Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

BLiSS: A Beat-Local, Signal-Quality-Informed Kalman Smoother for ECG Denoising under Non-Stationary Muscle Artifact, with Ectopic-Beat Reconstruction Assessment

Khanh Duy Tran

khanh1234tg@gmail.com

Vietnam National University Ho Chi Minh City https://orcid.org/0009-0006-5761-2683

Research Article

Keywords: Electrocardiogram denoising, Muscle artifact, Kalman smoother, Signal quality assessment, Ectopic beats

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11595235/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

BLiSS: A Beat-Local, Signal-Quality-Informed Kalman Smoother for ECG Denoising under Non-Stationary Muscle Artifact, with Ectopic-Beat Reconstruction Assessment

Khanh Duy Tran [1*]

[1*]School of Electrical Engineering, International University – Vietnam National University, Ho Chi Minh City, Vietnam.

Corresponding author(s). E-mail(s): khanh1234tg@gmail.com;

Abstract

ECG denoising under non-stationary muscle artifact (MA) must suppress noise while preserving ectopic morphology. We propose BLiSS, a beat-local, signal-quality-informed state-space smoothing algorithm. Following fixed wavelet-Wiener filtering and R-peak detection, raw-beat quality sets the measurement variance of a constant-velocity Kalman filter and Rauch–Tung–Striebel smoother, without neighbouring-beat morphology priors. BLiSS was compared with three Kalman-based and two EMG-focused algorithms, including adaptive wavelet-Wiener filtering with adaptive moving-average filtering (AWWF + AMAF), on non-paced MIT-BIH Arrhythmia recordings under constant-level, step, burst, and alternating MA. Targeted event-overlapping bursts assessed premature ventricular contraction (PVC) and supraventricular ectopic beat (SVEB) reconstruction. BLiSS achieved the largest average SNR gain among the evaluated methods for each contamination profile (4.44–5.38 dB), with the largest margins over AWWF + AMAF under step and burst MA. It also achieved the lowest mean PVC QRS-region normalized root-mean-square error (NRMSE; 0.217), PVC amplitude mean absolute error (0.194 mV), and SVEB QRS-region NRMSE (0.241). AWWF + AMAF achieved lower SVEB whole-beat NRMSE (0.501 versus 0.596) and lower P-peak timing error in a small atrial subgroup. Ablations supported beat-specific measurement weighting and backward smoothing; a preceding-beat prior increased PVC error. BLiSS therefore achieved the highest mean denoising gain and the lowest mean ectopic QRS-region errors, while whole-beat and P-wave preservation remained limitations.

Keywords: Electrocardiogram denoising, Muscle artifact, Kalman smoother, Signal quality assessment, Ectopic beats

1 Introduction

Electrocardiography (ECG) is widely used for the non-invasive assessment of cardiac electrical activity and the detection of arrhythmias [1]. However, ECG recordings are susceptible to baseline wander, electrode-motion artifacts, power-line interference, and electromyographic (EMG) contamination, hereafter referred to as muscle artifact [2, 3]. During ambulatory monitoring, skeletal-muscle activity can introduce transient, non-stationary contamination, while changes at the skin–electrode interface can produce distinct electrode-motion artifacts [3]. Such contamination can obscure waveform boundaries and compromise the reliability of ECG interval measurements and subsequent interpretation [1, 4]. This work focuses on muscle artifact, whose spectral overlap with cardiac activity makes noise suppression difficult without altering the underlying waveform [3, 5].

Bayesian state-space methods provide a framework for combining a model of signal dynamics with a model of measurement uncertainty. Under linear Gaussian assumptions, Kalman filtering and Rauch–Tung–Striebel (RTS) smoothing yield minimum mean-square error estimates under the specified model [6, 7], [8, Chaps. 4 and 8]. For ECG processing, McSharry et al. introduced a dynamical model based on a limit cycle and Gaussian functions representing characteristic waves [9]. Sameni et al. subsequently developed a nonlinear Bayesian denoising framework using extended Kalman filtering and smoothing and unscented Kalman filtering [2, 10]. Later developments included wave-based models for normal and ectopic beats [11], adaptive Kalman filtering [12], an expectation-maximization-based adaptive filter bank [13], hierarchical filtering with online learned patient-specific evolution priors [14], adaptive cubature filtering and smoothing [15], and dual augmented extended Kalman filtering with autoregressive noise models [16].

Nevertheless, abrupt morphological changes challenge models that rely on the dominant or expected beat shape. Sameni et al. discussed difficulties filtering premature ventricular contractions (PVCs) at low SNR, when both the phase-based model and the measurements may be unreliable [2]. Sayadi et al. introduced state variables for the P, QRS and T components and demonstrated improved tracking of spontaneous PVC morphology over the earlier two-state formulation [11]. These developments highlight the importance of accommodating ectopic morphology during denoising. Evaluations using real muscle artifact commonly scale noise excerpts to prescribed segment-level SNRs [11, 13, 15, 16]. Although these excerpts contain natural fluctuations and studies illustrate ectopic-beat preservation, segment-level performance summaries do not separately quantify reconstruction errors on individual ectopic beats. This study therefore complements continuous-record evaluation with targeted PVC and supraventricular ectopic beat (SVEB) reconstruction assessment, addressing ectopic morphologies relevant to arrhythmia analysis [17, 18].

Signal quality assessment offers another means of adapting processing to contamination. Li et al. combined ECG and arterial blood pressure information using signal quality indices (SQIs) and Kalman filtering for robust heart-rate estimation [19]. Smital et al. estimated a continuous SNR curve and distinguished signals suitable for full-waveform analysis, QRS detection only, or neither task [20]. Liu et al. developed rhythm-specific quality assessment to reduce false supraventricular and ventricular ectopic classifications in noisy ECGs [21]. Other studies examined artifact classification [22] and noise-dependent filter selection [23]. However, statistical SQIs can depend on the dataset and threshold selection [24]. Furthermore, quality assessment for accepting segments or judging classification reliability serves a different purpose from continuously controlling denoising strength within an individual beat.

Methods specifically targeting muscle artifact have also incorporated arrhythmic recordings. Li et al. combined adaptive wavelet-Wiener and moving-average filtering and demonstrated noise suppression on normal, arrhythmic, ventricular flutter, and exercise ECGs, including a PVC-containing example contaminated by real EMG noise [5]. Atanasoski et al. proposed an iterative regeneration method (IRM) and evaluated denoising using SNR improvement, waveform correlation, and amplitude correspondence at selected fiducial points [3]. These studies support evaluating noise suppression alongside waveform preservation, but do not provide separate quantitative reconstruction analyses for PVC and SVEB cohorts under controlled event-overlapping contamination. Such analyses are useful because aggregate recording-level performance can conceal errors affecting less frequent beat types. Moreover, premature atrial and ventricular beats differ in their regional morphology, motivating complementary whole-beat and QRS-region assessments [18].

To investigate these issues, we propose BLiSS, a beat-local, signal-quality-informed state-space smoothing algorithm for ECG enhancement under muscle artifact. A fixed front end based on wavelet-Wiener filtering [25] produces an initial estimate, on which an R-peak detector identifies beat locations. Within each beat window, the high-frequency power fraction and beat power computed from the raw ECG determine a quality-informed measurement variance, R_k. A constant-velocity state model is then applied using forward Kalman filtering and RTS smoothing [6, 7]. Its process-noise intensity, q, is estimated once at the beginning of each processing run by maximum likelihood, and the reconstructed beat windows are stitched into a continuous output. This formulation avoids imposing a preceding-beat morphology prior while allowing measurement uncertainty to vary between beats.

We evaluate BLiSS on all 44 non-paced recordings of the MIT-BIH Arrhythmia Database [26–28], using muscle artifact from the Noise Stress Test Database [29]. Constant-level, step, burst, and alternating contamination profiles assess continuous denoising performance, while targeted event-overlapping bursts assess PVC and SVEB reconstruction using regional and whole-beat waveform errors and PVC QRS amplitude error. To our knowledge, previous evaluations of Kalman-based and EMG-focused ECG denoisers have not jointly provided separate quantitative reconstruction assessments for PVCs and SVEBs under controlled non-stationary muscle artifact. This study addresses that gap.

The contributions of this work are as follows:

1. BLiSS combines wavelet-Wiener filtering with beat-local, quality-informed state-space smoothing without imposing a preceding-beat morphology prior. Ablations assess the effects of the Kalman state model and backward smoothing, beat-specific measurement weighting, process-noise estimation, and the preceding-beat prior; a timing sensitivity analysis examines the effect of R-peak detection.
2. A controlled benchmark across all 44 non-paced MIT-BIH Arrhythmia recordings compares BLiSS with Kalman-based [13, 14, 16] and EMG-focused [3, 5] methods under constant-level, step, burst, and alternating muscle artifact, using shared noise realizations and patient-clustered statistics.
3. A targeted evaluation quantifies PVC and SVEB waveform reconstruction and PVC amplitude preservation under event-overlapping muscle artifact, revealing class-specific filtering behavior that aggregate recording-level metrics may obscure.

2 Materials and Methods

2.1 Datasets and signal preparation

MIT-BIH Arrhythmia Database.

