Quantitative modelling of sleep dynamics

BY P. A. ROBINSON1,2,3,*, A. J. K. PHILLIPS1,4, B. D. FULCHER1,5, M. PUCKERIDGE1 AND J. A. ROBERTS1,2

1School of Physics, University of Sydney, Sydney, NSW 2006, Australia
2Brain Dynamics Center, Sydney Medical School—Western, University of Sydney, Westmead, NSW 2145, Australia
3Center for Integrated Research and Understanding of Sleep, 431 Glebe Point Road, Glebe, NSW 2037, Australia
4Division of Sleep Medicine, Brigham and Women’s Hospital, Harvard Medical School, 221 Longwood Avenue, Suite 438, Boston, MA 02115, USA
5Department of Physics, Clarendon Laboratory, Oxford University, Parks Road, Oxford OX1 3PU, UK

Arousal is largely controlled by the ascending arousal system of the hypothalamus and brainstem, which projects to the corticothalamic system responsible for electroencephalographic (EEG) signatures of sleep. Quantitative physiologically based modelling of brainstem dynamics theory is described here, using realistic parameters, and links to EEG are outlined. Verification against a wide range of experimental data is described, including arousal dynamics under normal conditions, sleep deprivation, stimuli, stimulants and jetlag, plus key features of wake and sleep EEGs.

Keywords: modelling; sleep–wake cycle; ascending arousal system; chronotype; electroencephalography

1. Introduction

Wake and sleep states of the brain are primarily due to the ascending arousal system (AAS) of the dorsal hypothalamus and brainstem, driven by homeostatic influences, circadian rhythms entrained to daily light cycles via the suprachiasmatic nucleus (SCN) and cortical feedbacks [1,2]. Sleep–wake dynamics also require the corticothalamic system and its network interactions to be considered, since this system generates the electroencephalographic (EEG) rhythms often used to characterize arousal states [3–9]. Figure 1 shows key brain structures, connections and feedbacks involved in arousal dynamics. The present paper concentrates on the AAS and SCN, and gives a brief outline of some EEG correlates.

*Author for correspondence (p.robinson@physics.usyd.edu.au).

One contribution of 11 to a Theme Issue ‘The complexity of sleep’.
Essential features of a realistic neurodynamic model are that it: (i) be based on physiology and anatomy, including salient features at many spatial and temporal scales, (ii) make quantitative predictions of measurable quantities, (iii) have parameters that relate directly to physiology and anatomy, and can be estimated independently, (iv) allow parameters to be deduced by fitting model predictions to data, and (v) be applicable to multiple phenomena and data types, rather than being a theory of a single phenomenon or type of experiment.

Neural field theories (NFTs) provide a natural basis for modelling and analysing multi-scale neural systems, consistent with these requirements. In NFTs, averages are taken over microscopic neural structure to obtain mean-field descriptions on scales from tenths of a millimetre up to the whole brain, incorporating representations of the anatomy and physiology of separate excitatory and inhibitory neural populations, nonlinear neural responses, multi-scale interconnections, synaptic, dendritic, cell-body and axonal dynamics, and feedbacks between structures [3,4,10–18]. Full NFTs are needed to model EEGs, but the simpler limit of neural mass theory (NMT) suffices to reproduce many aspects of arousal dynamics, as discussed in §2; in NMT, spatial dynamics are not resolved and each neural population is treated as a mass [7,10,19].

In §2, we outline our model’s physiological foundations, basic predictions and comparisons with experiment. Section 3 presents a brief summary.

