# A Data-Driven Gaussian Process Filter for Electrocardiogram Denoising

Mircea Dumitru [1] Qiao Li [1] Erick Andres Perez Alday [1] Ali Bahrami Rad [1] Joel Xue [1,2] Dave Albert [2] Gari D. Clifford [1,3] Reza Sameni [1]

[1]Department of Biomedical Informatics, School of Medicine, Emory University [2]AliveCor Inc. [3]Department of Biomedical Engineering, Georgia Institute of Technology

## Overview

A computationally efficient data-driven Gaussian Process ($\mathcal{GP}$) filter is proposed for ECG filtering, which outperforms state-of-the-art filters in signal-to-noise ratio (SNR) improvement and results in more accurate estimates of clinically important features of the ECG, such as the QT-interval.

## $\mathcal{GP}$-based ECG modeling

- **Model:** the noisy measurements $x(t)$ are modeled as a mixture of the clean ECG $s(t)$ and additive uncorrelated noise $n(t)$, at time instants $t$ for which measurements are available:

$$x(t) = s(t) + n(t), \quad t \in \{t_1 \dots t_N\}$$

- **Assumption:** the noise $n(t)$ and the ECG $s(t)$ are Gaussian processes:

$$s(t) \sim \mathcal{GP}\left(\mu_s(t), \kappa_s\left(t, t'\right)\right)$$
$$n(t) \sim \mathcal{GP}\left(\mu_n(t), \kappa_n\left(t, t'\right)\right)$$ , typically with $\mu_n(t) = 0$ and $\kappa_n\left(t, t'\right) = \sigma_n^2 \delta\left(t, t'\right)$

- **Note:** A $\mathcal{GP}$ is fully described by its mean and covariance matrix. Therefore, the ECG data-model is completely specified by the choice of the mean and covariance matrices of the $\mathcal{GP}$.
- Denoting $\boldsymbol{\mu_s} = [\mu_s(t)]_t$ as the signal mean, $\boldsymbol{K_s} = [\kappa_s\left(t, t'\right)]_{t,t'}$ as the signal covariance matrix (kernel), and $\boldsymbol{K_x}$ as the covariance matrix of the measurements, the optimal maximum a posteriori (MAP) estimator (filter) of the ECG is:

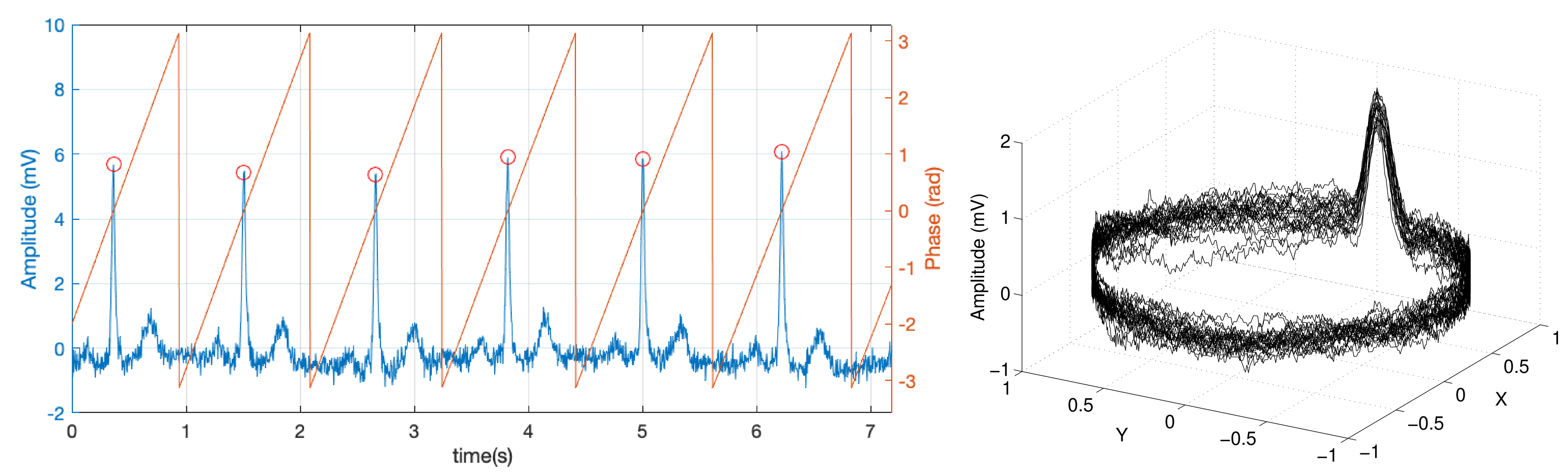
$$\widehat{\boldsymbol{s}} = \boldsymbol{K_s}\boldsymbol{K_x}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_s) + \boldsymbol{\mu}_s$$

