Comparing Traffic Speed Deflectometer and Falling Weight Deflectometer Data

Levenberg, Eyal; Pettinari, Matteo; Baltzer, Susanne; Christensen, Britt Marie Lekven

Published in:
Transportation Research Record

Link to article, DOI:
10.1177/0361198118768524

Publication date:
2018

Document Version
Peer reviewed version

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Comparing Traffic Speed Deflectometer and Falling Weight Deflectometer Data

(Accepted manuscript 18-00936)

Eyal Levenberg (corresponding author)
Department of Civil Engineering, Technical University of Denmark
Nordvej, Building 119
Kgs. Lyngby 2800, Denmark
Tel: +45 4525 1907
Email: eylev@byg.dtu.dk

Matteo Pettinari
The Danish Road Directorate
Guldalderen 12,
Hedehusene 2640, Denmark
Tel: +45 7244 7139
Email: map@vd.dk

Susanne Baltzer
The Danish Road Directorate
Guldalderen 12,
Hedehusene 2640, Denmark
Tel: +45 7244 3333
Email: sub@vd.dk

Britt Marie Lekven Christensen
Norwegian State Road Administration - East Region
Østensjøveien 34,
Oslo 0667, Norway
Tel: +47 9344 8513
Email: britt.christensen@vegvesen.no

Word count

<table>
<thead>
<tr>
<th>Section</th>
<th>Word Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract + Keywords</td>
<td>150</td>
</tr>
<tr>
<td>Text + Acknowledgement</td>
<td>4,175</td>
</tr>
<tr>
<td>References + Table titles + Figure captions</td>
<td>1,100</td>
</tr>
<tr>
<td>2 Tables + 6 Figures</td>
<td>2,000</td>
</tr>
<tr>
<td>Total</td>
<td>7,425</td>
</tr>
</tbody>
</table>
ABSTRACT
In recent years the pavement engineering community has shown increasing interest in shifting from a stationary Falling Weight Deflectometer (FWD) to moving testing platforms such as the Traffic Speed Deflectometer (TSD). This paper dealt with comparing TSD measurements against FWD measurements; it focused on the comparison methodology, utilizing experimental data for demonstration. To better account for differences in loading conditions between the two devices a new FWD deflection index was formulated first. This index served as reference/benchmark for assessing the corresponding TSD measurements. Next, a Taylor diagram was proposed for visualizing several comparison statistics. Finally, a modern agreement metric was identified and applied for ranking comparison results across different datasets. Overall, the suggested methodology is deemed generic and highly applicable to future situations, especially for assessing the worth of emerging device upgrades or improved interpretation schemes (or both).

Keywords: Falling Weight Deflectometer, Traffic Speed Deflectometer, Taylor diagram, Agreement metrics.

INTRODUCTION
The most common testing device for nondestructive evaluation of pavement condition is the Falling Weight Deflectometer (FWD). For a chosen test location, the FWD generates a vertical stress-pulse at the pavement surface - about 30 milliseconds in duration; this is achieved by dropping a mass and then blocking its fall in a controlled manner (1). The device records the time-history of the loading as well as the time-history of the resulting vertical surface velocities at several offset distances from the load center (2, 3). The velocity signals are internally integrated with respect to time to yield deflection time-histories. FWD measurements are usually employed to assess
mechanical layer properties by means of backcalculation (4-6). For this purpose, a pavement model is assumed wherein layers are treated as continuous having a priori known thicknesses.

Pavement management activities entail rapid inspection of structural condition on a network level. For such wide-area application backcalculation cannot always be performed. One reason is that layers are often distressed to a point that violates the continuity assumption in the modeling; another reason is that layer thicknesses are not readily available; finally, interpretation time is prohibitive, as ample data of drop-experiments are collected in field surveys. In light of these reasons the accepted FWD interpretation approach for pavement management is based on some index derived directly from the deflection peaks. The index is chosen such that it exhibits good statistical correlation with some definition of ‘structural capacity’ or ‘remaining life’ (7). The Surface Curvature Index $SCI_{300}$ is one such a parameter, it has units of length and denotes the difference between the central FWD peak deflection and the peak deflection at an offset of 300 mm (8, 9). It is calculated for a peak FWD load of 50 kN applied over a load-plate with a radius of 150 mm. The $SCI_{300}$ is roughly related to pavement fatigue life because, for a given system layering and loading, it correlates to the tensile strain at the bottom of the asphalt layer (10-12).

