# Uncertainty in heart rate complexity metrics caused by R-peak perturbations

Nicholas J. Napoli[a,b,*], Matthew W. Demas[c], Sanjana Mendu[c], Chad L. Stephens[d], Kellie D. Kennedy[d], Angela R. Harrivel[d], Randall E. Bailey[d], Laura E. Barnes[c,e]

[a] Dept. of Electrical and Computer Engineering, University Virginia, Charlottesville, Va, 22904, United States
[b] National Institute of Aerospace, Hampton, Va, 23681, United States
[c] Systems and Information Engineering, University of Virginia, Charlottesville, Va, 22904, United States
[d] NASA Langley Research Center, Hampton, Va, 23681, United States
[e] Data Science Institute, University of Virginia, Charlottesville, VA, 22904, United States

## ARTICLE INFO

## ABSTRACT

Heart rate complexity (HRC) is a proven metric for gaining insight into human stress and physiological deterioration. To calculate HRC, the detection of the exact instance of when the heart beats, the R-peak, is necessary. Electrocardiogram (ECG) signals can often be corrupted by environmental noise (e.g., from electromagnetic interference, movement artifacts), which can potentially alter the HRC measurement, producing erroneous inputs which feed into decision support models. Current literature has only investigated how HRC is affected by noise when R-peak detection errors occur (false positives and false negatives). However, the numerical methods used to calculate HRC are also sensitive to the specific location of the fiducial point of the R-peak. This raises many questions regarding how this fiducial point is altered by noise, the resulting impact on the measured HRC, and how we can account for noisy HRC measures as inputs into our decision models. This work uses Monte Carlo simulations to systematically add white and pink noise at different permutations of signal-to-noise ratios (SNRs), time segments, sampling rates, and HRC measurements to characterize the influence of noise on the HRC measure by altering the fiducial point of the R-peak. Using the generated information from these simulations provides improved decision processes for system design which address key concerns such as permutation entropy being a more precise, reliable, less biased, and more sensitive measurement for HRC than sample and approximate entropy.

## 1. Introduction

Heart rate complexity (HRC) have been established as an integral latent physiological indicator which have been utilized in numerous studies assessing cognitive workload [1–3], physiological health [4–8], autonomic nervous system insights [7], and analysis of depression [9]. HRC is often operationalized in the form of entropy measures, which are calculated using a person's heart rate variability (HRV) [4]. These measures require accurate, precise detection of individual heart beats. While there are many methods to measure heart rate, such as electrocardiograms (ECGs), photoplethysmograms, and optical heart rate monitors, ECGs are the only method that provides information about the QRS waveform - the so-called combination of the three deflections typically seen in the ECG - which corresponds to the depolarization of the heart's ventricles (a single heart beat). These QRS waveforms are critical to precisely calculating R-R intervals and thus HRV. HRV is then used to quantify the fluctuations in HRC. Thus, ECG telemetry stands as the only proven method for calculating HRC accurately. For these predictive technologies to provide insights, computational algorithms are required to examine variations from normal physiology [10]. Since many of these technologies require consistent monitoring, intermittent noise is inevitably introduced into the system. Thus noisy occurrences and detection of physiological anomalies generate ambiguities, leading to, false detection, inaccurate decision support, and alarm fatigue [11–13]. However, much of the previous, original work surrounding HRV detection was developed on cleaned, retrospective datasets

[4,5,14,15]; thus, we cannot assume that complexity algorithms perform appropriately on noisy and corrupt data.

Understanding the robustness of HRC measurements under noisy conditions allows for corrective computational approaches and system design [13]. These corrective measurements are becoming a paramount objective as systems are expected to work in real time which leads to high risk of signal corruption. In this investigation, we aim to fill this research gap by evaluating complexity algorithms under noisy and corrupted conditions.

**Prior Work.** One of the original HRC algorithm implementations was generated from Pincus's work which introduced approximate entropy [16]. These ideas, concepts, and applications used to evaluate heart rate entropy spawned a multitude of measures [15,17]. The two most utilized complexity measures since are approximate and sample entropy. In line with Pincus's, Moorman's, and Costa's prior work, these complexity measures have been used on retrospective data to demonstrate predictive model feasibility and practicality in classifying patients in clinical settings [18]. While extensive studies have been conducted on various HRC methods and applications in the past few decades, few studies exist that evaluate the influence of noise and erroneous behavior on these measures. These works have demonstrated that missed detection of QRS complexes (heart beats) could drastically alter downstream entropy measurements [13,19]. This was demonstrated by examining the effects of corruption through the false positive and false negative rates of the detection algorithms rather than looking at how the SNR itself alters entropy [13,19]. This was done to control for the false positive and false negatives in the detection algorithms. These approaches manipulate the HRV signal through the concatenation of sequences or random down and up sampling, which changes the entropy measurement associated with the HRC [13,19]. Although robust and accurate QRS complex detection algorithms are highly desirable and have been achieved through various types of machine learning approaches [5], QRS detection is not the only avenue through which erroneous entropy measurement can occur. ECG sampling frequency and R-peak interpolation have demonstrated importance as methodological considerations in obtaining consistent HRV signals [20]. These concerns underline the importance of the fiducial marker for the R-peak and its relationship to producing reliable HRV measurements, thus impacting the HRC. This can be depicted in Fig. 1, where the illustration of the ECG signal R-peak has an arbitrary confidence interval. Therefore, this induces uncertainty in the HRV signal below with complementary confidence intervals. Noise plays an integral role in where the fiducial marker for the R-peak is placed, which affects downstream applications. However, neither the extent to which noise alters HRV and HRC measurement through changes in the fiducial marker nor the downstream implications of these fiducial markers on statistical models have been characterized. Ultimately, current literature's focus has been on the advancement of complexity algorithms and their utility for classification, rather than understanding how and when these algorithms fail. Developing this understanding is critical for appropriate system design in preventing alarm fatigue, erroneous prediction, and decision support systems [11,13].

This work raises six relevant research questions (RQs): 1) How does the SNR, sampling, and the color of the noise alter the fiducial marker of the R-peak? 2) Is any particular entropy measurement more precise under corruption than others? 3) Does signal length contribute to a more robust measurement in the presence of corruption? 4) Can a precise entropy measurement still be statistically sensitive in differing HRC dynamics expected in study populations? 5) If the fiducial point is altered by noise, how does this effect the direction of the measured entropy calculation? 6) How much impact does noise and the fiducial marker downstream have on the statistical implications of a study?

**Challenges.** Due to the multiple stages involved in the process of calculating HRC, it is difficult to pinpoint exactly where noise alters an entropy measure's reliability. Noise can alter HRC reliability through numerous avenues such as False Positives for the QRS detection, false
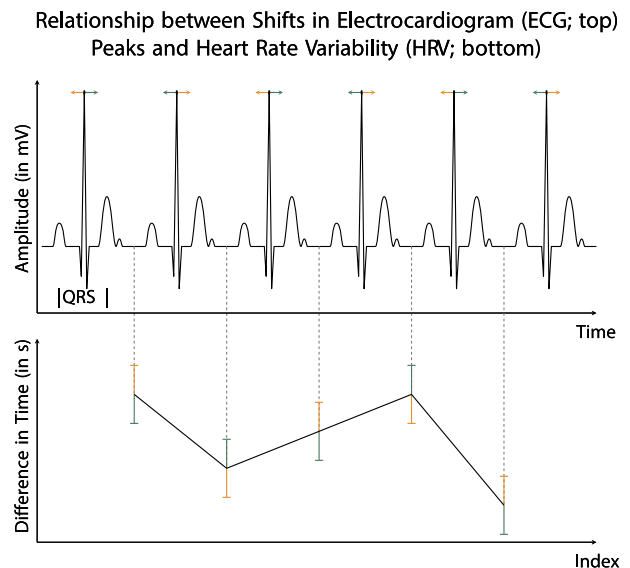
Relationship between Shifts in Electrocardiogram (ECG; top)
Peaks and Heart Rate Variability (HRV; bottom)



**Fig. 1.** Two simulated plots are depicted to illustrate how heart rate variability (HRV) is generated by the variation in timing between the "peaks" of the QRS complex, otherwise known as the R-wave. The top graphic is an ECG in which the R-peak's fiducial point can shift left or right (shown in green and orange). The bottom graphic is the HRV value that is the timing between R-peaks, which can potential increase or decrease (shown in green and orange). Therefore shifts within the identified location of R-wave peaks results in variation in the HRV time series (highlighted in green and orange). This Variation in the identified location of the R-peak from the QRS complex results in uncertainty in the heart rate variability (HRV) signal.

negatives of the QRS detection, or the QRS waveform fiducial point shifting. However, the robustness of current methods has only been evaluated by resampling HRV signals, thus only examining the false positive and false negative rate [19,21,22]. Therefore, the fundamental issue of how corruption in an ECG signal alters the HRV signal through slight shifts in fiducial points and QRS waveform timing has been neglected and is difficult to capture. Additionally, the numerous types of complexity measures produce different outputs with different scales, further complicating the validation process and comparative analysis.

