

## MIT Open Access Articles

*Point process temporal structure  
characterizes electrodermal activity*

The MIT Faculty has made this article openly available. **Please share**  
how this access benefits you. Your story matters.

**Citation:** Subramanian, Sandya et al. "Point process temporal structure characterizes electrodermal activity." *Proceedings of the National Academy of Sciences* 117, 42 (October 2020): 26422-26428. © 2020 National Academy of Sciences

**As Published:** <http://dx.doi.org/10.1073/pnas.2004403117>

**Publisher:** National Academy of Sciences

**Persistent URL:** <https://hdl.handle.net/1721.1/130247>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of Use:** Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



# Point process temporal structure characterizes electrodermal activity

Sandya Subramanian<sup>a,b,c</sup> , Riccardo Barbieri<sup>b,d,e</sup>, and Emery N. Brown<sup>a,b,c,d,f,1</sup>

<sup>a</sup>Harvard–Massachusetts Institute of Technology Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139;

<sup>b</sup>Department of Anesthesia, Critical Care, and Pain Medicine, Massachusetts General Hospital, Boston, MA 02114; <sup>c</sup>Institute of Medical Engineering and Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; <sup>d</sup>Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; <sup>e</sup>Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milan, Italy 20133; and <sup>f</sup>Picower Institute of Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Peter L. Strick, University of Pittsburgh, Pittsburgh, PA, and approved August 14, 2020 (received for review March 12, 2020)

**Electrodermal activity (EDA) is a direct readout of the body's sympathetic nervous system measured as sweat-induced changes in the skin's electrical conductance. There is growing interest in using EDA to track physiological conditions such as stress levels, sleep quality, and emotional states. Standardized EDA data analysis methods are readily available. However, none considers an established physiological feature of EDA. The sympathetically mediated pulsatile changes in skin sweat measured as EDA resemble an integrate-and-fire process. An integrate-and-fire process modeled as a Gaussian random walk with drift diffusion yields an inverse Gaussian model as the interpulse interval distribution. Therefore, we chose an inverse Gaussian model as our principal probability model to characterize EDA interpulse interval distributions. To analyze deviations from the inverse Gaussian model, we considered a broader model set: the generalized inverse Gaussian distribution, which includes the inverse Gaussian and other diffusion and nondiffusion models; the lognormal distribution which has heavier tails (lower settling rates) than the inverse Gaussian; and the gamma and exponential probability distributions which have lighter tails (higher settling rates) than the inverse Gaussian. To assess the validity of these probability models we recorded and analyzed EDA measurements in 11 healthy volunteers during 1 h of quiet wakefulness. Each of the 11 time series was accurately described by an inverse Gaussian model measured by Kolmogorov–Smirnov measures. Our broader model set offered a useful framework to enhance further statistical descriptions of EDA. Our findings establish that a physiologically based inverse Gaussian probability model provides a parsimonious and accurate description of EDA.**

electrodermal activity | point processes | statistics | autonomic nervous system | signal processing

**E**lectrodermal activity (EDA), which measures the electrical properties of the skin as changes in conductance, is mediated almost exclusively by the sympathetic branch of the autonomic nervous system (1). The skin continuously receives sympathetic innervation. Consequently, EDA is continuously present due to sweat glands filling and releasing sweat onto the skin. Changes in the level of filling occur in response to internal stimuli (physiological and psychological) and external stimuli such as threats or dramatic changes in ambient temperature. EDA is a component of the primal flight-or-fight response that is routinely used as a measure of the sympathetic nervous system activity in psychological studies, polygraph tests and studies of stress (1). Measures of EDA are now being developed as a neuromarketing tool to evaluate consumer responses to different products or promotions (2). For this reason, there is growing interest in the development of analysis methods to characterize EDA accurately.

EDA has a characteristic pattern consisting of two distinct components. There is a baseline or tonic component that drifts gradually with time. On top of the tonic component is a phasic component composed of pulse events that vary in amplitude, shape, and spacing. Sweat release is a pulsatile process because a

sufficient volume of sweat has to accumulate, fill the glands, and be released onto the skin to observe an EDA change. This integrate-and-fire nature of the phasic component is believed to represent fast changes in sympathetic nervous system activity. EDA activity can show large variation in both baseline activity and pulse activity within and between individuals.

Most current EDA analysis methods focus mainly on the phasic component to characterize sympathetic activity. These methods fall in two categories: rate-based methods and deconvolution methods. The rate-based methods specify a time window and estimate a moving average of the number of pulse events per time window (3–5). For the deconvolution methods, a single pulse shape is assumed, and the EDA signal is represented as the convolution of this pulse shape with neural inputs. For this technique, the neural input is deconvolved from the EDA while simultaneously fitting pulse shape parameters (6–10). Hence, deconvolution methods report the occurrence times and amplitudes of the pulse events. Although widely used, neither the rate-based nor the deconvolution methods are based on EDA's well-established physiology. Likewise formal statistical modeling is not used to characterize the interpulse interval dynamics (pulse rate and pulse times) or the pulse amplitudes.

The important advance that we report here is the application of elementary point process models based on physiology to characterize the dynamics of EDA interpulse intervals. These

## Significance

**Electrodermal activity (EDA) is a readout of the body's sympathetic nervous system measured as sweat-induced changes in the electrical conductance properties of the skin. Interest is growing in using EDA to track physiological conditions such as stress levels, sleep quality, and emotional states. The integrate-and-fire physiology underlying EDA production suggests that its interpulse intervals should obey an inverse Gaussian probability model. In an analysis of EDA data recorded in 11 healthy volunteers during quiet wakefulness, we established that the inverse Gaussian model accurately characterized the interpulse intervals. Our findings show a physiologically based statistical model provides a parsimonious and accurate description of EDA.**

Author contributions: S.S., R.B., and E.N.B. designed research; S.S. performed research; S.S., R.B., and E.N.B. contributed new reagents/analytic tools; S.S. analyzed data; and S.S., R.B., and E.N.B. wrote the paper.

