A Tri-Stage Approach for Predicting Construction and Demolition Waste Disposal: Hong Kong Case

Shuaiming Su[1], Xinjie Feng[1], Svetlana Besklubova[2], Muhammad Huzaifa Raza[1], Ray Y. Zhong[1]

[1]Department of Data and Systems Engineering, The University of Hong Kong, Hong Kong
[2] Department of Engineering, University of Cambridge, 7a JJ Thomson Avenue, Cambridge CB 0FA, UK
smsu@hku.hk; fengxinjie0201@connect.hku.hk; sb2837@cam.ac.uk; huzaifa@connect.hku.hk; zhongzry@hku.hk

Abstract - Urbanization-driven construction and demolition activities have caused a global surge in construction and demolition waste, posing severe environmental threats. As a high-density metropolis with intensive construction, Hong Kong faces acute construction and demolition waste management challenges, such as low recycling rates and illegal dumping. Accurate estimation of construction and demolition waste disposal volumes is crucial for proactive planning and resource allocation, yet conventional predictive models often struggle with small sample sizes, multi-factor interference, and the complex nonlinearity inherent in urban waste data. To resolve these issues, this study proposes a three-stage integrated forecasting framework combining the Grey Model, Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX), and Long Short-Term Memory neural networks (LSTM). The framework is validated using comprehensive metrics, including Mean Absolute Percentage Error, Root Mean Square Error, Rolling Prediction Stability, and Model Consistency Index. Results show the combined model outperforms individual models, achieving a Mean Absolute Percentage Error of 2.76%. Forecasts indicate Hong Kong's construction and demolition waste disposal volume will grow from 4,428 tonnes/day in 2023 to 4,710 tonnes/day in 2026. These findings provide a robust data-driven basis for optimizing Hong Kong's C&D waste management strategies, facilitating progress toward green urbanization and carbon neutrality goals.

Keywords –
Construction and Demolition Waste; Forecasting; Grey Model; ARIMAX; LSTM; Waste estimation; Hong Kong

1 Introduction

Urbanization and related construction and demolition (C&D) activities are leading to a rapid increase in C&D waste generation worldwide [1,2]. This surge poses significant environmental and societal challenges, including land resource depletion and water source contamination [3,4]. Research indicates that C&D waste accounts for as much as 40% of overall urban waste produced by extensive building and demolition projects, stemming from accelerated urbanization and urban rejuvenation efforts [5]. In Hong Kong, a high-density international metropolis characterized by frequent construction activities, C&D waste management presents particularly acute challenges [6]. In 2023, the total amount of solid waste disposed in Hong Kong's landfills reached 15,783 tonnes per day, of which 4,428 tonnes (approximately 28.1%) originated from construction activities. Despite the implementation of strict regulatory frameworks over recent decades, including various statutory and non-statutory policies, regulations, codes of practice, and programs, Hong Kong's C&D waste management still confronts significant obstacles [7]. These include low segregation and recycling rates, inadequate resource utilization, and frequent illegal dumping incidents. Moreover, the city's limited land resources and high urban density exacerbate the pressure on existing landfill facilities, making efficient waste management imperative for achieving green urbanization and carbon neutrality goals.

A critical gap in current construction and demolition waste management lies in the lack of reliable predictive tools for disposal volume estimation [8], which is essential for proactive planning and resource allocation. While regulatory frameworks are in place, existing forecasting efforts are constrained by two key issues. First, most related research is overly generalized and fails to accommodate the distinctive operational constraints of high-density urban contexts [9]. Second, even advanced techniques such as machine learning, though already applied in this domain, still struggle to address the small sample sizes, multi-factor interference, and complex nonlinearity inherent in urban waste data [10]. Collectively, these limitations hinder the development of targeted, effective strategies for accurate construction and demolition waste disposal volume prediction.

To address these challenges, this study proposes an integrated forecasting framework that leverages the complementary strengths of multiple predictive models to forecast C&D waste disposal volumes in Hong Kong [11]. The framework addresses three key characteristics inherent in urban solid waste data: small sample size, multifactor influence, and complex nonlinearity. The modeling strategy employs a three-stage approach. First, a Grey Model (GM) is established to capture the fundamental trend of the data, exploiting its advantage in handling limited historical observations [12,13]. Second, an Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model is constructed to integrate exogenous variables such as macroeconomic indicators (inflation rate) and demographic factors (population growth), thereby accounting for external drivers of waste generation. Third, an LSTM (Long Short-Term Memory) neural network model is applied to uncover nonlinear patterns and long-term dependencies within the temporal data [14]. These three sub-models are dynamically weighted through a linear combination mechanism to form an integrated forecasting system that balances interpretability and predictive accuracy.

