Visualize and Analyze Energy Data

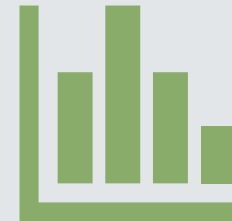*Insight into the Usage of the Electrical Sub-metering Devices*

Agenda

What is this analytics for ?

Do sub-meters provide homeowners the "useful" usage analytics?
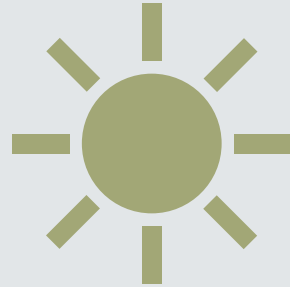
Can we predict future energy consumption from the existing data?

Can we improve the future data collection?

Description of the data

- Gathered in a house located in Sceaux (7km of Paris, France).

- 1,569,894 observations, 5 Attributes

- 1.1.2007 ~ 12.30.2009 (36 months)

- sub_metering_1: kitchen (dishwasher, oven, microwave)

- sub_metering_2: laundry room (washing-machine, tumble-drier, refrigerator, light)

- sub_metering_3: electric water-heater and air-conditioner

|	Date	Time	Sub_metering_1 (watt-hour)	Sub_metering_2 (watt-hour)	Sub_metering_3 (watt-hour)
1	2007-01-01	00:00:00	0	0	0
2	2007-01-01	00:01:00	0	0	0
.....	.....	.....	.....	.....	.....
1569894	2009-12-31	23:59:00	0	0	19

Statistics of the data

- Highest average energy consumption: submeter 3 (Electric water-heater and air-conditioner)

- Highest maximum energy consumption: submeter 1 (Kitchen)

- The minimum energy consumption of the three submeters are all 0.

- The Median of submeter 1 and submeter 2 are 0: over 50% of the time, there is no energy consumed measured by these two submeters

	Sub_metering_1	Sub_metering_2	Sub_metering_3
Min.	0	0	0
1st Qu.	0	0	0
Median	0	0	1.000
Mean	1.159	1.343	6.216
3rd Qu	0	1.000	17.000
Max.	82.000	78.000	31.000

Power Consumption of 3 submeters in day, week, year time period

Power Consumption of second Week in 2008 — Week

Power Consumption January 9th, 2008 — Day
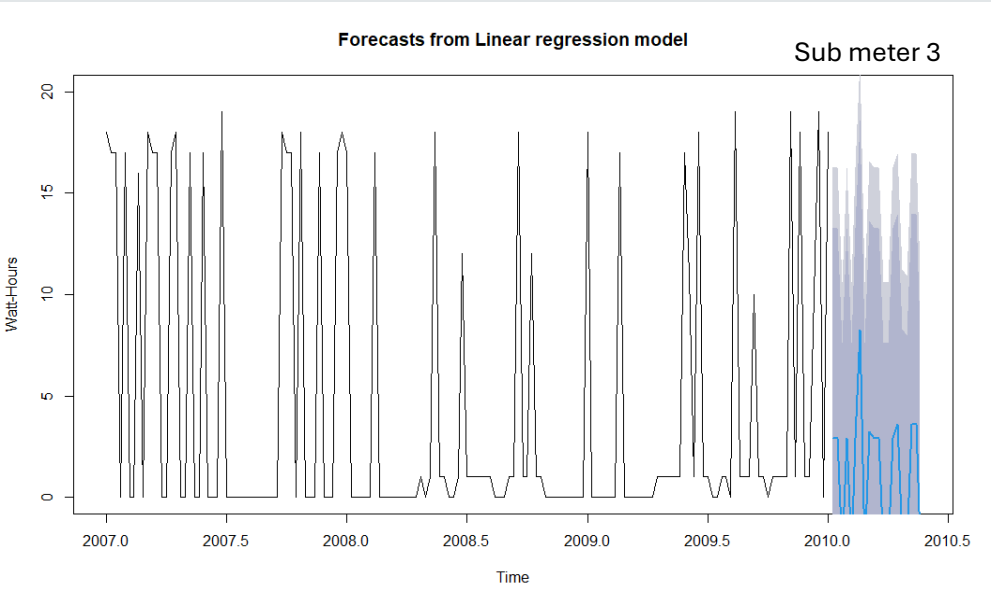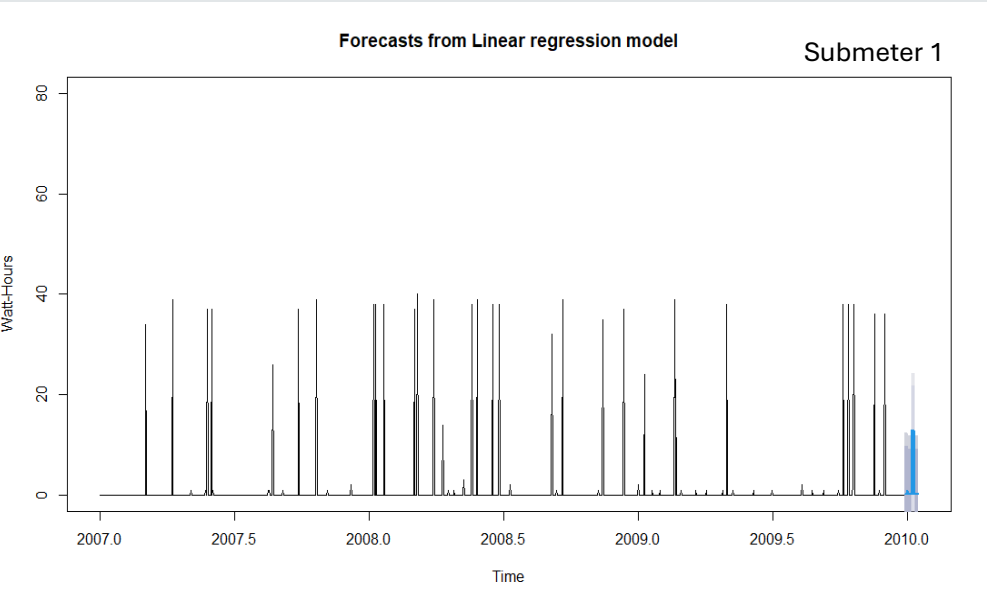

