Evaluating smartphone-based road roughness estimation systems in an urban area

Qiqin Yu[a], Yihai Fang[a], Richard Wix[b]
[a]Monash University, Australia
[b]Australian Road Research Board, Australia
qiqin.yu@monash.edu

Abstract. Road roughness indices indicate the condition and cost of road pavement. Traditional instruments are labour-intensive and limited in spatial coverage. Embedded sensors in smartphones are gaining popularity for estimating road roughness. Smartphone-based roughness index estimation (sRIE) systems were developed in the past years. However, there has been no objective evaluation of their performance under varying survey speeds, vehicle and mounting configurations in a metropolitan environment. This research selected three commercial sRIE Apps and tested them using a standardised evaluation framework. The results suggest that one App generates consistent results, but the accuracy of sRIE systems should be improved and be made more robust against varying practical settings.

1. Introduction

Roughness induces distress to the pavement surface and diminishes riding comfortability. Generally, pavement roughness assessment is needed to validate the quality of newly constructed pavement and monitor the existing pavement condition. The pavement roughness condition can be described using the international roughness index (IRI), a single roughness scale that enables the exchange of pavement roughness information internationally. The IRI is calculated by simulating the Quarter Car (QC) model traversing the road profile and accumulating the elevation deviation of the body (sprung mass) and the wheel (unsprung mass) over the travelled length (Sayers, 1995). Conventionally, measuring the IRI requires direct measurement of changes in road profile elevation, which involves expensive equipment and complex setup. Thanks to the prolificity and increasing sensing capabilities of smartphones, measuring pavement roughness using smartphone sensors has become a viable approach. Studies used the vehicle body response measured by a smartphone to estimate the road profile (Zhao et al., 2019), or trained machine learning models to directly estimate the IRI from the smartphone's response (Jeong et al., 2020). A detailed review of smartphone-based roughness index estimation (sRIE) methods is available in (Yu et al., 2022).

Studies evaluated the sRIE systems using various approaches. An sRIE was tested using a single vehicle type under different constant speeds (Wix, 2016). Similarly, two sRIE systems were evaluated by adopting constant speed and single vehicle type (Shah et al., 2017). In recent studies, varying speeds, mounting and vehicle types were introduced to mimic the real-world application environment of the sRIE system. For instance, sRIE systems were evaluated by adopting various mounting types under different constant speeds (Hossain et al., 2019), and using four different vehicle models and mounting locations (Botshekan et al., 2021). The commonly used statistic is the coefficients of the linear regression relationship between the smartphone and the reference instruments (Douangphachanh and Oneyama, 2014), while a direct spatial plot of the reference and smartphone measurements was applied to compare their performance (Xue et al., 2020). However, the existing studies applied their own evaluating statistical measures, which made the cross-comparison between different sRIE challenging, and as a result, there is no objective evaluation of the current sRIE systems. To address this gap, this study evaluates three commercial sRIE systems in a real-world application context, under the same configuration combinations, attributed by varying speed, vehicle model and mounting.

The rest of the paper is organised as follows. Section 2 elaborates on the adopted statistical measures. Section 3 demonstrates the experimental setup. Section 4 presents the results. Section 5 entails a discussion. Section 6 concludes the paper.

2. Evaluation statistical measures

sRIE systems are affected by surveying speed (Galagoda and Lanka, 2019), vehicle type (Islam et al., 2014), and mounting configuration (Bridgelall et al., 2019). The following statistical measures test the repeatability and accuracy performance of an sRIE system under the variation of these settings. Repeatability concerns the system's ability to produce the consistent smartphone-measured IRI (sIRI), while accuracy represents how close the results of the tested system are to the reference IRI (rIRI) values.

2.1 Repeatability test

The repeatability performance under each practical setting is reported using two statistical measures, namely the coefficient of variation (CoV) of each measuring segment and the R^2 of the $\mathrm{sIRI}_{\mathrm{mean}}$ vs $\mathrm{sIRI}_{\mathrm{individual}}$ linear regression model.

