Temperature plays a persistent and crucial role in power consumption.

r13250004 邵子軒

December 19, 2025

1 Data Descripition

Data resources : `https://archive.ics.uci.edu/dataset/849/power+consumption+of+tetouan+`

”We selected five variables as covariates, textbfTemperature, **Humidity**, **Wind Speed**, **General diffuse flows**, and **Diffuse Flows** with **Zone 1 Consumption** designated as the response variable. Figure 2 illustrates the distribution of these covariates against the response. As observed in Figure 2, the relationships exhibit non-linear patterns. Consequently, we applied the Nadaraya-Watson estimator to perform univariate regression of the response on each covariate; the results are presented in Figure 3. The distribution of Wind Speed appears relatively irregular. In contrast, for Temperature, Diffuse Flows, and General Diffuse Flows, we observe an increasing trend in power consumption as these values rise, disregarding the boundary regions where the estimation is subject to edge effects.

2 Semiparametric Model Analysis

We define the observed response variable y as **Zone 1 Consumption** and the observed covariate vector $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_5)^\top$ corresponding to **Temperature**, **Humidity**, **Wind Speed**, **General Diffuse Flows**, and **Diffuse Flows**.

$$E(y|\mathbf{x}) = m(\mathbf{x}) = g(\nu_\beta(\mathbf{x})) \tag{1}$$

The formulation of Equation (1) depends on the specific model employed. In the context of ADE in [5], it represents a single-index model, whereas for SIR [6] and MAVE [7], it corresponds to a multi-index model.

In our analysis, we employ ADE, SIR, and MAVE to estimate the coefficients in (1). For ADE and MAVE methods, we used the normal kernel and selected the bandwidth using the robust rule-of-thumb proposed in [1] and used the diagonal bandwidth matrix to estimate the density of covariates. (2) For SIR method, we slice y into $H = 100$

This method adjusts the standard deviation by incorporating the interquartile range (IQR) to mitigate the influence of outliers, defined as:

$$h = 0.9An^{-1/5}$$

where $A = \min(\hat{\sigma}, \mathrm{IQR}/1.34)$.

We use the **R** package **KernSmooth** to calculate the coefficients in (1). Due to the scale difference of covariates and the identifiability of β in ADE assumptions, we standardize the data before estimation and also standardize the $\hat{\beta}$ of ADE. We use the algorithm in [6] to estimate the covariance matrix of z instead of kernel estimation in order to reduce computational cost. In [7], the estimation of β incurs a high computational cost; therefore, we use kNN to reduce the computational burden, setting k=500.

Table 1: Comparison of Estimated Coefficients by Method

Variable	**ADE**	**SIR**	**MAVE**
Diffuse Flows	**0.7894**	0.0025	-0.0002
General Diffuse Flows	0.5890	0.0021	-0.0031
Wind Speed	0.1281	-0.0357	-0.2671
Humidity	-0.0864	-0.0078	-0.1012
Temperature	0.0781	**0.0575**	**0.9583**

Table 1 presents the estimation results from the three methods. In the ADE analysis, Diffuse Flows appear to be the dominant variable influencing the response y. In contrast, the results from SIR and MAVE identify Temperature as the primary covariate. This divergence in the identified significant variables is a noteworthy finding.

