ENERGY DATA ANALYSIS

Analysis and Recommendations

Kevin Burr

December 2019

WHAT WE WILL COVER

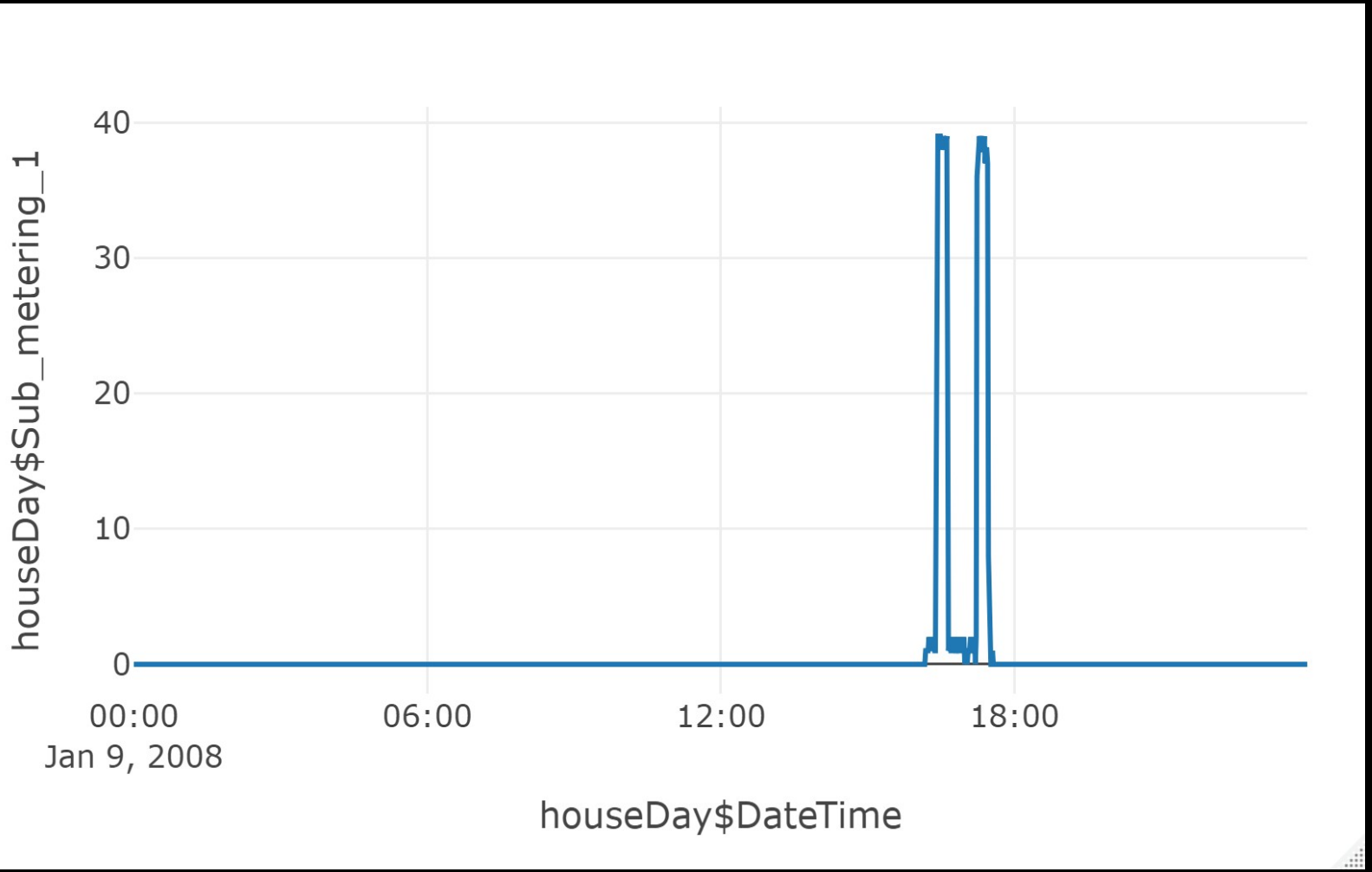
- Step 1: Analysis
- Step 2: Time Series Analysis
- Step 3: Linear Regression Forecasts
- Step 4: Decompositions (Trends and Seasonality)
- Step 5: Holt Winters Forecasts (Removing Seasonality)
- Useful Insight
- Summary Statement
- 5 Business Recommendations
- Lessons Learned

QUICK NOTE

- Sub Meter 1 corresponds to the Kitchen, containing a dishwasher, an oven and a microwave
- Sub Meter 2 corresponds to Laundry, containing a washing machine, tumbler dryer, refrigerator and light
- Sub Meter 3 corresponds to an Electric Water-Heater and an Air-Conditioner

