

Article

Carbon-Aware Digital-Twin Decision Support for Earthen Heritage: A Proof-of-Concept Case Study in Ushaiger, Saudi Arabia

Silvia Mazzetto ^{1,*}  and Haidar Hosamo ^{2,1} 

¹ Sustainable Architecture Laboratory, Department of Architecture, College of Architecture and Design, Prince Sultan University, Riyadh 12435, Saudi Arabia; haidar.hosamo@oslomet.no

² Department of Built Environment, Oslo Metropolitan University (OsloMet), 0130 Oslo, Norway

* Correspondence: smazzetto@psu.edu.sa

Abstract

Saudi Arabia's expanding rehabilitation of historic places requires decision tools that connect monitored performance with conservation priorities and carbon consequences. This study develops a carbon-aware heritage decision twin for a traditional mud-brick dwelling in Ushaiger using indoor–outdoor monitoring from September 2025 to September 2026. Quality control yielded 10,757 unique paired records, including 10,755 observations on a 17,109-step half-hour grid. A hybrid piecewise cubic Hermite interpolating polynomial (PCHIP)–Extreme Gradient Boosting (XGBoost 3.4.1) reconstruction was tested with artificial gaps of 7–61 days; the mean indoor-temperature mean absolute error (MAE) increased from 0.93 to 1.35 °C across these gap lengths. XGBoost with SHapley Additive exPlanations (SHAP 0.52.0) associated indoor temperature with accumulated outdoor forcing, with 3 and 24 h means receiving the largest contributions in the revised model. Non-dominated Sorting Genetic Algorithm II (NSGA-II) optimization at a fixed effective ventilation coupling reduced degree-hours above 35 °C by 2.62%; a joint coupling–persistence sensitivity produced 0.62–6.91% under the tested assumptions. Of 124 hot days, 45 entered the operational-ventilation class, 35 the combined or alternative reversible-measure classes, and 44 escalation review. Threshold and reconstruction-error tests showed that these counts remain scenario dependent. A screening life-cycle assessment (LCA) resulted in 9.7–29.1 kg carbon dioxide equivalent (CO₂e) for assisted ventilation versus 242.6–727.7 kg CO₂e for localized cooling within a fixed 588 h comparison window. The framework combines evidence confidence, explainable artificial intelligence (AI), passive-cooling opportunity, life-cycle carbon, and heritage compatibility to prioritize the least intrusive adequate action. The framework is presented as a single-building proof of concept; its decision architecture may support future multi-building applications, but its numerical thresholds and intervention rules require validation across additional Saudi heritage buildings, climatic contexts, and monitoring locations before portfolio-scale deployment.



Academic Editors: Sofia Agostinelli,
Daniele Groppi and Marco Casini

Received: 15 September 2026

Revised: 7 October 2026

Accepted: 15 October 2026

Published: 19 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: Ushaiger; earthen heritage; Saudi Arabia; digital twin; decision support; AI; XGBoost; SHAP; multi-objective optimization; LCA; embodied carbon; operational carbon; minimum intervention; adaptive reuse

1. Introduction

Saudi Arabia's heritage agenda is moving from preservation as an isolated conservation activity toward a broader model of rehabilitation, public access, adaptive reuse,

tourism, and living cultural heritage. Vision 2030 reporting emphasizes expanded restoration, documentation, and accessibility of historic places, while Heritage Commission program materials frame rehabilitation around cultural values, structural condition, compatibility of contemporary use, and the principle of adding or removing as little as possible [1,2]. The existence of a dedicated Saudi Historical Building Code (SBC 902) further reinforces that intervention decisions must be compatible with a specific regulatory context rather than treated as ordinary building retrofits [3].

This national context creates a decision problem. Earthen heritage buildings in central Saudi Arabia are culturally significant and inherently climate responsive, but their thermal behavior changes with season, building condition, opening state, solar exposure, and accumulated heat. Traditional Najdi settlements use thick mud-brick walls, compact morphology, shaded passages, courtyards, and controlled openings to moderate severe solar loads and large diurnal temperature swings [4–8]. Ushaiger Heritage Village is particularly relevant because restored and deteriorated dwellings coexist with active tourism, agricultural land, palm vegetation, open ground, and a wadi-edge microclimatic context [9].

Thermal mass can attenuate daytime extremes and improve passive survivability, but the same stored heat can delay nighttime recovery [10–14]. Consequently, a conservation manager needs more than an annual mean temperature or a comfort plot. The practical questions are sequential: Is the observed state reliable enough to act on? Is the building experiencing beneficial heat retention, ordinary hot conditions, or severe heat accumulation? Is cooler night air available? Can the problem be addressed through operation alone, or should a reversible intervention be tested? At what point should a case be escalated for detailed engineering and authority review?

Current research does not fully answer this sequence. Saudi studies have developed valuable frameworks for sustainable reuse and periodic evaluation of heritage buildings, including Ushaiger and Historic Jeddah [9,15,16]. Ushaiger also has a site-specific LCA benchmark comparing mud, limestone, decorative plaster, blended hydraulic cement, and ready-mix concrete for restoration [17]. This is an important local evidence base, but it is a material- and project-level assessment rather than a time-resolved operating rule. International research has separately demonstrated digital-twin thermal decision support [18], multi-criteria post-reuse evaluation [19], machine learning (ML)-assisted multi-objective retrofit optimization [20], and heritage-specific extensions of LCA that warn against treating cultural significance as an ordinary environmental metric [21]. In the reviewed literature, the missing link is, therefore, more specific than an absence of either AI or LCA: an operational framework is not yet evident that connects uncertain field evidence and state-dependent thermal need to the marginal life-cycle carbon of the least intrusive available action while keeping cultural significance and regulatory approval outside the optimizer.

A previous Ushaiger pilot linked monitoring, computer-aided design (CAD) documentation, simplified simulation, and digital-twin concepts [22]. The present work advances that pilot in six connected steps. First, extended monitoring gaps are reconstructed with quantified error while retaining a measured/reconstructed flag. Second, explainable AI identifies the thermal-memory variables governing the indoor response. Third, unsupervised learning converts the annual sequence into recurrent thermal states. Fourth, multi-objective optimization tests the conditional benefit and limitation of night ventilation. Fifth, these outputs are translated into a Saudi heritage action matrix that prioritizes monitoring, operational control, reversible interventions, and escalation review. Sixth, a screening-LCA gate compares the marginal product-stage modules A1–A3 and operational-energy module B6 burden of escalating actions using Ushaiger-specific material evidence and Saudi grid-carbon scenarios.

The novelty, therefore, lies less in applying AI or LCA as isolated tools and more in integrating evidence confidence, thermal-state diagnosis, cooling opportunity, marginal life-cycle carbon, and minimum-intervention action within one decision process. The carbon gate is deliberately subordinate to significance and safety: it asks which technically acceptable action carries the lowest additional carbon burden, but it never treats a material with low global warming potential (GWP) as permission to alter culturally significant fabric. This formulation also makes uncertainty part of the conservation decision itself: reconstructed periods can support seasonal planning and reversible trials, but they cannot independently justify fabric-altering action.

The study addresses the following seven research questions:

RQ1: Can extended monitoring gaps be reconstructed with quantified error rather than hidden by ordinary interpolation?

RQ2: How do linear and nonlinear data-driven models compare under seasonally blocked indoor-temperature validation?

RQ3: Which climatic and temporal variables dominate the learned indoor response?

RQ4: Can unsupervised learning reveal distinct thermal operating regimes that annual averages obscure?

RQ5: What trade-off between high-temperature exposure and intervention burden emerges when adaptive night ventilation is optimized?

RQ6: How can these outputs be translated into an evidence-gated, minimum-intervention decision framework relevant to heritage-building management in Saudi Arabia?

RQ7: How does marginal life-cycle carbon change the escalation hierarchy among operational, reversible, localized active, and fabric-based interventions?

2. Background and Research Gap

2.1. Thermal Performance of Earthen Buildings

Earthen construction moderates heat through a combination of thermal storage, delayed heat transfer, moisture exchange, and interaction with ventilation and solar exposure. Reviews of monitored earthen buildings show that wall thickness, moisture state, finishes, roof construction, openings, and climatic boundary conditions can change the apparent benefit of thermal mass substantially [10,23]. Hot-climate studies likewise show that large thermal inertia can attenuate daytime peaks without guaranteeing acceptable indoor conditions during prolonged heat events [11,12,14]. Recent two-year monitoring of historic adobe structures confirms the value of time lag and decrement factor as physical indicators rather than relying on mean temperature alone [24].

For Najdi heritage, the evidence base remains much thinner. Traditional dwellings combine thick mud-brick walls, compact urban form, courtyards, shaded circulation, and controlled openings, but their present-day condition introduces cracks, missing doors, roof damage, uncontrolled airflow, and changed occupancy [4–7]. The earlier Ushaiger pilot established short-term monitored evidence and a digital-twin feasibility workflow but explicitly lacked a seasonal record capable of testing whether the observed buffering persists across winter, transition, and extreme-summer conditions [22]. The first gap is, therefore, not a lack of knowledge that adobe has thermal mass; it is a lack of long-duration, state-resolved evidence for deteriorated Najdi buildings that distinguishes mean offset, amplitude attenuation, time lag, and nighttime heat release.

Beyond wall mass alone, the environmental performance of earthen heritage should be interpreted as the result of a coupled material–architectural–microclimatic system. Studies of adobe, cob, rammed earth, and other earth-based constructions show that thermal inertia, moisture exchange, and low-impact material cycles can provide important environmental benefits, but their effectiveness remains strongly dependent on material composition, mois-

ture state, ventilation, shading, and climatic exposure [25–29]. Courtyard configuration, shaded circulation and controlled natural ventilation further influence heat dissipation and indoor environmental conditions in hot climates, including Saudi and wider regional heritage contexts [30–33]. From a conservation and adaptive-reuse perspective, these environmental mechanisms should be considered together with material compatibility, minimum intervention and cultural continuity, favoring compatible repair, roof and envelope maintenance, reversible shading and controlled openings before more intrusive measures are introduced [15,34,35]. In Ushaiger, this broader perspective is particularly relevant at the agricultural and wadi edge, where vegetation, irrigation, exposed ground, soil moisture and local airflow may modify the boundary conditions affecting heat and moisture exchange; however, because these landscape drivers were not independently measured in the present campaign, they are treated as contextual microclimatic influences rather than demonstrated causal effects.

2.2. Monitoring Continuity and Data Reconstruction

Continuous heritage monitoring is difficult because sensors operate in buildings with restricted access, weak infrastructure, maintenance interruptions, and changing exposure. Building ML reviews increasingly identify preprocessing, missing-value treatment, temporal feature engineering, and validation design as determinants of model reliability rather than minor technical details [36]. This is especially important for long gaps: ordinary linear interpolation can create visually complete records while understating uncertainty and suppressing real seasonal structure.

The methodological gap is consequently one of evidence provenance. Building-AI studies commonly emphasize predictive accuracy on available records, whereas conservation decisions also require knowing whether each state is directly measured, reconstructed, or counterfactual. The present study addresses this by retaining a measured/reconstructed flag, validating reconstruction with deliberately withheld seasonal blocks, and preventing reconstructed evidence from independently triggering fabric-altering action. This treatment aligns data quality with decision authority rather than treating imputation as invisible preprocessing.

2.3. Heritage Digital Twins

Digital twins in heritage are moving beyond static building information modeling (BIM)-based heritage documentation toward sensor integration, predictive diagnostics, maintenance, and operational decision support. Systematic reviews, nevertheless, describe the field as immature, constrained by sparse empirical data, heterogeneous data capture, multi-source integration, incomplete validation, and limited performance-based management logic [37,38]. Earlier scan-to-model studies establish an important geometric evidence layer. Rolfsen et al. compared 3D laser-scanning data with BIM reference geometry for dimensional quality assurance, demonstrating how model-to-reality verification can expose deviations in the as-built state [39]. In a directly heritage-specific context, Shabani et al. combined expert preliminary survey with 3D laser scanning of historic timber log houses to document geometry, structural details, condition, and vulnerability-relevant evidence [40]. A recent heritage digital-twin case has further shown that sensed data and simulation can guide thermal interventions while preserving building fabric [18].

International applications to earthen heritage provide further relevant precedents. Puertas et al. developed a preventive-conservation Digital Twin for the rammed-earth Tower of Muhammad at the Alhambra in Granada [41], combining continuous monitoring, architectural inspections, nondestructive testing, numerical modeling, and data fusion. Similarly, Wang and Li established a dynamic Digital Twin for the rammed-earth Great Wall

in China [42], integrating point-cloud-derived deterioration indicators into a monitoring–assessment–warning framework for preventive conservation. These studies show the growing use of Digital Twin methods for earthen heritage in different geographical contexts, although they primarily emphasize structural condition and deterioration monitoring rather than the integrated thermal, operational, carbon, and minimum-intervention decision process developed in the present study. Together, these studies establish credible geometric, condition, and sensing foundations, while the reviewed heritage digital-twin literature still reports limited integration of longitudinal performance evidence with explicit intervention-level decision logic [37,38].

Building digital-twin research outside heritage provides stronger evidence that integrated sensing, BIM, probabilistic models, and ML can support operational decisions rather than visualization alone. Hosamo et al. demonstrated a digital-twin approach that linked BIM, real-time building data, occupant feedback, and predictive maintenance to decision-making for indoor performance [43]. That capability is transferable conceptually, but heritage adds constraints that ordinary facility management does not: authenticity, reversibility, minimum intervention, structural fragility, and authority review. The second major gap is, therefore, a lack of an uncertainty-aware heritage twin that converts environmental evidence into an explicit level of intervention, not merely a predicted temperature or an optimized set-point.

Parallel developments in monitored simulation and data-driven operational control further demonstrate how high-resolution environmental data can support predictive assessment and scenario testing. Data-driven ventilation and indoor-climate analysis have shown the potential of monitored environmental information to inform operational building decisions [44]. Similarly, monitored IDA ICE and dynamic building-modeling studies demonstrate the importance of reliable boundary conditions, calibration and scenario definition when evaluating indoor climate, retrofit performance and overheating risk [45–48]. For heritage applications, these precedents support using simulation as a subsequent validation and scenario-testing layer rather than as a replacement for measured evidence, particularly where conservation decisions depend on distinguishing directly observed conditions from reconstructed or counterfactual states.

2.4. Explainable AI and Optimization

ML and multi-objective optimization are increasingly combined to reduce the computational burden of retrofit exploration. Recent historic-building work integrates interpretable ML with multi-objective optimization across energy, cost, carbon, and comfort [20]. More generally, uncertainty-aware retrofit optimization now combines measured-data calibration, NSGA-II, XGBoost surrogates, and SHAP interpretation [49]. These studies demonstrate that explainability and Pareto analysis can make complex retrofit trade-offs more transparent.

The cited optimization studies remain dominated by conventional building-performance objectives, such as energy, cost, carbon, and comfort [20,49]. They do not explicitly test whether the same thermal service can be achieved through operation before equipment, through a reversible measure before fabric alteration, or whether reconstructed evidence is strong enough to justify escalation. Natural-ventilation research also confirms that performance depends strongly on timing, temperature difference, opening configuration, wind, and climatic context [50,51]. The optimization gap addressed here is, therefore, narrower and more conservation-specific: optimize the least intrusive feasible thermal response while treating heritage intervention burden and evidence confidence as decision constraints rather than as ordinary weighted objectives.

2.5. Saudi Heritage Decision Frameworks

Saudi heritage governance already supplies important normative boundary conditions. Heritage Commission program materials emphasize cultural-value identification, structural assessment, compatible use, conservation principles, and minimum intervention, while the Saudi Historical Building Code creates a dedicated regulatory context for historic structures [2,3]. Research on Historic Jeddah similarly calls for recurring assessment across function, fabric, safety, legal, economic, and heritage dimensions [16]. Ushaiger studies show that successful reuse depends on cultural, environmental, social, and economic performance together [9,15].

The reviewed Saudi heritage sources do not provide a quantitative daily operating rule between those principles and monitored thermal evidence [2,9,15,16]. A heritage manager can accept the principle of minimum intervention yet still lack a defensible trigger for moving from monitoring to opening control, reversible shading, use scheduling, localized cooling, or detailed retrofit design. Multi-criteria heritage methods can rank alternatives [19], but they are normally applied at project appraisal rather than updated from a continuous environmental state. The Saudi decision gap is, thus, an evidence-to-action translation problem: how to make routine monitoring informative enough for triage without allowing an algorithm to override cultural significance or regulatory review.

2.6. Life-Cycle Carbon in Heritage Interventions

The carbon evidence is also fragmented. Ushaiger already has a site-specific LCA comparing mud, limestone, decorative plaster, blended hydraulic cement, and ready-mix concrete for restoration using an 80 m² external-wall functional unit (FU) [17]. Saudi building LCA studies show that electricity-intensive operation can dominate life-cycle impacts in cooling-dependent buildings [52]. At the same time, heritage-LCA research warns that conventional environmental indicators do not capture authenticity, cultural significance, traditional-material specificity, or restoration logic adequately [21]. Green-heritage scholarship, therefore, places minimal intervention and prevention alongside environmental assessment rather than beneath it [53].

A further issue is temporal uncertainty. Hosamo et al. show that dynamic building LCA is affected by scenario, parameter, technology, operational, and future-grid uncertainties and argue for closer integration of BIM, monitoring, Internet of Things sensing, and predictive analytics in LCA [54]. That observation is directly relevant here because the carbon consequence of a thermal decision depends on when, how often, and for how many years an intervention operates. Across the reviewed LCA and heritage-decision literature, the unresolved gap is, therefore, not the absence of an LCA number; it is the lack of a rule that activates embodied- or operational-carbon comparison only after a thermal need has been verified and then asks whether the additional carbon is proportionate to the required heritage service [17,21,54].

2.7. Research Gap and Contribution

The literature, therefore, contains strong components but weak integration. Table 1 summarizes how the principal evidence streams lead to the seven research questions and clarifies what this study adds before introducing the methodological workflow.

Taken together, the research gap is a missing carbon-aware evidence-to-action bridge for Saudi earthen heritage. The contribution is an integrated decision architecture that combines the following:

1. Provenance-aware monitoring and reconstruction uncertainty;
2. Explainable AI thermal diagnosis;
3. State-dependent passive-cooling opportunity;

4. Multi-objective testing of a reversible operational strategy;
5. Marginal life-cycle carbon of escalation;
6. A minimum-intervention hierarchy in which significance, safety, compatibility, and regulatory review remain hard constraints.

Table 1. Literature-supported research gaps and this study's corresponding contribution.