As means of increasing FWD testing safety, testing efficiency, and spatial coverage of the pavement condition, and as means of minimizing traffic disruption during deflection measurements, the pavement engineering community has been pushing for the development of moving deflection-measuring platforms. One promising device in this category is the Traffic Speed Deflectometer (TSD). This device consists of a custom-built truck equipped with a linear array of Doppler lasers (13, 14). The lasers, mounted on a stiff beam, are oriented downward towards the pavement surface with a small incidence angle; they measure instantaneous velocities in the direction of the laser rays. These readings embody the truck’s travel speed as well as the pavement deflections due to the truck’s loading. After normalization w.r.t. travel speed (measured independently), deflection-slopes are obtained.
Given that FWDs have been in service for a long time while the TSD is newer and keeps evolving, there is a growing interest within the pavement community to quantify the agreement level between measurements performed by the two devices. The possibility of accessing FWD-based indices from a moving measurement platform offers considerable savings in all aspects of data collection for pavement condition evaluation. This ability is especially important for road authorities and pavement managers because maintenance decisions are often based on past experience gained with FWD deflection data. Therefore, at least until the new technologies mature, there is a need to relate back and carry out comparisons against FWD results.

Such comparisons have been recently carried out by several research groups (15-17). The main approach taken in these studies included: (i) measuring the same pavement section by both devices; (ii) calculating an index directly from the measurements of the two devices - such as the SCI\textsubscript{300}; and (iii) quantifying the differences between the two datasets by application of established/classical statistical tools such as Pearson’s correlation coefficient, Deming regression (18), and Bland-Altman plots (19).

This paper also deals with contrasting TSD measurements against FWD data. Similar to other studies it describes and compares data collected by both devices over the same pavement section. However, focus here is not placed on the comparison results but rather on the comparison methodology, so that the work remains relevant when newer devices or improved technologies emerge. Subsequently, the objective here is to suggest and demonstrate three methodology advances: (i) the formulation of a new FWD deflection index that better accounts for the differences in loading conditions between the two devices, and therefore better suited to serve as basis for statistical comparison (20); (ii) the graphical display of several statistics in a Taylor diagram to facilitate and enrich the comparative assessment (21); and (iii) the application of a modern agreement metric that embodies both correlation information and information on deviation magnitudes (22).
FIELD DATA
A newly constructed asphalt road, located close to the city of Slagelse (Denmark), was chosen for a testing campaign that included both TSD and FWD measurements. The designed pavement layering included (top to bottom): 190 mm asphalt concrete, 200 mm unbound granular base course (gravely material), and about 400 mm drainage subbase (sandy material). The Subgrade was visually classified as gravelly boulder clay, with 100% passing the 63 mm sieve, and about 50% passing the 0.063 mm sieve. The road was in pristine condition during testing, before any traffic was allowed, and before the final 50 mm wearing course was paved. All measurements were performed within five hours during a single day; the pavement surface temperature, as monitored by the TSD, was 17.6 °C on average, with a general increasing trend, fluctuating in the range of 14.4 °C and 20.1 °C.

A line was painted over the asphalt base layer along a stretch of 430 m to facilitate and guide the testing. Both the TSD and FWD device operators were instructed to measure along this line. Data collection was done in four steps according to the scheme shown in Figure 1 wherein TSD runs and run numbers are indicated by the shaded square markers. In Step 1, the TSD was driven at four different speeds: 20, 40, 60, and 70 km/h; this was repeated three times (Runs #1, #2, and #3). In step 2, FWD measurements were performed at either 5 m or 10 m intervals over 63 different stations - covering a distance of 350 m. Step 3 included more TSD measurements, with the device running at three different speeds: 40, 60, and 70 km/h; this was repeated twice (Runs #4 and #5). In the final and fourth Step, FWD measurements were resumed at 5 m intervals to complete the coverage of the 430 m road stretch, for a total of 80 test locations.

