Journal of Sleep Research

RESEARCH ARTICLE

Ordinal Sleep Depth: A Data-Driven Continuous Measurement of Sleep Depth

Erik-Jan Meulenbrugge[1,2] | Haoqi Sun[1,2] | Wolfgang Ganglberger[1,2] | Samaneh Nasiri[3,4] | Robert J. Thomas[2,5] | M. Brandon Westover[1,2]

[1]Department of Neurology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA | [2]Division of Sleep Medicine, Harvard Medical School, Boston, Massachusetts, USA | [3]Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA | [4]Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, Georgia, USA | [5]Division of Pulmonary, Critical Care and Sleep, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

Correspondence: M. Brandon Westover (bwestove@bidmc.harvard.edu)

Received: 20 November 2024 | **Revised:** 10 March 2025 | **Accepted:** 10 April 2025

Funding: Dr. Westover's group was supported by grants from the NIH (RF1AG064312, RF1NS120947, R01AG073410, R01HL161253, R01NS126282, R01AG073598, R01NS131347, R01NS130119), and NSF (2014431). Dr. Robert Thomas is supported by RF1NS120947.

Keywords: arousal probability | cognitive impairment | deep learning | EEG | sleep depth | sleep staging

ABSTRACT

Conventional sleep staging categorises sleep into discrete stages, which may not capture the continuous nature of sleep depth. We aimed to develop a data-driven continuous measure of sleep depth—ordinal sleep depth (OSD)—using a deep learning framework, and to evaluate its correlation with arousal probability and its association with age, sex, sleep-disordered breathing (SDB) and cognitive impairment. We used 21,787 polysomnography recordings from 18,116 unique patients. A convolutional neural network was trained on 3-s EEG segments to estimate sleep depth continuously, incorporating ordinal regression for the ordered nature of non-REM stages. OSD was compared with the odds ratio product (ORP). Correlations with sleep stages, Arousal Index and clinical variables were assessed. OSD showed a strong linear correlation with arousal probability (Pearson's $r = 0.994$), slightly outperforming ORP ($r = 0.923$). Both OSD and ORP reflected expected decreases in sleep depth with advancing age and demonstrated that females have significantly deeper sleep than males across several stages. OSD more accurately captured sleep depth reductions associated with SDB and increasing levels of cognitive impairment, showing significant reductions across all non-REM stages in patients with an increased level of cognitive impairment. OSD as a data-driven measure of sleep depth correlates strongly with arousal probability and effectively captures variations associated with age, sex, SDB and cognitive impairment. The results validate depth as an important dimension of sleep. OSD and ORP provide a nuanced understanding of sleep architecture with physiological and pathological implications.

1 | Introduction

In clinical practice, a fundamental step in analysing sleep electroencephalography (EEG) is to divide the recording into a series of nonoverlapping 30-s epochs and categorise the pattern of brain activity in each epoch into one of the five discrete 'stages', defined by the American Academy of Sleep Medicine (AASM) (Iber et al. 2007); wakefulness (W), non-rapid eye movement (NREM) stages N1, N2, N3 and rapid eye movement stage (REM). While this scheme facilitates manual scoring, it provides only a coarse characterisation of what, physiologically, is a continuous process (Picchioni et al. 2011). Estimates of regional and global sleep can in

Robert J. Thomas and M. Brandon Westover are co-senior authors.

fact be recognised to sub-second time resolutions (Andrillon and Oudiette 2023). A continuous approach to state characterisation with a finer resolution could more accurately quantify sleep depth and quality. For example, it is known that the arousal threshold increases gradually as sleep depth increases (Younes et al. 2015), which reinforces the notion that sleep depth lies along a continuum.

Machine learning approaches have been used to analyse large-scale sleep data sets; however, most have focused on getting computers to learn conventional AASM scoring (Sun et al. 2017; Biswal et al. 2018; Stephansen et al. 2018; Patanaik et al. 2018). Some sleep depth algorithms have been proposed based on features of the EEG, such as recurrence quantification analysis (Carrubba et al. 2012) and various forms of EEG spectral power analysis. (Younes et al. 2015; Asyali et al. 2007; Saastamoinen et al. 2006) Younes et al. (Hannun et al. 2019) developed a method to measure sleep depth based on consecutive 3-s EEG segments, called the odds ratio product (ORP). ORP shows reasonable correlation with conventional sleep stages, while all stages show widely varying sleep depth, demonstrating a limitation of the conventional characterisation. ORP demonstrates a better correlation with the Arousal Index (ArI), the probability of wakefulness or arousal within 30 s at a given state, than conventional sleep staging and thus is attractive as a continuous measure of sleep depth. ORP, though automated, is based on hand-engineered features and domain knowledge, which might not fully capture the complex nature of sleep. A data-driven approach could potentially offer a more nuanced, continuous measure of sleep depth.

This study aimed to develop a continuous measure of sleep depth directly from existing AASM-scored EEG signals in a data-driven way, without relying on human-designed features. To do so, we designed a convolutional neural network (CNN), which for every consecutive 3-s segment of EEG data produces a measure of how similar brain activity within the current segment is to brain activity in the awake state. We evaluated the correlation of OSD with conventional sleep stages, ArI and its association with age, sex, sleep-disordered breathing (SDB) and cognitive impairment, while additionally comparing OSD with ORP.

2 | Materials and Methods

2.1 | Data set

Retrospective analysis of the polysomnography (PSG) data was approved by the MGB (protocol # 2013P001024) and BIDMC Institutional Review Boards (# 2016P000058) without requiring additional consent for its use in this study. Data were recorded as part of routine clinical care in the Massachusetts General Hospital (MGH) Sleep Laboratory from 2009 to 2020. All PSGs used in the present analysis were recorded using equipment from Natus Neuro, CA, USA. EEG electrodes were placed in six locations according to the international 10–20 system, with each channel referenced to the contralateral mastoid: F3-M2, F4-M1, C3-M2, C4-M1, O1-M2 and O2-M1. Each PSG was annotated by one sleep technician. In total, annotations came from seven sleep technicians.

A total of 21,787 PSG recordings (18,116 unique patients) were used for this study. Before creating the development–validation split, we split the data in two categorised groups, 'No known cognitive decline' (19,302 recordings, 16,139 patients) and 'Known cognitive decline' (2485 recordings, 1977 patients) without patient overlap. Recordings from the 'No known cognitive decline' were split into 85% development and 15% validation, using a patient-wise split, while data from the 'Known cognitive decline' group were added to the validation set. Among the 'Known cognitive decline' group, 1914 recordings from 1479 patients involve self-reported memory issues without an mild cognitive impairment (MCI) or dementia diagnosis, 573 recordings from 459 patients are associated with a MCI diagnosis and 397 recordings from 316 patients indicate a diagnosis of dementia.

Ultimately, this led to the use of 15,940 records from 13,004 patients for development and 5847 records from 5112 unique patients for model validation. A representation of the split is shown in Figure 1. All reported results are generated from the records in the validation set. No patient overlap between the development and validation patients was present. The patient demographics are shown in Table 1.

2.2 | Pre-Processing

The EEG signals were sampled at 200 Hz, after which a notch filter at 60 Hz and a bandpass filter between 0.3 and 35 Hz were applied, following AASM guidelines. No further pre-processing was applied.

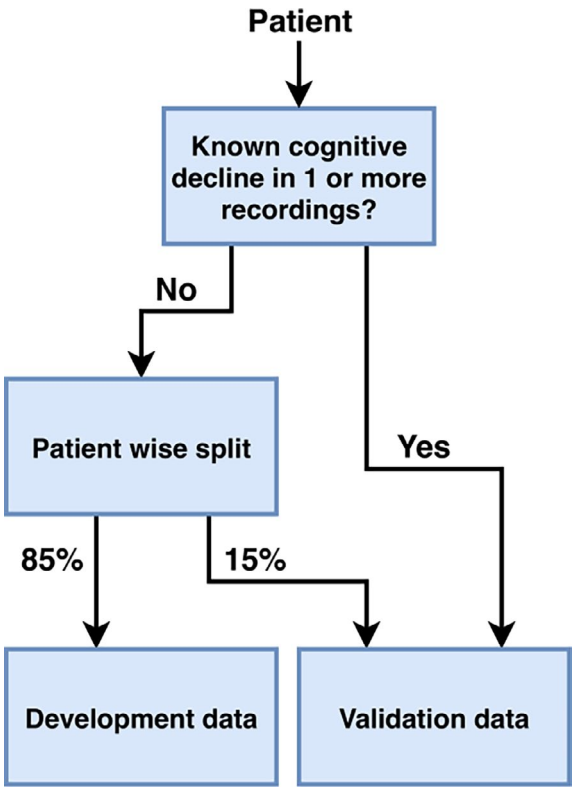

FIGURE 1 | Patient split: this figure visualises the decision flow used in the creation of the development and validation split. Patients are automatically assigned to the validation set if one or more recordings include cognitive decline. All other patients are split using an 85% development and 15% validation split.

TABLE 1 | Patient demographics.

	Development	Validation
Demographic characteristics		
Age in years (mean ± SD)	50.9 ± 16.6	59.1 ± 15.0
Gender, % man	57.4	57.2
BMI (mean ± SD)	31.7 ± 7.9	31.6 ± 7.4
Clinical characteristics		
PSG used (unique patients)	15,940 (13,004)	5847 (5112)
Unknown cognitive condition PSGs (unique patients)	9391 (7949)	1681 (1635)
No cognitive impairment PSGs (unique patients)	6549 (5320)	1282 (1235)
Symptomatic PSGs (unique patients)	0 (0)	1914 (1479)
MCI PSGs (unique patients)	0 (0)	573 (459)
Dementia PSGs (unique patients)	0 (0)	397 (316)
AHI (mean ± SD)	12.2 ± 14.5	13.8 ± 14.9
Arousal Index (mean ± SD)	9.1 ± 14.4	10.4 ± 15.0
Leg Movement Index (mean ± SD)	24.9 ± 38.5	32.2 ± 51.3
Periodic Leg Movement Index (mean ± SD)	15.8 ± 32.8	22.3 ± 45.8
PSG characteristics		
Recording time, hours (mean ± SD)	7.3 ± 1.0	7.3 ± 0.9
Sleep efficiency, % (mean ± SD)	82.2 ± 14.2	81.2 ± 14.7
WASO, minutes (mean ± SD)	65.5 ± 55.4	71.1 ± 57.7

Abbreviations: AASM, Academy of Sleep Medicine; AHI, Apnea–Hypopnea Index; ArI, Arousal Index; BIDMC, Beth Israel Deaconess Medical Center; BM, Benjamini–Hochberg; CNN, convolutional neural network; EEG, electroencephalography; MBG, Mass General Brigham; MGH, Massachusetts General Hospital; N1, non-rapid eye movement Stage 1; N2, non-rapid eye movement Stage 2; N3, non-rapid eye movement Stage 3; NREM, non-rapid eye movement stages; OLS, ordinal least squares; ORL, ordinal regression layer; ORP, odds ratio product; OSD, ordinal sleep depth; PSG, polysomnography; REM, rapid eye movement stage; SDB, sleep-disordered breathing; W, wakefulness.

