CINEFORUM
ISSN: 0009-7039
Vol. 66. No. S6, 2026 (Special Issue 6)

Comparative Analysis of Regression-Based Forecasting Models for State-Level E-Waste Generation Prediction in India.

Yashas Chaudhry

Step By Step School, Noida
Email: yashaschaudhry12@gmail.com.

Abstract
Electrical and electronic waste (or e-waste) constitute one of the most rapidly growing waste streams in India but, unfortunately, the accuracy of forecasting of state-wise generation of e-waste based on statistical analysis is seldom verified using real observations made after forecasting. In this research, two regression-based forecasting techniques, namely simple linear regression and second degree polynomial regression, are compared as forecasting methods for state-wise prediction of e-waste generation in India in the year 2024 based on historical data available with the Central Pollution Control Board (CPCB) from the years 2016 through 2023. The predictions have been validated using actual observations made in 2024 rather than goodness of fit, which makes this paper unique in comparison with many forecasting studies available. Polynomial regression resulted in slightly better mean absolute error (14,998 metric tonnes as opposed to 16,993 metric tonnes in linear regression case), and Pearson correlation of r = 0.893 as compared with r = 0.914 in the case of the linear regression method, although the linear regression technique produced slightly lower root-mean-square error and higher coefficient of determination, which was attributed to a single outlier resulting from polynomial regression for Uttarakhand. Accuracy at the state-level was quite different from the state to the state: three states - Tamil Nadu, Uttarakhand, and Maharashtra - were validated within five per cent error with the actual 2024 value, whereas the three other states, namely Andhra Pradesh, Himachal Pradesh, and Chhattisgarh, had forecasting error rates above 150%. These discrepancies seem to correlate more with the internal consistence of data series provided by particular states rather than with the used forecasting technique. It may be summarized that forecasting accuracy in Indian e-waste industry does not depend on mathematical complexity of used technique but more on the consistency of historical data provided. Thus, the most important outcome of this research is the following recommendation for State Pollution Control Boards to improve data-reporting processes rather than implement sophisticated predictive techniques
Keywords: *e-waste forecasting; regression analysis; India; validation study; data quality; state-level waste management; Central Pollution Control Board*.

INTRODUCTION

E-Waste or electronic waste includes electronic devices like computers, mobile phones, televisions, and household gadgets that have become waste products and have to be disposed of. E-waste has become one of the most rapidly growing waste streams internationally. The amount of e-waste generated globally has risen from 34 billion kg in 2010 to 62 billion kg in 2022 and is expected to grow to 82 billion kg by 2030 [1]. India currently ranks as the third largest producer of e-waste after China and the United States of America. The domestic production of e-waste in India is growing at a rate faster than the recyclers can cope with [1], [2].

As far as the response of the Indian government towards the above mentioned trend is concerned, there have been several rule changes with respect to the management of e-waste. The E-Waste (Management) Rules, 2016 for the very first time made it mandatory for the producers to make reports on an official basis, while these rules were significantly modified through the E-Waste (Management) Rules, 2022, that imposed the strict regime of Extended Producer Responsibility on producers, recyclers, dismantlers, and refurbishers making it compulsory for them to get registered along with their reporting of collection and processing volumes through an online platform maintained by the Central Pollution Control Board (CPCB) [3].

Notwithstanding the policy significance of such numbers, relatively little effort has been made thus far to project state-level generation of e-waste in India and to verify such projections using data unavailable at the time of model building. Previous Indian studies have been focused either on modeling aggregate trends, like the application of ARIMA models for modeling of processing capacity in Maharashtra [4], or on calculating waste generation rates by means of regression analysis with socio-economic parameters as predictors at the municipal or household level [5]. Internationally, forecasting of solid and electronic waste flows with the help of regression analysis has been

documented for Nepal [10], Mexico [9], Romania [8], and China. In such cases, model fitting and verification has generally been conducted in the same historical period, i.e. in-sample, and not out-of-sample, when the model has been tested using data generated after model development. The difference is critical because a good model fitting does not mean good forecasting and the opposite is even more true.

The current study aims to fill this void. Two independent regression models – ordinary least-squares linear regression and second-degree polynomial regression – are applied, on the basis of CPCB data compiled at the state level, from 2016 through 2023, to the following 13 Indian states exhibiting varying e-waste generation levels and industrial structures: Andhra Pradesh, Chhattisgarh, Gujarat, Haryana, Himachal Pradesh, Karnataka, Kerala, Maharashtra, Punjab, Rajasthan, Tamil Nadu, Telangana, and Uttarakhand. The predictions of both models for 2024 are then compared to the real value of that year, revealed in the subsequent CPCB report.