**Insight.** A systematic process of validation in which noise can be exactly replicated through the seeding of a random number generator and introduced to a signal provides control and reproducibility when applied to testing datasets. Monte Carlo simulations allow for an increased number of controlled iterations enabling the evaluation of the aleatory uncertainty from a set of selected parameters (e.g, Signal-to-Noise Ratio (SNR), type of entropy measurement) to form distributions that characterize errors caused by the fiducial R-peak. This controlled simulation enables rigorous analysis on how the fiducial points impose errors in the system. Here we strictly analyze shifts in the mean and variance associated with the fiducial point at which noise enters the system. Using these approaches, we can overcome scaling differences from different entropy measures by evaluating the percent error (PE) in the various reported entropy distributions with respect to that found without the introduced noise.

**Contributions.** We designed and implemented a Monte Carlo simulation to evaluate the effects of time window size and SNR level of ECG signals on the percent error of three entropy calculation methods applied to those ECG signals. This method provided a controlled framework for evaluating the effects of ECG corruption on the reliability of three entropy measures.

The contributions of this work are as follows:

1. We characterize how SNR alters the locations of the fiducial point of the R-peak via a probability mass function (PMF), and demonstrate that the fiducial point for the R-peak is most affected by white noise

at varying SNR levels and sampling frequency, as compared to pink noise.

2. We demonstrate that permutation entropy is more precise than both approximate and sample entropy (i.e., it has the lowest standard deviation).

3. We demonstrate that, regardless of the entropy measurement or type of noise, as the time window segment increases, the precision of the measured entropy improves.

4. We demonstrate that increased precision does not imply a lack of sensitivity for demonstrating significant differences in HRC.

5. We demonstrate that as the SNR increases (altering the fiducial marker), there is a directional change in the measured entropy.

6. We show the level of SNR that affects the likelihood of making a distinguishable distribution statistically indistinguishable.

All of these findings ultimately aid us in future design methodologies to reduce predictive modeling error downstream from upstream feature extraction methodologies.

## 2. Methods

The methods section is split into two sections. The first two parts of the methods, incorporate "Simulation Part A: Modeling Shifts of the Fiducial Point" and "Simulation Part B: Modeling Entropy Changes" which mainly utilize the PhysioNet data to address research questions RQ1 to RQ3. The third section of the methods, "Hypoxia HR Complexity Analysis", utilizes the NASA dataset and addresses RQ4 to RQ6.

### 2.1. Modeling Shifts of the fiducial point

This section outlines the Monte Carlo simulation implementation method shown in Fig. 2. These simulations were designed to study the effects of time window segmentation length and signal corruption resulting from multiple "colors" of noise on the prevalence of fiducial marker shifts on approximate, permutation, and sample entropy calculations.

#### 2.1.1. Normal sinus rhythm data

Single lead ECG data from 10 subjects (130 min, 128 Hz) selected from the MIT-BIH Normal Sinus Rhythm Database on PhysioNet [14] were utilized. The Normal Sinus Rhythm database was selected to reduce measurement variability that could be attributed to physiological dysfunction such as atrial fibrillation, ectopy, and other disorders [18]. Each subject's ECG signal was divided into 65, 26 and 13 intervals for 2, 5, and 10 min time windows, respectively. Following signal preprocessing, noise was added, and the prevalence of shifts in the fiducial

marker (i.e., the location of the R-peak) was collected to form a PMF.

#### 2.1.2. Signal preprocessing

After segmenting the ECG signals into intervals, a high pass and low pass finite impulse response Butterworth zero phase filter was applied to each time window. This produced our cleaned ECG signal, $X_S$. The zero phase filter avoids distortion in the phase of signal [23]. A key component for also avoiding fiducial marker shifting.

#### 2.1.3. PMFs of the fiducial shifts

The following section describes the procedure for creating a PMF of the prevalence of fiducial marker shifts caused by the addition of either white or pink noise to ECG signals.

*2.1.3.1. Step 1: add noise.* In Step 1, a white or pink noise signal, $X_N$, was added to the cleaned signal $X_S$ comprising signal-to-noise ratios (SNRs) from 2 to 20 in increments of 2. The SNR level $X_N$ added to $X_S$ is defined as

$$SNR = 10\log_{10}\sqrt{\frac{\sum (X_S \cdot \bar{X}_S)^2}{\sum (X_N \cdot \bar{X}_N)^2}},$$

(1)

where $\bar{X}_N$ and $\bar{X}_S$ are the complex conjugates of $X_N$ and $X_S$. A randomly seeded $X_N$ of each SNR was added to each subject's cleaned ECG signal interval 100, 250, or 500 times for each 2, 5, and 10 min interval, respectively. White noise is designed into the implementation of the Monte Carlo Simulation to evaluate the effects of electromagnetic interference, problematic sensors, or issues with wireless devices [24,25]. A uniform distribution was used to model white noise in the frequency domain by sampling from a random Gaussian distribution in the time domain sequence. The SNR is altered by adjusting the variance of the Gaussian distribution. Pink noise was selected to evaluate the effect of correlated noise typically associated with observation noise on the ECG [26]. Pink noise was modeled as a decreasing function ($1/f$) in the frequency domain and has close similarity to Brownian motion-like noise which is modeled as a decreasing function ($1/f^2$) and is related to electrode movement noise [26]. Pink noise was implemented using a noise generator package on MATLAB's file exchange service based on the theory for discrete simulations of colored noise by Kasdin [27].

*2.1.3.2. Steps 2 and 3: identify fiducial shifts and construct fiducial shift PMFs.* In Step 2, a QRS complex detection algorithm was applied to both the cleaned and noisy data to identify individual heart beats, R-R intervals, and heart rate variability (HRV) [28]. Using the Pan-Thompkins algorithm, the fiducial mark's location for the R-peak is determined from the rising edge of the waveform [28]. The locations of
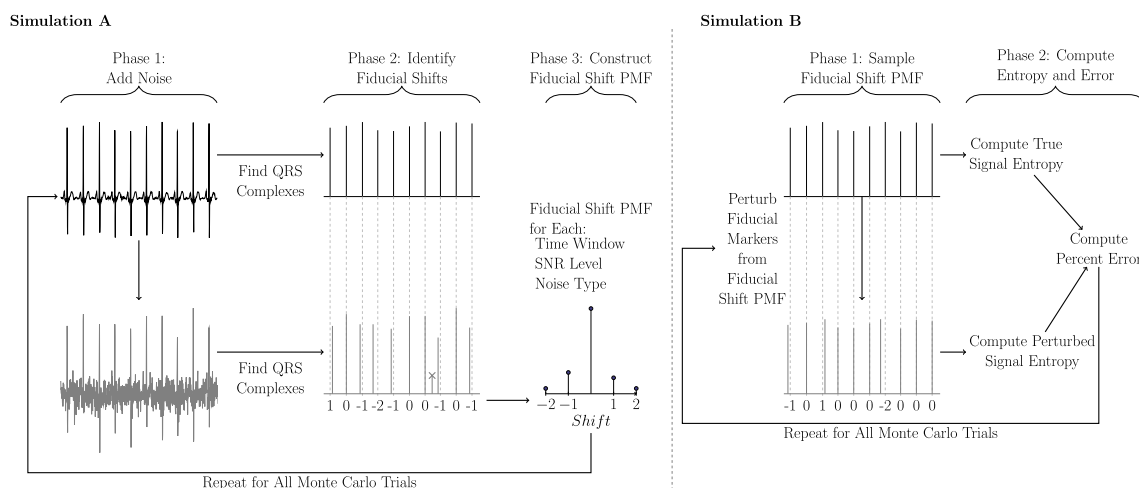


