# Time series

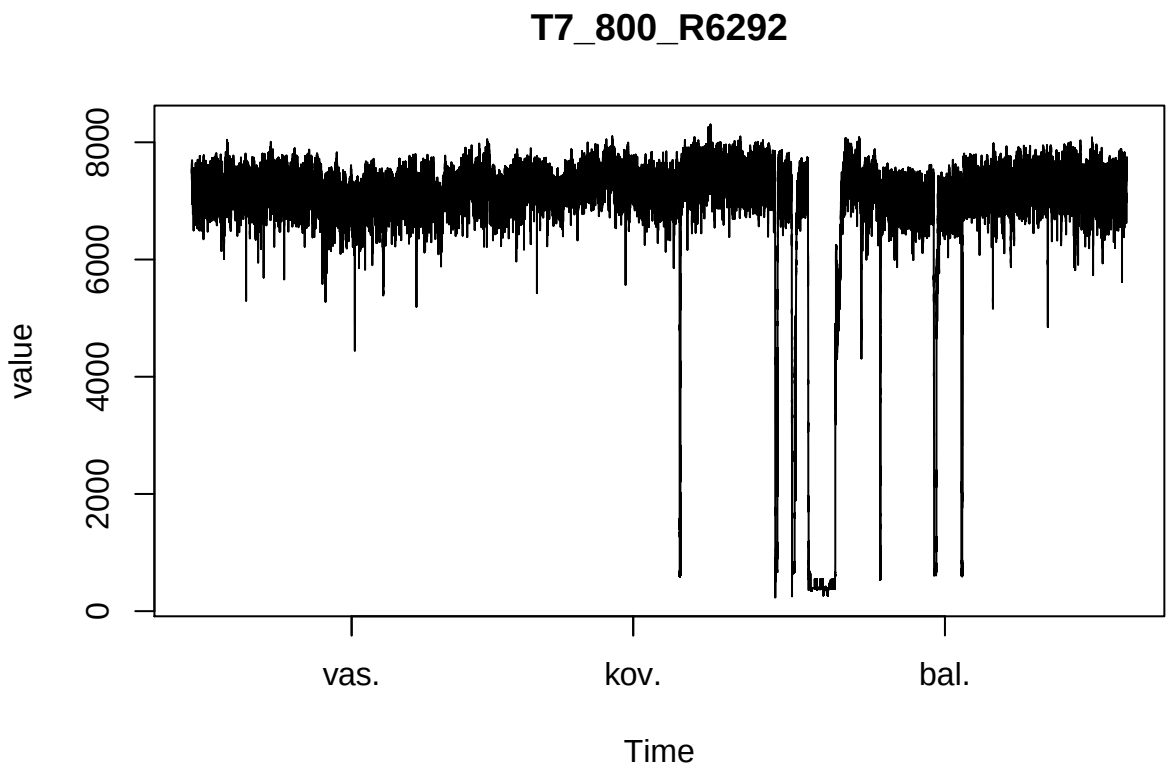
Robert Tarasevic

2020-04-06

## Contents

```r
library("tidyverse")
library("caret")
library("dplyr")
library("rhdf5")
library("zoo")
library("knitr")
library("kableExtra")
library("mgcv")
```

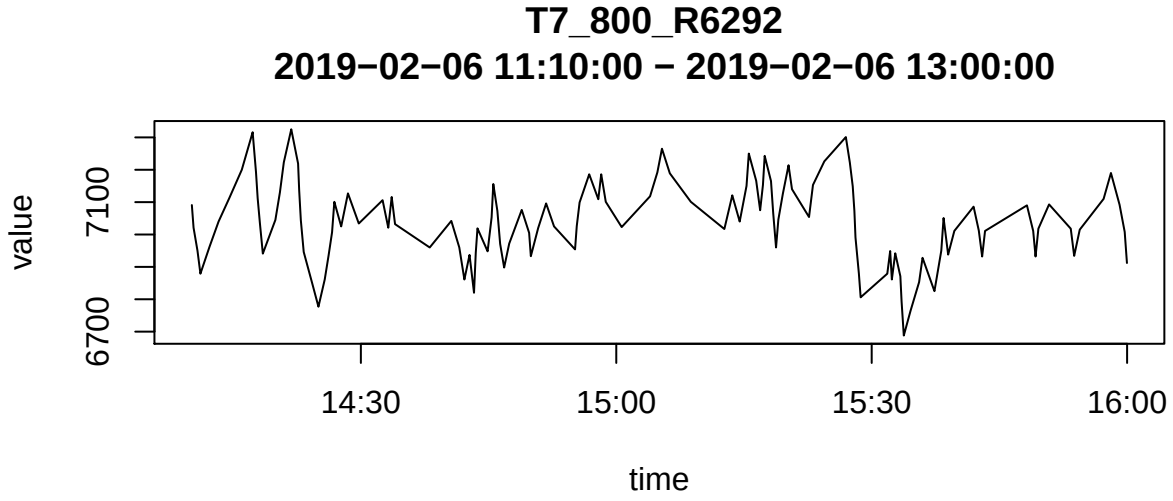
### Load data




```
## [1] "Only one graph showed because same sensor chosen"
```