Measurements from Steps 2, 3, and 4 were chosen for interpretation herein; these include all 80 FWD test locations, and TSD Runs #4 and #5 at 40, 60, and 70 km/h. A wider variation in pavement temperature was the main motivation for excluding Step 1. Nonetheless, so doing is inconsequential, as the work focuses on the comparison methodology.
TSD Measurements

The TSD device employed for the testing was a first generation model. The device had a heavy rear single-axle equipped with dual-wheels on each side. Under stationary conditions this axle exerted a total load of 100 kN. Four Doppler lasers (Polytec model OFV-353, December 2003) were incorporated for reading deflection-slopes; these were fixed to a stiff beam that was oriented in the travel direction and positioned in-between the dual-wheels on the right-hand side. Three out of the four lasers were placed relatively close to the dual-wheels with 100 mm spacing. The fourth laser served as reference and placed 3.6 m away, where the pavement deflection is assumed negligibly small.

Figure 2 presents a sketch of the TSD loading and measurement setup, focusing on the loaded zone and therefore displaying only three lasers. The tire-pavement contact areas are represented by shaded circles in Figure 2a and by shaded arrows in Figure 2b. The shown dimensions were obtained by direct measurement of the dual-tire assembly, with the diameter taken as the tire contact width. A Cartesian coordinate system is also included in this Figure, positioned in-between the two tires, with its origin located at the unloaded (undeformed) pavement surface (see Figure 2b). The $x$-axis points in the travel direction, the $y$-axis points in the transverse direction, and the $z$-axis points downward into the pavement medium.

Each time the TSD traversed the 430 m stretch of road, deflection-slope measurements from Lasers 1, 2 and 3 were collected, averaged over 10 m intervals, and ascribed to the interval center location. Based on these measurements the device provided an internally computed index, herein named $TSD_{300}$. This index has length units and resembles $SCI_{300}$; it denotes the difference between the deflection in-between the dual-wheels, i.e., at the origin of the coordinate system, and the deflection under Laser 3, i.e., at $x = 300 \text{ mm}$.

For computing $TSD_{300}$ an analytic two-dimensional model is employed by the device (23-25). According to this model the pavement system is represented by an infinitely long and weightless Euler-Bernoulli (EB) beam supported on a Winkler foundation. A vertical point-force
is employed to represent the loading of the dual-wheels assembly. Considering a rectangular coordinate system such that the \( x \)-axis coincides with the undeformed EB beam and the \( z \)-axis points downward towards the support, the point-force appears acting at the origin. Based on this description, the EB beam deflection-slope \( u'_z \) at any offset distance \( x \) from the applied force, is given by:

\[
u'_z = -A \sin(Bx)e^{-Bx} \quad (x \geq 0)
\]

wherein the constants \( A \) (unitless) and \( B \) (units of length\(^{-1}\)) are both positive, representing model parameters, such as: EB beam properties, Winkler foundation properties, and loading intensity. The numerical values of these constants are found by best-matching the three measured TSD deflection-slopes with calculated EB beam slopes at three offset distances: 100, 200 and 300 mm.

Once \( A \) and \( B \) are obtained, the EB beam deflection \( u_z \) at any offset distance \( x \) is given by the expression:

\[
u_z = \frac{A}{2B} e^{-Bx} \left( \sin(Bx) + \cos(Bx) \right) \quad (x \geq 0)
\]

from which the TSD\(_{300}\) index is derived as follows:

\[
\text{TSD}_{300} = u_z(x = 0) - u_z(x = 300\, \text{mm})
\]

This expression, consistent with all abovementioned mechanistic assumptions and modeling simplifications, is based on the premise that maximal deflection and zero slope occur at \( x = 0 \). In actuality, the deflection bowl generated by a moving load over a viscoelastic pavement system is not symmetric, and the maximal vertical surface displacement is expected to take place behind the axle, i.e., where \( x < 0 \) (26,27).

Figure 3 shows TSD\(_{300}\) values measured during Step 3 of the field testing campaign (refer to Figure 1). Figures 3a, 3b and 3c present, respectively, measurements collected at 40, 60, and 70 km/h. For each speed there are 44 data points from Run#4 (hollow markers) and 44 data points
from Run#5 (shaded markers). Included in the charts are average $TSD_{300}$ values (denoted as $Y$) and corresponding standard deviations (denoted as $\sigma_Y$).

