Toll Road Roughness Index Forecasting with Combination Grey Forecasting Model and Similarity Spatial Data

R Nurhadiansyah1, A Hadiana2
1Faculty of Postgraduate, Universitas Komputer Indonesia, Indonesia
2Department of Information Research Center of LIPI, The Indonesian Institute of Science

nurhadiansyah.ridwan@gmail.com, anahadiana.p2i.lipi@gmail.com

Abstract. The International Roughness Index (IRI) is used by toll road operators throughout the world as a main standard to quantify road surface roughness. IRI are performed to monitor the pavement conditions to evaluate pavements quality, and is a main supporting factor of safety and driving comfort. In Indonesian toll roads, IRI must be measured by annually with determinate value is $\leq 4 \text{ m/km}$ (unit of measurement). The purpose of this research is IRI forecasting on the Pondok Aren - Serpong toll road uses limited data history, testing results in 2013, 2015, 2016 and 2017 with Grey Forecasting Model (GM) method. Because of unavailability of testing results in 2014, this research tried to improve forecasting accuracy using the Similarity Spatial Data (SSD), is the IRI testing result on toll road that have similar characteristics with Pondok Aren - Serpong toll road. The final goal of this research is to determine how much influence the use of SSD in increasing the GM forecasting accuracy.

1. Introduction

Smother roads are required because they provide comfort and safety to road users [1]. One of the main factors supporting that is the level of roughness surface (International Roughness Index (IRI)). The pavement is measured by IRI developed by World Bank in the 1980's using Bump Integrator system (Figure 1). The IRI computation method converts the longitudinal and vertical profile data into a vehicle motion response using mathematical model [2]. Roughness of a road (or runway) is an important parameter which not only indicates the comfort level of ride over a pavement surface, but it is also related to vehicles vibration while in motion, the vehicle operating speed, wear and tear of the wheels with time, and the resulting operation cost of the vehicle [3]. By finding out the IRI value, it can be seen whether the road requires a maintenance only or quality improvement. Therefore, the result of IRI measurement can be used as a reference in determining the maintenance plan. The IRI summarize the roughness quality that impact vehicle response, and is most appropriate when a roughness measure is desired that relates to: overall vehicle operating cost, overall ride quality, dynamic wheel loads, and overall surface condition [4].

PT Margautama Nusantara (PT. MUN) is a company engaged in toll road service providers. Currently, PT. MUN operates many toll roads through company branches, one of them is Pondok Aren - Serpong toll road that connects Jakarta and South Tangerang. In its operation, a maintenance program is a mandatory activity to maintain the condition of road pavement. Still, in order to get an effective maintenance program so that the pavement is capable of serving traffic loads according to the age of the plan, a continuous assessment of the existing conditions of the pavement is needed. Because the measurement of IRI cannot be done at any time, a method is needed to predict the future value of IRI to
support the analysis and maintenance programs preparation. Based on the explanation above, this research is intended to predict the IRI value on Pondok Aren - Serpong toll road. Because of regulation reasons in Indonesia, the availability data of IRI measurement is very limited, are the results of 2013, 2015, 2016 and 2017 only. In this research, Grey forecasting model (GM) is selected as the right option of forecasting method. GM can give a solution to uncertainty model on limited data (there are at least four data) that’s usually called “partial known, partial unknown” [5]. However, the discontinuity of available data, it is feared that it will affect the accuracy of forecasting. Therefore, Similarity Spatial Data (SSD) is used in helping to improve the quality and continuity of data source.

SSD is a value or spatial data that has similar characteristics. The simplest way to find similarity between two categorical attributes is to assign a similarity of 1 if the values are identical and a similarity of 0 if the values are not identical. For two multivariate categorical data points, the similarity between them will be directly proportional to the number of attributes in which they match [6]. In this case, the SSD that used is the result of IRI measurements in 2018 on Pondok Aren - Ulujami toll road, that have similar characteristics and integrated directly with the Pondok Aren - Serpong toll road (see Figures 1 and 2).

