Authors
Hernandez-Cortes, Danae

Meng, Kyle

Weber, Paige

Publication Date
2026-05-21

Energy Institute WP 363

The Environmental Costs and Geography of U.S. Data Center Expansion

Danae Hernandez-Cortes, Kyle Meng, and Paige Weber

May 2026

The Environmental Costs and Geography of U.S. Data Center Expansion *

Danae Hernandez-Cortes, Kyle Meng, & Paige Weber

May 21, 2026

Abstract

Data centers powering artificial intelligence are growing rapidly across the United States, raising concerns among policymakers and local communities about their environmental and social costs. These costs depend on where data centers locate, and strategic siting has been proposed as a way to limit them. Using a comprehensive facility-level dataset of past, operational, and announced U.S. data centers, we characterize how the environmental and community characteristics of data center locations evolve over 2010–2030. We quantify how much of the projected growth in environmental damages can be attributed to changes in data center locations versus growth in power requirements. Decomposing projected carbon emissions and monetized local air pollution damages into scale and location-driven composition effects, we find that over 2010–2025, composition changes reduced carbon and local air pollution damages by 4 and 12% respectively. Over 2024–2030, scale accounts for approximately 97% of projected pollution growth, with composition changes contributing around 3%. We further find that data centers consistently locate in less densely populated areas, with planned facilities entering census tracts roughly five times less dense than tracts without data centers. Contrary to patterns documented for other disamenities, data centers are not systematically located in lower-income, higher-poverty, or higher non-White-share communities. Our results imply that policies nudging the location of data centers would do little to mitigate their aggregate environmental footprint, which is governed instead by the scale of buildout, and that permitting and community negotiations over siting will increasingly concentrate in the nation's least population dense communities.

*School of Sustainability and School for the Future of Innovation in Society, Arizona State University, Danae.Hernandez-Cortes@asu.edu; Meng: Bren School and Dept. of Economics, UC Santa Barbara, kmeng@bren.ucsb.edu; Weber: Energy and Resources Group, UC Berkeley, paigeweber@berkeley.edu. The authors have no financial or other conflicts of interest to disclose. This material is based upon work supported by Alfred P. Sloan Foundation under Grant No. G-2024-22430.

1 Introduction

The rapid expansion of artificial intelligence is driving an unprecedented buildout of data center infrastructure in the United States. This growth has prompted concern from policymakers, utility regulators, and local communities about the social costs of these facilities. The social and environmental impact of a data center depends on where it is located. And the ability of a data center to locate in a particular location depends on its ability to acquire local permits and an electricity purchase agreement. This paper uses a comprehensive data set of past, new, and announced U.S. data centers to study how environmental and social characteristics of data center locations have, and are planned to, evolve over 2010–2030. In doing so, we highlight that the key determinant of data centers' future national environmental costs is the scale of the buildout and not the spatial distribution of where centers are located. We also forecast which types of communities are likely to be at the center of future debates about data center siting.

This paper contributes to a growing literature quantifying the social and environmental footprint of data centers and AI from different disciplinary perspectives. Engineering-based work by Shehabi et al. (2016, 2024) estimate current energy using server purchases as a proxy for compute demand and projects future energy use under assumed changes in compute efficiency. Other work translates energy use estimates into monetized environmental and social impacts: Muller (2026) combines facility-level data on operational U.S. data centers with an integrated assessment model of the carbon and local pollution costs of electricity consumption, estimating that data centers generated approximately $25 billion in gross external damages in 2025, with Texas and Virginia alone accounting for 30% of the national total. Alvarez et al. (2026) estimate the local effects of data center entry on employment, establishments, house prices, and electricity prices, finding that data centers raise local electricity prices and generate measurable economic activity. Research has also demonstrated that forecasting the energy demand is complicated by the fact that AI energy demands vary significantly across use cases, for example, between text responses and image generation (Luccioni, Jernite and Strubell, 2024).

We contribute to this literature along several dimensions. First, we complement the literature that projects future energy consumption from data centers using model-based approaches, with a projection based solely on announcements from data center providers themselves, which provides a helpful outer bound estimate of data center capacity in the United States. The benefit of this approach is that it allows us to analyze location trends as indicated from the providers themselves. A drawback of this approach is that not all data centers announced will get built. While implied energy use from data centers in the data set aligns well with other estimates for 2023, the implied energy use is substantially higher than other estimates for 2030. This paper focuses on understanding trends in the spatial distribution and environmental footprints of data centers. We view the locational distribution of announced and planned data centers as informative for the locational distribution of those that are eventually built. Further, announced data centers will only get built if they acquire local permits and electricity purchase agreements needed to get online. Matching the locations of planned data center locations to socioeconomic and demographic characteristics tells us the types of communities where these permitting proceedings, debates, and negotiations will take place.

Existing work estimates the level of environmental damages from currently operating facilities (Muller, 2026). We use our newly available facility-level data covering current and announced data centers planned in the future to analyze trends in environmental costs and siting over time. We then extend our analysis of environmental costs with a formal scale-versus-composition decomposition of projected changes in carbon emissions and monetized local air pollution damages, using marginal emissions factors from Holland et al. (2024). This decomposition isolates the share of projected pollution growth attributable to growth in total data center energy consumption (scale effect) from the share attributable to shifts in the geographic distribution of data centers across locations with differing electricity grid emissions intensities (composition effect). We find that in the historical period, changes in locational composition of data centers reduced carbon emissions and local air pollution damages that would have otherwise occurred by 4 and 12%, respectively. Looking ahead to projected growth from 2024–2030, we find that scale effects account for roughly 97% of the projected increase in both carbon emissions and local air pollution damages, with compositional shifts in siting contributing to the remaining 3%.

These results have important policy implications for managing the social and environmental impacts of data centers. It is clear that the environmental footprints of data centers vary depending on where they are located: Appendix Figure A.2 and A.3 show the variation in carbon pollution and monetized damages from local air pollution across 13 electricity regions in the United States. This variation suggests opportunities to mitigate the carbon and local air pollution of data centers through strategic siting. Yet, our compositional analysis finds that the vast majority of the projected growth in national carbon and local air pollution attributable to data centers stem from scale effects. This suggests the need to consider not just the location of data center, but the social benefits and costs of data center growth overall. Further, our historical decomposition dispels the idea that locational changes are driving emissions growth.

We also document trends in socioeconomic and demographic characteristics of data center locations, past and planned. We show that new data centers announced to enter in 2026–2030 will be almost four times less population dense than current and previous data centers, and five times less dense than locations without. We also find that unlike other disamenities, data centers are not systematically locating in census tracts with lower income, higher shares of poverty, and higher shares of non-white population.[1] We interpret these findings in the context of the inverse U-shaped relationship between income and population density observed in our census tract-level data, where income initially rises as population density decreases, and then falls at very low levels of density.

The remainder of the paper proceeds as follows. Section 2 describes the data. Section 3 documents trends in data center entry and size. Section 4 presents our analysis of environmental impacts and decomposes projected emissions growth into scale and composition effects. Section 5 examines the socioeconomic and demographic characteristics of data center locations over time. Section 6 concludes.

2 Data & Methods

We use proprietary data from the *U.S. Data Center Dataset*, assembled by data provider Aterio. These data describe data centers that are currently operational, under construction, and announced for future construction in the United States. The data is sourced from data center providers, press releases, permitting documents, utility interconnection filings, investor presentations, and satellite imagery. Important for our analysis, the data set includes the location (latitude and longitude) of each data center, and an estimate of the total power nameplate capacity (MW) required to operate each data center, including information technology load and supporting infrastructure (i.e. cooling, lighting, mechanical systems). When power estimates are not available from public information, power requirements are estimated by Aterio using building square footage, data center type, and historical benchmarks from the provider or region.

We focus our analysis on trends observed in data center entry over 2010–2030. We denote data centers as past (2010–2023), new (2024, 2025), and planned (2026–2030) based on their entry year ("activation date" in the data). When estimating annual energy or emissions from data centers, we include energy requirements from data centers that entered prior to 2010. We do not observe data center retirements, so a data center that comes

[1] The environmental justice literature documents a host of disamenities that are more likely to locate in regions with lower incomes and larger non-white population share: Cain et al. (2024) provides a review of these findings in the last decade.

online in 2009, for example, is assumed to require energy throughout the study period. Appendix Figure A.1 shows the full time series of entry in the raw data, 2000–2048. The apparent surge of data center entries in 2000 is an artifact of the data set, which begins that year; we cannot distinguish data centers that entered in 2000 from those that entered earlier. Appendix Figure A.1 shows the modest entry that occurred between 2001–2010, with an average of three data centers entering per month in the U.S.,[2] and this statistic doubled from 2010–2023, with an average of six new data centers entering per month.

