RESEARCH ARTICLE

Exploration of the application of Grey-Markov models in the causality analysis of traffic accidents in roundabouts

Peijing Li[1]*, Jian Li[2]

1 Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan, United States of America, 2 College of Fashion and Design, Donghua University, Shanghai, China

* peijli@umich.edu

Abstract

We propose a multivariate Grey-Markov model to quantify traffic accident risk from different causality factors in roundabouts that is uniquely suited for the scarce and stochastic traffic crash data from roundabouts. A data sample of traffic crashes occurring in roundabouts in the U.S. State of Michigan from 2016 to 2021 was collected to investigate the capabilities of this modeling methodology. The multivariate grey model (MGM(1,4)) was constructed using grey relational analysis to determine the best dimensions for model optimization. Then, the Markov chain is introduced to address the unfitness of stochastic, fluctuating data in the MGM(1,4) model. Finally, our proposed hybrid MGM(1,4)-Markov model is compared with other models and validated. This study highlights the superior predictive performance of our MGM(1,4)-Markov model in fore-casting roundabout traffic accidents under data-limited conditions, achieving a 3.02% accuracy rate, in contrast to the traditional GM(1,1) model at 8.30% and the MGM(1,4) model at 4.47%. Moreover, incorporating human, vehicle, and environmental risk factors into a multivariate crash system yields more accurate predictions than merely aggregating crash counts.

Introduction

Modern roundabouts require traffic from all sides to yield circulating traffic, with sufficient horizontal curvature to slow entry and circulation speeds, and they are widely touted as an effective safety intervention to replace traditional intersections to lower accident rates and severities. This is in part due to their reduction in the number and types of conflicts between vehicles and/or pedestrians moving in different directions through the intersection [1, 2]. In the meantime, forecasting roundabout traffic accidents and identifying factors determining the severity of injuries to road users is an essential component of accident prevention decision-making and avoiding casualties or damage. However, the safety performance at roundabouts is affected by many uncertainty factors, such as the interaction of collision factors—humans, vehicles, and road environment- which negatively exhibits uncertain variability and ultimately complicates the performance prediction of roundabout traffic [3, 4]. Hence efficient methods are needed to capture this uncertainty, predict roundabout traffic crashes accurately

beyond empirical observations, and analyze their causality quantitatively, ideally with the help of emerging data collection and processing techniques, to inform future transportation policy and planning.

The most significant issue with modeling and predicting traffic crashes in roundabouts is the scarcity and stochasticity in the data for historical roundabout traffic accidents. This characteristic in such datasets, combined with the aforementioned fact that a variety of human, vehicle, and environmental factors all affect the safety performance of roundabouts [5, 6], matches the features of the grey prediction model well, which does not require much information and has simple calculations and high precision [7, 8]. Developed by Deng in 1982 [9], the grey theory deals with limited sample sizes (typically four or more samples) and poor information problems by employing ordinary differential equations to characterize the factors that may affect forecast accuracy. Since it provides a way to describe unknown systems under limited data conditions, it is widely used in engineering, economics, and science [10, 11]. Nonetheless, the grey model may yield less desirable performance when dealing with highly fluctuating data sequences during long-term prediction [12, 13].

In order to deal with datasets with more fluctuation like that of traffic crashes in roundabouts, we propose the integration of grey model techniques with Markov chains. Such a technique has already been proven to perform well with sparse and stochastic datasets with regards to determining causality and predicting future values. For example, Jin et al. [14] and Zou et al. [15] developed a one-factor Grey-Markov model to estimate the number of fatalities and economic losses associated with traffic accidents. These models, however, are all first-order grey models that takes into account only one independent variable. To date, few studies have given methods to determine the optimal dimension variables in multivariate Grey-Markov model prediction with limited data.

Driven by this question, this paper aimed to demonstrate the effectiveness of an optimized multivariable Grey-Markov model (MGM(1,n)-Markov) in fitting and predicting a stochastic and sparse roundabout crash dataset by comparing it with the classical GM(1,1) (grey model with one independent variable and one dependent variable) and optimized MGM(1,n) (grey model with multiple independent variable) models previously published in the literature [16]. The hybrid model consists of four steps. The first step involves performing a grey relational analysis using a dataset of traffic crashes at roundabouts from the Michigan Traffic Crash Facts dataset to measure the correlation between variables and the occurrence of traffic-related injuries. The next step is constructing a multivariate grey model for predictive theory by identifying the main elements or causal features describing collapse using grey relation analysis. As a third step, the residual error of the grey model prediction is modified using Markov theory, and as a final step, future trends are predicted.

The novelty aspects of this paper are summarized as follows:

1. Few studies have previously been published on the quantitative analysis of causality factors in roundabout traffic crashes.

2. We apply grey relational analysis to identify the weights and co-linearity between factors associated with accident occurrence in roundabouts, and combined with the Markov optimization model corrects the non-stationary data sequences.

3. A multivariable Grey-Markov model with sparse and stochastic data samples is proposed, which has not been applied to the prediction of roundabout traffic accidents.

To the best of our knowledge, there are only a few attempts to predict traffic accidents using multidimensional time series of influential factors. Most previous studies, particularly those focusing on roundabout traffic accidents, only examined a single variable, the number of

accidents, and did not consider the traffic system's multi-factor effects. Therefore, we aim to propose a multi-scale time-series roundabout crash prediction model that incorporates several human, vehicle, and road risk factors instead of relying solely on crash totals.

The remaining sections of the paper are organized as follows. We provide an overview of the state of the art of research in roundabout safety in the "Related works" section. The "Materials and methods" section presents the details of our methodology, which incorporates an optimized, multivariable MGM(1,n) model and Markov chains to update and optimize the grey model's predictive results. The "Results" section presents the results of utilizing this methodology to model roundabout traffic crashes in Michigan. Consequently, these results, methodological considerations, mechanisms, practical application recommendations, and limitations are discussed in the "Discussion" section. Finally, the "Conclusions and perspectives" section provides a conclusion to this paper.

Related works

In this section, we provide an extensive synthesis of the existing literature on the prevailing trends in roundabout development, emphasizing the application of predictive models for traffic crash forecasting.

Modern roundabouts, originating in the United Kingdom, have been designed to enhance traffic safety and flow [17]. While early designs presented constraints, introducing innovative traffic rules has improved capacity and safety. Research findings highlight the significant advantages of roundabouts for intermediate traffic demands, while signalized intersections exhibit superior performance in high-traffic scenarios [17, 18]. Roundabouts encourage low speeds, promoting safety, but possess unique crash profiles. Their implementation in the United States has led to decreased injury collisions and improved efficiency. However, the direct applicability of international research findings to the US context may be constrained due to various factors [18].

Current studies on traffic crash prediction focus on highways [6, 19–22], signalized intersections [23–25], and other roads [26, 27], whereas there has been little research on roundabouts in recent decades. Many approaches have been developed to solve traffic accident prediction problems, roughly categorized into three types: parametric models [28, 29] (e.g., ARIMA, grey prediction, spectral analysis), non-parametric models [6, 20, 21, 30–34] (e.g., non-parametric regression model, SVR, VAR, KNN, deep learning, parallel fp-growth mining, and types of neural network), and hybrid models [16, 22, 35] (e.g., Bayesian-neural network, fuzzy rule-based method, grey-ARIMA, and chaos-wavelet analysis-SVM). It is necessary to note that the methods above for predicting traffic collisions have higher requirements on the data sample size for training and verification and cannot capture data samples with considerable randomness [12]. However, the fact that roundabouts are an emerging mode of transportation in the United States and are not as widely deployed as in certain other countries determines that any available police-reported crash data would be highly sparse and stochastic. There is a gap in the current literature on a specific methodology to draw meaningful conclusions from such data samples.

Despite some progress in the above prediction models, certain research gaps in the existing literature still need to be explored in further detail. First, previous studies have overwhelmingly focused on conventional intersections, and relatively few studies have examined new roundabouts that may relieve traffic congestion to a great extent. Second, traditional prediction models commonly require normal data distribution and restrictions on sample size [8, 10]. In practice, roundabout traffic accidents consist of time series data that are highly nonlinear and scattered, making it difficult for historical data to be modeled and limiting prediction accuracy.

