b) to provide the formulation of large railway track data using big data analytics, (c) to provide direction on how to present the data (visualization of the results), (d) to provide practical applications for the railway and infrastructure industry, and (e) to provide a new direction in railway track data analysis.

Finally, I assume full responsibility for any errors in the book. The opinions presented in the book represent my experiences in civil infrastructure systems, machine learning, signal analysis, and probability analysis.

January, 2016

Nii O. Attoh-Okine
Newark, Delaware, USA

Acknowledgments

I would like to thank the staff of John Wiley & Sons, Inc., especially Susanne Steitz-Filler, for their time. I would also like to thank Dr. Allan Zarembski and Joe Palese and Hugh Thompson of FRA for their support and encouragement. Thanks also to my current and former graduate students Dr. Yaw Adu-Gyamfi, Dr. Offei Adarkwa, and Emmanuel Martey for offering constructive criticisms. Special thanks to Silvia Galvan-Nunez who additionally provided me support with the complex LaTex issues. I would also like to thank Erin Huston for editing the first draft of the book. Finally, as always, I would like to thank my family: my two children, Nii Attoh and Naa Djama; my wife, Rebecca, for providing the peace and excellent working environment; and my brother, Ashalley Attoh-Okine, an excellent actuary and energy expert, who introduced me to so many data analysis techniques, which have been part of my research over the years. I dedicate the book to the memory of my parents, Madam Charkor Quaynor and Richard Ayi Attoh-Okine, and my maternal grandparents, Madam Botor Clottey and Robert Quaynor.

1

Introduction

1.1 General

Currently, railroads collect enormous quantities of data through vehicle-based inspection cars, trackside (or wayside) monitoring systems, hand-held gauges, and visual inspections. In addition, these data are located geographically using the global positioning system (GPS). The data from these inspection systems are collected electronically by hand or using various sensors, video inspections, machine visions, and many other sources. Furthermore, the data are growing both in quantity and quality and are more precise and diverse. Data of extremely large sizes are difficult to analyze using traditional approaches since they may exceed the limits of a typical spreadsheet. The railway track data are present in diverse forms, including categorical, numerical, or continuous values. The general characteristics of the data dictate which type of method is appropriate for analysis. For example, categorical and nominal values are unsorted, while numerical and continuous values are assumed to be sorted or to represent ordinal data (Ramírez-Gallego et al., 2016).

The development of advanced sensors and information technology in railway infrastructure monitoring and control has provided a platform for the expansive growth of data. This has created a new paradigm in the processing, storing, streaming, and visualization of data and information. Furthermore, changes in technology include the possibility of installing sensors and smart chips in critical infrastructure to measure system performance, current condition, and other indicators of imminent failures. Many of the railway infrastructure components have communication capabilities that allow data to be uploaded on demand.

Big data is about extremely large volumes of data originating from various sources: databases, audio and video, millions of sensors, and other systems. The sources of data in some cases provide structured outputs, but most are unstructured, semi-structured, or poly-structured. These data are streaming in some cases with high velocity, and the data exposes at a higher speed or some speed as it is generated.

Big Data and Differential Privacy: Analysis Strategies for Railway Track Engineering, First Edition. Nii O. Attoh-Okine.

This chapter presents a general overview, basic description, and properties of deterministic and random data that are encountered in railway track engineering data and relies heavily on the data output based on the advances in sensors, information technology, high information technology, and development that has led to extremely massive data sets. These large data sets have made the traditional analytical techniques used for railway track maintenance and safety issues somewhat obsolete.

The data obtained in railway track monitoring are collected by different sensors, at different times and environmental conditions, at different frequencies, and at different resolutions. The outputs of these data have different characteristics: discrete or continuous, spatial or temporal, signal and images, and categorical and objective, among others. All these characteristics, properties, and the extreme volume of data collected have made traditional analytical techniques very inefficient; issues like visualization and data streaming, which are very critical in railway track maintenance and safety, are not adequately addressed. The traditional statistical techniques fail to scale up to the extremely large volumes of data collected by railway inspection vehicles and trackside monitoring devices. Therefore, the growing amount of data generated by railway track inspection activities is outpacing the current capacity to explore and interpret these data and hence appropriately addresses maintenance and safety issues.

