Applied Energy 427 (2027) 129270

Contents lists available at ScienceDirect

Applied Energy

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

Decarbonizing residential buildings: Evaluating heat pump solar and battery renovation pathways

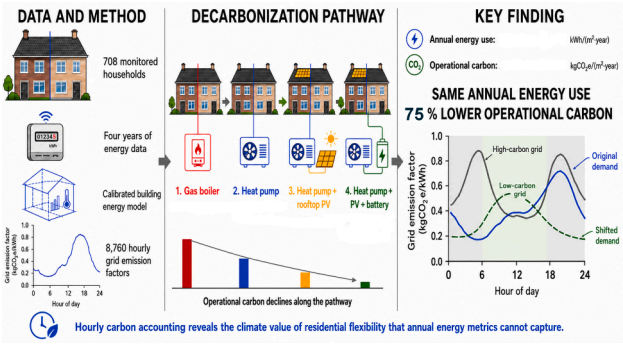
Marzyeh Khoshroo [a], Deepak Amaripadath [b], Tarek M.F. Ahmed [c], Shady Attia [a,*]

[a] Sustainable Building Design Lab, Department of UEE, Faculty of Applied Sciences, University of Liege, Liege, Belgium
[b] Building Energy Monitoring and Simulation, UR SPHERES, Department of Environmental Sciences and Management, University of Liege, Arlon, Belgium
[c] Environment, Northumbria University, Newcastle upon Tyne, UK

HIGHLIGHTS

- Heat-pump electrification cuts dynamic operational carbon by 67%.
- PV and battery storage reduce grid imports from 40.0 to 33.2 kWh/(m^2·a).
- The full renovation pathway lowers operational carbon by 75%.
- Net annual EUI remains unchanged across the three PV-storage scenarios.
- The temporal carbon performance gap widens from 3.0% to 30.3%.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Temporal carbon performance gap
Operational carbon emissions
Dynamic emission factors
Grid interaction
Heat pump electrification
North-Western Europe

ABSTRACT

Residential decarbonization requires renovation pathways that combine heating electrification, on-site renewable generation, and energy storage, yet their performance is commonly assessed using annual energy and carbon indicators that neglect the timing of grid interactions. This study evaluates five progressive renovation configurations for a representative Belgian semi-detached dwelling: a gas-boiler base case, heat-pump electrification, heat pump with rooftop photovoltaics (PV), and PV systems combined with 2 and 5 kWh battery storage. The benchmark dwelling was selected from 708 audited households and calibrated against four years of measured electricity and gas consumption, achieving monthly ASHRAE Guideline 14 criteria for both carriers. Operational carbon was evaluated using both the annual-average and measured hourly Belgian grid emission factors. Replacing the gas boiler with a reversible heat pump reduced gross delivered energy from 102.2 to 48.5 kWh/(m^2·a) and dynamic operational carbon from 18.46 to 6.14 kgCO$_2$e/(m^2·a). Adding PV and progressively larger batteries further reduced grid imports from 40.0 to 33.2 kWh/(m^2·a) and dynamic operational carbon to 4.64 kgCO$_2$e/(m^2·a), corresponding to a 75% reduction relative to the base case. However, annual indicators increasingly failed to represent these improvements. Net annual energy use remained identical at 31.7 kWh/(m^2·a) across all three PV configurations despite substantial reductions in grid imports and carbon emissions. Moreover, the discrepancy between annual-average and hourly carbon accounting increased from 3.0% for the base case to 30.3% for the heat-pump, PV, and 5 kWh battery configuration. This widening temporal carbon

* Corresponding author.
E-mail address: shady.attia@uliege.be (S. Attia).

https://doi.org/10.1016/j.apenergy.2026.129270
Received 19 August 2026; Received in revised form 6 October 2026; Accepted 8 November 2026
0306-2619/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

performance gap results from residual-import carbon intensification, as PV preferentially displaces cleaner midday electricity while remaining imports become concentrated in carbon-intensive winter and evening periods. The findings demonstrate that effective residential renovation assessment should evaluate not only annual energy reductions but also the timing and carbon intensity of grid interactions, particularly as buildings transition toward electrification, distributed generation, and storage.

Nomenclature

Symbols

A	conditioned floor area, m^2
E	delivered energy, kWh
EF	emission factor, $kgCO_2e/kWh$
g	solar factor of glazing,
n50	air change rate at 50 Pa, h^{-1}
OC	operational carbon, $kgCO_2e/(m^2 \cdot a)$
t	hour index (1 to 8760), –
U	thermal transmittance, $W/(m^2 \cdot K)$
η	seasonal conversion efficiency, –

Subscripts

a	annual
dyn	dynamic (hourly emission factors)
exp.	exported to the grid
gas	natural gas
imp	imported from the grid
net	net of on-site generation
stat	static (annual mean emission factor)
t	hour of the year

Abbreviations

ASHRAE	American Society of HeatingRefrigerating and Air-Conditioning Engineers
COP	coefficient of performance
CV(RMSE)	coefficient of variation of the root mean square error
DHW	domestic hot water
EPBD	Energy Performance of Buildings Directive
EPC	energy performance certificate
EPW	EnergyPlus weather file
EUI	energy use intensity
GHG	greenhouse gas
HDD	heating degree days
HP	heat pump
IQR	interquartile range
LCA	life cycle assessment
LHV	lower heating value
LWEF	load-weighted emission factor
MBE	mean bias error
NZEB	nearly zero-energy building
PEB	Performance Énergétique des Bâtiments (Belgian energy performance certification)
PV	photovoltaic
Q1Q3	first and third quartile
SOC	state of charge
SPF	seasonal performance factor
TMY	typical meteorological year
kWp	kilowatt peak (photovoltaic rated capacity)

1. Introduction

1.1. A shared building stock and a shared heating problem

Housing accounts for roughly 20% of energy consumption in the European Union [1], and the bloc has committed to cutting emissions by 55% by 2030 and reaching climate neutrality by 2050 [2]. The recast Energy Performance of Buildings Directive turns those commitments into obligations. It requires zero-emission standards for new buildings by 2028 and a staged renovation programme for the existing stock [3]. Maduta et al. [4] trace how the definition has moved from nearly zero-energy to zero-emission buildings and note that the transition shifts the burden of proof from energy to carbon. Rosenow [5] argues that demand-side measures carry more of that burden than supply-side policy usually assumes.

A substantial share of the Western European suburban housing stock consists of low-rise semi-detached and terraced dwellings with characteristics relevant to the renovation challenge examined here. Houses built in the two decades either side of 2000 are widespread across Belgium, the Netherlands, the United Kingdom and Ireland, as shown in Fig. 1. They share a construction period, a form, a floor area and, most importantly here, a heating system: a gas boiler serving both space heating and domestic hot water. Belgium alone has 64,535 of them from the 2000 to 2010 cohort [6], shown in Fig. 2. They are recent enough to have been built under energy regulation, old enough to fall inside the renovation wave, and numerous enough to matter to the national inventory.

Attia et al. [7] set out the design and compliance difficulties that arise when a stock built to one regulatory logic is asked to meet another, and those difficulties are now arriving in the heating system itself. Heat pump uptake in Belgium remains modest but is growing under policy support [8], so the cohort described above is precisely the group that will electrify over the next decade.

These dwellings also share a compliance regime. Performance is judged on annual primary energy intensity in kWh/(m2·a). That indicator was built for a stock heated by fossil fuel, where electricity was a small share of the total, and the building only ever drew from the grid. Under those conditions it carried one assumption nobody needed to question, namely that a kilowatt-hour is a kilowatt-hour whenever it is used.

1.2. What annual accounting cannot see

Electrification breaks that assumption. Once heating, hot water and cooling run on electricity, and once the dwelling also generates, its carbon outcome is tied to the hourly state of the power system rather than to its annual demand. In Belgium, a large low-carbon baseload holds the floor down, wind and solar vary, and gas plants fill the gap, so in 2024 the dirtiest hours carried roughly four times the intensity of the cleanest [9]. Absolute levels differ across the region [10], and Unnewehr et al. [11] show that open data now make comparable hourly intensity signals available for most European countries. The shape is shared wherever solar is significant: a midday trough with a morning and an evening peak.

The consequences for building assessment are established. Amaripadath et al. [12] found that nearly zero-energy buildings can still carry a substantial carbon footprint when demand lands in high-emission

periods, so a net zero energy building is not automatically a net zero carbon one. Bontekoe et al. [13] showed that choosing annual rather than hourly emission factors changes the retrofit conclusion by 25 to 50%, with the annual figure understating operational carbon. Zou et al. [14] reached the same conclusion for commercial buildings and showed that monthly factors do not repair the problem. Lou et al. [15] extended the argument over time, finding that retrofit rankings shift as the grid decarbonises, and Li et al. [16] showed that even envelope design choices change when dynamic factors are applied. Alvarez Florez et al. [17] set out hourly methods for assessing carbon and primary energy together in decarbonised buildings, and Bertini et al. [18] showed how strongly whole-life carbon results depend on the assumptions carried into a residential renovation.

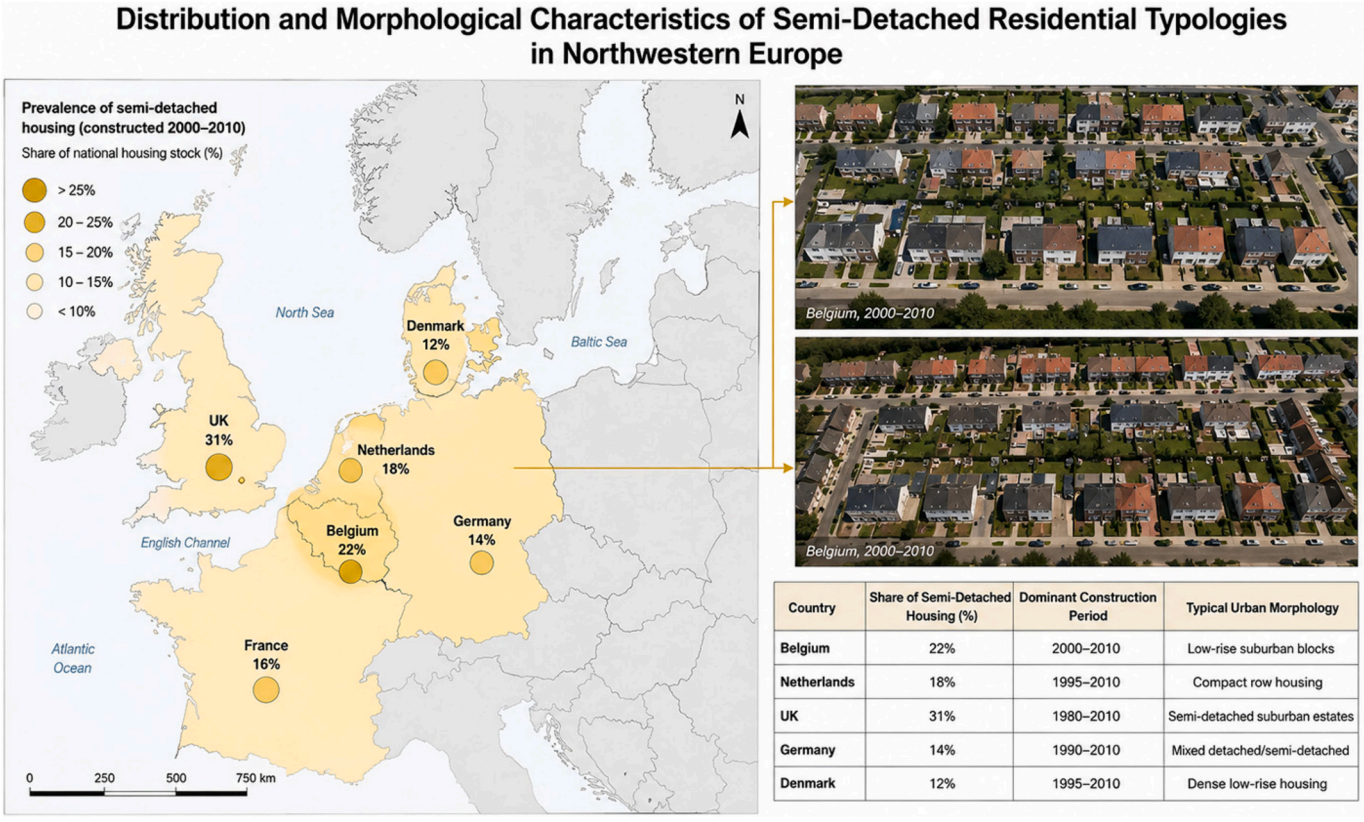
All of this was established for buildings that only consume. Rooftop solar changes the question, because it does not shave demand evenly. It removes midday imports in the sunnier half of the year and leaves mornings, evenings, and winter untouched. If midday electricity is cleaner than evening electricity, the imports that remain are more carbon-intensive than the ones displaced, and an annual average no longer describes them. Evidence for this has begun to gather at system scale. Brodnicke et al. [19] used optimisation to show that a constant annual factor can overestimate the abatement from building solar by a factor of two. Bistline and Watten [20] reached a similar conclusion for the United States using power sector capacity models. Both work on power systems. Neither describes what an energy performance certificate reports for a single dwelling, which is where the accounting rule is actually applied.

1.3. Approach and contribution

This study evaluates the energy and operational-carbon performance of progressive residential renovation pathways combining heating electrification, rooftop solar and battery storage. A benchmark dwelling was drawn from 708 households audited between 2021 and 2024, sits at the median of that sample, and was calibrated against four consecutive years of metered gas and electricity. This measured-data foundation follows the evidence of Bayer and Pruckner [21], who showed that data-driven assessment of heat-pump retrofits can produce materially different outcomes from assumption-driven approaches. Five configurations were assessed sequentially: the existing gas-boiler system, heat-pump electrification, heat pump with rooftop PV, and the same system with two battery capacities. For each configuration, the analysis quantifies delivered energy, grid imports and operational carbon, allowing the contribution of each successive renovation measure to be identified. Because electrification, PV and storage progressively alter when the dwelling interacts with the grid, operational carbon was additionally calculated using both measured hourly Belgian grid carbon intensity and its annual mean. This comparison tests whether conventional annual indicators continue to represent the carbon performance of increasingly grid-interactive renovation pathways. Two concepts are introduced to interpret how annual accounting represents these renovation outcomes. The temporal carbon performance gap is the difference between the two carbon calculations, expressed as a percentage of the annual result. Residual-import carbon intensification describes the mechanism behind that gap: behind-the-meter generation lowers the quantity of electricity imported while increasing the average carbon intensity of the imports that remain.

The gap grew along the retrofit pathway, from 3.0% for the gas boiler to 30.3% once solar and a 5 kWh battery were added. Over the same three solar configurations, the net annual energy use intensity stayed identical at 31.7 kWh/(m2·a) while grid imports fell 17% and operational carbon fell 14%. The carbon intensity of imported electricity rose from 0.1261 to 0.1394 kgCO2e/kWh, against an annual mean of 0.107.

In this research, four contributions follow. First, the study provides a



Country	Share of Semi-Detached Housing (%)	Dominant Construction Period	Typical Urban Morphology
Belgium	22%	2000–2010	Low-rise suburban blocks
Netherlands	18%	1995–2010	Compact row housing
UK	31%	1980–2010	Semi-detached suburban estates
Germany	14%	1990–2010	Mixed detached/semi-detached
Denmark	12%	1995–2010	Dense low-rise housing

Fig. 1. Distribution and representative morphology of semi-detached residential typologies constructed between 2000 and 2010 across Northwestern Europe. Left: prevalence of semi-detached housing by country. Right: representative Belgian suburban developments illustrating typical low-rise residential morphology.

calibrated benchmark for a widespread Western European suburban housing typology, checked against four years of measured consumption. Second, it quantifies the progressive energy, grid-import, and operational-carbon effects of a renovation pathway from gas heating through heat-pump electrification to rooftop PV and battery storage. Third, it demonstrates that the ability of annual indicators to represent these improvements deteriorates as the dwelling becomes more grid interactive, quantified through the temporal carbon performance gap and explained by residual-import carbon intensification. Fourth, it provides a verification approach for comparing the emission factor implied by each renovation step against the observed distribution of hourly grid factors and identifies a pathway toward time-resolved certification using interval-metering data.

The scope of the renovation pathway is deliberately limited to building energy-system interventions. The study evaluates the transition from gas heating to heat-pump electrification, followed by rooftop PV and battery storage, while retaining the calibrated building envelope, geometry and occupancy conditions across scenarios. It therefore does not assess whole-building deep renovation packages involving insulation, airtightness, glazing or other envelope measures. The assessment is likewise limited to energy and operational-carbon performance. The embodied carbon of the installed systems is not included (Section 5.4), and investment costs, energy prices, payback periods and life-cycle economic performance are outside the study boundary. The scenarios should consequently be interpreted as energy-system decarbonization pathways rather than techno-economic optimization or comprehensive deep-renovation packages.

One boundary should be stated at the outset. This study uses hourly average emission factors, which are the correct basis for attributional reporting and therefore for a certificate. They do not measure the marginal system emissions that a retrofit causes or avoids, which is a different question requiring different factors. Ben Amor et al. [22] set out methods for estimating hourly marginal factors, and Goldsworthy and Aryai [23] compare time-varying average and marginal factors in building analysis and show that the two can point to different conclusions. All results reported here are attributional.

2. Literature review

2.1. From annual efficiency to temporally resolved carbon

Building assessment in Europe was designed for a different problem from the one it now faces. When the framework took shape, heating was mostly fossil-fuelled, electricity was a small share of residential energy, buildings drew from the grid and never fed it, and annual demand was a reasonable summary of how a building performed. Under those conditions, the choice of an annual indicator was sound. It carried one assumption that nobody needed to question, namely that a kilowatt-hour is a kilowatt-hour whenever it is used.

Electrification breaks that assumption. Once heating, hot water, and cooling run on electricity, and once the building also generates, its carbon outcome is tied to the hourly state of the power system rather than to its annual demand. Zou et al. [14] showed for commercial buildings that the temporal resolution of the emission factor changes the carbon result materially, and that annual factors misstate it in ways monthly factors do not repair. Bontekoe et al. [13] reached the same conclusion for retrofit assessment, reporting differences of 25 to 50% between annual and hourly accounting.

One distinction matters throughout and is often blurred. Average emission factors describe the mean carbon content of electricity supplied at a given moment and are the appropriate basis for attributional accounting, which is what a certificate does. Marginal factors describe the emissions of the generator that to a change in load and answer a consequential question, namely what a retrofit causes the system to do. Ben Amor et al. [22] set out methods for estimating hourly marginal factors and show they behave differently from average factors. Gallardo et al. [24] review the whole family of choices and argue for matching the factor to the decision rather than treating one as generally superior. The present study uses hourly average factors and makes attributional claims only.