![Figure 1. IRI measurement using bump integrator.](image1)

![Figure 2. The integration of Pondok Aren - Serpong toll road and Pondok Aren - Ulujami toll road](image2)

2. Method
This research was conducted to compare the accuracy of the forecasting results of IRI Pondok Aren - Serpong toll road used the GM method with the main data sources in 2013, 2015, 2016 and 2017, which will be compared with forecasting results uses the main data in 2015, 2016, 2017 and utilization of SSD in 2018. The data was taken from PT. Margautama Nusantara (PT. MUN) and PT. Jalan Tol Lingkar Luar Jakarta (PT. JLJ) as a Pondok Aren - Serpong toll road operator and Pondok Aren - Ulujami toll
road operator. Based on the time period, IRI testing on the Pondok Aren - Ulujami toll road for the same year period is carried out earlier than the Pondok Aren - Serpong toll road that was in June 2018. With the continuity of data source in forecasting, expected to improve forecasting accuracy (See Figure 3).

**Figure 3.** Research methodology.

2.1. Grey Forecasting Model

In 1982, Grey System Theory was first developed by Deng, which focused on the problem of uncertainty and the limitations of data or information in conducting analysis, prediction and decision making [7]. GM is a mathematical method used for forecasting with limited data and information. GM involves the following steps:

- **Step 1.** Assuming the raw data series to be [8]:
  \[ x^{(0)} = (x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n)) \]  

- **Step 2.** Creating a sequence of first-order accumulated generating operation (AGO).
  \[ x^{(1)} = \sum_{i=1}^{1} x^{(0)}(i), \sum_{i=1}^{2} x^{(0)}(i), \sum_{i=1}^{3} x^{(0)}(i), ..., \sum_{i=1}^{n} x^{(0)}(i) \]  

- **Step 3.** Calculate Inverse Accumulated Generating Operation (IAGO).
  \[ x^{(1)}(k + 1) = x^{(1)}(k) + x^{(0)}(k + 1), \quad \text{for } k \geq 1 \]  

Later on, IAGO will be used to achieve restored value, or forecasting value in the future which is denoted as \( \tilde{x}^{(0)}(k + 1) \). From the data line \( X^{(1)} \) and then \( Z^{(1)} \) is obtained.

- **Step 4.** \( z^{(1)}(k) \) is the raise of the average value of the two data \( x^{(1)}(k) \).
  \[ z^{(1)}(k) = \frac{1}{2} \left[ x^{(1)}(k - 1) + x^{(1)}(k) \right], \quad k = 2, 3, ..., n \]  

- **Step 5.** Calculate \( a \) and \( b \) value. Each parameter \( a \) and \( b \) is called the development coefficient and grey action quantity, where to get the values of parameter \( a \) and \( b \), the least squares estimation method is used. Through the result of \( X^{(0)}, X^{(1)} \) and \( Z^{(1)} \) with \( X^{(0)} \) non-negative
calculation, if \( \hat{p} = \frac{[\alpha]}{[b]} \) is parameter model, \( Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, \quad B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \), then the least squares of \( x^{(0)}(x) + az^{(1)}(k) = b \) completes \( \hat{p} = [B^TB]^{-1}B^TY \). If \( a \leq 0.5 \), then the value of \( a \) and \( b \) can be calculated using:

\[
\begin{align*}
    a &= \frac{\sum_{k=2}^{n} z^{(1)}(k) \sum_{k=2}^{n} x^{(0)}(k) - (n - 1) \sum_{k=2}^{n} z^{(1)}(k) x^{(0)}(k)}{(n - 1) \sum_{k=2}^{n} [z^{(1)}(k)]^2 - [\sum_{k=2}^{n} z^{(1)}(k)]^2} \\
    b &= \frac{\sum_{k=2}^{n} [z^{(1)}(k)]^2 \sum_{k=2}^{n} x^{(0)}(k) - \sum_{k=2}^{n} z^{(1)}(k) x^{(0)}(k)}{(n - 1) \sum_{k=2}^{n} [z^{(1)}(k)]^2}
\end{align*}
\]  

- **Step 6.** Calculate \( f^{(1)}(k) \) value.

