A Water–Carbon Sustainability Framework for 24/7 Carbon-Free AI Datacenter Siting in the U.S.

Arif Ahmed
Power System Studies Consultant, PSC
Email: arif.ahmed@example.com

***Abstract*—Artificial intelligence (AI) datacenters are expanding rapidly across the United States and are placing growing pressure on electricity systems and freshwater resources. Most siting frameworks emphasize grid decarbonization but do not systematically account for hydrological risk, even though several low-carbon regions face severe baseline water stress (BWS). This paper develops a reproducible water–carbon sustainability framework that integrates state-level electricity carbon dioxide (CO_2) intensities from the U.S. Environmental Protection Agency's Emissions & Generation Resource Integrated Database (EPA eGRID 2023) and BWS values from the World Resources Institute (WRI) Aqueduct 4.0 Water Risk Atlas into a unified 2025 assessment. Using min–max normalization, a composite Water–Carbon Index (WCI), median-based quadrants, and regional aggregation, the framework characterizes joint carbon–water conditions across all U.S. states. Results reveal sharp stratification: only a few states, notably Maine and Washington, offer simultaneously low emissions and low water stress, while many solar-rich regions remain hydrologically constrained. Regional heatmaps and rankings highlight consistent patterns that are relevant for 24/7 carbon-free energy (CFE) planning. The framework provides a transparent, fully replicable basis for water-aware AI datacenter siting and can be extended to dynamic or technology-specific studies.**

***Index Terms*—Datacenters, Water Stress, Carbon Intensity, Siting Optimization, 24/7 CFE, Sustainability**

I. Introduction

The rapid growth of artificial intelligence (AI) workloads has accelerated the construction of large-scale datacenters across the United States. Recent assessments from the International Energy Agency (IEA) project that global datacenter electricity demand could exceed 900 TWh per year by 2030, with AI workloads a major driver [1], [2]. A single hyperscale facility can also withdraw millions of liters of freshwater annually for evaporative cooling [3], particularly in hot or arid climates. These combined pressures highlight the need for siting frameworks that explicitly account for both power-sector emissions and hydrological resilience.

Most current industry strategies prioritize carbon outcomes, especially in support of 24/7 carbon-free energy (CFE) commitments from major cloud providers such as Google and IBM [4], [5]. Under these commitments, datacenters seek to match hourly electricity use with carbon-free supply on the same grid. In practice, site selection is often guided by average grid CO_2 intensity or projected clean-energy shares. However, water scarcity is intensifying across large parts of the western United States [6], and several states with high solar or wind penetration also face severe baseline water stress (BWS). This decoupling means that a grid with favorable carbon characteristics may still be poorly suited for water-cooled AI infrastructure without substantial mitigation.

Recent studies have examined joint water and carbon impacts for digital infrastructure and power systems [7], [8], [2]. Some quantify water withdrawal factors for cooling and generation technologies, while others develop composite indices for individual basins or regions. However, much of this work remains highly localized, which limits its generalizability across the United States. Other studies rely on metrics such as basin-level hydrological simulations, technology-specific cooling-water intensities, or proprietary datacenter water-use factors [7], [8], [2], which cannot be reconstructed or validated using fully public datasets. In addition, many of these approaches do not align well with the broader multi-regional screening needs reported by hyperscale datacenter operators, who emphasize the importance of transparent and auditable indicators for early-stage siting decisions [4], [3].

To address these gaps, this paper develops a transparent, state-level water–carbon sustainability framework for AI datacenter siting in the U.S. The framework relies solely on open datasets: EPA eGRID 2023 electricity CO_2 intensities [9] and WRI Aqueduct 4.0 baseline water stress (BWS) [6]. A static 2025 snapshot is constructed and processed through simple, reproducible steps: normalization, a composite Water–Carbon Index (WCI), median-based quadrant classification, and regional aggregation. The resulting maps, rankings, and summaries form a national screening layer that integrates emissions and hydrological conditions, complementing (rather than replacing) detailed project-level economic and engineering studies.