Evidence Stream	What Is Established	Remaining Gap	Response in This Study
Earthen thermal performance	Thermal mass, time lag and attenuation are climate and condition dependent [10,24]	Limited annual, state-resolved evidence for deteriorated Najdi dwellings	Seasonal monitoring metrics and thermal-regime analysis
Monitoring and ML data quality	Missing-data treatment and preprocessing materially affect model robustness [36]	Long gaps are often filled without carrying uncertainty into the decision	Blind-tested reconstruction with measured/reconstructed provenance
Digital twins	BIM and laser scanning support geometric verification and heritage documentation, while sensor-connected twins support prediction and building decisions [37–40,43]	Heritage twins rarely translate multi-source evidence into an intervention level	Evidence-gated decision twin with explicit escalation states
Explainable optimization	SHAP and Pareto methods expose performance trade-offs [20,49]	Heritage intervention intensity and evidence confidence are usually outside the objective structure	Explainable thermal surrogate plus minimum-intervention NSGA-II screening
Saudi heritage management	Minimum intervention, compatibility and recurring assessment are established principles [2,16]	Reviewed sources do not provide a quantitative daily operating rule linking sensor evidence to those principles	Saudi heritage action matrix for monitoring, reversible action and escalation
Life-cycle carbon	Local Ushaiger material LCA and Saudi operational-carbon evidence exist [17,52]	Carbon is not linked to state-dependent thermal need and may conflict with cultural value	Marginal A1–A3/B6 carbon gate subordinate to heritage compatibility
Dynamic LCA and uncertainty	Time, technology and scenario assumptions can change building LCA results [54]	Static carbon comparisons do not reflect repeated seasonal control decisions	588 h seasonal and 20-year carbon sensitivity linked to the action ladder

The framework is a research decision-support layer rather than an official Heritage Commission or Saudi Building Code procedure.

3. Materials and Methods

3.1. Case Study and Digital Representation

The case study is a traditional mud-brick dwelling in Ushaiger Heritage Village, located close to the interface between compact settlement fabric, cultivated land, and the wadi system. At the time of monitoring, the dwelling was abandoned and partially deteriorated, with damaged or missing envelope elements and no regular occupant use. Consequently, no controlled opening schedule or occupancy pattern was available for direct logging; airflow through existing openings and envelope discontinuities, therefore, remained an uncontrolled boundary condition.

Figure 1 shows the location and ground-floor configuration, while Figure 2 links the architectural plan to the indoor and outdoor monitoring positions. The supplied 3D CAD

model, shown in Figure 3, provides the geometric layer of the digital representation. Its role as an evidence-organizing reference is conceptually consistent with scan/BIM verification workflows and heritage preliminary-survey practice, where geometric models support comparison, condition documentation, and subsequent assessment [39,40]. In the present study, this geometry is used to organize sensors, envelope zones, and future conservation scenarios; the new AI results are derived from the measured environmental record rather than from an uncalibrated physics-based simulation.

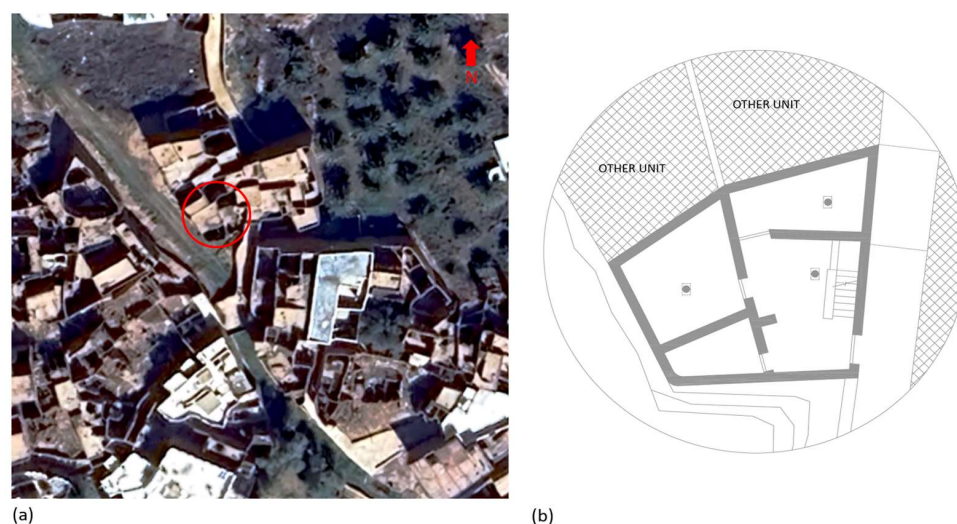


Figure 1. Location of the monitored wadi-edge dwelling and ground-floor spatial configuration derived from the architectural documentation. (a) View of the site; (b) Plan of the building.



Figure 2. Ground-floor monitoring layout and field photographs of the sensor locations. (a) External sensor location; (b) Internal sensor location.

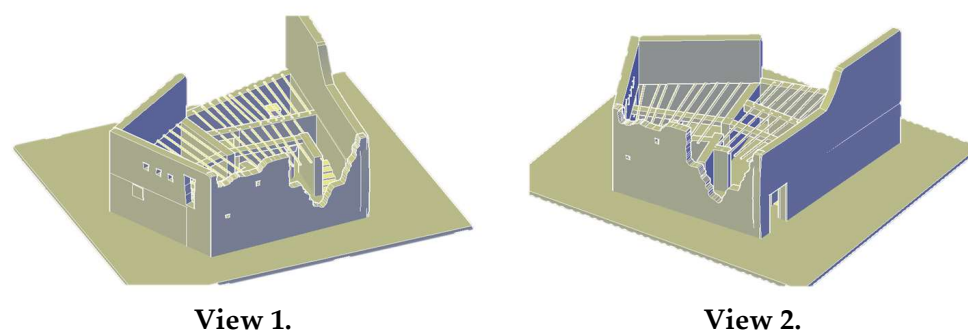


Figure 3. CAD-based 3D representation of the monitored dwelling. The geometry forms the spatial layer for the data-driven digital-twin workflow.

3.2. Monitoring and Data Quality

The monitoring platform follows the instrumental configuration reported for the previous Ushaiger campaign [22]. Indoor variables include air temperature, relative humidity (RH), carbon dioxide (CO₂), barometric pressure, and noise, while the outdoor module records air temperature and RH. The supplied exports have a nominal 30 min cadence. The consumer-grade sensor accuracy is approximately ± 0.3 °C for temperature and ± 3 percentage points for RH; the results are, therefore, interpreted primarily through repeated patterns, indoor–outdoor contrasts, and seasonal behavior rather than reference-grade metrology.

Six source files cover 10 September 2025 to 1 September 2026. A row was retained for paired environmental analysis when indoor temperature, outdoor temperature, indoor RH, outdoor RH, and indoor CO₂ were numeric. Sixty overlapping timestamps were removed, leaving 10,757 unique valid rows. Two isolated bridge records occurred on the hour rather than at 15/45; they were retained in the descriptive audit but excluded from the regular-grid reconstruction. The reconstruction grid, therefore, contains 17,109 30 min timestamps, of which 10,755 are measured, and 6354 require reconstruction. The resulting measured-data coverage and the two principal interruptions are shown in Figure 4.

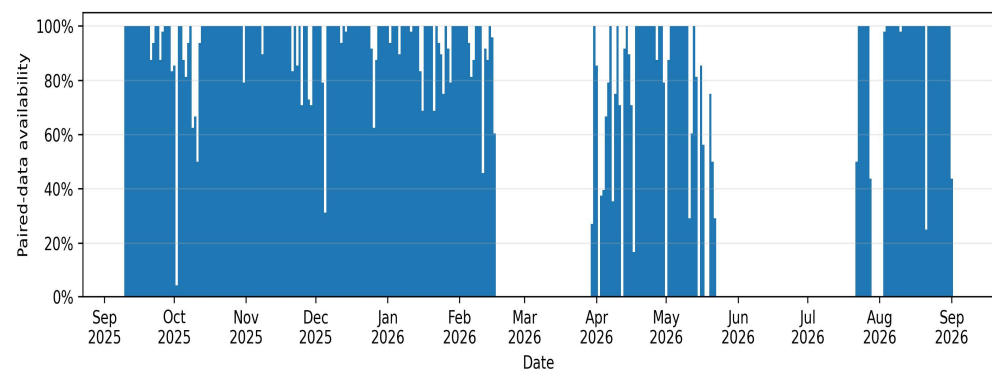


Figure 4. Measured daily paired-data availability before reconstruction. The two major interruptions are visible from mid-February to late March and from late May to late July 2026.

The deterministic analysis of the supplied exports is specified through implementation settings, withheld dates, daily classifications, and optimization outputs in Appendix B. Duplicate timestamps are resolved by retaining the first valid record in the ordered source files. The two off-grid bridge observations remain available for descriptive auditing but do not enter regular-grid models.

For supplementary climatic sensitivity analysis, hourly Open-Meteo regional climate data were supplied for the Ushaiger location over the monitoring period [55]. The external dataset included shortwave solar radiation, 10 m wind speed, and 10 m wind direction. These variables were used only as regional boundary-condition indicators and were not treated as direct site measurements because the regional data cannot resolve building-scale shading, courtyard airflow, vegetation, or opening conditions.

3.3. Hybrid Data Reconstruction

Long gaps were reconstructed as a slow daily component plus a learned diurnal residual. For indoor temperature, outdoor temperature, indoor RH, and outdoor RH, daily means were calculated from days containing at least 12 valid paired half-hour observations. PCHIP interpolates these daily anchors and is evaluated at calendar-day midnight; the resulting backbone value is applied to all timestamps in that day. This explicit convention avoids ambiguity about sub-daily evaluation of the seasonal trend.

The sub-daily residual $r_t = y_t - \bar{y}_d$ was then learned with XGBoost [56] using cyclic hour-of-day and day-of-year features and their first-order interactions. The completed value is

$$\hat{y}_t = \bar{y}_d + \hat{r}_t.$$

Measured values always replace modeled values where observations exist, so reconstruction does not modify the original record. Each timestamp also carries a measured/reconstructed flag. Because the PCHIP backbone uses observed daily states on both sides of each interruption, the procedure is a retrospective annual reconstruction for analysis and planning rather than an online forecasting algorithm. A live decision twin would require causal filtering or forward prediction that only uses information available up to the decision time.

The residual XGBoost models use 300 trees, a maximum depth of 3, a learning rate of 0.04, a subsample of 0.9, full feature sampling, an L2 regularization of 1, histogram tree construction, and a random seed of 2026. The predictors are sine/cosine hour and day-of-year terms and their four pairwise hour-year interactions. The residual training targets subtract the observed mean of the same calendar day. The reconstructed RH is bounded to 0–100%; the measured values are retained exactly.

Gap-length sensitivity was evaluated with 7-, 14-, 28-, 42-, and 61-day artificial interruptions. Six eligible windows per length were selected deterministically across the available dates; each had at least 75% paired observations and at least 12 measured records on every day before masking. All four variables were removed throughout each test window, and both the backbone and residual models were refitted from the remaining record. Errors are scored only against withheld measurements. Some long windows overlap, so they are stress tests rather than independent replicates. The eligible 28–61-day windows fall in October–February; the data do not permit an equally long measured summer test. Appendix B reports dates, coverage, and error by gap length.

3.4. Explainable Temperature Modeling

A separate indoor-temperature surrogate was fitted for interpretation. Its predictors are measured outdoor temperature and RH, trailing 3, 6, 12, and 24 h outdoor-temperature means, the trailing 24 h temperature range, and sine/cosine annual and diurnal terms. A row is eligible only when the target is measured, and every outdoor rolling window is complete with measured values; 8314 rows satisfy this rule. No indoor-temperature lag or reconstructed predictor is included. The retrospective blocked tests may use training records before and after a withheld week and are not forward-forecast tests.

Ridge regression, random forest, and XGBoost were evaluated on identical seven-day blocks starting on 20 September, 5 November, and 15 December 2025 and 15 January and 10 August 2026. Ridge uses standardized features and a regularization of 10. Random forest uses 300 trees, a maximum depth of 12, and a minimum leaf size of 5. XGBoost uses the reconstruction settings with 500 trees. SHAP is evaluated for all 8314 eligible measured targets using tree-path-dependent attribution [57]. Equal-season samples of 300, 900, and 2000 observations were drawn without replacement 30 times per size to test the rank stability. A separate SHAP summary uses the 1830 eligible measured observations above 35 °C. No June target is measured, so sampling cannot create June evidence.

3.5. Thermal Regime Identification

Seven standardized daily variables were clustered: mean indoor and outdoor temperature, their mean difference, indoor and outdoor diurnal amplitudes, and indoor and outdoor absolute humidity. Candidate cluster counts $k = 2$ –6 were compared with inertia, silhouette score, and the Davies–Bouldin index. Each fit uses 25 k-means initializations.

Stability was assessed with 20 additional seeds and 20 bootstrap resamples; bootstrap-fitted models predict labels for all original days, which are compared with the reference labels using the adjusted Rand index (ARI). The four-cluster solution is retained as an operational partition, while the statistical preference for a simpler solution is reported explicitly. Cluster names are assigned from centroid temperatures and offsets after fitting. These clusters and the A–D decision quadrants are separate classifications.

3.6. Multi-Objective Ventilation Optimization

The optimization stage tests a virtual, reversible night-ventilation intervention from 1 May to 1 September 2026. It is a counterfactual digital-twin scenario rather than a claim that the monitored building was operated according to the optimized schedule. The measured/reconstructed indoor series provides the no-intervention baseline, $T_b(t)$. A cooling-deviation state, δ_t , represents the incremental effect of a controlled opening strategy:

$$T_{cf,t} = T_{b,t} + \delta_t,$$

$$\delta_{t+1} = \begin{cases} \rho\delta_t + \eta(T_{out,t} - T_{cf,t}), & \text{ventilation active,} \\ \rho\delta_t, & \text{otherwise,} \end{cases}$$

where $\rho = 0.865$ is the previously reported reference half-hour persistence of residual indoor thermal deviations, and η is an effective coupling parameter rather than a directly measured air-change rate. Ventilation is permitted only within the candidate night window, when the outdoor air is cooler than the current counterfactual indoor state by at least the optimized trigger, ΔT_{tr} , the outdoor RH is below 70%, and the outdoor absolute humidity does not exceed the indoor absolute humidity by more than 2 g m^{-3} .

The effective coupling is a scenario assumption: values of 0.02–0.10 represent a correction of 2–10% of the instantaneous indoor–outdoor temperature difference per active half-hour step. They are not measured airflow rates or experimentally calibrated bounds. In an isolated first-order relaxation interpretation, this fraction corresponds to approximately 24.7–4.7 h e-folding times, but the full counterfactual also includes residual persistence. The reference values are $\eta = 0.061$ and $\rho = 0.865$. A joint grid tests $\eta = 0.02, 0.03, 0.05, 0.061, 0.07, 0.09$, and 0.10 against $\rho = 0.75, 0.80, 0.865, 0.90$, and 0.95 ; physical calibration remains necessary.

Three schedule variables are optimized: start hour (17:00–23:00, 0.5 h resolution), available duration (4–12 h, 0.5 h resolution), and cooling trigger (0.5–5.0 °C, 0.1-°C resolution). Effective coupling is fixed within each run. The two objectives are degree-hours above the screening temperature and active ventilation hours. Removing η from the burden objective prevents the same parameter from simultaneously defining physical efficacy and intervention cost. Active hours are an operating-effort proxy, not a quantified impact on heritage fabric. The reference run uses $\eta = 0.061$, with separate reoptimized runs at 0.03 and 0.09. The 35 °C threshold is a case-specific high-temperature exposure screen, not a formal comfort limit. A formal thermal-comfort model was not applied because the monitored dwelling was abandoned during the campaign and, therefore, had no stable occupant population, occupancy schedule, clothing level, metabolic activity, or subjective thermal-response data. In addition, the available monitoring did not include all environmental variables required for a robust PMV-type assessment, particularly local air speed and mean radiant temperature. The 35 °C metric is, therefore, used only as a first-stage heat-exposure screening indicator to identify periods requiring further assessment; it should not be interpreted as defining human thermal acceptability. For future occupied or tourism-related reuse, complement the screening layer with an adaptive-comfort assessment and field-based thermal-sensation feedback collected under actual patterns of use.

NSGA-II uses a population of 100 for 200 generations, simulated binary crossover probability of 0.9, distribution index of 15, polynomial mutation probability of 1 per individual, 1/3 per variable, and a distribution index of 20 [58]. Random seeds 2026–2030 are tested; duplicate genotypes are removed. Candidate schedules are rounded to the declared search resolutions before evaluation. Cooling generated at step t affects step $t + 1$, with zero initial deviation. The knee minimizes Euclidean distance to the ideal pair after each objective is normalized by its non-dominated-set range. Convergence is assessed using normalized hypervolume with reference point (1.05, 1.05). An exhaustive check evaluates all 10,166 schedules on the declared grid. The temperature screen is also varied to 30, 32, 37, and 40 °C, with schedule reoptimization at each value.

3.7. Heritage Decision-Support Framework

The final analytical layer translates the digital-twin outputs into a conservation screening decision. Figure 5 organizes this layer into three linked parts: non-negotiable core criteria, five sequential decision gates, and a five-level intervention-maturity ladder. Heritage significance, structural condition, minimum intervention and reversibility, SBC 902/authority review, and grid/LCA context constrain every decision before analytical outputs are allowed to trigger escalation. Table 2 shows how these Saudi heritage principles are converted into computational rules. The framework deliberately separates thermal need from permission to intervene: a severe thermal state can justify escalation for investigation, but it does not automatically justify alteration of significant fabric.

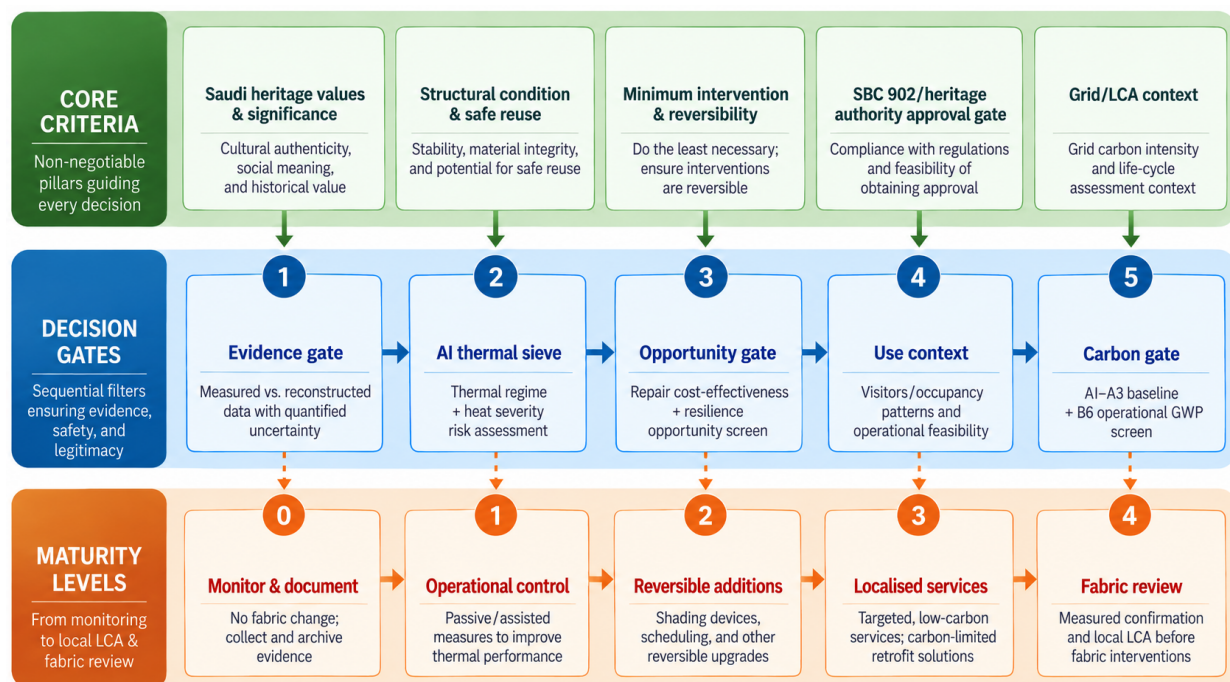


Figure 5. Carbon-aware Saudi heritage decision-support layer proposed in this study. Non-negotiable heritage and safety criteria constrain five sequential evidence, thermal, opportunity, use-context, and carbon gates; the resulting maturity ladder escalates from monitoring to operational control, reversible additions, localized services, and finally fabric review. The green upper layer shows the core criteria that guide all decisions. The blue middle layer shows the five decision gates: evidence, thermal assessment, opportunity, use context, and carbon. The orange lower layer shows the intervention levels, progressing from monitoring to operational control, reversible additions, localized services, and, finally, fabric review. The colors, therefore, separate conservation criteria, decision steps, and intervention levels.

