Evaluating the Environmental Impact of Hyperscale Data Centers in the U.S.

Francesca Dominici, PhD
Professor of Biostatistics, Population Health and Data Science
Harvard T.H. Chan School of Public Health
Director of the Harvard Data Science Initiative

HDSI | Harvard Data Science Initiative

Joint work with Gianluca Guidi, Tiziano Squartini, Callaway Sprinkle, Jonathan Gilmour Kevin Butler, Eric Bell, Scott Delaney, Falco J. Bargagli-Stoffi

Hyperscale Data Centers

Characteristics of Hyperscale Data Centers

Hyperscale data centers are the largest and most powerful class of data centers.

They are designed to handle huge computing workloads and data storage requirements.

Energy Demands of Data Centers

Data centers are naturally energy-intensive facilities.

Their main electricity needs come from computational power and cooling systems.

Specifications of Hyperscale Data Centers

They are large modular units spanning approximately 2,000 square meters each.

They have average power capacities exceeding 40 megawatts, enough to power tens of thousands of servers simultaneously.

At Amazon's Biggest Data Center, Everything Is Supersized for A.I.

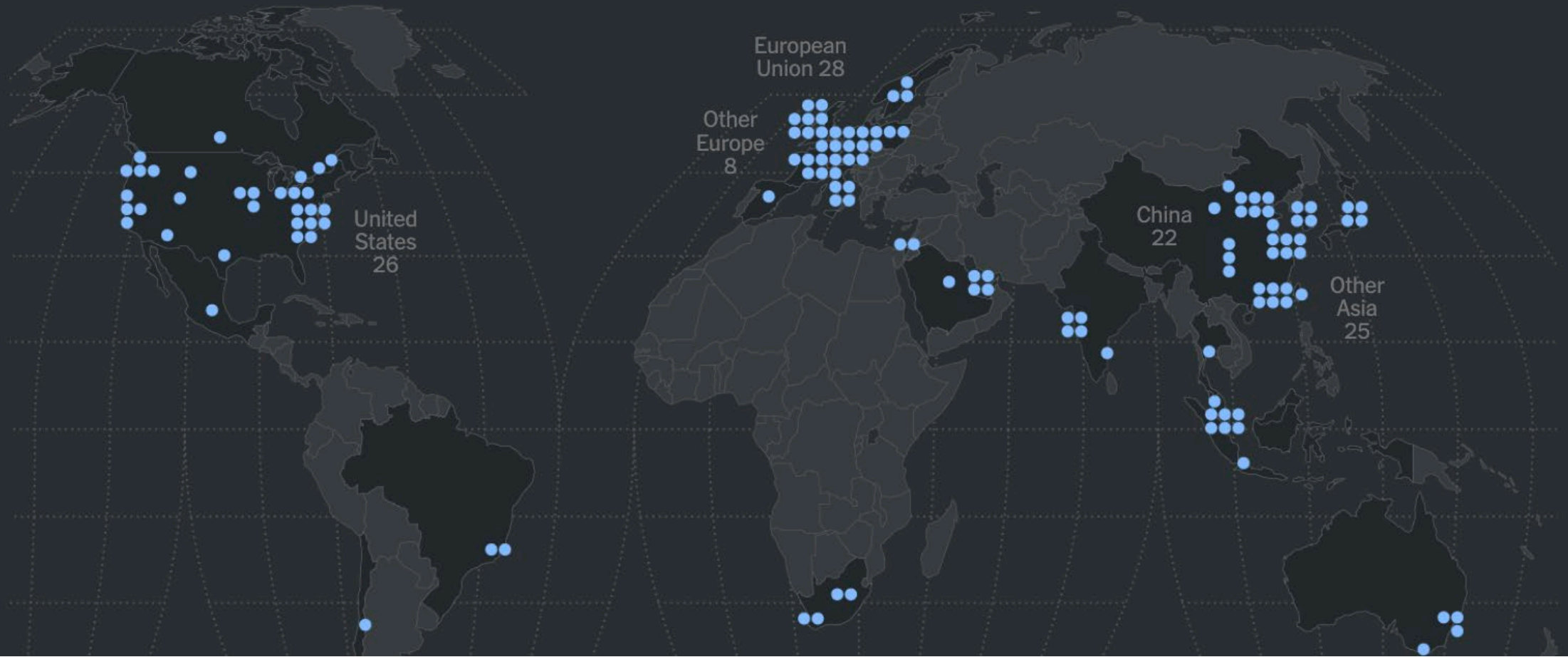
On 1,200 acres of cornfield in Indiana, Amazon is building one of the largest computers ever for work with Anthropic, an artificial intelligence start-up.

