Investigating the Effect of Outliers in Historical-Data on Grey Prediction Model for Power System State Estimation

A thesis submitted towards partial fulfillment of the requirements for the degree of

MASTER OF ENGINEERING

In

ELECTRICAL ENGINEERING

By

Prabal Samajdar

Class Roll Number: 002210802005
Examination Roll No.: M4ELE24005
Registration No.: 163480 of 2022-23
Academic Session: 2022-2024

Under The Guidance and Supervision of

Prof. (Dr.) Sunita Halder nee Dey

Professor
Department of Electrical Engineering,
Faculty of Engineering and Technology,
Jadavpur University, Kolkata – 700032,
West Bengal, India

Department of Electrical Engineering,
Faculty of Engineering and Technology,
Jadavpur University, Kolkata – 700032,
West Bengal, India

CERTIFICATE OF RECOMMENDATION

This is to certify that this dissertation entitled "**Investigating the Effect of Outliers in Historical-Data on Grey Prediction Model for Power System State Estimation**" has been carried out by **PRABAL SAMAJDAR** with Roll No. **002210802005** under my guidance and supervision and be accepted in partial fulfillment of the requirement for the degree of Master of Electrical Engineering. In my opinion this thesis, is worthy of its acceptance.

Signed by Supervisor of the Thesis-work

Prof. (Dr.) Sunita Halder nee Dey

Professor
Department of Electrical Engineering
Faculty of Engineering and Technology,
Jadavpur University, Kolkata – 700032,
West Bengal, India

Countersigned by

Prof. (Dr.) Biswanath Roy

Head
Department of Electrical Engineering
Faculty of Engineering and Technology,
Jadavpur University, Kolkata – 700032,
West Bengal, India

Prof. (Dr.) Dipak Laha

Dean
Faculty of Engineering and Technology,
Jadavpur University, Kolkata – 700032,
West Bengal, India

Department of Electrical Engineering,
Faculty of Engineering and Technology,
Jadavpur University, Kolkata – 700032,
West Bengal, India

CERTIFICATE OF APPROVAL*

The foregoing thesis is hereby approved as a creditable study of **Master of Electrical Engineering** and presented in a manner satisfactory for warranting its acceptance as a prerequisite to the degree for which it has been submitted. It is understood that by this approval, the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion herein but approve this thesis only for the purpose for which it is submitted.

Committee on Final Examination for Evaluation of the Thesis

Signatures of the Examiners

* Only in case the thesis is approved

Department of Electrical Engineering,
Faculty of Engineering and Technology,
Jadavpur University, Kolkata – 700032,
West Bengal, India

DECLARATION

I hereby certify that, except where due acknowledgement has been made, the work is solely of the candidate i.e. I alone. This thesis is a presentation of my original research work and has not been submitted previously, in whole or in part, to qualify for anyacademic award.

Furthermore, the content of the thesis is the result of work which has been carried out only since the official commencement date of the approved program. The work has been done under the guidance and supervision of ***Prof. (Dr.) Sunita Halder nee Dey***, Professor, Department of Electrical Engineering, Jadavpur University, Kolkata-700032, India. The pieces of information and data furnished above as well as in the report are authentic to the best of my knowledge and belief.

Name of the student	***: PRABAL SAMAJDAR***
Class Roll Number	***: 002210802005***
Registration Number	***: 163480 of 2022-23***
Examination Roll Number	***: M4ELE24005***
Title of Thesis	***: Investigating the Effect of Outliers in Historical-Data on Grey Prediction Model for Power System State Estimation***
Signature of the student	*:*

ACKNOWLEDGEMENT

It’s my pleasure to express gratitude from the core to everyone who has accompanied and assisted me in my thesis work. First and foremost, I would like to convey my heartfelt thankfulness to my mentor, **Prof. (Dr.) Sunita Halder nee Dey**, Professor, Department of Electrical Engineering, Jadavpur University, Kolkata-700032, India, for her invaluable guidance, suggestions, and encouragement throughout the discourse, which, by all means, empowered me in improving this work. It's been a privilege to work with her. Her encouragement and guardianship during the painful aftermath days of my father’s demise boosted my morale and confidence and without which, nothing perhaps would have been written in these pages.

I am indebted to **Prof. (Dr.) Biswanath Roy,** Head, Department of Electrical Engineering, Jadavpur University, Kolkata-700032, India and **Prof. (Dr.) Dipak Laha**, Dean, Faculty of Engineering and Technology, Jadavpur University, Kolkata-700032, India for their kind help and co- operation during the course of this thesis work.

Also special thanks to my batch-mates **Ranith Sarkar**, **Sajal Gayen**, **Soumik Ghosh**, **Sanju Sardar** and **Rajdeep Barua** whose support and companionship invaluably contributed to some quality days in the memory of my university-life.

Moreover, I would like to express my heartiest appreciation to all my family-members for their love and active support throughout the endeavor.

Date:

Prabal Samajdar
Student
Class Roll Number: 002210802005
Department of Electrical Engineering
Jadavpur University
Kolkata-700032
India

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

At the very dawn of delving into the titular subject matter, a little quanta of focus need to be put on load flow analysis. As known widely, to deploy load flow analysis, prerequisite can be either of the two forms namely information of active, reactive powers at the bus ends or knowledge of the said powers through transmission lines [1]. Naturally, such prerequisites are acquired by power measurements.

But, as soon as any sort of measurement gets into the scenario, errors follow it like its shadow and measuring powers is never an exception. An erroneous measurement will also lead to an erroneous load flow result and will completely spoil the objective of the analysis. Since no magical remedy is available in the market to make the measuring devices fully error-free, the onus is on statistical methods to reduce those errors as much as possible in calculations. State Estimation, under these circumstances, has been proved to be an excellent tool.

State Estimation of power system plays a key role to mitigate power measurement errors and thereby, it gives a usable estimate of power system states i.e. bus voltage magnitudes and the corresponding phase angles. Unfortunately, measurement error is not the only thing that affects the load flow analysis. Due to unanticipated failure of measuring gadgets and telemetering arrangements, any element in a set of measurement-data may become missing. This missing datum, which appears as a void, cannot be treated as an input even to the state estimation.

Therefore, State Estimation alone is insufficient to withstand the event of absence of any datum in its input parameters. So, first, the void needs to be replaced by some valid figure with the help of some relevant prediction methodology. In other words, existing vacancy of datum needs to be predicted first using historical data and then only, State Estimation may be exercised. Automatically, this brings into the scenario, the necessity of looking into suitable prediction techniques and their behaviors in the context of State Estimation of power system under different circumstances.

1.1 Literature Review

A thorough literature-survey finds out the commonly used present-day techniques for power-forecasting and they majorly include Fuzzy Prediction, Trend-Extrapolation, Neural Network Analysis, Analysis of Time-Series and Grey Prediction [2-6]. An application of fuzzy logic is worked out in [2] and it is equipped to adaptively predict wind power. The central point of this framework is that its performance can be optimally oriented through a fuzzy inference system and a tri-level adaptation functionality comprising of three trapezoidal fuzzy sets to track the inherent uncertainty of non-stationary wind power time series in different seasons as well as the uncertainty of the prediction model. Subject matter of [3] enlightens the linear exponential smoothening model of trend extrapolation by dint of mining electricity data features and its relations with environmental parameters and defines the electricity month-season ratio on the basis of trend extrapolation.

