

# Modelling Natural Gas Energy Production of United States

Yubin Cai<sup>1,\*</sup>, Yanqiao Deng<sup>1</sup>

<sup>1</sup> School of Science, Southwest University of Science and Technology, Mianyang 621000, China.

**Abstract.** In the transformation of the energy system, natural gas energy is regarded as a buffer energy. How to make a reasonable energy distribution and effectively predict its production is very significant. In the work of this paper, a grid-optimized fractional-order non-homogeneous grey model is used to predict the natural gas energy production in the United States and obtain reliable results. This paper first introduces the prediction method and prediction mechanism. Then the model is optimized to make the prediction effect more prominent. The natural gas energy prediction results show that this method has high prediction accuracy compared with other methods, which means that the method proposed in this paper can be used as an effective tool for short-term forecasting of natural gas production in the United States and play an auxiliary role in energy forecasting.

## 1 Introduction

Energy has become a problem that every country must pay attention to. Global energy demand is expected to increase by 35% in 2035[1]. With the popularization of low-carbon policies, the gradual transition of world energy from a high-carbon to a low-carbon system will reduce the burden of energy scarcity to a certain extent. Among them, natural gas has always been a very important energy source in the world. As a relatively clean fossil fuel, it not only occupies an important position in power generation, but is also a raw material for the manufacture of various chemical products[2]. The BP Statistical Review of World Energy pointed out that natural gas resources are likely to become the second largest energy source in the world[3]. Therefore, studying the natural gas energy production of the world's industrial power, the United States, has certain significance for energy development.

Among many research methods, the grey system theory proposed by Professor Deng stands out for its advantages of small sample size, small calculation workload and high prediction accuracy[4]. In the development of the grey model, Cui et al. constructed a new model to improve the prediction accuracy of the data sequence that approximates the characteristics of non-homogeneous exponential law[5]. Xie et al. obtained two new models by discretizing the traditional grey model and the non-homogeneous grey model[6]. In terms of the accumulative generation sequence of the grey model, Wu et al. proposed a fractional operator to improve the method of data accumulation[7], which provides a basis for the subsequent research on the fractional model of many scholars[8-9]. In terms of natural gas energy prediction, Li et al. used particle swarm optimization to optimize the non-homogeneous discrete model and achieved good results[10], and Zheng

et al. verified the optimization of the fractional Bernoulli model with the moth flame optimization (MFO) algorithm based on the production and consumption of natural gas[11].

The rest of this article is organized as follows. Section 2 introduces the non-homogeneous grey model, fractional accumulation operator and one-step rolling forecast mechanism. Section 3 introduces the method of optimizing nonlinear parameters. In Section 4, the proposed method is verified through the case of natural gas production(dry) in the United States and the Section 5 gives the conclusion.

## 2 The non-homogeneous grey model

The raw data sequence is  $Y^{(0)} = \{y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n)\}$ , and its first-order accumulation generating sequence is  $Y^{(1)} = \{y^{(1)}(1), y^{(1)}(2), \dots, y^{(1)}(n)\}$ , which  $y^{(1)}(k) = \sum_{i=1}^k y^{(0)}(i)$ .

And the NGM's first-order differential equation is:

$$\frac{dy^{(1)}(t)}{dt} + ay^{(1)}(t) = bt + c \tag{1}$$

where  $(a, b, c)^T$  is a grey parameter of the NGM.

The discrete differential equation of Eq 1 is:

$$y^{(0)}(t) + az^{(1)}(t) = bt + c \tag{2}$$

where  $z^{(1)}(t) = \frac{y^{(1)}(t) + y^{(1)}(t+1)}{2}$  is the sequence mean generated of consecutive neighbors of  $y^{(1)}(k)$ . Make:

\* Corresponding author: caiyubin4@163.com

$$\zeta = \begin{pmatrix} y^{(0)}(2) \\ y^{(0)}(3) \\ \vdots \\ y^{(0)}(n) \end{pmatrix}, \Phi = \begin{pmatrix} -z^{(1)}(2) & 2 & 1 \\ -z^{(1)}(3) & 3 & 1 \\ \vdots & \vdots & \vdots \\ -z^{(1)}(n) & n & 1 \end{pmatrix} \quad (3)$$