Coefficient of variation

The CoV of a segment measures the relative dispersion of the measurements around their mean (the ratio of standard deviation to the mean). This measures the variation of measurements in each segment.

For each segment, there is:

$$CoV = \frac{\sigma_n}{\overline{X_n}} \tag{1}$$

where

$\sigma_n = \sqrt{\frac{\sum_{i=1}^{N}(X_{ni}-\overline{X_n})^2}{N-1}}$; the sample standard deviation of measurements at the n$^{\mathrm{th}}$ segment

$\overline{X_n} = \frac{\sum_{i=1}^{N} X_{ni}}{N}$; the arithmetic mean of measurements at the n$^{\mathrm{th}}$ segment

N: the total number of repetitive runs

X_{ni}: the measurement on segment n from i$^{\mathrm{th}}$ run

CoV_{mean} is the average of the CoV of all segments in a route and indicates the system's performance on the entire route, and is calculated as:

$$CoV_{mean} = \frac{\sum_{n=1}^{n_s} CoV_n}{n_s} \tag{2}$$

where

CoV_n: the coefficient of variation of measurements at the n$^{\mathrm{th}}$ segment

n_s: total number of segments

Correlation with the mean measurements

In addition to the CoV_{mean}, the scatter plot of individual data points compared to the mean provides a visual representation of the consistency of the sRIE system. In this plot, the mean of the five measurements is plotted on the x-axis, while the five measurements from the smartphone are plotted on the y-axis. A linear regression model can be constructed between the individual sIRI values (dependent variable) and the mean of sIRI values (independent variable). The R^2 of the regression model is important to report, as it shows how closely individual measurements fit the regression line and explains the total residuals of the dependent variables to the model.

$$sIRI_{individual} = k_{(c,m)} \times sIRI_{mean} + b_{(c,m)} \tag{3}$$

where

$sIRI_{mean}$: the mean of IRI of five repetitive runs

$sIRI_{individual}$: the IRI of one particular run

$k_{(c,m)}, b_{(c,m)}$: regression coefficients for each vehicle (c) and mounting (m) setting

2.2 Accuracy test

Besides testing the repeatability, the sRIE system's measurement accuracy is vital. A quantitative accuracy test determines the proximity of the sRIE system measurements to the benchmark IRI derived from the reference tool. The accuracy results for each practical scenario are presented using two metrics, namely the average measurement error (ε_{mean}), and the R^2 of the sIRI vs rIRI linear regression model.

Average of measurement error

The average measurement error describes the difference between the sIRI and rIRI in an entire survey run. It is a percentage that directly reflects the accuracy of an sRIE system and is defined as:

$$\varepsilon_{mean} = \left| \frac{1}{n_s} \sum_{n=1}^{n_s} \frac{sIRI_{n_{mean}} - rIRI_n}{rIRI_n} \right| \tag{4}$$

where

$sIRI_{n_{mean}}$: Average of five sIRI on the nth segment

$rIRI_n$: reference IRI on the nth segment

n_s: total number of segments in a testing route

Correlation with the reference IRI

The scatter plot of sIRI data points versus the rIRI provides a visualisation of how sIRI measurements distribute with respect to the benchmark values. Using least squares regression, a linear regression model along with the coefficient of determination could be determined between the rIRI and the sIRI:

$$sIRI = k_{(c,m)} \times rIRI + b_{(c,m)} \tag{5}$$

where
$rIRI$: reference IRI of a segment
$sIRI$: smartphone measurements on a segment
$k_{(c,m)}, b_{(c,m)}$: regression coefficients for each vehicle (c) and mounting (m) setting

The coefficient of determination, or R^2, is a measure that provides information about the goodness of fit of a model. In the context of regression, it is a statistical measure of how well the regression line approximates the actual data.

3. Experiment

A field study was conducted to evaluate sRIE systems in an urban area. The selected route is from metropolitan Melbourne and has a total length of 10km. Specifically, the experiment adopted one reference instrument and three smartphone systems; two mounting locations and two vehicle types were tested. Five smartphone runs and three profiler runs were conducted.

