

Scaling of Algorithmic Bias in Pulse Oximetry with Signal-to-Noise Ratio

Jiaming Cao¹, Neil Ashim Mehta¹, Jingyi Wu², Sossena Wood², Jana M. Kainerstorfer², Pulkit Grover¹

Abstract—Recent work has noted a skin-color bias in existing pulse oximetry systems in their estimation of arterial oxygen saturation. Frequently, the algorithm used by these systems estimate a “ratio-of-ratios”, called the “ R -value”, on their way to estimating the oxygen saturation. In this work, we focus on an “SNR-related” bias that is due to noise in measurements. We derive expressions for the SNR-related bias in R -value estimation, and observe how it scales with the signal-to-noise ratio (SNR). We show that the bias can arise at two steps of R -value estimation: in estimating the max and min of a pulsatile signal, and, additionally in taking ratios to estimate the R -value. We assess the bias resulting from the combination of the two steps, but also separate out contributions of each step. By doing so, we deduce that the bias induced in max and min estimation is likely to dominate. Because the SNR tends to get worse with higher melanin concentration, our result provides a sense of scaling of this bias with melanin concentration.

I. INTRODUCTION

Pulse oximetry (PulseOx) is based on the technique of photoplethysmography (PPG) which optically estimates blood volume changes in tissue due to cardiac pulsations. PulseOx estimates arterial blood oxygen saturation non-invasively. It is used extensively to monitor patients’ health at home and in the clinics and ICUs. This paper is motivated by bias in PulseOx estimates when recording on patients with darker skin tones (e.g. [1], [2], [3], [4]). This bias is extremely unfortunate: PulseOx is a critical tool in deciding the need for a ventilator after acute respiratory distress syndrome (ARDS), e.g., due to Covid-19 or other respiratory illnesses.

Melanin, a high-absorbing and high-scattering skin pigment [5], is a known hindrance to PulseOx. Commonly, PulseOx systems inject light at 2 wavelengths into the tissue and measure the received light intensities. These are used to estimate arterial blood oxygenation by noting that the arterial component is the only one that is “pulsatile”, i.e., changes with the cardiac pulse. PulseOx systems output the “SpO₂ value”, i.e., the PulseOx estimate of the ground truth “SaO₂ value”, i.e., the ground truth arterial oxygenation (which can be obtained, e.g., through blood-gas measurements). The classic algorithm for obtaining the SpO₂ value (see, e.g. [6]) separates $I_{AC,i}$, the pulsatile “AC-component” (caused by

the cardiac pulsation), and $I_{DC,i}$, the non-pulsatile DC component (baseline intensity of the measured light), at the two wavelengths ($i = 1, 2$). Next, the ratio: $R = \frac{I_{AC1}/I_{DC1}}{I_{AC2}/I_{DC2}}$, called the “ R -value”, is computed. The AC components are normalized with the DC component at the respective wavelength (hence the term “ratio of ratios” for the R -value). This R -value is subsequently plugged into a “calibration curve”, whose output is the SpO₂ estimate. The division by the DC-component is intended to normalize for variations, including the effects of melanin (as explained through a derivation in [6]). However, as is evident from several classic [2], [3], [4] as well as recent works [1], [7], [8], [9], melanin-related bias is still substantial.

Why does this bias arise? Rigorous simulation analyses [10] show that, if melanin concentration is not accounted for, bias arises because the paths that the photons take are altered due to scattering by melanin particles. This altering of their paths affects the relationship between the received intensity of light and the transmitted intensity, not captured by a calibration curve that is the same for all participants. Consequently, calibration curves are biased, even when no measurement noise is present. We call this bias the “noiseless bias”. Our work identifies another source of bias in PulseOx estimates: an “SNR-related” bias, that arises due to sensor noise. In Section II, through theoretical results, we explain why the SNR-related bias arises, and obtain the scaling of bias with the SNRs at the two wavelengths in high-SNR situations. In Section III, we use simulations that add noise to real data to understand this scaling in practical situations.