Competing interest statement: A patent application was filed with S.S., R.B., and E.N.B. on July 15, 2020 (PCT/US2020/042031).

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>To whom correspondence may be addressed. Email: enb@neurostat.mit.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2004403117/-DCSupplemental>.

First published October 2, 2020.

parsimonious probability models provide a physiologically based framework for statistical analysis of EDA. The balance of this paper is organized as follows. In *Physiology and Theory*, we formulate generalized inverse Gaussian integrate-and-fire probability models derived from EDA's integrate-and-fire properties and, as alternatives, lognormal, exponential, gamma, and generalized inverse Gaussian non-integrate-and-fire probability models. In *Application and Results*, we use these models in the analysis of EDA recorded from 11 healthy subjects during quiet wakefulness. The *Discussion* describes the implications of our findings for future basic science and translational studies.

## Physiology and Theory

**The Anatomy and Physiology of Electrodermal Activity.** To develop our statistical models of EDA activity we review the anatomy and physiology of sweat production in the skin (1). Each sweat gland consists of three parts: the dermal gland, the duct that connects the gland to the skin surface, and the pore where the duct opens to the skin (Fig. 1A; 1). The dermal portion of the gland is innervated by sudomotor nerves, a part of the peripheral nervous system that is predominantly under sympathetic control (1). Sweat is produced in the gland by sympathetically induced abrupt increases in spiking activity (action potentials) in the sudomotor nerves. These electrical events are called sudomotor bursts. Sweat produced in response to these bursts accumulates in the duct. Once the duct is full, it pushes open the pore and spills onto the skin. The sweat either evaporates or is reabsorbed through the walls of the duct. With the duct now empty, the accumulation process begins again (1, 11, 12).

Sweat on the skin's surface increases its electrical conductance (inverse of resistance) because its salt content facilitates the propagation of electrical currents. Electrical conductance across the skin can be measured in a standard fashion by placing two electrodes on either the palm or fingers, applying a constant voltage, and measuring the current (1). The pulsatile effects of

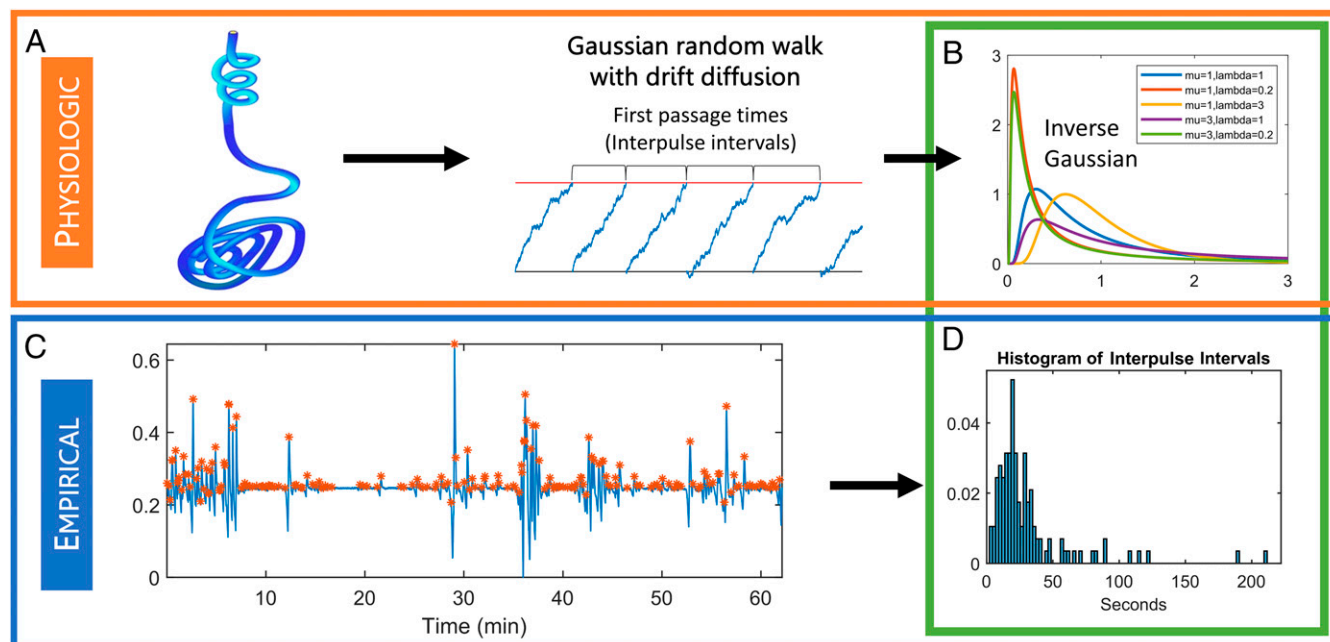
the sudomotor bursts measured at the skin are termed galvanic skin responses (GSRs). The second-to-second changes in skin sweating measured as second-to-second changes in skin conductance are termed EDA (1).

EDA measurements have two distinct components: tonic activity and phasic activity. Tonic activity represents generally ongoing EDA that reflects the background state of the EDA. The phasic activity reflects primarily the GSR or pulsatile events.

