**ARTICLE**    OPEN

Check for updates

# Digital oximetry biomarkers for assessing respiratory function: standards of measurement, physiological interpretation, and clinical use

Jeremy Levy [1,2], Daniel Álvarez [3,4,5], Aviv A. Rosenberg [6], Alexandra Alexandrovich[1], Félix del Campo [3,4,5] and Joachim A. Behar [1]✉

Pulse oximetry is routinely used to non-invasively monitor oxygen saturation levels. A low oxygen level in the blood means low oxygen in the tissues, which can ultimately lead to organ failure. Yet, contrary to heart rate variability measures, a field which has seen the development of stable standards and advanced toolboxes and software, no such standards and open tools exist for continuous oxygen saturation time series variability analysis. The primary objective of this research was to identify, implement and validate key digital oximetry biomarkers (OBMs) for the purpose of creating a standard and associated reference toolbox for continuous oximetry time series analysis. We review the sleep medicine literature to identify clinically relevant OBMs. We implement these biomarkers and demonstrate their clinical value within the context of obstructive sleep apnea (OSA) diagnosis on a total of $n = 3806$ individual polysomnography recordings totaling 26,686 h of continuous data. A total of 44 digital oximetry biomarkers were implemented. Reference ranges for each biomarker are provided for individuals with mild, moderate, and severe OSA and for non-OSA recordings. Linear regression analysis between biomarkers and the apnea hypopnea index (AHI) showed a high correlation, which reached $\overline{R}^2 = 0.82$. The resulting python OBM toolbox, denoted "pobm", was contributed to the open software PhysioZoo (physiozoo.org). Studying the variability of the continuous oxygen saturation time series using pbom may provide information on the underlying physiological control systems and enhance our understanding of the manifestations and etiology of diseases, with emphasis on respiratory diseases.

*npj Digital Medicine* (2021)4:1 ; https://doi.org/10.1038/s41746-020-00373-5

## INTRODUCTION

Pulse oximetry is routinely used for non-invasive monitoring of oxygen saturation levels. A low oxygen level in the blood means low oxygen in the tissues, which can ultimately lead to organ failure. Oximetry can be used to sporadically measure the oxygen saturation level during a medical examination or continuously monitor patients in the intensive care unit (ICU) or overnight for a polysomnography (PSG) study. Identification of digital biomarkers extrapolated from the oxygen saturation time series can support the diagnosis and continuous monitoring of patient pulmonary function to predict deteriorations (prognosis). Specifically, study-ing the variability of the oxygen saturation signal may provide information on the underlying physiological control systems. Furthermore, it may enhance our understanding of the manifesta-tion and etiology of diseases and identify digital oximetry biomarkers (OBMs) for the purpose of health monitoring. Sleep medicine makes standard usage of oximetry biomarkers, where overnight drops in oxygen saturation are characteristic of obstructive sleep apnea (OSA). Beyond the presence of OSA, the repetitive nocturnal hypoxemia may cause oxidative stress, contributing to the pathogenesis of cardiovascular morbidity[1]. Similarly, patients with advanced chronic obstructive pulmonary disease (COPD), and with no primary sleep-related breathing disorders, commonly exhibit overnight hypoxemia[2]. Yet, contrary to heart rate variability (HRV) measures, a field which has

benefited from the development of stable standards[3] and advanced toolboxes and software[4–6], there are currently no such standards and open tools for analyzing oxygen saturation time series in terms of its variability, dynamics, and the statistical characterization of specific patterns.

### This contribution

Research on the use of existing and development of new oximetry biomarkers has mainly focused on the diagnosis of OSA, as echoed by five recent reviews in this field[7–11]. Although this paper will naturally somewhat overlap with these reviews, we present a new comprehensive review focusing on the physiological inter-pretation and clinical use of oximetry biomarkers, in the spirit of the work of Malik et al.[3] in the field of HRV analysis. We also develop a complete Python toolbox (denoted "pobm") and software interface ("PhysioZoo OBM") for usage of these biomarkers, similar to our previous work in HRV analysis[6]. This will support rigorous research in oximetry time series analysis and ensure reproducibility of research. We apply the developed toolbox to a large dataset of overnight recordings in order to demonstrate its usability and clinical value in the context of OSA diagnosis. While the work mainly focuses on OBM developed in the field of sleep medicine, the reviewed OBMs can be applied to the analysis of continuous oximetry recordings for any other condition and we thus introduce a general purpose flow diagram

¹Faculty of Biomedical Engineering, Technion Institute of Technology, Haifa, Israel. ²Faculty of Electrical Engineering, Technion Institute of Technology, Haifa, Israel. ³Río Hortega University Hospital Valladolid, Valladolid, Spain. ⁴Biomedical Engineering Group, University of Valladolid, Valladolid, Spain. ⁵Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Valladolid, Spain. ⁶Department of Computer Science, Technion Institute of Technology, Haifa, Israel. ✉email: jbehar@technion.ac.il

for continuous oximetry analysis. We limit the scope of the biomarkers to single-channel oximetry analysis, thereby implicitly excluding OBMs that may require additional channels, such as airflow, to be engineered[12]. The paper defines categories of OBMs and a review of the literature to identify evidence-based OBMs in the field of sleep medicine as well as some additional suggested OBMs. The biomarkers are applied to a large PSG dataset totaling 3806 individual oximetry recordings.

## RESULTS

### pobm toolbox and PhysioZoo OBM interface

The pobm toolbox was implemented in Python. For the purpose of quality control, functions were benchmarked against comparative reference source code (Supplementary Table 1) or ranges published in the literature (Supplementary Table 2). For the comparison to range reported in the literature, we compared the order of magnitude of some biomarkers with those reported by other for the non-OSA group. Figure 1 shows the PhysioZoo OBM interface for oximetry analysis. In PhysioZoo OBM, a SpO$_2$ time series can be loaded (File → Open data file) and pre-filtered using one of the preprocessing filters introduced in section "Preprocessing". After computation, OBM can be exported together with standard data representation figures. The PhysioZoo software handles data in .txt, .mat (The MathWorks, Inc., Natick, MA, USA) and WFDB[13] formats. In addition, the PhysioZoo OBM enables oximetry analysis of multiple segments, thereby enabling tracking of temporal changes in oximetry measures for a given record. The pobm toolbox and PhysioZoo software are available at https://physiozoo.com/.

### Standard ranges for oximetry biomarkers

Tables 1 and 2 summarize the median and interquartile range for all the OBMs implemented in the PhysioZoo software for individuals participating in SHHS1. This provides a standard reference range for each oximetry biomarker. The null hypothesis of Kruskal–Wallis test was rejected for most biomarkers (43/44) with the smallest $p$ value obtained for Px, CA$_x$, and CT$_x$. Following the Dunn post hoc analysis, a total of 30 biomarkers were statistically discriminative between every pair of classes, i.e., $p < 0.05$ for all pairs (Tables 1 and 2).

### Added value in combining multiple biomarkers

When performing simple regression analysis for individual OBM against the AHI, CTM$_\rho$ with $\rho = 0.25$ achieved the highest goodness of fit ($\overline{R}^2 = 0.77$) followed by the ODI3 ($\overline{R}^2 = 0.74$). When combining the 10 oximetry biomarkers with the best score within a multivariable linear regression framework, the goodness of fit was further improved to $\overline{R}^2 = 0.82$ (Fig. 2).

## DISCUSSION

We showed that OBMs engineered from continuous oximetry recordings may provide discriminative information of groups of individuals suffering from respiratory disorders. Within the context of OSA, we found that CTM$_\rho$ with $\rho = 0.25$ had the highest $\overline{R}^2$ in estimating the AHI, with $\overline{R}^2 = 0.77$. Furthermore, we demonstrated that combining multiple oximetry biomarkers for estimating the AHI increased the $\overline{R}^2$ to 0.82. This highlights the complementary value in using multiple OBMs versus a single one.

Recent studies have shown that nocturnal hypoxemia correlates better with cardiovascular disease, cancer incidence, and mortality than traditional nocturnal respiratory disturbance indexes, such as the AHI[12,14]. This suggests that alternative nocturnal OBMs may provide important health information. Both intermittent hypoxia and sleep fragmentation are responsible for clinical manifestations and most related comorbidities of OSA[15]. In OSA, recurrent collapse of the upper airway leads to a reduced tidal volume and both intermittent hypoxemia and hypercapnia. In consequence, activity of the sympathetic nervous system increases and cortical arousals occur, leading to disrupted sleep architecture and restless sleep. In addition, repetitive hypoxemia–reoxygenation periods are linked to the production of free oxygen radicals, inflammation, and endothelial dysfunction[16]. In this regard, OBMs, such as CT90 and overnight mean and minimum saturation, have been significantly linked with dysfunction in cardiovascular modulation,



**Fig. 1  PhysioZoo OBM interface for oximetry time series analysis.** The analysis of an oximetry time series is shown. All the OBM biomarkers are computed for the selected window (in light green).

**Table 1.** Summary statistics (median and interquartiles Q1–Q3) for the General statistics, Complexity and Periodicity OBMs evaluated on the SHHS1 database.

