Power Prediction Model of Photovoltaic Plants Based on Improved GM Model

Yuze Zhang

School of Business Administration, Hohai University, Changzhou 213022, China

prettypeble@163.com

Abstract. In order to improve the prediction accuracy of GM (1, 1) model of photovoltaic (PV) plants, IAFSA is proposed and used to optimize the background values and initial values directly, and an improved GM (1, 1) prediction model is constructed in combination with the rolling data updating mode. The example of PV plants in service shows that the above-mentioned model is effective and complete, and can improve the prediction accuracy of the GM (1, 1) model.

1. Introduction

In 1982, Professor Deng Julong proposed the grey system theory to effectively solve the uncertain system problem of few samples and incomplete information [1]. The grey prediction model is one of the important contents of the grey system theory, and its core is the grey accumulative generation. The most commonly used grey prediction model is the GM (1, 1) model, which has the advantages of less required raw data, simple modeling process, convenient calculation, and verifiable accuracy, and it has been widely used in many fields such as energy, water conservancy, and economy, etc [2]. However, GM (1, 1) model still has its limitations, and it has strong applicability to modeling problems where the data changes are more gradual, and when the data sequence grows too fast or presents large fluctuations, the model's prediction results are not ideal and the accuracy is poor [3]. The reason for these is that the modeling mechanism of the GM (1, 1) model mainly has the following defects: (1) it neglects the change of α, usually it takes 0.5 as the value, which results in inevitable system error; (2) it defaults that GM (1, 1) model takes the initial point as the basis for solution, which results in that the prediction results deviated from the optimal effect of the modeling [4]. In view of above situations, many scholars have proposed various methods to improve the prediction accuracy of the GM (1, 1) model. (Xiao et al., 2014) improves the accuracy of the GM (1, 1) model by improving the smoothness of the raw data sequence, but the power-exponential function transformation process is quite complicated. (Zhou and He, 2013) proposed a grey model based on particle swarm optimization algorithm, but only to optimize the grey value development of the system. (Chen and Huang, 2013) used automatic optimization and weight-determination to select background values, and uses the least square method to improve the initial values, although it can improve the model prediction accuracy, it significantly increases the complexity of the model. (Pai et al., 2011) used the first and last data in the data sequence as the initial condition of the linear combination, and combined the optimization idea to solve the weight coefficient, but it still did not avoid the defect that the GM (1, 1) model takes the initial point as the basis of the solution.

In conclusion, this paper first describes the GM (1, 1) modeling process in detail, analyzes the inherent defects in its modeling mechanism, determines the parameters to be optimized in the model; second, it uses IAFSA to optimize the key parameters and combines rolling data updating mode to construct an improved GM (1, 1) prediction model; finally, the above improved model is applied to predict the power of actual PV plants in the chemical industry parks to further verify its effectiveness.

2. GM Modeling Process

Grey prediction is based on a grey model, in which single-order first-order linear differential equation GM is most commonly used. Set the raw data sequence as $\mathbf{x}^{(0)} = [x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)]$, use 1-AGO (Accumulated Generating Operation) to generate a first-order cumulative generation sequence, where:

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) \quad (k = 1,2,\cdots,n) \tag{1}$$

From equation (1), we can see that the sequence $x^{(1)}(k)$ exhibits an exponential growth, which exactly meets the requirement of the first-order differential equation. It can be considered that the $\boldsymbol{x}^{(1)}$ sequence satisfies the following first-order linear differential equation model:

$$\frac{\mathrm{d}\,\mathbf{x}^{(1)}}{\mathrm{d}\,t} + a\mathbf{x}^{(1)} = u \tag{2}$$

If the values of the parameters a and u are known, by directly solving equation (2) we can get:

$$\mathbf{Y}_n = \mathbf{BA} \tag{3}$$

Where,

$$\mathbf{Y}_n = \begin{pmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} -\frac{1}{2}[x^{(1)}(1) + x^{(1)}(2)] & 1 \\ -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(3)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x^{(1)}(n-1) + x^{(1)}(n)] & 1 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a \\ u \end{pmatrix}.$$

