Article

Model-Driven Analysis of ECG Using Reinforcement Learning

Christian O'Reilly [1,2,3,4]* , **Sai Durga Rithvik Oruganti** [1,2], **Deepa Tilwani** [1,2,3,4] **and Jessica Bradshaw** [3,4,5]

1   Artificial Intelligence Institute of South Carolina
2   Department of Computer Science and Engineering, University of South Carolina
3   Carolina Autism and Neurodevelopment Research Center, University of South Carolina
4   Institute for Mind and Brain, University of South Carolina
5   Department of Psychology, University of South Carolina
*   Correspondence: christian.oreilly@sc.edu

Abstract: Modeling is essential to understand better the generative mechanisms responsible for experimental observations gathered from complex systems. In this work, we are using such an approach to analyze the electrocardiogram (ECG). We present a systematic framework to decompose ECG signals into sums of overlapping lognormal components. We used reinforcement learning to train a deep neural network to estimate the modeling parameters from ECG recorded in babies of 1 to 24 months of age. We demonstrate this model-driven approach by showing how the extracted parameters vary with age. After correction for multiple tests, 10 of 24 modeling parameters showed statistical significance below the 0.01 threshold, with absolute Kendall rank correlation coefficients in the [0.27, 0.51] range. We presented a model-driven approach to the analysis of ECG. The impact of this framework on fundamental science and clinical applications is likely to be increased by further refining the modeling of the physiological mechanisms generating the ECG. By improving the physiological interpretability, this approach can provide a window into latent variables important for understanding the heart-beating process and its control by the autonomous nervous system.

Keywords: ECG; modeling; reinforcement learning; lognormal; autonomic nervous system; model-driven analysis

1. Introduction

Heart disease is the leading cause of death in America, responsible for 597,689 deaths in 2010, according to the US Centers for Disease Control and Prevention [1]. This number further increased to 693,021 in 2021 [2]. An average healthy human heart pumps approximately 2,000 gallons of blood daily[3], making this organ the driving force of the circulatory system. For that reason, it is crucial to track how the heart is performing. The electrocardiogram (ECG) is a simple and non-invasive way to record the heart's electrical activity due to the depolarization and repolarization of its membrane of muscle fibers during the cardiac cycle [4]. ECG is recorded by placing electrodes in direct contact with the skin to capture the electrical activity. ECG is one of the primary tools to check for irregular heart rhythm (arrhythmia) and help diagnose heart diseases, poor blood flow, and other health issues. Further, ECG activity and its interaction with the respiratory rhythm, a phenomenon known as respiratory sinus arrhythmia, provides a window into the regulation of the autonomic nervous system (ANS) with potential applications for a myriad of conditions associated with dysregulated ANS control. In neonatal care, because of the relationship between the immune system and the control of the heart rhythm by the ANS, ECG is also used to monitor the health of newborns and predict the occurrence of sepsis [5,6].

Over the years, many researchers have developed models to better understand cardiac physiology and its assessment through ECG, including approaches based on reaction-diffusion, oscillators, and transforms [7–9]. Each model has specific advantages and

disadvantages, so it is important to understand its properties before deploying it for specific applications.

Methods based on reaction diffusion and oscillators are similar. Although they adopt different perspectives, both involve dynamical models expressed as systems of Ordinary Differential Equations (ODE). The main drawback with dynamical models is that their parameters need to be limited within specific ranges of values for producing realistic ECG signals [7]. These models can exhibit chaotic behavior when their parameters are outside these ranges. Further, ranges that generate ECG signals with a particular rhythm can be incredibly narrow. The subspace of acceptable parameters can describe a complex manifold in a high-dimensional space (depending on the number of modeling parameters), so simple box constraints are generally insufficient for such applications. Our initial attempts at dynamically estimating parameters for such ODEs to track the modifications of ECG recordings (i.e., inverse modeling) without causing these models to enter a chaotic regime have revealed this problem to be thorny (unpublished results). Thus, although such models can be insightful for forward modeling (i.e., simulating realistic ECG signals), they are difficult to deploy for inverse modeling and inference (i.e., estimating modeling parameter values based on recorded ECG).

Transform-based modeling has also been proposed to study ECG signals. In this approach, each ECG cycle is decomposed into P, Q, R, S, and T waves, each modeled using elementary functions and represented by their corresponding Fourier series [9,10]. Such an approach may perform well on stable ECG (i.e., signals showing no variation from one cycle to the next). However, because it relies on Fourier series and its decomposition into infinite sinusoidal waves, it is likely to be of limited use to study temporal (as opposed to spectral) variations in the ECG signal from beat to beat.

Similarly, the individual wave of the PQRST complex can be decomposed and modeled directly in their native space (i.e., as time-series). For example, such an approach has been implemented using pairs of normal components [11]. However, a significant body of literature shows the practical and theoritical advantages of using lognormal rather than normal equations for decomposing biosignals. This approach has been pioneered in the context of human movement analysis [12]. In this paper, we build on that literature to develop a novel model of ECG using lognormal decomposition and reinforcement learning. This work paves the way for the analysis of ECG based on generative, mechanistic models. We aim to use this framework to gain further insights into latent variables involved in normal and abnormal ECG rhythms and better understand how the ANS controls cardiac activity.

In the following sections, we first introduce our modeling approach by reviewing 1) the biological process generating ECG signals and 2) the decomposition of biosignals into sums of lognormals. We then provide further details on our methods, including a description of the participant samples, the ECG preprocessing steps, the PQRST lognormal prototype used to constrain inverse modeling, and the deep reinforcement learning framework used to find optimal lognormal parameters. We then present results for an example use-case, showing how we can use this lognormal decomposition of ECG signal to study the developmental effect of ECG in infants from 1 to 24 months of age. We conclude this paper by discussing current limitations and future opportunities offered by this novel approach.

2. The generation of ECG

2.1. The cardiac rhythm and its modulation

Understanding the process leading to the generation of biologicals is essential for conducting powerful and meaningful analyses. ECG signals collected by surface leads positioned on the chest are due to the volume conduction of bioelectrical currents in the body. Thus, first, we need to understand how heartbeats lead to current dipoles to properly analyze ECG signals.

The sinoatrial node (SA) – a small region of the upper right atrium made of cells naturally exhibiting cyclic activity – drives the generation of the heart rhythm. The rhyth-

mogenicity of these cells is due to the constant depolarizing effect of leaky sodium channels. However, the time required for these cells to depolarize enough to reach their firing threshold can be increased or decreased by the opening or closing of calcium channels. By regulating the state of these calcium channels through neuromodulation, the autonomic nervous system can control the interval between two beats. This inter-beat interval is also known as the RR interval since it is generally defined as the duration between the peaks of two subsequent R waves. The balance between the sympathetic and the parasympathetic system controls the opening/closing of these channels through a molecular cascade involving second messengers. The sympathetic system acts through norepinephrine (neural origin) or epinephrine (distributed through blood circulation). This neuromodulator triggers the phosphorylation of calcium channels, thereby increasing the calcium inflow and causing earlier depolarization of SA cells (faster cardiac rhythm). The parasympathetic system acts in the opposite way through acetylcholine which inhibits this process, reducing calcium inflow and prolonging the inter-beat interval. Thus, the central control of the heart and its modulation by various physiological processes (e.g., respiratory sinus arrhythmia) can be captured by modeling the modulation of the calcium current in SA cells.