The MIT-BIH Arrhythmia Database contains 48 approximately 30-min, two-channel ambulatory ECG recordings from 47 subjects, digitized at 360 Hz with 11-bit resolution over a 10-mV range and accompanied by expert beat annotations [26, 28]. Four paced recordings (102, 104, 107, and 217) were excluded [17], leaving 44 recordings from 43 patients. The MLII lead was used, with signals expressed in mV and resampled to 250 Hz by polyphase filtering; annotation indices were adjusted accordingly. Following AAMI beat groupings [30], the normal class comprised N, L, R, e, and j, and the SVEB class comprised A, a, J, and S. PVCs (V) and ventricular escape beats (E) were treated separately.

MIT-BIH Noise Stress Test Database.

MA was obtained from the two-channel `ma` recording of the MIT-BIH Noise Stress Test Database (NSTDB) [29, 31]. Both channels were resampled to 250 Hz, and injected noise excerpts were drawn exclusively from the final 20% of each channel, approximately 6 min per channel.

Waveform timing references.

The QT Database (QTDB), comprising 105 fifteen-minute ECG excerpts sampled at 250 Hz with expert waveform-boundary annotations [32], supported configuration selection and validation of the wavelet delineator [4] used for secondary timing outcomes. Excluding the 13 QTDB excerpts overlapping the study cohort did not change the selected configuration. Expert P-wave annotations [33], available for 12 cohort recordings (100, 101, 103, 106, 117, 119, 122, 207, 214, 222, 223, and 231), provided reference P-peak positions for the atrial SVEB subgroup.

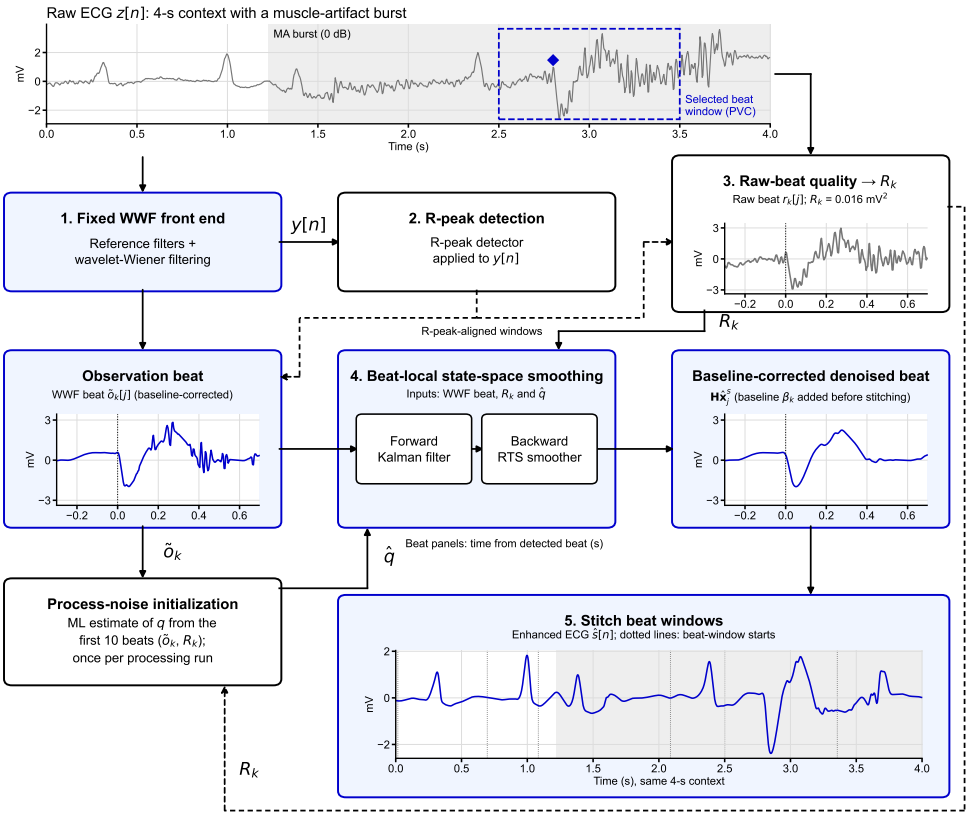


Fig. 1 BLiSS framework illustrated with a 4-s ECG excerpt containing a muscle-artifact burst at 0-dB nominal input SNR. R-peak-aligned windows, $[\tau_k - 0.30,\ \tau_k + 0.70)$ s, select corresponding WWF and raw beats. Raw-beat quality determines R_k; Kalman filtering and RTS smoothing reconstruct each beat before stitching. Grey shading marks the burst; the blue dashed box marks the selected PVC window.

2.2 BLiSS: beat-local, signal-quality-informed state-space smoother

BLiSS combines five modules (Fig. 1): fixed wavelet-Wiener filtering (WWF) filtering, R-peak detection, raw-beat quality assessment, beat-local smoothing, and continuous reconstruction. The WWF output supplies observation beats, while raw-beat quality determines R_k.

1. Fixed WWF front end.

The raw ECG, $z[n]$, first undergoes reference filtering to attenuate baseline wander and mains interference:

$$x = h_{\mathrm{N}} * h_{\mathrm{HP}} * z, \tag{1}$$

where h_{HP} is a linear-phase FIR high-pass filter (order 1000; cut-off 0.67 Hz), and h_{N} is a linear-phase FIR band-stop filter (order 1000; 59.2–60.8 Hz). The group delays are compensated.

A two-stage WWF uses four-level stationary wavelet transforms [25]: db4 coefficients with non-negative garrote thresholding produce a pilot estimate, and sym4 coefficients with pilot-derived Wiener gains produce $y[n]$.

At each detail level, the local noise scale is estimated as

$$\hat{\sigma}_m[n] = \frac{\mathrm{median}_{|i-n|\le 2f_s}\,|d_m[i]|}{0.6745}, \tag{2}$$

using a centred window of approximately 4 s. The first-stage threshold is $\lambda_m[n] = \hat{\sigma}_m[n]\sqrt{2\ln N}$, where N is the processed signal length. In the second stage, the detail coefficients of x are multiplied by

$$G_m[n] = \frac{u_m^2[n]}{u_m^2[n] + \hat{\sigma}_m^2[n]}, \tag{3}$$

where $u_m[n]$ denotes the corresponding pilot coefficient. Approximation coefficients are retained. The wavelet families, decomposition depth, and thresholding rule remain fixed; local noise estimates and Wiener gains depend on the signal.

2. R-peak detection and observation-beat extraction.

The R-peak detector (XQRS, PhysioNet WFDB Python) is applied to $y[n]$ with default settings [34]. Each detection τ_k defines a 250-sample window extending from 0.30 s before to 0.70 s after the detection:

$$\mathcal{B}_k = \{\tau_k - 75, \ldots, \tau_k + 174\}. \tag{4}$$

Windows extending beyond the recording are discarded.

The WWF observation beat is $o_k[j] = y[\tau_k - 75 + j]$, $j = 0, \ldots, 249$. A straight-line baseline is estimated from two fixed anchor segments, $\mathcal{I}_1 = [0, 20)$ and $\mathcal{I}_2 = [175, 195)$:

$$\beta_k[j] = \bar{o}_{k,1} + (\bar{o}_{k,2} - \bar{o}_{k,1})\frac{j - t_1}{t_2 - t_1}, \qquad \tilde{o}_k[j] = o_k[j] - \beta_k[j], \tag{5}$$

where $\bar{o}_{k,i}$ is the segment mean, with prescribed anchor locations $t_1 = 10$ and $t_2 = 185$. The baseline-corrected beat, $\tilde{o}_k$, is the input to the smoother.

3. Raw-beat quality and measurement variance.

The same detected window is extracted from the raw ECG and mean-centred: $r_k[j] = z[\tau_k - 75 + j] - \bar{z}_k$. Quality assessment therefore uses the contamination present before WWF filtering.

Beat power and high-frequency power fraction are computed as

$$P_k = \frac{1}{L}\sum_{j=0}^{L-1} r_k^2[j], \qquad h_k = \frac{\sum_{40\le f<125\,\mathrm{Hz}} S_k(f)}{\sum_{1\le f<125\,\mathrm{Hz}} S_k(f)}, \tag{6}$$

where $L = 250$ and $S_k(f)$ is the Hann-windowed periodogram. These quantities set one measurement variance for the entire beat:

$$R_k = R_0 + \kappa h_k P_k, \qquad R_0 = 1.5 \times 10^{-5}\ \mathrm{mV}^2, \qquad \kappa = \frac{39.5}{85}. \tag{7}$$

The bandwidth ratio $\kappa = 39.5/85 \approx 0.465$ maps the high-frequency proxy to 0.5–40 Hz under a locally white-noise approximation. Coloured MA and cardiac high-frequency activity make R_k a weighting proxy rather than a calibrated residual-noise estimate. It is recomputed independently for each beat.

4. Beat-local state-space smoothing.

The baseline-corrected observation beat is processed by a forward Kalman filter followed by a backward RTS smoother. A constant-velocity state model describes waveform amplitude and its change per sample [8, Chap. 4]:

$$\begin{aligned} \mathbf{x}_j &= \mathbf{F}\mathbf{x}_{j-1} + \mathbf{w}_j, & \mathbf{F} &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, & \mathbf{w}_j &\sim \mathcal{N}(\mathbf{0}, q\bar{\mathbf{Q}}), \\ \tilde{o}_k[j] &= \mathbf{H}\mathbf{x}_j + v_j, & \mathbf{H} &= \begin{bmatrix} 1 & 0 \end{bmatrix}, & v_j &\sim \mathcal{N}(0, R_k), \end{aligned} \tag{8}$$

with $\mathbf{x}_j = [x_j, \dot{x}_j]^\top$ and $\bar{\mathbf{Q}} = \begin{bmatrix} 1/3 & 1/2 \\ 1/2 & 1 \end{bmatrix}$.

