Sleep Medicine 98 (2022) 39–52

Contents lists available at ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

# Beyond traditional sleep scoring: Massive feature extraction and data-driven clustering of sleep time series

Nicolas Decat [a, 1], Jasmine Walter [b, c, 1], Zhao H. Koh [a, 1], Piengkwan Sribanditmongkol [a], Ben D. Fulcher [d], Jennifer M. Windt [b, c], Thomas Andrillon [b, e], Naotsugu Tsuchiya [a, f, g, *]

[a] School of Psychological Sciences, Turner Institute for Brain and Mental Health, Monash University, Melbourne, Victoria, Australia
[b] Philosophy Department, Monash Centre for Consciousness and Contemplative Studies, Monash University, Melbourne, Victoria, Australia
[c] Monash Centre for Consciousness and Contemplative Studies, Monash University, Melbourne, Victoria, Australia
[d] School of Physics, University of Sydney, Sydney, NSW, Australia
[e] Paris Brain Institute, Sorbonne Université, Inserm-CNRS, Paris, 75013, France
[f] Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology (NICT), Suita, Osaka, 565-0871, Japan
[g] Advanced Telecommunications Research Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288, Japan

ARTICLE INFO

*Article history:*
Received 11 October 2021
Received in revised form
23 May 2022
Accepted 2 June 2022
Available online 21 June 2022

*Keywords:*
Sleep scoring
Massive feature extraction
Time-series analysis
Clustering
Polysomnography (PSG)
Sleep physiology
Electroencephalography (EEG)

ABSTRACT

The widely used guidelines for sleep staging were developed for the visual inspection of electrophysiological recordings by the human eye. As such, these rules reflect a limited range of features in these data and are therefore restricted in accurately capturing the physiological changes associated with sleep. Here we present a novel analysis framework that extensively characterizes sleep dynamics using over 7700 time-series features from the *hctsa* software. We used clustering to categorize sleep epochs based on the similarity of their time-series features, without relying on established scoring conventions. The resulting sleep structure overlapped substantially with that defined by visual scoring. However, we also observed discrepancies between our approach and traditional scoring. This divergence principally stemmed from the extensive characterization by *hctsa* features, which captured distinctive time-series properties within the traditionally defined sleep stages that are overlooked with visual scoring. Lastly, we report time-series features that are highly discriminative of stages. Our framework lays the groundwork for a data-driven exploration of sleep sub-stages and has significant potential to identify new signatures of sleep disorders and conscious sleep states.



## 1. Introduction

### 1.1. Background

Sleep is underpinned by a rich repertoire of biological processes. Overnight polysomnography (PSG) combining electroencephalography (EEG), electromyography (EMG) and electrooculography (EOG) shows that concerted neurophysiological events occur repeatedly across the night. The observation of cyclic patterns suggests that the brain goes through several states during sleep, each one being characterized by distinct patterns of underlying neural activity. A set of rules for classifying these states was first developed by Rechtschaffen and Kales [1]. The authors designed these guidelines with awareness of their limitations. This scoring system was originally proposed as a reference point to facilitate comparisons of research findings among researchers and clinicians, expecting that the guidelines would be updated as research and technology evolved [2]. Despite their original intentions, these guidelines are used to this day by the American Academy of Sleep Medicine (AASM) and have undergone only a few updates. The AASM visual scoring rules currently serve as the gold standard for the assessment of sleep physiology and the diagnosis of sleep disorders.

### 1.2. Visual inspection of PSG data

The AASM sleep scoring manual provides guidelines for

* Corresponding author. School of Psychological Sciences, Turner Institute for Brain and Mental Health, Monash University, Melbourne, Victoria, Australia.
*E-mail address:* naotsugu.tsuchiya@monash.edu (N. Tsuchiya).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.sleep.2022.06.013
1389-9457/

categorizing sleep and wake states in five discrete stages (Wake, N1, N2, N3, REM). These rules are based on the visual inspection of neurophysiological signals. Sleep technicians typically examine patterns of characteristic rhythms in the neural signals as well as specific physiological events to classify 30-s segments of sleep. For example, REM sleep is defined by the AASM as a stage of low amplitude, mixed frequency activity accompanied by low chin muscle tone and rapid eye movements. These rules, based on visually recognizable markers, are easy to describe and to standardize across laboratories, playing a significant role in the development of sleep research over the decades.

### 1.3. Arbitrariness of scoring guidelines

Because the scoring guidelines have undergone only minimal revision since they were established, they mostly reflect the state of knowledge and practice from past decades [2]. A typical example is the 30-s window rule, which was introduced at the time brain signals were recorded on 30-cm wide polygraph pages moving at 1 cm/s [3]. Clearly, our current technology makes this obsolete and known micro-states [4,5] cannot be captured with this time scale. Regarding the PSG features that were selected for sleep scoring, some of them were chosen based on their visibility to human eyes in preprocessed time series, with minimal contribution from modern data analysis methods. In fact, at the time the AASM sleep scoring guidelines were published, 26 of the 29 visual scoring rules were determined by consensus among human scorers and were proposed without a complete validation of their biological relevance to sleep [6,7]. For instance, the 75-μV amplitude criterion for slow waves is arbitrary [6] and is not suited to the assessment of sleep quality in older people, who often do not have delta oscillations this large [8,9]. Furthermore, some of the most distinctive features specified by the AASM, such as sleep spindles in N2 or rapid eye movements in REM, are neither specific nor necessary to the scoring of the stage, which hinders a reliable and accurate description of sleep data. The arbitrariness and incompleteness of current scoring rules and analytic approaches is a known but unresolved limitation in the field of sleep research and medicine.

### 1.4. Inadequacy in characterizing sleep-stage transitions

One of the prominent problems that scorers often encounter regards epochs that are ambiguous and cannot be dealt with using the AASM criteria. Such ambiguities arise during sleep-stage transitions, where the underlying biological states presumably change gradually. Hence, interscorer disagreement mainly occurs between adjacent stages [10]. To cope with ambiguous epochs, the AASM recommends taking into account information from the neighboring epochs [11,12]. For example, a 30-s epoch with no characteristic EEG rhythms or transient events is scored as N2 if the epoch prior to or posterior to it shows sleep spindles. While this practice assumes continuity in the underlying sleep physiology and its dynamics, it is subject to highly variable criteria across scorers. Crucially, the analysis of sleep-stage transitions has revealed crucial insights in the diagnosis of several sleep disorders including narcolepsy [13–15], chronic fatigue syndrome [16], and insomnia [17]. Furthermore, this 'temporal smoothing' process tends to underestimate transitional states such as N1, which is not clearly identifiable visually but is associated with various processes like mental imagery [18] and memory processing [19]. The contextual information rules, which are used to cope with the inability to visually identify and verbalize distinctive features, are not suited to the definition of sleep-stage transitions.

### 1.5. Limited characterization of sleep time series

Despite all these known issues, traditional sleep scoring continues to rely on visually identifiable features in PSG signals. While raw EEG signals are noisy, non-stationary, and high dimensional, the traditional guidelines supplement visually identifiable features with only a few signal attributes, such as the dominant power in certain frequency bands. Thus, current guidelines do not fully capture the range of dynamical patterns that could underlie relevant changes in sleep physiology. In addition to representing a narrow range of signal properties, the visualization of certain features is undermined following EEG data preprocessing. Filtering, for example, makes it difficult for scorers to discern sleep spindles among the dominant slow wave activity in N3 [20]. Similarly, current guidelines for the referencing of EEG electrodes with mastoid electrodes favor global, centro-frontal components and could mask more local sleep features. Therefore, we might gain more understanding by characterizing sleep dynamics using the sophisticated range of modern time-series analysis algorithms.

### 1.6. Purpose of this study

The goal of the present study is to categorize sleep in a data-driven way, based on a wide range of their time-series properties. To achieve this goal, we first apply a massive feature extraction tool to sleep time series, using the highly comparative time-series analysis approach implemented in the *hctsa* software package, which contains a diverse set of over 7700 time-series features [21,22]. *hctsa* encodes scientific algorithms for time-series analysis in the form of *features* that each summarize a structural property of a time series in the form of a real number. It includes methods that were developed in and applied to a wide range of research areas, including Fourier and wavelet transforms, entropy, self-correlation and predictability, nonlinear time-series analysis, and fractal scaling. This large set of analyses has never before been applied systematically to PSG data.

Based on the results of these *hctsa* time-series analyses, we sought to organize sleep (i.e., N1, N2, N3, REM and intra-sleep wakefulness) independently of pre-defined scoring rules. To this end, we approached sleep scoring in an unsupervised way by clustering sleep time series based on the similarity of the dynamical properties captured by *hctsa*. Furthermore, we analyzed the data on an epoch-by-epoch basis to minimize subjective criteria related to the incorporation of contextual information. By taking this data-driven approach, which we refer to as feature-based clustering, our aim was not to replicate AASM visual scoring, but to go beyond it (Fig. 1). Our approach tackles the following question: if the AASM scoring guidelines did not exist today, how would we go about using modern methods to define the structure of sleep? While our approach draws on certain assumptions from the AASM (such as the existence of 5 distinct sleep stages), our method proceeds independently of the specific AASM guidelines for scoring/identifying these stages and is a first step towards moving beyond conventional sleep stages themselves.

We present our analyses in two parts. In the first part, we assess the extent to which the feature-based clusters correspond with visual (AASM) sleep scoring of the same epochs. In the second part, we examine cases in which feature-based clusters disagree with AASM labels to better understand what novel properties of sleep our approach can reveal. We also assess the ability of individual *hctsa* time-series features to distinguish AASM stages.