Power Consumption in 2008 — Year
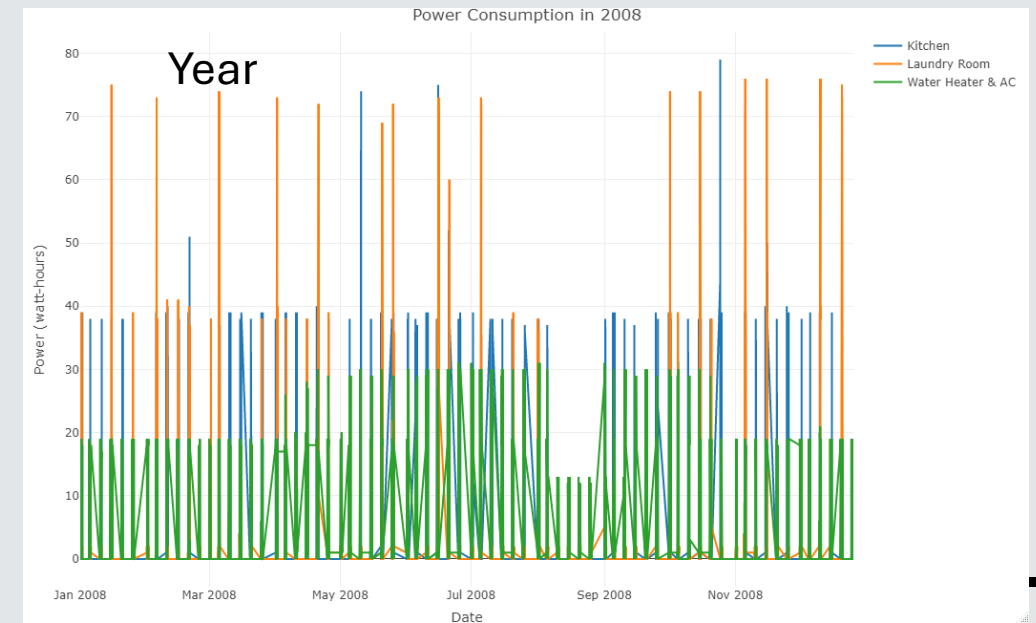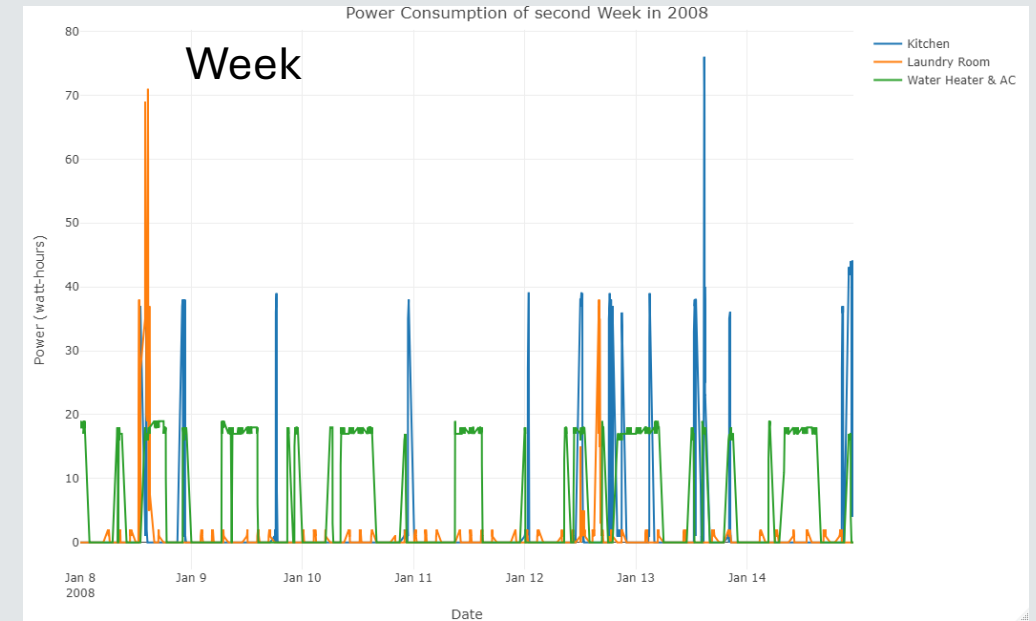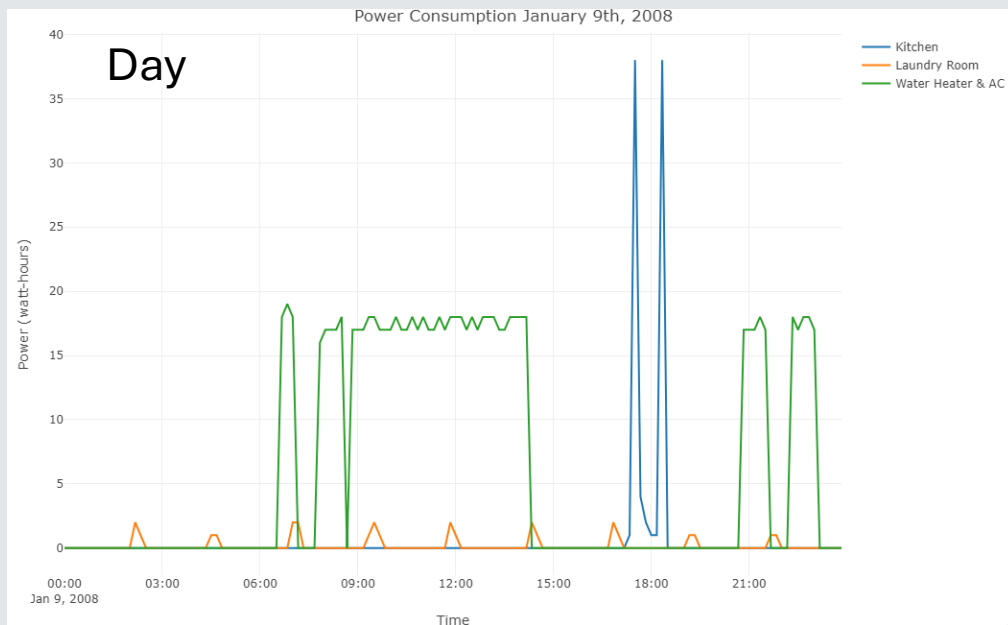
Power Consumption of 3 submeters during 2007-2009
