Sleep Medicine Reviews 63 (2022) 101611

Contents lists available at ScienceDirect

Sleep Medicine Reviews

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

TECHNICAL REVIEW

Representations of temporal sleep dynamics: Review and synthesis of the literature

Lieke WA. Hermans [a, b, *], Iris AM. Huijben [a], Hans van Gorp [a], Tim RM. Leufkens [b], Pedro Fonseca [b], Sebastiaan Overeem [a, c], Merel M. van Gilst [a, c]

[a] Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands
[b] Philips Research, Eindhoven, the Netherlands
[c] Sleep Medicine Center Kempenhaeghe, Heeze, the Netherlands

ARTICLE INFO

Article history:
Received 9 November 2021
Received in revised form
25 January 2022
Accepted 7 February 2022
Available online 17 February 2022

Keywords:
Sleep
Time dynamics
Measurements
Polysomnography
Representations
Sleep architecture
Artificial intelligence
Deep learning

SUMMARY

Sleep is characterized by an intricate variation of brain activity over time. Measuring these temporal sleep dynamics is relevant for elucidating healthy and pathological sleep mechanisms. The rapidly increasing possibilities for obtaining and processing sleep registrations have led to an abundance of data, which can be challenging to analyze and interpret. This review provides a structured overview of approaches to represent temporal sleep dynamics, categorized based on the way the source data is compressed. For each category of representations, we describe advantages and disadvantages. Standard human-defined 30-s sleep stages have the advantages of standardization and interpretability. Alternative human-defined representations are less standardized but offer a higher temporal resolution (in case of microstructural events such as sleep spindles), or reflect non-categorical information (for example spectral power analysis). Machine-learned representations offer additional possibilities: automated sleep stages are useful for handling large quantities of data, while alternative sleep stages obtained from clustering data-driven features could aid finding new patterns and new possible clinical interpretations. While newly developed sleep representations may offer relevant insights, they can be difficult to interpret in for example a clinical context. Therefore, there should always be a balance between developing these sophisticated sleep analysis techniques and maintaining clinical explainability.

Introduction

The sleeping brain generates specific patterns of electrophysiological activity. This activity changes according to organized cycles of alternating stadia, with a duration of 90–120 min [1]. Gold standard polysomnography (PSG) is primarily based on the non-invasive measurements of electrical activity from the cortex (i.e., EEG). Traditionally, signals obtained from PSG measurements are divided into time windows with a duration of 30 s, also called epochs. Then, based on standardized rules these epochs are manually assigned to either wake, REM sleep, or one of three categories of NREM sleep (N1-3). From these sleep stages, overall statistics can be calculated, such as total sleep time. These data are used in various ways, including the diagnostic process of a variety of sleep disorders.

The first whole-night sleep report was published in 1937. Since the discovery of dream sleep in 1952, much research has been focused on the basic neurophysiology of sleep [2]. Since the discovery of sleep apnea in the late 1960's, sleep recordings have also become very important for the diagnosis of a wide spectrum of sleep disorders.

Although PSG measurements have been used for a long time, understanding the functions and underlying mechanisms of sleep remains a large challenge [3]. More specifically, the exact mechanisms underlying several sleep disorders, such as insomnia and sleep-related movement disorders, are currently unknown [4,5]. One reason for the incomplete knowledge about sleep may be that the transition from the underlying physiological signals to sleep stages and/or statistics induces categorization and temporal compression, possibly not reflecting all relevant changes in sleep characteristics over the course of the night [6].

To obtain more information about temporal sleep dynamics, researchers have continuously been searching for both new

* Corresponding author. De Zaale, Eindhoven, the Netherlands.
E-mail address: l.w.a.hermans@tue.nl (L.WA. Hermans).

https://doi.org/10.1016/j.smrv.2022.101611

Abbreviations

AASM	American Academy of Sleep Medicine
AHI	apnea-hypopnea index
CPC	Cardio-pulmonary coupling
ECG	Electrocardiography
EEG	Electro-encephalography
EMG	Electromyography
EOG	Electro-oculography
GMM	Gaussian mixture model
HRV	Heart rate variability
ISO	Infraslow oscillations
LDA	Latent Dirichlet Allocation
NREM	Non-rapid eye movement
REM	Rapid eye movement
ORP	Odds ratio product
PCA	Principal component analysis
PPG	Photoplethysmography
PSG	Polysomnography
R&K	Rechtschaffen and Kales
Spo2	Blood oxygen saturation level

technologies to measure sleep and additional ways to analyze existing sleep recordings. The possibilities to study sleep dynamics have greatly increased since the availability of the first computer systems. The current spectacularly improved computational power has allowed for studying polysomnography signals on a higher resolution than 30 s [7,8]. Additionally, the recent increased interest towards implementing machine learning applications in healthcare has yielded new ways to analyze sleep [9,10]. Moreover, alternative measuring techniques using wearable devices are emerging, probably resulting in the availability of many more sleep recordings in the near future [11,12].

We now have increasing possibilities for both representing and measuring sleep, yielding important opportunities to learn more about both healthy and pathological sleep. However, we also face the challenge of translating this abundance of information to a manageable set of clinically relevant characteristics. This review provides an overview of existing approaches that can be used to represent the temporal dynamics of sleep. Within a structured framework, we distinguish different types of representations and describe their advantages and disadvantages. We provide considerations for using these representations in research to better understand sleep mechanisms, with a focus on sleep disordered populations. Finally, technical details regarding the various methods are described as supplementary data.

Methods

A systematic search was performed to obtain a complete overview of existing approaches that can be used to represent the temporal dynamics of sleep. Extensive literature searches were undertaken, initially in June 2020 and updated in December 2021, using the Pubmed, Scopus and Web of Science scientific databases. A broad selection of search terms was used to screen a comprehensive selection of entries (see supplemental section S1.1). Searches were adapted to each database, using appropriate database-specific indexing terms and syntax. Reference lists of selected papers were also screened. We excluded languages other than English and one-page conference abstracts, and we only considered representations that were used to assess sleep of human adults. Because of the extensive nature of the topic, we limited ourselves to including the most important articles containing relevant information about each type of representation, instead of providing an in-depth review of all papers belonging to a certain topic. Therefore, often-used representations, such as sleep spindles and automated sleep staging, do not contain an extensive list of citations.

Overview of sleep representations

Standard unprocessed source signals obtained by PSG include electroencephalography (EEG), electromyography (EMG), and electrooculography (EOG). A representation can be defined as a transformation of these source signals. Sleep analysis originally started with representations selected by people based on clinical knowledge. Two well-known examples include spectral power analysis of the EEG, and the assignment of standardized sleep stages. However, in fact, any transformation of the original data can be called a representation. The desired representation varies depending on the intended objective of the analysis. Often, representations have lost information compared to the original data: lossy data compression has taken place. Compression is often desirable, because sleep recordings typically yield large amounts of data. Note that, besides temporal compression, also spatial compression is induced, since signals from different areas of the brain are compressed into one representation. In this review, we focus on temporal sleep dynamics.

In the previous years, the data science community has made a rapid shift from 'conventional' human-defined representations to so-called data-driven representations learned by a machine. These are also increasingly used in the sleep field [10]. In this review, we describe representations based on their position on a spectrum ranging from human-defined to machine-learned, and the method that has been used to obtain them (Fig. 1). In each section, we introduce a different type of representation and provide a few examples of its applications for clinical research. At the end of each section, strengths and limitations of the representations are discussed. Additional methodological details are discussed in the supplementary data.

We should note that it is also possible to acquire alternative sleep measurements by recording physiological signals that change as a result of sleep, for example actigraphy, electrocardiography (ECG), respiration, and photoplethysmography (PPG) from the wrist. These alternative source signals are often used for automated surrogate sleep staging using wearables. Furthermore, occasionally they are used to obtain other types of representations, for example non-categorical representations based on heart rate variability (section Non-categorical representations). We chose to not classify the representations based on the source signal used, because often, particularly for machine-learned representations, the selected model is independent of the type of input signal. However, it is important to keep in mind that the choice of the input signal does influence the resulting representations.

Standardized sleep stages

Manual sleep staging

Manually scored sleep stages are based on the standardized classification of the signals measured during PSG, specifically wave form characteristics of EEG, EOG and EMG signals. Standard sleep stages have a fixed length of 30 s, dating back to the time that PSG recordings were printed on paper, based on the amount of signal that would fit on one sheet [13]. The first classification manual, which was published by Rechtschaffen and Kales (R&K) in 1968 [14], presents rules for the classification into either wake, REM or

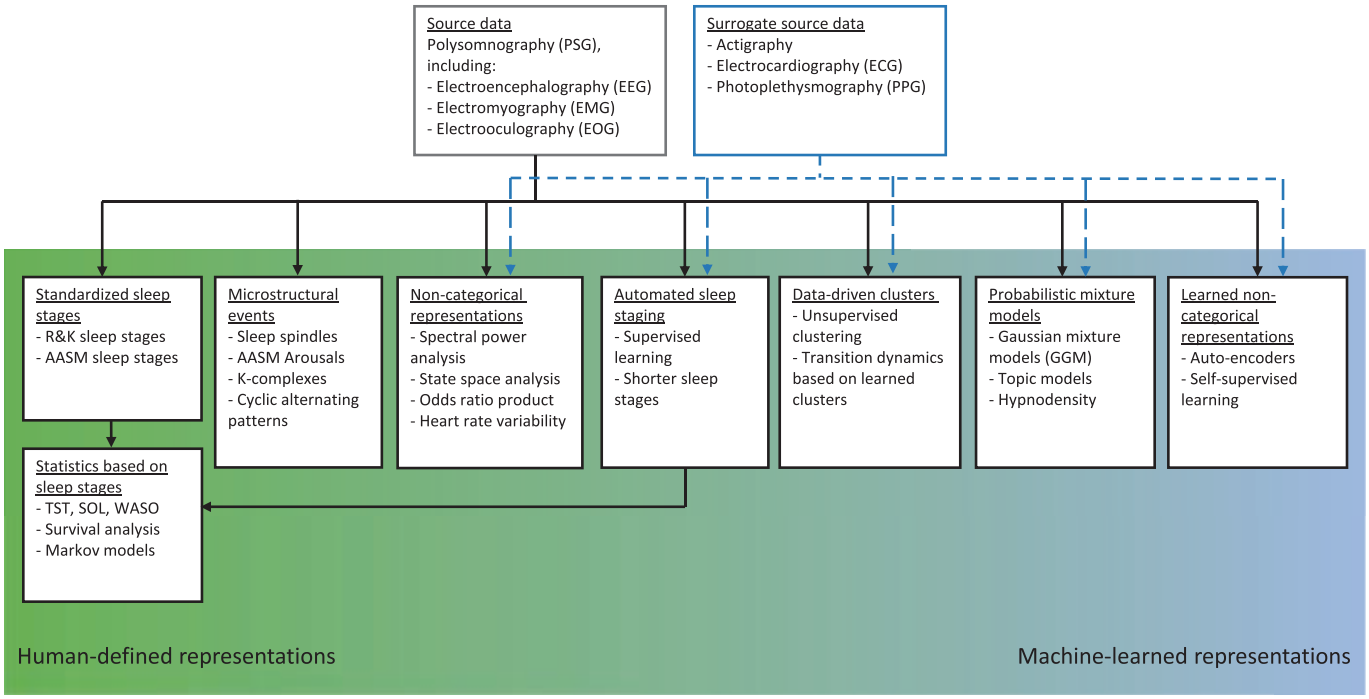

Fig. 1. An overview of approaches to represent temporal sleep dynamics. The blocks contain different types of representations, as well as some examples.

one out of four NREM sleep stages (S1–4), ranging from light to deep. Later, updated versions were published by the American Academy of Sleep Medicine (AASM) [15]. One of the major changes was combining of the two stages of deepest NREM sleep, resulting in three levels of NREM sleep (N1 - 3). Additionally, the stage 'movement time', which was previously used to exclude epochs that could not be categorized, was abolished. Because of their relatively low time resolution, human-defined sleep stages and their inferred statistics are often said to capture 'sleep macrostructure', as opposed to 'microstructure' that is described by representations with time resolutions higher than 30 s.

