

Improving electricity demand forecasting accuracy: a novel grey-genetic programming approach using GMC(1,N) and residual sign estimation

Flavian Emmanuel Sapnken

Laboratory of Technologies and Applied Science, University Institute of Technology, Douala, Cameroon

Benjamin Salomon Diboma

Higher Institute of Transport, Logistics and Commerce, Ambam, Cameroon

Ali Khalili Tazehkandgheshlagh

Faculty of Agricultural Engineering, University of Tehran, Tehran, Cameroon

Mohammed Hamaidi

Department of Maths, Faculty of Science and Technology, Ziane Achour University of Djelfa, Djelfa, Algeria

Prosper Gopdjim Noumo

Laboratory of Technologies and Applied Sciences, Douala, Cameroon

Yong Wang

School of Sciences, Southwest Petroleum University, Chengdu, China, and

Jean Gaston Tamba

Laboratory of Technologies and Applied Science, University Institute of Technology, University of Douala, Douala, Cameroon

Abstract

Purpose – This paper addresses the challenges associated with forecasting electricity consumption using limited data without making prior assumptions on normality. The study aims to enhance the predictive performance of grey models by proposing a novel grey multivariate convolution model incorporating residual modification and residual genetic programming sign estimation.

Design/methodology/approach – The research begins by constructing a novel grey multivariate convolution model and demonstrates the utilization of genetic programming to enhance prediction accuracy by exploiting the signs of forecast residuals. Various statistical criteria are employed to assess the predictive performance of the proposed model. The validation process involves applying the model to real datasets spanning from 2001 to 2019 for forecasting annual electricity consumption in Cameroon.

Findings – The novel hybrid model outperforms both grey and non-grey models in forecasting annual electricity consumption. The model's performance is evaluated using MAE, MSD, RMSE, and R^2 , yielding



We are particularly grateful to Mr David Horgan for his help in troubleshooting the algorithm implemented in [Section 3](#). His expertise in Python/Matlab programming was crucial in resolving the performance issues and ensuring reliable results.

This work was supported by the Natural Science Foundation of Sichuan Province (No. 2023NSFSC0428), the Central Government Funds of Guiding Local Scientific and Technological Development (No. 2023ZYD0004), the Sichuan National Applied Mathematics Center open fund (No. 2024-KFJJ-01-01), and the Chengdu Science and Technology Project (No. 2024-YF05-00323-SN).

values of 0.014, 101.01, 10.05, and 99% respectively. Results from validation cases and real-world scenarios demonstrate the feasibility and effectiveness of the proposed model. The combination of genetic programming and grey convolution model offers a significant improvement over competing models. Notably, the dynamic adaptability of genetic programming enhances the model's accuracy by mimicking expert systems' knowledge and decision-making, allowing for the identification of subtle changes in electricity demand patterns.

Originality/value – This paper introduces a novel grey multivariate convolution model that incorporates residual modification and genetic programming sign estimation. The application of genetic programming to enhance prediction accuracy by leveraging forecast residuals represents a unique approach. The study showcases the superiority of the proposed model over existing grey and non-grey models, emphasizing its adaptability and expert-like ability to learn and refine forecasting rules dynamically. The potential extension of the model to other forecasting fields is also highlighted, indicating its versatility and applicability beyond electricity consumption prediction in Cameroon.

Keywords Hybrid forecasting, Residual modification, GP sign estimation, Electricity demand, Cameroon

Paper type Research paper

1. Introduction

Accurate electricity demand forecasting is critical for ensuring a stable and reliable power supply, especially in developing economies like Cameroon (Tamba *et al.*, 2022). From illuminating homes and powering industries to driving technological advancements and ensuring healthcare access, its steady flow underpins economic development and social progress (Ugembe *et al.*, 2023). Yet, amidst this reliance on a seemingly abundant resource, lurks a hidden challenge: accurately predicting future demand.

For developing economies like Cameroon, grappling with rapid urbanization and burgeoning energy needs, the stakes of inaccurate electricity forecasting are particularly high (Guefano *et al.*, 2021). Power outages ripple through communities, disrupting businesses, jeopardizing essential services, and hindering economic growth (Dieudonne *et al.*, 2022). Conversely, oversupply translates into wasted resources and financial losses. Striking a delicate balance between these extremes necessitates a precise understanding of future electricity consumption patterns.

Traditional forecasting methods, while reliable in certain contexts, often struggle to capture the intricate dance of factors influencing electricity demand in evolving economies (Tamba *et al.*, 2018). Autoregressive integrating moving average (ARIMA) models, for instance, struggle with the inherent non-linearity and complex interdependencies within the data (Quartey-Papafio *et al.*, 2020; Li *et al.*, 2023). Regression models, while adept at capturing linear relationships, may miss the subtle yet influential nuances hidden within (Yildiz *et al.*, 2017).

In this landscape of forecasting uncertainties, grey models appear to offer a glimmer of hope (Xie and Wang, 2017). Tailored for scenarios with limited or incomplete data, they possess an inherent ability to model uncertainties and extract hidden patterns even from the sparsest of datasets (Sapnken *et al.*, 2023a). Their reliance on data generation processes and rolling forecasts makes them particularly well-suited for capturing the dynamic nature of electricity consumption.

However, while grey models excel in extracting insights from limited data, their inherent limitations in capturing intricate non-linear relationships persist. This is where genetic programming (GP) steps onto the stage, a powerful evolutionary algorithm capable of automatically generating non-linear programs that adapt to complex data patterns (Castelli *et al.*, 2015). By harnessing the strengths of both approaches, a novel hybrid grey-genetic model could emerge. This innovative hybrid model seamlessly blends the data-driven insights of grey models with the non-linear adaptability of GP. The convolutional grey multivariate model lays the foundation by capturing temporal dependencies and extracting underlying patterns from the data. GP then enters the fray, sculpting unique non-linear programs that evolve to fit the complexities of electricity consumption data, revealing otherwise hidden relationships between variables.

The GPGMC(1,N) model extends beyond just improving electricity demand forecasting in Cameroon. Its potential value lies in its broader application across various domains. By providing more accurate forecasts, the GPGMC(1,N) model empowers policymakers in developing economies to make informed decisions on infrastructure development, resource allocation, and energy security strategies. This translates to improved energy management for utilities, who can leverage the model for planning, optimizing generation schedules, and minimizing outages. Furthermore, the core principles of GPGMC(1,N) can be adapted to forecast water consumption, traffic patterns, and sales figures in data-scarce environments. This research, by demonstrating the effectiveness of GPGMC(1,N) in Cameroon's electricity sector, paves the way for broader application and adaptation across various sectors in developing economies, ultimately contributing to improved decision-making, resource management, and overall economic development.

In the following sections, we delve deeper into the theoretical underpinnings and practical implementation of this novel model, showcasing its effectiveness in predicting annual electricity consumption in Cameroon and paving the way for a future illuminated by precise foresight.

This study is driven by two overarching objectives, each aimed at illuminating the path towards enhanced electricity forecasting in Cameroon:

- (1) **Develop and implement the GPGMC(1,N) model:** The primary objective is to meticulously design and implement the GPGMC(1,N) model, fine-tuning its parameters and adapting it to the specific context of Cameroon's electricity data. This involves defining the population size and operators for GP, selecting appropriate grey model components, and optimizing the training process for enhanced performance.
- (2) **Evaluate the forecasting accuracy and explainability of the GPGMC(1,N):** With the model in place, the second objective revolves around rigorously evaluating its forecasting accuracy. We will compare the GPGMC(1,N) predictions to those of established models and assess its effectiveness in capturing temporal trends and seasonal variations. Furthermore, we will delve into the interpretability of the model, identifying the key non-linear relationships and variables that drive its predictions.

In the fast-growing field of electricity forecasting models, the GPGMC(1,N) is a major innovation that has carved out a special place for itself thanks to its characteristics and considerable contributions. The novelties of this study are threefold:

- (1) **Pioneering the convergence of GP and GMC(1,N):** To the best of our knowledge, this study marks the first-ever application of a hybridized model combining the strengths of GMC and GP for electricity forecasting. This novel approach transcends the limitations of existing methods, unlocking a new frontier in forecasting accuracy and adaptability.
- (2) **Empowering GMs with non-linear insights:** While adept at data-driven forecasting, grey models often struggle to capture the intricate non-linear relationships within electricity consumption data. The GPGMC(1,N) bridges this gap by seamlessly integrating GP's evolutionary search for non-linear patterns, enriching the model's understanding of complex dynamics.
- (3) **Tailoring data-scarce scenarios:** Cameroon, like many developing nations, faces challenges in securing extensive historical data for electricity consumption. The novel GPGMC(1,N), built upon the foundation of grey models, thrives in such data-

scarce environments, utilizing its robust data generation process to extract valuable insights from limited datasets.

2. Literature review

2.1 Previous studies

Research is booming in the field of predicting electricity consumption across various sectors and timeframes. As [Tamba et al. \(2018\)](#) point out, these approaches can be broadly categorized into three groups: 1) Statistical models: These rely on established statistical methods like the ARIMA ([Tarmanini et al., 2023](#)), XGBoost regression ([Wang et al., 2021](#)), adaptive decomposition, and Markov-chain mixture distribution ([Munkhammar et al., 2021](#)). Their strengths lie in their interpretability, simplicity, and ease of use ([Atalay et al., 2019](#); [Kapoor and Wichitaksorn, 2023](#)). However, they require expert knowledge to find relationships between variables and depend heavily on historical data. 2) Machine learning (ML) models: These utilize advanced algorithms like support vector machines ([Haq et al., 2021](#); [Yin et al., 2023](#)), artificial neural networks ([Wazirali et al., 2023](#)), LSTM networks ([Bilgili and Pinar, 2023](#)), and deep recurrent neural networks ([Abdulrahman et al., 2021](#)). They excel at handling complex calculations in electricity consumption forecasting and can deliver impressive results with large datasets. 3) Grey models: These are data-driven methods suitable for situations with limited information. They are particularly useful for short-term predictions and offer an alternative when other models lack sufficient data.

However, predicting electricity consumption becomes increasingly complex with rapid urbanization and industrialization, driven by fundamental changes in the industrial structure ([Yin et al., 2023](#)). This highlights the need for continuous research and development of even more sophisticated forecasting methods.