The grey parameter of the NGM satisfies:

$$\begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{pmatrix} = (\Phi^T \Phi)^{-1} \Phi^T \zeta \quad (4)$$

The solution of the first-order differential equation Eq 1 is:

$$\hat{y}^{(1)}(k) = \left( y^{(0)}(1) - \frac{\hat{b}}{\hat{a}} - \frac{\hat{c}}{\hat{a}^2} \right) e^{-\hat{a}(k-1)} + \frac{\hat{b}}{\hat{a}} k + \frac{\hat{c}}{\hat{a}^2}, k=1,2,\dots,n \quad (5)$$

Then the forecast results  $\hat{Y}^{(0)} = \{\hat{y}^{(0)}(1), \hat{y}^{(0)}(2), \dots, \hat{y}^{(0)}(n)\}$  of the NGM is:

$$\hat{y}^{(0)}(k) = \hat{y}^{(1)}(k) - \hat{y}^{(1)}(k-1), k = 2, 3, \dots, n \quad (6)$$

### 2.1 Fractional-order accumulation method

Change the accumulation method of the original sequence  $Y^{(0)} = \{y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n)\}$ , its fractional-order accumulation is  $Y^{(r)} = \{y^{(r)}(1), y^{(r)}(2), \dots, y^{(r)}(n)\}$ ,  $r$  is the fractional-order, which:

$$y^{(r)}(k) = \sum_{i=1}^k \frac{r(r+1) \cdots (r+k-i-1)}{(k-i)!} y^{(0)}(i), k=1,2,\dots,n \quad (7)$$

when  $r=1$ , the fractional-order accumulation degenerates to the one-order accumulation.

For the fractional-order accumulation operations, it can express:

$$y^{(0)}(k) = \sum_{i=1}^k \frac{-r(-r+1) \cdots (-r+i-1)}{(k-i)!} y^{(r)}(i), k=1,2,\dots,n \quad (8)$$

This paper used the fractional-order accumulation method to improve the prediction accuracy of the NGM.

### 2.2 Forecasting method

his paper uses a new time series forecasting method to improve the accuracy of the prediction, for a set of time series  $Y^{(0)} = \{y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n)\}$ , set the number of windows to  $\rho$ , divide  $Y^{(0)}$  into group  $n-\rho$  data  $\{Y_1^{(0)}, Y_2^{(0)}, \dots, Y_{n-\rho}^{(0)}\}$ , which:

$$\begin{cases} Y_1^{(0)} = \{y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(\rho)\} \\ Y_2^{(0)} = \{y^{(0)}(2), y^{(0)}(3), \dots, y^{(0)}(\rho+1)\} \\ \vdots \\ Y_{n-\rho+1}^{(0)} = \{y^{(0)}(n-\rho), y^{(0)}(2), \dots, y^{(0)}(n-1)\} \end{cases}, \rho < n \quad (9)$$

Modelling each set of data separately and predicting one-step backward, we use this  $n-\rho$  one-step forecast result as the one-step rolling forecast result of this time series  $\{\hat{y}_{one-step}^{(0)}(\rho+1), \hat{y}_{one-step}^{(0)}(\rho+2), \dots, \hat{y}_{one-step}^{(0)}(n)\}$ . The one-step rolling forecast mechanism can be represented in Fig 1.

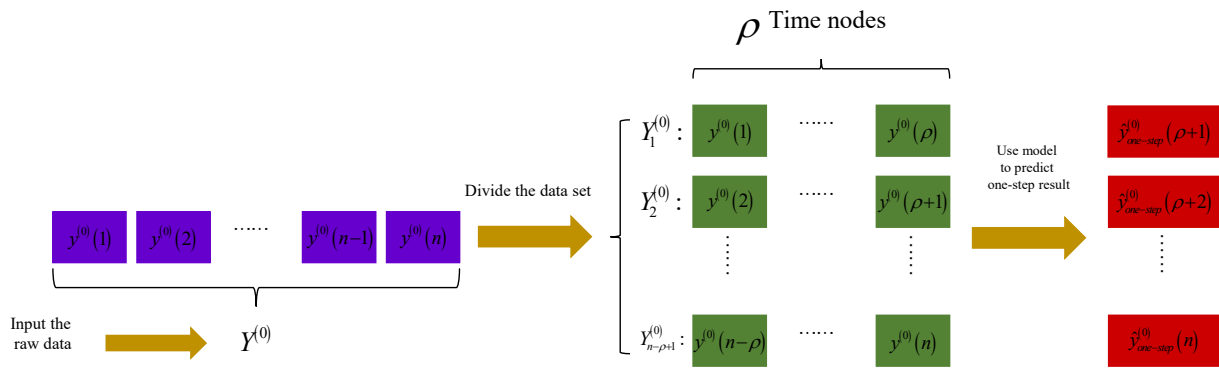
### 3 Parameter optimization based on grid search

