Analysis on Energy Consumption of Appliances in a Low-Energy House

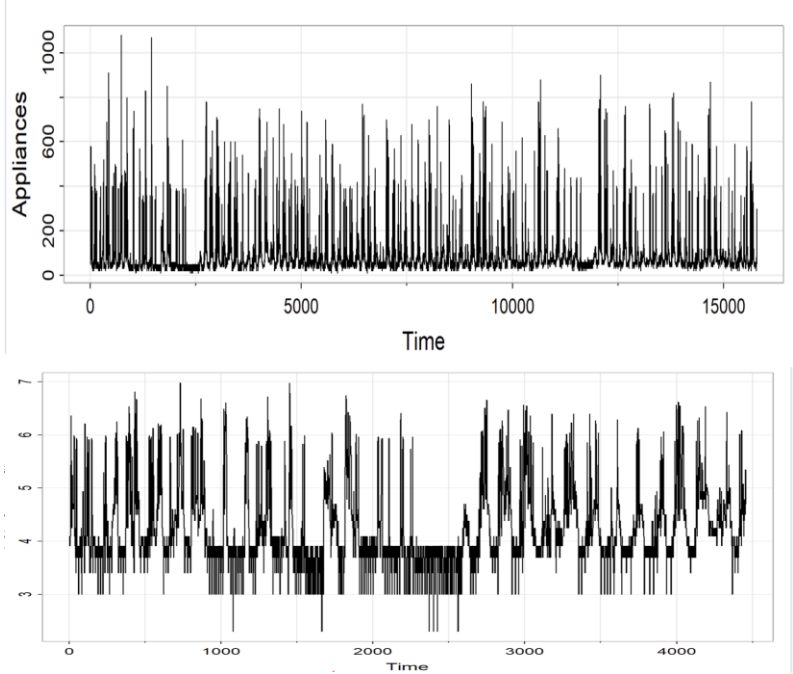
Introduction

- Energy consumption of low energy house(1/11 17:00 ~ 5/27 18:00, every 10 minutes)
- 19,735 observations, 29 input variables
- Temperature, humidity of 9 rooms
- Outside weather conditions (temperature, humidity, windspeed)
- GOAL : What factors would be influential in energy consumption, what kind of patterns may exist in energy consumption?

Variables	Description
date time	Year-month-day hour:minute:second
Appliances	Energy use in Wh
lights	Energy use of light fixtures in the house in Wh
T1	Temperature in kitchen area, in Celsius
RH_1	Humidity in kitchen area, in %
T2	Temperature in living room area, in Celsius
RH_2	Humidity in living room area, in %
T3	Temperature in laundry room area
RH_3	Humidity in laundry room area, in %
T4	Temperature in office room, in Celsius
RH_4	Humidity in office room, in %
T5	Temperature in bathroom, in Celsius
RH_5	Humidity in bathroom, in %
T6	Temperature outside the building (north side), in Celsius
RH_6	Humidity outside the building (north side), in %
T7	Temperature in ironing room, in Celsius
RH_7	Humidity in ironing room, in %
T8	Temperature in teenager room 2, in Celsius
RH_8	Humidity in teenager room 2, in %
T9	Temperature in parents room, in Celsius
RH_9	Humidity in parents room, in %
To	Temperature outside (from Chièvres weather station), in Celsius
Pressure	Pressure (from Chièvres weather station), in mm Hg
RH_out	Humidity outside (from Chièvres weather station), in %
Windspeed	Windspeed (from Chièvres weather station), in m/s
Visibility	Visibility (from Chièvres weather station), in km
Tdewpoint	Tdewpoint (from Chièvres weather station), °C
rv1	Random variable 1, nondimensional
rv2	Random variable 2, nondimensional

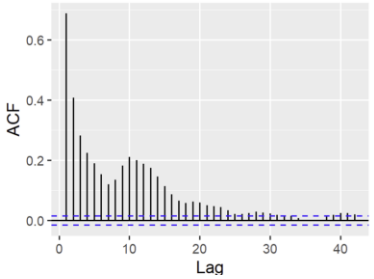
Method A - Regression Analysis

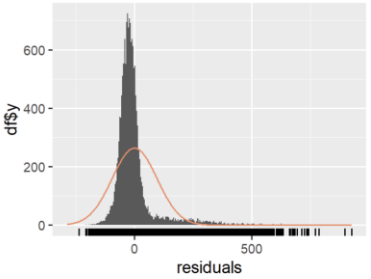
- No missing values
- No outlier removal or scaling performed
- Non-stationary ⟹ Log transformation