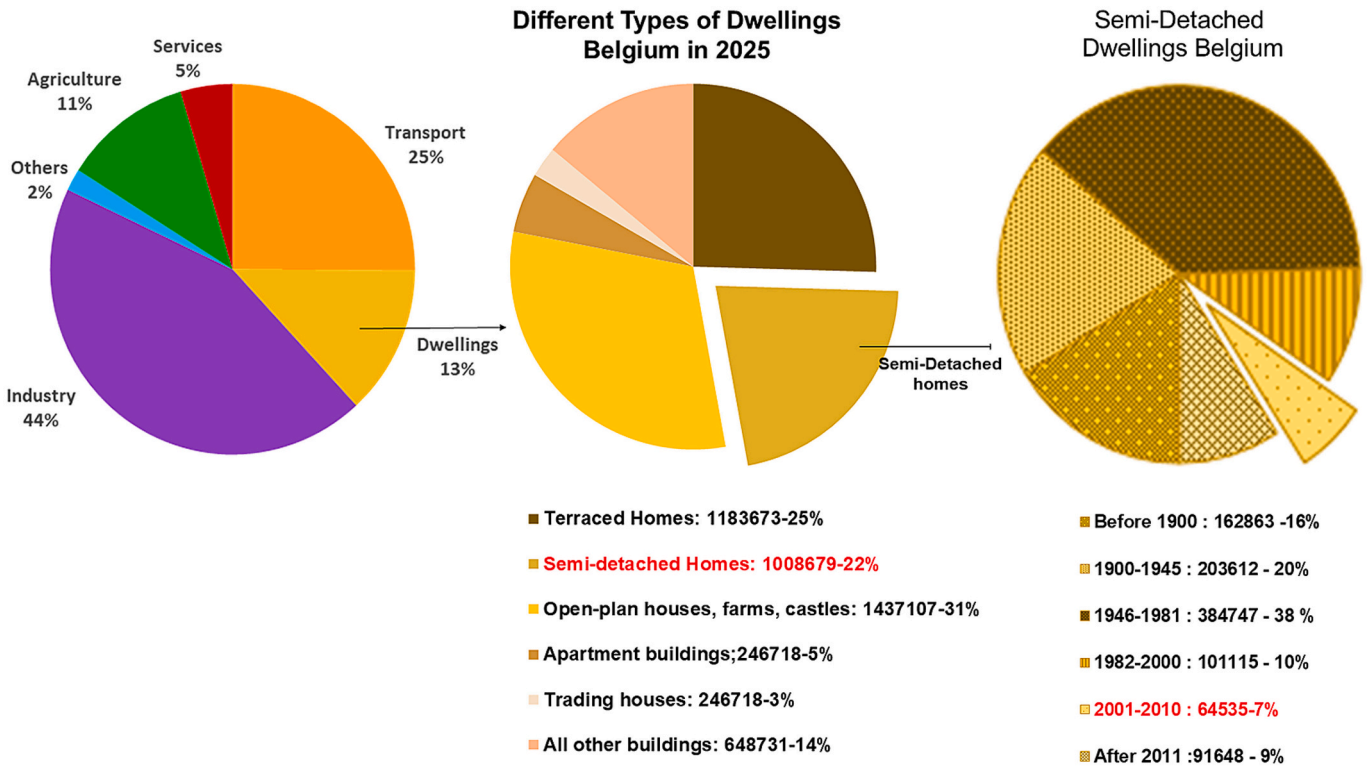


Fig. 2. Sectoral greenhouse gas emissions and distribution of the Belgian residential building stock by dwelling type and construction period in 2025.

The temporal question sits on top of an older one. Calculated ratings and measured consumption diverge widely for the same dwelling, and the reasons are well documented. Heo et al. [25] showed that calibration under uncertainty changes the ranking of retrofit options, so the starting point of any assessment is itself uncertain. Greer et al. [26] examined how operational emissions are handled inside whole-building life cycle assessment and found that method choices dominate the result.

The Dutch stock provides the clearest evidence on scale. Majcen et al. [27] compared certificates against metered consumption for roughly 200,000 dwellings and found that poorly labelled homes use much less energy than predicted while efficient homes use more, so the error does not merely scatter; it changes sign across the label bands. The same authors traced the causes in gas consumption specifically [28] and then modelled the heating prediction gap against building, household and behavioral characteristics [29]. Van den Brom et al. [30] showed that the gap varies systematically with household type as well as with the dwelling, and a later study of the Dutch non-profit stock found that empirical models estimate actual consumption better than the theoretical calculation does [31]. Taken together, these results mean that an annual certificate figure is already an uncertain quantity before any emission factor is applied to it. If the energy figure is uncertain and the factor is applied at the wrong resolution, the two errors do not cancel.

2.2. Electrification and the timing of residential loads

A heat pump does not simply reduce delivered energy. It converts a fuel demand that was invisible to the power system into an electrical demand with a definite shape. Fischer and Madani [32] reviewed heat pumps in smart grids and describe that shape clearly: concentrated in winter, with a morning recovery peak and an evening occupancy peak. Those are the hours when many European grids are at their most carbon-intensive, so the same annual kilowatt-hours carry different carbon depending on when the pump runs.

How much of that shape can be moved depends on the signal the household receives. Wilczynski et al. [33] modelled residential heat pumps under four tariff designs, from a flat rate to real-time pricing, and found the largest difference in activated flexibility between the flat and the real-time cases. Their result matters here for two reasons. It shows that the timing of heat pump electricity is genuinely movable rather than fixed by the weather, and it shows that what moves it is a price signal rather than a carbon signal. The two coincide only when the tariff happens to track carbon intensity, which is not generally the case.

Cooling behaves oppositely to heating. It is a summer load that peaks in the afternoon, which places it in the hours when solar output is high, and grid carbon intensity is usually low. Wang et al. [34] examined electrified communities and found that the energy, emission, and grid consequences of electrification depend on the combined shape of these loads rather than on their totals. For the dwellings considered here, the point is practical. Heating pulls demand into the dirtiest hours and cooling pushes it into the cleanest, so a reversible heat pump changes the carbon-weighted timing of the building in two directions at once.

That second load is growing. Dwellings in Central and Northern Europe were designed to keep heat in rather than out, and a recent multi-author review sets out how overheating arises in that stock and what it means for occupants and energy use [35]. Cooling degree days across the European Union have risen steadily since 1970 while heating degree days have fallen [36], and tools now exist to estimate community-scale cooling demand under a warming climate [37]. Rahif et al. [38] quantified how well different cooling strategies resist climate change overheating in a set of locations including Brussels, which is the climate of the benchmark used here.

Thermal storage adds a further degree of freedom. Reynders et al. [39] developed a general method for characterising the flexibility available in building thermal mass and storage, and showed that its usefulness depends on how long energy can be held. That horizon matters later in this article, because a store that shifts energy within a day cannot address a carbon gradient that runs between seasons.

The carbon consequences of heat pump adoption also depend on what is counted. Fadden et al. [40] reviewed life cycle impacts across residential and district heat pump systems and found wide variation driven by system boundary and by the assumed grid. This study takes the operational boundary only, which is what a certificate reports, and treats manufacturing and refrigerant effects as outside its scope.

2.3. Solar, self-consumption and the residual load

Studies of building-integrated solar conventionally report annual generation, self-consumption, self-sufficiency, imports, exports and an annual carbon saving. Salom et al. [41] set out load-matching and grid-interaction indicators that made these quantities comparable between studies, and they remain the standard vocabulary. What that vocabulary does not describe is which grid electricity the generation replaces.

This is a selection effect, and it is the heart of the problem. Rooftop solar does not remove a representative sample of a household's imports. It removes midday imports in the sunnier half of the year and leaves everything else untouched. If midday electricity is cleaner than evening electricity, then the imports that remain are more carbon-intensive than the ones that were displaced, and the average factor computed over the year no longer describes them.

Evidence for this has begun to accumulate at system scale. Brodnicke et al. [19] used optimisation with dynamic factors and export limits and found that a constant annual factor can overestimate the abatement from building solar by a factor of two. Bistline and Watten [20] reached a similar conclusion for the United States using capacity planning models, concluding that rooftop solar emission reductions are overstated when substitution effects are handled poorly. Cho and Honda [42] examined electrified detached houses with solar generation using an hourly emission profile and confirmed that the result depends on when the generation lands relative to demand.

The regulatory framing has moved in the same direction. Paulus and Borggrefe [43] showed that the value of shifting demand depends on when the shift happens rather than how much is shifted, which is the price analogue of the carbon argument made here. Support schemes across Northwestern Europe are now withdrawing one-to-one net metering, and an analysis of the Dutch scheme sets out both the distributional consequences and the reason for the change, namely that netting gives no incentive to self-consume or to store [44]. Once export is valued below import, the quantity that determines cost and carbon alike becomes the import profile rather than the annual balance. System studies have begun to design around that, pairing heat pumps with solar at settlement scale and sizing for self-consumption rather than for maximum generation [45].

These studies work at the level of a power system, a market or an optimisation problem. None of them describes what an energy performance certificate reports for a single dwelling, which is where the accounting rule is actually applied and where its consequences are felt by an owner.

2.4. Storage: Self-consumption or carbon-responsive operation

Residential batteries can be operated under three distinct control objectives that should be differentiated because they produce different energy, economic and carbon outcomes. Self-consumption control charges the battery from local PV surplus to reduce exports and subsequent grid imports, while cost-based control responds to electricity prices to minimize household expenditure. Carbon-aware control instead responds to time-varying grid carbon intensity and is explicitly designed to minimize emissions. These objectives do not necessarily

coincide, and the resulting battery operation can therefore differ substantially even for the same hardware.

Van Der Stelt et al. [46] and Uddin et al. [47] examined residential storage primarily from self-consumption and economic perspectives, while Schibuola and Tambani [48] quantified the ability of batteries to retain a larger share of on-site PV generation in retrofit applications. Wang et al. [49] considered carbon-responsive operation and showed that control based on time-varying emission signals can reduce operational carbon in an all-electric community. Pimm et al. [50], however, demonstrated that storage does not inherently reduce emissions, because the outcome depends on the carbon intensity of the electricity used for charging and the electricity displaced during discharge.

The interaction between these control objectives has been explored mainly through economic and operational studies. Lorenzi and Silva [51] showed that demand response and storage can partly substitute for one another in increasing PV self-consumption. Hoppmann et al. [52] and Nygard et al. [53] found that the economic viability of household batteries depends strongly on tariff structure, while Beaudin and Zareipour [54] highlighted the increasing complexity of home energy management once multiple objectives are combined. At the system scale, Dupont et al. [55] showed that storage can even increase emissions when charging and discharging shift electricity between periods supplied by generators with different carbon intensities. This mechanism is directly relevant to the household-scale question examined here. With residential battery deployment accelerating across Europe [56], the distinction between self-consumption, cost-based and carbon-aware control is becoming increasingly important for both building assessment and grid decarbonization.

The most direct precedent applies hourly emission factors to storage in net zero energy homes across the Group of Seven grids and reports a result that matters for this article: storage raises electricity use intensity through round-trip losses even where it lowers emissions, so the energy indicator and the carbon indicator move in opposite directions [57]. That is the same divergence examined here, arrived at from the storage side rather than the certification side.

The great majority of installed residential batteries follow the first objective, because that is what inverter manufacturers ship by default. What carbon benefit such a battery achieves, when it never sees a grid carbon signal, is therefore a practical question rather than a theoretical one, and it is not the same question as whether carbon-aware control works.

2.5. Certification, metering and the regulatory gap

The recast Energy Performance of Buildings Directive requires certificates to report annual final energy, on-site renewable production and operational greenhouse gas emissions [3]. It introduces the vocabulary of calculation intervals, self-use and exported energy, which is a clear step toward temporal accounting. The required indicators nevertheless remain annual, and member states retain wide discretion over method. The framework has therefore begun to ask for operational carbon disclosure without settling the temporal question that grid-interactive buildings raise.

The measurement side has moved faster than the rule. Crawley et al. [58] showed that smart meter data can support thermal performance ratings for dwellings, and Morewood [59] reviewed the data quality conditions under which monitored building performance can be trusted. Smertinas et al. [60] demonstrated digital certification of single-family houses in Denmark using measured data rather than assumed inputs. Belgium introduced 15-min interval metering for electricity in 2026, and the transmission system operator already publishes generation at the same resolution [9]. The European imbalance settlement period is likewise 15 min.

Policy has begun to describe what such data could support. The grid-interactive efficient building concept [61] treats a dwelling as a participant in the power system rather than a passive load, and Burger and Rosenow [62] survey the smart tariffs already operating in Europe to reward that participation. Carroll et al. [63] show that smart meter records can be assembled into reference load profiles suitable for stock-level analysis. Each of these depends on time-resolved metering, and each stops short of the certificate.

The consequence is a gap between what is measured and what is used. Quarter-hourly consumption is now recorded by law for metered dwellings, and quarter-hourly grid carbon intensity can be derived from published generation data. Certification does not use either. Two dwellings with the same annual balance and very different import profiles receive the same rating, and nothing in the current method can separate them.

2.6. Knowledge gap and research question

Three gaps remain. First, studies of temporally resolved building emissions have mostly examined buildings that only consume, or electricity systems in aggregate, with limited attention to how behind-the-meter generation changes the carbon composition of the imports that remain. Second, solar and storage studies commonly evaluate annual energy, self-consumption, cost or total carbon without isolating the accounting error that an annual average factor introduces. Third, this evidence has not been carried into the certification context, where annual netting can assign identical performance to systems with substantially different grid dependence. No framework currently connects calibrated residential demand, behind-the-meter flows, time-resolved grid carbon intensity, and certification-level indicators within a single assessment that can be verified internally.

This article addresses that gap through one question: How do progressive residential renovation pathways combining heat-pump electrification, rooftop PV and battery storage affect delivered energy, grid imports and operational carbon, and to what extent can conventional annual indicators capture these changes compared with time-resolved carbon accounting?

Two concepts are used throughout. The temporal carbon performance gap is the difference between operational carbon computed with hourly factors and the same quantity computed with the annual mean factor, expressed as a percentage of the annual result. It is not the familiar performance gap between modelled and measured consumption. Residual-import carbon intensification describes the second concept, namely that behind-the-meter generation reduces the quantity of imported electricity while raising the average carbon intensity of what remains. Together they produce the paradox this article examines. Electrification, solar, and storage lower both total imports and total operational carbon, and at the same time make every remaining imported kilowatt-hour dirtier.

3. Methodology

This study asks whether the timing of electricity use changes the carbon answer for a Belgian dwelling, and whether that answer shifts once the dwelling makes its own electricity. Four things are needed to answer it: a house that represents its stock, a model that behaves like the real house, grid data at the resolution the question demands, and a set of retrofit options to compare. Fig. 3 shows the workflow, and the sections below follow it from left to right. Supporting material sits in two appendices. Appendix A holds the average hourly grid carbon intensity for each month (Fig. A). Appendix B holds the reconstructed geometry, floor plans and section (Fig. B) and the occupancy and lighting schedules (Fig. C).

3.1. Case study and monitored data

The Sustainable Building Design Lab audited 708 households between 2021 and 2024, collecting smart meter readings for electricity and gas alongside building surveys. From that sample, a semi-detached

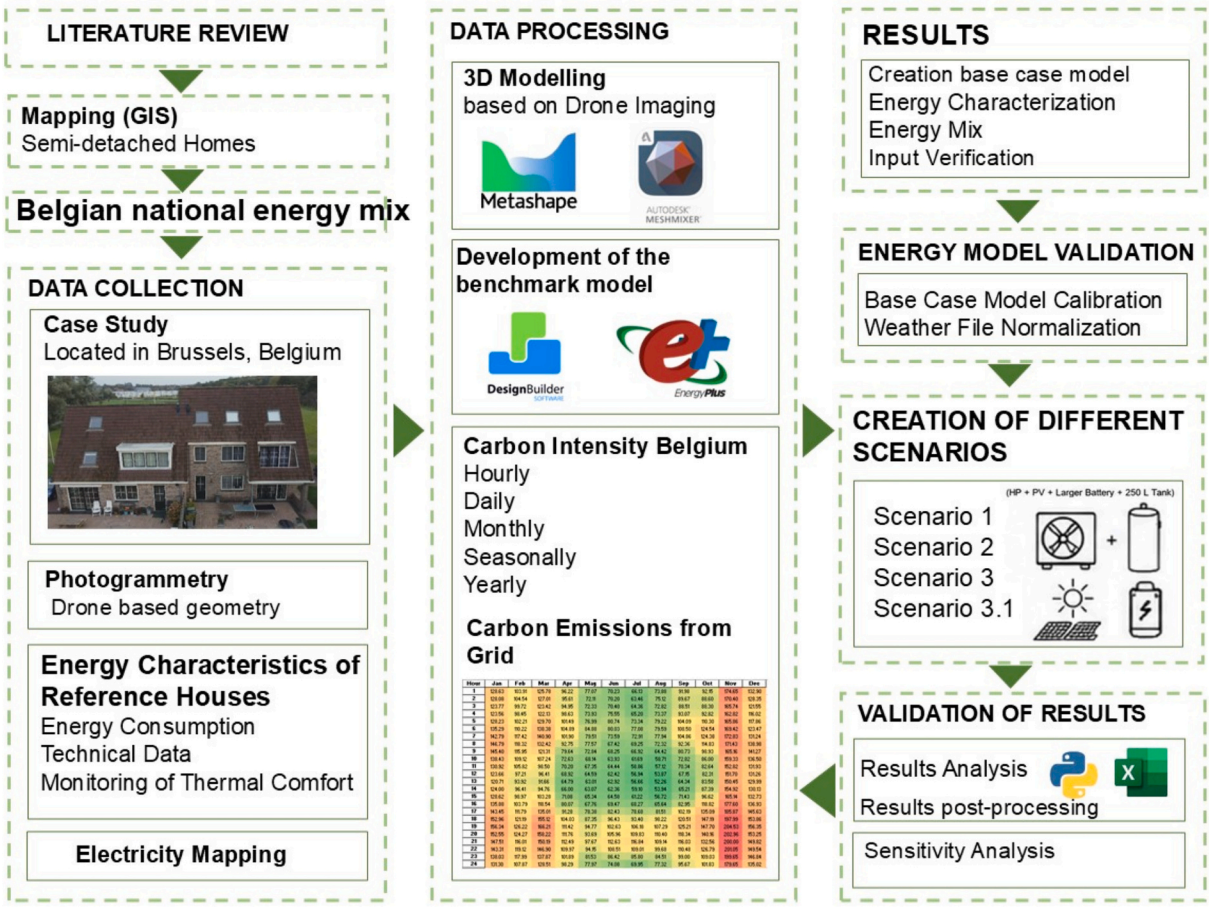


Fig. 3. Workflow for monitored residential energy modelling, dynamic carbon assessment and scenario analysis. The three columns cover data collection, data processing, and results with scenario construction and verification.

dwelling in Brussels was chosen as the archetype. It belongs to the group of 64,535 Belgian semi-detached houses built between 2000 and 2010 [6], and to the low-rise suburban form that spread across Northwestern Europe after 2000.

Selection used cadastral records, GIS mapping, a field survey, and the metered data together. The aim was a dwelling sitting near the middle of the sample on geometry, envelope and operation, rather than one that happened to be available. The result is a two-storey house with insulated cavity walls, a pitched roof, double glazing and mechanical ventilation. A gas boiler supplies both space heating and domestic hot water, and a 250 L tank acts as a buffer on the heating loop. There is no cooling. Four consecutive years of metered data give the calibration target, which matters because a model tuned to one year can reproduce that year and little else. Attia et al. [64] used the same archetype approach to build two benchmark models for Belgian post-war housing, and Attia et al. [65] extended it to renovated nearly zero-energy terraced dwellings.

