TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Jalilzadeh, A., Rafiee, A., Konstantinou, T., de Graaf, S., Scholten, H., & van Oosterom, P. J. M. (2026). DELFT-UBEM: An automated database-driven pipeline for district-scale building energy modeling, demonstrated on Dutch district retrofit. *Energy and Buildings*, *373*, Article 118345. https://doi.org/10.1016/j.enbuild.2026.118345

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

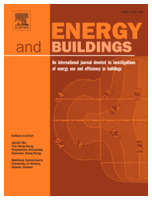
Energy & Buildings 373 (2026) 118345



Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enbuild



DELFT-UBEM: An automated database-driven pipeline for district-scale building energy modeling, demonstrated on Dutch district retrofit

Amin Jalilzadeh [a,*], Azarakhsh Rafiee [a], Thaleia Konstantinou [b], Stefan de Graaf [c], Henk Scholten [d], Peter J.M. van Oosterom [a]

[a] Section GIS Technology, Faculty of Architecture and the Built Environment, TU Delft, Delft, Netherlands
[b] Department of AE+T, Faculty of Architecture and the Built Environment, TU Delft, Delft, Netherlands
[c] Research Department, Geodan, Amsterdam, Netherlands
[d] Faculty of Economics, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

ARTICLE INFO

Keywords:
Urban building energy modeling
EnergyPlus
Building retrofit
City planning
Energy transition
Digital twin

ABSTRACT

District-scale energy transition planning requires building-level energy profiles, yet current Urban Building Energy Modeling (UBEM) workflows require substantial manual data preparation with no integrated path from building registries to scenario comparison. This paper presents DELFT-UBEM, an automated tool that generates dynamic simulation models from building registries and national performance standards through a lookup-table architecture that maps building age, function, and energy label to physical parameters, translating regulatory documents into simulation-ready parameters within a unified, database-driven pipeline. The pipeline shifts repeated project-level preparation into reusable database, lookup-table, and configuration components: once these components and a geographic boundary are available, data ingestion, geometric processing, archetype assignment, Input Data File (IDF) generation, parallel simulation, and post-processing execute as a connected automated workflow, producing analysis-ready results for downstream policy evaluation. Against national benchmarks, the model reproduces structural variation in Dutch residential heating demand across 35 archetypes without building-specific calibration. Comparison with measured postcode-level gas consumption across 63 postcodes (~3,000 buildings) corroborates this at aggregate scale, with residual errors systematically traceable to known data barriers and modeling limitations. To demonstrate downstream analytical capability, two contrasting Dutch districts were used to define custom retrofit and electrification applications, producing path-dependent marginal abatement cost curves with building-level sequencing across 3,404 upgrade steps. The paper discusses open challenges in UBEM validation standards and cross-tool comparability, highlighting how the pipeline's modular architecture supports continuous refinement. The complete pipeline, archetype library, and parameter mappings are published as open-source resources, and DELFT-UBEM is accessible as a web-based digital twin platform.

1. Introduction

Buildings account for approximately 40% of final energy consumption in the EU and are central to climate mitigation strategies [1]. Planning urban energy transitions (electrification, retrofits, local renewables) requires building-level energy profiles that capture how individual interventions interact across heterogeneous building stocks [2,3]. The 2024 recast of the EU Energy Performance of Buildings Directive [1] mandates building-level renovation roadmaps across all member states, creating urgent demand for scalable building energy modeling that current workflows cannot meet. Meeting this requirement means capturing building-level physics while scaling to hundreds or thousands of buildings, and stock-scale models can be organized by aggregation level and formulation. Top-down statistical models infer aggregate demand from climatic, socioeconomic, and stock indicators; they scale readily but do not retain the building-specific fabric and system detail that intervention analysis requires [4]. Bottom-up data-driven models learn consumption from measured or disclosed data and have matured rapidly, from city-scale prediction off partial benchmarking samples [5] to learned surrogates and hybrid physics couplings [6–8]. Two structural limits still constrain them for transition planning. First, models fitted to observed operation extrapolate poorly to deep retrofits or heating-system replacements that alter the learned relationships, so counterfactual evaluation still rests on an explicit physical backbone [9,10]. Second, they require

* Corresponding author.
E-mail addresses: A.Jalilzadeh@tudelft.nl (A. Jalilzadeh), A.Rafiee@tudelft.nl (A. Rafiee), T.Konstantinou@tudelft.nl (T. Konstantinou), stefan.de.graaf@geodan.nl (S. de Graaf), h.j.scholten@vu.nl (H. Scholten), P.J.M.vanOosterom@tudelft.nl (P.J.M. van Oosterom).

https://doi.org/10.1016/j.enbuild.2026.118345
Received 22 April 2026; Received in revised form 7 September 2026; Accepted 2 October 2026
Available online 5 October 2026
0378-7788/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

building-level metered or disclosed consumption that most jurisdictions do not publish, and where disclosure exists it covers only part of the stock [5], which limits transferability across stocks and climates [10]. In the Netherlands the finest openly published resolution is the postcode aggregate.

Bottom-up physics-based models represent thermal and system mechanisms explicitly and can therefore evaluate previously unobserved interventions, at the cost of substantially larger geometry, archetype, schedule, and parameterization requirements [11,12]. DELFT-UBEM adopts this family and drives it from open national data and standards rather than building-level meter data, sidestepping both data-driven limits while inheriting the physics family's own burden of data preparation and workflow effort, the barrier this paper targets. "Database-driven" here names that data architecture, not a predictive machine-learning method.

While single-building Building Energy Modeling (BEM) captures these physics in detail [13], scaling it to entire districts multiplies the data management, geometric processing, and parameterization effort by orders of magnitude [14]. Urban Building Energy Modeling (UBEM) addresses this scaling challenge by extending BEM principles to the district or city scale, using archetypes and shared assumptions to manage the combinatorial complexity of stock diversity [12,15]. The archetype approach (categorizing buildings into representative types defined by age, function, and construction) offers reliable stock-level estimates at the cost of individual building accuracy [3,15]. Established tools span two broad categories [14]: detailed physics-based platforms (e.g., City Building Energy Saver (CityBES) [16], URBANopt (URBANopt) [17]) that run full dynamic simulation engines like EnergyPlus (EnergyPlus) [18]; and reduced-order model-based tools (e.g., CitySim (CitySim) [19], City Energy Analyst (CEA) [20], TEASER [21]) that use resistance-capacitance thermal networks for computational efficiency.

Existing UBEM platforms automate different subsets of the modeling workflow, and can be compared along explicit criteria: which stages are automated (data ingestion from open datasets, geometry generation, standards-based parameterization, scenario generation, and downstream policy analysis), whether state passes between stages through a shared database or through exchanged files, whether stable building identifiers and parameters are carried end-to-end, and whether the tool is openly available and reproducible [14,22]. Assessed this way, established tools such as CityBES [16], URBANopt [17], CEA [20], TEASER [21], and CESAR-P [23], together with more recent frameworks including AutoBPS [24], UBEM.io [25], and GeoBEM [26], each automate several stages but share two recurring gaps. First, they exchange state between stages as files rather than through a database, so building identity, parameters, and provenance are not carried continuously; the platforms that do use a database core (CityBES, and partly GeoBEM) are closed rather than open. Second, model creation is automated more thoroughly than downstream use, with several tools stopping short of physically-constrained scenario generation or built-in policy analysis (TEASER, for instance, provides no built-in policy analysis). Each project reconstructs its own data pipelines, geometry processing routines, and archetype assignments from scratch [3,27]. The combinatorial complexity grows quickly: a modest district study comparing five retrofit scenarios across 1,000 buildings already requires 5,000 simulation configurations with tracked parameter provenance, tasks that current tools leave to ad-hoc scripts and spreadsheets [25,28]. A taxonomic review of 72 studies found that undocumented data models and workflow fragmentation render the vast majority of published UBEM results non-reproducible [29], a gap that structured, schema-enforced data exchange directly addresses [3], while systematic reviews of data acquisition confirm that limited data availability and the difficulty of determining stochastic occupant and operational inputs remain the primary bottlenecks at urban scale [30]. Recent work explores alternative automation strategies: automated generation of prototype building models from city-scale building datasets [24] and large language models to generate individual EnergyPlus models from natural language [31], alongside GPT-based data-driven UBEM for urban energy analytics and demand prediction [32]. These approaches, however, address model creation rather than the end-to-end data management, scenario orchestration, and downstream analysis that district-scale planning requires.

This workflow complexity, rather than simulation capability, is a principal barrier to UBEM adoption, confining it to specialists and limiting its uptake in municipal planning where building-level analysis is most needed [25,30,33,34]. This raises the question of whether the complete UBEM workflow can be automated end-to-end from open national data and regulatory standards alone, and what level of accuracy and analytical capability such an automated approach can deliver for district-scale energy transition planning.

To answer this question, this paper develops and validates DELFT-UBEM, an open-source tool that automates the complete UBEM pipeline. Rather than improving individual workflow stages, DELFT-UBEM orchestrates the complete pipeline as a single system. The paper proceeds through four steps:

1. Design and implement an end-to-end UBEM pipeline that generates district-scale EnergyPlus models from open national data and regulatory standards alone, shifting per-project preparation effort into reusable database, lookup-table, and configuration components, so that a new district requires only a configuration file and a geographic boundary.
2. Structure the pipeline around a unified, database-driven data architecture that replaces fragmented file-based workflows and handles automated geometry processing, standards-driven parameterization, and physically-constrained scenario generation within a single system. Among the tools surveyed, this pairing of a database-mediated architecture with an open-source release and automated ingestion from open national data is unique to DELFT-UBEM, and its "end-to-end" character denotes continuity of stable building identifiers, parameters, and results across all stages rather than the sequential execution of separately automated tools.
3. Validate the model at two complementary scales without building-specific calibration (archetype benchmarks and measured postcode consumption), characterizing the accuracy envelope of archetype-based UBEM and attributing residual errors to identified data gaps.
4. Apply the validated pipeline to two Dutch districts to compare envelope and electrification measures through path-dependent marginal abatement cost curves, and evaluate how the resulting intervention ranking depends on baseline assignment, heat-pump performance, grid carbon intensity, capital costs, discounting, and energy-price assumptions.

While the pipeline supports flexible parameter modification, this paper demonstrates it through the retrofit package mode, where predefined renovation levels derived from Dutch performance standards define the intervention space.

Section 2 opens with a complete roadmap of the pipeline and downstream analyzes; Section 4 examines accuracy boundaries and open challenges in UBEM validation and cross-tool comparability; and Section 5 summarizes findings and future directions.

2. Methodology

2.1. System architecture

DELFT-UBEM treats the complete workflow, from building registry to scenario comparison, as a single database-mediated pipeline (Fig. 1). For a prepared national implementation, each run is initiated with a geographic boundary and a versioned configuration file. The configuration records the input source, parameter-selection strategy and seed, per-building / per-archetype / stock-wide overrides, weather assignment, zoning and shading options, scenario assignment (including the missing-label rule), simulation workers, and output frequencies; economic and emission assumptions are supplied through the associated

analysis configuration (Appendix D enumerates the configurable fields and the values used for the reported studies). All building attributes, archetype parameters, and material properties are resolved from a relational database and versioned lookup tables.

The pipeline proceeds through four core operational stages, followed by a performance evaluation and two downstream analyzes, each detailed in its own subsection:

1. **Data management** — a shared PostgreSQL/PostGIS layer ingests national registries and user-provided inputs, and emits structured per-building records.
2. **Geometric processing** — adjacency classification, orientation, and shading analysis convert raw footprints into simulation-ready thermal zones.
3. **Archetype parameterization** — each building receives envelope, ventilation, Heating, Ventilation, and Air Conditioning (HVAC), Domestic Hot Water (DHW), infiltration, internal-loads, and weather parameters from lookup tables tied to national standards.
4. **Simulation and scenario management** — parallelized EnergyPlus execution, parsing into columnar Parquet storage, and physically-constrained scenario generation through the modification engine.
5. **Computational performance** (Section 2.5) — pipeline throughput and preparation-time benchmarks.
6. **Pipeline validation** (Section 3.1) — archetype-level benchmark against Voorbeeldwoningen 2022 (VW2022) reference dwellings and postcode-level comparison against measured Statistics Netherlands (Centraal Bureau voor de Statistiek) (CBS) consumption.
7. **District retrofit analysis** (Section 3.2) — baseline assignment, cost and emission framework, scenario comparison, and path-dependent marginal abatement cost curves for two Dutch districts.

The software separates a reusable workflow core (orchestration, geometry generation, Input Data File (IDF) assembly, simulation, parsing, and the scenario and Marginal Abatement Cost Curve (MACC) engines) from country-specific adapters for registry schemas, archetypes, regulatory mappings, weather, costs, carbon factors, and baseline classifications. Adaptation to a different national context therefore need not redesign the workflow core, but it requires substantial country-specific data engineering, expert translation of national standards, and independent validation; only the Dutch implementation is evaluated here. Extending the system to building technologies absent from the current library (e.g., district cooling or alternative fuel types) likewise requires adding the corresponding parameterization modules, not altering the pipeline itself. The thermal simulation relies on the EnergyPlus conduction transfer function method and detailed fenestration heat balance; the underlying physics models are described in [35].

2.1.1. Data management

File-based data exchange (spreadsheets, shapefiles, and configuration files processed sequentially) becomes error-prone at single-building scale and infeasible at stock scale [3,36]. DELFT-UBEM replaces this with a relational database (PostgreSQL/PostGIS (PostGIS)) that serves two roles: persistent storage of building attributes and geometric data, and spatial preprocessing of raw registry geometries into simulation-ready inputs. Spatial queries perform geometric preprocessing (adjacency, shading, orientation) across the entire building stock in bulk, without per-building processing loops. This preprocessing operates either on a pre-loaded national dataset (as in the present Dutch case study) or on demand for any user-supplied geographic boundary, producing structured per-building records that the IDF generation layer consumes directly via the Python libraries geomeppy (geometry translation) and eppy (IDF manipulation). Simulation configurations, which vary per scenario run, and simulation outputs are written to Apache Parquet files rather than stored in the database, because Parquet's columnar layout matches the read pattern of UBEM post-processing (selective column access across thousands of buildings) with built-in compression and predicate pushdown.[1] Lookup tables that map archetypes to physical parameters are maintained as CSV files for transparency and version control, and loaded into memory at runtime to avoid repeated file parsing during simulation. This database-mediated design is a methodological choice rather than an implementation convenience, and it follows established practice: geodatabases built on spatially enabled relational systems such as PostgreSQL/PostGIS are the standard infrastructure for managing and analyzing semantic 3D city models [37], and city-scale building energy studies have used them as the integrative data backbone since the Energy Atlas Berlin [38]. Beyond the bulk spatial preprocessing described above, three properties matter. First, provenance: common exchange formats such as shapefiles and GeoJSON enforce no schema for building properties, which leads to inconsistency between datasets [3], whereas the database types every attribute, and one stable identifier follows each building from registry through simulation outputs; combined with the versioned lookup tables and per-run logs of every assigned parameter, this responds directly to the finding that about 95% of published UBEM studies are not reproducible because data sources, data models, and workflows go undocumented [29]. Second, scenario management: scenario definitions are configuration records resolved against the same database state, so a five-scenario analysis of a thousand buildings remains a set of queries rather than a directory of manually tracked files. Third, interoperability: tool-specific and closed data formats have been shown to undermine data exchange in urban energy modeling [36], while a shared typed schema lets every pipeline stage, and external tools, read the same records. A disciplined file-based workflow can approximate several of these properties; the database-mediated design makes them structural rather than dependent on user convention. The database schema is detailed in Appendix B.

2.2. Geometric data processing and modeling

With building attributes and spatial data resolved, the pipeline constructs a stock-scale geometric abstraction for each building: registry polygons are retained for topology, orientation, and adjacency analysis, whereas the thermal geometry used in the reported EnergyPlus simulations is an oriented equivalent-rectangle prism derived from footprint area, perimeter, height, floor count, and directional exposure labels. Building geometry (form, orientation, adjacency, and external shading) directly affects thermal loads and solar gains [14]. Existing tools use shoebox abstractions that account for façade orientation [39], apply reduced-order monthly energy balances over City Geography Markup Language (CityGML) geometry [40], generate building geometry from prepared Geographic Information Systems (GIS) footprints within the URBANopt workflow [17], or use CAD-based urban massing in umi [41]. Urban shading typically requires separate tools or manual configuration [14,16]. The approach adopted here performs geometry generation, adjacency classification, and shading analysis within the database layer, so that all geometric operations use the same spatial data without intermediate file export.

2.2.1. Building geometry generation

Building footprints from national registries (e.g., the Dutch Basisregistratie Adressen en Gebouwen (BAG)) or custom datasets undergo three processing stages. PostGIS handles topological corrections, adjacency classification, and spatial queries; the geomeppy library then translates the processed geometries into EnergyPlus thermal zones.

[1] Alternatives were considered: the relational database introduces write contention from parallel simulations; Comma-Separated Values (CSV) lacks schema enforcement and requires full-file scans; and although EnergyPlus outputs per-building SQLite files, SQLite's row-oriented storage is inefficient for cross-building column aggregation.