The supporting initialization module in Fig. 1 estimates q from the first $K_0 = 10$ observation beats and their corresponding R_k values. It maximizes the forward-filter innovation likelihood [8, Chap. 12]:

$$\hat{q} = \arg\max_{q \in \mathcal{G}} \left\{ -\frac{1}{2} \sum_{k=1}^{K_0} \sum_{j=2}^{L-1} \left[\ln(2\pi S_{k,j}(q)) + \frac{\nu_{k,j}^2(q)}{S_{k,j}(q)} \right] \right\}, \tag{9}$$

where $\nu_{k,j}$ and $S_{k,j}$ are the innovation and its variance, and $\mathcal{G} = \{10^{-9+9i/36} : i = 0, \ldots, 36\}$ mV^2. The estimate $\hat{q}$ remains fixed for the run, while R_k varies between beats.

Each beat is initialized independently:

$$\hat{\mathbf{x}}_0^- = \begin{bmatrix} \tilde{o}_k[0] \\ \tilde{o}_k[1] - \tilde{o}_k[0] \end{bmatrix}, \qquad \mathbf{P}_0^- = 100\ \mathrm{diag}\left(\mathrm{Var}(\tilde{o}_k) + R_k,\ \mathrm{Var}(\Delta\tilde{o}_k) + 2R_k\right). \tag{10}$$

No state or morphology prior is transferred from the preceding beat.

The forward Kalman filter predicts the state and updates it using the current observation [6]:

$$\hat{\mathbf{x}}_j^- = \mathbf{F}\hat{\mathbf{x}}_{j-1}, \qquad \mathbf{P}_j^- = \mathbf{F}\mathbf{P}_{j-1}\mathbf{F}^\top + \hat{q}\bar{\mathbf{Q}}, \tag{11}$$

$$\nu_{k,j} = \tilde{o}_k[j] - \mathbf{H}\hat{\mathbf{x}}_j^-, \qquad S_{k,j} = \mathbf{H}\mathbf{P}_j^-\mathbf{H}^\top + R_k, \tag{12}$$

$$\mathbf{K}_j = \mathbf{P}_j^-\mathbf{H}^\top S_{k,j}^{-1}, \qquad \hat{\mathbf{x}}_j = \hat{\mathbf{x}}_j^- + \mathbf{K}_j\nu_{k,j},$$

$$\mathbf{P}_j = (\mathbf{I} - \mathbf{K}_j\mathbf{H})\mathbf{P}_j^-. \tag{13}$$

The initial prior is used directly at $j = 0$; prediction begins at $j = 1$.

The RTS pass then refines the filtered states using later observations within the same beat [7]:

$$\mathbf{C}_j = \mathbf{P}_j\mathbf{F}^\top(\mathbf{P}_{j+1}^-)^{-1}, \tag{14}$$

$$\hat{\mathbf{x}}_j^s = \hat{\mathbf{x}}_j + \mathbf{C}_j(\hat{\mathbf{x}}_{j+1}^s - \hat{\mathbf{x}}_{j+1}^-), \tag{15}$$

for $j = L-2, \ldots, 0$, starting from $\hat{\mathbf{x}}_{L-1}^s = \hat{\mathbf{x}}_{L-1}$. The baseline-corrected denoised beat shown in Fig. 1 is $\mathbf{H}\hat{\mathbf{x}}_j^s$.

5. Continuous reconstruction.

The removed baseline is restored before stitching:

$$\hat{s}_k[j] = \mathbf{H}\hat{\mathbf{x}}_j^s + \beta_k[j]. \tag{16}$$

In overlapping windows, the beat whose window starts later supplies the output samples:

$$\hat{s}[n] = \hat{s}_k[n - \tau_k + 75], \qquad n \in [\tau_k - 75,\, \min(\tau_k + 175, \tau_{k+1} - 75))\,. \tag{17}$$

The final beat contributes to the end of its valid window. Samples outside all valid windows retain the WWF output, $y[n]$.

Process-noise estimation requires 370 forward passes (37 grid values $\times$ 10 beats) once per run. Each beat is then filtered and smoothed using 2×2 state matrices. Centred front-end estimation and backward smoothing require future samples, making the pipeline non-causal.

2.3 Evaluation protocols and outcome measures

All methods processed identical inputs, $z = s_{\text{rec}} + \epsilon$, where s_{rec} is the original ECG recording and ϵ is added muscle artifact (MA). The reference was $s = \mathcal{H}\{s_{\text{rec}}\}$, using the filters in (1). Outputs underwent the same filtering before scoring. The reference recording was not assumed intrinsically noise-free.

Each mean-centred MA excerpt was scaled to a nominal input SNR of L dB:

$$\epsilon = \epsilon_0\sqrt{\frac{P_s}{P_{\epsilon_0}\,10^{L/10}}}, \qquad P_s = \frac{1}{|\Omega_s|}\sum_{n\in\Omega_s} s^2[n], \qquad P_{\epsilon_0} = \frac{1}{|\Omega_\epsilon|}\sum_{n\in\Omega_\epsilon}\big(\mathcal{H}\{\epsilon_0\}[n]\big)^2, \tag{18}$$

where Ω_s covers the processed record, window, or event context, and Ω_ϵ covers the noise excerpt. Excerpt positions and NSTDB channels were selected randomly using fixed seeds. Each processing unit received one noise realization, shared by all methods.

2.3.1 Continuous-record evaluation

Four schedules tested different patterns of MA exposure:

- **T1, constant:** each complete recording was evaluated at all 21 input SNRs from −5 to 15 dB in 1-dB steps.
- **T2, step:** three non-overlapping 10-min windows per recording were divided into segments with uniformly sampled durations of 20–90 s. Each segment received a randomly selected level from the same 21-level grid, with abrupt transitions.
- **T3, burst:** the same windows contained a 15-dB MA background interrupted by bursts at −5 to 10 dB in 1-dB steps. Burst durations were log-uniform over 1–60 s; the initial delay and inter-burst gaps were uniform over 30–60 s and 10–60 s, respectively. Bursts cross-faded with the background over 0.1 s.
- **T4, alternating:** after 5 min without added noise, MA alternated between 2-min on and off intervals, with 0.1-s fades. Each recording was evaluated at all 21 levels; off-intervals contained no added noise.

Each T2 segment, T3 burst, and T4 on-interval used a separately sampled noise excerpt. Scoring excluded the first 30 s and final 2 s of each recording or window. Table 1 summarized overall denoising across each contamination schedule, including the T3 background but excluding T4 off-intervals. In contrast, T3 burst-only scores assessed denoising during burst exposure (Fig. 2, Table 4, and block B of Table 5).

Level-specific scores excluded fades. T1/T4 used one score per recording and level. For T2/T3, samples at each level were pooled across each patient's windows because random level assignments produced variable, sometimes brief exposures (Fig. 2, Table 5).

2.3.2 Targeted ectopic-beat evaluation

Eligible ectopic beats had eight preceding and two following normal-class beats and occurred at least 60 s after the recording start and 20 s before its end. Up to 30 events per recording and class were randomly selected using a fixed seed, yielding 451 PVCs from 27 patients and 201 SVEBs from 21 patients (182 A and 19 a).

Each method processed an 80-s context, $[\tau_k - 60, \tau_k + 20)$ s, allowing initialization before the event. The primary exposure was a 1.2-s MA burst, $[\tau_k - 0.40, \tau_k + 0.80)$ s, with 0.1-s fades, evaluated at all 21 SNR levels. Scaling used the reference power of the full context, so the individual beat's measured SNR could differ from the nominal level. A matched normal beat from the same recording received the same noise exposure.

Reconstruction was assessed within a fixed QRS region, $\mathcal{Q}_k = [\tau_k - 0.06, \tau_k + 0.06)$ s, and a whole-beat region, $\mathcal{B}_k = [\tau_k - 0.30, \tau_k + 0.70)$ s. Both regions were anchored to the reference beat annotation, independently of detected R-peak timing. P-peak timing was assessed separately in 32 atrial SVEBs with expert P-wave annotations from five patients (records 100, 101, 103, 222, and 223).

2.3.3 Comparators and ablations

BLiSS was compared with the adaptive Kalman filter bank (AKFB) [13], hierarchical Kalman filtering (HKF) [14], adaptive dual augmented extended Kalman filtering

with a third-order autoregressive noise model (ADAEKF-AR3) [16], the iterative regeneration method (IRM) [3], and adaptive wavelet-Wiener filtering with adaptive moving-average filtering (AWWF + AMAF) [5]. HKF was ported from the authors' public code; AKFB, ADAEKF-AR3, IRM, and AWWF + AMAF were reimplemented from the published descriptions and parameter settings because code was unavailable or proprietary. Methods requiring beat locations used the same R-peak detections obtained from the WWF output.