In equation (3), the undetermined parameter is **A**, and the known quantities are $\mathbf{Y}_n$ and **B**. Since there are only two variables, a and u, while there are (n-1) equations and (n-1)>2, the equations have no solution. It can be approximated by the least squares fitting method and equation (3) can be rewritten as

$$\mathbf{Y}_n = \mathbf{B}\hat{\mathbf{A}} + \mathbf{E} \tag{4}$$

Where **E** is the error term.

To make $\min\left\|\mathbf{Y}_n - \mathbf{B}\hat{\mathbf{A}}\right\|^2 = \min\left(\mathbf{Y}_n - \mathbf{B}\hat{\mathbf{A}}\right)^T\left(\mathbf{Y}_n - \mathbf{B}\hat{\mathbf{A}}\right)$ established, according to the matrix derivation formula, we can further get:

$$\hat{\mathbf{A}} = \left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{Y}_n = \begin{pmatrix} \hat{a} \\ \hat{u} \end{pmatrix} \tag{5}$$

Substituting $\hat{a}$ and $\hat{u}$ obtained from equation (5) into the original differential equation, we can get:

$$\frac{\mathrm{d}\,\mathbf{x}^{(1)}}{\mathrm{d}\,t}+\hat{a}\mathbf{x}^{(1)}=\hat{u} \tag{6}$$

Solving equation (6), we can get:

$$x^{(1)}(t+1)=\left[x^{(1)}(1)-\frac{\hat{u}}{\hat{a}}\right]e^{-\hat{a}t}+\frac{\hat{u}}{\hat{a}} \tag{7}$$

Writing in *a* discrete form (because $x^{(1)}(1)=x^{(0)}(1)$), we can get:

$$x^{(1)}(k+1)=\left[x^{(0)}(1)-\frac{\hat{u}}{\hat{a}}\right]e^{-\hat{a}k}+\frac{\hat{u}}{\hat{a}} \qquad (k=0,1,2,\cdots) \tag{8}$$

Equations (7) and (8) are called GM-predicted time-response function models, which are further IAGO-reduced to get the GM prediction model of raw data sequence $\boldsymbol{x}^{(0)}$:

$$\begin{aligned}\hat{x}^{(0)}(k+1)&=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)\\&=\left(1-e^{\hat{a}}\right)\left[x^{(0)}(1)-\frac{\hat{u}}{\hat{a}}\right]e^{-\hat{a}k} \qquad (k=0,1,2,\cdots)\end{aligned} \tag{9}$$

Now the basic GM prediction model has been established, usually, relevant prediction criteria are used to verify the prediction results. If the accuracy of the original model fails to meet the standard, then the model needs to be further corrected or optimized. Under normal circumstances, the advantages and disadvantages of GM modeling are analyzed using the post-test difference test method. The specific steps are as follows:

Step 1. Calculating the residual error $e^{(0)}(k)$ and relative error $q(x)$ between $\boldsymbol{x}^{(0)}$ and $\hat{x}^{(0)}(k)$, respectively:

$$e^{(0)}(k)=\mathbf{x}^{(0)}-\hat{x}^{(0)}(k) \tag{10}$$

$$q(x)=\frac{e^{(0)}(k)}{x^{(0)}(k)} \tag{11}$$

Step2. Calculating the mean of $\boldsymbol{x}^{(0)}$ and its variance s_1;

Step3. Calculating the average value $\bar{q}$ of $e^{(0)}(k)$ and the variance s_2 of the residual error;

Step4. Calculating the variance ratio $C=s_2/s_1$;

Step5. Calculating the small probability error $P=P\{|e(k)|<0.6745s_1\}$;

Step6. Checking the table for accuracy testing, as shown in Table 1.

