Next Sustainability 8 (2026) 100375

Contents lists available at ScienceDirect

Next Sustainability

journal homepage: www.sciencedirect.com/journal/next-sustainability

Research article

Early action in transportation, electricity grid, commercial and residential buildings is essential for deep and near net-zero decarbonization

Peiyao Zhao [a],*,[1], Jimi Oke [b]

[a] Department of Chemical and Biomolecular Engineering, University of Massachusetts Amherst, MA, USA
[b] Department of Civil and Environmental Engineering, University of Massachusetts Amherst, MA, USA

ARTICLE INFO

Dataset link: https://github.com/narslab/manuscript-newhaven-emissions-forecast

Keywords:
Decarbonization
Scenario discovery
XGBoost
Emissions
Planning

ABSTRACT

In this work, we applied scenario discovery to explore emission reduction potentials and identify decarbonization pathways over multiple sectors, using the New Haven-Milford Metropolitan statistical area as a case study in regional planning. We first generated a diverse set of future emission states using Latin hypercube sampling (LHS), and then applied Gaussian mixture modeling to cluster outcomes based on emission percentage changes between 2021 and 2050 (EPC) and cumulative emissions percentage change (CEPC). This analysis identified four decarbonization scenarios: *Mild*, *Moderate*, *Deep*, and *Near Net-Zero*. We further examine the pathways associated with these decarbonization scenarios and the parameter combinations that could drive progress toward the decarbonization targets, as well as shifts in the relative importance of these activity parameters under different target-year assumptions. The results indicate that achieving *Near Net-Zero* requires parameter shifts exceeding 50% of their feasible sampled ranges for electrification rates, fuel economy (MPG), truck vehicle miles traveled (VMT), and commercial emissions, as well as earlier action in achieving this transition. Furthermore, we trained an eXtreme Gradient Boosting (XGBoost) model to predict EPC and CEPC. XGBoost model revealed the most relevant factors in each scenario. In particular, residential natural gas consumption is a key driver of emissions reduction in *Near-Net Zero*, followed by commercial emissions, highlighting an effort to improve the efficiency of residential and commercial building energy consumption. Overall, we demonstrate that not only the level of reduction of each activity parameter is important for emissions reduction, but also the pace of the parameter change over the prediction horizon.

1. Introduction

Cities are becoming increasingly important in shaping the future of carbon and climate in this century [1] and perhaps beyond. They contribute between 60 and 70% of global greenhouse gas (GHG) emissions, largely due to energy consumption, transportation and industrial activity [2–4], and this share continues to rise annually [5,6]. Crucially, urban centers are also leading efforts to develop and advocate for climate solutions [3]. Although global efforts set an ambitious climate target, the practical implementation of emissions mitigation often depends on decisions made at the national and regional level [7,8]. For example, the emissions profile can differ significantly between metropolitan regions, and thus a one-size-fits-all approach cannot be applied to every region, highlighting the need for localized strategies [9]. Therefore, it is important to discover climate mitigation scenarios or the decarbonization pathway to meet specific goals of reducing GHG emissions at the city and regional level, as these pathways can offer insight into exploring how certain reduction goals can be achieved through policy measures and actions for a variety of stakeholders such as policy makers, educational institutions, business companies, and society [10–13].

Historically, scenario planning has been used to facilitate decision-making under uncertainty. Classic scenario planning relies on exploring a small number of divergent or extreme futures. The usefulness of the insights gained from these efforts depends on the choice of scenario narrative and the quality of the respective assumptions made in each. These approaches, while still widespread, are inadequate for addressing multidimensional uncertainty in complex systems [14]. Although optimal strategies or solutions may be found in the enumerated deterministic scenarios and then compared in an effort to aid decision making, critical pathways or vulnerabilities are prone to be missed. Thus, traditional scenario planning is not well suited for robust decision-making (RDM) applications [15].

In the last two decades, scenario discovery (SD) [14,16,17] has emerged as a viable framework for analyzing multidimensional future

* Corresponding author.
E-mail address: peiyaozhao@umass.edu (P. Zhao).
[1] Lead contact.

https://doi.org/10.1016/j.nxsust.2026.100375
Received 24 October 2025; Received in revised form 16 March 2026; Accepted 13 June 2026
Available online 4 July 2026

outcomes across a set of given strategies for robust decision-making under deep uncertainty (see Fig. 1). Rather than the top-down approach of traditional scenario planning, SD relies on quantifying the uncertainty of key model input parameters governing the outcomes of interest. From the estimated distributions, representative joint outcomes (or "states") are generated through computational experimental design approaches, such as Latin hypercube sampling (LHS) [18–22] and Monte Carlo simulations [23]. The high-dimensional input spaces could be achieved to cover the full range of possible futures for exploration. These states are then evaluated by the model under the proposed strategies, which represent potential policy decisions. The resulting set of futures can then be classified by thresholding the performance of the future outcomes relative to the regret minimization strategy in each corresponding state (as exemplified by [24]) or clustered to more flexibly represent a variety of performance outcomes [25]. Data mining algorithms can then be used to uncover combinations of uncertain parameter inputs leading to high-interest outcomes. Such an approach allows the data to speak for itself and facilitates the identification of robust strategies (i.e., those with minimum failure under performance thresholds of interest). Vulnerabilities across strategies can also be analyzed to obtain robust policy pathways.

In environmental research, SD enables policymakers and entrepreneurs to identify the conditions under which strategies are most likely to succeed, providing robust pathways to achieve decarbonization targets. This is different from approaches that analyze emissions reduction pathways under existing scenarios, which focus on understanding outcomes given predetermined assumptions rather than uncovering the conditions that achieve a specific target [12,26–28]. Despite the strength of this approach and numerous developments and extensions throughout the years [14,16,24,29–34], SD has still seen limited use in exploring decarbonization pathways and emissions reduction scenarios with a few exceptions. For example, using a system dynamics model combined with scenario discovery via Monte Carlo simulations and sensitivity analysis, this work simulated multiple emissions mitigation pathways by shifting people's behavior from flights to trains [23]. While their approach is comprehensive and captures uncertainties by discovering the parameters that are most influential to mitigation strategies via sensitivity analysis, it still has limits in discovering which combination of assumptions of factor are most robust across many possible world. Another study on scenario discovery on emissions mitigation pathways looks at the energy system [35]. They generated a large ensemble of scenarios by varying eleven key uncertainties to capture a wide plausible futures and obtained the most strongly influence outcome via random forest. However, their work focuses on a single sector: transport or energy, while a multisectoral approach, including entire GHG inventories, could provide a more comprehensive understanding. Similarly, SD has been applied to discover the combinations of uncertain socioeconomic and technological factors that produce particular outcomes; however, it remains at an aggregated system level rather than analyzing the sector-specific contribution to the decarbonization pathways [25]. Moreover, even when studies use uncertainty sampling to generate multiple scenarios, they did not identify conditions that achieve specific targets to qualify as scenario discovery. For example, by running a wide range of simulations in combinations of public perception, technological change, and political responsiveness, five distinct emissions pathways were identified [13]. While their clustering highlights broad patterns and key drivers, it does not systematically examine the specific combinations of factors that produce each pathway.

In summary, there remains a dearth of studies on discovering robust emissions mitigation scenarios from a combination of activities that spanning multiple sectors including transportation, energy consumption, building, wastewater, and agriculture. In addition, no scenario discovery effort has considered the uncertainty in the pace of activity changes. In this study, we address this gap by applying SD in a forecasting context to discover robust decarbonization pathways and driving factors for emissions reduction. In particular, we also incorporate the target year for various mitigation actions as uncertain parameters. Furthermore, most studies that integrate forecast and scenarios generate a fixed number of scenarios under certain assumptions and then project emissions trajectories for those scenarios instead of focusing on uncertainties [36–38]. Building on our previous study [39], which focused on identifying influential parameters for robust regional mitigation strategies using scenario discovery as an exploratory tool, in this work, we reconstruct the emissions trajectory for each scenario by back-calculating from model outputs generated from scenario discovery and producing an inferred emissions pathway, which allows us to quantify temporal patterns and cumulative impacts of different scenarios. This current work sequentially combines emissions forecasting with scenario discovery to capture the trajectory of uncertainty, which has not been reported to our knowledge in previous scenario discovery studies.

2. Objectives and framework

The primary objective of this paper is to study the typology of the exploratory scenarios. Specifically, we use SD to generate and classify a large ensemble of plausible decarbonization trajectories under deep uncertainty in sectoral activity transitions. The goal is not to rank strategies by robustness, but rather to identify emergent pathway archetypes across sectors. Recent scenario studies use integrated modeling approaches to explore cross-sector interactions and long-term sustainability transitions [40,41]. Unlike this study, our study applies scenario discovery to identify typologies of emissions trajectories and the parameter conditions associated with each pathway.

Specifically, we used an Autoregressive Integrated Moving Average (ARIMA) model framework to forecast base case (business-as-usual) emissions from 2021 to 2050 (the "Base Case" scenario). We then applied Latin Hypercube Sampling (LHS) to generate 32-dimensional 2 million states representing unique combinations of possible changes in key activity changes by a given target. We then estimated the emissions for each case, generating a corresponding number of futures, and clustered these based on the percentage of emissions change from 2050 to 2021 (EPC) and the percentage of cumulative emissions change (CEPC). Finally, we applied an XGBoost model to predict emissions outcomes and to uncover priority actions relevant to each scenario. A diagram of the framework is shown in Fig. 1.

3. Results

3.1. Data-driven discovery yields four decarbonization scenarios

Clustering the futures yielded four decarbonization scenarios, namely: *Mild*, *Moderate*, *Deep* and *Near Net-Zero*. We plot the CEPC versus EPC in Fig. 2. Boxplots of each of these metrics by scenario are shown in Fig. 3. We observe distinct emissions profiles across each of the scenarios.

The *Mild* scenario presents the smallest change from the base case, while *Near-Net Zero* represents the most significant reduction, highlighting the impact that intense action has on the reduction of cumulative long-term emissions. The importance of different activity parameters vary in each scenario, as shown in Fig. 4. For example, *Near-Net Zero* cases are characterized by stronger reductions in industrial and commercial sector emissions, lower residential gas consumption, lower vehicle miles traveled (VMT), and cleaner electricity grids relative to the *Mild* and *Moderate*. Additionally, the timing of these transitions also differs among different scenarios. *Near-Net Zero* generally indicates earlier action across sectors, with significant shifts in residential natural gas consumption, commercial and industrial emissions, as well as truck and car VMT reduction occurring closer to 2030–2035. In contrast, *Mild* pathways have most changes until later decades, often closer to 2035–2040. The distribution of the timing further emphasizes that both pace and scale of the activity parameters change differentiate the four scenarios.

