# A Systematic Method for Preprocessing and Analyzing Electrodermal Activity

Sandya Subramanian, *Student Member, IEEE,* Riccardo Barbieri, *Senior Member, IEEE,* and Emery N. Brown, *Fellow, IEEE*

***Abstract*— Electrodermal activity (EDA) is a measure of sympathetic tone using sweat gland activity that has applications in research and clinical medicine. We previously identified never-before-seen statistical structure in EDA. However, there is no systematic method to preprocess and analyze EDA data to capture such statistical structure. Therefore, in this study, we analyzed the data of two healthy volunteers while awake and at rest. We used a systematic process that takes advantage of the tail behavior of various statistical distributions to ensure capturing the point process structure in EDA. We verified the presence of this temporal structure in a new dataset of subjects. Our results demonstrate for the first time that point process structure of EDA pulses can be identified across multiple datasets using a systematic method that is still rooted in the underlying physiology.**

## I. Introduction

Electrodermal activity (EDA) captures changes in skin conductance as a result of sweat gland activity. EDA is not only part of the "fight-or-flight" response, but it is also uniquely *under the sole control of the sympathetic system*. EDA can be simply measured with a few electrodes on the fingers, which explains its popularity as a "neuromarketing" tool. While EDA also has many potential applications in both research and clinical medicine, there is a lack of accurate statistical models for it based in physiology. We addressed this in our previous work by showing that point process models can be used [1]. However, there is also a lack of systematic methods for preprocessing and analyzing EDA data. In this study, we sought to address this need.

Existing approaches to EDA analysis, like our point process methods, take advantage of the pulse-like activity in the data, called galvanic skin responses or GSRs. However, they do not investigate the temporal structure of the GSRs. Simplistic metrics such as number of fluctuations in skin conductance (NFSC) consist of counting oscillations in EDA larger than a pre-set threshold and averaged across time [2-3]. While easy to calculate, this method averages sympathetic dynamics across time and involves setting a magnitude threshold across subjects. On the other hand, deconvolution-based methods assume that all GSRs have the same shape, which can be used to deconvolve the underlying nerve input to the sweat gland [4-8]. However, GSRs vary widely in shape even within a single subject's data, indicating that the fundamental physiological assumption of the model is incorrect. In addition, these algorithms are computationally expensive and the resulting estimate of neural input cannot be validated. Finally, none of these methods include a standardized preprocessing pipeline.

In our previous work, we investigated EDA data from subjects under propofol sedation and demonstrated that there was statistically significant temporal structure in the data using point process methods [9]. We also discussed how the underlying physiology of sweat glands that yields discrete GSRs lends itself to point process methodology, similar to neural spiking and heartbeat generation. Finally, we showed that such temporal structure can be used to model instantaneous dynamics [1]. This work furthers our previous work in three key ways, by 1) verifying the previous observations made about EDA temporal structure in another set of subjects under different conditions, 2) proposing a systematic method for preprocessing EDA data and extracting pulses to accurately capture the temporal structure, and 3) providing deeper insight into the previously uncovered statistical structure.

## II. Physiology & Data

The sudomotor nerve is the peripheral sympathetic nerve innervating sweat glands. When it increases its firing, sweat diffuses upwards towards the skin surface through the duct of the sweat gland. Because sweat contains many ions, it creates a low resistance path through the skin, which results in a transient increase in skin conductance, the GSR [10]. GSRs (which look like pulses) are part of the phasic component of EDA, which strongly correlates with dynamic sympathetic activity. EDA also has an underlying tonic skin conductance level that evolves on the order of minutes and is related to factors like skin tone, skin thickness, and hydration [10]. EDA, when measured, aggregates the discrete activity of several hundred sweat glands due to nerve firing. When multiple sweat glands are simultaneously active, this leads to parallel low-resistance pathways through the skin, further increasing conductance.

In this study, we analyzed data from two healthy volunteers ages 25 and 26. All data were collected under protocol approved by the Massachusetts Institute of Technology (MIT) Institutional Review Board. For both

*This work was supported by a grant from the Picower Institute for Learning and Memory, Cambridge, MA and the National Science Foundation Graduate Research Fellowship.

SS is with the Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (phone: 616-406-6087; e-mail: sandya@mit.edu).

RB is with the Department of Electronics, Informatics and Engineering, Politecnico di Milano, Milano, Italy. He is also with MGH and Massachusetts Institute of Technology, MA (barbieri@neurostat.mit.edu).

ENB is with the MGH Department of Anesthesia, Critical Care, and Pain Medicine, MIT Department of Brain and Cognitive Sciences, MIT Institute for Medical Engineering and Science, and Picower Institute for Learning and Memory, Cambridge, MA. (enb@neurostat.mit.edu).

