axioms

Article

Novel Fractional Grey Prediction Model with the Change-Point Detection for Overseas Talent Mobility Prediction

Peng Jiang [1], Geng Wu [2,*], Yi-Chung Hu [2], Xue Zhang [1] and Yining Ren [1]

[1] School of Business, Shandong University, Weihai 264209, China
[2] Department of Business Administration, Chung Yuan Christian University, Taoyuan 320314, Taiwan
* Correspondence: wgoodbye@yeah.net

Abstract: Overseas students constitute the paramount talent resource for China, and, hence, overseas talent mobility prediction is crucial for the formulation of China's talent strategy. This study proposes a new model for predicting the number of students studying abroad and returning students, based on the grey system theory, owing to the limited data and uncertainty of the influencing factors. The proposed model introduces change-point detection to determine the number of modeling time points, based on the fractional-order grey prediction model. We employed a change-point detection method to find the change points for determining the model length, based on the principle of new information priority, and used a fractional order accumulated generating operation to construct a grey prediction model. The two real data sets, the annual number of students studying abroad and returning students, were employed to verify the superiority of the proposed model. The results showed that the proposed model outperformed other benchmark models. Furthermore, the proposed model has been employed to predict the tendencies of overseas talent mobility in China by 2025. Further, certain policy recommendations for China's talent strategy development have been proposed, based on the prediction results.

Keywords: students studying abroad; returned students; change-point detection; fractional grey prediction model

Citation: Jiang, P.; Wu, G.; Hu, Y.-C.; Zhang, X.; Ren, Y. Novel Fractional Grey Prediction Model with the Change-Point Detection for Overseas Talent Mobility Prediction. *Axioms* **2022**, *11*, 432. https://doi.org/10.3390/axioms11090432

Academic Editor: Lifeng Wu

Received: 25 July 2022
Accepted: 22 August 2022
Published: 26 August 2022

1. Introduction

1.1. Background

Due to rapid economic development and the deepening internationalization of China, increasing numbers of Chinese students prefer studying abroad and, likewise, increasing numbers of overseas talent are seeking a return to China [1,2]. The return of overseas talent has significantly contributed to China's rapid development, and, hence, these overseas students are an important resource of high-level talents [3]. Moreover, with intensified global competition for international talent, attracting more overseas students to serve their home country has become an urgent issue for all developing countries, particularly China [4]. Therefore, the return of overseas talent has become a prime focus of research.

Talent return is essentially talent mobility. The current research on talent mobility primarily focuses on the influence factors [5], innovation and entrepreneurship [3,4], knowledge management [6–8], and performance management [3,9], whereas few researchers have concentrated on the challenges of prediction. Accurate predictions of the talent management are crucial for China in helping the policy makers to comprehend the flow of overseas talent in future, and, henceforth, to design appropriate policies. Therefore, research on the prediction of Chinese overseas talent mobility is highly significant.

This study focuses on the prediction of Chinese overseas talent mobility, which includes the prediction of the number of Chinese students who study abroad and who return to China. Since the predicted data set is a complex system, its tendencies of development and external influences are uncertain. Hence, this study employs the grey prediction model to explore the issue of prediction of Chinese overseas talent mobility.

1.2. Literature Review

Grey theory is a theoretical approach to deal with uncertain information proposed by Professor Deng in 1982 [10]. The grey prediction technique is one of the main parts of grey theory and is an important branch of the modern prediction theory system. Grey prediction models can be divided into univariate models, such as GM(1,1), and multivariate models, such as GM(1,N), and most of the other models are developed based on the two models [11]. During the past 40 years, grey prediction techniques have been widely used in various fields, such as energy [12,13], traffic [14,15], economy [16,17], agriculture [18], tourism [19,20], medicine [21,22], and the environment [23,24].

The core of the grey prediction model lies in its modeling based on dynamic system equations, using accumulation generation and differential equations to characterize the system evolution patterns [25]. It has certain advantages in solving practical problems [11]. Compared with traditional statistical analysis models and machine learning models, grey models not only do not require statistical assumptions, but also do not need large amounts of data [26]. In addition, the grey model is easy to implement and only requires at least four points for modeling to obtain satisfactory results [27].

However, there are still some limitations of the grey prediction model to be improved, and we focus on the implementation of the principle of new information priority and the determination of the model length in this study. The traditional grey prediction models employ the whole data set for modeling. However, the trends in the whole data set may have certain changes in the process of time series generation, owing to the perturbation of the external information. Thus, it is difficult to obtain a satisfactory performance for new data prediction by using old data, when there are dissimilar trends between the old and new data. The principle of the new information priority for the grey prediction models proposed by Professor Deng can be employed to solve this challenging problem [28].

The three current approaches that can accomplish the principle of new information priority for the grey prediction models are the grey models with a rolling mechanism, such as rolling-GM(1,1) [29], segment grey models, such as SGM(1,1) [30], and fractional order grey prediction models, such as fractional GM(1,1) [31] and fractional Hausdorff GM(1,1) [32]. The grey prediction model with a rolling mechanism primarily implements the principle of prioritizing new information by continuously discarding old data and adding new data. However, the modeling length of this method is not conclusive. Since the grey model requires at least four points for modeling [27], Akay and Atak [29] have employed four-point modeling, whereas Wang et al. [33] utilized recent data with exponential growth for rolling modeling. The studies of Yuan et al. [34] and Liu et al. [35] adopted different modeling lengths. The problem of uncertainty of the modeling length i.e., the old data that affects the modeling accuracy, cannot be solved. The segment grey model intercepts the most recent continuous segment s of the whole data set for modeling, with the length l of segment s, and $l \geq 4$. Then, the length l that minimizes the mean absolute percentage error (MAPE) is selected as the optimal length for the best input subset. However, the method is only based on the principle of minimizing MAPE, and when the l is long, the selected modeled data segment may still have varying tendencies. Further, the method is similar to the brute-force parametric method in machine learning, which is time-consuming and laborious. The fractional grey prediction model assigns different weights to the time points by introducing the fractional order accumulated generating operation. Although the fractional order enables the principle of new information priority [31], the old data still need to be taken into account in the model fitting process, which affects the model fitting and prediction.