| Biomarker | Non-OSA ($n = 1195$) | Mild ($n = 1303$) | Moderate ($n = 833$) | Severe ($n = 475$) | Kruskal test |
|---|---|---|---|---|---|
| **General statistics** | | | | | |
| AV | 95.90 (94.77–96.93) | 95.14 (94.01–96.09) | 94.43 (93.36–95.41) | 93.51 (92.23–94.67) | $p < 0.001$[‡] |
| MED | 96.00 (95.00–97.00) | 95.00 (94.00–96.00) | 95.00 (93.00–96.00) | 94.00 (93.00–95.00) | $p < 0.001$[‡] |
| Min | 79.00 (72.00–85.00) | 79.00 (74.00–84.00) | 79.00 (73.00–83.00) | 76.00 (70.00–81.00) | $p < 0.001$ |
| SD | 1.42 (1.08–1.90) | 1.55 (1.29–1.90) | 1.81 (1.50–2.19) | 2.33 (1.88–3.04) | $p < 0.001$[‡] |
| RG | 20.00 (14.00–27.00) | 19.00 (14.00–25.00) | 20.00 (15.00–26.00) | 22.00 (18.00–29.00) | $p < 0.001$ |
| Pxx | 91.00 (88.00–93.00) | 90.00 (88.00–92.00) | 89.00 (86.00–90.12) | 85.00 (81.00–88.00) | $p < 0.001$[‡] |
| Mx | 2.14 (1.18–3.82) | 3.44 (1.94–5.51) | 5.84 (3.49–8.96) | 10.98 (6.78–17.00) | $p < 0.001$[‡] |
| ZCxx | 435.00 (346.00–540.00) | 481.00 (391.00–579.00) | 557.00 (463.50–638.00) | 690.00 (592.00–812.50) | $p < 0.001$[‡] |
| $\Delta I$ | 0.36 (0.30–0.42) | 0.47 (0.41–0.54) | 0.66 (0.57–0.76) | 1.03 (0.86–1.44) | $p < 0.001$[‡] |
| **Complexity** | | | | | |
| ApEn | 0.17 (0.13–0.20) | 0.21 (0.17–0.26) | 0.28 (0.23–0.34) | 0.39 (0.31–0.48) | $p < 0.001$[‡] |
| LZ | 2389.00 (2241.00–2511.00) | 2435.00 (2314.00–2547.00) | 2492.00 (2374.50–2587.50) | 2587.00 (2483.50–2698.00) | $p < 0.001$[‡] |
| CTMx | 0.92 (0.90–0.93) | 0.89 (0.88–0.91) | 0.86 (0.84–0.87) | 0.79 (0.73–0.82) | $p < 0.001$[‡] |
| SampEn | 0.06 (0.04–0.08) | 0.08 (0.06–0.11) | 0.12 (0.09–0.15) | 0.21 (0.14–0.31) | $p < 0.001$[‡] |
| DFA | 1.14 (0.99–1.37) | 1.35 (1.18–1.54) | 1.70 (1.49–1.98) | 2.47 (2.05–3.24) | $p < 0.001$[‡] |
| **Periodicity** | | | | | |
| $PRSAD_c$ | 0.56 (0.53–0.60) | 0.56 (0.54–0.59) | 0.59 (0.56–0.63) | 0.67 (0.61–0.78) | $p < 0.001$[‡] |
| $PRSAD_{ad}$ | 1.55 (1.34–1.89) | 1.83 (1.58–2.16) | 2.42 (2.01–2.97) | 3.68 (2.86–4.94) | $p < 0.001$ |
| $PRSAD_{os}$ | 0.11 (0.09–0.13) | 0.13 (0.11–0.15) | 0.17 (0.14–0.20) | 0.25 (0.20–0.33) | $p < 0.001$[‡] |
| $PRSAD_{sb}$ | 0.01 (0.00–0.03) | 0.03 (0.02–0.05) | 0.07 (0.05–0.10) | 0.14 (0.09–0.21) | $p < 0.001$[‡] |
| $PRSAD_{sa}$ | 0.02 (0.01–0.04) | 0.04 (0.03–0.06) | 0.08 (0.05–0.11) | 0.14 (0.09–0.20) | $p < 0.001$ |
| AC($\times 10^4$) | 18.20 (17.83–18.54) | 17.97 (17.58–18.29) | 17.69 (17.30–18.04) | 17.33 (16.90–17.79) | $p < 0.001$ |
| PSD_total | 91.49 (89.48–93.71) | 94.65 (92.36–97.49) | 99.81 (96.76–103.93) | 109.27 (103.81–118.61) | $p < 0.001$[‡] |
| PSD_band | 3.22 (2.78–3.91) | 4.28 (3.66–5.06) | 6.18 (5.18–7.39) | 9.46 (7.60–13.42) | $p < 0.001$[‡] |
| PSD_ratio | 0.04 (0.03–0.04) | 0.05 (0.04–0.05) | 0.06 (0.05–0.07) | 0.09 (0.07–0.11) | $p < 0.001$[‡] |
| PSD_peak | 0.02 (0.02–0.02) | 0.03 (0.02–0.03) | 0.04 (0.03–0.05) | 0.06 (0.05–0.10) | $p < 0.001$[‡] |

The symbol ‡ indicate when the result of the Dunn post hoc test between all pairs of groups was $p < 0.05$. Results presented for the categories general statistics, complexity, and periodicity.

arterial hypertension, atrial fibrillation, increased insulin resistance, higher incidence of lung cancer, and worst prognosis after myocardial infarction, as well as higher risk of post-surgery complications in OSA patients[17]. However, it remains very unclear which oximetric biomarkers (or combinations of biomarkers) are most predictive of clinical endpoints, such as metabolic and cardiovascular diseases. For example, the average duration and morphology of the events are not considered in routine OSA diagnosis. This is limiting, since longer apnea or hypopnea events will likely result in increased desaturation (in length and depth), which will likely result in more hypoxic stress, leading to more severe cardiovascular consequences. At the same time, longer desaturation events may result in a decrease in AHI, that is, a lower number of events per hour. Thus the relationship between duration and morphology of events and a clinical endpoint (e.g., cardiovascular complication) remains unclear. For this reason, additional desaturation biomarkers may provide valuable information on disease phenotyping. This has been suggested, for example, in the work of Kulkas et al.[18], who showed, albeit on a very small population sample ($n = 19$), that additional oximetry biomarkers, i.e., duration and morphology related, enhance OSA phenotyping.

Nocturnal hypoxemia can be present in many respiratory diseases that are either acute or chronic. For example, OSA patients show cyclic desaturation–resaturation episodes during the night, which are linked with partial or complete obstruction of the upper airway, leading to the well-known chronic intermittent hypoxia pattern. On the other hand, COPD patients show slower and longer desaturations linked with sustained hypoventilation, mainly during rapid eye movement (REM) sleep, leading to a state of nocturnal chronic hypoxemia. Characterization of hypoxemia is therefore different during REM sleep. In subjects with OSA, oxygen desaturation indices (ODIs) of 3 and 4%, as well as mean, minimum, and CT90, are widely used in sleep medicine. On the contrary, criteria for classifying a COPD patient as a nocturnal desaturator are not well established. Showing at least one episode with saturation <90% lasting for >5 min and reaching a minimum saturation of at least 85% has been proposed[19], while some authors define nocturnal desaturators as patients with CT90 ≥30%[20]. This example highlights that a standard to characterize and quantify some respiratory conditions, such as COPD, using nocturnal oximetry, remain to be defined. Usage of multiple OBM may also enable to identify patterns for different apnea types such as central apnea versus mixed apnea versus obstructive apnea as ongoing research studies such as the SomnaPatch intend (Somnarus Inc., ClinicalTrials.gov Identifier: NCT02034175). Machine learning algorithms will play an important role in engineering models that can learn complex combinations of OBM for the purpose of regression or classification tasks. Such models will uncover the OBM combinations that best reflect the unique patterns of a given condition.

Because respiratory conditions may possess different oximetry patterns/dynamics and oximetry recordings may be of different

**Table 2.** Summary statistics (median and interquartiles Q1–Q3) for the Desaturations and Hypoxic Burden OBMs evaluated on the SHHS1 database.

| Biomarker | Non-OSA ($n = 1195$) | Mild ($n = 1303$) | Moderate ($n = 833$) | Severe ($n = 475$) | Kruskal test |
|---|---|---|---|---|---|
| Desaturations | | | | | |
| $DL_\mu$ | 39.44 (34.38 to 45.23) | 41.95 (37.00 to 47.31) | 42.10 (37.24 to 46.98) | 38.04 (33.30 to 43.11) | $p < 0.001$ |
| $DL_\sigma$ | 23.04 (18.65 to 26.82) | 22.94 (19.62 to 26.16) | 21.68 (18.69 to 24.36) | 18.10 (15.10 to 21.04) | $p < 0.001$ |
| $DL_\sigma$ | 4.93 (4.11 to 6.06) | 4.60 (4.12 to 5.27) | 4.53 (4.13 to 5.16) | 5.00 (4.39 to 6.02) | $p < 0.001^\ddagger$ |
| $DDmax_\sigma$ | 2.63 (1.43 to 3.97) | 2.24 (1.48 to 3.26) | 2.20 (1.57 to 2.95) | 2.48 (1.84 to 3.34) | $p < 0.001$ |
| $DD100_\mu$ | 9.47 (8.20 to 10.91) | 9.38 (8.42 to 10.50) | 9.57 (8.54 to 10.63) | 10.28 (8.98 to 11.88) | $p < 0.001$ |
| $DD100^\sigma$ | 3.24 (1.85 to 5.06) | 2.78 (1.90 to 4.03) | 2.60 (1.95 to 3.58) | 2.74 (2.11 to 3.84) | $p < 0.001$ |
| $DS_\mu$ | −0.22 (−0.26 to 0.18) | −0.19 (−0.23 to 0.16) | −0.18 (−0.21 to 0.16) | −0.20 (−0.25 to 0.17) | $p < 0.001^\ddagger$ |
| $DS_\sigma$ | 0.14 (0.11 to 0.19) | 0.13 (0.10 to 0.16) | 0.11 (0.09 to 0.14) | 0.11 (0.09 to 0.14) | $p < 0.001^\ddagger$ |
| $DAmax_\mu$ | 104.74 (80.48 to 137.64) | 104.55 (87.47 to 126.95) | 108.16 (90.44 to 130.04) | 115.22 (96.08 to 144.56) | $p < 0.001$ |
| $DAmax_\sigma$ | 87.52 (53.67 to 144.54) | 82.72 (58.97 to 122.55) | 85.36 (64.67 to 117.12) | 94.62 (70.51 to 127.90) | $p < 0.001$ |
| $DA100_\mu$ | 280.75 (223.36 to 352.90) | 304.38 (251.74 to 362.10) | 315.47 (258.35 to 377.54) | 313.76 (262.60 to 380.12) | $p < 0.001$ |
| $DA100_\sigma$ | 169.66 (120.29 to 262.66) | 175.29 (133.78 to 233.17) | 177.53 (134.72 to 222.65) | 175.48 (135.08 to 225.74) | $p > 0.05$ |
| $TD_\mu$ | 1760.82 (1122.78 to 2804.90) | 940.56 (673.65 to 1373.38) | 429.00 (321.28 to 617.51) | 181.61 (117.70 to 268.42) | $p < 0.001^\ddagger$ |
| $TD_\sigma$ | 2365.50 (1660.51 to 3388.55) | 1457.2 (1092.66 to 2014.68) | 812.64 (635.57 to 1066.14) | 375.70 (222.57 to 544.07) | $p < 0.001$ |
| Hypoxic burden | | | | | |
| $POD_x$ | 1.72 (1.69 to 01.76) | 10.20 (9.60 to 11.00) | 18.23 (18.11 to 18.79) | 31.10 (30.02 to 33.65) | $p < 0.001^\ddagger$ |
| $AODmax$ | 0.06 (0.03 to 0.10) | 0.11 (0.07 to 0.16) | 0.24 (0.16 to 0.36) | 0.61 (0.37 to 1.05) | $p < 0.001^\ddagger$ |
| $AOD100$ | 0.15 (0.09 to 0.26) | 0.31 (0.20 to 0.47) | 0.70 (0.46 to 1.00) | 1.65 (1.04 to 2.77) | $p < 0.001^\ddagger$ |
| $CT_x$ | 0.51 (0.15 to 1.34) | 0.76 (0.30 to 1.69) | 1.42 (0.58 to 3.29) | 4.48 (1.68 to 11.58) | $p < 0.001^\ddagger$ |
| $CA_x$ | 0.45 (0.13 to 1.10) | 0.64 (0.25 to 1.42) | 1.21 (0.51 to 2.79) | 3.84 (1.42 to 9.86) | $p < 0.001^\ddagger$ |

The symbol ‡ indicate when the result of the Dunn post hoc test between all pairs of groups was $p < 0.05$. Results presented for the categories desaturation measures and hypoxic burden.
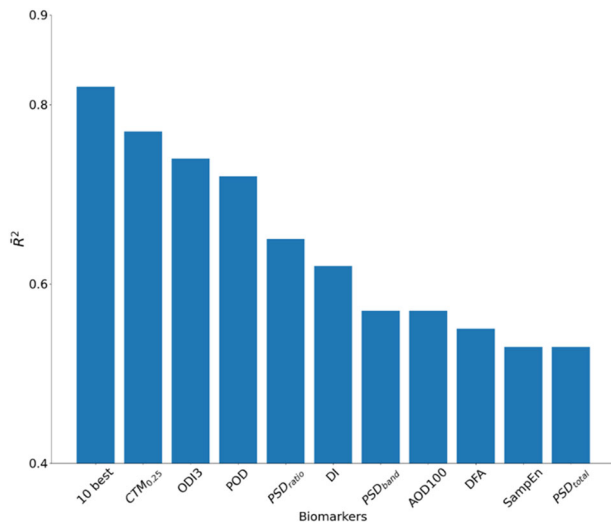


**Fig. 2 Linear correlation between oximetry biomarkers and the AHI.** The adjusted $\bar{R}^2$ is reported for every biomarker and for the 10 biomarkers with the highest $\bar{R}^2$.

durations, it is important to define a general methodology for continuous oximetry time series analysis using the OBM toolbox. The suggested flow for such analysis is illustrated in Fig. 3. Following these steps, performance statistics relevant to the task at hand should be reported and a clear discussion should be delivered regarding the biomarkers that were most relevant to the data-driven model including interpretation about the underlying physiology.