The evaluation framework extends beyond conventional accuracy metrics such as Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) by introducing innovative stability and consistency indicators. Specifically, Rolling Prediction Stability (RPS) measures the model's resistance to fluctuations across different time windows, while Model Consistency Index (MCI) assesses the alignment of forecasted trends with actual patterns. This multi-angle validation mechanism ensures the scientific rigor and reliability of the forecasting results, providing robust support for decision-making in waste management planning.

This study processes historical C&D waste disposal data from 1999 to 2023, sourced from the Hong Kong Environmental Protection Department (HKEPD), to generate future forecasts extending to 2026 [14]. The results directly support strategic decision-making in Hong Kong's C&D waste management, including planning treatment facilities, optimizing transfer systems, and formulating evidence-based policies. By providing accurate and reliable data-driven insights, this research contributes to the sustainable development of Hong Kong's urban environment and advances the city's progress toward green urbanization and carbon neutrality objectives.

2 Model Development

2.1 Model Selection Rationale

Single predictive models exhibit inherent limitations when applied independently to construction and demolition waste disposal volume forecasting, which cannot be ignored when addressing the characteristics of urban solid waste data. The Grey Model GM(1,1) is highly sensitive to abrupt fluctuations in time series data, making it prone to large prediction deviations when faced with sudden policy adjustments or unforeseen incidents. The ARIMAX model relies on strict linear assumptions, failing to capture the complex nonlinear relationships between C&D waste generation and its driving factors. LSTM neural networks demand sufficiently high-quality training samples to ensure model validity, and their performance deteriorates significantly when dealing with small sample datasets common in urban waste management research.

To overcome these drawbacks, this study adopts a combined modeling strategy that integrates the complementary advantages of GM(1,1), ARIMAX, and LSTM models. The linear weighting-based integrated framework balances the strengths of each sub-model while mitigating their individual limitations, thereby improving the robustness and accuracy of C&D waste disposal volume forecasts. The specific role and value of each sub-model in the integrated system are as follows:

- The GM(1,1) model serves as the base trend predictor, leveraging its strength in processing small sample datasets. It weakens the randomness of raw data through accumulation generation operations, effectively capturing the long-term monotonic growth trend of C&D waste disposal volumes. The posterior variance test of the model also provides strict validation criteria to ensure the rigor of the baseline prediction results.
- The ARIMAX model undertakes the task of integrating exogenous variables, incorporating macroeconomic indicators and demographic factors into the forecasting framework through a dynamic regression structure. It addresses the impact of external drivers on waste generation, enhancing the interpretability of the model. The Akaike Information Criterion is used to optimize model parameters, avoiding subjective bias in parameter setting.
- An LSTM neural network is responsible for mining nonlinear patterns and long-term dependencies in time series data. Its unique gating mechanism effectively solves the gradient vanishing problem of traditional recurrent neural networks, enabling it to identify latent correlations between C&D waste data and factors such as policy changes and unforeseen events. The introduction of dropout layers controls model complexity and prevents overfitting, while bootstrapping technology quantifies prediction uncertainty to provide risk early warning for decision-making.

In summary, this decomposition-integration strategy combines the trend-fitting capability of GM(1,1), the exogenous variable integration capability of ARIMAX, and the nonlinear pattern mining capability of LSTM. The dynamic weighting mechanism objectively assigns weights according to the prediction performance of each sub-model, reducing the probability of extreme prediction deviations and forming a more reliable C&D waste disposal volume forecasting system.

2.2 GM(1,1) Model Development

The Grey prediction model GM(1,1) is a forecasting method specifically designed for systems with small samples and incomplete information. Its core idea is to uncover hidden regularities in raw data through data generation techniques. In this study, the GM(1,1) model is applied to forecast the time series of C&D waste generation volume. It weakens the randomness of the original sequence using accumulation generation technology to reveal inherent exponential change patterns.