**Statistical Models of EDA Pulses.** We investigate four statistical models to characterize the times between GSR events recorded during EDA measurements: generalized inverse Gaussian, lognormal, gamma, and exponential. The choice of the generalized inverse Gaussian model follows directly from the physiology described previously. The phasic component of the EDA measurements is comprised of pulsatile events. The times between these pulse events are governed by sympathetic stimulation of the glands, the sweat production and its accumulation in the glands, sweat release on the skin, and sweat reabsorption and evaporation. We represent this four-step sequence of sweat accumulation in the gland and its release onto the skin as an integrate-and-fire process defined by a Gaussian random walk with drift diffusion (Fig. 1B). It is well known that the times between firing events for this elementary diffusion process obey an inverse Gaussian probability model (13–15). Therefore, we chose as the first model the generalized form of the inverse Gaussian probability density defined for an interpulse interval time  $x > 0$  as

$$f(x|\psi, \chi, \lambda) = \frac{(\psi/\chi)^{\lambda/2}}{2K_{\lambda}((\psi\chi)^{1/2})} x^{\lambda-1} \exp\left[-\frac{1}{2}\left(\psi x + \frac{\chi}{x}\right)\right], \quad [1]$$

where  $-\infty < \lambda < \infty$ ,  $\psi \geq 0$ ,  $\chi \geq 0$ , and  $K_{\lambda}$  is the modified Bessel function of the third kind with index  $\lambda$  (16). The generalized



**Fig. 1.** A summary of our model, including both physiologic and empirical components and how they align. (A) An illustration of how sweat gland physiology can be modeled as a Gaussian random walk with drift diffusion, which suggests that the times between first passage events (interpulse intervals) should follow (B) an inverse Gaussian distribution (examples shown; for density, see *SI Appendix*, Eq. S1). Sweat gland image credit: mikrostoker © 123RF.com. (C) Extracted GSR events (marked by red asterisks) from EDA data, and (D) a histogram of interpulse interval data, to which we fit generalized inverse Gaussian (diffusion and nondiffusion), lognormal, gamma, and exponential probability models.

**Table 1. Results: Fitted parameters for generalized inverse Gaussian diffusion models**

| Subject | Number of pulses extracted | GIG diffusion models |         |        | Identifiable distribution? |
|---------|----------------------------|----------------------|---------|--------|----------------------------|
|         |                            | Lambda               | Chi     | Psi    |                            |
| S01     | 242                        | −0.5                 | 9.2706  | 0.0427 | Inverse Gaussian           |
| S02     | 122                        | −0.5                 | 18.3391 | 0.0238 | Inverse Gaussian           |
| S03     | 112                        | −0.5                 | 18.6703 | 0.0178 | Inverse Gaussian           |
| S04     | 223                        | −0.5                 | 16.8932 | 0.0655 | Inverse Gaussian           |
| S05     | 348                        | −0.5                 | 11.0195 | 0.0973 | Inverse Gaussian           |
| S06     | 97                         | −0.5                 | 11.7719 | 0.0075 | Inverse Gaussian           |
| S07     | 182                        | −0.5                 | 27.1432 | 0.0645 | Inverse Gaussian           |
| S08     | 344                        | −0.5                 | 15.8727 | 0.1450 | Inverse Gaussian           |
| S09     | 299                        | −0.5                 | 10.2546 | 0.0701 | Inverse Gaussian           |
| S10     | 111                        | −0.5                 | 42.7494 | 0.0369 | Inverse Gaussian           |
| S11     | 125                        | −0.5                 | 21.4675 | 0.0240 | Inverse Gaussian           |

For each subject, the parameters of the best fitted generalized inverse Gaussian diffusion model are shown. If the model is identifiable as a known distribution based on the parameter values, that is indicated in the rightmost column. GIG, generalized inverse Gaussian.

inverse Gaussian is a flexible family of probability models that includes the inverse Gaussian model as a special case  $\lambda = -\frac{1}{2}$ . For  $\lambda \leq 0$ , the generalized inverse Gaussian probability density gives diffusion (integrate-and-fire) models other than the inverse Gaussian. For  $\lambda > 0$ , the generalized inverse Gaussian density gives nondiffusion (non-integrate-and-fire) models. As  $\chi \rightarrow 0$ , the generalized inverse Gaussian yields the special case of a gamma probability model with parameters  $\lambda > 0$  and  $2^{-1}\psi \geq 0$ . When  $\lambda = 1$  as  $\chi \rightarrow 0$ , the generalized inverse Gaussian becomes the exponential probability density (16). Hence, by fitting the generalized inverse Gaussian model we can study systematically how well EDA inter-pulse interval time series can be characterized by a broad class of integrate-and-fire and non-integrate-and-fire models.

To broaden the class of alternatives to integrate-and-fire models, we also consider the lognormal distribution. Its probability density function is defined as

$$f(x|\mu, \sigma) = \frac{1}{(2\pi x^2 \sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(\log x - \mu)^2\right\}, \quad [2]$$

where  $-\infty < \mu < \infty$  and  $\sigma > 0$ . The lognormal has been widely used as an empirical model to characterize interevent distributions in point process time series (16).

We postulate that under stable, or approximately stable, background conditions the inverse Gaussian model will provide an accurate description of the inter-GSR (interpulse) event times as the integrate-and-fire model is a plausible description of skin sweat production. This assumes that the group activity of the sweat glands shows a coherent behavior. If there is variation in the behavior of the individual glands, even when the background state is stable, then it is possible that the elementary inverse Gaussian model may not apply.