Therefore, the key to accurate forecasting is choosing an appropriate method, especially in cases of incomplete data. Finally, relatively few studies on roundabout crashes have considered aggregated, empirical crash counts without adequately addressing the contributing explanatory variables [14, 36]. Numerous factors affect or explain such collisions, including driver behavior, weather conditions, traffic flow, and road structure [37–39]. Inevitably, most of these data-driven approaches are summary-statistics-based joint analyses, which are generally challenging to specify the insights of the prediction results. Such a literature gap calls for further research incorporating driver, vehicle, and environmental factors to analyze and predict roundabout traffic collisions with limited data.

Materials and methods

Data source

We used the data source of MTCF (Michigan Traffic Crash Facts) [40] yearly incidence data for roundabout traffic crashes from 2016 to 2021 to compile the roundabout traffic crash data in this paper. Note that this third-party data source is compiled by the Michigan State Police Office of Highway Safety Planning (OHSP) and the University of Michigan Transportation Research Institute (UMTRI) from all police reports of traffic crashes in the state of Michigan; and the authors of this article do not have and cannot provide direct access to the raw dataset. However, the dataset, as well as detailed breakdowns of crash figures by year and other explanatory variables, are available through MTCF's public query tool at https://www. michigantrafficcrashfacts.org/querytool. Time frame-wise, we focus on traffic crashes from 2016 to 2021 since earlier records did not identify whether a collision occurred within a roundabout. Individual data items within this dataset were digitized from police reports of accidents.

According to Liu et al. [41], index screening should follow three principles, namely measurability, representativeness, and comparability. In accordance with this principle, we then selected eight explanatory variables (i.e., snow-covered, head-on crash on left turn, sideswipe, distraction, injury, median, buses/trucks, and rainy) to reflect the severity and probability of traffic accident risk (see Table 1). These eight variables are the only ones that provide continuous samples of non-trivial and non-zero data over the time period, given the sparsity and stochasticity of the data sample to begin with. Using this data set, we investigate the impact of each contributing element on the annual number of crashes at roundabouts compared to traditional intersections.

Note that in the Table 1 above, X_1 refers to crashes occurring on snow-covered roads, X_2 refers to head-on crashes against another vehicle performing a left turn, X_3 refers to a sideswipe crash, X_4 refers to crashes with distracted drivers, X_5 refers to crashes resulting in injuries, X_6 refers to crashes in the median of roads, X_7 refers to crashes involving buses and/or trucks, and X_8 refers to crashes in rainy weather.

Table 1. Explanatory variables of traffic crash data in Michigan roundabouts.

Year	Total crashes	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
2016	489	20	1	146	23	48	5	25	52
2017	1510	73	3	535	48	125	19	97	154
2018	1501	78	3	523	41	116	16	86	112
2019	1864	85	5	642	57	146	19	162	176
2020	1300	61	2	416	49	103	19	96	77
2021	1730	79	1	618	56	132	23	131	116

Methodology overview

The purpose of this section is to propose an optimized multivariate Grey-Markov model for predicting short-term roundabout traffic crashes. Specifically, it combines grey relational analysis, the Multivariate Grey-Markov model, and the model evaluation framework for our predictor of traffic accidents (see Fig 1).

Grey relational analysis

The basic idea of grey relational analysis is to judge the closeness of the relationship based on the similarity of the geometric shapes of the sequence curves. The higher the similarity, the greater the correlation. Traditional statistical methods have many drawbacks, such as requiring large amounts of data, typical sample distributions, and difficulty meeting practical needs. Grey relational analysis overcomes these shortcomings and only requires four data points without the need to meet typical distribution rules.

In many instances, roundabout accidents are caused by multiple potentially inter-connected factors. At the same time, numerous variables affect the occurrence and characteristics of accidents. We are interested in determining how each factor variable affects the accident variable to develop a reliable model for accident prediction. Typically, an impact factor for a potential accident is defined as a temporally ordered sequence of variables that characterize trends of the respective factor.

A variable sequence represents a time series of multiple variables that must be simulated and predicted to predict traffic accidents. We use grey relational analysis [42] to determine whether there is a compact connection between sequence curves based on their geometric similarity. Roundabout accident variables and their impact factors are com-pared to determine the coefficients.

Following grey relational analysis theory, let us suppose that a roundabout traffic system contains m time-series variables. The first of these variables describe the number of crashes, and the rest describe the various causality characteristics. Furthermore, a total of n measurements of those variables are taken at different discrete points in time. These time series can be represented as follows, where i refers to a certain time-series variable and k refers to one of the
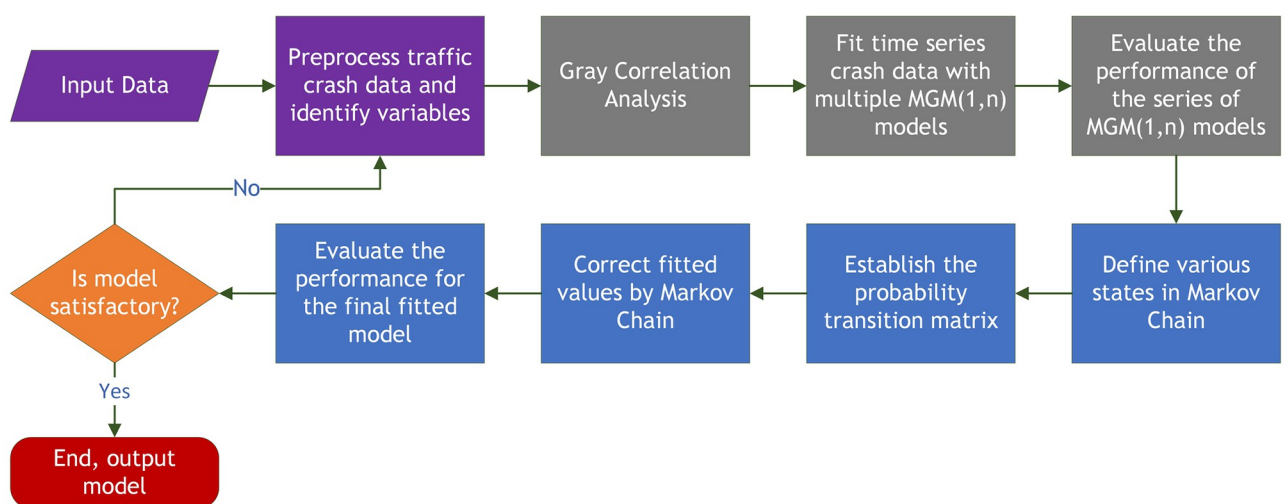

Fig 1. Flow chart of the optimized multivariate Grey-Markov model.

measurements of the variable over time.

$$x_i = \{x_i(k) | k = 1, 2, \ldots, n\} = \{x_i(1), x_i(2), \ldots, x_i(n)\}, i = 1, 2, \ldots, m \tag{1}$$

We can further define that x_1 refers to the dependent time-series variable of traffic crash numbers, and that x_j, $j = 2, 3, \ldots, m$ refers to the various independent causality characteristics that may be correlated to the outcome of traffic crash statistics [43]. Then the grey correlation between the number of crashes observed (x_1) and an independent variable (x_i) can be expressed as r_i in the following expression.

$$r_i = \frac{1}{n} \sum_{k=1}^{n} \zeta_i(k), \text{ where } \zeta_i(k) = \frac{\min_s \min_t |x_1(t) - x_s(t)| + \rho \max_s \max_t |x_1(t) - x_s(t)|}{x_1(k) - x_i(k) + \rho \max_s \max_t |x_1(t) - x_s(t)|} \tag{2}$$

Note that $\rho \in [0, 1]$ refers to a "grey correlation coefficient," $\zeta_i(k)$ provides the grey correlation coefficient at time point k between x_0 and x_i, $\min_s \min_t |x_1(t) - x_s(t)|$ and $\max_s \max_t |x_1(t) - x_s(t)|$ denotes the two-level minimum and maximum discrepancy, between $x_1(t)$ and any of $x_2(t), \ldots, x_m(t)$, respectively. We can then define r_i to be the grey correlation degree coefficient. The higher the value of r_i, the stronger the correlation is between x_1 and x_i.