Data centers that are projected to come online in the future are based on announcements from data center providers themselves. Appendix Figure A.1 shows the data center entry for all future years included in the data, with the latest projected entry occurring in 2048. The tapering of projected entry in later years indicates fewer announcements at longer times horizons, not necessarily a projected decline in entry. We exclude from our analysis data centers projected to enter after 2030. We note that our data is based on announcements as of March 2026, and future announcements for entering in 2026–2030 could be forthcoming.

After excluding data centers that are missing an activation date, our final data set includes 5,105 unique data center campuses. We observe data center power requirements for roughly half of those. We impute power requirements for data centers without capacity information using state-by-year averages. Roughly 12% data centers' power requirements are further imputed based on annual means when state-by-year averages are not available. We estimate total power consumed by data centers in the United States by multiplying data center capacity by hours in the year (8760) and a load factor, which represents how much of the total capacity of the data center is being used at any point in time. Using a load factor of 0.67–1.00 together with an estimate of total 2023 electricity sales in the United States from U.S. Energy Information Administration (2025), we find that data centers consumed 3.8–5.7% of total 2023 U.S. electricity. This estimate is comparable to Shehabi et al. (2024), which found that data centers consumed 4.4% (176 TWh) of 2023 U.S. electricity. Muller (2026) estimates that data centers consumed 250 TWh in 2025: in 2025 we estimate data center energy consumption of 230 and 343 TWh at a load factor of 0.67 and 1.00, respectively. The analysis in section 4 assumes a load factor of 0.67.

On the other hand, our projections of future energy use are much larger than those of other model-based projections. McKinsey & Company (2024) projects that data centers in the United States will consume 606 terawatt-hours (TWh) by 2030. Shehabi et al. (2024) projects consumption between 325 and 580 TWh in 2028. The International Energy Agency (2025) projects U.S. consumption will increase by up to 240 TWh in 2030 com-

[2] This statistic excludes data centers in 2000 as those data centers entered in or before 2000.

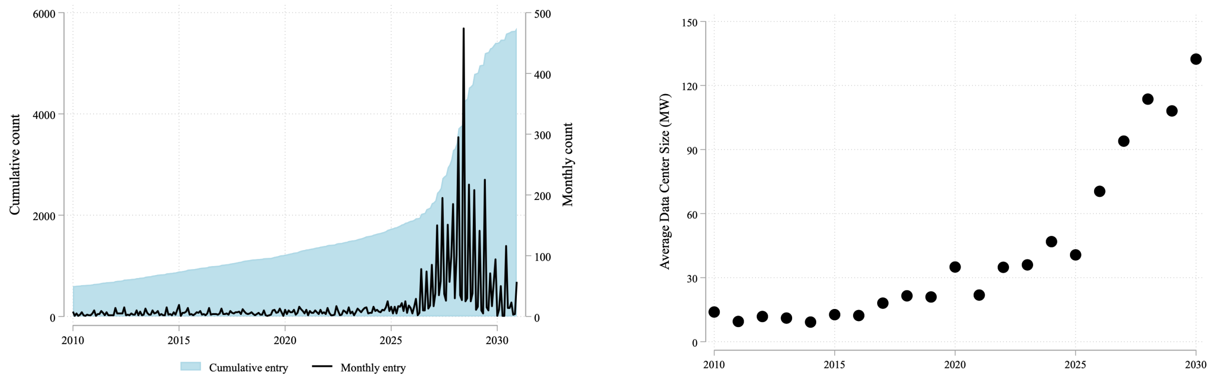
pared to 2024. The implied energy use in 2030 if all the planned data centers in our data set get built is 2,086 TWh. We view this value as an outer bound energy estimate for 2030 as there are many reasons why a planned data center would not come to fruition.

3 Entry trends

Data center entry is poised to increase dramatically in 2026–2030 compared to 2010–2025. Figure 1(a) plots the monthly count of data center entry over time in the United States. This figure shows the increase in data centers in 2026–2030 with an average of 65 new data centers per month, compared to an average of 7 per month during 2010–2025, representing an 9-fold increase in entry. In 2023–2025 Virginia led with the most annual data center entry, around 30 new data centers per year. Toward the end of the period Texas takes over as the state with the largest annual projected entry, averaging 155 new data centers per year during 2028–2030.

New and planned data centers are not just growing in count; they are also getting bigger in power requirements. Figure 1(b) shows the average size of data centers over this time period in terms of the data center nameplate capacity, measured in megawatts (MW). A data center's capacity characterizes the maximum electricity the facility can draw at any moment, and if the data center were to run at maximum capacity, one could estimate power demand in MWh by multiplying its capacity (MW) by the time utilized. Figure 1(b) shows an increase in data center size starting in 2024, continuing through 2030. Data centers that entered in 2010–2023 had an average capacity of 21 MW, while data centers entering 2024–2030 are 5 times as large, around 100 MW.

Figure 1: Data center entry (count) and size (MW) over time, 2010–2030



(a) Entry (count)

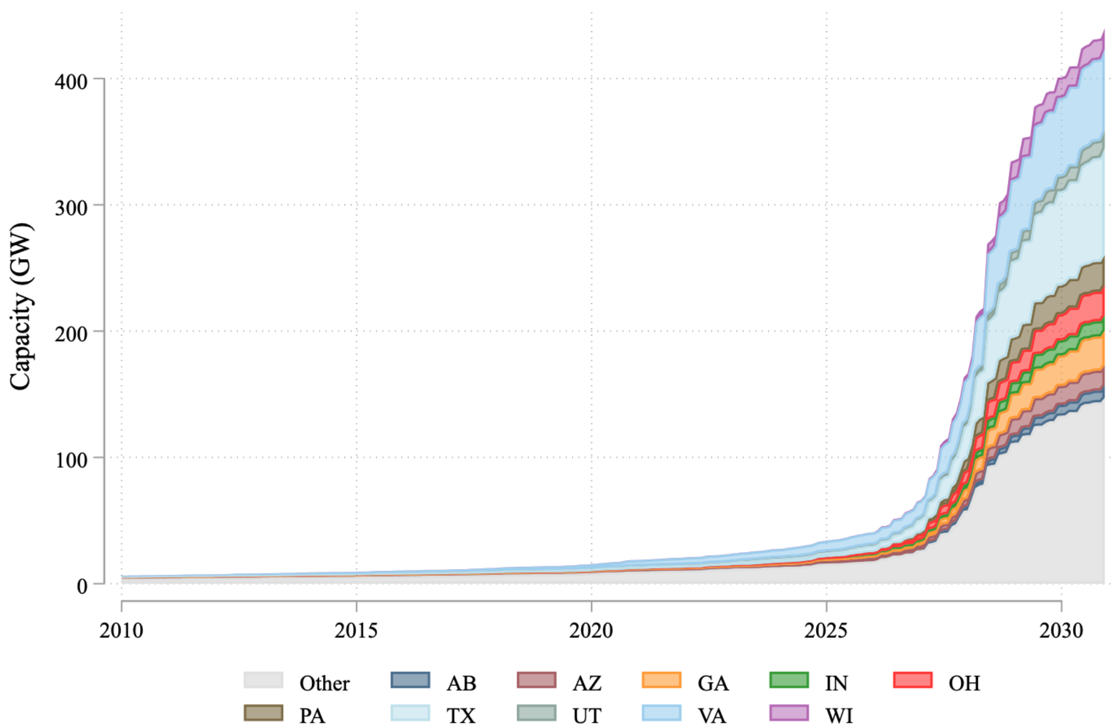
(b) Average size (MW)

NOTES: Panel (a) plots the number of monthly data centers entering in black on the left y-axis, with the cumulative number of data centers in light blue on the right axis. Panel (b) shows the average size (nameplate capacity, MW) of data centers over time.

Next, we construct cumulative energy requirements implied by the stated capacity of the projected data centers. Figure 2 plots cumulative capacity for data centers over 2010–2030, delineated among states with the largest data center capacity as of the end of 2030, which are the following states in order of their ranking from most to least: Texas, Virginia, Georgia, Ohio, Pennsylvania, Utah, Wisconsin, Arizona, Alabama, Indiana .[3] Figure 2 shows an increase in capacity from roughly 0.5 GW at the end of 2010 to 430 GW by the end of 2030. We reiterate that to the extent not all announced data centers get built, this would be an outer bound estimate. On the other hand, this is based on announcements as of March 2026, and more announcements for entry during this period could be forthcoming.