The above analysis shows that these methods do not consider the varying trends between the old and new data for the modeling based on the new information priority principle. The fractional-order grey model considers the old data for modeling, and the rolling and segment grey models require multiple modeling, which enhances the modeling complexity.

1.3. Contributions

This study proposes a method that utilizes change-point detection for determining the model length for the grey prediction model construction, based on the principle of new information priority. Since change-point detection can detect abrupt changes in time series data [36], we apply change-point detection to identify different trends in time series data sets, and, then, the new data with the same trend are retained, whereas the old data with different trends are discarded. The retained data are utilized to construct the grey prediction model, and the length of the retained data is the modeling length. Since the retained data are all new data, the method also satisfies the principle of new information priority.

In brief, this study proposes a fractional grey prediction model based on change-point detection for overseas talent mobility prediction. This model uses change-point detection to select the modeling data for determining the modeling length, and applies fractional order accumulated generating operation to construct the grey prediction model. The contributions from this study are given as follows.

(1) We studied overseas talent mobility prediction by employing the grey prediction model for the prediction of overseas talent mobility.

(2) We propose a method for determining the modeling length of the FGM(1,1) model using change-point detection based on new information priority.

(3) We propose a high-precision model for Chinese overseas talent mobility prediction. We also provide certain feasible suggestions for the relevant management decision-making departments, based on the prediction results.

The remaining part of this paper is organized as follows. Section 2 introduces the GM(1,1) and FGM(1,1), and Section 3 describes the proposed grey prediction model. Section 4 examines the proposed model for the prediction of Chinese students studying abroad and returning students to China. Conclusion and future work are briefed in Section 5.

2. GM(1,1) and FGM(1,1)

2.1. GM(1,1)

GM(1,1) is a basic for the grey prediction model, and the first "1" and the last "1" implies the first-order differential equation and one dependent variable, respectively [37]. The modeling steps for GM(1,1) are given below.

(1) Set $X^{(0)}$ as an original non-negative sequence,

$$\mathrm{X}^{(0)} = \left(x_1^{(0)}, x_2^{(0)}, \cdots, x_n^{(0)}\right) \tag{1}$$

(2) Convert $X^{(0)}$ to the first-order accumulation sequence $X^{(1)}$ by 1-order accumulated generating operation (1-AGO),

$$\mathrm{X}^{(1)} = \sum_{i=1}^{k} x_i^{(0)},\ k = 1, 2, \cdots, n \tag{2}$$

(3) Set $Z^{(1)}$ as an immediately adjacent mean generating sequence of $X^{(1)}$,

$$\mathrm{Z}^{(1)} = 0.5 \times x_k^{(1)} + 0.5 \times x_{k-1}^{(1)},\ k = 2, 3, \cdots, n \tag{3}$$

(4) Construct grey differential equations as

$$x_k^{(0)} + a \cdot z_k^{(1)} = b,\ k = 2, 3, \cdots, n \tag{4}$$

(5) Estimate the parameters a and b by least squares method

$$[a, b]^T = \left(B^T B\right)^{-1} B^T Y \tag{5}$$

where

$$B = \begin{bmatrix} -z_2^{(1)} & 1 \\ -z_3^{(1)} & 1 \\ \vdots & \vdots \\ -z_n^{(1)} & 1 \end{bmatrix}, Y = \begin{bmatrix} x_2^{(0)} \\ x_3^{(0)} \\ \vdots \\ x_n^{(0)} \end{bmatrix} \tag{6}$$

(6) Obtain the time response series of the grey differential equation,

$$\hat{x}_k^{(1)} = \left(x_1^{(0)} - \frac{b}{a} \right) \cdot e^{-a(k-1)} + \frac{b}{a},\ k = 2, 3, \cdots, n \tag{7}$$

(7) The final prediction value $x_k^{(0)}$ is obtained by 1-order reverse AGO,

$$\hat{x}_k^{(0)} = \hat{x}_k^{(1)} - \hat{x}_{k-1}^{(1)},\ k = 2, 3, \cdots, n \tag{8}$$

2.2. FGM(1,1)

FGM(1,1) is a fractional version of GM(1,1) with a fractional order accumulation. Therefore, the modeling process of FGM(1,1) is basically the same as that of GM(1,1). The difference is that FGM(1,1) uses fractional order accumulation and inverse accumulation operation to grey and whiten the model, respectively. The fractional order accumulation sequence $X^{(r)}$ is obtained through the r-order accumulated generating operation (r-AGO),

$$X^{(r)} = \sum_{i=1}^{k} \binom{k-i+r-1}{k-i} x_i^{(0)},\ k = 1, 2, \cdots, n \tag{9}$$

where

$$\binom{k-i+r-1}{k-i} = \frac{(k-i+r-1)(k-i+r-2)\cdots(r+1)r}{(k-i)!} \tag{10}$$

The fractional grey differential equation is constructed as

$$x_k^{(0)} + a \cdot z_k^{(r)} = b,\ k = 2, 3, \cdots, n \tag{11}$$

The time response series of the fractional grey differential equation is obtained as,

$$\hat{x}_k^{(r)} = \left(x_1^{(0)} - \frac{b}{a} \right) \cdot e^{-a(k-1)} + \frac{b}{a},\ k = 2, 3, \cdots, n \tag{12}$$

Consequently, the final prediction value $x_k^{(0)}$ is obtained by the r-order reverse AGO,

$$\hat{x}_k^{(0)} = \sum_{i=1}^{k} \binom{k-i-r-1}{k-i} \hat{x}_i^{(r)},\ k = 2, 3, \cdots, n \tag{13}$$

It is noteworthy that the FGM(1,1) model is equivalent to the GM(1,1) model when $r = 1$.