Multi-variable Grey modeling

Variable and sequence definition.   Let us denote the dependent variable sequence from the previous section as $x_1^{(0)}$ and the independent variable sequences from the previous section as $x_j^{(0)}$ with $j = 2, 3, \ldots, m$. Then given $i = 1, 2, \ldots, m$, we can apply the first-order accumulated generation operation (1-AGO) to $x_i^{(0)} = \{x_i^{(0)}(1), x_i^{(0)}(2), \ldots x_i^{(n)}\}$ to obtain an "accumulated" times series $x_i^{(1)}$ as follows.

$$x_i^{(1)}(k) = \sum_{p=1}^{k} x_i^{(0)}(p), k = 1, 2, \ldots, n \tag{3}$$

Meanwhile, we can then define the adjoining main generated sequence of $x_i^{(1)}$ as $x_i^{(2)}$ [44].

$$x_i^{(2)}(k) = \frac{1}{2}\left(x_i^{(1)}(k) + x_i^{(1)}(k-1)\right), k = 2, 3, \ldots, n \tag{4}$$

We can then define the vector of all time-series variables, their 1-AGO forms, and their adjoining main sequences at the discrete time point $k \in [1, n]$ as follows.

$$X^{(0)}(k) := \left[x_1^{(0)}(k), x_2^{(0)}(k), \ldots, x_m^{(0)}(k)\right]^T \tag{5}$$

$$X^{(1)}(k) := \left[x_1^{(1)}(k), x_2^{(1)}(k), \ldots, x_m^{(1)}(k)\right]^T \tag{6}$$

$$X^{(2)}(k) := \left[x_1^{(2)}(k), x_2^{(2)}(k), \ldots, x_m^{(2)}(k)\right]^T \tag{7}$$

In order to fit a Multivariable Grey Model to, and to predict the value of the time series variables of $X^{(0)}$ and $X^{(1)}$, we can also introduce the following two vectors of continuous time

functions that correspond to the model fitted values with regards to $X^{(0)}$ and $X^{(1)}$, respectively.

$$\tilde{X}^{(0)}(t) := \left[\tilde{x}_1^{(0)}(t), \tilde{x}_2^{(0)}(t), \ldots, \tilde{x}_m^{(0)}(t)\right]^T \tag{8}$$

$$\tilde{X}^{(1)}(t) := \left[\tilde{x}_1^{(1)}(k), \tilde{x}_2^{(1)}(t), \ldots, \tilde{x}_m^{(1)}(t)\right]^T \tag{9}$$

Grey model equation system definition. The Multivariable Grey model can be defined by the following system of first-order differential equations.

$$\frac{\mathrm{d}\tilde{X}^{(1)}}{\mathrm{d}t} = A\tilde{X}^{(1)} + B \tag{10}$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix}, B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Note that the initial condition is defined as $\tilde{X}^{(1)}(1) := X^{(1)}(1)$.

We can acquire the values of A and B as follows. Let us first define α_i as the i-th row vector from the matrix $[A|B]$, or

$$\alpha_i = [a_{i1}, a_{i2}, \ldots, a_{im}, b_i]^T, i = 1, 2, \ldots, m \tag{11}$$

We can then compute the values for each α_i through the following expression.

$$\alpha_i = (P^T P)^{-1} P^T Q_i \tag{12}$$

where

$$P = \begin{bmatrix} x_1^{(2)}(2) & x_2^{(2)}(2) & \ldots & x_m^{(2)}(2) \\ x_1^{(2)}(3) & x_2^{(2)}(3) & \ldots & x_m^{(2)}(3) \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(2)}(n) & x_2^{(2)}(n) & \ldots & x_m^{(2)}(n) \end{bmatrix}, Q_i = \begin{bmatrix} x_i^{(0)}(2) \\ x_i^{(0)}(3) \\ \vdots \\ x_i^{(0)}(n) \end{bmatrix}$$

Grey model solution. The explicit, analytical solution to Eq 10 can thus be expressed as follows.

$$\tilde{X}^{(1)}(t) = e^{At}X^{(1)}(1) + A^{-1}(e^{At} - I)B \tag{13}$$

where I is the $m \times m$ identity matrix, and the matrix exponential e^{At} can be computed using Taylor series decomposition as follows.

$$e^{At} = I + \sum_{j=1}^{\infty} \frac{A^j t^j}{k!} \tag{14}$$

Finally we can recover the prediction values for the time series variables using their predicted

1-AGO forms. For instance, the following expression can be used if one wishes to predict the value of $\tilde{x}^{(0)}(m+1)$ given the predicted time series data for the first m time points.

$$\tilde{X}^{(1)}(m+1) = e^{A(m+1)}X^{(1)}(1) + A^{-1}(e^{A(m+1)} - I)B \tag{15}$$

$$\tilde{x}^{(0)}(m+1) = \tilde{x}^{(1)}(m+1) - \tilde{x}^{(1)}(m) \tag{16}$$

Markov chain modeling

In this section only, let x and $\tilde{x}$ denote the raw and predicted sequences of n variables obtained from the aforementioned Multivariable Grey Models, respectively.

State determination.   We can first compute the following residual sequences between the raw and predicted data values.

$$\delta = \{\delta(k)|k = 2, 3, \ldots, n\}, \text{ where } \delta(k) = \frac{x(k) - \tilde{x}(k)}{\tilde{x}(k)} \tag{17}$$

Then we can define the following Markov state transitions based on the above residual values. Note that each state S_j has a lower bound residual value of L_j and upper bound of U_j, defined as follows.

$$L_j = \min[\delta(i)] + \frac{j-1}{n}(\max[\delta(i)] - \min[\delta(i)]) \tag{18}$$

$$U_j = \min[\delta(i)] + \frac{j}{n}(\max[\delta(i)] - \min[\delta(i)]) \tag{19}$$

State transition matrix.   In order to correct for the inaccuracies of the output values $\tilde{x}$ of the Multivariable Grey Model, we would apply a linear transformation to each element of $\tilde{x}$ depending on which Markov state its residual value falls under. We can then define a $n{\times}n$ Markov transition probability matrix that defines such linear transformations.

$$P(k) = \begin{bmatrix} P_{11}(k) & P_{12}(k) & \ldots & P_{1n}(k) \\ P_{21}(k) & P_{22}(k) & \ldots & P_{2n}(k) \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1}(k) & P_{n2}(k) & \ldots & P_{nn}(k) \end{bmatrix}, \text{ where } P_{ij}(k) = \frac{Q_{ij}(k)}{Q_i(k)} \tag{20}$$

$P_{ij}(k)$ in the above matrix, named the "k-step transition matrix," represents the probability for the residual of the predicted time series variable at time $m+k$ to transition to S_j from the S_i state attained at time m. $Q_{ij}(k)$ represents the number of state transitions from S_i to S_j already observed from time m to $m+k$ for any m. $Q_i(k)$ represents the number of observations where the residual falls within state S_i.

It is also worthwhile noting that each of the row vectors of $P(k)$ produces a sum up to 1, or that $\sum_j P_{ij}(k) = 1$.

Correction of predicted value.   Finally, for the Multivariable Grey Model output $\tilde{x}(k)$ whose residual is determined to be at state $S_j = [L_j, U_j]$, we can produce an updated prediction

value of $\tilde{y}(k)$ as follows [45].

$$\tilde{y}(k) = \tilde{x}(k)\left(1 + \frac{L_j + U_j}{2}\right) \tag{21}$$

Model evaluation

The proposed method is evaluated on a prediction task to compare the proposed modeling method with other alternative methods to compare its performance. In this process, we use different numeric metrics to measure the model fit and prediction accuracy, including mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE) [46].

The following are the formulae used to evaluate those specific metrics.

$$MAPE_i(k) = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{x_i(k) - \tilde{x}_i(k)}{x_i(k)}\right| \times 100\% \tag{22}$$

$$MAE_i(k) = \frac{1}{n}\sum_{k=1}^{n}|x_i(k) - \tilde{x}_i k| \times 100\% \tag{23}$$

$$RMSE_i(k) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_i(k) - \tilde{x}_i k)^2} \times 100\% \tag{24}$$

Where $x_i(k)$ and $\tilde{x}_i(k)$ represented the actual and predicted time series variable values at time k, respectively.

Results

This section presents the results of analyzing the Michigan Traffic Crash Facts dataset with the Grey-Markov modeling method as described in the previous section. The model fitness and its prediction performance are also to be compared against alternative models presented in the literature.

Grey relational analysis

Before choosing appropriate forecast factors, we shall first calculate the grey incidence degree. The factors with distinct lower grey relational degrees will be omitted, reducing calculation and complexity to minimize the errors. According to Eq 2, and given synthetic coefficient $\rho = 0.5$ [42], the degrees of influence of each factor impacting the causality of crashes are shown as specified in the following Fig 2.