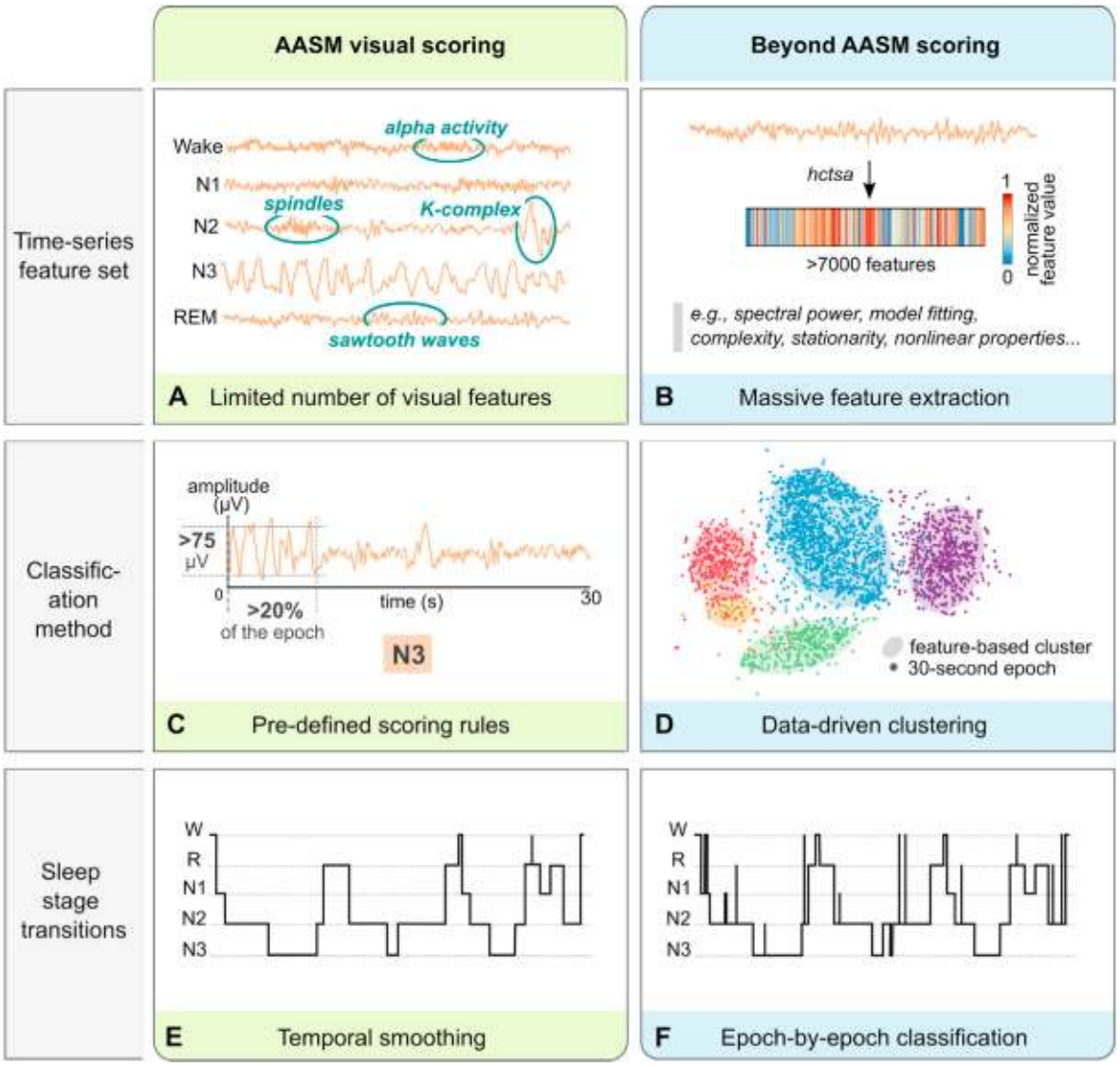


**Fig. 1. Our feature-based clustering framework seeks to provide a better description of sleep.** The present study introduces a sleep classification framework that represents a first step towards expanding our understanding of sleep physiology. Our framework (right column) differs from the traditional AASM scoring method (left column) in three main ways: the number of features extracted (first row), the classification method (second row) and the scoring of sleep-stage transitions (third row). **A**) The traditional AASM scoring method is based on a limited number of visual time-series features, such as alpha activity, sleep spindles, K-complexes and sawtooth waves. **B**) To go beyond our understanding of sleep, we need to broaden the characterization of sleep time series. We use highly comparative time-series analysis (*hctsa*) to extract a diverse set of properties using measures of data distribution, correlation properties, model fitting and others. *hctsa* transforms each 30-s epoch into a single vector containing the output values produced by all *hctsa* features for this epoch. **C**) The AASM scoring manual provides pre-defined rules for the classification of sleep stages. We illustrate the example with a N3 epoch, which is scored as such when slow wave oscillations of >75 μV amplitude occupy more than 20% of the sleep epoch. Note that both the 75-μV criterion and the 30-s timescale are not based on any scientific basis (see Introduction). **D**) Our framework approaches sleep scoring in an unsupervised way by clustering sleep time series based on the similarity of their properties captured by *hctsa*. We used *k*-means clustering to organize sleep in five feature-based clusters. **E**) To deal with visually ambiguous epochs, which are mostly encountered in sleep-stage transitions, human scorers are encouraged to use the surrounding epochs. This results in a temporal smoothing of the sleep structure. **F**) Unlike traditional scoring, our framework scores each 30-s epoch in isolation, thus preserving the temporal structure of the data. Altogether, these three methodological components represent a first step towards a novel data-driven understanding of sleep, which complements the traditional approach.

## 2. Results

### *2.1. Part 1. Correspondence between feature-based clustering and traditional sleep stage architecture*

To objectively assess the correspondence between how sleep epochs are organized using (i) time-series features and (ii) AASM labels, we developed an automated procedure that combines unsupervised clustering, cluster matching, and cross-validation (Fig. 2; see Methods; *Feature-based clustering*). We applied *hctsa* to 12 full-night PSG datasets, each containing a preprocessed EEG, EOG, and EMG derivation. Through this process, the EEG, EOG and EMG time series of each 30-s epoch is converted into a single vector containing the output values produced by all *hctsa* features for this epoch (Fig. 2a). The feature values which generated non-real numbers or errors were filtered out and the remaining valid values were normalized. After forming a balanced dataset (Fig. 2b and c), we grouped 70% of the data (training set) into 5 clusters based on their *hctsa* feature values (Fig. 2d). Using our sequential maximum matching algorithm, we assigned each unlabeled training cluster to one of the five AASM sleep stages (Fig. 2e). We named the five training clusters C1, C2, C3, CR and CW, as those that

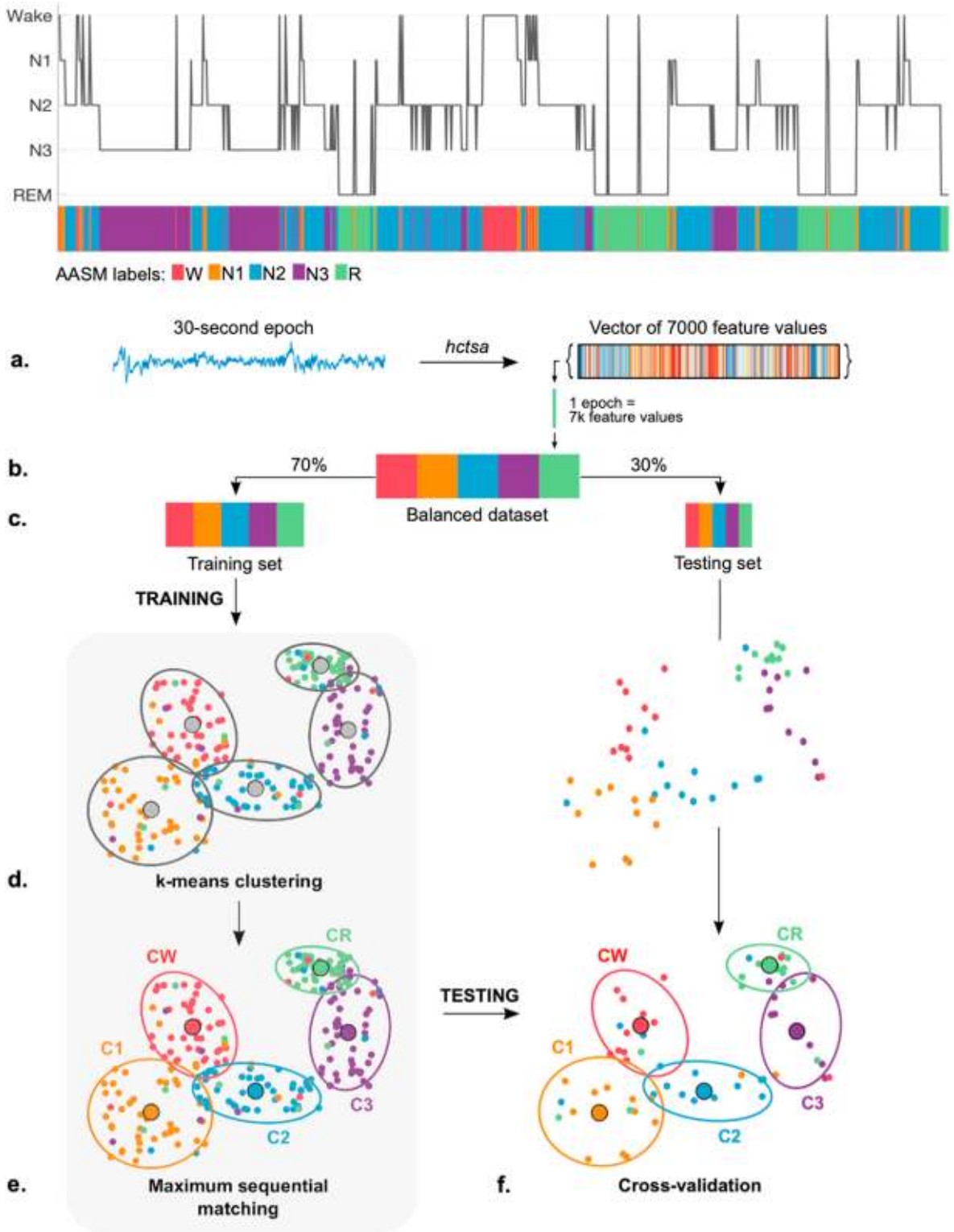


**Fig. 2. Computational steps in our unsupervised feature-based clustering approach**. a) We applied *hctsa* on PSG data, which converted each 30-s epoch into a single vector containing the output values produced by all *hctsa* features for this epoch. b) To equally represent each AASM stage, we formed a balanced dataset by random subsampling. c) We randomly assigned 70% and 30% of the balanced dataset into a training and testing set, respectively. d) We used the feature values of the training set for *k*-means clustering (k = 5). Black ellipses depict clusters. Their centroids are depicted as large gray circles. Each small colored dot indicates the given AASM label, representing the vector of feature values from one training epoch. e) Next, we mapped each unlabeled cluster onto one of the five AASM sleep stages based on the most frequently matched AASM label in each cluster using our sequential maximum matching algorithm. We named the 5 clusters as CW, C1, C2, C3 and CR, which were matched with W, N1, N2, N3 and REM, respectively. f) After completion of training, each test epoch was assigned to the nearest cluster (i.e., cross-validation). To assess the overlap between cluster decisions and the AASM labels, we computed the percentage of test epochs whose AASM label matches the cluster it was assigned to (e.g., N2 epoch assigned to C2). The entire procedure was iterated 100 times per dataset. See Methods (*Feature-based clustering*) for details on each step of the procedure.

most closely mapped onto N1, N2, N3, REM and wake, respectively. Then, we assigned the remaining 30% data (testing set) based on their distance to the training clusters (Fig. 2f). We used the proportion of match between the AASM labels and the cluster labels of the test data (cross-validation) through 100 iterations to assess the correspondence between the two methods.