Firstly, let the original time series of target data be the specific sequence:

$$x^{(0)} = \left(x^{(0)}(1),\ x^{(0)}(2),\ \ldots,\ x^{(0)}(n)\right) \tag{1}$$

Then, weaken the randomness of the original sequence through first-order accumulation generation (1-AGO):

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), \qquad k = 1,2,\ldots,n \tag{2}$$

This operation transforms a non-negative original sequence into a monotonically increasing sequence, revealing the system's internal pattern. Next, calculate the mean sequence of consecutive neighbors as the background value for the grey differential equation:

$$z^{(1)}(k) = \frac{1}{2}\left[x^{(1)}(k) + x^{(1)}(k-1)\right], \quad k = 2,3,\ldots,n \tag{3}$$

Subsequently, establish the fundamental equation of the GM(1,1) model:

$$x^{(0)}(k) + az^{(1)}(k) = b \tag{4}$$

where a is the development coefficient, and b is the grey action quantity. After that, estimate parameters using the least squares method:

$$\begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y \tag{5}$$

where

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \tag{6}$$

Then, establish the time response function for the accumulated sequence:

$$\hat{x}^{(1)}(k) = \left[x^{(0)}(1) - \frac{b}{a}\right] e^{-a(k-1)} + \frac{b}{a} \tag{7}$$

Following this, obtain the predicted values of the original sequence through inverse accumulation (IAGO):

$$\hat{x}^{(0)}(k) = \begin{cases} x^{(0)}(1) & k = 1 \\ x^{(1)}(k) - \hat{x}^{(1)}(k-1) & k = 2,3,\cdots,n \end{cases} \tag{8}$$

Further, calculate the forecasted value for the next T years:

$$\hat{x}^{(0)}(n+T) = \left[\hat{x}^{(1)}(n+T) - \hat{x}^{(1)}(n+T-1)\right] \tag{9}$$

Finally, evaluate model accuracy using the posterior variance test, including the posterior variance ratio (C) and Small error probability (P):

$$C = \frac{S_2}{S_1} \tag{10}$$

$$P = P(|e(k) - \bar{e}| < 0.6745 S_1) \tag{11}$$

where $e(k) = x^{(0)}(k) - \widehat{x^{(0)}}(k)$ stands for the prediction residual.

2.3 ARIMAX Model Development

The ARIMAX model is an extension of the ARIMA model that integrates exogenous variables to enhance the explanatory and predictive power for complex systems. The ARIMA model, short for Autoregressive Integrated Moving Average model, combines the Autoregressive (AR) and Moving Average (MA) models with a differencing process—the core step to convert a non-stationary time series into a stationary one. It relies solely on the historical data of the time series itself (after stationarization) to forecast its future trends.

The AR model is expressed as:

$$Y_{t+1} = c + \alpha_1 Y_t + \alpha_2 Y_{t-1} + \cdots + \alpha_p Y_{t-p+1} + \varepsilon_{t+1} \tag{12}$$

The MA model is expressed as:

$$Y_{t+1} = u + \varepsilon_{t+1} + \theta\varepsilon_{t+1} \tag{13}$$

$$\varepsilon_t = Y_{t+1} - Y^*_{t+1} \tag{14}$$

The ARIMA model is expressed as

$$Y_{t+1} = c + \alpha_1 Y_t + \alpha_2 Y_{t-1} + \cdots + \alpha_p Y_{t-p+1} + \varepsilon_{t+1} + \varepsilon_{t+1} + \theta\varepsilon_t \tag{15}$$

Since c and ε_{t+1} are both constant terms, they can be combined. The long-term trend prediction function of u is performed by the AR model as a whole, so u does not appear in the equation. Therefore, the final expression of the ARIMA model is derived as:

$$Y_{t+1} = c + \alpha_1 Y_t + \alpha_2 Y_{t-1} + \cdots + \alpha_p Y_{t-p+1} + \theta\varepsilon_t \tag{16}$$

ARIMA model construction mainly includes data stationarity test, stationarity transformation using differencing to transform a non-stationary series into a stationary one, and parameter determination. The ARIMA model requires determining the values of p, d, q. Among them, d can be determined by the order of differencing required to achieve stationarity. p and q can be determined from the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the differenced series. ARIMA model accuracy testing generally employs goodness-of-fit tests, specifically testing whether the error sequence between the original data and the fitted data conforms to white noise (i.e., no significant autocorrelation). If the residual sequence is not white noise, indicating residual autocorrelation exists, the model needs further refinement.

