



OPEN

A clustering fractional-order grey model in short-term electrical load forecasting

Xiang Yu^{1,4}✉, Lihua Lu^{1,4}✉, Jianming Qi^{1,4}, Yuchen Qian^{1,4}, Lisen Zhao^{1,4}, Chang Tan^{1,4}, Yangquan Chen^{2,4} & Zhigang Han^{3,4}

Short-term electrical load series forecast plays an essential role in energy demand management, however power consumption data are non-stationary, nonlinear and multi-dimensional series, leaving prediction a difficult task. Recently, fractional-order partial differential equations are attracting attention as they have been successfully utilized to describe power consumption behaviors in complex electrical systems and power grids. In this paper, a clustering fractional order predictive model called C-FGM is introduced for short-term electrical load forecast missions. The novelty of the C-FGM is that it initiates a parameter α to describe the accumulative weather trends of multiple clustering sub-series, and this parameter is also assigned to a fractional-order partial differential equation to depict the previous power series. Hyper parameters of these equations are then sent to a global optimization algorithm to reduce predictive errors. Simulation results on two electricity datasets demonstrated that our algorithm can learn from datasets hyper parameters inside equations and produce forecast values efficiently. Compared with contemporary models such as LSTM and the Transformer, C-FGM clearly achieved a higher accuracy (MAPE from 1.97 to 4.67%, outperforms LSTM whose average MAPE is 4.34% and Transformer whose average MAPE is 5.42%). This satisfactory performance suggests that our data-driven model can be used as an effective tool for real time forecasting missions.

Keywords Fractional-order differential equations, Data driven modeling, Short-term electrical load forecast, Grey system

Recently, energy demand management has become a focus in government planning due to the inadequacy of energy resources, exponential growth in energy supply demand, and scarcity of sustainable and clean energy supplies. Short-term electrical load series forecast¹ is a task in daily management of electric equipment, and it generates power demand data in the future based on limited number of historical records. An accurate and reliable short-term load forecaster is necessary to support significant decisions in large energy management systems such as electricity generation and scheduling, electricity purchase, power grid management, etc. Many administrations and agencies built their data portal^{2,3} for researchers, and they are looking forward for an authentic and stable electric load forecaster inside their systems. However, it is challenging to build an accurate forecaster with only day-ahead power meter readings. The non-static and nonlinear properties of these power data collections impede researchers from developing an accurate forecasting model. In addition, exogenous factors such like weather, season and social activities may affect the performance of our model as well. Thus, a desired forecasting model should not only consider the nonlinear characteristics of incoming data series, but also deal with the environment efforts such as temperature, season period to sufficiently address its capability in modern power grid applications.

Short-term electrical load series forecast models can be categorized broadly into three groups from the techniques they used for prediction: (i) statistic models; (ii) deep learning models; and (iii) hybrid models. Statistic models such as the famous autoregressive integrated moving average (ARIMA)⁴ and LSTM⁵ are suitable for prediction of time series with linear characteristics. The second group of models emerged rapidly within the past decades due to a revolutionary architecture called the Transformer⁶. Equipped with ground-breaking techniques such as embedding and self-attention, these models have shown great capabilities in many fields such

¹School of Electronics and Information, Shanghai Dianji University, Shuihua Road, Shanghai 201306, China.

²Department, University of California, Merced, CA 95301, USA. ³College of Geography and Environmental Science, Henan University, Kaifeng 475001, Henan, China. ⁴These authors contributed equally: Xiang Yu, Lihua Lu, Jianming Qi, Yuchen Qian, Lisen Zhao, Chang Tan, Yangquan Chen and Zhigang Han. ✉email: yux@sdju.edu.cn; luhua@sdju.edu.cn

as electricity forecast, traffic flow prediction and weather modeling. This category of models are well developed for multi-dimensional long range data predictions. For example, the ETTh1 data set has six dimensions, each representing data from an individual sensor. Scientists designed a novel architecture called Informer⁷ to analyze this data set. Informer achieved a sound performance on ETTh1 which exceeded most of contemporary models at that time. The third group, the hybrid models, combine the advantage of several techniques to obtain results with higher accuracy. Such type of models include GRU⁸ and LightGBM⁹. These models clustered data samples into different groups, and apply algorithms with different configurations on these groups to obtain optimum results. These models have been proved to have higher accuracy in time-series data leader boards such as M4 Competition¹⁰ and their ideas of clustering and optimum training has been utilized in this manuscript as well.

Among existing methods for short-term electrical load series forecast, fractional-order grey models (FGM)^{11–15} have regained our interests due to their capability to provide highly accurate predictions and their hybrid nature to work with other models. For example, Yin¹⁶ introduced a multivariate NGBM to solve short-term power load problems. In literature^{17–19} the authors constructed a grey system with fractional order accumulation efforts. By integrating the concept of fractional order calculus into a grey system, this model addressed issues of uncertainty and partial information together. Another fractional neural grey system model was introduced in²⁰, where it combines neural networks with fractional order grey system to enhance the accuracy of predictions and system analysis. By using the particle swarm optimization algorithm to determine the optimal fractional order, this model has demonstrated high accuracy in forecasting industrial power consumption. One more approach^{21,22} to fractional grey system realization is introducing non-singular exponential kernel. Based on a new fractional order accumulation generator, this model estimates parameters through the least squares method and uses meta-heuristic algorithms to determine the model's order. Further interesting application is an augmented fractional accumulation grey model²³. This model improves upon traditional grey models through structural expansion, parameter optimization, and a rolling mechanism. It has been applied to forecast and analyze the trends in industrial power consumption and compared with seven other benchmark prediction models, demonstrating the best performance in simulation and prediction metrics. To summarize, FGMs^{24–27} have demonstrated their capability to provide highly accurate predictive outputs and they are suitable to be placed in a hybrid electric forecasting model as well.

In this article, the authors design and build a novel channel-wise fractional-order grey model named as Clustering Fractional-Grey-Model (C-FGM). A hypothesis was firstly made that a power load pattern, along with its temperature attributes, can be sufficiently explained and simulated with a series of FGMs. With properly set parameters such as fractional order α and nonlinear term β , FGMs can fit electricity meter readings accurately. A number of FGM modules should learn from different portions of existing datasets and generate predictions. The learning process of FGM modules referred to the accumulated temperature lagging effect which was proposed by Li and other researchers^{4,9}. Our algorithm clustered load data firstly from day- before accumulated weather statistics. Unlike traditional approaches which correlated data from two individual dimensions (weather plane and load plane), our clustering technique was able to divide data into respective time series while keep their coherence on both planes.

The main contributions are:

1. By clustering historical load data into different groups according to their day- ahead accumulated weather deviation level, significantly achieved a higher level of model accuracy (channel wise MAPEs ranging from 1.97 to 4.67%, outperforms LSTM whose average MAPE is 4.34% and Transformer whose average MAPE is 5.42%);
2. By Fitting real consumption data series with fractional-order equations, showing statistic distributions of hyper parameters, and simplifying optimization process;
3. Putting forward a data-driven hybrid time-series model in energy demand management.

Rest of this manuscript is organized as follows. "Methods" section discussed necessary mathematical modeling, equations and algorithms implemented in our model. "Experimental setup" section depicted applied data sets, simulation settings, software and scientific packages utilized by C-FGM algorithm. "Results and discussion" section analyzed simulation results on four electricity datasets for discussion. "Conclusion and future work" section summarized the work and brought forward further discussions.

Methods

C-FGM overview

In this section we present a flow chart of C-FGM. Our model consists of three sub sections: data clustering, fractional-order modeling and model optimization. Its system diagram is shown in Fig. 1. In data clustering phase, power consumption data are firstly clustered according to their accumulated temperature records. Then each clustered data arrays are fitted with a corresponding fractional order differential equations. Coefficients inside these equations are calculated and updated in optimization stage. Once predictions are within boundaries (in our algorithm, error rate less than 5 percent), C-FGM terminates and prints out the forecasted values.