Although the predictive accuracy of ML is well established, it still has some significant weaknesses. For instance, ML forecasts for electricity demand can be data-hungry, prone to overfitting, and offer limited insight into why consumption changes ([Sapnken et al., 2023c](#)). This can hinder their performance and usefulness for grid operators. Similarly, statistical models for electricity demand forecasts have difficulty adapting to non-linear changes and require high-quality data, which limits their accuracy in dynamic environments ([Ungureanu et al., 2021](#)).

According to [Xie and Wang \(2017\)](#), grey models are very useful for energy forecasts when data is scarce or complex. Their simplicity makes them interpretable and easy to implement, even with limited information, while their robustness allows uncertainty and data gaps to be managed ([Qian and Sui, 2021](#); [Lei et al., 2024](#)). Although grey models (GMs) do not always achieve the maximum accuracy of complex models, they are a powerful option for reliable forecasts when data or expertise is scarce.

Given its effectiveness in producing precise predictions for incomplete information systems, the grey system theory introduced by [Deng \(1982\)](#) has found application in various domains, as indicated by [Xie and Wang \(2017\)](#). Just to cite a few, [Qian and Sui \(2021\)](#) proposed an intelligent grey model for forecasting renewable energy demand. [Akay and Atak \(2007\)](#) proposed an adjusted grey prediction model that integrates a new condition and rolling mechanism. [Wang et al. \(2023\)](#) introduced an innovative fractional cumulative operator to devise a structure-adaptive fractional derivative grey prediction model aimed at predicting China's overall energy consumption. There are many other grey models that have been developed and applied in other fields other than energy. [Xie and Wang \(2017\)](#) work summarizes them well.

While simple and data-efficient, univariate grey models (GM(1,1)) raise concerns. They neglect key external factors, leading to inaccurate forecasts when those factors change ([Wang et al., 2022b](#)). Sensitive to data imperfections, GM(1,1) models struggle with outliers

and gaps, skewing predictions (Qian and Sui, 2021). Their focus on linear relationships misses complex dynamics, limiting accuracy in non-linear scenarios as demonstrated by Sapnken and Tamba (2022). Lacking sophisticated feature engineering, GM(1,1) models may overlook subtle patterns (Zhao *et al.*, 2023). Validating their suitability is challenging due to their simplicity, and overfitting risks producing misleading forecasts. To achieve reliable and accurate predictions, researchers have considered combining them with other models (Guefano *et al.*, 2021), incorporating external factors (Tien, 2012), or restructuring their framework for more flexibility (Wang *et al.*, 2023).

GM(1,N) models can overcome most of the weakness of the basic GM(1,1) model and has a number of strengths, including its ability to handle small sample sizes, to handle non-linear relationships between variables and to deal with missing data (Ye *et al.*, 2024). Nevertheless, the efficacy of the residual series within GM(1,N) is contingent on the prevalence of data points exhibiting consistent signs, a scenario that tends to be infrequent when the number of observations is limited (Min *et al.*, 2012; San Cristóbal *et al.*, 2015). Regrettably, GM(1,N) often associates the progression of the target variable with $(N - 1)$ independent variables, considering factors that could influence system variations, as noted by Wang *et al.* (2022a). This issue was initially highlighted by Tien (2012), who demonstrated that GM(1,N) is essentially a causal model and is generally ineffective for forecasting purposes. Nevertheless, enhanced versions like GMC(1,N) have demonstrated their effectiveness in forecasting, as supported by studies conducted by Wu *et al.* (2018), Ding and Li (2020) and Yin and Mao (2023).

To improve the accuracy of predicting residual signs in Grey Models (GMs), researchers have explored enhanced residual sign estimators. For instance, Hsu and Chen (2003) introduced an upgraded GM that incorporates both residual ANNs sign estimation and residual modification for electricity consumption forecasts. In a similar vein, Hsu (2003) modified the residual of the GM(1,1) model, employing Markov-chain sign estimation to predict the value of the global integrated circuit industry. Building on these approaches, this study introduces an innovative method that combines residual modification and genetic programming (GP) sign estimation to refine the precision of the residual sign estimator.

GP, a strategy for evolving functions, is effective in performing designated tasks (Gil-Gala *et al.*, 2023). Similar to genetic algorithms (GAs), GP utilizes crossover, mutation, and reproduction rules to find optimal solutions (Ong *et al.*, 2005). What sets GP apart is its ability to perform well without assuming specific relationships between dependent and independent variables, making it particularly suitable for small datasets. GP offers two main advantages: it can derive a mathematical equation through regression analysis, and it can express a mathematical expression using a parse tree technique. Consequently, GP emerges as an efficient tool for achieving optimal residual sign estimation.

2.2 Summary, contributions and novelty

None of the aforementioned studies has combined GP and GMC(1,N) despite the benefits they could offer. This research seeks to address a gap in the existing knowledge by integrating residual modification and residual GP sign estimation to enhance a grey prediction model. The motivation behind creating this model lies in two primary reasons. Firstly, when dealing with a limited time-series dataset, employing GP can significantly enhance the precision of predicting residual signs. Secondly, compared to ANNs, GP demonstrates superior accuracy and reliability in constructing forecasting models, as evidenced by previous studies (Ong *et al.*, 2005; Huang *et al.*, 2006). The study utilizes electricity consumption data from Cameroon as an empirical case to showcase the effectiveness of the proposed model. They include:

- (1) **Enhancing policy formulation:** With reliable and accurate forecasts provided by the GPGMC(1,N), policymakers can design and implement informed strategies for infrastructure development, resource allocation, and energy security measures. This translates into more efficient management of energy resources, contributing to economic growth and sustainable development.
- (2) **Improving grid stability:** Accurate predictions of electricity demand can help optimize grid operations, enabling the anticipation of peak loads and facilitating the allocation of resources to prevent outages and ensure stable power supply. This empowers consumers and industries alike, fostering confidence and promoting economic activity.
- (3) **Paving the way for future models:** The success of the GPGMC(1,N) in Cameroon can serve as a springboard for further research and development in the field of grey modelling and hybrid forecasting approaches. This opens doors for exploring the application of similar models in other data-scarce contexts and for diverse forecasting challenges beyond electricity demand.

The remaining sections of this paper are divided as follows: The [Section 3](#) outlines the methodology and principles used; [Section 4](#) presents the simulations carried out, the results obtained and the related discussions; This is followed by [Section 5](#), which reveals the significance of the findings, challenges to be overcome and the opportunities to be seized. Finally, the [Section 5](#) concludes the paper and casts a glance towards future work.

3. Methodological framework

3.1 The standard GMC(1,N) model

Consider a system of N variables (called sequences) defined as follows:

$$X^{(0)} = \{X_1^{(0)}; X_2^{(0)}, \dots, X_N^{(0)}\} \tag{1}$$

where $X_1^{(0)}$ represents the dependent variable (electricity demand) and $X_i^{(0)} (i = 2, \dots, N)$ represent the independent variables (prices, income, household expenses and number of subscribers). Assume there is a strong correlation between $X_1^{(0)}$ and $X_i^{(0)} (i = 2, \dots, N)$. Moreover, we make the assumption that the length of the sequence for each variable $X_i^{(0)} (i = 1, \dots, N)$ is n , so:

$$X_i^{(0)} = \{x_i^{(0)}(1), x_i^{(0)}(2), \dots, x_i^{(0)}(n)\}, i = 1, 2, \dots, N. \tag{2}$$

The GMC model, which involves multivariate grey convolution, relies on the mean sequence and 1-AGO (1st order accumulating generation operator). $X_i^{(1)} (i = 1, 2, \dots, N)$ represent 1-AGO sequences of $X_i^{(0)} (i = 1, \dots, N)$ and are established as follows:

$$x_i^{(1)}(k) = \sum_{p=1}^k x_i^{(0)}(p), k = 1, 2, \dots, n \tag{3.1}$$

$$X_i^{(1)} = \{x_i^{(1)}(1), x_i^{(1)}(2), \dots, x_i^{(1)}(n)\}, i = 1, 2, \dots, N. \tag{3.2}$$

1-AGO defined in [Eq. \(3.1\)](#) is identical to the one defined in conventional GM(1,N) model. [Eq. \(4.1\)](#) illustrates the mean sequences generated by successive terms of $X_i^{(1)}$:

$$Z_i^{(1)} = \{z_i^{(1)}(1), z_i^{(1)}(2), \dots, z_i^{(1)}(n)\}, i = 1, 2, \dots, N. \tag{4.1}$$

$Z_i^{(1)}$ ($i = 1, 2, \dots, N$) is defined by Eq. (4.2):

$$z_i^{(1)}(k) = 0.5(x_i^{(1)}(k-1) + x_i^{(1)}(k)), k = 2, 3, \dots, n; i = 1, 2, \dots, N \tag{4.2}$$

Assuming that $X_i^{(0)}$ ($i = 2, \dots, N$) and $X_1^{(0)}$ are still defined as in Eqs. (1) and (2), respectively, and Eq. (3) still defines the 1-AGO sequences $X_i^{(1)}$ ($i = 1, \dots, N$). Then, the GMC(1,N) model is given by Eq. (5):

$$\frac{dx_1^{(1)}(t)}{dt} + \alpha x_1^{(1)}(t) = u + \sum_{i=2}^N \beta_i x_i^{(1)}(t) \tag{5}$$

The expression $\sum_{i=2}^N \beta_i x_i^{(1)}(t)$ involves the derivative term represented by β_i coefficients, while $-\alpha$ and u are the development coefficient and parameters for GMC(1,N), respectively.