Kalman-module ablations compared WWF alone (K0), random walk + RTS (K1), forward-only constant velocity (K2), constant velocity + RTS (K3, BLiSS), and constant acceleration + RTS (K4), with the front end, beat timing, R_k, stitching, noise realizations, and scoring held fixed. These variants used the targeted 1.2-s bursts rather than T3; ΔSNR was computed within the whole-beat windows of ectopic events and matched normal beats (Fig. 7).

Additional ablations replaced R_k with its first-10-beat mean, either retaining BLiSS's $\hat{q}$ or re-estimating it; used moment-based q; or added a preceding-beat morphology prior. All configurations shared noise realizations. Annotated R peaks provided a timing sensitivity analysis using reference information unavailable during normal operation.

2.3.4 Outcome measures

Continuous denoising was assessed by SNR improvement:

$$\mathrm{SNR}(u;\Omega) = 10\log_{10}\frac{\sum_{n\in\Omega} s^2[n]}{\sum_{n\in\Omega}(u[n]-s[n])^2}, \qquad \Delta\mathrm{SNR} = \mathrm{SNR}(\hat{s};\Omega) - \mathrm{SNR}(\tilde{z};\Omega), \tag{19}$$

where $\tilde{z} = \mathcal{H}\{z\}$, $\hat{s}$ is the reference-filtered method output, and Ω is the scoring interval.

Waveform reconstruction error was measured by

$$\mathrm{NRMSE}_{\mathcal{R}} = \frac{\|\hat{s}_{\mathcal{R}} - s_{\mathcal{R}}\|_2}{\|s_{\mathcal{R}}\|_2}, \qquad \mathcal{R} \in \{\mathcal{Q}_k, \mathcal{B}_k\}. \tag{20}$$

PVC QRS amplitude error was

$$A(u) = \max_{n\in\mathcal{Q}_k} |u[n] - \mathrm{median}_{m\in\mathcal{P}_k}\, u[m]|\,, \qquad \varepsilon_A = |A(\hat{s}) - A(s)|, \tag{21}$$

where $\mathcal{P}_k = [\tau_k - 0.30, \tau_k - 0.25)$ s. The largest absolute deflection, $A(u)$, accommodates upright and negative-dominant QRS complexes without waveform delineation. The mean of ε_A is reported as amplitude MAE.

The primary endpoints were continuous-record ΔSNR, following established denoising evaluations [2, 5, 13], PVC QRS-region NRMSE, and SVEB whole-beat NRMSE. The waveform endpoints complement noise suppression by assessing reconstruction fidelity [1, 3]. Secondary endpoints were PVC QRS amplitude MAE and SVEB QRS-region NRMSE. P-peak timing and PVC QRS-duration MAE were exploratory.

Timing measurements used the QTDB-selected wavelet delineator [4, 32] described in Section 2.1. For the atrial subgroup, P-peak error was

$$\varepsilon_{\mathrm{P}} = |\hat{t}_{\mathrm{P}} - t_{\mathrm{P}}|, \tag{22}$$

where $\hat{t}_{\mathrm{P}}$ is the delineated peak and t_{P} the expert annotation [33]. Peaks within 150 ms of the reference were considered located. Errors were averaged within each event over noise levels with located peaks and reported alongside the located fraction. Thirty of the 32 eligible events contributed timing measurements, with coverage varying by method and level. Reference-ECG delineation quantified baseline measurement error. These descriptive subgroup results were not generalized to all 201 SVEBs.

PVC QRS duration was measured with the same delineator on $\hat{s}$ and s. Let $\hat{t}$ and t denote the respective intervals from QRS onset to offset:

$$\varepsilon_D = |\hat{t} - t|. \tag{23}$$

Errors were averaged within each event across levels with valid measurements on both signals. Reference durations were delineator-derived rather than expert-annotated, and wide-QRS delineation was not independently validated.

2.3.5 Statistical analysis

All tables report means with 95% confidence intervals (CIs) from a patient-clustered bootstrap with 2000 resamples. Patients were sampled with replacement, retaining all analysis units of each sampled patient; records 201 and 202 formed one cluster. The statistic was recomputed for each resample, and its 2.5th and 97.5th percentiles defined the CI.

Analysis units were recordings for Table 1 (averaged across 21 levels for T1/T4 or three windows for T2/T3) and events for Tables 2–3 (averaged across levels; timing outcomes used valid measurements only). Table 4 used recordings for T2, windows for T3 with burst samples pooled within each window, and events for targeted outcomes. In Table 5, blocks A–B used the patient-level T2/T3 scores defined above, and blocks C–D used event scores at each level. Blocks were summarized separately.

Point estimates averaged analysis units, giving greater weight to patients with more units, except in patient-level T2/T3 summaries. Bootstrap draws were shared across methods with the same patient set. Bold marks the best observed mean; comparisons were descriptive, and neither rankings nor CI overlap constituted a formal test.

3 Results and Discussion

3.1 Continuous-record denoising

BLiSS had the highest observed mean ΔSNR under all four MA schedules (Table 1): 4.44, 5.38, 4.68, and 4.55 dB for constant, step, burst, and alternating MA, respectively. AWWF + AMAF followed closely (4.36–5.04 dB), then IRM (3.60–4.07 dB), whereas AKFB and HKF remained below 1 dB. The margin over AWWF + AMAF was largest for step and burst MA, the schedules with abrupt changes in noise level.

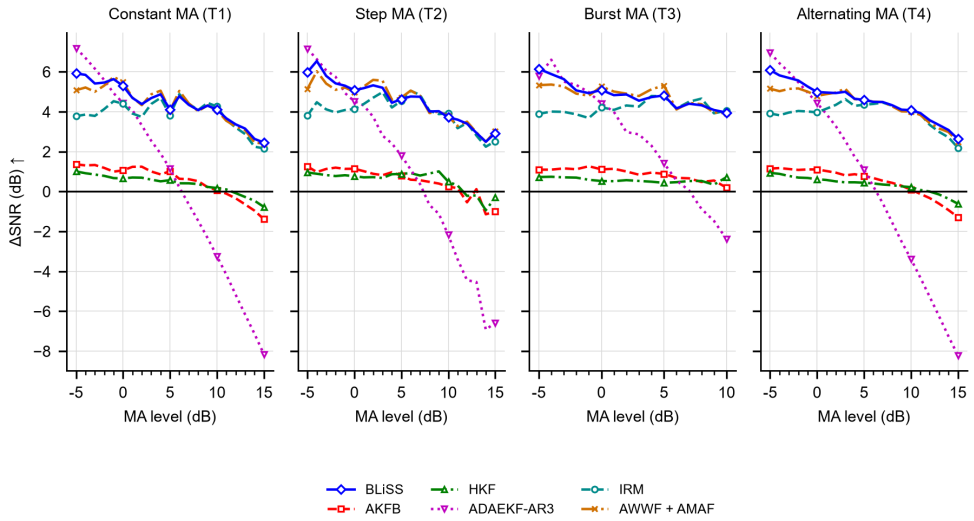


Fig. 2 Continuous-record ΔSNR by MA level for constant (T1), step (T2), burst (T3; burst intervals only) and alternating (T4; noisy intervals) schedules. Curves show point estimates over 1-dB steps; uncertainty is reported in Tables 1 and 5.

Across input SNRs (Fig. 2), BLiSS gained most at strong contamination. ADAEKF-AR3 achieved the largest improvement at −5 dB in step MA (7.15 dB; Table 5), but its performance declined sharply with increasing SNR, reaching approximately −8 dB at 15 dB in the constant and alternating schedules. This distortion of mildly contaminated signals explains its low level-averaged ΔSNR (Table 1). From 0 dB upward, AWWF + AMAF or IRM had the highest mean at most representative T2/T3 levels, with BLiSS within 0.5 dB of the best mean (blocks A–B of Table 5).

Fig. 3 illustrates the mechanism in a burst-MA window of record 100. R_k rose from approximately 10^{-4} to 10^{-2} mV^2 within the first beats of each burst and returned to its background level after the burst ended, so smoothing strength followed the beat-level contamination. During the bursts, the 1-s error of BLiSS was comparable to that of IRM and AWWF + AMAF.

Thus, level-averaged performance depended on maintaining gains across input SNRs rather than on peak gain under the strongest contamination: BLiSS, IRM, and AWWF + AMAF remained beneficial throughout, whereas the low-SNR advantage of ADAEKF-AR3 came at the cost of distorting mildly contaminated signals (Fig. 2, Table 5).