The main contributions of this paper are threefold:

- A reproducible water–carbon sustainability index constructed entirely from public data, with an explicit weighting structure that emphasizes hydrological risk while retaining a clear physical interpretation.
- A set of state- and region-level visualizations and rankings (scatter plots, heatmaps, composite bars, and tables) that identify where joint water–carbon performance is most favorable and where siting would require significant mitigation.
- A compact methodology that practitioners can readily extend to dynamic, hourly, or technology-specific siting analyses for 24/7 CFE datacenters, including integration with cooling-system design and long-run transmission planning.

979-8-3315-8138-1

II. Methodology

The methodology develops a transparent, state-level framework for jointly evaluating electricity-sector carbon intensity and hydrological scarcity across the United States. The approach integrates two public datasets, processes them into a consistent mid-2020s snapshot, normalizes the underlying indicators, constructs a composite Water–Carbon Index (WCI), and applies both median-based quadrant classification and regional aggregation to reveal broader spatial patterns. The subsections that follow describe the data inputs, processing pipeline, normalization procedure, index formulation, categorical grouping, and the visualization and ranking outputs that support the analysis.

A. Data Sources

Two public datasets form the basis of the analysis:

- **EPA eGRID 2023** [9]: state-level average electricity CO_2 intensity (kg/MWh), derived from verified plant-level generation, fuel mix, and emissions data.
- **WRI Aqueduct 4.0** [6]: baseline water stress (BWS), a unitless 0–5 indicator measuring the ratio of total withdrawals to renewable freshwater availability.

The analysis covers all 50 U.S. states for state-level results. Washington, D.C. is excluded because it does not operate a generating balancing authority. For regional analyses, Alaska is omitted because it does not belong to any of the four U.S. Census regions used in this study. This separation keeps the regional medians consistent and avoids misinterpretation when comparing multi-state aggregates. For each state, a representative CO_2 intensity and baseline water-stress value is obtained by aggregating the underlying datasets to the state level.

B. Static 2025 Snapshot and Processing Pipeline

The analysis uses the most recent publicly available datasets (EPA eGRID 2023 for state-level CO_2 intensity and WRI Aqueduct 4.0 for baseline water stress) to construct a static snapshot that is representative of mid-2020s conditions, centered conceptually on the year 2025. The processing pipeline consists of three steps:

1) Construct state-level CO_2 and BWS datasets from eGRID 2023 and Aqueduct 4.0.
2) Merge the datasets and compute normalized indicators and a composite index.
3) Apply quadrant classification, perform regional aggregation, and generate visualizations and tables.

All processing is implemented in Python using `pandas`, `NumPy`, and `matplotlib`. Every step is scripted and fully reproducible.

C. Normalization of Metrics

Let C_i denote the state-level CO_2 intensity (kg/MWh) for state i, and let W_i denote the corresponding baseline water stress (BWS) score. To place these quantities on a common scale, min–max normalization is applied to each metric:

$$\hat{X}_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}, \tag{1}$$

where X_i is either C_i or W_i, and $X_{\min}$ and $X_{\max}$ are the minimum and maximum values across all states for that metric. This produces normalized indicators $\hat{C}_i$ and $\hat{W}_i$ bounded in $[0, 1]$, with 0 representing the most favorable observed condition and 1 the most constrained. Min–max normalization is used instead of z-scores to retain this intuitive "0 to 1" interpretation and to avoid negative values when combining indicators in the composite index.

D. Composite Water–Carbon Index (WCI)

Since CO_2 intensity and baseline water stress capture different aspects of siting suitability, a single metric is useful for summarizing their combined effect. To provide this integrated measure, a composite Water–Carbon Index (WCI) is defined as

$$\mathrm{WCI}_i = w_C \hat{C}_i + w_W \hat{W}_i, \tag{2}$$

where $\hat{C}_i$ and $\hat{W}_i$ are the normalized carbon and water indicators. The weights are set to $w_C = 0.35$ and $w_W = 0.65$, reflecting the greater operational sensitivity of water-cooled datacenters in basins with high water stress. Lower values of WCI indicate more favorable combined conditions.

