26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Simulation of residential energy-efficiency star rating and thermal requirement using NatHERS certificate data

C-H Wang[a] and AM Rojas[a]

[a] CSIRO Energy, Clayton, Victoria, Australia
Email: chi-hsiang.wang@csiro.au

Abstract: In a study to investigate the interaction between residential energy consumption and electric grids for a net zero-carbon future, we develop statistical models needed for dwellings of unknown energy efficiency star rating and thermal requirement. The needs stem from common situations requiring time-series residential energy requirement but only a proportion of the dwellings in the area had sufficient information available to the study. The developed models address the limited availability of real-life energy consumption data and how we can leverage the available data from a proportion of the dwellings in a study area.

The area studied is in postcode 3216, Geelong, Victoria (Figure 1), and the dwellings considered are detached residential housing of building class 1A, National Construction Code of Australia. A data set of 2,126 rated dwellings is extracted from the CSIRO database of certificates assessed by the Nationwide House Energy Rating Scheme (NatHERS) accredited tools. Another data set of 26,562 un-rated dwellings of the building stock in the postcode 3216 is collated from (a) Geelong City Council up to 2013, and (b) OpenStreetMap after 2013. The two data sets are fully-joined (i.e. keeping all distinct data points in both but deleting duplicates) to obtain 27,876 detached dwellings for energy-efficiency modelling and estimation.

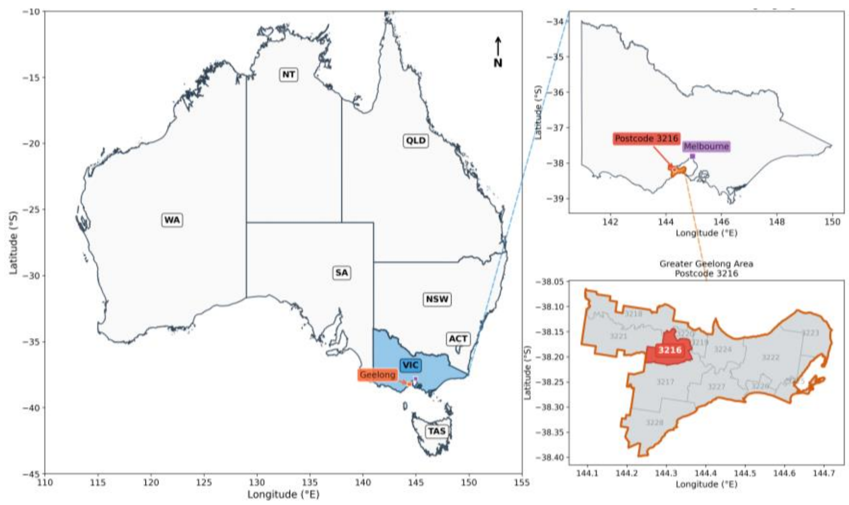


Figure 1 Hierarchical maps showing the area of postcode 3216, Geelong, Victoria, Australia

Because of the differences in statistical characteristics, we divide the rated data into three dwelling groups: (a) 'B6S' for dwellings constructed prior to 2012, (b) 'N6S' for dwellings constructed in or after 2012, and (c) 'R6S' for dwellings renovated in or after 2012, and whether the energy-efficiency star rating is less or greater than six stars. For B6S, however, only 4 of the 148 rated dwellings are rated ≥ 6 stars, thus we assume provisionally that it has the same statistical characteristics as the N6S ≥ 6 stars, leading to a total of five analysis data sets. Beta regression of star rating on one of two predictors available in the data set: Year of construction and distance to city centre, is conducted for each of the five data sets. The total thermal requirement (i.e. heating + cooling requirements) of a star rating can be obtained by interpolation of the NatHERS star bands for dwellings rated in [0.5, 10] stars. For dwellings of zero star, the total requirement is treated as a random variable and the CSIRO zero-star certificate data (206 certificates) in climate zone 60 is fitted to the gamma probability distribution.