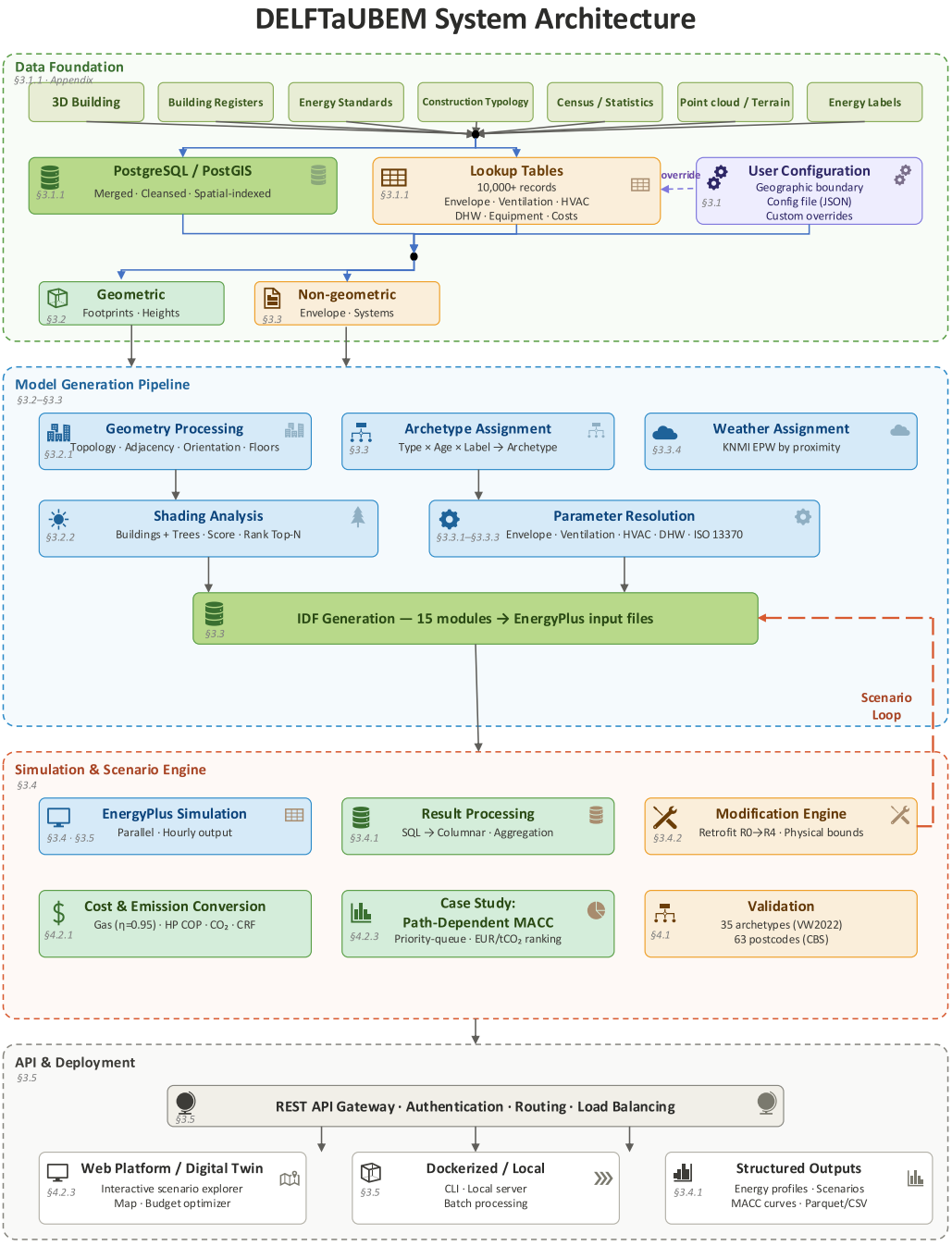


Fig. 1. Process flow of the DELFT-UBEM pipeline.

The process begins with topological corrections on raw footprints using PostGIS spatial functions.[2] Boundary segments then undergo adjacency analysis. Each building polygon is decomposed into boundary segments using ST_DumpSegments. Segments are classified into cardinal directions (N/E/S/W) by computing the relative angle between each segment's azimuth and the building's main orientation axis, with ±45° tolerance bands. For each direction, shared and façade edge lengths are summed separately; a direction is classified as 'Shared' when its total shared length exceeds its façade length, using a 10% overlap threshold against neighboring building geometries. This produces four boundary condition labels per building that map directly to EnergyPlus wall surface boundary conditions: 'Shared' → Adiabatic (no sun/wind exposure), 'Façade' → Outdoors (sun/wind exposed).

Building orientation is determined from the longest edge of the minimal oriented bounding rectangle, normalized to $[0°, 180°)$. This primary azimuth θ_{main} defines the coordinate rotation applied during geometry preparation:

$$x' = x\cos(\theta) - y\sin(\theta), \quad y' = x\sin(\theta) + y\cos(\theta) \tag{1}$$

Floors are stacked on an identical rectangular plan, rotated to the primary azimuth (Eq. (1)) and inheriting the directional boundary classifications; optional perimeter-core zoning is supported, and the present

[2] ST_MakeValid resolves self-intersections, ST_Dump() handles multi-part geometries, and ST_ForceRHR enforces consistent vertex ordering.

study uses single-zone modeling throughout. This prismatic abstraction preserves the surface-to-volume ratio, orientation, and shared-wall fractions that principally shape the stock-level heat balance while avoiding the topology-repair failures of direct footprint extrusion. Vertically varying forms such as setbacks, podium-tower massing, pilotis, and floor-by-floor plan changes are not represented; buildings are prismatic with flat thermal roofs, and pitched-roof attributes inform retrofit costing and PV area rather than zone geometry.

2.2.2. Modeling shading effects in urban building energy simulations

To integrate shading effects from neighboring buildings and vegetation, DELFT-UBEM uses tree crown polygons and heights derived from the Dutch national Actueel Hoogtebestand Nederland (AHN) airborne LiDAR point cloud (Boomregister crown segmentation; crown height taken as the highest LiDAR return per crown).[3] Spatial Structured Query Language (SQL) queries identify, score, and rank candidate obstructions for optional conversion by eppy into EnergyPlus shading objects. The database workflow calculates critical attributes for each candidate-façade pair: geometry, distance, relative height, azimuth angle, and seasonal transmittance for vegetation. Proximity-based spatial functions eliminate distant obstructions while differentiating opacity: buildings receive full opacity whereas trees use seasonal transmittance values reflecting foliage changes. Azimuth calculations filter objects beyond ±90° relative to façade orientation, retaining only forward-facing candidates.

Candidate obstructions are ranked by a geometric score that approximates the obstruction elevation angle as seen from the façade midpoint:

$$S = \begin{cases} \frac{\Delta h}{d} & \text{if } \Delta h > 0 \text{ and } d > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where Δh is the height of the obstruction above the façade and d is the horizontal distance. This ratio is equivalent to $\tan(\alpha)$ for the elevation angle, providing a computationally efficient proxy for detailed ray-tracing. The ranking combines this score with transmittance values and seasonal sun path geometry (derived from the site latitude and the EnergyPlus Weather (EPW) weather file) to prioritize obstructions with the greatest annual shading impact. Obstruction geometry is evaluated on locally flat terrain: the height difference Δh uses relative registry heights, not absolute elevations, and terrain or building-base elevation enters no other solar or insolation calculation. This suits the near-flat Dutch context; in areas with significant relief, ground-elevation differences between neighboring parcels would alter obstruction angles. The data model already carries per-building terrain elevation (3DBAG `b3_h_maaiveld`), so adding a terrain term to Δh is a data-mapping extension rather than a redesign. The ranking supports retaining the top-N obstructions per façade, and the exported obstruction tables retain all forward-facing candidates for downstream selection. The validation and district results reported here include window-level shading (archetype-specific blinds at solar-irradiance setpoints); neighboring-building and tree obstruction shading is computed and stored in the database but was not activated in the reported simulations. When the module is invoked, its output (geometries in local coordinates, transmittance values, and seasonal metadata) is written to the database for optional downstream use by the IDF generation pipeline (Fig. 2).

2.3. Archetype parameterization from national standards

With geometry established, each building requires physical parameters (envelope, ventilation, infiltration, DHW, internal loads, ground temperatures, and window shading) to complete the IDF. All baseline parameters derive from national datasets (Tables B.1 and B.3 in Appendix B): the VW2022 dataset [42] for envelope and ventilation configurations, and NTA 8800 (NTA8800) for ventilation classifications, DHW methods, and infiltration airtightness ($q_{v;10}$ forfait values tabulated per construction-year band and dwelling type, e.g. 1.44 L/s·m^2 for a pre-1945 detached house down to 0.12 for a post-2015 apartment), with Tweedekamer 2020 and TABULA serving as independent reference benchmarks and CBS statistics filling only non-thermal gaps (equipment loads). Archetype keys absent from any lookup halt processing with an explicit error; no silent defaults are applied. The pipeline indexes all parameters by building function, construction period, and renovation scenario. Each parameter is stored as a plausible minimum–maximum range: the archetype benchmark (Section 3.1.1) uses deterministic midpoint assignment, while the postcode validation and district analyzes draw each parameter uniformly within its range under a fixed random seed (seed 42), reflecting within-archetype heterogeneity while remaining exactly reproducible. The demand spread these ranges induce is bounded by the within-archetype variation shown in Fig. 5.

These standards produce parameters for monthly steady-state energy balance calculations (thermal resistances, ventilation classes, regulatory schedules), not for hourly dynamic simulation. The core engineering task is translating each parameter into its EnergyPlus equivalent: R-values into multi-layer constructions, ventilation classes into fan objects with pressure and efficiency, and annual equivalent full-load hours into hourly profiles. The accuracy of these translations is evaluated against national benchmarks and measured consumption. Table B.4 lists each translation, its receiving EnergyPlus object, and its governing assumption. The translations with the largest effect on annual heating demand are the ground-boundary temperature, the infiltration treatment, and the heat-recovery effectiveness; the ground-boundary and infiltration effects are quantified in Section 3.1.1, while heat-recovery effectiveness, which matters most for modern buildings, remains the most uncertain translation. Translating regulatory parameters into simulation-ready inputs remains an evolving process; as measurement data and validated archetypes accumulate, these mappings are progressively refined. Fig. 3 shows the IDF generation workflow.

2.3.1. Envelope and fenestration

Each opaque element is represented as a multi-layer construction assembled from a predefined materials library (Fig. 3, right path). The library contains layer-by-layer definitions for each element type (exterior wall, shared wall, roof, and ground floor) across all energy label scenarios, with each layer specified by thickness, conductivity, density, and specific heat ranges. An uninsulated masonry wall, for instance, resolves to an outer brick leaf, a cavity air layer, and an inner brick leaf; a well-insulated flat roof assembles as EPDM membrane, PIR insulation, vapor barrier, reinforced concrete, and interior plaster. Under retrofit scenarios, insulation layers are appended or thickened to achieve the upgraded R-value while structural layers remain fixed. A lookup-based thermal-bridge surcharge ($\Delta U = 0.012$–$0.30\,\mathrm{W\,m^{-2}\,K^{-1}}$), indexed by envelope element, construction period, and building function, with dwelling-type detail for residential buildings, is added to the U-value of exterior-wall, roof, and ground-floor constructions; the implementation realizes the corrected U-value by reducing the resistance of the highest-resistance layer. Shared walls, interior walls, inter-floor constructions, and windows receive no direct surcharge. The library ships with predefined configurations for all standard scenarios; users can select from these or define custom layer compositions for site-specific conditions.

2.3.2. HVAC and ventilation

Heating and cooling demands are modeled with `ZoneHVAC:IdealLoadsAirSystem`, which isolates envelope and ventilation effects from plant-level details, appropriate for stock-level analysis where the building fabric dominates the energy balance.

Its output represents ideal thermal energy delivered to the zone rather than fuel or electricity consumed by real plant. The abstraction

[3] https://www.ahn.nl/

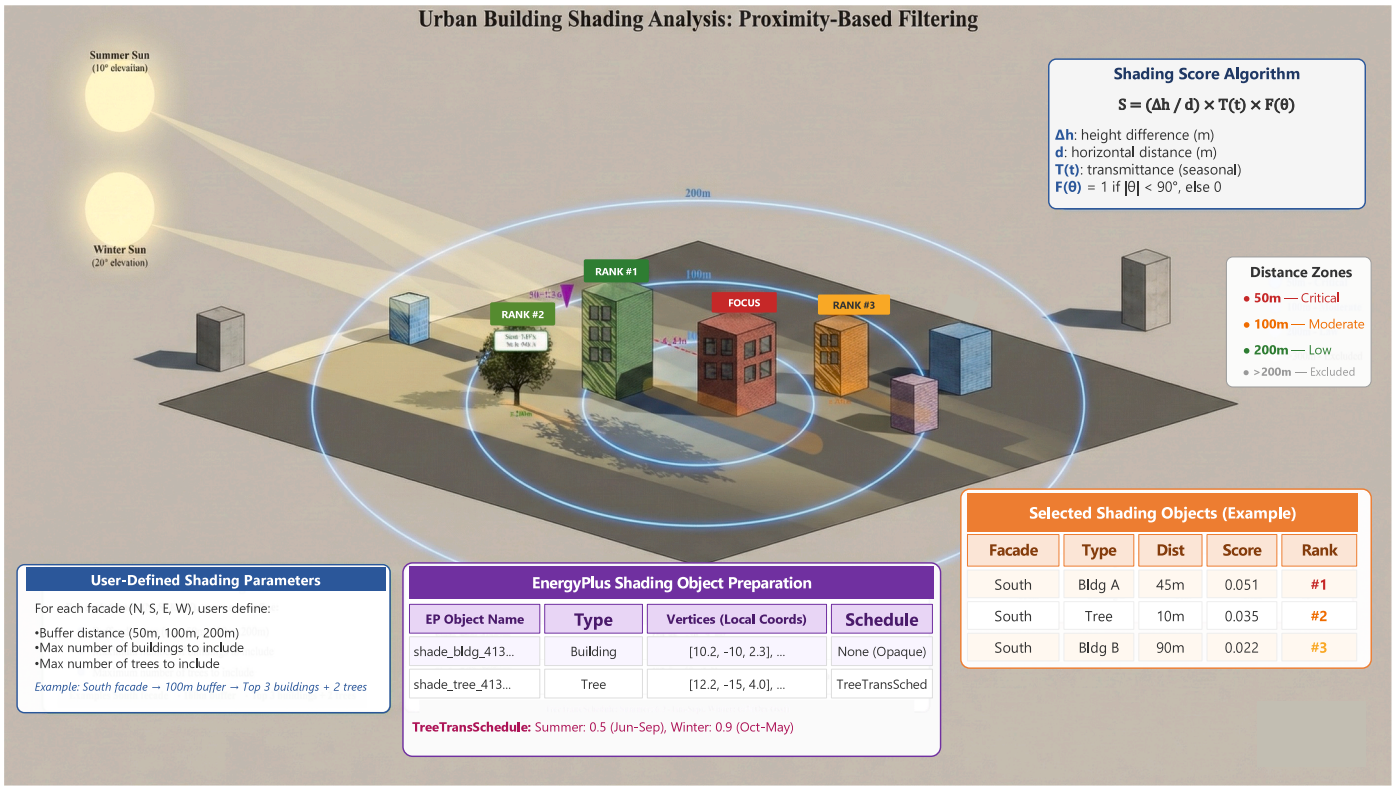


Fig. 2. Database preprocessing workflow for identifying, filtering, scoring, and ranking candidate external obstructions. The module assumes locally level terrain and was not activated in the EnergyPlus simulations reported in this study.

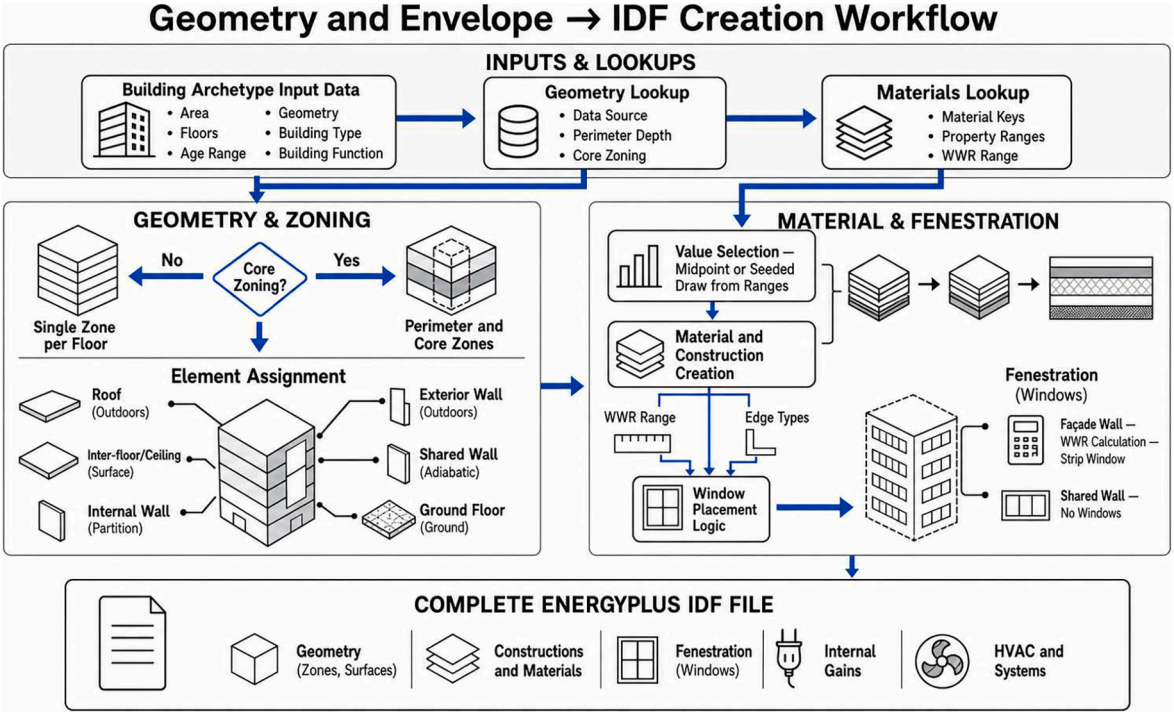


Fig. 3. IDF creation workflow: Geometry/zoning path (left) handles zone configuration and element assignment; material/fenestration path (right) maps thermal properties and window placement. Outputs merge into simulation-ready EnergyPlus files.

omits part-load efficiency, distribution and emission losses, control dynamics, defrost, auxiliary (fan and pump) electricity, and cycling; plant efficiency enters only in post-processing, where ideal demand is converted to fuel. Ventilation is treated analogously: fan electricity of mechanical systems (B–D) is not metered, and heat recovery is an ideal sensible effectiveness rather than a controlled unit with bypass.

Four ventilation configurations reflect the range of Dutch building practice: natural ventilation through openable windows and trickle vents (System A), mechanical supply with natural exhaust (System B), mechanical exhaust with passive air inlets (System C), and balanced mechanical ventilation with heat recovery (System D). Renovation scenarios progressively upgrade ventilation systems along this sequence, with heat recovery sensible effectiveness ranging from zero (no recovery) to values consistent with counterflow exchangers.

For naturally ventilated buildings (System A), infiltration and ventilation are physically the same airflow. To prevent double-counting (since EnergyPlus treats `ZoneInfiltration` and `ZoneVentilation` as additive), the pipeline compares the infiltration air change rate against the ventilation requirement. When infiltration alone satisfies the requirement, no `ZoneVentilation:DesignFlowRate` object is created; when it falls short, a supplementary ventilation flow bridges the deficit, capped at 50% of the infiltration rate.

For mechanically ventilated buildings (Systems B, C, and D), the fan pressure field partially suppresses envelope leakage. The reduced infiltration rate follows the NTA8800 quadrature interaction:

$$\dot{q}_{inf,red} = \sqrt{\dot{q}_{inf}^2 + \dot{q}_{mech}^2} - \dot{q}_{mech} \tag{3}$$

which yields an effective infiltration rate below the unassisted value. The maximum permitted reduction increases with unassisted leakage from 10% to 28%; the calibration status of the upper limit is described in Section 3.1.1. The airtightness value $q_{v;10}$ is applied directly as the EnergyPlus design infiltration rate (Flow/Area), with weather coefficients normalized to unity at a reference condition of a 10 K indoor-to-outdoor temperature difference and 3 m/s wind speed; this effective-infiltration convention (rather than an explicit conversion from the 10 Pa test pressure to operating pressures) is adopted deliberately. The sensitivity of heating demand to the infiltration level is quantified. Heating and cooling demand is calculated with `ZoneHVAC:IdealLoadsAirSystem` for all four ventilation configurations; ventilation is represented separately. Systems A–C represent scheduled ventilation through `ZoneVentilation:DesignFlowRate`; System A adds only a capped supplementary flow where infiltration falls short of the requirement. System D links `DesignSpecification:OutdoorAir` to the IdealLoads unit and uses its native sensible heat-recovery fields. Night cooling uses separate `ZoneVentilation:DesignFlowRate` objects and is not part of the A–D classification. Latent heat-recovery effectiveness is zero.