**Fig. 2.** Flow chart process of simulation.

each R-peak extracted from the noisy signal were compared to the locations of the R-peaks extracted from the filtered signal. In Step 3, the prevalence of noisy R-peaks within ± 39.1 ms of their filtered counterparts were collected to form a PMF. This particular time increment was due to our sampling frequency of 128 Hz, thus $F_s^{-1} = 128^{-1} = 7.81 \ ms$. False positives (i.e., noisy R-peaks detected outside of this window) and false negatives were not included in the construction of these PMFs. This process of creating PMFs was repeated for all subjects' time intervals (2, 5, 10 min), resulting in PMFs for each noise type (white or pink), and for each SNR level (2,4, …,20) equaling a total of 60 PMFs (i.e., 3 Time Intervals x 10 SNR Levels x 2 Types of Noise).

We then use visual inspection of PMFs and the corresponding tabulated shift frequency values to evaluate the effects of different SNRs and noise types on the prevalence of fiducial marker shifts.

### 2.2. Simulation B: Modeling Entropy Changes

The following section discusses the process of calculating the effects on entropy measurement resulting from perturbing the location of R-peaks extracted from filtered (cleaned) signals by an amount sampled from the PMFs constructed during Simulation A.

#### 2.2.1. Sample fiducial shift PMF

In Step 1, the location of R-peaks of subjects' filtered ECG signal intervals were perturbed with a probability determined from the fiducial shift PMFs constructed in Simulation A. A HRV signal was then computed for both the filtered and perturbed signals.

#### 2.2.2. Compute entropy and percent error

In Step 2, three entropy measures were applied to each of the HRV signals–approximate entropy [16], permutation entropy [17,29], and sample entropy [15].

*Approximate entropy.* is defined as

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r), \tag{2}$$

where $\Phi^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \log C_i^m(r)$. $C_i^m(r)$ is the number of matches to template $i$ of length $m$ within a tolerance of $r$ (*including* self matches) [16].

*Sample Entropy* is defined as

$$SampEn(m, r, N) = -\ln \left[ \frac{A^m(r)}{B^m(r)} \right] \tag{3}$$

where $A^m(r)$ and $B^m(r)$ are the probabilities of two sequences matching for $m + 1$ and $m$ data points respectively within a tolerance of $r$ (*excluding* self matches) [15].

*Permutation Entropy.* employs rank order (i.e., not exact distance) to quantify time series similarity. This is defined as

$$H_n = - \sum_{j=1}^{n!} p_j' \log_2 p_j' \tag{4}$$

where $p_j'$ is the proportion of the occurrence of the $j^{th}$ template (of length $m$) in the signal [17,29]. Approximate [16] and sample entropy [15] both use distance as the measurement to examine similarity in time series to quantify entropy. The main difference between the two is that unlike approximate entropy, sample entropy excludes self matches. On the other hand, permutation entropy employs rank order (i.e., not exact distance) to quantify time series similarity [17,29]. In Step 2, these three entropy values were computed for both the filtered and perturbed HRV signals and compared to calculate percent error 100, 250, or 500 times for each 2, 5, or 10 min interval, respectively; each noise type (white or pink); and each SNR level (2,4, …,20). The entropy parameters utilized in calculating the sample and approximate entropy values were $m = 2$ & $r = 0.15\sigma$ and $m = 3$ for permutation entropy order.

#### 2.2.3. Entropy precision

This subsection addresses how we compare the entropy precision for RQ2 and RQ3. RQ2 is achieved by computing the standard deviation of each time interval and then applied the natural logarithm to the data to obtain relatively equal variances and reasonably symmetric distributions. We then tested for the differences in the mean log standard deviation of the percent error using a one-way ANOVA. We then address RQ3 from Simulation B data, by constructing conditional distributions of the entropy percent error for time window sizes of 2, 5, and 10 min and for both pink and white noise by collapsing over SNR, subject, and time interval number. Using visual comparison and descriptive statistics, the effects of window size on entropy reliability were addressed for both white and pink noise.

### 2.3. Hypoxia HRC analysis

The experience of hypoxia is known to result in autonomic-nervous-system-driven changes (sympathetic arousal) in the cardiac and respiratory systems [30]. We utilize a dataset which characterizes hypoxic responses because of these known physiological changes in sympathetic arousal to address research questions RQ4 to RQ6.

#### 2.3.1. NASA hypoxia data

The dataset was collected by a research team at NASA Langley Research Center (LaRC) who subjected 49 volunteers (all with current hypoxia training certificates) to normobaric hypoxia to study the impact on aircraft pilot performance [31,32]. All participants consented to take part of the study as approved by the Institutional Review Board of NASA LaRC. Subjects in the study experienced simulated altitudes of Sea Level (21% $O_2$) and 15,000 feet (11.2% $O_2$) induced by an Environics, Inc. Reduced Oxygen Breathing Device (ROBD-2). Thus each subject was exposed to a non-hypoxic and hypoxic bout each lasting 10 min in which ECG data was collected at 256 Hz.

#### 2.3.2. Evaluating entropy sensitivity

To evaluate the sensitivity of entropy measures in detecting mild hypoxia for RQ4, we applied approximate, permutation, and sample entropy to the ECG data collected during the final 2 min and final 5 min of non-hypoxic and hypoxic exposures. Ten minute time segments were excluded from this analysis because hypoxic exposures were only 10 min long and thus not all subjects demonstrated indicators of mild hypoxia (e.g., Sp02 < 80%) until several minutes into the 15,000 ft exposures. Due to small sample size and non-normality, the non-parametric Wilcoxon rank-sum test was employed to test for differences between the hypoxic and non-hypoxic cohorts with smaller p-values indicating a greater ability for an entropy calculation method to discriminate between hypoxic and non-hypoxic states. The initialized parameters used in calculating the entropy measurements were the same as in Simulation B (The parameters used for sample and approximate entropy are $m = 2$ & $r = 0.2\sigma$. The parameters used for permutation entropy are $m = 3$ using the 'order' method when equal values were present).

#### 2.3.3. Entropy biases in hypoxic subjects

Using the NASA hypoxia ECG dataset and the PMFs of the fiducial marker shifting produced in Simulation A, we corrupted the HRV signals by altering the locations of the R-peaks for 2 and 5 min time segments at various SNR levels. Each subject and their respective cohort's (hypoxia and non-hypoxia) fiducial markers were corrupted 250 times to examine the raw change in the entropy distribution. We pooled the approximate, permutation, and sample entropy values for all subjects' Monte Carlo trials for both hypoxic and non-hypoxic exposures at each SNR level and time window size (2 and 5 min). Kernel density plots of each pooled distribution were obtained and visually inspected for skewness. For each conditional pooled distribution, the mean entropy value was computed and qualitatively assessed for general trends in the

data to address RQ5.

### 2.3.4. Cohort discriminability

We address RQ6 through an approach similar to that of RQ3. Using the NASA hypoxia dataset and the probability distributions associated with the fiducial shifts, we altered the fiducial markers of the R-peaks in the ECG signals to simulate corruption for each subject. This process was simulated 250 times for each time window segment length, cohort, and entropy measurement at each SNR level. During each simulation trial, the non-parametric Wilcoxon rank-sum test evaluated differences between the hypoxic and non-hypoxic distributions with the null-hypothesis significance level set at 0.05. After all simulation trials were complete, the percentage of tests rejecting the null-hypothesis (i.e., having $p < 0.05$) was collected. A greater percentage of null-hypothesis rejections indicates a greater ability to resolve hypoxic from non-hypoxic states.