Based on the developed statistical models, Monte Carlo simulation is carried out to generate star ratings and the associated thermal requirements of the un-rated dwellings. These will then be used to simulate hourly time-series of heating and cooling requirements of each dwelling needed in our study of the interaction between residential electric energy consumption and electricity grids for a net-zero future.

Keywords: *Residential energy efficiency, star rating, thermal requirement, statistical model*

1. INTRODUCTION

Studies involving energy efficiency and consumption of homes in a community require the information of energy-efficiency performance of dwellings. Past energy consumption from sources such as smart metres provides valuable information to understand household energy requirements, conduct informed analysis, and derive useful consumption models. However, acquiring smart meter time-series data is usually difficult because of proprietary restriction and confidentiality of private information. One avenue to obtain such information in Australia is making use of the energy-efficiency certificate data produced by software accredited by the Nationwide House Energy Rating Scheme (NatHERS). However, typically only a small fraction of dwellings in a community was assessed as it is labour-intensive and time-consuming to input thousands of necessary house characteristics to the software. To overcome the data scarcity issue, one viable approach is to develop statistical energy-efficiency models using the limited number of assessed dwellings and use Monte Carlo simulation to generate energy-efficiency performance measures for un-assessed dwellings.

For an ongoing study of the interaction between residential electricity consumption and electric grids (refer to a companion abstract (Rojas *et al*., 2025)), this paper presents statistical modelling and Monte Carlo simulation for generating the energy efficiency star rating as well as the associated total (heating + cooling) thermal requirements of houses, i.e. building class 1A as specified in the National Construction Code (NCC). Detached and semi-detached houses account for approximately 70 % of Australian homes (Australian Bureau of Statistics, 2022).

Since the early 2000s, the energy efficiency of new homes and major renovations in Australia has had to be rated to demonstrate compliance with the NCC energy efficiency requirements and gain approval for construction. On behalf of all states and territories, NatHERS is responsible for improving and developing an energy efficiency rating scheme and accredits four rating tools (Commonwealth of Australia, 2022a): AccuRate Home, BERS Pro, FirstRate5, and HERO for this purpose. The assessed energy performance is expressed by an energy efficiency star rating ranging from 0 to 10 stars; a higher star rating represents a more energy efficient home and hence lower thermal requirement. All the tools use a CSIRO-developed computational engine, Chenath (Commonwealth of Australia, 2022a), for estimating thermal requirement and star rating.

As the requirements of energy-rating certificate for building approval were only adopted by all Australian states and territories since 2003, the energy efficiency of many dwellings in the current building stock has not been rated. For energy consumption studies at a community scale, the star ratings and the associated thermal requirements of un-rated dwellings need to be estimated by statistical methods. As a demonstration, we choose the area located in postcode 3216, Geelong, Victoria, covered by NatHERS climate zone 60 (Commonwealth of Australia, 2022b). The study area of postcode 3216 (Figure 1) occupies mainly the south west corner of the City of Greater Geelong, spread approximately between 2 and 13 kilometres to the Town Hall, which is used as the proxy location of the central business district.

2. DATA

Star-rating certificates have been collected by CSIRO with the support of Sustainability Victoria since May 2016 with monthly updates. Useful summary statistics and analysis results for heating and cooling energy requirements as well as star ratings from the collected certificates have been made available on the Australia Housing Data website (CSIRO, 2025) hosted by CSIRO. The certificates were submitted by independent assessors (inevitably the input data would be subject to inadvertent entry errors) and the NatHERS Administrator had no control over the collection of the data.

The data set with star ratings used in this study for postcode 3216 was extracted from the CSIRO certificate database. The initial query returned a data set of 4,752 dwellings. After data cleaning and tidying, 2,126 certificates were available for analysis. Another data set related to the building stock in the postcode was obtained by collating the data sets from the Geelong City Council (City of Greater Geelong, 2024) for dwellings up to 2013 and from OpenStreetMap (OpenStreetMap contributors, 2024) for dwellings from 2013 onwards. After data tidying, a total of 26,562 residential dwellings classified as building class 1A by the NCC were obtained in the collated data set. Of the 26,562 dwellings in the building stock, 812 have an accredited star rating. The remaining 25,750 dwellings without star rating will have their star rating estimated based on the probability models developed in Section 3.