Similarly to Table 1, X_1 refers to crashes occurring on snow-covered roads, X_2 refers to head-on crashes against another vehicle performing a left turn, X_3 refers to a sideswipe crash, X_4 refers to crashes with distracted drivers, X_5 refers to crashes resulting in injuries, X_6 refers to crashes in the median of roads, X_7 refers to crashes involving buses and/or trucks, and X_8 refers to crashes in rainy weather.

The figure above shows us that the grey relation scores for roundabout accidents, driver distraction, and crashes involving buses/trucks are not high enough. In other words, factors with grey incidence values smaller than a set threshold of 0.7 can be omitted [42].
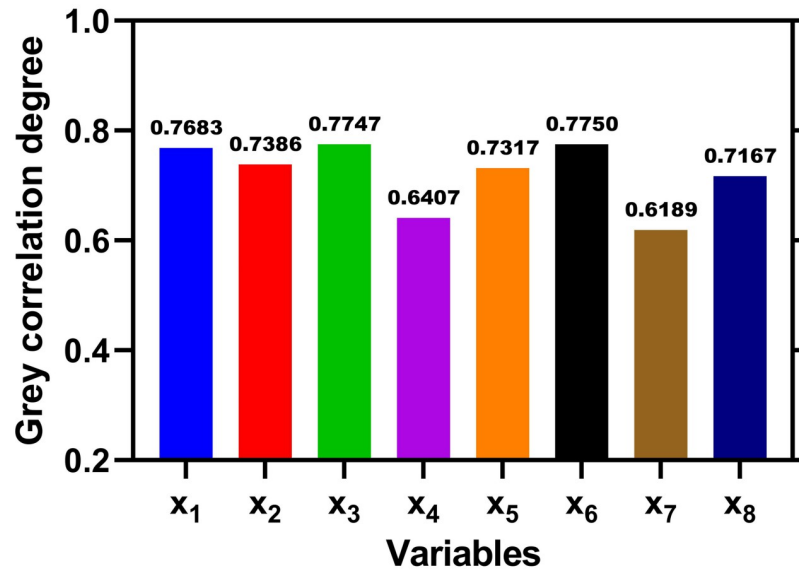
Fig 2. Grey relational grades of each independent variable associated with traffic crashes in roundabouts.

Results of model selection

The potential numbers of causality factors to traffic crashes in roundabouts are admittedly numerous, and it would be unrealistic and inefficient to consider every possible factor in a single model. Therefore, we decide that only a certain number of factors that are enough correlated with the occurrence of traffic crashes in roundabouts are to be considered. Our metric for this degree of correlation is "grey correlation factor." We also determined that only those causality factors with grey correlation factors over a threshold of 0.7 would be considered as the "main factors."
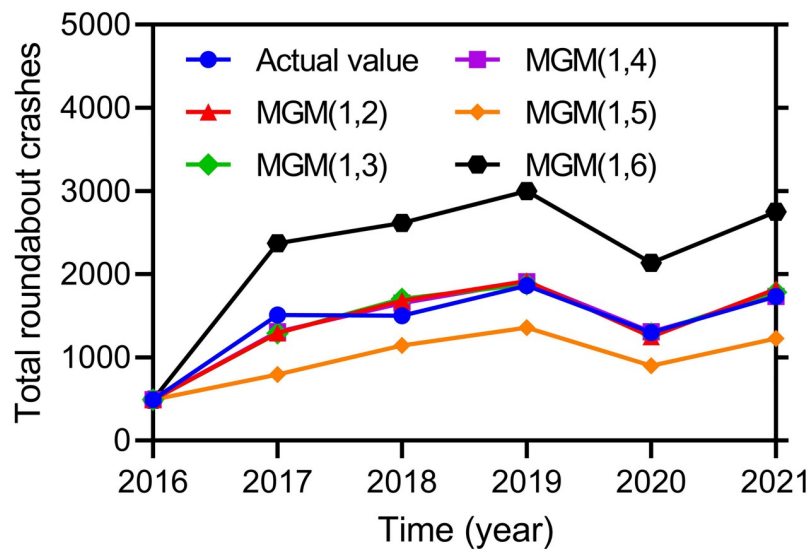

Fig 3. Fitted accuracies of different Multivariable Grey (MGM(1,n)) models.

Different Multivariable Grey (MGM(1, n)) models can be constructed with any number (n) of independent variables that have the highest ranking in their correlation grades, as provided in the previous section. In order to obtain the ideal number of independent variables for fitting our dataset, we can construct multiple Multivariable Grey models using a range of independent variables from 2 to 6. The MAPE percentage errors are computed for each model for each of the fitted values that it outputs as shown in Eq 22. The results of this evaluation process is shown below in Fig 3.

Similarly to Table 1, X_1 refers to crashes occurring on snow-covered roads, X_2 refers to head-on crashes against another vehicle performing a left turn, X_3 refers to a sideswipe crash, X_4 refers to crashes with distracted drivers, X_5 refers to crashes resulting in injuries, X_6 refers to crashes in the median of roads, X_7 refers to crashes involving buses and/or trucks, and X_8 refers to crashes in rainy weather.

Also note the following:

1. The input parameters for the MGM(1,2) model include X_3, X_6.

2. The input parameters for the MGM(1,3) model include X_1, X_3, X_6.

3. The input parameters for the MGM(1,4) model include X_1, X_2, X_3, X_6.

4. The input parameters for the MGM(1,5) model include X_1, X_2, X_3, X_5, X_6.

5. The input parameters for the MGM(1,6) model include X_1, X_2, X_3, X_5, X_6, X_8.

Due to the collinearity between different variables and the redundancy among variables, the MGM(1,n) model will have difficulty making accurate predictions due to confounding variables exerting other effects depending on the interference. Similarly, the most influential factors largely reflect the roundabout traffic collapse system changes. Referring to Fig 3, we then calculated the MAPE for MGM(1,2), MGM(1,3), MGM(1,4), MGM(1,5), and MGM(1,6) model as 6.39%, 5.50%, 4.47%, 26.43%, and 52.56%, respectively. Thus, the MGM(1,4) model with the highest prediction accuracy was selected as the best representation of the influential variables based on a comparison of prediction accuracy across models.

Hybrid, optimized MGM(1,4)-Markov model

A hybrid model was further developed based on the MGM (1,4) model. According to the relative error (R_E) between the actual value and the fitted value, we divided the relative error into three state intervals:

The state S_1 signifies underestimation of the fitted value, with the relative error satisfying the following condition.

$$\min(R_E) \leq R_E \leq \min(R_E) + \frac{|\min(R_E)| + |\max(R_E)|}{3} \tag{25}$$

The state S_2 is the normal state of the fitted value, with the relative error satisfying the following condition.

$$\min(R_E) + \frac{|\min(R_E)| + |\max(R_E)|}{3} \leq R_E \leq \max R_E - \frac{|\min(R_E)| + |\max(R_E)|}{3} \tag{26}$$

Table 2. Markov states of MGM(1,4) output values.

Year	Actual value	MGM(1,4)	Relative error	State
2016	489	489	0.0000	S_2
2017	1510	1308	0.1338	S_3
2018	1501	1653	-0.1013	S_1
2019	1864	1910	-0.0247	S_1
2020	1300	1308	-0.0062	S_2
2021	1730	1734	-0.0023	S_2

https://doi.org/10.1371/journal.pone.0287045.t002

The state S_3 signifies overestimation of the fitted value, with the relative error satisfying the following condition.

$$\max R_E - \frac{|\min(R_E)| + |\max(R_E)|}{3} \leq R_E \leq \max R_E \tag{27}$$

According to the relative error of the actual value to the fitted value, three statuses were identified: $S_1 := [-0.1013, -0.0229)$, $S_2 := [-0.0229, 0.0554)$, and $S_3 := [0.0554, 0.1338]$. The status divisions from the 2016 to 2021 series are shown in Table 2.

Based on the above state distributions, we can compute the following 1-, 2-, and 3-step probability matrices of Markov state transitions, as specified in Eq 22.

$$P(1) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \end{bmatrix}, P(2) = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}, P(3) = \begin{bmatrix} \frac{3}{8} & \frac{3}{8} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{8} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \tag{28}$$

Based the a 3-step transition probability matrices from Eq 28, we can compute the following predicted state transitions of the yearly roundabout crash numbers as in Table 3. Our calculations are based on the three most recent values and use different transfer steps to calculate the predicted values.