Oximetry biomarkers may vary significantly with the technology used (transmission versus reflectance) as well as by the manufacturer. Most oximeters use two light-emitting diodes (LED) that face a translucent part of the body, such as the fingertip or earlobe, and a photodiode that receives light rays. In most cases, one LED is red and the second infrared. The oximeter includes a processor that calculates the oxygen saturation using the ratio between the amount of light that was emitted and the amount that was received at each wavelength. Oximeters may be transmissive or reflective. In a transmissive oximeter, the photo-diode and the LEDs are placed on opposite sides of the measurement site and the light passes through the site. In a reflective oximeter, the LEDs are placed on the same side and the light is reflected to the photodiode across the measurement site.

During the current coronavirus disease 2019 (COVID-19) pandemic, many individuals with suspected or confirmed, but mild, COVID-19 are told to monitor their symptoms at home or from government-managed locations. Hospitalization is only an option if there is a medical need. Monitoring the blood oxygen level may be a meaningful way to remotely monitor individuals with mild COVID-19[21]. It could also be used for continuous monitoring of patients in the ICU with pneumonia, a common complication of COVID-19. However, there is a lack of smart algorithms that can exploit the information encrypted within these oxygen saturation physiological time series. The development of such algorithms will facilitate the continuous monitoring of COVID-19 ICU patients in predicting deteriorations. It remains to be determined how the information contained in the oxygen saturation physiological time series can be exploited. Are trends or absolute values or the occurrence of specific patterns the most meaningful information for identification of the disease and prediction of its course? The pobm toolbox developed in this publication can support researching novel biomarkers for diagnosis and prognosis of COVID-19.

Additional OBMs, such as kernel entropy, bispectrum, and wavelet, among others, should be considered and added to the library in future works. Although we demonstrated the usage of the PhysioZoo OBM resource on OSA, there is a need to assess the value of these biomarkers for other respiratory disorders.

Typical oximetry biomarkers used in clinical practice include the ODI and CT90. While these indices are standardized, to some extent, and interpretable, they fail to capture important pathophysiological characteristics. We reviewed evidence-based oximetry biomarkers,
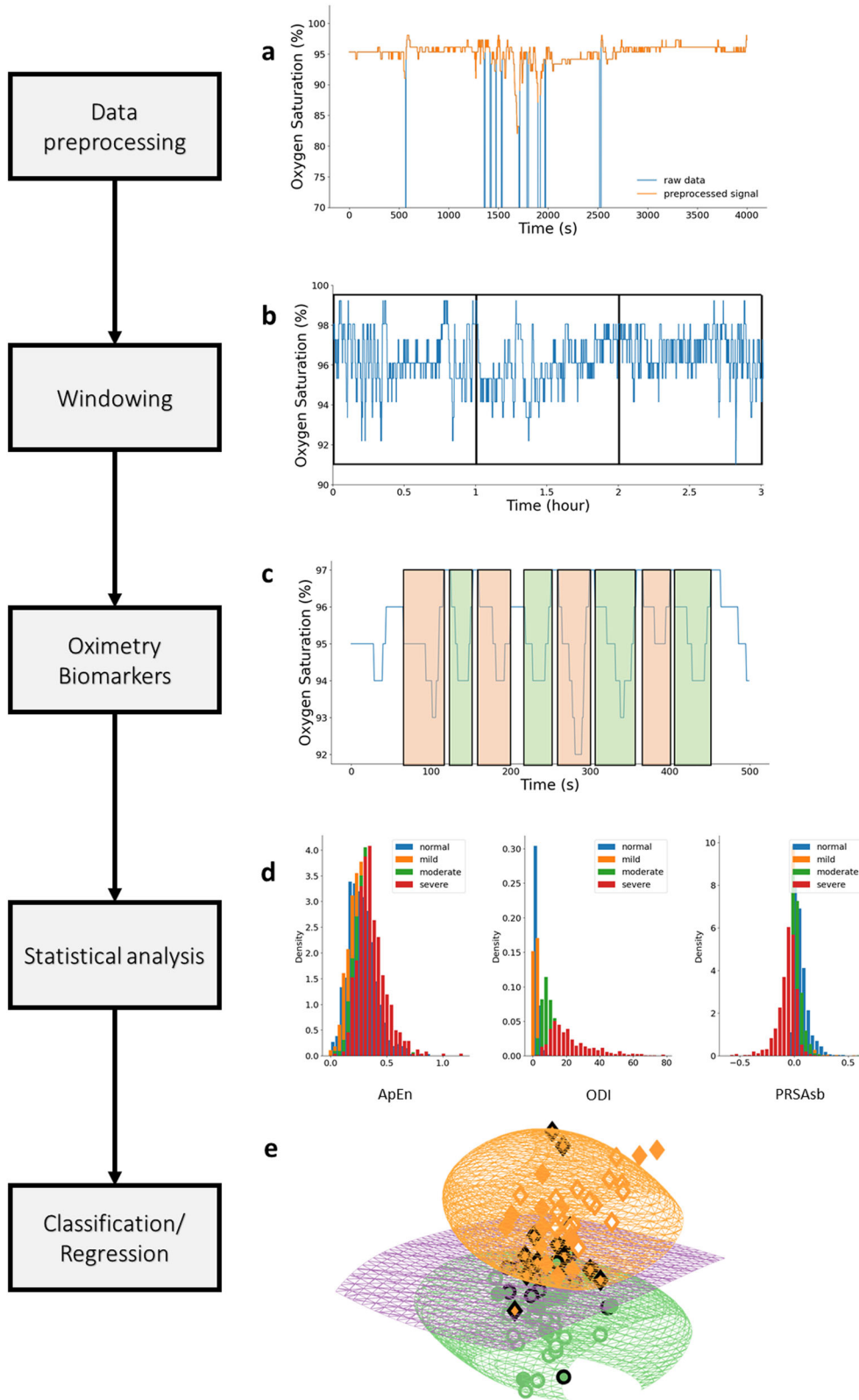
Fig. 3  **Flowchart for continuous oximetry time series analysis. a** The raw data are first preprocessed using some of the routines presented in section "Preprocessing", to remove any non-physiological values that are likely caused by noise. **b** The recording is then windowed. The size of the analysis windows and whether to consider overlapping windowing is an important consideration and will be research dependent. **c** Oximetry biomarkers are then computed using the OBM toolbox. **d** A statistical analysis is performed to obtain a preliminary understanding of the biomarker behavior for the different classes of interest. **e** Finally, a machine learning model can be used to combine the biomarkers for a regression or classification task.

suggested a classification system, and created a unique resource (pobm toolbox and PhysioZoo OBM interface) for performing oximetry time series analysis. This resource can be applied to gain novel physiological, clinical, and epidemiological insights.

## METHODS

### Oximetry biomarker categorization

Various categorizations of OBMs have been previously suggested[10,22]. We introduce a five-category classification scheme of our own, which we believe best reflects the literature and usage of these biomarkers in medical practice.

(i)   *General statistics:* these are time-based statistics describing the oxygen saturation time series data distribution.
(ii)  *Complexity*: quantifies the presence of long-range correlations in non-stationary time series.
(iii) *Periodicity*: quantifies consecutive events to identify periodicity in the oxygen saturation time series.
(iv)  *Desaturations:* time-based descriptive measures of the desaturation patterns occurring throughout the time series.
(v)   *Hypoxic burden*: time-based measures quantifying the overall degree of hypoxemia imposed on the heart and other organs during the recording period.

A comprehensive summary of the OBMs reviewed and implemented in this research is presented in Table 3 for the general statistics, complexity, and periodicity categories and in Table 4 for the desaturation measures and hypoxic burden categories. A total of 44 oximetry biomarkers were engineered. A glossary with variables symbols and definition is presented in Supplementary Table 3.

### Preprocessing

Raw oximetry data is often associated with missing values and artefacts caused, for example, by motion of the oximeter or lack of proper contact between the finger and the probe. Therefore, the toolbox includes an option for two preprocessing filters:

*Delta filter.* A delta filter is applied to the $SpO_2$ time series, in which, when two consecutive samples are $>x\%/s$ apart, they are considered non-physiological and are discarded. By default, $x = 4\%/s$ apart as in the work of Taha et al.[23]. As an example, applying the delta filter to the $SpO_2$ time series is shown in Fig. 4a.

*Block of data filter.* An error value is considered a value $<50\%$. For each error value, a small block of data of length $x$ s (default is $x = 20$ s) around it is discarded. Once the small blocks are removed, the mean is computed for each block of data of length 100 s around the original error value. The mean of the overall $SpO_2$ signal is also computed. Each block with a mean $<6\%$ smaller than the overall mean is discarded. This technique was used by Buekers et al.[24]. As an example, applying the block of data filter to the $SpO_2$ time series is shown in Fig. 4b.

### General statistics

Average (AV): average of $SpO_2$ values. Median (MED): median of $SpO_2$ values. Min (Min): minimum of $SpO_2$ values representing the physiological minimum of the $SpO_2$. Standard deviation (SD): standard deviation of $SpO_2$ values. Range (RG): the difference between the maximal and minimal $SpO_2$ values. Percentile (Px): The $x$th percentile of $SpO_2$ values. BelowMedian (Mx): Percentage of the signal $x\%$ below median oxygen saturation.