2.2. PQRST waves

The ECG signal is constituted of the repetition of a stereotypical waveform, referred to as the PQRST complex (see Figure 1 for a visual representation). Each heartbeat is responsible for the generation of one such complex. As previously mentioned, cells in the SA node, located on the top-right portion of the right atrium, initiate this process by firing an action potential. This action potential propagates outwardly from one cell to its neighboring atrial myocardial cells through gap junctions, generating wave of depolarization. The far-field caused by this wave can be modeled as coming from an equivalent electrical dipole that points toward the lower left side of the heart. This process is responsible for the positive P-wave on a traditional ECG with electrodes placed as shown in Figure 1. This progation mediated by gap-junction is of moderate speed, causing a relatively low-amplitude, wide-spread positive deflection. A resistive fibrous membrane separates the myocardial cells from the atrial and the ventricular part of the heart, stopping the depolarization wave in the atria from propagating to the ventricles. The atrioventricular (AV) node provides the only window for this depolarization to travel toward the ventricles. The cells of AV node are specialized for slow conduction, allowing for the delay between atrial and ventricular contraction required to support efficient pumping. Because the propagation through the AV node involves a relatively small number of cells, their contribution to electrical currents at distant locations (i.e., at the site of the electrodes) is insufficient to generate an equivalent dipole visible on the ECG. Thus, this phase of the cardiac rhythms results in the flat line characteristic of the PQ segment (also sometimes referred to as the PR segment since the Q component is not always visible on ECG).

As the depolarization slowly crosses the AV Node, it reaches the fast-conducting His-Purkinje network, and triggers the depolarization of the lower-left section of the ventricular septum. The resulting wave travels toward the upper part of the septum, causing an equivalent depolarization dipole pointing toward the negative ECG lead and reflected as the fast and low-amplitude negative Q wave. This component is followed immediately by the main depolarization of the ventricles, causing a short-lived and large-amplitude equivalent dipole pointing toward the positive lead and visible on the ECG as the R wave. This wave then propagates to the basal region of the ventricle, generating an upward-pointing dipole identifiable as the negative S wave. Then, myocardial cells exhibit a sustained depolarization plateau because their repolarizing potassium current gets counteracted by the opening of their calcium channels (depolarizing). This plateau is reflected on the ECG as the electrical silence during the ST segment. At the end of this segment, the calcium channels close, and the outward potassium currents repolarize the myocardial cells. This repolarization starts with the cells on the outer layer of the ventricular tissues, propagating inward and generating a negative equivalent dipole pointing toward
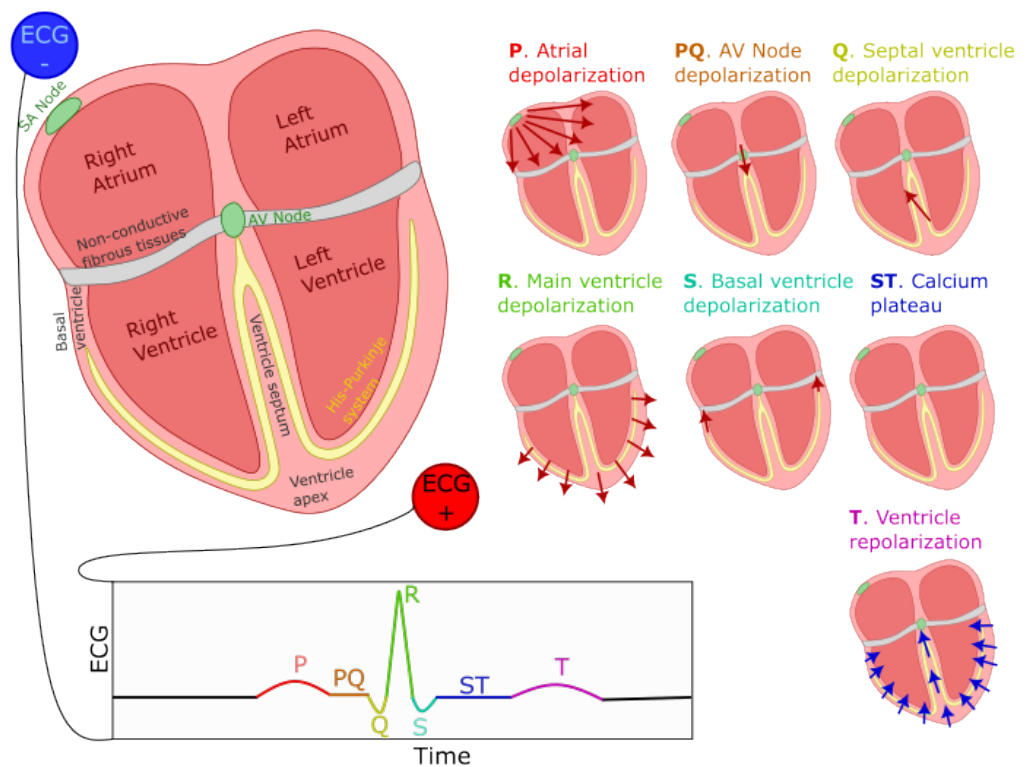
Figure 1. Generation of the ECG PQRST complex. The left-right image shows the positioning of ECG electrodes. Seven smaller representations of the heart illustrate the direction and polarity of electrical dipoles generated at each phase of the heartbeat. The bottom of the figure displays the resulting PQRST waveform.

the negative lead responsible for the T component of the ECG. This sequence of events is depicted in Figure 1.

Further, we should note that atrial repolarization happens as the ventricle depolarizes. Since the ventricles contain many more cells than the atrium, it is responsible for a much larger equivalent dipole. This simultaneous activity *obfuscates* the effect of the atrium repolarization on the ECG.

The description of this process shows that ECG events offer a direct window into latent neural and ionic properties, as illustrated by the direct relationship between the width of the ST segments and the dynamics of the calcium channels generating the depolarization plateau in myocardial cells. We can exploit this relation when inferring from ECG the properties of the cardiac process and its autonomic control.

3. Materials & Methods

3.1. Lognormal modeling

For this work, we used an approach similar to [11] but adopting a sum of time-shifted and scaled lognormal density functions rather than scaled normal density functions. We favor lognormal over normal functions for such modeling for multiple reasons. First, lognormal transfert functions are causal. They have a support that starts at a specific time t_0 (see (1)), as opposed to an $]-\infty, +\infty[$ support for the normal equation. Further, the use of sums of lognormal functions for modeling biological time-series has been well-developed for the study of motor control [13–16]. The biological relevance of the lognormal as a transfer function is well-rooted mathematically through the Central Limit Theorem, as applied to biological networks of subprocesses coupled through proportionality. This contrasts with the response of a system made of independent and additive subprocesses, which converges, under the same theorem, toward a normal response. These ideas have been developed in detail in the context of the Lognormality Principle [17,18].

$$\Lambda_j(t; P_j) = \frac{D_j exp\left(\frac{[ln(t-t_{0j})-\mu_j]^2}{-2\sigma_j^2}\right)}{\sigma_j(t-t_{0j})\sqrt{2\pi}} \tag{1}$$

$$P_j = \{\mu_j, \sigma_j, D_j, t_{0j}\} \tag{2}$$

In terms of interpreting the meaning of lognormal parameters, t_0 (in seconds) represents the time when the lognormal process was initiated. Since lognormal-shaped waveforms are emerging from the convergence of small effects associated with many coupled subsystems, we can see t_0 as the moment the first cells generated an action potential and triggered an avalanche of subsequent action potentials in the network of neighboring cells. The parameter D represents the amplitude of the equivalent electrical dipole created by this process. More precisely, it is equal to the integration over time of the depolarization or repolarization waves (in $V * s$). The parameters μ and σ represent, on a logarithmic scale, the time delay and response time of the network of excitable cells. Therefore, these parameters are emergent properties due to the speed of propagation of the electrical waves in these networks. We expect their values to reflect biological properties that modulate the propagation speed, such as the strength of gap-junction coupling between neighboring cells.

In this approach, a PQRST complex is constituted of six partly overlapping lognormal equations (1). The P, Q, R, and S waves are all represented by a single equation. For expediency, we accounted for the negative skew of the T wave by representing it with the subtraction of two lognormals (noted T_+ and T_-). Further investigation will be required to understand the physiological underpinning of this negative skew and, consequently, the most biologically-relevant way to model it.

3.2. ECG Sample

We collected 150 ECG recordings from 40 infants between 1 week and 24 months of age. A small device containing an ECG sensor (Actiwave Cardio; CamNTech) with two electrodes was placed on the infants' chest. Raw ECG signals were recorded in a laboratory setting while the infants participated in different naturalistic experimental paradigms involving self-initiated interaction with objects or their parents.