There may be a mixture of inverse Gaussian models which could produce either longer or shorter interevent intervals relative to a single inverse Gaussian model. These differences could be manifested by either heavier or lighter tails relative to an inverse Gaussian model. To model heavy tail distributions, we consider the lognormal and the gamma distributions, whereas to model the light tail distributions, we consider the gamma and the exponential distributions. The gamma is flexible in that it can have either a heavier or a lighter tail than the inverse Gaussian depending on the values of the parameters (17). Although the exponential distribution is a special case of the gamma, it represents the null model of an underlying Poisson process governing GSR-event production.

We quantify the tail behavior of a probability density  $f(x)$  by evaluating the settling rate, which is defined as

$$SR_f = \lim_{x \rightarrow \infty} \frac{f(x)}{1 - F(x)}, \quad [3]$$

where  $F(x)$  is the corresponding cumulative distribution function. The settling rate is the limit of the hazard function as  $x$  tends to infinity (17). By this definition, a distribution is commonly classified as light, medium, or heavy tailed based on whether the settling rate is infinite, positive but finite, or zero, respectively. The more slowly (rapidly) the tail of a distribution settles, the heavier (lighter) the tail. We can divide our set of probability models into four medium-tailed (exponential, gamma, inverse Gaussian, and generalized inverse Gaussian) and one heavy-tailed distribution (lognormal). The four medium-tailed distributions are distinguished relative to each other based on their parameter values, which determine the respective settling rates (17, 18).

## Application

**Experimental Data.** Our experimental protocol was approved by the Massachusetts Institute of Technology Institutional Review Board, and all subjects provided written informed consent. We collected EDA data from 12 healthy volunteers (six men) between the ages of 22 and 34 while awake and at rest (19). Electrodes were connected to the second most distal phalange of the second and fourth digits of each subject's nondominant hand. Approximately 1 h of EDA data were collected at 256 Hz. Subjects were seated upright and instructed to remain awake. They were allowed to read, meditate, or watch something on a laptop or tablet but not to use the instrumented hand. We assumed skin and ambient temperature were constant for the duration of the experiment. One subject's data were not included in the analysis because we learned, after the data collection, that this individual occasionally experienced a Raynaud's type phenomenon. This would affect the quality of the EDA data. Data from the remaining 11 subjects were analyzed using MATLAB 2017a.

**Data Preprocessing and EDA Pulse Selection.** Preprocessing consisted of two steps: 1) detecting and removing artifacts and 2) isolating the phasic component. Artifact detection was done based on the derivative of the time series since large rapid changes are physiologically impossible for skin conductance. Artifact removal was done in two parts, first correcting for artifact-related large magnitude changes in the remainder of the signal and then interpolating the few seconds around the artifact itself. Then a low-pass

FIR filter was used to estimate and remove the slow-moving tonic component of the signals, thereby, isolating the phasic component. The data preprocessing is described in further detail in Subramanian et al. (20).

Since there are underlying tonic fluctuations, absolute pulse amplitude alone is insufficient to extract pulses reliably. Therefore, we computed locally adjusted amplitudes for all detected peaks using the MATLAB function *findpeaks*. The *findpeaks* algorithm computes a prominence or relative amplitude for each peak, by adjusting the amplitude of each peak as the height above the highest of neighboring valleys on either side. The valleys are chosen based on the lowest point in the signal between the peak and the next intersection with the signal of equal height on either side. With this method, a peak with small absolute amplitude can be rewarded in its prominence value if it is in a region of data with low activity. We then used a threshold on this prominence value to account for the varying baseline, instead of on the absolute amplitude.

We used a prominence threshold of 0.005 to extract peaks across all subjects, unless this resulted in too few or too many pulses for each hour-long recording. This was primarily verified by visual inspection of extracted pulses, as well as rough estimates of 60 and 360 pulses as generous bounds for what should be expected of 1 h of data at rest with little external stimulation. This corresponds to one pulse every 10 to 60 s on average (4, 21–26). In the case that too few pulses were extracted, we gradually reduced the prominence threshold by 0.001 until the number of pulses exceeded 100 and no obvious pulses were missed by visual inspection (verified by three different viewers independently). In the case too many pulses were extracted, we gradually increased the prominence threshold by 0.001 until the number of pulses was fewer than 360 and the extracted pulses were distinguishable from sensor noise by visual inspection (verified by three different viewers independently). If changing the threshold in increments of 0.001 resulted in drastic changes in the number of pulses each time, we reduced the increment to 0.0005. Although not fully automated at this stage, the design of this method to extract pulses attempted to take into account the wide variation in baseline levels of EDA activity seen across subjects. The extracted peaks included smaller peaks that other methods would generally ignore as noise. However, we chose to include them in the analysis.

**Statistical Model Fitting and Comparison.** First, we restricted the parameter space to  $\lambda \leq 0$  for the generalized inverse Gaussian model to find the best integrate-and-fire model for each data series. To allow for statistical improvements by capturing deviations from the integrate-and-fire models, we fit four other probabilities models to each data series: generalized inverse

Gaussian non-integrate-and-fire ( $\lambda > 0$ ), lognormal, gamma, and exponential models. We fit all models by maximum likelihood (27, 28). We assessed goodness-of-fit by using KS plots and by computing Akaike's information criterion (AIC), defined as

$$\text{AIC} = -2\log f(\hat{\theta}_{ML}) + 2p, \quad [4]$$

where  $f(\hat{\theta}_{ML})$  is the likelihood evaluated at the maximum likelihood parameter estimates and  $p$  is the number of parameters. A lower AIC indicates a better fit.