After optimizing the NGM by using the fractional accumulation operation, how to determine the value of the fractional parameters  $r$  becomes the key to optimization. Searching in the parameter grid at intermediate intervals can ensure that the obtained parameters are valid. When the parameter grid is divided into small enough, the optimal approximate parameters in the grid can be obtained, and then the searched parameters are used for modeling prediction.

To adjust the parameters during the modeling process, the data set needs to be divided into a training set and a validation set. Taking  $Y_1^{(0)} = \{y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(\rho)\}$  as an example,  $Y_1^{(0)}$  data set is divided into two parts: training set and validation set, marked as  $Y_{1,train}^{(0)} = \{y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(v)\}$  and  $Y_{1,valid}^{(0)} = \{y^{(0)}(v+1), y^{(0)}(v+2), \dots, y^{(0)}(\rho)\}$ .  $Y_{1,train}^{(0)}$  is used to initialize the prediction model,  $Y_{1,valid}^{(0)}$  is used to verify whether the searched parameter is the optimal approximate parameter in the parameter grid, that is,  $\hat{Y}_{1,valid}^{(0)} = \{\hat{y}^{(0)}(v+1), \hat{y}^{(0)}(v+2), \dots, \hat{y}^{(0)}(\rho)\}$  predicted by parameter  $r$  is used to verify whether the optimal search condition is satisfied with this equation:

$$\min \theta = \sum_{i=1}^{\rho-v} [y^{(0)}(v+i) - \hat{y}^{(0)}(v+i)]^2 \quad (10)$$

The search algorithm in this section is summarized as Algorithm 1.



**Figure 1** The one-step rolling forecast mechanism

**Algorithm 1**

- 1 Input: the original data sequence  $Y^{(0)}$ , forecasting model  $mdl$ , parameters  $r$  to be optimized, parametric grid interval  $S$ , the number of windows  $\rho$
- 2 Divide  $Y^{(0)}$  into  $n - \rho$  parts:  $\{Y_1^{(0)}, Y_2^{(0)}, \dots, Y_{n-\rho}^{(0)}\}$
- 3 for  $i=1; i \leq n - \rho + 1; i = i + 1$  do:
- 4 For  $Y_i^{(0)}$ , divide  $Y_i^{(0)}$  into two parts: training set  $Y_{i,train}^{(0)}$  and validation set  $Y_{i,valid}^{(0)}$
- 5 for  $j$  in  $S$  do:
- 6  $r_i = j$
- 7 Substitute the parameter values  $r_i$  at this time into the model  $mdl$  for modelling
- 8 Use the model  $mdl$  to forecast the validation set to get the forecast value  $\hat{Y}_{i,valid}^{(0)}$
- 9 Calculate  $\theta$  of the forecast result  $\hat{Y}_{i,valid}^{(0)}$  and validation set  $Y_{i,valid}^{(0)}$
- 10 end
- 11 Take the value of  $r_i$  when  $\hat{\theta}$  is minimum as the optimal parameter
- 12 end
- 13 Get the fitted model group:  $\{mdl_1, mdl_2, \dots, mdl_{n-\rho}\}$
- 14 Use  $\{mdl_1, mdl_2, \dots, mdl_{n-\rho}\}$  to predict backwards by one-step to form the forecast result  $\{\hat{y}_{one-step}^{(0)}(\rho+1), \hat{y}_{one-step}^{(0)}(\rho+2), \dots, \hat{y}_{one-step}^{(0)}(n)\}$
- 15 return the one-step forecast result  $\{\hat{y}_{one-step}^{(0)}(\rho+1), \hat{y}_{one-step}^{(0)}(\rho+2), \dots, \hat{y}_{one-step}^{(0)}(n)\}$