STEP 1: ANALYSIS

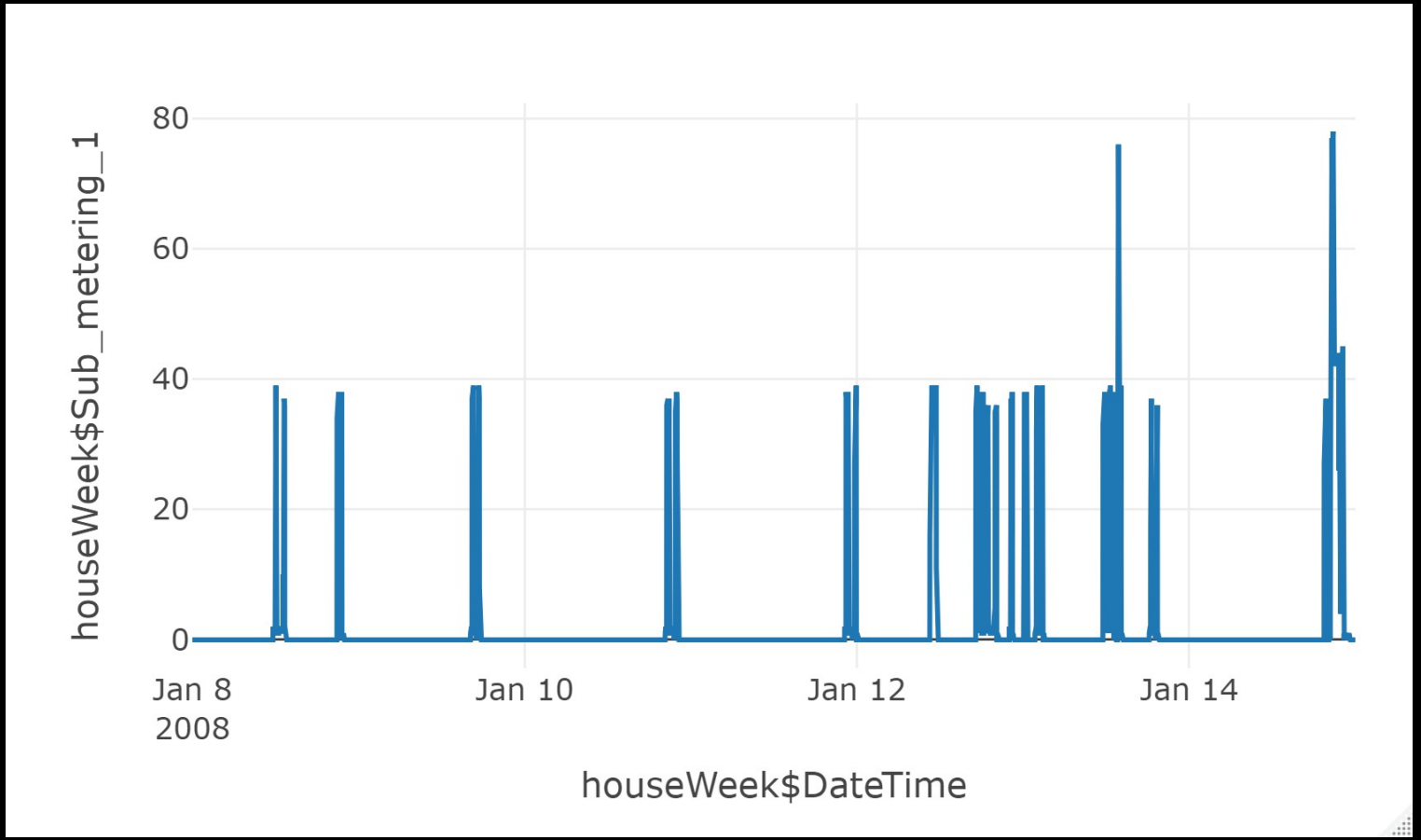
3 SLICES OF THE DATA



One Day of consumption
January 9, 2008
Sub Meter 1

STEP 1: ANALYSIS

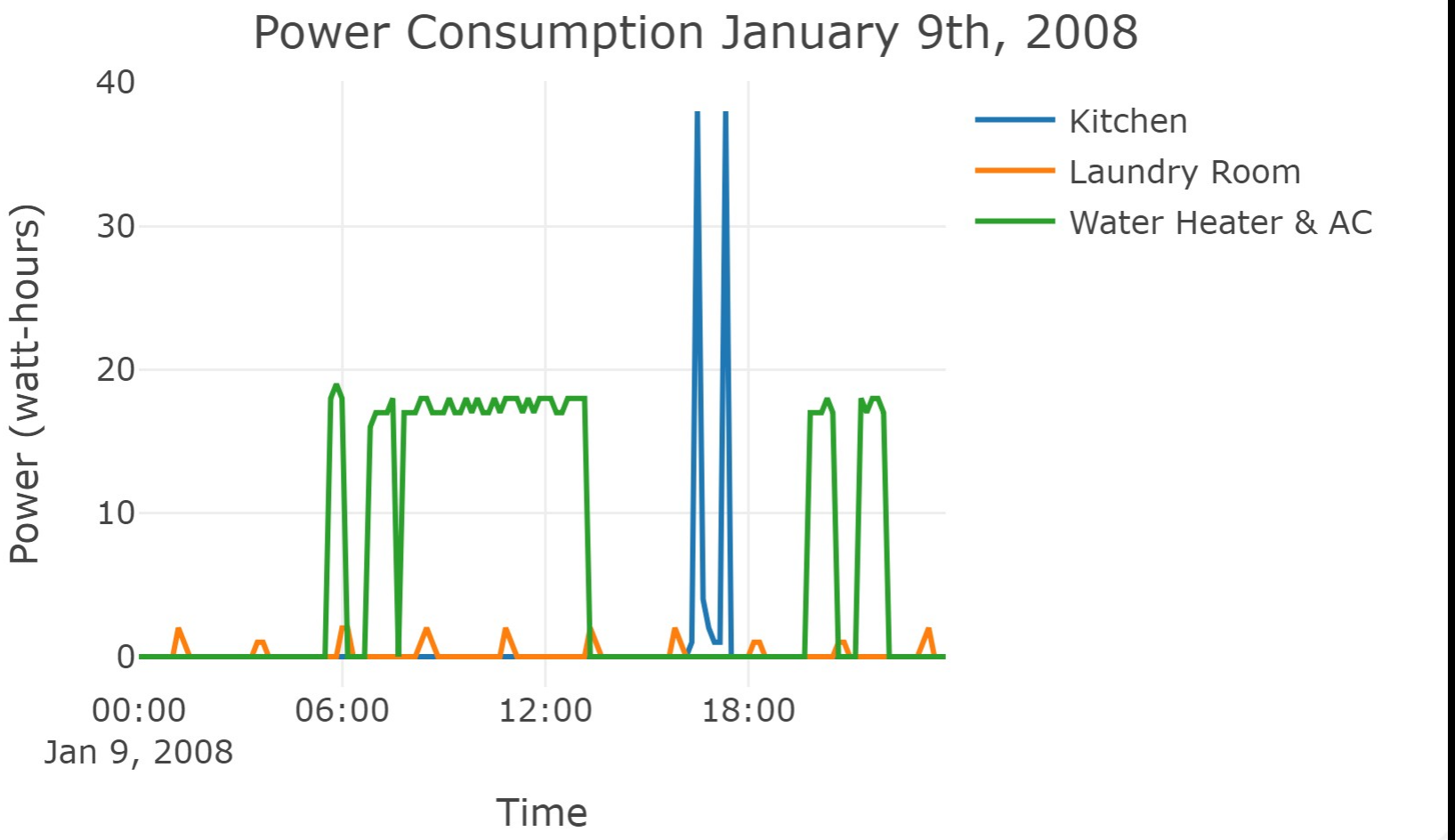
3 SLICES OF THE DATA



One Week of consumption
January 8, 2008
Sub Meter 1

STEP 1: ANALYSIS

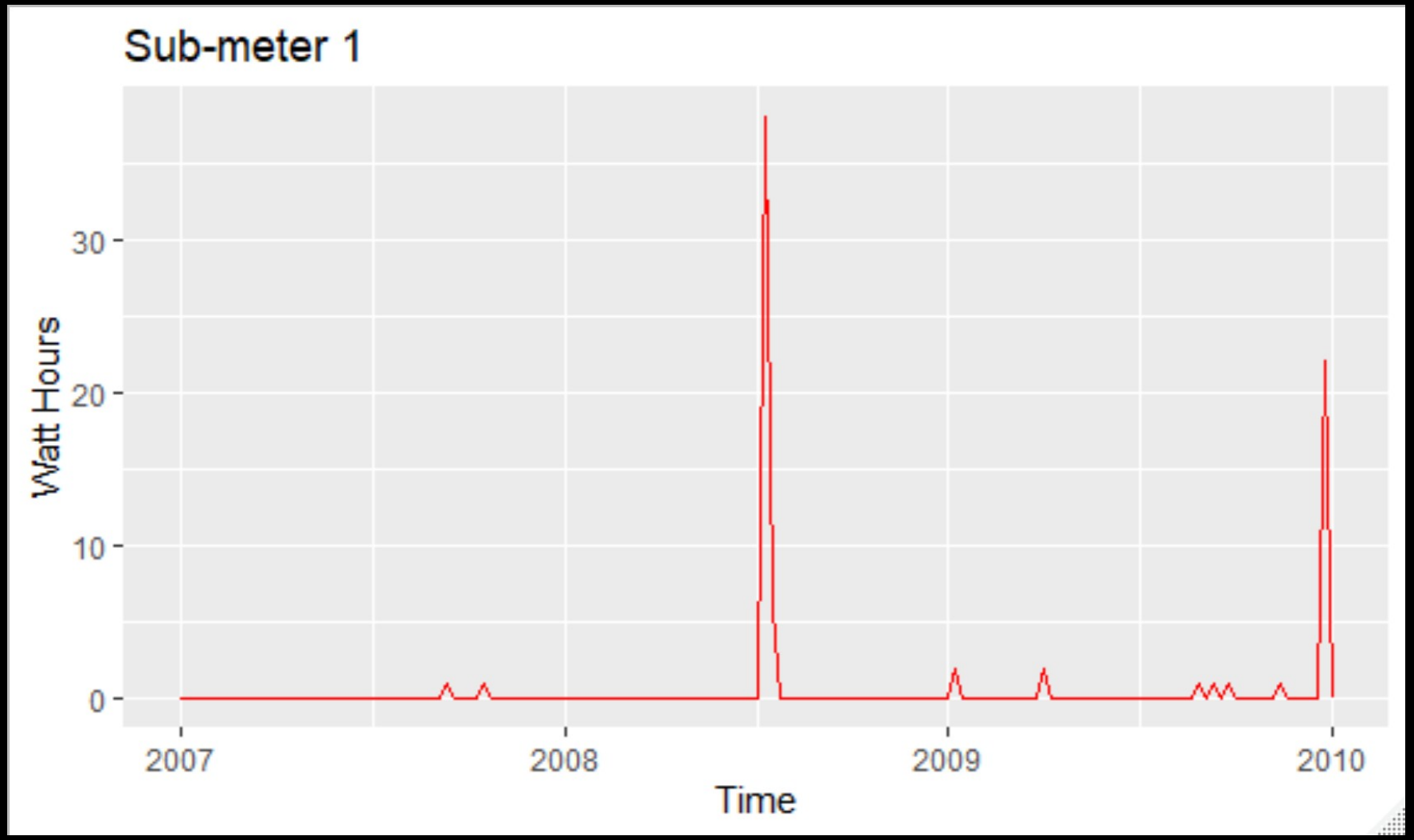
3 SLICES OF THE DATA



One Day of consumption
January 9, 2008
Sub Meter 1, 2 and 3
Every Ten Minutes (Freq)