[3] Translating these capacities into energy requires assuming about how many hours data centers are run, and a "load factor", or an assumption about what share of the total capacity of the data center is used. We make these assumptions later when estimating environmental costs of data centers. Since both are linear transformations the growth in capacity is informative on the growth in energy.

Figure 2: Data center capacity requirements, top states



NOTES: The figure plots cumulative data center capacity among states with the most planned data center capacity by the end of 2030.

4 Environmental impacts

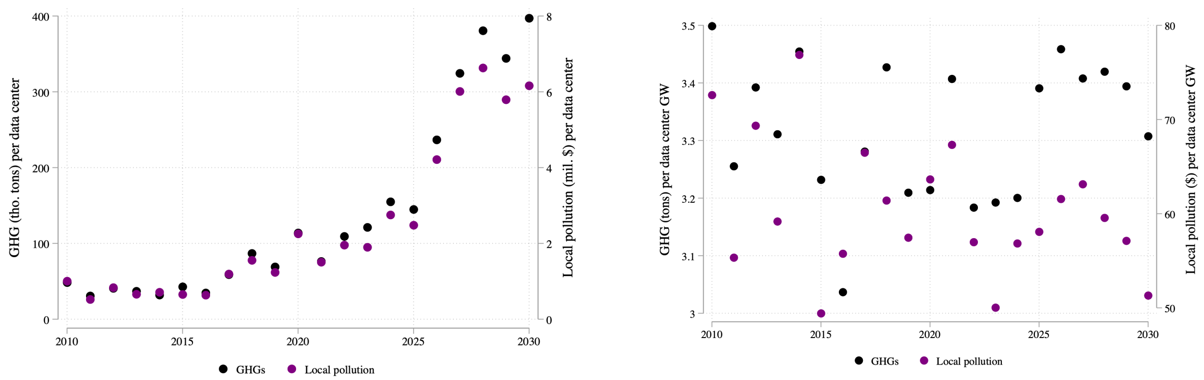
Next we study trends in the pollution impacts of data centers, in terms of both carbon emissions and monetized local air pollution. To do so, we make use of regional marginal emissions estimates from Holland et al. (2024), which provides by season (winter/summer) and by hour estimates of the CO_2 (tons) and local air pollution damages (dollars) generated from an additional unit of electricity demand in 13 different regions in the United States. We obtain average annual emissions intensities by equally weighting hour-season specific estimates. The emission intensities reflect grid conditions as of year 2023 and are shown in Appendix Figure A.2 for carbon emissions and Figure A.3 for monetized local air pollution. Other work has used these type of marginal emissions estimates to calculate the total pollution costs of data centers, for example, Muller (2026). Our aim is to study trends in emissions intensities of data centers over time, asking if data centers have been entering into more or less emission intensive places.

Figure 3 plots the average carbon emissions and monetized local air pollution intensity of a new data center over time with two perspectives. Figure 3(a) plots emissions per unique data center, which are obtained by dividing the total carbon and air pollution damages in each year by the total number of data centers that enter in that year. As a result, this figure does not normalize for the fact that the average size of a data center is growing over time, which is shown in Figure 1(b). Figure 3(a) shows large increases in the carbon and local pollution impact per data center over time: the average annual GHG emissions of a data center was 69 thousand tons per year in 2010–2023, increasing by almost 5-fold to 283 thousand tons per year in 2024–2030. Annual local air pollution damages also increased around four times in 2024–2030 compared to 2010–2023: local air pollutions costs per data center were $1.2 million per year in 2010–2023 and $4.9 million per year for 2024–2030.

Figure 3(b) now normalizes by data center size. The figure shows the emissions intensity per GW of data center capacity, estimated as the total carbon emissions and local air pollution damages in each year divided by the total capacity in each year. Thus, variation in this figure is generated entirely through changes in the locational composition of data centers, e.g., if data centers are moving into places with higher or lower marginal carbon emissions intensities and local pollution damages. This figure shows an ambiguous relationship between emissions intensity per unit of data center power for carbon as well as local air pollution. This illustrates that the increasing trend in emissions per data center in Figure 3(a) is being driven by increases in data center power requirements, and not a shift in entry locations to more emissions and local pollution intensive electricity grids.

This point is further explored in our next decomposition analysis.

Figure 3: Pollution per data center by count and power, 2010–2030



(a) Pollution per data center

(b) Pollution per data center GW

NOTES: Panel (a) shows annual pollution in GHG tons and monetized air pollution damages per number of new data centers. Panel (b) shows annual pollution in GHG tons and monetized air pollution damages per GW capacity of new data centers.

4.1 Decomposition analysis

Next we study the relative roles of data center growth (i.e., increase in power requirements, MWh) versus changes in the locations of where data centers are entering (i.e., changes in emissions intensities of location, $\frac{CO_2(tons)}{MWh}$, $\frac{MWh}{\$\ local\ pollution}$). In the context of decomposition analysis, the former is referred to as a "scale" effect, and the later is a "compositional" effect. This decomposition allows us to ask: how much of data centers emissions footprint is being driven by changes in the scale of the deployment versus changes in the emissions intensities of their locations. Let j index data centers and t index years. Total annual carbon emissions E_t from data centers can be modeled as:

$$E_t = \sum_j \phi_j \cdot \delta_{jt} \cdot Q_t \tag{1}$$

where $Q_t = \sum_j E_{jt}$ is total data center energy consumption, $\delta_{jt} = E_{jt}/Q_t$ is data center j's share of total energy consumption, and ϕ_j is the time-invariant grid emissions intensity at data center j's location, $\frac{CO_2\ tons}{MWh}_j$. The change in total emissions between years 0 and t decomposes as:

$$\Delta E_t = \underbrace{\sum_j \phi_j \bar{\delta}_j \Delta Q_t}_{\text{Scale}} + \underbrace{\sum_j \phi_j \bar{Q}_t \Delta \delta_{jt}}_{\text{Composition}} \quad (2)$$

where $\bar{\delta}_j = (\delta_{j0} + \delta_{jt})/2$, $\bar{Q}_t = (Q_0 + Q_t)/2$, $\Delta Q_t = Q_t - Q_0$, and $\Delta \delta_{jt} = \delta_{jt} - \delta_{j0}$. Because this analysis uses time invariant emissions intensities, this decomposition is exact with no residual. The scale term captures emissions growth due to rising total data center energy consumption, weighted by the average geographic distribution of that consumption. The composition term captures emissions changes due to shifts in the geographic distribution of data center energy, weighted by each location's emissions intensity.

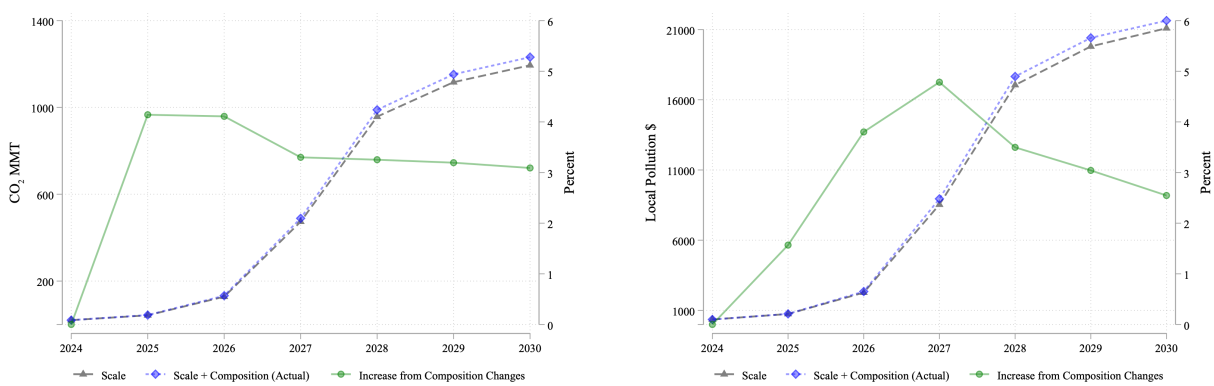
We perform the decomposition for carbon emissions and monetized local air pollution. For local air pollution, the left-hand side of equation 1 is total annual air pollution damages in dollars, and ϕ_j is the time-invariant local pollution intensity at data center j's location in dollars per MWh. Figure 4a shows the results of the decomposition analysis for carbon emissions using 2024 as the base year. This figure shows the relative contribution of scale and compositional changes to the annual data center carbon emissions compared to emissions in 2024. We chose 2024 as the base year, given the trends shown in Figure 1(a), which shows the increase in data center entry starting in 2025. The gray line represents the annual carbon emissions that would occur from changes in data center energy demand only, holding the composition of data center locations fixed as of 2024. The blue line shows the actual total projected annual carbon emissions, which combines scale and compositional effects. The green line shows the percent of carbon emissions increase that occurs due to compositional changes, e.g. changes in the locational composition of data centers.