Speed. A constant travelling speed was not maintained during the survey, as the route contains traffic lights. However, the operator maintained the speeds to a maximum of 60 and 80 km/h, respectively. Five runs were conducted at each speed.

Vehicle and mounting setup. Two vehicle models employed in the experiments are a Ford Ranger (denoted as U for ute) and a Volkswagen Golf (denoted as H for hatchback). The tested sRIE systems were mounted on both windshield (denoted as W) and dashboard (denoted as D), as shown in Figure 1(a).

(a) (b)

Figure 1 (a) the in-cabin set up and (b) the ground-truth survey vehicle

Reference instrument. The Australian Road Research Board (ARRB) inertial profiler was chosen to obtain the ground-truth road profile and IRI. The profiler is equipped on the survey vehicle, as shown in Figure 1(b).

4. Results

This section demonstrates the results of the experiment. To provide a direct illustration of the smartphone's results, the sIRI and rIRI results from one run are shown in Figure 2. The practical setting is "Dash| Windshield at 60 km/hr."

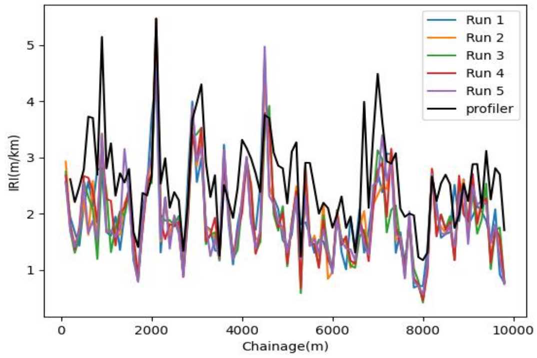

Figure 2 Plot of sIRI and rIRI measurements of Wellington Rd (App3)

4.1 Repeatability test

Coefficient of variation. The mean coefficient of variation (CoV_{mean}) represents the average of the coefficient of variation for all segments within a test route, demonstrating the relative spread of the five measurements around their average. The results under all practical settings are presented in Table 1. The CoV_{mean} sits in a range of 7.80 to 21.64, and a lower CoV_{mean} suggests a better repeatability performance. Among the three systems, App3 produced CoV_{mean} of approximately10 in practical settings.

Table 1 CoV_{mean} of three Apps

Speed (km/h)	Vehicle	Mounting	App1	App2	App3
60	H	D	14.158	16.743	12.042
		W	13.662	11.479	11.397
	S	D	12.063	20.511	7.804
		W	12.411	11.393	9.454
80	H	D	16.754	16.678	11.301
		W	15.152	17.599	13.575
	S	D	13.830	21.640	8.967
		W	13.766	12.816	9.089
		Mean	13.975	16.107	10.453

Correlation with the mean measurements. The R^2 of the linear regression model between the individual sIRI values (dependent variable) and the $sIRI_{mean}$ values (independent variable) demonstrates the measurement consistency. The results of the sIRI - $sIRI_{mean}$ regression are shown in Figure 3 to Figure 5, with each plot displaying measurements of sIRI vs $sIRI_{mean}$ under two survey speeds, differentiated by colours. The vehicle and mounting configuration is denoted on top and to the left of the plots. The axial is limited by an IRI value of 8 mm/m. “SciPy” was used for regression analysis with the coefficients k and β estimated by “least square estimates”, minimizing the sum of squared residuals in the sample (Hastie et al., 2021).

App 1. The R^2 values are approximately 0.7 which is significantly lower than the other two Apps.

App 2. The R^2 is higher when the smartphone is mounted on the windshield. Survey speed does not make a significant impact.

App 3. App3 demonstrated the best repeatability performance, as evidenced by significantly higher R^2 values. The performance is consistent across different practical settings.