## 3. Evaluation and discussion

In this section, we address the following research questions regarding the effects of ECG corruption on entropy dynamics which were introduced in Section I:

**RQ1** How do varying levels of SNR, noise, and sampling frequencies affect the fiducial point of the R-peak?
**RQ2** If the fiducial point of the R-peak shifts, does a single entropy method demonstrate superior precision from the proposed three types of entropy calculations that measure HRC?
**RQ3** Does increasing the HRV signal time segment length enhance HRC variance (entropy dynamics)?
**RQ4** Is increased precision associated to lack of sensitivity for demonstrating significant differences in HRC?
**RQ5** Is there a directional change in the measured entropy from corruption in which the entropy increases, decreases, or does only the variance symmetrically increase providing no actual change in complexity?
**RQ6** At what simulated SNR level do fiducial shifts render two previously distinguishable distributions indistinguishable?

### 3.1. RQ1 — Fiducial R-peak marker, type of noise, and SNR

We hypothesized that the fiducial marker will change as a function of the SNR, type of noise, and sampling rate. Each row in Tables 2–4, and, 4 present a PMF of the fiducial marker shifting for white, pink noise, and sampling, respectively, as a function of the targeted SNR. The PMF was generated by phase 1 of the simulation shown in Fig. 2 which describes the probability of the fiducial marker shifting. Each shift in the fiducial marker is a single discretized point or index in the time series equivalent to $F_s^{-1}$. For the case where the sampling rate of the discretized signal is 128 Hz, the index shifts in increments of 7.81 ms.

**Table 2**
Pink noise Monte Carlo simulation summary of probability distribution of the fiducial shift (Trials = ∼ 65,000, $F_s$ = 128 Hz).

| Target | Distribution of Fiducial Shift by Milliseconds | | | | |
|---|---|---|---|---|---|
| SNR | −15.62 ms | −7.81 ms | 0 ms | 7.81 ms | 15.62 ms |
| 2 | 0.0 | 4.63 | 90.39 | 4.95 | 0.0 |
| 4 | 0.0 | 3.68 | 92.41 | 3.89 | 0.0 |
| 6 | 0.0 | 2.94 | 93.97 | 3.07 | 0.0 |
| 8 | 0.0 | 2.33 | 95.22 | 2.42 | 0.0 |
| 10 | 0.0 | 1.85 | 96.21 | 1.92 | 0.0 |
| 12 | 0.0 | 1.47 | 96.99 | 1.53 | 0.0 |
| 14 | 0.0 | 1.18 | 97.60 | 1.21 | 0.0 |
| 16 | 0.0 | 0.94 | 98.10 | 0.96 | 0.0 |
| 18 | 0.0 | 0.75 | 98.48 | 0.76 | 0.0 |
| 20 | 0.0 | 0.60 | 98.79 | 0.60 | 0.0 |

**Table 3**
White noise Monte Carlo simulation summary of probability distribution of the fiducial shift (Trials = ∼ 65,000, $F_s$ = 128 Hz).

| Target | Distribution of Fiducial Shift by Milliseconds | | | | |
|---|---|---|---|---|---|
| SNR | −15.62 ms | −7.81 ms | 0 ms | 7.81 ms | 15.62 ms |
| 2 | 0.01 | 9.74 | 78.9 | 11.21 | 0.10 |
| 4 | 0.0 | 8.02 | 82.90 | 9.02 | 0.04 |
| 6 | 0.0 | 6.52 | 86.29 | 7.16 | 0.01 |
| 8 | 0.0 | 5.23 | 89.13 | 5.62 | 0.0 |
| 10 | 0.0 | 4.20 | 91.35 | 4.44 | 0.0 |
| 12 | 0.0 | 3.33 | 93.15 | 3.49 | 0.0 |
| 14 | 0.0 | 2.65 | 94.57 | 2.79 | 0.0 |
| 16 | 0.0 | 2.10 | 95.71 | 2.17 | 0.0 |
| 18 | 0.0 | 1.66 | 96.60 | 1.73 | 0.0 |
| 20 | 0.0 | 1.37 | 97.21 | 1.41 | 0.0 |

**Table 4**
White noise Monte Carlo simulation summary of probability distribution of the fiducial shift (Trials = ∼ 13,750, $F_s$ = 256 Hz).

| Target | Distribution of Fiducial Shift by Milliseconds | | | | |
|---|---|---|---|---|---|
| SNR | −7.81 ms | −3.91 ms | 0 ms | 3.91 ms | −7.81 ms |
| 2 | 0.01 | 20.75 | 52.48 | 23.25 | .0244 |
| 4 | 0.0 | 19.02 | 57.42 | 21.76 | 0.01 |
| 6 | 0.0 | 17.05 | 62.50 | 19.61 | 0.01 |
| 8 | 0.0 | 14.79 | 67.71 | 17.15 | 0.0 |
| 10 | 0.0 | 12.54 | 72.85 | 14.47 | 0.0 |
| 12 | 0.0 | 10.49 | 77.55 | 11.90 | 0.0 |
| 14 | 0.0 | 8.59 | 81.77 | 9.61 | 0.0 |
| 16 | 0.0 | 6.96 | 85.40 | 7.62 | 0.0 |
| 18 | 0.0 | 5.58 | 88.37 | 6.03 | 0.0 |
| 20 | 0.0 | 4.42 | 90.84 | 4.72 | 0.0 |

**Table 1**
White noise (WN) Entropy movement (Trials = ∼ 250).

| Time | Target | Approximate Entropy | | Permutation Entropy | | Sample Entropy | |
|---|---|---|---|---|---|---|---|
| **Mins** | **SNR** | $H_\mu$ | $NH_\mu$ | $H_\mu$ | $NH_\mu$ | $H_\mu$ | $NH_\mu$ |
| 2 | 2 | 0.781 | 0.750 | 0.824 | 0.847 | 1.277 | 1.414 |
| 2 | 6 | 0.771 | 0.747 | 0.818 | 0.843 | 1.252 | 1.407 |
| 2 | 10 | 0.776 | 0.742 | 0.811 | 0.838 | 1.230 | 1.396 |
| 2 | 14 | 0.767 | 0.740 | 0.806 | 0.835 | 1.216 | 1.390 |
| 2 | 18 | 0.764 | 0.738 | 0.803 | 0.833 | 1.207 | 1.387 |
| 5 | 2 | 0.896 | 0.912 | 0.896 | 0.912 | 1.378 | 1.474 |
| 5 | 6 | 0.891 | 0.909 | 0.891 | 0.910 | 1.355 | 1.458 |
| 5 | 10 | 0.887 | 0.907 | 0.887 | 0.907 | 1.339 | 1.446 |
| 5 | 14 | 0.885 | 0.906 | 0.885 | 0.906 | 1.330 | 1.439 |
| 5 | 18 | 0.883 | 0.904 | 0.883 | 0.904 | 1.322 | 1.432 |

**Table 5**
Pink noise Monte Carlo simulation summary of probability distribution of the fiducial shift (Trials = $\sim$ 13,750, $F_s = 256 Hz$).

| Target | Distribution of Fiducial Shift by Milliseconds | | | | |
|---|---|---|---|---|---|
| SNR | $-7.81$ ms | $-3.91$ ms | 0 ms | 3.91 ms | $-7.81$ ms |
| 2 | 0.0 | 11.98 | 74.02 | 13.64 | 0.0 |
| 4 | 0.0 | 10.62 | 77.23 | 11.99 | 0.0 |
| 6 | 0.0 | 9.13 | 80.63 | 10.17 | 0.0 |
| 8 | 0.0 | 74.02 | 84.42 | 8.13 | 0.0 |
| 10 | 0.0 | 5.89 | 87.67 | 6.40 | 0.0 |
| 12 | 0.0 | 4.72 | 90.24 | 5.01 | 0.0 |
| 14 | 0.0 | 3.76 | 92.22 | 4.00 | 0.0 |
| 16 | 0.0 | 2.97 | 93.83 | 3.18 | 0.0 |
| 18 | 0.0 | 2.37 | 95.08 | 2.53 | 0.0 |
| 20 | 0.0 | 1.90 | 96.06 | 2.02 | 0.0 |