First-order differencing is defined as:

$$\Delta Y_{t+1} = Y_{t+2} - Y_{t+1} \tag{17}$$

Second-order differencing is defined as:

$$\begin{aligned} \Delta Y_{t+1} = \Delta(Y_{t+2} - Y_{t+1}) \\ = (Y_{t+2} - Y_{t+1}) - (Y_{t+1} - Y_t) \\ = Y_{t+2} - 2Y_{t+1} + Y_t \end{aligned} \tag{18}$$

ARIMAX(p, d, q) model is:

$$\emptyset_p(L)(1-L)^d y_t = c + \sum_{i=1}^{k} \beta_i x_{it} + \theta_q(L)\varepsilon_t \tag{19}$$

Specific exogenous variable indicators are selected based on the target situation:

$$X_t = \begin{bmatrix} a_t \\ \vdots \\ z_t \end{bmatrix} \tag{20}$$

The Akaike Information Criterion (AIC) is further used for optimal parameter search:

$$AIC = 2k - 2\ln(\hat{L}) \tag{21}$$

The search space for optimal parameters is as follows: p ∈ {0, 1, 2, 3, 4} (Autoregressive order); d ∈ {0, 1} (Differencing order); q ∈ {0, 1, 2, 3, 4} (Moving average order).

The model requires forecasting and interval estimation. Conditional expectation forecasts and 95% confidence intervals are:

$$\hat{y_t} = E(y_t \mid y_{1:t-1}, X_{1:t}) \tag{22}$$

$$CI_{95\backslash\%} = \hat{y_t} \pm 1.96 \times SE(\hat{y_t}) \tag{23}$$

2.4 LSTM Model Development

LSTM is a special type of Recurrent Neural Network (RNN) specifically designed to solve the vanishing gradient problem in traditional RNNs when processing long sequence data. In C&D waste forecasting, LSTM can effectively capture long-term dependencies and nonlinear patterns in time series, with its core being the gating mechanism.

The LSTM cell regulates information flow through three gating structures (Forget Gate, Input Gate, Output Gate). The Forget Gate determines what information to discard from the cell state:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{24}$$

The Input Gate determines what new information to store in the cell state:

$$\begin{cases} i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C_t} = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \end{cases}. \tag{25}$$

Update the current cell state content to form a new memory:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C_t} \tag{26}$$

The Output Gate determines what information to output from the cell state:

$$\begin{cases} o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t \odot tanh(C_t) \end{cases}. \tag{27}$$

where σ is the Sigmoid activation function outputting values in [0,1], $\odot$ is the Hadamard product (element-wise multiplication), x_t is the network input at time t (normalized C&D waste volume), h_t is the output of the LSTM cell at time t, C_t is the state of the memory cell at time t, and W_f,W_i,W_C,W_o are weight matrices with corresponding bias vectors b_f, b_i, b_C, b_o.

2.5 Combined Forecasting Model

The combined forecasting model integrates the strengths of multiple independent forecasting models to build a more robust and accurate framework. This study employs a linear weighting combined method, fusing the forecasts from the GM(1,1), ARIMAX, and LSTM models to leverage their complementary advantages.

The core mathematical expression of the combined forecasting model is:

$$\hat{y}_t^{comb} = w_0 + \sum_{i=1}^{3} w_i \hat{y}_t^{(i)} \tag{28}$$

Optimal weights are determined by minimizing the Mean Squared Error (MSE) between the combined forecast and the actual values:

$$\min_{w} \sum_{t=1}^{n} \left(y_t - \left(w_0 + \sum_{i=1}^{3} w_i \hat{y}_t^{(i)} \right) \right)^2 \tag{29}$$

The matrix form solution to this optimization problem is:

$$w = (X^T X)^{-1} X^T Y \tag{30}$$

The design matrix X and response vector Y are:

$$X = \begin{bmatrix} 1 & \hat{y}_1^{GM} & \hat{y}_1^{ARIMAX} & \hat{y}_1^{LSTM} \\ 1 & \hat{y}_2^{GM} & \hat{y}_2^{ARIMAX} & \hat{y}_2^{LSTM} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \hat{y}_n^{GM} & \hat{y}_n^{ARIMAX} & \hat{y}_n^{LSTM} \end{bmatrix} \tag{31}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{32}$$

The variance of the combined forecasting model can be expressed as:

$$CI_{95\%} = \hat{y}_t^{comb} \pm 1.96 \times \sqrt{Var(\hat{y}_t^{comb})} \quad (33)$$

where $\hat{y}_t^{comb}$ is the combined forecast value for period t, $\hat{y}_t^{(1)}$ is the GM(1,1) model forecast value (trend dominant), $\hat{y}_t^{(2)}$ is the ARIMAX model forecast value (exogenous variable response), $\hat{y}_t^{(3)}$is the LSTM model forecast value (nonlinear pattern capture).