Table 2. Translation of Saudi heritage decision principles into the proposed digital-twin logic.

Heritage Decision Principle	Digital-Twin Implementation	Decision Consequence
Cultural value and significance	Heritage significance remains an external constraint rather than an optimization variable inferred from sensors	Thermal benefit alone cannot justify modification of significant fabric
Structural condition and safe reuse	Structural assessment is a hard gate before Level 4 action	A thermal alert can trigger inspection, not structural intervention
Compatible contemporary function	Use context includes visitor timing, occupancy, and feasible opening periods	Operational scheduling can be tested before physical alteration
Minimum intervention and reversibility	Escalation ladder progressing from monitoring and operational control to reversible additions, localized services, and, only when necessary, fabric review	Least-intrusive effective action is prioritized
Documentation and traceability	Every timestamp retains measured/reconstructed provenance and daily evidence confidence	Reconstructed periods support planning but require measured confirmation for irreversible action
Carbon proportionality	Marginal A1–A3 and B6 GWP are screened only after thermal need and action feasibility are established	Where heritage-compatible actions provide similar service, the lower-carbon option is preferred
Regulatory compliance	SBC 902/authority review sits outside and above the analytical optimizer	The digital twin supports screening and prioritization; it does not replace approval

Daily heat severity is quantified as

$$H_d = \sum_{t \in d} \max(T_{in,t} - 35, 0) \Delta t,$$

where $\Delta t = 0.5$ h. The daily night-cooling opportunity, C_d , is the number of hours between 20:00 and 07:00 that satisfy a fixed diagnostic operating screen: an outdoor temperature of at least 1.40 °C below the indoor temperature, an outdoor RH below 70%, and an outdoor absolute humidity of no more than 2 g m^{−3} above the indoor absolute humidity. The 35 °C value remains a study-specific high-temperature exposure indicator, not a statutory comfort threshold.

The 20:00–07:00 opportunity window and 1.40 °C temperature advantage are held fixed as diagnostic definitions when the reference controller is revised. They describe the available cooling opportunity and are not asserted to be the newly selected optimum. The 35 °C value was selected to retain comparability with the initial severe-heat screening case; it was not estimated as a physiological or statutory boundary. Threshold sensitivity reports both locally recalculated hot-day medians and the fixed 35 °C median splits.

For the 124 days with $H_d > 0$, the completed-series medians, $H^* = 37.741$ degree-hours/day and $C^* = 6.5$ h/day are used as case-specific screening splits rather than universal policy limits. Four hot-day decision classes follow: (A) lower heat severity with higher cooling opportunity, favoring operational night ventilation; (B) higher severity with higher opportunity, favoring ventilation plus reversible solar control; (C) lower severity with lower opportunity, favoring reversible shading and/or use scheduling; and (D) higher severity with lower opportunity, triggering escalation review and consideration of localized services in low-significance spaces. Days with $H_d = 0$ remain at Level 0 routine monitoring.

A second gate captures evidence confidence using the fraction of each day that is directly measured: high confidence >75%, moderate confidence 25–75%, and low confidence <25%. Low-confidence reconstructed states may trigger monitoring, additional sensing, or

reversible trials, but they are not treated as sufficient evidence for irreversible fabric change. Any Level 4 envelope or structural intervention remains conditional on field inspection, significance assessment, structural evaluation, and applicable authority/code review.

Evidence confidence uses measured records divided by 48 nominal half-hour positions per calendar day, including the partial final day. Class A denotes $H \leq H^*$ and $C \geq C^*$; B denotes $H > H^*$ and $C \geq C^*$; C denotes $H \leq H^*$ and $C < C^*$; and D denotes $H > H^*$ and $C < C^*$. Equalities are, therefore, assigned deterministically. The partial final day is retained for transparency, with incomplete-day evidence flagged.

The opportunity gate in Figure 5 asks whether repair/maintenance, passive night cooling, reversible shading, or use scheduling can provide an adequate and resilient response before services or fabric change are considered. “Repair cost-effectiveness” is used here as a qualitative proportionality screen rather than a monetized life-cycle-cost calculation, because project-specific repair quantities and Saudi cost data were not available. The maturity levels, therefore, represent escalation authority, not a numerical ranking in which a higher level is automatically preferable. The framework is visually coded by function: green identifies non-negotiable conservation constraints, blue identifies the sequential analytical decision gates, and orange identifies the escalating intervention-maturity levels.

Uncertainty propagation perturbs reconstructed values only. First, systematic indoor-temperature shifts of -0.97 and $+0.97$ °C test the previously used error scale. Second, 200 draws resample contiguous seven-day, four-variable residual blocks from the 28–61-day blind tests, preserving within-block temporal and cross-variable dependence; RH remains bounded to 0–100%. The reference class splits and controller are held fixed. Annual and June/July A/D counts and ventilation benefit are recalculated. The resulting 5th–95th percentile envelopes are scenario intervals rather than calibrated confidence intervals, since long summer gaps have no matching ground truth.

3.8. Screening LCA and Carbon Constraint

The carbon layer is a transparent screening LCA rather than a full comparative building LCA. Its purpose is to test whether the intervention ladder remains rational after adding life-cycle carbon to the thermal decision, not to declare environmental equivalence between fundamentally different conservation actions. The screen, therefore, quantifies only the life-cycle modules for which defensible intervention-scale information is available: A1–A3 for newly added material and B6 for electricity used by assisted ventilation or localized cooling. Transport, installation, maintenance other than the stated shading replacement, refrigerant leakage, heating, ventilation, air-conditioning (HVAC)-embodied impacts, and end-of-life modules are excluded because project-specific quantities and equipment specifications are not yet available. These exclusions make the active-cooling results conservative rather than punitive.

For an intervention scenario, s , the incremental cradle-to-gate material burden is calculated from the quantity of each added product and its A1–A3 emission factor:

$$GWP_{A1-A3,s} = \sum_{i=1}^n q_{i,s} EF_{i,A1-A3},$$

where $q_{i,s}$ is the scenario quantity expressed in the reference unit of the corresponding environmental product declaration (EPD) or LCA dataset, and $EF_{i,A1-A3}$ is its cradle-to-gate GWP factor. For the reversible timber-screen example, the material quantity is obtained from area and equivalent solid thickness, $q = At = 10 \text{ m}^2 \times 0.015 \text{ m} = 0.15 \text{ m}^3$, per installation.

Annual operational carbon for an electricity-using intervention is calculated as

$$GWP_{B6,yr} = P_{el} h_{op} EF_{grid},$$

where P_{el} is the electrical input (kW), h_{op} is the annual operating time (h), and EF_{grid} is the grid emissions intensity (kg CO₂e/kWh). The common operating window is 588 h, a fixed reference comparison window independent of the revised controller's active hours. This common window isolates the carbon consequence of power intensity rather than assuming that every option would operate continuously. Saudi Electricity Company (SEC) reports a 2024 grid emissions intensity of 0.55 t CO₂e/MWh (numerically 0.55 kg CO₂e/kWh) and a 2030 target of 0.40t CO₂e/MWh [59]; both are evaluated.

For a decision horizon, H , and a replaceable material intervention with service cycle, L_s , the number of installations and the cumulative A1–A3 burden are

$$N_s = \left\lceil \frac{H}{L_s} \right\rceil, \quad GWP_{A1-A3,H} = N_s GWP_{A1-A3,s}$$

The corresponding screening total over the horizon is

$$GWP_{screen,H} = GWP_{A1-A3,H} + H GWP_{B6,yr}$$

Equation is applied only to modules explicitly included for each scenario. For the timber shade, $GWP_{B6,yr} = 0$ in the present screen; for assisted ventilation and localized cooling, equipment A1–A3 is unavailable and is, therefore, set outside the quantified boundary rather than assumed to be zero. Consequently, the active-service totals are lower-bound screening values.

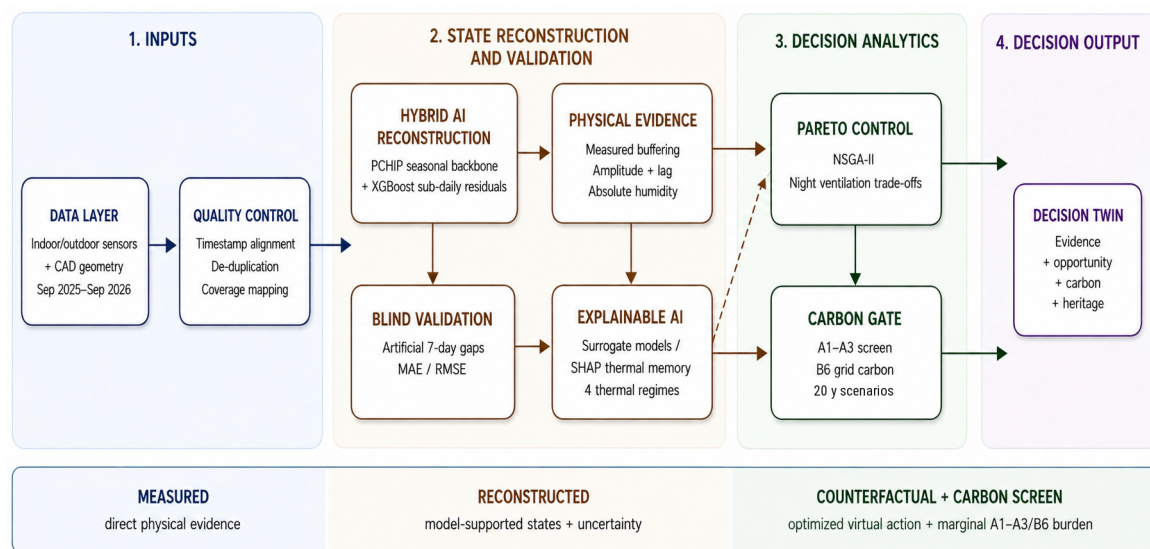
Three intervention families are screened. Assisted night ventilation is parameterized at 30–90 W to represent low-power fans where secure natural opening is insufficient. Localized active cooling is parameterized at 0.75–2.25 kW as a deliberately broad service-power sensitivity for a limited refuge/occupied zone; these are scenario inputs, not measured equipment ratings. A reversible timber solar screen is represented by 10 m² of shading with 15 mm equivalent solid timber thickness (0.15 m³ per installation). The Stora Enso EPD guide reports an A1–A3 fossil GWP of 33.9 kg CO₂e/m³ for Classic Sawn timber [60]. This value is a fossil GWP (GWP-fossil) manufacturing proxy, not a total GWP (GWP-total) result under European Standard EN 15804:2012+A2:2019 [61] including biogenic-carbon accounting; it is, therefore, used here only as a positive-emissions screening input and should not be interpreted as the complete climate balance of timber. A 10-year service cycle is assumed, resulting in two installations over the 20-year decision horizon. The published Ushaiger material LCA is retained separately as a local contextual benchmark rather than added to the shading/energy scenarios, because its products have different functions and its FU is an 80 m² external wall over a 100-year building life [17].

The assumptions used in the screening calculations are summarized in Table 3. They are intentionally visible so that future project-specific EPDs, equipment powers, operating hours, service lives, and transport distances can replace the present screening inputs without changing the decision architecture.

Figure 6 summarizes the complete analytical sequence. It separates four stages: monitored/CAD inputs and quality control; uncertainty-aware state reconstruction and physical validation; explainable AI, Pareto control, and the carbon gate; and a final heritage decision twin. The lower provenance band distinguishes direct measurements from reconstructed states and from counterfactual/carbon-screen outputs so that evidence type remains visible when decisions are interpreted.

Table 3. Scope and assumptions for the carbon-aware intervention screening.

Input/Scope	Screening Value	Sensitivity/Source	Decision Interpretation
Carbon comparison window	588 h/year	Fixed common runtime for power-intensity comparison	Common runtime for comparing power intensity
Saudi grid B6 factor	0.55 kg CO ₂ e/kWh	SEC 2024 reported intensity [59]	Current-grid screening case
2030 grid B6 factor	0.40 kg CO ₂ e/kWh	SEC 2030 target [59]	Decarbonized-grid sensitivity
Assisted ventilation	30–90 W	Parametric low-power fan range	Operational/reversible Level 1 action
Localized cooling	0.75–2.25 kW	Parametric limited-zone power range	Level 3 service escalation; equipment embodied carbon excluded
Reversible timber shade	10 m ² , 15 mm equivalent timber; 10 y cycle	33.9 kg CO ₂ e/m ³ A1–A3 fossil GWP [60]	Level 2 reversible-addition screen
Ushaiger material benchmark	Published 80 m ² external-wall FU	Site-specific LCA [17]	Context for Level 4 material review; not functionally interchangeable
Excluded modules	Transport and construction (A4–A5); maintenance, repair, replacement, and refurbishment (B2–B5); end-of-life (C1–C4); equipment and refrigerants	Data unavailable at intervention-design stage	Reported values are screening burdens, not complete whole-life totals

**Figure 6.** Workflow from fragmented monitoring to a carbon-aware heritage decision twin. Measured sensor/CAD inputs pass through quality control, hybrid reconstruction and blind validation, physical and explainable thermal analysis, Pareto control and carbon screening before being combined as evidence, opportunity, carbon, and heritage information in the decision output.

4. Results

4.1. Annual Reconstruction and Validation

The completed time series contains 17,109 nominal half-hour steps: 10,755 measured (62.9%) and 6354 reconstructed (37.1%). Figure 7 distinguishes measured and reconstructed

indoor and outdoor temperatures with separate line colors, including minor gaps, as well as the two principal interruptions. The revised reconstruction preserves every available measured value.

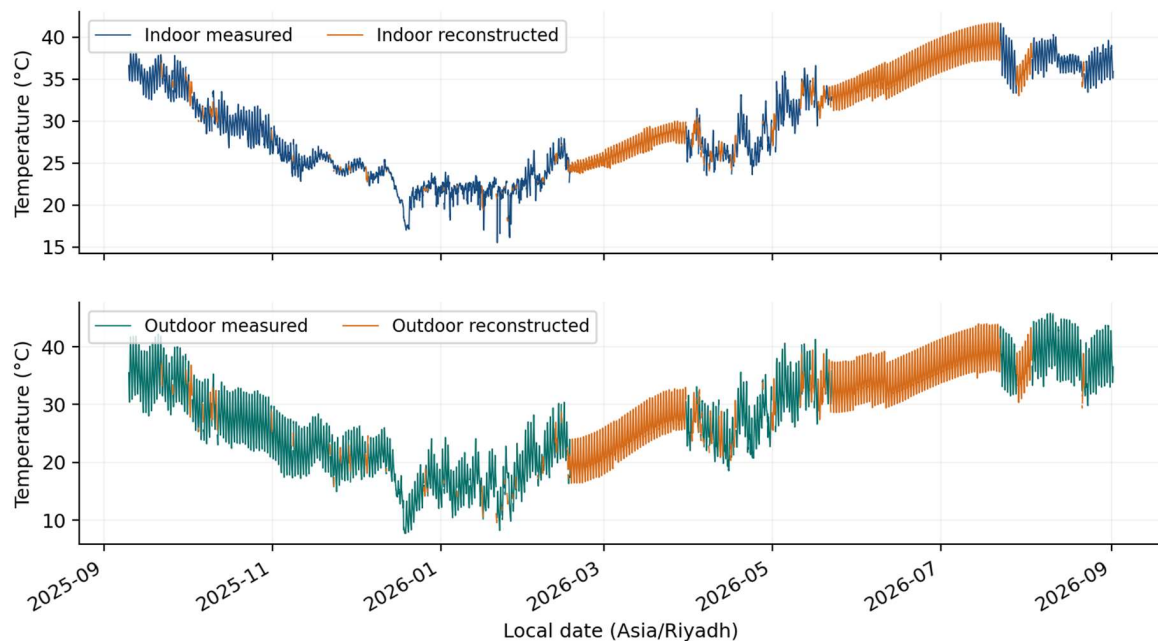


Figure 7. Completed temperature record. Blue/green traces denote measured indoor/outdoor temperatures, and orange traces denote reconstructed values. Line segments are broken when their evidence category changes.

Across the six seven-day tests, indoor-temperature reconstruction has a mean MAE of 0.93 °C, median MAE of 0.87 °C, and mean RMSE of 1.11 °C. Mean indoor MAE increases to 1.14, 1.31, and 1.35 °C at 28, 42, and 61 days, respectively. At 61 days, outdoor-temperature mean MAE reaches 3.08 °C and outdoor-RH mean MAE 14.99 percentage points. Table 4 reports the short-block results; Table A1 reports all gap lengths. The longer-gap uncertainty is, therefore, larger, particularly for external forcing and moisture. The absence of long measured summer windows prevents direct validation of the May–July interruption.

Table 4. Blind seven-day gap-reconstruction performance across six seasonal blocks.

Variable	Mean MAE	Median MAE	Mean RMSE	Scope
Indoor temperature (°C)	0.93	0.87	1.11	Short-block empirical error; longer tests in Table A1
Outdoor temperature (°C)	1.26	1.31	1.50	Short-block empirical error; longer tests in Table A1
Indoor RH (percentage points)	3.27	3.36	4.23	Short-block empirical error; longer tests in Table A1
Outdoor RH (percentage points)	5.10	5.42	6.29	Short-block empirical error; longer tests in Table A1

Table 5 reports monthly means and measured coverage for the revised reconstruction. June is entirely reconstructed and is estimated at 35.7 °C indoors and 34.6 °C outdoors; July is estimated at 38.2 and 38.0 °C, with 19.2% measured coverage. These estimates describe model-supported monthly states and do not replace unobserved measurements. Temperature and RH uncertainty are propagated separately to decision outcomes.

Table 5. Completed monthly environmental statistics. Measured (%) denotes coverage within the available monitoring window; the September periods are partial months.