**FWD Measurements**

As indicated by Figure 1, the FWD device was operated at intervals of 5 m or 10 m along the 430 m test road. The device used was Dynatest Model 8002 equipped with a load-plate radius of 150 mm and nine geophones at the following offsets: 0, 200, 300, 450, 650, 900, 1200, 1500, and 1800 mm. Four separate FWD drops were executed in every test location, with only the last three deflection basins employed for analysis. In actuality, peak applied loads were very close to 50 kN; nonetheless, deflection peaks were linearly normalized to correspond to a peak load of exactly 50 kN. The admissibility of this procedure is based on the premise that nonlinear effects w.r.t. loading level are small.

**COMPARISON METHODOLOGY**

The purpose here is to compare the level of similarity between TSD and FWD measurements. Since the same exact road stretch was tested, results are expected to be similar, but only if a common pavement index is employed for the comparison. Subsequently, given that the TSD device reports $TSD_{300}$ index values at 10 m intervals, the comparison approach commences by constructing/calculating $TSD_{300}$ values from FWD measurements, also at 10 m intervals. Once this is completed, statistical tools become applicable to judge closeness of the datasets.

**Analysis of FWD Deflections**

The analysis presented hereafter aims at calculating $TSD_{300}$ from FWD measurements. To achieve this there is a need to consider and account for the dissimilarity in loading configuration between the two devices. FWD deflections are the result of loading a single circular plate with a peak force
of 50 kN. On the other hand, the TSD involves a dual-wheels assembly loaded (nominally) to 50 kN, with a center-to-center tire spacing of 335 mm (see Figure 2).

Subsequently, the idea advocated here is to virtually place a second FWD plate with a spacing of 335 mm (center-to-center) as means of imitating the TSD loading configuration. This is done assuming a linear pavement response for which superposition is applicable. With reference to Figure 2, such an analysis requires accessing FWD peak deflections at two nonstandard offsets. The first offset is 167.5 mm (=335/2), corresponding to the midpoint between the dual-tires. The second offset is 343.6 mm (=\sqrt{167.5^2+300^2}) corresponding to the point under Laser 3. The difference between these two deflections is deemed equivalent to the TSD\textsubscript{300} index.

For each FWD drop, nine FWD deflection peaks \(d_0...d_8\) were measured, corresponding to nine offset distances \(x_0...x_8\), with \(x_0 = 0\) and \(x_{i+1} > x_i\). A continuous and smooth deflection basin function \(w(x)\) was defined by means of a piecewise cubic interpolant passing through the \([x_i,d_i]\) pairs:

\[
w(x) = \begin{cases} 
  w_1(x) & x_0 \leq x < x_1 \\
  w_2(x) & x_1 \leq x < x_2 \\
  \vdots & \vdots \\
  w_8(x) & x_7 \leq x \leq x_8 
\end{cases} \quad (4)
\]

wherein \(x\) is the lateral (or offset) coordinate with origin under the load-plate center, and:

\[
w_i(x) = d_{i-1} + B_i(x-x_{i-1}) + C_i(x-x_{i-1})^2 + D_i(x-x_{i-1})^3 \quad (5)
\]

The constants \(B_i\)'s, \(C_i\)'s and \(D_i\)'s are spline constants, determined according to the usual requirements, i.e., \(w_i(x_i) = d_i\), \(w_i'(x_i) = w_{i+1}'(x_i)\), \(w_i''(x_i) = w_{i+1}''(x_i)\), and \(w_8''(x_8) = 0\). To enforce zero deflection-slope under the load, i.e., \(w_i'(x_0) = 0\), a condition that \(B_i = 0\) was also be included in the formulation. After defining \(\Delta d_i = d_i - d_{i-1}\) and \(\Delta x_i = x_i - x_{i-1}\), the resulting equation set for obtaining the spline constants is:
For each measured deflection basin in Steps 2 and 4 of the field testing campaign (see Figure 1), Equation 6 was solved, spline constants were determined, and a continuous deflection function was generated according to Equation 4. Then, the two sought nonstandard deflections were found by evaluating \( w(x = 167.5\, \text{mm}) \) and \( w(x = 343.6\, \text{mm}) \), and \( \text{TSD}_{500} \) was calculated as the difference between the two. The above description is graphically presented in Figure 4 for a randomly chosen deflection set.