Intuitively, reduced SNR would contribute to higher *variability* in PulseOx estimates. It might be less obvious why it would contribute to a bias. In Section II, we show that SNR-related bias arises at two steps of R -value estimation: in estimating the AC-amplitudes by computing max and min of a pulsatile signal, and, additionally in taking ratios to estimate the R -value. We assess the bias resulting from the combination of the two steps, but also separate out the contributions of each step. By doing so, we show that the bias induced in max and min estimation is likely to dominate at high SNRs, but is present only when the SNRs at the two wavelengths are substantially different. This bias term rises roughly as $\frac{1}{\sqrt{\text{SNR}}}$ (see Eq. (5) for the exact expression). Our simulations in Section III confirm these predictions: the bias rises roughly linearly with noise standard deviation (i.e., as $\frac{1}{\sqrt{\text{SNR}}}$), and, additionally, the bias is small when the SNRs at the two wavelengths are equal.

How does the SNR-related bias connect with skin color? As

This work was supported by the “2021 Engineering Approaches to Responsible Neural Interface Design” Award by Facebook Reality Labs.

¹Jiaming Cao, Neil Ashim Mehta, and Pulkit Grover are with Department of Electrical and Computer Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 15213, USA. {jiamingc, neilashm, pgrover}@andrew.cmu.edu

²Jingyi Wu, Sossena Wood, Jana M. Kainerstorfer are with Department of Biomedical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 15213, USA. {jingyiwu, scwood, jkainers}@andrew.cmu.edu

noted, melanin is highly absorbing and scattering [5], and causes different signal attenuation at different wavelengths [11]. Indeed, simulation analyses in [12] show that the SNR is lower for participants with higher melanin. Interestingly, even seemingly small biases in R -value estimates, induced e.g. at high SNRs, can induce fairly high biases in SpO₂ estimates. To see this, note that the calibration curve is approximately linear (with a negative slope), as long as SpO₂ is only varied by a few percentage points. Thus, with a change in sign (due to the negative slope), the bias in an R -value estimate is approximately linearly proportional to the bias in the respective SpO₂ estimate. To understand this slope, in typical calibration curves (e.g. see [13]), a small bias of ≈ 0.03 in the R -value estimate can affect a 1% bias in the SpO₂ estimate, which is the average bias observed in PulseOx for darker skinned participants in [8].

II. HOW DOES BIAS SCALE WITH SNR? A THEORETICAL RESULT

As discussed, we focus on the classic algorithm that is used in many commercial systems [14]. For simplicity of analysis, we assume that the DC components are estimated without bias or error, and thus they are not used in our analysis. In simulations in Section III, we also estimate the DC components, and show that the scaling of the bias in R -value estimates with SNR remains qualitatively similar even with the inclusion of the DC component.

First, in Section II-A, we introduce preliminary results, which we use subsequently in Section II-B to assess the bias and its scaling with SNR, where the assumptions in bias assessment are clearly stated.

A. Lemmas that help estimate the bias

Lemma 1 (Max/min bias): For uniform, independent and identically distributed (iid) noise, $Z_i \sim \mathbb{U}[-\Delta, \Delta]$, the estimate $\hat{Z}_{\max} = \max_i \{Z_i\}$ for $i = 1, \dots, n$ has the bias:

$$\text{Bias}(\hat{Z}_{\max}) = \frac{(n-1)\Delta}{n+1}, \quad (1)$$

and variance

$$\text{Var}(\hat{Z}_{\max}) = \frac{4n\Delta^2}{(n+1)^2(n+2)}. \quad (2)$$

Proof: We will now evaluate the bias for $\hat{X}_{\max} = \max_i \{X + Z_i\}$, where $X > \Delta$ is a constant. Since the computed bias does not depend on X , post computation of the bias, we will simply subtract X from $\mathbb{E}[\hat{X}_{\max}]$ to arrive at the bias in $\hat{Z}_{\max}$.

