Accepted Manuscript



Application of a novel nonlinear multivariate grey Bernoulli model to predict the tourist income of China

Xin Ma, Zhibin Liu, Yong Wang

PII: S0377-0427(18)30470-9
DOI: https://doi.org/10.1016/j.cam.2018.07.044
Reference: CAM 11830

To appear in: *Journal of Computational and Applied Mathematics*

Received date : 10 June 2017
Revised date : 21 July 2018

Please cite this article as: X. Ma, Z. Liu, Y. Wang, Application of a novel nonlinear multivariate grey Bernoulli model to predict the tourist income of China, *Journal of Computational and Applied Mathematics* (2018), https://doi.org/10.1016/j.cam.2018.07.044

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Available online at www.sciencedirect.com

ScienceDirect

J COMPUT APPL MATH 00 (2018) 1–13

COMPUT APPL MATHlogo

www.elsevier.com/locate/procedia

Application of a novel nonlinear multivariate grey Bernoulli model to predict the tourist income of China

Xin Ma[a,c,*], Zhibin Liu[b], Yong Wang[b,c,**]

[a]School of Science, Southwest University of Science and Technology, Mianyang, China
[b]School of Science, Southwest Petroleum University, Chengdu, China
[c]State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu, China

Abstract

The nonlinear grey Bernoulli model, which is often called the NGBM(1, 1), appeals considerable interest of research due to its effectiveness in time series forecasting. The success of the NGBM(1, 1) model proves that the nonlinear Bernoulli equation can be efficient to build grey prediction models. This paper is mainly aiming at building a more general nonlinear grey prediction model based on the Bernoulli equation and the modelling procedures of the GMC(1, n), which is called the nonlinear grey Bernoulli multivariate model, abbreviated as the NGBMC(1, n). The NGBMC(1, n) can be converted to the NGBM(1, 1) and the GMC(1, n) with different power parameter, thus it can be regarded as an extension of these models. Two cases studies of forecasting the domestic income and foreign currency earning have been carried out to validate the effectiveness of the NGBMC(1, n) model, comparing to the novel linear multivariate grey models, including the GMC(1, n), DGM(1, n) and RDGM(1, n), and the classical ARIMA model. The advantages of the NGBMC(1, n) model in nonlinear time series forecasting over the existing linear models have been shown in the numerical results.

© 2018 Published by Elsevier Ltd.

Keywords: Bernoulli equation, Nonlinear grey model, NGBMC(1, n) model, Tourist industry, Emerging industry

1. Introduction

The grey system theory was proposed by Deng in 1982 [1], which is based on the idea of "Grey Box" and aiming at solving uncertain problems with small samples. The prediction models play an key role in the grey system theory, which are often called the grey models (GM).

The basic GM(1, 1) model is the most important grey model, which is easy to use and efficient in time series prediction, and it also represents the main idea of the methodology of the grey models. Numerous grey prediction models have been developed based on the similar methodology of the GM(1, 1) in recent years. Such as the NGM model [2], FGM(1, 1) [3], TDPGM(1, 1) [4], INDGM [5], seasonal GM(1, 1) [6], grey polynomial model [7], etc. However, Deng has pointed out that the GM(1, 1) presents a formulation with "partially differential and partially difference". The "differential" means its whitening equation is a differential equation and its response function used for forecasting is the analytic solution of its whitening

*E-mail address:cauchy7203@gmail.com (X. Ma)
**E-mail address:ywangsc@qq.com (Y. Wang)

ACCEPTED MANUSCRIPT

equation; the "difference" means its basic equation is actually a difference equation derived from its whitening equation. Such inconsistency makes the GM(1, 1) inaccuracy in many cases, and many researchers have proposed a lot of methods to improve the GM(1, 1) model [8–11].

The GM(1, n) model is a more general model than the GM(1, 1). By taking the linear combination of the reliance sequences (or input series) as a constant, the GM(1, n) essentially shares the same modelling steps of the GM(1, 1) model. But such treatment has been proved to be wrong by Tien in 2012 [12]. Tien has presented that the analytic solution of the whitening equation of the GM(1, n) model has a convolution integral, and then he applied the trapezoid formula to discretize the convolution integral, and the case studies shown by Tien have indicated that this treatment is much better than the conventional GM(1, n) model. This method has been developed in the later researches, the developments of the DGMD(1, 1, 1) [13], DGDMC(1, n) [14], IGDMC(1, n) [15] and FGMC(1, n) [16] all share a similar formulation, which presents to have a convolution integral. With the successes of these grey models, the modelling procedures of the GMC(1, n) can be regarded as a novel methodology of the grey models in our opinion. And it should be mentioned that some novel methods to improve the GM(1, n) model or GMC(1, n) model based on the recursive method[17, 18], revised background values [19], nonlinear least squares method (NLS)[20], and new accumulation schemes [21, 22] have also been presented to be efficient to improve the grey models, but the solution of the improved models are essentially a discrete convolution. Based on the above descriptions, we may regard the grey models with convolution integral as a novel and efficient branch of the grey models.

However, the above grey models are all essentially linear models, and obviously they could be inefficient to deal with nonlinear problems. Deng has pointed out in this book [23] that the nonlinear grey models should be an important part of the grey system theory as the nonlinear phenomena widely exist in the real world. For instances, the nonlinear flow has long been a hot point in the geology science and petroleum engineering [24, 25] and the other energies [26, 27]. Some important indicators are often nonlinearly related in the water chemistry [28]. The times series of the wind speed data are often presented to be nonlinear thus only the nonlinear methods such as neural networks, support vector machines can be used for accurate prediction in these fields [29–31]. Even some times the nonlinear phenomena can not be described with a simple mathematical model, and expensive experiments have to be carried out [32, 33]. The grey Verhulst model could be taken as the initial research of the nonlinear grey model, which can be seen in the book of Liu, *et.al*. [34]. Actually, Deng has discussed that the grey power model is a general nonlinear univariate grey model, but he has not discussed the details on the modelling procedures. The whitening equation of the grey power model is essentially a Bernoulli equation, thus Chen *et.al*. [35] renamed this model as the nonlinear grey Bernoulli model NGBM(1, 1). The practice examples presented by Chen have illustrated the effectiveness of the NGBM(1, 1) model and its advantages over the traditional GM(1, 1) model. The NBGM(1, 1) and its improved models have been applied in many fields in recent years [36–44]. However, the Bernoulli equation has not been introduced to the multivariate grey models.

As mentioned above, the multivariate grey models are more general than the univariate models, the research of nonlinear multivariate models may be more important to the grey system theory. At present, there exist several multivariate nonlinear grey models, such as the NGMC(1, n) by Wang [45] and the kernel regularized nonlinear GMC(1, n) [46], *etc*. The novel methodologies developed from these models have also been used to build other nonlinear models. For instance, the NGM(1, n) [47] shares a similar form with the NGMC(1, n) model, and the kernel regularized formulation has been used to build the KRNGM model[48]. It is interesting to notice that although these models have different formulations, the main ideas of building such nonlinear models share a similar pattern: introduce a nonlinear formulation at first, and then translate it into a linear form. Then the existing methodologies of building the linear multivariate grey models can be used. And this pattern can also be found in some non-grey models, such as the kernel-based nonlinear extension of the Arps model in [49], *etc*.. This pattern will also be used in this paper.