The New York Times

The Global A.I. Divide

WHERE A.I. DATA CENTERS ARE LOCATED

Only 32 nations, mostly in the Northern Hemisphere, have A.I.-specialized data centers.



403 Hyperscale data centers and 3318 energy supplier power plants in the US (May 2024 to April 2025)

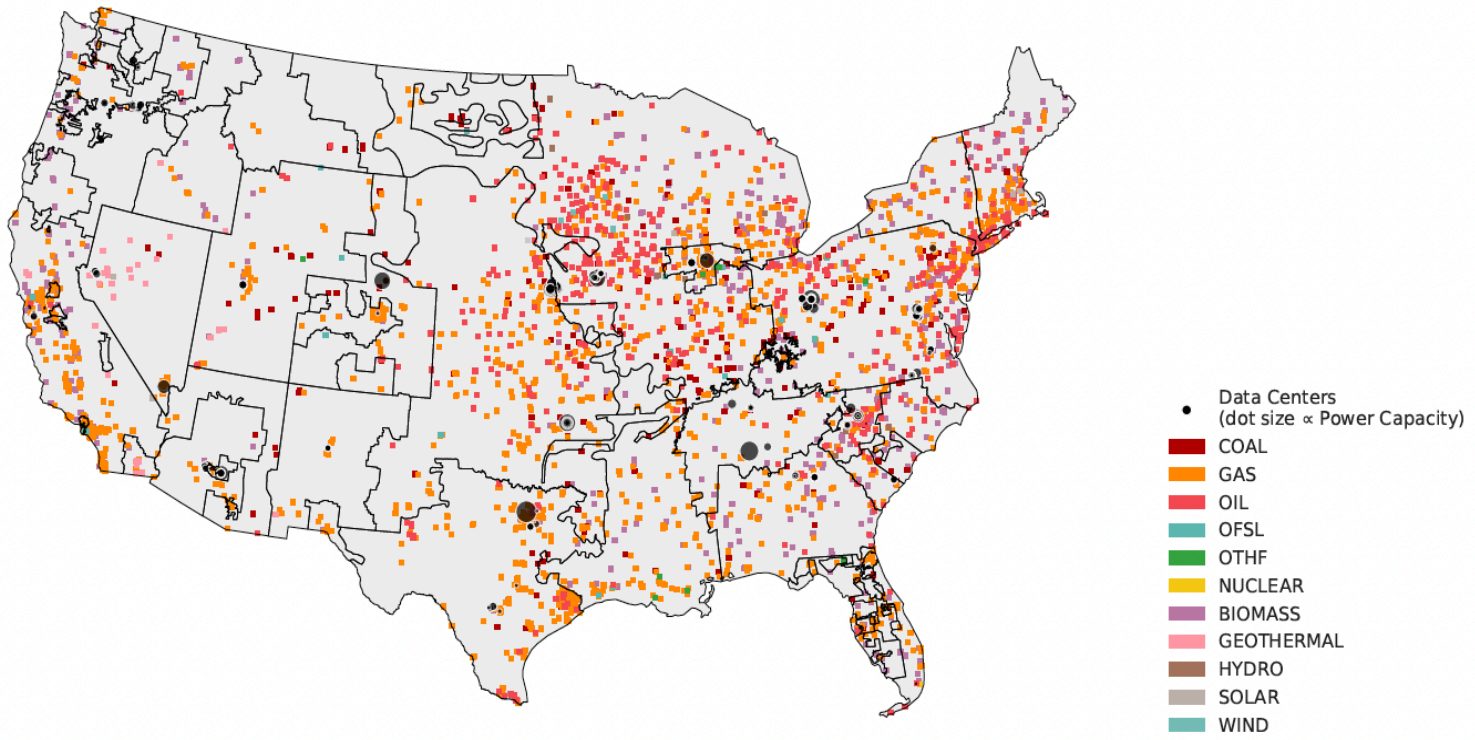


Fig. 1. Geographic distribution of hyperscale data centers and power plants in the contiguous US, overlaid with balancing authority regions. This figure shows the 403 hyperscale data centers and 3,318 operational power plants included in our analysis for the study period from May 2024 to April 2025. The map is displayed at the balancing authority (BA) level, representing regions where electricity supply and demand are managed in real time. The size of each hyperscale data center marker is proportional to its power capacity, while power plants are colored by their primary fuel type.

Scientific questions

1. What are the electricity consumption, sources, and attributable CO2 emissions of those 403 data centers?

2. What is the fuel mix of the power plants supplying electricity to data centers?

3. Which states have the highest CO2 emissions attributable to data centers?

Hint: With a data pipeline that can answer those questions, we make informed decisions, such as: Where should I place a data center? Where should I intervene on the power grid? How can we decarbonize this sector?