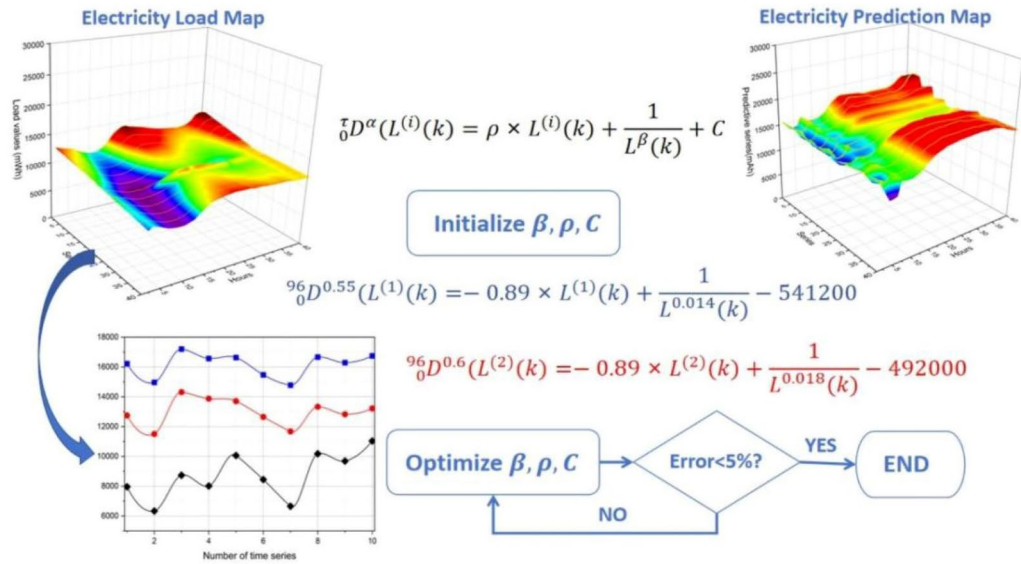


Fig. 1. C-FGM process overview.

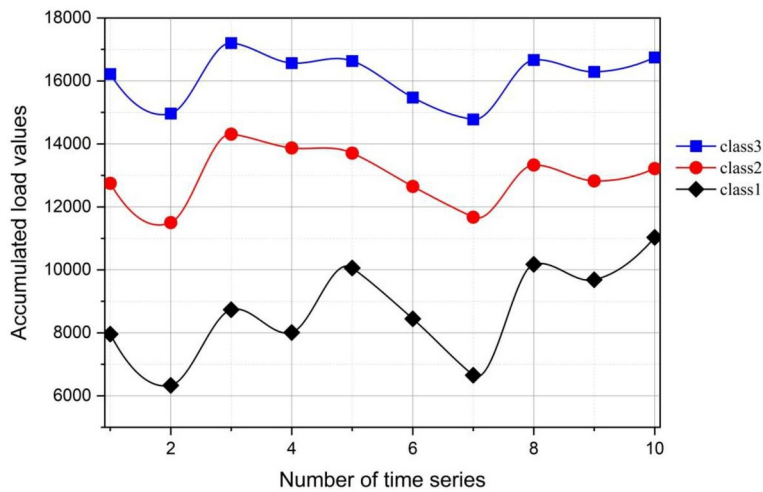


Fig. 2. Clustered load series.

Data clustering

In this section we present our clustering strategy on the existing datasets. The original load sequence is named as $L^{(0)}(k)$, where k is the index with a scope $k <= N$ (N is the length of the load series). Each consumption record is paired with a temperature reading as $T^{(0)}(k)$. According to our assumption, a load data will be put into the i th load series named $L^{(i)}(k)$, and in our simulation $i \in [1, 10]$. Our data clustering policy is described in Definition 1. As is shown in Fig. 2, three different load series are clustered out from original electricity recordings.

Definition 1 A load data $L^{(0)}(k)$ belongs to an i th load series $L^{(i)}(k)$ when it satisfies both Eqs. (1) and (2):

$$L^{(0)}(k) \in \{L^{(i)}(j), j \in [1, \tau]\} \tag{1}$$

$$\sum_{j=1}^{\tau} T^{(0)}(j - \tau) \in [\Delta T \times (i - 1), \Delta T \times i] \tag{2}$$

Symbol	Meaning	Value	Unit
L(0) (k)	Original load data series	[3162.4, ..., 3744.6]	kWh
L(i) (k)	Clustered load data series		kWh
D α (L(i) (k))	α th Caputo fractional-order equation		
α	Fractional-order of Caputo equation	[0.5, 0.55, ..., 0.95]	NA
β	Fractional-order of AGO on L(0) (k)	[0.10, 0.11, ..., 0.99]	NA
ρ	Coefficient on β th AGO	[-0.9, 0.9]	NA
C	Constant in C-FGM equation	[-4,00,000, 400,000]	NA
τ	Accumulation effort parameter	96	(5 min)

Table 1. Notations in C-FGM algorithm.

where $\Delta T = \frac{\sum_{k=1}^N T^{(0)}(k)}{N}$ is a mean value for the temperature series, and τ is a lagging effort parameter. Through this clustering operation, all power consumption data are classified into one of the 10 groups, and each group may contain multiple incidences of $L^{(i)}(k)$ for equation fitting and evaluation.

C-FGM

The contents included equation deduction steps, C-FGM flow chart, and its detailed executing procedures. The notations used throughout this part are described in Table 1.

The analytic steps for C-FGM are described as follows:

Assertion 1: In a given $L^{(i)}(k)$ data series, the fractional-order α is proportional to the temperature accumulation index i , and the value of load can be depicted using Eq. (3):

$${}_{\tau}D^{\alpha}(L^{(i)}(k)) = \rho \times L^{(i)}(k) + \frac{1}{L^{\beta}(k)} + C \tag{3}$$

where $D^{\alpha}(L^{(i)}(k))$ is an α th fractional order differential Caputo equation, and $L^{\beta}(k)$ is an β th accumulated generation operator (AGO). These two terms can be expanded using Eqs. (4) and (5) respectively:

$${}_{\tau}D^{\alpha}(L^{(i)}(k)) = \frac{1}{\Gamma(1-\alpha)} \int_0^{\tau} \frac{d(L(k,t))}{dk} dt \tag{4}$$

$$L^{\beta}(k) = \sum_{j=1}^k \frac{\Gamma(\beta+k-j)}{\Gamma(k-j+1)\Gamma(\beta)} L^{(i)}(j); k \in [1, \tau] \tag{5}$$

Because the short-term electrical load data to be fitted in our article are not continuous, Eq. (4) is now replaced by its discrete form in Eq. (6).

$${}_{\tau}D^{\alpha}(L^{(i)}(k)) = \frac{\tau}{\Gamma(\alpha) \times k} \sum_{j=1}^k \frac{L^{(i)}(j) - L^{(i)}(j-1)}{j^{\alpha} - (j-1)^{\alpha}} \tag{6}$$

Assertion 2.

Assume a function $Y(k)$ as defined in Eq. (7), then our predictive data $L^{(i)}(k)$ can be expressed as by $Y(k-1)$ and $Y(k)$ in Eq. (8):

$$Y(k) = \frac{\tau}{\Gamma(\alpha) \times k} \sum_{j=1}^k \frac{L^{(i)}(j) - L^{(i)}(j-1)}{j^{\alpha} - (j-1)^{\alpha}} \tag{7}$$

$$Y(k) = Y(k-1) + \frac{L^{(i)}(k) - L^{(i)}(k-1)}{k^{\alpha} - (k-1)^{\alpha}} \times \frac{\tau}{\Gamma(1-\alpha) \times k} \tag{8}$$

The initial condition of $Y(k)$ for the i th load series is defined in Eq. (9):

$$Y(1) = \frac{\tau}{\Gamma(\alpha)} \times L^{(i)}(1) \tag{9}$$

Now $Y(k)$ can be applied in Eq. (3), and we have Eq. (10):

$$Y(k) = \rho \times L^{(i)}(k) + \frac{1}{L^{\beta(k)}} + C \tag{10}$$

Equation (10) is a linear approximation for our α th fractional order Caputo equation, and C-FGM is able to fit real data series with it. With an optimal set of values $[\alpha, \beta, \rho, C]$ given, a predictive series $L_p^{(i)}(k)$ can be given by reversing Eq. (11).

$$L_p^{(i)}(k) = L_p^{(i)}(k - 1) + \left[\frac{k^\alpha - (k - 1)^\alpha}{\tau} \right] \times [Y(k) - Y(k - 1)] \times \Gamma(1 - \alpha) \times k \tag{11}$$

C-FGM flow chart

A detailed flow chart is given in Fig. 3, and a step-by-step description is written in Algorithm1 for readers and technicians to reproduce our simulation. A data set is firstly segmented into multiple data arrays, where each array contains power load data within a certain period (e.g. 48 h). Temperature records within each array are accumulated, and then arrays will be clustered due to their accumulated values. Each array will be assigned with a fractional term α , then a fractional order partial differential equation is initiated to fit this array. Predictions will be optimized after this equation is set, and the performance of this equation will be validated using mean absolute percentage error (MAPE) of this array.