Most recordings were sampled at 1024 Hz (N=107). ECG recorded at a different rate (512 Hz: N=13; 128 Hz: N=30) were upsampled to 1024 Hz to ensure consistency. We segmented the data according to markers labeling the beginning and the end of a given experimental condition. Although we did not study the effect of experimental conditions for this paper, segmenting signals according to these markers ensured that we kept only valid ECG sections. For example, initial data points recorded before the start of the experiment and before the sensor was placed on the infant's chest were automatically discarded by this procedure. Most of the 555 segments obtained are of 5-25 minutes duration, but their distribution has a long tail, with the longest segment being 150 minutes long (percentiles (min): 0: 0.25; 10: 4.74; 25: 8.35; 50: 10.03: 75: 16.23; 90: 25.12; 100: 150.25).

3.3. Preprocessing

Heartbeats were automatically detected using HeartPy [19]. We rejected five segments because HeartPy could not process the signal (i.e., it threw an exception) or because it detexted less than 20 beats. For the remaining segments, we epoched and normalized the beat duration to make beats comparable. Considering t_1, t_2, and t_3 three subsequent R peaks, the epoched and normalized version of the peak corresponding to t_2 is obtained by 1) linearly interpolating the ECG between t_1 and t_2 over 250 regularly-spaced samples, 2) doing the same for the segment from t_2 to t_3, and 3) concatenating these two segments.

After this process, all beats are defined as 500-sample epochs starting at the previous R-peak   211
and ending at the next R-peak, with their own R-peak centered in that window. For plotting,   212
we map these epochs to the [-1, 1] interval and refer to the variable along that dimension   213
as the *normalized time*. For each segment, we computed a *mean beat* ($< x >$) by averaging   214
across these epochs. We characterized the stability of the PQRST profile within a segment   215
by computing the following signal-to-noise (SNR) ratio: 216

$$SNR = 10 * log_{10}\left(\frac{\sum_{i=1}^{500} < x_i >^2}{\sum_{i=1}^{500}(x_i - < x_i >)^2}\right) \tag{3}$$

Since the goal of this study is not to develop preprocessing techniques to clean noisy   217
ECG signals, we rejected all segments with an $SNR < 5dB$ (N=116). Figure 2 illustrate   218
how the signal quality varies with SNR. Eliminating noisy PQRST complexes allows us to   219
concentrate on the specific contribution of this work, i.e., the model-driven analysis of ECG.   220
This exclusion criterion left 434 valid segments for analysis.   221
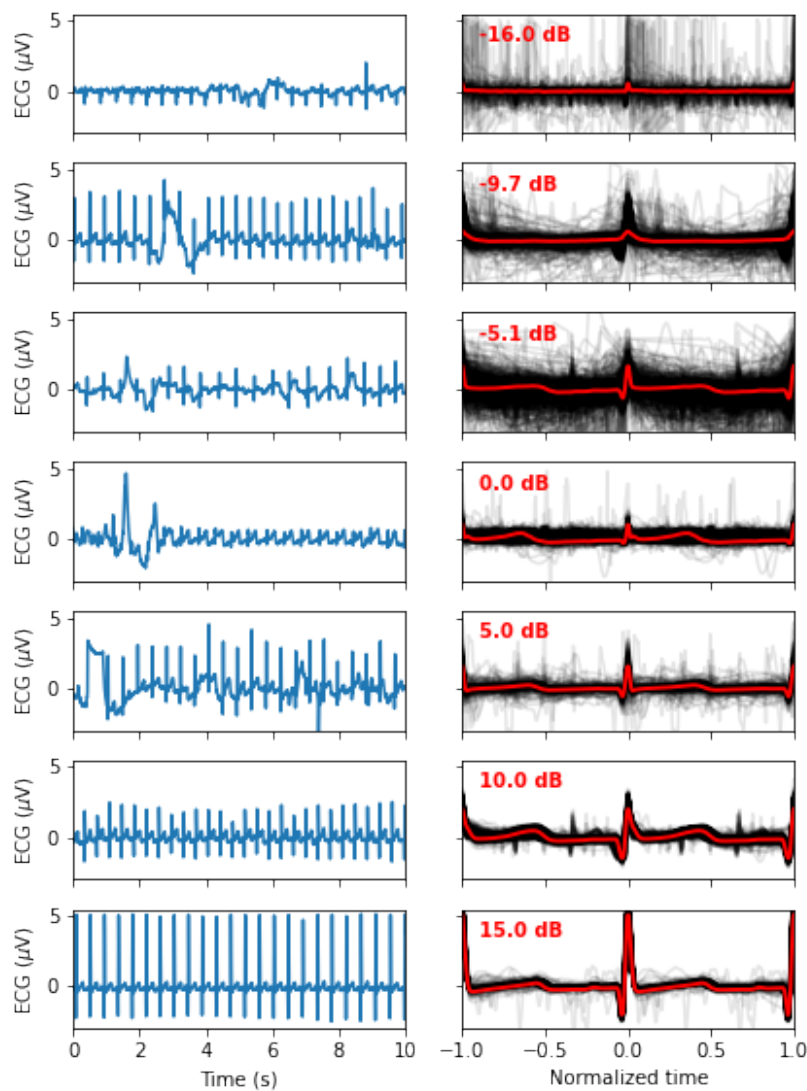

Figure 2. Signal quality as a function of the SNR. The left panels show 10s of the original signals, whereas the right panel shows all the normalized beats (in light grey) along with the mean beat (red). From top to bottom, we display as examples the recordings with an SNR closest to -15, -10, -5, 0, 5, 10, and 15 dB.

3.4. The PQRST lognormal prototype

The inverse modeling (i.e., curve fitting) constitutes a significant difficulty when modeling time-series with sums of lognormal functions. Various algorithms have been developed to extract parameters for such models [20–23]. These extractors perform reasonably well, but the difficulty of this regression problem leaves a place for improvement. For our application, we want to benefit from the ability to map lognormal components between time-series provided by the prototype-based extraction [21]. We can adopt this approach where modeling time-series with a stereotypical shape, as is the case for the PQRST complex. With that method, we embed our *prior* knowledge of the problem and our hypotheses about the generation of the observed signals by defining a starting solution (i.e., a *prototype*) made by a series of lognormal components. Then, we obtain the final solutions by adjusting this prototype to every sample, fitting the lognormal parameters to account for individual differences.

However, one issue with the least-square regression of prototypes is that such optimal solutions, if not properly constrained, can drift away from the conceptual modeling of the generative process captured by the prototype. For example, in this case, a lognormal component initially set in the prototype to model the P components can drift and end up modeling slight noise-related deviations in high-amplitude components such as the R component. Although this is not a problem for obtaining a good curve fitting, it becomes an issue when studying the distributions of the parameters, e.g., associated with the P component. That is, although potentially providing an optimal curve fitting, such optimization may not hit the right balance between fitting the observation and providing a plausible solution in terms of the biological or physiological mechanisms of the process being modeled.