3.2. Geometry and model development

Facade and roof images were taken with a DJI Mini 2 and processed in Agisoft Metashape through image alignment, point cloud generation, mesh building and surface trimming, then exported to DesignBuilder. Measured facade dimensions, glazing ratios and roof shape from the site survey were used to correct the reconstruction, and neighbouring buildings were added as shading surfaces. The resulting geometry, plans and section are shown in Fig. B in Appendix B. Table 1 gives the geometry and envelope.

The model was built in DesignBuilder with EnergyPlus 9.4.0.002 and twelve thermal zones, ten of them conditioned. The conditioned floor area of 121.11 m2 is the denominator for every intensity in this paper, for both carriers. EnergyPlus was chosen for its sub-hourly heat balance and its ability to report at the resolution this study needs [66]. Heating of the base case was modelled for the gas boiler, and fuel conversion is applied at the efficiencies in Table 6.

Table 1
Geometry and envelope of the benchmark model, from the field survey and the reconstruction.

Parameter	Value	Parameter	Value
Conditioned floor area	121.11 m^2	Glazing, whole-window U	2.8 W/(m^2·K)
Total floor area	169.42 m^2	Solar factor, g	0.79
Conditioned volume	323.75 m^3	Window area	20 m^2
Zones, total and conditioned	12 and 10	Mechanical ventilation	421.5 m^3/h
External wall U	0.25 W/(m^2·K)	Thermal bridges	Not modelled
Pitched roof U	0.23 W/(m^2·K)	Ground temperature	18 °C, constant
Ground floor U	0.30 W/(m^2·K)	Cooling, Base Case	None
Attic floor U	0.25 W/(m^2·K)	Occupants	4

3.3. Occupancy, gains and operation

Occupancy follows a four-person household. The dwelling is occupied overnight and in the evening on weekdays, empty between 09:00

Table 2
Occupancy, internal gains and setpoints used in the model.

Parameter	Value	Source
Occupants	4 persons	Field survey
Occupied, weekday	18:00 to 08:00	Monitored
Unoccupied, weekday	09:00 to 15:00	Monitored
Occupied, weekend	All hours	Monitored
Heating setpoint	23 °C	Calibration
Heating setback	18 °C	Calibration
Cooling setpoint	25 °C	Assumed
Cooling season	1 June to 30 September	Assumed
Fresh air rate	10 L/s per person	EN 16798–1
Domestic hot water draw	64.5 L/day	Calibration
Lighting power density	3.20 W/m^2	Calibration
Plug loads, installed	1208 W	Calibration
Circulators, installed	386 W	Calibration
Infiltration	0.597 ach	Calibration

Table 3
Holiday periods, during which the dwelling follows the weekend profile.

Period	Dates	Days
Easter	30/03 to 05/04	7
Summer	01/08 to 15/08	15
All Saints	28/10 to 03/11	7
Christmas	24/12 to 01/01	9
Total		38

and 15:00, and occupied all day at weekends. Thirty-eight holiday days follow the weekend pattern. Hot water is drawn on a morning and evening profile. Table 2 lists the operational settings and Table 3 the holiday periods. The hourly profiles are shown in Fig. C in Appendix B.

3.4. Calibration

Calibration compared simulated and metered consumption month by month for both carriers, using Mean Bias Error and the Coefficient of Variation of the Root Mean Square Error against the ASHRAE Guideline 14 monthly limits of plus or minus 5% and 15% [67]. The sequence followed the evidence-based approach reviewed by Coakley et al. [68]. Using measured data at monthly and finer resolution to constrain a whole-building model follows the case study of Raftery et al. [69], and Rouchier [70] sets out when measured data can constrain a model without overfitting it.

Six parameters were adjusted, each held inside a range set by evidence rather than by the target. Table 4 lists them with their bounds and the reason for each. Two independent checks were used alongside the monthly statistics. The summer gas floor, measured between June and August when no space heating occurs, gives the hot water demand on its own, so it tests the split between heating and hot water rather than the annual total. The measured July electricity of 5.27 kWh/day sets a ceiling on appliance and standby load, since no appliance schedule can exceed the whole dwelling consumption in the month with the least lighting.

One reference weather year for Brussels was used for all simulations, so the model reproduces the average conditions of the monitoring period rather than the weather of each separate year. Calibration performance is therefore the difference between one simulated year and the four-year measured mean. Weather effects between years are handled separately by degree day normalisation to ISO 15927-6 [71]. Calibration results are reported in Section 4.

Table 4
Calibration parameters, their bounds and the evidence that sets them.

Parameter	Bound applied	Evidence
Lighting power density	2.0 to 6.0 W/m^2	Typical residential range
Plug loads	Below the July ceiling	Measured July consumption
Circulators	Boiler plus buffer loop	System as surveyed
Hot water draw	40 to 60 L/person/day	Belgian household surveys
Infiltration	0.4 to 0.8 ach	Envelope age and glazing type
Heating setpoint	20 to 23 °C	Occupant interview

3.5. Grid carbon intensity

Hourly Belgian grid carbon intensity for 2024 came from Electricity Maps [9], converted from UTC to local time and aligned to the 8760-h simulation year. Noussan et al. [10] show why hourly resolution matters for this kind of accounting. The series runs from 0.022 to 0.294 kgCO2e/kWh with an annual mean of 0.107. Hour of day means range from 0.086 to 0.134, and the 95th to 5th percentile ratio is 4.13. The daily low falls between 12:00 and 13:00 and the high between 18:00 and 20:00. Monthly means range from 0.075 in July to 0.175 in November. Fig. 4 shows these patterns, and Fig. A in Appendix A gives the average daily profile for each month.

Both carriers are treated on a direct emissions basis. Electricity uses the hourly series above and its annual mean of 0.107 kgCO2e/kWh. Natural gas uses a fixed factor of 0.2016 kgCO2e/kWh on a lower heating value basis. This matches the boundary used in energy performance certification and in national greenhouse gas inventories, which is the policy setting this study addresses. Lifecycle factors for the Belgian grid are reported by Attia [72] using the technology values of the IPCC Fifth Assessment Report [73].

The factors used here are hourly average factors. They describe the mean carbon content of electricity supplied in a given hour and are the appropriate basis for attributional accounting, which is what an energy performance certificate does. They are not marginal factors, which describe the emissions of the generator responding to a change in load and answer a consequential question about what a retrofit causes the power system to do. Ben Amor et al. [22] set out methods for estimating hourly marginal factors, and Gallardo et al. [24] argue that the factor should be matched to the decision rather than treated as generally superior. All results in this article are attributional.

3.6. Operational carbon

Dynamic operational GHG emissions or carbon multiplies each hour of grid import by the factor for that hour and sums across the year, with gas added separately:

$$OC_{dyn} = \Sigma_{t=1}^{8760} \left(E_{imp,t} \times EF_{el,t}\right) + E_{gas} \times EF_{gas} \quad (1)$$

Static operational carbon applies the annual mean factor to the same energy flows:

$$OC_{stat} = E_{imp} \times EF_{el,2024} + E_{gas} \times EF_{gas} \quad (2)$$

Here $E_{imp,t}$ is electricity imported during hour t in kWh, $EF_{el,t}$ the matching hourly factor in kgCO_2e/kWh, and E_{gas} annual gas use. Both results are divided by conditioned floor area and reported in kgCO_2e/(m^2·a).

The difference between the two is the temporal carbon performance gap, expressed as a percentage of the annual result:

$$TCPG = \left(OC_{dyn} - OC_{stat}\right)/OC_{stat} \times 100 \quad (3)$$

This is not the performance gap between modelled and measured consumption. It is the difference between the carbon reported with annual average accounting and the carbon obtained by matching imports to time-resolved grid carbon intensity.

Two conventions matter. Operational carbon comes only from grid imports, so solar electricity used in the house lowers emissions by lowering imports, while exported electricity earns no credit. Net annual energy use intensity works the other way and credits all generation against gross delivered energy, whether it is used on-site or not. It is reported as an energy balance only and carries no carbon claim. The two

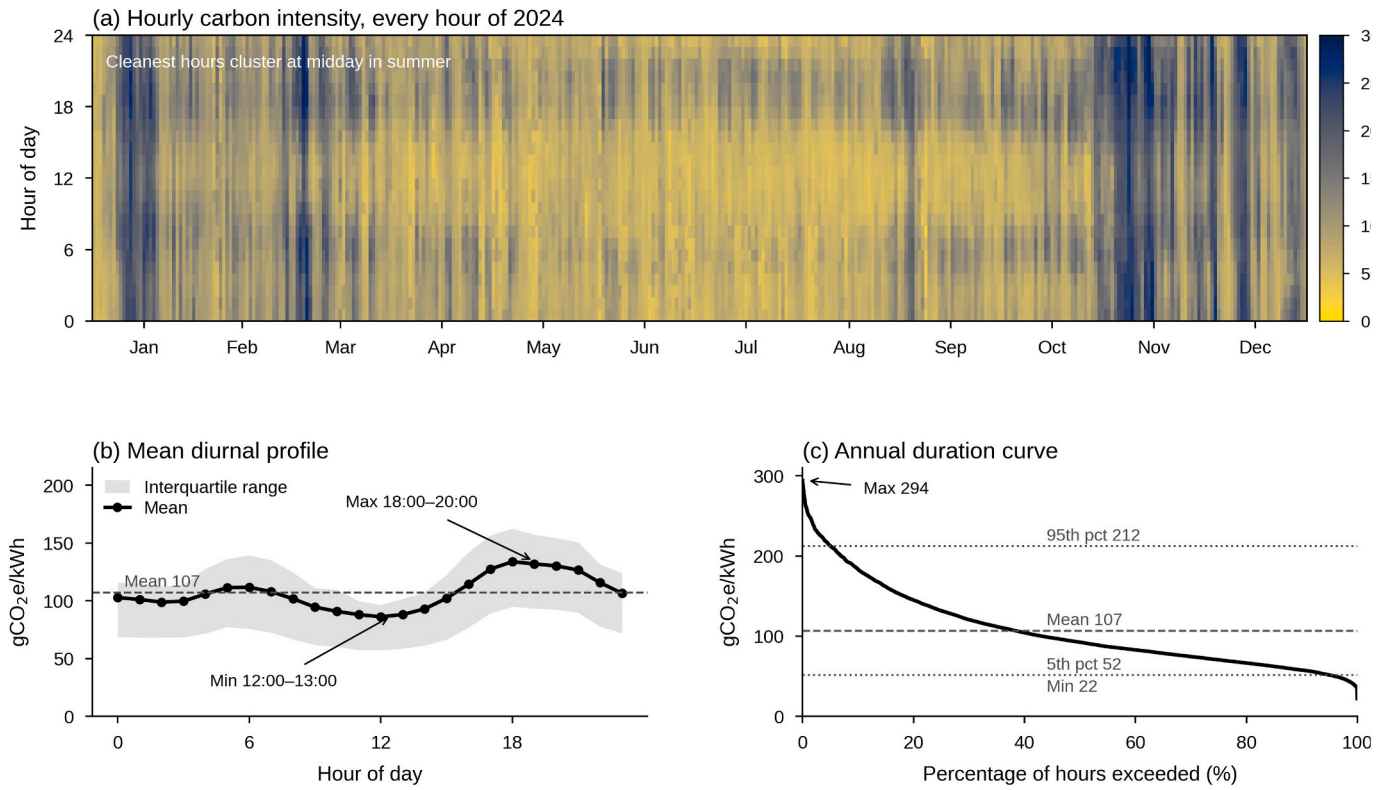


Fig. 4. Belgian grid carbon intensity in 2024. (a) All 8760 h, day of year against hour of day. (b) Mean daily profile with the interquartile range across the year. (c) Annual duration curve with the 5th and 95th percentiles marked. Source: Electricity Maps [9], converted to local time.

measures separate once a generation is present, and measuring that gap is one aim of this study. The import-based convention also matches Belgian market rules, since one-to-one net metering ended in Flanders in 2021 and in Wallonia in January 2024 [74].

3.7. Scenarios

The scenario selection was grounded in contemporary Belgian market conditions and observed household renovation activity rather than representing hypothetical technology combinations. Evidence from Belgian residential networks on combined heat-pump and PV deployment [75], recent large-scale evidence on Belgian household renovation activity [76], and 2025 market data on residential heat-pump uptake [77] support the market relevance of the selected technologies. The technology choices and capacities were additionally informed by the household survey underpinning the present study. Ritter et al. [78] further show that the uptake and activation of residential energy flexibility are shaped by household-level economic, behavioral, regulatory, and market conditions across European countries. Together, this evidence supports a plausible incremental renovation pathway from heating electrification to on-site generation and electrical storage.

The technical configuration is further supported by established evidence on heat-pump electrification in low-carbon grids [32], thermal-storage characterization [39], and residential battery storage at comparable capacities [46,79]. The analysis is deliberately bounded to building energy-system renovation: envelope characteristics are held constant across all configurations to isolate the effects of heating electrification, on-site solar generation, and electrical storage.

The five renovation configurations were implemented and simulated in EnergyPlus using the calibrated Base Case as the common reference model. The heat pump, PV system, and battery storage were introduced sequentially according to the scenario definitions in Table 5, while the calibrated building geometry, envelope, occupancy, internal gains, and demand schedules were retained across scenarios. Heat-pump performance was represented using the seasonal performance parameters reported in Table 6, while PV generation was calculated from the plane-of-array irradiance of the simulation weather file and battery operation followed the hourly self-consumption control described below. This consistent modelling framework isolates the effect of each successive renovation measure while preserving the calibrated building demand and operating conditions. The battery follows a solar self-consumption rule at an hourly timestep. Surplus generation charges it within its rated power and usable capacity, and it discharges to meet remaining demand under the same limits. Charging from the grid is not allowed. Because charging depends on solar surplus, the battery moves electricity within a day and not between seasons. It is a self-consumption device, not a carbon-responsive controller, and the results should be read that way.

Table 5
The five investigated configurations.

Scenario	Heating and hot water	Cooling	PV	Battery	Buffer tank
Base Case	Gas boiler	None	None	None	250 L
Scenario 1	Heat pump	Yes	None	None	250 L
Scenario 2	Heat pump	Yes	1.9 kWp	None	250 L
Scenario 3	Heat pump	Yes	1.9 kWp	2.0 kWh, 1.0 kW	250 L
Scenario 3.1	Heat pump	Yes	1.9 kWp	5.0 kWh, 2.5 kW	250 L

Table 6
System specifications and emission factors.

Parameter	Value	Notes / basis
Simulation engine	EnergyPlus 9.4.0.002	
Interface	DesignBuilder 7.3.1.003	
Zone timestep	2 per hour	
Boiler efficiency, space heating	0.94	
Boiler efficiency, hot water	0.94	
Heat pump SPF, space heating	3.20	Representative retrofit value, benchmarked against Daikin Altherma 3 Belgian manufacturer data; SCOP ≈ 3.26 to 3.32 at 55 °C, average climate
Heat pump SPF, hot water	2.50	Conservative model assumption for higher-temperature DHW operation; consistent with lower efficiency expected at elevated water temperatures
Heat pump SEER, cooling	3.50	Conservative model assumption for reversible air-to-water heat pump cooling
Buffer tank volume	250 L	
Buffer tank standing loss	1.1 kWh/(m^2·a)	Manufacturer-derived value
Cooling setpoint and season	25 °C, June to September	
PV rated capacity	1.90 kWp	
PV tilt and azimuth	35°, south	
PV performance ratio	0.85	Model assumption representing aggregate PV system losses
PV specific yield	1072 kWh/(kWp·a)	
Battery round-trip efficiency	0.90	Representative residential Li-ion storage assumption
Battery state-of-charge range	0 to 100% of usable capacity	Model definition; 2 and 5 kWh capacities refer to usable, not nominal, storage capacity
Grid charging	Not permitted	
Electricity emission factor, hourly	0.022 to 0.294 $kgCO_2e$/kWh	
Electricity emission factor, annual mean	0.107 $kgCO_2e$/kWh	
Natural gas emission factor	0.2016 $kgCO_2e$/kWh	

The 250 L buffer tank is retained in all five configurations. In the Base Case, its standing losses are implicitly included in the metered gas consumption used for calibration. In Scenarios 1 to 3.1, the same buffer tank serves the heat pump, where it provides hydraulic separation and supports defrost and minimum run time. A standing loss of 1.1 kWh/(m^2·a), derived from manufacturer data for a comparable 250 L buffer vessel, was included explicitly in the heat-pump scenarios. Because this loss is small and distributed approximately uniformly over the heating season, it affects annual-average and hourly carbon calculations similarly and therefore primarily influences absolute energy and carbon results rather than the temporal carbon performance gap. The buffer tank is not operated for load shifting in any scenario, and no flexibility benefit is attributed to it.

The SPF values in Table 6 are model assumptions and were not measured in the case study dwelling. The space heating value of 3.20 was set slightly below the manufacturer SCOP of 3.26 to 3.32 for the Daikin Altherma 3 at 55 °C in the average climate. Field studies often report lower values than rated ones. O'Hegarty et al. [80] reviewed 378 air-to-water heat pumps and found an average in-use seasonal performance of 2.59. Gleeson and Lowe [81] compared European field trials and showed that the reported values depend strongly on how the system boundary of the SPF is defined. The most recent Fraunhofer ISE field study of 77 heat pumps in existing homes in Germany reported a mean SPF of 3.4 for air-source units, with a range of 2.6 to 4.9 [82]. The values used here therefore sit inside the measured range, but not at its lower end. Section 5.4 discusses how lower values would affect the results.

3.8. Verification and transferability

Three checks were run before any result was reported. First, the hourly energy balance was verified for each scenario, confirming that grid imports plus useful solar equal electricity demand in every hour, and that generation equals direct use plus battery charging plus exports. Second, the annual mean of the hourly emission factor series was checked against the annual factor in Eq. (2), because a mismatch would mean the comparison measured the system boundary rather than the time resolution. Third, the emission factor implied by each step of the scenario ladder was computed as the change in operational carbon divided by the change in grid imports, then tested against the observed spread of hourly factors. A step that displaces electricity dirtier than the dirtiest hour of the year is not possible, and this check finds such a result at once. All three are reported in Section 4.

The calibrated Base Case provides a benchmark for examining renovation pathways in the Belgian semi-detached housing context. The methodological framework is transferable to other residential stocks, but the numerical results are specific to the building characteristics, occupancy, climate, and grid conditions represented here. Application to other countries or dwelling typologies therefore requires local weather and hourly grid emission data together with appropriate building, system, and occupancy inputs. The scenario framework itself, namely the progressive transition from gas heating to heat-pump electrification, rooftop PV, and battery storage, can be reproduced consistently across such contexts. Section 5 discusses the implications for other Western European grids.

All post-processing was done in Python. The calculation script, the hourly demand profiles, and the emission factor series are available so that every reported value can be regenerated from the inputs.