Materials and Methods

01

Dataset Compilation and Validation

Our analysis began with compiling and validating a dataset of 403 HDCs.

This was achieved using private data providers, web scraping, and satellite imagery.

02

Power Capacity Estimation

Missing power capacities were estimated using a Gradient Boosted Regression Tree model.

03

Electricity Consumption Calculation

Annual electricity consumption was calculated by multiplying the power power capacity, hours hours per year, and a a utilization rate of 66.3%.

The utilization rate was determined empirically.

04

CO2 Emissions Estimation

Each HDC was assigned to its balancing authority and corresponding power plants.

CO2 emissions were estimated using an energy generation-weighted model and EPA emission factors.

05

Carbon Intensity Computation

Carbon intensity was computed as grams of CO2 emitted per kilowatt-hour consumed.

This computation was performed at various geographic levels.

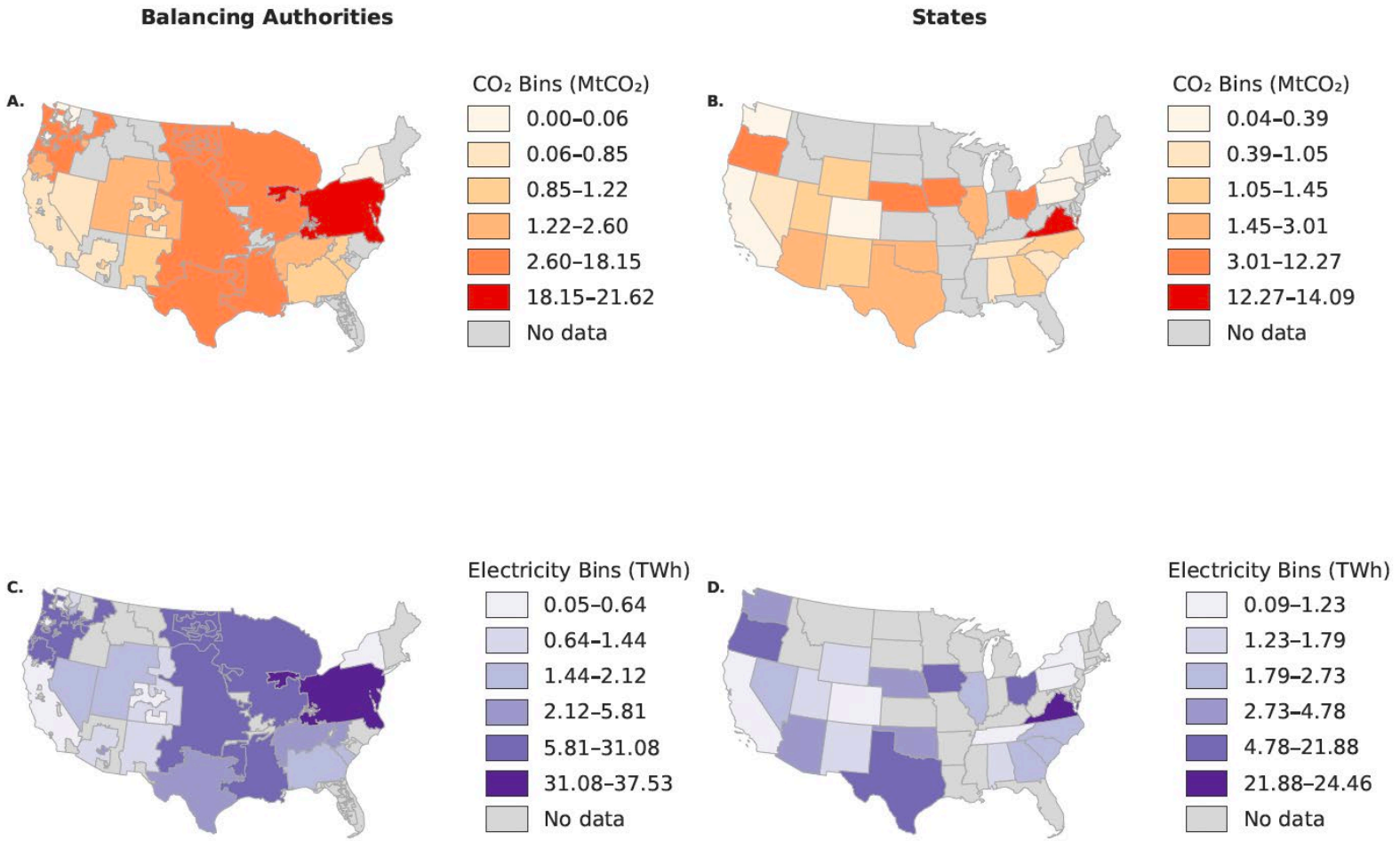


Fig. 2. Hyperscale Data Center electricity consumption and CO_2 emissions. (Left column, A and C) The balancing authority (BA) region in which a hyperscale data center is located determines the mix of power plants that supply its electricity and thus its attributable emissions. See fig.S.4.1 for BA regions and corresponding names. (Right column, B and D) Maps at the state level show electricity consumption and emissions for which the hyperscale data centers within the state are responsible for. Color bins represent percentile-based ranges: 0–20%, 20–40%, 40–60%, 60–80%, 80–99%, and 99–100%.