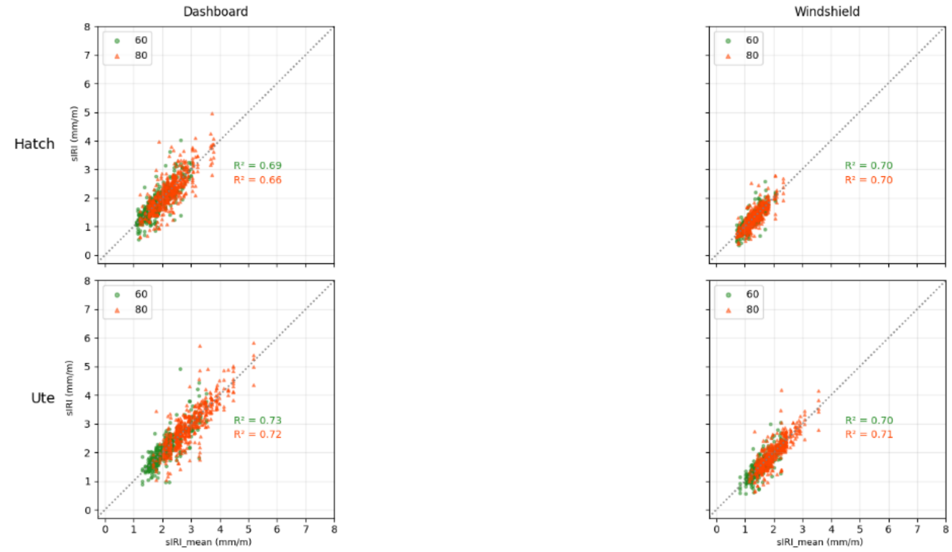

Figure 3 sIRI vs sIRImean plot (App1)

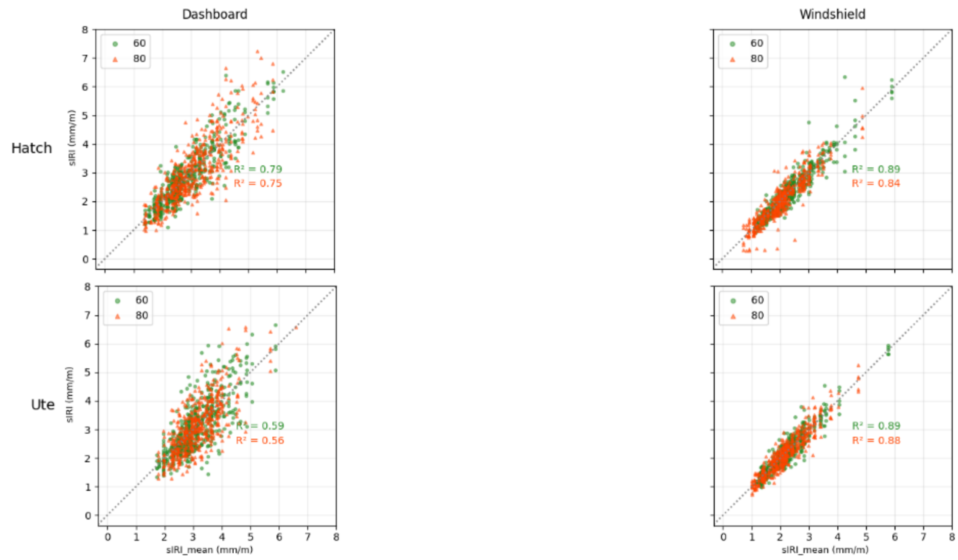

Figure 4 sIRI vs sIRImean plot (App2)