To help constrain our prototype, we used upper and lower bounds for the PQRST complexes. These bounds are defined analytically from box constraints on the lognormal parameters (4-7) (see [20] for the derivation of these bounds). The Figure 3 illustrates the envelope of all PQRST complexes generated by this *sigma-lognormal* (i.e., sum of lognormals) model, given the bounds we selected (see Table 1). We chose these bounds as a trade-off between encompassing the mean PQRST complexes of all recordings (except some apparent outliers) while constraining the possible solutions to minimize the proportion of the parameter space that can produce PQRST time-series that are not physiologically plausible. These bounds were derived from the parameters P_P of the prototype $\Sigma\Lambda_P$ such that $P_\pm = P_P \pm 0.2|P_P|$, except for 1) the value of the $D_\pm$ parameters being ceiled or floored to 0 when their interval intersects 0, and 2) for $D_{T+} = D_{T+P} + 0.4D_{T+P}$. The first exception ensures that the qualitative interpretation of the different waves remains intact. A sign inversion for a D parameter would mean changing a depolarization wave into a repolarization wave, or vice-versa. Such a conversion is not possible in normal physiological conditions. For example, the P wave can only be due to the depolarization of the myocardial cells of the atrium. We will never observe a hyperpolarizing wave, rather than depolarizing wave, being caused by an inversion of this process. Alternatively, such polarity reversal could be due to technical issues, such as a misplacement of the electrodes. However, these acquisition problems should be managed when preprocessing the recordings rather than being captured by the model parameters. We used the second exception for a pragmatic reason; we needed to allow the D_T parameter to vary within a wider range to include most of the recorded PQRST complexes within the established bounds. With respect to the $|P_P|$ term, the $|...|$ notation denotes the absolute value of P_P, and the indices in P_P, P_+, and D_{TP} stand for "prototype", "upper bound", and "T wave of the prototype", respectively. Other indices are derived using the same logic. The prototype was obtained by manually fitting a sigma-lognormal curve that captures the general trend of our sample of beat profiles. Its parameters are given in Table 2.

$$\Lambda_- = \min \begin{cases} \Lambda(t - t_{0+}; \mu_+, \sigma_-) \\ \Lambda(t - t_{0+}; \mu_+, \sigma_+) \\ \Lambda(t - t_{0-}; \mu_+, \sigma_+) \\ \Lambda(t - t_{0-}; \mu_-, \sigma_+) \\ \Lambda(t - t_{0-}; \mu_-, \sigma_-) \end{cases} \tag{4}$$

$$\Lambda_+ = \begin{cases} 0 & if \quad t \le t_{0-}] \\ \Lambda(t - t_{0-}; \mu_-, \sigma_+) & if \quad t \in]t_{0-}, t_{0-} + e^{\mu_- - \sigma_+}] \\ \Lambda(t - t_{0-}; \mu_-, \sigma = \mu_- - ln(t - t_{0-1})) & if \quad t \in]t_{0-} + e^{\mu_- - \sigma_+}, t_{0-} + e^{\mu_- - \sigma_-}] \\ \Lambda(t - t_{0-}; \mu_-, \sigma_-) & if \quad t \in]t_{0-} + e^{\mu_- - \sigma_-}, t_{0-} + e^{\mu_- - \sigma_-^2}] \\ \Lambda(e^{\mu_- - \sigma_-^2}; \mu_-, \sigma_-) & if \quad t \in]t_{0-} + e^{\mu_- - \sigma_-^2}, t_{0+} + e^{\mu_- - \sigma_-^2}] \\ \Lambda(t - t_{0+}; \mu_-, \sigma_-) & if \quad t \in]t_{0+} + e^{\mu_- - \sigma_-^2}, t_{0+} + e^{\mu_-}] \\ \Lambda(t - t_{0+}; \mu = ln(t - t_{0+}), \sigma_-) & if \quad t \in]t_{0+} + e^{\mu_-}, t_{0+} + e^{\mu_+}] \\ \Lambda(t - t_{0+}; \mu_+, \sigma_-) & if \quad t \in]t_{0+} + e^{\mu_+}, t_{0+} + e^{\mu_+ + \sigma_-}] \\ \Lambda(t - t_{0+}; \mu_+, \sigma = ln(t - t_{0+}) - \mu_+) & if \quad t \in]t_{0+} + e^{\mu_+ + \sigma_-}, t_{0+} + e^{\mu_+ + \sigma_+}] \\ \Lambda(t - t_{0+}; \mu_+, \sigma_+) & if \quad t > t_{0+} + e^{\mu_+ + \sigma_+} \end{cases} \tag{5}$$

$$\Sigma\Lambda_- = \sum_i D_{i-} \begin{cases} \Lambda_{i-} & if \quad D_{i-} > 0 \\ \Lambda_{i+} & else \end{cases} \tag{6}$$

$$\Sigma\Lambda_+ = \sum_i D_{i+} \begin{cases} \Lambda_{i+} & if \quad D_{i+} > 0 \\ \Lambda_{i-} & else \end{cases} \tag{7}$$

Table 1. Parameter bounds. $D_\pm$ are in μV and $t_{0\pm}$ are using normalized time.

	μ_-	μ_+	σ_-	σ_+	t_{0-}	t_{0+}	D_-	D_+
P	-2.4	-1.6	0.08	0.12	-0.288	-0.192	0	3.6
Q	-3.6	-2.4	0.32	0.48	-0.096	-0.064	-60	0
R	-3.6	-2.4	0.2	0.3	-0.054	-0.036	0	96
S	-4.2	-2.8	0.32	0.48	0.012	0.018	-12	0
T_+	-1.2	-0.8	0.32	0.48	0.12	0.18	0	204
T_-	-1.2	-0.8	0.184	0.276	0.176	0.264	-144	0

3.5. Reinforcement learning for parameter extraction

For this project, we used the Stable-Baselines3 and OpenAI Gym Python packages to implement a deep reinforcement learning approach for fitting our lognormal model to PQRST time-series. We aim to improve over the grid optimization algorithm previously used to adjust the sigma-lognormal prototype to recorded time-series. We adopted reinforcement learning for this problem because it can find optimal solutions to problems for which the "true" or "best" solution is not known in advance, as long as we can define a measure of the quality of a solution. In our case, the root-mean-square fitting error can serve as such a measure. As opposed to classical supervised learning, reinforcement learning is not limited by the quality of the training data (i.e., it can adapt and become better than experts at a given task). Further, the reinforcement learning approach is interesting because the active development of Bayesian reinforcement learning [24]. We expect future Bayesian extensions to provide a more appropriate way to integrate prior knowledge into the estimation process by using prior distributions in the definition of the prototype instead of point values and box constraints (see the Discussion section).

Figure 4 shows a classic schematic representation of reinforcement learning, illustrating the interplay between the learning agent and its environment. In that paradigm, the policy
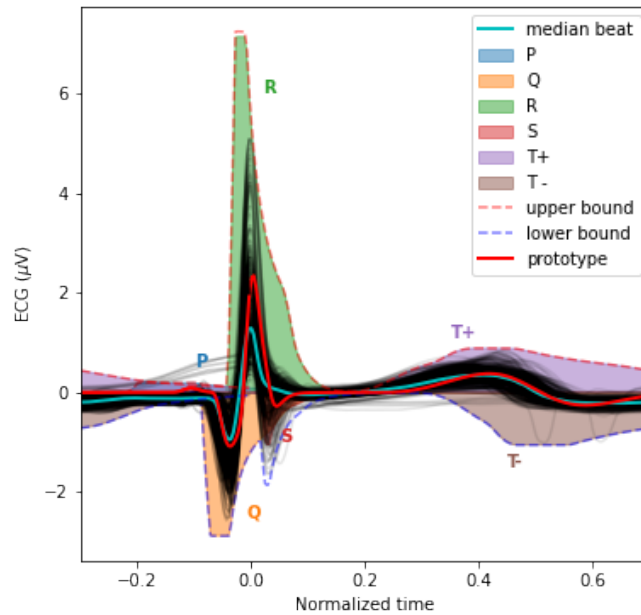
Figure 3. This plot overlays the mean PQRST for each valid segment (light grey) and their median (cyan). We also overlaid the sigma-lognormal prototype (red) and the upper (dashed red) and lower (dashed blue) envelop of the PQRST time-series that can be generated by the model when using parameters within the $\Sigma\Lambda_-$ and $\Sigma\Lambda_+$ bounds specified in Table 1. The space covered by each component is also shown using different colors.

maps the current state of the environment to the probability of performing given actions. This policy captures the transition probability of a Markovian decision process, as it does not depend on past states. In this context, the learning task boils down to optimizing the policy so that actions leading to the highest rewards have the highest probability of being executed. To implement such an approach, the space of possible actions, the observations space (i.e., the observable environment states), and the reward must first be defined.

