

**Universidade Federal de Minas Gerais**

Departamento de Fisiologia e Biofísica

Programa de Pós-Graduação em Fisiologia e Farmacologia

Desenvolvimento e validação de um *Toolbox* de código aberto para classificação de sono de roedores

Renan Augusto Viana Mendes

Belo Horizonte - MG

2021

RENAN AUGUSTO VIANA MENDES

Desenvolvimento e validação de um *Toolbox* de código aberto para classificação de sono de roedores

Dissertação apresentada ao Programa de Pós-Graduação em Fisiologia e Farmacologia da Universidade Federal da Universidade Federal de Minas Gerais, como parte das exigências para a obtenção do título de Mestre em Ciências.

Área de Concentração: Neurociências

Orientador: Prof. Dr. Cleiton Lopes Aguiar

Coorientador: Prof. Dr. Eduardo Mazoni

Andrade Marçal Mendes

Belo Horizonte, Minas Gerais

2021

043 Mendes, Renan Augusto Viana.  
Desenvolvimento e validação de um Toolbox de código aberto para  
classificação de sono de roedores [manuscrito] / Renan Augusto Viana Mendes.  
- 2021.  
135 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Cleiton Lopes Aguiar. Coorientador: Prof. Dr. Eduardo  
Mazoni Andrade Marçal Mendes.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de  
Ciências Biológicas. Programa de Pós-Graduação em Fisiologia e Farmacologia.

1. Fisiologia e Farmacologia. 2. Sono/classificação. 3. Ritmo Circadiano. 4.  
Hipocampo. 5. Eletrofisiologia. 6. Eletromiografia. I. Aguiar, Cleiton Lopes. II.  
Mendes, Eduardo Mazoni Andrade Marçal. III. Universidade Federal de Minas  
Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 612:615

|

Ficha catalográfica elaborada pela bibliotecária Fabiane C M Reis - CRB 6 - 2680



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS BIOLÓGICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM FISIOLOGIA E FARMACOLOGIA

FOLHA DE APROVAÇÃO DA DEFESA DA DISSERTAÇÃO N° 538 DE RENAN AUGUSTO VIANA MENDES

"Desenvolvimento e Validação de um Toolbox de Código Aberto Para Classificação de Sono de Roedores"

**RENAN AUGUSTO VIANA MENDES**

Dissertação de Mestrado defendida e aprovada, no dia **23 de abril de 2021**, pela Banca Examinadora constituída pelos seguintes professores:

**Prof. Dr. Danilo Barbosa Melges**, ENGENHARIA/Universidade Federal de Minas Gerais

**Prof. Dr. Rodrigo Neves Romcy Pereira**, ICE/Universidade Federal do Rio Grande do Norte

**Prof. Dr. Eduardo Mazoni Andrade Marçal Mendes**, ENGENHARIA/Universidade Federal de Minas Gerais  
– Coorientador

**Prof. Dr. Cleiton Lopes Aguiar**, ICB/Universidade Federal de Minas Gerais - Orientador

Belo Horizonte, 23 de abril de 2021.

Assinatura dos membros da banca examinadora:



Documento assinado eletronicamente por **Cleiton Lopes Aguiar**, Professor do Magistério Superior, em 26/04/2021, às 16:25, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Eduardo Mazoni Andrade Marçal Mendes**, Professor do Magistério Superior, em 27/04/2021, às 09:35, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Danilo Barbosa Melges**, Professor do Magistério Superior, em 27/04/2021, às 14:49, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rodrigo Neves Romcy Pereira**, Usuário Externo, em 02/06/2021, às 16:01, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0685384** e o código CRC **F7045265**.



## **Agradecimentos**

Gostaria de agradecer primeiramente a minha família que me deu tudo que eu precisei para conseguir chegar aqui. Gostaria de dedicar esse trabalho ao esforço de gerações da minha família, que desde a sua origem tem feito de tudo para que as novas gerações pudessem gozar de oportunidades que antes ficavam apenas no pensamento. É um privilégio fazer parte dessa família e todos os dias eu entrego o meu máximo para honrar tudo o que já foi feito.

Agradeço também ao meu orientador, Prof. Dr. Cleiton Lopes Aguiar que me deu oportunidades ímpares para minha carreira, das quais tive aprendizados que levarei para a vida toda. Sou muito grato por todos os seus ensinamentos que variaram de instruções técnicas sobre metodologias científicas e sobre programação até a conceituação do que é elegância no fazer científico. Por fim, agradeço pela paciência e pela confiança em mim. Agradeço também o meu coorientador Prof. Dr. Eduardo Mazoni, que desde a minha iniciação científica tem fornecido importante auxílio nas discussões dos resultados e na captação dos financiamentos para pesquisa.

Gostaria de deixar meu profundo agradecimento também a dois colegas de trabalho, amigos e referências, Ikaro Jesus Silva Beraldo e Francisco de Assis Carvalho Carmo. Os dois, praticamente desde o início da minha carreira científica, me ensinaram e me ajudaram muito em tudo, além de atuarem de forma direta ou indiretamente neste trabalho. Se não fossem os dois, não chegaria nem perto de onde estou.

Agradeço também ao Dr. Flávio Mourão por ter me acolhido nos meus primeiros dias no laboratório, por ter tido extrema paciência para me ajudar a entender um pouco sobre ciência. Se não fosse sua ajuda, não sei se teria conseguido começar minha iniciação científica. Agradeço ao Dr. Daniel Medeiros, que me ajudou imensuravelmente, em especial na minha adaptação à Suécia e na minha participação no projeto em colaboração com o Instituto Karolinska. Além desses, agradeço ao Dr. Leonardo Guarnieri pelas excelentes discussões e rodadas de café que contribuíram largamente para meu entendimento sobre a ciência, mas também sobre a vida, e

agradeço também ao M.e Leonardo Rakauskas Zacharias pelas discussões nas análises e pela disponibilização do registro feito em rato para este trabalho.

I would like to thank Dr. Gilberto Fisone for accepting me as a Research Assistant at his lab at the Karolinska Institute, in Sweden, and for giving me this exceptional opportunity to work in one of the best research institutes in the world. A special thanks to Carina Plewnia who dedicated part of her time providing me essential technical and theoretical information for my work and scientific development.

Por fim, agradeço também à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), à Swedish Foundation for International Cooperation in Research and Higher Education (STINT) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo financiamento necessário para o desenvolvimento deste projeto.

## Lista de abreviaturas e siglas

6-OHDA - 6-hidroxidopamina-HCl

AUC - Área sob a curva

CNN - Convolutional neural network

EEG - Eletroencefalografia

EMG - Eletromiografia

FN - *False negative*

FP - *False positive*

GMM - Modelo de Misturas de Gaussianas

LFP - Potencial de campo local

MFB - Feixe prosencefálico basal (via nigroestriatal)

MLE - Estimação de Verossimilhança Máxima (*maximum-likelihood estimation*)

NREM - *non rapid eye movement*

NSQ - Núcleo supraquiasmático

REM - *rapid eye movement*

RMS – *Root mean squared*

ROC - Característica de Operação do Receptor

SWA - Atividade em ondas lentas (*Slow Wave Activity*)

TC - Tálamo-corticais

TN - *True negative*

TP - *True Positive*

## Sumário

1. Resumo .....	10
1.1 - Em português .....	10
1.2 - Em inglês .....	11
2. Revisão da Literatura e Objetivos .....	12
2.1 - Ritmos circadianos e sono .....	12
2.2 - Memória, hipocampo e sono.....	21
2.3 - Classificação do ciclo sono-vigília .....	23
2.4 Objetivos gerais .....	27
2.5 Objetivos específicos .....	27
3. Métodos .....	27
3.1 - Sujeitos experimentais .....	30
3.2 - Procedimentos cirúrgicos .....	30
3.3 - Aquisição do dado .....	31
3.4 - Pré-processamento .....	32
3.5 - Extração de componentes .....	34
3.6 - Descrição do algoritmo: Modelo de Mistura de Gaussianas .....	35
3.7 - Classificação Manual .....	37
3.8 - Cálculo da curva ROC e das AUC .....	37
4. Resultados .....	41
4.1 - Valores do RMS do EMG apresenta informações importantes para diferenciar principalmente sono da vigília .....	42
4.2 - Valores da razão Teta/Delta do LFP apresenta informações importantes para diferenciar principalmente sono NREM do sono REM .....	44
4.3 - As distribuições de Teta/Delta e RMS do EMG flutuam ao longo do tempo coerentemente com as etapas do ciclo circadiano .....	47
4.4 - O Modelo de Mistura de Gaussianas (GMM) cria grupos com épocas de características similares .....	49
4.5 - A definição da classificação manual do usuário como um padrão ouro da classificação .....	51

4.6 - O limiar de probabilidade posterior mais adequado é definido como o de melhor acurácia para cada estado .....	54
4.7 - Todas as épocas classificadas em um mesmo estado apresentam características coerentes entre si e exibem as características típicas de seus respectivos estados.....	59
4.8 - Os estados flutuam ao longo do tempo coerentemente com o ciclo circadiano.....	62
4.9 - As características finais dos estados são consistentes nos diferentes registros testados .....	64
4.10 - A <i>Toolbox</i> demonstrou elevada acurácia na classificação de registros de animais modelo da doença de Parkinson .....	67
4.11 - Um registro feito em rato também apresentou elevada acurácia na classificação.....	71
4.12 - As análises feitas pela <i>Toolbox</i> atuam como controle sobre a qualidade do dado coletado .....	73
5. Discussão .....	78
5.1 - A relação Teta/Delta do sinal do hipocampo e o RMS do EMG como parâmetros principais da classificação .....	79
5.2 - O nosso método de classificação personaliza os limites de cada estado em cada registro, mas sempre respeitando suas características esperadas .....	81
5.3 - Classificação manual como parte ativa no processo de classificação .....	82
5.4 - A nossa <i>Toolbox</i> possui uma série de estratégias que favorecem a implementação e reprodutibilidade das classificações .....	83
5.5 - Aplicações da <i>Toolbox</i> para diversos modelos murinos e da doença de Parkinson.....	84
5.6 - Próximos passos .....	85
6. Referências .....	86
7. Apêndice 1 - Interface Gráfica .....	97
8. Apêndice 2 - Instruções para utilização do classificador de estágios do ciclo sono-vigília .....	107

## 1. Resumo

### 1.1 - Em português

Um corpo crescente de evidências tem demonstrado que o sono exerce diversas funções primordiais, tais como homeostase e consolidação de memórias. Muitos trabalhos atualmente tentam desvendar os mecanismos desses processos, mas ainda existem perguntas básicas a serem respondidas. De fato, descobertas recentes sobre a relação entre diversas doenças e alterações de sono reforçam a importância desse campo de estudo. Porém, ao revisarmos a literatura, fica evidente que a maioria das estratégias de classificação de sono apresentam diversos problemas, a saber: (1) as estratégias utilizadas não estão bem descritas, (2) o acesso e implementação das ferramentas disponíveis são difíceis, (3) os classificadores desenvolvidos não atingem um acurácia satisfatório, ou (4) as classificações manuais, quando utilizadas, assumem papel secundário na classificação, atuando como única estratégia de classificação ou apenas para contabilizar o acurácia. Como consequência direta, os estudos sobre a arquitetura do sono ainda enfrentam problemas de reprodutibilidade. Neste trabalho, nós desenvolvemos uma *Toolbox* de classificação do ciclo sono-vigília i) que utiliza a classificação manual dos usuários para atingir um alto acurácia nas classificações, ii) que apresenta interface gráfica de fácil implementação e livre acesso, iii) que retorna ao usuário figuras controle das etapas, e iv) que pode ser utilizada em murinos, tanto saudáveis e como submetidos a modelos de doenças. Portanto, essa *Toolbox* pode ser uma ferramenta importante para o avanço das pesquisas sobre a fisiologia e sobre os possíveis impactos de diferentes doenças no ciclo circadiano.

Palavras-chave: sono, classificação, ciclo circadiano, hipocampo, eletrofisiologia, eletromiografia, potencial de campo local

## 1.2 - Em inglês

A growing body of evidences has shown that sleep has several key functions, such as homeostasis and memory consolidation. Many works today try to unravel the mechanisms of these processes, but there are still basic questions to be answered. In fact, recent discoveries about the relationship between different diseases and sleep disorders reinforce the importance of this field of study. However, when reviewing the literature, it is evident that most sleep classification strategies present several problems, namely: (1) the strategies used are not well described, (2) access and implementation of the available tools are difficult, (3) the classifiers developed do not achieve satisfactory accuracy, or (4) manual classifications, when used, assume a secondary role in the classification, acting as the classification strategy itself or just to calculate the accuracy. As a direct consequence, studies on sleep architecture still face reproducibility problems. In this work, we developed a *Toolbox* for classification of the sleep-wake cycle that i) uses the manual classification to achieve a high accuracy in the classifications, ii) presents an easy-to-implement and open-source graphical interface, iii) returns control figures of each step to the user, and iv) can be used in rodents in both healthy and disease models. Therefore, this *Toolbox* can be viewed as an important tool for the advance of research on physiology and on the possible impacts of different diseases on the circadian cycle.

Keywords: sleep, classification, circadian cycle, hippocampus, electrophysiology, electromyography, local field potential

## **2. Revisão da Literatura e Objetivos**

Diversos estudos sobre a fisiologia do ciclo circadiano em humanos alertam para o grande impacto do estilo de vida moderno sobre a qualidade do sono (Ohayon 2002). A maior exposição à luz (artificial) e a maior exposição a situações de estresse impactam grande parte da população mundial, e que, como consequência direta disso, enfrentam prejuízos no sono (Gaston, Visser, and Hölker 2015). Estudos demográficos indicam que pelo menos 30% da população mundial adulta sofrem com algum tipo de problema no sono (Brownlow, Miller, and Gehrman 2020). Estudos estimam que os custos associados a distúrbios no sono chegam a aproximadamente 100 bilhões de dólares (Wickwire, Shaya, and Scharf 2016). A atenção crescente para os impactos na arquitetura do ciclo circadiano causados por doenças também eleva a necessidade de investir em estudos no campo (K. Pavlova and Latreille 2019). É fundamental que modelos animais sejam utilizados para que os mecanismos subjacentes aos diversos prejuízos, incluindo os de saúde mental, decorrentes destas alterações de sono sejam melhor compreendidos. Como será demonstrado adiante, a classificação de diferentes estágios de sono não é algo trivial. As ferramentas disponíveis para classificação automática de estágio de sono apresentam problemas em relação: (1) à sua facilidade de execução por usuários não especializados em linguagens de programação, (2) ao fato de não possuir seu código aberto e, portanto, não oferecer possibilidades de flexibilizar o programa para demandas mais específicas, (3) ao seu custo computacional muito alto e (4) ao seu baixo acurácia. Tais problemas, a maioria deles relacionados à falta de transparência, são possíveis fatores que contribuem para a crise de reprodutibilidade que muitas áreas de pesquisa enfrentam atualmente, incluindo a neurofisiologia experimental.

### **2.1 - Ritmos circadianos e sono**

Seres vivos com diferentes níveis de complexidade, variando de bactérias a humanos, exibem flutuações em sua atividade metabólica e comportamento ao longo das 24 horas do dia (Anafi, Kayser, and Raizen 2019). Tais variações constituem os ritmos circadianos, ou seja, ritmos cujo período possuem duração de cerca de um dia.



Cada etapa dessa flutuação apresenta suas características típicas, que podem ser utilizadas como seus marcadores fisiológicos, e que, curiosamente, são altamente conservadas ao longo da filogenia dos seres vivos (Anafi, Kayser, and Raizen 2019). Por exemplo, a maioria das espécies apresentam períodos de inatividade, no qual seu metabolismo e sua responsividade a estímulos é reduzida (Siegel 2008). Esses períodos inativos podem ocorrer durante a noite ou durante o dia, mas são sempre substituídos por períodos de maior atividade e maior responsividade aos estímulos, e tudo isso dentro de aproximadamente 24 horas (Siegel 2008). As medusas (gênero *Cassiopea*) apresentam atividade muscular mais intensa durante o dia e menos durante a noite. Curiosamente, forçar o movimento das medusas durante a noite, reduz sua responsividade no dia seguinte, o que sugere um ajuste homeostático a este período (Satterlie 2011). Mosca das frutas (gênero *Drosophila*) também apresentam períodos de imobilidade e baixa responsividade durante a noite, e que se perturbados, também levam a redução da locomoção no período ativo (durante o dia) subsequente, novamente indicando uma função homeostática deste período (Kaiser and Steiner-Kaiser 1983). Em humanos, a redução da atividade metabólica pode ser atingida por meio de diferentes mecanismos (Anafi, Kayser, and Raizen 2019). Por exemplo, o uso de um anestésico geral leva a um estado de menor atividade metabólica com redução na responsividade, porém este estado não é regulado homeostaticamente. O único período que apresenta, concomitantemente, ajuste homeostático, redução metabólica e de diminuição da responsividade é o estado denominado sono. Portanto, parâmetros de quiescência metabólica, grau de responsividade e presença de ajuste homeostático são geralmente utilizados para determinar se um animal está em estado de sono ou acordado. Uma interessante característica que também é atribuída ao sono é a reversibilidade rápida. Diferentemente do coma, por exemplo, no qual o retorno à vigília depende de vários fatores, durante o sono um forte estímulo olfativo já pode ser o suficiente para retomar total vigilância (Anafi, Kayser, and Raizen 2019). Porém, a característica de destaque para o sono é a regulação homeostática, pois é exclusiva desse estado, o que sugere sua extrema importância (ver quadro 1).

Quadro 1 - Propriedades comportamentais do sono

Quiescent behavioural state	Reduced responsiveness?	Rapid reversibility?	Homeostatically regulated?
Sleep <sup>a</sup>	Yes	Yes	Yes
Quiet wakefulness	No	Yes	No
Torpor or hibernation	Yes	No	No
Tonic immobility	No	No	No
Stupor or coma	Yes	No	No
General anaesthesia	Yes	No	No

<sup>a</sup>Sleep is the only state to fulfil all three behavioural criteria; therefore, these behavioural tests have been used to distinguish sleep from other quiescent behavioural states.

*Adaptado de Ron C. Anafi, Matthew S. Kayser e David M. Raizen, 2018*

Diversos parâmetros fisiológicos apresentam ritmos circadianos e podem atuar como preditores dos estados. A temperatura do corpo durante a vigília se mantém em média a 37°C, mas reduz aproximadamente 1°C durante o sono (Figura 1-A). Os hormônios do crescimento (Figura 1-B) e do cortisol (Figura 1-C) também elevam seus níveis plasmáticos durante o dia e reduzem durante a noite. A melatonina, um hormônio responsável pelo controle dos outros hormônios descritos acima e da temperatura, também apresenta ciclagem circadiana (Figura 1-D). A melatonina é produzida na glândula pineal, mas sua síntese é controlada pelo núcleo supraquiasmático (NSQ). Na presença de luz, as células ganglionares fotossensíveis da retina são ativadas, levando à ativação do NSQ, que por sua vez, inibe a produção da melatonina. Na ausência de luz, o NSQ libera a pineal para produzir a melatonina, induzindo sono (Liu et al. 2016). Devido ao seu papel no controle do ciclo circadiano, a melatonina tem sido utilizada como tratamento para diversas doenças, como distúrbios no sono e até mesmo câncer (Figura 2).

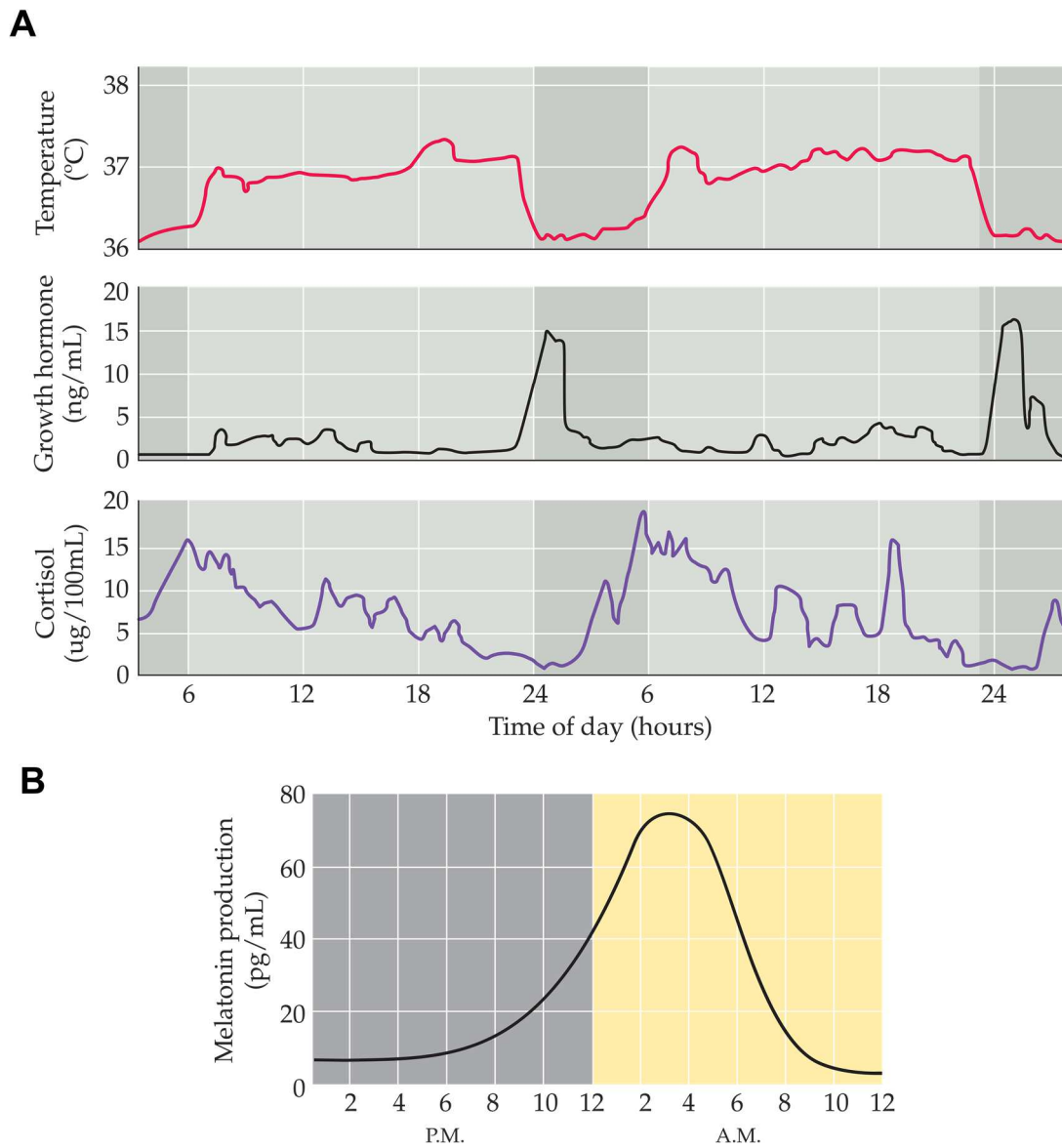


Figura 1. A: Flutuações da temperatura do corpo (gráfico superior), da concentração plasmática do hormônio do crescimento (gráfico no meio) e do cortisol (gráfico inferior) ao longo de dias; B: Flutuação da concentração plasmática de melatonina ao longo do dia. Adaptado de Neuroscience, Sixth edition, Purves et al, 2018.

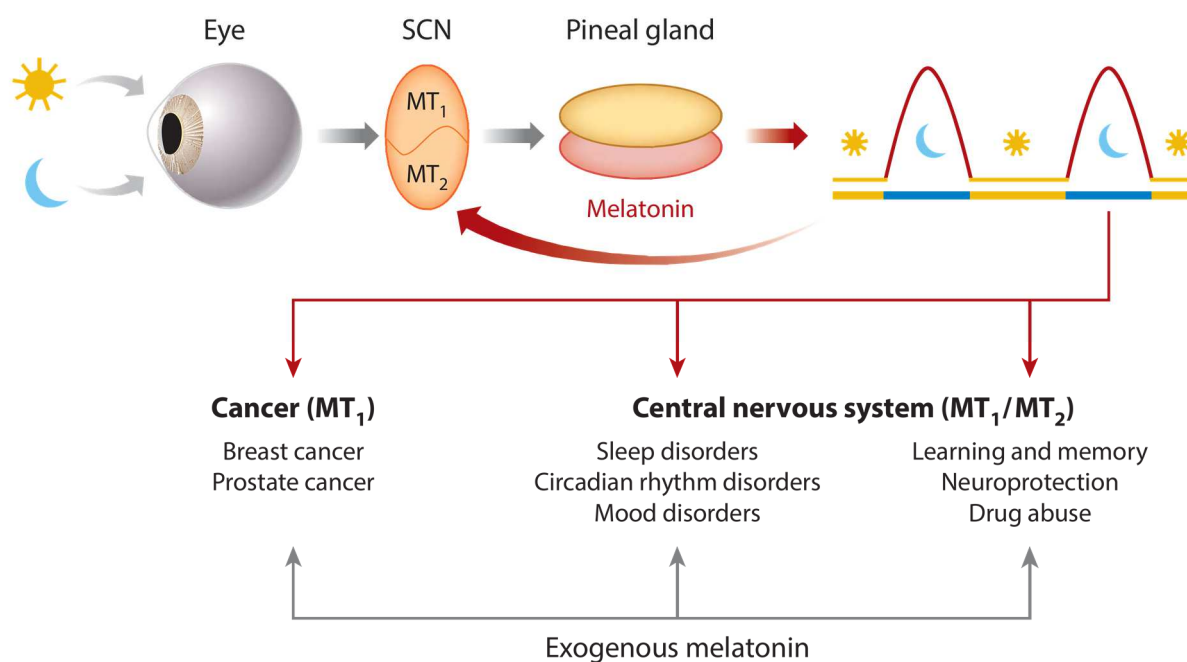
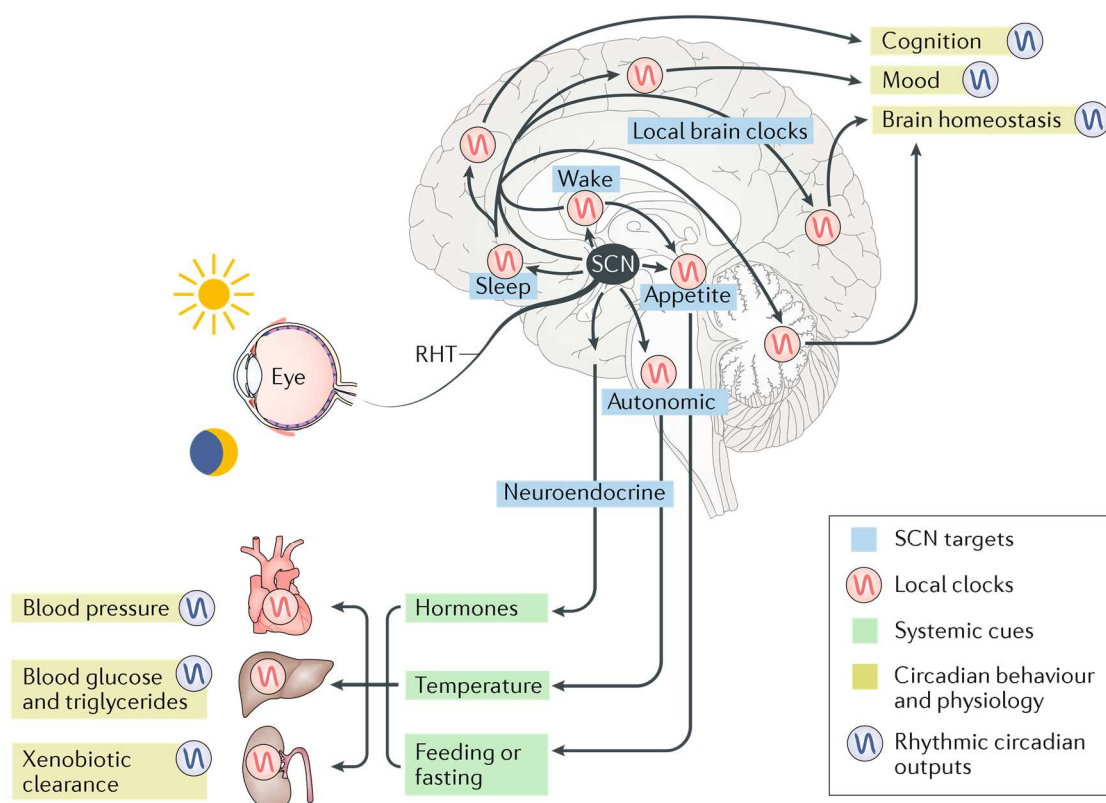


Figura 2. Ciclo da melatonina e algumas aplicações terapêuticas. A melatonina é produzida na glândula pineal, que é modulada pelo NSQ. Seguindo o ritmo circadiano, aumenta sua concentração plasmática durante a noite. Os receptores da melatonina, chamados de MT<sub>1</sub> e MT<sub>2</sub>, são expressos no próprio NSQ, mas também nos diversos tecidos e são responsáveis pelos efeitos celulares do hormônio (acoplados a proteínas G inibitórias). Atualmente têm sido propostas terapias com administração de melatonina no combate a diversos tipos de câncer, distúrbios no ciclo circadiano entre outros. Adaptado de Jiabei Liu et al, 2016.