According to Table 3, the traffic crashes at Michigan roundabouts in 2022 were most likely in S_2. Accordingly, the predicted revised GM(1,1)–Markov chain value in 2022 based on the above prediction steps can be obtained as follows.

$$\tilde{y} = (1 + 0.5 \times (-0.029 + 0.0554)) \times 1730 \approx 1758 \tag{29}$$

Our study suggests that the number will likely increase slightly next year, implying that taking steps to minimize the likelihood is important.

Table 3. Predicted Markov state transitions of yearly traffic crash figures.

Year	Initial state	Transfer step	S_1	S_2	S_3
2021	S_2	1	0	1/2	1/2
2020	S_2	2	1/2	1/4	1/4
2019	S_1	3	3/8	3/8	1/4
Total			7/8	9/8	1

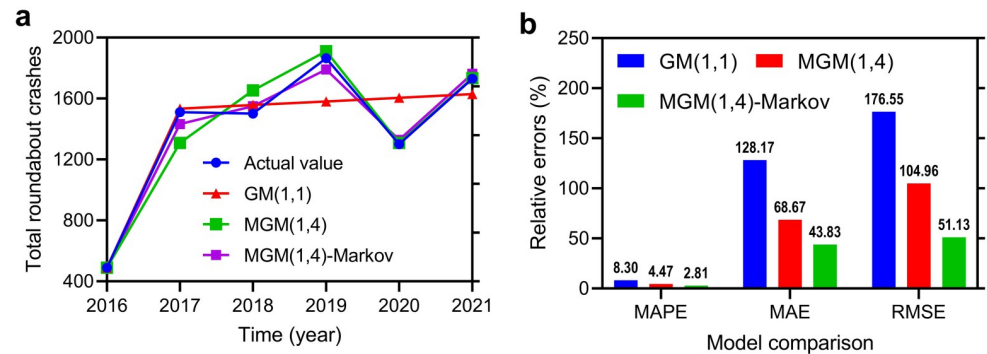https://doi.org/10.1371/journal.pone.0287045.t003

Fig 4. Comparison between existing GM [47] and MGMs [48] in literature and the MGM-Markov model. (a) Model-fitted total roundabout crashes; (b) Relative error values.

Model comparison

A hybrid approach combining the theory of grey systems and Markov chains has been used to predict roundabout accidents. Time series data from 2016 to 2021 were fitted using the GM (1,1) model [47], MGM(1,4) model [48], and hybrid MGM(1,4)-Markov model, as shown in Fig 4.

According to the results, the average MAPE values of GM(1,1) models, MGM(1,4) models, and Grey-Markov models were 8.30%, 4.47%, and 2.81%, respectively. The MAE values of the GM(1,1), MGM(1,4), and Grey-Markov models were 128.17%, 68.67%, and 43.38%, respectively. The RMSE values of the GM(1,1), MGM(1,4), and Grey-Markov models were 176.55%, 104.96%, and 51.13%, respectively. The fitting results indicate that the grey Markov model outperforms the single grey prediction in accuracy. When applying the grey model to compare actual incident data with the predicted values, we find that the relative errors are high and incongruous. However, after correcting the predictions of the grey model using a Markov chain process, we can enhance the prediction accuracy by combining multiple prediction models. Thus, the hybrid Grey-Markov model is more suitable for predicting the number of incidents from 2016 to 2021.

Discussion

To our knowledge, this is one of the first attempts to predict roundabout traffic crashes via an optimized multivariate Grey-Markov model. Our proposed MGM(1,n)-Markov model represents a valid and efficient method for predicting round-about traffic crashes in Michigan under insufficient data, temporal fluctuations, and trend changes. In summary, the findings of this study may assist policymakers in formulating policies to improve roundabout traffic safety and potentially reduce crashes. Given the results of this study, the remainder of this section elaborates on gaps in the existing literature and priority areas that should be addressed in future studies of traffic crashes at roundabouts.

Summary of evidence

First, we identify eight roundabout traffic crash impact factors and then apply the grey relational algorithm to prioritize these impact factors. Roundabout traffic accident risk can be assessed from various perspectives using these indicators, which have proven effective. Roundabout traffic crashes are influenced by multiple sources of information, including humans, vehicles, and surrounding elements, each of which interacts with the others and influences the

outcome [3, 4, 22, 39]. Meanwhile, annual traffic crash prediction can be viewed as a problem associated with the grey system because several factors (i.e., humans, vehicles, and surroundings) influence roundabout traffic crashes, but the precise relationships between these factors are, at least within the scope of this study, unclear. Among the factors influencing roundabout traffic accidents in Michigan, crashes on the median strip (grey correlation 0.8186) have the greatest impact. The constructed model can predict these factors with high accuracy. The traffic administration can use the results of this study to control the impact factors and reduce the likelihood of traffic accidents at roundabouts.

Second, we found that multivariate optimization (MGM(1,n)) outperforms uni-variate optimization (GM(1,1)) because multiple variables are simultaneously considered. Our analysis indicates that the one-dimensional GM(1,1) model performed well with a good prediction, while the optimized four-dimensional MGM(1,4) model performed much better with a lower average relative error. Hybrid MGM(1,4)-Markov models can better fit the nonlinear functions of raw time sequences than single grey models and MGM(1,4) models, making them effective tools for preventing traffic accidents at roundabouts. There is evidence that the hybrid MGM (1,4)-Markov model produced more accurate predictions and that the forecasted results may prove useful in assisting management in dealing with traffic issues. A Markov chain is well-suited to deal with intra-sequence fluctuations since the grey model is well-suited to exponential sequences. The results of this study indicate that the hybrid model is more accurate in terms of fitting data than the basic GM(1,1) model. Despite the reduction in accuracy, the Markov model provides interval-based predictions and improves prediction accuracy. In general, four-dimensional MGM(1, n) is a more accurate prediction model based on relevant variables' influence than other models.

Finally, Markov chains have been proven to be a valid methodological approach to enhance the fitting accuracy of the MGM(1,n) model through modification of the prediction results. A Markov chain process is characterized by the fact that it does not require a large amount of data; Only a limited amount of information is needed to make a prediction [49]. This may be because a Markov probability matrix can be used to determine the transfer rules of states, while the MGM(1,n) model can reveal the predicted data's development trend. With the combination of the two models, the resulting information from the raw data can be adequately evaluated, resulting in high prediction accuracy and stochastic volatility. Prior studies [43, 50] has focused primarily on optimizing background values; however, residual modification models are rarely applied to the MGM(1,n) model. The prediction accuracy of traditional residual modification models has been improved by developing modified residual modification models such as the Grey model, Markov chains, Fourier series, genetic programming, and neural networks. Based on comparisons of the results of the MGM(1,n)-Markov model, the GM (1,1) model, and the optimized MGM(1,n) model, the first model is more accurate, with a prediction error of 3.02%, a much smaller error than the alternative models of 8.30% and 4.47%, which provides evidence that the method is feasible. To date, only a few existing studies, such as Zouhair et al. [51] and Govindan et al. [52], the Grey-Markov predictive models, have been used to predict maritime incidents and traffic volumes.

Interpretation of results

Predicting roundabout accidents based on limited data is essential in traffic management and decision-making when prioritizing road safety projects. There have been some qualitative or quantitative studies of the factors contributing to these accidents in the past, but a clear and complete picture is difficult to obtain due to incomplete information and additional factors that interact [53]. The following reasons may explain these findings.

The first interpretation concerns exploratory studies with limited sample sizes. This is mainly due to the dynamic and time-varying nature of roundabout traffic systems and finite volume data, indicating a non-stationary stochastic process with definite trends. Typically, previous models require a certain amount of historical data to make predictions, which are not guaranteed to be accurate with limited data [54]. Therefore, it is not possible to use these models to predict road accidents, as they are not suitable for that purpose. Moreover, models based on older data are less accurate and add uncertainty than models based on more recent data, resulting in reduced estimation accuracy [55]. In future studies, it would be beneficial to consider ideal data collection times for the grey model.

The second explanation involves factors that affect predictive accuracy, which was complex before. A grey relational analysis of traffic crashes shows that the importance of the grey relational rank differs between reference and comparison series. This suggests that additional influences influence roundabout traffic crashes, and each influence plays a distinct role in influencing these crashes. Interactions between humans, vehicles, and surroundings also lead to more uncertainty in the prediction results and large data fluctuations. On the other hand, we found that a single grey model (i.e., GM(1,1) and MGM(1,n)) is not optimal in terms of prediction accuracy. In this case, this can be explained by the model requiring exponential regularity for the data sequence, and a relatively high degree of stochastic fluctuations is not acceptable [12, 13].

