Joint Morphology and Baseline Estimation for the 12-Lead ECG via an Interpretable Cost Function

Lukas Tenbrink[1,2], Konstantinos D Rizas[1], Axel Loewe[2]

[1] Department of Cardiology, LMU Klinikum, Munich, Germany
[2] Institute of Biomedical Engineering, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Abstract

Traditional ECG baseline removal faces a frequency trade-off: aggressive cutoffs destroy T wave morphology because the T wave and baseline overlap in frequency, while conservative cutoffs leave contaminating drift. We propose a per-lead template fitting algorithm that jointly estimates beat morphology and the baseline by gradient descent on a cost function encoding simple ECG invariants. The ECG recording is modeled as a smoothly-drifting baseline plus beats temporally localized around each R peak, with the residual treated as noise. On 200 LUDB records, the fit removed baseline wander with a $+10.0$ *dB SNR gain at 0 dB input while keeping T wave morphology stable (Pearson* $\geq$ 0.95*), and reproduced simulated ground-truth morphology at Pearson 0.98 on 90 MedalCare-XL records across four pathology classes. Morphology and baseline are exposed as separate signals: repolarization analysis can use T wave shape that the preprocessing has not deformed, and baseline dynamics remain available in their own right.*

1. Introduction

Accurate T wave morphology is central to clinical biomarkers concerned with repolarization analysis. The QT interval underlies Long QT Syndrome diagnosis [1] and the ICH E14 regulatory regime for drug-safety screening [2]. The ST segment separates STEMI from NSTEMI treatment pathways [3]. Periodic Repolarization Dynamics (PRD), an established predictor of sudden cardiac death and all-cause mortality [4, 5], depends on the shape of the T wave in a beat-to-beat vector representation over recordings of at least 30 minutes. Traditional high-pass filters used for baseline removal face a challenge: aggressive cutoffs destroy T wave morphology because the T wave and baseline drift overlap in frequency, while conservative cutoffs leave residual drift that deforms the T wave shape. Better preprocessing techniques without this trade-off are needed for downstream analysis.

Two families of preprocessors dominate the field. Rule-based cascades such as Laguna's analytical delineator [6] and ECGDeli [7] remove baseline and noise and then annotate in sequence. We argue that each step in such a cascade solves a harder problem than it needs to, because the later steps have not yet produced their outputs: baseline removal is easier when the beats have been subtracted, and beat subtraction is easier when the baseline is already known. This is especially relevant when dealing with real-world data, which is heterogeneous: any rule-based algorithm can fall apart if the data does not adhere to its rules, explicit or implicit. Therefore, the more rules an algorithm has, the more it is limited to data similar to the data it was developed on. Typical end-to-end neural delineators [8] fuse the steps but at the cost of interpretability: effective modern models often use millions of parameters, which are hard both to understand and to verify conclusively. To our knowledge, neither family routinely checks that its outputs preserve the inter-lead algebraic identities — Einthoven's law $\mathrm{II} = \mathrm{I} + \mathrm{III}$ and the Goldberger augmented-lead equations — that vectorcardiographic analysis and multi-lead fusion depend on.

We propose a per-lead template fitting algorithm that jointly estimates beat morphology and the baseline by gradient descent on a cost function encoding a few uncontentious facts about the ECG. Because every term is a soft constraint weighed against every other, the fit finds a single decomposition that respects them all at once rather than chaining them sequentially. The model has roughly 5,000 trainable parameters — three orders of magnitude smaller than typical neural delineators. The decomposition preserves Einthoven's law on LUDB records even though inter-lead dependence is not exploited during fitting, and exposes morphology and baseline as separate outputs, so that each can be investigated independently.

2. Method

The recording is modeled as beats plus baseline: a template placed at each detected R peak, and a smoothly drifting signal underneath. Both are estimated jointly by

gradient descent; what neither explains is left as residual noise. Fitting is a per-record optimization: every parameter comes from that one recording, with no training set and nothing learned across records. The inputs are a single unfiltered 10-second 12-lead ECG record at 500 Hz, R peak indices from any upstream detector, and the beat clusters described below. The two estimates are the model's explicit outputs, both with the same length and lead count as the input: a beat morphology signal `sig_morph` and a baseline signal `baselines`, one per lead; their sum is the reconstruction of the recorded signal. Beats are clustered by DBSCAN on band-pass filtered (4–45 Hz) QRS-centered windows, with one template learned per cluster per lead and shared by every beat in that cluster. Clean 10-second recordings typically reduce to one or two near-identical clusters. Initial templates are the mean of extracted beat windows over each cluster, decomposed into a Daubechies-5 wavelet detail vector via PyWavelets [9]; the baseline for each lead is initialized to zero at a coarser wavelet scale. Both detail vectors are trainable parameters reconstructed to the time domain through an inverse-DWT layer [10]. The wavelet basis parameterizes the unknowns rather than decomposing the recorded signal: it imposes a multiresolution prior — coarse scales for the baseline, finer scales for the template — at a fraction of the parameters a per-sample representation would need. The model is lead-agnostic — 12-lead batching is a convenience — and has ≈5,000 trainable parameters per single-cluster 12-lead fit (≈420/lead, scaling linearly), with each additional cluster adding ≈3,000 more. Figure 1 shows the architecture.

