

Lifecycle Intelligence for Carbon Energy Water Footprint Optimization in Artificial Intelligence Systems

Kuldeep Singh Kaswan

Galgotias University

Avadhesh Kumar

Galgotias University

Praveena Rajendra

Taylor's University

Ramdas Khomane

`ramdas.khomane@sitnagpur.siu.edu.in`

Symbiosis International University

Article

Keywords: sustainable AI, lifecycle inventory, energy accounting, water consumption, carbon intensity, shared infrastructure, uncertainty propagation, multi-objective routing

Posted Date: October 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11295417/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Lifecycle Intelligence for Carbon Energy Water Footprint Optimization in Artificial Intelligence Systems

Kuldeep Singh Kaswan¹, Avadhesh Kumar², Praveena Rajendra³, Ramdas Khomane⁴

¹School of Basic Sciences, Galgotias University, Greater Noida, India - 201310, UP, India

²School of Computer Science and Engineering, Galgotias University, Greater Noida, India - 201310, UP, India

³ Department of Mechanical Engineering, Taylor's University, Malaysia

⁴Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune-440008, India

Kuldeep Singh Kaswan: kaswankuldeep@gmail.com, ORCID: 0000-0003-0876-0330

Avadhesh Kumar: pvcdrak@galgotiasuniversity.edu.in, ORCID: 0000-0002-9469-9611

Praveena Rajendra: praveenarajendra.jnu07@gmail.com, ORCID: 0000-0001-7473-3598

Ramdas Khomane: ramdas.khomane@sitnagpur.siu.edu.in, ORCID: 0000-0002-8806-7351

Corresponding Author: *Ramdas Khomane - ramdas.khomane@sitnagpur.siu.edu.in, *Kuldeep Singh Kaswan - kaswankuldeep@gmail.com

Abstract

The environmental comparison of artificial intelligence (AI) systems is context-dependent regarding lifecycle boundaries, the sharing and allocation of infrastructure, the accounting of electricity use and the accounting of water use. A low operational carbon value can coexist with substantial upstream water consumption or a large unallocated hardware burden. This paper develops a lifecycle framework in which primitive resource events are assigned once to workload stages and translated into energy, greenhouse-gas, and water inventories through explicit, versioned factors. The framework separates physical water consumption from scarcity characterization, operational electricity from upstream energy carriers, and attributional reporting from consequential scheduling. We derive conservation under allocation, an exact temporal aggregation bias identity, an identifiability condition for incomplete inventories, a shared-infrastructure allocation model, and a service-volume break-even condition. A constrained minimax routing extension evaluates conflicting operational objectives without collapsing physical units into an unexplained score. A fully specified synthetic study covers three hypothetical regions, 168 hourly intervals, 24 demand profiles, four routing policies, and 10,000 uncertainty draws. For the base case of 3.36 million requests, carbon-only routing reduces serving carbon by approximately 50.5% relative to the reference, but increases serving water consumption by approximately 124.3%. Once fixed lifecycle burdens are included, the corresponding changes are approximately -4.2% and +9.6%. Through this specific situation, the minimax policy is redefined, so no enhancement over the minimax reference is claimed. This study does not exhibit accounting accuracy but rather accounting consistency and conditional decision behavior. A new specification is obtained that makes it easy to implement and gives specification points for the field validation work which is required for future work.

Keywords: sustainable AI; lifecycle inventory; energy accounting; water consumption; carbon intensity; shared infrastructure; uncertainty propagation; multi-objective routing.

1. INTRODUCTION

AI services foster the linkage among models and legacy physical infrastructure. AI services link to ongoing physical infrastructures. This means burdens in data acquisition, experimentation, training, adaptation, inference and storage as well as network transfer, manufacturing and retirement – all at different locations and at different times. Ignoring part of the system just results in a measurement of accelerator activity only, and, in multiplying any annual's electricity total with a single environmental factor, some temporal dependence is hidden. The early research and training for NLP established and focused on the development spend [1] and later on hardware and location decisions [2]. When repeated inference is a significant part of a model's operational history, this motivation continues to be salient.

More recent primary evidence is out from carbon only. Host energy, idle capacity and facility overhead are included in Google's production serving study [6]. Astrobotic's lifecycle disclosure addresses upstream impacts, as well as uncertainty of accelerators inventories [7]. A UNU INWEH report from 2026 [17] concentrates on energy related carbon, water and land burdens. This leads to a narrower question here, namely how an inventory of workloads can be maintained if their events, ownership boundaries and decision purposes are different?

1.1 Background

Energy is an inventory item, CO₂ equivalents are a characterization of GHG emissions, and 'water' refers to withdrawal, use, or a weight on the scarcity of water. These quantities are not the same but related. The use of electricity is different based on the computing and cooling setup. In addition to electricity supply and embodied emissions, carbon relies on. Water can rely on direct cooling, upon generation of electric power upstream, and upon processes in the value chains. But a common output unit is not sufficient to ensure that mismatched boundaries are comparable.

1.2 Research Problems

Produce an estimated stage-resolved footprint given heterogeneous records for a model and its descendants, that (i) reconcile to physical totals, (ii) carries the provenance of records for allocations and factors, (iii) reports uncertainty instead of silently assuming that missing values are zero and (iv), supports decisions for routing simultaneously under two (or more) competitive resource constraints. The main challenge is a meter observes equipment consumption and desired result is part of a model service. To map one to the other, there must be a declared allocation rule.

1.3 Research Gaps

The gap addressed in **table 1** is operational integration, which connects resource measurement, lifecycle inheritance, contextual factors, and uncertainty in a common event ledger. For many of the component methods, existing carbon trackers (e.g., [3–6]), water-estimation methods (e.g., [7–9]), lifecycle studies, and regional scenario analyses exist. This study does not claim that joint accounting, proportional allocation, minimax optimization, or Monte Carlo propagation are individually new. Its proposed contribution is the explicit integration with conservation checks and the honest separation of verified quantities from scenario assumptions. Table 1 presents the research questions tied to specific analytical or computational evidence. None requires treating synthetic requests as observed production traffic.

Table 1. Research Gaps

Question	Objective	Evidence produced
RQ1: Can shared burdens be allocated without duplication?	Define a conservation-preserving event map	Proposition 1; allocation example; reconciliation invariant
RQ2: When does averaging distort a footprint?	Derive temporal and location-related errors	Covariance identity; water-factor equation
RQ3: Can reducing carbon increase water consumption?	Compare feasible policies with identical demand	Four policies; 24 profiles; Figs. 1–4
RQ4: Which uncertainties remain unidentified?	Separate input variability from missing information	Rank condition; uncertainty intervals; Table 9

1.4 Contributions

Firstly, a primitive-event ledger is defined which involves an allocation matrix that keeps track of total burdens throughout the lifecycle and across model descendants. Secondly, it provides an exact error expression for aggregation over the time dimension without using weights, and a condition where missing information will not be replicated. Third, it is formulated as a minimax routing problem with explicit baseline scales and service constraints, which are both dimensionless and removed from the background. Third, it provides a dimensionless minimax routing formulation with explicit background

scales and service constraints. Fourthly, it completes an open scenario experiment, including negative results, ablations and uncertainty analysis. Finally, it provides the source of calculation of the HTML artefact so that each scenario quantity can be recalculated.

Successful, quality-compliant request per unit of functional unit is 1,000. This unit requires the accompanying Use (Tokens), Latency, Task Success and Modality in a deployment study. In this case, every synthetic request is considered to be successful and is the same across all policies - so the service quality is determined by construction, not measured. This restriction hampers the ability to compare the results with that of commercial models.

2. LITERATURE REVIEW

De Vries-Gao [1] analyzed the carbon and water footprints of data centres and discussed their potential implications for the rapidly expanding use of artificial intelligence. Khan et al. [2] explored green artificial intelligence strategies and analyzed strategies to lower energy usage in AI systems. Hankendi et al. [3] discussed the need to achieve transparency in advancing sustainable data centres and the carbon neutral Artificial Intelligence. The projected growth of AI-related carbon emissions in China offered by Chen et al. [4] suggested the AI carbon footprint would significantly expand after the national carbon peak in China. Following the review by Lei et al. [5] of the water use of the data-centre workloads, the most important factors influencing the total water-use were determined. Lal & You [6] reviewed the progress and obstacles on the path to AI infrastructure scaling and energy and climate goals.

