Hobo-report

Nele Stackelberg: Matrikel-Nr. 3548205

25.1.2018

1 Introduction

During the Christmas break 2017/2018 the temperature and light intensity at different locations in Freiburg was measured. For the measurment a HOBO-sensor of the type "HOBO® Pendant Temperature/Light Data Logger (Part # UA-002-XX)" was used. First a quality check was performed on the logged data. This Quality control excludes erroneous data. The exclusion of data causes gaps in the data, which are filled with data of a regression model. To calculate the regression model, a nearby weather-stations data is used. As a result the temperature trend in the measurement period is presented, including the accuracy of the measurment itself.

1.1 Description of Hobo location

Figure 1: The Placement of the Hobo-sensor.

The information about the sensor location of this report are listed in table 1. The loggger was programmed to log temperature in °C and light intensity in lux every 5 minutes. Unlike the instructions of the manufaturer the logger had not been mounted horizontally, but vertically. This might have an evect on the light measurements. Furthermore the position at the balcony may be inflenced by the scarce opening of the door and therefore rise of temperature. The opening of the door only occured in the first week of measurement and between the 27nd 30th of december.

Table 1: Informations about the Hobo-sensor.

Hobo-logger 10350051	
measurement period (from - to):	15.12.2017 - 08.01.2018
location:	Eschholzstr. 70, Freiburg
longitude / latitude:	7.834819 / 47.997663
position:	W-faced 2.floor balcony

Hobo-logger 10350051	
influence by the sun:	never exposed to the sun
other possible influences:	temperature inflenced by the scarce opening of the door

2 Method

The data of the Hobo-sensor was read out with the Hobo-software and exported to a txt-file.

The software "R" (R Core Team 2016) is than used to read in the data and wragle it. For this purpose the R-packages "tidyverse" (Wickham 2017), "lubridate" (Spinu, Grolemund, and Wickham 2017) and "readr" (Wickham, Hester, and Francois 2017) were used. Because the logger was moved on the first and last day of measurement, all data of these dates had to be removed. The actual analysis takes only place for the period 16.12.2017 - 07.01.2018.

2.1 Quality-control

The measured data is qualitiy-checked to detect missing data and errors. Occuring errors could than also be corrected. The Quality control (QC) provides data with a higher standart of accuracy (Zahumenský 2004). The QC includes three checks. The plausible value check (including the plausible rate of change), the maximum variability check and the minimum variability check. All Data-sets that fail a check are flagged. In this report the QC is performed for the temperature data.

First the data was checked for the plausible rate of change of the values. A change of Temerature greater than 1 K whithin 5 minutes was set as threshold value. Furthermore there shouldn't be any temperatures higher than 60 °C or lower than -80 °C.

Next the maximum variability was checked based on standard deviation. The maximum variability was calculated with the following equation (after 2004):

$$max.var = |T_i - T_{i-1}| + |T_i - T_{i+1}|$$

Data points will be flagged if the maximum variability is smaller than four times the standart deviation of the last 10 values.

$$max.var \leq (4 * \sigma)$$

For each value, the standart deviation was generated from the value itself and its prior ten values. This means, the standart deviation was calculated from 11 fife-minute values (55 minutes in total) of the temperature.

Further the minimum variability of the data has to be checked. Data points of the temperature-Data are flagged, if the range of the data of the previous 60 minutes is smaller than 0.1 K.

2.2 Excluding the flagged data

All datasets that fail one of the previous tests are flagged. Nevertheless it is visually analysed which values failed the specific tests, to get an overview of the types of occuring errors.

The previous flagging-tests didn't react for the first 12 values of the series, because there is missing data to calculate the minimum variability of the previous hour. Therefore those values are checked manually.

The hourly temperatures are calculated as average from the fife-minute data. In the hourly time series the flagged data is excluded if there were at least two flagged values per hour. It doesn't matter if a single value failed one or more checks. Excluded values are replaced by NAs.