3 Results

3.1 Evaluation Metrics

To comprehensively assess the forecasting performance of different models, this study establishes a multidimensional evaluation framework encompassing forecast accuracy, stability, and consistency metrics. The Mean Absolute Percentage Error (MAPE) measures the relative deviation between forecasted and actual values, with the calculation formula shown in Equation (34). The evaluation standard indicates that MAPE < 10% represents excellent prediction performance.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (34)$$

The Root Mean Square Error (RMSE) quantifies the average deviation between forecasted and actual values, providing an absolute measure of prediction accuracy as shown in Equation (35).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (35)$$

Beyond accuracy metrics, the Rolling Prediction Stability (RPS) evaluates the volatility of forecasting results across different time windows. The calculation formulas are presented in Equations (36) and (37), where σ_k and μ_k represent the standard deviation and mean of forecast values within window k, respectively. The evaluation standard specifies that RPS < 0.15 indicates high stability.

$$RPS_k = \frac{\sigma_k}{\mu_k} \quad (36)$$

$$RPS = \frac{1}{m}\sum_{k=1}^{m} RPS_k \quad (37)$$

The Model Consistency Index (MCI) measures the synchronicity between forecasted and actual value changes through correlation analysis, as defined in Equations (38) and (39), where ρ denotes the correlation coefficient.

$$MCI_k = \rho(y_k, \hat{y}_k) \quad (38)$$

$$MCI = \frac{1}{m}\sum_{k=1}^{m} MCI_k \quad (39)$$

To synthesize the performance across multiple dimensions, the Model Ranking Index (MRI) aggregates ranking results from all evaluation metrics, as shown in Equation (40), where Rank$j$ (I) represents the ranking of a model on indicator j (with rank 1 indicating the best performance). A smaller MRI value indicates superior overall performance.

$$MRI = \frac{1}{k}\sum_{j=1}^{k} Rank_j(I) \quad (40)$$

3.2 Datasets Introduction

Table 1 Data used in the evaluation

Year	C&D waste volume (tonnes/day)	HK Pop. Annual Growth Rate (%)	HK Inflation Rate (%)	HK Population (person)
1999	7868	0.0096	0	6610000
2000	7475	0.0088	-0.0369	6665000
2001	6408	0.0074	-0.0166	6714300
2002	10202	0.0044	-0.0298	6744100
2003	6728	-0.002	-0.0267	6730800
2004	6595	0.0078	-0.0027	6783500
2005	6556	0.0044	0.0083	6813200
2006	4125	0.0064	0.02	6857100
2007	3158	0.0086	0.0204	6916300
2008	3092	0.006	0.043	6957800
2009	3121	0.0022	0.0058	6972800
2010	3584	0.0073	0.0229	7024200
2011	3331	0.0067	0.0531	7071600
2012	3440	0.011	0.0405	7150100
2013	3591	0.004	0.0434	7178900
2014	3942	0.007	0.0442	7229500
2015	4200	0.0085	0.0299	7291300
2016	4422	0.0062	0.0241	7336600
2017	4207	0.0077	0.0149	7393200
2018	4081	0.008	0.0241	7452600
2019	3946	0.0074	0.0288	7507900
2020	3418	-0.0036	0.0025	7481000
2021	3646	-0.0091	0.0157	7413100
2022	4128	0.008026332	0.0188	7472600
2023	4428	0.004081578	0.017	7503100

This study utilizes C&D waste data (tonnes per day) obtained from the Hong Kong Environmental Protection Department, covering the period from 1999 to 2023, including the information for year, C&D amount, population annual growth rate, inflation rate, and population size, as shown in Table 1. This dataset presents three typical challenges, which are consistent with the previous statement :

1) Small sample size: only 25 annual observations are available;
2) Multi-factor interference: C&D waste is jointly affected by population, economy, and external uncertainties;
3) Complex nonlinearity: the waste generation series shows strong fluctuations and non-stationarity.