Initial values in C-FGM should be carefully deployed to avoid unnecessary calculations. As discussed within "Experimental results on data set A" and "Statistic analysis on data set A" sections, parameter β is related with the frequency of data fluctuations, while parameter ρ is associated with the range of fluctuations. In simulations, the starting point of β will be given 0.5 gradually decrease to 0.1, if our target array is smooth. Otherwise we will select a β higher than 0.5 to fit for our highly vacillating candidate. The parameter ρ will be given an initial value of one for data fitting, then steadily moving to -1 as well.

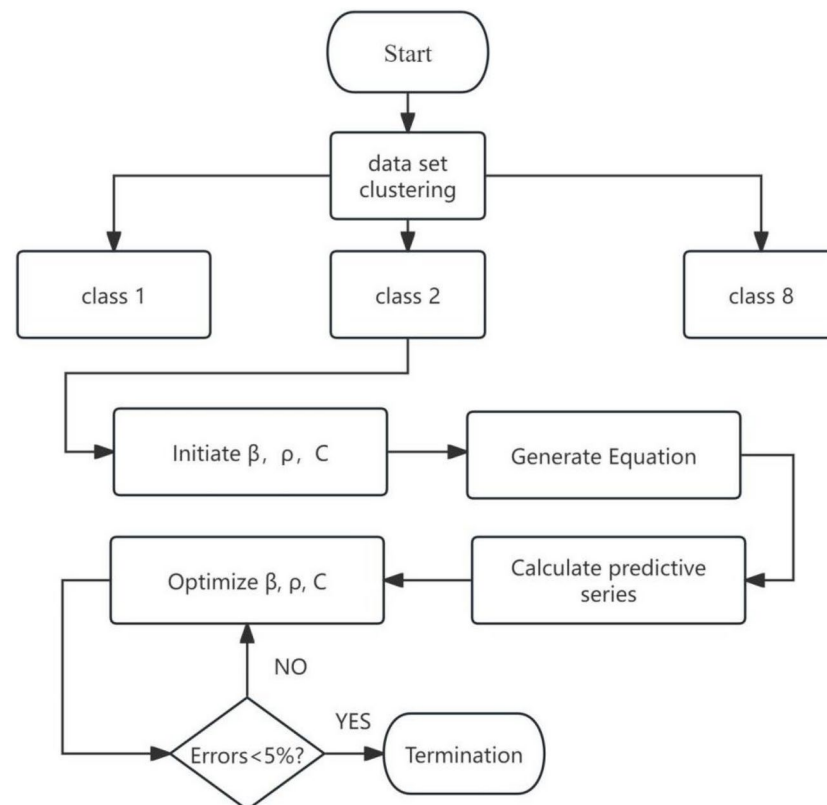


Fig. 3. C-FGM flow chart.

Require: $L_1^{(i)}, L_2^{(i)}, \dots, L_n^{(i)}$: clustered j^{th} load series;
Require: $\alpha(i)$: fractional order of Caputo equation
Require: $\beta(i)$: fractional order of AGO equation
Require: τ : of weather accumulation period
Ensure: $L_{1,p}^{(i)}, L_{2,p}^{(i)}, \dots, L_{n,p}^{(i)}$: clustered i^{th} load predictions;
Ensure: MAPE: mean error rate for i^{th} load series
Ensure: ρ : coefficient on β^{th} AGO
Ensure: C: constant in C-FGM equation

- 1: Initialize $[\rho, C]$ values
- 2: Calculate β^{th} AGO series $L^\beta(k)$ using equation (2);
- 3: Calculate intermediate series $Y(k)$ using equation (9) and (10);
- 4: Calculate predictive load series using equation (11);
- 5: Calculate mean absolute error rates using the following equation: $\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{I_{p,i} - I_{k,i}}{I_{k,i}} \right| \times 100$;
- 6: **while** an $[\rho, C]$ has been not found which satisfies the terminating criteria $\text{MAPE}_j = 5\%$ **do**
- 7: **if** $[\rho, C]$ from line 1 cannot generate a predictive series which satisfies the requirement $\text{MAPE}_j = 5\%$ **then**
- 8: the algorithm will return to line 1 with a different $[\rho, C]$ combination
- 9: **end if**
- 10: **end while**

Algorithm 1. C-FGM algorithm.

Experimental setup

Data source

The data sources directly applied in our simulation were described in Table 2: data set overview. The first data set was published by Australian Energy Market Operator (AEMO) on its website. The data set contained more than 80,000 rows of load data and temperature recordings, ranging from January 2006 until December 2010. It kept residential electricity power recordings of Australian citizens. The second file came from Global Energy Forecasting Competition (GEFCOM) 2012, held by IEEE society. This data set contained data recordings from different power stations, and in experiment the authors only chosen a certain station's consumption data for modeling.

Software implementation

The programming language inside C-FGM is PYTHON with version 3.9, and the essential software packages used are SCIPY with version 1.11.3 and NUMPY with 1.26.0. Three scientific modules are SCIKIT-LEARN with version 1.3.2, SKLEARN-LINEAR-MODEL-MODIFICATION 0.0.11 and STATSMODELS 0.14.0. C-FGM utilizes PANDAS library with version 2.1.1 to access raw data files.

Code availability

C-FGM software is publicly available on Github²⁸, with all codes and implementations available for research. Simulation results are also available per request.

Data set A		Data set B	
Name	Australia Load Data set	Name	GEFCOM 2012 Load Data set
Publisher	Australian Energy Market Operator	Publisher	Global Energy Forecasting Competition 2012
Size	87,648 rows, 3 columns	Size	32,992 rows, 28 columns
Features	Load, Temperature	Features	Load, Temperature, Station drhongtao, Glider, Pierre. (2012)
Citation	AEMO Electricity Data Model Monthly Archive ²	Citation	Global Energy Forecasting Competition 2012 Load Forecasting. https://kaggle.com/c/glob-energy-forecasting-competition-2012-load-fo

Table 2. Data set overview.

MAPE				MAE			
Class	C-FGM(ours)	LSTM	Transformer	Class	C-FGM(ours)	LSTM	Transformer
1	259.62	456.62	456.34	1	2.87	4.74	5.01
2	199.77	463.57	575.17	2	2.42	5.09	6.44
3	153.56	359.37	439.16	3	2.03	4.34	5.39
4	151.44	341.15	884.92	4	2.04	4.34	7.43
5	131.54	398.12	514.68	5	1.91	5.33	3.16
6	154.94	536.58	458.52	6	1.80	5.91	5.25
7	158.71	347.20	579.19	7	2.00	3.87	6.25
8	240.39	365.28	484.47	8	2.97	3.91	5.26
9	380.67	602.80	862.85	9	4.67	6.14	8.46

Table 3. C-FGM/LSTM/Transformer simulation reports on data set A.

