A Stochastic Hybrid Method to Forecast Operating Reserve: Comparison of Fuzzy and Classical Set Theory

ARASH ASRARI,1 AMIN KARGARIAN,2 MOHAMMAD HOSSEIN JAVIDI,3 MOHAMMAD MONFARED,3 and SAEED LOTFIFARD1

1Department of Electrical Engineering and Computer Science, University of Central Florida, Orlando, Florida, USA
2Department of Electrical and Computer Engineering, Mississippi State University, Starkville, Mississippi, USA
3Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

Abstract Accurate operating reserve forecasting helps the system operator to make decisions contributing to the security of the power system. It also helps market participants to adopt proper strategic bidding for the day-ahead ancillary services market to enhance their financial profit. This article proposes a stochastic hybrid method to forecast the operating reserve requirement in day-ahead electricity markets. At the first stage, based on using a modified Gray model, the day-ahead operating reserve is forecasted. In order to improve the accuracy of the operating reserve forecasting, at the next stage, a Markov chain model is used to predict the forecasting error of the Gray model. These two models are linked to each other using two different approaches—classical and fuzzy. The proposed approach is verified by the historical data of the operating reserve for spring and autumn seasons in the Khorasan Electricity Network located in Khorasan Province, Iran.

Keywords operating reserve, Gray model, Markov chain model, fuzzy approach

1. Introduction

In electricity markets, ancillary services have an essential role in secure and reliable operation of the system [1]. In most restructured power systems, one basic responsibility of the independent system operator (ISO) is to provide ancillary services through commercial contracts with market participants [1, 2]. Among six different ancillary services defined in FERC Order 888, the operating reserve (OR) is one important service. The OR consists of a spinning reserve (SR) as well as a non-SR. The SR refers to the extra generating capacity of the generators already connected to the power system, and the...
non-SR is the supplementary generating capacity not currently connected to the system and can be brought on-line after a short period of time [3].

Accurate short-term prediction of the OR is indispensable for the ISO to make appropriate and timely decisions leading to effective and economical management of the power system. Many ISOs calculate the day-ahead OR requirement through deterministic methods. For example, in the California electricity market, the OR is considered to be 6.5–7.5% of the California ISO (CAISO) system load forecast [4]. In the Khorasan Electricity Network, located in Khorasan Province, Iran, the OR is taken as 6% of the relevant ISO system load prediction [5]. Despite their simplicity, the deterministic methods may cause insecure and uneconomical power systems operation originating from sizeable underestimation or overestimation of the OR. Despite a variety of publications related to load forecasting [6–10] and electricity price forecasting [11–15], in competitive electricity markets, the prediction of the OR through complex and intelligent methods is in its initial steps, encouraging researchers to concentrate on this area more than the past [4, 16–18]. An adaptive wavelet neural network to predict the day-ahead SR of the California electricity market was presented in [16]. This method forecasted the SR of the next day with 3.71% mean absolute percentage error (MAPE) for a winter test week. In [3], the method presented in [16] was improved to predict the SR and OR of the California electricity market with 3.47% and 1.37% MAPEs, respectively.

Faria et al. presented an artificial neural network (ANN) methodology implemented with a multi-agent-based electricity market simulator to predict the day-ahead SR of California electricity market [17]. This method has predicted the SR of the next day with a 3.67% MAPE for a winter test week. An SR prediction strategy composed of mutual-information-based feature selection and a novel neural-network-based forecast engine was proposed in [18].

This article presents a stochastic hybrid method for forecasting the day-ahead OR in electricity markets. The proposed method includes the Gray model (GM) and Markov chain model linked together on the basis of the classical and fuzzy approaches. To evaluate the efficiency of this hybrid method, it is applied on Khorasan Electricity Network [5], which contains the Toos, Neyshaboor, Hybrid Cycle Shirvan, Mashhad, Shariati, Hybrid Cycle Shariati, Ghaen, Hybrid Cycle Kaveh, and Ferdowsi power plants. The main contributions of this article are

1. 24 GMs are assigned to 24 hours of a day to improve accuracy of forecasting.
2. the result of load forecasting is used as one of the inputs of the proposed method to decrease forecasting error.
3. classical and fuzzy approaches are employed to set a link between the GM and Markov chain model.
4. the strategy based on which membership vectors of the Markov chain model are computed to correct the error of GM forecasting.