To visualize how the large number of *hctsa* time-series features vary across a night of sleep, we show a color-coded feature matrix representing the feature values for each 30-s EEG, EOG and EMG time series from one subject over one night (Fig. 3). The matrix displays the normalized feature values for 5946 valid features on 1014 epochs in chronological order. Below the matrix, we show the hypnogram generated from the AASM labels supplied by consensus between three trained scorers. We observe that the hypnogram generally corresponds with patterns of feature values, notably for the EEG and EOG. More specifically, notice that the short periods of wake (e.g., during the 3rd hour; green arrows at the top) correspond with narrow bands of feature values that contrast with neighboring epochs of N2 sleep. This broad correspondence can be explained by the apparition of high-amplitude hyper-synchronized slow waves at the transition from wakefulness to deep sleep [11,12], resulting in a sharp increase in the autocorrelation of the EEG signal. The fact that many *hctsa* features reflect these wake-NREM transitions could thus stem from their sensitivity to the signal autocorrelation. As a result, the *hctsa* feature matrix broadly tracks the sleep architecture described by visual sleep scoring. However, we also note instances of potential discordances between the two methods. For example, during the continuous REM period (marked by a red bar at the top), multiple patterns of feature values are identifiable. This hints at heterogeneity in time-series properties within conventional sleep stages, which we examine in more detail later.

### 2.2. Feature-based clusters substantially overlap with visual sleep scoring

To what extent do feature-based clusters correspond with AASM sleep stages? Unlike traditional scoring, our framework organizes sleep epochs without considering their temporal context. Furthermore, it is unclear what correlation structures in the feature space would emerge from this specific type of data, which makes it difficult to predict whether the application of uniformly-weighted features will produce an AASM-like structure.

To quantify the degree of correspondence between the two approaches, we computed the mean overlap between the two classifications across datasets (Fig. 4a). While the highest overlap was observed for N3 (77.4% of the N3 epochs were assigned to C3), the lowest degree of agreement was obtained for N2 (only 43.1% of the N2 epochs were attributed to C2). The majority of N2 mismatch was due to the assignment into C3 (20.9%). We also note high off-diagonal overlap between Wake, N1, and REM epochs (e.g., 27.6% of N1 assigned to CR, 26.6% of Wake to C1, 21.3% of REM to C1). Overall, this level of overlap demonstrates the ability of our stage mapping algorithm to attribute each cluster to the corresponding AASM stage without *a priori* information on how sleep stages are defined.

To illustrate what each data-driven cluster looks like, we show exemplar EEG time series located near the cluster centers from one dataset (Fig. 4b–f). Feature-based clusters exhibit visual features that are descriptive of AASM sleep stages (e.g., CW and C1 are characterized by high alpha activity, C2 by the presence of K-complexes, C3 by slow-wave activity). This analysis demonstrates that our data-driven approach reveals—without any use of AASM labels—a cluster structure in the data, and this structure reflects the key visual aspects that characterizes traditional scoring.

### 2.3. Part 2. Revealing the source of divergence between feature-based clustering and traditional scoring

#### 2.3.1. Contextual information does not fully explain the observed discrepancies

While the feature-based clusters and traditional sleep-stage labels largely overlap, our novel framework also revealed interesting divergences. Our visual inspection of the epochs with mismatched labels suggested that instances of disagreement may stem from the use of contextual information in AASM scoring. While human scorers are encouraged to use neighboring epochs to assist in labeling ambiguous epochs, our algorithm analyzes each individual epoch in isolation.

To test the possibility that the observed divergences are due to the use of contextual information in human scoring, we recruited 4 expert sleep scorers ('No-context scorers') and asked them to score a subset of the epochs from the same data and under the same conditions as our feature-based clustering approach. Specifically, we first equated the number of epochs across sleep stages, then presented each epoch in isolation and in random order (see Methods; *Visual scoring task*). Under these settings, no-context scorers agreed with the trained scorers for 67.5% of the epochs. This is lower than the ~80% interscorer agreement obtained in sleep studies [10], implying that contextual information may explain ~12.5% of the disagreement. Importantly, no-context scorers agreed with feature-based clustering in 41.1% of cases, similar to the level of agreement with trained scorers (45.2%), suggesting that the use of contextual information does not fully explain the disagreement between feature-based clusters and the implementation of AASM scoring rules.

### 2.4. Feature-based clustering points to potential sub-stages within AASM sleep stages

To explore the source of discrepancy other than contextual information, we show, in Fig. 5, the same feature matrix plotted in Fig. 3 but with epochs grouped according to the AASM labels (see bottom annotations). Additionally, within each AASM label, we reordered epochs according to the cluster assigned by the algorithm (CW, C1, C2, C3, and CR; see colorbar). Fig. 5 shows that, within each AASM stage, epochs assigned to different clusters coincide with visually distinct patterns of EEG and EOG feature values (less so for EMG). This observation is most striking within N2. Feature-based clustering assigned N2 epochs to five different clusters, each displaying distinct patterns in feature values. Fig. 6 shows a visualization of the N2 epochs (same data as in Figs. 3 and 4) across subjects using *t*-distributed stochastic neighbor embedding (*t*-SNE; see Methods; *t-SNE*). In 10 out of 12 subjects, we observe a heterogeneous structure (marked by dotted lines) within N2. In other words, according to our feature-based clustering approach, N2 is far from being a uniform state.

We conclude that the divergence between feature-based clustering and AASM scoring is driven not just by the lack of contextual information in our approach compared to traditional scoring, but also by the detection of neurophysiological heterogeneity within each AASM stage. The heterogeneity of N2 and other AASM stages has been reported in previous studies and precise characterizations in sub-stages have been proposed (see Discussion). This heterogeneous structure might reflect physiologically distinct periods of sleep that are overlooked with traditional scoring.

### 2.5. Identification of the most discriminative features

Lastly, we aimed to illustrate how *hctsa* can uncover new types of time-series features for the classification of sleep data. While

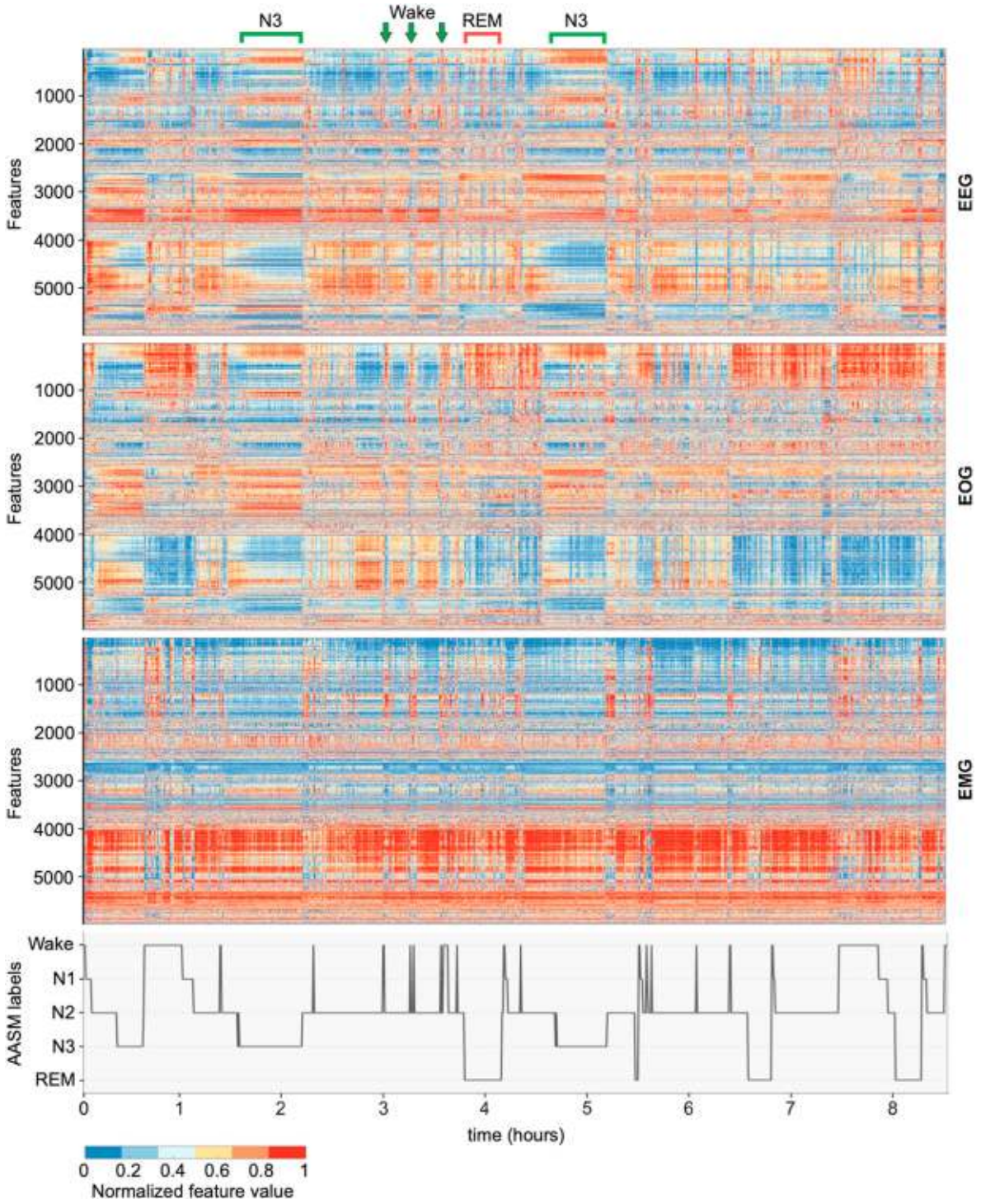


**Fig. 3. Time-series properties broadly track visual sleep scoring**. The matrix is derived from applying 5946 *hctsa* features to one full-night dataset (ID: 1800439). Each column corresponds to one of the 1014 30-s epochs recorded by an EEG, EOG, and EMG (top, middle, and bottom, respectively). The x-axis shows time (in hours) from the first non-wake epoch. Each row corresponds to one *hctsa* feature. To help interpretation, we order features in EEG by their correlation-based similarity using average linkage clustering [21,22]. We order features in EOG and EMG using the same ordering as in EEG. Each pixel encodes a normalized feature value from 0 to 1 (see colorbar below the matrix). Below the feature matrix, we show the hypnogram (the labels derived by consensus between three trained scorers in accordance with the AASM visual scoring guidelines). The hypnogram is broadly consistent with different patterns in the feature matrix (notably for the EEG and EOG matrices).