The energy efficiency of Australian residential construction has gradually improved from practically no consideration before 1990 to a minimum energy-efficiency star rating of 7 since 2023. In particular, a minimum of 6-star requirement was implemented from 2012 to 2022. As such, the building stock dwellings are categorised into: (a) N6S: New dwellings, constructed between 2012 and 2022; (b) R6S: Dwellings renovated between 2012 and 2022; and (c) B6S: Dwellings constructed prior to 2012.

The building-stock data set provides only two statistically significant predictors: year of construction and distance to city centre; furthermore, some dwellings have missing predictor values. The missing predictor values and the unknown dwelling group (N6S or R6S) associated with post-2011 construction are imputed by the method of multivariate imputation by chained equation using the R package mice (van Buuren & Groothuis-Oudshoorn, 2024). The imputed data set is used for statistical modelling of star rating as described in the next section.

3. METHOD

Even though the NatHERS star bands provide a look-up table for the relationship between star rating and total requirement, the minimum rating specified is 0.5 star (corresponding to a maximum of 797 MJ/m^2 for climate zone 60) and any requirement greater than that is regarded as zero star. That is, the energy required by a zero-star dwelling could be any value > 797 MJ/m^2, hence it may be treated as a random variable modelled by a probability distribution.

This section presents the analysis and modelling of star rating and the corresponding total (i.e. heating + cooling) requirements of detached (NCC class 1A) dwellings. The un-rated dwellings constructed prior to and including 2011 are classified as B6S dwellings, and from 2012 onwards are either N6S or R6S. It is noted that the statistical model fitting performed in this study uses the entire analysis data set rather than splitting it into training and validation data sets as typically carried out when conducting machine-learning model fitting. The main reasons are that: (a) The star rating has lower and upper bounds, for which beta or logistic regression can usually model the data well. Moreover, no excessive non-linearity or irregularity is observed when plotted against likely predictors (i.e. construction year and distance to city centre); (b) Very few useful predictors are present in the data set. After accounting for statistical significance, each statistical model uses either zero or one linear predictor only. The zero-predictor models have no R^2 values, whereas the one-predictor models have low R^2 (< 0.1); i.e. these models are unlikely to suffer from over-fitting, to which splitting a data set into training and validation data sets is to run counter; and (c) The zero-predictor and one-predictor models are equivalent to estimating the mean and the trend line, respectively. The accuracy of the estimates is quantified by their standard errors, which decrease as the sample size increases. Splitting the original data set into training and validation data sets in effect reduces the sample size, leading to a deterioration in the accuracy of the estimated mean and trend line.

3.1. Star rating

For probability modelling, it is noted that the star ratings are distributed differently for dwellings rated < 6 than for those rated ≥ 6 stars, hence it makes sense to model stars in the domains of [0, 6) and [6, 10] with different distribution functions. In addition, each dwelling group has a unique proportion of dwellings corresponding to the two distributions. Because the NatHERS certificate database is the only available information on the dwellings' star ratings for the study area, we assume that the dwelling-group composition of the database is a representative sample of the building stock. Accordingly, the proportions estimated from the analysis data set are listed in Table 1, in which $p_{Đ1}$ and $p_{Đ2}$ denote the proportions of ratings in $[0,6)$ and $[6,10]$, respectively. Similarly, the proportions of N6S and R6S as determined from the analysis data set are 0.87 and 0.13, respectively.