By hypothesizing that the right-hand side (RHS) of Eq. (5) can be expressed as a function $f(t)$ Tien (2012) initially estimates Eq. (5) as the subsequent difference equation.

$$x_1^{(0)}(t) + \alpha z_1^{(1)}(t) = u + \sum_{i=2}^N \beta_i z_i^{(1)}(t), t = 2, \dots, n \tag{6}$$

This involves performing integrals on both sides of Eq. (5) over the range from $(t-1)$ to t and subsequently employing the trapezoid formula for the remaining unspecified terms. Eq. (6) represents a set of linear equations that can be expressed in matrix format as:

$$BA = Y \tag{7}$$

where:

$$\left\{ \begin{array}{l} A = \begin{pmatrix} \alpha \\ \beta_2 \\ \vdots \\ \beta_N \\ u \end{pmatrix} \in \mathbb{R}^{(N+1) \times 1} \\ B = \begin{pmatrix} -z_1^{(1)}(2) & z_2^{(1)}(2) & \dots & z_N^{(1)}(2) & 1 \\ -z_1^{(1)}(3) & z_2^{(1)}(3) & \dots & z_N^{(1)}(3) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -z_1^{(1)}(n) & z_2^{(1)}(n) & \dots & z_N^{(1)}(n) & 1 \end{pmatrix} \in \mathbb{R}^{(n-1) \times (N+1)} \\ Y = \begin{pmatrix} x_1^{(0)}(2) \\ x_1^{(0)}(3) \\ \vdots \\ x_1^{(0)}(n) \end{pmatrix} \in \mathbb{R}^{(n-1) \times 1} \end{array} \right. \tag{8}$$

The least-squares method can be employed to calculate the parameters $A = [\alpha, \beta_2, \dots, \beta_N, u]^T$, provided that the matrix $B^T B$ is invertible.

$$A = (B^T B)^{-1} B^T Y \tag{9}$$

Furthermore, Tien (2012) employed the starting condition $\hat{x}_1^{(1)}(t = 1) = x_1^{(0)}(1)$ to derive the time response function [1] of the GMC(1,N) model as presented in Eq. (5):

$$\hat{x}_1^{(1)}(t) = x_1^{(0)}(1)e^{\alpha(1-t)} + \int_{\tau=1}^{\tau=t} e^{\alpha(\tau-t)} f(\tau) d\tau \tag{10}$$

The convolution integral is situated on the right-hand side of Eq. (10), posing challenges in deriving a direct expression. Fortunately, we can employ numerical integrations to estimate the outcome. The trapezoid formula is an easy-to-use method [2] that produces accurate time response function, as seen in Eq. (11):

$$\hat{x}_1^{(1)}(t) = x_1^{(0)}(1)e^{\alpha(1-t)} + 0.5h(t) \sum_{\tau=2}^t (f(\tau)e^{\alpha(\tau-t)} + f(\tau-1)e^{\alpha(\tau-t-1)}); t \geq 2 \tag{11}$$

$h(t)$ in Eq. (11) represents the unit step defined as:

$$h(t) = \begin{cases} 0, & t < 2 \\ 1, & t \geq 2 \end{cases}$$

Finally, by using inverse 1-AGO, it is possible to determine the predicted value $\hat{x}_1^{(0)}(t)$:

$$\hat{x}_1^{(0)}(t) = \hat{x}_1^{(1)}(t) - \hat{x}_1^{(1)}(t - 1), t \geq 2$$

For the sake of brevity, we have moved the analysis of the proof of superiority of the GMC model over the conventional GM(1,1,N) model to the appendix, and we have also highlighted the weaknesses of the GMC(1,N) model.

3.2 Genetic programming-based GMC(1,N) model

The residual series is the disparity between the target values $x_1^{(0)}(t)$ and the anticipated values $\hat{x}_1^{(0)}(t)$. To enhance the forecasting precision of the GMC(1,N) model, it is essential to develop a residual GMC(1,N) model. When original GMC(1,N) and residual GMC(1,N) are combined, the updated forecasted values are obtained. Nevertheless, the effectiveness of the residual series is influenced by the quantity of data points sharing the same sign. If there are fewer than four data points displaying the same sign, it is not feasible to construct the residual GMC(1,N) model.

Hsu and Chen (2003) proposed a forecasting technique using grey models in 2003. Their method combined modifying residuals with estimating their signs using ANNs. While effective, ANNs require large datasets and struggle with justifying the hidden layer complexity. This research, therefore, presents an improved version of the GMC(1,N) model that combines residual modification with estimating signs using GP. This new approach aims to boost the accuracy of predicting residual signs. More details about the proposed model's construction can be found in subsection 3.2.1.

3.2.1 Residual GMC(1,N) model. Denote the residual sequence's initial absolute values as $\epsilon^{(0)}$, which is given by,

$$\epsilon^{(0)} = \{ \epsilon_1^{(0)}; \epsilon_2^{(0)}, \dots, \epsilon_N^{(0)} \}, i = 1, 2, \dots, N \tag{12}$$

where,

$$\epsilon^{(0)}(k) = |x_1^{(0)}(k) - \hat{x}_1^{(0)}(k)|, k = 2, 3, \dots, n \tag{13}$$

Using Eqs. (1)–(10), the GMC(1,N) model of $\hat{\epsilon}^{(0)}$ can be created. Eq. (14.1) provides the forecast of accumulated residual series model, from which inverse AGO calculates $\epsilon^{(0)}$ as stated in Eq (14.2),

$$\hat{\epsilon}_1^{(1)}(t) = \epsilon_1^{(0)}(1)e^{\alpha(1-t)} + 0.5h(t) \sum_{\tau=2}^t (f(\tau)e^{\alpha(\tau-t)} + f(\tau-1)e^{\alpha(\tau-t-1)}); t \geq 2 \tag{14.1}$$

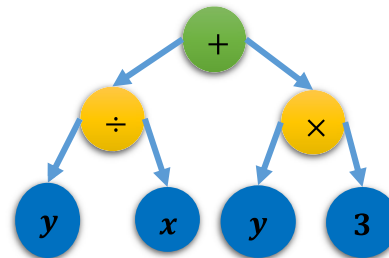
$$\hat{\epsilon}_1^{(0)}(t) = \hat{\epsilon}_1^{(1)}(t) - \hat{\epsilon}_1^{(1)}(t-1), t \geq 2 \tag{14.2}$$

3.2.2 Model for estimating GP residual sign. GP is a method developed by Koza (1994) for data forecasting and grouping, applicable in the realm of computer programs. It has found utility in symbolic regression and the identification of model structures. The key principles—crossover, mutation, and reproduction—share similarities with GAs (Nyathi and Pillay, 2018). Unlike GAs, GP employs a generic parse tree representation rather than the binary logic numbers (0 and 1) of genetic states. Consequently, GP has gained substantial popularity compared to conventional linear forecasting techniques, owing to its adeptness in navigating complex non-linear domains. Additionally, GP is extensively employed in practical scenarios, including forecasting coastal algal blooms, constructing credit scoring models, and simulating rainfall-runoff processes.

The operators ($\{+, -, \log, exp\}$), trigonometric functions ($\{sin, cos, tan\}$), and conditional statements (If, then, while) are among GP functions and statements. The GP parse tree in Figure 1 can be used to describe $3y + x/y$ as a straightforward example.

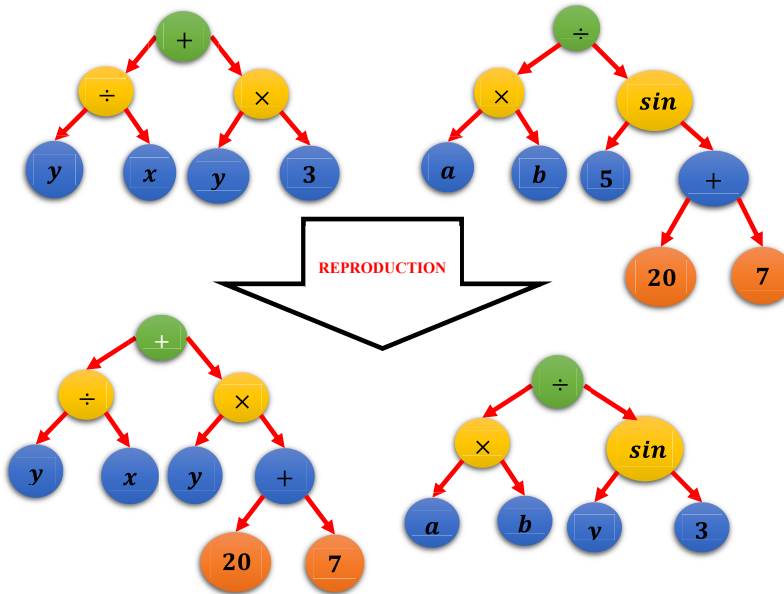
Additionally, by combining a generic parse tree with symbolic regression, the GP operation system may provide an ideal prediction function. Figure 2 depicts an example of the crossover operator in GP. Unlike ANNs, GP is versatile in its application across various sample sizes. Moreover, in the process of selecting input variables, GP autonomously identifies the variables that carry the most significance in contributing to the model.

In this study, GP is employed instead of ANN sign estimation for predicting the sign of the residual, deviating from the approach taken by Hsu and Chen (2003) in developing a residual sign estimator. A forecasting equation can be obtained when the GP model is utilized, in



Source(s): Authors' work

Figure 1. Illustration of a program structured in a genetic programming parse tree



Source(s): Authors' work

Figure 2.
Illustration of
crossover operator in
genetic programming

addition to being able to build a forecasting model for a limited data set. Symmetric mean absolute percentage error (SMAPE) (Kim and Kim, 2016) can be used as the objective function to lower the GP forecasting error.

$$\text{minimize } \frac{100\%}{n} \sum_{t=1}^n \frac{|e_t|}{\left(\left|\hat{x}_1^{(0)}(t)\right| + \left|x_1^{(0)}(t)\right|\right) / 2} \quad (15)$$

where $e_t = \hat{x}_1^{(0)}(t) - x_1^{(0)}(t)$, n stands for the length of the test dataset and $\hat{x}_1^{(0)}(t)$ and $x_1^{(0)}(t)$, respectively, refer to the projected load value and actual load value of the test data. Given that SMAPE is theoretically bound by a number of conditions, including parameter value range and original data, it was chosen as the objective function.

This research tackles predicting the ups and downs (positive and negative signs) of residual series using a two-step GP model. It's like a coin toss: heads (positive) or tails (negative). First, the model introduces a "dummy variable" called $\zeta(t)$. It's like a flag that indicates whether the residual for a specific year (t) is positive ($\zeta(t) = 1$) or negative ($\zeta(t) = 0$). Next, the model predicts the future sign of the residual ($\zeta(t + 1)$) based on the past two signs ($\zeta(t - 1)$) and $\zeta(t)$. It's like making an educated guess about the next coin toss based on the previous two. The model's "brain" (the GP parameters) are listed in Table 1. In essence, this two-stage GP model uses the past to predict the future fluctuations in residuals, ultimately revealing the ups and downs in the data.

The t^{th} year residual's sign, $\sigma(t)$, can be written as follows:

$$\sigma(t) = \begin{cases} -1, & \zeta(t) = 0 \\ 1, & \zeta(t) = 1 \end{cases}, \quad t \geq 1 \quad (16)$$

Thus, with Eqs. (1)–(16), the proposed forecasting approach, referred to as GPGM(1,N) (Multivariate grey convolution model powered with chaos-based multigene genetic programming), may be generated as:

$$\hat{x}_1^{(0)}(t) = \hat{x}_1^{(0)}(t) \Big|_{GMC(1,N)} + \sigma(t)\hat{\varepsilon}_1^{(0)}(t) \Big|_{MGGP} \quad t \geq 1 \tag{17}$$

The flowchart of the proposed prediction model is shown in Figure 3.