A KS plot compares the rescaled quantiles from the fit of the estimated probability model with the quantiles of an exponential distribution with rate 1 using the time-rescaling theorem (29). This theorem states that any point process can be rescaled to an exponential distribution with rate 1 using its conditional intensity (hazard) function. The KS distance computes the maximum distance between the quantiles of the rescaled data and the uniform distribution, which is a simple transform of an exponential distribution with rate 1. A smaller KS distance indicates that the model is more similar to the structure observed in the data. We computed 95% confidence intervals (5% significance cutoffs) for the KS plot and compared the KS distances across models (30). A KS distance that is within (outside) the 95% confidence intervals suggests that the model offers (fails to offer) a reasonably accurate description of the data.

We also compared the models using a tail behavior analysis, in which the settling rates of the models were compared to determine the heaviness of the tails of the distributions. We hypothesized that each EDA interpulse interval distribution results from the activity of multiple sweat glands. This could lead to deviations from the inverse Gaussian with slightly heavier or lighter tails. These deviations can be captured statistically by right-skewed models such as the lognormal, generalized inverse Gaussian and gamma that together allow for more flexibility in tail behavior.

## Results

**Extraction of EDA Pulses.** Fig. 1C shows an example of an excerpt of extracted pulses for Subject S07. This includes pulses large enough to be used in most analyses as well as those that are much smaller and usually either smoothed out or ignored as noise. We included both types of pulses for all subjects and did not distinguish between them. The majority of subjects showed appreciable fluctuations in the tonic component across time (*SI Appendix, Fig. S1*). Across the 11 subjects, the total number of pulses in the 1-h time window ranged between 97 and 348, including the distantly spaced smaller pulses (*SI Appendix, Fig. S2*).

**Table 2. Results: AIC for all models**

| Subject | Prominence threshold | Number of pulses extracted | GIG diffusion AIC | GIG nondiffusion AIC | Lognormal AIC | Gamma AIC    | Exponential AIC |
|---------|----------------------|----------------------------|-------------------|----------------------|---------------|--------------|-----------------|
| S01     | 0.005                | 242                        | 1,700             | 1,736                | <b>1,698</b>  | 1,783        | 1,781           |
| S02     | 0.005                | 122                        | 989               | 1,021                | <b>981</b>    | 1,050        | 1,049           |
| S03     | 0.005                | 112                        | 916               | 963                  | <b>896</b>    | 995          | 996             |
| S04     | 0.005                | 223                        | 1,670             | 1,647                | 1,645         | <b>1,634</b> | 1,679           |
| S05     | 0.023                | 348                        | 2,333             | 2,290                | 2,286         | <b>2,262</b> | 2,337           |
| S06     | 0.004                | 97                         | <b>832.6</b>      | 870                  | 832.8         | 889          | 900             |
| S07     | 0.005                | 182                        | 1,430             | 1,417                | <b>1,406</b>  | 1,407        | 1,458           |
| S08     | 0.005                | 344                        | <b>2,179</b>      | 2,184                | 2,183         | 2,214        | 2,299           |
| S09     | 0.01                 | 299                        | 2,102             | 2,060                | <b>2,034</b>  | 2,041        | 2,084           |
| S10     | 0.0025               | 111                        | 976               | <b>973</b>           | 974           | 977          | 998             |
| S11     | 0.005                | 125                        | 1,082             | 1,080                | <b>1,070</b>  | 1,089        | 1,093           |

The best performing model per subject is in bold. The final prominence threshold used and number of pulses extracted is also indicated for each subject. GIG, generalized inverse Gaussian.



Table 3. Results: KS distance for all models

| Subject | Significance cutoff | GIG diffusion KS distance | GIG nondiffusion KS distance | Lognormal KS distance | Gamma KS distance | Exponential KS distance |
|---------|---------------------|---------------------------|------------------------------|-----------------------|-------------------|-------------------------|
| S01     | 0.087               | <b>0.057*</b>             | 0.0790*                      | 0.065*                | 0.0791*           | 0.085*                  |
| S02     | 0.123               | 0.073*                    | 0.111*                       | <b>0.040*</b>         | 0.119*            | 0.120*                  |
| S03     | 0.128               | 0.121*                    | 0.156                        | <b>0.054*</b>         | 0.159             | 0.169                   |
| S04     | 0.091               | 0.077*                    | 0.050*                       | 0.043*                | <b>0.023*</b>     | 0.102                   |
| S05     | 0.073               | 0.067*                    | 0.053*                       | 0.031*                | <b>0.030*</b>     | 0.122                   |
| S06     | 0.138               | 0.071*                    | 0.136*                       | <b>0.050*</b>         | 0.134*            | 0.177                   |
| S07     | 0.101               | 0.088*                    | 0.074*                       | 0.057*                | <b>0.048*</b>     | 0.146                   |
| S08     | 0.073               | <b>0.025*</b>             | 0.0295*                      | 0.0301*               | 0.047*            | 0.120                   |
| S09     | 0.079               | 0.078*                    | 0.0512*                      | <b>0.028*</b>         | 0.0511*           | 0.100                   |
| S10     | 0.129               | 0.050*                    | <b>0.032*</b>                | 0.042*                | 0.055*            | 0.109*                  |
| S11     | 0.122               | 0.052*                    | 0.032*                       | <b>0.028*</b>         | 0.056*            | 0.077*                  |

The best performing model per subject is in bold, and all models under significance cutoff are marked with an asterisk. The significance cutoff was computed based on the number of pulses extracted. GIG, generalized inverse Gaussian.

The final prominence thresholds used ranged between 0.0025 and 0.023.