Table 1 Comparison table of GM accuracy testing

Grade	Relative error *q*	Variance ratio *C*	Small probability error *P*
Grade 1 (good)	< 0.01	$C \le 0.35$	$P \ge 0.95$
Grade 2 (qualified)	< 0.05	$0.35 < C \le 0.50$	$0.80 \le P < 0.95$
Grade 3 (barely)	< 0.10	$0.5 < C \le 0.65$	$0.70 \le P < 0.80$
Grade 4 (unqualified)	> 0.20	$C \ge 0.65$	$P < 0.70$

Two indicators of good extrapolation ability of the GM prediction model are: the posterior variance is smaller than *C* as much as possible, the small probability error *P* is as large as possible. Generally, in the model evaluation indicators, it requires that the variance s_1 of the observation data is large, and the variance s_2 of the prediction residual error is small, therefore the smaller the *C* value is, the better. In addition, if the relative deviation $|e(k)|/s_1$ is required to be less than 0.6745, then it allows the absolute value of the deviation $|e(k)|$ increase with it, therefore the larger the *P* value is, the better. Based on this, the two indicators, *C* and *P* are used in this paper to test the accuracy of the prediction model of the PV plants generated power.

3. GM defects and its optimization

At present, GM optimization mainly starts from the following two aspects:

3.1. GM background value optimization

From equation (5) we can know that, the parameters $\hat{a}$ and $\hat{u}$ in GM are closely related to the structure of the background value $z^{(1)}$. In the interval [*k*-1, *k*], derivation of $\frac{\mathrm{d}\mathbf{x}^{(1)}}{\mathrm{d}t}+\hat{a}\mathbf{x}^{(1)}=\hat{u}$ yields:

$$x^{(1)}(k)-x^{(1)}(k-1)+\hat{a}\int_{k-1}^{k}x^{(1)}(t)\,\mathrm{d}t=\hat{u} \tag{12}$$

$$z^{(1)}(k)=\int_{k-1}^{k}x^{(1)}(t)\,\mathrm{d}t \tag{13}$$

From the Lagrange mean value theorem, the general form of the background value can be constructed as:

$$z^{(1)}(k)=\alpha x^{(1)}(k-1)+(1-\alpha)x^{(1)}(k) \tag{14}$$

Where, $\alpha\in(0,1)$.

The following relationship exists between α and $\hat{a}$: $\alpha=\frac{1}{\hat{a}}-\frac{1}{e^{\hat{a}}-1}$. However, the traditional GM simply takes α=0.5 and ignores the change of α. When it takes α=0.5, it will cause prediction failure at the time $|\hat{a}|$ is large.

3.2. GM boundary value optimization

From Equation (3), Equation (5), Equation (8) and Equation (9), we can see that: (1) Precondition of GM is $\hat{x}^{(1)}(1)=x^{(0)}(1)$; (2) Since $x^{(0)}(1)$ does not participate in the construction of **B** and **Y**, this value is not related to the solution of the parameters $\hat{a}$ and $\hat{u}$ in the model, but this value affects the index correction effect of the GM prediction results. Set the boundary value correction formula to $\hat{x}^{(1)}(1)=x^{(0)}(1)+\theta$, where θ is the correction value of the boundary value. Meanwhile, considering that the GM prediction result is a result obtained by least-squares fitting, its value does not necessarily include the point $(1, x^{(0)}(1))$. Therefore, if the system boundary value is forced to be selected as $x^{(0)}(1)$, it means that restricting above fitting curve to pass the point $(1, x^{(0)}(1))$ is lack of theoretical basis. From this, it can be seen that the parameters affecting GM prediction result of the PV plants power of chemical industrial parks are two, respectively are α and θ. In order to accurately estimate the values of above parameters, an appropriate chemical industrial park PV plants power prediction objective function must be established, and intelligent optimization algorithms should be used to solve the parameters. Therefore, this paper chooses the IAFSA to optimize and solve the parameters.