3.3 Evaluation criteria

The accuracy of energy predictions plays a crucial role in managing and planning electrical grids. To evaluate how well predicted energy values (denoted as $\hat{x}_1^{(0)}(t)$) match actual values (denoted as $x_1^{(0)}(t)$), statisticians use various criteria. Four common ones are:

- **Mean Absolute Error (MAE) (Eq. (18)):** This measures the average absolute difference between predicted and actual values (de Myttenaere *et al.*, 2016). A lower MAE indicates better accuracy, as it means predictions are closer to reality on average.

$$MAE = \frac{1}{k} \sum_{t=1}^k |e_t| \tag{18}$$

- **Mean Squared Deviation (MSD) (Eq. (19)):** Similar to MAE, MSD calculates the average squared difference between predictions and actual values (Shi *et al.*, 2022). However, it penalizes larger errors more heavily than smaller ones, emphasizing the importance of accurate forecasts for extreme values.

$$MSD = \frac{1}{k} \sum_{t=1}^k (e_t)^2 \tag{19}$$

- **Root Mean Squared Error (RMSE) (Eq. (20)):** This is simply the square root of MSD (Karunasingha, 2022). It provides the error in the same units as the original data, making it easier to interpret.

$$RMSE = \sqrt{MSD} \tag{20}$$

Table 1.
Parameter signs of genetic programming

Parameter	Value
Explanatory factor	$\zeta(t-1)\zeta(t)$
output factor	$\zeta(t+1)$
Function set	$+, -, \div, \log, exp$
Objective function	$min(SMAPE)$
Maximum number of generation	100
Population size	50
Mutation probability	30%
Crossover probability	30%

Source(s): Authors' work

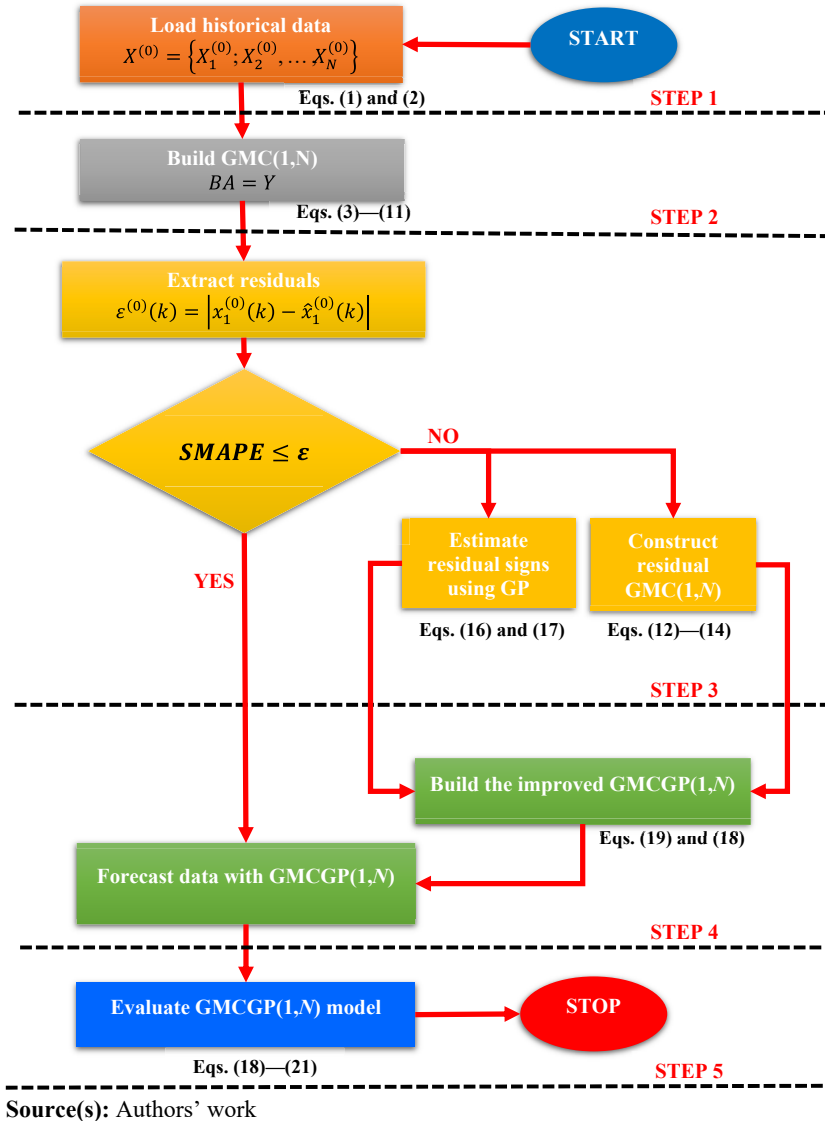


Figure 3. Steps taken to build the proposed GPGMC(1,N) model

- **Coefficient of Determination (R^2)** (Eq. (21)): This measures the proportion of variance in the actual values that can be explained by the predicted values (Taira *et al.*, 2023). A value of 1.0 indicates perfect prediction, while 0.0 implies no correlation. R^2 helps assess how well the model captures the underlying trends in the data.

$$R^2 = 1 - \frac{\frac{1}{k} \sum_{t=1}^k (x_1^{(0)}(t) - \hat{x}_1^{(0)}(t))^2}{\left(x_1^{(0)}(t) - \bar{x}_1^{(0)}(t)\right)^2} \quad (21)$$

By analysing these different criteria, researchers and engineers can gain valuable insights into the strengths and weaknesses of their energy forecasting models. They can then use this information to refine their models and improve the accuracy of their predictions, leading to more efficient and reliable energy systems.

Equations (18)-(21) introduce key criteria for analysing energy prediction accuracy. $x_1^{(0)}(t)$ represents actual consumption, $\bar{x}_1^{(0)}(t)$ is the average of $x_1^{(0)}(t)$. While MAE, MSD, and RMSE are common statistics for evaluating energy forecasts, their interpretations differ.

MAE and RMSE, focusing on the average prediction error, provide similar insights but in different units. Both range from 0 to infinity and don't differentiate between overestimations and underestimations. However, the square root nature of RMSE gives it a unique characteristic: it prioritizes large errors heavily. This makes RMSE particularly valuable when minimizing significant discrepancies is crucial. Interestingly, MAE can be used to set boundaries for RMSE:

- (1) $MAE \leq RMSE$: This guarantees that RMSE will always be greater than or equal to MAE. If all prediction errors are equal, they coincide ($RMSE = MAE$).
- (2) $MAE \cdot \sqrt{k} \geq RMSE$: This inequality reveals the biggest potential difference between the two metrics, occurring when all prediction errors stem from a single sample. In this scenario, RMSE for that sample can be as high as MAE multiplied by the square root of the sample size (k).

In essence, these relationships highlight the trade-offs between MAE and RMSE. While MAE offers a straightforward average error, RMSE emphasizes larger discrepancies, making it a valuable tool for situations where minimizing significant errors is paramount.

4. Simulation results and discussions

This section dives into the simulation results obtained with GPGMC(1,N) model and some alternatives. The modelling process follows three steps. First, we meticulously acquire data on electricity demand and relevant factors. This data undergoes cleaning and transformation to ensure compatibility with the GPGMC(1,N) model. After training the model with historical data, we fine-tune its parameters to optimize performance. Finally, we obtain future values of influencing factors by applying a simple regression of each factor over time because the models do not directly predict the influencing factors.

We built all the models and fed them training data as shown in Figure 3. For the BPNN model, we stuck with the default settings. The effectiveness of these models can significantly differ based on various parameters; however, identifying the optimal ones for forecasting can be a time-consuming process. Therefore, we compare them in a realistic setting where future outcomes are unknown, and the evaluation resources are limited.

We ran each model through 10 simulations and fine-tuned them to fit the data. Then, we gave them a 50-run stability test. The simulations were executed using Matlab R2021a on a personal computer equipped with 8 gigabytes of RAM and an AMD Ryzen 3 3200U processor with Radeon Vega Mobile Gfx operating at 2.60 Hz.

4.1 Dataset and data source

We used a combination of data sources, including: 1) Electricity consumption (in GWh) and subscriber data. These data came directly from [ENEO-Cameroon \(2023\)](#), the national electricity company; 2) Economic data: Information on real income per capita (FCFA/habitant), price of electricity (FCFA/kWh) and final household expenditure (FCFA) were drawn from [World Bank \(2021\)](#) indicators. All these datasets are shown in [Table 2](#) (**Source:** [ENEO-Cameroon \(2023\)](#)).

To ensure accurate results, we divided the data into two sets: 1) Training set, which comprises 70% of the overall dataset (that is from 2001 to 2013). This data was used to build and train the model; 2) Validation set, made up of the remaining 30% (from 2014 to 2019). The splitting ratio 70:30 was chosen in order to achieve a balance between training the model and having enough data for validation. Ultimately, the validated model was utilized to project electricity consumption for the timeframe extending from 2020 to 2030.

4.2 Results and discussion

A rigorous comparison was conducted between the developed GPGMC(1,N) and some competing grey models. ARIMA and BPNN were employed as benchmark models representing alternative non-grey approaches. Additionally, a linear regression (LR) model was implemented to validate the consistency of independent variables with reality. The LR model's analysis of variable signs aligns with real-world expectations. Positive values across final household expenses, number of subscribers and real per capita income reveal a direct link between electricity consumption and these factors (see [Table 3](#)). For instance, rising income translates to higher appliance spending, ultimately driving up household electricity

Year	Electricity demand	Income per capita	Number of subscribers	Average price of electricity	Final household expenditure
2000	3541.0	462778.8	451325	48.29	4.91E+12
2001	3382.5	477282.0	452142	49.06	5.17E+12
2002	3174.2	493389.3	488213	49.93	5.58E+12
2003	3206.7	503295.1	504265	51.93	5.97E+12
2004	3508.8	533537.6	507415	51.86	6.26E+12
2005	3693.2	533734.5	527157	47.68	6.52E+12
2006	3822.2	555381.1	537265	47.55	7.03E+12
2007	3788.1	572278.0	571856	47.24	7.45E+12
2008	4080.2	614275.4	614256	46.47	8.12E+12
2009	3901.5	620786.1	660325	47.79	8.56 E+12
2010	4159.6	636560.5	711214	47.75	9.09 E+12
2011	5336.0	662148.7	707235	48.19	9.53 E+12
2012	5541.0	691571.8	709849	50.29	1.03 E+13
2013	5757.0	723878.4	852024	52.06	1.10 E+13
2014	5994.8	761680.0	887105	52.43	1.19 E+13
2015	5850.9	784835.3	927008	54.93	1.29 E+13
2016	6536.5	808509.7	969360	53.86	1.35 E+13
2017	6785.2	827498.4	1012964	50.18	1.42 E+13
2018	6896.6	852329.2	1015626	50.05	1.51 E+13
2019	6998.4	883231.7	1019021	50.24	1.62 E+13
2020	7215.6	912647.8	1021554	49.47	1.70 E+13

Note(s): Electricity demand is in GWh; Income per capita is in FCFA/habitant; Average price of electricity is in FCFA/kWh while Final household expenditure is in FCFA

Source(s): Authors' work

Table 2.
Datasets used as inputs
in this study

use. Similarly, the sign of the price is in line with the economic theory of supply and demand, according to which a rise in the price of electricity would result in a fall in demand.

