Artifact Characterisation:

Detection of artifacts in monitoring data in the Paediatric Intensive Care

---

Anne van den Elzen
673237

Supervisors:
Prof. dr. Arno Siebes
dr. Ad Feelders
Ruben Zoodsma, MSc
dr. Lex van Loon

Department Information and Computing Sciences
Master Applied Data Science
Utrecht University
The Netherlands
July 11, 2024

Abstract

There is a need for annotated medical data that can be used to develop and evaluate smart algorithms. Synthetic data can help provide this, but is often too sterile to use. We aim to detect artifacts in real PICU data and describe these artifacts, such that the synthetic data can be realistically corrupted by artificial artifacts.
Unsupervised methods for artifact detection often rely on a filtering technique or simple cut-off rules. For our artifact detection method we will build on this idea and expand it. We will detect different types of artifacts in mean blood pressure, oxygen saturation and heart rate. We will focus on a artifact caused by blood sampling and dips in oxygen saturation and heart rate and describe characteristics of the detected artifacts. We have found a total of 678 blood sample events in a total of 2666 hours of observations, a total of 1110 artifact dips in oxygen saturation and 1946 artifact dips in heart rate in a total of 2782 hours of observations .With a amplitude and duration mean (standard deviation) of 95 (72) mmHg, 9.8 (8.0) % and 18.0 (14.5) bpm for blood sample events, oxygen saturation dips and heart rate dips respectively. And a duration of 77 (83), 26.6 (20.6), 12.3 (9.8) seconds. The blood sample events are likely uniformly distributed in time, for the artifacts in oxygen saturation and heart rate there is likely a time dependence.
With these results we have a better understanding of what the characteristics of artifacts in vital parameters look like for PICU data.

Contents

1 Introduction

At the Paediatric Intensive Care Unit (PICU), where the are children in critical conditions, alarms are omnipresent. These alarms have the goal to inform the critical care nurses that the patients condition is deteriorating and actions need to be taken. These monitoring systems are designed to be very sensitive and not very specific. Research shows that the false alarm rate in intensive care units range from 0.7 to 0.99 (Hever et al., 2020). These frequent alarms are not only annoying, but also leads to alarm fatigue among health care personnel (Drew et al., 2014). Which may have serious negative consequences for both patients and nursing personnel. So, there is a demand for smart alarming systems to decrease the false alarms at the PICU.

To develop these smart alarming systems, large amounts of data are needed. Preferable with annotations whether a change is a true clinical deterioration or an artifact. At the PICU the monitoring data is saved, but there are no annotations available. Without these annotations it is difficult to use these datasets for the training, testing and validating of smart alarm algorithms since annotating retrospectively is very difficult.

Realistic synthetic vital sign data could be a solution for this problem. Biophysical models can be used to model vital signs that fit a certain illness. The down-side of synthetic data that is generated with biophysical models is that it lacks realism. It is very sterile and misses the artifacts present in real monitoring data. These artifacts can originate from different factors, such as disconnected arterial lines, movement, coughing, blood sampling. While especially these artifacts result in extreme vital sign values that cause false alarms. So, it is important to enrich the synthetic data with these artifacts. To be able to realistically add artifacts to synthetic data, we need to detect artifacts in real PICU monitoring data and describe their frequency, shape and timing.

Large-scale studies about artifact detection in critical care data are still rare. So, there is no consensus about what kind of artifact detection methods lead to good results for this population. Currently, most used artifact detection/correction methods use cut-off filters or moving mean/median filters (Du, Glick, & Tung, 2019) (Charbonnier & Portet, 2012) (Hoorweg, Pasma, van Wolfswinkel, & de Graaff, 2018). But it is unclear whether these methods are really suitable for the PICU population we are studying.

In this project we will develop two artifact detection methods, the first is especially for blood sample events and the second uses a baseline correction and patient and vital sign specific cut-off filters. The artifacts that are detected using these methods will be described in amplitude, duration, frequency and timing.

So, the main objective of this project was to detect artifacts in PICU monitoring data and characterise them. These characteristics could in turn be used to enrich the synthetic data and get a better insight in artifacts present in the data. We will focus on an artifact in the mean blood pressure cause by blood sampling and artifacts in the oxygen saturation and heart rate.

The set up of this report is as follows. In Section 2 the data set and the method used to detect artifacts will be described. First, the method for detecting blood sample events will be discussed in Section 2.2. In Section 2.3 a method is proposed to detect dip shaped artifacts in oxygen saturation and heart rate. Section 3 gives the results for the detection methods and what characteristics the detected artifacts have. In Section 4 we will give a summary of our findings and what implementations they can have. In Section 5 the limitations of our research will be discussed and recommendations for future research are given. Additional information can be found in the Appendices.

2 Methods

2.1 Data description

The dataset that was used for this project is provided by the Wilhelmina Children's Hospital (WKZ). A fully anonymized version of the data was used. The patient population consisted of 68 infants with cCHD admitted to the Paediartic Intensive Care (PICU) of the University Medical Centre Utrecht between 2002 and 2018. Data is collected of 5 vital parameters in a frequency of 1 observation per second. The included parameters are saturation (SpO_2), regional cerebral saturation (rSO_2) in both hemispheres, invasive mean, systolic and diastolic arterial blood pressure (MBP, SBP and DBP), respiratory rate (RR), heart rate (HR) and the ventilation status. Patients were excluded if the only observations available were of the regional cerebral saturation.

Furthermore, we have some more details about the patients. Most importantly, age and weight when they got admitted to the PICU.

The model used to produce the synthetic data outputs the following parameters: mean blood pressure, SpO_2, heart rate and respiratory rate. Therefore, we will only focus on these parameters.

Patients were excluded if they had less than 2 hours ($n = 2$) of total observations available and if the only observations available were of the regional cerebral saturation ($n = 17$). So, in total we had data of 49 patients. In Table 1 the median and Inter Quartile Range (IQR) of the patient specific characteristics and their vital signs are given. There is no further pre-processing since we want to detect noise and abnormalities. If patients do not have observations available for one of the parameters of interest, that patient is excluded for that specific analysis. More details for this are given in the Section 3.

Table 1: Summary statistics of the characteristics of the patients and the vital parameters.

Characteristics of Patients			
Patient specific (median (IQR))			
weight at $t = 0$ (g)	3450 (3175:3750)		
age at $t = 0$ (days)	10 (7, 17)		
available data (hours)	56.2 (6.1: 84.4)		
Vital parameters (median, (IQR))			
	Total	averaged over patients	percent missing values over patients(%)
Heart rate (beats per minute)	146 (135:157)	144 (132:153)	5.0 (0.4: 24.5)
Respiratory rate (breaths per minute)	34 (28: 39)	34 (28: 40)	11.1 (1.6: 42.2)
SpO_2 (%)	92.3 (80.4: 97.3)	90.3 (86.8: 92.7)	9.9 (1.5: 32.9)
Mean Blood Pressure (mmHg)	51.8 (46.4:58.3)	52.5 (48.0: 58.0)	23.3 (1.3: 48.6)

2.2 Detection of blood samples using mean blood pressure measurement

Arterial blood pressure waveforms are frequently corrupted by artifacts, caused by events such as transducer flushing, drawing a blood sample and physical movement. For now we focus on one type of artifact, namely the typical shape in mean blood pressure caused by taking a blood

sample. In the dataset we have three parameters that describe the blood pressure. We have the systolic blood pressure (SBP), the mean blood pressure (MBP) and the diastolic blood pressure (DBP).The DBP and SBP are both measurements directly from the monitor and the MBP is calculated using these two measures. A common method to estimate the MBP is by taking the DBP and adding one third of the difference between the SBP and the DBP.

The artifact type we focus on can be described in the following general way: The is a sudden and rapid increase in MBP to some maximum, followed by a immediate and fast decrease. In this period the MBP is higher then the SBP. There could be a second peak in short period after the first peak, due to flushing of the arterial line. For the period where the MBP is higher then the systole there is a period of consecutive missing values for the SBP in the dataset available. Before or after the maximum value of the MBP there is a period with missing values for the MBP. A real example of this pattern is shown in Figure 1.

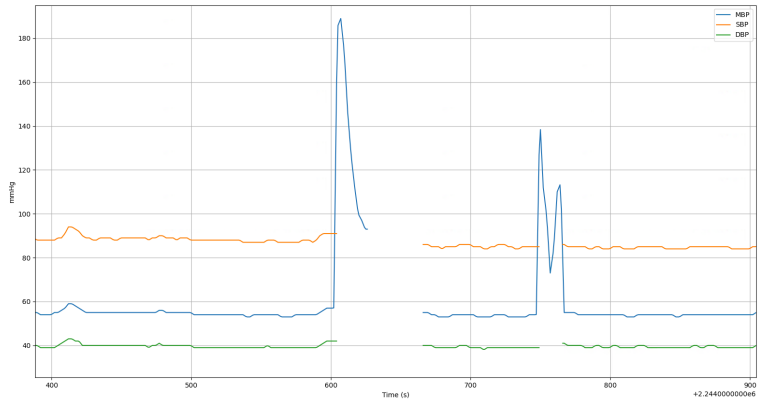


Figure 1: Example of blood sample event.

We want descriptive statistics of the shape that is shown above. In order to do this we need to find time intervals in the dataset where the MBP and the SBP match the pattern described above.

To find intervals that match the pattern described above we use the following method. Fill in all the missing values for SBP by propagating the last valid observation until the next valid observation. This is done, because during the peak of the MBP there are often missing values for the whole duration of this peak or even a little longer. A lot of peaks would not be found if the missing SBP values would not be imputed.

Let $\mathcal{T} := \{t \in \{0, ...T\} \mid MBP(t) > SBP(t)\}$ be the set with time points where the MBP is higher then the SBP. Then for all $t \in \mathcal{T}$ we take the interval $[t - k, t + k]$ as segment we are interested in, where k is a predefined constant. Note that intervals can overlap. If intervals overlap, they are joined together. This means that all segments can have different lengths, with a minimum length of $2k$ seconds.

For each blood sample event we want to characterise the shape. To do this we use duration and amplitude of the blood sample event and the median, minimum and maximum value in the interval. We use the number of missing values in the SBP parameter in the interval as a proxy for the duration of the blood sample event. The amplitude is defined as the difference between the maximum and median MBP value in the segment.

Per patient we are interested in the absolute number of detected artifacts and the relative number of artifacts, which is defined as the total number of artifacts that were detected for the

patient divided by the total duration of the available observations.

Parameter setting

The only predefined constant in this method is the search window size. We choose to use $k = 60$. It is empirically found that results for duration are consistent for non-extreme cases. Adding to that it is plausible that abnormal patterns (i.e. the number of missing values in the SBP parameter) are still caused by the same blood sample event if they occur in this window.

2.3 Dip Characterization for oxygen saturation and heart rate

For critically ill patients a decrease in vital parameters such as oxygen saturation and heart rate can be an indication that the patients clinical state is deteriorating. In this section we are looking for artifacts that have a rapid change from a normal value to a minimum value and transition back to the normal value of before the decrease. If this happens in a short amount of time and to extreme values this dip is most likely not because of physiological reasons and is thus an artifact we are interested in to find and describe.

(Charbonnier & Portet, 2012) proposed a self-tuning adaptive trend extraction method for process monitoring and diagnosis with as application to reduce the false alarms in Neonate Intensive Care Units. NICUs care for new born babies who are premature or very ill. This population is similar to the patient population in our research. Among other things they define a method to detect artifacts (or step changes). The idea of using deviation from a baseline signal and the numerical derivative we will use in this project are inspired by their approach.

We choose for a non-parametric approach to detect artifacts since we want to make no assumptions about the distribution of the data. With this kind of medical time series such assumptions often do not hold. Therefore a more straight forward and understandable method with no assumptions about the underlying distribution is proposed.

In order to detect a artifact dip we need to define such a dip precisely. We assume an artifact dip has the following characteristics:

- From the start of the dip a decrease is observed, until a minimum point followed by an increase back to the initial level.
- The minimum value of the dip is less then a threshold from the baseline. This threshold is fixed based on the signal variability.
- This dip is an abrupt change. The slope from the start point till the minimum point, (or from the minimum point till the end point) should be more extreme than expected.
- The maximum duration of the dip is D seconds. Where D is chosen beforehand.

For more robust detection of dips in the monitoring data, brief insignificant fluctuations are removed by using a rolling median filter. Median filters are a easy and understandable way to remove these insignificant fluctuations while maintaining the the overall baseline (Mäkivirta, Koski, Kari, & Sukuvaara, 1991). Let $y_N(t)$ be the results of a median filter applied to x, defined as:

$$y_N(t) = \text{Median}\left(x(t-N), x(t-N+1), ..., x(t+N-1), x(t+N)\right)$$

where N is the window size. A centered median filter is used to ensure that the variations in the median filtered signal are aligned with the variation in the data.

We want to detect dips that pass a certain threshold. For example, we might be interested in dips where the minimum is below 95% for oxygen saturation (Ray, Rogers, Raman, & Peters, 2017). In the data set we have measurements for critically ill patients and therefore there are a lot of fluctuations, both caused by clinical deterioration and artifacts. We are only interested in the latter. If we would use the same threshold for everyone we have poor results for patients that have a low oxygen saturation in general. We therefore use the deviation from a baseline to detect artifacts. Artifacts are usually more abrupt than changes that are due to a physiological cause. By using the deviation from the baseline we focus on the more abrupt changes. To determine the baseline a rolling centered median is used with a (large) window size $N_{baseline}$.

Let the residual signal to be the difference between the median filtered signal y_N and the baseline $y_{Nbaseline}$, given by

$$res(t) = y_{Nbaseline}(t) - y_N(t)$$

The value of the residual signal will be compared to a threshold. This threshold is defined as $\text{mean}(res) - \lambda\sigma$. Where σ is the sample standard deviation of the residual signal and λ is a predefined constant. If the residual signal passes the threshold, then this point is part of a dip we are interested in. From this minimum point that passes the threshold we search back to find the begin point of the dip. The begin point t_{begin} is a time point before the minimum where $res(t) \geq -\alpha$ and $t - t_{begin}$ is minimal, where $\alpha > 0$ is a predefined constant. Similarly, for the end point t_{end} of the dip, we look for time points after the minimum where $res(t) \geq -\alpha$ and $t_{end} - t$ is minimal. For dips that occur at the very beginning of the signal, it might be the case that there exists no t satisfying the above conditions. In that case, we find no beginning of the dip and discard it. Similarly, we discard dips where no end point is found because the dip is close to the end of the available observations.

Artifact dips are abrupt changes with a large amplitude. With the above conditions we have found the dips that have a large amplitude. To ensure that the dips we found are abrupt we add a second constraint.

Let $res_d(t) = res(t) - res(t-1)$ denote the numerical derivative of the residue, then let σ_d be the sample standard deviation for the numerical derivative of the residue. Define the maximum slope of a dip to be

$$p = \max\left(\left|\frac{res(t_{begin}) - res(t_{min})}{t_{begin} - t_{min}}\right|, \left|\frac{res(t_{min} - res(t_{end}))}{t_{min} - t_{end}}\right|\right), \tag{1}$$

where $t_{min} = \text{argmin}_{t \in [t_{begin}, t_{end}]} res(t)$.

Then if

$$p \geq \lambda_d \sigma_d, \tag{2}$$

,with λ_d a predefined constant, holds the dip is abrupt.

Lastly, we impose a constraint on the duration of the dip. That is $t_{end} - t_{begin} < D$ to be detected as artifact.

The above described method is summarized in pseudo-code in 1.

Algorithm 1 Dip detection

Initialize λ, α, λ_d, D
Let $V := \{t \in \{0, 1, ..., T\}$ such that $res(t) < \lambda\sigma\}$
for $t \in V$ **do**
  for $0 < k < t$ **do**
    if $res(k) \geq -\alpha$ and $t - k$ is minimal **then**
      $t_{begin} \leftarrow k$
    end if
  end for
  for $t < j < T$ **do**
    if $res(j) \geq -\alpha$ and $j - t$ is minimal **then**
      $t_{end} \leftarrow j$
    end if
  end for
  Set $t_{min} = \text{argmin}_{t \in [t_{begin}, t_{end}]} res(t)$
  Set $p_1 = \frac{res(t_{begin}) - res(t_{min})}{t_{begin - t_{min}}}$.
  Set $p_2 = \frac{res(t_{begin}) - res(t_{min})}{t_{begin - t_{min}}}$.
  if $\max(|p_1|, |p_2|) \geq \lambda_d \sigma_d$ and $t_{end} - t_{begin} < D$ **then**
    Dip is an abrupt change and a possible artifact.
  end if
end for

A vectorized approach is used for the implementation of Algorithm 1, using the `numpy` package in Python to make the algorithm more efficient.

Now that we have located the dips, we want to know the characteristics of the dip. We are interested in the duration, the amplitude, the timing, the start and end value of the dip and the slope.

The duration of dip is equal to $t_{end} - t_{start}$; The amplitude of a dip is difference between the minimal value $\min_{t \in [t_{begin}, t_{end}]} res(t)$ and the mean value of $res(t_{begin})$ and $res(t_{end})$.

Parameter setting

The constants that should be defined beforehand are window sizes N and $N_{baseline}$, the recover parameter α, the sensitivity parameters λ for the amplitude of the dip and λ_d for the abruptness and finally the maximum duration D.

The parameter settings are given in Table 2.

Parameter settings		
	SpO_2	HR
N	2	2
$N_{baseline}$	150	150
α	2	2
λ	4	4
λ_d	4	5
D	100	100

Table 2: Parameter values used for dip characterization

The parameters for both oxygen saturation and heart rate are set to the same values. Except for λ_d. There are more fluctuations and dips in the heart rate parameter that are not because of artifacts, but have a physiological cause. Which is expected considering that the patient all are diagnosed with Critical congenital heart disease. To make the artifact detection more robust for these rapid changes, the threshold for the slope is set higher. So, a dip has to be more extreme then expected to be seen as a artifact compared to the oxygen saturation dips.

We have set λ to 4. Meaning that the dip should be at least 4 standard deviations from the mean. This is quite a conservative choice. We chose to set it to 4 to be more sure that the dips that we detect are in fact artifact dips.