The main purpose of this paper is to build a novel nonlinear multivariate grey model based on the Bernoulli equation, by translating the Bernoulli equation into a linear model, we will use the methodology of the GMC(1, n) model to derive its parameter estimation and solution. The novel model will be called the nonlinear grey Bernoulli multivariate model with convolution integral, abbreviated as NGBMC(1, n) in the rest of this paper.

The outline of this paper is organized as follows: A brief overview of the existing GMC(1, n) model is

ACCEPTED MANUSCRIPT

presented in Section 2; the modelling procedures of the novel NGBMC(1, n) model are shown in Section 3; two case studies of the tourist industry of China are carried out in Section 4, including the comparison of the effectiveness of the NGBMC(1, n) model and the existing linear grey models; discussions and conclusions are drawn in Section 5.

2. The existing GMC(1, n) model

In this section we give a very brief overview of the GMC(1, n) model. Within the assumption that observations $\left(X_1^{(0)}, X_2^{(0)}, ..., X_n^{(0)}\right)$ from a certain dynamic system with $n-1$ inputs $\left(X_2^{(0)}, ..., X_n^{(0)}\right)$ and an output $X_1^{(0)}$ are available at interval of time, Tien[50] suggests GMC(1, n) model to be the following linear differential equation

$$\frac{dX_1^{(1)}(rp+t)}{dt} + b_1 X_1^{(1)}(rp+t) = \sum_{i=2}^{n} b_i X_i^{(1)}(t) + u, \tag{1}$$

where

$$X_1^{(1)}(rp+t) = \sum_{k=rp+1}^{rp+t} X_1^{(0)}(k), t = 1, 2, ..., r + rf \tag{2}$$

and

$$X_i^{(1)}(t) = \sum_{k=1}^{t} X_i^{(0)}(k), t = 1, 2, ..., r + rf \tag{3}$$

are the 1-AGO series of $X_1^{(0)}$ and $X_i^{(0)}(i = 2, 3, ..., n)$, respectively; r is the number of entries to build the model, rp is a period of delay, rf is the number of entries to be forecasted, and $b_1, b_2, ..., b_n$ and u are model parameters to be estimated. The equation (1) is also called the whiten equation of the GMC(1, n) model. If the bias u is omitted and $rp = 0$, the GMC(1,n) model yields to the conventional GM(1,n) model; when $n = 1$ and $rp = 0$, the GMC(1,n) model (1) yields to the GM(1, 1) model[34].

The approximation of $\frac{dX_1^{(1)}(rp+t)}{dt}$ is taken as

$$\frac{dX_1^{(1)}(rp+t)}{dt} \approx \lim_{\Delta t \to 1} \frac{X_1^{(1)}(rp+t) - X_1^{(1)}(rp+t-\Delta t)}{\Delta t} = X_1^{(1)}(rp+t) - X_1^{(1)}(rp+t-1) = X_1^{(0)}(rp+t),$$

and the background values of $X_1^{(1)}(rp+t)$ and $X_i^{(1)}(rp+t)$ are defined as

$$Z_1^{(1)}(rp+t) = \frac{1}{2}[X_1^{(1)}(rp+t) + X_1^{(1)}(rp+t-1)] \tag{4}$$

and

$$Z_i^{(1)}(t) = \frac{1}{2}[X_i^{(1)}(t) + X_i^{(1)}(t-1)]. \tag{5}$$

And then the differential equation (1) can be approximately rewritten as the following difference equation

$$X_1^{(0)}(rp+t) + b_1 Z_1^{(1)}(rp+t) = \sum_{i=2}^{n} b_i Z_i^{(1)}(t) + u. \tag{6}$$

The least squares estimation of the parameters of GMC(1, n) by t from 1 to r is the least squares solution of Eq.(6), which can be written as

$$[b_1, b_2, ..., b_n, u]^T = (B^T B)^{-1} B^T Y_R, \tag{7}$$

where

$$B = \begin{bmatrix} -Z_1^{(1)}(rp+2) & Z_2^{(1)}(2) & ... & Z_n^{(1)}(2) & 1 \\ -Z_1^{(1)}(rp+3) & Z_2^{(1)}(3) & ... & Z_n^{(1)}(3) & 1 \\ \vdots & \vdots & & & \vdots \\ -Z_1^{(1)}(rp+r) & Z_2^{(1)}(r) & ... & Z_n^{(1)}(r) & 1 \end{bmatrix}, Y_R = \begin{bmatrix} X_1^{(0)}(2) \\ X_1^{(0)}(3) \\ \vdots \\ X_1^{(0)}(r) \end{bmatrix}.$$

ACCEPTED MANUSCRIPT

The response function for $\hat{X}_1^{(1)}(rp+t)$ can be obtained by solving the whiten equation (1) as

$$\hat{X}_1^{(1)}(rp+t) = X_1^{(0)}(rp+1)e^{-b_1(t-1)} + \int_1^t e^{-b_1(t-\tau)} f(\tau)dt, \tag{8}$$

where

$$f(t) = \sum_{i=2}^{n} b_i X_i^{(1)}(t) + u. \tag{9}$$

In recent researches, the the convolution integral of the right hand of Eq.(8) is often discretized using the Gaussian formula, and the 1-AGO series can also be evaluated by the following discrete function

$$\hat{X}_1^{(1)}(rp+t) = X_1^{(0)}(rp+1)e^{-b_1(t-1)} + \sum_{\tau=2}^{t} \left\{ e^{-b_1(t-\tau+\frac{1}{2})} \cdot \frac{1}{2}[f(\tau) + f(\tau-1)] \right\}. \tag{10}$$

The predicted value $\hat{X}_1^{(0)}(rp+t)$, which is also called the restored value in the conventional grey system theory, is given as

$$\hat{X}_1^{(0)}(rp+t) = \hat{X}_1^{(1)}(rp+t) - \hat{X}_1^{(1)}(rp+t-1). \tag{11}$$

As presented above, once given the sample data, the system parameters in Eq.(1) can be obtained. And then the output series can be predicted with the system parameters and input series using the discrete function (10) and (11).

3. The proposed NGBMC(1, n) model

Notably, the existing NGBMC(1, n) model is essentially a linear model, which limits its applicability. In this section we will introduce a novel nonlinear grey model based on the modelling method of the GMC(1, n).