Action space: Our model has 24 parameters. We consider these as latent variables. The parameter space is defined as a bounded box in $\mathbf{R}^{24}$, and the upper bound $\Sigma\Lambda_+$ and the lower bound $\Sigma\Lambda_-$ have already been defined in the previous section. We define a reference point as the middle of this bounded box, i.e., $\Sigma\Lambda_{ref} = (\Sigma\Lambda_+ + \Sigma\Lambda_+)/2$. At the start of the fitting process, the estimated solution is set equal to $\Sigma\hat{\Lambda}_0 = \Sigma\Lambda_{ref}$, where the index in $\Sigma\hat{\Lambda}_n$ indicates the step number, with 0 standing for the initial estimate. The action space is a box $A_{24} = \{(a_1, a_2, ..., a_{24}) | a_1, a_2, ..., a_{24} \in [-0.01, 0.01]\}$. At each step of the fitting process, the

Table 2. PQRST Sigma-Lognormal prototype. $D_\pm$ are in μV and $t_{0\pm}$ are using normalized time.

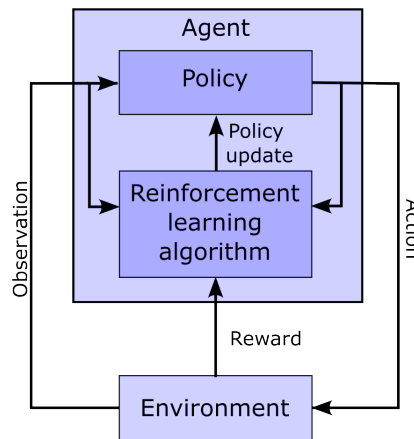|	μ	σ	t_0	D
P	-2.0	0.1	-0.24	3
Q	-3.0	0.4	-0.08	-50
R	-3.0	0.25	-0.045	80
S	-3.5	0.4	0.015	-10
T_+	-1.0	0.4	0.15	150
T_-	-1.0	0.23	0.22	-120

Figure 4. Schematic representation of the interaction between the agent and the environment in the context of reinforcement learning.

action $a \in A_{24}$ taken by the agent is used to update the estimated solution following the rule $\Sigma\hat{\Lambda}_{i+1} = \Sigma\hat{\Lambda}_i + a\Sigma\Lambda_{ref}$. However, we note that even within this relatively narrow box, the sigma-lognormal parameters can take values such that the order of the PQRST lognormal components gets changed (i.e., the Q component may move after the R component). Since such an inversion is contrary to the intent of the model, we further constrain this ordering. To this end, we first define the order of lognormal components by the timing of their peak. For a lognormal defined by parameters P_j (as defined in (2)), this peak can be shown to happen at time $t_{pj} = t_{0j} + e^{\mu_j} e^{-\sigma_j^2}$. Thus, we constrain the impact of the action on P_j so that it is canceled whenever it would results in $t_{pj+1} < t_{pj+1}$ (i.e., the action is such that it would cause the timing of the peaks of two consecutive components to get inverted), with j and $j+1$ referring to two consecutive lognormal components in the set of $\{P, Q, R, S, T_+, T_-\}$ components. We further limit the action so that the resulting parameters remain within the box defined by $\Sigma\Lambda_+$ and $\Sigma\Lambda_-$.

Observation space: The observation space for this model is a dictionary containing values for the estimated parameters and the fitting errors. The space for the fitting errors is a box of dimension 250. The PQRST signals are normalized to 500 points equally spaced between -1 and +1 (see the Preprocessing section), but we observe only half of this interval, between -0.3 and 0.7. ECG signals in the remaining portion of the normalized time interval are mostly flat and contain little information. We set lower and upper bounds for that box at 1.5 times the minimum and maximum values observed for each of these points across all recorded PQRST time-series. Observed values for the fitting error are taken as the difference between the PQRST time-series being analyzed and the signal synthesized from $\Sigma\hat{\Lambda}_i$. The observation space for the estimated parameters is the 24-dimensional box bounded by $\Sigma\Lambda_-$ and $\Sigma\Lambda_+$, and the observed values are $\Sigma\hat{\Lambda}_i$. We further normalized the observation space, as is generally recommended when using a heterogeneous observation space. We confirmed that the normalization of the observation space resulted in a significant improvement in performance for our application.

Reward: In reinforcement learning, the reward is the objective function to maximize, and it depends on the action performed. Thus, reinforcement learning algorithms aim at finding the action that maximizes the reward associated with any given state. For our application, the reward is defined as the difference in fitting SNR before and after the action was taken.

Training: The optimization of a PQRST time-series terminates when the algorithm reaches its maximal number of steps (1000) or when the agent fails to find a solution (i.e., a set of parameter values) improving over its best SNR for 100 consecutive steps. Every time the optimization of a PQRST time-series terminated, another time-series was picked at random, and a new optimization was started. We trained our model for a total of 3,000,000 steps, at

which point performances were stable, and no additional training seemed likely to provide any advantage.

For learning, we implemented a deep reinforcement learning solution based on the Proximal Policy Optimization (PPO) [25] scheme. This relatively new algorithm has been designed to control the size of the steps taken when doing the gradient-based optimization of the policy. Large steps in the optimization process can result in instability as they can push the optimizer in a distant subspace with a flat objective function, potentially trapping the optimizer in an unproductive and low-reward region of the solution space. Specifically, in this approach, the update rule for the policy is defined by the loss function

$$L(\theta) = \mathbb{E}\left[\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}\hat{A}_t\right] = \mathbb{E}\left[r_t(\theta)\hat{A}_t\right] \tag{8}$$

where π represents the probabilistic policy parameterized with a new (θ) and an old (θ_{old}) set of parameters, a_t is the action selected at time step t, s_t is the state of the environment at time t, and $\hat{A}_t$ is the advantage associated with a state, defined as the expected total reward discounted using a factor γ. This variable captures how much choosing the action a_t when in state s_t is better than randomly taking an action according to the current policy π_θ, in terms of discounted reward (i.e., the future discounted reward when using a discounting factor γ is given by $r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + ...$). We can formulate recurrently the value of a state in terms of expected future discounted reward using the Bellman equations [26]:

$$V(s_t|\pi_\theta) = \mathop{\mathbb{E}}_{\substack{a_t \sim \pi_\theta, \\ s_{t+1} \sim P}}[r(s_t, a_t) + \gamma V(s_{t+1})] \tag{9}$$

We can similarly define an action-value function which is the same as $V(s_t|\pi_\theta)$, but for a known action:

$$Q(s_t, a_t|\pi_\theta) = \mathop{\mathbb{E}}_{s_{t+1} \sim P}\left[r(s_t, a_t) + \gamma \mathop{\mathbb{E}}_{a_{t+1} \sim \pi_\theta}[Q(s_{t+1}, a_{t+1})]\right] \tag{10}$$

Then, we can describe the advantage as a subtraction of these two functions

$$\hat{A}_t = Q(s_t, a_t|\pi_\theta) - R(s_t|\pi_\theta) \tag{11}$$

Since the policy update specified by the rule (8) can be large (e.g., for $\pi_{\theta_{old}}(a_t|s_t) \approx 0$) and results in unstable learning, the PPO algorithm clip this function

$$L^C(\theta) = \mathbb{E}\left[min\left(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t\right)\right] \tag{12}$$

with epsilon set to a small value (e.g., around 0.2).

Hyperparameter tuning: We attempted to improve the learning and convergence of the PPO algorithm by tuning the hyperparameters using the Optuna-based [27] stable-baselines3-zoo package. The first attempt with 500 trials of 100,000 iterations failed to improve upon the default hyperparameterization provided with stable-baselines3's implementation of PPO. We tried a second round with 1000 trials of 3,000,000 iterations with similar results. Consequently, we used the default hyperparameterization for PPO that comes with stable-baselines3.

Deep reinforcement learning and network architecture: The PPO algorithm is implemented using deep learning and the architecture illustrated in Figure 5.