2.3 Filling gaps with regression model

The result of the quality-control are data gaps (NA values) in the hourly temperature time series. To fill these gaps with values, temperature-data of a nearby wheather-station is used. The closest weather-station to the logger of this report is the Weinbauinstitut (WBI), that provides its houlry temperature data for free.

A simple linear regression model is fitted. In the model the data of the WBI acts as predictor, the Hobo-data as response variable. To assess the quality of the model, the fit is evaluated visually, the residuals are analysed and the coefficient of determination is assessed (see Dormann 2013).

3 Results

The wragled fife-minute data of the temperature is illustrated in figure 2 The light intensity is included in the plot as the color. In the plot the temperature shows dayly variances (warmer during the day, colder at night) and even stronger variances in the course of the weeks. High light intensities seem to appear at high temperatures, but not for all temperture peeks.

A plot of the light intensity shows the day-night pattern (3). The dayly maximum values had been around 2500 lux at many days. But compared to those days there are also a few days that have very high lux intensities, with 18600 lux as the maximum recorded value. A value around 20000 lux is described on wikipedia as “Shade illuminated by entire clear blue sky, midday”. The lux values therefore show that the logger wasn’t exposed to sun during the measurment period. No sun exposure is required to get valuable data.

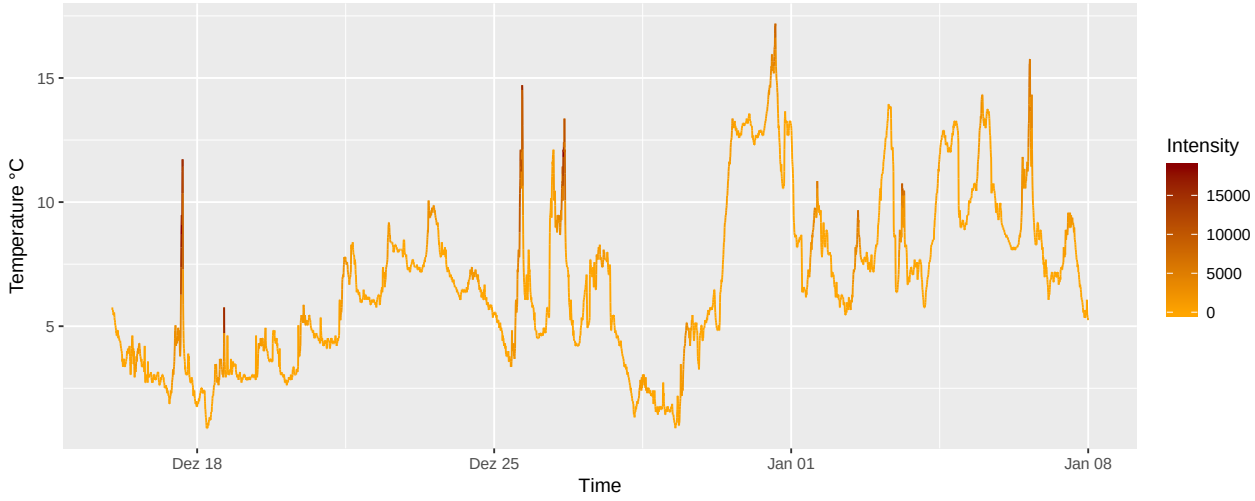

Figure 2: Temperature in °C, measured by the Hobo-sensor. The color refers to the measured light intensity (lux).

3.1 Quality control

The plausible value check showed that the sensor only recorded plausible values. The following summary of the Temperature data show amongst others the minimum and maximum-value of temperature. For this range of values the HOBO® Pendant describes an accuracy of ~0.45 K and a resolution of ~0.15 K (Compter Corporation 2009).

