Department for
Energy Security
& Net Zero

Inter-Model Comparison Summary Report

Home Energy Model Validation

December 2023

Acknowledgements

The Home Energy Model has been developed for the Department for Energy Security & Net Zero by a consortium led by the Building Research Establishment (BRE), including AECOM, Sustenic, University of Strathclyde's Energy Systems Research Unit, Kiwa Ltd., Loughborough University Enterprises Limited, Chris Martin and John Tebbit.

Quality assurance has been undertaken by a consortium led by Etude, including Levitt Bernstein, Julie Godefroy Sustainability, and UCL.

Document reference: HEM-VAL-01

Document version: v1.0

Issue date: 13/12/23

© Crown copyright 2023

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3 or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk.

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

Any enquiries regarding this publication should be sent to us at: homeenergymodel@beis.gov.uk

Contents

Executive Summary

The Home Energy Model (HEM) will replace the existing Standard Assessment Procedure (SAP) for the rating of dwellings' energy use. The new model will be used to demonstrate compliance with the Future Homes Standard, and to produce Energy Performance Certificates. The model will therefore underpin a large number of government policies, making it of critical importance to the delivery of our housing quality and Net Zero objectives.

To ensure accurate simulation of homes energy use, the HEM has been validated using a range of different techniques. This paper summarises the findings and recommendations of the inter-model validation workstream. The inter-model validation compared the HEM with two established modelling software packages:

- The Passivhaus Planning Package (PHPP), a building energy model which is regarded as demonstrably accurate for modelling of high-performance homes in the field.
- Environmental Systems Performance – Research (ESP-r), a building energy model which offers a high time resolution and is known to have good building physics accuracy.

In addition, the HEM was compared to SAP 10.2 to understand how and where it offers improvements.

The inter-model comparison (IMC), comprising a small sample of dwelling archetypes, was undertaken over two phases. Phase 1 used an earlier version of HEM. The simulations across all software packages were fully aligned, meaning input values have been adjusted to represent the same design and environmental conditions in each model. The results were used to validate the building physics algorithms of the HEM, and to inform further development to the consultation version of HEM.

Phase 2 used the consultation version of HEM. Models within this phase applied specific conventions, assumptions, and normalised external conditions, collectively referred to as the model's standardisation. The aim of standardisation is to best represent the actual conditions in the field e.g. how occupants actually heat their homes. It is important to note the standardisation of each model differ. The standardisations were introduced in stages to demonstrate the relative impact of each component.

To assess the validity of the HEM, a range of metrics were applied. These included internal operative air temperature, space heating demand and energy use, each at various time resolutions. Statistical analysis was undertaken on the differences in results between the models and compared with published guidance on acceptable thresholds for such differences (ASHRAE Guideline AG14). As these thresholds were originally used for comparison with real world data, they were not treated as a definitive target for the Home Energy Model, but rather used as a reference.

In Phase 1, the Home Energy Model was shown to agree with its comparators, being within the threshold values for most of the statistical indices applied. The differences between the space heating demand of the HEM and the comparator models varied between archetypes. For example, in some archetypes the HEM predicted a higher demand than the comparator models but for others predicted a lower demand than one or both the comparator models.

Where statistical threshold values were not achieved, the differences in calculations methodology were understood to be the driver and were explainable. These differences were accepted as not being of a high priority (i.e. their impacts were small and not expected to significantly change the HEM outputs) to resolve pre-consultation. Instead, they were considered areas for further investigation and possible future development post-consultation. The Phase 1 findings were used to refine the HEM. These refinements were made prior to Phase 2 of this IMC.

Phase 2 demonstrated that agreement with PHPP reduced after the standardisations were applied. This is expected, as each model has a different representation of actual conditions. PHPP is considered as an accurate representation of monthly conditions in the field for new dwellings. Therefore, further consideration of the accuracy of the HEM:FHS standardisation may be required.

ESP-r applied the same standardisation as the HEM; however, agreement also reduced between the two models. This indicated that the HEM core methodology is sensitive to the conditions (e.g. heating pattern, set point and gains profile) particularly when greater variability of conditions was introduced.

Phase 1 of this IMC study has indicated that the HEM core methodology is suitably aligned with the comparator models; however, under its intended Future Homes Standard application, as was the case in Phase 2, the level of agreement was much reduced. Therefore, it is recommended that further testing of the HEM with the intended FHS normalisation is undertaken, to ensure it is suitably valid. This recommendation should also be extended to any further standardisations that may be considered for other contexts, in particular EPCs.

Specific areas have also been identified for further validation, including a wider range of characteristics (beyond characteristics representative of new build dwellings), ventilation systems and heat pumps.

The consultation version of the HEM has been found to be sufficiently valid as a building physics engine; however, further work will be required to ensure validity is maintained for the intended application to rate both new and existing dwelling energy use. The Home Energy Model will continue to be validated against established models throughout its development to improve its accuracy, including further examination of older building typologies.

1. Introduction

This technical report forms part of the documentation for the government's new Home Energy Model and serves as supplementary information for the Home Energy Model consultation and Home Energy Model: Future Homes Standard assessment consultation.

Validation is the process of checking how well a model meets its goals and user needs. Our intent is for the Home Energy Model (HEM) to simulate building energy use as realistically as possible. Therefore, we have undertaken a series of validation exercises to determine whether this is the case, assessing the model's performance against:

1. Comparable modelling tools,
2. Monitoring data from real dwellings, and
3. Laboratory test data.

These exercises represent a significant increase in the rigour of the development process as compared to SAP. At the time of writing, the HEM validation process is ongoing; therefore, the results presented in this report demonstrate the current status of the validation. We hope that by demonstrating a robust and transparent validation process, we can give users confidence in the new Home Energy Model and demonstrate improvement over SAP 10.2.

This report summarises the work undertaken to date on the first of these three exercises. We have modelled a set of archetype homes using both the Home Energy Model and well-validated existing models and then compared the results. Validation against other models is very valuable as it enables comparison on a detailed component-by-component basis, rather than comparing only "headline" outputs such as overall heating demand. This helps spot errors and allows us to judge the significance and relative strengths of different approaches.

When identifying and explaining points of difference, none of the models is necessarily assumed to be giving the "correct answer"; however, the comparator models chosen are well-validated and established tools.

In this validation exercise, we compared the Home Energy Model[1] against:

- The Passivhaus Planning Package (PHPP)[2], a building energy model which is regarded as demonstrably accurate for modelling of high-performance homes in the field.
- Environmental Systems Performance – Research (ESP-r)[3], a building energy model which offers a high time resolution and is known to have good building physics accuracy.

[1] HEM version 0.14 (Phase 1) and version 0.23 (Phase 2)
[2] PHPP version 10.3
[3] ESP-r version 13.3.15

- In addition, we compared the Home Energy Model to SAP 10.2 to understand how and where it offers improvements.

1.1. Comparing Building Energy Models

Model comparison, along with comparison to test cell and test building data, has been used extensively in the past[4] [5] and there are well established processes for testing building energy modelling tools[6].

Despite the programmes that have been undertaken over the years to test and refine modelling tools[7] [8], such tests rarely lead to perfect agreement. Differences in the modelling philosophy (e.g. dynamic, steady state), underpinning building physics algorithms and data structures, and the phenomena captured in each tool mean that outputs are never perfectly aligned and indeed irreconcilable discrepancies still exist even at the end of an extensive comparison process.

Different models have been designed for different purposes and have different strengths and weaknesses. Even when representing the same dwelling, models accounting for different effects and processes may reach different conclusions. Different models may perform better or worse than one another (in the sense of closely matching real measurements) under different circumstances – there is no single axis of quality and different approaches can bring trade-offs. It is therefore important to compare multiple types of approach across a range of circumstances, to build a robust picture.

Key properties of the models used in this exercise are summarised in Table 1 below.

[4] Judkoff, R., & Neymark, J. (1995). International Energy Agency building energy simulation test (BESTEST) and diagnostic method (No. NREL/TP-472-6231). National Renewable Energy Lab.(NREL), Golden, CO (United States).
[5] Neymark, J., & Judkoff, R. (2002). International Energy Agency Building Energy Simulation Test and Diagnostic Method for Heating, Ventilating, and Air-Conditioning Equipment Models (HVAC BESTEST); Volume 1: Cases E100-E200 (No. NREL/TP-550-30152). National Renewable Energy Lab., Golden, CO.(US).
[6] ANSI/ASHRAE. 2011. ANSI/ASHRAE Standard 140-2011, Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs. Atlanta, GA: ASHRAE.
[7] Judkoff, Ron, and Joel Neymark. What Did They Do in IEA 34/43? Or How to Diagnose and Repair Bugs in 500,000 Lines of Code. No. NREL/CP-550-44978. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2009.
[8] Erkoreka, A., Gorse, C., Fletcher, M., & Martin, K. (2016). EBC Annex 58 Reliable Building Energy Performance Characterisation based on full scale dynamic measurements.

	ESP-r	**PHPP**	**SAP 10.2**	**HEM**
Thermal zones	Multiple, unrestricted	Single zone	Multiple – restricted to living and non-living zones	Multiple, unrestricted – however the FHS assessment will be restricted to a living and non-living zone only.
Simulation type	Dynamic	Steady state	Steady state	Steady state
Simulation timestep	High resolution (i.e. half-hourly or less).	Monthly	Monthly	High resolution (down to half-hourly).
Simulation areas	Building physics, system performance of PV array (but not heating systems etc).	Building physics and system performance of various services.	Building physics and system performance of various services.	Building physics and system performance of various services.
Validation	Demonstrated to have good accuracy vs field test data. Comparable to other detailed simulations tools when subject to the Building Energy Simulation Test (BESTEST) and other empirical trials.	Has demonstrated accurate simulation space heating demand and annual energy use of high-performance dwellings in field trials.	The current calculation method for Part L compliance and EPC ratings. It is not regarded to provide accurate simulation of dwellings in the field. It is included here not for comparison, rather as a reference model.	N/A – new model
Role	Comparator model for dynamic calculations of zonal operative temperature, solar energy, and elemental energy flux.	Comparator model for monthly space heating demand, elemental energy flux and annual energy use and demand.	Reference model to demonstrate the step change expected for compliance modelling under the FHS.	N/A

Table 1 Summary of the modelling packages applied in this inter-model validation study.

The key metrics analysed for HEM, and the comparator models are as follows:

- Annual space heating demand, which is defined as the heat emitter output required to maintain each model zone's set point;
- Internal operative temperature, which is taken here as the average of the dry bulb temperature of internal air and the mean radiant temperature of all internal surfaces within each modelled zone; and
- Dwelling annual energy balance, inclusive of all forms of heat gains and heat losses.

Statistical indices were calculated, based on the differences between the space heating and internal operative temperature metrics predicted by the HEM, PHPP and ESP-r. The statistical testing was undertaken on the monthly outputs of HEM and PHPP and the half-hourly outputs of HEM and ESP-r. In addition to this further metrics were analysed, these included:

- The annual solar gain for each modelled zone for HEM and ESP-r, PHPP was compared at a whole dwelling level;
- The annual losses through infiltration and ventilation
- The annual losses through the main building fabric elements
- The annual losses through thermal bridging

The initial target for the statistical indices (i.e. the maximum differences between the comparison of results between models) was set based on ASHRAE Guideline 14 (AG14). These ASHRAE criteria apply to comparisons between modelling results and real-life data; therefore, performance against the criteria is useful for information, but it has its limitations when applied, as here, in a comparison between models.

Although the aim of the inter-model comparison (IMC) was for HEM to output values within the statistical index tolerance levels for both ESP-r and PHPP, the individual statistical indices achieved between HEM and the other modelling packages were considered as a guide, any observed deviations could then be used to check the HEM core calculation algorithms.

1.2. Archetype Modelling

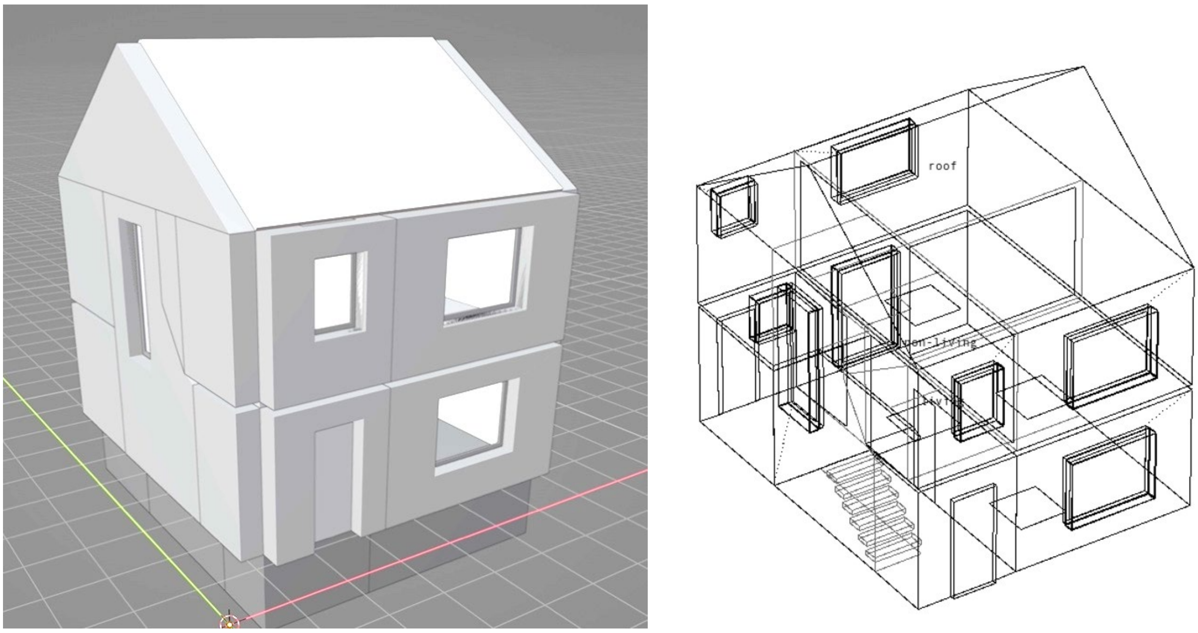
For the IMC, a total of five archetypes were modelled, which were agreed between the modelling teams and DESNZ at the workstream planning stage. These were:

- A detached house (DE),
- A semi-detached house (SD),
- A terrace house (TE),
- A mid-floor deck access flat (DA),
- A mid-floor Victorian-era flat (VF).

These archetypes are fictional but realistic dwellings, chosen to encompass a range of build types. The Victorian flat archetype was chosen to have relatively poor fabric quality; however, the other archetypes were all specified at current new build fabric standards, to ensure relevance to the Future Homes Standard.

The detailed ESP-r model reports, on which the geometry for the other models have been based, are available in Appendix A. Detailed drawings of the archetypes were also produced and are available in Appendix B.

Figure 1 Detached House archetype geometry, extracted from ESP-r modelling report.



It was later agreed to supplement the five archetypes above with a simple ‘shoebox’ model. This additional case further supported the analysis of differences between the core building physics algorithms. This model comprised an archetype with four walls (tested at various orientations in comparison to its environment), a roof and floor each with the same internal surface area. A single window was also included. The purpose of the shoebox was to focus on building physics alone, and so no building service systems were included.

The shoebox analysis included a detailed assessment of solar energy transmitted through transparent elements per orientation, the impact of using different heat transfer coefficients, changes to measurement conventions for diffuse shading calculations, and changes to thermal mass conventions. Learning from this analysis was fed into the main IMC workstream.

In each archetype, Zone 1 is defined as the living space (i.e. the living room) and any contiguous areas of the dwelling connected to this space (i.e. those spaces which were not separated by doors from the living space).

Zone 2 is defined as all other heated spaces within the dwelling.

PHPP is a single zone model comprising both living and non-living zones, and so calculates a single heat balance for the entire dwelling. SAP10.2 also calculates a single heat balance based on the whole dwelling average temperature, as calculated from the two separate thermally controlled zones.

1.3. Plan of work

Since the intention was to isolate the underlying workings from the models' differing standardised assumptions and input conventions, all models underwent an initial input alignment process to ensure like-for-like comparison and test the underlying workings - i.e. the HEM core engine. Broadly speaking, this is Phase 1 of the validation process. This alignment was then progressively relaxed as the testing went on, to examine the effects of key standardisations (like the choice of heating setpoint temperature). Broadly speaking, this is Phase 2 of the validation process. Note therefore that the model results being compared in this report are not what each model (including the HEM) would produce if used normally in the field (except in the final "2D" runs – see below).

In the early stages of the exercise, ESP-r was the primary comparator model. Notably, only ESP-r and the Home Energy Model can simulate an unheated, "free-floating" dwelling, as both SAP10.2 and PHPP automatically assume that dwelling heat demand is both defined and met. As standard assumptions are re-introduced, PHPP takes a more prominent role.

Phase 1: Validating the core engine

The HEMv0.14 was used for this phase of analysis. Models were run with the same or equivalent inputs and assumptions for all packages. Three sub-phases were undertaken, each with a different standardisation. Phase 1A compared an unheated dwelling, 1B a continuously heated dwelling and 1C an intermittently heated dwelling. A comparison of monthly space heating demands, monthly average internal air temperatures and half-hourly operative temperatures was undertaken. The purpose of the comparison was to validate that the building physics algorithms within the HEM produced reasonable estimates of internal temperature and space heating demand. The three different standardised conditions applied aimed to highlight different sensitivities between the models.

