POLITECNICO
MILANO 1863

Bayesian Learning and Montecarlo Simulation

Energy Efficiency

Moein Taherinezhad, 10935476
Trygve Myrland Tafjord, 11077296

July 18, 2025

Contents

1 Introduction

1.1 Background and motivation

With extreme temperatures becoming more frequent across the globe, improving the energy efficiency of buildings has never been more important. Whether the goal is to lower energy costs or to reduce environmental impact, optimizing heating and cooling systems can have significant benefits. In this project, we explore a dataset that offers valuable insights into how building design influences energy consumption, with a primary goal to model and predict the Heating Load for residential buildings using a Bayesian statistical framework.

1.2 Dataset description

The data are related to energy analysis using 12 different building shapes simulated in Ecotect. The dataset comprises 768 samples with 8 features, and two real valued responses. The features *Orientation* and *Glazing Area Distribution* are categorical values. Specifically:

X1 Relative Compactness	X6 Orientation
X2 Surface Area	X7 Glazing Area
X3 Wall Area	X8 Glazing Area Distribution
X4 Roof Area	y1 Heating Load
X5 Overall Height	y2 Cooling Load

2 Data Exploration and Preprocessing

An exploration of the dataset was conducted to understand its structure, identify key relationships, and prepare it for modeling. This section details the initial inspection and the subsequent preprocessing pipeline.

The dataset consists of 768 observations and 10 variables. An initial inspection confirmed that there were no missing (NA) or blank values requiring imputation or removal. The data types are primarily numeric, with the categorical variable `Orientation` holding four different values, and the categorical variable `Glazing_Area_Distribution` five different values. No outliers are detected.

2.1 Key Observations

An analysis of the relationships between variables revealed several key insights that directly informed our modeling strategy.

First, an examination of the correlation matrix in figure 1 highlighted significant correlation among the predictors. Notably, `Relative_Compactness` exhibit a very strong negative correlation of $r = -0.99$ with the `Surface_Area` feature, and a strong positive correlation of $r = -0.83$ with `Overall_Height`. This level of correlation among predictors necessitates a formal model selection procedure. It is worth noting that `Glazing_Area_Distribution`

and `Orientation` exhibits low correlation with the other features, confirming that these features are categorical.

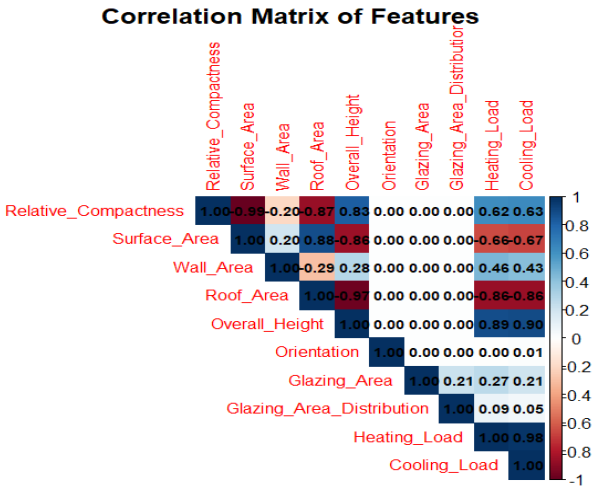


Figure 1: Correlation Matrix of Features

Second, the distribution of the response variable, `Heating_Load`, is distinctly non-normal. Figure 2 presents the kernel density estimate of the `Heating_Load` variable from the Energy Efficiency dataset. This plot provides a smooth approximation of the distribution, allowing for an intuitive understanding of its overall shape.

The distribution appears to be **bimodal**, with two prominent peaks. This suggests the existence of two distinct groups of buildings, those with lower heating demands, potentially due to efficient design or insulation, and those with higher demands.

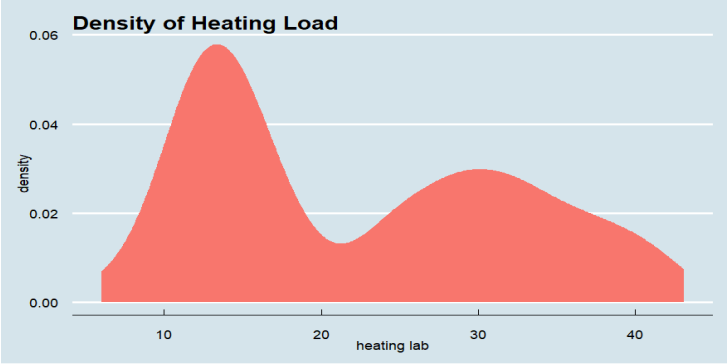


Figure 2: Density Of Heating Load

There is a dip between the two peaks, indicating a relative scarcity of observations with intermediate heating requirements. The distribution is clearly non-normal, which may have implications for downstream modeling tasks, particularly for methods that assume normality. In such cases, alternative approaches such as stratified modeling or the use of mixture distributions may be more appropriate.