Power Plants Supplying Electricity to HDCs

Total Power Plants Identified

3,318

Number of HDCs Supplied

403

Fossil Fuel Dependency Dependency

60%

Nuclear Power Contribution

29%

Renewable Energy Share

11.2%

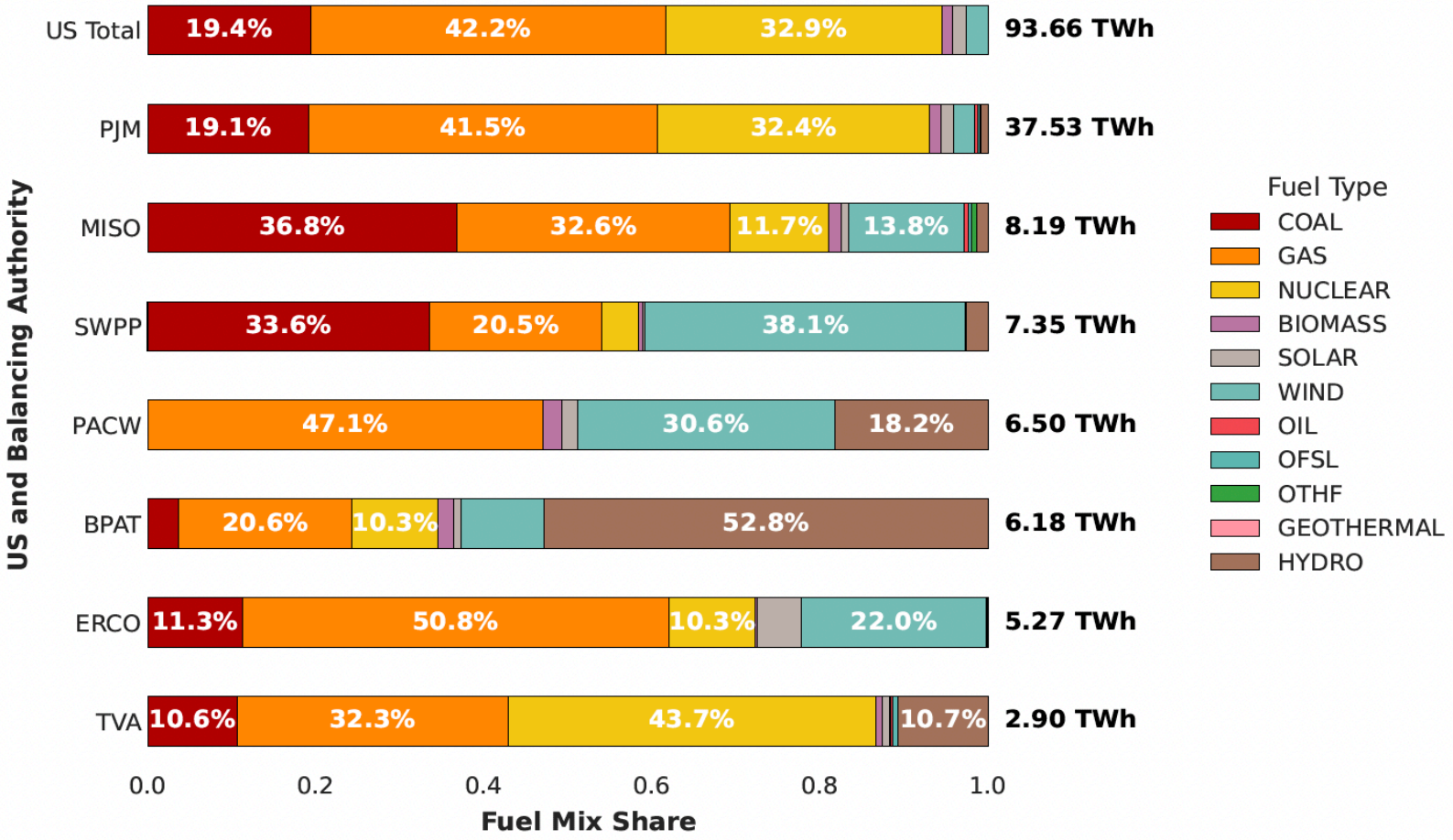


Fig. 4. Fuel mix of power plants supplying electricity for hyperscale US data centers. The top bar represents the distribution of fuel types used by the power plants supplying electricity for hyperscale US data centers in our study. The bottom bars show the largest balancing authorities ranked by aggregated power capacity of hyperscale data centers (shown on the vertical axis), and the amount of electricity produced per fuel type. See fig.S.4.1 for BA regions and corresponding names.

Carbon Emissions Attributable to Hyperscale Data Centers

Total CO2 Emissions from HDCs

52.69M

The total CO2 emissions attributable attributable to the 403 hyperscale hyperscale data centers (HDCs) amounted to 52.69 million metric metric tons.

Proportion of US Carbon Emissions

1.10%

This represents approximately 1.10% 1.10% of the total US carbon emissions from electricity consumption in 2023.