Phase 2: Testing the impact of conventions and set assumptions, including the FHS assessment wrapper

The HEMv0.23 was used for this phase of analysis. This is the consultation version of the HEM and incorporates fixes and updates to bugs and core algorithms based on the recommendations made in Phase 1. Models were run with the same dwelling designs as Phase 1 but interpreted in line with each modelling packages' conventions. Multiple sub-phases were undertaken, including a variation testing phase (2A variations). These variations were used to demonstrate how altering single input values impact the internal temperature and space heating demand, providing further validation of the building physics algorithms.

Comparisons of the estimated energy use were undertaken to validate whether the HEM accurately simulated domestic services. The purpose of the sub-phases 2B,2C and 2D was to demonstrate how the standardisation of assumptions influenced the comparison of modelling packages, following the alignment level achieved at Phase 1 when all inputs and conventions are aligned. To provide clarity on the impact of each convention and standardised input, the phases were further subdivided, allowing each component of the standardisation to be introduced in turn. By Phase 2D full standardisation was achieved. Therefore, this final phase can be considered a comparison of the intended application of each model I.e. a certifiable PHPP model, Part L SAP 10.2 model and a HEM:FHS model.

Further detail on the modelling undertaken within Phases 1 and 2 of this IMC is outlined in sections 2 & 3 below.

2. Phase 1

In the first phase of the exercise, we concentrate on the core building physics algorithms of each model, isolated from choices of input conventions and different standardised assumptions which can cause very similar models to give different outputs. This requires a careful override of some initial stages of the model calculations, such as replacing the various sources of incidental gains with a single figure common to all models. Once the model inputs have been aligned to the greatest extent possible, any differences observed can be attributed to more fundamental methodological issues.

Three runs were carried out on the five dwelling archetypes in Phase 1:

Phase 1A: A simulated year of "free-floating" temperatures, with no heating system activity and with fixed constant internal gains. This run was carried out in ESP-r and the Home Energy Model only, as the other models cannot easily represent a free-floating dwelling. Since there is no energy consumption, the focus of the comparison is entirely on the internal temperature and core gains-losses balance in the two models.

Phase 1B: A simulated year of continuous heating to a set point of 21oC and fixed constant internal gains. This included the 4 models. The comparison of results was expanded to include the monthly space heating demand, internal temperature, and core gains-losses balance for all four models. A full description of the comparison metrics employed in Phase 1 is provide in Table 2, below.

Phase 1C: A simulated year of intermittent heating to a set point of 21oC and fixed constant internal gains. The comparisons were undertaken as per Phase 1B.

Full details of the parameterisation and input alignment for each run can be found in Appendix C.

In addition to these three runs, further simulations were carried out on the **Shoebox** archetype to facilitate exploration of the models' differing treatments of solar gains.

In all of Phase 1, the following indices and associated thresholds (from ASHRAE Guideline 14) were applied to the space heating demand, average monthly internal air temperature and half hourly internal operative temperatures as a measure of closeness between outputs:

For Space heating, average monthly internal air temperature and half hourly internal operative temperature -

- The coefficient of variation of the root mean square error (Cv(RMSE)) – tolerance target level of 30% for half-hourly data or 15% for monthly data.
- The normalised mean bias error (NMBE) – tolerance target level of +/- 10% for half-hourly data or +/- 5% for monthly data.

For half hourly internal operative temperature only -

- The mean absolute error (MAE) – tolerance target of +/- 1oC for hourly data.
- The root mean square error (RMSE) - tolerance target of +/- 1.5oC for hourly data.

The key comparisons considered were between the monthly outputs of HEM and PHPP and the half-hourly outputs of HEM and ESP-r.

Throughout this report the statistical indices have been presented in a tabular format with colour coding applied. Table cells coloured green indicate that the AG14 thresholds stated above have been satisfied; red colouration indicates an exceeded threshold.

Again, no modelling package has been considered a representation of the true real-world data. As such, the individual statistical indices achieved between HEM and the other modelling packages were considered as a guide only; any observed deviations could then be used to check the HEM core calculation algorithms.

Table 2 – Key output metrics for Phase 1

	ESP-r	**PHPP & SAP10.2**	**HEM**
Space heating	Heat emitter output per zone and for whole dwelling	Heat emitter output for whole dwelling	Heat emitter output per zone and for whole dwelling
Operative temperature of air point	Typical weekly air point temperature profile (summer/winter) per zone	N/A – temperature variations not calculated	Typical weekly air point temperature profile (summer/winter) per zone
Solar gain	Total solar energy entering per zone and for whole dwelling	Total solar energy entering dwelling	Total solar energy entering per zone and for whole dwelling
Internal gain	Total energy from internal sources per zone and for whole dwelling	Total energy from internal sources	Total energy from internal sources per zone and for whole dwelling
Fabric loss/gain	Convective energy transfer from air point	Total energy transfer from external boundary	Convective energy transfer from air point and from external boundary (both calculated)
Ventilation and infiltration loss/gain	Total energy from combined active and passive air changes	Total energy from combined active and passive air changes	Total energy from combined active and passive air changes
Thermal bridging loss/gain	Total energy from thermal bridging per zone and for whole dwelling	Total energy from thermal bridging	Total energy from thermal bridging per zone and for whole dwelling

ESP-r in Phase 1

ESP-r modelling was undertaken first as the basis of archetype construction, and the other models were aligned with the ESP-r input values and assumptions. When comparing ESP-r with the other modelling packages it is important to note the following:

- Whereas all the other models include the void as part of the roof build up, ESP-r models the roof void dynamically. The result of this is that the roof void temperature fluctuates significantly above the external air temperature due to the solar absorption occurring on the roof. ESP-r models the roof void as a separate unheated zone to understand effects of heat transfer between thermally controlled zones and dynamic temperatures of the roof void, this was expected to have an impact on the DE, SD and TE archetypes.
- ESP-r gives the energy balance for the zone air point and describes the various gains and losses affecting the air point temperature. This includes convective heat exchanges with all opaque and transparent surfaces, internal gains, infiltration and ventilation exchanges, thermal bridge exchanges, heat storage in the air volume (usually negligible) and output of the heater.
- Solar, long-wave radiant and convective heat exchange associated with bounding surfaces are treated as distinct, are time-varying and include thermal storage effects, so it is difficult to derive a composite ‘loss’ figure for a surface, as the heat transfer ‘through’ a construction is complex, e.g. it’s possible to have simultaneous heat gain at the inside and outside surfaces (with heat being stored) and vice versa. The most pragmatic way to derive a ‘heat loss’ metric for constructions that mimics a typical U-value calculation, was to sum all the other gains and losses associated with the zone, integrated over time, with the remainder being the loss attributable to bounding surfaces.
- A key difference between sub-hourly dynamic modelling packages such as ESP-r and low-resolution steady state models, such as the HEM, is the assumption in the latter that the temperature will not rise above the set point. As a result, the steady state monthly models would not account for losses and gains occurring when the ambient internal air temperature is above the set point, although these are accounted for indirectly within the unutilised gains for both PHPP and SAP 10.2. ESP-r accounts for such losses and gains throughout the test period, and also accounts for the longer-term impacts of such heat transfer through the thermal mass of the air and building elements.

PHPP in Phase 1

PHPP conventions state that archetypes should be modelled using the external dimensions. Additionally, the ventilated volume of the archetype is calculated differently depending on the internal space definition. To align with the other modelling packages PHPP has been modelled using the same internal dimensions and internal air volumes as the other modelling packages. This was expected to impact the heat loss occurring through infiltration, ventilation, thermal bridging and building fabric.

PHPP building physics algorithms calculate the space heating demand for a single whole-dwelling zone and assume continuous heating to the target set point. Therefore, the comparison between PHPP and other models was limited to Phase 1B only.

The following were applied for Phase 1B:

- The dwelling, volume, heat loss element areas and thermal bridging lengths were all set using internal dimensions, defined as the measurements between internal surfaces. As the boundary for heat loss was moved from the external surface of all building elements to the internal surface, the thermal mass parameter was corrected to only account for half of the mass as it was assumed that mass was equally distributed either side of the insulation layer.
- PHPP requires input thermal bridge psi values (total linear heat loss in W/m.K) to represent the heat loss on the external surface boundary; however, for consistency the internal surface values applied in the other modelling packages have been applied.
- PHPP requires an input value of the air change hours measured at a test pressure difference of 50Pa with the external environment. This is then corrected to a standard pressure difference through the application of wind protection coefficients. The impacts of these coefficients were removed by manually inputting a total air change rate in m3/h at the standard pressure difference into PHPP.
- PHPP applies a dirt correction factor of 0.95 to all transparent building elements, which reduces the solar gains transmitted. This is not done in the other models and was removed for Phase 1B comparison purposes.
- PHPP U-value calculations apply default surface resistance values depending on the pitch of the element face and its interaction with either internal or external environments. The U-values that were set within the modelling already included the effects of surface resistances. As such, PHPP default values for internal and external surface resistances (Rsi and Rse) were set to zero.

SAP 10.2 in Phase 1

SAP 10.2 contains several default values and correction factors which are applied within the Part L 2021 normalisation. These defaults and correction factors were removed. The following actions were taken:

- SAP requires an input value of air permeability measured in m3/h.m2 measured at a test pressure difference of 50Pa with the external environment. This is then corrected to a standard (lower) pressure difference. This was overridden by manually inputting the total air change rate in m3/h at the standard pressure difference.
- SAP wind correction factors which impact monthly infiltration rates were set to zero so that the infiltration input value was maintained throughout the test period.
- For dwellings with whole house extract ventilation, which was assumed for the Phase 1 archetypes, SAP calculates the effective air change rate from the ventilation system as

a function of the infiltration occurring within each monthly calculation. As such, if the infiltration is sufficient to provide the required air change rate to meet the fresh air requirements set by Part F of the building regulations, then corrections occur for the assumed ventilation system and no ventilation is provided. This calculation was removed from the SAP algorithm to ensure that the input value for ventilation was maintained throughout the test period.

- SAP applies U-value corrections on all windows to account for the impact of drawn curtains. The default assumption that curtains or blinds were drawn some of the time resulted in the opening area having a lower rate of heat transfer. To ensure that the input U-value for the windows was used by SAP, the effect of the reduction factor was removed by manually increasing the input U-value given to SAP by an equivalent to the reduction factor.
- In SAP, a general shading factor is applied to all windows, regardless of size, location, or orientation. As such the impacts of shading structures such as reveals and balcony overhangs are not accurately accounted for. SAP conventions state that the default of “average/unknown” shading should be applied for all compliance calculations, which introduces a shading factor of 0.77 from SAP 10.2 Table 6d. To overcome this correction the shading factor was set to 1 (no shading). Despite this, because the other models can account for near and distant shading in detail, SAP was expected to differ in its solar gains results.

HEM in Phase 1

The FHS assessment wrapper calculations were removed from the HEM. The air point energy balances of HEM and ESP-r are directly comparable; however, it was not appropriate to compare these energy balances with those produced by PHPP or SAP 10.2. To address this, the BRE developed a further HEM energy balance at the external dwelling boundary, which is comparable to PHPP and SAP 10.2.

To facilitate the Phase 1 IMC, the following actions were taken:

- The convective and radiative fractions of the direct electric heating system applied in the 1B and 1C modelling runs were revised to match the ESP-r values.
- Building elements were assigned the absorptivity values from ESP-r report; however, the HEM core engine assumes an emissivity of 0.9 for all materials within each element, this differed to ESP-r where the emissivity values of each element layer were explicitly set within the input assumptions.
- The HEM ground floor temperature requires the internal air temperature. As the HEM cannot dynamically calculate the ground floor temperature, a monthly average internal temperature profile was assumed, this was based on the SAP 10.2 predicted internal monthly temperatures for the same dwelling design. In practice the ground temperature would fluctuate in response to heat transfer through the ground floor construction and in response to environmental conditions, as is the case in ESP-r.

- The HEM calculates the wind driven infiltration using the hourly wind speed values. As such, to align with the other models, a constant infiltration air change rate was assumed, which was unaffected by the impacts of the wind.

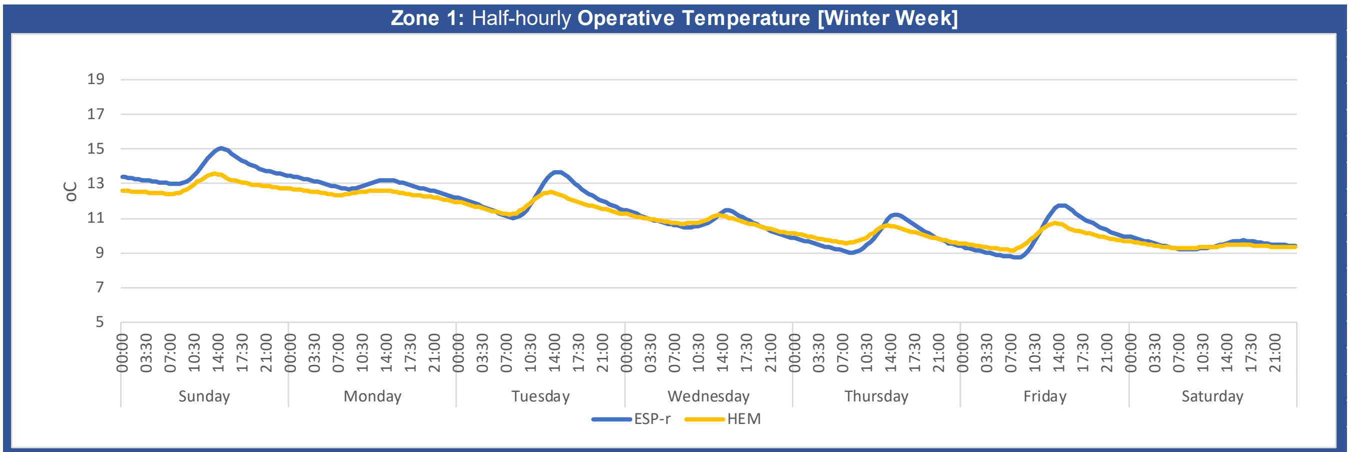
Details on how the modelling assumptions and conventions were broken down into stages and how this was applied to the comparator models is provided in the tables below. A full summary of the input assumptions is available in Appendix C.

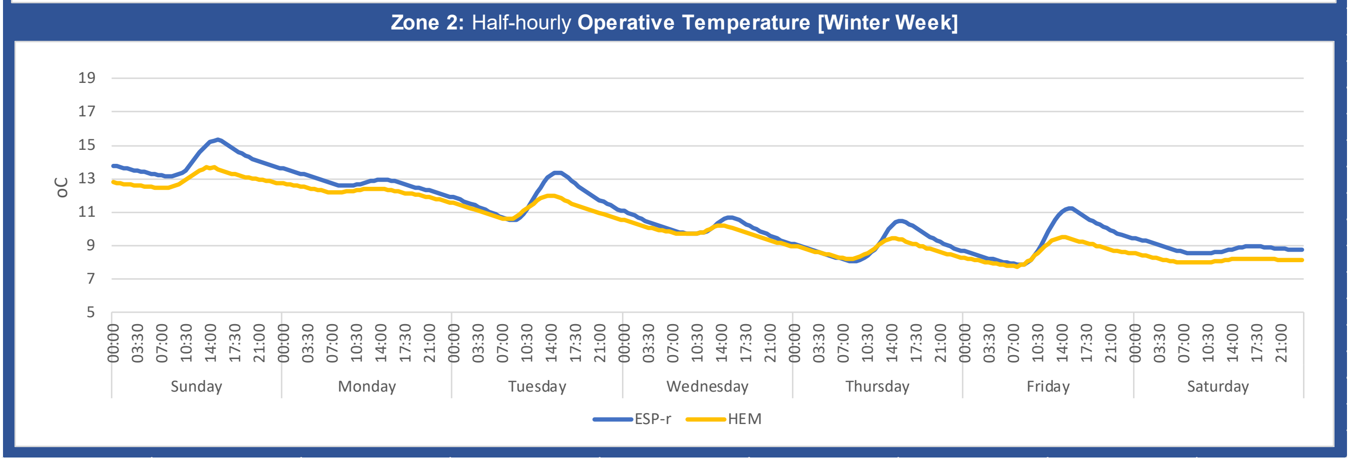
2.1. Phase 1 Results

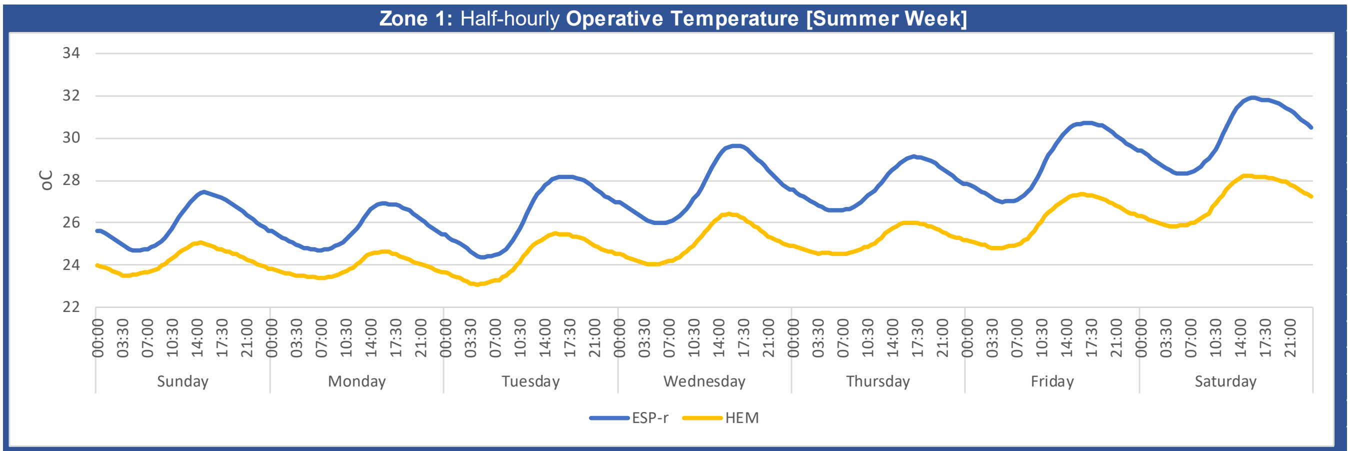
Here we present a selection of outputs for each of the Phase 1 runs. For each sub-phase a full set of results is presented for the DE archetype at the beginning of each respective section. These include the half-hourly internal operative temperature comparisons between ESP-r and the HEM, annual and monthly space heating demands of all models, and the statistical indices of difference between models. Excerpts from the other archetype results are included only where there are significant deviations from DE results. Full results from all runs can be found in Appendix D.