Parameter extraction: Once trained, we used this system to extract lognormal parameters for all PQRST complexes from all segments (N=751,510). For the final extraction, we increased the maximum number of iterations to 2000 and the maximum number of iterations with no progress to 200 to maximize our chances of obtaining solutions with high SNR. Beats fitted with an $SNR < 5dB$ were excluded (18.28%), and the remaining beats were averaged in turns within segments and then within recordings. We rejected time points
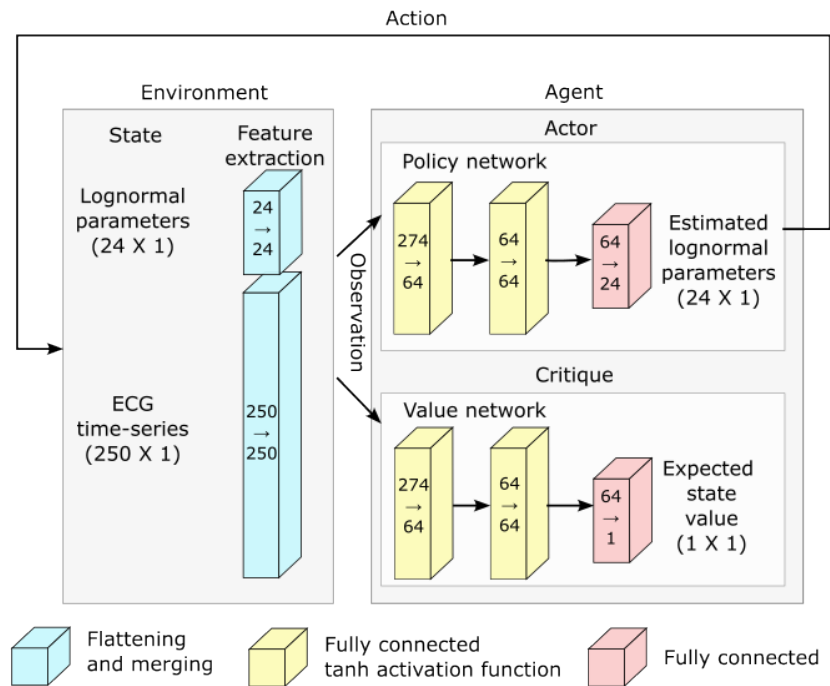
Figure 5. Network architecture used for deep reinforcement learning.

with recordings from less than 8 participants to ensure sufficient sample size for reliable statistical estimations. We used the remaining mean parameter values for statistical analysis (N=121; per age (months): 1: 9, 2: 13, 3: 19, 4: 21, 6: 14, 9: 19, 12: 17, 15: 9). We did not use a hold-out method or cross-validation for this application since our ECG model was fixed and we wanted to train an estimator that would provide the best parameter fitting. Our goal was not to release a generalizing extractor that would perform well on a new dataset, to benchmark classification metrics (e.g., accuracy), or any similar application that would require such controls.

3.6. Parameter denormalisation

Since we estimated lognormal parameters in "normalized time", we transformed them back into normal time before statistical analysis. For a lognormal profile to remain numerically identical after the compression of its time variable t by a factor α such that we have a new time variable $t^* = \alpha t$, we must adjust the original parameters P such that the new values P^* are defined as follows: $\{\mu^*, \sigma^*, t_0^*, D^*\} = \{\mu + log(\alpha), \sigma, \alpha t_0, \alpha D\}$. Therefore, from the parameters P estimated using normalized time t, we computed the non-normalized parameters using $\alpha = (t_3 - t_1)/2$, where t_1 and t_3 are the time of the R peaks for the preceding and following heartbeats, as previously defined (see section Preprocessing).

3.7. Software

All analyses were made in Python. Preprocessing was performed mainly using HeartPy and MNE-Python and the standard Python scientific programming stack (NumPy, SciPy, Matplotlib, Seaborn, Pandas).

4. Results

As a proof of concept for this approach, we investigated if the modeling parameters are sensitive to a factor expected to impact systematically the ECG: age. Figure 6 illustrates the relationship between every modeling parameter and the age of the infant participants. Almost half of the parameters (10/24) show a statistically significant relationship with age with a p-value lower than 0.01 after correction for multiple tests. Further, from the plot,

some of these relationships (e.g., t_0 for the S component or mu for the T_- component) are very clear, with very low p-values and large correlation coefficients.

Using a mixed-effect model with the age as a fixed effect and the subjects as a random effect, we predicted the modification of the PQRST complex as a function of age. Figure 7 compares the prediction and the PQRST time-series averaged across subjects using the mean and the median. As demonstrated by this figure, the effect of age is easier to summarize and understand using the model prediction than direct averaging of time-series. This can reflect the "cleaning" of the data by imposing a linear variation of the parameters, effectively smoothing out natural intra- and inter-subject variability that may not be related to age. It should be kept in mind, however, that whether or not such a linear relationship is appropriate should be further validated. From some of the plots in Figure 6, an exponential relationship may better capture the variation of some of these parameters. This observation is particularly true of the t_0 parameters, which is not surprising as time parameters like reaction times tend to be distributed logarithmically. We can make a similar argument for amplitude parameters (i.e., D).

5. Discussion

In this paper, we aimed to demonstrate how the analysis of ECG can benefit from using a model-based approach. Most processes relevant to fundamental science and practical applications in medicine and neuroscience present us with ill-posed problems due to the complexity of the generative processes, the large quantity of uncontrolled latent variables, and the relative sparsity of experimental measurements. To progress in the study of such systems, we need to constrain our analysis using prior knowledge. We can best exploit such knowledge by embedding it in the structure and parameterization of our models. This process effectively injects our knowledge into our analysis by operationalizing our hypotheses about the generative mechanisms that caused the observed signals.

We plan to further develop this framework along three lines of investigation. First, here we demonstrated this approach using a partially mechanistic model (e.g., lognormals are associated with known electrical processes in the heart) and partially phenomenological (e.g., we used the subtraction of two lognormals to accommodate the reverse asymmetry of the T wave). The full power of such modeling can be reached only when using rich mechanistic models. Phenomenological models can be useful in practical applications. For example, they may offer ways to summarize the data that help improve performances for biomarkers in clinical applications. However, we argue that for fundamental research, a model is as good as its relationship to the underlying mechanisms. By developing models that mechanistically map latent variables to observable signals, we gain ways to make inferences on previously inaccessible variables. Refining such models by properly capturing the various biological mechanisms is an incremental and protracted process. In our case, we have good reasons to assume the lognormality of the different waves of the PQRST complex [17,18]. Nevertheless, further analyzing how the physiology and anatomy of the heart impact the shape and lognormality of these waves would improve the interpretability and conceptual validation of this approach. This investigation will, for example, require a more in-depth analysis of the propagation of polarization/depolarization waves along the atrial and ventricular myocardial cells, and the potential impact of border effects as these waves travel through these spatially bounded structures.

