Transportation Research Part C 36 (2013) 116–124

Contents lists available at ScienceDirect

Transportation Research Part C

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

Quality control of weigh-in-motion data incorporating threshold values and rational procedures

Derong Mai *, Rod E. Turochy, David H. Timm

Department of Civil Engineering, 238 Harbert Engineering Center, Auburn University, AL 36849, United States

ARTICLE INFO

Article history:
Received 28 March 2013
Received in revised form 25 August 2013
Accepted 25 August 2013

Keywords:
Weigh-in-motion system
WIM data
Quality control
Threshold check
Rational check

ABSTRACT

Weigh-in-motion (WIM) stations constitute a key source of traffic data for use in mechanistic–empirical pavement design. One of the major improvements provided by the Mechanistic–Empirical Pavement Design Guide (MEPDG) is in traffic characterization. Instead of converting all truck axles to 18,000 lb equivalent single axles (ESALs), the Mechanistic–Empirical Pavement Design Guide (MEPDG) simulates every truck axle, and the associated stresses and strains imposed on the pavement structure, from a wide range of axle load spectra (ALS). This paper presents an objective approach to quality control (QC) of WIM data that includes threshold checks that detect implausible values of individual variables in the truck weight records and rational checks that examine patterns in axle load distributions and relationships among the variables. Instead of using subjective visual comparisons of gross vehicle weight (GVW) distributions, this research implements a peak-range check, peak-shift check, and correlation analysis to quantify the ALS comparison process of rational checks. A number-of-axles check that calculates the average number of axles per vehicle class is also introduced herein. The entire QC procedure has been applied to three years of data from 12 WIM stations in Alabama that used bending plate sensors. As a result, 23.8% of data were filtered out, and all data from one WIM station were removed. Therefore, QC of WIM data is strongly recommended, regardless of the extent of WIM system calibration. Furthermore, it is also recommended that the rational checks module be integrated in the data collection process for rapid detection of systematic errors.

1. Introduction

Numerous types of sensors are collecting data on roadways every day. These sensors provide an enormous amount of traffic data on every aspect of the transportation system. However, data rich does not necessarily mean information rich. For this reason, the transportation engineering community has devoted countless efforts to translate data into useful information in the field. On the other hand, the quality of traffic data is equally important because data of poor quality could result in incorrect interpretation and non-optimized use of resources.

These ideas apply to data collected from weigh-in-motion (WIM) systems. WIM data are widely used in pavement design, bridge design, law enforcement analysis and other fields, but its quality can be questionable. Even though WIM calibration recommendations through the Long Term Pavement Performance (LTPP) Program suggest local government agencies or data collectors calibrate WIM stations regularly, it is suspected that WIM stations may not be routinely calibrated (LTPP, 2001). Furthermore, WIM calibration may not be able to address random errors which are common in WIM data.

* Corresponding author. Tel.: +1 334 220 9983; fax: +1 334 844 6290.
E-mail address: dzm0005@auburn.edu (D. Mai).

0968-090X/$ - see front matter
http://dx.doi.org/10.1016/j.trc.2013.08.012

To generate traffic inputs required by the Mechanistic–Empirical Pavement Design Guide (MEPDG) in an efficient way, the TrafLoad software was developed in 2004 as part of the NCHRP Project 1-39 to serve as a principal source of traffic inputs for MEPDG (Wilkinson, 2005). In recent years, since little documentation has been published on quality control (QC) procedures for WIM data, some WIM data users may rely on TrafLoad to perform QC on their data. However, this is risky because TrafLoad only performs rudimentary checks for valid site IDs and lanes and direction values, and does not provide a sophisticated QC procedure (Wilkinson, 2005). A past study that evaluated systematic bias indicated that MEPDG pavement life estimation is highly sensitive to WIM data (Haider et al., 2012). To minimize the potential for "garbage in and garbage out" problems in later data analysis, QC of WIM data is crucial.