2. Neural mass model of arousal

(a) Ascending arousal system model

Arousal states are mainly governed by the wake-promoting nuclei of the AAS of the brainstem and hypothalamus, which project diffusely to the corticothalamic system [20]. The most important nuclei to model in the AAS are well established from physiological investigations, and are shown in the green block in figure 1.
These include the wake-promoting monoaminergic (MA) group and the sleep-promoting ventrolateral preoptic nucleus (VLPO), which inhibit one another, resulting in flip-flop dynamics if they interact strongly—only one is active at a time, and it suppresses the other [2] to form the sleep-wake switch. In wake the MA is dominant, and in sleep the VLPO is dominant. State transitions are driven by inputs to the sleep-wake switch, including the circadian drive $C$ from the SCN and the homeostatic sleep drive $H$ from buildup of metabolic by-products (probably adenosine, Ad, but possibly others) in wake and their clearance in sleep [21]. Cholinergic (ACh) and orexinergic (Orx, not shown in figure 1) inputs to the MA group are also present [20,22].

Most models of human sleep have been either non-mathematical (e.g. based on sleep diaries) or abstract (mathematical, but not derived directly from physiology). The widely known two-process model is of the latter form, and includes circadian and homeostatic influences [23]. The relatively recent advances in sleep neurophysiology have enabled the development of physiologically based models [24–26], and here we use NMT to model the dynamics of the AAS nuclei, following Phillips & Robinson [19], who argued that (i) since the system spends little time in transitions, the generation rate of $H$ has just two values, one for wake and one for sleep, (ii) the clearance rate of $H$ is proportional to $H$ with a characteristic time scale $\chi$, and (iii) the production rate of $H$ is $\mu Q_m$, where $\mu$ is a constant and $Q_m$ serves as a proxy for arousal state. These steps yield equations for $H$ and the mean soma voltages $V_a$ in the MA group ($a = m$) and VLPO ($a = v$):

$$\tau \frac{dV_v}{dt} + V_v = \nu_{vm} Q_m + D, \quad (2.1)$$
$$\tau \frac{dV_m}{dt} + V_m = \nu_{mv} Q_v + A, \quad (2.2)$$
$$\chi \frac{dH}{dt} = -H + \mu Q_m \quad (2.3)$$
and

$$D = \nu_{vc} C + \nu_{vh} H, \quad (2.4)$$

where the time constants $\tau$ of the responses are assumed equal, and the $\nu_{ab}$ are coupling strengths to neurons $a$ from neurons in population $b$. Firing rates $Q_j$ are defined by a sigmoidal function of the $V_j$, with

$$Q_j = S(V_j) = \frac{Q_{max}}{1 + \exp[(\theta - V_j)/\sigma^3]}, \quad (2.5)$$

where $Q_{max}$ is the maximum possible rate, $\theta$ is the mean firing threshold relative to resting and $\sigma^3 \pi/\sqrt{3}$ is its standard deviation [14].

The total excitatory drive $D$ to the VLPO comprises $C$ and $H$, where $C$ can be interpreted as the SCN firing rate and $H$ is a firing rate change due to somnogenic effects; in both cases only the products $\nu_{vc} C$ and $\nu_{vh} H$ influence $D$. When simple entrainment can be assumed, $C \approx c_0 + \cos(\Omega t)$, where $c_0$ is a constant offset, $\Omega = 2\pi/(24\text{ h})$, and the amplitude of $C$ is absorbed into $\nu_{vc}$. In other cases, such as jetlag and modelling chronotype, $C$ is modelled using the oscillator model of Forger et al. [27], which includes (i) a light processing component (process L), and (ii) a van der Pol oscillator component (process P). In process L, retinal
photoreceptors are converted from ready to activated state by photons, at a rate \( a \), dependent on light intensity \( I \). Activated photoreceptors are converted back to ready at a constant rate \( b \). The fraction \( n \) of activated photoreceptors follows
\[
\frac{dn}{dt} = \lambda[a(1 - n) - \beta n],
\]
where \( a = a_0(I/I_0)^p \) [28]. The resultant photic drive \( B \) is proportional to the rate \( a(1 - n) \) at which photoreceptors are activated, with
\[
B = aG(1 - bx)(1 - by)(1 - n),
\]
where \( b \) and \( G \) are constants and equation (2.7) embodies circadian phase dependence of light sensitivity [28]. The circadian oscillator (process P) follows
\[
\frac{dU}{dt} = g(x - \frac{4x^3}{3}) - y\left[(\frac{\tau_0}{\tau_c})^2 + kB\right]
\]
and
\[
\frac{d\theta}{dt} = x + B,
\]
where \( y \) is the pacemaker activity, \( x \) is a complementary variable, \( k \) sets the photic drive strength, \( \tau_c \) is intrinsic period, \( \tau_0 \) is constant and \( \gamma \) determines the oscillator stiffness. As input to the AAS, we use \( C = y + c_0 \). The physiological parameters in the above model components have the nominal values in table 1, determined by physiological constraints from the literature and comparison with a restricted set of experiments [19,29,30]. The theory then predicts the results of a wide range of experiments.