Define $F_{\max}(t) := \Pr(\hat{X}_{\max} \leq t)$, the cumulative distribution function (cdf) of $\hat{X}_{\max}$. Then,

$$\begin{aligned} \mathbb{E}[\hat{X}_{\max}] &= \int_0^\infty (1 - F_{\max}(t)) dt \\ &= X + \Delta - \frac{2\Delta}{n+1}, \end{aligned}$$

Finally, after subtracting X , the bias in $\hat{Z}_{\max}$ is:

$$\text{Bias} = \Delta - \frac{2\Delta}{n+1} = \frac{n-1}{n+1}\Delta. \quad (3)$$

Variance can be calculated similarly. ■

Remarks: For applying Lemmas 1 in practice, one can, for instance, examine the number of time-points at which the (noiseless) signal is relatively constant, and plug in that value of n . Alternatively, for a slowly varying signal, the result can be rederived under assumptions of bounded slope. The result shows that the bias in maximum estimation using the estimate $\hat{X}_{\max}$ increases linearly with Δ (for a fixed n). For $n = 0$, there is no bias, as there is only one sample and that sample itself is the estimate of the maximum. However, this requires sampling the process at the (unknown) time at which the maximum is attained.

Lemma 2 (Ratio-bias with biased numerator, denominator):

Assuming the bias of the estimators of y and x is b_x and b_y , respectively, with independent, mean-0 additive noises z_y and z_x of variances σ_y^2 and σ_x^2 , the resulting ratio estimator $\hat{R} = \frac{\hat{y}}{\hat{x}}$ has a bias:

$$\mathbb{E}[\hat{R}] - R \approx \frac{(b_y - Rb_x)}{x} \left(1 - \frac{b_x}{x}\right) + \frac{R\sigma_x^2}{x^2}. \quad (4)$$

Proof: The derivation builds on the textbook derivation of ratio bias, see e.g. [15, Ch. 4]. The overall bias in $\hat{R}$ is:

$$\begin{aligned} &\mathbb{E}\left[\frac{y + b_y + z_y}{x + b_x + z_x}\right] - R \\ &= \mathbb{E}\left[\frac{y + b_y + z_y - R(x + b_x + z_x)}{x(1 + \frac{b_x + z_x}{x})}\right] \\ &\stackrel{(a)}{\approx} \frac{1}{x} \mathbb{E}[(y + b_y + z_y - Rx - Rb_x - Rz_x)] \\ &= \frac{1}{x} (b_y - Rb_x) \left(1 - \frac{b_x}{x}\right) + \frac{R\sigma_x^2}{x^2}, \end{aligned}$$

where (a) uses the approximation that $\frac{1}{1+\eta} \approx 1 - \eta$ for $\eta \ll 1$. ■

B. Using lemmas to estimate bias and its scaling with SNR in PulseOx estimates

Now, to apply the results in Section II-A to assess bias in commonly used (two-wavelength) PulseOx systems, we make the following assumptions (working backwards from PulseOx calculation to light absorbance in the two wavelengths). First, we assume that the bias in R -value estimate is linearly proportional to the bias in the PulseOx estimate (with an opposite sign). Second, enabled² by Lemma 1, we assume that the estimate $\hat{I}_{AC,j}$ of $I_{AC,j}$, the absorbance in the j -th wavelength, has a bias $b_j = \alpha\Delta_j$ for some $\alpha > 0$, and

¹Here, this approximation holds best when the bias and noise are much smaller than the signal value, which can be thought of as a high SNR situation.

²A good choice of n from Lemma 1 is not known here as it depends on sampling rate and its relation to how fast the signal changes. Instead, the parameters α and γ capture the effect of n because functions of n form multiplicative constants for bias and variance.

variance $\sigma_j^2 = \gamma \Delta_j^2$. We also assume that the noise that caused these biases is independent for the two wavelengths.

These assumptions allow us to use Lemma 2. The assumptions hold under limited conditions. The first assumption holds when the arterial O₂ saturation (SaO₂) only varies in a small interval, and is more accurate when SaO₂ takes a value that is not very high (i.e., below 95%). The second assumption relies on independence of noise, which holds, e.g., if the dominant noise is thermal noise.

Under these assumptions, utilizing Lemma 2, the bias in estimating ratio R (which is approximately negative of bias in SpO₂ estimate) is given by:

$$\begin{aligned} \text{Bias} &\approx \alpha R \left(\frac{\Delta_1}{I_{AC,1}} - \frac{\Delta_2}{I_{AC,2}} \right) \left(1 - \frac{\alpha \Delta_2}{I_{AC,2}} \right) + \frac{R \sigma_2^2}{I_{AC,2}^2} \\ &= \frac{\alpha R}{\sqrt{\gamma}} \left(\frac{1}{\sqrt{\text{SNR}_1}} - \frac{1}{\sqrt{\text{SNR}_2}} \right) \left(1 - \frac{\alpha}{\sqrt{\gamma \text{SNR}_2}} \right) \\ &\quad + \frac{R}{\text{SNR}_2}, \end{aligned} \quad (5)$$

where $\text{SNR}_i = I_{AC,i}^2 / \sigma_i^2 = I_{AC,i}^2 / (\gamma \Delta_i^2)$ and $\sigma_i^2 = \gamma \Delta_i^2$.