2.3 | Deep Learning Network Architecture

We adapted a deep neural network architecture proposed by Hannun et al. for classifying ECG signals (Hannun et al. 2019). Our model consists of five convolutional blocks, each with a one-dimensional convolutional layer followed by batch normalisation, rectified linear unit activation, max pooling and a 10% dropout layer. For convolutional layers, the number of filters increases by 32 for each consecutive block, starting at 64. The convolutional kernel size of 17 is consistent across all blocks. For the batch normalisation layer, the momentum parameter was set to 0.99. The max pooling layer down-samples the signal by a factor of 2. After the five convolution blocks, the data are flattened, and two branches are created. Each branch takes the flattened layer as input and has 1 dense layer with 512 nodes followed by a dense layer of 5 nodes and a classification layer. The first branch has a categorical sleep staging objective; the second branch has an ordinal regression objective. A schematic representation of the model architecture is given in Figure 2.

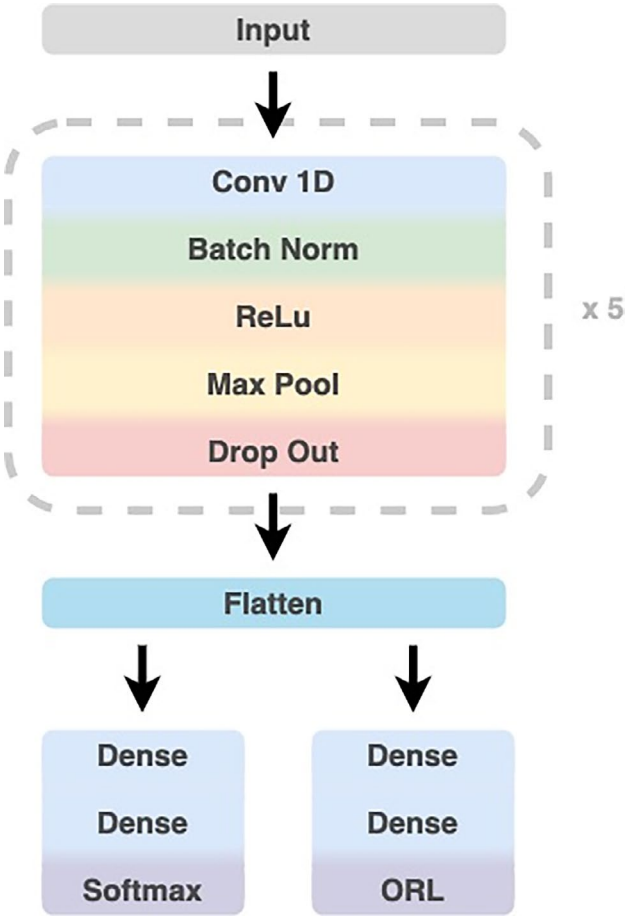

FIGURE 2 | Architecture of the deep learning model: The model processes a 3-s input EEG signal through five convolutional blocks, each with a one-dimensional convolutional layer (kernel size 17), batch normalisation, rectified linear unit activation, max pooling (down-sampling by 2) and a dropout layer (10%). As the signal passes through the blocks, it is progressively reduced to a set of features, which are flattened and reduced to 500 nodes. This multi-task network then splits into two branches: one for categorical sleep staging with a Softmax layer, and the other for ordinal regression, capturing the ordered nature of NREM sleep stages.

We trained the model using two complementary objectives: (1) main task: an *ordinal regression objective*, which encourages the model to learn that sleep stages are ordered, that is, NREM

sleep stages increase in depth: W < N1 < N2 < N3 (i.e., NREM sleep stages are ordinal variables) and (2) auxiliary task: a *classification objective*, which encourages the model to estimate the likelihood that a sleep expert would classify the 30-s epoch in which the current 3-s EEG segment is embedded as N1, N2, N3, R or W, treating these as categorical variables. To train the categorical sleep staging branch, a Softmax layer with categorical cross-entropy loss is used, while a separate loss function is applied for ordinal sleep staging. Unlike categorical cross-entropy, the ordinal loss accounts for the ranked order of sleep stages. REM is excluded from the ordinal loss calculation by setting the class weights of REM samples to zero, as REM and NREM sleep are regulated by distinct physiological systems (Nielsen 2000). Although the final model can be evaluated on REM epochs, they do not influence parameter tuning for the ordinal loss function.

For the ordinal sleep staging branch, an ordinal regression layer (ORL) was created. The ORL has a trainable weight matrix $\boldsymbol{W}$, bias b and set of thresholds μ (between 0 and 1) that discretize a continuous value into stages.

The ORL can be mathematically described using the equation: (Rennie and Srebro 2005)

$$P(y|z) = \sigma(-\mathrm{d}(z, \boldsymbol{\mu})) = \sigma(-\mathrm{d}(\boldsymbol{W}\boldsymbol{x} + b, \boldsymbol{\mu})) \quad (1)$$

where $\sigma(\cdot)$ is the softmax function, $d(\cdot)$ is a distance measure, $\boldsymbol{W} \in \mathbb{R}^I$ is the learnable weight matrix, $b \in \mathbb{R}$ is the learnable bias, $\boldsymbol{x} \in \mathbb{R}^I$ is the data from the previous layer and $\boldsymbol{\mu} \in \mathbb{R}^4$ is the vector of learnable thresholds. For interpretability we introduce the variable $\boldsymbol{z}$ which denotes $\boldsymbol{W}\boldsymbol{x} + b$. The similarity measure we used is

$$\mathrm{d}(z, \boldsymbol{\mu}) = \left[|\mu_1 - z| \quad \cdots \quad |\mu_4 - z| \right]^T \quad (2)$$

We fixed μ_1 to 0 and μ_4 to 1 and allowed μ_2 and μ_3 to vary while being monotonic ($\mu_1 < \mu_2 < \mu_3 < \mu_4$). The fixation of μ_1 and μ_4 is necessary to ensure one unique optimal solution to the maximum likelihood estimation of Equation (1), because d in invariant to adding a constant shift to μ. Each μ is connected to an AASM NREM sleep stage with μ_1 for N3 and μ_4 for wakefulness.

The OSD initially uses 3-s epochs. As a post-processing step, OSD is smoothened by a moving average filter of 30 s, using a 3-s step size.

2.4 | Acquiring ORP

We ran ORP on the validation data using the Cerebra MY ORP API (Younes 2024). Available ORP reports were downloaded, and ORP and ORP30 were extracted using code provided by Cerebra. To align with Cerebra's reports, we used ORP30, the average of 30 s of ORP calculated using non-overlapping 30 s windows.

2.5 | Statistical Analysis

To validate the sleep depth measures obtained from the OSD and ORP algorithms and to investigate their associations with clinical variables such as age, sex, Apnea–Hypopnea Index (AHI) and cognitive impairment, we conducted several statistical analyses. Our aim was to assess the performance of these algorithms in reflecting sleep depth and to determine their sensitivity to known physiological and pathological factors affecting sleep. Shown p-values are corrected for multiple comparisons using the Benjamini–Hochberg (BH) method.

2.6 | Evaluation of Algorithm Performance

We evaluated the performance of the OSD and ORP algorithms by examining their correlations with the conventional sleep staging (hypnogram) and the ArI. To estimate the variability and confidence intervals of the correlation coefficients, we employed a bootstrapping procedure. Specifically, we performed 1000 rounds of bootstrapping, where in each round, we randomly sampled subjects with replacement from the validation data set to create a bootstrap sample. This method provides robust non-parametric estimates of the correlations and their confidence intervals.

For each bootstrap sample, we calculated Spearman's rank correlation coefficient (Spearman's rho) between the sleep depth measures (OSD and ORP) and the hypnogram stages. This analysis was conducted at both 3- and 30-s resolutions to align with the native timeframes of both the algorithms and the hypnogram. Spearman's rho is a non-parametric measure that assesses the strength and direction of the monotonic association between two ranked variables, allowing us to evaluate how well the sleep depth measures reflect the conventional sleep stages.

Additionally, we calculated Pearson's correlation coefficient (Pearson's r) between the average sleep depth measures and the ArI across subjects in each bootstrap sample. Pearson's r assesses the linear relationship between two continuous variables, providing insight into how well the sleep depth measures predict the frequency of arousals. The relationship between sleep depth and the ArI was additionally assessed using a quadratic curve-linear correlation.

2.7 | Association With Age, Sex and AHI

To investigate associations between sleep depth measures and age, sex and AHI while controlling for potential confounders, we performed partial correlation analyses. Additionally, the shown p-values are adjusted for multiple comparisons using the BH method.

2.7.1 | Age and Sex

We analysed subjects with AHI less than 5 (indicating no significant SDB) and no known cognitive impairment or symptoms. This stratification minimised confounding effects from SDB and cognitive impairment. When assessing the effect of age on sleep depth, we controlled for sex; conversely, when assessing the effect of sex, we controlled for age. Age was treated as a continuous variable and sex as a categorical variable.