Loss function. The loss function has six terms, with fixed weights chosen during development of the method and detailed in the following.

Reconstruction (1 term): Mean squared error between `sig_morph + baselines` and the raw signal.

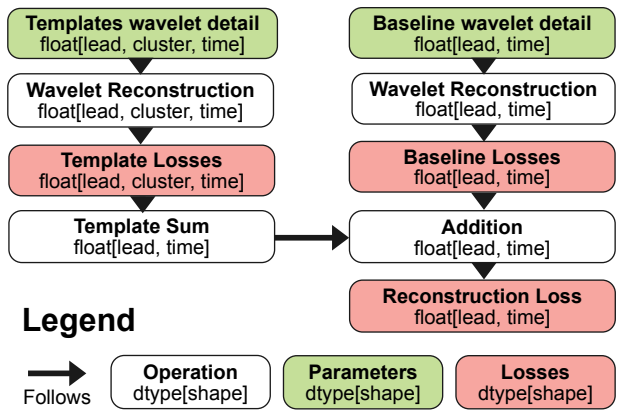

Figure 1. Model architecture: trainable wavelet detail vectors feed through an inverse-DWT layer into a full-record prediction, compared to the raw recording under the six cost terms, grouped into the three loss boxes shown.

Beat template regularization (4 terms): Two L1 decay penalties push the template toward zero before the P wave (-0.25 s after R) and after the T wave ($+0.65$ s); two L2-across-leads shrinkage penalties act on the pre- and post-QRS regions. These anchor the template at the isoelectric line: without them, regular templates can absorb baseline signal.

Baseline regularization (1 term): In 1-second sliding windows, we penalize the standard deviation of the baseline's sample-to-sample difference. This gives the baseline inertia: it can drift at any rate, but changing that rate is expensive, so the optimizer prefers baselines that do not develop beat-synchronous oscillations. The term discourages high-frequency baseline but does not forbid it, unlike traditional 0–2 Hz high-pass filters which must be bounded to avoid corrupting T wave morphology.

Optimization. We optimize the template and baseline detail vectors jointly with Adam at learning rate 0.025 for 300 steps. A single 10-second record fits in 4–5 minutes of wall-clock time on a single laptop CPU. Hyperparameters are fixed across all experiments and datasets, with no per-record tuning. The source code and evaluation scripts are open source.

Evaluation was performed using 3 datasets. **LUDB** [11]: 200 × 10 s × 12-lead × 500 Hz, with expert P/QRS/T onset/peak/offset annotations per lead. We used all 200 records for the cross-lead consistency experiment and records 1–20 for the noise robustness experiment. QRS peak indices came from the lead-II expert annotations. **MIT-BIH NSTDB** [12]: baseline wander (`bw`) and electrode motion (`em`) noise records at 360 Hz. We resampled to 500 Hz before superposition. **MedalCare-XL** [13]: simulated 12-lead ECG records from openCARP [14] with ground-truth morphology by construction. We used 30 records from the sinus test subset plus 20 records each from `avblock`, `lbbb`, and `rbbb` subsets for the fidelity check.

3. Results

Cross-lead consistency check. As a consistency check, we verified that the fit does not violate the inter-lead algebraic identities. We computed the RMS of the Einthoven residual II − I − III on the raw signal and fitted morphology for each of the 200 LUDB records (Figure 2). The fitted morphology reduced the residual on 199 of 200 records; the one exception had a raw residual already below 0.002 mV. The Goldberger augmented-lead identities showed the same pattern.

Noise robustness. We added NSTDB baseline wander and electrode motion at target input SNRs of 0, 6, and 12 dB to 20 LUDB records ($N = 120$ conditions total) and re-fitted.

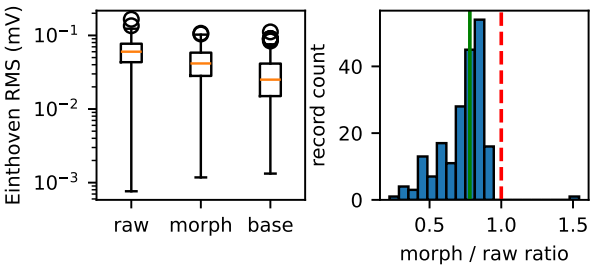

Figure 2. Einthoven residual RMS (mV) on 200 LUDB records. Left: boxplot for raw, fitted morphology, and fitted baseline. Right: morph/raw ratio histogram; red dashed = no change, green = median.