The maximum duration of the dips is set to 100 seconds. This is a bit longer then what the maximum expected duration of an artifact dip is, which is about 60 seconds. This is done, because dips with a very large amplitude or that are very sudden, but last longer then initial thought are still detected.

$\alpha = 2$ is chosen because we want to allow of the dip to not start exactly at the baseline and after the decrease increase back to the exact level as before the dip. It is empirically tested that with a lower α there are dips that remain undetected because the value does not recover back closer to the baseline within the maximum duration.

2.4 Relative timing of artifacts

Patients in this data set are critically ill when they enter the PICU. Over time the patient will become more stable. It is therefore expected that there will be more artifacts in the beginning of their recording. For example blood samples might be needed more frequent in the beginning to test the patients condition or there is more movement due to pain in the beginning of their PICU stay.

It is therefore interesting to investigate whether more artifacts are detected in a certain period of a patients stay. If so this can be taking into account when corrupting the synthetic data with artifacts. Since, all patient have a different amount of observations timing for a event is scaled compared to the total duration of observations. This is done by dividing the index of the start of the event by the total number of observation for that patient. The relative timing is then an real number in the interval $[0, 1]$. By doing this there is a better comparison between the timing of events per patient and the distribution of the relative timing can be analysed.

3 Results

In this Section the results of the characteristics of the detected artifacts will be given. Moreover, one patient was randomly chosen to use for a validation of our artifact detection. This patient had a total of 31 hours and multiple detected artifacts for all three vital signs, MBP, oxygen saturation and heart rate. An expert stated per detected artifact whether this was correctly identified as artifact. They furthermore indicated whether there were episodes that should have been detected as artifact, but were missed. We say there is a true detection if the expert and the artifact detection agreed. The true detection rate is the number of true detection's divided by the number of total to be detected artifacts.

There were 12 blood sample events detected. One was a true blood sample event. The rest were falsely identified as blood sample events, because of a loss of signal. The were no blood sample events identified by the expert that were missed. The true detection rate was $1/1 = 1$.

For oxygen saturation there were 21 dips detected as artifact. 19 out of the 21 dips were correctly identified as artifact, so 2 dips were falsely identified as artifacts. In total 27 artifacts were identified by the expert, so 8 artifact dips were missed. So, the true detection rate is 19/27 = 0.70. Three of the missed artifacts were not identified because the slope was not steep enough. One because it was at the end of the observations and the rest because there were missing values around the dips, so no start or end could be detected.

For heart rate there were 5 dips detected as artifact. All 5 were correctly identified as artifact. In total 11 artifacts were identified by the expert. So, the true detection rate is the 5/11 = 0.45. However, there was a lot of variance in the heart rate of this patient. It was difficult to say whether a dip was an artifact or not. A more zoomed in window should be used to identify whether the dip is an artifact or not and then still it is difficult to say whether the dip is an artifact or not.

The full validation document can be found in Appendix B.

3.1 Detected blood sample events using MBP

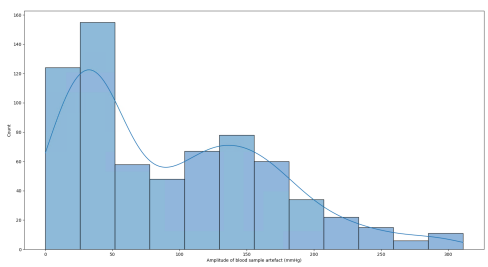
For this analysis 43 different patients were included. 6 patients were excluded since there were no SBP measurement available. In total 714 artifacts that were found in a total of 2675 hours of observations. Missing values were not taken into account for this. The number of available observation differed per patient. On average the total observation time is 62 hours per patient (standard deviation of 57.5 hours) and a minimum of 2.1 hours and a maximum of 209.4 hours.

The mean search window was 334 seconds (standard deviation of 447.0 seconds). The median search window was 254 seconds and the interquartile range was 244:316. The maximum search window was 10565 seconds. This is a lot longer then expected. When inspecting the MBP and SBP of the patient where extremely long search windows were used, this was because of a loss of signal. Because these influence the summary statistics a lot. We decided to only look at the artifact that were found in a window with a size less then the 95% percentile, which is 638 seconds. With this constraint on the search window, we have $n = 678$ artifacts left. Per patient on average there are 0.43 blood sample events per hour with an standard deviation of 0.39. The median number of blood sample events per patient per hour is 0.3 with an interquartile range of 0.19:0.51. A summary of the artifact characteristics are shown in Table 3.

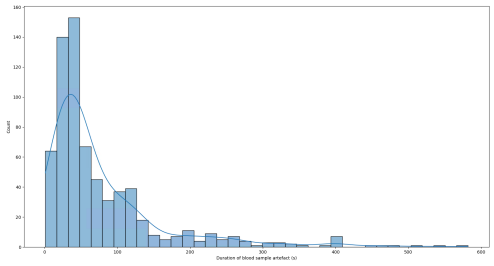
Table 3: mean, standard deviation (std), median and interquartile range (IQR) for the blood sample event.

	Amplitude (mmHg)	Duration (s)	Minimum (mmHg)	Median (mmHg)	Maximum (mmHg)
mean (std)	95 (72)	77 (83)	30 (30)	54 (26)	149 (75)
median (IQR)	79 (34:149)	46 (29: 98)	39 (2:48)	50 (45:57)	137 (86:201)

The correlation between the amplitude and the duration of the artifact is tested with a two-sided Pearson test. The Pearson product-moment correlation coefficient is -0.017 with a corresponding p-value of 0.65. So, the an α-level of 0.05 we do not reject the null-hypothesis. Hence, with this test there is not enough evidence to conclude that the correlation between the amplitude and the duration is non-zero. The joint distribution of amplitude and duration is also shown in Appendix 12.



(a) Distribution of blood sample artifact amplitude.



(b) Distribution of blood sample artifact duration.

Figure 2: Distribution of the amplitude and the duration of the detected artifacts.

The standard deviation of the patient average amplitude to the total average amplitude is 37.7 mmHg. The average standard deviation within a patient is 68.8 mmHg. For the duration the standard deviation of the patient average duration to the total average total is 71.5 seconds. The average standard deviation within a patient is 121.6 seconds. So, there is more variation for duration and amplitude within a patient then between the patients.

Relative timing of blood sampling

In Figure 3 a histogram is shown with the frequency of the relative timing of a detected artifact. The relative timing for all 678 events are used.

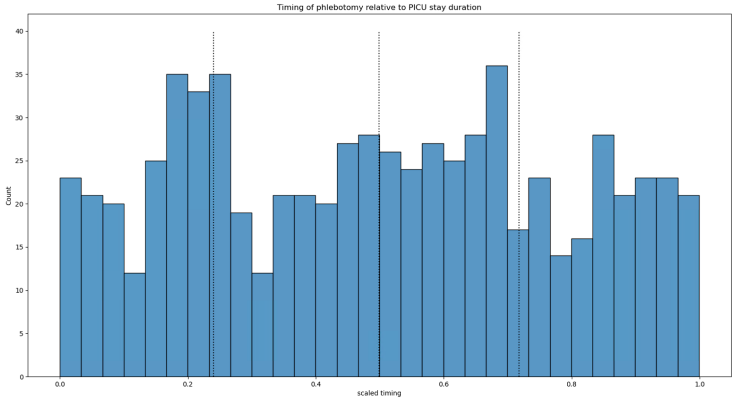


Figure 3: Histogram of the relative timing of the start of artifact. With three vertical lines indicating the 25%, 50% and 75% quantiles.

By visually inspecting this histogram we see no apparent pattern. By eye-balling this histogram it might be the case that the relative timing is uniformly distributed.

To quantify whether the distribution of the relative timing is close to a uniform distribution with as interval $[0, 1]$ we use a one-sample Kolmogorov-Smirnov test. We test whether the 678 data points look like they could have been drawn from a uniform distribution between $[0, 1]$. The

results are: Kolmogorov–Smirnov statistic is -0.439 with a corresponding p-value of 0.127. With an alpha-level of 0.05 we do not reject the null-hypothesis. Hence, with this test there is not enough evidence to conclude that the distribution is not uniform.

3.2 Dips in heart rate and oxygen saturation

In this subsection the results for the detected dips are presented. First, we will discuss the results for the dips detected in the oxygen saturation and next in the heart rate.

Oxygen saturation

For this analysis 49 different patients were included. The total hours of available observations are 2782.2. On average the total observation time per patient is 56.8 hours (standard deviation of 56.15). With a minimum of 2.1 hours and a maximum of 209.4 hours.

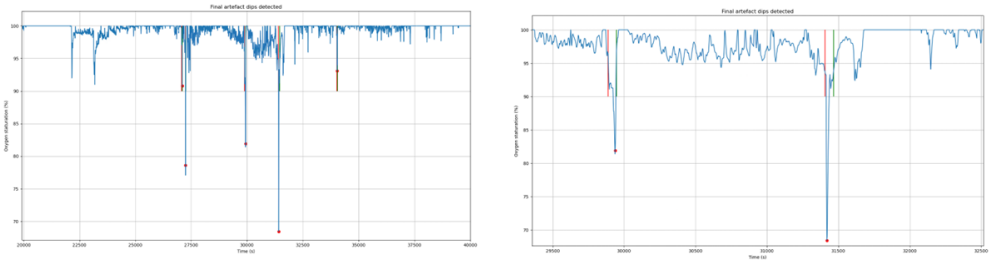
The threshold that the residual signal needs to pass in order to detect a dip is dependent on the mean residual signal and the standard deviation of the residual signal. In Table 4 descriptive statistics are given about the residual signal.

Table 4: mean, standard deviation (std), median and interquartile range (IQR) for the oxygen residual signal.

	mean(res)	σ	mean(res$_d$)	σ_d
mean, std	-0.07, 0.19	1.67, 0.82	0.00, 0.00	0.21, 0.10
median (IQR)	-0.05 (-0.13:0.00)	1.66 (1.11:2.05)	0.00 (0.00: 0.00)	0.18 (0.15: 0.29)

There are 2646 dips that pass the threshold in total, with $\lambda = 4$ and no constraint on the slope or duration. There is one patients where the residual signal never passes the threshold. On average there are 1.23 (std: 0.82) dips per hour where the minimum value of the dip is below the threshold.

When also looking at the second constraint 2 there are 1180 dips that still satisfy. Then dips that satisfy 2 and where the duration is less then $D = 100$ we have 1110 dips. There were 5 patients were there were no artifact dips detected. For the patients that did have detected artifact dips, on average there were 0.48 artifacts per hour with a standard deviation of 0.35. The median number of artifacts per hour is 0.41 and the interquartile range is 0.22: 0.52. Details about the duration and the amplitude of the dip are shown in Table 5. In Figure 4a an example of detected dips are shown. In Figure 4b a close up for a dip is shown.



(a) Example of the dips in the measured signal that are detected in oxygen saturation. The red and green vertical line indicate the begin and end of the dip, respectively. The red point indicates the point where the deviation from the baseline is maximum.

(b) Zoomed into the largest dip of the segment in Figure 4a.

Figure 4: Examples of the detected dips in oxygen saturation.

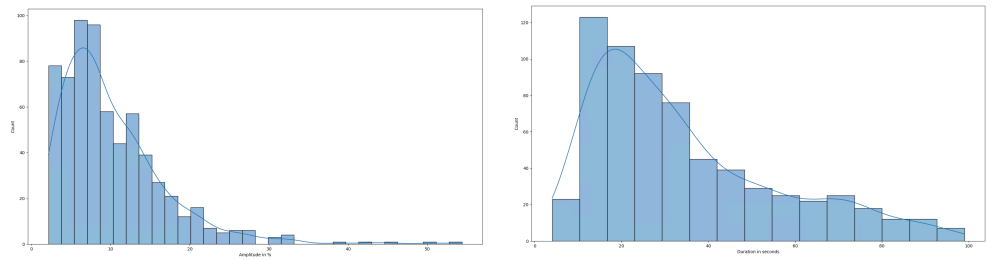
Summary statistics for the detected dips are shown in Table 5.

Table 5: Summary statistics of the duration and the amplitude of the dips detected in oxygen saturation. Where start value is the measured oxygen saturation at the beginning of a dip and minimum value is the minimum measured oxygen saturation during the dip.

	Duration (s)	Amplitude (%)	Start value (%)	minimum value (%)
mean (standard deviation)	26.6 (20.6)	9.8 (8.0)	88.1 (10.4)	78.2 (12.96)
median (IQR)	19 (12:32)	7.8 (5.0:12.2)	90.7 (84.1: 96.6)	79.8 (70.2:89.0)
minimum	2	0.2	52.2	1.5
maximum	99	95.5	100	97.6

The mean maximum slope in the dip is 2.1 % per seconds (std 3.0% per seconds). For 645 dips slope is from the start point of the dip until the minimum is larger the from the minimum point until the end point of the dip. For the other 465 dips it is the other way around.

The distributions for the amplitude and the duration are also shown in Figures 5a and 5b.



(a) Distribution of the amplitude (%) of the dips detected in SpO_2

(b) Distribution of the duration(s) of the dips detected in SpO_2

Figure 5: Distribution of amplitude and duration for dips detected in oxygen saturation.

on average the standard deviation of the duration within a patient is 20.4 seconds and the average standard deviation between patients mean duration is 11.8 seconds. On average the standard deviation for the amplitude within a patient is 11.2 % and the average standard deviation between patients mean amplitude is 7.3 %.

The correlation between the amplitude and the duration of the dips is tested with a two-sided Pearson test. The Pearson product-moment correlation coefficient is 0.34 with a corresponding p-value of 0.000. With an α-level of 0.05 we reject the null-hypothesis that states that the correlation coefficient is 0. The joint distribution of amplitude and duration is also shown in Appendix A.

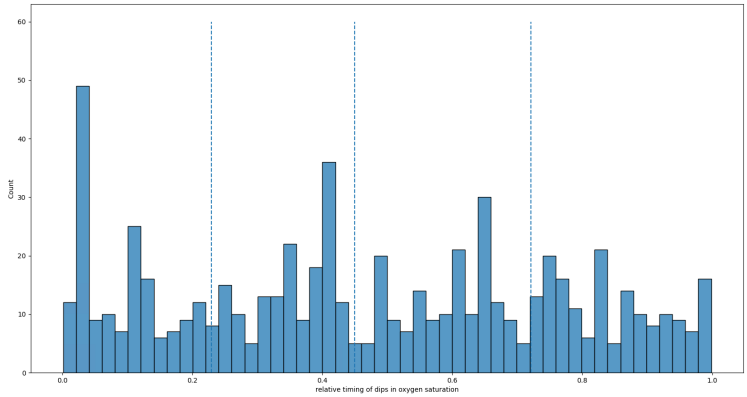


Figure 6: The distribution of the relative timing of the start of a dip in oxygen saturation.

As was also the case for the relative timing of the artifacts due to blood sampling, this histogram does not reveal an apparent pattern. A one-sample Kolmogorov-Smirnov test gives the following result: Kolmogorov-Smirnov statistic is 0.076 with a corresponding p-value of 0.000. With an α-level of 0.05 reject the null-hypothesis that states that the relative timing samples are drawn from a uniform distribution between $[0, 1]$.

Heart rate

For this analysis all 49 different patients were included. The total hours of available observations are 2782.2. On average the total observation time per patient is 56.8 hours (standard deviation of 56.2). With a minimum of 2.1 hours and a maximum of 209.4 hours. For the heart rate parameter there is in general more deviation from the baseline and more sudden dips, that are not artifacts but have a physiological cause. This makes it more difficult to detect artifact dips. Therefor the results for the artifact dips found with the proposed method are bound to be poor and we can say with less confidence that the detected dips are in fact artifacts.

Table 6: mean, standard deviation (std), median and Inter Quartile Range (IQR) for the heart rate residual signal.

	mean(res)	σ	mean(res_d)	σ_d
mean (std)	0.07 (0.44)	3.6 (2.2)	0.00 (0.00)	0.67 (0.33)
median (IQR)	0.01 (-1.2: 0.20)	3.4 (1.80: 4.70)	0.00 (0.00:0.00)	0.62 (0.38: 0.89)

In total 4561 dips are detected that pass the threshold, with $\lambda = 4$.

When also looking at the slope constraint there are 1954 dips that still satisfy. The number of dips that satisfy the derivative constraint and have a duration less then 100 seconds is 1946. 5 patient had no detected dips and the average number of dips per hour is 0.76 (std 0.74) with a median of 0.49 dips per hour and an interquartile range of 0.23: 1.05.

Example of the detected dips are shown in Figures 7, 8 and 9. In general there are three types of dips detected. The first one can be described as a period in the measured signal where the heart rate is quite stable before and after the dip and the dip it self has a large amplitude and a short duration, this is likely an artifact. The second type is not really a dip, but rather a short sensor dysfunction where the measured heart rate is equal to 0. The last type of dips that were detected are dips that are similar for a longer period of time and occur fast after each other. This type is likely not an artifact, but have a physiological cause.

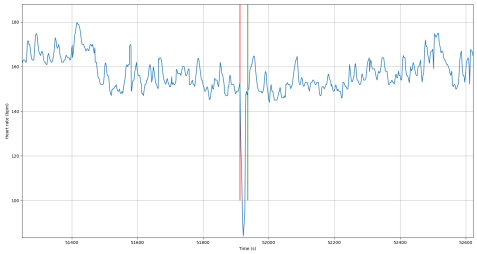


Figure 7: Sudden decrease in HR followed by a rapid decrease back to a normal value. That is due to an artifact.

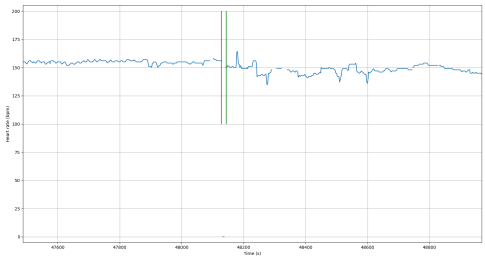


Figure 8: Example of sensor dysfunction in HR. Where HR goes from a normal value to 0.