2. The GM

The Gray system theory was introduced by Deng in 1989 [19], who called any random process a “gray” process if all the variables involved vary within certain amplitudes. The normal format of a GM is represented by \(GM(n, h)\), in which \(n\) and \(h\) signify the order of differential equations and number of variables, respectively. Usually, \(n\) is set to be 1 to simplify the process of GM construction. The procedure to build a \(GM(1, h)\) is described as follows [20].
a. Set up the original sequences:

\[
\begin{align*}
x^{(0)}_1 &= \left( x^{(0)}_1(1), x^{(0)}_1(2), \ldots, x^{(0)}_1(n) \right), \\
x^{(0)}_2 &= \left( x^{(0)}_2(1), x^{(0)}_2(2), \ldots, x^{(0)}_2(n) \right), \\
& \vdots \\
x^{(0)}_h &= \left( x^{(0)}_h(1), x^{(0)}_h(2), \ldots, x^{(0)}_h(n) \right),
\end{align*}
\]  

(1)

where \( x_1 \) represents the main sequence, and \( x_2 \) through \( x_h \) are reference sequences. Having a greater number of reference sequences increases the accuracy of the GM, although increasing the amount of \( h \) increases the model complexity.

b. Set up accumulated generating operation (AGO) sequences from the original ones. Due to the poor regularity of the original sequences, the AGO technique is used in the GM to decrease the uncertainty of raw data:

\[
\begin{align*}
x^{(1)}_1(k) &= \sum_{i=1}^{k} x^{(0)}_1(i) & (k = 1, 2, \ldots, n), \\
x^{(1)}_2(k) &= \sum_{i=1}^{k} x^{(0)}_2(i) & (k = 1, 2, \ldots, n), \\
& \vdots \\
x^{(1)}_h(k) &= \sum_{i=1}^{k} x^{(0)}_h(i) & (k = 1, 2, \ldots, n).
\end{align*}
\]  

(2)

c. Form GM(1, \( h \)):

\[
\frac{dx^{(1)}_1}{dk} + ax^{(1)}_1 = \sum_{i=2}^{h} b_i x^{(1)}_i(k),
\]  

(3)

where \( a \) is a developing coefficient, and \( b_2 \) to \( b_h \) are control variables.

d. Solve for \( a \) and \( b_2 \) to \( b_h \) by the least-square method (LSM):

\[
\begin{bmatrix}
a \\
b_2 \\
\vdots \\
b_h
\end{bmatrix} = (B^T B)^{-1} (B^T Y),
\]  

(4)
where

\[
B = \begin{bmatrix}
-0.5 \left( x_1^{(1)}(1) + x_1^{(1)}(2) \right) & x_2^{(1)}(2) & \cdots & x_h^{(1)}(2) \\
-0.5 \left( x_1^{(1)}(2) + x_1^{(1)}(3) \right) & x_2^{(1)}(3) & \cdots & x_h^{(1)}(3) \\
\vdots & \vdots & \ddots & \vdots \\
-0.5 \left( x_1^{(1)}(n-1) + x_1^{(1)}(n) \right) & x_2^{(1)}(n) & \cdots & x_h^{(1)}(n)
\end{bmatrix},
\]

(5)

\[
Y = \begin{bmatrix}
x_1^{(0)}(2) \\
x_1^{(0)}(3) \\
\vdots \\
x_1^{(0)}(n)
\end{bmatrix}
\]

\[
\hat{x}_1^{(1)}(k + 1) = e^{-a_k} \left[ x_1^{(0)}(1) + \int_0^k \left( \sum_{i=2}^h b_i x_i^{(1)}(t) \right) e^{a_t dt} \right]
\]

\[(k = 0, 1, \ldots, n-1, \ldots). \quad (6)\]

For instance, when using GM(1, 2), the solution of the GM equation is

\[
\hat{x}_i^{(1)}(k + 1) = \left( x_1^{(0)}(1) - \frac{b_2}{a} x_2^{(1)}(k) \right) e^{-a_k} + \frac{b_2}{a} x_2^{(1)}(k)
\]

\[(k = 0, 1, \ldots, n-1, \ldots). \quad (7)\]

f. Compute the inverse AGO (IAGO) to return the forecast data into their original environment:

\[
\hat{x}_1^{(0)}(k + 1) = \hat{x}_1^{(1)}(k + 1) - \hat{x}_1^{(1)}(k) \quad (k = 1, 2, \ldots, n - 1, \ldots), \quad (8)
\]

\[
\hat{x}_1^{(0)}(1) = x_1^{(0)}(1)
\]

where \(\hat{x}^{(0)}(k + 1)\) stands for the fitted data of training sequence provided that \(1 \leq k \leq n - 1\), and it represents the predicted future data if \(k \geq n\).