outside the scope of the goal of inferring the structure of sleep time series using unsupervised methods, we aimed to demonstrate the potential application of time-series feature extraction for tackling supervised learning settings for sleep analysis problems. We demonstrate the approach by aiming to deduce the most discriminative time-series features for each pair of AASM sleep stages (see Supplementary Text 1 for more details on the methods and results). Our approach identified measures that recapitulated some of the methods used in the sleep-scoring literature, and flagged novel features. For classifying Wake/N2, Wake/N3, N1/N3,

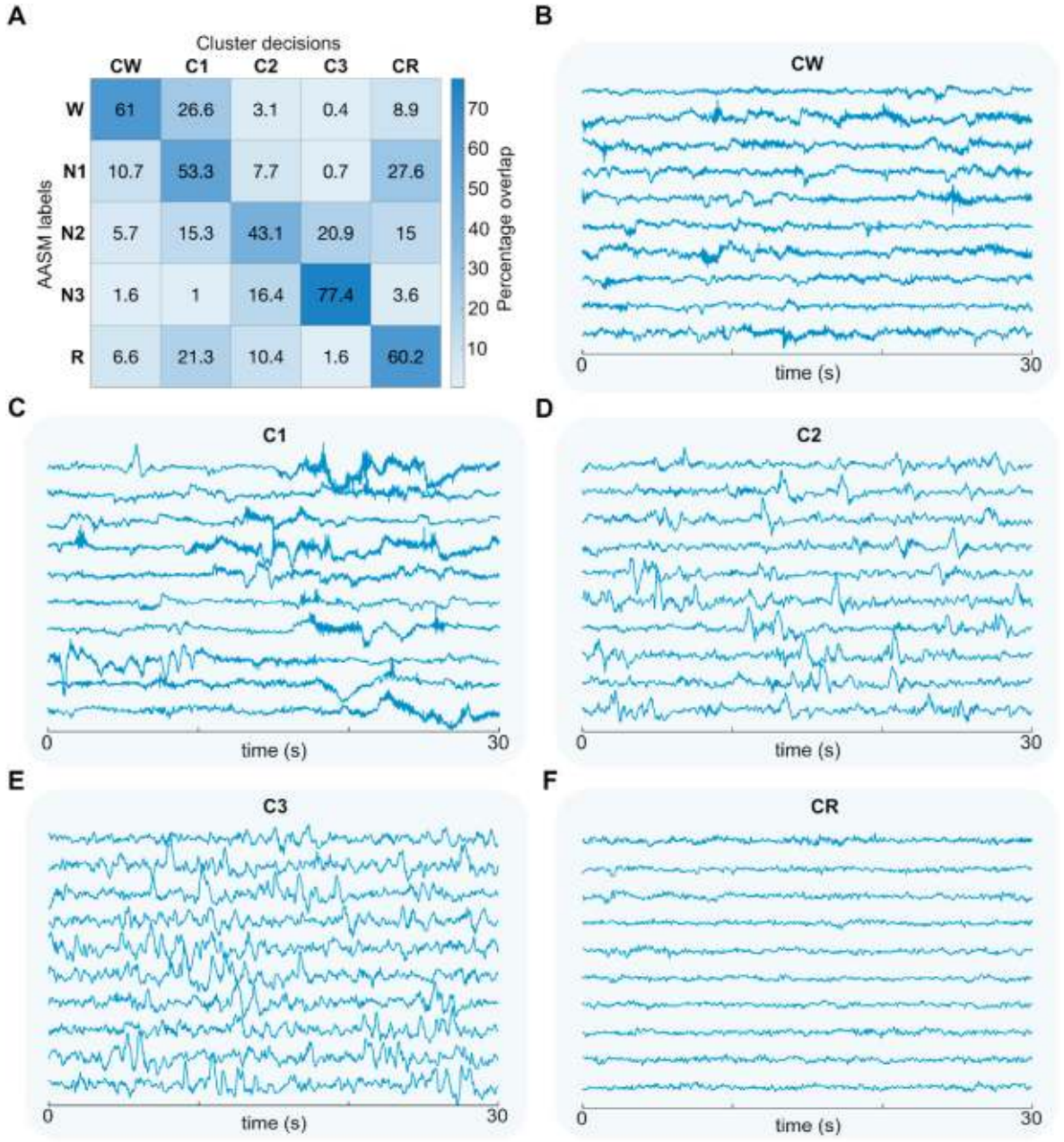


**Fig. 4. Overlap between the AASM visual scoring and our feature-based clustering across stages**. **A**) Overlap matrix between our feature-based clustering results on the test data (columns: CW, C1, C2, C3, and CR) and the AASM labels (rows: Wake, N1, N2, N3, and REM). The number in each cell represents the mean overlap (in %) across the 12 datasets. **B**) CW. **C**) C1. **D**) C2. **E**) C3. **F**) CR. For each cluster, we selected 10 EEG time series (from one dataset; ID: 1800005) that were closest to the cluster centroids. Each line corresponds to the EEG time-series at electrode C3 for one of these 30-s-long epochs.

N1/REM, N2/REM, and N3/REM, the most discriminative time-series features measured spectral properties, such as Fourier spectral analysis, power in some frequency bands, and logarithmic power. The slope of the linear fit (i.e., scaling exponent) of the power spectrum was the most discriminative feature across all pairs of stages (83.2%; percentage accuracy of this single feature averaged across all pairs and across 100 iterations). This feature could distinguish stages associated with poor classification performance using classical EEG properties, such as N2/REM (70.3%) and N1/REM (68.0%). The most discriminative features also included a range of novel measures that are not widely used to discriminate between AASM sleep stages. This includes distributional moments (e.g., ninth-order moment; 83% for N1/N2), outlier distribution (e.g., fit to beta distribution; 71% for Wake/N1) and variance (e.g., mean absolute deviation; 87% for N2/N3; Fig. S1). Incorporating these novel features into automated sleep scoring could improve performance while providing additional insights into sleep physiology (see Discussion).

## 3. Discussion

We present a sleep time-series analysis framework that goes beyond traditional visual sleep scoring. To do so, we 1) leverage over 7700 features to comprehensively characterize sleep time series, 2) cluster time series in an unsupervised way, rather than classifying them according to the AASM guidelines, and 3) analyze

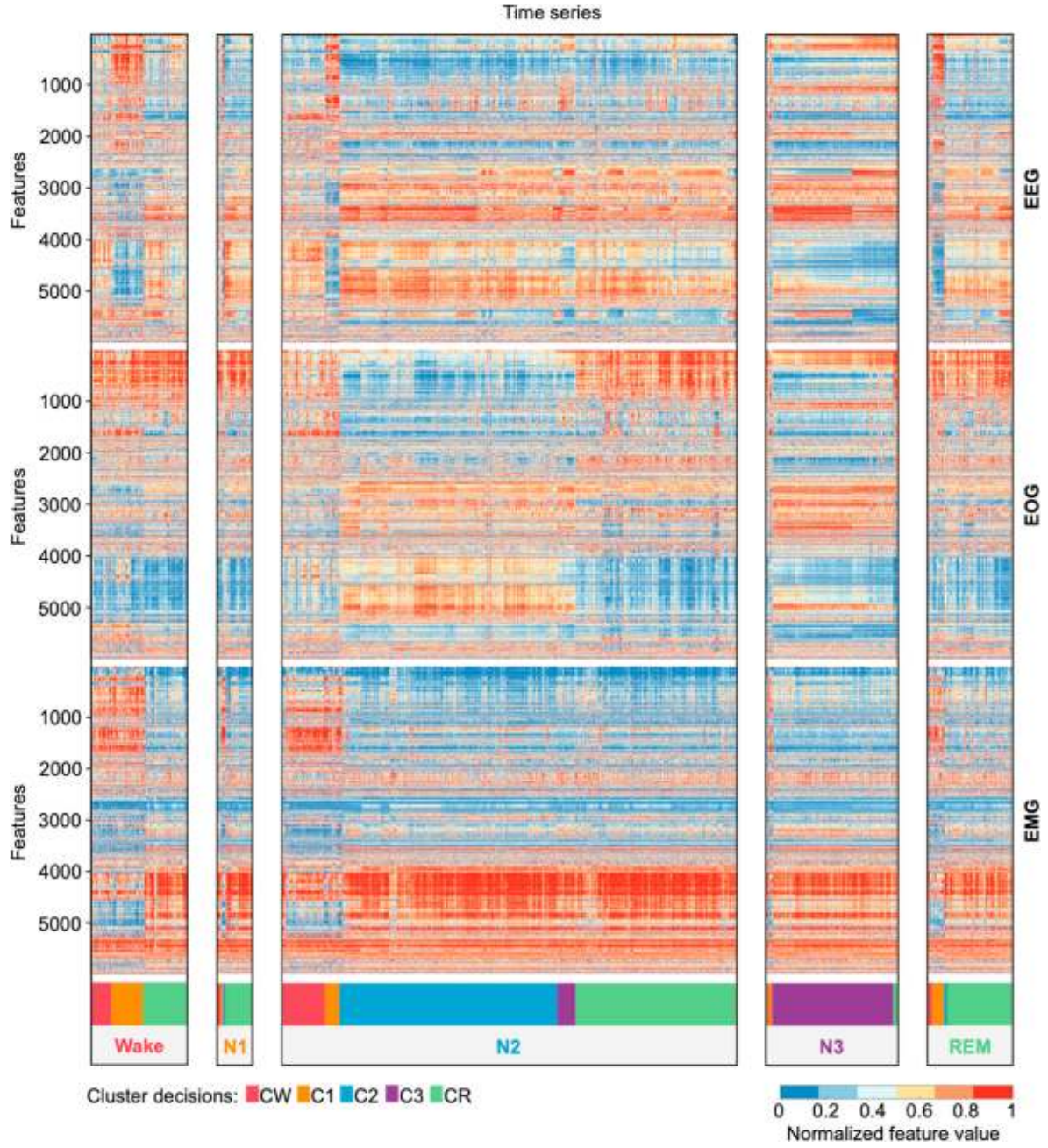


**Fig. 5. Feature matrix points to heterogeneity within each AASM sleep stage.** The same 1014 epochs in Fig. 3 are plotted (see legend of Fig. 3), except for a different ordering across the x-axis. From the left to right, we organized columns in five matrices according to the original AASM labels (Wake, N1, N2, N3, and REM). Within each AASM label, we further ordered columns according to feature-based cluster assignment (CW, C1, C2, C3, and CR; see colorbar and legend). Disagreements between the AASM labels and feature-based clustering clearly correspond to distinct patterns in the feature matrix.

the data without assuming the temporal context to resolve ambiguities arising in the scoring of sleep-stage transitions.