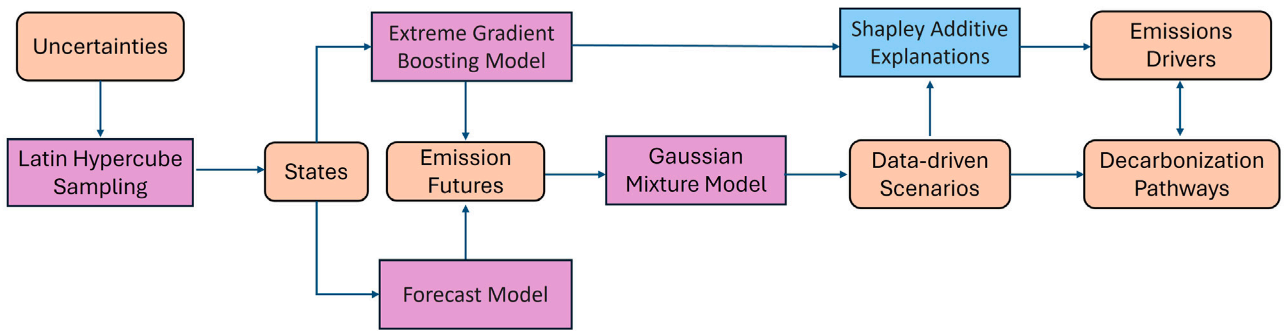


Fig. 1. Flowchart describing the research framework, including data sampling, state and future generation, emissions forecast modeling, outcome clustering, and scenario evaluation.

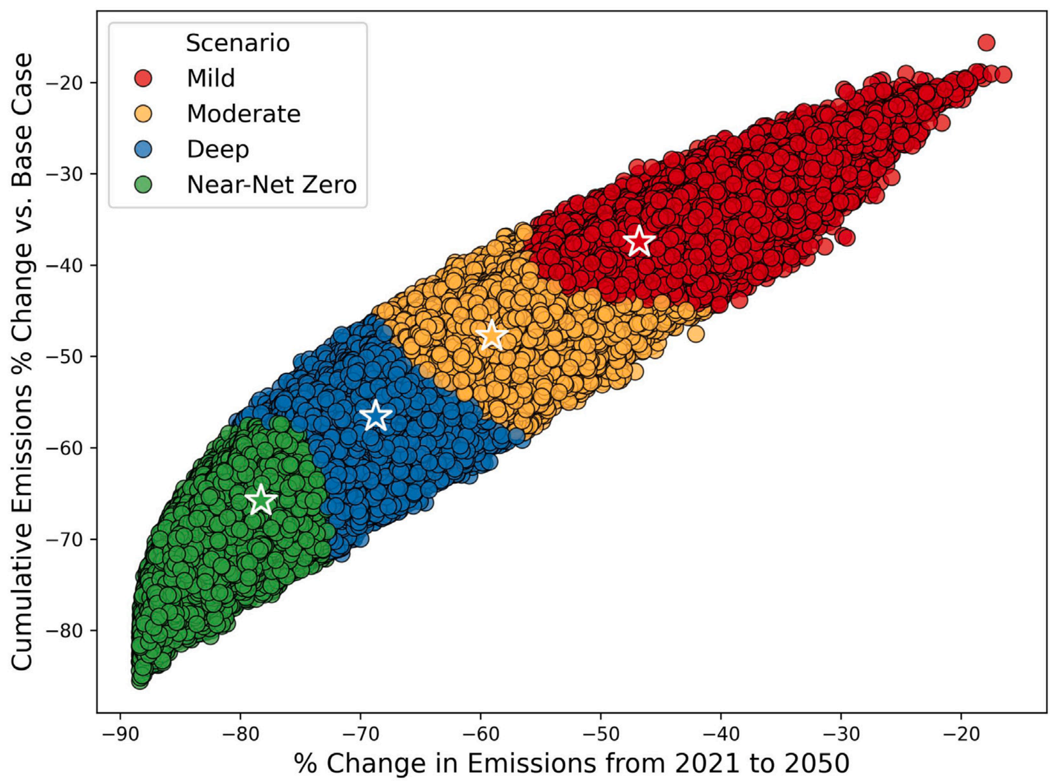


Fig. 2. Scatter plot of the four decarbonization scenarios discovered (Mild, Moderate, Deep, and Near-Net Zero), with star markers indicating the centroid of each, based on the mean value of the clustering metrics (% change in emissions from 2021 to 2050 and cumulative emissions % change vs. base case).

3.2. Scenarios are distinguished by pace and scale of activity changes

The decarbonization pathway relies on multisectoral contributions, which vary depending on the scenarios (see Fig. 5). We find that higher electrification rates, improved fuel economy (MPG), reduced truck VMT, and commercial emissions, as well as e-Grid emissions factors are strongly associated with *Deep* and *Near-Net Zero*. Specifically, an electrification rate increase of approximately 0.5% (±0.25%), a 0.75% improvement in fuel economy (MPG), a 0.75% reduction in commercial emissions, and a 0.6% decrease in truck VMT are required to achieve *Near-Net Zero*. In addition, *Near-Net Zero* requires earlier and more aggressive transformations in most sectors, particularly in transportation electrification, industrial emissions, and decarbonization of the electricity grid. In contrast, *Mild* reflects delayed or less ambitious changes, with relatively modest parameter changes by the last year on the horizon. The results illustrate that a successful decarbonization pathway requires not only sector-specific action but also cross-cutting coordination, investment, and equity-focused implementation.

3.3. Residential sector is the largest emitter by 2050

Although emissions in other sectors decline and stabilize by 2050, residential sector emissions show little or no change throughout the 2021–2050 period, as shown in Fig. 6. In *Near-Net Zero* and *Deep* scenarios, mobile combustion emissions fall rapidly and soon plateau. Electricity consumption and commercial emissions also witness rapid decreases and a large reduction by 2050. Surprisingly, the residential sector remains almost unchanged over the planning horizon and becomes the largest contributor to emissions by 2050 across all sectors in both scenarios, likely due to constant energy use patterns and limited mitigation measures targeting household energy consumption, highlighting the need to decarbonize residential buildings and prioritize domestic energy demand.

3.4. Grid, buildings, and freight are most relevant to emissions changes

SHapley Additive exPlanations (SHAP) values were plotted for each scenario to investigate whether there were differences in the

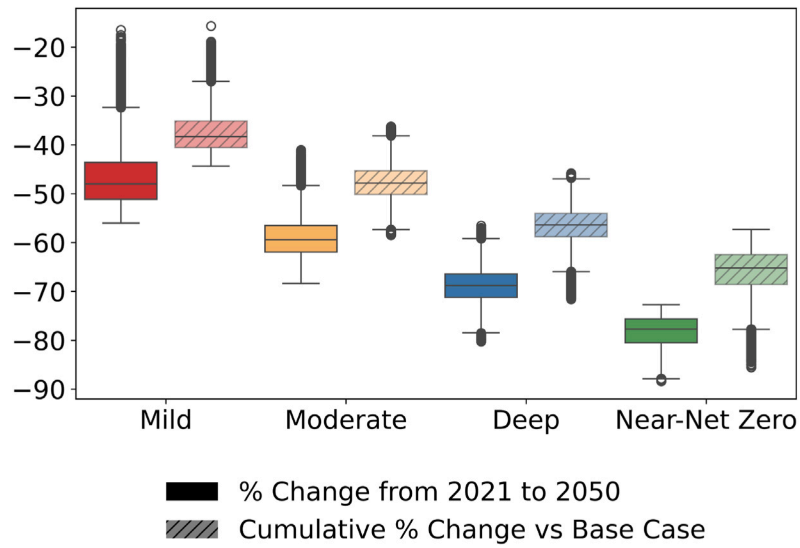


Fig. 3. Boxplot of the values of two emissions metrics (% change from 2021 to 2050 and cumulative % change vs. base case) by scenario types.

classification of activity parameters among the top ten in each type in the prediction model to predict. We observe that the driving factors for EPC and CEPC differ significantly (see Figs. 7 and 8 A to D). In terms of CEPC, we found that the year of achieving the commercial emissions target is ranked as the first in *Near-Net Zero*, followed by electricity consumption and emissions factors from the e-grid (as shown in Fig. 7). The target year for truck VMT and truck VMT are among the five most influential factors. For the rest of the scenario, electricity consumption and e-grid emissions factors are the two strongest contributors.

For EPC (as shown in Fig. 8), the consumption of residential natural gas outweighs other factors in most scenarios. Commercial emissions are among the most significant in *Near-Net Zero* and *Deep*, followed by the year in which the target value in commercial emissions is reached as well as the e-grid emissions factors. Truck VMT, electricity consumption, car VMT, industrial emissions are among the few highly influential factors. These results indicate that targeted energy management is needed in residential sectors, as it is among the largest contributors to deep decarbonization, which corroborates the results in Fig. 5.

4. Discussion

4.1. Robust decarbonization requires prioritizing actions across multiple sectors

The four representative decarbonization scenarios (*Near-Net Zero*, *Deep*, *Moderate*, *Mild*) reveal multisector pathways to decarbonization. For example, achieving *Near-Net Zero* requires substantial progress across several sectors, including promoting electrification, improving vehicle fuel economy, and reductions in truck VMT. This suggests that coordinated policy interventions across different sectors (transportation, freight, and building construction) are important in achieving emissions reduction. In addition to identifying sectoral contributions, the scenario discovery framework also provides insight into the early progress in key sectors, particularly in transportation, grid decarbonization, and improvement in fuel economy. Delaying the implementation of these measures reduces the range of policy options available to achieve the target.

4.2. Residential emissions mitigation and early commercial action identified

Residential natural gas consumption is the most significant factor in emissions reduction, as shown in Fig. 8. Besides, we found that there is a growing contribution of residential emissions over the planning horizon, which suggests that residential energy consumption becomes an increasingly important bottleneck to achieve deep decarbonization targets. Based on this, a combination of policy levers to address this problem includes promoting building electrification, targeting retrofit programs for inefficient housing stocks, establishing stronger building energy codes for new building construction, switching fuels, and financial incentives to help the cost barriers associated with building retrofits and expansion and installation of heat pumps [42]. Those measures can ensure that the residential sector keeps pace with the progress of decarbonization in the transportation and electricity systems.

In addition, we also found that the target year for achieving commercial emissions reduction strongly influences cumulative emissions percentage change (CEPC), which suggests that the timing of decarbonization in the commercial sector is essential to determine the overall emissions outcomes and the delay in reaching emissions reduction may limit the ability to achieve the system-wide decarbonization targets. From a policy standpoint, this highlights that early intervention or action in commercial sector such as commercial building electrification may reduce long-term mitigation pressures on other sectors.