2.3.3. Ground temperature correction

Per-building equivalent ground-boundary temperatures are computed following ISO 13370, using the characteristic dimension:

$$B' = \frac{A_{floor}}{0.5 \times P_{exposed}} \tag{4}$$

where A_{floor} is the floor area and $P_{exposed}$ is the exposed perimeter derived from the adjacency classification. This ground-boundary treatment substantially affects calculated heating demand, particularly for older buildings with uninsulated ground floors.

2.3.4. Internal loads, DHW, shading, and weather

Internal load schedules must translate annual energy totals from standards into hourly profiles. For lighting and equipment, the pipeline combines three inputs: a peak power density (W/m^2), a reference 24-hour usage shape (hourly fractions for weekdays and weekends), and an annual operating target expressed as equivalent full-load hours. A scale factor adjusts the usage shape so that the integrated annual hours match the target:

$$f_{scale} = \frac{\mathrm{EFLH}_{target}}{52 \times (5\,\bar{h}_{wd} + 2\,\bar{h}_{we}) \times 24} \tag{5}$$

where $\bar{h}_{wd}$ and $\bar{h}_{we}$ are the mean hourly fractions for weekdays and weekends respectively, and individual hourly values are capped at 1.0 after scaling. Monthly multipliers further modulate the profiles to capture seasonal variation (reduced lighting in summer, holiday closures for institutional buildings), producing a resolved 8,760-hour `Schedule:Compact` object per zone.

DHW demand is computed from occupant count and per-capita consumption for residential buildings, and from area-based intensities for non-residential functions. The daily volume is distributed across morning and evening peaks using integral-normalized schedules that ensure the simulated draw matches the calculated demand exactly.

Occupant metabolic heat gains are not modeled in the baseline configuration; lighting, equipment, thermostat, and occupancy schedules are deterministic. This omission slightly over-estimates heating demand, with a proportionally larger effect in well-insulated dwellings, where internal gains form a larger share of the heat balance. Window shading activates at archetype-specific solar irradiance setpoints. Weather data uses KNMI reference station EPW files assigned by geographic proximity.

2.4. Simulation management and scenario generation

Once simulations complete, the pipeline must connect parameter changes to their energy outcomes across thousands of buildings and multiple scenario variants. In manual workflows, this connection is maintained through ad-hoc file naming and spreadsheets, limiting both reproducibility and the number of scenarios a study can consider [43]. The pipeline treats simulation, result parsing, and scenario modification as stages of a single automated cycle: EnergyPlus simulations execute in parallel, each building's results are parsed into a unified columnar representation, and scenario modifications operate on the parsed parameter state rather than on raw files.

2.4.1. Result parsing and scenario definition

Each EnergyPlus simulation produces an SQLite database of time-series results and the IDF that generated them. At stock scale, this yields thousands of isolated files with no cross-building structure. The parsing stage reads these SQLite outputs and unifies them into Parquet files indexed by building identifier and scenario variant, so that comparing a retrofit variant against its baseline is a dataframe join rather than a file-matching exercise. A variable-aware aggregation layer converts sub-hourly output to hourly, daily, and monthly resolution using physically appropriate operators: summing extensive quantities (energy, mass flow) and averaging intensive quantities (temperature, humidity).

A scenario specifies parameter transformations across a building's physical domains (envelope, ventilation, systems, schedules), where each parameter influences demand through a different pathway: wall insulation reduces transmission losses; window replacement changes both U-value and Solar Heat Gain Coefficient (SHGC) with opposing effects on heating and cooling. Scenarios can target the full stock uniformly or specific subsets through a hierarchical assignment logic that matches buildings by identifier, energy label, or archetype pattern, with unmatched buildings retaining their baseline configuration.

2.4.2. Modification engine

Unconstrained parameter modification produces physically inconsistent models: ventilation rates below code minimums, insulation thicknesses beyond constructible limits, or window properties that violate optical relationships. The modification engine embeds domain constraints so that every generated scenario remains physically feasible.

The engine operates through category-specific handlers for envelope, HVAC, ventilation, infiltration, lighting, and DHW. Each handler applies predefined retrofit packages that bundle physically coherent upgrades (R0–R4 levels) derived from national envelope and ventilation standards. All transformations are clamped to material-specific bounds inferred from material type (insulation thickness is constrained to 0.02–0.40 m, vapor barriers to 0.1–5 mm), preventing configurations that EnergyPlus would reject. Coupled constraints are enforced after individual modifications: thermal resistance must remain consistent with the thickness–conductivity pair ($R = t/\lambda$), and visible transmittance is clamped to satisfy $\tau_{vis} \leq$ SHGC.

Operating mode determines which parameters are modifiable. In retrofit mode, only physically realizable interventions are permitted (insulation thickness, glazing replacement, ventilation system upgrade), while intrinsic material properties (density, specific heat, absorptance) are locked. This prevents scenarios that achieve energy savings through non-physical parameter changes rather than actual building interventions. The retrofit packages are defined as sequential levels, where each level inherits and extends the previous level's envelope and ventilation upgrades. The specific Dutch-code levels applied in this study are defined in Section 3.2.1.

2.5. Computational performance

Pipeline throughput determines whether the framework can scale to stock- and city-level analyzes. Two metrics characterize its performance: end-to-end preparation time (from geographic boundary to simulation-ready models) and parallel simulation throughput. Human preparation effort and computational wall time are treated as separate quantities. For context, the UBEM.io Evanston case study [25] reports approximately 214 person-hours of expert effort for conventional workflows, reduced to approximately 38 person-hours with UBEM.io; the large majority of those hours fall on data preparation and archetype assignment, the stages DELFT-UBEM automates entirely. The equivalent preparation here requires a configuration file and a geographic boundary from the user, with the machine-side steps (ingestion, geometry, parameterization, scenario definition) completing in under 15 minutes for a district of comparable size (1,363 buildings in the Evanston pilot against 1,282 in Holten-Kom; Fig. 4). Both rest on comparable upstream assumptions: [25] assume the municipality already supplies cleaned building footprints and height data, and the DELFT-UBEM figure likewise assumes a prepared national implementation. Direct comparison is limited by differences in stock, scope, and national context, but the order-of-magnitude reduction consistently reflects the elimination of manual data preparation, which dominates conventional UBEM project timelines rather than simulation itself.

Computational throughput was measured at two scales on a workstation-class machine (Apple M4 Max, 14 cores, 36 GB RAM). A worker-scaling benchmark on a single postcode (59 buildings, one scenario) isolates parallelization behavior: at 8 workers the pipeline (IDF generation, EnergyPlus simulation, result parsing, and aggregation) sustains 0.77 buildings/s, with simulation scaling near-linearly and result parsing the largest per-building cost (Table 1). At district scale, the archived Holten-Kom production runs (1,282 buildings per scenario, 10 parallel workers) record wall times of 5,870 to 5,988 s per scenario (mean 98.8 min, or 4.63 s per building-scenario), with result parsing and aggregation to Parquet, rather than EnergyPlus simulation, the principal cost at that scale. The three archived envelope-upgrade scenarios (R1 to R3) total approximately 4.9 h on the same workstation; heat-pump electrification is evaluated in post-processing from hourly loads and requires no additional simulation. Parquet compression reduces the 17 MB of raw EnergyPlus output per building to 0.7 MB, keeping storage manageable at stock scale. Model preparation itself, from geographic boundary to simulation-ready IDF models, completes in minutes for the two case study districts combined (1,960 buildings); end-to-end wall time is dominated by EnergyPlus execution and result parsing, which scale with building count as reported above; processing time for larger stocks (city-scale, >10,000 buildings) is discussed in [35].

Table 1
Worker-scaling benchmark on a single postcode (59 buildings, one scenario) by worker count. Apple M4 Max, 14 cores, 36 GB RAM.

	2 workers	4 workers	8 workers	12 workers
Wall time (s)	120	92	76	74
Rate (bld/s)	0.49	0.64	0.77	0.80
Speedup	1.00×	1.30×	1.57×	1.62×
IDF + simulation (s/bld)	1.28	0.82	0.56	0.51
SQL/IDF parsing (s/bld)	0.63	0.63	0.63	0.64
Aggregation (s/bld)	0.05	0.05	0.05	0.05
Raw output (MB/bld)	17.0	17.0	17.0	17.0
Parsed Parquet (MB/bld)	0.7	0.7	0.7	0.7

3. Results

All results reported below include archetype-specific window-shading controls. The external-obstruction step described in Section 2.2 was not activated in the reported simulations, so shading from neighboring buildings and trees does not affect any result in this section.

3.1. Pipeline validation

The pipeline is validated at two complementary scales: individual archetypes against national heating benchmarks, and aggregated postcodes against measured consumption. Both validations target annual energy, the resolution at which measured consumption is publicly available for the Dutch stock. This validation scale reflects data availability, not the model's temporal resolution: the pipeline resolves every building hourly, and hourly profiles drive the heat-pump COP conversion and capacity sizing of the retrofit scenarios. Extending to operational studies (peak-shaving, demand-response) is a matter of validating these hourly profiles against metered interval data, not of model capability. The framework is positioned as an annual, stock-scale estimation and scenario-comparison tool.

3.1.1. Heating demand benchmark

The first test verifies simulated annual heating demand against the VW2022 reference dwelling dataset [42], a government-published collection of standardized dwelling configurations with calculated energy performance per NTA 8800, which provides benchmark values for 35 residential archetypes (5 building types × 7 construction periods). No building-specific or postcode-specific calibration was performed. The NTA8800 $q_{v;10}$ forfait values are standards-derived inputs and were not fitted. The leakage-dependent infiltration-suppression bound permits a reduction of 10–28%; its 28% upper limit was calibrated once against the VW2022 reference dwellings and then held fixed in all subsequent simulations. The VW2022 comparison is therefore not an independent test of that limit; the postcode consumption data were not used to set it.

Because heating demand depends on building geometry as well as envelope properties (floor area, height, orientation, and adjacency conditions vary across the real stock), each of the 35 archetypes was simulated in four geometric configurations spanning the typical range of Dutch housing, varying floor area, story count, cardinal orientation, and shared-wall adjacency, giving 140 residential simulations. The four configurations were averaged within each archetype, and all reported CV-RMSE and NMBE values, baseline and sensitivity alike, were calculated across the resulting 35 archetype means.

Across all 35 archetypes, the model achieves a CV-RMSE of 17.3% and NMBE of - 9.6% on archetype-averaged values, with 27 of 35 archetypes falling within the ±15% band (Fig. 5). The small transparent markers show individual geometric variants; their spread illustrates how strongly

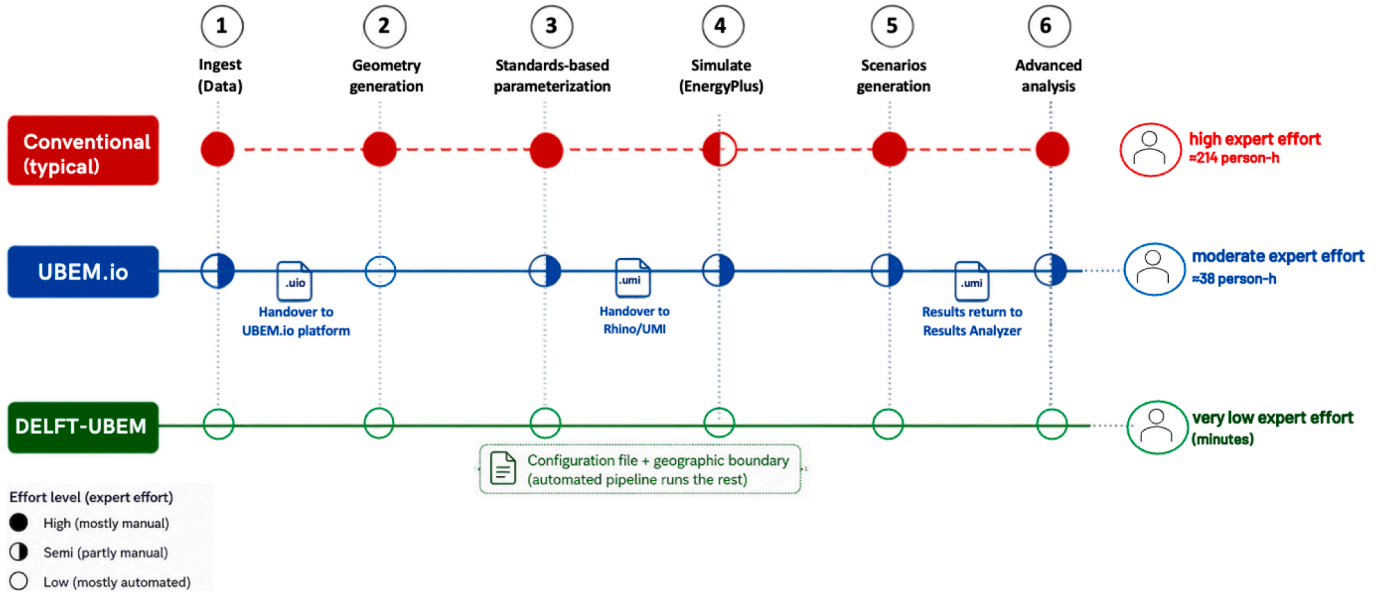


Fig. 4. Expert effort and workflow continuity across three approaches. Circles indicate the level of expert involvement at each stage, and labeled markers show where a manual handover or file exchange interrupts the workflow. Conventional and UBEM.io effort estimates are those reported by [25] for the Evanston case study (1,363 buildings).

floor area and story count affect heating intensity for the same envelope specification.

The negative NMBE indicates systematic under-prediction, despite the omission of occupant metabolic heat gains (Section 2), which would push predictions in the opposite direction. This is consistent with geometric representation (corner exposure, story count, shared-wall assumptions), rather than envelope parameterization, being the leading contributor to the remaining error; the sources of this accuracy are examined below. Corner houses show the largest deviations (-19% to -29% for pre-1991), because the VW2022 reference assumes fully exposed facades while the model's bounding-box geometry produces a lower surface-to-volume ratio for this typology. Apartments exhibit similar under-prediction combined with the highest within-archetype variance (standard deviations up to 57 kWh/m^2), reflecting sensitivity to shared-wall assumptions and floor count. In contrast, modern detached and semi-detached houses houses (post-2006) over-predict by 12–15%: for well-insulated buildings, internal gains constitute a larger fraction of the heat balance, and the model's conservative assumptions (following NTA8800, with no occupant objects) inflate net heating demand.

The CV-RMSE of 17.3% exceeds the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Guideline 14 [44] threshold of 15%, which targets calibrated single-building models tuned against measured data. Comparable UBEM studies without building-specific calibration report similar or larger deviations [45,46], and the model's CV-RMSE (17.3%) is below the 33–37% space-heating-demand reduction evaluated in the district scenarios (Section 3.2). Critically, the expected rank ordering (pre-1945 highest demand, post-2015 lowest) is preserved across all five building types, confirming that the parameterization captures the structural variation in the Dutch stock even where individual archetype deviations are large.

Two kinds of choice govern this accuracy: how the pipeline resolves each building's inputs, and how it translates steady-state standards into dynamic inputs. On the input side, the strongest measured driver is the registered energy label, which the pipeline maps to per-building construction; postcodes with more registered A–B labels are significantly more accurate (49% good labels among passing versus 35% among failing postcodes, $p = 0.020$; label share correlates with gas deviation at $r = -0.41$, $p < 0.001$; Section 3.1.2). Registry-derived story count and shared-wall adjacency form the second input driver, consistent with the geometry-dominated deviations of corner houses and apartments noted above. On the translation side, the assumption with the largest measured effect is the ISO 13370 ground-boundary temperature: without it, floor heat loss through undisturbed-soil boundary conditions inflates total heating demand by 20–35% for older archetypes with uninsulated ground floors, shifting the model from over-prediction to the slight under-prediction observed in Fig. 5. The sensible heat-recovery effectiveness, which matters most for modern buildings, remains the most uncertain translation; the sensitivity to the infiltration level is quantified below. Accuracy is therefore bounded by registry completeness: 36–85% of buildings per postcode lack a registered label and fall back to default or imputed labels, and the registry carries no data on the actual heating system, renovation history, or occupancy that drive most of the residual deviation. The infiltration sensitivity test multiplied the final Flow per Zone Floor Area input written to `ZoneInfiltration:DesignFlowRate` by 0, 0.5, and 1.5, with 1.0 denoting the baseline; the multiplier is applied after the ventilation–infiltration interaction has been evaluated, so it perturbs the EnergyPlus infiltration input without regenerating the ventilation objects. At each setting all 140 residential cases were simulated, the four configurations averaged within each archetype, and the metrics calculated across the same 35 archetype means as the baseline. Removing infiltration raises CV-RMSE from 17.3% to 52% (NMBE from -9.6% to -47%); across the ±50% sensitivity range the model moves from under-prediction (CV-RMSE 34%, NMBE -28% at half infiltration) to over-prediction (16%, + 9% at 1.5 times), with the bias changing sign between the baseline and the 1.5 setting. The factors 0.5 and 1.5 bracket the standards-derived value symmetrically by ±50%, a range chosen to represent the airtightness spread among dwellings that share one forfait class. The zero setting removes infiltration entirely, to show how much of the modeled heat loss it carries. The four settings do not establish a continuous optimum, but they show that the reported accuracy depends materially on the infiltration input, which we note as a limitation.