The aforementioned calculations were carried out for all test locations using the last 3 drops out of the 4 executed FWD drops. The three calculated \( \text{TSD}_{500} \) values were then averaged to arrive at a representative index value for the test location. Ultimately, \( \text{TSD}_{500} \) from FWD were reported at intervals of 10 m along the experimental road stretch to correspond to the reported TSD results. For this purpose, FWD tests taken 5 m before and after the location of interest were factored-in with a relative weight of 25%. For example, the \( \text{TSD}_{500} \) from FWD at location 180 m was composed of 50% of the index measured at location 180 m, plus 25% of the index measured at location 175 m plus 25% of the index measured at location 185 m. Doing so somewhat imitates the TSD device, wherein continuous data were averaged over 10 m long road sections. The final set of \( \text{TSD}_{500} \) values calculated from FWD according to the above-described procedure is shown in Figure 3d. This chart includes a total of 44 values serving as reference for the corresponding

\[
\begin{bmatrix}
\Delta x_1^2 & \Delta x_1^3 \\
2\Delta x_1 & 3\Delta x_1^2 & -1 \\
2 & 6\Delta x_1 & -2 \\
\Delta x_2 & \Delta x_2^2 & \Delta x_2^3 \\
1 & 2\Delta x_2 & 3\Delta x_2^2 & -1 \\
2 & 6\Delta x_2 & -2 \\
\ldots & \ldots & \ldots & \ldots \\
\Delta x_7 & \Delta x_7^2 & \Delta x_7^3 \\
1 & 2\Delta x_7 & 3\Delta x_7^2 & -1 \\
2 & 6\Delta x_7 & -2 \\
\Delta x_8 & \Delta x_8^2 & \Delta x_8^3 \\
2 & 6\Delta x_8 & -2 \\
\end{bmatrix}
\begin{bmatrix}
C_1 \\
D_1 \\
B_2 \\
C_2 \\
D_2 \\
\vdots \\
B_7 \\
C_7 \\
D_7 \\
B_8 \\
C_8 \\
D_8 \\
\end{bmatrix} = \begin{bmatrix}
\Delta d_1 \\
0 \\
0 \\
\Delta d_2 \\
0 \\
\vdots \\
0 \\
\Delta d_7 \\
0 \\
0 \\
\Delta d_8 \\
0 \\
\end{bmatrix}
\]  

(6)
TSD\textsubscript{300} values measured by the TSD device during any run (with a given speed). Also included in the Figure are the average (denoted as \( \bar{X} \)) and standard deviation (denoted as \( \sigma_X \)). It is clear from Figure 3 that the TSD results exhibit a larger spread (higher variability) as compared to the FWD derived results. If both devices had identical measurement resolution and accuracy, and given that the TSD provides readings averaged over 10 m intervals, higher variability would be expected in the FWD measurements.

Visualization of Comparison Statistics

In what follows, \( X_n \)'s denote values of TSD\textsubscript{300} derived from FWD (Figure 3d) with \( n = 1 \ldots N \) and \( N = 44 \). Denoted as \( Y_n \)'s are values of TSD\textsubscript{300} directly reported by the TSD device (Figures 3a, 3b and 3c), under a given measurement speed (40, 60, or 70 km/h) and for a given Run (#4 or #5).

A popular statistical tool for comparing two datasets is the Pearson product-moment correlation coefficient:

\[
r = \frac{\sum_{n=1}^{N} (X_n - \bar{X})(Y_n - \bar{Y})}{N \sigma_X \sigma_Y}
\]

(7)

wherein \( \bar{X} \) and \( \bar{Y} \) are the respective means of \( X_n \) and \( Y_n \), while \( \sigma_X \) and \( \sigma_Y \) are the corresponding standard deviations. The \( r \) statistic is a dimensionless metric that indicates the degree or strength of the linear dependence between the \( X_n \)'s and the \( Y_n \)'s. Values of \( r \) range between -1 to 1, and describe both the linear relationship strength and the linear relationship direction.