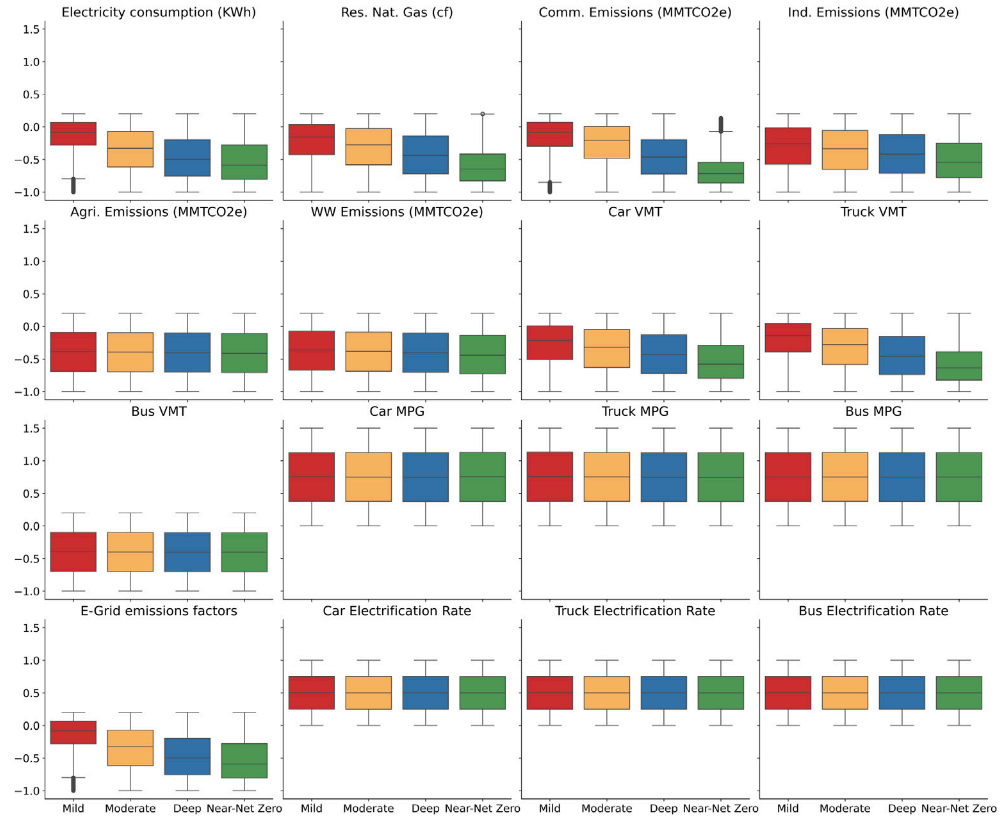
4.3. Study limitations

One limitation of this work is that it relies on aggregated and top-down data and simplified behavioral and technological assumptions, which may not fully capture local heterogeneity in energy use, behavioral responses, and technology adoption. Those limitations are due to the lack of sufficient and granular data. Future work could address this by integrating more comprehensive datasets, higher-resolution spatial information, and empirically grounded behavioral models to better represent real-world system dynamics. In addition, trade-off between cost and decarbonization targets for a more robust analysis of pathways should also be considered.

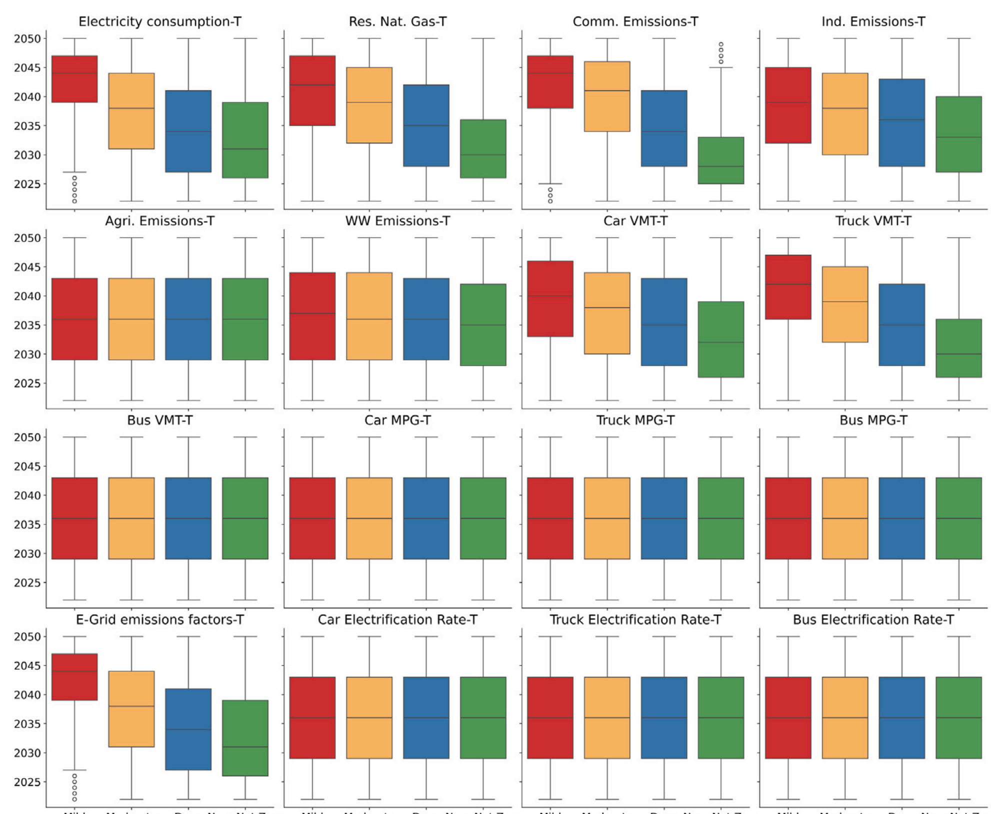
While this study incorporates electricity consumption associated with electric vehicle adoption in the model, a more detailed investigation of how changes in electricity and electric vehicle charging demand interact with other strategies will be conducted, as explored in this study [43]. We plan to explore interaction effects among key parameters, such as the combined impact of electrification and reductions in vehicle miles traveled, to identify potential nonlinear synergies.

5. Conclusions

This study applied scenario discovery and backcasting emissions forecasts as exploratory tools to identify potential decarbonization



(a)



(b)

Fig. 4. Boxplots of activity parameter values and their corresponding target years for achieving emissions reductions across scenario types.

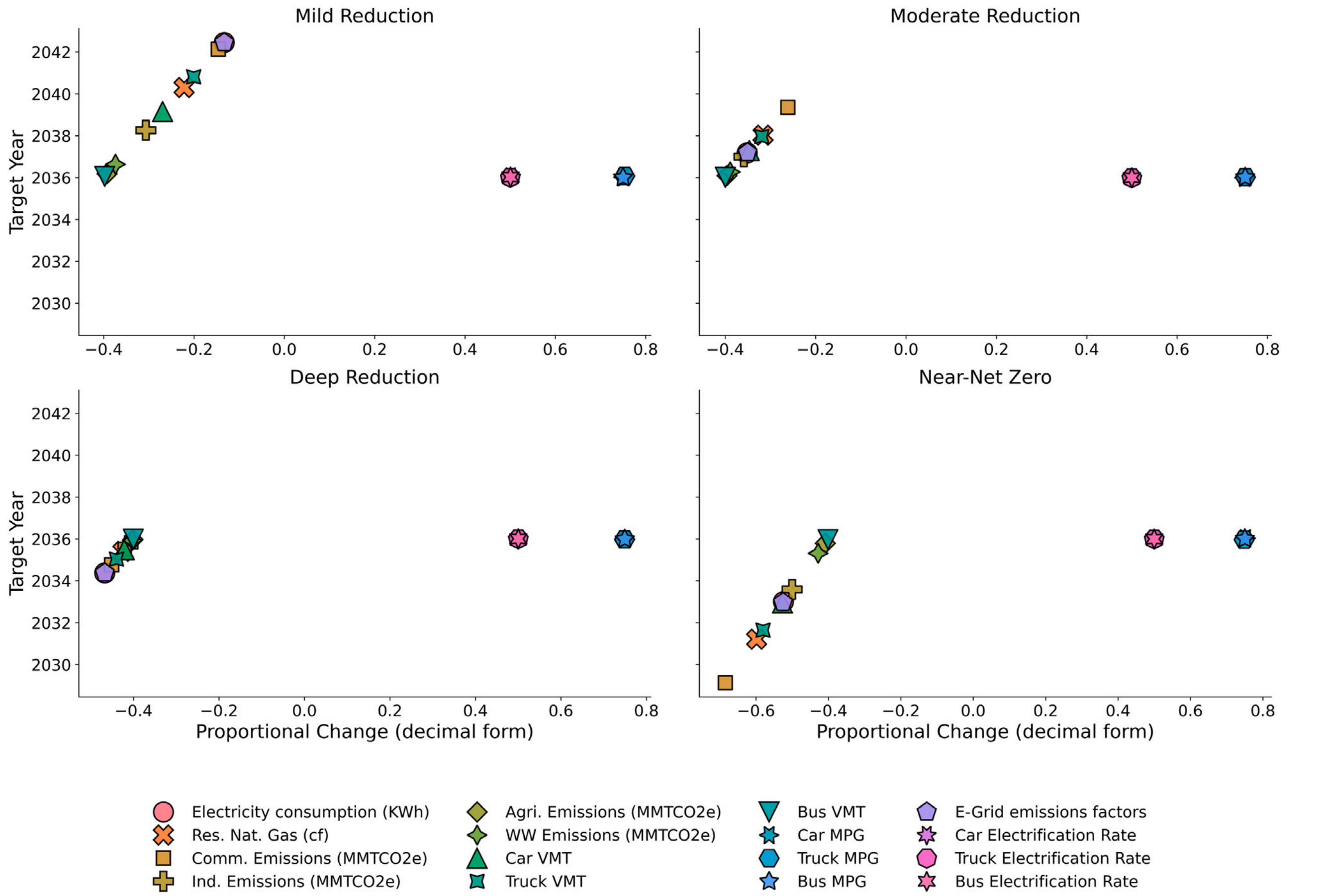


Fig. 5. Decarbonization pathways: the year reaching the target are indicated on the y-axes while corresponding activity parameter changes are shown on the x-axes for each of the scenarios.

pathways. We identified four scenarios: *Mild*, *Moderate*, *Deep*, and *Near Net-Zero*, each corresponding to a different decarbonization target. Each scenario demonstrates different combinations of activity parameters and the year of reaching that targets. *Near Net-Zero* is achieved when these parameters move beyond the midpoint of their explored decarbonization bounds, accompanied by an earlier timing of these transitions. We also found that the residential emissions stay consistent over the planning horizon while emissions from other sector decrease. This suggests a need to prioritize residential energy and emissions management. Furthermore, our XGBoost model highlights the need for greater emissions mitigation efforts in the residential sector. Future work will further investigate the cost–benefit analysis of these decarbonization targets and the potential synergies and interaction effects among different activity parameters. While this study focuses on specific metropolitan regions, the framework could be used to study other urban areas with similar socioeconomic and energy system characteristics. By combining the exploratory scenario generation and planning with policy insights from the model outcomes, this framework can support the climate planning efforts to achieve ambitious emissions targets with uncertainties in the future development.

6. Data & methods

We simulated emissions under varying growth or decay trajectories of activity parameters and factors in all sectors. Firstly, we projected a “Base Case” scenario using the ARIMA-based framework from year 2021 to 2050. Then, we applied LHS to sample uncertain growth rates and the time for target achievement for 16 activity parameters (32 vectors in total) to a diverse set of futures. Afterwards, we applied cluster analysis based on two metrics: emissions percentage change from 2050 to 2021 (EPC) and cumulative emissions percentage change (CEPC). This is to identify the emissions patterns and discover effective emissions reduction pathways.