4. Results

Results follow the order of the method. Section 4.1 places the benchmark within the national stock. Section 4.2 tests whether the model reproduces the metered dwelling. Section 4.3 reports the calibrated characteristics. Sections 4.4 and 4.5 give the energy outcomes of the five configurations, Section 4.6 the carbon outcomes and the gap between annual and hourly accounting, and Section 4.7 the verification checks.

4.1. The benchmark and the national stock

Semi-detached housing is found across Belgium rather than in one region. Fig. 5 maps the stock by municipality, both in total and for the 2000 to 2010 period. The cohort built in that decade numbers 64,535 dwellings [6]. Densities are highest around the northern and central urban belts and lowest in the southern provinces, but no part of the country is without it. This matters for what follows because the selected typology represents a substantial and geographically distributed segment of the Belgian housing stock, while the subsequent analysis remains a detailed case-based assessment of one calibrated benchmark dwelling.

The 708 households audited between 2021 and 2024 are a subset of that population. Consumption spans a factor of five, from about 50 to 250 kWh/(m2·a). Gas dominates at every point in the ranking, and the electricity band is close to uniform across the sample. Households differ mainly in how much gas they burn, not in how much electricity they use.

That finding shapes the scenario design. If almost all the variation sits in space heating, then replacing the heating system is the measure with the widest reach, and the electrical baseline against which solar

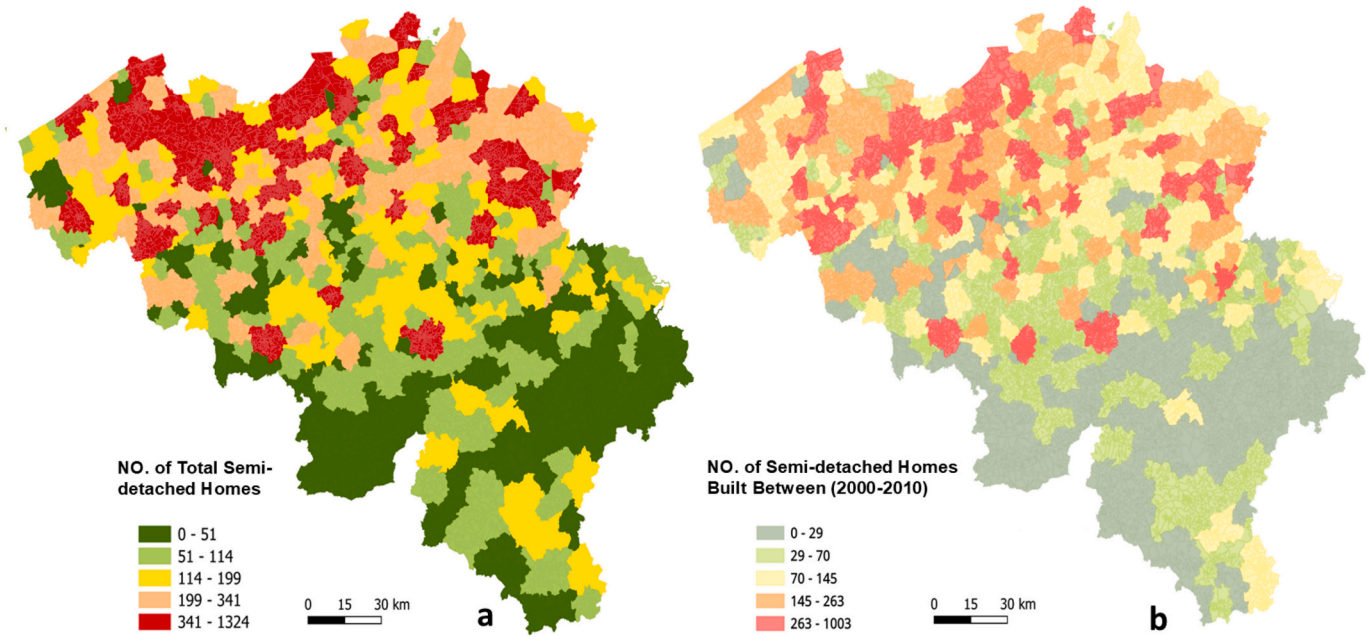


Fig. 5. Distribution of semi-detached dwellings across Belgian municipalities. (a) Total number. (b) Number built between 2000 and 2010.

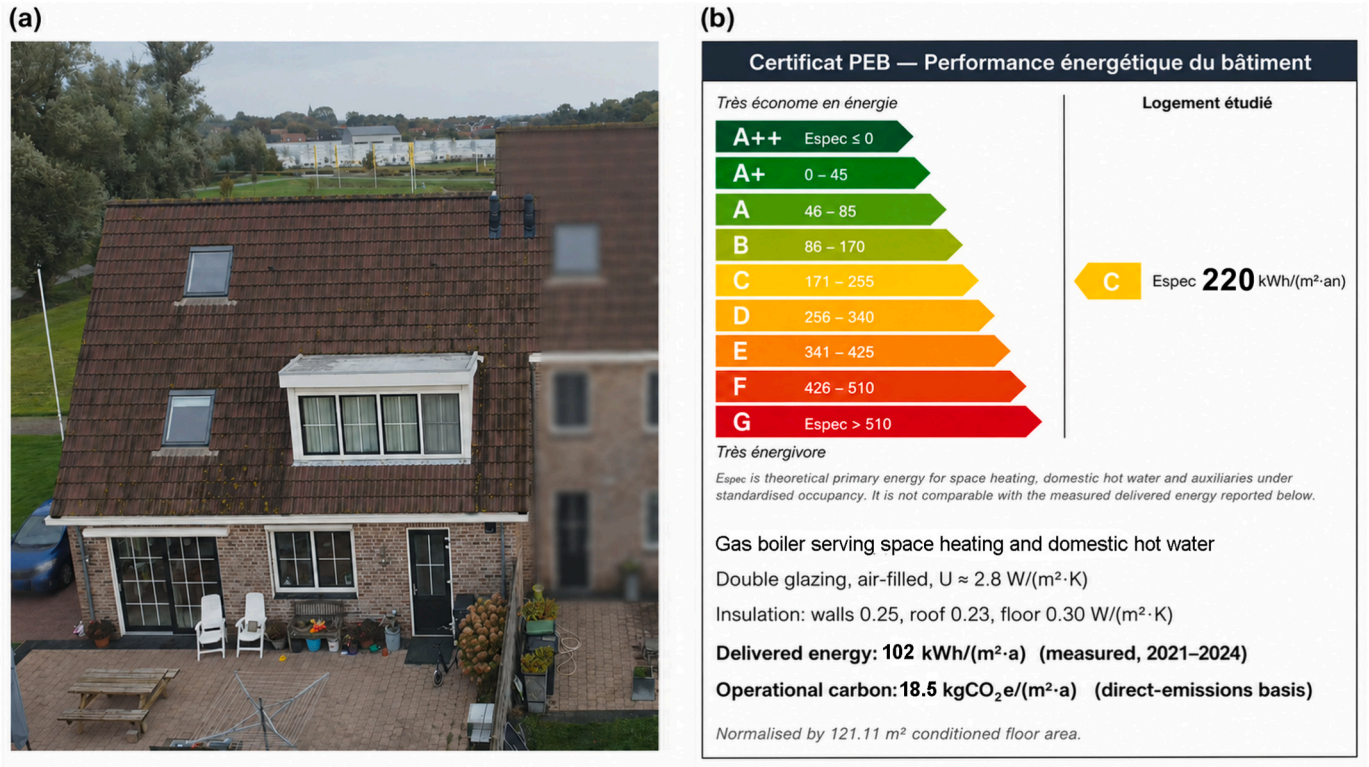


Fig. 6. The benchmark dwelling. (a) Rear elevation. (b) Belgian energy performance certificate with the measured delivered energy and operational carbon. Values normalised by 121.11 m^2 conditioned floor area.

and storage work is nearly the same from one dwelling to the next. The benchmark used here sits at rank 354 of 708, the median of the audited sample, with delivered energy of 102 kWh/(m2·a). It is a middle dwelling rather than a convenient one. Fig. 6 shows the dwelling and its energy certificate. Fig. 7 ranks them by annually delivered energy and shows the two carriers separately.

4.2. Calibration

Four consecutive years of metered electricity and gas were available. Fig. 8 shows the monthly record. Both carriers repeat closely from year to year. December electricity varies by only 0.28 kWh/m2 across the four years and January gas by 0.75 kWh/m2. That stability indicates settled occupancy over the monitoring period and makes the four-year mean a sound calibration target.

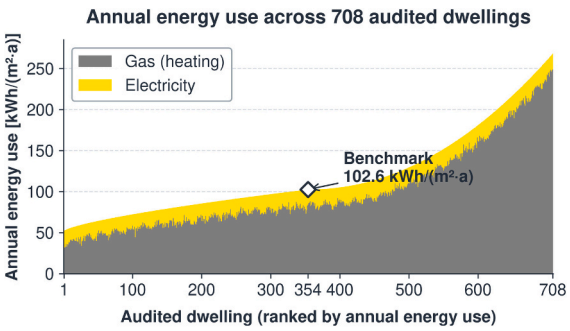


Fig. 7. Annual delivered energy across the 708 audited dwellings, ranked by total consumption, with gas and electricity shown separately. The benchmark is marked at rank 354.

Six parameters were adjusted within the bounds set in Section 3.4. The final values are lighting at 3.20 W/m2, plug loads at 1208 W installed, circulators at 386 W, a hot water draw of 64.5 L/day, infiltration at 0.597 ACH, and a heating setpoint of 23 °C. No electric heating was added, because none exists in the dwelling.

The calibrated model meets ASHRAE Guideline 14 for both carriers at a monthly resolution [67]. Electricity gives a mean bias error of +1.3% and a CV(RMSE) of 7.1%. Gas gives 0.0% and 7.7%. Both sit within the limits of ±5% and ±15%. Table 7 reports the monthly comparison and Fig. 9 places the benchmark and the model against the distribution of the 708 audited dwellings.

Two independent checks support the annual statistics. The summer gas floor, measured between June and August when no space heating occurs, gives hot water demand on its own. The model returns 1634 kWh against a measured 1617 kWh, so the split between heating and hot water is reproduced and not only the annual total. The measured July electricity of 5.27 kWh/day sets a ceiling on appliance and standby load, and the calibrated appliance total sits inside it.

The largest residuals fall in the shoulder months and in July. October gas is 20% low and July electricity 20% high on a small base. These follow from the single reference weather year, which cannot reproduce the start of any particular heating season. Both annual totals and both dispersion statistics are well inside tolerance.

4.3. Calibrated benchmark characteristics

Table 8 reports the benchmark as calibrated. These values are a result rather than an input, because four of them were settled by the calibration rather than by the survey. Two are worth noting. Lighting at 3.20 W/m2 is close to current residential practice, and infiltration at

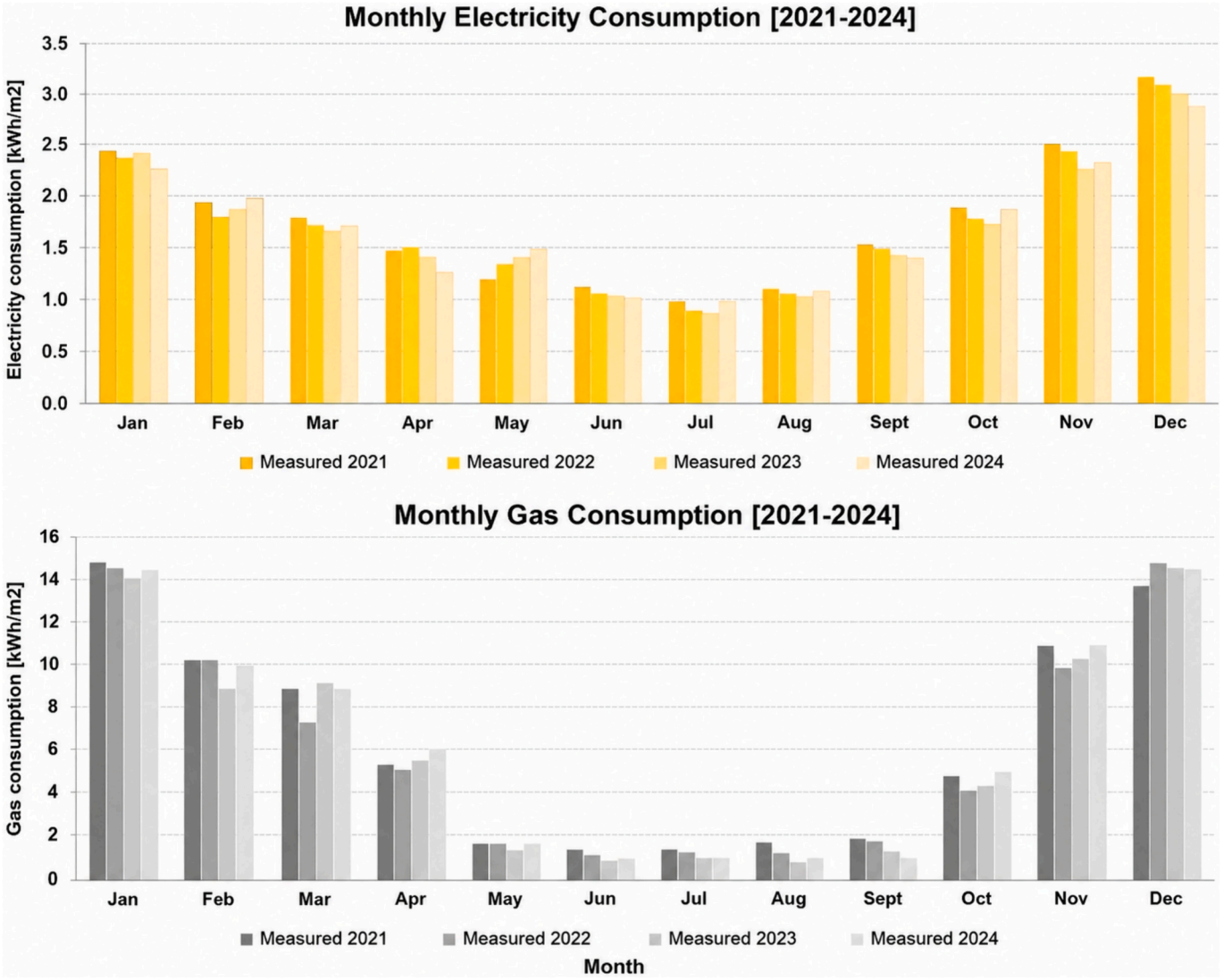


Fig. 8. Measured monthly electricity and natural gas consumption of the benchmark dwelling, 2021 to 2024.

Table 7
Calibration of the benchmark model against the measured record, 2021 to 2024 mean. One reference weather year was used for all simulations, so the simulated column reports a single annual result against a four-year measured mean.

Month	Electricity measured (kWh)	Electricity simulated (kWh)	Deviation (%)	Gas measured (kWh)	Gas simulated (kWh)	Deviation (%)
January	405	415	+2.4	1749	1779	+1.7
February	325	300	−7.8	1187	1138	−4.1
March	292	245	−16.2	1031	1132	+9.8
April	241	226	−6.2	656	659	+0.5
May	235	218	−7.4	185	136	−26.4
June	180	179	−0.6	130	96	−26.2
July	163	197	+20.4	135	109	−19.2
August	182	186	+2.0	139	136	−2.2
September	246	241	−2.3	173	168	−2.9
October	309	302	−2.3	542	435	−19.7
November	404	414	+2.3	1266	1324	+4.6
December	514	533	+3.8	1736	1818	+4.7
Annual	3497	3454	−1.2	8928	8929	0.0
MBE			+1.3			0.0
CV(RMSE)			7.1			7.7

0.597 ACH is consistent with clear double glazing in metal frames and an envelope of this period. The circulator load of 386 W reflects a boiler serving both heating and hot water together with a buffer loop, and it is what carries the winter shape of the electricity profile.

4.4. Base case energy and carbon

The calibrated base case delivers 102.2 kWh/(m2·a), of which 73.7 is gas, and 28.5 is electricity. Space heating accounts for 60.2, and hot water for 13.5, so heating and hot water together are 72% of delivered energy. On the electrical side, appliances lead at 13.0, followed by circulators at 7.7 and lighting at 7.9. Table 9 gives the breakdown.

Applying the natural gas factor of 0.2016 kgCO2e/kWh and the annual mean electricity factor of 0.107 kgCO2e/kWh gives a static operational carbon of 17.92 kgCO2e/(m2·a). Gas contributes 14.87 of that and electricity 3.05, so gas carries 83% of base case operational carbon. That share sets the ceiling on what changing the heating system alone can achieve.

4.5. Scenario energy performance

Replacing the gas boiler with a reversible heat pump cuts gross delivered energy from 102.2 to 48.5 kWh/(m2·a), a fall of 53%. Heating and hot water together drop from 73.7 kWh/(m2·a) of gas to 22.8 of electricity. Cooling adds 2.2 kWh/(m2·a), a service the base case does not provide at all, so the comparison gives the electrified scenarios a summer benefit the gas boiler cannot match.

The 250 L buffer tank is present in all five configurations and therefore does not differentiate the renovation scenarios. In the Base Case, its standing losses are implicitly included in the metered gas consumption used for calibration. In Scenarios 1 to 3.1, the same buffer tank serves the heat pump, with a standing loss of 1.1 kWh/(m^2·a) included in the model, derived from the manufacturer. Because this loss is small and distributed approximately uniformly over the heating season, it affects annual-average and hourly carbon calculations similarly. Its influence is therefore primarily on the absolute energy and carbon results rather than on the temporal carbon performance gap.

Adding solar changed the picture in a way that annual and hourly indicators describe differently. Gross delivered energy stayed at 48.5 kWh/(m2·a) in every generating scenario, because the dwelling still needs the same energy. Net annual energy use intensity fell to 31.7 and then stayed identical across Scenarios 2, 3 and 3.1, because annual netting credits all generation regardless of when it occurs and the array is the same size in all three. Grid imports told a different story, falling from 48.5 to 40.0, 36.2 and 33.2 kWh/(m2·a) as storage was added.

Solar utilisation explains that difference. Without storage, the dwelling used 50.5% of its generation directly. The 2 kWh battery raised that to 75.7% and the 5 kWh battery to 95.2%, while exports fell from 8.3 to 4.1 and 0.8 kWh/(m2·a). Cooling helps here, because it falls in the same hours as the generation. Self-sufficiency, the share of demand met on site, rose from 17.5% to 25.3% and 31.4%. Table 10 gives the flows.

4.6. Operational carbon

Operational carbon fell across the ladder, from 18.46 kgCO2e/(m2·a) in the base case to 6.14 with the heat pump, 5.38 with solar, and 4.95 and 4.64 with the two batteries. The heat pump delivered the largest single reduction at 67%, which follows from gas carrying 83% of base case carbon. Solar cut a further 12%, and the two batteries 8% and 14% relative to Scenario 2. Table 11 reports the full balance and Fig. 10 shows it.

A key finding is that the two accounting methods increasingly diverge along the renovation pathway. Hourly accounting produced higher operational carbon than annual-average accounting in every configuration.

The load-weighted factor of imported electricity explains why. It rose steadily from 0.1261 kgCO2e/kWh in the base case to 0.1394 in Scenario 3.1, and exceeded the annual mean of 0.107 in every case. The electricity this dwelling imports is consistently more carbon-intensive than the Belgian annual average and becomes more so as generation is added.