The modest differences between the blue and gray lines indicate that compositional changes to data center locations over time play a small role in the overall year to year changes. As shown in the green line, compositional changes account for just 3% of 2030 emissions growth compared to 2024 emissions, with scale effects accounting for the other 97%. This figure also shows that if all announced and planned data centers are built, carbon emissions from data centers will increase by 61-fold. We interpret this as an outer bound, since we do not expect all announced and planned data centers to get built. However, if the locations of announced data centers are representative of those that do get built, our estimates of the relative roles of scale versus compositional effects will remain approximately valid.[4]

[4]The decomposition is exact at any scale of build-out, and total emissions growth ΔE_t is linear in the share of announced capacity that gets built. The relative shares of the scale and composition terms shift only

Figure 4b shows a similar decomposition for monetized local air pollution. If all announced and planned data centers get built, our projections indicate that local air pollution damages from data centers will increase by 60-fold in 2030 compared to 2024. We estimate that compositional changes accounted for 2.5% of this growth, with scale changes accounting for the other 97.5%. As with our estimates of energy use and carbon emissions, this is an outer bound to the extent that not all announced and planned data centers will get built, but a lower bound to the extent that data center announcements are forthcoming. In either case, if the locations of announced-planned data centers are representative of the distribution of the locations of the data centers that eventually get built, our estimates of the role of scale versus compositional effects are still approximately valid.

Figure 4: Scale versus composition effects of data center emissions, 2024–2030



(a) Carbon emissions

(b) Monetized local air pollution

NOTES: The figure shows the contribution of scale–growth in energy requirements from data centers–and composition–the emissions intensities of their locations–to annual carbon emissions in panel (a) and monetized local air pollution damages in panel (b), using 2024 as the base year.

The specification above uses 2024 as the base year, given the projected growth in data center entry, as shown in Figure 1 begins in 2025. We also perform the same analysis using 2010 as the base year, which allows us to study scale and compositional changes when comparing outcomes in 2030 to those in 2010. The results are provided in Appendix Figures A.4. These figures show that compositional changes in data center locations over 2010–2030 lead to a reduction in emissions, compared to a counterfactual in which data center growth occurred with the same geographical composition as in 2010. Specifically, compositional changes in locations resulted in a 1.5% reduction in carbon emissions and

slightly with this share, through the nonlinear dependence of location shares, $\delta_{jt} = E_{jt}/Q_t$, on aggregate energy, Q_t.

an almost 11% reduction in monetized local air pollution in 2030 compared to 2010. In other words, data centers entering in 2011 and beyond were in relatively cleaner places compared to the distribution data centers in 2010. Finally, we perform a decomposition over the period 2010–2025 to study emissions and local pollution growth from new data centers compared to 2010. The results provided in Appendix Figure A.5 show that compared to 2010, the distribution of data center locations have lower levels of pollution intensity, such that compositional changes lead to a decrease in carbon emissions in 2025 compared to 2010 by around 4% and a decrease in local pollution by around 12%.

In sum, while the specific results vary depending on the period studied (e.g. from 2010–2025 vs. 2010–2030 vs. 2024–2030), all three perspectives dispel the idea that the observed and projected growth in data center pollution is driven by compositional changes. Compared to a counterfactual where data centers 2010 location composition persists, the 2025 and 2030 locational distributions reduce pollution. When comparing the projected growth in pollution from planned 2030 data centers, the locational composition increases pollution, but only by around 3% for both carbon emissions and local air pollution damages.

5 Socioeconomic and demographic composition

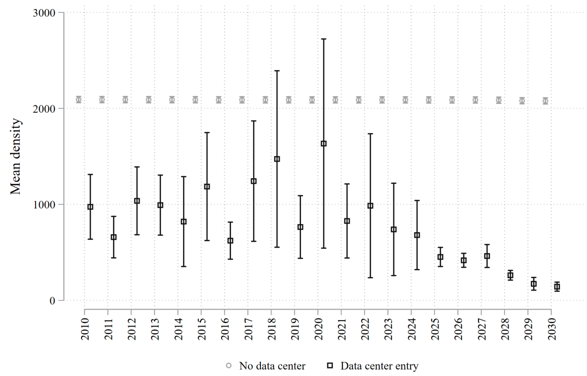
Next we study trends in the socioeconomic and demographic characteristics of the places data centers are entering. To do so, we match data center locations to the characteristics of their census tracts. We use data from the 2020 Decennial Census to obtain tract-level population density, share of Non-Hispanic white, median income, and poverty-share. We fix these characteristics to the year 2020, so that our analysis of the trends in location characteristics is driven only by changes in locations of data centers and not changes in locations characteristics over time.

Figure 5 plots the mean (circles) and 95% confidence intervals (bars) of different location characteristics among the census tracts where data centers enter in each year in black. The gray data series plot the mean (circles) and 95% confidence intervals (bars) among census tracts without data centers. Figure 5(a) shows that data centers have been consistently entering census tracts with lower population densities over the entire period. Perhaps more notable, new and planned data centers are trending towards increasingly less dense tracts over the 2010–2030 period. Appendix Table A.1 quantifies this trend change: data centers that are projected to enter in 2026–2030 locate in tracts with a mean population density of 383 pop / sq. km, whereas from 2010–2023 the mean density of tracts where data centers entered was 1,534 pop / sq. km. Appendix Figure A.6 shows the mean and confidence intervals of the share of rural population, showing a similar trend as with population density.

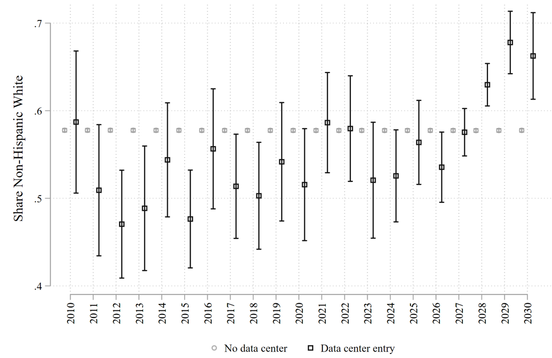
Figure 5(b) shows the share of non-Hispanic white population among census tracts where data centers have and are planned to locate. The figure shows that in many years tracts with and without data centers do not have statistically different non-Hispanic white shares, though on average they enter in tracts with lower shares of non-Hispanic white population from 2010–2027. In 2028–2030 planned data centers are projected to enter locations with statistically higher shares of non-Hispanic white populations. Figure 5(c) shows that data centers enter locations on average with higher median income, and this difference is statistically precise in most years. Notably, the gap in income decreases from 2025–2030. By 2029, data centers enter census tracts with median incomes that are comparable to census tracts that do not see new data centers. Figure 5(d) shows that on average, data centers tend to locate in places with lower shares of households in poverty, though this difference is not statistically significant over the 2010–2018 period.

The relationship between data center locations and income and poverty share runs counter to traditional expectations about where disamenities locate. Canonical economic theory (Rosen, 1974) hypothesizes that features of the environment that are perceived as disamenities to households would be located in places with lower income. This mechanism works in tandem with the fact that conditional on other factors impacting firm profitability, firms seek lower land prices. Appendix Figure A.7 sheds light on this finding. Recall that Table A.1 shows that the mean population density (pop / sq km) among census tracts with data centers moves from 1,534 for past data centers (2010–2023) to 610 for new data centers (2024–2025) to 383 for planned data centers (2026–2030). Appendix Figure A.7 plots the relationship between income and density for all U.S. census tracts. This shows that in the range of densities where new or planned data centers are located, income and density have a quadratic relationship. As densities initially decrease, income increases, but decreases in density beyond around 500 (pop / sq km) come with reductions in average tract-level income. This is consistent with populations in slightly less dense suburban regions having higher median incomes, but populations in increasingly less dense and more rural-locations having lower incomes. Appendix Figure A.6 plots the share of the census tract population that is rural based on the U.S. Census Bureau's definition. This figure shows a sharp increase in the rural population share among data centers entering in 2027–2030. In sum, Figure 5(a),(c), and (d), and Appendix Figure A.7 and A.6, show that data centers are moving to census tracts that are increasingly rural and on the right side of the inverse U-shaped function describing income and density.