## $\mathcal{GP}$-based ECG modeling challenges

- The choice of the covariance matrix $\boldsymbol{K_s}$ is not evident (*analytical kernels proposed in the literature [2], [3], require optimizing over a large number of hyper-parameters*).
- The inversion of $\boldsymbol{K_x}$ required by the filter restricts the use of the model to short ECG signals (the complexity of matrix inversion is $\mathcal{O}(N^3)$).

## A data-driven $\mathcal{GP}$-based filter

- **The proposed $\mathcal{GP}$ scheme:** The kernel is built based on the sample covariance matrix corresponding to the ECG in the *phase domain* [6], via a transformation $\boldsymbol{\Theta} \in \mathbb{R}^{\mathcal{T} \times N}$, which maps each ECG beat to $\mathcal{T}$ bins in the phase domain (regardless of the heart rate).




(a) ECG and ECG phase (b) Phase wrapped ECG

Figure 1. A typical ECG and the ECG phase that maps each ECG cycle to $[-\pi, \pi]$ (cf. [6])

**Algorithm:** 1) detect the R-peaks; 2) transform the ECG to the phase domain; 3) calculate the $\mathcal{GP}$ mean vector and covariance matrix; 4) estimate the noise covariance; 5) apply the filter.

## Data-driven $\mathcal{GP}$ priors selections

- A transformation matrix $\boldsymbol{\Theta}$ maps the ECG to the mean "*binned-heartbeat*" $\boldsymbol{x}_{\text{ph}}$ in the phase domain:

$$\boldsymbol{x}_{\text{ph}} = \boldsymbol{\Theta}\boldsymbol{x} \qquad \left[\mathsf{S}_{s_{\text{ph}}} = \boldsymbol{\Theta}\boldsymbol{K}_s\boldsymbol{\Theta}^T\right].$$

- The transformation is chosen such that $\boldsymbol{\Theta}\boldsymbol{\Theta}^T = \boldsymbol{I}_{\mathcal{T}}$.
- In phase domain, the sample covariance matrices are linked by

$$\mathsf{S}_{s_{\text{ph}}} = \mathsf{S}_{x_{\text{ph}}} - \tilde{\sigma}_n^2 \boldsymbol{I}_{\mathcal{T}}.$$

- The Kernel choice is $\boldsymbol{K}_s = \boldsymbol{\Theta}^T \mathsf{S}_{s_{\text{ph}}} \boldsymbol{\Theta}$ and the mean choice is $\boldsymbol{\mu}_s = \boldsymbol{\Theta}^T \boldsymbol{\mu}_{\text{ph}}$, where $\boldsymbol{\mu}_{\text{ph}}$ is the average ECG beat in the phase domain (both obtained in a data-driven non-parametric approach).
- The inversion is performed via the matrix inversion lemma:

$$\boldsymbol{K}_x^{-1} = \sigma_n^{-2}(\boldsymbol{I}_N - \boldsymbol{\Theta}^T\boldsymbol{\Theta}) + \boldsymbol{\Theta}^T \mathsf{S}_{x_{\text{ph}}}^{-1} \boldsymbol{\Theta}.$$

- The filter is data-driven and requires the inversion of a $\mathcal{T} \times \mathcal{T}$ matrix (typically $\mathcal{T} \ll N$):

$$\widehat{\boldsymbol{s}} = \boldsymbol{\Theta}^T(\boldsymbol{I}_{\mathcal{T}} - \tilde{\sigma}_n^2 \mathsf{S}_{x_{\text{ph}}}^{-1})\boldsymbol{\Theta}(\boldsymbol{x} - \boldsymbol{\mu}_s) + \boldsymbol{\mu}_s.$$