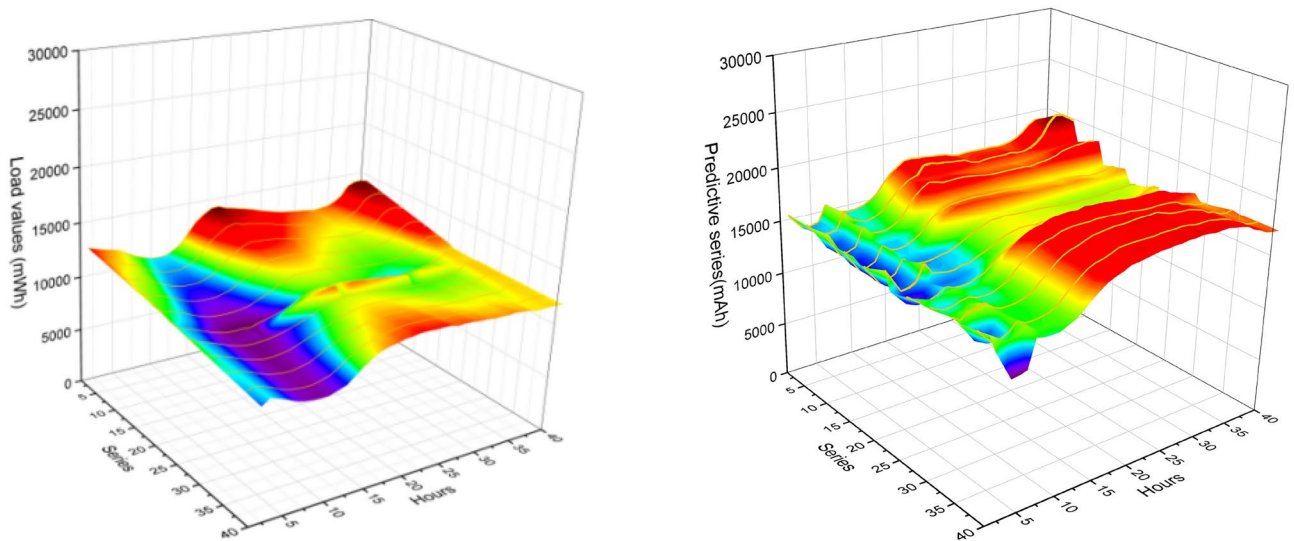


Fig. 4. 3D wave surface behavior on data set A.

Results and discussion

Experimental results on data set A

C-FGM simulations were carried on Australia electricity data set (Table 3) ranging from Jan. 1st 2006 to Dec. 31st 2010. Figure 4 displayed the three dimensional data series viewings of this data set (left) and our algorithm output(right). Eight curved lines were plotted on the surfaces showing traces of real load data and their predictions for various clustering classes. The viewings proved that C-FGM was sufficient to imitate most of the data series form correctly and caught up with trends of their fluctuation. This is because C-FGM’s embedded fractional-order equations memorized previous data series and produced their corresponding forecasts with a step-forward coefficient term, and then produced new output values.

For example, in class 1, C-FGM’s Caputo equation is (12):

$${}^96_0 D^{0.55}(L^{(1)}(k)) = -0.89 \times L^{(1)}(k) + \frac{1}{L^{0.014}(k)} - 541200 \tag{12}$$

In class 2, C-FGM’s function changed into (13):

$${}^96_0 D^{0.6}(L^{(2)}(k)) = -0.89 \times L^{(2)}(k) + \frac{1}{L^{0.018}(k)} - 495000 \tag{13}$$

Simulation results (shown in Table 3) indicated that ranging from one class to the next one. Observations from this table clearly demonstrated our algorithm has achieved better performance against two counterpart algorithms. C-FGM provided predictive outputs with absolute errors ranging from 130 to 380 kWh, while in case of LSTM and Transformer, their errors increased from 340 to 880 kWh. In terms of absolute percentage error, our algorithm has reached 1.91% to 4.67%, better than its two competitors. The average percentage error

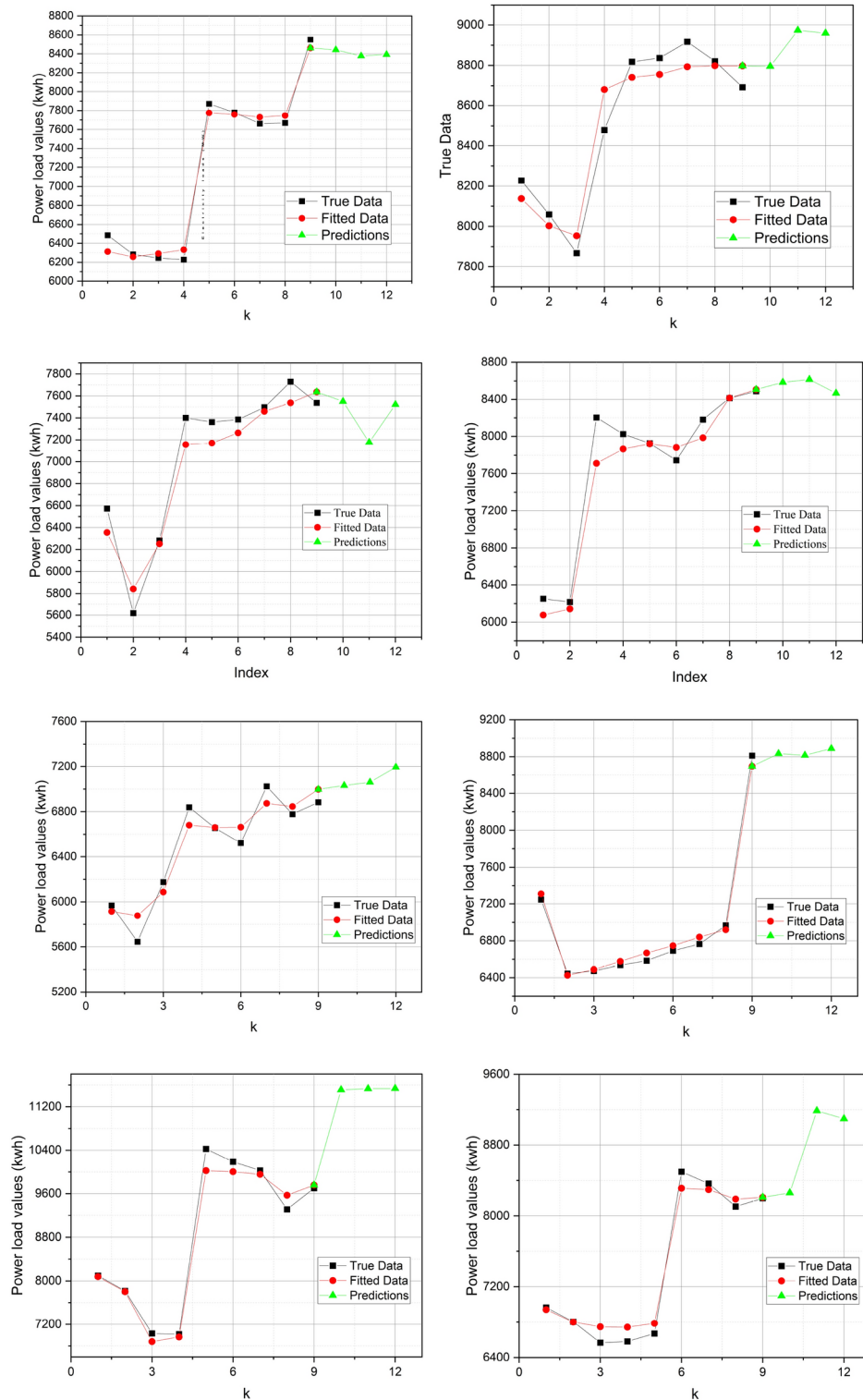


Fig. 5. C-FGM simulations on data set A, class 1–8.

for LSTM is 4.34% and 5.42% for Transformer. Thus, we can conclude that on this data set, our algorithm’s performance is much better than our counterparts.

Statistic analysis on data set A

In C-FGM, parameter α was determined as soon as the data series were clustered, and it is linearly correlated with accumulated temperatures. The proof was given in Table 3: in class 1 and class 2, when temperatures were dropped between 10 to 12.5 Fahrenheit, and it was Spring in Australia, α was given a value of 0.55 in C-FGM.