3. The Proposed Grey Prediction Model

3.1. Change-Point Detection

Change-point detection intends to find out the transition points that produce changes in the process of time series generation, and is extensively employed in medical, financial, meteorological, and other fields [38]. We employ the environmental time series change point detection (EnvCpt) method proposed by Beaulieu and Killick [39] for change-point detection, a method which has proven its effectiveness in the fields of environment [39], meteorology [40], and tourism [41]. The EnvCpt method utilizes the maximum likelihood estimation to estimate the change-points, besides selecting the model with the minimum Akaike information criterion (AIC) as the best fitting model. Then, it relies on a pruned exact linear time (PELT) algorithm [42] for obtaining the optimal number of change-points. Another reason for choosing the EnvCpt method in this study is its easy implementation, and the "EnvCpt" package in R can automatically apply the EnvCpt method for change-point detection with 12 different models, comprising "Trend cpt + AR(2)", "Trend cpt + AR(1)", "Trend cpt", "Mean cpt + AR(2)", "Mean cpt + AR(1)", "Mean cpt", "Trend + AR(2)", "Trend + AR(1)", "Trend", "Mean + AR(2)", "Mean + AR(1)", and "Mean" [43]. The "Trend cpt + AR(2)" denotes the multiple change-points in the trend with the second-order autoregression, it is calculated as:

$$y_t = \begin{cases} \lambda_1 + \beta_1 t + \varphi_1 y_{t-1} + \varphi_1' y_{t-2} + e_t & t \leq c_1 \\ \lambda_2 + \beta_2 t + \varphi_2 y_{t-1} + \varphi_2' y_{t-2} + e_t & c_1 < t \leq c_2 \\ \vdots & \vdots \\ \lambda_m + \beta_m t + \varphi_m y_{t-1} + \varphi_m' y_{t-2} + e_t & c_{m-1} < t \leq n \end{cases} \tag{14}$$

where

y_t is the time series,

λ and β are the intercept and trend parameters, respectively,

φ is the autocorrelation coefficients,

e_t is the normal-distributed white noise errors,

c is the timing of the changepoints between segments,

n is the length of the time series.

Due to the technical limitations, only the above model is presented in this section, and details of the remaining 11 models can be found in the description of the "EnvCpt" package [43] and in [39].

3.2. Combining Change-Point Detection and FGM(1,1)

This study combines change-point detection with FGM(1,1) to propose a new grey prediction model suitable for prediction of Chinese overseas study talent, abbreviated as CPD-FGM(1,1). The proposed model uses change-point detection to determine the modeling length and FGM(1,1) to construct the CPD-FGM(1,1) model. The specific modeling steps are given as follows.

(1) Determine the modeling length by employing the change-point detection. First, we employ the EnvCpt method to find the n change points of the original time series. Then, the original time series is divided into n + 1 sub-sequences. If the length of the last subsequence l_{n+1} is greater than, or equal to, s, then the length of the CPD-FGM(1,1) model l is equal to l_{n+1}. Further, if the length of the last subsequence l_{n+1} is less than s, then the length of the CPD-FGM(1,1) model l is equal to the sum of l_{n+1} and l_n, while ensuring that l is also greater than s, and so forth. Since FGM(1,1) requires at least four points for modeling [16], we set s equal to 4. Furthermore, if n is equal to 0, the length of the CPD-FGM(1,1) model is equal to the length of the original time series.

(2) Construct the proposed model by using FGM(1,1). The modeling process of CPD-FGM(1,1) is identical to the FGM(1,1) in Section 2.2. Furthermore, the optimal fractional order is estimated by particle swarm optimization (PSO). The PSO algorithm is one of the most common tools for parameter optimization of grey prediction models [44,45], and has some advantages over other meta-heuristics in terms of optimization capability, stability, and robustness [46]. We used the EvoloPy framework in Python proposed by Faris et al. [47] to implement the PSO algorithm. The following were the specific parameters of PSO: the maximum number of iterations was 100, the particle number was 50, the search dimension was 1, and the others were the default.

Figure 1 illustrates the construction process of the proposed CPD-FGM(1,1) model.

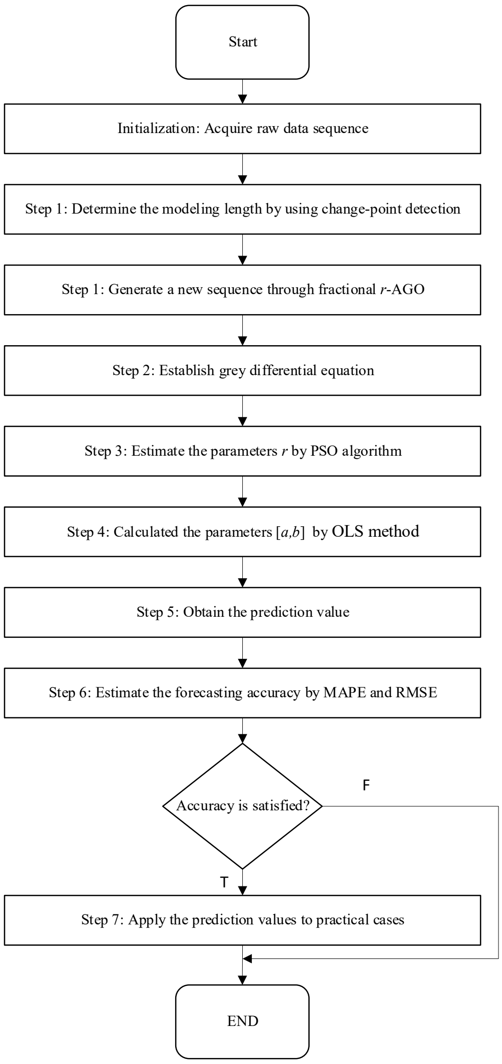

Figure 1. The process of the CPD-FGM(1,1) model.

3.3. Validation of the CPD-FGM(1,1)

We used the mean absolute percentage error (MAPE) to evaluate the accuracy of prediction models, since it is a benchmark for model evaluating [26]. MAPE is calculated as

$$MAPE = \frac{1}{n}\sum_{k=1}^{n}\frac{|x_k - \hat{x}_k|}{x_k} \times 100\% \tag{15}$$

The criterion for MAPE is shown in Table 1.

Table 1. The criterion for MAPE.

MAPE (100%)	Accuracy
0–10	High
10–20	Good
20–50	Reasonable
>50	Inaccurate

4. Experimental Research

The primary data employed in this study included the number of both annual Chinese students studying abroad and the returned students. The time period covered the years 2000 to 2019 owing to the availability of data. We exploited the data from 2000 to 2016 for the model fitting and from 2017 to 2019 for the ex-post testing. The dataset was procured from the National Bureau of Statistics of China (http://www.stats.gov.cn/, accessed on 19 May 2022). Figure 2 shows the dataset used in this study. The number of students studying abroad and the returned students have both grown significantly over the past decade.