4. IAFSA-GM-based parameter optimization and rolling data updating

4.1. IAFSA optimized GM parameters

First, establish the IAFSA-optimized objective function of PV plants power prediction model, in actual error test, the minimum average relative error of the prediction result is also an important indicator, namely:

$$\min_{\alpha,\theta}\left[\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{x}^{(0)}(i)-x^{(0)}(i)}{x^{(0)}(i)}\right|\%\right] \tag{15}$$

Therefore, in this section, IAFSA is used to optimize the objective function shown in equation (15), obtain the parameters α and θ in GM, and predict the output power of the PV modules [9]. The optimal parameter values α and θ are solved and their values are substituted into the GM so that a new data prediction model can be established [10].

4.2. GM rolling data updating

If at current time $t=k$, GM modeling is performed on all data of past time before the current time point, then the model is a continuous time function. In principle, GM can extend from the initial value to any time in the future. For an intrinsic grey system, unknown uncertainties will continue to enter the system over time to affect the system. Therefore, the larger the prediction time scale and the larger the grey scale, the smaller the actual meaning of the GM prediction value. Based on this, this paper uses a rolling data updating mode, that is to predict the short-term power based on IAFSA-GM, using rolling mode to continuously update the model data set, so as to improve the accuracy of PV module power prediction as much as possible.

Assuming k is the current time point, use the IAFSA-GM short-term power prediction to predict the PV module output power at time k in the time interval [$k+N_cT$, $k+N_TT$] within *Tcycle* resolution period T, where $k+N_cT$ is the prediction start time; $k+N_TT$ is the prediction end time; *Tcycle* is the number of prediction cycles, and its value is $(N_T-N_c)T$. Based on this prediction data, only the prediction results in the interval [$k+N_cT$, $k+(N_c+1)T$] are extracted, i.e., the "target period", and the power output prediction of the PV module at the time of $k+N_cT$ is taken as the initial state of the next prediction period (namely the time period interval [$k+(N_c+1)T$, $k+(N_T+1)T$]), then using IAFSA-GM to re-predict the power; at the same time, the data value with the earliest time (i.e., the data at time k) is deleted in order not to add additional calculation amount, thereby keeping the dimension of the entire sequence unchanged. Using this rolling data updating method can continuously correct the prediction bias of the model and can finally predict the trend in future time period with characteristic data that can reflect the output power of PV modules.

5. Example analysis

In order to verify the effectiveness and practicability of the above-mentioned IAFSA-GM-based PV plants power prediction model, the GM method and the IAFSA-GM method were used for simulation prediction. The experimental data selected in this paper comes from a PV plant in the Huai'an Yanhua Industrial Park in Jiangsu Province (118.68 East longitude, 33.73 North latitude). Table 2 shows the actual operation data of the PV plant from 09:00 to 11:15 (the sampling interval is 15 minutes) from June 10 to June 11, 2014, and the above data are input to the GM and IAFSA-GM respectively to predict the PV power generation of the next time period.

Table 2 PV power generation data

Power generation data	**Sampling point**				
	1	2	3	4	5
Data 1 / MW	8.97	9.92	10.97	12.03	13.58
Data 2 / MW	17.25	17.93	18.34	18.90	20.71
Power generation data	**Sampling point**				
	6	7	8	9	10
Data 1 / MW	15.99	18.23	20.94	24.66	28.07
Data 2 / MW	23.45	22.18	25.65	27.02	28.18