This research has been structured with respect to five questions. Can regression-based models predict the generation of e-waste at state level accurately? If yes, then which form, linear or polynomial, yields better performance. What significance does the quality and continuity of the time series data have on forecast accuracy? Which states are suitable for prediction based on their e-waste generation pattern? Lastly, how this type of prediction can help in e-waste management in India? The rest of the paper has been structured as follows. Section 2 provides an overview of literature related to e-waste and municipal solid waste forecasting. Section 3 identifies the research gap addressed by this study. Section 4 details the methods used for data processing and modelling. Section 5 provides the results of validation. Sections 6 and 7 provide discussion and conclusions respectively.

2. Literature Review

2.1 Material Flow Analysis Approaches to E-Waste Estimation

Material Flow Analysis (MFA) has historically been the dominant methodology for estimating e-waste generation, particularly at the national scale. MFA treats the electrical and electronic equipment sector as a stock-and-flow system in which sales data are combined with assumed product lifespan distributions to estimate the volume of equipment reaching end-of-life in any given year. A systematic review of forty-one MFA-based e-waste studies found that the overwhelming majority were conducted at the national rather than sub-national level, reflecting the difficulty of obtaining disaggregated sales and lifespan data for smaller administrative units [6]. A subsequent review extended this synthesis, noting that MFA's reliance on assumed lifespan distributions makes it sensitive to the quality of input assumptions rather than to any inherent weakness in the modelling technique itself, a sensitivity that mirrors the data-quality dependency identified in the present study for regression-based approaches [7].

2.2 Grey and Time-Series Forecasting Models

Where historical time series are short or the underlying process is not well characterised by a fixed functional form, grey forecasting models, particularly the GM(1,1) formulation, and Auto-Regressive Integrated Moving Average models have been widely applied to waste forecasting problems, since both can operate on limited data while capturing short-term trend dynamics. In the Indian context, an ARIMA model fitted to Maharashtra's e-waste processing capacity between 2016 and 2023 projected an average processing load of approximately 163,563 metric tonnes per year through 2030, illustrating the use of time-series methods for single-state, single-model forecasting exercises [4]. Such models are attractive where autocorrelation in the series is strong, but they are typically applied and reported without subsequent validation against real observations collected after the forecast was made, leaving open the question of how such projections perform once the forecast horizon has actually elapsed.

2.3 Regression-Based Forecasting of Waste Generation

Regression analysis remains the most widely used statistical tool for waste generation forecasting internationally, owing to its interpretability and modest data requirements relative to machine-learning alternatives. An empirical regression model built from household-level data in Dhanbad, India, linked solid waste generation rates to socio-economic and demographic predictors and remains one of the more frequently cited India-specific regression studies in this domain [5]. Outside India, comparable regression-based forecasting has been reported for Kathmandu, Nepal, where a linear regression model relating population to municipal solid waste generation produced a correlation coefficient of 0.956 [10]; for a Mexican municipality, where a multiple linear regression model achieved a comparable coefficient of determination [9]; and for Romania, where prognostic regression tools were benchmarked directly against one another for municipal solid waste forecasting, one of the few studies

in this literature to systematically compare multiple regression formulations rather than reporting a single fitted model [8]. None of these studies, however, tested their regression forecasts against waste-generation figures observed in a year that postdated the training period, which is the specific methodological step this study contributes.

2.4 National Versus State-Level Forecasting in the Indian Context

A further limitation specific to the Indian data landscape is that officially reported e-waste generation, as distinct from collection and processing, is estimated nationally from producer sales data and average equipment lifespans rather than measured directly at the state level; the Ministry of Housing and Urban Affairs has confirmed in a parliamentary response that state-wise generation estimates in this form are not compiled [11]. The figures used in the present study, and in comparable Indian analyses, are therefore CPCB-compiled collection and processing records rather than independently estimated generation totals, a distinction that has direct bearing on data continuity, since collection and processing volumes depend on the operational status of registered recyclers and dismantlers in a given state and year, and are consequently more susceptible to reporting gaps than a demand-side sales estimate would be.

3. Research Gap

As can be seen from the existing literature, there is an array of methodologies used here: MFA, grey models, ARIMA, and regression; however, a very high dependence on the use of fit statistics in samples to prove model adequacy is evident. There are very few instances of out-of-sample validation using real-world data in this body of literature, even less so at the sub-national level in India. The current research attempts to bridge this gap by employing linear and polynomial regressions trained solely on 2016-2023 data and tested using 2024 data reported by CPCB.

4. Materials and Methods

4.1 Data Source and Scope

State-level data of collection and processing of e-waste presented in CPCB annual reports between 2016 and 2024 forms part of the data that is used for the present study. Such a data set consists of all the states and UTs in India in addition to an overall Indian data set. The dataset that has been used has 13 states which are Andhra Pradesh, Chhattisgarh, Gujarat, Haryana, Himachal Pradesh, Karnataka, Kerala, Maharashtra, Punjab, Rajasthan, Tamil Nadu, Telangana, and Uttarakhand. The generation rate of e-waste for each state varies from below 200 metric tonnes per year to above 100,000 metric tonnes per year.