### Chosen time period for signals

```
## First signal subset:  2019-02-06 11:10:00 - 2019-02-06 13:00:00
```




```
## Second signal subset:  2019-02-06 16:00:00 - 2019-02-06 18:00:00
```

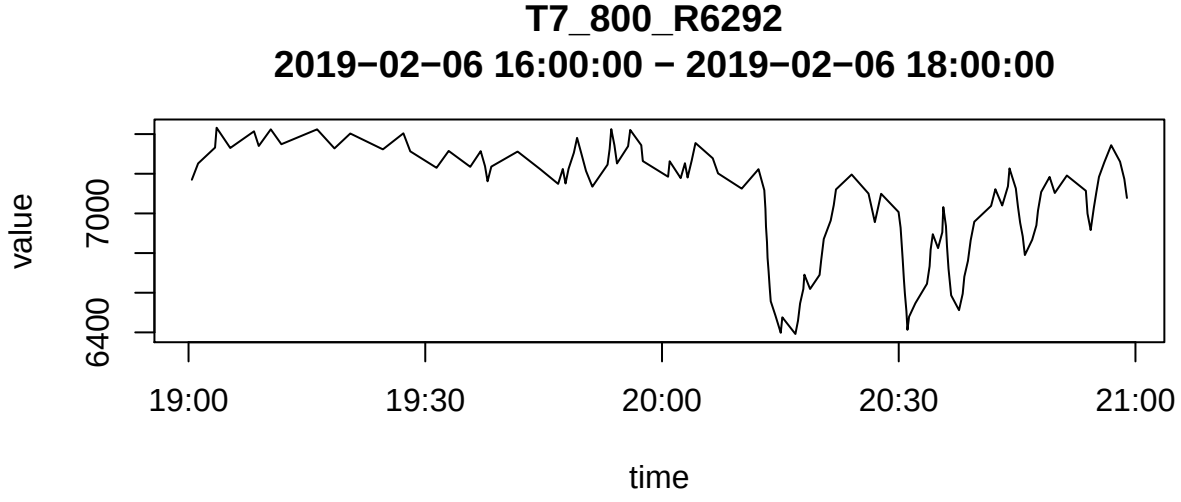


Table 1: Statistics table

| | length | t_mode | v_mode | sd | mean | kurt | skew | rms | entropy | cf | sf | cross_r |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| sub 1 | 126 | 12 | 6879 | 123.49 | 7037.98 | -0.20 | -0.12 | 7037.98 | 4.84 | 1.04 | 1.00 | 39 |
| sub 2 | 137 | 12 | 7344 | 279.48 | 7032.68 | -0.72 | -0.60 | 7032.68 | 4.92 | 1.06 | 1.00 | 10 |
| raw 1 | 115711 | 4 | 384 | 2055.88 | 6472.54 | 4.08 | -2.40 | 6472.54 | 11.58 | 1.22 | 1.05 | 698 |
| raw 2 | 115711 | 4 | 384 | 2055.88 | 6472.54 | 4.08 | -2.40 | 6472.54 | 11.58 | 1.22 | 1.05 | 698 |