Scatterplots of \( X_n \) vs. \( Y_n \) are included in Figure 5 for the different TSD runs, along with the associated Pearson correlation coefficient. As can be seen, in all considered cases a positive correlation was attained, with \( r \) values in the range of 0.335 to 0.698. All \( r \) values were subjected to a statistical significance test to assess whether that they are larger than zero. This significance test was carried out assuming \( r \) follows a Student's \( t \)-distribution with \( N - 2 \) degrees of freedom. Accordingly, \( t \)-values were calculated in each case with the expression: \( r \sqrt{\frac{N - 2}{(1 - r^2)}} \) and
then the significance level was obtained from the associated $t$-distribution based on a one-tailed check. All $r$ values for the data shown in Figure 5 were confirmed to be positive, with a statistical significance that is better than 1.3%. The verbal descriptions appearing in the charts (i.e., weak, moderate, and strong) are based on the categorization suggested in Evans (28). This categorization, see Table 1, is context dependent; it is offered here as one possible means for interpreting correlation magnitudes.

The Pearson correlation coefficient is scale-blind as it does not convey information on the deviation magnitudes. A popular agreement metric that overcomes this drawback, and considers dimensional information, is:

$$E = \sqrt{\frac{\sum_{n=1}^{N} (X_n - Y_n)^2}{N}}$$ (8)

in which $E$ is the root-mean-square error (RMSE); it essentially aggregates the squared (individual) differences between $X_n$ and $Y_n$ across the population. In effect, two different error types are represented in $E$, namely: pattern mismatch ($E'$) and mean mismatch ($\bar{E}$). These two error types may be decomposed according to the formula $E^2 = \bar{E}^2 + E''^2$ wherein $\bar{E} = \bar{Y} - \bar{X}$ and $E'$ is the centered RMSE metric:

$$E' = \sqrt{\frac{\sum_{n=1}^{N} ((X_n - \bar{X}) - (Y_n - \bar{Y}))^2}{N}}$$ (9)

The standard deviations ($\sigma_x$ and $\sigma_y$) and the Pearson product-moment correlation coefficient ($r$) are related to the centered RMSE metric ($E'$) according to the expression $E''^2 = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y$. This expression bears similarity to the trigonometric law of cosines; it led Taylor (21) to propose a diagram that simultaneously depicts these different statistics in a single chart. Shown in Figure 6 is such a Taylor diagram, representing data from Figures 3 and 5.

In this diagram, standard deviations are depicted on both the abscissa and ordinate while the Pearson correlation coefficient $r$ is represented by azimuthal positions w.r.t. the coordinate
origin such that the abscissa coincides with $r = 1$ (i.e., perfect linear correlation) and the ordinate coincides with $r = 0$ (i.e., no linear correlation). Points representing the different TSD runs are shown in the chart: triangular markers represent measurements collected at 40 km/h, square markers represent a measurement speed of 60 km/h, and measurement speed of 70 km/h is represented by circular markers. The distance of these points from the coordinate origin represents their standard deviation and is indicated by dashed arcs. The reference FWD dataset is depicted with a solid circular marker on the abscissa at a standard deviation $\sigma_x = 4.57 \mu m$ (refer to Figure 3d). Radial distances from the reference point are indicated by dotted arcs and represent centered RMSE metric values (Equation 9).

Overall, the Taylor diagram makes it easy and intuitive to assess similarity levels between TSD and FWD measurements. In the specific examples shown, three features are immediately apparent from the graphical depiction: (i) the standard deviations of the $TSD_{300}$ values (dashed arcs), residing between 8\(\mu \text{m}\) and 10\(\mu \text{m}\), exhibit a slight tendency to increase with increasing measurement speed; (ii) the centered RMSE metric values (dotted arcs) appear insensitive to TSD measurement speed; and (iii) the correlation between TSD and FWD (azimuthal positions) somewhat improves as TSD measurement speed increases.