3.1. Representation of the NGBMC(1, n) model and its linearization

With the same definitions of the 1-AGO and the given sample, the NGBMC(1, n) model is represented as the following nonlinear differential equation

$$\frac{dX_1^{(1)}(rp+t)}{dt} + b_1 X_1^{(1)}(rp+t) = \left(\sum_{i=2}^{n} b_i X_i^{(1)}(t) + u \right) \left(X_1^{(1)}(rp+t) \right)^{\gamma}, \tag{12}$$

where γ is a nonlinear parameter. Obviously, the NGBMC(1, n) yields to the GMC(1, n) model when $\gamma = 0$. When $n = 1$ and $rp = 0$, the NGBMC(1, n) model can be written as follows:

$$\frac{dX_1^{(1)}(t)}{dt} + b_1 X_1^{(1)}(t) = u \left(X_1^{(1)}(t) \right)^{\gamma}. \tag{13}$$

The Eq.(13) is the representation of the NGBM(1, 1) model introduced in the reference [35].

Actually, the NGBMC(1, n) model can be converted to a linear model when $\gamma \neq 0$ and $\gamma \neq 1$. Firstly we multiply the (12) by $\left(X_1^{(1)}(rp+t) \right)^{-\gamma}$, then we have

$$\left(X_1^{(1)}(rp+t) \right)^{-\gamma} \frac{dX_1^{(1)}(rp+t)}{dt} + b_1 \left(X_1^{(1)}(rp+t) \right)^{1-\gamma} = \sum_{i=2}^{n} b_i X_i^{(1)}(t) + u. \tag{14}$$

Set

$$y(rp+t)^{(1)} = \left(X_1^{(1)}(rp+t) \right)^{1-\gamma}, \tag{15}$$

ACCEPTED MANUSCRIPT

and notice that

$$\frac{dy^{(1)}(rp+t)}{dt} = (1-\gamma)\left(X_1^{(1)}(rp+t)\right)^{-\gamma}\frac{dX_1^{(1)}(rp+t)}{dt},$$

which is also

$$\frac{dX_1^{(1)}(rp+t)}{dt} = \frac{1}{1-\gamma}\left(X_1^{(1)}(rp+t)\right)^{\gamma}\frac{dy^{(1)}(rp+t)}{dt}. \tag{16}$$

By substituting (15) and (16) into (12), we have the linearized form of the NGBMC(1, n) model as follows

$$\frac{dy^{(1)}(rp+t)}{dt} + b_1(1-\gamma)y^{(1)}(rp+t) = (1-\gamma)\sum_{i=2}^{n} b_i X_i^{(1)}(t) + (1-\gamma)u. \tag{17}$$

The above equation (17) is called the linearization of the NGBMC(1, n) model. It can be clearly seen that the linearization (17) is mathematically equivalent to the GMC(1, n) model (1), thus we can easily obtain the parameters estimation and the solution of the NGBMC(1, n) using the corresponding results of the GMC(1, n).

3.2. Parameters estimation of the NGBMC(1, n) model

Obviously, an approximate linear discrete form of the NGBMC(1, n) model is needed to derive its parameters estimation. In the existing GMC(1, n) model such discrete form is obtained using the approximation of the $\frac{dX_1^{(1)}(rp+t)}{dt}$ and the background values of the $X_i^{(1)}(rp+t), i = 2, ..., n$. But in this paper we will use another process based on integral. Considering the integral of the linearization (17) in the interval $[t-1, t]$, we have

$$\int_{t-1}^{t} dy^{(1)}(rp+\tau) + b_1(1-\gamma)\int_{t-1}^{t} y^{(1)}(rp+\tau)d\tau = (1-\gamma)\sum_{i=2}^{n} b_i \int_{t-1}^{t} X_i^{(1)}(t)d\tau + (1-\gamma)\int_{t-1}^{t} ud\tau. \tag{18}$$

The integrals can be computed as follows with new notations

$$\begin{aligned} y^{(0)}(rp+t) &= \int_{t-1}^{t} dy^{(1)}(rp+\tau) \\ &= y^{(1)}(rp+t) - y^{(1)}(rp+t-1) \\ &= \left(X_1^{(1)}(rp+t)\right)^{1-\gamma} - \left(X_1^{(1)}(rp+t-1)\right)^{1-\gamma}, \end{aligned} \tag{19}$$

$$\begin{aligned} Z_y(rp+t) &= \int_{t-1}^{t} y(rp+\tau)d\tau \\ &= \frac{1}{2}\left(y^{(1)}(rp+t) - y^{(1)}(rp+t-1)\right) \\ &= \frac{1}{2}\left[\left(X_1^{(1)}(rp+t)\right)^{1-\gamma} + \left(X_1^{(1)}(rp+t-1)\right)^{1-\gamma}\right], \end{aligned} \tag{20}$$

and

$$\int_{t-1}^{t} ud\tau = u. \tag{21}$$

Substituting the (19)-(21) and (5) into (22), we can obtain the discrete form of the NGBMC(1, n) model concisely as

$$y^{(0)}(rp+t) + b_1(1-\gamma)Z_y(rp+t) = (1-\gamma)\sum_{i=2}^{n} b_i Z_i^{(1)} + (1-\gamma)u. \tag{22}$$

Once the parameter γ is given, the parameters estimation of the NGBMC(1, n) model can be easily obtained using a similar way of the GMC(1, n) model as follows

$$[b_1, b_2, ..., b_n, u]^T = (B^T B)^{-1} B^T Y_R \tag{23}$$

ACCEPTED MANUSCRIPT

where

$$B = (1-\gamma)\begin{bmatrix} -Z_y(rp+2) & Z_2^{(1)}(2) & \dots & Z_n^{(1)}(2) & 1 \\ -Z_y(rp+3) & Z_2^{(1)}(3) & \dots & Z_n^{(1)}(3) & 1 \\ \vdots & \vdots & & & \vdots \\ -Z_y(rp+r) & Z_2^{(1)}(r) & \dots & Z_n^{(1)}(r) & 1 \end{bmatrix}, Y_R = \begin{bmatrix} y^{(0)}(rp+2) \\ y^{(0)}(rp+3) \\ \vdots \\ y^{(0)}(rp+r) \end{bmatrix}.$$

3.3. Computing the the 1-AGO series $\hat{X}_1^{(1)}$ and the restored values $\hat{X}_1^{(0)}$

The solution of the whitening equation (12) can be indirectly obtained by solving its linearization (17). As the linearization (17) is mathematically equivalent to the GMC(1, n) model (1), we can easily obtain its solution by simply changing some notations as

$$\hat{y}^{(1)}(rp+t) = y^{(0)}(rp+1)e^{-b_1(1-\gamma)(t-1)} + (1-\gamma)\int_1^t e^{-b_1(1-\gamma)(t-\tau)} f(\tau)dt, \tag{24}$$

where $f(t)$ is defined in (9). And similarly the discrete form can be written as

$$\hat{y}^{(1)}(rp+t) = y^{(0)}(rp+1)e^{-b_1(1-\gamma)(t-1)} + (1-\gamma)\sum_{\tau=2}^{t}\left\{e^{-b_1(1-\gamma)(t-\tau+\frac{1}{2})}\cdot\frac{1}{2}[f(\tau)+f(\tau-1)]\right\}. \tag{25}$$