## Implementation

- In the simplified scheme, only the sample variances are taken into account and the filter becomes:

$$\widehat{\boldsymbol{s}} = \boldsymbol{\Theta}^T \mathbf{P} \boldsymbol{\Theta}(\boldsymbol{x} - \boldsymbol{\mu}_s) + \boldsymbol{\mu}_s, \quad \mathbf{P} = \text{diag}\left[\dots\ 1 - \rho_i\ \dots\right], \quad \rho_i = \frac{\tilde{\sigma}_n^2}{\sigma^2_{x_{\text{ph}_i}}} = \frac{\tilde{\sigma}_n^2}{\sigma^2_{s_{\text{ph}_i}} + \tilde{\sigma}_n^2}.$$

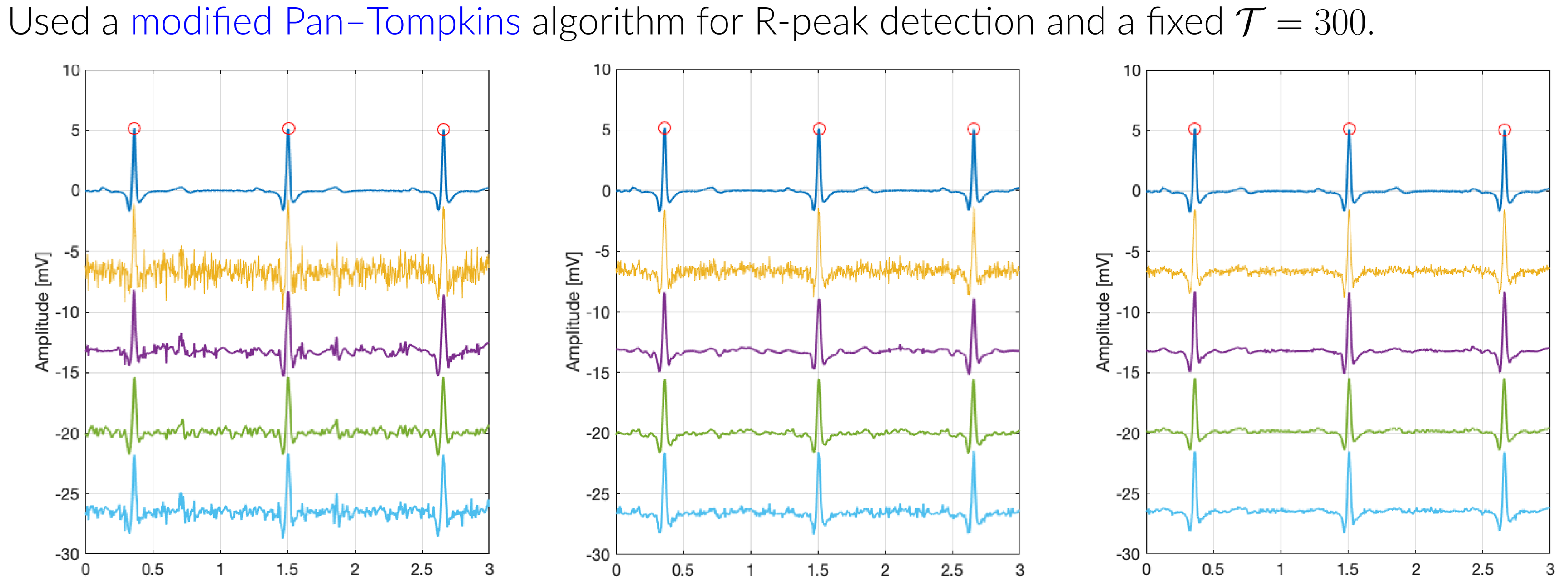
- In accordance with Bayesian filtering terminology, we refer to the above as $\mathcal{GP}$ posterior-based, and the corner case of extremely noisy ECG ($\tilde{\sigma}_n^2 \to \infty$) as $\mathcal{GP}$ prior-based.

## Advantages

- The model's kernel is calculated in a data-driven manner (no *ad hoc* parametric kernels are required).
- As compared with conventional $\mathcal{GP}$, the computational cost is significantly reduced.
- The model requires only *two* hyper-parameters: 1) the number of phase-domain bins $\mathcal{T}$ (which can be fixed); 2) the noise covariance (which is estimated via maximum evidence).

## Results

- **Dataset:** The PhysioNet QT Database [5]. **Benchmark:** a wavelet denoiser, which outperformed other non-model-based filtering schemes in previous research [4]
- Used a modified Pan–Tompkins algorithm for R-peak detection and a fixed $\mathcal{T} = 300$.