E. Median-Based Quadrant Classification

To complement the continuous WCI metric with a simple categorical view, states are grouped according to whether their CO_2 intensity and baseline water stress fall above or below national median values. Let C_{med} and W_{med} denote the medians of C_i and W_i across all states. Each state is assigned to one of four quadrants:

- Q1: $C_i \leq C_{\mathrm{med}}$ and $W_i \leq W_{\mathrm{med}}$ (low carbon, low stress),
- Q2: $C_i \leq C_{\mathrm{med}}$ and $W_i > W_{\mathrm{med}}$ (low carbon, high stress),
- Q3: $C_i > C_{\mathrm{med}}$ and $W_i > W_{\mathrm{med}}$ (high carbon, high stress),
- Q4: $C_i > C_{\mathrm{med}}$ and $W_i \leq W_{\mathrm{med}}$ (high carbon, low stress).

Using medians rather than ad-hoc thresholds ensures that each quadrant contains a balanced and interpretable set of states, making cross-group comparisons more robust and informative.

F. Regional Aggregation

To reveal broader spatial patterns, states are grouped into the four U.S. Census regions. For each region, median values of the normalized indicators $\hat{C}_i$, $\hat{W}_i$, and the composite index WCI_i are computed. Using medians provides a stable summary that is less sensitive to extreme state-level values and yields a clearer picture of regional tendencies. These aggregated metrics help identify large-scale differences that matter when datacenter operators evaluate multi-state portfolios or plan clusters of interconnected cloud regions.

979-8-3315-8138-1

G. Visualization and Ranking

The analytical pipeline produces a standardized set of visual summaries, including (i) a CO_2–BWS scatter plot with quadrant boundaries, (ii) a state-level heatmap of normalized indicators, (iii) composite WCI bar charts, (iv) regional median summaries, and (v) tables ranking the ten best and ten worst states by WCI. Together, these outputs provide a consistent and interpretable view of the joint water–carbon landscape across the United States and form the basis of the Results section.

III. Results and Discussion

A. Joint Distribution of CO_2 and Water Stress

Fig. 1 shows the joint distribution of state-level CO_2 intensity and baseline water stress, with median thresholds defining the four quadrants described in Section II. The national medians are 347 kg/MWh for CO_2 intensity and 1.86 for BWS.

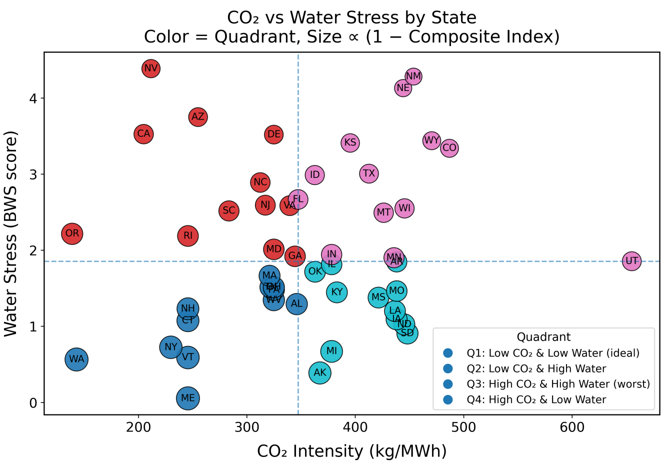

Fig. 1: State-level CO_2 intensity vs. baseline water stress with median-based quadrants.

Using the national medians (347 kg/MWh for CO_2 intensity and 1.86 for BWS), the 50 states are grouped into four quadrants. Because one state lies exactly on the median boundary, the effective grouping reflects 49 states falling strictly above or below the thresholds. This yields four balanced and interpretable categories that reveal clear patterns in the joint distribution of carbon intensity and water stress.

Q1 (Low CO_2, Low Water Stress) contains twelve states, including Maine (ME), Washington (WA), Vermont (VT), New York (NY), Connecticut (CT), and New Hampshire (NH). These states pair comparatively clean electricity mixes with secure hydrological conditions, representing the most favorable baseline environments for water-aware AI datacenter siting.

Q2 (Low CO_2, High Water Stress) also includes twelve states, such as Oregon (OR), Rhode Island (RI), Georgia (GA), Maryland (MD), North Carolina (NC), California (CA), Arizona (AZ), and Nevada (NV). Although their grids are relatively low-carbon, these states face significant water scarcity, making cooling-system selection a primary design constraint for large-scale AI deployments.