Table 1 Proportions of dwellings rated < 6 ($p_{Đ1}$) and ≥ 6 ($p_{Đ2}$) stars

Dwelling group	$p_{Đ1}$	$p_{Đ2}$
B6S	0.97	0.03
N6S	0.01	0.99
R6S	0.70	0.30

Let $Đ_1$ and $Đ_2$ denote the distribution functions of star ratings corresponding to the domains [0, 6) and [6, 10], respectively. Assuming that they follow the beta distribution parameterised as follows:

$$f_{Đ_j,T}\left(s;\mu_{Đ_j,T},\phi_{Đ_j,T},\nu_{Đ_j,T}\right)=\begin{cases} p_{Đ_1}\dfrac{f_{\text{xbetax}}(s/s_\theta;\mu_{Đ_1,T},\phi_{Đ_1,T},\nu_{Đ_1,T})}{s_\theta}, & s<s_\theta; \\ p_{Đ_2}\dfrac{f_{\text{xbetax}}\left(\frac{s-s_\theta}{10-s_\theta};\mu_{Đ_2,T},\phi_{Đ_2,T},\nu_{Đ_2,T}\right)}{(10-s_\theta)}, & s\geq s_\theta. \end{cases} \quad (1)$$

where s is the star-rating value, $s_\theta = 6$ is the star-rating threshold delineating $Đ_1$ and $Đ_2$, $\mu_.$ and $\phi_.$ are the mean and precision parameters, respectively, $\nu_.$ is the mean of the exponentially-distributed exceedance parameter for the underlying beta distribution on $[-\nu_., 1+\nu_.]$ that is censored to $[0,1]$, the subscript $j = 1$ or 2, $T \in$ (B6S, N6S, R6S), and $f_{\text{xbetax}}(\cdot)$ is the continuous mixtures of extended-support beta distribution (Kosmidis & Zeileis, 2025). The advantage of the extended-support beta distribution is its ability to account for a finite probability of 0-star rating that could occur amongst B6S dwellings.

The Geelong City Council and OpenStreetMap data sets provide only two statistically significant predictors: year of construction, Y, and distance to the city centre, D. For beta regression, Y and D are scaled to $Y_t = Y/100$ and $D_t = D/1000$. Then the beta regression parameters $\mu_{Đ_j,T}$ and $\phi_{Đ_j,T}$ are assumed to be:

$$\begin{aligned} g\left(\mu_{Đ_j,T}\right) &= \begin{cases} \beta_0, & \text{if no predictor is significant;} \\ \beta_0 + \beta_1 Y_t, & \text{if } Y_t \text{ is significant;} \\ \beta_0 + \beta_1 D_t, & \text{if } D_t \text{ is significant.} \end{cases} \\ h\left(\phi_{Đ_j,T}\right) &= \gamma_0 \end{aligned} \quad (2)$$

where $g(\cdot): (0,1) \mapsto \mathbb{R}$ and $h(\cdot): (0,\infty) \mapsto \mathbb{R}$ are link functions mapping, respectively, $\mu_{Đ_j,T}$ and $\phi_{Đ_j,T}$ to the real line. In this study, we use the logit and log links for $g(\cdot)$ and $h(\cdot)$, respectively. The beta regression is carried out by the R package betareg (Kosmidis & Zeileis, 2025). Figure 2 compares the star-rating histograms of the three dwelling groups to the fitted probability distributions. Note that the B6S group has only 148 certificates, hence it has a more ragged histogram than the other two. This leads to higher modelling uncertainty as will be seen also in Figure 5.