In the experiments, GM(1,1) uses only the first two columns, while ARIMAX and LSTM use all variables listed in Table 1.

3.3 Performance Comparison

The performance evaluation results for the four forecasting models are presented in Table **2**. The data from each model's forecasted results were substituted into the evaluation formulas, with a rolling window size of 3 years for RPS and MCI calculations. The results reveal distinct performance characteristics across different modeling approaches.

Table 2. Performance evaluation results for each model

Model	RMSE	MAPE (%)	RPS	MCI
GM(1,1)	1248.38	22.01	0.033	0.188
ARIMAX	1337.83	16.28	0.119	0.270
LSTM	523.64	7.08	0.075	0.897
Combined	195.89	2.76	0.097	0.897

The RMSE and MAPE metrics demonstrate that the LSTM and Combined models significantly outperform traditional statistical approaches. Notably, both LSTM and combined models satisfy the criterion of MAPE $\leqslant$ 10%, indicating superior forecasting accuracy. All models exhibit RPS values below 0.15, confirming high stability and low volatility in their predictions. The MCI values reveal that LSTM and combined models both exceed 0.8, demonstrating strong consistency between forecasted trends and actual waste disposal patterns.

Table 3. Model rankings per metric

Model	RMSE	MAPE (%)	RPS	MCI
GM(1,1)	3	4	1	4
ARIMAX	4	3	4	3
LSTM	2	2	2	2
Combined	1	1	3	1

To facilitate a comprehensive comparison, models were ranked for each metric, with rank 1 indicating the best performance and rank 4 the worst. The specific rankings are presented in Table 3.

Based on the ranking data, the MRI values were calculated as follows: $MRI_{GM(1,1)} = 3.0$, $MRI_{ARIMAX} = 3.5$, $MRI_{LSTM} = 2.0$, and $MRI_{Combined} = 1.5$. The combined forecasting model achieves the lowest MRI value, confirming its superiority in comprehensive performance. This result validates the effectiveness of integrating complementary strengths from GM(1,1), ARIMAX, and LSTM models through a linear weighting framework.

3.4 Forecasting Results for 2024-2026

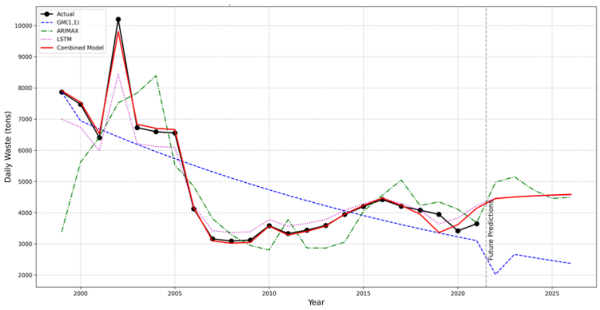

Figure 1. Daily waste disposal prediction results based on the comparison of different models

Utilizing the combined forecasting model identified as the highest quality approach, projections for Hong Kong's C&D waste disposal were generated for the period 2024-2026. The forecasting results indicate a gradual recovery trajectory in construction activities, with disposal volumes increasing from 4,558 tonnes per day in 2024 to 4,651 tonnes per day in 2025, and further to 4,710 tonnes per day in 2026. This represents a total increment of 282 tons per day compared to the 2023 baseline of 4,428 tons per day, reflecting an anticipated 6.4% growth over the three years. The upward trend aligns with expectations of deepening implementation of national economic stimulus policies and the gradual recovery of Hong Kong's construction industry. Figure 1 illustrates the comparative performance of all four models against historical data and future projections, demonstrating the combined model's superior fitting capability and reasonable extrapolation characteristics.

4 Discussions

According to the forecast, Hong Kong's C&D waste disposal volume is projected to rise steadily from 4,428 tons/day in 2023 to 4,710 tons/day in 2026, with an incremental growth of 282 tons/day over three years. This continuous growth poses severe challenges to the city's waste management system, including potential overload risks at core landfills, aggravated cross-district transportation pressure, prolonged waste retention, and associated environmental pollution issues.