In Part 1, we showed that the sleep structure revealed by our approach substantially overlapped with the traditional AASM sleep stages defined by visual scoring (Fig. 3), suggesting that both approaches captured a genuine physiological structure of sleep. However, we also observed important points of divergence between the two methods (Fig. 4a). In Part 2, we examined potential sources of the divergence. While it is partially explained by contextual information, it is also reflected by the fact that *hctsa* features depicted heterogeneity in time-series properties within each of the AASM sleep stages (Figs. 5 and 6). Finally, we identified discriminative properties of high-performing features. In the rest of the Discussion, we identify various directions our framework can take to pick up low-hanging fruits in its applications.

### *3.1. Towards establishing a general and objective ground truth of sleep stages*

The present framework represents a complementary approach to the existing method aiming to provide a more accurate and objective definition of the sleep structure. In line with past fMRI [23] and EEG [24] studies revealing latent states within traditional stages by means of machine-learning algorithms, our novel data-

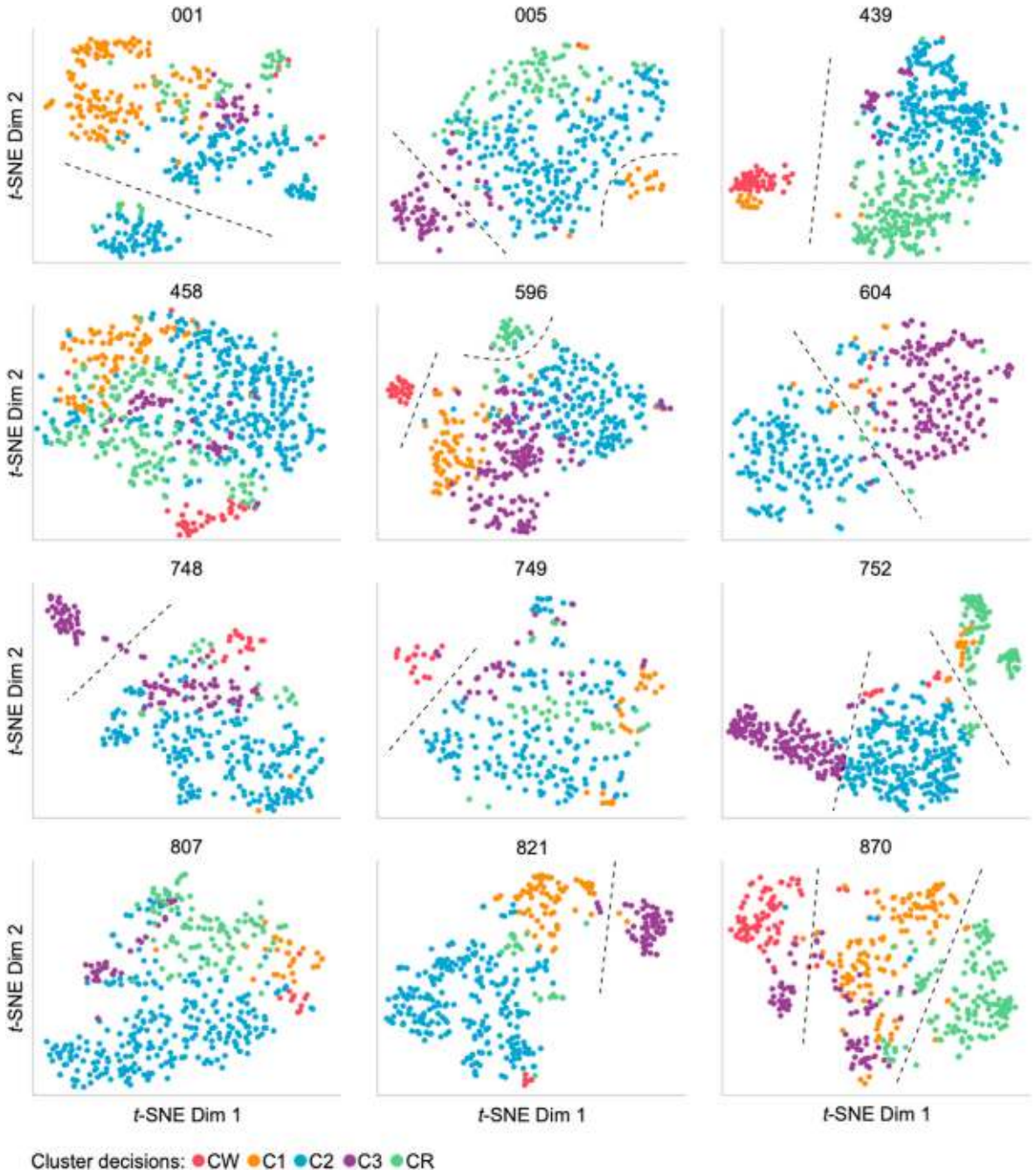


**Fig. 6. Heterogeneity in time-series properties within N2.** *t*-SNE two-dimensional projection of N2 time series (EEG, EOG, and EMG) for each of 12 datasets. Units are arbitrary. The title refers to the dataset index. Cluster assignments (indicated by the color of each dot) were obtained separately for each dataset as in Fig. 5 (CW, C1, C2, C3, CR; see legend). For 10 of the 12 datasets, we observe a heterogeneous structure within the N2 stage, as illustrated by black dotted lines. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

driven approach will expand our understanding of sleep physiology, which has been difficult to achieve with traditional methods.

As an estimate of what the traditional method would tell us, we used AASM standards and examined how our framework differently characterized and organized sleep data. Although the AASM visual scoring is the gold standard, we do not consider this to be the ground truth of sleep physiology. As we argued in the Introduction, conventional scoring provides an inadequate description of the physiological changes that occur during sleep. Given the lack of ground truth, our study raises the question of what a better sleep classification approach would involve and, more specifically, what "better" means in this context. Our data-driven approach, equipped with a large number of features, offers the possibility to fine-tune a wide range of parameters. Depending on the purpose of the study, therefore, these parameters can be adjusted. This includes, for example, selecting features that are the most sensitive to a specific sleep stage, sleep function or sleep pathology. In this regard, our approach has significant potential to contribute to a more complete understanding of the nature of sleep in a flexible manner.

One interesting future research direction is to apply our framework to data from the same subjects across multiple nights. By dividing the data into training and test nights, we can apply

similar cross-validations as we established in our framework (Fig. 2). Such studies would be an ideal way to estimate the sensitivity and specificity of our framework, providing a benchmark for further studies.

The confirmation of the stability and precision of our framework, in comparison with the traditional AASM scoring approach, can also be tested across subjects. An initial assessment of the generalizability of feature-based clusters suggests that similar data-driven structures are observed across subjects (see Fig. S2). These preliminary findings support the idea that these clusters reflect a physiologically relevant structure that generalizes across subjects. Follow-up investigations using a larger sampling size are needed to provide further insights into between-subject similarities in sleep structure. Such a generalization study should also include populations that are more heterogeneous than the cohort we used (e.g., wider age range). How much individual variability will our framework reveal on its application to massive healthy population data? Can it explain variance in mental and physiological differences in a more sensitive and specific way than traditional approaches? Our framework provides a first step towards answering these questions.

### 3.2. Refining traditional sleep stages

In Part 2, we took an initial step towards the discovery of sub-stages that are overlooked with the traditional scoring. We focused on N2 because this stage had the lowest level of agreement between our approach and AASM scoring (43.1%; Fig. 4a) and represented the largest amount of the data available in each dataset (Fig. 5). *t*-SNE projection exhibited heterogeneity within N2 in a visually striking way and in the majority of datasets (Fig. 6).

Variability within N2 was previously reported. Sleep spindle activity does not occur uniformly within N2, but their occurrence is modulated across and within sleep cycles [25] and is dependent on local temporal context [26], which may reflect fluctuations in sleep depth within N2 [27]. Along this line, based on autonomic and hormonal activity, the division of N2 has been proposed as a quiet type (before the transition into N3) and an active type (preceding REM sleep) [28]. From a physiological perspective, it is not surprising if N2, which occupies as much as 50% of total sleep [29], is not physiologically homogeneous.

While we focused on N2 in Fig. 6, we believe that this observation is not restricted to this stage. The feature matrix presented in Fig. 5 suggests variability in time-series properties within every AASM sleep stage. To the extent that these variations reflect underlying changes of sleep, our findings indicate that sleep staging can be refined to unveil more subtle sleep-stage distinctions. These distinctions may be defined within AASM sleep stages, similar to how REM sleep can be decomposed into a phasic and tonic period [30], each of which is associated with different physiological and cognitive processes [31]. Sub-stages may also not necessarily align with coarse sleep stages but instead emerge as transitional stages, similar to how N1 briefly appears between the wake state and N2.