\[
f^{(1)}(k) = \left( x^{(0)}(1) - \frac{b}{a} \right) e^{-a(k-1)} + \frac{b}{a}
\]  

- **Step 7.** Calculation of forecasting values using equations below:

\[
f^{(0)}(k) = f^{(1)}(k) - f^{(1)}(k - 1)
\]

### 2.2. Similarity Spatial Data

Spatial data is data that has geographic reference with two important parts that make it different from other data, namely location information and attribute information. Spatial data is information about a physical object that can be represented by numerical values in a geographic coordinate system. The characteristics of spatial data (dependence and heterogeneity) often void the attractive properties of standard statistical techniques [9].

Similarity data is data that has attribute similarity, namely property similarity or characteristic of an object. The simplest way to find similarity between two categorical attributes is to assign a similarity of 1 if the values are identical and a similarity of 0 if the values are not identical. For two multivariate categorical data points, the similarity between them will be directly proportional to the number of attributes in which they match [10]. From the explanation above, the conclusion is that Similarity Spatial Data (SSD) is a spatial data that has resemblance or similarity to property attributes or characteristic of its object.

### 2.3. Forecasting Accuracy

To demonstrate the efficiency of the proposed forecasting model, this study adopts the residual error test method to compare the actual and forecast values [8]. After achieving the forecast results using GM, to see the accuracy of the forecasting required Residual es series, Residual error and Average Residual Error [5]. For example, \( x^{(0)}(k) \) is the real data and \( \hat{x}^{(0)}(k) \) is the forecasting uses GM, Residual series defined as:

\[
e^{(0)}(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), \quad \text{for } k \geq 2
\]

\[
\text{Residual Error} = \frac{[e^{(0)}(k)]}{[x^{(0)}(k)]}
\]

\[
\text{Average Residual Error} = \frac{1}{n} \sum_{k=1}^{n} \left| \frac{e^{(0)}(k)}{x^{(0)}(k)} \right|
\]

\[
\text{Accuracy} = 100\% - \frac{1}{n} \sum_{k=1}^{n} \left| \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \right| \times 100\%
\]

To find out the level of forecasting accuracy, it can be seen in the following Table 1 [5]:
Table 1. Criteria of GM forecasting accuracy

| Accuracy Level | Criteria | Residual Error | Accuracy |
|---------------|----------|---------------|---------|
| I (Very Good) |          | 1%            | 90%     |
| II (Good)     |          | 5%            | 80%     |
| III (Fair)    |          | 10%           | 70%     |
| IV (Poor)     |          | 20%           | 60%     |

3. Results and Discussion

3.1. Main Data

The main data used in this research is the actual result of IRI measurements in the Pondok Aren - Serpong toll road period in 2013, 2015, 2016, and 2017. The data is presented in the overall average (Figure 4) and lane average (Figure 5). Based on the overall average value, IRI on Pondok Aren - Serpong toll road in good condition (≤ 4 m/km), but an increase in IRI value starts from 2015. That shows there is a deterioration in the level of road services. While based on the lane average, the worst IRI value is Lane 1. However, the most significant decrease in the level of road service is Lane 2 (B) and Lane 3 (B). A and B are terms used in directions name, where (A) is a normal direction and (B) is an opposite direction. In this case, (A) is a Serpong direction, and (B) is a Pondok Aren direction.

Figure 4. The result of IRI measurements average on the Pondok Aren - Serpong toll road
Similarity Spatial Data (SSD)

SSD utilization in this research are IRI measurements in 2018 on the Pondok Aren - Ulujami toll road, is intended to obtain the minimum data required by GM method. The combination of main data and SSD can be seen on Figures 6 and 7. Figure 6 shows the overall average and Figure 7 shows the lane average. With the continuity of data (after the combining of main data and SSD), it is expected to increase the results of forecasting accuracy.

Figure 5. The result of IRI measurements lane average on the Pondok Aren - Serpong toll road

Figure 6. Average value from main data and SSD
3.3. **GM Forecasting Result**

The following are the results of forecasting the IRI value of the Pondok Aren - Serpong toll road in 2018 which is carried out using the main data in the form of IRI test results in 2013, 2015, 2016 and 2017 as a whole and the lane average (see Tables 2 and 3).