3. Building the GM

3.1. Proposed Strategy to Build GM(1, 2)

According to Eq. (6), GM(1, 1) has an exponential solution. In order to cover the fluctuation of the OR signal, GM(1, 2) has been selected to forecast the OR. In this
study, two test weeks in the spring and autumn of the year 2010 in Khorasan Electricity Network were chosen. For each test week, the OR data of 20 previous days are considered as the training data; this means that \(20 \times 24 = 480\) training samples are used to forecast the 24 OR samples of the next day. The training process is repeated for every day of each test week. Therefore, a new model is executed for each day of each test week. Although it sounds simpler to report the forecasting result of just the next day when attempting to predict the day-ahead OR, the reported results can hardly be taken as strong indicators of the productivity of the method, since the OR signal in the next day is likely to have exceptionally soft variations or, on the other hand, dramatically volatile fluctuations. So, seven days of a week are selected for day-ahead OR prediction. The proposed method is consecutively run seven times. This short training time is enough for effective extraction of the data trend [14] and for obtaining an accurate forecast. In order to evaluate the performance of the proposed method, the weekly MAPE (WMAPE), is used:

\[
WMAPE = \frac{100}{168} \sum_{i=1}^{168} \frac{|x(i) - y(i)|}{x(i)},
\]

where \(x\) and \(y\) signify the actual and predicted OR data, respectively. For spring, the OR data from 21 April to 10 May 2010 are used as training samples, and OR data from 11–17 May 2010 are considered as test samples [5]. In setting up the GM, the strategy adopted here builds 24 separate GMs, representing 24 hours of a day. In order to build up each of the 24 GM\((1,2)\)s, \(h = 2\) is considered and Eqs. (1) through (8) are followed to find \(a\) and \(b\) and solve the differential equation of each GM\((1,2)\). But the main question arising here is how the main and reference sequences are determined. The main sequence contains the OR data of the corresponding hour from 21 April to 10 May 2010 since the goal is to predict the OR samples of 11 May. Considering that GM\((1,2)\) is used in this phase, each of the 24 models requires only 1 reference sequence. A satisfactory reference sequence should have the highest correlation with the main one. So, the accuracy of the Gray forecasting model is enhanced. Figure 1 shows the correlation of each OR sample with 50 other OR data relating to the previous 50 hours. The largest correlation is between the OR data for each hour and the OR sample of the previous hour. Therefore, for setting up each GM\((1,2)\), in this study, the OR samples of the previous hour are considered as the reference sequence of the relevant model.

For instance, to develop the first GM\((1,2)\), the OR data from hour 00:00 of 22 April to 10 May 2010 are considered as the main sequence, and the OR data from hour 23:00 of 21 April to 9 May 2010 are considered as the reference sequence. Each of the 24 GMs forecasts only 1 single OR sample. By developing this strategy, the complexity of predicting OR data can be overcome.

Adopting this procedure, the OR of the other six days of the test week is predicted. Figure 2 shows the forecast results of the spring test week. The WMAPE of GM\((1,2)\) to predict the OR data from 11–17 May 2010 is 7.1557\%. Despite the fact that the 24 GM\((1,2)\)s accurately predict the OR peaks of some days, such as Wednesday and Saturday, the OR peaks of some other days, such as Friday and Monday, are not predicted correctly. Moreover, the prediction of the OR trend in many hours is not as convincing as that of the first OR peak attributed to Sunday or the third OR peak related to Monday. In the next section, utilization of GM\((1,3)\) is suggested for considering the
load data as one of the reference sequences to increase the accuracy of the forecasting model.