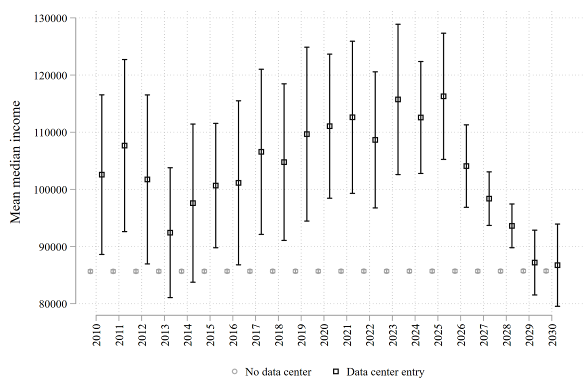
Figure 5: Location characteristics of entering data centers, 2010–2030



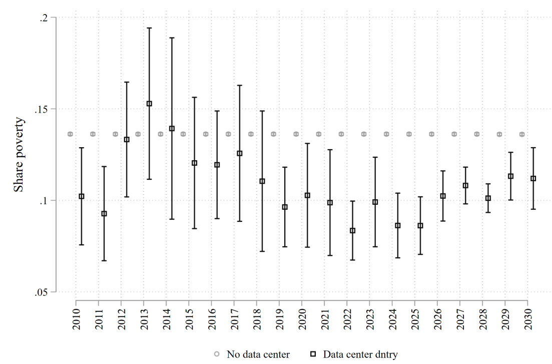
(a) Population density



(b) Share non-Hispanic white



(c) Median income



(d) Share poverty

NOTES: This figure plots the mean (circles) and 95% confidence intervals (bars) of census tracts where data centers are entering in black, and the mean (circles) and 95% confidence intervals (bars) of census tracts without data centers in gray. Socio-economic characteristics come from the 2020 Decennial Census.

6 Conclusion

The rapid buildout of data center infrastructure in the United States has the potential to reshape local landscapes, communities, and electricity grids. This paper documents trends in data center locations, finding that they have consistently located in lower population density areas, and that this trend is accelerating: facilities projected to come online between 2026 and 2030 will locate in census tracts with population densities five times lower than census tracts without data centers, and nearly four times lower than data centers that entered between 2010 and 2023. Contrary to our expectations about where disamenities typically locate, data centers are not being located in places that are particularly low income or with high poverty rates. Further, toward the end of the study period planned data centers are entering tracts with statistically higher share non-Hispanic white population. Together, our findings describe where the locus of future permitting debates and community-level negotiations over data center entry will occur: in the United States' least population dense and increasingly rural communities. Local regulatory capacity, from legal expertise, planning staff, and institutional knowledge, needed to negotiate favorable terms with large industrial entrants likely scales with population. The shift toward smaller, more rural host communities may then place future data center negotiations in localities least equipped to secure beneficial terms on conditions for entry.

This paper also contributes to the literature on the environmental costs of data centers. We show that our data set and methodology leads to estimates of data center energy consumption in recent years that are highly comparable to other research (Shehabi et al., 2024; Muller, 2026). We project the growth in data center energy and emissions based on data centers announced to come online and decompose this growth into scale and compositional effects. This approach allows us to study whether the projected growth is due to entry in more emissions intensive locations, or increases in entry. Our aim with this analysis was to understand how strategic siting can be used as a policy tool to manage the environmental costs of data centers. Our results suggest that this type of nudging would have a negligible impact on the overall growth of data center induced emissions. We find that compositional shifts in the geographic distribution of data center accounts for just 3% of projected growth in carbon emissions between 2024 and 2030, with an analogous 2.5% share for monetized local air pollution damages. The remaining 97% and 97.5% of projected carbon and local pollution growth is attributable to the sheer scale of expansion in data center power capacity.

There are several important caveats for our analysis. First, our projections in 2026–2030 are based on announcements of planned data centers from providers themselves. Not all of these data centers will get built, and we exclude data centers that are announced after March 2026. That said, to the extent that the locations of announced data centers are representative of those ultimately constructed, this would have a negligible impact on the findings of our decomposition analysis, nor a large impact on our findings describing the local socioeconomic and demographic characteristics of data center locations. Next, we assume data centers do not retire once they come online, which could modestly bias our estimates upward, though given the rapid demand for data centers among end using AI companies, this assumption seems reasonable. Another important caveat regards our use of marginal emissions intensity estimates. The estimates reflect grid conditions as of 2023, and the grid may look different in 2030, in particular as new electricity generation and transmission investments respond to data center entry. Finally, the local pollution estimates in our paper do not indicate which communities receive these damages. Future work estimating these impacts can identify where local pollution costs of data centers accrue.

References

Alvarez, Fernando E., David Argente, Joyce Chow, and Diana Van Patten. 2026. "Data Centers and Local Economies in the Age of AI: A Shift–Share Approach." National Bureau of Economic Research NBER Working Paper 35194.

Cain, Lucas, Danae Hernandez-Cortes, Christopher Timmins, and Paige Weber. 2024. "Recent findings and methodologies in economics research in environmental justice." *Review of Environmental Economics and Policy*, 18(1): 116–142.

Holland, Stephen P, Erin T Mansur, Valentin Verdier, and Andrew J Yates. 2024. "Regularization from economic constraints: A new estimator for marginal emissions."

International Energy Agency. 2025. "Energy and AI." International Energy Agency World Energy Outlook Special Report, Paris. Licence: CC BY 4.0. U.S. data centre electricity consumption projected to increase by around 240 TWh (up 130%) by 2030 relative to 2024; see "Energy demand from AI" chapter.

Luccioni, Alexandra Sasha, Yacine Jernite, and Emma Strubell. 2024. "Power Hungry Processing: Watts Driving the Cost of AI Deployment?" Rio de Janeiro, Brazil:ACM.

Manson, Steven, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. 2024. "IPUMS National Historical Geographic Information System: Version 19.0 [dataset]."

McKinsey & Company. 2024. "Artificial Intelligence Power Demand in Charts." `https://www.mckinsey.com/featured-insights/week-in-charts/ais-power-binge`, Accessed: 2026-05-19.

Muller, Nicholas Z. 2026. "Measuring the Impact of Data Centers in the United States Economy: Monetary Damage from Air Pollution and Greenhouse Gas Emissions." National Bureau of Economic Research.

Rosen, Sherwin. 1974. "Hedonic prices and implicit markets: product differentiation in pure competition." *Journal of political economy*, 82(1): 34–55.

Shehabi, Arman, Sarah J. Smith, Alex Hubbard, Alex Newkirk, Nuoa Lei, Md Abu Bakar Siddik, Billie Holecek, Jonathan Koomey, Eric Masanet, and Dale Sartor. 2024. "2024 United States Data Center Energy Usage Report." Lawrence Berkeley National Laboratory LBNL-2001637, Berkeley, California. Accessed: 2025-12-04.

