

Minimizing outliers on cardiovascular signals: an open-source solution

Gabriel Bazo (✉ gabrielbazo@usp.br)

Universidade de São Paulo <https://orcid.org/0000-0002-2205-9420>

Taiz Francine Brasil da Silva

Universidade de São Paulo <https://orcid.org/0000-0001-5618-9656>

Fernando Morgan de Aguiar Corrêa

Universidade de São Paulo <https://orcid.org/0000-0003-4067-9524>

Short Report

Keywords: cardiovascular signals, outliers, automated process, open-source, pharmacology

Posted Date: May 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-528311/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Context: Intravenous cardiovascular recording of conscious animals is susceptible to outlier's presence, due to freely movement and manipulations. These outliers can interfere on heart rate variability results, indicating erroneous results of sympathetic or parasympathetic modulation.

Objective: Develop an automated computational approach to minimize the presence of outliers in cardiovascular recorded signals.

Method: An application was developed according to the problem addressed based on free-use web frameworks.

Results: The use of the proposed application detected and minimized respectively 1% and 0.97% of points outliers in signals of systolic arterial pressure (SAP) and pulse interval (PI) from a representative blood pressure recording. As until then the work of minimization of outliers was carried out manually; the use of the new application considerably reduced time spent analyzing the data.

Conclusion: The proposed algorithm can detect and minimize interferent points, reducing chances of an erroneous interpretation about the cardiovascular modulation by the autonomic nervous system. The method can also significantly reduce the time of manual point-to-point screening performed by researchers.

1. Introduction

The blood pressure (BP) recording is widely used in biomedical cardiovascular research. The three most-used methods for this purpose are: tail plethysmography (non-invasive), and intra-arterial catheter or radiotelemetry (invasive). The invasive method through surgery for insertion of intra-arterial catheters is considered more accurate when compared to the non-invasive method. Radiotelemetry is used to measure the BP for longer periods. Initially, invasive methods were used mainly in anesthetized animals and later became commonly performed in conscious rats, since anesthesia interferes with normal BP values. This transition allowed the achievement of data closer to real values [1]. However, registration in non-anesthetized rodents is more susceptible to interfering fluctuations caused by freely movement and manipulations with experimental animals. The data-acquisition software currently used, offer tools for demarcation and identification of the moment that interferences that are non-related to the experimental protocol occurred in the recording, allowing the subsequent analysis and minimization of the outliers from the database. Presently, this analysis and cleaning are performed entirely by hand.

One of the techniques used to assess the modulation exerted by the autonomic nervous system (ANS) on cardiovascular parameters is the heart rate variability (HRV) [2], which requires segments of BP recording that present the greatest possible stationarity. In this case, the tools for analysis and detections offered by data-acquisition software cannot be used, since the cleaning of the outliers causes a break in the

record's timeline, a factor that is essential for opening the record in the program that performs HRV (Cardioseries v.2.7 [3]).

Today, in order to circumvent this limitation, the record can be converted to .txt format and the screening of outliers is performed manually, a fact that demands a lot of time, especially in cases of extensive records, which can generate up to 2,000 points per second when signal is acquired at 2000Hz. In this sense, automated tools are widely used and related in the literature, and offer the possibility of optimizing time-consuming activities and reduce work overload [4].

2. Problem Presentation

Given the differences in the cardiovascular modulation between anesthetized and conscious rats, it is essential to obtain these signals in conscious and freely moving animals, allowing the study of the cardiovascular response under different conditions, such as at rest in baseline condition and even in face of challenges, such as emotional stressors. However, the freely movement of the animal considerably increases the occurrence of interferences resulting from experimental manipulation and behavior responses of the animal attempting to cope with the stressful situation. Taking into account the need to study the modulation exerted by the ANS on the cardiovascular system in both conditions, the mitigation of outliers becomes extremely necessary, as its presence could be understood as a BP fluctuation caused by sympathetic or parasympathetic modulation, impairing data analysis and interpretation.

Thus, the development and use of automated routines could reduce time expend with the manual work, and generate greater homogeneity and reliability of the results.

3. A Proposed Approach To Minimize Outliers

In order to optimize and assist in the present question, a web-based application (systems designed for use through a web-browser) was developed. For that purpose, the Streamlit v0.581[5] framework was used, which is a set of classes developed based on Python v3.8.22 [6], and it was chosen due to its characteristics; this framework is aimed at the development of applications focused on analysis and exploration of data. Its use is also interesting given that the creation time of the application is short when compared to other methods, and it is customizable enough to meet the needs of the targeted application.