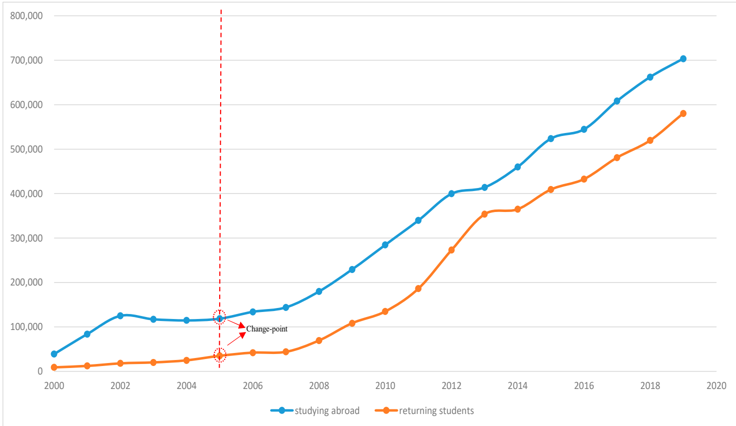

Figure 2. Historical annual number of Chinese students studying abroad and returned students (unit: person).

To prove the superior performance of the proposed CPD-FGM(1,1) model, it was compared with other prediction models, including ARIMA [48], GM(1,1) [49], FGM(1,1) [49], segment FGM(1,1)(SFGM(1,1)) [50] and CPD-GM(1,1).

4.1. Results for the Students Studying Abroad

We used the "EnvCpt" package in R to auto detect the change-point for the number of students studying abroad. The AIC was utilized to select the best fit model, as shown in Figure 3. Thus, "Trend cpt + AR(2)" model was the best fit model, and it located the change-point at the sixth point, as shown in Figure 2. Thus, the first six points (2000–2005) and the last eleven points (2006–2016) of the dataset had different trends, and, as described in Section 3.2, we dropped the data from 2000 to 2005 and selected the data from 2006 to 2016 for CPD-FGM(1,1) modeling. The optimal r = 0.326232 was estimated by PSO algorithm to construct the best fit model given below.

$$\hat{x}_k^{(0.326232)} = (134000 + 1395239.467) \cdot e^{0.0479(k-1)} - 1395239.467$$

Table 2 shows the fitting and forecasting results of the CPD-FGM(1,1) and the other modes for comparison. All of the models achieved high accuracy in terms of MAPE (<10%) for the model fitting. SFGM(1,1) had the best accuracy in terms of MAPE (1.78%), followed by CPD-FGM(1,1) in terms of MAPE (4.76%). For the ex-post testing, CPD-FGM(1,1) and SFGM(1,1) achieved a high accuracy in terms of MAPE, and the former (0.93%) was better than the latter (7.39%). Although the proposed CPD-FGM(1,1) model in this study is not as well fitting as the SFGM(1,1) model, its ex-post testing ability exceeded other models, including the SFGM(1,1) model. In this study, the ex-post testing capability was the primary basis for testing the predictive power of the model. Thus, we concluded that the proposed CPD-FGM(1,1) model was more suitable for forecasting the number of the students studying abroad.

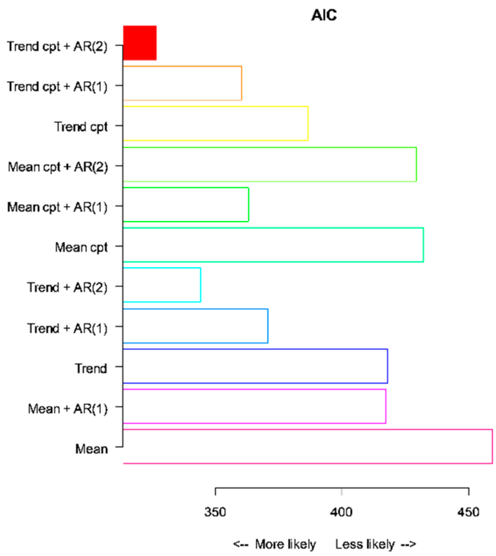

Figure 3. AIC values of different change-point models for the students studying abroad.

Table 2. The results for students studying abroad prediction.

Year	Raw Data	ARIMA	GM(1,1)	FGM(1,1)	SFGM(1,1)	CPD-GM(1,1)	CPD-FGM(1,1)
2000	38,989	38,972	38,989	38,989			
2001	83,973	83,925	84,348	83,973			
2002	125,179	128,957	96,053	95,875			
2003	117,307	166,385	109,382	109,304			
2004	114,682	109,435	124,561	124,547			
2005	118,515	112,057	141,846	141,878			
2006	134,000	122,348	161,530	161,593		134,000	134,000
2007	144,000	149,485	183,945	184,028		190,191	165,328
2008	179,800	154,000	209,471	209,563		215,680	204,846
2009	229,300	215,600	238,538	238,628		244,584	245,289
2010	284,700	278,800	271,640	271,714	284,700	277,362	286,242
2011	339,700	340,100	309,335	309,378	339,700	314,533	327,855
2012	399,600	394,700	352,261	352,255	386,931	356,685	370,340
2013	413,900	459,500	401,143	401,067	430,227	404,486	413,900
2014	459,800	428,200	456,809	456,637	470,588	458,693	458,718
2015	523,700	505,700	520,200	519,901	508,452	520,165	504,962
2016	544,500	587,600	592,387	591,925	544,079	589,874	552,790
MAPE		7.09	9.77	9.77	1.78	8.27	4.76
2017	608,400	565,300	674,591	673,922	577,654	668,926	602,351
2018	662,100	586,100	768,203	767,273	609,323	758,572	653,791
2019	703,500	606,900	874,805	873,550	639,211	860,232	707,252
MAPE		10.76	17.09	16.94	7.39	15.60	0.93

4.2. Results for the Returned Students

The change-point detection was the same as in Section 4.2 for the number of returned students. The best fit model "Trend cpt + AR(2)" was selected by AIC, as shown in Figure 4, and it located the change-point at the sixth point, as shown in Figure 2. Thus, we selected the data from 2006 to 2016 for CPD-FGM(1,1) modeling, and the optimal r = −0.0561162 was estimated by PSO algorithm to construct the best fit model given below.