Literature presented in [4] shows the competence of the artificial neural networks named deep learning network and the deep belief network in wind power forecasting. In order to ensure the prediction accuracy, the numerical weather prediction data, actual wind speed data of wind farm and wind power data are used as the input of the prediction mechanism. Taking into account the uncertainty of numerical weather prediction data, the *K*-means clustering algorithm is also used to find the largest historical samples which have the greatest influence on forecasting accuracy. The stuff placed at [5] of bibliography section proposes and establishes the load forecasting techniques by utilizing time series analysis and neural network methods and by harvesting data from Thailand. Time series analysis utilizes autoregressive integrated moving average and seasonal autoregressive integrated moving average procedures. In addition, neural networks cover not only artificial neural network but also long-short term memory based recurrent neural network. Additionally, the Average True Range index is adapted to enhance the performance of recurrent neural network.

Contemplating the properties of the conventional Grey prediction model with a few input variables, a multivariable Grey prediction model is proposed in [6]. This new model rectifies the multivariable residual error and thereby improves prediction accuracy to a large extent. Following this modification, the prediction model is verified and validated with the historical load data in a certain area. The results show that the prediction accuracy of the improved prediction model is twice that of the traditional prediction model. Multiple Linear Regression (MLR) [7], like the Grey Prediction Model, is also a strong predictive procedure used in power system studies. It is used to find relationships between the target and predictors by fitting a linear equation to the observed data. It can also detect variables that are strongly correlated with the target variable. But unlike the Grey Prediction Model, the approach of MLR requires a large amount of historical data which are used to calculate regression coefficients for the latter case [7].

Out of all these prediction-making ways, Grey Prediction Model is both simple and least historical data-requiring simultaneously, hence preferred over the other prediction methods [8]. Grey Prediction Model is based on Grey Systems Theory which was propounded by Professor Deng in 1982. This model amends the defects of using historical data directly to build up a time series prediction model. It makes a holistic and dynamic analysis of multiple factors and reflect the dynamically changing relationship between the study variables series and the related factors. This method is successfully applied in numerous segments of energy and power systems, e.g. electricity price as well as consumption forecasting [9-11], predicting reliability parameters [12], wind-power prediction [13,14], life prediction of electric power transmission cable of ship [15] and the list goes on.

1.2 Objective

Despite having such large quantity of applications of the Grey Prediction Model in widespread segments of power system, there are very few studies related to its application in power system state estimation with respect to preprocessing of historical data. To fill this gap, this simple, yet powerful approach of Grey Prediction is effectively presented in [16]. But in [16], the historical data utilized are not crowded with outliers. So, how Grey Prediction reacts while facing outlier-enriched data for state estimation, is an area that needs further investigation. The present work, which is represented from [8], extends an endeavor to contribute to such investigation.

[16] gives a comparison of the predictions made by GM(1,1) with the values of the same given by central tendency (Mean in this case). If historical data values lie within a fairly smaller range, then mean-value of them may be capable enough to give a very good prediction. As soon as outliers come into the data, mean-values cannot be treated as a fair prediction since even a single outlier may significantly deviate the mean from the one in normal scenarios. In these cases, median values are of great use as they push the largest and smallest data-values to the ends of the data-sequence that is intermediately required for finding out median.

Suppose, number of data present in historical-data-set is N, out of which, M number(s) of data may be treated as outliers. If M is less than N/2, then median of the said N numbers of data may serve as a good enough prediction of future value. Once M equals or exceeds N/2, the median value also is of no more uses. So, in this case, a convenient, fast and less data-hungry predicting tool is required and the same is provided by GM(1,1). How efficiently the case M greater than or equals N/2 may be handled by the Grey Prediction model, GM(1,1), is at the core of this literature. For incorporating simplicity in calculations, the condition M=N/2 is chosen here and predictions are done for state estimation. All the predicted values of missing recorded data and the estimated values of states with respect to GM(1,1) and Mean and Median of historical data are also compared to study the relative accuracy and validity of GM(1,1) in dealing with outliers for power system state estimation.

CHAPTER 2

THEORY

2.1 Grey Prediction Model, GM(1,1)

Among the family of Grey Prediction Models, GM(1,1) is the most basic and most frequently used univariate model [8] and is able to acquire high prediction accuracy despite requiring small sample size. The steps of establishing GM(1,1) model, as presented in [8], are as follows:

1) Original time sequence with n samples is expressed as:

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots , x^{(0)}(n)) \quad (1)$$

where $X^{(0)}$ is a non-negative sequence and n is the sample size of the data.

2) First-order accumulative generation operation is used to convert chaotic series $X^{(0)}$ into monotonically increasing series $X^{(1)}$ as:

$$X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots , x^{(1)}(n)) \quad (2)$$

where

$$x^{(1)}(i) = \sum_{k=1}^{i} x^{(0)}(k) \quad (3)$$

3) The generated mean sequence $Z^{(1)}$ of $X^{(1)}$ is defined as:

$$Z^{(1)} = (z(2), z(3), \dots, z(n)) \quad (4)$$

where, for $k = 2,3, \dots, n$

$$z(k) = 0.5\, x^{(1)}(k) + 0.5\, x^{(1)}(k-1) \quad (5)$$

4) Predicted value of the primitive data at time point (k+1) is extracted by the following expression as depicted in [8]:

$$x_{prd}^{(0)}(k+1) = (1-e^{a})[\, x^{(0)}(1) - b/a\,] \quad (6)$$

where

$$[\, a \;\; b\,]^{T} = (\boldsymbol{B}^{T}\boldsymbol{B})^{-1}\boldsymbol{B}^{T}\boldsymbol{Y} \quad (7)$$

Values of the matrices $\boldsymbol{Y}$ and $\boldsymbol{B}$ are calculated by the following expressions as depicted in [8]:

$$\boldsymbol{Y} = [\, x^{(0)}(2) \quad x^{(0)}(3) \quad \dots \quad x^{(0)}(n)\,]^{T} \quad (8)$$

$$\boldsymbol{B} = \begin{bmatrix} -z(2) & 1 \\ -z(3) & 1 \\ \vdots & \vdots \\ -z(n) & 1 \end{bmatrix} \quad (9)$$

So, expressing in words instead of algebraic equations, it can be said that Grey Prediction Model, by virtue of equation (6), tries to give a generalized expression of future value of the very initial time data set; i.e. if there are 'k' number of data in the primitive data set, the model forms a general rule to determine 'k+1'th value; this generalized value is formed with the help of all the data present in that set.

To check the mathematical validity of the above cited outcomes, a pure index sequence is treated as the primitive data set in [17] and such validity is established with the help of Discrete Grey Model (DGM). [17] also enumerates the parallelism of Discrete Grey Model (DGM) and GM(1,1) and infers that DGM model and GM(1, 1) model can be considered as different expressions of the same model.

2.2 State Estimation by WLS Method

Estimation of the values of state variables is done iteratively [18], starting with either available erroneous measurements or initially assumed data. Estimate of state vector ***X*** at (p+1)th iteration can be determined by Weighted Least Squares (WLS) method [19] as follows:

$$\boldsymbol{X_{est}}^{(p+1)} = \boldsymbol{X_{est}}^{(p)} + \boldsymbol{\Delta}\, \boldsymbol{X_{est}}^{(p)} \tag{10}$$

where

$$\boldsymbol{X} = [\, x_1 \;\; x_2 \;\; x_3 \ldots \,]^{\mathrm{T}} \tag{11}$$

In (11), x_1, x_2, x_3 etc. are state variables. In (10), $\boldsymbol{X_{est}}^{(p)}$ is the estimate of state vector $\boldsymbol{X}$ at p^{th} iteration, $\boldsymbol{\Delta}\, \boldsymbol{X_{est}}^{(p)}$ is change to be added with $\boldsymbol{X_{est}}^{(p)}$ to calculate $\boldsymbol{X_{est}}^{(p+1)}$, which is the estimate of state vector $\boldsymbol{X}$ at $(p+1)^{\text{th}}$ iteration. $\boldsymbol{\Delta}\, \boldsymbol{X_{est}}^{(p)}$ is calculated, as explained in [19], by dint of the following expression:

$$\boldsymbol{\Delta}\, \boldsymbol{X_{est}}^{(p)} = (\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\begin{bmatrix} z_1 - f_1(\boldsymbol{X}) \\ z_2 - f_2(\boldsymbol{X}) \\ \vdots \end{bmatrix} \tag{12}$$