The application operating process was designed to be as simple and automatic as possible, just importing the data and exporting results; all intermediate steps are realized automatically. To guarantee the correct functioning of the application, the data must be contained in a xls file (Microsoft Excel Spreadsheets), without labels in the columns (only numerical data are necessary) and the information must respect the following order: time, blood pressure and pulse interval (PI) (it is possible to adjust for other types of data; this sequence applies for this specific case described here).

The application consists of a single sliding screen with all available options. First, the number of rows and columns contained in the imported file is displayed; after that, two tables display the first five lines of

data and descriptive information about each column, such as mean, standard deviation (std), minimum (min), maximum (max) and quantiles. Finally, two graphs are presented, representing the PI and the BP over time. This exploratory analysis of the data allows to verify if the imported file is the correct one and the characteristics of the data. For this step, the Pandas v.1.0.3 package [7] was used.

After these steps, the application automatically suggests cutoff points to minimize outliers according to the distribution of the variable; by default, the algorithm suggests the use of percentile 1 and 99 (z-score - 2.32 and 2.32 of a normal distribution, percentile function available in the Numpy v1.18.4 package [8]), but it is also possible to customize the points, increasing or decreasing, according to the criteria established by the user. With the defined cut-off points, data processing occurs automatically without the need of any previous command. A table is displayed with the new data characteristics, post processing (mean, std, min, max, quantiles) and two overlapping graphs, with the original and the post-processing data, allowing the visualization of the changes. If necessary, it is possible to return to the step of defining cleaning point values and changes them, and the processing will automatically be redone, generating new tables and graphs, so this process can be repeated as many times as necessary.

The operating logic of the algorithm that performs the data processing is:

1. If the value of the analyzed point is greater than the maximum value defined for processing, this value will be replaced by the mean of the two preceding points.
2. If the value of the analyzed point is less than the minimum value defined for processing, this value will be replaced by the mean of the two following points.

For this, the algorithm traverses the vector of points from a loop *i* according to the range of the data, ranging from *i* + 1 to *i* - 2, according to the logic below:

This logic applied to minimize the extreme points was the same established to perform manual corrections. The last step is the export of the processed data, which is performed by a specific function. The user only needs to click on download for the processed data to be stored on the local disk.

The application can be accessed for free at: <https://signalproc.herokuapp.com/>. Other information, such as the codes used in the development, data file for tests and more information are available on GitHub (<https://github.com/gbazo/signalproc>).

4. Results

For representative example, a segment of recording containing 6461 points (Fig. 3) of SAP, PI and time was used. Applying the proposed approach, 65 SAP points (Fig. 3A) and 63 PI points were detected (Fig. 3B). These number of points represents a 1% correction of the PAS values and 0.97% correction of the PI one.

The automation of this outlier's minimization process reduced the time of analyses and the workload, since this correction was normally performed manually, point-to-point. The use of an automatic tool also can reduce the discrepancies due to processing performed by different researchers.

5. Conclusion

The use of the automatic method can detect and minimize the artifact outliers in biological signals used in studies of cardiovascular system modulation. This approach can reduce the chance that HRV detects the interfering oscillations as being variations resulting from the modulation exercised by the ANS, thus allowing a better analysis and interpretation of data. In addition, automation of the process turns the workflow more efficient, allowing the correction of interferences to be quickly performed. It is worth mentioning that the proposed method can be applied to other types of data and/or problems.

Declarations

Conflict of Interest:

The authors declare no conflict of interest.

References

1. S. Parasuraman, R. Raveendran, Measurement of invasive blood pressure in rats., *J. Pharmacol. Pharmacother.* 3 (2012) 172–7. <https://doi.org/10.4103/0976-500X.95521>.
2. C. Cerutti, M.P. Gustin, C.Z. Paultre, M. Lo, C. Julien, M. Vincent, J. Sassard, Autonomic nervous system and cardiovascular variability in rats: a spectral analysis approach, *Am. J. Physiol. Circ. Physiol.* 261 (1991) H1292–H1299. <https://doi.org/10.1152/ajpheart.1991.261.4.H1292>.
3. D.P. Martins Dias, *Cardio Series*, (2019)..
4. R. Parasuraman, V. Riley, Humans and automation: Use, misuse, disuse, abuse, *Hum. Factors*. 39 (1997) 230–253.
5. A. Treuille, T. Teixeira, A. Kelly, Streamlit, (2018)..
6. G. Van Rossum, Python programming language., in: USENIX Annu. Tech. Conf., 2007: p. 36.
7. W. McKinney, *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*, " O'Reilly Media, Inc.", 2012.
8. T.E. Oliphant, *A guide to NumPy*, Trelgol Publishing USA, 2006.