These tables demonstrate that as the SNR decreases, there is a greater probability of altering the fiducial marker. As a function of the type of noise applied to the signal, white noise is the most corruptive in altering the fiducial marker. In terms of sampling frequency, we analyzed the two data sets with respect to their sampling frequency (Fs) and how the fiducial point is altered. The MIT-BIH data set has a Fs = 128 Hz and the NASA data set was sampled at 256 Hz. We can visually note that across all levels of SNR and types of noise the high sampling frequency has a greater probability of fiducial marker shifting with respect to their SNR level. However if compare the contextual meaning of the fiducial marker between the two sampling frequencies, the error of the timing differences (3.91 ms vs 7.81 ms) is smaller. This is because one fiducial shift for 256 Hz is equivalent 3.91 ms vs 7.81 ms for the 128 Hz sampling. The higher sampling frequency of 256 Hz has essentially zero instances of the fiducial marker extending out 7.81 ms. From the perspective of implementing entropy measurement that rely on timing distances, such as sample or approximate entropy, this may seem beneficial for improving robustness of the measurement. However, these beat-to-beat timing distances are typically standardized across the entire sequence of beats. Thus, magnitude of time becomes less meaningful but the overall amount of fiducial shifts become more impactful with respect to altering entropy measurements. Thus, higher sampling rates are more sensitivity to corrupting such entropy measurements.

### 3.2. RQ2 — Type of entropy and precision

Due to the fiducial point of the R-peak shifting, we address how precision (i.e., standard deviation, $\sigma$) can alter the proposed three types of entropy calculations that measure HRC. Tables 6 and 7 are generated

from Part B of the Simulation discussed in Fig. 2 that develops distributions based on the percent change for their respective entropy measurement, SNR, Time Segment, and type of noise. To address RQ2 we can first visually compare the standard deviations of Tables 6 and 7 where for each row in the table, all the variables (e.g., SNR, Time Segments) are held constant except for the type of entropy measurement (e.g., sample entropy). We demonstrate that the standard deviation of the permutation entropy is respectively lower than either approximate or sample entropy across all SNR levels, time segments, and type of noise. Thus from the Monte Carlo simulation, we demonstrated that permutation entropy provides a more precise HRC measurement than either approximate or sample entropy. This finding is likely due to the manner in which permutation entropy is calculated versus sample and approximate entropy. Permutation entropy analyzes HRV by rank order of the R-R interval timing rather than their distance criteria like sample and approximate entropy. Outside of false positives and missed detections of the QRS wave, corruption of the fiducial shift and the timing of QRS wave indices has a stronger effect on distance than on the rank order.

For both white and pink noise, the one-way ANOVA indicated that there were significant differences in the mean log standard deviation of the percent error of the three entropy measures ($F(2,31575) = 12440$, $p < 0.05$ and $F(2,30957) = 11156$, $p < 0.05$ for white and pink noise respectively). Post-hoc comparisons using the Tukey-Kramer procedure revealed that all entropy methods produced significantly different mean log standard deviations of the percent error ($p < 0.05$) for both white and pink noise with permutation entropy having the lowest mean log standard deviation of the percent error followed by approximate entropy and then sample entropy.

### 3.3. RQ3 — Noisy time segments and HRC variance

Increased time segments provide more data which should aid in reducing the aleatory uncertainty for entropy measurements. Therefore we hypothesize that increased time segment lengths at a fixed SNR value would result in more stable and precise HRC measurements. Through the proposed Monte Carlo Simulation (Part B) in Fig. 2, the SNR values were kept fixed but the time windows varied by 2, 5 and 10 min for pink and white noise. By examining Tables 6 and 7 at their respective SNR levels, we note that the standard deviation and expectation decreases as the window size increases for both white and pink noise.

Based on the Monte Carlo simulated samples, a pictorial representation of the distribution was created using kernel density estimation for white noise SNR levels of 2 and 20 at 2 and 10 min segments shown in Figs. 4 and 3. Fig. 4 visually demonstrated Sample Entropy

**Table 6**
Pink noise entropy percent change distribution: (Trials = $\sim$ 65,000).

| Time | Target | Approximate Entropy | | Permutation Entropy | | Sample Entropy | |
|---|---|---|---|---|---|---|---|
| Mins | SNR | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ |
| 2 | 2 | 0.34 | 5.90 | 1.83 | 1.38 | 7.68 | 9.23 |
| 2 | 6 | 0.27 | 5.14 | 1.50 | 1.23 | 5.89 | 9.10 |
| 2 | 10 | 0.34 | 3.95 | 1.05 | 1.01 | 4.02 | 6.25 |
| 2 | 14 | 0.20 | 3.12 | 0.68 | 0.84 | 2.48 | 4.96 |
| 2 | 18 | 0.04 | 2.45 | 0.47 | 0.68 | 1.65 | 4.31 |
| 5 | 2 | 1.52 | 4.39 | 1.65 | 0.93 | 5.50 | 7.85 |
| 5 | 6 | 1.18 | 3.89 | 1.35 | 0.81 | 4.25 | 7.16 |
| 5 | 10 | 0.82 | 2.82 | 0.93 | 0.64 | 2.97 | 4.98 |
| 5 | 14 | 0.53 | 1.98 | 0.62 | 0.50 | 1.95 | 3.32 |
| 5 | 18 | 0.36 | 1.49 | 0.42 | 0.39 | 1.28 | 2.50 |
| 10 | 2 | 4.78 | 3.22 | 1.61 | 0.59 | 6.99 | 2.50 |
| 10 | 6 | 3.77 | 2.63 | 1.32 | 0.52 | 5.45 | 2.09 |
| 10 | 10 | 2.50 | 1.88 | 0.90 | 0.40 | 3.55 | 1.66 |
| 10 | 14 | 1.64 | 1.33 | 0.60 | 0.32 | 2.31 | 1.29 |
| 10 | 18 | 1.09 | 1.04 | 0.40 | 0.26 | 1.50 | 1.05 |

**Table 7**
White noise entropy percent change distribution: (Trials = ~ 65,000).

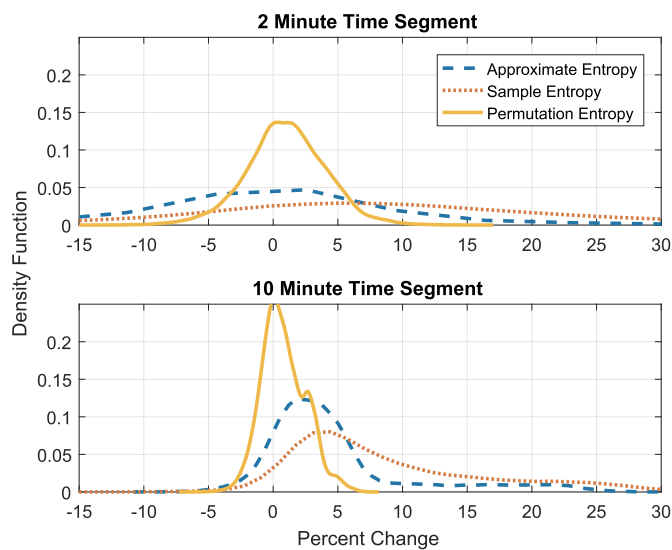| Time | Target | Approximate Entropy | | Permutation Entropy | | Sample Entropy | |
|------|--------|------|------|------|------|------|------|
| Mins | SNR | μ | σ | μ | σ | μ | σ |
| 2 | 2 | − 0.34 | 11.21 | 3.45 | 2.07 | 18.84 | 14.63 |
| 2 | 6 | 0.21 | 8.35 | 2.75 | 1.74 | 12.92 | 11.75 |
| 2 | 10 | 0.40 | 6.29 | 1.98 | 1.46 | 8.65 | 9.84 |
| 2 | 14 | 0.32 | 4.78 | 1.39 | 1.19 | 5.71 | 7.84 |
| 2 | 18 | 0.26 | 3.77 | 0.94 | 0.98 | 3.63 | 6.64 |
| 5 | 2 | 3.31 | 8.26 | 3.21 | 1.65 | 14.24 | 13.38 |
| 5 | 6 | 2.46 | 6.41 | 2.57 | 1.32 | 9.50 | 11.54 |
| 5 | 10 | 1.72 | 4.96 | 1.88 | 1.04 | 6.42 | 9.03 |
| 5 | 14 | 1.15 | 3.73 | 1.33 | 0.80 | 4.24 | 6.78 |
| 5 | 18 | 0.85 | 2.52 | 0.89 | 0.62 | 2.97 | 4.20 |
| 10 | 2 | 12.7 | 8.07 | 3.23 | 1.09 | 19.20 | 5.24 |
| 10 | 6 | 8.54 | 5.58 | 2.59 | 0.86 | 13.10 | 3.87 |
| 10 | 10 | 5.75 | 3.80 | 1.89 | 0.66 | 8.50 | 2.82 |
| 10 | 14 | 3.80 | 2.69 | 1.32 | 0.51 | 5.49 | 2.18 |
| 10 | 18 | 2.48 | 1.90 | 0.90 | 0.40 | 3.54 | 1.66 |