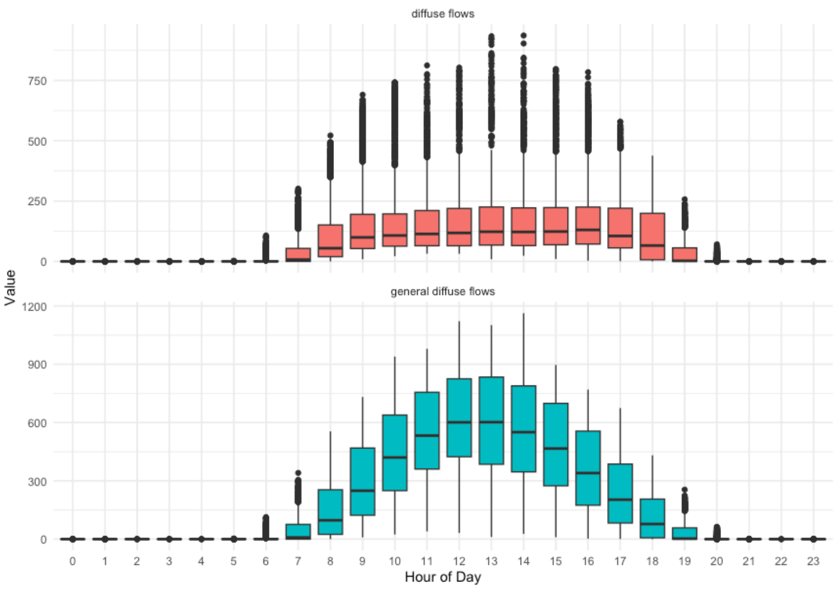

Figure 1: Boxplots of Diffuse Flows and General Diffuse Flows by Hour of Day

The basic idea of ADE in [5] is rather than trying to fit the entire curve (like SLS), it is better to directly calculate the 'local slope' at each point and then average these slopes together. This average direction is the direction of β. The distinct divergence in coefficient estimation between the ADE and MAVE methods, as presented in Table 1, can be seemed in Figure 1. Figure 1 reveal a structure characterized by high variance and a significant number of outliers, indicating rapid fluctuations and non-smooth behavior during daylight hours. This characteristic strongly aligns with the objective function of the Average Derivative Estimation (ADE) method. Since ADE focuses on estimating the derivatives of the regression function, it is inherently sensitive to local rates of change. Consequently, ADE assigns the highest weight (0.7894) to Diffuse Flows, interpreting the high-frequency variations and ”noise” as critical derivative information that explains instantaneous shifts in power consumption.

In contrast, SIR relies on the eigendecomposition of the covariance matrix of the slice means $E(z|y)$. Consequently, it captures the global trend of the inverse regression curve rather than the local changes. This makes SIR less sensitive to the high-frequency fluctuations (derivatives) observed in Diffuse Flows, which are explicitly captured by ADE. Instead, SIR prioritizes directions that maximize the between-slice variance, aligning more closely with stable predictors like Temperature.

On the other hand, MAVE aims to minimize the estimated conditional variance given the true direction. Unlike the other two methods, it explicitly considers the link function g. In terms of its algorithm, it iteratively updates the weights, allowing for a more effective estimation of β. From Figures 4 and 5, we can see that the trend of power consumption and temperature during the daytime is similar. Therefore, the MAVE estimation appears to be more reasonable.

3 Time Varying Coefficient Model

We consider the varying coefficient model in [8],

$$y = \beta_0(t) + \sum_{p=1}^{5} x_p \beta_p(t). \tag{2}$$

To further analyze the results observed in Table 1, where ADE identifies solar radiation as important while MAVE prioritizes temperature, we relax the assumption that the influence of variables (coefficient β) is fixed. Instead, we assume that these coefficients vary as a function of ”Time (Hour) (t).” By observing the evolution of these coefficients over time, we aim to capture the alignment between the temporal trend of Temperature and Power Consumption. The methods discussed in Section 2 fail to capture the distinct diurnal characteristics—specifically the differences between day and night—of the covariate coefficients, and are therefore unable to detect this similarity. We construct the model using the **mgcv** package, implementing a GAM with cyclic cubic regression splines and setting the number of knots to 12.

By shifting focus between the Top-Left (s(Hour)) and Top-Middle (Temperature) panels of Figure 6, we can observe distinct structural similarities. The baseline power curve (Top-Left) illustrates the dominant rhythm of human activity: climbing from early morning, peaking in the afternoon, and maintaining high levels into the evening, representing a smooth, long-wavelength fluctuation. Similarly, the temperature effect curve (Top-Middle), exhibits a smooth, large-wave characteristic. Analyzing the temporal evolution reveals that between 06:00 and 12:00, as baseline power rises, the Temperature coefficient undergoes significant variation—initially descending and then ascending—indicating a major shift in power sensitivity to Temperature during morning hours. Crucially, during the afternoon period from 12:00 to 18:00, while Power Consumption sustains high levels, the temperature coefficient also stabilizes at a relatively high point. This demonstrates a clear alignment between the waveform frequency of temperature and that of Power Consumption.

4 Summary

This analysis investigates the environmental covariates influencing Zone 1 power consumption using semiparametric dimension reduction techniques and a Time-Varying Coefficient Model. Initially, the study applied three methods—Average Derivative Estimation (ADE), Sliced Inverse Regression (SIR), and Minimum Average Variance Estimation (MAVE)—for analysis.