Based on the established metrics of MAE, MAD, and RMSE, where lower values indicate superior performance, the GPGMC(1,N) model emerged as the most effective predictor of annual electricity consumption in Cameroon. Table 4 shows that the combination of GP and GMC(1,N) is more effective in forecasting electricity demand. This is why the predictive curve of the GPGMC(1,N) fits the real data almost perfectly (see Figure 4). The results equally show that while GPGMC(1,N) exhibited a slight margin over ARIMA, BPNN, and competing grey models, it's training phase required a longer duration (1263 ms). When the residuals of each model are plotted on the same graph, it is clear that those of the GPGMC(1,N) model are by far the smallest for the entire data set (see Figure 5). This is further proof that the GPGMC(1,N) predictive curve fits the real data well. The graphical representation of the absolute percentage errors (APEs) generated by each model reinforces this last proof of the superiority of the GPGMC(1,N). Indeed, as Figure 6 shows, the APEs of predictions obtained with the GPGMC(1,N) model are the lowest of all in each forecasting period.

Comprehensive analysis revealed that GPGMC(1,N) secured the first position in terms of overall efficiency, exceeding all alternative models. Conversely, GM(1,N)-VAR(p) and BPNN demonstrated the lowest performance, although they compensated for this by requiring minimal training time. Table 4 clearly identifies the most efficient predictive model for each performance metric, with GPGMC(1,N) consistently occupying the top positions. The reasons why GPGMC(1,N) take more time for training is because GP evolves entire programs or structures. This makes fitness evaluation more complex and computationally expensive, as the entire program needs to be executed and its output evaluated against the desired outcome. Additionally, GP can grow in size and complexity as they evolve, further increasing the computational cost of evaluation. Despite this apparent weakness, the new model nevertheless manages to produce excellent results when all the validation criteria are considered.

Table 3.
Summary of estimated coefficients obtained with LR model

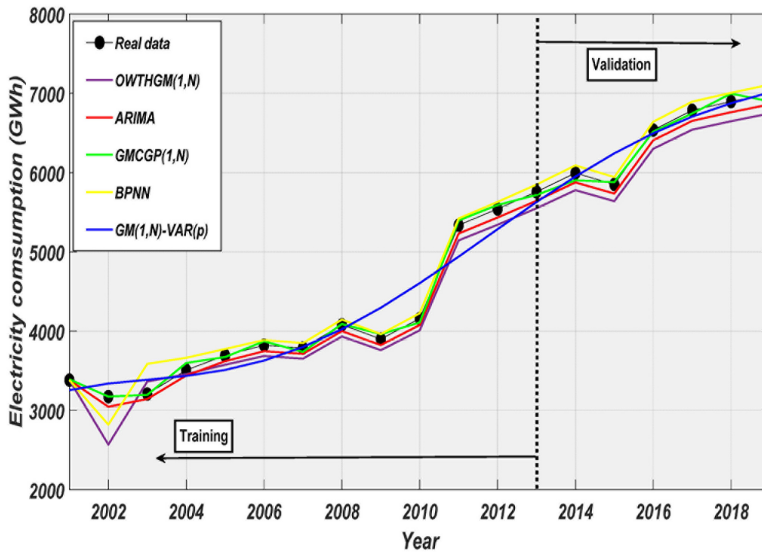
Variables	Coef	t-stat	p-value
Constant	0.0033	4.3280	0.0000
Electricity price	-0.0562	-4.1600	0.0000
Real income per capita	0.0418	3.4480	0.0006
Number of subscribers	0.1251	4.3510	0.0000
Final household expenditure	0.0104	3.8870	0.0001
R^2	98.65%		
Adjusted- R^2	95.23%		

Source(s): Authors' work

Table 4.
Performance statistics obtained during the training and test phases

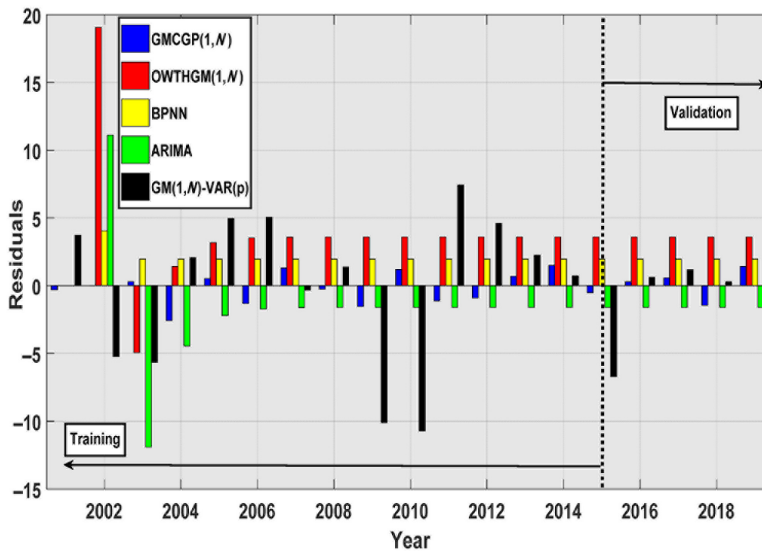
Model	MAE	Training phase			Test phase			
		MSD	RMSE	R^2	MAE	MSD	RMSE	R^2
ARIMA	0.0572	395.62	19.89	0.90	0.0337	287.56	16.96	0.92
BPNN	0.0690	401.24	20.03	0.88	0.0588	429.76	20.73	0.89
OWTHGM(1,N)	0.0651	661.85	25.73	0.87	0.0571	531.20	23.05	0.88
GM(1,N)-VAR(p)	0.0983	944.14	30.73	0.75	0.0970	1086.74	32.97	0.73
GPGMC(1,N)	0.0140*	101.01*	10.05*	0.99*	0.0112*	116.78*	10.81*	0.99*

Note(s): *Indicates the best statistic
Source(s): Authors' work



Source(s): Authors' work

Figure 4.
Fit curves of each
model during the
training and
testing phase



Source(s): Authors' work

Figure 5.
Trends of residuals
over the training
and testing periods
for the entire
dataset

The stability of the GPGMC(1,N) model is a crucial aspect of its performance. To ensure reliability, we employed a rigorous testing procedure that involved multiple simulations and fine-tuning:

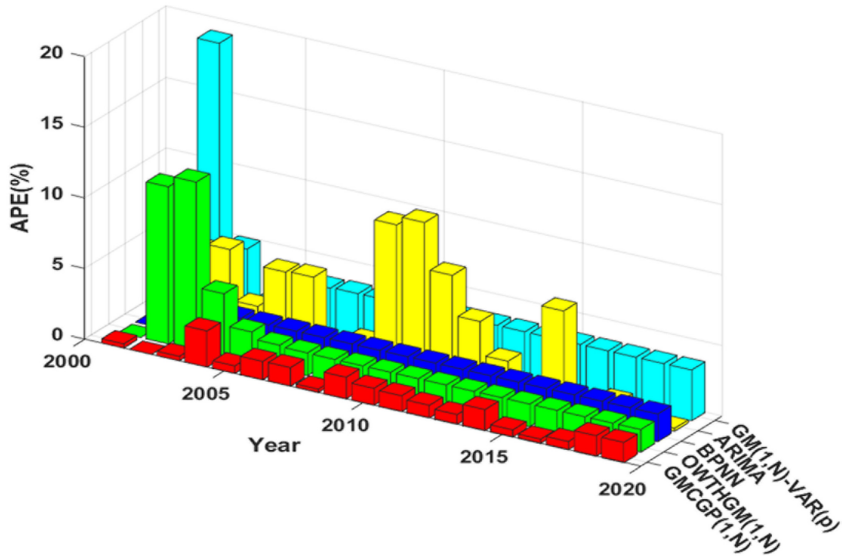


Figure 6.
Absolute percentage
error (APE)
distribution for all
models

Source(s): Authors' work

- (1) **Step 1: Multiple simulations (10 runs):** We repeated the training process for the GPGMC(1,N) model and all competing models ten separate times. This helped account for the inherent variability present in training algorithms. Each run utilized a random split of the data between the training and validation sets.
- (2) **Step 2: Fine-tuning for optimal performance:** Following each training run, we fine-tuned the model parameters to achieve the best possible performance on the validation set. This fine-tuning process helped ensure the models were optimized for the specific characteristics of the electricity demand data in Cameroon.
- (3) **Step 3: Stability testing with 50 runs:** After the training and fine-tuning phases, we conducted a 50-run stability test. In each run, the already optimized model was used to forecast electricity consumption on the validation set (data from 2014 to 2019). By analysing the variation in forecasting performance across these 50 runs, we were able to assess the model's stability.

The results of the stability tests are presented in [Table 5](#). These results suggest the GPGMC(1,N) model demonstrates greater stability compared to the competing models. The GPGMC(1,N) model has a lower average MAE (0.0125) compared to the competing models, indicating its forecasts are generally closer to the actual values on average. Moreover, the standard deviation of MAE for GPGMC(1,N) (0.0010) is also the lowest, suggesting less variation in forecast accuracy across the 50 runs. Finally, the maximum forecast error [3] for GPGMC(1,N) (2.50) is also the lowest, implying a lower probability of significant deviations from actual values in any single test run.