Existe um corpo considerável de evidências mostrando que o NSQ é capaz de sincronizar a expressão de genes do relógio biológico de diferentes regiões do cérebro e do corpo (Figura 3), por meio do controle de suas funções neuroendócrinas e da sinalização autonômica, além do papel regulatório do comportamento do animal, o que demonstra seu papel de marca-passo principal do ciclo circadiano (Hastings, Maywood, and Brancaccio 2018). Como consequência, a expressão desses genes também poderia ser utilizada como marcador das etapas do ciclo circadiano (Koronowski and Sassone-Corsi 2021). Atualmente existem diversos trabalhos mostrando que uma circuitaria complexa no cérebro interage de maneira direta ou indireta com o NSQ na indução dos estados do ciclo circadiano (A. Adamantidis and Lüthi 2018), tendo a participação de diversos sistemas de neurotransmissão, como o (1) dopaminérgico (Luo et al. 2018; Oishi et al. 2017) ou (2) colinérgico (Zant et al. 2016), (3) diversas estruturas como sistema reticular ascendente do tronco encefálico (Kroeger et al. 2017) ou (4) subpopulações neuronais de diversas regiões como o

hipotálamo lateral (Burdakov 2020) ou a área tegumentar ventral (Yu et al. 2019), além das importantes contribuições dos astrócitos espalhados por todo o encéfalo (Bjorness et al. 2016). Porém, estas interações são complexas e não são exclusivas de um estado, o que torna difícil utilizá-los como marcadores das etapas do ciclo (Scammell, Arrigoni, and Lipton 2017).



*Figura 3. O NSQ recebe inervação direta do trato retino hipotalâmico (RHT) e sincroniza sua atividade com a captação de luz, ou seja, com o ciclo circadiano. O NSQ projeta para várias partes do cérebro e do corpo, que contém seus próprios relógios moleculares. O NSQ sincroniza diretamente os relógios biológicos do cérebro e sincroniza os relógios do corpo indiretamente por meio dos outros hormônios, da flutuação da temperatura e dos comportamentos modulados pelo NSQ. Adaptado de Michael H. Hastings, Elizabeth S. Maywood and Marco Brancaccio, 2018*

O desenvolvimento da técnica do registro eletroencefalográfico (EEG) permitiu compreender que o sono é constituído de diferentes fases com padrões de atividade oscilatória no cérebro bastante distintos: (1) o sono com movimento rápido dos olhos (*Rapid Eye Movement* - REM, também chamado de sono paradoxal) e (2) o sono não-REM (*Non Rapid Eye Movement* - NREM, também chamado de sono de ondas lentas).

Essas oscilações podem ser captadas a partir de eletrodos posicionados sobre o crânio, e, portanto, captam o resultado de um somatório de fluxos de cargas nos neurônios próximos do eletrodo (Buzsáki, Anastassiou, and Koch 2012; Herreras 2016; Kajikawa and Schroeder 2011). O avanço das técnicas de registro por meio do implante de eletrodos e/ou arranjo de eletrodos em áreas profundas do cérebro (Herreras 2016) aliado a recentes contribuições da optogenética possibilitaram investigações mais diretas em regiões específicas do cérebro e até sobre o comportamento de subpopulações neuronais específicas de cada região, e têm demonstrado uma complexa interação entre diversas áreas do cérebro na geração desses ritmos oscilatórios. Diversos trabalhos têm suportado a hipótese de que os neurônios em certas áreas do cérebro sincronizam seus disparos em ritmos característicos e dependentes do estado do animal (Benchenane et al. 2010; Diba et al. 2014; Harris 2005), que são captados pelos eletrodos e podem atuar como marcadores de processos biológicos de extrema relevância fisiológica (Drieu, Todorova, and Zugaro 2018; Marshall et al. 2020; de Vries, Slagter, and Olivers 2020) e também podem definir os estados ao longo do ciclo circadiano (A. R. Adamantidis 2019).

Durante o sono NREM, as redes encefálicas e hipocâmpais reduzem a frequência de disparos (Miyawaki and Diba 2016; Miyawaki, Watson, and Diba 2019), e assumem um padrão de atividade de elevada sincronia (Vyazovskiy et al. 2009), o que faz com que os eletrodos de registros captem flutuações de baixa frequência e elevada amplitude do sinal dentro da faixa de 1 a 4 Hz (limites variam entre os autores), atividade denominada *Slow Wave Activity (SWA)* ou *Delta* (Figura 4-B). Ao longo de uma noite de sono, as oscilações Delta decaem de amplitude e apresentam efeito rebote em caso de privação (Vyazovskiy et al. 2009). Dada sua predominância espectral durante o sono NREM, as oscilações Delta atuam como um importante marcador eletrofisiológico desse estágio do sono. As oscilações Delta são geradas a partir da orquestração do circuito cortical pelo tálamo, no qual destaca-se o papel das fibras tálamo-corticais (TC) que passam a exibir um padrão fásico de disparos, consequência do platô despolarizante induzido pelos canais de cálcio sensíveis a tensão, forçando toda a rede cortical a entrar em sincronia (Crunelli et al. 2018).

Durante o sono REM, os neurônios em específico do hipocampo aumentam sua frequência de disparos em comparação com o sono NREM (Grosmark et al. 2012; Miyawaki, Watson, and Diba 2019) e sincronizam na faixa de *Teta* (6 a 10 Hz, variando de acordo com os autores), que é uma atividade captada também no córtex (Figura 4-C). As oscilações Teta no hipocampo emergem da ativação rítmica dos neurônios GABAérgicos do septo medial, que projetam para a rede inibitória do hipocampo e forçam o hipocampo a sincronizar em sua frequência (Colgin 2016). O septo medial possui um mecanismo intrínseco de disparo dentro da faixa de Teta (Varga et al. 2008), o que o torna um importante marcapasso deste ritmo. Portanto, as oscilações Teta atuam como marcadores do estágio de sono REM, mas não são exclusivos desse estado, uma vez que também estão presentes durante a vigília (Drieu and Zugaro 2019). Os ciclos Teta durante sono REM e vigília marcam posições temporais de prevalência para disparos de neurônios, atuando como sinalizadores de uma orquestração na dinâmica de atividade dos neurônios (Buzsáki 2010). Sendo assim, destaca-se o papel da Teta para processos de codificação e consolidação de memórias (Colgin 2016). As redes corticais e hipocampais, durante o sono REM e vigília, também apresentam a atividade em *Low e High Gamma* (30 a 60 Hz e 60 a 90 Hz, respectivamente), que surgem como reflexo da ativação de circuitarias locais (Buzsáki and Wang 2012), e se acoplam às oscilações Teta (Sirota et al. 2008), sendo também importantes para a formação de memórias (Drieu and Zugaro 2019). Durante a vigília, as diferentes entradas sensoriais e seus processamentos privilegiam a potencialização das sinapses (Tononi and Cirelli 2014). Mecanismos já bem estudados de potencialização a longo prazo (LTP) nas sinapses do hipocampo (Nicoll 2017) demonstram que as etapas de codificação de memórias culminam com uma predominância desses processos para o aumento da excitabilidade (Citri and Malenka 2008). Durante o sono NREM o cérebro tende a contrabalançar predomínio (Tononi and Cirelli 2020), privilegiando os processos de depressão sináptica por um mecanismo também muito bem descrito para a circuitaria do hipocampo (Lamprecht and LeDoux 2004), chamado de depressão a longo prazo (LTD).

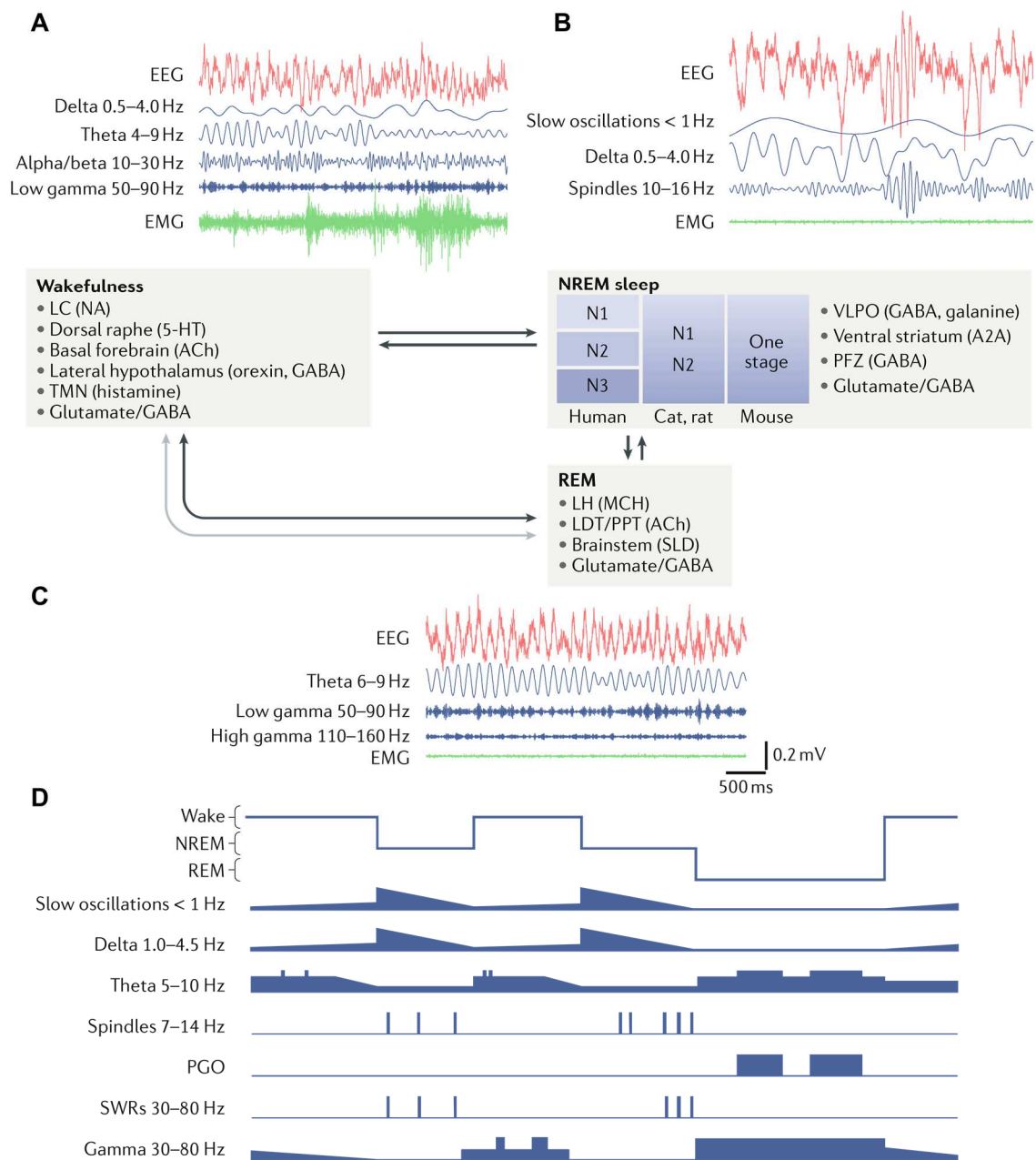


Figura 4. Resumo das oscilações típicas de cada estado. A: Exemplo de traçado típico do estado acordado, com sinal filtrado em Delta, Teta, Alpha/Beta e Low Gamma. Atenção para o sinal do eletromiograma EMG. LC, Locus Coeruleus; TMN, Núcleo Tuberomamilar; B: Mesmo para o sono NREM. VLPO, área pré-óptica ventrolateral; PFZ, Zona Parafacial; C: Mesmo para sono REM. LH, Hipotálamo Lateral; LDT, Tegumento látero-dorsal; D: Dinâmica ao longo do tempo da potência nas oscilações associadas à classificação. PGO, oscilações ponto-genículo-occipitais. Adaptado de Antoine R. Adamantidis, Carolina Gutierrez Herrera e Thomas C. Gent, 2019.



## 2.2 - Memória, hipocampo e sono

Uma das estruturas do sistema nervoso central de mamíferos que tem atraído muito interesse é o hipocampo. Esta região é fundamental para o processo de formação de memórias episódicas (Drieu and Zugaro 2019) e apresenta atividades oscilatórias bastante distintas, dependendo da fase do ciclo sono-vigília que o indivíduo se encontra. Tais variações de padrões de atividade têm sido extensamente investigadas para compreender melhor o papel do sono em processos de formação de memória (A. R. Adamantidis 2019). Localizada no lobo temporal (Figura 5), a formação hipocampal basicamente compreende o giro denteado (GD) e as regiões chamadas de *Cornu Ammonis* (CA), divididas didaticamente de acordo com o perfil de proteínas expressadas (Hainmueller and Bartos 2020). O hipocampo apresenta uma rede densa de neurônios enfileirados em camadas extremamente organizadas que lhes confere a possibilidade de serem orquestradas de forma síncrona (Drieu and Zugaro, 2019). Em resumo, a circuitaria possui duas vias principais:

- a via direta (também chamada de via perforante), na qual axônios partem da camada III do córtex entorrinal e atingem diretamente os dendritos de CA1 (Hainmueller and Bartos 2020). Esta via tem sido relacionada à entrada de novas informações para o hipocampo durante codificação do que é explorado no ambiente (Takahashi et al. 2014);
- A via indireta (também chamada de via tri sináptica), na qual axônios do córtex entorrinal projetam primeiramente ao GD. Do GD partem as fibras musgosas até CA3, que atua como um importante integrador das informações de entrada para o hipocampo (Zhang, Schlögl, and Jonas 2020). De CA3 partem axônios, chamados de colaterais de Schaffer, para CA1. Esta via está relacionada à recuperação de informações codificadas previamente, atuando como um integrador de memórias (Cabral et al. 2014).

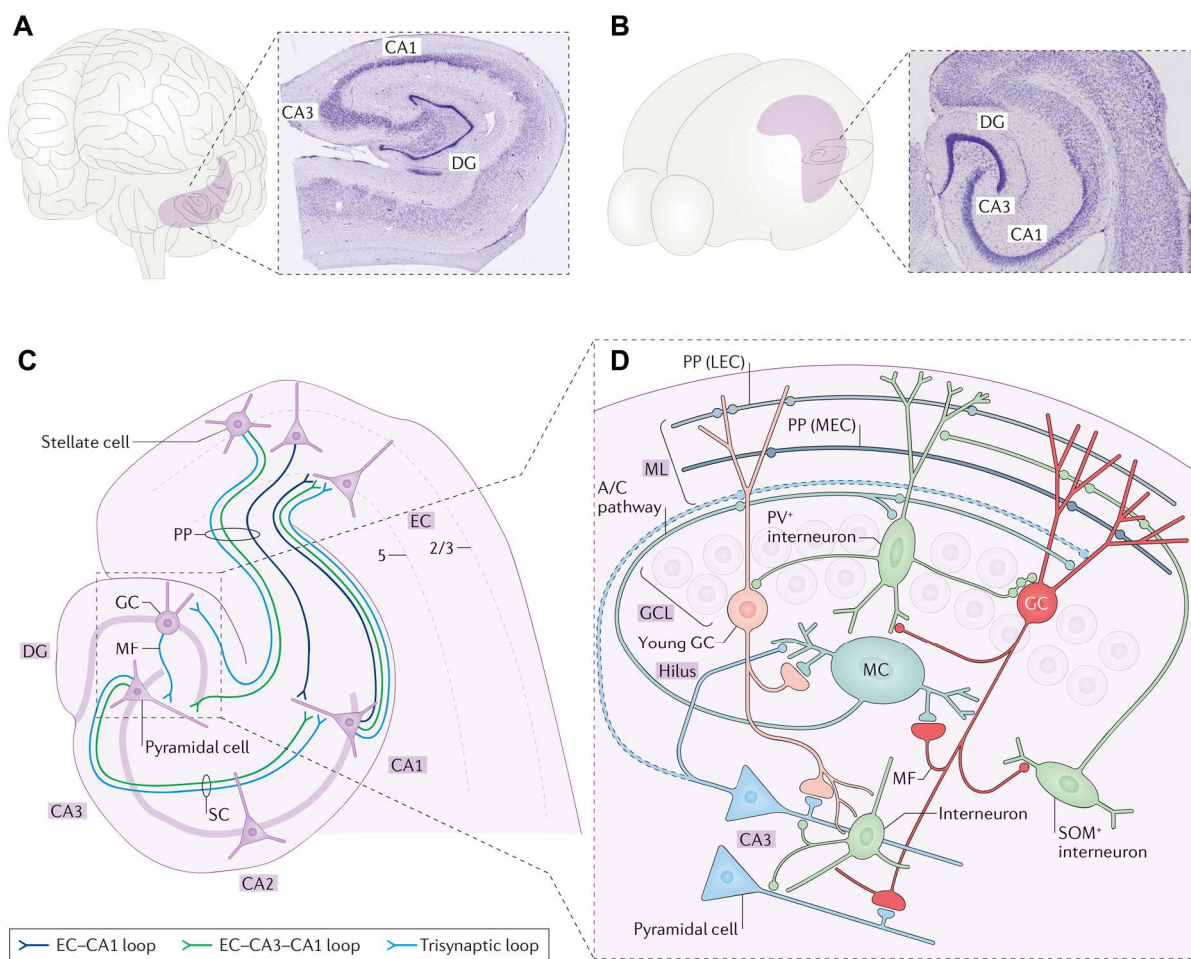


Figura 5. Organização anatômica do hipocampo. A: Hipocampo em humanos, com fatia colorida por Nissl; B: O mesmo para o cérebro de roedor; C: Desenho esquemático da formação hipocampal e suas principais conexões. Células granulares (GC) do giro denteado (DG) enviam axônios chamados de fibras musgosas (MF) para CA3. As células piramidais de CA3 enviam projeções chamadas de colaterais de Schaffer (SC) para CA1, que projeta de volta para o córtex entorrinal (EC); D: Microcircuitaria do DG amplificada, demonstrando conexões internas e as saídas para CA3. Neurônios no DG formam uma densa camada (GCL) composta por neurônios parvalbumina positivos (PV +), células cesta (basket cells) e neurônios com projeções axo-axônicas. Embrulhadas pela GCL estão as células musgosas (MC) e neurônios que expressam somatostatina (SOM +). A/C pathway, via associativa-comissural; LEC e MEC, córtex entorrinal lateral e medial, respectivamente; ML, camada molecular.

O hipocampo durante o sono NREM e em períodos pós-consumatórios durante a vigília, apresenta eventos oscilatórios de alta frequência denominados *sharp-wave ripples* (150 a 250 Hz) que são importantes para a consolidação de memórias (Klinzing, Niethard, and Born 2019) e também contribuem para a homeostase sináptica (Norimoto et al. 2018). Por sua vez, o córtex durante o sono NREM gera as

*spindles* (10 a 15 Hz) por um mecanismo de gênese similar ao da Delta (Crunelli et al. 2018). O acoplamento sequencial de *ripples-Delta-spindles* são importantes marcadores do processo de consolidação de memórias (Buzsáki 2015). A geração artificial de delta (Maingret et al. 2016) e de *spindles* (Latchoumane et al. 2017) aumentam a incidência desse acoplamento e resultam em melhora cognitiva, como também a inibição de *ripples* leva a prejuízos de memória (Girardeau et al. 2009).

### **2.3 - Classificação do ciclo sono-vigília**

Em experimentação científica animal, as classificações do ciclo sono-vigília são baseadas nas características eletrofisiológicas típicas dos estados (Katsageorgiou et al. 2015). Períodos com elevada atividade muscular são critérios decisivos para caracterizar o estado acordado, enquanto que períodos de silenciamento da maioria dos músculos caracterizam o sono (A. R. Adamantidis 2019). Alguns trabalhos calculam a raiz do valor quadrático médio (em inglês *root mean squared* - RMS) do EMG, obtendo um único valor que representa toda a atividade muscular do animal no período registrado (Durán et al. 2018; Soltani et al. 2019). As oscilações Teta e Delta são predominantes nos estados de sono REM e NREM respectivamente (A. R. Adamantidis 2019), e a razão Teta/Delta tem sido frequentemente utilizada como parâmetro principal para diferenciar estes estados (Durán et al. 2018; Katsageorgiou et al. 2015; Norimoto et al. 2018; Soltani et al. 2019), pois também resume a atividade de ambas oscilações em um único valor. Com isso, é possível calcular um único valor que represente ambas atividades do cérebro e músculos durante a classificação.

Os pesquisadores nesses estudos devem lidar com uma série de desafios nas classificações, como (1) registros geralmente feitos por horas ou até mesmo por dias (Souza et al. 2018), (2) arquiteturas do ciclo sono-vigília complexas, que apresentam variações diárias, dependentes da idade, sexo e do modelo animal estudado (Cirelli and Tononi 2015; Mander, Winer, and Walker 2017), e (3) ainda precisam considerar que doenças podem causar perturbações não usuais em padrões oscilatórios ou na atividade dos músculos que são difíceis de serem detectadas, como as espículas interictais em epilepsia (Mendes et al. 2019) e o distúrbios do sono REM na doença de Parkinson (De Castro Medeiros et al. 2019), respectivamente.

Na literatura, é possível encontrar uma diversidade de métodos de classificação de sono e vigília baseados em padrões oscilatórios e atividade muscular de roedores (Quadro 2). Existem aqueles que são resumidos em classificações manuais dos experimentadores (Durán et al. 2018; Koike et al. 2017; Neckelmann et al. 1994), sendo que para estes casos, estima-se que um tempo aproximadamente equivalente ao do registro é necessário para a classificação (Brankačk et al. 2010). Outros métodos combinam processamentos automáticos com verificações manuais, e são mais comuns, compreendendo boa parte dos métodos disponíveis (Faust et al. 2019). Porém muitos dos métodos não relatam o acurácia dos classificadores ou quando o fazem, atingem eficácia relativamente baixa para o sono REM (Figura 6). Em outros estudos, o código utilizado ou não está disponível (*código fechado*) ou não é de fácil uso, especialmente por pesquisadores que não são especialistas em linguagens de programação como Matlab e Python (Vladimir et al. 2020; Yamabe et al. 2019; Yildirim, Baloglu, and Acharya 2019). Como consequência direta, os estudos sobre a arquitetura do sono ainda enfrentam problemas com a reprodutibilidade dos achados (Frommlet 2020).

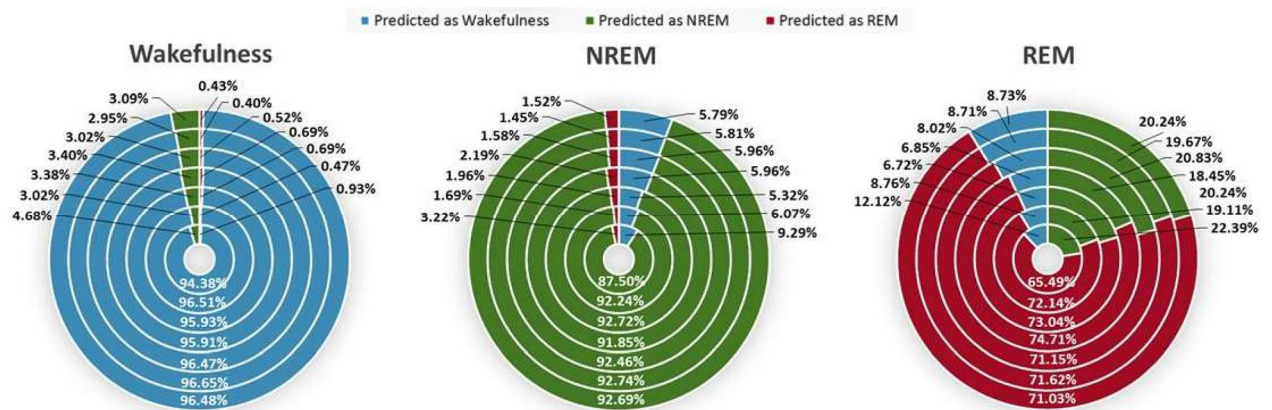


Figura 6. Acurácia de diferentes métodos testados. Cada círculo concêntrico corresponde a um dos métodos. Os métodos dispostos de dentro para fora são: DTree (Brankačk et al. 2010), k-NN (Rytönen, Zitting, and Porkka-heiskanen 2011), LDA (Brankačk et al. 2010), NB (Rytönen, Zitting, and Porkka-heiskanen 2011), l-SVM (Zeng et al. 2012), RBF-SVM (Katsageorgiou et al. 2015), MLP (Katsageorgiou et al. 2015). Adaptado de Vasiliki-Maria Katsageorgiou et al, 2015.

Quadro 2 - Trabalhos que envolveram classificações do ciclo sono-vigília

Autores	Modelo Animal	Dado registrado	Método de classificação	Tamanho do registro	Código aberto?	Reporta acurácia?
---------	---------------	-----------------	-------------------------	---------------------	----------------	-------------------

Durán et al, <b>SLEEPJ</b> 2018	Rato	Córtex e Hip	Classificação manual	24 horas	Não	Não
Q. Vo et al, <b>Parkinson's Disease</b> 2014	Rato	EEG + EOG + EMG	Limiares arbitrários	24 horas	Não	Não
Miyawaki et al, <b>Sci Rep</b> 2019	Rato	LFP CA1 + EMG	Limiares arbitrários	24 horas	Não	Não
Boyce et al, <b>Science</b> 2016	Camundongo	EEG + CA1 + EMG	Classificação manual	4 horas	Não	Não
Vyazovskiy et al, <b>Neuron</b> 2009	Rato	EEG + CA1 + EMG	Classificação manual	2 a 3 semanas	Não	Não
Norimoto et al, <b>Science</b> 2018	Camundongo	CA1 + Video	Limiar arbitrário	4 a 6 horas	Não	Não
Koike et al, <b>J. Neuro.</b> 2017	Rato	LFP + EEG + EMG	Classificação manual	6 horas	Não	Não
Yamamoto e Tonegawa, <b>Neuron</b> 2017	Camundongo	Hip + MEC + Video	Limiares pré-definidos	30 - 45 min	Não	Não
Rothschild et al, <b>Nat. Neurosci.</b> 2016	Camundongo	EEG + EMG	Software pago	24 horas	Não	Não
Kroeger et al, <b>J. Neuro.</b> 2017	Camundongo	EEG + EMG	Software pago	12 horas	Não	Não
Cho et al, <b>Neuron</b> 2017	Camundongo	EEG + EMG	Software pago	12 horas	Não	Não
Saito et al, <b>Front. Neurosci.</b> 2018	Camundongo	EEG + EMG	Classificação manual	24 horas	Não	Não
Iyer et al, <b>Npj Parkinsons Dis.</b> 2019	Rato	EEG + EOG + EMG	Limiares arbitrários	24 horas	Não	Não

---

Fonte: Elaboração própria.