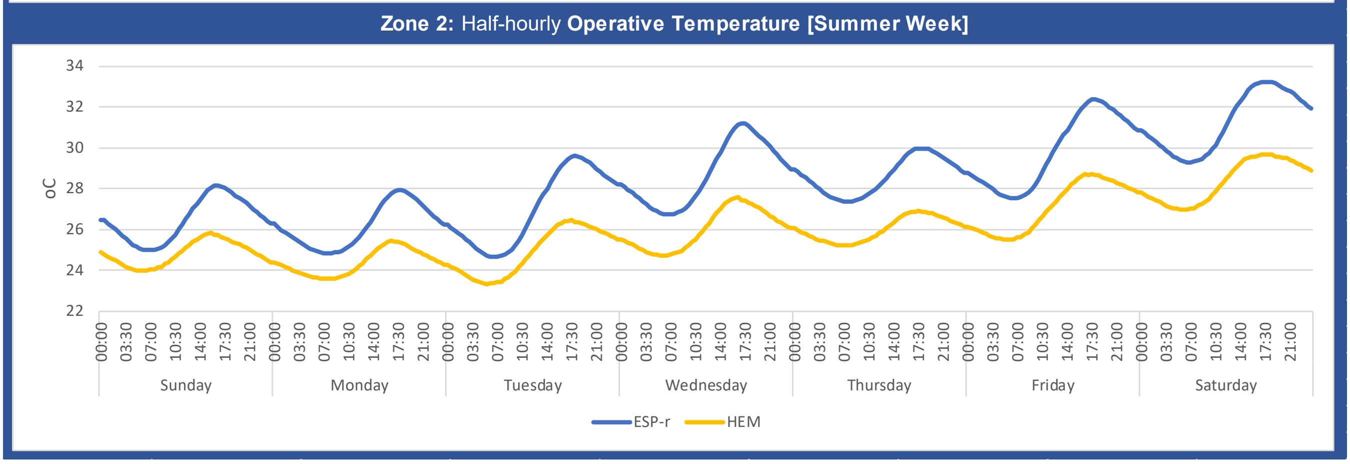
2.1.1. Phase 1A – free-floating temperature

Detached House









	ESP-r Vs HEM			
	CV(RMSE)	NMBE	RMSE (oC)	MAE (oC)
Monthly Space Heating (kWh/yr)	0%	0%	0%	0%
Monthly Average Internal Air (dry bulb) Temperature [average for whole dwelling] (°C)	7%	6%	0	0
Zone 1: Half-hourly Operative Temperature [Winter Week]	5%	2%	0.9	0.9
Zone 2: Half-hourly Operative Temperature [Winter Week]	7%	6%	1.1	1.0
Zone 1: Half-hourly Operative Temperature [Summer Week]	9%	9%	1.8	1.8
Zone 2: Half-hourly Operative Temperature [Summer Week]	9%	9%	1.7	1.7

**Note that grey cells indicate that no comparison was undertaken.*

The internal temperature profiles in winter between ESP-r and HEM were well aligned. The statistical indices for the half-hourly operative temperatures in both Zone 1 and 2 for the winter and summer weeks were within the threshold values outlined by ASHRAE AG14 for CV(RMSE) and NMBE. Threshold values were exceeded for the MAE of the typical winter week. The rate and magnitude of the daily temperature fluctuations were observed to be different between the HEM and ESP-r. This was understood to be a consequence of the different approaches applied for simulating the inertia of building elements and the rate of transfer of stored energy with the air point.

The RMSE and MAE for both Zone 1 and Zone 2 in the summer week were exceeded the threshold value, suggesting larger differences between the models when there is greater environmental solar energy and when external temperatures are higher. It was concluded that dynamic nature of the solar transmittance calculations and external surface HTCs was a driver for the difference. Differences in the way thermal mass is accounted for in the models was also suspected to be a driver. A future area of development of the HEM could be to understand how the current steady state methods differ to the dynamic calculations within ESP-r.

The statistic indices produced for the SD, TE, DA and VF were equal in magnitude to those observed for the DE above. The TE archetype produced the closest alignment between ESP-r and the HEM. The Zone 2 half-hourly operative temperature for the typical summer week was the only value which exceeded the AG14 threshold. This was assumed to be due to the same reason as noted for the DE archetype above. Close alignment of the TE models was expected due to the minimal solar shading and high proportion of adiabatic surface area.

2.1.2. Phase 1B – continuous heating

Detached House

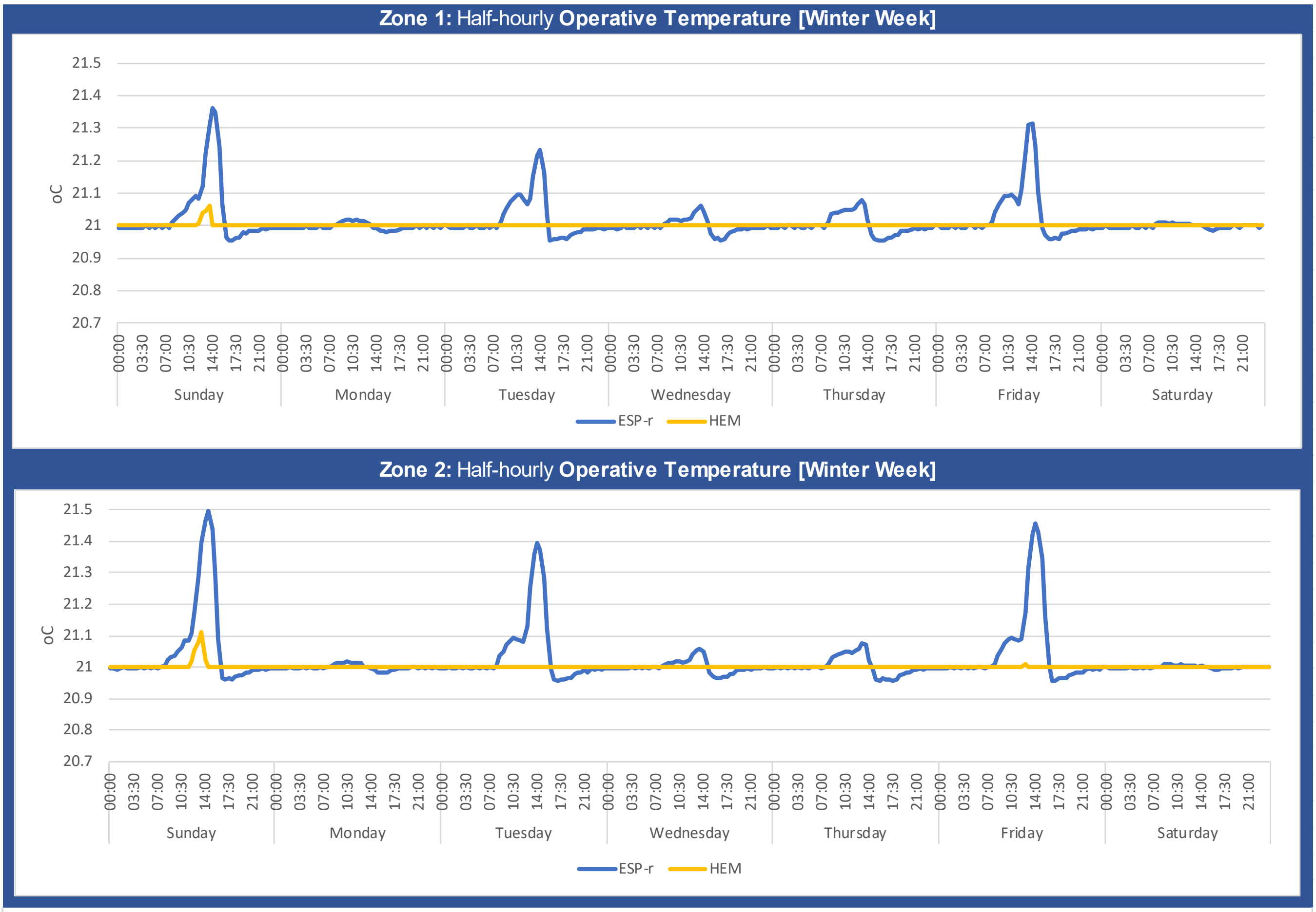


Figure 1B_DE Zone Half-hourly Operative Temperature Profile in Typical Winter Week for ESP-r and HEM

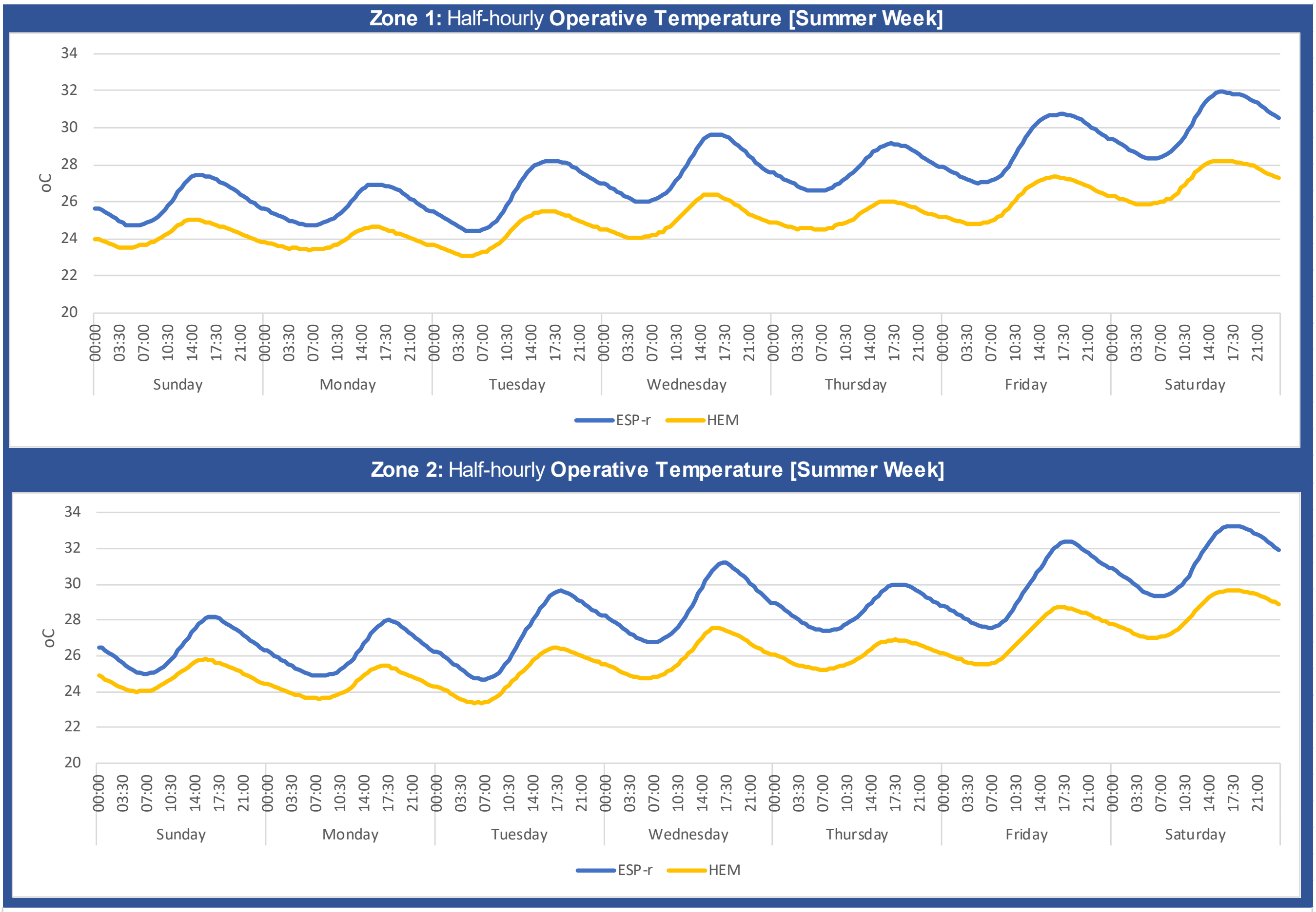


Figure 1B_DE Zone Half-hourly Operative Temperature Profile in Typical Summer Week for ESP-r and HEM

Table 1B_DE Annual Energy Demand

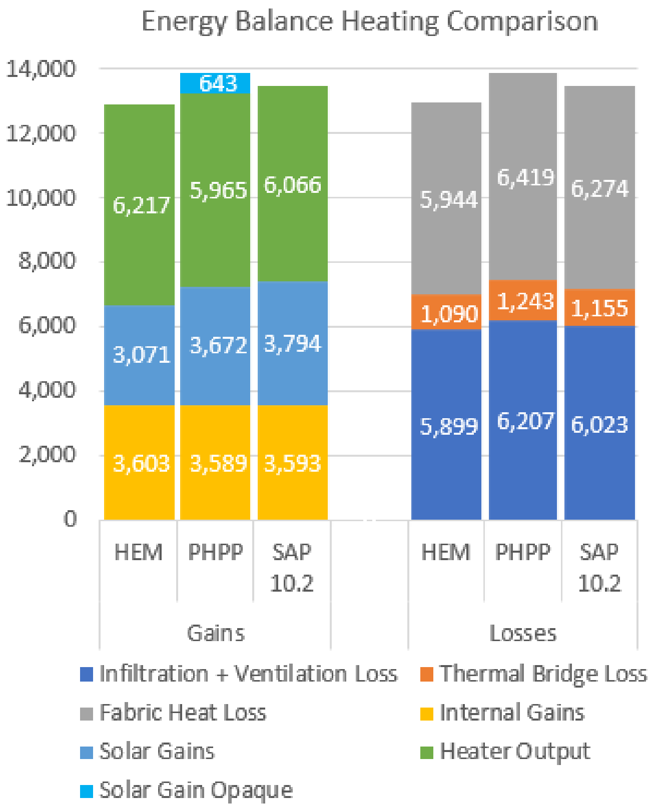
		1B_DE			
		ESP-r	PHPP	SAP10.2	HEM
Annual Energy Demand (kWh/yr)	Space Heating	5,311	6,052	5,967	6,036
	Hot Water				
	Cooling				
	Lighting				
	Ventilation				

Table 1B_DE Monthly Space Heating Demand

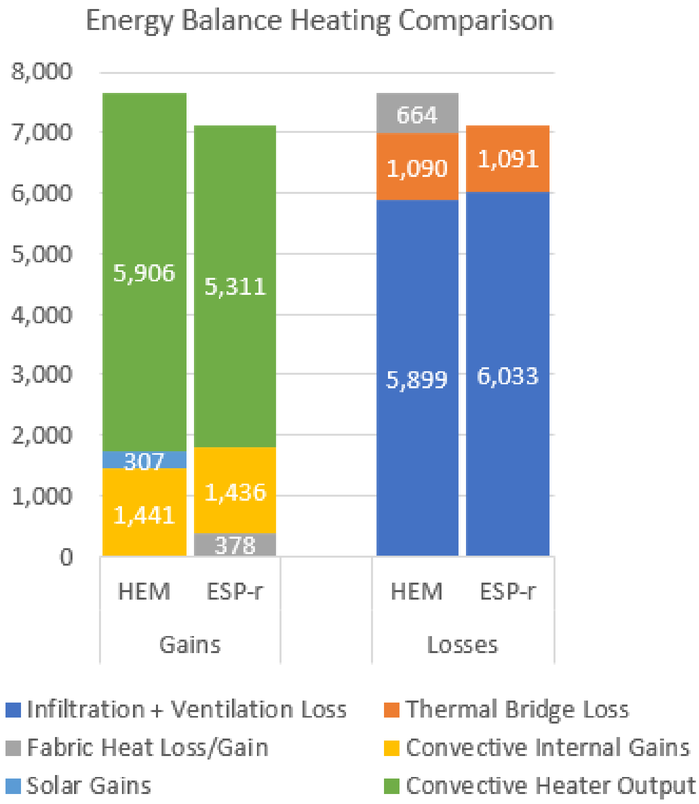
		1B_DE			
		ESP-r	PHPP	SAP10.2	HEM
Monthly Space Heating Demand (kWh)	January	1055.0	1071.0	1110.4	1086.7
	February	776.0	815.0	865.9	834.2
	March	572.0	659.0	731.1	666.5
	April	393.0	487.0	443.2	510.3
	May	163.0	261.0	218.4	272.8
	June	5.0	87.0	67.5	22.8
	July	0.0	34.0	18.7	0.4
	August	0.0	29.0	26.0	0.0
	September	54.0	191.0	119.1	133.4
	October	390.0	469.0	428.6	491.6
	November	818.0	846.0	798.6	885.9
	December	1086.0	1102.0	1139.8	1131.6