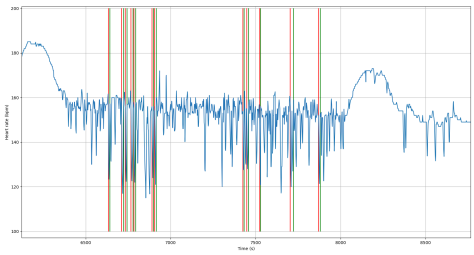


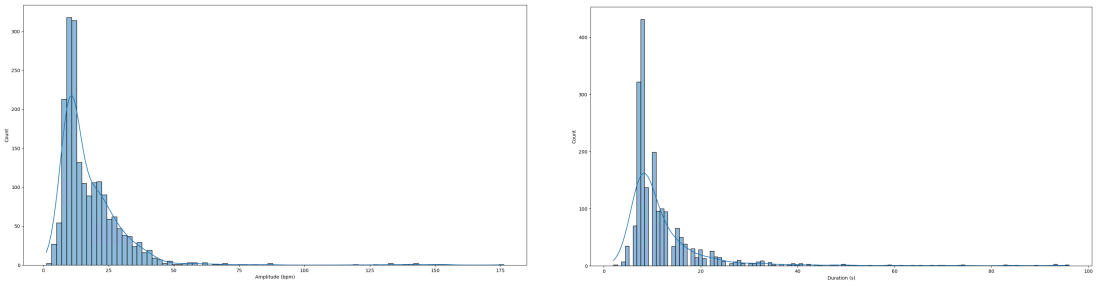
Figure 9: Example of dips where it is unclear whether they are indeed artifacts.

Details about the duration and the amplitude of the detected dips are shown in Table 7.

Table 7: Descriptive statistics of the amplitude and duration of the dips detected in the HR parameter.

	Duration (s)	Amplitude (bpm)	Start value (bpm)	Minimum value (bpm)
mean (std)	12.3 (9.8)	18.0 (14.5)	152 (15.2)	134.3 (19.8)
median (IQR)	9 (8: 13)	13.1 (10.1: 22.2)	150 (144:159)	135.3 (127.8: 143.0)
minimum	2.0	1.2	24	0
maximum	96.0	176	216	200

The mean maximum slope in the dip is 5.8 bpm per seconds with a standard deviation of 6.4. For 1461 dips the slope from the start point of the dip until the minimum is larger than the slope from the minimum point until the end point of the dip. For the other 485 dips it is the other way around. In Figure 10a and 10b the distribution of the amplitude and the duration of the dips are shown.



(a) Distribution of the amplitude (bpm) of the dips detected in HR.

(b) Distribution of the duration (s) of the dips detected in HR.

Figure 10: Distribution of amplitude and duration for dips detected in heart rate.

On average the standard deviation of the duration within a patient is 10.9 seconds and the average standard deviation between patients mean duration is 12.4 seconds. On average the standard deviation for the amplitude within a patient is 9.8 bpm and the average standard deviation between patients mean amplitude is 34.2 bpm. The correlation between the amplitude and the duration of the dips is tested with a two-sided Pearson test. The Pearson product-moment correlation coefficient is 0.43 with a corresponding p-value of 0.000. With an α-level of 0.05 we reject the null-hypothesis that states that the correlation coefficient is 0.

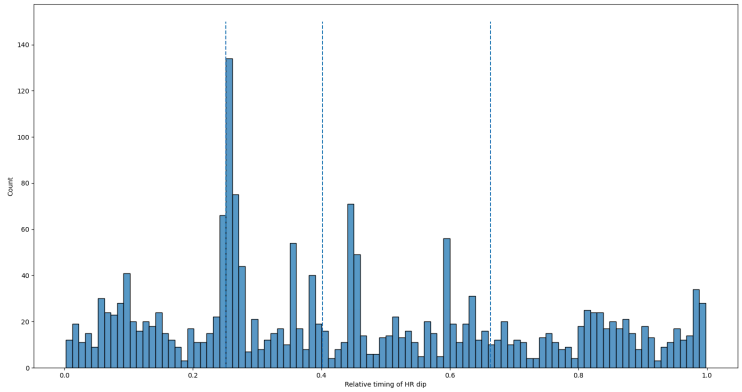


Figure 11: The distribution of the relative timing of the start of a dip in heart rate.

For the relative timing we have again preformed a one-sample Kolmogorov-Smirnov test. The test statistic is 0.12 and with a corresponding p-value of 0.000. Hence, with an α-level of 0.05 we reject the null-hypothesis that states that the relative timing samples are drawn from a uniform distribution between $[0, 1]$.

4 Conclusion

In this project we focused on finding artifacts in real monitoring data from the PICU in Wilhelmina Children's Hospital. We proposed two methods to identity two types of artifacts and described the characteristics of the detected artifacts.

The first method focused on artifacts due to blood sampling. We used a simple method based on the MBP and SBP to detect the blood sample events. These events show a rapid increase in the MBP. To realistically add artificial blood sample events in synthetic data MBP peaks should be added with a average amplitude of 95 mmHg and an average duration of 77 seconds, with on average 0.43 blood sample events per hour. For the amplitude the distribution has two peaks. One between 25 and 50 mmHg and one between 125 and 150 mmHg. It is unclear why there are two peaks. When visually inspecting some of the events where the amplitude was between 25 and 50 mmHg and between 125 and 150 mmHg they appear to have a similar shape. An explanation could be that the MBP peaks with an amplitude between 125 and 150 mmHg are due to the blood sampling events itself and the MBP peaks with an amplitude between 25 and 50 mmHg are due to the flushing of the arterial line. There is no correlation between the duration of the blood sample event and the amplitude of the MBP peak. There is more variation in duration and amplitude of the blood sample event within a patient then between patients. So, the duration and amplitude are not patient specific. Furthermore, it is likely that there is no time dependence for the when the blood sample event happens. So, artificial blood sample events can be added to the synthetic data uniformly in time. Which is different then what we would have expected.

The second proposed method to detect artifacts are for sudden and large dips in oxygen saturation and heart rate. It was found that on average 0.48 artifact dips occur per hour of monitoring data from patient. The dips median measured oxygen saturation is 90.7% at the begin of a dip, which is considered to be normal value for this patient population. The median for the minimum value for measured oxygen saturation during the dip is 79.8 %. The median duration for such a dip is 19 seconds. The average variation for both amplitude and duration of the dip within patient is higher then between patients. Meaning that amplitude and duration dips are not patient specific. There is a positive relation between amplitude and duration. In general the larger the amplitude of the dip, the longer the duration. However, the spread is big so we cannot say this with confidence. The relative timing of the detected oxygen saturation dips are likely not uniformly distributed between $[0, 1]$ indicating that there is a time dependence for the oxygen saturation dips. In the distribution we see that there dips are more frequently found in the beginning of the observations. This is inline with what we expected.

The same dip detection method we used for oxygen saturation is also used for heart rate. The dips that are detected using this method are a mix of artifacts and dips that are cause by physiological changes. The description of the dips is therefore not a description of the artifact in the measured heart rate. On average 0.76 dips occur in hour per patient. Which is more then average number of dips found in oxygen saturation. The median value of the start point of the dip is 150 bpm, which is close to the overall median heart rate. The minimum value is 135. The average standard deviation for both amplitude and duration of the dip within patients is lower then between patients. This indicates that we indeed detected a lot of dips that have a physiological cause and thus have patient specific characteristics. The relative timing of the dips are not uniformly distributed, indicating a time dependence. This is expected, since the clinical condition of a patient is usually worse at the beginning of their PICU stay.

All in all, we have proposed a method to detect different types of artifacts. We provided insights in the characteristics of the artifacts in the vital parameters mean blood pressure, oxygen

saturation and heart rate. These results give a first idea what certain kinds of artifacts, namely blood sample events and dip like artifacts in oxygen saturation an heart rate, look like and what their frequency is in PICU data. And in turn could eventually be used to realistically add artificial artifacts to synthetic data.

5 Discussion

In this Section we will discuss the limitations of this study and make recommendations for future study.

First we will discuss some general limitation of this project. Then we will focus on limitation specific for the method we used. Lastly, we will provide a list of other types of artifacts that could be detected and described to get a better and more complete description of the types of artifacts present in PICU data.

For this project there were no annotations available stating whether a time point is an artifact or not. If this would have been available, the proposed artifact detection method could have been evaluated in more detail. Without this validation of the method, we are not able to say whether using the describing statistics given in this project lead to realistic artificial artifacts for synthetic data. We have asked an expert to check the artifact detection for one randomly selected patient. This patient had a very variable heart rate and loss of signal for the blood pressure. This showed that our method is not robust and still needs to be improved.

Furthermore, we have used methods where it was needed to fix certain constants beforehand. We have set these constants in such a way that after a quick empirical test the results seemed reasonable. it is advised to further look into this and see how much the final result changes. Due to the scope of this project we have not looked at this. Suppose an annotated dataset was available, a random selection of patients could be used to fine tune these parameters (Jakob et al., 2000).

All the artifact detection methods we used performed bad when the signal quality is low. In this project we searched for specific shapes of artifacts. Because of this the detection methods perform bad when the signal quality is low. There are several signal quality indices (SQI's) designed for vital signs (Li, Mark, & Clifford, 2008), (Sun, Reisner, & Mark, 2006) (Zong, Moody, & Mark, 2004). The information of different SQI's can be used to know what kind of artifacts could be present in a segment and a more efficient search for artifacts can be done.

We will now discuss the limitations per artifact detection method.

For the blood sample event detection we looked for time points where the MBP was higher then the SBP and used a window to get characteristics of this blood sample event. The search windows overlap if time points where the MBP is higher then the SBP are closer to each other then 60 seconds. If there was such an overlap the search windows were merged, since we assumed that this would still be due to the same blood sample event. However, by doing this the search windows could get extremely large if a patients records for MBP and SBP were very noisy or there was loss of signal. Especially the estimation of the duration was prone to error because of this. Since, we used the number of missing values as a proxy for the duration. If a search window got very large due to irregularities and missing value in MBP and SBP this also resulted in a very long estimation for the duration. For future research it is advised to use another estimation for the duration of a blood sample event.For example by locating the beginning of the MBP peak and the end of the peak and only use that segment to describe the characteristics.

For the artifact dip detection for oxygen saturation and heart rate we made clear assumption about what an artifact dip looks like. By doing so, we restricted ourselves in our analysis. For example step-changes, high-frequency artifacts and square-wave artifact forms we not taken into account. We also focused on artifacts where there was an decrease in the parameters. Our method can be easily extended to also detect artifacts where there is increase in the parameter.

Especially for heart rate our proposed artifact detection model performed badly, since there are dips in the heart rate parameter that satisfy all our artifact criteria but are in fact not artifacts. We have prevented to detect a lot of dips by increasing the sensitivity constant, λ_d, however because of this we also missed artifacts and still labeled dips as artifacts that are likely not artifacts. It is advised to look into other methods to detect artifacts in the heart rate parameter. For example by looking into unsupervised pattern recognition methods, for example clustering, using more complex filtering methods, such as a Kalman filter for example (Li et al., 2008) or a decision tree (Tsien et al., n.d.). If an annotated dataset would be available in the future using supervised pattern recognition method could be promising (Chen et al., 2016) (Maleczek et al., 2024).

The aim of this project was get insight in irregularities in real vital parameter data of PICU patients. We choose to look for certain irregularities. Below we provide a summary of other irregularities, including but not limited to, that can be investigated in the real data to enrich the synthetic data:

- Missing data: Sensors can be removed from the patient if for example the parents want to hold them. This results in large periods of missing data. These periods could be described in duration, inter arrival time and whether they are time specific.
- In this project we looked whether the occurrence of an artifact was time dependent by looking at the relative timing of the event. It is recommended to also look whether artifacts are dependent on the time of the day or night.
- Different shapes of artifacts, such as high frequency noise, step changes, loss of signal, peaks.

References

Charbonnier, S., & Portet, F. (2012). A self-tuning adaptive trend extraction method for process monitoring and diagnosis. *Journal of Process Control*, *22*(6), 1127–1138. Retrieved from http://dx.doi.org/10.1016/j.jprocont.2012.03.010 doi: 10.1016/j.jprocont.2012.03.010

Chen, L., Dubrawski, A., Wang, D., Fiterau, M., Guillame-Bert, M., Bose, E., ... Hravnak, M. (2016, 7). Using Supervised Machine Learning to Classify Real Alerts and Artifact in Online Multisignal Vital Sign Monitoring Data. *Critical care medicine*, *44*(7), 456–63. doi: 10.1097/CCM.0000000000001660

Drew, B. J., Harris, P., Zègre-Hemsey, J. K., Mammone, T., Schindler, D., Salas-Boni, R., ... Hu, X. (2014). Insights into the problem of alarm fatigue with physiologic monitor devices: A comprehensive observational study of consecutive intensive care unit patients. *PLoS ONE*, *9*(10). doi: 10.1371/journal.pone.0110274

Du, C. H., Glick, D., & Tung, A. (2019, 6). Error-checking intraoperative arterial line blood pressures. *Journal of clinical monitoring and computing*, *33*(3), 407–412. doi: 10.1007/s10877-018-0167-7

Hever, G., Cohen, L., O'Connor, M. F., Matot, I., Lerner, B., & Bitan, Y. (2020). Machine learning applied to multi-sensor information to reduce false alarm rate in the ICU. *Journal*

of Clinical Monitoring and Computing, *34*(2), 339–352. Retrieved from https://doi.org/10.1007/s10877-019-00307-x doi: 10.1007/s10877-019-00307-x

Hoorweg, A.-L. J., Pasma, W., van Wolfswinkel, L., & de Graaff, J. C. (2018, 2). Incidence of Artifacts and Deviating Values in Research Data Obtained from an Anesthesia Information Management System in Children. *Anesthesiology*, *128*(2), 293–304. doi: 10.1097/ALN.0000000000001895

Jakob, S., Korhonen, I., Ruokonen, E., Virtanen, T., Kogan, A., & Takala, J. (2000). Detection of artifacts in monitored trends in intensive care. *Computer Methods and Programs in Biomedicine*, *63*(3), 203–209. doi: 10.1016/S0169-2607(00)00110-3

Li, Q., Mark, R. G., & Clifford, G. D. (2008). Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiological Measurement*, *29*(1), 15–32. doi: 10.1088/0967-3334/29/1/002

Mäkivirta, A., Koski, E., Kari, A., & Sukuvaara, T. (1991). The median filter as a preprocessor for a patient monitor limit alarm system in intensive care. *Computer Methods and Programs in Biomedicine*, *34*(2-3), 139–144. doi: 10.1016/0169-2607(91)90039-V

Maleczek, M., Laxar, D., Kapral, L., Kuhrn, M., Abulesz, Y.-T., Dibiasi, C., & Kimberger, O. (2024, 7). A Comparison of Five Algorithmic Methods and Machine Learning Pattern Recognition for Artifact Detection in Electronic Records of Five Different Vital Signs: A Retrospective Analysis. *Anesthesiology*, *141*(1), 32–43. doi: 10.1097/ALN.0000000000004971

Ray, S., Rogers, L., Raman, S., & Peters, M. J. (2017). Liberal oxygenation in paediatric intensive care: retrospective analysis of high-resolution SpO2 data. *Intensive Care Medicine*, *43*(1), 146–147. doi: 10.1007/s00134-016-4606-y

Sun, J. X., Reisner, A. T., & Mark, R. G. (2006). A signal abnormality index for arterial blood pressure waveforms. *Computers in Cardiology*, *33*, 13–16.

Tsien, C. L., Kohane, I. S., Mcintosh, N., Program, H. I., Ma, B., & Infirmary, R. (n.d.). Building ICU Artifact Detection Models With More Data in Less Time BWH / MGH Harvard Affiliated Emergency Medicine Residency , Boston MA . b Children ' s.

Zong, W., Moody, G. B., & Mark, R. G. (2004, 9). Reduction of false arterial blood pressure alarms using signal quality assessment and relationships between the electrocardiogram and arterial blood pressure. *Medical & biological engineering & computing*, *42*(5), 698–706. doi: 10.1007/BF02347553

Appendix A Joint distribution for amplitude and duration

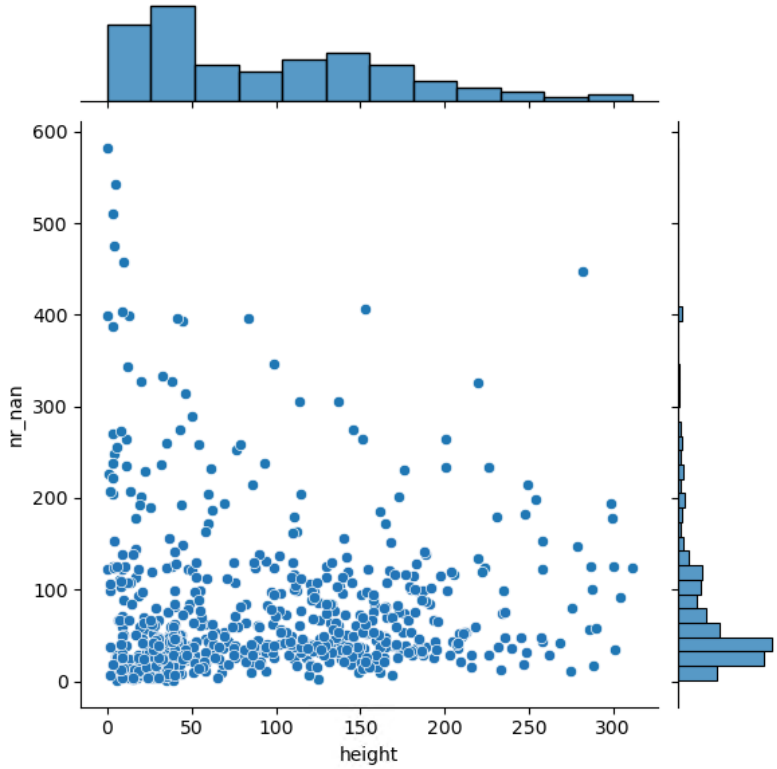


Figure 12: Joint distribution of amplitude and duration for blood sample event.