Then we can compute the 1-AGO series $\hat{X}_1^{(1)}$ using

$$\hat{X}_1^{(1)}(rp+t) = \left(\hat{y}^{(1)}(rp+t)\right)^{\frac{1}{1-\gamma}}. \tag{26}$$

And the restored values $\hat{X}_1^{(0)}$ can also be computed using the 1-IAGO

$$\hat{X}_1^{(0)}(rp+t) = \hat{X}_1^{(1)}(rp+t) - \hat{X}_1^{(1)}(rp+t-1). \tag{27}$$

3.4. Computational steps of the NGBMC(1, n) model

The computational steps of NGBMC(1, n) model can be summarized as follows

Step 1: Compute the 1-AGO series $\left(X_1^{(1)}, X_2^{(1)}, ..., X_n^{(1)}\right)$ of the given data set $\left(X_1^{(0)}, X_2^{(0)}, ..., X_n^{(0)}\right)$;

Step 2: Construct the matrices B and Y defined in (23) with a given γ and the definitions in (5) and (19)-(21) , and then compute the parameters $[b_1, b_2, ..., b_n, u]$ using the least squares estimation (23);

Step 3: Compute the series $\hat{y}^{(1)}$ using (25);

Step 4: compute the 1-AGO series $\hat{X}_1^{(1)}$ using (26);

Step 5: Compute the restored values $\hat{X}_1^{(0)}$ using the 1-IAGO (28).

4. Applications

In this section, the NGBMC(1, n) model will be applied in forecasting the domestic income and the foreign currency earnings of the tourist industry of China in order to assess its effectiveness in time series prediction, comparing to the existing linear multivariate grey prediction models, including the GMC(1, n) [50], RDGM(1, n) [18] and DGM(1, n)[51]. The ARIMA model [52] is used for the comparison as well. The raw data are all collected from the official website of the National Bureau of Statistics of China [1].

The mean absolute percentage error (MAPE) is used to evaluate the accuracy of the models, which is defined as

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{X}_1^{(0)}(rp+t) - X_1^{(0)}(rp+t)}{X_1^{(0)}(rp+t)}\right| \times 100\%. \tag{28}$$

[1] http://www.stats.gov.cn/

ACCEPTED MANUSCRIPT

The determination of the nonlinear parameter γ is quite simple. In this section, we search the γ in the interval $[-0.99, 0.99]$ by step of 0.01. The optimal γ is determined when the MAPE for fitting of the NGBMC(1, n) model achieves the minimum value. As for the parameter p (order for auto regression part) and q (order for moving average part) are determined by the AIC and BIC criteria, which is described at https://ww2.mathworks.cn/help/econ/choose-arma-lags.html. The raw data are directly used for modelling as the samples used in the cases are very small, thus the parameter D is set to be 0 in all the cases.

4.1. Case 1: Predicting the domestic income of the tourist industry of China

The latest data of the domestic income of the tourist industry of China from 2006 to 2015 are listed in Table 1, including the domestic income ($X_1^{(0)}$, unit: 10^9 CNY) and the number of trips ($X_2^{(0)}$, unit: 10^8). The first five pairs are used for building the models, including the GMC(1, 2), DGM(1, 2), RDGM(1, 2), ARIMA and the NGBMC(1, 2), and the last five pairs are used for testing. In this case, the parameters for ARIMA are $p = 1$ and $q = 0$ and the optimal nonlinear parameter of NGBMC(1, 2) is $\gamma = -0.45$, respectively.

The detailed results of each model and the MAPE for fitting and prediction are shown in Table 2 and the modelling values are also plotted in Fig.1. It can be seen that the RDGM(1, 2) shows the best performance in fitting, which is only 0.15%, the NGBMC(1, 2) performs worse than it, the ARIMA performs worse than the other grey models except the GMC(1, 2) model, which performs worst. But in terms of prediction, the MAPE of NGBMC(1, 2) is minimum in these models, the others' are extremely large. This indicates the results of the other four models are not acceptable at all. The Fig.1 presents a clearer picture that only the NGBMC(1, 2) can accurately predict the future domestic income of the tourist industry of China. In summary, the NGBMC(1, 2) performs best in this case study.

Table 1: Raw data of the domestic income (DI) of the tourist industry of China from 2006 to 2015

Year	DI (10^8 CNY)	Trips(10^8)
2006	6229.7	13.94
2007	7770.6	16.1
2008	8749.3	17.12
2009	10183.7	19.02
2010	12579.77	21.03
2011	19305.39	26.41
2012	22706.22	29.57
2013	26276.12	32.62
2014	30311.86	36.11
2015	34195.05	40

Table 2: The details of the prediction results of the domestic income of the tourist industry of China by GMC(1, 2), DGM(1, 2), RDGM(1, 2) and NGBMC(1, 2)

Year	DI (10^8 CNY)	GMC(1, 2)	Error(%)	DGM(1, 2)	Error(%)	RDGM(1, 2)	Error(%)	ARIMA		NGBMC(1, 2)	Error(%)
2006	6229.7	6229.70	0	6229.7	0	6229.7	0	6229.7	0	6229.7	0
2007	7770.6	7072.07	-8.99	7750.71	-0.26	7779.63	0.12	7790.60	0.26	7710.37	-0.78
2008	8749.3	7267.85	-16.93	8755.84	0.07	8740.02	-0.11	7780.60	-11.07	8425.41	-3.70
2009	10183.7	6899.53	-32.25	10133.19	-0.50	10202.23	0.18	8759.30	-13.99	10035.54	-1.45
2010	12579.77	4915.30	-60.93	12469.01	-0.88	12602.34	0.18	10193.70	-18.97	12237.59	-2.72
MAPE			29.77		0.43		0.15		11.07		2.16
2011	19305.39	-1881.40	-109.75	14941.12	-22.61	16226.46	-15.95	12589.77	-34.79	15115.06	-21.71
2012	22706.22	-20387.05	-189.79	19471.70	-14.25	22548.69	-0.69	12599.76	-44.51	18778.94	-17.30
2013	26276.12	-65921.27	-350.88	29845.61	13.58	36397.18	38.52	12609.76	-52.01	23369.01	-11.06
2014	30311.86	-175472.54	-678.89	56238.65	85.53	69023.50	127.71	12619.76	-58.37	29120.29	-3.93
2015	34195.05	-436664.50	-1376.98	127156.15	271.86	148494.64	334.26	12629.76	-63.07	36332.13	6.25
MAPE			541.26		81.56		103.43		50.55		**12.05**

