ENVIRONMENTAL RESEARCH
COMMUNICATIONS

INCLUSIVE PUBLISHING TRUSTED SCIENCE

Part of the ENVIRONMENTAL RESEARCH SERIES

PURPOSE-LED PUBLISHING™

LETTER • OPEN ACCESS

Scenario discovery framework aids robust regional emissions mitigation planning

To cite this article: Peiyao Zhao *et al* 2026 *Environ. Res. Commun.* **8** 031008

Environmental Research Communications

OPEN ACCESS

RECEIVED
11 June 2025

REVISED
23 February 2026

ACCEPTED FOR PUBLICATION
27 February 2026

PUBLISHED
17 March 2026

LETTER

Scenario discovery framework aids robust regional emissions mitigation planning

Peiyao Zhao[1,*], **Joseph Mega**[2], **Mohammed Abdalazeem**[3], **Mahsa Arabi**[3], **Camille V Barchers**[2] and **Jimi Oke**[3]

[1] Department of Chemical Engineering, University of Massachusetts Amherst, MA, United States of America
[2] Department of Landscape Architecture and Regional Planning, University of Massachusetts Amherst, MA, United States of America
[3] Department of Civil and Environmental Engineering, University of Massachusetts Amherst, MA, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: peiyaozhao@umass.edu

Keywords: greenhouse gas inventory, scenario discovery, mitigation planning, data-driven

Supplementary material for this article is available online

Abstract

We present a comprehensive accounting and scenario discovery framework for sector-specific regional greenhouse gas emissions. Three regions in Connecticut, namely Bridgeport, Hartford, and New Haven, served as case studies. Our inventory approach combines bottom-up and top-down approaches, relying on readily available data sources. Via scenario discovery, we explore a range of plausible emissions outcomes in a stylized parameter space under reasonably bounded assumptions, which results in robust decarbonization scenarios at three levels (*Mild*, *Moderate*, and *Major*). The data-driven scenario narratives, accompanied by sensitivity analyses, highlight the importance of truck mileage, grid emissions factors, residential fossil fuel use, and commercial emissions for mitigation planning. Our results demonstrate that regional variability and multisectoral approaches are relevant for robust mitigation. Ultimately, we expect that this framework can be readily deployed to other regions and can serve as an exploratory screening tool to identify key areas of focus for equitable and effective emissions mitigation planning.

1. Introduction

Given the growing climate crisis, there is an urgent need for relevant and targeted greenhouse gas (GHG) emissions mitigation planning. To this end, GHG emissions inventories are important for understanding emissions drivers and, consequently, developing effective mitigation strategies. The Intergovernmental Panel on Climate Change (IPCC), launched by the United Nations, provides scientific assessments on climate change and mitigation strategies, emphasizing the need for emissions inventories to track progress and inform policy decisions [1–3]. In response to these efforts, several countries are actively developing comprehensive inventories and mitigation strategies [4, 5]. An important international effort was the Paris Agreement, signed in 2016, which aims to restrict the rise in global average temperature to well below 2 °C above pre-industrial levels [6]. Yet, today, this agreement has been severely challenged as nations grapple with the complex realities and the deep uncertainties of economic and socio-political implications of mitigation [7].

Nationwide [8–10] and statewide (or provincial) [11–14] GHG emissions inventories have been well established. Yet, metropolitan and municipal inventories are arguably more crucial for identifying targeted mitigation strategies [15, 16]. Such strategies can be targeted to address specific community needs while promoting stakeholder and public participation in climate action efforts [17–19]. In a notable study, inventories for 100 US metropolitan areas were developed using a unified accounting framework, shedding light on emissions trends across different urban landscapes for the years 2002 and 2014 [20]. Another large-scale effort inventoried emissions for nearly 300 prefecture-level cities in China, using a 2005 benchmark [21]. Similar efforts in

Europe include inventories in Portugal [22], Sweden and Italy [23], each focusing on a single region. Other efforts identifying primary emissions contributors in a given region include Anji County in China [24], Bangkok [25], San Francisco [26], and the San Luis Valley of Colorado [27].

Traditional scenario planning has historically been used to explore plausible emissions futures by developing narratives based on key driving forces [28, 29]. However, this traditional top-down approach depends largely on fixed, limited and extreme assumptions, which often fail to capture the full complexity of the system. Scenario discovery (SD) was then developed to address these limitations using an approach that enables robust decision-making under deep uncertainty [30–33]. In SD, states representing possible joint outcomes of uncertain parameters are generated using computational experimental design approaches and then evaluated under possible strategies [34]. The resulting ensemble of futures is then mined to obtain relevant scenarios based on specified performance metrics [35, 36]. Ultimately, SD finds data-driven robust clusters of future outcomes that can then be used to define narratives of interest and aid decisinon-making. We note that scenario discovery is distinct from the scenario-based forecasting approach which uses statistical models (such as, multiple regression analysis and autoregressive integrated moving average) to predict future emissions trends based on exogenous variables or external factors such as energy use and climatic variables [37–41]. More specifically, scenario discovery focuses on discovering robust clusters of future outcomes given specified performance targets (which may or may not include forecasting), as it minimizes the chance of overlooking important combinations of factors [42–44]. Meanwhile, traditional scenario-based forecasting emphasizes computing or predicting target variables based on a pre-determined scenario narrative.

SD has been successfully applied for robust decision-making in applications such as water resource planning [31], energy supply [45], and climate adaptation [46–48]. Despite these efforts, no study has yet investigated robust regional mitigation approaches based on emissions inventories under deep uncertainty. Most existing urban emissions inventories are based on deterministic models that assume static activity and fixed emission factors [49]. In contrast, the scenario discovery approach can capture systematic variability and offer insights for robust emissions mitigation planning. There is also a dearth of research focused on identifying key drivers of regional multisector greenhouse gas emissions and discovering climate change mitigation scenarios using scenario discovery.

To address these gaps, we make the following contributions in this paper. Significantly, we integrate an inventory estimation approach with a data-driven scenario discovery to obtain actionable decarbonization scenarios. This integration of detailed inventory estimates under deep uncertainty for robust mitigation scenarios is a first in the literature. Our integrated framework also facilitates mitigation planning at multiple decarbonization levels, as we demonstrate with three targets. Furthermore, we explicitly account for uncertainty in activities and influencing factors across multiple sectors. We demonstrate the outcomes of our framework using three regional case studies in the US state of Connecticut. The variability in the results further highlight the importance of regional-level decision-making for more effective emissions mitigation.

2. Methodology

First, we developed a detailed and replicable mixed accounting approach to estimate regional GHG inventories, combining bottom-up estimation based on activity data, in addition to the top-down method of downscaling state-level emissions using appropriate indicators. Second, we also conducted a quality assessment of the inventory to validate our inventory results. We further developed an interactive dashboard to promote engagement and planning [4]. Third, we applied a scenario discovery approach to obtain data-driven scenarios that shed light on actionable scenarios that lead to robust emissions reduction outcomes. The flowchart of our research approach is shown in figure 1. As an aid to the reader, all acronyms and symbols used in this paper are described in table 1.

2.1. Study areas

Three planning regions in the northeastern US state of Connecticut, namely Bridgeport, Hartford, and New Haven, served as the case study areas. The Bridgeport region comprises the Bridgeport-Stamford-Norwalk metropolitan statistical area (MSA) and the towns of Bridgewater and New Milford. The Hartford region comprises the Hartford-east Hartford-Middletown MSA and the towns of Colchester, Lyme, and Old Lyme. Finally, the New Haven region is the New Haven-Milford MSA. Maps of the planning regions are shown in figure 2. Population, housing, and gross domestic product (GDP) are summarized in Supplementary Information (SI) table 7. (Heat maps of population, forestry and agricultural area distributions are shown in SI figure 2.)

[4] Greenhouse gas emissions interactive dashboard: ctghginventory.narslab.org

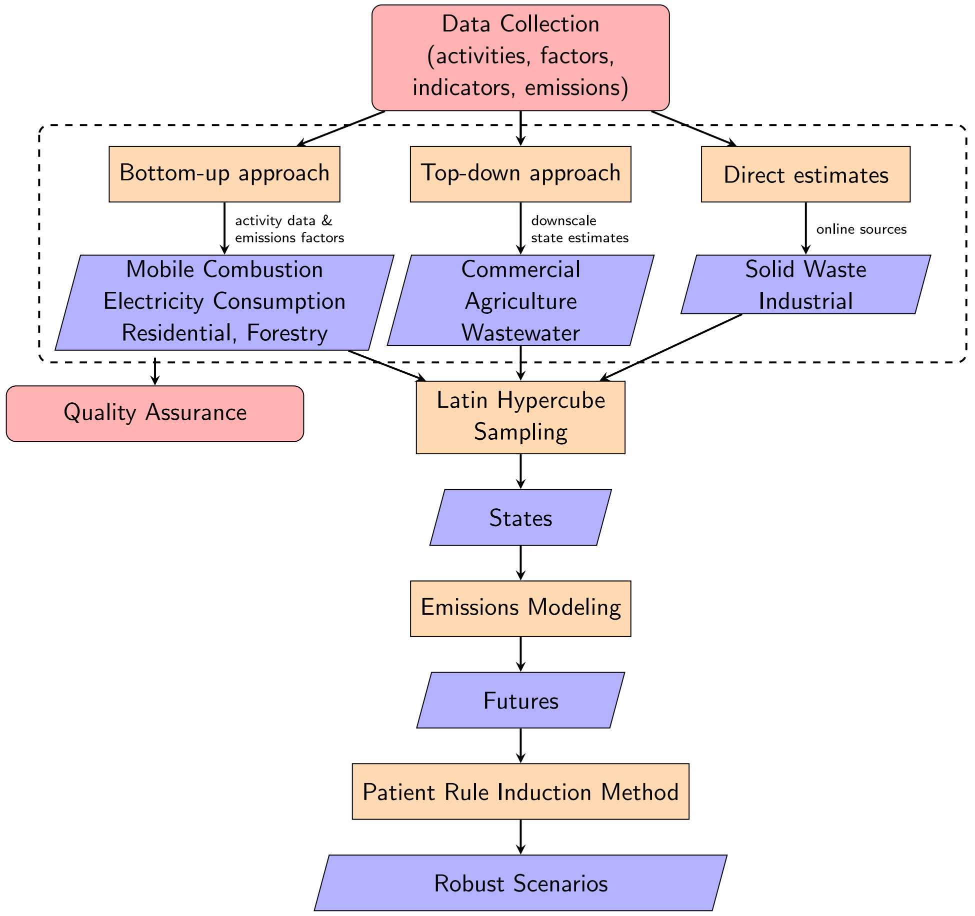

Figure 1. Flowchart describing the regional GHG emissions inventory and scenario discovery framework.