The delineation of sub-stages can also be achieved using much shorter time scales than the traditional 30-s time window. As the present study sought to compare the structure emerging from feature-based clustering with that of traditional scoring, we adopted the 30-s window as the unit of our analysis. A previous study showed that the AASM scoring can be successfully replicated using time windows down to 5 s [32]. Our framework is highly flexible too and can be applied at shorter time scales, which is likely to be optimal for the reliable identification of sub-stages, as demonstrated for example within N1 [4]. The temporal refinement is also most suited to the incorporation of short-lived events, such as sleep spindles or K-complexes, which may be associated with the ability to report dream content [33,34]. Application of our automated framework therefore promises highly productive areas of future research by refining the temporal scales of the analysis.

Our approach is also scalable in space. In fact, the EEG technology (as well as other physiological variables, other than EOG and EMG) has progressed tremendously in quality and quantity since the time of the original development of the AASM guidelines. Especially, the high quality and high-density EEG (60 to 256 channels) are more and more adopted in sleep research in both healthy and clinical populations. Higher spatial resolution in EEG has led to discoveries that have reshaped our conceptualization of sleep physiology, with the characterization of NREM slow waves as traveling waves [35], the identification of two types of NREM slow waves [36], the presence of slow waves in REM sleep [37] and even in wakefulness [38], or the identification of inter-hemispheric differences in slow wave activity [39]. This recent research has revealed the importance of the local aspects of sleep [40]. This notion of “local sleep” challenges the view that sleep is a global phenomenon and raises the question of locally defined sub-stages [41]. While these studies point to the importance of analyzing all available channels of EEG, the AASM guidelines ignore much of the spatial information. The incorporation of spatial information in our framework is highly feasible and represents another low-hanging fruit for its application.

### 3.3. Broadening the characterization of sleep time series

To our knowledge, a comprehensive and interdisciplinary library of time-series analysis, such as *highly comparative time-series analysis* (*hctsa* [21,22]), has never been systematically applied to sleep EEG data. While sleep scoring studies typically investigate the significance of single measures, *hctsa* offers the opportunity to assess the discriminative power of over 7700 features at once (for information on the level of redundancy of the *hctsa* library, see Fig. S3). In the last part of the paper, we illustrated how our approach can help identify highly discriminative features for the distinction of the different pairs of stages. These features could incorporate additional meaningful information in sleep scoring and refine the current description of sleep architecture.

While some may be concerned that our approach will drown researchers in the massive number of features in space and time, this problem can be overcome by leveraging machine learning techniques [32], a highly promising interdisciplinary research agenda (e.g., expediting and replicating traditional sleep scoring by humans [42]). While this is not what we aimed for in the present work, we see a similar approach can be usefully incorporated within our framework.

Our data-driven approach recapitulated some of the existing sleep scoring literature, but also identified less frequently used features that have also been described as discriminative in past automatic scoring studies. An interesting feature highlighted by our analyses is the slope of the linear fit to the power spectrum—the most discriminative feature across all 10 pairs of stages—which has been recently used to track changes in consciousness [43] or vigilance state [44]. We also found highly performing features that have rarely received attention, including visibility graphs, symbolic motifs, automutual information, distributional shape and multiscale entropy. It is unclear, however, if the specific features we identified generalize to a larger and more diverse dataset. Future work needs to test if our results replicate among a larger and more diverse population.

Our approach examined all of these features at once without cherry-picking which ones to test and report. Selective testing and reporting of a single feature in an individual study runs a significant risk in the replication crisis. This is a pervasive problem across the

scientific fields, but in particular, it is a significant problem in the young field of data-analysis intense neuroscience (e.g., Ref. [45]). Such risk can be minimized using time-series feature extraction tools like *hctsa*.

### 3.4. Future implications

As we have already mentioned, there are already various feasible and impactful research projects that can arise from our framework. As such, this paper is a first step towards the establishment of a purely data-driven sleep classification method to discover sleep sub-stages that are not visible to the human eye.

We also note that our framework (which focuses just on the univariate dynamics of individual system elements) can be extended to incorporate interactions between pairs of elements of the system represented by the full multivariate time series [46]. Quantifying pairwise dependences between pairs of time series, e.g., using methods like correlation and coherence, has been attracting attention in signal processing of the task-related EEG analysis, but not utilized in sleep staging. Systematic investigations of temporally lagged correlations between electrodes are likely to characterize functional and effective connectivity among brain areas, which would arrive at a more reliable and fine-grained definition of sleep stages based on brain mechanisms.

In terms of application of our framework, we believe that clinical applications are highly promising. The AASM visual scoring rules were originally developed for healthy subjects. Conventional scoring may be too coarse to capture relevant structural patterns in patients with sleep disorders, whose sleep is often fragmented [47]. Moreover, sleep architecture is generally characterized by an abnormal time course of sleep stages in sleep disorders [48], therefore the use of contextual information is less adequate for the classification of transitional epochs in patients. Undoubtedly, establishing an entirely new scoring approach for the clinical field is an arduous task. The diagnosis and treatment of sleep pathologies strongly relies on the way sleep is traditionally scored [12]. A thorough process of validity and reliability assessment is necessary before proposing a useful method that clinicians can trust. Our framework is an initial step towards this goal, and could first be used as a complement to traditional classification methods to expand our understanding of sleep pathologies.

Indeed, our framework, as a complement to existing scoring conventions, can promote understanding of the PSG correlates of a certain sleep state or phenomenon. In sleep state misperception or paradoxical insomnia, conventionally scored PSG recording may indicate that the patients are asleep or display normal sleep patterns, while they report being awake or having barely slept [49,50]. Current PSG analysis is still poor at reflecting the objective and biological aspects of sleep and subjective ratings of sleep quality [51,52]. In-depth analysis of PSG with *hctsa* could bridge this gap, linking objective data and subjective reports for a better understanding of sleep state misperception and showing how it differs from normally perceived sleep [53].

Another promising application of our framework is to find neural signatures of various conscious states, especially in the context of dream research [54]. In fact, the presence or absence of dreaming does not map onto one specific AASM stage [55–57]. It is highly plausible that some features (or their combinations) in *hctsa* can help find reliable brain signatures of dreaming.

Our novel framework, which applied massive time-series feature extraction to sleep data, represents an initial step to go beyond traditional sleep staging. While our framework provided broadly consistent classification results, it revealed finer structures than the traditional approach. Once validated with more extensive within- and across-participants validation projects, our framework could represent a complementary method to standard norms that generalizes across diverse populations.

## 4. Methods

### 4.1. Feature-based clustering

#### 4.1.1. Pre-processing and selection of CCSHS data

We selected full-night polysomnographic recordings from 12 participants (6 female & 6 male) from an open-source dataset from the Cleveland Children's Sleep and Health Study (CCSHS) [58–62]. The CCSHS cohort gathers polysomnographic recordings from children and adolescents. In this study, we focused on the participants whose age ranged from 17.2 to 18.9 years (mean = 17.7, std = 0.47; Table S1), which is the upper range of the age distribution in the database. The dataset is available for download at https://www.sleepdata.org. Each recording contained electroencephalograms (C3/C2 and C4/C1, A1, & A2), bilateral electrooculograms, a bipolar submental electromyogram, sampled at 256 Hz for the EMG and 128 Hz for the EEG. We selected 1 EOG derivation (LOC-A2: electrode placed next to the left canthus and referenced to the opposite mastoid, A2), 1 EMG derivation (EMG1-EMG2, bipolar derivation over the chin) and 1 left central scalp EEG derivation reference to the opposite (right) mastoid (C3-A2). This central EEG derivation is optimal for the detection of both sleep spindles and slow waves. To ensure a consistent sampling rate across the three selected channels, we downsampled the EMG channel from 256 Hz to 128 Hz. We used MATLAB (R2015b) to process CCSHS data. It should be noted that artifacts could influence the clustering of sleep data and additional pre-processing steps aiming at mitigating their influence could be added, prior to *hctsa* feature extraction and epoch clustering.

### 4.2. Feature extraction and selection

We performed feature extraction on MATLAB 2015b using *hctsa* v0.95. This version includes 7749 features that are derived from diverse scientific methodological literatures (such as physics, seismology, economics, neuroscience, and artificially generated simulated data). Among the 7749 *hctsa* features, some returned special values (non-real values or fatal errors). We excluded these features using the *hctsa* function *TS_Normalize* [22]. Using this function, we removed the features that generated special values in more than 20% of the time series of each dataset. On average, we removed 1747 features per dataset (min = 1174, max = 1981).

The range of values generated by the remaining *hctsa* features varied across features. Therefore, we normalized values using the Scaled Robust Sigmoid (SRS) transformation implemented in *TS_Normalize*. SRS is a nonlinear transformation that uses median and interquartile ranges instead of the mean and standard deviation. SRS is robust against the influence of outliers and it scales the transformed data into a range between 0 and 1 (negative infinity maps to 0 and positive infinity maps to 1). The normalization by SRS of each feature value can be defined as:

$$f(x) = \frac{1}{1 + e^{(-x-m)/(iqr/1.35)}}$$

where $m$ is the median value produced by the given feature across time series and $iqr$ is the interquartile range across time series.

### 4.3. Balanced datasets

Based on the AASM labels provided by the trained scorers, we trimmed the 12 selected datasets (Table S1) to exclude any periods

of wakefulness that occurred directly before sleep onset (defined as the first non-wake epoch) or directly after the final non-wake epoch. This allowed us to focus on the intra-sleep periods of wakefulness.

The five sleep stages did not occur in equal proportions throughout the night. Table S1 gives the proportion of each AASM stage, which is similar to the known proportion of sleep states (e.g., N1 comprises only 5% of sleep whereas N2 accounts for approximately 50% [29]). To deal with the imbalance, we subsampled the datasets to equalize the number of epochs for each stage as explained in Fig. 2b. In each iteration of cross validation, we subsampled N epochs from each sleep stage, where N is the number of epochs in the least frequent stage for each dataset. We thus considered 5N epochs in total. We have selected the datasets with $N \geq 37$ (minimum total epochs = 185). We repeated the above subsampling procedure 100 times on all available data. This assured us that the result of our analysis reflects the nature of the entire dataset.

### 4.4. k-means clustering

We used *k*-means clustering (MATLAB function *kmeans* with Replicates = 50, Max Iterations = 500 from Statistics and Machine Learning Toolbox version 11.2). This algorithm groups data based on its Euclidean distance to the center location for each of *k* clusters [63]. We used it due to its simplicity, efficiency and wide presence in the literature [64]. To enable comparison with the 5-way classification of the AASM guidelines, we set *k* to be 5. However, by no means does our framework impose this number.