STEP 2: TIME SERIES

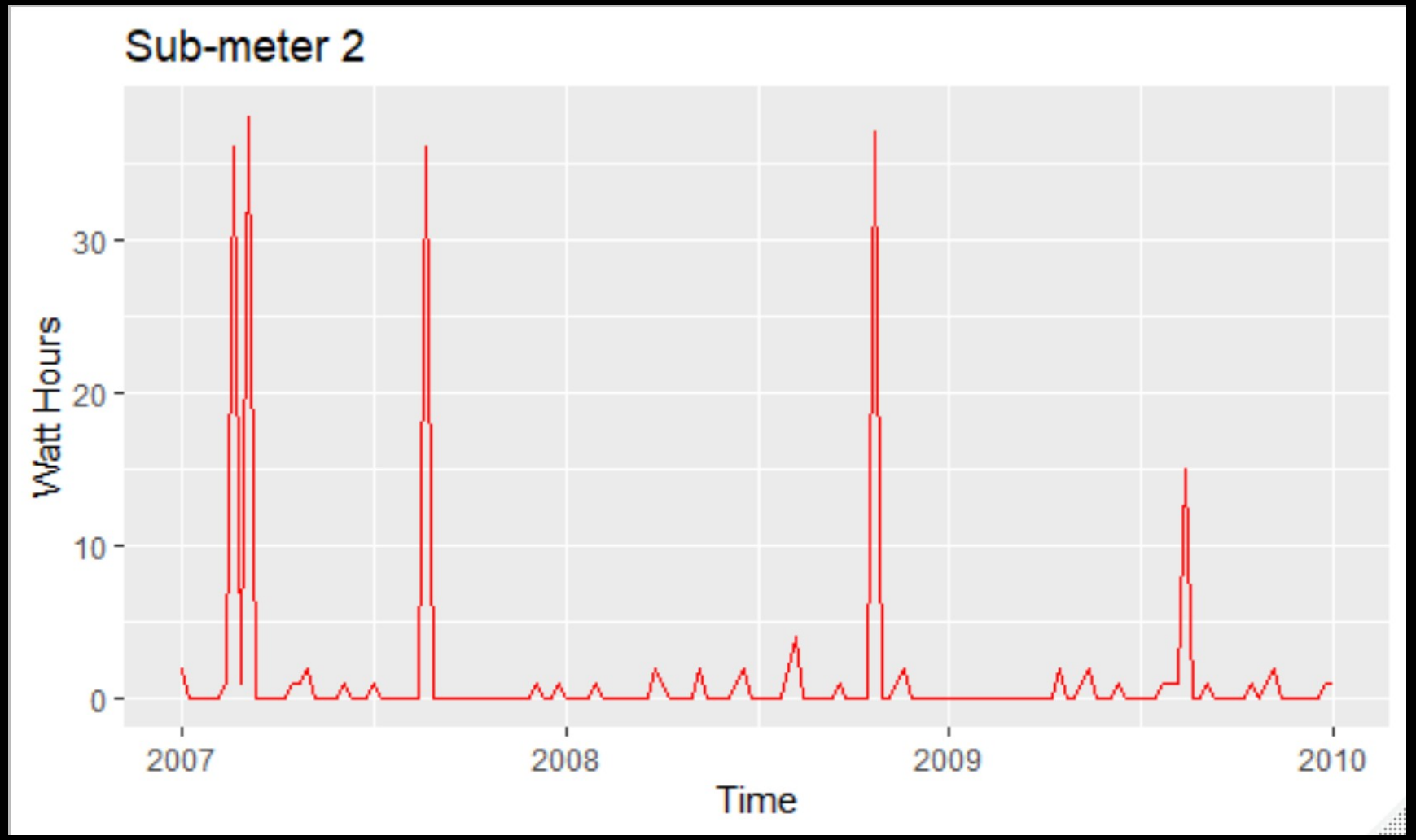
3 SLICES OF THE DATA



Time Series Consumption
Mondays 8pm
2007-2010
Sub Meter 1

STEP 2: TIME SERIES

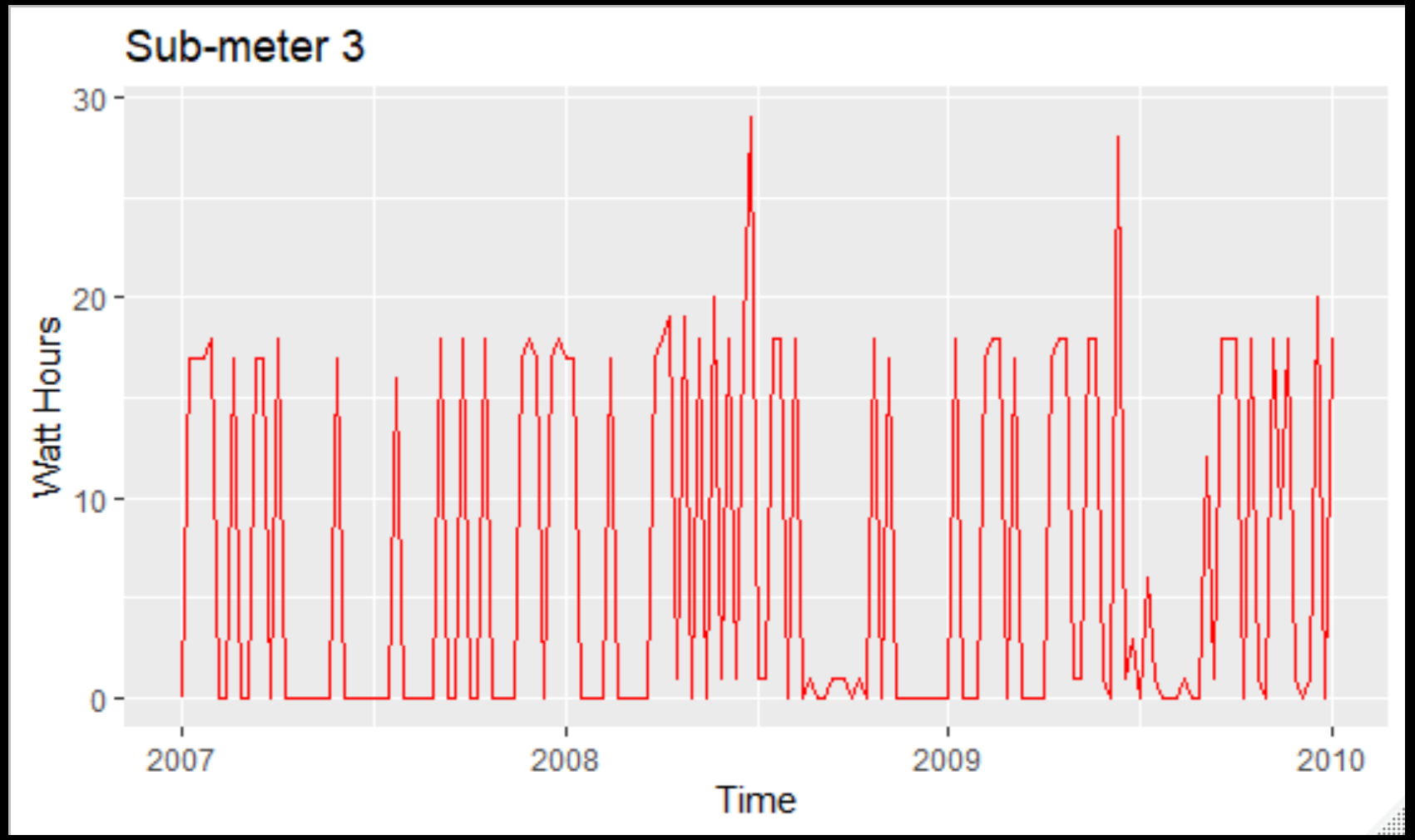
3 SLICES OF THE DATA



Time Series Consumption
Mondays 8pm
2007-2010
Sub Meter 2

STEP 2: TIME SERIES

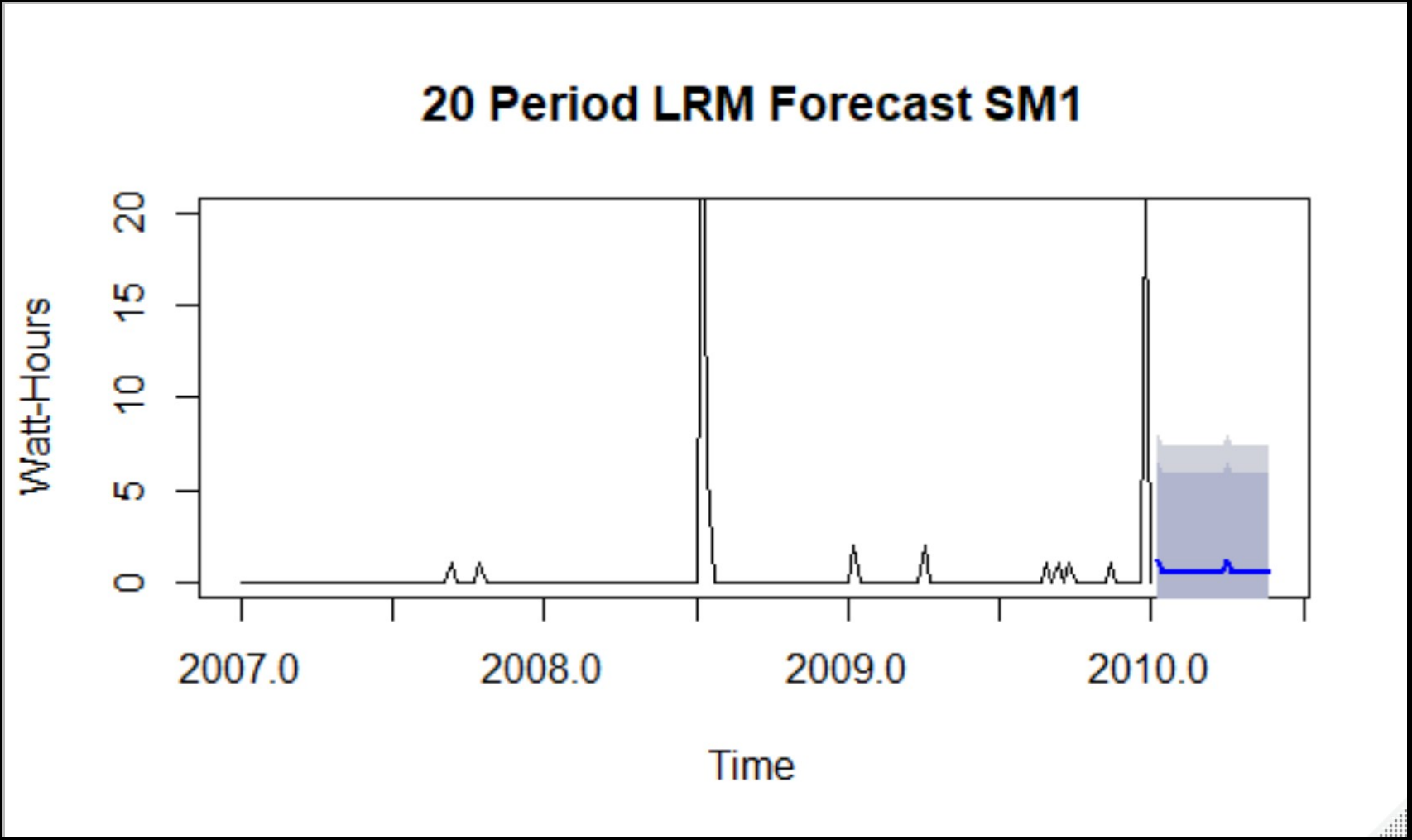
3 SLICES OF THE DATA



Time Series Consumption
Mondays 8pm
2007-2010
Sub Meter 3

STEP 3: LINEAR REGRESSION FORECASTS

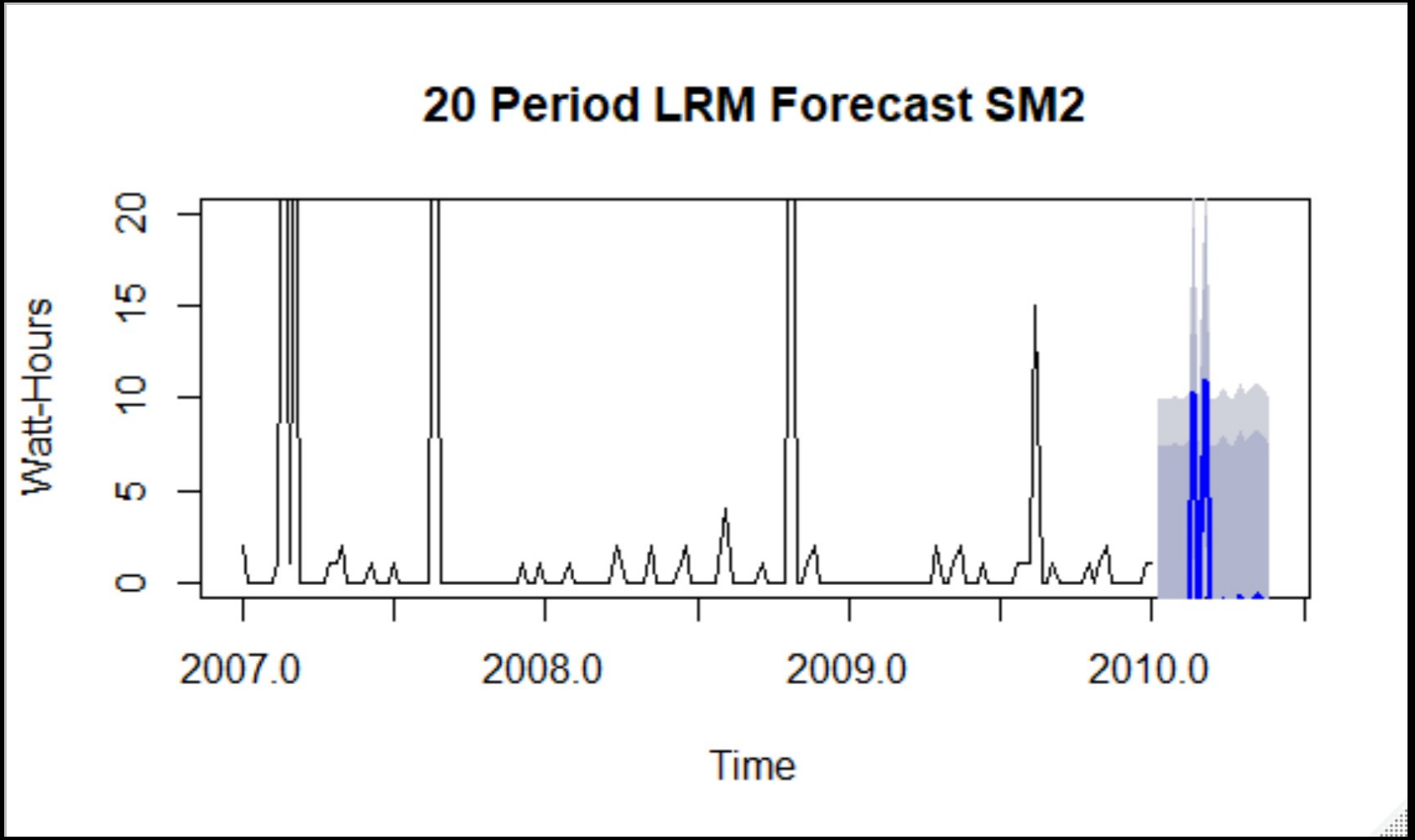
3 SLICES OF THE DATA



Time Series Consumption
Weekly 2007 - 2009
Sub Meter 1