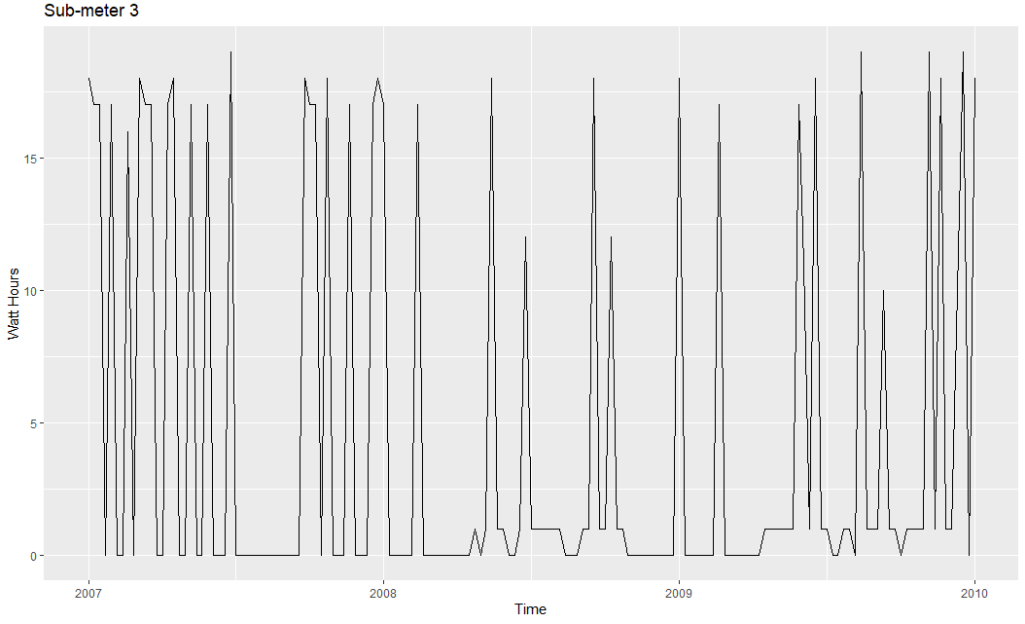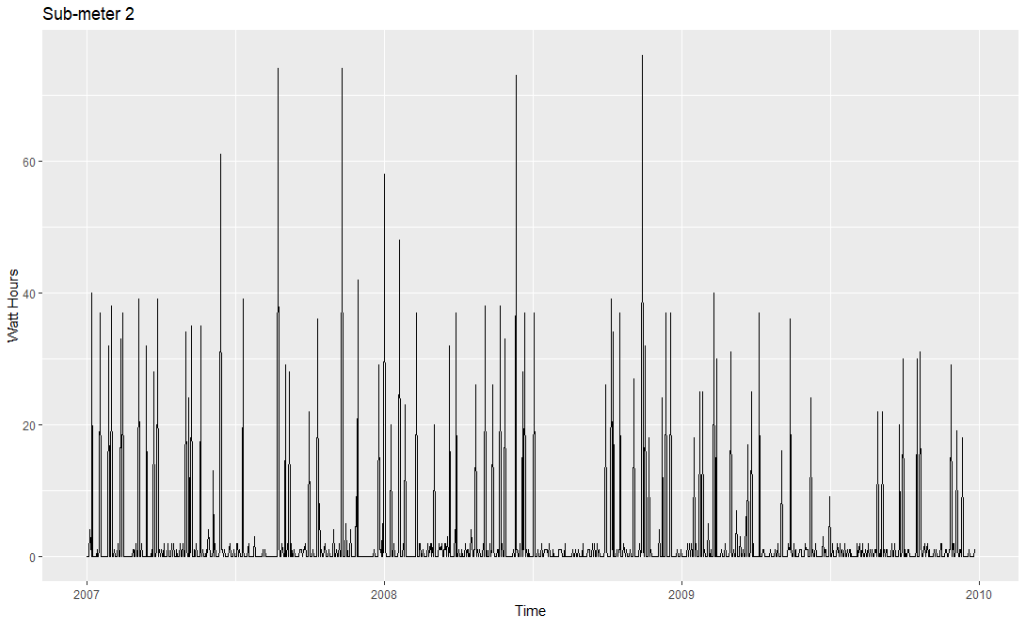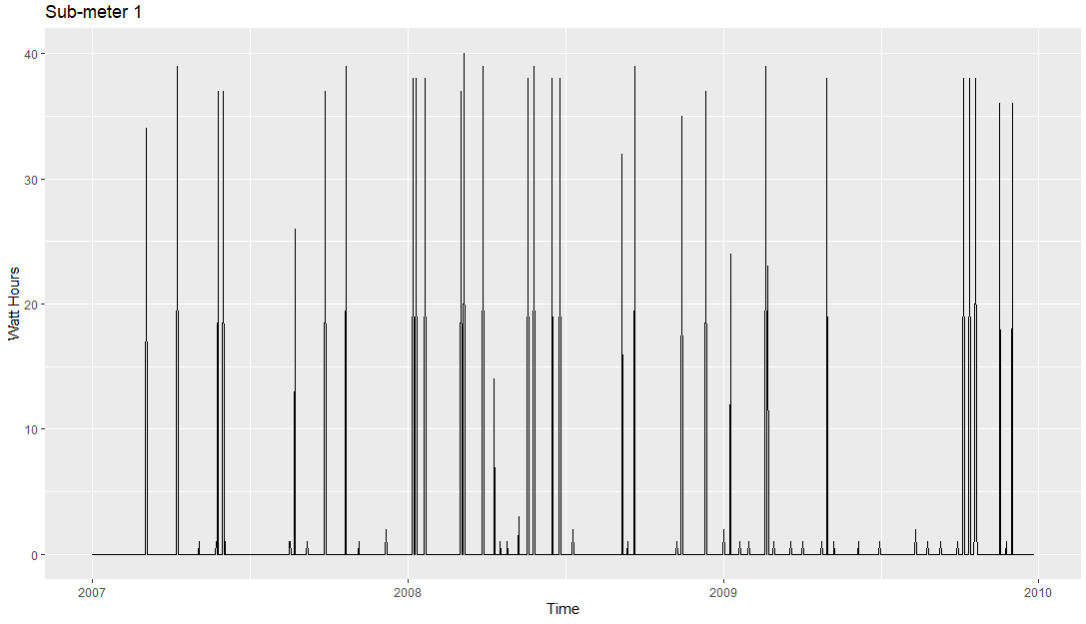
Linear Regression Forecast of 3 submeters

