Field Experiment for Accuracy Verification of the Devices Measuring International Roughness Index in TRUE Project: Five Years Report

K Tomiyama1*, H Nakamura2, H Mashito3, M Jomoto4 and S Kameyama5

1Division of Civil and Environmental Engineering, Kitami Institute of Technology, 165 Koen-cho, Kitami, 090-8507, Japan
2NIPPO Corporation, 6-70 Mitsuhashi, Nishi-ku, Saitama, 331-0052, Japan
3Toa Road Corporation, 315-126 Kaname, Tsukuba, 300-2622, Japan
4Taisei Rotec Corporation, 1456 Oaza-Kamiya, Konosu, 365-0027, Japan
5Hokkaido University of Science, 7-Jo 15-4-1 Maeda, Teine, Sapporo, 006-8585, Japan

*Corresponding author: tomiyama@mail.kitami-it.ac.jp

Abstract. A lot of devices measuring the International Roughness Index (IRI) has recently been developed and applied in response to the pavement inspection purposes. Consequently, the accuracy of measured IRIs obtained with different devices on the same road surface has become a practical concern for providing a consistent inspection result. Against this background, the specified non-profit organization Pavement Diagnosis Researchers Group (PDRG) in Japan has undertaken a project regarding the experiment to harmonize and compare Test methods for surface Roughness Under actual road Environment (TRUE project) since 2014. The TRUE project has so far conducted the roughness measuring experiments three times over five years at Hokkaido, Japan. This paper introduces the features of the TRUE project such as participated devices, test site selection, and data recording and reporting procedures. A benchmark testing for obtaining reference profiles is also provided to make the project reliable. In this paper, the results of experiments conducted in the last five years are statistically described as well. Finally, this paper introduces the potential of the TRUE project such as a new device grouping and the accuracy evaluation report for device certification for contributing to further improvement of a device participated in the project.

1. Introduction

Against the background that pavement infrastructures have aged, the Bureau of Public Road (BPR) of Ministry of Land, Infrastructure, Transport and Tourism (MLIT), Japan has issued a strategy for inspection of the whole road stock and has provided a tentative implementation guide for evaluating roughness by the International Roughness Index (IRI) since 2013. Then, BPR has officially published the guidelines for pavement inspection using IRI since 2016 [1]. According to the guidelines, road administrators are obliged to maintain their pavements effectively and efficiently with establishing a maintenance cycle consisted of inspection, diagnosis, treatment, and recording during the service period of a pavement. These strategies have accelerated the development of a large variety of roughness measuring devices corresponding to the needs for inspection purposes [2].

On this account, the reliability of the devices for obtaining the same result with different devices on the same profile has become an important matter to monitor the surface condition periodically in a life
interval of pavements. Against this background, the specified non-profit organization Pavement Diagnosis Researchers Group (PDRG) in Japan has launched the experiment to harmonize and compare Test methods for surface Roughness Under actual road Environment (TRUE project) since 2014. A purpose of the project is to improve technologies of surface measurement under actual road environment by (a) supporting the experiment implementation, (b) analyzing the data obtained in experiments, and (c) reporting and publishing the outcomes of activities. PDRG has so far conducted roughness measuring experiments three times in the last five years as part of the TRUE project to verify the accuracy of devices measuring IRI. This paper provides the basic idea of the TRUE project and describe the results of three experiments. Not all of the devices used in Japan, but this study includes a number of them. The information presented in this paper is intended to provide a benchmark of the reliability for associating different devices.

2. Feature of TRUE project

2.1. Participated devices

One of the well-known experiments to harmonize roughness measurement devices was performed as part of the EVEN project by the Permanent International Association of Road Congresses - World Road Association (PIARC), which was conducted in the United States, Japan, and Europe [3-5]. Seven devices participated in the second experiment of the project conducted in Japan in 1998 [4]. After 16 years, in the TRUE project, thirty-four devices including high- and low-speed devices participated in the first experiment in 2014. Following the first experiment, twenty-eight devices took part in the second and third experiment held in 2016 and 2018, respectively. Table 1 summarizes the number of the participated devices. An example of the devices is shown in figure 1. This enthusiastic participation in every experiment infers a great concern and demand for the roughness inspection and its methods. In addition, after the profile measurement, the project investigated the structural properties of the pavements by a Falling Weight Deflectometer (FWD).