**Table 3.** List of digital oximetry biomarkers for the categories: general statistics, complexity, and periodicity.

| | Biomarker | Definition | Unit |
|---|---|---|---|
| *General statistics* | | | |
| 1 | AV | Blood oxygen saturation ($SpO_2$) mean | % |
| 2 | MED | $SpO_2$ median | % |
| 3 | Min | $SpO_2$ min | % |
| 4 | SD | $SpO_2$ standard deviation | % |
| 5 | RG | $SpO_2$ range | % |
| 6 | Px | $x^{th}$ percentile $SpO_2$ value, by default $x = 1$ | % |
| 7 | Mx | Percentage of the signal at least $x\%$ below median oxygen saturation, by default $x = 2$, used by Deviaene et al.[22] | % |
| 8 | ZCx | Number of zero-crossing points at the $x\%$ $SpO_2$ level[25], by default $x = AV$ | nu |
| 9 | ΔIx | Delta index[26], by default $x = 12$ s. | % |
| *Complexity* | | | |
| 10 | ApEn | Approximate entropy[28] with, by default, $m = 1$, $r = 0.25$ times the standard deviation of the data | nu |
| 11 | LZ | Lempel–Ziv complexity[31] | nu |
| 12 | $CTM_\rho$ | Central tendency measure[37] with radius $\rho$, by default $\rho = 0.25$ | nu |
| 13 | SampEn | Sample entropy[29] with, by default, $m = 1$, $r = 0.25$ | nu |
| 14 | DFA | Detrended fluctuation analysis[36] with, by default, $n = 20$ | % |
| *Periodicity* | | | |
| 15 | $PRSAD_c$ | Phase-rectified signal averaging (PRSA) capacity[22,38]. With $d$ the fragment duration, by default $d = 10$ | % |
| 16 | $PRSAD_{ad}$ | PRSA amplitude difference[22,38,57]. With $d$ the fragment duration, by default $d = 10$ | % |
| 17 | $PRSAD_{os}$ | PRSA overall slope[38,57]. With $d$ the fragment duration, by default $d = 10$ | %/s |
| 18 | $PRSAD_{sb}$ | PRSA slope before the anchor point[38,57]. With $d$ the fragment duration, by default $d = 10$ | %/s |
| 19 | $PRSAD_{sa}$ | PRSA slope after the anchor point[38,57]. With $d$ the fragment duration, by default $d = 10$ | %/s |
| 20 | AC | Autocorrelation | $\%^2$ |
| 21 | PSD_total | The integral of the power spectral density (PSD) function[45] | % |
| 22 | PSD_band | The integral of the PSD function within the band $0.014-0.033$ Hz[45] | % |
| 23 | PSD_ratio | The integral of the PSD function within the band $0.014-0.033$ Hz with respect to the total integral[45] | nu |
| 24 | PSD_peak | Peak amplitude of the PSD function within the band $0.014-0.033$ Hz[45] | % |

**Table 4.** List of digital oximetry biomarkers for the categories: desaturation measures and hypoxic burden.

| | Biomarker | Definition | Unit |
|---|---|---|---|
| *Desaturation measures* | | | |
| 25 | $ODI_x$ | The oxygen desaturation index as defined in Jung et al. and Behar et al.[50,51], by default $x = 3$ | event/h |
| 26 | $DL_\mu$ | Mean of desaturations length (e.g., related work in Kulkas et al.[47]) | s |
| 27 | $DL_\sigma$ | Standard deviation of desaturations length (e.g., related work in Kulkas et al.[47]) | $s^2$ |
| 28 | $DDmax_\mu$ | Mean of desaturations depth (e.g., related work in Kulkas et al.[47]) | % |
| 29 | $DDmax_\sigma$ | Standard deviation of desaturation depth (e.g. related work in Kulkas et al.[47]) | $\%^2$ |
| 30 | $DD100_\mu$ | Mean of desaturations depth using 100% $SpO_2$ level as baseline (e.g., related work in Terrill et al.[10]) | % |
| 31 | $DD100_\sigma$ | Standard deviation of desaturations depth using 100% $SpO_2$ level as baseline (e.g., related work in Terrill et al.[10]) | $\%^2$ |
| 32 | $DS_\mu$ | Mean of the desaturation slope | %/s |
| 33 | $DS_\sigma$ | Standard deviation of the desaturation slope | $(\%/s)^2$ |
| 34 | $DAmax_\mu$ | Desaturation area defined as the mean of the desaturation areas using the maximum $SpO_2$ value in each desaturation event as baseline[47] | %*s |
| 35 | $DAmax_\sigma$ | Standard deviation of desaturation area[47] | $(\%*s)^2$ |
| 36 | $DA100_\mu$ | Desaturation area: mean of desaturation area under the 100% $SpO_2$ level as baseline[47] | %*s |
| 37 | $DA100_\sigma$ | Standard deviation of desaturation area under the 100% $SpO_2$ level as baseline[47] | $(\%*s)^2$ |
| 38 | $TD_\mu$ | Mean of time between two consecutive desaturation events | sec |
| 39 | $TD_\sigma$ | Standard deviation of time between two consecutive desaturation events | $s^2$ |
| *Hypoxic burden* | | | |
| 40 | $POD_x$ | Time of oxygen desaturation event, normalized by the total recording time[18], by default $x = 3$ | s |
| 41 | $AOD_{max}$ | The area under the oxygen desaturation event curve[18], using the maximum $SpO_2$ value as baseline and normalized by the total recording time | % |
| 42 | $AOD_{100}$ | Cumulative area of desaturations under the 100% $SpO_2$ level as baseline and normalized by the total recording time | % |
| 43 | $CT_x$ | Cumulative time below the $x$% oxygen saturation level, by default $x = 90$. Introduced by Olson et al.[55] | % |
| 44 | $CA_x$ | Integral of $SpO_2$ below the $x$ $SpO_2$ level normalized by the total recording time, by default $x = AV$ | % |

*ZeroCrossing (ZCx).* ZeroCrossing (ZCx) is the number of zero-crossing points, used by Xie et al.[25], using the $x$% $SpO_2$ level as baseline. A crossing point is considered as two consecutive samples of the $SpO_2$ signal, one lower than the baseline and the second greater, or vice versa. This biomarker helps to understand how the signal oscillates around a baseline. The intuition is that a $SpO_2$ time series from a patient with OSA as compared to that of a non-OSA patient will oscillate more around the baseline because of the presence of desaturations and then reach a higher value. A common baseline used for this biomarker is the mean of the signal (i.e., by default, $x = AV$). ZCx is defined as:

$$ZCx = \sum_{i=10}^{N_{SpO_2}-1} ZC_i(x), \qquad (1)$$

$$ZC_i(x) = \begin{cases} 1 & \text{if}(SpO2_i - x)(SpO2_{i+1} - x) < 0 \\ 0 & \text{else} \end{cases}, \qquad (2)$$

where $N_{SpO2}$ is the number of samples of the $SpO_2$ time series.

*Delta index (ΔIx).* $\Delta Ix$[26] corresponds to the sum of the absolute variations between two successive points divided by the number of intervals. The original intuition was that $SpO_2$ oscillations, induced by repeated apnea resumption of ventilation sequence, will lead to a high ΔI, while COPD-induced prolonged desaturations or nearly constant $SpO_2$ would lead to a low ΔI. In the original paper of Pepin et al.[26], recordings from a total of 160 consecutive patients referred for PSG were used to set the threshold of the ΔI distinguishing between OSA and non-OSA. When testing on the prospective group of patients, i.e., $n = 36$ patients with $p = 34$ nights of recordings for each patient, they obtained a sensitivity of 0.75 and specificity of 0.86. In Magalang et al.[27], the ΔI was the best predictor ($r^2 = 0.60$) for the AHI versus other OBMs including the ODI. The ΔI index is defined as:

$$\Delta Ix = \frac{1}{N_{window}} \cdot \sum_{i=1}^{N_{window}} |SpO2\_window_{i+1} - SpO2\_window_i|, \qquad (3)$$

where $SpO2\_window_i$ is the average of the level of oxygen saturation for the

window $i$ of length $x$ s, and $N_{window}$ is the number of windows. In their original work, Pepin et al.[26] used $x = 12$ s. In our implementation, signals are re-sampled to 1 Hz by default, so, by default, a window will contain 12 samples.

## Complexity measures

Regularity quantifies how often similar patterns are observed in the oximetry signal[10]. In the context of physiological time series analysis, approximate entropy (ApEn)[28] and sample entropy (SampEn)[29] have commonly been used as measures of the unpredictability (opposite of regularity). OSA individuals typically have less regular oximetry patterns, leading to higher ApEn and SampEn values as compared to non-OSA individuals. Loss of physiological complexity may be better captured by using other measures that can detect and quantify the presence of long-range correlations in non-stationary time series[30] with measures such as the Lempel–Ziv complexity (LZ)[31]. Fractal objects, generated by stochastic or nonlinear deterministic mechanisms[30], may also be used to capture complexity, as they show self-similarity, i.e., the smaller-scale structure resembles the larger-scale form[32]. Detrended fluctuation analysis (DFA) has commonly been used for fractal analysis in the field of physiological time series analysis.

*Approximate entropy.* ApEn is a biomarker introduced in Pincus et al.[28], which aims to capture the irregularity in the signal, with higher values indicating higher irregularity. This biomarker is very useful in the detection of OSA, as high randomness is associated with high values of the biomarker. Thus apneas and hypopneas are associated with high ApEn values. ApEn($m$, $r$, $N$) can be defined as:

$$ApEn = \varphi^m(r) - \varphi^{m+1}(r), \qquad (4)$$

$$\varphi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln\left(\frac{N^m(i)}{(N-m+1)}\right), \qquad (5)$$

where $N^m(i)$ is the number of windows of length $m$ for which the distance from the window beginning at the index $i$ is lower than or equal to $r$. The
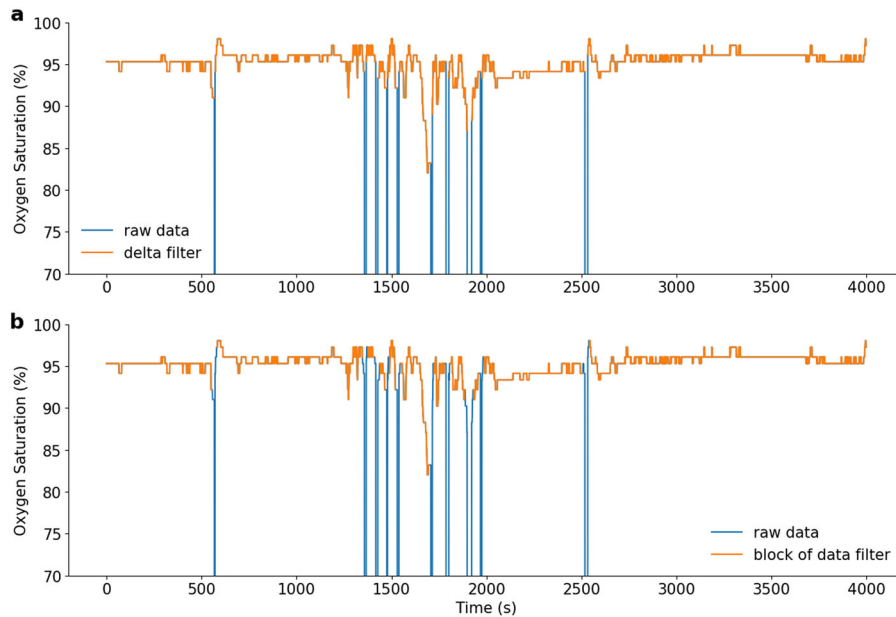
**Fig. 4 Example of preprocessing approaches. a** shows the time series preprocessed with the delta filter technique, and **b** shows the same time series preprocessed with the block of data technique.

distance between two windows can be defined as:

$$d(X(i), X(j)) = \max_{1 \le k \le m} |x(i + k - 1) - x(j + k - 1)|. \tag{6}$$

This biomarker was first used in the context of OSA diagnosis from oximetry data, in a study by Hornero et al.[33], with $m = 1$, $r = 0.25 \cdot \sigma$, where $\sigma$ is the standard deviation of the data. The database was composed of $SpO_2$ time series from subjects showing symptoms of sleep disordered breathing categorized into OSA-positive and OSA-negative groups according to the gold standard PSG. ApEn was used to diagnose OSA and reached 82.09% sensitivity and 86.96% specificity on a test set composed of $n = 113$ individuals.