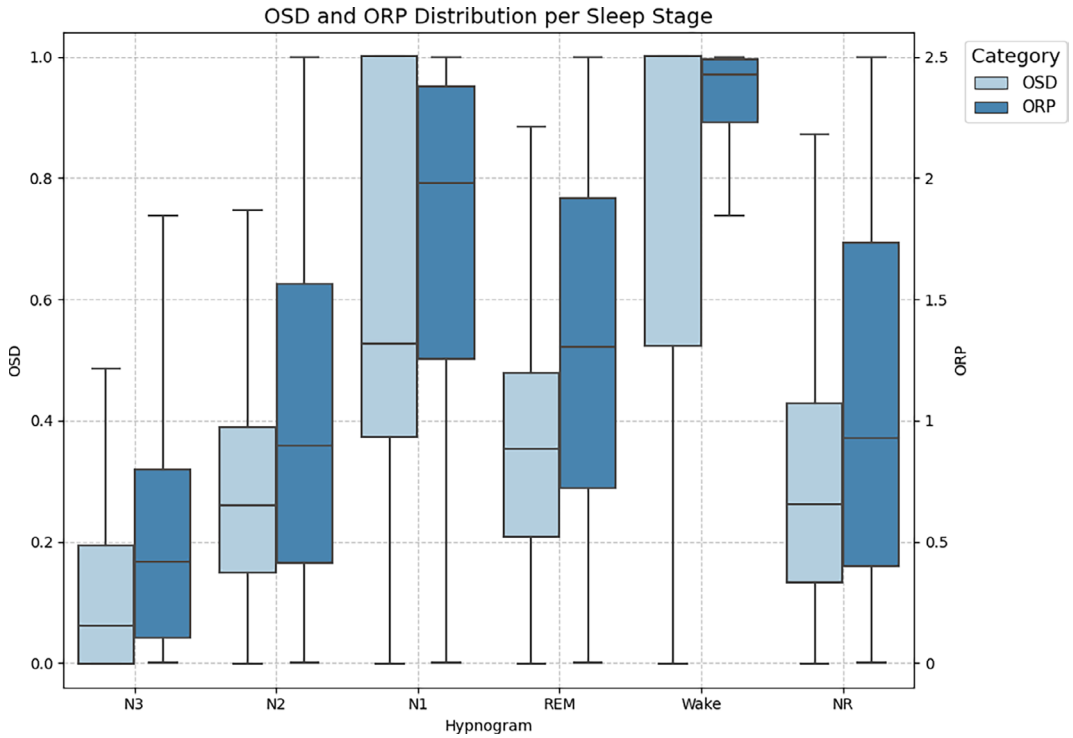

FIGURE 3 | Comparison of OSD and ORP values across AASM sleep stages: this figure shows the distributions of OSD and ORP values for each AASM-defined sleep stage (wake, N1, N2, N3 and REM) as boxplots. Both OSD and ORP demonstrate an increase in sleep depth across NREM stages (N1, N2, N3). ORP exhibits a larger spread in stages N3, N2, REM and NREM, while OSD shows wider ranges in N1 and Wakefulness. A two-sample *t*-test (with ORP values scaled from 0 to 1 to match OSD's scale) confirmed that OSD and ORP distributions are statistically different within each sleep stage ($p < 0.01$).

2.7.2 | Sleep-Disordered Breathing

To assess the influence of SDB on sleep depth, we included subjects without cognitive impairment or symptoms and performed partial correlation analyses between AHI and sleep depth measures, controlling for age and gender. This approach allowed us to isolate the effect of SDB severity on sleep depth. Furthermore, we categorised SDB into REM-dominant SDB, NREM-dominant SDB and non-dominant SDB. With a study being classified as dominant, the ratio 'AHI in NREM' to 'AHI in REM' should be bigger than 2 for a study to be classified as NREM-dominant or smaller than 0.5 for REM-Dominant SDB. Non-dominant SDB is defined as the case not satisfying either REM dominant or NREM dominant.

2.7.3 | Association With Cognitive Impairment

We explored how sleep depth measures vary with different levels of cognitive impairment. For this analysis, we selected a subset of the validation data set consisting of studies with AHI less than five and known cognitive status to ensure that SDB did not confound the analysis of cognitive impairment.

Due to the ordinal nature of cognitive impairment levels, we used ordinal least squares (OLS) regression to model the association between sleep depth measures and cognitive status. We assigned numerical labels to cognitive status categories: No Cognitive Impairment = 0, Symptomatic = 1, MCI = 2 and Dementia = 3. Age and gender were included as covariates in the regression model to adjust for their potential effects on sleep depth. By regressing the sleep depth measures on the cognitive status labels, while controlling for age and gender, we assessed whether there was a significant association between increasing levels of cognitive impairment and changes in sleep depth.

3 | Results

3.1 | Sleep Depth in Each Sleep Stage

Both OSD and ORP show an increasing sleep depth along with N1, N2 and N3. Figure 3 visualises OSD and ORP values per AASM sleep stage. ORP has a larger spread in N3, N2, REM and NREM sleep. OSD showed wider ranges in N1 and Wakefulness. The missing whiskers are due to the unevenly distributed values, making parts of the boxplot extremely compressed, seemingly making them nonexistent during visualisation. The distribution per AASM sleep stages in ORP and OSD was tested using a linear mixed effects model, and a post hoc Tukey honest significant difference test was applied. This showed that all stages within the sleep depth algorithms are significantly different from each other, as well as that ORP and OSD are statistically different from each other in all stages with $p < 0.001$ while adjusting for multiple comparisons. ORP was re-scaled to a range of 0–1 for

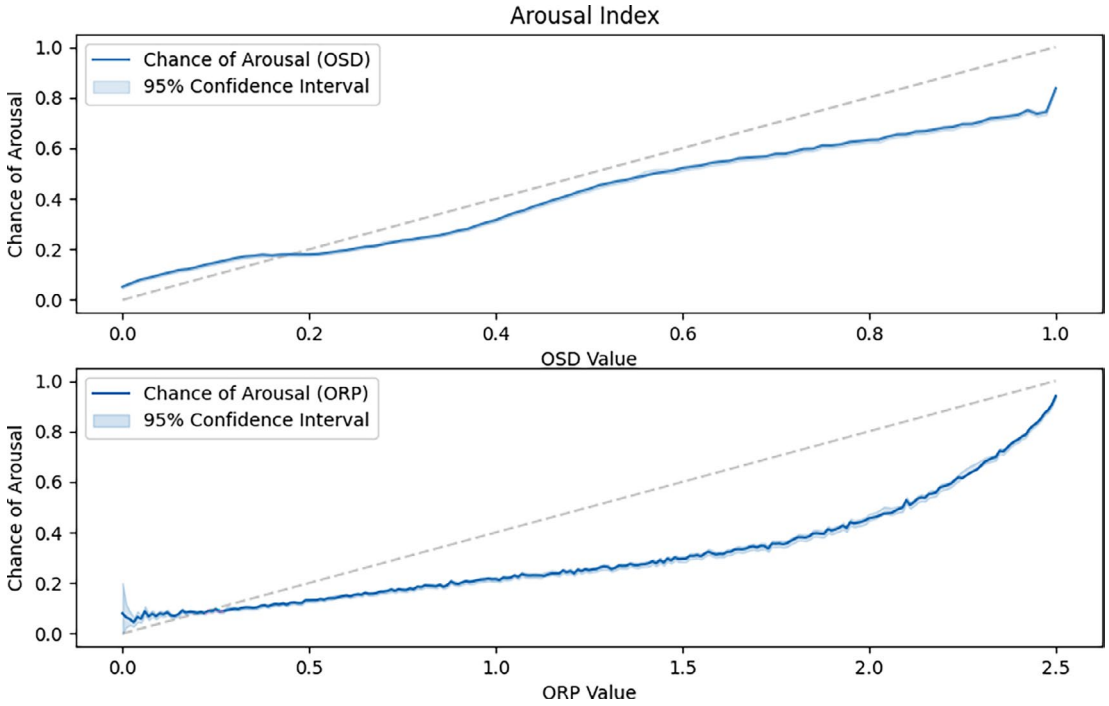

FIGURE 4 | The Arousal Index: both OSD (upper panel) and ORP (lower panel) demonstrate a strong linear correlation with the Arousal Index, indicating that higher values of these algorithms correspond to lighter sleep and an increased likelihood of arousal. The grey dotted line represents a perfect linear correlation. The Pearson's correlation coefficients and the curve-linear r^2 (median [95% CI]) between each algorithm and the Arousal Index are 0.994 [0.992, 0.995] and 0.987 [0.985, 0.989] for OSD and 0.923 [0.918, 0.928] and 0.950 [0.945, 0.954] for ORP. Both underscoring the robustness of these associations.

this comparison. All OSD results on the Development set are presented in Appendix.

3.2 | Sleep Depth and the Arousal Index

Both algorithms showing strong linear and curve-linear correlation with the ArI. Pearson's correlation (median [95% CI]) of OSD and ORP, with ArI is 0.994 [0.992, 0.995] and 0.923 [0.918, 0.928], respectively. The quadratic curve-linear fit yields an r^2 of 0.987 [0.985, 0.989] for OSD and 0.950 [0.945, 0.954] for ORP. ArIs with confidences intervals are shown in Figure 4.

3.3 | Influence of Age and Gender on Sleep Depth

Both OSD and ORP indices showed significant associations with age across most sleep stages. The age distribution, grouped in 10-year bins, per sleep stage for both OSD and ORP is shown in Figure 5. For OSD, age was positively correlated with sleep depth reductions across all stages, indicating shallower sleep with advancing age: N3, N2, N1, REM and NREM. ORP displayed a similar trend, with age associated with significantly lighter sleep in N1, REM and NREM. Age was correlated with deeper sleep in N3. The statistical findings of age, gender, AHI and cognitive impairment are summarised in Figure 6.

Gender also significantly impacted sleep depth when adjusted for age. OSD showed that female subjects had deeper sleep in N3, N2, REM and NREM. In ORP, females displayed deeper sleep across all stages: N3, N2, N1, REM and NREM.

3.4 | Influence of AHI on Sleep Depth

The influence of SDB was analysed by categorising SDB into NREM-dominant, REM-dominant and non-dominant SDB patterns. For non-dominant SDB, OSD showed to have a significant influence on sleep depth in N2 and NREM stages, with lighter sleep associated with increased SDB. Similarly, ORP showed significant associations in N2 and NREM.

NREM-dominant SDB was significantly positively associated with OSD sleep depth in N2 and in total NREM. ORP showed significant positive associations in N2 and total NREM. REM-dominant SDB was associated with decreased sleep depth in N2 and REM for OSD. N3, N2 and NREM were associated with lighter sleep with increasing SDB severity for ORP.

3.5 | Sleep Depth and Cognitive Impairment

Cognitive impairment was associated with lighter sleep across multiple stages. OSD showed reductions in sleep depth in N3, N2, N1 and NREM with increasing cognitive impairment. ORP reflected a significant reduction in sleep depth in N1 for individuals with cognitive impairment.