## Statistics table

**Some statistics explained:**

**Signal length -** *Data points number in signal*
**Time mode -** *Mode of time change frequency*
**Value mode -** *Most frequent value*
**Sd -** *Standard deviation*
**Mean -** *Average value*
**n5, n25, n75, n95 -** *Quantile values*
**Kurtosis -** *Kurtosis quantifies the peak value of the PDF (positive kurtosis tells that there is lot of data in tails, negative - little data in tails*
**Skewness -** *Skewness quantifies the asymmetry behavior of vibration signal through its PDF*
**RMS -** *Root mean square value changes faster then mean*
**Entropy -** *Amount of uncertainty in an entire probability distribution*
**Crest factor (cf) -** *Ratio of the instantaneous peak amplitude of a waveform, to its root mean square RMS value*
**Shape factor (sf) -** *Shape factor is a value that is affected by an object's shape but is independent of its dimensions*
**Mean crossing -** *Mean value crossing count*

---

## Correlation

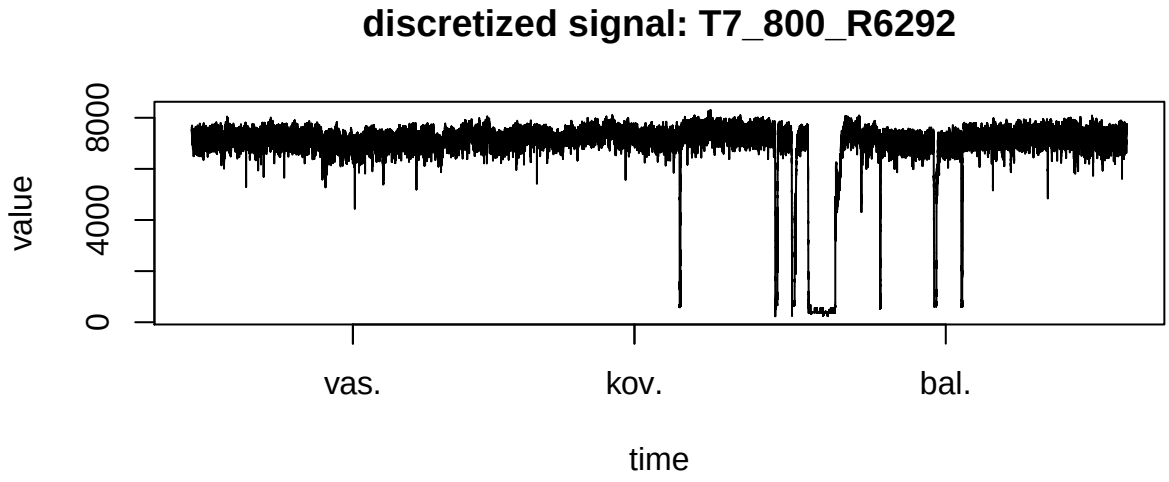
**Correlation between sub-sets**

```
## [1] -0.08195883
```

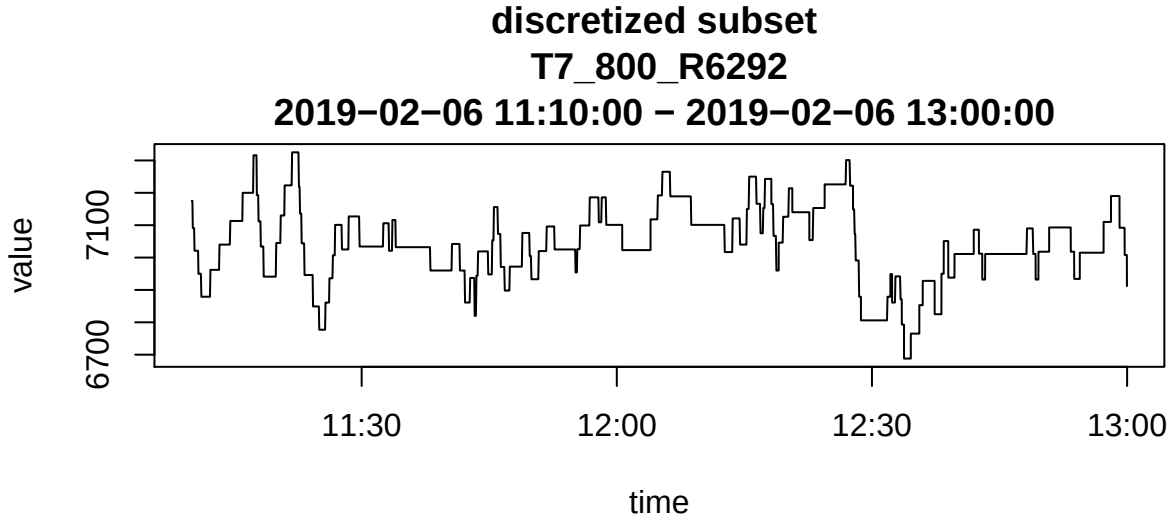
**Correlation between signals**

```
## [1] 1
```