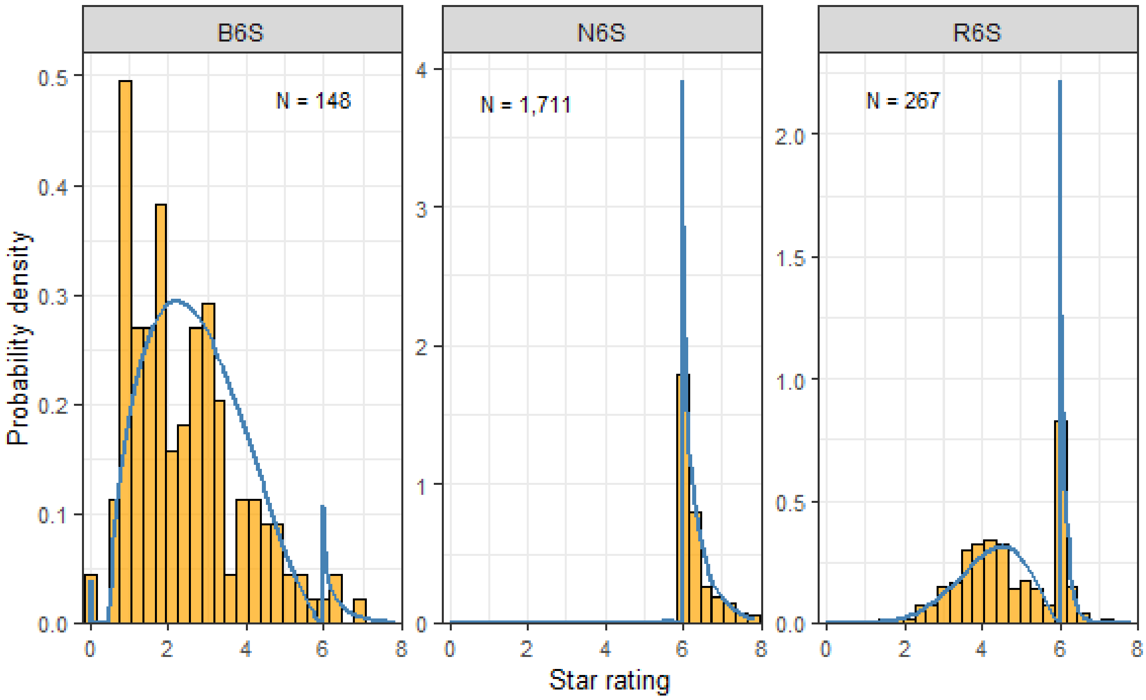


Figure 2 Histograms and fitted probability densities of star ratings of N6S and R6S dwellings

The estimated parameters are listed in Table 2, in which $\ln(\nu)$ is the natural logarithm of ν. The extended support of the distribution is assumed to be exponentially distributed (Kosmidis & Zeileis, 2025) on the left and right of $(0,1)$. Table 2 shows that Y_t is highly significant to the distribution Đ1 of B6S, and Y_t is highly significant to the distribution Đ1 of R6S. For B6S, since there are only four dwellings with a star rating ≥ 6 stars, we assume provisionally that its probability distribution Đ2 is the same as Đ2 of N6S.

Recalling that the star bands have a minimum of 0.5 star, and dwellings requiring a total thermal energy > 797 MJ/m^2 are given zero star, hence no star values $\in (0,0.5)$. This creates a difficulty for beta regression of star rating $\in \{[0,10]\backslash(0,0.5)\}$, where the set $A\backslash B$ means "A and not B." To overcome the discontinuity issue we shift 0 to a value close to but less than 0.5, say 0.49. That is, for beta regression of B6S < 6 stars, first shift star 0 to 0.49, then normalise the star-rating support $S \in [0.49,6)$ to $S_t \in [0,1)$ by