The reason is the temporal structure reported in Section 3.5. Grid carbon intensity reaches its daily low between 12:00 and 13:00, and its high between 18:00 and 20:00, and its monthly mean runs from 0.075 in July to 0.175 in November. Solar generation falls in exactly the hours and months when the grid is cleanest. Self-consumption therefore displaces the cleanest imports first and leaves a residual concentrated in winter evenings. Fig. 11 shows the hourly consequence.

The battery acts almost entirely in the evening. Of the carbon it displaces, 89% falls between 15:00 and 21:00, because it charges on solar surplus in the middle of the day and discharges within the same day. It cannot move energy from a sunny July to a dark January, which is where the largest carbon gradient lies.

Across Scenarios 2, 3 and 3.1, the net annual energy use intensity was identical at 31.7 kWh/(m2·a). Over the same three configurations, grid imports fell 17%, and operational carbon fell 14%. An indicator that reports no change across three physically different systems is not measuring what determines their carbon performance. Two mechanisms

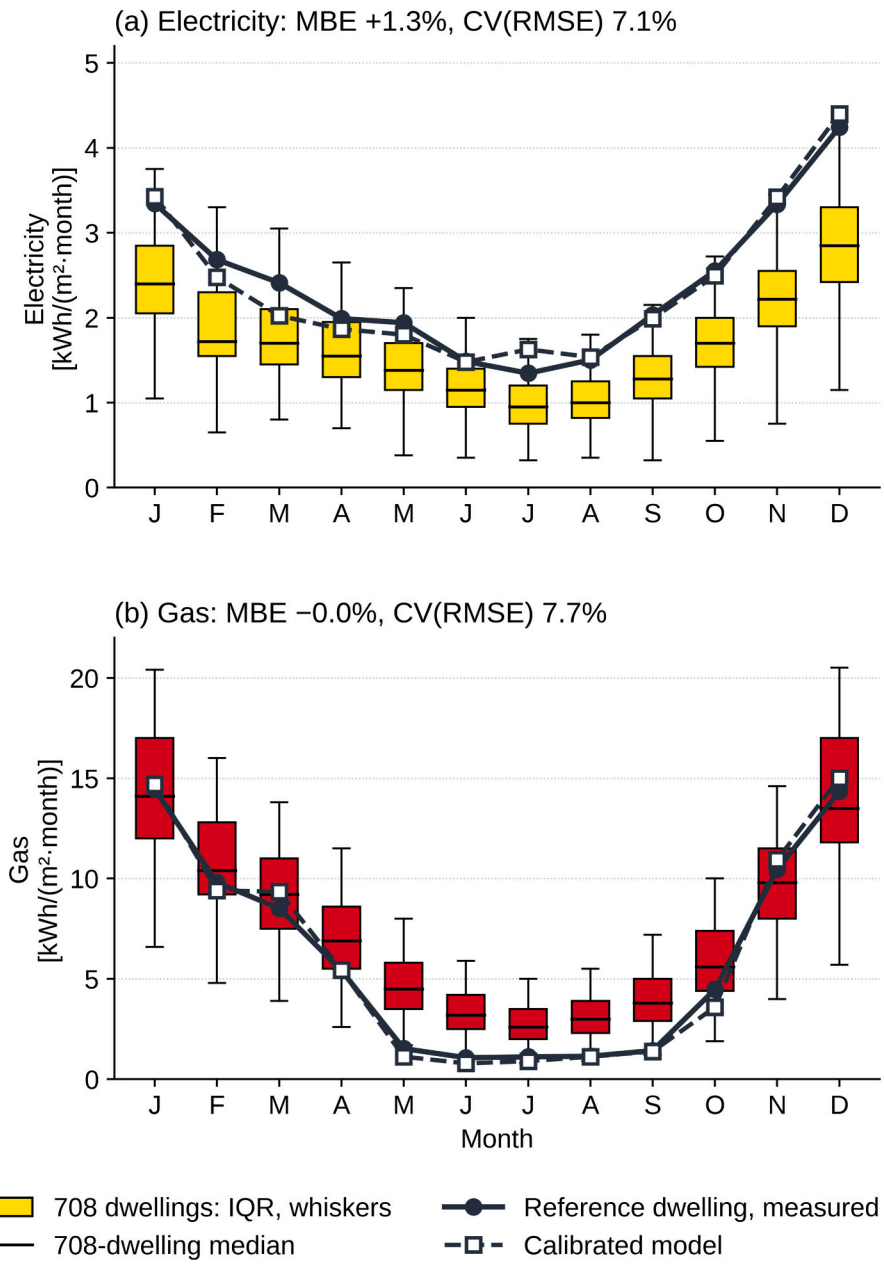


Fig. 9. Monthly consumption of the benchmark dwelling and the calibrated model against the distribution of the 708 audited dwellings. (a) Electricity. (b) Natural gas.

Table 8
Calibrated characteristics of the benchmark dwelling. Values marked as calibrated were settled against the metered record; the remainder come from the field survey and the geometry reconstruction.

Parameter	Value	Source
Conditioned floor area	121.11 m^2	Reconstruction
Total floor area	169.42 m^2	Reconstruction
Conditioned volume	323.75 m^3	Reconstruction
External wall U	0.25 W/(m^2·K)	Survey
Pitched roof U	0.23 W/(m^2·K)	Survey
Ground floor U	0.30 W/(m^2·K)	Survey
Glazing, whole-window U	2.8 W/(m^2·K)	Survey
Solar factor, g	0.79	Survey
Occupants	4 persons	Survey
Infiltration	0.597 ach	Calibrated
Lighting power density	3.20 W/m^2	Calibrated
Plug loads, installed	1208 W	Calibrated
Circulators, installed	386 W	Calibrated
Hot water draw	64.5 L/day	Calibrated
Heating setpoint and setback	23 °C and 18 °C	Calibrated

Table 9
Simulated annual energy performance of the Base Case, normalised by 121.11 m^2 conditioned floor area.

End use	kWh/a	kWh/(m^2·a)	Share (%)
Space heating (gas)	7291	60.2	58.9
Domestic hot water (gas)	1634	13.5	13.2
Catering (gas)	5	0.0	0.0
Lighting	956	7.9	7.7
Appliances	1573	13.0	12.7
Circulators and fans	925	7.7	7.5
Total gas	8929	73.7	72.1
Total electricity	3454	28.5	27.9
Total delivered energy	12,383	102.2	100.0

Table 10
Solar utilisation and grid interaction. Self-consumption is generation retained on site divided by total generation. Self-sufficiency is the useful solar contribution divided by gross electricity demand. Both are computed after storage losses. All values in kWh/(m^2·a) except the ratios.

Quantity	Scenario 2	Scenario 3	Scenario 3.1
Gross electricity demand	48.5	48.5	48.5
Solar generation	16.8	16.8	16.8
Solar used directly	8.5	8.5	8.5
Battery charge	0.0	4.2	7.5
Battery discharge	0.0	3.8	6.8
Storage losses	0.0	0.4	0.8
Useful solar contribution	8.5	12.3	15.2
Exports to grid	8.3	4.1	0.8
Grid imports	40.0	36.2	33.2
Self-consumption (%)	50.5	75.7	95.2
Self-sufficiency (%)	17.5	25.3	31.4

drive this. Annual netting credits exported electricity at the same value as imported electricity, which has not matched Belgian market rules since one-to-one net metering ended [74]. And separately, the electricity that self-consumption displaces is cleaner than the annual average, so applying an annual mean factor to the residual imports understates its carbon content. The first makes the indicator blind to storage. The second makes it optimistic about generation.

4.7. Verification

The hourly energy balance closed in every configuration. Grid imports plus useful solar equalled electricity demand, and generation equalled direct use plus battery charging plus exports, with a worst annual residual of 9×10–13 kWh.

The simple annual mean of the hourly factor series was confirmed against the annual factor used in Eq. (2). Both come from the same dataset on the same accounting basis, so the static and dynamic results differ only in time resolution.

The factor implied by each step of the ladder was computed as the change in operational carbon divided by the change in grid imports. The three steps gave 0.090, 0.112 and 0.107 kgCO2e/kWh, all within 16% of the annual mean and all well inside the observed range of hourly factors, whose 99th percentile is 0.252 and whose maximum is 0.294. Table 12 reports the check. The implied factors do not rise as storage grows, which is the behaviour physics requires, since a battery displaces the most carbon-intensive hours it can reach first and each further unit of capacity must reach into cleaner hours.

5. Discussion

5.1. Key findings and recommendations

Four findings come out of this work. The first is that the temporal carbon performance gap grows along the retrofit pathway. Hourly accounting gave higher operational carbon than the annual mean factor in every configuration, and the gap widened from 3.0% for the gas boiler to 30.3% once solar and a 5 kWh battery were added. The direction agrees with the established literature. Bontekoe et al. [13] reported that annual factors understate operational carbon by 25 to 50% for retrofit assessment, and Zou et al. [14] found that temporal resolution changes the carbon result materially in commercial buildings. What is new here is the progression. The error is small while the dwelling burns gas and grows by an order of magnitude as it electrifies and generates, so the accounting method matters least in the buildings that need assessing least.

The second finding explains the first and is the mechanism this article names. Residual-import carbon intensification means that behind-the-meter generation reduces how much electricity a dwelling imports while raising the average carbon content of what remains. Belgian grid carbon intensity is lowest between 12:00 and 13:00 and highest between 18:00 and 20:00, and it runs from 0.075 kgCO2e/kWh in July to 0.175 in November. Solar generates in exactly the hours and months when the grid is cleanest, so self-consumption removes the cleanest imports first. The load-weighted factor of imported electricity rose from 0.1261 to 0.1394 kgCO2e/kWh as generation was added and stayed above the annual mean of 0.107 throughout. This produces the paradox at the centre of the article. Electrification, solar and storage

Table 11
Annual energy balance and operational carbon across the five configurations. Static carbon applies the 2024 annual mean factor of 0.107 $kgCO_2e$/kWh to gross grid imports. Dynamic carbon applies measured hourly factors to hourly imports. The load-weighted factor is dynamic carbon divided by grid imports. Energy in kWh/(m^2·a), carbon in $kgCO_2e$/(m^2·a).

Scenario	Gross EUI	Solar	Net EUI	Grid imports	Static	Dynamic	Difference	Load-weighted factor
Base Case	102.2	0.0	102.2	28.5	17.92	18.46	+3.0%	0.1261
Scenario 1	48.5	0.0	48.5	48.5	5.19	6.14	+18.4%	0.1267
Scenario 2	48.5	16.8	31.7	40.0	4.28	5.38	+25.6%	0.1344
Scenario 3	48.5	16.8	31.7	36.2	3.87	4.95	+27.8%	0.1368
Scenario 3.1	48.5	16.8	31.7	33.2	3.56	4.64	+30.3%	0.1394

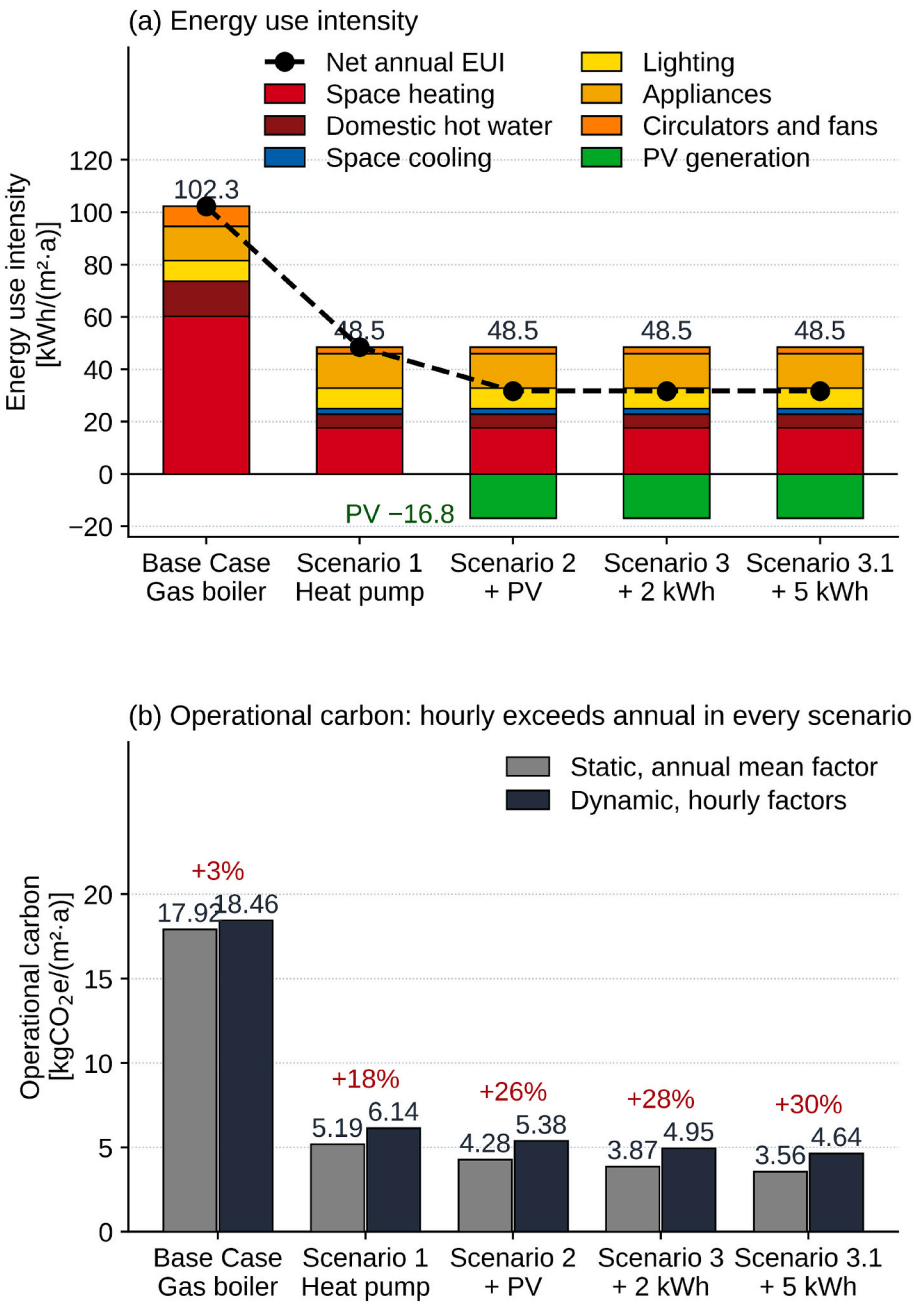


Fig. 10. Energy use intensity and operational carbon across the five configurations. (a) End-use composition of gross delivered energy with net annual energy use intensity overlaid. (b) Static and dynamic operational carbon, with the percentage by which the dynamic result exceeds the static one.

lower both total imports and total operational carbon, and at the same time make every remaining imported kilowatt-hour dirtier.

The third finding is that the compliance indicator cannot see any of this. Net annual energy use intensity was identical at 31.7 kWh/(m2·a) across three physically different systems, while grid imports fell 17% and operational carbon fell 14%. An indicator that reports no change across those three cases is not measuring what determines their carbon performance. Two certified dwellings with the same annual balance and very different import profiles are indistinguishable under the current method.

The fourth finding concerns the battery. It is a self-consumption device and not a flexibility asset. It charges on solar surplus in the middle of the day and discharges the same evening, so 89% of the carbon it displaces falls between 15:00 and 21:00. It cannot move energy from a sunny July to a dark January, which is where the largest carbon gradient sits. Wang et al. [49] showed that a controller responding to time-varying emissions can lower operational carbon in an all-electric community, and Pimm et al. [50] cautioned that storage does not always reduce emissions at all. The battery modelled here never sees a grid carbon signal, so its carbon benefit follows from raising self-consumption rather than from any response to the grid.

Recommendations.

For EPBD implementation. Add an hourly operational carbon indicator alongside the annual primary energy figure. The two measure different things, and one does not replace the other. Annual primary energy remains a sound measure of how much energy a building needs. It says nothing about when that energy is drawn, and after electrification that is what decides the carbon. The recast Directive already introduces calculation intervals, self-use and exported energy into its vocabulary [3], so the step is smaller than it appears.

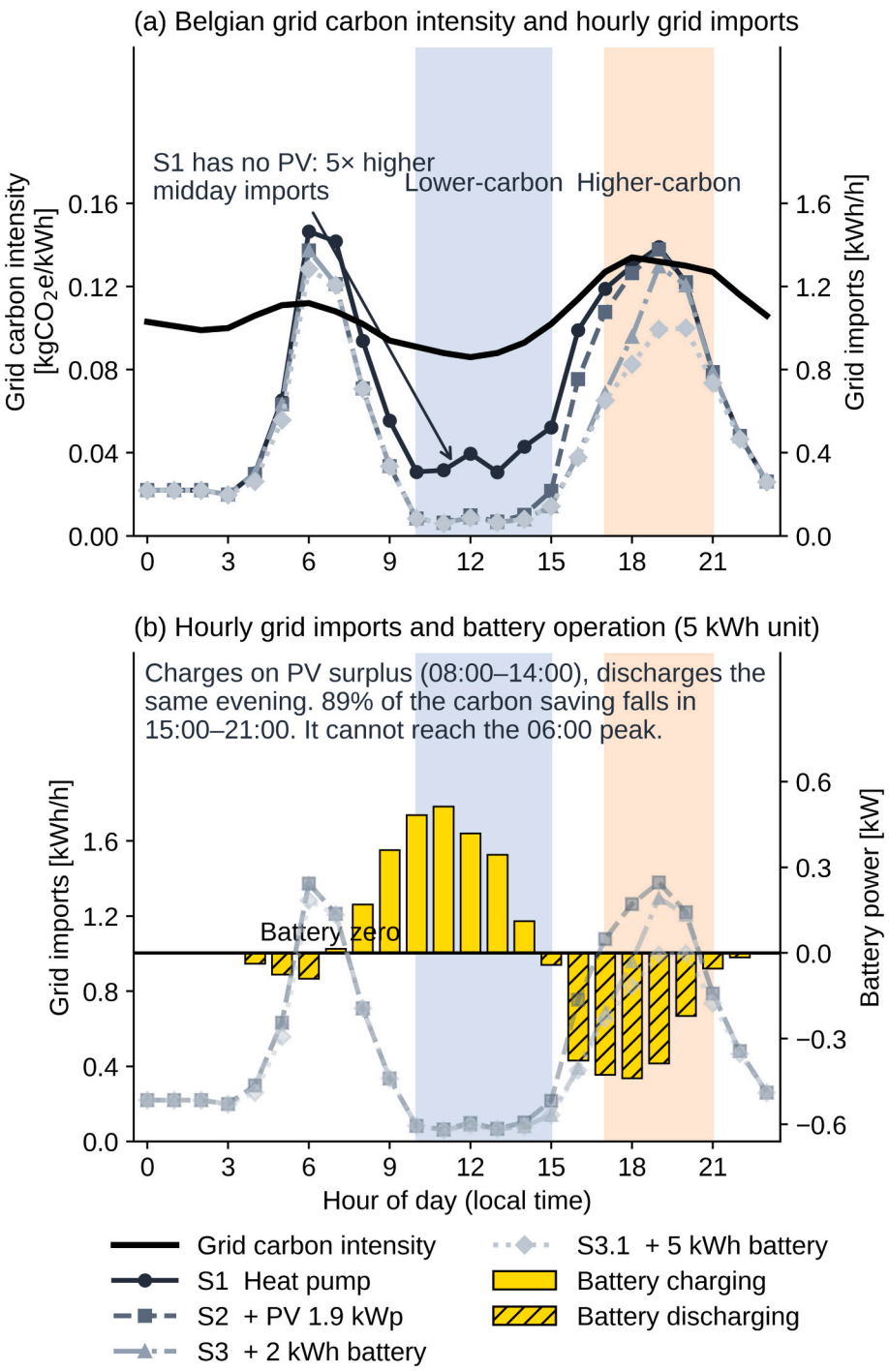


Fig. 11. Hourly grid interaction across the five configurations. (a) Belgian grid carbon intensity with mean hourly grid imports. (b) Grid imports and battery operation for the 5 kWh unit.