*Sample entropy.* This is a non-linear biomarker that quantifies the irregularity in the data and has less bias compared to ApEn. It was used in the original work of Richman and Moorman[29], who proved the robustness of sample entropy within the context of physiological time-series analysis (on neonatal HRV). This biomarker has also been used by Behar et al.[6] for HRV analysis across different mammals. A pseudo-code for the implementation of SampEn is provided in Supplementary Methods.

*LZ complexity.* This biomarker was introduced by Lempel and Ziv[31] in 1976. Within the context of $SpO_2$ analysis, LZ evaluates the degree of complexity of spatiotemporal patterns in the $SpO_2$ signal. It has been largely used in the domain of medicine, especially in the domain of biomedical signal analysis, such as in the work of Amigó et al.[34], who used it on electroencephalogram time series, or by Álvarez et al.[35] to discriminate between OSA and non-OSA individuals. For the later work, it resulted in a sensitivity of 86.5%, a specificity of 77.6%, and an accuracy of 82.9%, when tested on a population of $n = 187$ patients, including 147 males and 40 females. A pseudo-code for the LZ measure is available in Supplementary Methods.

*Detrended fluctuation analysis.* DFA is a scaling analysis method that aims to represent the autocorrelation properties of the signal. A major advantage of this method is its robustness against non-stationarity of the signal. This biomarker was introduced by Peng et al.[36] to identify crossover behavior in signals. Larger fluctuations typical of repetitive desaturations lead to a higher DFA profile, while near-constant or slow, longer desaturations result in lower profiles. A pseudo-code for the implementation of DFA is provided in Supplementary Methods.

*Central tendency measure ($CTM_\rho$).* $CTM_\rho$ is a non-linear method first proposed by Cohen et al.[37], with the goal of assessing the degree of

variability in cardiac physiological data. The higher the variability in the $SpO_2$ signal (i.e., more desaturations/apneas), the lower the $CTM_\rho$. Indeed, as $CTM_\rho$ measures the number of points within a circular region of radius $\rho$, the higher the variability/dispersion the lower the number of points within the circle so the lower the $CTM_\rho$. This biomarker was used in the study of Álvarez et al.[35] on a dataset composed of $n = 187$ patients, with $\rho = 0.25$ for the purpose of OSA diagnosis. Their analysis resulted in a sensitivity of 90.01% and a specificity of 82.9%. $CTM_\rho$ is calculated as:

$$CTM_\rho = \frac{\sum_{i=1}^{N_{SpO_2}-2} \delta\rho(i)}{N_{SpO_2} - 2}, \tag{7}$$

$$\delta\rho(i) = \begin{cases} 1 & \text{if } \sqrt{(SpO_2(i+2) - SpO_2(i+1))^2 + (SpO_2(i+1) - SpO_2(i))^2} < \rho \cdot \\ 0 & \text{else} \end{cases} \tag{8}$$

## Periodicity measures

Consecutive apneic events create some periodicity in the oxygen saturation time series. This periodicity can be quantified through techniques, such as frequency analysis, phase-rectified signal averaging (PRSA), and autocorrelation.

*Phase-rectified signal averaging.* PRSA is a signal processing technique introduced by Bauer et al.[38] to detect and quantify quasi-periodic oscillations in a noisy non-stationary signal. The method also identifies patterns in increasing and decreasing regions of the signal. A PRSA window can be defined as:

$$\bar{x}(k) = \frac{1}{M} \sum_{i=1}^{M} X_i[k], \quad \text{for } -L \le k < L, \tag{9}$$

where $X_i$ is the window of length $2L$ around the anchor point $x(i)$ and $M$ is the number of anchor points. An anchor point is a decreasing point in the signal: $x(i) < x(i-1)$, such that the decreasing part (negative slope) of the desaturation is always within the window. It can also be defined as increasing points in order to investigate patterns in the resaturation part of the event. Figure 5 shows an example of PRSA computation on an oximetry time series for $L = 10$ and $M = 10$ anchor points.

Within the context of OSA diagnosis, PRSA biomarkers were used in the study of Deviaene et al.[22] and evaluated on three datasets: the Sleep Heart Health Study (SHHS) dataset[39–42], the Apnea-ECG[43] dataset, and a third set recorded at the sleep laboratory of the University Hospital Leuven. Five PRSA biomarkers were found significant and these are the ones that we implemented[22].
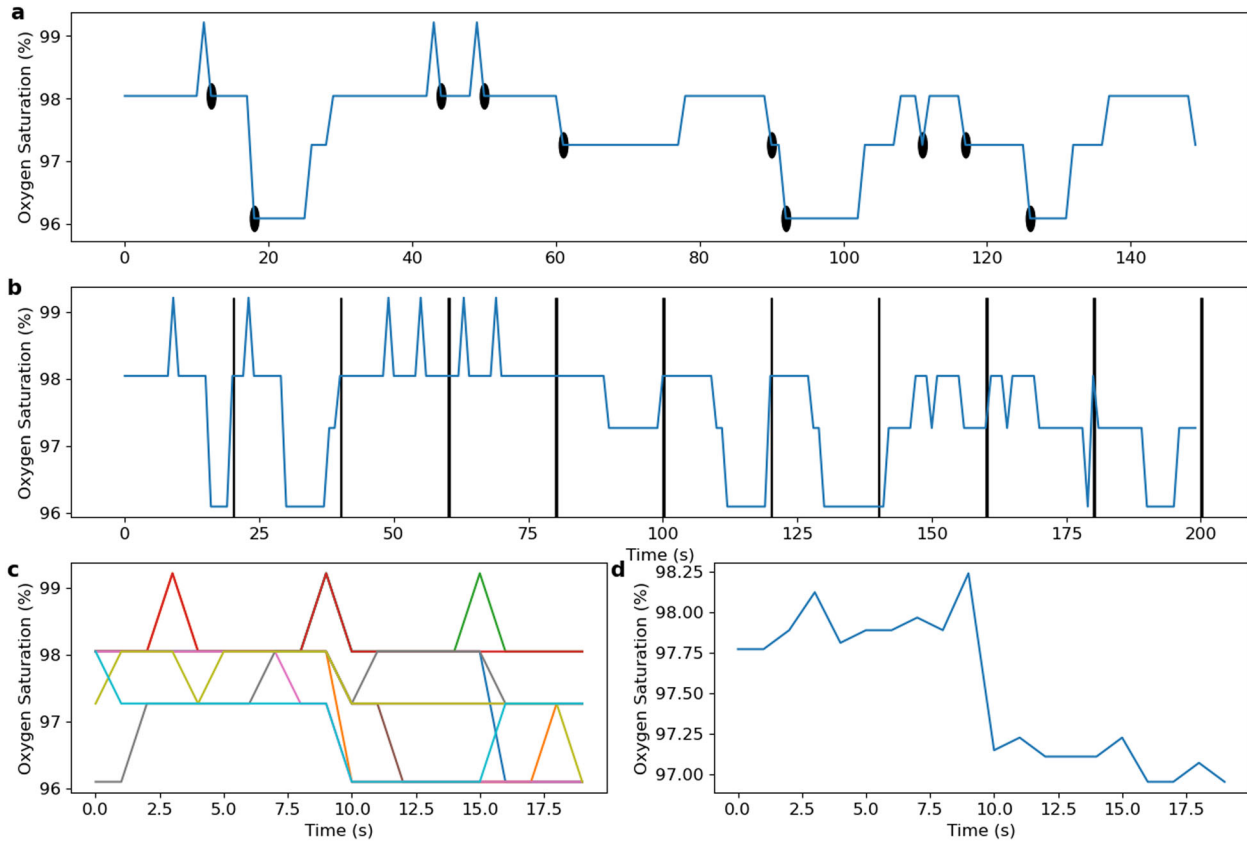
**Fig. 5 PRSA analysis of the oximetry time series.** In **a**, the oximetry time series, with anchor points highlighted. In **b**, concatenation of PRSA windows, defined around each anchor points. In **c**, PRSA windows of top of each other. In **d**, the average of all the PRSA windows.

- $PRSAd_c$, the capacity of the window, defined as

$$PRSAd_c = \frac{\overline{x}(0) + \overline{x}(1) - \overline{x}(-1) - \overline{x}(-2)}{4}. \quad (10)$$

- $PRSAd_{ad}$, the amplitude differences; this is the difference between max and min values.
- $PRSAd_{os}$, the overall slope of the window; the window is linearly approximated, and the slope is retained.
- $PRSAd_{sb}$, the slope before the anchor point.
- $PRSAd_{sa}$, the slope after the anchor point.

*Autocorrelation.* $AC(k)$ measures the degree of correlation between values of the same variable. This is achieved by computing the correlation between the original $SpO_2$ time series and a shifted version of it. The analysis of AC can be used to find repeating patterns, such as a periodic signal. Mathematically, AC can be defined as:

$$AC(k) = \sum_{i=1}^{N-k} SpO_2(i) * SpO_2(i+k). \quad (11)$$

*Power spectral analysis (power spectral density ($PSD_{total}$, $PSD_{band}$, $PSD_{ratio}$, $PSD_{peak}$).* In Zamarrón et al.[44], the authors analyzed the PSD curve of the oximetry time series. They defined a spectral band of interest for oximetry analysis within the context of OSA as 0.014–0.033 Hz. Zamarrón et al.[45] assessed PSD biomarkers on a total of 250 subjects between the ages of 21 and 82 years and obtained a sensitivity of 78.2% and a specificity of 89.0%. Figure 6 shows the differences in the spectral signal between a non-OSA and an OSA patient.

- $PSD_{total}$ corresponds to the area defined by the power spectrum:

$$PSD_{total} = \sum_{i=1}^{NFFT} X(i), \quad (12)$$

where $X$ is the amplitude of the PSD function, estimated by the Welch's method[46] using a hamming window, and NFFT is the number of points in the PSD signal.

- $PSD_{band}$ corresponds to the energy within the band 0.014–0.033 Hz:

$$PSD_{band} = \sum_{i=N_1}^{N_2} X(i), \quad (13)$$

where $N_1$ and $N_2$ are the limits of the summation between 0.014 and 0.033 Hz.

- $PSD_{ratio}$ corresponds to the ratio between the power (area) within the spectral band 0.017–0.033 Hz and $PSD_{total}$.

$$PSD_{ratio} = \frac{\sum_{i=N_1}^{N_2} X(i)}{\sum_{i=0}^{NFFT} X(i)}, \quad (14)$$

- $PSD_{peak}$ corresponds to the peak amplitude of the PSD within the band 0.014–0.033 Hz.

$$PSD_{peak} = \max_{N_1 < i < N_2} \{X(i)\}, \quad (15)$$

where $i$ is the index of the power spectrum signal.