Third, an analysis of the relationships between variables revealed several key insights that directly inform the modeling strategy. Scatter plots of geometric features against `Heating_Load` (Figure 6) show clear trends. Specifically, `Relative_Compactness` exhibits a strong positive correlation with `Heating_Load`, which aligns with the physical principle

that a lower surface-area-to-volume ratio reduces heat loss. Similarly, the data indicate that buildings with larger `Surface_Area` and `Roof_Area` values tend to be associated with lower heating loads.

However, the relationships are not purely linear. Multiple plots reveal distinct data clusters which appear to separate buildings into high- and low-efficiency groups. This clustering effect aligns with the bimodal distribution observed in the response variable and suggests that while a linear model is a valid starting point, it may not capture the full complexity of the data. Nonetheless, these clear trends validate the use of a linear regression framework as a primary analytical approach for this study.

2.2 Data Preprocessing Pipeline

To ensure the data is suitable for regression analysis and robust model evaluation, a four-step preprocessing pipeline is implemented:

1. **Logarithmic Transformation:** To address a positive skew observed in the exploratory analysis, a logarithmic transform is applied to the `Relative_Compactness` and `Wall_Area` features to produce more symmetric distributions.
2. **Categorical Variable Encoding:** The categorical predictors, `Orientation` and `Glazing_Area_Distribution`, are converted into dummy variables using one-hot encoding so they could be correctly interpreted by the linear model.
3. **Data Partitioning:** The dataset is partitioned into a training set (80% of observations) and a test set (20%). This separation is crucial for out-of-sample validation, allowing for an unbiased assessment of the model's ability to generalize to new data.
4. **Feature Scaling:** Exploratory analysis revealed that the numeric predictors operated on vastly different scales. To prevent features with larger magnitudes from dominating the model, the numeric predictors in the training set were standardized to have a mean of zero and a standard deviation of one.

3 Model specification and posterior analysis

This section specifies the Bayesian linear regression models used to predict building energy efficiency. Given the 14 potential predictors, we face significant model uncertainty, which we address by adopting a Bayesian framework that formally incorporates this uncertainty into the analysis. To ensure our conclusions are robust, we explore a range of prior distributions, including a non-informative reference prior (approximated by BIC) and several variants of Zellner's g-prior. We also employ MCMC methods to navigate the large model space, allowing for robust inference that does not depend on a single "best" model.

3.1 Initial Model Selection

In a regression context with p potential predictors, there are 2^p possible models. Given the 14 predictors in our preprocessed dataset (after encoding categorical variables), we face a space of $2^{14} = 16,384$ competing models. The primary challenge is to identify a good model that explains the variability in Heating Load without overfitting, especially

given the multicollinearity among geometric predictors identified in the exploratory data analysis.

To address this, we adopt an approach using the `BAS` package. Initially a regression linear model is created. For the regression coefficients (β) within each model, we use the common objective choice of Zellner's g-prior, calibrated with the Unit Information Prior ($g = n$), which ensures the prior contains about the same amount of information as a single observation. For the model space, we assign a uniform prior, making all 16,384 models equally likely *a priori*. Based on this regression model, we use BAS to get the models with the highest posterior probabilities.

This analysis revealed significant model uncertainty. The model with the highest posterior probability (HPM) achieved a probability of only 35%, with the next two most likely models attaining probabilities of 19% and 17%. The absence of a single dominant model confirms that several predictors are close substitutes due to high correlation, making it risky to rely on any single model for inference.

This uncertainty motivates a direct comparison of predictive strategies. We performed an out-of-sample validation on the held-out test set to compare the predictive performance of three estimators: the Bayesian Model Averaging (BMA) estimate, the Highest Posterior Model (HPM) estimate, and a "Full Model" that includes all 14 predictors. The results are summarized in Table 1.