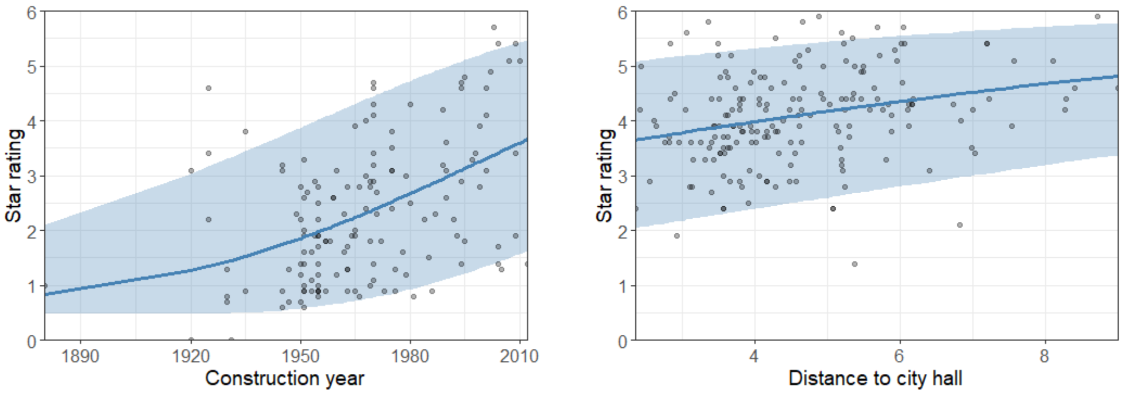
$$S_t = \frac{S - 0.49}{5.51} \quad (3)$$

To transform S_t back to S, take the inverse of Eq. (3), and shift $S = 0.49$ back to 0. For $S \in (0.49,0.5)$, round the value to $S = 0.5$.

Figure 3 plots the 90 % prediction intervals of star rating versus construction year for B6S rated < 6 stars as well as the star rating versus distance to city centre for R6S rated < 6 stars. In both cases, the star rating has positive trends with the respective predictors, implying that the more recent the construction year or the farther from the city centre, the higher the star rating in general.

Table 2 The beta regression coefficients of star rating for the three dwelling groups

Term	Estimate	Std error	Statistic	p-value
Dwelling group: B6S Distribution: Đ1				
β_0	-44.949	7.235	-6.213	0.000
Y_t	2.249	0.367	6.128	0.000
γ_0	1.496	0.120	12.463	0.000
$\ln(\nu)$	-4.974	0.938	-5.300	0.000
Dwelling group: N6S Distribution: Đ1				
β_0	1.627	0.215	7.565	0.000
γ_0	1.975	0.325	6.077	0.000
Dwelling group: N6S Distribution: Đ2				
β_0	-2.313	0.031	-75.509	0.000
γ_0	1.887	0.039	48.850	0.000
Dwelling group: R6S Distribution: Đ1				
β_0	0.105	0.178	0.590	0.555
D_t	0.146	0.038	3.888	0.000
γ_0	2.195	0.099	22.148	0.000
Dwelling group: R6S Distribution: Đ2				
β_0	-3.456	0.135	-25.579	0.000
γ_0	3.094	0.185	16.686	0.000



(a) B6S (star < 6) (b) R6S (star < 6)

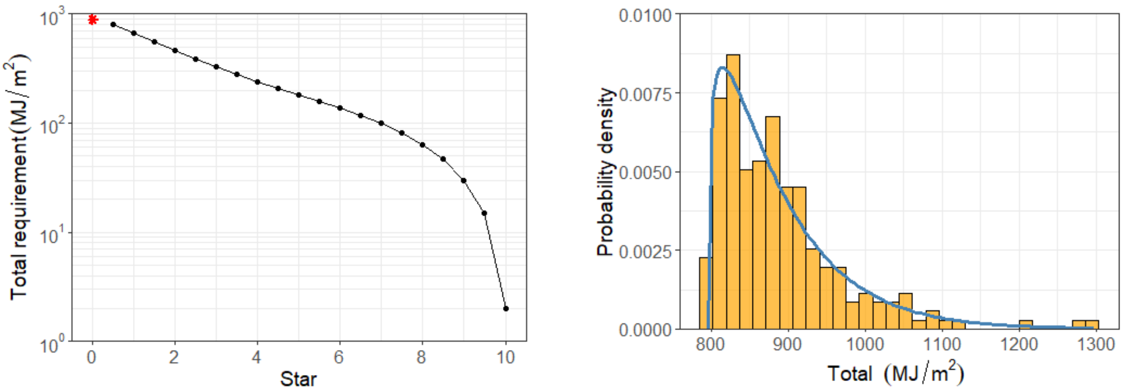
Figure 3 Prediction intervals of star rating versus predictors