### Desaturation measures

Desaturations can occur as a consequence of conditions such as sleep disordered breathing, and can be characterized by descriptors such as their lengths and depths. For example, a study by Kulkas et al.[47] studied the gender difference in the distribution of the desaturation lengths, depths, and areas caused by hypopnea and apnea events. Desaturations are not only caused by apnea or hypopnea events[48] and thus desaturation events and their statistical descriptors may capture the expression of other conditions during sleep.
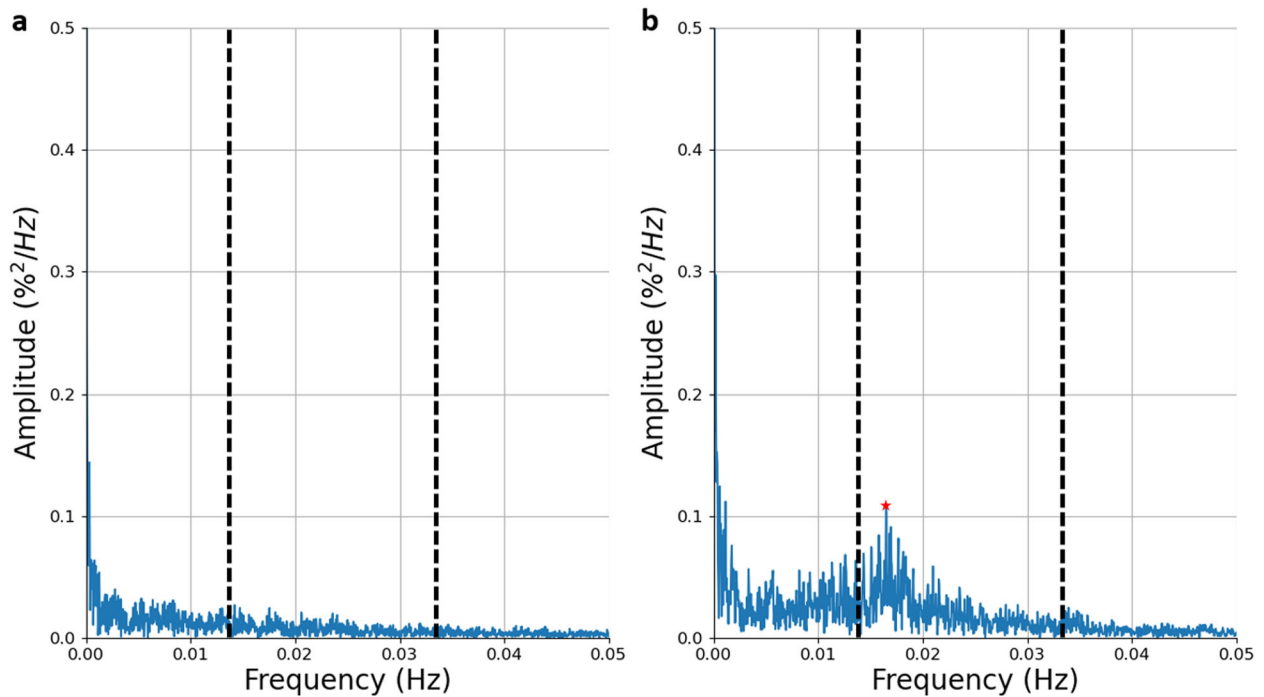
**Fig. 6 Power spectral density (PSD) of the oximetry time series.** Example on **a** a non-OSA individual and **b** a patient with moderate OSA. The band 0.014–0.033 Hz is delineated by the vertical dashed lines. In **b**, the peak within this band is located at $f = 0.017$ Hz, as highlighted by the red star.
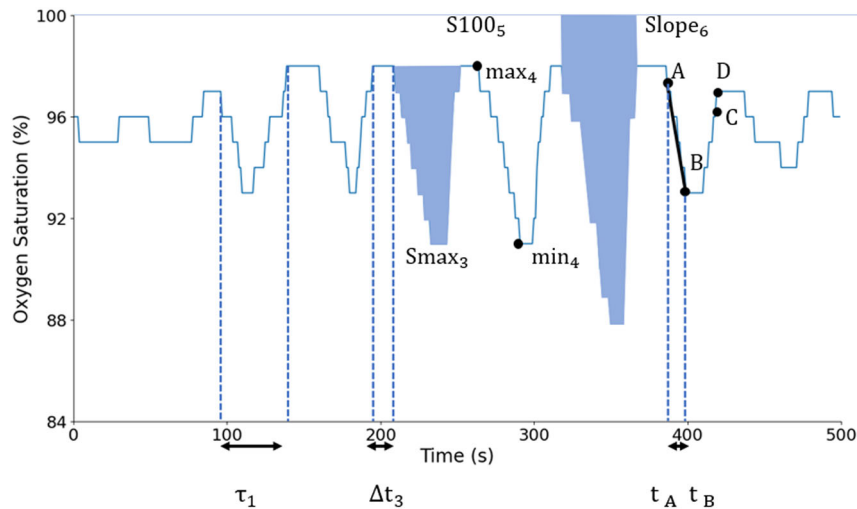


**Fig. 7 Desaturation biomarkers.** A, B, C, and D are the fiducials points delimiting a desaturation event.

*Oxygen desaturation index ($ODI_x$).* The $ODI_x$ corresponds to the average number of desaturation events per hour. A desaturation is defined as a $SpO_2$ drop of $x$% below the baseline. The $ODI_x$ is a widely used measure in the field of sleep medicine, where desaturations are characteristic of apnea and hypopnea events[49]. Indeed, obstruction of the airway leads to reduced entrance of oxygenated air to the lungs, which leads to a drop in oxygenated hemoglobin until airway patency is restored. These manifest as transient hypoxemic events or desaturations. Traditionally $x = 3$ or 4%. There exist many implementations of the ODI with variable definitions. In the present work, the implementation of the $ODI_x$ detection algorithm was developed and validated by Behar et al.[50], building on the parent model and desaturation definition of Jung et al.[51]. Specifically, the model of Jung et al.[51] defines three fiducial points A, B, and C, to determine the occurrence of a desaturation. Fiducial point A is defined as the point where the $SpO_2$ value decreases by ≥1 and ≤3%, fiducial point B as the

value that reaches a minimum of at least 3% below A, and fiducial point C as the point where the $SpO_2$ value returns to a level either 1% below A or 3% above B. Some additional constraints are imposed, including that fluctuations in consecutive $SpO_2$ values should be <1% between A and B and >−1% between B and C. Finally, the time interval between A and C must be ≥10 and ≤60 s[51]. In the original paper, 90 s was used as the time limit, but to ensure capture of resaturation, 60 s is used here. An additional fiducial point D was defined as a point posterior to C at which the $SpO_2$ time series reaches a level of at least 1% below A and where the time interval between A and D is ≤60 s. From the detected desaturation events, several oximetry biomarkers can be computed (Fig. 7). Within our context of single-channel $SpO_2$ analysis, i.e., when no reference EEG channel is available to quantify sleep time, the $ODI_x$ is defined as:

$$ODI_x = \frac{N_{desat}}{TRT}, \tag{16}$$

where $N_{\text{desat}}$ is the number of desaturations in the signal and TRT is the total recording time in hours.

*Desaturation length (DL$_\mu$, DL$_\sigma$).* DL$_\mu$ is the mean and DL$_\sigma$ is the standard deviation across the entire length of the desaturation. These biomarkers contain information about the duration of the desaturation events in the SpO$_2$ signal. Indeed, ODI$_x$ only considers the number of desaturations but does not consider their length. The length of a desaturation is particularly important because it reflects how long an individual is under hypoxic stress. Furthermore, desaturations of various lengths will lead to a higher DL$_\sigma$. DL$_\mu$ and DL$_\sigma$ are defined as:

$$DL_\mu = \frac{1}{N_{\text{desat}}} \sum_{i=1}^{N_{\text{desat}}} \tau_i, \tag{17}$$

$$DL_\sigma = \sqrt{\frac{\sum_{i=1}^{N_{\text{desat}}} (\tau_i - DL_\mu)^2}{N_{\text{desat}}}}, \tag{18}$$

where $\tau_i$ corresponds to the duration of the *i*th oxygen desaturation event.

*Desaturation depth (DDmax$_\mu$, DDmax$_\sigma$, DD100$_\mu$, DD100$_\sigma$).* For a single desaturation, the desaturation depth is computed as the maximal minus the minimal SpO$_2$ value within the desaturation event, from point A to B. DDmax$_\mu$ is the mean and DDmax$_\sigma$ is the standard deviation across all individual desaturation depths. A similar pair of biomarkers can be engineered by computing the depth with respect to the 100% SpO$_2$ level, i.e., 100% minus the minimal SpO$_2$ value of a desaturation event. These are, respectively, denoted DD100$_\mu$ and DD100$_\sigma$ for the mean and the standard deviation computed across all desaturation depths. The idea of quantifying the desaturation depth has also been used by Terrill[10]. The desaturation depth may be an important factor for determination of the severity of OSA because it will reflect, for a given desaturation, the level of the hypoxic stress imposed. Desaturation depth also varies with sleep stages, and so there is value in capturing its variation across desaturations[52]. Furthermore, alternation of soft and deep desaturations will lead to high values of DDmax$_\sigma$ and DD100$_\sigma$. The two pairs of biomarkers are defined as:

$$DDmax_\mu = \frac{1}{N_{\text{desat}}} \sum_{i=1}^{N_{\text{desat}}} max_i - min_i, \tag{19}$$

$$DD100_\mu = \frac{1}{N_{\text{desat}}} \sum_{i=1}^{N_{\text{desat}}} 100 - min_i, \tag{20}$$

$$DDmax_\sigma = \sqrt{\frac{\sum_{i=1}^{N_{\text{desat}}} (max_i - min_i - DDmax_\mu)^2}{N_{\text{desat}}}}, \tag{21}$$

$$DD100_\sigma = \sqrt{\frac{\sum_{i=1}^{N_{\text{desat}}} (100 - min_i - DD100_\mu)^2}{N_{\text{desat}}}}, \tag{22}$$

where $max_i$ is the maximum value of desaturation *i* and $min_i$ is the minimum value of desaturation *i*. The variables $max_i$ and $min_i$ are illustrated in Fig. 7. The desaturation depth can also be described by extending the ODI to any threshold *x* and studying the cumulative frequency of the desaturations as a function of $x$[53].

*Desaturation slope (DS$_\mu$, DS$_\sigma$).* The downslope of the signal is calculated for each desaturation. The decreasing phase of the desaturation is linearly approximated. DS$_\mu$ is the mean and DS$_\sigma$ is the standard deviation of the slopes over all desaturation events. These biomarkers consider the slope of the desaturation, which is a different factor than the number, duration, or depth of the desaturations. Indeed, OSA and other pathologies may lead to sharp drops in SpO$_2$, which would lead to high DS$_\mu$ value. The slope of a specific desaturation can be written as:

$$Slope_i = \frac{B - A}{t_B - t_A}, \tag{23}$$

where $(A, t_A)$ is the point of inflexion (amplitude and timestamp) of the desaturation, and $(B, t_B)$ is the minimum point of the desaturation. Accordingly, the mean and standard deviation of slopes are computed as:

$$DS_\mu = \frac{1}{N_{\text{desat}}} \sum_{i=1}^{N_{\text{desat}}} Slope_i, \tag{24}$$

$$DS_\sigma = \sqrt{\frac{\sum_{i=1}^{N_{\text{desat}}} (Slope_i - DS_\mu)^2}{N_{\text{desat}}}}. \tag{25}$$

$Slope_i$ can be seen in Fig. 7.