3.2. Utilization of GM(1, 3) to Consider Load Data

Increasing the number of reference sequences improves the accuracy of the GM while increasing the complexity of the forecasting method. Under such circumstances, the most influential decision relies on adopting the most relevant reference sequences. One of the reference sequences is the OR data of the previous hour. In this section, GM(1, 3) is suggested for use to forecast the OR. Hence, two reference sequences are required. Although one might be tempted to opt for the OR data of two hours ago as the second reference sequence, a higher correlation was found in several pilot trials between the current OR data and the current load sample than even the one between the current OR data and the OR sample of the previous hour. Therefore, load samples were adopted as the other reference sequence. Figure 3 presents the training OR data and OR samples from 21 April to 10 May, together with the load samples during this period. The trend of load data is very similar with that of OR samples.

When ANNs are employed for prediction, it is common to normalize the inputs of the network. Here, normalizing the inputs of GM(1, 3) is not permitted, since there is a non-linear exponential function in the solution of the GM. Consequently, the samples that are not introduced by megawatts, such as price, cannot be used as the reference sequence of GM when predicting the OR.
Now the question not yet dealt with is: which load samples should be adopted as the reference sequence—those related to previous hours or the current ones? OR samples have a close relationship with the load data corresponding to the same time. Since, in a real situation, there is no load data of the next hour, it is necessary to use the forecast load data as the second reference sequence of GM\(^{1,3}\). It is worth noting is that the accurate prediction of load data is easier than that of the OR, since the fluctuations of load data are considerably softer compared with that of OR samples. Figure 3 confirms this issue; therefore, this study adopted GM\(^{1,2}\) to forecast the load data of Khorasan Electricity Network from 11–17 May 2010 [21].

Having selected reference and main sequences, the 24 GMs must be built up as in the previous section. When the OR samples of the next day are predicted, each of the 24 GM\(^{1,3}\)s has the responsibility to predict just one OR sample corresponding to that particular hour. Figure 4 indicates the forecast OR data of the spring test week by GM\(^{1,3}\) together with the actual OR samples. The WMAPE of the GM\(^{1,3}\) predicting model is 5.6090%, which is smaller than that of GM\(^{1,2}\) (7.1557%). Adding just one appropriate reference sequence significantly improved the accuracy of the forecasting model. The performance of GM\(^{1,3}\) in predicting some sharp peaks, like the second peak on Friday, has been more accurate in comparison with that of GM\(^{1,2}\) (Figures 2 and 4). However, GM\(^{1,3}\)’s prediction of the last peak on Sunday is less accurate compared with GM\(^{1,2}\) performance.

The question arising here is that although the overall performance of GM\(^{1,3}\) in forecasting the OR samples is more accurate than that of GM\(^{1,2}\), why does the
Figure 3. Training OR and load samples. (color figure available online)

Figure 4. Forecasting results of GM(1,3) considering load data. (color figure available online)
prediction of some sharp peaks, such as the last one on Sunday, become less accurate when adding the load sequence to the predicting model and making use of GM(1, 3) rather than GM(1, 2). The answer to this question relies on two facts.

1. The predicted rather than actual load samples are used. Although the prediction of load data is more accurate than that of OR, there might be some hours in which the fluctuation of the load sample is sharper than that of OR.
2. There is no guarantee that the correlation of load and OR samples in all hours of study is high. In such hours, making use of the load sample as the reference data will lead to a lower accuracy in OR prediction.

In the next section, the setbacks arising from the use of a single forecasting model are solved through the introduction of the Markov chain model.

4. Markov Chain Model Developed by Classical and Fuzzy Approaches

In the previous section, it was demonstrated that the use of a single forecasting method like the GM is short of convincing accuracy in predicting the fluctuating trend of OR. The most important reason of this shortcoming originates from the anomalies arising from the uncertain nature of the OR and load data and the dependency of load data on the variations of temperature. The initial suggestion one can offer is the adaptation of the temperature as an additional reference sequence, thus developing GM(1, 4), with the OR data, the forecast load sample, and temperature as the first, second, and third reference sequences, respectively. Close scrutiny of this proposition renders it impossible, since temperature is not expressed in MW and it is not permissible to normalize the inputs of GM. Even if this were possible, prediction of temperature would still be accompanied with uncertainties. Moreover, this proposition can hardly result in a fundamental correction in the productivity of the forecasting method, for there would still be only one single predicting model. An alternative procedure offered here is the application of the Markov chain model as a stochastic method for correcting the forecasting results of GM(1, 3).