For national and regional certification bodies. Stop crediting exported electricity at the same value as imported electricity. That convention has not matched Belgian market rules since one-to-one net metering ended, in Flanders in 2021 and in Wallonia in January 2024 [74]. Report grid imports as a certified quantity in their own right, because imports drive the carbon result and net energy use intensity hides them. Smertinas et al. [60] show that digital certification from measured data is already workable at national scale.

For industry. Size storage for self-consumption and describe it that way. The carbon case and the self-consumption case point the same way here, but only because the battery reduces imports, not because it targets carbon-intensive hours. A supplier claiming a carbon benefit from a solar-charged battery should be able to show which hours it displaces. The test in Table 12 does that in a single calculation.

5.2. *Battery control and the carbon result*

The battery in Scenarios 3 and 3.1 follows a self-consumption rule. It charges only from solar surplus and discharges to meet the next demand. Two other control strategies are common, and both would change the results.

Cost-minimising control under a dynamic tariff charges the battery

Table 12
Emission factor implied by each step of the scenario ladder, tested against the observed distribution of hourly factors. A step displacing electricity more carbon-intensive than the dirtiest hour of the year would indicate an error in the accounting.

Step	Imports displaced	Carbon displaced	Implied factor	Ratio to annual mean	Inside observed range
Scenario 1 to 2	8.5	0.762	0.090	0.84	Yes
Scenario 2 to 3	3.8	0.427	0.112	1.05	Yes
Scenario 3 to 3.1	3.0	0.318	0.107	1.00	Yes

when electricity is cheap and discharges it when electricity is expensive. If grid charging is allowed, the battery can also charge at night or on windy days, and it can work in winter, when there is little solar surplus. The carbon effect then depends on how well the price follows the carbon intensity. In Belgium, cheap hours often coincide with high solar or wind output, which are also low-carbon hours, so part of the price signal points in the right direction. The match is not reliable, however. Wilczynski et al. [33] show that the tariff design decides how much flexibility a household activates, and Fleschutz et al. [83] found that price-based load shifting increased emissions in 8 of 20 European countries at the carbon prices of 2017 to 2019. A cost-minimising battery could therefore reduce carbon more or less than the self-consumption battery modelled here, depending on the tariff and the year.

Carbon-forecast control charges the battery in forecast low-carbon hours and discharges it in forecast high-carbon hours. Its main advantage is that it can charge from the grid in winter, when the dirtiest imports occur. A simple condition shows when grid charging saves carbon. Charging in hour a and discharging in hour b reduces emissions only if $EF_a < \eta_rt \times EF_b$, where η_rt is the round-trip efficiency. With $\eta_rt = 0.90$, the charging hour must be at least 10% cleaner than the discharging hour. In 2024, the mean daily profile ran from 0.086 $kgCO_2e/kWh$ at midday to 0.134 in the evening, a spread of 55%, so this condition is met on an average day with a wide margin. Wang et al. [49] showed that control based on time-varying emission signals lowers operational carbon in an all-electric community. Under carbon-forecast control, we would therefore expect lower dynamic carbon than in Scenarios 3 and 3.1, and a smaller temporal carbon performance gap, because the battery would target the winter evening hours that drive the gap. Such a battery still could not move summer solar into winter, where the largest gradient lies (0.075 $kgCO_2e/kWh$ in July against 0.175 in November).

Two cautions apply to both alternatives. First, grid charging adds round-trip losses, so it raises gross delivered energy and net annual EUI while it lowers carbon. The energy and carbon indicators would then move in opposite directions, as Amaripadath et al. [57] reported for batteries in net zero energy homes. Second, the expected savings hold under hourly average factors. Under marginal factors, the result can be different, and storage can even increase emissions [50,55] (Section 5.3).

Neither alternative was simulated in this study. The self-consumption rule was kept because most residential batteries are sold with this setting (Section 2.4), so it reflects how batteries are used today. A comparison of the three strategies on the same calibrated model is part of the future work in Section 5.5.

5.3. Average versus marginal emission factors

All results in this article are attributional. They use hourly average factors, which describe the mix of electricity supplied in each hour. A consequential assessment asks a different question: which extra or avoided emissions does the retrofit cause in the power system? It uses marginal factors, which describe the generators that respond to a change in demand [22,23]. The two can give different answers. Siler-Evans et al. [84] showed for the United States that marginal factors follow a different daily pattern from average factors, and that average factors can misestimate avoided CO_2 emissions by up to 35%. Fleschutz et al. [83] showed that average factors can misjudge the emissions avoided by demand-side measures in European countries as well. Hawkes [85] and Gagnon and Cole [86] add a further distinction between short-run marginal factors, which reflect current dispatch, and long-run marginal factors, which also account for the power plants built or closed in response to new demand.

Would marginal accounting change the ranking of the pathway? For the heat pump step, it is unlikely. The heat pump replaces gas burned at 0.2016 $kgCO_2e/kWh$ in a boiler with an efficiency of 0.94. At an SPF of 3.20, it saves carbon on space heating as long as the marginal factor stays below about 0.69 $kgCO_2e/kWh$ ($0.2016 \times 3.20 / 0.94$). For hot water, at an SPF of 2.50, the limit is about 0.54 $kgCO_2e/kWh$. Belgium closed its last coal plant in 2016, and gas plants now provide most of its flexible generation, so the marginal unit in most heating hours is likely to be a gas plant, with a factor below these limits. Only if coal or lignite plants in neighbouring countries were on the margin for many heating hours would the heat pump lose its advantage. The solar and battery steps are less certain. Under marginal accounting, midday solar may displace a gas plant rather than the clean average mix, which would make solar worth more than the average factor suggests. In hours of renewable surplus, however, the marginal unit may be curtailed wind or solar, and the avoided emissions would then be close to zero.

Does residual-import carbon intensification hold under marginal accounting? It depends on the daily shape of the marginal factor. The mechanism needs the midday imports that solar displaces to be cleaner than the imports that remain. For average factors, the Belgian data show this clearly. Marginal factors do not have to follow the same shape. Siler-Evans et al. [84] found regions where marginal factors were highest at night and others where they peaked in the early morning. If the Belgian marginal factor were flat over the day, the intensification effect would become weak. If it were higher at midday than in the evening, the effect would reverse, and solar would look better, not worse, than annual accounting suggests. We therefore do not claim that the mechanism holds under marginal factors. It is a finding for attributional accounting, which is the basis of energy performance certificates. Testing it with hourly marginal factors for Belgium, estimated with the methods of Ben Amor et al. [22], is a clear next step (Section 5.5).

5.4. Strengths and limitations

Four things make this study stronger than a single case result. The first is that it does not stand alone. Three separate lines of work now point the same way using methods that share nothing with each other. Bistline and Watten [20] used power sector capacity planning models for the United States and concluded that rooftop solar emission reductions are overstated when substitution effects are handled poorly. Brodnicke et al. [19] used mixed-integer optimisation and found that a constant annual factor can overestimate the abatement from building solar by a factor of two. This study used a calibrated building model checked against four years of metered consumption in a European regulatory setting. Different scales, different tools, same direction. A finding that survives that kind of triangulation is much harder to dismiss than one resting on a single method.

The second is that this study provides a way to check its own answer, which the field currently lacks. Most dynamic carbon studies report a reduction and offer no means of testing whether it is physically possible. The check in Table 12 works out the emission factor implied by each step of the scenario ladder and compares it against the observed spread of hourly factors. If a step appears to displace electricity dirtier than the dirtiest hour of the year, the accounting is wrong somewhere. The three steps here gave 0.090, 0.112 and 0.107 kgCO2e/kWh against a maximum observed hourly value of 0.294, so all three pass. The test

costs one line of arithmetic, and any study in this area can apply it.

The third is the way the model was calibrated. Most calibrated building models are tuned until they match one annual number, which leaves the split between end uses unchecked. A model can reproduce annual gas exactly while attributing it wrongly between heating and hot water. Morewood [59] set out why data quality conditions of this kind decide whether monitored performance can be trusted. Two constraints used here do not depend on the annual total at all. The summer gas floor, measured between June and August when no space heating happens, gives hot water demand on its own, and the model returned 1634 kWh against a measured 1617 kWh. The measured July electricity of 5.27 kWh/day sets a ceiling that no appliance schedule can exceed. The model is therefore pinned on its composition and not only on a single figure it was fitted to.

The fourth is where the argument is aimed. Bistline and Watten [20] speak to national emission accounting. Brodnicke et al. [19] speak to optimal system design. Neither addresses what an energy performance certificate reports to a homeowner, a lender, or a regulator. This study does, and that is where the accounting rule has practical consequences. Annual accounting attributes a larger carbon benefit than the timing and carbon intensity of the avoided imports support.

Five limitations bound these results. The first is scope. One dwelling was modelled over one reference weather year, so there is no spread across the stock and no variation between years. Fig. 7 shows where the benchmark sits within the 708 audited dwellings, which gives some assurance, but a single case is a single case.

The second is the accounting boundary, and it needs stating plainly. This study uses hourly average emission factors, which are the right basis for attributional reporting and therefore for a certificate. They do not measure the marginal system emissions caused or avoided by a retrofit. Ben Amor et al. [22] set out methods for estimating hourly marginal factors, and Gallardo et al. [24] argue that the factor should be matched to the decision rather than treated as generally superior. A consequential question about what this retrofit causes the Belgian power system to do would need marginal factors and would give a different number. Section 5.3 discusses how the ranking and the mechanism could change. The results here should be read as attributional throughout. The pathway ranking reported here is based on energy and operational-carbon performance rather than economic optimality. Capital costs, electricity and gas tariffs, maintenance, equipment replacement and payback were not included. These factors may alter the preferred sequence or sizing of renovation measures, particularly for PV and battery storage, whose incremental carbon benefits diminish as the pathway progresses. Future work should therefore combine time-resolved carbon assessment with life-cycle cost analysis to identify pathways that are both carbon-effective and economically viable.

A third consideration bounds transferability rather than validity. The mechanism needs a grid whose cleanest hours are sunny hours. Belgium has one, because solar output now depresses midday carbon intensity against a large low-carbon baseload. Not every grid works that way, and hourly signals for most European countries show different shapes [11].

In wind-dominated grids such as Denmark and Ireland, the cleanest hours are windy hours. Wind output has no fixed daily pattern and is usually higher in winter. Solar self-consumption would then remove midday imports that are not much cleaner than the rest, so the intensification effect would be weak and the solar part of the gap would be small. The heat pump part of the gap could also shrink, or even become negative, because heating demand falls in the windy months when the grid is cleaner than its annual mean.

In nuclear-dominated grids such as France, the daily profile is flatter and the annual mean is low. The absolute gap in $kgCO_2e/(m^2 \cdot a)$ would be small. The relative gap would not necessarily disappear, because winter peaks are still met partly by gas plants and imports, so heating loads would still carry more carbon than the annual mean suggests. The part of the gap caused by solar would be smaller, since midday hours are not much cleaner than other hours.

In hydro-dominated grids such as Norway, carbon intensity is very low in all hours, and the cleanest hours may fall at night, when demand is low. Solar would then displace imports that are no cleaner, or even dirtier, than the imports that remain. The intensification effect would disappear or reverse, and both the absolute and the relative gap would be small.

These are expected directions and not computed results. The method can test them directly by keeping the calibrated Belgian dwelling and replacing the Belgian hourly series with the 2024 series of each country, as proposed in Section 5.5. The method transfers everywhere. The specific result transfers where the daily carbon minimum is driven by solar.

A fourth limitation is the efficiency of the heat pump. The SPF values of 3.20 for space heating and 2.50 for hot water are assumptions, and field data suggest that real systems can perform lower (Section 3.7). The model applies a constant SPF, so a lower value raises heat pump electricity in every hour by the same factor and does not change its hourly shape. With an SPF of 2.80 for space heating and 2.20 for hot water, for example, heat pump electricity for these two uses would rise from 22.8 to 26.0 $kWh/(m^2 \cdot a)$, and gross delivered energy from 48.5 to 51.7 $kWh/(m^2 \cdot a)$. Most of this extra electricity (2.5 of 3.2 $kWh/(m^2 \cdot a)$) is for space heating and falls in the heating season, when the grid is dirtier than its annual mean and the solar array produces little. To test the effect on carbon, the hourly calculation was repeated with these lower values (Table 13). Grid imports rose by 3.1 to 3.2 $kWh/(m^2 \cdot a)$ and dynamic operational carbon by about 0.43 $kgCO_2e/(m^2 \cdot a)$ in every scenario. The temporal carbon performance gap changed by 0.5 percentage points or less, from 18.4% to 18.8% in Scenario 1 and from 30.3% to 30.5% in Scenario 3.1. The extra heat pump electricity has almost the same carbon intensity as the imports it joins, so it adds carbon without changing their average. A lower SPF therefore raises the carbon of every scenario by about the same amount but leaves the gap, and its growth along the pathway, almost unchanged. Heat-pump electrification remains the largest single reduction, at 64% instead of 67% below the Base Case, and the full pathway reduces operational carbon by 73% instead of 75%.

A fifth limitation is the system boundary. This study assesses operational carbon only, which is what an energy performance certificate reports. The heat pump, PV array, inverter and batteries also carry embodied carbon from manufacture, transport, replacement and end of life. These emissions were not counted, so the net carbon benefit of each step is overestimated. The effect is likely to matter most for the later steps. The two batteries save only 0.43 and 0.74 $kgCO_2e/(m^2 \cdot a)$ of operational carbon compared with Scenario 2, and they will likely need replacing during the life of the building. Evidence from the Dutch housing stock shows that the embodied emissions of renovation can offset a large part of its operational savings [87]. Fadden et al. [40] likewise found that the life cycle impact of heat pump systems depends

Table 13
Sensitivity to lower heat pump SPF values: 2.80 for space heating and 2.20 for hot water, instead of 3.20 and 2.50. All other inputs are as in Table 11. Energy in $kWh/(m^2 \cdot a)$, carbon in $kgCO_2e/(m^2 \cdot a)$. The last column gives the change in the temporal carbon performance gap compared with Table 11, in percentage points.

Scenario	Gross EUI	Net EUI	Grid imports	Static	Dynamic	Difference	Change in gap
Scenario 1	51.7	51.7	51.7	5.53	6.58	+18.8%	+0.4
Scenario 2	51.7	34.9	43.1	4.61	5.80	+25.9%	+0.3
Scenario 3	51.7	34.9	39.3	4.21	5.38	+27.9%	+0.1
Scenario 3.1	51.7	34.9	36.3	3.89	5.07	+30.5%	+0.2

strongly on the system boundary. A whole-life assessment could therefore change the preferred sequence or size of the measures, for example by favouring a smaller battery or no battery. The results reported here should not be read as the net decarbonisation potential of the pathway.

5.5. Implications for practice and future work

The dwellings this study speaks to are numerous and share a recognisable set of properties. They are low-rise, one or two storeys with a pitched roof, semi-detached or terraced, and built between roughly 1995 and 2010. Walls and roofs were insulated to the regulations of that period, so envelope losses are moderate rather than severe, with wall U-values near 0.25 and roof values near 0.23 W/(m2·K). Glazing is the weak point, usually clear double units in metal frames at around 2.8 W/(m2·K). Airtightness is ordinary. Heating and hot water come from a gas boiler. Roofs are simple, unshaded and large enough for solar. Belgium alone holds 64,535 of them from the 2000 to 2010 decade [6], and the same form dominates the suburban stock of the Netherlands, the United Kingdom and Ireland.

Two things are happening to this group at once. Heating demand is falling, because the envelope is already reasonable and the climate is warming. Cooling demand is rising, because summers are hotter and these dwellings have large glazed areas, lightweight upper floors under a pitched roof, and no shading. The reversible heat pump modelled here already met 7.6 kWh/(m^2·a) of useful cooling demand, corresponding to approximately 2.2 kWh/(m^2·a) of delivered electricity at the modelled SEER of 3.5. Fischer and Madani [32] describe how the timing of heat pump loads shapes their grid consequences, and cooling adds a second load with the opposite seasonal pattern. The balance of what matters is therefore shifting. Because envelope renovation was outside the study boundary, the present results do not establish the relative performance or cost-effectiveness of additional envelope measures. Future work could assess how measures such as improved glazing, airtightness and solar shading interact with the energy-system pathways examined here.

What remains is the system. Heating, cooling, hot water, generation and storage now decide both the energy and the carbon outcome of these dwellings, and they decide it through the timing of electricity use rather than the size of the annual demand. This is a real change in where the engineering effort belongs. It also changes what the certificate should measure, because an indicator built to describe envelope performance is not the right instrument for a building whose carbon depends on when it draws power.

The data needed for that shift already exists. Belgium introduced 15-min interval metering for electricity in 2026, and the transmission system operator publishes generation at the same resolution [9]. The European imbalance settlement period is likewise 15 min. Nothing in this study required a special measurement campaign. It used hourly consumption of a kind a utility already holds and an hourly grid series that is published openly. Crawley et al. [58] showed that smart meter data can support dwelling performance ratings, and Smertinas et al. [60] demonstrated digital certification from measured data at national scale. A certification body could compute an hourly carbon indicator today from data collected anyway. The obstacle is not measurement or cost. It is that no rule asks for it.

Five lines of work follow. The first is temporal. The Belgian grid will keep getting cleaner, and its daily profile will flatten as storage and interconnection grow, so the gap measured here will change. Running the same benchmark against projected emission factors would show whether the effect grows, shrinks or reverses as the system decarbonises. The second is control. The same calibrated model should compare self-consumption, cost-minimising and carbon-forecast control of the battery (Section 5.2). Quarter-hourly settlement and dynamic tariffs make genuine carbon-aware operation possible, where a solar-surplus rule cannot be. Wilczynski et al. [33] show that the flexibility a household actually activates depends on the tariff it faces, with the widest difference between a flat rate and real-time pricing, so the signal matters as much as the hardware. Schibuola and Tambani [48] quantify how much generation storage can retain. A battery charging against a carbon forecast rather than against sunshine would reach the winter evening hours that the one modelled here cannot touch. The third is geographic. A first step is to rerun the Belgian benchmark with the 2024 hourly series of wind-, nuclear- and hydro-dominated grids such as Denmark, Ireland, France and Norway (Section 5.4). The methodological framework can be transferred to other countries by combining local weather and hourly grid emission data with representative building, system and occupancy inputs. Applying the framework across Belgium, the Netherlands, the United Kingdom and Ireland would turn the Belgian case study into a comparative assessment and test how strongly the findings depend on differences in building stock, climate and grid characteristics.