**4 Case study**

In this section, we use monthly natural gas production (dry) of U.S. (from Jan-2017 to Aug-2020) to verify the above prediction method. In this case, to highlight the performance advantages of FNGM, we compare and analyze GM, DGM and their fractional accumulation form models FGM, FDGM with FNGM. In the first part of this section, four evaluation indicators are used to quantify the predictive performance of each model. The

forecast results are analyzed in the last part of this section.

**4.1 Model evaluation metrics**

We use evaluation metrics to quantify the predictive performance of the model. For a data set  $Y^{(0)} = \{y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n)\}$ , the predicted set marked as  $\hat{Y}^{(0)} = \{\hat{y}^{(0)}(1), \hat{y}^{(0)}(2), \dots, \hat{y}^{(0)}(n)\}$  is obtained by the model prediction, which is used to

reflect the predictive ability of the model compared to  $Y^{(0)}$ . The definitions of two evaluation metrics used in this paper are introduced in Table 1 [12].

**Table 1** The definitions of evaluation metrics

Metrics	Expression
Mean absolute percentage error(MAPE)	$MAPE = \frac{1}{n} \sum_{i=1}^n \left  \frac{y^{(0)}(i) - \hat{y}^{(0)}(i)}{y^{(0)}(i)} \right  \times 100\%$
Root mean squared percentage error(RMSPE)	$RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{y^{(0)}(i) - \hat{y}^{(0)}(i)}{y^{(0)}(i)} \right)^2} \times 100\%$

### 4.2 Forecast result analysis

The monthly natural gas production (dry) of U.S. is used to make predictions, which has 44 times nodes and 12 designated windows marked as  $\rho = 12$ . The predicted

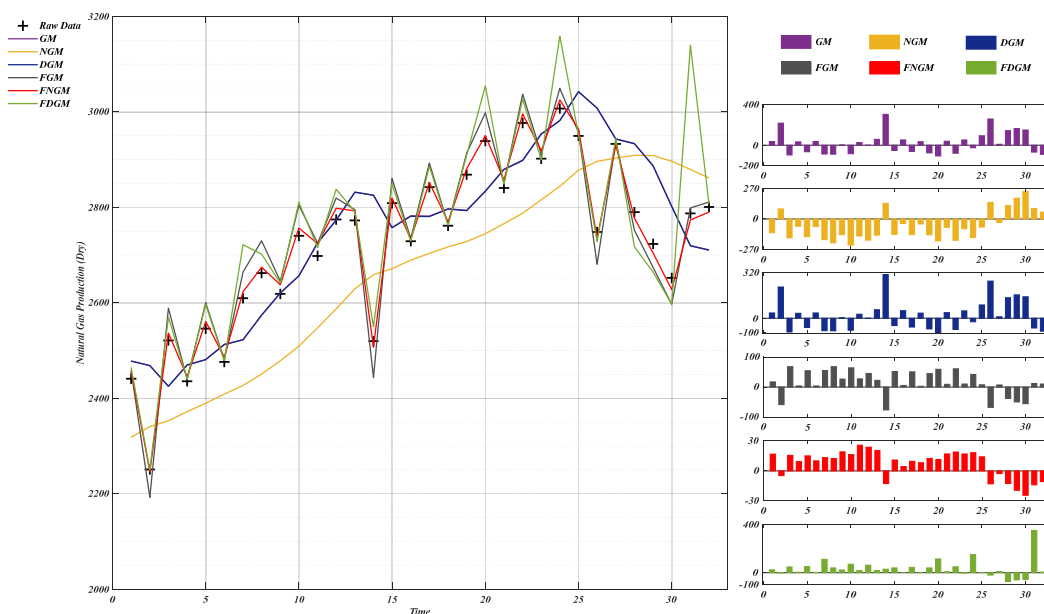
value of a total of 32 times nodes (Jan-2018 to Aug-2020) are obtained by the one-step rolling prediction method, and the prediction results and evaluation metrics values of each model are shown in Table 2. Comparing the values of the evaluation metrics, the MAPE and RMSPE calculated using the prediction results of FNGM are 0.52% and 0.56%, respectively. Therefore, the FNGM has better predictive performance.