Plociennik et al. [7] have performed the review on life-cycle assessment (LCA) methods for AI application and pinpointed key research gaps and future prospects. Luo et al. [8] studied the energy performance and the life cycle of the data centre desiccant-assisted indirect evaporative cooling system. Burden shifting in data-centre cooling has been studied by Brahmabhatt et al. [9] and the optimal impacts of water, energy and carbon localized at urban infrastructure level. The research on radiative free cooling was carried out by Aili et al. [10] for the purpose of lowering energy and water consumption in data centres. Isler-Kaya and Karaosmanoglu in their published study [11] have done Life Cycle Assessment of a cooling system which is more climate friendly for data centres and assessed the environmental performance of the system.

Al Kez et al. [12] discussed the sustainability issues challenges of digitalisation and Internet data centers, and upraised their environmental impacts. Walk et al. [13] explored the artificial intelligence field of sustainability and showed that computer vision can help to achieve sustainable smart product-service systems. Ahmad et al. [14] did a review of the role of artificial intelligence in the field of sustainable energy and talked about the status, challenges and opportunities of AI in the field of sustainable energy. The researcher, Kar et al. [15] conducted a systematic review of the literature on artificial intelligence and sustainability and explored the roles that AI could play in sustainable development. By reviewing the literature, Wu et al. [16] discussed the innovations of AI for energy saving and explored the potential of using AI for enhancing energy efficiency in Chinese listed companies.

Cao et al., [17] proposed a digital-twin adaptive capacity provisioning method for carbon-aware data centres, and also assessed the method's effectiveness on sustainable resource management. Huang et al. [18] discussed the power-management optimization in data centre systems with respect to power-supply systems and explored how to improve operational efficiency. Nassar et. al. 2019 [19] has presented an extensive environmental footprint evaluation method of a data centre and analysed several dimensions of sustainability. Hu et al. [20] cover robust spatiotemporal optimal scheduling of distributed data centres with multiple uncertainties in mind. To develop a formal modeling and analysis framework for a small-scale data centre integrated with renewable energy, Samaye et al. [21] used timed automata.

Vanderbauwhede [22] simulated the concept of geographic load-shifting with carbon awareness and explored the concept of spatial distribution of workloads based on a carbon awareness. Hasan and others mentioned in this article [23] reviewed the ways carbon emissions can be measured for machine learning, and discussed some of the approaches that could be used to measure the carbon footprint of machine learning workloads. Wang et al. [24] explored how to more effectively manage energy and computational workload for geo-distributed data centres, while studying how to coordinate resources across geo-distributed locations. Canali et al. [25] suggested a distributed optimization solution to reduce the energy expenses in the cloud data centres due to migration of workloads, data transmission and computation resources. For inter-data-centre networks with elastic optical infrastructure, Zhang et al. [26] explored the problem of energy-aware virtual-machine management.

Alanazi et al. [27] studied the possibilities of peak shaving and locked-in energy avoidance to reduce consumption of data centres. Zhou et al. [28] designed an energy efficient Virtual machine allocation algorithm for the IoT application in cloud data centre called AFED-EF. To reduce the energy consumption of data-centres, Wang et al. [29] proposed a frequency adaptive virtual data-centre embedding approach. Pantazoglou et al. [30] proposed a decentralized workload management approach and analyzed energy-efficient workload allocation in enterprise cloud environment.

Algorithm 1: (J-MEGA) Joint Minimax Environmental and Green Allocation Algorithm

Input: $T = 168$ hours, 3data centers, 24lifecycle windows.

Time vector

$$t = 0, 1, \dots, 167, h_t = t \cdot 24$$

PUE

$$P_{d,t} = P_d, P = [1.15, 1.30, 1.20]$$

Carbon intensity

$$C_{d,t} = C_d \left[1 + 0.20 \sin \left(\frac{2\pi(h_t - \phi_d)}{24} \right) \right]$$

where,

$$C = [0.15, 0.45, 0.30], \phi = [2, 8, 14].$$

Water intensity

$$W_{d,t} = W_d \left[1 + 0.15 \sin \left(\frac{2\pi(h_t - 4)}{24} \right) \right]$$

where,

$$W = [1.20, 0.25, 0.60].$$

Virtualization/energy factor

$$V_{d,t} = [2.0, 0.4, 0.8].$$

Environmental coefficient

$$K_{d,t} = [P_{d,t}, P_{d,t}, C_{d,t}W_{d,t} + P_{d,t}V_{d,t}].$$

Normalization

$$B = [1.2, 0.36, 1.56]$$

Minimax score

$$S_{d,t} = \max_{m \in \{1,2,3\}} \frac{K_{d,t,m}}{B_m}.$$

Policy selection

$$r_t^* = \arg \min_d S_{d,t}$$

for the **Joint Minimax** policy.

Other policies are obtained as:

$$r_t^C = \arg \min_d K_{d,t,2},$$

$$r_t^W = \arg \min_d K_{d,t,3}.$$

Request load

For lifecycle window j :

$$n_{j,t} = 20000 \left(1 + \frac{j}{30} \right) \left[1 + 0.25 \sin \left(\frac{2\pi(h_t - j8)}{24} \right) \right].$$

Utilization

$$u_{j,t} = 0.35 + 0.55 \frac{n_{j,t}}{\max_t n_{j,t}}.$$

IT energy

$$e_{j,t} = \frac{n_{j,t}}{1000} \left(0.18 + \frac{0.12}{u_{j,t}} \right).$$

Serving impact

$$I_j = \sum_{t=1}^{168} e_{j,t} K_{r_t,t}.$$

Total lifecycle impact

$$T_j = I_j + F,$$

where F represents fixed embodied/environmental impacts.

Impact per 1,000 requests

$$I_j^{1000} = \frac{T_j}{N_j} \times 1000, \quad N_j = \sum_t n_{j,t}.$$

Monte-Carlo uncertainty

$$z_i \sim \mathcal{N}(0, 1),$$

$$\sigma_i = \sqrt{\ln(1 + \text{CV}_i^2)},$$

$$M_i = \exp(\sigma_i z_i).$$

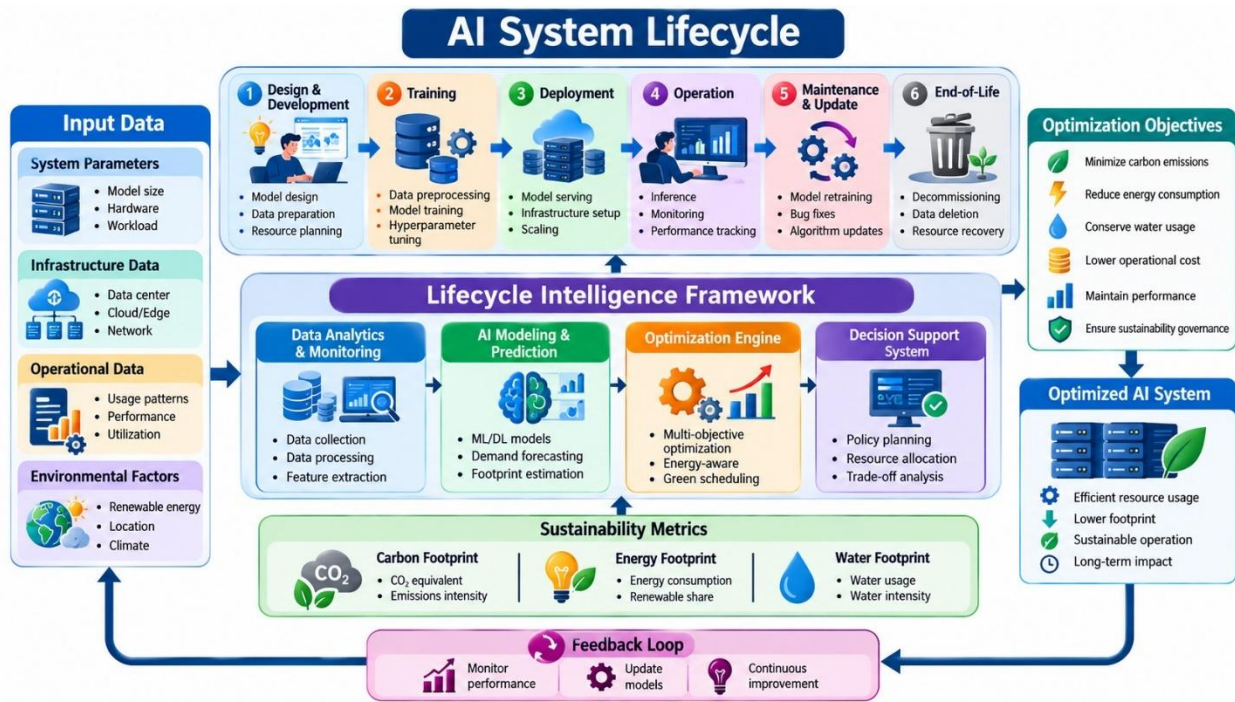
For 10,000 samples:

$$T^{(q)} = T \odot M,$$

and uncertainty intervals are obtained from

$$Q_{2.5\%}, Q_{50\%}, Q_{97.5\%}.$$

In the **figure 1** proposed Lifecycle Intelligence Framework integrates AI system lifecycle stages, input data, sustainability metrics, predictive modelling, multi-objective optimization, decision support, and feedback mechanisms to jointly minimize carbon emissions, energy consumption, and water usage while maintaining AI performance, reducing operational costs, and enabling continuous sustainable optimization.