Method A - Regression Model 1

- Linear regression with every input variable
- Eight insignificant variables(T5, RH_5, RH_6, etc.)
- ACF plot : many values above the blue line
- Ljung-Box plot : H0(auto correlations are zero) rejected






```
Residual standard error: 95.66 on 15761 degrees of freedom
Multiple R-squared:  0.1745,    Adjusted R-squared:  0.1731
F-statistic: 128.1 on 26 and 15761 DF,  p-value: < 2.2e-16
```

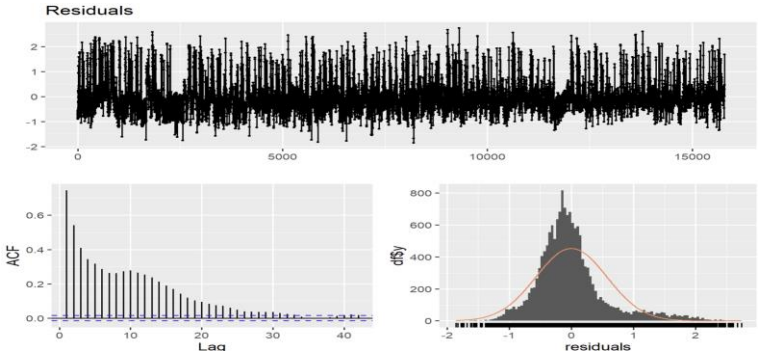
```
        Ljung-Box test

data:  Residuals
Q* = 14836, df = 10, p-value < 2.2e-16
```

Method A - Regression Model 2

- 24 input variables, including time^2 value
- Log transformation on the output variable(energy consumption)
- Increase R^2 from 0.1745 to 0.2882
- ACF plot : many values above the blue line
- Ljung-Box plot : H0(auto correlations are zero) rejected
- Linear assumption may not hold with other variables
- Lower AIC, BIC, higher R^2 : Model 2 is better

```
> AIC(reg1)
[1] 188845.2
> AIC(reg2)
[1] 27158.14
```

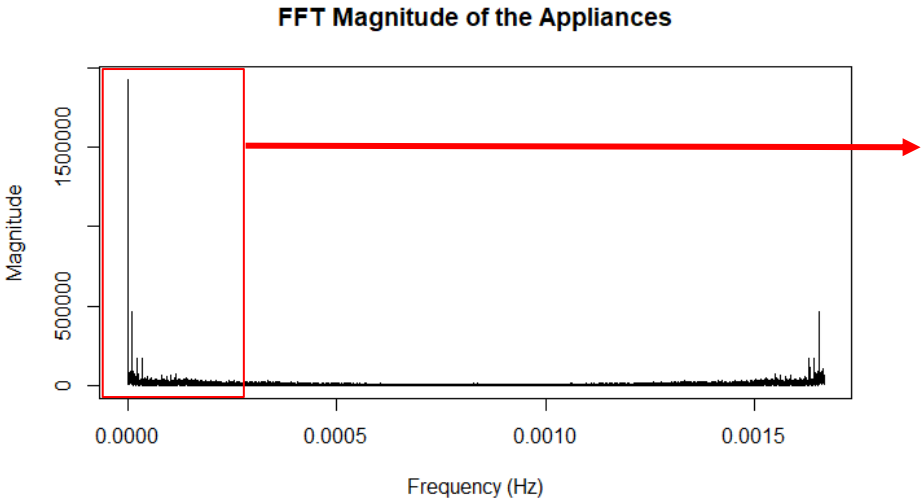
```
> BIC(reg1)
[1] 189059.9
> BIC(reg2)
[1] 27357.48
```