Strengths and limitations

Sleep stages have been studied for many years, and they were defined based on sleep characteristics that are important in clinical practice, making them easy to interpret. Furthermore, they have the important advantage of standardization, and of compatibility with results achieved from sleep research in the past, as virtually all reference and normative values are based on standard sleep stages. On the other hand, manual sleep scoring is very labor intensive and requires considerable training and experience (for automated sleep scoring see chapter 7). Moreover, the average interrater agreement of human scorers is approximately 82% [16], leading to potentially large differences in interpretation of the sleep stages. Additionally, sleep stages may be influenced by differences between R&K and AASM scoring rules [17]. For example, according to AASM, a transition from N2 to N1 should be scored in the presence of an arousal, while this is not the case according to R&K, leading to difference in the scored percentages of NREM2 [17]. This difference is relevant for older publicly available datasets that were still annotated with R&K labels. Together these limitations of reliability of sleep stages lead to the conclusion that manually scored sleep stages are a far from perfect gold standard [18,19]. Apart from reliability issues, both the temporal aspect (i.e., the need for sleep stages with a length of exactly 30 s) and the categorical aspect (i.e., the need to distinguish exactly five sleep/wake stages) may be questioned.

Statistics based on sleep stages

The representations described in this chapter are derived from classical sleep stages assigned to 30s-epochs. The statistics can be derived irrespective of the way the sleep stage labels were generated. The majority of the statistics are derived from manually scored (section Standardized sleep stages) or automatically generated (section Automated sleep staging) epoch-based stages, and are described here. In section Data-driven clusters, we describe overview statistics based on machine-learned clusters that can be seen as alternative sleep stages.

Whole night overview statistics

Sleep stages are usually depicted in a hypnogram, showing the progression of the sleep stages over time (Fig. 2). Additionally, standard overview statistics can be calculated (Table 1). These statistics, often referred to as 'sleep architecture', are typically assessed in clinical care and have been used in many scientific studies. They are very useful to obtain a quick overview of abnormalities in the whole-night sleep structure. The overview statistics often result from maximal temporal compression; often, only one number is provided to characterize a whole night of sleep.

Sleep transition dynamics

Analysis of sleep state transition dynamics considers the way specific sleep stages or combinations of sleep stages alternate over the night. Here we refer to sleep stages or combinations of sleep stages as 'states'. The traditional overview statistics listed in Table 1 mostly reflect quantities of sleep stages, while it can be also very informative to assess their distribution over the night. For example, a percentage of a certain sleep stage does not tell us whether that particular sleep stage occurs in a few long uninterrupted trains or a large number of short fragments [20].

Models describing sleep state transition dynamics usually involve one out of two key questions: 1) how long does a certain

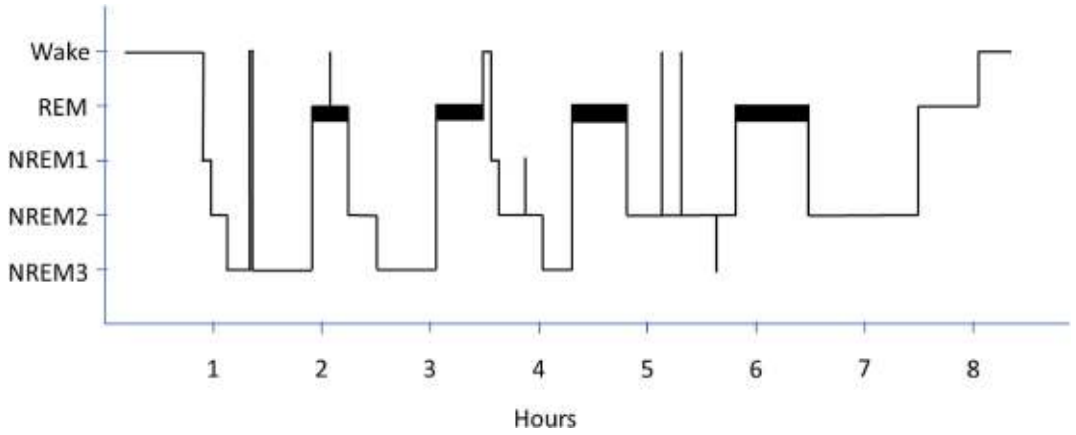

Fig. 2. Idealized example of a classical hypnogram, reflecting the progression of sleep stages over the night, based on 30-s epoch annotation.

Table 1
Often-used sleep description statistics according to AASM criteria.

Sleep statistic	Unit	Description
Total Sleep Time (TST)	Time (minutes)	Summed duration of all REM and NREM epochs in a sleep episode, from lights off to lights on.
Time in Bed (TIB)	Time (minutes)	Total amount of time that a person spends in bed, regardless of whether or not they are sleeping.
Sleep Efficiency (SE)	Percentage	Proportion of time in bed that is actually spent sleeping, calculated as TST/TIB x 100%.
Sleep Onset Latency (SOL)	Time (minutes)	Amount of time between bedtime or lights off time and the start of the first epoch of any sleep stage
Wake After Sleep Onset (WASO)	Time (minutes)	Summed duration of epochs scored as wake occurring after sleep onset.
# awakenings	Number	Total number of awakenings during the sleep recording. An awakening is defined as one or more consecutive wake epochs.
REM sleep latency	Time (minutes)	Amount of time between sleep onset and the first epoch scored as REM sleep
N1; %N1 (also applicable for N2, N3, and REM)	Percentage	Total duration of sleep stages scored as N1; part of total sleep time (or time in bed) that consists of N1, calculated as the summed duration of the epochs scored as N1/TST (or TIB).

sleep state typically last before transiting to another sleep state? and 2) by what sleep state is it followed? For the first question survival analyses may be used, the second question can be answered using Markov models.

Survival analysis

Survival analysis concerns analyzing the expected duration of time until one or more events happen. In sleep research one for example examine the survival dynamics of the combined NREM sleep stages within a sleep recording [21], i.e., the time it takes before trains of NREM sleep epochs (also called bouts) are interrupted by either wake or REM sleep (Fig. 3). This approach yields a list of lifetimes of NREM bouts. A similar strategy could be followed for wake, REM sleep, sleep in general, or specific NREM sleep stages. Often, a distribution is fitted, so that the lifetimes of these bouts are summarized with one or two parameters for each sleep recording. For instance, in case the bout lifetimes can be modelled using an exponential distribution, the average bout duration could be used as a summarizing parameter (see S2.1).

Often, survival dynamics are compared between healthy sleepers and patients [21–24], where decreased lifetimes of certain sleep states can be interpreted as a decreased stability of that state. For example, Norman et al. showed a decreased stability of sleep (i.e., combined NREM and REM sleep) in people with sleep apnea [20]. Our research group showed that people with insomnia had less stable NREM sleep compared to healthy sleepers [24].

Markov models

Besides lifetimes, the transitions between specific states are also of interest. Analysis of sleep stages often involves calculating the number of transitions from one state to another per hour of sleep, for example the number of transitions from N2 to N3 per hour [25,26]. Alternatively, one can calculate probabilities of changing to the different states, given the observation of a current sleep state [27–32]. For example, the probability to progress to N3 could be expressed if the current state is N2. Calculating such probabilities for the AASM sleep stages yields a 5×5 table, called a transition matrix (for analyses see S2.2). A chain of sleep stages over the night is then essentially modelled using a first-order Markov chain: a process that can be fully described by the probability of changing to another state from a given state (Fig. 4). Higher-order Markov chains can also be used to model sleep (see S2.2).

The transition matrix can be used to compare patient populations. For instance, Ferri et al. showed that a lower probability of REM to N2 transitions may be a distinguishing feature of people with narcolepsy type 1 [31]. Additionally, Wei et al. showed that people with insomnia had a larger probability of transiting from N2 to wake or N1 compared to healthy controls, implying lower N2 stability [32].

Models incorporating external factors

It is also possible to utilize more complex models for a more realistic interpretation of sleep. These models often involve combinations of survival and transition dynamics [33–35]. Some of these models additionally consider clock time and time asleep factors [35,36], which could potentially be useful to assess homeostatic and circadian abnormalities in people with sleep disorders.

Yetton et al. constructed a Bayesian network to model the probability of the current sleep stage given the type and duration

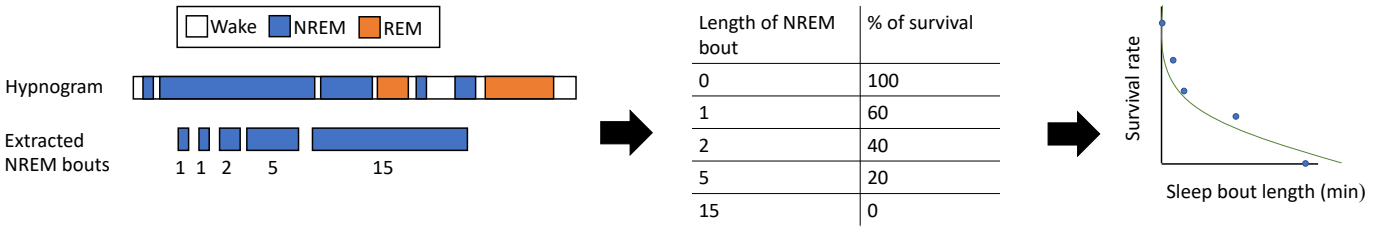

Fig. 3. Survival analysis of NREM sleep (fictional example). The lengths of all uninterrupted trains of consecutive NREM sleep epochs are empirically obtained from the hypnogram, yielding a list of the lifetimes of all NREM bouts within the sleep recording. These lifetimes can be plotted as a function of the percentage of uninterrupted NREM sleep bouts of a certain length L, based on counting the percentage of bouts with a length > L. These data points can then be used for further analysis, for example by fitting a distribution (green line). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