Alguns grupos se propuseram a desenvolver ferramentas exclusivamente focadas na classificação do ciclo circadiano (Quadro 3). Algumas dessas ferramentas se baseiam em simples definições de limiares arbitrários auxiliados por

representações gráficas (Gross et al. 2009; Silva-Pérez et al. 2020), que ainda apresentam alta suscetibilidade a vieses dos experimentadores e os tornam de difícil reprodução. Um grande número de trabalhos recentes adotou algoritmos de redes neurais, como o *convolutional neural network* (CNN), para extrair informações complexas dos registros de EEG ou LFP e utilizá-las para a classificação (Barger et al. 2019; JIANG, MA, and WANG 2019; Li and Guan 2021; Maurice Abou Jaoude, Haoqi Sun, Kyle R. Pellerin, Milena Pavlova, Rani A. Sarkis, Sydney S. Cash, M. Brandon Westover\* 2018; Vladimir et al. 2020; Yamabe et al. 2019; Yildirim, Baloglu, and Acharya 2019). Porém, ferramentas como esta extraem padrões do dado que não estão sob domínio do usuário ou que não necessariamente têm relevância biológica (Alex and Lon 2019). Além disso, requerem um poder computacional e conhecimento técnico não condizente com a maioria dos laboratórios. Em um trabalho recente foram necessárias 160 horas, utilizando computador de alta acurácia, apenas para a etapa de treinamento do modelo, para então fazer a classificação (Yamabe et al. 2019).

Quadro 3 - Métodos de classificação já descritos na literatura

Autores	Dado de entrada	Algoritmo principal	Tem GUI?	É código aberto?
Masato Yamabe et al, <b>Sci Rep</b> 2019	EEG + EMG (base de dados)	Redes Neurais	Não	Não
Vladimir Svetnik et al, <b>Journal of Neuroscience Methods</b> 2020	EEG + EMG (base de dados)	Redes Neurais	Não	Não
Shelly Crisler et al, <b>Journal of Neuroscience Methods</b> 2007	EEG + EMG	SVM	Não	Não
Manuel Silva-Pérez et al, <b>J. Sleep Res.</b> 2020	EEG + EMG	Limiares arbitrários	Não	Não
Thomas Lampert et al, <b>Journal of Neuroscience Methods</b> 2015	EEG + acelerômetro	Classificadores múltiplos	Não	Não
Brooks et. al. <b>Journal of Neuroscience Methods</b> 2009	EEG + EMG	Limiares arbitrários	Sim	Sim
Jurij Brankačk et al,	EEG + EMG	Análise de	Não	Não

<b>Brain Research</b> 2010		Discriminante Linear		
Zeke Barger et al, <b>Plos One</b> 2019	EEG + EMG	Redes Neurais	Sim	Sim
Vance Gao et al, <b>Journal of Neuroscience Methods</b> 2015	EEG + EMG	Classificadores múltiplos	Não	Não

Fonte: Elaboração própria.

Curiosamente, o acurácia dos métodos desenvolvidos e descritos na literatura, quando demonstrados, sempre são calculados comparando dados de classificações manuais com a classificação final. Isso faz com que os dados da classificação por especialistas sejam subutilizados e colocados em uma posição passiva para a classificação, quando poderiam assumir papel ativo e crucial. Diferentes estratégias matemáticas foram propostas para o cálculo do acurácia, como cálculos da taxa de verdadeiros e falsos positivos (Hoo, Candlish, and Teare 2017), ou o coeficiente *Kappa* (Raadt et al. 2019), que apenas servem para demonstrar o grau de proximidade da classificação final com a classificação do usuário.

## 2.4 Objetivos gerais

Nosso objetivo foi desenvolver uma *Toolbox* código fechado para a classificação do ciclo sono-vigília de registros feitos em modelos murinos.

## 2.5 Objetivos específicos

1. Desenvolver um método de classificação dos estados acordado, sono NREM e REM;
2. Testar a relevância da classificação manual como parte ativa da classificação para atingir o melhor acurácia possível;
3. Desenvolver uma interface gráfica de fácil implementação e de código aberto;
4. Implementar uma série de *scripts* capazes de gerar figuras que possibilitam avaliar desde o acurácia da classificação como a qualidade do sinal adquirido;
5. Testar o acurácia do classificador em registros de roedores saudáveis e modelos da doença de Parkinson;

### 3. Métodos

A *Toolbox* foi desenhada visando a maior acessibilidade possível para os diferentes usuários. O usuário deverá definir alguns parâmetros, em especial para as figuras. Porém, ressaltamos que para utilizar a *Toolbox*, é necessário que o dado esteja concatenado e no formato *MAT*. Com isso, o usuário deverá informar:

- Frequência de amostragem;
- Horários (local) de início e término de registros (esses horários serão necessários para algumas figuras);
- Frequência de ruído da rede elétrica local;
- As bandas de frequência que serão somadas à Teta, para então serem divididas por Delta. Ao fim, o algoritmo vai calcular a distribuição em *z-score*;
- Pasta destino para todas as figuras;

Foi preparado um arquivo em *.pdf*, disponível para todos, no qual é explicado o que o usuário deve fazer para conseguir utilizar a *Toolbox*. Nele também são apresentadas todas as figuras que são geradas durante a classificação (No PDF, essas figuras pertencem ao registro utilizado como treinamento do GMM, que por sua vez treina o modelo em todas as classificações). Este documento está como apêndice deste trabalho (Apêndice 2).

Tendo essas informações, o usuário pode utilizar a *Toolbox*. Ela iniciará a classificação, passando por todas as etapas automaticamente (Figura 7) até que a etapa da classificação manual se inicie. Feito isso, o algoritmo vai automaticamente finalizar as análises.

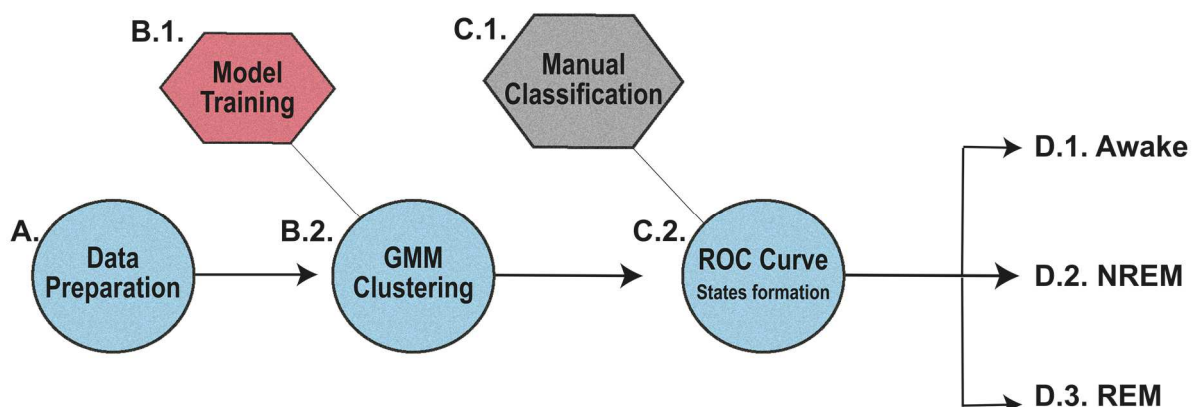




Figura 7. Fluxograma demonstrando as etapas da classificação.

A Figura 7 apresenta um fluxograma com as etapas da classificação, que são, resumidamente:

- *A. Preparação do dado: o dado coletado foi fragmentado em épocas menores e encaminhado para uma etapa de pré-processamento para extrair flutuações da linha de base e excluir faixa de frequência com ruído da rede elétrica local. Depois são os extraídos parâmetros do sinal registrado que serão utilizados na classificação;*
- *B.1. Treinamento do Modelo: um registro previamente classificado foi utilizado como treinamento do nosso algoritmo, visando aumentar acurácia da Toolbox na classificação;*
- *B.2. Agrupamento pelo Modelo de Misturas de Gaussianas (GMM): ferramenta utilizada para calcular a probabilidade de cada época pertencer a cada estado;*
- *C.1. Manual Classification: o usuário deve classificar manualmente um conjunto de épocas, que serão utilizadas na próxima etapa da classificação;*
- *C.2. Curva ROC (Formação dos estados): os dados da classificação manual são comparados com os grupos formados pelo GMM para definir qual o valor mínimo de probabilidade que as épocas devem possuir para serem classificadas em um dos estados (valor varia de um estado para outro);*
- *D.1 ao 3: as épocas são agrupadas nos estados e disponibilizadas para o usuário para futuras análises.*

Por fim, todas informações de entrada e saída da *Toolbox* são apresentadas no quadro abaixo:

Quadro 4. Entradas e saídas da *Toolbox*

Inputs do usuário	Outputs da <i>Toolbox</i>
Dado bruto coletado	Dado fragmentado
Horários de início e fim do registro	Valores de potência de cada época
Frequência de amostragem	Parâmetros das Misturas das Gaussianas
Bandas de frequência para ser utilizadas	Classificação final de todas as épocas
Frequência do ruído de rede local	Probabilidade posterior de cada época

Máxima verossimilhança

Limiaries de probabilidade escolhidos para cada estado

Taxas de verdadeiro e falso positivo

Área sob a curva (TP x limiaries e FP x limiaries)

Distribuições em *z-score* dos parâmetros escolhidos

---

Fonte: Elaboração própria

### 3.1 - Sujeitos experimentais

Os dados utilizados para implementação da *Toolbox* foram coletados de um rato Wistar macho adulto e de 8 camundongos *C57BL6* fêmeas adultas.

Os camundongos, adquiridos do Charles River Laboratories (Netherlands), foram designados aleatoriamente para os grupos controle (4 animais) e grupo modelo da doença de Parkinson (4 animais). Os animais foram mantidos em ciclo claro-escuro de 12 horas, sob condições estáveis de temperatura (a 22°C), com água e comida à vontade. Todos os experimentos com estes animais foram conduzidos no Instituto Karolinska, na Suécia, e receberam aprovação do comitê de ética local, em acordo com o *European Community Council Directive*, de 24 de Novembro de 1986 (86/609/EEC).

Um único rato Wistar macho foi utilizado, com peso inicial entre 200 e 350 gramas, e foi fornecido pelo Biotério Central da Universidade de São Paulo, campus de Ribeirão Preto. O animal foi habituado ao ambiente do biotério experimental com alternância de ciclo claro/escuro de 12 horas e temperatura média de 24°C. Todos os procedimentos experimentais utilizados neste animal foram aprovados *pelo Comitê de Ética em Uso de Animais (CEUA)* da FMRP sob o protocolo 16/2016.

### 3.2 - Procedimentos cirúrgicos

Para gerar o modelo da doença de Parkinson, camundongos receberam injeção de 6-hydroxydopamine-HCl (6-OHDA; Sigma-Aldrich, Sweden AB) no feixe

prosencefálico basal (via nigroestriatal, em inglês MFB). Os animais receberam também injeções subcutâneas de Temgesic® (0.1 mg/kg) como analgésico antes do procedimento cirúrgico. A anestesia foi induzida com isoflurano a 4%, e depois mantida com isoflurano a 2%. O animal anestesiado foi posicionado no estereotáxico acoplado a uma manta aquecedora para manter a temperatura fisiológica (David Kopf Instruments, Tujunga, CA).

A 6-OHDA (3,75µg/µL) foi dissolvida em 0,02% de ácido ascórbico e salina. Injetamos, em cada animal, 1 µl dessa solução a uma taxa de 0,2 µl/min segundo as seguintes coordenadas para a MFB (mm): anteroposterior (AP), -1.2; mediolateral (ML), -1.2; dorsoventral (DV), -4.8 (usando o bregma como referência). A agulha foi deixada posicionada no local por mais 5 minutos para correta dissolução da droga no tecido.

Uma semana após a injeção de 6-OHDA, os camundongos passaram pela cirurgia de implante de eletrodo (tungstênio *coberto* com teflon, uma unidade) em CA1 do hipocampo (AP - 1.7, ML - 1.2, DV - 1.5) para potencial de campo local (LFP), inserido ipsilateralmente à lesão. Outro eletrodo foi posicionado sob o músculo do pescoço para registro eletromiográfico (EMG). Um parafuso foi inserido no osso occipital e utilizado como referência do sinal. Um outro parafuso, inserido no parietal contralateral à injeção, serviu de apoio para estabilidade do arranjo do conector dos eletrodos ao sistema de registro. Ao fim da cirurgia, todo o arranjo foi coberto por resina acrílica.

O rato foi anestesiado com injeções de ketamina (100mg/kg, i.p. em NaCl 0,15 M) e xilazina (10mg/kg, i.p. em NaCl 0,15 M) e ajustado em manta aquecedora. O animal foi então posicionado no estereotáxico (KOPF; KOPF Instruments, EUA) para implante de eletrodo monopolar de registro (tungstênio encapado com teflon, 45 µm de diâmetro; AM-Systems, EUA) em CA1 do hipocampo (AP -5.4, ML -4.6, DV - 2.5) para registro do LFP e outro posicionado sob o músculo do pescoço para registro do EMG. Dois parafusos foram inseridos no osso occipital, um para referência do sinal e outro estabilidade do capacete de acrílico construído.

### 3.3 - Aquisição do dado

Após completa recuperação da cirurgia (peso normal), os camundongos foram individualmente alocados para a caixa de registro (caixa comum), com acesso a água e comida *ad libitum*. Os animais passaram por 12 horas de habituação, já plugados ao sistema de registro, mas no dia seguinte, iniciamos os registros às 19 horas, durando no total 25 horas. Ao início das análises, os dados referentes à primeira hora dos registros foram excluídos, sendo utilizadas 24 horas de registro para as análises.

O sistema de aquisição da Biopac software (MP160 System) foi utilizado para os registros em camundongo. O canal de registro do LFP do hipocampo foi coletado com taxa de amostragem de 2000 pontos por segundo, filtrado de 1 a 300 Hz e com ganho de 5000 vezes. Já o canal do EMG foi filtrado de 1 a 500 Hz, e com as mesmas configurações restantes.

O rato foi alocado para uma caixa comum por uma semana após o implante, onde teve acesso a água e comida à vontade, tendo seu peso monitorado até a recuperação de seu nível normal. Um dia antes do registro, o rato foi habituado à caixa de registro (caixa comum) por aproximadamente 4 horas. No dia seguinte iniciamos o registro às 10 horas da manhã até aproximadamente as 15 horas da tarde.

O rato foi registrado usando um amplificador A-M Systems 3600 (A-M Systems, EUA) acoplado ao sistema de aquisição Powerlab Daq 16/30 (AD Instruments). O canal do LFP e do EMG foram coletados a 4 000 pontos por segundo, filtrados de 0,1 a 1000 Hz e amplificados em 1 000 vezes.

Todas as análises foram implementadas usando MATLAB (MathWorks, 2019b). Todos os registros passaram pelas mesmas etapas descritas a seguir.

### 3.4 - Pré-processamento

O pré-processamento da *Toolbox* inicia com a fragmentação dos registros em épocas para serem analisadas individualmente (Figura 8). O tamanho das épocas

pode ser ajustado pelo usuário, mas neste trabalho optamos por épocas de 10 segundos, uma vez que são discutidas como ideais para balancear a complexidade dos sinais registrados com a eficácia das classificações (Brankačk et al. 2010; Lampert et al. 2015). Em seguida, algumas funções do MATLAB foram utilizadas para a padronização do dado de entrada para o classificador, sendo elas:

- *detrend*: utilizada para retirar flutuações não fisiológicas da linha de base de ambos canais, como consequência de eventuais momentos de instabilidade do sinal referência;
- *decimate*: utilizada para reduzir a amostragem para 1000 pontos por segundo;
- *eegfilt2* (Delorme and Makeig 2004): filtragem exclusiva do sinal do EMG entre 85 a 300 Hz;

Sendo assim, os registros feitos em camundongo, após fragmentados, somaram um total de 8640 épocas (provenientes das 24 horas). O registro em rato obteve um total de 1827 épocas (proveniente das aproximadamente 5 horas de registro).

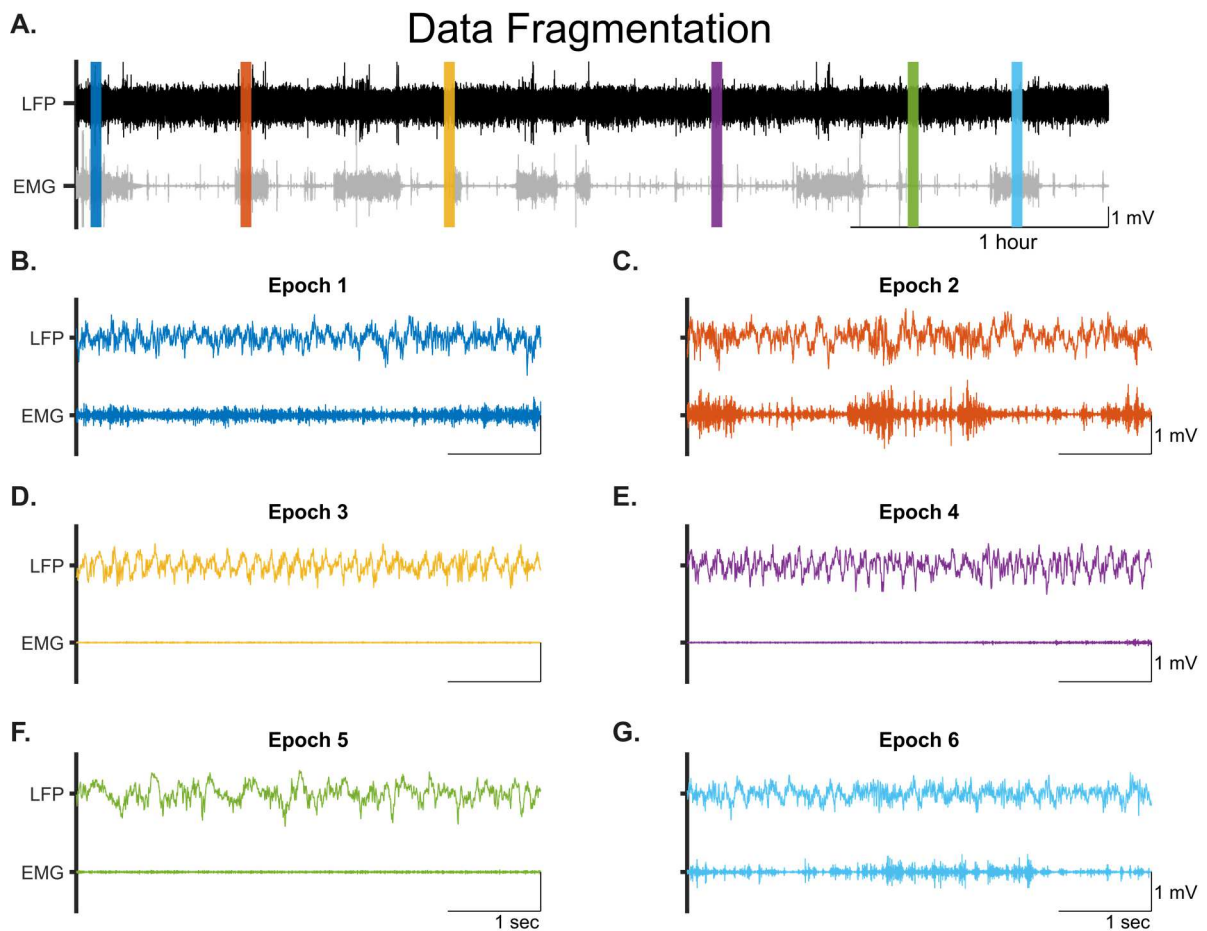


Figura 8. Fragmentação do registro. A: 4 horas escolhidas aleatoriamente de um dos registros, com LFP do hipocampo acima e sinal do EMG abaixo. Barras coloridas representam a posição das épocas também aleatoriamente selecionadas e apresentadas abaixo; B: Época representativa com atividade elevada no EMG; C: Mesmo que B; D: Época com EMG baixo e oscilação Teta proeminente; E: Mesmo que D; F: Época com EMG silenciado e oscilação Delta proeminente; G: Mesmo que B.

### 3.5 - Extração de componentes

Para o LFP, extraímos os valores de potência em cada época utilizando a função *pwelch* (Cohen 2014). A resolução de frequência para a função foi 2000 unidades, entre 1 e 500 Hz. O tamanho da janela foi a metade da taxa de amostragem (500 pontos), sem sobreposição das janelas. Os valores foram depois normalizados pelo somatório de todos os valores calculados pela mesma época. A distribuição das potências normalizadas foi integrada dentro das seguintes bandas de frequências:

- *Delta*: 1 a 4 Hz;

- *Teta*: 6 a 10 Hz;
- *Beta*: 10 a 30 Hz;
- *Low Gamma*: 30 a 50 Hz;
- *High Gamma*: 60 a 90 Hz;

Para cada época, calculamos a razão entre o valor de potência em Teta dividido pelo valor em Delta (chamado de agora em diante de razão ou relação Teta/Delta). Por fim, calculamos a distribuição em *z-score* dos valores de cada época dessa razão, segundo a fórmula a seguir:

$$z = \frac{x - \underline{x}}{S}$$

Onde  $x$  é o valor da relação Teta/Delta da época,  $\underline{x}$  é o valor de média da relação para todas as épocas do registro e  $S$  é o valor do desvio padrão.

Para o EMG, calculamos a raiz do valor quadrático médio (em inglês *root mean squared* - RMS) dos pontos em cada época, como na fórmula a seguir:

$$RMS = \sqrt{\frac{1}{N} \sum_{n=1}^N |x_n|^2}$$

Onde  $N$  é o número de pontos e  $x$  são os valores de amplitude registrados pelo eletrodo. Ao fim deste cálculo, cada época terá um único valor representando a atividade do EMG ao longo dos 10 segundos. Com isso, calculamos a distribuição em *z-score* dos valores de RMS do EMG das épocas.

Definimos como as variáveis para a classificação das épocas seus respectivos valores (*z-score*) da relação Teta/Delta do hipocampo e do RMS do EMG.

### 3.6 - Descrição do algoritmo: Modelo de Mistura de Gaussianas (*Gaussian Mixture Model - GMM*)

Um Modelo de Mistura é um modelo probabilístico que tem como objetivo sugerir grupos para as unidades de um conjunto de dados. Com isso, diferentemente de métodos de agrupamento conhecidos como *hard clustering algorithms* onde essas unidades são separadas em grupos, neste modelo é calculada a probabilidade de cada unidade de pertencer a cada um dos grupos (Nasraoui 2019). Este cálculo utiliza uma abordagem bayesiana que depende das características dos grupos e dos valores das unidades.

No caso do GMM, é assumido que cada grupo corresponde a uma distribuição gaussiana, ou seja, cada grupo possui suas unidades variando em torno de uma média e de um desvio padrão. Para o caso de mais de uma variável, é calculado a média de cada grupo nas variáveis e a matriz de covariância. O GMM atribui  $k$  grupos, que neste trabalho foram definidos pelo usuário, com valores independentes de média e covariância (Geoffrey McLachlan 2000). A função que descreve as gaussianas é representada como:

$$P(x) = \sum_{i=1}^k \alpha_i \cdot N(x|\mu_i, \Sigma_i)$$

Onde  $\mu_i$  e  $\Sigma_i$ , são a média e a matriz de covariância da gaussiana  $i$ ,  $\alpha_i$  é a probabilidade de  $x$  pertencer à Gaussiana  $i$ .

A função do MATLAB *fitgmdist* foi utilizada neste trabalho para os cálculos do GMM. A configuração nativa da função inicia o algoritmo calculando os valores iniciais de média e covariância de cada grupo a partir do *K-means*. Para este presente trabalho, primeiro classificamos um registro utilizando o algoritmo da *Toolbox*, e elegemos essa classificação inicial como treinamento do GMM. O treinamento consistiu em definir os valores de média e covariância deste primeiro registro como os



valores iniciais na função *fitgmdist* para todos os outros registros e em substituição do *K-means*.

O GMM utiliza o algoritmo chamado de *Expectation Maximization* (Sataloff, Johns, and Kost 2008) para redefinir as características das gaussianas de acordo com novas observações. É um processo dividido em duas etapas (como o nome sugere):

- *Expectation*: na qual é calculada a probabilidade da nova observação pertencer à distribuição;
- *Maximization*: na qual, utilizando probabilidade bayesiana, são calculados os novos valores de média e covariância da distribuição para que a observação de fato pertença à gaussiana;

Esse processo é repetido até que o valor da função de verossimilhança seja máximo (maximum-likelihood estimation - MLE), ou seja, até que os valores de média e covariância não se alterem mais para aquele conjunto de dados. Com os valores finais definidos, calculamos a probabilidade (posterior) de cada época pertencer a cada um dos grupos.

No caso do conjunto de dados utilizados neste trabalho, cada unidade é uma época de 10 segundos, cada gaussiana (grupo) é um estado (acordado, sono NREM e sono REM) e as variáveis utilizadas (características) são a razão Teta/Delta e o RMS do EMG (*z-score*), ou seja, cada estado apresenta um valor de média para a Teta/Delta e para o EMG e respectiva matriz de covariância. Os valores iniciais foram o registro de treinamento e cada nova observação no EM corresponde a uma época do registro analisado.

### **3.7 - Classificação Manual**

A classificação manual é a única etapa na qual o usuário deve intervir ativamente na classificação, com a finalidade de gerar um padrão ouro da classificação. A *Toolbox* se baseia neste padrão ouro para definir o limiar de probabilidade de cada estado (demonstrado na seção 3.8). Maiores detalhes desta etapa são apresentados na seção da Interface gráfica (Apêndice 1). Para isso, a

*Toolbox* seleciona épocas aleatoriamente e as apresenta para o usuário. De cada época, são apresentados:

- o sinal bruto do LFP e do EMG;
- PSD do LFP (gráfico de Densidade Espectral, em inglês Power Spectrum Density).

A classificação manual ocorre até que 0,5% do total de épocas sejam classificadas em cada estado. O que significa que para este trabalho, os registros em camundongo de 24 horas (8640 épocas no total) tiveram 40 épocas por estado classificadas manualmente (120 épocas classificadas manualmente por registro), e no registro feito no rato, 8 épocas (36 épocas classificadas manualmente).

Neste trabalho, as classificações manuais de todos os registros foram feitas por um experimentador que, utilizando a interface gráfica da *Toolbox*, completou a classificação manual dos registros de camundongo com aproximadamente uma hora em média, enquanto que o do registro em rato, cerca de 20 minutos.

### **3.8 - Cálculo da curva ROC (Característica de Operação do Receptor, em inglês *Receiver Operating Characteristic*) e das AUC (área sob as curvas).**

O limiar de probabilidade posterior foi encontrado comparando as épocas classificadas como cada estado pelo GMM e as classificadas manualmente pelo usuário. Essa comparação foi feita por meio do cálculo da taxa de verdadeiros positivos (TPR) e falso positivos (FPR) que cada limiar de probabilidade posterior atingiria caso fosse escolhido.

As TRP foram calculadas usando a classificação manual como referência, ou seja, do total de épocas classificadas manualmente em cada estado, quantas destas foram também classificadas pelo GMM no mesmo estado em cada um dos limiares. Este cálculo, para cada limiar de um determinado estado, pode ser expresso da seguinte forma:

$$\text{Taxa de verdadeiros positivos (TPR)} = \text{TP} / (\text{TP} + \text{FN})$$

Onde, o verdadeiro positivo (do inglês *True Positives* – TP) representa, das épocas classificadas manualmente num determinado estado, quantas também foram classificadas pelo GMM no mesmo estado. Já o falso negativo (do inglês *False Negatives* – FN) representa, das épocas que foram manualmente classificadas num determinado estado, quantas foram classificadas em outro estado pelo GMM.

Já a FPR, novamente considerando um único limiar em um dos estados, foi calculada da seguinte forma:

$$\text{Taxa de Falsos Positivos (FPR)} = \text{FP} / (\text{FP} + \text{TN})$$

Onde o falso positivo (do inglês *False Positives* – FP) representa, do total de épocas classificadas manualmente, quantas foram classificadas num estado diferente do analisado pela classificação manual, mas foram classificadas no estado analisado pelo GMM. O verdadeiro negativo (do inglês *True Negative* – TN) representa, das épocas classificadas manualmente, quantas foram classificadas em um estado diferente do analisado em ambos, classificação manual e GMM.