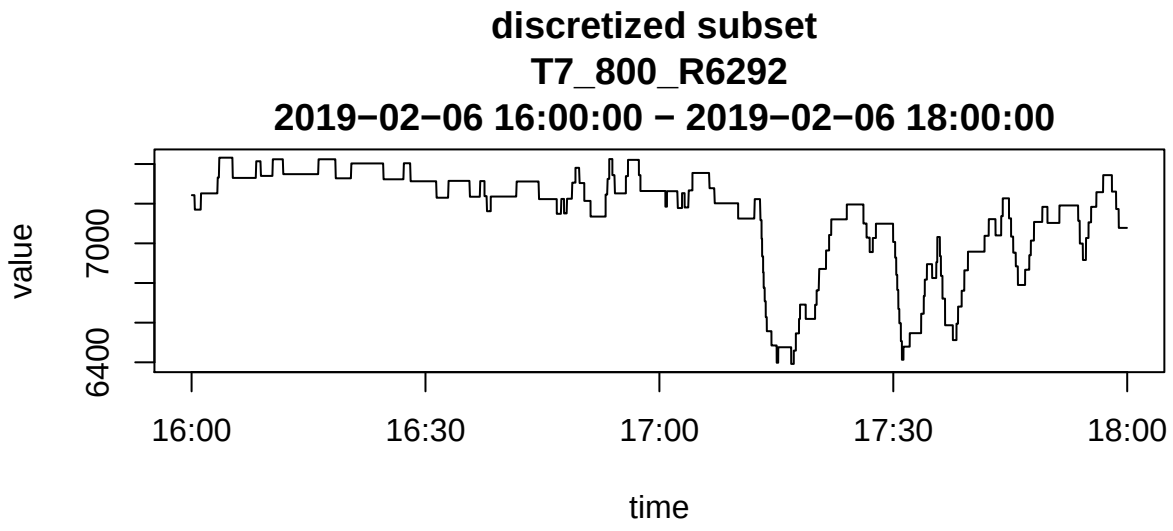
**Data discretization**




```
## [1] "Only one graph showed because same sensor chosen"
## First signal subset:  2019-02-06 11:10:00 - 2019-02-06 13:00:00
```

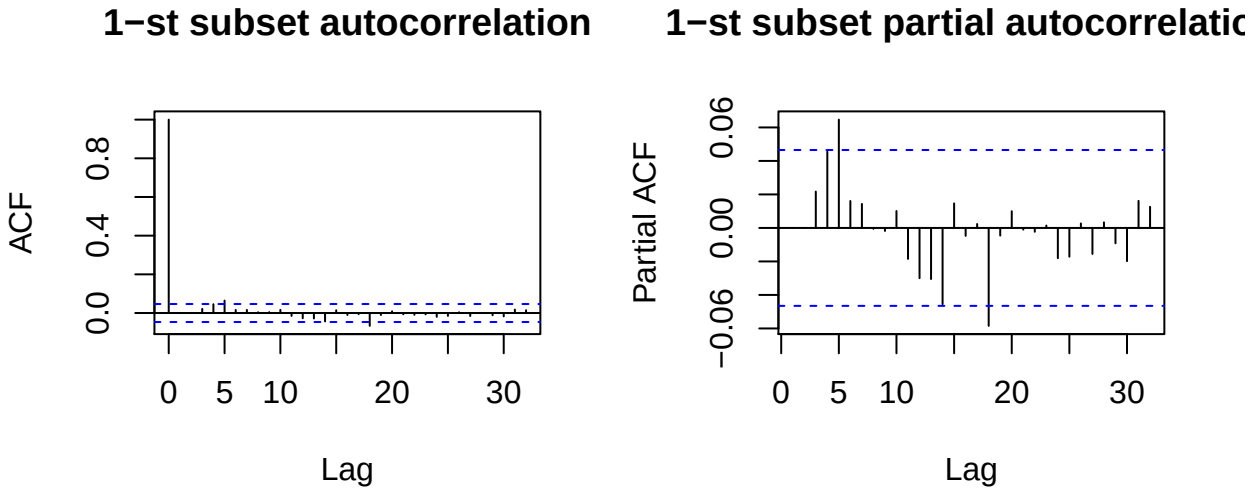



```
## Second signal subset:  2019-02-06 16:00:00 - 2019-02-06 18:00:00
```


discretized subset
T7_800_R6292
2019−02−06 16:00:00 − 2019−02−06 18:00:00
value
7000
6400
16:00
16:30
17:00
17:30
18:00
time