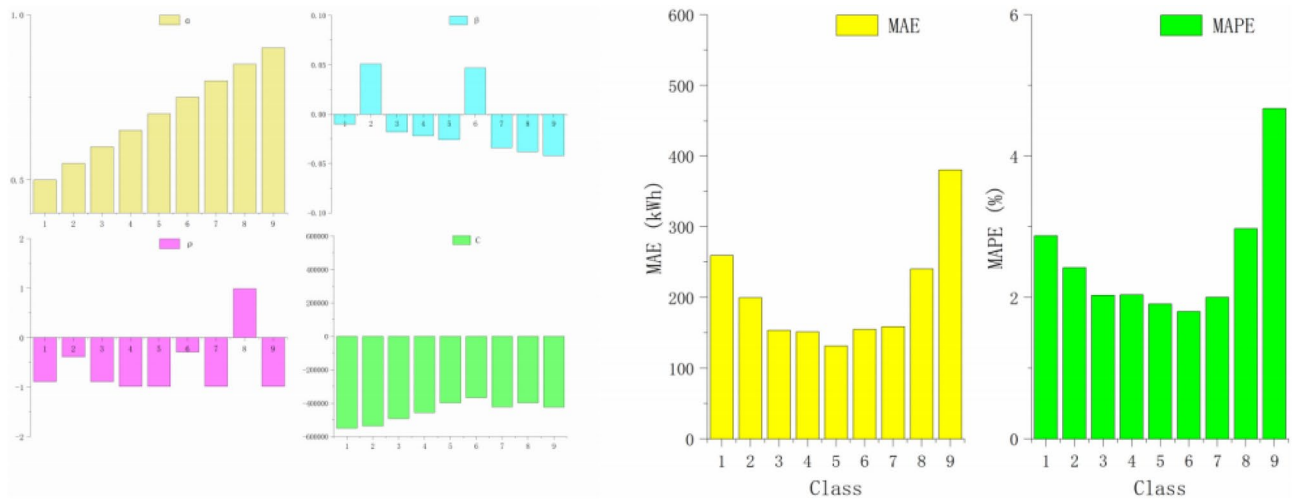


Fig. 6. Statistic analysis on data set A Left: hyper-parameter distribution Right: MAE and MAPE distribution.

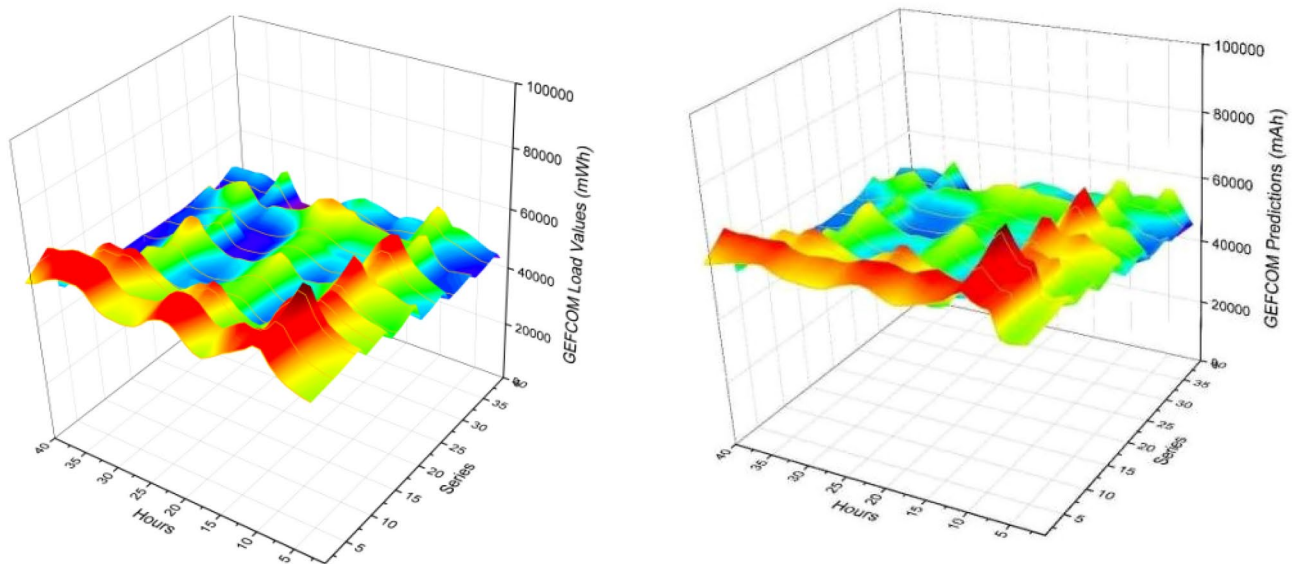


Fig. 7. 3D wave surface behavior on Data set B.

While in case of class 4 and class 6, the weather became warmer in Australia, α was assigned higher values such as 0.7 and 0.8. This linear attribute of α greatly extended C-FGM's applicability on different scenarios. As readers may learn from C-FGM's simulation results on data set B: coefficient α recorded on Table 4 were identical compared with Table 3, and its performance was still satisfactory. Thus, a conclusion can be made that despite of different power consumption patterns, as long as data are clustered according to accumulated weather levels, C-FGM is able to provide reliable and robust predictions based on linearly preset fractional order coefficient α .

Served as an AGO coefficient, value of β in C-FGM reflected the nonlinear and dynamic characteristic of raw data itself. A review of Eqs. (12), (13) and Fig. 5a,b explained why values of β reflected the slopes of data curves. In our simulations, the values of β on data set B were almost one hundred higher than on data set A, the reason is because data set B were much more dynamic and nonlinear, as demonstrated on Fig. 5. Also an examination of Figs. 5 and 6 demonstrated that smooth data series, such as class 1 and class 2 in data A, share lower numerical values of β . While steep slopes and surfaces in data set B own much higher values of β . From lab experiments, β can be concluded as a slope velocity coefficient inside C-FGM.

Experimental results on data set B

The other data set applied in our simulation is GEFCOM electricity data set ranging from Jan. 1st 2006 to Dec. 31st 2010. Figure 7 displayed three dimensional data series behaviors of this data set.

A contrast between Figs. 4 and 7 clearly indicated that data clustering was effective in both scenarios, as different clusters exhibited some extents of similarities. However, data set B exhibits a more frequent fluctuating

Class 1 (Fig. 8a): $\alpha=0.5$; $\beta=0.3$					Class 2 (Fig. 8b): $\alpha=0.55$; $\beta=0.28$				
Date	Temp	Load	Pred	Error (%)	Date	Temp	Load	Pred	Error (%)
21-02-2008	29	29,160	29,144	0.05	02-02-2004	24	28,567	28,710	0.50
21-02-2008	31	32,624	32,538	0.26	02-02-2004	21	20,317	20,570	1.24
21-02-2008	30	28,714	27,867	2.95	02-02-2004	23	23,879	23,414	1.95
21-02-2008	32	27,521	27,101	1.53	02-02-2004	20	22,422	22,860	1.95
22-02-2008	34	28,497	26,603	6.65	03-02-2004	34	18,849	19,613	4.05
22-02-2008	36	27,274	25,962	4.81	03-02-2004	36	16,523	17,507	5.96
22-02-2008	37	25,656	25,325	1.29	03-02-2004	35	16,180	16,123	0.35
22-02-2008	38	24,661	24,548	0.45	03-02-2004	37	16,746	16,786	0.24
Class 4 (Fig. 8d): $\alpha=0.65$; $\beta=0.3$					Class 6 (Fig. 8f): $\alpha=0.75$; $\beta=0.6$				
Date	Temp	Load	Pred	Error (%)	Date	Temp	Load	Pred	Error (%)
11-11-2007	52	21,497	20,499	4.64	27-10-2006	71	19,476	19,857	1.96
11-11-2007	55	21,329	21,409	1.31	27-10-2006	74	18,407	19,110	3.82
11-11-2007	53	20,543	20,260	1.38	27-10-2006	65	17,632	18,088	2.59
11-11-2007	56	19,178	19,196	0.09	27-10-2006	64	17,359	17,460	0.59
12-11-2007	63	19,042	18,038	5.27	28-10-2006	71	17,576	16,281	7.37
12-11-2007	66	18,940	18,833	0.57	28-10-2006	74	18,138	17,278	4.74
12-11-2007	65	18,140	18,145	0.03	28-10-2006	65	17,561	17,344	1.23
12-11-2007	69	17,002	17,104	0.60	28-10-2006	64	17,019	16,621	2.34

Table 4. C-FGM simulation report on data set B.

behavior compared with data set A. Visual inspection of Fig. 10b simulations(right) brought us an impression that C-FGM effectively learned these up and downs, and we can conclude that these fractional order equations were righteous solutions in fitting load data series.