1.2 Track Components

The term "tracks" includes superstructure, substructure, and special structures (Figure 1.1). The superstructure is made of rails, ties, fasteners, turnouts, and crossings, while the substructure consists of ballast, subballast, the subgrade, and other drainage facilities. The superstructure and substructure are separated by the tie–ballast interface.

The main purpose of the railway track structure is to provide a safe and economical train transportation system through guiding the vehicle and transmitting loads through the track components to the subgrade. The carrying

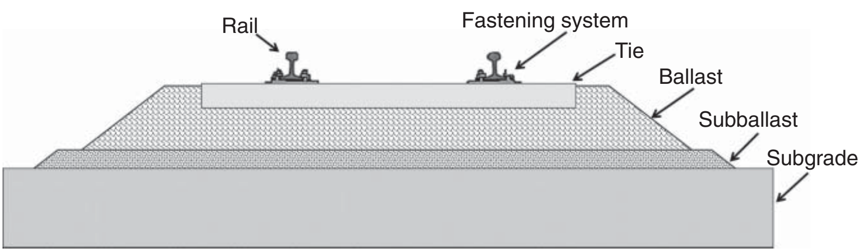

Figure 1.1 Track structure components

capacity and long-term durability of the track structure highly depend on how the superstructure and substructure respond to and interact with each other when subjected to moving trains and environmental factors (Selig and Waters, 1994; Kerr, 2003).

The function of different rail components has been presented by various authors, such as Hay (1982), Selig and Waters (1994), Esveld (2001), Kerr (2003), Sadeghi (2010), and Tzanakakis (2013). The aim of this section is to summarize this function. The rails are the longitudinal steel members that are placed on spaced ties to guide the train wheels evenly and continuously. Their strength and stiffness must be sufficient to maintain a steady shape and smooth track configuration and to resist various forces (vertical, lateral, and longitudinal) by vehicles. The rails also in some cases serve as electrical conductors for the signal circuit and also as a groundline for the electric locomotive power circuit. The profile of the rail surface (transverse and longitudinal) and wheel surface has a major influence on the operation of the vehicles on the track, and track defects may in some instances create and cause large dynamic loads that lead to derailment and safety issues, as well as accelerated degradation.

Most steel rail sections are connected either by bolted joints or by welding. The bolted joints create several problems, including rough riding track, undesirable vibration, and additional impact loads, among others; hence, the use of continuous welded rail (CWR) has been the better solution. CWR attempts to address some of the disadvantages of the bolted joints, which have its own set of maintenance requirements.

The rail fastener systems, or fastenings, include all the components that connect the rail to the tie, with the tie plate, spike, and anchor for wood ties and clip, insulator, and elastic fasteners for concrete ties. The function of the fastenings is to retain the rail against the ties and resist vertical, lateral, longitudinal, and overturning movements of the rail. They also serve as wheel load impact attenuation, increasing track elasticity, as well as electrical isolation between rails.

For concrete tie tracks, rail pads are installed on rail supporting points to reduce and transfer the stress and dynamic forces from the rail to the ties, and they reduce the interaction force between the rail and the ties (Choi, 2014). The pads also provide adequate resistance to longitudinal and rotational movement of the rail and provide a conforming layer between the rail and tie to avoid contact areas of high pressure. From a dynamic point of view, the rail pads tend to influence overall track stiffness.