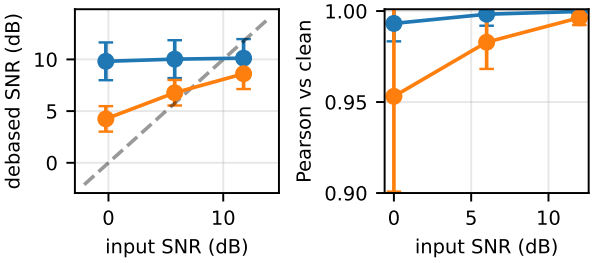

Figure 3. Left: output SNR after fitted-baseline subtraction vs. input SNR (dashed line $y = x$). Right: morphology-output Pearson vs. the clean-input fit on the same record. Blue = baseline wander (`bw`), orange = electrode motion (`em`).

Denoising via baseline subtraction. Subtracting the fitted baseline from the noisy input preserved the original beat content and removed the low-frequency contamination (Figure 3). On `bw` at 0 dB input, this lifted output SNR from -0.23 dB to 9.82 dB ($\Delta = +10.04$ dB); at 6 and 12 dB input the output stabilized near $\sim$10 dB. On `em` at 0 dB, $\Delta = +4.47$ dB. A one-tailed sign test at 0 dB yielded $p < 10^{-6}$ for both noise types.

Morphology stability. Pearson correlation between each noisy-input morphology and the clean-input morphology from the cross-lead-consistency fits was 0.993–0.9998 for `bw` and 0.953–0.996 for `em` across the three input SNRs (nearly noise-invariant).

Fidelity on MedalCare-XL. Real ECG records have no known true morphology. We therefore fit to 90 simulated records from MedalCare-XL (30 sinus rhythm plus 20 each of AV block, left bundle branch block, and right bundle branch block) and compared the fitted reconstruction (`sig_morph + baselines`) to the ground-truth simulated signal. Pearson correlation was 0.98 ± 0.01 pooled across all 90 records, with per-class means of 0.979 (sinus), 0.996 (AVB), 0.976 (LBBB), and 0.985 (RBBB).

Table 1. ecgpuwave fiducial localization error and matched beat counts vs. LUDB expert annotations, on 200 records in four preprocessing modes. Bold marks the best value in each RMS row.

Fiducial	raw	debased	reconst.	morph.
QRS peak RMS (ms)	5.94	5.72	4.70	**4.46**
QRS matches	1814	1812	1718	1624
P peak RMS (ms)	**6.17**	6.42	6.78	6.50
P matches	1358	1356	1016	955
T peak RMS (ms)	10.02	10.56	**9.64**	9.96
T matches	1234	1218	1450	1449

Simulated ECGs are smoother than real recordings, so this is a complementary check; it confirmed the fit reproduced morphology where the ground truth is known by construction.

Downstream fiducial recovery. We ran ecgpuwave, the WFDB/PhysioNet ecosystem's [15, 16] classical P/QRS/T wave delineator, on four variants of each of the 200 LUDB records and compared the detected fiducial points to LUDB expert annotations on lead II (Table 1).

Baseline subtraction alone (`debased = raw − baselines`) was approximately neutral on clean LUDB: its value was visible under superimposed NSTDB noise (above). Using the fitted morphology (`morph = sig_morph`) reduced mean QRS peak RMS by 25% and recovered more T wave detections (1449 vs. 1234), at the cost of slightly worse P peak localization because the P-timing boundary penalty pulled individual P positions toward the cluster average. Median QRS RMS dropped from 1.22 ms (raw) to 0 ms in both fitted modes, indicating that on most records ecgpuwave matched expert fiducials within a single 2 ms sample. The mean T peak RMS improvement was small (4% for reconst, essentially flat for morph) and the median degraded, as cluster averaging blurred the T peak position by the same mechanism that affected P localization. All inter-mode differences in Table 1 fall within the 11–13 ms inter-reader uncertainty of point-based annotation [17].

Residual spectrum. One template per cluster implicitly averages beat-to-beat morphology variation. The power spectral density of the residual `raw − sig_morph − baselines`, averaged across all 200 LUDB records, was only 7.2%, 3.6%, and 4.3% of raw energy in the beat-coherent bands (0.5–4, 4–15, 15–50 Hz) and 55.1% in the 50–125 Hz noise band. The fit preserved beat-coherent content at clinically relevant frequencies.

4. Discussion

The Einthoven consistency check confirms that the per-lead fit does not introduce cross-lead artifacts despite operating independently on each lead. The baseline inertia regularization avoids the frequency trade-off of classical

filters: the fit removes baseline wander without deforming T wave morphology. This supports shape-based biomarkers such as PRD, affected by T wave distortion. Separately, baseline wander itself can carry clinically relevant information, and keeping morphology and baseline as distinct outputs opens possibilities for biomarkers and risk scores that use both signals independently.