Q3 (High CO_2, High Water Stress) is the largest quadrant with thirteen states, including Indiana (IN), Minnesota (MN), Florida (FL), Wisconsin (WI), Montana (MT), Utah (UT), Texas (TX), Kansas (KS), Colorado (CO), and New Mexico (NM). These states exhibit the least favorable conditions for water-cooled datacenters, combining both elevated carbon intensity and severe hydrological stress.

Q4 (High CO_2, Low Water Stress) contains twelve states, including Alaska (AK), Michigan (MI), South Dakota (SD), North Dakota (ND), Iowa (IA), Louisiana (LA), Kentucky (KY), and Mississippi (MS). While these states benefit from comparatively secure water resources, their electricity systems remain heavily fossil-dependent; ongoing decarbonization policies may improve their future suitability.

Overall, the quadrant classification highlights a pronounced national stratification: water stress is the dominant factor shaping regional suitability, while carbon intensity introduces additional differentiation among otherwise hydrologically similar states.

B. State-Level Heatmap

The state-level heatmap provides a compact visual comparison of the two underlying drivers of the Water–Carbon Index. By normalizing both indicators to a common 0–1 scale, the figure allows direct visual assessment of how each state performs relative to all others and highlights the structural differences between carbon intensity and hydrological stress across the U.S. grid. Ordering the states from best to worst composite performance further clarifies how these two factors jointly shape the national landscape for 24/7 CFE siting.

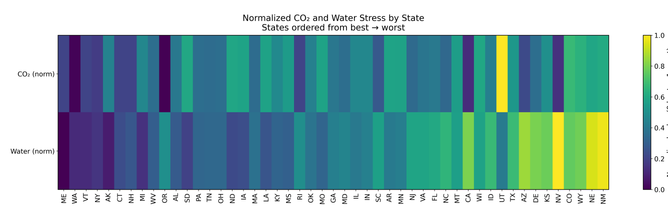

Fig. 2: Normalized CO_2 intensity and baseline water stress for all states, ordered from lowest to highest WCI. Both indicators are scaled to $[0, 1]$ for comparability across metrics.

Figure 2 highlights the joint distribution of the normalized indicators across states, ordered from the most to the least favorable composite performance. A clear asymmetry emerges between the two metrics: while normalized CO_2 values span a relatively narrow band, baseline water stress varies substantially across the full 0–1 range. This pattern reinforces the dominant role of hydrological conditions in shaping the overall WCI and explains the sharper separation of states observed in the ranking tables.

The left portion of the heatmap, corresponding to the best-performing states by WCI, shows consistently low water-stress values with only modest variation in CO_2 intensity. As the ordering progresses toward higher WCI values, water stress increases sharply and becomes the primary contributor to deterioration in composite performance. Several states

979-8-3315-8138-1

near the right end of the scale exhibit simultaneously high water stress and moderate-to-high carbon intensity, indicating combined constraints that are unfavorable for water-cooled AI infrastructure.

Overall, the heatmap reveals that while state-level carbon intensity does influence composite performance, baseline water stress is the more decisive factor driving variation across the United States.

C. Composite Index and Rankings

The composite WCI provides a single, comparable measure of joint carbon and water conditions across the United States. Figure 3 displays the distribution of WCI values for all states, ordered from best to worst. The range is wide: several states cluster near the lower bound (0.05–0.20), while the least favorable states exceed 0.75. This spread reflects the highly uneven geographical distribution of hydrological stress, which strongly shapes the overall index.

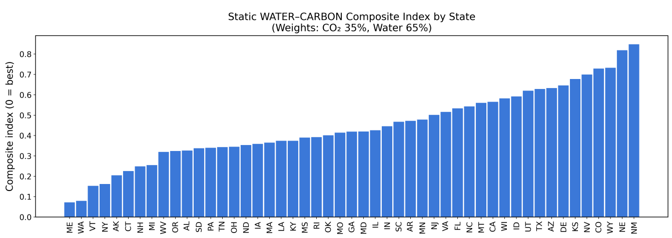

Fig. 3: Composite Water–Carbon Index for all states. Lower values indicate more favorable joint conditions.

The top-performing states (Table I) share a common structural feature: secure water availability. Maine (ME), Washington (WA), and Vermont (VT) achieve the lowest WCI values despite differing generation mixes. New York (NY) and Alaska (AK) also rank favorably; although AK has moderately higher CO_2 intensity, its low hydrological stress offsets this effect. The relative stability of water resources across these states places them at a strong advantage for water-cooled datacenter deployment.