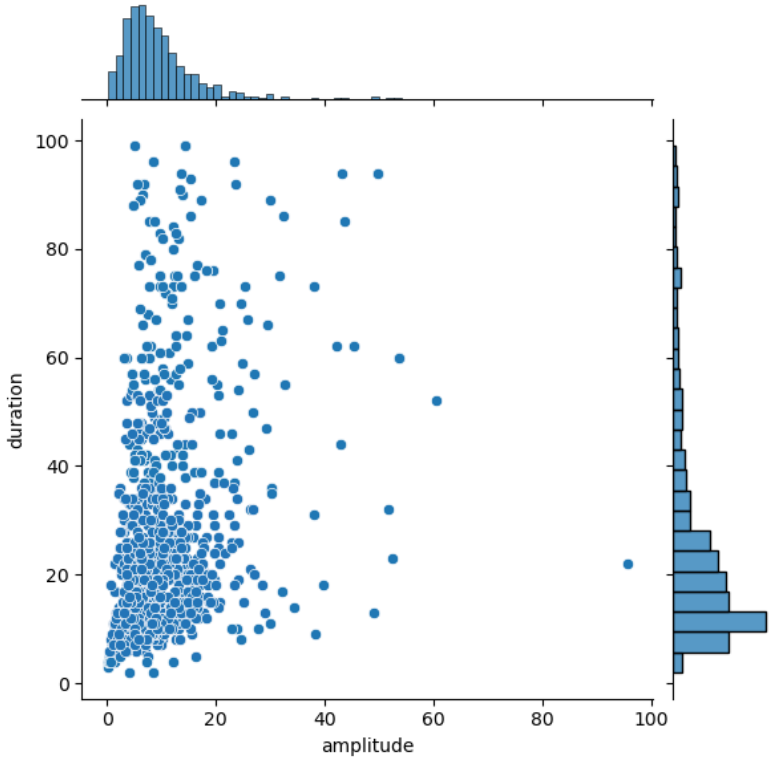


Figure 13: Joint distribution of amplitude and duration for oxygen saturation dips.

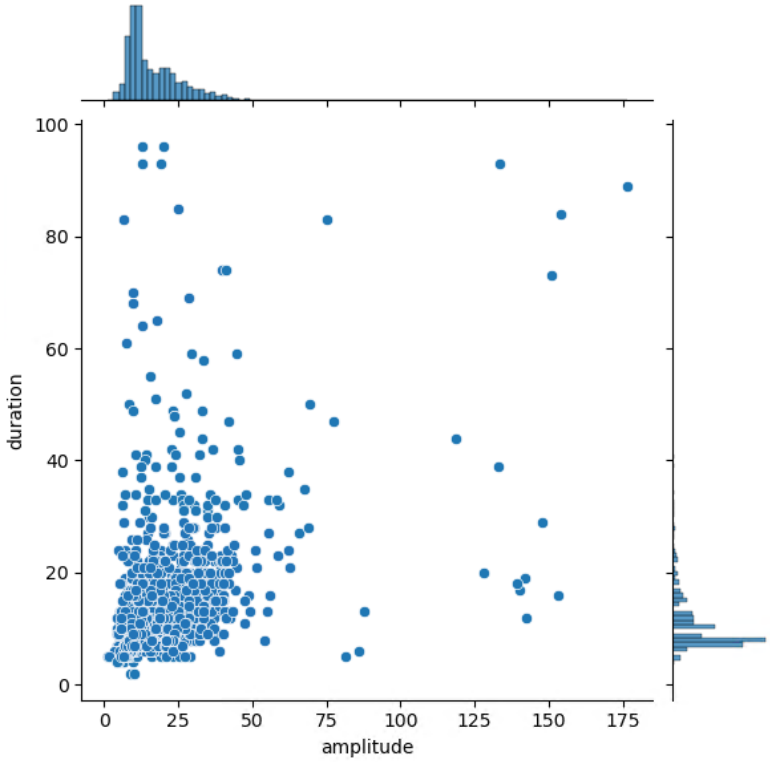


Figure 14: Joint distribution of amplitude and duration for heart rate

Appendix B Validation artifacts

Below we have included the documents used for the validation of our method. Note that these are in Dutch.

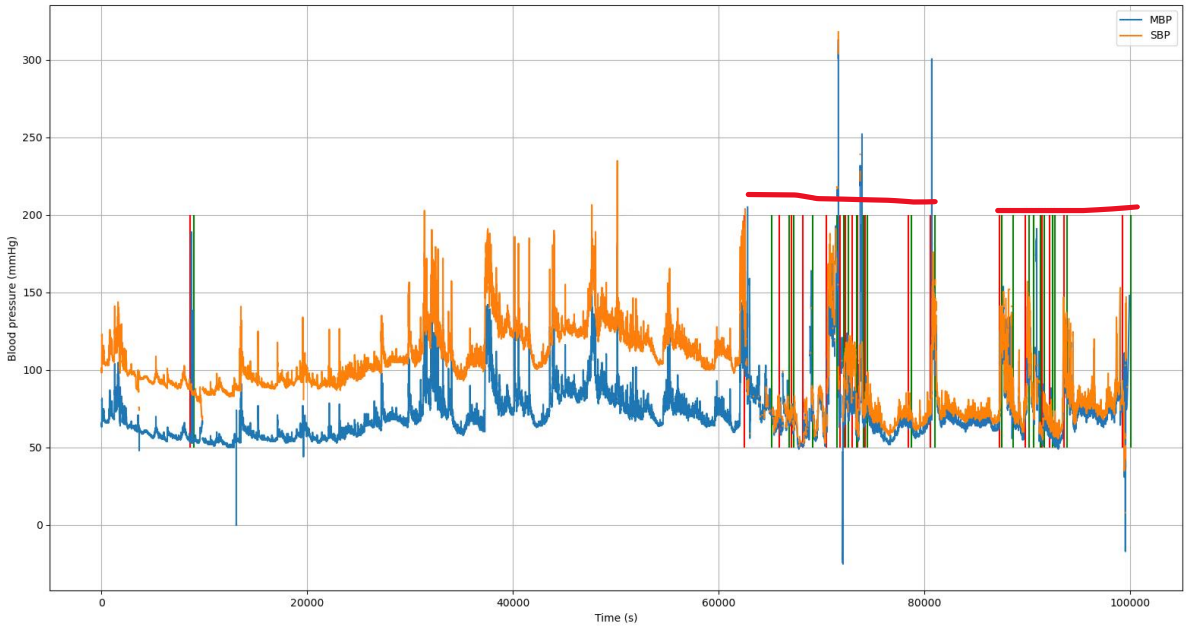
We will give an explanation of the document set-up. Plots of the parameters are shown (MBP and SBP for the blood sample events, oxygen saturation and heart rate). First the total observations are shown and then we zoomed into certain segments. We have in total 30.1 hours of observations and zoomed into segments of 5.5 hours. For each detected artifact the expert had to indicated whether or not he agreed that this was an artifact, by making yes or no boldface.(This is the sentence: Dip x is juist als artefact gedetecteerd: ja / nee). After indicating in a segment whether the detected artifacts were correct, the expert indicated whether artifacts were missed and if so at which time stamp (this is done in red).

For the blood sample events time periods are marked red where there is a loss of signal. All the blood sample events that were detected here are likely false.

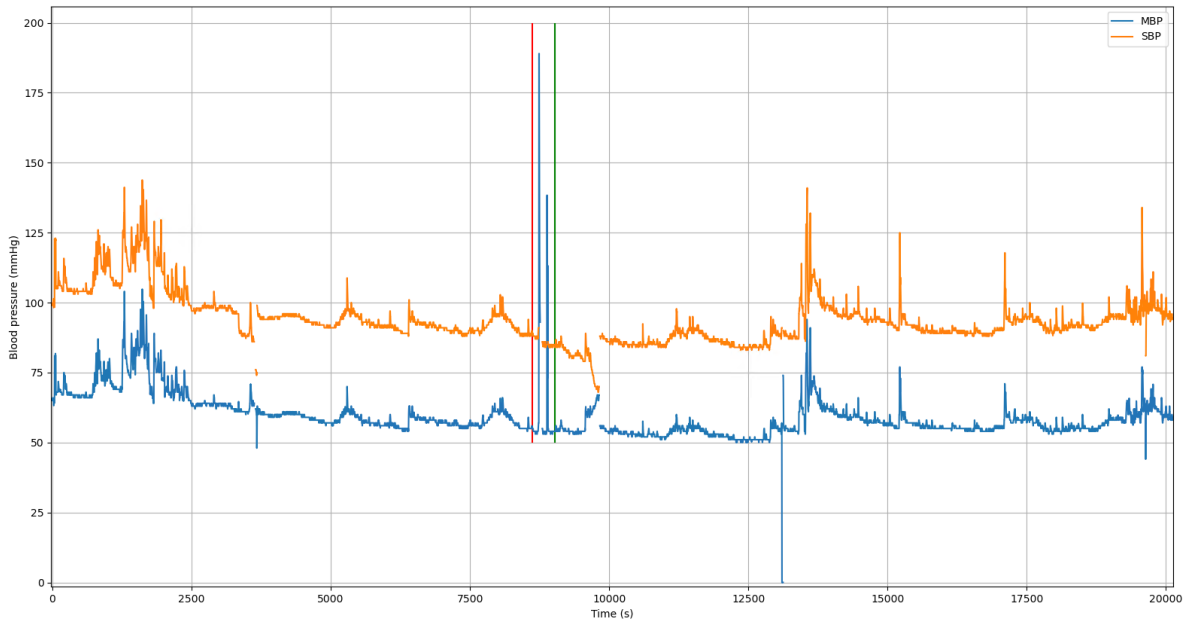
For heart rate the expert indicated with red blocks where there are likely artifacts.