3.2 Ectopic-beat reconstruction

Fig. 4 shows an inverted PVC (record 109) under a −5 dB burst. BLiSS reconstructed the negative QRS with a QRS-region NRMSE of 0.61, second to IRM (0.51) and lower than AWWF + AMAF (0.66), ADAEKF-AR3 (0.73), HKF (0.86), and AKFB (0.87). IRM and ADAEKF-AR3 showed spurious high-amplitude deflections

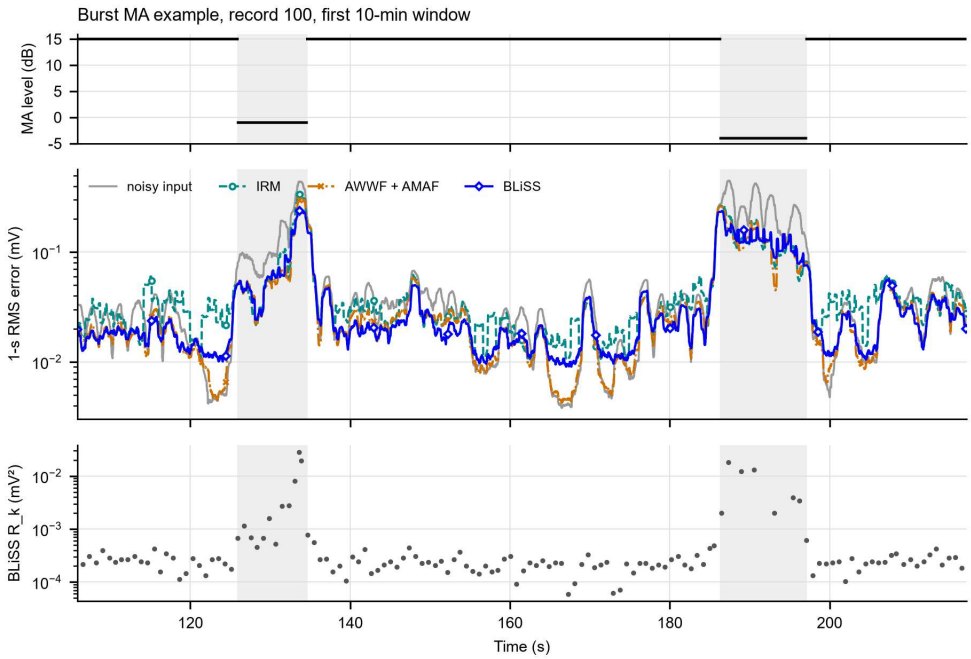


Fig. 3 Burst-MA example (record 100, first 10-min window): MA schedule, 1-s RMS error of the noisy input, IRM, AWWF + AMAF and BLiSS, and the BLiSS measurement variance R_k; bursts shaded.

Table 1 Continuous-record MA denoising across 44 recordings from 43 patients. Values are means [95% CIs].

Method	Constant MA ΔSNR (dB) ↑	Step MA ΔSNR (dB) ↑	Burst MA ΔSNR (dB) ↑	Alternating MA ΔSNR (dB) ↑
BLiSS (proposed)	**4.44** [4.18, 4.68]	**5.38** [5.13, 5.63]	**4.68** [4.41, 4.92]	**4.55** [4.32, 4.77]
AKFB	0.54 [0.37, 0.68]	0.98 [0.90, 1.06]	0.70 [0.54, 0.85]	0.47 [0.33, 0.60]
HKF	0.39 [0.11, 0.59]	0.74 [0.57, 0.88]	0.40 [0.19, 0.56]	0.36 [0.10, 0.56]
ADAEKF-AR3	0.46 [-0.19, 1.13]	3.81 [3.16, 4.44]	0.34 [-0.45, 1.14]	0.37 [-0.28, 1.01]
IRM	3.87 [3.69, 4.05]	4.07 [3.94, 4.20]	3.60 [3.42, 3.77]	3.90 [3.76, 4.04]
AWWF + AMAF	4.36 [4.15, 4.57]	5.04 [4.85, 5.23]	4.39 [4.16, 4.63]	4.37 [4.21, 4.52]

in the post-PVC ST–T region, while IRM, ADAEKF-AR3 and AWWF + AMAF also showed spurious deflections near the P wave of the following beat. Across all 451 PVCs

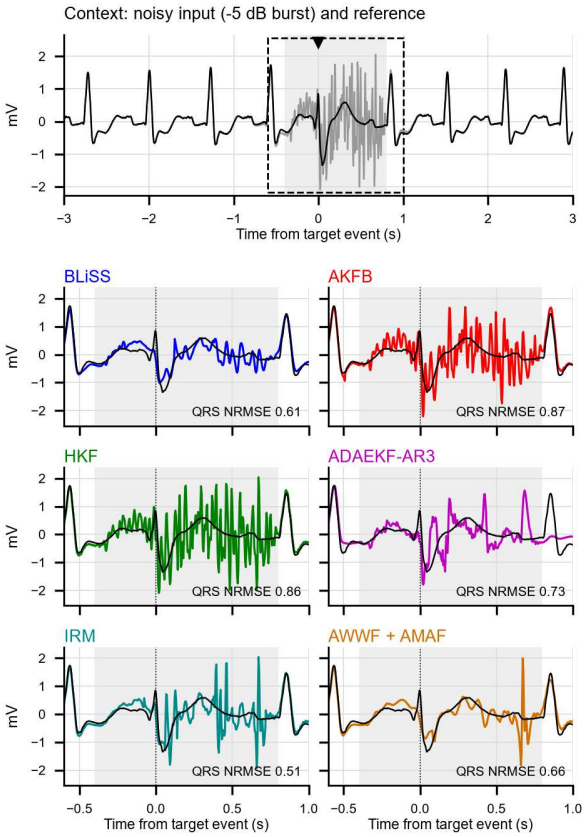


Fig. 4 PVC example (record 109, inverted QRS) under a 1.2-s, −5 dB on-event MA burst. Top: noisy input and reference over ±3 s; bottom: each method (colour) with the reference (black) over the zoom interval. Annotation: QRS-region NRMSE.

(Table 2), BLiSS had the lowest mean QRS-region NRMSE (0.217) and QRS amplitude MAE (0.194 mV), followed by AWWF + AMAF (0.231; 0.204 mV) and IRM (0.242; 0.237 mV); the confidence intervals of these three methods overlapped. BLiSS had the lowest QRS-region NRMSE at four of five representative levels (Table 5C). The exploratory delineator-derived QRS-duration MAE favoured AWWF + AMAF (45.9 ms), IRM (51.5 ms), and ADAEKF-AR3 (59.3 ms) over BLiSS (67.1 ms).

Fig. 5 shows the most premature atrial SVEB (record 222) under the same exposure. AWWF + AMAF yielded the lowest whole-beat NRMSE (0.68), followed by BLiSS (0.82) and ADAEKF-AR3 (0.84). ADAEKF-AR3 distorted the following beat, IRM introduced a spurious deflection before it, and most methods, including BLiSS, reconstructed this following beat, which also lay within the burst, less accurately than the target SVEB. Across all 201 SVEBs (Table 3), AWWF + AMAF achieved the lowest whole-beat NRMSE (0.501), followed by IRM (0.561) and BLiSS (0.596), whereas BLiSS had the lowest QRS-region NRMSE (0.241 versus 0.268–0.491). AWWF +

Table 2 PVC reconstruction under an event-overlapping MA burst: 451 PVCs from 27 patients. The 1.2-s burst, $[\tau_k - 0.40, \tau_k + 0.80)$ s, was evaluated from −5 to 15 dB in 1-dB steps. Values are means [95% CIs].

Method	QRS-region NRMSE ↓	QRS amplitude MAE (mV) ↓	QRS-duration MAE (ms) ↓
BLiSS (proposed)	**0.217** [0.162, 0.297]	**0.194** [0.164, 0.227]	67.1 [60.5, 73.3]
AKFB	0.283 [0.223, 0.368]	0.277 [0.232, 0.327]	76.7 [66.7, 86.2]
HKF	0.307 [0.251, 0.367]	0.317 [0.266, 0.377]	75.9 [67.6, 84.1]
ADAEKF-AR3	0.949 [0.839, 1.046]	0.669 [0.565, 0.817]	59.3 [52.0, 66.8]
IRM	0.242 [0.195, 0.294]	0.237 [0.199, 0.275]	51.5 [45.7, 56.6]
AWWF + AMAF	0.231 [0.184, 0.291]	0.204 [0.176, 0.232]	**45.9** [41.0, 50.4]

Table 3 SVEB reconstruction under the same burst conditions as Table 2: 201 SVEBs from 21 patients (182 A and 19 a). P-peak timing uses the atrial subgroup defined in Section 2.3. Values are means [95% CIs].

Method	Whole-beat NRMSE ↓	QRS-region NRMSE ↓	Atrial P-peak error (ms) ↓
BLiSS (proposed)	0.596 [0.554, 0.666]	**0.241** [0.221, 0.274]	41.2 [30.1, 49.2]
AKFB	0.688 [0.639, 0.773]	0.270 [0.244, 0.309]	42.8 [27.9, 53.2]
HKF	0.734 [0.673, 0.840]	0.286 [0.253, 0.335]	39.3 [24.4, 49.3]
ADAEKF-AR3	0.648 [0.588, 0.742]	0.491 [0.423, 0.587]	**31.9** [26.5, 39.0]
IRM	0.561 [0.518, 0.639]	0.269 [0.244, 0.310]	35.0 [25.0, 39.8]
AWWF + AMAF	**0.501** [0.466, 0.563]	0.268 [0.245, 0.304]	32.3 [24.5, 36.9]

AMAF was best from 0 dB upwards and ADAEKF-AR3 at −5 dB (Table 5D). In the atrial subgroup, P-peak errors of AWWF + AMAF, ADAEKF-AR3, and IRM (32–35 ms) were closer to the reference delineation floor (29.5 ms) than that of BLiSS (41.2 ms), close to the noisy-input error (40.5 ms). The located fraction was 0.94 on the reference ECG and 0.82 on the noisy input; method-specific fractions were BLiSS 0.77, AKFB 0.82, HKF 0.80, ADAEKF-AR3 0.94, IRM 0.83, and AWWF + AMAF 0.80. Timing errors therefore describe detected P peaks and must be interpreted alongside detection coverage.