Second, here we used box constraints on the parameter space for our model. Such binary (as opposed to probabilistic) constraints are relatively unrealistic and unrefined. Box constraints enforce the idea that the probability of the parameters taking a given value goes from being null to uniformly distributed for an infinitesimal difference in parameter values. Such a discontinuity is not a fair representation of our prior knowledge or assumptions. Further, box constraints generally do not accurately capture the subspace of plausible or acceptable solutions. The importance of this problem increases with dimensionality. The proportion of space occupied by the "corners" of the box (as opposed to its center) increases with the number of dimensions. These corners represent portions of the subspace where
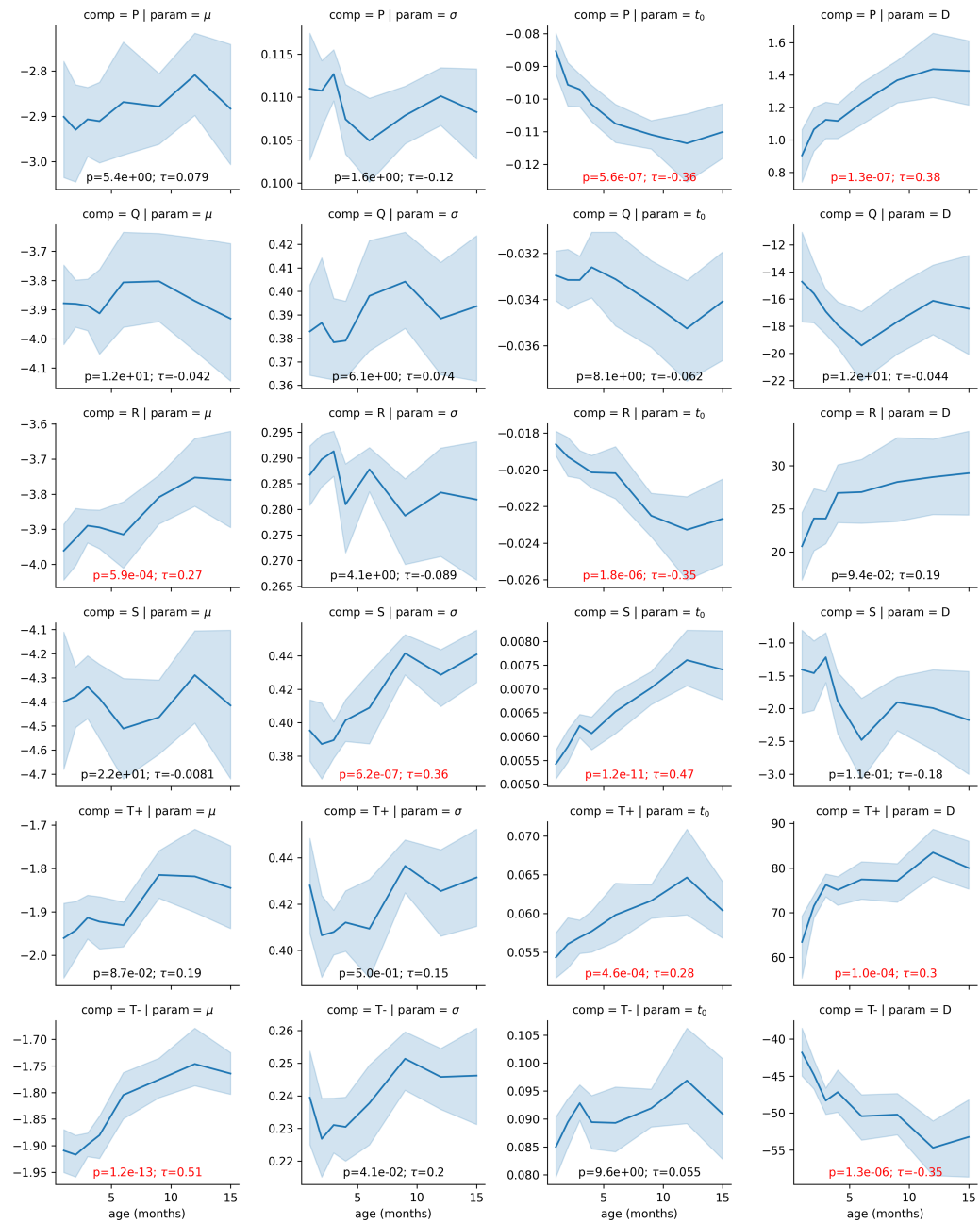
Figure 6. Correlation between the age and the modeling parameters. The light blue region denotes the 95% confidence interval obtained with bootstrapping. Overlaid on each panel is the Kendall rank correlation coefficient (τ) and its corresponding p-value corrected for 24 independent tests using Bonferroni's correction, both in red if significant (i.e., $p < 0.01$). Parameters D are expressed in $\mu V * s$ (i.e., multiplied 1e6), for convenience.
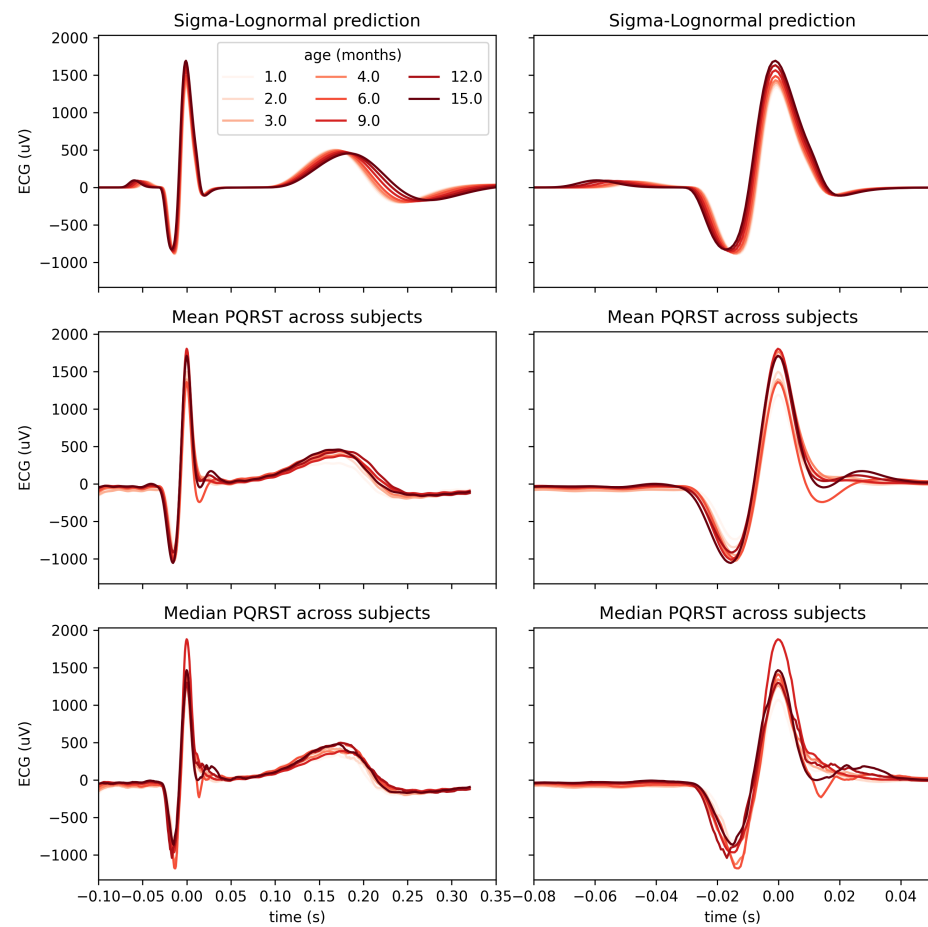
Figure 7. Comparison of PQRST complexes predicted from Sigma-Lognormal modeling (top panel), and mean (middle panel) and median (bottom panel) PQRST across participants, for each time point with at least 8 participants. Both columns show the same data; the right columns only zoom further on the P peak to better appreciate the modulation of its amplitude by age.

multiple parameters have extreme values (i.e., edge cases), and are, therefore, likely to be associated with biologically-implausible solutions. This effect is not negligible in 24 dimensions. For example, whereas a unit-size ball occupies 52% of a unitary cube, a unit-size hyperball occupies 0.00000001% of a unit-size hypercube in a 24-dimensional space. We propose to integrate a Bayesian approach to constrain the estimation of parameters using prior distributions. Aside from being more realistic, this approach for constraining modeling parameters allows more easily incremental integration of new knowledge by adjusting the Bayesian priors. It also integrates more naturally into hierarchical models, where such priors can be generated as an output of another part of the model. Since we do not expect the manifold supporting plausible solutions to be box-shaped, such prior distributions would ideally be a multivariate distribution rather than a series of 24 univariate distributions.

Third, although we consider our approach highly relevant for many applications related to cardiology, we primarily aim to develop this model to support the use of ECG as a window into the state of the autonomic nervous system. Thus, we plan to further model the generation of the heartbeat sequence, as opposed to modeling independent PQRST, as proposed in this paper. The heartbeat is controlled by the balance between two opposing systems: the sympathetic (epinephrine and norepinephrine) and the parasympathetic (acetylcholine). Both systems influence the cardiac process by modulating the opening and closing of calcium channels in myocardial cells. By modeling the impact of these inputs on the heart, the analysis of ECG could give us further insights into autonomic control and

might open the door to clinical applications in conditions where this control is atypical, such as autism spectrum disorder [28].