In (12), z_i is the measurement given by i^{th} meter i.e. i^{th} measurement and $f_i\,(\boldsymbol{X})$ is function of $\boldsymbol{X}$ which is used to calculate the value being measured by the i^{th} measurement. For example, we know that current = potential difference / resistance = f (say) i.e. in this example current is a function (f) of both potential difference and resistance. Now, we can achieve the value of current either by measuring the same with ammeter or by calculating it with the help of the function f, provided that the values of independent variables are available. In the similar fashion, WLS method requires both the i^{th} measurement and the function $f_i\,(\boldsymbol{X})$ which can calculate the entity supplied by i^{th} measurement; in other words, they are giving measured value and calculated value respectively. Values of ***R*** and ***H*** are calculated with the help of the following expressions respectively:

$$\boldsymbol{R} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots \\ 0 & \sigma_2^2 & 0 & \cdots \\ 0 & 0 & \sigma_3^2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{13}$$

$$\boldsymbol{H} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} & \cdots \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{14}$$

In (13), σ_i^2 is the variance for i^{th} measurement. The iterations terminate if the norm of difference between two successive estimated values lies within a specified tolerance [1]. The algorithm of this iterative process is as follows:

Step 1: Choose initial guess i.e. starting value of ***X*** (it may be obtained from load flow study) .

Step 2: Read measurements i.e. z_i .

Step 3: Set iteration count, $p = 0$ and set $\boldsymbol{X_{est}}^{(p)}$ = Initial guess of ***X*** .

Step 4: Calculate f_i (***X***), ***R*** and ***H*** by (13), (14) .

Step 5: Compute $\boldsymbol{\Delta X_{est}}^{(p)}$ by (12) .

Step 6: Compute updated value of $\boldsymbol{X}$ i.e. $\boldsymbol{X_{est}}^{(p+1)}$ by (10) .

Step 7: Calculate tolerance i.e. $\| \boldsymbol{X_{est}}^{(p+1)} - \boldsymbol{X_{est}}^{(p)} \|$.

Step 8: Check whether tolerance is less than a desired specified number. If so, go to step 10 else go to step 9 .

Step 9: Increment iteration count by 1 i.e. $p = p + 1$ and go to step 4 .

Step 10: Terminate iterations.

CHAPTER 3

CASE ANALYSIS

3.1 Simulation

In order to explore the efficacy of GM(1,1) for handling outlier-enriched data with respect to state estimation of power system, a three-bus power system, as detailed in [20], is chosen and same is illustrated below in Fig. 1.

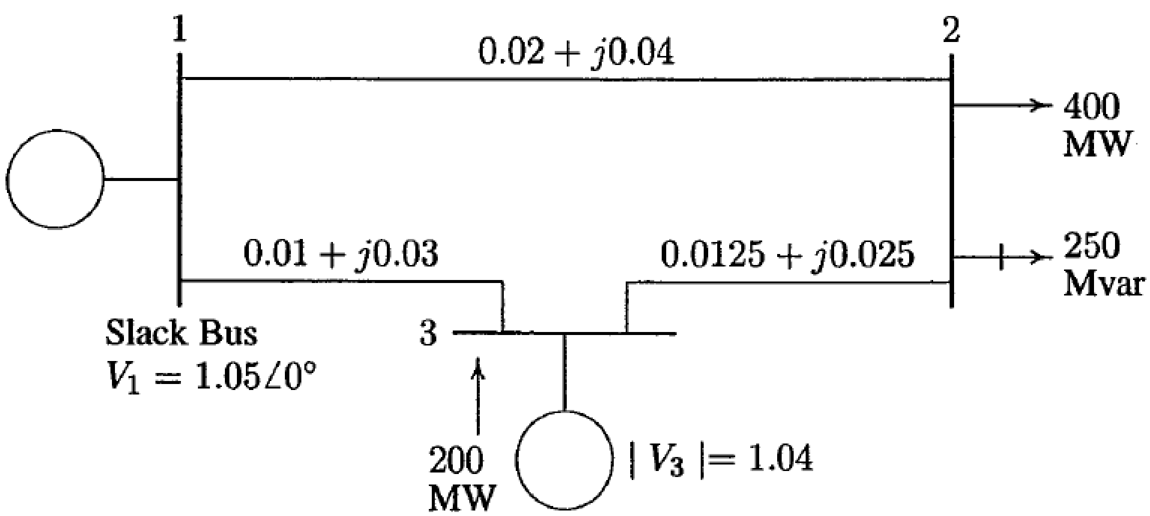

Fig. 1. Three-bus power system under consideration.

As seen in Fig. 1, the system consists of one swing bus (bus-1), one PQ bus (bus-2) and one PV bus (bus-3). Line Parameters of this system are systematically shown TABLE I.

TABLE I. LINE DATA

FROM BUS	TO BUS	R (pu)	X (pu)
1	2	0.02	0.04
1	3	0.01	0.03
2	3	0.0125	0.025

Making use of the above-mentioned data, at the very beginning, steady-state load flow analysis by Newton-Raphson method [1] is performed on the system using MATLAB script program. The results obtained are then validated with the outcomes of the simulation of Fig. 1 in SIMULINK platform of MATLAB software. Fig. 2 illustrates the said simulation. These results are now considered as the true values of the unknown quantities.

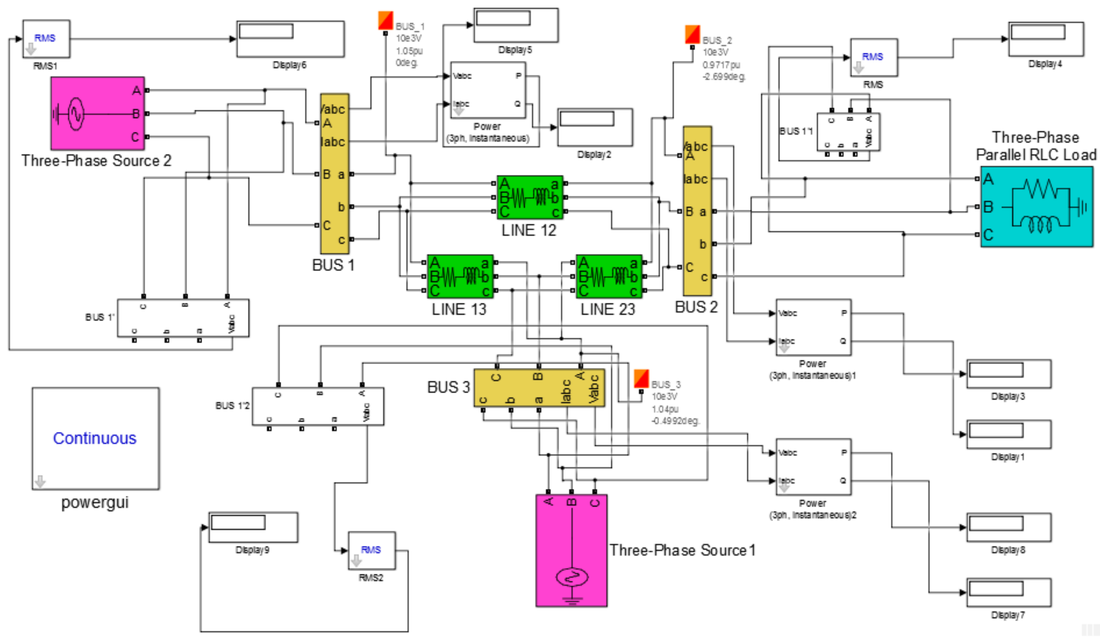

Fig. 2. Simulation of the three-bus power system under consideration in SIMULINK.