The Quality control results in the flags that are displayed in table 2. To see the effect of the quality checks, the flagged values are shown in figure 4. The unrealistic values are flagged and the leftover data is of good

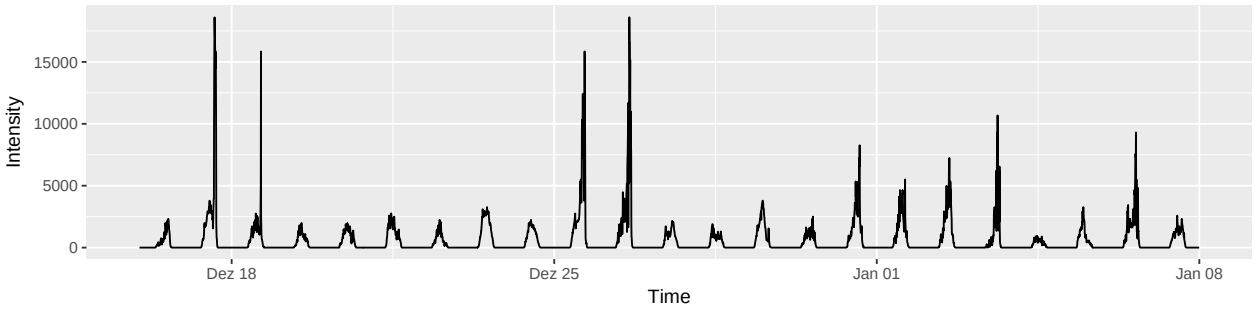

Figure 3: time series of the light intensity (lux)

quality. For the hourly Quality-controled temperature-data some indices are calculated. The relative amount of NA-values is 15.04 %. The dayly average Minimum (5.16 °C), Mean (6.62 °C) and Maximum (8.41 °C) are within a realistic range. Furthermore the value of the dt90 is 1.27, an index to characterize rapid temperature changes.