$$\hat{x}_k^{(-0.0561162)} = (42000 + 922488.9462) \cdot e^{0.030611153(k-1)} - 922488.9462$$

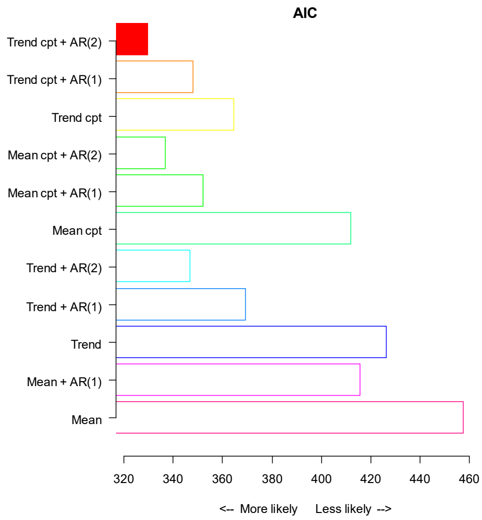

Figure 4. AIC values of different change-point models for the returned students.

Table 3 shows the fitting and ex-post testing results of the CPD-FGM(1,1) and the other modes for comparison. Only the SFGM(1,1) model achieved high accuracy in terms of MAPE (0.52%) for the model fitting. For the ex-post testing, the CPD-FGM(1,1) and ARIMA models achieved high accuracy in terms of MAPE, and the former performed the best in terms of MAPE (0.72%). Thus, we concluded that the proposed CPD-FGM(1,1) model was more suitable for forecasting the number of returned students.

Table 3. Results for returned students prediction.

Year	Raw Data	ARIMA	GM(1,1)	FGM(1,1)	SFGM(1,1)	CPD-GM(1,1)	CPD-FGM(1,1)
2000	9121	9116.921	9121	9121			
2001	12,243	12,248.26	34,537.3	19,429.6			
2002	17,945	15,365	42,485.7	30,473.2			
2003	20,152	23,647	52,263.4	42,085.8			
2004	24,726	22,359	64,291.3	54,181.2			
2005	34,987	29,300	79,087.3	66,706.1			
2006	42,000	45,248	97,288.4	79,624.6		42,000	42,000
2007	44,000	49,013	119,678	92,910.8		97,429.5	74,337.5
2008	69,300	46,000	147,221	106,545		117,575	108,177
2009	108,300	94,600	181,102	120,512		141,885	143,527
2010	134,800	147,300	222,781	134,800		171,222	180,356
2011	186,200	161,300	274,052	149,398		206,625	218,652
2012	272,900	237,600	337,123	164,298		249,348	258,417
2013	353,500	359,600	414,708	179,494	353,500	300,905	299,661
2014	364,800	434,100	510,149	194,978	367,331	363,122	342,406
2015	409,100	376,100	627,554	210,745	404,670	438,203	386,675
2016	432,500	453,400	771,979	226,792	433,790	528,808	432,500
MAPE		11.31	92.51	56.72	0.52	28.49	21.91
2017	480,900	455,900	949,642	243,113	577,654	638,148	479,914
2018	519,400	479,300	1,168,190	259,706	609,323	770,095	528,953
2019	580,300	502,700	1,437,040	276,566	639,211	929,325	579,659
MAPE		8.76	123.34	50.60	15.86	47.04	0.72

4.3. Robustness of the Proposed Model

To illustrate the robustness of the proposed model in this study, we verified the prediction accuracy of the model for different test sets. We took the students studying abroad as an example for proving the robustness of the proposed CPD-FGM(1,1) model in comparison with other prediction models. The results are shown in Figure 4. The horizontal coordinates are the different test sets and the vertical coordinates are MAPE. For example, the CPD-FGM(1,1) model obtained the minimum MAPE of the ex-post testing in the test set from 2017 to 2019, while the ARIMA performed best in the test set from 2015 to 2019 (see Figure 5).

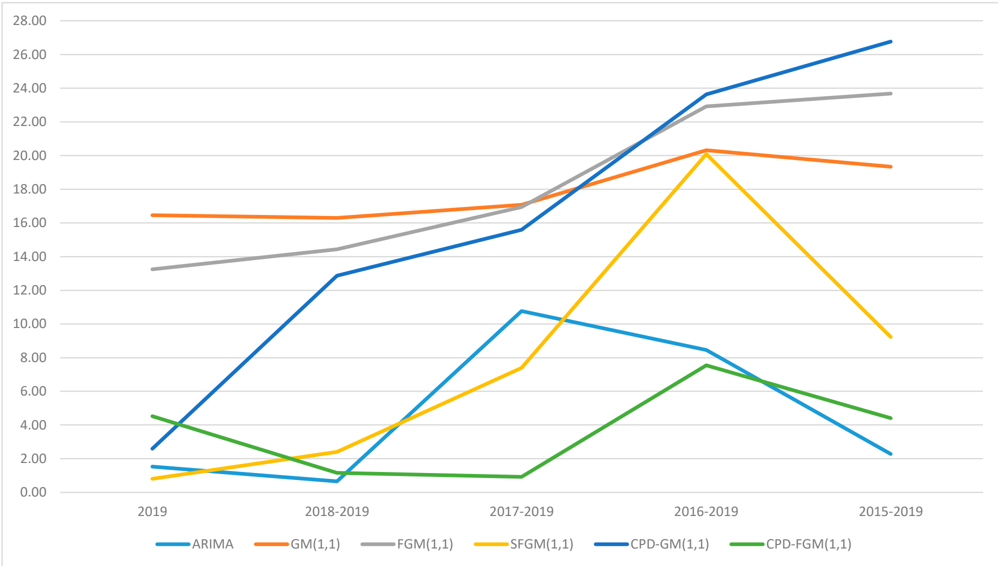

Figure 5. The MAPE of the six comparison models for different test sets.