20 Week Forecast

STEP 3: LINEAR REGRESSION FORECASTS

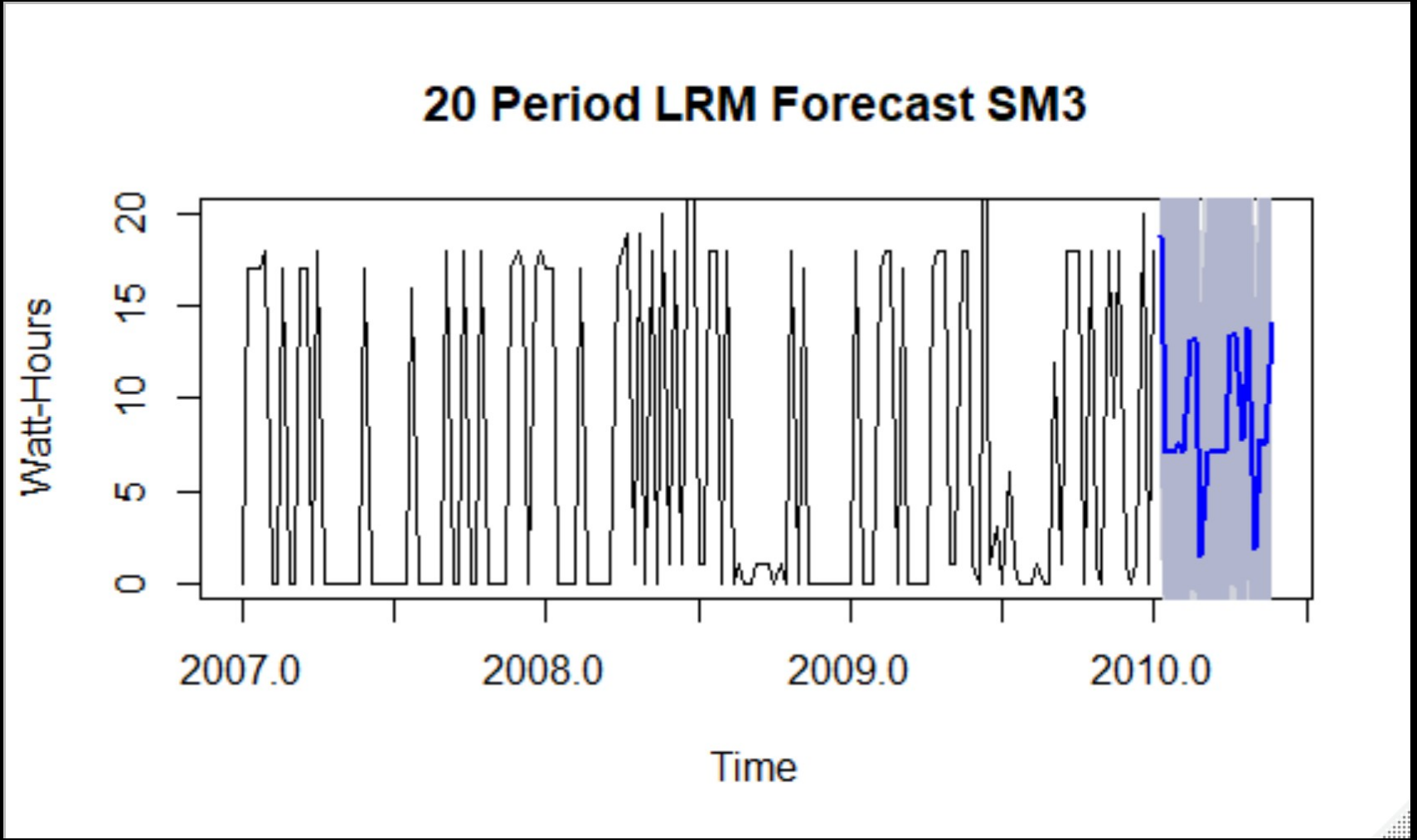
3 SLICES OF THE DATA



Time Series Consumption
Weekly 2007 - 2009
Sub Meter 2
20 Week Forecast

STEP 3: LINEAR REGRESSION FORECASTS

3 SLICES OF THE DATA



Time Series Consumption
Weekly 2007 - 2009
Sub Meter 3
20 Week Forecast

STEP 3: LINEAR REGRESSION FORECASTS

3 SLICES OF THE DATA

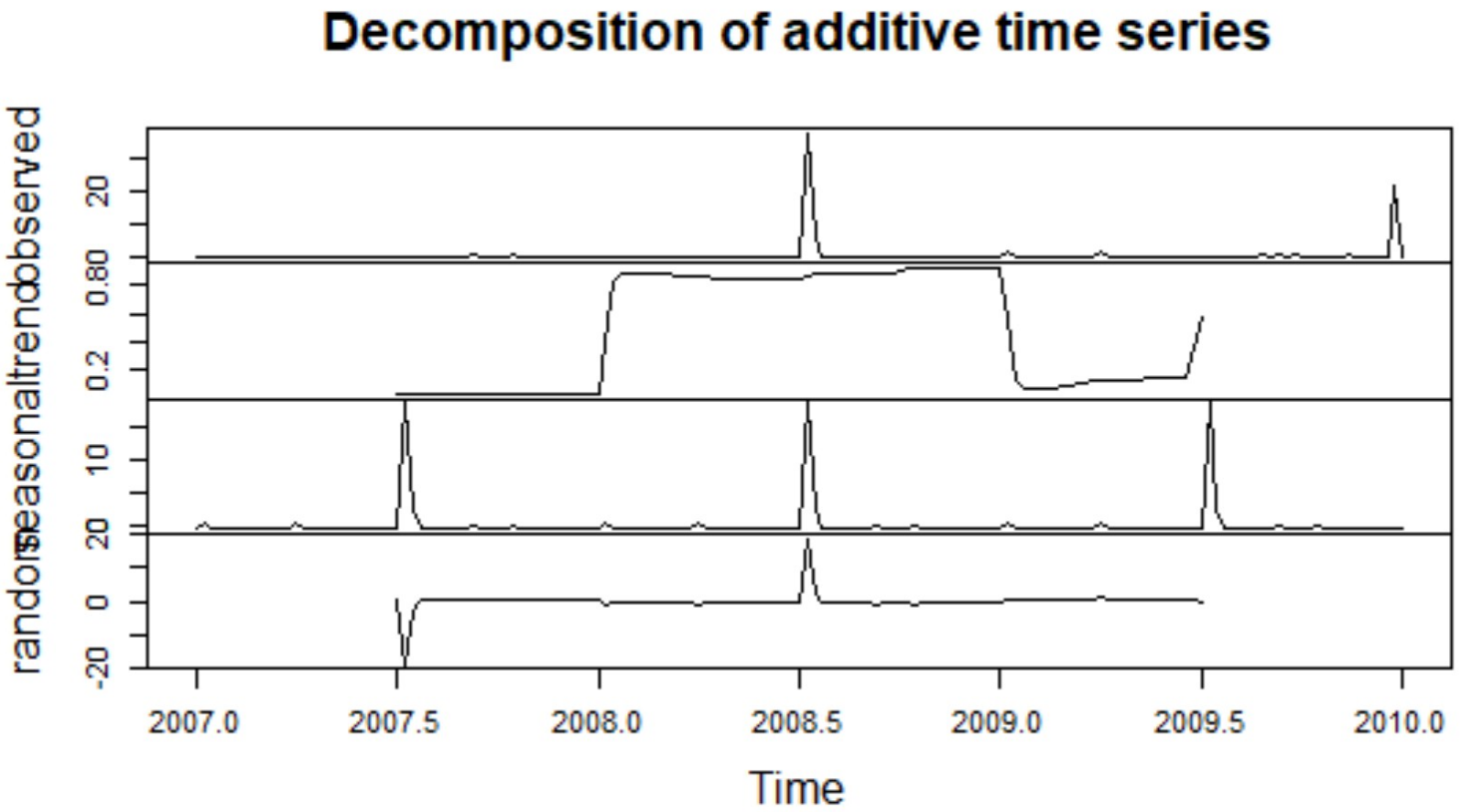
Model	R2	RMSE
TimeSeries Forecast for Sub Meter 1	0.3248	2.889331
TimeSeries Forecast for Sub Meter 2	0.3152	4.867134
TimeSeries Forecast for Sub Meter 3	0.263	7.362187