Ties are transverse beams resting on ballast and support. They span below and tie together two rails. The main functions of ties are as follows:

- Uniformly transfer and distribute loads from the rail to the ballast
- Hold the fastening system to maintain proper track gage
- Restrain the lateral, longitudinal, and vertical rail movement by anchorage of the superstructure to the ballast

- Provide a cant to the rails to help develop proper wheel–rail contact by matching the inclination of the conical wheel shape
- Provide an insulation layer
- Allow fast drainage of fluid
- Allow for proper ballast maintenance

Ballast is the layer of crushed stone placed at the top layer of the substructure in which the tie is embedded. It is an elastic support and transfers forces from the rail and tie to the subballast. As some of its functions, it

- Distributes load from ties uniformly over the subgrade
- Anchors the track in place against lateral, vertical, and longitudinal movements
- Absorbs shock from the dynamic load
- Allows suitable global and local track settlement
- Avoids freezing and melting (thawing) problems by frost action
- Allows for proper drainage
- Allows for maintenance of the track geometry

The subballast is the layer between the ballast and the subgrade. As some of its functions, it

- Reduces the stress at the bottom of the ballast layers to a reasonable level to protect the subgrade
- Migrates fines from the subgrade to the upper layer of the ballast
- Protects the subgrade from the ballast
- Permits drainage of water that might otherwise flow upward from the subgrade

The subgrade is the last support of the track systems and, in some cases, is the existing soil at the location, unless the existing formation is very weak. In the case of a weak existing formation, techniques like stabilization and modification of the existing elevation use more appropriate soil. The addition of geosynthetic material has been used to improve the subgrade performance and bearing capacity. Its main functions are the following:

- Provide support to the track structure
- Bear and distribute the resultant load from the train vehicle through the track structure
- Provide sufficient drainage

1.3 Characteristics of Railway Track Data

Railway track data are similar to data from other infrastructures. Its characteristics include the following:

- *Massive Data Sets*. Railway track data collection and monitoring has resulted in extremely large data sets for infrastructure monitoring. In some cases, the actual data are processed and only the reduced version is stored, while in most cases smaller amounts of data are stored for further analysis.
- *Unstructured Data, Heterogeneous Databases*. Some of the railway track data are stored in databases. In most cases, different agencies and countries have different data formats, different database management systems, and different data manipulation algorithms. Most of these databases are evolving, which in some cases makes analysis and data mining across them challenging. Some of the databases include unstructured images, plots, and tables, as well as links to other transportation and infrastructure documents of the agency. This can be challenging in terms of both analysis and reporting.
- *Information in the Form of Images*. The analysis of railway track, in terms of both rail and geometry defects, by its very nature deals with issues associated with the extraction of meaningful information from massive amounts of railway track images, thus opening a new direction in railway track analysis.
- *Poor Quality of Data*. Railway track data analysis, especially the image data, in most cases is of poor quality due to the railway track environment and sensor noise. In some cases, data are missing or input incorrectly. Furthermore, the data from different sources can vary in terms of quality. Also, the railway inspectors may in some cases have incomplete knowledge about the mechanism and initiation of different defects. This may lead to inconclusive reporting and analysis.
- *Multiresolution and Multisensor Data*. Several different sensors are used to collect different information and data. This may create a situation where several images may have different resolutions over time. Therefore, care must be taken so that the change in resolution can be included.
- *Noisy Data*. Noisy data cannot be avoided in railway track data collections. Methods of reducing the noise in data need to be implemented during the preprocessing of the data for further analysis. For example, shadows and orientations of the vehicle collecting the data can have an impact on the images. Therefore, poor illumination can have a major impact on the obtained image.
- *Missing Data*. The risk of missing data is always present in railway track data collection; this is mostly due to sensor malfunction. Filling the gaps can be a daunting task. Again care must be taken with how missing data is included.
- *Streaming Data*. Some of the data sets collected during railway monitoring can be streaming in nature; that is, a constant stream of data is being collected and received. This requires a specialized set of analyses different from the chunk data methods used in traditional analysis.

More broadly, the data can either be random or deterministic. The random data is shown in Figure 1.2, and the deterministic data is shown in Figure 1.3, as presented by Bendat (1998).