6. Conclusions

By embedding our knowledge into a forward model of ECG, we demonstrated how a sum of overlapping lognormal can be used to analyze the ECG and the properties of its PQRST complex. We also provided a framework to estimate the values of modeling parameters using deep reinforcement learning and constrained this optimization to preserve parameter interpretability (i.e., parameters are not allowed to take values that would change their qualitative interpretation). Finally, we used the effect of age to test the sensitivity of the modeling parameters to factors systematically affecting ECG. In our discussion, we highlighted some directions to expand on this research. We are confident that as scientists and engineers come to refine their understanding of the complexity of the various natural systems they are tackling, the systematic embedding of our knowledge into mechanistic models and the adoption of a model-driven analytic approach will increasingly reveal itself to be the surest way forward.

Abbreviations

The following abbreviations are used in this manuscript:

ANS	autonomic nervous system
AV	atrioventricular
ECG	electrocardiogram
ODE	Ordinary Differential Equations
PPO	Proximal Policy Optimization
RR	inter-beat interval
SA	sinoatrial node
SNR	signal-to-noise ratio

References

1. Murphy, S.L.; Xu, J.; Kochanek, K.D. Deaths : final data for 2010, 2013.
2. Ahmad, F.B. Provisional Mortality Data — United States, 2021. *MMWR. Morbidity and Mortality Weekly Report* **2022**, *71*. https://doi.org/10.15585/mmwr.mm7117e1.
3. Medicine, M. Anatomy of a Human Heart, 2019.
4. Becker, D.E. Fundamentals of Electrocardiography Interpretation. *Anesthesia Progress* **2006**, *53*, 53–64. https://doi.org/10.2344/0003-3006(2006)53[53:FOEI]2.0.CO;2.

5.   Lake, D.E.; Fairchild, K.D.; Moorman, J.R. Complex signals bioinformatics: evaluation of heart rate characteristics monitoring as a novel risk marker for neonatal sepsis. *Journal of Clinical Monitoring and Computing* **2014**, *28*, 329–339. https://doi.org/10.1007/s10877-013-9530-x.

6.   Fairchild, K.D.; Aschner, J.L. HeRO monitoring to reduce mortality in NICU patients. *Research and Reports in Neonatology* **2012**, *2*, 65–76. Publisher: Dove Press, https://doi.org/10.2147/RRN.S32570.

7.   Quiroz-Juárez, M.A.; Jiménez-Ramírez, O.; Vázquez-Medina, R.; Breña-Medina, V.; Aragón, J.L.; Barrio, R.A. Generation of ECG signals from a reaction-diffusion model spatially discretized. *Scientific Reports* **2019**, *9*, 19000. https://doi.org/10.1038/s41598-019-55448-5.

8.   Cardarilli, G.C.; Di Nunzio, L.; Fazzolari, R.; Re, M.; Silvestri, F. Improvement of the Cardiac Oscillator Based Model for the Simulation of Bundle Branch Blocks. *Applied Sciences* **2019**, *9*, 3653. Number: 18 Publisher: Multidisciplinary Digital Publishing Institute, https://doi.org/10.3390/app9183653.

9.   Bala, M.; B., A. Simulation Of Pathological ECG Signal Using Transform Method. *Procedia Computer Science* **2020**, *171*, 2121–2127. https://doi.org/10.1016/j.procs.2020.04.229.

10. Kubicek, J.; Penhaker, M.; Kahankova, R. Design of a synthetic ECG signal based on the Fourier series. In Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2014, pp. 1881–1885. https://doi.org/10.1109/ICACCI.2014.6968312.

11. Awal, M.A.; Mostafa, S.S.; Ahmad, M.; Alahe, M.A.; Rashid, M.A.; Kouzani, A.Z.; Mahmud, M.A.P. Design and Optimization of ECG Modeling for Generating Different Cardiac Dysrhythmias. *Sensors (Basel, Switzerland)* **2021**, *21*, 1638. https://doi.org/10.3390/s21051638.

12. Plamondon, R.; Marcelli, A.; Ferrer-ballester, M.A. *The Lognormality Principle and its Applications in e-security, e-learning and e-health*; Vol. 88, World Scientific, 2020.

13. Plamondon, R.; Feng, C.; Woch, A. A kinematic theory of rapid human movement. Part IV: a formal mathematical proof and new insights. *Biological Cybernetics* **2003**, *89*, 126–138. https://doi.org/10.1007/s00422-003-0407-9.

14. Plamondon, R. A kinematic theory of rapid human movements: Part III. Kinetic outcomes. *Biological Cybernetics* **1998**, *78*, 133–145. https://doi.org/10.1007/s004220050420.

15. Plamondon, R. A kinematic theory of rapid human movements. Part II. Movement time and control. *Biological Cybernetics* **1995**, *72*, 309–320. https://doi.org/10.1007/BF00202786.

16. Plamondon, R. A kinematic theory of rapid human movements. Part I. Movement representation and generation. *Biological Cybernetics* **1995**, *72*, 295–307. https://doi.org/10.1007/BF00202785.

17. Plamondon, R.; O'Reilly, C.; Rémi, C.; Duval, T. The lognormal handwriter: learning, performing, and declining. *Frontiers in psychology* **2013**, *4*, 945. Publisher: Frontiers.

18. Plamondon, R.; Marcelli, A.; Ferrer, M.A.; (Firm), W.S., Eds. *The lognormality principle and its applications in e-security, e-learning and e-health*; Number v.88 in Series in machine perception and artificial intelligence, World Scientific Publishing Co. Pte. Ltd: Singapore ; Hackensack, NJ, 2021. Meeting Name: International Workshop on the Lognormality Principle and its Applications.

19. van Gent, P.; Farah, H.; van Nes, N.; van Arem, B. HeartPy: A novel heart rate algorithm for the analysis of noisy signals. *Transportation Research Part F: Traffic Psychology and Behaviour* **2019**, *66*, 368–378. https://doi.org/10.1016/j.trf.2019.09.015.

20. O'Reilly, C.; Plamondon, R. A Globally Optimal Estimator for the Delta-Lognormal Modeling of Fast Reaching Movements. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **2012**, *42*, 1428–1442. https://doi.org/10.1109/TSMCB.2012.2192109.

21. O'Reilly, C.; Plamondon, R. Prototype-Based Methodology for the Statistical Analysis of Local Features in Stereotypical Handwriting Tasks. In Proceedings of the 2010 20th International Conference on Pattern Recognition, 2010, pp. 1864–1867. ISSN: 1051-4651, https://doi.org/10.1109/ICPR.2010.460.

22. O'Reilly, C.; Plamondon, R. Development of a Sigma–Lognormal representation for on-line signatures. *Pattern recognition* **2009**, *42*, 3324–3337. Publisher: Pergamon.

23. Djioua, M.; Plamondon, R. A New Algorithm and System for the Characterization of Handwriting Strokes with Delta-Lognormal Parameters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2009**, *31*, 2060–2072. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, https://doi.org/10.1109/TPAMI.2008.264.

24. Ghavamzadeh, M.; Mannor, S.; Pineau, J.; Tamar, A. Bayesian Reinforcement Learning: A Survey. *Found. Trends Mach. Learn.* **2015**, *8*, 359–483. https://doi.org/10.1561/2200000049.

25. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms, 2017. arXiv:1707.06347 [cs], https://doi.org/10.48550/arXiv.1707.06347.

26. Bellman, R. *Dynamic Programming*, 1 ed.; Princeton University Press: Princeton, NJ, USA, 1957.

27. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019.

28. Taylor, E.C.; Livingston, L.A.; Callan, M.J.; Ashwin, C.; Shah, P. Autonomic dysfunction in autism: The roles of anxiety, depression, and stress. *Autism* **2021**, *25*, 744–752, [https://doi.org/10.1177/1362361320985658]. PMID: 33491461, https://doi.org/10.1177/1362361320985658.