*Desaturation area (DAmax$_\mu$, DAmax$_\sigma$, DA100$_\mu$, DA100$_\mu$).* The area of the desaturation is computed for each desaturation. DA100$_\mu$ is the mean and DA100$_\sigma$ is the standard deviation of the area across all individual desaturations, taking 100% SpO$_2$ as baseline. The area can also be computed by taking the maximum SpO$_2$ value of individual desaturations as baseline. DAmax$_\mu$ is the mean and DAmax$_\sigma$ is the standard deviation of the area across all individual desaturations. Whereas the *Desaturation length* biomarker considers the duration of the events (reflecting the time under hypoxic stress) and the *desaturation depth* biomarkers consider the depth of events (reflecting the strength of hypoxia), the desaturation area factorizes both the depth and the length of the desaturations. The two pairs of biomarkers can be mathematically written as:

$$DA100_\mu = \frac{1}{N_{\text{desat}}} \sum_{i=1}^{N_{\text{desat}}} S100_i, \tag{26}$$

$$DA100_\sigma = \sqrt{\frac{\sum_{i=1}^{N_{\text{desat}}} (S100_i - DA100_\mu)^2}{N_{\text{desat}}}}, \tag{27}$$

$$DAmax_\mu = \frac{1}{N_{\text{desat}}} \sum_{i=1}^{N_{\text{desat}}} Smax_i, \tag{28}$$

$$DAmax_\sigma = \sqrt{\frac{\sum_{i=1}^{N_{\text{desat}}} (Smax_i - DAmax_\mu)^2}{N_{\text{desat}}}}, \tag{29}$$

where $Smax_i$ is the area of the specific desaturation event integrated from the maximal (max) value of the desaturation event and $S100_i$ is the area of the specific desaturation event integrated from 100%. $Smax_i$ and $S100_i$ can be seen in Fig. 7.

*Time between desaturation (TD$_\mu$, TD$_\sigma$).* The average and standard deviation of time elapsed between two consecutive desaturation events can be used to capture some aspect of the temporal distribution of desaturation events. The two biomarkers can be computed as:

$$TD_\mu = \frac{1}{N_{\text{desat}} - 1} \sum_{i=2}^{N_{\text{desat}}} \Delta t_i, \tag{30}$$

$$TD_\sigma = \sqrt{\frac{\sum_{i=2}^{N_{\text{desat}}} (\Delta t_i - TD_\mu)^2}{N_{\text{desat}} - 1}}, \tag{31}$$

where $\Delta t_i$ is the time elapsed between desaturation *i* and desaturation $i - 1$. $\Delta t_i$ can be seen in Fig. 7.

## Measures of the hypoxic burden

*The percentage of oxygen desaturation events (POD$_x$).* The POD$_x$ is the overall duration of all desaturations, normalized by the total recording time. It was introduced by Kulkas et al.[18] in order to estimate the severity of OSA from the SpO$_2$ time series. It was used in the work of Watanabe et al.[54] to study the prognostic importance of novel oxygen desaturation measures in heart failure and central sleep apnea population samples. Non-survivors had a higher POD$_x$ compared with survivors ($19 \pm 13$ versus $11 \pm 6.4\%$; $p = 0.001$). By contrast, non-survivors did not differ significantly from survivors with respect to the AHI and CT90%. An adjusted logistic regression analysis revealed that the POD$_x$ was the best independent predictor of mortality. In the work by Kulkas et al.[18], the biomarker was computed on a dataset collected from 160 male patients with different levels of AHI severity. The correlation between AHI and POD$_x$ was high: $r^2 = 0.87$[18]. The PODx can be mathematically defined as:

$$POD_x = 100 \cdot \frac{\sum_{i=1}^{N_{\text{desat}}} \tau_i}{TRT}, \tag{32}$$

where $\tau_i$ (Fig. 7) corresponds to the duration of each oxygen desaturation event and *x* to the level of the desaturation. In their original publication, Watanabe et al.[54] set $x = 4\%$.
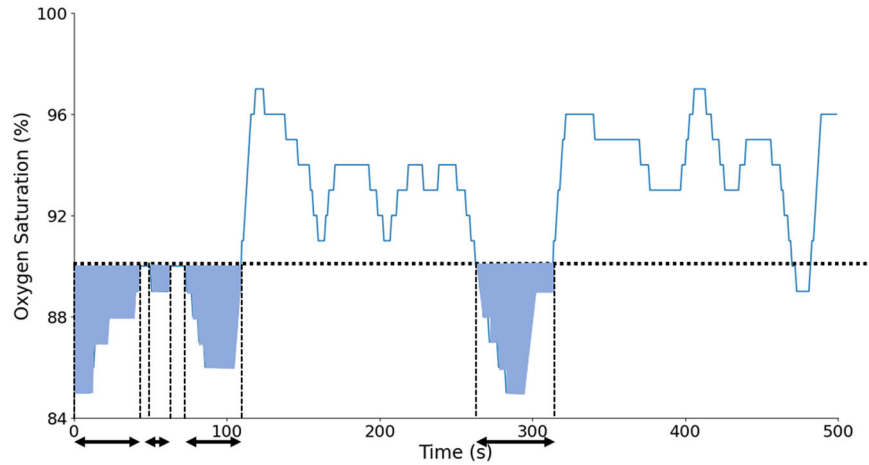
**Fig. 8 Illustration of CA90 and CT90 biomarkers.** The sum of all the blue areas corresponds to CA90, whereas the sum of all the length arrows corresponds to CT90.

*The area under the oxygen desaturation curve (AODmax, AOD100).* The $AOD_x$ was introduced by Kulkas et al.[18] in the same context as the $POD_x$, for the estimation of the sleep apnea–hypopnea syndrome. It was used in the work of Watanabe et al.[54], along with the POD biomarker. Survivors in this study appeared to have lower $AOD_x$ than non-survivors ($0.16 \pm 0.2$ versus $0.26 \pm 0.2$; $p = 0.08$). It represents the sum of the area of each desaturation event divided by TRT. This index was demonstrated to be an independent modulator of increased epicardial fat volume (EFV) in an acute myocardial infarction population sample[14] ($n = 105$). EFV is associated with adverse cardiovascular events after myocardial infarction. In the work of Kulkas et al.[18], this biomarker appeared to have moderate correlation with AHI: $r^2 \in [0.581-0.689]$, $p < 0.001$. It can be mathematically defined as:

$$\text{AODmax} = 100 \cdot \frac{\sum_{i=1}^{N_{\text{desat}}} \text{Smax}_i}{\text{TRT}}, \tag{33}$$

$$\text{AOD100} = 100 \cdot \frac{\sum_{i=1}^{N_{\text{desat}}} \text{S100}_i}{\text{TRT}}. \tag{34}$$

$\text{Smax}_i$ and $\text{S100}_i$ are illustrated in Fig. 7.

*Cumulative time ($CT_x$).* Percentage of the time spent below the $x$% oxygen saturation level. Typically, CT90 is used[55] but other thresholds such as 80 or 84% have also been assessed[27]. This biomarker is evaluated on the overall signal, i.e., not only on the desaturation events, and it might consequently capture hypoxic behaviors that are different from the desaturation events found in OSA. The biomarker is illustrated in Fig. 8. It is mathematically defined as:

$$CT_x = 100 \cdot \frac{\sum_{i=1}^{N_{\text{SpO2}}} t(x)_i}{\text{TRT} * \text{fs}}, \tag{35}$$

$$t(x)_i = \begin{cases} 1 & \text{if } \text{SpO2}_i < x \\ 0 & \text{else} \end{cases}, \tag{36}$$

where fs is the sampling frequency of the signal.

*Cumulative area ($CA_x$).* Total area under the $x$% oxygen saturation level. This biomarker was introduced by Watanabe et al.[54], with $x = 90\%$. Indeed, OSA patients tend to have a greater area under the baseline $x$ than non-OSA patients and then get a higher value for this biomarker. The biomarker is illustrated in Fig. 8. It can be defined as:

$$CA_x = 100 \cdot \frac{\sum_{i=1}^{N_{\text{SpO2}}} (x - \text{SpO2}(x)_i)}{\text{TRT} * \text{fs}}, \tag{37}$$

$$\text{SpO2}(x)_i = \begin{cases} \text{SpO2}_i & \text{if } \text{SpO2}_i < x \\ x & \text{else} \end{cases}. \tag{38}$$

### Evaluation database

In order to demonstrate the usability of the implemented oximetry biomarkers and to define some normality ranges, we used the SHHS[39–42]

database. SHHS was a multi-center cohort study conducted by the National Heart Lung & Blood Institute (ClinicalTrials.gov Identifier: NCT0000527) to determine the cardiovascular and other consequences of sleep-disordered breathing. In all, 6441 men and women, aged ≥40 years, were enrolled between November 1, 1995 and January 31, 1998. Institutional review board from the Technion-IIT Rappaport Faculty of Medicine was obtained under number 62-2019 in order to use this database for research. The variable "ahi_a0h3a" was used for the AHI in order to define the classes. To elaborate this variable, the AHI was computed as the average number of all apneas and hypopneas (with oxygen desaturation >3% or an arousal) per hour of sleep and following the American Academy of Sleep Medicine (AASM) 2012 rules[56]. OSA severity was defined with respect to the AHI, i.e., mild ($5 \le \text{AHI} < 15$), moderate ($15 \le \text{AHI} < 30$) or severe ($\text{AHI} \ge 30$). The Nonin XPOD 3011 pulse oximeter (Nonin, USA) was used for recording. The signal was sampled at 1 Hz and with a resolution of ±0.01%. The OBM were computed for patients with available recordings and at least 4 h of continuous $SpO_2$ tracing. This resulted in a total of 3806 individual patient recordings out of 5793 patients who participated in the first study visit (SHHS1). Among them, there were 1195 non-OSA, 1303 with mild OSA, 833 with moderate OSA, and 475 with severe OSA. OBMs were evaluated on these recordings in order to report reference ranges for each OBM.

### Statistical and regression analysis

The median and interquartile range of the $SpO_2$ biomarkers were computed for the following classes: non-OSA, mild, moderate, and severe OSA. Kruskal–Wallis test with post hoc analysis was performed. Statistical significance or non-significance was indicated as "$p < 0.05$", "$p < 0.001$", or "$p > 0.05$". Dunn post hoc analysis was performed between each pair of the classes. Multivariable linear regression was performed to assess the added value in combining OBM for the purpose of estimating the AHI. To this end, linear regression was performed between individual and combined sets of biomarkers and the AHI. For each model, the adjusted $R^2$ ($\overline{R}^2$) score was reported. $\overline{R}^2$ is defined as:

$$\overline{R}^2 = 1 - \frac{(1 - R^2)(\text{size} - 1)}{\text{size} - \text{pred} - 1}, \tag{39}$$

where pred is the number of predictors, and size is the total sample size.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## DATA AVAILABILITY

The SHHS database used in this research may be requested from sleepdata.org.

## CODE AVAILABILITY

The source code used in this research is available at physiozoo.com.