ACCEPTED MANUSCRIPT

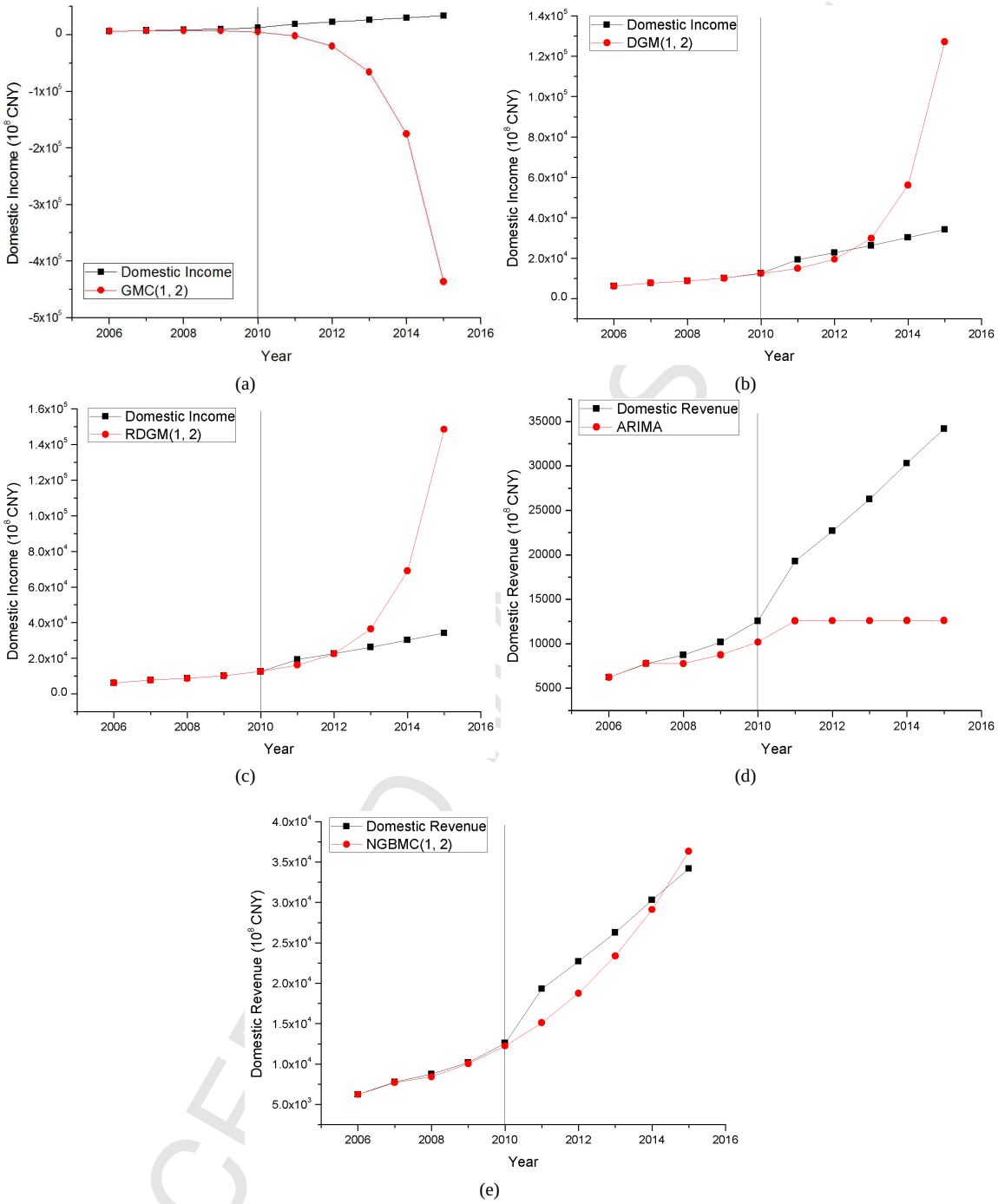


Fig. 1: The prediction results of the domestic income of the tourist industry of China. (a) GMC(1, 2); (b) DGM(1, 2); (c) RDGM(1, 2); (d) NGBMC(1, 2)

4.2. Case 2: Predicting the foreign currency earnings of the tourist industry of China

The latest data of the foreign currency earnings of the tourist industry of China from 2006 to 2015 are listed in Table 3, including the foreign currency earnings ($X_1^{(0)}$, unit: 10^6 USD) and the number of inbound

ACCEPTED MANUSCRIPT

trips ($X_2^{(0)}$, unit: 10^8). The first five pairs are used for building the models, including the GMC(1, 2), DGM(1, 2), RDGM(1, 2), ARIMA and the NGBMC(1, 2), and the last five pairs are used for testing. In this case, the parameters of ARIMA are $p = 1$ and $q = 0$, which is the same to case 1, and the optimal nonlinear parameter of NGBMC(1, 2) is $\gamma = -0.43$, respectively.

The detailed results of each model and the MAPE for fitting and prediction are shown in Table 4 and the modelling values are also plotted in Fig.2. It can be seen that the performances of these models differ more largely in this case. It is quite clear to see that the GMC(1, 2), RDGM(1, 2) and ARIMA all perform worse than NGBMC(1, 2) model for fitting and prediction. The DGM(1, 2) performs best in fitting with the MAPE 0.05% and the NGBMC(1, 2) model performs best in prediction, but the MAPE of prediction for DGM(1, 2) is 28.36%, which is only 1.31% larger than NGBMC(1, 2). Maybe it is not easy to say which is better between the DGM(1, 2) and NGBMC(1, 2) in this case with such small difference. But a very interesting results can be found in Fig.2. It can be clearly seen that the foreign currency earnings continuously grow from 2011 to 2015, only the NGBMC(1, 2) model correctly catches such tendency, the trend reflected by the other models are all wrong.

Table 3: Raw data of the foreign currency earnings (FCE) of the tourist industry of China from 2006 to 2015

Year	FCE (10^6 USD)	Trips(10^8)
2006	33949	1.249421
2007	41919	1.318733
2008	40843	1.300274
2009	39675	1.264759
2010	45814	1.337622
2011	48464	1.354235
2012	50028	1.324053
2013	51664	1.290778
2014	56913	1.284983
2015	113650	1.338204

Table 4: The details of the prediction results of the foreign currency earnings (FCE) of the tourist industry of China by GMC(1, 2), DGM(1, 2), RDGM(1, 2), ARIMA and NGBMC(1, 2)

Year	FCE (10^8 USD)	GMC(1, 2)	Error(%)	DGM(1, 2)	Error(%)	RDGM(1, 2)	Error(%)	ARIMA	Error(%)	NGBM(1, n)	Error(%)
2006	33949	33949	0	33949	0	33949	0	33949	0	33949	0
2007	41919	6.59E+05	1.47E+03	41915.06	-0.01	42059.62	0.34	40450.54	-3.50	42035.66	0.28
2008	40843	4.14E+07	1.01E+05	40859.20	0.04	40011.47	-2.04	40168.60	-1.65	38504.12	-5.73
2009	39675	2.68E+09	6.76E+06	39645.39	-0.07	42456.51	7.01	40341.91	1.68	40064.67	0.98
2010	45814	1.74E+11	3.80E+08	45850.35	0.08	37920.04	-17.23	40530.05	-11.53	44312.06	-3.28
MAPE			9.67E+07		0.05		6.65		4.59		2.57
2011	48464	1.13E+13	2.33E+10	40141.58	-17.17	55819.41	15.18	39541.20	-18.41	51008.68	5.25
2012	50028	7.33E+14	1.46E+12	44397.93	-11.25	5469.37	-89.07	40551.60	-18.94	60736.43	21.40
2013	51664	4.75E+16	9.20E+13	37495.65	-27.42	141031.96	172.98	40388.85	-21.82	74250.66	43.72
2014	56913	3.08E+18	5.42E+15	44693.99	-21.47	-236494.61	-515.54	40415.07	-28.99	92447.49	62.44
2015	113650	2.00E+20	1.76E+17	40340.75	-64.50	811208.54	613.78	40410.84	-64.44	116433.76	2.45
MAPE			3.63E+16		28.36		281.31		30.52		**27.05**