A detailed inspection can be carried on Table 4 and Fig. 9 separately by viewing numerical results, specific values and trend lines. In comparison with data set A, data set B brought forward much higher values of β and ρ , because the candidate series in data set B increased, and the level of vibrations were bigger. However, our algorithm exhibited great flexibility under these two scenarios. This simulation further validated C-FGM's compliance on short term load forecast missions.

Statistic analysis on data set B

In our simulation on the second data set, parameter α (Fig. 9: Left-above) was uniformly distributed inside range [0.5–0.95] and because it was correlated with accumulated temperatures. Readers may observe this interesting attribute of α when we applied our algorithm on the second data set. The distribution of β (Fig. 9: Left-Down) in C-FGM reflected the nonlinear and dynamic characteristic of raw data itself. Readers may conclude that after clustering, the data series exhibit similar nonlinear dynamic behaviors, showing a certain level of similarities. To summarize, the hyper-parameter distributions on our second data set simulations demonstrated a certain level of stability, and the authors believe our model exhibited a good level of stability and robustness on this mission. The diagram of Fig. 9 on the right clearly revealed the error rates in both absolute level (MAE, unit kWh) and the relative level (MAPE, in percentage). Our algorithm's outputs achieved a mean absolute error rate of 620 kWh, and a mean absolute percentage error of 3.67. Compared with benchmark models such as LSTM and Transformer, the accuracy of our algorithm is 20 percent higher.

Comparison against other forecasting methods

In order to validate C-FGM's potential on short-term forecasting tasks, the authors chose two contemporary algorithms: LSTM and Transformer for performance evaluation. Both of these two models were running on data set A and B with sufficient training and testing epochs. Then, according to C-FGM's specification (eight different data series), LSTM and Transformer's calculations were rearranged and grouped for comparison.

Figure 10 demonstrated predictive data series of our algorithm against two other counterparts on data set A and B, respectively. The visualizations clearly indicated that C-FGM outperformed against two other algorithm in providing stable and less vibrating data series. Both LSTM and transformer models produced unsteady predictions, leaving their reliability under doubt. This was caused by insufficient number of data samples, especially for transformer-based systems. The authors also provided a flowchart in Fig. 11 for readers to understand the difference between C-FGM and Transformer diagram.

Numerical outputs were recorded in Table 3 including three algorithms' forecasting values and the error rates against true values. It can be seen that under most of the scenarios C-FGM delivered predictions with less error rates. Compared with contemporary models such as LSTM and the Transformer, C-FGM clearly achieved a higher accuracy (MAPE from 1.97 to 4.67%, outperforms LSTM whose average MAPE is 4.34% and Transformer whose average MAPE is 5.42%).

Although both LSTM and Transformer are superior time series predictive models, their performance is under our expectation in this scenario. The reason is because they are restricted in this task, on both these data

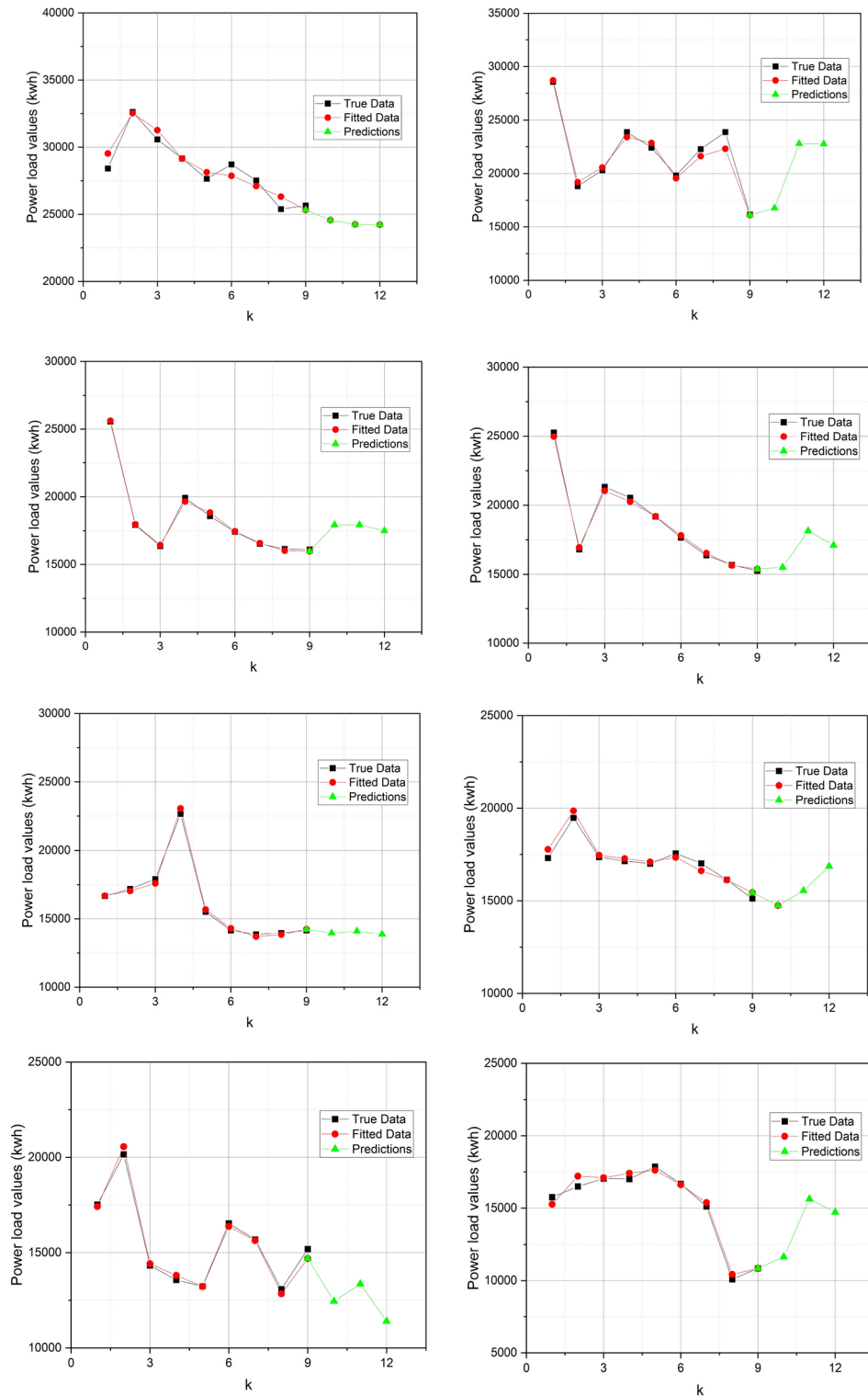


Fig. 8. C-FGM simulations on data set B, class 1–8.

sets and the hyper-parameter settings. For example, in AUS data set, there are only 3 columns available: date, temperature, load, while load was used as target. This brought difficulty to Transformer, as it needs sufficient data to feed its attention modules. If we add more columns for it such as month, weekend, price. Transformer will achieve a better performance. Similarly for LSTM, if it cannot be given enough training data (similar patterns), it will forget some data characteristics. The authors believe that the fractional order grey model is a better choice in this scenario, as a grey system is naturally born to learn data patterns without sufficient data. Another reason is related to hyper-parameter settings: our model can only be given 96 data points as maximum, or the fractional