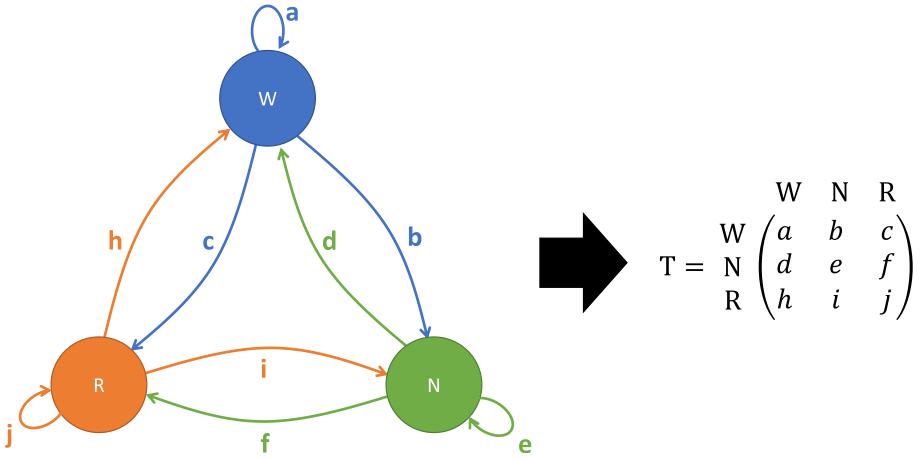

Fig. 4. Examples of a first order Markov model for sleep. In this example, we show a Markov model with three states: wake (W), REM sleep (R), and Non-REM sleep (N). On the right, the transition matrix with 3 × 3 transitions is shown.

of the previous sleep stages; time of day; time asleep; and individual characteristics such as age, sex and BMI [36]. Interestingly, the authors concluded that information about time of day and time slept did not add much predictive power to the model in case the previous two sleep stages were used to predict the next one [36]. The authors speculate that this could be caused by the fact that specific patterns of sleep stages occur at different times of the night [36].

Strengths and limitations

General overview statistics are very useful because a whole night can be summarized by a few parameters. However, important information, such as the degree of sleep fragmentation, is often lost. Modelling sleep transition dynamics yield additional information about sleep fragmentation. Sleep transition dynamics represent a night based on another representation: the sleep stages. Analysis of transition dynamics is potentially very sensitive to subtle differences between the scoring of sleep stages, because short events can have a large effect on the outcome of the analysis. For example, an awakening with a length of one epoch could result in counting two distinct sleep fragments instead of one. This would in turn largely affect the average length of the sleep bouts. In a study into transition dynamics of people with narcolepsy, Zhang et al. indeed reported strong interrater variability regarding the scoring of short awakenings during the night [22]. When researching survival dynamics in datasets annotated by multiple scorers, we would therefore recommend specifically assessing inter-scorer reliability of sleep state transition dynamics.

Microstructural events

Polysomnographically measured sleep is characterized by recurring recognizable features in the EEG, such as k-complexes, spindles, and arousals. Some of these events are seen as characteristics of the different sleep stages, and are therefore part of the AASM scoring rules. Furthermore, they can be used independently to characterize sleep dynamics. Sleep spindles are sudden bursts of oscillatory brain activity, which mostly occur during N2. K-complexes are sharp large-amplitude waveforms that also occur during N2. Sleep spindles and k-complexes both are believed to play a role in memory consolidation, and in protecting the sleep from external stimuli [37,38]. Arousals are sudden increases of the EEG frequency [39], presumably reflecting short sleep disruptions. Because microstructural events occur on a small timescale, they provide information of high temporal resolution. For many microstructural events, information is available about their origin in the brain. Additionally, models are available about the underlying regulation processes that could result in such oscillations [7], but this falls outside the scope of this review. Microstructural events can be used to calculate overview statistics, where they are often summarized as number of events per hour. Similarly to overview statistics of sleep stages (e.g., the number of awakenings per hour; see also section Whole night overview statistics), temporal information is then compressed to one characteristics for a full night of sleep.

Sleep spindles

For research into sleep disorders, sleep spindles are the most studied micro-events.

Decreased numbers of sleep spindles and lower spindle frequencies have been shown in patients with OSA [40,41]. Furthermore, lower spindle activity predicted larger increases in insomnia symptoms in response to stress, and therefore possibly play a role in the pathophysiology of insomnia [42]. However, consistent differences of sleep spindle density have not been found between people with insomnia and healthy sleepers [40]. Potentially, studying other characteristics of sleep spindles (e.g., average length, average frequency, etc.) could increase our understanding of sleep disorders [40]. Medications that affect sleep, such as benzodiazepines and benzodiazepine agonist, also affect spindles [43,44].

Because of the presumed involvement of microstructural events in sleep regulation, their temporal organization is of particular interest. An example of a temporal organization of microstructural events is the cyclic alternating pattern (CAP).

Cyclic alternating patterns

Terzano et al. introduced CAP based on the observation that microstructural events usually occur in bursts [45]. They divided sleep into an A phase and a B phase. Phase A involves phasic events that visually stand out from the background EEG, including delta bursts, K-complex sequences, vertex sharp transients, polyphasic bursts and arousals. Phase B represents tonic non-transient background activity. Phase A can in turn be subdivided into three subtypes (see S3.1). Sleep is classified as CAP if an alternating A and B phase can be observed, and as nonCAP in case only phase B is present. The most extensively used statistic that can be calculated from CAP is CAP rate, which is calculated as the percentage ratio of total CAP time to total NREM sleep time (consisting of CAP + nonCAP). As all constituents of phase A can be seen as sleep disturbances on the microstructural level, CAP rate is seen as a measure of arousal instability, where higher CAP rates are associated with poorer sleep quality [46].

Altered CAP rates have been observed in a wide range of pathological sleep conditions, including sleep apnea, insomnia, and periodic limb movements [46–48]. Moreover, sleep-related events such as sleepwalking, bruxism and epileptic activity seem to have a higher prevalence during phase A, while sleep apneas typically occur during phase B [46].

Tonic and phasic REM sleep

The distinction between phasic and tonic sleep has also been used for eye movements during REM sleep. During phasic REM sleep, frequent eye movements occur [49]. During tonic REM sleep, one can observe distinct low-voltage theta 'background' activity at the EEG electrodes, and a decrease of EMG amplitude [49]. Tonic REM sleep has a lower arousal threshold than phasic REM sleep [49]. This distinction is potentially important for research into abnormal sleep [49]. For example, violent behaviors during REM sleep behavioral disorders more often occur during phasic REM sleep, while epileptic events more often occur during tonic REM sleep [49]. Studying the difference between tonic and phasic REM sleep could help elucidating the underlying mechanisms of these disorders.

Interdependency of oscillations and micro-events

There is emerging evidence that oscillations and micro-events within the sleep EEG can influence each other. The sleep EEG contains large-amplitude infraslow oscillations (ISO) with a frequency of 0.02–0.2 Hz, which are not detectable in standard clinical sleep recordings [50–52]. These ISO were shown to be strongly synchronized with faster oscillations, i.e., the amplitude of the faster oscillations depended on the phase of the ISO [52]. Additionally, the phase of the ISO modulated the occurrence of events such as sleep spindles, k-complexes and interictal epileptic activity [52,53]. It has been speculated that ISO have a sleep-regulating function, and that perturbed slow oscillations may cause ill-timed arousals in people with sleep disorders [51].

Strengths and limitations

Microstructural events usually have a high degree of standardization and reveal information at high temporal resolution. Additionally, the fact that individual events have a presumed function and origin in the brain makes them interpretable. Therefore, microstructural events are often studied in fundamental research into the function of sleep. Additionally, arousals and CAP A phases can be seen as sleep disturbances with a higher resolution compared to awakenings, and can therefore provide a more detailed view on sleep fragmentation. Important disadvantages of microstructural events are the fact that this information can only be obtained from standard PSG measurements and the need for manual scoring, which is laborious and time-consuming. Automated scoring algorithms have been developed for various micro-events, such as sleep spindles and CAP [54–56]. However, these algorithms are not as widely used as automated sleep staging algorithms, due to the difficulty of exactly mimicking manually scored events. For example, there are still important differences between manually and automatically scored sleep spindles [57]. Automated scoring algorithms for phasic and tonic REM sleep have the additional problem that there is no real gold standard, due to the lack of a clear and uniform definition (i.e., definitions of REMs vary across manual scorers [58], and the distinction between phasic and tonic REM epochs have been made in various ways). This lack of gold standard also limits the comparability between manually scored studies.

Non-categorical representations

Non-categorical representations, in contrary to earlier discussed representations, do not categorize sleep epochs into discrete categories like sleep stages, or presence or absence of micro-events. On the contrary, a finer quantized scale is used to represent a transformation of the source signals. A spectral representation of an EEG is a common example. Because non-categorical representations do not rely on the assignment of categories (which induces information loss), they can reflect small differences between sleep recordings that are not noted upon comparing traditional sleep stages. Furthermore, they can provide a more gradual impression of changes throughout the night, compared to the abrupt changes that are inherent to sleep stages.

Non-categorical representations are mostly based on transformations of the EEG signal. Alternatively, metrics based on heart rate variability (HRV) have been employed.

EEG-based representations

Spectral power analysis

Spectral power analysis of the EEG is a widely used strategy [59,60]. The spectral content of an EEG signal can be divided into standard frequency bands [59] (Table 2). Usually, spectral power amplitude calculations are based on 30s-epochs, similar to the AASM sleep stages. The average spectral power during the night or during a specific sleep stage can then be compared between groups of people. Sometimes, changes over consecutive sleep cycles are analyzed. As an alternative for dividing spectral power into frequency bands, it is also possible to directly model and analyze the slope of the

Table 2
Spectral power bands often used in sleep analysis.

Frequency band	Frequency (Hz)	Description
Delta	~ 0.5–4	The slowest of the classically described brain waves, with a high amplitude. Epochs with more than 20% delta waves, with a minimum amplitude of 75 μV, are classified as N3.
Theta	~4–8	Theta activity is mostly found during N1 and N2. They reflect the gradual slowing of EEG waves that can be observed when someone is falling asleep.
Alpha	~8–15	Alpha activity can be seen in someone who is relaxed and awake. They are apparent when the eyes are closed.
Sigma	~12–16	Frequency range in which sleep spindles can be seen. Sleep spindles are mainly found during N2.
Beta	~16–31	Beta activity is associated with normal waking consciousness. They also occur during REM sleep, of which the EEG signal is similar to wake activity.

amplitude/frequency content [61,62]. If modelled correctly, this approach can potentially provide the advantage of using less parameters to describe the same behavior, thus avoiding type I errors [62]. Another widely used approach is analyzing non-linear dynamics of the EEG, often involving calculating either fractal dimensions or entropy from the EEG (see S4.1). However, these types of representations are not often used in research into sleep disorders.

Changes from wake via light sleep to deep NREM sleep are accompanied by a gradual slowing of spectral power. Different frequency bands have distinct physiological interpretations, e.g., lower frequency bands seem to be associated with a lower chance of waking up from external stimuli [63] and higher frequency bands seem to reflect more wake-like brain activity [60]. People with insomnia often present more high-frequency EEG power during NREM sleep compared to healthy sleepers, a finding that is often interpreted as a sign of hyperarousal and alertness [64,65].