3.6 | Sleep Depth Correlates With the Hypnogram

For OSD, the correlation (median [95% CI]) with the hypnogram was 0.678 [0.673, 0.682] and 0.755 [0.750, 0.759] for the 3 and 30 s windows, respectively, in comparison to 0.596 [0.591,

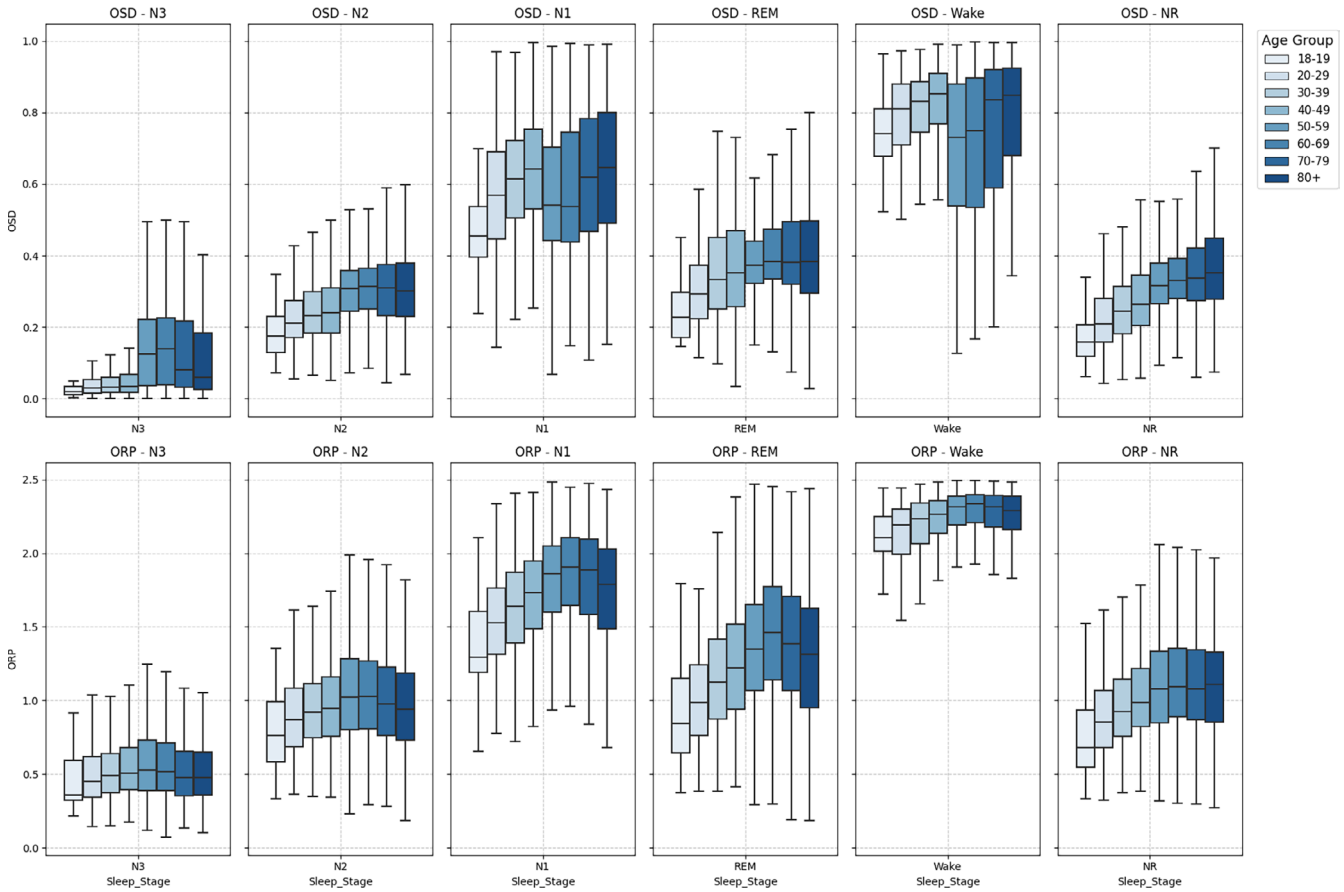

FIGURE 5 | Sleep distribution per AASM stage for age per algorithm: The rows are showing OSD and ORP, respectively. Both algorithms are showing a strong overall trend of lighter sleep with increasing age, with OSD showing fluctuations in N3, N1 and Wake, while having strong correlations in N2, REM and NREM. ORP shows fewer fluctuations yet an increase in sleep depth from age 50 going upwards in N3, N2, N1 and REM.

0.600] and 0.715 [0.710, 0.720] for ORP. The distribution among patients can be seen in the boxplots shown in Figure 7. A full night of OSD and ORP together with the hypnogram is shown in Figure 8. Additionally, the average spectrogram using the mean of the frontal, central and occipital electrodes is visualised. For OSD and ORP, the hypnogram is shown in black, where REM is annotated in red. The light blue prediction is the non-averaged prediction, and the darker blue is the final prediction. It can be seen that OSD and ORP show high correlation to the hypnogram, OSD and ORP displaying a gradual change at sleep stage transition. It is seen that ORP shows relatively lighter sleep in REM compared to OSD. Additionally, the light grey-blue predictions of ORP, which visualise the non-averaged 3-s bins, show a higher variance than the unaveraged OSD values, indicating a lower coherence between the current and neighbouring segments in ORP compared to OSD.

4 | Discussion

In this study, we developed a novel, data-driven continuous measure of sleep depth called OSD by training a deep neural network on a large clinical data set of sleep EEG recordings. Our findings demonstrate that, similar to ORP, OSD has a strong linear relationship with the probability of arousal and effectively captures variations in sleep depth associated with age, gender, SDB and cognitive impairment. Notably, OSD showed lighter sleep in patients with dementia, who exhibit increased delta and theta activity, aligning with existing literature on neurodegenerative diseases (Neto et al. 2015; Kavcic et al. 2016). Our results independently validate the basic idea of sleep depth.

Our approach differs from prior methods such as the ORP (Younes et al. 2015), which relies on hand-engineered features based on EEG spectral power. While both OSD and ORP aim to quantify sleep depth continuously, OSD is entirely data-driven and leverages deep learning to capture fine-grained and morphological EEG features. This results in notable differences between the two measures. OSD values have less overlap between the AASM sleep stages and exhibit smoother transitions in sleep depth, indicating greater temporal stability (Figure 3). These characteristics suggest that OSD may be more sensitive to subtle changes in sleep architecture. However, the AASM stages have substantial arbitrariness, and variances between OSD and ORP may not be clinically meaningful or point to important biological features that are hidden by current standard approaches.

Age-related changes in sleep are well documented, with older adults experiencing less deep sleep and increased fragmentation (Li et al. 2018; Ohayon 2011; Lichstein 2004; Tang et al. 2017; Vysata et al. 2012). OSD reflected this phenomenon by showing significant decreasing sleep depth trends in N3, N2, N1, REM and overall NREM sleep with advancing age. However, ORP showed an unexpected increase in sleep depth in N3 with age

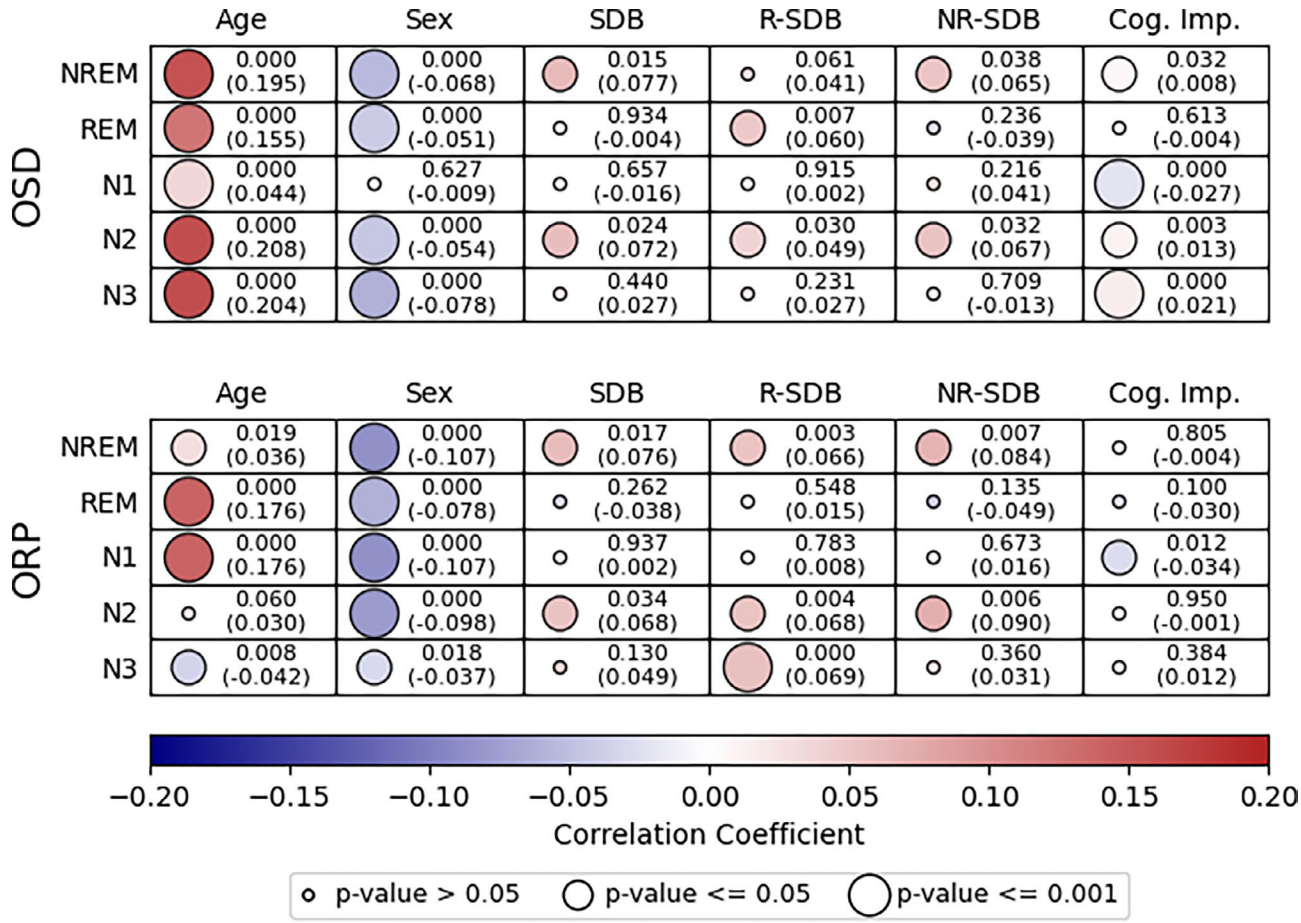

FIGURE 6 | Sleep influencing factor on OSD and ORP: The P-values per test are summarised in this table. The rows represent the sleep stage (dependent variable) per algorithm whereas the columns indicate the independent variable. The *p*-values for Age, Sex and AHI, R-AHI, NR-AHI are acquired by partial correlation, whereas the *p*-values for cognitive impairment are acquired by ordinal least square regression. Cog. Imp = different levels of cognitive impairment; NR-SDB = NREM-dominant SDB; R-SDB = REM-dominant SDB; With SDB = non-dominant SDB.

and an increased sleep depth with increasing age in N1, REM and NREM. Age-related behaviour of ORP, at this level of granularity, is not yet described in the literature. Younes et al. (Younes et al. 2021) showed a significant increase in sleep depth for ORP in NREM and REM using 15-year bins, which aligns with our findings. The increase in sleep depth in N3 may be due to ORP's reliance on EEG spectral power. ORP captures the EEG spectrum in four bins: 0.3–2.3, 2.7–6.3, 7.3–14.0 and 14.3–35.0 Hz (Younes et al. 2015). The delta range is captured in two bins, with one of them also capturing theta activity. This might be a design choice that is not be able to fully capture age-related EEG changes. In contrast, OSD's deep learning framework allows it to account for a broader range of EEG characteristics, providing a sleep depth measure consistent with known age-related alterations in sleep.