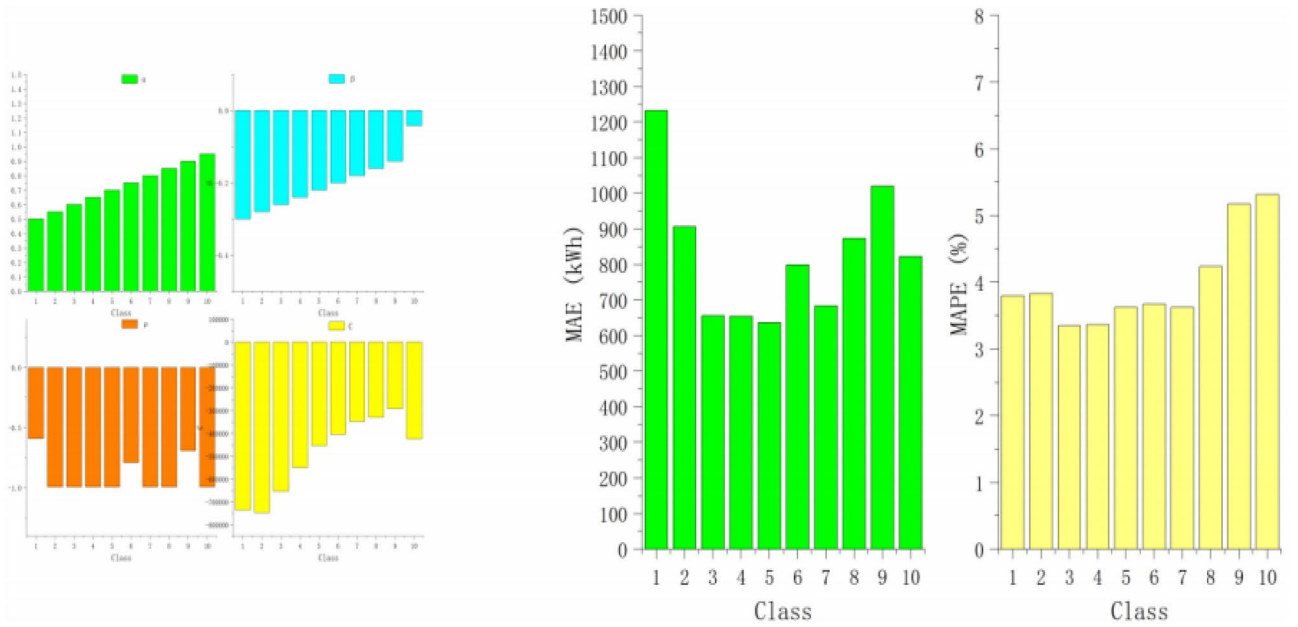


Fig. 9. Statistic analysis on data set A Left: hyper-parameter distribution Right: MAE and MAPE distribution.

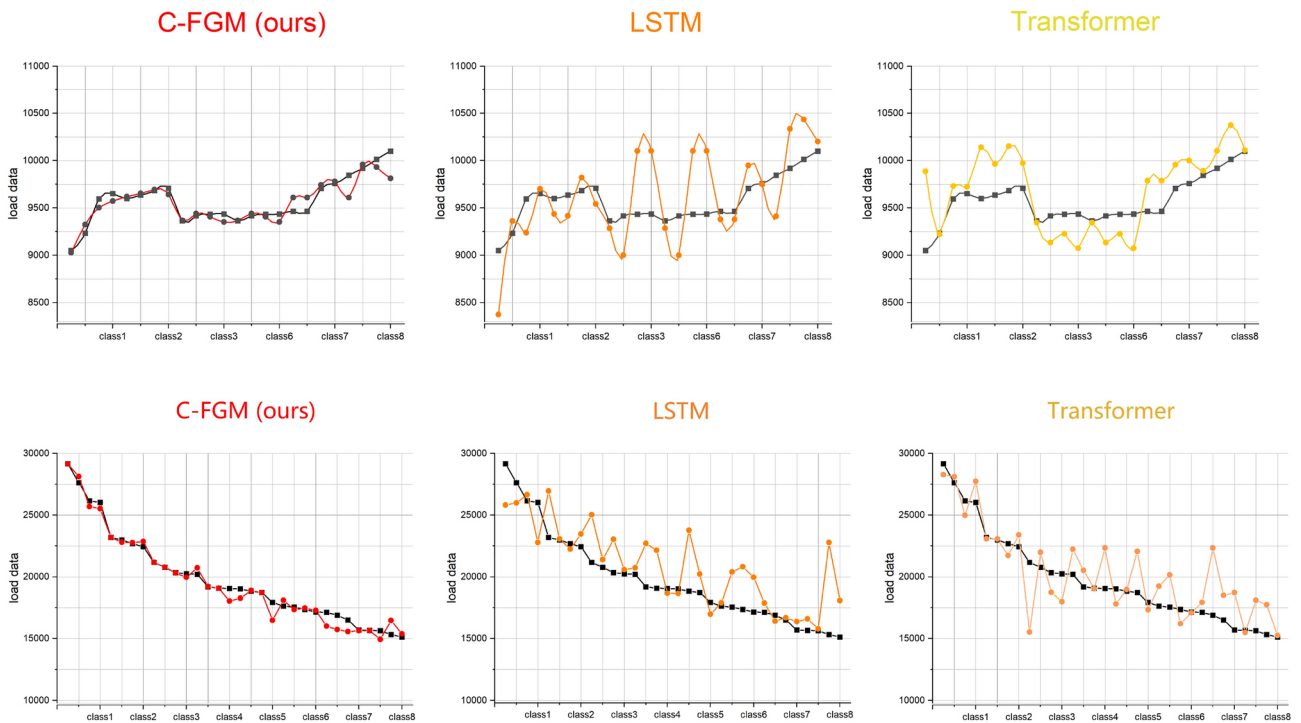


Fig. 10. C-FGM simulations on data set B, class 1–8.

order output will explode due to its exponential operation. While for Transformer series, their batch size can be set up to 256 or 512 data series. In order to compare with our model, this mission requests the maximum bulk input to be 96 points, and this may lead to a degrade of the Transformer performance.

Conclusion and future work

This article proposes a clustering fractional order grey model named as C-FGM for short term load forecasting. The originality of C-FGM can be summarized as follows:

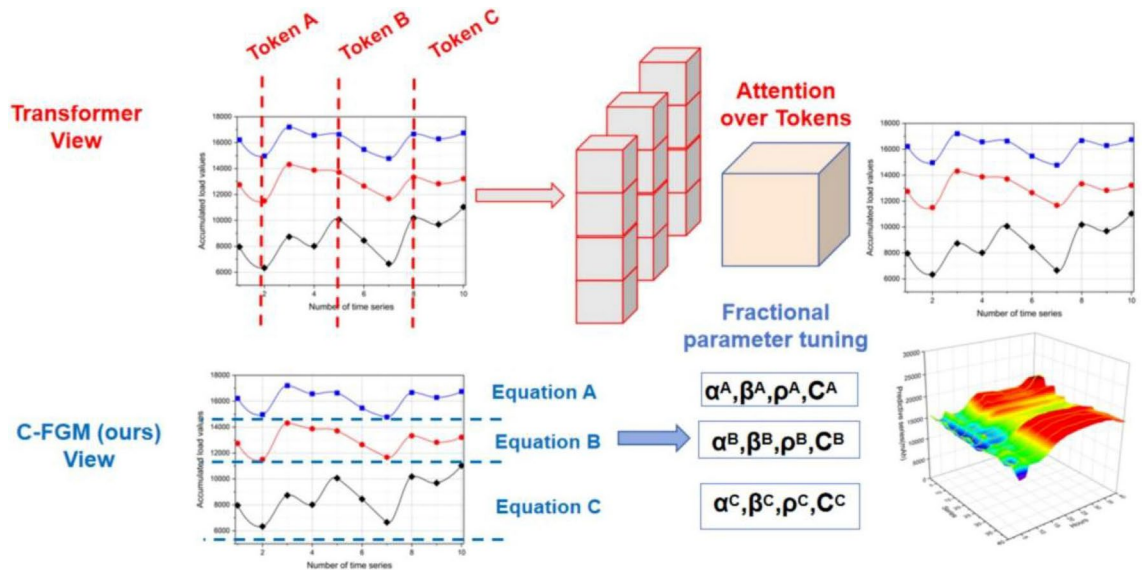


Fig. 11. C-FGM data flow versus transformer data flow.