**Developed by the authors. The underlying framework and conceptual structure are authors' original work; AI assistance was used only to assist with visual representation and infographic design.*

Figure 1. Framework AI System Lifecycle

3. MATERIALS AND METHODS

The study combines deductive model construction with controlled numerical experiments. The analytical component defines units, event boundaries, allocation constraints, and uncertainty propagation. The computational component varies traffic profiles and routing while holding service requirements and fixed lifecycle costs constant. There are no training of a new neural architecture and no physical power or water measurement. The tests establish internal consistency and conditional trade-offs, not external predictive validity.

3.1 Dataset / Data Collection

Two evidence layers are maintained. Layer A contains verified publications and official data portals used to justify definitions and identify field-data sources. Layer B contains the synthetic hourly input array generated by Eqs. (18) – (20). No eGRID, EIA, Aqueduct, or private production rows were downloaded or used to parameterize the reported experiments. Their roles and restrictions are made explicit below so that a valid source link is not mistaken for an executed data ingestion.

Table 2 source-to-analysis mapping distinguishes usable public resources from data actually analyzed. An empirical replication must archive exact releases and extraction scripts rather than relying on these live portal pages.

Table 2. Dataset Sources

Source / link	Available evidence	Resolution / caveat	Use in this paper
EPA eGRID (Emissions & Generation Resource Integrated Database (eGRID) US EPA)	Electricity generation and emissions	US historical aggregates; not hourly marginal signals	Field-data pathway only
EIA-923 Form EIA-923 detailed data with previous form data (EIA-906/920) - U.S.	Power-plant operating data	Cooling-water coverage and missingness require checks	Upstream-water pathway only

Energy Information Administration (EIA)			
WULCA / AWARE Home - WULCA	Water scarcity characterization	Basin/time and method version matter	Characterization definition
CodeCarbon Code Carbon	Energy/carbon estimation software	Measured and modeled components depend on setup	Telemetry option; not executed

3.2 Data Preprocessing

If using measured deployments, ensure that the timestamps are recorded in UTC, time zones are not changed and reject duplicate meter/event identifiers. Only include power inside valid intervals; mark missing intervals and use a clearly separate model to estimate power. To convert to kWh, divide the power value (in watts) times the number of seconds by 3.6×10^6 . The cost of cooling electric power cannot be billed once at a facility boundary, but rather already shows up as a meter reading, and should not be counted another PUE multiplier. The readings from water should be retained as withdrawal, return flow or consumption.

Add electricity factors/physical supply areas and intervals and include water characterization/basin for the actual water using process. Water taken from a basin may not be automatically defined as actual water used in a data centre by a power plant. Retain a “source identifier”, “release date”, “units”, and “uncertainty class” for each benchmarked factor and include a “boundary description” on each one. Unlike synthetic inputs the finite value, positivity, shape and dimensional checks are done during array construction and do not require imputation.

3.3 Proposed Method

Let J denote AI workloads, S the lifecycle stages, R be a region, T be a time interval and K be a primitive inventory event. There is a unique identifier for each event, inventor vector b_k for it, a time interval a physical location and a measurement/proxy flag. The lowest burden that can be admitted to the ledger without it overlapping another burden is primitive event. A parent facility meter cannot be a primitive total along with child meter(s).

$$b_k = (E_k, C_k, W_k)^T; E [kWh], C [kg CO_2e], W [L consumed]. (1)$$

Energy must have a carrier vector, for instance purchased electricity, fuel, and heat use must also be recorded separately – it is permitted to add these together as one energy as final energy, provided there is consistent conversion and a consistent system boundary. Primary-energy demand is not an added component in final electric power but is separately identified. In the numerical study all energy inputs included are considered electrical energy and reported E are cumulative electricity for declared processes which is not a comprehensive primary-energy input.

$$F_{js} = \sum_{k=1}^K A_{js,k} b_k; A_{js,k} \geq 0, \sum_{j,s} A_{js,k} = 1. (2)$$

Proposition 1 –conservation. The total amount of inventory remains constant if every primitive event has only one column in A and each column contains a number that is 1. Let finite sums be interchanged as follows:

$$\sum_{j,s} F_{js} = \sum_k (\sum_{j,s} A_{js,k}) b_k = \sum_k b_k. (3)$$

The result for each inventory component is independent. Does not prove causal correctness of a chosen allocation rule. If a work is not assigned to the unassigned or non-AI residual category and is modeled for less than the life of the asset, the unassigned or non-AI residual must still remain. If you remove that

category and re-normalize over the incomplete users, the unobserved burdens transfer to the observed users.

Table 3. Example allocation unit test (not included in the routing experiment). A child portfolio does not have a direct correspondence to the parent event portfolios; instead, they are only receiving a portion of the burden.

Table 3. Units Allocation

Primitive event	Model A share	Model B share	Residual share	Conservation check
Shared pretraining: 10,000 kg CO ₂ e	0.50 → 5,000	0.30 → 3,000	0.20 → 2,000	10,000
Hardware manufacture: 4,000 kg CO ₂ e	0.25 → 1,000	0.50 → 2,000	0.25 → 1,000	4,000
Cooling consumption: 2,000 L	0.40 → 800	0.40 → 800	0.20 → 400	2,000

For an inherited model family, pretrained, fine-tuned and serving are separate events. When models are copied, compressed, and the deployed model's identifiers move, so do the identifiers of their child elements. A new model number does not result in an additional manufacturing event. If known, the Allocation factors may be changed as a consequence of the lifetime demand and as a result, the inventory versions must be distinguishable from the original and the revised factors.

Temporal boundary choice is also available. A retrospective service inventory could allocate development to actual output over the life of the project, and a prospective project inventory could apply forecasts of demand. These two should be substituted but not put silently in place of each other. We deliberately choose a training heavy test case, for this paper all declared fixed costs are allocated to one single 7-day synthetic service campaign, and investigate how the interpretation changes as a function of life time output.

3.4 Proposed Model -operational energy, carbon, and water

Let $P_{it,j}(t)$ denote allocated IT power for workload j , in watts. For sampled data the trapezoidal rule approximates the integral, with intervals expressed in seconds. Define p_{rt} as facility PUE, c_{rt} as location-based electricity greenhouse-gas intensity, w_{rt} as on-site water consumption per IT kWh, and v_{rt} as upstream water consumption per delivered facility kWh.

$$E_{it,j} = (1/3.6 \times 10^6) \int P_{it,j}(t) dt; \quad E_{fac,jrt} = p_{rt} E_{it,jr}. \quad (4)$$

$$C_{op,j} = \sum_{rt} E_{fac,jrt} c_{rt} + C_{fuel,j} + C_{refrigerant,j}. \quad (5)$$

$$W_{op,j} = \sum_{rt} E_{it,jrt} w_{rt} + \sum_{rt} E_{fac,jrt} v_{rt}. \quad (6)$$

Since IT energy is the declared denominator of the direct WUE, then it is used in the multiplication in this case, as in equation (6). If a source provides liters per facility kWh, then the first term should be based on facility energy. Only disjoint processes can be used for the direct and upstream factors. If the direction water is being used for cooling, then the same water would be added twice to the water factor—once as a water and once as an all-in water. Fuel and refrigerant emissions are assumed zero in the numerical experiment, not omitted from the general specification.

$$W_{scar,j} = \sum_{b,t} C F_{bt} W_{jbt} / 1000 \quad [m^3 \text{ world-equivalent}]. \quad (7)$$

Here b indexes the basin of each water-consuming process. Equation (7) is a characterization result when compatible AWARE factors are used [15]; it is not a physical volume. Withdrawals and returns require separate inventories. Aqueduct risk scores [14] can inform siting context but must not be substituted for CF. Because this study has no measured basin allocation or downloaded factors, scarcity-weighted numerical results are not reported.

$$C_{exact} - C_{annual} = T \text{Cov}_T(E_t, c_t), \quad C_{annual} = T \bar{E} \bar{c}. \quad (8)$$

Proposition 2-temporal aggregation bias: Define $\text{Cov}_T(E, c) = T^{-1} \Sigma (E_t - \bar{E}) (c_t - \bar{c})$.