Sex differences in sleep patterns have been reported, with females generally experiencing deeper sleep characterised by higher sleep efficiency, less wakefulness after sleep onset, less N1 sleep and more N3 sleep (Carrier et al. 2017; van den Berg et al. 2009; Roehrs et al. 2006; Goel et al. 2005; Bixler et al. 2009). Both OSD and ORP demonstrated that females have significantly deeper sleep in N3, N2, REM and overall NREM sleep compared to males. However, ORP also indicated deeper sleep in N1 for females, a finding not observed in OSD.

SDB is known to disrupt sleep architecture and reduce sleep depth. Both OSD and ORP showed significant reductions in sleep depth in N2 and NREM sleep during general and NREM-dominant SDB. Interestingly, ORP indicated an increase in sleep depth in REM during NREM-dominant SDB, which may seem counterintuitive since REM sleep is not typically deeper to compensate for lighter NREM sleep (Patel et al. 2024; Bonsignore et al. 2024). Although EEG slowing in REM is reported in untreated OSA patients compared to their controls. (Morisson et al. 2001, 1998) OSD might decrease (increased sleep depth) based on the REM slowing. Whether slowed REM sleep in OSA patients resembles deeper REM sleep is yet to be determined. OSD, however, showed consistent reductions in sleep depth across N2, REM and NREM during REM-dominant SDB, aligning with the understanding that SDB affects multiple sleep stages. This suggests that OSD may more accurately reflect the impact of SDB on sleep depth across different stages.

Cognitive impairment and neurodegenerative diseases like dementia are associated with altered sleep patterns, including reductions in deep sleep (Neto et al. 2015; Kavcic et al. 2016; Kim 2024; Song et al. 2015). OSD demonstrated significant reductions in sleep depth across N3, REM *and NREM stages. Additionally, it showed sleep depth increase in N1 with increasing levels of cognitive impairment, whereas ORP only showed

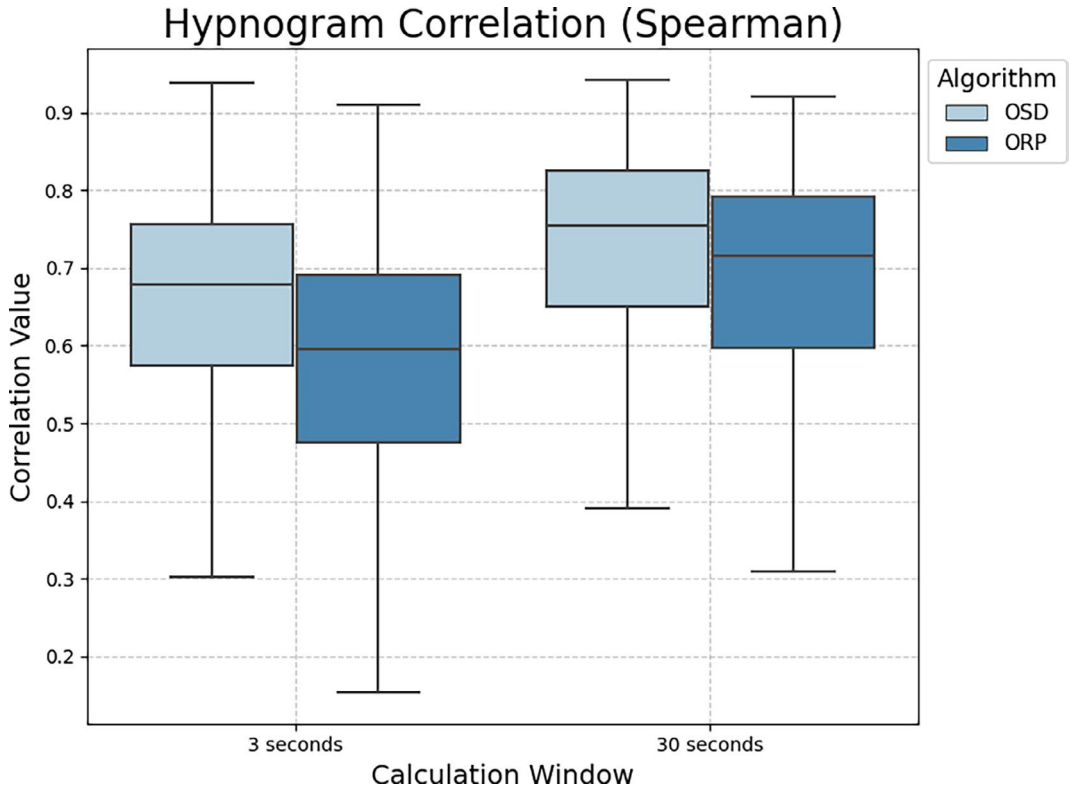

FIGURE 7 | Spearman correlation: The patient-wise Spearman correlation distribution with the hypnogram for both OSD and ORP. In the 3-s calculation windows, native to OSD and ORP, correlations of 0.678 and 0.596 are shown for OSD and ORP, respectively. Using 30 s windows, native to the hypnogram, correlations of 0.755 and 0.715 are found for OSD and ORP, respectively.

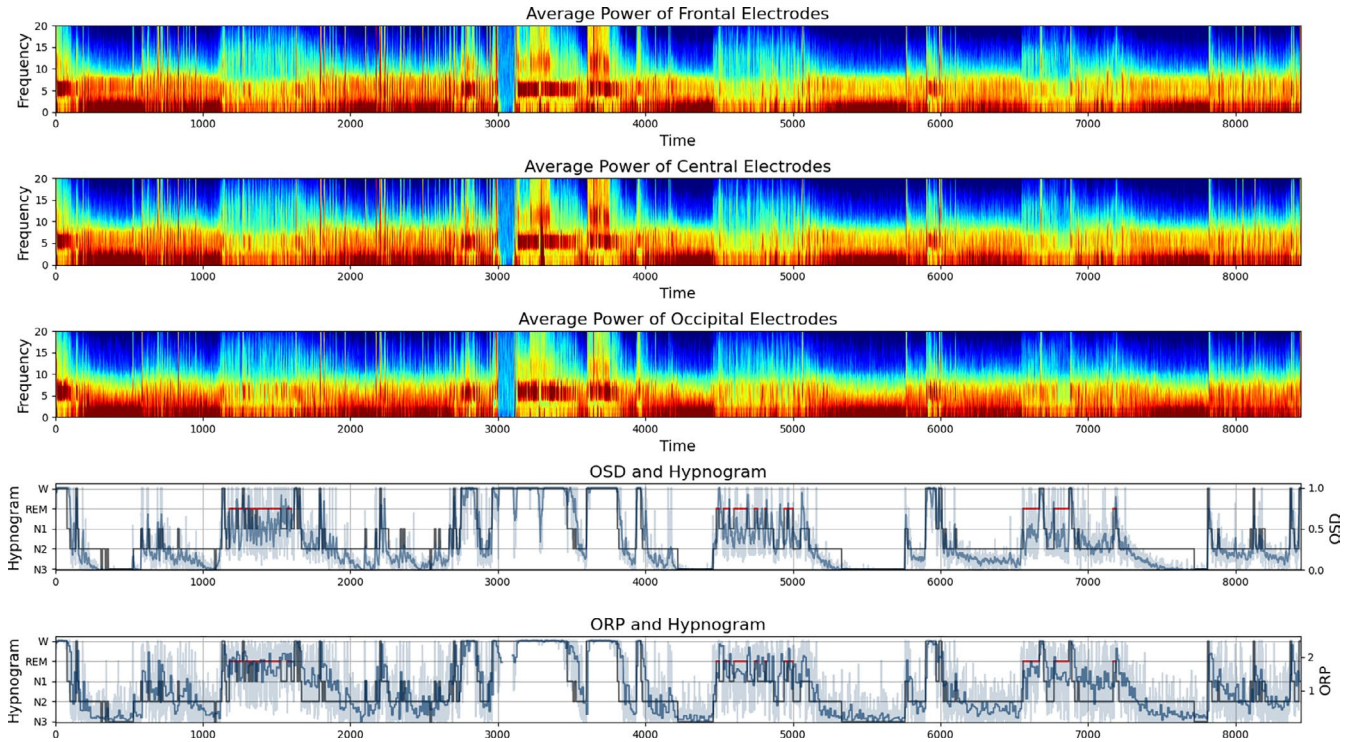

FIGURE 8 | The sleep depth algorithms in relation to the hypnogram: a representation of the OSD and ORP with their hypnograms. The first three panels are showing the power spectral density for the frontal, central and occipital electrodes, respectively. Whereas the lower fourth panel shows the OSD without smoothing in light blue and the smoothed/final OSD in dark blue. The hypnogram is shown in black where REM sleep is annotated in red. The fifth panel is similar to the fourth except now the ORP is shown.

significant increase in N1. The reduction in N3, REM and NREM and increased sleep depth in N1 does show overall lighter sleep, which suggests that less time is spent in the deeper stages. The reduction of time spent in N3, N2 and the increase in N1 time aligns with prior literature. (Neto et al. 2015; Kavcic et al. 2016; Kim 2024; Song et al. 2015) This underscores OSD's potential advantages in capturing the complex interplay between sleep depth and cognitive decline. By incorporating fine-grained EEG features beyond spectral power, OSD may provide new insights into how neurodegenerative processes affect sleep architecture.

The assessment of sleep depth requires some important additional considerations. The AASM arousal requires 3 s of EEG acceleration but both shorter and longer transients are important. Arousals are modified by drugs, time of night, stage and stage of sleep and have numerous concomitant networked transients in motor, autonomic and respiratory domains. It may be insightful and important to see how OSD/ORP functions when autonomic transients are the primary arousal equivalents (heart rate acceleration, blood pressure surges, photoplethysmographic pulse wake amplitude reductions). The AASM 30-s epoch was born out of pragmatism for manual scoring. Current approaches to automated scoring remain constrained by this 'rule'. There are advantages to variation of this standard. For example, Cyclic Alternating Pattern is best viewed at higher compressions, 60–90 s. Differentiation of phasic and tonic REM sleep can be easier at longer than 30 s time scales. The OSD used a 3-s epoch in part as ORP uses that duration, allowing direct comparisons with an established measure of continuous sleep depth. The dynamics of spindles, <1 Hz slow-oscillations and other graphoelements of sleep all have their own optimal analysis windows. Sleep-compatible electrical brain activity can be identified at very fine sub-second resolutions, is dynamic and likely encodes depth information—areas that are worth research exploration. Sleep depth research is likely to be relevant to a wide range of neurological (e.g., neurodegeneration, epilepsy) and sleep disorders (e.g., insomnia, narcolepsy, parasomnia).