The fourth is accounting. The scenarios should be rerun with hourly marginal factors for Belgium [22,84] to test whether residual-import carbon intensification also appears under consequential accounting (Section 5.3). The fifth is the system boundary. A whole-life assessment that adds the embodied carbon of the heat pump, PV system and batteries [40,87] would show whether the later steps of the pathway save more carbon than they cause.

6. Conclusions

This study evaluated how progressive building energy-system renovation pathways, from gas heating through heat-pump electrification to rooftop PV and battery storage, affect delivered energy, grid imports and operational carbon, and whether conventional annual indicators adequately capture these changes compared with time-resolved carbon accounting. The analysis used a representative Belgian semi-detached dwelling selected from 708 audited households and calibrated against four years of measured gas and electricity consumption. Five configurations were assessed under consistent building, occupancy and operating conditions, allowing the effects of successive energy-system interventions to be isolated.

The renovation pathway produced substantial reductions in both energy use and operational carbon. Replacing the gas boiler with a reversible heat pump reduced gross delivered energy from 102.2 to 48.5 kWh/(m^2·a) and dynamic operational carbon from 18.46 to 6.14 $kgCO_2e/(m^2·a)$, making electrification the dominant decarbonization step. Adding rooftop PV and battery storage further reduced grid dependence, with imports declining from 40.0 kWh/(m^2·a) without storage to 36.2 and 33.2 kWh/(m^2·a) with 2 and 5 kWh batteries, respectively. Dynamic operational carbon ultimately fell to 4.64 $kgCO_2e/(m^2·a)$, representing a 75% reduction relative to the Base Case.

The results also reveal an important limitation of annual performance indicators. Net annual EUI remained fixed at 31.7 kWh/(m^2·a) across all three PV configurations, despite grid imports falling by 17% and operational carbon by 14%. At the same time, the temporal carbon performance gap increased progressively from 3.0% in the Base Case to 18.4% following heat-pump electrification, 25.6% with PV, and 30.3% with the 5-kWh battery. This divergence arises because behind-the-meter generation and storage change not only how much electricity is imported but also when it is imported. The load-weighted carbon intensity of the remaining imports increased from 0.1261 to 0.1394 $kgCO_2e$/kWh as the dwelling became increasingly grid interactive.

These findings have two implications for residential decarbonization. First, heat-pump electrification, rooftop PV and battery storage can form an effective sequence of building energy-system interventions, although the incremental operational-carbon benefit diminishes along the pathway. Second, annual energy balances alone become increasingly insufficient for distinguishing the performance of electrified, generating and storage-equipped dwellings. Assessment frameworks should therefore complement annual energy indicators with grid imports and time-resolved operational carbon so that the temporal consequences of renovation are visible.

The numerical results remain specific to the benchmark dwelling, its

occupancy, the Brussels climate and the Belgian grid conditions represented in the analysis. The study does not evaluate envelope renovation, embodied carbon, investment costs, energy prices, payback periods or life-cycle economic performance. It uses hourly average emission factors, so its results are attributional. It should therefore be interpreted as an assessment of building energy-system decarbonization pathways rather than whole-building deep renovation or techno-economic optimization. Nevertheless, the framework can be transferred by combining representative local building and occupancy characteristics with local weather and time-resolved grid emission data. Future work should extend this approach across dwelling typologies and electricity systems and investigate carbon-responsive control of heat pumps and batteries. Such extensions would establish when time-resolved carbon assessment materially changes renovation decisions and how it could be incorporated into emerging performance certification frameworks.

CRediT authorship contribution statement

Marziyeh Khoshroo: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Deepak Amaripadath:** Writing – review & editing, Validation, Supervision. **Tarek M.F. Ahmed:** Writing – review & editing, Validation, Methodology. **Shady Attia:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors express their thanks to all survey respondents. The Walloon Region partially funded this research under the call ‘Actions de Recherche Concertées 2019 (ARC)’ and the project OCCuPANt, on the Impacts Of Climate Change on the indoor environmental and energy PerformAnce of buildiNgs in Belgium during summer. The authors would like to acknowledge the Walloon Region and Liege University for their funding, gratefully. We would also like to recognise the Sustainable Building Design Lab for the use of monitoring equipment in this research and the valuable support during the experiments and data analysis.

We would like to acknowledge the Sustainable Building Design Lab for the use of the state-of-the-art Super COmputeR ProcessIng wOrkstatioN (SCORPION) with a processor of 64 cores, 128 threads, and a 256 MB cache for computing power and performance. The 128GB (4 x 32GB) of memory (RAM) and NVIDIA GeForce graphics card of 24GB allowed to perform the intensive computations.

Appendix A

Hour	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1:00	128.63	103.91	125.78	96.22	77.07	70.23	66.13	73.88	91.98	92.15	174.65	132.90
2:00	128.08	104.54	127.01	95.61	72.11	70.20	63.46	75.12	89.67	88.60	170.40	128.35
3:00	123.77	99.72	123.42	94.95	72.33	70.40	64.36	72.82	88.51	88.30	165.74	121.55
4:00	123.56	98.45	122.13	98.63	73.93	75.55	65.20	73.37	93.07	92.82	162.82	116.02
5:00	128.23	102.21	129.70	101.49	76.99	80.74	73.34	79.22	104.09	110.30	165.86	117.86
6:00	135.29	110.22	138.38	104.89	84.88	80.03	77.08	79.59	108.50	124.54	169.42	123.47
7:00	142.79	117.42	140.90	101.90	79.51	73.59	72.91	77.94	104.86	124.38	172.83	131.24
8:00	146.79	118.32	132.42	92.75	77.57	67.42	69.25	72.32	92.36	114.83	171.43	138.98
9:00	145.40	115.95	121.31	79.64	72.84	68.25	66.92	64.42	80.73	98.93	165.16	141.27
10:00	138.43	109.12	107.24	72.63	68.14	63.93	61.69	58.71	72.82	86.00	159.33	136.50
11:00	130.92	105.82	98.50	70.20	67.35	64.44	58.86	57.12	70.34	82.64	152.82	131.93
12:00	123.66	97.21	96.41	68.92	64.59	62.42	56.94	53.87	67.15	82.31	151.70	131.26
13:00	120.71	93.92	91.66	64.79	63.01	62.92	56.66	52.26	64.34	83.58	150.45	129.99
14:00	124.00	96.41	94.76	66.00	63.07	62.36	59.10	53.94	65.21	87.39	154.92	130.13
15:00	128.62	98.97	103.28	71.08	65.34	64.58	61.22	56.72	71.43	96.62	165.14	132.73
16:00	135.88	103.79	118.54	80.07	67.76	69.47	68.27	65.64	82.95	118.82	177.60	136.93
17:00	143.45	111.79	135.01	91.28	78.38	82.43	78.68	81.51	102.19	135.09	185.87	145.63
18:00	152.96	121.19	155.12	104.03	87.35	96.43	93.40	98.22	120.51	147.19	197.99	153.86
19:00	156.34	126.22	166.21	111.42	94.77	102.63	106.18	107.29	125.21	147.70	204.53	156.35
20:00	152.55	124.27	158.22	111.76	93.69	105.96	109.83	110.40	118.34	140.16	202.96	153.25
21:00	147.51	116.01	150.19	112.49	97.67	112.63	116.84	109.14	116.03	132.56	200.00	149.82
22:00	143.31	116.12	146.90	109.97	94.15	108.51	109.01	99.68	110.48	126.79	201.05	149.54
23:00	138.03	117.99	137.87	101.89	81.53	86.42	85.80	84.51	99.00	109.03	199.65	146.84
0:00	131.30	107.87	128.51	98.29	77.97	74.08	69.95	77.32	95.67	101.83	179.65	135.02
Avg gCO2e	136.26	109.18	127.06	91.70	77.17	78.15	75.46	76.46	93.14	108.86	175.08	136.31

Fig. A. Average hourly electricity grid carbon intensity for each month in Belgium 2024 expressed in gCO_2e/kWh.

Appendix B

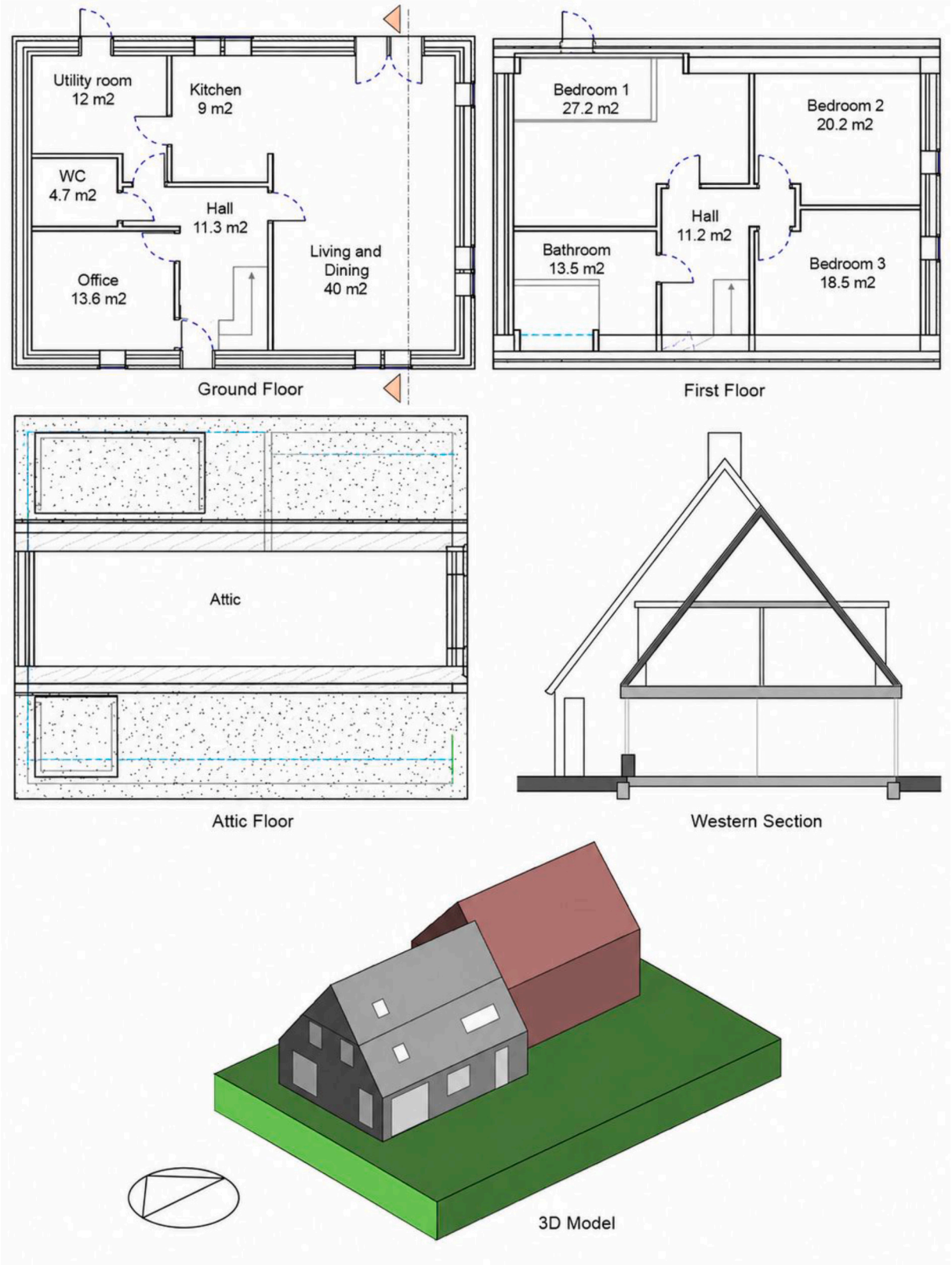


Fig B. Representative Belgian semi-detached residential buildings and benchmark reference house model, including housing stock distribution, 3D geometry, floor plans, and building section.

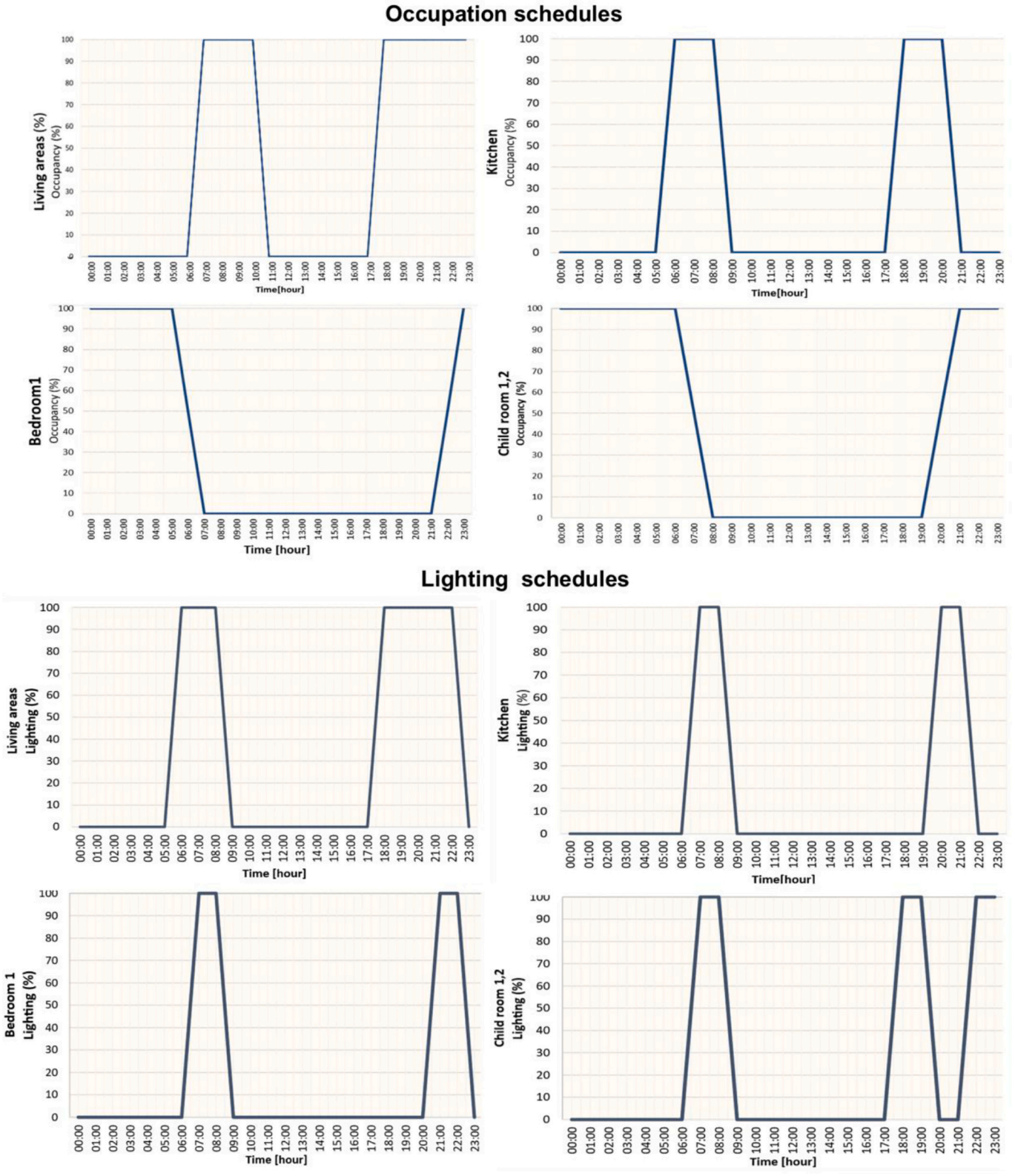


Fig C. Occupancy and lighting operation schedules adopted for the calibrated benchmark residential model.

Data availability

Data is available: Khoshroo, Marziyeh; Attia, Shady, 2025, "Development of a nearly zero-energy semidetached house in Belgium", https://doi.org/10.7910/DVN/GOLIDT, Harvard Dataverse, V1

References

[1] International Energy Agency. Overview of energy efficiency indicators: IEA energy end-use and efficiency trends. 2026. . Available at: International Energy Agency (accessed 18 August 2026).

[2] European Commission. *State of the union: Commission raises climate ambition* [press release]. 2026. . Available at: European Commission Press Corner (accessed 18 August 2026).

[3] European Parliament and Council of the European Union. Directive (EU) 2024/1275 of 24 April 2024 on the energy performance of buildings (recast). *Official Journal of the European Union*; 2024. . Available at: EUR-Lex (accessed 18 August 2026).

[4] Maduta C, D'Agostino D, Tsemekidi-Tzeiranaki S, Castellazzi L. From nearly zero-energy buildings (NZEBs) to zero-emission buildings (ZEBs): current status and future perspectives. Energ Buildings 2025;328:115133. https://doi.org/10.1016/j.enbuild.2024.115133.

[5] Rosenow J. Beyond supply: the case for decarbonising energy demand. PLOS Climate 2025;4(3):e0000590. https://doi.org/10.1371/journal.pclm.0000590.

[6] Statbel. Cadastral statistics of the building stock. 2026. . Available at: Statbel (accessed 18 August 2026).

[7] Attia S, Eleftheriou P, Xeni F, Morlot R, Ménézo C, Kostopoulos V, et al. Overview and future challenges of nearly zero energy buildings (nZEB) design in southern Europe. Energ Buildings 2017;155:439–58. https://doi.org/10.1016/j.enbuild.2017.09.043.

[8] European Commission, Joint Research Centre. Belgium: Status of the heat pump market *(country fiche JRC137131)*. Publications Office of the European Union; 2024. . *[*Available at: European Commission JRC (accessed 18 August 2026)..

[9] Electricity Maps. Belgium. 2026. . Available at: Electricity Maps (accessed 18 August 2026)..

[10] Noussan M, Fambri G, Negro V, Chiaramonti D. Hourly electricity CO_2 intensity profiles based on the real operation of large-scale natural gas combined cycle cogeneration plants. Energy 2024;312:133424. https://doi.org/10.1016/j.energy.2024.133424.

[11] Unnewehr JF, Weidlich A, Gfüllner L, Schäfer M. Open-data based carbon emission intensity signals for electricity generation in European countries: top down vs. bottom up approach. Cleaner Energy Syst 2022;3:100018. https://doi.org/10.1016/j.cesys.2022.100018.

[12] Amaripadath D, Sailor DJ, Bertini A, Barker M, Attia S. Are net zero energy buildings necessarily also net zero emission buildings? Time-integrated analysis using dynamic grid emission factors. Build Environ 2025;283:113367. https://doi.org/10.1016/j.buildenv.2025.113367.

[13] Bontekoe E, Schade J, Erikkson L, Tsarchopoulos P, Lampropoulos I, Van Sark W. On the discrepancy of using annual or hourly emission factors for power generation to estimate CO_2 reduction of building retrofitting. Energ Buildings 2024;319:114499. https://doi.org/10.1016/j.enbuild.2024.114499.