A Markov chain is a special case of a Markov process, which is, in its turn, a special case of a stochastic process. A random process $X_n$ is called a Markov chain if

$$P(X_{n+k} = q_{n+k} \mid X_n = q_n, \ldots, X_1 = q_1) = P(x_{n+k} = q_{n+k} \mid X_n = q_n),$$

where $q_1, q_2, \ldots, q_n, \ldots, q_{n+k}$ take discrete values. To make use of the Markov chain method to modify the output of the GM, it is essential to appropriately determine the strategy of defining membership vectors whose members correspond to the different states. It is also necessary to specify the nature of each state. This article suggests the relative errors between the actual training OR data and the GM’s fitted data as the samples determining the boundary of different states. The significant question arising here is whether such relative errors should be classified to different states using the classical theory or the fuzzy approach.

Table 1 shows the actual training OR samples corresponding to the first GM together with the fitted data predicted by GM(1, 3) for these training samples. The relative errors between these two kinds of values are the foundation of determining the classical and fuzzy membership vectors. In this study, only three states are selected to avoid more...
complexity of the method. The relative errors related to days 3 and 4 are considerably
greater than the others. The reason of such a conspicuous difference lies in the fact
that the GM requires some time to gradually adapt itself to the training samples. Such
adaptation depends on receiving feedback from reference sequences; that is why GM
cannot follow such fluctuating trends, because it is deprived of any reference sequence
and can only predict the signals with exponential behavior [22]. It should be mentioned
that the relative error related to day 2 equals 0 as a result of the basic assumption of
Eq. (8).

In this stage, the membership vectors must be determined, either through a classical
approach or a fuzzy approach. In order to apply the classical approach, it is found to be
more justifiable to exclude the exceptionally larger relative errors of days 3 and 4 to reduce
classification error. Then, the minimum–maximum interval of the remaining errors is
divided into three equal classes in such a way that each relative error absolutely belongs
to only one single state. The fifth column of Table 1 shows the classical membership
vectors of the relative errors.

In order to apply the fuzzy theory, the application of triangle membership functions
is suggested because of their simplicity and productivity. A general scheme of a triangle
fuzzy space is illustrated in Figure 5. There are odd numbers of fuzzy functions in a
triangle fuzzy space.
Excluding the relative errors of days 3 and 4, the remaining interval is divided into three equal periods and the fuzzy functions are determined. Equation (11) indicates the three triangle membership functions corresponding to hour 00:00:

\[
    u(k, 1) = \begin{cases} 
    1 & \varepsilon(k) \leq -1 \\
    -\frac{1}{2}(\varepsilon(k) - 1) & -1 \leq \varepsilon(k) \leq 1 \\
    0 & \text{otherwise} 
\end{cases},
\]

\[
    u(k, 2) = \begin{cases} 
    \frac{1}{2}(\varepsilon(k) + 1) & -1 \leq \varepsilon(k) \leq 1 \\
    -\frac{1}{2}(\varepsilon(k) - 3) & 1 \leq \varepsilon(k) \leq 3 \\
    0 & \text{otherwise} 
\end{cases},
\]

\[
    u(k, 3) = \begin{cases} 
    \frac{1}{2}(\varepsilon(k) - 1) & 1 \leq \varepsilon(k) \leq 3 \\
    1 & \varepsilon(k) \geq 3 \\
    0 & \text{otherwise} 
\end{cases},
\]

where \( u(k, m) \) stands for the membership degree of the \( k \)th relative error for each of the three classes, and \( \varepsilon(k) \) is the relative error corresponding to each training data. Applying the aforementioned fuzzy functions to the relative errors yields the fuzzy membership vectors, shown in the sixth column of Table 1. It should be mentioned that this procedure is repeated for the results of the other 23 GM(1, 3)s.

The next step to apply the Markov chain model to correct the prediction of the GM is building up the transition probability matrix. The matrix in Eq. (12) indicates
the probability that each state is transferred to another state for each interval:

$$P = \begin{bmatrix}
  p_{11} & p_{12} & \cdots & p_{1m} \\
  p_{21} & p_{22} & \cdots & p_{2m} \\
  \vdots & \vdots & \ddots & \vdots \\
  p_{m1} & p_{m2} & \cdots & p_{mm}
\end{bmatrix},$$  \hspace{1cm} (12)

where \( p_{ij} \) represents the probability based on which state \( i \) can be transferred to state \( j \) in one step. Such a probability is calculated as follows:

$$p_{ij} = \frac{m_{ij}}{M_i},$$  \hspace{1cm} (13)

where \( m_{ij} \) signifies the number of state \( i \) transferred into state \( j \) by one step, and \( M_i \) stands for the number of the appearances of state \( i \).