To analyze the prediction performance of each model more intuitively, we made a comparison chart of the prediction results of the models in Figure 2, which shows that FGM and FNGM have relatively better prediction performance among the six prediction models. Further analyzing the error analysis chart on the right, the FGM error value is within the interval  $[-100, 100]$ , and most of the error values exceed 50, while the error interval of FNGM is  $[-30, 30]$ , which is much smaller than that of FGM, and the error interval of the other four prediction models is wider. Therefore, the prediction performance of FNGM is superior.

**Table 2** The forecast results and evaluation metrics of the GM, NGM, DGM, FGM, FNGM and FDGM

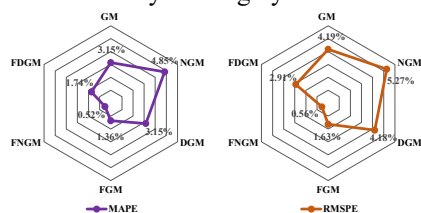
Date	Raw Data	GM	NGM	DGM	FGM	FNGM	FDGM
Jan-18	2441.06	2478.36	2318.13	2477.93	2457.98	2457.85	2464.96
Feb-18	2250.91	2468.57	2340.64	2468.40	2192.39	2246.20	2250.68
Mar-18	2521.15	2425.54	2352.92	2425.00	2588.89	2536.58	2570.24
Apr-18	2435.68	2470.14	2372.19	2469.54	2439.22	2444.98	2440.48
May-18	2545.81	2481.39	2389.52	2480.79	2600.19	2560.73	2596.45
Jun-18	2476.15	2513.00	2408.94	2512.41	2480.13	2486.27	2478.96
Jul-18	2609.65	2522.92	2426.76	2522.32	2664.30	2622.92	2721.73
Aug-18	2662.29	2574.45	2450.88	2573.79	2729.91	2674.59	2702.28
Sep-18	2618.67	2620.95	2478.68	2620.26	2645.48	2637.69	2641.37
Oct-18	2740.27	2656.94	2508.90	2656.30	2804.33	2756.51	2810.76
Nov-18	2698.18	2726.10	2547.50	2725.45	2725.08	2723.95	2715.73
Dec-18	2774.45	2773.43	2587.14	2772.92	2819.56	2798.03	2838.12
Jan-19	2772.56	2832.06	2629.33	2831.67	2794.94	2792.86	2791.17
Feb-19	2519.76	2825.44	2658.97	2825.26	2443.36	2507.08	2549.97
Mar-19	2808.73	2757.97	2671.74	2757.09	2860.43	2819.50	2849.17
Apr-19	2729.03	2782.21	2689.39	2781.39	2734.54	2733.39	2729.08
May-19	2842.30	2781.59	2703.60	2780.76	2893.10	2851.94	2886.21
Jun-19	2761.43	2797.11	2717.30	2796.41	2764.31	2769.58	2762.34
Jul-19	2868.63	2793.60	2728.33	2792.90	2913.39	2881.02	2909.49
Aug-19	2938.97	2834.40	2744.86	2833.67	2997.96	2950.35	3054.44
Sep-19	2840.56	2879.83	2765.84	2879.04	2849.83	2857.51	2848.80
Oct-19	2976.72	2898.88	2787.68	2898.13	3037.24	2995.57	3027.01

Nov-19	2901.94	2953.92	2815.70	2953.15	2912.26	2918.85	2899.69
Dec-19	3006.93	2982.25	2844.03	2981.58	3049.23	3025.04	3158.43
Jan-20	2949.60	3042.59	2877.85	3042.06	2956.97	2963.59	2951.99
Feb-20	2748.37	3007.81	2896.31	3007.56	2680.81	2735.41	2728.72
Mar-20	2932.99	2943.53	2903.30	2942.84	2939.87	2930.24	2943.68
Apr-20	2790.01	2933.72	2909.08	2933.12	2752.23	2777.54	2717.51
May-20	2723.38	2887.41	2908.50	2886.68	2673.88	2704.04	2664.42
Jun-20	2652.06	2802.44	2896.74	2801.80	2596.90	2627.72	2597.32
Jul-20	2787.35	2719.86	2879.21	2719.20	2798.82	2773.48	3139.58
Aug-20	2800.76	2710.83	2861.82	2710.15	2811.02	2790.08	2806.74
MAPE		3.15%	4.85%	3.15%	1.36%	<b>0.52%</b>	1.74%
RMSPE		4.19%	5.27%	4.18%	1.63%	<b>0.56%</b>	2.91%