2.2. Data sources

We accounted for emissions from seven sectors: *Mobile Combustion*, *Electricity Consumption*, *Solid Waste*, *Stationary Combustion* (comprising *Residential*, *Commercial*, and *Industrial* subsectors), *Agriculture*, *Wastewater Treatment*, and *Forestry*, which sequesters carbon dioxide [50, 51]. Emissions from *Electricity Consumption* are considered as indirect emissions while the rest of the sectors are direct emissions. Direct emissions are released from activities that occur within the location of attribution, while indirect emissions are those generated in a different location to enable activities in the place of attribution [52, 53].

We gathered relevant activity and indicator data, as well as available emissions estimates, from various sources (summarized in table 2). We obtained daily vehicle miles traveled (VMT) for each of the relevant counties from the Connecticut Department of Transportation (CTDOT) [54], and vehicle-type distributions from the Federal Transit Administration [55]. Vehicle fuel economy data were obtained from the EPA's Local Government Greenhouse Gas Inventory Tool (LGGIT) [56]. County-level electricity consumption data were obtained from EnergizeCT [57]. *Industrial* emissions and *Solid Waste* emissions, including those from solid waste combustion and landfill methane emissions, were directly acquired from the EPA Facility Level Information on GreenHouse gases Tool (FLIGHT) database [58](Further details are provided in SI tables 1 and 2). Data collected for the *Residential* subsector included household fuel consumption distributions from the American Community Surveys (ACS) [60] and statewide fuel consumption from the Energy Information Association (EIA) [59]. Commercial building footprints were obtained via OSMnx, a Python package [61]. Agriculture land area, wastewater facility counts, and forestry data were obtained from the United States Department of Agriculture (USDA) [63], EPA National Pollutant Discharge Elimination System (NPDES) permit dataset [64], and the Connecticut Land Cover database [65], respectively.

We considered 2021 as the reporting (base) year of the inventory, as it had full coverage among all the sectors. (Base year activity data are shown in SI table 6.) Corresponding GHG-specific emissions factors (EFs) for each activity obtained from United States EPA Emissions Hub [66] are summarized in SI table 8.

Table 1. List of mathematical symbols and their meanings, with symbols for emissions estimation and scenario discovery separated by a horizontal line.

Category	Symbol	Description
Sector	$\mathcal{A}$	Agriculture
	$\mathcal{E}$	Electricity consumption
	$\mathcal{F}$	Forestry
	$\mathcal{M}$	Mobile combustion
	$\mathcal{S}$	Stationary combustion
	$\mathcal{S}_r$	Residential subsector
	$\mathcal{S}_c$	Commercial subsector
	$\mathcal{S}_i$	Industrial subsector
	$\mathcal{W}$	Solid waste
	$\mathcal{T}$	Wastewater treatment
Parameter	A	Total area (of building footprint, forestry, fertilizer land)
	a_C	Amount of carbon dioxide sequestrated per unit of carbon (3.67)
	η_v	Fuel economy (miles per gallon) for vehicle type v
	ψ_g	Global warming potential for greenhouse gas g
	N_s	Number of sectors
	r_C	Carbon sequestration factor (2.23 metric tons of carbon per hectare)
	V_v^R	Vehicle miles traveled (VMT) for vehicle type v in region R
	ε_ϕ	Effectiveness of fertilizer ϕ
	π_s^R	Regional proportions of building footprint, fertilizer, and wastewater facilities relative to state totals
	$\eta_{\ell,\phi}$	Percentage of nitrogen lost in fertilizer ϕ due to evaporation
	η_ϕ	Percentage of nitrogen in fertilizer ϕ
	$\rho_{s,g}$	GHG emissions factor for sector s and gas g
Indicators	ϕ	Fertilizer type (manure, organic, synthetic)
	f	Fuel type (natural gas: nat; propane: pro; heating oil: oil)
	$\mathcal{R}$	Regional level
	$\mathcal{S}$	Statewide level
	s	Sector
	v	Vehicle type (passenger car: car; bus: bus; truck: tru; motorcycle: mot)
Variables	$G_{s,g}^R$	GHG emissions g from fuel combustion in sector s in region R
	G_0	Total GHG emissions in baseline state
	G_n	Total GHG emissions in state n (future)
	ΔG_n	Relative change in emissions with respect to the baseline emissions G_0
	$N_{h,f}$	Number of households using a specific fuel type
	$N_{\mathcal{W}}$	Number of wastewater treatment facilities
	Q_s^R	Quantity of given activity in region R for sector s
	Q_s^S	Quantity of given activity at state level for sector s
Scenario Discovery	B_k	Box from patient rule induction method
	L_k^R	A subset of input parameters defining box k
	$\boldsymbol{x}_n \in \mathbb{R}^D$	Vector of uncertain parameters in state n
	z_n^R	Percentage change in emissions relative to the baseline
	ρ_k	Density of box k
	τ_n	Categorical indicators for decarbonization targets
	$\mathcal{I}_\tau^R$	Subset of states in region R within target range τ
	Ω_k^R	Coverage of box k

2.3. Emissions calculation

We estimated regional greenhouse gas emissions using a physical equation combining bottom-up approaches (*Mobile Combustion*, *Electricity Consumption*, and *Residential* subsector), a top-down approach of downscaling statewide estimates (*Commercial* subsector, *Agriculture*, *Forestry* and *Wastewater Treatment* sectors) and direct use of available regional estimates (*Solid Waste* and *Industrial* subsector). These three components are summarized in the subsections below, while further details on sector-specific emissions calculations are provided in appendix Appendix B.

2.3.1. Bottom-up approach

For a given activity (e.g. miles driven by a certain vehicle type, or amount of electricity consumed), the relevant emissions factors (EFs) were multiplied to obtain the greenhouse gas emissions estimate:

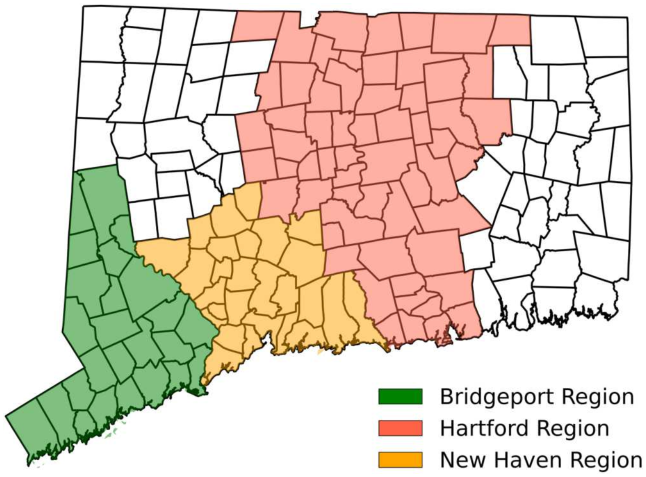

Figure 2. Map of town boundaries and planning region areas in Connecticut.

Table 2. GHG emissions sectors, input data, and sources.

Sector	Inputs	Year	Source
Mobile Combustion	Vehicle miles traveled (VMT)	1966-2022	CTDOT [54]
	Statewide vehicle type distribution	2021	FHWA, DOT [55]
	Vehicle fuel efficiency	2021	LGGIT [56]
Electricity Consumption	Electricity consumption	2016-2022	EnergizeCT [57]
Solid Waste	Landfill emissions	2010-2022	FLIGHT [58]
	Waste combustion emissions	2010-2022	FLIGHT [58]
Stationary Combustion	Household heating fuel consumption	2021	EIA [59]
	Number of households using each heating fuel	2010-2021	ACS [60]
	Commercial building footprint	2021	OSMnx, Python [61]
	Statewide commercial emissions	1990-2021	DEEP [62]
	Industrial facility emissions	2010-2022	FLIGHT [58]
Agriculture	Land area of fertilizer type	2017	USDA [63]
	Nitrogen content and lose	2021	LGGIT [56]
	Statewide agricultural emissions	1990-2021	DEEP [62]
Wastewater Treatment	Number of wastewater treatment facilities	2021	CT NPDES Permits [64]
	Statewide wastewater emissions	1990-2021	DEEP [62]
Forestry	Forest land area	2015	2015 Land Cover [65]
	Carbon sequestration factor	2021	LGGIT [56]

$$G_{s,g}^{R} = \rho_{s,g} Q_{s,g}^{R} \tag{1}$$

where $G_{s,g}^{R}$ are the direct emissions g from fuel combustion in sector s and region R, ρ is the GHG-specific emissions factor—the amount of a given greenhouse gas emitted per unit of a certain activity— and $Q_{s,g}^{R}$ is the quantitative measure of the activity. Three types of GHG are considered in our inventory: carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O). These gases are converted to carbon dioxide equivalents (CO_2e) based on their respective global warming potentials (ψ_g) obtained from IPCC [67] (see SI section 1.2 for detail). We then obtained greenhouse gas emissions G_s^R for sector s and region R (in MMTCO$_2$e) as:

$$G_s^R = \sum_g \psi_g G_{s,g}^{R} \tag{2}$$

2.3.2. Top-down approach

For sectors where low-level activity data were not available, we downscaled statewide emissions using the proportion of a relevant indicator in the region relative to the state (e.g. fertilizer-treated land, number of wastewater treatment facilities) to obtain regional emissions G_s^R as follows:

Table 3. Parameter changes across explored states relative to baseline activity values, showing upper and lower percentage bounds).