Table 1: Out-of-Sample Predictive Performance Comparison

Model	**MSE**	**Num. Outside 95% Interval**	**Percent Outside**
Highest Posterior Model (HPM)	9.08	16	10.46%
Bayesian Model Averaging (BMA)	9.10	16	10.46%
Full Model	9.15	17	11.11%

Both the Highest Posterior Model (HPM) and Bayesian Model Averaging (BMA) estimators improved on the Full Model's predictive accuracy, achieving lower Mean Squared Error (MSE) and better prediction interval coverage by removing non-informative predictors.

The performances of HPM and BMA were nearly identical, though HPM's MSE was marginally lower on this specific test set. However, given the significant model uncertainty, BMA is generally the more robust choice because it formally incorporates this uncertainty by averaging over the model space instead of relying on a single model.

3.2 Bayesian Multiple Linear Regression

In this section, we present the Bayesian formulation of the multiple linear regression model applied to the Energy Efficiency dataset. Our goal is to incorporate prior beliefs about the model parameters and derive the posterior distribution using Bayes' theorem. We employ a reference prior to perform a default or baseline analysis. This approach serves as a bridge between Bayesian and frequentist methodologies and facilitates direct comparison of their results.

The Bayesian model begins with the same linear formulation as in the classical frequentist

approach. For each observation i, the model is expressed as:

$$y_i = \beta_0 + \beta_1 x_{\text{RC},i} + \beta_2 x_{\text{SA},i} + \beta_3 x_{\text{WA},i} + \beta_4 x_{\text{RA},i} + \beta_5 x_{\text{OH},i} + \beta_6 x_{\text{OR},i} + \beta_7 x_{\text{GA},i} + \beta_8 x_{\text{GAD},i} + \varepsilon_i$$

where:

- y_i denotes either the *Heating Load* or *Cooling Load* for the i-th building,
- $x_{\text{RC},i}$ is the Relative Compactness,
- $x_{\text{SA},i}$ is the Surface Area,
- $x_{\text{WA},i}$ is the Wall Area,
- $x_{\text{RA},i}$ is the Roof Area,
- $x_{\text{OH},i}$ is the Overall Height,
- $x_{\text{OR},i}$ is the Orientation,
- $x_{\text{GA},i}$ is the Glazing Area,
- $x_{\text{GAD},i}$ is the Glazing Area Distribution,
- ε_i is the random error term.

We assume the error terms are normally distributed: $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. In the Bayesian setting, priors are specified for the regression coefficients and variance term, and inference is carried out through the posterior distribution.

3.2.1 Bayesian Information Criterion

For Bayesian inference, we must define prior distributions for all model parameters. Since each energy efficiency response value y_i (either *Heating Load* or *Cooling Load*) is continuous, we assume that the error terms ε_i are independent and identically distributed as a Normal distribution:

$$\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

We also assign priors to the regression coefficients $\beta_0, \beta_1, \ldots, \beta_8$, where each β_j corresponds to one of the eight building features (e.g., Relative Compactness, Surface Area, etc.). Under the reference prior framework for multiple linear regression, we assume:

$$\pi(\beta_0, \beta_1, \ldots, \beta_8 \mid \sigma^2) \propto 1, \quad \pi(\sigma^2) \propto \frac{1}{\sigma^2}$$

This prior is non-informative and represents a flat distribution over the coefficients, while assigning a scale-invariant prior to the error variance.

From the marginal posterior summaries, we observe that the probability of each coefficient being non-zero is effectively 1. This is expected, as the model is forced to include all predictors without exclusion. Notably, under the Bayesian centered model, the posterior mean of the intercept β_0 equals the sample mean of the response variable. For instance, for the *Heating Load* response, the posterior mean of β_0 is 22.30, which differs from the OLS-estimated intercept due to centering.

Figure 3 presents the posterior distributions of the coefficients β_1 to β_8.