TABLE I: Top 10 Most Sustainable States (Lowest WCI)

State	CO_2 (kg/MWh)	BWS	WCI
ME	245.56	0.05	0.07
WA	142.77	0.56	0.08
VT	245.56	0.59	0.15
NY	229.79	0.73	0.16
AK	367.33	0.39	0.21
CT	245.56	1.08	0.23
NH	245.56	1.23	0.25
MI	378.12	0.67	0.25
WV	324.81	1.35	0.32
OR	138.72	2.22	0.32

The lowest-performing states (Table II) are dominated by the Southwest, Mountain West, and portions of the Great Plains. These states typically exhibit both elevated CO_2 intensity and severe water scarcity. New Mexico (NM), Nebraska (NE), Wyoming (WY), and Colorado (CO) all exceed WCI 0.70. Even states with relatively low CO_2 intensity, such as Nevada (NV) and Arizona (AZ), rank poorly because water stress strongly outweighs their carbon advantage.

TABLE II: Bottom 10 Least Sustainable States (Highest WCI)

State	CO_2 (kg/MWh)	BWS	WCI
UT	655.12	1.86	0.62
TX	412.63	3.01	0.63
AZ	254.90	3.75	0.63
DE	324.81	3.52	0.65
KS	395.32	3.41	0.68
NV	211.46	4.39	0.70
CO	486.86	3.34	0.73
WY	470.52	3.44	0.73
NE	444.12	4.13	0.82
NM	453.70	4.28	0.85

D. Regional Patterns

For broader spatial analysis, states are grouped according to the U.S. Census Bureau regional structure. The four regions used in this study are:

- **Northeast:** ME, NH, VT, MA, RI, CT, NY, NJ, PA
- **Midwest:** OH, IN, IL, MI, WI, MN, IA, MO, ND, SD, NE, KS
- **South:** DE, MD, DC, VA, WV, NC, SC, GA, FL, KY, TN, MS, AL, OK, TX, AR, LA
- **West:** ID, MT, WY, CO, NM, AZ, UT, NV, WA, OR, CA

For each region, median values of the normalized indicators $\hat{C}_i$ and $\hat{W}_i$, together with the composite Water–Carbon Index WCI_i, are computed. Medians provide a stable regional characterization that limits the influence of extreme state-level values and supports multi-state datacenter siting decisions.

Figure 4 shows the median normalized CO_2 and water-stress levels for each region, while Figure 5 reports the resulting regional WCI values.

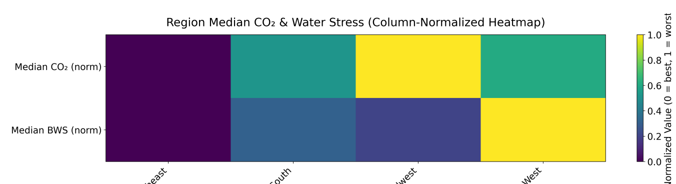

Fig. 4: Regional medians of normalized CO_2 intensity and baseline water stress.

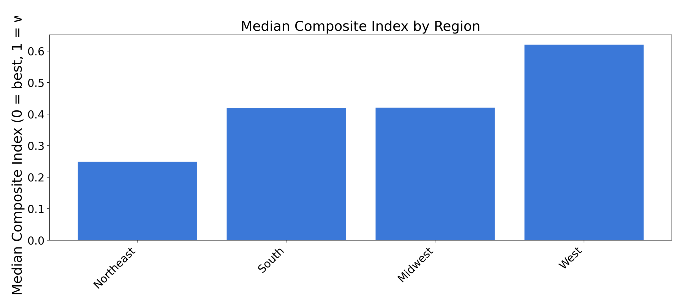

Fig. 5: Regional composite WCI based on median state performance.

The Northeast emerges as the most favorable region overall, combining low carbon intensity with generally secure hydrol-

979-8-3315-8138-1

ogy. The Midwest and South occupy intermediate positions: both regions show moderate composite values, reflecting a balance between relatively stable water availability and the presence of legacy thermal generation in several states.