O limiar escolhido foi o que apresentasse menor distância ao ponto (0,1) no gráfico (Figura 17), calculado pela distância euclidiana de cada limiar a este ponto. O ponto (0,1) representa o valor máximo da taxa de verdadeiro positivo e mínima de falso positivo, apontando para a melhor acurácia possível para o classificador, dada às informações disponíveis (Hoo, Candlish, and Teare 2017).

A curva ROC apresenta a TPR e FPR para um conjunto de limiares definidos, e foi representado como um gráfico com as três informações (Figura 9-A). Porém, para o cálculo da área sob a curva (AUC), dividimos o gráfico em dois: AUC considerando a TPR pelos diferentes limiares (Figura 9-B) e considerando a FPR pelos diferentes limiares (Figura 9-C). Utilizamos a função *trapz* do MATLAB para o cálculo das AUCs, sendo que ela aproxima a integral de um intervalo para a área de um trapézio, sendo os diferentes limiares como base desse trapézio. Como consequência, os estados com elevada acurácia na classificação apresentarão os maiores valores de AUC TP e os menores de AUC FP (Figura 9-A, B e C), enquanto que os estados com pior

acurácia (Figura 9-D) apresentarão AUC TP baixa (Figura 9-E) e AUC FP alta (Figura 9-F).

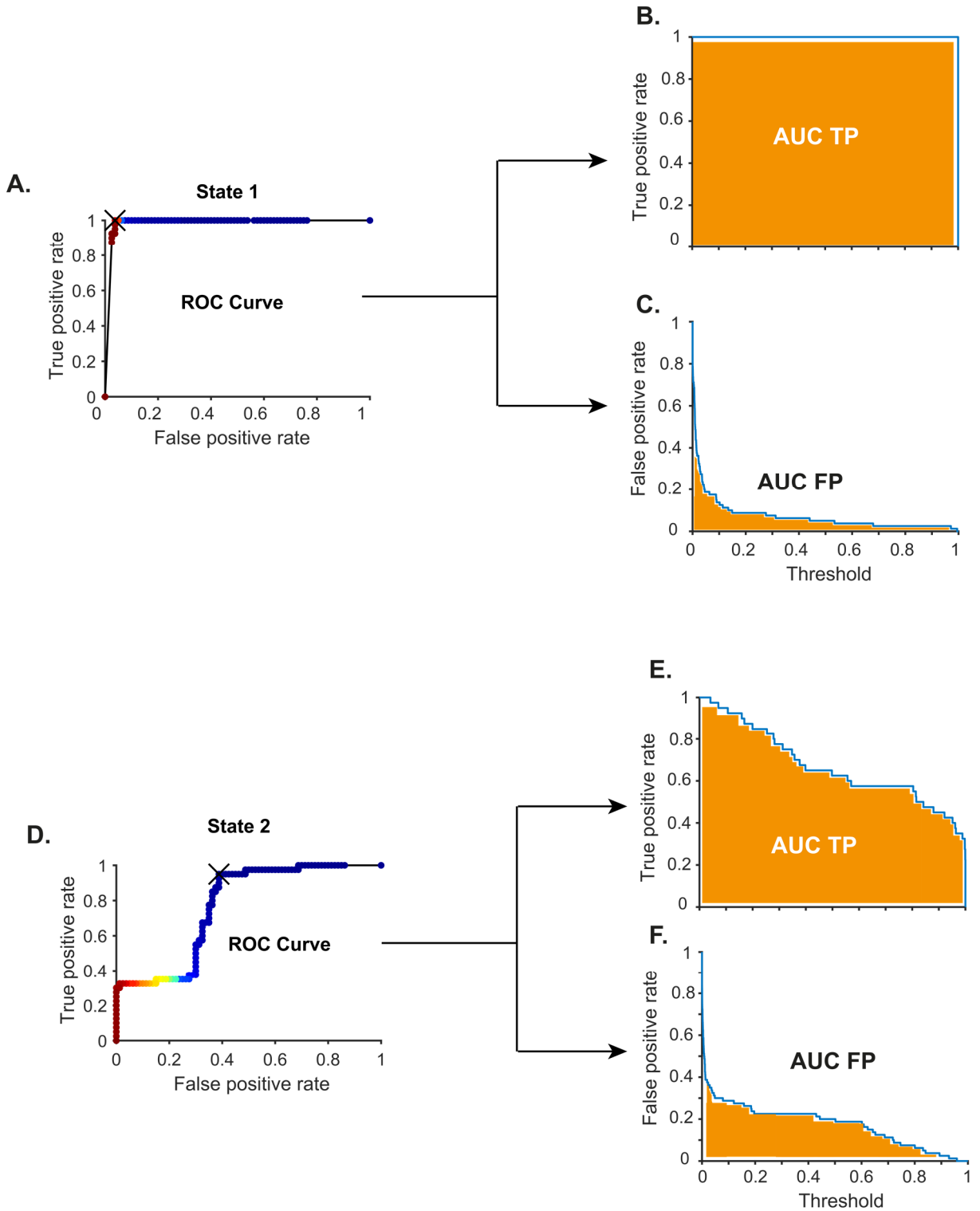


Figura 9. Demonstração do cálculo da área sob a curva pelo método trapezoidal. A: Curva ROC de um estado representativo com alta acurácia. Limiar escolhido marcado no gráfico; B: Área calculada sob a curva da taxa de verdadeiros positivos pelos diferentes limiares; C: Mesmo que em B, para taxa de

falso positivos pelos diferentes limiares; D: Mesmo que A, para um estado representativo com baixa acurácia; E: mesmo que em B; F: Mesmo que C.

#### 4. Resultados

Algoritmos de classificação variam de simples definições arbitrárias de limiares a complexas redes neurais (Quadros 2 e 3). Porém, definições arbitrárias são altamente susceptíveis à vieses dos experimentadores, e dificultam a reprodução dos achados, enquanto que a implementação desses algoritmos de redes neurais geralmente necessita de um especialista em programação e são de difícil implementação, o que normalmente não é a realidade da maioria dos laboratórios não especializados no assunto.

Neste estudo, desenvolvemos uma *Toolbox* para classificação de fases do ciclo sono-vigília de roedores que apresenta bom acurácia, é de livre acesso e possui interface gráfica que facilita sua implementação e favorece a reprodutibilidade dos achados, além de oferecer ao usuário representações gráficas e numéricas da classificação. Uma estratégia foi desenvolvida para otimizar o uso da classificação manual feita pelo usuário, tornando-a parte ativa na classificação. A *Toolbox* também oferece subsídios para que o usuário faça comparações entre diferentes registros de animais saudáveis e de modelos de doenças, sendo uma ferramenta útil para diferentes estudos sobre o ciclo circadiano.

Para desenvolver tal ferramenta, um conjunto de dados coletados pelo nosso grupo, em parceria com o grupo do Prof. Dr. Gilberto Fisone do Instituto Karolinska, na Suécia, foi utilizado. Foram obtidos registros do LFP em CA1 do hipocampo e do EMG do músculo do pescoço de 4 camundongos C57 por 24 horas, sempre iniciados às 19 horas e utilizamos esse conjunto de dados para a confecção da ferramenta. Em seguida, sua acurácia foi testada em registros de animais modelo da doença de Parkinson, os quais após a injeção de 6-OHDA, foram também implantados com eletrodos para LFP e EMG, e passaram pelas mesmas etapas de registro. Por fim, o método proposto foi aplicado em um registro feito em rato sob as mesmas condições, para investigar o acurácia da *Toolbox* também neste outro modelo murino.

Um registro representativo foi escolhido para as figuras desta seção, mas todos os registros passaram pelas mesmas análises, e algumas comparações foram feitas. A análise do dado e desenvolvimento do algoritmo foram feitos em MATLAB®. Descrevendo resumidamente a etapa de pré-processamento do dado, o sinal foi fragmentado em épocas de 10s, e em seguida extraímos a potência do sinal do hipocampo em cada época e somamos seus valores em bandas de frequência descritas classicamente na literatura. Nesta seção apresentamos o que foi utilizado para a classificação e quais foram os resultados finais da classificação dos registros, sendo apresentadas (1) uma caracterização dos dados obtidos (seções 4.1 e 4.2), seguidas da (2) descrição e avaliação da *Toolbox* em si (seções 4.3 até 4.12).

Anexados a este trabalho, estão a descrição completa da interface gráfica desenvolvida e um passo-a-passo para sua utilização.

#### **4.1 - Valores do RMS do eletromiograma são tipicamente utilizados para diferenciar sono de vigília**

Em situações fisiológicas, a presença de tônus muscular e/ou da movimentação é uma característica determinante do estado acordado. Além disso, durante o sono REM, ocorre completa atonia da maioria dos músculos esqueléticos. Na literatura são utilizadas diferentes estratégias para determinar tais estados comportamentais, a exemplo, rastreamento de vídeo do comportamento do animal (Nath et al. 2019), análise de dados coletados de acelerômetros instalados na cabeça dos animais (Pinnell, Almajidy, and Hofmann 2016), ou ainda, eletrodos implantados sob o músculo do pescoço do animal (Silvani et al. 2017). Neste trabalho, implantamos eletrodos simples sob o músculo do pescoço dos animais e utilizamos o dado do EMG para auxiliar na classificação de estados ao longo do ciclo sono-vigília. A Figura 10-A mostra um exemplo de traçado bruto de EMG que, na literatura, é tipicamente atribuído ao estado acordado. Em contrapartida, quando os animais estão em repouso ou dormindo, seus músculos estão relaxados, e o sinal de EMG é de baixa amplitude (Figura 10-B e C).

O cálculo do RMS do sinal do EMG de uma época resume, em um número, a atividade muscular (no caso, do músculo do pescoço) que o animal teve durante o

período relativo aquela época. Como esperado, este valor é proporcional à atividade do animal, ou seja, é mais alto quanto maior a atividade do animal e menor quanto menor a atividade do animal (Figura 10-D). Sendo assim, concluímos que o RMS do EMG de cada época deve ser usado para resumir as informações necessárias para distinguir períodos de vigília dos de sono.

A distribuição em *z-score* dos valores de RMS mantém essa relação, centraliza a distribuição em zero por subtrair cada valor pela média (Figura 10-D), e ainda torna as escalas padronizadas, o que facilita a comparação entre diferentes momentos de um registro (Figura 10-E). Considerando um registro inteiro (Figura 10-F), é possível distinguir épocas com valores mais elevados de EMG (acima de 0 *z-score*), típicas do estado acordado, das épocas com valores mais baixos, típicas do estado de sono (NREM e REM). Por fim, a distribuição em *z-score* é consistente entre os diferentes registros (Figura 10-G), que por sua vez facilita a comparação entre diferentes registros.

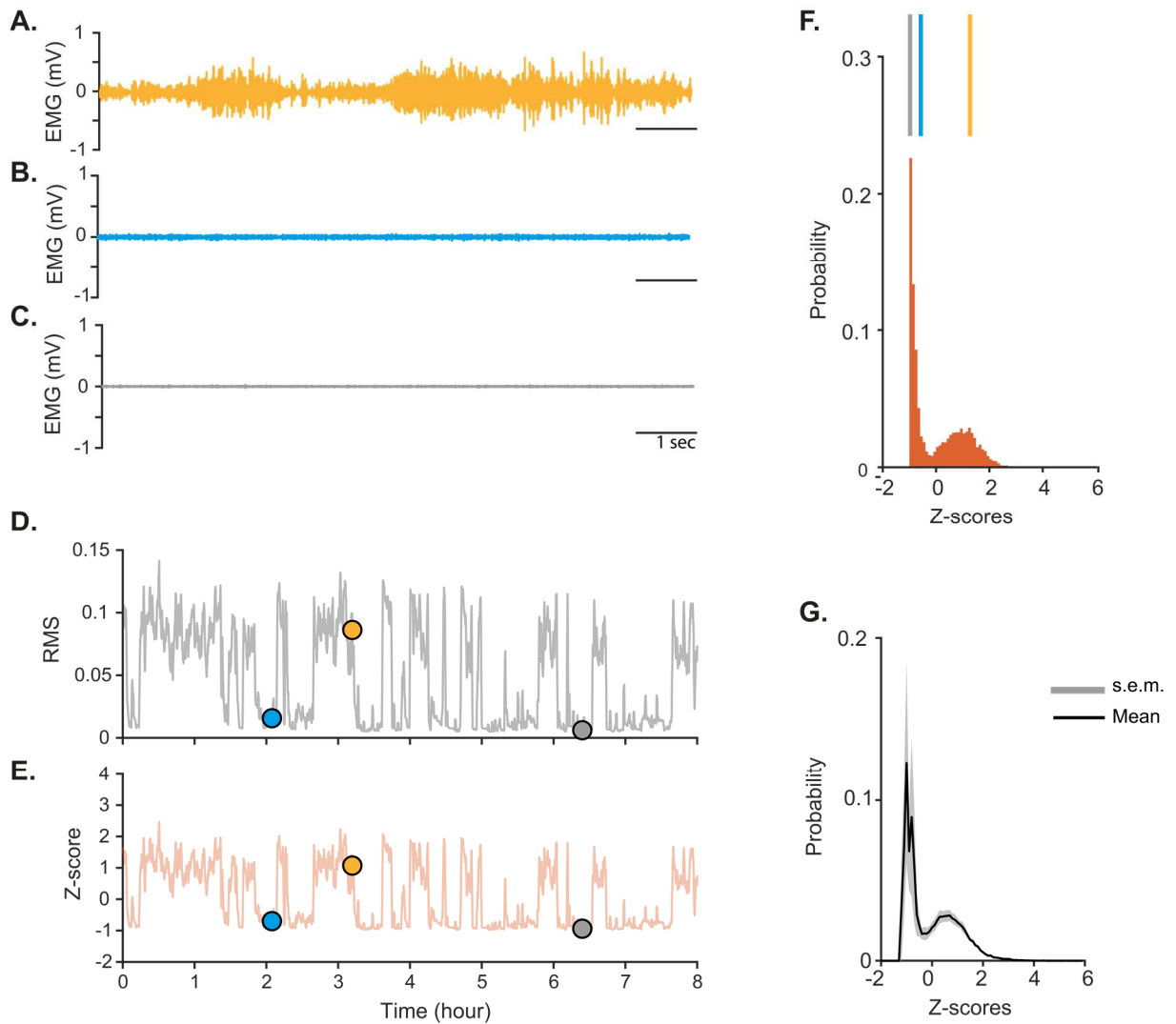


Figura 10. Cálculos sobre o EMG. A: Sinal bruto durante estado acordado (amarelo); B: Sinal bruto durante sono NREM (azul); C: Sinal bruto durante sono REM (cinza); D: Distribuição dos valores de RMS durante 8 horas aleatoriamente selecionadas; E: Distribuição em z-score dos mesmos valores de RMS. Bolas representam o valor de cada época em cada gráfico; F: Histograma com a distribuição em z-score dos valores de RMS de cada época. Cada barra representa o valor relativo de cada uma das épocas selecionadas; G. Média e erro da média das distribuições em z-score dos valores de RMS do EMG dos registros. S.E.M.: erro padrão.

#### 4.2 - Razão Teta/Delta calculada a partir do LFP hipocampal é tipicamente utilizada para diferenciar sono NREM do sono REM

O desenvolvimento das técnicas de registro eletrofisiológico do cérebro levou à divisão e organização das etapas de sono em sono NREM e sono REM, que são definidas a partir das características espectrais do sinal eletrofisiológico (A. R. Adamantidis 2019). O hipocampo é uma região que, devido à sua disposição



anatômica e organização de sua circuitaria, apresenta ritmos bem definidos e típicos para cada estado comportamental (Drieu and Zugaro 2019; De La Prida 2020). Sendo assim, as análises espectrais do LFP coletado nesta região tem um grande potencial preditor desses estados.

Dentre as épocas com menores valores de RMS do EMG, foi possível detectar épocas com predominância espectral na faixa de Delta (Figura 11-C e D), típicas de sono NREM. Além disso, observamos a existência de épocas com potência concentrada na faixa de Teta (Figura 11-E e F), tipicamente classificadas como sono REM.

A divisão entre a potência na faixa de Teta pela potência em Delta resultou em valores únicos que representam ambos padrões oscilatórios. Esses padrões são típicos de cada etapa do sono (A. R. Adamantidis 2019), haja visto que períodos de sono REM apresentam os maiores valores possíveis para essa relação, ao passo que o sono NREM, os menores (Figura 11-G). A normalização por *z-score* da relação Teta/Delta torna a distribuição dos valores de Teta/delta consistente entre os diferentes registros (Figura 11-H), permitindo comparações entre diferentes animais.

Em suma, o fator principal para sugerir distinção do estado acordado em relação ao sono é o valor elevado do EMG. Por outro lado, a relação Teta/Delta é responsável por sugerir a distinção do sono NREM para o sono REM. Curiosamente, o sinal do hipocampo durante a vigília pode apresentar distribuição espectral intermediária em relação aos estados de sono (Figura 11-A e B), pois apresenta coexistência de oscilações delta e Teta. O Quadro 5 resume as principais características de cada estado comportamental, de acordo com a literatura, a saber:

- Acordado: os maiores valores do RMS do EMG *z-score*;
- Sono NREM: os menores valores de EMG e da relação Teta/Delta;
- Sono REM: os menores valores de EMG e os maiores da relação Teta/Delta;

Quadro 5. Esquema da relação entre as variáveis e os respectivos estados

	Acordado	Sono NREM	Sono REM
Potência em Teta	Intermediária	Baixa	Alta

Potência em Delta	Intermediária	Alta	Baixa
Relação Teta/Delta	Intermediária	Baixa	Alta
RMS do EMG	Alto	Baixo	Baixo

---

Fonte: Elaboração própria

Portanto, nossa *Toolbox* define como modelo padrão da classificação a relação Teta/Delta do sinal do hipocampo associada à atividade do EMG. Porém, caso opte, o usuário pode somar outras bandas de frequência a essa relação, tendo como opção as distribuições em *z-score* da relação:

- Teta + Beta / Delta;
- Teta + Low Gamma / Delta;
- Teta + High Gamma / Delta;
- Teta + Gamma (low + high) / Delta;
- 6 a 90 Hz / Delta;

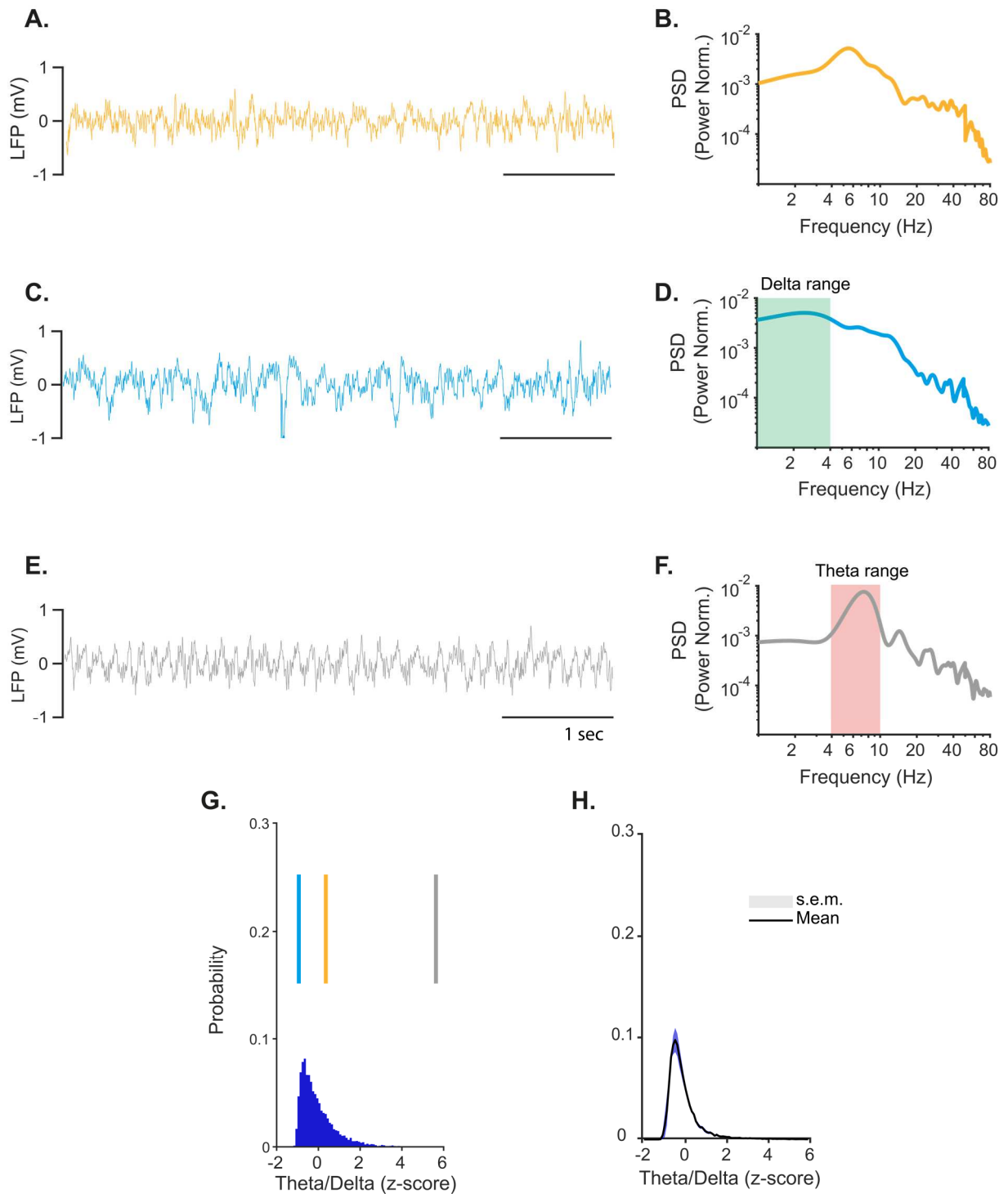


Figura 11. Cálculos sobre o LFP. A: 5 segundos representativos com sinal do hipocampo do estado acordado; B: PSD calculada para a época em A; C: Mesmo que A, para sono NREM; D: PSD da época em C, com limites de Delta coloridos em verde; E: Mesmo que A, para sono REM; F: PSD da época em E, com limites de Teta coloridos em vermelho; G: Histograma com a distribuição em z-score dos valores de Teta/Delta de cada época. Cada barra representa o valor relativo de cada uma das épocas selecionadas; H: Média e erro da média das distribuições em z-score dos valores de RMS do EMG dos registros.

### **4.3 - Flutuação dos valores de Teta/Delta e RMS do EMG ao longo do tempo mostraram-se coerentes com as etapas do ciclo circadiano**

O ciclo circadiano dos roedores é dividido em uma fase mais ativa e uma fase mais inativa. A fase ativa é também denominada de fase escura, e varia aproximadamente entre as 19 até as 7 horas, período no qual os animais passam maior parte do tempo acordados (Soltani et al. 2019). Já a fase inativa, também denominada de fase clara, e geralmente variando das 7 até as 19 horas, os animais tendem a dormir mais (Soltani et al. 2019), passando maior parte do tempo em sono NREM e, com o avançar do tempo, aumentam os episódios de sono REM. Durante o processo de classificação, é imprescindível checar se as variáveis utilizadas para a classificação acompanhem essa flutuação.

Para nosso conjunto de dados, épocas com os maiores valores no EMG (Figura 12-A, eixo X) estão presentes ao longo das 24 horas de registro (Figura 12-B), porém seus maiores valores estão mais concentrados durante a fase escura, sendo maiores que 0 por cerca de 76% do tempo (Figura 12-C) e refletindo uma maior atividade muscular dos animais. Isso sugere que os animais despenderam maior tempo no estado acordado, como esperado. Em média, o valor do RMS EMG para a fase escura foi +0,41 (*z-score*), enquanto que para a fase clara é -0,42 (*z-score*), período no qual os valores de EMG superiores a 0 ocupam apenas 37% do tempo (Figura 12-C).

Usando como referência a distribuição da relação Teta/Delta por todo o registro (Figura 12-A, eixo Y), a fase escura exibiu a maioria dos valores superiores que a média (Figura 12-C), pontuando em média 0,11 (*z-score*) enquanto que a média das 24 horas foi de aproximadamente 0 (*z-score*). Esse valor médio provavelmente refletiu períodos de maior exploração dos animais, haja visto que variam conjuntamente com o EMG. Esses são os períodos com maior frequência das épocas candidatas ao estado acordado (Figura 12-C, em amarelo).

Durante a fase clara, os valores da razão Teta/Delta, em média, reduziram para -0,11 (*z-score*). As épocas candidatas ao sono NREM são encontradas em períodos

com os valores mais baixos da razão Teta/Delta e do EMG (Figura 12-C, em azul). Em compensação, as épocas candidatas ao sono REM podem ser localizadas em períodos com picos da relação Teta/Delta (Figura 12-B, em cinza), atingindo valor máximo de 14,87 (z-score), e períodos de valores negativos no EMG.

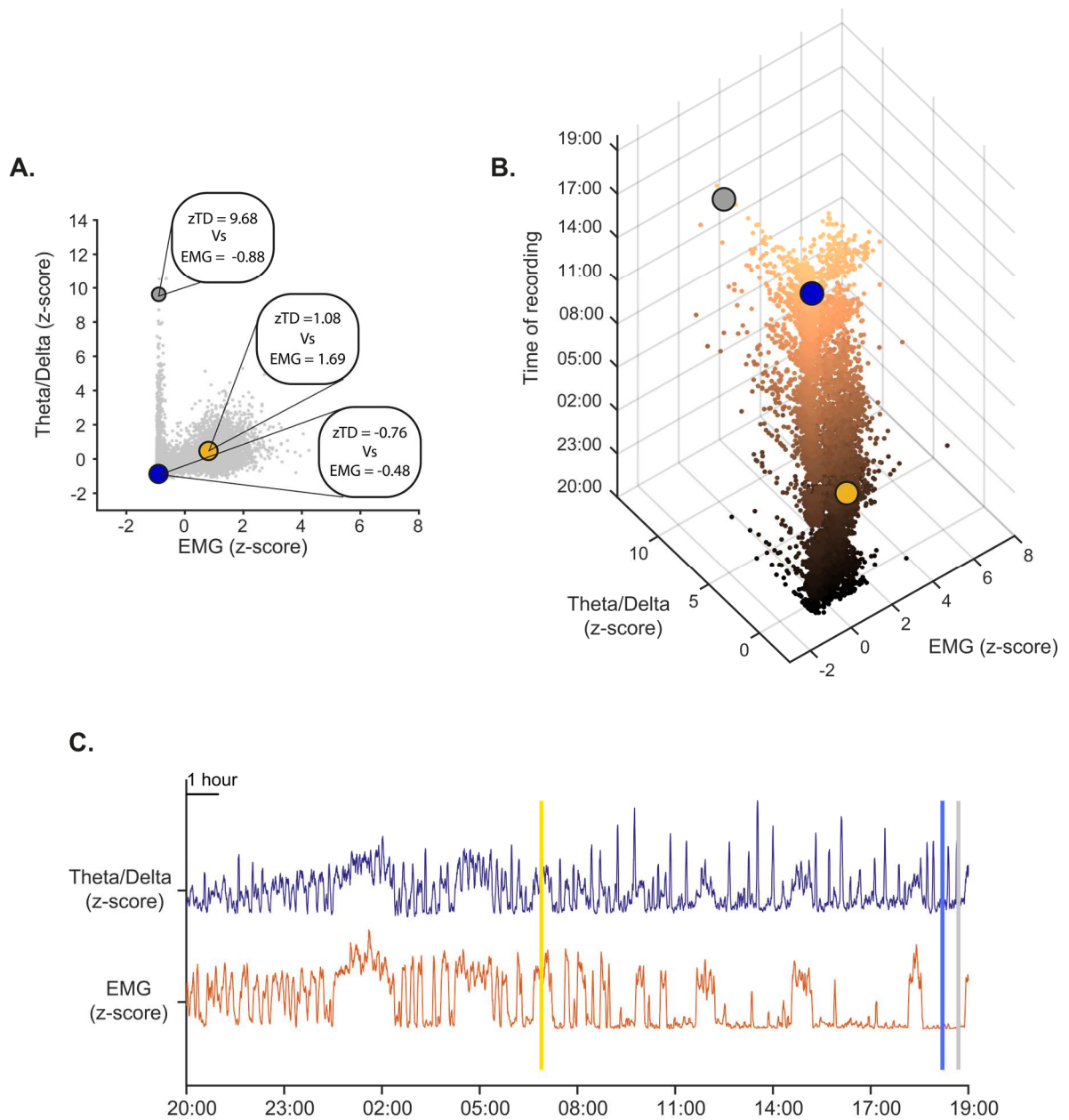


Figura 12. Distribuição ao longo tempo de épocas candidatas ao estado acordado (amarelo), sono NREM (azul) e sono REM (cinza). A: Gráfico de Dispersão com os valores z-score da relação Teta/Delta e do RMS do EMG de cada época. Nos balões, esquemas dos valores esperados em cada época; B: Mesmo em A, porém disposta ao longo do tempo do registro. Intensidade da cor também representa o tempo, com as primeiras horas em mais escuro e as últimas mais claro; C: Disposição ao longo do

*tempo das variáveis separadamente. Note que o marcador com o nome de cada variável foi colocado na posição relativa à média da distribuição da respectiva variável.*

#### **4.4 - O Modelo de Mistura de Gaussianas (GMM) cria grupos com épocas de características similares**

Após definirmos os principais parâmetros e as características mais marcantes de cada estado, foi necessário escolher um método capaz de agrupar as épocas de 10s em grupos coesos e condizentes com o que está descrito na literatura.