For all states, data up to 2023 were employed to develop forecasting models, whereas 2024 served as the year held out for validation of the models. The division was fixed prior to model fitting; hence, the comparison presented in Section 5 involving 2024 serves as a true out-of-sample test.

4.2 Data Preprocessing and the Role of Imputation

These underlying data for the CPCB returns include many instances where an empty cell exists since the recyclers registered in that state did not report any separate annual figures for that year, or since the earlier reporting years prior to 2022 rules were less detailed on state-wise basis. In such cases, the working database includes the latest available value or uses linear interpolation between the years where values have been reported, and maintains a boolean variable denoting whether a value has been reported or obtained using this procedure. This paper does not drop or re-estimate these flagged values since doing so will hide exactly the data quality issue which this paper seeks to address. Consequently, the share of affected years within the training period varies across states, ranging from zero for Tamil Nadu, Andhra Pradesh, and Telangana to two for Chhattisgarh, Haryana, Himachal Pradesh, and Rajasthan, and this information is utilized in the Discussion section of the paper.

4.3 Forecasting Models

For each state, two different regression models have been separately formulated using 2016 to 2023 data, with year of reporting as the independent variable and tons of electronic waste produced annually as the dependent variable.

In linear regression, the association of variables is expressed by a straight line:

$\hat{y} = \beta_0 + \beta_1 \cdot t$

where $\hat{y}$ is the estimate of e-waste generation, t is the year of report, and β_0 and β_1 are the coefficients estimated using ordinary least squares method.

The second regression specification extends the first one with the addition of a squared term, accounting for non-linear growth when it is the case:

$\hat{y} = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2$

Both regressions were conducted using the scikit-learn library for Python 1.5. Polynomial features were constructed using scikit-learn's Polynomial Features. Neither regularization nor weighting or excluding of any outliers was performed apart from the pre-processing stage described in Section 4.2 in order to show the results as is.

4.4 Validation Framework

A single-point forecast for 2024 was provided by each of the fitted models and compared against the respective actual value. Comparison was done on two different levels. On the aggregate level, forecasts for all thirteen states were compared against the respective actual values to determine the error and correlation for the two fitted models. On the individual level, the better of the two models in terms of having the lower absolute percentage error was chosen as the representative forecast for that particular state, categorized into one of four categories: excellent (absolute percentage error less than 5%), good (5%-20%), moderate (20%-50%), or poor (greater than 50%).

4.5 Performance Metrics

The following five statistics have been used to measure forecasting accuracy, calculated separately for each of the two models using the set of thirteen states:

$MAE = (1/n)\ \Sigma|y_i - \hat{y}_i|$

The MAE criterion is not mentioned directly in the literature reviewed, but its equivalent, an average error criterion, forms the foundation of the Romanian benchmarking study by Ghinea et al. [8]. The advantage of MAE being that it measures the error in the same unit of tonnes as the data without giving unnecessary emphasis to outliers.

$RMSE = \sqrt{[(1/n)\ \Sigma(y_i - \hat{y}_i)^2]}$

None of the papers studied give the value for RMSE because all the models were used in-sample fitting instead of out-of-sample validation. This measure is useful due to the fact that its value includes squared values of residuals, so larger misses will be punished more than in the case with MAE; RMSE and MAE were used in this research.

$\text{Percentage Error} = (|y_i - \hat{y}_i| / y_i) \times 100$

The measure of percentage error does not appear in the literature referenced in the paper, as they prefer either aggregate correlation or fit statistics over case-by-case errors. The benefit of the measure is that it expresses the error in relation to the actual value for each state, thus allowing states of vastly different sizes to be compared on the same scale.

$R^2 = 1 - [\Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y})^2]$

As stated above, R^2 is used for the goodness-of-fit of the model in sample. This coefficient was used in this unusual way in the study from Mexico [9]. It was calculated on the 2024 out-of-sample forecasts rather than on the in-sample data, in order to estimate how well the model captured the real variance between states.

$r = \Sigma(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) / \sqrt{[\Sigma(y_i - \bar{y})^2\ \Sigma(\hat{y}_i - \bar{\hat{y}})^2]}$

This particular piece of data is available from the research paper from Nepal [10], which utilized a linear regression analysis for which the value of r is 0.956. Pearson's r gives an idea of the linear relationship between the actual value and predicted value independent of any scale or bias.

Where, y_i refers to the actual value in the year 2024 of state i, $\hat{y}_i$ refers to the corresponding forecasted value, $\bar{y}$ and $\bar{\hat{y}}$ refer to the means of the actual values and forecasted values, respectively, while n refers to the number of

states (n = 13). Mean absolute error and root-mean-square error are measured in metric tons; percentage error is measured per state; the coefficient of determination and the Pearson correlation coefficient are unitless numbers.