As shown in Figure 4, the proposed CPD-FGM(1,1) model achieved high prediction accuracy in all test sets and obtained the minimum mean MAPE value. Compared with the grey model, the proposed CPD-FGM(1,1) model obtained the minimum MAPE in all the test sets, except 2019. Compared with the non-grey model, ARIMA, the proposed CPD-FGM(1,1) model obtained the minimum mean MAPE and standard deviation, although it was less than ARIMA in three of all test sets. Therefore, we concluded that the proposed CPD-FGM(1,1) model had strong robustness and high accuracy for ex-post testing.

4.4. Prediction of the Students Studying Abroad and the Returned Students, from 2020 to 2025

We used the proposed CPD-FGM(1,1) model to predict the students studying abroad and returned students from 2020 to 2025. The change-point for the dataset of the students studying abroad was located at the 6th and 11th points, and the dataset for the returned students was located at the 12th, by using the same method as given in Section 4.1. Thus, we employed the data to construct the prediction model from 2011 to 2019 for the students studying abroad, and the data from 2012 to 2019 for the returned students. The results are shown in Table 4. By 2025, the number of the students studying abroad and the returned students will reach 1,236,210 and 1,061,480 in China, respectively. Compared to 2019, the rates of increase were 9.85% and 10.59% for the next 6 years, for the number of the students studying abroad and the returned students, respectively, and the ratio of the latter to the former also showed an increasing trend.

Figure 6 shows that the ratio of the talent returning to China has increased significantly after 2000 and reached a high-point (85%) in 2013, followed by a small decrease. According to our results, this ratio will steadily increase in the future and is expected to reach a new high (86%) in 2025.

The interest of Chinese students in choosing to study abroad is increasing, and, at the same time, a concurrently increasing number of overseas students prefer to return to China. This study proposes several suggestions for the relevant departments to make reference to the planning and management of overseas study abroad. (1) The government should provide a high standard of public services, reduce administrative obstacles and complicated procedures, and increase the convenience for overseas personnel to return to their home country. (2) Provide the students with more favorable talent policies, such as start-up funds and housing subsidies. (3) Continue to deepen the reform and strengthen economic development. The economic factor is still important for attracting talent back to China, and it is possible to attract more talent back by continuing to strengthen economic development.

Table 4. Results of the prediction for students studying abroad and returned students, from 2020 to 2025.

	Students Studying Abroad		Returned Students	
	Raw Data	Predict Values	Raw Data	Predict Values
2011	339,700	339,700		
2012	399,600	399,598	272,900	272,900
2013	413,900	428,068	353,500	353,500
2014	459,800	463,280	364,800	370,529
2015	523,700	503,649	409,100	399,243
2016	544,500	548,924	432,500	434,911
2017	608,400	599,228	480,900	476,556
2018	662,100	654,858	519,400	524,079
2019	703,500	716,212	580,300	577,728
2020		783,768		637,943
2021		858,076		705,300
2022		939,751		780,485
2023		1,029,480		864,291
2024		1,128,020		957,620
2025		1,236,210		1,061,480

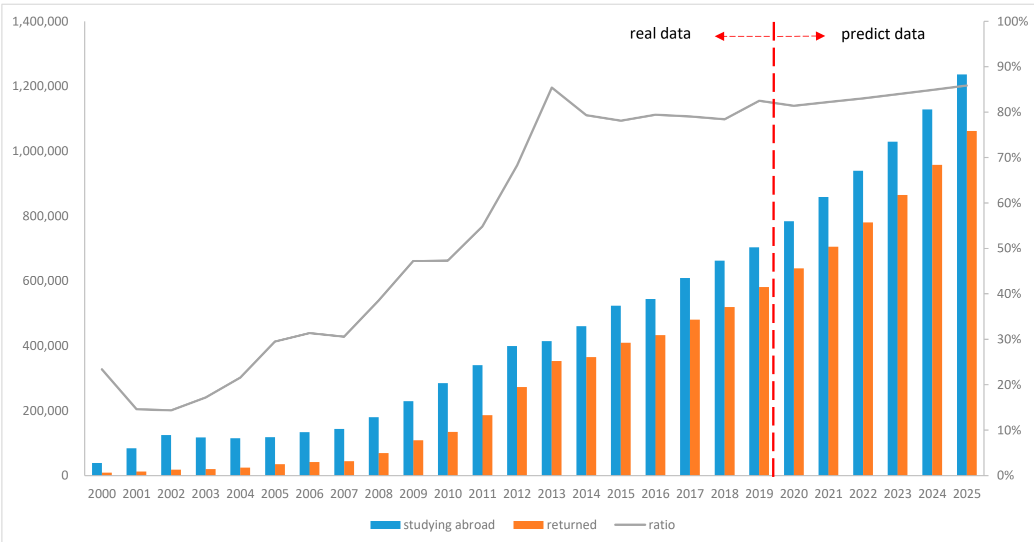

Figure 6. Ratio of returned students to students studying abroad.

5. Conclusions and Future Research

Overseas students are an important high talent resource for China, and their management is an important part of talent management. Accurate prediction of overseas talent mobility is helpful for talent strategy formulation and promotion of the internationalization of the workforce. We propose a fractional grey prediction model with change-point detection to predict the number of students studying abroad and returned students, annually. The fractional grey prediction model is suitable, owing to the limited and unclear influencing factors for the dataset in this study. Change-point detection was employed to determine the modeling length for the proposed prediction model. Considering the final prediction results, the prediction accuracy of the proposed CPD-FGM(1,1) model outperformed alternative models. The results demonstrated the effectiveness of using change-point detection to determine the length of the fractional grey prediction model.

We forecasted the number of students studying abroad and returned students, from 2000 to 2025 using the proposed CDP-FGM(1,1) model. It showed that the number of students studying abroad and returned students will steadily increase and reach one million by 2023 and 2025, respectively. It indicates that the interest of Chinese students for going overseas is increasing, and, with the rapid development of China's economy, increasing overseas students prefer a return to China for work.

Compared with other grey prediction models, the proposed CPD-FGM(1,1) model has the following advantages: (1) It can avoid the influence of old data on new data in the modeling process. (2) It has strong robustness and high prediction accuracy. (3) It is simple and easy to implement. The CPD-FGM(1,1) model still has some limitations, such as the fact it is a univariate prediction model, which cannot fit the influence of exogenous variables, and is not applicable to data with seasonality.