Label	Parameter	Baseline activity value				Range	
		Bridgeport	Hartford	New Haven	Units	Lower bound	Upper bound
x_{n1}	Passenger car vehicle miles traveled	2.90	4.25	2.81	billion miles	−1	0.2
x_{n2}	Trucks vehicle miles traveled	4.00	5.86	3.87	billion miles	−1	0.2
x_{n3}	Passenger car fuel efficiency	24.1	24.1	24.1	mpg	0	1.5
x_{n4}	Grid emissions factors	543.85	543.85	543.85	lb/MWh	−1	0.2
x_{n5}	Residential fuel oil consumption	83.10	115.90	74.62	million gallons	−1	0.2
x_{n6}	Residential propane consumption	13.43	19.33	9.98	million gallons	−1	0.2
x_{n7}	Residential natural gas consumption	10447.55	14203.39	9988.50	million cf	−1	0.2
x_{n8}	Commercial Emissions	0.73	1.60	0.96	$MMTCO_2e$	−1	0.2
x_{n9}	Industrial Emissions	0.11	0.35	0.34	$MMTCO_2e$	−1	0.2

$$G_s^R = \pi_s^R G_s^S \tag{3}$$

where π_s^R represents the proportion of an indicator (such as building footprint, agricultural fertilizer amount, or wastewater facility) at the regional level relative to that at the state level, while G_s^S are the statewide sector-specific emissions.

2.3.3. Direct estimates

For the *Solid Waste* sector and *Industrial* subsector, GHG estimates ($G_{\mathcal{W}}^R$ and $G_{\mathcal{S}_i}^R$, respectively) were obtained directly from existing datasets, requiring no further estimation.

2.3.4. Total baseline emissions

The total baseline emissions is given by the physical equation summing up contributions from all sectors. Thus,

$$G_0^R = \sum_g \left[\sum_v \left[\frac{\rho_{\mathcal{M},g,v} V_v^R}{\eta_v^{\mathbb{I}(g=CO_2)}} \right] + \rho_{\mathcal{E},g} Q_{\mathcal{E}}^R + \sum_f \rho_{\mathcal{S}_r,g,f} Q_f^R \right] + G_{\mathcal{S}_c}^R + G_{\mathcal{S}_i}^R + G_{\mathcal{A}}^R + G_{\mathcal{T}}^R + G_{\mathcal{W}}^R + G_{\mathcal{F}}^R \tag{4}$$

In the above equation, $g \in$, $v \in \{car, tru, bus, mot\}$ indicates vehicle type (passenger car, truck, bus, and motorcycle, respectively), $\rho_{\mathcal{M},g,v}$ is the GHG-specific and vehicle-specific mobile combustion emissions factor (EF) and V_v^R the VMT for vehicle type v in region R. $\rho_{\mathcal{E},g}$ is the GHG-specific EF for electricity generation while $Q_{\mathcal{E}}^R$ is the electricity consumed in region R. $G_{\mathcal{S}_c}^R$ and $G_{\mathcal{S}_i}^R$ are the regional GHG emissions for the commercial and industrial subsectors. The last four terms in the equation ($G_{\mathcal{A}}^R$, $G_{\mathcal{T}}^R$, $G_{\mathcal{W}}^R$, $G_{\mathcal{F}}^R$) are the regional GHG emissions for the commercial and industrial subsectors, and for agriculture, solid waste, wastewater and forestry sectors, respectively.

2.4. Scenario discovery

We sampled nearly 2 million *states* (equally likely combinations of uncertain parameters). In each region, we then evaluated all states using the physical equation shown in the previous section, resulting in a collection of *futures*. The performance metric of interest was the percentage change in emissions in that state relative to the baseline. We defined 3 performance target levels (*Mild*, *Moderate* and *Major*) and then implemented a clustering algorithm to find robust scenarios, the analyses of which provided insights into relevant activity parameters to guide mitigation planning efforts.

2.4.1. State generation

We identified nine activity parameters as uncertain (see table 3). These parameters were implemented as factors indicating percentage change in the corresponding activities. We then used the Latin Hypercube sampling method to obtain plausible joint outcomes across these 9 parameter ranges. Each of these outcomes is referred to as a state $\boldsymbol{x}_n \in \mathbb{R}^D$, where $D = 9$. The set of states is given as:

$$\mathcal{S} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \tag{5}$$

Thus, we generated $|\mathcal{S}| = 1{,}953{,}125$ (5^9) states for each of the 3 regions.

2.4.2. Evaluation of futures

We then evaluated each state $\boldsymbol{x}_n$ using the physical emissions calculation described in the previous section, resulting in corresponding emissions futures z_n^R. For each sector, we multiplied the relevant activity parameter by its corresponding uncertainty x_{nd} (in terms of proportion). We define G_n^R as the emissions computed via the physical equation for state n:

$$\begin{aligned} G_n^R = \sum_g \left[\sum_v \left[(1 + x_{n1}^{\mathbb{I}(v=\mathrm{car})} + x_{n2}^{\mathbb{I}(v=\mathrm{tru})} + x_{n3}^{\mathbb{I}(v\in\{\mathrm{bus,mot}\})}) \frac{\rho_{MC,g,v} V_v^R}{[1 + x_{n3}^{\mathbb{I}(v=\mathrm{car})}]\eta_v^{\mathbb{I}(g=\mathrm{CO}_2)}} \right] \right. \\ + (1 + x_{n4})\rho_{\mathcal{E},g} Q_{\mathcal{E}}^R + \sum_f [(1 + x_{n5}^{\mathbb{I}(f=\mathrm{oil})} + x_{n6}^{\mathbb{I}(f=\mathrm{pro})} + x_{n7}^{\mathbb{I}(f=\mathrm{nat})})\rho_{\mathcal{S}_r,g,f} Q_f^R] \\ + (1 + x_{n8}) G_{\mathcal{S}_c}^R + (1 + x_{n9}) G_{\mathcal{S}_i}^R + G_{\mathcal{A}}^R + G_{\mathcal{T}}^R + G_{\mathcal{W}}^R + G_{\mathcal{F}}^R \end{aligned} \quad (6)$$

Thus baseline emissions, G_0^R, are obtained when $x_{nd} = 0\ \forall\, d$. The change in the total emissions for each state n is then defined as

$$\Delta G_n^R = G_n^R - G_0^R \quad (7)$$

We can then write the equation for ΔG_n^R as:

$$\begin{aligned} \Delta G_n^R = \sum_g \left[\sum_v \left[(x_{n1}^{\mathbb{I}(v=\mathrm{car})} + x_{n2}^{\mathbb{I}(v=\mathrm{tru})} + x_{n3}^{\mathbb{I}(v\in\{\mathrm{bus,mot}\})}) \rho_{M,g,v} \frac{V_v^R}{[1 + x_{n3}^{\mathbb{I}(v=\mathrm{car})}]\eta_v^{\mathbb{I}(g=\mathrm{CO}_2)}} \right] \right. \\ + x_{n4}\rho_{\mathcal{E},g} Q_{\mathcal{E}}^R + \sum_f [(x_{n5}^{\mathbb{I}(f=\mathrm{oil})} + x_{n6}^{\mathbb{I}(f=\mathrm{pro})} + x_{n7}^{\mathbb{I}(f=\mathrm{nat})})\rho_{\mathcal{S}_r,g,f} Q_{\mathrm{f}}^R] + G_{\mathcal{S}_c}^R x_{n8} + G_{\mathcal{S}_i}^R x_{n9} \end{aligned} \quad (8)$$

Ultimately, we define each region-specific future outcome z_n^R as the percentage change in the emissions computed in state n relative to that of the baseline. Thus:

$$z_n^R = \%\Delta G_n^R = \frac{\Delta G_n^R}{G_0^R} \times 100\%. \quad (9)$$

We investigated three performance ranges in terms of percentage emissions reduction: $[-30\%, -50\%]$, $[-50\%, -70\%]$, and $[-70\%, -90\%]$, corresponding to *Mild*, *Moderate* and *Major* decarbonization, respectively. Specifically, for each future, we assigned a categorical indicator τ_n according to which threshold range the outcome falls into:

$$\tau_n = \begin{cases} 1, & \mathrm{if} z_n^R \in [-30\%, -50\%], \\ [2pt]2, & \mathrm{if} z_n^R \in [-50\%, -70\%], \\ [2pt]3, & \mathrm{if} z_n^R \in [-70\%, -90\%], \end{cases} \quad (10)$$

For each region, the subset $\mathcal{I}_\tau^R$ identifies the states $\boldsymbol{x}_n$ that result in futures within the specified performance threshold category τ_n:

$$\mathcal{I}_\tau^R = \{\ \boldsymbol{x}_n | \tau_n\ \}, \quad \tau_n \in \{1, 2, 3\} \quad (11)$$

2.4.3. Patient rule induction method

We applied the Patient Rule Induction Method (PRIM), a rule-based statistical algorithm [68], to identify clusters (or boxes) of futures corresponding to the intended target $\mathcal{I}_\tau^R$. For each region, PRIM generates a collection of multidimensional boxes $\mathcal{B}^R = \{B_1^R, B_2^R, \ldots\}$, where each box B_k^R is defined by a range of limiting constraints on a region-specific subset of input parameters $L_k^R \subseteq \{1, \ldots, D\}$, where D is the length of the uncertainty vector:

$$B_k^R = \{\ \boldsymbol{x}_n | a_{nd}^R \leqslant x_{nd} \leqslant b_{nd}^R,\ \ d \in L_k^R\ \}. \quad (12)$$

Boxes are iteratively generated starting from those with the highest coverage (and lowest density) to those with the highest density possible [32, 35, 68]. Coverage Ω_k is defined as the ratio of the total number of outcomes of interest in the box B_k^R to the total number of outcomes of interest across the entire sample space $\mathcal{X}$:

$$\Omega_k^R = \frac{\sum_{\boldsymbol{x}_n \in B_k^R} \mathbb{I}(\boldsymbol{x}_n \in \mathcal{I}_\tau^R)}{\sum_{\boldsymbol{x}_n \in \mathcal{X}} \mathbb{I}(\boldsymbol{x}_n \in \mathcal{I}_\tau^R)}, \quad (13)$$

Successive boxes are then obtained by iteratively reducing the limits L_k in the dimension that maximizes density [68]. In this context, density ρ_k^R is defined as the ratio of the total number of the outcomes of interests in a box B_k^R to the number of cases in that box. Thus,