**Findings from Statistical Model Comparison.** For all 11 subjects, the optimal generalized inverse Gaussian diffusion model was an inverse Gaussian, indicated by  $\lambda=-0.5$  in Table 1. This inverse Gaussian was always within the significance cutoff according to KS distance, indicating that it provides an accurate description of the data and supporting our hypothesis that the pulsatile sweat release events of EDA can be modeled as an integrate-and-fire process. When allowing for deviations from the inverse Gaussian, for 9 of the 11 subjects, one or more of the lognormal, gamma, or generalized inverse Gaussian nondiffusion models was able to improve statistically on the fit of the inverse Gaussian (Tables 2 and 3; Fig. 2; and *SI Appendix, Figs. S3–S7*). The exponential was never a better fit than the inverse Gaussian, and was only within the significance cutoff for 4 of the 11 subjects. This suggests that for the majority of subjects, the exponential model did not offer an accurate description of the data.

Both AIC and KS distance were in agreement for the best fit models for the data of 8 of the 11 subjects. This corresponded to 1 generalized inverse Gaussian diffusion model (S08), 1 generalized inverse Gaussian nondiffusion model (S10), 4 lognormal (S02, S03, S09, and S11), and 2 gamma (S04 and S05) models. For S01 and S06, the AICs and KS distances suggested different best fit distributions between inverse Gaussian and lognormal. For S07, they identified different best fit distributions between lognormal and gamma. It is reasonable to expect that the results from AIC and from KS distance will not match exactly since

different metrics are intended to capture different aspects of model fits. However, the fact that they agree across the majority of subjects reinforces that there is specific statistical structure in the data that can be captured with a parsimonious model.

The second phase of comparing the models was analyzing the tail behavior (Table 4) using the settling rates estimated for each of the five models for all subjects. Across all 11 subjects, the lognormal always had the heaviest tail since it is commonly classified as a heavy-tailed distribution. The generalized inverse Gaussian integrate-and-fire model had the next heaviest tail, indicated by the next smallest settling rate. The lognormal was the only other class of models besides the generalized inverse Gaussian diffusion models that was within the significance cutoff for all subjects. This suggests that deviations from the inverse Gaussian tended toward heavier tails that were best captured by the lognormal.

Among the remaining models, the generalized inverse Gaussian non-integrate-and-fire models always had lighter tails than the integrate-and-fire models. The tails of the gamma were even lighter. Omitting the exponential, since it fit poorly in the majority of cases, the remaining models—lognormal, generalized inverse Gaussian (diffusion and nondiffusion), and gamma—together provide a systematic framework to: 1) evaluate the presence of inverse Gaussian structure in EDA using diffusion models; and 2) enhance statistical descriptions using models capable of capturing a range of tail behavior properties.

Among the three subjects for whom the AIC and KS distance disagreed on the best fit model (subjects S01, S06, and S07), there was only one case (S07) in which this disagreement

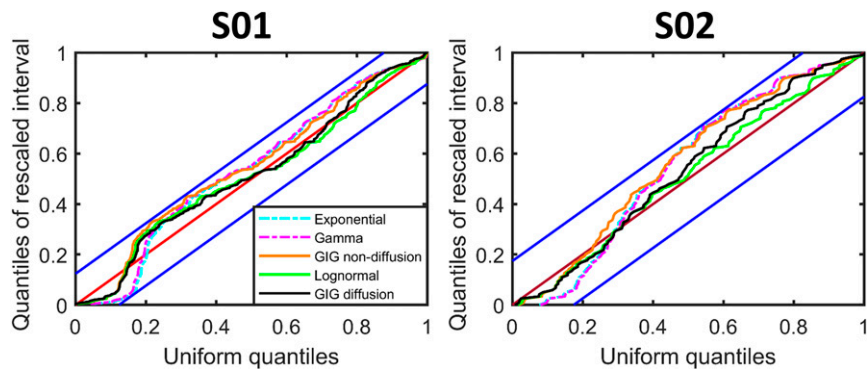


Fig. 2. KS plots of interpulse interval data for subjects S01 and S02 with 95% confidence bounds. All five models are shown against each other, with KS distances for each. A smaller KS distance, along with remaining fully within the 95% confidence bounds, indicates a better fit. The KS distances are ordered in each case from best to worst fit model. LogN, lognormal; GIG, generalized inverse Gaussian; Exp, exponential.

**Table 4. Settling rates for all distributions**

|                          | GIG diffusion  | GIG nondiffusion | Lognormal   | Gamma  | Exp                        |
|--------------------------|--|------------------|---|--|----------------------------|
| Density                  | $0.5(\psi\chi^{-1})^{(\lambda/2)}(K_{\lambda}(\psi\chi)^{1/2})^{-1}\chi^{\lambda-1}\exp(-0.5(\psi\chi + \chi\chi^{-1}))$ |                  | $(2\pi\chi^2\sigma^2)^{-1/2}\exp(-0.5\sigma^{-2}(\ln\chi - \mu)^2)$ | $\beta^{\alpha}(\Gamma(\alpha))^{-1}\chi^{\alpha-1}e^{-\beta\chi}$ | $\lambda e^{-\lambda\chi}$ |
| Settling rate ( $SR_f$ ) |  | $\frac{\psi}{2}$ | 0   | $\beta$  | $\lambda$                  |
| Tail classification      |  | Medium           | Heavy   | Medium   | Medium                     |
| S01                      | 0.0214   | 0.0553           | 0   | 0.071  | 0.068                      |
| S02                      | 0.0119   | 0.0299           | 0   | 0.035  | 0.036                      |
| S03                      | 0.0089   | 0.0244           | 0   | 0.026  | 0.031                      |
| S04                      | 0.0328   | 0.0639           | 0   | 0.120  | 0.062                      |
| S05                      | 0.0487   | 0.0957           | 0   | 0.184  | 0.094                      |
| S06                      | 0.0038   | 0.0164           | 0   | 0.017  | 0.025                      |
| S07                      | 0.0323   | 0.0566           | 0   | 0.107  | 0.049                      |
| S08                      | 0.0725   | 0.1203           | 0   | 0.198  | 0.096                      |
| S09                      | 0.0350   | 0.0764           | 0   | 0.142  | 0.083                      |
| S10                      | 0.0185   | 0.0332           | 0   | 0.056  | 0.029                      |
| S11                      | 0.0120   | 0.0288           | 0   | 0.044  | 0.033                      |