Expanding the product gives $\Sigma E_t c_t - T \bar{E} \bar{c}$, proving Eq. (8). The average-factor estimator is exact only when covariance vanishes, or when its intensity is explicitly energy-weighted. Positive dependence causes underestimation by an unweighted mean. The same identity applies to electricity-related water factors. A published annual factor is not necessarily the arithmetic average of an unavailable hourly series, so Eq.(8) isolates one mechanism rather than explaining every mismatch. Location-based inventories describe attributed purchased-electricity emissions. A consequential scheduling estimate instead needs a defensible marginal response of the power system, including effects of shifting demand. The experiment uses exogenous average-intensity signals; reported differences are therefore attributional. Renewable certificates, offsets, and restoration claims are not subtracted from physical electricity or water quantities. Alternative contractual reporting can be added as a separate ledger view.

3.5 Proposed Method / Model - allocation, uncertainty, and break-even

For asset a , let h_{ja} represent attributable occupied capacity-hours, H_a the total allocable lifetime capacity-hours, and B_a its embodied vector. An idle residual preserves unused capacity. A use-based rule is $\alpha_{ja} = h_{ja}/H_a$, with $\Sigma_j h_{ja} \leq H_a$. This is an unambiguous generic rule that isn't uniquely privileged – reserved capacity rules, peak-demand rules are responding to different questions of attribution.

$$F_j = F_{op,j} + \Sigma_a \alpha_{ja} B_a + F_{other,j}, \alpha_{res,a} = 1 - \Sigma_j \alpha_{ja} \quad (9)$$

$$I_j = 1000 F_j / N_{ok,j}; \quad N_{ok,j} = \Sigma_i 1\{quality_i \geq q_0, latency_i \leq \ell_0\} \quad (10)$$

If the nucleotide costing was only in the numerator then the bad work would be rewarded. All admitted life cycle burden, such as retries, discarded experiments, and failed requests include in the numerator, useful output in the denominator. Model comparisons should be based on quality thresholds that are kept stable, as should be the value of tokens per request.

$$\text{Var}(F) \approx J \Sigma_\theta J^T, J_{ik} = \partial F_i / \partial \theta_k \quad (11)$$

$$\text{Var}(E_c) \approx c^2 \sigma_E^2 + E^2 \sigma_c^2 + 2Ec \text{Cov}(E, c) \quad (12)$$

Equation (11) is a first-order expansion around nominal inputs. It fails to be reliable for very large non-linear perturbations or disconnections in routes, which is why scenario sampling is desired. A shared factor will impact more than one job at a time; hence the importance of correlation. If a new grid factor would be drawn for each of these requests, it would be artificial continuation of the uncertainty loss as the number of these requests increases. Figure 4 coordinates a correlation sensitivity calculation with the unattached-factor Monte Carlo model actually used.

$$y = H\theta; F = G\theta. \text{Fidentifiable from } y \Leftrightarrow \ker(H) \subseteq \ker(G). \quad (13)$$

Proposition 3 Inventory identifiability. Then, if h in $\ker(H)$ then, the observed data for both θ and $\theta+h$ is the same. They have different targets for inventories. Hence, the target is unique only when all these Gh 's are nulls. On the other hand, the inclusion implies that all possible y sharing θ share F as well. An operational electricity mode measurement can only be used in combination with other constraints in

order to recover the missing embodied-water coefficient. Reporting zero uncertainty on an unknown number is solely an unfortunate result of the exhaustive approach that might be adopted. It is logically incorrect to report zero uncertainty in a quantity whose value is not known.

$$F_A(N) = B_A + Ne_A; N^* = (B_B - B_A)/(e_A - e_B). \quad (14)$$

Equation (14) is component wise and assumes equal quality, constant per-request burdens, and fixed B. It follows by setting $\widehat{F}_A(N) = F_B(N)$ and rearranging. A positive break-even exists only when numerator and denominator share a sign; zero denominator requires a separate case. Carbon and water generally have different thresholds, so a single universal “green break-even” is not defined. If the efficient system causes additional demand, the volume must also change; this rebound is outside the fixed-demand comparison.

3.6 Decision formulation

Let x_{jrt} be the proportion of workload j routed to region r in interval t . Let a_{qrt} be the operational burden per TT kWh for indicator $q \in \{E, C, W\}$, equal to $(p, pc, w + pv)$. Choose fixed positive normalization scales $d = (1.20, 0.36, 1.56)$, with units matching a . A normalized minimax program is:

$$\min_{\mathbf{x}, z} z \text{ s. t. } \sum_{jrt} x_{jrt} E_{it} j t a_{qrt} / D_q \leq z \forall q; \sum_r x_{jrt} = 1, \mathbf{x} \geq 0. \quad (15)$$

$$\sum_j x_{jrt} n_{jt} \leq K_{rt}; x_{jrt} = 0 \text{ if region } r \text{ violates latency or quality constraints.} \quad (16)$$

D_q is a fixed total- budget normalization for Eq. (15), whereas d_q normalizes per- kWh coefficients in the hourly comparator below. Equation (15) is a linear program when costs and capacity are exogenous. Its Lagrangian contains multipliers $\lambda_q \geq 0$ on resource constraints, with stationarity in z requiring $\sum \lambda_q = 1$. Thus resource weights arise as dual variables of the worst normalized burden; they do not assert equivalence of litres and kilograms.

$$r_t^* = \operatorname{argmin}_r \max_q \{a_{qrt}/d_q\}. \quad (17)$$

The executed experiment uses the deterministic hourly rule (17), not the fractional, capacity-coupled program (15). All traffic is routed to one region per hour; capacities are assumed sufficient and quality/latency identical. The local rule is optimal for its own hourly discrete objective but is not claimed to solve the global minimax program. Migration energy, queueing, and endogenous grid response are excluded.

3.7 Experimental Setup

Table 4 these are not measurements of any specific site or country, but rather are objective scenarios defined by the designer. The labels of the regions denote the suggested compromise and are not to be applied for geographical suggestions.

Table 4. Measurement of Hypothetical Region

Hypothetical region	PUE p	Base (kg/kWh)	c	Base w (L/IT- kWh)	v (L/facility- kWh)	Carbon phase ϕ
A: lower-carbon / higher-water	1.15	0.15		1.20	2.00	2 h
B: higher-carbon / lower-water	1.30	0.45		0.25	0.40	8 h
C: intermediate reference	1.20	0.30		0.60	0.80	14 h

$$c_{rt} = \bar{c}_r [1 + 0.20 \sin(2\pi(h_t - \phi_r)/24)], \quad w_{rt} = \bar{w}_r [1 + 0.15 \sin(2\pi(h_t - 4)/24)] \quad (18)$$

$$\eta_{jt} = 20,000(1 + j/30)[1 + 0.25\sin(2\pi(h_t - (j \bmod 8))/24)], \quad j = 0, \dots, 23. \quad (19)$$

$$u_{jt} = 0.35 + 0.55 \eta_{jt} / \max_t \eta_{jt}, \quad E_{itjt} = (\eta_{jt}/1000)(0.18 + 0.12/u_{jt}). \quad (20)$$

$h_t = t \bmod 24$ for $t = 0, \dots, 167$. Request values are the number of requests/reads/writes that are expected to happen during a particular time slot. They may be fractional at the individual hour level, and are summed to create the denominator. The utilization relation isn't a "fitted" server power curve: it's a prescribed algebraic relationship that can be found in the literature. The progressive check for idled amortization ensures a moderated pathway to investigate utilisation impacts.

3.8 Metrics, Baselines, and Implementation

Report absolute lifecycle E, C, W; serving-only values; per-1,000-successful-request intensities; component shares; and paired relative changes $\Delta_q = 100(F_{policy,q} - F_{ref,q}) / F_{ref,q}$. Reductions = negative values. Nor 'normalised worst coefficient' is an environmental impact unit, therefore it is an IRO internal objective. Reconciliation residuals checked on a per primitive event and per indicator basis. Coverage without a source is not a numerical uncertainty and must be reported separately.

Table 5 indicates the fixed operational stages use $p = 1.25, c = 0.40, w = 0.50$, and $v = 1.20$. The manufacturing entry is not electricity generated by electricity generating facilities of the data centres but electricity generated to produce goods. It is one-off in nature and doesn't get a further operational multiplier.