Lower Values of RMSE indicate a better fit

R2 measures how close the data is to the fitted linear regression line. 100% is a perfect fit.

STEP 4: DECOMPOSITIONS

3 SLICES OF THE DATA

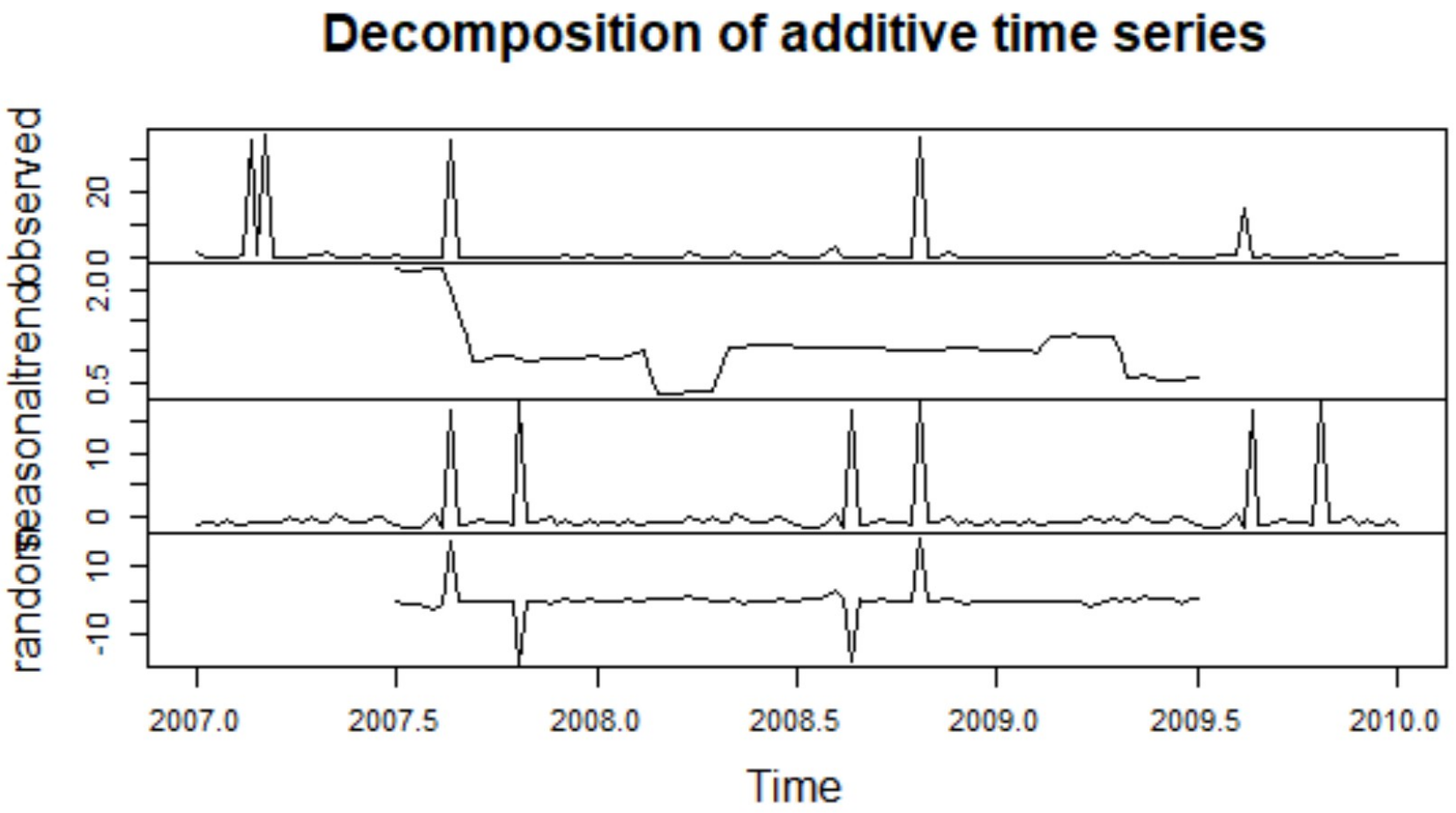


Observed
Trend
Seasonal
Random

Mondays 8pm
2007-2010
Sub Meter 1

STEP 4: DECOMPOSITIONS

3 SLICES OF THE DATA

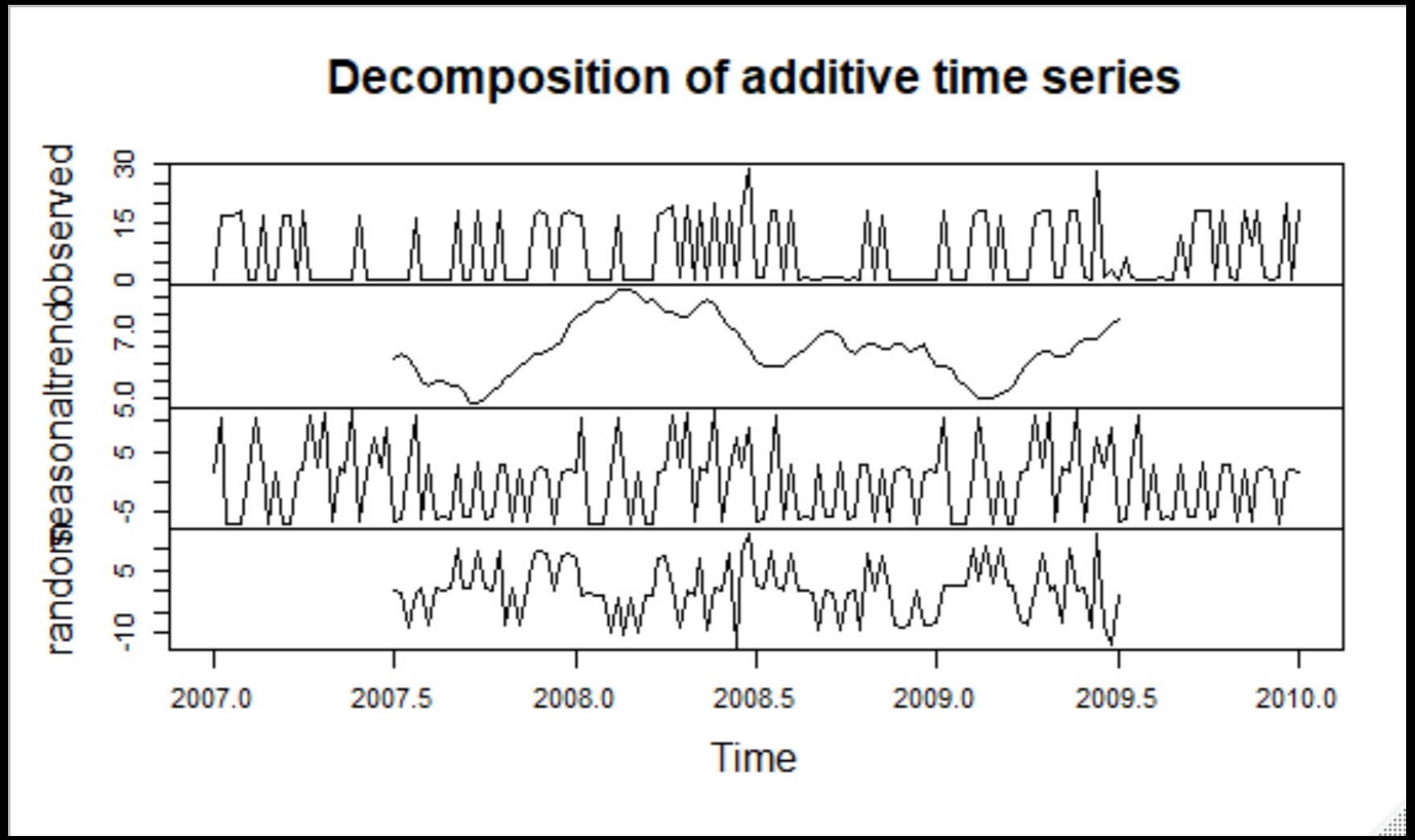


Observed
Trend
Seasonal
Random

Mondays 8pm
2007-2010
Sub Meter 2

STEP 4: DECOMPOSITIONS

3 SLICES OF THE DATA



Observed
Trend
Seasonal
Random

Mondays 8pm
2007-2010
Sub Meter 3

STEP 4: DECOMPOSITIONS

3 SLICES OF THE DATA

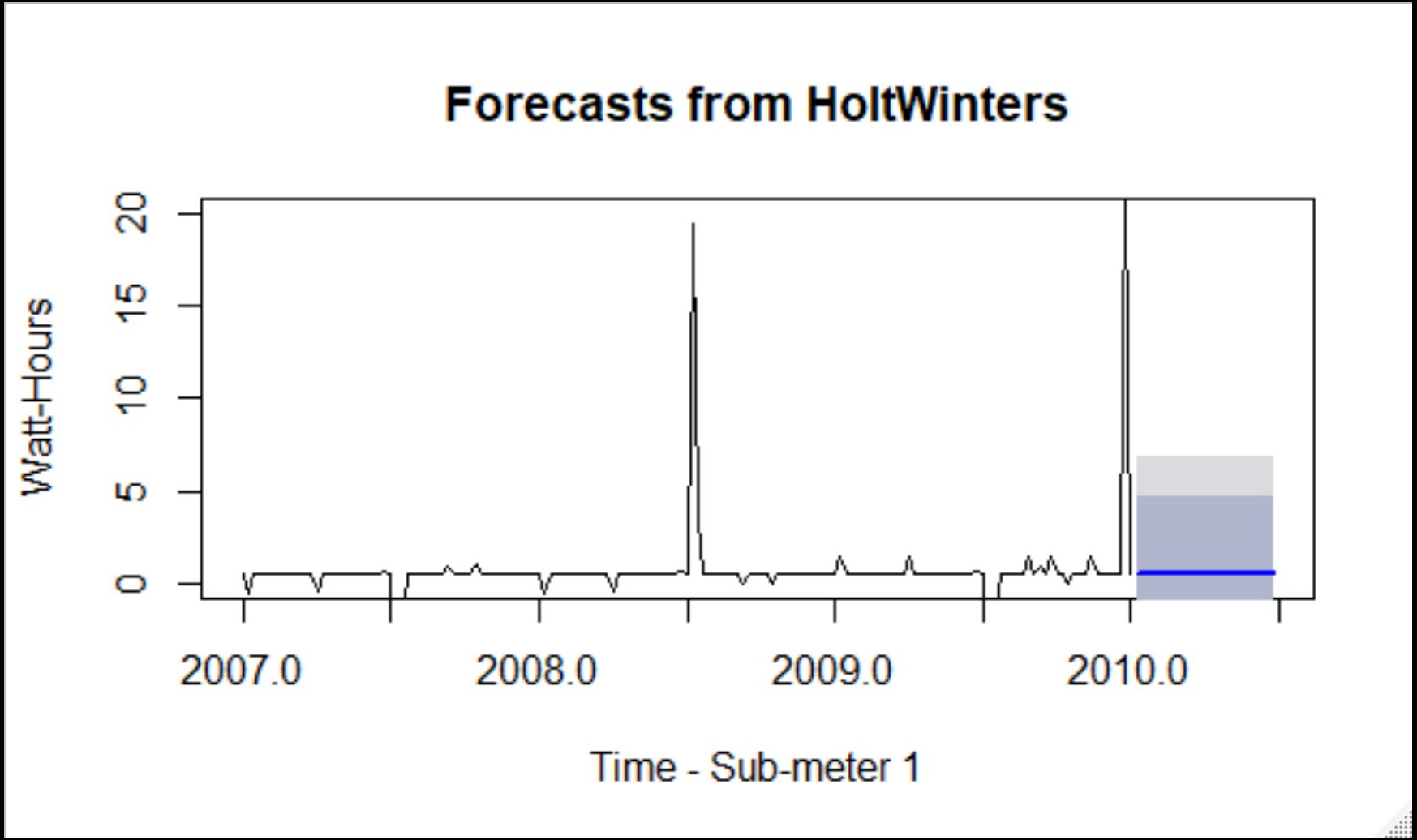
	SM1 Length	SM2 Length	SM3 Length
X	• 157	• 157	• 157
Seasonal	• 157	• 157	• 157
Trend	• 157	• 157	• 157
Random	• 157	• 157	• 157