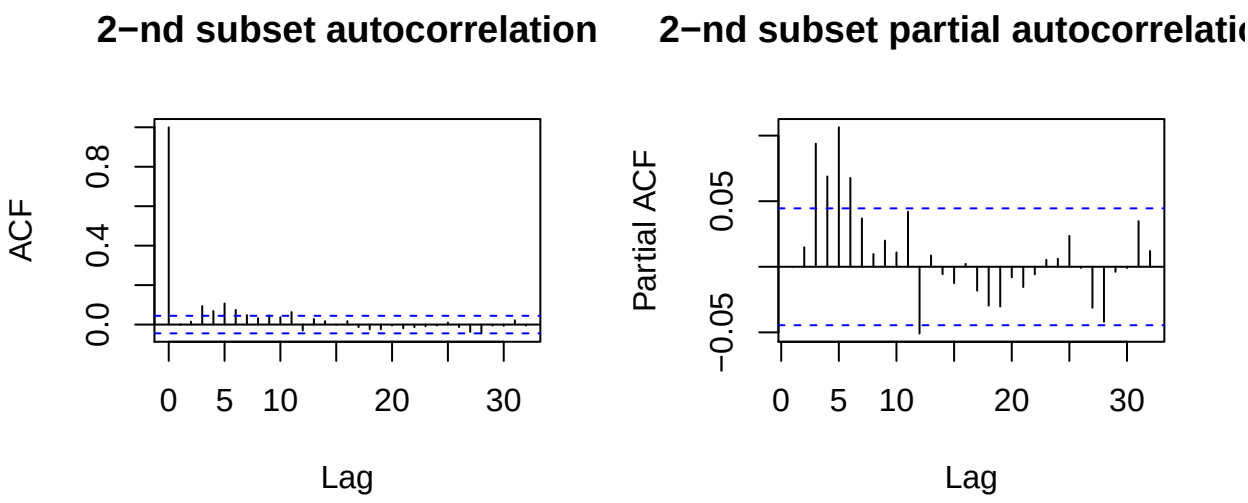
### Auto-correlation

**Auto correlation** is the correlation of a time series with its own past and future values
if data is strongly seasonal - peaks will apear with seasonality period
**Partial autocorrelation** at lag k is the correlation that results after removing the effect of any correlations due to the terms at shorter lags