In this study, only one state is assigned to each relative error to build up the transition matrix. This action is straightforward with the classical approach. The only thing that needs to be done is to determine the class to which the error absolutely belongs. For instance, the state assigned to the relative error of day 10 would be class 2 (Table 1). In the case of the fuzzy approach, the more reasonable option is using the greatest membership principle presented in Eq. (14), which means the state whose probability of occurrence is the maximum should be selected:

$$U(x) = \bigvee_{k=1}^{m} \{u_{A_k}(\varepsilon(k))\}.$$  \hspace{1cm} (14)

In this equation, \( \varepsilon(k) \) relatively belongs to \( A_k \), where \( A_1, A_2, \ldots, A_m \) refer to \( m \) fuzzy sets on \( u \). So \( \varepsilon(k) \) refers to each relative error of GM, and \( U(.) \) stands for the function determining the state of \( x \); \( u(.) \) represents the fuzzy membership vector corresponding to \( \varepsilon(k) \).

For example, the state assigned to the relative error of day 10 is class 1 (Table 1). Following this process, the transition probability matrix can be calculated for the classical and the fuzzy approach corresponding to hour 00:00 as follows:

$$P_{\text{Classic}} = \begin{bmatrix}
  0.2 & 0.6 & 0.2 \\
  0.1 & 0.6 & 0.3 \\
  1 & 0 & 0
\end{bmatrix},$$  \hspace{1cm} (15)

$$P_{\text{Fuzzy}} = \begin{bmatrix}
  0.286 & 0.571 & 0.143 \\
  0.25 & 0.375 & 0.375 \\
  1 & 0 & 0
\end{bmatrix}.$$

This procedure is followed for the results of the other 23 GMs to obtain the relevant transition matrices. When the membership vectors and the transition probability matrix are obtained, the following formula can be used to predict the next state of relative error:

$$F(\varepsilon(n + 1)) = F(\varepsilon(n))P = [u_{A_1}(\varepsilon(n + 1)), \ldots, u_{A_m}(\varepsilon(n + 1))].$$  \hspace{1cm} (16)
Each component of $F(\varepsilon(n + 1))$ shows the membership degree of the next relative error to each fuzzy state at time step $n + 1$. To clarify the application of the equation, an example will be given. As can be observed from the last row of Table 1, the membership vector for both classical and fuzzy approaches is $(0, 0, 1)$. Obviously, these vectors can be different for the other 23 hours. Making use of the classical and fuzzy transition matrices related to hour 00:00, the predicted next state for day 21 for both the classical and fuzzy approaches by the Markov chain model is calculated as follows:

$$
F_{\text{Classic}}(\varepsilon(21)) = F_{\text{Classic}}(\varepsilon(20))P = (0, 0, 1)
\begin{bmatrix}
0.2 & 0.6 & 0.2 \\
0.1 & 0.6 & 0.3 \\
1 & 0 & 0
\end{bmatrix} = (1, 0, 0),
$$

$$
F_{\text{Fuzzy}}(\varepsilon(21)) = F_{\text{Fuzzy}}(\varepsilon(20))P = (0, 0, 1)
\begin{bmatrix}
0.286 & 0.571 & 0.143 \\
0.25 & 0.375 & 0.375 \\
1 & 0 & 0
\end{bmatrix} = (1, 0, 0).
$$

(17)

Although the transition matrices of the classical and fuzzy approaches are different, the predicted state of day 21 for both approaches is equal. It should be emphasized that such a coincidence may not occur for the other 23 Markov chain models corresponding to the other 23 hours, so the next membership vector can now be predicted. This time a relative error must be assigned to the forecast membership vector. From among the relevant strategies to accomplish this, it was found to be reasonable to use the weight sum method described as follows:

$$
\varepsilon(n + 1) = \frac{1}{2} \sum_{i=1}^{m} (f_i(\varepsilon(n + 1)) (\varepsilon_{i-1} + \varepsilon_i)),
$$