The third interpretation involves a difference in methodology considerations. It is possible to characterize roundabout traffic accidents as time series data affected by human interaction, vehicles, and surrounding factors. Choosing a suitable forecasting algorithm is vital to traffic accident forecasting. The traditional forecasting methods (e.g., parametric models [28, 29] and non-parametric models [20, 30–33, 56]) play an essential role in actual work; however, there are limitations due to the distribution and size of the sample. While grey prediction can overcome insufficient data-driven shortcomings of traditional estimation models, it is relatively poor at predicting roundabout traffic crashes, which are largely fluctuating data series influenced by humans, vehicles, and surroundings. It has been demonstrated that Markov chains apply to large fluctuations in data series; thus, by combining them with the grey prediction model, limitations due to a lack of data and changes in the raw data sequence can be overcome. The usage of its forecast for roundabout traffic crashes can deal with data sequences that fluctuate highly, as well as takes into account the interaction factors (e.g., humans, vehicles, surroundings) with a great scientific and practical method. Accordingly, the hybrid MGM(1,n)-Markov model has a high level of accuracy in its prediction for roundabout traffic crashes from 2016 to 2021 (Fig 3).

Our finding was that the single GM(1,1) or MGM(1,n) models were found to have a large relative error of prediction in some years, reflecting the fluctuation of roundabout traffic crashes. However, after error testing, the model's accuracy was qualified, indicating that it is suitable for predicting roundabout traffic accidents. The error in the prediction value needs to be corrected. A hybrid Grey-Markov chain model is constructed, and the Markov chain method is used to modify the predictions after making them. As a result of reducing prediction error and performing an accuracy test, we found that the Grey-Markov chain model can make more accurate predictions. Our results suggest that the Grey-Markov chain model performs significantly better than the single GM(1,1) model, as shown in Fig 3. Studies have shown that the Markov chain achieves high accuracy by correcting predictions' fitted curves to match the fluctuations of the actual values.

Nevertheless, a note should be made regarding the number of state partitions included in the Grey-Markov model. State divisions determine the accuracy of their predictions, which are not standardized but depend on the amount of historical data. More states should be split if

there is a large amount of historical data, resulting in greater prediction accuracy. Large state divisions have enhanced prediction accuracy, yet, too many states also lead to small samples per state and low transition probabilities. Therefore, a reasonable division of states number is recommended to improve prediction accuracy.

Conclusions and perspectives

This paper constructed a hybrid-optimized Grey-Markov model and compared it with two additional existing models to predict roundabout traffic crashes in Michigan from 2016 to 2021. The optimized multivariate Grey-Markov model outperforms other grey models for predicting roundabout traffic accidents, achieving higher prediction accuracy. We can also conclude that similar approaches to that undertaken in this study can become suitable for solving problems related to long-term prognosis and fluctuations in data, as demonstrated in the experiments.

A summary of the key findings of this study are listed as follows.

1. Roundabout traffic accidents can be caused by various factors, including humans, vehi-cles, and the surrounding environment. Our findings demonstrate that the leading major causes of accidents in Michigan roundabouts are ranked as follows: crash on median (0.7750), sideswipe (0.7747), snow-covered road (0.7683), head-on crash on left turn (0.7386), crash with injury (0.7317), crash in rainy weather (0.7167), distraction driver (0.6407), and buses/trucks crash (0.6189) through grey relational analysis.

2. In the context of roundabout traffic accidents, the proposed hybrid Grey-Markov model is suitable for short-term predictions with limited sample data. Grey system theory can handle uncertainty in the data, and Markov chain processes enhance its overall accuracy by correcting its predictions. For a limited sample of cases, the hybrid multivariate Grey-Markov model performs exceptionally well, demonstrating a close match between predicted and actual values. Thus, roundabout accident prediction with limited data is effectively applied with a mean relative error of 3.02%.

3. Several countermeasures have been developed in response to these risk factors affecting roundabout crashes, urging the government, society, and drivers to work together. Further studies in the vehicle region may benefit from examining more detailed incident data to identify causal relationships between roundabout collisions.

The proposed prediction method makes several practical contributions to roundabout traffic. First, we offered a hybrid roundabout traffic crash prediction model that combines grey systems theory and Markov chains, which has significant engineering applications to assist policymakers in predicting roundabout traffic accidents with limited data. To our knowledge, there have been no extensive studies of such combined prediction methods; Second, our proposed approach provides a detailed analysis of how impact factors relate to roundabout traffic crashes. The grey relational analysis shows that the top three causes of roundabout accidents in Michigan over the past six years were crashes in the median, head-on collisions in left turns, and sideswipes from the same direction. Therefore, authorities should consider the findings of this study when planning future traffic enforcement priorities for roundabouts, considering the limited number of enforcement officers available. Finally, the results provide a basis for preventing roundabout traffic crashes in the future and similar methodological applications (e.g., maritime, mining, and civil aviation accidents). In other words, this approach simultaneously examines human, vehicle, and surrounding factors and provides insights into fundamental issues regarding future roundabout traffic and practical countermeasures.

We propose the following countermeasures based on the characteristics and predictions of traffic collisions at roundabouts:

1. Governments should establish appropriate mechanisms to formulate specific policies regarding roundabout traffic safety based on the attributes of roundabouts and analyze the available data to extract valuable information.

2. Law enforcement should intensify their efforts at intersections by enforcing strict sanctions against violators to control the reckless behavior of drivers. Decision makers may benefit from the results of this study in dealing with roundabouts and other traffic control tools. This study should, however, be extended to further locations in Michigan with different characteristics.

3. Vehicle drivers should pay adequate attention to traffic safety. We found that driver inattention or distraction is one of the leading causes of roundabout accidents. Therefore, regular publicity must inform drivers about unsafe practices that can lead to roundabout accidents and serious consequences. Traffic safety must be taken very seriously, and all regulations must be strictly followed.

There are a few limitations that must be considered in this study. First, the roundabout crash data collected for this study were derived from official reports with limited observations, which could be biased, and the data should be interpreted cautiously. Due to the limited sample size, we may only be able to provide preliminary results since the information gathered may not be sufficient and a larger database on roundabouts is necessary. Second, roundabout traffic crashes are influenced by multiple factors, yet we only consider eight key variables. Further research will be conducted to incorporate additional roundabout traffic crash factors into the model and to improve the predictive model from the point of view of the mechanisms involved. Third, the hybrid Grey-Markov model is a predictive approach with limited data conditions, and further sensitivity analysis is necessary to determine the optimal data sample size and state partition to improve the prediction accuracy. Finally, the data are from Michigan and may not apply directly to settings in other states of the U.S. or other countries. Additionally, this improved model may be extended to alternative fields (e.g., maritime, mining, and civil aviation accidents); however, additional theoretical and applied studies are needed.

Despite the hybrid MGM(1,4)-Markov model performing well in predicting round-about traffic accidents, further expansion and verification of the grey forecasting model is expected in future studies. Further research is needed to combine other artificial intelligence approaches to solve uncertain system modeling for long-term forecasting effects. By expanding model construction and relaxing parameter setting and model construction restrictions, artificial intelligence can be used to compensate for the traditional grey model's obvious linear feature, thus compensating for its shortcomings.

Acknowledgments

Author Contributions

Conceptualization: Peijing Li.

Data curation: Peijing Li.

Formal analysis: Peijing Li, Jian Li.

Investigation: Peijing Li.

Methodology: Peijing Li, Jian Li.

Resources: Peijing Li.

Software: Jian Li.

Supervision: Peijing Li.

Validation: Peijing Li, Jian Li.

Visualization: Jian Li.

Writing – original draft: Peijing Li.

Writing – review & editing: Peijing Li, Jian Li.