3.2. Total thermal requirement

The NatHERS star bands map the total thermal requirement to star rating for the 69 climate zones (Commonwealth of Australia, 2019). Since the NatHERS stars $\in [0.5,10]$ are rounded to one decimal point, each star value actually corresponds to an interval of the requirement values. However, for a given climate zone, if the logarithm of total requirement is linearly interpolated by the star rating, any un-rounded star $\in [0.5,10]$ will correspond to a unique requirement value. With the aid of the interpolated relationship, the un-rounded star ratings estimated according to the preceding subsection can be used to estimate the thermal requirement. Therefore, we approximate the logarithmic requirement of climate zone 60 as a piecewise linear function of the NatHERS star, as shown by the black line in Figure 4(a).

The 0.5 star in climate zone 60 has a maximum requirement of 797 MJ/m^2, and anything larger than that is rated zero star. As a result, the thermal requirement of zero-star dwellings may be treated as a random variable without an upper bound. Since the analysis data set from postcode 3216 has only two dwellings rated zero star, we use the 206 zero-star dwellings from climate zone 60 available in the CSIRO certificate data set. The histogram of the zero-star total requirements is shown in Figure 4(b). It has an estimated mean and standard deviation of 889 MJ/m^2 and 82.7 MJ/m^2, respectively. The red asterisk in Figure 4(a) shows the estimated mean requirement.

After subtracting the zero-star total requirement values by 797 MJ/m^2, the data set is fitted by maximum likelihood estimation method to the gamma distribution (Ang & Tang, 2007), leading to an estimated shape parameter $a = 1.24$ and rate parameter $b = 0.0135$.



(a) NatHERS star band (b) Histogram of zero-star total requirement

Figure 4 NatHERS star band for climate zone 60 and histogram of zero-star total energy requirement

The star ratings of un-rated dwellings can be generated by the models developed in Section 3, and the total thermal requirement can be determined by the interpolated NatHERS star band or, for zero-star rating, generated by the gamma probability model. The histograms in Figure 5 show that the simulated requirements for N6S and R6S approximate the observed well, while the simulated requirements for B6S tend to be more conservative. This may be attributed to the fact that there were only 148 B6S rated dwellings for analysis, which is invariably plagued by significant epistemic uncertainty. The collection of more certificates in the CSIRO database will lead to a reduction of such uncertainty and improvement of the star-rating model.

4. CONCLUSIONS

To overcome the issue of data scarcity on residential energy-efficiency performance, we have developed statistical models that use to generate energy star rating and the associated thermal requirement for un-rated dwellings, using the building stock in postcode 3216, Geelong, Victoria, as a case study. Since the statistical modelling techniques are generic, they can be applied to other geographical areas. They can also be used for re-analysis when data sets with additional predictors become available; e.g. a potentially useful predictor is the status of housing tenure.

The analysis results show that, on the average, the more recent the construction year for B6S rated < 6 stars and the farther from the city centre for R6S rated < 6 stars, the higher the star rating. Our next step is to use the simulated star rating and thermal requirement to generate hourly whole-of-home energy consumption needed in the analysis of the impact of residential energy-efficiency intervention on electric grids.