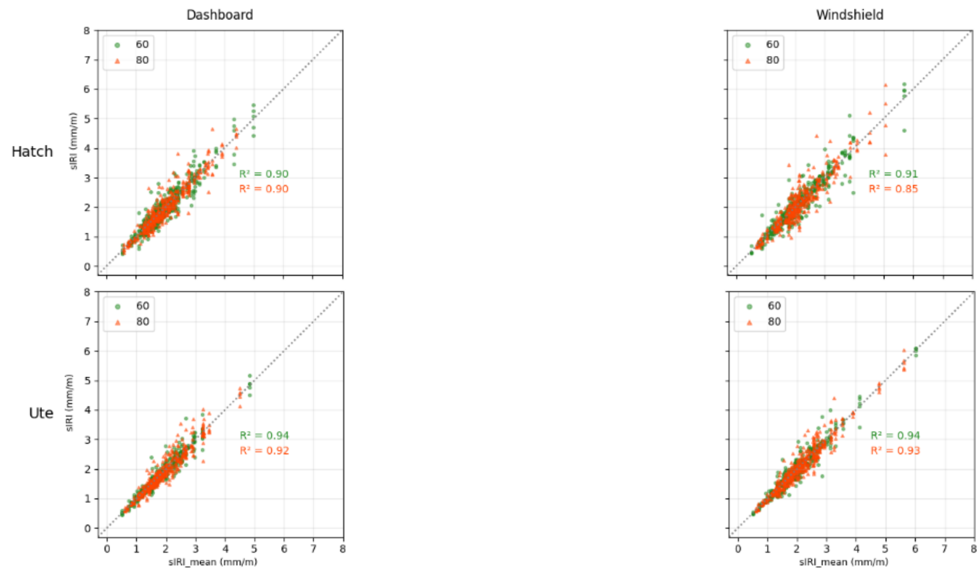

Figure 5 sIRI vs sIRImean plot (App3)

Accuracy test

Average of measurement error. The ε_{mean} is calculated as the percentage difference between sIRI and rIRI measurements, and it reflects the overall accuracy performance of the sRIE systems. As Table 2 shows, the ε_{mean} sites in a range of 0.085 to 0.536, and a lower ε_{mean} indicates a better accuracy. Among the three Apps, App2 produced a lower ε_{mean}.

Table 2 ε_{mean} of three Apps

Speed (km/h)	Vehicle	Mounting	App1	App2	App3
60	H	D	0.274	0.142	0.259
		W	0.536	0.085	0.208
	U	D	0.151	0.269	0.313
		W	0.392	0.125	0.262
80	H	D	0.095	0.253	0.251
		W	0.471	0.201	0.154
	U	D	0.175	0.305	0.280
		W	0.207	0.138	0.212
		Mean	0.288	0.189	0.242

Correlation with the $rIRI$. Linear regression models were created between sIRI and rIRI measurements, as shown in Figure 6 to Figure 8. Each figure displays plots of sIRI vs rIRI under varying practical settings, with speed distinguished by colours. The dotted 45° line within the plot represents a perfect correlation between the smartphone system and the profiler. It is noted that the closer the slope is to one and the interest is to zero, the better the smartphone system generates measurements comparable to the reference instrument.

App 1. With increasing speed, sIRI measurements tend to be higher, as indicated by the data points for higher speeds situated in the plot's upper region.

App 2. The performance under two speeds is comparable. The R^2 values of the windscreen are significantly higher than that of the dashboard.

App 3. Overall, the R^2 values are higher than that of the other two Apps and are consistent in four practical settings.