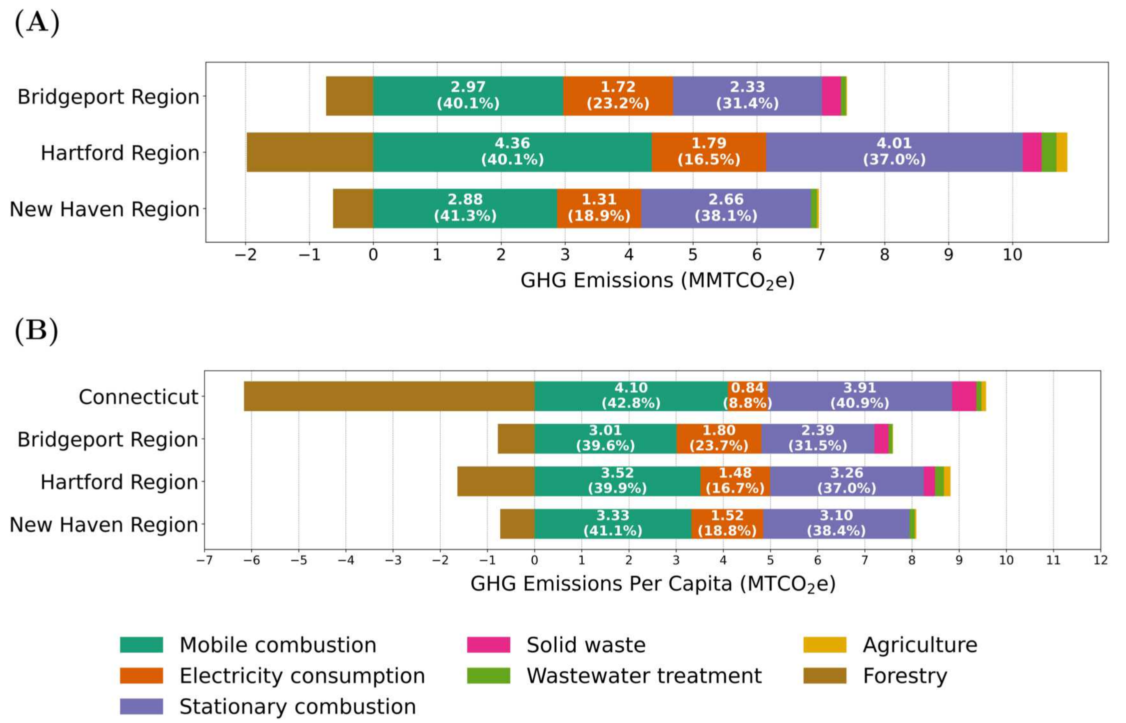

Figure 3. Emissions plots: (A) Total GHG emissions across all sectors, (B) Emissions per capita. (The carbon sequestration is not included in the total carbon emissions calculation; only the net carbon emissions are considered.).

$$\rho_k^R = \frac{\sum_{\boldsymbol{x}_n \in B_k^R} \mathbb{I}(\boldsymbol{x}_n \in \mathcal{I}_\tau^R)}{\sum_{\boldsymbol{x}_n \in B_k^R} 1}. \tag{14}$$

The optimal box is then selected to maximize density, coverage and interpretability (i.e. the ideal box would be restricted by the fewest possible number of uncertain dimensions in order to provide a useful narrative). Such a box then yields a high-interest data-driven scenario narrative [31]—analyses of which can guide future mitigation planning. We also term the discovered scenarios as robust, given that they span futures that balance the success tradeoff between coverage and density.

3. Results

3.1. Baseline greenhouse gas emissions

Percentage emissions contributions by sector are shown along with those of the entire state of Connecticut in figure 3(B). In general, *Mobile Combustion* had the highest percentage contribution relative to the total emissions in the three regions and the state (40%–43%). However, the relative contributions of *Electricity Consumption* at the regional level (23%, 17%, and 19% in Bridgeport, Hartford, and New Haven, respectively) were much higher compared to the state-level contribution of 9% from the same sector. The reverse is observed in *Stationary Combustion*, where the regional contributions are 32%, 37%, and 38% (in Bridgeport, Hartford, and New Haven, respectively), which are lower than the statewide share of 41%. Sequestered carbon dioxide from *Forestry* accounts for approximately −10%, −18%, and −9% of the total emissions in the Bridgeport, Hartford, and New Haven regions, respectively. (A detailed breakdown of forest area by type is illustrated in SI figure 1).

Overall, regional per-capita emissions, accounting for sequestered carbon emissions from *Forestry*, were 6.8, 7.2, and 7.4 $MTCO_2e$ for Bridgeport, Hartford, and New Haven, respectively. In contrast, per-capita emissions for the entire state amounted to 3.4 $MTCO_2e$, indicating greater intensity in these regions. However, we observe lower regional per-capita emissions in the *Mobile Combustion*, *Stationary Combustion* and *Solid Waste* sectors compared to the statewide values. The reverse is the case for *Electricity Consumption*.

Further, we find that fossil fuel combustion in either *Mobile Combustion* or *Stationary Combustion* accounted for the greatest emissions in all regions, as shown in figure 4. Among all the activities, diesel-based trucks contributed the highest emissions (25% of total emissions) across all three regions. After truck emissions, residential buildings have the second highest emissions among all activities. These comprise the combustion of

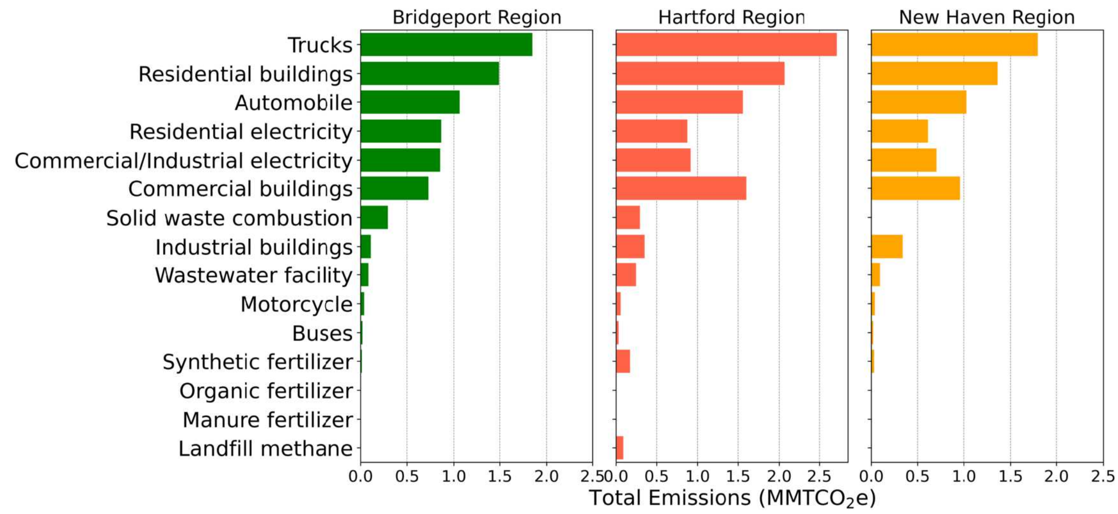

Figure 4. Total emissions in each region broken down by activities.

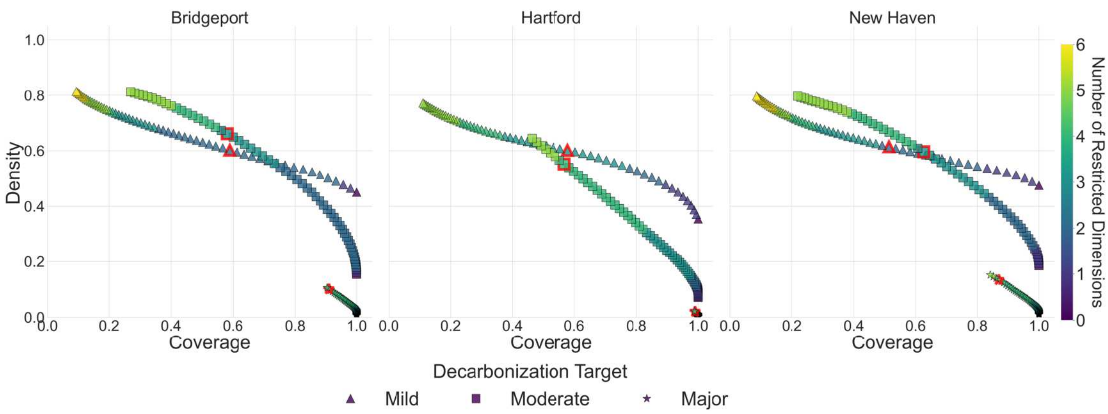

Figure 5. PRIM trade-off curves are shown for three decarbonization scenarios across different regions, with selected scenarios in each region outlined in red.

natural gas, heating oil, and propane for heating and cooking. Among the mobile sources, the next highest was automobile passenger cars, with buses contributing the lowest amount of emissions among other vehicles.

3.2. High-interest data-driven decarbonization scenarios

For each of the three decarbonization targets and in each of the three regions, we selected the optimal robust scenarios from the tradeoff curves generated by PRIM (see figure 5). In the *Mild* and *Moderate* ranges, we identified pareto-optimal scenarios at about 60% density and coverage across all 3 regions. In the *Major* cases, however, it was impossible to achieve a density beyond 20%, which points to the difficulty of achieving this level of decarbonization, as the number of relevant futures is fewer. Attaining these scenarios would thus require more effort and focus. Figure 6 shows the distribution of emissions for each decarbonization scenario across the three regions. Figure 7 illustrates how the features of uncertainty parameters restrict each of the selected scenarios. Narrower ranges indicate greater influence on the constrained scenario space. Four features consistently show as significant: truck VMT, electricity grid emissions factors, commercial emissions, and residential fuel oil consumption.

3.3. Sensitivity of emissions change to the activity parameters

We analyze the sensitivity of the changes in emissions to activity data, as shown in figure 8. The analysis reveals the relative impact of various activity parameters on total greenhouse gas emissions across the three regions. Truck VMT has the largest impact, with 10% change in truck travel distances, leading to 3% change in total emissions across all three regions. The grid emissions factor, indicating electricity generation impact, also plays a major role, which suggests that the shifts in grid decarbonization (e.g.renewable energy adoption) could lead

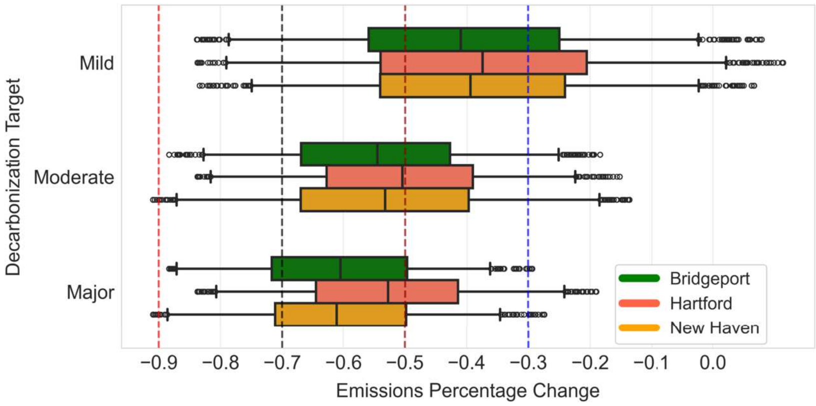

Figure 6. Box plot of emissions change distribution in the selected scenarios.