3.1.2. *Postcode-level validation against measured consumption*

The archetype benchmark (Section 3.1.1) validates individual building types against calculated reference values. Building-level measured

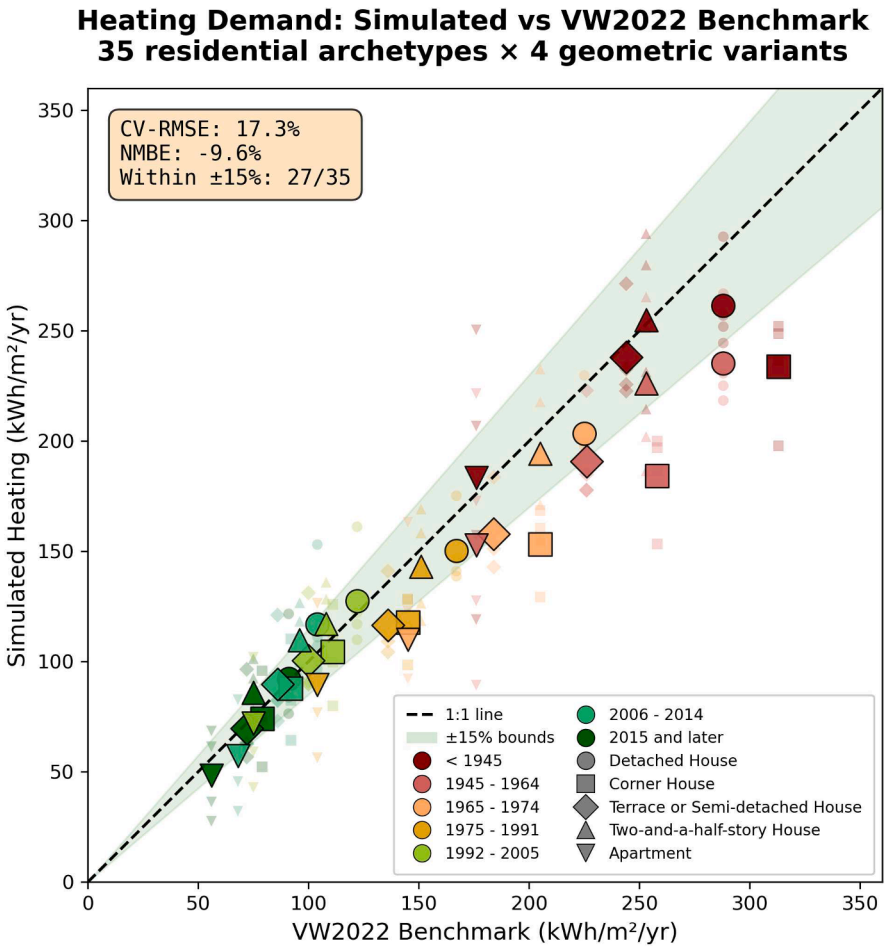


Fig. 5. Simulated vs. VW2022 benchmark heating demand for 35 residential archetypes, each simulated in 4 geometric variants (small markers) and averaged (large markers). Marker shape indicates building type; color indicates construction period. Green band: ±15% of benchmark.

energy data are not publicly available in the Netherlands; the finest resolution is Statistics Netherlands (CBS)[4] or grid-operator postcode-level consumption statistics [47]. To test whether the model reproduces this measured consumption for real building stocks, 63 six-digit postcodes (approximately 3,000 buildings) were selected from the national database to span diverse distributions of energy labels, construction periods, building types, residential/non-residential composition, and grid operator coverage (Liander and Enexis), so that the validation reflects the heterogeneity of the Dutch building stock rather than a single regional context.

Simulated gas consumption per dwelling was compared against CBS statistics for 2022, using the fuel conversion described in Section 3.2.1. All buildings were simulated with one common typical-year weather file and gas-fired heating; no building-specific or climate-zone adjustment was applied, so spatial climatic variation is not represented.

Aggregate Results. The model over-predicts gas consumption by 12.6% on average and under-predicts electricity by 9.4% (Table 2). At a threshold of ±30% (adopted in the absence of established benchmarks for postcode-level stock models), 65% of postcodes pass for gas and 81% for electricity (Fig. 6).

The gap between mean absolute and signed deviation reflects opposing biases: the model assigns gas heating universally, while a growing share of post-2015 dwellings use heat pumps, shifting consumption between fuel types. The share of post-2015 buildings is significantly correlated with gas deviation ($r = -0.34$, $p = 0.007$), as is the share of buildings with registered labels A or B ($r = -0.41$, $p < 0.001$).

Table 2
Deviation statistics between simulated and CBS-reported energy consumption across 63 postcodes.

	Gas	Electricity
Mean deviation	+12.6%	−9.4%
Median deviation	+12.4%	−10.5%
Mean absolute dev.	29.8%	19.0%
Pass rate (±30%)	65%	81%
Range	−72 to +104%	−56 to +68%

Deviation Sources. Three systematic factors explain most of the deviation between simulated and measured consumption. The largest outliers coincide with district heating or alternative energy networks: postcodes such as Heerlen 6433JR (+74%) show extreme over-prediction because the model assigns gas heating to buildings that are actually served by geothermal or district heat supply. Energy label coverage is a second major driver. Between 36% and 85% of buildings per postcode lack a registered energy label and received a default label B assignment in this postcode-validation baseline (the district analysis in Section 3.2 instead imputes missing labels from construction year); where the true distribution skews toward lower labels, the model under-predicts heating de-

[4] https://opendata.cbs.nl/statline/

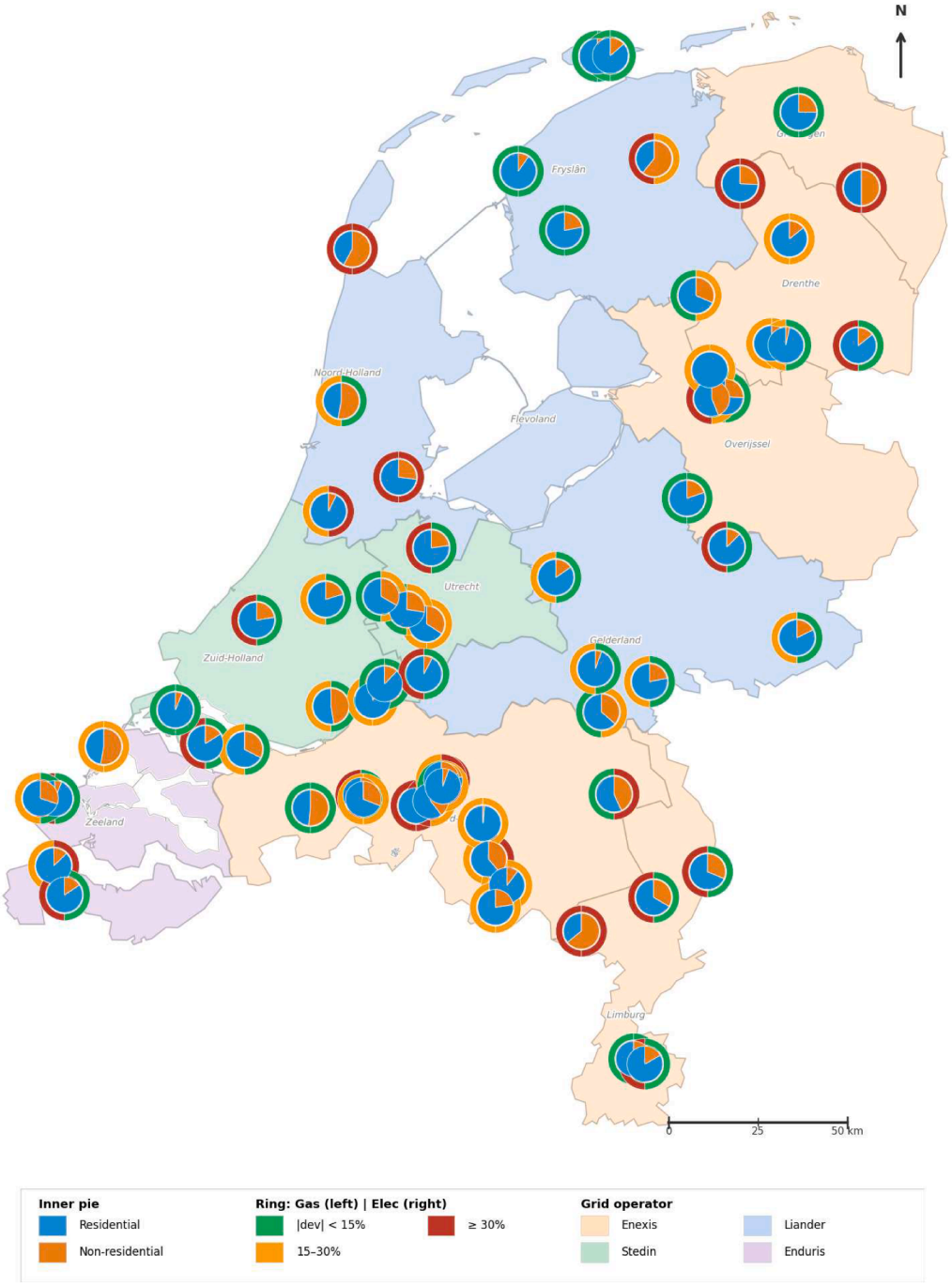


Fig. 6. Spatial distribution of 63 validation postcodes across the Netherlands. Inner pie shows residential (blue) vs. non-residential (orange) building share; ring color indicates deviation from CBS measured consumption for gas (left half) and electricity (right half): green <15%, orange 15–30%, red >30%. Background shading indicates grid operator territory. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

mand. This effect is statistically significant: postcodes passing the ±30% threshold have 49% good labels (A–B) on average, versus 35% for failing postcodes ($p = 0.020$, Mann-Whitney U test). Third, non-residential building diversity amplifies deviation magnitude without predicting its direction: postcodes with >50% non-residential buildings show nearly double the mean absolute deviation of predominantly residential postcodes (45% vs. 25%), reflecting the limited non-residential archetype coverage.

The remaining variance is attributable to factors absent from the building registry: renovation history, actual heating systems, occupant behavior, and district heating connections. For postcodes dominated by gas-heated residential stock of mixed age (the majority of the Dutch building stock), the model reproduces measured consumption with a median deviation of 12–13% (Table 2) without building-specific adjustment.

3.2. District retrofit analysis

The validated pipeline is applied to the Dutch building stock using BAG building records [48] and the archetype parameterization described in Section 2.3. Residential typologies account for approximately 85% of the stock (Fig. 7), justifying the stronger data coverage for residential archetypes. Within the residential stock, terraced and semi-

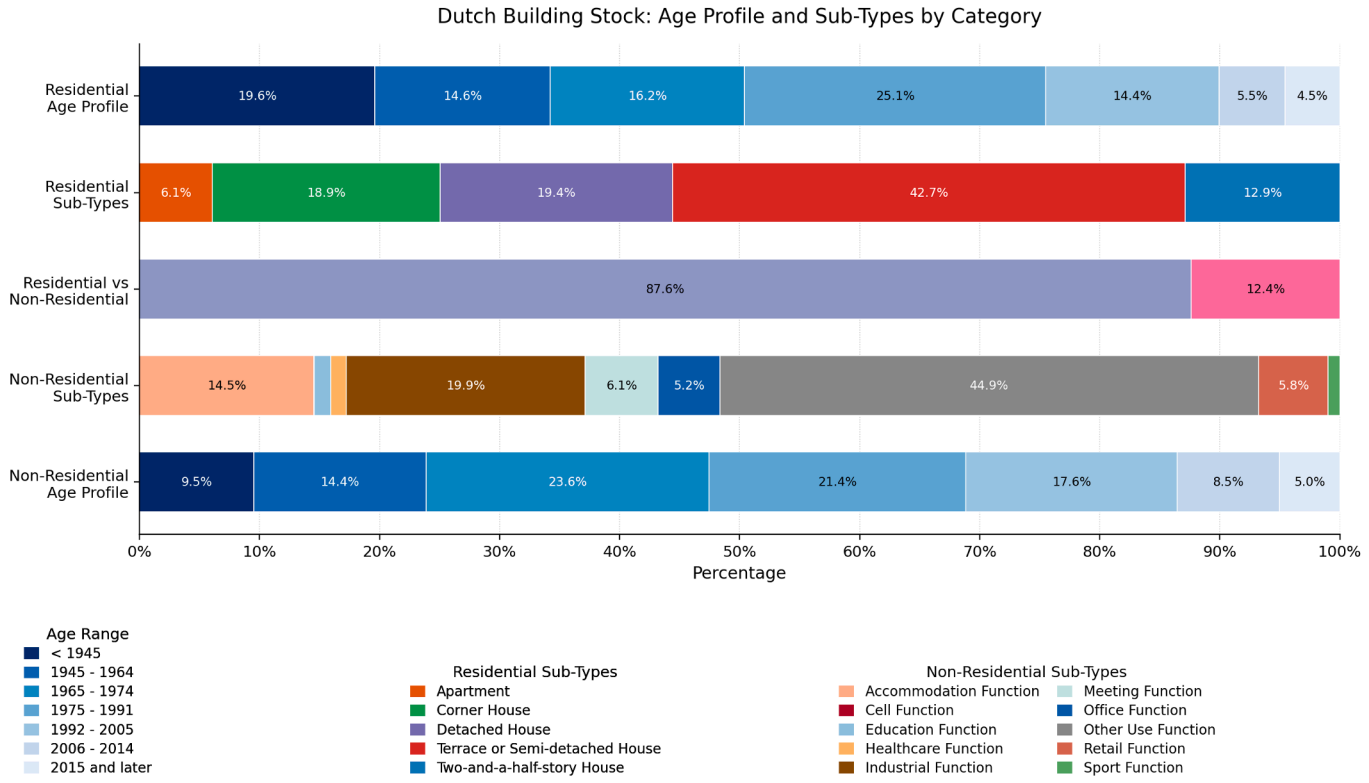


Fig. 7. Percentage distribution of Dutch building stock by age profile and subtypes for residential/non-residential categories. Data source: 2023 BAG dataset.

detached houses form the largest subtype, and pre-1975 construction represents the majority; these older, poorly insulated buildings carry the highest heating intensities and constitute the primary target for retrofit intervention.

Having established aggregate accuracy (median gas deviation 12%, with larger errors traceable to heat pump and district heating substitution), we now apply the pipeline to a policy-relevant downstream analysis: how should two districts prioritize building-level retrofit investments?

To evaluate how building-level simulation informs municipal energy transition planning (*wijkaanpak*), two contrasting Dutch districts were analyzed: *Chassé* (Breda), a dense urban district combining pre-war terraced streets (Oud-Boeimeer) with a modern redevelopment on a former military site (Chassépark, 2001–2007), totalling 678 buildings (535 residential), and *Holten-Kom* (Rijssen-Holten), a small eastern town with 1,282 buildings (1,184 residential) spanning the full Dutch construction period range. Both districts were processed with identical parameterization; only the geographic boundary differs.

3.2.1. Setup and evaluation framework

Baseline Assignment and Retrofit Packages. The retrofit packages (Table 3) follow the VW2022 envelope specifications and are mapped to energy labels by the authors to link building registry data to intervention depth: labels G/F start at R0 (original envelope), E/D at R1, C at R2, and B together with all A grades at R3; R4 denotes the electrified package (the R3 envelope combined with a heat pump) and is therefore an upgrade target rather than a baseline state. Buildings lacking a registered label (42–46%) are assigned a label from construction year, 2015 or later to A, 1994–2014 to B, and earlier to C, exploiting the strong correlation between construction period and registered performance in the Dutch stock. Because most unlabeled buildings predate 1994, this places the majority at an R2 baseline and the remainder at R3, retaining an envelope-upgrade step for the older unlabeled stock (Section 4.5).

Each level is a prerequisite for the next, and the MACC enforces this sequencing per building. Target Rc values follow VW2022 prescriptions for each energy label class. Ranges for R0/R1 span pre-1945 through 1992–2005 archetypes; post-2006 buildings whose baseline already exceeds R2–R4 targets retain their original values. R2 and R3 share the same deep-insulation envelope — the upgrade from R2 to R3 is ventilation only — and R4 adds heat pump substitution to the R3 envelope.

This assignment places approximately 77% of buildings in both districts at R2 or R3 baseline levels (Fig. 8). Under this baseline assignment, electrification is the largest remaining modelled transition step for most buildings.

Cost and Emission Framework. Retrofit costs are derived from the RVO Kostenkentallen 2025 [49], the Dutch government's official renovation cost benchmarks. Envelope measures are costed per square meter of affected surface; ventilation and air sealing per dwelling. Investment costs are computed incrementally: a building currently at R1 pays only the R1→R2 cost increment, not the full R2 package.

Thermal demand is converted to gas consumption using a condensing boiler efficiency $\eta_b = 0.95$ and the standard gas calorific value of 9.769 kWh/Nm^3. Gas volumes represent theoretical maxima before the prebound effect: in comparable European housing stocks, measured consumption averages about 30% below the calculated rating, and the gap widens in the poorest-rated dwellings [50]. Prebound adaptation lowers both baseline consumption and achievable savings, so absolute EUR/tCO_2e values are lower bounds. Because prebound is concentrated in poorly insulated dwellings [50], it reduces envelope savings more than electrification savings, so occupant adaptation is more likely to shift the ranking toward electrification than to reverse it.

For the reported analysis, heat-pump electricity is computed from hourly heating demand using an assumed fixed seasonal COP of 3.5. The temperature-dependent form below, in which COP scales with the difference between supply and outdoor temperature, is retained as a consistency check and yields seasonal values of the same order:

$$\mathrm{COP}(T_{\mathrm{out}}) = \eta_C \frac{T_{\mathrm{supply}}}{\max(T_{\mathrm{supply}} - T_{\mathrm{out}},\ 5)} \tag{6}$$

Table 3
Retrofit package definitions with target envelope and system parameters. Rc in m^2K/W, U in W/m^2K. The Labels column gives the energy label each package corresponds to. Baseline assignment places buildings at R0 to R3 only: R4 requires a heat pump and is therefore an upgrade target rather than a baseline state.