	Multiple R-squared	Residual standard error
Submeter 1	0.3831	6.871
Submeter 2	0.3151	6.059
Submeter 3	0.3207	9.111

Forecasts from Linear regression model — Submeter 2

Forecasts from Linear regression model — Submeter 1

Forecasts from Linear regression model — Sub meter 3
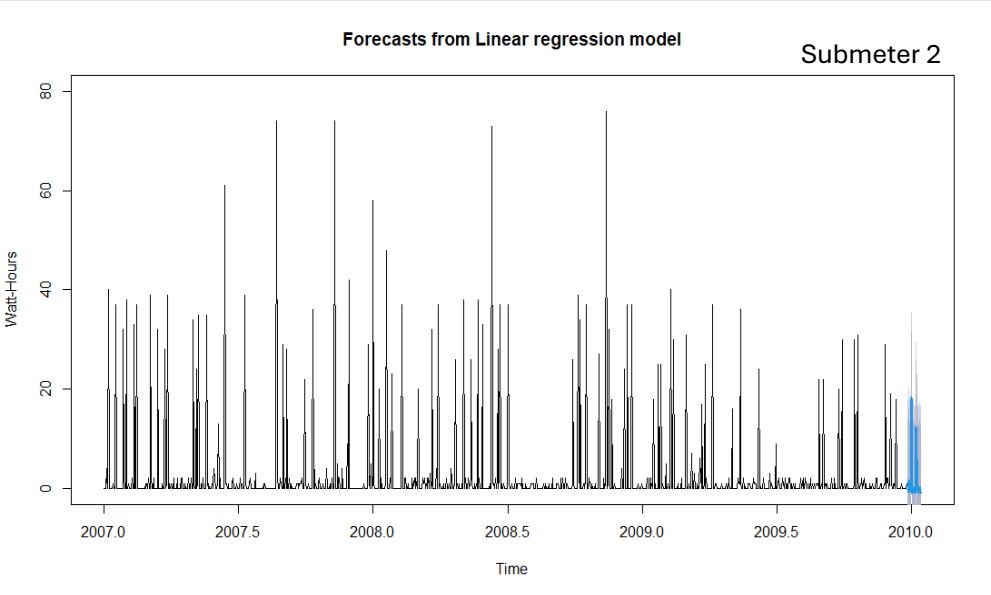
Decomposition Visualizations of 3 Submeters