| Device Type          | Date         | FY 2014 (Sep. 17-19, 2014) | FY 2016 (Sep. 14-16, 2016) | FY 2016 (Oct. 12-14, 2018) | Total |
|----------------------|--------------|-----------------------------|-----------------------------|-----------------------------|-------|
| High-speed Devices   |              | 20                          | 15                          | 12                          | 47    |
| Low-speed Devices    |              | 14                          | 13                          | 16                          | 43    |
| Total                |              | 34                          | 28                          | 28                          | 90    |

Figure 1. An Example of Participated Devices: (a) Dipstick, (b) Inertial Profiler, (c) Mobile Mapping System (MMS), (d) Walking Profiler.
2.2. Test Section
The experiments of the TRUE project have been performed on prefectural roads with the cooperation of Hokkaido prefecture of Japan. Two test sites categorized into arterial (high volume) and residential (low volume) roads including two sections for each site have been established in consideration of the variety of roughness level. Each test section is 200 m long with 20 m and 5 m additional extents from the beginning and end points, respectively. Table 2 summarizes specifications of the test sites with the IRIs calculated from the reference profiles called the true profiles by the standard Golden-Car simulation. Only in the second experiment, one extra section was added in the Site 1 for further analyses, however, this paper does not deal with this section.

| Overview |
|----------|

| Table 2. Specifications of the Test Sites |
|------------------------------------------|
| Site | 1 | 2 |
| Road Category | Arterial (High Volume) Road | Residential (Low Volume) Road |
| Driving Speed (km/h) | 40, 50, and 60 km/h | 20, 30, and 40 km/h |
| Length (m) | 200 | 200 |

2.3. Data Recording and Reporting Procedures
In the experiment, participated devices made three repeated runs for each test section. The high-speed devices were also required to obtain data at the following different operating speeds: 40 km/h, 50 km/h, and 60 km/h on Site 1 and 20 km/h, 30 km/h, and 40 km/h on Site 2. The participants were asked to report IRI and profile measurements with data format of “.csv” and “.xlsx”, respectively, in accordance with specified formats of the project. The IRI data were reported with a fixed interval of 200 m and a minimum step available for each device. The TRUE project received profile and IRI measurements obtained with high-speed profilers (vans, trucks, cars, etc.) and hand-operated low-speed profilers, and IRI measurements obtained with response type systems after the experiment.

3. Reference "True Profile" Measurement

3.1. Measurement procedure
A reference profile called "true profile" for each test section was accurately measured by the same manner for every experiment. A true profile consists of the combination of measurements obtained with three different reference devices such as a Rod and Level, a static Dipstick, and a hand-operated profiler as shown in Figure 2 with basic three steps described in the following subsections.

3.1.1. Step 1. The static Dipstick measures the difference in height which is sensitive to the IRI calculation at an interval of 0.25 m.
3.1.2. Step 2. The Rod and Level measurements compensate longer wavelength ranges than the static Dipstick at an interval of 10 m.
3.1.3. **Step 3.** The hand-operated profiler interpolates shorter wavelength ranges than the static Dipstick at an interval of no longer than 0.05 m.

![Figure 2](image.png)

**Figure 2.** A Schematic Illustration of Measuring a True Profile

3.2. **Benchmark testing**

A benchmark device has been tested to validate the true profiles obtained with the above-mentioned procedure prior to the experiment. As a benchmark testing, we applied three levels of Rod and Level devices and total station (TS) as shown in table 3 to provide a ground truth. Figure 3 shows how do they work. In the testing, a longitudinal profile was measured for 105 m long at an interval of 0.05 m by using each device. Although the most reliable device in table 3 is Digital Level of level 1, it is inconvenient considering workability on residential roads. Thus, a benchmark device was selected from the remaining devices by assuming the Digital Level of level 1 as the reference.

| Device                        | Level | Name/Resolution       | Resolution |
|-------------------------------|-------|-----------------------|------------|
| Digital Level                 | 1     | DiNi 0.3 (Trimble)/    | 0.01 mm    |
| Bar-code Levelling Staff      | 1     | DL-502 (TOPCON)/      |            |
| Digital Level                 | 2     | AT-M3 (TOPCON)/       |            |
| Bar-code Levelling Staff      | 2     |                       | 0.1 mm     |
| Auto Level                    | 3     | TS15 1" (Leica)/      | 0.1 mm     |
| Levelling Staff               | -     |                       | 1 mm       |
| Total Station                 | 1     |                       |            |