Thus, BLiSS achieved the lowest mean QRS-region errors for both ectopic types and PVC amplitude MAE, whereas AWWF + AMAF had lower SVEB whole-beat and

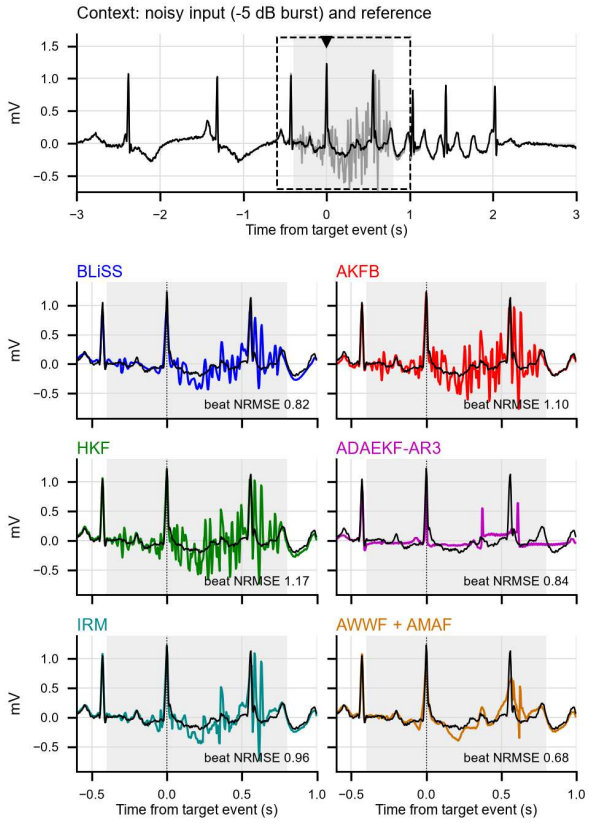


Fig. 5 SVEB example (record 222, most premature atrial beat) under the same exposure as Fig. 4. Annotation: whole-beat NRMSE.

subgroup P-peak timing errors. QRS-region gains therefore did not extend uniformly to the rest of the beat (Tables 2 and 3).

3.3 Ablations

Among the Kalman-module variants (Fig. 7), the constant-velocity RTS smoother used by BLiSS (K3) provided the highest burst ΔSNR over most levels. With WWF alone (K0) or forward-only filtering (K2), PVC QRS-region NRMSE at −5 dB was approximately 0.70, compared with 0.44 for K3, confirming the contribution of the backward pass. The random-walk model (K1) performed similarly to K3 at −5 dB but lost ΔSNR as SNR increased, whereas the higher-order constant-acceleration model (K4) did not improve on K3.

Replacing beat-specific R_k with a fixed value reduced ΔSNR by about 0.5 dB in step MA and 0.84 dB within T3 bursts and increased PVC and SVEB errors (Table 4). Re-estimating $\hat{q}$ for the fixed R produced nearly identical results, so refitting the process noise did not compensate for the loss of beat-level weighting. Moment-based q

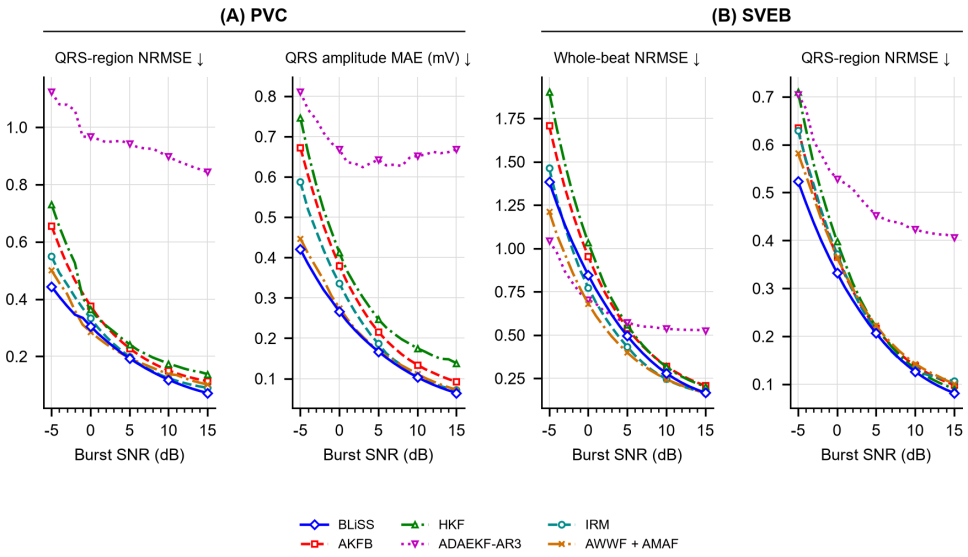


Fig. 6 Targeted ectopic-beat reconstruction versus burst SNR. (A) PVC: QRS-region NRMSE and QRS amplitude MAE. (B) SVEB: whole-beat NRMSE and QRS-region NRMSE. Curves show point estimates; uncertainty is reported in Tables 2, 3 and 5.

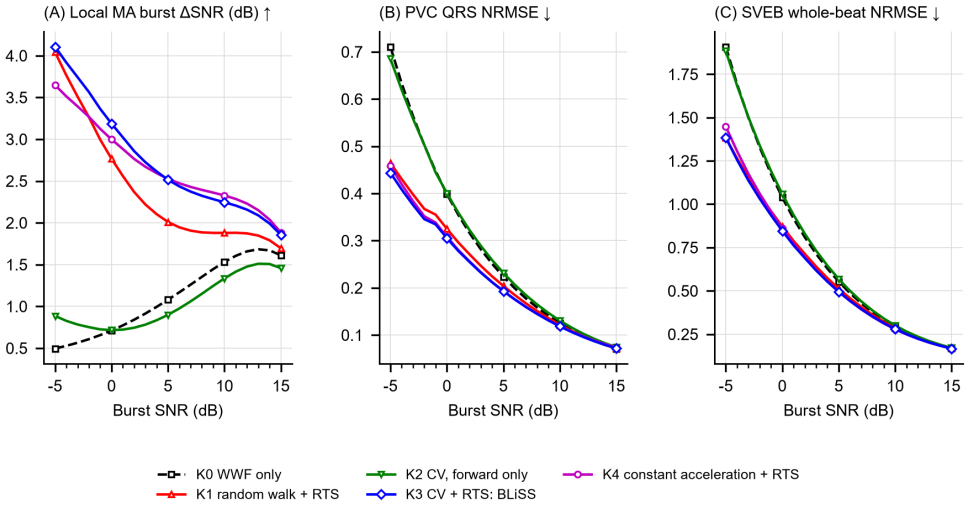


Fig. 7 Kalman-module ablation on the 1.2-s targeted bursts: K0, WWF only; K1, random walk + RTS; K2, constant velocity, forward only; K3, constant velocity + RTS (BLiSS); K4, constant acceleration + RTS.

Table 4 Effects of measurement weighting, process-noise estimation, morphology priors, and beat timing on BLiSS performance. Values are means [95% CIs].

Configuration	T2 ΔSNR (dB) ↑	T3 burst ΔSNR (dB) ↑	PVC QRS-region NRMSE ↓	SVEB whole-beat NRMSE ↓
BLiSS: full configuration	5.38 [5.13, 5.63]	5.26 [4.98, 5.54]	0.217 [0.162, 0.297]	0.596 [0.554, 0.666]
Fixed R, same $\hat{q}$ as BLiSS	4.85 [4.60, 5.10]	4.42 [4.19, 4.66]	0.272 [0.206, 0.362]	0.703 [0.646, 0.795]
Fixed R, re-estimated $\hat{q}$	4.84 [4.61, 5.08]	4.40 [4.17, 4.63]	0.273 [0.206, 0.362]	0.703 [0.647, 0.795]
Moment-based q	3.10 [2.71, 3.52]	3.30 [2.94, 3.66]	0.304 [0.224, 0.383]	0.585 [0.550, 0.639]
With preceding-beat morphology prior	5.32 [5.11, 5.52]	5.23 [4.97, 5.48]	0.314 [0.243, 0.414]	0.627 [0.573, 0.729]
Annotated R peaks (timing sensitivity)	5.37 [5.12, 5.61]	5.24 [4.97, 5.52]	0.207 [0.162, 0.268]	0.591 [0.550, 0.656]

reduced ΔSNR by approximately 2 dB and increased PVC error, although its SVEB whole-beat NRMSE was slightly lower than BLiSS (0.585 versus 0.596). Adding a preceding-beat morphology prior preserved ΔSNR but increased PVC QRS-region NRMSE from 0.217 to 0.314 and SVEB whole-beat NRMSE from 0.596 to 0.627. Annotated R peaks changed performance only marginally, indicating that detector timing had a limited effect under the evaluated conditions.

Together, the ablations support beat-specific measurement weighting and backward smoothing as contributors to BLiSS performance. The constant-velocity RTS model balanced noise suppression and reconstruction error, without a consistent benefit from the higher-order model. Moment-based q slightly reduced SVEB whole-beat error but lowered denoising gain and PVC fidelity. The preceding-beat prior increased ectopic reconstruction errors despite only small changes in overall ΔSNR (Fig. 7, Table 4). These trade-offs show that overall denoising gain alone is insufficient to guide model selection when ectopic-beat preservation is also required.