Increase Since 2018

5x

This is more than five times the total total emissions reported for HDCs in HDCs in 2018.

Highest Emissions by State

24.46M

Virginia had the highest CO2 emissions attributable to HDCs, amounting to 24.46 million metric metric tons.

Significant State Contributions

5.82M

Ohio followed with 5.82 million metric tons of CO2 emissions attributable to HDCs.

- 52.69 M represents the annual CO_2 emissions of a major U.S. city or a sizable portion of the U.S. aviation industry.

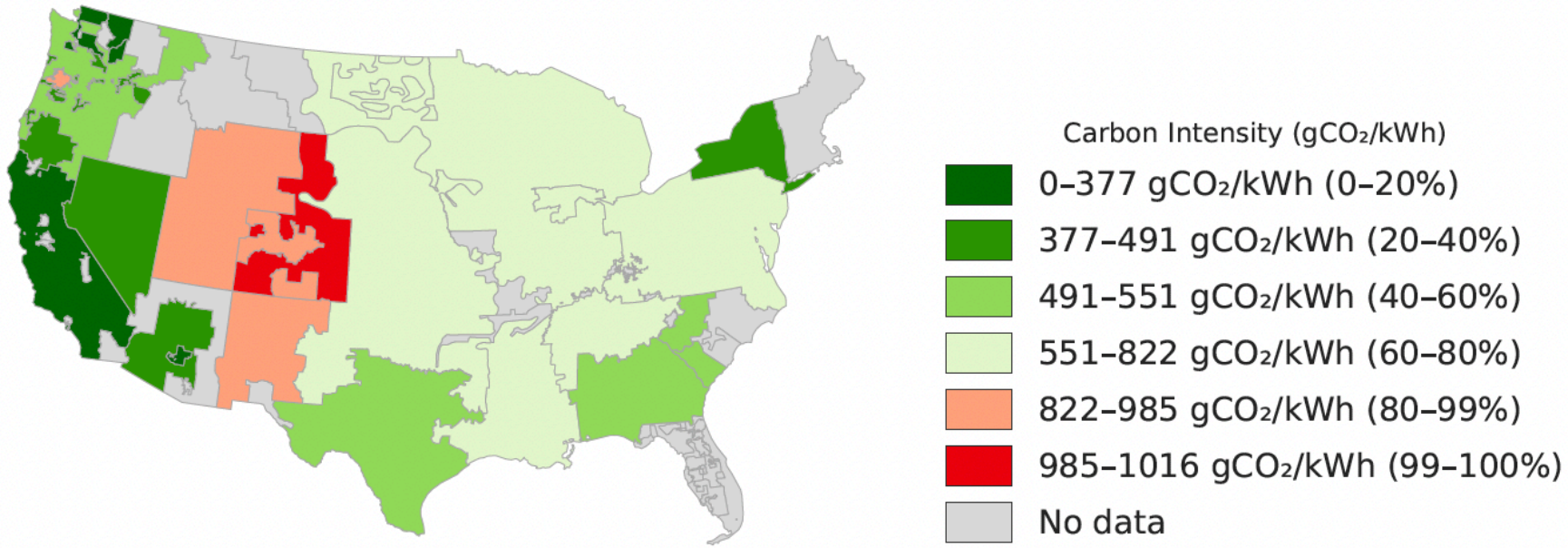


Fig. 3. Carbon intensities of electricity consumption for hyperscale US data centers by balancing authority. Carbon intensity is defined as the amount of carbon dioxide emissions produced per unit of electricity generated, or consumed, and is expressed in units such as grams of CO_2 per kilowatt-hour (gCO_2/kWh) for electricity generation. The figure shows HDCs' carbon intensity for electricity consumption at the balancing authority level, in grams of CO_2 per kWh. Color bins represent percentile-based ranges: 0–20%, 20–40%, 40–60%, 60–80%, 80–99%, and 99–100%.