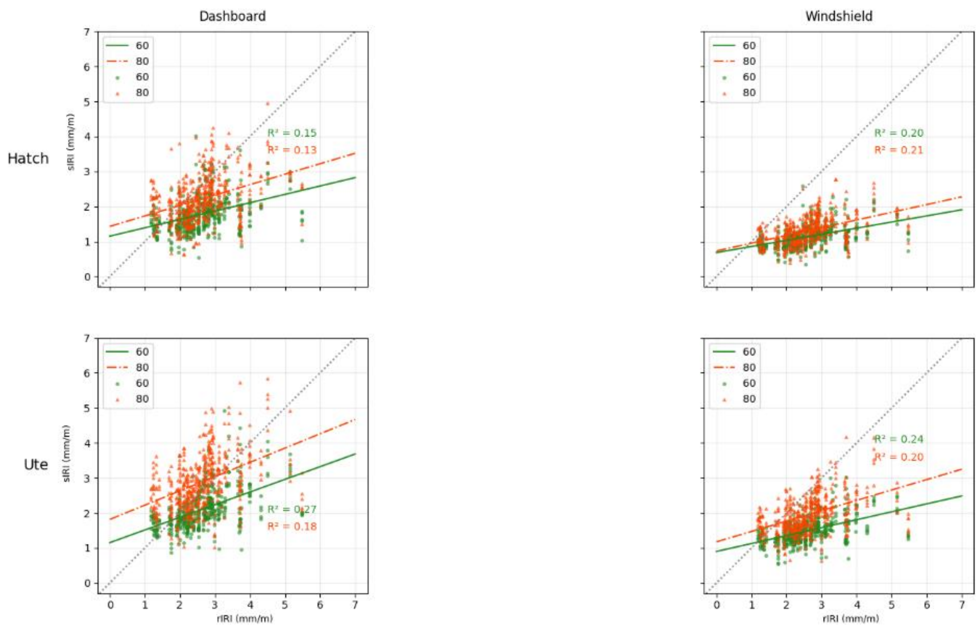

Figure 6 sIRI vs rIRI and the regression line (App1)

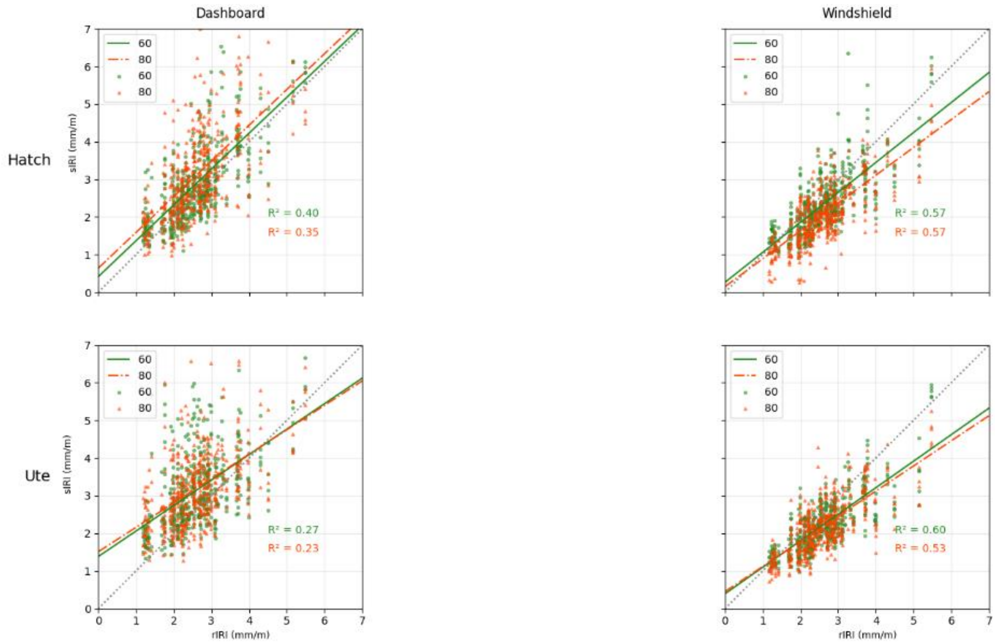

Figure 7 sIRI vs rIRI and the regression line (App2)