Table 1B_DE Operative and Internal Air Temperatures

		1B_DE			
		ESP-r	PHPP	SAP10.2	HEM
Annual Internal Air (dry bulb) Temperature [average for whole dwelling] (°C)		22.3	21.0	21.0	21.9
Half-Hourly Operative Temperature	Zone 1: Winter Week] (oC)	21.2	21.0	21.0	21.2
	Zone 2: [Winter Week] (oC)	21.0	21.0	21.0	21.0
	Zone 1: [Summer Week] (oC)	25.1	21.0	21.0	23.7
	Zone 2: [Summer Week] (oC)	25.6	21.0	21.0	24.0



Phase 1B - DE - Energy Balance Comparison Graphs at the External Boundary



Phase 1B - DE - Air Node Energy Balance Comparison Graphs

Table 1B_DE Statistical indicies for CV(RMSE), NMBE, RMSE, MAE for all inter-model comparisons

		ESP-r Vs HEM				PHPP Vs HEM		SAP10.2 Vs HEM		ESP-r Vs PHPP		SAP 10.2 vs PHPP		SAP10.2 Vs ESP-r	
		CV(RMSE)	NMBE	RMSE (oC)	MAE (oC)	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE
Monthly Space Heating (kWh/yr)		15%	-13%	0	0	7%	0%	10%	-1%	15%	-13%	9%	-1%	15%	12%
Monthly Average Internal Air (dry bulb) Temperature [average for whole dwelling] (°C)		3%	2%	0	0	7%	-4%	7%	-4%	10%	6%	0%	0%	10%	-6%
Half-hourly Operative Temperature	Zone 1: Half-hourly Operative Temperature [Winter Week]	0%	0%	0.1	0.0	0	0	0	0	0	0	0	0	0	0
	Zone 2: Half-hourly Operative Temperature [Winter Week]	0%	0%	0.1	0.1	0	0	0	0	0	0	0	0	0	0
	Zone 1: Half-hourly Operative Temperature [Summer Week]	9%	9%	1.9	1.8	0	0	0	0	0	0	0	0	0	0
	Zone 2: Half-hourly Operative Temperature [Summer Week]	9%	9%	1.8	1.7	0	0	0	0	0	0	0	0	0	0

The half-hourly operative temperature profiles for the typical winter week displayed greater fluctuation in the ESP-r results, this can be explained by the set point control applied within ESP-r. It is important to note that in Figure 1B_DE for the HH Operative Temperature for Typical Winter Week, the y-axis is set at a high resolution and the differences observed are proportionally small. The method by which the mean radiant temperature is calculated at each timestep meant that the exact operative temperature set point is not always achieved. No notable difference with the Phase 1A results was observed for the operative temperatures in the typical summer week.

The HEM, PHPP and SAP 10.2 predicted similar annual space heating demands at Phase 1B. The monthly space heating demand CV(RMSE) and NMBE values are within the AG14 thresholds for all three model comparisons. ESP-r was a relative outlier at Phase 1B. The dynamic HTCs assumed for the external boundary were expected to reduce the rate of fabric heat loss relative to the other models. The HEM, PHPP and SAP 10.2 all derive the HTC from the European Committee for Standardisation (CEN) recommended approach in ISO 13370:2017. In addition to HTCs, the way that ESP-r calculates the solar transmittance into the internal environment differed to the other models. This was expected to drive differences between ESP-r and the other models. Further details on this difference are discussed in the shoebox analysis section below.

To understand the impact of these methodological differences between ESP-r and the steady state models, a light touch modelling scenario was considered, although this was not a formal part of the validation work plan. This suggested that when CEN HTCs were applied in ESP-r and the impacts of glazing were removed, the alignment of ESP-r and the other models improved. Therefore, a formalised validation exercise with ESP-r could be undertaken as part of ongoing HEM validation to understand the value of applying dynamic HTCs and explicitly calculating the transparent element characteristics.

The energy balance graphs indicated alignment in the thermal flux of the HEM with the comparator models. It should be noted a direct comparison of ESP-r with PHPP and SAP 10.2 was not possible due to the ESP-r energy balance being produced exclusively for the internal air point. Two key differences were observed between the models. The combined heat loss from ventilation and infiltration differed most significantly between the models. This was expected due to the different average internal air temperatures reported in Table 1B_DE Operative and Internal Air Temperatures. The fabric heat loss/gain of ESP-r and the HEM differed significantly. This was again concluded to be a result of the dynamic HTCs and explicit calculation of window characteristics.

The statistical indices calculated for the SD, TE and VF archetypes were similar in magnitude to those achieved for the DE archetype. A notable difference was that the HEM vs PHPP CV(RMSE) and NMBE values for both monthly space heating and average internal air temperature exceeded the AG14 threshold for the TE archetype. This indicated that agreement between these models may be sensitive to low heat loss rates. The TE archetype was found to have the lowest heat loss parameter (W/m2.K) of all the archetypes. .

Deck Access Flat

Table 1B_DA Statistical indicies for CV(RMSE), NMBE, RMSE, MAE for all inter-model comparisons

		ESP-r Vs HEM				PHPP Vs HEM		SAP10.2 Vs HEM		ESP-r Vs PHPP		SAP 10.2 vs PHPP		SAP10.2 Vs ESP-r	
		CV(RMSE)	NMBE	RMSE (oC)	MAE (oC)	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE
Monthly Space Heating (kWh/yr)		7%	4%	0	0	23%	13%	29%	-19%	20%	-10%	38%	-32%	32%	-22%
Monthly Average Internal Air (dry bulb) Temperature [average for whole dwelling] (°C)		1%	-1%	0	0	6%	-4%	6%	-4%	6%	3%	0%	0%	6%	-3%
Half-hourly Operative Temperature	Zone 1: Half-hourly Operative Temperature [Winter Week]	0%	0%	0.0	0.0	0	0	0	0	0	0	0	0	0	0
	Zone 2: Half-hourly Operative Temperature [Winter Week]	1%	0%	0.2	-0.1	0	0	0	0	0	0	0	0	0	0
	Zone 1: Half-hourly Operative Temperature [Summer Week]	4%	-4%	1.1	-1.1	0	0	0	0	0	0	0	0	0	0
	Zone 2: Half-hourly Operative Temperature [Summer Week]	4%	4%	0.9	0.9	0	0	0	0	0	0	0	0	0	0

For the DA archetype the magnitude of the CV(RMSE) indices for all comparisons, excluding ESP-r vs HEM, exceeded the AG14 thresholds. This indicated a general divergence between all models when significant and complex shading structure were included in the model. The same indices for the ESP-r vs HEM comparison were within the threshold, suggesting that the HEM replicated the complex shading calculations of ESP-r with reasonable accuracy. The high resolution and dynamic nature of the ESP-r calculations are expected accurately simulate the impacts of external solar on the internal air temperature and space heating demand. Differing monthly calculations of solar energy and the impacts of hear shading in both PHPP and SAP 10.2 was a primary driver for the higher index values observed. The lower level of alignment indicated by the CV(RMSE) indices for the HEM and the monthly steady state models should not be an area of concern when alignment with ESP-r have been demonstrated.

2.1.3. Phase 1C – intermittent heating

Detached House

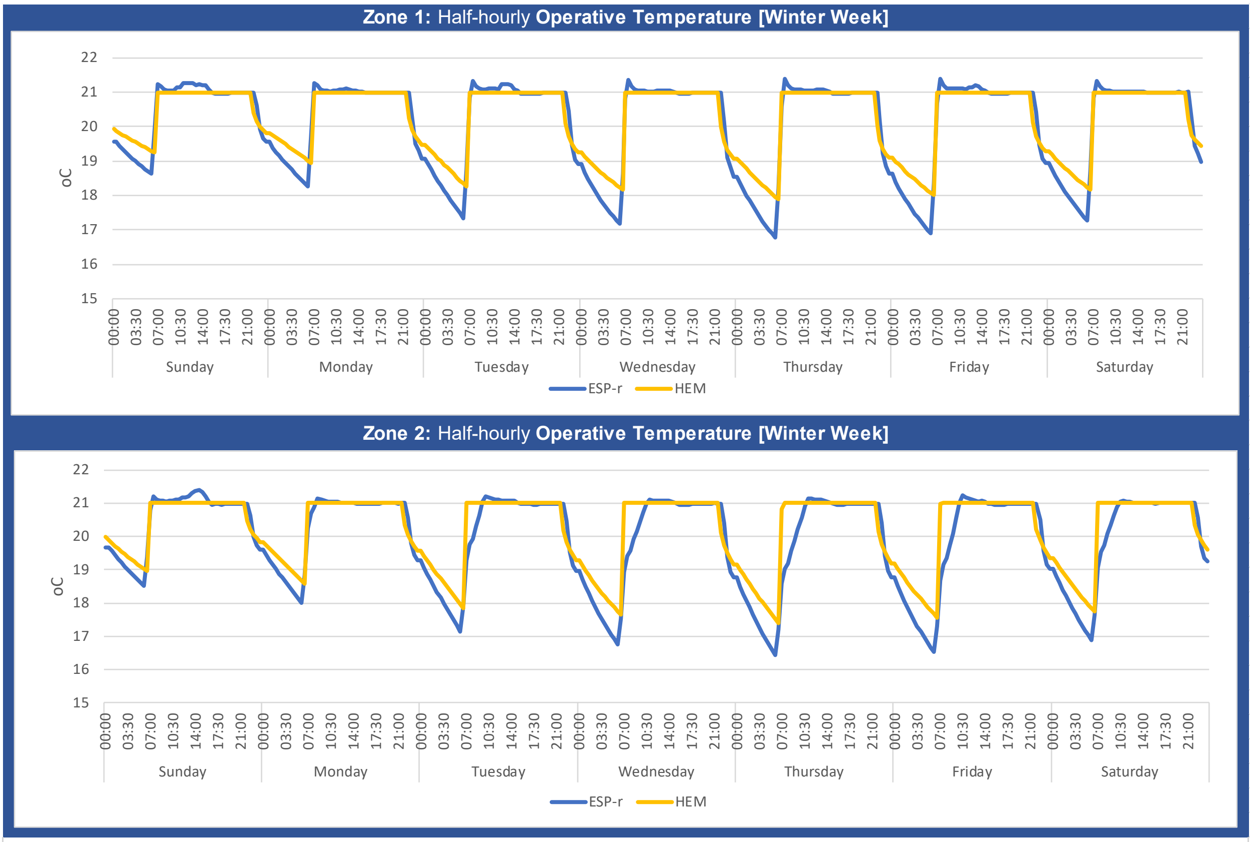


Figure 1C_DE Zone Half-hourly Operative Temperature Profile in Typical Winter Week for ESP-r and HEM

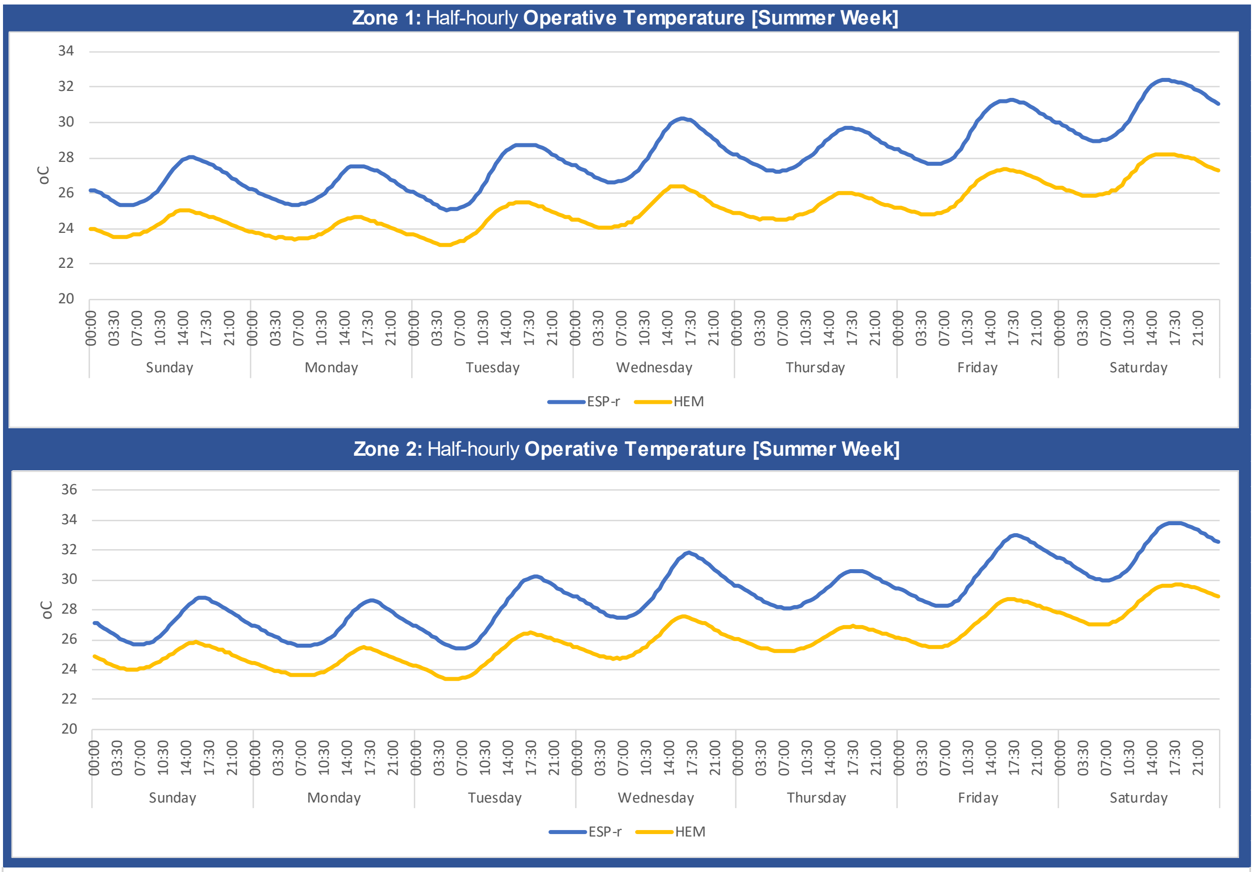


Figure 1C_DE Zone Half-hourly Operative Temperature Profile in Typical Summer Week for ESP-r and HEM

Table 1C_DE Annual Energy Demand

		1C_DE			
		ESP-r	PHPP	SAP10.2	HEM
Annual Energy Demand (kWh/yr)	Space Heating	3,558		5,547	5,648
	Hot Water				
	Cooling				
	Lighting				
	Ventilation				

Table 1C_DE Monthly Space Heating Demand

		1C_DE			
		ESP-r	PHPP	SAP10.2	HEM
Monthly Space Heating Demand (kWh)	January	771.0		1027.5	1079.8
	February	540.0		801.9	780.1
	March	354.0		678.7	613.3
	April	229.0		414.3	462.5
	May	75.0		207.2	240.3
	June	1.0		65.5	13.7
	July	0.0		18.5	0.2
	August	0.0		25.6	0.0
	September	21.0		114.0	106.0
	October	204.0		399.7	449.8
	November	566.0		739.6	832.9
	December	797.0		1054.5	1068.9

Table 1C_DE Operative and Internal Air Temperatures

		1C_DE			
		ESP-r	PHPP	SAP10.2	HEM
Annual Internal Air (dry bulb) Temperature [average for whole dwelling] (°C)		22.1		21.0	21.5
Half-Hourly Operative Temperature	Zone 1: Winter Week] (oC)	19.1			19.6
	Zone 2: [Winter Week] (oC)	19.6			19.9
	Zone 1: [Summer Week] (oC)	25.7			23.7
	Zone 2: [Summer Week] (oC)	26.2			24.0

Table 1C_DE Statistical indicies for CV(RMSE), NMBE, RMSE, MAE for all inter-model comparisons

		ESP-r Vs HEM				SAP10.2 Vs HEM		SAP10.2 Vs ESP-r	
		CV(RMSE)	NMBE	RMSE (oC)	MAE (oC)	CV(RMSE)	NMBE	CV(RMSE)	NMBE
Monthly Space Heating (kWh/yr)		54%	-45%	0	0	10%	-2%	50%	44%
Monthly Average Internal Air (dry bulb) Temperature [average for whole dwelling] (°C)		5%	3%	0	0	7%	-2%	11%	-5%
Half-hourly Operative Temperature	Zone 1: Half-hourly Operative Temperature [Winter Week]	2%	-1%	0.3	-0.1	0	0	0	0
	Zone 2: Half-hourly Operative Temperature [Winter Week]	3%	-1%	0.3	0.0	0	0	0	0
	Zone 1: Half-hourly Operative Temperature [Summer Week]	11%	11%	2.4	2.4	0	0	0	0
	Zone 2: Half-hourly Operative Temperature [Summer Week]	11%	11%	2.4	2.3	0	0	0	0

The comparison with PHPP has been removed from this phase as an intermittent heating pattern cannot be simulated by PHPP. I introduction of intermittent heating did not cause exceedance of the statistical thresholds for half-hourly comparisons in the typical winter week; however, they were exceeded for the monthly space heating demand. Inspection of the typical winter week profiles indicated that the models disagree on the rate of cool off between heating periods. The two most likely causes for this disagreement were the modelling of inertia of the thermal mass and the rate of the heat loss, which may differ due to the approaches to surface HTCs. It was noted that the HEM was less well aligned, particularly for space heating demand, with ESP-r under the intermittent heating scenario. This may have significant bearing on the intended HEM FHS application and is analysed further in Phase 2.