	Seasonal	Trend	Random
Submeter 1	Length 1091	Length 1091	Length 1091
Submeter 2	Length 1090	Length 1090	Length 1090
Submeter 3	Length 157	Length 157	Length 157
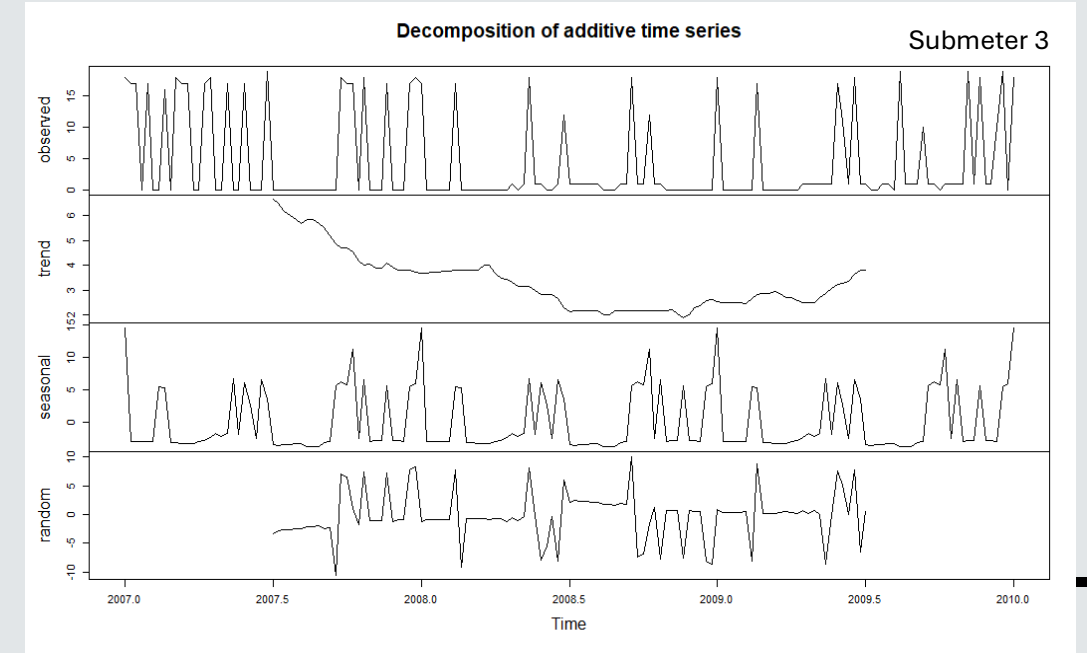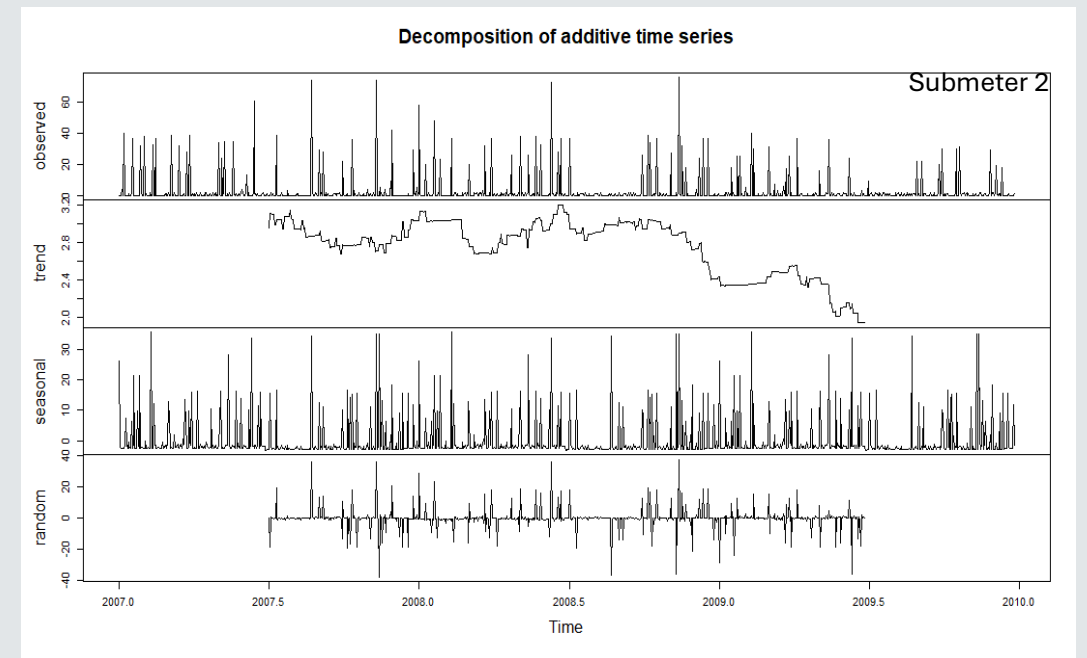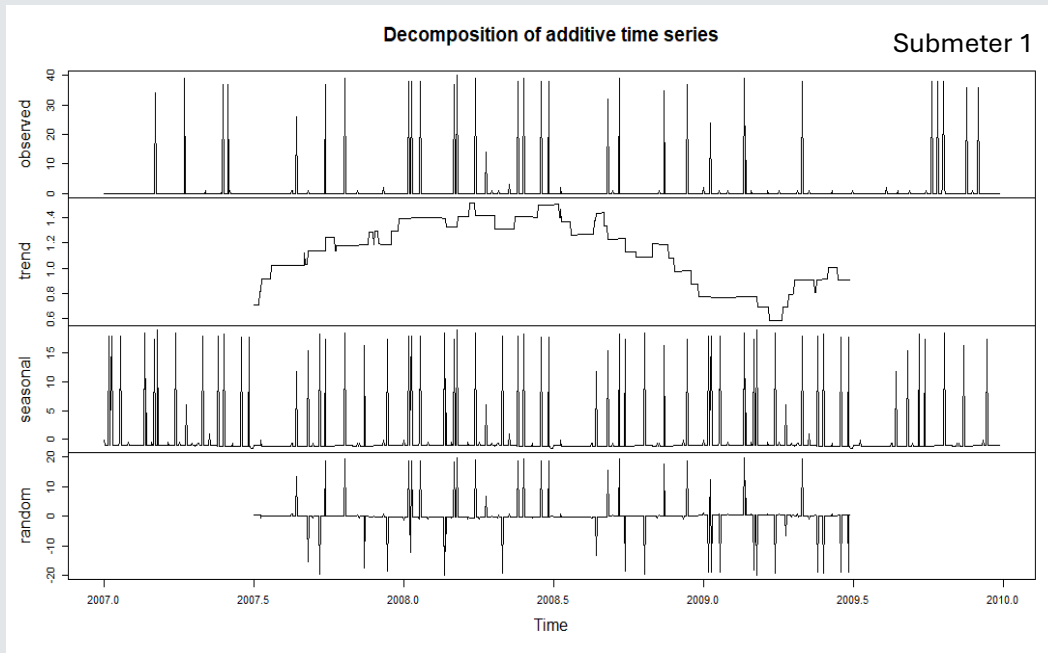
Holt Winters Forecasting and Analysis