6.1. Data

We collected activity parameters to calculate the GHG emissions from the following sectors: *Mobile Combustion*, *Electricity Consumption*, *Stationary Combustion* (comprising *Residential*, *Commercial*, and *Industrial* subsectors), *Agriculture*, and *Wastewater*. The data sources for those activity parameters are summarized in Table 1.

6.2. Base case forecast using ARIMA

We calculated the base case emissions trajectory, also called “Base Case” scenario, denoted as $E^{\text{base}}(t)$ for years $t \in \{2021, \dots, 2050\}$ from predicted activity parameters and factors from 2021 to 2050 using a time series ARIMA model. The model is specified as $ARIMA(p, d, q)$, where p indicates the number of autoregressive (AR) terms (lagged observations), d represents the order of nonseasonal differences required to make the time series stationary, and q stands for the order of moving average (MA) processes. The model equation is thus:

$$x_0(t) = \mathcal{ARIMA}\big(x_0; t, p, d, q\big) = \alpha + \sum_{j=1}^{p} \phi_j x_0^{(d)}(t-j) + \varepsilon_t + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} \tag{1}$$

$$x_0^{(d)}(t) = x_0^{(d-1)}(t) - x_0^{(d-1)}(t-1), \quad d \geq 1 \tag{2}$$

$$x_0^{(0)}(t) \equiv x_0(t) \tag{3}$$

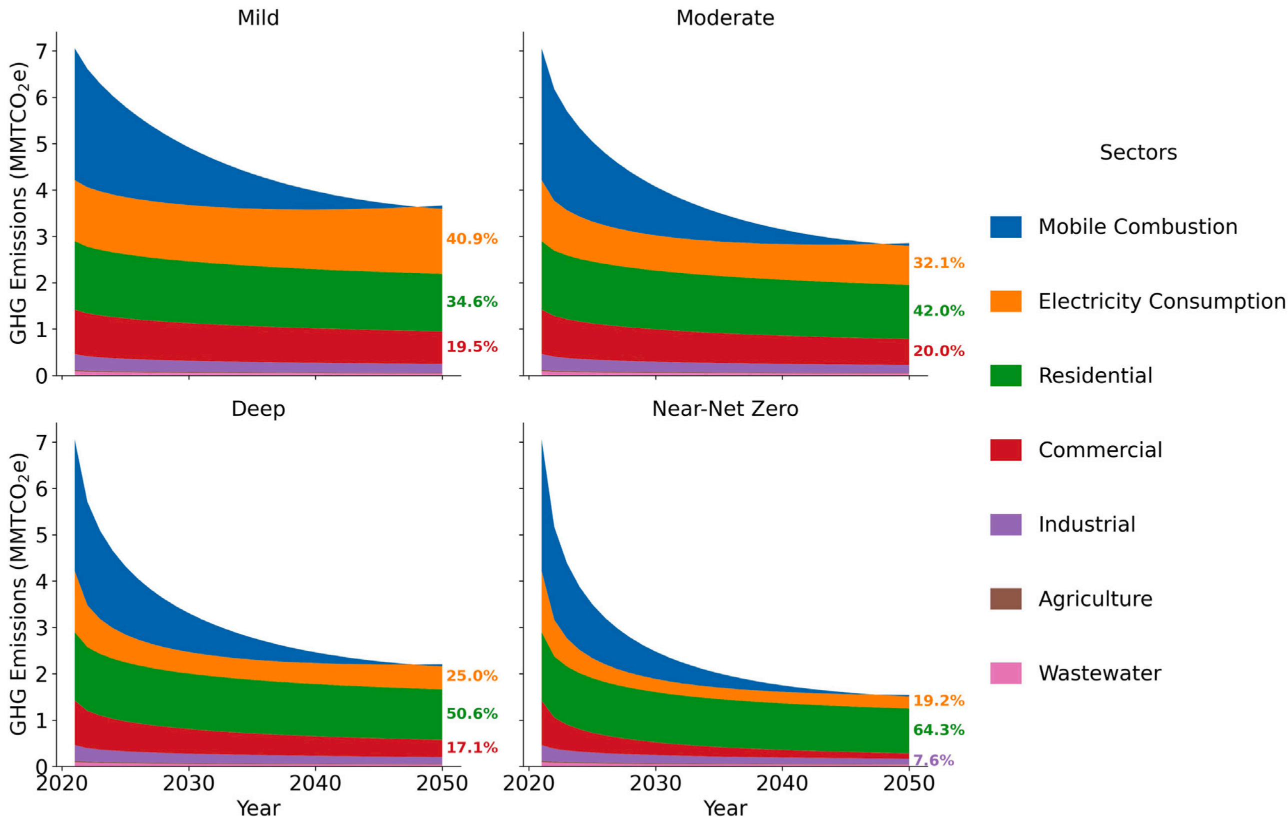


Fig. 6. Sectoral breakdown of projected emissions trajectories for each decarbonization scenario across the planning horizon from 2021 to 2050.

Table 1
GHG emissions sectors, input data, and sources.

Sector	Inputs	Year	Source
Mobile Combustion	Vehicle miles traveled (VMT)	1966–2022	CTDOT [44]
	Statewide vehicle type distribution	2021	FHWA, DOT [45]
	Vehicle fuel efficiency	2021	LGGIT [46]
Electricity Consumption	Electricity consumption	2016–2022	EnergizeCT [47]
Stationary Combustion	Household heating fuel consumption	2021	EIA [48]
	Number of households using each heating fuel	2010–2021	ACS [49]
	Statewide commercial emissions	1990–2021	DEEP [50]
	Industrial facility emissions	2010–2022	FLIGHT [51]
Agriculture	Statewide agricultural emissions	1990–2021	DEEP [50]
Wastewater	Statewide wastewater emissions	1990–2021	DEEP [50]

$x_0(t)$ denotes the activity parameters and associated factors used to calculate the base case emissions, where the subscript 0 indicates the base case level and t represents time in the range $[0, T]$ (2021–2050). α is the intercept (constant term), ϕ_j are the autoregressive coefficients ($j \in \{1, \dots, p\}$), ε_t is the error term (or innovation) at time t ($\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$), and θ_k are the moving average or lagged forecast error coefficients with ($k \in \{1, \dots, q\}$). We then used the fitted model to forecast activity parameters and factors from 2021 to 2050, assuming no additional policies or interventions applied. The coefficient and performance of ARIMA model are listed in Table 2. We selected the optimal value for ARIMA(p, d, q) model using a grid-search procedure, which was then systematically evaluated using $p \in \{0, \dots, 8\}$, $d \in \{0, 3\}$, and $q \in \{0, \dots, 8\}$. For each parameter combination, we assessed the performance of the model using maximum likelihood on the training set (70% of observations), and then generated the forecast for the remaining 30% (test set). We examined the performance of the forecast using mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean squared error (RMSE). We chose the final order of the model (p^*, d^*, q^*) based on the combination that minimized the MAPE outside the sample in the test set. The coefficients for the ARIMA model, including the autoregressive, moving average terms, and the intercept, are presented in Table 3.

6.3. Scenario discovery design

6.3.1. Latin hypercube sampling of uncertain parameters

We jointly sampled 32 uncertain parameters, consisting of 16 activity parameter values in year 2050 $x_{i,u}^T$, and 16 target years at which these levels are reached ($\tau_{i,u}$) from LHS, with i indicating each state and u indicating different types of uncertain parameters. Thus, each state s_i can be expressed as:

$$s_i = [\boldsymbol{x}_{i,u}^{\tau}, \boldsymbol{\tau}_{i,u}] = \left[x_{i,1}^T, x_{i,2}^T, \dots, x_{i,16}^T, \tau_{i,1}, \tau_{i,2}, \dots \tau_{i,16}\right] \tag{4}$$

The lower and upper bounds of these uncertain parameters are shown in Table 4.

The parameter ranges (upper and lower bound) used in the Latin Hypercube Sampling were selected to reflect plausible transition pathways consistent with policy targets and findings from the recent energy

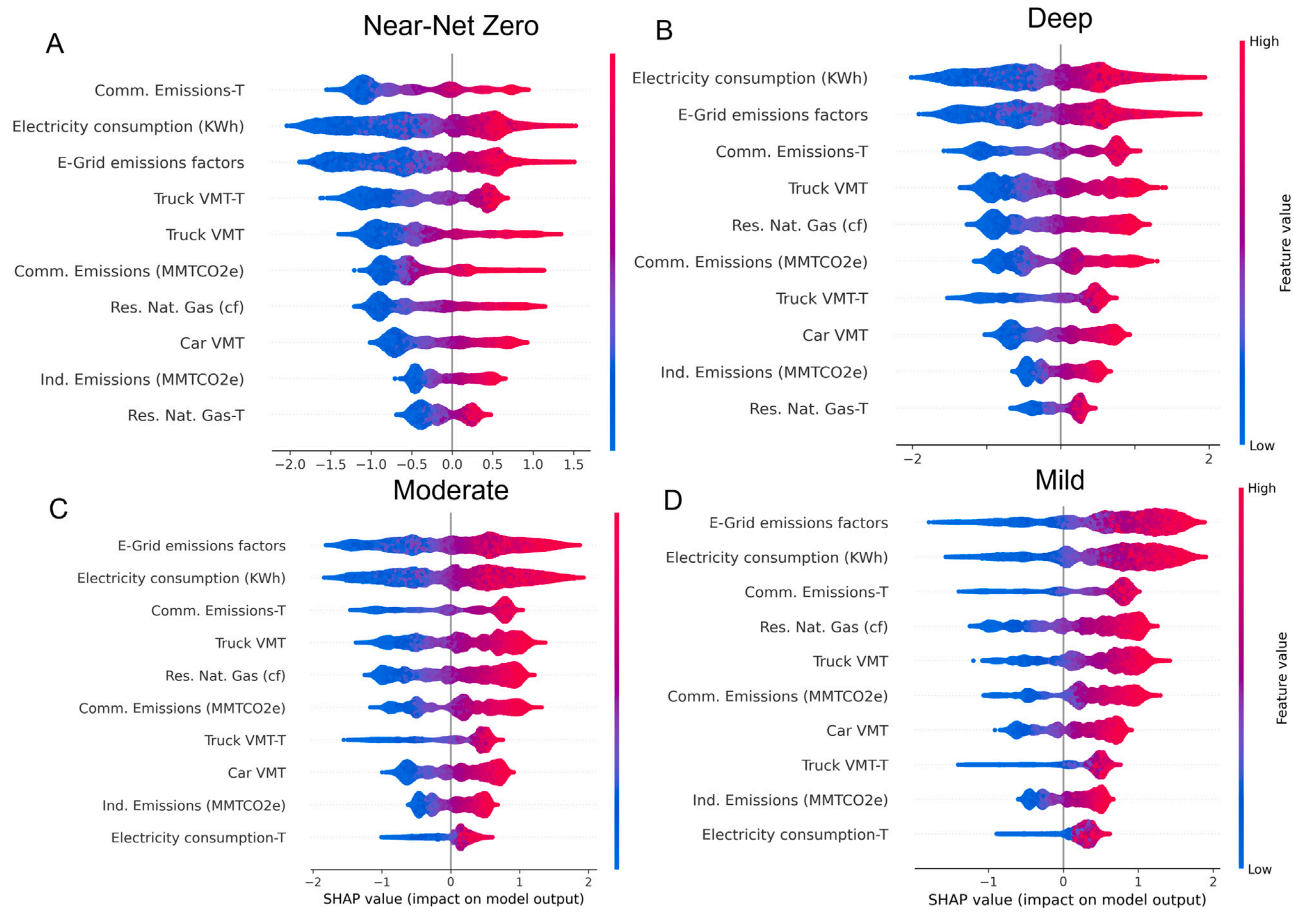


Fig. 7. Indicators relevant to the prediction of cumulative emissions percentage change versus base case (CEPC) ranked based on SHAP values from most to least important: (A)–(D) violin plots of the top 10 indicators for each scenario, individually.