Table 5. Fixed Operational Stages

Fixed stage	IT electricity (kWh)	Facility electricity (kWh)	Carbon (kg CO ₂ e)	Water (L)
Data preparation	120	150	60	240
Search / discarded trials	1,800	2,250	900	3,600
Training	3,600	4,500	1,800	7,200
Adaptation	600	750	300	1,200
Storage / ancillary IT	180	225	90	360
Retirement processing IT	30	37.5	15	60
Allocated manufacturing inventory	—	2,000 ¹	1,200	8,000
Total fixed burden	6,330	9,912.5	4,365	20,660

3.9 Baseline Methods / Comparative Methods

Carbon-only chooses $\argmin_r(p_r, c_{rt})$ Water-only chooses $\argmin_r(w_r + p_r v_r)$. Joint minimax follows Eq. (17). All four have the same fixed ledger, identical IT Traffic model. Distinctly, boundary ablations rid of the embedding weights, exclude lifecycle stages or disregard PUE or upstream water. These ablations have only been deliberately incomplete estimators, not Code Carbon or Carbon tracker. A software performance ranking would require using those tools in conjunction with a common measurement methodology.

3.10 Implementation Details

Using Python on its own with NumPy and Matplotlib creates the arrays, regions for their analysis, burdens and renders vector figures. Embedded in this HTML is the calculation source, which can be downloaded from the HTML's tool bar. It functions completely without any private APIs, commercial-model experiments are not necessary. The random generator is some number generator from NumPy, which has default random number generator + seed. For the deterministic study, the size of the request interval is 24×168 and the regional coefficient vector is studied with 3 vectors, as well as 4 policies.

The multiplicative factors of competing policies are determined by 10,000 uncertainties. Median-one lognormal multipliers are produced from 5 standard normal independent variables that are used to generate these emissions. The produced multipliers are the median-one lognormal multipliers for the emissions of serving energy, serving carbon intensity, combined serving-water intensity, embodied carbon, and embodied water, which are shown in the lognormal distribution below. Their coefficients of variation are 0.10, 0.15, 0.25, 0.30, and 0.30. The costs of costs: operational burden and embodied electricity are fixed. As such, the uncertainty intervals are partial and conditional—they do not cover all uncertainty in the lifecycle.

4. RESULTS

In **figure 2**, all the values of Lifecycle footprint increase over the elapsed days, and show that the more the AI systems run, the more electricity they use, carbon emissions they emit, and water they consume.

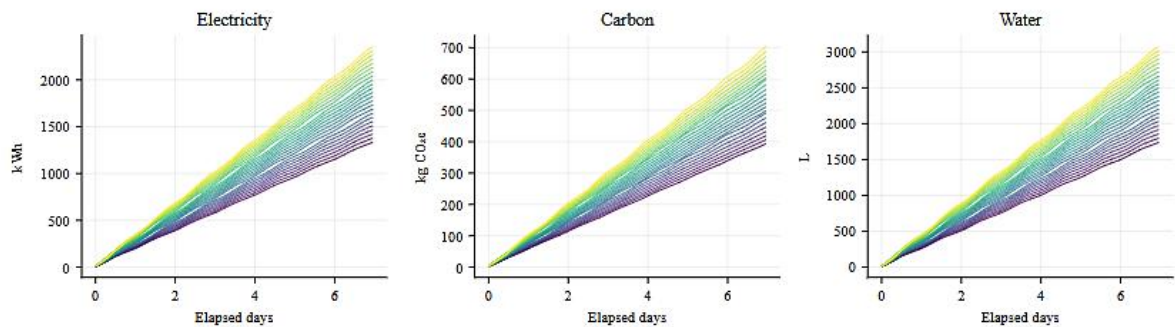


Figure 2. the purpose of the synthetic cumulative serving burdens is to work out the number of demands that can be met by the combined policy: lines from purple to yellow indicate the index of the demand profile (0–23). Time is shared by the three panels, but they are otherwise independent. Each line incorporates 168 hours’ worth of events, which are alternate scenarios, not 24 measured systems.

This is because the cumulative trajectories are an increasing sequence (due to the absence of offsets in the ledger and nonnegative consumption). They are separated according to the designed traffic-volume multiplier and the small diurnal curvature is caused by utilization and the environment. The visualization has many lines for the purpose of showing profile sensitivity without any of the lines being a stochastic duplicate. Cumulative serving curves do not have a fixed lifecycle burden, adding in the fixed vector translates each curve without changing the increment.

Table 6 about all entries are based on the same base workload of 3.36 million requests. Minimax and Reference C are the same since for all 168 intervals in the TAMiT table, region C gets the lowest declaration of the hourly worst normalized coefficient.

Table 6. Minimax Policy Profile

Policy (profile 0)	Serving E (kWh)	Serving C (kg)	Serving W (L)	Total C (kg)	Total W (L)
Reference C	1,333.42	393.52	1,738.14	4,758.52	22,398.14
Carbon-only	1,277.86	194.80	3,898.54	4,559.80	24,558.54
Water-only	1,444.54	643.94	857.57	5,008.94	21,517.57
Joint minimax	1,333.42	393.52	1,738.14	4,758.52	22,398.14

The full lifecycle total is 11,245.92 kWh, 4,758.52 kg CO₂e, and 22,398.14 L. Dividing by 3,360 units of 1,000 requests gives 3.3470 kWh, 1.4162 kg CO₂e, and 6.6661 L per functional unit. These intensities represent all of the allocated manufacturing and development loads of a short service campaign; hence, they are higher than serving-only intensities. They are not a measure of production or estimate to mount a commercial chat bot.

Carbon-only subjects region A to serving carbon, which drops by 50.50% compared to the reference, whereas water-only subjects region B to serving water, which goes up by 124.29% compared to the

reference. Water-only results in lowering serving water by 50.66% and augmenting serving carbon by 63.63%. This is a direct result of the scenario factors and not merely an empirical observation of actual region grids. It ensures that the actions for a carbon objective do not end up being interpreted as water objective.

The fixed component supplies about 91.7% of the reference carbon and 92.2% of the reference water. Those large fractions make total-lifecycle changes much smaller than serving-only changes. Reporting a serving reduction as an equal percentage lifecycle reduction would therefore be incorrect in this experiment. Lifecycle scope changes the denominator as well as the list of included processes.

4.1 Comparative Analysis

Figure 2 should not be interpreted as evidence that expanding demand is environmentally beneficial. Absolute serving burdens increase with traffic; only the fixed burden per request decreases. The distinction between absolute totals and normalized efficiency is essential for evaluating rebound. In this study profile volume increases from 3.36 million to 5.936 million requests without changing the fixed inventory, while eight demand phase shifts repeat across profile indices.

The **figure 3** shows that as requests per lifecycle window increase, both carbon and water footprints decrease, with the carbon-only and water-only strategies reducing their respective footprints while the joint-minimization strategy provides a balanced reduction across both environmental impacts.

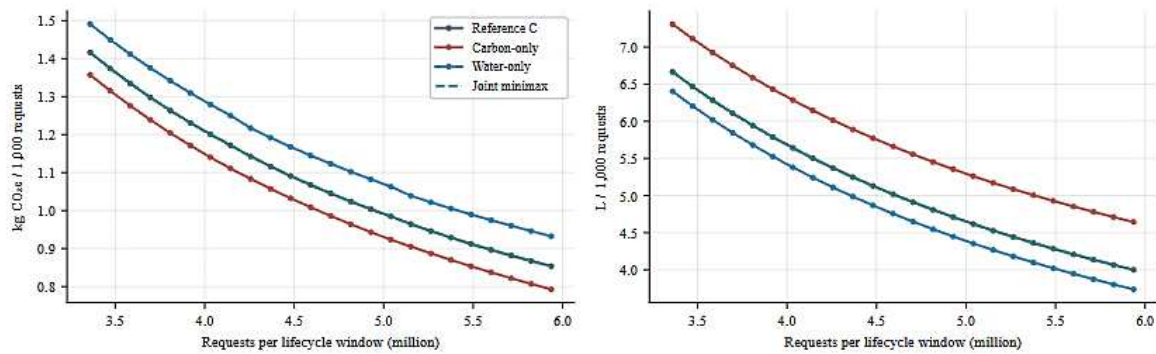


Figure 3. Lifecycle carbon and water intensity across 24 designed profiles under all four policies. Reference C and joint minimax overlap exactly; the equality is a result, not missing data. Lower intensity at larger service volume is driven mainly by amortization of the same fixed burden.

Table 7 shows the lifecycle changes that uses Reference C as the denominator and preserve all fixed stages. The joint rule offers a compromise relative to single-objective routing but provides no improvement over this already-intermediate reference.

Table 7. Policies Lifecycles

Policy	Δ lifecycle E (%)	Δ lifecycle C (%)	Δ lifecycle W (%)	Region-hours A/B/C
Reference C	+0.00	+0.00	+0.00	0 / 0 / 168
Carbon-only	-0.49	-4.18	+9.65	168 / 0 / 0
Water-only	+0.99	+5.26	-3.93	0 / 168 / 0
Joint minimax	+0.00	+0.00	+0.00	0 / 0 / 168

The minimax rule is not robust in any way. Scaling dC up gives one less restriction on carbon; the same for water in the case of dW. Such scales should be clear expressions of decision preferences, or resource budgets, and should always be attached to an expression of a selected point being “balanced.” If there is an equal number of measurement units, but the physical measurements are incompatible, there is nothing really to interpret.