Suppose the true values at bus *i* are expressed as follows: active power injection is P_i, reactive power injection is Q_i, voltage magnitude is V_i and phase angle of voltage is δ_i. Then, figuratively, these true values (on 100 MVA and 10 kV bases) are enlisted in TABLE II.

TABLE II. LOAD FLOW ANALYSIS RESULT

i	P_i **(p.u.)**	Q_i **(p.u.)**	V_i **(p.u.)**	δ_i **(Degree)**
1	2.1840	1.4080	1.0500	0.0000
2	-4.0000	-2.5000	0.9716	-2.6964
3	2.0000	1.4620	1.0400	-0.4988

After having the knowledge of the true power-injection values, the following steps are pursued:

STEP-1

A random sequence of six numbers is generated whose values lie within ±10 percent of their corresponding true values (i.e. P_1, P_2, P_3, Q_1, Q_2 and Q_3). This sequence is treated as the set of errors.

STEP-2

The said sequence, when added with the corresponding true values, gives an artificially created measurement data-set. Since, this artificial data-set contains random errors, hence it is very much realistic and resembles a feasible situation as surrogate data.

STEP-3

24 more numbers of such data-sets are created by iterating step 1 and 2 consecutively 24 times and hence, total 25 such data-sets are prepared which serve as normal data.

STEP-4

25 more measurement-data-sets are also created, but this time, with values randomly 2 to 3 times their corresponding true values (i.e. P_1, P_2, P_3, Q_1, Q_2 and Q_3). They serve as outliers.

So, now, for each of P_1, P_2, P_3, Q_1, Q_2 and Q_3, there are 25 normal data (which contain some errors) and 25 outlying data/outliers (whose values are far beyond possible range of values).

STEP-5

Another set of 25 random numbers, lying among 1 to 50, are generated and they serve as the positions of the 25 outliers; i.e. these positions are populated with the 25 outliers and rest of the positions are populated with normal data.

So, now, there lies a set of 50 artificially generated historical data for each of P_1, P_2, P_3, Q_1, Q_2 and Q_3.

STEP-6

Now, these 6 historical data-sets, each with 50 data, are fed to the GM(1,1) algorithm, as explained in chapter 2, by MATLAB software and the latest value i.e. the 51st datum is predicted for each of P_1, P_2, P_3, Q_1, Q_2 and Q_3.

For the sake of comparison, mean and median are also calculated for each.

STEP-7

The procedures undergone in steps 1 to 6 are now repeatedly executed 49 more times (so total 50 times execution) to calculate 49 predicted values (so total 50 predicted values including the value of the very first iteration) with respect to the three different methods namely GM(1,1), mean and median.

3.2 Results

The procedures, as mentioned in the aforesaid section, yield following outcomes in pu and on 100 MVA base and are illustrated in Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig. 7 and Fig. 8.

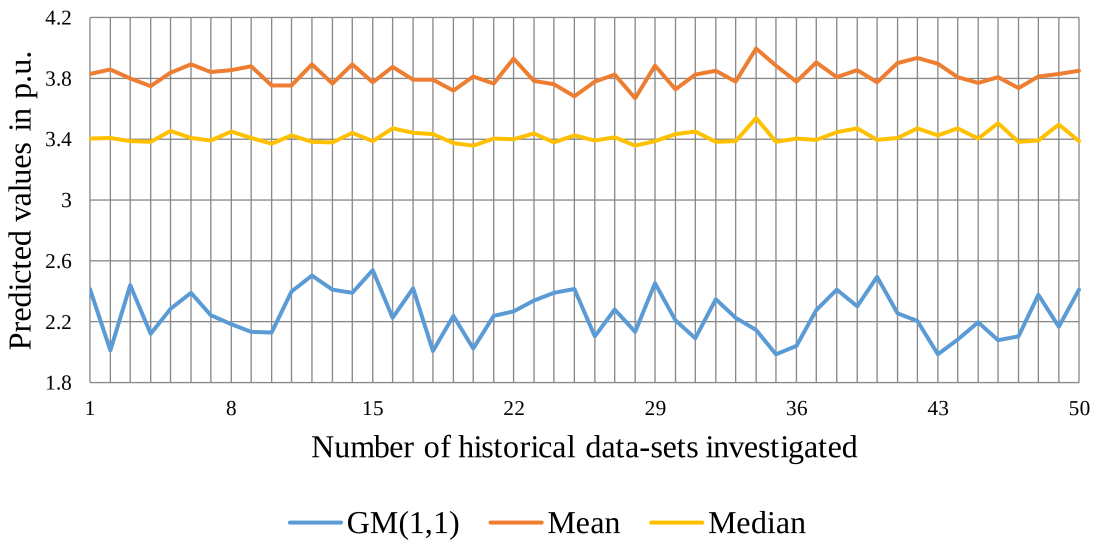

Fig. 3. Predicted values of P_1 in the three different methods.

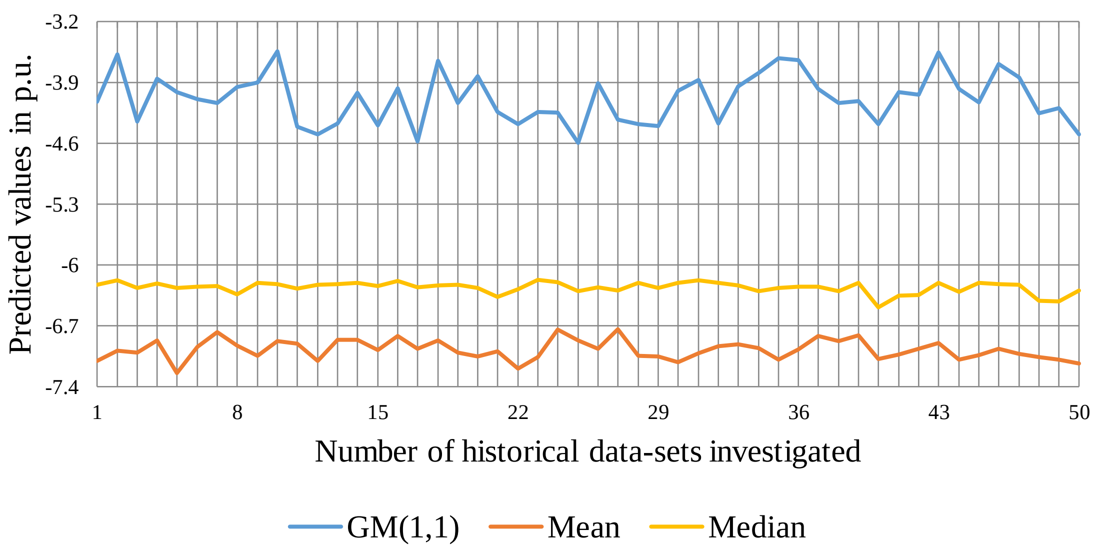

Fig. 4. Predicted values of P_2 in the three different methods.