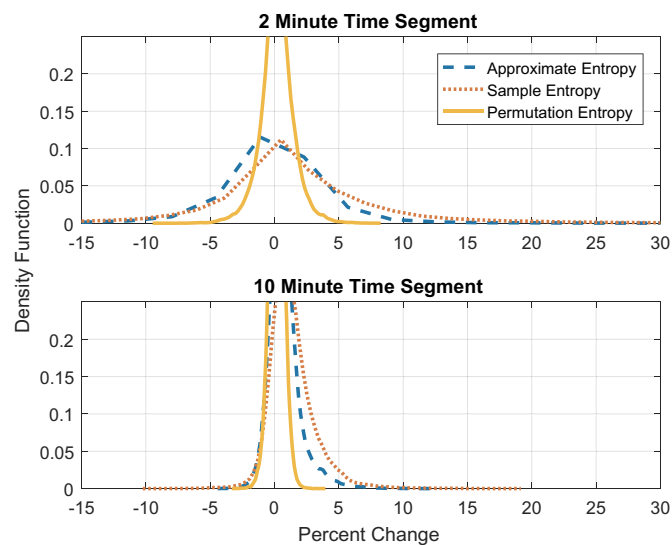**Fig. 3.** Percent change entropy distribution (SNR = 20 db, white noise).



**Fig. 4.** Percent change entropy distribution (SNR = 2 db, white noise).

extending beyond a 30% change from its original entropy calculation. However, as the time segment increases we can note the mass of the distribution becomes more centralized, thus increasing the precision of the entropy measurement. This effect of increasing the windowing segment length is consistent for all three entropy calculation methods. In regards to real-time health monitoring, we are met with a trade-off between time resolution and the precision of the entropy measurement. For example, as we aim to characterize a subject's physiology within a smaller time frame, the precision of the measurement worsens. Thus, when designing these systems, one should consider the level of imprecision which can exist before unacceptable Type 1 or Type 2 errors are encountered in the system.

### 3.4. RQ4 — HRC sensitivity insights

Although some HRC measurements may not offer ideal precision or stability (e.g., Sample and Approximate Entropy), we hypothesize that this lack of precision is to due higher sensitivity in distinguishing features between two cohorts. Therefore the increased variation in the measurement, which results in a decrease in precision may provide more information, may aid in demonstrating independence between two classes or distributions when noise is not present.

Table 8 provides the p-value for the Wilcoxon Rank Sum Test for the three proposed entropy measurements at 2 and 5 min, which evaluates the hypoxic and non-hypoxic cohorts of the presented NASA experiment. In Table 8, we note that permutation entropy had a lower p-Value for both sets and for all time segments than approximate and sample entropy, demonstrating independence between two classes or distributions. Based on the results and discussion of RQ3, permutation entropy was also shown to be the most precise measurement as compared to the other proposed entropy measurements. Thus, these results do not fully support our hypothesis that these imprecise measurements are more sensitive to noise but are better at distinguishing independence. Hence, there is no trade-off between increased precision and the ability to distinguish the two distributions of hypoxia and non-hypoxia.

**Table 8**
Significance level (p-value) of Wilcoxon Rank Sum test between hypoxic and non-hypoxic cohorts.

| Time | Permutation | Approximate | Sample |
|------|-------------|-------------|--------|
| 2 min | **p = 0. 0038** | $p = 0.3445$ | **p = 0. 0233** |
| 5 min | **p = 0. 0045** | $p = 0.1229$ | **p = 0. 0121** |

## 3.5. RQ5 — Directional changes in entropy

As noise alters the fiducial point, we aim to understand how this affects the direction of the entropy measurements. For example, as noise increases and alters the fiducial point, does the measured entropy increase, decrease, or is there no apparent trend (symmetric)? We hypothesized that increased perturbations of the fiducial point from increased noise would make the HRV signal less predictive, thus increasing the entropy measured in the signal. The results provided in RQ2 and RQ3, which indicate a general positive shift in the reported expectation ($\mu$) of the calculated percent change, show evidence to support this hypothesis. This positive shift is also demonstrated by Figs. 4 and 3, where the percent change of the distributions is skewed further to the right as SNR increases.

However, in order to properly support our hypothesis RQ5, we provide the raw calculated values which were averaged over subjects and trials. More specifically, Table 1 provides the mean value of these distributions specifically associated to either the hypoxic ($H_\mu$) or non-hypoxic ($NH_\mu$) case across SNR levels and time windows. We can note a consistent linear increase of mean entropy as noise increases from the initial uncorrupted dataset for all time segments and entropy measurements. This finding is critical since healthier individuals (i.e., an individual with a lower sympathetic response) have higher entropy measurements which produce more complex heart rate dynamics and higher variability. This informs us that if corruption occurs and the fiducial point is altered, subjects will appear healthier than they actually are.

## 3.6. RQ6 — Indistinguishable independent distributions

We demonstrated in RQ5 that there is a clear directional change in the measured entropy when white noise is introduced into the signal. Because the noise does not alter the measured entropy distribution symmetrically (e.g., Fig. 4) and thus shifts the entire distribution in one direction, we must then inquire about the implications of how this statistically affects our readings and to what degree it affects statistical independence. For example, if a subject has a lower entropy measurement, we perceive them to have a higher amount of physiological stress (i.e., increased sympathetic response). However if noise is introduced into the signal, causing an increase in entropy, what is the likelihood that the subject would statistically be viewed as someone not under hypoxic stress?

Since hypoxia is known to cause an increase in sympathetic response [33], we address this research question through an approach similar to that of RQ4 using the NASA hypoxia dataset. We hypothesized that as SNR increases, it becomes increasingly difficult to discriminate ($\alpha = 0.05$) between the two cohorts of hypoxic and non-hypoxic HRC measurements. The simulation was iterated 250 times for each window time segment, cohort, type of entropy measurement, and SNR for their respective cohorts, hypoxia and non-hypoxia. These simulations in Table 9, pink noise was only added to the hypoxia cohort data. Pink noise was utilized rather than white noise, because the magnitude of corruption is less severe demonstrate in the prior Tables 4 and 5. However, in Table 10 the simulation added noise to both the hypoxia and non-hypoxia cohorts. The percentage of statistical tests that were significant from the 250 simulations (% *Signif*.) are shown in each column of Tables 9 and 10.