The relationship between the intensity of using and serving water to carbon intensity is shown in **figure 4**. In all feasible segments, the intensity of the use and service of water decreases with broadening of the

carbon intensity measure, identified by the coefficient of variation (CV). The bounds of the carbon use intensity measure expand as the correlation between carbon and energy-related factors grows, which means that, under stronger positive correlation, there is greater variation in the water intensity factor bound.

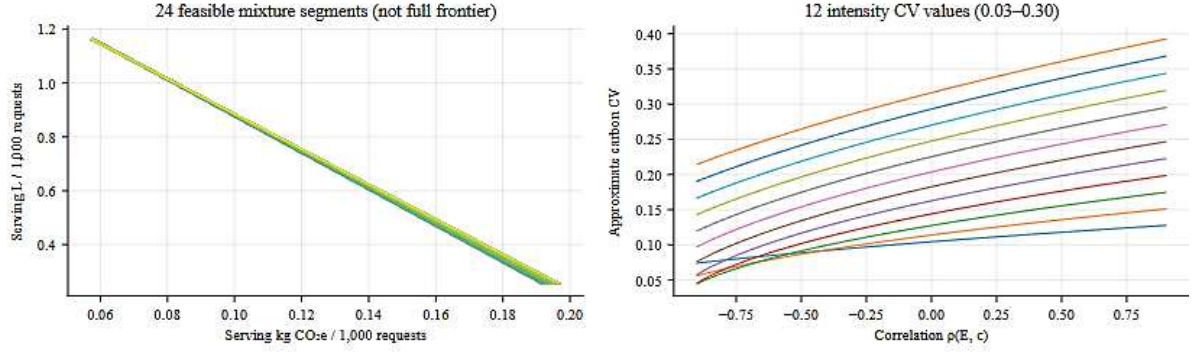


Figure 4. Left: 24-line segments show their convex mixtures of carbon-only and water-only inventories of serving size. They are realistic mixtures (not a calculated global Pareto frontier), based on the assumed divisible traffic. Right: 12 analytical curves, showing the variation of the carbon uncertainty with correlation, for CVs of intensity from 0.03 to 0.30.

After normalisation, the profiles almost overlap along some of the mixture lines, since the difference mostly lies in scaling. Just a rich, dense curve does not lead to rich empirical evidence. The right panel follows $CV_C \approx \sqrt{(CV_E^2 + CV_c^2 + 2\rho CV_E CV_c)}$ with $CV_E=0.10$, derived from Eq. (12). If it turns out there is a positive correlation, uncertainty is increased; if it turns out that there is a negative correlation, that can reduce uncertainty. This sensitivity experiment is not a distribution of different factors as described in Section 4.4.

4.2 Ablation Study

The ablations test the margin of a learned model, not disable sub-models. Let S = vector of servings according to Reference C and B = manufacturing vector. The complete expression is $F = F_{\text{fixed}} + S$. is removed from the text “No embodied”. “Serving only” keeps S . “No upstream water” excludes electricity related water from the fixed operational stages and serving but include direct cooling and embodied water. “Ignore serving PUE” recalculates serving for $p=1$, without changing the “fixed” ledger.

Table 8. The 0/1 parse determines all of the ablations. They reveal missing quantities, not error in measurement when compared to an actual measurement, physical ground truth or standard; smaller totals of ablated quantities are not efficiency gains.

Table 8. The 0/1 parse determines all of the ablations.

Estimator / ablation	E (kWh)	C (kg CO ₂ e)	W (L)	Interpretation
Full lifecycle	11,245.92	4,758.52	22,398.14	Declared reference
No embodied	9,245.92	3,558.52	14,398.14	Missing allocated manufacture
Serving only	1,333.42	393.52	1,738.14	Excludes every fixed stage
No upstream water	11,245.92	4,758.52	11,836.40	Incomplete water boundary
Ignore serving PUE	11,023.69	4,692.94	22,220.35	Facility overhead omitted in serving

Figure 5 illustrates that idle coefficients show a non-linear decrease with increasing utilization whereas allocated embodied carbon multipliers show a rapid decrease in utilization with increased requests per lifetime in the case of resources, with defined lower limits, but present consistently higher values for short lifetimes.

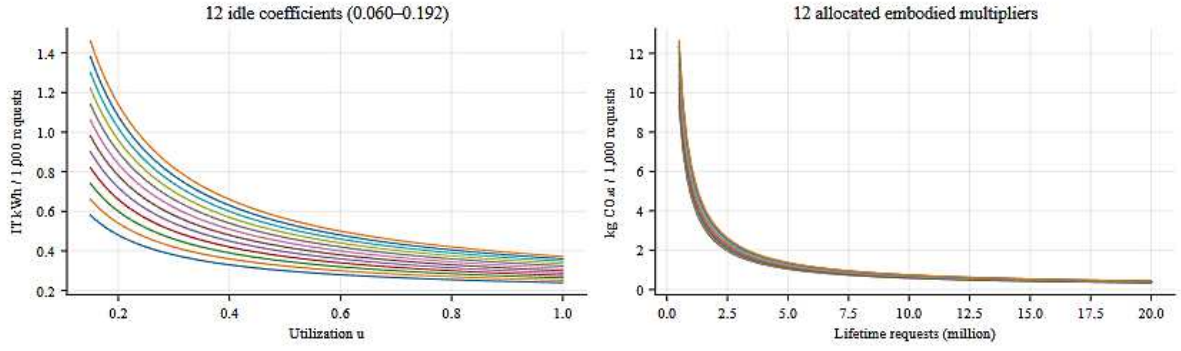


Figure 5. Left: 12 analytical utilization curves use $e(u) = 0.18 + a/u$ with $a = 0.060, \dots, 0.192$ kWh per 1,000 requests. Right: 12 carbon-amortization curves use $I(N) = 1,000(1,200m + 4,000)/N + 0.12$ with $m = 0.5, \dots, 1.875$. These auxiliary sweeps are independent of the main study constants.

In case $u > 0$ the model is decreasing and convex: Its derivatives are $de/du = -a/u^2$ and second derivative $2a/u^3$. Coaders showed that, consequently, the same actual improvement in the utilization of the element translates to a greater corresponding reduction in energy at a lower utilization, on the basis of this algebraic model. This is excluding the addition for power caps, memory bottlenecks, batching latency and bumps in model quality. When the operational rate is fixed, the right panel follows the $dI/dN = -1,000B/N^2$ rule. It distinguishes between the effect of service volume over the lifetime of the equipment and embodied-inventory uncertainty. The function (or the strictly monotone increasing branch of the function) is not asymptotic to the level of intensity 0 but rather the assumed operational intensity. Realized output-inventory and periodic reconciliation helps to avoid that reporting situation because an accounting policy is constantly plugging in the forecast lifetime volume, generating an artificially low current intensity. Embodied impacts are eliminated by excluding them from every policy, adding 1,200 kgCO_{2e} and 8,000 L. The absolute routing contrast does not change, since these amounts are considered to be typical. The percentage of the total lifecycle impact is however different. This shows another reason to have both absolute differences and boundary matched percentage differences next to optimization claims.

4.3 Statistical Analysis / Significance Testing

These 24 profiles represent the certain design points and do not reflect a population of deployed AI systems. Classical significance testing of these profiles, then, would give a false empirical interpretation. No p-values, no statistically significant real world superiority is reported. Rather, we tackle explicit uncertainty about input and empirical evaluate policies with a single drawn-in randomness.

$$M_k = \exp(\sigma_k Z_k), Z_k \sim N(0,1), \sigma_k = \sqrt{1n(1 + CV_k^2)}. \quad (21)$$

$$C^{(b)} = C_{\text{fixed,op}} + B_C M_{BC}^{(b)} + S_C M_E^{(b)} M_c^{(b)}, \quad \Delta^{(b)} = F_{\text{joint}}^{(b)} - F_{\text{ref}}^{(b)}. \quad (22)$$

For electricity use $E^{(b)} = E_{\text{fixed}} + S_E M_E$; for water use $W^{(b)} = W_{\text{fixed,op}} + B_W M_{BW} + S_W M_E M_w$. It's not the median of a sum, it's the median of the multipliers; therefore a sampled median of a sum need not equal its nominal value. These draws represent the inventory uncertainty along a route where the policy choices have been taken as exogenous as opposed to endogenous, with the result being a selection of the route from noisy forecasts.