Table 2
Summary of ARIMA model configurations and performance by sector.

Sector	Activity parameters	(p,d,q)	# of local observations	Test MAPE (%)
Mobile combustion	Car VMT	(0,1,0)	40	1.00
	Truck VMT	(6,1,4)	40	4.53
	Bus VMT	(1,2,7)	40	5.21
Electricity consumption	Electricity consumption	(0,0,1)	7	3.61
Stationary combustion (residential)	Natural gas consumption	(2,2,2)	32	2.52
Stationary combustion (commercial)	Emissions	(4,0,0)	32	5.09
Stationary combination (industrial)	Emissions	(1,0,0)	32	9.6
Agriculture	Emissions	(1,2,0)	32	4.6
Wastewater	Emissions	(1,1,4)	32	0.33

transition literature. For activity variables such as sector emissions and vehicle miles traveled, we chose the range of −100% to 20% relative change by 2050 to capture a wide range of potential states, including aggressive mitigation measures, as well as moderate growth reflecting continued economic and population growth. We selected the range for the technology improvement parameters including vehicle fuel economy (MPG), from 0 to 150%, which shows long-term efficiency improvements. We sampled the electrification rates for cars, trucks, and buses between 0 and 100%, which represents the full spectrum from limited electrification to complete electrification in deep decarbonization scenarios. Finally, we allowed the electricity grid emissions factors to decrease by up to 100% to represent the complete decarbonization of the power sector, consistent with net-zero electricity pathways discussed in recent work on scenario-based energy system analyzes [52,53].

6.3.2. Backcasting emissions trajectories for all futures

We modeled the trajectory of activity parameters $x^t_{i,u}$ over time for each future using the exponential decay function applied to the ARIMA model prediction in each time step t:

$$x^t_{i,u} = x^t_{0,u} \cdot \exp\left(-\gamma_i \cdot t\right) \tag{5}$$

This ensures that $x^t_{i,u} = x^T_{i,u}$ when $t = \tau_{i,u}$, regardless of whether the target is hit before 2050. Thus, the growth or decay rate γ is computed as:

$$\gamma_i = -\frac{1}{\tau_{i,u} - 2021} \cdot \ln\left(\frac{x^T_{i,u}}{x^T_{0,u}}\right) \tag{6}$$

$x^T_{i,u}$ is the projected value of activity parameter u in 2050 in the future i, while $x^T_{0,u}$ is the projected value in 2050 for the Base Case scenario. $\tau_{i,u}$ is the target year in which the desired activity value $x^T_{i,u}$ is reached.

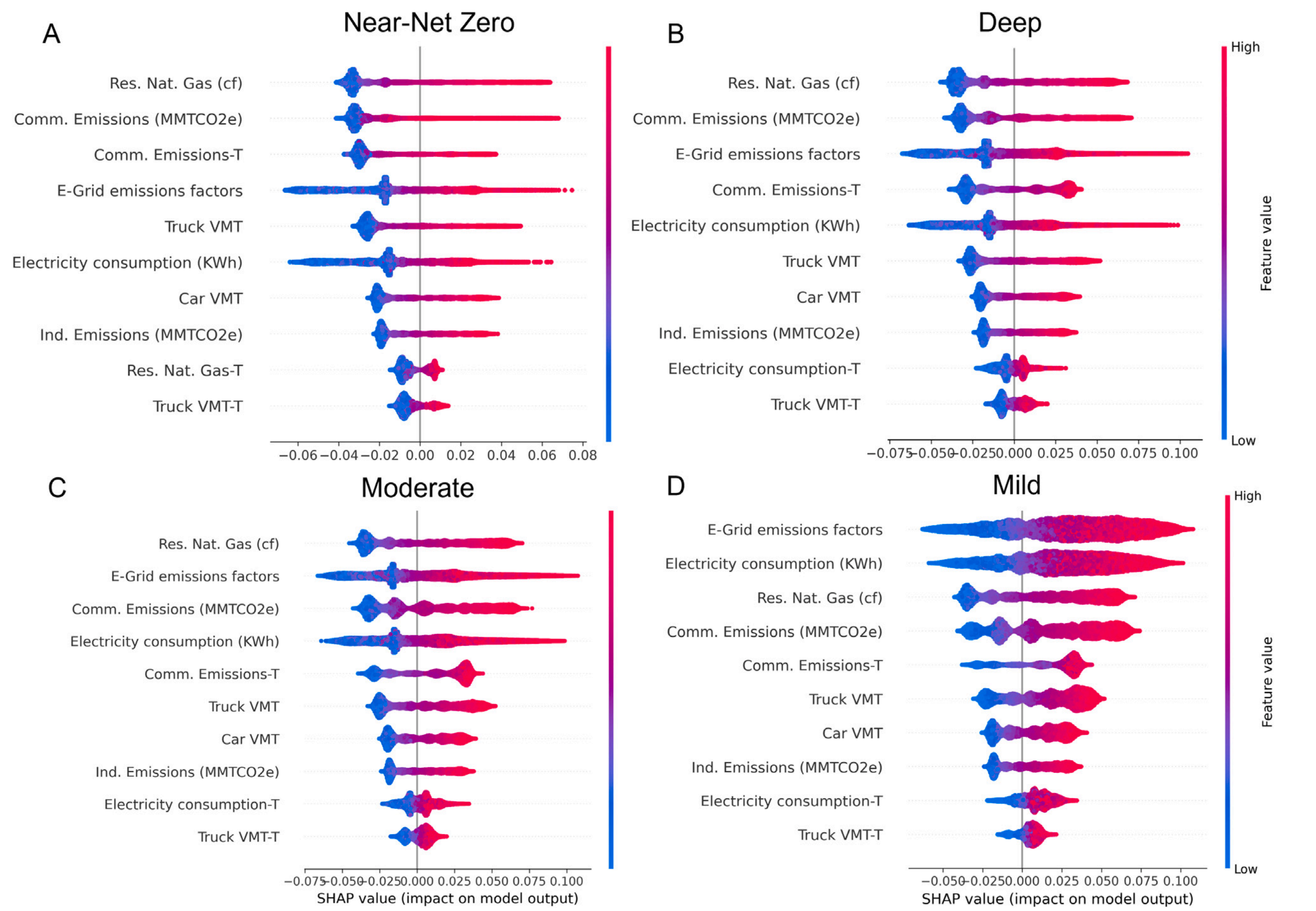


Fig. 8. Indicators relevant to the prediction of emissions percentage change from 2050 to 2021 (EPC) ranked based on SHAP values from most to least important: (A)–(D) violin plots of the top 10 indicators for each scenario, individually.

6.3.3. Sensitivity test across decay functions

To assess the sensitivity and robustness of the exponential functions, we conducted the same study using other two functional forms to model the temporal evolution of emissions: linear and logistic. These functions differ in the assumptions about the pace and acceleration of emissions reductions over time. The linear function assumes a constant rate of change, the exponential function models a continuously compounding rate of the change, and the logistic function shows a saturating effect as emissions approach a lower bound, as shown in Fig. 9. By applying these functions to the same set of activity parameters, we quantified how variations in functional form influence the projected timing and magnitude of emissions reductions.

The linear function is given in Eq. (8), and the logistic function is given in Eq. (9).

$$x^t_{i,u} = x^t_{0,u} - \gamma_i \cdot t \tag{7}$$

where $x^t_{0,u}$ is the initial emissions level, and γ_i is the linear reduction rate.

$$x^t_{i,u} = x_{\min} + \frac{x_{\max} - x_{\min}}{1 + \exp\left(-\gamma_i \cdot (t - t_0)\right)} \tag{8}$$

where $x_{\min}$ and $x_{\max}$ are the minimum and maximum achievable emissions, respectively. γ_i is the logistic rate parameter, and t_0 is the inflection point at which the reduction rate begins to slow.

We compared the sensitivity results generated using the three trajectory functions shown in Fig. 10. The exponential function produced the largest average emissions reduction, while the outputs from the linear and logistic functions were nearly identical. Therefore, in this study, we proceeded with the exponential function to compute the emissions trajectories.

6.3.4. Emissions futures calculation

For sectors with a given activity (e.g. miles driven by a certain vehicle type or amount of electricity consumed), such as *Mobile Combustion*, *Electricity Consumption*, and *Residential* subsector, the relevant emissions factors were multiplied to obtain the GHG emissions estimate shown below. This equation can be applied to all projected activity parameters and factors to calculate the emissions for the sector mentioned above.

$$G^s_{i,t} = \rho^s_{t,i} x^t_{i,u} \tag{9}$$

where $G^s_{i,t}$ are the direct emissions from fuel combustion in year t in the future state i for sector s, ρ is the specific GHG emissions factor — the amount of a given GHG emitted per unit of a certain activity — and $x^s_{t,i}$ is the quantitative measure of activity.

For the remaining sectors (*Commercial*, *Industrial*, *Agriculture*, and *Wastewater*), the projections directly represent the emissions and therefore do not require additional calculation.

The total emissions, which represent each future state, are calculated by summing the projected emissions across all sectors.

$$G_{i,t} = \sum_s G^s_{i,t} \tag{10}$$

6.3.5. Emissions metrics

We calculated two metrics: percentage change in emissions between 2050 and 2021 (EPC) and cumulative emissions percentage change relative to the base case ($CEPC_i$). The equations for calculating these metrics are shown below. The histogram of the distribution of EPC and CEPC is shown in Fig. 11.

$$EPC_i = \frac{G_{i,T} - G_0}{G_0} \tag{11}$$

Table 3
Estimated ARIMA coefficients and statistical significance by sector.