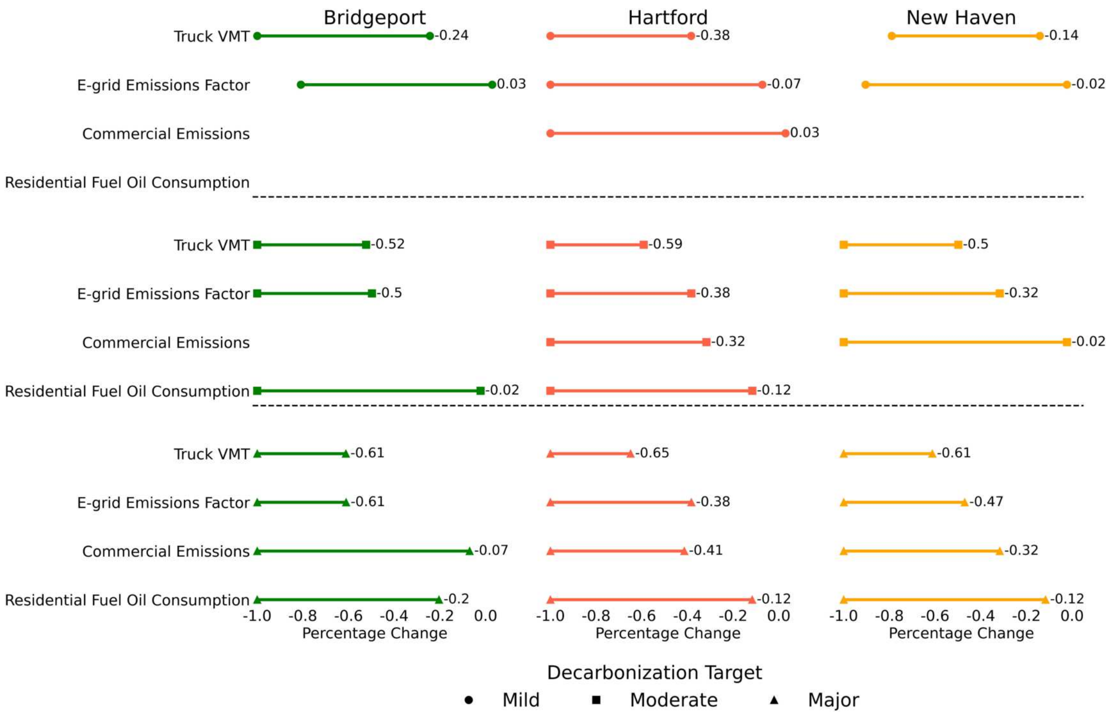

Figure 7. Restricted dimensions of the decarbonization scenarios.

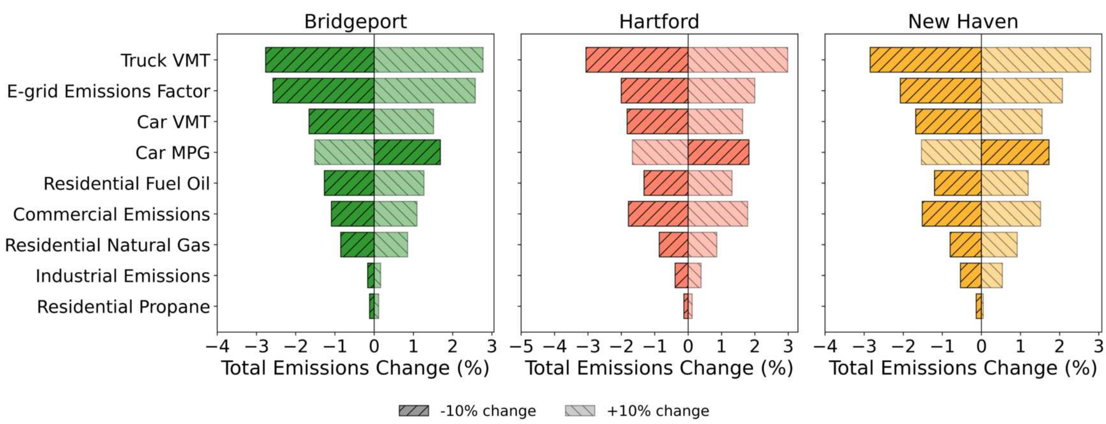

Figure 8. Sensitivity analysis of uncertainty parameter.

to emissions reduction. Other transport-related factors are car VMT and car MPG, which highlight the important contribution from mobile sectors. Meanwhile, reducing commercial emissions plays a more influential role in Hartford and New Haven than Bridgeport region. Industrial emissions and emissions from residential propane usage are less dominant but still relevant.

4. Discussion

4.1. Regional emissions inventory yield patterns that enable mitigation planning

The emissions inventories for the three planning regions shows spatial variations in both total and per-capita GHG emissions. Hartford, as the hub for transportation, the finance center, and the industry-intensive area [69], has the highest total emissions and per-capita emissions among the three, especially in the transportation sector due to high travel demand and a greater number of vehicle miles traveled. These observations highlight the importance of region-specific inventories, as they may yield patterns that enable more effective near-term emissions mitigation planning. By considering contributions at a more granular level (as shown in figure 4), we can also identify relevant activities that can be targeted for emissions reduction. Mitigation strategies should be prioritized in the areas of trucks and automobiles emissions, as they contributes to the highest amount of emissions due to heavy reliance on fossil fuels for freight and personal mobility. Stationary combustion emissions, including residential emissions across all three regions and commercial in the Hartford region, should also be prioritized as key targets for mitigation. Measures such as vehicle and building electrification, fuel efficiency improvements, and adoption of green heating and cooling technologies in buildings, are required.

4.2. Robust scenarios are informed by multi-sectoral approaches

Across all three decarbonization targets examined, we find that activities across multiple sectors are significant in defining scenario narratives. This demonstrates that multi-sectoral approaches may be crucial for mitigation planning. In *Mild* decarbonization scenarios, electricity grid emissions factors and truck VMT are the primary indicators of emissions reduction across all three regions, as shown in figure 7. In contrast, commercial emissions are only significant in the Hartford region. For *Moderate* decarbonization, both truck VMT and the grid emissions factor continue to play an important role across all regions, with Hartford and Bridgeport requiring reductions of more than 50% to meet targets. Additionally, Hartford should focus on commercial emissions and residential fuel consumption for mitigation, while Bridgeport should focus on residential fuel oil consumption. In both Bridgeport and New Haven, reductions of over 50% in truck VMT and grid emissions factor are required. The Commercial subsector also contributes significantly to emissions in New Haven. Further, transitioning to electric and renewable energy sources for facilities and vehicles will be most effective in mitigating emissions in this region.

4.3. Truck mileage and grid emissions factors are most critical to regional emissions

The sensitivity results corroborate our findings from the scenario discovery process regarding the two significant factors—truck VMT and grid emissions factor. In the Hartford and New Haven planning regions, the commercial sector also contributes a significant portion of total emissions. Low-carbon freight technologies such as shifting to battery electric vehicle and hydrogen fuel cell vehicles [70], use of biofuels [71], and optimizing fleet charging in locations with the lowest grid emission factors [72] can lead to benefits in freight sector emissions mitigation. Further, coordination of building electrification with regional grid decarbonization and the high penetration of renewable are essential for grid and building decarbonization [73, 74].

4.4. Limitations of the study

We note that the futures examined in this scenario discovery experiment do not incorporate forecasts. We also note that cross-sector interactions among uncertain parameters are not considered. Thus, while we cannot explain outcomes or provide pathways, we can still examine implications of plausible outcomes and understand the combinations of activities leading to such. Finally, we do not distinguish the various classes of uncertainty in each of the parameters examined, but this is a viable path for future work. Further avenues of research include the incorporation of dynamic considerations, in addition to relevant exogenous variables, structural system transformation, and endogenous sectoral feedback, to render this framework useful for identifying robust mitigation strategies and for longer-term decision-making use cases.

5. Conclusion

We developed a comprehensive framework for estimating regional GHG inventories, treating 2021 as the reference year. We also created a dashboard for cross-region comparisons and conducted quality assurance tests to evaluate data integrity. Via our data-driven scenario discovery framework which explores a stylized partitioned parameter space, we find that truck mileage, grid emissions factors, residential fuel oil consumption, and commercial emissions are important features in emissions mitigation narratives. Furthermore, these robust scenarios narratives demonstrate the importance of multisectoral approaches for mitigation. The results also suggest that major emissions reduction (70%–90%) demands synchronized efforts across transportation, buildings and the grid.

Overall, this study reveals significant regional variability in emissions patterns and dynamics, highlighting the need for region-specific mitigation strategies. Our inventory approach and scenario discovery framework can be readily deployed across other regions in the US, and also globally, given the existence of similar datasets. It can also be adapted for real-time or annual monitoring purposes. Thus, we expect that our framework can serve as a valuable starting point for identifying and investigating equitable mitigation scenarios.

Funding

We acknowledge funding from the South Central Regional Council of Governments (SCRCOG), the Naugutuck Valley Council of Governments (NVCOG) through an Environmental Protection Agency (EPA) Climate Pollution Reduction Grant (CPRG). We also acknowledge partial support from the NSF (Award Number 2325956).

Conflict of interest

The authors declare no conflict of interest.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/narslab/tracking-msa-ghg.