A divergence in variable significance was observed among the methods: due to its sensitivity to local high-frequency fluctuations, ADE identified 'Diffuse Flows' as the dominant variable, whereas SIR and MAVE identified 'Temperature' as the most critical factor. Subsequently, this study constructed a Time-Varying Coefficient Model. The results demonstrate that the influence of covariates is not fixed but varies significantly over time. Specifically, the temperature coefficient aligns closely with power consumption trends during peak daytime hours. This finding not only validates the results of MAVE and SIR but also explains the local fluctuation characteristics captured by ADE, underscoring the necessity of incorporating the temporal dimension in power load modeling.

References

[1] Bernard W Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, London, 1986.

[2] Wolfgang Härdle. *Nonparametric and semiparametric models*. Springer Science & Business Media, 2004.

[3] Hidehiko Ichimura. Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of econometrics*, 58(1-2):71–120, 1993.

[4] MP Wand. Fast computation of multivariate kernel estimators. *Journal of Computational and Graphical Statistics*, 3(4):433–445, 1994.

[5] Wolfgang Härdle and Thomas M Stoker. Investigating smooth multiple regression by the method of average derivatives. *Journal of the American statistical Association*, 84(408):986–995, 1989.

[6] Ker-Chau Li. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327, 1991.

[7] Yingcun Xia, Howell Tong, Wai Keung Li, and Li-Xing Zhu. An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64(3):363–410, 2002.

[8] Trevor Hastie and Robert Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 55(4):757–779, 1993.