Table 10. 10,000 draws to get Central 95% model based uncertainty intervals. They are not confidence intervals for a measured population mean, but rather the exact zero paired difference is a consequence of the identical paths and common uncertainty factors.

Indicator (profile 0)	2.5th percentile	Median	97.5th percentile	Paired joint-reference
-----------------------	------------------	--------	-------------------	------------------------

Electricity (kWh)	11,010.91	11,246.09	11,536.73	0 in every draw
Carbon (kg CO ₂ e)	4,220.90	4,767.45	5,701.10	0 in every draw
Water (L)	18,823.90	22,449.43	28,577.65	0 in every draw

The **figure 6** shows S-shaped cumulative probability distributions for electricity, carbon, and water footprints, indicating rapid transitions around their central resource-use levels followed by saturation toward a cumulative probability of 1.0, with increasing resource thresholds corresponding to higher cumulative probabilities.

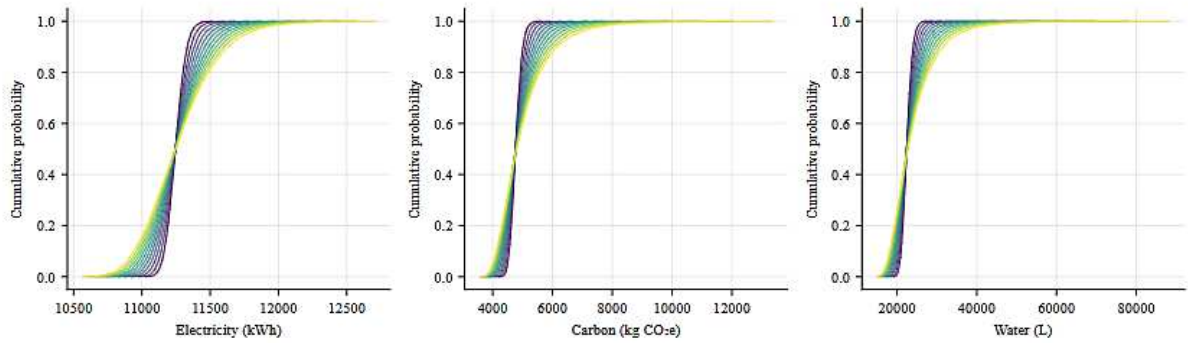


Figure 6. The sensitivity in the log-space scale of uncertainty is 0.5 to 2.0 times the baseline log-space standard deviations and is depicted in 12 empirical distribution curves per indicator. Evaluating input uncertainty (not repeated field measurements) is reflected in the width of all the curves, which have used the same 10,000 standard-normal draws.

A higher number of draws would make the quantiles more accurate, but would not make the input parameters of the problem appear with less uncertainty. Except for uncertainty in fixed operational energy, factor geography, the quality of a model and utilization in the future, the five-factor model would not have been excluded. These exclusions are more significant than adding decimal precision to the quantiles reported. For future field study the unit of replication should be a deployment/day or separate workload executions and matched model, hardware, task, output length. Temporal dependence may also be preserved in the interval using the block bootstrap and factor uncertainty may be shared using paired contrasts. The meaning of the statistical testing would justify itself only after the choice is made of the sampling target and the measurement protocol is specified.

5. DISCUSSIONS

The methodological finding is that there are questions different routing objectives and lifecycle inventories can answer. The routing experiment results in a different location of the servers, while the development and manufacturing costs do not change. This is why, in the case of a 50.5% cut in serving carbon, the top life cycle carbon cut is a mere 4.2%. Thus, if the optimisation doesn't change the underlying operational advantage, a correct accounting framework can help minimise the impact of the optimisation, it may seem at least. It is also good to have reference to the joint policy, that equated to Reference C. It makes sure that a proposed framework doesn't necessarily gain an invented trump, just because it's called "proposed. In this case the value of this is explicit treatment of conflicting goals, not reducing all to the same level; and then a tracing of a result. Different "interpretations" of the normalization scale, location, a different service constraint, etc. would yield a different route, and could require new experimentation, not an extrapolation from that experiment.

5.1 Comparison with Previous Studies

Development expenditure is treated as motivational training cost work [1] and reproducible metadata is an extension of the principles of systematic reporting [3]. They have a wider scope that corresponds with the BLOOM lifecycle analysis [5] and feature non-accelerator electricity because it corresponds to a

production-service measurement [6]. The addition of the possibility of carbon–water tension is in line with the coupled regional questions explored [11] and the 2026 UNU report [17]. These studies are not numerically validated against because of their different workloads and scope.

5.2 Practical Implications / Limitations

The first step in a practical implementation is to select a site-metres hierarchy then establish links between jobs and model lineage to the lowest common non-overlapping events. Quality of a model might be accompanied by the stage totals, source coverage and uncertainty intervals. A procurement comparison must indicate the inclusion of manufacturing and development, the type of water used (consumption or withdrawal) and the inclusion of electricity emissions as attributed or as marginal. While some information used to identify individuals and proprietary data may be anonymized with privacy-preserving aggregates, reconciliation totals should be auditable.

The greatest drawback is lack of concrete evidence. Therefore, the hypothetical regional factors intentionally create a trade-off but on their own do not determine how often or in what size and magnitude this occurs in reality. Assumptions of embodied values, a rather short campaign, imposing nature of quality equality. Capacity limits, migration costs, network topology, data from water basins, and end-user devices are not represented. The three indicators are showing mineral depletion, toxicity, land use and e-waste are off of the scope of the three indicators. The framework will accept such extensions, but these have not been tested here.

Allocation is not a cause. If nominal hardware lifespan is extended, the loading on the annual burden would be smaller but the actual emissions of hardware production would be the same. Similarly, if a workload is being transferred, the area under the curve changes due to that transfer but we would need a consequential power-system analysis to determine the actual overall emissions. The reported uncertainty doesn't account for the lack of constraints on the size of processes that never took place.

6. DECLARATIONS

6.1 Author Contributions

Kuldeep Singh Kaswan contributed to research supervision, conceptualization, methodology development, experimental design, data analysis, interpretation of results, and manuscript preparation and revision. Avadhesh Kumar contributed to technical guidance, methodology validation, experimental supervision, analysis, and manuscript review. Praveena Rajendra contributed to literature investigation, data collection and preparation, experimentation, data organization, software implementation, data curation, visualization, formal analysis, and experimental validation. Ramdas Khomane contributed to interpretation of results, experimental analysis, manuscript preparation and revision, research supervision, technical guidance, methodology validation, and manuscript review and revision.

6.2 Funding

This research is not funded by any organization.

6.3 Data Availability Declaration

The data used in this study are publicly available benchmark datasets. Details of the datasets, along with their respective sources, are provided in **Section 8** as Dataset Sources.

6.4 Competing Interest Declaration

The authors declare that they have no competing interests.

6.5 Conflict of Interest

The authors declare that they have no conflict of interest.

6.6 Code Availability

The source code developed and used in this study is available from the authors.

7. CONCLUSION AND FUTURE WORK

This paper identifies an auditable lifecycle approach for electricity-, carbon-, and water-store inventories for AI. Conservation, temporal aggregation bias, identifiability and break-even derivations explain those cases where a numerical result is defensible. The realized synthetic study demonstrated a transposition of loads even considering a single object and that percentage comparisons will significantly change when costs are fixed over a lifetime. It also provides a neutral output when the result of the joint minimax matches the result from the reference with the declared inputs:

Further research is needed to include external power consumption data, process specific embodied inventories, include upstream water connecting to basins where processed water is consumed and test routing for capacity and latency constraints. The accuracy of measurement requires an independent field dataset. No novelty review and journal specific formatting should be done before submitting.