![Figure 3](image.png)

**Figure 3.** Candidate Devices for Benchmark Testing: (a) Digital Level of Level 1 and Level 2 and (b) Auto Level of Level 3 and Total Station (TS) of Level 1.
Figure 4 indicates the result of benchmark testing compared in terms of (a) longitudinal profile, (b) amplitude gain, (c) profile agreement [6], and (d) IRI. As shown in figure 4 (a) for the profile comparison, Auto Level of level 3 had insufficient resolution for the elevation measurement as the profile shape was not smooth. This result negatively reflects to the amplitude gain, profile agreement for the roughness profile, and IRI as shown in figure 4 (b), (c), and (d), respectively. The testing result also indicates that TS is inaccurate on the profile agreement and IRI as shown in figure 4 (c) and (d), respectively. This negative result is due mainly to the heat haze on the ground. In contrast, Digital Level of level 2 realize well correlated result with Digital Level of level 1 as shown in figure 4 with less operation effort. Thus, in the TRUE project, we decided to use Digital Level of level 2 which has the resolution of 0.1 mm for elevation measurement as a benchmark device.

![Figure 4. Result of the Benchmark Testing for Comparing (a) Longitudinal Profile, (b) Amplitude Gain, (c) Profile Agreement, and (d) IRI.](image)

4. Statistical Description of Experiment Result
A standard practice for analyzing data in the TRUE project statistically describes the influence of operating speed, repeatability as precision, and reproducibility as well as portability of the measurements as bias in terms of the IRI. Table 4 summarizes the viewpoints of the analysis such as the type of error (random or systematic), comparison factor (within or between devices), comparison index (mean or standard deviation), and data averaging method.

4.1. Influence of Operating Speed
Since the influence of operating speed is assumed to be a systematic error within a device, it can be evaluated by the standard deviation of IRI data obtained by the different operating speeds. The high-speed devises made three repeated runs at the specific speeds on each site for assessing the influence of operating speed on the measurement result. The operating speeds were 40 km/h, 50 km/h, and 60 km/h on Site 1 and 20 km/h, 30 km/h, and 40 km/h on Site 2. Note that the operating speeds on Site 2 in 2018 were only 20 km/h and 30 km/h due to the environmental problem around the test site. Figure 5 shows the standard deviation of IRI obtained at different speeds on each site in each year. As shown in the figure, the mean of the standard deviation of IRI for the influence of operating speed increases with increasing IRI values, which is similar trend in 2014 and 2016. However, no relationship was observed in 2018 especially for Site 2. This result is caused by which a wide variety of devices participated in the experiment in 2018. The influence of operating speed was within 10% precision of...
the measured IRI values. For example, the influence tends to be less than 0.2 mm/m when the measured IRI is 2.0 mm/m.

Table 4. Viewpoints of Data Analysis

| Term of Description             | Type of Error | Comparison Factor | Comparison Index   |
|---------------------------------|---------------|-------------------|--------------------|
| Influence of Speed              | Systematic    | Within            | Standard Deviation |
| (only for high-speed devices)    | An ability to repeat the measures on different operation speeds | Deviation from the average obtained with repeated runs |
| Repeatability                   | Random        | Within            | Standard Deviation |
| (Precision)                     | An ability to repeat the measures with a same profiler | Deviation from the average obtained with repeated runs |
| Reproducibility and Portability | Systematic    | Between           | Mean Value         |
| (Bias)                          | An ability to repeat the measures with a different profiler | Deviation from the average obtained with an expected value |

Figure 5. Influence of Operating Speed for High-speed devices in FY 2014, FY 2016, and FY 2018.

4.2. Repeatability

Repeatability is the ability to obtain repeat measures with the same device at the same time [7], which is the random error corresponding to the precision within the device. Repeatability can be evaluated by the standard deviation of IRI measurements obtained by repeated runs. Figure 6 indicates the repeatability of IRI measurements obtained with the high- and low-speed devices. As shown in the figures, the standard deviation for repeatability increases with increasing IRI values, which is similar trend in 2014 and 2016. However, no relationship was observed in 2018 especially for Site 2 due to the participation of a wide variety of devices. For the high-speed devices shown in figure 6 (a), the repeatability was within 10% precision of the measured IRI values. For example, the repeatability of high-speed devices tends to be less than 0.2 mm/m when the measured IRI is 2.0 mm/m. This trend is similar to the influence of operating speed. Thus, the repeatability corresponds to the influence of operating speed with respect to the within comparison factor.