5 Appendix

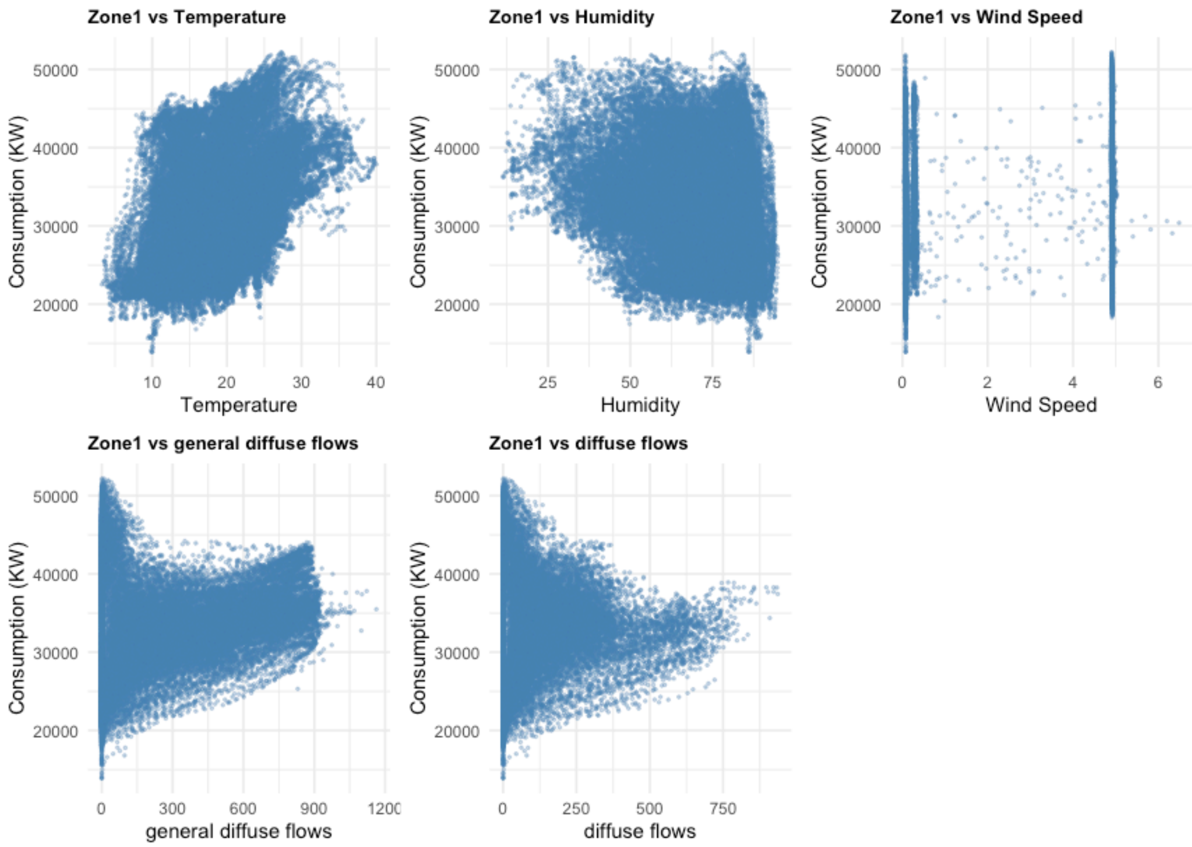

Figure 2: Scatter Plots of Zone 1 Consumption against Five Covariates

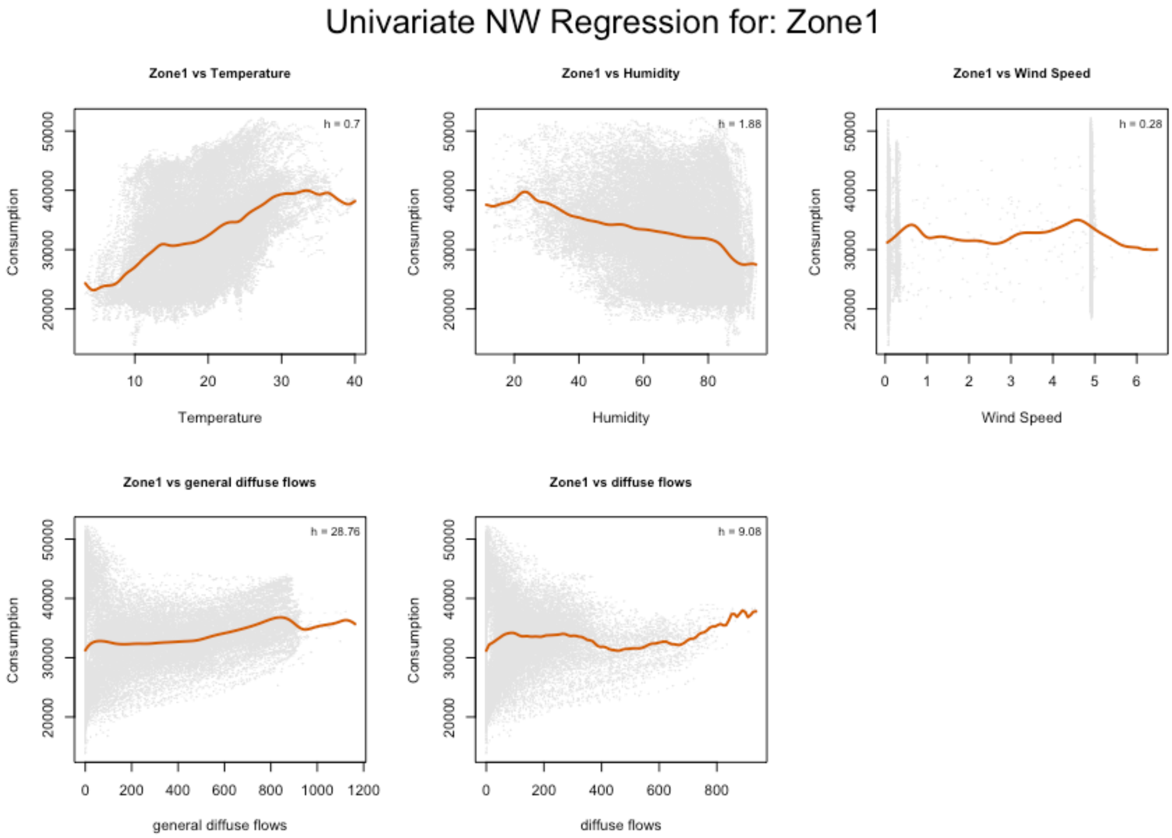

Figure 3: Univariate Nadaraya-Watson Regression Results for Zone 1

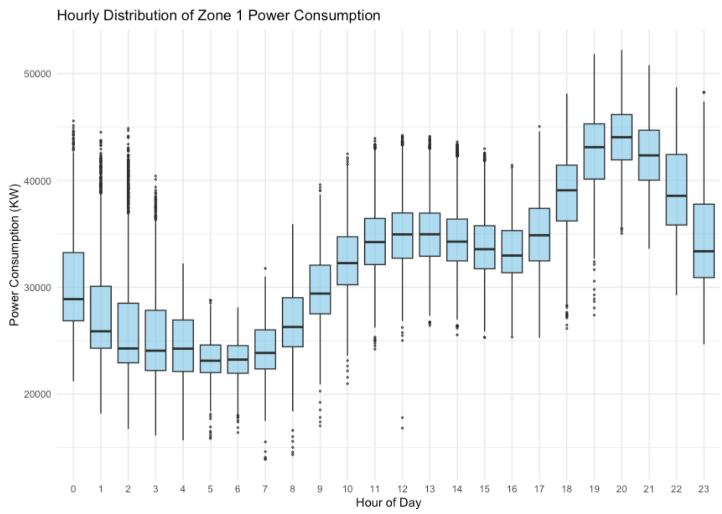