Appendix A. Quality assurance

To ensure the accuracy and reliability of the estimated greenhouse gas (GHG) emissions for each sector, we implemented a comprehensive quality assurance (QA) process. This involved a comprehensive assessment of the discrepancies between local greenhouse gas emissions estimates and those generated through Quality Control (QC) processes [75]. We considered two key metrics: signed bias and variance:

$$\text{Bias} = \frac{1}{N_{\text{sec}}} \sum_{s=1}^{N_{\text{sec}}} \frac{E_s^L - E_s^Q}{\frac{1}{N_{\text{sec}}} \sum_{s=1}^{N_{\text{sec}}} E_s^Q} \times 100\% \tag{A1}$$

$$\mu = \frac{1}{N_{\text{sec}}} \sum_{s=1}^{N_{\text{sec}}} \frac{E_s^Q - E_s}{E_s} \tag{A2}$$

$$\text{Variance} = \frac{1}{N_{\text{sec}}} \sum_{s=1}^{N_{\text{sec}}} \left(\frac{E_s^Q - E_s}{E_s} - \mu \right)^2 \times 100\% \tag{A3}$$

where N_{sec} is the total number of sectors, E_s^L is the local emissions estimate from the inventory, E_s^Q is the QC emissions estimate and μ is the mean percentage error. The resources used for local estimates and QC estimates are provided in SI table 9. Figure A1 shows the summary of signed biases and variances between local estimates and QC estimates, respectively (Further details are provided in SI table 10). The local and QC estimates for different sectors are summarized in table 11 of SI.

In *Mobile Combustion Miles Per Year (MPY)*, biases are relatively low, with Bridgeport showing the largest at −10%. This is due to differences in how VMT is estimated, as local estimates were derived from regional data provided by CTDOT, whereas QC estimates use population-based scaling. *Mobile Combustion Miles Per Gallon (MPG)* shows minimal bias across all areas (less than 1%). *Electricity Consumption* in the commercial and industrial subsectors presents a high bias, particularly in Hartford (−46%), due to population-adjusted scaling

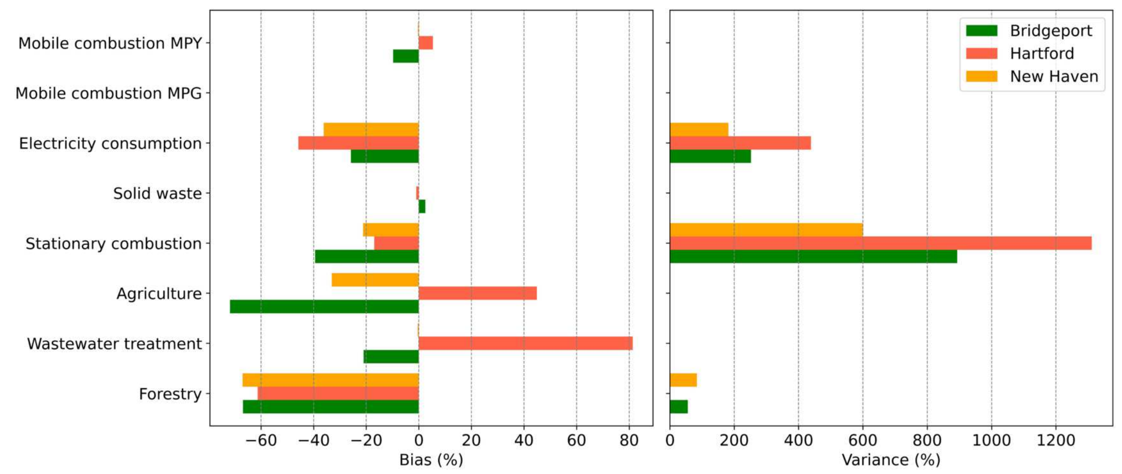

Figure A1. Summary of signed biases and variances.

from county-wide data. Conversely, *Solid Waste* shows minimal discrepancies, as both estimates rely on similar data sources (FLIGHT). Significant biases arise in *Stationary Combustion*, with Bridgeport showing the highest at −39%, due to different data scaling methods. The highest bias is observed in *Agriculture* (−72% for Bridgeport), attributed to differences in how land area and emissions data were scaled. *Wastewater Treatment* shows minimal bias in New Haven but high bias (81%) in Hartford due to differences in scaling methodologies. Finally, *Forestry* shows large discrepancies (up to −67%) due to the contrasting bottom-up and top-down approaches. Across sectors, bias and variance tend to increase when comparing top-down QC estimates to bottom-up local estimates. Furthermore, biases can result from utilizing varied data sources for QC and local estimates, or because downscaling statewide data may not always ensure comparability with county-wide data gathered through a bottom-up approach.

Appendix B. Sector-specific emissions

B.1. Mobile combustion

Mobile Combustion emissions were calculated from estimates of vehicles miles traveled (VMT), vehicle type distributions (see SI table 3 for details), vehicle type-specific fuel economies (see SI table 4) and emissions factors (see SI table 8). The vehicle types include automobile, bus, truck, and motorcycle while the categories in the fuel economy data are passenger car, light truck, heavy-duty vehicle, and motorcycle. To address the naming mismatch in vehicle type categories between the two datasets-vehicle type distribution and vehicle fuel economy, we standardized the groups by assuming that all automobiles correspond to gasoline-powered passenger cars, all trucks correspond to gasoline-powered light trucks, and all buses correspond to diesel-powered heavy-duty vehicles. We also assumed that motorcycles consume gasoline. In the absence of regional-level vehicle type distribution, we assumed the statewide distribution for each region.

The CO_2 EFs are based on gallons of fuel, while those for CH_4 and N_2O are based directly on VMT. Thus, mobile combustion emissions $G^R_{MC,g}$ for greenhouse gas g are given by:

$$G^R_{MC,g} = \begin{cases} \sum_{v\in\mathcal{V}} \rho_{MC,g,v}\frac{V^R_v}{\eta_v}, & \text{if} g = CO_2 \\ \sum_{v\in\mathcal{V}} \rho_{MC,g,v}V^R_v, & \text{if} g \in \{CH_4, N_2O\} \end{cases} \tag{B1}$$

where η_v is fuel economy (miles/gallon) and $\rho_{MC,g,v}$ are the fuel-specific EFs for each greenhouse gas g.

B.2. Electricity consumption

Electricity Consumption emissions, $G^R_{\mathcal{E},g}$, were calculated from kilowatt-hour usage $Q^R_{\mathcal{E}}$ (from residential, industrial, and commercial locations):

$$G^R_{\mathcal{E},g} = \rho_{\mathcal{E},g}Q^R_{\mathcal{E}} \tag{B2}$$

where $\rho_{\mathcal{E},g}$ are grid EFs and $Q^R_{\mathcal{E}}$ is the amount of electricity consumed (MWh) in region R.

B.3. Stationary—Residential

For the *Residential* subsector, we calculated the quantity consumed Q_f^R of each fuel f (natural gas, propane, heating oil) in region R by scaling down the statewide consumption Q_f^S with the ratio of households utilizing a specific fuel type $N_{h,f}^R$ in a given region to the total number of households consuming the same fuel at the state level $N_{h,f}^S$:

$$Q_f^R = \frac{N_{h,f}^R}{N_{h,f}^S} Q_f^S \tag{B3}$$

We then computed the emissions $G_{\mathcal{S}_r g}^R$ as:

$$G_{\mathcal{S}_r g}^R = \sum_{f \in \mathcal{F}} \rho_{\mathcal{S}_r g, f} Q_f^R \tag{B4}$$

where $\rho_{\mathcal{S}_r g, f}$ are the EFs for each heating fuel (kg per gallon).

B.4. Stationary—Commercial

We computed regional *Commercial* emissions $G_{\mathcal{S}_c g}^R$ by scaling down statewide commercial building emissions $G_{\mathcal{S}_c}^S$ based on the proportion of the commercial building footprint in each region $A_{\mathcal{S}_c}^R$ relative to the state total $A_{\mathcal{S}_c}^S$. Thus,

$$G_{\mathcal{S}_c}^R = \frac{A_{\mathcal{S}_c}^R}{A_{\mathcal{S}_c}^S} G_{\mathcal{S}_c}^S \tag{B5}$$

B.5. Agriculture

Only fertilizer-based emissions were considered in *Agriculture* sector [56]. Assuming that agricultural emissions $G_{\mathcal{A}}^R$ in each region are directly proportional to the effective area of land under fertilizer treatment, we estimated $G_{\mathcal{A}}^R$ as:

$$G_{\mathcal{A}}^R = \frac{\sum_\phi \varepsilon_\phi A_\phi^R}{\sum_\phi \varepsilon_\phi A_\phi^S} G_{\mathcal{A}}^S \tag{B6}$$

where ε_ϕ is the effectiveness of fertilizer $\phi \in$ {manure, organic, synthetic}, A_ϕ^R is the area of land treated by fertilizer ϕ in region R, A_ϕ^S is the statewide land area treated by fertilizer ϕ and $G_{\mathcal{A}}^S$ are the statewide agricultural emissions. The fertilizer effectiveness ε_ϕ is the proportion of nitrogen loss in one type of fertilizer relative to that across all types [56]:

$$\varepsilon_\phi = \frac{\eta_\phi \eta_{\ell,\phi}}{\sum_\phi \eta_\phi \eta_{\ell,\phi}} \tag{B7}$$

where η_ϕ is the percentage of nitrogen and $\eta_{\ell,\phi}$ is the percentage of nitrogen lost due to volatilization in fertilizer ϕ (see SI table 5 for details on nitrogen content and loss in those three different types of fertilizers).

B.6. Wastewater treatment

Wastewater Treatment emissions $G_{\mathcal{T}}^R$ were calculated by downscaling statewide wastewater emissions based on the proportion of facilities in each region relative to the state total. Thus:

$$G_{\mathcal{T}}^R = \frac{N_{\mathcal{T}}^R}{N_{\mathcal{T}}^S} G_{\mathcal{T}}^S \tag{B8}$$

where $N_{\mathcal{T}}^R$ and $N_{\mathcal{T}}^S$ are the numbers of wastewater treatment facilities in region R and across the state S, respectively.

B.7. Forestry

We estimated sequestered emissions $G_{\mathcal{F}}^R$, from *Forestry* as follows:

$$G_{\mathcal{F}}^R = -r_c A_\phi^R a_C \tag{B9}$$

where r_c is the carbon sequestration factor, A_ϕ^R is the forested area (hectare) at the regional level and a_C is the amount of carbon dioxide sequestered per unit of carbon.

Appendix C. Interactive dashboard development

We developed an interactive dashboard[5] to visualize and analyze GHG emissions data for the case study areas [76]. The dashboard serves as a comprehensive resource for stakeholders to explore emissions across the 7 sectors considered. The dashboard uses Apache Superset, which allows for cross-filtering and efficient data retrieval [77].