As shown in figure 6 (b), the low-speed devices were more repeatable than the high-speed devices. This result can be easily anticipated because it is easier to trace the same line between repeated runs with low-speed devices than high-speed ones. The repeatability of low-speed devices except Site 1-2 in 2018 was within 1% precision of the measured IRI values. For example, the repeatability of low-speed devices tends to be less than 0.02 mm/m when the measured IRI is 2.0 mm/m.
4.3. Reproducibility and Portability

Reproducibility is the ability to repeat the measures with a different device of the same basic design [7] which is the systematic error corresponding to the bias of measurements between devices. According to the literature [7], for profilers, reproducibility is not of as much interest as portability that is the ability to repeat the measures with completely different profiler design. The standard for portability is generally the true profile. On this account, this study evaluates the portability of IRI measurements as shown in Figure 7. The figures indicate the percentile levels of the mean of measured IRIs for the repeated runs obtained with the participated devices corresponding to the reference IRIs obtained from the true profiles. The spread of a box-plot along horizontal axis denotes the reproducibility of measures. The degree of reproducibility increases with decreasing the spread. The deviation of a mean from a line of identity represents the portability of measures. The degree of portability increases with decreasing the deviation. This representation contributes to demonstrating the reliability of a candidate device in pavement inspection. For the high-speed devices, the 50th - 75th percentile devices were within the error of 10%, while some devices exceeded the error of 30%. On the other hand, the 25th – 50th percentiles of the low-speed devices were possible to achieve accurate measurement within the error of less than 10%.

5. Potential of the TRUE Project

5.1. Device Grouping

Roughness measurement devices have traditionally been categorized into the following four classifications on the basis of how directly their measures pertain to the IRI [8].

- (a) Class 4 – A roughness measure is not reproducible or stable with time, and can only be compared to IRI by subjective estimation.
- (b) Class 3 – A measure obtained from a response type system is calibrated to the IRI scale by correlation with reference measures from a Class 1 or 2 system.
- (c) Class 2 – A profile-based method is used that is reproducible and stable with time, and that is calibrated independently of other roughness measuring instruments.
- (d) Class 1 – A profile-based method similar to Class 2 is used. A profile-based measurement qualifies as a Class 1 measure if it is so accurate that further improvements in accuracy would not be apparent.

However, in recent years, a wide variety of roughness measuring devices have been developed and served for surface inspection activities. Thus, some newly developed devices are hard to be categorized into the traditional classifications. Against this background, we have proposed the new grouping criteria for roughness measurement devices by how the device measures IRI as shown in table 5 to promote the development of diverse devices.
Figure 7. Reproducibility and Portability of IRI Measurements in (a) High-speed Devices of FY 2014, (b) Low-speed Devices of FY 2014, (c) High-speed Devices of FY 2016, (d) Low-speed Devices of FY 2016, (e) High-speed Devices of FY 2018, and (f) Low-speed Devices of FY 2018.

Table 5. New Grouping Criteria for Roughness Measurement Devices

| Group | Requirements                              | Traditional Class |
|-------|-------------------------------------------|-------------------|
| I     | Subjective Visual Inspection/Ride Experience | On Vehicle        |
|       | Subjective Visual Inspection              | By Walk           |
| II    | Static                                    | Measuring Elevation Directly<sup>1</sup> |
|       |                                            | Measuring Elevation by Inclinometer<sup>2</sup> |
| III   | Profile-based Method Low-speed            | Non-contact Sensor(s)<sup>3</sup> |
|       |                                            | Contact Sensor(s)<sup>4</sup> |
|       |                                            | Dedicated Device(s) |
|       |                                            | Multi-purpose Device(s)<sup>5</sup> |
| IV    | Dynamic High-speed                        | Non-contact Sensor(s)<sup>3</sup> |
|       |                                            | Contact Sensor(s)<sup>4</sup> |
|       |                                            | Dedicated Device(s) |
|       |                                            | Multi-purpose Device(s)<sup>5</sup> |
| V     | Response Type                             | Non-contact Sensor(s)<sup>3</sup> |
|       |                                            | Contact Sensor(s)<sup>4</sup> |
|       |                                            | Dedicated Device(s) |
|       |                                            | Multi-purpose Device(s)<sup>5</sup> |
| VI    | Otherwise                                 | -                 |