5. Results

5.1 Aggregate Model Performance

Table 1 shows the combined performance of both models in predicting values for all the thirteen states through a comparison of the total of thirteen predictions in 2024 made by each model against the corresponding actual value. The polynomial regression model had a lower mean absolute error as compared to the linear regression model since it had 14,998 metric tonnes while the other model had 16,993 metric tonnes. This shows that, on average, the polynomial model performed better in predicting values because the predictions it made were closer to the actual 2024 values. Linear regression, on the other hand, had a lower root-mean-square error and a higher coefficient of determination — 21,644 metric tonnes and 0.816, respectively, as opposed to 23,325 metric tonnes and 0.787 for the polynomial model.

Table 1. Aggregate Forecasting Performance: Linear Regression VS Polynomial Regression

Metric	**Linear Regression**	**Polynomial Regression (Degree 2)**
Mean Absolute Error (MT)	16,992.67	14,998.00
Root Mean Square Error (MT)	21,644.12	23,324.64
Coefficient of Determination (R^2)	0.816	0.787
Pearson Correlation Coefficient (r)	0.914	0.893

The reason why the polynomial regression wins in the mean absolute error whereas linear regression is victorious in the root mean square error and R^2 is explained by the higher sensitivity of RMSE and R^2 to larger individual errors in comparison with MAE due to squaring of individual values when calculating these two indicators. The reason is demonstrated by the overestimation of the prediction for Uttarakhand by the polynomial regression by approximately 50,000 metric tonnes in comparison with the real value of 2024 (Section 5.2).

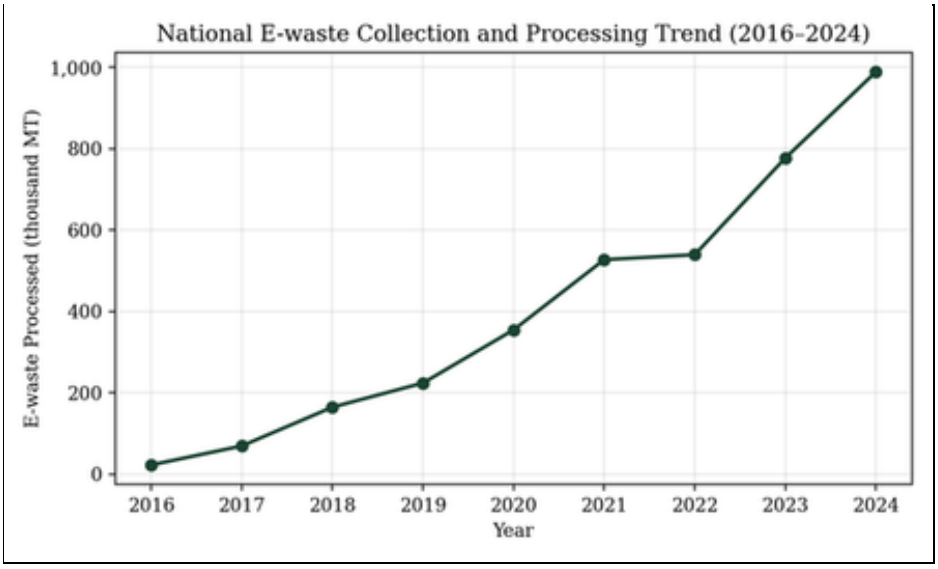

Figure 1. National E-waste Collection and Processing Trend, 2016–2024 (All-India Aggregate, CPCB Data)