References

1. Elvik R. Road safety effects of roundabouts: A meta-analysis. Accident Analysis & Prevention. 2017 Feb; 99:364–371. Available from: https://www.sciencedirect.com/science/article/pii/S0001457516304560. https://doi.org/10.1016/j.aap.2016.12.018 PMID: 28064101

2. Park M, Lee D, Park JJ. An Investigation of the Safety Performance of Roundabouts in Korea Based on a Random Parameters Count Model. Journal of Advanced Transportation. 2018 Oct; 2018:e5628293. Publisher: Hindawi. Available from: https://www.hindawi.com/journals/jat/2018/5628293/. https://doi.org/10.1155/2018/5628293

3. Kountouriotis GK, Merat N. Leading to distraction: Driver distraction, lead car, and road environment. Accident Analysis & Prevention. 2016 Apr; 89:22–30. Available from: https://www.sciencedirect.com/science/article/pii/S0001457515301846. https://doi.org/10.1016/j.aap.2015.12.027 PMID: 26785327

4. Zhang L, Zhang M, Tang J, Ma J, Duan X, Sun J, et al. Analysis of Traffic Accident Based on Knowledge Graph. Journal of Advanced Transportation. 2022 Aug; 2022:e3915467. Publisher: Hindawi. Available from: https://www.hindawi.com/journals/jat/2022/3915467/. https://doi.org/10.1155/2022/3915467

5. Montella A. Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. Accident Analysis & Prevention. 2011 Jul; 43(4):1451–1463. Available from: https://www.sciencedirect.com/science/article/pii/S0001457511000443. https://doi.org/10.1016/j.aap.2011.02.023 PMID: 21545878

6. Yang Y, Yuan Z, Meng R. Exploring Traffic Crash Occurrence Mechanism toward Cross-Area Freeways via an Improved Data Mining Approach. Journal of Transportation Engineering, Part A: Systems. 2022 Sep; 148(9):04022052. Publisher: American Society of Civil Engineers. Available from: https://ascelibrary.org/doi/10.1061/JTEPBS.0000698.

7. Xie N. Explanations about grey information and framework of grey system modeling. Grey Systems: Theory and Application. 2017 Aug; 7(2):179–193. Available from: https://www.emerald.com/insight/content/doi/10.1108/GS-05-2017-0012/full/html.

8. Liu S, Forrest J, Yang Y. A brief introduction to grey systems theory. Grey Systems: Theory and Application. 2012 Aug; 2(2):89–104. Available from: https://www.emerald.com/insight/content/doi/10.1108/20439371211260081/full/html.

9. Deng Jl. Control problems of grey systems. Systems & Control Letters. 1982 Mar; 1(5):288–294. Available from: https://www.sciencedirect.com/science/article/pii/S016769118280025X. https://doi.org/10.1016/S0167-6911(82)80025-X

10. Bezuglov A, Comert G. Short-term freeway traffic parameter prediction: Application of grey system theory models. Expert Systems with Applications. 2016 Nov; 62:284–292. Available from: https://www.sciencedirect.com/science/article/pii/S095741741630313X. https://doi.org/10.1016/j.eswa.2016.06.032

11. Yin MS. Fifteen years of grey system theory research: A historical review and bibliometric analysis. Expert Systems with Applications. 2013 Jun; 40(7):2767–2775. Available from: https://www.sciencedirect.com/science/article/pii/S0957417412012122. https://doi.org/10.1016/j.eswa.2012.11.002

12. Comert G, Begashaw N, Huynh N. Improved grey system models for predicting traffic parameters. Expert Systems with Applications. 2021 Sep; 177:114972. Available from: https://www.sciencedirect.com/science/article/pii/S0957417421004139. https://doi.org/10.1016/j.eswa.2021.114972

13. Duan H, Xiao X, Xiao Q. An inertia grey discrete model and its application in short-term traffic flow prediction and state determination. Neural Computing and Applications. 2020 Jun; 32(12):8617–8633. Available from: https://doi.org/10.1007/s00521-019-04364-w

14. Jin X, Zheng J, Geng X. Prediction of Road Traffic Accidents Based on Grey System Theory and Grey Markov Model. International Journal of Safety and Security Engineering. 2020 Apr; 10(2):263–268. Available from: http://www.iieta.org/journals/ijsse/paper/10.18280/ijsse.100214.

15. ZOU X, YUE WL, LI C. Road Traffic Accident Prediction using an SCGM (1,1)c-Markov Model. Asian Transport Studies. 2018 Mar; 5(1):191–205.

16. Liu Y, Wu C, Wen J, Xiao X, Chen Z. A grey convolutional neural network model for traffic flow prediction under traffic accidents. Neurocomputing. 2022 Aug; 500:761–775. Available from: https://www.sciencedirect.com/science/article/pii/S0925231222006567. https://doi.org/10.1016/j.neucom.2022.05.072

17. Kamla J, Parry T, Dawson A. Roundabout Accident Prediction Model: Random-Parameter Negative Binomial Approach. Transportation Research Record. 2016 Jan; 2585(1):11–19. Publisher: SAGE Publications Inc. Available from: https://doi.org/10.3141/2585-02

18. Gbologah FE, Guin A, Rodgers MO. Safety Evaluation of Roundabouts in Georgia. Transportation Research Record. 2019 Jul; 2673(7):641–651. Publisher: SAGE Publications Inc. Available from: https://doi.org/10.1177/0361198119843265.

19. Liu Q, Li C, Jiang H, Nie S, Chen L. Transfer learning-based highway crash risk evaluation considering manifold characteristics of traffic flow. Accident Analysis & Prevention. 2022 Apr; 168:106598. Available from: https://www.sciencedirect.com/science/article/pii/S0001457522000343. https://doi.org/10.1016/j.aap.2022.106598 PMID: 35180467

20. Rahim MA, Hassan HM. A deep learning based traffic crash severity prediction framework. Accident Analysis & Prevention. 2021 May; 154:106090. Available from: https://www.sciencedirect.com/science/article/pii/S0001457521001214. https://doi.org/10.1016/j.aap.2021.106090 PMID: 33740462

21. Yang Y, He K, Wang Yp, Yuan Zz, Yin Yh, Guo Mz. Identification of dynamic traffic crash risk for cross-area freeways based on statistical and machine learning methods. Physica A: Statistical Mechanics and its Applications. 2022 Jun; 595:127083. Available from: https://www.sciencedirect.com/science/article/pii/S0378437122001273. https://doi.org/10.1016/j.physa.2022.127083

22. Yang Y, Wang K, Yuan Z, Liu D. Predicting Freeway Traffic Crash Severity Using XGBoost-Bayesian Network Model with Consideration of Features Interaction. Journal of Advanced Transportation. 2022 Apr; 2022:e4257865. Publisher: Hindawi. Available from: https://www.hindawi.com/journals/jat/2022/4257865/. https://doi.org/10.1155/2022/4257865

23. Mitra S, Bhowmick D. Status of signalized intersection safety-A case study of Kolkata. Accident Analysis & Prevention. 2020 Jun; 141:105525. Available from: https://www.sciencedirect.com/science/article/pii/S0001457520306254. https://doi.org/10.1016/j.aap.2020.105525 PMID: 32334155

24. Park HC, Yang S, Park PY, Kim DK. Multiple membership multilevel model to estimate intersection crashes. Accident Analysis & Prevention. 2020 Sep; 144:105589. Available from: https://www.sciencedirect.com/science/article/pii/S0001457520303791. https://doi.org/10.1016/j.aap.2020.105589 PMID: 32593780

25. Lee J, Abdel-Aty M, Cai Q. Intersection crash prediction modeling with macro-level data from various geographic units. Accident Analysis & Prevention. 2017 May; 102:213–226. Available from: https://www.sciencedirect.com/science/article/pii/S0001457517301070. https://doi.org/10.1016/j.aap.2017.03.009 PMID: 28340414

26. Lu P, Zheng Z, Ren Y, Zhou X, Keramati A, Tolliver D, et al. A Gradient Boosting Crash Prediction Approach for Highway-Rail Grade Crossing Crash Analysis. Journal of Advanced Transportation. 2020 Jun; 2020:e6751728. Publisher: Hindawi. Available from: https://www.hindawi.com/journals/jat/2020/6751728/. https://doi.org/10.1155/2020/6751728

27. Panicker AK, Ramadurai G. Injury severity prediction model for two-wheeler crashes at mid-block road sections. International Journal of Crashworthiness. 2022 Mar; 27(2):328–336. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/13588265.2020.1806644.

28. Ihueze CC, Onwurah UO. Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria. Accident Analysis & Prevention. 2018 Mar; 112:21–29. Available from: https://www.sciencedirect.com/science/article/pii/S0001457517304542. https://doi.org/10.1016/j.aap.2017.12.016 PMID: 29306685

29. Deretić N, Stanimirović D, Awadh MA, Vujanović N, Djukić A. SARIMA Modelling Approach for Forecasting of Traffic Accidents. Sustainability. 2022 Apr; 14(8):4403. Available from: https://www.mdpi.com/2071-1050/14/8/4403. https://doi.org/10.3390/su14084403

30. Čubranić Dobrodolac M, Švadlenka L, Čičević S, Dobrodolac M. Modelling driver propensity for traffic accidents: a comparison of multiple regression analysis and fuzzy approach. International Journal of Injury Control and Safety Promotion. 2020 Apr; 27(2):156–167. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17457300.2019.1690002.