Limitations and future work. The single-template-per-cluster assumption absorbs beat-to-beat morphology differences into the residual; the residual spectrum analysis verifies this does not discard beat-coherent content at clinically relevant frequencies, but applications investigating respiration, T wave alternans, or atrial rhythm irregularities need a richer decomposition. We plan cost-model extensions including respiration as a coupled modulation term and TWA as an alternating-template variant. The per-lead independence is by design, as this allowed us to test the assumption that our model does not destroy clinically relevant signal through a violation of invariants. In future work, we plan to exploit signal correlation using the vectorcardiographic framework.

5. Conclusion

Joint-cost optimization with terms rooted in ECG physiology is a viable framework for ECG preprocessing: it removes baseline wander without deforming T wave morphology, reproduces ground-truth morphology, and preserves the inter-lead algebraic identities of the 12-lead system without enforcing this algorithmically. It is compatible with rule-based downstream annotators, transparent where end-to-end neural delineators are opaque, and extensible to further cost-model variants for bespoke analysis or better generalizability.

References

[1] Priori SG, Blomström-Lundqvist C, Mazzanti A, Blom N, Borggrefe M, Camm J, et al. 2015 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death. European Heart Journal 2015;36(41):2793–2867.

[2] ICH E14 Clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs - Scientific guideline. Technical report, ICH.

[3] Ibanez B, James S, Agewall S, Antunes MJ, Bucciarelli-Ducci C, Bueno H, et al. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation. European Heart Journal 2018;39(2):119–177.

[4] Rizas KD, Nieminen T, Barthel P, Zürn CS, Kähönen M, Viik J, et al. Sympathetic activity–associated periodic repolarization dynamics predict mortality following myocardial infarction. Journal of Clinical Investigation 2014; 124(4):1770–1780.

[5] Bauer A, Klemm M, Rizas KD, Hamm W, Von Stülpnagel L, Dommasch M, et al. Prediction of mortality benefit based on periodic repolarisation dynamics in patients undergoing prophylactic implantation of a defibrillator. The Lancet 2019;394(10206):1344–1351.

[6] Laguna P, Jané R, Caminal P. Automatic Detection of Wave Boundaries in Multilead ECG Signals: Validation with the CSE Database. Computers and Biomedical Research 1994; 27(1):45–60.

[7] Pilia N, Nagel C, Lenis G, Becker S, Dössel O, Loewe A. ECGdeli - An open source ECG delineation toolbox for MATLAB. SoftwareX 2021;13:100639.

[8] Jimenez-Perez G, Alcaine A, Camara O. Delineation of the electrocardiogram with a mixed-quality-annotations dataset using convolutional neural networks. Scientific Reports 2021;11(1):863.

[9] Lee GR, Gommers R, Waselewski F, Wohlfahrt K, O'Leary A. Pywavelets: A python package for wavelet analysis. Journal of Open Source Software 2019;4(36):1237. URL `https://doi.org/10.21105/joss.01237`.

[10] Leiderman T, Ben Ezra Y. Information bottleneck driven deep video compression—ibopendvcw. Entropy 2024; 26(10).

[11] Kalyakulina AI, Yusipov II, Moskalenko VA, Nikolskiy AV, Kosonogov KA, Osipov GV, et al. LUDB: A New Open-Access Validation Tool for Electrocardiogram Delineation Algorithms. IEEE Access 2020;8:186181–186190.

[12] Moody GB, Muldrow WE, Mark RG. The MIT-BIH Noise Stress Test Database, 1992.

[13] Gillette K, Gsell MAF, Nagel C, Bender J, Winkler B, Williams SE, et al. MedalCare-XL: 16,900 healthy and pathological 12 lead ECGs obtained through electrophysiological simulations, 2022.

[14] Plank G, Loewe A, Neic A, Augustin C, Huang YL, Gsell MA, et al. The openCARP simulation environment for cardiac electrophysiology. Computer Methods and Programs in Biomedicine 2021;208:106223.

[15] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000; 101(23):e215–e220.

[16] Xie C, McCullum L, Johnson A, Pollard T, Gow B, Moody B. Waveform Database Software Package (WFDB) for Python. PhysioNet 2023;Version 4.1.0.

[17] Panicker GK, Karnad DR, Natekar M, Kothari S, Narula D, Lokhandwala Y. Intra- and interreader variability in QT interval measurement by tangent and threshold methods in a central electrocardiogram laboratory. Journal of Electrocardiology 2009;42(4):348–352.

Address for correspondence:

Lukas Tenbrink
Department of Cardiology, LMU Klinikum
Marchioninistr. 15, 81377 Munich, Germany
lukas.tenbrink@gmail.com