8. DATASET SOURCES

1. EPA eGRID : <https://www.epa.gov/egrid>
2. EIA-923 : <https://www.eia.gov/electricity/data/eia923/>
3. WULCA / AWARE: <https://wulca-waterlca.org/what-is-aware/download-aware-factors/>
4. CodeCarbon : <https://codecarbon.io/>

References

- [1] de Vries-Gao, “The carbon and water footprints of data centers and what this could mean for artificial intelligence,” *Patterns*, vol. 7, no. 1, Art. no. 101430, 2026, doi: 10.1016/j.patter.2025.101430.
- [2] S. Khan, N. S. Naz, T. Mazhar, M. U. Tariq, T. Shahzad, S. Guizani, and H. Hamam, “Green AI techniques for reducing energy consumption in AI systems,” *Array*, vol. 29, Art. no. 100652, 2026, doi: 10.1016/j.array.2025.100652.
- [3] Hankendi, A. K. Coskun, and B. K. Sovacool, “Why transparency matters for sustainable data centers and carbon-neutral artificial intelligence (AI),” *iScience*, vol. 28, no. 11, Art. no. 113705, 2025, doi: 10.1016/j.isci.2025.113705.
- [4] Z.-M. Chen, Q. Xiong, J. Duan, J. Ma, Z. Chen, and S. Guo, “AI carbon footprint in China sets to double post-2030 carbon peaking,” *Energy Economics*, vol. 150, Art. no. 108880, 2025, doi: 10.1016/j.eneco.2025.108880.
- [5] N. Lei, J. Lu, A. Shehabi, and E. R. Masanet, “The water use of data center workloads: A review and assessment of key determinants,” *Resources, Conservation and Recycling*, vol. 219, Art. no. 108310, 2025, doi: 10.1016/j.resconrec.2025.108310.
- [6] Lal and F. You, “Advances and challenges in energy and climate alignment of AI infrastructure expansion,” *Advances in Applied Energy*, vol. 20, Art. no. 100243, 2025, doi: 10.1016/j.adapen.2025.100243.
- [7] Plociennik, P. Watjanatepin, K. Van Acker, and M. Ruskowski, “Life Cycle Assessment of Artificial Intelligence Applications: Research Gaps and Opportunities,” *Procedia CIRP*, vol. 135, pp. 924–929, 2025, doi: 10.1016/j.procir.2025.01.079.
- [8] J. Luo, K. Yang, Z. Li, Y. Lizhong, Y. Zhang, G. Ye, and H. Yang, “Life cycle and energy performance assessment of a novel desiccant-assisted indirect evaporative cooling system for data

centers,” *Journal of Cleaner Production*, vol. 554, Art. no. 148095, 2026, doi: 10.1016/j.jclepro.2026.148095.

- [9] P. Brahmabhatt, M. Hemmati, J. Vergara-Zambrano, V. M. Charitopoulos, and S. Avraamidou, “Uncovering the burden shifting of data center cooling: System-level water-energy-carbon optimization for urban infrastructure,” *Journal of Cleaner Production*, vol. 575, Art. no. 149201, 2026, doi: 10.1016/j.jclepro.2026.149201.
- [10] Aili, W. Long, Z. Cao, and Y. Wen, “Radiative free cooling for energy and water saving in data centers,” *Applied Energy*, vol. 359, Art. no. 122672, 2024, doi: 10.1016/j.apenergy.2024.122672.
- [11] Isler-Kaya and F. Karaosmanoglu, “Life cycle assessment of a climate-friendly data center cooling device,” *Energy and Buildings*, vol. 288, Art. no. 113006, 2023, doi: 10.1016/j.enbuild.2023.113006.
- [12] Al Kez, A. M. Foley, D. Lavery, D. F. Del Rio, and B. Sovacool, “Exploring the sustainability challenges facing digitalization and internet data centers,” *Journal of Cleaner Production*, vol. 371, Art. no. 133633, 2022, doi: 10.1016/j.jclepro.2022.133633.
- [13] J. Walk, N. Köhl, M. Saidani, and J. Schatte, “Artificial intelligence for sustainability: Facilitating sustainable smart product-service systems with computer vision,” *Journal of Cleaner Production*, vol. 402, Art. no. 136748, 2023, doi: 10.1016/j.jclepro.2023.136748.
- [14] T. Ahmad, D. Zhang, C. Huang, H. Zhang, N. Dai, Y. Song, and H. Chen, “Artificial intelligence in sustainable energy industry: Status Quo, challenges and opportunities,” *Journal of Cleaner Production*, vol. 289, Art. no. 125834, 2021, doi: 10.1016/j.jclepro.2021.125834.
- [15] K. Kar, S. K. Choudhary, and V. K. Singh, “How can artificial intelligence impact sustainability: A systematic literature review,” *Journal of Cleaner Production*, vol. 376, Art. no. 134120, 2022, doi: 10.1016/j.jclepro.2022.134120.
- [16] Wu, H. Li, and B. Yuan, “AI-driven sustainable energy saving: Pathways for enhancing energy efficiency in Chinese listed firms,” *Applied Energy*, vol. 401, Art. no. 126607, 2025, doi: 10.1016/j.apenergy.2025.126607.
- [17] Z. Cao, R. Wang, X. Zhou, R. Tan, and others, “Adaptive Capacity Provisioning for Carbon-Aware Data Centers: A Digital Twin-Based Approach,” *IEEE Transactions on Sustainable Computing*, vol. 10, no. 4, pp. 730–743, 2025, doi: 10.1109/TSUSC.2025.3526192.
- [18] H. Huang, W. Lin, J. Lin, and K. Li, “Power Management Optimization for Data Centers: A Power Supply Perspective,” *IEEE Transactions on Sustainable Computing*, vol. 10, no. 4, pp. 784–803, 2025, doi: 10.1109/TSUSC.2025.3542779.
- [19] Nassar and R. Bashroush, “Toward a More Effective and Comprehensive Assessment of Data Center Environmental Impact,” *IEEE Transactions on Sustainable Computing*, vol. 11, no. 2, pp. 187–198, 2026, doi: 10.1109/TSUSC.2026.3666670.
- [20] J. Hu, J. Lu, J. M. Guerrero, and J. Cao, “Robust Spatiotemporal Optimal Scheduling for Distributed Data Centers With Multiple Uncertainties,” *IEEE Transactions on Sustainable Computing*, vol. 11, no. 2, pp. 173–186, 2026, doi: 10.1109/TSUSC.2026.3665882.
- [21] Samaye, G. Sassatelli, and A. Gamatié, “Formal Modeling and Analysis of Small-Scale Data Centers Integrating Renewable Energy Using Timed Automata,” *IEEE Transactions on Sustainable Computing*, vol. 11, no. 1, pp. 57–71, 2026, doi: 10.1109/TSUSC.2025.3645150.
- [22] W. Vanderbauwhede, “Modeling Scenarios for Carbon-Aware Geographic Load Shifting of Compute Workloads,” *IEEE Transactions on Sustainable Computing*, vol. 11, no. 4, pp. 402–413, 2026, doi: 10.1109/TSUSC.2026.3702249.
- [23] S. M. Hasan, T. Islam, M. Saif, K. R. Ahmed, et al., “Carbon Emission Quantification of Machine Learning: A Review,” *IEEE Transactions on Sustainable Computing*, vol. 10, no. 6, pp. 1085–1102, 2025, doi: 10.1109/TSUSC.2025.3578834.

- [24] X. Wang, R. Wu, L. Liu, C. Yi, et al., “Joint Energy and Computation Workload Management for Geo-Distributed Data Centers,” *IEEE Transactions on Green Communications and Networking*, vol. 9, no. 4, pp. 2115–2128, 2025, doi: 10.1109/TGCN.2025.3559505.
- [25] Canali, L. Chiaraviglio, R. Lancellotti, and M. Shojafar, “Joint Minimization of the Energy Costs From Computing, Data Transmission, and Migrations in Cloud Data Centers,” *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 2, pp. 580–595, 2018, doi: 10.1109/TGCN.2018.2796613.
- [26] L. Zhang, T. Han, and N. Ansari, “Energy-Aware Virtual Machine Management in Inter-Data-Center Networks Over Elastic Optical Infrastructure,” *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 1, pp. 305–315, 2018, doi: 10.1109/TGCN.2017.2771724.
- [27] S. Alanazi, M. Dabbagh, B. Hamdaoui, M. Guizani, and N. Zorba, “Reducing Data Center Energy Consumption Through Peak Shaving and Locked-In Energy Avoidance,” *IEEE Transactions on Green Communications and Networking*, vol. 1, no. 4, pp. 551–562, 2017, doi: 10.1109/TGCN.2017.2744602.
- [28] Z.-H. Zhou, M. Shojafar, M. Alazab, J. Abawajy, and F.-F. Li, “AFED-EF: An Energy-Efficient VM Allocation Algorithm for IoT Applications in a Cloud Data Center,” *IEEE Transactions on Green Communications and Networking*, 2021, doi: 10.1109/TGCN.2021.3067309.
- [29] Z. Wang, C. Guo, S. K. Bose, and G. Shen, “Frequency-Adaptive VDC Embedding to Minimize Energy Consumption of Data Centers,” *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 1, pp. 447–461, 2022, doi: 10.1109/TGCN.2021.3105934.
- [30] S. Pantazoglou, G. Tzortzakis, and A. Delis, “Decentralized and Energy-Efficient Workload Management in Enterprise Clouds,” *IEEE Transactions on Cloud Computing*, vol. 4, no. 2, pp. 196–209, 2016, doi: 10.1109/TCC.2015.2464817.