C.1. Dashboard provides insights that support policy-making decisions.

Our inventory dashboard can serve as a tool providing local governments and organizations with actionable data to inform GHG emissions reduction efforts. Figure C1 is an example view of the dashboard, which presents an overview of GHG emissions across different sectors and regions, allowing users to compare emissions per capita and understand the distribution of emissions by sector. The dashboard can also support social science research related to science communication, including public engagement, surveys, information dissemination, and behavioral responses to scientific information. Ultimately, the dashboard framework, as it is accessible and user-friendly, can help inform the public about climate change and mitigation strategies.

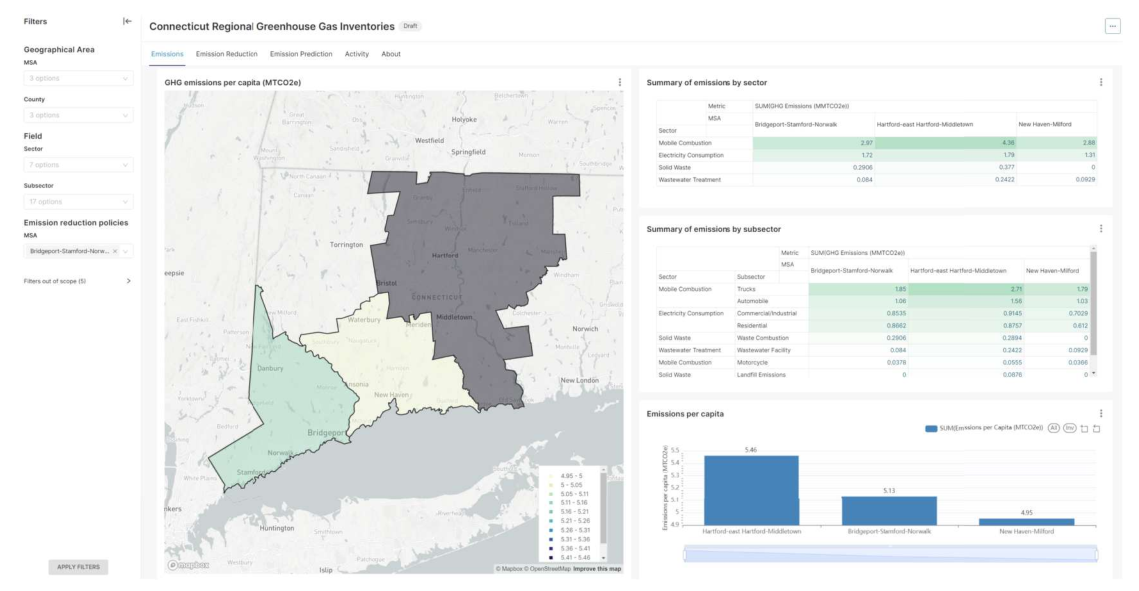

Figure C1. Screenshots of the interactive dashboard.

Appendix D. Emissions percentage change

The histogram of the emissions reduction outcomes in all futures is shown in figure D1. Summary statistics of the histograms are provided in table D1.

[5] https://ctghginventory.narslab.org/

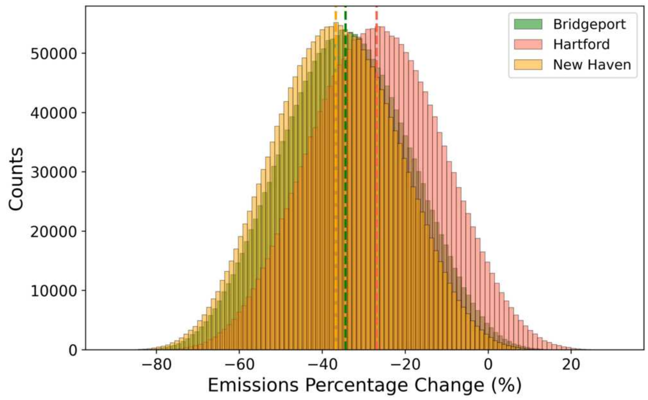

Figure D1. Histogram of emissions percentage change in three MSAs, based on activity data generated using Latin Hypercube sampling.

Table D1. Summary statistics of the histogram of emissions percentage change for Bridgeport, Hartford, and New Haven.

Region	Mean	Median	Min	Max	Std Dev
Bridgeport	−34.41	−34.44	−88.27	20.89	15.13
Hartford	−26.93	−26.95	−83.70	31.44	15.64
New Haven	−36.74	−36.76	−90.93	18.62	14.79

Author contributions

Peiyao Zhao 0000-0001-7866-6249
Conceptualization (equal), Data curation (lead), Formal analysis (lead), Investigation (lead), Methodology (lead), Software (lead), Visualization (lead), Writing – original draft (lead), Writing – review & editing (lead)

Joseph Mega
Writing – original draft (supporting)

Mohammed Abdalazeem 0000-0003-4660-8030
Software (supporting), Methodology (supporting)

Mahsa Arabi 0000-0003-1784-0906
Methodology (supporting), Writing – review & editing (supporting)

Camille V Barchers 0000-0002-7653-2949
Writing – review & editing (supporting), Conceptualization (equal), Supervision (supporting)

Jimi Oke 0000-0001-6610-445X
Conceptualization (equal), Formal analysis (supporting), Funding acquisition (equal), Methodology (supporting), Project administration (lead), Supervision (lead), Writing – review & editing (equal)

References

[1] Pirani A *et al* 2024 Scenarios in IPCC assessments: lessons from AR6 and opportunities for AR7 *npj Climate Action* **3** 1–7
[2] Zhang L *et al* 2024 Global nitrous oxide emissions from livestock manure during 1890–2020: an IPCC tier 2 inventory *Global Change Biol.* **30** e17303
[3] Calvin K *et al* 2023 *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H Lee and J Romero (Eds.)]* (Intergovernmental Panel on Climate Change (IPCC)) pp 35–115
[4] Umemiya C and White M K 2024 National GHG inventory capacity in developing countries—a global assessment of progress *Climate Policy* **24** 164–76
[5] Bun R *et al* 2024 Tracking unaccounted greenhouse gas emissions due to the war in Ukraine since 2022 *Sci. Total Environ.* **914** 169879
[6] Meinshausen M *et al* 2022 Realization of Paris agreement pledges may limit warming just below 2 °C *Nature* **604** 304–9

[7] Forner C and Waskow D 2023 *Beyond the Paris Agreement: The Next Phase of International Cooperation on Climate Action* (World Resources Institute) (https://www.wri.org/technical-perspectives/beyond-paris-agreement-next-phase-international-cooperation-climate-action)
[8] Hockstad L and Hanel L 2018 *Inventory of U.S. Greenhouse Gas Emissions and Sinks. Environmental System Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) (United States)* (cdiac:EPA-EMISSIONS) (https://doi.org/10.15485/1464240)
[9] Kato A, Gurney K R, Roest G S and Dass P 2023 Exploring differences in FFCO2 emissions in the United States: comparison of the vulcan data product and the EPA national GHG inventory *Environ. Res. Lett.* **18** 124043
[10] Seymore R, Inglesi-Lotz R and Blignaut J 2014 A greenhouse gas emissions inventory for South Africa: a comparative analysis *Renew. Sustain. Energy Rev.* **34** 371–9
[11] Young B, Birney C and Ingwersen W W 2024 Dataset of 2012–2020 US national- and state-level greenhouse gas emissions by sector *Data in Brief.* **53** 110173
[12] Rose A, Neff R, Yarnal B and Greenberg H 2005 A greenhouse gas emissions inventory for pennsylvania *J. Air Waste Manage. Assoc.* **55** 1122–33
[13] He G, Jiang W, Gao W and Lu C 2024 Unveiling the spatial-temporal characteristics and driving factors of greenhouse gases and atmospheric pollutants emissions of energy consumption in Shandong province, China *Sustainability* **16** 1304
[14] Wei Y, Zhang X, Xu M and Chang Y 2023 Greenhouse gas emissions of meat products in china: a provincial-level quantification *Resour. Conserv. Recycl.* **190** 106843
[15] Ramaswami A, Hillman T, Janson B, Reiner M and Thomas G 2008 A demand-centered, hybrid life-cycle methodology for city-scale greenhouse gas inventories *Environmental Science & Technology* **42** 6455–61
[16] Ahn D Y, Goldberg D L, Coombes T, Kleiman G and Anenberg S C C O 2023 2 emissions from C40 cities: citywide emission inventories and comparisons with global gridded emission datasets *Environ. Res. Lett.* **18** 034032
[17] Blackhurst M, Matthews H S, Sharrard A L, Hendrickson C T and Azevedo I L 2011 Preparing US community greenhouse gas inventories for climate action plans *Environ. Res. Lett.* **6** 034003
[18] Markolf S A, Matthews H S, Azevedo I M L and Hendrickson C 2018 The implications of scope and boundary choice on the establishment and success of metropolitan greenhouse gas reduction targets in the United States *Environ. Res. Lett.* **13** 124015
[19] Zhao P and Oke J 2025 Typology of US metropolitan areas illuminates per-capita mobile greenhouse gas emissions impacts and informs mitigation strategies *Environ. Res. Lett.* **20** 084075
[20] Markolf S A, Matthews H S, Azevedo I L and Hendrickson C 2017 An integrated approach for estimating greenhouse gas emissions from 100 US metropolitan areas *Environ. Res. Lett.* **12** 024003
[21] Cai B *et al* 2019 A benchmark city-level carbon dioxide emission inventory for China in 2005 *Appl. Energy* **233** 659–73
[22] Gomes J, Nascimento J and Rodrigues H 2008 Estimating local greenhouse gas emissions—a case study on a portuguese municipality *Int. J. Greenhouse Gas Control* **2** 130–5
[23] Marchi M, Pulselli F M, Mangiavacchi S, Menghetti F, Marchettini N and Bastianoni S 2017 The greenhouse gas inventory as a tool for planning integrated waste management systems: a case study in central Italy *J. Clean. Prod.* **142** 351–9
[24] Cai Z, Hu L, Chen D, Zhang Y and Fang X 2024 Structural characteristics and drivers of greenhouse gas emissions at county-level and long-time scales: a case study of the Anji County, China *J. Environ. Sci.* **140** 319–30
[25] Sugsaisakon S and Kittipongvises S 2024 *Impacts of the Nationally Determined Contribution to the 1.5 °C Climate Goal and Net-Zero Target on Citywide Greenhouse Gas Emissions: A Case Study on Bangkok, Thailand* (Environment, Development and Sustainability) (https://doi.org/10.1007/s10668-024-04579-5)
[26] Jones C M and Kammen D M 2015 *A Consumption-Based Greenhouse Gas Inventory of San Francisco Bay Area Neighborhoods, Cities and Counties: Prioritizing Climate Action for Different Locations* (UC Berkeley) (https://escholarship.org/uc/item/2sn7m83z)
[27] Dubinsky J and Karunanithi A T 2017 Greenhouse gas accounting of rural agrarian regions: the case of San Luis Valley *ACS Sustainable Chemistry & Engineering* **5** 261–8
[28] Finn D and Miller N 2022 *Scenario Planning using Climate Data: New Tools Merging Science and Practice* (Lincoln Institute of Land Policy) (https://books.google.com/books/about/Scenario_Planning_Using_Climate_Data_New.html?id=qAo30AEACAAJ)
[29] Avin U 2007 *Using Scenarios to Make Urban Plans* (Lincoln Institute of Land Policy) (https://www.tandfonline.com/doi/full/10.1080/01944360903167653)
[30] Lempert R J, Groves D G, Popper S W and Bankes S C 2006 A general, analytic method for generating robust strategies and narrative scenarios *Manage. Sci.* **52** 514–28
[31] Groves D G and Lempert R J 2007 A new analytic method for finding policy-relevant scenarios *Glob. Environ. Change* **17** 73–85
[32] Bryant B P and Lempert R J 2010 Thinking inside the box: a participatory, computer-assisted approach to scenario discovery *Technol. Forecast. Soc. Change* **77** 34–49
[33] Lempert R J 2019 Robust decision making (RDM) *Decision Making under Deep Uncertainty: From Theory to Practice* ed V A W J Marchau, W E Walker, P J T M Bloemen and S W Popper (Springer International Publishing) pp 23–51
[34] McKay M D, Beckman R J and Conover W J 1979 A comparison of three methods for selecting values of input variables in the analysis of output from a computer code *Technometrics* **21** 239–45
[35] Kwakkel J H and Cunningham S C 2016 Improving scenario discovery by bagging random boxes *Technol. Forecast. Soc. Change* **111** 124–34
[36] Lempert R, Bryant B and Bankes S March 2008 *Comparing Algorithms for Scenario Discovery WR-557-NSF* RAND Corporation, Santa Monica, CA 1–34 (https://www.rand.org/pubs/working_papers/WR557.html)
[37] Ozdemir M, Pehlivan S and Melikoglu M 2024 Estimation of greenhouse gas emissions using linear and logarithmic models: a scenario-based approach for Turkiye's 2030 vision *Energy Nexus* **13** 100264
[38] Zhao R, Pandyaswargo A H and Onoda H 2024 A bottom-up approach for greenhouse gas emission estimation at the community level: a case study in Japan *Energy* **307** 132530
[39] Abolghasemzadeh H, Zekri E and Nasseri M 2024 Regional-scale energy-water nexus framework to assess the GHG emissions under climate change and development scenarios via system dynamics approach *Sustainable Cities and Society* **111** 105565
[40] McCollum D and Yang C 2009 Achieving deep reductions in us transport greenhouse gas emissions: scenario analysis and policy implications *Energy Policy* **37** 5580–96
[41] Espinosa Valderrama M, Cadena Monroy ÁI and Behrentz E 2019 Challenges in greenhouse gas mitigation in developing countries: a case study of the colombian transport sector *Energy Policy* **124** 111–22
[42] Guivarch C, Lempert R and Trutnevyte E 2017 Scenario techniques for energy and environmental research: an overview of recent developments to broaden the capacity to deal with complexity and uncertainty *Environ. Modelling Softw.* **97** 201–10
[43] Kwakkel J H 2019 A generalized many-objective optimization approach for scenario discovery *Futures & Foresight Science* **1** e8