3.4 Discussion

BLiSS achieved the highest observed recording-level mean ΔSNR (Table 1), although its advantage varied with contamination level. At −5 dB, it exceeded AWWF + AMAF by approximately 0.8 dB in both step and burst MA; from 0 dB upward, AWWF + AMAF matched or exceeded BLiSS at most representative levels, with differences within 0.5 dB (blocks A–B of Table 5). Table 1 therefore reflects level-averaged performance rather than superiority at every level. Moever, ADAEKF-AR3 led under the strongest step contamination but distorted mildly contaminated signals, producing negative ΔSNR at high input SNR. The present evaluation does not isolate which modelling or filtering assumptions caused this behaviour.

The ablations support beat-specific measurement weighting: replacing R_k with a fixed variance reduced step-MA ΔSNR from 5.38 to 4.85 dB, below AWWF + AMAF (5.04 dB), and increased ectopic reconstruction errors (Table 4). This is consistent with using signal-quality information to weight estimates under varying noise [19, 20].

Table 5 Level-specific denoising and ectopic-beat reconstruction under MA. Values are means [95% CIs]. N/A: 15 dB is the T3 background level, not a burst level.

Method	-5 dB	0 dB	5 dB	10 dB	15 dB
A. T2 step: ΔSNR (dB) ↑					
BLiSS	5.98 [5.40, 6.64]	**5.08** [4.61, 5.56]	4.63 [4.20, 5.06]	3.72 [3.29, 4.12]	2.93 [2.31, 3.51]
AKFB	1.26 [1.05, 1.48]	1.14 [0.95, 1.35]	0.79 [0.63, 0.97]	0.24 [-0.06, 0.53]	-1.01 [-1.53, -0.55]
HKF	0.96 [0.78, 1.16]	0.77 [0.59, 0.94]	0.89 [0.41, 1.48]	0.50 [-0.15, 1.00]	-0.30 [-0.98, 0.31]
ADAEKF-AR3	**7.15** [6.71, 7.62]	4.52 [3.77, 5.25]	1.80 [0.70, 2.96]	-2.17 [-3.23, -1.13]	-6.57 [-7.77, -5.42]
IRM	3.80 [3.41, 4.20]	4.13 [3.63, 4.61]	4.55 [4.06, 5.05]	**3.93** [3.40, 4.48]	2.52 [1.83, 3.19]
AWWF + AMAF	5.14 [4.57, 5.75]	4.99 [4.38, 5.56]	**4.71** [4.21, 5.21]	3.77 [3.35, 4.18]	**3.02** [2.32, 3.67]
B. T3 burst intervals: ΔSNR (dB) ↑					
BLiSS	**6.15** [5.44, 6.93]	5.09 [4.66, 5.54]	4.82 [4.31, 5.36]	3.95 [3.51, 4.37]	N/A
AKFB	1.10 [0.87, 1.33]	1.11 [0.93, 1.32]	0.89 [0.70, 1.08]	0.19 [-0.14, 0.51]	N/A
HKF	0.73 [0.53, 0.92]	0.53 [0.32, 0.76]	0.44 [0.10, 0.71]	0.73 [0.46, 0.99]	N/A
ADAEKF-AR3	5.79 [5.23, 6.32]	4.42 [3.65, 5.14]	1.42 [0.22, 2.44]	-2.39 [-3.58, -1.22]	N/A
IRM	3.89 [3.46, 4.35]	4.23 [3.78, 4.66]	4.78 [4.21, 5.39]	**4.05** [3.46, 4.61]	N/A
AWWF + AMAF	5.33 [4.72, 6.02]	**5.27** [4.71, 5.83]	**5.31** [4.67, 5.96]	3.93 [3.43, 4.41]	N/A
C. Targeted PVC: QRS-region NRMSE ↓					
BLiSS	**0.444** [0.335, 0.601]	0.306 [0.220, 0.439]	**0.193** [0.143, 0.267]	**0.118** [0.090, 0.155]	**0.072** [0.058, 0.087]
AKFB	0.655 [0.499, 0.877]	0.376 [0.293, 0.498]	0.228 [0.182, 0.293]	0.151 [0.125, 0.186]	0.113 [0.096, 0.135]
HKF	0.731 [0.551, 0.971]	0.364 [0.310, 0.425]	0.242 [0.196, 0.299]	0.175 [0.133, 0.228]	0.138 [0.096, 0.191]
ADAEKF-AR3	1.123 [0.938, 1.390]	0.967 [0.860, 1.065]	0.942 [0.828, 1.049]	0.897 [0.785, 1.003]	0.844 [0.738, 0.944]
IRM	0.549 [0.442, 0.676]	0.333 [0.269, 0.405]	0.198 [0.160, 0.241]	0.124 [0.100, 0.149]	0.090 [0.073, 0.110]
AWWF + AMAF	0.501 [0.355, 0.738]	**0.286** [0.238, 0.343]	0.195 [0.165, 0.229]	0.138 [0.115, 0.163]	0.102 [0.079, 0.126]
D. Targeted SVEB: whole-beat NRMSE ↓					
BLiSS	1.385 [1.281, 1.524]	0.847 [0.784, 0.945]	0.495 [0.457, 0.559]	0.280 [0.260, 0.318]	0.167 [0.156, 0.187]
AKFB	1.707 [1.574, 1.912]	0.954 [0.882, 1.072]	0.540 [0.501, 0.607]	0.321 [0.299, 0.361]	0.208 [0.192, 0.237]
HKF	1.903 [1.750, 2.160]	1.034 [0.942, 1.179]	0.558 [0.513, 0.637]	0.317 [0.289, 0.370]	0.203 [0.183, 0.241]
ADAEKF-AR3	**1.044** [0.960, 1.165]	0.703 [0.635, 0.807]	0.573 [0.508, 0.667]	0.536 [0.470, 0.625]	0.526 [0.456, 0.621]
IRM	1.462 [1.351, 1.665]	0.774 [0.710, 0.884]	0.433 [0.400, 0.491]	0.246 [0.228, 0.278]	0.167 [0.156, 0.186]
AWWF + AMAF	1.214 [1.123, 1.372]	**0.681** [0.632, 0.767]	**0.399** [0.371, 0.448]	**0.245** [0.228, 0.275]	**0.165** [0.153, 0.182]

Section 3.2 showed that BLiSS achieved the lowest mean QRS-region errors, whereas AWWF + AMAF had lower SVEB whole-beat and atrial-subgroup P-peak timing errors. These differences may partly reflect the methods' assumptions. IRM uses similar surrounding beats [3], which may constrain reconstruction of isolated morphologies. AWWF + AMAF filters within and between QRS complexes differently [5]; its boundary selection may affect wide PVCs, although this mechanism was not tested directly. Its mean PVC QRS-region NRMSE was only slightly higher than BLiSS (0.231 versus 0.217), with overlapping CIs. BLiSS uses within-beat smoothness without a neighbouring-beat morphology prior; adding such a prior increased PVC error in the ablation. These findings support assessing ectopic waveform fidelity separately from overall noise suppression.

Limitations.

BLiSS constants (R_0, K_0, beat windows, and baseline anchors) and reference-anchored evaluation windows ($\pm$60 ms for QRS and 1 s for the whole beat) were set heuristically, without systematic sensitivity analysis. Fixed windows may not cover wide or atypical complexes. P-peak timing included only 30 atrial SVEBs from five patients; PVC QRS duration used delineator-derived reference boundaries, with the delineator selected mainly on normal beats [4, 32]. These exploratory endpoints do not establish clinical interval accuracy [1]. Target-beat scoring also omitted distortions in following beats (Figs. 4 and 5). Finally, MA came from a single database [29]; performance on naturally contaminated wearable ECG remains to be evaluated.

Future work should evaluate BLiSS on wearable ECG recordings with naturally occurring artifacts and during long-term arrhythmia monitoring. Algorithmic extensions should incorporate information from preceding and following beats while preserving abrupt changes in P-wave morphology, complex QRS waveforms, and ST-segment shape.

4 Conclusion

BLiSS combines beat-local Kalman–RTS smoothing with measurement weighting derived from raw-beat signal quality. It achieved the highest observed mean ΔSNR across four MA schedules and the lowest mean QRS-region errors for PVCs and SVEBs, although no method led at every noise level or endpoint. AWWF + AMAF achieved lower SVEB whole-beat and atrial-subgroup P-peak timing errors than BLiSS. Ablations supported beat-specific weighting and backward smoothing, while a preceding-beat prior increased ectopic reconstruction error despite small changes in overall ΔSNR. These findings show that noise suppression and ectopic waveform preservation require separate assessment. Future work should evaluate BLiSS on naturally contaminated wearable recordings and incorporate neighbouring-beat information while preserving abrupt P-wave, QRS, and ST-segment changes.

Declarations

- **Funding** Not applicable.

- **Conflict of interest** The author declares no conflict of interest.
- **Ethics approval and consent to participate** Not applicable; this study used only publicly available, de-identified databases.
- **Consent for publication** Not applicable.
- **Data availability** All data are publicly available from PhysioNet [27, 31–33].
- **Code availability** Available from the corresponding author on reasonable request.