Month	Measured (%)	Indoor T (°C)	Outdoor T (°C)	Difference (°C)	Indoor RH (%)	Outdoor RH (%)
2025-09	97.6	35.3	34.9	+0.4	17.7	22.3
2025-10	90.5	30.1	28.2	+1.9	20.1	26.4
2025-11	95.8	25.3	22.0	+3.3	22.9	30.7
2025-12	95.0	22.4	17.6	+4.8	38.2	58.1
2026-01	94.6	21.7	16.2	+5.5	28.5	41.8
2026-02	51.5	24.6	21.4	+3.2	27.4	37.5
2026-03	4.0	27.4	26.0	+1.4	27.0	34.9
2026-04	77.5	27.1	26.3	+0.8	38.0	52.0
2026-05	46.9	32.6	32.9	−0.3	21.9	27.3
2026-06	0.0	35.7	34.6	+1.1	20.3	25.1
2026-07	19.2	38.2	38.0	+0.2	17.6	22.7
2026-08	91.0	37.1	38.4	−1.3	17.7	22.4
2026-09	100.0	35.9	35.3	+0.6	19.9	25.5

4.2. Measured Thermal Buffering

The original observations remain the evidential basis for physical interpretation. Across all 10,757 unique valid paired observations, mean indoor and outdoor temperatures were 28.38 and 26.32 °C. Indoor temperature ranged from 15.5 to 41.6 °C, compared with 7.7–45.8 °C outdoors. Figure 8 shows only measured temperature data and, therefore, preserves visible interruptions.

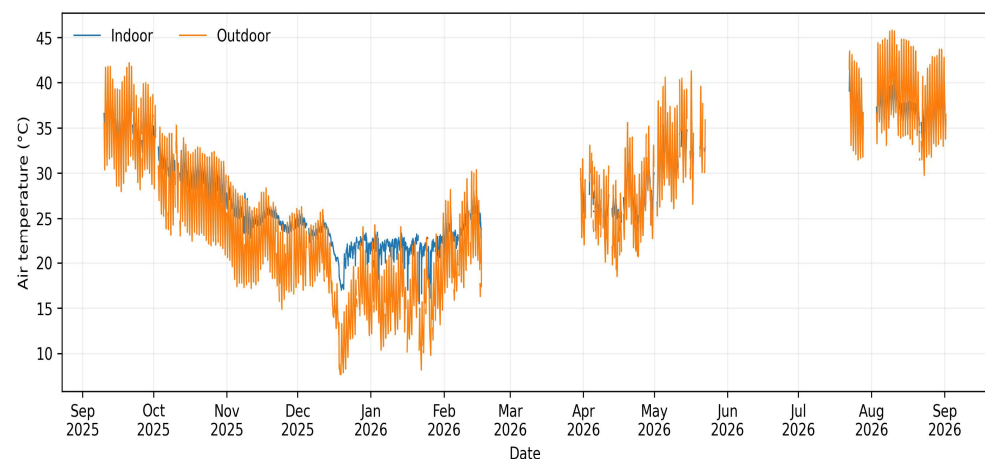


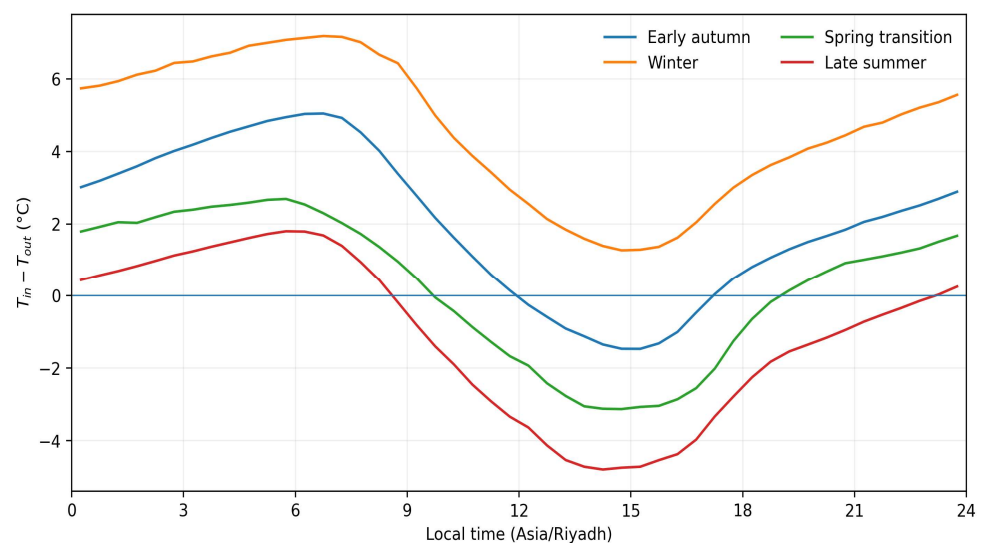
Figure 8. Measured indoor and outdoor temperatures. Lines are broken across missing periods; no reconstructed values are used in this figure.

Phase-based analysis of the measured record shows a seasonal reversal in envelope function. The winter mean indoor temperature was 4.55 °C higher than the outdoor temperature, and the indoor diurnal amplitude was only 0.99 °C compared with 6.70 °C outdoors, corresponding to a damping ratio of 0.15. In late summer, the mean indoor offset changed to −1.07 °C, while the indoor amplitude remained only 3.67 °C compared with 9.75 °C outdoors. Peak delays from the phase-mean profiles were approximately 1–2 h. The phase-specific damping and lag metrics are summarized in Table 6.

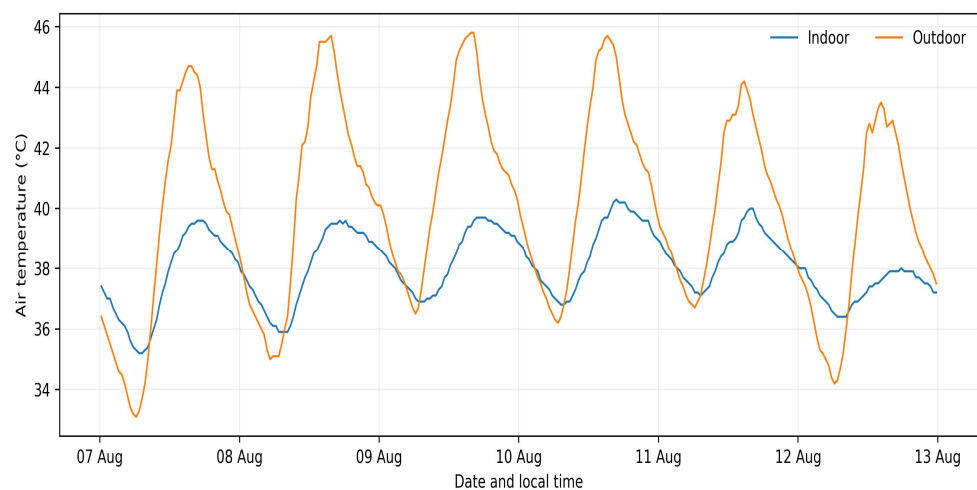
Table 6. Measured thermal-buffering metrics by monitoring phase.

Phase	<i>n</i>	Mean ΔT (°C)	A_{in} (°C)	A_{out} (°C)	f_d	Peak Lag (h)
Early autumn	3709	+2.04	2.73	9.16	0.30	1.0
Winter	3513	+4.55	0.99	6.70	0.15	2.0
Spring transition	1875	+0.18	3.17	8.54	0.37	1.5
Late summer	1660	−1.07	3.67	9.75	0.38	2.0

Figure 9 shows the corresponding hourly reversal between daytime attenuation and nighttime heat release.

**Figure 9.** Measured mean diurnal indoor–outdoor temperature difference. Warm-phase afternoons show strong attenuation, while positive nighttime values show delayed heat release.

During late summer, 306 measured observations occurred with $T_{out} \geq 42$ °C. The mean outdoor temperature in those periods was 43.38 °C, and the mean indoor temperature was 38.65 °C, producing a 4.74 °C mean attenuation. Conversely, when the late-summer outdoor temperature was at or below 34 °C, the interior remained 2.12 °C warmer on average. The extreme sequence in Figure 10 demonstrates that the envelope protects against daytime peaks but stores enough heat to limit nighttime recovery.

**Figure 10.** Measured extreme-summer sequence around the highest-recorded outdoor temperature.

The measured absolute humidity averaged 6.99 g m^{-3} indoors and 8.41 g m^{-3} outdoors. This confirms that the lower indoor RH is accompanied by a lower vapor content and is not solely a temperature artefact. The absolute-humidity correlation weakens during extreme summer, indicating stronger indoor–outdoor decoupling, although the available instrumentation cannot isolate moisture sorption, infiltration, ventilation, and local landscape effects. Figure 11 shows the measured/derived absolute-humidity sequence used for this interpretation.

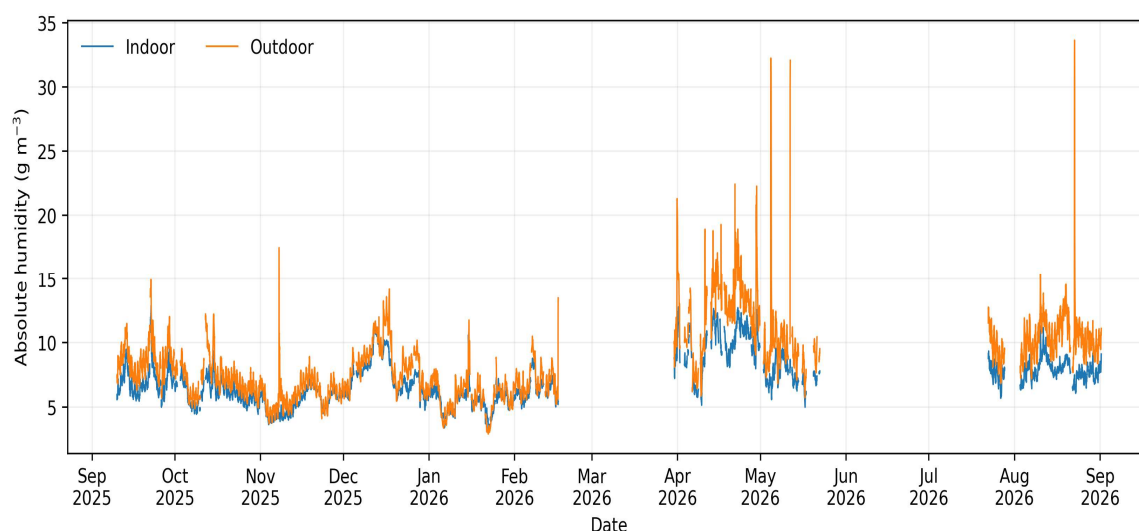


Figure 11. Measured/derived indoor and outdoor absolute humidity during available monitoring periods.

4.3. Explainable AI Results

Table 7 and Figure A1 report the revised identical-fold comparison. Random forest produces the lowest mean and median MAE (0.64 and 0.47 °C), compared with 0.82 and 0.72 °C for Ridge and 0.81 and 0.57 °C for XGBoost. XGBoost is retained as the explanation model and not claimed to be the most accurate model. The paired climate-expanded XGBoost comparison is discussed in Section 5.7.

Table 7. Indoor-temperature surrogate performance across five seasonally blocked validation weeks.

Model	Mean MAE (°C)	Median MAE (°C)	Mean RMSE (°C)	Role
Ridge	0.82	0.72	0.97	Linear reference
Random forest	0.64	0.47	0.84	Lowest errors in revised tests
XGBoost	0.81	0.57	0.96	SHAP explanation model

SHAP for all 8314 eligible measured targets assigns mean absolute contributions of 1.62 °C to the 3 h outdoor-temperature mean, 1.32 °C to the 24 h mean, 0.97 °C to the 6 h mean, and 0.96 °C to the instantaneous outdoor temperature (Figure 12). The 12 h mean contributes 0.09 °C in this implementation. The 3 h feature ranks first in all 90 season-stratified repeated samples and in the measured high-temperature subset; rank correlations with the equal-season reference are at least 0.991 (Table A3). The association with accumulated outdoor forcing is reproducible within this model; a uniquely dominant 12 h response is not supported. Correlated rolling features share predictive information, so SHAP does not identify a unique physical time constant.

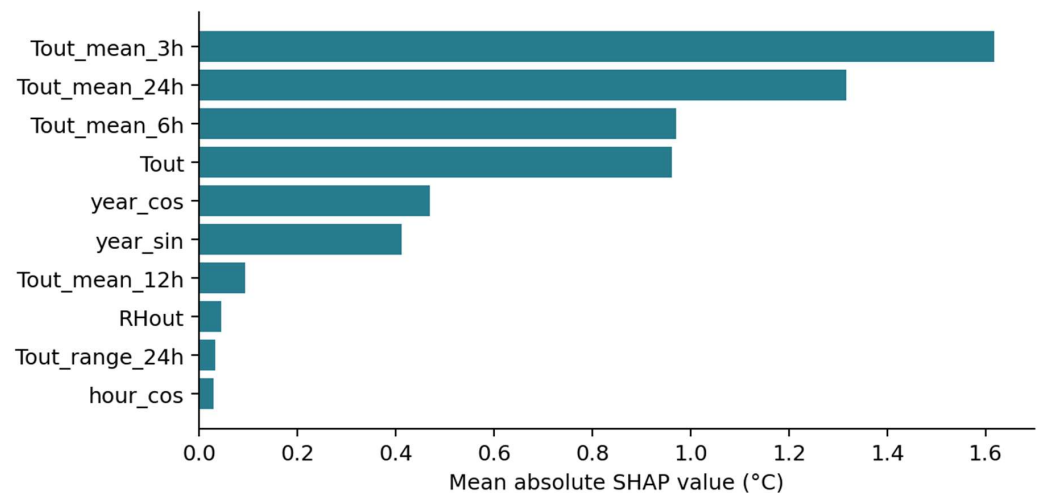


Figure 12. Mean absolute SHAP contributions for all eligible measured-target observations in the revised XGBoost surrogate. Correlated rolling means are predictive associations, not independent causal effects.

4.4. Thermal Regime Classification

The four-cluster operational partition contains 99 winter heat-retention days, 34 mild-transition days, 79 hot-transition days, and 145 extreme-heat days (Table 8 and Figure 13). Winter centroids are 23.4 °C indoors and 18.8 °C outdoors, with a +4.59 °C offset. Extreme-heat centroids are 35.7 °C indoors and outdoors, with a near-zero mean offset, while indoor and outdoor daily amplitudes are 3.9 and 9.1 °C. The silhouette and Davies–Bouldin criteria favor $k = 2$ over $k = 4$ (0.398 versus 0.290 and 0.999 versus 1.191). For $k = 4$, seed ARI has a median of 1.000 and a minimum of 0.968; bootstrap ARI has a median of 0.840 and a minimum of 0.649. These diagnostics support reproducibility of the fitted operational partition but do not establish four objectively optimal or transferable regimes (Table A2 and Figure A4).

Table 8. Thermal regimes identified from standardized daily variables.

Regime	Days	Outdoor Mean (°C)	Indoor Mean (°C)	Difference (°C)	Outdoor Amplitude (°C)	Indoor Amplitude (°C)
Winter heat retention	99	18.8	23.4	+4.59	7.7	2.0
Mild transition	34	24.4	26.1	+1.73	7.2	2.4
Hot transition	79	26.7	28.3	+1.55	8.6	3.0
Extreme-heat buffering	145	35.7	35.7	−0.01	9.1	3.9

4.5. Ventilation Optimization

The revised hot-season baseline contains 4838.1 degree-hours above 35 °C and 1886.5 exposure hours (Figure 14). At $\eta = 0.061$ and $\rho = 0.865$, the reference knee starts at 17:30, has a 10 h availability window, and uses a 1.50 °C trigger. It reduces degree-hours by 2.62% and exposure hours by 2.84%, with 361.5 active ventilation hours. Figure 15 shows 90 unique non-dominated schedules in the seed-2026 final set. All five seeds produce identical knee objective values; one run selects an equivalent 18:00/9.5 h window. Exhaustive evaluation of 10,166 schedules results in the same knee degree-hours and active hours, although several schedules share those objectives. Absolute final-50-generation hypervolume changes are below 0.006% (Table A7 and Figure A5).

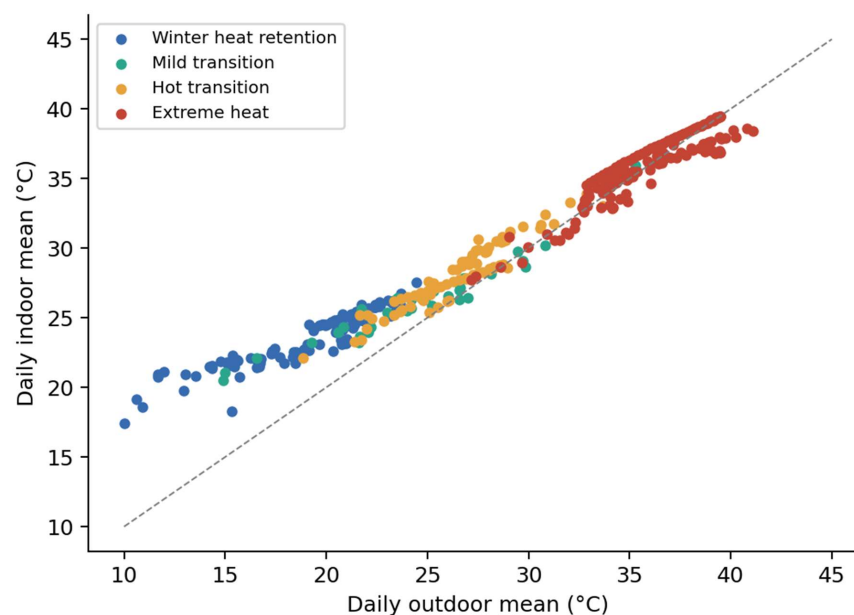


Figure 13. Daily indoor versus outdoor temperature colored by AI-derived thermal regime. The dashed line represents equal indoor and outdoor mean temperatures.

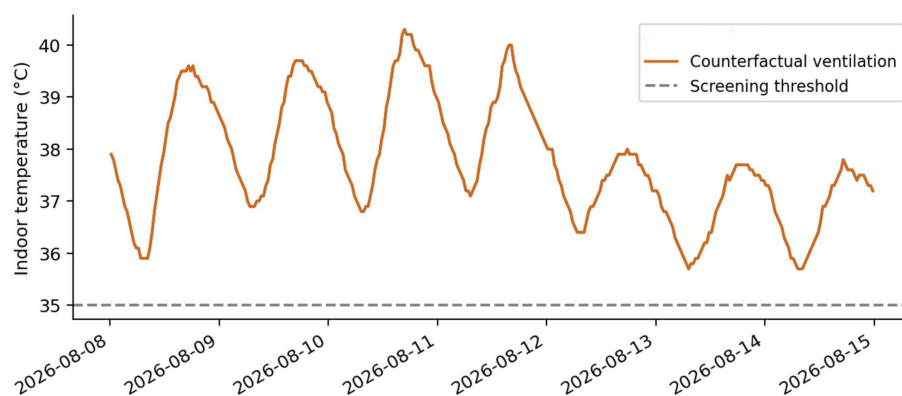


Figure 14. Available August hot-week baseline and the revised counterfactual controller. The counterfactual is an assumed ventilation scenario, not an observed treatment.