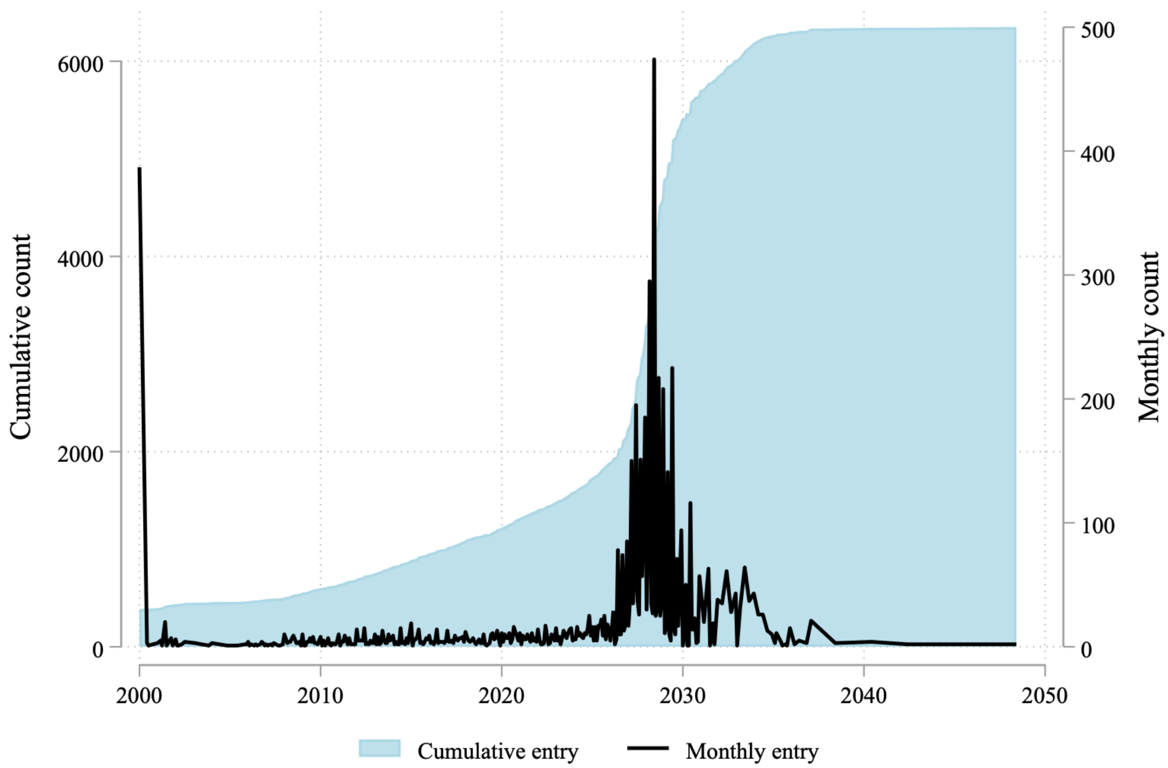
Shehabi, Arman, Sarah J. Smith, Dale A. Sartor, Richard E. Brown, Magnus Herrlin, Jonathan G. Koomey, Eric R. Masanet, Nathaniel Horner, Inês L. Azevedo, and William Lintner. 2016. "United States Data Center Energy Usage Report." Lawrence Berkeley National Laboratory LBNL-1005775, Berkeley, California.

U.S. Energy Information Administration. 2025. "Electric Power Annual, Table 1.2: Summary Statistics for the United States." U.S. Department of Energy Electric Power Annual. With data for 2024; released October 16, 2025. Accessed 2026-05-18.

Appendix

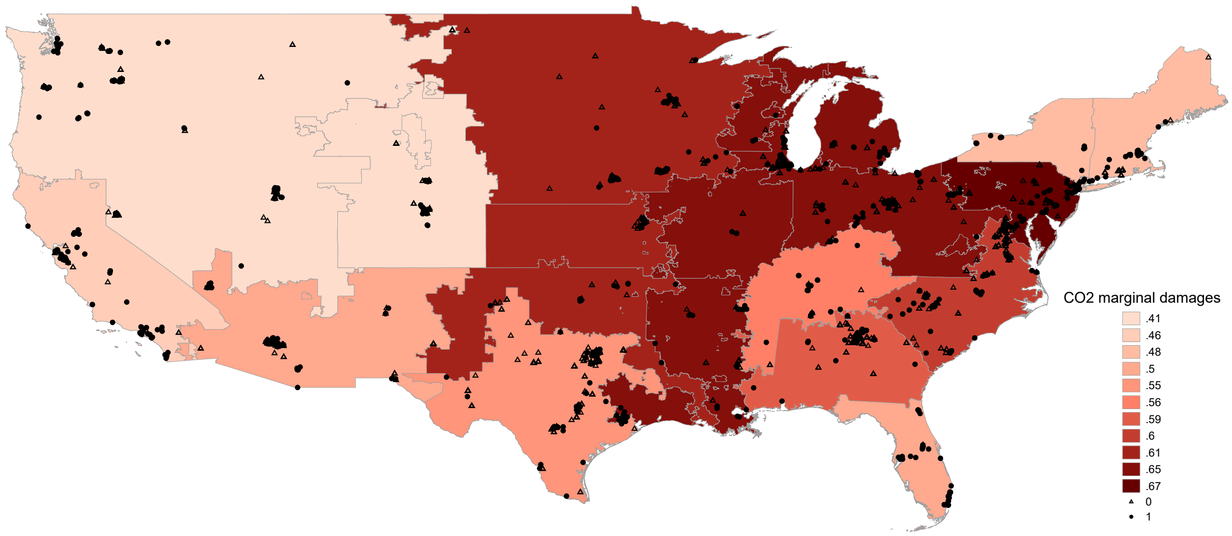
A Appendix Exhibits

Figure A.1: Data center entry (count) over time, 2000–2048



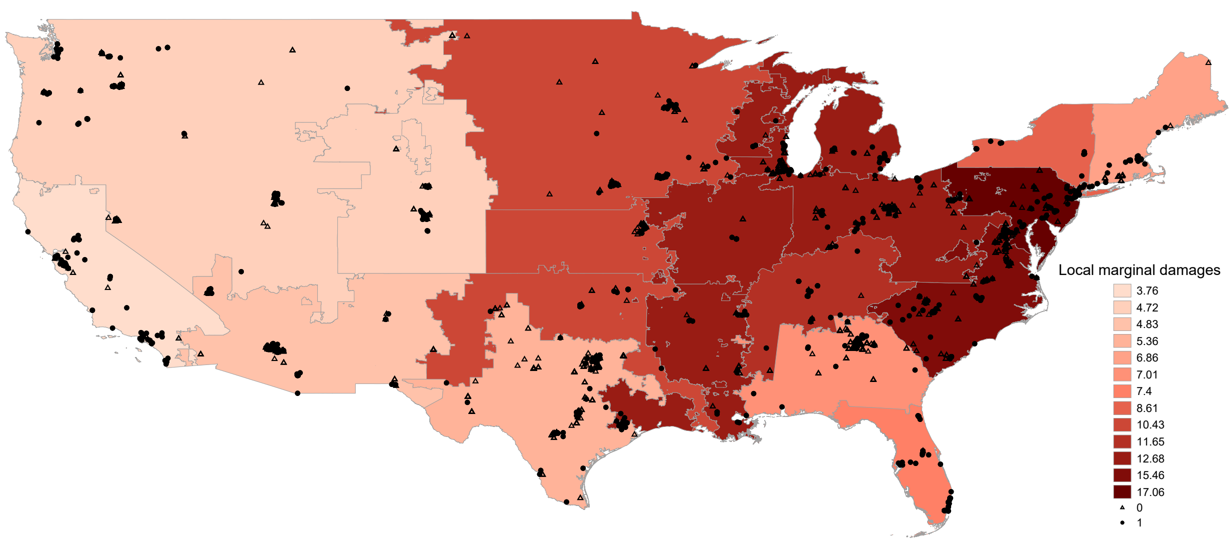
Notes: This figure plots the number of monthly data centers entering in black on the left y-axis, with the cumulative number of data centers in light blue on the right axis. The first year of the data is 2000, and the large spike in that year represents all data centers active in that year, which may have entered in or before 2000. Future projections reflect provider announcements as of March 2026, and so tapering in later years indicates fewer announcements at longer times horizons, not necessarily a projected decline in entry.

Figure A.2: Data centers and CO_2 Emissions Intensities



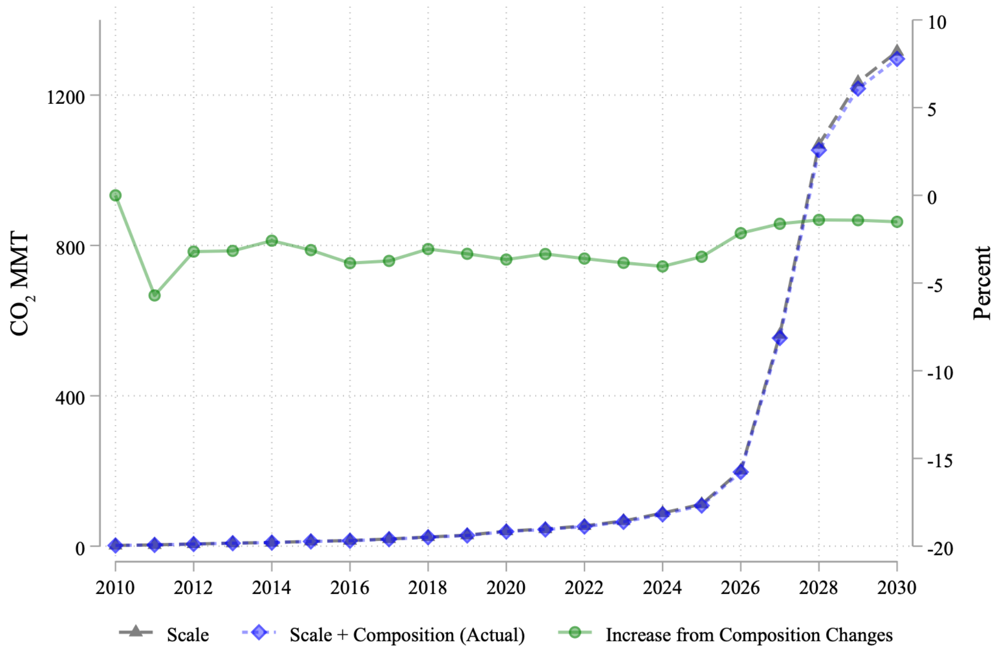
Notes: This figure shows data center locations, active (circle) and proposed (triangle) together with marginal CO_2 emissions factors, estimated by Holland et al. (2024) across 13 electricity regions in the United States.

