Temporal Dynamics of a City's Power Consumption

April 1, 2025

1 Background

An efficiently run electrical grid is characterized by a balance between power supply and demand that errs on the side of (minimal) overproduction to avoid service disruptions to its customers. Maintaining such a balance is challenging as demand fluctuates in response to weather, patterns of consumer behavior, pricing and variation in population size, among other factors. Compounding these, existing systems are increasingly challenged as populations grow and extreme weather events become more common, making accurate forecasts of consumption patterns all the more critical.

Here we have a year's worth of data on power consumption in Tetouan, Morocco. The data set provides power consumption in each of three zones at 10 minute intervals (n=52,417), date and time stamps and a collection of weather metrics. The data are described in Salam and Hibaoui (2018), where the authors utilize these data as a case study (see Section III) to evaluate a collection of machine learning algorithms, an exercise that we will here asiduously avoid.

2 Analysis Guidelines

- Set aside the December 2017 power consumption data from Zones 1 and 2 and the November *and* December power consumption data from Zone 3 as a "test set."
- Use the remaining data ("training set") to construct a model for power consumption conditional on the weather data, zone and time.
- Describe estimated associations between power consumption, zone and over time.
- *Once you have finished model building*, use your model to forecast the test data. Report RMSE (and any other metric you see as relevent) for these predictions.

3 Data

The data can be found in the file `powerConsumption.csv`. See Table 1 for the first several observations and formating. The file contains the following variables:

DateTime Date and time of the measurements.

Temp Temperature in celsius.

Humidity Relative humidity (percentage).

WindSpeed Wind speed (units unknown).

GDF General diffuse flows. A measure of soil water flow (units unknown).

DF Diffuse flows (as above, specifics unknown).

Zone1 Power consumption in Zone 1.

Zone2 Power consumption in Zone 2.

Zone3 Power consumption in Zone 3.

	DateTime	Temp	Humidity	WindSpeed	GDF	DF	Zone1	Zone2	Zone3
1	1/1/2017 0:00	6.56	73.80	0.08	0.05	0.12	34055.70	16128.88	20240.96
2	1/1/2017 0:10	6.41	74.50	0.08	0.07	0.09	29814.68	19375.08	20131.08
3	1/1/2017 0:20	6.31	74.50	0.08	0.06	0.10	29128.10	19006.69	19668.43
4	1/1/2017 0:30	6.12	75.00	0.08	0.09	0.10	28228.86	18361.09	18899.28
5	1/1/2017 0:40	5.92	75.70	0.08	0.05	0.09	27335.70	17872.34	18442.41
6	1/1/2017 0:50	5.85	76.90	0.08	0.06	0.11	26624.81	17416.41	18130.12

Table 1: First six observations of the data set.

4 Assignment

- Conduct an exploratory data analysis for class on **Monday April 7**. It may be useful to experiment with frequentist models at this stage in your analysis. Begin preparing your project slide deck by describing this analysis, highlighting any issues you encountered and unresolved questions you may have, and summarizing your preliminary inferences and discoveries. Be prepared to share these in class.

- *Add* to your project slides *about* five slides describing progress on your Bayesian analysis, highlighting assumptions, shortcomings and obstacles you have encountered (if any). Be prepared to begin presenting these in class **Monday April 14**. Turn in your full set of slides (EDA + model–based analysis) on **Wednesday April 16**.

- Describe your analysis in a formal report of up to four pages (writing and display equations); all tables and figures should appear in a separate appendix and be referenced by name in the main text. Your report should highlight all relevant aspects of your analysis (exploratory and modeling) and include graphical and numerical summaries that aid in communicating your results. You may also include code and other supporting material in the supplemental appendix. Due at the start of class **Wednesday April 16**.