Our study has limitations that warrant consideration. During the training of OSD, REM sleep was not directly used to guide the ordinal regression objective because REM is regulated by different physiological mechanisms compared to NREM sleep and does not fit naturally along a continuum of sleep depth (Nielsen 2000). Technically, we addressed this by setting the class weight for REM to zero in the ordinal loss function, allowing the model to learn REM sleep patterns passively while focusing on NREM stages for sleep depth estimation. Although OSD provides sleep depth estimates during REM epochs, users should interpret these values cautiously. Additionally, OSD is currently constrained by conventional sleep staging as defined by the AASM. Future research could explore more independent approaches that move beyond traditional staging to develop novel sleep measures closely related to health outcomes. It is possible that the optimal sleep epoch length is something not yet considered and can be estimated by data driven approaches. Additionally, we did not examine night-to-night variability. Given that sleep depth can fluctuate across nights due to various intrinsic and extrinsic factors, future research should explore the consistency of these measures over multiple nights.

In conclusion, OSD is a data-driven measure of sleep depth that correlates strongly with arousal probability and effectively captures variations associated with age, sex, SDB and cognitive impairment. By leveraging deep learning to analyse EEG data without reliance on hand-engineered features, OSD offers a nuanced understanding of sleep architecture. Despite the reported differences between OSD and ORP, these results also validate the general approach of ORP, as most of the differences are not likely clinically meaningful at the individual subject level. Both algorithms have the potential to show individual differences in sleep vulnerability to disruptors such as sleep disruptive breathing. These capabilities to reflect subtle EEG characteristics may enhance physiological studies of sleep disorders, contribute to defining sleep quality and improve our understanding of sleep's role in neurological diseases. Future work will investigate OSD's potential applications in conditions such as insomnia and its value in clinical assessments of sleep health.

Author Contributions

Erik-Jan Meulenbrugge: data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft. **Haoqi Sun:** conceptualization, data curation, methodology, supervision, software, validation, writing – review and editing. **Wolfgang Ganglberger:** data curation, validation, writing – review and editing. **Samaneh Nasiri:** writing – review and editing. **Robert J. Thomas:** conceptualization, writing – review and editing. **M. Brandon Westover:** conceptualization, data curation, funding acquisition, methodology, project administration, resources, supervision, validation, writing – review and editing.

Conflicts of Interest

Dr. Westover is a co-founder, scientific advisor and consultant to Beacon Biosignals and has a personal equity interest in the company. Dr. Thomas reports: (1) patent and licence to MyCardio LLC, for ECG-spectrogram; (2) grant support, licence and intellectual property (licensed patents) from DeVilbiss-Drive Healthcare; (3) unlicensed patent for a device regulation CO_2 during positive pressure therapy, for central sleep apnea; (4) patent submitted for respiratory self-similarity for estimation of expressed high loop gain; (5) patent submitted for Enhanced Expiratory Rebreathing Space for treatment of high loop gain sleep apnea; (6) general sleep medicine consulting: GLG Councils, Guidepoint Global; (5) Consultant-Jazz Pharmaceuticals. While this article was being written, E. Meulenbrugge transitioned to Onera Health. Onera Health did not provide any financial nor non-financial support for this work.

Data Availability Statement

All data and code that are used to conduct this research will be freely available at BDSP.IO. Please refer to the README file in the repository for guidance.

References

Andrillon, T., and D. Oudiette. 2023. "What Is Sleep Exactly? Global and Local Modulations of Sleep Oscillations all Around the Clock." *Neuroscience and Biobehavioral Reviews* 155: 105465. https://doi.org/10.1016/j.neubiorev.2023.105465.

Asyali, M. H., R. B. Berry, M. C. K. Khoo, and A. Altinok. 2007. "Determining a Continuous Marker for Sleep Depth." *Computers in Biology and Medicine* 37, no. 11: 1600–1609. https://doi.org/10.1016/j.compbiomed.2007.03.001.

Biswal, S., H. Sun, B. Goparaju, M. B. Westover, J. Sun, and M. T. Bianchi. 2018. "Expert-Level Sleep Scoring With Deep Neural Networks." *Journal of the American Medical Informatics Association* 25, no. 12: 1643–1650. https://doi.org/10.1093/jamia/ocy131.

Bixler, E. O., M. N. Papaliaga, A. N. Vgontzas, et al. 2009. "Women Sleep Objectively Better Than Men and the Sleep of Young Women Is More Resilient to External Stressors: Effects of Age and Menopause." *Journal of Sleep Research* 18, no. 2: 221–228. https://doi.org/10.1111/j.1365-2869.2008.00713.x.

Bonsignore, M. R., E. Mazzuca, P. Baiamonte, B. Bouckaert, W. Verbeke, and D. A. Pevernagie. 2024. "REM Sleep Obstructive Sleep Apnoea." *European Respiratory Review* 33, no. 171: 230166. https://doi.org/10.1183/16000617.0166-2023.

Carrier, J., K. Semba, S. Deurveilher, et al. 2017. "Sex Differences in Age-Related Changes in the Sleep-Wake Cycle." *Frontiers in Neuroendocrinology* 47: 66–85. https://doi.org/10.1016/j.yfrne.2017.07.004.

Carrubba, S., P. Y. Kim, D. E. McCarty, A. L. Chesson, C. Frilot, and A. A. Marino. 2012. "Continuous EEG-Based Dynamic Markers for Sleep Depth and Phasic Events." *Journal of Neuroscience Methods* 208, no. 1: 1–9. https://doi.org/10.1016/j.jneumeth.2012.04.018.

Goel, N., H. Kim, and R. P. Lao. 2005. "Gender Differences in Polysomnographic Sleep in Young Healthy Sleepers." *Chronobiology International* 22, no. 5: 905–915. https://doi.org/10.1080/07420520500263235.

Hannun, A. Y., P. Rajpurkar, M. Haghpanahi, et al. 2019. "Cardiologist-Level Arrhythmia Detection and Classification in Ambulatory Electrocardiograms Using a Deep Neural Network." *Nature Medicine* 25, no. 1: 65–69. https://doi.org/10.1038/s41591-018-0268-3.

Iber, C., S. Ancoli-Israel, A. L. Chesson, and S. Quan. 2007. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. American Academy of Sleep Medicine.

Kavcic, V., B. Zalar, and B. Giordani. 2016. "The Relationship Between Baseline EEG Spectra Power and Memory Performance in Older African Americans Endorsing Cognitive Concerns in a Community Setting." *International Journal of Psychophysiology* 109: 116–123. https://doi.org/10.1016/j.ijpsycho.2016.09.001.

Kim, K. Y. A. 2024. "What the Changes in Sleep Architecture Tell You About Cognitive Decline—An Editorial." *Sleep* 47, no. 10: zsae180. https://doi.org/10.1093/sleep/zsae180.

Li, J., M. V. Vitiello, and N. S. Gooneratne. 2018. "Sleep in Normal Aging." *Sleep Medicine Clinics* 13, no. 1: 1–11. https://doi.org/10.1016/j.jsmc.2017.09.001.

Lichstein, K. L., ed. 2004. *Epidemiology of Sleep: Age, Gender, and Ethnicity*. Lawrence Erlabaum Assoc.

Morisson, F., A. Décary, D. Petit, G. Lavigne, J. Malo, and J. Montplaisir. 2001. "Daytime Sleepiness and EEG Spectral Analysis in Apneic Patients Before and After Treatment With Continuous Positive Airway Pressure." *Chest* 119, no. 1: 45–52. https://doi.org/10.1378/chest.119.1.45.

Morisson, F., G. Lavigne, D. Petit, T. Nielsen, J. Malo, and J. Montplaisir. 1998. "Spectral Analysis of Wakefulness and REM Sleep EEG in Patients With Sleep Apnoea Syndrome." *European Respiratory Journal* 11, no. 5: 1135–1140. https://doi.org/10.1183/09031936.98.11051135.

Neto, E., E. A. Allen, H. Aurlien, H. Nordby, and T. Eichele. 2015. "EEG Spectral Features Discriminate Between Alzheimers and Vascular Dementia." *Frontiers in Neurology* 6: 25. https://doi.org/10.3389/fneur.2015.00025.

Nielsen, T. A. 2000. "A Review of Mentation in REM and NREM Sleep: 'Covert' REM Sleep as a Possible Reconciliation of Two Opposing Models." *Behavioral and Brain Sciences* 23, no. 6: 851–866. https://doi.org/10.1017/s0140525x0000399x.

Ohayon, M. M. 2011. "Epidemiological Overview of Sleep Disorders in the General Population." *Sleep Medicine Research* 2, no. 1: 1–9. https://doi.org/10.17241/smr.2011.2.1.1.

Patanaik, A., J. L. Ong, J. J. Gooley, S. Ancoli-Israel, and M. W. L. Chee. 2018. "An End-To-End Framework for Real-Time Automatic Sleep Stage Classification." *Sleep* 41, no. 5: zsy041. https://doi.org/10.1093/sleep/zsy041.

Patel, A. K., V. Reddy, K. R. Shumway, and J. F. Araujo. 2024. "Physiology, Sleep Stages." In *StatPearls*. StatPearls Publishing.

Picchioni, D., S. G. Horovitz, M. Fukunaga, et al. 2011. "Infraslow EEG Oscillations Organize Large-Scale Cortical–Subcortical Interactions During Sleep: A Combined EEG/fMRI Study." *Brain Research* 1374: 63–72. https://doi.org/10.1016/j.brainres.2010.12.035.

Rennie, J. D. M., and N. Srebro. 2005. "Loss Functions for Preference Levels: Regression With Discrete Ordered Labels." In *Proceedings of the IJCAI Multidisciplinary Workshop on Advances in Preference Handling*, 1–6. AAAI Press.

Roehrs, T., A. Kapke, T. Roth, and N. Breslau. 2006. "Sex Differences in the Polysomnographic Sleep of Young Adults: A Community-Based Study." *Sleep Medicine* 7, no. 1: 49–53. https://doi.org/10.1016/j.sleep.2005.05.008.

Saastamoinen, A., E. Huupponen, A. Värri, J. Hasan, and S. L. Himanen. 2006. "Computer Program for Automated Sleep Depth Estimation." *Computer Methods and Programs in Biomedicine* 82, no. 1: 58–66. https://doi.org/10.1016/j.cmpb.2006.02.003.

Song, Y., T. Blackwell, K. Yaffe, et al. 2015. "Relationships Between Sleep Stages and Changes in Cognitive Function in Older Men: The MrOS Sleep Study." *Sleep* 38, no. 3: 411–421. https://doi.org/10.5665/sleep.4500.

Stephansen, J. B., A. N. Olesen, M. Olsen, et al. 2018. "Neural Network Analysis of Sleep Stages Enables Efficient Diagnosis of Narcolepsy." *Nature Communications* 9, no. 1: 5229. https://doi.org/10.1038/s41467-018-07229-3.