State	Annual CO2e Emissions *MTons*	Annual Electricity Use *TWh*	Number of Hyperscale Data Center *#*	Hyperscale Data Center Averages per State			
				Electricity Demand *TWh*	CO2e emissions *Mtons*	Electricity Intensity *MW/sq.ft.*	Facility Size *thousand sq.m*
Virginia	14.09	24.46	142	0.17	0.1	153	18
Ohio	5.82	10.1	38	0.27	0.15	174	25
Iowa	5.62	8.19	26	0.32	0.22	136	42
Oregon	5.6	12.72	56	0.23	0.1	173	23
Nebraska	3.22	4.04	8	0.5	0.4	195	51
Texas	2.68	5.27	22	0.24	0.12	140	30
Oklahoma	2.64	3.31	8	0.41	0.33	163	44
Illinois	1.55	2.69	8	0.34	0.19	119	35
Arizona	1.52	3.69	12	0.31	0.13	132	40
Wyoming	1.43	1.41	2	0.7	0.71	794	18

Table 1: **Statistics for the top ten states by CO_2 emissions attributable to hyperscale data center electricity consumption.** The emissions and annual electricity values reflect those attributable to hyperscale data centers located within each state. However, due to the structure of the electricity grid and the role of balancing authorities (BA), these emissions may not align precisely with the physical locations of the data centers, and do not occur necessarily within the state borders, but rather in the BAs borders, see fig.S.4.1 for BA regions and corresponding names.

Science Initiative and Dominici Lab warns of serious health impacts and associated costs from Balico, LLC's proposed gas plant in Pittsylvania County, Virginia

Researchers assessed how fine particulate matter pollution from the proposed 3,500-megawatt gas-fired power plant would harm

NEWS | JUNE 4, 2025

Data center defeated

How Pittsylvania said, 'No, thank you,' to a massive gas-fired power plant and data cent campus.

It's hard to miss the message from Pittsylvania County residents who united against the formerly proposed polluting projects in their community. F displayed big banners and many other signs of protest across the county. (Cornelius Lewis/SELC)