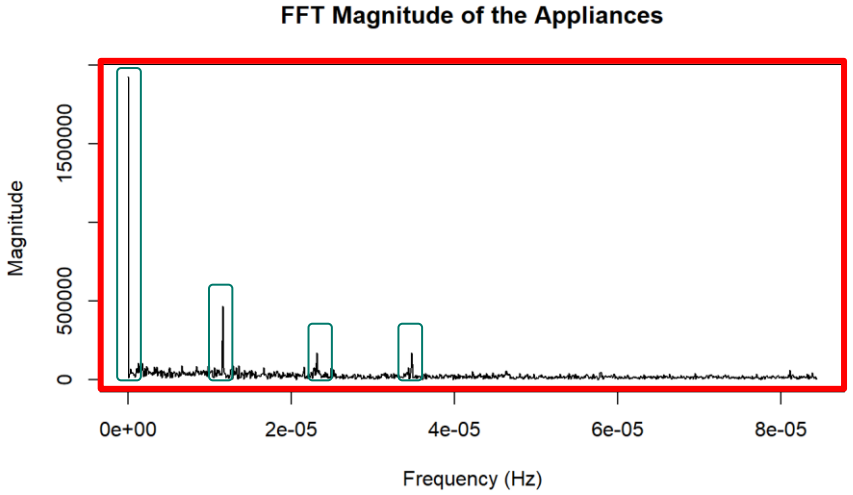



```
Residual standard error: 0.5714 on 15763 degrees of freedom
Multiple R-squared:  0.2882,    Adjusted R-squared:  0.2871
F-statistic:   266 on 24 and 15763 DF,  p-value: < 2.2e-16
```

```
> predict(reg2, df[15788:15793,])
   15788    15789    15790    15791    15792    15793
4.554286 4.543875 4.558177 4.564981 4.748895 4.613458
> log(df[15788:15793,2])
[1] 5.703782 5.913503 6.380123 5.768321 5.736572 5.560682
```