This percentage is intended to provide contextual meaning of the likelihood that the two cohorts are statistically independent as a function of SNR, windowed time segment and type of entropy measurement. From Table 9, we can observe that there is a much smaller percentage of statistical independence when noise is only introduced to the hypoxia cohort. We already know from Table 8, that when no noise is introduced, that permutation entropy for a 2 min time segment is significant, ($p = 0.0038$). However, when noise only corrupts a 2 min segment of hypoxic data with an SNR of 10, we expect the two classes to

**Table 9**
Evaluating the uncertainty in indistinguishable independent distributions when pink noise enters the hypoxia cohort: (Fs = 256 Hz, Trials = ~ 250).

| Time | Target | Approx. Entropy | Perm. Entropy | Sample Entropy |
|------|--------|-----------------|---------------|----------------|
| Mins | SNR | %Signif. | %Signif. | %Signif. |
| 2 | 2 | 0% | 0% | 0% |
| 2 | 6 | 0% | 0.% | 0% |
| 2 | 10 | 0% | 14.8% | 9.2% |
| 2 | 14 | 0% | 86.4% | 47.6% |
| 2 | 18 | 0% | 99.2.0% | 85.2% |
| 5 | 2 | 0% | 0% | 0% |
| 5 | 6 | 0% | 0% | 0% |
| 5 | 10 | 0% | 24.4% | 76.6% |
| 5 | 14 | 0% | 99.6% | 100% |
| 5 | 18 | 0% | 100% | 100% |

**Table 10**
Evaluating the uncertainty in indistinguishable independent distributions when noise enters both cohorts: (Fs = 256 Hz, Trials = ~ 250).

| Time | Target | Approx. Entropy | Perm. Entropy | Sample Entropy |
|------|--------|-----------------|---------------|----------------|
| Mins | SNR | %Signif. | %Signif. | %Signif. |
| 2 | 2 | 0% | 39.6% | 25.2% |
| 2 | 6 | 0% | 72.4% | 41.6% |
| 2 | 10 | 0% | 96.0% | 69.2% |
| 2 | 14 | 0% | 99.6% | 86.8% |
| 2 | 18 | 0% | 100% | 92.0% |
| 5 | 2 | 0% | 50.8% | 100% |
| 5 | 6 | 0% | 86.8% | 100% |
| 5 | 10 | 0% | 100% | 100% |
| 5 | 14 | 0% | 100% | 100% |
| 5 | 18 | 0% | 100% | 100% |

be independent only 14.8%. This means that out of the 250 simulations that we ran, only 37 generated a p-value less than 0.05. Similarly, in Table 10, where both hypoxic and non-hypoxic cohorts are corrupted, 240 out of the 250 simulations generated a p-value less than 0.05. Thus, we expect to achieve significance 96.0% of the time when both signals are corrupted. This discrepancy between how noises enters the two cohorts is directly related to the impact on how the noise shifts the means of the distributions. Hypoxia readings have a lower entropy measurement, since the subject is under stress. However if noise is present strictly for the hypoxia data, it biases the distribution into making the collected data look as if the subject is not under stress and thus non-hypoxic. Therefore, it becomes more difficult to statistically distinguish between the two cohorts. Machine learning approaches specifically designed to detect anomalies in the data, such as single class Support Vector Machines (SVMs) [34,35], are designed to learn the decision boundaries strictly on a single class. Thus as incoming data is altered by noise and the distribution is being shifted outside the classifier boundary, the algorithm will classify it as an anomaly causing a false positive. This finding is paramount if we know that hypoxia HRV signals are being corrupted, and can be combated by introducing noise into the control cohort (i.e., non-hypoxia). This injection of noise can appropriately adjust the results of our statistical inference (re-train our ML algorithm for more generalized classifier boundaries to incorporate various noise levels) or we can simply bias the distribution based on your incoming SNR. It is also worth noting that although the parameter $r = .20\sigma$ was used for the simulations in Tables 9 and 10, we explored $r = .15\sigma$ as the distance metric for similarity for both sample and approximate entropy. We found that although both parameters were significant prior to signal corruption, the more stringent criteria of $r = .15\sigma$ performed dramatically worse than $r = .20\sigma$. This finding suggests that as noise is introduced these distance criteria should be loosened to achieve independence.

## 4. Conclusion

Current literature about handling HRC calculations focuses on reducing the false positive and false negative rates of detecting the QRS waveform in the ECG signal. The location of the R-peak is typically extrapolated from the detected QRS waveform and little attention is given to how errors regarding the location of the R-peak can alter HRC readings. In this paper we present a Monte Carlo simulation framework that evaluates the effects of ECG signal corruption on the fiducial point of the R-peak and how it effects HRC measurements when using sample, approximate, and permutation entropy.

Through our findings with Monte Carlo simulations, we are able to characterize PMF distributions and how the fiducial point shifts based on signal quality of the ECG. White noise was shown to cause higher perturbations in the fiducial point of the R-peak when compared to pink noise. This characterization allowed us to run additional Monte Carlo trials in order to evaluate changes in the precision of the proposed entropy measurement, in which permutation entropy is demonstrated to be most precise during corruption. From these findings, we utilized a secondary data set that addresses the sensitivity of demonstrating a statistical difference between hypoxia vs non-hypoxia caused by altered heart rate dynamics from the autonomic nervous system. This analysis showed that permutation entropy not only had better precision under noisy environments but was also sensitive statistically for 2 and 5 min time segments between the two cohorts. Where as, Approximate entropy was not significant for either 5 or 10 min and sample entropy was only significant for 5 min time segments. This work then demonstrated that as perturbations of the R-peak increased, as a function decreasing SNR, the entropy of the signal increased. We demonstrated that corrupted ECG signals entropy calculations have the potential to have biased means. However, permutation entropy showed to have a stronger precision and sensitivity which was able to still out perform sample and approximate entropy. Thus, sample and approximate entropy have a greater likelihood of showing inaccurate statistical changes in heart dynamics during hypoxia for 2 min segments.

From the perspective of system and experimental design, these findings are specifically pivotal for mobile ECG devices in which HRC calculations are performed and have a higher likelihood of corruption. We recommend that device developers fully characterize the level of noise and type of noise that can be introduced into an ECG system since the statistical inferences can be drastically altered. If researchers are not vigilant of the level of corruption and statistical inference goals of the work, the implications can be catastrophic. For example, in real-time health care systems that utilized HRC as a discriminant feature to detect physiological complications and severe corruption was present, the patient's calculated HRC could result in misinformed diagnoses of health failure. This data would suggest that the patient is healthier than normal. However, noise does not have to be irreparably detrimental. But the system or experimental designer, needs to maintain an equivalent and consistent level of SNR though the entire data collection process. This countermeasure will insure the robustness of the statistical inferences, as demonstrated in RQ6. If SNR cannot be regulated in this manner, researchers should at least be aware of how and when noise is injected into the data. This improved vigilance to noise would allow researchers to adjust for the shift in the mean HRC measurement (demonstrated in RQ5) to reduce the risk of such mis-classifications.

From the physiological perspective, our study affirms that during hypoxia HRC dynamics decrease. Although there is strong evidence that inducing hypoxia causes sympathetic arousal [33], it is fully not clear what primarily drives the decrease in HRC during hypoxia. Within the HRV community, there is a strong consensus that HRV is driven by the vagal nerve, possibly due to it directly interfacing with the parasympathetic system. However, there are many physiological systems in the body that can influence HRV in which other bodies of work point to how sympathetic response can alter HRV [36–38]. Within the NASA hypoxia experiment, vagal tone can be influenced by its peripheral chemoreceptors, induced by the changes in oxygen and carbon dioxide in blood. Thus, we can imagine that this is likely not a univariate problem in which parasympathetic influence alone drives HRC. However, the core of this work is to understand and characterize how small noise-induced perturbations in the HRV signal can influence our physiological interpretation of results and their potentially catastrophic implications for real-time health monitoring systems. From the analytics and computing aspect, we should examine more robust methods to inform us of autonomic changes in body. As we begin to introduce more low-power wireless sensor systems for long term physiological monitoring where data corruption is inevitable and easily introduced, these methods will become paramount. Permutation entropy is a prime example of a more robust and informative analytic approach that demonstrates the actual measurable distance between beats is not fundamentally critical for differentiating the two cohorts. However, the order and direction of how the beat intervals change by the HR slowing down or speeding up are key characteristics for quantifying autonomic nervous response. Since permutation entropy is not reliant on precise measurements of the timing distance of the beat-to-beat intervals, it is innately more robust to noise by examining the structure of the HRV time series. Future work should continue to explore the structural components of HRV and ECG signals that may not require precise measurements. Multiscale entropy (MSE), in which the HRV structure is analyzed by averaging the HRV signal over time, may provide an even more robust method for overcoming ECG signal noise. Thus, this work sets the foundation on how noise can critically influence HRC calculations and discusses the downstream implications it can have on statistical modeling techniques for physiological insight. Furthermore, we addressed fundamental design questions, allowing researchers to evaluate the merits of different entropy measurements, suggestions on how to handle ECG fiducial point corruption, and ideal time window segments based on the type of environment imposed on their telemetry system and applications.