### 4.5. Sequential maximum matching

When comparing the unsupervised clustering results with the AASM labeling, we assigned each cluster into one of the 5 AASM sleep stages (Fig. 2e), using our "sequential maximum matching" algorithm. This algorithm matches each cluster to one of the five AASM sleep stages based on the most represented AASM label in the cluster. For details, see Fig. S4. When there was a tie, we selected a cluster according to the cluster order randomly generated by the k-means algorithm.

### 4.6. Visual scoring task

During standard visual scoring, human scorers have access to a full night of temporally sequenced data. This allows them to rely on temporal context to score the epoch, especially when it is ambiguous (e.g., is it preceded/followed by N2, N3 or REM?). To compare the epoch-by-epoch discrimination in our approach with traditional scoring, we created a task where the surrounding context was eliminated for human scorers. We recruited 4 scorers ('No-context scorers'), who had substantial experience in sleep scoring according to the AASM guidelines (all from the sleep laboratory at Monash University including one of the authors, TA). They scored isolated 30-s epochs one at a time in a 5-way forced choice manner using our custom-made sleep scoring experimental software (MATLAB, Psychtoolbox). In each trial, the software displayed an isolated 30-s signal consisting of 1 EEG (C3-A2), EOG (LOC-A2) and EMG (EMG1-EMG2). Participants were asked to score these epochs in accordance with the AASM sleep scoring guidelines and to the best of their ability. The number of epochs for each sleep stage was equalized. Two participants scored a subset of randomly selected epochs (n = 37 trials per stage, in total n = 185) from a balanced dataset (ID: 1800001) and the two others scored epochs (n = 41 trials per stage, in total n = 205) from a different balanced dataset (ID: 1800005).

### 4.7. t-SNE

To provide a visualization of N2 time series from the *hctsa* feature space, we used *t*-distributed stochastic neighbor embedding (*t*-SNE), which projects high-dimensional data into a low-dimensional space [65]. We performed a two-dimensional *t*-SNE projection with MATLAB's function *tsne* (default parameters) using as input the EEG, EOG and EMG feature values of the N2 time series.

## Funding

National Health and Medical Research Council (APP1183280) (NT, TA, BDF).

National Health and Medical Research Council (GNT1183280) (BDF).

Australian Research Council (DP180104128 and DP180100396) and Grant-in-Aid for Transformative Research Areas (B) (HDM5Y) from Japan Society for the Promotion of Science (NT).

International Brain Research Organization Post-Doc fellowship and Long-Term Fellowship from the Human Frontier Science Program (LT000362/2018-L) (TA).

Australian Research Council Discovery Early Career Researcher Award (DE170101254) (JMW).

## Data and materials availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

## CRediT authorship contribution statement

**Nicolas Decat:** Software, Formal analysis, Writing – review & editing, Visualization. **Jasmine Walter:** Formal analysis, Writing – original draft. **Zhao H. Koh:** Software, Formal analysis, Writing – original draft, Visualization. **Piengkwan Sribanditmongkol:** Methodology. **Ben D. Fulcher:** Software, Writing – review & editing, Supervision. **Jennifer M. Windt:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Thomas Andrillon:** Formal analysis, Writing – review & editing, Supervision. **Naotsugu Tsuchiya:** Conceptualization, Methodology, Writing – review & editing, Supervision.

## Declaration of competing interest

The authors declare that they have no competing interests.

## Acknowledgements

We thank Tom Drummond for co-supervizing PS for her thesis, Dror Cohen for his pilot analysis using the *hctsa* method on another set of sleep data, Jeroen van Boxtel for discussion on the visual scoring task, Jay Kim for constructing a preliminary version of the visual scoring task, Romain Holzmann and Ursula Voss for providing the data, whose results will be published elsewhere.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sleep.2022.06.013.

## References

[1] Rechtschaffen A, Kales A. A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects. Brain Information