Remarks: The first term in (5) scales as $\sqrt{1/\text{SNR}}$, or as the noise standard deviation, if the ratios of the two SNRs are kept constant and the SNRs are increased proportional to each other. This is because the term $(1 - \alpha/\sqrt{\gamma \text{SNR}_2})$ approaches 1 at high SNR. The first term dominates the second term, which scales as $1/\text{SNR}$, at high SNR. However, if the two SNRs are equal, the first term is zero, and the second term dominates (and the bias would also be smaller). This motivates our simulation scenarios in Section III, when we examine how bias scales with SNR for real data corrupted by simulated noise. There, we worsen the SNR while maintaining the noise standard deviations as scalar multiples of each other.

III. SIMULATIONS-BASED ASSESSMENT OF HOW BIAS SCALES WITH SNR

This section uses noise-corrupted real data to study effects of SNRs at the two wavelengths on bias in the estimated R -value. While Section II establishes that a bias is added, and characterizes it theoretically, it makes several assumptions on the signal, and has unknown parameters (α, γ). Further, it ignores the effect of I_{DC} . Simulations in this section thus take us a step closer to practice, and also help us understand if the results obtained in Section II are useful guiding principles in practice.

“Ground truth” data: To create a realistic simulation, we acquired near-infrared spectroscopy (NIRS) data, which is the basis of photoplethysmography (PPG) data [16], from a healthy participant’s finger in transmission mode, smoothened it by filtering, and utilized it as the ground truth data on which noise is added to study the effects of SNR. The data collection and analysis was approved by Carnegie Mellon University’s Institutional Review Board. The data was collected over a 10-minute period with a sampling frequency of 79.47 Hz at 2 wavelengths, 730 nm and 830 nm, indexed by 1 and 2, respectively. An ISS Imagent system (ISS Inc., Champaign,

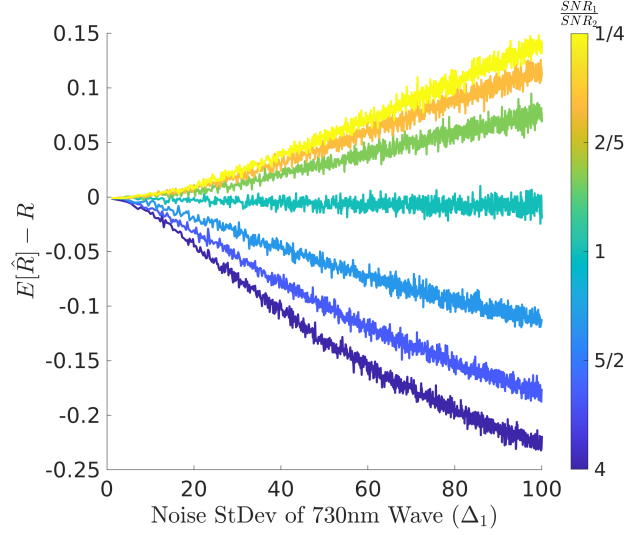


Fig. 1. Estimated bias in R -value for different noise standard-deviations and SNRs. Predictably, as the noise standard-deviation increases (i.e. SNR decreases), the magnitude of bias in R -value estimates increases. The sign of bias depends on the relative SNR ratio between the 2 wavelengths. The bias for matched SNRs is shown in red. Consistent with theoretical results in Section II (specifically, Eq. (5)), the bias is small when the SNRs are equal, and rises linearly with the standard deviation for unequal SNRs when the SNR ratio, $\text{SNR}_1/\text{SNR}_2$, is kept constant.

IL) was used for the data acquisition. Minimum-order FIR lowpass and highpass filters with a commonly used passband (0.66-15 Hz) [17], designed to attain a -60dB attenuation in the stopband, were employed.