A lower settling rate indicates a heavier tail. GIG, generalized inverse Gaussian; Exp, exponential.

predicted very different tail behavior. In this case, the AIC and KS distance were divided between the lognormal and the gamma as the best model, one of which predicted a very heavy tail while the other predicted a very light tail. However, the tail of the interpulse interval distribution may have been relatively under-represented for this subject, so the fit was primarily based on the shape near the mode, which both the lognormal and gamma distributions fit well. With a data collection period longer than 1 h, we would most certainly arrive at a more accurate description of the tail behavior.

## Discussion

In this study, we used EDA data collected from 11 healthy volunteers at rest to test our hypothesis that EDA contains highly regular statistical structure that is consistent with integrate-and-fire physiology that describes sweat production and release. To do that, we fit probability models to the EDA time series and quantified the goodness of fit using AIC and KS distance. We also assessed tail behavior by computing the settling rates. Our findings show that for each of the 11 data series, the model fits and tail behavior were consistent with integrate-and-fire sweat gland physiology. There was also room for statistical improvements to capture deviations due to EDA data reflecting the simultaneous activity of many sweat glands.

The physiology of sweat gland activity predicts that the interpulse intervals for sweat gland pulses should obey an inverse Gaussian distribution. This is an elementary model for processes where there is a build-up of a continuous variable to a threshold. Crossing the threshold leads to an observed event. Here, the build-up is the accumulation of sweat in the gland in response to sympathetic stimulation. The observed event is the GSR pulse measured as EDA. Our results show that the EDA data are consistent with an inverse Gaussian model for all subjects (Table 3), and the inverse Gaussian is also the optimal integrate-and-fire probability model ( $\lambda = -0.5$ ) from the entire class of generalized inverse Gaussian integrate-and-fire models fit to these data (Table 1).

We refined our hypothesis further by considering that measured EDA is the aggregation of data from hundreds of sweat glands. This predicts that the interpulse intervals could likely follow a mixture of inverse Gaussian models. This mixture could deviate from a single inverse Gaussian in tail behavior, which can be captured by other non-integrate-and-fire models. Lighter tails were modeled as gamma or generalized inverse Gaussian non-integrate-and-fire models, whereas heavier tails were modeled as lognormal (Table 3). More of the data are consistent with the heavier tail lognormal distribution, suggesting more frequent

longer interpulse intervals across sweat glands than would be predicted by a homogeneous inverse Gaussian model. Furthermore, we did not observe any multimodal structure in the EDA suggesting a degree of coherent behavior among the sweat glands in the recording area. The non-integrate-and-fire models provide a systematic framework to make individualized improvements in the statistical fits. However, the fact remains that there is an inverse Gaussian model that is an accurate statistical description of the data for each subject. This reinforces the idea that the statistical structure in the data are fundamentally guided by the physiology of sweat gland activity which can be approximated well as an elementary integrate-and-fire process which we took to be a Gaussian random walk with drift diffusion.

Our results link directly the physiology of sweat glands and the statistical structure of the EDA data collected at the skin surface. Current detailed signal processing methods for EDA analysis require significant computational complexity (6–10). However, looking to the physiology provided a principled framework by which to drastically reduce model dimensionality—all of our models had only one, two, or three parameters—and increase the accuracy of the data description. This result has implications for understanding and tracking the sympathetic activity in the autonomic nervous system in a more informed way.

Several important extensions are possible in future work. We will study EDA pulse amplitudes along with interpulse intervals, by taking into account that both arise from the same integrate-and-fire process. Having established a natural point process structure in the EDA time series under approximately stable conditions, we can now study their dynamics over longer time periods, by applying history-dependent inverse Gaussian models like those developed for heart rate variability (31–35). These more detailed models could include other relevant covariates such as skin and environmental temperature. We will also study EDA in other contexts, such as under different emotional conditions, during sleep, in response to painful stimuli, and under general anesthesia. Our findings provide a principled, physiologically based approach for extending EDA analyses to these more complex and important applications.

**Data Availability.** Anonymized electrodermal activity time series and code for analysis have been deposited in PhysioNet (<https://physionet.org/>) (19).

**ACKNOWLEDGMENTS.** We thank the Massachusetts Institute of Technology Clinical Research Center staff. This work was partially funded by funds from the Picower Institute for Learning and Memory, the NSF Graduate Research Fellowship Program, and NIH Award P01-GM118629 (to E.N.B.).