(18)

where $\varepsilon_{i-1}$ and $\varepsilon_i$ are, respectively, the minimum and maximum relative errors of training data in class $i$. $f_i(\varepsilon(n + 1))$ refers to the $i$th component of the predicted $F(\varepsilon(n + 1))$ by the Markov chain model. $\varepsilon(n + 1)$ stands for the relative error corresponding to the predicted membership vector. This predicted $\varepsilon(n + 1)$ is the forecast error between the predicted OR of hour 00:00 on day 21 by GM(1,3) and the actual relevant OR. That is why the proposed application of the Markov chain model has helped to correct the predictions made by GM(1,3).

Now, in the final step, the relative error predicted by the Markov chain model must be applied to the predicted OR by GM(1,3) to obtain a more accurate prediction of OR:

$$
\hat{y}(n + k) = \frac{\hat{x}(n + k)}{1 - \varepsilon(n + k)},
$$

(19)

where $\hat{x}(n + k)$ and $\hat{y}(n + k)$ signify the predicted values by GM(1,3) and GM(1,3)-fuzzy-Markov models, respectively.

This process should be followed for the other 23 GM(1,3)s to complete OR prediction for day 21. This procedure should then be repeated for the other six days of the test week. Figure 6 shows the forecast result of the GM(1,3)-classical-Markov model along with that of the GM(1,3)-fuzzy-Markov model for the spring test week. The WMAPE of the GM(1,3)-classical-Markov model and that of the GM(1,3)-fuzzy-Markov model
are, respectively, 5.4366% and 2.8081%. Considering that the WMAPE of GM(1, 3) is 5.6090%, the application of the classical approach to link the Gray and Markov chain models does not seem particularly productive. On the other hand, the conspicuous difference between the WMAPEs of the GM(1, 3) and GM(1, 3)-fuzzy-Markov model substantiates the influential role of the triangle fuzzy functions in correcting the prediction of the GM through the application of the Markov chain model.

It should be noted at this juncture that other more complicated fuzzy functions may bring about greater correction of forecasting results supplied by GM(1, 3). The exact procedure whereby these fuzzy functions are employed for reserve prediction could be a fruitful line of further research.

A challenging issue regarding the classification of relative errors is why the considerable relative error of day 16 cannot be ignored for the same reason the noticeable errors of days 3 and 4 were ignored. The answer lies in the different sources of these large errors: In determining the boundary of the classification, the errors of days 3 and 4 were excluded, since the GM was in its initial learning process and the large relative error could not be regarded as a reliable classifying criterion. But the large relative error of day 16 is not related to the bad training process of the GM, since the relative errors of the days before and after day 16 are convincingly small. The main reason of such a large relative error originates from the conspicuous jump in the SR between days 15 and 16.
Therefore, if the error of day 16 is excluded from the errors determining the boundary of the classification, the Markov chain model is practically deprived of learning such a noticeable jump.

There is something interesting about the capability of the Markov chain model: This model predicts uncertainties, such as the occurrence of sharp peaks, provided that such sharp fluctuations exist in the training data. To clarify this issue, an example is given. Considering the last two OR peaks of Figure 6 relating to hours 16:00 and 20:00 of Monday, it is observed that the GM-fuzzy-Markov’s prediction of the sharper peak (related to hour 20:00 of Monday) is more accurate than that related to hour 16:00 of Monday. But at first glance, it seems that the prediction of the sharper peak will be more difficult. The reason of this paradox is related to the existence of a history of comparable sharp variations during the 20-day training period for hour 20:00, while the history of training samples for hour 16:00 consists of less noticeable jumps. Figure 7 indicates the rate of growth of OR samples between each two consecutive training days for hours 16:00 and 20:00. There are quite conspicuous fluctuations in the history of the training data for hour 20:00 accompanied with the slight variations in this period. For example, notice the growth rate of 27.8% between days 4 and 5 and that of 16.5% between days 18 and 19 and, on the other hand, the growth percentage of 0.4% between days 13 and 14. This can be taken to mean that there are enough variations of training samples with different ranges in the 20-day training period relating to hour 20:00. In contrast, the history of fluctuations corresponding to hour 12:00 is limited to variations up to 5%, except those related to the interval of days 11 and 12 and that of 15 and 16. Lack of enough such sharp variations in the training samples corresponding to GM(1.3) for hour 16:00 brings about the weakness of the Markov chain model to accurately correct the GM prediction related to hour 16:00. As can be vividly perceived from Figure 6, other sharp peaks are