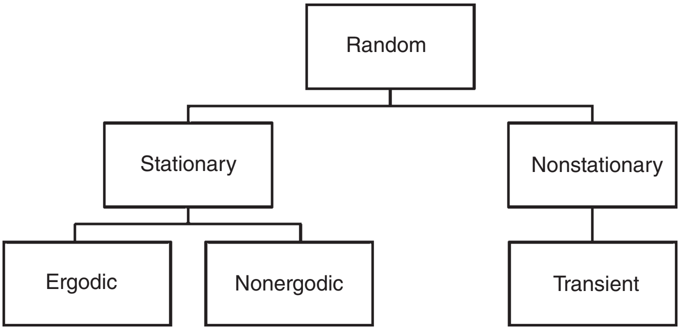

Figure 1.2 Classification of random data (Bendat (1998). Reproduced with their permission of John Wiley & Sons)

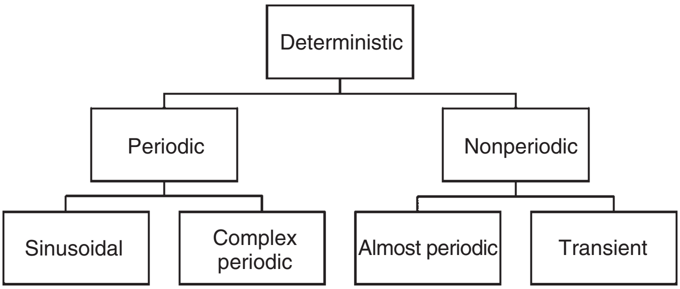

Figure 1.3 Classification of deterministic data (Bendat (1998). Reproduced with their permission of John Wiley & Sons)

Table 1.1 shows the general taxonomy of big data methods in railway engineering.

1.4 Railway Track Engineering Problems

Generally, railway track engineering problems can be classified into two groups according to Santamarina and Fratta (1998): (a) forward problems and (b) inverse problems. Table 1.2 shows the group of problems that fall under the two categories. For forward problems, the major objective is to design systems to satisfy predefined performance criteria. Also, convolution forms part of the forward problems. In convolution, the input is known, the type of system is known, and the only unknown is the output.

Inverse problems can either be (a) system identification where the input and output are known but the system is unknown or (b) deconvolution where the input is unknown, while the system and output are known. Figure 1.4 shows a generic representation of general civil engineering problems, including railway

Table 1.1 Taxonomy of big data in railway engineering.

Analysis domain	Sources	Characteristics	Approaches	Comments
Structured data	Field data collection, sensors, data from scientific experiments	Structured records, real time	Data mining, statistical analysis	All infrastructure systems need field data
Unstructured data	Extreme events, sensors	Unstructured records, mixture of variables	Anomaly detection	Infrastructure inspection reports, specification updates
Text analytics	Logs, email, corporate documents, government rules and regulations, text content of web pages, citizen feedback and comments	Unstructured, rich textual, context, semantic, language dependent	Document presentation, NLP, information extraction, topic model, summarization, categorization, clustering, question answering, option mining	Early detection
Multimedia analytics	Corporation-produced multimedia, user-generated multimedia, surveillance	Image, audio, video, massive, redundancy, semantic gap	Summarization, annotation, indexing and retrieval, recommendation, event detection	Early detection
Mobile analytics	Mobile apps, RFID sensors		Monitoring, location-based mining	

Table 1.2 Engineering problems.

Forward problems		Inverse problems	
System design	**Convolution**	**System identification**	**Deconvolution**
The system is designed to satisfy performance criteria (controlled output for estimated input)	Input: known System: known Output: unknown	Input: known System: unknown Output: known	Input: unknown System: known Output: known

(Santamarina and Fratta., 1998) Reproduced with the permission of ASCE.

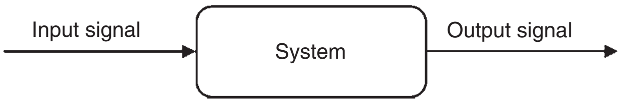

Figure 1.4 Engineering signals

track problems. But there can be different situations, including (a) single-input–output relationships as shown on the generic representation and (b) multiple-input–output relationships. Therefore, depending on the structural and objective analyses, there are different assumptions and analyses. Systems can be divided into two broad groups, (a) linear and (b) nonlinear. The linear systems can be further divided into constant parameter and time-varying systems (Bendat, 1998).