Para isso, escolhemos o Modelo de Misturas de Gaussianas (GMM) como algoritmo base para esta etapa. O GMM assume que cada estado possui uma distribuição normal em relação a cada variável utilizada para a classificação, ou seja, o GMM assume que cada estado apresenta épocas com valores próximos à média da relação Teta/Delta e do EMG, e dentro dos seus limites de covariância (Geoffrey McLachlan 2000).

O algoritmo chamado de *Expectation and maximization* (EM) é uma etapa importante do GMM, pois é responsável pela personalização dos valores de média e covariância dos estados em cada registro. Porém, para iniciar essa personalização, o algoritmo pode partir de uma classificação inicial feita pelo *K-means* (modelo não treinado) ou de um modelo inicial (modelo treinado). Após a etapa de EM, o GMM retorna a probabilidade de cada época pertencer a cada estado (seção 3.6), oferecendo a possibilidade da formação de grupos de acordo com o valor mínimo de probabilidade escolhido.

Analisando a formação dos grupos do modelo não treinado, o GMM forma grupos com características diferentes daquelas utilizadas nas definições dos estados, como por exemplo um grupo com os maiores valores da razão Teta/Delta e os menores valores de EMG é candidato ao sono REM, porém as épocas com valores altos de EMG (Figura 13-C), que por sua vez são características do estado acordado, sugerem que este grupo não deve ser utilizado, uma vez que, caso escolhêssemos essa proposta de grupos, a classificação seria de baixa acurácia.

Para direcionar a formação dos grupos pelo GMM, utilizamos um registro de 24 horas previamente classificado como modelo inicial do EM. Este modelo inicial é chamado de Treinamento do GMM. Como resultado final da formação dos grupos com o modelo treinado (Figura 13), as épocas com valores mais elevados de EMG são agrupadas com elevada probabilidade posterior (Figura 13-D), enquanto que as épocas com os menores valores de Teta/Delta e EMG formam um outro grupo (Figura 13-E). As épocas com os maiores valores para a razão Teta/Delta e os menores para o EMG foram agrupadas em um terceiro grupo (Figura 13-F).

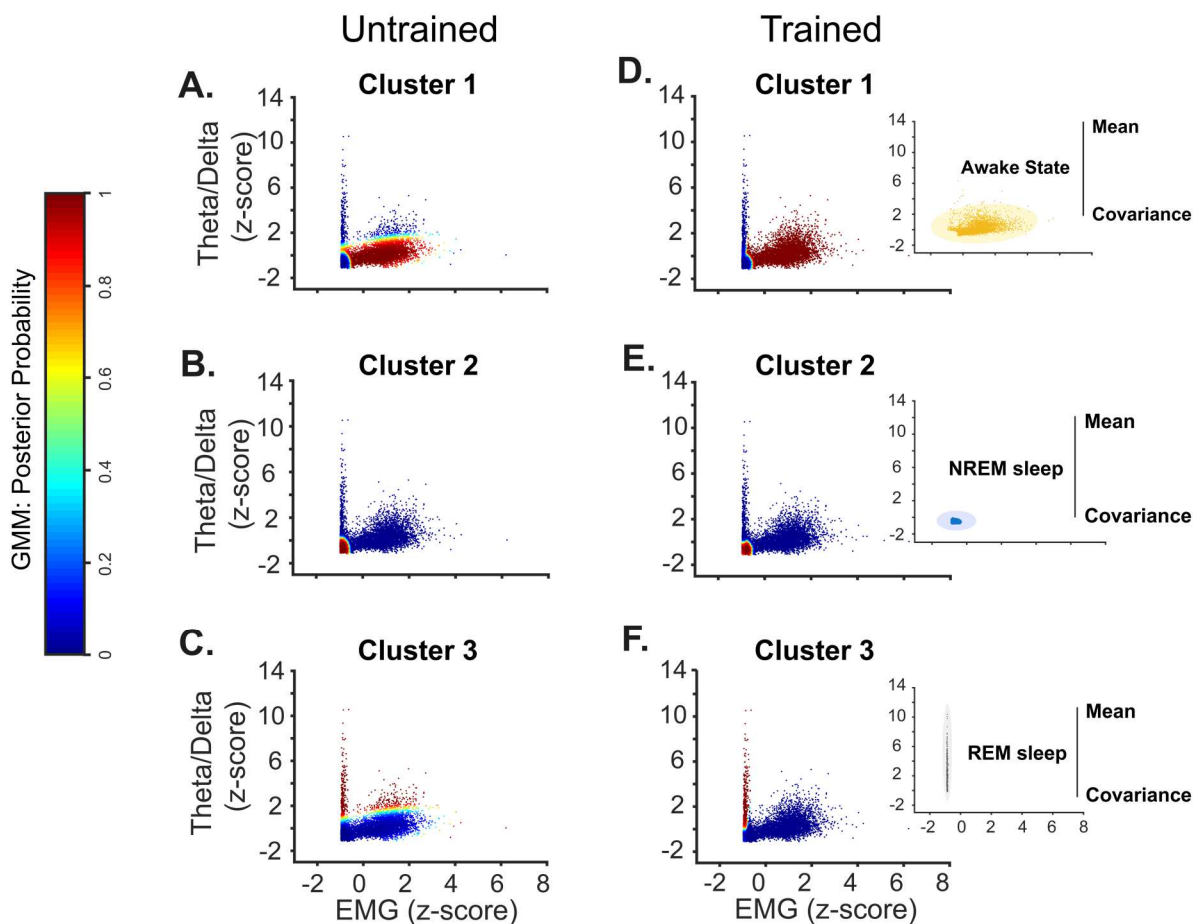


Figura 13. Treinamento do modelo. A: Probabilidade posterior calculada para cada época para grupo 1, sem treinamento; B: Mesmo para grupo 2; C: Mesmo para possível grupo 3; D: Probabilidade posterior calculada para o grupo 1, porém com treinamento. Valores de média de cada variável e matriz de covariância dos dados de treinamento são usados pelo GMM; E: Mesmo para grupo 2; F: Mesmo para possível grupo 3.

#### 4.5 - Definindo a classificação manual do usuário como um padrão ouro da classificação

Diversos trabalhos utilizam a classificação manual como método único para a classificação (Quadro 2). Porém, essa abordagem é impraticável na maioria dos estudos sobre ciclo circadiano, haja visto que registros longos (por exemplo de 24 horas) tomariam um tempo proporcionalmente grande para serem classificados manualmente. Outros métodos utilizam a classificação manual como uma referência para medir a acurácia da classificação, mas muitos desses trabalhos não descrevem completamente a metodologia e raramente reportam os resultados desta etapa.

Ao avaliarmos a literatura, constatamos que a classificação manual geralmente é feita por meio de análises visuais do sinal bruto e fragmentado em épocas com tamanhos variados e que dependem do método escolhido. Já foi reportado que épocas de 10 segundos são ideais para classificação de estados por contemplar melhor as diversas características dos sinais registrados (Brankač et al. 2010; Lampert et al. 2015). Sendo assim, para nossa classificação manual, várias épocas de 10s foram selecionadas aleatoriamente e apresentadas para um único experimentador, que as classificou manualmente. A interface gráfica (Apêndice 1) sempre apresentará ao usuário um conjunto de épocas aleatoriamente selecionadas que deverão ser classificadas até completar o número definido de épocas e dependente do tamanho do registro. Para os registros de 24 horas feitos em camundongos, de um total de 8640 épocas, 40 foram classificadas em cada estado e 120 no total. No registro em rato, de um total de 1827 épocas, 8 foram classificadas em cada estado e 36 no total. Ressaltamos que, uma vez atingido esse número, a *Toolbox* continuará a classificação, mas, ao fim do processo, o usuário terá acesso a todas essas épocas classificadas manualmente. Além disso, o usuário poderá classificar épocas como transição entre os estados e poderá utilizá-las para análises futuras, mas essas informações não serão utilizadas para a classificação da *Toolbox*. Para a classificação, são apresentados o sinal bruto do LFP e do EMG de cada época, além da PSD do LFP. Maiores detalhes sobre esta etapa são apresentados na seção 3.7 dos métodos e na apresentação da interface gráfica (Apêndice 1).



Para verificar a classificação manual, analisamos os valores médios de potência das épocas classificadas em cada um dos estados. Como esperado, as épocas classificadas apresentaram características espectrais típicas de cada estado, como potência elevada em Teta durante sono REM e Delta durante sono NREM (Tabela 1). As épocas classificadas como sono REM apresentam, em média, potência elevada em Teta e baixas em Delta. As classificadas como sono NREM, em média, apresentam potência baixa em Teta e alta em Delta. Por fim, as classificadas como estado acordado apresentaram distribuição intermediária para ambas: Teta e Delta (Figura 14-A).

Tabela 1. Valor máximo de potência normalizada e a respectiva frequência

		Acordado	Sono NREM	Sono REM
Teta	Potência Normalizada	$3,6 \times 10^{-3}$	$2,9 \times 10^{-3}$	$5,4 \times 10^{-3}$
	Frequência no Pico (Hz)	6,1	6	6,75
Delta	Potência Normalizada	$3 \times 10^{-3}$	$5,3 \times 10^{-3}$	$2,1 \times 10^{-3}$
	Frequência no Pico (Hz)	3,9	2,3	4

Fonte: Elaboração própria

Como consequência dessas características espectrais, as épocas classificadas como sono REM são aquelas com os maiores valores da relação Teta/Delta, enquanto que as classificadas como NREM, apresentam os menores. As épocas classificadas como acordado apresentam distribuição intermediária entre ambos, mas possuem os maiores valores para o EMG (Figura 14-B). Para o sono NREM e REM, os valores do EMG são os menores entre as épocas classificadas. A classificação ao longo do tempo (Figura 14-C) demonstra que as épocas classificadas como sono REM são mais presentes nas últimas horas de registro (aproximadamente a partir das 14 horas), enquanto que as classificadas como acordado são mais presentes nas primeiras horas de registro (até as 23:00), o que reflete os períodos ativos da fase escura e inativos da fase clara.

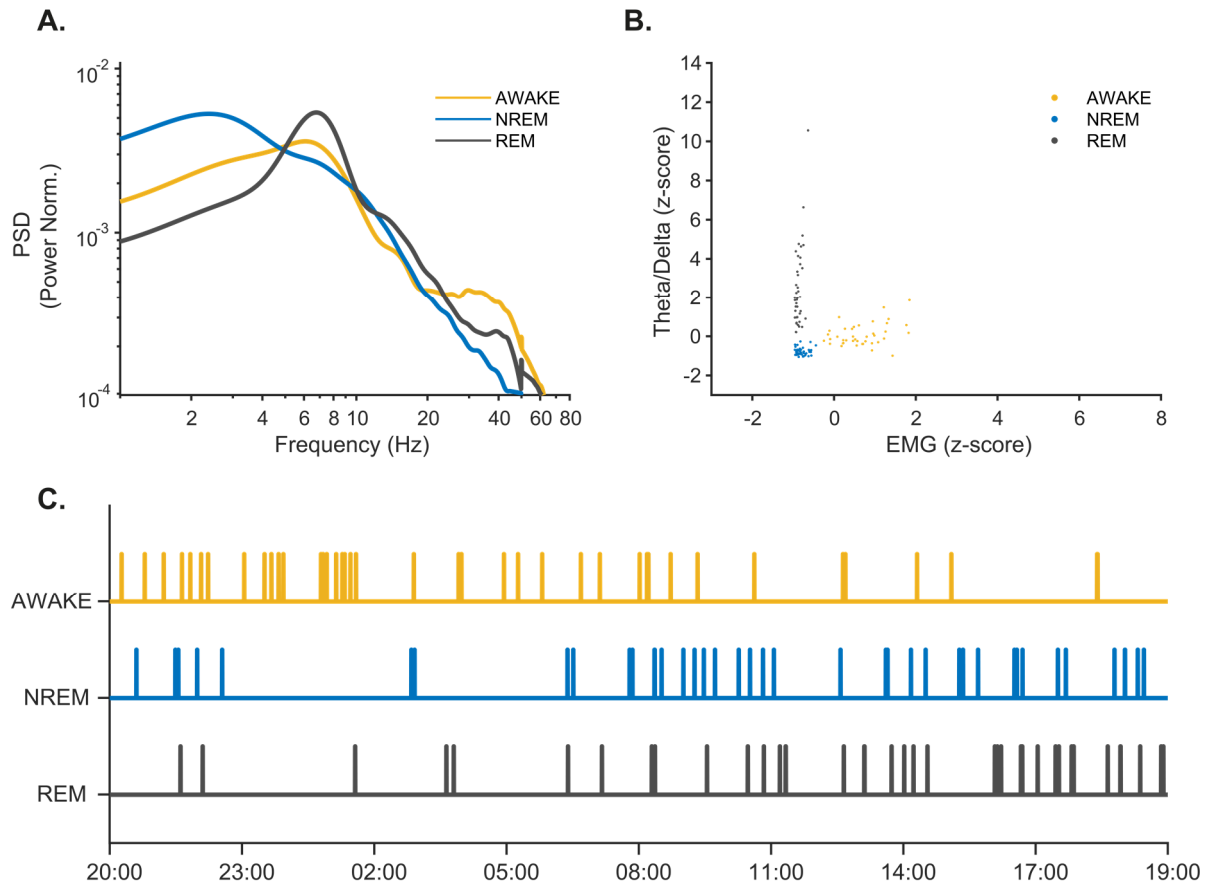


Figura 14. Resultados da Inspeção visual. A. A PSD da média de todas as épocas classificadas pelo usuário em cada um dos três estados, representados pelas cores amarela para o estado acordado, azul para sono NREM e cinza para sono REM; B. Posição das épocas classificadas pelo usuário no gráfico segundo as distribuições em z-score da relação Teta/Delta e do RMS do EMG; C. Posição no tempo das épocas classificadas pelo usuário.

Essas análises (Figura 14) oferecem para o usuário uma forma de controle das épocas classificadas manualmente, dando subsídio para que possam ser utilizadas como padrão ouro da classificação. Os rótulos gerados nessa etapa serão necessários para definir qual será o limiar de probabilidade ideal para clusterização dos estados (próxima seção).

#### 4.6 - Definição do limiar de probabilidade para cada estado

A formação de grupos a partir de um conjunto de dados exige que valores mínimos ou máximos (limiares) sejam escolhidos para definir os seus limites. No caso dos métodos de classificação do ciclo circadiano, os limiares são definidos para as variáveis escolhidas para a classificação e os grupos são os estados comportamentais. Portanto, é nesta etapa que as épocas candidatas devem ser designadas (ou não) a um determinado estado.

Vários métodos de classificação disponíveis na literatura definem limiares arbitrários para cada estado. Estes, por muitas das vezes, são escolhidos por meio de inspeções visuais de gráficos, não envolvendo cálculos objetivos. Definições como estas podem introduzir algum tipo de viés à classificação e/ou dificultar a reprodução desses achados. Neste trabalho nós desenvolvemos uma nova estratégia para a definição dos limiares, como será detalhado a seguir.

O GMM, como discutido anteriormente, é um modelo probabilístico que assume distribuições gaussianas para cada estado. Com isso, ao final de seus cálculos, o GMM retorna, para cada época de 10s, qual é a sua probabilidade de pertencer a cada estado. Para definir o limiar de probabilidade, nós utilizamos as épocas da classificação manual e calculamos a taxa de verdadeiros positivos e falso positivos que cada estado possuiria caso escolhêssemos cada um dos limiares de probabilidade posterior possíveis. Com isso, a classificação manual passa a ser parte ativa da classificação, pois é ela que define o acurácia da classificação para cada limiar de probabilidade possível, garantindo o maior acurácia possível.

A Figura 15 demonstra como seria a formação dos estados caso a escolha do limiar fosse de 0%. Com este limiar, todas as épocas seriam classificadas em todos os estados (Figura 15-D, E e F), uma vez que todas elas têm, pelo menos, 0% de chance de pertencer a cada um dos estados dado as características de cada estado e delas mesmas (Figura 15-A, B e C). Com esse limiar, obteremos uma taxa de 100% de verdadeiros positivos para os 3 estados (Figura 15-J, K e L), haja visto que todas as épocas classificadas como acordado pelo padrão ouro também foram classificadas como acordado pelo GMM (Figura 15-G, H e I). Porém, a taxa de falsos positivos

também é máxima, uma vez que todas as épocas classificadas pelo usuário como sono NREM e como sono REM também foram classificadas como acordado pelo GMM (Figura 15-J, K e L). Este problema se estende para todos os outros estados, como pode ser observado na Figura 15.

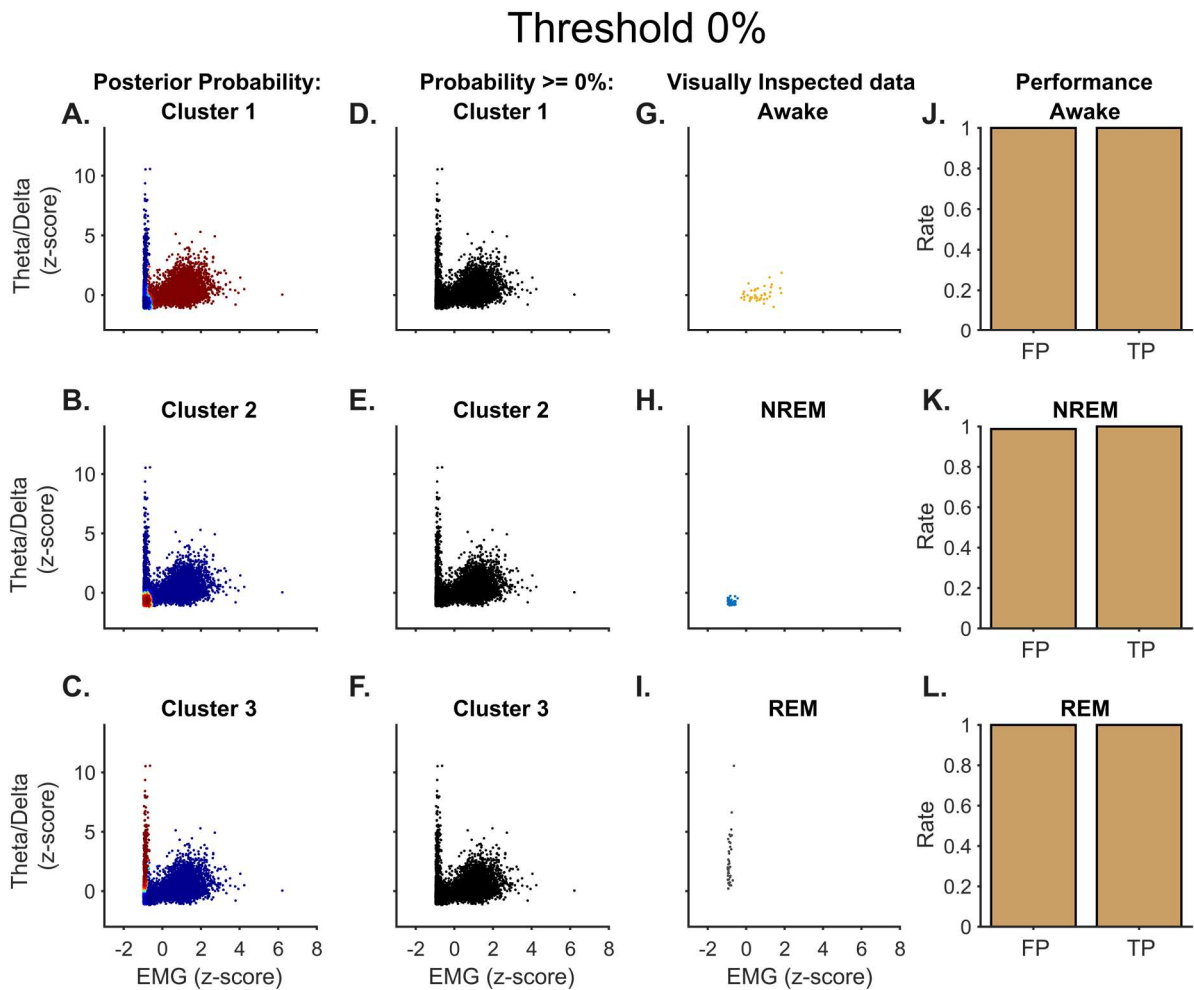


Figura 15. Limiar de probabilidade posterior arbitrariamente escolhido como de 0%. A: Probabilidade posterior calculada para o grupo 1; B: Probabilidade posterior calculada para o grupo 2; C: Probabilidade posterior calculada para o grupo 3; D: Épocas com probabilidade igual ou superior ao limiar escolhido para grupo 1; E: Mesmo que D, para o grupo 2; F: Mesmo que D, para o grupo 3; G: Épocas classificadas pelo usuário como acordado; H: Épocas classificadas pelo usuário como sono NREM; I: Épocas classificadas pelo usuário como sono REM; J: Taxa de falso (FP) e verdadeiro positivos (TP) calculadas para o estado acordado; K: Taxa de falso (FP) e verdadeiro positivos (TP) calculadas para o sono NREM; L: Taxa de falso (FP) e verdadeiro positivos (TP) calculadas para o sono REM.

Com um limiar de 75%, as épocas escolhidas para cada estado já passam a apresentar um padrão mais próximo do esperado (Figura 16-D, E e F). Quando confrontadas com a classificação manual (Figura 16-G, H e I), obteríamos uma elevada taxa de verdadeiros positivos (acima de 0,9) e baixa de falsos positivos (abaixo de 0,1) para os três estados (Figura 16-J, K e L).

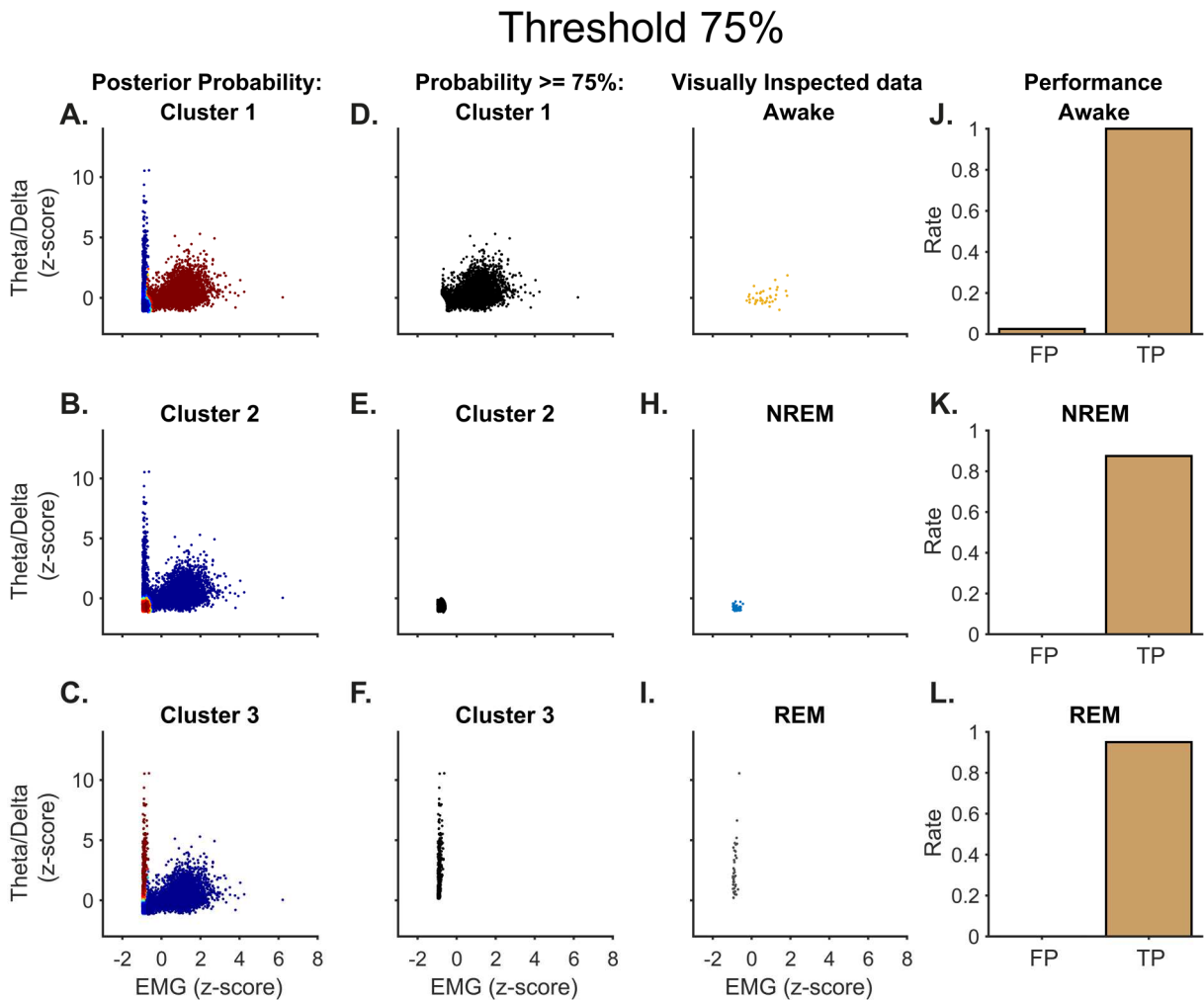


Figura 16. Limiar de probabilidade posterior artificialmente escolhido como de 75%. A: Probabilidade posterior calculada para o grupo 1; B: Probabilidade posterior calculada para o grupo 2; C: Probabilidade posterior calculada para o grupo 3; D: Mesmo que A, para o grupo 1; E: Mesmo que D, para o grupo 2; F: Mesmo que D, para o grupo 3; G: Épocas classificadas pelo usuário como acordado; H: Épocas classificadas pelo usuário como sono NREM; I: Épocas classificadas pelo usuário como sono REM; J: Taxa de falso (FP) e verdadeiro positivos (TP) calculadas para o estado acordado; K: Taxa de falso (FP) e verdadeiro positivos (TP) calculadas para o sono NREM; L: Taxa de falso (FP) e verdadeiro positivos (TP) calculadas para o sono REM.

Repetindo esse cálculo para diferentes limiares (de 0 a 100%) e computando as taxas de falsos positivos e verdadeiros positivos, construímos uma curva ROC (Característica de Operação do Receptor, em inglês). O limiar escolhido foi aquele que apresentou a maior taxa de verdadeiros positivos e a menor taxa de falsos positivos possível para o registro, segundo o método da menor distância ao ponto (0,1) (Unal 2017). Adotando esta estratégia, os estados acordado e sono NREM apresentaram taxa de 100% de verdadeiros positivos, enquanto que o sono REM apresentou de 97,5%. Os estados acordado e sono REM obtiveram 0% de falsos positivos, enquanto que o sono NREM, 2,5% (Figura 17-D, E e F).