Sector	Activity	(p,d,q)	Coefficient	Value	**p-value**
Mobile combustion	Car VMT	(0,1,0)	σ^2	5.00e−04	0.00
	Truck VMT	(6,1,4)	ϕ_1	−0.04	0.99
			ϕ_2	0.79	0.55
			ϕ_3	0.10	0.98
			ϕ_4	0.20	0.89
			ϕ_5	−0.05	0.92
			ϕ_6	−0.05	0.89
			θ_1	0.30	0.99
			θ_2	−0.82	0.97
			θ_3	−0.31	0.97
			θ_4	−0.19	0.97
			σ^2	3.00e−04	0.96
	Bus VMT	(1,2,7)	ϕ_1	−0.43	0.05
			θ_1	0.24	0.20
			θ_2	−0.14	0.71
			θ_3	0.08	0.85
			θ_4	0.12	0.85
			θ_5	−0.17	0.78
			θ_6	0.20	0.58
			θ_7	−0.14	0.79
			σ^2	4.60e−07	0.01
Electricity consumption	Electricity consumption	(0,0,1)	c	5.45e+06	0.00
			ϕ_1	0.28	1.00
			σ^2	1.08e+10	0.00
Stationary combustion (residential)	Natural gas	(2,2,2)	c	0.56	0.00
			ϕ_1	−0.36	0.33
			ϕ_2	−0.21	0.48
			θ_1	−4.69e−05	1.00
			θ_2	−1.00	1.00
			σ^2	0.17	1.00
Stationary combustion (commercial)	Emissions	(4,0,0)	c	−0.05	0.97
			ϕ_1	0.51	0.10
			ϕ_2	0.35	0.25
			ϕ_3	0.00	0.99
			ϕ_4	−0.20	0.75
			σ^2	0.68	0.00
Stationary combustion (industrial)	Emissions	(1,0,0)	c	0.72	0.37
			ϕ_1	0.86	0.00
			σ^2	0.16	0.02
Agriculture	Emissions	(1,2,0)	ϕ_1	−0.62	0.00
			σ^2	3.50	0.00
Wastewater	Emissions	(1,1,4)	ϕ_1	0.17	0.79
			θ_1	0.33	1.00
			θ_2	−0.04	1.00
			θ_3	0.33	1.00
			θ_4	1.00	1.00
			σ^2	9.00e−03	1.00

Table 4
Uncertainty bounds for the 16 activity parameters used in Latin Hypercube Sampling (LHS). For transparency and reproducibility, lower and upper proportional change bounds are explicitly reported for each parameter, even when identical across groups. The target year range ($\tau_{i,u}$) is consistent across all parameters: 2021–2050.

Label	Parameter	% Change lower	% Change upper	Year lower	Year upper
1	Electricity Consumption	−1	0.2	2021	2050
2	Residential Natural Gas Consumption	−1	0.2	2021	2050
3	Commercial Emissions	−1	0.2	2021	2050
4	Industrial Emissions	−1	0.2	2021	2050
5	Agricultural Emissions	−1	0.2	2021	2050
6	Wastewater Emissions	−1	0.2	2021	2050
7	Car VMT	−1	0.2	2021	2050
8	Truck VMT	−1	0.2	2021	2050
9	Bus VMT	−1	0.2	2021	2050
10	Car MPG	0	1.5	2021	2050
11	Truck MPG	0	1.5	2021	2050
12	Bus MPG	0	1.5	2021	2050
13	E-Grid Emissions Factors	−1	0.2	2021	2050
14	Car Electrification Rate	0	1	2021	2050
15	Truck Electrification Rate	0	1	2021	2050
16	Bus Electrification Rate	0	1	2021	2050

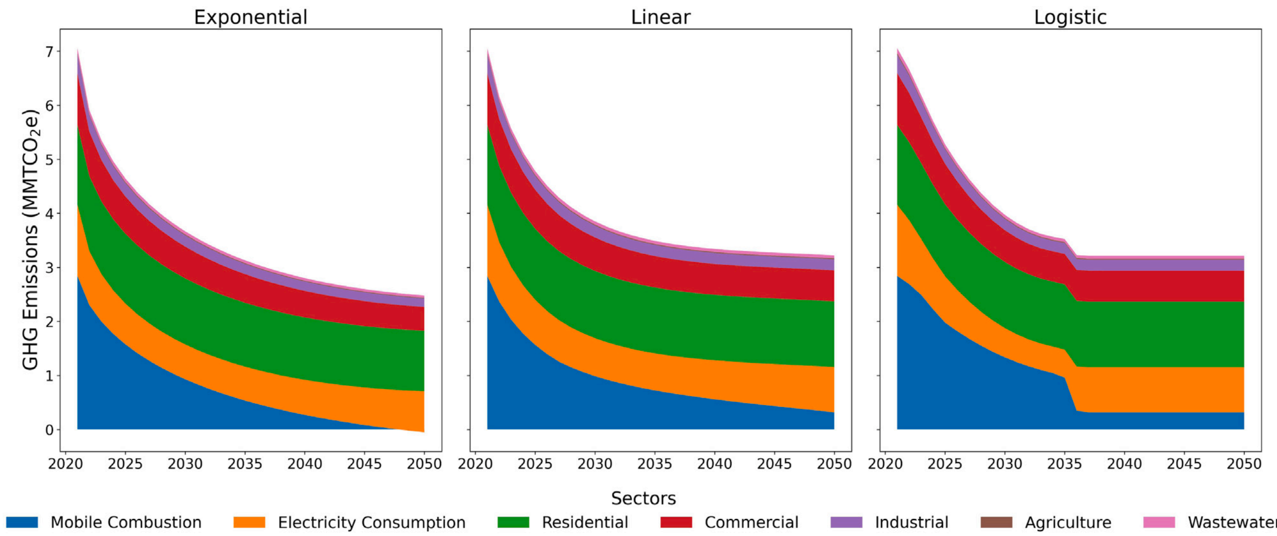


Fig. 9. Sectoral projections of GHG emissions trajectories from 2021 to 2050 under exponential, linear, and logistic decay functions.

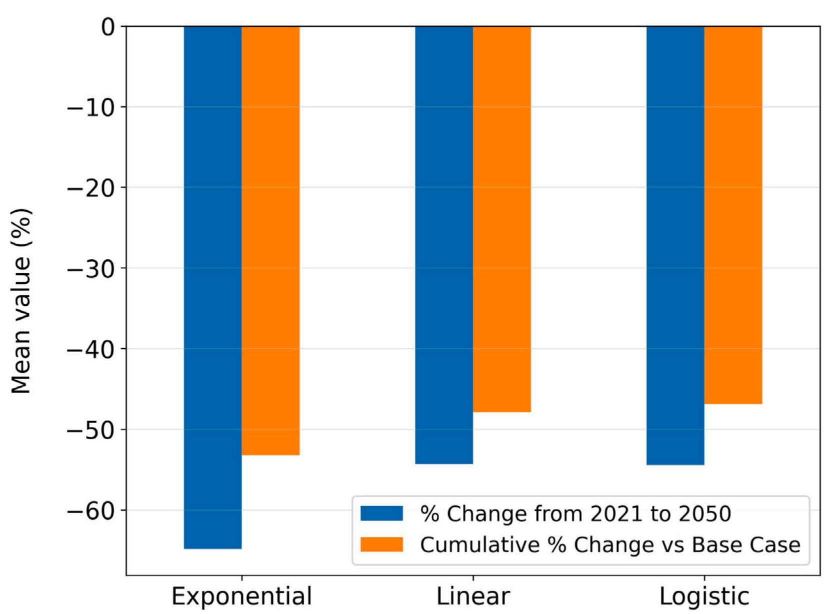


Fig. 10. Comparison of cumulative emissions percentage change versus percentage change in emissions from 2050 to 2021 computed using exponential, linear, and logistic decay functions.

Table 5
Performance metrics of the multi-output XGBoost model for the two emissions reduction outcomes: percentage change in emissions from 2021 to 2050 (EPC), cumulative emissions percentage change vs. base case (CEPC).

Dataset	EPC		CEPC	
	R^2	MAPE	R^2	MAPE
Training	0.9942	0.0098	0.9932	0.0120
Test	0.9941	0.0099	0.9931	0.0122

$$CEPC_i = \frac{\sum_{t=0}^{T} G_{i,t} - \sum_{t=0}^{T} G_{0,t}}{\sum_{t=0}^{T} G_{0,t}} \tag{12}$$

where $G_{0,t}$ is the base case emissions in year t and $G_{i,T}$ is the total emissions in year T in each future i. We note that the emissions in year 0 (2021) is the same across all futures (including the base case), hence we simply denote this as G_0.

6.3.6. Cluster analysis

We clustered all futures into groups based on their emissions outcomes (EPC and CEPC). We used the Gaussian Mixture Model (GMM) approach to perform the clustering analysis, which assumes that the data are generated from a mixture of K Gaussian distributions [54,55]. The GMM is specified as:

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{13}$$

where $\boldsymbol{x}$ is a multivariate observation (factor scores of each metro area), $\boldsymbol{\theta}$ is the vector of model parameters $(\boldsymbol{\pi}, \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\})$ and π_k is the weight of the mixture with $\sum_{k=1}^{K} \pi_k = 1$ and $0 \leq \pi_k \leq 1$. $\mathcal{N}(\boldsymbol{x}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the multivariate Gaussian, where $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mean and covariance matrix of the kth Gaussian component, respectively. We estimated $\boldsymbol{\theta}$ via the expectation–maximization algorithm [56]. Following parameter estimation, the posterior probability of membership (also known as the *responsibility*) r_{nk} of each group k for observation $\boldsymbol{x}_n$ is obtained as

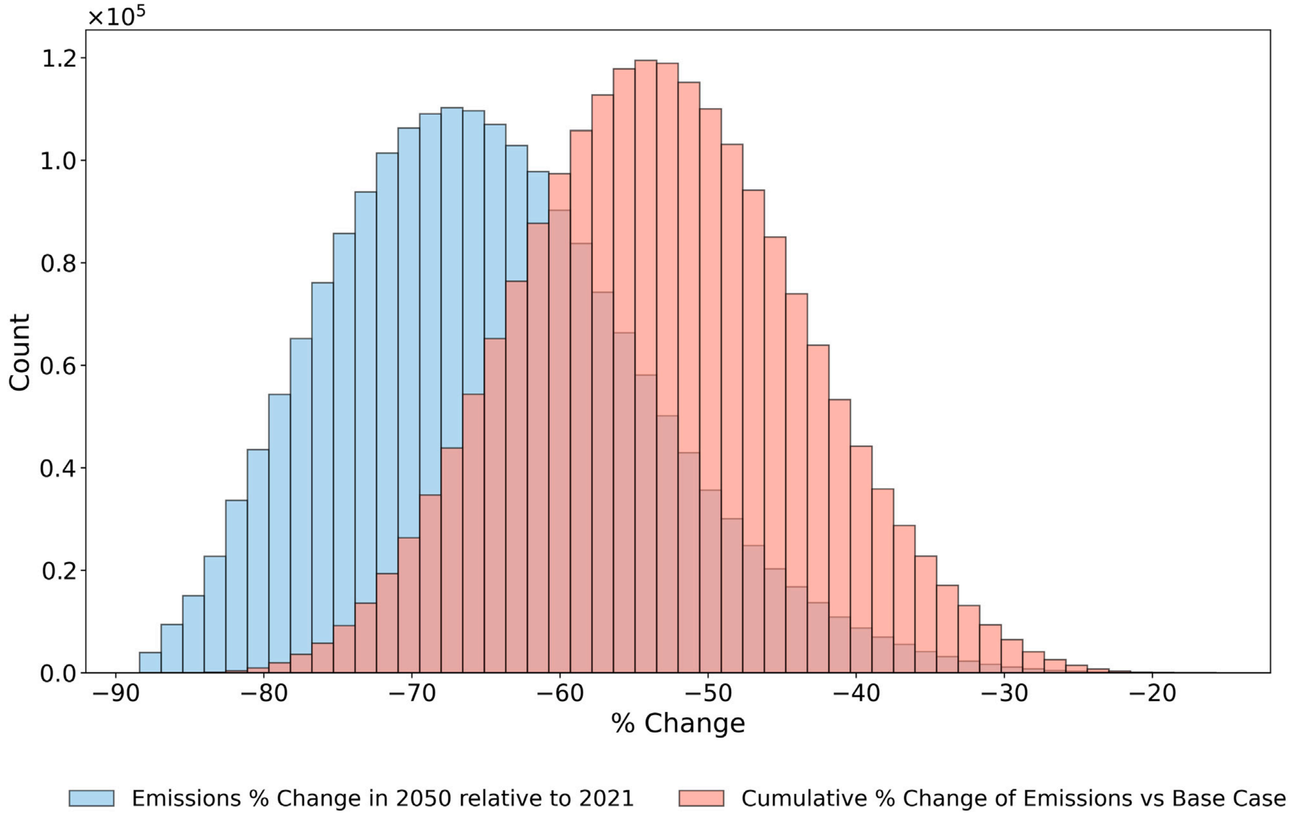


Fig. 11. Histogram of emissions percentage change from 2050 to 2021 (EPC) and cumulative emissions percentage change (CEPC) for all generated future emissions (scenarios).