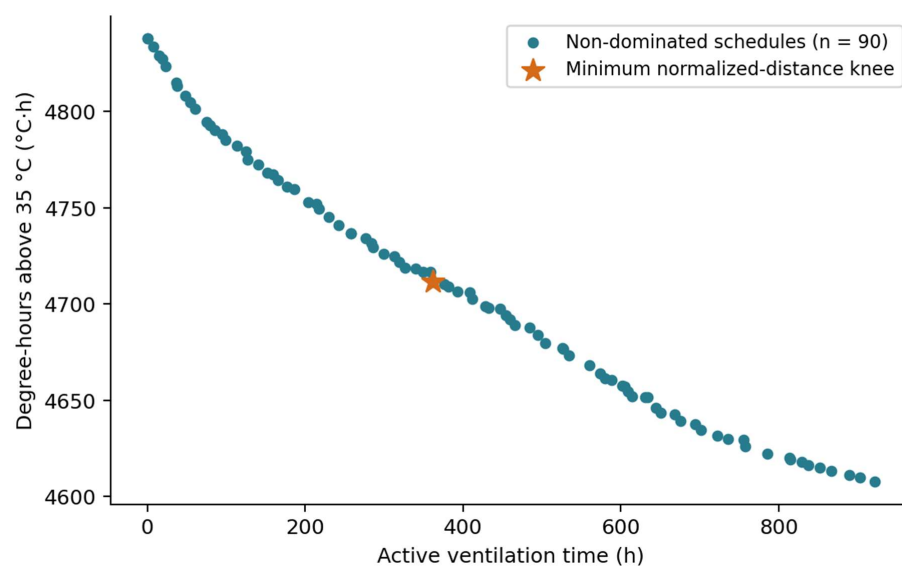


Figure 15. Revised Pareto front at fixed $\eta = 0.061$ and $\rho = 0.865$. Blue points show the 90 unique non-dominated schedules in the seed-2026 final set; the orange star identifies the minimum normalized-distance knee. The burden axis is the active time without a coupling multiplier.

The joint η – ρ sensitivity with the reference schedule fixed produces degree-hour reductions from 0.62% to 6.91% (Table 9). At the reference persistence, reoptimizing at $\eta = 0.03$ and 0.09 results in 1.51% and 3.46% reductions, respectively. The exact percentage is parameter dependent, while the tested scenarios consistently indicate modest passive relief rather than elimination of seasonal heat exposure. These are conditional estimates of operational potential, not measured intervention efficacy or a guaranteed physical upper bound. Field or calibrated airflow/thermal-model validation is required before implementing a schedule.

Table 9. Reduction in degree-hours above 35 °C (%) across the joint coupling–persistence grid, with the reference schedule and trigger held fixed.

η per Step	$\rho = 0.750$	$\rho = 0.800$	$\rho = 0.865$	$\rho = 0.900$	$\rho = 0.950$
0.020	0.62	0.75	1.06	1.42	2.90
0.030	0.90	1.08	1.51	2.01	3.99
0.050	1.39	1.65	2.27	2.99	5.57
0.061	1.64	1.93	2.62	3.40	6.07
0.070	1.82	2.13	2.84	3.67	6.35
0.090	2.16	2.48	3.21	4.02	6.80
0.100	2.30	2.61	3.36	4.15	6.91

4.6. Heritage Decision-Support Outcomes

Figure 16 places the 124 hot days against daily heat exposure and the fixed night-cooling opportunity indicator. The revised local medians are $H^* = 37.741$ degree-hours/day and $C^* = 6.5$ h/day. They are screening splits for this building and temperature threshold, not national compliance limits. The k-means thermal regimes do not determine the four decision quadrants.

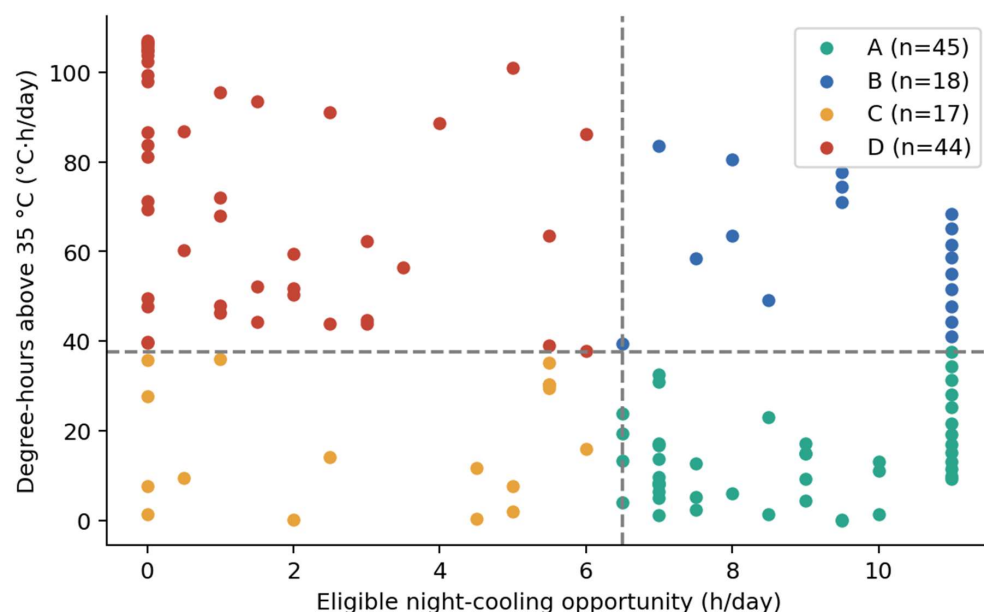


Figure 16. Ushaiger hot-day decision matrix. The four classes separate days that can initially be addressed through operation from those needing reversible solar/use measures or escalation review. Median splits are case-specific analytical thresholds, not regulatory limits.

Table 10 reports 233 routine-monitoring days and 124 hot days. The hot-day counts are 45 in Class A (36.3%), 18 in Class B (14.5%), 17 in Class C (13.7%), and 44 in Class D (35.5%).

Classes A–C, therefore, account for 80 hot days (64.5%) that initially favor operational or reversible responses. Table 10 includes low-, moderate-, and high-confidence counts and percentages within each class; these are coverage categories rather than probability estimates of model accuracy.

Table 10. Ushaiger decision classes and evidence confidence derived from the completed annual digital-twin sequence.

Decision Class	Days	Hot-Day Share (%)	Low Confidence	Moderate Confidence	High Confidence	Initial Action
0—Routine monitoring	233	–	54 (23.2%)	26 (11.2%)	153 (65.7%)	Maintain monitoring
A—Operational response	45	36.3	27 (60.0%)	1 (2.2%)	17 (37.8%)	Test opening control
B—Combined passive response	18	14.5	14 (77.8%)	0 (0.0%)	4 (22.2%)	Ventilation and reversible solar control
C—Shading or use scheduling	17	13.7	2 (11.8%)	3 (17.6%)	12 (70.6%)	Reversible shading or use scheduling
D—Escalation review	44	35.5	17 (38.6%)	1 (2.3%)	26 (59.1%)	Detailed assessment before localized service or fabric change

Confidence percentages use the number of days within each row as the denominator. High coverage is >75%, moderate coverage 25–75%, and low coverage < 25% of 48 nominal daily half-hour positions. Partial-day coverage is not treated as a full day of evidence.

Within Class D, 26 days (59.1%) have high measured coverage, 1 (2.3%) moderate coverage, and 17 (38.6%) low coverage. The severe, low-opportunity state is, therefore, present in predominantly measured days, although reconstructed days materially affect annual totals. Low-confidence Class D states justify additional sensing and reversible trials, not irreversible fabric intervention.

Figure 17 shows the seasonal distribution: June contains 22 Class A and 8 Class B days, all dependent on reconstruction; July contains 17 Class D days and August 26. Sensitivity to the temperature threshold changes the number of hot days from 183 at 30 °C to 25 at 40 °C. With threshold-specific medians, Class A spans 32.6–44.0% of hot days and Class D 31.5–40.0%; reoptimized ventilation reduction spans 0.95–2.62%. Counts using fixed 35 °C splits are reported separately in Table A5. The 35 °C counts are, therefore, a scenario snapshot, not invariant intervention requirements.

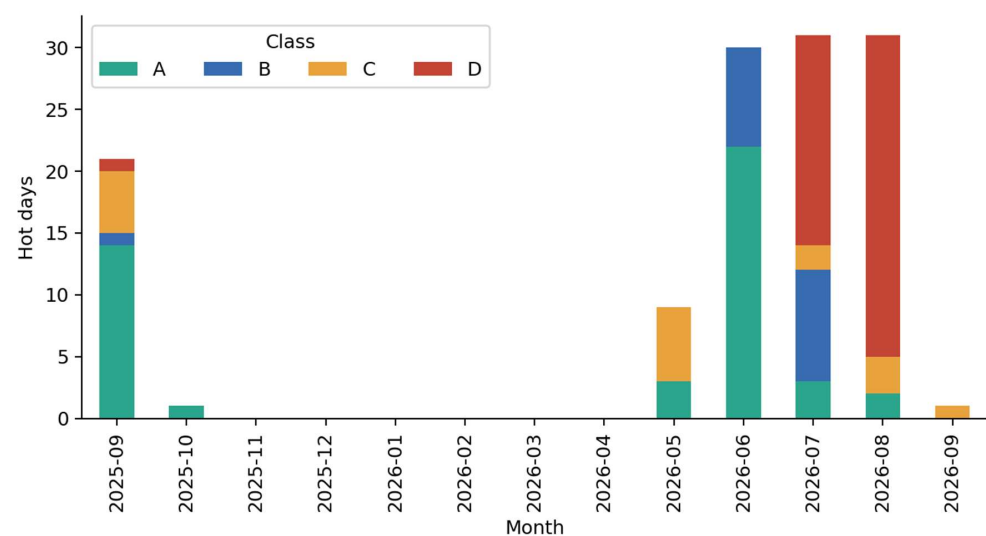


Figure 17. Monthly decision-class counts for the revised reconstruction at 35 °C. June has no measured target observations; uncertainty envelopes for June and July are reported in Table A6.

For Saudi heritage management, the present results demonstrate building-scale decision support and suggest a possible future portfolio-scale application. At building scale, the framework provides an operating rule for openings, visitor timing, and reversible heat-control trials. Subject to multi-building validation, the same indicators could in the future support triage across monitored heritage buildings, with sites repeatedly exhibiting high-confidence Class D states prioritized for detailed hygrothermal modeling, structural/condition surveys, conservation design, and funding. This turns continuous sensing into a resource-allocation tool rather than a dashboard alone.

With fixed decision splits, -0.97 and $+0.97$ °C shifts in reconstructed indoor temperatures change hot-day totals to 117 and 136, Class A totals to 34 and 53, and Class D totals to 51 and 33. The opposite A/D movements reflect changes in both severity and the indoor–outdoor cooling advantage. Across 200 joint residual-block scenarios, the 5th–95th percentile envelopes are approximately 115–129 hot days, 23–41 Class A days, and 37–57 Class D days; the fixed controller’s degree-hour benefit spans 1.00–4.14%. June and July outcomes vary substantially (Table A6). These envelopes are stress scenarios; the long-gap summer uncertainty remains uncalibrated.

4.7. Life-Cycle Carbon Screening

The carbon screen changes the meaning of escalation because the same 588 h thermal-management window produces very different A6 burdens depending on the power intensity. Figure 18 compares assisted ventilation and localized cooling at the reported 2024 SEC grid intensity; Table 11 also reports the 2030 grid-intensity target. Under the 0.55 kg CO₂e/kWh case, 30–90 W assisted ventilation produces only 9.7–29.1 kg CO₂e per hot season, whereas 0.75–2.25 kW localized cooling produces 242.6–727.7 kg CO₂e. At the central sensitivity points, the 1.50 kW cooling case produces 485.1 kg CO₂e compared with 19.4 kg CO₂e for a 60 W fan: a factor of 25 for the same operating hours. The 2030 target reduces both burdens proportionally but does not reverse their ranking.

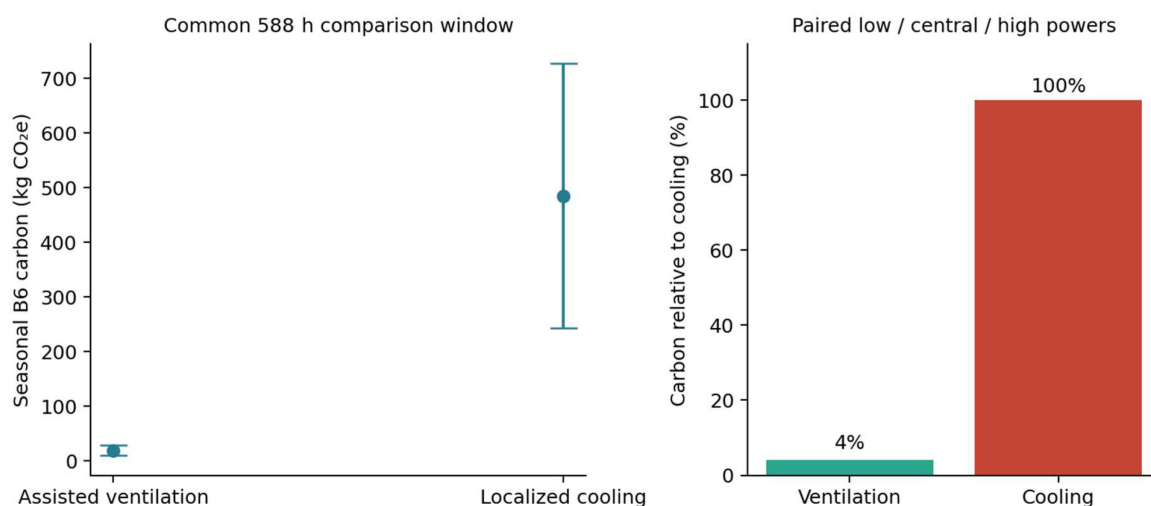


Figure 18. Annual B6 carbon at 0.55 kg CO₂e/kWh within a fixed 588 h comparison window. Error bars cover the stated low/high-power assumptions. Paired low-, central-, and high-ventilation/cooling powers each give ventilation 4% of cooling carbon, or 96% less. Thermal services are not assumed equivalent; equipment-embodied impacts remain excluded.

At paired low, central, and high powers, ventilation has 4% of localized-cooling B6 carbon, corresponding to a 96% reduction or a 25-fold ratio. Pairing is essential: independently choosing endpoints produces different ratios. This comparison quantifies power intensity over equal hours and does not demonstrate equivalent delivered thermal comfort.

The 588 h carbon reference is retained as a standardized scenario rather than substituted for the revised controller's 361.5 active hours.

Table 11. Operational GWP sensitivity for the fixed 588 h reference comparison window.

Scenario	Electricity (kWh/Season)	GWP at 0.55 kg CO ₂ e/kWh	GWP at 0.40 kg CO ₂ e/kWh
Assisted ventilation, 30 W	17.64	9.7 kg	7.1 kg
Assisted ventilation, 60 W	35.28	19.4 kg	14.1 kg
Assisted ventilation, 90 W	52.92	29.1 kg	21.2 kg
Localized cooling, 0.75 kW	441	242.6 kg	176.4 kg
Localized cooling, 1.50 kW	882	485.1 kg	352.8 kg
Localized cooling, 2.25 kW	1323	727.7 kg	529.2 kg

The numerical values underlying Figure 18 are reported in Table 11 so that the screening calculation remains auditable. The table also shows that grid decarbonization lowers operational carbon by about 27% between the two SEC intensity cases, but a high-power intervention remains substantially more carbon intensive than a low-power operational measure.

Figure A2 shows the long-horizon implications. Repeating the same seasonal operating window for 20 years produces 0.19–0.58 t CO₂e for assisted ventilation at the reported grid intensity and 0.14–0.42 t CO₂e under the 2030 target. Adding two assumed reversible timber-screen installations produces positive-emissions screening sums of approximately 0.20–0.59 and 0.15–0.43 t CO₂e, respectively; the timber increment uses GWP-fossil and is not a complete GWP-total balance including biogenic carbon. In contrast, the operational component of localized cooling reaches 4.85–14.55 t CO₂e at the reported grid intensity and 3.53–10.58 t CO₂e under the target. Because the active-cooling calculation excludes equipment production, replacement, refrigerant leakage, and maintenance, these figures should be interpreted as a lower-bound screening burden rather than complete whole-life GWP.

A second decision scale concerns fabric intervention. Figure A3 places the dynamic operating scenarios beside a site-specific material context without treating the materials as functionally interchangeable. The previously published Ushaiger LCA reports A1–A3 fossil GWP values of 46.48 kg CO₂e for decorative plaster, 94.46 for limestone, 169.22 for mud, 170.00 for ready-mix concrete, and 189.03 for blended hydraulic cement for its stated 80 m² wall FU [17]. These values are summarized in Table 12. Their decision value is contextual: a lower numerical GWP cannot justify replacing authentic mud with a material of different physical, hygrothermal, visual, or cultural function.

Table 12. Published Ushaiger A1–A3 fossil-GWP values used as the local fabric-intervention benchmark [17].

Material/Product in Published Ushaiger LCA	A1–A3 Fossil GWP (kg CO ₂ e per Stated 80 m ² FU)	Use in This Study
Decorative plaster	46.48	Contextual restoration-material benchmark
Limestone	94.46	Contextual restoration-material benchmark
Mud	169.22	Traditional-material benchmark; authenticity remains a separate constraint
Ready-mix concrete	170.00	Modern-material benchmark; not assumed compatible with earthen heritage
Blended hydraulic cement	189.03	Modern-material benchmark; not assumed compatible with earthen heritage

The resulting carbon-aware action hierarchy is summarized in Table 13. It does not replace the thermal decision matrix; it adds a proportionality test after a need has been identified. The practical implication is that Classes A–C should normally remain operational or reversible while the thermal service is still adequate, because their marginal carbon burden is small relative to sustained active cooling. Class D can justify localized active cooling where use, health, collections, or visitor protection requires it, but the decision twin can now expose the carbon cost of that escalation and restrict it to the smallest zone and time window that delivers the required service. Level 4 fabric change remains a conservation-design decision supported by local LCA evidence, not an automatic response to high temperature.

Table 13. Carbon-aware interpretation of the heritage action ladder.