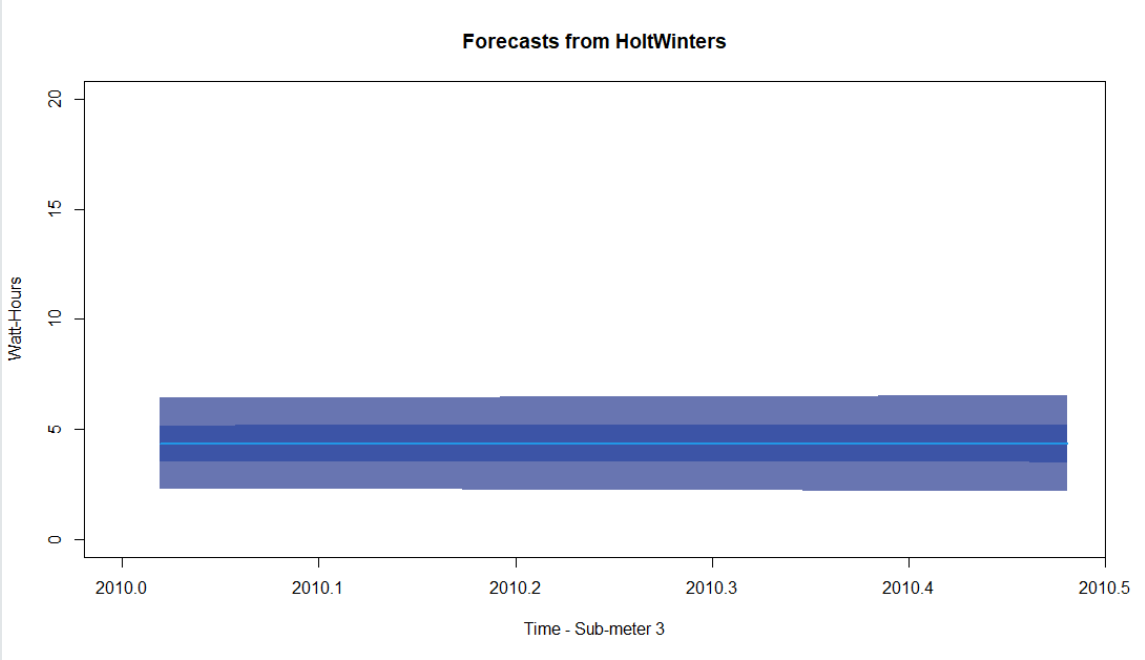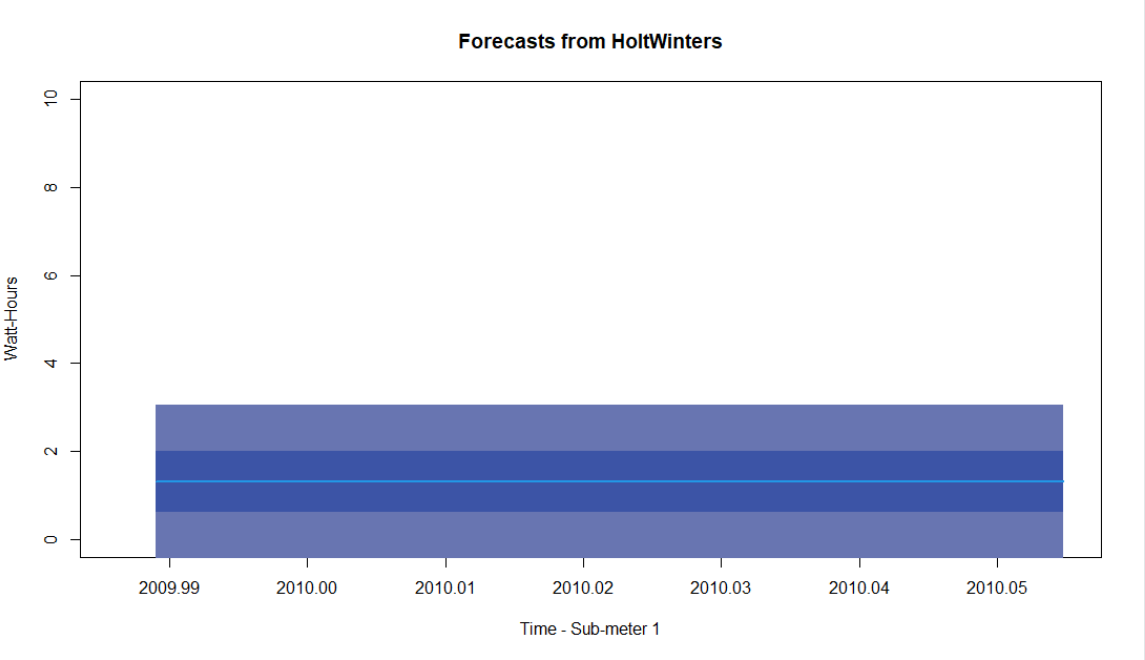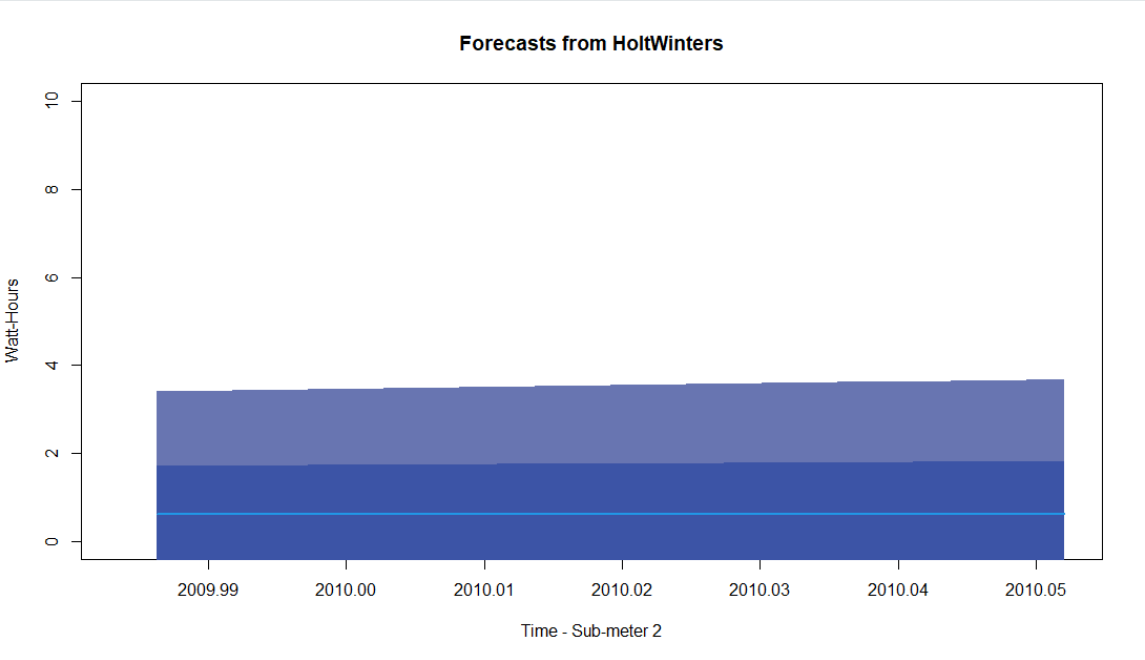
Analyzing Results of the Predictions