(a) SNR = 0dB (b) SNR = 6dB (c) SNR = 12dB

Figure 2. An ECG from the PhysioNet QT Database at different SNRs denoised by the proposed filters and a wavelet denoiser [4]. From top to bottom: the original ECG, noisy, wavelet, $\mathcal{GP}$ prior-based and $\mathcal{GP}$ posterior-based

## SNR improvement results

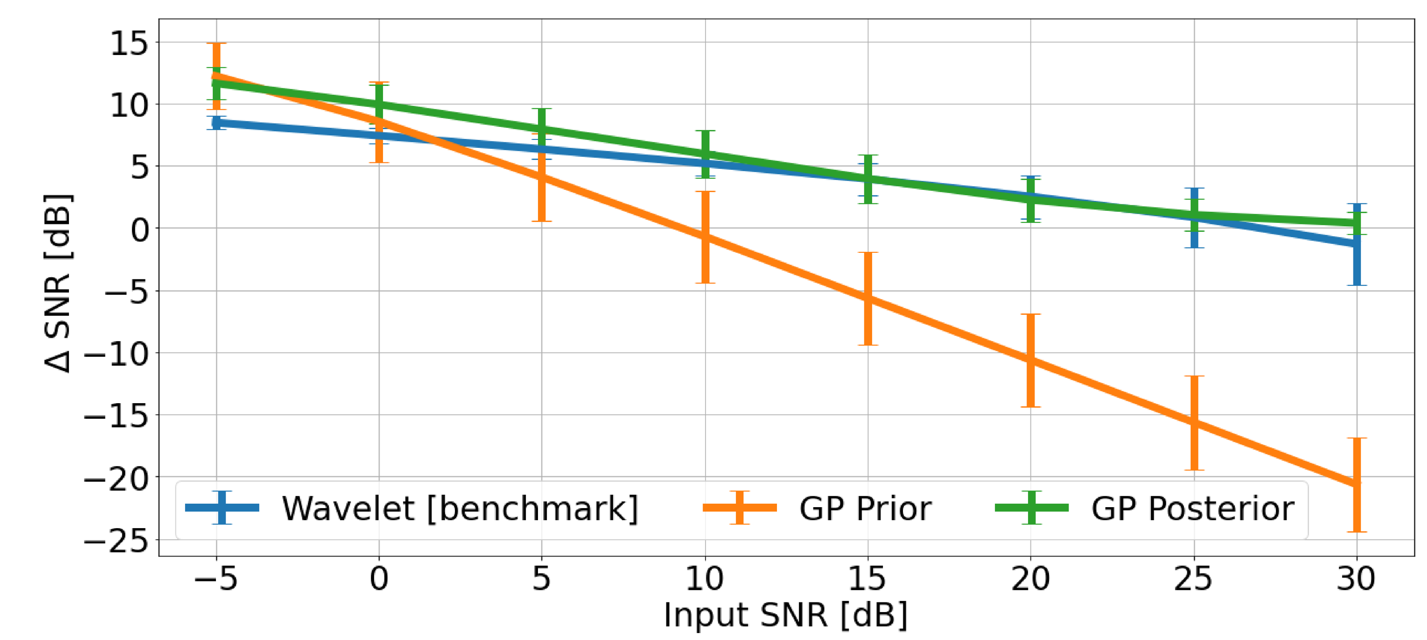



Figure 3. SNR improvement using the proposed $\mathcal{GP}$ filter and the benchmark wavelet denoiser [4]. The curves correspond to the average SNR improvement across all samples of the PhysioNet QT Database, in leads I and II (whenever available) and with 5 repetitions using different random noise instances per record.

## QT estimation performance

- The proposed filter also preserves clinically important features of the ECG, such as the QT-interval.
- We use the algorithms detailed in [1] to compare the QT-intervals estimated from clean ECG and the ones estimated from the denoised ECG using the proposed and benchmark wavelet denoisers.