978-1-5386-1311-5/19/$31.00 ©2019 IEEE

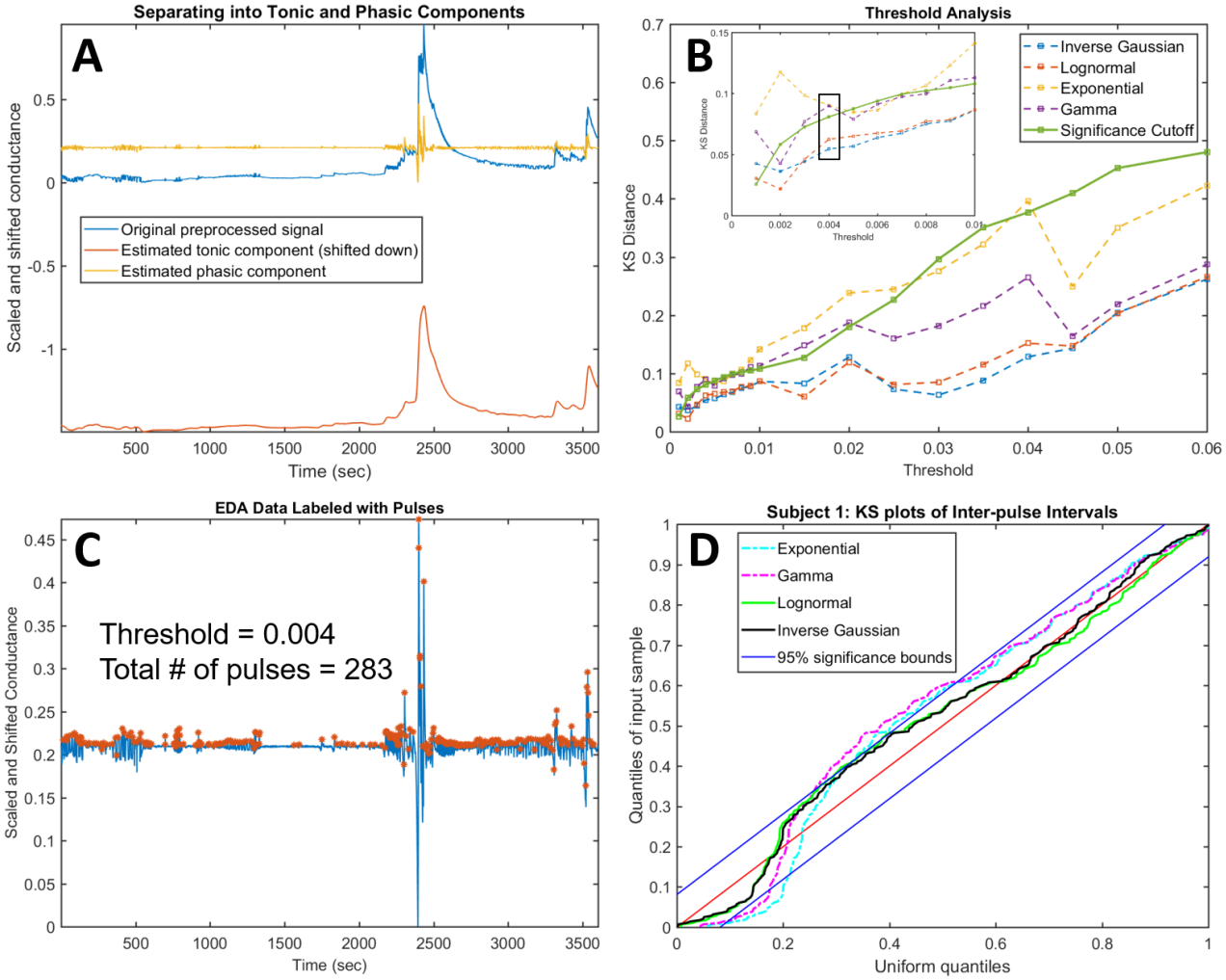


**Figure 1.** The preprocessing and analysis of EDA data from Subject 1. (A) The separation of tonic and phasic components from the artifact-removed data. (B) Threshold analysis showing the KS-distances of different distributions at various thresholds and highlighting the final threshold that was chosen. The inset plot is a zoom-in of the lower-threshold region for clarity. (C) The EDA data with identified pulses labeled. (D) The KS-plots showing the goodness-of-fit of all of the distributions at the final threshold.

subjects, approximately one hour of EDA data was collected at 256 Hz from the second and fourth digits of their non-dominant hand using dry electrodes while they were seated and at rest. The subjects were allowed to read, meditate, or watch something on a laptop or tablet, but were instructed not to sleep or write with the instrumented hand. All data were analyzed using Matlab R2017a.