(b) Normal sleep dynamics, deprivation and recovery

The first key result from the model is that the steady states of equations (2.1)–(2.4) display a ‘fold’ when plotted against the sleep drive \( D \). The upper and lower branches represent wake and sleep, respectively, with an unstable branch in between. Cycles of \( D \) cause the system to move around the hysteresis loop shown in figure 2, with saddle-node bifurcations from wake to sleep and back again. Near-stable ghost states are located just beyond these bifurcations [32].
The existence of different thresholds for sleep-to-wake and wake-to-sleep transitions is the result of mutual inhibition between MA and VLPO, and is a characteristic shared with the two-process model. The foundation of the two-process model is thus clarified by physiologically based modelling. An important distinction from the two-process model is that the present model’s state passes both thresholds during normal cycling rather than staying strictly between the thresholds, as shown in figure 3. Furthermore, because our model’s states are defined in terms of firing rates and voltages, state transitions are not instantaneous, but have their own time scale determined by $\tau$ [29].

The present model has been applied to study the effects of total sleep deprivation and recovery by fixing $V_m$ and $V_v$ to their wake values during deprivation, then releasing them. The results reproduce observed time courses of recovery sleep and recovery of sleep latency to baseline, as in figure 4, and predicts that initiating sleep near normal bedtime minimizes required recovery sleep [29].

(c) Stimuli

External stimuli to the MA or VLPO are modelled as perturbations to the drives $A$ and $D$ [31]. Stimuli excite the MA [34] via an additional term $\mathcal{A}$ on the right of equation (2.2), causing an excursion from equilibrium. To simulate sleep fragmentation by auditory stimuli [35], $\mathcal{A}$ is applied to perturb the model from the sleep branch to higher $V_m$. For small $\mathcal{A}$, the system returns quickly to the sleep branch, but if $\mathcal{A}$ is large enough, the return is via the wake ghost, where the system lingers in a brief awakening [31]. Hence, we define the arousal threshold to be the magnitude of the drive $|\mathcal{A}|$ required to perturb the system to the wake ghost and cause awakening. A linear fit to a clinical auditory decibel scale then allows us to predict the arousal threshold versus time since sleep onset.
Figure 3. Comparison of (a) the Phillips–Robinson (PR) model to (b) the two-process model. Thresholds for transitions are labelled $T_1$ and $T_2$. Homeostatic drive (process S) increases during wake until $T_2$ is reached and the PR model moves to the sleep state; similarly, transitions to wake occur at $T_1$. Equivalently, the total sleep drive $D$ for the PR model is shown in (c), with the two-process model in analogous form in (d). Thresholds $D_1$ and $D_2$ for state transitions are constant. Adapted from Phillips & Robinson [29].