This study restricted itself to exploring the univariate forecasting problem while excluding the influence of exogenous variables. Thus, a multivariate grey forecasting model that can reflect the influence of exogenous variables on the number of students studying abroad and returned students, will be a key research direction in the future. The influence of COVID-19 on the dataset used in this study was not considered in this research. Therefore, we could incorporate the situational prediction method to enhance the credibility of the prediction results in a future work. Furthermore, we will explore the applicability of the CDP-FGM(1,1) model to other areas, such as tourism demand or energy forecasting.

Author Contributions: Conceptualization, G.W. and P.J.; methodology, Y.-C.H. and G.W.; formal analysis, G.W.; data curation, Y.R. and X.Z.; writing—original draft preparation, G.W. and P.J.; writing—review & editing, G.W. and P.J.; funding acquisition, P.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by China Federation of overseas Chinese (19CZQK20), Social Science Foundation of Shandong Province (21DRKJ03), Key R & D projects (Soft science) in Shandong Province (2021RZB05024), Humanities and Social Sciences project of Shandong University (IFWF2029), and the financial support from the Youth Scholars Program of Shandong University, Weihai.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: (http://www.stats.gov.cn/, accessed on 19 May 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lin, D.; Zheng, W.; Lu, J.; Liu, X.; Wright, M. Forgotten or not? home country embeddedness and returnee entrepreneurship. *J. World Bus.* **2019**, *54*, 1–13. [CrossRef]
2. National Bureau of Statistics. *China Statistical Yearbook*; China Statistics Press: Beijing, China, 2020.
3. Dai, O.; Liu, X. Returnee entrepreneurs and firm performance in chinese high-technology industries. *Int. Bus. Rev.* **2009**, *18*, 373–386. [CrossRef]
4. Zhang, C.; Guan, J. Returnee policies in China: Does a strategy of alleviating the financing difficulty of returnee firms promote innovation? *Technol. Forecast. Soc. Chang.* **2021**, *164*, 120509. [CrossRef]
5. Scott, A.J. Jobs or amenities? Destination choices of migrant engineers in the USA*: Migrant engineers. *Pap. Reg. Sci.* **2010**, *89*, 43–63. [CrossRef]
6. Yi, L.; Wang, Y.; Upadhaya, B.; Zhao, S.; Yin, Y. Knowledge spillover, knowledge management capabilities, and innovation among returnee entrepreneurial firms in emerging markets: Does entrepreneurial ecosystem matter? *J. Bus. Res.* **2021**, *130*, 283–294. [CrossRef]
7. Tzeng, C.-H. How foreign knowledge spillovers by returnee managers occur at domestic firms: An institutional theory perspective. *Int. Bus. Rev.* **2018**, *27*, 625–641. [CrossRef]
8. Bai, W.; Johanson, M.; Martín Martín, O. Knowledge and internationalization of returnee entrepreneurial firms. *Int. Bus. Rev.* **2017**, *26*, 652–665. [CrossRef]
9. Lin, Y.-H.; Chen, C.-J.; Lin, B.-W. The dual-edged role of returnee board members in new venture performance. *J. Bus. Res.* **2018**, *90*, 347–358. [CrossRef]
10. Deng, J.-L. Control problems of Grey systems. *Syst. Control Lett.* **1982**, *1*, 288–294. [CrossRef]
11. Dang, Y.G.; Wang, Z.X.; Qian, W.Y.; Xiong, P.P. *Grey Prediction Techniques and Methods*; Science Press: Beijing, China, 2016.
12. Hu, Y.-C. Grey prediction with residual modification using functional-link net and its application to energy demand forecasting. *Kybernetes* **2017**, *46*, 349–363. [CrossRef]