SAP 10.2 and ESP-r also experienced a divergence in monthly space heating due to the introduction of intermittent heating, this may indicate the limitation of a steady state model simulating an intermittent daily heating profile.

2.1.4. Shoebox Analysis

This analysis was targeted towards demonstrating how the differences in solar gain methodology between the models were impacting operative temperatures and space heating demand.

One area in which it was challenging to achieve full alignment between the four models was the solar transmittance values. HEM, SAP 10.2 and PHPP all apply a constant g-value to windows, applicable at normal incidence, and each applies a correction factor to represent the average annual reflection, consistent with ISO 52016.

ESP-r, however, applies varying rates of reflection based on the incident angle at any one timestep, calculated as a function of the direct transmittance value at a normal angle of incidence. Additionally, the ESP-r calculation of solar energy transmittance was broken down into direct shortwave transmittance through each layer of the transparent element (i.e. outer pane of glass, air gap, and the inner pane of glass) as well as solar absorption leading to longwave radiation emitted from the internal surface of the transparent building element. A visual representation of the differences in solar energy transfer through transparent elements is presented in Appendix E.

The monthly total shortwave radiation entering the Shoebox as well as the half hourly profile of shortwave radiation was plotted for a North, East, South and West orientation. The impacts of near shading structures, such as window reveals were also included to understand the impacts. The solar incidence values were taken as the total radiation *after* the impacts of absorption by the glass panes and air gap, incident reflection and near shading. The radiation values take account of direct, diffuse and ground reflected shortwave radiation.

Results

The degree to which all models agreed on the solar energy gain (unshaded) varied significantly with orientation[9]. Detailed results are presented in Appendix F. It was observed that more solar energy was transmitted to the internal environment through Northern and Eastern elevations within the HEM. This was considered to be a contributing factor of the typical summer week half-hourly operative temperature difference presented in Phase 1A and 1B. This effect was more clearly identified in the Shoebox model as the glazing was a high proportion of the external element area and was limited to a single orientation. The same impact is expected in the main archetypes; however, the evidence of this is less clear due to their complexity.

When near shading was introduced, SAP 10.2 had the highest annual solar transmittance. This was demonstrated in Figure SB2 and is attributed to the lack of detailed shading calculations within SAP 10.2. For the high-resolution models, shading was observed to have a more significant impact, particularly for south orientated glazing. Annual shading factors for the south orientation were calculated to be 77%, 76% and 92% for the HEM, ESP-r and PHPP respectively. It is expected that this relationship may change for different weather files as the high-resolution models will be sensitive to the incident angle of any direct solar radiation during the timesteps where direct solar is available.

A seasonal difference in solar energy transmitted was observed between ESP-r and the HEM. This was attributed to the difference in the calculation of incident reflection. For the South orientation, the difference observed to be greater in summer months and lower in winter months.

[9] The analysis of the half-hourly shortwave solar radiation indicated a discrepancy for specific orientations. This was attributed to an issue with the format of the Energy Plus Weather (.epw) applied. This was investigated further but is outside of the scope of this report. A summary of the issue is presented in Appendix G.

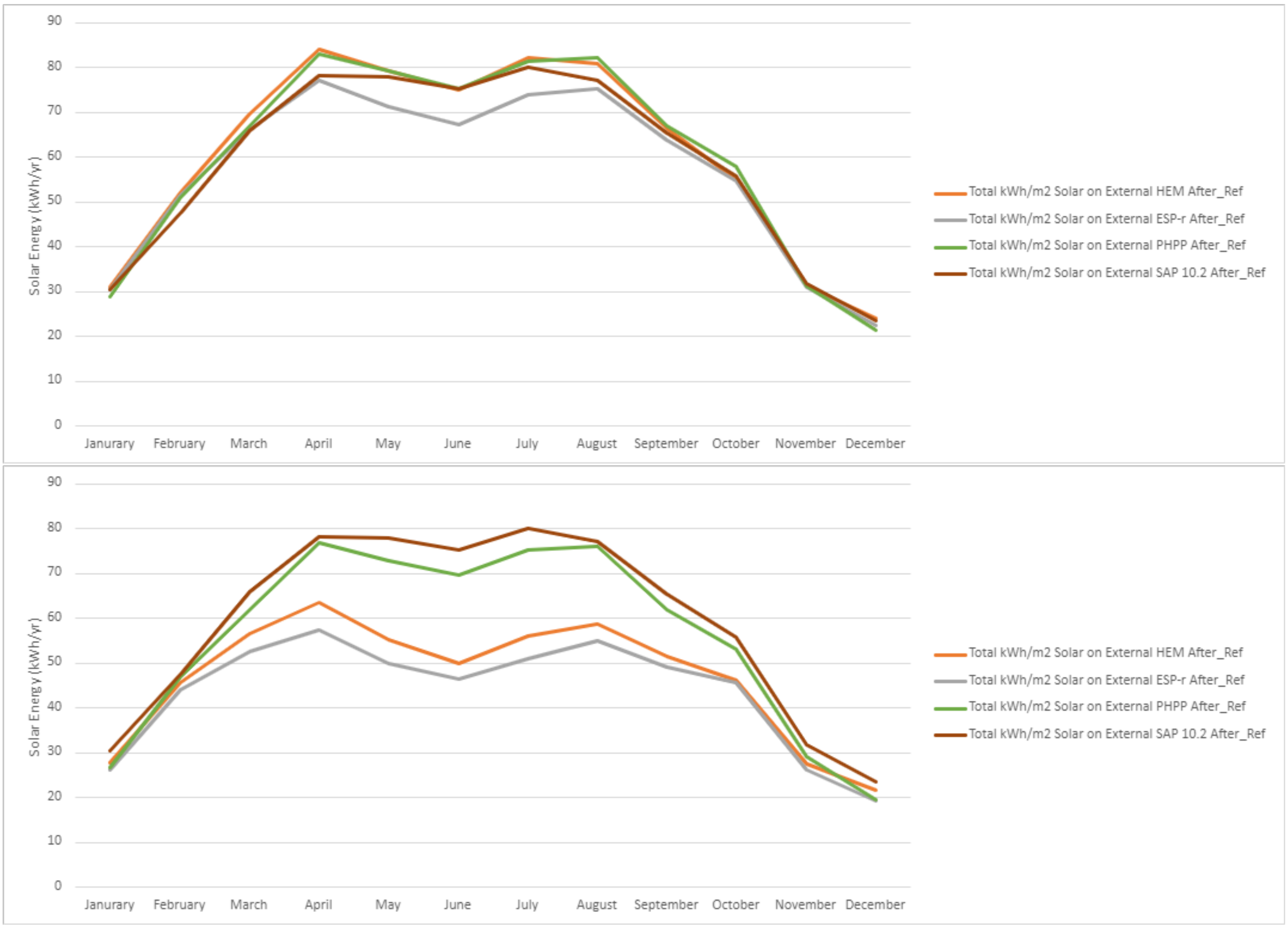


Figure SB1 – Shoebox comparison of total monthly shortwave solar radiation entering through South facing transparent element (Above – no near shading; Below – with near shading)

Figure SB2 – Shoebox comparison of total monthly shortwave solar radiation entering through North facing transparent element (Above – no near shading; Below – with near shading)

2.2. Phase 1 Conclusions

Phase 1 of this IMC applied a fully aligned modelling approach to validate the building physics algorithms of the HEM. The HEM was compared with ESP-r and PHPP, as well as with SAP 10.2. Results have also indicated how well the comparator models aligned with one another.

The HEM has been shown to partly agree with its comparators. It is within the threshold values recommended by ASHRAE AG14 for many of the statistical indices of closeness and difference that were calculated, although again noting that these thresholds are for comparison with in-use data. Where threshold values were not achieved, the differences in calculation methodology were understood to be the driver and were explainable. These differences were accepted as not being of a high enough priority to resolve during this phase of the validation. Instead, they were considered areas for further investigation and possible future development.

The results presented in the above sections have informed the below considerations for future HEM validation exercises and focus areas for model development.

- A difference in the calculation of thermal inertia and flux of building elements between the HEM and ESP-r was noted to impact the rate and magnitude of temperature fluctuation in the half-hourly operative temperature profile comparisons. The HEM method of applying the areal heat capacity (in J/m2.K) of each building element and the mass distribution class as per BS EN ISO 52016-1:2017 is less granular than the dynamic layer by layer calculation applied in ESP-r. Layer specific thermal mass calculation could be considered as a means to model the thermal mass impacts in more detail. This could result in more accurate simulation of thermal mass effects.
- The magnitude of solar reflection on transparent building elements was found in ESP-r to differ depending on the time of day and season. Consideration should be given to calculating the reflection from the incident angle at each timestep or at a greater frequency than the annual average currently applied in HEM. This could also be expanded to explicitly calculate further characteristics of transparent elements, such as direct transmittance, absorption, and pane temperature.
- The dynamically calculated HTCs in ESP-r predicted lower levels of heat loss compared to static HTCs. Further testing of the differences between CEN and dynamically calculated HTCs should be considered and the appropriateness of the CEN HTCs should be explored.
- Future validation work should also consider the inclusion of dwellings with low heat loss rates relative to the heated volume. The Phase 1B results indicated that agreement between models exceeded the threshold values for the archetype with the lowest area specific heat loss rate.

3. Phase 2

This Phase of work aimed to demonstrate how the standardisation of each individual model can impact the alignment achieved at Phase 1: in Phase 2 of the exercise, we move from the “same design, aligned inputs” paradigm to a “same design but some standardisation” paradigm, where the underlying dwelling archetypes are the same as before, but these are interpreted by each model in a way that is progressively more similar to how they would operate when used normally in the field I.e. with their conventions and standardised inputs.

Note ESP-r does not possess standardised conventions in the same way as the other models, and so plays a lesser role as a benchmark model as the phase progresses. ESP-r was aligned with the assumptions being used in the Home Energy Model for each run to demonstrate how a DSM would behave under the same assumptions.

Note also that, as detailed in the Plan of Works section, Phase 1 work informed further development for HEM. Therefore, Phase 2 simulations used the consultation version of HEM, incorporating further development compared to the minimum viable product version used in Phase 1 simulations. In some instances, results between two phases are therefore not exactly comparable.

Compared to Phase 1, by calculating the same output metrics and statistical test indices for differences between the models it was possible to track the scale of difference caused by the various components of standardisation. In addition, energy usage was calculated to assess differences in the representation of systems across the 4 models. These energy use figures were not subject to statistical analysis; however, where large differences were identified this has been highlighted as a potential area for future HEM development and validation.

The number of archetypes was reduced to include only the DE, DA and VF as these were considered a broad enough sample to test the standardisations.

Five sub-phases were carried out in Phase 2, different groupings of archetypes were included depending on the nature of the comparison:

Phase 2A: A simulated year of continuous heating to a set point of 21oC and typical daily internal gain profile.

Phase 2A Variations: As per Phase 2A, and with the same metrics assessed, with the addition of energy use of building services, where these were simulated. A summary of the variations tested is provided in the below table.

Run reference	Description
AP5	Infiltration input based on standardised test value at 50 pascal pressure difference air permeability result of 5m3/h.m2

AP3	Infiltration input based on standardised test value at 50 pascal pressure difference air permeability result of 3m3/h.m2
AP1	Infiltration input based on standardised test value at 50 pascal pressure difference air permeability result of 1m3/h.m2
ORI	Archetype was rotated 90 clockwise
OVH	Overhangs were introduced to all glazing [700mm deep overhangs 100mm above and set to the opening width]
FHS	Element thermal performance aligned to the FHS consultation specification to represent a high-performance dwelling
EXI	Element thermal performance aligned to the VF archetype specification to represent an existing dwelling
MVH	Whole house balanced mechanical ventilation introduced.
PVA	PV array introduced with a panel area equal to 40% of the dwelling footprint area

Phase 2B: A simulated year of continuous heating to a set point of 21oC applying the annual internal gains profile produced by the FHS standardisation. Further three sub-phases were introduced where SAP 10.2 and PHPP both ran with their own internal gain calculations, occupancy calculations and standardised internal gains, instead of the FHS internal gains.

Phase 2C: Comprised of a further three sub-phases which introduce the standardisation of heating set points, programmes, and weather. Subsequently to this phase being undertaken a revision to the FHS standardised Zone 2 set point was undertaken. As such, the results of this phase were superseded by Phase 2D and are not presented in this report.

Phase 2D: A rerun of the Phase 2C subphases, with a Zone 2 set point of 20oC (but still 21oC for Zone 1). These models were run using the same direct electric heating system as the earlier phases. In addition, a heat pump variation was run.

A summary how each model's standardisation was introduced is provided below. Full details of the parameterisation and input alignment for each run can be found in Appendix A.

Phase	Description	Difference with previous phase
2A	Annual simulation with continuous heating, 21oC set points in all zones, a repeating daily gains profile.	• Each packages respective modelling conventions were applied. • Varied gains profile introduced in place of the continuous gains.
2B-a	Annual simulation with continuous heating, 21oC set points in all zones, annual FHS gains profile and LeedsTRY2020High50 weather file applied.	• FHS standardised weather introduced. • FHS calculated gains profile introduced. • AD F required ventilation rates applied. • Geometric inputs adjusted to match architectural drawings in place of ESP-r values.
2B-b	Annual simulation with continuous heating, 21oC set points in all zones, native gains calculated using the FHS number of occupants and LeedsTRY2020High50 weather file applied.	• Each model's gains calculated based on the FHS generated number of occupants.
2B-c	Annual simulation with continuous heating, 21oC set points in all zones, native gains calculated using native occupancy and LeedsTRY2020High50 weather file applied.	• Each model's occupancy calculations enabled, and associated gains.
2B-d	Annual simulation with continuous heating, 21oC set points in all zones, default gains and LeedsTRY2020High50 weather file applied.	• All gains in each model set to their set assumptions and conventions.
2C-a to 2C-c	Runs superseded by 2D runs following decision to revise Zone 2 set point from 18oC to 20oC.	
2D-a	Annual simulation with continuous heating, default set points, default gains and LeedsTRY2020High50 weather file applied.	• Native set points applied (PHPP 20oC, all other models Zone 1 21oC and Zone 2 20oC)

2D-b	Annual simulation with default heating programmes, default set points, default gains and LeedsTRY2020High50 weather file applied.	• Native heating programmes applied (PHPP continuous, all other models intermittent)
2D-c	Annual simulation with default heating programmes, default set points, default gains and default weather I.e. all models run as they would in their real application, with their conventions and set inputs: • HEM-FHS • Certifiable PHPP model • SAP10.2 for Part L calculations	• Native weather files applied.

A summary of the detailed conventions applied, where these differ to the approach adopted for Phase 1, for each modelling package are listed below.

SAP 10.2

- Correction factors for wind speed impacts on infiltration were enabled.
- Infiltration inputs were based on the air permeability (m3/h/m2) at a pressure difference of 50 pascals.
- The calculation of reduced ventilation rates during periods of high infiltration was enabled.
- The Table 6d default shading factor of 0.77 was applied to all transparent building elements, as is required for Part L 2021 calculations.
- For all gains scenarios, the monthly average rate of gain was calculated in W.

PHPP

- The geometry was revised to represent the external dimensions.
- Infiltration inputs were based on air changes per hour calculated at a pressure difference of 50 pascals.
- Default window protection coefficients were applied.
- A window dirt factor of 0.95 was introduced.
- For all gains scenarios, the annual area specific rate of gain was calculated in W/m2.

ESP-r

- Replicated the HEM inputs for all comparisons.

HEM

- The wind driven infiltration calculation module was enabled.
- Infiltration inputs were based on air changes per hour calculated at a pressure difference of 50 pascals.
- The calculation of reduced ventilation rates during periods of high infiltration was enabled.

Neither modelling package has been considered a representation of the true real-world data. As such, the individual statistical indices achieved between HEM and the other modelling packages were considered as a guide only; any observed deviations could then be used to check the HEM core calculation algorithms and the impact of set inputs and conventions.

3.1. Phase 2 Results

Here we present a selection of outputs for each of the Phase 2 runs. The DE results have been presented for the Phase 2A and 2D-c runs. Due to the similarity of results at Phase 1, the DE can be considered to provide representative results for both the SD and TE archetypes. The DA and VF have distinctly different characteristics and have been presented separately where the results present alternative conclusions about the impact of the various model standardisations. The full set of results can be found for the DE, DA and VF archetypes in Appendix D.