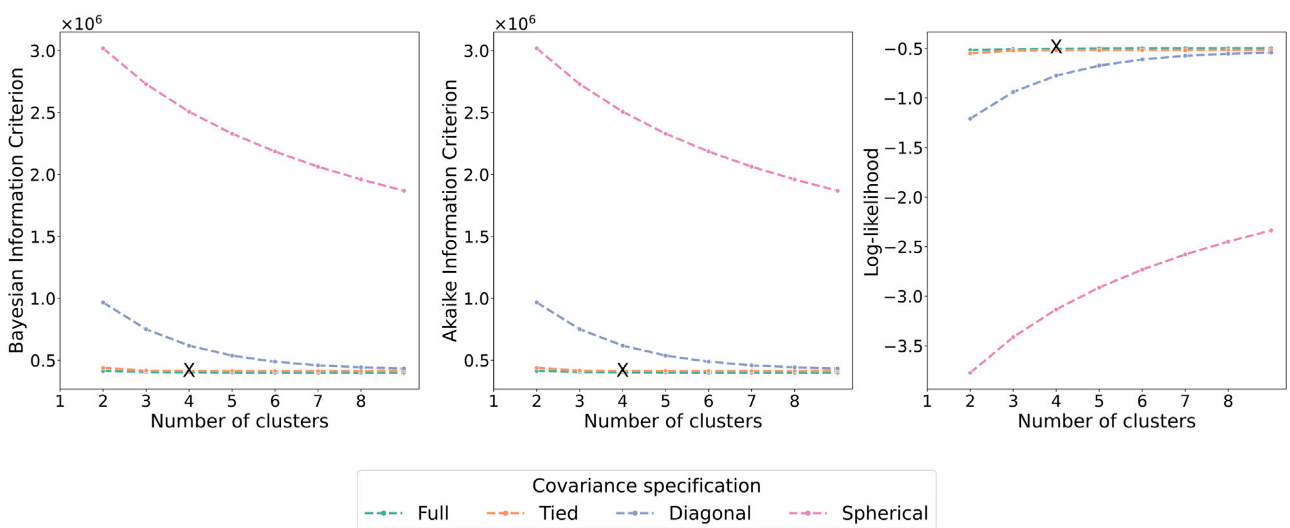


Fig. 12. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), and Log-Likelihood of grid search results for Gaussian Mixture Model (GMM) with number of components ranging from 2 through 10, and four different covariance specifications.

$r_{nk} = p(z_n = k|\boldsymbol{x}_n, \theta)$. The underpinning of this is a latent categorical variable $z \in \{1, \dots, K\}$ that generates the observations $\boldsymbol{x}$. Observations are then assigned to the responsibility-maximizing cluster.

To determine the optimal number of clusters (components), we performed a grid search, evaluating various numbers of clusters and covariance specifications (spherical, tied, diagonal, and full) to identify the model with the highest log-likelihood score and the lowest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). We chose 4 clusters and a full covariance specification that yielded the highest log-likelihood, BIC, and AIC value, as shown in Fig. 12.

6.4. Predictive modeling

Using 16 activity parameters and their target years as input, we fitted an Extreme Gradient Boosting (XGBoost) model [57] to predict and explain emissions reduction outcomes. We considered two response variables—the percentage change in emissions from 2021 to 2050 (EPC) and the cumulative emissions percentage change versus base case (CEPC), which then followed by implementing a multi-output regression framework. This framework uses the MultiOutputRegressor

Table 6
List of mathematical symbols and their meanings.

Symbol	Meaning
d	Differencing order
$G_{t,i}^{s}$	Direct emissions from fuel combustion in year t in future state i for sector s
p	Number of autoregressive terms
q	Number of moving average processes
s	Sector
T	Year of 2050
t	Any given year in the planning horizon
u	Types of uncertain parameters, with values ranging from 1 to 16
$x_0(t)$	Activity parameters and their target years for Base Case emissions calculation
$x_{i,u}^{t}$	Activity parameters in a specific year
$x_{0,u}^{t}$	Differenced time series data at time t
α	Intercept term in ARIMA model
γ	Decay rate
ϵ_t	Error (innovation) term at time t in ARIMA model
θ_k	Moving average coefficients
$\rho_{t,i}^{s}$	Sector-specific emissions factors for GHG
τ	The year the activity parameter hits the target value
ϕ_j	Autoregressive coefficients

wrapper, which trains a separate XGBoost model for each target variable. Subsequently, we partitioned the dataset into training and testing subsets using an 85:15 split and calculated the model performance using multiple complementary metrics, including the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) for both training and test datasets. The performance metrics of the model are summarized in Table 5.

The relevant indicators were analyzed to predict emissions reduction using the SHAP framework, which measures the influence of a feature on the model outcome for each individual data point [58]. Each point in the plot represents the SHAP value for a specific instance, indicating how much that feature pushes the prediction up or down. The color gradient from blue to pink encodes the feature values, with higher values generally having a more significant positive impact on the model's output. The mathematical symbols used throughout this paper are listed in Table 6.

CRediT authorship contribution statement

Peiyao Zhao: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Jimi Oke:** Conceptualization, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Resource availability

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the Lead Contact, Peiyao Zhao (peiyaozhao@umass.edu).

Materials availability

This study did not generate new unique materials.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Peiyao Zhao reports financial support was provided by US Environmental Protection Agency. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the U.S. Environmental Protection Agency (U.S. EPA) Climate Pollution Reduction Grants (CPRG) program.

Data and code availability

All the data and code used in generating the results and figures in this paper are publicly available at https://github.com/narslab/manuscript-newhaven-emissions-forecast.

References

[1] J.J.T. Connolly, The implications of climate gentrification for urban climate action, Nat. Clim. Chang. 15 (4) (2025) 341–344, http://dx.doi.org/10.1038/s41558-025-02265-6.

[2] Cities have a key role to play in tackling climate change – here's why | UNDP climate promise, 2023, https://climatepromise.undp.org/news-and-stories/cities-have-key-role-play-tackling-climate-change-heres-why.

[3] H. Karimipour, V.W.Y. Tam, K.N. Le, H. Burnie, A greenhouse-gas emission reduction toolkit at urban scale, Sustain. Cities Soc. 73 (2021) 103103, http://dx.doi.org/10.1016/j.scs.2021.103103.

[4] A.C. O'Regan, M.M. Nyhan, Towards sustainable and net-zero cities: A review of environmental modelling and monitoring tools for optimizing emissions reduction strategies for improved air quality in urban areas, Environ. Res. 231 (2023) 116242, http://dx.doi.org/10.1016/j.envres.2023.116242.

[5] Chapter 8: urban systems and other settlements, 2025.

[6] M. Luqman, P.J. Rayner, K.R. Gurney, On the impact of urbanisation on CO2 emissions, Npj Urban Sustain. 3 (1) (2023) 6, http://dx.doi.org/10.1038/s42949-023-00084-2.

[7] J. Rogelj, M. den Elzen, N. Höhne, T. Fransen, H. Fekete, H. Winkler, R. Schaeffer, F. Sha, K. Riahi, M. Meinshausen, Paris agreement climate proposals need a boost to keep warming well below 2 °c, Nature 534 (7609) (2016) 631–639, http://dx.doi.org/10.1038/nature18307.

[8] M. Meinshausen, J. Lewis, C. McGlade, J. Gütschow, Z. Nicholls, R. Burdon, L. Cozzi, B. Hackmann, Realization of Paris agreement pledges may limit warming just below 2 °c, Nature 604 (7905) (2022) 304–309, http://dx.doi.org/10.1038/s41586-022-04553-z.

[9] P. Zhao, J. Oke, Typology of US metropolitan areas illuminates per-capita mobile greenhouse gas emissions impacts and informs mitigation strategies, Environ. Res. Lett. 20 (8) (2025) 084075, http://dx.doi.org/10.1088/1748-9326/adf07b.

[10] A. Gambhir, M. George, H. McJeon, N.W. Arnell, D. Bernie, S. Mittal, A.C. Köberle, J. Lowe, J. Rogelj, S. Monteith, Near-term transition and longer-term physical climate risks of greenhouse gas emissions pathways, Nat. Clim. Chang. 12 (1) (2022) 88–96, http://dx.doi.org/10.1038/s41558-021-01236-x.

[11] A. Gambhir, J. Rogelj, G. Luderer, S. Few, T. Napp, Energy system changes in 1.5 °c, well below 2 °c and 2 °c scenarios, Energy Strat. Rev. 23 (2019) 69–80, http://dx.doi.org/10.1016/j.esr.2018.12.006.

[12] P. Achakulwisut, P. Erickson, C. Guivarch, R. Schaeffer, E. Brutschin, S. Pye, Global fossil fuel reduction pathways under different climate mitigation strategies and ambitions, Nat. Commun. 14 (1) (2023) 5425, http://dx.doi.org/10.1038/s41467-023-41105-z.