1. W. Boucsein, *Electrodermal Activity*, (Springer, 2012).
2. Lynne Bell et al., Beyond self-report: A review of physiological and neuroscientific methods to investigate consumer behavior. *Frontiers in Psychology* **9**, 10.3389/fpsyg.2018.01655 (2018).
3. H. Storm et al., Skin conductance correlates with perioperative stress. *Acta Anaesthesiol. Scand.* **46**, 887–895 (2002).
4. H. Storm, M. Shafiei, K. Myre, J. Raeder, Palmar skin conductance compared to a developed stress score and to noxious and awakening stimuli on patients in anaesthesia. *Acta Anaesthesiol. Scand.* **49**, 798–803 (2005).
5. A. Sano, R. W. Picard, R. Stickgold, Quantitative analysis of wrist electrodermal activity during sleep. *Int. J. Psychophysiol.* **94**, 382–389 (2014).
6. M. Benedek, C. Kaernbach, Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology* **47**, 647–658 (2010).
7. M. Benedek, C. Kaernbach, A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* **190**, 80–91 (2010).
8. D. M. Alexander et al., Separating individual skin conductance responses in a short interstimulus-interval paradigm. *J. Neurosci. Methods* **146**, 116–123 (2005).
9. A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, L. Citi, cvxEDA: A convex optimization approach to electrodermal activity processing. *IEEE Trans. Biomed. Eng.* **63**, 797–804 (2016).
10. R. Faghih et al., “Characterization of fear conditioning and fear extinction by analysis of electrodermal activity” in 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), J. Patton, Ed., (IEEE Engineering in Medicine and Biology Society, 2015), pp. 7814–7818.
11. M. Kunitomo, K. Kirnō, M. Elam, B. G. Wallin, Neuroeffector characteristics of sweat glands in the human hand activated by regular neural stimuli. *J. Physiol.* **442**, 391–411 (1991).
12. M. Kunitomo, K. Kirnō, M. Elam, T. Karlsson, B. G. Wallin, Non-linearity of skin resistance response to intraneural electrical stimulation of sudomotor nerves. *Acta Physiol. Scand.* **146**, 385–392 (1992).
13. E. Schrödinger, Zur theorie der fall-und steigversuche an teilchen mit brownischer bewegung. *Phys. Z.* **16**, 289–295 (1915).
14. G. L. Gerstein, B. Mandelbrot, Random walk models for the spike activity of a single neuron. *Biophys. J.* **4**, 41–68 (1964).
15. R. Chhikara, J. Folks, *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*, (Marcel Dekker, New York, 1989).
16. S. Iyengar, Q. Liao, Modeling neural activity using the generalized inverse Gaussian distribution. *Biol. Cybern.* **77**, 289–295 (1997).
17. L. Halliwell, Classifying the tails of loss distributions. *Casualty Actuar. Soc.* **2**, 1–27 (2013).
18. J. Folks, R. Chhikara, The inverse Gaussian distribution and its statistical application—A review. *J. Royal Stat. Soc. Ser. B (Methodological)* **40**, 263–289 (1978).
19. S. Subramanian, R. Barbieri, E. Brown. Electrodermal activity of healthy volunteers while awake and at rest (version 1.0). PhysioNet. <https://doi.org/10.13026/art2540>. Deposited 4 August 2020.
20. S. Subramanian, R. Barbieri, E. N. Brown, “A systematic method for preprocessing and analyzing electrodermal activity” in *Proceedings of the 41st IEEE International Conference on Engineering in Medicine and Biology*, R. Barbieri, Ed. (IEEE Engineering in Medicine and Biology Society, 2019), pp. 6902–6905.
21. R. G. Hallin, H. E. Torebjörk, Afferent and efferent C units recorded from human skin nerves in situ. A preliminary report. *Acta Soc. Med. Ups.* **75**, 277–281 (1970).
22. K. E. Hagbarth, R. G. Hallin, A. Hongell, H. E. Torebjörk, B. G. Wallin, General characteristics of sympathetic activity in human skin nerves. *Acta Physiol. Scand.* **84**, 164–176 (1972).
23. R. G. Hallin, H. E. Torebjörk, Single unit sympathetic activity in human skin nerves during rest and various manoeuvres. *Acta Physiol. Scand.* **92**, 303–317 (1974).
24. B. G. Wallin, Sympathetic nerve activity underlying electrodermal and cardiovascular reactions in man. *Psychophysiology* **18**, 470–476 (1981).
25. B. G. Wallin, J. Fagius, Peripheral sympathetic neural activity in conscious humans. *Annu. Rev. Physiol.* **50**, 565–576 (1988).
26. A. Uncini, S. L. Pullman, R. E. Lovelace, D. Gambi, The sympathetic skin response: Normal values, elucidation of afferent components and application limits. *J. Neurol. Sci.* **87**, 299–306 (1988).
27. Y. Pawitan, *In All Likelihood*, (Clarendon Press, Oxford, 2013).
28. B. Jorgensen, *Statistical Properties of the Generalized Inverse Gaussian Distribution*, (Springer-Verlag, New York, 1982).
29. E. N. Brown, R. Barbieri, V. Ventura, R. E. Kass, L. M. Frank, The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput.* **14**, 325–346 (2002).
30. D. Daley, D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*, (Springer, 2007).
31. R. Barbieri, E. C. Matten, A. A. Alabi, 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.* **288**, H424–H435 (2005).
32. R. Barbieri, E. N. Brown, Analysis of heartbeat dynamics by point process adaptive filtering. *IEEE Trans. Biomed. Eng.* **53**, 4–12 (2006).
33. R. Barbieri, E. N. Brown, Application of dynamic point process models to cardiovascular control. *Biosystems* **93**, 120–125 (2008).
34. G. Valenza et al., Instantaneous monitoring of heart beat dynamics during anesthesia and sedation. *J. Comput. Surg.* **3**, 1–18 (2014).
35. S. Subramanian, R. Barbieri, E. N. Brown, “A point process characterization of electrodermal activity” in *Proceedings of the 40th IEEE International Conference on Engineering in Medicine and Biology*, J. Patton, Ed. (IEEE Engineering in Medicine and Biology Society, 2018), pp. 37–40.