Decision Level	Preferred Thermal Response	Carbon Evidence from Screening	Decision Rule
Level 0	Monitor/document	Negligible incremental action carbon	Do not intervene without verified need
Class A/Level 1	Manual opening or low-power assisted night ventilation	9.7–29.1 kg CO ₂ e/season at 2024 SEC intensity	Prefer operation while safe night-cooling opportunity exists
Classes B–C/Level 2	Ventilation plus reversible shading/use scheduling	Approx. 0.20–0.59 t CO ₂ e positive-emissions screen over 20 y for fan + two timber-screen cycles	Add reversible solar control before permanent fabric change
Class D/Level 3	Smallest feasible localized conditioned refuge or service	4.85–14.55 t CO ₂ e over 20 y B6-only screening	Escalate only with verified thermal need; minimize zone, power, and runtime
Level 4	Fabric/construction review	Use Ushaiger-specific LCA plus project-specific EPDs and quantities	Carbon informs, but significance, compatibility, safety, and authority approval remain hard constraints

5. Discussion

5.1. Uncertainty-Aware Digital Twin

The principal advance of this study is not the production of a continuous-looking curve. It is the explicit separation of three evidence classes: direct measurements, AI-reconstructed states, and counterfactual optimized states. This distinction responds to data-quality and validation concerns identified in building and heritage digital-twin reviews, where heterogeneous observations and modeled states require explicit provenance [36–38]. Here, measured-only analyses establish the physical behavior of the earthen envelope; reconstructed data extend seasonal continuity; and optimization is performed only after uncertainty and model assumptions are stated.

The revised long-block tests show why a weekly error cannot characterize a multi-week interruption: mean indoor MAE rises from 0.93 °C at seven days to 1.35 °C at 61 days, while outdoor-temperature and RH errors rise more sharply. Overlapping cool-season tests establish gap-length stress sensitivity but do not validate an unmeasured summer sequence. Monthly reconstructed states and decision counts are, therefore, accompanied by scenario envelopes, and no point estimate for June is given the evidential status of a measurement.

5.2. Thermal Memory and Explainable AI

Measured damping and delayed heat release remain the primary physical evidence for thermal buffering. The revised explanation model assigns its largest SHAP contributions to 3 and 24 h outdoor-temperature means, and repeated seasonal samples preserve that ranking. This supports an association with accumulated external forcing, but the allocation

among strongly correlated rolling means is model dependent. Feature rank alone, therefore, does not identify a thermal time constant or causal heat-transfer mechanism.

The revised identical-fold comparison favors random forest rather than XGBoost on both the mean and median error. XGBoost is retained to examine an explicit nonlinear surrogate with SHAP, while both the linear baseline and the better-performing ensemble remain visible. Retrospective seasonal holdouts reveal error variation across available regimes; they do not establish prospective prediction skill or performance during entirely unmeasured months.

5.3. State-Dependent Thermal Performance

The four clusters convert the seasonal story into discrete operating states. Winter is dominated by beneficial heat retention, whereas the extreme-heat regime is dominated by amplitude attenuation with little mean cooling advantage. The hot-transition and mild-transition states lie between those extremes. This state dependence is important for conservation because the best intervention is unlikely to be constant throughout the year. Measures that aggressively increase heat loss could undermine winter protection, while summer strategies need to target stored heat and solar gains without removing the beneficial thermal capacity of the earthen fabric.

Cluster diagnostics prefer a two-way separation statistically. Four states are used only to describe operationally useful subdivisions of the observed seasonal trajectory. Their high seed agreement and lower bootstrap agreement show that numerical repeatability under initialization does not remove sensitivity to the observed days. The four-regime names and centroids require new validation and recalibration in another building.

The regime framework also provides a practical structure for future physics-based calibration. Instead of fitting one model to the entire year, a future IDA Indoor Climate and Energy model or other hygrothermal model can be calibrated and independently checked within representative regime windows. Parameters related to thermal capacity, infiltration, openings, roof exposure, and shading can then be assessed against amplitude, offset, and lag targets specific to each state.

5.4. Limits of Night Ventilation

The revised ventilation calculation produces a 2.62% reduction at the reference coupling and persistence, while the joint parameter envelope spans 0.62–6.91%. This difference prevents the precise reference percentage from being treated as a robust field-effect estimate. The direction of benefit and its limited scale persist across the tested assumptions, but neither a measured air-change rate nor a formal physical cooling ceiling has been established. The new active-hours objective also separates operating effort from the assumed cooling coupling.

For conservation practice, the implication is that night ventilation should be combined with measures that reduce daytime heat storage or increase safe heat rejection. Candidate scenarios include compatible roof repair, reversible external or courtyard shading, solar-gain reduction at openings, and ventilation strategies triggered by both temperature and moisture. These options should be tested in a calibrated physics-based model before intervention because the present coupling coefficient is an effective scenario parameter, not a measured air-change rate. The thermal indicators used in this study should also be distinguished from formal occupant-comfort assessment. The 35 °C exposure metric was selected as a simple screening variable for identifying severe heat accumulation in an unconditioned and currently abandoned heritage dwelling; it does not represent a comfort boundary for residents or visitors. PMV-based assessment would require additional information on metabolic rate, clothing, local air speed, and mean radiant temperature, while an adaptive-

comfort interpretation is most meaningful when actual occupants are present and able to adapt their behavior and report their thermal perception. Because these conditions were absent during the monitoring campaign, applying either framework retrospectively would imply a level of human-response precision that the available evidence cannot support. For future adaptive reuse, particularly where Ushaiger buildings accommodate visitors, staff, or residents, the proposed decision framework should, therefore, incorporate a second human-centered layer. This should combine operative-environment measurements with occupancy and use schedules and short thermal-sensation/acceptability surveys. In that future application, the present 35 °C exposure indicator would remain an initial screening signal, while comfort-based evidence would determine whether observed thermal states actually produce unacceptable conditions for specific user groups and periods of use.

5.5. Saudi Heritage Decision Implications

The decision-support layer addresses a different question from conventional retrofit optimization. Standard optimization asks which design minimizes energy, discomfort, cost, or carbon. Heritage practice must first ask whether a physical intervention is justified at all. The Heritage Commission's emphasis on cultural values, structural safety, compatible reuse, and minimum intervention [2] makes this sequencing particularly relevant. The proposed ladder, therefore, treats the optimizer as a screening component inside a conservation process, not as an autonomous design authority.

At the 35 °C reference screen, Classes A–C contain 80 of 124 hot days (64.5%); Class D contains 44 (35.5%). Twenty-six Class D days have high measured coverage. However, threshold changes and multivariable reconstruction-error scenarios change the class totals materially. The framework, therefore, uses repeated measured severe states to prioritize assessment, with reconstructed states informing sensing priorities and reversible trials; it does not convert a single threshold-based count into a mandatory retrofit decision.

The framework also creates a transparent link between Saudi adaptive-reuse research and digital-twin analytics. Existing Saudi assessment models emphasize periodic evaluation of function, authenticity, repair, safety, and legal/economic considerations [9,16]. The proposed digital layer can support these assessments with time-resolved environmental evidence while retaining human and regulatory decisions about cultural significance. This division of roles is important: sensors can quantify heat severity and AI can estimate missing states, but neither can infer the cultural acceptability of replacing a roof element, modifying an opening, or adding a visible shading device.

Although the present framework is developed and tested on one principal monitored dwelling, an independent thermal study previously conducted on a different mud-brick house within Ushaiger Heritage Village provides useful comparative evidence [22]. In that study, short-term monitoring during August and September 2025 showed that the earthen envelope reduced daytime heat transmission and attenuated diurnal temperature fluctuations. During the August campaign, the mean indoor temperature was 38.12 °C compared with 40.22 °C outdoors. The indoor diurnal temperature variation remained approximately 3–4 °C compared with approximately 12–15 °C outdoors, while the indoor thermal response lagged by approximately 1–3 h relative to external temperature peaks. Table 14 compares the main thermal-buffering indicators for the two independently monitored Ushaiger dwellings.

This comparison provides preliminary intra-site corroboration rather than full external validation. Both dwellings are located within Ushaiger Heritage Village and, therefore, share broadly similar climatic, material, and cultural conditions. Furthermore, the previously published study was based on short summer monitoring windows and was explicitly presented as a pilot feasibility investigation rather than a fully validated digital-twin as-

essment. Accordingly, the agreement between the two cases strengthens confidence in the repeatability of the observed thermal-buffering mechanism within Ushaiger, but it does not demonstrate transferability to different heritage settlements, climatic regions, building typologies, or conservation conditions.

Table 14. Comparison of the measured thermal-buffering behavior in two independently monitored Ushaiger dwellings.

Indicator	Present Study	Previously Monitored Ushaiger Dwelling [22]
Monitoring scope	September 2025–September 2026	August–September 2025
Hot-season indoor–outdoor mean temperature difference	−1.07 °C, late summer	−2.10 °C, August
Indoor diurnal temperature amplitude	3.67 °C	approx. 3–4 °C
Outdoor diurnal temperature amplitude	9.75 °C	approx. 12–15 °C
Estimated indoor peak lag	approx. 2 h	approx. 1–3 h
Main observed behavior	Reduced amplitude and delayed response	Reduced amplitude and delayed response

The proposed transferability should, therefore, be understood at the level of the decision architecture rather than as the direct transfer of numerical thresholds between buildings. The transferable components are the sequence of evidence-quality assessment, thermal-state diagnosis, passive-cooling opportunity screening, carbon proportionality, and conservation-based escalation. By contrast, numerical parameters such as thermal-regime centroids, high-temperature exposure thresholds, night-cooling opportunity, effective ventilation coupling, and intervention performance must be recalibrated using building-specific monitoring and condition information. Applying it to another heritage building would, therefore, require an initial monitoring period, evidence-quality assessment, local thermal-state characterization, and validation of the relevant operational thresholds before using the action matrix for decision support.

Subject to further multi-building and multi-context validation, the method could be developed in the future as a portfolio triage protocol for wider Saudi heritage applications.

A practical portfolio deployment would, therefore, proceed in stages:

1. Install a standardized low-cost monitoring package;
2. Establish the proportion of measured versus reconstructed evidence;
3. Derive building-specific thermal states and passive-cooling opportunity;
4. Test reversible operational responses;
5. Escalate only those buildings that repeatedly exhibit high-confidence, high-severity, and low-opportunity conditions.

Buildings with different morphology, conservation state, use pattern, or climatic exposure would retain the same decision sequence but would not be assumed to share the Ushaiger numerical thresholds. Before portfolio-scale implementation, the framework should, therefore, be tested across multiple buildings and multiple indoor and outdoor measurement points, including different orientations, envelope conditions, levels of deterioration, use patterns, and Saudi climatic regions. Such validation should determine whether the decision classes remain stable, whether building-specific recalibration is sufficient, and whether the proposed escalation logic performs consistently when spatial thermal variability is explicitly represented.

A standardized low-cost sensor package could provide the environmental layer; the evidence gate would distinguish directly observed from reconstructed conditions; the action matrix would identify buildings repeatedly entering severe low-opportunity states; and only those cases would progress to detailed simulation, significance assessment, condition survey, and SBC 902/authority review. Such a sequence aligns digitalization with conservation capacity building and with the broader Saudi emphasis on documenting, rehabilitating, and keeping heritage places accessible as living assets [1].

5.6. Carbon-Aware Intervention Escalation

The screening LCA adds a decision dimension that is absent from a thermal-only optimizer. Saudi building LCA literature has already shown why this matters: in electricity-intensive buildings, the operational phase can dominate life-cycle impacts [52]. The present heritage case reaches the same conclusion in a more targeted form. A 1.50 kW localized-cooling scenario operated for the same 588 h window has 25 times the B6 carbon of a 60 W assisted-ventilation scenario. Over repeated hot seasons, this difference accumulates to tons of CO₂e even before the active system's equipment and refrigerant burdens are counted.

Grid decarbonization improves all electricity-using options but does not remove the proportionality question. SEC reports 0.55 t CO₂e/MWh for 2024 and targets 0.40 t CO₂e/MWh by 2030 [59]. Applying the lower factor reduces the 20-year cooling screen from 4.85–14.55 to 3.53–10.58 t CO₂e, yet low-power ventilation remains roughly an order of magnitude lower. This result strengthens the minimum-intervention logic with an environmental argument: when operational or reversible measures provide adequate service, they conserve both the fabric and carbon budget.

The Ushaiger material benchmark adds a second layer of local relevance. The site-specific LCA demonstrates that restoration materials carry distinct A1–A3 and maintenance burdens [17]. Nevertheless, those values cannot be interpreted as a simple ranking of acceptable heritage substitutes. Heritage-LCA research explicitly identifies cultural value, traditional materials, and restoration-specific processes as gaps in conventional LCA frameworks [21]. The decision twin, therefore, keeps heritage significance as a hard external constraint and uses carbon only to compare technically and culturally admissible actions. This separation is central to the proposed novelty: the optimum is not the action with the smallest GWP, but the least intrusive admissible action whose thermal service is adequate and whose marginal carbon is proportionate to the verified need.

5.7. Wadi-Edge and Moisture Effects

The agricultural and wadi-edge setting remains a plausible boundary-condition influence, but the new AI analysis does not convert that context into causality. Local wind, radiation, irrigation, soil moisture, and comparative core-settlement weather were not measured.

Supplementary regional climate information comes from the supplied Open-Meteo CSV [55]. It contains 8568 uninterrupted hourly records from 10 September 2025 00:00 to 1 September 2026 23:00 in Asia/Riyadh (UTC+3), with shortwave radiation, 10 m wind speed, and wind direction. Its reported grid location is 25.25° N, 45.25° E at 704 m, approximately 11.4 km from the Netatmo location (25.33949° N, 45.19436° E). The CSV does not identify its underlying weather model or archive query. Accordingly, it is described as Open-Meteo regional climate data; an exclusive ERA5 provenance or resolution is not asserted. Regional variables cannot reproduce façade shading, courtyard airflow, or opening-level conditions.

Regional variables were interpolated in local time onto the half-hour timestamps and compared only with measured paired observations ($n = 10,755$). Shortwave radiation correlates with outdoor temperature at $r = 0.305$, indoor temperature at $r = 0.150$, and indoor–outdoor temperature difference at $r = -0.509$. Wind speed has $r = -0.125$ with that

difference. These descriptive associations are compatible with daytime buffering and weak regional wind correspondence but do not identify landscape or airflow causality. Table 15 presents the source characteristics and measured-only correlations.

Table 15. Supplementary Open-Meteo regional climate characteristics and associations with measured thermal response.

Regional Variable	Full-Record Characteristic	Measured Thermal Association	Scope
Shortwave radiation	Mean = 259.4 W/m ² across all hours; maximum = 1042 W/m ²	$r = -0.509$ with $T_{in} - T_{out}$	Regional indicator, not façade radiation
10 m wind speed	Mean = 3.79 m/s; maximum = 12.99 m/s	$r = -0.125$ with $T_{in} - T_{out}$	Does not resolve opening airflow
10 m wind direction	0–360°; modeled with sine and cosine	No directional causal effect inferred	Underlying archive not identified in CSV

Adding regional radiation, wind speed, and sine/cosine wind direction to XGBoost uses exactly the same eligible target rows, blocked dates, and training exclusions as the base model. Mean MAE changes from 0.806 to 0.734 °C, median MAE from 0.570 to 0.558 °C, and mean RMSE from 0.961 to 0.896 °C (Table A8). The improvement is modest and does not remove the need for direct site measurements.

Outdoor RH is also the least accurately reconstructed variable. Any claim that the wadi or vegetation directly causes the indoor moisture pattern would, therefore, exceed the evidence. A next campaign should add synchronized edge/core stations, wind speed and direction, shortwave radiation, and irrigation logs so that microclimatic drivers can enter the twin as measured predictors rather than contextual hypotheses.

5.8. Study Limitations

The study has nine principal limitations. First, the sensors are consumer-grade and were not reference-calibrated on site. Second, 37.1% of the regular annual grid is reconstructed; reconstructed values are useful for seasonal continuity but are not equivalent to observations. Third, the hybrid reconstruction is retrospective because the seasonal PCHIP backbone uses observations on both sides of each long interruption, and it cannot recreate unobserved short-lived weather extremes. The added 7–61-day stress tests quantify sensitivity to gap length, but long measured summer windows are unavailable and the tested cool-season windows overlap. Their errors do not fully calibrate June/July uncertainty. The method should, consequently, not be interpreted as an online forecasting algorithm or as a substitute for observations within the long gaps. Fourth, the four-regime k -means solution is exploratory and physically interpretable but is supported by seed/bootstrap diagnostics but remains statistically less separated than $k = 2$ and has not been shown to be transferable. Fifth, the indoor-temperature surrogate represents one monitored location and does not capture spatial gradients across the dwelling. Sixth, the night-ventilation optimization is counterfactual and uses an effective coupling parameter; it must be validated experimentally or in a calibrated airflow/thermal model before implementation.

Seventh, the study does not contain direct site measurements of local wind, solar radiation, opening state, or occupancy, limiting causal attribution and control-model fidelity. Supplementary Open-Meteo regional climate data were, therefore, used to examine the sensitivity of the thermal interpretation to regional solar-radiation and wind conditions. The archive model is not identified in the supplied CSV; these regional variables provide climatic context but cannot reproduce local façade exposure, shading, courtyard airflow, vegetation

effects, or air movement through individual openings. Direct solar-radiation sensing, local wind measurement, opening-state logging, and occupancy monitoring, therefore, remain priorities for future field campaigns. The absence of occupants also means that no thermal-sensation, acceptability, or tourist-experience feedback could be collected. Consequently, the study does not claim that the 35 °C screening threshold represents occupant comfort or discomfort. Future monitoring of occupied or visitor-accessible heritage buildings should combine environmental measurements with adaptive-comfort evaluation and structured user feedback so that thermal-risk screening can be linked explicitly to human experience.

Eighth, the LCA is an intervention-screening exercise rather than a complete comparative assessment compliant with International Organization for Standardization requirements. It uses scenario equipment powers, a common 588 h runtime, an external timber GWP-fossil proxy, SEC grid-intensity cases, and a 20-year intervention horizon; it omits most transport, installation, maintenance, end-of-life, HVAC-embodied, and refrigerant impacts, and the timber proxy is not a full GWP-total result with biogenic-carbon accounting. The active-cooling carbon values are, therefore, conservative lower bounds, while the timber-shading result should be replaced with project-specific Saudi sourcing and EPD data at design stage.

Ninth, external generalizability remains limited because the complete decision-support framework is developed and tested on one principal dwelling. A previously published monitoring study of a second mud-brick dwelling in Ushaiger [22] provides preliminary cross-building corroboration, showing comparable attenuation of outdoor temperature fluctuations and delayed indoor thermal response. However, because both buildings are located within the same settlement and climatic context, this comparison represents intra-site corroboration rather than independent validation across different contexts. The transferability of the framework, its thermal-regime classifications, and its intervention thresholds should, therefore, be tested across buildings with different geometries, orientations, conservation states, envelope conditions, occupancy patterns, and climatic regions before broader generalization is attempted.

These limitations define the next digital-twin stage: reference-sensor co-location, multi-room sensing, weather and radiation measurements, opening-state logging, regime-based physics calibration, project-specific LCA inventory development, a field test of the Pareto-selected ventilation rule, and systematic multi-building validation across Ushaiger and other Saudi earthen heritage contexts. The current AI/LCA layer is, therefore, a decision-support and hypothesis-generation component, not a substitute for physical validation or conservation design.