**Figure 2** The forecast results of each models and their error comparison

We can clearly see the comparison of the predictive capabilities of the six models through Figure 3 formed from the calculated evaluation metrics value. The FNGM takes advantage of the forecast of natural gas production of U.S., indicating that the one-step rolling forecast method and the method of evaluating fractional parameters by grid search are very reliable for improving the prediction accuracy of the grey model.



**Figure 3** The MAPE and RMSPE comparison of the GM, NGM, DGM, FGM, FNGM and FDGM

## 5 Conclusion

This paper uses the mechanism of one-step rolling prediction to model and predict the grey model. Then we use grid search to evaluate the parameter values to obtain the optimal approximate parameters for fractional accumulation models with parameters in the model. In the first part, the theory of NGM, the concept of fractional accumulation operator and the one-step rolling prediction mechanism are briefly described. The following paper describes how to set the optimal approximation of the parameters. Finally, it was applied to the prediction of natural gas production (dry) in the United States. Under the same prediction mechanism, the FNGM with

fractional-order parameters showed superior prediction performance than the other five models. The prediction results show that the method proposed in this paper can be used as an effective tool for energy prediction.

## Acknowledgments

This research was supported by the Talent cultivation and subject competition innovation fund project of the School of Science, Southwest University of Science and Technology(No. LX2020002).

## References

1. Simões-Filho I. BP Energy Outlook: 2017 Edition[C]//New Energy Landscape: Impacts for Latin America, 6th ELAEE/IAEE Latin American Conference, April 2-5, 2017. International Association for Energy Economics, 2017.
2. U.S. Energy Information Administration, 2019a. Electric power monthly [Online]. Available. <https://www.eia.gov/electricity/monthly/> [dataset].
3. BP (British Petroleum). BP statistical review of world energy 2019[J]. Available online at, 2019.
4. Cui J, Dang Y G, Liu S F. Novel grey forecasting model and its modeling mechanism[J]. Control and Decision, 2009, 24(11): 1702-1706.
5. Xie N, Wang R, Chen N. Measurement of shock effect following change of one-child policy based on grey forecasting approach[J]. Kybernetes, 2018.
6. Xie N, Liu S. Interval grey number sequence prediction by using non-homogenous exponential discrete grey forecasting model[J]. Journal of Systems Engineering and Electronics, 2015, 26(1): 96-102.
7. Wu L, Liu S, Yao L, et al. Grey system model with the fractional order accumulation[J]. Communications in Nonlinear Science and Numerical Simulation, 2013, 18(7): 1775-1785.
8. Wu W, Ma X, Zeng B, et al. Application of the novel fractional grey model FAGMO (1, 1, k) to predict China's nuclear energy consumption[J]. Energy, 2018, 165: 223-234.
9. Mao S, Kang Y, Zhang Y, et al. Fractional grey model based on non-singular exponential kernel and its application in the prediction of electronic waste precious metal content[J]. ISA transactions, 2020, 107: 12-26.
10. Li N, Wang J, Wu L, et al. Predicting monthly natural gas production in China using a novel grey seasonal model with particle swarm optimization[J]. Energy, 2021, 215: 119118.
11. Zheng C, Wu W Z, Xie W, et al. A MFO-based conformable fractional nonhomogeneous grey Bernoulli model for natural gas production and consumption forecasting[J]. Applied Soft Computing, 2021, 99: 106891.
12. Xiang, X., Ma, X., Ma, M., Wu, W., & Yu, L. "Research and application of novel Euler polynomial-driven grey model for short-term PM10 forecasting." Grey Systems: Theory and Application (2020).