The regression has the multiple R-squared value 0.9645, which shows that the linear model is very suitable. The plot of the residuals (figure 5) of a good model shouldn't show any pattern. For this model it doesn't show an obvious pattern, thats why it is used to fill the data gaps.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.893   4.311   6.471   6.760   8.581  17.190
```

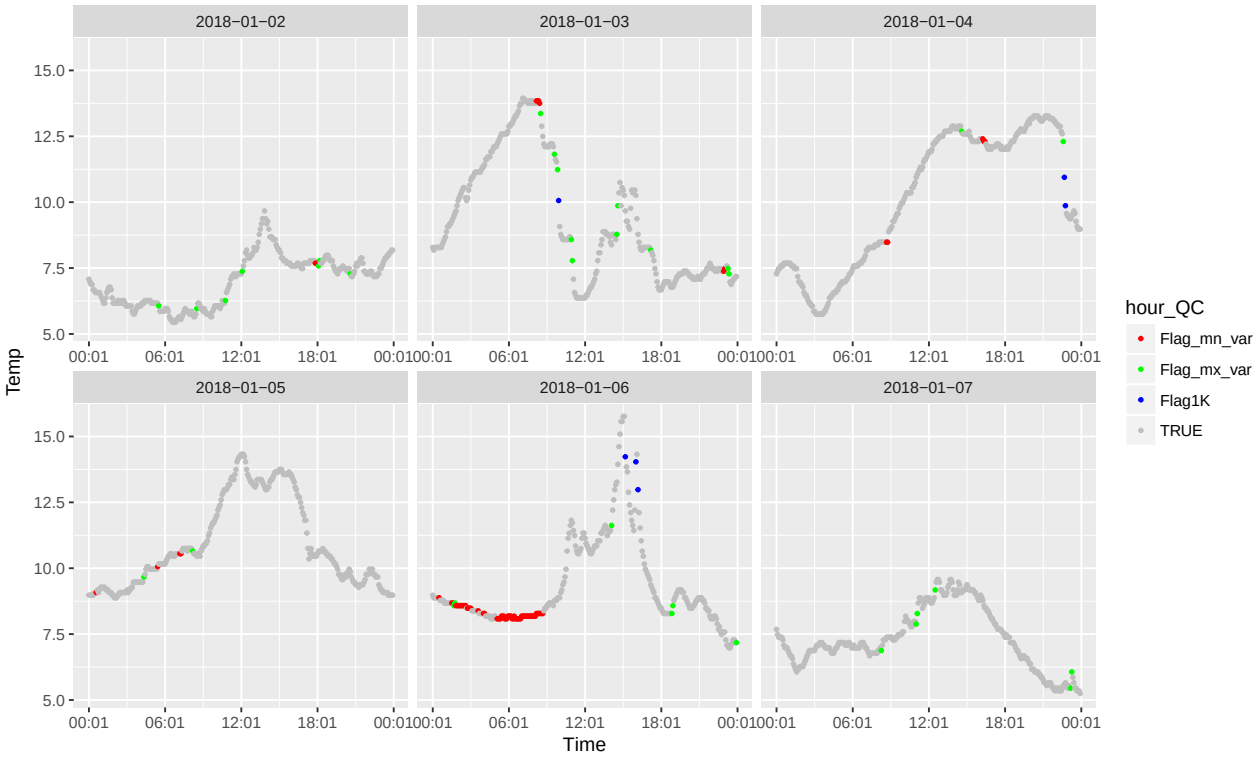

Figure 4: Some example dates, where the temperature values are colored according to their flaggs. TRUE means that they are not flagged.

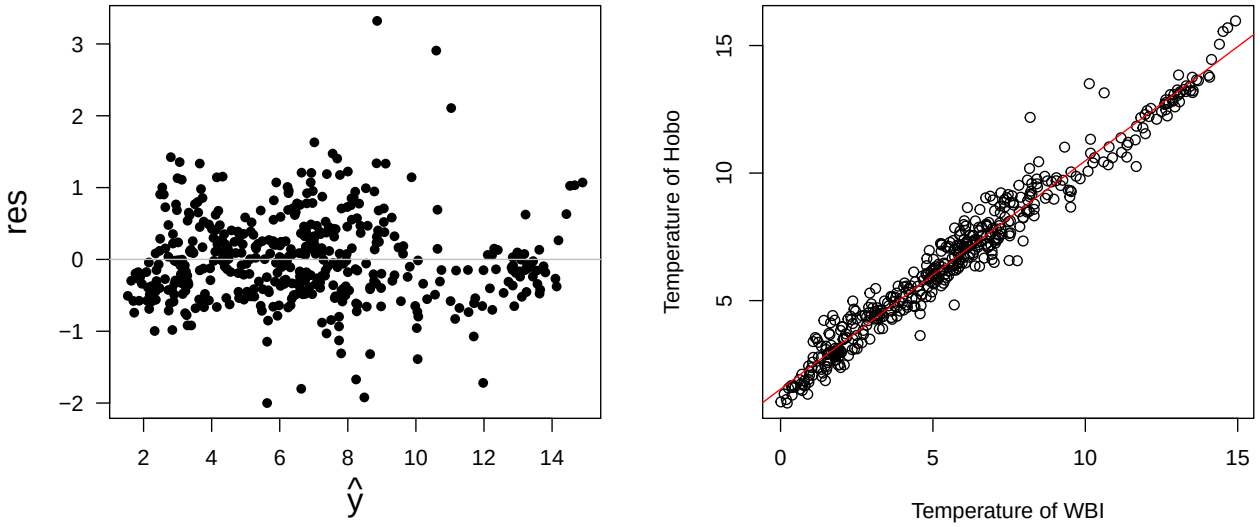

Figure 5: Analysis of the regression model: Plot of the residuals on the left and on the right the WBI-data (predictor) vs. the HOBO-data (response) with the linear model (red line)

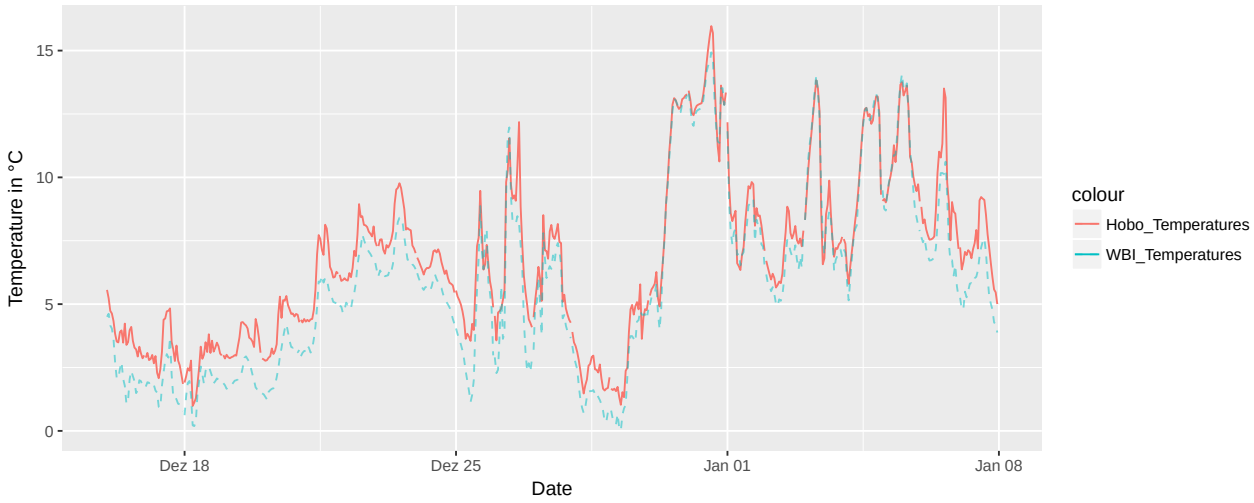

Figure 6: Temperature of the Hobo-sensor. The gaps in the timeline have been filled with the regression model. The WBI-data, which is displayed dashed in the figue served as predictor.

Table 2: Number of the computed Flaggs for every step of the Quality check.

Flag	Amount of Flags
More than 1 K temperature change in 5 minutes	26
Maxumum variability based on standart deviation	176
Minimum variability whithin 60 min lower than 0.1 K	216
Flags in hourly data	83

4 Discussion

The data collection with the Hobo-sensor is a solid method to measure temperature and light intensity. Mainly because it measures automatic. Though the placement of the sensor on a balcony and its meaning should be discussed. Furthermore the threshold values of the Quality checks could be adjusted to eliminate less values. The extend of the adjustment