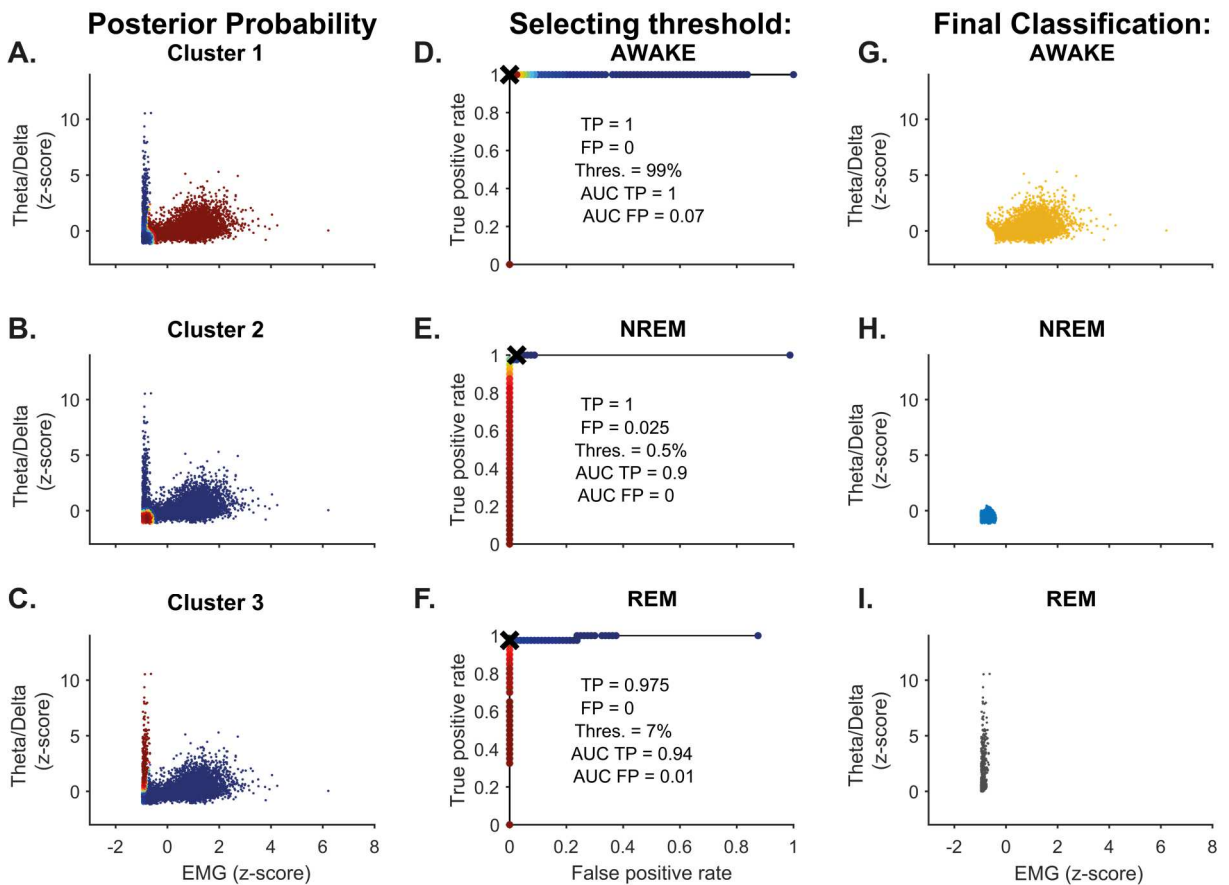


Figura 17. Limiar de probabilidade posterior escolhido pela menor distância ao ponto (0,1). A: Probabilidade posterior calculada para o grupo 1; B: Probabilidade posterior calculada para o grupo 2; C: Probabilidade posterior calculada para o grupo 3; D: Curva ROC calculada para estado acordado. Taxa de verdadeiro positivos (TP), falso positivos (FP), limiar escolhido (Thres.). Demonstrados no meio do gráfico o valor da área sobre a curva (AUC) para a TP versus os limiares e FP versus os limiares; E: Mesmo que D, para sono NREM; F: Mesmo que D, para sono REM; G: Épocas classificadas como acordado; H: Épocas classificadas como sono NREM; I: Épocas classificadas como sono REM.

Como forma auxiliar de demonstrar a acurácia da classificação, calculamos também a área sob a curva (AUC) considerando uma por vez, a taxa de verdadeiros positivos (AUC TP) e depois a taxa de falso positivos (AUC FP), ambos para os diferentes limiares possíveis. Um registro cuja classificação apresentou acurácia elevado deverá apresentar AUC TP próximo de 1, demonstrando que, para os diferentes limiares escolhidos, a TP foi alta (acima de 0,9). Ainda, este mesmo registro deverá apresentar AUC FP próximo de 0, indicando que para todos os limiares escolhidos a TF foi baixa (abaixo de 0,1). Em combinação neste caso, ambas informações dizem que o GMM gerou grupos com forte similaridade com os da classificação manual, ou seja, classificou as épocas com elevado acurácia.

O estado acordado apresentou valor máximo para AUC TP (1 a.u.) e um valor muito baixo para AUC FP (0.07 a.u.), sugerindo que o acurácia do classificador foi máximo em reconhecer as épocas pertencentes e não pertencentes a este estado. O estado NREM apresentou AUC TP de 0,9 e AUC FP de 0, portanto consideramos a acurácia excelente. Por fim, a também elevada AUC TP e a praticamente nula AUC FP para o sono REM demonstram que as épocas foram, em sua maioria, corretamente designadas para seus respectivos estados.

Com o limiar definido para cada grupo, as épocas foram então destinadas para seus respectivos estados (Figura 17-G, H e I). Caso uma determinada época possua valor de probabilidade menor que o limiar escolhido para os 3 estados, ela será designada ao estado para o qual possui maior probabilidade. Caso uma época tenha probabilidade maior que o limiar para mais de um estado, ela é designada ao estado de maior probabilidade.

Ao fim dessa etapa, todas as épocas são destinadas exclusivamente a um dos 3 estados e o usuário terá acesso a essa classificação por meio de um arquivo no formato do MATLAB e em uma planilha, podendo utilizá-la para futuras análises.

#### 4.7 - Todas as épocas classificadas em um mesmo estado apresentam características coerentes entre si e exibem as características típicas de seus respectivos estados

Visando investigar se as épocas classificadas formam grupos com características coerentes com as esperadas para cada estado e que justificam o acurácia elevado da classificação apresentada neste registro (Figura 18-A), calculamos a média de potência de todas as épocas classificadas em cada um dos estados (Figura 18-B). Em média, as épocas classificadas como sono REM apresentam os maiores valores da faixa de Teta, enquanto que as de sono NREM apresentam os maiores valores da faixa de Delta. Por sua vez, o estado acordado apresenta os valores intermediários para ambas bandas de frequência (Figura 18-B e Tabela 2).

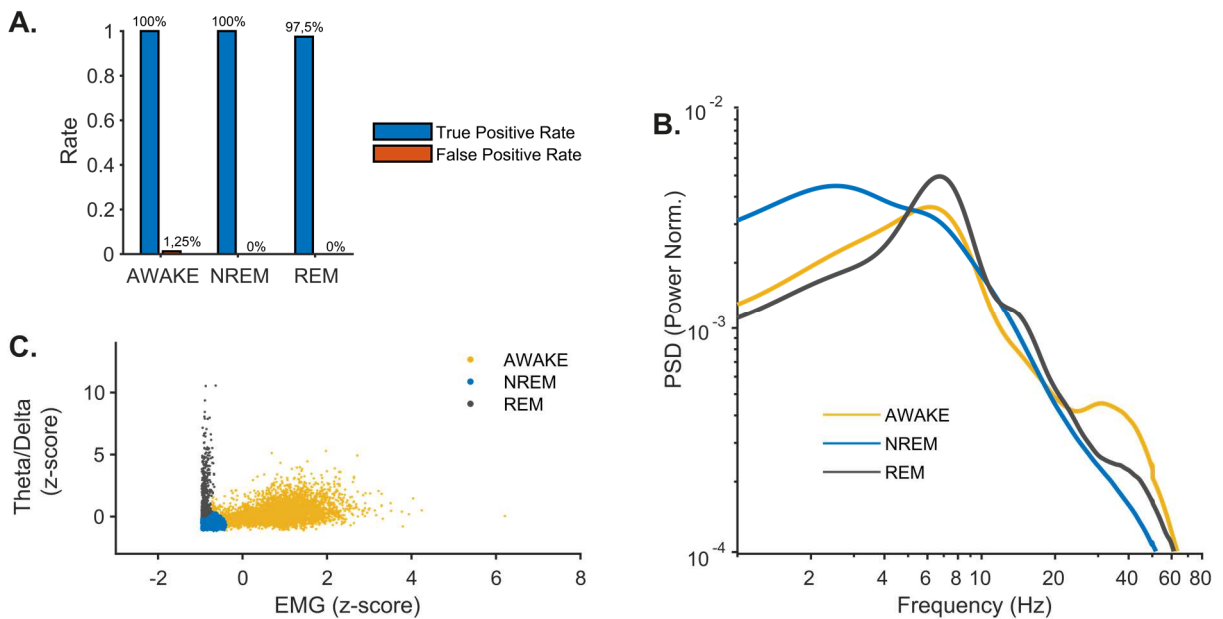


Figura 18. Acurácia final. A. Taxa de verdadeiro positivos (em azul) e Falso positivos (em vermelho) obtidas para esse registro; B. Média das PSDs das épocas classificadas em cada estado; C. Gráfico de dispersão das épocas conforme seus valores de z-score da relação Teta/Delta e z-score do RMS do EMG, e coloridas de acordo com a classificação;



Tabela 2. Valor máximo de potência normalizada e a respectiva frequência

		Acordado	Sono NREM	Sono REM
Teta	Potência Normalizada	$3,6 \times 10^{-3}$	$3,3 \times 10^{-3}$	$5 \times 10^{-3}$
	Frequência no Pico (Hz)	6,5	6	6,75
Delta	Potência Normalizada	$2,9 \times 10^{-3}$	$4,5 \times 10^{-3}$	$2,3 \times 10^{-3}$
	Frequência no Pico (Hz)	3,95	2,55	3,9

Fonte: Elaboração própria

Como consequência da classificação, as épocas formaram grupos coesos e que respeitam as características de cada estado (Figura 18-C). Observamos que as épocas designadas para o estado acordado apresentaram os maiores valores de EMG (Figura 19-A), e foram agrupadas com a maior média para este parâmetro, além de um valor baixo (menor que 1) para a média dos desvios absolutos (Tabela 3). Por outro lado, as épocas classificadas como sono NREM apresentaram valores baixos para EMG e Teta/Delta (Figura 19-B), tendo a menor média da relação Teta/Delta (Tabela 3). Já para as épocas classificadas como sono REM, observamos os menores valores de EMG médio e de média dos desvios (Tabela 3). Como esperado, as épocas em REM apresentaram a maior média para a razão Teta/Delta, porém obtiveram um valor alto para a média dos desvios (Tabela 3).

Tabela 3. Dispersão dos estados

		Acordado	Sono NREM	Sono REM
Razão Teta/Delta	Média	0,33	-0,61	1,71
	Média dos desvios absolutos	0,59	0,21	1,4
RMS do EMG	Média	0,91	-0,82	-0,89
	Média dos desvios absolutos	0,55	0,11	0,04

Fonte: Elaboração própria

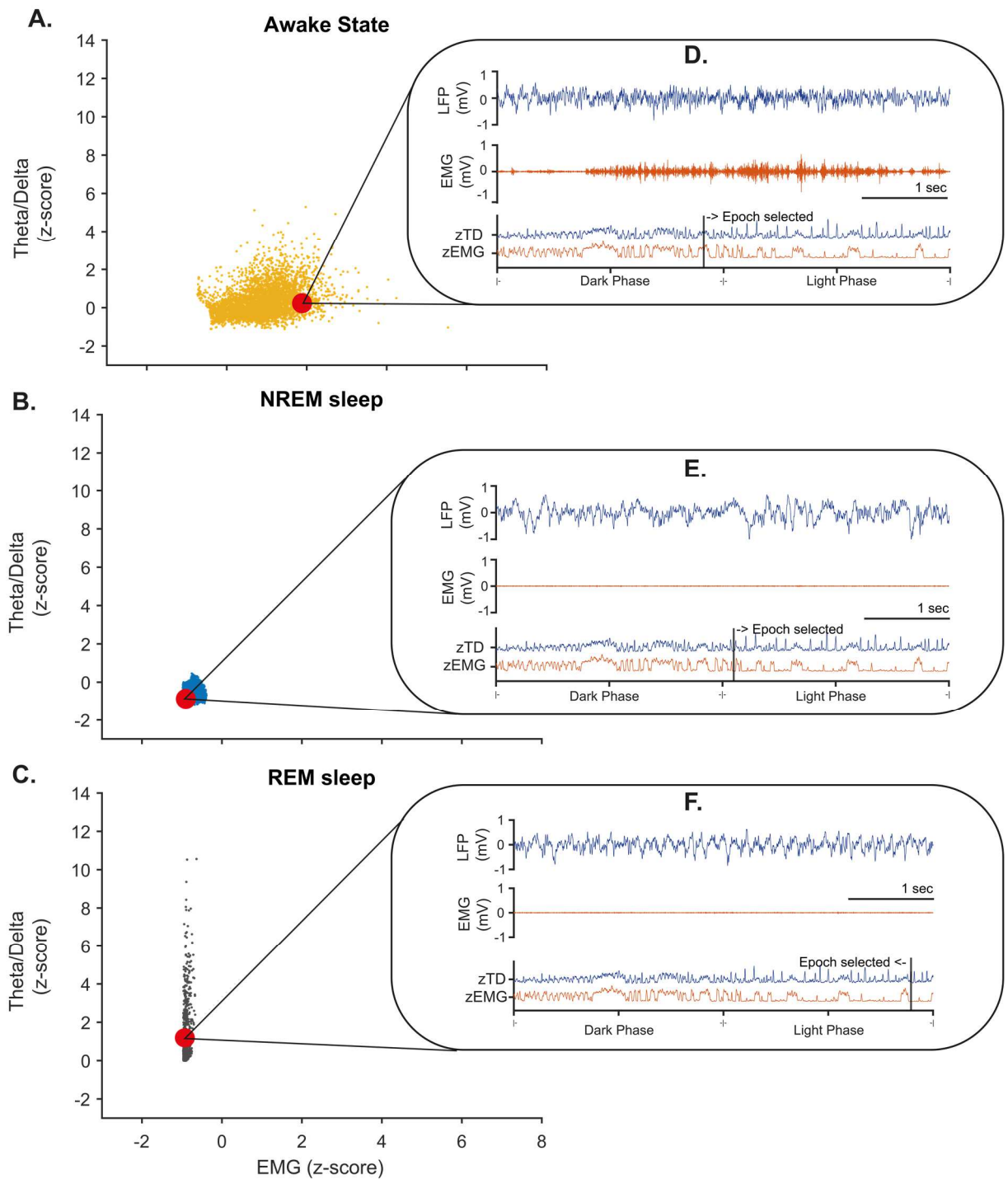


Figura 19 Classificação Final. A: Épocas classificadas como acordado; B: Épocas classificadas como sono NREM; C: Épocas classificadas como sono REM; D: Época representativa do estado acordado demonstrando, de cima para baixo, o sinal do hipocampo e do EMG (por 5 segundos) e a posição relativa da época na distribuição ao longo do tempo da relação Teta/Delta (z-score da relação Teta/Delta, em z-score) e do RMS do EMG (z-score do RMS do EMG, em z-score); E: Mesmo que em D, mas para o sono NREM; F: Mesmo que em D, mas para sono REM.

#### **4.8 - Os estados flutuam ao longo do tempo coerentemente com o ciclo circadiano**

Na seção 4.3, demonstramos que os valores de EMG e Teta/Delta das épocas flutuam ao longo das 24 horas de registro e representam o comportamento dos animais de forma coerente com o esperado para cada etapa do ciclo circadiano. Como as épocas classificadas formaram grupos com características típicas de cada estado (seção anterior), esperamos que, quando analisada ao longo do tempo, elas acompanhem também a arquitetura do ciclo circadiano.

A Figura 20 apresenta um gráfico de dispersão dos valores de Teta/Delta e EMG ao longo do tempo para as épocas classificadas. Observando esta distribuição durante as 12 horas iniciais do registro (20:00 às 08:00), ou seja, a fase escura, é possível observar uma maior quantidade de épocas classificadas como estado acordado (aproximadamente 60% do tempo total). Já nas 12 horas subsequentes (fase clara) as épocas foram classificadas mais frequentemente como sono NREM e REM, totalizando 62% do tempo (52% e 10%, respectivamente). Esses resultados foram condizentes com o esperado para cada fase do ciclo circadiano.

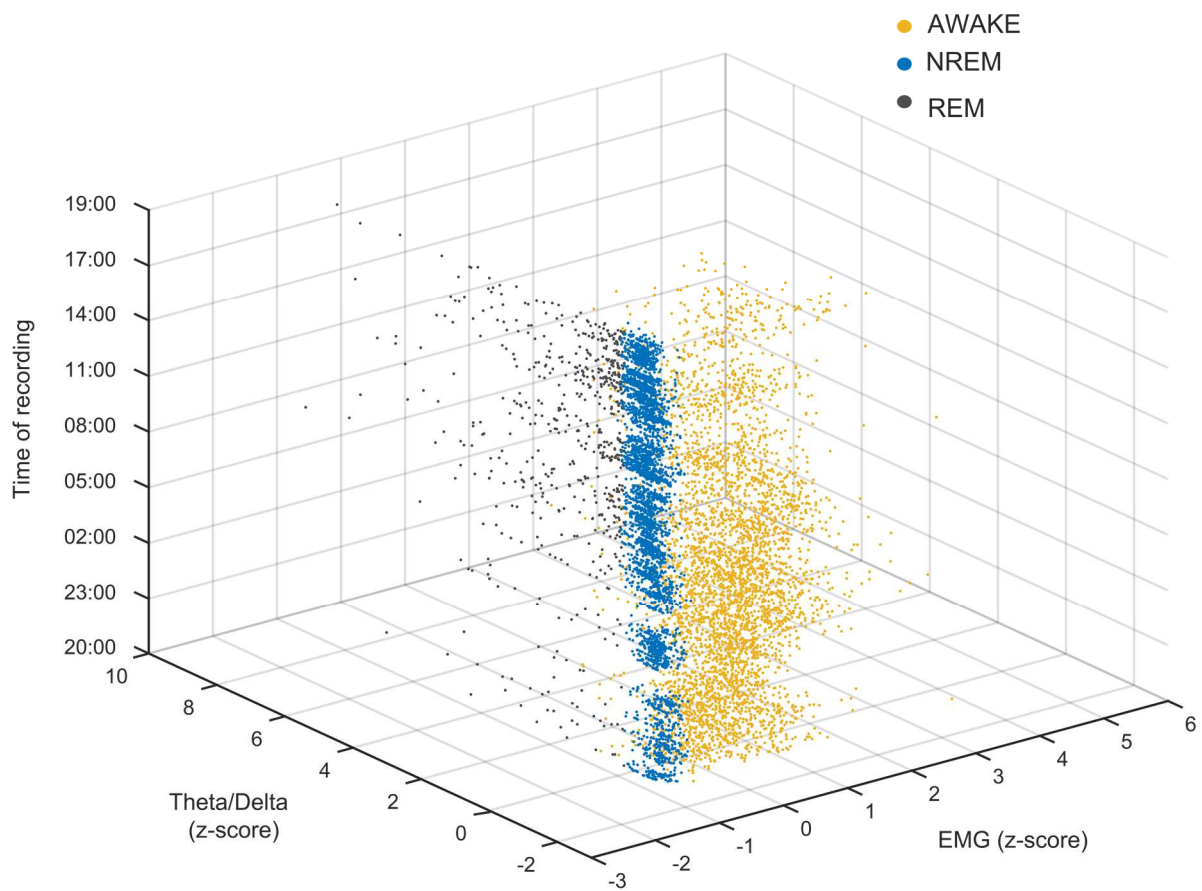


Figura 20. Classificação das épocas em gráficos de dispersão. Distribuição das épocas classificadas em cada um dos estados e seus valores de z-score da relação Teta/Delta, do RMS do EMG e da posição no tempo;

Durante a fase escura, além de ocuparem maior parte do tempo, os períodos em estado acordado foram mais longos, durando em média aproximadamente 15 minutos (Figura 21-A). Já na fase clara, a arquitetura dos estados apresentou-se mais fragmentada, porém com maior parte do tempo em sono, com aproximadamente 62% do tempo em sono NREM e 11% em sono REM (Figura 21-B).

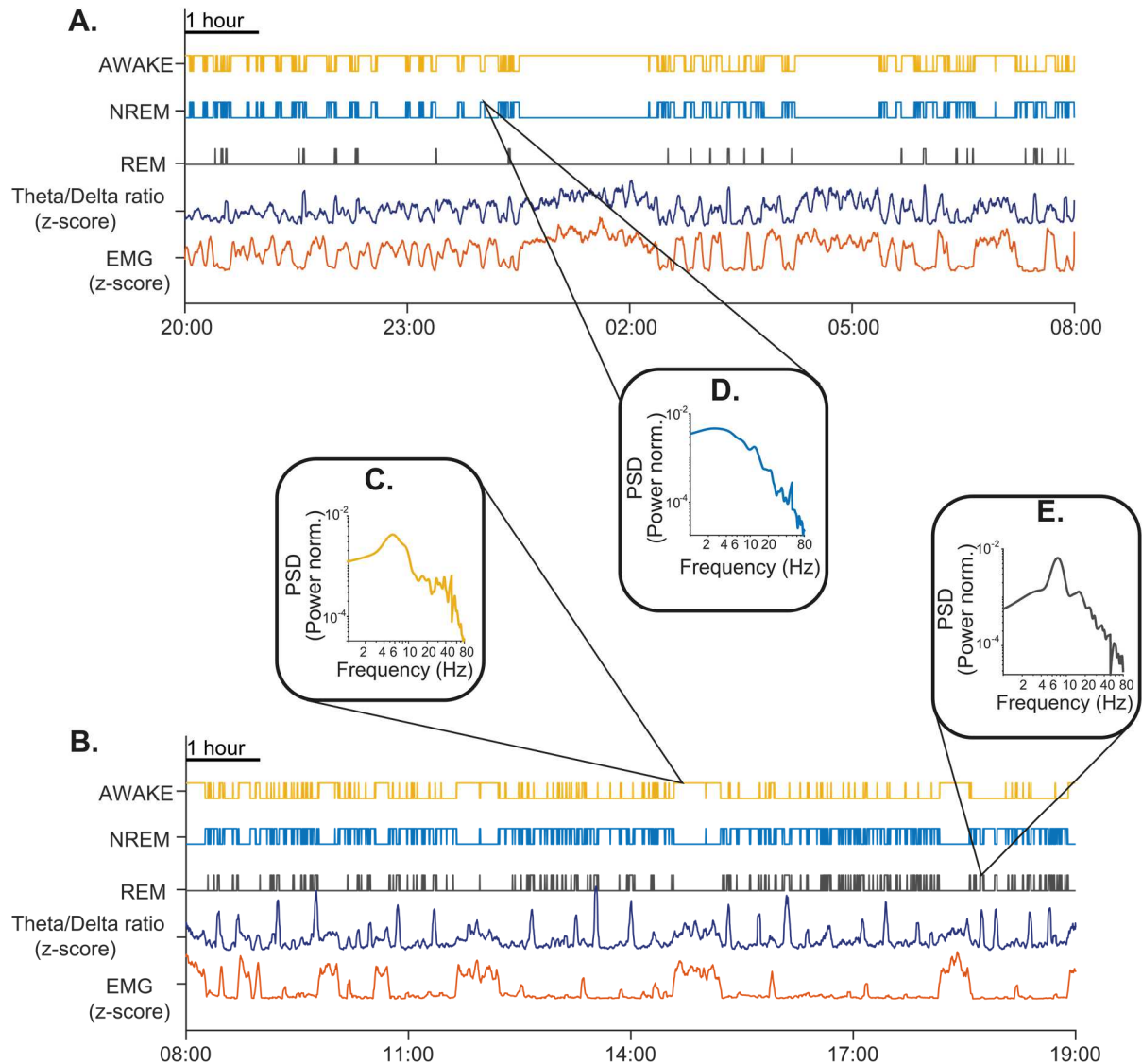


Figura 21. Classificação ao longo do tempo. A. De cima para baixo, distribuição das épocas classificadas como acordado, sono NREM, sono REM, e dos z-scores da relação Teta/Delta e do RMS do EMG, durante a fase escura em A e fase clara em B. PSD de uma época representativa do estado acordado em C. do sono NREM em D. e do sono REM em E.

#### 4.9 - As características finais dos estados são consistentes nos diferentes registros testados

A distribuição em z-score dos valores de RMS do EMG e da relação Teta/Delta de cada época tornou possível uma comparação direta entre diferentes registros, sendo uma ferramenta importante para análises do resultado das classificações em diferentes registros.

Considerando todos os registros analisados, as épocas classificadas como acordado obtiveram os maiores valores de média para o EMG (Figura 22-A e D). As classificadas como NREM obtiveram os menores valores em média da razão Teta/Delta e do EMG (Figura 22-B e E). Por fim, as classificadas como REM exibiram os maiores valores em média para razão Teta/Delta (Figura 22-C) mas com uma elevada dispersão (Figura 22-F), provavelmente devido à sua distribuição (Figura 19-C). Além disso, obtiveram baixos valores em média para EMG (Figura 22-C).

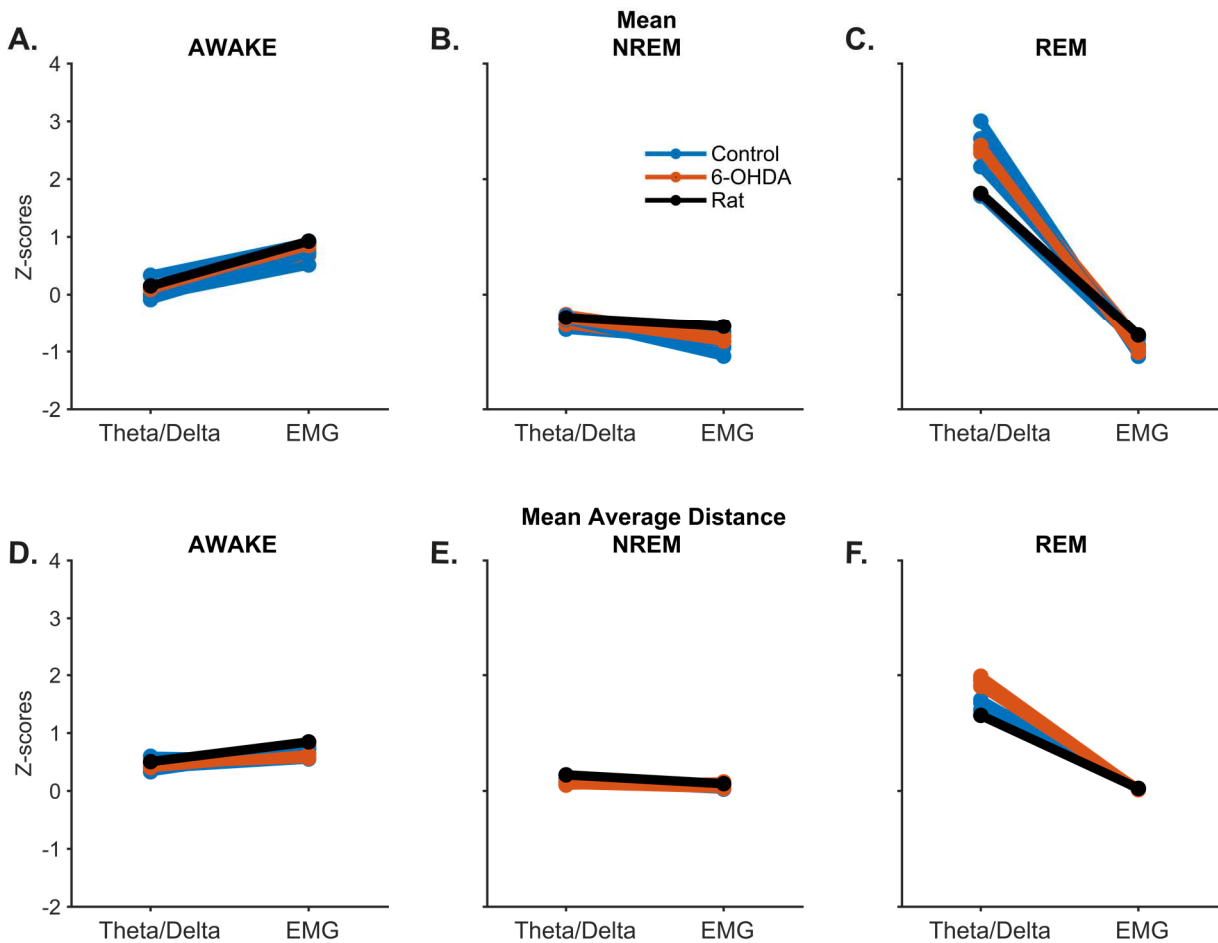


Figura 22. Dispersão das épocas classificadas em cada um dos estados. A: Média dos valores da razão Teta/Delta para o estado acordado dos registros em camundongo controle (azul), 6-OHDA (vermelho) e em rato (preto); B: Mesmo que A, para o estado NREM; C: Mesmo que A, para o estado REM; D: Média dos desvios absolutos da relação Teta/Delta e do RMS do EMG para o estado acordado dos registros; E: Mesmo que D, para o estado NREM; F: Mesmo que D, para o estado REM.