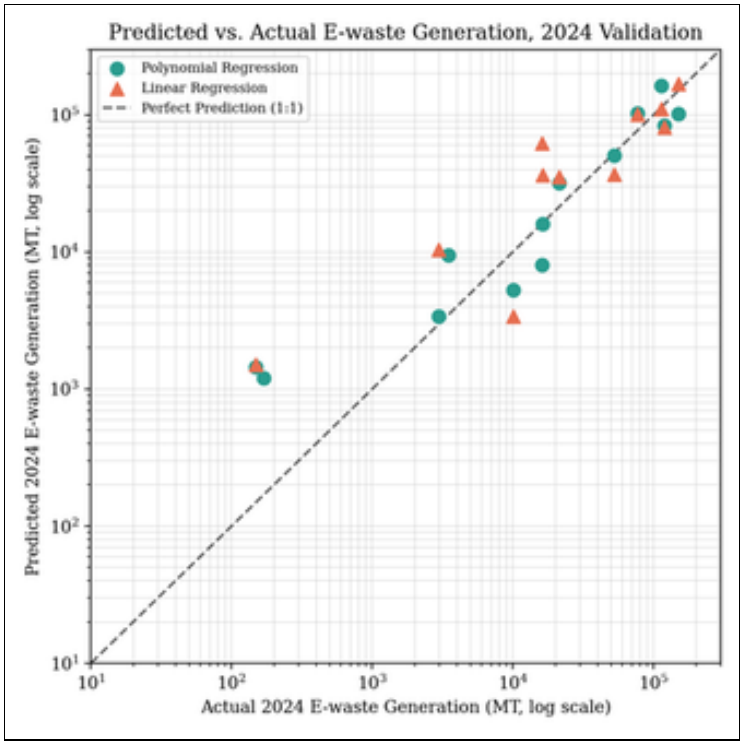

Figure 2. Predicted versus Actual 2024 E-Waste Generation for Both Models (log–log scale, dashed line indicates perfect prediction)

5.2 State-Level Validation Results

Table 2 presents the actual value of 2024, predicted values from the two models, percentage errors, and better performing model along with its validation status for each of the thirteen states. There exists a lot of variability in the results. There are three states where the error was below 5% using their better performing model and hence fall in the excellent category. There are two states which have errors between 12% to 15%, thereby falling in the good category. There are four states whose error lies between 30% to 49% and hence falls in the moderate category. Finally, there are four states whose error exceeds 50% and the worst case being that of Andhra Pradesh where the error from the best model is 874.95%.

Table 2. State-Level Validation Results, 2024 (Sorted by Validation Category)

State	Actual 2024 (MT)	Linear Pred. (MT)	Poly. Pred. (MT)	Linear Err. (%)	Poly. Err. (%)	Better Model	Category
Tamil Nadu	16,360.82	36,381.48	15,972.88	122.37	2.37	Polynomial	Excellent
Uttarakhand	113,562.26	110,811.55	163,572.76	2.42	44.04	Linear	Excellent
Maharashtra	52,597.80	36,714.54	50,800.40	30.20	3.42	Polynomial	Excellent
Haryana	149,653.64	167,989.29	101,184.14	12.25	32.39	Linear	Good
Punjab	2,972.09	10,421.02	3,406.23	250.63	14.61	Polynomial	Good
Telangana	119,187.98	80,950.83	83,275.24	32.08	30.13	Polynomial	Moderate
Gujarat	76,879.92	101,070.59	103,207.90	31.47	34.25	Linear	Moderate
Kerala	10,079.63	3,386.57	5,258.54	66.40	47.83	Polynomial	Moderate
Rajasthan	21,345.48	35,439.92	31,762.73	66.03	48.80	Polynomial	Moderate

Karnataka	16,151.77	62,048.06	8,051.68	284.16	50.15	Polynomial	Poor
Chhattisgarh	3,500.92	−21,761.07	9,465.64	721.58	170.38	Polynomial	Poor
Himachal Pradesh	169.27	−573.27	1,196.47	438.67	606.84	Linear	Poor
Andhra Pradesh	148.97	1,498.26	1,452.38	905.75	874.95	Polynomial	Poor

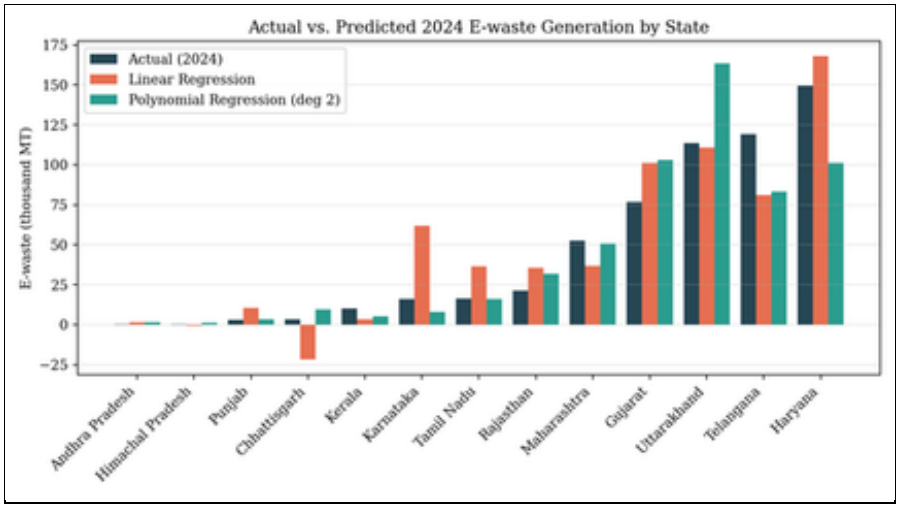

Figure 3. Real vs Predicted E-waste generation in 2024 by State (Both Models, sorted by real values)