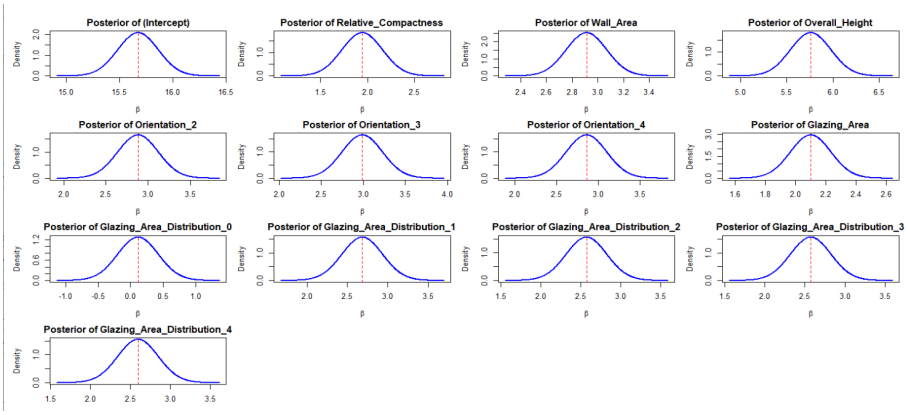


Figure 3: Posterior distributions of the regression coefficients under reference prior

We further report posterior means, standard deviations, and 95% credible intervals for all coefficients. These summaries provide valuable interpretability. For example, we find with 95% certainty that *Heating Load* increases as *Glazing Area* rises from 0.1 to 0.4. On the other hand, predictors like *Orientation* and *Glazing Area Distribution* exhibit credible intervals that include zero, suggesting they may be excluded in a simplified model without substantial loss of explanatory power.

Each subplot shows the marginal posterior density of a regression coefficient, derived using Bayesian inference with a reference prior. These densities summarize our updated beliefs about the coefficient values after observing the data.

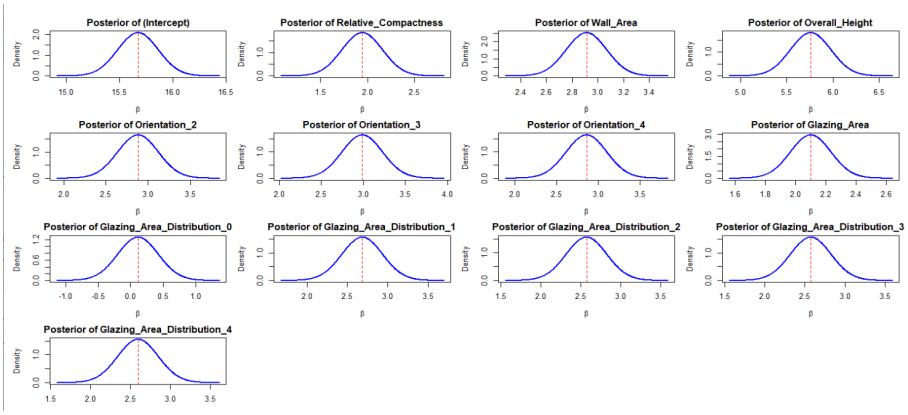


Figure 4: Posterior densities of regression coefficients under the BIC prior. Red dashed lines indicate posterior means.

Figure 4 presents the posterior densities of the regression coefficients obtained using Bayesian Model Averaging (BMA) under the Bayesian Information Criterion (BIC) prior. Each subplot corresponds to one model coefficient, with the horizontal axis representing the coefficient value and the vertical axis representing its posterior density. The red dashed vertical line denotes the posterior mean of each coefficient.

The posterior distributions for continuous predictors such as `Relative_Compactness`, `Wall_Area`, `Overall_Height`, and `Glazing_Area` exhibit sharp, symmetric bell-shaped curves, indicating low uncertainty and strong influence on the target variable. In contrast, some of the categorical predictors (e.g., `Orientation_j`, `Glazing_Area_Distribution_j`)

display wider or flatter posterior distributions, suggesting greater uncertainty or weaker contributions.

Overall, the BIC prior encourages model sparsity by penalizing model complexity, leading to the shrinkage of less informative coefficients toward zero. This visualization supports the identification of influential predictors in the Bayesian linear regression framework and aligns with the model selection principle imposed by the BIC criterion.

The posterior analysis shows that key architectural features such as `Relative_Compactness`, `Glazing_Area`, and `Overall_Height` play a significant role in determining the energy efficiency of buildings. In contrast, features like `Orientation` exhibit high uncertainty and may be excluded in simplified models. Bayesian modeling thus not only offers predictive capability but also allows for a probabilistic understanding of the importance and uncertainty of each predictor. We discuss Bayesian model comparison using the Bayesian Information Criterion (BIC), which approximates the log marginal likelihood in large-sample settings. Under BIC, the prior on the coefficient vector $(\beta_0, \ldots, \beta_8)^T$ is considered uniformly flat, equivalent to using the reference prior conditional on σ^2.