Major parts of railway track data are in the form of signals and images; therefore, a deeper understanding of analytical issues for signals and images is needed to analyze and interpret railway track data. A major issue related to track images is the presence of noise, which tends to affect the overall images if it is not properly reduced or accounted for. Therefore, efficient algorithms are needed to reduce noise in railway track images before further analysis can be done.

Table 1.3 shows examples of different track inspection technologies and their level of maturity. Railway track conditions are, in most cases, evaluated using the characteristics of track geometry wave form and vehicle dynamic response to the track. Also, in some cases, images from high definition cameras are also collected. It is apparent that to obtain the true picture of the railway track condition, there should be methods that can go beyond traditional statistical analysis. An efficient method is one that can perform the mining of the data, reduce noise from the wave forms, and combine data and information from different sources to provide a clear understanding of what maintenance activities to perform and how to satisfy all safety requirements.

Table 1.3 Track inspection technologies.

Component	Inspection	Base technology	Maturity	Measurement	Maintenance activity	Level of automation
Rail	Profile	Laser/camera	Mature	Wear, cant, lip, GF angle, profile/contact	Wear, grinding, cant, lip	Very high
	Corrugation	Inertial or chord	Moderate	Wavelength and amplitude	Grinding	High
	Internal fatigue	Ultrasonic	Mature	Location and size of flaw	Replacement	High
	Surface fatigue	High definition camera	New	Location and size of flaw	Grinding	Moderate
	Surface cracks	Eddy current	New	Location, length, and density of cracks	Grinding	Low
Joint bars	Cracks, bolts	High definition camera	New	Location of failure	Replacement	Moderate
Ties	Cracks, plate cut	High definition camera	New	Length, width, and density of cracks	Replacement	Low
	Rot	Back scatter X-ray	Very new	Internal density	Replacement	Low
	GRMS	Force/displacement	Mature	Tie/fastener interface strength/stiffness	Replacement	High
Fasteners	Condition, existence	High definition camera	New	Missing or damaged	Replacement	Low
	GRMS	Force/displacement	Mature	Tie/fastener interface strength/stiffness	Replacement	High
Turnouts/ diamonds	Rail geometry	Laser/camera	New	Relative heights and dimensions, points, frog	Replacement, grinding, welding	Moderate
	Track geometry	Inertial	Mature	Relative rail-to-rail relationship	Tamping	High
	Component condition	High definition camera	New	Missing or damaged	Replacement, tighten	Low

(Continued)

Table 1.3 (Continued)

Component	Inspection	Base technology	Maturity	Measurement	Maintenance activity	Level of automation
Geometry	Geometry car	Inertial or contact	Mature	Relative rail-to-rail relationship	Tamping	High
	Absolute geometry	Inertial	New	Absolute rail-to-rail relationship	Tamping	Low
	Vertical track interaction	Inertial	New	Vertical vehicle response to track	Tamping	High
Ballast/ subballast/ subgrade	Ground-penetrating radar	Radar	Moderate	Ballast depth and condition	Tamping, undercutting	Low
	Cone penetrometer	Force/displacement	Moderate	Ballast/subballast depth, condition, and strength	Tamping, undercutting	Low
	Vertical track deflection	Force/displacement	New	Track stiffness	Tamping, undercutting	Moderate
	Profile	Lidar	New	Ballast profile	Add ballast and tamp	Moderate
Clearance	Envelope	Lidar/laser	Moderate	Surrounding clearance and obstructions		Moderate

1.5 Wheel–Rail Interface Data

The wheel–rail contacts at the interface between the wheel and rail determine in part the reliability of railway systems. Tzanakakis (2013) presented in Figure 1.5 the different outcomes and effects of wheel–rail contact. The rail vehicles are supported, accelerated, and decelerated by contact forces acting on extremely small wheel–rail contact areas (around 1 cm^2).