Todos os registros utilizados apresentaram elevado acurácia na classificação confirmado pela taxa de verdadeiro e falso positivos (Figura 23-A, B e C) e também pelas áreas sob as curvas ROCs (Figura 23-D). Em média, os animais passaram a

maior parte do tempo acordados durante a fase escura e mais tempo dormindo (sono REM+NREM) durante a fase clara (Figura 23-E).

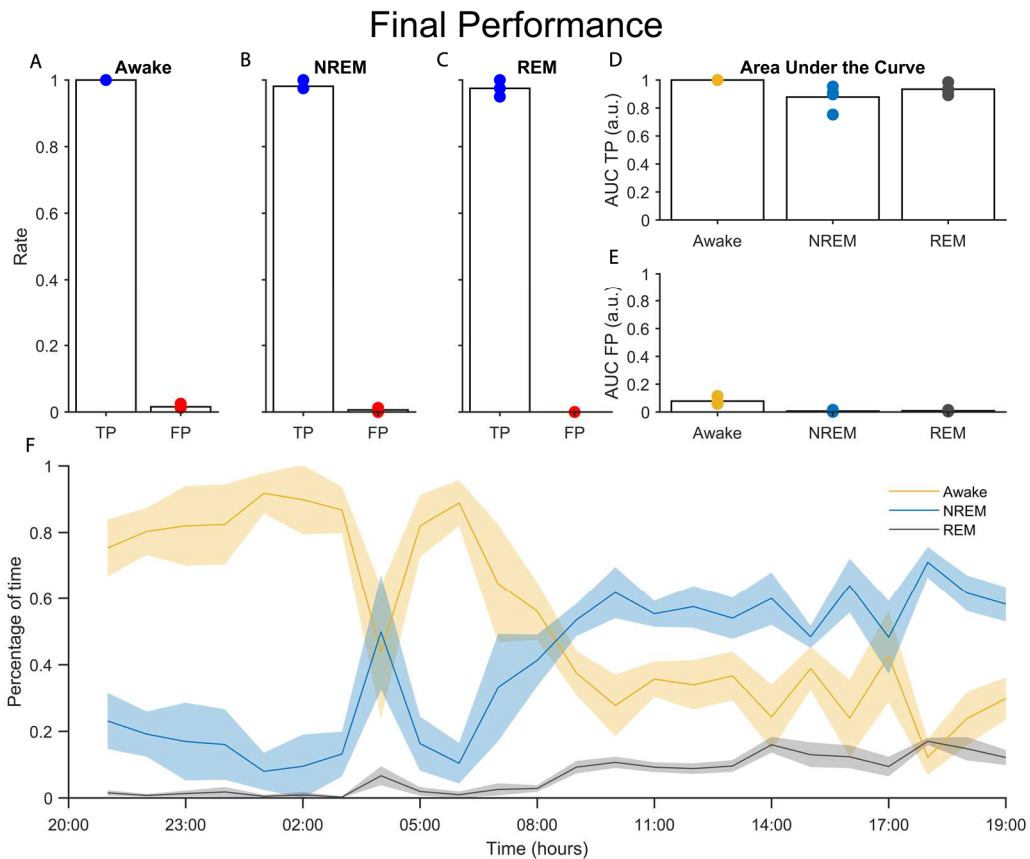


Figura 23. Acurácia geral dos registros em camundongo controle. A: Taxa de verdadeiro positivos (azul) e falso positivos (vermelho) para o estado acordado; B: Mesmo para sono NREM; C: Mesmo para sono REM; D: Valor da área sob a taxa de verdadeiros positivos pelos diferentes limiares, calculada para cada estado; E: Mesmo que D, para taxa de falsos positivos para cada limiar; F: Porcentagem do tempo em cada estado a cada hora (porcentagem da primeira hora, i. e. 20:00 às 21:00, demonstrada às 21:00). Média e erro da média representados como linha e área colorida, respectivamente.

#### 4.10 - A *Toolbox* demonstrou acurácia elevada na classificação de registros de animais modelo da doença de Parkinson

A doença de Parkinson (DP) é a segunda doença neurodegenerativa mais comum no mundo e é classicamente associada a neurodegeneração da via nigroestriatal gerando sintomas motores, como a bradicinesia e o tremor (Braak et al. 2003; Del Tredici and Braak 2016). Porém, trabalhos recentes têm demonstrado que a doença pode afligir múltiplos sistemas de neurotransmissão no sistema nervoso

central. Tais alterações parecem estar subjacentes aos diversos sintomas não-motores, tais como prejuízos do sono, sonolência excessiva e distúrbios no sono REM (Bastide et al. 2015; De Castro Medeiros et al. 2019; Gros and Videnovic 2017; Schapira, Chaudhuri, and Jenner 2017). Modelos murinos experimentais da DP têm sido amplamente utilizados na tentativa de esclarecer melhor os possíveis mecanismos desses sintomas não-motores (Carvalho et al. 2017; Ciric et al. 2019; Vo et al. 2014).

No presente estudo, também testamos nossa *Toolbox* em registros realizados em um modelo clássico de DP induzido em camundongos (D. L. Commins, R. A. Shaughnessy, K. J. Axt, G. Vosmer 1990; Kirik, Rosenblad, and Björklund 1998; Stott and Barker 2014). Quatro animais foram injetados com 6-OHDA na via nigroestriatal, e uma semana após, eletrodos foram implantados em CA1, ipsilateral à injeção, e no músculo do pescoço para EMG. Fizemos registros de 24 horas desses animais e utilizamos a nossa *Toolbox* para fazer a classificação desse conjunto de dados.

Inicialmente, escolhemos um registro como representativo para demonstrar a acurácia da *Toolbox* (Figura 24-A). Após a classificação, as PSDs médias de cada estado mantiveram as características espectrais esperadas (Figura 24-B), apresentando as épocas em sono REM com elevada potência em Teta e baixa em Delta, enquanto que sono NREM baixa potência em Teta e elevada em Delta (Tabela 3).

Tabela 3. Valor máximo de potência normalizada e a respectiva frequência

		Acordado	Sono NREM	Sono REM
Teta	Potência Normalizada	$4,6 \times 10^{-3}$	$3,5 \times 10^{-3}$	$7,1 \times 10^{-3}$
	Frequência no Pico (Hz)	6,5	6	6,85
Delta	Potência Normalizada	$3,1 \times 10^{-3}$	$4,1 \times 10^{-3}$	$1,8 \times 10^{-3}$
	Frequência no Pico (Hz)	4	2,6	4

Fonte: Elaboração própria

Como demonstrado também na Figura 22, as épocas classificadas formaram grupos coesos com as características típicas de cada estado. As classificadas como



acordado obtiveram os maiores valores de EMG, enquanto que as em sono NREM os menores valores de EMG e da razão Teta/Delta, enquanto que, esta última teve seus maiores valores para o sono REM (Figura 24-C).

A classificação ao longo do tempo indicou maior presença do estado acordado durante a fase escura, com 74% do tempo (Figura 24-D). Durante a fase clara (Figura 24-E), o animal esteve dormindo por 58% do tempo (50% sono NREM e 8% sono REM).

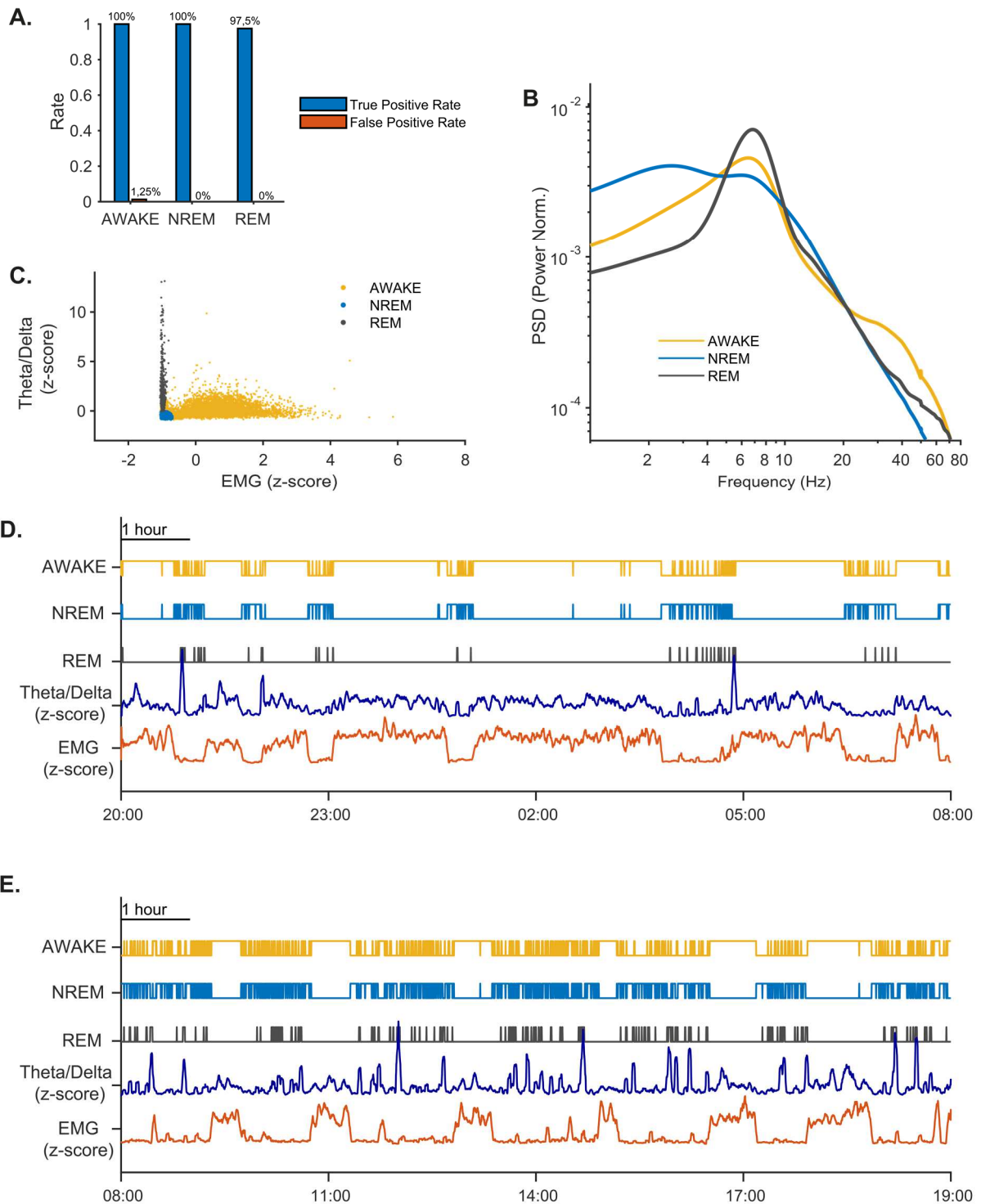


Figura 24. Registro de animais modelo da doença de Parkinson. A. Taxa de verdadeiro positivos (em azul) e Falso positivos (em vermelho) obtidas para esse registro; B. Média das PSDs das épocas classificadas em cada estado; C. Gráfico de dispersão das épocas conforme seus valores de z-score da relação Teta/Delta e z-score do RMS do EMG, e coloridas de acordo com a classificação; D. De cima para baixo, distribuição das épocas classificadas como acordado, sono NREM, sono REM, e dos z-score da relação Teta/Delta e z-score do RMS do EMG, durante a fase escura; E: Mesmo que D para a fase clara.

Todos os 4 registros de animais submetidos a DP utilizados neste trabalho apresentaram elevada taxa de verdadeiro positivos e baixa taxa de falsos positivos (Figura 25-A, B e C), com pontuação ideal no cálculo da área sob a curva (Figura 25-D). Em média, os animais DP, assim como os animais controle, passaram maior parte do tempo da fase escura acordados. Durante a fase clara, ficaram mais tempo dormindo (Figura 25-E).

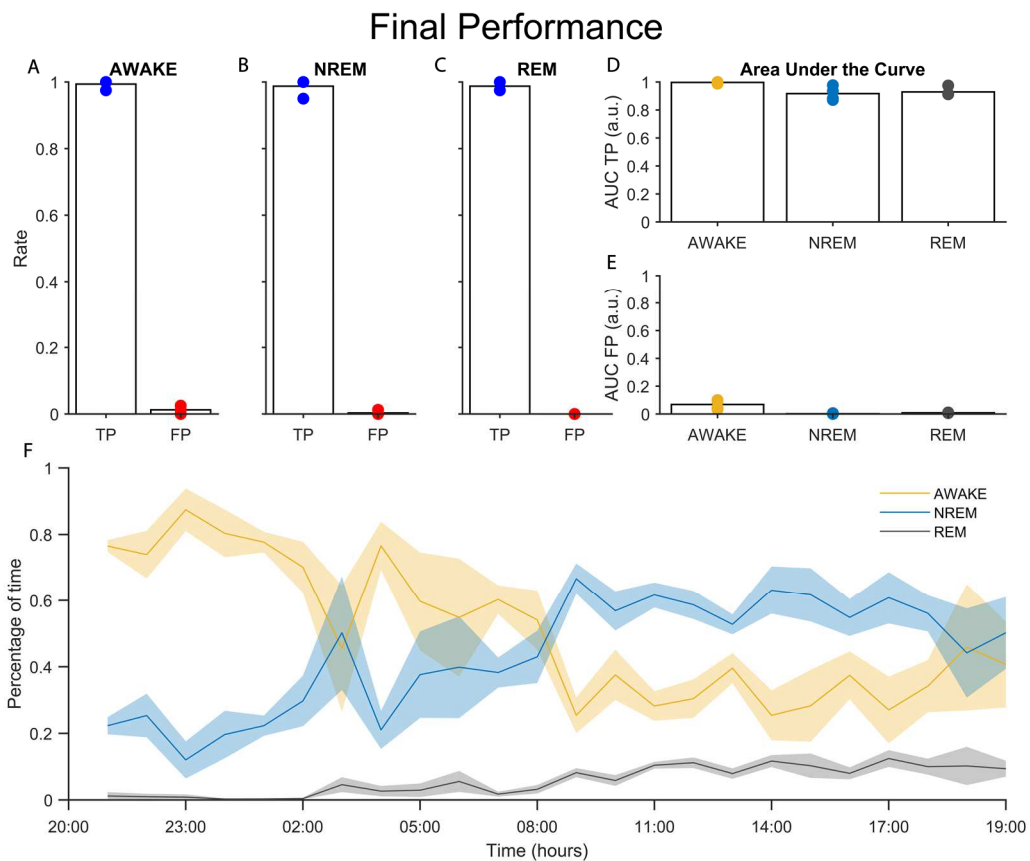


Figura 25. Acurácia geral dos registros em camundongo 6-OHDA. A: Taxa de verdadeiro positivos (azul) e falso positivos (vermelho) para o estado acordado; B: Mesmo para sono NREM; C: Mesmo para sono REM; D: Valor da área sob a taxa de verdadeiros positivos pelos diferentes limiares, calculada para cada estado; E: Mesmo que D, para taxa de falsos positivos para cada limiar; F: Porcentagem do tempo em cada estado a cada hora (porcentagem da primeira hora, i. e. 20:00 às 21:00, demonstrada às 21:00). Média e erro da média representados como linha e área colorida, respectivamente.

Já foi demonstrado que animais 6-OHDA dormem mais durante a fase escura, o que caracteriza um quadro de sonolência excessiva, sintoma também encontrado em pacientes da DP (Fantini et al. 2018; Stefani and Högl 2019). Nossos resultados preliminares ainda não reproduziram estatisticamente o mesmo efeito (n=4;

porcentagem de tempo acordado durante a fase escura dos animais controle vs animais lesão;  $p = 0,4876$ ; Mann-Whitney T test). Nesta avaliação preliminar, não reproduzimos esse quadro de sonolência excessiva, porém este estudo ainda está em andamento e mais animais ainda serão registrados e analisados.

Tabela 4. Porcentagem média de tempo em cada estado dos registros controle e 6-OHDA.

	Controle		6-OHDA	
	Fase Escura	Fase Clara	Fase Escura	Fase Clara
Acordado	77%	30%	68%	33%
Sono NREM	21%	58%	30%	57%
Sono REM	2%	11%	2%	10%

Fonte: Elaboração própria

#### 4.11 - Avaliação do acurácia da *Toolbox* para registro realizado em rato Wistar

Ratos têm sido utilizados como modelos animais para estudo de sono, memória (Buzsáki and Tingley 2018; Colgin 2020), e também são utilizados para indução de modelos experimentais de várias doenças, como Parkinson (De Castro Medeiros et al. 2019) e Epilepsia (Mendes et al. 2019). Porém, os trabalhos que conseguem demonstrar e comparar resultados de classificação do ciclo circadiano em modelos normais e/ou de doenças em diferentes espécies são muito escassos, principalmente devido a incompatibilidades na metodologia de classificação.

Para testar a acurácia da *Toolbox* na classificação de um registro feito em rato, implantamos um eletrodo em CA1 do hipocampo e um no músculo do pescoço para registros eletrofisiológicos. O registro foi feito durante 5 horas da fase clara. O dado foi fragmentado em épocas de 10 segundos e utilizamos a nossa *Toolbox* para a classificação. Nossos resultados mostram que a nossa *Toolbox* realizou a classificação com 100% de acurácia para os estados acordado e sono NREM, enquanto que uma taxa de 87,5% para o sono REM. A taxa de falso positivo foi de 0% para os estados acordados e sono REM, enquanto que para o sono NREM, foi de 6,25% (Figura 26-A). Análises espectrais do sinal do hipocampo revelaram que, em média, as épocas apresentaram as características esperadas de cada estado (Figura 26-B). Considerando os valores do EMG, as épocas de ambas etapas do sono

obtiveram os menores valores, enquanto que as classificadas como acordado, os maiores (Figura 26-C). A classificação disposta ao longo do tempo (Figura 26-D) demonstra que cerca de 39% das épocas foram classificadas como acordadas, enquanto que 53% foram classificadas como sono NREM e 9% como sono REM. Como esperado para um registro na fase clara, por maior parte do tempo o animal foi classificado como dormindo. Esses resultados sugerem que a *Toolbox* possa também ser utilizada em ratos, porém são necessários mais registros para futuras comparações e robustez.

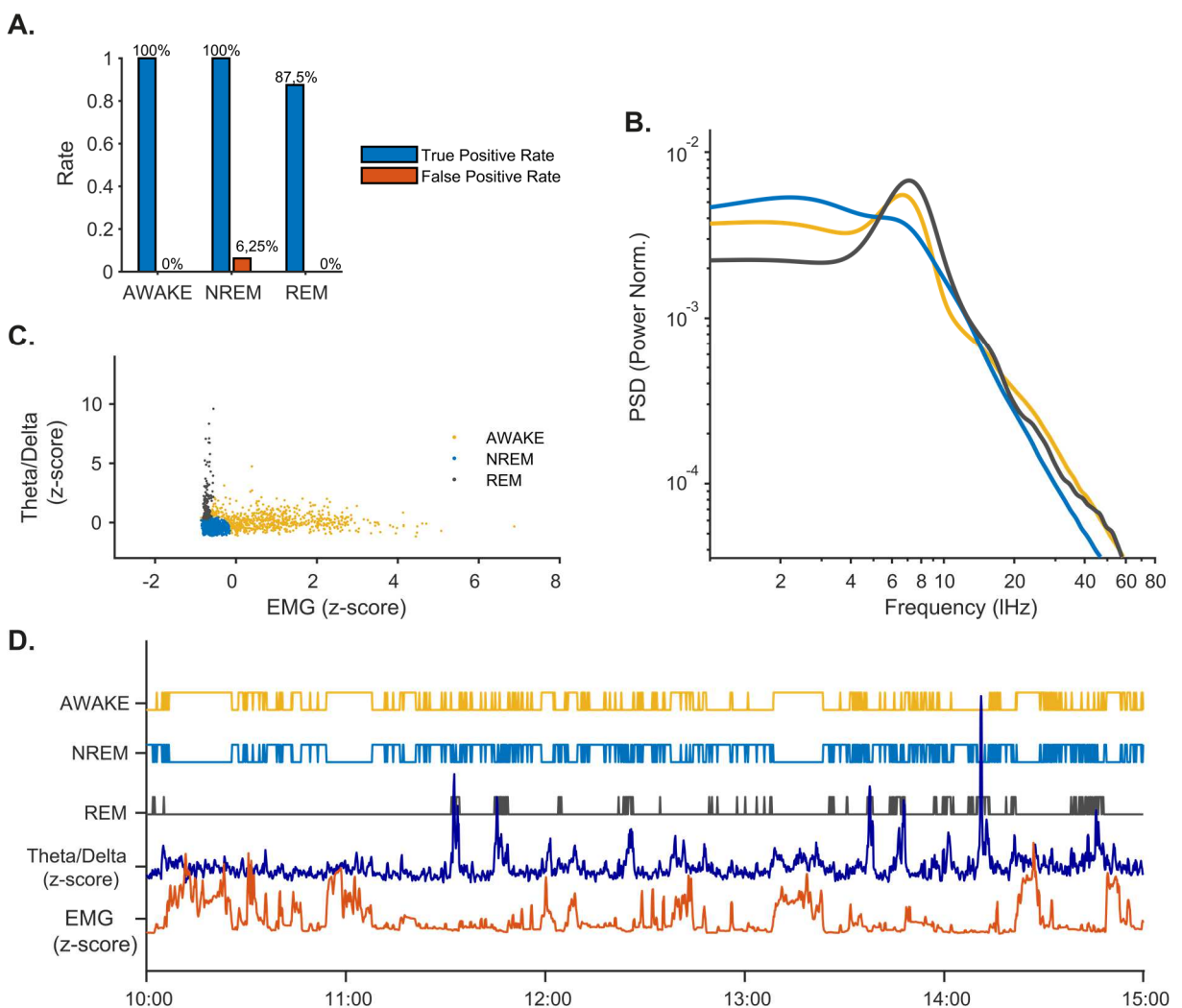


Figura 26. Registro rato. A: Taxa de verdadeiros positivos (em azul) e falsos positivos (em vermelho) obtidas para esse registro; B. Média das PSDs das épocas classificadas em cada estado; C. Gráfico de dispersão das épocas conforme seus valores de z-score da relação Teta/Delta e z-score do RMS do EMG, e coloridas de acordo com a classificação; D. De cima para baixo, distribuição das épocas classificadas como acordado, sono NREM, sono REM, e dos z-score da relação Teta/Delta e z-score do RMS do EMG.

#### **4.12 - As análises feitas pela *Toolbox* atuam como controle sobre a qualidade do dado coletado**

Registros eletrofisiológicos *in vivo* são tecnicamente difíceis de serem coletados com alta qualidade. Fatores que variam de problemas no implante dos eletrodos a interferências da rede elétrica do laboratório onde a coleta está sendo feita podem impactar negativamente na qualidade do sinal adquirido. Foram testadas se as análises feitas pela *Toolbox* oferecem subsídios para julgar a qualidade do dado e definir um padrão mínimo de exigências sem que haja interferência nos resultados finais. Para isso, utilizamos um registro de 24 horas em camundongo controle, mas no qual o eletrodo de registro do hipocampo foi mal colocado, impossibilitando a coleta do dado. Como comparação, utilizamos o registro representativo de um camundongo saudável que foi apresentado neste trabalho.

A curva ROC calculada para cada estado combina os resultados da classificação do GMM com os da classificação manual do usuário, sendo uma ferramenta importante para avaliar a acurácia do GMM (Figura 9-A, B e C), mas também para anunciar classificações aberrantes que exigem futuras análises (Figura 9-D, E e F). O cálculo da área sob a curva indica se a distribuição do dado segue o padrão da classificação manual (Bruijn 2011). Valores dessa área superiores a 0.8 a.u. para a AUC TP e menores que 0.2 para AUC FP (Figura 27-A, B e C) indicam maior similaridade entre as classificações. Valores próximos a 0.5 a.u. para ambas AUC TP e FP indicam que o GMM não conseguiu distinguir falso positivos de verdadeiro positivos, representando o pior acurácia possível. Valores menores que 0.2 para AUC TP e maiores que 0.8 para AUC FP indicam que o GMM formou grupos com características contrárias aos das épocas classificadas manualmente (Figura 28-A, B e C). Os dois últimos casos demonstram que os grupos formados pelo GMM não apresentam as características esperadas, como consequência dos valores individuais das épocas, indicando má qualidade do dado coletado.

Ao fim da classificação, figuras com épocas representativas são geradas e tornando possível analisar o sinal do hipocampo (Figura 27-D) e EMG (Figura 27-E) dos registros e estabelecer comparações com o registro problemático (Figura 28-D e E). Escolhendo uma época com um dos maiores valores da relação Teta/Delta e um

dos menores valores para o EMG (Figura 27-F), características do sono REM, é esperado que o sinal do hipocampo apresenta atividade em Teta proeminente (Figura 27-J), com valores de amplitude variando aproximadamente entre -0.5 e 0.5 mV (Figura 27-D).