Level	Wall Rc	Roof Rc	Floor Rc	Win. U	Ventilation	Heating	Labels
R0	0.19–2.5	0.35–2.5	0.15–2.5	5.1–2.9	A (natural)	Gas boiler	G, F
R1	0.23–2.5	0.38–2.5	0.18–2.5	3.1–2.8	B (nat. supply + mech. exh.)	HR107 boiler	E, D
R2	1.7	3.5	3.5	1.4	C (mech. exhaust)	HR107 boiler	C
R3	1.7	3.5	3.5	1.4	D (balanced + HRV)	HR107 boiler	B
R4	1.7	3.5	3.5	1.4	D (balanced + HRV)	Heat pump	A and above

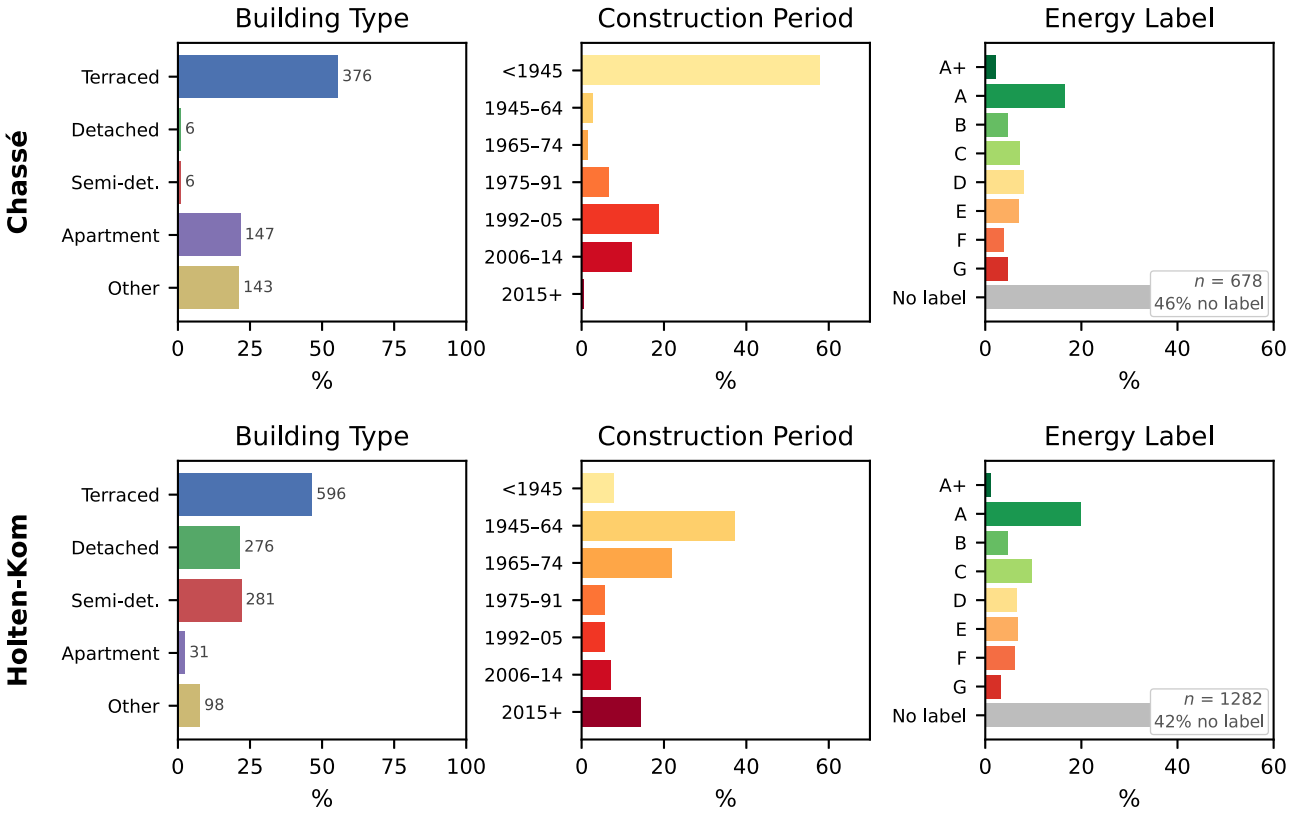


Fig. 8. Building stock characterization: type distribution and construction period (left), registered energy labels and assigned baseline R-levels (right) for Chassé and Holten-Kom. NaN = no registered label (imputed from construction year; predominantly R2).

with $T_{supply} = 308.15$ K (35 °C), $\eta_C = 0.40$ (an assumed Carnot quality grade), bounded between 1.5 and 6.0. Heat pump capacity is sized from the 98th percentile of each building's non-zero hourly heating load at the target envelope level, multiplied by a 1.10 safety factor and rounded up to the next standard capacity class.

Annual space-heating-related greenhouse-gas emissions are calculated on a common operational boundary that excludes upstream fuel-chain and infrastructure emissions for both carriers: 1.779 kg CO_2e/Nm^3 for natural-gas combustion and 0.270 kg CO_2e/kWh for the electricity-production component of the 2024 Dutch grid mix [51]. For heat-pump steps, net abatement is the gas-related CO_2e eliminated minus the generation-related CO_2e of heat-pump electricity.

Investment costs are annualized using the capital recovery factor:

$$\mathrm{CRF} = \frac{r(1+r)^n}{(1+r)^n - 1} \tag{7}$$

at a real discount rate $r = 3\%$ following EU Delegated Regulation 244/2012 [52], the cost-optimal framework in force when the analysis was carried out, with $n = 25$ years for envelope measures (CRF = 0.057) and $n = 15$ years for heat pumps (CRF = 0.084). Component lifetimes are conservative relative to EN 15459-1 reference values [53]. The headline metric is annualized capital cost per annual ton of CO_2e avoided (EUR/tCO_2e). It deliberately excludes operating energy expenditure, maintenance, and purchase subsidies (e.g. Investeringssubsidie Duurzame Energie en Energiebesparing (ISDE)), ranking measures by investment intensity from a public-planning perspective rather than by net household cost; the subsidized variant is examined in Section 3.2.4.

3.2.2. *District scenario outcomes*

Three scenarios were evaluated for each district: the heterogeneous baseline (each building at its label-assigned level), All-R3 (every building brought to R3: those below R3 receive the staged upgrades required, those already at R3 are unchanged), and R4 (deep envelope plus heat pump electrification). Table 4 summarizes the outcomes. The All-R3 transition reduces annual space-heating demand by 33–37%. Heat pump electrification eliminates gas consumption entirely (assuming full electrification without hybrid boiler retention), reducing CO_2e emissions by approximately 73–75%; residual emissions arise from grid electricity consumed by heat pumps. Total investment is €19.4M for Chassé and €31.4M for Holten-Kom, of which 79–80% corresponds to the staged envelope and ventilation measures through R3 and 20–21% heat pump installation (approximately €36,000 and €27,000 per residential dwelling respectively).

Table 4
District stock summary and scenario comparison.

Metric	Chassé	Holten-Kom
Stock characteristics		
Buildings (residential / non-res.)	535 / 143	1,184 / 98
Total floor area (m^2)	380,952	209,542
Baseline R0 / R1 / R2 / R3	58 / 101 / 249 / 270	120 / 172 / 574 / 416
Mean baseline heating (kWh/m^2)	76	102
Scenario outcomes		
Baseline heating (GWh/yr)	12.3	15.7
All-R3 heating (GWh/yr)	7.7	10.6
Baseline gas (kNm^3/yr)	1,330	1,694
R4 gas (kNm^3/yr)	0	0
R4 electricity (GWh/yr)	2.2	3.0
Baseline CO_2e (tCO_2e/yr)	2,365	3,014
R4 CO_2e (tCO_2e/yr)	597	816
Total investment (€M)	19.4	31.4

Fig. 9 decomposes these outcomes. The stock composition (a) shows that both districts are dominated by post-1945 construction at R2–R3 baseline levels, confirming the pattern in Fig. 8. The energy pathway (b) shows that the All-R3 transition reduces gas consumption by 4.8–5.4 GWh/yr, while R4 electrification replaces the remaining gas with 2.2–3.0 GWh/yr of electricity. The median abatement cost (d) varies by upgrade type: heat-pump electrification (~530–575 EUR/tCO_2e) and ventilation improvements (R2→R3, ~600–610 EUR/tCO_2e) form the low-cost tier, while deep envelope insulation (R1→R2) costs 1,100–2,400 EUR/tCO_2e depending on building size, consistent with analyzes questioning the cost-effectiveness of deep envelope retrofits in moderate climates [54], and with IEA findings on cost-effective district-level renovation [55]. Recent UBEM-based retrofit studies confirm that building-level variation drives practical investment decisions: multi-scale analyzes show that urban context modulates retrofit efficacy across districts [56], while data fusion approaches improve stock-level retrofit characterization [57].

3.2.3. Path-dependent marginal abatement cost curve

These district totals mask building-level variation in modeled capital abatement costs. For each building, the pipeline generates all incremental upgrade steps from its current level to R4 (e.g., a building at R1 produces three steps: R1→R2, R2→R3, R3→R4). The marginal abatement cost curve (MACC) [58] pools all steps across all buildings and ranks them by EUR/tCO_2e, but enforces prerequisite chains: a building's R2→R3 step only becomes available after its R1→R2 step has been selected. This path dependency prevents infeasible investment sequences (Fig. 10); the unit costs underlying these curves are documented in Appendix C. The resulting curve represents *technical* abatement potential under physical-feasibility and prerequisite-sequencing constraints; the prerequisite chains are the model's representation of measure compatibility. It is not a prediction of realizable municipal adoption: construction feasibility, ownership structure, heritage restrictions, occupant disruption, installer capacity, and financing are not modeled. The engine does support budget-capped selection (Fig. 11) as a first step toward pathway construction. Abatement costs use official RVO point-estimate benchmarks (Appendix C); their absolute values carry benchmark uncertainty, but the large separation between the cheapest interventions (ventilation, electrification) and the most expensive (deep envelope) keeps the ranking robust to plausible benchmark variation.

Nearly all interventions exceed the EU ETS carbon price (≈80 EUR/tCO_2e, 2024); only three large-building ventilation upgrades in Holten-Kom (62–74 EUR/tCO_2e) fall below, indicating, under the annualized-capital metric, that building-sector decarbonization requires policy instruments beyond carbon pricing alone, consistent with findings from national transition studies [59].

To make these building-level results accessible to municipal planners, DELFT-UBEM is available both as a containerized REST API that can be deployed on any server and as a digital twin web platform that serves an interactive 3D interface (Fig. 11). The API layer, built on Flask, accepts geographic boundaries, scenario definitions, and custom building data; the web platform renders building-level results on a 3D city model. The technical architecture, including the API design and 3D visualization pipeline, is detailed in [35].

3.2.4. Sensitivity of the abatement ranking

The ranking of measure classes was tested against the assumptions it depends on by recomputing the Holten-Kom marginal abatement cost curve under varied conditions. Grid decarbonization strengthens the electrification finding: under a hypothetical lower-grid-carbon sensitivity setting of 0.21 kg CO_2e/kWh, relative to the 0.270 operational baseline, the cost per ton of the heat-pump steps falls substantially, because their residual emissions shrink, while envelope-step costs are unaffected. In the adverse direction, a conservative high-carbon grid (0.45), the minimum-to-maximum span of the RVO cost benchmarks, and a discount rate raised from 3% to 5% each move costs but never move heat-pump steps out of the low-cost tier or deep-envelope steps into it. Because the headline metric is capital-only, it is invariant to energy prices by construction; applying the ISDE heat-pump subsidy only widens the electrification margin. The seasonal heat-pump performance assumed here (a fixed COP of 3.5) lies within the range a temperature-dependent Carnot model produces for this climate, and its influence on the ranking is bounded by the far larger grid-carbon variation already tested. Across every combination, including a combined adverse case (high-carbon grid, minimum envelope cost, and a 5% discount rate), heat-pump electrification retained a lower marginal abatement cost than deep-envelope insulation for the already-insulated majority of the stock.

4. Discussion

Automating the full UBEM pipeline changes what becomes practical at urban scale: the district retrofit analysis presented here reused the prepared registry, lookup, geometry, and simulation infrastructure rather than reconstructing a model pipeline; expert effort remained necessary to define scenarios, select cost and emission assumptions, check outputs, and interpret results, because the data architecture already carried each building from registry to simulation-ready model. This section examines what that automation covers and where it falls short: in tool capability, in modeling accuracy, and in the policy conclusions the results can support.

4.1. What automation actually means in UBEM

The term "automation" covers a wide range of claims in the UBEM literature. A tool that runs batch simulations calls itself automated; so does one that converts GIS footprints to geometry. End-to-end automation, where each stage's output becomes the next stage's input without manual intervention, is a different and stronger claim. A pipeline that automates geometry but requires manual schedule entry still concentrates the labor burden precisely where non-specialists are least equipped to handle it.

This stage-level framing is not incidental. Published reviews confirm that for most established tools, "the managing of input data is completely developed with other software and methodologies" [14]. No standard workflow connects these tools [33], and archetype definition has remained largely arbitrary, resting on deterministic assumptions and the modeler's own judgement [60]. Of the eight steps that a canonical UBEM project traverses (feasibility assessment, scope definition, GIS preparation, baseline model creation, simulation execution, upgrade definition, scenario analysis, and reporting), only simulation execution is consistently automated across established platforms.

DELFT-UBEM targets the remaining seven steps through a single config-driven pipeline. In practice this changes the unit of work: a new district is a boundary polygon, a new policy question is a configuration edit, and a 1,282-building district is generated and simulated in

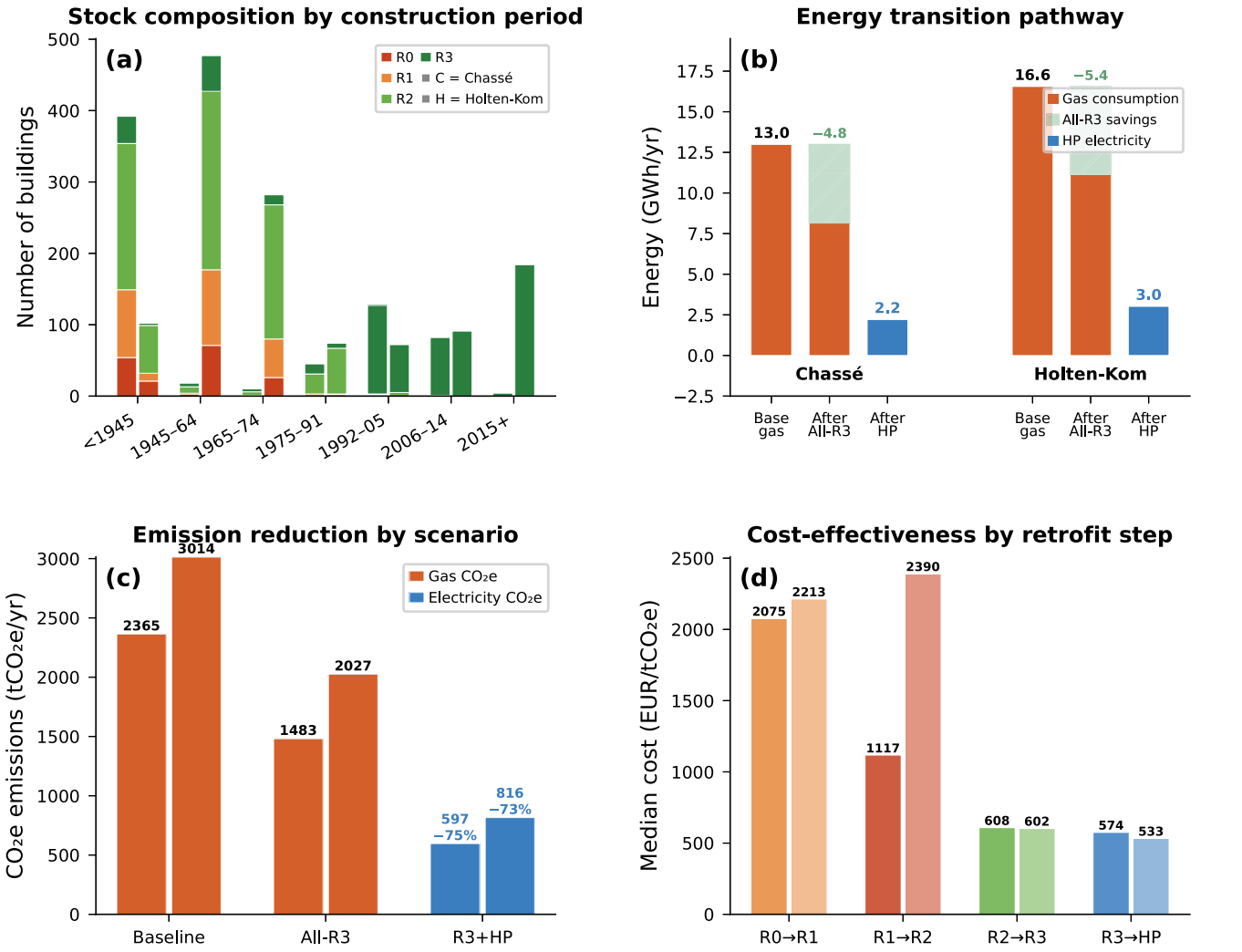


Fig. 9. District energy transition summary. (a) Building stock by construction period and baseline R-level. (b) Energy pathway from gas baseline through All-R3 to R4 electrification; hatched area shows All-R3 savings, blue bar shows residual heat-pump electricity. (c) CO_2e emissions by scenario; percentages indicate total reduction from baseline. (d) Median abatement cost (EUR/tCO_2e) by upgrade step. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

about 11 minutes on a workstation, with full result extraction dominated by parsing (Section 2.5). The deterministic, rule-based approach trades adaptability of case-by-case expert judgment for reproducibility and scalability.

4.2. *From building standards to simulation parameters*

Rich documentation for building energy modeling exists across European member states: national performance standards, EU directives, technical regulations, material databases. The gap is not in the existence of these sources but in translating them into simulation-ready parameters, a process that requires expertise in both building physics standards and simulation engine conventions, and whose results are rarely preserved in reusable form.

The challenge is compounded by the mismatch described: national standards target monthly steady-state calculations, and converting their parameters into hourly simulation objects requires engineering assumptions documented nowhere in the standards themselves. Where sources conflict (as when TABULA and NTA8800 prescribe different infiltration values for the same building type), the modeler must make judgment calls that go unrecorded.

The archetype library developed for this study (35 residential archetypes across 5 retrofit scenarios, with documented mappings from VW2022 and NTA8800 to EnergyPlus objects) provides one attempt to organize and share this effort. Published alongside the tool, it makes explicit the hundreds of translation decisions that typically remain buried in project-specific scripts: which standard provides each parameter, where assumptions fill gaps, and which values lack authoritative sources entirely. Non-residential archetypes expose these gaps most clearly: a single “office” archetype must span buildings with fundamentally different internal load profiles, and reference data for validation is sparse.