Meymand et al. (2016) presented a comprehensive survey on the topic. The paper discusses well-known theories for modeling normal contacts based on Hertzian and non-Hertzian methods and tangential contacts based on Kalker's linear theory and Polach theory.

Track irregularities tend to produce different magnitudes of force on the track. These forces on the track can result in three types of loads: (a) vertical, (b) lateral, and (c) longitudinal. Lateral loads are transverse to the track, while longitudinal loads are parallel to the track. Depending on their nature, loads can be (a) static loads, (b) quasi-static loads, and (c) dynamic loads. The dynamic loads may cause

- Irregularities in the track geometry
- Wear of the running surface
- Discontinuities on the running surface, which includes switches and frogs
- Dynamic forces, which appear in two categories: P1 and P2 forces. Frequencies of P1 forces range between 100 and 2000 Hz, are mainly impact forces.

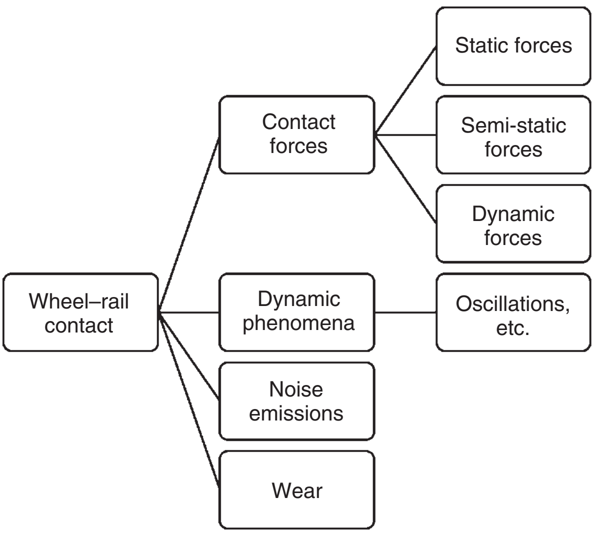

Figure 1.5 Wheel–rail contact impacts (Tzanakakis (2013). Reproduced with the permission of Springer)

P1 forces can cause, among other things, bolt hole failure and cracking of concrete ties and have minimal effects on the ballast or subgrade. P2 forces contribute to the degradation of track geometry and are classified in the frequency range between 30 and 100 Hz.

The contact between the wheel and the rail is the basic constitutive element of the railway dynamics (Table 1.4). For modeling purposes, two aspects are considered (Trzaska, 2012): (a) the geometric or kinematical relations of the wheel–rail contact and (b) the contact mechanical relations for the calculation of the contact forces. The wheel–rail contact provides insight into the formation of corrugation and other rail defects, like wear, crack growth, and others. The wear depends on tangential forces and creep age at the contact patch. Using mathematical analysis, it is possible to build a comprehensive and functional understanding of wheel–rail interaction, suspension and suspension component behavior, simulation, and experimental validation. This is beyond the scope of the current analysis.

Figure 1.6 shows various wheel–rail interfaces and their effects.

In wheel–rail contact there are three "zones" of contact, namely, Region A, Region B, and Region C, as shown in Figure 1.7. Region A is the contact between the central region of the rail crown and the wheel thread (conicity, hollow wear, and thermal loads), Region B is the extreme reference gage corner contact of the two-point contact, and, finally, Region C is the field side contact. At Region A,

Table 1.4 Vertical track forces.

Cause	Force	Symptom
Impact at rail welds	Rail: P1+P2	Rail fatigue failure Corrugations Pad degradation Tie cracking/movement Ballast degradation Weld fatigue
Vehicle/track interaction	Quasi-static Dynamic forces	Track geometry deterioration Rail failure/fatigue Ballast failure/degradation Subgrade failure/degradation
Wheel irregularities	Wheel: P1+P2	Tie cracking Rail breaks Wheel cracks Ballast deterioration

Tzanakakis (2013). Reproduced with the permission of Springer.