Results of Summary on decomposed time series

The information provided from summary was useless

STEP 5: HOLT WINTERS FORECASTS

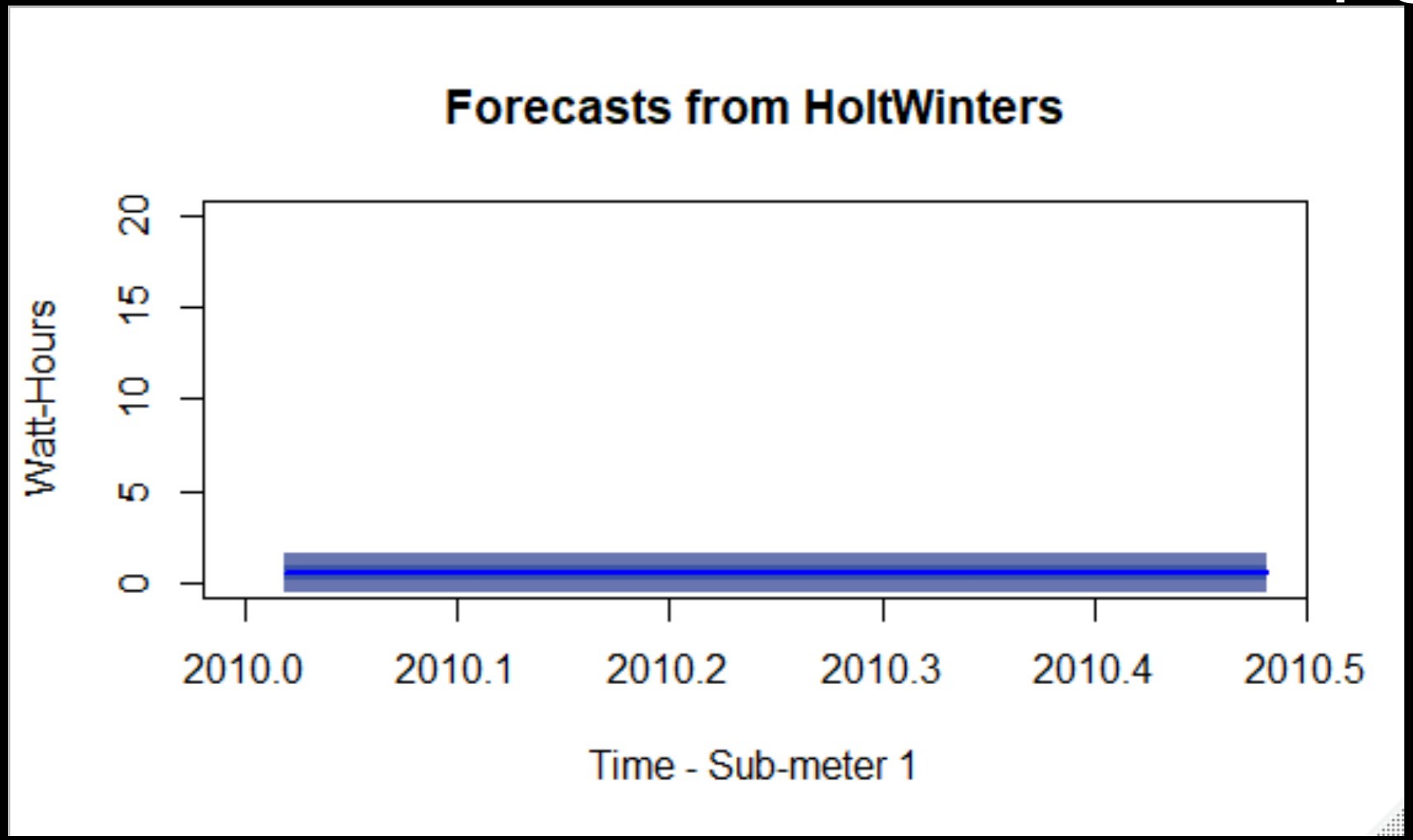
3 SLICES OF THE DATA



Seasonal Values Removed from Time Series for Forecasting

Mondays 8pm
2007-2010
Sub Meter 1

STEP 5: HOLT WINTERS FORECASTS

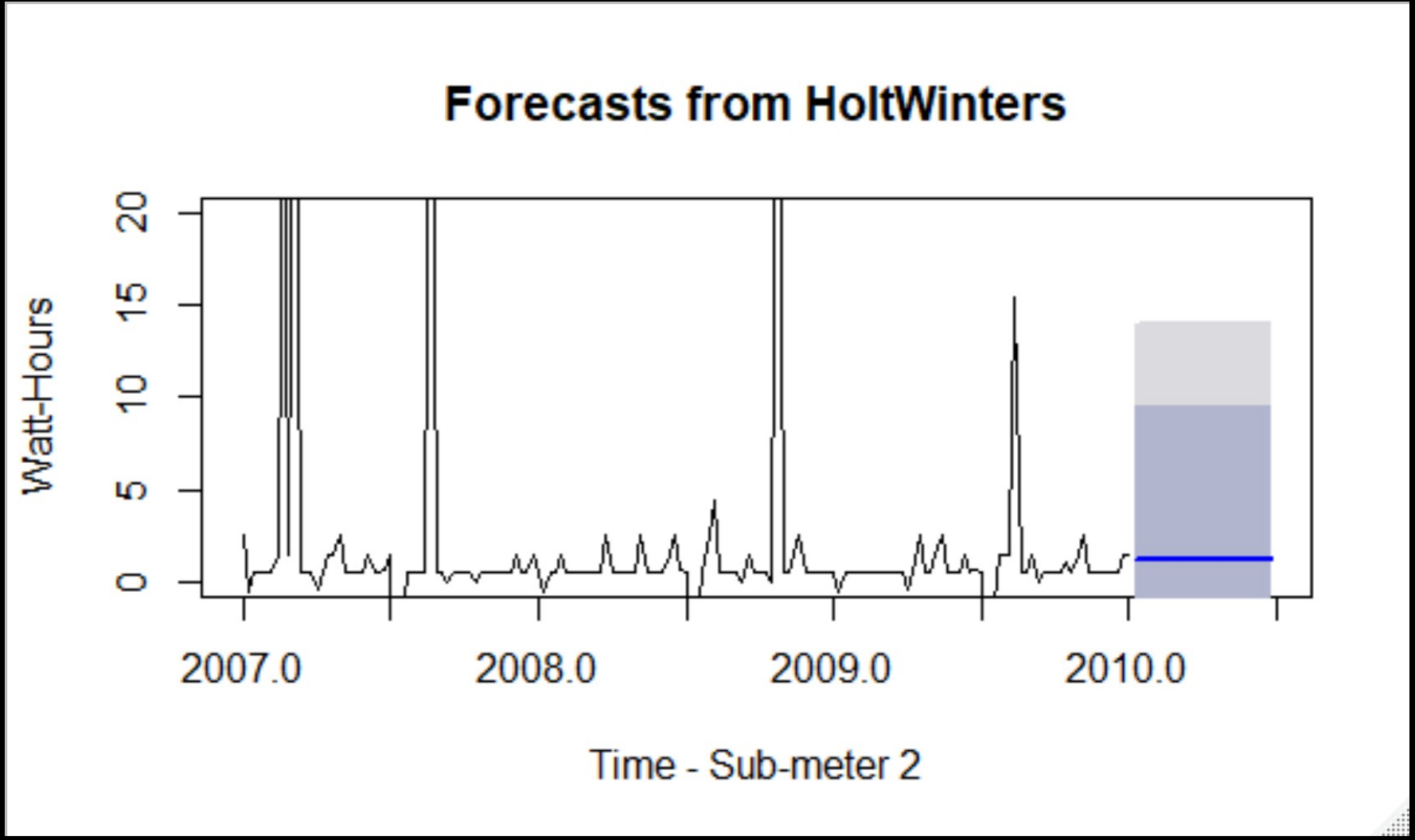


3 SLICES OF THE DATA

Seasonal Values Removed from Time Series for Forecasting – Only the Forecast Values Presented