Total observaties gemiddelde bloeddruk en systole bloeddruk

(Rood heb ik gemarkeerd omdat hier het signaal erg rommelig is) Window 4 en 5

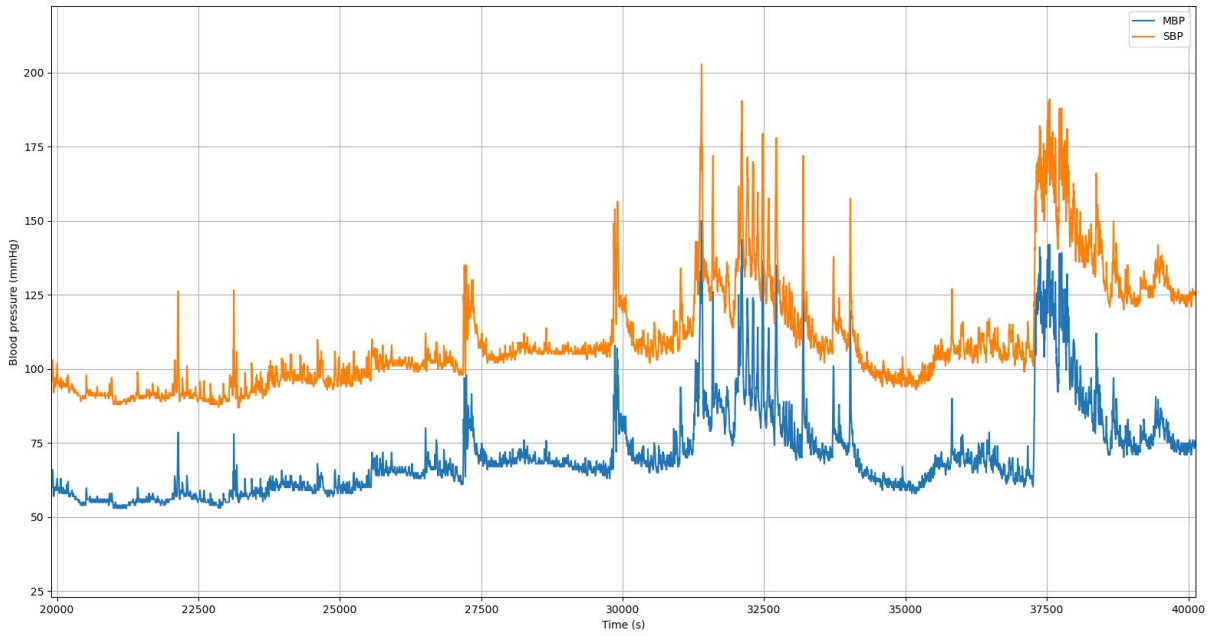


Window 1

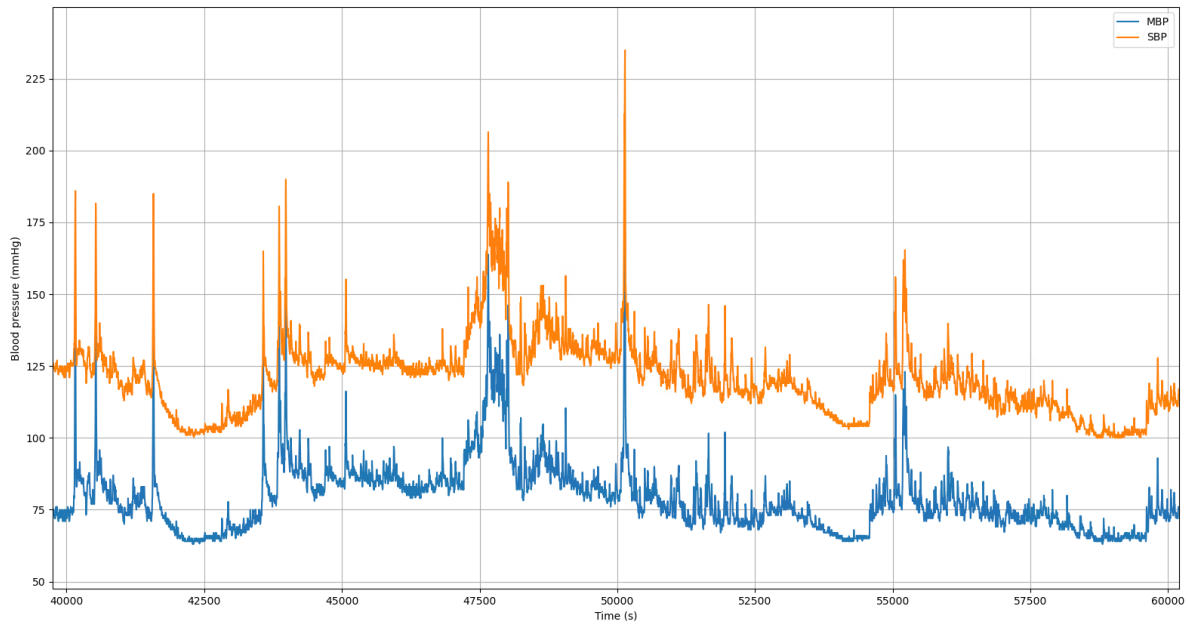


Juist als bloed afname gedetecteerd:**ja** / nee

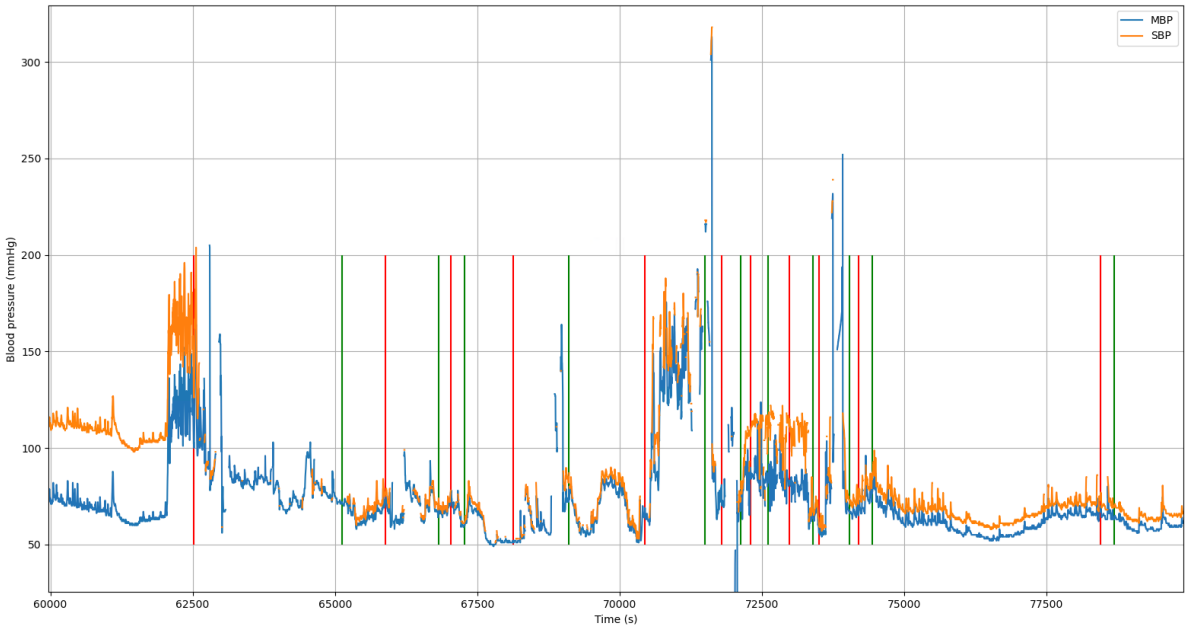
Window 2, niks



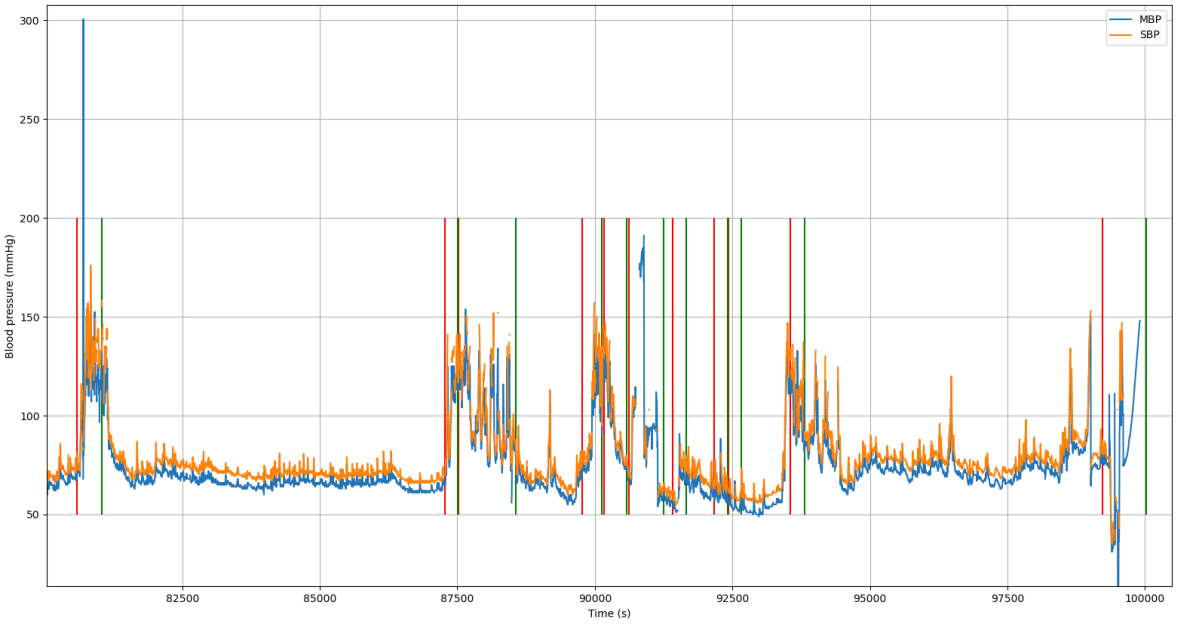
Window 3, niks



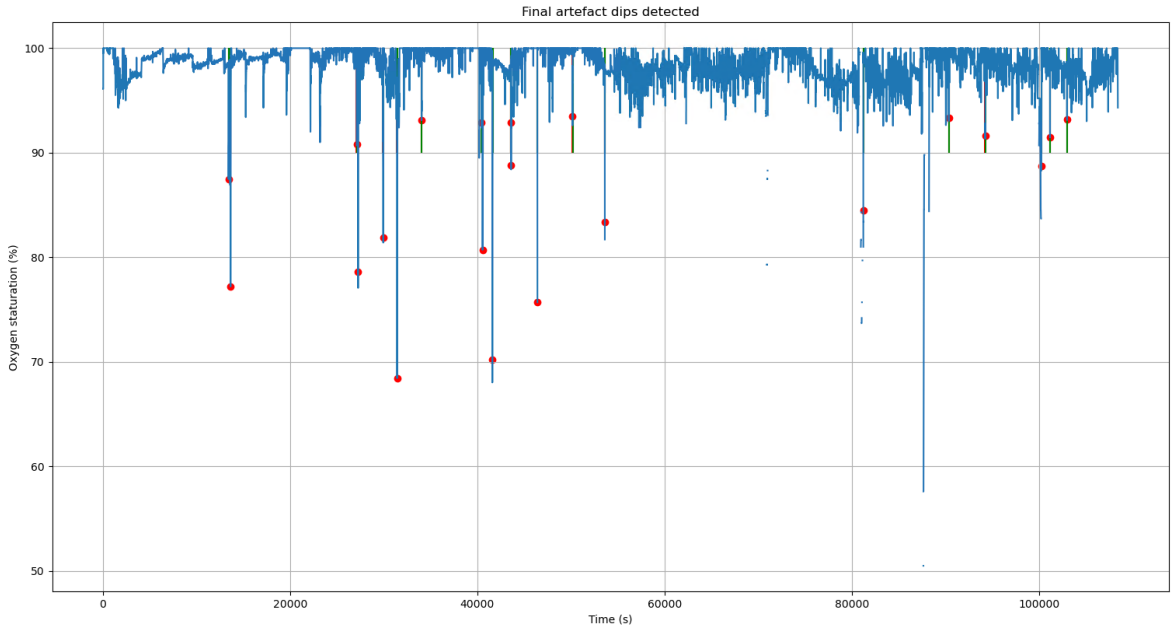
Windw 4 Fout vanwege andere artefacten



Window 5 Fout vanwege andere artefacten

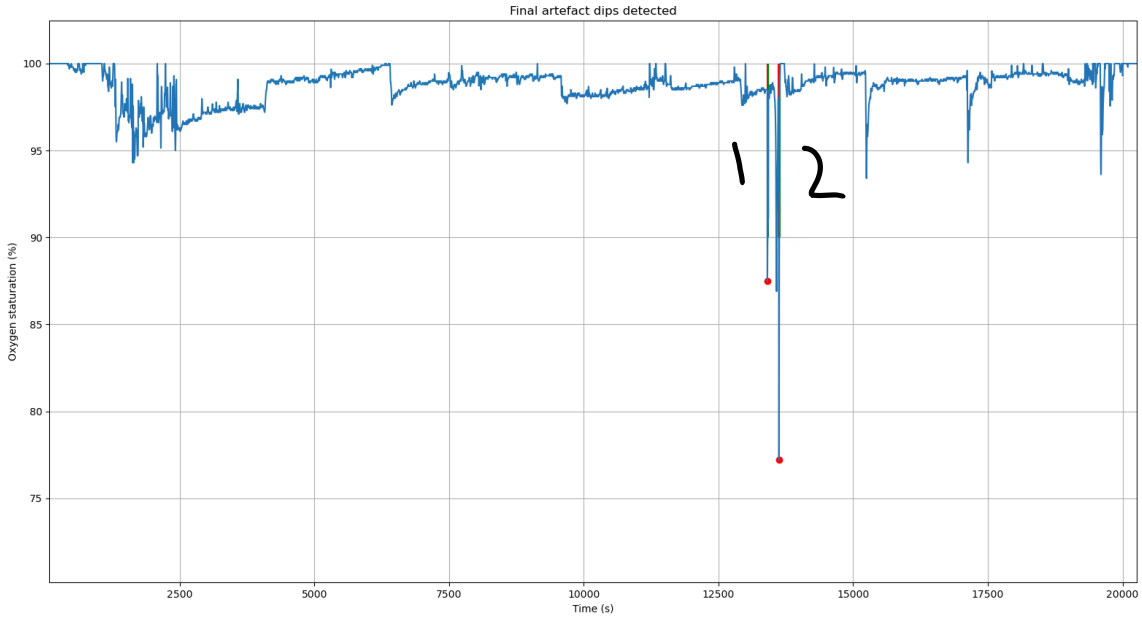


Alle observaties van patient met pseudo_id 12 voor zuurstof saturatie met dips

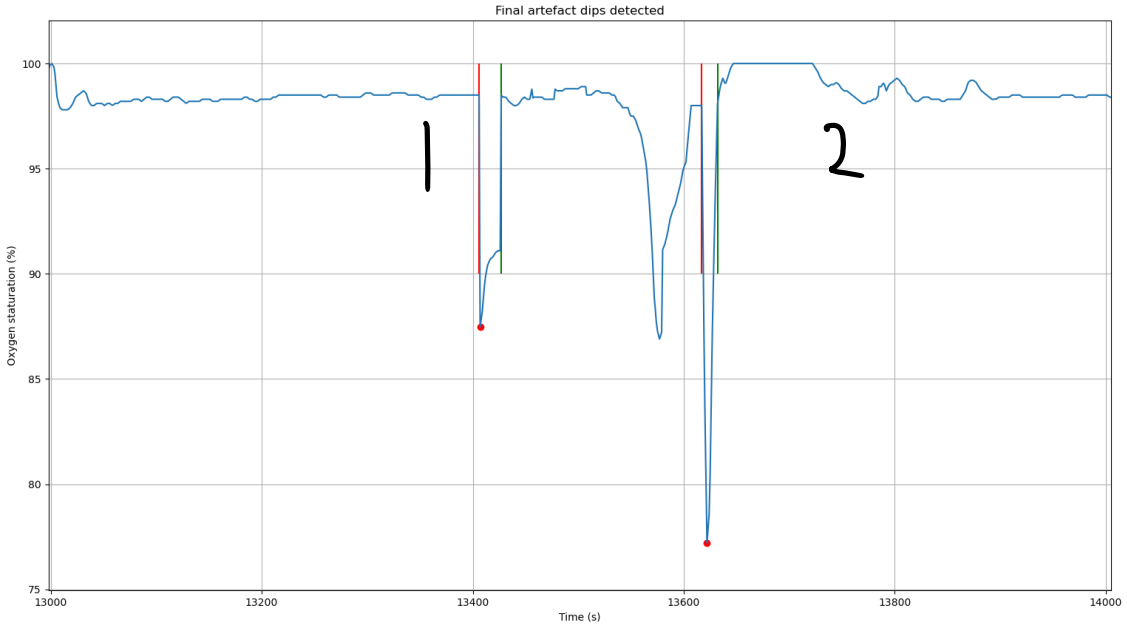


21 artefacten in 30.1 uur aan observaties

Window 1, nummer artefacten gedetecteerd: 2



Zoom op artefacten in window 1

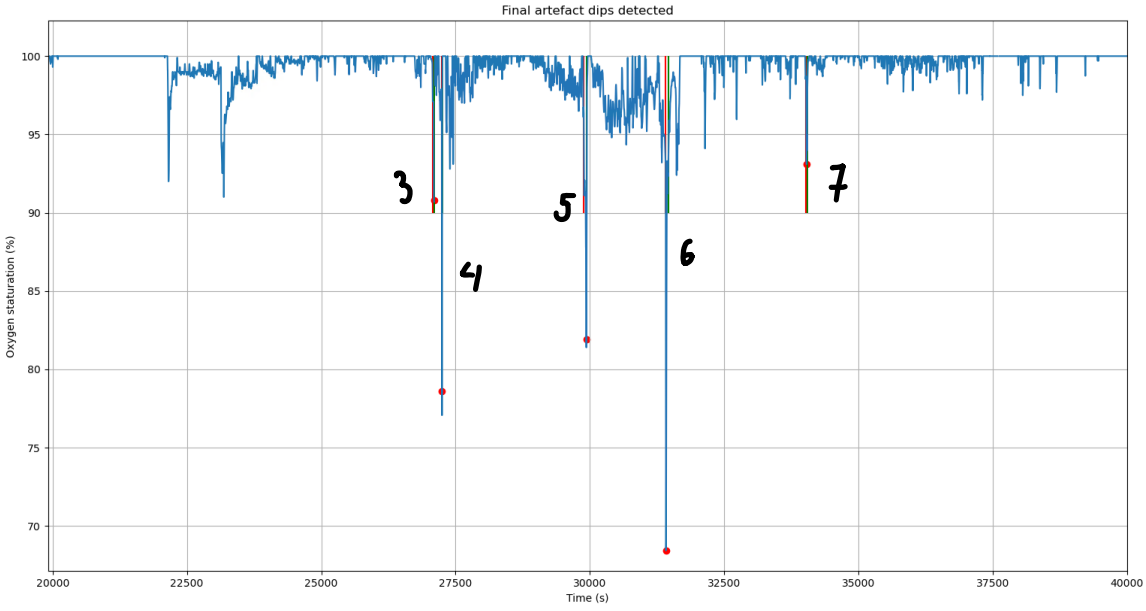


Dip 1 is juist als artefact gedetecteerd: **ja** / nee
Dip 2 is juist als artefact gedetecteerd: **ja** / nee

Er zijn artefacten gemist in window 1: **ja** / nee

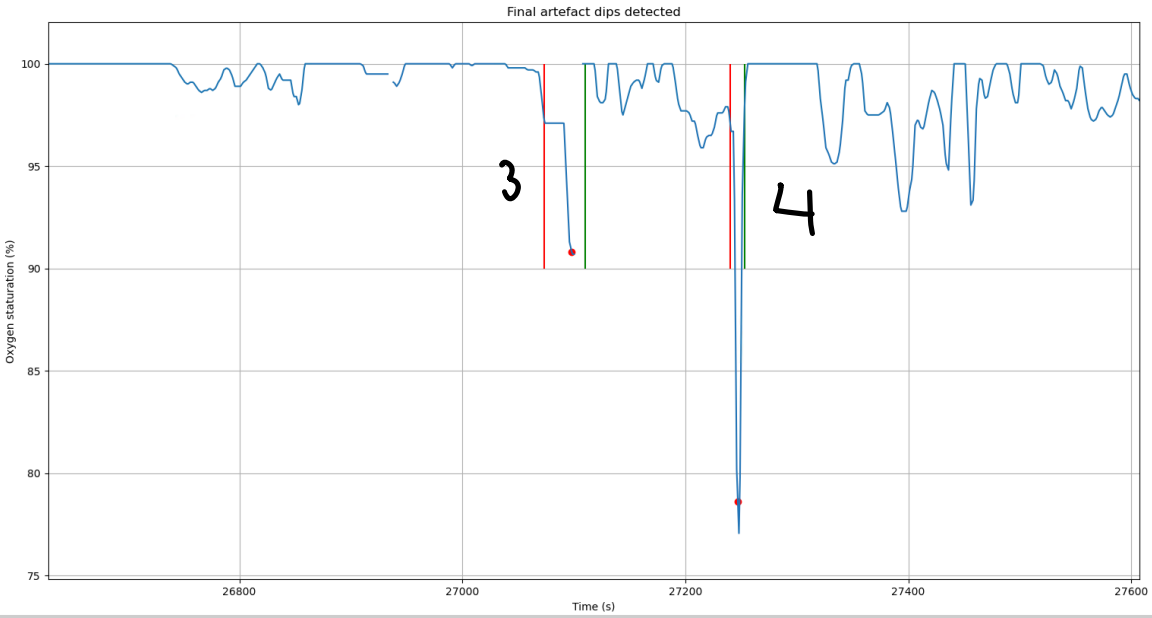
Zo, ja: ongeveer op tijdspunt:**15.100 & 17.000 & 19.000**

Window 2, nummer artefacten gedetecteerd: 5



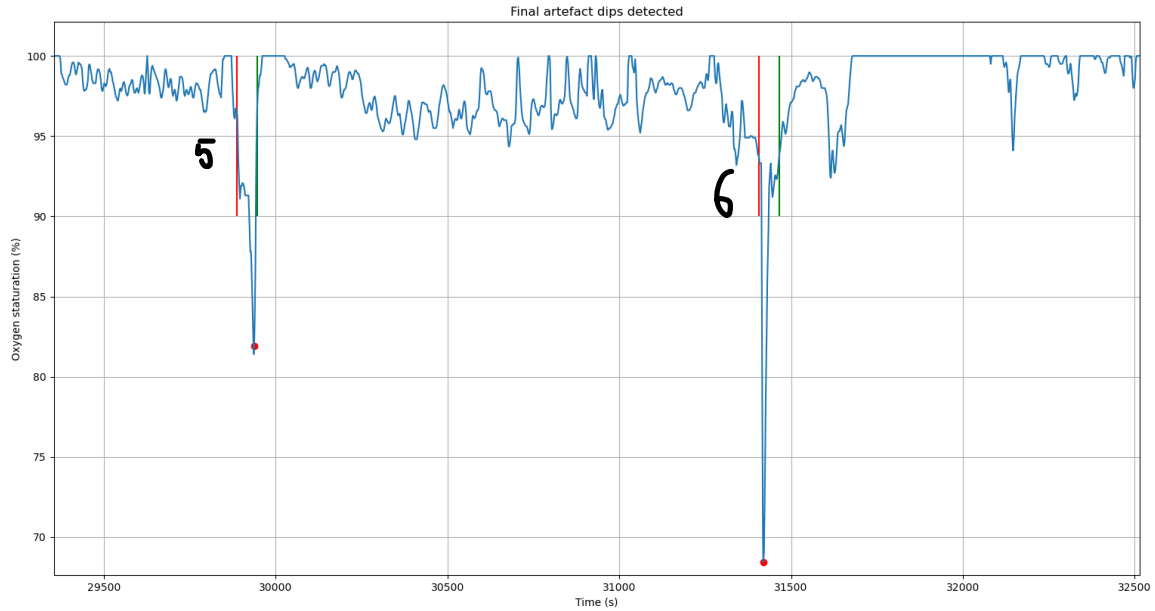
Zoom op artefacten in window 2

Eerste 2 dips



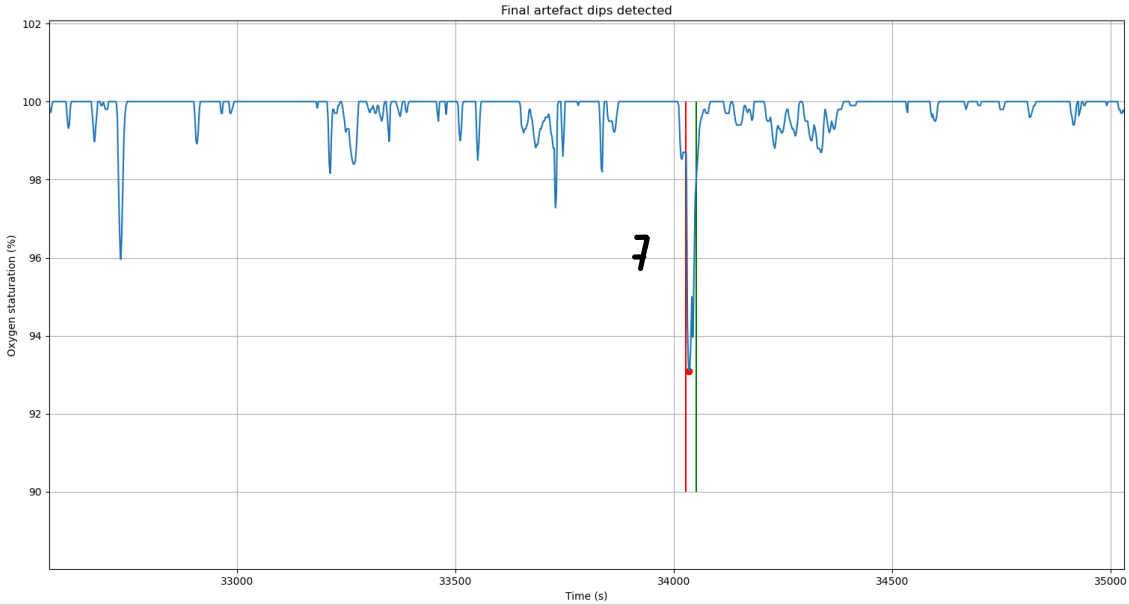
Dip 3 is juist als artefact gedetecteerd: **ja** / nee
Dip 4 is juist als artefact gedetecteerd: **ja** / nee