Representations of sleep depth: odds ratio product and bispectral index

Spectral power has also been used to express sleep depth [63]. Younes et al. introduced the odds ratio product (ORP), a measure based on 'arousability', i.e., the chance of waking up from a certain sleep stage [66]. This measure was based on a representation of the spectral amplitudes of 3-s time windows within four frequency bands (see S4.2). The amplitudes within the frequency bands were assigned a rank between 0 and 9, and combining these four numbers resulted in numbers between 0000 and 9999, which all were assigned to a different odds ratios. Another index of sleep depth is the bispectral index (BIS), which is an empirically derived parameter based on a combination of EEG parameters from the time and frequency domain [67]. BIS is only available in commercial software.

Younes et al. assessed the ORP during the first 9 s after arousals in people with obstructive sleep apnea [68]. In some patients the ORP normalized faster than in others, which could be an explanation for the fact that people with the same apnea-hypopnea index (AHI) can experience very large variations in the severity of their complaints [68].

State space analysis

Schoch et al. divided polysomnographic recordings into 5-s epochs and plotted each epoch in two-dimensional space based on two fixed frequency ratios calculated from the EEG signal [69]. Colors were assigned to epochs based on manually assigned AASM sleep labels. Epochs assigned to certain sleep stages naturally form clusters, as AASM scoring rules are partially based on EEG spectral power. This visualization was used to 'track' the trajectories between the 5-s epochs over the night, which can be visualized by plotting lines between subsequent epochs. The authors refer to this approach as 'state space analysis'.

It was found that the clusters of datapoints labelled as REM and wake were more merged for people with narcolepsy compared to healthy sleepers, indicating that REM and wake had more similar spectral characteristics. Furthermore, patients with narcolepsy showed an increased number of transitions between datapoints labelled as wake and sleep [69]. Such two-dimensional state space analyses could also be based on other features than spectral power, and potentially provide interesting insights.

Heart rate variability and cardiopulmonary coupling

In the previous paragraphs, we discussed representations based on EEG. However, non-categorical representations can also be based on alternative source signals, for example heart rate. Heart rate variability (HRV) considers the physiological variability of the length of the interval between heart beats over time [70]. There are multiple indices that can reflect heart rate variability (HRV). Most commonly used methods express HRV in either the time domain (for example by calculating the root mean square of the successive differences between heart rates) or the frequency domain (for example using spectral power analysis). It has been speculated that high frequency (0.15–0.4 Hz) HRV is modulated by the parasympathetic system, while low frequency (0.04–0.15 Hz) components of HRV are modulated by both the parasympathetic and the sympathetic system [71]. That way, analyzing different frequency bands could provide additional information about the regulation of sleep by the central nervous system. HRV changes over the consecutive sleep stages, providing the basis for automated sleep staging based on cardiac signals (see also chapter 7) [12]. A related family of metrics is cardiopulmonary coupling (CPC), which is a measure of the coupling between cardiac and respiratory systems [72]. This metric was introduced as a way to evaluate autonomic sleep regulation in people with reduced heart-rate variability [72]. CPC can be calculated based on two time series derived from the ECG: fluctuations of heart rate over time, and a surrogate respiration signal, obtained from respiration-induced fluctuations of the ECG [72]. CPC is then the product of the coherence and cross-spectral power of these two signals [72].

Evaluation of HRV and CPC has been used for research in people with OSA, who consistently show alterations of autonomic nervous activity, although the exact underlying mechanisms are not known [73]. Similar processes possibly also play a role in other sleep disorders, such as insomnia and periodic limb movements [74]. Fluctuations of HRV and CPC seem to co-occur with CAP, indicating that CPC and HRV changes could reflect sleep instability [75,76].

Strengths and limitations

Non-categorical representations can provide high-resolution information. However, calculating overview statistics based on non-categorical representations is a bit more difficult. Therefore, we often see that they are combined with sleep stage information and analyzed based on 30s-epochs. Non-categorical continuous representations have the advantage of not requiring manual scoring. However, most of them have the limitation of being based on exclusively the EEG signal. Analyzing signals obtained from

wearable sleep measurements, for example based on HRV, may provide a more scalable alternative.

Automated sleep staging

Supervised machine learning

Although standardized sleep staging rules were initially intended for manual classification, nowadays a plethora of automated sleep staging algorithms are available. These algorithms are all forms of supervised learning, because they are trained to reproduce ground-truth labels. Throughout the years, many different algorithms have been developed (for reviews see: [18,77]). Recently deep learning techniques have become popular to automatize complex tasks, because they can achieve high performance in the presence of large amounts of data. These algorithms also have been successful for sleep staging. Algorithms that are able to model time relations, for example when incorporating long-short-term memory (LSTM) cells, have been particularly successful, probably because sleep staging for a specific epoch is strongly related to the surrounding epochs. Deep learning models are often trained on untransformed data [18]. However, manual feature selection based on AASM guidelines in combination with deep learning networks have been employed as well to improve interpretability, for example by the commercially available Somnolyzer algorithm. Automated sleep staging algorithms based on standard PSG recordings achieve accuracies of 70–80%, which is a very good performance if one keeps in mind that the inter-scorer agreement of human scorers is approximately 80%.

Furthermore, there is an increased availability of automated sleep staging algorithms based on alternative sleep measurements using wearables (for a review, see: [78]). Older surrogate measurements such as actigraphy allow for distinguishing between sleep and wake, based on the assumption that movement indicates wake and stillness indicates sleep. Although actigraphy in general provides good results, it is important to keep in mind that it has a low specificity for sleep, i.e., non-moving wakefulness can be misinterpreted as sleep [79]. Methods based on autonomic activity, for example HRV, rely on the expression of sleep stages on autonomic nervous system activity. Such algorithms already allow for the distinction of wake and three different sleep stages (REM, N1-2 and N3) [12]. Accuracies of 59–76% have been reported for such algorithms [12].

Sleep staging using reduced epoch lengths

Automatic scoring algorithms can also be used to obtain sleep stage assignments for shorter epochs, e.g., with a length of a few seconds [80,81]. Perslev et al. predicted sleep stages with even shorter epoch lengths up to a frequency of 7680 epochs/minute [82]. The authors built a random forest model to distinguish patients with sleep apnea from healthy controls, based on the occurrence of triplet transitions between sleep stages (i.e., the frequency of combinations of three consecutive unique sleep stages, for example N3, N1, N2), obtained from scoring algorithms with different sleep stage lengths. The performance of the classifier was better for short sleep stages than for longer sleep stages, indicating that the high-resolution sleep stages seemed to provide additional relevant information about pathological sleep characteristics.

Strengths and limitations

Automated sleep staging saves a lot of effort and time compared to manual scoring. Furthermore, because automated sleep staging is deterministic, the same input signals will always result in the same scoring. Automatically scored sleep stages are more flexible compared to manually scored sleep stages, because they can be inferred from other measurements than standard polysomnography, yet they still have the same standardization and interpretation advantages as gold standard sleep stages. However, exactly mimicking manually scored sleep stages is not a straightforward task. Large datasets are required to obtain algorithms with a good performance. Also, it is important to combine datasets from patients and healthy people and from different sleep hospitals and scorers, to avoid overfitting. Existing algorithms were often trained on cohorts with narrow characteristics, for example including only a limited number of sleep disorders. This, together with the black-box nature of these algorithms, may reduce the willingness of clinicians to fully embrace automated sleep staging [9].

Shorter sleep stages can offer an additional increase of resolution. However, the validity of these sleep stages is difficult to evaluate, due to a lack of a gold standard. We will see that this disadvantage is also applicable for other data-driven representations discussed in the next sections.

Data-driven clusters

Data-driven clusters are an alternative for classical rule-based sleep staging, and are typically obtained by unsupervised learning algorithms. Unsupervised learning aims to reveal intricate patterns or clusters in the data in a purely data-driven fashion (as opposed to supervised learning, see section Automated sleep staging, which leverages sleep stage annotations). Resulting clusters could thus be interpreted as alternative sleep stages, and potentially reflect patterns that have not been observed by humans.

Unsupervised clustering

Several strategies have been reported to obtain data-driven clusters from PSG recordings [81,83–85]. These strategies can be summarized as follows. The source signals are first divided into epochs (which are typically 30 s or shorter), after which one or multiple features are extracted from each epoch. Features used for generating data-driven sleep clusters in the literature include – among others - complexity of the EEG signal [83,85], spectral power of the EEG signal [81,84], and cross-correlation between the left and right EOG channels [81], but any type of feature derived from sleep measurements could be used as input for most clustering algorithms. An extensive overview of common features derived from EEG signals is provided by Motamedi-Fakhr et al. [8]. Lastly, some form of clustering is applied on the derived features. Clustering algorithms group data based on (dis) similarities between datapoints. Clustering techniques can in general be subdivided into methods that *a priori* set the number of clusters and assign data points to each of these clusters [84], and strategies that have to decide upon a stopping criterium for a hierarchical method that iteratively merges or splits clusters. This stopping criterium is often defined as the point where a fixed number of clusters is reached [83,85], but it could also be defined using an information-driven criterium [86], such as AIC or BIC.

As a means of validation, found clusters are often compared to standard AASM sleep stages. Frilot et al. found four data-driven sleep clusters [85]. When comparing these clusters to AASM sleep stages, they found that the AASM-defined wake and REM stages were heterogeneous: each corresponded to multiple data-driven clusters [85]. Comparing different types of wake-related clusters also revealed a difference between wake before and after sleep on the one hand, and WASO on the other hand [85].

Clustering based on fMRI data

Data-driven clusters can be derived from any type of signal, including alternative source signals. For example, Stevner et al. used an unsupervised algorithm to identified clusters based on fMRI data [6]. The authors compared the obtained clusters with AASM stages of a simultaneously recorded PSG signal, and analyzed inferred transition dynamics (see also section Strengths and limitations. For an introduction of transition dynamics).

The authors show that N1 did not correspond to any of the data-driven sleep stages [6]. The transition map did, however, indicate the existence of a different transitional state between wake and sleep. Additionally, similar to conclusions by Frilot et al. [85], results indicated that WASO and wake before sleep onset or after last awakening may be two different states with different brain activity and different transition dynamics [6]. This finding may explain sleep inertia, i.e., the sleepiness and impaired cognitive performance immediately after awakening [6]. The presence of different types of brain activity during wakefulness may also provide an explanation for the fact that people with insomnia seem to experience awakenings differently than healthy sleepers.

Strengths and limitations

Alternative sleep stages can be used to recognize complex patterns in the data, that are sometimes difficult to find for human observers. Unsupervised machine learning has the additional advantage that no manually scored labels are required to train the algorithm, which is potentially beneficial in large datasets, and in sleep measurements obtained by wearables.

In general, methods based on machine learning can be based on any type of signal. However, an inherent difficulty with unsupervised machine learning is the lack of a reference truth, making it hard to evaluate the performance of the algorithm and the chosen features. However, there are ways to assess the quality of the distinction between the found clusters, for example quantification of sleep stage class separability in high-dimensional spaces [87]. Another difficulty is that data-driven clusters are not standardized, and different clusters can be found depending on the input data, features, and clustering algorithm. Furthermore, interpretation can be challenging. From the presented examples of the work of Frilot et al. and Stevner et al., it can be seen that comparing data driven clusters to AASM stages aids interpretability.