Mondays 8pm
2007-2010
Sub Meter 1

STEP 5: HOLT WINTERS FORECASTS

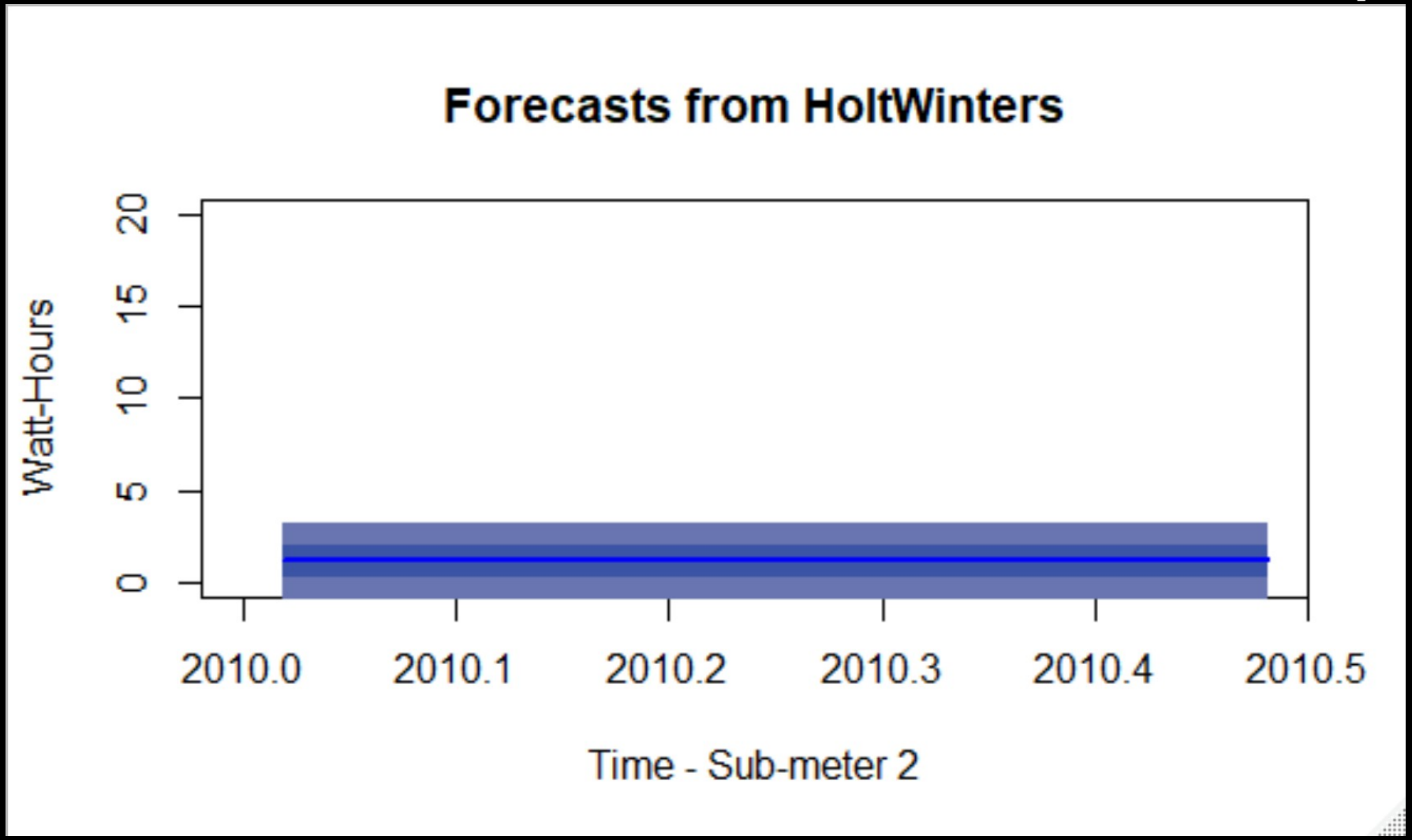


3 SLICES OF THE DATA

Seasonal Values Removed from Time Series for Forecasting

Mondays 8pm
2007-2010
Sub Meter 2

STEP 5: HOLT WINTERS FORECASTS

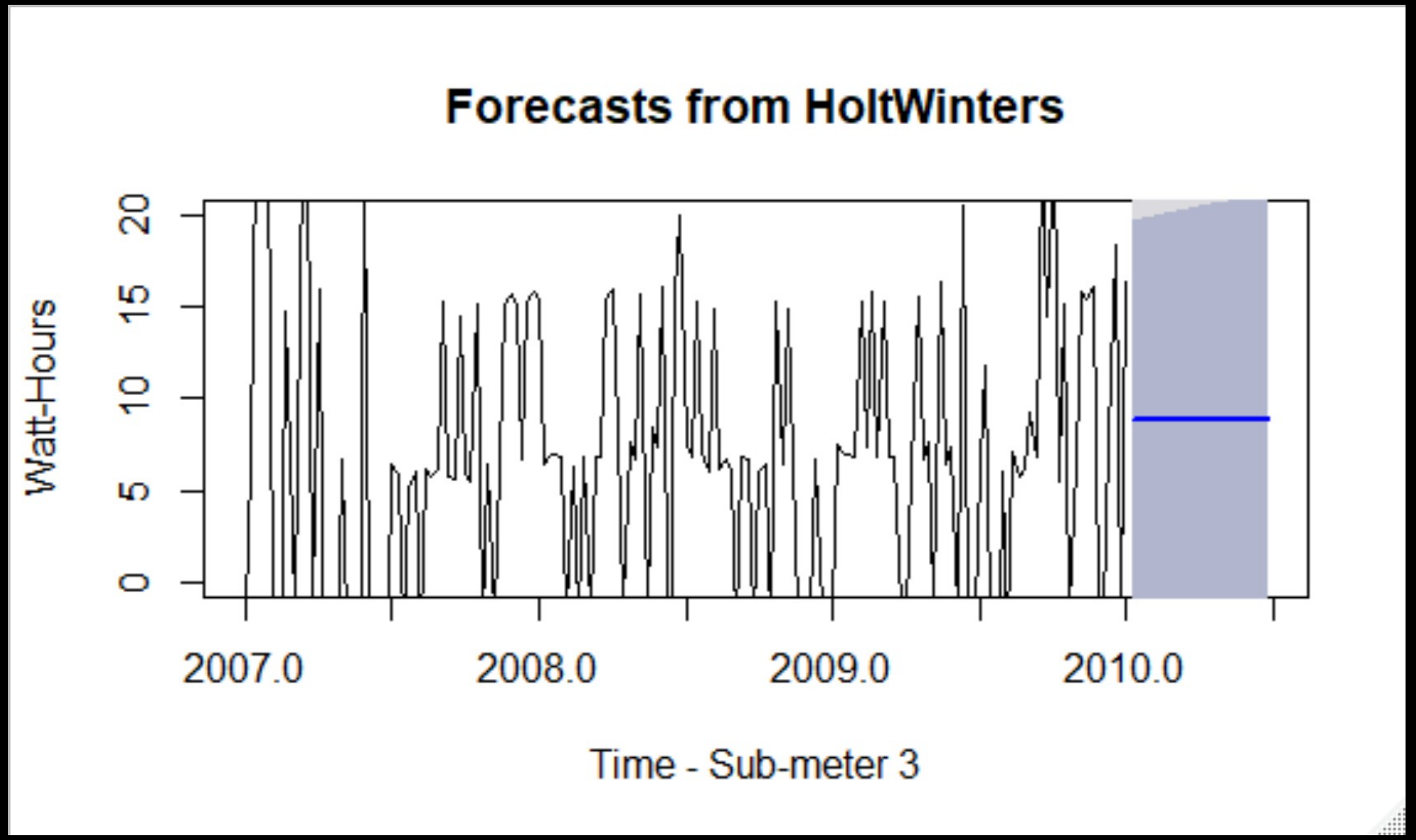


3 SLICES OF THE DATA

Seasonal Values Removed from Time Series for Forecasting – Only the Forecast Values Presented