## III. Methods and Results

### A. Preprocessing

We preprocessed the data as follows. First, artifacts were detected by identifying large, sharp drops or rises in the data, defined as when the absolute value of the derivative of the signal exceeded 10. Since it is physiologically impossible for sweat glands to activate that quickly and on that scale, this method can easily isolate what are likely motion-related artifacts. These artifacts were corrected by vertically shifting the data after each such artifact to match with the data beforehand and interpolating the few seconds of signal around the drop or rise itself with a cubic spline.

Then we isolated and removed the tonic component of the data to leave only the phasic component that contains sympathetic dynamics. To estimate the tonic component of the data, which is known to change on the order of several minutes, we applied an FIR low pass filter with cutoff around $1x10^{-12}$ Hz and filter length of 8192 samples. Then we subtracted the estimated tonic component from the overall signal to yield the phasic component. Figures 1A and 2A show the decomposition of EDA data for two subjects into phasic and tonic components. The tonic component has been shifted down on the y-axis for ease of visualization.

### B. Extracting Pulses

To extract pulses from the phasic component of the EDA data, we chose not to apply a universal magnitude threshold to the signal, because pulses can vary drastically not only between subjects, but even within a single subject across time. Therefore, we used the Matlab function *findpeaks*, which finds local maxima based on 'prominence', which takes into account local context around each peak [11]. For each identified peak, prominence was calculated by first

