Hierarchical Analysis of Power Consumption Data

Rose Mary Chittilappilly - Sayalee Joshi
chittilr@uci.edu - sayalesj@uci.edu

1 Introduction

The efficient management of power consumption in urban areas is crucial for sustainability and resource allocation. This study employs a hierarchical modeling approach to analyze power consumption patterns in Tetouan City, Morocco, a region characterized by diverse environmental conditions across different zones. Understanding the factors influencing power consumption, such as temperature, humidity, and zone-specific effects, is essential for effective urban planning and energy management strategies.

2 Problem statement

The objective of this study is to investigate how various environmental factors and spatial variations across different zones influence power consumption in Tetouan City. Our dataset has three zones, and hierarchical models help in checking variability across these zones. These models improve generalization and provide more robust inferences, offering flexibility and better interpretability. By leveraging a hierarchical modeling framework, we aim to quantify the impact of temperature, humidity, and zone-specific effects on power consumption patterns observed over a decade.

3 Dataset

The dataset used in this analysis is sourced from the UCI Machine Learning Repository, titled "Power Consumption of Tetouan City". It comprises detailed records of power consumption readings from January 2001 to December 2012, recorded at 10-minute intervals across three distinct zones within Tetouan City. Additionally, the dataset includes measurements of temperature, humidity, wind speed, and other relevant variables. Prior to analysis, missing values in the dataset were handled to ensure data integrity and reliability.

4 Exploratory Data Analysis

Our analysis begins with an exploration of power consumption data. The histogram (1) provided initial insights into the relationships between variables such as temperature, humidity, and power consumption, highlighting potential outliers and patterns. The correlation analysis (2), revealed significant correlations among these variables and guided the selection of predictors for subsequent modeling. Time series plots (3) were then employed to examine how power consumption varies over time across different zones within the city, elucidating spatial variations and temporal trends that are crucial for understanding energy demand dynamics in urban settings.

5 Details of the Model

To capture the hierarchical structure of the data, a Bayesian hierarchical model was employed. The model includes fixed effects such as temperature and humidity, which directly influence power consumption across all zones. Additionally, random effects were incorporated to account for zone-specific variations in power consumption patterns, ensuring the model captures spatial dependencies and heterogeneity across the city.

The hierarchical Bayesian model used to analyze power consumption is formulated as:

$$\text{Power}_{ij} = \beta_0 + \beta_1 \cdot \text{Temperature}_{ij} + \beta_2 \cdot \text{Humidity}_{ij} + u_{\text{Zone}_j} + \epsilon_{ij}$$

where:
- Power_{ij}: Power consumption for observation i in zone j.
- β_0: Intercept.
- β_1: Coefficient for temperature.
- Temperature_{ij}: Temperature for observation i in zone j.
- β_2: Coefficient for humidity.
- Humidity_{ij}: Humidity for observation i in zone j.
- u_{Zone_j}: Random effect for zone j, $u_{\text{Zone}_j} \sim N(0, \sigma^2_{\text{Zone}})$
- ϵ_{ij}: Error term.
- σ^2_{Zone} is the variance of the zone-specific random effect.

In our analysis of power consumption patterns in Tetouan City, Morocco, we initially began with a simple regression model to identify the most influential covariates explaining power consumption variability. This initial stage focused on understanding the effects of temperature, humidity, and other environmental factors on energy usage across the city.
Upon establishing the significance of these factors, we progressed to a more sophisticated hierarchical modeling approach. This hierarchical model introduced random effects to account for variations in power consumption specific to different zones within Tetouan City. By incorporating this hierarchical structure, we aimed to capture and quantify the spatial dependencies and heterogeneities that exist across the city, thereby refining our understanding of how local environmental conditions impact energy demand.
This sequential approach—from a basic regression model to a hierarchical framework—allowed us to systematically enhance the accuracy and interpretability of our analysis. It enabled a comprehensive exploration of both global and localized factors influencing power consumption, providing insights essential for effective urban energy management strategies.

6 Diagnostics

- Rhat: The potential scale reduction factor (Rhat) was used to assess the convergence of the Markov Chain Monte Carlo (MCMC) chains. Rhat values close to 1 indicate satisfactory convergence, ensuring reliable estimation of model parameters and credible intervals. In our analysis, Rhat values were consistently 1 across all parameters, confirming robust convergence.