6. Conclusions

This study advances the Ushaiger project from long-term monitoring to an evidence-gated digital twin designed for heritage action rather than visualization alone.

First, the annual grid preserves 10,755 measured and 6354 reconstructed timestamps. The blind gap-length tests extend to 61 days; the mean indoor-temperature MAE rises from 0.93 °C at seven days to 1.35 °C at 61 days. Reconstructed states retain provenance, and long-gap summer uncertainty remains uncalibrated despite the added stress tests.

Second, the measured observations confirm state-dependent thermal buffering. The revised XGBoost interpretation associates indoor temperature with accumulated outdoor forcing, with the 3 and 24 h means receiving the largest SHAP contributions. Repeated seasonal sampling supports the revised ranking, while correlated-feature attribution does not establish a unique thermal time constant.

Third, the reference counterfactual controller reduces degree-hours above 35 °C by 2.62% and exposure hours by 2.84%, with 361.5 active hours. Five NSGA-II seeds and an

exhaustive schedule-grid check agree on the knee objective values. The joint η – ρ test spans a 0.62–6.91% degree-hour reduction; these are conditional operational-potential estimates requiring physical calibration before implementation.

Fourth, at the 35 °C reference screen, 233 days remain in routine monitoring and 124 are classified as hot: 45 Class A days, 35 combined Class B/C days, and 44 Class D days. Twenty-six Class D days have high measured coverage. Threshold variation and reconstruction-error propagation change the counts, especially in June and July; the classes, therefore, guide staged assessment rather than invariant intervention prescriptions.

Fifth, the decision matrix makes minimum intervention computationally explicit without pretending to quantify heritage value. The digital twin determines evidence confidence, thermal severity, and passive-cooling opportunity; the conservation process retains authority over cultural significance, structural safety, compatible use, and regulatory approval. Reconstructed data can trigger monitoring and reversible trials, while fabric-altering interventions require measured confirmation, field assessment, and applicable heritage/code review.

Sixth, the screening LCA makes carbon proportionality explicit. Under the SEC 2024 grid-intensity case, 30–90 W assisted ventilation contributes 9.7–29.1 kg CO₂e per 588 h hot-season window, while 0.75–2.25 kW localized cooling contributes 242.6–727.7 kg CO₂e. Repeated over 20 years, the positive-emissions screening sum for ventilation plus two reversible timber-shading cycles is approximately 0.20–0.59 t CO₂e, compared with 4.85–14.55 t CO₂e for localized cooling before equipment-embodied impacts; the timber component uses GWP-fossil and is not a full biogenic-carbon balance. The SEC 2030 grid target reduces these values but does not reverse the intervention hierarchy.

Comparison with a previously monitored independent dwelling within Ushaiger provides preliminary cross-building support for the repeatability of the observed thermal-buffering behavior, particularly reduced indoor diurnal amplitude and delayed thermal response. Nevertheless, this evidence remains intra-site and does not fully validate the proposed decision-support framework externally. Future research should, therefore, extend the approach to multiple heritage buildings and different Saudi climatic and morphological contexts in order to evaluate the transferability of model parameters, thermal classifications, and intervention thresholds. Accordingly, the present study proposes a transferable decision architecture, while multi-building validation is needed to establish the transferability of its numerical parameters and intervention thresholds.

The resulting novelty is an integrated decision process combining monitoring, uncertainty-aware reconstruction, explainable AI, thermal-regime identification, optimization, screening LCA/carbon proportionality, and conservation action. For Saudi Arabia, the study provides a proof-of-concept pathway that, after multi-building and multi-point validation, could support future prioritization of operational measures, reversible interventions, localized services, detailed engineering studies, and funding across a wider heritage portfolio. Future work should validate the decision classes across multiple Najdi buildings; add direct site measurements of solar radiation and wind, along with opening-state and occupancy monitoring; connect the data-driven layer to calibrated hygrothermal simulation; replace screening LCA inputs with project-specific Saudi EPDs and measured equipment performance; and field-test the selected ventilation strategy. Only after this multi-building and multi-context validation should portfolio-scale deployment be considered.

Author Contributions: Conceptualization, S.M. and H.H.; methodology, S.M. and H.H.; software, H.H.; validation, H.H.; formal analysis, H.H.; investigation, S.M. and H.H.; resources, S.M. and H.H.; data curation, H.H.; writing—original draft preparation, S.M. and H.H.; writing—review and editing, S.M. and H.H.; visualization, S.M. and H.H.; funding acquisition, S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Prince Sultan University, RIC (Research and Initiative Center), grant number: SEED Project n. 208 2024–2026. Prince Sultan University, RIC (Research and Initiative Center), Riyadh, Saudi Arabia, funded the APC and financial incentives.

Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request, subject to Prince Sultan University and heritage-site permissions.

Acknowledgments: The authors gratefully acknowledge Prince Sultan University and the RIC (Research and Initiative Center) for covering the APC charges, providing financial incentives, and supporting research collaboration at the SaLab (Sustainable Architecture Laboratory), College of Architecture and Design. The authors thank the College of Architecture and Design, Department of Architecture, the Sustainable Architecture Laboratory (SaLab) and participants in SEED Project n. 208 2024–2026 for all their technical support and knowledge transfer during collaboration and research.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Secondary Figures

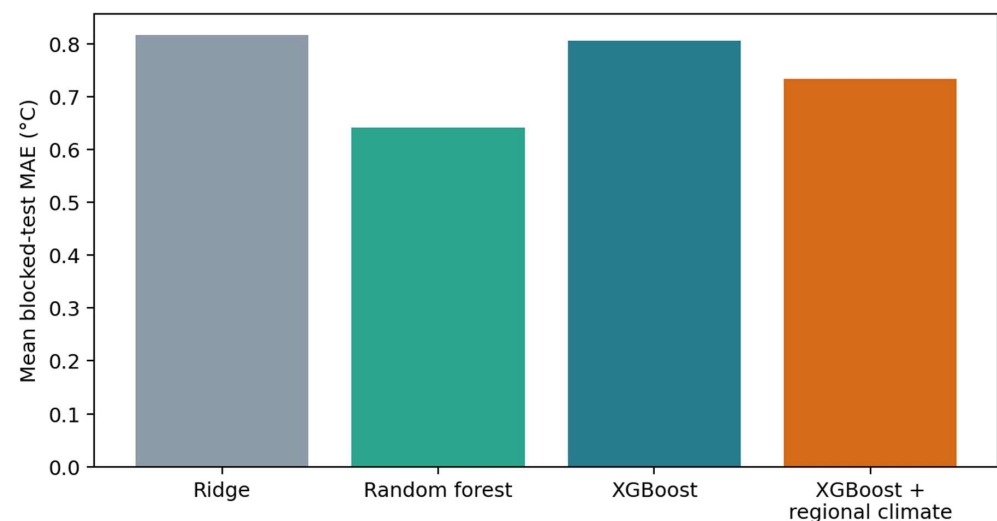


Figure A1. Revised mean MAE across the same five seven-day blocked tests. The fourth model adds regional radiation, wind speed, and cyclic wind direction to XGBoost.

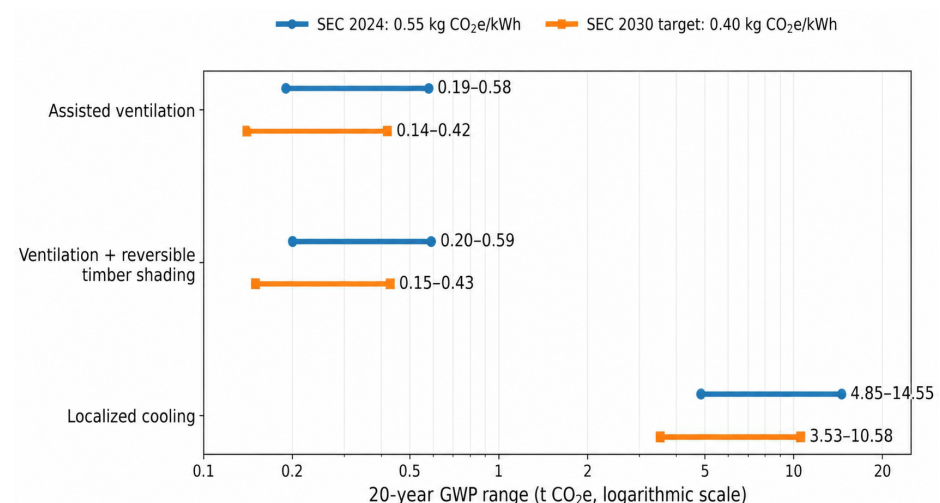


Figure A2. Twenty-year positive-emissions screening of escalating thermal-management actions. The reversible-shading package includes two 0.15 m³ sawn-timber installations represented with an A1–A3 GWP-fossil proxy rather than full GWP-total biogenic accounting; localized active cooling includes B6 electricity only and is, therefore, a conservative lower-bound comparison.

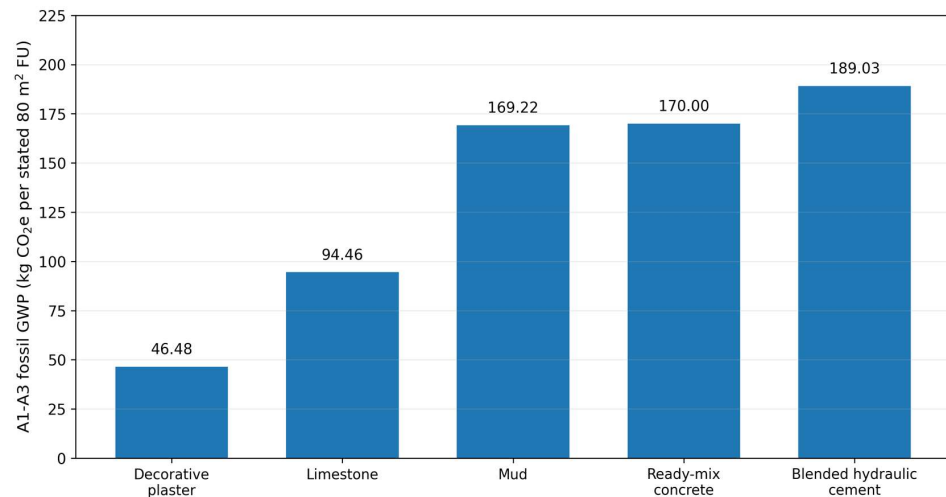


Figure A3. Site-specific Ushaiger restoration-material A1–A3 fossil-GWP benchmark reproduced numerically from the published local LCA [17]. The products serve different restoration functions and are shown as contextual carbon evidence, not as direct substitutes.

Appendix B. Reanalysis Methods and Robustness Diagnostics

All tables in this appendix derive from the supplied six Netatmo exports and regional CSV using the declared revision script. The raw audit is unchanged. Source measurements are never modified. Overlapping validation windows are not independent samples, and scenario percentiles are not calibrated summer confidence limits.

Table A1. Gap-length sensitivity of blind reconstruction.

Gap (Days)	Windows	Tin Mean MAE (°C)	Tin Median MAE (°C)	Tout Mean MAE (°C)	RHin Mean MAE (pp)	RHout Mean MAE (pp)
7	6	0.93	0.87	1.26	3.27	5.10
14	6	0.83	0.87	1.55	3.20	5.13
28	6	1.14	1.10	2.08	5.98	11.57
42	6	1.31	1.04	2.15	7.17	11.11
61	6	1.35	1.25	3.08	7.79	14.99

Each length uses six deterministically selected windows from weekly candidate starts, excluding the first/last seven days and requiring $\geq 75\%$ measured coverage plus ≥ 12 observations per day. All fields are masked and models refitted. Seven-day starts: 17 September, 29 October, 3 December, and 31 December 2025; 4 February and 12 August 2026. Full start/end dates and scored counts are supplied in validation_windows.csv. The 28-, 42-, and 61-day windows are restricted to October–February and overlap substantially.

Table A2. Cluster separation and stability.

k	Inertia	Silhouette	Davies–Bouldin	Seed ARI Median/min	Bootstrap ARI Median/min
2	1379.3	0.398	0.999	1.000/1.000	0.961/0.849
3	1137.4	0.360	1.187	1.000/0.986	0.902/0.524
4	964.3	0.290	1.191	1.000/0.968	0.840/0.649
5	856.4	0.316	1.105	0.956/0.508	0.561/0.399
6	755.6	0.286	1.232	0.930/0.914	0.682/0.524

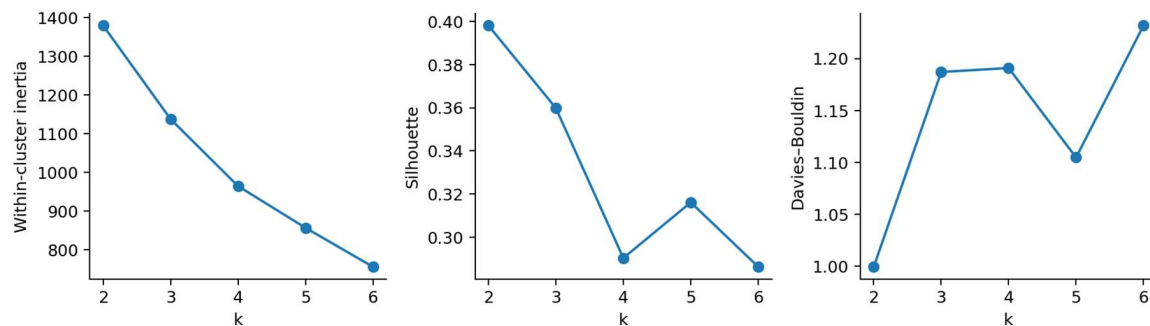


Figure A4. Inertia, silhouette, and Davies–Bouldin diagnostics for $k = 2$ – 6 . The four-state solution is operational rather than the statistically best-separated partition.

Table A3. SHAP sampling robustness.

Sample Size	Repeats	Sampling	Most Frequent First Feature	First-Rank Frequency	Minimum Rank Correlation
300	30	Equal seasons	3 h outdoor mean	30/30	0.991
900	30	Equal seasons	3 h outdoor mean	30/30	0.991
2000	30	Equal seasons	3 h outdoor mean	30/30	0.991

All 8314 eligible measured targets are attributed before sampling. DJF, MAM, JJA, and SON have 2595, 1256, 1449, and 3014 targets, respectively. The high-temperature measured subset has 1830 targets and the same leading feature. June has no measured target; a balanced seasonal sample, therefore, cannot validate June. Correlations compare full feature rankings with the equal-season reference, with strongly correlated outdoor means interpreted cautiously.

Table A4. Counterfactual assumptions and objective definitions.

Item	Value or Definition	Interpretation
η reference	0.061 per half-hour	Effective scenario coupling, not measured air-change rate
η sensitivity	0.02, 0.03, 0.05, 0.061, 0.07, 0.09, 0.10	2–10% driving-difference correction per active step; uncalibrated range
ρ reference and grid	0.865; grid 0.75, 0.80, 0.865, 0.90, 0.95	Reported reference persistence retained; joint scenario sensitivity
Thermal objective	Degree-hours above each tested threshold	Counterfactual high-temperature exposure
Burden objective	Actual active hours, with η fixed within each run	Operating effort separated from effective coupling
Moisture eligibility	Outdoor RH < 70%; $AH_{out} \leq AH_{in} + 2 \text{ g/m}^3$	Baseline moisture screen; no counterfactual moisture dynamics
Decision opportunity	20:00–07:00; $T_{out} \leq T_{in} - 1.40 \text{ } ^\circ\text{C}$	Fixed diagnostic definition independent of the revised optimum
Carbon runtime	588 h, fixed reference	Common power comparison; revised controller uses 361.5 h

Shares use hot days at each threshold. The diagnostic cooling opportunity, coupling, persistence, and moisture rules remain fixed; the three schedule variables are reoptimized for each temperature threshold. Fixed 35 °C splits result in A/D counts of 29/67, 19/58, 45/44, 36/18, and 7/0 at 30, 32, 35, 37, and 40 °C, respectively. This comparison separates

threshold effects from changes in the local median splits. Full B/C counts and schedules are supplied in threshold_sensitivity.csv.

Table A5. Temperature-threshold sensitivity with locally recalculated medians.

Threshold (°C)	Hot Days	Class A	Class D	H*/C*	Reoptimized Reduction (%)
30	183	64 (35.0%)	58 (31.7%)	119.70/7.0	2.03
32	155	56 (36.1%)	51 (32.9%)	87.35/6.5	2.55
35	124	45 (36.3%)	44 (35.5%)	37.74/6.5	2.62
37	89	29 (32.6%)	28 (31.5%)	14.20/5.5	1.88
40	25	11 (44.0%)	10 (40.0%)	4.75/1.0	0.95

Table A6. Reconstruction-error propagation with fixed class splits and controller.

Scenario	Hot Days	A Days	D Days	Reduction (%)	June A/D	July A/D
Indoor shift −0.97 °C	117	34	51	0.42	16/2	1/23
Indoor shift +0.00 °C	124	45	44	2.62	22/0	3/17
Indoor shift +0.97 °C	136	53	33	5.78	16/0	4/7
Joint residual blocks 5th	115.0	23.0	37.0	1.00	2.0/0.0	1.0/9.0
Joint residual blocks median	122.0	31.0	47.0	2.45	9.0/4.0	2.0/17.0
Joint residual blocks 95th	129.0	41.0	57.0	4.14	18.0/11.0	5.0/23.0

Systematic shifts change reconstructed indoor temperatures only. The 200 joint draws perturb all four reconstructed environmental fields using seven-day blocks drawn from the ≥ 28 -day validation residuals. Measured timestamps, reference H*/C*, and the selected controller remain fixed. Fractional day counts are empirical percentile interpolation. These envelopes measure the tested scenario sensitivity; they are not calibrated probabilistic confidence intervals for the missing summer observations.

Table A7. NSGA-II seed agreement and convergence.

Seed	Start/Duration/Trigger	Degree-Hours	Active Hours	Reduction (%)	Schedules	Final-50 HV Change (%)
2026	17.5/10.0/1.5	4711.32	361.5	2.62	90	+0.00442
2027	17.5/10.0/1.5	4711.32	361.5	2.62	93	+0.00083
2028	18.0/9.5/1.5	4711.32	361.5	2.62	87	−0.00402
2029	17.5/10.0/1.5	4711.32	361.5	2.62	94	+0.00545
2030	17.5/10.0/1.5	4711.32	361.5	2.62	88	+0.00531

Population 100; 200 generations; SBX crossover probability of 0.9, distribution index of 15; polynomial mutation individual probability of 1.0, per-variable probability of 1/3, distribution index of 20; random real-valued sampling; rank/crowding survival; and duplicate-genotype elimination. The retained schedules are rounded to 0.5 h, 0.5 h, and 0.1 °C. Hypervolume scales degree-hours by the no-intervention baseline and active hours by available monitoring hours. The seed-2026 knee has a zero degree-hour and active-hour difference from the exhaustive 10,166-schedule-grid knee. Schedule timings are not uniquely identifiable where eligibility patterns and objectives coincide. The final-set convergence is not proof of optimum outside the declared grid or physical validity.