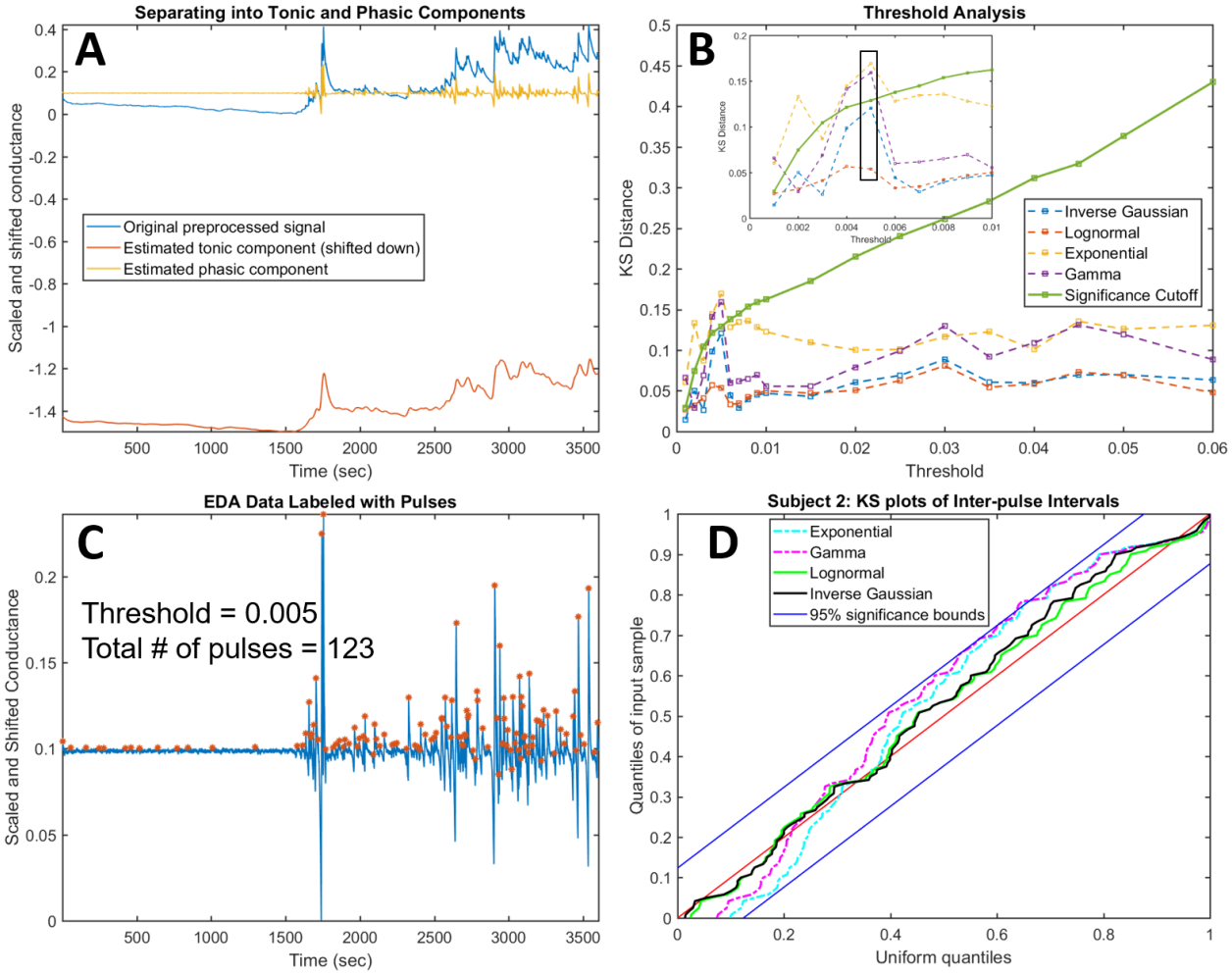


**Figure 2.** The preprocessing and analysis of EDA data from Subject 2. (A) The separation of tonic and phasic components from the artifact-removed data. (B) Threshold analysis showing the KS-distances of different distributions at various thresholds and highlighting the final threshold that was chosen. The inset plot is a zoom-in of the lower-threshold region for clarity. (C) The EDA data with identified pulses labeled. (D) The KS-plots showing the goodness-of-fit of all of the distributions at the final threshold.

drawing a horizontal line from the peak on either side and determining at what distance the line either intersected with the signal again or reached one of the endpoints of the signal. Then, for that interval on either side of the peak, the minimum value of the signal was found (a "valley"). The prominence of the peak was the height of the peak computed from the larger of the two valleys (one on either side of the peak). Once all peak prominences were calculated, an appropriate threshold for prominence was determined.

### *C. Determining the correct threshold*

To determine the appropriate prominence threshold, we employed a process in which we used the fits of several distributions (inverse Gaussian, lognormal, gamma, and exponential) across a range of thresholds as an indicator of whether we were accurately capturing the temporal structure in the EDA data. For each threshold and distribution, we computed a KS-distance after rescaling the inter-pulse intervals using the time-rescaling theorem [12]. The time-rescaling theorem states that any point process can be rescaled to a Poisson process with rate 1 using the density of the appropriate distribution. The KS-distance computes the maximum distance between the quantiles of the rescaled data and the uniform distribution, which is a simple transform of a rate 1 Poisson process done for visual simplicity. The KS-distances for each distribution at each threshold were plotted along with the 95% significance cutoff given by the number of pulses at that threshold, as shown in Figures 1B and 2B. The significance cutoff increases with increasing threshold, as the number of pulses decreases. A KS distance below the significance cutoff shows that the theoretical and empirical distributions are not statistically significantly different from each other [9].

From this plot, we determined the upper and lower bounds of the appropriate threshold. For the upper bound, we used the point at which all tested distributions fell under the significance cutoff as a marker of having too few pulses for one hour of data. For the lower bound, we identified the point at which light or medium tailed distributions such as the gamma and especially the exponential are no longer under the significance cutoff. Within the acceptable range of thresholds, we used the threshold at which the KS-distance for the gamma and exponential distributions is as far above the significance cutoff as possible. The final thresholds we

used for both subjects are marked with black boxes in Figures 1B and 2B. In addition, the identified pulses are labeled on the phasic EDA in Figures 1C and 2C.

### *D. Verifying Goodness-of-Fit*

We verified the goodness-of-fit of all distributions at the final threshold with KS-plots, shown in Figures 1D and 2D for both subjects. The KS plots show the quantiles of rescaled inter-pulse intervals against the quantiles of the uniform distribution, transformed from a Poisson process with rate 1. The 95% significance bounds are also shown.

## IV. Discussion

This work expands on our previous work by: 1) verifying the presence of similar temporal structure in a completely different dataset (awake healthy volunteers at rest) and 2) outlining a systematic process to preprocess and analyze EDA data, including verifying the presence of this physiologically justified temporal structure. The process relies heavily on the properties of several distributions, namely the inverse Gaussian, lognormal, gamma, and exponential, which differ specifically in their tail behavior. The inverse Gaussian and lognormal distributions are heavier tailed distributions with densities that "fall off" at a slower rate than the exponential and gamma. We were able to take advantage of this to ensure that we captured the temporal structure in EDA without sensor noise [9].