has to be balanced with the needed quality of the data. For example, the standart deviation was generated from the last 10 5-minute values of the temperature. The standart deviation could either be calculated from the last 10 values, the resent value excluded, or from the last 10 values plus the resent value. The second option means that the maximum variability has to be calculated from 11 values. This slight difference is of relevance regarding the different amount of flagged data.

Fortunately there was very good data available from a nearby weather station. The regression model produced qualitative data to fill the gaps in the Hobo time series.

References

Compter Corporation. 2009. *HOBO Pendant Temperature/Light Data Logger*. Onset Computer Corporation.

Dormann, Carsten F. 2013. *Parametrische Statistik : Verteilungen, Maximum Likelihood Und Glm in R*. Statistik Und Ihre Anwendungen. Berlin ; Heidelberg: Springer Spektrum.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Spinu, Vitalie, Garrett Grolemund, and Hadley Wickham. 2017. *Lubridate: Make Dealing with Dates a Little Easier*. https://CRAN.R-project.org/package=lubridate.

Wickham, Hadley. 2017. *Tidyverse: Easily Install and Load the 'Tidyverse'*. https://CRAN.R-project.org/package=tidyverse.

Wickham, Hadley, Jim Hester, and Romain Francois. 2017. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Zahumenský, Igor. 2004. "Guidelines on Quality Control Procedures for Data from Automatic Weather Stations," COMMISSION for basic systems,. WORLD METEOROLOGICAL ORGANIZATION. https://www.wmo.int/pages/prog/www/IMOP/meetings/Surface/ET-STMT1_Geneva2004/Doc6.1%282%29.pdf.