3.2.2 Zellner's g-prior

Zellner's g-prior has been widely used in Bayesian model selection and Bayesian model averaging. Now the question is, how do we pick g? The Bayes factor depends on g. There are some solutions which appear to lead to reasonable results in small and large samples, based on empirical results with real data as well as theory. In the following examples, we let the prior distribution of g depend on n, the size of the data.

There are several prior formulations that have been shown, both theoretically and empirically, to lead to reasonable results for Bayesian linear regression, especially in small and large samples. In the following examples, we explore prior distributions for the shrinkage factor g, where the prior on g is allowed to depend on n, the number of observations in the dataset.

Unit Information Prior: One commonly used choice is the *unit information prior*, where we set $g = n$. This implies $n/g = 1$, meaning the prior information is equivalent to one observation. Under this prior, we need only specify the prior mean β_0 for the intercept term, and assign priors to the remaining predictor coefficients $\beta = (\beta_1, \beta_2, \ldots, \beta_8)^T$.

Zellner–Siow Cauchy Prior: However, taking $g = n$ ignores uncertainty in the selection of g. Since g is not known a priori, an alternative is to model n/g as a random variable, creating a hierarchical prior. One such approach is the *Zellner–Siow Cauchy prior*, where we assume:

$$\frac{n}{g} \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$$

This introduces heavier tails and allows for greater shrinkage flexibility, especially in small-sample scenarios.

Hyper-g/n Prior: Another flexible alternative is the *hyper-g/n prior*, where the transformation is applied as:

$$\frac{1}{1 + (n/g)} \sim \text{Beta}\left(\frac{a}{2}, \frac{b}{2}\right)$$

Here, a and b are hyperparameters controlling the degree and uncertainty of shrinkage, or regularization. The Bayes factor under this prior can be analytically expressed in terms of hypergeometric functions, making it particularly tractable for model comparison.

Model Selection Implications: These priors are often used in Bayesian model selection procedures to compute posterior inclusion probabilities (PIPs) for each coefficient. Figure 5 displays the PIP values of all regression coefficients β_1 through β_8 under the different g-prior formulations discussed.

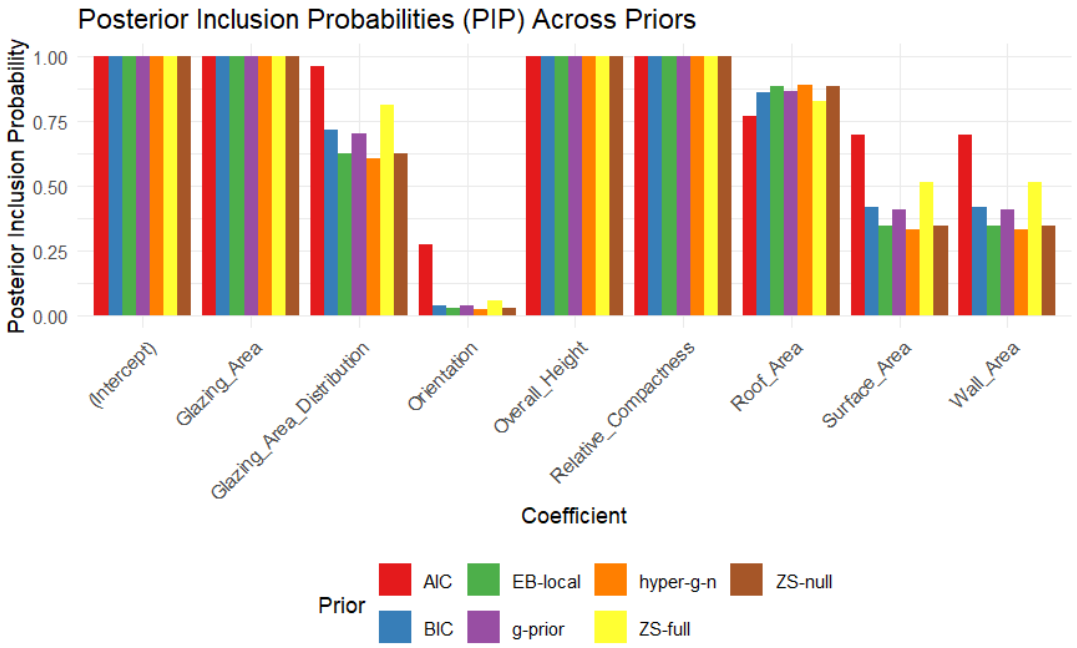


Figure 5: Posterior inclusion probabilities (PIPs) for coefficients β_1 through β_8 under different g-prior formulations

In Figure 5, each group of bars corresponds to a specific regression coefficient, while the colors represent different prior distributions used in Bayesian model averaging. The height of each bar reflects the posterior inclusion probability (PIP) of the associated predictor variable under a given prior.