13. Wu, G.; Hu, Y.-C.; Chiu, Y.-J.; Tsao, S.-J. A new multivariate Grey prediction model for forecasting China's regional energy consumption. *Environ. Dev. Sustain.* **2022**, 1–21. [CrossRef] [PubMed]
14. Song, Z.; Feng, W.; Liu, W. Interval prediction of short-term traffic speed with limited data input: Application of Fuzzy-Grey combined prediction model. *Expert Syst. Appl.* **2022**, *187*, 115878. [CrossRef]
15. Liu, Y.; Wu, C.; Wen, J.; Xiao, X.; Chen, Z. A Grey convolutional neural network model for traffic flow prediction under traffic accidents. *Neurocomputing* **2022**, *500*, 761–775. [CrossRef]
16. Chen, C.-I.; Chen, H.L.; Chen, S.-P. Forecasting of foreign exchange rates of Taiwan's major trading partners by novel nonlinear Grey Bernoulli model NGBM(1,1). *Commun. Nonlinear Sci. Numer. Simul.* **2008**, *13*, 1194–1204. [CrossRef]
17. Hu, Y.-C. A Multivariate Grey prediction model with Grey relational analysis for bankruptcy prediction problems. *Soft Comput.* **2020**, *24*, 4259–4268. [CrossRef]
18. Li, B.; Zhang, S.; Li, W.; Zhang, Y. Application progress of Grey model technology in agricultural science. *Grey Syst. Theory Appl.* **2022**. [CrossRef]
19. Jiang, P.; Hu, Y.-C. Constructing interval models using neural networks with non-additive combinations of grey prediction models in tourism demand. *Grey Syst. Theory Appl.* **2022**, *in press*. [CrossRef]
20. Tang, X.; Xie, N.; Hu, A. Forecasting annual foreign tourist arrivals to China by incorporating firefly algorithm into fractional non-homogenous discrete Grey model. *Kybernetes* **2022**, *51*, 676–693. [CrossRef]
21. Ceylan, Z. Short-term prediction of COVID-19 spread using Grey rolling model optimized by particle swarm optimization. *Appl. Soft Comput.* **2021**, *109*, 107592. [CrossRef]
22. Seneviratna, D.M.K.N.; Rathnayaka, R.M.K.T. Hybrid Grey exponential smoothing approach for predicting transmission dynamics of the COVID-19 outbreak in Sri Lanka. *Grey Syst. Theory Appl.* **2022**, *in press*. [CrossRef]
23. Jiang, H.; Kong, P.; Hu, Y.-C.; Jiang, P. Forecasting China's CO_2 emissions by considering interaction of bilateral FDI using the improved Grey multivariable verhulst model. *Environ. Dev. Sustain.* **2021**, *23*, 225–240. [CrossRef]
24. Wang, M.; Wu, L.; Guo, X. Application of Grey model in influencing factors analysis and trend prediction of carbon emission in Shanxi province. *Environ. Monit. Assess.* **2022**, *194*, 542. [CrossRef] [PubMed]
25. Liu, S.; Dang, Y.; Fang, Z.; Xie, N. *Grey System Theory and Application*; Science Press: Beijing, China, 2010.
26. Hu, Y.-C.; Jiang, P.; Tsai, J.-F.; Yu, C.-Y. An optimized fractional Grey prediction model for carbon dioxide emissions forecasting. *Int. J. Environ. Res. Public Health* **2021**, *18*, 587. [CrossRef] [PubMed]
27. Xie, N.; Wang, R. A historic review of Grey forecasting models. *J. Grey Syst.* **2017**, *29*, 1–29.
28. Julong, D. Introduction to Grey system theory. *J. Grey Syst.* **1989**, *1*, 1–24.
29. Akay, D.; Atak, M. Grey prediction with rolling mechanism for electricity demand forecasting of Turkey. *Energy* **2007**, *32*, 1670–1675. [CrossRef]
30. Sun, X.; Sun, W.; Wang, J.; Zhang, Y.; Gao, Y. Using a Grey–Markov model optimized by Cuckoo search algorithm to forecast the annual foreign tourist arrivals to China. *Tour. Manag.* **2016**, *52*, 369–379. [CrossRef]
31. Wu, L.; Liu, S.; Yao, L.; Yan, S.; Liu, D. Grey system model with the fractional order accumulation. *Commun. Nonlinear Sci. Numer. Simul.* **2013**, *18*, 1775–1785. [CrossRef]
32. Chen, Y.; Lifeng, W.; Lianyi, L.; Kai, Z. Fractional hausdorff Grey model and its properties. *Chaos Solitons Fractals* **2020**, *138*, 109915. [CrossRef]
33. Wang, J.; Jiang, H.; Zhou, Q.; Wu, J.; Qin, S. China's natural gas production and consumption analysis based on the multicycle Hubbert model and rolling Grey model. *Renew. Sustain. Energy Rev.* **2016**, *53*, 1149–1167. [CrossRef]
34. Yuan, C.; Zhu, Y.; Chen, D.; Liu, S.; Fang, Z. Using the GM(1,1) model cluster to forecast global oil consumption. *Grey Syst. Theory Appl.* **2017**, *7*, 286–296. [CrossRef]
35. Liu, L.; Wang, Q.; Wang, J.; Liu, M. A rolling Grey model optimized by particle swarm optimization in economic prediction: PSO-RGM in economic prediction. *Comput. Intell.* **2016**, *32*, 391–419. [CrossRef]
36. Brodsky, E.; Darkhovsky, B.S. *Nonparametric Methods in Change Point Problems*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1993; Volume 243.
37. Liu, S.; Lin, Y. *Grey Information: Theory and Practical Applications*; Springer: London, UK, 2006.
38. Aminikhanghahi, S.; Cook, D.J. A survey of methods for time series change point detection. *Knowl. Inf. Syst.* **2017**, *51*, 339–367. [CrossRef]
39. Beaulieu, C.; Killick, R. Distinguishing trends and shifts from memory in climate data. *J. Clim.* **2018**, *31*, 9519–9543. [CrossRef]
40. You, S.-H.; Jang, E.J.; Kim, M.-S.; Lee, M.-T.; Kang, Y.-J.; Lee, J.-E.; Eom, J.-H.; Jung, S.-Y. Change point analysis for detecting vaccine safety signals. *Vaccines* **2021**, *9*, 206. [CrossRef] [PubMed]
41. Yang, M.; Han, C.; Cui, Y.; Zhao, Y. COVID-19 and mobility in tourism cities: A statistical change-point detection approach. *J. Hosp. Tour. Manag.* **2021**, *47*, 256–261. [CrossRef]
42. Killick, R.; Fearnhead, P.; Eckley, I.A. Optimal detection of changepoints with a linear computational cost. *J. Am. Stat. Assoc.* **2012**, *107*, 1590–1598. [CrossRef]
43. Killick, R.; Beaulieu, C.; Taylor, S.; Hullait, H. EnvCpt: Detection of Structural Changes in Climate and Environment Time Series. Available online: https://CRAN.R-project.org/package=EnvCpt (accessed on 2 July 2022).
44. Wu, L.; Gao, X.; Xiao, Y.; Yang, Y.; Chen, X. Using a novel multi-variable Grey model to forecast the electricity consumption of Shandong province in China. *Energy* **2018**, *157*, 327–335. [CrossRef]

45. Ding, S.; Li, R. A new multivariable Grey convolution model based on Simpson's rule and its applications. *Complexity* **2020**, *2020*, 1–14. [CrossRef]
46. Ding, S.; Li, R.; Wu, S. A novel composite forecasting framework by adaptive data preprocessing and optimized nonlinear Grey Bernoulli model for new energy vehicles sales. *Commun. Nonlinear Sci. Numer. Simul.* **2021**, *99*, 105847. [CrossRef]
47. Faris, H.; Aljarah, I.; Mirjalili, S.; Castillo, P.A.; Merelo, J.J. EvoloPy: An open-source nature-inspired optimization framework in Python. In Proceedings of the the 8th International Joint Conference on Computational Intelligence, Porto, Portugal, 9–11 November 2006; SCITEPRESS-Science and Technology Publications: Porto, Portugal, 2016; pp. 171–177.
48. Yuan, C.; Liu, S.; Fang, Z. Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model. *Energy* **2016**, *100*, 384–390. [CrossRef]
49. Wu, L.; Li, N.; Yang, Y. Prediction of air quality indicators for the Beijing-Tianjin-Hebei region. *J. Clean. Prod.* **2018**, *196*, 682–687. [CrossRef]
50. Hu, Y.-C.; Wu, G.; Jiang, P. Tourism demand forecasting using nonadditive forecast combinations. *J. Hosp. Tour. Res.* **2021**, 109634802110478. [CrossRef]