To address C&D waste management dilemmas, the Hong Kong government has implemented a series of policies across multiple dimensions. Since 2006, the Waste Disposal (Charges for Disposal of Construction Waste) Regulation has imposed uniform disposal fees on inert and non-inert waste, complemented by a transfer station franchising system, which initially reduced landfill intake by 28% from 2006 to 2010. In spatial planning, the Hong Kong 2030+ plan incorporates landfill expansion into the Northern Metropolis Reclamation Project, authorizing the requisition of buffer zones to alleviate saturation pressure at existing sites. For resource utilization, the Waste Blueprint for Sustainable Use of Resources 2035 sets a 70% C&D waste recovery target, mandating recycled aggregate usage in government projects and issuing relevant technical specifications to foster local recycled material factories.

However, these policies exhibit notable limitations that hinder their long-term effectiveness. The static disposal fee standard has remained unadjusted for over a decade, lagging far behind levels in Singapore and Tokyo, and failing to curb waste growth effectively. Regulatory blind spots persist in rural sporadic construction projects, leading to rampant illegal dumping. Landfill expansion plans conflict with ecological conservation areas such as the Mai Po Wetland, triggering resident litigation and project delays. Additionally, the lack of incentives for private projects to adopt recycled aggregates results in 90% of recycled products relying on government procurement, while misalignment between technical specifications and building codes increases developers' compliance costs.

To tackle these challenges and adapt to the projected waste increment, targeted optimization strategies are imperative. First, establishing a dynamic disposal network with real-time IoT-GPS monitoring can enable adaptive load allocation across the three major landfills, directing incremental waste to underutilized facilities like the South East New Territories Landfill, and implementing regional transfer subsidies to activate idle capacity. Second, reforming the disposal charging system by linking fees to annual inflation rates and expanding coverage to rural small-scale projects, combined with AI-powered violation detection, can strengthen regulatory effectiveness and reduce illegal dumping. Third, spatial-ecological synergistic planning should be prioritized, requiring landfill operators to conduct ecological remediation at a 1:1.2 ratio of filled areas in exchange for permit extensions, alongside mandatory real-time leachate monitoring and slope stability audits. Finally, deploying innovative technologies such as blockchain-based full-chain traceability systems and landfill gas collection facilities, coupled with tax incentives for advanced sorting technologies, can enhance resource recovery efficiency and mitigate the "pollution-warming" positive feedback loop.

In essence, Hong Kong's C&D waste management requires a shift from a passive, expansion-oriented model to a proactive, resource-driven framework, integrating technological innovation, policy adjustment, and ecological protection to achieve sustainable urban development goals.

5 Conclusion

This study proposes a three-stage integrated forecasting framework combining the GM(1,1), ARIMAX, and LSTM models to address the small sample size, multi-factor influence, and complex nonlinearity challenges in forecasting Hong Kong's construction and demolition waste disposal volumes. This framework capitalizes on the complementary strengths of its sub-models: the GM(1,1) model captures the fundamental data trend with limited historical observations; the ARIMAX model incorporates exogenous variables to enhance the interpretability of external drivers; the LSTM model identifies nonlinear patterns and long-term temporal dependencies. These sub-models are integrated via a linear weighting mechanism, with optimal weights determined by minimizing the mean squared error between combined forecasts and actual values.

A multidimensional evaluation framework, including accuracy, stability, and consistency metrics, validates the combined model's superiority. It achieves the lowest Mean Absolute Percentage Error (MAPE = 2.76%) and Root Mean Square Error (RMSE = 195.89 tonnes/day), as well as high stability and consistency. The Model Ranking Index (MRI) further confirms that the combined model outperforms all individual models in comprehensive performance. Forecasting results indicate that Hong Kong's C&D waste disposal volume will rise steadily from 4,428 tonnes/day in 2023 to 4,710 tonnes/day in 2026, representing a 6.4% three-year growth. These findings provide data-driven support for C&D waste management decisions such as treatment facility planning and transfer system optimization.

The proposed framework exhibits good scalability and generalization. Its modular design allows flexible adaptation to cities with different data availability and urban characteristics, while the data-driven weighting mechanism can be readily recalibrated for new contexts. The model is particularly applicable to other high-density metropolises with similar construction and demolition waste generation patterns.