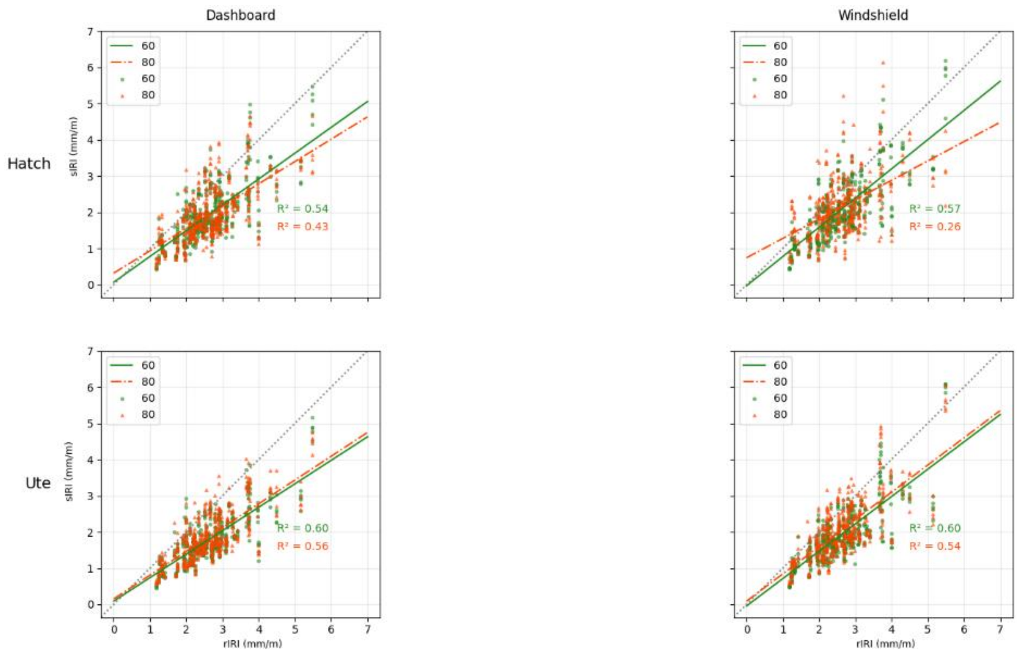

Figure 8 sIRI vs rIRI and the regression line (App3)

5. Discussion

This study's main goal was to assess the repeatability and accuracy of three commercial sRIE systems in a practical application setting. Using the framework proposed by (Yu et al., 2023), the evaluation was undertaken in realistic circumstances, including varying speeds, vehicle types, and mounting configurations.

The results of the repeatability tests showed that App3 consistently demonstrated the best performance in both the CoV and coefficient of determination (R^2) for the sIRI vs sIRI mean linear regression model. It was observed that the repeatability performance of App1 and App3 is consistent in the four vehicle-mounting configurations since their statistical measure values are close. However, for App2, the repeatability performance on windshield mounting is better than that on the dashboard, as the dashboard $\mathrm{CoV_{mean}}$ approximately doubles that of the windscreen; and the dashboard R^2 is 0.2 less the windscreen R^2. It was also noticed that the two survey speeds do not significantly affect the repeatability of all three Apps.

In terms of accuracy, the average measurement error ($\varepsilon_{\mathrm{mean}}$) results showed that App2 achieved the best performance. This suggests that App2 is more accurate in estimating pavement roughness, as its sIRI measurements were closer to the reference IRI (rIRI) values. However, the correlation with the rIRI results revealed that App3 displayed a higher R^2 values across different practical settings. Furthermore, it was observed that survey speed affects the R^2 of App1, where the R^2 of the regression line of 80km/h is significantly lower than that of the 60km/h. Meanwhile, both App2 and App3 generate close $\varepsilon_{\mathrm{mean}}$ and R^2 values in two survey speeds.

These findings provide valuable insights into the applicability in different scenarios. For instance, App3 should be selected where repeatability and consistency are of primary importance, while App2 or App3 are both preferred in situations where accuracy is the primary concern. While the pre-survey setup was completed in accordance with the guidelines provided by the sRIE system developers, there is a difference between the mounting setup suggested by the developer and what is adopted in the experiment. For instance, App3 requires the smartphone to be tagged on the vehicle windscreen using an armless mount. However, such a setup was not selected since the identical mounting configuration was adopted for three Apps. In summary, the results suggested that:

- The best repeatability performance of the sRIE system achieved a self-regression R^2 value of 0.90 across all practical settings.
- In accuracy test, App3 could achieve an R^2 value of 0.60, correlated with the reference instrument.
- The accuracy performance is to be improved, with a special focus on accommodating varying practical settings attributed to vehicle model and mounting configuration.

This study has limitations. First, it only adopted two options of survey speed, vehicle model and mounting configuration, and these are limited practical settings that do not fully represent the range of real-world scenarios that the sRIE systems may be tested under. Secondly, the mounting setup of App3 does not exactly follow the instructions from the App developer, which may induce noises to the raw signal and lead to discrepancies in the intended performance.

In terms of future research, the mechanistic properties of the mountings could be identified, and be integrated to alleviate the mounting variation's impact on the results. Similarly, vehicle uncertainties could be reduced by pre-calibration that considers vehicle geometric and suspension properties. The differences between various vehicles could be empirically identified from a real field experiment that covers a range of typical passenger vehicles.