Mondays 8pm
2007-2010
Sub Meter 2

STEP 5: HOLT WINTERS FORECASTS



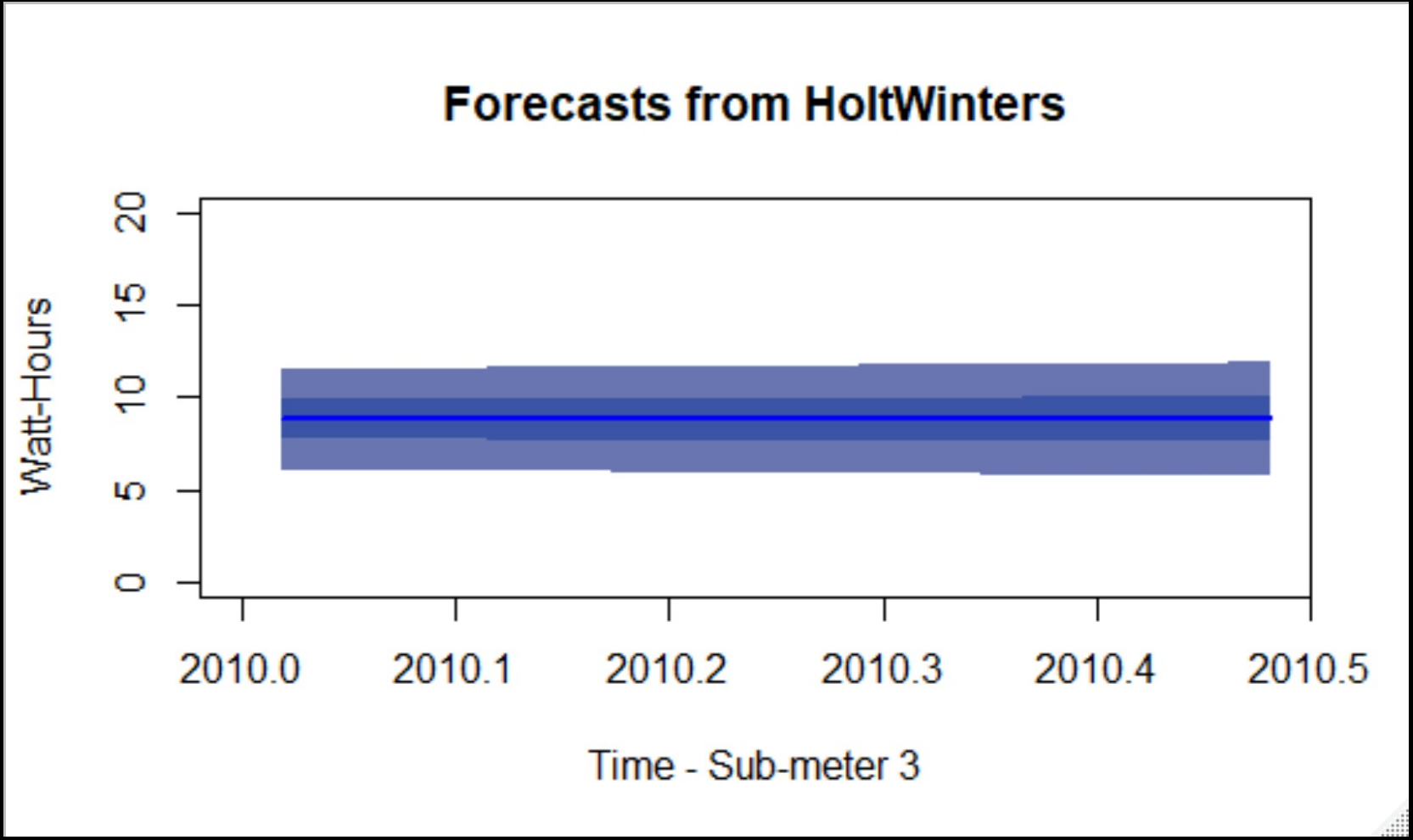
3 SLICES OF THE DATA

Seasonal Values Removed from Time Series for Forecasting

Mondays 8pm
2007-2010
Sub Meter 3

STEP 5: HOLT WINTERS FORECASTS

3 SLICES OF THE DATA



Seasonal Values Removed from Time Series for Forecasting – Only the Forecast Values Presented

Mondays 8pm
2007-2010
Sub Meter 3

USEFUL INSIGHT

- Based on a small subset observed, the kitchen is primarily utilized in the evening, probably preparing dinner and cleaning up.
- The utilities (sample from Jan so likely heater) run early morning to mid day and again in the evening. So, likely the break observed on one day is the owner leaving the home and turning the heater off while gone
- The laundry usage was low and fairly evenly spreadout, suggesting either generally low usage
- Looking at one day at 8p across 3 years, usage was generally low and steady with the exception of two spikes, one mid-2008 and one end of 2009. These spike could possibly represent high power consumption for failing kitchen components taking longer to complete their cycles. It could also mean a huge party was thrown in the house and a lot of dishes needed to be cleaned. This is an opportunity for more data relative to each component on a sub meter.
- Similar to above, looking at the laundry on Monday's a 8p, there are spike in usage that could represent occurrences where the owner decided to do laundry during our observation period. Otherwise, there is low usage at this time of day.
- The utilities (heater and air) are the largest consumers of power over the three years observed. They are consistently in use on Monday's at 8p. It is observed there are a couple of spikes during the summer. This is likely due to hotter than average weather outside requiring more energy to cool house.

USEFUL INSIGHT

- Nothing too surprising on the forecast of usage at first. The kitchen and laundry are general low and consistent. The one interesting thing to maybe look more closely at is the forecast usage for the utilities (heater and air). The forecast suggests lower energy consumption in the period forecast. Weather data, or predicted weather data, might be interesting data to look at as an input to our model (think Farmer's Almanac??)
- Looking at the error statistics, it could be inferred our estimated usage could be close to 30% accurate. So, to make this information useful we probably should look at more data, or suggest more detailed monitoring equipment by component on each submeter.
- Seasonal observations suggest an increase in power consumption on the whole during the winter months and summer months. If we were to look at historical weather information, we might find this correlates to colder weather in the winter requiring more heating and warmer weather in the summer requiring more cooling.
- Interestingly it looks like the owner tended to use more power in the kitchen over the course of 2008. It would be interesting to see if the power consumption is a particular component in the kitchen that spiked usage, and then to know if any components were replaced in the start of 2009 leading to more efficient energy consumption

USEFUL INSIGHT

- It is very interesting that smoothing the data by removing seasonality prior to predicting usage by submeter resulted in predictions that are very constant. This suggests that seasonality is actually a very important part of the puzzle in helping home owners become more efficient in use of power.

SUMMARY STATEMENT

The goal of this project has been met.

We arrived at a handful of interesting recommendations for business to consider to help home owners consume less power, or be more effective at metering usage.

We also arrived at a handful of interesting recommendations for business to consider in terms of new data to help with better predictive analytics, or product enhancements to collect better usage data.

5 BUSINESS RECOMMENDATIONS

1) Collecting weather data during the observed periods would be really interesting to map energy consumption relative to weather events. It could confirm or deny certain seasonal patterns, help predict usage better during certain periods of the year, and help the owner plan for higher or lower bills based on weather events.
2) Collecting individual component usage data (washer, dryer, air conditioner, etc) would provide insight into times of day used, energy used by that specific component, and others such that we could infer things like unusual increased power consumption indicates a particular component may require service, or confirm what we believe to be true like the air conditioner is used in the summer and heater in the winter, or times of data kitchen appliances tend to be used.
3) Install individual appliance power meters on each sub meter, or create more sub meters to get a more granular view of component power usage
4) Given some of the power usage patterns observed, it could be useful to provide programmable power meters for major appliances and components like the air conditioner and heater, the two largest consumers of power in the house, to try to optimize cooling and heating, based on the weather outside and the likelihood of home occupancy at any particular time of day.
5) Work with homeowners to validate whether or not appliances and components were replaced during this period of observation. That could help correlate abnormalities in usage with defect equipment for future predictions

LESSONS LEARNED

- It’s important to sit down and write out a good description of the subset before trying to write any code so I understand what it is I am really trying to do.
- There is a lot of power in one line of R code. It’s important to clearly understand what each option means and how it impacts the outcome of the function calls.
- It is really important to follow the pipeline steps in order. Getting something out of order makes it really difficult to debug.
- Good variable names in pipelines as big as this one are critical to ensuring I don’t accidentally overwrite important often used variables with data I need for processing.
- The analysis part of almost more art than science. There is a lot of additional follow up that comes from initial assessments like this. But it’s really interesting to see how this effort at least starts to point us in productive directions for further exploration rather than random meandering through a spreadsheet.

QUESTIONS

-