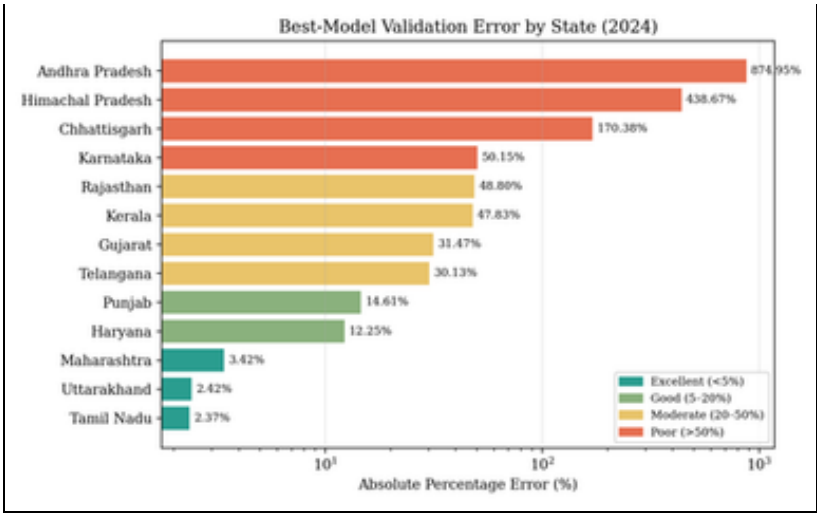

Figure 4. Best Model Validation Errors per state, 2024 (Coloured by Validation Category)

5.3 Distribution of Better-Performing Models

Polynomial regression fitted nine states (Tamil Nadu, Maharashtra, Punjab, Telangana, Kerala, Rajasthan, Karnataka, Chhattisgarh, and Andhra Pradesh) out of 13, while linear regression fitted the remaining states (Uttarakhand, Haryana, Gujarat, and Himachal Pradesh). It is important to note here that polynomial regression performed well in three out of four poor category states, which indicates that the extra curvature parameter in polynomial regression helped in reducing the error due to erratic underlying data series; this will be explained in Section 6.

6. Discussion

6.1 Why Polynomial Regression Outperforms on Some Metrics but Not Others

The reason why the polynomial regression model is preferred under MAE but the linear model is preferred under RMSE and R^2 lies in how sensitive these two metrics are to a single large error. RMSE squares the residuals before averaging them, and R^2 is calculated from the same squared residuals, so both are much more affected by one bad prediction than MAE is. Here, the one bad prediction is Uttarakhand. Its e-waste generation rose from 2,165 metric tonnes in 2018 to 134,255 metric tonnes in 2023, and using this trend, the polynomial model overestimated the 2024 value by nearly 50,000 metric tonnes, while the linear model was off by only 2.42%. If Uttarakhand were removed from the sample, the RMSE gap between the two models would shrink considerably. MAE does not get pulled around by this kind of single large error in the same way, and by that measure the polynomial model still comes out ahead on the rest of the states. So no single metric on its own gives the full picture of which model performs better.

6.2 Why Some States Validated with High Accuracy

The three states in the excellent category – Tamil Nadu, Uttarakhand, and Maharashtra – have in common that their training-period series, even though not completely immune to year-to-year variations, have followed a consistent pattern during the 2016–2023 period without any sudden change. The training period series of Uttarakhand shows an almost monotonic increase in seven of its eight training periods, a pattern which is quite straightforward for linear regression to predict. The training period series of Maharashtra, on the other hand, increases and speeds up after increasing mildly until 2018–2020, a curvature pattern that can be captured better by polynomial regression than linear regression, which assumes an average trend per training period, leading to a relatively high error rate of 30.20% in the former as opposed to 3.42% in the latter. Similarly, the training period series of Tamil Nadu shows an upward trend to a peak between 2019–2022 and then drops off slightly, which too is a pattern that is better suited to polynomial regression than linear regression. Moreover, these states have a zero or one imputed training year each.

6.3 Why Some States Produced Extremely High Errors

These four poor-category states exhibit at least two separate causes of high error, and these should be distinguished from one another instead of simply categorizing high error as a homogenous occurrence. The training data for the

state of Chhattisgarh displays the imputation artifact in its most concrete sense; while the values for 2016, 2017, and 2018 are all recorded at 62,387.85 metric tonnes (of which the first two are marked as imputed values), there is an abrupt drop down to 26.99 metric tonnes in 2019, marking a difference of more than three orders of magnitude, followed by values that swing from about 130 to 7,371 metric tonnes for the rest of the training period. In such a case, a regression model, whether linear or polynomial, will essentially just be averaging between the fabricated flat segment and the actual new regime; this is apparent in the extremely high errors obtained using either model on this particular state (721.58% in the case of linear and 170.38% in the case of polynomial).