[44] Gerst M D, Wang P and Borsuk M E 2013 Discovering plausible energy and economic futures under global change using multidimensional scenario discovery *Environ. Modelling Softw.* **44** 76–86

[45] Moksnes N, Rozenberg J, Broad O, Taliotis C, Howells M and Rogner H 2019 Determinants of energy futures—a scenario discovery method applied to cost and carbon emission futures for south american electricity infrastructure *Environ. Res. Commun.* **1** 025001

[46] Shortridge J and Guikema S 2016 Scenario discovery with multiple criteria: an evaluation of the robust decision-making framework for climate change adaptation - shortridge - 2016 - risk analysis - wiley online library *Risk Analysis* **36** 2298–312

[47] Greeven S, Kraan O, Chappin ÉJL and Kwakkel J H 2016 The emergence of climate change mitigation action by society: an agent-based scenario discovery study *Journal of Artificial Societies and Social Simulation* **19** 9

[48] Zhou X *et al* 2023 Towards the carbon neutrality of sludge treatment and disposal in China: a nationwide analysis based on life cycle assessment and scenario discovery *Environ. Int.* **174** 107927

[49] Blackhurst M and Matthews H S 2022 Comparing sources of uncertainty in community greenhouse gas estimation techniques *Environ. Res. Lett.* **17** 053002

[50] United States EPA 2022 Learn More about Official State Greenhouse Gas Inventories [Data and Tools] (https://www.epa.gov/ghgemissions/learn-more-about-official-state-greenhouse-gas-inventories)

[51] US EPA OAR 2017 Inventory of U.S. Greenhouse Gas Emissions and Sinks [Reports and Assessments] (https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks)

[52] Li Y *et al* 2024 A review on carbon emission accounting approaches for the electricity power industry *Appl. Energy* **359** 122681

[53] Kongboon R, Gheewala S H and Sampattagul S 2022 Greenhouse gas emissions inventory data acquisition and analytics for low carbon cities *J. Clean. Prod.* **343** 130711

[54] Connecticut DoT 2024 Traffic Monitoring Volume and Classification Information Traffic Count Data (https://portal.ct.gov/DOT/PP_SysInfo/Traffic-Monitoring)

[55] Policy and Governmental Affairs OoHPI 2024 Highway statistics series - policy | federal highway administration (https://www.fhwa.dot.gov/policyinformation/statistics.cfm)

[56] US EPA OAR 2017 Local greenhouse gas inventory tool [data and tools] (https://www.epa.gov/statelocalenergy/local-greenhouse-gas-inventory-tool)

[57] Connecticut Statewide Energy Efficiency Dashboard 2024 (https://www.ctenergydashboard.com/Login.aspx)

[58] Environmental Protection Agency FLIoGgT 2023 EPA Facility Level GHG Emissions Data (https://ghgdata.epa.gov/ghgp/main.do)

[59] EIA CS 2024 Connecticut Profile (https://www.eia.gov/state/print.php?sid=CT)

[60] United States Census Bureau 2024 ACS Demographic and housing estimates - Census Bureau Table; eco (https://data.census.gov/table/ACSDP5Y2021.DP05?g=040XX00US09_310XX00US14860,25540,35300&d=ACS%205-Year%20Estimates%20Data%20Profiles)

[61] Boeing G 2017 OSMnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks *Comput. Environ. Urban Syst.* **65** 126–39

[62] Connecticut DoEEP *et al* 2024 CT greenhouse gas inventory reports. connecticutCT Greenhouse Gas Inventory Reports (https://portal.ct.gov/DEEP/Climate-Change/CT-Greenhouse-Gas-Inventory-Reports)

[63] United States Department of Agriculture NASS 2024 QuickStats Ad-hoc Query Tool (https://quickstats.nass.usda.gov/results/C02AC8A9-A90D-3D2A-ABA6-98BC8A87838AUSDA/NASS)

[64] US EPA R 2016 Connecticut NPDES Permits [Collections and Lists] (https://www.epa.gov/npdes-permits/connecticut-npdes-permits)

[65] Bonsack K *et al* 2022 CT Land Cover Numbers and Charts | Center for Land Use Education and Research (https://clear.uconn.edu/projects/landscape/ct-stats/)

[66] US EPA OAR 2015 (https://www.epa.gov/climateleadership/ghg-emission-factors-hub) GHG Emission Factors Hub [Overviews and Factsheets]

[67] Eggleston H S, Buendia L, Miwa K, Ngara T and Tanabe K 2006 2006IPCC Guidelines for National Greenhouse Gas Inventories (https://www.osti.gov/etdeweb/biblio/20880391)

[68] Friedman J H and Fisher N I 1999 Bump hunting in high-dimensional data *Stat. Comput.* **9** 123–43

[69] Hartford, Connecticut. Wikipedia 2025 2025-11-11T23:08:29Z (https://en.wikipedia.org/w/index.php?title=Hartford,_Connecticut&oldid=1321668744)

[70] Aryanpur V and Rogan F 2024 Decarbonising road freight transport: the role of zero-emission trucks and intangible costs *Sci. Rep.* **14** 2113

[71] Cabrera-Jiménez R, Mateo J M, Jiménez L and Pozo C 2025 Prospective life-cycle assessment of sustainable alternatives for road freight transport *Renew. Sustain. Energy Rev.* **211** 115243

[72] Lu J, Shan R, Kittner N, Hu W and Zhang N 2023 Emission reductions from heavy-duty freight electrification aided by smart fleet management *Transportation Research Part D: Transport and Environment* **121** 103846

[73] Zhong Y, Li R and Cai W 2024 Coordinating building electrification with regional grid decarbonization: comparison of carbon emissions of different scheduling strategies *Energy* **303** 131942

[74] Arbabzadeh M, Sioshansi R, Johnson J X and Keoleian G A 2019 The role of energy storage in deep decarbonization of electricity production *Nat. Commun.* **10** 3413

[75] US EPA OW 2022 Quality Assurance and Quality Control [Collections and Lists] (https://www.epa.gov/choose-fish-and-shellfish-wisely/quality-assurance-and-quality-control)

[76] Zhao P 2024 Connecticut Regional Greenhouse Gas Inventories (https://ctghginventory.narslab.org/superset/dashboard/da18b1c8-cf34-4d45-8d58-0e2a43fa9289/?permalink_key=619Q7ywZaJ8&standalone=true)

[77] Welcome | Superset 2024 (https://Superset.apache.org/)