Having confirmed the validity and stability of the GPGMC(1,N) model, we utilized it to predict the electricity demand in Cameroon for the duration spanning 2020 to 2030. The outcomes of the forecast are detailed in [Table 6](#). It can be seen that electricity demand in Cameroon will rise from 7659.9 GWh in 2020–10797.5 GWh in 2030, which represents a record

level of demand for the country. It is therefore necessary for public authorities and policy makers to take measures in order to meet this demand over the coming years.

5. Significance of this study and implications for policy

The results of this study raise challenges to ensure reliable and affordable electricity access for all, but also open a window of opportunity to drive economic growth and foster sustainable development. Major challenges, opportunities and implication for policy are discussed in this section.

5.1 Challenges

Cameroon is a developing country with a growing population and a booming economy, with a population growth rate of 2.6% and GDP growth of 3.54%. This demographic and economic growth is leading to an increase in demand for electricity. Recent studies have revealed that demand for electricity in Cameroon is expected to continue to increase over the next years (Guefano *et al.*, 2021; Dieudonne *et al.*, 2022; Sapnken *et al.*, 2023b). This increase in electricity demand presents a number of challenges for the Cameroonian government. Firstly, it will require significant investment in electricity generation, transmission and distribution infrastructure. This includes the expansion of existing power plants, the construction of new power stations (potentially from renewable energy sources) and the modernization of the electricity network to cope with the increased load (SND30, 2020).

Secondly, the growth in demand for electricity could jeopardize the country's energy security. Cameroon currently relies heavily on hydroelectricity, with more than 60% of its (Tamba *et al.*, 2022), which is a renewable energy source but can be vulnerable to droughts and climate change. Diversifying the energy mix with more resilient sources, such as solar and wind power, can improve the country's energy security and reduce its dependence on imported fossil fuels (Sapnken *et al.*, 2024).

Model	Mean MAE (all runs)	Standard deviation of MAE	Maximum forecast error (random run)
GPGMC(1,N)	0.0125	0.0010	2.500
ARIMA	0.0150	0.0025	3.00
BPNN	0.0130	0.0030	5.025
OWTHGM(1,N)	0.0139	0.0038	4.618
GM(1,N)-VAR(p)	0.0603	0.0104	9.022

Source(s): Authors' work

Table 5.
Results of
stability tests

Year	Electricity demand	Year	Electricity demand
2020	7659.9	2026	9601.3
2021	7799.8	2027	9789.8
2022	8368.1	2028	10209.6
2023	8479.5	2029	10543.6
2024	8962.9	2030	10797.5
2025	9208.6		

Note(s): Electricity demand values are in GWh

Source(s): Authors' work

Table 6.
Cameroon's forecast
demand for electricity
from 2020 to 2030
based on GPGMC(1,N)

Finally, the rising cost of electricity can have a negative impact on the lives of Cameroon citizens, particularly those on low incomes. It is important to put in place policies to mitigate this impact, such as subsidies targeted at low-income households or programmes to promote energy efficiency (Jacques Fotso *et al.*, 2023).

5.2 Opportunities

Despite these challenges, the growth in demand for electricity also presents opportunities for Cameroon. Firstly, increasing access to electricity can support economic growth and job creation (Ayuketah *et al.*, 2023). Businesses and industries need electricity to run, and access to electricity can stimulate productivity and innovation. Secondly, investment in renewable energy and energy efficiency can create new jobs in sectors such as installation, maintenance and research. These jobs can be particularly beneficial for young people and women. Finally, the transition to renewable energy sources can help to protect the environment (Kouer and Meukam, 2023). Renewable energies do not produce greenhouse gases, which can help combat climate change.

5.3 Policy considerations

To meet the challenges and seize the opportunities presented by the growth in demand for electricity, the Cameroonian government needs to take a number of policy measures. These include.

- (1) The development of a long-term national energy plan that sets out strategies to meet future demand, diversify the energy mix and ensure affordability.
- (2) The implementation of regulatory frameworks that encourage investment in renewable energy and energy efficiency, attract private sector participation and ensure fair competition in the energy sector.
- (3) The implementation of demand-side management programmes that encourage consumers and businesses to adopt energy-saving practices and technologies, reducing peak demand and improving grid stability.
- (4) Developing social safety net programmes to protect vulnerable populations from the impact of rising electricity costs.

By taking proactive steps to address the challenges and opportunities presented by growing electricity demand, Cameroon can ensure a sustainable and prosperous future for its citizens.

6. Conclusions and future work

A new hybrid grey prediction approach powered by genetic programming, abbreviated GPGMC(1,N), has been proposed in this paper. This approach combines the advantages of grey prediction models, which are capable of handling limited and non-normal data, with the advantages of genetic programming, which can be used to improve forecast accuracy. More specifically, GP allows the signs of the forecast residuals to be estimated more rigorously. This allows the GMC(1,N) model, on the one hand, to better capture long-term trends in the data and, on the other hand, to better capture short-term variations in the data.

The hybrid GPGMC(1,N) approach was applied to real electricity demand data in Cameroon. The results showed that this combination is able to improve the accuracy of electricity demand forecasts compared to traditional grey prediction models. The results have important implications for research and practice in electricity demand forecasting. They suggest that hybrid approaches that combine the advantages of different types of prediction

models can be used to improve forecast accuracy. Future research could explore the use of the approach for electricity demand forecasting in other contexts. It could also explore the use of other optimization techniques, such as ML, to further improve forecast accuracy.

The success of GPGMC(1,N) in electricity forecasting could also open doors for its application in diverse fields with limited data and volatile fluctuations. Potential areas include water demand forecasting for efficient water resource management, traffic congestion prediction for optimized traffic flow control, and even stock market trend analysis for informed investment decisions. The data-sparseness resilience and the adaptive learning capabilities of GP make GPGMC(1,N) a promising tool for enhancing forecasting accuracy across various domains.

Notes

1. The work of (Tien, 2012) offers a comprehensive elucidation of the process used to derive the solution.
2. One alternative approach for computing the convolution integral in Eq. (10) involves the utilization of high-precision numerical integration techniques, such as the Gaussian formula as discussed by (Ma and Liu, 2016).
3. This metric indicates the largest deviation between a predicted and actual value in any given run. A stable model will not exhibit drastic fluctuations in the maximum error across the tests.

References

- Abdulrahman, M.L., Ibrahim, K.M., Gital, A.Y., Zambuk, F.U., Ja'afaru, B., Yakubu, Z.I. and Ibrahim, A. (2021), "A review on deep learning with focus on deep recurrent neural network for electricity forecasting in residential building", *Procedia Computer Science*, Vol. 193, pp. 141-154, doi: [10.1016/j.procs.2021.10.014](https://doi.org/10.1016/j.procs.2021.10.014).
- Akay, D. and Atak, M. (2007), "Grey prediction with rolling mechanism for electricity demand forecasting of Turkey", *Energy*, Vol. 32 No. 9, pp. 1670-1675, doi: [10.1016/j.energy.2006.11.014](https://doi.org/10.1016/j.energy.2006.11.014).
- Atalay, S.D., Calis, G., Kus, G. and Kuru, M. (2019), "Performance analyses of statistical approaches for modeling electricity consumption of a commercial building in France", *Energy and Buildings*, Vol. 195, pp. 82-92, doi: [10.1016/j.enbuild.2019.04.035](https://doi.org/10.1016/j.enbuild.2019.04.035).
- Ayuketah, Y., Gyamfi, S., Diawuo, F.A. and Dagoumas, A.S. (2023), "A techno-economic and environmental assessment of a low-carbon power generation system in Cameroon", *Energy Policy*, Vol. 179, 113644, doi: [10.1016/j.enpol.2023.113644](https://doi.org/10.1016/j.enpol.2023.113644).
- Bilgili, M. and Pinar, E. (2023), "Gross electricity consumption forecasting using LSTM and SARIMA approaches: a case study of Türkiye", *Energy*, Vol. 284, 128575, doi: [10.1016/j.energy.2023.128575](https://doi.org/10.1016/j.energy.2023.128575).
- Castelli, M., Vanneschi, L. and De Felice, M. (2015), "Forecasting short-term electricity consumption using a semantics-based genetic programming framework: the South Italy case", *Energy Economics*, Vol. 47, pp. 37-41, doi: [10.1016/j.eneco.2014.10.009](https://doi.org/10.1016/j.eneco.2014.10.009).
- de Myttenaere, A., Golden, B., Le Grand, B. and Rossi, F. (2016), "Mean absolute percentage error for regression models", *Neurocomputing*, Vol. 192, pp. 38-48, doi: [10.1016/j.neucom.2015.12.114](https://doi.org/10.1016/j.neucom.2015.12.114).
- Deng, J.-L. (1982), "Control problems of grey systems", *Systems and Control Letters*, Vol. 1 No. 5, pp. 288-294, doi: [10.1016/S0167-6911\(82\)80025-X](https://doi.org/10.1016/S0167-6911(82)80025-X).
- Dieudonne, N.T., Armel, T.K.F., Vidal, A.K.C. and Rene, T. (2022), "Prediction of electrical energy consumption in Cameroon through econometric models", *Electric Power Systems Research*, Vol. 210, 108102, doi: [10.1016/j.epsr.2022.108102](https://doi.org/10.1016/j.epsr.2022.108102).
- Ding, S. and Li, R. (2020), "A new multivariable grey convolution model based on simpson's rule and its applications", *Complexity*, Vol. 2020, pp. 1-14, doi: [10.1155/2020/4564653](https://doi.org/10.1155/2020/4564653).