[13] F.C. Moore, K. Lacasse, K.J. Mach, Y.A. Shin, L.J. Gross, B. Beckage, Determinants of emissions pathways in the coupled climate–social system, Nature 603 (7899) (2022) 103–111, http://dx.doi.org/10.1038/s41586-022-04423-8.
[14] D.G. Groves, R.J. Lempert, A new analytic method for finding policy-relevant scenarios, Glob. Environ. Chang. 17 (1) (2007) 73–85, http://dx.doi.org/10.1016/j.gloenvcha.2006.11.006.
[15] S. Bankes, Exploratory modeling for policy analysis, Oper. Res. 41 (3) (1993) 435–449, http://dx.doi.org/10.1287/opre.41.3.435.
[16] R.J. Lempert, D.G. Groves, S.W. Popper, S.C. Bankes, A general, analytic method for generating robust strategies and narrative scenarios, Manag. Sci. 52 (4) (2006) 514–528, http://dx.doi.org/10.1287/mnsc.1050.0472.
[17] R.J. Lempert, Robust decision making (RDM), in: V.A.W.J. Marchau, W.E. Walker, P.J.T.M. Bloemen, S.W. Popper (Eds.), Decision Making under Deep Uncertainty: from Theory To Practice, Springer International Publishing, Cham, 2019, pp. 23–51, http://dx.doi.org/10.1007/978-3-030-05252-2_2.
[18] M.D. McKay, R.J. Beckman, W.J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics 21 (2) (1979) 239–245, http://dx.doi.org/10.2307/1268522, arXiv:1268522.
[19] M.M. Saad, R.P. Menon, U. Eicker, Supporting decision making for building decarbonization: developing surrogate models for multi-criteria building retrofitting analysis, Energies 16 (16) (2023) 6030, http://dx.doi.org/10.3390/en16166030.
[20] X. Liu, J. Liu, C. Dou, Uncertainty analysis of provincial carbon emission inventories: a comparative assessment of emission factors sources, Sustainability 17 (11) (2025) 4787, http://dx.doi.org/10.3390/su17114787.
[21] D. van der Mensbrugghe, A latin hypercube sampling utility: With an application to an integrated assessment model, J. Glob. Econ. Anal. 8 (1) (2023) http://dx.doi.org/10.21642/JGEA.080102AF.
[22] P. Jouannais, M. Marchand-Lasserre, M. Douziech, P. Pérez-López, When is agrivoltaism environmentally beneficial? a consequential LCA and scenario discovery approach, Environ. Sci. Technol. 59 (23) (2025) 11540–11550, http://dx.doi.org/10.1021/acs.est.5c04479.
[23] D. Álvarez-Antelo, P. López-Muñoz, L. Llases, A. Lauer, Towards a sustainable mobility lifestyle: exploring the flight to rail shift through model-based behavioural change scenarios, Ecol. Econom. 230 (2025) 108498, http://dx.doi.org/10.1016/j.ecolecon.2024.108498.
[24] B.P. Bryant, R.J. Lempert, Thinking inside the box: A participatory, computer-assisted approach to scenario discovery, Technol. Forecast. Soc. Change 77 (1) (2010) 34–49, http://dx.doi.org/10.1016/j.techfore.2009.08.002.
[25] M.D. Gerst, P. Wang, M.E. Borsuk, Discovering plausible energy and economic futures under global change using multidimensional scenario discovery, Thematic Issue on Innovative Approaches to Global Change Modelling, Environ. Model. Softw. 44 (2013) 76–86, http://dx.doi.org/10.1016/j.envsoft.2012.09.001.
[26] M. Browning, J. McFarland, J. Bistline, G. Boyd, M. Muratori, M. Binsted, C. Harris, T. Mai, G. Blanford, J. Edmonds, A.A. Fawcett, O. Kaplan, J. Weyant, Net-zero CO2 by 2050 scenarios for the United States in the energy modeling forum 37 study, Energy Clim. Chang. 4 (2023) 100104, http://dx.doi.org/10.1016/j.egycc.2023.100104.
[27] L. Wang, X. Xue, Z. Zhao, Y. Wang, Z. Zeng, Finding the de-carbonization potentials in the transport sector: Application of scenario analysis with a hybrid prediction model, Environ. Sci. Pollut. Res. 27 (17) (2020) 21762–21776, http://dx.doi.org/10.1007/s11356-020-08627-1.
[28] H. Yuan, M. Ma, N. Zhou, Z. Ma, C. Zhang, Assessing provincial carbon budgets for residential buildings to advance net-zero ambitions, Carbon Neutrality 5 (1) (2026) 5, http://dx.doi.org/10.1007/s43979-025-00159-2.
[29] C. Hamarat, J.H. Kwakkel, E. Pruyt, E.T. Loonen, An exploratory approach for adaptive policymaking by using multi-objective robust optimization, Simul. Model. Pr. Theory 46 (2014) 25–39, http://dx.doi.org/10.1016/j.simpat.2014.02.008.
[30] D. Hadka, J. Herman, P. Reed, K. Keller, An open source framework for many-objective robust decision making, Environ. Model. Softw. 74 (2015) 114–129, http://dx.doi.org/10.1016/j.envsoft.2015.07.014.
[31] J.H. Kwakkel, M. Jaxa-Rozen, Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes, Environ. Model. Softw. 79 (2016) 311–321, http://dx.doi.org/10.1016/j.envsoft.2015.11.020.
[32] J.H. Kwakkel, M. Haasnoot, W.E. Walker, Comparing robust decision-making and dynamic adaptive policy pathways for model-based decision support under deep uncertainty, Environ. Model. Softw. 86 (2016) 168–183, http://dx.doi.org/10.1016/j.envsoft.2016.09.017.
[33] N. Moksnes, J. Rozenberg, O. Broad, C. Taliotis, M. Howells, H. Rogner, Determinants of energy futures—a scenario discovery method applied to cost and carbon emission futures for south American electricity infrastructure, Environ. Res. Commun. 1 (2) (2019) 025001, http://dx.doi.org/10.1088/2515-7620/ab06de.
[34] W. Cao, S. Wang, M. Xu, Optimal scheduling of virtual power plant based on Latin hypercube sampling and improved CLARA clustering algorithm, Processes 10 (11) (2022) 2414, http://dx.doi.org/10.3390/pr10112414.
[35] J. Wessel, G. Iyer, T. Wild, Y. Ou, H. McJeon, J. Lamontagne, Large ensemble exploration of global energy transitions under national emissions pledges, Earth's Futur. 12 (10) (2024) http://dx.doi.org/10.1029/2024EF004754, e2024EF004754.
[36] M.J. Gidden, K. Riahi, S.J. Smith, S. Fujimori, G. Luderer, E. Kriegler, D.P. van Vuuren, M. van den Berg, L. Feng, D. Klein, K. Calvin, J.C. Doelman, S. Frank, O. Fricko, M. Harmsen, T. Hasegawa, P. Havlik, J. Hilaire, R. Hoesly, J. Horing, A. Popp, E. Stehfest, K. Takahashi, Global emissions pathways under different socioeconomic scenarios for use in CMIP6: A dataset of harmonized emissions trajectories through the end of the century, Geosci. Model. Dev. 12 (4) (2019) 1443–1475, http://dx.doi.org/10.5194/gmd-12-1443-2019.
[37] D. Niu, K. Wang, J. Wu, L. Sun, Y. Liang, X. Xu, X. Yang, Can China achieve its 2030 carbon emissions commitment? scenario analysis based on an improved general regression neural network, J. Clean. Prod. 243 (2020) 118558, http://dx.doi.org/10.1016/j.jclepro.2019.118558.
[38] L. Zhu, Z. Li, X. Yang, Y. Zhang, H. Li, Forecast of transportation CO2 emissions in shanghai under multiple scenarios, Sustainability 14 (20) (2022) 13650, http://dx.doi.org/10.3390/su142013650.
[39] P. Zhao, J. Oke, C. Barchers, J. Mega, M.A. Mohammed, M. Arabi, Scenario discovery framework Aids Robust Regional emissions mitigation planning, Environ. Res. Commun. (2026) http://dx.doi.org/10.1088/2515-7620/ae4b94.
[40] M. Ma, S. Zhang, J. Liu, R. Yan, W. Cai, N. Zhou, J. Yan, Building floorspace and stock measurement: A review of global efforts, knowledge gaps, and research priorities, Nexus 2 (3) (2025) http://dx.doi.org/10.1016/j.ynexs.2025.100075.
[41] S. Zhang, M. Ma, N. Zhou, J. Yan, W. Feng, R. Yan, K. You, J. Zhang, J. Ke, Estimation of global building stocks by 2070: unlocking renovation potential, Nexus 1 (3) (2024) 100019, http://dx.doi.org/10.1016/j.ynexs.2024.100019.
[42] Challenges and opportunities in the global net-zero building sector, Cell Rep. Sustain. 1 (8) (2024) 100154, http://dx.doi.org/10.1016/j.crsus.2024.100154.
[43] H. Yuan, M. Ma, N. Zhou, Y. Deng, J. Liu, S. Zhang, Z. Ma, Planning future charging infrastructure for private EVs: A city-scale assessment of demand and capacity, Adv. Appl. Energy 21 (2026) 100258, http://dx.doi.org/10.1016/j.adapen.2025.100258.
[44] Connecticut, Department of Transportation, Traffic monitoring volume and classification information traffic count data, 2024, https://portal.ct.gov/DOT/PP_SysInfo/Traffic-Monitoring.
[45] Policy and Governmental Affairs, Office of Highway Policy Information, Highway statistics series - policy | federal highway administration, 2024, https://www.fhwa.dot.gov/policyinformation/statistics.cfm.
[46] U.S. EPA, OAR, Local greenhouse gas inventory tool, 2017, https://www.epa.gov/statelocalenergy/local-greenhouse-gas-inventory-tool.
[47] Connecticut statewide energy efficiency dashboard, 2024.
[48] EIA, Connecticut State, Connecticut profile, 2024, https://www.eia.gov/state/print.php?sid=CT.
[49] U.S. Census Bureau, eco. DP05: ACS DEMOGRAPHIC AND HOUSING ESTIMATES - Census Bureau Table.
[50] Connecticut, Department of Energy & Environmental Protection, CT Greenhouse Gas Inventory Reports, Tech. Rep., Connecticut, 2024.
[51] Environmental Protection Agency, Facility Level Information on GreenHouse gases Tool, EPA facility level GHG emissions data, 2023, https://ghgdata.epa.gov/ghgp/main.do.
[52] Y. Wang, M. Ma, N. Zhou, Z. Ma, Paving the way to carbon neutrality: evaluating the decarbonization of residential building electrification worldwide, Sustain. Cities Soc. 130 (2025) 106549, http://dx.doi.org/10.1016/j.scs.2025.106549.
[53] R. Yan, N. Zhou, M. Ma, C. Mao, India's residential space cooling transition: decarbonization ambitions since the turn of millennium, Appl. Energy 391 (2025) 125929, http://dx.doi.org/10.1016/j.apenergy.2025.125929.
[54] E. Patel, D.S. Kushwaha, Clustering cloud workloads: k-means vs Gaussian mixture model, Procedia Comput. Sci. 171 (2020) 158–167, http://dx.doi.org/10.1016/j.procs.2020.04.017.
[55] T.I. Khan, N. Sakib, M.M. Hassan, S. Ide, Gaussian mixture model in clustering acoustic emission signals for characterizing osteoarthritic knees, Biomed. Signal Process. Control. 87 (2024) 105510, http://dx.doi.org/10.1016/j.bspc.2023.105510.
[56] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the *EM* algorithm, J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (1) (1977) 1–22, http://dx.doi.org/10.1111/j.2517-6161.1977.tb01600.x.
[57] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge Discovery and Data Mining, in: KDD '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 785–794, http://dx.doi.org/10.1145/2939672.2939785.
[58] S. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, 2017, http://dx.doi.org/10.48550/arXiv.1705.07874, arXiv:1705.07874.