References

[1] Kligfield, P., Gettes, L.S., Bailey, J.J., Childers, R., Deal, B.J., Hancock, E.W., van Herpen, G., Kors, J.A., Macfarlane, P., Mirvis, D.M., Pahlm, O., Rautaharju, P., Wagner, G.S.: Recommendations for the standardization and interpretation of the electrocardiogram: Part I: The electrocardiogram and its technology. Journal of the American College of Cardiology **49**(10), 1109–1127 (2007) https://doi.org/10.1016/j.jacc.2007.01.024

[2] Sameni, R., Shamsollahi, M.B., Jutten, C., Clifford, G.D.: A nonlinear Bayesian filtering framework for ECG denoising. IEEE Transactions on Biomedical Engineering **54**(12), 2172–2185 (2007) https://doi.org/10.1109/TBME.2007.897817

[3] Atanasoski, V., Petrovic, J., Popovic Maneski, L., Miletic, M., Babic, M., Nikolic, A., Panescu, D., Ivanovic, M.D.: A morphology-preserving algorithm for denoising of EMG-contaminated ECG signals. IEEE Open Journal of Engineering in Medicine and Biology **5**, 296–305 (2024) https://doi.org/10.1109/OJEMB.2024.3380352

[4] Martínez, J.P., Almeida, R., Olmos, S., Rocha, A.P., Laguna, P.: A wavelet-based ECG delineator: Evaluation on standard databases. IEEE Transactions on Biomedical Engineering **51**(4), 570–581 (2004) https://doi.org/10.1109/TBME.2003.821031

[5] Li, Y., Su, Z., Chen, K., Zhang, W., Du, M.: Application of an EMG interference filtering method to dynamic ECGs based on an adaptive wavelet-Wiener filter and adaptive moving average filter. Biomedical Signal Processing and Control **72**, 103344 (2022) https://doi.org/10.1016/j.bspc.2021.103344

[6] Kalman, R.E.: A new approach to linear filtering and prediction problems. Journal of Basic Engineering **82**(1), 35–45 (1960) https://doi.org/10.1115/1.3662552

[7] Rauch, H.E., Tung, F., Striebel, C.T.: Maximum likelihood estimates of linear dynamic systems. AIAA Journal **3**(8), 1445–1450 (1965) https://doi.org/10.2514/3.3166

[8] Särkkä, S.: Bayesian Filtering and Smoothing. Cambridge University Press, Cambridge (2013). https://doi.org/10.1017/CBO9781139344203

[9] McSharry, P.E., Clifford, G.D., Tarassenko, L., Smith, L.A.: A dynamical model for generating synthetic electrocardiogram signals. IEEE Transactions on Biomedical Engineering **50**(3), 289–294 (2003) https://doi.org/10.1109/TBME.2003.808805

[10] Sameni, R., Shamsollahi, M.B., Jutten, C.: Filtering electrocardiogram signals using the extended Kalman filter. In: Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 5639–5642 (2005). https://doi.org/10.1109/IEMBS.2005.1615765

[11] Sayadi, O., Shamsollahi, M.B., Clifford, G.D.: Synthetic ECG generation and Bayesian filtering using a Gaussian wave-based dynamical model. Physiological Measurement **31**(10), 1309–1329 (2010) https://doi.org/10.1088/0967-3334/31/10/002

[12] Vullings, R., de Vries, B., Bergmans, J.W.M.: An adaptive Kalman filter for ECG signal enhancement. IEEE Transactions on Biomedical Engineering **58**(4), 1094–1103 (2011) https://doi.org/10.1109/TBME.2010.2099229

[13] Danandeh Hesar, H., Mohebbi, M.: An adaptive Kalman filter bank for ECG denoising. IEEE Journal of Biomedical and Health Informatics **25**(1), 13–21 (2021) https://doi.org/10.1109/JBHI.2020.2982935

[14] Revach, G., Locher, T., Shlezinger, N., van Sloun, R.J.G., Vullings, R.: HKF: Hierarchical Kalman filtering with online learned evolution priors for adaptive ECG denoising. IEEE Transactions on Signal Processing **72**, 3990–4006 (2024) https://doi.org/10.1109/TSP.2024.3443875

[15] Danandeh Hesar, H., Danandeh Hesar, A.: Adaptive augmented cubature Kalman filter/smoother for ECG denoising. Biomedical Engineering Letters **14**(4), 689–705 (2024) https://doi.org/10.1007/s13534-024-00362-7

[16] Danandeh Hesar, H., Danandeh Hesar, A.: Adaptive dual augmented extended Kalman filtering of ECG signals. Measurement **239**, 115457 (2025) https://doi.org/10.1016/j.measurement.2024.115457

[17] de Chazal, P., O'Dwyer, M., Reilly, R.B.: Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Transactions on Biomedical Engineering **51**(7), 1196–1206 (2004) https://doi.org/10.1109/TBME.2004.827359

[18] García-Isla, G., Mainardi, L., Corino, V.D.A.: A detector for premature atrial and ventricular complexes. Frontiers in Physiology **12**, 678558 (2021) https://doi.org/10.3389/fphys.2021.678558

[19] Li, Q., Mark, R.G., Clifford, G.D.: Robust heart rate estimation from multiple

asynchronous noisy sources using signal quality indices and a Kalman filter. Physiological Measurement **29**(1), 15–32 (2008) https://doi.org/10.1088/0967-3334/29/1/002

[20] Smital, L., Haider, C.R., Vítek, M., Leinveber, P., Jurák, P., Němcová, A., Smíšek, R., Maršánová, L., Provazník, I., Felton, C.L., Gilbert, B.K., Holmes, D.R.: Real-time quality assessment of long-term ECG signals recorded by wearables in free-living conditions. IEEE Transactions on Biomedical Engineering **67**(10), 2721–2734 (2020) https://doi.org/10.1109/TBME.2020.2969719

[21] Liu, J., Zhou, X., Liu, X., Wang, X., Zhou, J.: A rhythm-specific ECG signal quality assessment framework for robust cardiac health monitoring of AI-based arrhythmia classifier. In: Proceedings of the IEEE Biomedical Circuits and Systems Conference (BioCAS), pp. 1–5 (2023). https://doi.org/10.1109/BioCAS58349.2023.10388946

[22] Liu, H., Zhang, S., Gamboa, H., Xue, T., Zhou, C., Schultz, T.: Taxonomy and real-time classification of artifacts during biosignal acquisition: A starter study and dataset of ECG. IEEE Sensors Journal **24**(6), 9162–9171 (2024) https://doi.org/10.1109/JSEN.2024.3356651

[23] Rahman, S., Yearwood, J., Karmakar, C.: Design and evaluation of a knowledge-based ECG noise filtering framework. Scientific Reports **16**, 2429 (2026) https://doi.org/10.1038/s41598-025-32249-7

[24] Rahman, S., Karmakar, C., Natgunanathan, I., Yearwood, J., Palaniswami, M.: Robustness of electrocardiogram signal quality indices. Journal of the Royal Society Interface **19**(189), 20220012 (2022) https://doi.org/10.1098/rsif.2022.0012

[25] Smital, L., Vítek, M., Kozumplík, J., Provazník, I.: Adaptive wavelet Wiener filtering of ECG signals. IEEE Transactions on Biomedical Engineering **60**(2), 437–445 (2013) https://doi.org/10.1109/TBME.2012.2228482

[26] Moody, G.B., Mark, R.G.: The impact of the MIT-BIH arrhythmia database. IEEE Engineering in Medicine and Biology Magazine **20**(3), 45–50 (2001) https://doi.org/10.1109/51.932724

[27] Moody, G.B., Mark, R.G.: MIT-BIH Arrhythmia Database (version 1.0.0). PhysioNet (1992). https://doi.org/10.13026/C2F305

[28] Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K., Stanley, H.E.: PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation **101**(23), 215–220 (2000) https://doi.org/10.1161/01.CIR.101.23.e215

[29] Moody, G.B., Muldrow, W.E., Mark, R.G.: A noise stress test for arrhythmia detectors. In: Computers in Cardiology, vol. 11, pp. 381–384 (1984). https://georgebmoody.com/publications/nst-cinc-1984.pdf

[30] Association for the Advancement of Medical Instrumentation: ANSI/AAMI EC57:2012 Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms. AAMI, Arlington, VA (2012). https://doi.org/10.2345/9781570204784

[31] Moody, G.B., Muldrow, W.E., Mark, R.G.: The MIT-BIH Noise Stress Test Database (version 1.0.0). PhysioNet (1992). https://doi.org/10.13026/C2HS3T

[32] Laguna, P., Mark, R.G., Goldberger, A.L., Moody, G.B.: A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. In: Computers in Cardiology, vol. 24, pp. 673–676 (1997). https://doi.org/10.1109/CIC.1997.648140

[33] Němcová, A., Smíšek, R., Vítek, M., Smital, L.: MIT-BIH Arrhythmia Database P-Wave Annotations (version 1.0.0). PhysioNet (2018). https://doi.org/10.13026/C2108F

[34] Xie, C., McCullum, L., Johnson, A., Pollard, T., Gow, B., Moody, B.: Waveform Database Software Package (WFDB) for Python. PhysioNet (2023). https://doi.org/10.13026/JT5E-GQ15