For computing a ground-truth R -value, we split the data into non-overlapping 2-second windows over which we estimate the max and min of the signals to approximate the AC amplitude of the j -th wavelength, $I_{AC,j}$. These AC components are normalized by their respective DC components, $I_{DC,j}$, estimated by taking the sample mean of the signal in the 2-second window. The R -value is averaged over all such windows ($m=295$):

$$R = \frac{1}{m} \sum_{i=1}^m \frac{(I_{AC,1}/I_{DC,1})^{(i)}}{(I_{AC,2}/I_{DC,2})^{(i)}}. \quad (6)$$

This value is assumed to represent the true parameter, R , in the bias calculation, $\mathbb{E}[\hat{R}] - R$.

Simulated noisy data: We proceed to simulate the effect of worsening SNR via the addition of uniform noise independent and identically distributed across time at each wavelength. The noise at the two wavelengths is generated independently and the noise variance is varied modeling the effect of worsening SNR, where the SNR of the j -th wavelength is defined as:

$$\text{SNR}_j = \frac{\bar{I}_{AC,j}^2}{\sigma_j^2}, \quad (7)$$

where $\bar{I}_{AC,j}$ is estimated as the difference between the average of all peaks and the average of all valleys of signal j . Since it is, in general, hard to predict which of the two wavelengths

would have higher SNR in practice, we assessed the bias-noise relationship at various SNR_1/SNR_2 ratios, ranging from 1/4 to 4 (shown by the color bar in Fig. 1).

Estimating R-value from noisy data: The filtering and amplitude estimation process from above is performed now on this noisy data, and the R -value estimated is averaged over all 2-second windows to compute the average estimated value:

$$\mathbb{E}[\hat{R}] = \frac{1}{m} \sum_{i=1}^m \frac{(\hat{I}_{AC,1}/\hat{I}_{DC,1})^{(i)}}{(\hat{I}_{AC,2}/\hat{I}_{DC,2})^{(i)}}. \quad (8)$$

Results: Fig. 1 plots the bias resulting from subtracting R from $\mathbb{E}[\hat{R}]$. On the x -axis is σ_1 , the noise standard deviation at 730 nm. The SNR-ratio value (indicated by the color of the curve) determines the noise variance at the 830 nm wavelength. Consistent with (5) in Sec. II-B, at equal SNRs for the two wavelengths, the bias is small. Further, the magnitude of the R -value bias scales approximately linearly with σ_1 (for fixed SNR-ratios), especially so for low and moderate values of σ_1 , again, consistent with (5).

IV. DISCUSSIONS AND CONCLUSION

In this work, we demonstrate both mathematically and using simulations that when the SNR of the PPG signals worsen, the bias of R -value estimation, hence SpO_2 estimation, increases. However, there are nuances: when both channels have similar SNR, the increase in bias magnitude is small. When both channels have very different SNRs, the bias magnitude is large. These results are consistent across theory and simulations. Because darker skinned participants have lower SNR for the same intensity of injected light, This suggests that the SpO_2 readings from highly pigmented subjects can be highly inaccurate and thus bias clinical decisions related to their health.

In practice, our assumptions in Sections II and III might be simplistic (e.g., iid noise). Nevertheless, our results explain when this SNR-related bias can be large, and how knowledge of the underlying data statistics (e.g. through online estimation of signal parameters or characterization of skin melanin concentration) can be used to correct this bias.

For simplicity, we focus for the classic algorithm [6], that estimates R -values on their way to SpO_2 estimation. While this algorithm is commonly used in practice, other algorithms that directly invert the modified Beer-Lambert law could also be employed. Relatedly, some PulseOx systems use multiple (> 2) wavelengths. Assessment of bias for such alternate algorithms and/or multiple wavelengths is left for future work.

What our study reveals is that, in addition to addressing the “noiseless bias”, discussed in Section I (and studied, e.g., in [10]), we also need to address the “SNR-related bias” that is the focus of our work. Removal of this SNR-related bias appears to require estimation of SNR, which most current systems do not perform. True equity might requires a change in hardware as well. E.g., if the injected light intensity is

high enough, these biases will be sufficiently small that they can be tolerated (e.g. $<0.05\%$ in PulseOx estimates).