Service; 1968.
[2] Himanen S-L, Hasan J. Limitations of rechtschaffen and Kales. Sleep Med Rev 2000 Apr;4(2):149–67. https://doi.org/10.1053/smrv.1999.0086.
[3] Loomis AL, Harvey EN, Hobart GA. Distribution of disturbance-patterns in the human electroencephalogram, with special reference to sleep. J Neurophysiol 1938 Sep 1;1(5):413–30. https://doi.org/10.1152/jn.1938.1.5.413.
[4] Tanaka H, Hayashi M, Hori T. Statistical features of hypnagogic EEG measured by a new scoring system. Sleep 1996 Nov;19(9):731–8. https://doi.org/10.1093/sleep/19.9.731.
[5] Brodbeck V, Kuhn A, von Wegner F, Morzelewski A, Tagliazucchi E, Borisov S, Michel CM, Laufs H. EEG microstates of wakefulness and NREM sleep. Neuroimage 2012;62(3):2129–39. https://doi.org/10.1016/j.neuroimage.2012.05.060.
[6] Parrino L, Ferri R, Zucconi M, Fanfulla F. Commentary from the Italian Association of Sleep Medicine on the AASM manual for the scoring of sleep and associated events: for debate and discussion. Sleep Med 2009 Aug;10(7):799–808. https://doi.org/10.1016/j.sleep.2009.05.009.
[7] Schulz H. Rethinking sleep analysis. J Clin Sleep Med 2008 Apr 15;4(2):99–103. https://doi.org/10.5664/jcsm.27124.
[8] Feinberg I, Koresko RL, Heller N. EEG sleep patterns as a function of normal and pathological aging in man. J Psychiatr Res 1967 Jun 1;5(2):107–44. https://doi.org/10.1016/0022-3956(67)90027-1.
[9] Feinberg I, March JD, Floyd TC, Fein G, Aminoff MJ. Log amplitude is a linear function of log frequency in NREM sleep EEG of young and elderly normal subjects. Electroencephalogr Clin Neurophysiol 1984 Aug 1;58(2):158–60. https://doi.org/10.1016/0013-4694(84)90029-4.
[10] Rosenberg RS, Van Hout S. The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring. J Clin Sleep Med 2013 Jan 15;9(1):81–7. https://doi.org/10.5664/jcsm.2350.
[11] Iber C, Ancoli-Israel S, Chesson A, Quan SF. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. Westchester: American Academy of Sleep Medicine; 2007 Jan.
[12] Silber MH, Ancoli-Israel S, Bonnet MH, Chokroverty S, Grigg-Damberger MM, Hirshkowitz M, et al. The visual scoring of sleep in adults. J Clin Sleep Med 2007 Mar 15;3(2):121–31. https://doi.org/10.5664/jcsm.26814.
[13] Christensen JAE, Carrillo O, Leary EB, Peppard PE, Young T, Sorensen HBD, et al. Sleep-stage transitions during polysomnographic recordings as diagnostic features of type 1 narcolepsy. Sleep Med 2015 Dec;16(12):1558–66. https://doi.org/10.1016/j.sleep.2015.06.007.
[14] Liu Y, Zhang J, Lam V, Ho CKW, Zhou J, Li SX, et al. Altered sleep stage transitions of REM sleep: a novel and stable biomarker of narcolepsy. J Clin Sleep Med 2015 Aug 15;11(8):885–94. https://doi.org/10.5664/jcsm.4940.
[15] Pizza F, Vandi S, Iloti M, Franceschini C, Liguori R, Mignot E, et al. Nocturnal sleep dynamics identify narcolepsy type 1. Sleep 2015 Aug 1;38(8):1277–84. https://doi.org/10.5665/sleep.4908.
[16] Kishi A, Natelson BH, Togo F, Struzik ZR, Rapoport DM, Yamamoto Y. Sleep-stage dynamics in patients with chronic fatigue syndrome with or without fibromyalgia. Sleep 2011 Nov 1;34(11):1551–60. https://doi.org/10.5665/sleep.1396.
[17] Wei Y, Colombo MA, Ramautar JR, Blanken TF, van der Werf YD, Spiegelhalder K, et al. Sleep stage transition dynamics reveal specific stage 2 vulnerability in insomnia. Sleep 2017 Sep 1;40(9). https://doi.org/10.1093/sleep/zsx117.
[18] Lemyre A, Belzile F, Landry M, Bastien CH, Beaudoin LP. Pre-sleep cognitive activity in adults: a systematic review. Sleep Med Rev 2020 Apr;50:101253. https://doi.org/10.1016/j.smrv.2019.101253.
[19] Stenstrom P, Fox K, Solomonova E, Nielsen T. Mentation during sleep onset theta bursts in a trained participant: a role for NREM stage 1 sleep in memory processing? International Journal of Dream Research 2012;5(1):37–46.
[20] Andrillon T, Nir Y, Staba RJ, Ferrarelli F, Cirelli C, Tononi G, Fried I. Sleep spindles in humans: insights from intracranial EEG and unit recordings. J Neurosci 2011;31(49):17821–34. https://doi.org/10.1523/JNEUROSCI.2604-11.2011.
[21] Fulcher BD, Little MA, Jones NS. Highly comparative time-series analysis: the empirical structure of time series and their methods. J R Soc Interface 2013 Jun 6;10(83):20130048. https://doi.org/10.1098/rsif.2013.0048.
[22] Fulcher BD, Jones NS. hctsa: a computational framework for automated time-series phenotyping using massive feature extraction. Cell Syst 2017 Nov 22;5(5):527–31. https://doi.org/10.1016/j.cels.2017.10.001. e3.
[23] Stevner ABA, Vidaurre D, Cabral J, Rapuano K, Nielsen SFV, Tagliazucchi E, Laufs H, Vuust P, Deco G, Woolrich MW, Van Someren E, Kringelbach ML. Discovery of key whole-brain transitions and dynamics during human wakefulness and non-REM sleep. Nat Commun 2019;10(1):1035. https://doi.org/10.1038/s41467-019-08934-3. Published 2019 Mar 4.
[24] Koch H, Christensen JA, Frandsen R, Zoetmulder M, Arvastson L, Christensen SR, Jennum P, Sorensen HB. Automatic sleep classification using a data-driven topic model reveals latent sleep states. J Neurosci Methods 2014;235:130–7. https://doi.org/10.1016/j.jneumeth.2014.07.002.
[25] Andrillon T, Nir Y, Staba RJ, Ferrarelli F, Cirelli C, Tononi G, et al. Sleep spindles in humans: insights from intracranial EEG and unit recordings. J Neurosci 2011 Dec 7;31(49):17821–34. https://doi.org/10.1523/JNEUROSCI.2604-11.2011.
[26] Purcell SM, Manoach DS, Demanuele C, Cade BE, Mariani S, Cox R, et al. Characterizing sleep spindles in 11,630 individuals from the national sleep research resource. Nat Commun 2017 Jun 26;8:15930. https://doi.org/10.1038/ncomms15930.
[27] Himanen S-L, Virkkala J, Huhtala H, Hasan J. Spindle frequencies in sleep EEG show U-shape within first four NREM sleep episodes. J Sleep Res 2002 Mar;11(1):35–42. https://doi.org/10.1046/j.1365-2869.2002.00273.x.
[28] Brandenberger G, Ehrhart J, Buchheit M. Sleep Stage 2: an electroencephalographic, autonomic, and hormonal duality. Sleep 2006 Jan 1;28:1535–40. https://doi.org/10.1093/sleep/28.12.1535.
[29] Carskadon M, Rechtschaffen A. Monitoring and staging human sleep. In: Perspect Psychol Sci; 2005. p. 1359–77. https://doi.org/10.1016/B0-72-160797-7/50123-3.
[30] Molinari S, Foulkes D. Tonic and phasic events during sleep: psychological correlates and implications. Percept Mot Skills 1969 Oct 1;29(2):343–68. https://doi.org/10.2466/pms.1969.29.2.343.
[31] Simor P, van der Wijk G, Nobili L, Peigneux P. The microstructure of REM sleep: why phasic and tonic? Sleep Med Rev 2020 Aug;52:101305. https://doi.org/10.1016/j.smrv.2020.101305.
[32] Stephansen JB, Olesen AN, Olsen M, Ambati A, Leary EB, Moore HE, et al. Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. Nat Commun 2018 Dec;9(1):5229. https://doi.org/10.1038/s41467-018-07229-3.
[33] Nielsen T, Carr M, Blanchette-Carrière C, Marquis L-P, Dumel G, Solomonova E, et al. NREM sleep spindles are associated with dream recall. Sleep Spindles & Cortical Up States 2017 Mar 1;1(1):27–41. https://doi.org/10.1556/2053.1.2016.003.
[34] Siclari F, Bernardi G, Cataldi J, Tononi G. Dreaming in NREM sleep: a high-density EEG study of slow waves and spindles. J Neurosci 2018 Sep;10. https://doi.org/10.1523/JNEUROSCI.0855-18.2018.
[35] Massimini M, Huber R, Ferrarelli F, Hill S, Tononi G. The sleep slow oscillation as a traveling wave. J Neurosci 2004 Aug 4;24(31):6862–70. https://doi.org/10.1523/JNEUROSCI.1318-04.2004.
[36] Siclari F, Bernardi G, Riedner BA, LaRocque JJ, Benca RM, Tononi G. Two distinct synchronization processes in the transition to sleep: a high-density electroencephalographic study. Sleep 2014 Oct 1;37(10):1621–37. https://doi.org/10.5665/sleep.4070.
[37] Bernardi G, Betta M, Ricciardi E, Pietrini P, Tononi G, Siclari F. Regional delta waves in human Rapid Eye Movement sleep. J Neurosci 2019 Apr 3;39(14):2686–97. https://doi.org/10.1523/JNEUROSCI.2298-18.2019.
[38] Andrillon T, Burns A, Mackay T, Windt J, Tsuchiya N. Predicting lapses of attention with sleep-like slow waves. Nat Commun 2021;12(1):1–12. https://doi.org/10.1038/s41467-021-23890-7.
[39] Tamaki M, Bang JW, Watanabe T, Sasaki Y. Night watch in one brain hemisphere during sleep associated with the first-night effect in humans. Curr Biol 2016 May 9;26(9):1190–4. https://doi.org/10.1016/j.cub.2016.02.063.
[40] Siclari F, Tononi G. Local aspects of sleep and wakefulness. Curr Opin Neurobiol 2017 Jun;44:222–7. https://doi.org/10.1016/j.conb.2017.05.008.
[41] Windt JM. Consciousness in sleep: how findings from sleep and dream research challenge our understanding of sleep, waking, and consciousness. Philos Compass 2020;15(4):e12661. https://doi.org/10.1111/phc3.12661.
[42] Fiorillo L, Puiatti A, Papandrea M, Ratti P-L, Favaro P, Roth C, et al. Automated sleep scoring: a review of the latest approaches. Sleep Med Rev 2019 Dec;48:101204. https://doi.org/10.1016/j.smrv.2019.07.007.
[43] Colombo MA, Napolitani M, Boly M, Gosseries O, Casarotto S, Rosanova M, et al. The spectral exponent of the resting EEG indexes the presence of consciousness during unresponsiveness induced by propofol, xenon, and ketamine. Neuroimage 2019 Apr 1;189:631–44. https://doi.org/10.1016/j.neuroimage.2019.01.024.
[44] Lendner JD, Helfrich RF, Mander BA, Romundstad L, Lin JJ, Walker MP, et al. An electrophysiological marker of arousal level in humansHaegens S, Colgin LL, Piantoni G, editors. Elife 2020 Jul 28;9:e55092. https://doi.org/10.7554/eLife.55092.
[45] Botvinik-Nezer R, Holzmeister F, Camerer CF, Dreber A, Huber J, Johannesson M, et al. Variability in the analysis of a single neuroimaging dataset by many teams. Nature 2020 Jun;582(7810):84–8. https://doi.org/10.1038/s41586-020-2314-9.
[46] Cliff OM, Lizier JT, Tsuchiya N, Fulcher BD. Unifying pairwise interactions in complex dynamics. arXiv preprint arXiv:2201.11941. 2022 Jan 28.
[47] Stepanski EJ. The effect of sleep fragmentation on daytime function. Sleep 2002 May 1;25(3):268–76. https://doi.org/10.1093/sleep/25.3.268.
[48] Schlemmer A, Parlitz U, Luther S, Wessel N, Penzel T. Changes of sleep-stage transitions due to ageing and sleep disorder. Phil Trans Math Phys Eng Sci 2015 Feb;13:373. https://doi.org/10.1098/rsta.2014.0093 (2034):20140093.
[49] Bixler EO, Kales A, Leo LA, Slye T. A comparison of subjective estimates and objective sleep laboratory findings in insomniac patients. Sleep research 1973:2–143.
[50] Harvey AG, Tang NKY. (Mis)perception of sleep in insomnia: a puzzle and a resolution. Psychol Bull 2012 Jan;138(1):77–101. https://doi.org/10.1037/a0025730.
[51] Kaplan KA, Hirshman J, Hernandez B, Stefanick ML, Hoffman AR, Redline S, et al. When a gold standard isn't so golden: lack of prediction of subjective sleep quality from sleep polysomnography. Biol Psychol 2017 Feb;123:37–46. https://doi.org/10.1016/j.biopsycho.2016.11.010.
[52] Andrillon T, Solelhac G, Bouchequet P, Romano F, Le Brun M-P, Brigham M, et al. Revisiting the value of polysomnographic data in insomnia: more than meets the eye. Sleep Med 2020 Feb;66:184–200. https://doi.org/10.1016/j.sleep.2019.12.002.

[53] Lecci S, Cataldi J, Betta M, Bernardi G, Heinzer R, Siclari F. Electroencephalographic changes associated with subjective under- and overestimation of sleep duration. Sleep 2020 Nov 12;43(11):zsaa094. https://doi.org/10.1093/sleep/zsaa094.
[54] Windt JM, Noreika V. How to integrate dreaming into a general theory of consciousness–a critical review of existing positions and suggestions for future research. Conscious Cognit 2011 Dec;20(4):1091–107. https://doi.org/10.1016/j.concog.2010.09.010.
[55] Nielsen TA. A review of mentation in REM and NREM sleep: "covert" REM sleep as a possible reconciliation of two opposing models. Behav Brain Sci 2000 Dec;23(6):851–66. https://doi.org/10.1017/s0140525x0000399x. ; discussion 904-1121.
[56] Noreika V, Valli K, Lahtela H, Revonsuo A. Early-night serial awakenings as a new paradigm for studies on NREM dreaming. Int J Psychophysiol 2009 Oct;74(1):14–8. https://doi.org/10.1016/j.ijpsycho.2009.06.002.
[57] Siclari F, LaRocque JJ, Postle BR, Tononi G. Assessing sleep consciousness within subjects using a serial awakening paradigm. Front Psychol 2013. https://doi.org/10.3389/fpsyg.2013.00542. 0.
[58] Dean DA, Goldberger AL, Mueller R, Kim M, Rueschman M, Mobley D, et al. Scaling up scientific discovery in sleep medicine: the national sleep research resource. Sleep 2016 May 1;39(5):1151–64. https://doi.org/10.5665/sleep.5774.
[59] Hibbs AM, Storfer-Isser A, Rosen C, Ievers-Landis CE, Taveras EM, Redline S. Advanced sleep phase in adolescents born preterm. Behav Sleep Med 2014 Sep 3;12(5):412–24. https://doi.org/10.1080/15402002.2013.825838.
[60] Rosen CL, Larkin EK, Kirchner HL, Emancipator JL, Bivins SF, Surovec SA, et al. Prevalence and risk factors for sleep-disordered breathing in 8- to 11-year-old children: association with race and prematurity. J Pediatr 2003 Apr;142(4):383–9. https://doi.org/10.1067/mpd.2003.28.
[61] Spilsbury JC, Storfer-Isser A, Drotar D, Rosen CL, Kirchner HL, Redline S. Effects of the home environment on school-aged children's sleep. Sleep 2005 Nov;28(11):1419–27. https://doi.org/10.1093/sleep/28.11.1419.
[62] Zhang G-Q, Cui L, Mueller R, Tao S, Kim M, Rueschman M, et al. The national sleep research resource: towards a sleep data commons. J Am Med Inf Assoc 2018 Oct 1;25(10):1351–8. https://doi.org/10.1093/jamia/ocy064.
[63] Jain AK. Data clustering: 50 years beyond K-means. Pattern Recogn Lett 2010 Jun 1;31(8):651–66. https://doi.org/10.1016/j.patrec.2009.09.011.
[64] Huang JZ, Ng MK, Rong H, Li Z. Automated variable weighting in k-means type clustering. IEEE Trans Pattern Anal Mach Intell 2005 May;27(5):657–68. https://doi.org/10.1109/TPAMI.2005.95.
[65] Maaten L van der, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res 2008;9(86):2579–605.