<sup>1</sup> e.g. Rod and Level, <sup>2</sup> e.g. Dipstick, <sup>3</sup> e.g. Lase, Ultrasonic, etc., <sup>4</sup> e.g. Wheel(s), <sup>5</sup> e.g. Smartphone Device(s)
5.2. Accuracy Evaluation Report for Device Certification

As early mentioned, one of the purposes of the TRUE project is to improve technologies of surface measurement devices. In order to achieve this purpose, the project certified the accuracy of the devices participated in the latest experiment conducted in 2018 by publishing the evaluation report shown in figure 8. The report includes the evaluation result of accuracy for a device in terms of the profile agreement and IRI with respect to the repeatability, portability, and influence of operating speed when the high-speed device. Detailed values of IRI are also reported as an appendix. This report contributes to further improvement of a device participated in the project by receiving the feedback of accuracy evaluation based on the well-established true profiles.

![Accuracy Evaluation Report for Device Certification](image)

**Figure 8.** Accuracy Evaluation Report for Device Certification.

6. Discussion and Conclusions

According to the pavement inspection guidelines using IRI published by the BPR in 2016 in Japan, a lot of devices measuring the IRI has recently been developed and applied in response to the inspection purposes. Consequently, the accuracy of measured IRIs obtained with different devices on the same road surface has become a practical concern for providing a consistent inspection result. In the light of this background, a project regarding the experiment to harmonize and compare Test methods for surface Roughness Under actual road Environment (TRUE project) has been launched since 2014.

The TRUE project has so far conducted the roughness measuring experiments three times at Hokkaido, Japan in September 2014, September 2016, and October 2018. This paper has provided the features of the TRUE project such as participated devices, test site selection, and data recording and reporting procedures. The features have been kept through the experiments conducted since the last five years. To make the project reliable, a benchmark device was tested to validate the procedure for obtaining true profiles prior to the experiment. As a result, in the TRUE project, Digital Level of level 2 which has the resolution of 0.1 mm for elevation measurement was selected as a benchmark device.

This paper has also statistically described the results of three experiments conducted in the last five years. The analysis of the data obtained in the experiments has focused on repeatability and
reproducibility as well as portability of the IRI measurements obtained with the participated devices. The influence of operating speed on the IRI measurements has also been considered for the high-speed devices.

Finally, this paper has introduced the potential of the TRUE project that is a new device grouping and the accuracy evaluation report for device certification for contributing to further improvement of a device participated in the project.

References
[1] The Bureau of Public Road of Ministry of Land, Infrastructure, Transport and Tourism 2016 Pavement Inspection Guidelines (in Japanese)
[2] Tomiyama K, Nakamura H, Mashito H, Moriishi K, Jomoto M and Watanabe K 2018 Comparison of Different Roughness Measuring Devices in True Project Proc. 8th Symposium on Pavement Surface Characteristics (SURF2018)
[3] Schmidt B 2001 EVEN Project: Experiment to Compare and Harmonize Methods for Assessment of Longitudinal and Transverse Evenness of Pavement Transp. Res. Rec. No 1764 pp 221-231
[4] Kawamura A, Takahashi M and Inoue T 2001 Basic Analysis of Measurement Data from Japan in EVEN Project Transp. Res. Rec. No 1764 pp 232-242
[5] Descornet G, Berlemont B and Martin J M 2001 Study of Precision of Transverse Evenness Measurements in FILTER Experiment: Forum of European National Highway Research Laboratories Investigation on Longitudinal and Transverse Evenness of Roads Experiment Transp. Res. Rec.No 1764 pp 210-220
[6] Karamihas S M and Gillespie T D 2002 Development of Cross Correlation for Objective Comparison of Profiles, Publication UMTRI-2002-36 The University of Michigan Transportation Research Institute
[7] Sayers M W and Karamihas S M 1998 The Little Book of Profiling, - Basic Information about Measuring and Interpreting Road Profiles The University of Michigan
[8] Sayers M W 1990 Profiles of Roughness Transp. Res. Rec. No 1260 pp 106-111