Probabilistic mixture models

Following standard human-defined sleep scoring rules, exactly one sleep stage is assigned to each 30 s of data. The same holds true for data-driven (discrete) clusters, which were discussed in Section Data-driven clusters. In contrast, probabilistic mixture models represent sleep as a continuous mixture of several clusters over time (see Fig. 5 for an example). Probabilistic mixture models possibly present additional information, because multiple clusters can be assigned simultaneously, and because the probability of a sleep stage is now represented in a non-categorical way. Like the continuous representations described in section Non-categorical representations, this allows for reflecting gradual changes.

We can distinguish models that predict probabilities of data-driven clusters and models that predict probabilities of the AASM stages. In sleep literature, probabilities of data-driven clusters are often modelled using Gaussian mixture models (GMM) and/or topic models.

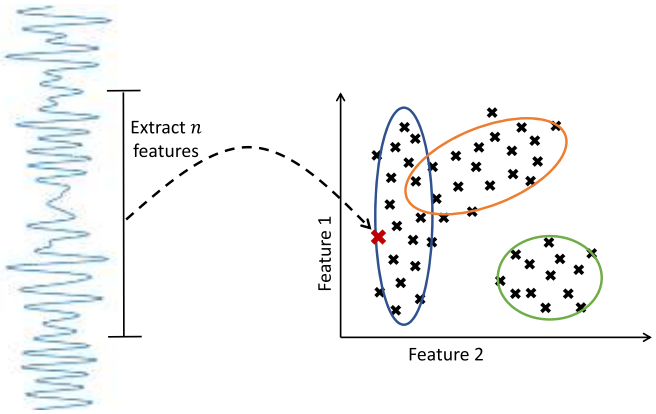

Fig. 5. Gaussian Mixture Model (GMM). Here, we show an illustrative GMM with K = 3 Gaussian distributions (i.e., clusters), trained on two features. Usually, these models are trained based on features that are extracted from the data. For each set of features within one epoch, the GMM yields a probability that this point belongs to each of the K clusters.

Gaussian mixture models

A GMM is a probabilistic model that assumes that the data can be represented as a mixture of K Gaussian distributions, where K – which may be considered the number of clusters - should be chosen in advance (Fig. 5). While any data representation can in theory be used to fit a GMM, often features are used. For each set of features within one epoch, the GMM yields a probability that it belongs to each of the K clusters.

Flexer et al. were the first to model human sleep as a mixture of different probabilities of wakefulness, deep sleep and rem sleep [80,88]. They used a combination between a GMM and a hidden Markov model to reflect time relations between sleep stages [80,88]. Cesari et al. aimed to predict healthy aging from sleep data using a GMM ($K = 3$) [89]. The resulting clusters were interpreted as typical wake, deep sleep, and light sleep, and it was shown that aging is better captured in the statistics of the GMM, than in statistics of traditional automatic sleep scoring classifiers [89]. Mainly the predicted deep sleep probability resulted to be an indicator for aging. Lewandowski et al. trained a GMM ($K = 20$) on the parameters of an auto-regressive model [90]. Interestingly, the authors did not only maximize the GMM fit, but also the fit with R&K labels, and some spindle labels defining the amount of presence of spindle activity. By taking into account these labels, the authors aimed to find micro-states/clusters that correlate well with one of the R&K sleep stages [90]. As a result, multiple micro-states together could encompass exactly one R&K stage, making the definition of R&K stages more fine-grained. Certain combinations of micro-states were found to correlate better with sleep quality metrics than the time spent in R&K stages [90].

Topic models

Topic models originate from the field of natural language processing, where different text documents span different (and possibly multiple) topics. Each word in a language is assigned to each of the topics with a certain probability, thereby creating topics that are a probabilistic representation of the underlying words (see S5.1 for a comparison between GMMs and topic models).

Esbroeck et al. translated the use of topic models, specifically Latent Dirichlet Allocation (LDA), to sleep data [91] (Fig. 6), where quantized frequency band information was used to form 'words'. A resulting *topic mixture diagram* reveals the probability of each of the learned topics/clusters at each epoch. The model was trained for

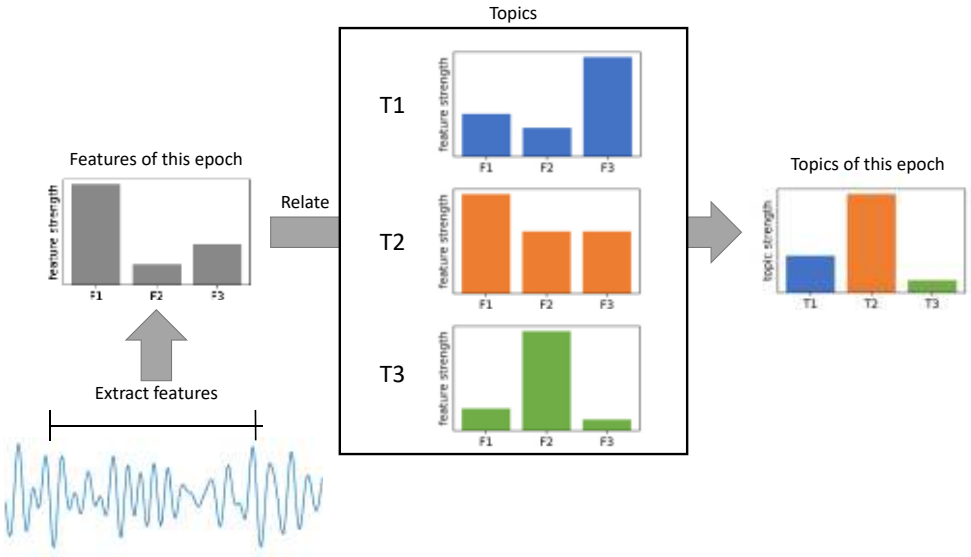

Fig. 6. Topic model. A topic model is based on features that are extracted from the data. In this example, three features are extracted from an epoch (which can be thought of as 'words' in a conventional topic model). We also have three topics. Each topic has a distribution over the features associated with it. By relating the extracted features to each topic, we can find a second distribution: how much of each topic is present in this epoch. The probability over the topics can then be plotted over the night similar to Fig. 7.

each patient separately, resulting in patient-specific topics. Patient-specific sleep stages (a possibility that is not restricted to topic models only) are probably well-suited to model individuals with altered physiology, but they are also difficult to compare across individuals. Models fit on a representative sample from the population, yielding the same sleep stages for each participant, are more suitable for generalization and comparison. Koch et al. trained LDA models on a group of healthy participants [92]. The authors showed that the six obtained latent sleep stages could be used to distinguish patients with REM sleep behavior disorder and Parkinson's disease from controls [92].

Probabilities of AASM sleep stages - hypnodensity

Stephansen et al. predicted probabilistic mixtures of AASM stages, by training a neural network to predict annotated AASM labels [93]. Such a prediction entails a probability for each of the classes. For the task of automated sleep staging, the sleep stage with the highest probability is normally selected and compared to the ground truth label. The authors propose to directly analyze these probabilities of the AASM stages, and referred to the resulting visualization as a *hypnodensity* plot [93]. An example of a hypnodensity plot is shown in Fig. 7.

Based on the sleep stage probabilities, the authors manually generated features reflecting mixing/dissociation of sleep stages, and features expected to predict narcolepsy based on prior work. These features were fed to a classifier, with the goal to develop biomarkers to distinguish narcolepsy from healthy controls. A high sensitivity and specificity of 93 and 91% were found in a test dataset. The most discriminating feature was an early occurrence of a mixture between wake, N2 and REM, reflecting the known sleep stage dissociation in people with narcolepsy [93]. Krauss et al. proposed to use the auto- and cross-correlations of the sleep stage probabilities in the hypnodensity plot as measures for

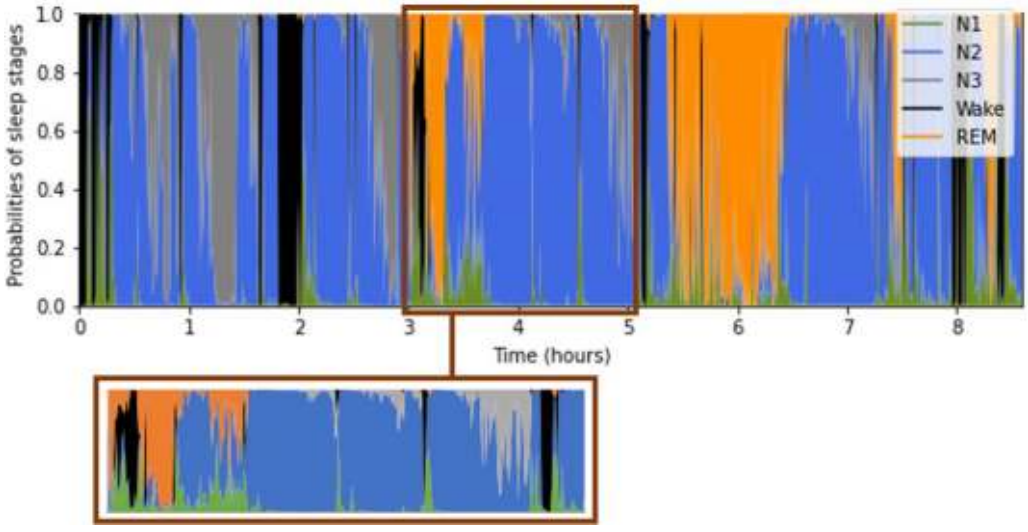

Fig. 7. Example of a hypnodensity plot, showing probabilities of sleep stages over time. Combinations of probabilities of sleep stages can be present at the same time. For example, between 3 and 5 h, multiple instances of combinations of N2 and N3 probabilities are visible (see also magnification below).

identifying pathological disturbances in patients' sleep cycle length [94].

Strengths and limitations

Probabilistic mixture models have similar advantages and disadvantages compared to the learned sleep stages described in section Data-driven clusters, with the possible additional advantage of added information due to the probabilities of the sleep stages. Over time, the focus of machine learning for clustering in general has shifted from simpler models such as GMMs to deep learning models, which are more flexible and do not assume a Gaussian distribution. For sleep-specific applications, the more recently developed hypnodensity approach provides interpretability advantages compared to GMMs and topic models, because hypnodensity algorithm outputs probabilities that can be directly linked to AASM sleep stages, and have been proven clinically useful for the diagnosis of narcolepsy [94]. Regardless of the method that was used to obtain the probabilities, the exact way they probabilities should be interpreted, remains an open question. Stephansen et al. noticed that their probabilities corresponded to uncertainty between different scorers [93]. It is not known if spread probabilities could also reflect mixture, i.e., the presence of different sleep stages at the same time.

Learned non-categorical representations

In the previous sections we discussed supervised machine learning to obtain AASM sleep stages (section Automated sleep staging) and unsupervised machine learning to obtain data-driven clusters, that can be seen as alternative sleep stages (section Data-driven clusters). The goal of both of these methods is to make a prediction about the input data (i.e., a class or a cluster).