Sun, H., J. Jia, B. Goparaju, et al. 2017. "Large-Scale Automated Sleep Staging." *Sleep* 40: 10. https://doi.org/10.1093/sleep/zsx139.

Tang, J., Y. Liao, B. C. Kelly, et al. 2017. "Gender and Regional Differences in Sleep Quality and Insomnia: A General Population-Based Study in Hunan Province of China." *Scientific Reports* 7, no. 1: 43690. https://doi.org/10.1038/srep43690.

van den Berg, J. F., H. M. E. Miedema, J. H. M. Tulen, A. Hofman, A. K. Neven, and H. Tiemeier. 2009. "Sex Differences in Subjective and Actigraphic Sleep Measures: A Population-Based Study of Elderly Persons." *Sleep* 32, no. 10: 1367–1375. https://doi.org/10.1093/sleep/32.10.1367.

Vysata, O., J. Kukal, A. Prochazka, L. Pazdera, and M. Valis. 2012. "Age-Related Changes in the Energy and Spectral Composition of EEG." *Neurophysiology* 44, no. 1: 63–67. https://doi.org/10.1007/s11062-012-9268-y.

Younes, M. 2024. "My ORP – ORP Sleep Architecture." https://docs.cerebraportal.com/.

Younes, M., A. Azarbarzin, M. Reid, D. R. Mazzotti, and S. Redline. 2021. "Characteristics and Reproducibility of Novel Sleep EEG Biomarkers and Their Variation With Sleep Apnea and Insomnia in a Large Community-Based Cohort." *Sleep* 44, no. 10: zsab145. https://doi.org/10.1093/sleep/zsab145.

Younes, M., M. Ostrowski, M. Soiferman, et al. 2015. "Odds Ratio Product of Sleep EEG as a Continuous Measure of Sleep State." *Sleep* 38, no. 4: 641–654. https://doi.org/10.5665/sleep.4588.

Appendix

OSD Results on the Development Data

Sleep Depth in Each Sleep Stage

Both OSD shows an increasing sleep depth along with N1, N2 and N3. Figure A1 visualises OSD values per AASM sleep stage. The distribution per AASM sleep stages in OSD was tested using a mixed linear model, and a post hoc Tukey honest significant difference test was applied.

This showed that all stages within the OSD are significantly different from each other, $p < 0.001$ while adjusting for multiple comparisons.

Sleep Depth is Correlated With Arousal Index

OSD shows strong linear and curve-linear correlation with the Arousal Index (ArI). Pearson's correlation (median [95% CI]) of OSD with ArI is 0.994 [0.994, 0.995]. The quadratic curve-linear holds an r^2 of 0.996 [0.996, 0.997] for OSD. The ArI with confidences intervals is shown in Figure A2.

Influence of Age and Gender on Sleep Depth

OSD indices showed significant associations with age across most sleep stages. The age distribution, grouped in 10-year bins, per sleep stage for both OSD and ORP are shown in Figure A3. For OSD, age was positively correlated with sleep depth reductions across all stages, indicating shallower sleep with advancing age: N3, N2, N1, REM and NREM. Gender also significantly impacted sleep depth when adjusted for age. OSD showed that female subjects had deeper sleep in N3, N2, REM and NREM.

Influence of AHI on Sleep Depth

The influence of sleep-disordered breathing (SDB) was analysed by categorising SDB into NREM-dominant, REM-dominant and non-dominant SDB patterns. For non-dominant SDB, OSD showed to have a significant influence on sleep depth in N2 and NREM stages, with lighter sleep associated with increased SDB. Similarly, ORP showed significant associations in N2 and NREM.

NREM-dominant SDB was significantly positively associated with OSD sleep depth in N2 and in total NREM. ORP showed significant positive associations in N2 and total NREM. REM-dominant SDB was associated with decreased sleep depth in N2 and REM for OSD. N3, N2 and NREM were associated with lighter sleep with increasing SDB severity for ORP. The statistical results for age, gender and SDB are summarised in Figure A4.

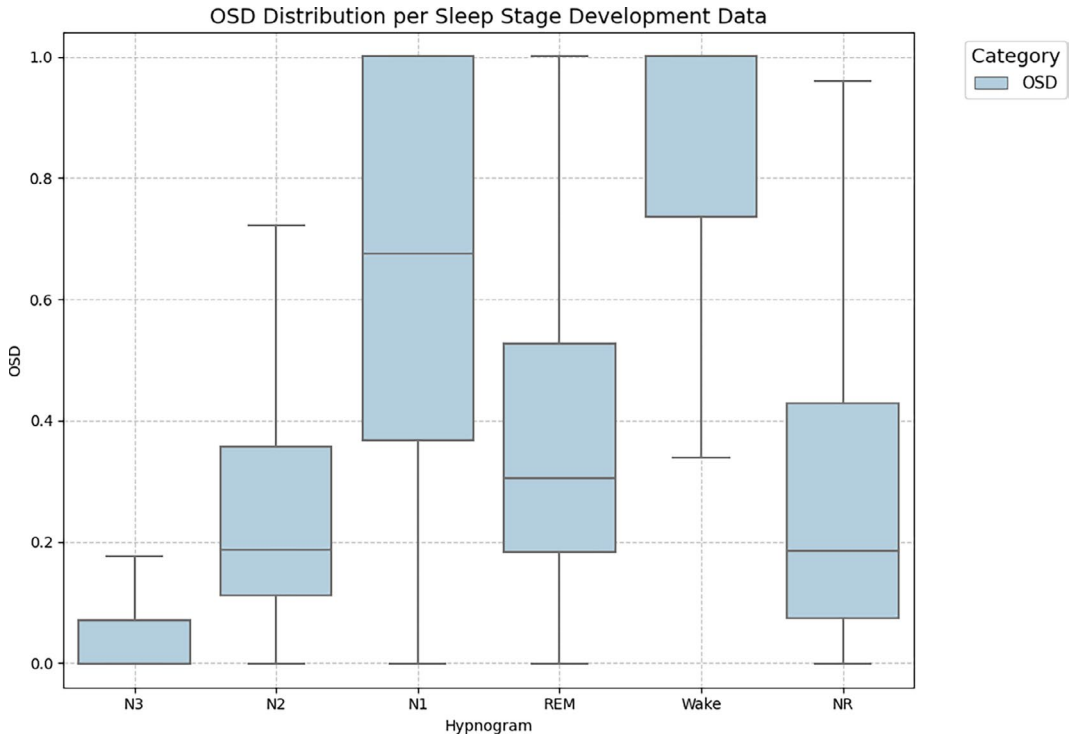

FIGURE A1 | OSD values across AASM sleep stages: this figure shows the distributions of OSD values for each AASM-defined sleep stage (Wake, N1, N2, N3 and REM) as boxplots. OSD demonstrates an increase in sleep depth across NREM stages (N1, N2, N3).

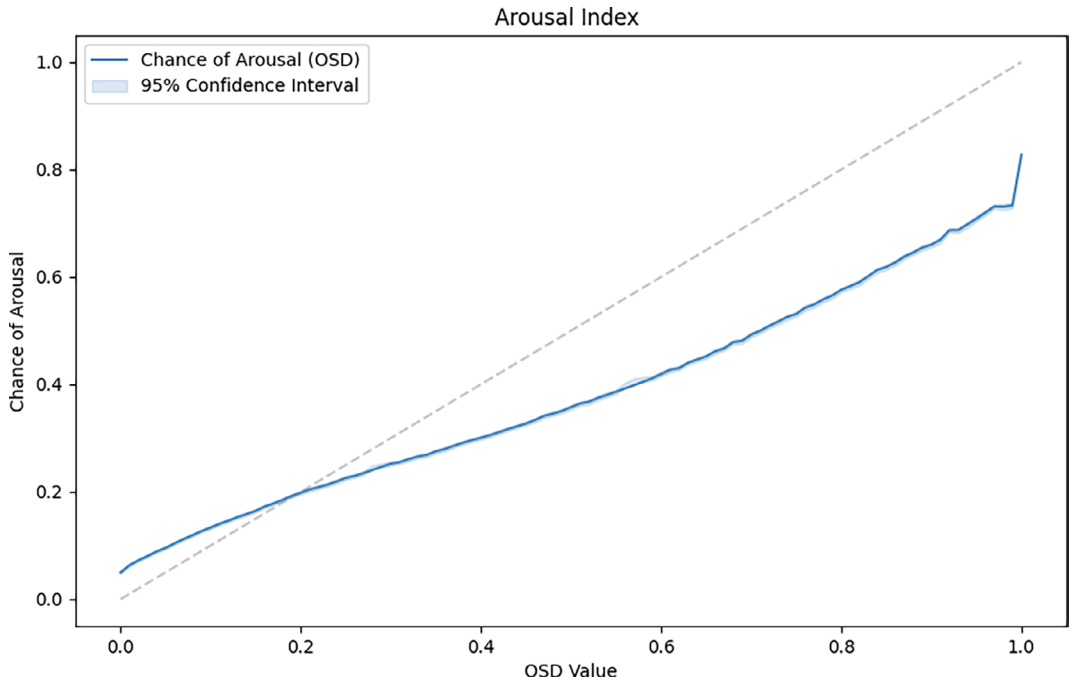

FIGURE A2 | The Arousal Index: OSD demonstrates a strong linear correlation with the Arousal Index, indicating that higher values of these algorithms correspond to lighter sleep and an increased likelihood of arousal. The grey dotted line represents a perfect linear correlation. Pearson's correlation coefficients (median [95% CI]) between each algorithm and the Arousal Index are 0.994 [0.994, 0.995]. The quadratic curve-linear holds an r^2 of 0.996 [0.996, 0.997].