The same gap affects validation data: measured energy consumption at building level is unavailable in the Netherlands, and at postcode level it conflates gas-heated buildings with heat pump and district heating

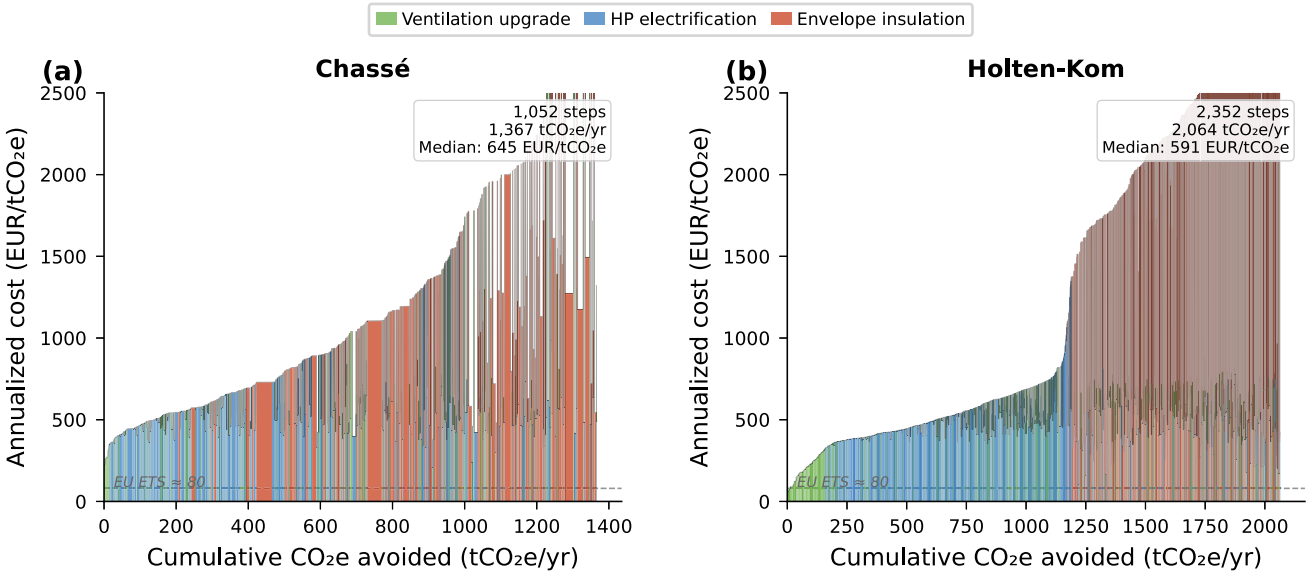


Fig. 10. Path-dependent building-level marginal abatement cost curves (technical potential). Each bar represents one upgrade step for one building, colored by type: ventilation upgrade (green), heat pump electrification (blue), and envelope insulation (red). Dashed line: EU ETS carbon price (≈80 EUR/tCO_2e). Chassé: 1,052 steps, 1,367 tCO_2e/yr total abatement potential, median 645 EUR/tCO_2e. Holten-Kom: 2,352 steps, 2,064 tCO_2e/yr, median 591 EUR/tCO_2e. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

connections. Finding consistent, scale-appropriate validation data across building types and climate zones remains a primary obstacle for any UBEM study.

4.3. Toward standardized UBEM evaluation

Current UBEM tool comparisons focus on supported features (which simulation engine, which geometry sources, which output formats) without standardized benchmarks for the outcomes that matter: accuracy at different scales, automation completeness, input burden, and output utility. The result is that each project defines its own validation criteria, compares against different benchmarks, and reports metrics at inconsistent scales. When three established tools were applied to the same Berlin district, heated-area estimates varied by 1.9× across tools, and within a single tool the number of buildings actually simulated ranged from 365 to 646, driven by differences in data ingestion and archetype assignment rather than simulation physics [61]. If tools cannot agree on how many buildings a district contains, comparing their energy predictions is premature.

Existing validation standards do not transfer to urban scale. ASHRAE Guideline 14 [44], designed for calibrated single-building models (Section 3.1.1), provides no guidance for stock-level accuracy, where aggregation effects cause district-level predictions to appear accurate (often 1–10% deviation) while individual building errors can exceed 100% [45]. No consensus exists on acceptable accuracy at intermediate scales, nor on how error distributions should be characterized beyond point estimates [46,62]. The gap between predicted and measured performance remains a persistent challenge across all scales [63].

Several evaluation dimensions deserve standardization beyond simulation accuracy alone. *Automation completeness*: distinguishing batch simulation from end-to-end automation (pipelines requiring minimal expert intervention from geographic boundary to scenario comparison). *Input burden*: what data does a user need to provide, and what does the tool resolve automatically? *Preparation effort*: elapsed time from project initiation to first simulation results, including data acquisition and model setup. *Output resolution*: temporal (annual, monthly, hourly), spatial (stock-average, archetype-level, building-level), and physical (heating only, all end-uses, demand profiles). *Extensibility*: how readily can a developer add new building systems, archetypes, or analytical methods without modifying the core pipeline? *Reproducibility*: are data sources, archetype definitions, and parameter mappings documented sufficiently for independent replication? Standardized reference districts (publicly available datasets with geometry, archetypes, and measured consumption across climate zones) would enable comparison across these dimensions. Several of these dimensions, notably automation completeness and reproducibility, are the criteria along which the Introduction compares existing tools; the broader aim is to standardize them across studies rather than for a single comparison.

This paper contributes toward these needs by documenting the complete data transformation pipeline and releasing the source code, archetype library, lookup tables, and parameter mappings upon publication so that the automation and reproducibility claims made here can be independently verified.

4.4. International comparison and context dependence

Reported retrofit percentages depend on climate, initial stock condition, intervention scope, heating-system efficiency, and the accounting boundary, so cross-study comparison establishes context rather than validation. Ang et al. [64] model seed neighborhoods in eight municipalities spanning diverse climates: deep retrofits reduce whole-building site energy-use intensity by 32–66% in seven of them, Braga excepted because its scenario includes future air-conditioning uptake, and cut operational carbon by 34–84% before further grid decarbonization. The endpoints are not directly comparable with the present result. The Middlebury and Kiel packages substitute the heating system, raising average heating COP from about 0.9–1.0 to 3.0–3.5, and report whole-building site energy, whereas the 33–37% reported here is a space-heating demand reduction with the gas boiler retained, for a stock in which 77% of buildings already start at R2 or R3 (Section 3.2).

Package definition also affects the reported reduction. In Varberg, Johari et al. [65] report that renovation to Swedish building-code envelope requirements reduced aggregate annual residential energy demand from 193 to 182 GWh/yr (6%), whereas a Passive House package with deeper envelope and window targets and mechanical heat recovery of at least 75% reduced it to 110 GWh/yr (43%); the authors identify the heat-recovery criterion as the most influential difference. All-R3 also

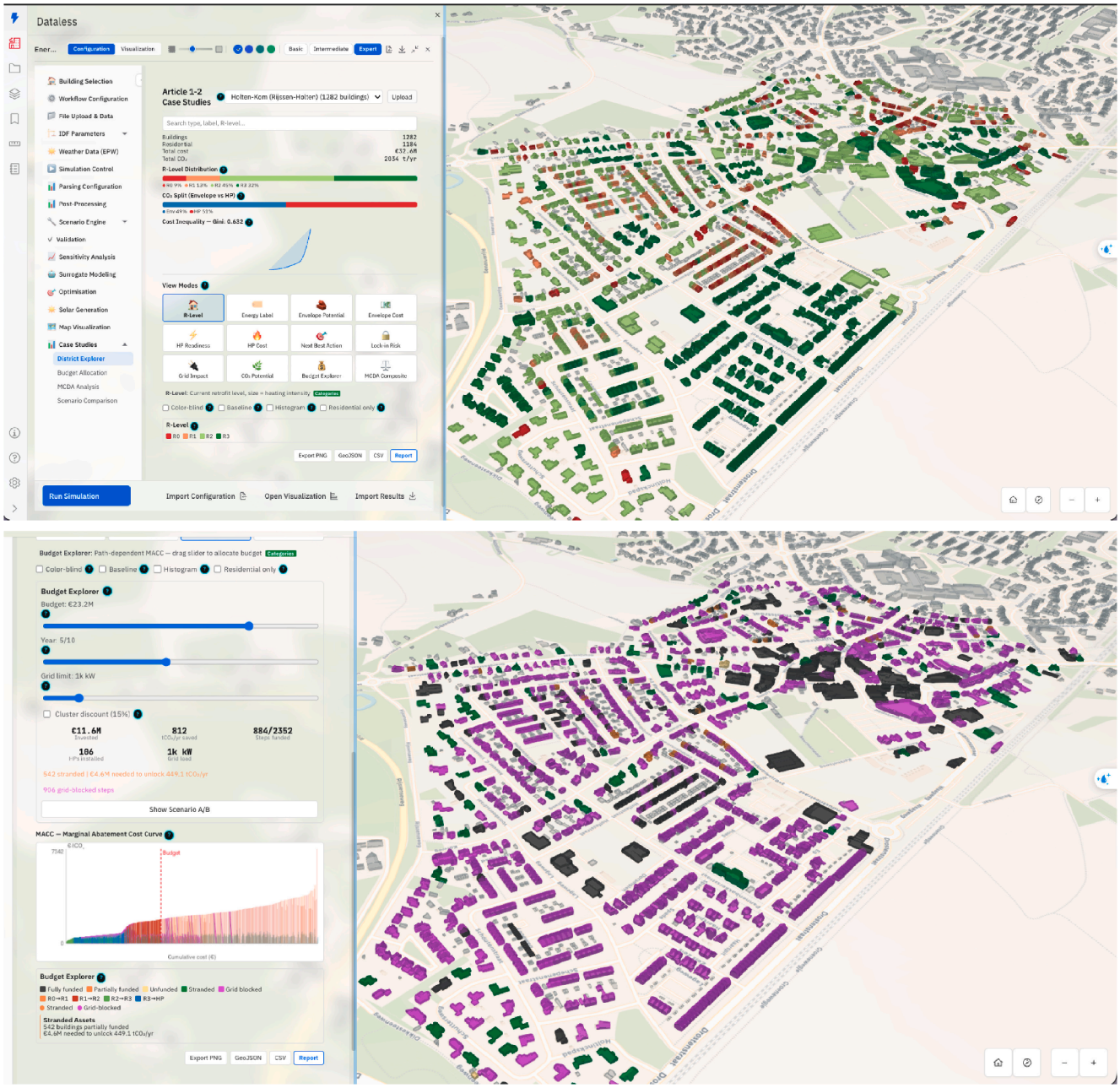


Fig. 11. Interactive planning platform for district energy transition analysis. (a) Scenario configuration for the Holten-Kom case study, showing baseline assignment and retrofit scenario selection across 1,282 buildings. (b) Budget explorer with marginal abatement cost curve: the platform selects cost-optimal retrofit sequences (purple buildings) within a user-defined budget, reporting aggregate savings, investment, and CO_2e reduction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

includes heat recovery for buildings upgraded to R3 but, as defined in Section 3.2, applies only the measures each building needs to reach R3.

End-use composition also changes which measures are attractive. Across 202,391 buildings in Hong Kong, Zhang et al. [66] find that chiller and lighting upgrades reduce peak electricity demand by up to 6.86% and 3.44%, respectively; under their financial assumptions lighting yields the strongest returns, rooftop photovoltaics achieve internal rates of return of 5–10%, and envelope and window retrofits are economically unattractive. For an East Portland neighborhood of more than 43,000 homes, Hong et al. [67] find that heat-pump electrification reduces annual CO_2e emissions by 38% but raises peak electricity demand by approximately 9%; combining electrification with efficiency measures reduces emissions by 48% and peak demand by approximately 6%. Their emission factors include upstream fuel-chain stages, whereas the present study uses a common operational boundary. The R4 scenario concerns space heating and does not evaluate coincident network peaks. Its ranking should therefore be interpreted within the Dutch climate, stock condition, grid mix, intervention scope, and accounting boundary rather than as a universal retrofit priority.

4.5. *Limitations and research outlook*

The principal challenge in building DELFT-UBEM was not the simulation engine or the pipeline architecture but locating, collecting, and reconciling the data needed to populate it. Dutch building performance is documented across dozens of regulatory documents, datasets, and

technical standards in different formats and at different levels of detail, and resolving contradictions between overlapping sources consumed the majority of the development effort. Every UBEM study repeats some version of this work, and the inconsistency it introduces undermines cross-study comparability. A community-maintained hub — where regulatory parameters from different national contexts are mapped to simulation-engine inputs in a common, versioned format aligned with existing schemas such as the CityGML Energy ADE [36] and the TABULA/EPISCOPE archetype classification — would allow multiple UBEM tools to share the same upstream data effort rather than each reimplementing it independently.

Coverage of the current implementation is bounded: residential archetypes are treated with reasonable depth, but extending to all building functions, physical domains, and construction periods is incomplete. Adaptation to non-Dutch contexts does not require redesigning the core (orchestration, geometry generation, IDF assembly, simulation, parsing, and the scenario and MACC engines are country-agnostic), but it does require country-specific data and expert mapping effort: a building registry with footprint, construction year, function, and height; archetype parameter tables (the TABULA/EPISCOPE library covers some twenty European countries [68] as a starting point); ventilation, DHW, and infiltration mappings from the national EPB method; an energy-label register for baseline assignment; local weather-station EPW files; and national cost benchmarks. Only the Dutch implementation has been validated; the scarcest transfer ingredient is a VW2022-style reference-dwelling dataset, needed for benchmark validation rather than for modeling itself. The Energy Performance of Buildings Directive (EPBD) recast makes the required inputs increasingly mandatory across member states.

The external-obstruction module assumes locally level terrain. It was not activated in the reported simulations, so the influence of neighboring-building and tree shading on solar gains, annual demand, and retrofit outcomes was not quantified.

The principal modeling simplifications (single-zone thermal modeling on equivalent-rectangle prismatic massing, deterministic schedules, no occupant metabolic gains, and ideal-loads HVAC, where reported demands are heat delivered to zones and plant efficiency is applied in post-processing, whose influence Section 3.2.4 examines) are deliberate choices that prioritize stock-scale scalability over single-building fidelity; the lookup-table architecture isolates each so that a multi-zone model or stochastic occupant generator could replace the current defaults without altering the rest of the pipeline. Occupant-related uncertainty extends beyond metabolic gains to schedules, setpoint behavior, and prebound adaptation. The two dominant effects act on different parts of the stock: prebound adaptation is concentrated in poorly insulated dwellings, so it reduces envelope savings more than electrification savings and shifts the ranking toward electrification; omitted metabolic gains, by contrast, have a proportionally larger effect in well-insulated dwellings, where they over-estimate the demand that electrification addresses. The net influence on the ranking is therefore bounded rather than one-directional. The abatement curves represent technical potential, bounded further by adoption factors outside the model.

Future work targets three directions: coupling hourly demand profiles with grid topology models to identify distribution network bottlenecks that the Dutch *wijkaanpak* requires [2,59,65], completing the sensitivity, surrogate-modeling, and optimization extensions for which the modification engine already provides infrastructure [69,70], and packaging the pipeline as a Python library with Large Language Model (LLM)-assisted model configuration to lower the adoption barrier further. These demand-side extensions can be paired with community-level supply-side optimization — reinforcement-learning energy management and multi-objective battery dispatch [71,72] — to close the loop between building-level demand profiles and microgrid operation.

5. Conclusion

The primary barrier to UBEM adoption is not simulation capability but the workflow surrounding it: data acquisition, geometric processing, archetype parameterization, and scenario management each require specialist effort that is rarely preserved or transferable between projects. This paper asked whether that entire workflow can be automated from national registries and building performance standards alone.

The DELFT-UBEM pipeline demonstrates that this is possible. Where national building registries and codified performance standards exist, a geographic boundary and a configuration file are sufficient to generate district-scale EnergyPlus models.

Validation at two scales (archetype benchmarks and postcode-level measured consumption) confirms that this standards-driven parameterization reproduces structural variation in Dutch residential heating demand without building-specific calibration. The validation identifies systematic deviations associated with geometry representation, heating-system assignment, label coverage, and weather input; where controlled ablations exist (Section 3.1.1), individual contributions are quantified, and the modular architecture allows each to be addressed, without requiring changes to other pipeline stages.

The district retrofit analysis demonstrates the practical value of end-to-end automation. The analysis itself revealed that for the studied districts, electrification rather than further insulation is the dominant transition lever for the majority of the stock, a ranking that holds across the tested ranges of grid carbon intensity, discount rates, measure costs, and heat-pump subsidy (Section 3.2.4), and that describes modeled technical potential for the studied stocks under the stated capital-cost metric rather than a universal retrofit priority.

The principal frontiers are data coverage, validation standardization, and generalization. Energy label gaps affect baseline assignment for nearly half the stock. The field lacks agreed benchmarks for evaluating UBEM tools across comparable dimensions (automation scope, data requirements, output resolution, validation transparency), limiting meaningful cross-tool comparison. Regarding generalization, the lookup-table architecture isolates country-specific parameters, separating the reusable workflow core from country-specific registry, standards, weather, economic, and validation adapters (Section 4.5); adaptation requires no core redesign, but transferability beyond the Netherlands remains unvalidated and requires substantial national data and expert mapping work. As the EU EPBD recast demands building-level renovation roadmaps across all member states, the infrastructure demonstrated here (automated, open, and standards-driven) provides one operational path toward delivering them at the scale and speed that municipal energy transition planning requires.

CRediT authorship contribution statement

Amin Jalilzadeh: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization; **Azarakhsh Rafiee:** Writing – review & editing, Supervision, Resources; **Thaleia Konstantinou:** Writing – review & editing, Supervision; **Stefan de Graaf:** Writing – review & editing, Resources, Data curation; **Henk Scholten:** Writing – review & editing, Resources; **Peter J.M. van Oosterom:** Writing – review & editing, Supervision.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used Claude AI in order to improve grammar, enhance readability, and polish draft texts. The AI was used as a language enhancement tool to improve the clarity

and flow of the manuscript. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Funding

This work is part of the DATALESs project (with project number 482.20.602) jointly financed by the Netherlands Organization for Scientific Research (NWO), and the National Natural Science Foundation of China (NSFC).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: This work is part of the DATALESs project (with project number 482.20.602), with 5 work packages defined as PhD topics, jointly financed by the Netherlands Organization for Scientific Research (NWO), and the National Natural Science Foundation of China (NSFC).

Data availability

Upon publication of this article, the DELFT-UBEM source code, archetype library, lookup tables with per-row source attribution, parameter mappings, and the configuration files used for the reported runs will be released publicly as a versioned, tagged repository release at https://github.com/amin-jalilzadeh-tu/FPLUS_10, under the licenses stated in the repository. Extended documentation and user manuals will accompany the release at https://delftaubem.com. Because the lookup library comprises approximately 5,300 parameter records, the appendices reproduce only selected representative tables; the complete versioned set will be provided in the repository. Where third-party standards restrict redistribution, the repository provides source metadata and transformation scripts rather than reproducing protected source material. An interactive web-based platform for scenario analysis and visualization is accessible at https://dataless.beta.geodan.nl/. Simulation input data are derived from open national registries. Processed datasets generated during this study are available from the corresponding author upon reasonable request.

Appendix A. Database schema

The pipeline's spatial data are held in two core database tables: the primary building table (Table A.1), which stores registry attributes, geometry, boundary conditions, and the energy label per building, and the shading-objects table (Table A.2), which stores the computed obstruction geometry per façade.

A.1. Primary building table

A.2. Shading objects table

Table A.1
Primary building database schema.