### Table 2. The average results of GM forecasting

| Types of Data | Period | \( x^{(0)} \) | IRI Value (m/km) |
|---------------|--------|---------------|------------------|
| Main Data     | 2013   | \( x^{(0)}(1) \) | 2.830            |
| Main Data     | 2015   | \( x^{(0)}(2) \) | 2.195            |
| Main Data     | 2016   | \( x^{(0)}(3) \) | 2.354            |
| Main Data     | 2017   | \( x^{(0)}(4) \) | 2.829            |
| **Forecasting Result** | 2018 | \( x^{(0)}(5) \) | **2.480**        |

### Table 3. The lane average results of GM forecasting

| Lane (Direction) | Period (2018 Forecasting) |
|------------------|---------------------------|
|                  | 2013 | 2015 | 2016 | 2017 |               |
| Lane 1 (A)       | 2.918 | 2.353 | 2.578 | 2.884 | **2.654** |
| Lane 2 (A)       | 2.857 | 2.164 | 2.383 | 2.690 | **2.448** |
| Lane 3 (A)       | 2.812 | 2.151 | 2.357 | 2.712 | **2.435** |
| Lane 1 (B)       | 2.933 | 2.427 | 2.592 | 2.887 | **2.698** |
| Lane 2 (B)       | 2.760 | 2.061 | 2.132 | 2.950 | **2.353** |
| Lane 3 (B)       | 2.699 | 2.011 | 2.081 | 2.852 | **2.288** |
3.4. GM and SSD Forecasting Result
The following are the results of forecasting the IRI value of the Pondok Aren - Serpong toll road in 2018 which is carried out using the main data in the form of IRI test results in 2013, 2015, 2016 and 2017 and SSD utilization in 2018 on an overall average and lane average. This calculation intended to compare forecasting accuracy can be seen in Tables 4 and 5 below.

### Table 4. Average results of GM forecasting and SSD utilization

| Types of Data         | Period  | \(x^{(0)}\) | IRI Value (m/km) |
|-----------------------|---------|-------------|-----------------|
| Main Data             | 2015    | \(x^{(0)}(1)\) | 2.195           |
| Main Data             | 2016    | \(x^{(0)}(2)\) | 2.354           |
| Main Data             | 2017    | \(x^{(0)}(3)\) | 2.829           |
| Similarity Spatial Data| 2018 | \(x^{(0)}(4)\) | 2.436           |
| **Forecasting**       | 2018    | \(x^{(0)}(5)\) | **2.638**       |

### Table 5. Lane average results of GM forecasting and SSD utilization

| Lane (Direction) | Period | 2015 | 2016 | 2017 | 2018 (SSD) | 2018 Forecasting (m/km) |
|------------------|--------|------|------|------|------------|-------------------------|
| Lane 1 (A)       | 2015   | 2,353| 2,578| 2,884| 2,683      | **2.827**               |
| Lane 2 (A)       | 2015   | 2,164| 2,383| 2,690| 2,599      | **2.632**               |
| Lane 3 (A)       | 2015   | 2,151| 2,357| 2,712| 2,266      | **2.565**               |
| Lane 1 (B)       | 2015   | 2,427| 2,592| 2,887| 2,578      | **2.817**               |
| Lane 2 (B)       | 2015   | 2,061| 2,132| 2,950| 2,324      | **2.539**               |
| Lane 3 (B)       | 2015   | 2,011| 2,081| 2,852| 2,168      | **2.448**               |

3.5. Forecasting Accuracy
Testing the accuracy of forecasting results is done by calculating the residual value, residual error and average residual error. Comparison of GM forecasting results using the main data and utilization of SSD produces can be seen in Table 6 and Figure 8 below.