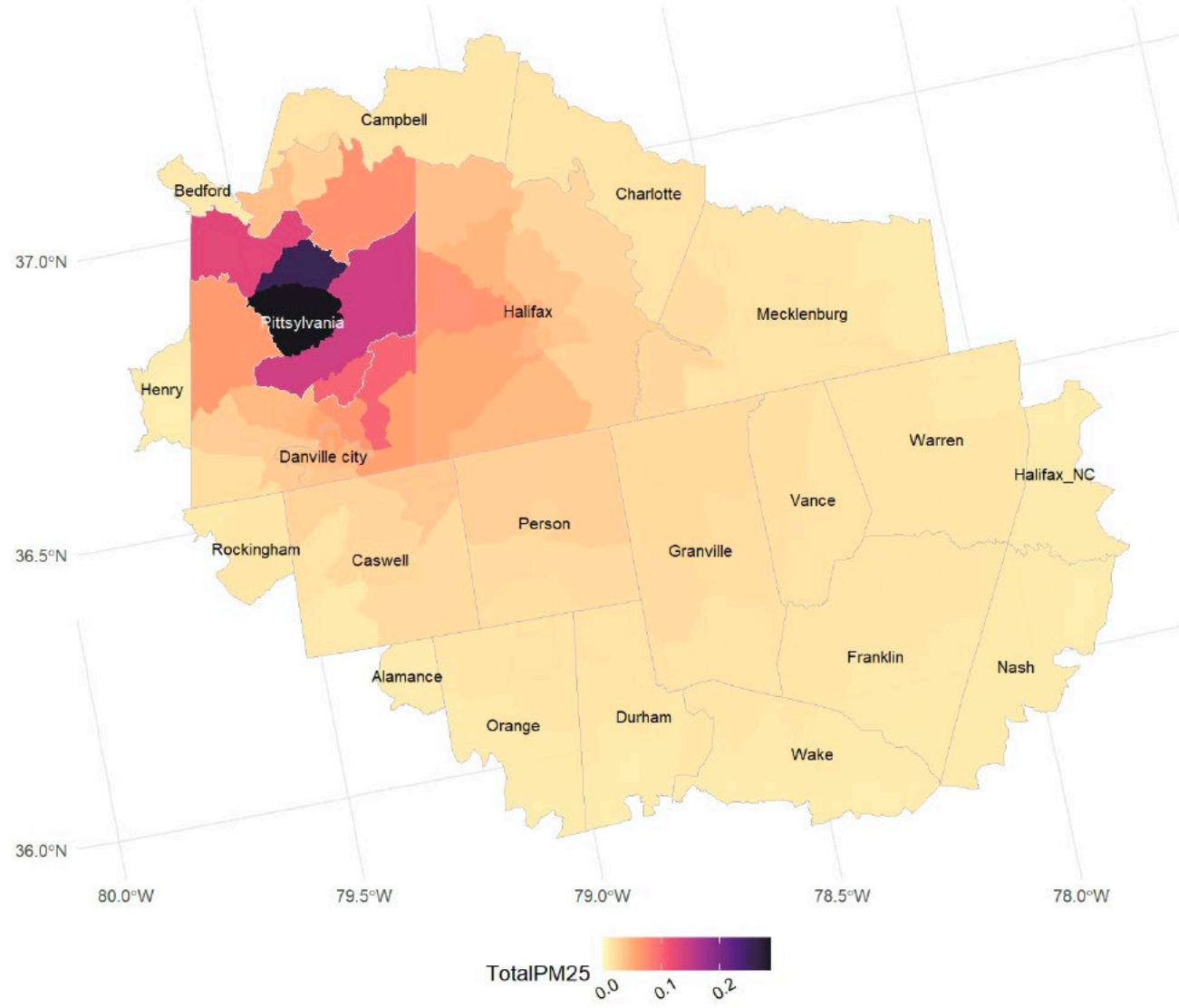


Figure 2. Zoomed-in map of estimated total $PM_{2.5}$ concentrations (μg/m³) from power plant emissions across Census tracts in Virginia and North Carolina with concentrations $\geq$ 0.01 μg/m³, based on InMAP simulations. Census tracts with concentrations $\geq$ 0.1 μg/m³ are enclosed within a white boundary.

Population Exposure and Environmental Justice Concerns

1.28M

Population Experiencing PM2.5 Increase

Model results indicate over 1.28 million people people would experience experience at least a 0.01 0.01 µg/m³ increase in PM2.5.

17,600

Residents in Pittsylvania Pittsylvania County Exposed

More than 17,600 residents in Pittsylvania County are exposed to PM2.5 increases above 0.1 µg/m³.

$49,647 7

Median Household Income Income in Affected Areas Areas

Median household incomes in the most affected Census tracts are are roughly half the state state average ($49,647 vs. vs. $100,268).

18.51%

Poverty Rate in Affected Areas

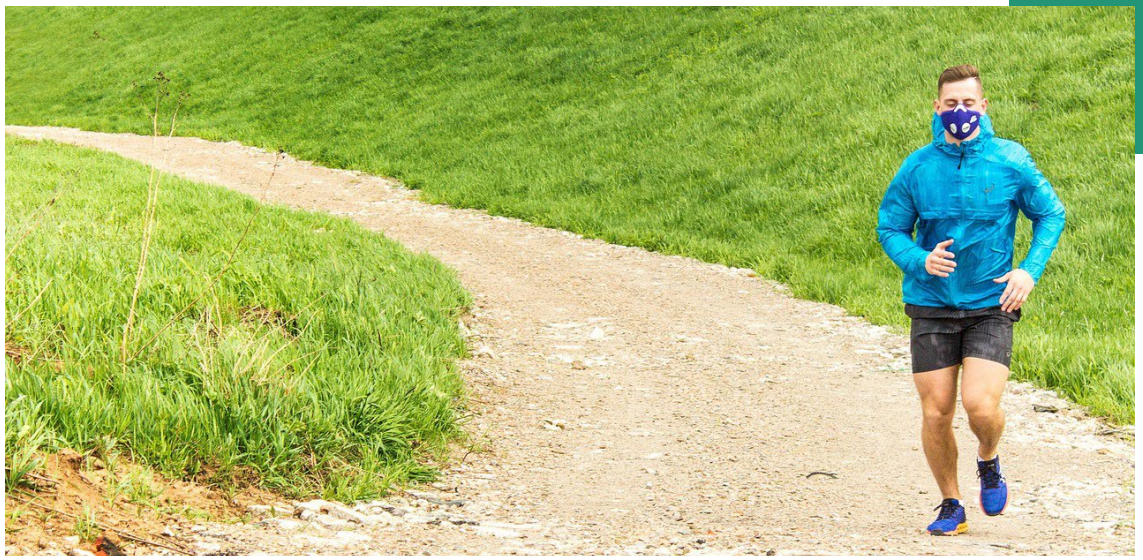
Poverty rates in the most most affected Census tracts are nearly double double Virginia's average average (18.51% vs. 10.16%).