1. utilizes a series of fractional-order partial differential equations to describe the three dimensional wave-like behavior of current power consumption data;
2. integrates a fractional order Caputo equation into a grey model and provides numerical deductions theoretically;
3. verifies a linear effort between fractional order α and day-ahead weather accumulations;
4. reveals the relationship between free parameters β , ρ , C and the accuracy of our predictions.

Although C-FGM model is an appropriate solution in short-term electrical load forecasting, it has also several disadvantages as well:

1. currently our model can only accept a time series of less than 96 data points. The root cause is because the fractional equation has an exponential component inside and once the index exceeds 96, will result in a numeric expression to be out of bound.

Thus our algorithm has to make predictions based on a memory of less than 96 data records. Recent time-series forecasters based on the Transformer architecture or LLMs usually execute on data series with batch size 256 or higher (thousands of tokens). This shortcoming obviously restricts our model's applications.

2. our clustering strategy only considers temperature variance in a time-series data set, while more and more researchers tend to analyze multi-dimensional datasets such as ETTh1 (7 horizons), PEMS04 (307 horizons). The authors are currently implementing our algorithm inside STGCNs to build forecasters on spatial-temporal data sets. We have a strong belief that once properly implemented our clustering strategy will bring a higher level of predictive accuracy.

Data availability

The datasets generated during current study are available in the Github repository <https://github.com/Neural-SIR/Algorithms/releases/tag/v1.0.0-official>.

Received: 22 April 2024; Accepted: 10 February 2025

Published online: 20 February 2025

References

1. Mohan, N., Soman, K. P. & Kumar, S. S. A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model. *Appl. Energy* **232**, 229–244. <https://doi.org/10.1016/j.apenergy.2018.09.190> (2008).
2. Australian Energy Market Operator (AEMO). Electricity Data Model Monthly Archive. <https://visualizations.aemo.com.au/aemo/newweb/index.html/mms-data-model>.
3. The US Energy Information Administration (EIA) Data Portal, <https://eia.gov/opendata/>.
4. Li, B., Lu, M., Zhang, Y. & Huang, J. A weekend load forecasting model based on semi-parametric regression analysis considering weather and load interaction. *Energies* **12**, 3820. <https://doi.org/10.3390/en12203820> (2019).
5. Lindemann, B. et al. A survey on long short memory networks for time series prediction. *Procedia CIRP* **99**, 650–655 (2021).

6. Vaswani, A. et al. *Attention Is All You Need*. *Advances in Neural Information Processing Systems* <https://arxiv.org/abs/1706.03762> (2017).
7. Zhou, H. Informer: beyond efficient transformer for long sequence time-series forecasting. In *AAAI*, vol. 35, no 12, <https://doi.org/10.1609/aaai.v35i12.17325> (2021).
8. Li, G. D., Masuda, S. & Nagai, M. The prediction model for electrical power system using improved hybrid optimization model. *Electr. Power Energy Syst.* **44**, 981–987. <https://doi.org/10.1016/j.i.jepes.2012.08.047> (2013).
9. Bian, H., Wang, Q., Xu, G. & Zhao, X. Load forecasting of hybrid deep learning model considering accumulated temperature effect. *Energy Rep.* **8**, 205–215. <https://doi.org/10.1016/j.egy.2021.11.082> (2022).
10. Makridakis, S., Spiliotis, E. & Assimakopoulos, V. The M4 Competition: 100,000 time series and 61 forecasting methods. *Int. J. Forecast.* <https://doi.org/10.1016/j.ijforecast.2019.04.014> (2019).
11. Chen, L. P. et al. Asymptotic behavior of fractional-order nonlinear systems with two different derivatives. *J. Eng. Math.* <https://doi.org/10.1007/s10665-023-10272-9> (2023).
12. Kayacan, E., Ulutas, B. & Kaynak, O. Grey system theory-based models in time series prediction. *Expert Syst. Appl.* **37**, 1784–1789. <https://doi.org/10.1016/j.eswa.2009.07.064> (2010).
13. Fan, K., Liu, J., Sun, B., Wang, J. & Li, Z. Some basic theorems and formulas for building fractal nonlinear wave models. *Alex. Eng. J.* **81**, 193–199. <https://doi.org/10.1016/j.aej.2023.09.001> (2023).
14. 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.* **13**, 1194–1204. <https://doi.org/10.1016/j.cnsns.2006.08.008> (2008).
15. Deng, J. L. Introduction to grey system theory. *J. Grey Syst.* **1**, 1–24 (1989).
16. Chen, Y. & Mao, S. H. Fractional multivariate grey Bernoulli model combined with improved wolf algorithm: Application in short-term power load forecasting. *Energy* **269**, 126844. <https://doi.org/10.1016/j.energy.2023.126844> (2023).
17. Li, H. et al. A novel fractional-order grey prediction model: a case study of Chinese carbon emissions. *Environ. Sci. Pollut. Res.* **30**, 110377–110394. <https://doi.org/10.1007/s11356-023-29919-2> (2023).
18. Wu, L., Liu, S., Fang, Z. & Xu, H. Properties of the GM(1,1) with fractional order accumulation. *Appl. Math. Comput.* **252**, 287–293. <https://doi.org/10.1016/j.amc.2014.12.014> (2015).
19. Huang, H., Tao, Z. & Liu, J. Exploiting fractional accumulation and background value optimization in multivariate internal grey prediction model and its application. *Eng. Appl. Artif. Intell.* **104**, 104360. <https://doi.org/10.1016/j.engappai.2021.104360> (2021).
20. Zhou, W., Li, H. & Zhang, Z. A novel rolling and fractional-ordered grey system model and its application for predicting industrial electricity consumption. *J. Syst. Sci. Syst. Eng.* **33**, 207–231. <https://doi.org/10.1007/s11518-024-5590-3> (2024).
21. Xie, W., Liu, C. & Wu, W. Z. A novel fractional grey system model with non-singular exponential kernel for forecasting enrollments. *Expert Syst. Appl.* **219**, 119652. <https://doi.org/10.1016/j.eswa.2023.119652> (2023).
22. Ma, X. & Liu, Z. The kernel-based nonlinear multivariate grey model. *Appl. Math. Model.* **56**, 217–238. <https://doi.org/10.1016/j.apm.2017.12.010> (2018).
23. Xiao, Q., Shan, M., Gao, M., Xiao, X. & Goh, M. Parameter optimization for nonlinear grey Bernoulli model on biomass energy consumption prediction. *Appl. Soft Comput.* **95**, 106538. <https://doi.org/10.1016/j.asoc.2020.106538> (2020).
24. Wan, K., Bin, L., Zhou, W., Zhu, H. & Ding, S. A novel time-power based grey model for nonlinear time series forecasting. *Eng. Appl. Artif. Intell.* **105**, 104441. <https://doi.org/10.1016/j.engappai.2021.104441> (2021).
25. Duan, H. M. & Wang, G. Partial differential grey model based on control matrix and its application in short-term traffic flow prediction. *Appl. Math. Model.* **116**, 763–785. <https://doi.org/10.1016/j.apm.2022.12.012> (2023).
26. Wu, W. A time power-based grey model with comfortable fractional derivative and its applications. *Chaos Solitons Fract.* **155**, 111657. <https://doi.org/10.1016/j.chaos.2021.111657> (2022).
27. Zhang, Y. H. A time power-based grey model with Caputo fractional derivative and its application to the prediction of renewable energy consumption. *Chaos Solitons Fract.* **164**, 112750. <https://doi.org/10.1016/j.chaos.2022.112750> (2022).
28. C-FGM project code repository, public on Jan. 10th <https://github.com/Neural-SEIR/Algorithms/releases/tag/v1.0.0-official> (2024).

Acknowledgements

This work is part of a project supported by the National Science Foundation for Scientists of China (Grant No. 42371433).

Author contributions

All authors contributed to the study conception and model design. Material preparation, data collection and analysis were performed by Yuchen Qian, Lisen Zhao and Chang Tan. The first draft of the manuscript was written by Xiang Yu and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to X.Y. or L.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025