![Figure 7](image-url)

**Figure 7.** Growth rate of OR data between each two consecutive training days for hours 16:00 and 20:00.
Table 2  
WMAPE (%) in Khorasan Electricity Network for the two test weeks of study

<table>
<thead>
<tr>
<th>Electricity network</th>
<th>Test weeks</th>
<th>GM(1, 2) considering load data</th>
<th>GM(1, 3)-classical-Markov</th>
<th>GM(1, 3)-fuzzy-Markov</th>
<th>ARIMA</th>
<th>MLPNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Khorasan Spring</td>
<td>7.156</td>
<td>5.609</td>
<td>5.437</td>
<td>2.808</td>
<td>4.377</td>
<td>3.803</td>
</tr>
<tr>
<td>Autumn</td>
<td>8.539</td>
<td>6.950</td>
<td>6.004</td>
<td>3.781</td>
<td>5.791</td>
<td>5.180</td>
</tr>
<tr>
<td>Average</td>
<td>7.847</td>
<td>6.279</td>
<td>5.720</td>
<td>3.294</td>
<td>5.084</td>
<td>4.491</td>
</tr>
</tbody>
</table>

accurately predicted by GM(1, 3)-fuzzy-Markov except for two peaks, one of which is again related to hour 16:00 on Friday due to the same above-stated reason.

5. Comparison of the Proposed Method with Other Techniques

Table 2 shows the comparison of the suggested methods for spring and autumn test weeks. The training data of the autumn test week is related to OR samples from 22 October to 10 November 2010, and the test data are the OR samples from 11–17 November 2010 [5].

A multilayer perceptron neural network (MLPNN) and an autoregressive integrated moving average (ARIMA) model were simulated in [21] to compare their prediction with the efficiency of a Gray-fuzzy-Markov model. For building the MLPNN, a three-layer neural network was simulated, which has three inputs in the input layer and has eight and two neurons in the first and second hidden layers. A similar MLPNN was also used in [23] to show its efficiency in electricity price forecasting. This study again used these two popular models for comparison. As can be seen in Table 2, the proposed model forecasts the day-ahead OR more accurately in comparison to MLPNN and ARIMA models.

It should be noted that the training process of the proposed method for OR prediction in Khorasan Electricity Network takes 0.6 sec on a PC with Intel(R) core2 Duo CPU E7500 with 2.93 GHz and 2-GB RAM, which is the consequence of the non-iterative nature of this hybrid model.

6. Conclusion

In this article, a stochastic hybrid method is proposed for forecasting the OR requirement of the Khorasan Electricity Network. First described was how 24 GMs are assigned to the 24 hours of a day to analyze the data of each hour separately. Secondly, the procedure was explained whereby the forecast load is used as one of the inputs of the proposed method to increase the accuracy of OR prediction. The third stage described the way the fuzzy approach was applied to set a link between the Gray and Markov models. Finally, the way the membership vectors are calculated was portrayed, based on the relative errors of the Gray forecasting model, to build up the transition probability matrix of the Markov chain model.

It was found that the application of the fuzzy approach results in a substantial correction on GM forecasts (about 3% in the WMAPE), while the classical approach
could at best only marginally affect a correction (less than 1% in the WMAPE). It should be emphasized again that a fuzzy model or a Markov chain model was not used as an OR forecaster. The Markov chain model is only used here to forecast the next fuzzy state of the relative errors, and the fuzzy approach is merely used as a clustering method to set a link between the GM and the Markov chain model. Consequently, a comparison between OR forecasting capabilities of a fuzzy model or a Markov chain model as predictors and the proposed model is beyond the scope of this article and can be pursued as another useful line of further research.

Acknowledgments
The authors would like to thank Khorasan Regional Electricity Company (KREC), especially Mehdi Oloomi Buygi, for providing the operating reserve data of the Khorasan Electricity Network.

References
2. FERC, “Promotion of wholesale competition through open access non-discriminatory transmission services by public utilities and recovery of stranded costs by public utilities and transmitting utilities,” FERC Order No. 888, April 1996.