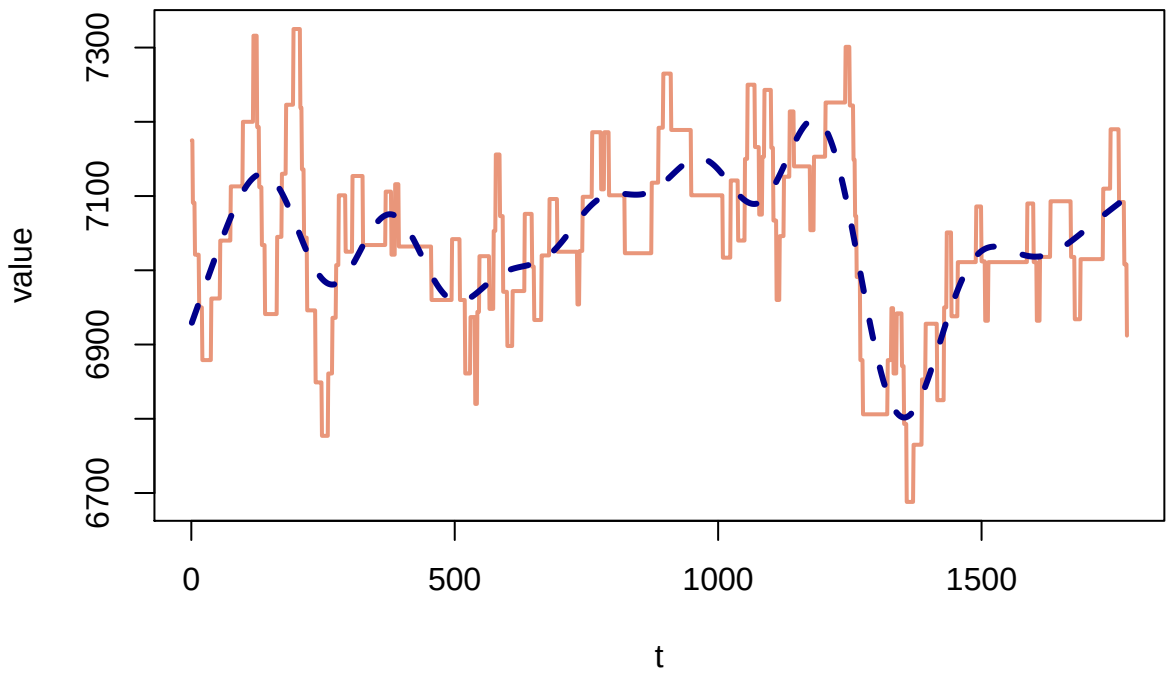








**Generalized Additive Model**



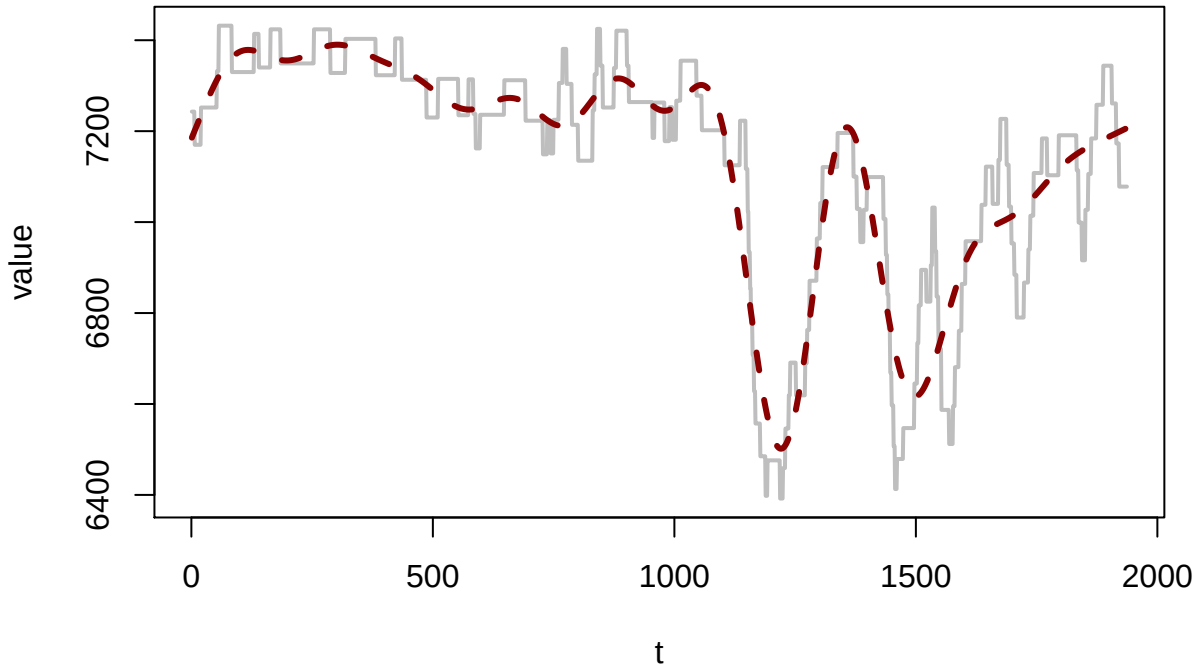



Table 2: Generalized additive method coefficients

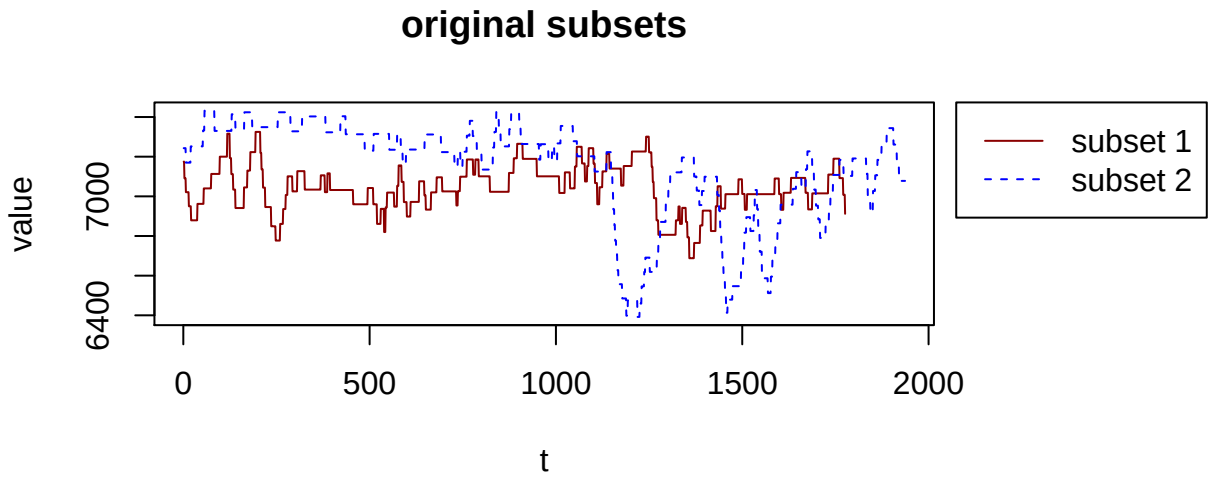
| | model.coefficients | model2.coefficients |
|---|---|---|
| (Intercept) | 7041.694 | 7142.776 |
| s(t).1 | 182.927 | -830.923 |
| s(t).2 | -258.267 | 823.207 |
| s(t).3 | 163.705 | 248.025 |
| s(t).4 | -382.138 | -955.976 |
| s(t).5 | -110.540 | 348.579 |
| s(t).6 | 459.813 | 376.510 |