From the results, we observe that the features `Glazing_Area`, `Overall_Height`, and `Relative_Compactness` exhibit PIPs close to 1 across all priors. This indicates strong agreement across all prior choices that these variables are highly relevant and should be included in the model. This aligns well with previous model selection findings.

`Roof_Area` also maintains a high PIP (above 0.75), while `Surface_Area` and `Wall_Area` show more variation across priors. Specifically, under the `BIC`, `ZS-null`, and `hyper-g/n` priors, their PIPs drop below 0.5, suggesting these methods are more conservative regarding their inclusion.

`Glazing_Area_Distribution` presents moderate inclusion probability under most priors but shows considerable variation. Meanwhile, `Orientation` is consistently given very low inclusion probability by all priors, indicating weak evidence for its relevance in the model.

Comparing across priors, we see that the `BIC` prior is the most conservative, tending to

include fewer predictors, whereas `AIC` and `ZS-full` are more liberal in variable inclusion. The `EB-local` and `hyper-g/n` priors fall somewhere in between, reflecting adaptive behavior based on model fit and uncertainty.

3.2.3 Bayesian model with markov chain monte carlo (MCMC) sampler

To further explore the model space in a Bayesian framework, we applied the `bas.lm` function from the `BAS` package using `method = "MCMC"`. For the prior distribution on the regression coefficients, we used the Zellner–Siow Cauchy prior, which introduces heavier tails and offers greater robustness in variable selection.

We configured the MCMC sampler to continue until either the number of unique models visited exceeded 2^p, or the number of iterations surpassed 2×2^p, whichever was smaller. Here, $p = 8$ corresponds to the number of predictor variables in our model.

To assess whether the MCMC algorithm sufficiently explored the model space, we used the `diagnostics()` function to generate convergence plots. The left panel in Figure 7 shows a comparison of posterior inclusion probabilities (PIP) estimated using MCMC versus uniform sampling. The right panel compares the posterior model probabilities under both sampling methods.

Ideally, points should fall on the 45-degree reference line in both plots. While most do, a few notable deviations are observed. This suggests that although convergence is reasonable, additional MCMC iterations could improve the stability and accuracy of the posterior estimates.

The Bayesian Model Averaging (BMA) approach was applied to predict the heating load based on the selected building features. The model achieved a Mean Squared Error of 9.09 on the test dataset. These metrics indicate a reasonable predictive accuracy,

Furthermore, the spread of the residuals seems relatively constant across the range of fitted values, implying that the assumption of homoscedasticity (constant variance) holds. However, we also observe a few labeled points, specifically observations 558, 291, 294, and 287, that deviate substantially from the main cloud of residuals. These may be considered potential outliers and warrant further investigation.

3.3 Regression Results

Table 2: Comparison of Bayesian Regression Models

Model Description	Hyperparameters	Test MSE	Points Outside 95% Interval	% Outside
Conjugate Model	$\beta \mid \sigma^2 \sim \mathcal{N}(\mathbf{0}, \sigma^2(\mathbf{I} \cdot 0.1)^{-1})$, $\sigma^2 \sim \mathcal{IG}(1,1)$	9.403	17	11.11%
BAS (Unit Info g-prior)	prior="g-prior", $g = n$, modelprior=uniform()	9.078	16	10.46%
BAS (Zellner-Siow Prior)	prior="JZS", modelprior=uniform()	9.095	17	11.11%
BAS (Hyper-g/n Prior)	prior="JZS", $a = 3$, modelprior=uniform()	9.095	17	11.11%
BAS (Empirical Bayes Prior)	prior="JZS", modelprior=uniform()	9.095	17	11.11%
BAS (BIC Prior)	prior="BIC", modelprior=uniform()	9.095	16	10.46%
BAS (AIC Prior)	prior="AIC", modelprior=uniform()	9.095	16	10.46%
BAS (MCMC Sampler)	prior="ZS-null", modelprior=uniform(), method="MCMC"	9.086	16	10.46%