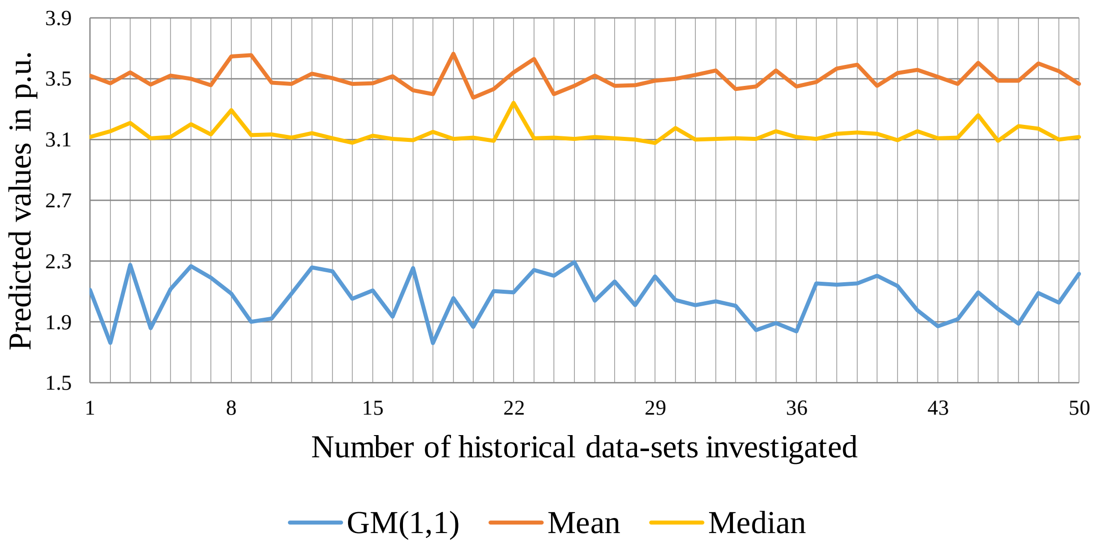

Fig. 5. Predicted values of P_3 in the three different methods.

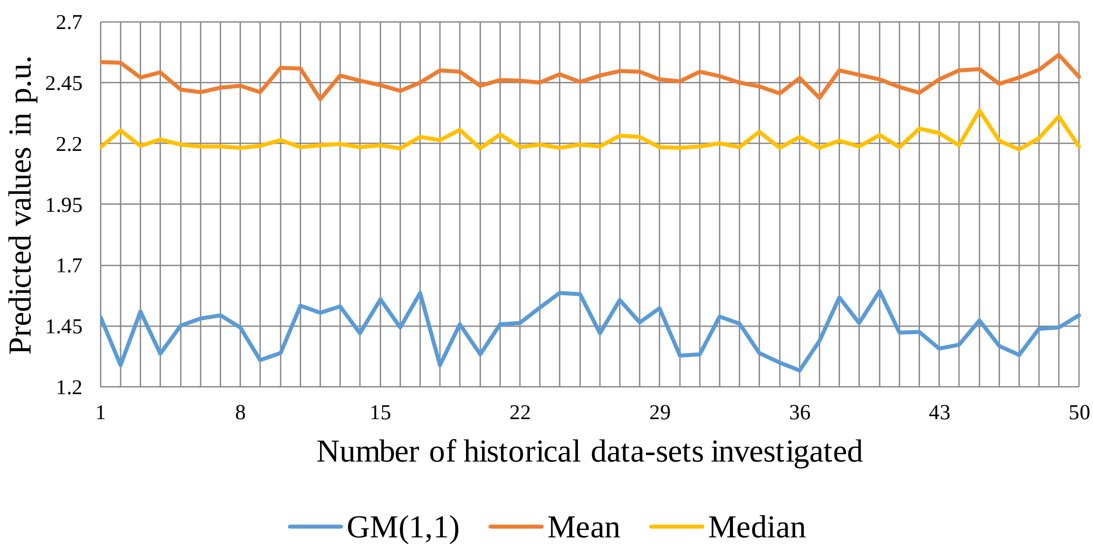

Fig. 6. Predicted values of Q_1 in the three different methods.

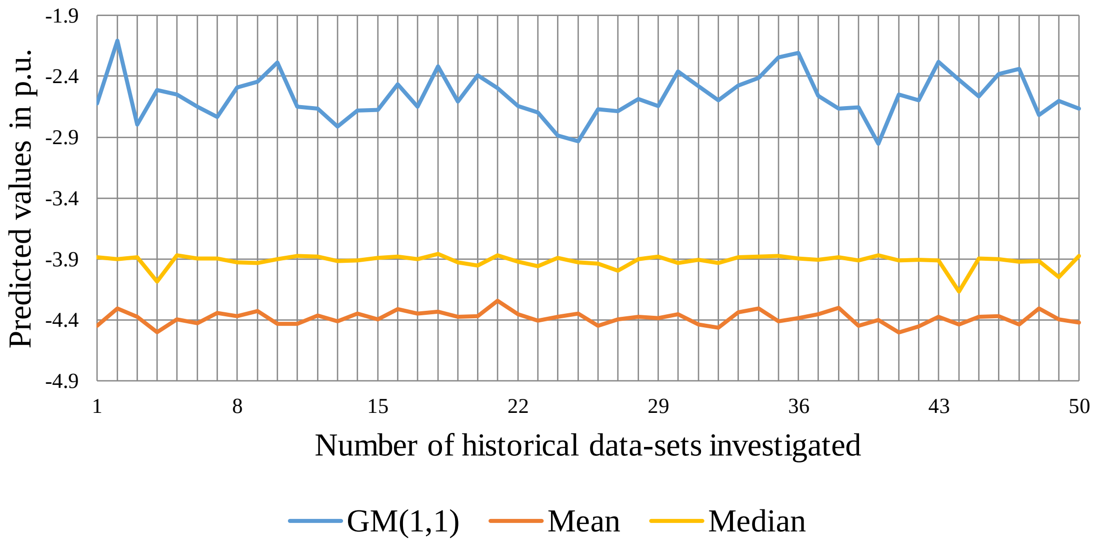

Fig. 7. Predicted values of Q_2 in the three different methods.

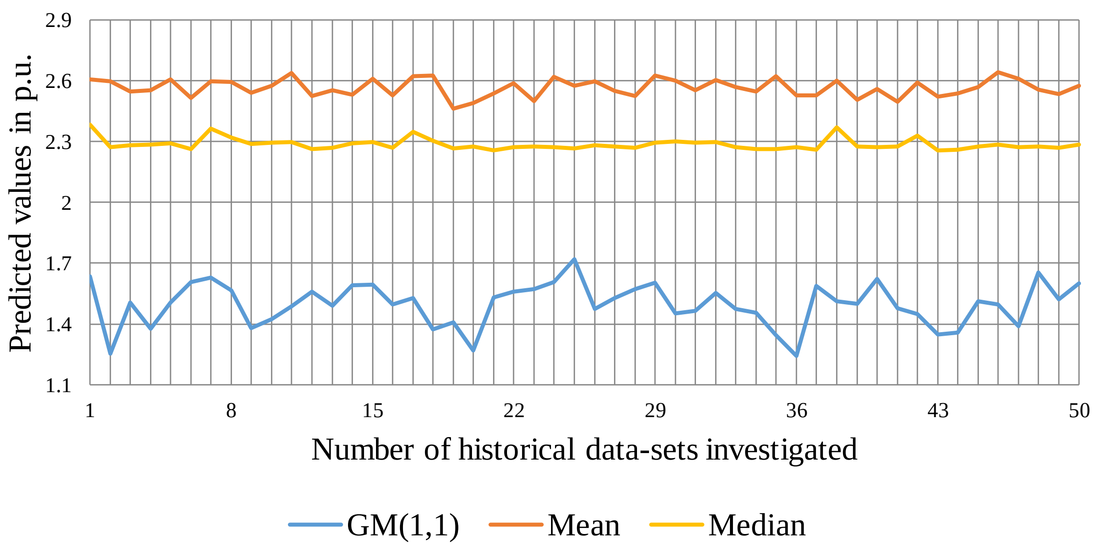

Fig. 8. Predicted values of Q_3 in the three different methods.