Figure 4. Sleep latencies after deprivation for (a) model and (b) experiment (mean ± s.e.m.) of Lamond et al. [33]. Triangles are for 9h time in bed (TIB) after 63h deprivation (long deprivation). Other curves are after 39h deprivation, with 9h TIB (squares), 6h TIB (circles), plus 7h TIB (dotted), for which there are no experimental data. In (a), sleep latencies are measured at the end of each day, for baseline (B) and recovery days N1, N2, …. In (b), latencies are measured for B, deprivation (SD) and recovery days R1, R2, …. Adapted from Phillips & Robinson [29].
Figure 5. Model predictions of arousal threshold. (a) The model arousal threshold $A$ (mV) (full circles) agrees well with experimental auditory arousal thresholds $I_e$ (dB) (open circles) measured over a normal night of sleep [36]. (b) The model also predicts the arousal threshold $I_e$ (dB) (full circles) and body temperature $I_b$ ($^\circ$C) (open circles) in a sleep fragmentation study [37]. Data are circles, model predictions are curves, and the 9 dB offset in $I_e$ between frames is due to different experimental conditions and subjects. Adapted from Fulcher et al. [31].

[31], as seen in figure 5. This procedure can be used for other stimuli, including pharmacological agents, that can be represented in terms of drives to the MA and VLPO (cf. §2e).

(d) Wake effort and fatigue

Ordinarily, as $D$ increases past its wake–sleep bifurcation value ($\approx 2.5$ mV), the system enters sleep. By applying an external drive, it is possible to keep the system in the wake ghost to maintain wakefulness. We term this additional drive ‘wake effort’ $W$ and add it to the MA drive $A$ on the right of equation (2.2). This drive tends to keep the MA active and thus opposes the effects of $D$ on the system. The effort required to remain awake is zero on the wake branch, and increases with $D$ in the wake ghost, consistent with common experience. Simulating sleep deprivation in this way produces a $W$ time series that can be compared with experimental measures. We have confirmed our hypothesized correlation between $W$ and subjective fatigue levels with multiple comparisons to data [31], as illustrated in figure 6. Performance levels are also expected to correlate with $W$, although the relationship may be nonlinear and task-dependent, e.g. due to motivational input from the limbic system that decreases that required from the cortex and improves performance [40,41].

(e) Caffeine

Caffeine is a competitive antagonist of adenosine (Ad), since it competes for Ad receptor sites in the brain, partially masking its effects. It also reduces inhibition of basal forebrain ACh by Ad, increasing the firing rate of ACh nuclei [42], and the value of $A$ in equation (2.2). These effects are modelled by the replacements

$$\nu_{vh} \rightarrow \nu_{vh}[1 - \zeta_H Z_C(t)] \quad \text{and} \quad A \rightarrow A + \zeta_A Z_C(t), \quad (2.10)$$

Phil. Trans. R. Soc. A (2011)
Figure 6. Model fits to clinical subjective fatigue data. Data are shown with circles, and model fits with a dashed line. (a) Subjective ‘effort’ data from a 7-day sleep deprivation study of Pasnau et al. [38]. Simulated wake effort $W$ over the same period is rescaled by a constant factor to reproduce the observations. (b) Subjective fatigue from Fröberg et al. [39] linearly scaled to the model’s $W$ drive. Adapted from Fulcher et al. [40].

Figure 7. Examples of model dynamics with and without a caffeine dose of 200mg at 22.00 (true clock time resets to zero at 24.00 on the time axis). The caffeine case is shown solid, while a baseline case (no caffeine) is shown dashed. In the second frame, the effective homeostatic drive felt by the subject who takes caffeine is shown grey. Adapted from Puckeridge et al. [43].

where $Z_C$ is the concentration of caffeine in milligrams per kilogram, and $\zeta_A$ and $\zeta_H$ are constants determined from comparison with experiment [43]. Puckeridge et al. [43] modelled caffeine pharmacokinetics by assuming that it is absorbed and eliminated at rates proportional to dose and concentration, respectively, giving

$$Z_C(t) = \gamma_C[e^{-k_e(t-t_0)} - e^{-k_a(t-t_0)}],$$

for $t \geq t_0$, where $k_a \approx 10^{-3}$ s$^{-1}$ and $k_e \approx 4.5 \times 10^{-5}$ s$^{-1}$ are rates of absorption and elimination, respectively, and $\gamma_C$ and $t_0$ are the size and time of dose.