**Modern Agreement Metric**

The Taylor diagram depicts several centered pattern error statistics. It does not, however, offer a single similarity metric needed for performing inter-comparison or for ranking different cases. To this end, it is instructive to introduce a modern agreement metric $\lambda$, suggested originally in the field of Biometrics for evaluating method reproducibility (29), and highlighted recently by Duveiller et al. (22). The $\lambda$ metric summarizes the closeness of two datasets in a single index, considering both correlation and bias. It is dimensionless, symmetric, and relatively 'cheap' to compute. For positive Pearson correlation coefficient values it is given by the expression:
\[
\lambda = \frac{2r}{\sigma_x + \sigma_y + (\bar{X} - \bar{Y})^2} \quad (r \geq 0)
\]

(10)

where the purpose of the denominator is to ‘penalize’ the value of \( r \) for increasing additive bias or multiplicative bias (or both) between the datasets. As can be seen, if two datasets \( X_n \) and \( Y_n \) have identical means and standard deviations, then the \( \lambda \) metric becomes equal to the Pearson correlation coefficient. In all other cases, for which the means or standard deviations (or both) are dissimilar, the \( \lambda \) metric is smaller than \( r \). Therefore, \( \lambda \) in Equation 10 ranges from zero to unity.

Calculated \( \lambda \) values are included in the charts in Figure 5 just beneath the \( r \) values. As expected, they are different and always smaller than \( r \). Based on these \( \lambda \) values Table 2 ranks the agreement level between FWD and TSD across the different datasets collected in the field experiment; it also presents agreement rankings based on the Pearson correlation coefficient.

As can be seen in Table 2, the rankings differ. The highest level of \( \lambda \) is reported for TSD Run#5 at 60 km/h (\( \lambda = 0.393 \)). If only correlation level was taken as an agreement criterion then Run#4 at 70 km/h (\( r = 0.698 \)) would be ranked as best. The lowest level of \( \lambda \) is reported for TSD Run#4 at 60 km/h (\( \lambda = 0.255 \)). If only correlation level was taken as an agreement criterion then Run#4 at 40 km/h (\( r = 0.335 \)) would be ranked as worst. Given that \( \lambda \) embodies both correlation and bias, rankings based on this metric are deemed superior and more appropriate than rankings based on \( r \). It may also be noticed from Table 2 that the three top ranked agreement levels refer to the higher measurement speeds.

**SUMMARY AND COMMENTS**

The introduction of new pavement evaluation devices require comparison against accepted methods so that existing knowledge and experience can remain useful - at least during some transition period. In this connection, the paper was concerned with comparing TSD and FWD measurements taken along the same road stretch. The work focused on the analysis approach rather
than on the comparison results. This is because comparison outcomes depend on the specific TSD and FWD devices and operational modes employed during data collection. Comparison outcomes also depend on the tested pavement conditions because the quality of measured values, especially if small, may be influenced by finite sensing resolution.

Herein, comparison was based on a deflection index named $TSD_{300}$ which is internally calculated, and subsequently reported, when operating the TSD device. The first interpretation step involved estimating $TSD_{300}$ from FWD, with the results serving as a reference/benchmark dataset for subsequent evaluation. This was done as an initial attempt to account for the differences in loading configuration between the devices. The estimation of $TSD_{300}$ from FWD presumed linear quasi-static pavement response, and considered only load and deflection peaks.

The second interpretation step herein employed involved a Taylor diagram as a tool for visualizing dissimilarity between datasets. Finally, a modern metric $\lambda$ was proposed for quantifying the closeness of two datasets and therefore rank calculated agreement levels across different cases. Both statistical tools are generic and therefore applicable for evaluating other/different datasets. In particular, they appear to be highly suited for displaying trends in agreement resulting from employing better/improved interpretation algorithms. Similarly, they should also be effective in demonstrating the merits of changing operational settings or modifying data collection procedures. Moreover, although not the focus herein, the $\lambda$ metric is well suited for quantifying device reproducibility ($30$), e.g., the similarity level between two TSD run results. Hence, the proposed methodology is deemed highly suited for the task of quantifying the value of forthcoming device upgrades.

In future work, more sophisticated efforts may be envisioned to enhance comparison validity wherein systematic differences between devices are further reduced prior to any statistical assessment. The following aspects related to loading conditions should be considered as candidate for interpretation improvement: (i) area of loaded zones, i.e., diameter of FWD load-plate vs. TSD tire imprint size; (ii) contact stress distribution uniformity; (iii) type of loading, i.e., FWD impact
vs. TSD rolling wheels; and (iv) oscillation of TSD load due to vehicle dynamics in combination with pavement roughness. Moreover, the following pavement modeling aspects should be considered: (i) employ identical mechanical model for FWD and TSD measurements when estimating $TSD_{300}$, e.g., abandon the EB beam model; and (ii) employ more faithful mechanical representation of the pavement layers, e.g., nonlinear response, time-dependence, thermal-sensitivity, and inertia effects. The above list of proposed ideas for reducing systematic differences (prior to agreement testing) implicitly advocate using mechanical layer properties as basis for comparison - given that they are device independent.