## References

[1] G. Berntson, J. Stowell, Ecg artifacts and heart period variability: don't miss a beat!, Psychophysiology 35 (1) (1998) 127–132.

[2] S. Laborde, E. Mosley, J.F. Thayer, Heart rate variability and cardiac vagal tone in psychophysiological research " recommendations for experiment planning, data analysis, and data reporting, Front. Psychol. 8 (2017) 189–197.

[3] C. Schubert, M. Lambertz, R. Nelesen, W. Bardwell, J.-B. Choi, J. Dimsdale, Effects of Stress on Heart Rate Complexity a Comparison Between Short-Term and Chronic Stress vol. 80, (2008), pp. 325–332.

[4] N.T. Liu, L.C. Cancio, J. Salinas, A.I. Batchinsky, Reliable real-time calculation of heart-rate complexity in critically ill patients using multiple noisy waveform sources, J. Clin. Monit. Comput. 28 (2) (2014) 123–131.

[5] N. Napoli, L. Barnes, A Dempster-Shafer approach for corrupted electrocardiograms signals, Twenty-ninth International Florida Artificial Intelligence Research Society Conference, 2016.

[6] A.H. Khandoker, H.F. Jelinek, M. Palaniswami, Identifying diabetic patients with cardiac autonomic neuropathy by heart rate complexity analysis, Biomed. Eng. Online 8 (2009) 3.

[7] J. Sztajzel, Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system, Swiss Med. Wkly. 134 (35–36) (2004) 514–522.

[8] Y. Zhong, K.-M. Jan, K.H. Ju, K.H. Chon, Quantifying cardiac sympathetic and parasympathetic nervous activities using principal dynamic modes analysis of heart rate variability, Am. J. Physiol. Heart Circ. Physiol. 291 (3) (2006) H1475–H1483.

[9] S.J.-J. Leistedt, P. Linkowski, J. Lanquart, J.E. Mietus, R. Davis, A.L. Goldberger, M. Costa, Decreased neuroautonomic complexity in men during an acute major depressive episode: analysis of heart rate dynamics, Transl. Psychiatry 8 (2011) 1–7.

[10] Q. Zheng, C. Chen, Z. Li, A. Huang, B. Jiao, X. Duan, L. Xie, A novel multi-resolution svm (mr-svm) algorithm to detect ecg signal anomaly in we-care project, 2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC), 2013, pp. 1–6.

[11] A. Ukil, S. Bandyopadhyay, C. Puri, A. Pal, Heart-trend: an affordable heart condition monitoring system exploiting morphological pattern, 2016 IEEE

International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 6260–6264.

[12] G. Clifford, W. Long, G. Moody, P. Szolovits, Robust parameter extraction for decision support using multimodal intensive care data, Phil. Trans. Royal Soc. Lond. A: Math. Phys. Eng. Sci. 367 (1887) 411–429 2009.

[13] N.T. Liu, A.I. Batchinsky, L.C. Cancio, J. Salinas, The impact of noise on the reliability of heart-rate variability and complexity analysis in trauma patient, Comput. Biol. Med. 43 (11) (2013) 1955–1964.

[14] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215.

[15] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am. J. Physio. Heart Circ. Physio. 278 (6) (2000) 2039–2049.

[16] S.M. Pincus, Approximate entropy as a measure of system complexity, Proc. Natl. Acad. Sci. Unit. States Am. 88 (6) (1991) 2297–2301.

[17] M. Riedl, A. Müller, N. Wessel, Practical considerations of permutation entropy, Eur. Phys. J. Spec. Top. 222 (2) (2013) 249–262.

[18] M. Costa, A. L. Goldberger, C.-K. Peng, Multiscale entropy analysis of complex physiologic time series, Phys. Rev. Lett. 89.

[19] K. Kim, H. Baek, Y. Lim, K. Park, Effect of missing rr-interval data on nonlinear heart rate variability analysis, Comput. Methods Progr. Biomed. 106 (3) (2012) 210–218.

[20] R.J. Ellis, B. Zhu, J. Koenig, J.F. Thayer, Y. Wang, A careful look at ecg sampling frequency and r-peak interpolation on short-term measures of heart rate variability, Physiol. Meas. 36 (9) (2015) 1827.

[21] K.K. Kim, Y.G. Lim, J.S. Kim, K.S. Park, Effect of missing rr-interval data on heart rate variability analysis in the time domain, Physiol. Meas. 28 (12) (2007) 1485.

[22] K. Kim, J. Kim, Y. Lim, K. Park, The effect of missing rr-interval data on heart rate variability analysis in the frequency domain, Physiol. Meas. 30 (10) (2009) 210–218.

[23] F. Gustafsson, Determining the initial states in forward-backward filtering, IEEE Trans. Signal Process. 44 (4) (1996) 988–992.

[24] C. Chew, E. Zahedi, Limb cardiovasulature system identification using adaptive filtering, Int. Fed. Med. Biol. Proc. (2006) 406–409.

[25] H. Limaye, V. Deshmukh, Ecg noise sources and various noise removal techniques: a survey, Int. J. Appl. Innovation Engin. Manag. 5 (2016) 406–409.

[26] G.D. Clifford, F. Azuaje, P. McSharry, Ecg statistics, noise, artifacts, and missing data, Advanced Methods and Tools for ECG Data Analysis, Artech House, Inc., Norwood, MA, USA, 2006, pp. 70–71 Ch. 3.

[27] N.J. Kasdin, Discrete simulation of colored noise and stochastic processes and $1/f$ alpha; power law noise generation, Proc. IEEE 83 (5) (1995) 802–827.

[28] J. Pan, W.J. Tompkins, A real-time QRS detection algorithm, IEEE Trans. Biomed. Engin. BME 32 (3) (1985) 230–236.

[29] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series, Phys. Rev. Lett. 88 (17).

[30] D. Gradwell, Hypoxia and hyperventilation, in: D. Gradwell, D.J. Rainford (Eds.), Ernstings's Aviation Med, fourth ed., CRC Press, 2006Ch. 3.

[31] C. Stephens, K. Kennedy, B. Crook, R. Williams, M. Last, P. Schutte, Mild normobaric hypoxia exposure for human-autonomy system testing, Proc. of the Human Factors and Ergonomics Society Annual Meeting 61.

[32] C. Stephens, K. Kennedy, N. Napoli, M. Demas, B. Crook, R. Williams, M. Last, P. Schutte, Effects on task performance and psychophysiological measures of performance during normobaric hypoxia exposure, Proc. of the Int. Symp. on Aviation Psychology Dayton, OH.

[33] R. Hainsworth, M.J. Drinkhill, M. Rivera-Chira, The autonomic nervous system at high altitude, Clin. Auton. Res. 17 (2007) 1319.

[34] J. Su, Y. Long, X. Qiu, S. Li, D. Liu, Anomaly Detection of Single Sensors Using OCSVM_KNN, Springer Int.l Pub., Cham, 2015, pp. 217–230.

[35] Y. Chen, J. Qian, V. Saligrama, A new one-class svm for anomaly detection, 2013 IEEE Int. Conf. on Acoustics, Speech and Signal Proc., 2013, pp. 3567–3571.

[36] L. E. Virgilio-Silva, C. A. Aguiar-Silva, H. C. Salgado, R. Fazan, The role of sympathetic and vagal cardiac control on complexity of heart rate dynamics, Am. J. of Physio Heart and Circ. 32.

[37] M. Weippert, M. Behrens, A. Rieger, K. Behrens, Sample entropy and traditional measures of heart rate dynamics reveal different modes of cardiovascular control during low intensity exercise, Entropy 16 (11) (2014) 5698–5711.

[38] A.E. Draghici, J.A. Taylor, The physiological basis and measurement of heart rate variability in humans, J. Physiol. Anthropol. 35 (2016) 1–8.