However, machine learning is not only restricted to supervised learning and unsupervised data-driven clustering. There is also a growing line of research to representation learning models, that aim to find a lower-dimensional data representation. Such models could be understood as more advanced (i.e., data-driven and non-linear) variants of principle component analysis (PCA). A lower-dimensional representation of such a model is typically the output of a sub-model (or *layer*), which are stacked after each other. Recently, learned representations have gained popularity in many different research fields, including for example medical imaging processing [95].

Learned lower-dimensional representations might also provide clinically relevant insights in the structure of sleep recordings. The extent to which these representations contain useful information depends on the type of model, the training strategy, and the data used to train the model.

We will discuss two types of representation learning models that have been investigated for sleep recordings; unsupervised auto-encoder models (Fig. 8, left), and models that rely on self-supervised learning (Fig. 8, right); a subset of unsupervised learning.

Auto-encoders

Autoencoders aim to reconstruct input data as good as possible, given a bottleneck in the middle of the stack of layers. The bottleneck is necessary to ensure that the model learns a meaningful compressed representation, which is often referred to as the *latent space* or *embedding*. A shallow and linear autoencoder has been shown to reveal the principle components of PCA in its latent space. In practice, however, the first few principle components are often not descriptive enough due to high complexity of the data. Over the years, autoencoders have therefore become more complex (including more layers, non-linearities, and convolutional operations to exploit spatial or temporal relations in the data) in an attempt to find better lower-dimensional representations.

Such low(er)-dimensional data representations have in sleep research typically been used as input to any type of classifier. Feng et al. for example show that sleep apnea can be detected by training a classifier on the latent space features of a specific type of autoencoder [96]. However, interpretability of such features remains a challenge due to the lack of direct relations to known features. Al-Hussaini et al. aim to slightly bridge this gap, by relating learned features to scorer-defined characteristics via a similarity metric, in order to train interpretable classifiers, such as decision trees, for automatic sleep staging [97].

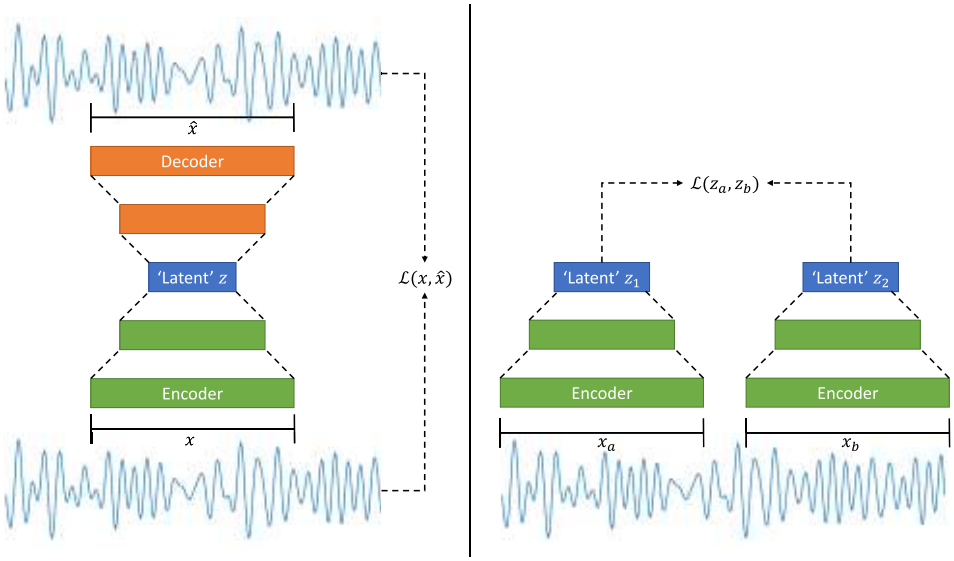

Fig. 8. Left: Auto-encoder: An autoencoder aims to reconstruct the input data as good as possible, given a bottleneck in the middle of the stack of layers. This bottleneck is necessary to ensure that the models learn a meaningful compressed representation. The representation at the bottleneck of the ideally reveals the underlying latent factors of variation that are present in the data. Right: Self-supervised learning: Self-supervised models leverage structures in the data (here: time relations) to find a representation of the data. In this example, the time relation between windows is used to learn this representation.

Self-supervised learning

The training objective of an autoencoder is not guaranteed to result in the most optimal embedding for a required task, which lead to the invention of alternative self-supervised training strategies. Models that are trained in a self-supervised fashion aim to perform a task that relies on structures in the data, for example, temporal relations (Fig. 8, right). Even though self-supervised models are neither guaranteed to find optimal embeddings, thanks to the conceptually different way of training, both models might result in distinct representations, serving different goals.

Banville et al. leverage three different self-supervised tasks and an auto-encoder to learn hidden representations of EEG signals [98]. The authors compared the performance of sleep stage classification using the resulting embeddings, based on the idea that good performance on this task can be used as an indicator that a learned representation still contains useful information [98]. The classification performance on auto-encoder embeddings was much lower than the performance on the self-supervised features, indicating that the self-supervised models preserved more relevant sleep structure-related information. We note, however, that the tested autoencoder had a simple architecture, while more complex architectures have been available. It also remains to be investigated how the choice for a specific self-supervised task influences the learned representations. Banville et al. reduced the dimensionality of the embeddings from two of the three self-supervised models even more by using a common visualization technique, to finally get a 2D representation that is suitable for visual analyses [98]. Transitions between AASM stages were found to be smooth in this 2D space, suggesting no hard class boundaries between these stages. Moreover, sleep apnea patients were found to cluster together in these 2D visualizations, as well as age groups.

Strengths and limitations

Like unsupervised data-driven clusters, representation learning methods have the advantage that no manually labeled data is needed. Even if labels would be available, they are usually not used to obtain learned continuous representations, because it is hypothesized that hidden representations not based on labels can serve as a more general representation of sleep than representations that are (possibly only) optimal for the task of sleep staging.

The field of learned continuous representations is rather young, and still in full development. It remains to be investigated whether recent developments in the aforementioned models may be suitable to learn more representations of sleep recordings that truly reveal new clinical insights. Thanks to their flexibility (i.e., being unsupervised and fully data-driven), they could be of potential use to represent and compress relevant information from multiple-night recordings in large and/or heterogenous datasets. Similarly as for the learned clusters and mixtures sleep clusters described in sections Automated sleep staging and Data-driven clusters, it remains an open research question how to best evaluate the quality of a learned representation. Furthermore, a future challenge would be to relate these representations to interpretable clinical features.

Discussion

Although PSG measurements have been used for a long time to study sleep, understanding the exact mechanisms underlying pathological sleep remains challenging. In spite of reliability issues, epoch-based assignment of sleep stages and derived statistics remain the gold standard for describing sleep, especially in the clinical setting. Although these conventional metrics have the advantages of standardization, interpretability and compatibility with previous research results, they possibly do not reflect all relevant changes of sleep characteristics over the course of the night. This has led researchers to explore a plethora of alternative representations of temporal sleep dynamics. These different representations each have their own advantages and disadvantages, and their validity depends on the goal of the analysis. One of those advantages is that new representations often offer a higher temporal resolution. Furthermore, they may reflect a different number of stages if needed, and/or offer non-categorical information. Additionally, recently developed machine-learned representations bring the benefit of faster, automated sleep staging, and allow for analyzing large datasets obtained from various measurement modalities, for example multi-night recordings obtained using wearables. Lastly, these methods can also be used to find new patterns from the data that are possibly too complex to be identified by human observers.

The rise of new representations for clinical research is accompanied with three major challenges: 1) validation, 2) clinical relevance and 3) interpretability. First, using new representations for clinical research requires validation to make sure that a representation embodies what we want it to represent. The most straightforward way to validate is comparing to a gold standard. Manually scored sleep stages function as a gold standard for automated sleep stages, but still, it is sometimes difficult to evaluate performance in a good way. Standard metrics to evaluate performance, such as accuracy and class-balanced accuracy, reflect the general agreement over a full night of sleep, but do not provide information about specific derived features that may be important (for example sleep fragmentation). Sleep stages are also an imperfect and ambiguous gold standard, due to variability between and within human scorers. Therefore, no 100% accuracy can be obtained. If we are looking for alternative representations that offer additional information that is not yet available, validation inherently becomes more difficult, because of the lack of a gold standard. In these cases, new representations should ideally be selected based on clinical relevance and interpretability.

Evaluating clinical relevance is important because sleep measurements provide a very rich data source, and the number of representations that can be inferred is near endless. Not all of those are clinically relevant. Since researchers are often looking for representations that can help to understand sleep disorders, the relevance of new representations is often demonstrated by showing differences between healthy sleepers and people with a sleep disorder [22,27,47]. Here, artificial intelligence could play an important role in feature selection, for example to identify which parameters distinguish best between healthy and sleep disorders. We should note that judging clinical relevance based on the presence of significant differences between groups has a risk of being circular, i.e., assuming that there should be a difference between groups could lead to the conclusion that the representation is clinically relevant, even if there was no actual clinically relevant difference. This should be taken into account. Another, possibly complementary, strategy could be correlating the representations to the perception of the sleeper, for example to find out which sleep characteristics influence someone's perceived sleep quality [99,100], or to study daytime sleepiness in relation to sleep transition dynamics [101]. This type of research has mainly been performed for overview statistics based on sleep stages, but not for many alternative representations.

The previously described strategies to obtain clinical relevance also aid interpretability. Interpretability could be further improved by combining different types of representations with known characteristics. For example, we often see that data-driven sleep stages are compared to AASM sleep stages to identify which extra information is provided by the machine [6,85,90]. Additionally, interpretability can be improved by providing reference values. Sleep representations typically show large variation with age and gender,

and possibly also between nights. Data from normal sleepers with different characteristics, as well as people with sleep disorders, is required to find reference values. Multiple night recordings using wearable sensors can help to learn more about normal night-to-night variation. However, it is important to keep in mind that the type of source signal may influence the resulting representation.

Understanding night-to-night variability is not only a prerequisite to interpret and understand sleep representations; it also potentially contains clinically relevant information. Further research is needed to find suitable representations reflecting clinically relevant multiple-night information. One other important target for future research is the development of representations reflecting information about the time of the night. Sleep parameters inevitably involve data compression over time. For example, the average NREM sleep bout duration provides information about sleep fragmentation, but it does not answer the question whether this fragmentation changes over the course of the night. Such information could reflect circadian and homeostatic processes, which are possibly altered in people with sleep disorders. One useful approach could be the construction of models incorporating time of the night or time asleep as a parameter [36].

In summary, there are many different ways to represent time dynamics of sleep. These representations can be complementary used to describe different aspects of pathological sleep. Because of the large number of possible representations, future research should focus on finding a balance between developing sophisticated sleep analysis techniques and maintaining clinical interpretability. To find clinically meaningful representations, a strong collaboration between clinical and technological research is needed.