The preceding graphs clearly show the following facts:

- The predicted values of P_1, P_2, P_3, Q_1, Q_2 and Q_3 are satisfactorily close to their corresponding true values when GM(1,1) model is applied for the predictions.

- Mean values of the historical data are much away from the true values because of the presence of outliers in data-set.

- Predictions offered by median values are better than those by mean values in each case; still they are much inferior to the results obtained by GM(1,1) as the number of outliers is half of the number of data present in each historical-data-set. From the concepts of median, it may be commented that predictions done by median values would have been even worse if there were more outliers in the historical data.

Therefore, it can be inferred that superior predictions of P_1, P_2, P_3, Q_1, Q_2 and Q_3, are made possible by GM(1,1); this, in turn eases the task of state estimation by providing well-preprocessed raw materials.

The predicted values of P_1, P_2, P_3, Q_1, Q_2 and Q_3 are to be passed through the weighted least squares state estimation considering $\sigma_i^2 = 1$ for all measurements since all the errors, which were artificially introduced to construct surrogate data sets, lie within ±10% of their corresponding true values in cases of normal measurements. The states to be estimated here are δ_2, δ_3 and V_2 and the Jacobian matrix in this case, as determined by (15), is as follows:

$$\boldsymbol{H} = \begin{bmatrix} \frac{\partial P_1}{\partial \delta_2} & \frac{\partial P_1}{\partial \delta_3} & \frac{\partial P_1}{\partial V_2} \\ \frac{\partial P_2}{\partial \delta_2} & \frac{\partial P_2}{\partial \delta_3} & \frac{\partial P_2}{\partial V_2} \\ \frac{\partial P_3}{\partial \delta_2} & \frac{\partial P_3}{\partial \delta_3} & \frac{\partial P_3}{\partial V_2} \\ \frac{\partial Q_1}{\partial \delta_2} & \frac{\partial Q_1}{\partial \delta_3} & \frac{\partial Q_1}{\partial V_2} \\ \frac{\partial Q_2}{\partial \delta_2} & \frac{\partial Q_2}{\partial \delta_3} & \frac{\partial Q_2}{\partial V_2} \\ \frac{\partial Q_3}{\partial \delta_2} & \frac{\partial Q_3}{\partial \delta_3} & \frac{\partial Q_3}{\partial V_2} \end{bmatrix} \qquad (15)$$

In this study, outliers are assumed to be caused only due to abnormalities in the power system and not by errors in measurement. So, in other words, the measurement errors for all data including those at the positions of outliers in each data-set would have been ±10% of their corresponding true values if there were no anomalies in the operation of the power system. Since all the measuring devices here have same range of dispersion (i.e. ±10% error for each) in their measurements, hence they possess equal meter accuracy and therefore they may be considered to have equal variances (σ_i^2). Now, in the calculations explained in the preceding sections, the role of σ_i^2 is limited only to constructing the ***R*** matrix, and in this matrix, all elements of the principal diagonal become equal since all σ_i^2 values are equal. Hence, here ***R*** matrix can be expressed as an Identity Matrix (of the same order) multiplied by a constant coefficient (σ_i^2 in this case). In other words, we are taking out σ_i^2 ‘common’ from ***R*** matrix. Since this constant term remains same for all calculations and their all iterations, its value does not put any impact on inferences of these calculations as only the relative values (for example, relative percentage errors in the three different methods) are of interest in this context and not the absolute values. Therefore, for simplicity in the calculations, ***R*** matrix can be considered as an Identity Matrix of the same order putting the variance term as unity i.e. $\sigma_i^2 = 1$ for all i, is a fair supposition. Since inverse of an Identity Matrix is the matrix itself, hence due to this supposition, calculation of $\boldsymbol{R}^{-1}$ becomes much easier and so is the calculation of $\boldsymbol{\Delta X_{est}}^{(p)}$.

When the unknown states are estimated, they clearly reveal that the performance of GM(1,1), as expressed in terms of percentage errors in estimated final values of δ_2, δ_3 and V_2 and as depicted in Fig. 9, Fig. 10 and Fig. 11, surpasses the other two methods (i.e. mean and median) by substantial margin. Percentage error, in this context, has been determined by dividing the magnitude of difference between estimated and true values, by the magnitude of true value and thereafter multiplying the result with 100. Fig. 9, Fig. 10 and Fig. 11 show that, in case of GM(1,1), percentage errors remain fairly under 10%, 20% and 1% for δ_2, δ_3 and V_2 respectively in almost all cases whereas for mean and median, they are much higher.

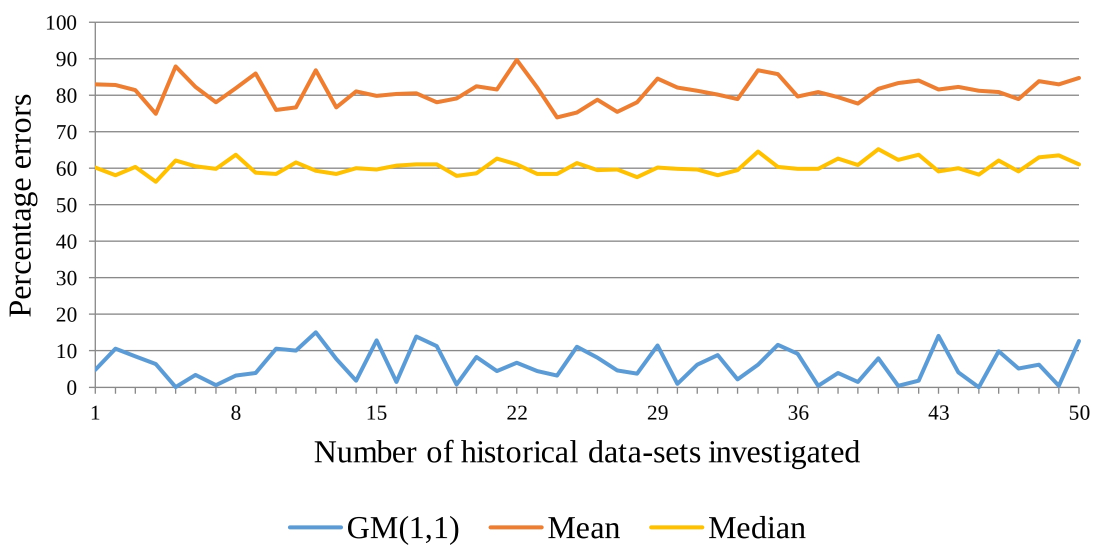

Fig. 9. Percentage errors in estimating δ_2 in the three different methods.

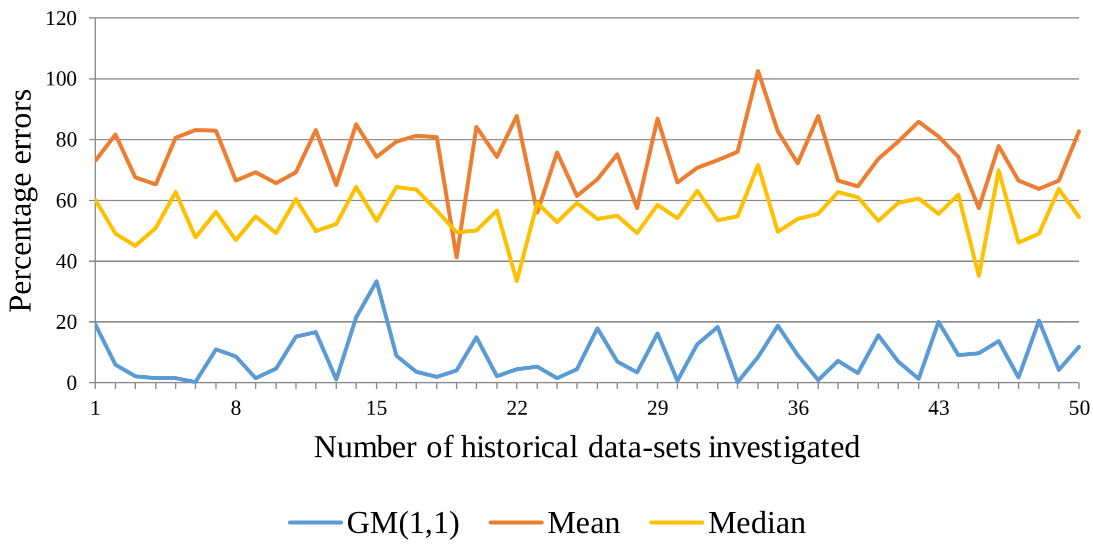

Fig. 10. Percentage errors in estimating δ_3 in the three different methods.