- ENEO-Cameroon (2023), "Electricity rates", *Decision by ARSEL, to Set New Electricity Tariffs*, available at: <https://eneocameroon.cm/> (accessed 6 December 2023).
- Gil-Gala, F.J., Sierra, M.R., Mencía, C. and Varela, R. (2023), "Surrogate model for memetic genetic programming with application to the one machine scheduling problem with time-varying capacity", *Expert Systems with Applications*, Vol. 233, 120916, doi: [10.1016/j.eswa.2023.120916](https://doi.org/10.1016/j.eswa.2023.120916).
- Guefano, S., Tamba, J.G., Azong, T.E.W. and Monkam, L. (2021), "Forecast of electricity consumption in the Cameroonian residential sector by Grey and vector autoregressive models", *Energy*, Vol. 214, 118791, doi: [10.1016/j.energy.2020.118791](https://doi.org/10.1016/j.energy.2020.118791).
- Haq, E.U., Huang, J., Xu, H., Li, K. and Ahmad, F. (2021), "A hybrid approach based on deep learning and support vector machine for the detection of electricity theft in power grids", *Energy Reports*, Vol. 7, pp. 349-356, doi: [10.1016/j.egy.2021.08.038](https://doi.org/10.1016/j.egy.2021.08.038).
- Hsu, L.-C. (2003), "Applying the grey prediction model to the global integrated circuit industry", *Technological Forecasting and Social Change*, Vol. 70 No. 6, pp. 563-574, doi: [10.1016/S0040-1625\(02\)00195-6](https://doi.org/10.1016/S0040-1625(02)00195-6).
- Hsu, C.-C. and Chen, C.-Y. (2003), "Applications of improved grey prediction model for power demand forecasting", *Energy Conversion and Management*, Vol. 44 No. 14, pp. 2241-2249, doi: [10.1016/S0196-8904\(02\)00248-0](https://doi.org/10.1016/S0196-8904(02)00248-0).
- Huang, J.-J., Tzeng, G.-H. and Ong, C.-S. (2006), "Two-stage genetic programming (2SGP) for the credit scoring model", *Applied Mathematics and Computation*, Vol. 174 No. 2, pp. 1039-1053, doi: [10.1016/j.amc.2005.05.027](https://doi.org/10.1016/j.amc.2005.05.027).
- Jacques Fotso, W., Mvogo, G. and Bidiassé, H. (2023), "Household access to the public electricity grid in Cameroon: analysis of connection determinants", *Utilities Policy*, Vol. 81, 101514, doi: [10.1016/j.jup.2023.101514](https://doi.org/10.1016/j.jup.2023.101514).
- Kapoor, G. and Wichitaksorn, N. (2023), "Electricity price forecasting in New Zealand: a comparative analysis of statistical and machine learning models with feature selection", *Applied Energy*, Vol. 347, 121446, doi: [10.1016/j.apenergy.2023.121446](https://doi.org/10.1016/j.apenergy.2023.121446).
- Karunasingha, D.S.K. (2022), "Root mean square error or mean absolute error? Use their ratio as well", *Information Sciences*, Vol. 585, pp. 609-629, doi: [10.1016/j.ins.2021.11.036](https://doi.org/10.1016/j.ins.2021.11.036).
- Kim, S. and Kim, H. (2016), "A new metric of absolute percentage error for intermittent demand forecasts", *International Journal of Forecasting*, Vol. 32 No. 3, pp. 669-679, doi: [10.1016/j.ijforecast.2015.12.003](https://doi.org/10.1016/j.ijforecast.2015.12.003).
- Kouer, J.P. and Meukam, P. (2023), "Power generation scenarios for Cameroon: valorisation of biomass for the reduction of electricity transmission and the mitigation of greenhouse gas emissions by 2050", *Process Safety and Environmental Protection*, Vol. 180, pp. 487-510, doi: [10.1016/j.psep.2023.10.022](https://doi.org/10.1016/j.psep.2023.10.022).
- Koza, J.R. (1994), "Genetic programming as a means for programming computers by natural selection", *Statistics and Computing*, Vol. 4 No. 2, doi: [10.1007/BF00175355](https://doi.org/10.1007/BF00175355).
- Lei, D., Li, T., Zhang, L., Liu, Q. and Li, W. (2024), "A novel time-delay neural grey model and its applications", *Expert Systems with Applications*, Vol. 238, 121673, doi: [10.1016/j.eswa.2023.121673](https://doi.org/10.1016/j.eswa.2023.121673).
- Li, Y., Wu, K. and Liu, J. (2023), "Self-paced ARIMA for robust time series prediction", *Knowledge-Based Systems*, Vol. 269, 110489, doi: [10.1016/j.knosys.2023.110489](https://doi.org/10.1016/j.knosys.2023.110489).
- Liu, S. and Lin, Y. (2011), *Grey Systems: Theory and Applications*, 1st ed., Springer-Verlag, Berlin, Heidelberg.
- Ma, X. and Liu, Z. (2016), "Research on the novel recursive discrete multivariate grey prediction model and its applications", *Applied Mathematical Modelling*, Vol. 40 Nos 7-8, pp. 4876-4890, doi: [10.1016/j.apm.2015.12.021](https://doi.org/10.1016/j.apm.2015.12.021).
- Min, G.U.O., Jinhui, L.A.N., Juanjuan, L.I., Zongshu, L.I.N. and Xinrong, S.U.N. (2012), "Traffic flow data recovery algorithm based on gray residual GM (1, N) model", *Journal of Transportation Systems Engineering and Information Technology*, Vol. 12 No. 1, pp. 42-47, doi: [10.1016/S1570-6672\(11\)60183-9](https://doi.org/10.1016/S1570-6672(11)60183-9).

- Munkhammar, J., van der Meer, D. and Widén, J. (2021), "Very short term load forecasting of residential electricity consumption using the Markov-chain mixture distribution (MCM) model", *Applied Energy*, Vol. 282, 116180, doi: [10.1016/j.apenergy.2020.116180](https://doi.org/10.1016/j.apenergy.2020.116180).
- Nyathi, T. and Pillay, N. (2018), "Comparison of a genetic algorithm to grammatical evolution for automated design of genetic programming classification algorithms", *Expert Systems with Applications*, Vol. 104, pp. 213-234, doi: [10.1016/j.eswa.2018.03.030](https://doi.org/10.1016/j.eswa.2018.03.030).
- Ong, C.-S., Huang, J.-J. and Tzeng, G.-H. (2005), "Building credit scoring models using genetic programming", *Expert Systems with Applications*, Vol. 29 No. 1, pp. 41-47, doi: [10.1016/j.eswa.2005.01.003](https://doi.org/10.1016/j.eswa.2005.01.003).
- Qian, W. and Sui, A. (2021), "A novel structural adaptive discrete grey prediction model and its application in forecasting renewable energy generation", *Expert Systems with Applications*, Vol. 186, 115761, doi: [10.1016/j.eswa.2021.115761](https://doi.org/10.1016/j.eswa.2021.115761).
- Quartey-Papafio, T.K., Javed, S.A. and Liu, S. (2020), "Forecasting cocoa production of six major producers through ARIMA and grey models", *Grey Systems: Theory and Application*, Vol. 11 No. 3, pp. 434-462, doi: [10.1108/GS-04-2020-0050](https://doi.org/10.1108/GS-04-2020-0050).
- San Cristóbal, J.R., Correa, F., González, M.A., de Navamuel, E.D.R., Madariaga, E., Ortega, A., López, S. and Trueba, M. (2015), "A residual grey prediction model for predicting s-curves in projects", *Procedia Computer Science*, Vol. 64, pp. 586-593, doi: [10.1016/j.procs.2015.08.570](https://doi.org/10.1016/j.procs.2015.08.570).
- Sapnken, F.E. and Tamba, J.G. (2022), "Petroleum products consumption forecasting based on a new structural auto-adaptive intelligent grey prediction model", *Expert Systems with Applications*, Vol. 203, 117579, doi: [10.1016/j.eswa.2022.117579](https://doi.org/10.1016/j.eswa.2022.117579).
- Sapnken, F.E., Ahmat, K.A., Boukar, M., Nyobe, S.L.B. and Tamba, J.G. (2023a), "Learning latent dynamics with a grey neural ODE prediction model and its application", *Grey Systems: Theory and Application*, Vol. 13 No. 3, pp. 488-516, doi: [10.1108/gS-12-2022-0119](https://doi.org/10.1108/gS-12-2022-0119).
- Sapnken, F.E., Hamaidi, M., Hamed, M.M., Hassane, A.I. and Tamba, J.G. (2023b), "An optimal wavelet transform grey multivariate convolution model to forecast electricity demand: a novel approach", *Grey Systems: Theory and Application*, Vol. 14 No. 2, pp. 233-262, doi: [10.1108/gS-09-2023-0090](https://doi.org/10.1108/gS-09-2023-0090).
- Sapnken, F.E., Hamed, M.M., Soldo, B. and Gaston Tamba, J. (2023c), "Modeling energy-efficient building loads using machine-learning algorithms for the design phase", *Energy and Buildings*, Vol. 283, 112807, doi: [10.1016/j.enbuild.2023.112807](https://doi.org/10.1016/j.enbuild.2023.112807).
- Sapnken, F.E., Posso, F., Kibong, M.T. and Tamba, J.G. (2024), "The potential of green hydrogen fuel as an alternative in Cameroon's road transport sector", *International Journal of Hydrogen Energy*, Vol. 49, pp. 433-449, doi: [10.1016/j.ijhydene.2023.08.339](https://doi.org/10.1016/j.ijhydene.2023.08.339).
- Shi, L., Shen, L. and Chen, B. (2022), "Complementary mean square deviation and stability analyses of the widely linear recursive least squares algorithm", *Digital Signal Processing*, Vol. 122, 103357, doi: [10.1016/j.dsp.2021.103357](https://doi.org/10.1016/j.dsp.2021.103357).
- SND30 (2020), *Stratégie Nationale de Développement 2020-2030 : Pour la transformation structurelle et le développement inclusif*, 1st ed., MINEPAT, Yaoundé.
- Taira, K., McInnes, D. and Zhang, L. (2023), "How many data points and how large an R-squared value is essential for Arrhenius plots?", *Journal of Catalysis*, Vol. 419, pp. 26-36, doi: [10.1016/j.jcat.2023.01.033](https://doi.org/10.1016/j.jcat.2023.01.033).
- Tamba, J.G., Essiane, N., Sapnken, E.F., Koffi, F.D., Nsouand, J.L., Soldo, B. and Njomo, D. (2018), "Forecasting natural gas: a literature survey", *International Journal of Energy Economics and Policy, Econjournals*, Vol. 8 No. 3, p. 216.
- Tamba, J.G., Sapnken, F.E., Azong, T.W.E., Guefano, S., Lele, A.F. and Monkam, L. (2022), "An overview of electricity in Cameroon: current status, influential factors and government actions", *International Journal of Energy Economics and Policy*, Vol. 12 No. 4, pp. 470-481, doi: [10.32479/ijeep.13024](https://doi.org/10.32479/ijeep.13024).
- Tarmanini, C., Sarma, N., Gezegin, C. and Ozgonenel, O. (2023), "Short term load forecasting based on ARIMA and ANN approaches", *Energy Reports*, Vol. 9, pp. 550-557, doi: [10.1016/j.egy.2023.01.060](https://doi.org/10.1016/j.egy.2023.01.060).