Figure 7 shows model output for a subject with habitual bedtime of 23.00 who takes 200 mg of caffeine at 22.00. Caffeine delays sleep onset, shortens its duration and raises $A_d$ concentration because (i) $D$ decreases, shifting the system away from sleep, and (ii) the increase in $A$ stabilizes wake at larger $D$. There is a smaller delay in waking because $C$ grows in early morning.

The model successfully matches clinical data of Penetar et al. [44], who followed subjective sleepiness over 60 h total sleep deprivation. They compared the effectiveness of 600 mg of caffeine against a placebo given on the 49th hour
Figure 8. Model output versus clinical fatigue data during sleep deprivation. Subjects took caffeine or placebo at 54h (49 h total deprivation). Experimental data are shown as squares, triangles and circles, for $S$ before either dose, after caffeine and after placebo, respectively. Simulated values are shown solid and dashed using equation (2.12) for subjects with low caffeine sensitivity who took or did not take caffeine, respectively. Adapted from Puckeridge et al. [43].

of the study. Sleepiness was assessed via the Stanford sleepiness scale [45]. From figure 8, the score $S$ is found to be linearly related to $D$, with

$$S = c_1 D + c_2,$$

where $c_1 \approx 0.23 \text{ mV}^{-1}$ and $c_2 \approx 2.33$ [43]. Comparison with experiment in figure 8 shows that both exhibit a steady increase in $S$ with superposed circadian oscillations. They concur that 600mg of caffeine reduces $S$ by approximately 2.5 and that its effects are lost after about 15h ($\approx 3$ decay times).

(f) Orexin and narcolepsy

The excitatory effect of orexin on MA is modelled above as a static contribution to $A$ on the right of equation (2.2). Reducing $A$, as in narcoleptics, is found to reduce the width of the hysteresis loop, shortening the distance between the two state transition thresholds. In the presence of noisy input, this makes the system more susceptible to transitions from wake to sleep and back, as in narcolepsy, consistent with the hypothesis that orexin stabilizes the sleep–wake switch [2]. We simulate neural and environmental noise by adding Gaussian white noise to the drives in equations (2.1) and (2.2) with the correct amplitudes to yield 1.0 mV root-mean-square responses if $v_{nw} = v_{vm} = 0$. Decreasing $A$ in the presence of noise is found numerically to result in more time spent in transitions between wake and sleep. The model also predicts lower $Q_m$ in wake, which may contribute to the excessive daytime sleepiness of narcoleptics.

(g) Chronotype and jetlag

An individual’s prefered sleep time defines their chronotype. Morning types prefer to sleep and wake earlier in the day, while evening types prefer to sleep and wake later, and these are believed to be inherent traits [46,47]. Significant
Figure 9. Entrained midsleep time versus entrained circadian phase (both relative to normal, where positive values indicate advance). Model predictions of sleep phase versus intrinsic circadian phase are shown for variations of circadian period $\tau_c$ (solid line) and amplitude $\nu_c$ (dotted line) and of homeostatic time constant $\chi$ (dashed line).

age-related changes in chronotype have been identified, including eveningness in adolescence [48], and morningness in old age [49]. Chronotype differences may be caused by mutations in circadian clock genes [50], or variations in intrinsic circadian period [51] or sleep homeostasis [52]. Physiologically based models can help identify the mechanisms involved [53] by examining the parameter dependences of sleep times and angle of circadian entrainment relative to the daily light cycle. In figure 9, we show the effects on sleep phase of varying circadian period, homeostatic time constant and circadian amplitude under sinusoidal light input $I = 5000[1 + \sin(\omega t)]/2$ lux.