Our previous work indicated that we need to identify smaller "micropulses" in the data that are traditionally excluded from analysis to accurately capture the temporal structure in EDA data. We also found that this temporal structure results in right-skewed heavy tailed distributions, because there are many regions with widely spaced micropulses resulting in long inter-pulse intervals (also seen in Figures 1C and 2C) [1]. Setting too high a prominence threshold will likely miss the micropulses. On the other hand, setting too low a threshold will likely also include 'pulse-like' activity that results from sensor noise and mask the true temporal structure of the EDA. Therefore, we optimized the threshold determination by identifying where the data were fit well by the inverse Gaussian and lognormal distributions but not as well by the exponential or gamma distributions, as shown in Figures 1B and 2B.

Figures 1D and 2D show that for both subjects, the inverse Gaussian and lognormal distributions remain fully within the significance bounds, while the exponential and gamma do not. The fact that the KS plots of the exponential and gamma distributions have a steeper slope and exceed the significance bounds specifically around the upper quantiles reflects that they decrease faster in the tail compared to the inverse Gaussian and lognormal. Both the inverse Gaussian and the lognormal distributions have been used to model RR-interval and neural inter-spike intervals [12-13], and in our previous work, we highlighted how the physiology of a GSR in the sweat gland is similar in nature, lending support to the fits of these distributions to EDA data [1]. Pulse selection must be optimized using temporal structure in the data because the characteristics of pulses can change drastically even for the same subject at different times. More importantly, the statistical structure in the data is the most direct link to the underlying physiology of the sweat gland.

## V. Conclusion

There are three main conclusions from this analysis: 1) The statistical structure we previously saw in EDA data from subjects under sedation is also seen in awake subjects at rest, similarly when including the smaller pulses that are usually disregarded. 2) We can isolate this temporal structure using a systematic process that identifies pulses by taking into account local context and takes advantage of the differing tail behavior of various right-skewed distributions. 3) The success of this process verifies that the regions of EDA data with widely spaced micropulses which make up the tail of the inter-pulse interval distribution are crucial to the point process temporal structure in EDA. In our future work, we will investigate having a time-varying prominence threshold as part of our point process model to more precisely capture instantaneous dynamics in these regions.


## Acknowledgment

SS thanks the MIT Clinical Research Center staff.


## References


[1] S. Subramanian, R. Barbieri, and E. N. Brown, "A Point Process Characterization of Electrodermal Activity," *Proc. 40th IEEE International Conf on Eng in Biol and Med (EMBC)*, Jul. 2018.

[2] H. Storm, K. Myre, M. Rostrup, O. Stokland, M. D. Lien, and J. C. Raeder, "Skin conductance correlates with perioperative stress," *Acta Anaesthesiol Scand*, vol. 46, no. 7, pp. 887–895, Aug. 2002.

[3] H. Storm, M. Shafiei, K. Myre, and J. Raeder, "Palmar skin conductance compared to a developed stress score and to noxious and awakening stimuli on patients in anaesthesia," *Acta Anaesthesiol Scand*, vol. 49, no. 6, pp. 798–803, Jul. 2005.

[4] M. Benedek and C. Kaernbach, "Decomposition of skin conductance data by means of nonnegative deconvolution.," *Psychophysiology*, vol. 47, no. 4, pp. 647–58, Jul. 2010.

[5] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity.," *Journal of neuroscience methods*, vol. 190, no. 1, pp. 80–91, Jun. 2010.

[6] Greco, Alberto et al. "cvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing." *IEEE Transactions on Biomedical Engineering* 63 (2016): 797-804.

[7] Faghih, Rose et al. "Characterization of fear conditioning and fear extinction by analysis of electrodermal activity." *Proc. 40th IEEE International Conf on Eng in Biol and Med (EMBC)*, Jul. 2015.

[8] D. M. Alexander, C. Trengove, P. Johnston, T. Cooper, J. P. August, and E. Gordon, "Separating individual skin conductance responses in a short interstimulus-interval paradigm.," *Journal of neuroscience methods*, vol. 146, no. 1, pp. 116–23, Jul. 2005.

[9] D. Daley and D. Vere-Jones, An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure. Springer, 2007.

[10] Boucsein W. *Electrodermal Activity*. New York: Springer; 2012.

[11] MATLAB and Statistics Toolbox Release 2017a, The MathWorks, Inc., Natick, Massachusetts, United States.

[12] E. N. Brown, R. Barbieri, V. Ventura, R. E. Kass, and L. M. Frank, "The Time-Rescaling Theorem and its Application to Neural Spike Train Data Analysis," *Neural Computation*, vol. 14, pp. 325-346, 2001.

[13] R. Barbieri, E. C. Matten, A. A. Alabi, and E. N. Brown, "A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability," *Am. J. Physiol. Heart Circ. Physiol*, vol. 288, no. 1, pp. H424–435, Jan. 2005.