Figure A.3: Data centers and Local Pollution Damage Intensities

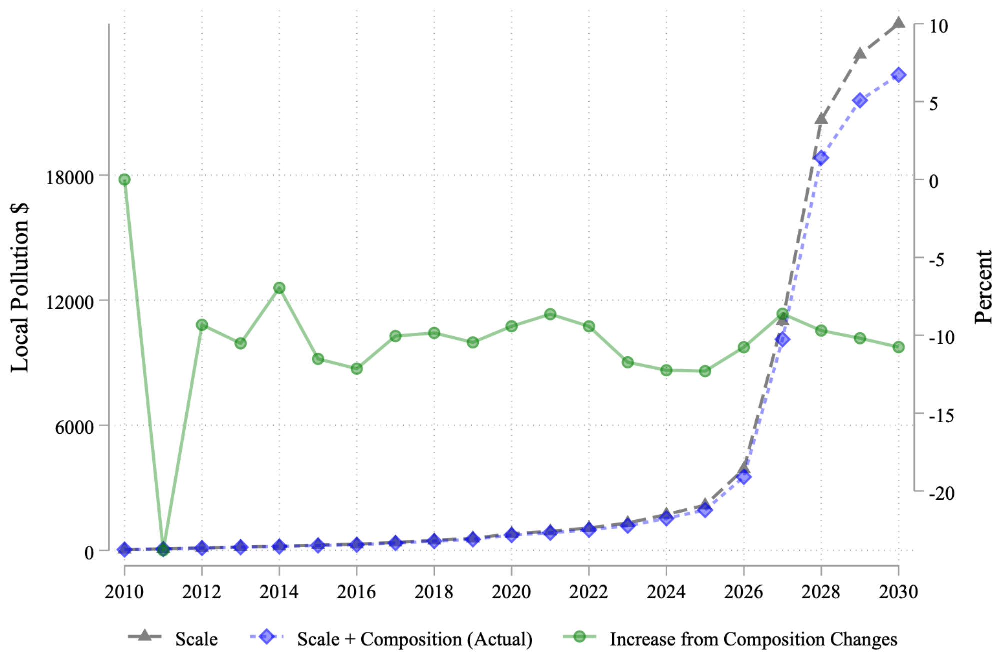


Notes: This figure shows data center locations, active (circle) and proposed (triangle) together with marginal damages ($) from local air pollution, estimated by Holland et al. (2024) across 13 electricity regions in the United States.

Figure A.4: Scale versus composition effects, 2010–2030



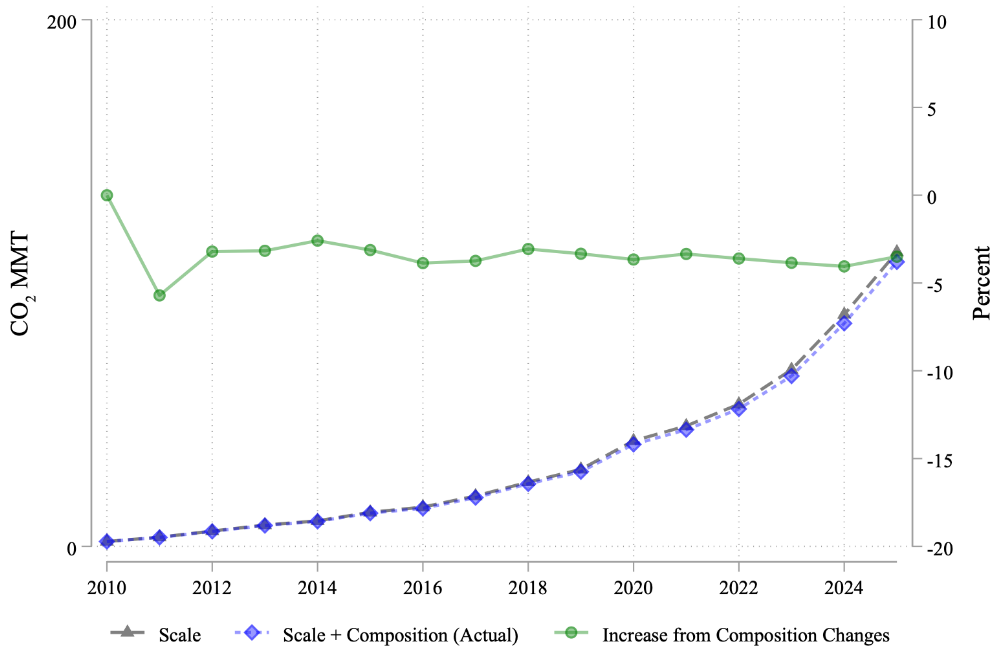
(a) CO_2 emissions



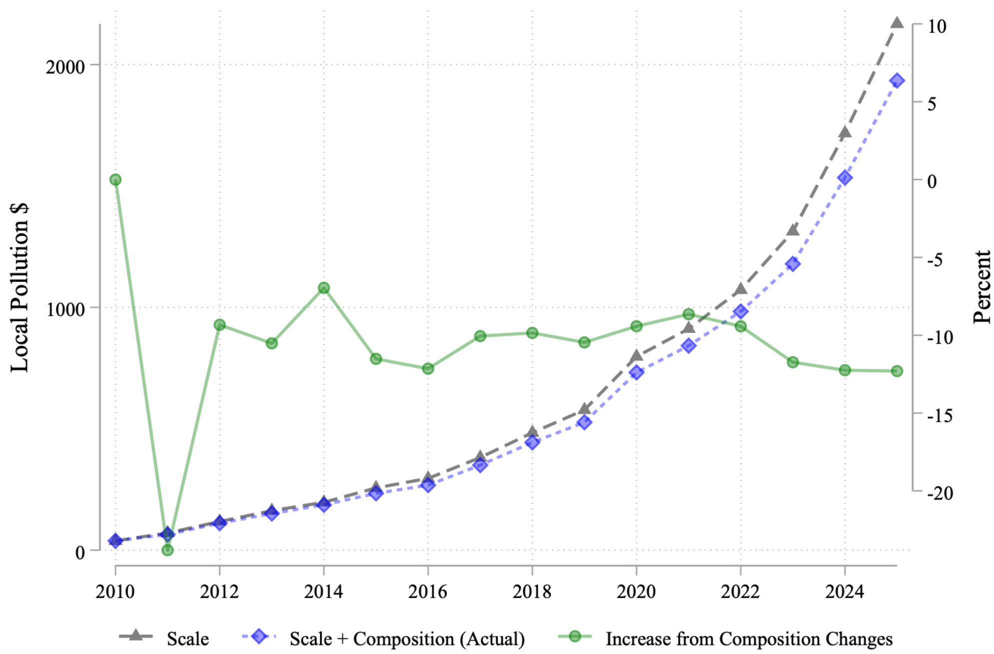
(b) Monetized local air pollution damages

NOTES: The figure shows the contribution of scale–growth in energy requirements from data centers–and composition–the emissions intensities of their locations–to the carbon emissions and monetized annual local air pollution damages of data centers, using 2010 as the base year.