25.49%

Black Residents Overrepresentation

Black residents are overrepresented in the most affected Census tracts (25.49% vs. 18.90%).

Economic Costs of PM2.5 Exposure



$110k

Cost per Ton of PM2.5

Each additional ton of PM2.5 from power plants is plants is estimated to increase healthcare costs by costs by approximately $110,000 due to morbidity morbidity and mortality.

326.53 tons tons

Proposed Plant Emissions Emissions

The proposed plant is expected to emit 326.53 tons of PM2.5 annually.

$31M-$48M

Annual Health Costs

Annual health-related costs are projected to start near $31 million and and rise to $48 million by 2040.

$625M

Cumulative Costs by 2040

Cumulative health-related costs are estimated to reach $625 million by 2040. 2040.

Conclusions

Electricity Consumption

US hyperscale data centers consume approximately 93.66 TWh of electricity annually.
This significant consumption highlights their substantial energy demand.

CO2 Emissions Contribution

These data centers contribute over 52 million metric tons of CO2 emissions annually.
This accounts for about 1.10% of US electricity-related emissions in 2023.

Carbon Intensity

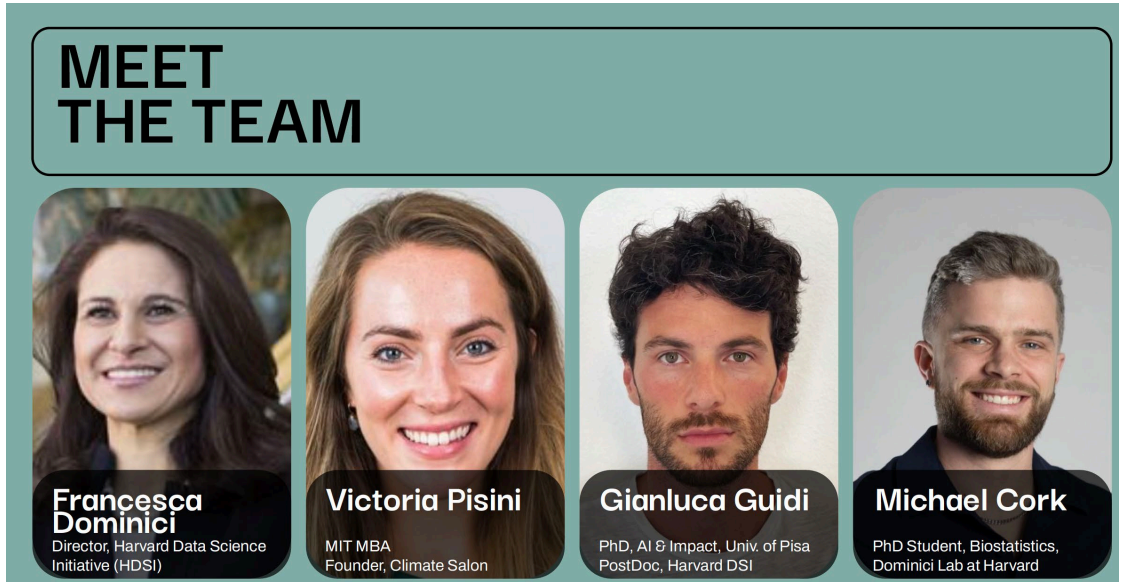
The carbon intensity of hyperscale data centers is centers is 52% higher than the national average. average.
This is due to their location in regions with more carbon-intensive power generation.

Impact of AI and Cloud Computing

The rapid growth of AI and cloud computing is is driving increased demand and environmental environmental impact.
This trend underscores the need for sustainable practices in the sector.

Policy and Monitoring Needs

Data-driven methodologies integrating diverse data diverse data sources and satellite validation are are essential for accurate impact assessment.
Policymakers must use these insights to develop strategies to mitigate the environmental footprint of hyperscale data centers.



Applied Impact Inc.

JUNE 2025

We offer a transparent, science-based impact analysis tool that helps data center developers and communities collaboratively forecast health impacts and healthcare costs, addressing concerns about insufficient data and alternative energy options. Unlike other solutions, our platform uniquely integrates health outcomes into project planning and actively involves all stakeholders to reduce delays and build trust.