ACKNOWLEDGEMENT

The authors would like to acknowledge the Measurement Division of the Danish Road Directorate for their help in carrying out the FWD and TSD measurements. Also acknowledged is Mr. Alexander Schimmer Molin for his contribution in identifying relevant literature on comparative statistics.
REFERENCES


LIST OF TABLE TITLES AND FIGURE CAPTIONS

TABLE 1 Verbal interpretation of Pearson correlation offered for comparing TSD and FWD data

TABLE 2 Ranking of TSD and FWD agreement level based on the $\lambda$ metric (Equation 10) and based on the Pearson correlation coefficient $r$ (Equation 7)

FIGURE 1 Testing campaign for generating TSD and FWD comparison data.

FIGURE 2 TSD loading and measurement setup: (a) plan view, and (b) side view.

FIGURE 3 Values of $TSD_{300}$ index reported at 10 m intervals for the same tested road: (a) acquired from TSD device at 40 km/h, (b) acquired from TSD device at 60 km/h, (c) acquired from TSD device at 70 km/h, and (d) acquired from interpretation of FWD deflections.

FIGURE 4 Calculation of $TSD_{300}$ from FWD by means of spline interpolation; the $d_i$’s (rotated square markers) are measured FWD peak deflections normalized to a peak force of 50 kN; the $w_i$’s (solid line), are components of a fitting spline curve with $x$ as the offset coordinate (see Equation 4).

FIGURE 5 Cross-plots of measured $TSD_{300}$ data versus $TSD_{300}$ obtained from FWD testing.

FIGURE 6 Taylor diagram of $TSD_{300}$ index statistics for comparing FWD and TSD measurements at three different speeds: 40 km/h (triangular markers), 60 km/h (square markers), and 70 km/h (circular markers).
TABLE 1 Verbal interpretation of Pearson correlation offered for comparing TSD and FWD data

<table>
<thead>
<tr>
<th>Correlation range</th>
<th>Verbal description (27)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00 ↔ 0.19</td>
<td>Very weak</td>
</tr>
<tr>
<td>0.20 ↔ 0.39</td>
<td>Weak</td>
</tr>
<tr>
<td>0.40 ↔ 0.59</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.60 ↔ 0.79</td>
<td>Strong</td>
</tr>
<tr>
<td>0.80 ↔ 1.00</td>
<td>Very strong</td>
</tr>
</tbody>
</table>
TABLE 2 Ranking of TSD and FWD agreement level based on the $\lambda$ metric (Equation 10) and based on the Pearson correlation coefficient $r$ (Equation 7)

<table>
<thead>
<tr>
<th>TSD Run #</th>
<th>Speed [km/h]</th>
<th>Agreement ranking based on $\lambda$</th>
<th>Agreement ranking based on $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>40</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
FIGURE 1 Testing campaign for generating TSD and FWD comparison data.
FIGURE 2 TSD loading and measurement setup: (a) plan view, and (b) side view.
FIGURE 3 Values of $TSD_{300}$ index reported at 10 m intervals for the same tested road: (a) acquired from TSD device at 40 km/h, (b) acquired from TSD device at 60 km/h, (c) acquired from TSD device at 70 km/h, and (d) acquired from interpretation of FWD deflections.
FIGURE 4 Calculation of $TSD_{300}$ from FWD by means of spline interpolation; the $d_i$'s (rotated square markers) are measured FWD peak deflections normalized to a peak force of 50 kN; the $w_i$'s (solid line), are components of a fitting spline curve with $x$ as the offset coordinate (see Equation 4).
FIGURE 5 Cross-plots of measured $TSD_{300}$ data versus $TSD_{300}$ obtained from FWD testing.
FIGURE 6 Taylor diagram of TSD$_{300}$ index statistics for comparing FWD and TSD measurements at three different speeds: 40 km/h (triangular markers), 60 km/h (square markers), and 70 km/h (circular markers).