## REFERENCES

1. Lavie, L. Obstructive sleep apnoea syndrome - an oxidative stress disorder. *Sleep Med. Rev.* **7**, 35–51 (2003).
2. Budhiraja, R., Siddiqi, T. A. & Quan, S. F. Sleep disorders in chronic obstructive pulmonary disease: etiology, impact, and management. *J. Clin. Sleep Med.* **11**, 259–270 (2015).
3. Malik, M. et al. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Eur. Hear. J.* **17**, 354–381 (1996).
4. Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-aho, P. O. & Karjalainen, P. A. Kubios HRV - heart rate variability analysis software. *Comput. Methods Prog. Biol.* **113**, 210–20 (2014).
5. Vest, A. N. et al. Benchmarking heart rate variability toolboxes. *J. Electrocardiol.* **50**, 744–747 (2017).
6. Behar, J. A. et al. PhysioZoo: a novel open access platform for heart rate variability analysis of mammalian electrocardiographic data. *Front. Physiol.* **9**, 1–14 (2018).
7. Uddin, M. B., Chow, C. M. & Su, S. W. Classification methods to detect sleep apnea in adults based on respiratory and oximetry signals: a systematic review. *Physiol. Meas.* **39**, 2018 (2018).
8. del Campo, F. et al. Oximetry use in obstructive sleep apnea. *Expert Rev. Resp. Med.* **12**, 665–681 (2018).
9. Mendonça, F., Mostafa, S. S., Ravelo-García, A. G., Morgado-Dias, F. & Penzel, T. A review of obstructive sleep apnea detection approaches. *IEEE J. Biomed. Health Inform.* **23**, 825–837 (2019).
10. Terrill, P. I. A review of approaches for analysing obstructive sleep apnoea-related patterns in pulse oximetry data. *Respirology* **25**, 475–485 (2020).
11. Rashid, N. H. et al. The value of oxygen desaturation index for diagnosing obstructive sleep apnea: a systematic review. *Laryngoscope.* https://doi.org/10.1002/lary.28663 (2020).
12. Azarbarzin, A. et al. The hypoxic burden of sleep apnoea predicts cardiovascular disease-related mortality: the osteoporotic fractures in Men Study and the Sleep Heart Health Study. *Eur. Heart J.* **40**, 1149–1157 (2019).
13. Goldberger, A. L. et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101**, E215–E220 (2000).
14. Linz, D. et al. Nocturnal hypoxemic burden is associated with epicardial fat volume in patients with acute myocardial infarction. *Sleep Breath.* **22**, 703–711 (2018).
15. Dewan, N. A., Nieto, F. J. & Somers, V. K. Intermittent hypoxemia and OSA: implications for comorbidities. *Chest* **147**, 266–274 (2015).
16. Orrù, G. et al. Obstructive Sleep Apnea, oxidative stress, inflammation and endothelial dysfunction-an overview of predictive laboratory biomarkers. *Eur. Rev. Med. Pharmacol. Sci.* **24**, 6939–6948 (2020).
17. Suen, C. et al. Sleep study and oximetry parameters for predicting postoperative complications in patients with OSA. *Chest* **155**, 855–867 (2019).
18. Kulkas, A., Tiihonen, P., Julkunen, P., Mervaala, E. & Töyräs, J. Novel parameters indicate significant differences in severity of obstructive sleep apnea with patients having similar apnea-hypopnea index. *Med. Biol. Eng. Comput.* **51**, 697–708 (2013).
19. Fletcher, E. C., Miller, J., Divine, G. W., Fletcher, J. G. & Miller, T. Nocturnal oxyhemoglobin desaturation in COPD patients with arterial oxygen tensions above 60 mm Hg. *Chest* **92**, 604–608 (1987).
20. Resta, O. et al. Sleep related O2 desaturation in COPD patients with normoxaemia and mild hypoxyaemia. *Boll. Soc. Ital. Biol. Sper.* **74**, 91–98 (1998).
21. Quaresima, V. & Ferrari, M. COVID-19: efficacy of prehospital pulse oximetry for early detection of silent hypoxemia. *Crit. Care* **24**, 501 (2020).
22. Deviaene, M. et al. Automatic screening of sleep apnea patients based on the SpO2 signal. *IEEE J. Biomed. Health Inform.* **23**, 607–617 (2019).
23. Taha, B. H. et al. Automated detection and classification of sleep-disordered breathing from conventional polysomnography data. *Sleep* **20**, 991–1001 (1997).
24. Buekers, J. et al. Wearable finger pulse oximetry for continuous oxygen saturation measurements during daily home routines of patients with chronic obstructive pulmonary disease (COPD) over one week: observational study. *J. Med. Internet Res.* **21**, 1–14 (2019).
25. Xie, B. & Minn, H. Real-time sleep apnea detection by classifier combination. *IEEE Trans. Inf. Technol. Biomed.* **16**, 469–477 (2012).
26. Pepin, J. L., Levy, P., Lepaulle, B., Brambilla, C. & Guilleminault, C. Does oximetry contribute to the detection of apneic events? Mathematical processing of the SaO2 signal. *Chest* **99**, 1151–1157 (1991).
27. Magalang, U. J. et al. Prediction of the apnea-hypopnea index from overnight pulse oximetry. *Chest* **124**, 1694–1701 (2003).
28. Pincus, S. M. Approximate entropy as a measure of system complexity. *Proc. Natl Acad. Sci. USA* **88**, 2297–2301 (1991).
29. Richman, J. S. & Moorman, J. R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ.* **278**, H2039–H2049 (2000).
30. Goldberger, A. L., Peng, C. K. & Lipsitz, L. A. What is physiologic complexity and how does it change with aging and disease? *Neurobiol. Aging* **23**, 23–26 (2002).
31. Lempel, A. & Ziv, J. On the complexity of finite sequences. *IEEE Trans. Inf. Theory* **22**, 75–81 (1976).
32. Ary, L., Goldberger, D. R. R. & B., J. Chaos and fractals in human physiology. *Sci. Am.* **262**, 42–49 (1990).
33. Hornero, R., Álvarez, D., Abásolo, D., Del Campo, F. & Zamarrón, C. Utility of approximate entropy from overnight pulse oximetry data in the diagnosis of the obstructive sleep apnea syndrome. *IEEE Trans. Biomed. Eng.* **54**, 107–113 (2007).
34. Amigó, J. M., Szczepański, J., Wajnryb, E. & Sanchez-Vives, M. V. Estimating the entropy rate of spike trains via Lempel-Ziv complexity. *Neural Comput.* **16**, 717–736 (2004).
35. Álvarez, D., Hornero, R., Abásolo, D., Del Campo, F. & Zamarrón, C. Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection. *Physiol. Meas.* **27**, 399–412 (2006).
36. Peng, C. K., Havlin, S., Stanley, H. E. & Goldberger, A. L. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos* **5**, 82–87 (1995).
37. Cohen, M. E., Hudson, D. L. & Deedwania, P. C. Applying continuous chaotic modeling to cardiac signal analysis. *IEEE Eng. Med. Biol. Mag.* **15**, 97–102 (1996).
38. Bauer, A. et al. Phase-rectified signal averaging detects quasi-periodicities in nonstationary data. *Phys. A Stat. Mech. Appl.* **364**, 423–434 (2006).
39. Zhang, G. Q. et al. The National Sleep Research Resource: towards a sleep data commons. *J. Am. Med. Inform. Assoc.* **25**, 1351–1358 (2018).
40. Quan, S. et al. The Sleep Heart Health Study: design, rationale, and methods. *Sleep* **20**, 1077–1085 (1997).
41. Redline, S. et al. Methods for obtaining and analyzing unattended polysomnography data for a multicenter study. *Sleep Heart Health Res. Group Sleep* **21**, 759–767 (1998).
42. Dean, D. A. et al. Scaling up scientific discovery in sleep medicine: The National Sleep Research Resource. *Sleep* **39**, 1151–1164 (2016).
43. Penzel, T., Rg, G. B. M., Goldberges, M. A. L. & Peter, H. The Apnea-ECG Database. *Comput. Cardiol.* **27**, 255–258 (2000).
44. Zamarrón, C., Gude, F., Barcala, J., Rodriguez, J. R. & Romero, P. V. Utility of oxygen saturation and heart rate spectral analysis obtained from pulse oximetric recordings in the diagnosis of sleep apnea syndrome. *Chest* **123**, 1567–1576 (2003).
45. Zamarrón Sanz, C., Romero, P. V., Rodriguez, J. R. & Gude, F. Oximetry spectral analysis in the diagnosis of obstructive sleep apnoea. *Clin. Sci.* **97**, 467–473 (1999).
46. Welch, P. D. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **9**, 2077 (1967).
47. Kulkas, A., Duce, B., Leppänen, T., Hukins, C. & Töyräs, J. Gender differences in severity of desaturation events following hypopnea and obstructive apnea events in adults during sleep. *Physiol. Meas.* **38**, 1490–1502 (2017).
48. Behar, J. A. From sleep medicine to medicine during sleep: a new paradigm. *Sleep* **43**, 1–3 (2020).
49. Behar, J. A. et al. Feasibility of single channel oximetry for mass screening of obstructive sleep apnea. *EClinicalMedicine* **11**, 81–88 (2019).
50. Behar, J. A., Palmius, N., Penzel, T., Bittencourt, L. & Tufik, S. Single-channel oximetry monitor versus in-lab polysomnography oximetry analysis: does it make a difference? *Accept. Physiol. Meas.* **41**, 4 (2020).
51. Jung, D. W. et al. Real-time automatic apneic event detection using nocturnal pulse oximetry. *IEEE Trans. Biomed. Eng.* **65**, 706–712 (2018).
52. Findley, L. J., Wilhoit, S. C. & Suratt, R. M. Apnea duration and hypoxemia during REM sleep in patients with obstructive sleep apnea. *Chest* **87**, 432–436 (1985).
53. Terrill, P. I., Dakin, C., Edwards, B. A., Wilson, S. J. & MacLean, J. E. A graphical method for comparing nocturnal oxygen saturation profiles in individuals and populations: application to healthy infants and preterm neonates. *Pediatr. Pulmonol.* **53**, 645–655 (2018).
54. Watanabe, E. et al. Prognostic importance of novel oxygen desaturation metrics in patients with heart failure and central sleep apnea. *J. Card. Fail.* **23**, 131–137 (2017).
55. Olson, L. G., Ambrogetti, A. & Gyulay, S. G. Prediction of sleep-disordered breathing by unattended overnight oximetry. *J. Sleep Res.* **8**, 51–55 (1999).
56. Berry, R. B. et al. Rules for scoring respiratory events in sleep: Update of the 2007 AASM manual for the scoring of sleep and associated events. *J. Clin. Sleep Med.* **8**, 597–619 (2012).

57. Deviaene, M. et al. Pulse oximetry markers for cardiovascular disease in sleep apnea. In *2019 Computing in Cardiology (CinC)* 3–6 (IEEE, 2019).

## AUTHOR CONTRIBUTIONS

J.A.B. conceived and designed the research; J.L. created the toolbox and analyzed the data. D.A. and F.d.C. provided clinical guidance on reviewing evidence-based biomarkers, the conduct of the analysis, and interpretation of the results. A.A.R. and A.A. provided technical guidance on the elaboration of the resource, i.e., toolbox structuring and its integration in the PhysioZoo resource. J.A.B. and J.L. drafted the manuscript; J.L. prepared the figures; all authors edited and revised the manuscript and approved the final version.

## COMPETING INTERESTS

J.A.B. holds shares in SmartCare Analytics Ltd. The remaining authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41746-020-00373-5.

**Correspondence** and requests for materials should be addressed to J.A.B.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.