3.1.1. Phase 2A

Table 2A_DE Annual Energy Demand

		2A_DE			
		ESP-r	PHPP	SAP10.2	HEM
Annual Energy Demand (kWh/yr)	Space Heating	4,738	5,255	4,470	4,563
	Hot Water				
	Cooling				
	Lighting				
	Ventilation				

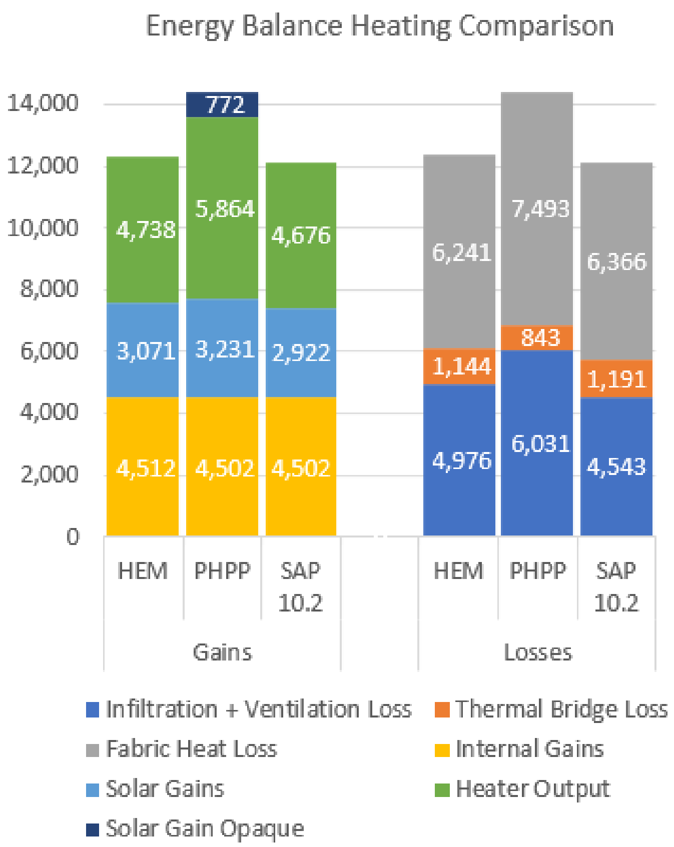
Table 2A_DE Monthly Space Heating Demand

		2A_DE			
		ESP-r	PHPP	SAP10.2	HEM
Monthly Space Heating Demand (kWh)	January	977.0	1032.0	909.2	931.4
	February	706.0	767.0	704.8	732.4
	March	499.0	580.0	584.2	514.4
	April	333.0	381.0	323.1	348.5
	May	122.0	133.0	143.0	120.3
	June	1.0	14.0	0.0	3.3
	July	0.0	2.0	0.0	0.0
	August	0.0	1.0	0.0	0.0
	September	33.0	87.0	0.0	33.9
	October	318.0	387.0	293.6	298.8
	November	742.0	804.0	605.8	626.7
	December	1008.0	1067.0	906.2	953.5

Table 2A_DE Operative and Internal Air Temperatures

		2A_DE			
		ESP-r	PHPP	SAP10.2	HEM
Annual Internal Air (dry bulb) Temperature [average for whole dwelling] (°C)		22.6	21.0	21.0	22.4
Half-Hourly Operative Temperature	Zone 1: Winter Week] (oC)	21.1	21.0	21.0	21.2
	Zone 2: [Winter Week] (oC)	21.0	21.0	21.0	21.0
	Zone 1: [Summer Week] (oC)	25.9	21.0	21.0	25.4
	Zone 2: [Summer Week] (oC)	26.5	21.0	21.0	25.6

Phase 2A - DE - Air Node Energy Balance Comparison Graphs



Phase 2A - DE - External Boundary Energy Balance Comparison Graphs

Table 2A_DE Statistical indicies for CV(RMSE), NMBE, RMSE, MAE for all inter-model comparisons

		ESP-r Vs HEM				PHPP Vs HEM		SAP10.2 Vs HEM		ESP-r Vs PHPP		SAP 10.2 vs PHPP		SAP10.2 Vs ESP-r	
		CV(RMSE)	NMBE	RMSE (oC)	MAE (oC)	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE
Monthly Space Heating (kWh/yr)		11%	4%	0	0	19%	14%	8%	-2%	12%	-10%	23%	-16%	16%	-6%
Monthly Average Internal Air (dry bulb) Temperature [average for whole dwelling] (°C)		1%	1%	0	0	11%	-7%	11%	-7%	12%	7%	0%	0%	12%	-7%
Half-hourly Operative Temperature	Zone 1: Half-hourly Operative Temperature [Winter Week]	0%	0%	0.1	0.0	0	0	0	0	0	0	0	0	0	0
	Zone 2: Half-hourly Operative Temperature [Winter Week]	1%	0%	0.2	0.1	0	0	0	0	0	0	0	0	0	0
	Zone 1: Half-hourly Operative Temperature [Summer Week]	5%	4%	1.0	0.9	0	0	0	0	0	0	0	0	0	0
	Zone 2: Half-hourly Operative Temperature [Summer Week]	5%	4%	1.0	0.9	0	0	0	0	0	0	0	0	0	0

The main differences observed between the Phase 2A and Phase 1B results were due to the introduction of the varied gains profile, which caused all models to reduce their space heating demand. The alignment between PHPP and the other models reduced primarily due to the change in modelling conventions. It is important to note that the alignment between the PHPP and other models achieved under Phase 1 was not expected to continue at Phase 2 due to the convention change. However, the PHPP output is still expected to be a reasonable measure of real-world space heating demand as its application during Phase 2 represents a model which has been broadly validated against measured data.

A significant change was noted for both HEM and SAP 10.2, where the space heating demand reduced relative to the Phase 1B results, suggesting that the impact of the conventions was as significant in SAP10.2 and in HEM, and more significant than both ESP-r and PHPP. The resulting index values for comparisons of SAP 10.2 with PHPP and ESP-r increased.

The statistical indices achieved for the HEM and ESP-r comparison improved for all metrics. It is possible that the HEM algorithms, which began to adjust the rate of infiltration loss relative to the wind speed in Phase 2, counterbalanced the known differences in HTC calculations reported in Phase 1. The energy balances for the HEM and ESP-r also supported this conclusion.

Deck Access Flat

Table 2A_DA Statistical indicies for CV(RMSE), NMBE, RMSE, MAE for all inter-model comparisons

		ESP-r Vs HEM				PHPP Vs HEM		SAP10.2 Vs HEM		ESP-r Vs PHPP		SAP 10.2 vs PHPP		SAP10.2 Vs ESP-r	
		CV(RMSE)	NMBE	RMSE (oC)	MAE (oC)	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE
Monthly Space Heating (kWh/yr)		42%	30%	0	0	114%	88%	22%	-13%	81%	-63%	127%	-99%	59%	-43%
Monthly Average Internal Air (dry bulb) Temperature [average for whole dwelling] (°C)		5%	-4%	0	0	20%	-14%	20%	-14%	16%	10%	0%	0%	16%	-10%
Half-hourly Operative Temperature	Zone 1: Half-hourly Operative Temperature [Winter Week]	0%	0%	0.0	0.0	0	0	0	0	0	0	0	0	0	0
	Zone 2: Half-hourly Operative Temperature [Winter Week]	1%	-1%	0.2	-0.1	0	0	0	0	0	0	0	0	0	0
	Zone 1: Half-hourly Operative Temperature [Summer Week]	8%	-8%	2.1	-2.1	0	0	0	0	0	0	0	0	0	0
	Zone 2: Half-hourly Operative Temperature [Summer Week]	1%	-1%	0.2	0.1	0	0	0	0	0	0	0	0	0	0

The DA archetype experienced a more significant increase, than in the other archetypes, to the statistical indices achieved at Phase 1 between all models I.e. more discrepancies. Two core drivers for this were observed, the significantly higher rate of internal heat gain assumed at Phase 2A, and the convention changes impacting the shading. The internal heat gains were proportionally higher than the other archetypes. The resulting space heating demand was distinctly lower meaning that differences between models appeared more significant.

The HEM and SAP 10.2 conventions adjust the ventilation rate dependent on the infiltration levels; this results in lower ventilation losses during timesteps with high infiltration. Therefore, in smaller dwellings where the infiltration rate was higher due to the small internal volume, ventilation losses were typically lower in the HEM and SAP 10.2 than in the other models. Therefore, the appropriateness of the calculations in the HEM and SAP 10.2 may need to be considered, particularly where the convention for calculating ventilation and infiltration losses is sensitive to both dwelling size and the base infiltration rate. It should be noted that this applied only to dwellings with extract only ventilation systems, those with balanced continuous supply and extract ventilation systems would not be impacted by any changes to these calculations.

3.1.2. Phase 2A – Variations

Table 2A_DE to 2A_DE_OVH Statistical indicies for CV(RMSE), NMBE, RMSE, MAE for all inter-model comparisons

Monthly Space Heating (kWh/yr)	ESP-r Vs HEM		PHPP Vs HEM		SAP10.2 Vs HEM		ESP-r Vs PHPP		SAP 10.2 vs PHPP		SAP10.2 Vs ESP-r	
	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE
2A_DE	11%	4%	19%	14%	8%	-2%	12%	-10%	23%	-16%	16%	-6%
2A_DE_AP5 [Equivalent T50 Air Permeability Rate of 5.0m3/h.m2]	10%	4%	38%	31%	13%	-8%	32%	-28%	49%	-39%	21%	-12%
2A_DE_AP3 [Equivalent T50 Air Permeability Rate of 3.0m3/h.m2]	8%	1%	31%	25%	9%	-2%	28%	-23%	36%	-27%	16%	-4%
2A_DE_AP1 [Equivalent T50 Air Permeability Rate of 1.0m3/h.m2]	12%	-7%	16%	10%	13%	5%	21%	-17%	15%	-5%	21%	13%
2A_DE_ORI [Archetype rotated 90 degree clockwise on its horizontal axis]	11%	5%	21%	17%	10%	-1%	14%	-12%	24%	-18%	17%	-6%
2A_DE_FHS [Thermal resistance values aligned with FHS consultation]	12%	0%	17%	11%	12%	-8%	13%	-11%	26%	-19%	18%	-8%
2A_DE_EXI [Thermal resistance values aligned with VF specificaiton]	17%	-16%	7%	7%	22%	-3%	23%	-22%	22%	-10%	27%	12%
2A_DE_MVH [Continous MVHR employed]	0%	0%	19%	-4%	25%	19%	0%	0%	32%	23%	0%	0%
2A_DE_PVA [Part L 2021 notional PV array introduced]	0%	0%	19%	14%	8%	-2%	0%	0%	23%	-16%	0%	0%
2A_DE_HTP [Heat Pump Heating and Hot Water introduced]	0%	0%	20%	15%	12%	-3%	0%	0%	26%	-18%	0%	0%
2A_DE_OVH [700mm overhangs introduced on all openings]	11%	6%	20%	15%	267%	-200%	10%	-8%	263%	-200%	267%	-200%

The DE archetype statistical indices show that when comparing HEM and ESP-r the variations that exceeded the AG14 thresholds were for the AP1, ORI, EXI and OVH. This supported the conclusions from Phase 1 and 2A which outlined that methodological difference in the calculation of infiltration losses, solar shading and HTCs may be an area of difference between the two models.

Between the HEM and PHPP the base case 2A run was in exceedance of the thresholds and this was also the case for all variations, except for the EXI. The largest changes to both the CV(RMSE) and NMBE values occurred for the AP5, AP3 and AP1 runs. This again supported the conclusion that differences in the combined infiltration and ventilation heat losses may be an area for further validation work. The level of alignment between ESP-r and PHPP at 2A was maintained for the ORI, FHS and OVH variations; however, the index values increased for the AP5, AP3, AP1 and EXI. For the EXI case it was noted that no two models achieve the threshold for both monthly space heating index values.

A notable outlier in the results is the SAP 10.2 comparisons for the OVH variation; this was attributed to the limited methodology applied by SAP 10.2 for solar shading. The energy generation of the PVA models was considered. The DE archetype demonstrated little variation between the models

3.1.3. Phase 2B

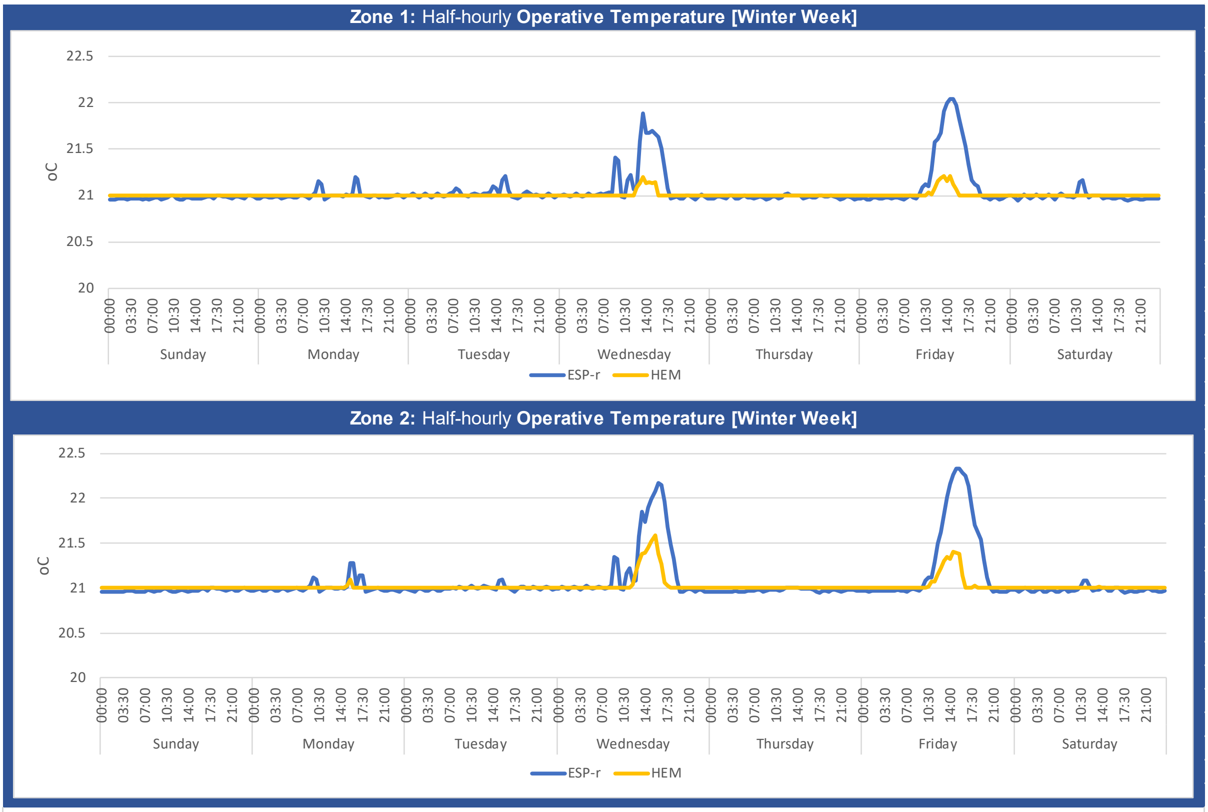


Figure 2Ba_DE Zone Half-hourly Operative Temperature Profile in Typical Winter Week for ESP-r and HEM

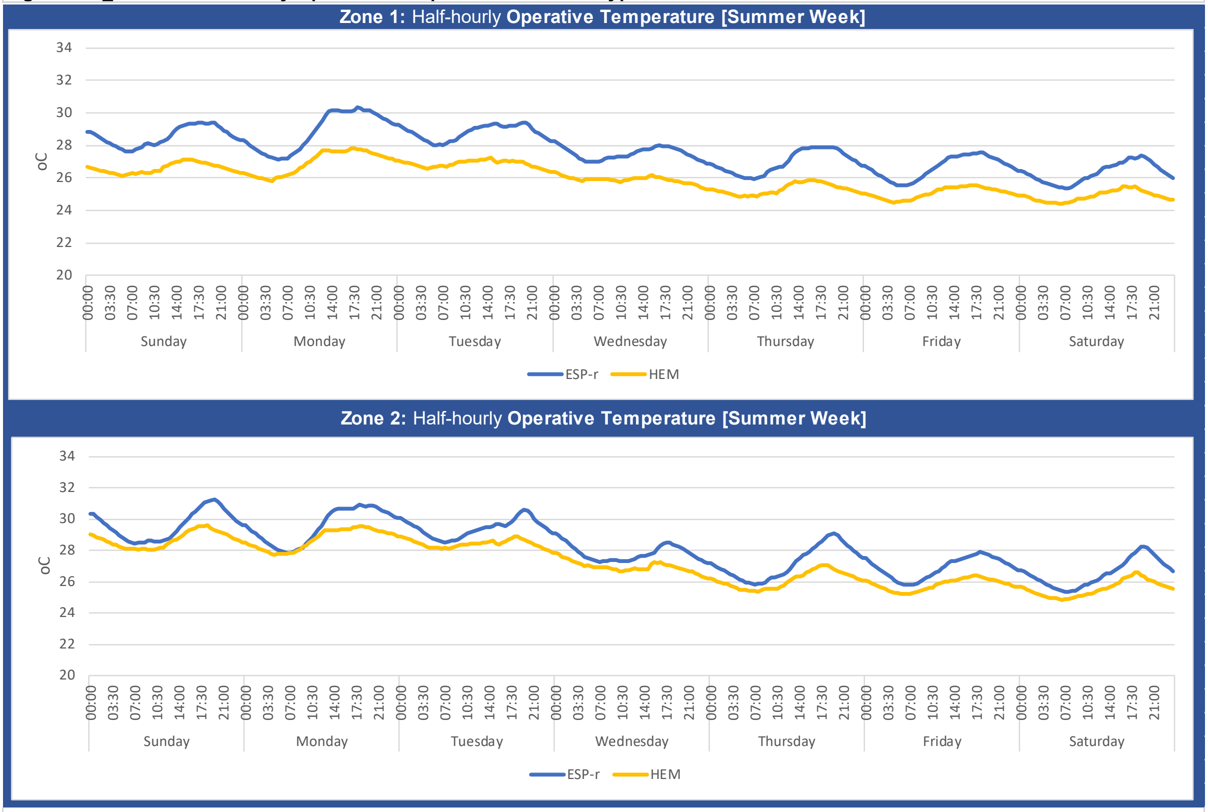


Figure 2Ba_DE Zone Half-hourly Operative Temperature Profile in Typical Summer Week for ESP-r and HEM