The West region ranks as the least favorable on a median basis. This result is driven not by the hydro-rich Pacific Northwest, which performs well individually, but by the high water-stress conditions in the more arid parts of the West (for example, NV, AZ, UT, and NM). These states raise the region's composite WCI substantially and offset the strong performance of Washington, Oregon, and northern California.

Overall, these regional contrasts highlight that while national decarbonization efforts continue to reduce carbon intensity across much of the United States, *water availability remains the dominant spatial constraint*. Hydrological scarcity is highly localized and has an outsized influence on regional suitability for 24/7 CFE-aligned datacenter siting.

E. Implications for 24/7 CFE Siting

The Water and Carbon Index provides a practical way to screen datacenter locations by combining water availability and carbon performance. Regions with low index values, such as the Northeast and parts of the Midwest, offer favorable conditions for round the clock clean energy strategies due to their moderate carbon intensity and comparatively secure hydrology.

In contrast, several states in the West face significant water limitations. Even though carbon performance in the region continues to improve with the growth of solar resources, the lack of water remains a firm constraint on new datacenter development. In these areas, new facilities may need to rely on dry cooling, hybrid cooling, reclaimed water, or careful site selection to avoid high stress watersheds. The South also exhibits elevated water stress in many states, although the range of carbon intensity is more varied.

These findings highlight the importance of incorporating water considerations directly into datacenter planning. While carbon intensity is falling across much of the country, water scarcity remains highly localized and uneven. A combined view of water and carbon helps identify suitable locations, avoid long term hydrological risks, and select cooling technologies that support sustainable growth in digital infrastructure.

IV. Conclusion

This paper introduced a transparent framework for evaluating water and carbon sustainability for AI datacenter siting in the United States. Publicly available datasets from EPA eGRID 2023 and WRI Aqueduct 4.0 were combined to construct a static snapshot for 2025. The analysis applied consistent preprocessing steps, including normalization, formation of a combined Water and Carbon Index, quadrant classification of joint conditions, and regional aggregation across the U.S. Census structure.

The results indicate that only a small group of states, led by Maine and Washington, combine low carbon intensity with low water stress. Many solar-rich states remain located in highly water-stressed basins, while several hydrologically secure states continue to depend on fossil generation. The regional aggregation reinforces these patterns. The Northeast emerges as the most favorable region due to its low carbon intensity and comparatively secure hydrology, while the West contains many of the most water-constrained areas in the country, which elevates its overall WCI despite the strong individual performance of the Pacific Northwest.

Future extensions of this work include combining the static results with hourly emissions and clean-energy profiles from forward-looking datasets such as NREL Cambium, incorporating technology-specific cooling water requirements, applying climate-adjusted hydrological projections, and integrating economic metrics such as marginal abatement costs or transmission congestion. These additions would enable a more comprehensive multi-criteria siting toolkit to support the next generation of AI infrastructure.

References

[1] I. E. Agency, "Data centres and data transmission networks – tracking clean energy progress," 2024. [Online]. Available: https://www.iea.org/reports/data-centres-and-data-transmission-networks

[2] P. Henderson *et al.*, "Green ai: Environmental costs and policy implications of artificial intelligence," *Communications of the ACM*, 2023.

[3] Microsoft, "Environmental sustainability report 2023," 2023. [Online]. Available: https://www.microsoft.com/en-us/sustainability

[4] Google, "24/7 carbon-free energy by 2030: Methodology and progress," 2021. [Online]. Available: https://www.gstatic.com/gumdrop/sustainability/247-carbon-free-energy.pdf

[5] IBM, "Sustainable data center operations: Ibm 2023 progress report," 2023. [Online]. Available: https://www.ibm.com/sustainability

[6] World Resources Institute, "Aqueduct 4.0: Water risk atlas," 2023. [Online]. Available: https://www.wri.org/aqueduct

[7] Y. Liu *et al.*, "Siting data centers for sustainable ai: Integrating water and carbon footprints," *Nature Sustainability*, 2023.

[8] M. Ashby and R. Zhou, "The dual impact of cooling and energy mix on datacenter environmental performance," *Applied Energy*, 2022.

[9] U.S. Environmental Protection Agency, "egrid 2023 technical guide and data," 2023. [Online]. Available: https://www.epa.gov/egrid

979-8-3315-8138-1