Zoom dip 5 en 6



Dip 5 is juist als artefact gedetecteerd: **ja** / nee
Dip 6 is juist als artefact gedetecteerd: **ja** / nee

Zoom op dip 7:

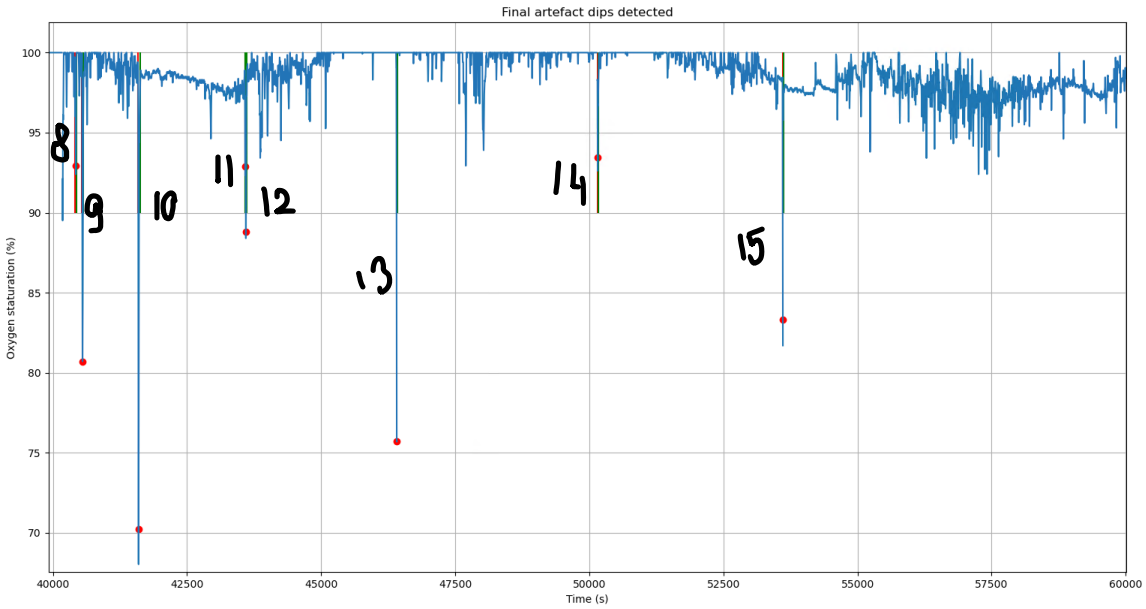


Dip 7 is juist als artefact gedetecteerd: **ja** / nee

Er zijn artefacten gemist in window 2: ja / **nee**

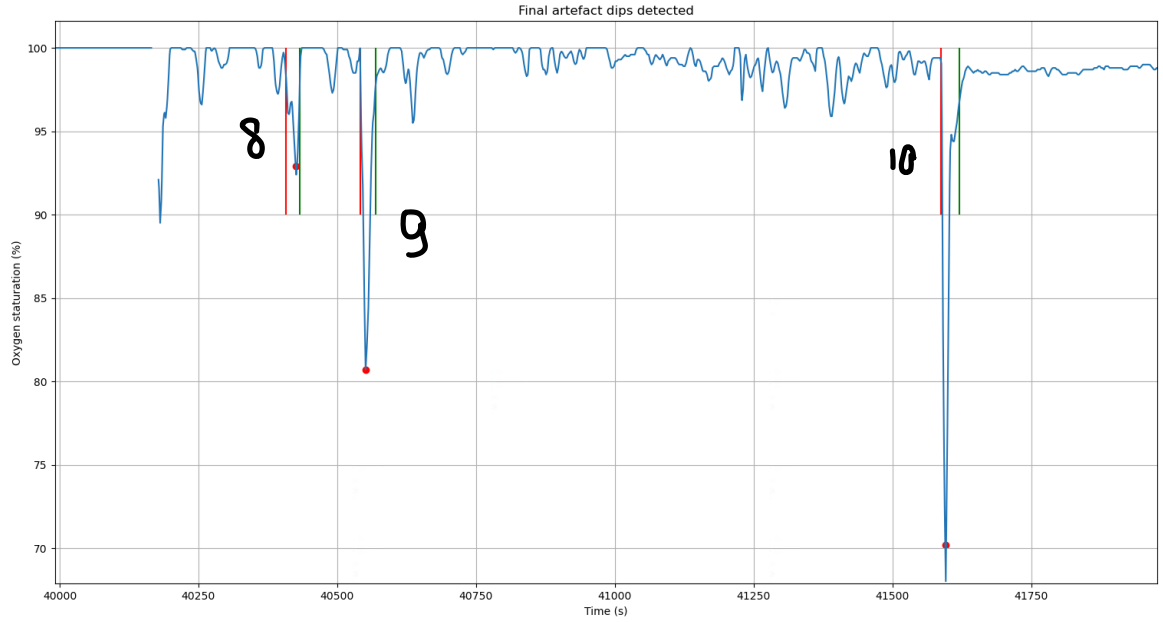
Zo, ja: ongeveer op tijdspunt:

Window 3, nummer artefacten gedetecteerd: 8



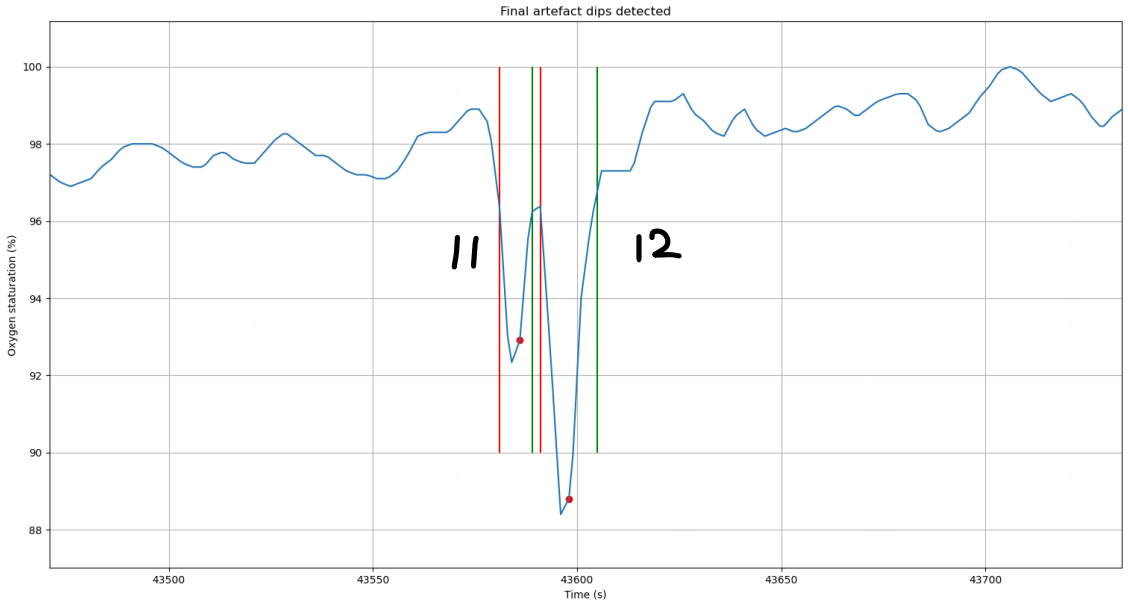
Zoom op artefacten in window 2

Zoom op eerste 3 dips van de window



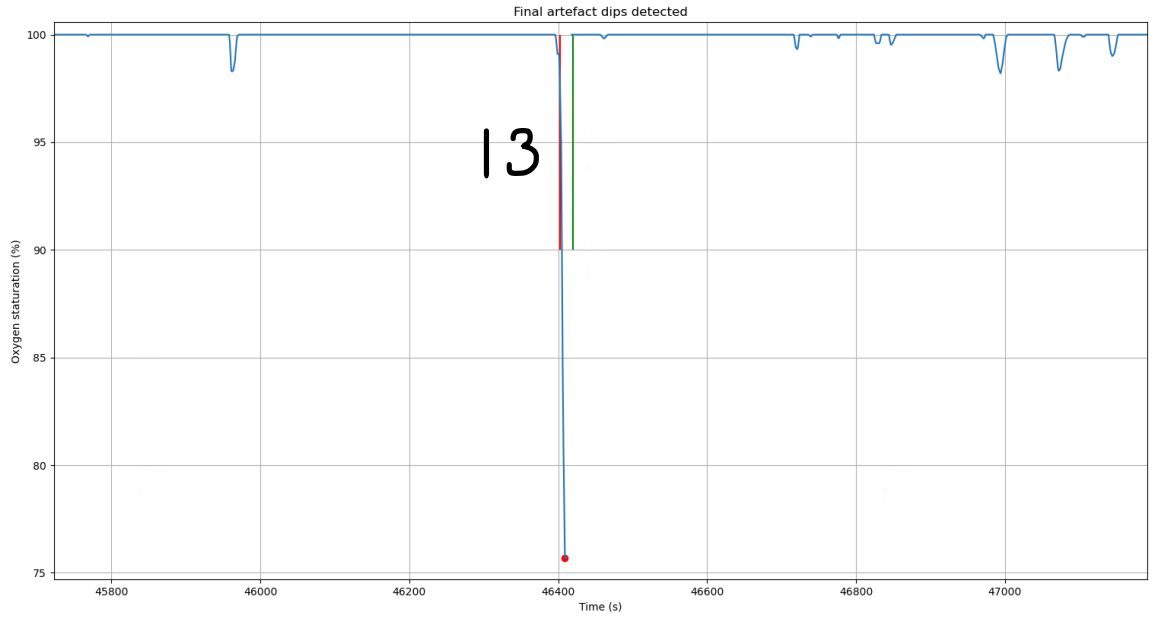
Dip 8 is juist als artefact gedetecteerd: ja / **nee**
Dip 9 is juist als artefact gedetecteerd: **ja** / nee
Dip 10 is juist als artefact gedetecteerd: **ja** / nee

Zoom op dip 4de en 5de dip in



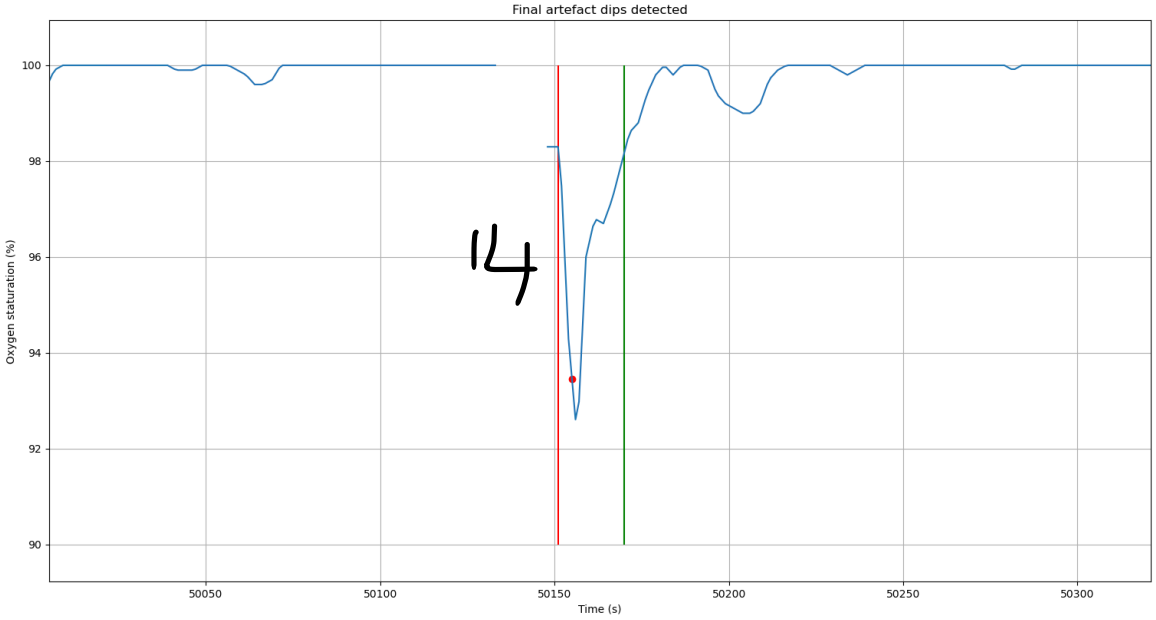
Dip 11 is juist als artefact gedetecteerd: ja / **nee**
Dip 12 is juist als artefact gedetecteerd: **ja /** nee

Zoom op dip 13



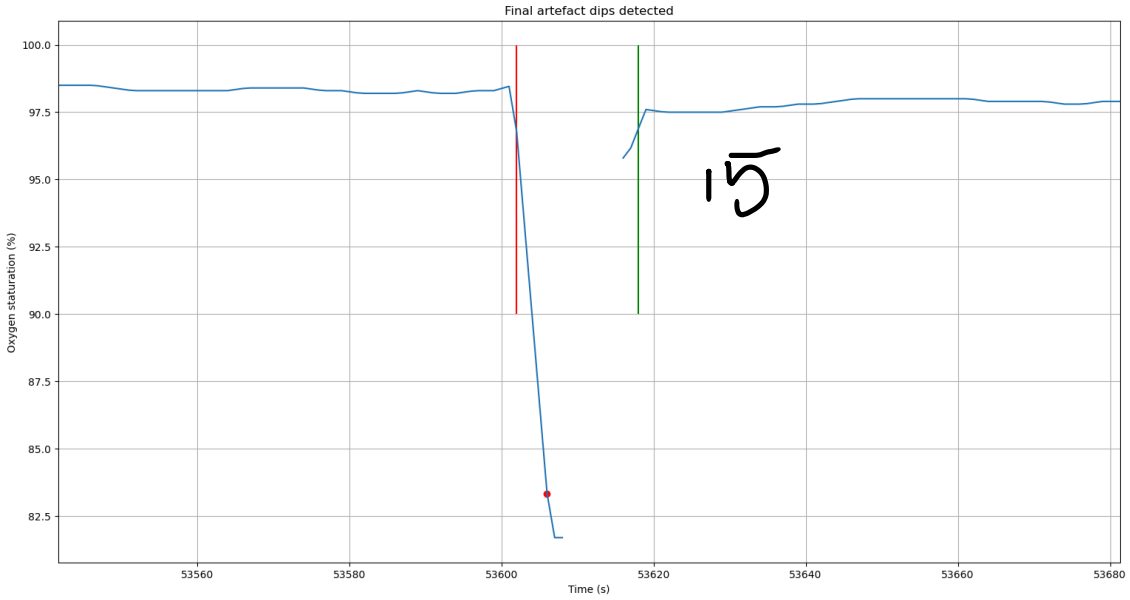
Dip 13 is juist als artefact gedecteerd: **ja** / nee

Zoom op dip 14



Dip 14 is juist als artefact gedetecteerd: **ja** / nee

Zoom op dip 15

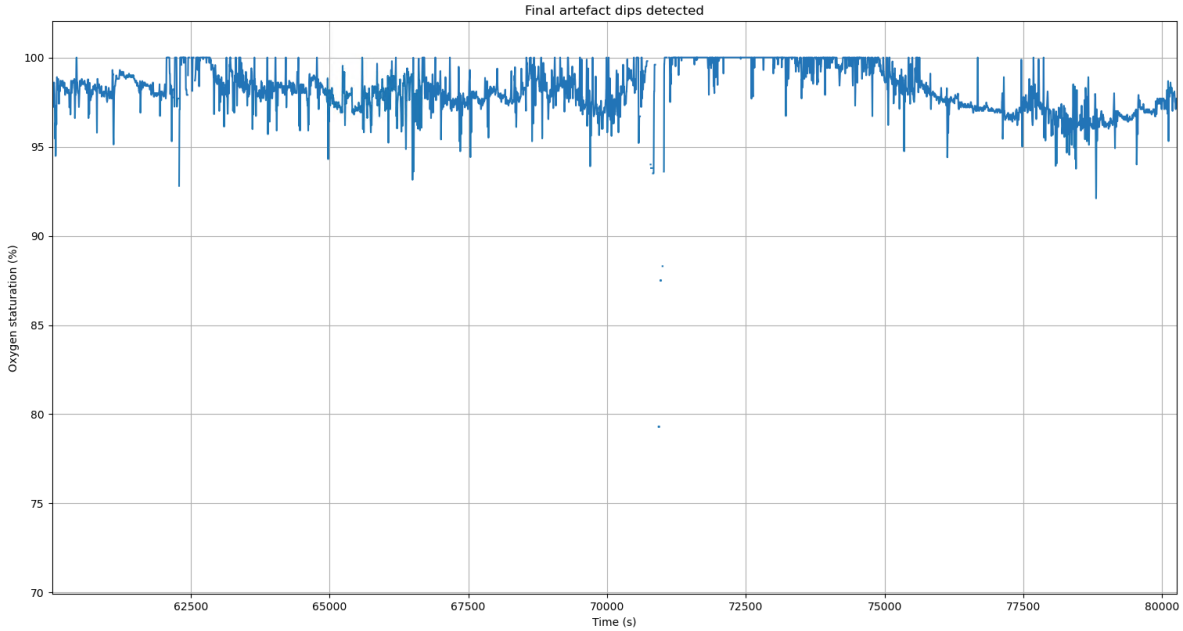


Dip 15 is juist als artefact gedetecteerd: **ja** / nee

Er zijn artefacten gemist in window 3: ja / **nee**

Zo, ja: ongeveer op tijdspunt:

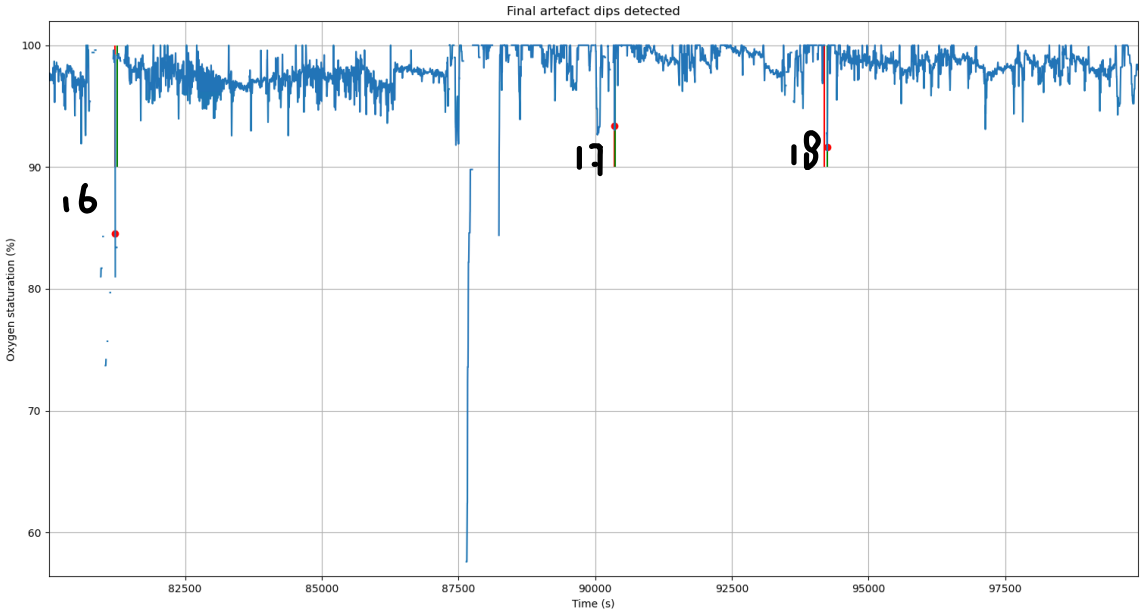
Window 4, 0 artefacten gedetecteerd



Er zijn artefacten gemist in window 4: ja / **nee**

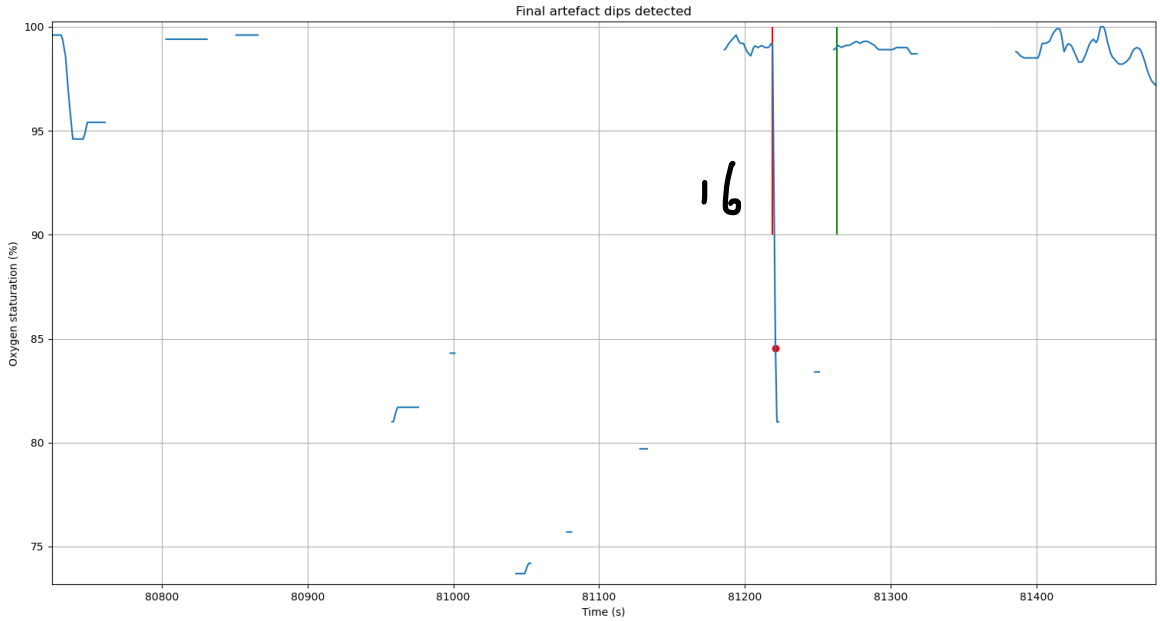
Zo, ja: ongeveer op tijdspunt:

Window 5, nummer artefacten gedetecteerd: 3



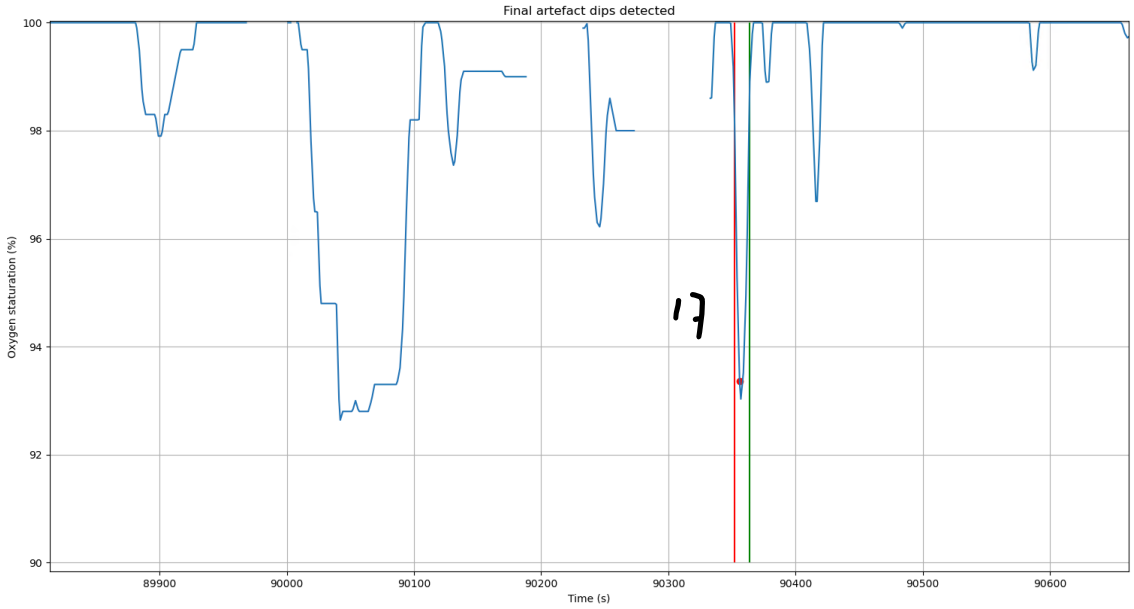
Zoom op artefacten

Dip 16:



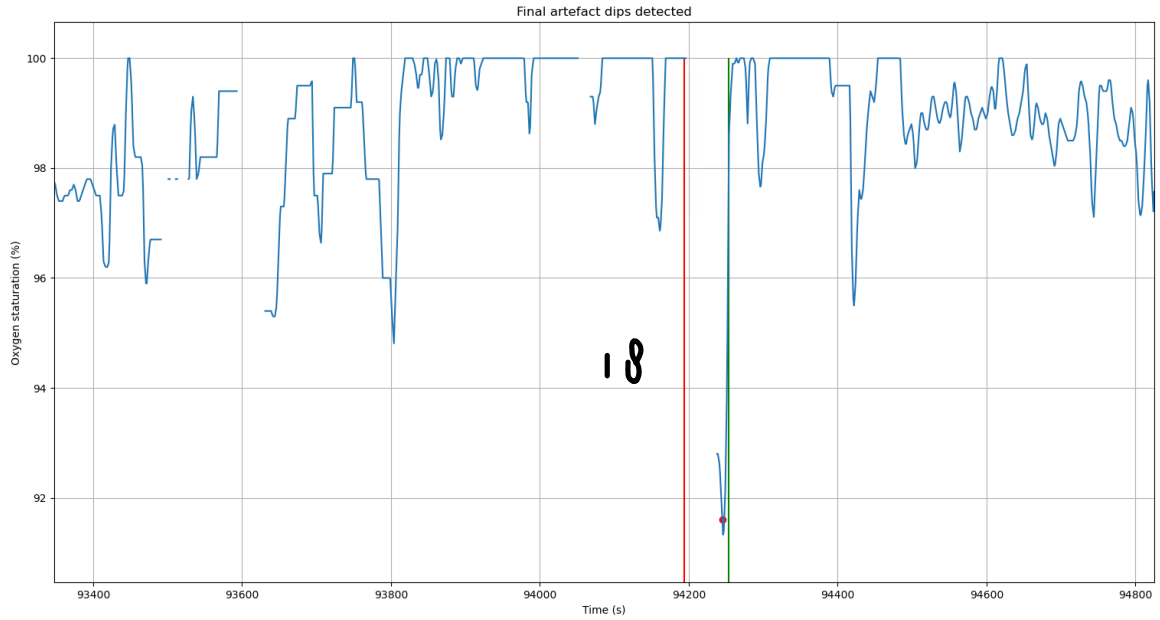
Dip 16 is juist als artefact gedetecteerd: **ja** / nee

Dip 17:



Dip 17 is juist als artefact gedetecteerd: **ja** / nee

Dip 18:

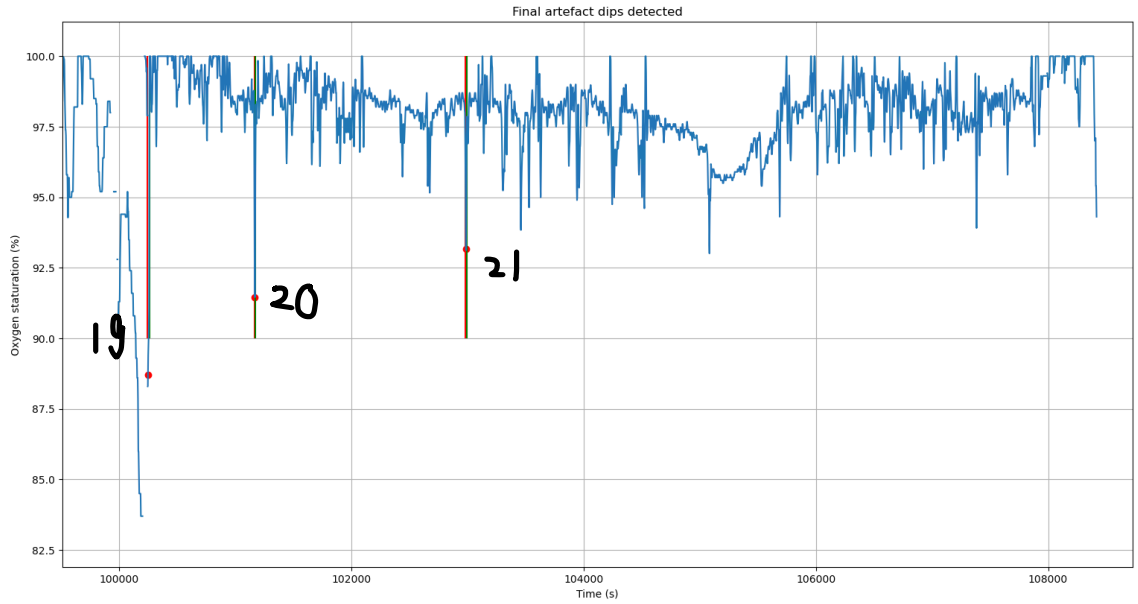


Dip 18 is juist als artefact gedetecteerd: **ja** / nee

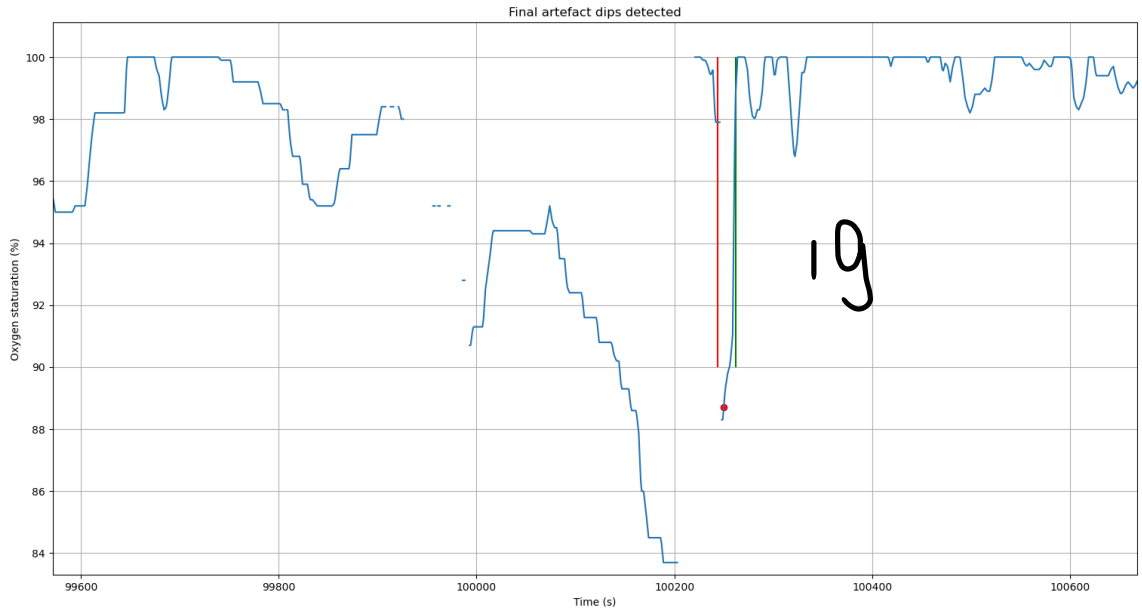
Er zijn artefacten gemist in window 5: **ja** / nee

Zo, ja: ongeveer op tijdspunt: **Voorafgaand aan dip 16 een rij (vermoedelijk door Loss of Signal). Ook enkelen rond 87.500.**

Window 6, nummer artefacten gedetecteerd: 3

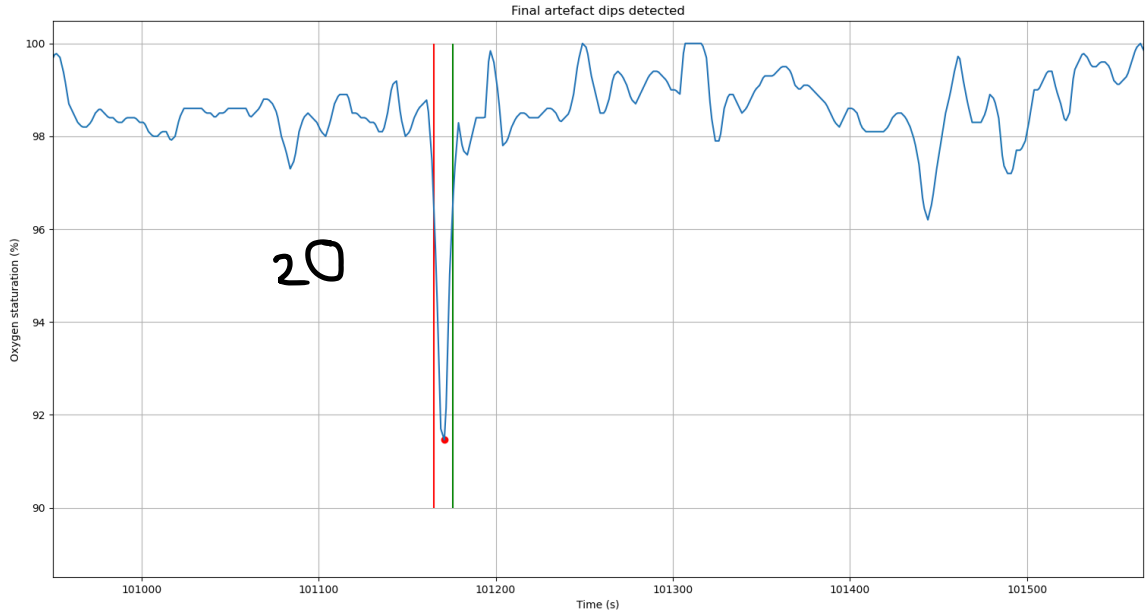


Dip 19:



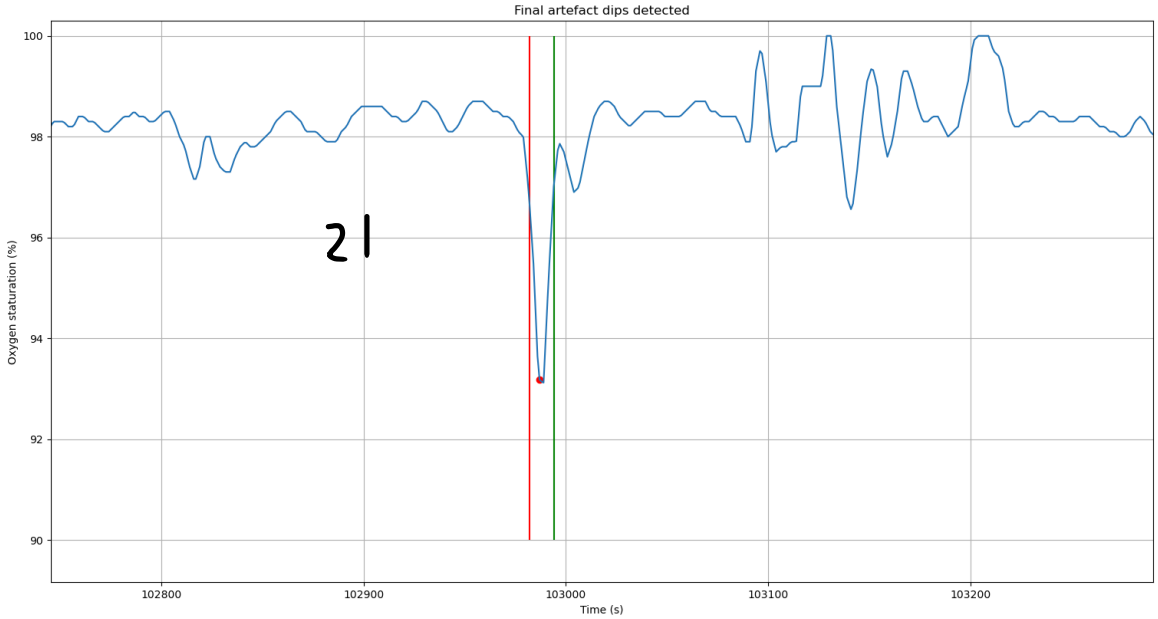
Dip 19 is juist als artefact gedetecteerd: **ja** / nee