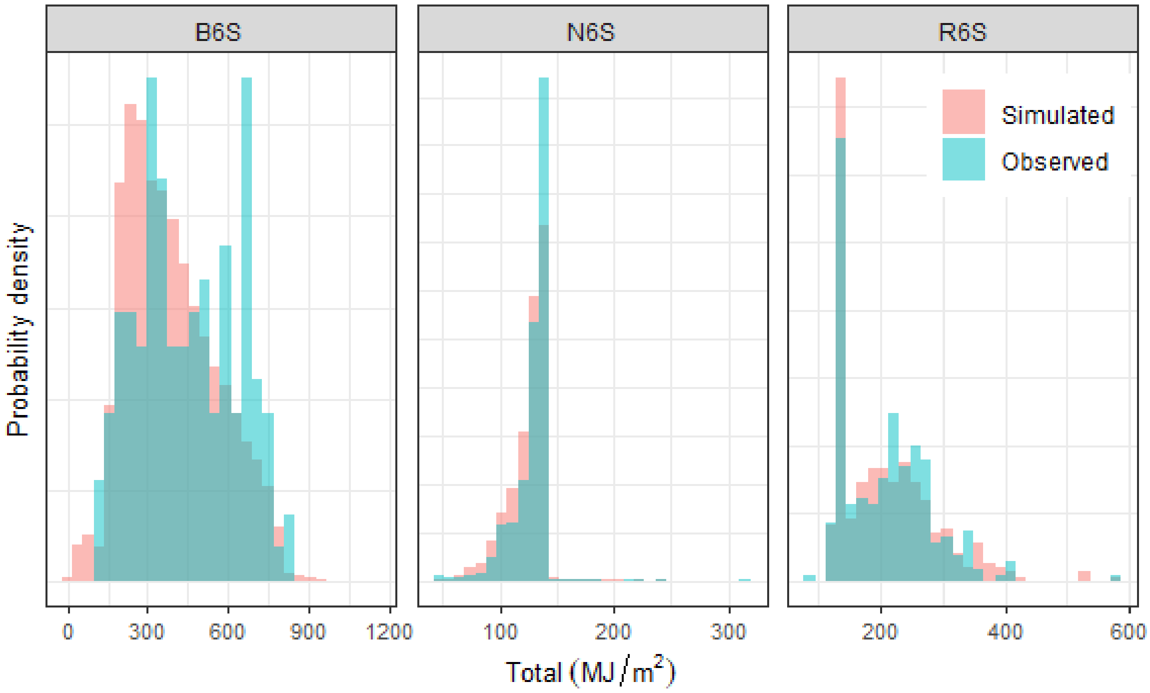


Figure 5 Histograms of simulated and observed total thermal requirements

REFERENCES

Ang, A. H.-S., & Tang, WH. (2007). Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Australian Bureau of Statistics. (2022). Snapshot of Australia: A picture of the economic, social and cultural make-up of australia on census night, 10 august 2021. Retrieved from https://www.abs.gov.au/statistics/people/people-and-communities/snapshot-australia/latest-release

City of Greater Geelong. (2024). Roofprints - City of Greater Geelong. Retrieved from https://www.geelongdataexchange.com.au/explore/dataset/roofprints-city-of-greater-geelong/information/

Commonwealth of Australia. (2019). Star band criteria. Retrieved from https://www.nathers.gov.au/sites/default/files/2019-10/NatHERS%20Star%20bands.pdf

Commonwealth of Australia. (2022a). Benchmark tool. Retrieved from https://www.nathers.gov.au/WholeofHome/software-development

Commonwealth of Australia. (2022b). NatHERS climate zones and weather files. Retrieved from https://www.nathers.gov.au/nathers-accredited-software/nathers-climate-zones-and-weather-files

CSIRO. (2025). Australian housing data. Retrieved from https://ahd.csiro.au/

Kosmidis, I., & Zeileis, A. (2025). Extended-support beta regression for [0, 1] responses. Journal of the Royal Statistical Society Series C: Applied Statistics, qlaf039. https://doi.org/10.1093/jrsssc/qlaf039

OpenStreetMap contributors. (2024). Planet dump retrieved from https://planet.osm.org . Retrieved from https://www.openstreetmap.org

Rojas, AM, Wang, C-H and Wright, A. (2025). Evidence-based energy futures: Conceptualisation of a national simulation platform for assessing residential construction code impacts. The 26th International Congress on Modelling and Simulation (MODSIM2025). Adelaide, South Australia.

van Buuren, S., & Groothuis-Oudshoorn, K. (2024). Mice: Multivariate imputation by chained equations. Retrieved from https://github.com/amices/mice