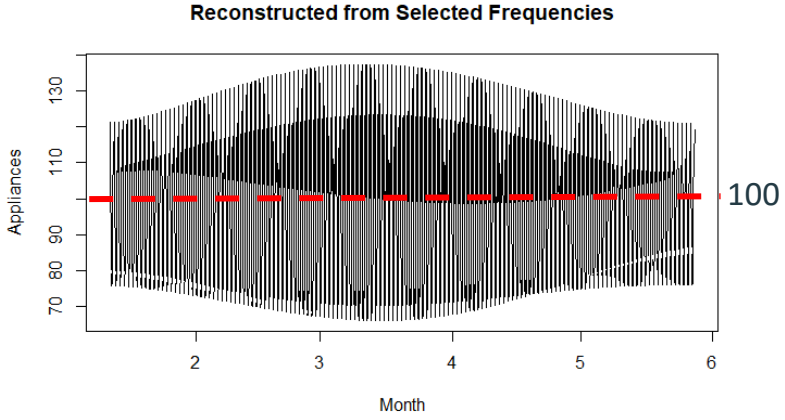
Method B - Spectral Analysis

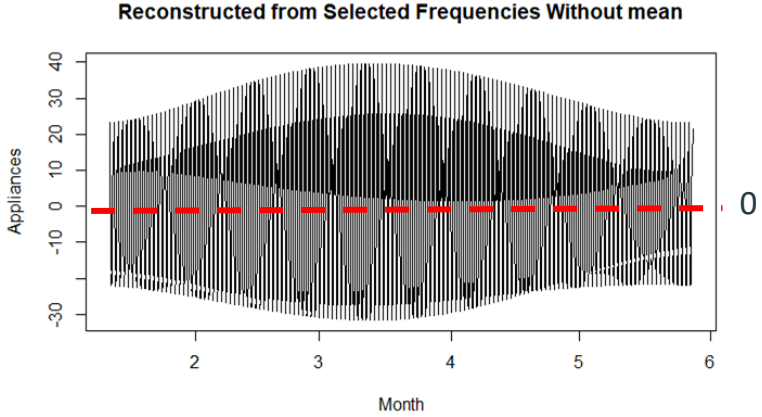


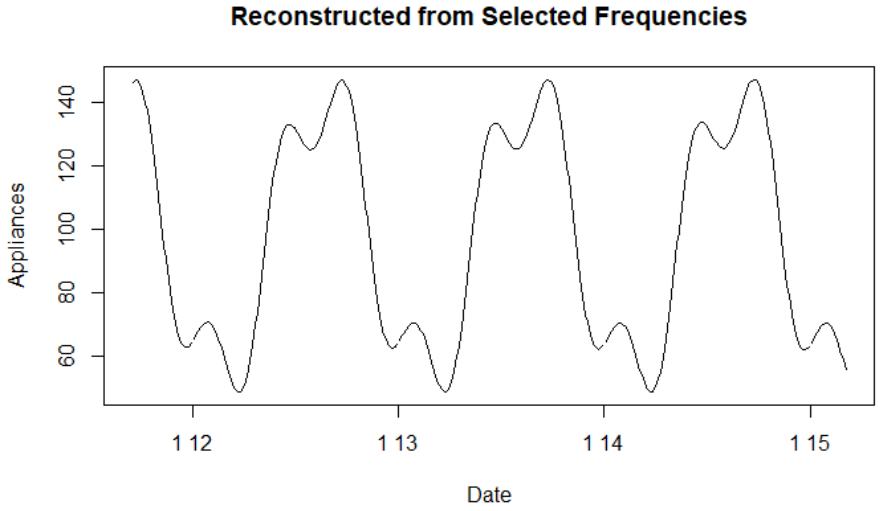


frequencies <dbl>	magnitudes <dbl>
0.000000e+00	1928010.0
1.156997e-05	466585.5
3.470991e-05	171567.9
2.313994e-05	170487.8

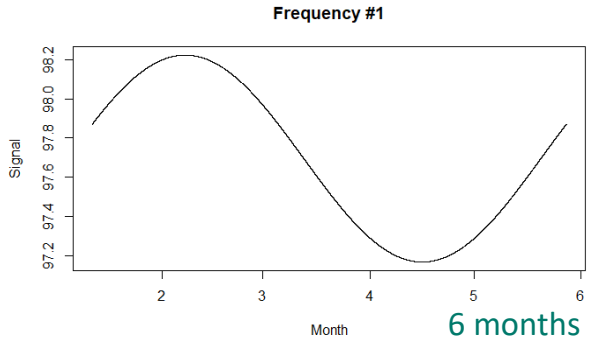
Results - Spectral Analysis

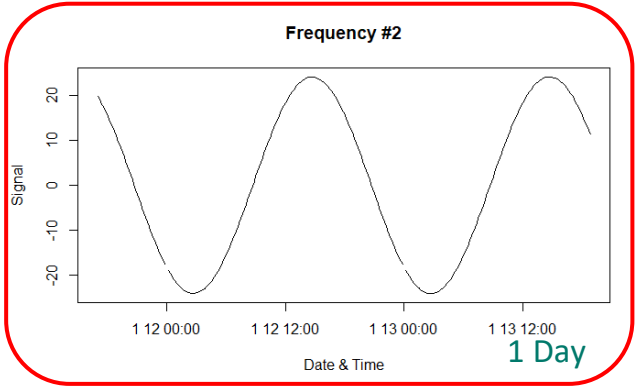


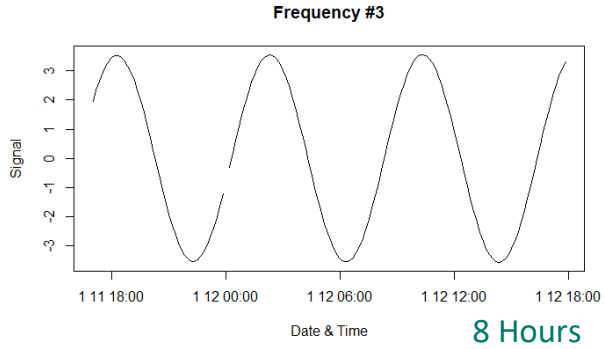


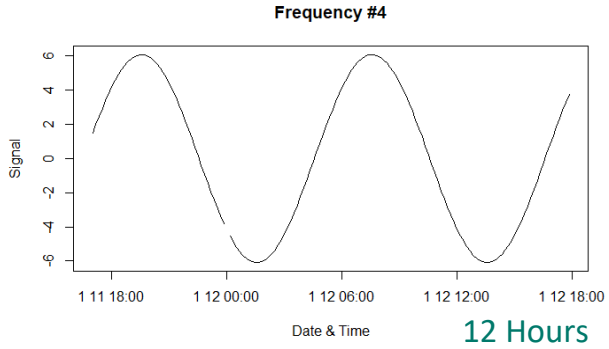


Results - Spectral Analysis









The result of spectral analysis aligns with our expectation.

Except for the mean (Frequency #1), predominant cycle of the data is 1 Day.

Conclusion - How to improve R^2?

- **Limitations**
  - Could not get white noise
  - Low R^2 of regression model
- **Further analysis**
  - Use the filtered time series
  - Complex models(SVM, Randomforest, etc)