### Table 6. Forecasting results comparison of GM forecasting and SSD utilization

| Period | \(k\) | Types of Data | IRI Value | GM Forecasting | GM Forecasting and SSD Utilization |
|--------|-------|---------------|-----------|----------------|-----------------------------------|
|        |       |               |           | Forecast Result | Residual Value | Residual Error | Forecast Result | Residual Value | Residual Error |
| 2013   | 1     | Main Data     | 2.830     | -              | -              | -              | -              | -              | -              |
| 2015   | 2     | Main Data     | 2.195     | 2.537          | -0.342         | 15.56%         | -              | -              | -              |
| 2016   | 3     | Main Data     | 2.354     | 2.518          | -0.164         | 6.96%          | 2.647          | -0.294         | 12.48%         |
| 2017   | 4     | Main Data     | 2.829     | 2.499          | 0.331          | 11.69%         | 2.644          | 0.185          | 6.53%          |
| 2018   | 5     | SSD           | 2.436     | -              | -              | -              | 2.641          | -0.205         | 8.42%          |
| 2018   | 6     | Main Data     | 2.869     | 2.480          | 0.389          | 13.57%         | 2.638          | 0.231          | 8.04%          |
| Average Residual Error |       |               |           |                |                |                |                |                | 11.94%        |
| **Forecasting Accuracy** | |               |           |                |                |                |                |                | **88.06%**   |

**Forecasting Accuracy** 91.13%
The forecasting results of the IRI value of Pondok Aren - Serpong toll road using main data in 2013, 2015, 2016, and 2017 achieved an accuracy rate 88.06%. Meanwhile, the forecasting results using main data in 2015, 2016, 2017, and SSD utilization in 2018 achieved accuracy rate 91.13%.

4. Conclusion
Based on the discussion about impact of the use of Similarity Spatial Data (SSD) in the roughness index (IRI) forecasting of Pondok Aren - Serpong toll road used the Grey Forecasting Model method with data limitations and discontinuities, it can be concluded: The continuity of actual data with the minimum amount required in forecasting using the GM method can reduce the accuracy of forecasting. Sustainable use of SSD can improve the quality of data sources and the accuracy of forecasting. Based on the results of this research, in the case of spatial data forecasting in the form of an inequality index on Pondok Aren - Serpong toll road, SSD utilization can increase forecast accuracy by 3.07%. Eventually, this research was able to answer that the utilization of Similarity Data in forecasting using the Grey Forecasting Model can increase the accuracy of forecasting. The better the quantity and quality of the data that used, then the better the accuracy of forecasting.

References
[1] Al-Rousan, T., Asi, I., & Baker, A. A. (2010, October). Roughness evaluation of Jordan highway network. In 24th ARRB Conference, Melbourne, Australia, 41.
[2] Abiola, O. S., Kupolati, W. K., & Odunfa, S. O. (2014). Predicting visual pavement score from international roughness index. In International Conference on Biological Civil and Environmental Engineering (BCEE-2014).
[3] Yero S A, Hainin M R and Yacoob H 2012 Determination of Surface Roughness Index of Various Bituminous Pavements. IJRRAS 13, 98
[4] Sayers, M. W. (1998). The little book of profiling: basic information about measuring and interpreting road profiles.
[5] Liu, S., & Forrest, J. Y. L. (2010). Grey systems: theory and applications. Springer Science & Business Media.
[6] Chandola, V., Borah, S., & Kumar, V. (2007). Similarity Measures for Categorical Data–A Comparative Study. Technical Report 07-022, Department of Computer Science & Engineering, University of Minnesota.
[7] Deng, J. L. (1982). Control problems of grey systems. Sys. & Contr. Lett., 1(5), 288-294.
[8] Huang, Y. L., & Lee, Y. H. (2011). Accurately forecasting model for the stochastic volatility data in tourism demand. *Modern economy*, 2(05), 823.

[9] Anselin, L. (1989). What is Special About Spatial Data? Alternative Perspectives on Spatial Data Analysis (89-4).

[10] Chandola, V., Borah, S., & Kumar, V. (2007). *Similarity Measures for Categorical Data--A Comparative Study*. Technical Report 07-022, Department of Computer Science & Engineering, University of Minnesota.