[14] Zou A, Raftery P, Roberts M, Wang Y, Lv G, Duarte C, et al. Temporal resolution matters: evaluating carbon emission factors for accurate accounting in commercial buildings. Appl Energy 2026;412:127655. https://doi.org/10.1016/j.apenergy.2026.127655.

[15] Lou Y, Ye Y, Yang Y, Zuo W. Long-term carbon emission reduction potential of building retrofits with dynamically changing electricity emission factors. Build Environ 2022;210:108683. https://doi.org/10.1016/j.buildenv.2021.108683.

[16] Li C, Pan Y, Liu Z, Liang Y, Yuan X, Huang Z, et al. Optimal design of building envelope towards life cycle performance: impact of considering dynamic grid emission factors. Energ Buildings 2024;323:114770. https://doi.org/10.1016/j.enbuild.2024.114770.

[17] Álvarez Flórez L, Péan T, Salom J. Hourly based methods to assess carbon footprint flexibility and primary energy use in decarbonized buildings. Energ Buildings 2023;294:113213. https://doi.org/10.1016/j.enbuild.2023.113213.

[18] Bertini A, Al-Obaidy M, Dasse M, Amaripadath D, Gobbo E, Attia S. Parametrization of variables affecting the whole life carbon performance of nearly zero energy residential building renovation. Build Environ 2025;278:113013. https://doi.org/10.1016/j.buildenv.2025.113013.

[19] Brodnicke L, Ganter A, Tröber S, Sansavini G, Vulic N. Dynamic grid emission factors and export limits reduce emission abatement and cost benefits of building PV systems. Energ Buildings 2024;323:114772. https://doi.org/10.1016/j.enbuild.2024.114772.

[20] Bistline JET, Watten A. Emissions reductions of rooftop solar are overstated by approaches that inadequately capture substitution effects. Nat Clim Chang 2025;15(11):1173–5. https://doi.org/10.1038/s41558-025-02459-y.

[21] Bayer DR, Pruckner M. Data-driven heat pump retrofit analysis in residential buildings: carbon emission reductions and economic viability. Appl Energy 2024;373:123823. https://doi.org/10.1016/j.apenergy.2024.123823.

[22] Ben Amor S, Sgarciu S, Batz Lineiro T, Musgens F. Advanced models for hourly marginal CO_2 emission factor estimation: a synergy between fundamental and statistical approaches. Appl Energy 2025;397:126265. https://doi.org/10.1016/j.apenergy.2025.126265.

[23] Goldsworthy M, Aryai V. On the use of time varying marginal and time varying average emissions factors in building energy analysis. Sustainable Energy, Grids and Networks 2023;35:101103. https://doi.org/10.1016/j.segan.2023.101103.

[24] Gallardo A, Dumitrascu L, Dandres T, Macdonald I, Knudsen H. Ten questions concerning operational greenhouse gas (GHG) emissions accounting for electricity use in buildings. Build Environ 2026;300:114689. https://doi.org/10.1016/j.buildenv.2026.114689.

[25] Heo Y, Choudhary R, Augenbroe GA. Calibration of building energy models for retrofit analysis under uncertainty. Energ Buildings 2012;47:550–60. https://doi.org/10.1016/j.enbuild.2011.12.029.

[26] Greer F, Raftery P, Horvath A. Considerations for estimating operational greenhouse gas emissions in whole building life-cycle assessments. Build Environ 2024;254:111383. https://doi.org/10.1016/j.buildenv.2024.111383.

[27] Majcen D, Itard LCM, Visscher H. Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: discrepancies and policy implications. Energy Policy 2013;54:125–36. https://doi.org/10.1016/j.enpol.2012.11.008.

[28] Majcen D, Itard L, Visscher H. Actual and theoretical gas consumption in Dutch dwellings: what causes the differences? Energy Policy 2013;61:460–71. https://doi.org/10.1016/j.enpol.2013.06.018.

[29] Majcen D, Itard L, Visscher H. Statistical model of the heating prediction gap in Dutch dwellings: relative importance of building, household and behavioural characteristics. Energ Buildings 2015;105:43–59. https://doi.org/10.1016/j.enbuild.2015.07.009.

[30] van den Brom P, Meijer A, Visscher H. Performance gaps in energy consumption: household groups and building characteristics. Build Res Inf 2018;46(1):54–70. https://doi.org/10.1080/09613218.2017.1312897.

[31] van der Bent HS, van den Brom PI, Visscher HJ, Meijer A, Mouter N. The energy performance of dwellings of Dutch non-profit housing associations: modelling actual energy consumption. Energ Buildings 2021;253:111486. https://doi.org/10.1016/j.enbuild.2021.111486.

[32] Fischer D, Madani H. On heat pumps in smart grids: a review. Renew Sust Energ Rev 2017;70:342–57. https://doi.org/10.1016/j.rser.2016.11.182.

[33] Wilczynski EJ, Chambers J, Patel MK, Worrell E, Pezzutto S. Assessment of the thermal energy flexibility of residential buildings with heat pumps under various electric tariff designs. Energ Buildings 2023;294:113257. https://doi.org/10.1016/j.enbuild.2023.113257.

[34] Wang J, El Kontar R, Jin X, King J. Electrifying high-efficiency future communities: impact on energy, emissions, and grid. Adv Appl Energy 2022;6:100095. https://doi.org/10.1016/j.adapen.2022.100095.

[35] Taylor J, McLeod R, Petrou G, Hopfe C, Mavrogianni A, Castaño-Rosa R, et al. Ten questions concerning residential overheating in central and northern Europe. Build Environ 2023;234:110154. https://doi.org/10.1016/j.buildenv.2023.110154.

[36] Erbey MM, Tumen Ozdil NF. Forecasting the European Union's space cooling potential using the cooling degree-day factor (1970–2024). Energy 2026;344:140113. https://doi.org/10.1016/j.energy.2026.140113.

[37] Saikia P, Corcoran L, Ugalde-Loo CE, Abeysekera M. A cooling demand estimator for housing communities in a warming world. Appl Energy 2025;377:124597. https://doi.org/10.1016/j.apenergy.2024.124597.

[38] Rahif R, Hamdy M, Homaei S, Zhang C, Holzer P, Attia S. Simulation-based framework to evaluate resistivity of cooling strategies in buildings against overheating impact of climate change. Build Environ 2022;208:108599. https://doi.org/10.1016/j.buildenv.2021.108599.

[39] Reynders G, Diriken J, Saelens D. Generic characterization method for energy flexibility: applied to structural thermal storage in residential buildings. Appl Energy 2017;198:192–202. https://doi.org/10.1016/j.apenergy.2017.04.061.

[40] Fadden L, Radcliffe J, Mehta N. Decarbonising residential heating: A systematic review of life cycle impacts for residential and district heat pump systems *[Conference abstract]*. EGU General Assembly 2026, EGU26-1219. Doi:. 2026. https://doi.org/10.5194/egusphere-egu26-1219.

[41] Salom J, Marszal AJ, Widén J, Candanedo J, Lindberg KB. Analysis of load match and grid interaction indicators in net zero energy buildings with simulated and monitored data. Appl Energy 2014;136:119–31. https://doi.org/10.1016/j.apenergy.2014.09.018.

[42] Cho M, Honda T. CO_2 emissions in electrified detached houses with photovoltaic generators: an hourly emission-profile approach with demand-side management and battery integration. Energy Convers Manage: X 2026;30:101642. https://doi.org/10.1016/j.ecmx.2026.101642.

[43] Paulus M, Borggrefe F. The potential of demand-side management in energy-intensive industries for electricity markets in Germany. Appl Energy 2011;88(2):432–41. https://doi.org/10.1016/j.apenergy.2010.03.017.

[44] Masciandaro C, Mulder M, Kesina M. Distributional effects of the Dutch net-metering scheme for residential solar panels. Energy Econ 2025;151:108891. https://doi.org/10.1016/j.eneco.2025.108891.

[45] Mihetec M, Stunjek G, Krajačić G, Mikulčić Krnjaja G. A suburban energy transition: deploying heat pump–photovoltaic systems as a cost-effective alternative to district heating in low-density settlement. Energy 2026;360:141630. https://doi.org/10.1016/j.energy.2026.141630.

[46] Van Der Stelt S, AlSkaif T, Van Sark W. Techno-economic analysis of household and community energy storage for residential prosumers with smart appliances. Appl Energy 2018;209:266–76. https://doi.org/10.1016/j.apenergy.2017.10.096.

[47] Uddin K, Gough R, Radcliffe J, Marco J, Jennings P. Techno-economic analysis of the viability of residential photovoltaic systems using lithium-ion batteries for energy storage in the United Kingdom. Appl Energy 2017;206:12–21. https://doi.org/10.1016/j.apenergy.2017.08.170.

[48] Schibuola L, Tambani C. Impact of electrical energy storage on the performance of photovoltaic systems in building retrofits. Energ Buildings 2026;351:116755. https://doi.org/10.1016/j.enbuild.2025.116755.

[49] Wang J, Munankarmi P, Maguire J, Shi C, Zuo W, Roberts D, et al. Carbon emission responsive building control: a case study with an all-electric residential community in a cold climate. Appl Energy 2022;314:118910. https://doi.org/10.1016/j.apenergy.2022.118910.

[50] Pimm AJ, Palczewski J, Barbour ER, Cockerill TT. Using electricity storage to reduce greenhouse gas emissions. Appl Energy 2021;282:116199. https://doi.org/10.1016/j.apenergy.2020.116199.

[51] Lorenzi G, Silva CAS. Comparing demand response and battery storage to optimize self-consumption in PV systems. Appl Energy 2016;180:524–35. https://doi.org/10.1016/j.apenergy.2016.07.103.

[52] Hoppmann J, Volland J, Schmidt TS, Hoffmann VH. The economic viability of battery storage for residential solar photovoltaic systems: a review and a simulation model. Renew Sust Energ Rev 2014;39:1101–18. https://doi.org/10.1016/j.rser.2014.07.068.

[53] Nygård HS, Ottesen SØ, Skonnord OH. Profitability analyses for residential battery investments: a Norwegian case study. Energies 2024;17(16):4048. https://doi.org/10.3390/en17164048.

[54] Beaudin M, Zareipour H. Home energy management systems: a review of modelling and complexity. Renew Sust Energ Rev 2015;45:318–35. https://doi.org/10.1016/j.rser.2015.01.046.

[55] Dupont C, Heijmans-Kuryatnikova O, Ghiassi-Farrokhfal Y, Bunn D. The unintended carbon impacts of large-scale electricity storage. Manuf Serv Oper Manag 2025. https://doi.org/10.1287/msom.2024.1411.

[56] SolarPower Europe. European battery market outlook 2026–2030. 2026. . Available at: SolarPower Europe (accessed 18 August 2026)..

[57] Amaripadath D, Sailor DJ, Khan A, Bertini A, Attia S. Achieving low-carbon electricity operations in net zero energy residential buildings through plug-in battery energy storage systems. J Energy Storage 2026;153:120964. https://doi.org/10.1016/j.est.2026.120964.

[58] Crawley J, McKenna E, Gori V, Oreszczyn T. Creating domestic building thermal performance ratings using smart meter data. Buildings & Cities 2020;1(1):1–13. https://doi.org/10.5334/bc.7.

[59] Morewood J. Building energy performance monitoring through the lens of data quality: a review. Energ Buildings 2023;279:112701. https://doi.org/10.1016/j.enbuild.2022.112701.

[60] Smertinas J, Van Hove MYC, Bacher P, Madsen H. Digital energy performance certification of single-family houses in Denmark. Sustain Cities Soc 2026;141:107283. https://doi.org/10.1016/j.scs.2026.107283.

[61] Neukomm M, Nubbe V, Fares R. Grid-interactive efficient buildings *(report no. DOE/GO-102019-5227)*. U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, building technologies office; 2019. . *Available at: U.S. Department of Energy report (*accessed 18 August 2026)..

[62] Burger J, Rosenow J. Flexibility forward: a survey of smart tariffs optimising electric vehicles and heat pumps in Europe. Electr J 2025;38(4):107520. https://doi.org/10.1016/j.tej.2025.107520.

[63] Carroll, P., Silva, F., Tahir, F., O'Regan, B., *&* Lyons, P. *(*2025*).* Smart meter reference load profiles and peak demand models. *In* A. A. Juan, J. Faulin, *&* D. Lopez-Lopez *(Eds.),* Decision sciences (DSA ISC 2024) *lecture notes in Computer science, Vol.* 14778, 361–375. Springer. *Doi:* https://doi.org/10.1007/978-3-031-78238-1_33.

[64] Attia S, Mustafa A, Giry N, Popineau M, Cuchet M, Gulirmak N. Developing two benchmark models for post-world war II residential buildings. Energ Buildings 2021;244:111052. https://doi.org/10.1016/j.enbuild.2021.111052.

[65] Attia S, Canonge T, Popineau M, Cuchet M. Developing a benchmark model for renovated, nearly zero-energy, terraced dwellings. Appl Energy 2022;306:118128. https://doi.org/10.1016/j.apenergy.2021.118128.

[66] Crawley DB, Lawrie LK, Winkelmann FC, Buhl WF, Huang YJ, Pedersen CO, et al. EnergyPlus: creating a new-generation building energy simulation program. Energ Buildings 2001;33(4):319–31. https://doi.org/10.1016/S0378-7788(00)00114-6.

[67] American Society of Heating, *Refrigerating and Air-Conditioning Engineers. (*2023*).* ASHRAE guideline 14-2023: measurement of energy, demand, and water savings. ASHRAE *[*Available at: ASHRAE standards and guidelines (accessed 18 August 2026). (ASHRAE).

[68] Coakley D, Raftery P, Keane M. A review of methods to match building energy simulation models to measured data. Renew Sust Energ Rev 2014;37:123–41. https://doi.org/10.1016/j.rser.2014.05.007.

[69] Raftery P, Keane M, Costa A. Calibrating whole building energy models: detailed case study using hourly measured data. Energ Buildings 2011;43(12):3666–79. https://doi.org/10.1016/j.enbuild.2011.09.039.

[70] Rouchier S. Solving inverse problems in building physics: an overview of guidelines for a careful and optimal use of data. Energ Buildings 2018;166:178–95. https://doi.org/10.1016/j.enbuild.2018.02.009.

[71] International Organization for Standardization. Hygrothermal performance of buildings: Calculation and presentation of climatic data. Part 6: Accumulated temperature differences (degree-days) (ISO 15927-6:2007). *2007. . Available at: ISO 15927-6:2007 (accessed 18 August 2026)*..

[72] Attia S. Belgian electricity grid greenhouse gas emission factors: historical reconstruction 1950 to 2025 and future projections to 2050 *(version 1.1). Sustainable building design laboratory, University of Liège*; 2026. . *ISBN 978-2-930909-30-1. Available at: ORBi, University of Liège (*accessed 18 August 2026)..

[73] Intergovernmental Panel on Climate Change. Annex III: Technology-specific cost and performance parameters. In: Edenhofer O, Pichs-Madruga R, Sokona Y, Farahani E, Kadner S, Seyboth K, Adler A, Baum I, Brunner S, Eickemeier P, Kriemann B, Savolainen J, Schlömer S, von Stechow C, Zwickel T, Minx JC, editors. Climate change 2014: Mitigation of climate change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press*; *2014*. p. *1329–56*.

[74] Gautier A, Jacqmin J, Poudou J-C. The prosumers and the grid. J Regul Econ 2018;53(1):100–26. https://doi.org/10.1007/s11149-018-9350-5.

[75] Protopapadaki C, Saelens D. Heat pump and PV impact on residential low-voltage distribution grids as a function of building and district properties. Appl Energy 2017;192:268–81. https://doi.org/10.1016/j.apenergy.2016.11.103.

[76] Gepts B, Nuyts E, Verbeeck G. Where is the potential for deep (energy) renovations? A large-scale study of the renovation activities of Belgian households between 2018 and 2022. Energ Buildings 2025;328:115150. https://doi.org/10.1016/j.enbuild.2024.115150.

[77] Climafed & Frixis. Warmtepompmarkt herpakt zich: Vier kwartalen groei en uitzicht op doorbraak in 2026. . Climafed. Published February 3*, 2026.* Available from: *https://www.climafed.be/nl/nieuws/n-60-warmtepompmarkt-herpakt-zich-vier-kwartalen-groei-en-uitzicht-op-doorbraak-in-2026/*; 2026 *(*accessed 18 August 2026*)*..

[78] Ritter, D., Rohringer, C., Afshari, A., Andrés-Chicote, M., Hernandez-Cruz, P., Duenzen, K., Bauknecht, D., *&* Attia, S. *(*2026*).* Activating the electrical energy flexibility of residential thermal systems: an analysis of non-technical barriers in six countries. Energy Inform, 9, 59. *Doi:* https://doi.org/10.1186/s42162-026-00673-2.

[79] Schram W, Louwen A, Lampropoulos I, Van Sark W. Comparison of the greenhouse gas emission reduction potential of energy communities. Energies 2019;12(23):4440. https://doi.org/10.3390/en12234440.

[80] O'Hegarty R, Kinnane O, Lennon D, Colclough S. Air-to-water heat pumps: review and analysis of the performance gap between in-use and product rated performance. Renew Sust Energ Rev 2022;155:111887. https://doi.org/10.1016/j.rser.2021.111887.

[81] Gleeson CP, Lowe R. Meta-analysis of European heat pump field trial efficiencies. Energ Buildings 2013;66:637–47. https://doi.org/10.1016/j.enbuild.2013.07.064.

[82] Fraunhofer Institute for Solar Energy Systems ISE. Fraunhofer ISE research project completed: Heat pumps provide climate-friendly heating in existing buildings [Press release]. https://www.ise.fraunhofer.de/en/press-media/press-releases/2025/fraunhofer-ise-research-project-completed-heat-pumps-provide-climate-friendly-heating-in-existing-buildings.html; 2025. accessed 29 September 2026.

[83] Fleschutz M, Bohlayer M, Braun M, Henze G, Murphy MD. The effect of price-based demand response on carbon emissions in European electricity markets: the importance of adequate carbon prices. Appl Energy 2021;295:117040. https://doi.org/10.1016/j.apenergy.2021.117040.

[84] Siler-Evans K, Azevedo IL, Morgan MG. Marginal emissions factors for the U.S. electricity system. Environ Sci Technol 2012;46(9):4742–8. https://doi.org/10.1021/es300145v.

[85] Hawkes AD. Long-run marginal CO_2 emissions factors in national electricity systems. Appl Energy 2014;125:197–205. https://doi.org/10.1016/j.apenergy.2014.03.060.

[86] Gagnon P, Cole W. Planning for the evolution of the electric grid with a long-run marginal emission rate. iScience 2022;25(3):103915. https://doi.org/10.1016/j.isci.2022.103915.

[87] Zhang C, Hu M, Sprecher B, Sacchi R, Yang X, Yang S, et al. Revealing the interplay between decarbonisation, circularity, and cost-effectiveness in building energy renovation. Nat Commun 2025;16:7153. https://doi.org/10.1038/s41467-025-62442-1.