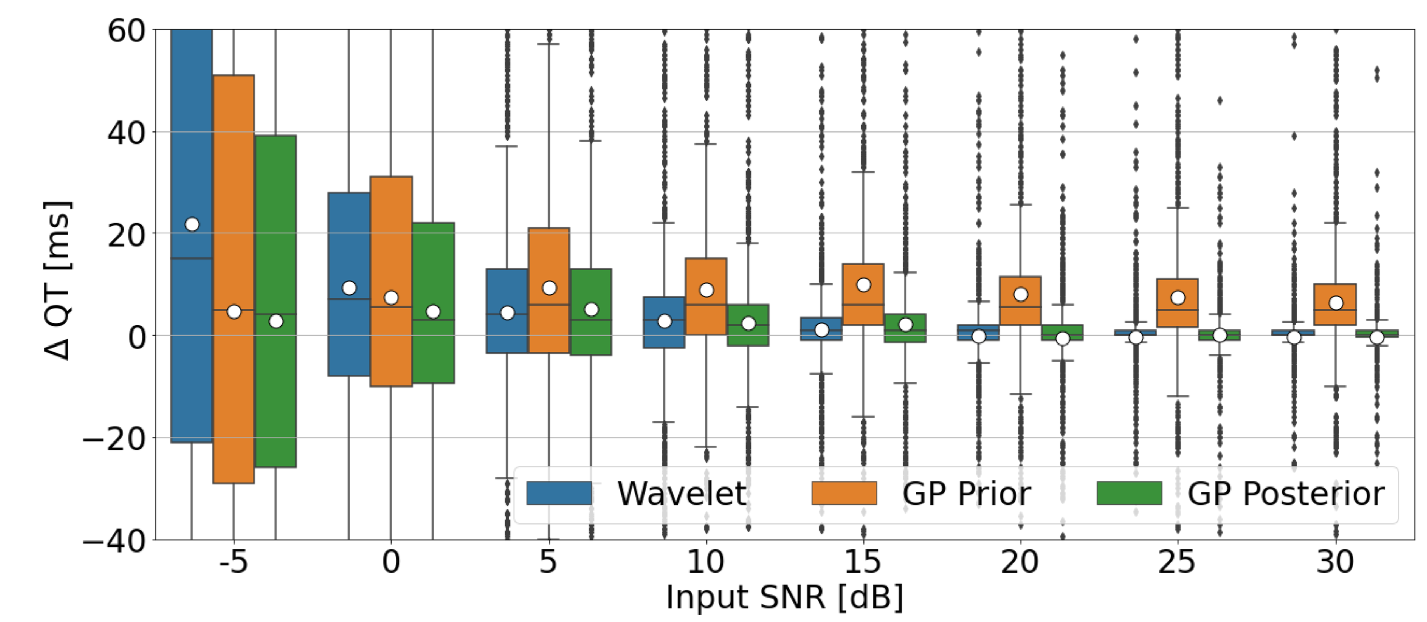



Figure 4. The distribution of the QT-interval estimation differences between the outputs of the proposed and benchmark algorithms minus the QT-interval estimated from the clean ECG (at different SNR). The $\mathcal{GP}$ posterior-based filter outperforms the prior-based and the state-of-the-art wavelet denoiser, in terms of estimation error bias and variance.

## Conclusion

- Compared with the state-of-the-art non-model-based ECG denoisers, the proposed data-driven $\mathcal{GP}$ posterior-based filter has superior performance in terms of SNR improvement and QT-interval parameter estimation accuracy.

## References

[1] Q. Li, M. Dumitru, and E.A. Perez Alday, et al. QT-Interval Estimation Improved with Fusion of Multiple Automated Algorithms. In *ISCE*, 2022.

[2] M. Niknazar, B. Rivet, and C. Jutten. Fetal ECG extraction from a single sensor by a non-parametric modeling. In *EUSIPCO*, 2012.

[3] B. Rivet, M. Niknazar, and C. Jutten. Non parametric modelling of ECG: Applications to denoising and single sensor fetal ECG extraction. In *LVA/ICA 2012*, 2012.

[4] R. Sameni. Online filtering using piecewise smoothness priors: Application to normal and abnormal electrocardiogram denoising. *Signal Processing*, 133, 2017.

[5] R. Sameni. *The Open-Source Electrophysiological Toolbox (OSET), version 3.14*, 2018. URL `https://github.com/alphanumericslab/OSET`.

[6] R. Sameni, C. Jutten, and M. B. Shamsollahi. Multichannel Electrocardiogram Decomposition using Periodic Component Analysis. *IEEE TBME*, 55(8):1935–1940, 2008.

**Funding/Disclosures:** This study was funded by unrestricted funding from Alivecor, the National Institute of Biomedical Imaging and Bioengineering (NIBIB) under NIH grant R01EB030362, and the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR002378.

*save a copy!*

 mircea.dumitru@dbmi.emory.edu