The results from the various Bayesian models are remarkably stable, particularly those implemented with the BAS package. The BAS models consistently produced Test MSE values around 9.08-9.10. The percentage of test points falling outside the 95 percentage predictive interval was also highly consistent, ranging from 10.46

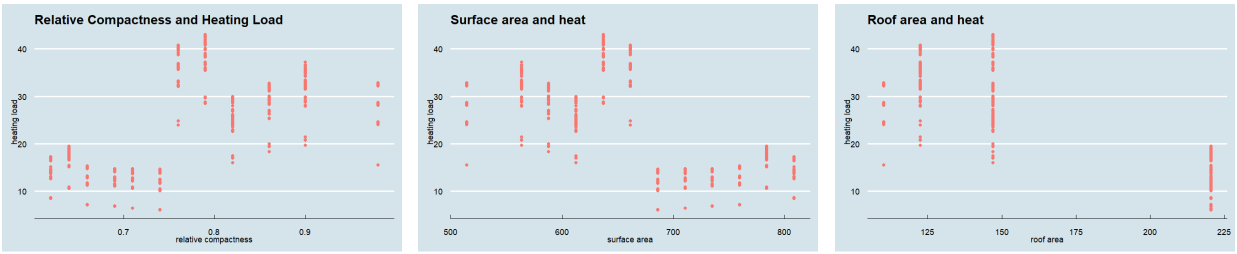
The Conjugate Model performed slightly worse, with an MSE of 9.40. This is likely because two predictors, `Surface_Area` and `Roof_Area`, were manually removed to manage computational complexity. Notably, several BAS models (including Zellner-Siow, BIC, and AIC) yielded the exact same MSE of 9.095, indicating a strong consensus on the Highest Posterior Model (HPM) for prediction.

This collection of models effectively serves as a sensitivity analysis. The stability of the performance metrics demonstrates that the model's conclusions are robust and not overly sensitive to the choice of prior variance for the coefficients. The MCMC-based sampler further confirmed these results and showed proper convergence. The observation that over 10 percent of points fell outside the 95 percent predictive interval suggests the model's predictive uncertainty might be slightly underestimated.

A Appendix

List of Figures

A.1 Additional Figures



(a) Heat Load/Relative Compactness (b) Heat Load/Surface Area (c) Heat Load/Roof Area

Figure 6: Scatter plots

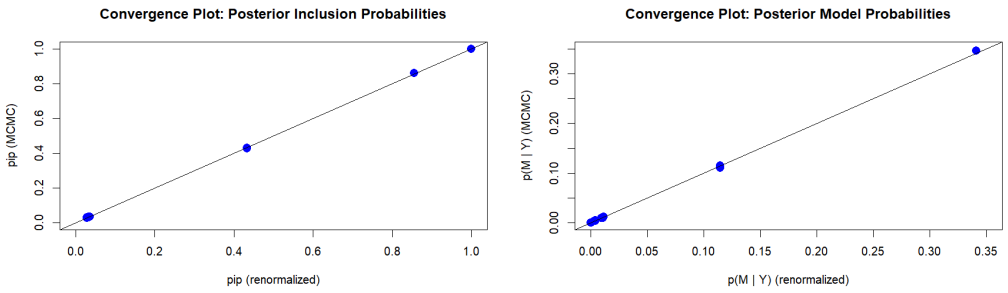


Figure 7: Comparison of posterior distributions under two different prior choices for Heating_Load

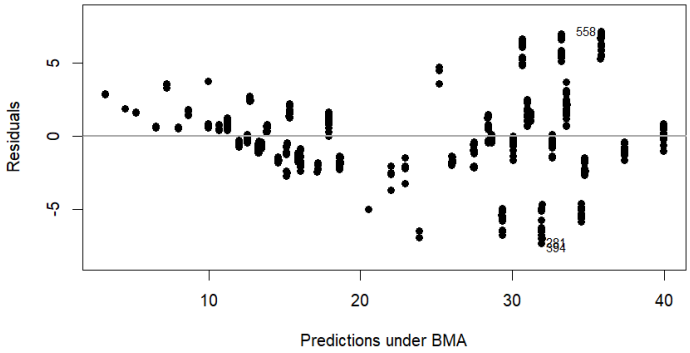


Figure 8: Residuals versus fitted values using BMA