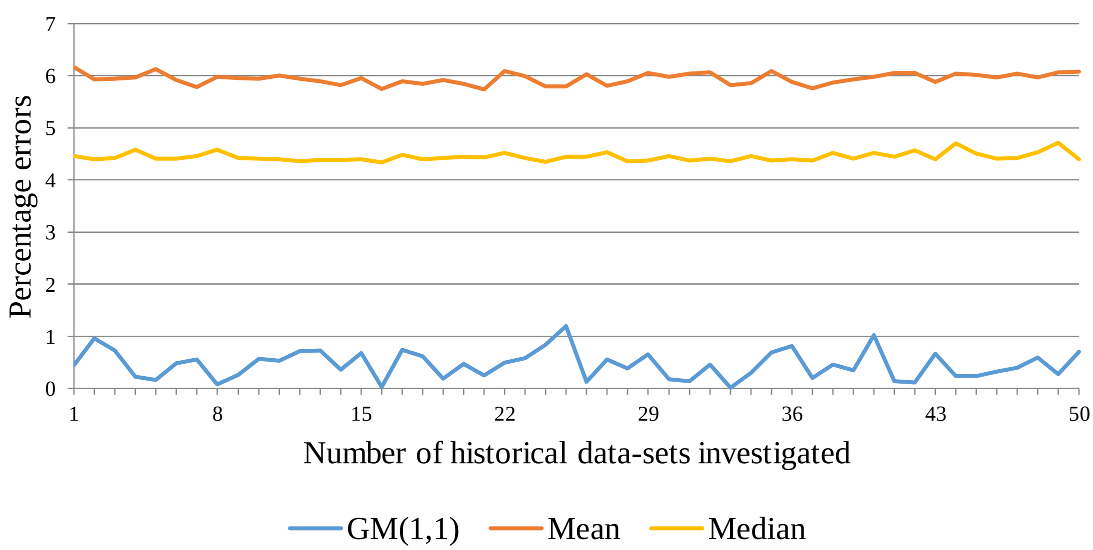

Fig. 11. Percentage errors in estimating V_2 in the three different methods.

The power system under study has been presumed to operate without any disturbances and contingencies at the instant of recording the 51st measurement (which, due to equipment malfunction, was recorded as a void or missing datum). Therefore, the load flow results achieved at the inception of the case-analysis are modeled as the true values at the 51st instant of its operation for calculating errors. All calculations have been performed in MATLAB software.

CHAPTER 4

CONCLUSION

4.1 Epilogue

The presented case analysis clearly portrays the performance-supremacy of GM(1,1) as far as prediction from outlier-infested historical data is concerned. It is able to produce good predictions which, in return, give a good SE of the power system under study. As already demonstrated in Fig. 9, Fig. 10 and Fig. 11, percentage errors in estimating δ_2, δ_3 and V_2 are significantly smaller for GM(1,1) than the other two methods. Average percentage error, for each method, is denoted by dint of the following expression:

$$e_{avg} = \frac{1}{n} \sum_{k=0}^{n} \left| \frac{E_k - T_k}{T_k} \right| \times 100 \tag{16}$$

where e_{avg} is average percentage error in estimated states, n is number of investigated data-sets of P_1, P_2, P_3, Q_1, Q_2 and Q_3 collectively, E_k is the estimated value with respect to k^{th} data-sets and T_k is the corresponding true value. By virtue of amalgamation of Fig. 9, Fig. 10 and Fig. 11, the average percentage errors in estimated states i.e. δ_2 (State 1, say), δ_3 (State 2, say) and V_2 (State 3, say), as calculated by (16) putting $n = 50$, are realized in Fig. 12.

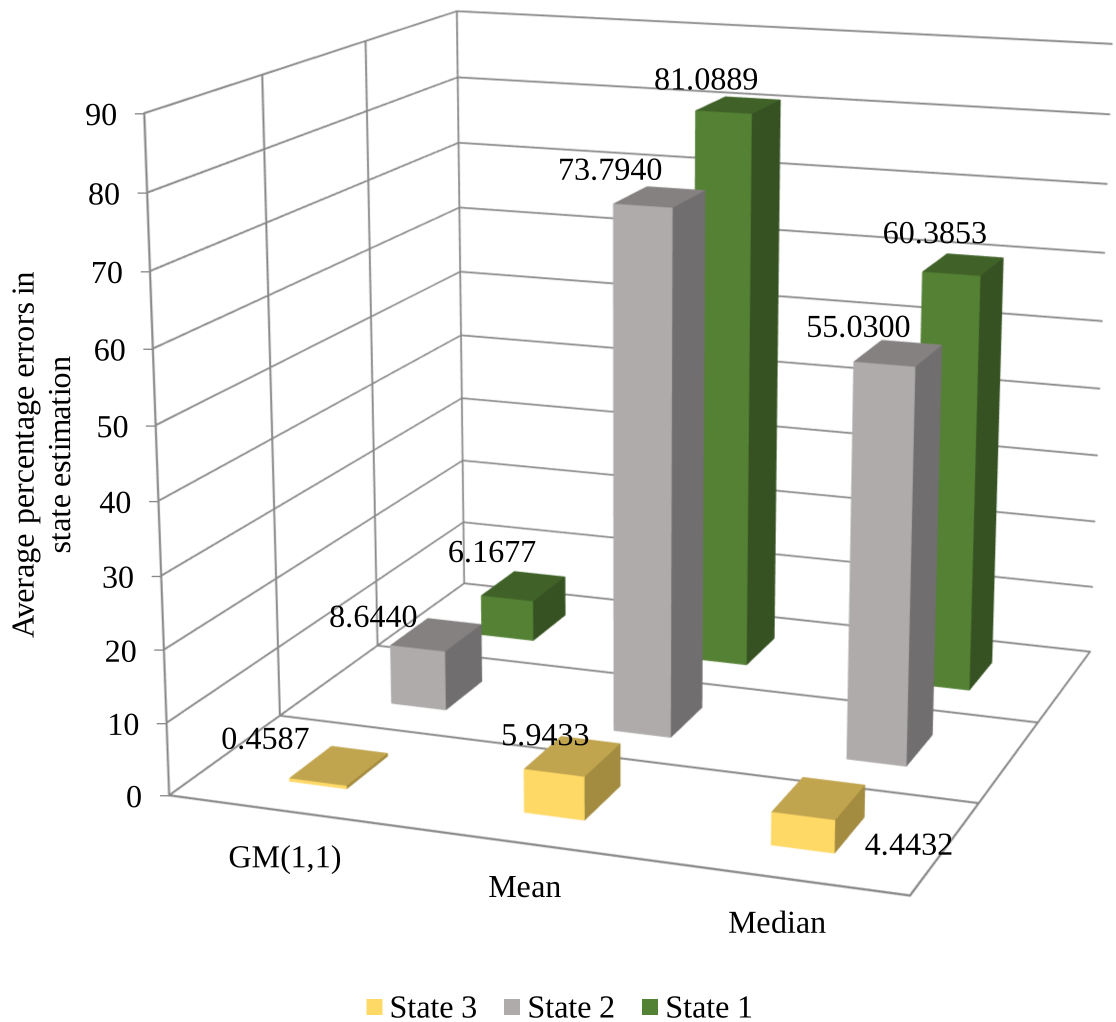

Fig. 12. Average percentage errors in estimating δ_2 (State 1), δ_3 (State 2) and V_2 (State 3) in the three different methods.