Figure 4: Hourly Distribution of Zone 1 Power Consumption

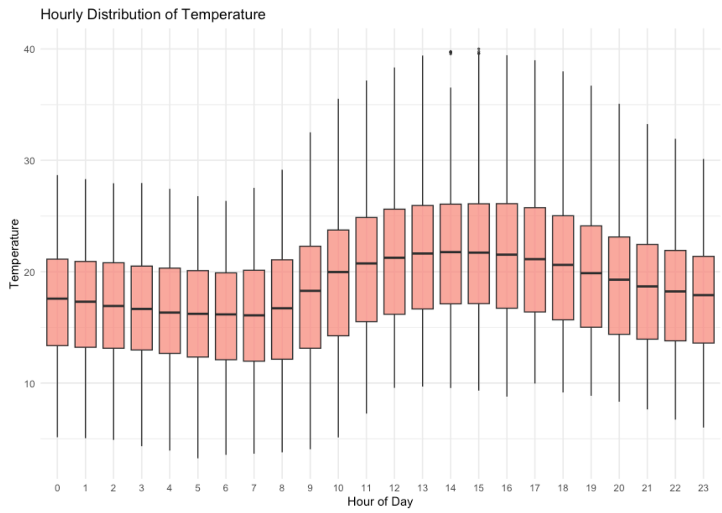

Figure 5: Hourly Distribution of Temperature

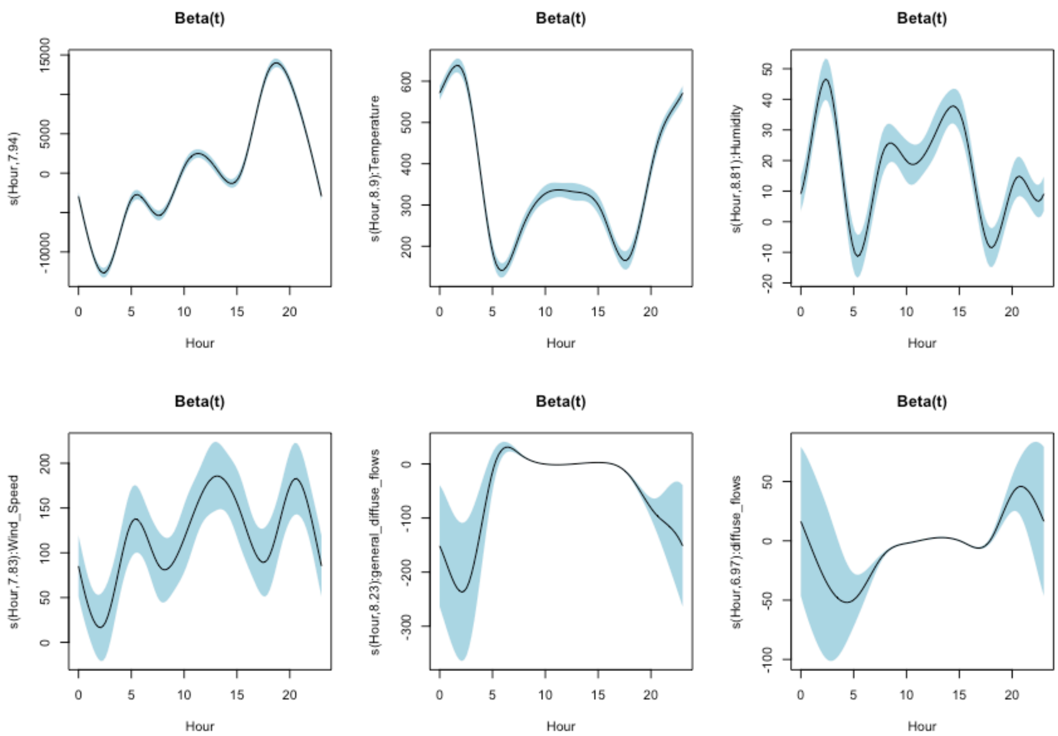

Figure 6: Estimated Time-Varying Coefficients ($\beta(t)$) over 24 Hours