- Tien, T.-L. (2012), "A research on the grey prediction model GM (1, n)", *Applied Mathematics and Computation*, Vol. 218 No. 9, pp. 4903-4916, doi: [10.1016/j.amc.2011.10.055](https://doi.org/10.1016/j.amc.2011.10.055).
- Ugembe, M.A., Brito, M.C. and Inglesi-Lotz, R. (2023), "Electricity access and unreliability in the creation of sustainable livelihoods in Mozambique", *Energy for Sustainable Development*, Vol. 77, 101330, doi: [10.1016/j.esd.2023.101330](https://doi.org/10.1016/j.esd.2023.101330).
- Ungureanu, S., Topa, V. and Cziker, A.C. (2021), "Deep learning for short-term load forecasting— industrial consumer case study", *Applied Sciences*, Vol. 11 No. 21, 10126, MDPI, doi: [10.3390/app112110126](https://doi.org/10.3390/app112110126).
- Wang, Y., Sun, S., Chen, X., Zeng, X., Kong, Y., Chen, J., Guo, Y. and Wang, T. (2021), "Short-term load forecasting of industrial customers based on SVM and XGBoost", *International Journal of Electrical Power and Energy Systems*, Vol. 129, 106830, doi: [10.1016/j.ijepes.2021.106830](https://doi.org/10.1016/j.ijepes.2021.106830).
- Wang, M., Wang, W. and Wu, L. (2022a), "Application of a new grey multivariate forecasting model in the forecasting of energy consumption in 7 regions of China", *Energy*, Vol. 243, 123024, doi: [10.1016/j.energy.2021.123024](https://doi.org/10.1016/j.energy.2021.123024).
- Wang, Y., Zhang, Y., Nie, R., Chi, P., He, X. and Zhang, L. (2022b), "A novel fractional grey forecasting model with variable weighted buffer operator and its application in forecasting China's crude oil consumption", *Petroleum*, Vol. 8 No. 2, pp. 139-157, doi: [10.1016/j.petlm.2022.03.002](https://doi.org/10.1016/j.petlm.2022.03.002).
- Wang, Y., Sun, L., Yang, R., He, W., Tang, Y., Zhang, Z., Wang, Y. and Sapnken, F.E. (2023), "A novel structure adaptive fractional derivative grey model and its application in energy consumption prediction", *Energy*, Vol. 282, 128380, doi: [10.1016/j.energy.2023.128380](https://doi.org/10.1016/j.energy.2023.128380).
- Wazirali, R., Yaghoubi, E., Abujazar, M.S.S., Ahmad, R. and Vakili, A.H. (2023), "State-of-the-art review on energy and load forecasting in microgrids using artificial neural networks, machine learning, and deep learning techniques", *Electric Power Systems Research*, Vol. 225, 109792, doi: [10.1016/j.epr.2023.109792](https://doi.org/10.1016/j.epr.2023.109792).
- World Bank (2021), "World development indicators | DataBank", available at: <https://databank.worldbank.org/source/world-development-indicators> (accessed 27 May 2022).
- Wu, L., Gao, X., Xiao, Y., Yang, Y. and Chen, X. (2018), "Using a novel multi-variable grey model to forecast the electricity consumption of Shandong Province in China", *Energy*, Vol. 157, pp. 327-335, doi: [10.1016/j.energy.2018.05.147](https://doi.org/10.1016/j.energy.2018.05.147).
- Xie, N. and Wang, R. (2017), "A historic review of grey forecasting models", *Journal of Grey System*, Vol. 29 No. 4, pp. 1-29.
- Ye, J., Li, Y., Meng, F. and Geng, S. (2024), "A novel multivariate time-lag discrete grey model based on action time and intensities for predicting the productions in food industry", *Expert Systems with Applications*, Vol. 238, 121627, doi: [10.1016/j.eswa.2023.121627](https://doi.org/10.1016/j.eswa.2023.121627).
- Yildiz, B., Bilbao, J.I. and Sproul, A.B. (2017), "A review and analysis of regression and machine learning models on commercial building electricity load forecasting", *Renewable and Sustainable Energy Reviews*, Vol. 73, pp. 1104-1122, doi: [10.1016/j.rser.2017.02.023](https://doi.org/10.1016/j.rser.2017.02.023).
- Yin, C. and Mao, S. (2023), "Fractional multivariate grey Bernoulli model combined with improved grey wolf algorithm: application in short-term power load forecasting", *Energy*, Vol. 269, 126844, doi: [10.1016/j.energy.2023.126844](https://doi.org/10.1016/j.energy.2023.126844).
- Yin, H., Tang, Z. and Yang, C. (2023), "Predicting hourly electricity consumption of chillers in subway stations: a comparison of support vector machine and different artificial neural networks", *Journal of Building Engineering*, Vol. 76, 107179, doi: [10.1016/j.job.2023.107179](https://doi.org/10.1016/j.job.2023.107179).
- Zhao, X., Ma, X., Cai, Y., Yuan, H. and Deng, Y. (2023), "Application of a novel hybrid accumulation grey model to forecast total energy consumption of Southwest Provinces in China", *Grey Systems: Theory and Application*, Vol. 13 No. 4, pp. 629-656, doi: [10.1108/gs-02-2023-0013](https://doi.org/10.1108/gs-02-2023-0013).

Appendix

According to the results derived in Section 3.1, we can draw two major analyses, which are as follows:

- If the grey system is a single-variable model, meaning when N equals 1, the second component on the right-hand side of Eq. (5) diminishes, leading to a simplified model:

$$\frac{dx_1^{(1)}(t)}{dt} + \alpha x_1^{(1)}(t) = u$$

This particular differential equation corresponds to the image equation of the traditional GM(1,1) model. By fixing $f(\tau)$ to a fixed value u , we are able to derive the following from Eq. (10):

$$\begin{aligned} x_1^{(1)}(1) &= x_1^{(0)}(1)e^{\alpha(1-t)} + \int_{\tau=1}^{\tau=t} e^{\alpha(\tau-t)} u d\tau \\ &= x_1^{(0)}(1)e^{\alpha(1-t)} + \frac{u}{\alpha} (e^{\alpha t} - e^{\alpha}) e^{-\alpha t} \\ &= \frac{u}{\alpha} + \left(x_1^{(0)} - \frac{u}{\alpha}\right) e^{\alpha(1-t)} \end{aligned}$$

This serves as the time response function for the GM(1,1) model. The parameters α and u can still be determined through the least-squares method outlined in Eq. (7). However, the matrix B is altered, leading to:

$$B = \begin{pmatrix} -z_1^{(0)}(2) & 1 \\ -z_1^{(0)}(3) & 1 \\ \vdots & \vdots \\ -z_1^{(0)}(n) & 1 \end{pmatrix} \in \mathbb{R}^{(n-1) \times 2} \tag{A.1}$$

- If the model parameter $u = 0$, the GMC(1,N) model presented by Eq. (5) reduces to Eq. (A.2).

$$\frac{dx_1^{(1)}(t)}{dt} + \alpha x_1^{(1)}(t) = \sum_{i=2}^N \beta_i x_i^{(1)}(t) \tag{A.2}$$

The essential formula for the GM(1,N) model is represented by Eq. (13). In the earlier studies conducted by Liu and Lin (2011, pages 107–147), it was established that the RHS of Eq. (A.2) remains a constant. The temporal response function of the multivariate GM(1,N) model can be accurately derived from the traditional GM(1,1) model as:

$$\hat{x}_1^{(1)}(1) = \left(x_1^{(0)}(t) - \frac{1}{\alpha} \sum_{i=2}^N \beta_i x_i^{(1)}(t)\right) e^{\alpha(t-1)} + \frac{1}{\alpha} \sum_{i=2}^N \beta_i x_i^{(1)}(t).$$

Unlike the GMC(1,N) model, the GM(1,N) model features one less parameter in the matrix $\tilde{A} = [\alpha, \beta_2, \dots, \beta_N]^T$. The values of these parameters can also be determined using the least-squares method. B also demonstrates a distinction, namely:

$$\tilde{B} = \begin{pmatrix} -z_1^{(1)}(2) & x_2^{(1)}(2) & \dots & x_N^{(1)}(2) \\ -z_1^{(1)}(3) & x_2^{(1)}(3) & \dots & x_N^{(1)}(3) \\ \vdots & \vdots & \ddots & \vdots \\ -z_1^{(1)}(n) & x_2^{(1)}(n) & \dots & x_N^{(1)}(n) \end{pmatrix} \in \mathbb{R}^{(n-1) \times N}$$

The preceding analysis unmistakably indicates that GMC(1,N) outperforms the conventional GM(1,N) model in the specified domains:

- (1) The configuration of the GMC(1,N) model is consistent with the classical GM(1,1), while the standard GM(1,N) cannot be transformed into the GM(1,1) model.
- (2) Within the traditional GM(1,N) model, the inaccurate treatment of the driving term as a constant is evident, leading to a flawed time response function for the GM(1,N) model.
- (3) When calculating $A = [\alpha, \beta_2, \dots, \beta_N, u]^T$, the GM(1,N) model is more appropriate. In contrast, the GMC(1,N) model incorporates the summation of values from the second column to the N^{th} column of B , while the conventional GM(1,N) model considers these terms as fixed constants.

However, the GMC(1,N) model still has at least three shortcomings, which we mention below:

- (1) In order to estimate $A = [\alpha, \beta_2, \dots, \beta_N, u]^T$ from the GMC(1,N) model, the difference equation (Eq. (6)) is solved using the least squares method. The time response function, however, is produced using Eq. (10) (the first-order differential equation). Eq. (6) and Eq. (10) are roughly equivalent, but they are fundamentally different. Due to this parameter mismatch, the GMC(1,N) model may become unstable.
- (2) Similar to the conventional GM(1,N) model, the GMC(1,N) model also functions as a factor and state model. Nevertheless, the simplicity of the GMC(1,N) model's structure remains a limitation. While several modifications have been introduced to enhance the model's structure, there has been a lack of investigation into the linear influence of time t on its performance. This gap in research may contribute to the model's suboptimal prediction accuracy.
- (3) Finally, a significant portion of studies has overlooked the possibility that the input data might include irrelevant information, thereby causing overfitting or underfitting issues in the prediction stage.

The previous analysis lead to the conclusion that the GMC(1,N) model still has flaws related to parameter estimation mismatch and that it is too basic to handle systems found in real-world settings. These glaring flaws motivate to develop an efficient GMC(1,N) model that fixes them.

Corresponding author

Flavian Emmanuel Sapnken can be contacted at: sapnken.emmanuel@gmail.com

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com