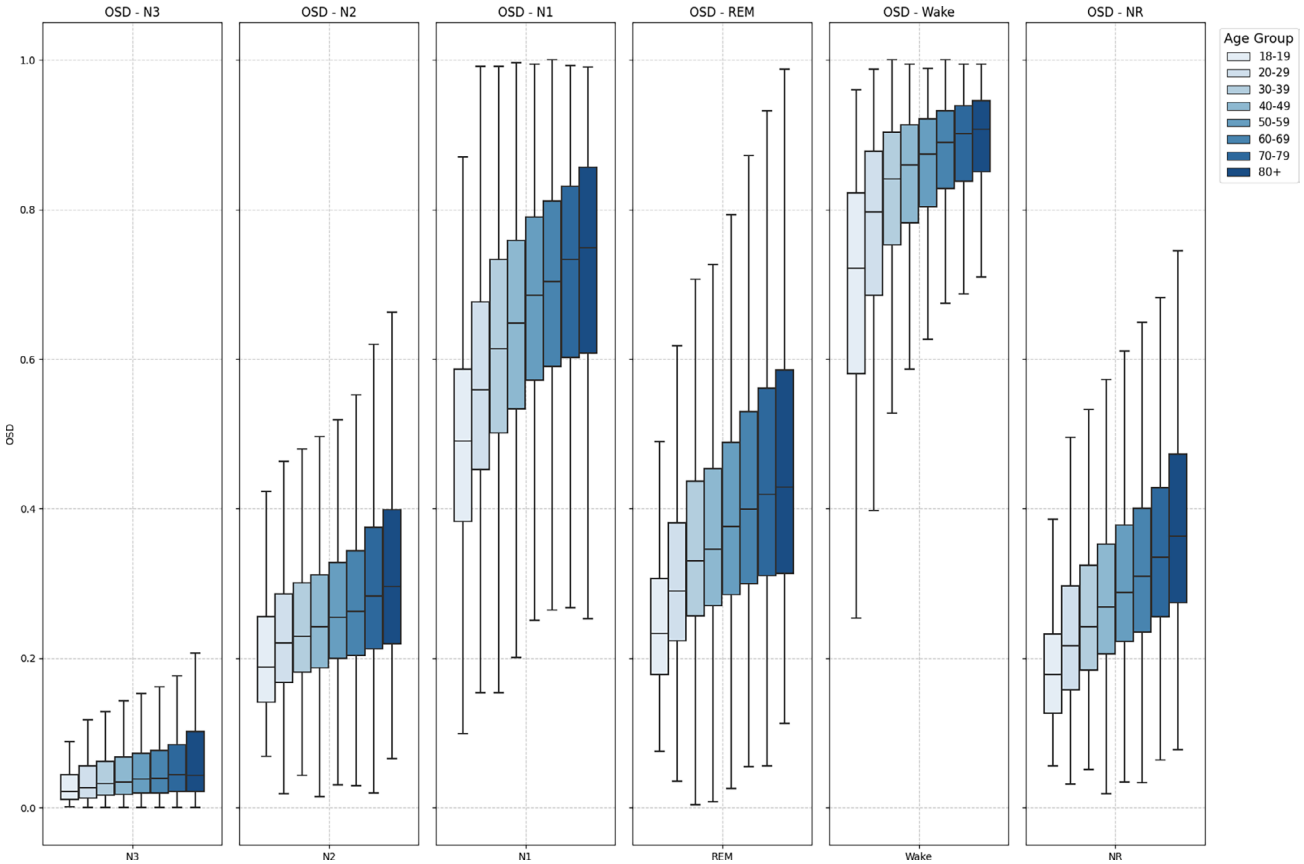

FIGURE A3 | Sleep distribution per AASM stage for age per algorithm: OSD shows a strong overall trend of lighter sleep with increasing age.

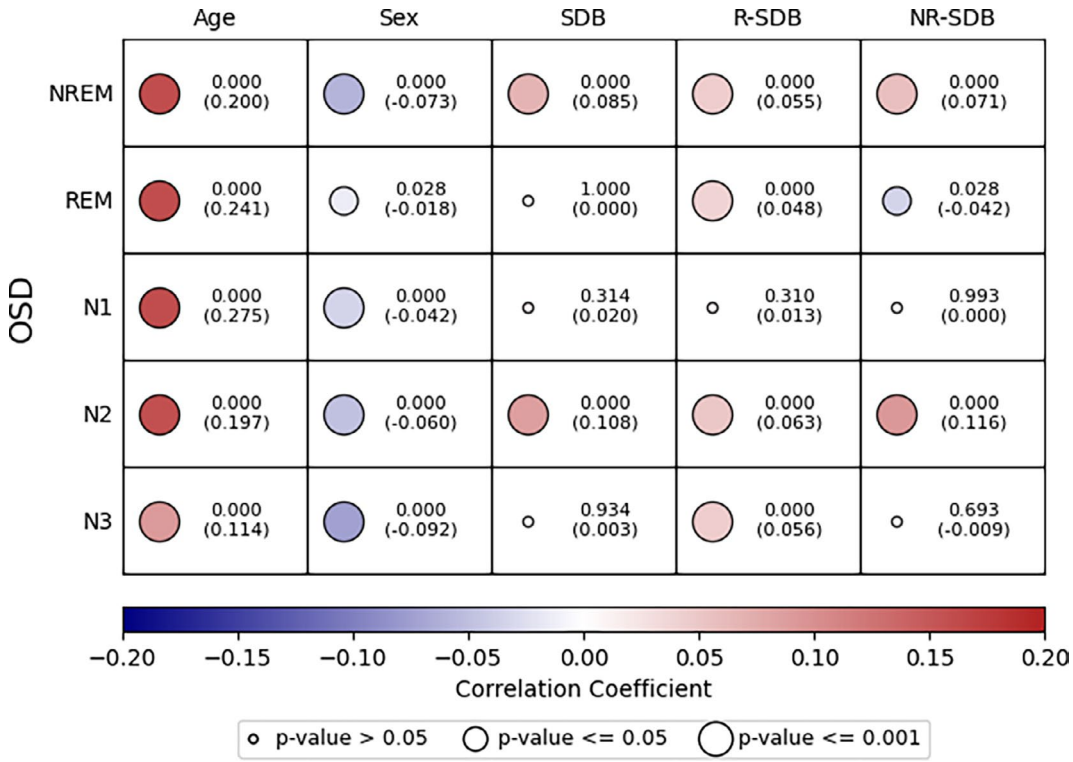

FIGURE A4 | Sleep influencing factor on OSD and ORP: the *p*-values per test are summarised in this table. The rows represent the sleep stage (dependent variable) per algorithm, whereas the columns indicate the independent variable. The *p*-values are acquired by partial correlation. R-SDB = REM-dominant SDB; NR-SDB = NREM-dominant SDB; With SDB = non-dominant SDB.

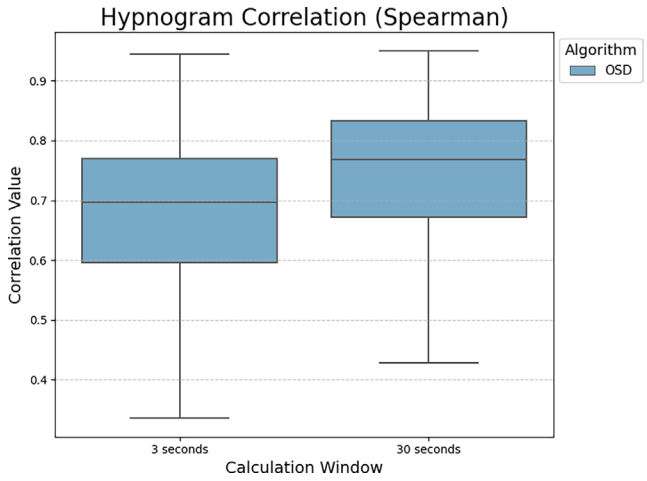

FIGURE A5 | Spearman correlation: The patient-wise Spearman correlation distribution with the hypnogram for OSD. The Spearman correlation (median [95% CI]) is 0.696 [0.694, 0.699] and 0.768 [0.766, 0.771] for the 3- and 30-s windows, respectively.

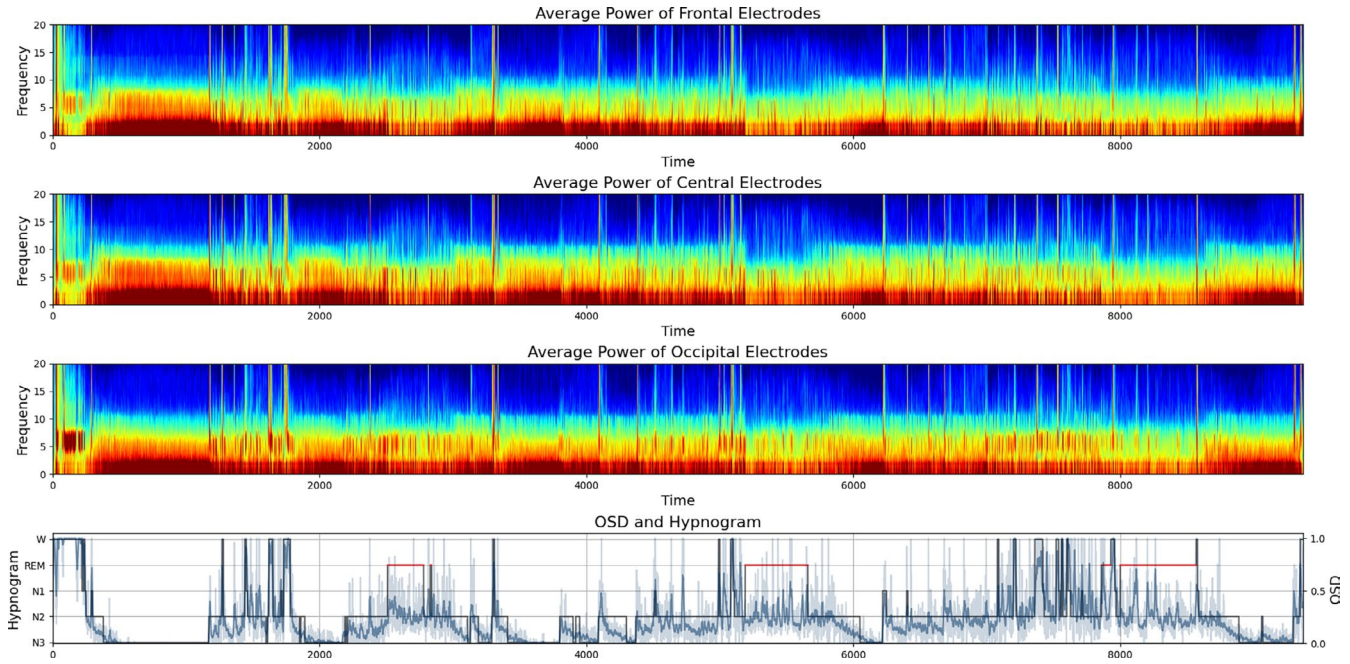

FIGURE A6 | The sleep depth algorithms in relation to the hypnogram: The first three panels are showing the power spectral density for the frontal, central and occipital electrodes, respectively. Whereas the lower fourth panel shows the OSD without smoothing in light blue and the smoothed/final OSD in dark blue. The hypnogram is shown in black where REM sleep is annotated in red.

Sleep Depth Correlates With the Hypnogram

For OSD, the correlation (median [95% CI]) with the hypnogram was 0.696 [0.694, 0.699] and 0.768 [0.766, 0.771] for the 3 and 30 s windows, respectively. The distribution among patients can be seen in the boxplots shown in Figure A5. A full night of OSD with the hypnogram is shown in Figure A6. Additionally, the average spectrogram using the mean of the frontal, central and occipital electrodes is visualised. The hypnogram is shown in black, where REM is annotated in red. The light blue prediction is the non-averaged prediction, and the darker blue is the final prediction. OSD shows a high correlation with the hypnogram as well as a gradual change at sleep stage transition.