6. Conclusion

This study evaluated three commercial sRIE systems using an evaluation framework. The tested systems surveyed a 10km road in an urban area and the measurement results were compared to that surveyed by a profiler. Results showed that App3 is preferred when both the repeatability and consistency are of primary importance, while App2 or App3 may be selected in situations where the measurement accuracy is critical. This study demonstrates the effectiveness of testing multiple sRIE systems using a standardised evaluation framework, and the results suggest that special focus to be made improving the sRIE system's robustness against different vehicle models and mounting configurations.

References

Botshekan, M., Asaadi, E., Roxon, J., Ulm, F.J., Tootkaboni, M., Louhghalam, A., 2021. Smartphone-enabled road condition monitoring: From accelerations to road roughness and excess energy dissipation. Proc. R. Soc. A Math. Phys. Eng. Sci. 477. https://doi.org/10.1098/rspa.2020.0701

Bridgelall, R., Hough, J., Tolliver, D., 2019. Characterising pavement roughness at non-uniform speeds using connected vehicles. Int. J. Pavement Eng. 20, 958–964. https://doi.org/10.1080/10298436.2017.1366768

Douangphachanh, V., Oneyama, H., 2014. Using smartphones to estimate road pavement condition, in: International Symposium for Next Generation Infrastructure. https://doi.org/10.14453/isngi2013.proc.16

Galagoda, D.Y., Lanka, S., 2019. Smartphone Applications for Pavement Roughness Computation of Sri Lankan Roadways. J. East. Asia Soc. Transp. Stud. 13, 2581–2597. https://doi.org/10.11175/easts.13.2581

Hastie, T., Tibshirani, R., James, G., Witten, D., 2021. An Introduction to Statistical Learning (2nd Edition), Springer Texts.

Hossain, M., Tutumluer, E., Nikita, Cole, G., 2019. Evaluation of android-based cell phone applications to measure international roughness index of rural roads, in: International Conference on Transportation and Development. pp. 309–318. https://doi.org/10.1061/9780784482575.034

Islam, S., Buttlar, W.G., Aldunate, R.G., Vavrik, W.R., 2014. Use of cellphone application to measure pavement roughness, in: Int&DI Congress 2014: Planes, Trains, and Automobiles. pp. 553–563. https://doi.org/10.1061/9780784413586.053

Jeong, J.H., Jo, H., Ditzler, G., 2020. Convolutional neural networks for pavement roughness assessment using calibration-free vehicle dynamics. Comput. Civ. Infrastruct. Eng. 35, 1209–1229. https://doi.org/10.1111/mice.12546

Sayers, M.W., 1995. On the calculation of International Roughness Index from longitudinal road profile. Transp. Res. Rec.

Shah, A., Zhong, J., Ly, J., 2017. Adopting smartphone technology to supplement road asset performance monitoring. 17th Int. Flex. pavement Conf.

Wix, R., 2016. Measuring road roughness with a smartphone – horses for courses? 27th ARRB Conf. people, places Oppor. 1, 13p.

Xue, K., Nagayama, T., Zhao, B., 2020. Road profile estimation and half-car model identification through the automated processing of smartphone data. Mech. Syst. Signal Process. 142. https://doi.org/10.1016/j.ymssp.2020.106722

Yu, Q., Fang, Y., Wix, R., 2023. Evaluation framework for smartphone-based road roughness index estimation systems. Int. J. Pavement Eng. 24. https://doi.org/10.1080/10298436.2023.2183402

Yu, Q., Fang, Y., Wix, R., 2022. Pavement roughness index estimation and anomaly detection using smartphones. Autom. Constr. 141, 104409. https://doi.org/10.1016/j.autcon.2022.104409

Zhao, B., Nagayama, T., Xue, K., 2019. Road profile estimation, and its numerical and experimental validation, by smartphone measurement of the dynamic responses of an ordinary vehicle. J. Sound Vib. 457, 92–117. https://doi.org/10.1016/j.jsv.2019.05.015