The model reproduces entrainment phenomena, including recovery from jetlag as seen in figure 10. Consistent with experiment, the model favours phase delay over phase advance, as demonstrated by the direction of re-entrainment after a 12 h shift. Gating of light by arousal state is found to affect recovery time significantly, so it is important to assess individual light exposure.

(h) Electroencephalographic aspects

EEG signatures of sleep are largely due to activity in the corticothalamic system in figure 1. NFT has been successfully used to predict the resulting spectra [4,14]. For example, figure 11a shows excellent agreement of predicted and observed EEG spectra over several decades. The features reproduced include alpha and beta peaks and asymptotic low- and high-frequency behaviours; key differences between wake and sleep spectra can also be reproduced, including the strong increase in low-$f$ activity in sleep, where our corticothalamic NFT predicts steepening of the spectrum from $1/f$ towards $1/f^3$, as seen in figure 11b [5,6,15]. Each feature is related to anatomy and physiology: $1/f$ or $1/f^3$ behaviours are signatures of marginally stable, near-critical dynamics, which allow complex behaviour [4,14], while the high-$f$ fall-off results from low-pass
filtering by synaptic and dendritic dynamics. Corticothalamic and intrathalamic loop resonances account for the alpha, beta and spindle peaks, and for the correlated changes in spectral peaks between sleep and wake [4–6,54].

It has been shown that the state and physical stability of the corticothalamic system can be approximately represented in a three-dimensional space with axes $x$, $y$, $z$ that parametrize dimensionless cortical, corticothalamic and intrathalamic gains, respectively [4]. (These gains are modulated by the AAS, via connections in figure 1 that remain to be modelled in detail.) Parameters corresponding to stable arousal states lie in a stability zone (SZ) seen in figure 12. Normal states lie in the SZ, as shown for alert eyes open (EO) to deep sleep, including relaxed eyes closed.
Figure 12. Brain stability and time series. (a) Stability zone. The surface is shaded according to instability frequency, as labelled (blue, spindle; green, alpha; red, theta), with the transparent front right-hand face corresponding to a slow-wave instability. Approximate locations are shown of alert eyes open (EO), relaxed eyes closed (EC), sleep stage 2 (S2) and stage 4 (S4) states, with each state located at the top of its bar, whose $x$, $y$ coordinates can be read from the grid. (b) Simulated time series corresponding to EO, EC, S2 and S4, approximating sensory inputs as white noise. (a,b) Adapted from Robinson et al. [4,6], respectively. (Online version in colour.)

(EC) and sleep stages 1–4 (S1–S4), as shown in figure 12a, corresponding to the typical time series in figure 12b [4]. Zones outside the SZ appear to correspond to epileptic seizures [4,55].

3. Summary and discussion

Physiologically based theories of brain dynamics can incorporate physiology and anatomy across the many scales necessary to reproduce a wide range of neural phenomena. They achieve this for physiologically realistic parameters, and yield numerous predictions that accord with observations in linear and nonlinear regimes in a way that unifies disparate subfields and permits parameter determination via fits of model predictions to experiment.

Neural mass and neural field models are well suited to exploring the mechanisms that control arousal and related EEG phenomena. The models reviewed here provide a framework for additional applications to phenomena, such as chronic sleep deprivation, and sleep in other species [56]. They also lay the foundation for further generalization and integration of additional physiology, in particular, inclusion and calibration of more realistic circadian inputs via the SCN to treat shiftwork, jetlag and chronotypes, and inclusion of Orx nuclei. Incorporation of the diffusely projecting output from ACh nuclei to cortex and thalamus will also be necessary to enable key corticothalamic parameters that determine EEGs to be set by the AAS, and other interactions (figure 1) will also be needed in future modelling.
The Australian Research Council, The National Health and Medical Research Council, The Westmead Millennium Institute and the National Space Biomedical Research Institute through NASA NCC 9-58 supported this work.

References


P. A. Robinson et al.