There are a few WIM data QC procedures that have been introduced at the federal level. LTPP applies its QC procedure (LTPP, 2001) to SPS WIM sites before its annual publication (LTPP, 2012); the Traffic Monitoring Guide (FHWA, 2001) published by the Federal Highway Administration (FHWA) focuses on calibration of WIM systems during system installation and maintenance. The Quality Control Procedure for Weigh-in-Motion Data (Nichols and Bullock, 2004) and the WIM Data Analyst's Manual (Quinley, 2010) introduce QC methods for agencies at different levels. Studies conducted for state DOTs in North Carolina (Ramachandran et al., 2011), Kentucky (Southgate, 1990), Oregon (Pelphrey and Higgins, 2006) and Arkansas (Wang, 2009; Nguyen, 2010) detailed their QC procedures and criteria. In the 1990s, Southgate (1990) found a logarithmic relationship between steering axle load and the first axle spacing (longitudinal distance between steering axle and the next axle group) to adjust systematic errors of weight data, and data from the static weight station were used as the calibration target. The Arkansas DOT QC process (Wang, 2009; Nguyen, 2010) followed the LTPP procedure (LTPP, 2001) that monitored peak patterns of tandem axles and percentages of overweight GVW. The study conducted for Oregon DOT (Pelphrey and Higgins, 2006) illustrated the use of acceptable ranges to identify and remove errors, but it was observed that these range checks could not filter out replicate identical records, and it was necessary to use gross vehicle weight (GVW) distributions to manually look for visual distinctions such as repeated records, spurious outliers, and other inconsistencies. In the NCDOT QC procedures, the premise of rational checks was that GVW distributions of the same vehicle classification in different months maintain a very stable pattern. Then, manual checks, visual interpretation and local knowledge were used to identify abnormal patterns caused by systematic errors (Ramachandran et al., 2011). However, these abnormal patterns are evaluated subjectively and had not been statistically quantified.

2. Objective and scope of work

The objective of this paper is to demonstrate an unbiased approach to QC of WIM data. Major improvements of this QC procedure include a data-file-size check, quantified axle load spectra (ALS) comparisons and number-of-axles check. This QC procedure has been applied to raw data from 12 WIM stations that use bending plate sensors in Alabama from 2006 to 2008 before further developing data for use with the MEPDG.

3. Methodology

The QC process developed in this study consists of two types of approaches to ensuring data validity: threshold checks and rational checks. WIM data with implausibly low or high values can be readily identified, for example, a semi-trailer with speed over 120 mph. Threshold checks are used to filter them out. However, some systematic errors cannot be detected merely by examining their values; to detect these errors, rational checks that examine axle load distributions and relationships among them are developed. The overall QC procedure is shown in Fig. 1.

In the first phase of the QC procedure, a file-size check is conducted on a monthly basis. Then, an out-of-range check inspects values of every row of data within these files. In the second phase, the ALS comparison module consists of a peak-range check, peak-shift check and correlation analysis, by looking at data on a monthly basis. Finally, the number-of-axles check examines station-wide axle groups per vehicle (AGPV) inputs. Each check is discussed in more detail below.

3.1. Threshold check

The threshold check phase consists of two steps: (1) eliminating dataset file-size outliers and (2) deleting out-of-range values. The file-size check is used to detect severe file size drops which represent substantial amounts of missing data. These drops might be due to WIM system failure, road maintenance, rehabilitation and so on. However, regardless of the abnormal circumstances which lead to a file-size outlier, disrupted truck traffic counting and weighing should not be used for pavement design purposes. Therefore, monthly datasets with file-size outliers should be eliminated. In the second step of the threshold check, an out-of-range check is applied to detect and remove extreme values caused by random errors.

3.1.1. File-size check

A file-size check is recommended by the FHWA's *WIM Data Analyst's Manual* (Quinley, 2010); however, no detailed procedures are discussed. The file-size check developed herein assumes that file size has a positive linear relationship with the volume of truck traffic counted, and a file-size outlier indicates WIM system errors or abnormal circumstances occurred on the road. The approach uses three years of WIM data to compare file sizes and makes no assumption on the abnormal