| s(t).7 | -221.561 | -18.056 |
| s(t).8 | 272.068 | -1200.304 |
| s(t).9 | 234.727 | 179.715 |
| s(t).10 | 64.967 | -160.544 |
| s(t).11 | 288.297 | -1041.629 |
| s(t).12 | 398.487 | -75.881 |
| s(t).13 | -188.467 | 0.011 |
| s(t).14 | 410.885 | -1469.601 |
| s(t).15 | -147.774 | 282.820 |
| s(t).16 | 3.768 | -322.895 |
| s(t).17 | 120.766 | 629.964 |
| s(t).18 | 876.167 | 1733.116 |
| s(t).19 | 648.390 | 871.507 |

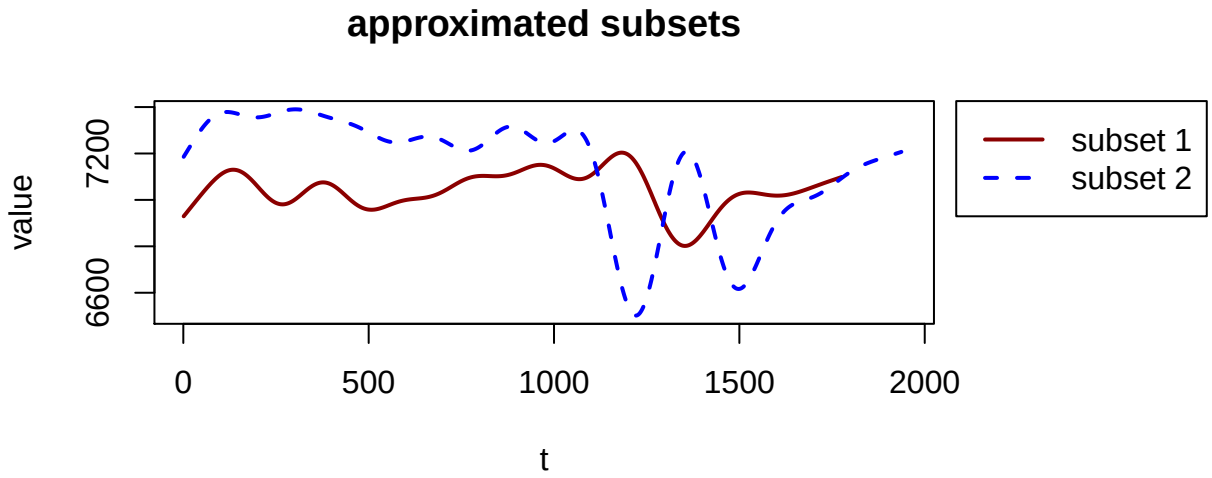
Generalized additive model formula:

- $g(E(Y)) = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_m(x_m) + \varepsilon$
  - $x_1, x_2, \ldots, x_m$ - are independent variables
  - $\beta_0$ - is an intercept
  - $f_1, f_2, \ldots, f_m$ - are unknow smooth functions (splines)
  - $\varepsilon$ - is an random error

In our case formula is :

- $g(E(Y)) = \beta_0 + f_1(x_1) + \varepsilon$,
  - $f(x_1) = \sum_{j=1}^{k} b_j(x_1)\beta_j$,
  - $b_j(x_1)\beta_j$ - smooth spline function.




---

**Correlation between approximated signals**

```
## [1] -0.0308847
```

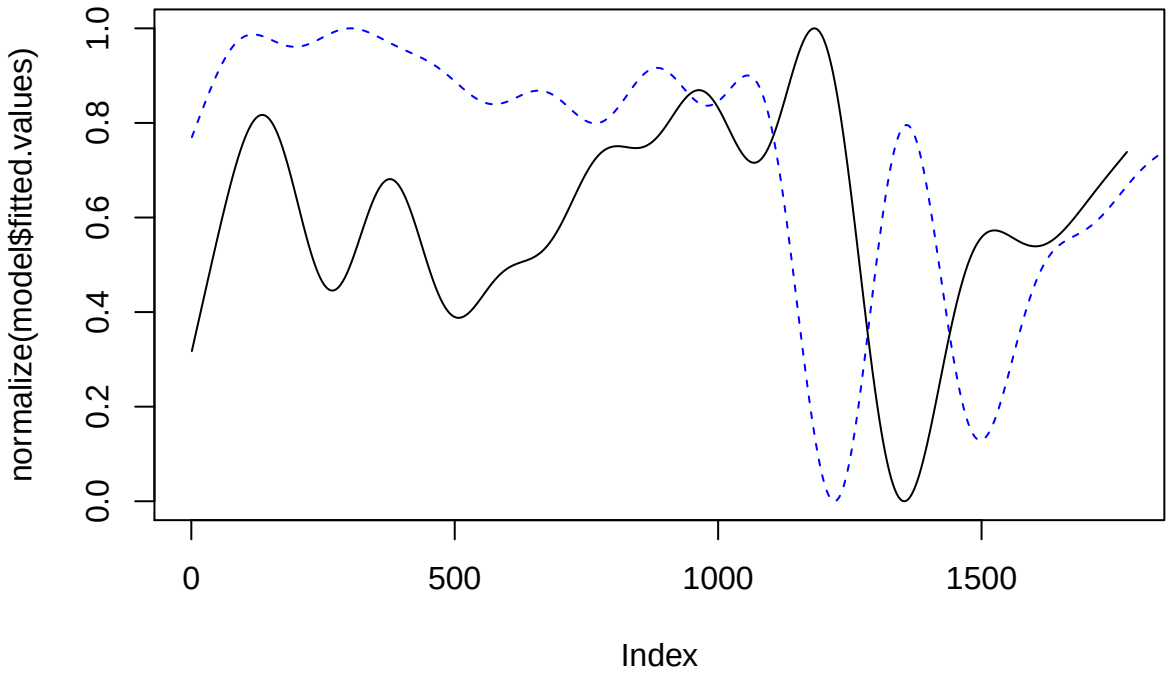
We can make further signal predictions based on this models.

```
#
#
# #+ eval=F
# ## create a chart for original data with 'plotly'
  plotly::plot_ly(
    x = d1$timestamp
    , y = d1$value
    , type = 'scatter'
```

```
    , mode = 'lines'
  )
```

```
## TypeError: Attempting to change the setter of an unconfigurable property.
## TypeError: Attempting to change the setter of an unconfigurable property.
```

```
plot(normalize(model$fitted.values), type = 'l')
lines(normalize(model2$fitted.values), col = 4, lty=2)
```


normalize(model$fitted.values)
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
Index