- ESS Numbers: Effective Sample Size (ESS) measures were computed to evaluate the precision of parameter estimates. High ESS values for both Bulk ESS and Tail ESS indicated sufficient sampling efficiency, confirming robust inference from the Bayesian hierarchical model.

- Credible Intervals: We obtained 95% credible intervals for the coefficient estimates for this model. Credible intervals provide a range of values for model parameters, reflecting the uncertainty inherent in Bayesian estimation. (Table 1, Figure 4)

Parameter	**Estimate**	**Est. Error**	**l-95% CI**	**u-95% CI**	**Rhat**	**Bulk ESS**	**Tail ESS**
sd(Intercept)	24942.90	11361.91	12091.88	55679.65	1.00	1237	2051
Intercept	-1382.74	517.25	-2404.18	-389.71	1.00	5373	4611
Temperature	286.60	3.57	279.60	293.53	1.00	4668	4571
Humidity	-35.83	1.57	-38.88	-32.79	1.00	4634	4677
sigma	5682.65	16.10	5651.27	5714.16	1.00	6425	4314

Table 1: Summary of model parameters

- Plots: We assessed mixing and convergence using posterior probability checks (Figure 7) and trace plots over 4 chains (Figure 6). The posterior probability check visually compares the simulated data generated by the model to the observed data, ensuring the model accurately captures the true data distribution and identifying any systematic biases. In trace plots, we monitor the evolution of parameters across chains, aiming for well-mixed and overlaid traces that indicate convergence. These diagnostics validate our Bayesian model's ability to analyze power consumption patterns across different zones, focusing on chain mixing and parameter estimation. Importantly, we observe effective mixing of chains, indicating thorough exploration and sampling of the parameter space. Additionally, we confirm convergence and present the posterior probability check, demonstrating satisfactory agreement between simulated and empirical data distributions.

7 Results

Model Summary:
The model incorporates the effects of Temperature and Humidity while accounting for potential zone-specific variations in consumption patterns through random intercepts for each zone.
The intercept β_0 represents the expected power consumption when all predictors (Temperature and Humidity) are zero. In this case, it has a negative value indicating that, at zero

temperature and humidity, the base power consumption is predicted to be -1382 plus the specific zone intercept units. However, this scenario is not be realistic since humidity cannot actually be zero in real-world situations.
The temperature coefficient β_1 indicates that for each unit increase in temperature, the power consumption is expected to increase by 286.60 units, holding all other variables constant. The small standard error and narrow 95% credible interval suggest that this effect is estimated very precisely.
The humidity coefficient β_2 indicates that for each unit increase in humidity, the power consumption is expected to decrease by 35.83 units, holding all other variables constant. Similar to Temperature, the effect is estimated precisely with a small standard error and narrow 95% credible interval.
The standard deviation of zone σ^2_{Zone} value represents the standard deviation of the random intercepts for the different zones. A large value indicates that there is substantial variability in the baseline power consumption across different zones.
Overall City-wide Effects:
The estimated coefficients for fixed effects indicated significant associations between temperature and power consumption. There is a positive relationship between power consumption and Temperature. This means that higher temperatures lead to increased power consumption in Tetouan City.
There is a negative relationship between power consumption and Humidity. Which means, higher humidity is associated with lower power consumption. Humidity also showed a statistically significant but smaller effect on power consumption, with higher humidity levels associated with slight reductions in energy usage.
Zone-Specific Effects:
The model captures zone-specific variations in power consumption through random intercepts. These random intercepts are added to the fixed intercept to account for zone-specific deviations in power consumption. Zones with higher estimates deviate more from the city-wide average consumption. From the posterior summary, we can see that Zone 1 deviates the most from the city-wide average and Zone three deviates the least. (Figure 5)

8 Conclusion

In conclusion, our study utilized hierarchical modeling to analyze power consumption patterns across different zones in Tetouan City, Morocco, over a significant time span. Through the exploratory data analysis, we identified key environmental factors—such as temperature and humidity—that significantly influence energy demand variations across zones. Our Bayesian hierarchical model provided robust estimates of these relationships between environmental conditions and energy consumption.
The model's diagnostics, including posterior probability checks, trace plots, and credible intervals, confirmed its reliability and effectiveness in capturing the inherent uncertainties in energy demand modeling. We observed satisfactory convergence of MCMC chains, ensuring stable parameter estimates and validating the model's adequacy in explaining observed data.
The analysis highlights how environmental conditions and spatial differences impact power consumption in urban areas. Using hierarchical modeling, we uncover how temperature, hu-

midity, and geographic location interact to influence energy usage patterns. These findings are crucial for devising focused strategies to enhance energy efficiency and sustainability efforts in Tetouan City.