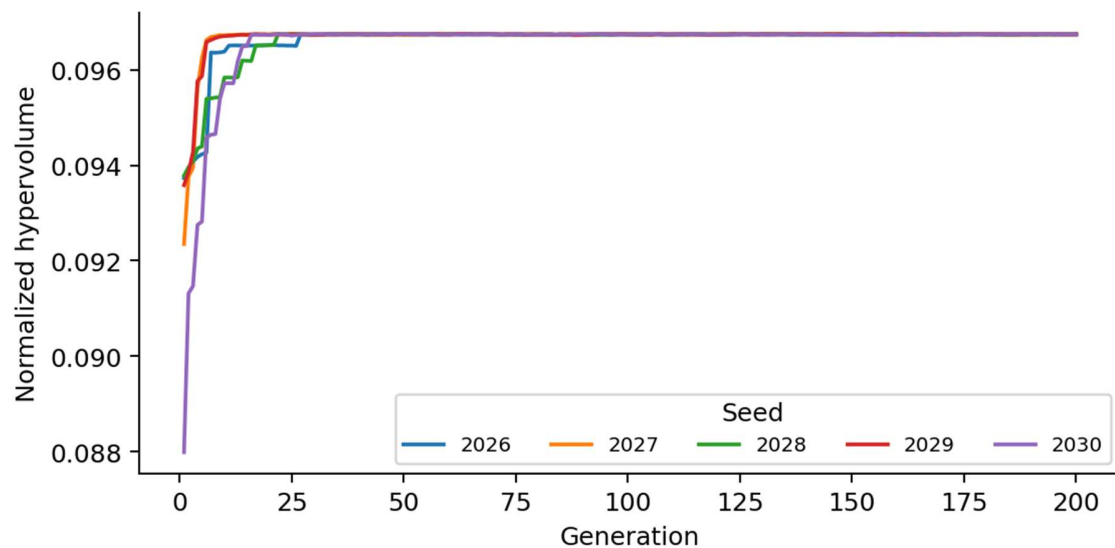


Figure A5. Hypervolume trajectories for the five declared NSGA-II seeds. Small final changes and agreement with the exhaustive grid support numerical convergence within the declared scenario.

Table A8. Paired regional-climate model comparison.

Model	Mean MAE (°C)	Median MAE (°C)	Mean RMSE (°C)	Eligibility
XGBoost	0.806	0.570	0.961	Same five blocks and 8314 eligible targets
XGBoost + regional climate	0.734	0.558	0.896	Same five blocks and 8314 eligible targets

Wind direction is converted to sine/cosine components before time interpolation to avoid the $0^\circ/360^\circ$ discontinuity. These components, radiation, and wind speed are interpolated from hourly regional records to half-hour timestamps. Both models use the same measured targets, measured outdoor history, folds, exclusions, and XGBoost parameters. Full block-level errors and sample sizes are supplied in `surrogate_by_block.csv`. Source metadata do not independently establish ERA5 provenance.

Software versions: numpy 2.3.5, pandas 2.2.3, scipy 1.17.0, scikit-learn 1.8.0, xgboost 3.4.1, shap 0.52.0, numba 0.68.0, pymoo 0.6.2, and matplotlib 3.10.8. The analysis script uses a random seed of 2026 unless a listed stability seed replaces it. Input filenames and source timestamps are retained in the supplementary records.

References

1. Saudi Vision 2030. Vision 2030 Annual Report 2025. Available online: https://www.vision2030.gov.sa/media/ecdjfopq/vision2030_annual_report_2025_en.pdf (accessed on 2 September 2026).
2. Heritage Commission, Ministry of Culture, Kingdom of Saudi Arabia. Rehabilitation of Historic Buildings and Adaptation to Contemporary Use: Best Practices and Models. Available online: <https://engage.moc.gov.sa/imprint-of-heritage/practices/> (accessed on 2 September 2026).
3. Saudi Building Code National Committee. The Saudi Historical Building Code (SBC 902). Available online: https://sbc.gov.sa/ar/BC/Documents/tableofcontent2024/SBC%20902/SBC902_CR_241224-FA.pdf (accessed on 2 September 2026).
4. Al-Hemiddi, N.A.; Megren Al-Saud, K.A. The Effect of a Ventilated Interior Courtyard on the Thermal Performance of a House in a Hot-Arid Region. *Renew. Energy* **2001**, *24*, 581–595. [\[CrossRef\]](#)
5. Al-Oraier, F. The Thermal Analysis of Traditional Adobe Dwellings in Riyadh City, Saudi Arabia. Ph.D. Thesis, Cardiff University, Cardiff, UK, 2005.
6. Alnaim, M.M. The Hierarchical Order of Spaces in Arab Traditional Towns: The Case of Najd, Saudi Arabia. *World J. Eng. Technol.* **2020**, *8*, 347–366. [\[CrossRef\]](#)
7. Alnaim, M.M. Dwelling Form and Culture in the Traditional Najdi Built Environment, Saudi Arabia. *Open House Int.* **2021**, *46*, 595–614. [\[CrossRef\]](#)

8. I Dwidar, S. Bioclimatic Architecture for Heritage Residential Buildings in the Kingdom of Saudi Arabia Environmental Design Approach Towards Providing Thermal Comfort in Future Buildings. *JES J. Eng. Sci.* **2019**, *47*, 868–882. [[CrossRef](#)]
9. Mazzetto, S. Fostering National Identity Through Sustainable Heritage Conservation: Ushaiger Village as a Model for Saudi Arabia. *Heritage* **2025**, *8*, 4. [[CrossRef](#)]
10. Carrobé, A.; Rincón, L.; Martorell, I. Thermal Monitoring and Simulation of Earthen Buildings. A Review. *Energies* **2021**, *14*, 2080. [[CrossRef](#)]
11. Ben-Alon, L.; Rempel, A.R. Thermal Comfort and Passive Survivability in Earthen Buildings. *Build. Environ.* **2023**, *238*, 110339. [[CrossRef](#)]
12. Alayed, E.; Bensaid, D.; O'Hegarty, R.; Kinnane, O. Thermal Mass Impact on Energy Consumption for Buildings in Hot Climates: A Novel Finite Element Modelling Study Comparing Building Constructions for Arid Climates in Saudi Arabia. *Energy Build.* **2022**, *271*, 112324. [[CrossRef](#)]
13. Wakil, M.; El Mghari, H.; Idrissi Kaitouni, S.; El Amraoui, R. Thermal Energy Performance of Compressed Earth Building in Two Different Cities in Moroccan Semi-Arid Climate. *Energy Built Environ.* **2024**, *5*, 800–816. [[CrossRef](#)]
14. Idrissi Kaitouni, S.; Charai, M.; Es-sakali, N.; Mghazli, M.O.; El Mankibi, M.; Uk-Joo, S.; Ahachad, M.; Brigui, J. Energy and Hygrothermal Performance Investigation and Enhancement of Rammed Earth Buildings in Hot Climates: From Material to Field Measurements. *Energy Build.* **2024**, *315*, 114325. [[CrossRef](#)]
15. Mazzetto, S. Urban Heritage Conservation in Saudi Arabia Toward Sustainable Reuses. *Adv. Sci. Technol.* **2024**, *137*, 79–86. [[CrossRef](#)]
16. Alhojaly, R.A.; Alawad, A.A.; Ghabra, N.A. A Proposed Model of Assessing the Adaptive Reuse of Heritage Buildings in Historic Jeddah. *Buildings* **2022**, *12*, 406. [[CrossRef](#)]
17. Mazzetto, S. Comparative Life Cycle Assessment of Traditional and Modern Materials in Heritage Building Restoration: A Case Study from Ushaiger Village. *Sustainability* **2025**, *17*, 25. [[CrossRef](#)]
18. Cheng, J.C.P.; Zhang, J.; Kwok, H.H.L.; Tong, J.C.K. Thermal Performance Improvement for Residential Heritage Building Preservation Based on Digital Twins. *J. Build. Eng.* **2024**, *82*, 108283. [[CrossRef](#)]
19. Mohamed, B.; Marzouk, M. Post-Adaptive Reuse Evaluation of Heritage Buildings Using Multi-Criteria Decision-Making Techniques. *J. Build. Eng.* **2025**, *99*, 111485. [[CrossRef](#)]
20. Wang, Y.; Zhen, Y.; Ning, Q. Four-Objective Optimization of Historic Building Renovation Considering Climate Change: A Trade-Off in Energy, Economy, Environment, and Thermal Comfort. *J. Build. Eng.* **2025**, *110*, 113017. [[CrossRef](#)]
21. Fokaides, P.A. Development of a Heritage Life Cycle Assessment (H-LCA) Framework: Integrating Sustainability Metrics and Cultural Preservation in Heritage Buildings. *Energy Rep.* **2025**, *14*, 4437–4448. [[CrossRef](#)]
22. Mazzetto, S.; Alnaim, M.M. Thermal Performance of Earthen Architecture in Ushaiger, Saudi Arabia: A Pilot Digital-Twin Feasibility Study. *Sustainability* **2026**, *18*, 3634. [[CrossRef](#)]
23. Yang, W.; Xu, J.; Lu, Z.; Yan, J.; Li, F. A Systematic Review of Indoor Thermal Environment of the Vernacular Dwelling Climate Responsiveness. *J. Build. Eng.* **2022**, *53*, 104514. [[CrossRef](#)]
24. Heracleous, C.; Panagiotou, R.; Ioannou, I.; Michael, A.; Philokyprou, M. Hygrothermal Performance Monitoring and Indoor Thermal Evaluation in Adobe Structures. *Energy Build.* **2025**, *346*, 116049. [[CrossRef](#)]
25. Shukla, A.; Tiwari, G.N.; Sodha, M.S. Embodied Energy Analysis of Adobe House. *Renew. Energy* **2009**, *34*, 755–761. [[CrossRef](#)]
26. Ben-Alon, L.; Loftness, V.; Harries, K.A.; DiPietro, G.; Hameen, E.C. Cradle to Site Life Cycle Assessment (LCA) of Natural vs Conventional Building Materials: A Case Study on Cob Earthen Material. *Build. Environ.* **2019**, *160*, 106150. [[CrossRef](#)]
27. Belarbi, Y.E.; Ferroukhi, M.Y.; Issaadi, N.; Poullain, P.; Bonnet, S. Assessment of Hygrothermal Performance of Raw Earth Envelope at Overall Building Scale. *Energy Build.* **2024**, *310*, 114119. [[CrossRef](#)]
28. Ramezannia, A.; Gocer, O.; Bashirzadeh Tabrizi, T. The Life Cycle Assessment of Stabilized Rammed Earth Reinforced with Natural Fibers in the Context of Australia. *Constr. Build. Mater.* **2024**, *416*, 135034. [[CrossRef](#)]
29. Khaksar, A.; Tabadkani, A.; Mofidi Shemirani, S.M.; Hajirasouli, A.; Banihashemi, S.; Attia, S. Thermal Comfort Analysis of Earth-Sheltered Buildings: The Case of Meymand Village, Iran. *Front. Archit. Res.* **2022**, *11*, 1214–1238. [[CrossRef](#)]
30. Dwidar, S.; Sirror, H.; Derbali, A.; Abdelgawad, D.; Abdelsattar, A. Importance If Internal Courtyards in Designing Historical and Contemporary Masjid. *J. Islam. Archit.* **2022**, *7*, 356–363. [[CrossRef](#)]
31. Salameh, M.; Abu-Hijleh, B.; Touqan, B. Impact of Courtyard Orientation on Thermal Performance of School Buildings' Temperature. *Urban Clim.* **2024**, *54*, 101853. [[CrossRef](#)]
32. Basunbul, A.I.; Braham, W.W. Climate Responsive Lessons from an Architectural and Ethnographic Study of Hijazi Traditional Dwellings in Jeddah, Saudi Arabia. *Front. Built Environ.* **2023**, *9*, 1127615. [[CrossRef](#)]
33. Sirror, H. Lessons Learned from the Past: Tracing Sustainable Strategies in the Architecture of Al-Ula Heritage Village. *Sustainability* **2024**, *16*, 5463. [[CrossRef](#)]
34. Ascione, F.; de Rossi, F.; Vanoli, G.P. Energy Retrofit of Historical Buildings: Theoretical and Experimental Investigations for the Modelling of Reliable Performance Scenarios. *Energy Build.* **2011**, *43*, 1925–1936. [[CrossRef](#)]

35. Frasca, F.; Verticchio, E.; Cornaro, C.; Siani, A.M. Performance Assessment of Hygrothermal Modelling for Diagnostics and Conservation in an Italian Historical Church. *Build. Environ.* **2021**, *193*, 107672. [\[CrossRef\]](#)
36. Mu, W.; Cardelli, R.; Ferrari, S. Data Preprocessing Techniques for Machine Learning Towards Improving Building Energy Performance: A Systematic Review. *Energies* **2026**, *19*, 1561. [\[CrossRef\]](#)
37. Vuoto, A.; Funari, M.F.; Lourenço, P.B. Shaping Digital Twin Concept for Built Cultural Heritage Conservation: A Systematic Literature Review. *Int. J. Archit. Herit.* **2023**, *18*, 1762–1795. [\[CrossRef\]](#)
38. Hou, H.; Lai, J.H.K.; Wu, H.; Wang, T. Digital Twin Application in Heritage Facilities Management: Systematic Literature Review and Future Development Directions. *Eng. Constr. Archit. Manag.* **2023**, *31*, 3193–3221. [\[CrossRef\]](#)
39. Rolfsen, C.N.; Lassen, A.K.; Han, D.; Hosamo, H.; Ying, C. The Use of the BIM-Model and Scanning in Quality Assurance of Bridge Constructions. In *ECPPM 2021-eWork and eBusiness in Architecture, Engineering and Construction*; CRC Press: Boca Raton, FL, USA, 2021; pp. 357–360.
40. Shabani, A.; Hosamo, H.; Plevris, V.; Kioumarsis, M. *A Preliminary Structural Survey of Heritage Timber Log Houses in Tønsberg, Norway*; International Centre for Numerical Methods in Engineering: Barcelona, Spain, 2021; ISBN 978-84-123222-0-0.
41. Puertas, E.; Ávila, F.; García-Macías, E.; Gallego, R. Preventive Preservation of Rammed Earth Historical Heritage Through Continuous Monitoring, Architectural Inspections, and Data Fusion. *Buildings* **2024**, *14*, 3294. [\[CrossRef\]](#)
42. Wang, T.; Li, Z. Identification and Risk Assessment of Erosion Characteristics of the Rammed-Earth Great Wall from the Perspective of Preventive Conservation. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2025**, X-M-2–2025, 335–339. [\[CrossRef\]](#)
43. Hosamo, H.H.; Nielsen, H.K.; Kraniotis, D.; Svennevig, P.R.; Svidt, K. Improving Building Occupant Comfort through a Digital Twin Approach: A Bayesian Network Model and Predictive Maintenance Method. *Energy Build.* **2023**, *288*, 112992. [\[CrossRef\]](#)
44. Hosamo, H.; Mazzetto, S. Data-Driven Ventilation and Energy Optimization in Smart Office Buildings: Insights from a High-Resolution Occupancy and Indoor Climate Dataset. *Sustainability* **2025**, *17*, 58. [\[CrossRef\]](#)
45. Hilliaho, K.; Lahdensivu, J.; Vinha, J. Glazed Space Thermal Simulation with IDA-ICE 4.61 Software—Suitability Analysis with Case Study. *Energy Build.* **2015**, *89*, 132–141. [\[CrossRef\]](#)
46. Hilliaho, K.; Nordquist, B.; Wallentén, P.; Hamid, A.A.; Lahdensivu, J. Energy Saving and Indoor Climate Effects of an Added Glazed Facade to a Brick Wall Building: Case Study. *J. Build. Eng.* **2016**, *7*, 246–262. [\[CrossRef\]](#)
47. Nord, N.; Tereshchenko, T.; Woszczek, A.; Naess, J.S.; Sandberg, N.H.; Pahlavan, H.M.; Brattebø, H. A Dynamic Modelling Approach to Explore Zero Emission Building Stock Opportunities towards 2050—Case Study of a University Campus. *Energy Build.* **2024**, *325*, 115024. [\[CrossRef\]](#)
48. Schade, J.; Ylmén, P.; Wallenberg, N.; Thorsson, S. Overheating in a Common Swedish Residential Multi-Story Building under Current and Future Climate—Integrating Trees to Combat Overheating. *City Environ. Interact.* **2025**, *27*, 100206. [\[CrossRef\]](#)
49. Lin, D.; Wu, T.; Liu, K.; Liu, Y.; Zhao, L.; Xu, X. Explainable Multi-Objective Optimization for Building Retrofit Considering Uncertainty. *Energy* **2026**, *342*, 139679. [\[CrossRef\]](#)
50. Pan, Y.; Zhong, W.; Zheng, X.; Xu, H.; Zhang, T. Natural Ventilation in Vernacular Architecture: A Systematic Review of Bioclimatic Ventilation Design and Its Performance Evaluation. *Build. Environ.* **2024**, *253*, 111317. [\[CrossRef\]](#)
51. Bagasi, A.A.; Calautit, J.K. Experimental Field Study of the Integration of Passive and Evaporative Cooling Techniques with Mashrabiya in Hot Climates. *Energy Build.* **2020**, *225*, 110325. [\[CrossRef\]](#)
52. Alhazmi, H.; Alduwais, A.K.; Tabbakh, T.; Aljamlani, S.; Alkahlan, B.; Kurdi, A. Environmental Performance of Residential Buildings: A Life Cycle Assessment Study in Saudi Arabia. *Sustainability* **2021**, *13*, 3542. [\[CrossRef\]](#)
53. Elnaggar, A. Nine Principles of Green Heritage Science: Life Cycle Assessment as a Tool Enabling Green Transformation. *Herit. Sci.* **2024**, *12*, 7. [\[CrossRef\]](#)
54. Hosamo, H.; Coelho, G.B.A.; Buvik, E.; Drissi, S.; Kraniotis, D. Building Sustainability through a Novel Exploration of Dynamic LCA Uncertainty: Overview and State of the Art. *Build. Environ.* **2024**, *264*, 111922. [\[CrossRef\]](#)
55. Open-Meteo. Historical Weather API Documentation. Available online: <https://open-meteo.com/en/docs/historical-weather-api> (accessed on 4 October 2026).
56. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
57. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation, Inc.: Diego, CA, USA, 2017; Volume 30, pp. 4765–4774.
58. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [\[CrossRef\]](#)
59. Saudi Electricity Company. Sustainability Report 2024. Available online: <https://www.se.com.sa/en/Investors/Column3/ESG/ESG-Reports/> (accessed on 2 September 2026).

60. Stora Enso. Stora Enso EPD Guide. Available online: <https://www.storaenso.com/-/media/documents/download-center/documents/sustainability/stora-enso-epd-guide.pdf> (accessed on 2 September 2026).
61. EN 15804:2012+A2:2019; Sustainability of Construction Works—Environmental Product Declarations—Core Rules for the Product Category of Construction Products. European Committee for Standardization (CEN): Brussels, Belgium, 2019.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.