Dip 20:



Dip 20 is juist als artefact gedetecteerd: **ja** / nee

Dip 21:

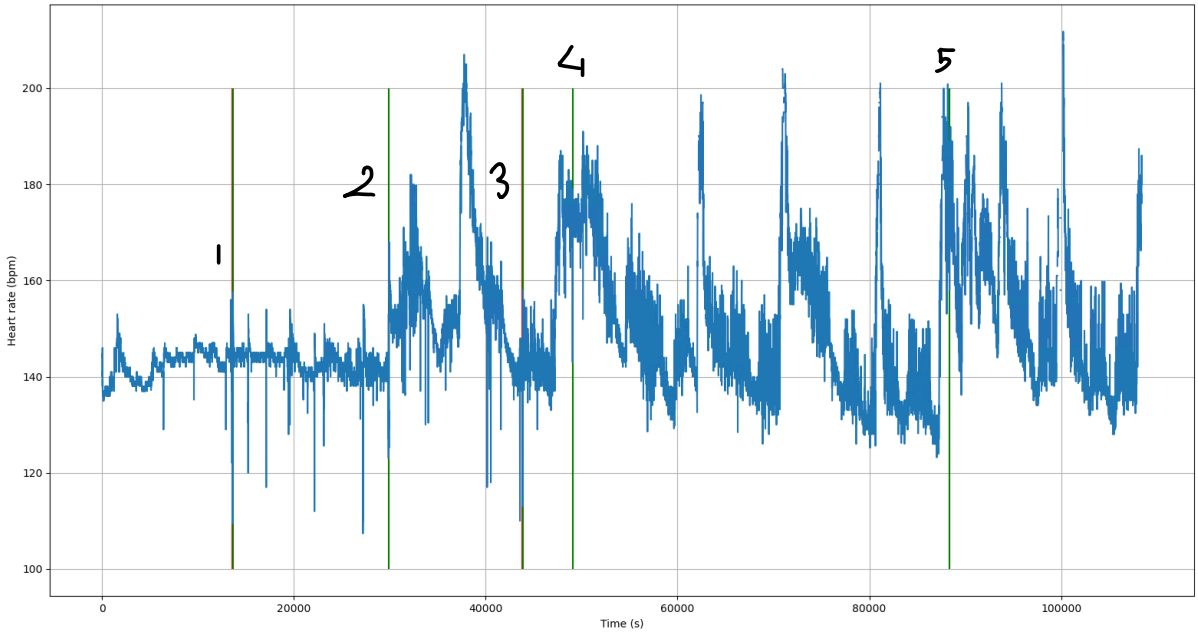


Dip 21 is juist als artefact gedetecteerd: **ja** / nee

Er zijn artefacten gemist in window 6: **ja** / nee

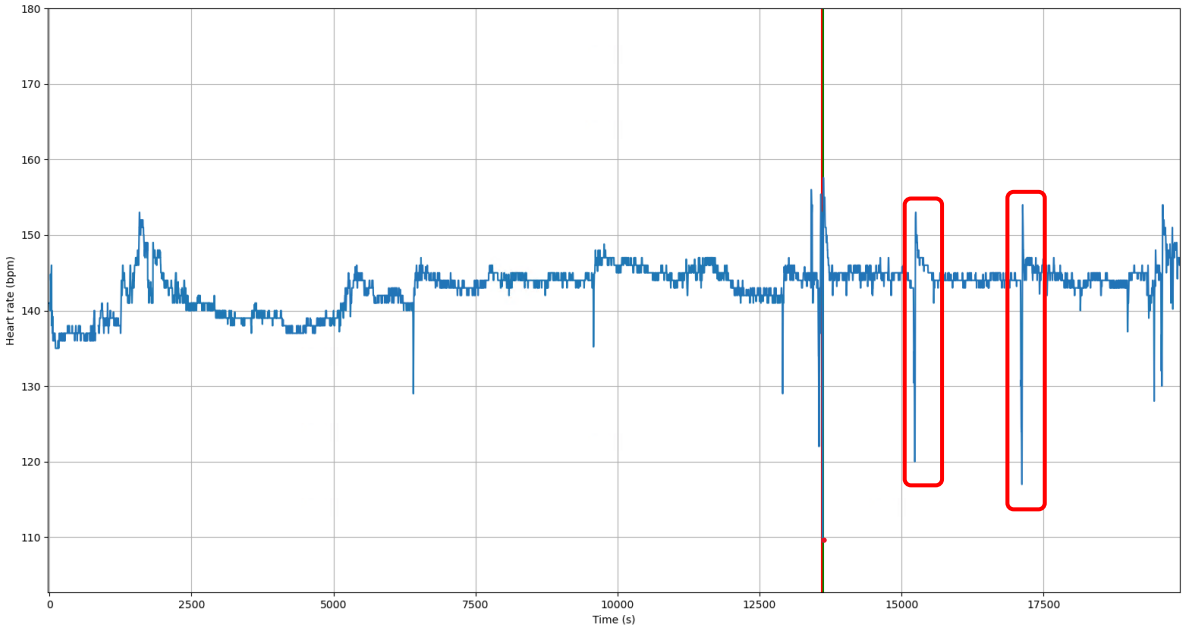
Zo, ja: ongeveer op tijdspunt: **Voorafgaand aan punt 19, op 106.000 en aan het einde van de window.**

Totale observatie HR : 5 dips gedetecteerd



Total aantal dips

Window 1, 1 dip gedetecteerd

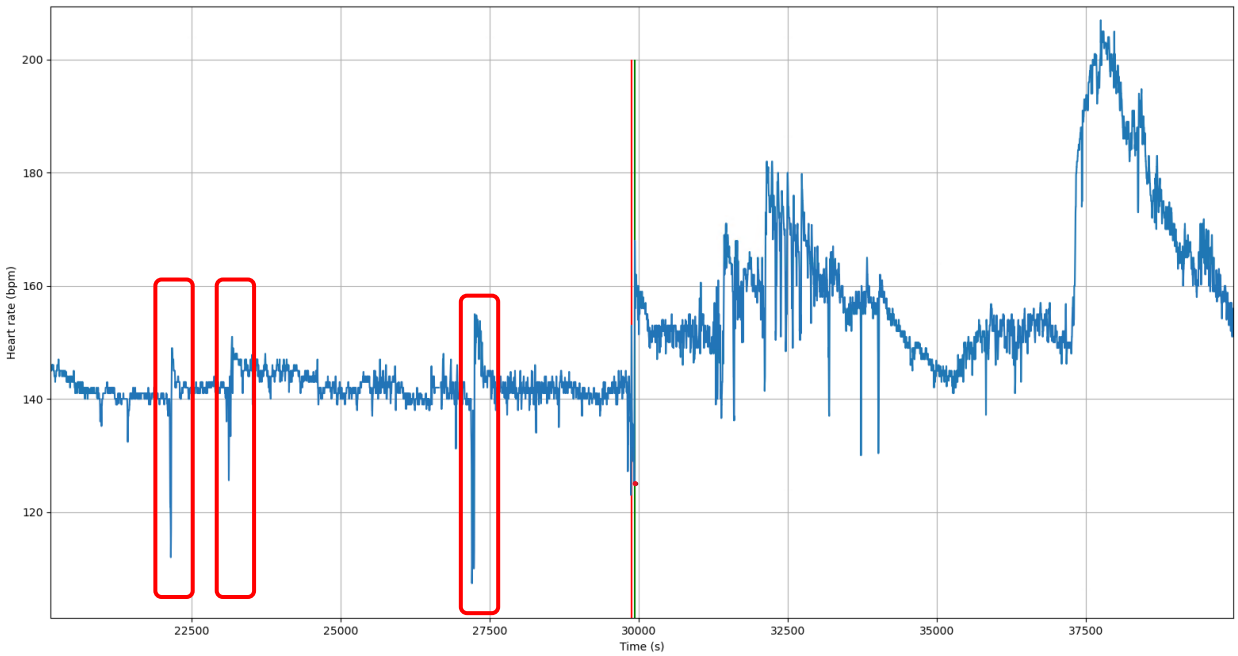


Dip 1 is juist als artifact: **ja** / nee

Er zijn artifacten gemist in deze window 1: **ja** / nee

Zo ja, welk tijdspunt ongeveer:

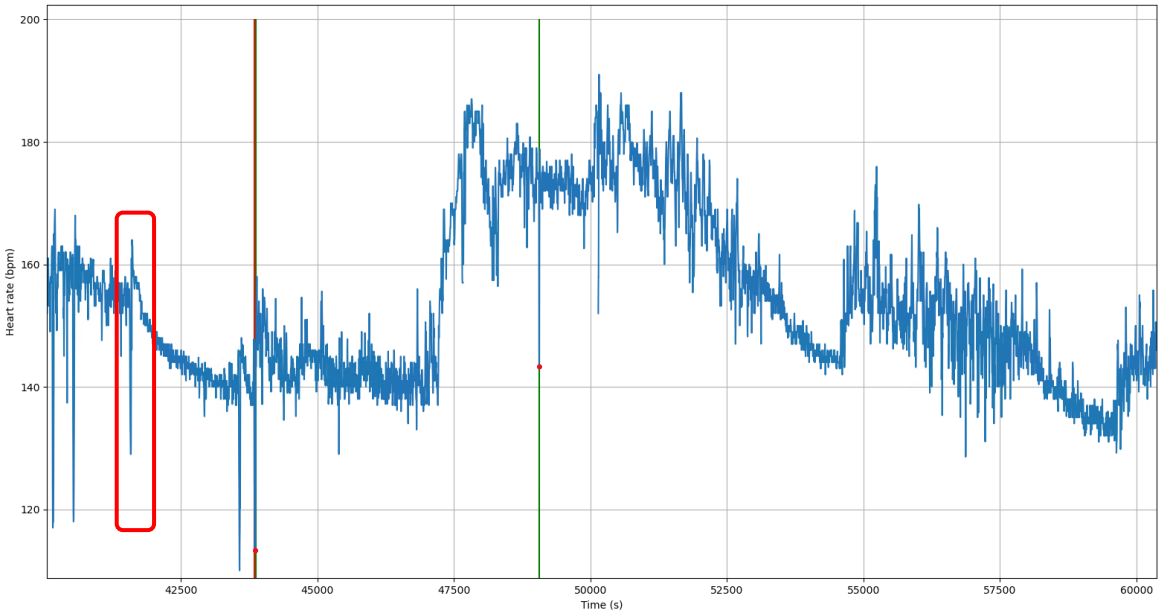
Window 2



Dip 2 is juist als artefact: **ja** / nee

Er zijn artefacten gemist in window 2: **ja**/ nee

Wndow 3

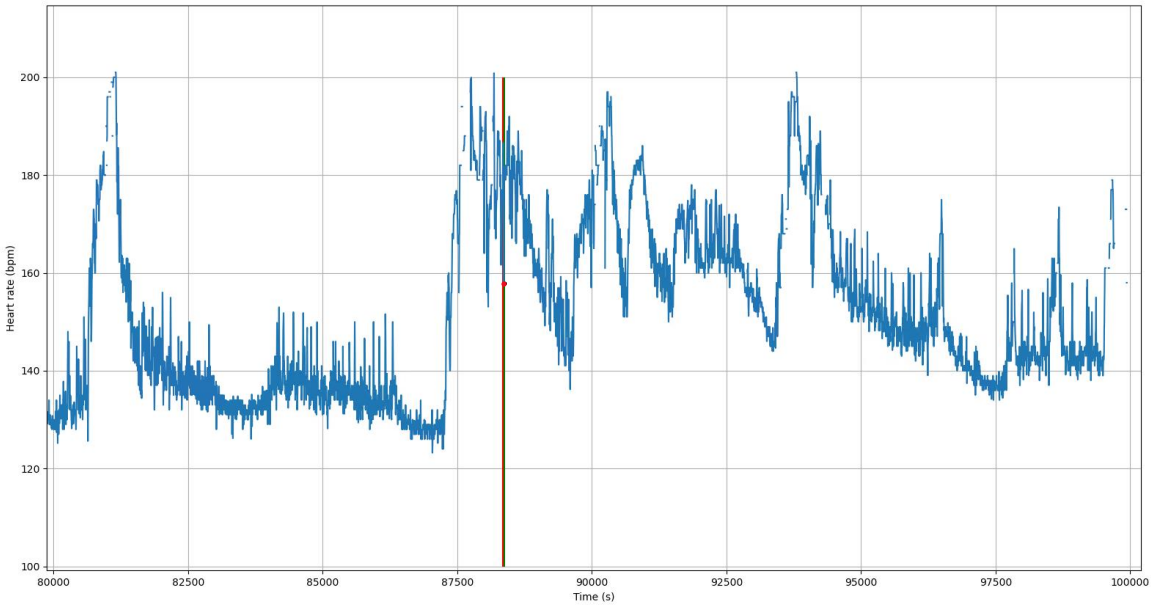


Dip 3 is juist als artefact: **ja**/ nee

Dip 4 is juist als artefact: **ja**/ nee

Er zijn artefacten gemist in window 3

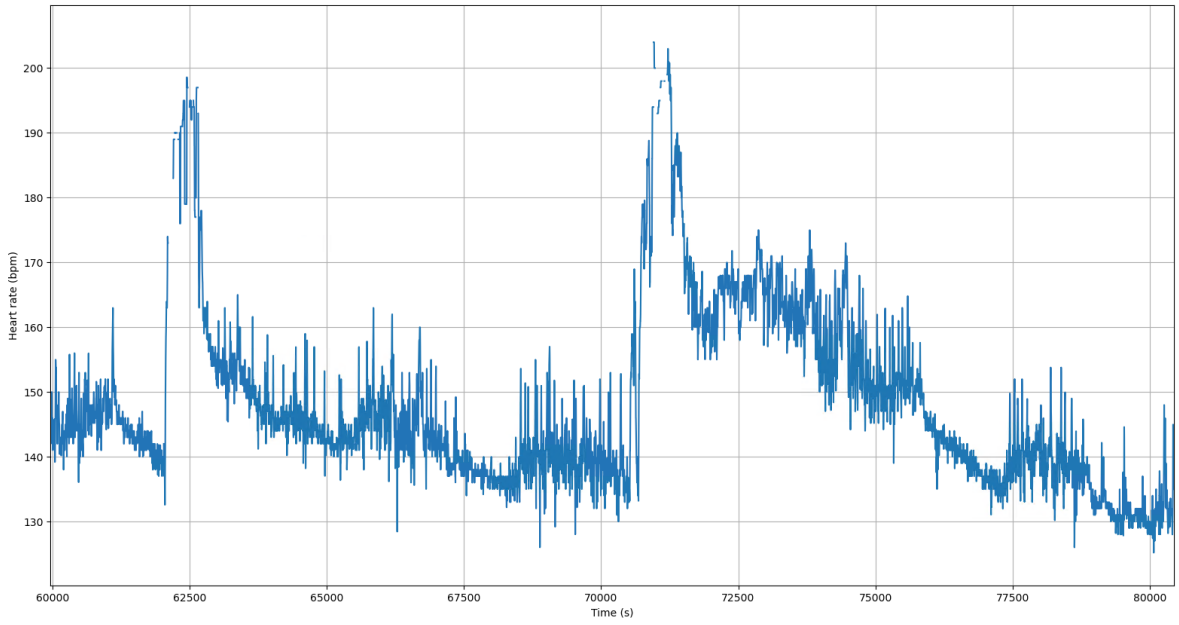
Window 4



Dip 5 is correct als artefact: **ja** / nee

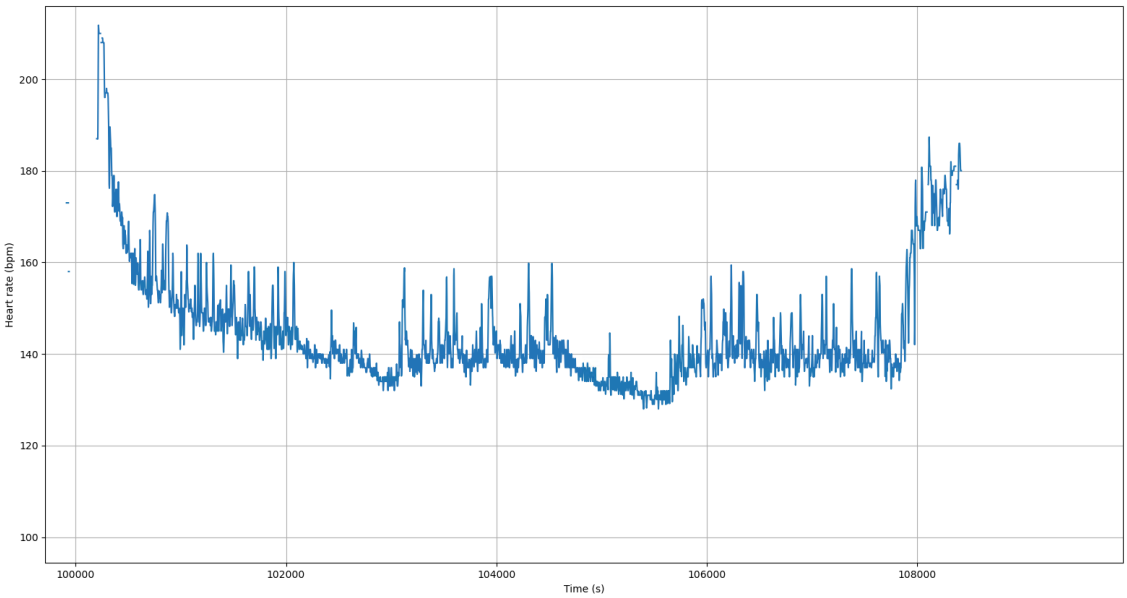
Er zijn artefacten gemist in window 4: ja / **nee**
Zo ja, waar ongeveer

Window 5



Er zijn artefacten gemist in window 5: ja / **nee**
Zo ja, waar ongeveer:

Window 6



Er zijn artefacten gemist in window 6: ja / **nee**
zo ja, waar ongeveer:

Appendix C Code

The code used for this project will be available on my GitHub. https://github.com/AnnieTheWannie/Artifact-Detection/tree/main