References

[1] A. Salam, A. E. Hibaoui, Comparison of Machine Learning Algorithms for the Power Consumption Prediction : - Case Study of Tetouan city , International Renewable and Sustainable Energy Conference, 2018.

[2] Nick Jenkins, Bayesian Data Analysis with brms, 2021.

[3] Paul Buerkner, Bayesian regression models using Stan.

A Figures

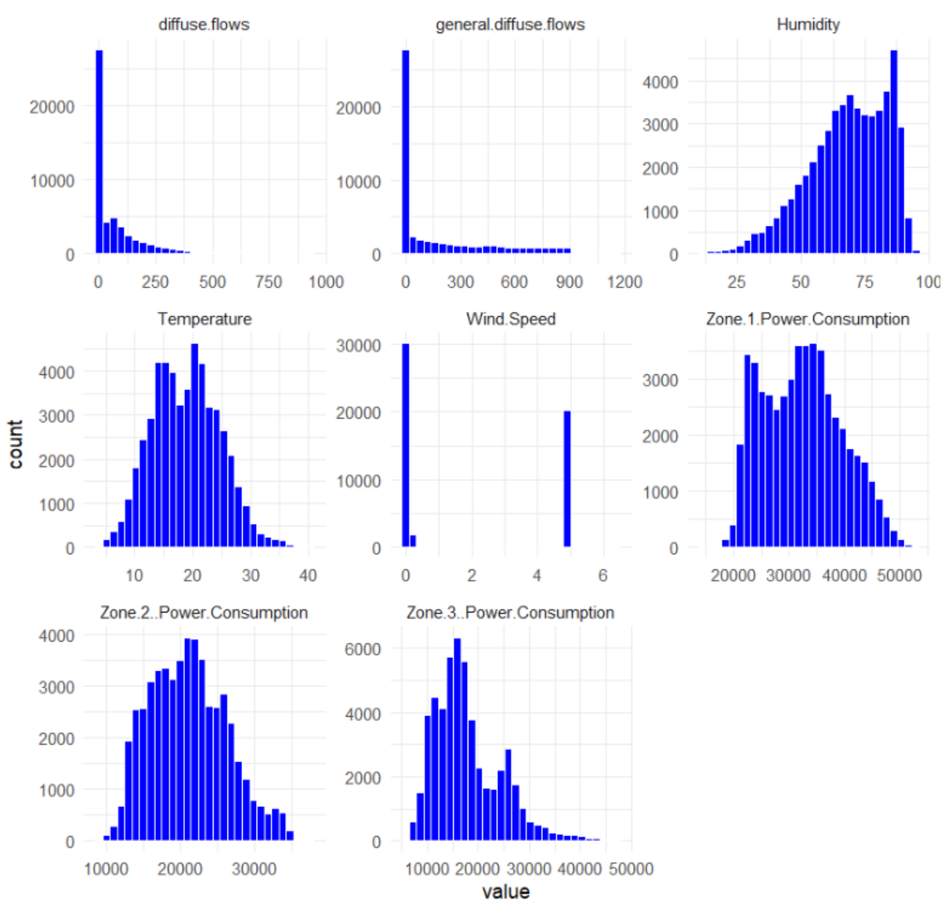

Figure 1: Histogram

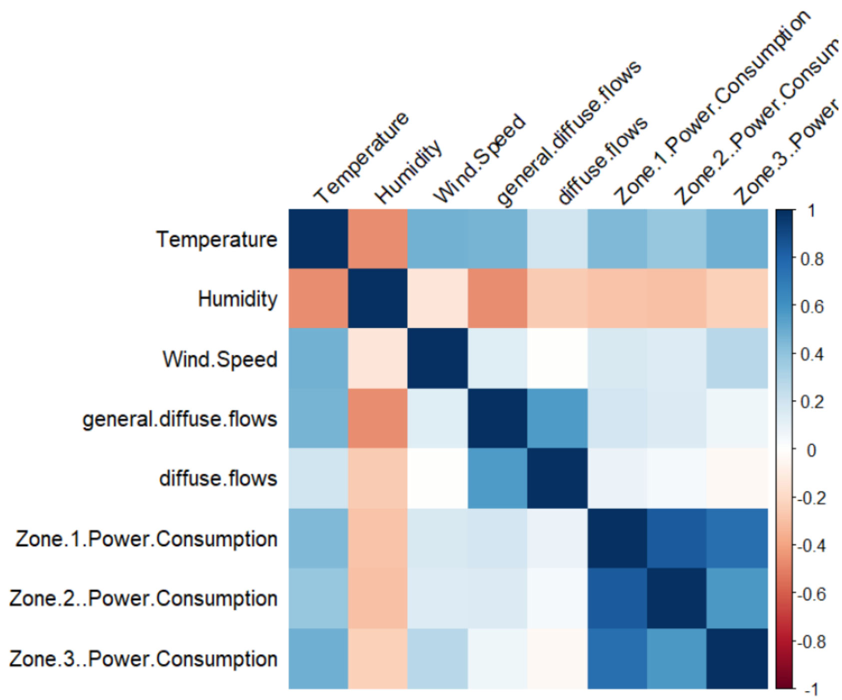

Figure 2: Correlation Heatmap

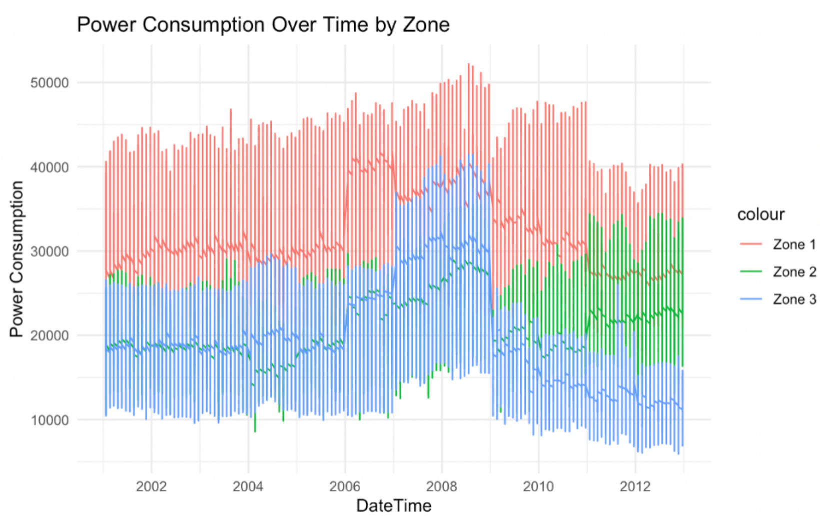

Figure 3: Power Consumption over time across different zones

```
Family: gaussian
  Links: mu = identity; sigma = identity
Formula: PowerConsumption ~ Temperature + Humidity + (1 | Zone)
   Data: long_time_power_data (Number of observations: 62208)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
         total post-warmup draws = 8000

Multilevel Hyperparameters:
~Zone (Number of levels: 3)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept) 24942.90  11361.91 12091.88 55679.65 1.00     1237     2051

Regression Coefficients:
            Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept   -1382.74    517.25 -2404.18  -389.71 1.00     5373     4611
Temperature   286.60      3.57   279.60   293.53 1.00     4668     4571
Humidity      -35.83      1.57   -38.88   -32.79 1.00     4634     4677

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma  5682.65     16.10  5651.27  5714.16 1.00     6425     4314

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Figure 4: Output