Field	Type	Description
Identifiers		
ogc_fid	Integer	Unique record identifier
pand_id	String	National building registry ID (BAG)
postcode	String	Postal code for spatial aggregation
Geometric Attributes		
area	Float	Ground floor area [m^2]
perimeter	Float	Building footprint perimeter [m]
height	Float	Building height [m]
gem_hoogte	Float	Average height from 3D model [m]
gem_bouwlagen	Integer	Number of floors
x, y	Float	Coordinates in national grid (RD New)
lon, lat	Float	Coordinates in WGS84
Roof Attributes		
b3_dak_type	String	Roof type (flat, pitched, multiple)
b3_opp_dak_plat	Float	Flat roof surface area [m^2]
b3_opp_dak_schuin	Float	Sloped roof surface area [m^2]
Classification Attributes		
bouwjaar	Integer	Construction year
age_range	String	Construction period category
building_function	String	Primary function (residential/non-residential)
residential_type	String	Dwelling type (e.g., Corner House, Apartment)
non_residential_type	String	Non-residential function (e.g., Office, Retail)
verblijfsobjecten	Integer	Number of dwelling units in building
meestvoorkomendelabel	String	Most common energy label (A–G)
Boundary Conditions		
north_side	String	North façade exposure (facade/shared)
east_side	String	East façade exposure (facade/shared)
south_side	String	South façade exposure (facade/shared)
west_side	String	West façade exposure (facade/shared)
building_orientation	Float	Primary orientation angle [degrees]
building_orientation_cardinal	String	Cardinal direction (N, NE, E, etc.)
Simulation Parameters		
average_wwr	Float	Window-to-wall ratio [0–1]

Table A.2
Shading objects database schema.

Field	Type	Description
ep_id	Integer	Unique shading object identifier
focus_ogc_fid	Integer	Reference to target building
edge_label	String	Façade receiving shade (north, east, etc.)
object_id	Integer	Identifier of shading object
object_type	String	Type of obstruction (building/tree)
object_height	Float	Height of shading object [m]
distance_to_facade	Float	Distance to target façade [m]
height_diff	Float	Height difference from target building [m]
angle_deg	Float	Azimuth angle relative to façade [degrees]
shading_score	Float	Computed shading impact score
transmittance	Float	Base solar transmittance [0–1]
transmittance_summer	Float	Summer transmittance (vegetation) [0–1]
transmittance_winter	Float	Winter transmittance (vegetation) [0–1]

Appendix B. Data sources

The tables in this appendix reproduce selected representative entries from the lookup library; the complete set of lookup tables, archetype definitions, and parameter mappings will be released with the source code upon publication (see Data availability).

B.1. Data sources overview

Table B.1
Data sources, their role in the simulation pipeline, and key parameter ranges.

Source	Role in Pipeline	Key Parameters (ranges)
Voorbeeldwoningen 2022 [42]	Primary residential parameterization	Wall Rc 0.19–4.50, Roof Rc 0.22–6.00, Floor Rc 0.15–3.50 m^2K/W; Window U 1.40–5.10 W/m^2K; SHGC 0.40–0.80; Heat demand 13–401 kWh/m^2
TK2020 [73]	Reference benchmarks for infiltration, heating capacity, insulation targets (validation)	Heating capacity 2,982–32,070 W; Heat demand 32–209 kWh/m^2
NTA8800 [74]	Ventilation, infiltration, DHW, schedules, HVAC efficiency	Infiltration forfait $q_{v;10}$ 0.12–1.44 $L/s{\cdot}m^2$ (residential) and 0.15–0.90 (non-residential), by year band and type; Boiler eff. 0.65–0.94; COP 1.6–4.7; HRV 55–85%; DHW 12.5 kWh/m^2
TABULA [68]	Cross-validation, gap-filling	ACH 0.05–0.40 h^{-1}; HRV 55–90%; Heat demand 16–225 kWh/m^2
3DBAG [75]	Building geometry (LoD1/LoD2), heights, roof types	Footprint, height, roof surface areas for 10M buildings
BAG [48]	Building registry: function, construction year, address	Building identifiers, postcode, bouwjaar
RVO Kostenkentallen [49]	Retrofit cost data (Appendix C)	Envelope, ventilation, heat pump costs per dwelling type
CBS [47]	Postcode validation (Section 3.1.2)	Postcode-level gas/electricity consumption
Liander [76]	Postcode validation	Postcode-level annual gas/electricity consumption for small connections
TNO [77]	Non-residential parameterization	Office, retail, hospitality: indicative annual energy use and floor/roof/facade areas by dominant use function
EP-Online [78]	Energy label validation	NTA8800-based energy performance certificates for registered NL buildings
Enexis [79]	Postcode validation (complementary to Liander)	Postcode-level gas/electricity consumption
KNMI [80]	Weather input (EPW file)	Hourly temperature, radiation, wind, humidity; nearest of 13 reference stations per building
Guerra-Santin & Silvester [81]	Occupancy and heating schedules	Dutch occupancy profiles and heating setpoint patterns
ISO 10,456 [82]	Materials library thermal properties	Design thermal conductivity values for building materials
BENG reference buildings [83]	Post-2020 archetype definitions	Nearly Zero Energy Building standards
Bouwbesluit 2012 [84]	Airtightness requirements	Dutch building code

B.2. Dwelling type mapping

Table B.2
Dwelling type mapping between sources. Cost class: egw = single-family house (*eengezinswoning*), mgw = apartment (*meergezinswoning*).

BAG Type	VW Code	TK Code	TABULA	Model Type	Cost
woonfunctie	vrij	Vrijstaand	SFH	Detached House	egw
woonfunctie	2-1	–	SFH	Semi-detached housesz	egw
woonfunctie	rij (tuss)	Tussenwoning	TH	Mid-terrace House	egw
woonfunctie	rij (hoek)	Hoekwoning	TH	Corner House	egw
woonfunctie	gal/port/mais	Galerij/Portiek	MFH/AB	Apartment	mgw

B.3. Complete pipeline data flow

Table B.3
Complete data flow from sources to ENERGYPLUS objects. Source types: D = Direct, Der = Derived, A = Assumed, C = Compiled from multiple sources.

Parameter Domain	Type	Primary Source	Lookup CSV	IDF Module	Records
Envelope (Rc, U, SHGC)	D/Der	VW2022	residential_materials (351), non_res_materials (701)	fenez	1,052
Ventilation systems	D	VW2022, NTA8800	ventilation_params	ventilation	440
Infiltration (qv10)	D	NTA8800 forfait	ventilation_params (forfait table)	ventilation	299
HVAC setpoints	D	NTA8800	hvac_setpoints	HVAC	106
DHW consumption	D	NTA8800	dhw_params	DHW	33
Lighting density	A	NEN / NTA8800	lighting_params	Elec	32
Equipment loads	A	CBS statistics	equipment_params	eequip	49
Schedules	A	NTA8800 profiles	hvac_schedules	Multiple	106
Ground temperature	Der	ISO 13370	groundtemp_params	tempground	24
Window shading	A	NTA8800 + eng. judgment	shading_params	wshading	67
Geometry (WWR)	D/Der	VW2022 + BAG	geometry_params	geomz	30
Retrofit packages	D	VW2022	retrofit_packages_VW2022	Scenario engine	176
Retrofit costs	D	RVO Kostenkentallen 2025	envelope_costs (97), hp_costs (18)	Cost module	~115
Energy benchmarks	D	VW2022 + TK2020	heating_benchmarks (603), cooling/dhw/elec	Validation	~2,500
Materials library	C	Multiple	materials_library	fenez	297
Total					~5,300

B.4. Regulatory parameter to ENERGYPLUS object mapping

Table B.4
Mapping from each regulatory parameter to its receiving ENERGYPLUS object and the governing translation assumption.

Parameter	ENERGYPLUS object	Key translation assumption
Envelope (Rc, U)	`Construction` (multi-layer)	era-typical layer build-ups; lookup-based ΔU 0.012–0.30 W/m^2K thermal-bridge surcharge on opaque exterior constructions
Zone conditioning (A–D)	`ZoneHVAC: IdealLoadsAirSystem`	ideal zone heating and cooling demand for all four systems; conversion to fuel or electricity in post-processing
Scheduled ventilation (A–C)	`ZoneVentilation: DesignFlowRate`	System A supplies only the flow not met by infiltration and creates no object when none is required; Systems B and C represent mechanical supply or exhaust airflow
Balanced ventilation (D)	`DesignSpecification: OutdoorAir` linked to `ZoneHVAC: IdealLoadsAirSystem`	scheduled outdoor air through IdealLoads; native sensible heat recovery; latent effectiveness zero
Night cooling	`ZoneVentilation: DesignFlowRate`	separate natural-ventilation objects on the night-cooling schedule; not the representation of ventilation class A–D
Infiltration ($q_{v;10}$)	`ZoneInfiltration: DesignFlowRate` (Flow/Area)	$q_{v;10}$ applied directly as the design rate; weather coefficients normalized to 10 K / 3 m/s; Systems B–D use a leakage-dependent maximum permitted reduction of 10–28%
HVAC setpoints	`ThermostatSetpoint: DualSetpoint`	uniform day/night pattern
DHW	`WaterHeater:Mixed`	floor area → occupancy bands; integral-normalized draw schedule
Lighting / equipment	`Lights` / `ElectricEquipment` + `Schedule:Compact`	EFLH-scaled daily shapes with monthly multipliers
Ground temperature	`Site:GroundTemperature: BuildingSurface` (ISO 13370)	$B' = A/(0.5\,P_{exposed})$; indoor 18 °C
Window shading	`WindowShadingControl`	archetype-specific irradiance setpoints

Appendix C. Retrofit cost data

C.1. Dwelling size classification

Table C.1
Dwelling size classification for cost differentiation.

Type	Code	Small (min)	Medium (mid)	Large (max)
Houses (detached, semi, corner, terraced)	egw	<80 m^2	80–120 m^2	>120 m^2
Apartments (flat, maisonette, gallery, porch)	mgw	<50 m^2	50–80 m^2	>80 m^2

C.2. Envelope costs

Table C.2
Envelope retrofit costs (EUR/m^2). R2–R4 share the same deep-insulation target. Pitched roof costs are identical for egw and mgw.

Element	Level	Intervention	egw min	egw mid	egw max	mgw min	mgw mid	mgw max
Wall	R1	Cavity fill mineral wool	15.23	13.66	12.33	15.15	14.20	13.64
Wall	R2–R4	External wall insulation	211.74	209.68	206.45	214.57	212.26	209.27
Floor	R1	Underfloor insulation basic	53.31	52.27	50.68	52.93	51.68	51.11
Floor	R2–R4	Deep floor insulation	76.84	76.85	75.04	76.57	75.89	75.36
Pitched roof	R1	Interior roof insulation	96.26	95.61	94.67		—	
Pitched roof	R2–R4	Deep interior insulation	99.86	99.21	98.28		—	
Flat roof	R1	Exterior roof insulation	256.75	247.57	247.26	248.18	246.53	247.07
Flat roof	R2–R4	Deep roof insulation	317.73	308.55	308.23	307.99	307.51	306.76
Window	R1	HR++ glass replacement	246.33	234.37	219.88	314.05	284.84	274.88
Window	R2–R4	Triple glazing + frame	770.22	764.18	754.95	814.68	799.64	792.08

C.3. Ventilation costs

Table C.3
Ventilation system costs (EUR/dwelling).

Level	System	Intervention	egw	mgw
R1	C2	MV-unit replacement	590	590
R2	C2	Full mech. exhaust from natural	3,683	3,476
R3–R4	D2	Central MVHR from natural	7,941	7,734

This appendix documents the unit costs used in the MACC analysis (Section 3.2): envelope, ventilation, and heat-pump unit costs are listed in Tables C.2–C.4 respectively. All envelope costs are from RVO Kostenkentallen 2025 (construction costs including 21.6% overhead), differentiated by dwelling type: single-family houses (egw) and apartments (mgw), each in three size classes (Table C.1); the registry-to-model dwelling-type mapping is given in Table B.2. Heat pump costs are from CBS 2023–2024 market data with ISDE 2025/2026 subsidy rates. ISDE subsidy amounts are tabulated for reference; the abatement cost curves use gross installed costs (subsidy excluded), and the subsidized variant appears in the sensitivity analysis (Section 3.2.4).

C.4. Heat pump costs

Table C.4
Heat pump installed costs and Sizing (EUR/building).

Size (kW)	Category	Equipment	Installation	Total	Net cost
4	residential	2,800	1,200	4,000	1,875
6	residential	3,500	1,500	5,000	2,425
8	residential	4,500	2,200	6,700	3,675
10	residential	5,500	2,800	8,300	4,825
12	residential	6,500	3,500	10,000	6,075
14	residential	7,500	4,500	12,000	7,625
16	residential	8,500	5,500	14,000	9,175
20	small_commercial	10,000	6,000	16,000	10,275
25	small_commercial	11,500	7,500	19,000	12,150
30	small_commercial	13,500	9,000	22,500	14,525
40	small_commercial	18,000	11,000	29,000	18,775
50	small_commercial	22,000	13,000	35,000	22,525

Appendix D. Configuration reference

Table D.1 lists the fields the user defines, grouped by pipeline stage, together with the values used for the reported study; fields left unset fall back to the defaults shown.

Table D.1
User-defined configuration fields and their reported-run values.

Group	What the user defines	Reported-run value
Study area	Bounding box, postcode list, or building-ID list, or a user-supplied building CSV	Postcode / district boundary
Pipeline stages	Per-stage switches (geometry, parameterization, simulation, post-processing, analysis)	All enabled
Scenario assignment	Rule mapping buildings to scenarios: default, per-archetype, or per-energy-label, including the missing-label rule	Per-energy-label; missing labels imputed by construction year
Parameter selection	Assignment strategy and random seed	Midpoint (benchmark); seeded draw, seed 42 (postcode, district)
Parameter overrides	Optional overrides per building, per archetype, or stock-wide, as a fixed value or a range	None
Weather	Nearest KNMI station, a fixed weather file, or a target climate year	Nearest-station assignment (district); single typical-year file (postcode validation)
Simulation	Parallel workers, per-run timeout, and output reporting frequency	8–10 workers; hourly and monthly output
Retrofit packages	Selection of predefined renovation levels (R0–R4) or custom parameter modifications	Predefined R0–R4
Shading	Window-shading solar-irradiance setpoints (archetype-assigned, user-overridable); building and tree obstruction shading is precomputed in the database and not configured per run	Archetype setpoints; obstruction shading not activated
Economic analysis	Emission factors, energy prices, discount rate, component lifetimes, cost tier (min/avg/max), and subsidy on/off	Grid 0.270 kg CO_2e/kWh, gas 1.779 kg CO_2e/Nm^3; discount 3%; 25/15 yr; average cost; subsidy off

References

[1] European Parliament and Council of the European Union, Directive (EU) 2024/1275 on the energy performance of buildings (recast), 2024, . Official Journal of the European Union, L 2024/1275, published 8 May 2024; entered into force 28 May 2024, https://eur-lex.europa.eu/eli/dir/2024/1275/oj/eng.

[2] A. Jalilzadeh, A. Rafiee, P. van Oosterom, T. Konstantinou, A comprehensive review and framework on the applications of digital twins for energy transition at the district level, Renew. Sustain. Energy Rev. 234 (2026) 116872. https://doi.org/10.1016/j.rser.2026.116872

[3] T. Hong, Y. Chen, X. Luo, N. Luo, S.H. Lee, Ten questions on urban building energy modeling, Build. Environ. 168 (2020) 106508. https://doi.org/10.1016/j.buildenv.2019.106508

[4] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, M. Djurovic-Petrovic, A review of bottom-up building stock models for energy consumption in the residential sector, Build. Environ. 45 (7) (2010) 1683–1697. https://doi.org/10.1016/j.buildenv.2010.01.021

[5] C.E. Kontokosta, C. Tull, A data-driven predictive model of city-scale energy use in buildings, Appl. Energy 197 (2017) 303–317. https://doi.org/10.1016/j.apenergy.2017.04.005

[6] Y. Chen, Z. Deng, T. Hong, Automatic and rapid calibration of urban building energy models by learning from energy performance database, Appl. Energy 277 (2020) 115584. https://doi.org/10.1016/j.apenergy.2020.115584

[7] A. Nutkiewicz, Z. Yang, R.K. Jain, Data-driven urban energy simulation (DUE-S): a framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow, Appl. Energy 225 (2018) 1176–1189. https://doi.org/10.1016/j.apenergy.2018.05.023

[8] A. Nutkiewicz, B. Choi, R.K. Jain, Exploring the influence of urban context on building energy retrofit performance: a hybrid simulation and data-driven approach, Adv. Appl. Energy 3 (2021) 100038. https://doi.org/10.1016/j.adapen.2021.100038

[9] C. Deb, A. Schlueter, Review of data-driven energy modelling techniques for building retrofit, Renew. Sustain. Energy Rev. 144 (2021) 110990. https://doi.org/10.1016/j.rser.2021.110990

[10] T. Hong, Z. Wang, X. Luo, W. Zhang, State-of-the-art on research and applications of machine learning in the building life cycle, Energy Build. 212 (2020) 109831. https://doi.org/10.1016/j.enbuild.2020.109831

[11] D.B. Crawley, J.W. Hand, M. Kummert, B.T. Griffith, Contrasting the capabilities of building energy performance simulation programs, Build. Environ. 43 (4) (2008) 661–673. Part Special: Building Performance Simulation, https://doi.org/10.1016/j.buildenv.2006.10.027

[12] C.F. Reinhart, C. Cerezo Davila, Urban building energy modeling - a review of a nascent field, Build. Environ. 97 (2016) 196–202. https://doi.org/10.1016/j.buildenv.2015.12.001

[13] H.-X. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (6) (2012) 3586–3592. https://doi.org/10.1016/j.rser.2012.02.049

[14] M. Ferrando, F. Causone, T. Hong, Y. Chen, Urban building energy modeling (UBEM) tools: a state-of-the-art review of bottom-up physics-based approaches, Sustain. Cities Soc. 62 (2020) 102408. https://doi.org/10.1016/j.scs.2020.102408

[15] C. Cerezo, J. Sokol, S. AlKhaled, C. Reinhart, A. Al-Mumin, A. Hajiah, Comparison of four building archetype characterization methods in urban building energy modeling (UBEM): a residential case study in Kuwait city, Energy Build. 154 (2017) 321–334. https://doi.org/10.1016/j.enbuild.2017.08.029

[16] Y. Chen, T. Hong, M.A. Piette, Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis, Appl. Energy 205 (2017) 323–335. https://doi.org/10.1016/j.apenergy.2017.07.128

[17] R.E. Kontar, B. Polly, T. Charan, K. Fleming, N. Moore, N. Long, D. Goldwasser, URBANopt: an open-source software development kit for community and urban district energy modeling, in: 2020 Building Performance Analysis Conference and SimBuild Co-organized by ASHRAE and IBPSA-USA, vol. 9 of *SimBuild Conference*, ASHRAE/IBPSA-USA, Virtual, 2020, pp. 293–301. https://publications.ibpsa.org/conference/paper/?id=simbuild2020_C035.