4.3. Short discussions on the results

According to the results shown above, it can be seen that the DGM(1, 2) and RDGM(1, 2) are generally better than the GMC(1, 2) model. This is coincident to the results of the previous works [18, 51], which have proved that the RDGM(1, n) and DGM(1, n) are generally more accurate than GMC(1, n) model. It is also very interesting to see that the best fitting accuracy are achieved by RDGM(1, 2) and DGM(1, 2) in case 1 and case 2, respectively, but they all perform worse in prediction than NGBMC(1, 2) in all the cases. This indicates the nonlinearity of NGBMC(1, 2) is more effective than the existing linear grey models.

It should be noticed that the ARIMA performs the worst in all the cases. It is well known that the ARIMA is designed for stationary time series, and it usually needs large amount of sample data. But the

ACCEPTED MANUSCRIPT

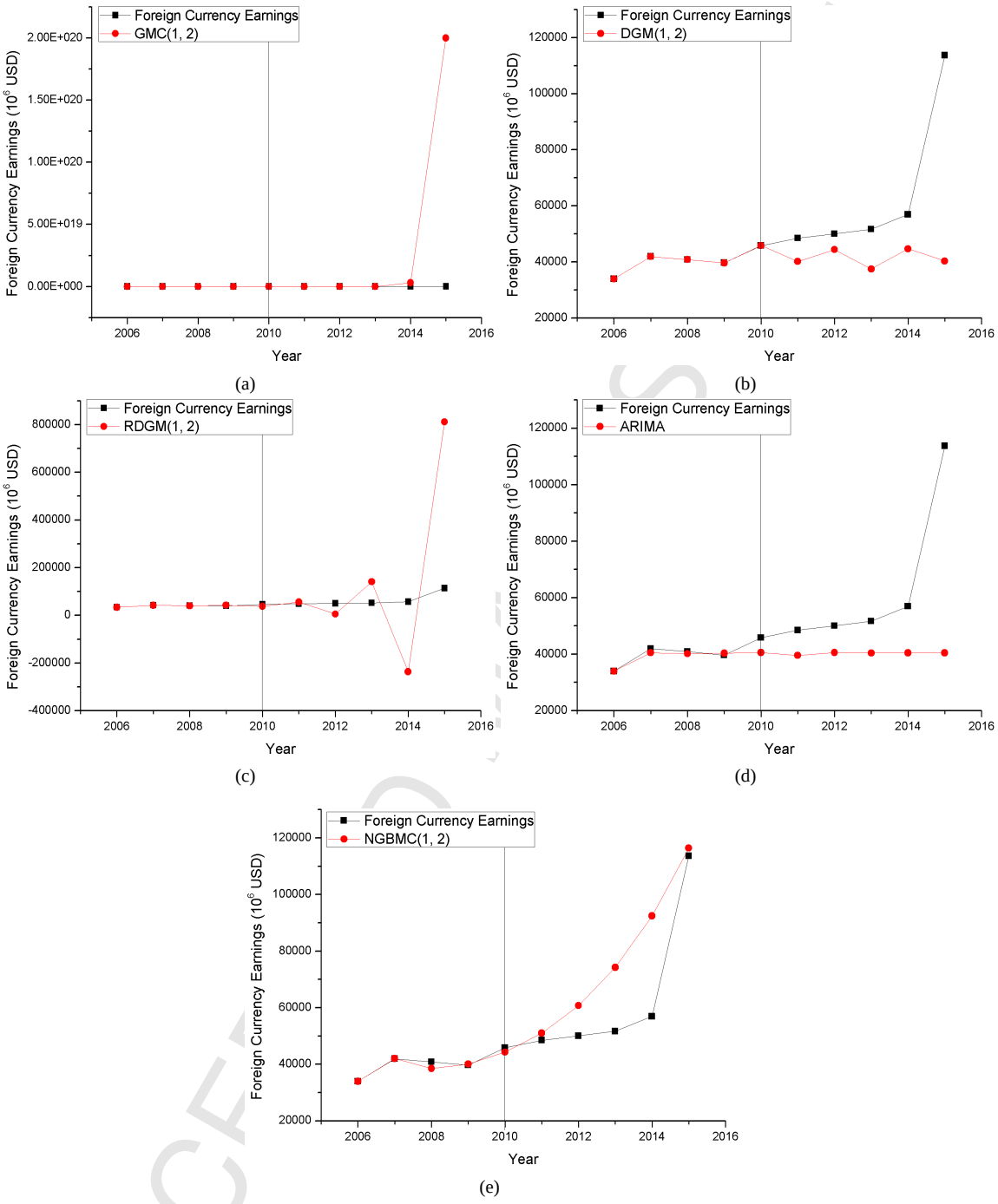


Fig. 2: The prediction results of the foreign currency earnings of the tourist industry of China. (a) GMC(1, 2); (b) DGM(1, 2); (c) RDGM(1, 2); (d) ARIMA; (e) NGBMC(1, 2)

data size of the case studies is very small, which leads to inaccuracy of the ARIMA. On the other hand, the ARIMA is also essentially a linear model as it shares the linear formulation, which is another reason why the ARIMA underperforms the NGBMC(1, 2) in the cases presented above.

ACCEPTED MANUSCRIPT

5. Conclusions

In this paper we have proposed a novel nonlinear grey model by introducing the Bernoulli equation and using the modelling methodology of the GMC(1, n) model. It can be seen that the NGBMC(1, n) model can be regarded as an extension of the NGBM(1, 1) model, but it doesn't share the same modelling procedures of the NGBM(1, 1). Instead, we take the idea of the existing nonlinear models and the whitening equation of the NGBMC(1, n) has been firstly translated to a linear equation, and such linear form is mathematically equivalent to the GMC(1, n) model. It can be seen that this operation makes it much easier to build the NGBMC(1, n) model, as its linearization shares so many similarities with the GMC(1, n) model, even some results of GMC(1, n) can be directly used. The results of the case studies indicate that the NGBMC(1, n) outperforms the existing multivariate linear models, and its accuracy are generally much more higher than the linear models in these cases. In summary, the NGBMC(1, n) model is an easy-to-use and efficient nonlinear grey model.