The states of Himachal Pradesh and Andhra Pradesh provide another illustration of a different cause of extrapolation error: even though there are only a few imputed years within the dataset, the figures are highly volatile. For instance, Himachal Pradesh's values jump from 140.9 metric tonnes in 2019 to 72.94 in 2020, 373.2 in 2021, 2,106.12 in 2022, and 159.68 in 2023. In a similar fashion, the values for Andhra Pradesh go from being in the single or double digits between 2016 and 2019 to 2,021.19 metric tonnes in 2021 and then fall back to 148.97 metric tonnes in 2024. As a consequence of these trends, both linear and polynomial extrapolations cannot be made, resulting in respective errors of 438.67% and 874.95% – again, the best-performing model has the worst error estimate possible. Therefore, these errors should be attributed to the volatility of the generation and reporting process itself instead of the poor quality of data and possibly due to fluctuations in activity or infrequent reporting by a few major recyclers. The state of Karnataka provides another example, going from approximately 7,000 metric tonnes in 2016 to 96,192 metric tonnes in 2020 and then dropping back down to 16,000-21,000 metric tonnes for the remainder of the training period, a swing that produced a best-model error of 50.15% and placed it in the poor category alongside the other three states discussed above.

6.4 Why Data Continuity Matters More Than Model Choice

When considered together, Section 6.2 and 6.3 demonstrate how the hypothesis of this research stands true. The states that demonstrated consistent and steady reporting throughout the period had high levels of validation regardless of the type of regression used in the analysis. On the other hand, states whose reporting histories had either imputation discontinuities or genuine fluctuations had low validation levels regardless of whether a linear or polynomial regression was used in the analysis; however, switching between the types of regressions increased the magnitude of errors among the poor performers, but never brought their validation levels closer to the level of the better performers. This can be supported by observing the correlation of volatility measures and forecast error in the analysis of all thirteen states: states with low year-on-year volatility had low forecast errors whereas states with high year-on-year volatility had high forecast errors.

The most obvious example is, however, still the comparison of Tamil Nadu, which has a training coefficient of variation equal to 0.34 and a best-model error of 2.37%, and Andhra Pradesh, which has a coefficient of variation equal to 1.25 and a best-model error of 874.95%. The implication for the CPCB and State Pollution Control Board practice is that the single most effective action that one can take to increase forecasting accuracy is not the implementation of a more complex model, but rather the mitigation of reporting discontinuities such as those seen in Chhattisgarh, as well as the analysis of structural volatility in Andhra Pradesh and Himachal Pradesh, which could very well be the consequence of inconsistent registration/compliance with the E-Waste (Management) Rules of 2022 [3].

6.5 Comparison with Previous Studies

Such an awareness of data quality requirements is also supported by – albeit seldom stated as explicitly – other studies in the waste forecasting literature. For instance, Ghinea et al., when assessing the relative performances of several regression-based models used to forecast Romanian municipal solid waste, showed that models' performances significantly varied based on the stability of the time series and not based on finding the 'best performing' model [8]. The similar findings were observed above in the variation in the performance of certain models in the state-by-state analysis. While Kumar and Samadder managed to achieve high fitting of their regression model for Dhanbad at the household level through conducting a survey of primary data in well-controlled conditions, the present study was conducted using multi-source data from different agencies where no independent verification of data completeness can be done [5]. Islam and Huda's literature reviews of Material

Flow Analysis, in turn, show that MFA outputs are only as reliable as the assumptions about the sales and life span used for the modeling [6], [7].

6.6 Contribution to Future E-Waste Forecasting

Out-of-sample validation, as employed in this paper, with training done using 2016-2023 data only and the test done using real values for 2024, as far as the authors know, is hardly ever conducted in the Indian e-waste literature, where model adequacy is typically indicated by means of in-sample fit criteria only. In other words, in this case, one cannot assume that the model fit statistics for 2016-2023 years are an adequate criterion for future prediction. In case the value of 2024 depends on certain circumstances, not captured by a linear trend, but not taken into account, such as, for instance, changes in recycler registration, the prediction made for 2024 might be wrong. Thus, in the future, such predictions should be validated with respect to their out-of-sample accuracy rather than in-sample fit.

7. Conclusion

This research was an attempt to find out whether regression-based forecasts for e-waste generation at state level in India could be accurate enough to be useful, which type of regression works better, how dependent is the accuracy on data quality, which states are good for accurate forecasting and what can be done for policy making by using this method. However, the results are quite conditional. Indeed, regression-based forecasts can be accurate enough, as in three out of thirteen states the difference between forecasted and real value of e-waste generation in 2024 does not exceed 5%, but regression forecasts can also turn out to be completely wrong, with three states having error rates exceeding 150% and one state having an error rate exceeding 800%. Polynomial regression provides slightly better results in terms of mean absolute error and Pearson correlation, whereas linear regression shows better performance regarding root-mean-square error and coefficient of determination.