As seen in Fig. 12, maximum average percentage error occurring in case of GM(1,1) out of all the three states, lies even under 10%, whereas those for median and mean lie above 60% and 80% respectively. This deduces the dominance of GM(1,1) over the other two schemes to maneuver outliers.

4.2 Future Scopes

As elaborated in the previous pages, this thesis work revolves around testing the capability of GM(1,1) to handle input power-data that contain outliers. It is no more surprising here that, in what order or sequence the entities of input data lie, is of paramount importance and so is the gravity of outliers' positions.

In the study that has been put forth here, there are 25 numbers of outliers out of 50 numbers of input data. In 50 test runs, those 25 outliers are placed at random positions within 1 to 50 to achieve outcomes as much unbiased as possible. But it is a mathematical fact that 25 outliers can populate 50 spaces in 50! / (50-25)! ways which is a magnificent number. Therefore, many possibilities of permutations of outliers'-positioning are left untouched in this literature which may be looked into in the future.

Apart from that, the way GM(1,1) has behaved in this investigation may alter if there are changes in deliberately introduced error-percentages. In this study, the error percentages are strictly within –10% to 10% for all valid measurements (i.e. excluding the outliers); this situation may definitely alter if we alter the range of errors (say –15% to 15% or any other feasible range) and/or deploy different error ranges for different measurements/measuring devices; for example, measuring device 1 may have ±15% error whereas measuring device 2 may have ±9% error and so on. This scenario will challenge GM(1,1) with more difficult prediction-tasks by imposing variances other than $\sigma_i^2 = 1$. This is also a broad area that may be investigated in upcoming days.

Differences may also occur if state estimation techniques other than weighted least squares are used. Such variations may be inspected in future endeavors to explore new avenues. Besides all these, research may be done to modify the steps of GM(1,1) to enhance their effect specifically with respect to outlier-infested data.

It is also noteworthy that the prediction model can be tested with different types of outliers; in this study, outliers which are highly larger only (two to three times of the corresponding normal values in this case), are considered. Outliers, which are very much smaller than normal values or a combination of smaller and larger outliers may also be put under the microscope of this trial of prediction in relation with state estimation of power system.

Last but not the least, this study made use of a very simple three bus power system since it was uncertain in the beginning whether the test would be fruitful and produce desired results. Now, once we have achieved some favorable outcomes, the research may be optimistically extended towards more complex power systems.

BIBLIOGRAPHY

[1] A. Chakrabarti and S. Halder, Power System Analysis: Operation and Control, 3rd Edition. PHI Learning Pvt. Ltd., 2010.

[2] B. Khorramdel, C. Y. Chung, N. Safari and G. C. D. Price, "A Fuzzy Adaptive Probabilistic Wind Power Prediction Framework Using Diffusion Kernel Density Estimators," in IEEE Transactions on Power Systems, vol. 33, no. 6, pp. 7109-7121, Nov. 2018.

[3] Y. Wu, Zhongmei Pan, Xuan Luo, Jian Gao and Yuhan Zhang, "A hybrid forecasting method of electricity consumption based on trend extrapolation theory and LSSVM," 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Xi'an, 2016.

[4] X. Zheng, X. Qi, H. Liu, X. Liu and Y. Li, "Deep Neural Network for Short-Term Offshore Wind Power Forecasting," 2018 OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO), Kobe, Japan, 2018.

[5] T. Panapongpakorn and D. Banjerdpongchai, "Short-Term Load Forecast for Energy Management Systems Using Time Series Analysis and Neural Network Method with Average True Range," 2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP), Bangkok, Thailand, 2019.

[6] F. Gao, "Application of Improved Grey Theory Prediction Model in Medium-Term Load Forecasting of Distribution Network," 2019 Seventh International Conference on Advanced Cloud and Big Data (CBD), Suzhou, China, 2019

[7] L. Wu, S. You, J. Dong, Y. Liu and T. Bilke, "Multiple Linear Regression Based Disturbance Magnitude Estimations for Bulk Power Systems," 2018 IEEE Power & Energy Society General Meeting (PESGM), Portland, OR, USA, 2018, pp. 1-5.

[8] P. Samajdar and S. Halder nee Dey, "Investigating the Effect of Outliers in Historical-Data on Grey Prediction Model for Power System State Estimation," 2023 3rd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET), Patna, India, 2023.

[9] J. Yang, "Electricity Price Forecast Based on Metabolic GM(1,1) Model under Wind Power Development Prospect," 2022 IEEE 9th International Conference on Power Electronics Systems and Applications (PESA), Hong Kong, Hong Kong, 2022

[10] J. Kluabwang, S. Kothale and S. Yukhalang, "Using Basic Grey Prediction Model to Forecast Electricity Consumption of ASEAN," 2019 International Conference on Power, Energy and Innovations (ICPEI), Pattaya, Thailand, 2019, pp. 82-85.

[11] J. Li, S. Feng, T. Zhang, L. Ma, X. Shi and X. Zhou, "Study of Long-Term Energy Storage System Capacity Configuration Based on Improved Grey Forecasting Model," in IEEE Access, vol. 11, pp. 34977-34989, 2023.

[12] Z. Han, C. Zhou, S. Gan and X. Li, "Prediction of Original Reliability Parameters in Power Systems Based on Improved Discrete Grey Model," 2023 6th Asia Conference on Energy and Electrical Engineering (ACEEE), Chengdu, China, 2023, pp. 205-210.

[13] Y. Zhang, H. Sun and Y. Guo, "Wind Power Prediction Based on PSO-SVR and Grey Combination Model," in IEEE Access, vol. 7, pp. 136254-136267, 2019.

[14] T. H. M. El-Fouly, E. F. El-Saadany and M. M. A. Salama, "Grey predictor for wind energy conversion systems output power prediction," in IEEE Transactions on Power Systems, vol. 21, no. 3, pp. 1450-1452, Aug. 2006.

[15] L. Cai, J. Lin and X. Liao, "Life Prediction of Ship CXF Cable Using a Non-Equidistant Grey Model With Small Samples," in IEEE Transactions on Power Delivery, vol. 37, no. 6, pp. 5094-5101, Dec. 2022.

[16] K. Li, X. Zhang, X. Liu, G. He, J. Lu and M. Zhong, "Research on Power System State Estimation Technology Considering Data Filling Technology," 2020 IEEE Sustainable Power and Energy Conference (iSPEC), Chengdu, China, 2020, pp. 2570-2576.

[17] Nai-ming Xie, Si-feng Liu, Discrete grey forecasting model and its optimization, Applied Mathematical Modelling, Volume 33, 2009, pp. 1173-1186.

[18] C. Hernández and P. Maya-Ortiz, "Comparison between WLS and Kalman Filter method for power system static state estimation," 2015 International Symposium on Smart Electric Distribution Systems and Technologies (EDST), Vienna, Austria, 2015, pp. 47-52.

[19] Allen J. Wood, Bruce F. Wollenberg, Gerald B. Sheblé, Power Generation, Operation, and Control, 3rd Edition. John Wiley & Sons, Inc., 2014. pp. 403-435.

[20] Hadi Saadat, Power System Analysis, 3rd Edition. PSA Publishing.,2010. p. 256.