However, there still exist some issues which have not been solved in this paper. For example, the error variance and prediction interval with $1 - \alpha$ confidence (e.g. 95%). It should be noticed that such validation criteria are very commonly used for the statistical models, but rarely used for the grey models (e.g. no discussions on the error variance and prediction interval can be found in the references cited in this paper [2–10, 12–23, 34–48, 50, 51, 53–59]). Thus, in our opinion, the future works can be oriented on the theoretical analysis on these issues. On the other hand, the NGBMC(1, n) could also be regarded as a representation of the modelling methodology for the nonlinear grey models, within which more nonlinear models can be built with high accuracy. And in our next works, we will introduce the state-of-the-art intelligent optimization methods such as the PSO [60], multi-objective evolutionary optimizers [61] to develop the new framework of the intelligent nonlinear grey models.

6. Acknowledgement

The authors sincerely thank the editors and the anonymous for their insightful comments which helped to improve this work. Especial thanks are due to the Prof. Marc Goovaerts, who will always live in our memory.

This work is supported by the Open Fund (PLN201710) of State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation (Southwest Petroleum University) and Doctoral Research Foundation of Southwest University of Science and Technology (no. 16zx7140).

References

[1] D. Ju-Long, Control problems of grey systems, Systems & Control Letters 1 (5) (1982) 288–294.
[2] J. Cui, S. feng Liu, B. Zeng, N. ming Xie, A novel grey forecasting model and its optimization, Applied Mathematical Modelling 37 (6) (2013) 4399 – 4406.
[3] T.-L. Tien, A new grey prediction model FGM(1, 1), Mathematical and Computer Modelling 49 (7C8) (2009) 1416 – 1426.
[4] X. Ma, Z. Liu, Application of a novel time-delayed polynomial grey model to predict the natural gas consumption in China, Journal of Computational and Applied Mathematics 324 (2017) 17–24. doi:10.1016/j.cam.2017.04.020.
[5] H. Duan, X. Xiao, L. Pei, Forecasting the short-term traffic flow in the intelligent transportation system based on an inertia nonhomogenous discrete gray model, Complexity 2017 (2017) 1–17.
[6] Z.-X. Wang, Q. Li, L.-L. Pei, A seasonal GM (1, 1) model for forecasting the electricity consumption of the primary economic sectors, Energy 154 (2018) 522–534.
[7] B. Wei, N. Xie, A. Hu, Optimal solution for novel grey polynomial prediction model, Applied Mathematical Modelling 62 (2018) 717 – 727.
[8] Z.-x. Wang, Y.-g. Dang, S.-f. Liu, An optimal GM(1, 1) based on the discrete function with exponential law, Systems Engineering-Theory & Practice 2 (2008) 61–67.
[9] P.-Y. Chen, H.-M. Yu, Foundation settlement prediction based on a novel NGM model, Mathematical Problems in Engineering 2014.
[10] B. Zeng, H. Duan, Y. Bai, W. Meng, Forecasting the output of shale gas in China using an unbiased grey model and weakening buffer operator, Energy 151 (2018) 238–249.
[11] J. Wang, P. Du, H. Lu, W. Yang, T. Niu, An improved grey model optimized by multi-objective ant lion optimization algorithm for annual electricity consumption forecasting, Applied Soft Computing.doi:https://doi.org/10.1016/j.asoc.2018.07.022.
[12] T.-L. Tien, A research on the grey prediction model GM(1, n), Applied mathematics and computation 218 (9) (2012) 4903–4916.