The main value of the present research lies in providing evidence that forecasting in this field depends not on the complexity of the model but rather on the coherence and consistency of reporting done by a certain state. States with training data indicating one trend (either constant growth or accelerating growth) could be predicted with equal success by both models, while states with training data showing either an artifact created due to imputation or year-on-year variance in reported figures could be predicted unsuccessfully by both models, and changing from linear to polynomial model was not of much help there. The conclusion to draw from this is that model complexity does not play the role of a limiting factor for the accuracy of predictions made in the current Indian e-waste reporting practice.

This would mean that in terms of policy planning, the better place to channel investments meant to enhance the predictive accuracy of e-waste management in states will be through efforts geared towards enhancing the consistency and integrity of reports produced by the CPCB and State Pollution Control Boards, especially on the performance of recyclers in terms of registrations and returns filing in line with EPR, rather than investing in increasingly complex methods of prediction in the face of imperfect data.

Only after the reporting record of a state becomes somewhat consistent can the marginal decision about the prediction method begin to matter.

The present analysis has been restricted in its scope to thirteen states in one country, by its use of two regression specifications, and by the validation process carried out on one held-out year rather than through a longer rolling forecast period. Further research could profitably explore the application of the present validation process to autoregressive and gray forecasting models, such as were employed in previous analysis of Maharashtra's processing capacity [4], to determine if there is any real benefit of accuracy provided by the additional modeling process, and use the same coefficient of variation and year-to-year variability analysis for identifying states that can be reliably predicted before choosing a specific method. An extension of the data set over additional years as the CPCB provides further return data will permit testing of the reliability of the current conclusions over more than one validation year.

..

References

1. C. P. Baldé, R. Kuehr, T. Yamamoto, R. McDonald, E. D'Angelo, S. Althaf, G. Bel, O. Deubzer, E. Fernandez-Cubillo, V. Forti, V. Gray, S. Herat, S. Honda, G. Iattoni, D. S. Khetriwal, V. Luda di Cortemiglia,

Y. Lobuntsova, I. Nnorom, N. Pralat, and M. Wagner, Global E-waste Monitor 2024. Geneva/Bonn: International Telecommunication Union (ITU) and United Nations Institute for Training and Research (UNITAR), 2024.

2. V. Forti, C. P. Baldé, R. Kuehr, and G. Bel, The Global E-waste Monitor 2020: Quantities, Flows, and the Circular Economy Potential. Bonn/Geneva/Rotterdam: United Nations University/United Nations Institute for Training and Research, International Telecommunication Union, and International Solid Waste Association, 2020.
3. Ministry of Environment, Forest and Climate Change, Government of India, "E-Waste (Management) Rules, 2022," Notification No. G.S.R. 801(E), 2 Nov. 2022.
4. W. A. Bagwan, "Electronic waste (e-waste) generation and management scenario of India, and ARIMA forecasting of e-waste processing capacity of Maharashtra state till 2030," Waste Management Bulletin, vol. 1, no. 4, pp. 41–51, 2024, doi: 10.1016/j.wmb.2023.08.002.
5. A. Kumar and S. R. Samadder, "An empirical model for prediction of household solid waste generation rate – A case study of Dhanbad, India," Waste Management, vol. 68, pp. 3–15, 2017, doi: 10.1016/j.wasman.2017.07.034.
6. M. T. Islam and N. Huda, "Application of Material Flow Analysis (MFA) in Electronic Waste (E-Waste) Management: A Review," Proceedings, vol. 2, no. 23, p. 1457, 2018, doi: 10.3390/proceedings2231457.
7. M. T. Islam and N. Huda, "Material flow analysis (MFA) as a strategic tool in E-waste management: Applications, trends and future directions," Journal of Environmental Management, vol. 244, pp. 344–361, 2019.
8. C. Ghinea, E. N. Drăgoi, E. D. Comăniță, M. Gavrilescu, T. Câmpean, S. Curteanu, and M. Gavrilescu, "Forecasting municipal solid waste generation using prognostic tools and regression analysis," Journal of Environmental Management, vol. 182, pp. 80–93, 2016, doi: 10.1016/j.jenvman.2016.07.026.
9. J. A. Araiza-Aguilar, M. N. Rojas-Valencia, and R. A. Aguilar-Vera, "Forecast generation model of municipal solid waste using multiple linear regression," Global Journal of Environmental Science and Management, vol. 6, no. 1, pp. 1–14, 2020, doi: 10.22034/gjesm.2020.01.01.
10. A. Khanal, "Forecasting municipal solid waste generation using linear regression analysis: A case of Kathmandu Metropolitan City, Nepal," Multidisciplinary Science Journal, 2023.
11. T. Sahu, Minister of State, Ministry of Housing and Urban Affairs, Government of India, reply to Rajya Sabha Unstarred Question No. 2384, 16 Dec. 2024. [Online]. Available: https://sansad.in/getFile/annex/266/AU2384_9gmDUV.pdf