Table 3 Comparison between GM and IAFSA-GM simulation results

Data selection	**Prediction model**	***a***	***b***	***α***	***θ***
Data 1	GM	-0.136	7.458	0.5	—
	IAFSA-GM	-0.130	7.137	0.171	5.077
Data 2	GM	-0.0624	15.702	0.5	—
	IAFSA-GM	-0.0621	15.660	0.459	3.282
Data selection	**Prediction model**	**Fitting error / %**	**Variance ratio *C***	**Small probability error *P***	**One-step prediction error / %**
Data 1	GM	2.1441	0.0652	1	1.6743
	IAFSA-GM	2.0720	0.0763	1	1.4852
Data 2	GM	2.8830	0.2212	1	1.7790
	IAFSA-GM	2.8020	0.2121	1	1.2081

When IAFSA is used to optimize the GM parameters, the initialization range of the parameters to be optimized is set by the traditional GM algorithm. The simulation results based on the above models are shown in Table 3. Solving the existing mechanism defects in GM modeling helps to improve the prediction accuracy of the model. At the same time, this paper uses IAFSA to perform optimization and iteration on the parameters of GM, and then use the minimum average relative error as the objective function to find the best background value and the correction value of the boundary value, so that the accuracy of the prediction value is higher and more in line with the actual change. In addition, from the above prediction results we can see that, the experimental process of this paper only collects operational data from 10 chemical industrial park PV plants, the one-step prediction error of GM and IAFSA-GM was controlled within 2%, which further proved that GM can effectively resolve the uncertain system problem with few samples and incomplete information.

6. Conclusions

In this paper, IAFSA is used to optimize the background values and boundary values of GM prediction model to obtain the optimal parameter values and reduce the influence of raw data volatility; this paper further adopts the rolling data mode to continuously update the data set, so as to improve the prediction accuracy of the model as much as possible. The simulation analysis and research are carried out based on the actual operation data of a PV plant in a chemical industrial park. The examples show that, compared with the traditional GM prediction model, the proposed method can effectively improve the prediction accuracy of the model.

7. References

[1] Zhao H., Zhao H., Guo S. (2016). Using GM (1, 1) Optimized by MFO with Rolling Mechanism to Forecast the Electricity Consumption of Inner Mongolia [J]. Applied Sciences,

6 (1): 1-18.
[2] Yuan C.Q. , Chen D. (2016). Effectiveness of the GM (1, 1) model on linear growth sequence and its application in global primary energy consumption prediction [J]. Kybernetes, 45 (9): 1472-1485.
[3] Yu Z., Yang C., Zhang Z., et al. (2015). Error correction method based on data transformational GM (1, 1) and application on tax forecasting[J]. Applied Soft Computing, 37 (C): 554-560.
[4] Liu J., Xiao X., Guo J., et al. (2014). Error and its upper bound estimation between the solutions of GM (1, 1) grey forecasting models [J]. Applied Mathematics & Computation, 246 (C): 648-660.
[5] Xiao X.P., Guo H., Mao S.H. (2014). The modeling mechanism, extension and optimization of grey GM (1, 1) model [J]. Applied Mathematical Modelling, 38 (5-6): 1896-1910.
[6] Zhou W., He J.M. (2013). Generalized GM (1, 1) model and its application in forecasting of fuel production[J]. Applied Mathematical Modelling, 37 (9): 6234-6243.
[7] Chen C.I., Huang S.J. (2013). The necessary and sufficient condition for GM (1, 1) grey prediction model[J]. Applied Mathematics & Computation, 219(11): 6152-6162.
[8] Pai T.Y., Ho C.L., Chen S.W., et al. (2011). Using Seven Types of GM (1, 1) Model to Forecast Hourly Particulate Matter Concentration in Banciao City of Taiwan [J]. Water Air & Soil Pollution, 217 (1-4): 25-33.
[9] Xian S., Zhang J., Xiao Y., et al. (2017). A novel fuzzy time series forecasting method based on the improved artificial fish swarm optimization algorithm [J]. Soft Computing, 2017 (10): 1-11.
[10] Cao J., Zhao X., Li Z., et al. (2017). Modified artificial fish school algorithm for free space optical communication with sensor-less adaptive optics system [J]. Journal of the Korean Physical Society, 71 (10): 636-646