31. Liu Y, Wang R. Study on network traffic forecast model of SVR optimized by GAFSA. Chaos, Solitons & Fractals. 2016 Aug; 89:153–159. Available from: https://www.sciencedirect.com/science/article/pii/S096007791500329X. https://doi.org/10.1016/j.chaos.2015.10.019

32. Li P, Abdel-Aty M, Yuan J. Real-time crash risk prediction on arterials based on LSTM-CNN. Accident Analysis & Prevention. 2020 Feb; 135:105371. Available from: https://www.sciencedirect.com/science/article/pii/S0001457519311108. https://doi.org/10.1016/j.aap.2019.105371 PMID: 31783334

33. Xie Y, Lord D, Zhang Y. Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis. Accident Analysis & Prevention. 2007 Sep; 39(5):922–933. Available from: https://www.sciencedirect.com/science/article/pii/S0001457507000073. https://doi.org/10.1016/j.aap.2006.12.014 PMID: 17306751

34. Yang Y, Tian N, Wang Y, Yuan Z. A Parallel FP-Growth Mining Algorithm with Load Balancing Constraints for Traffic Crash Data [preprint]. Research Square; 2022. Available from: https://doi.org/10.21203/rs.3.rs-1311180/v1

35. Zheng X, Liu M. An overview of accident forecasting methodologies. Journal of Loss Prevention in the Process Industries. 2009 Jul; 22(4):484–491. Available from: https://www.sciencedirect.com/science/article/pii/S0950423009000552. https://doi.org/10.1016/j.jlp.2009.03.005

36. Roshandel S, Zheng Z, Washington S. Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis. Accident Analysis & Prevention. 2015 Jun; 79:198–211. Available from: https://www.sciencedirect.com/science/article/pii/S0001457515000871. https://doi.org/10.1016/j.aap.2015.03.013 PMID: 25841161

37. Shibata A, Fukuda K. Risk factors of fatality in motor vehicle traffic accidents. Accident Analysis & Prevention. 1994 Jun; 26(3):391–397. Available from: https://www.sciencedirect.com/science/article/pii/0001457594900132. https://doi.org/10.1016/0001-4575(94)90013-2 PMID: 8011052

38. Theofilatos A, Yannis G. A review of the effect of traffic and weather characteristics on road safety. Accident Analysis & Prevention. 2014 Nov; 72:244–256. Available from: https://www.sciencedirect.com/science/article/pii/S0001457514001900. https://doi.org/10.1016/j.aap.2014.06.017 PMID: 25086442

39. Yang Y, Yuan ZZ, Sun DY, Wen XL. Analysis of the Factors Influencing Highway Crash Risk in Different Regional Types Based on Improved Apriori Algorithm. Advances in Transportation Studies. 2019; 49. Available from: https://trid.trb.org/view/1667714.

40. The Michigan Office of Highway Safety Planning. Michigan Traffic Crash Facts. Ann Arbor, MI: University of Michigan;. Available from: https://www.michigantrafficcrashfacts.org/querytool.

41. Liu Y, Huang X, Duan J, Zhang H. The assessment of traffic accident risk based on grey relational analysis and fuzzy comprehensive evaluation method. Natural Hazards. 2017 Sep; 88(3):1409–1422. Available from: https://doi.org/10.1007/s11069-017-2923-2.

42. Liu S, Zhang H, Yang Y. Explanation of terms of grey incidence analysis models. Grey Systems. 2017; 7(1):136–142. Num Pages: 7 Place: Bingley, United Kingdom Publisher: Emerald Group Publishing Limited. Available from: https://www.proquest.com/docview/2088803258/abstract/D502E5A1E0454797PQ/1. https://doi.org/10.1108/GS-11-2016-0045

43. Lao T, Chen X, Zhu J. The Optimized Multivariate Grey Prediction Model Based on Dynamic Background Value and Its Application. Complexity. 2021 Feb; 2021:6663773. Publisher: Hindawi. Available from: https://www.hindawi.com/journals/complexity/2021/6663773/. https://doi.org/10.1155/2021/6663773

44. Ye J, Dang Y, Li B. Grey-Markov prediction model based on background value optimization and central-point triangular whitenization weight function. Communications in Nonlinear Science and Numerical Simulation. 2018 Jan; 54:320–330. Available from: https://www.sciencedirect.com/science/article/pii/S1007570417302113. https://doi.org/10.1016/j.cnsns.2017.06.004

45. Jia Zq, Zhou Zf, Zhang Hj, Li B, Zhang Yx. Forecast of coal consumption in Gansu Province based on Grey-Markov chain model. Energy. 2020 May; 199:117444. Available from: https://www.sciencedirect.com/science/article/pii/S036054422030551X. https://doi.org/10.1016/j.energy.2020.117444

46. Li J, Wang Y, Li J, Jiang R. Forecasting the Impact of the COVID-19 Outbreak on China's Cotton Exports by Modified Discrete Grey Model with Limited Data. AATCC Journal of Research. 2023 Jan;

p. 247234442211479. Available from: http://journals.sagepub.com/doi/10.1177/24723444221147966.

47. Mao M, Chirwa EC. Application of grey model GM(1, 1) to vehicle fatality risk estimation. Technological Forecasting and Social Change. 2006 Jun; 73(5):588–605. Available from: https://www.sciencedirect.com/science/article/pii/S0040162504001064. https://doi.org/10.1016/j.techfore.2004.08.004

48. Li W, Zhao X, Liu S. Traffic Accident Prediction Based on Multivariable Grey Model. Information. 2020 Mar; 11:184.

49. Grewal JK, Krzywinski M, Altman N. Markov models—Markov chains. Nature Methods. 2019 Aug; 16 (8):663–664. Number: 8 Publisher: Nature Publishing Group. Available from: https://www.nature.com/articles/s41592-019-0476-x.

50. Xie Nm, Liu Sf. Discrete grey forecasting model and its optimization. Applied Mathematical Modelling. 2009 Feb; 33(2):1173–1186. Available from: https://www.sciencedirect.com/science/article/pii/S0307904X08000206. https://doi.org/10.1016/j.apm.2008.01.011

51. Zouhair F, Kerby J. Forecasting U.S. Maritime Incidents using the Grey-Markov Model. Transportation Research Record. 2021 Sep; 2675(9):1621–1630. Publisher: SAGE Publications Inc. Available from: https://doi.org/10.1177/03611981211009219

52. Govindan K, Ramalingam S, Broumi S. Traffic volume prediction using intuitionistic fuzzy Grey-Markov model. Neural Computing and Applications. 2021 Oct; 33(19):12905–12920. Available from: https://doi.org/10.1007/s00521-021-05940-9

53. Liu S, Yang Y, Xie N, Forrest J. New progress of Grey System Theory in the new millennium. Grey Systems. 2016; 6(1):2–31. Num Pages: 30 Place: Bingley, United Kingdom Publisher: Emerald Group Publishing Limited. Available from: https://www.proquest.com/docview/2085699464/abstract/843F4E38298B40EEPQ/1. https://doi.org/10.1108/GS-09-2015-0054

54. Hassouna FMA, Al-Sahili K. Practical Minimum Sample Size for Road Crash Time-Series Prediction Models. Advances in Civil Engineering. 2020 Dec; 2020:e6672612. Publisher: Hindawi. Available from: https://www.hindawi.com/journals/ace/2020/6672612/. https://doi.org/10.1155/2020/6672612

55. Chen L, Lin W, Li J, Tian B, Pan H. Prediction of lithium-ion battery capacity with metabolic grey model. Energy. 2016 Jul; 106:662–672. Available from: https://www.sciencedirect.com/science/article/pii/S0360544216303401. https://doi.org/10.1016/j.energy.2016.03.096

56. Cai Q, Abdel-Aty M, Yuan J, Lee J, Wu Y. Real-time crash prediction on expressways using deep generative models. Transportation Research Part C: Emerging Technologies. 2020 Aug; 117:102697. Available from: https://www.sciencedirect.com/science/article/pii/S0968090X20306124. https://doi.org/10.1016/j.trc.2020.102697