Figures

```

for i in data.index:
    if i > 1 and i < (len(data)-2):
        if (data.var[i] >= max(range_values)):
            c = ((data.var[i-1] + data.var[i-2]) / 2)
            data.var[i] = c
            i += 1
        elif (data.var[i] <= min(range_values)):
            c = ((data.var[i+1] + data.var[i+2]) / 2)
            data.var[i] = c
            i += 1
        else:
            pass
    else:
        pass

```

Figure 1

Fragment of the proposed algorithm that minimizes extreme outliers.



Figure 2

Representation of the workflow from the animal used, through the data acquisition, application of the proposed approach and final analysis of the data.

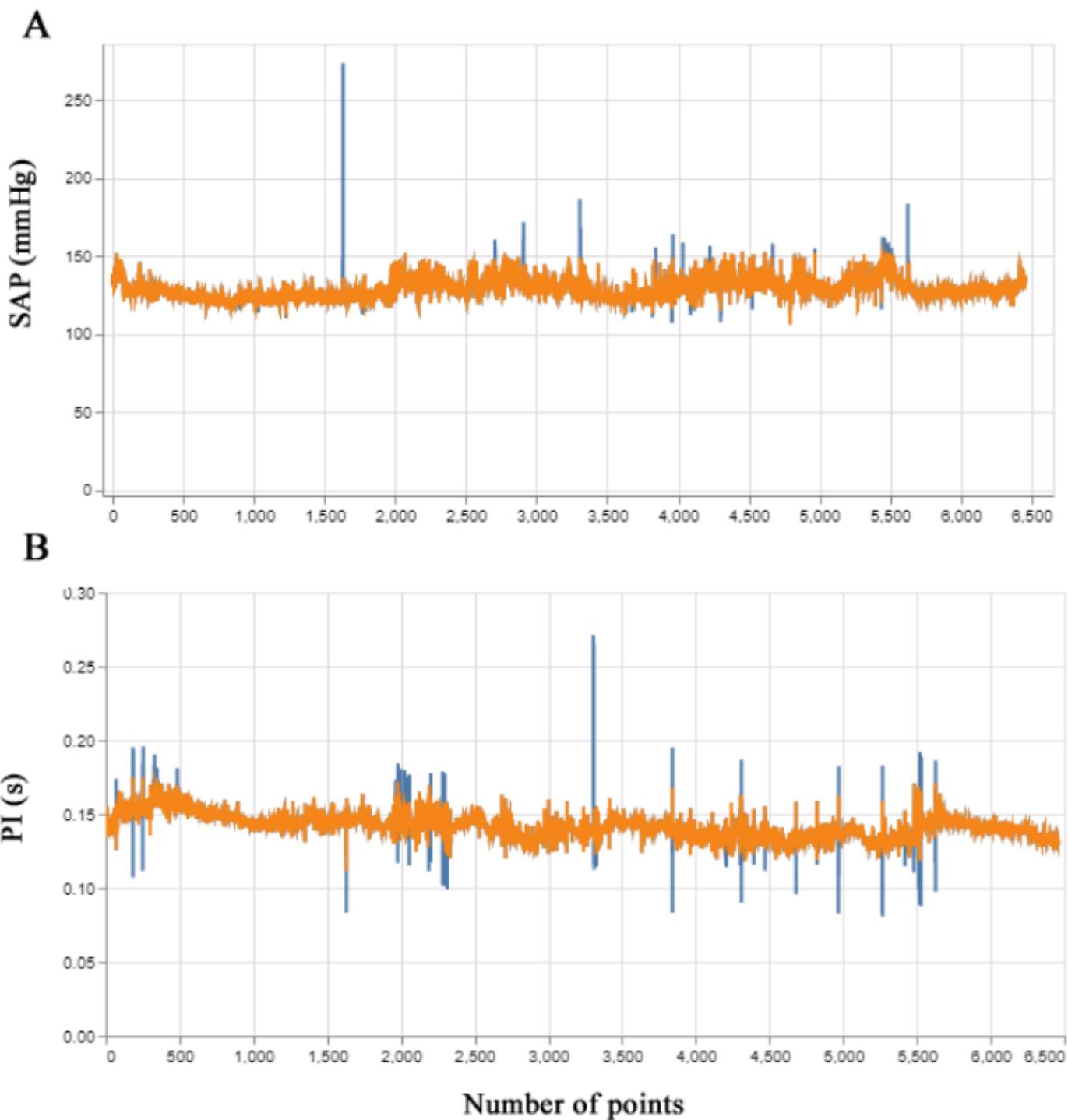


Figure 3

(A) Graphical representation of a segment of intra-arterial recording of systolic arterial pressure (SAP) and (B) pulse interval (PI) of a single rat, before and after the processing by the developed algorithm. The blue signs represent the data before and the orange the post-processing one. The data was overlapped to allow a better visualize of the changes.