Practice points

- Various methods can be used complementary to represent different aspects of temporal sleep dynamics. The choice of the representation should depend on the goal of the analysis, the size of the dataset, and the sleep disorder under consideration.
- For exploratory research, if comparing obtained sleep characteristics with reference values is desirable, standardly used representations such as AASM sleep stages, overview statistics based on those sleep stages, and spectral power analysis can be useful.
- If sleep fragmentation is of specific interest for the sleep disorder under consideration, sleep stage transition dynamics can be used.
- For small datasets, in case there is a need for information with a high time resolution, and for research related to memory and brain function, often-used representations include microstructural events and non-categorical information such as spectral power, in combination with standard AASM sleep stages.
- For large datasets and sleep measurements obtained with wearables, supervised machine learning can be leveraged to obtain automatically scored AASM sleep stages. Additionally, non-categorical information based on autonomic parameters, for example heart rate variability, can be of use for wearable sleep measurements.
- Alternatively, for large datasets, unsupervised machine learning can be used to find new patterns that may reflect underlying mechanisms of sleep disorders.

Research agenda

- There is a need for new representations that are suitable to describe and quantify dynamic changes in sleep structure over the course of the night
- There is also a need for representations that can reflect variation between nights
- Data from normal sleepers with different characteristics, as well as people with sleep disorders, is required to find references values for the different representations.
- The increased availability of multiple-night measurements based on wearable sleep measurements, generating large amounts of data, requires scalable representations that can be automatically generated.
- For new machine-learned representations, a gold standard is often lacking. Therefore, strategies are necessary for selecting which representations are valid and clinically relevant.
- Future research should facilitate interpretability of new and existing representations by correlating these representations to the perception of the sleeper, for example perceived sleep quality, daytime functioning, tiredness and/or sleep misperception.

Conflicts of interest

This work has been performed in the IMPULS framework of the Eindhoven MedTech Innovation Center (e/MTIC, incorporating Eindhoven University of Technology, Philips Research, Sleep Medicine Center Kempenhaeghe). The funders had no role in the study design, decision to publish, or preparation of the manuscript. Philips Research provided support in the form of a salary for authors LWAH, TRML and PF, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. This activity is in part funded by the PPS program research and innovation of the Dutch ministry of Economic affairs and Climate.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.smrv.2022.101611.

References

[1] Le Bon O. Relationships between REM and NREM in the NREM-REM sleep cycle: a review on competing concepts. Sleep Med 2020;70:6–16.
[2] Dement WC. The study of human sleep: a historical perspective. Thorax 1998 Oct 1;53(suppl 3). S2 LP-S7.
[3] Krueger JM, Frank MG, Wisor JP, Roy S. Sleep function: toward elucidating an enigma. Sleep Med Rev 2016;28:46–54.
[4] Feige B, Baglioni C, Spiegelhalder K, Hirscher V, Nissen C, Riemann D. The microstructure of sleep in primary insomnia: an overview and extension. Int J Psychophysiol 2013;89(2):171–80.
[5] Gabryelska A, Bia P. Association between excessive daytime sleepiness , REM phenotype and severity of obstructive sleep apnea. 2020. p. 1–6.
[6] Stevner ABA, Vidaurre D, Cabral J, Rapuano K, Nielsen SFV, Tagliazucchi E, et al. Discovery of key whole-brain transitions and dynamics during human wakefulness and non-REM sleep. Nat Commun 2019 Mar;10(1):1035.
[7] Olbrich E, Claussen JC, Achermann P. The multiple time scales of sleep dynamics as a challenge for modelling the sleeping brain. Philos Trans R Soc A Math Phys Eng Sci 2011;369:3884–901.
[8] Motamedi-fakhr S, Moshrefi-torbati M, Hill M, Hill CM, White PR. Biomedical Signal Processing and Control Signal processing techniques

applied to human sleep EEG signals — a review. Biomed Signal Process Control 2014;10:21–33.
[9] Goldstein CA, Berry RB, Kent DT, Kristo DA, Seixas AA, Redline S, et al. Artificial intelligence in sleep medicine : background and implications for clinicians. J Clin Sleep Med 2020;16(4):609–18.
*[10] Watson NF, Fernandez CR. Artificial intelligence and sleep: advancing sleep medicine. Sleep Med Rev 2021:101512.
[11] Chinoy ED, Cuellar JA, Huwa KE, Jameson JT, Watson CH, Bessman SC, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. Sleep 2020 Dec 30;44(5):1–16.
[12] Wulterkens BM, Fonseca P, Hermans LWA, Ross M, Cerny A, Anderer P, et al. It is all in the wrist: wearable sleep staging in a clinical population versus reference polysomnography. Nat Sci Sleep 2021 Jun 28;13:885–97.
[13] Schulz H. Rethinking sleep analysis. J Clin Sleep Med 2008;4(2).
[14] Rechtschaffen A, Kales A. A manual of standardized techniques and scoring system for sleep stages of human subjects, vol. 12. Washington: DC US Gov Print Off; 1968. NIH Public.
[15] American Academy of Sleep Medicine. In: Medicine AA of S, editor. The international classification of sleep disorders. 2nd ed. Illinois: Westchester; 2006.
[16] Danker-Hopfe H, Anderer P, Zeithofer J, Boeck M, Dorn H, Gruber G, et al. Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. J Sleep Res 2009 Mar 1;18(1):74–84.
[17] Moser D, Anderer P, Gruber G, Parapatics S, Loretz E, Boeck M, et al. Sleep classification according to AASM and rechtschaffen & Kales: effects on sleep scoring parameters. Sleep 2009 Feb;32(2):139–49.
*[18] Fiorillo L, Puiatti A, Papandrea M, Ratti P, Favaro P, Roth C, et al. Automated sleep scoring : a review of the latest approaches. Sleep Med Rev 2019;48:101204.
[19] Lechat B, Scott H, Naik G, Hansen K, Nguyen DP, Vakulin A, et al. New and emerging approaches to better define sleep disruption and its consequences. Front Neurosci 2021 Oct 7;15:751730.
*[20] Norman R, Scott M, Ayappa I, Walsleben J, Rapoport D. Sleep continuity measured by survival curve analysis. Sleep 2006;29(12):1625–31.
[21] Klerman EB, Wang W, Duffy JF, Dijk DJ, Czeisler CA, Kronauer RE. Survival analysis indicates that age-related decline in sleep continuity occurs exclusively during NREM sleep. Neurobiol Aging 2013;34(1):309–18.
[22] Zhang X, Kantelhardt JW, Dong XS, Krefting D, Li J, Yan H, et al. Nocturnal dynamics of sleep-wake transitions in patients with narcolepsy. Sleep 2017 Feb;40(2).
[23] Bianchi MT, Cash SS, Mietus J, Peng C-K, Thomas R. Obstructive sleep apnea alters sleep stage transition dynamics. PLoS One 2010 Jun;5(6):e11356.
[24] Hermans LWA, Regis M, Fonseca P, Hoondert B, Leufkens TRM, Overeem S, et al. Sleep-wake survival dynamics in people with insomnia. Nat Sci Sleep 2021 Mar 12;13:349–60.
[25] Hansen MH, Kornum B, Jennum PJ. Sleep transitions in hypocretin deficient narcolepsy. Eur J Neurol 2016 Jun;23(2, SI):48.
[26] 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.
[27] 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;34(11):1551–60.
[28] Bianchi MT, Eiseman NA, Cash SS, Mietus J, Peng C-K, Thomas RJ. Probabilistic sleep architecture models in patients with and without sleep apnea. J Sleep Res 2012;21(3).
[29] Bizzotto R, Zamuner S, De Nicolao G, Karlsson MO, Gomeni R. Multinomial logistic estimation of Markov-chain models for modeling sleep architecture in primary insomnia patients. J Pharmacokinet Pharmacodyn 2010 Apr;37(2):137–55.
[30] Kemp B, Kamphuisen HA. Simulation of human hypnograms using a Markov chain model. Sleep 1986;9(3):405–14.
[31] Ferri R, Pizza F, Vandi S, Iloti M, Plazzi G. Decreased sleep stage transition pattern complexity in narcolepsy type 1. Clin Neurophysiol 2016 Aug;127(8):2812–9.
[32] 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;40(9).
[33] Bizzotto XR, Zamuner S. Higher Neural Functions and Behavior Analysis of variability in length of sleep state bouts reveals memory-free sleep subcomponents consistent among primary insomnia patients. 2020. p. 1836–51.
[34] Basner M, Glatz C, Griefahn B, Penzel T, Samel A. Aircraft noise: effects on macro- and microstructure of sleep. Sleep Med 2008 May;9(4):382–7.
[35] Bizzotto R, Zamuner S. Analysis of variability in length of sleep state bouts reveals memory-free sleep subcomponents consistent among primary insomnia patients. J Neurophysiol 2018 May;119(5):1836–51.
[36] Yetton BD, McDevitt EA, Cellini N, Shelton C, Mednick SC. Quantifying sleep architecture dynamics and individual differences using big data and Bayesian networks. PLoS One 2018;13(4):1–27.
[37] Lüthi A. Sleep spindles: where they come from, what they do. Neuroscientist 2014;20(3):243–56.
[38] Halász P. K-complex, a reactive EEG graphoelement of NREM sleep: an old chap in a new garment. Sleep Med Rev 2005 Oct;9(5):391–412.
[39] Bonnet MH, Carley DW, Carskadon MA, Easton PA. EEG arousals : scoring rules and examples. A preliminary report from the sleep disorders atlas task force of the American sleep disorder association. Sleep 1992;15. July 2015.
[40] Weiner OM, Dang-Vu TT. Spindle oscillations in sleep disorders: a systematic review. Neural Plast 2016;2016.
[41] Mohammadi H, Aarabi A, Rezaei M, Khazaie H, Brand S. Sleep spindle characteristics in obstructive sleep apnea syndrome (OSAS). Front Neurol 2021;12:134.
[42] Dang-Vu TT, Salimi A, Boucetta S, Wenzel K, O'Byrne J, Brandewinder M, et al. Sleep spindles predict stress-related increases in sleep disturbances. Front Hum Neurosci 2015;9:68.
[43] Zhang J, Yetton B, Whitehurst LN, Naji M, Mednick SC. The effect of zolpidem on memory consolidation over a night of sleep. Sleep 2020 Nov;43(11).
[44] Plante DT, Goldstein MR, Cook JD, Smith R, Riedner BA, Rumble ME, et al. Effects of oral temazepam on sleep spindles during non-rapid eye movement sleep : a high-density EEG investigation. Eur Neuropsychopharmacol 2015;25(10):1600–10.
[45] Terzano MG, Mancia D, Salati MR, Costani G, Decembrino A, Parrino L. The cyclic alternating pattern as a physiologic component of normal NREM sleep. Sleep 1985;8(2):137–45.
*[46] Parrino L, Ferri R, Bruni O, Terzano MG. Cyclic alternating pattern (CAP): the marker of sleep instability. Sleep Med Rev 2012;16(1):27–45.
[47] Terzano MG, Parrino L, Boselli M, Spaggiari MC, Di Giovanni G. Polysomnographic analysis of arousal responses in obstructive sleep apnea syndrome by means of the cyclic alternating pattern. J Clin Neurophysiol Off Publ Am Electroencephalogr Soc 1996 Mar;13(2):145–55.
[48] Parrino L, Boselli M, Buccino GP, Spaggiari MC, Di Giovanni G, Terzano MG. The cyclic alternating pattern plays a gate-control on periodic limb movements during non-rapid eye movement sleep. J Clin Neurophysiol Off Publ Am Electroencephalogr Soc 1996 Jul;13(4):314–23.
[49] Ermis U, Krakow K, Voss U. Arousal thresholds during human tonic and phasic REM sleep. J Sleep Res 2010 Sep 1;19(3):400–6.
[50] Watson BO. Cognitive and physiologic impacts of the infraslow oscillation. Front Syst Neurosci 2018 Oct 16;12:44.
[51] Lecci S, Fernandez LMJ, Weber FD, Cardis R, Chatton J-Y, Born J, et al. Coordinated infraslow neural and cardiac oscillations mark fragility and offline periods in mammalian sleep. Sci Adv 2017 Feb 8;3(2):e1602026–e1602026.
[52] Vanhatalo S, Palva JM, Holmes MD, Miller JW, Voipio J, Kaila K. Infraslow oscillations modulate excitability and interictal epileptic activity in the human cortex during sleep. Proc Natl Acad Sci USA 2004;101(14):5053–7.
[53] Lázár ZI, Dijk D-J, Lázár AS. Infraslow oscillations in human sleep spindle activity. J Neurosci Methods 2019 Mar 15;316:22–34. 2018/12/17.
[54] Tramonti Fantozzi MP, Faraguna U, Ugon A, Ciuti G, Pinna A. Automatic cyclic alternating pattern (CAP) analysis: local and multi-trace approaches. PLoS One 2021 Dec 2;16(12). e0260984–e0260984.
[55] O'Reilly C, Nielsen T. Automatic sleep spindle detection: benchmarking with fine temporal resolution using open science tools. Front Hum Neurosci 2015 Jun 24;9:353.
[56] Hartmann S, Baumert M. Automatic A-phase detection of cyclic alternating patterns in sleep using dynamic temporal information. IEEE Trans Neural Syst Rehabil Eng 2019;27(9):1695–703.
[57] Lacourse K, Yetton B, Mednick S, Warby SC. Massive online data annotation, crowdsourcing to generate high quality sleep spindle annotations from EEG data. Sci Data 2020;7(1):190.
[58] Agarwal R, Takeuchi T, Laroche S, Gotman J. Detection of rapid-eye movements in sleep studies. IEEE Trans Biomed Eng 2005 Aug;52(8):1390–6.
[59] Thakor NV, Tong S. Advances in quantitative electroencephalogram analysis methods, vol. 6. Annual Review of Biomedical Engineering; 2004. p. 453–95.
[60] Zhao W, Van Someren EJW, Li C, Chen X, Gui W, Tian Y, et al. EEG spectral analysis in insomnia disorder: a systematic review and meta-analysis. Sleep Med Rev 2021 Jan;59:101457.
[61] Lendner JD, Helfrich RF, Mander BA, Romundstad L, Lin JJ, Walker MP, et al. An electrophysiological marker of arousal level in humans. Elife 2020 Jul;9.
[62] Bódizs R, Szalárdy O, Horváth C, Ujma PP, Gombos F, Simor P, et al. A set of composite, non-redundant EEG measures of NREM sleep based on the power law scaling of the Fourier spectrum. Sci Rep 2021 Jan 21;11(1):2041.
[63] McKinney SM, Dang-Vu TT, Buxton OM, Solet JM, Ellenbogen JM. Covert waking brain activity reveals instantaneous sleep depth. PLoS One 2011 Mar;6(3):e17351.
[64] Spiegelhalder K, Regen W, Feige B, Holz J, Piosczyk H, Baglioni C, et al. Increased EEG sigma and beta power during NREM sleep in primary insomnia. Biol Psychol 2012;91(3):329–33.
[65] Perlis ML, Smith MT, Andrews PJ, Orff H, Giles DE. Beta/gamma EEG activity in patients with primary and secondary insomnia and good sleeper controls. Sleep 2001;24(1):110–7. 1.
*[66] Younes M, Ostrowski M, Soiferman M, Younes H, Younes M, Raneri J, et al. Odds ratio product of sleep EEG as a continuous measure of sleep state. Sleep 2015;38(4):641–54.