Figure A.5: Scale versus composition effects, 2010–2025



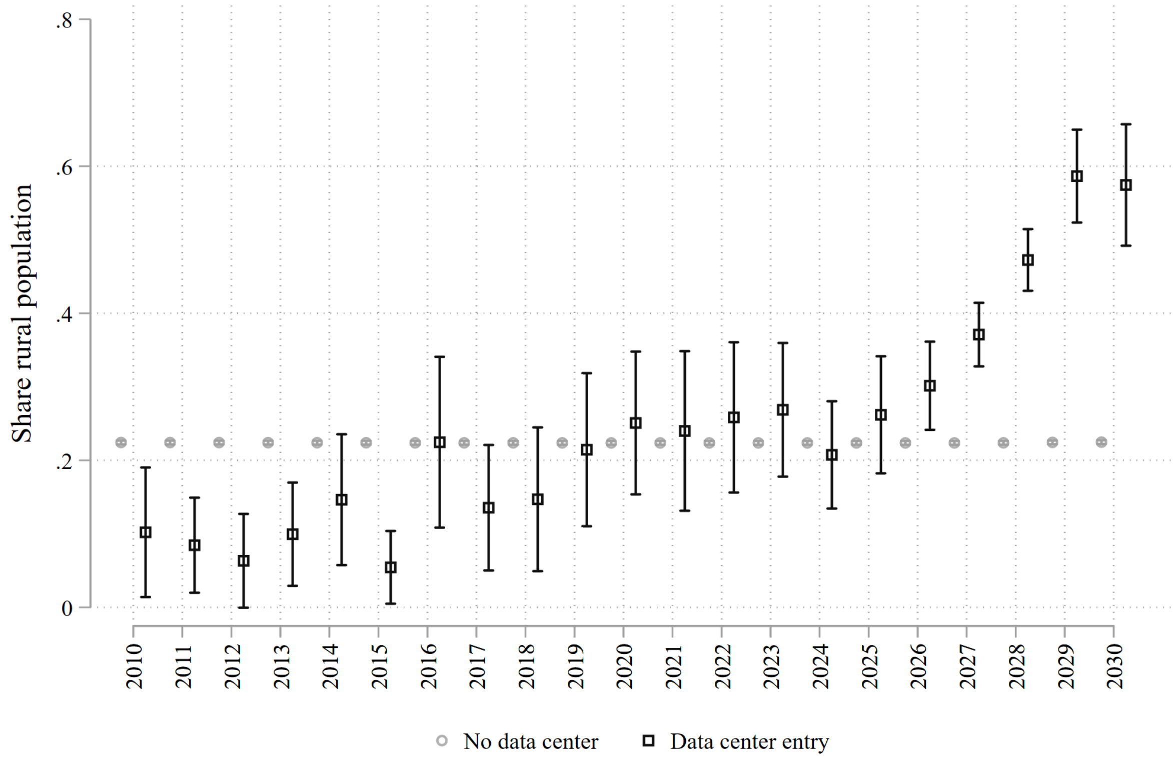
(a) CO_2 emissions



(b) Monetized local air pollution damages

NOTES: The figure shows the contribution of scale–growth in energy requirements from data centers–and composition–the emissions intensities of their locations–to the carbon emissions and monetized annual local air pollution damages of data centers, using 2010 as the base year, studying growth through 2025.

Figure A.6: Location characteristics of entering data centers: share of rural population



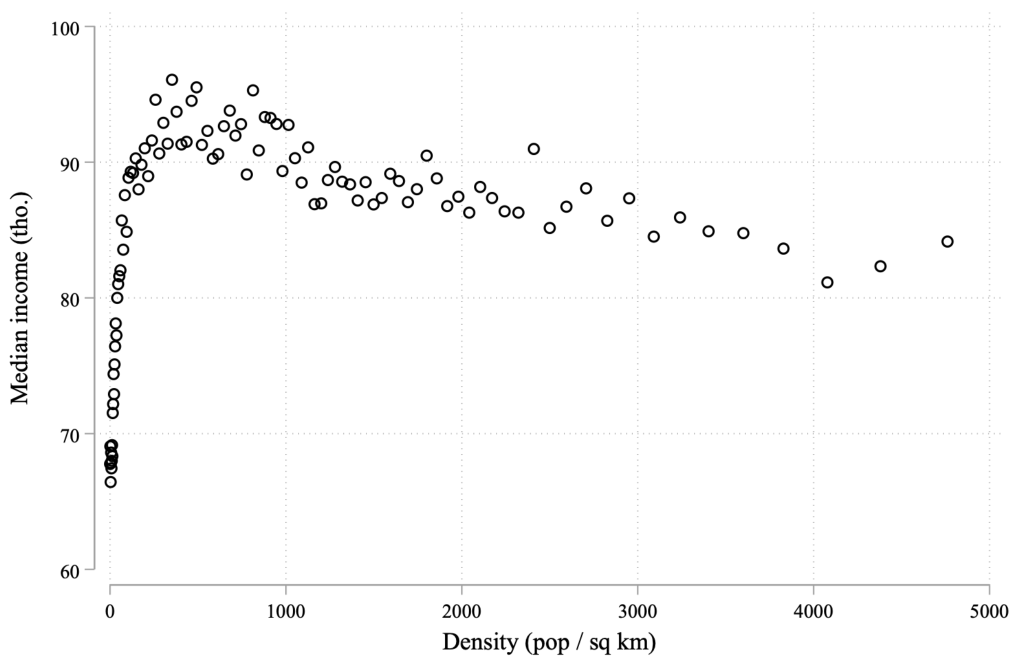
NOTES: This figure plots the mean (circles) and 95% confidence intervals (bars) of census tracts where data centers are entering in black, and then mean (circles) and 95% confidence intervals (bars) of census tracts without data centers in gray. Rural population share is obtained by dividing total rural population by total population in the census tract from the 2020 Decennial Census, both obtained from IPUMS NHGIS, based on the U.S. Census definition. (Manson et al., 2024).

Table A.1: Characteristics of census tracts with and without data center by entry period

	(1) CT w/data center (all U.S.)	(2) CT w/o data center (all U.S.)	(3) Difference p-value
		Past data centers (2010–2023)	
Density (pop/sq km)	1,534	2,080	0.000
	(3,116)	(4,745)	
Total census tracts	728	84,332	85,060
		New data centers (2024–2025)	
Density (pop/sq km)	610	2,078	0.000
	(1,360)	(4,737)	
Total census tracts	142	84,918	85,060
		Planned data centers (2026–2030)	
Density (pop/sq km)	383	2,089	0.000
	(889)	(4,750)	
Total census tracts	705	84,355	85,060
		Past data centers (2010–2023)	
Median income	95,216	85,646	0.000
	(47,055)	(41,362)	
Total census tracts	698	83,130	83,828
		New data centers (2024–2025)	
Median income	112,908	85,681	0.000
	(49,475)	(41,392)	
Total census tracts	139	83,689	83,828
		Planned data centers (2026–2030)	
Median income	91,608	85,678	0.000
	(38,617)	(41,440)	
Total census tracts	686	83,142	83,828
		Past data centers (2010–2023)	
Poverty share	0.14	0.14	0.891
	(0.13)	(0.11)	
Total census tracts	706	83,573	84,279
		New data centers (2024–2025)	
Poverty share	0.09	0.14	0.000
	(0.08)	(0.11)	
Total census tracts	139	84,140	84,279
		Planned data centers (2026–2030)	
Poverty share	0.11	0.14	0.000
	(0.09)	(0.11)	
Total census tracts	689	83,590	84,279
		Past data centers (2010–2023)	
Non-Hispanic White share	0.50	0.58	0.000
	(0.24)	(0.29)	
Total census tracts	725	84,056	84,781
		New data centers (2024–2025)	
Non-Hispanic White share	0.55	0.58	0.220
	(0.24)	(0.29)	
Total census tracts	141	84,640	84,781
		Planned data centers (2026–2030)	
Non-Hispanic White share	0.60	0.58	0.009
	(0.26)	(0.29)	
Total census tracts	699	84,082	84,781
		Past data centers (2010–2023)	
Rural population share	0.08	0.23	0.000
	(0.23)	(0.38)	
Total census tracts	725	84,056	84,781
		New data centers (2024–2025)	
Rural population share	0.22	0.23	0.966
	(0.36)	(0.38)	
Total census tracts	141	84,640	84,781
		Planned data centers (2026–2030)	
Rural population share	0.41	0.22	0.000
	(0.42)	(0.38)	
Total census tracts	699	84,082	84,781

NOTES: This table shows average location characteristics–population density, median income, poverty share, share non-Hispanic White–by data entry period–past (2010–2023), new (2024–2025) and planned (2026–2030), with the standard deviation shown in parentheses. Characteristics are based on the 2020 Decennial Census. Rural population share is obtained by dividing total rural population by total population in the census tract from the 2020 Decennial Census, both obtained from IPUMS NHGIS, based on the U.S. Census definition. (Manson et al., 2024).

Figure A.7: Population density and median income



NOTES: This figure shows a binned scatter plot with 100 quantiles of census tract population density and median income among all census tracts with population density less than 5000 people per sq. km.