[18] D.B. Crawley, L.K. Lawrie, F.C. Winkelmann, W.F. Buhl, Y.J. Huang, C.O. Pedersen, R.K. Strand, R.J. Liesen, D.E. Fisher, M.J. Witte, J. Glazer, EnergyPlus: creating a new-generation building energy simulation program, Energy Build. 33 (4) (2001) 319–331. https://doi.org/10.1016/S0378-7788(00)00114-6

[19] D. Robinson, F. Haldi, J. Kämpf, P. Leroux, D. Perez, A. Rasheed, U. Wilke, CITYSIM: comprehensive micro-simulation of resource flows for sustainable urban planning, in: Proceedings of the Eleventh International IBPSA Conference (Building Simulation 2009), Glasgow, 2009, p. 1083–1090. https://publications.ibpsa.org/proceedings/bs/2009/papers/bs2009_1083_1090.pdf.

[20] J.A. Fonseca, T.-A. Nguyen, A. Schlueter, F. Marechal, City energy analyst (CEA): integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts, Energy Build. 113 (2016) 202–226. https://doi.org/10.1016/j.enbuild.2015.11.055

[21] P. Remmen, M. Lauster, M. Mans, M. Fuchs, T. Osterhage, D. Müller, TEASER: an open tool for urban energy modelling of building stocks, J. Build. Perform. Simul. 11 (1) (2018) 84–98. https://doi.org/10.1080/19401493.2017.1283539

[22] F. Johari, G. Peronato, P. Sadeghian, X. Zhao, J. Widén, Urban building energy modeling: state of the art and future prospects, Renew. Sustain. Energy Rev. 128 (2020) 109902. https://doi.org/10.1016/j.rser.2020.109902

[23] K. Orehounig, L. Fierz, J. Allan, S. Eggimann, N. Vulic, A. Bojarski, CESAR-P: a dynamic urban building energy simulation tool, J. Open Source Softw. 7 (78) (2022) 4261. https://doi.org/10.21105/joss.04261

[24] Z. Deng, Y. Chen, J. Yang, F. Causone, AutoBPS: a tool for urban building energy modeling to support energy efficiency improvement at city-scale, Energy Build. 282 (2023) 112794. https://doi.org/10.1016/j.enbuild.2023.112794

[25] Y.Q. Ang, Z.M. Berzolla, S. Letellier-Duchesne, V. Jusiega, C. Reinhart, UBEM.io: a web-based framework to rapidly generate urban building energy models for carbon reduction technology pathways, Sustain. Cities Soc. 77 (2022) 103534. https://doi.org/10.1016/j.scs.2021.103534

[26] S. Zhang, L. Chen, L. Xu, Z. Wang, GeoBEM: a geospatial computing empowered framework for urban-scale building energy modeling, Sustain. Cities Soc. 121 (2025) 106203. https://doi.org/10.1016/j.scs.2025.106203

[27] U. Ali, M.H. Shamsi, C. Hoare, E. Mangina, J. O'Donnell, Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis, Energy Build. 246 (2021) 111073. https://doi.org/10.1016/j.enbuild.2021.111073

[28] C. Cerezo Davila, C.F. Reinhart, J.L. Bemis, Modeling boston: a workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets, Energy 117 (2016) 237–250. https://doi.org/10.1016/j.energy.2016.10.057

[29] A. Malhotra, J. Bischof, A. Nichersu, K.-H. Hafele, J. Exenberger, D. Sood, J. Allan, J. Frisch, C. van Treeck, J. O'Donnell, G. Schweiger, Information modelling for urban building energy simulation—A taxonomic review, Build. Environ. 208 (2022) 108552. https://doi.org/10.1016/j.buildenv.2021.108552

[30] C. Wang, M. Ferrando, F. Causone, X. Jin, X. Zhou, X. Shi, Data acquisition for urban building energy modeling: a review, Build. Environ. 217 (2022) 109056. https://doi.org/10.1016/j.buildenv.2022.109056

[31] G. Jiang, Z. Ma, L. Zhang, J. Chen, EPlus-LLM: a large language model-based computing platform for automated building energy modeling, Appl. Energy 367 (2024) 123431. https://doi.org/10.1016/j.apenergy.2024.123431

[32] S. Choi, S. Yoon, GPT-based data-driven urban building energy modeling (GPT-UBEM): concept, methodology, and case studies, Energy Build. 325 (2024) 115042. https://doi.org/10.1016/j.enbuild.2024.115042

[33] D. Kong, A. Cheshmehzangi, Z. Zhang, S.P. Ardakani, T. Gu, Urban building energy modeling (UBEM): a systematic review of challenges and opportunities, Energy Effic. 16 (6) (2023) 69. https://doi.org/10.1007/s12053-023-10147-z

[34] N. Buckley, G. Mills, C. Reinhart, Z.M. Berzolla, Using urban building energy modelling (UBEM) to support the new European union's green deal: case study of Dublin Ireland, Energy Build. 247 (2021) 111115. https://doi.org/10.1016/j.enbuild.2021.111115

[35] A. Jalilzadeh, A. Rafiee, S. De Graaf, T. Konstantinou, P.J.M. Van Oosterom, Automated district-level energy demand modeling using EnergyPlus empowered by digital twin technology, in: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. X-G-2025, 2025, pp. 397–404. https://doi.org/10.5194/isprs-annals-X-G-2025-397-2025

[36] G. Agugiaro, J. Benner, P. Cipriano, R. Nouvel, The energy application domain extension for cityGML: enhancing interoperability for urban energy simulations, Open Geospat. Data Softw. Stand. 3 (2018) 2. https://doi.org/10.1186/s40965-018-0042-y

[37] Z. Yao, C. Nagel, F. Kunde, G. Hudra, P. Willkomm, A. Donaubauer, T. Adolphi, T.H. Kolbe, 3DCityDB - a 3D geodatabase solution for the management, analysis, and visualization of semantic 3D city models based on CityGML, Open Geospat. Data Softw. Stand. 3 (2018) 5. https://doi.org/10.1186/s40965-018-0046-7

[38] R. Kaden, T.H. Kolbe, City-wide total energy demand estimation of buildings using semantic 3D city models and statistical data, ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. II-2/W1 (2013) 163–171. https://doi.org/10.5194/isprsannals-II-2-W1-163-2013

[39] T. Dogan, C. Reinhart, Shoeboxer: an algorithm for abstracted rapid multi-zone urban building energy model generation and simulation, Energy Build. 140 (2017) 140–153. https://doi.org/10.1016/j.enbuild.2017.01.030

[40] R. Nouvel, K.-H. Brassel, M. Bruse, E. Duminil, V. Coors, U. Eicker, D. Robinson, SimStadt, a new workflow-driven urban energy simulation platform for CityGML city models, in: Proceedings of International Conference CISBAT 2015 Future Buildings and Districts Sustainability from Nano to Urban Scale, LESO-PB, EPFL, 2015, pp. 889–894. https://doi.org/10.5075/epfl-cisbat2015-889-894

[41] C. Reinhart, T. Dogan, A. Jakubiec, T. Rakha, A. Sang, Umi – an urban simulation environment for building energy use, daylighting and walkability, in: Proceedings of Building Simulation 2013: 13th Conference of IBPSA, vol. 13 of *Building Simulation*, IBPSA, Chambery, France, 2013, pp. 476–483. https://doi.org/10.26868/25222708.2013.1404

[42] RVO, Voorbeeldwoningen 2022: Woningkenmerken en Voorbeeldberekeningen, Technical Report, Rijksdienst voor Ondernemend Nederland, The Hague, Netherlands, 2022. Reference dwelling dataset containing 51 example dwellings with calculated energy performance..

[43] H. Jia, A. Chong, Eplusr: a framework for integrating building energy simulation and data-driven analytics, Energy Build. 237 (2021) 110757. https://doi.org/10.1016/j.enbuild.2021.110757

[44] ASHRAE, Guideline 14–2014: Measurement of Energy, Demand, and Water Savings, 2014. https://www.ashrae.org/technical-resources/standards-and-guidelines.

[45] A. Oraiopoulos, B. Howard, On the accuracy of urban building energy modelling, Renew. Sustain. Energy Rev. 158 (2022) 111976. https://doi.org/10.1016/j.rser.2021.111976

[46] P. Fennell, S. Ebrahimigharehbaghi, E. Mata, G. Kokogiannakis, S. Amrith, S. Ignatiadou, S. Lo Piano, Urban building energy models: how can we improve the treatment of uncertainty for energy policy decision-making?, Environ. Res. Commun. 7 (2) (2025) 022002. https://doi.org/10.1088/2515-7620/ad9438

[47] CBS, Energielevering aan woningen en bedrijven naar postcode, 2023, https://www.cbs.nl/nl-nl/maatwerk/2023/46/energielevering-aan-woningen-

en-bedrijven-naar-postcode. Average natural gas and electricity deliveries per PC6 postcode area, 2014–2022; accessed 2025.
[48] B. Dukai, H. Ledoux, J. Stoter, 3D Registration of Buildings and Addresses (BAG) 2020, 2020. https://doi.org/10.4121/13088933.v1
[49] Rijksdienst voor Ondernemend Nederland (RVO), Actualisatie 2025 bouw- en investeringskosten verduurzamingsmaatregelen (prijspeil mei 2025), 2025, https://regelhulpenvoorbedrijven.nl/kostenkentallen/. Cost benchmarks for sustainability measures in residential and utility buildings, price level May 2025.
[50] M. Sunikka-Blank, R. Galvin, Introducing the prebound effect: the gap between performance and actual energy consumption, Build. Res. Inf. 40 (3) (2012) 260–273. https://doi.org/10.1080/09613218.2012.690952
[51] RVO, CO_2-emissiefactoren, 2024, https://www.co2emissiefactoren.nl/. 2024 Dutch operational emission factors: natural-gas combustion 1.779 kg CO_2e/Nm^3 and grid-electricity production 0.270 kg CO_2e/kWh.
[52] European Commission, Commission delegated regulation (eu) no 244/2012 supplementing directive 2010/31/EU by establishing a comparative methodology framework for calculating cost-optimal levels of minimum energy performance requirements for buildings and building elements, 2012. Official Journal of the European Union, L 81, 21.3.2012, pp. 18–36; CELEX 32012R0244. Repealed with effect from 31 December 2025 by Commission Delegated Regulation (EU) 2025/2273, https://eur-lex.europa.eu/eli/reg_del/2012/244/oj/eng.
[53] CEN, EN 15459-1:2017 energy performance of buildings – economic evaluation procedure for energy systems in buildings – Part 1: calculation procedures, Module M1-14, 2017. European Committee for Standardization, Brussels; published 2017-05-31, https://knowledge.bsigroup.com/products/energy-performance-of-buildings-economic-evaluation-procedure-for-energy-systems-in-buildings-calculation-procedures-module-m1-14.
[54] R. Galvin, Deep energy efficiency renovation of Germany's residential buildings: is this as economically viable as Germany's policymakers and popular promoters often claim?, Energy Effic. 17 (2024) 47. https://doi.org/10.1007/s12053-024-10227-8
[55] IEA EBC, Annex 75: cost-effective building renovation at district level combining energy efficiency and renewables, Technical Report, International Energy Agency, 2023. Project Summary Report, published 2023, https://annex75.iea-ebc.org/Data/publications/EBC_PSR_Annex75.pdf.
[56] L.E. Excell, A. Nutkiewicz, R.K. Jain, Multi-scale retrofit pathways for improving building performance and energy equity across cities: a UBEM framework, Energy Build. 324 (2024) 114931. https://doi.org/10.1016/j.enbuild.2024.114931
[57] S.R. Penaka, K. Feng, T. Olofsson, A. Rebbling, W. Lu, Improved energy retrofit decision making through enhanced bottom-up building stock modelling, Energy Build. (2024) 114378. https://doi.org/10.1016/j.enbuild.2024.114378
[58] F. Kesicki, P. Ekins, Marginal abatement cost curves: a call for caution, Clim. Policy 12 (2) (2012) 219–236. https://doi.org/10.1080/14693062.2011.582347
[59] PBL, Klimaat- en energieverkenning 2023, Technical Report, Planbureau voor de Leefomgeving, The Hague, Netherlands, 2023. Annual climate and energy outlook for the Netherlands, https://www.pbl.nl/publicaties/klimaat-en-energieverkenning-2023.
[60] J. Sokol, C. Cerezo Davila, C.F. Reinhart, Validation of a Bayesian-based method for defining residential archetypes in urban building energy models, Energy Build. 134 (2017) 11–24. https://doi.org/10.1016/j.enbuild.2016.10.050
[61] F. Rehmann, M. Mosteiro-Romero, C. Miller, R. Streblow, Enhancing urban energy modeling: a case study of data acquisition, enrichment, and evaluation in Berlin, Energy Build. 346 (2025) 116070. https://doi.org/10.1016/j.enbuild.2025.116070
[62] O. Vera-Piazzini, M. Scarpa, Building energy model calibration: a review of the state of the art in approaches, methods, and tools, J. Build. Eng. 86 (2024) 108287. https://doi.org/10.1016/j.jobe.2023.108287
[63] P. de Wilde, The gap between predicted and measured energy performance of buildings: a framework for investigation, Autom. Constr. 41 (2014) 40–49. https://doi.org/10.1016/j.autcon.2014.02.009
[64] Y.Q. Ang, Z.M. Berzolla, S. Letellier-Duchesne, C.F. Reinhart, Carbon reduction technology pathways for existing buildings in eight cities, Nat. Commun. 14 (2023) 1689. https://doi.org/10.1038/s41467-023-37131-6
[65] F. Johari, O. Lindberg, U.H. Ramadhani, F. Shadram, J. Munkhammar, J. Widén, Analysis of large-scale energy retrofit of residential buildings and their impact on the electricity grid using a validated UBEM, Appl. Energy 361 (2024) 122937. https://doi.org/10.1016/j.apenergy.2024.122937
[66] S. Zhang, L. Chen, Q. Huang, X. Wang, Z. Wang, An integrated energy–carbon–economic assessment of building decarbonization at city scale, Appl. Energy 424 (2026) 128406. https://doi.org/10.1016/j.apenergy.2026.128406
[67] T. Hong, S.H. Lee, W. Zhang, H. Li, K. Sun, J. Kace, Informing electrification strategies of residential neighborhoods with urban building energy modeling, Build. Simul. 18 (1) (2025) 3–18. https://doi.org/10.1007/s12273-024-1214-6
[68] T. Loga, B. Stein, N. Diefenbach, TABULA building typologies in 20 European countries — making energy-related features of residential building stocks comparable, Energy Build. 132 (2016) 4–12. https://doi.org/10.1016/j.enbuild.2016.06.094
[69] M. Manfren, B. Nastasi, Parametric performance analysis and energy model calibration workflow integration-a scalable approach for buildings, Energies 13 (3) (2020). https://doi.org/10.3390/en13030621
[70] J. Zhang, N. Liu, S. Wang, A parametric approach for performance optimization of residential building design in Beijing, Build. Simul. 13 (2) (2020) 223–235. https://doi.org/10.1007/s12273-019-0571-z
[71] D. Zendehdel, G. Ferro, E. De Santis, A. Rizzi, Degradation-aware energy management in residential microgrids: a reinforcement learning framework, in: Proceedings of the 17th International Joint Conference on Computational Intelligence (IJCCI 2025), SCITEPRESS – Science and Technology Publications, Marbella, Spain, 2025. DOI to be added once published in the SCITEPRESS digital library.
[72] D. Zendehdel, E. De Santis, A. Capillo, P. Odonkor, A. Rizzi, Multi-objective battery dispatching using an enhanced SAC algorithm, in: Proceedings of the 2025 International Joint Conference on Neural Networks (IJCNN), 2025, pp. 1–8. https://doi.org/10.1109/IJCNN60510.2025.11227829
[73] Nieman Raadgevende Ingenieurs, Rapport standaard en streefwaardes bestaande woningbouw, Technical Report 20190115/15645, Rijksdienst voor Ondernemend Nederland, 2021. Commissioned by RVO; defines insulation standard and target values for existing Dutch residential buildings, https://open.overheid.nl/repository/ronl-63c7a9e9-12f8-4ca5-9a97-66c3c9e68baa/1/pdf/rapport-standaard-en-streefwaarden-bestaande-woningbouw-.
[74] NEN, NTA 8800: energieprestatie van gebouwen – bepalingsmethode, 2020, (Nederlandse Norm). Dutch energy performance calculation standard, effective January 2021, replacing NEN 7120, https://www.nen.nl/nta-8800-2020-nl-272896.
[75] R. Peters, B. Dukai, S. Vitalis, J. van Liempt, J. Stoter, Automated 3D reconstruction of LoD2 and LoD1 models for all 10 million buildings of the Netherlands, Photogramm. Eng. Remote Sens. 88 (3) (2022) 165–170. https://doi.org/10.14358/PERS.21-00032R2
[76] Liander, Open data kleinverbruik: verbruiksgegevens gas en elektriciteit, 2025, https://www.liander.nl/over-ons/open-data. Postcode-level annual gas and electricity consumption for small connections, updated annually.
[77] J.M. Sipma, Verrijkte BAG ter ondersteuning van lokale energetische vraagstukken, Technical Report TNO 2023 P10648, TNO, 2023. Enriched BAG dataset with energy-relevant building attributes, commissioned by RVO, https://resolver.tno.nl/uuid:7007a888-bde5-4b5a-8cb3-4ef3e53b557f.
[78] RVO, EP-Online: energieprestatiecertificaten register, 2024, https://www.ep-online.nl/. Public register of energy performance certificates for all Dutch buildings under NTA 8800.
[79] Enexis Netbeheer, Open data kleinverbruiksgegevens, 2025, https://www.enexis.nl/over-ons/wat-bieden-we/open-data. Postcode-level annual gas and electricity consumption for small connections.
[80] KNMI, Uurgegevens van het weer in Nederland, 2024, https://www.knmi.nl/nederland-nu/klimatologie/uurgegevens. Hourly weather data; Schiphol (station 240) used to generate EnergyPlus EPW file.
[81] O. Guerra-Santin, S. Silvester, Development of Dutch occupancy and heating profiles for building simulation, Build. Res. Inf. 45 (4) (2017) 396–413. https://doi.org/10.1080/09613218.2016.1160563
[82] International Organization for Standardization, ISO 10456:2007 building materials and products – hygrothermal properties – tabulated design values and procedures for determining declared and design thermal values, Technical Report, ISO, Geneva, Switzerland, 2007. Design thermal conductivity and resistance values for building materials, https://www.iso.org/standard/40966.{html}.
[83] DGMR, BENG referentiegebouwen, Technical Report E.2015.1371.00.R001, Rijksdienst voor Ondernemend Nederland, 2016. Nearly Zero Energy Building reference models for Dutch building code compliance, corrected November 2017, https://www.rvo.nl/sites/default/files/2017/02/Referentiegebouwen%20BENG.pdf.
[84] Rijksoverheid, Bouwbesluit 2012, 2012, https://rijksoverheid.nl/onderwerpen/bouwregelgeving/bouwbesluit-2012. Dutch building code; defines minimum airtightness and insulation requirements.