ACCEPTED MANUSCRIPT

[13] C.-K. Chen, T.-L. Tien, The indirect measurement of tensile strength by the deterministic grey dynamic model DGDM(1, 1, 1), International journal of systems science 28 (7) (1997) 683–690.
[14] T.-L. Tien, The deterministic grey dynamic model with convolution integral DGDMC(1, n), Applied Mathematical Modelling 33 (8) (2009) 3498–3510.
[15] T.-L. Tien, The indirect measurement of tensile strength for a higher temperature by the new model IGDMC(1, n), Measurement 41 (6) (2008) 662–675.
[16] T.-L. Tien, The indirect measurement of tensile strength by the new model FGMC(1, n), Measurement 44 (10) (2011) 1884–1897.
[17] B. Zeng, C. Luo, S. Liu, Y. Bai, C. Li, Development of an optimization method for the GM(1, n) model, Engineering Applications of Artificial Intelligence 55 (2016) 353 – 362.
[18] X. Ma, Z.-b. Liu, Research on the novel recursive discrete multivariate grey prediction model and its applications, Applied Mathematical Modelling 40 (7) (2016) 4876–4890.
[19] B. Zeng, C. Li, Improved multi-variable grey forecasting model with a dynamic background-value coefficient and its application, Computers & Industrial Engineering 118 (2018) 278–290.
[20] L.-L. Pei, Q. Li, Z.-X. Wang, The NLS-Based nonlinear grey multivariate model for forecasting pollutant emissions in China, International journal of environmental research and public health 15 (3). doi:10.3390/ijerph15030471.
[21] L. Wu, Z. Zhang, Grey multivariable convolution model with new information priority accumulation, Applied Mathematical Modelling 62 (2018) 595 – 604.
[22] L. Wu, X. Gao, Y. Xiao, Y. Yang, X. Chen, Using a novel multi-variable grey model to forecast the electricity consumption of shandong province in china, Energy 157 (2018) 327 – 335. doi:https://doi.org/10.1016/j.energy.2018.05.147.
[23] D. Julong, Fundations of grey theory, Huazhong University of Science and Technology Press, 2002, in Chinese.
[24] Y. Wang, X. Yi, Transient pressure behavior of a fractured vertical well with a finite-conductivity fracture in triple media carbonate reservoir, Journal of Porous Media 20 (8) (2017) 707–722. doi:10.1615/JPorMedia.v2018.30.
[25] Y. Wang, X. Yi, Flow modeling of well test analysis for a multiple-fractured horizontal well in triple media carbonate reservoir, International Journal of Nonlinear Sciences and Numerical Simulationdoi:10.1515/ijnsns-2016-0075.
[26] W. Cai, F. Liu, O. Dinolov, J. Xie, P. Liu, J. Tuo, Energy benchmarking rules in machining systems, Energy 142 (2018) 258–263.
[27] W. Cai, C. Liu, , C. Zhang, M. Ma, W. Rao, W. Li, K. He, M. Gao, Developing the ecological compensation criterion of industrial solid waste based on emergy for sustainable development, Energy 157 (2018) 940–948.
[28] Y. Hu, E. Mackay, O. Vazquez, O. Ishkov, et al., Streamline simulation of barium sulfate precipitation occurring within the reservoir coupled with analyses of observed produced-water-chemistry data to aid scale management, SPE Production & Operations.
[29] J. Fan, X. Wang, L. Wu, H. Zhou, F. Zhang, X. Yu, X. Lu, Y. Xiang, Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in china, Energy Conversion and Management 164 (2018) 102–111.
[30] P. Du, J. Wang, W. Yang, T. Niu, Multi-step ahead forecasting in electrical power system using a hybrid forecasting system, Renewable Energy 122 (2018) 533–550.
[31] J. Wang, W. Yang, P. Du, Y. Li, Research and application of a hybrid forecasting framework based on multi-objective optimization for electrical power system, Energy 148 (2018) 59–78.
[32] Y.-B. Li, Y. Chen, W.-F. Pu, H. Gao, B. Bai, Experimental investigation into the oxidative characteristics of tahe heavy crude oil, Fuel 209 (2017) 194–202.
[33] J. Fan, B. Chen, L. Wu, F. Zhang, X. Lu, Y. Xiang, Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions, Energy 144 (2018) 903–914.
[34] S. Liu, Y. Lin, J. Y. L. Forrest, Grey systems: theory and applications, Springer, 2010.
[35] C.-I. Chen, H. L. Chen, S.-P. Chen, Forecasting of foreign exchange rates of Taiwans major trading partners by novel nonlinear grey Bernoulli model NGBM(1, 1), Communications in Nonlinear Science and Numerical Simulation 13 (6) (2008) 1194 – 1204.
[36] C.-I. Chen, Application of the novel nonlinear grey Bernoulli model for forecasting unemployment rate, Chaos, Solitons & Fractals 37 (1) (2008) 278 – 287.
[37] J. Zhou, R. Fang, Y. Li, Y. Zhang, B. Peng, Parameter optimization of nonlinear grey Bernoulli model using particle swarm optimization, Applied Mathematics and Computation 207 (2) (2009) 292 – 299.
[38] L.-C. Hsu, A genetic algorithm based nonlinear grey Bernoulli model for output forecasting in integrated circuit industry, Expert Systems with Applications 37 (6) (2010) 4318 – 4323.
[39] C.-I. Chen, P.-H. Hsin, C.-S. Wu, Forecasting Taiwans major stock indices by the Nash nonlinear grey Bernoulli model, Expert Systems with Applications 37 (12) (2010) 7557 – 7562.
[40] Z.-X. Wang, K. W. Hipel, Q. Wang, S.-W. He, An optimized NGBM(1,1) model for forecasting the qualified discharge rate of industrial wastewater in China, Applied Mathematical Modelling 35 (12) (2011) 5524 – 5532.
[41] P.-H. Hsin, C.-I. Chen, Application of game theory on parameter optimization of the novel two-stage Nash nonlinear grey Bernoulli model, Communications in Nonlinear Science and Numerical Simulation 27 (1C3) (2015) 168 – 174.
[42] Z.-X. Wang, An optimized Nash nonlinear grey Bernoulli model for forecasting the main economic indices of high technology enterprises in China, Computers & Industrial Engineering 64 (3) (2013) 780 – 787.
[43] L. Zhang, Y. Zheng, K. Wang, X. Zhang, Y. Zheng, An optimized nash nonlinear grey Bernoulli model based on particle swarm optimization and its application in prediction for the incidence of Hepatitis b in xinjiang, china, Computers in Biology and Medicine 49 (2014) 67 – 73.
[44] J. Lu, W. Xie, H. Zhou, A. Zhang, An optimized nonlinear grey Bernoulli model and its applications, Neurocomputing 177 (2016) 206 – 214.
[45] Z.-X. Wang, Nonlinear grey prediction model with convolution integral NGMC and its application to the forecasting of Chinas industrial emissions, Journal of Applied Mathematics 2014.
[46] X. Ma, Z. Liu, Y. Wei, X. Kong, A novel kernel regularized nonlinear GMC(1, n) model and its application, Journal of Grey

ACCEPTED MANUSCRIPT

System 28 (3) (2016) 97–109.
[47] Z.-X. Wang, D.-J. Ye, Forecasting Chinese carbon emissions from fossil energy consumption using non-linear grey multivariable models, Journal of Cleaner Production 142, Part 2 (2017) 600 – 612.
[48] X. Ma, Y.-s. Hu, Z.-b. Liu, A novel kernel regularized nonhomogeneous grey model and its applications, Communications in Nonlinear Science and Numerical Simulation 48 (2017) 51–62.
[49] X. Ma, Z. Liu, Predicting the oil production using the novel multivariate nonlinear model based on Arps decline model and kernel method, Neural Computing and Applications 29 (2) (2018) 579–591.
[50] T.-L. Tien, The indirect measurement of tensile strength of material by the grey prediction model GMC(1, n), Measurement Science and Technology 16 (6) (2005) 1322.
[51] X. Ma, Z. Liu, Predicting the oil field production using the novel discrete GM(1, n) model, The Journal of Grey System 27 (4) (2015) 63–74.
[52] G. E. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, Time series analysis: forecasting and control, John Wiley and Sons, 2015.
[53] N. Xie, N. Chen, Flexible job shop scheduling problem with interval grey processing time, Applied Soft Computing 70 (2018) 513–524.
[54] Z.-X. Wang, K. W. Hipel, Q. Wang, S.-W. He, An optimized NGBM(1, 1) model for forecasting the qualified discharge rate of industrial wastewater in china, Applied Mathematical Modelling 35 (12) (2011) 5524 – 5532.
[55] N.-m. Xie, S.-f. Liu, Discrete grey forecasting model and its optimization, Applied Mathematical Modelling 33 (2) (2009) 1173–1186.
[56] X. Ma, Z. Liu, The GMC (1, n) model with optimized parameters and its application, Journal of grey system 29 (4) (2017) 122–138.
[57] C.-R. Su, C.-K. Chen, W.-L. Liu, H.-Y. Lai, Estimation for inner surface geometry of furnace wall using inverse process combined with grey prediction model, International Journal of Heat and Mass Transfer 52 (15) (2009) 3595–3605.
[58] W.-Y. Wu, S.-P. Chen, A prediction method using the grey model GMC(1, n) combined with the grey relational analysis: a case study on internet access population forecast, Applied Mathematics and Computation 169 (1) (2005) 198–217.
[59] Z.-X. Wang, A predictive analysis of clean energy consumption, economic growth and environmental regulation in china using an optimized grey dynamic model, Computational Economics (2015) 1–17.
[60] Z. Zeng, J. Xu, S. Wu, M. Shen, Antithetic method-based particle swarm optimization for a queuing network problem with fuzzy data in concrete transportation systems, Computer-Aided Civil and Infrastructure Engineering 29 (10) (2014) 771–800.
[61] P. Du, J. Wang, Z. Guo, W. Yang, Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting, Energy Conversion and Management 150 (2017) 90–107.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

- A novel nonlinear Bernoulli grey model NGBMC(1, n) is proposed.
- The NGBMC(1, n) model is more general than the existing NGBM(1, 1) and GMC(1, n)
- The NGBMC(1, n) model is quite easy-to-use and more efficient than the existing linear models
- The NGBMC(1, n) is successfully applied in tourist industry of China
- A new methodology of nonlinear grey models is provided

ACCEPTED MANUSCRIPT