Despite its contributions, this study has certain limitations. The current forecasting framework is built on historical data from 1999 to 2023, and its adaptability to sudden policy adjustments or extreme economic fluctuations requires further validation. Future research can incorporate real-time monitoring data to enhance the dynamic responsiveness of the forecasting model. Extending the framework to other high-density urban contexts will also help verify its regional universality and transferability, providing more generalized support for sustainable C&D waste management in metropolitan areas.

Acknowledgements This work was supported by RGC Research Impact Fund (R7036-22), RGC GRF project (17202124), and Public Policy Research Funding Scheme (PPRFS) (2024.A8.154.24D).

References

[1] Z. Wu, A.T.W. Yu, C.S. Poon, Promoting effective construction and demolition waste management towards sustainable development: A case study of Hong Kong, Sustainable Development 28 (6) (2020) pp. 1713-1724, https://doi.org/10.1002/sd.2119.

[2] W. Lu, V.W.Y. Tam, Construction waste management policies and their effectiveness in Hong Kong: A longitudinal review, Renewable and Sustainable Energy Reviews 23 (2013) pp. 214-223, https://doi.org/10.1016/j.rser.2013.03.007.

[3] S. Su, Y. Yang, S. Besklubova, J. Song, X.T.R. Kong, R.Y. Zhong, Digital Twin-Enabled Reverse Supply Chain for Building Demolition Waste Management, Computers & Industrial Engineering 211 (2026) pp. 111605, https://doi.org/10.1016/j.cie.2025.111605.

[4] W. Zhao, J.L. Hao, G. Gong, T. Fischer, Y. Liu, Applying digital technologies in construction waste management for facilitating sustainability, Journal of Environmental Management 373 (2025) 123560, 10.1016/j.jenvman.2024.123560.

[5] C.S. Poon, A.T.W. Yu, L.H. Ng, On-site sorting of construction and demolition waste in Hong Kong, Resources, Conservation and Recycling 32 (2) (2001) pp. 157-172, https://doi.org/10.1016/S0921-3449(01)00052-0.

[6] E.P.D.o.H. Kong, Monitoring of solid waste in Hong Kong - waste statistics for 2023, 2024. https://www.wastereduction.gov.hk/sites/default/files/resources_centre/waste_statistics/msw2023_eng.pdf (Accessed date:

[7] H. Duan, J. Li, Construction and demolition waste management: China's lessons, Waste Management & Research 34 (5) (2016) pp. 397-398, 10.1177/0734242X16647603.

[8] M.S. Aslam, B.J. Huang, L.F. Cui, Review of construction and demolition waste management in China and USA, Journal of Environmental Management 264 (2020) pp. 13, 110445, http://doi.org/10.1016/j.jenvman.2020.110445.

[9] S. Besklubova, E. Kravchenko, B.Q. Tan, R.Y. Zhong, A feasibility analysis of waste concrete powder recycling market establishment: Hong Kong case, Environmental Impact Assessment Review 103 (2023) pp. 107225, https://doi.org/10.1016/j.eiar.2023.107225.

[10] M. Jafari, E. Mousavi, Machine learning-based prediction of construction and demolition waste generation in developing countries: a case study, Environmental Science and Pollution Research (2024) 10.1007/s11356-024-34527-9.

[11] M.R. Islam, G. Kabir, K.T.W. Ng, S.M. Ali, Yard waste prediction from estimated municipal solid waste using the grey theory to achieve a zero-waste strategy, Environmental Science and Pollution Research 29 (31) (2022) pp. 46859-46874, 10.1007/s11356-022-19178-y.

[12] W. Lu, J. Lou, C. Webster, F. Xue, Z. Bao, B. Chi, Estimating construction waste generation in the Greater Bay Area, China using machine learning, Waste Management 134 (2021) pp. 78-88, 10.1016/j.wasman.2021.08.012.

[13] Y. Yi, J. Liu, M. Cristina Lavagnolo, A. Manzardo, Evaluating the carbon emission reduction in construction and demolition waste management in China, Energy and Buildings 324 (2024) 114932, 10.1016/j.enbuild.2024.114932.

[14] M.S. Jassim, G. Coskuner, N. Sultana, S.M.Z. Hossain, Forecasting domestic waste generation during successive COVID-19 lockdowns by Bidirectional LSTM super learner neural network, Applied soft computing 133 (2023) pp. 109908, 10.1016/j.asoc.2022.109908.