Table 2Ba_DE Annual Energy Demand

		2Ba_DE			
		ESP-r	PHPP	SAP10.2	HEM
Annual Energy Demand (kWh/yr)	Space Heating	4,794	4,945	6,119	5,248
	Hot Water				1,483
	Cooling				
	Lighting				119
	Ventilation				110

Table 2Ba_DE Annual Energy Use

		2Ba_DE			
		ESP-r	PHPP	SAP10.2	HEM
Annual Energy Use (kWh/yr)	Space Heating	4794.0	4944.5	6119.1	5247.8
	Hot Water				2569.4
	Cooling				
	Lighting				119.2
	Ventilation				109.6
	Unregulated Uses				2217.9
	Total	4794.0	4944.5	6119.1	10263.9

Table 2Ba_DE Monthly Space Heating Demand

		2Ba_DE			
		ESP-r	PHPP	SAP10.2	HEM
Monthly Space Heating Demand (kWh)	January	979.0	971.0	1082.0	1046.7
	February	780.0	712.0	912.5	870.0
	March	613.0	532.0	803.6	662.7
	April	333.0	341.0	500.9	392.8
	May	81.0	115.0	224.8	127.5
	June	12.0	10.0	72.9	15.7
	July	2.0	1.0	24.2	4.5
	August	0.0	1.0	15.7	0.0
	September	66.0	89.0	183.2	81.4
	October	299.0	385.0	454.2	325.8
	November	672.0	773.0	794.3	736.0
	December	957.0	1013.0	1051.0	984.8

Table 2Ba_DE Operative and Internal Air Temperatures

		2Ba_DE			
		ESP-r	PHPP	SAP10.2	HEM
Annual Internal Air (dry bulb) Temperature [average for whole dwelling] (°C)		22.4	21.0	21.0	22.1
Half-Hourly Operative Temperature	Zone 1: Winter Week] (oC)	21.3			21.3
	Zone 2: [Winter Week] (oC)	21.0			21.0
	Zone 1: [Summer Week] (oC)	28.3			26.5
	Zone 2: [Summer Week] (oC)	28.8			26.7

Table 2Ba_DE Statistical indicies for CV(RMSE), NMBE, RMSE, MAE for all inter-model comparisons

		ESP-r Vs HEM				PHPP Vs HEM		SAP10.2 Vs HEM		ESP-r Vs PHPP		SAP 10.2 vs PHPP		SAP10.2 Vs ESP-r	
		CV(RMSE)	NMBE	RMSE (oC)	MAE (oC)	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE
Monthly Space Heating (kWh/yr)		11%	-9%	0	0	16%	-6%	18%	15%	13%	-3%	27%	21%	27%	24%
Monthly Average Internal Air (dry bulb) Temperature [average for whole dwelling] (°C)		2%	1%	0	0	8%	-5%	8%	-5%	10%	6%	0%	0%	10%	-6%
Half-hourly Operative Temperature	Zone 1: Half-hourly Operative Temperature [Winter Week]	1%	0%	0.0	0.0	0	0	0	0	0	0	0	0	0	0
	Zone 2: Half-hourly Operative Temperature [Winter Week]	1%	0%	0.0	0.0	0	0	0	0	0	0	0	0	0	0
	Zone 1: Half-hourly Operative Temperature [Summer Week]	7%	6%	2.0	2.0	0	0	0	0	0	0	0	0	0	0
	Zone 2: Half-hourly Operative Temperature [Summer Week]	4%	4%	1.0	0.9	0	0	0	0	0	0	0	0	0	0

Table 2Ba_DE to 2Bd_DE Statistical indicies for CV(RMSE), NMBE, RMSE, MAE for all inter-model comparisons

Monthly Space Heating (kWh/yr)		ESP-r Vs HEM		PHPP Vs HEM		SAP10.2 Vs HEM		ESP-r Vs PHPP		SAP 10.2 vs PHPP		SAP10.2 Vs ESP-r	
		CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE
Detached	2Ba_DE	11%	-9%	16%	-6%	18%	15%	13%	-3%	27%	21%	27%	24%
	2Bb_DE			18%	11%	11%	1%	25%	-20%	19%	-10%	14%	10%
	2Bc_DE			19%	12%	12%	-1%	27%	-21%	23%	-14%	13%	8%
	2Bd_DE			19%	11%	10%	3%	26%	-20%	18%	-9%	15%	12%
Deck Access Flat	2Ba_DA	6%	1%	16%	7%	57%	-48%	17%	-6%	68%	-55%	58%	-49%
	2Bb_DA			26%	20%	100%	-83%	26%	-18%	119%	-98%	101%	-84%
	2Bc_DA			28%	22%	101%	-83%	28%	-20%	121%	-100%	101%	-84%
	2Bd_DA			34%	28%	88%	-73%	33%	-27%	114%	-96%	89%	-74%
Victorian Flat	2Ba_VF	28%	21%	24%	19%	22%	20%	13%	2%	12%	1%	12%	-1%
	2Bb_VF			35%	29%	11%	9%	14%	-9%	26%	-20%	21%	-12%
	2Bc_VF			36%	31%	12%	11%	15%	-10%	26%	-20%	20%	-10%
	2Bd_VF			34%	28%	15%	13%	14%	-8%	21%	-15%	17%	-7%

With the FHS predicted internal gain applied, the CV(RMSE) index value between the HEM and ESP-r was only marginally impacted, indicating that geometric updates did not have a significant impact. Generally the indices reported for all model comparisons at Phase 2Ba were worse than 2A, which suggested that the change to the gains profile, background ventilation rate and weather data exacerbated the differences identified in the Phase 1 analysis.

As the PHPP and SAP 10.2 were run using their own internal gains and occupancy calculations (Phase 2B-b), the divergence between these models with both the HEM and ESP-r increased. Proportionally the largest divergence occurred when both PHPP's and SAP 10.2's own internal gains calculations were employed, despite still using the FHS occupancy level. This was the case in all archetypes. This outlined a significant difference in the way that gains are calculated from a set occupancy rate between PHPP, SAP 10.2 and the HEM. For 2B-d the differences between models were similar in magnitude to those observed for 2B-b. Therefore, it can still be concluded that the interpretation of internal gains is still a core driver of difference between PHPP, SAP 10.2 and the HEM, regardless of whether these gains are calculated, or the standard gains value is applied.

3.1.4. Phase 2D

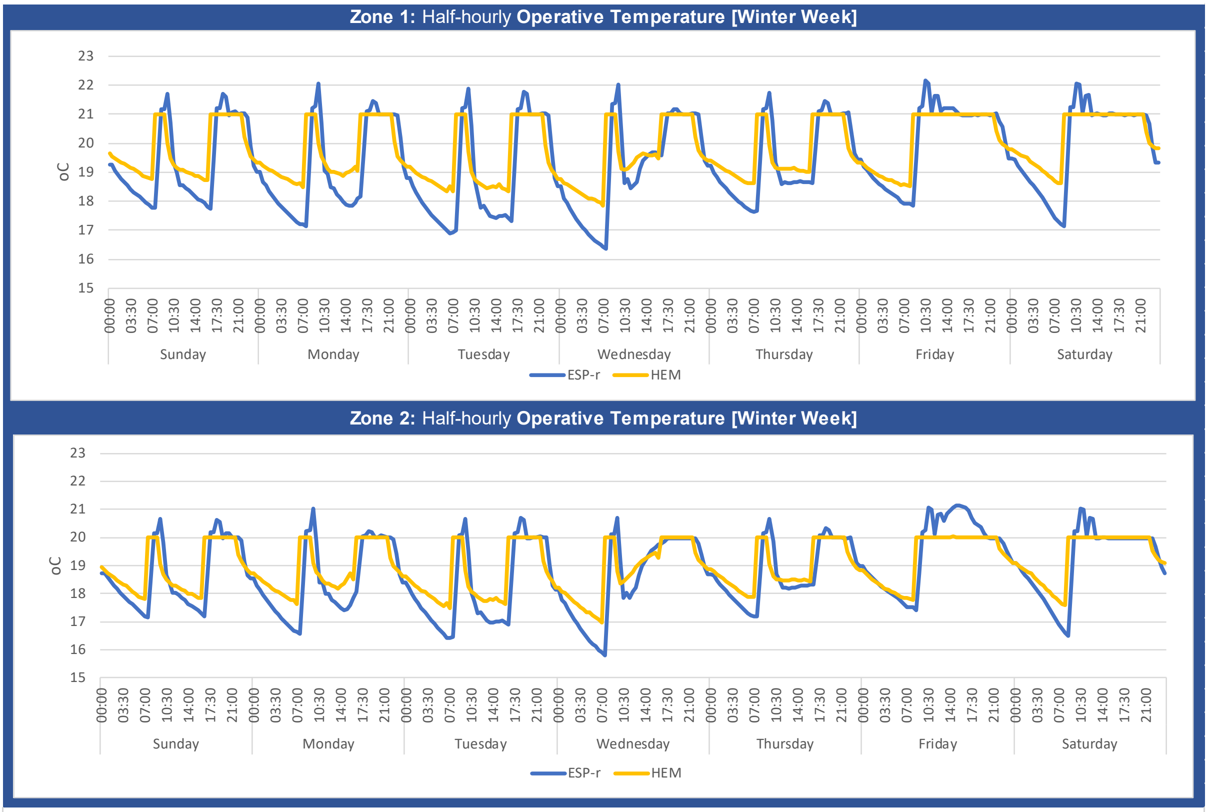


Figure 2Dc_DE Zone Half-hourly Operative Temperature Profile in Typical Winter Week for ESP-r and HEM

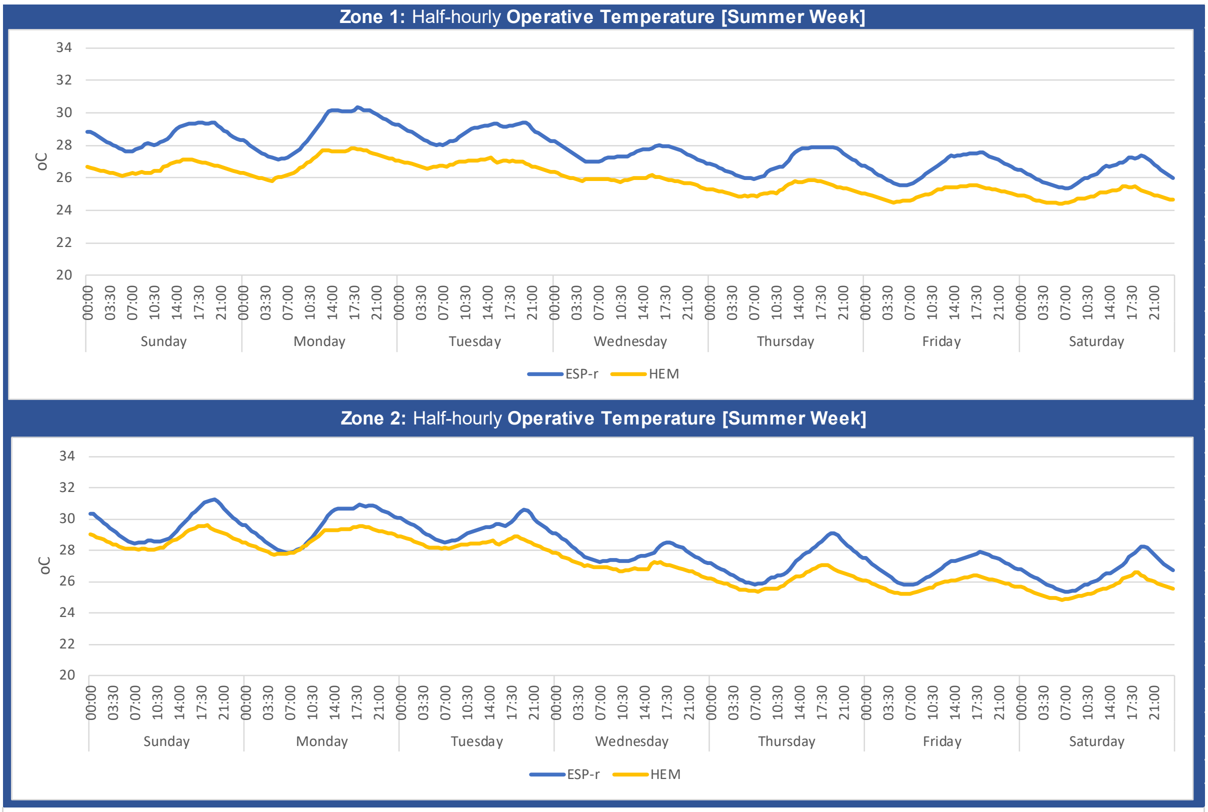


Figure 2Dc_DE Zone Half-hourly Operative Temperature Profile in Typical Summer Week for ESP-r and HEM

Table 2Dc_DE Annual Energy Demand

		2Dc_DE			
		ESP-r	PHPP	SAP10.2	HEM
Annual Energy Demand (kWh/yr)	Space Heating	3,575	5,465	4,081	4,143
	Hot Water		1,027	1,255	1,483
	Cooling		0	129	0
	Lighting		72	n/a	119
	Ventilation		58	n/a	110

Table 2Dc_DE Annual Energy Use

		2Dc_DE			
		ESP-r	PHPP	SAP10.2	HEM
Annual Energy Use (kWh/yr)	Space Heating	3575.0	5464.8	4080.5	4140.4
	Hot Water		1770.8	2089.5	2569.4
	Cooling		0.0	0.0	n/a
	Lighting		72.0	189.3	119.2
	Ventilation		99.0	130.7	109.6
	Unregulated Uses		1306.2	2761.4	2217.9
	Total	3575.0	8712.7	9251.4	9156.4

Table 2Dc_DE Monthly Space Heating Demand

		2Dc_DE			
		ESP-r	PHPP	SAP10.2	HEM
Monthly Space Heating Demand (kWh)	January	773.0	1045.0	772.1	851.3
	February	601.0	819.0	606.5	710.3
	March	456.0	683.0	529.5	530.8
	April	216.0	428.0	333.1	284.1
	May	35.0	157.0	176.9	68.9
	June	3.0	15.0	0.0	4.2
	July	0.0	1.0	0.0	0.4
	August	0.0	0.0	0.0	0.0
	September	27.0	55.0	0.0	43.7
	October	191.0	412.0	309.1	228.8
	November	519.0	785.0	556.8	596.3
	December	754.0	1063.0	796.4	823.7

Table 2Dc_DE Operative and Internal Air Temperatures

		2Dc_DE			
		ESP-r	PHPP	SAP10.2	HEM
Annual Internal Air (dry bulb) Temperature [average for whole dwelling] (°C)		21.2	20.0	19.2	21.1
Half-Hourly Operative Temperature	Zone 1: Winter Week] (oC)	18.8			19.3
	Zone 2: [Winter Week] (oC)	19.2			19.6
	Zone 1: [Summer Week] (oC)	28.3			26.5
	Zone 2: [Summer Week] (oC)	28.8			26.7

Table 2Dc_DE Statistical indicies for CV(RMSE), NMBE, RMSE, MAE for all inter-model comparisons

	ESP-r Vs HEM				PHPP Vs HEM		SAP10.2 Vs HEM		ESP-r Vs PHPP		SAP 10.2 vs PHPP		SAP10.2 Vs ESP-r	
	CV(RMSE)	NMBE	RMSE (oC)	MAE (oC)	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE
Monthly Space Heating (kWh/yr)	18%	-15%	0	0	34%	27%	17%	-2%	51%	-42%	39%	-29%	22%	13%
Monthly Average Internal Air (dry bulb) Temperature [average for whole dwelling] (°C)	2%	0%	0	0	11%	-5%	12%	-9%	13%	6%	6%	-4%	13%	-10%
Zone 1: Half-hourly Operative Temperature [Winter Week]	5%	-1%	1.5	0.2	0	0	0	0	0	0	0	0	0	0
Zone 2: Half-hourly Operative Temperature [Winter Week]	5%	-1%	1.3	0.3	0	0	0	0	0	0	0	0	0	0
Zone 1: Half-hourly Operative Temperature [Summer Week]	7%	6%	2.0	2.0	0	0	0	0	0	0	0	0	0	0
Zone 2: Half-hourly Operative Temperature [Summer Week]	4%	4%	1.0	0.9	0	0	0	0	0	0	0	0	0	0

Table 2Bd_DE to 2Dc_DE Statistical indicies for CV(RMSE), NMBE, RMSE, MAE for all inter-model comparisons

Monthly Space Heating (kWh/yr)		ESP-r Vs HEM		PHPP Vs HEM		SAP10.2 Vs HEM		ESP-r Vs PHPP		SAP 10.2 vs PHPP		SAP10.2 Vs ESP-r	
		CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE	CV(RMSE)	NMBE
Detached	2Bd_DE	11%	-9%	19%	11%	10%	3%	26%	-20%	18%	-9%	15%	12%
	2Da_DE			17%	7%	11%	0%	22%	-16%	18%	-7%	13%	9%
	2Db_DE	18%	-15%	30%	21%	21%	-10%	45%	-36%	44%	-31%	16%	5%
	2Dc_DE			34%	27%	17%	-2%	51%	-42%	39%	-29%	22%	13%
Deck Access Flat	2Bd_DA	6%	1%	34%	28%	88%	-73%	33%	-27%	114%	-96%	89%	-74%
	2Da_DA			30%	22%	94%	-77%	29%	-21%	116%	-95%	96%	-78%
	2Db_DA	8%	5%	56%	45%	82%	-66%	52%	-40%	127%	-104%	87%	-70%
	2Dc_DA			62%	50%	70%	-57%	57%	-46%	122%	-100%	75%	-61%
Victorian Flat	2Bd_VF	28%	21%	34%	28%	15%	13%	14%	-8%	21%	-15%	17%	-7%
	2Da_VF	26%	19%	29%	23%	10%	8%	13%	-4%	22%	-15%	20%	-11%
	2Db_VF	21%	15%	49%	40%	12%	0%	31%	-26%	51%	-40%	26%	-15%
	2Dc_VF			55%	46%	16%	8%	36%	-31%	47%	-38%	20%	-7%

Under the fully standardised scenario (2D-c) it was observed that all comparison exceeded at least one of the CV(RMSE) or NMBE thresholds for monthly space heating. This demonstrates that the standardisation of the models is likely to be the most significant driver of differences between models in their intended application, the HEM:FHS version included.