- Compare to the Linear Regression Forecast images, the Holt Winters Forecasting images show a very consistent forecast for all three sub-meters. This may be because we remove the seasonal and random components from the original data.

- We may need to increase the frequency of data during the time period to increase the forecast accuracy.

- We may need to increase the length of observation time to increase the forecast accuracy.

The Goal of this Project and the Recommendations

- Do sub-meters provide homeowners the "useful" usage analytics? Yes, we can provide the power consumption insights of different parts of the house by using sub-meters. We can also predict the power usage of different submeters.

- In the original data, washing-machine, tumble-drier, refrigerator and light were grouped together. I would recommend to put refrigerator in submeter 1, which contained the appliances in the kitchen. I also recommend to put the energy consumed of lights in a separate group.

- In the original data, electric water-heater and air-conditioner were grouped together. I would recommend put these two in separate submeters.

- The power usage trend can let the homeowners know when and which part of the house use the most power and help them to male a plan the save the energy.

Thank you!

Presented by Wanyun Ho
howanyun@gmail.com

Lesson Learned

- The time series analysis is very useful to analyze the data recorded in a consistent time and help to forecast the data in the future.

- The visualization skills of R in this task is useful for compare different data in the same time period.

- I will read more details about the time series analysis to apply this analysis skill to more different kind of data, such as weather forecasting, price trend etc.

- The decomposition method is useful to only focus on "TREND" of the data, such as revenue and profit forecasting.