```
                                          Estimate    Est.Error          Q2.5         Q97.5
b_Intercept                            -1382.74226   517.247665   -2404.17701    -389.70909
b_Temperature                            286.59523     3.568820     279.60267     293.52933
b_Humidity                               -35.83459     1.570229     -38.87798     -32.79484
sd_Zone__Intercept                     24942.89605 11361.909155   12091.87617   55679.65009
sigma                                   5682.64737    16.095149    5651.27212    5714.15519
Intercept                               1536.15919   498.498130     548.85760    2485.67812
r_Zone[1.Power.Consumption,Intercept]  30550.32348   499.922681   29588.69825   31532.85128
r_Zone[2..Power.Consumption,Intercept] 19211.24081   500.567104   18243.33381   20207.72277
r_Zone[3..Power.Consumption,Intercept] 16243.80205   499.048091   15284.59633   17231.18690
lprior                                 -1076.88095    26.134536   -1128.29583   -1026.54831
lp__                                 -627130.01062     2.137616 -627135.14003 -627126.87872
```

Figure 5: Posterior Summary

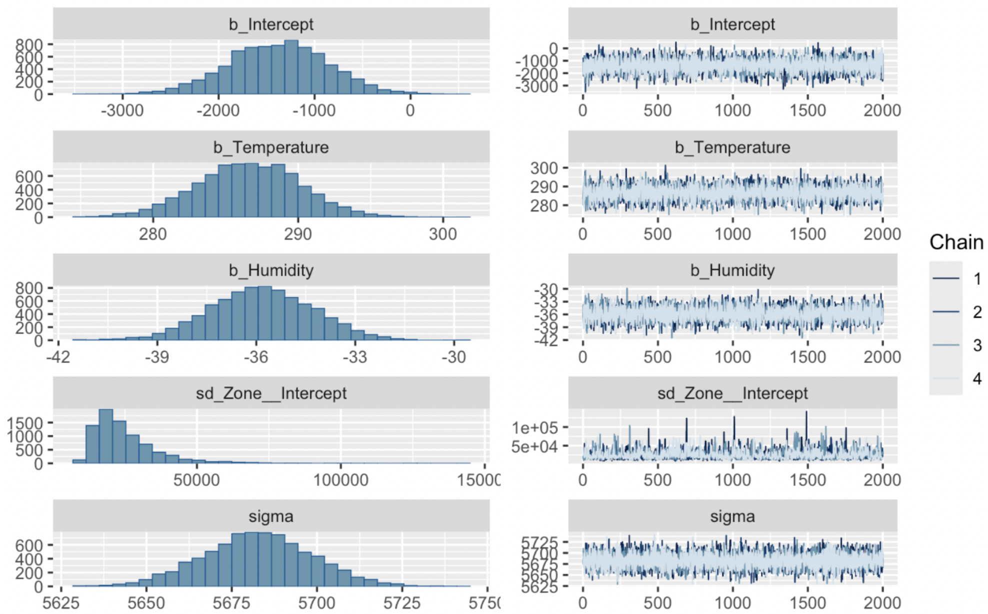

Figure 6: Plot of the model

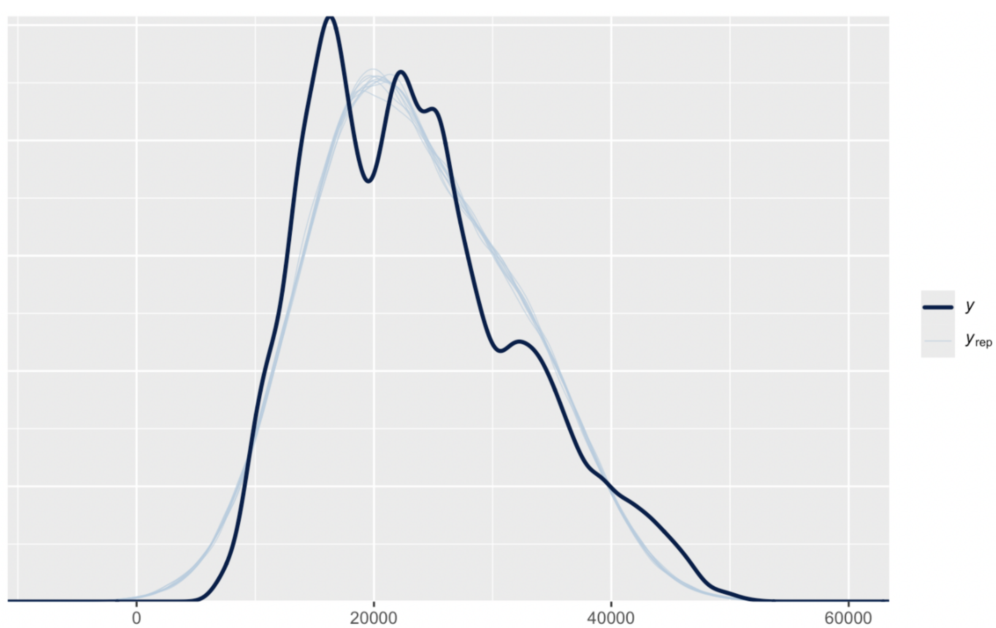

Figure 7: Posterior Probability Check