As a headline comparison, it can be noted that, compared to Phase 1B, the standardisation has a much smaller effect in PHPP (reduction of space heating demand by about 10% in the DE archetype) than in SAP10 and HEM (reduction by about one third). This is partly attributed to the continuous heating assumption within the PHPP standardisation.

More detailed observations can also be made. The most notable change to the statistical indices of all model comparisons occurred with the introduction of standardised heating hours. A change to the heating hours was expected to have a significant impact, particularly as it was understood that the way that each model simulated the rate of cool down and heat loss when at lower internal temperatures differed. For example, the differences in thermal mass calculation and HTCs noted in Phase 1 between ESP-r and the HEM have been exacerbated by the introduction of the FHS intermittent heating programme. This is evident in both Zone 1 and Zone 2 half-hourly internal operative temperature profiles for the typical winter week. The worsening of index values between PHPP and the other models was expected to be driven by the difference in heating programme, which meant that the periods of low heat loss between heating periods was not replicated by PHPP's continuous heating simulation.

A notable change in the alignment was observed because of the standardisation of weather. The comparatively high average dry bulb external temperature applied by the FHS standardisation resulted in lower monthly space heating demands for both ESP-r and the HEM relative to SAP 10.2 and PHPP.

The annual energy use produced indicated that PHPP predicted the lowest energy use for all categories: its lower space heating demand prediction is balanced by lower calculated energy use for domestic hot water, ventilation, lighting, and unregulated energy uses. For ventilation, PHPP required a lower ventilation rate because of the conventions for calculating ventilated volume differing to SAP 10.2 and the HEM. It was noted that the HEM run hours were based on continuous operation; however, this may not match the annual ventilation rate used to calculate the annual ventilation heat loss. For lighting the assumed average lumens output of the lighting systems differed between models and was the core driver of differences. The unregulated energy use was understood to be different due to the type of calculation used by each model. PHPP employs a bottom-up calculations based on an assumed mix of appliances. Both SAP 10.2 and the HEM use a top-down approach informed by EFUS datasets which is based on a different standardised appliance mix and usage pattern. The HEM and SAP 10.2 differed as the dataset used to inform the top-down calculations were based on different survey years.

More significant differences in energy use were observed for the space heating and hot water systems. The differences in space heating energy use were a direct result of the factors discussed above in relation to the space heating demand. However, the hot water demands were found to differ because of the standardised occupancy assumptions and consumption rates. Each model simulated the distribution losses and the standing losses of the assumed hot water cylinder. Primary pipework losses were not considered due to a direct acting immersion being assumed. Differences in the distribution pipework losses were expected due to technical differences in the algorithms and the usage patterns. PHPP was observed to experience lower losses than the HEM due to the average number of tapping events per person being lower than the FHS assumption, 6 tapping events per person per day in PHPP against an average of 11.5 calculated for the FHS. As a result, the energy lost from dead leg losses after each tapping event was expected to be higher in the HEM. SAP 10.2 applied a default 15% loss; therefore, it is not sensitive to actual design and non-comparable to the HEM or PHPP.

The storage tanks losses also differed because of methodology and standardisation. Both PHPP and the HEM assumed a 60oC storage temperature, whereas SAP 10.2 assumed 55oC.The HEM calculated the standing losses based on the explicitly calculated tank temperature in any one timestep. PHPP calculated this based on a constant 60oC condition. The HEM standing loss calculation is expected to be the most accurate due to its high resolution and used of the predicted tank temperature. It was noted that the additional losses expected for the PHPP standing loss calculation counterbalanced the lower distribution losses resulting from the lower tapping event frequency. The result was that PHPP and the HEM predicted the same proportional hot water system losses, which were 72% and 73% respectively.

The difference in energy uses other than for space heating is understood; however, some further investigation of the algorithms used to calculate energy use for the normalised dwelling conditions is recommended. These energy uses cumulatively account for significant proportion of the total dwelling energy consumption and should be given similar weight in future validation work to the space heating demand that has been extensively tested in this study. Furthermore, these energy uses directly impact the space heating demand through provision of internal gains.

3.1.5. Phase 2D HP

Table 2Dc_HP_DE Annual Energy Demand

Annual Energy Demand (kWh/yr)		2Dc_HP_DE			
		ESP-r	PHPP	SAP10.2	HEM
	Space Heating		5465.0	4541.2	5714.2
	Hot Water		1027.0	1854.7	1483.2
	Cooling		0.0	0.0	0.0
	Lighting		72.0	189.3	119.2
	Ventilation		58.0	130.7	109.6

Table 2Dc_HP_DE Annual Energy Use

Annual Energy Use (kWh/yr)		2Dc_HP_DE			
		ESP-r	PHPP	SAP10.2	HEM
	Space Heating		1844.0	2177.0	1147.1
	Hot Water		570.0	712.4	1536.3
	Cooling		0.0	0.0	0.0
	Lighting		72.0	189.3	119.2
	Ventilation		58.0	130.7	109.6
	Unregulated Uses		1306.0	4019.0	2217.9
	Total		3850.0	7228.4	5130.1

Table 2Dc_HP_DE Monthly Space Heating Demand

Monthly Space Heating Demand (kWh)		2Dc_HP_DE			
		ESP-r	PHPP	SAP10.2	HEM
	January		1045.0	947.9	1226.2
	February		819.0	676.4	950.6
	March		683.0	597.6	674.5
	April		428.0	385.1	357.3
	May		157.0	214.6	82.0
	June		15.0	0.0	4.2
	July		1.0	0.0	0.5
	August		0.0	0.0	0.0
	September		55.0	0.0	60.0
	October		412.0	363.9	296.0
	November		785.0	628.4	854.0
	December		1063.0	909.4	1208.8

Table 2Dc_HP_DE Operative and Internal Air Temperatures

		2Dc_HP_DE			
		ESP-r	PHPP	SAP10.2	HEM
Annual Internal Air (dry bulb) Temperature [average for whole dwelling] (°C)			20.0	19.2	21.2
Half-Hourly Operative Temperature	Zone 1: Winter Week] (oC)				19.8
	Zone 2: [Winter Week] (oC)				20.1
	Zone 1: [Summer Week] (oC)				26.5
	Zone 2: [Summer Week] (oC)				26.8

The introduction of a ASHP serving a wet distribution system, which is less responsive than the direct electric heating assumed in the previous phases, had a notable impact on the space heating demand in both SAP 10.2 and the HEM. The difference in the HEM space heating demand was understood to be due to the higher mean internal temperature, caused by the introduction of a setback temperature of 18oC in the HEM. The setback was introduced to represent the more continuous operation of the heating system expected for heat pumps. Additionally, the low responsiveness of the system resulted in unmet demand being reported, this highlighted that the way that the HEM presents heating demand may be disproportionally impacted by unmet demand and should be an area for further development for the HEM. The space heating demand of SAP 10.2 increased due to the higher mean internal temperature, which was caused in turn by the longer heating system running hours calculated from Appendix N of the SAP 10.2 methodology.

When considering the heat pump SCOP reported in each model, it is important to note the system boundary. PHPP reports both a H1[10] and H3[11] system boundary, SAP 10.2 a H3 system boundary. At the time of writing, it was not possible to obtain the equivalent SCOP from the HEM. This is also the case for SCOP for hot water generation. Further validation exercises should be completed for heat pump systems once the HEM can calculate the SCOPs for the same system boundary as the comparator models. It should also be noted that the appropriateness of using identical heat pump units may have an impact on the reported efficiency, particularly where the space heating demands between the models differs.

Comparisons between the different models also became complex for the heat pump runs. The emitter circuit characteristics and heat pump controls were observed to be key sensitivities between the phases. For the continuous heating scenario, the emitter circuit was sufficiently sized; however, when intermittent heating was introduced, unmet demand was reported. As such, it was not appropriate to compare the space heating energy use with the space heating demand results of the HEM as this does not account for any unmet demand.

3.2. Phase 2 Conclusions

Phase 2 of this IMC applied a ‘same design, but with conventions and standardised inputs’ modelling approach to demonstrate how the HEM with its FHS standardised assumptions compared to the standardised comparator model PHPP. ESP-r was retained in this phase of comparison to ‘mimic’ the HEM and test the impact of the FHS standardisation on the building physics algorithms validated in Phase 1 . Results indicated how the comparator models aligned after each core component of the standardisation was introduced. The comparison

[10] COP(H1) = (heat pump heat output) / (electricity input for heat pump* + electricity for circulation pump on borehole or ground array)

[11] COP(H3) = (heat pump heat output + backup heater heat output) / (electricity input for heat pump* + electricity for circulation pump on borehole or ground array + backup heater electricity use)

between the HEM and SAP 10.2 were presented as context on how compliance calculations will differ when the HEM is adopted as the national methodology.

The comparison between the HEM and the comparator models at various Phases of the IMC indicated that testing with the intended FHS normalisation should be a particular focus of future validation exercises. The intent to validate the suitability of the HEM for compliance modelling should demonstrate reasonable levels of accuracy through its intended application. This IMC study has indicated that, under certain applications, the HEM is suitably aligned with the comparator models; however, under its intended application the level of agreement was much reduced. Therefore, it is recommended that further testing of the HEM applies the intended FHS normalisation to ensure that the perceived general application of the HEM is suitably valid. This recommendation should also be extended to any further normalisations that may be considered for other compliance calculations, in particular EPCs.

Specific areas have also been identified for further validation, including a wider range of characteristics (beyond characteristics representative of new build dwellings), and ventilation systems – see details in the Limitations section.

4. Limitations

Some limitations to the validation exercise and how closely their outputs can be compared are inherent to each model, as detailed for each model at the start of the Phase 1 section.

In addition, and as noted in the Introduction to this report, the validation exercise is on-going, and this report presents the current status of HEM development and validation. By its nature, validation is iterative, so further areas to look into are identified through each step, and gradually resolved. Key limitations in the scope of the current exercise include:

- A relatively limited sample of 5 archetypes in Phase 1, and 3 in Phase 2, most of them representative of new build characteristics. The modelling that has been completed on the VF archetype has demonstrated that alignment between the HEM and the comparator models is poorer for existing buildings. Although many of the observations of methodological differences between the models may contribute to this difference, it is recommended that further testing is conducted between the HEM and the comparator models, most notably ESP-r.
- In addition, validation exercises with field data for older dwellings may also prove useful to validate that the HEM can accurately predict the space heating demand of existing dwellings with higher rates of heat loss.
- So far, most of the focus has been on space heating demand, internal temperature and building physics. While some validation of systems has been carried out, this is more limited. More interrogation of how systems are modelled will need to be carried out.
- Alignment of the models when differing levels of air tightness was applied indicated a difference in the core algorithms. The current HEM assumption may reduce the estimated benefit of an MVHR system compared to a natural ventilation or decentralised mechanical extract solution.
- This approach also highlighted that the energy consumption calculated for a continuous operation extract ventilation system may be based on a different ventilation rate that is used for the heat loss calculations. Furter validation work should be conducted on ventilation and infiltration impacts. This could include gathering evidence on the operation of WHEV and NV in existing homes to inform the assumptions within the HEM. Consideration could also be given to the calculation module for ventilation system energy consumption to ensure that it applies the same assumption as the ventilation module for the annual running hours.

5. Overall conclusions and recommendations

The analysis contained within this report outlines how the HEM has been found to differ from the established and validated comparator models ESP-r and PHPP. The differences with SAP 10.2 have also been documented to understand how compliance calculations will differ when the HEM is adopted as the national methodology.

The recommendations of this study relate exclusively to the HEM, its future development, and its application with FHS standardisation. Conclusions relating to further research or recommendations for the other modelling packages in the IMC were developed outside of this report.

In Phase 1 the Home Energy Model was shown to agree with its comparators, being within the threshold values for most of the statistical indices applied. The differences between the space heating demand of the HEM and the comparator models varied between archetypes. For example, the HEM predicted demand was higher the ESP-r but less the PHPP figure the 1B DE and DA runs; for the 1B SD TE and VF runs, the HEM was higher than both ESP-r and PHPP.

Where statistical threshold values were not achieved, the differences in calculations methodology were understood to be the driver and were explainable. These differences were accepted as not being of a high priority to resolve during this phase of the validation. Instead, they were considered areas for further investigation and possible future development. The Phase 1 findings were used to refine the HEM. These refinements were made prior to Phase 2 of this IMC.

Phase 2 demonstrated that agreement with PHPP reduced after the standardisations were applied. This is expected to some extent, as each model carries different standardisations. PHPP is considered as an accurate representation of conditions in the field. Therefore, further consideration of the accuracy of the HEM:FHS standardisation may be required.

Each phase of this IMC study informed a series of recommendations for future validation exercises and development of the HEM. These are listed below. It is understood that the suitability of methodological changes discussed above will need to be considered alongside the intended uses of the HEM. Therefore, potential implications of any changes to the usability of the HEM, and the availability of information and evidence to support compliance calculations must be considered.

- Consider how better alignment with ESP-r could be achieved by altering the input parameterisation of thermal mass for building elements. This would be a departure from the BS EN 52016-1:2017 approach, with a number of alternatives available.
- Closer alignment with ESP-r's solar gain calculations could be achieved by more accurately accounting for variation in reflectivity of windows, depending on the incident

angle of sunlight. This also means going beyond the methodology in BS EN 52016-1:2017. The static reflectivity value taken from the Standard is lower than that used in PHPP and the time-average derived from the varying ESP-r values, leading to higher solar gains overall.

- Accounting for inter-zone heat transfers would potentially improve alignment of heat demand with both ESP-r and PHPP. BS EN 52016-1:2017 does provide a methodology to account for this effect.
- Undertaking further validation of the heat transfer coefficient values applied within the Home Energy Model and considering the appropriateness of dynamically calculated heat transfer coefficients (HTCs), as used in ESP-r.
- So far, most of the focus has been on space heating demand, internal temperature and building physics. While some validation of systems has been carried out, this is more limited. More interrogation of systems, especially heat pumps, is needed. To enable this validation, the HEM outputs should be adjusted to allow for the calculation of SCOP at both a H1 and H3 system boundaries.
- Alignment of the models when differing levels of air tightness was applied reflected a difference in the core algorithms. The current Home Energy Model approach may reduce the estimated benefit of an MVHR system compared to a natural ventilation or decentralised mechanical extract solution.
- Further work is needed to develop the Home Energy Model's treatment of infiltration and ventilation, to improve alignment and internal consistency with the heat loss calculations.
- Further work is need on the HEM with the FHS standardisation against more representative data from real dwellings to ensure it is valid for in intended application of rating both new build and existing dwelling energy use.
- Comparison of different heating systems and a range of heating schedules to further explore the performance of these under the FHS assessment assumptions.

The consultation version of the HEM has been found to be sufficiently valid as a building physics engine; however, further work will be required to ensure validity is maintained for the intended application to rate both new and existing dwelling energy use. The Home Energy Model will continue to be validated against established models throughout its development to improve its accuracy, including further examination of older building typologies.

This publication is available from: https://www.gov.uk/government/publications/home-energy-model-validation-documentation