ACKNOWLEDGMENTS

The authors would like to thank Shidhartho Roy and Lara Abdelmohsen for many useful discussions as well as their kind help with data collection.

REFERENCES

- [1] Michael W Sjoding, Robert P Dickson, Theodore J Iwashyna, Steven E Gay, and Thomas S Valley. Racial bias in pulse oximetry measurement. *New England Journal of Medicine*, 383(25):2477–2478, 2020.
- [2] Philip E Bickler, John R Feiner, and John W Severinghaus. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *The Journal of the American Society of Anesthesiologists*, 102(4):715–719, 2005.
- [3] EB Wassenaar and JGH Van den Brand. Reliability of near-infrared spectroscopy in people with dark skin pigmentation. *Journal of clinical monitoring and computing*, 19:195–199, 2005.
- [4] John R Feiner, John W Severinghaus, and Philip E Bickler. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. *Anesthesia & Analgesia*, 105(6):S18–S23, 2007.
- [5] Weiye Song, Lei Zhang, Steve Ness, and Ji Yi. Wavelength-dependent optical properties of melanosomes in retinal pigmented epithelium and their changes with melanin bleaching: a numerical study. *Biomedical optics express*, 8(9):3966–3980, 2017.
- [6] Hyun Jae Baek, JaeWook Shin, and Jageol Cho. The effect of optical crosstalk on accuracy of reflectance-type pulse oximeter for mobile healthcare. *Journal of healthcare engineering*, 2018.
- [7] Eric Raphael Gottlieb, Jennifer Ziegler, Katharine Morley, Barret Rush, and Leo Anthony Celi. Assessment of racial and ethnic differences in oxygen supplementation among patients in the intensive care unit. *JAMA internal medicine*, 182(8):849–858, 2022.
- [8] Ashraf Fawzy, Tianshi David Wu, Kunbo Wang, Matthew L Robinson, Jad Farha, Amanda Bradke, Sherita H Golden, Yanxun Xu, and Brian T Garibaldi. Racial and ethnic discrepancy in pulse oximetry and delayed identification of treatment eligibility among patients with covid-19. *JAMA internal medicine*, 182(7):730–738, 2022.
- [9] Ana M Cabanas, Macarena Fuentes-Guajardo, Katina Latorre, Dayneri León, and Pilar Martín-Escudero. Skin pigmentation influence on pulse oximetry accuracy: a systematic review and bibliometric analysis. *Sensors*, 22(9):3402, 2022.
- [10] Mohammed Shahriar Arefin, Alexander P Dumont, and Chetan A Patil. Monte carlo based simulations of racial bias in pulse oximetry. In *Design and Quality for Biomedical Technologies XV*, volume 11951, pages 8–12. SPIE, 2022.
- [11] Nikiforos Kollias. The spectroscopy of human melanin pigmentation. *Journal of Investigative Dermatology*, 102(2):268–268, 1994.
- [12] Subhasri Chatterjee and Panayiotis A Kyriacou. Monte carlo analysis of optical interactions in reflectance and transmittance finger photoplethysmography. *Sensors*, 19(4):789, 2019.
- [13] Ali İhsan Bülbül and Serdar Küçük. Pulse oximeter manufacturing & wireless telemetry for ventilation oxygen support. *International Journal of Applied Mathematics Electronics and Computers*, pages 211–215, 2016.
- [14] Edward D Chan, Michael M Chan, and Mallory M Chan. Pulse oximetry: understanding its basic principles facilitates appreciation of its limitations. *Respiratory medicine*, 107(6):789–799, 2013.
- [15] Sharon L Lohr. *Sampling: design and analysis*. Chapman and Hall/CRC, 2021.
- [16] Meir Nitzan, Salman Noach, Elias Tobal, Yair Adar, Yaacov Miller, Eran Shalom, and Shlomo Engelberg. Calibration-free pulse oximetry based on two wavelengths in the infrared—a preliminary study. *sensors*, 14(4):7420–7434, 2014.
- [17] Norbert Stuban and Masatsugu Niwayama. Optimal filter bandwidth for pulse oximetry. *Review of Scientific Instruments*, 83(10):104708, 2012.