[67] Sleigh JW, Andrzejowski J, Steyn-Ross A, Steyn-Ross M. The bispectral index: a measure of depth of sleep? Anesth Analg 1999;88(3).
[68] Younes M, Hanly PJ. Immediate postarousal sleep dynamics: an important determinant of sleep stability in obstructive sleep apnea. J Appl Physiol 2016 Apr;120(7):801–8.
*[69] Schoch SF, Werth E, Poryazova R, Scammell TE, Baumann CR, Imbach LL. Dysregulation of sleep behavioral states in narcolepsy. Sleep 2017;40(12).
[70] Stein PK, Pu Y. Heart rate variability , sleep and sleep disorders. Sleep Med Rev 2012;16(1):47–66.
[71] Nano M-M, Fonseca P, Vullings R, Aarts RM. Measures of cardiovascular autonomic activity in insomnia disorder: a systematic review. PLoS One 2017;12(10):e0186716.
[72] Thomas RJ, Mietus JE, Peng C-K, Goldberger AL. An electrocardiogram-based technique to assess cardiopulmonary coupling during sleep. Sleep 2005 Sep;28(9):1151–61.
[73] Kim J Bin, Seo BS, Kim JH. Effect of arousal on sympathetic overactivity in patients with obstructive sleep apnea. Sleep Med 2019 Oct;62:86–91.
[74] Nano M, Fonseca P, Overeem S, Vullings R, Aarts RM. Lying awake at night: cardiac autonomic activity in relation to sleep onset and maintenance. Front Neurosci 2020;2020(13):1405.
[75] Thomas RJ, Wood C, Bianchi MT. Cardiopulmonary coupling spectrogram as an ambulatory clinical biomarker of sleep stability and quality in health, sleep apnea, and insomnia. Sleep 2018 Feb;41(2).
[76] Ferri R, Parrino L, Smerieri A, Terzano MG, Elia M, Musumeci SA, et al. Cyclic alternating pattern and spectral analysis of heart rate variability during normal sleep. J Sleep Res 2000 Mar;9(1):13–8.
[77] Faust O, Razaghi H, Barika R, Ciaccio EJ, Acharya UR. A review of automated sleep stage scoring based on physiological signals for the new millennia. Comput Methods Progr Biomed 2019;176:81–91.
[78] de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep Technology in clinical and research settings. Med Sci Sports Exerc 2019 Jul;51(7):1538–57.
[79] Marino M, Li Y, Rueschman MN, Winkelman JW, Ellenbogen JM, Solet JM, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep 2013 Nov 1;36(11):1747–55.
[80] Flexer A, Dorffner G, Sykacekand P, Rezek I. An automatic, continuous and probabilistic sleep stager based on a hidden markov model. Appl Artif Intell 2002 Mar 1;16(3):199–207.
[81] Koch H, Jennum P, Christensen JAE. Automatic sleep classification using adaptive segmentation reveals an increased number of rapid eye movement sleep transitions. J Sleep Res 2019 Apr;28(2):e12780.
*[82] Perslev M, Darkner S, Kempfner L, Nikolic M, Jennum PJ, Igel C. U-Sleep: resilient high-frequency sleep staging. NPJ Digit Med 2021;4(1):72.
[83] Wielek T, Lechinger J, Wislowska M, Blume C, Ott P, Wegenkittl S, et al. Sleep in patients with disorders of consciousness characterized by means of machine learning. PLoS One 2018;13(1):e0190458.
[84] Tian JY, Liu JQ. Automated sleep staging by a hybrid system comprising neural network and fuzzy rule-based reasoning. Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf 2005;2005:4115–8.
[85] Frilot 2nd C, McCarty DE, Marino AA. An original method for staging sleep based on dynamical analysis of a single EEG signal. J Neurosci Methods 2018 Oct;308:135–41.
[86] Boone DS. Determination of the number of clusters in a data set: a stopping rule clustering algorithm comparison. Int J Strat Decis Sci 2011;2:1–13.
[87] Prabhudesai KS. Topic modeling for inferring brain states from electroencephalography (EEG) signals. 2018.
*[88] Flexer A, Gruber G, Dorffner G. A reliable probabilistic sleep stager based on a single EEG signal. Artif Intell Med 2005 Mar;33(3):199–207.
*[89] Cesari M, Stefani A, Mitterling T, Frauscher B, Sch SV. Sleep modelled as a continuous and dynamic process predicts healthy ageing better than traditional sleep scoring. 2021. p. 77.
[90] Lewandowski A, Rosipal R, Dorffner G. Extracting more information from EEG recordings for a better description of sleep. Comput Methods Progr Biomed 2012 Dec;108(3):961–72.
[91] Van Esbroeck A, Westover B. Data-driven modeling of sleep states from EEG. Conf Proc IEEE Eng Med Biol Soc 2012;2012.
[92] Koch H, Christensen JAE, Frandsen R, Arvastson L, Christensen SR, Sorensen HBD, et al. Classification of iRBD and Parkinson's patients using a general data-driven sleep staging model built on EEG. In: 2013 35th Annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE Engineering in Medicine and Biology Society Conference Proceedings; 2013. p. 4275–8.
*[93] Stephansen JB, Olesen AN. Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. Nat Commun 2018:1–15.
[94] Krauss P, Metzner C, Joshi N, Schulze H, Traxdorf M, Maier A, et al. Analysis and visualization of sleep stages based on deep neural networks. Neurobiol Sleep Circadian Rhythm 2021;10:100064.
[95] Chowdhury A, Rosenthal J, Waring J, Umeton R. Applying self-supervised learning to medicine: review of the state of the art and medical implementations. Informatics 2021;8.
[96] Feng K, Qin H, Wu S, Pan W, Liu G. A sleep apnea detection method based on unsupervised feature learning and single-lead electrocardiogram. IEEE Trans Instrum Meas 2021;70:1–12.
[97] Al-Hussaini I, Xiao C, Westover MB, Sun J. {SLEEPER:} interpretable sleep staging via prototypes from expert rules. Proc 4th Mach Learn Healthc Conf 2019;(10):721–39. abs/1910.vol. 0.
[98] Banville H, Chehab O, Hyvärinen A, Engemann D, Jul ML, Paris-saclay U. Uncovering the structure of clinical EEG signals with self-supervised learning. J Neural Eng 2021;18(4).
[99] Rosipal R, Lewandowski A, Dorffner G. In search of objective components for sleep quality indexing in normal sleep. Biol Psychol 2013;94(1):210–20.
[100] 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;2017(123):37–46.
[101] Waechter M, Kantelhardt JW, Bonsignore MR, Bouloukaki I, Escourrou P, Fietze I, et al. Unique sleep-stage transitions determined by obstructive sleep apnea severity, age and gender. J Sleep Res 2020 Apr;29(2).