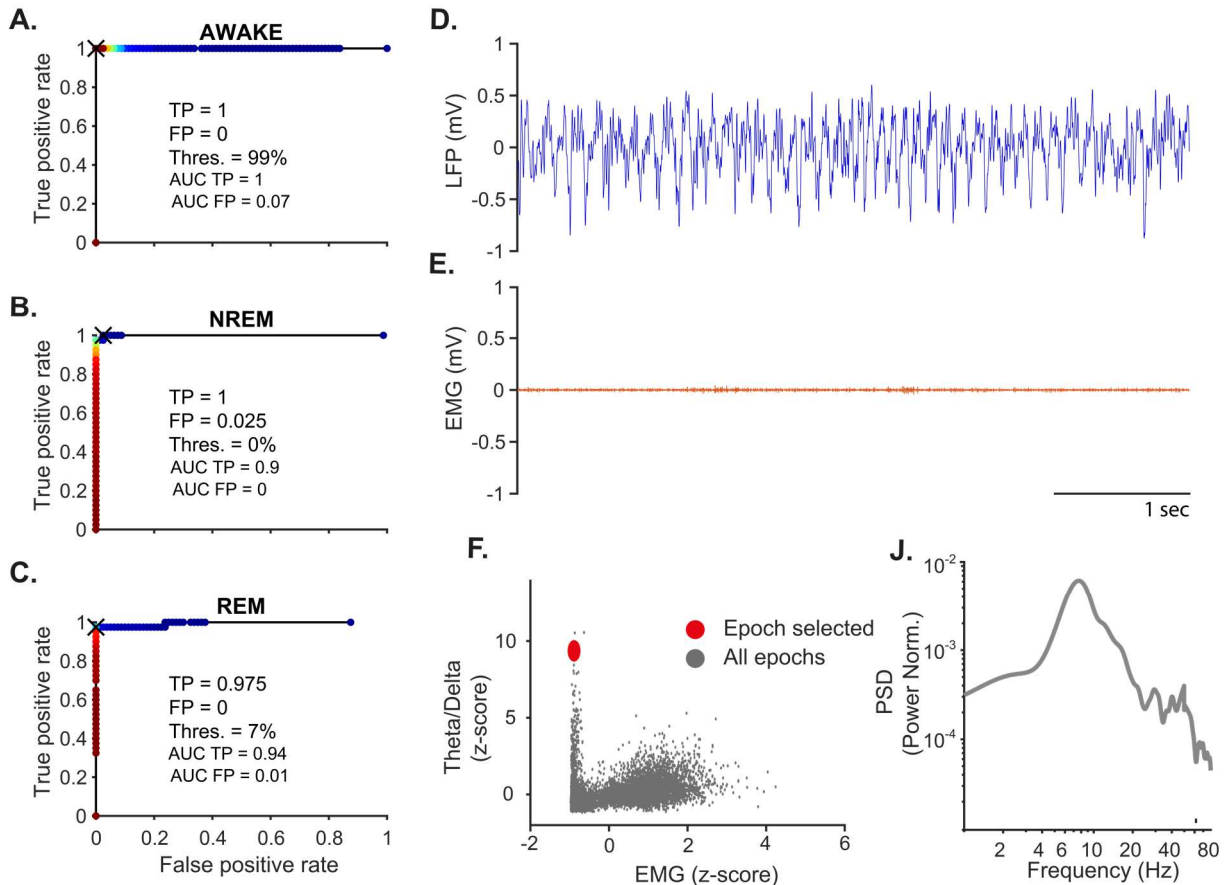


Figura 27. Dado com elevada qualidade. A: Curva ROC para o estado acordado, com sua respectivos taxas de verdadeiro (TP) e falso positivos (FP), limiar de probabilidade escolhido (Thres.) e área sob a curva (AUC) descritos no meio do gráfico; B: mesmo que A, para sono NREM; C: mesmo que A, para sono REM; D: Sinal do hipocampo da época com valor máximo da relação Teta/Delta e mínimo do EMG possível; E: Sinal do EMG da mesma época em D; F: Distribuição das épocas no gráfico de dispersão. Época selecionada em vermelho; J: PSD do sinal do hipocampo da mesma época selecionada.

Comparando com a época anterior com uma de características semelhantes para o registro problemático (Figura 28-F), é possível perceber que o sinal do hipocampo dessa época (Figura 28-D) não apresenta a elevada amplitude esperada (Figura 28-J). Como consequência, o valor máximo da relação Teta/Delta, dentre os de menores valores de EMG, chega a aproximadamente 5 z-score, enquanto que em registros ideais, esse valor pode variar de 10 a 20 z-score (Figura 22-C).

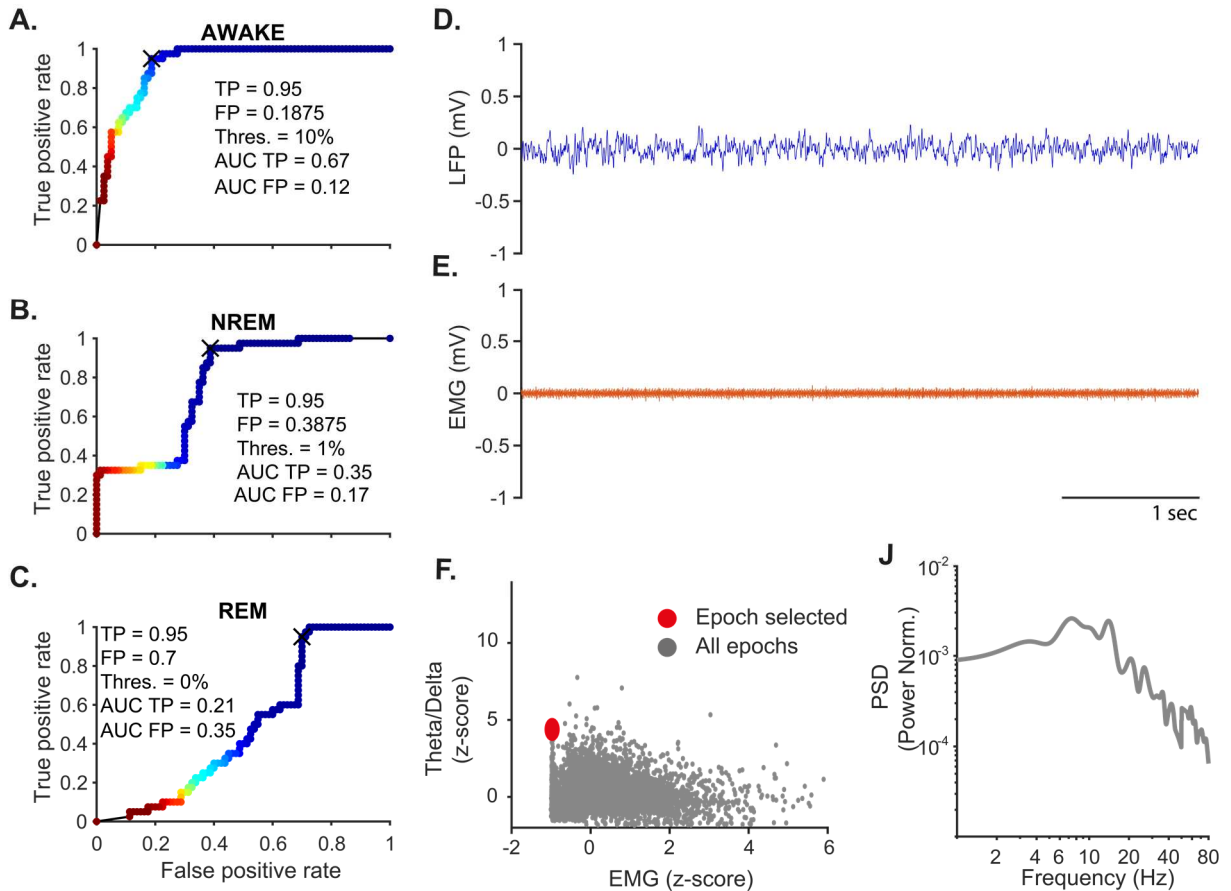


Figura 28 - Dado com qualidade baixa. A: Curva ROC para o estado acordado, com sua respectivos taxas de verdadeiro (TP) e falso positivos (FP), limiar de probabilidade escolhido (Thres.) e área sob a curva (AUC) descritos no meio do gráfico; B: mesmo que A, para sono NREM; C: mesmo que A, para sono REM; D: Sinal do hipocampo da época com valor máximo da relação Teta/Delta e mínimo do EMG possível; E: Sinal do EMG da mesma época em D; F: Distribuição das épocas no gráfico de dispersão. Época selecionada em vermelho; J: PSD do sinal do hipocampo da mesma época selecionada.

Com o eletrodo mal posicionado, atingindo o córtex ao invés do hipocampo, a distribuição do dado é comprometida (Figura 29), fazendo com que o GMM gere grupos com características não esperadas, resultando num acurácia não satisfatório da classificação (Figura 28-A).



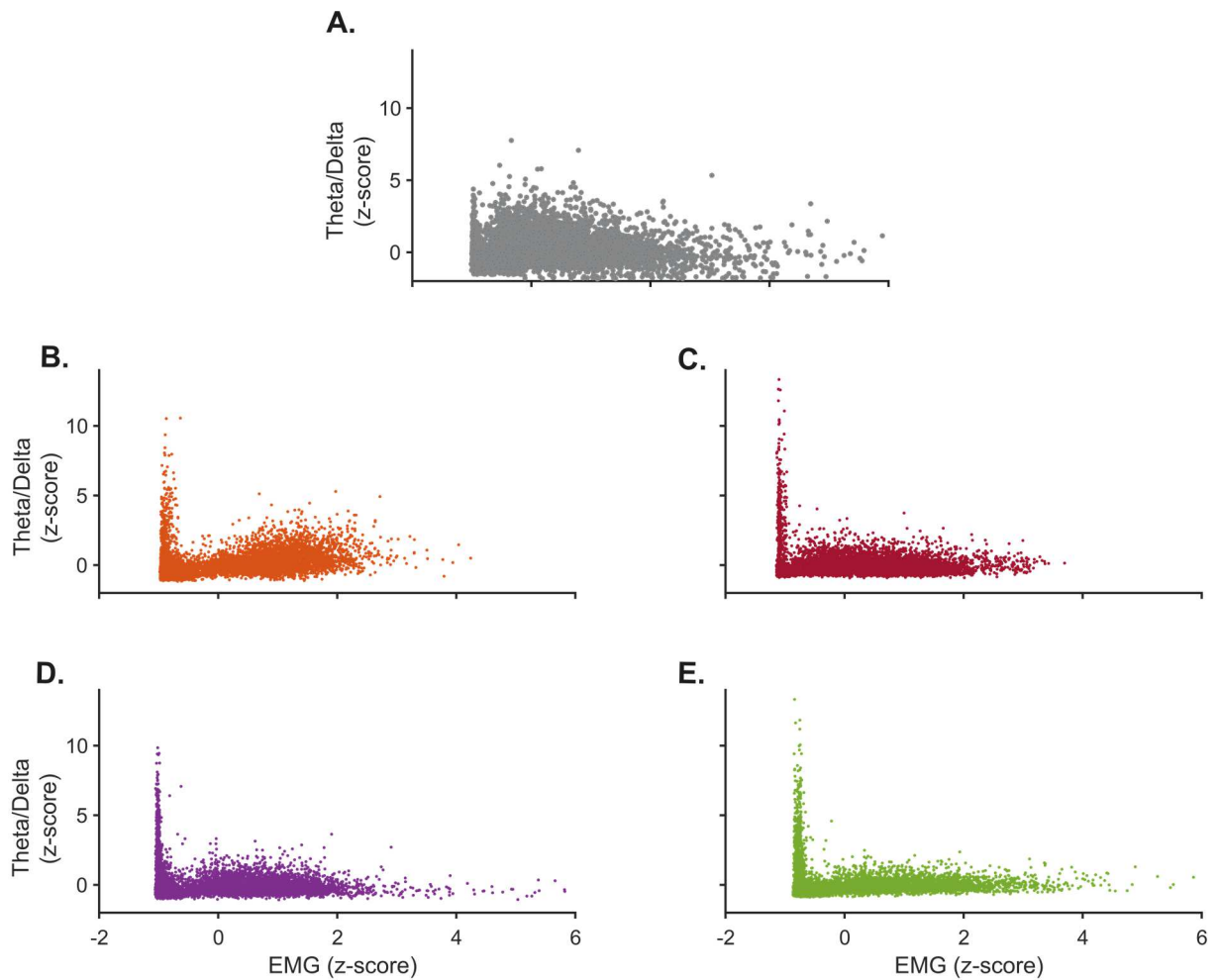


Figura 29. Comparando distribuições dos registros controle. A: Gráfico de dispersão de registro com sinal do hipocampo com qualidade inferior; B: Gráfico de dispersão com sinal do hipocampo com qualidade ideal. Este foi o registro utilizado nas figuras 1 a 13 deste trabalho; C: Outro registro com qualidade ideal; D: Outro registro com qualidade ideal; E: Outro registro com qualidade ideal;

Além das análises sobre a distribuição do dado, as análises finais sobre a média dos valores de potência das épocas classificadas em cada estado também revelam problemas no registro, com ausência do pico em Teta na média das janelas de sono REM (Figura 30-B). Neste caso, o GMM classificou equivocadamente como acordado épocas com os menores valores de EMG (Figura 30-C).

Em combinação, essas análises ajudam o usuário a determinar quais registros deverão ser excluídos do trabalho por não atingirem um padrão mínimo de qualidade, podendo impactar negativamente na interpretação dos resultados obtidos.

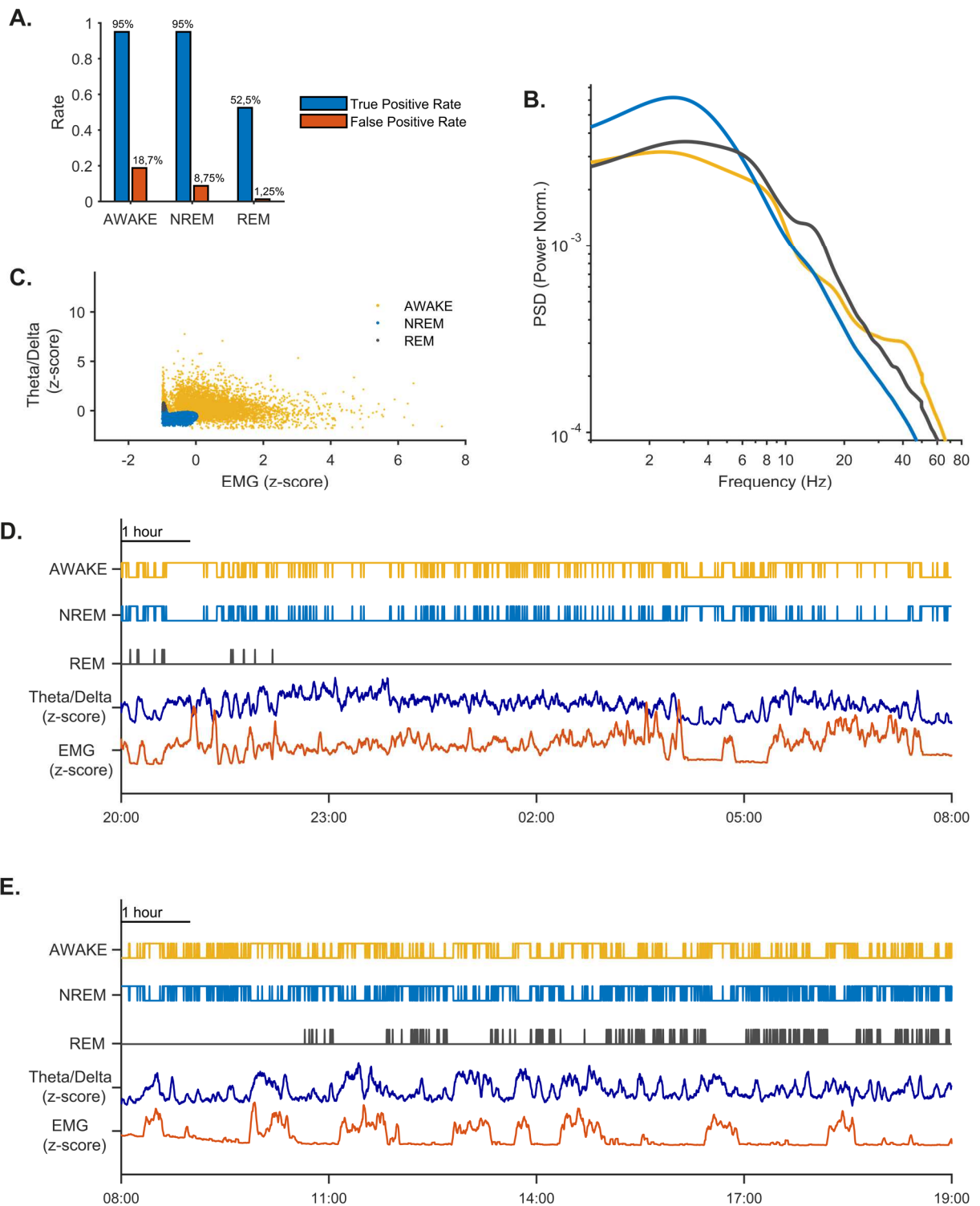


Figura 30. Registro de baixa qualidade. A. Taxa de verdadeiro positivos (em azul) e Falso positivos (em vermelho) obtidas para esse registro; B. Média das PSDs das épocas classificadas em cada estado; C. Gráfico de dispersão das épocas conforme seus valores de z-score da relação Teta/Delta e z-score do RMS do EMG, e coloridas de acordo com a classificação; D. De cima para baixo, distribuição das épocas classificadas como acordado, sono NREM, sono REM, e dos z-score da relação Teta/Delta e z-score do RMS do EMG, durante a fase escura em a. e fase clara em b.

## 5. Discussão

Os estudos sobre o ciclo circadiano em modelos animais entraram em uma nova era graças ao desenvolvimento de ferramentas altamente eficientes na coleta, armazenamento e análise de um grande volume de dados. Os avanços especificamente nas técnicas de registros eletrofisiológicos são de grande contribuição, pois possibilitaram a descoberta e caracterização de padrões oscilatórios que podem ser utilizados como marcadores de estados comportamentais ao longo do ciclo sono-vigília (A. R. Adamantidis 2019). Vários trabalhos propuseram métodos de classificação desses estados, utilizando estratégias que variam de fragmentação dos dados em épocas seguidas de suas classificações manuais até o desenvolvimento de algoritmos complexos para classificações automáticas (Quadros 2 e 3). Contudo, esses trabalhos falham em reportar detalhadamente os métodos utilizados e o acurácia do classificador utilizados. Dentre as estratégias disponíveis de classificação que reportam suas taxas de verdadeiros e falsos positivos, pode-se notar que a maioria possui acurácia sub-ótimo, em especial para classificação de sono REM (Katsageorgiou et al. 2015). Além de um acurácia satisfatório, um classificador deve também permitir reprodutibilidade. Para isso, é imprescindível que haja o máximo de transparência em relação aos métodos utilizados. Especificamente para estudos do sono, as estratégias utilizadas para a classificação dos estados precisam estar bem detalhadas nos métodos. Na seção de resultados, representações gráficas capazes de ilustrar didaticamente aspectos específicos dos principais achados obtidos são muito importantes. Nesse contexto, é necessário que existam ferramentas de acesso livre que possuam interface gráfica para garantir sua fácil utilização e que sejam capazes de realizar a classificação de estados e gerar uma série de figuras para avaliação do acurácia do classificador e controle de qualidade do conjunto de dados em questão. Trabalhos recentes têm empregado algoritmos com redes neurais para seus métodos, porém esses algoritmos são de difícil utilização por pesquisadores não especializados em programação (Vladimir et al. 2020; Yamabe et al. 2019; Yildirim, Baloglu, and Acharya 2019) e, por muitas das vezes, extraem padrões complexos dos dados que são de difícil interpretação (Alex and Lon 2019). Além disso, essas estratégias exigem um poder computacional elevado e sua implementação requer

profissionais com formação em programação, reduzindo o alcance destas ferramentas à comunidade científica.

Neste contexto, nosso trabalho se propôs a desenvolver e verificar uma *Toolbox* para a classificação do ciclo sono-vigília capaz de (1) apresentar uma acurácia satisfatória e mensurável, no qual a classificação manual tem papel decisivo, que é (2) igualmente eficaz em diferentes espécies de roedores e (3) em diferentes modelos murinos de doenças, (4) além de ser de fácil implementação por meio de uma interface gráfica, (5) de livre acesso e (6) compatível com a maioria dos computadores. Além disso, nosso classificador (7) gera representações gráficas e numéricas que possibilita controle e reprodução dos achados. A seguir, discutiremos os principais aspectos dos nossos achados.

### **5.1 - A relação Teta/Delta do sinal do hipocampo e o RMS do EMG como parâmetros principais da classificação**

A definição de quais informações devem ser extraídas dos dados para serem utilizadas na classificação é uma das etapas mais importantes para o desenvolvimento de um método de classificação eficiente (Katsageorgiou et al. 2015). A definição de estados comportamentais ao longo do ciclo sono-vigília é baseada classicamente na atividade do EMG e em características espectrais do sinal do registro EEG cortical ou dos registros de LFP (A. R. Adamantidis 2019).

O uso do EMG como dado importante para classificação já foi alvo de críticas, na qual alega-se que o eletrodo não é capaz de captar movimentos sutis do animal (Lampert et al. 2015). Porém, para o melhor de nossos esforços, não encontramos nenhum trabalho que investigou sistematicamente essa hipótese. Além disso, um eletrodo posicionado sob o músculo do pescoço, que são responsáveis pelos movimentos da cabeça, são indispensáveis para qualquer comportamento típico da vigília e dificilmente não captaria as variações mais relevantes para definir que se o animal está acordado (Silvani et al. 2017). Se implantado corretamente, apresentará as mudanças no sinal de forma coerente com o comportamento do animal (seção 4.1).

O aprimoramento dos sistemas de registro e das ferramentas de análise de dados possibilitaram a descoberta de diferentes padrões de atividade do LFP ou do EEG que são caracterizados no domínio da frequência (Cohen 2014). Essas oscilações podem ser utilizadas para a classificação do ciclo sono-vigília, como no caso de Brooks A. Gross et al, 2009 que recorreram à potência em diferentes bandas de frequência isoladamente para a classificação. Por outro, a razão Teta/Delta tem sido amplamente utilizada para as classificações, pois ela auxilia na diferenciação entre sono NREM (com alta potência em Delta e baixa em Teta) e sono REM (com baixa potência em Delta e alta em Teta). Porém, trabalhos recentes têm desafiado a ideia de que análises no domínio da frequência são suficientes para resumir as características mais essenciais para a classificação, alegando que o cálculo de potência esconderia informações que poderiam ser relevantes para a descoberta e caracterização de novos estados (Faust et al. 2019). Por exemplo, modelos de redes neurais que extraem padrões mais complexos e variados diretamente do sinal bruto são propostos como alternativas. Esses trabalhos encontram padrões que posteriormente são utilizados como argumentos para a formação de novos estados do ciclo circadiano (Katsageorgiou et al. 2018). Contudo, o presente estudo optou por realizar a classificação dos estados utilizando essas análises no domínio da frequência pois são bem suportados por estudos que descrevem seus mecanismos de gênese e sobre seu papel para memória durante sono (Buzsáki and Tingley 2018; Colgin 2016; Drieu and Zugaro 2019; Girardeau and Zugaro 2011; De La Prida 2020; Marshall et al. 2020).

Contudo, variações patológicas nas estruturas responsáveis pela gênese das oscilações utilizadas na classificação, ou ainda nos músculos registrados pelo EMG podem comprometer fortemente o acurácia da classificação (De Castro Medeiros et al. 2019; Mendes et al. 2019). Pensando nisso, na atual versão da *Toolbox*, o usuário pode acoplar outras bandas de frequência à classificação para além da razão Teta/Delta, o que pode ser útil para melhorar o acurácia do classificador. Nas próximas versões da *Toolbox*, nós vamos implementar melhorias para a classificação, nas quais o usuário poderá variar entre e/ou combinar a utilização da relação Teta/Delta e o RMS do EMG com o número de *zero crossings* e o RMS do sinal de LFP, que já foi utilizado para a classificação de sono em trabalhos com animais anestesiados com uretana (Guyon et al. 2020; Lopes-aguiar et al. 2020).

## **5.2 - O método proposto de classificação personaliza os limites de cada estado em cada registro, mas sempre respeitando suas características esperadas**

O modelo treinado do GMM parte de um registro previamente classificado como molde dos estados e depois personaliza os valores de média e a matriz de covariância (Geoffrey McLachlan 2000) da razão Teta/Delta e do RMS do EMG de cada estado de acordo com a distribuição do registro em questão. Conseqüentemente, épocas que possuam valores dentro dos limites definidos para um estado em um registro, podem não ser suficientes para a classificação nesse mesmo estado em um outro registro. Por exemplo, o menor valor de EMG que uma época classificada como acordado assumiu no registro representativo deste trabalho foi de -0.73 *z-score* (Figura 19-A), enquanto que em outro registro, também classificado neste trabalho (não apresentado individualmente), esse valor foi de -0.93 *z-score*. Importante ressaltar que, em todos os registros utilizados neste trabalho, a média das épocas classificadas como acordado obtiveram os maiores valores entre os 3 estados (Figura 22). Isto sugere que o nosso método é sensível às diferentes variações dos registros, mais ainda respeita as características típicas de cada estado. Essas variações podem surgir como consequência do comportamento do animal durante o registro ou de variações na qualidade do sinal coletado como reflexo do posicionamento do eletrodo.

Ao fim da classificação, todas as épocas foram designadas a um dos três estados (e sempre serão para qualquer registro que o usuário utilizar). Com isso, as épocas de transição (Gross et al. 2009; Koike et al. 2017) entre os estados serão definidas como pertencentes a algum dos estados, e a decisão de para qual estado uma determinada época de transição será designada dependerá da distribuição de épocas do registro e do limiar de probabilidade escolhido. Caso seja do interesse do usuário investigar os períodos de transição, o usuário pode utilizar os valores de probabilidade de cada época para redefinir como transição aquelas que apresentarem valores próximos ao limiar escolhido entre 2 estados.

Outras estratégias foram utilizadas para separar as épocas nos 3 estados (acordado, sono NREM e sono REM), tendo como base os valores de Teta/Delta e RMS do EMG (em alguns a potência do sinal), como exemplo por meio do *K-means*

(Rytkönen, Zitting, and Porkka-heiskanen 2011) ou como na maioria dos casos, por limiares arbitrários auxiliados com diferentes representações gráficas (Quadro 2). Porém, essas estratégias apresentam limitações que acreditamos serem importantes para o tipo de dado que trabalhamos. No caso do *K-means*, o algoritmo vai tender a formar grupos que apresentem tamanhos iguais (Jain 2010), o que é um grande problema haja visto que, num período de 24 horas, é esperado que o animal passe aproximadamente 50% do tempo acordado, 40% do tempo em sono NREM e 10% do tempo em REM. Já o caso dos limiares arbitrários, como discutido anteriormente, são altamente susceptíveis ao viés dos experimentadores e geralmente são de difícil reprodução. Essas limitações em combinação com o que foi descrito na seção de resultados favoreceram a escolha do GMM como modelo matemático para a classificação.

### **5.3 - Classificação manual como parte ativa no processo de classificação**

A classificação manual é uma etapa crucial do nosso método, pois é a partir dela que calculamos a curva ROC e, conseqüentemente, é a partir dela que definimos os limiares de probabilidade para cada estado. Essa abordagem para a classificação ainda não foi utilizada em nenhum trabalho publicado dentre os que tivemos acesso, e além disso, a classificação manual, quando utilizada, se resume em duas funções: ou é como o próprio método de classificação de todo o registro (Boyce, Glasgow, and Williams 2016; Durán et al. 2018; Saito et al. 2018; Vyazovskiy et al. 2009); ou atua como referência para a classificação e cálculo do acurácia dos classificadores (Iyer et al. 2019; Yamabe et al. 2019). Acreditamos que esta etapa esteve sempre subutilizada, pois atuava sempre apenas para dimensionar a acurácia dos classificadores, enquanto que no nosso trabalho ela atua ativamente para extrair a melhor classificação que o método pode oferecer em cada registro.

O número de épocas manualmente classificadas nos registros de camundongo, foi de 40 épocas por estado (0,5% do total de épocas). Curiosamente, a relação entre o número de épocas classificadas manualmente em um estado e o total de épocas classificadas no mesmo estado varia. Como exemplo, no registro representativo deste trabalho, foram 4134 épocas classificadas pela *Toolbox* como acordado (52% do total, aproximadamente), 3914 como sono NREM (40% do total, aproximadamente)

enquanto que o sono REM foram 592 (8% do total, aproximadamente). Estas épocas foram confrontadas com 40 épocas em cada estado classificadas manualmente. Os resultados mostram que nosso método apresentou alta acurácia e classificou bem, inclusive, a fase de sono REM, que é frequentemente reportada com menor acurácia por outros classificadores (Katsageorgiou et al. 2015).

#### **5.4 - A *Toolbox* proposta possui uma série de estratégias que favorecem a implementação e reprodutibilidade das classificações**

As figuras geradas pela *Toolbox* (Apêndice 2) foram desenhadas com o intuito de oferecer ao usuário o controle completo das etapas da classificação, uma vez que são de interpretação simples e direta. Uma estratégia que facilitou este processo foi a normalização das distribuições de cada variável em *z-score*, uma vez que este cálculo padroniza a escala das distribuições mantendo a relação entre seus valores individuais. Com isso, as comparações entre diferentes registros podem ocorrer de forma direta e de fácil interpretação. Zeke Barger et al, 2019 contestaram o uso do *z-score* alegando que a padronização nas escalas das distribuições das variáveis poderia ser uma das razões do baixo acurácia em modelos de redes neurais desenvolvidos para a classificação, uma vez que, com as escalas padronizadas, se os valores de uma época qualquer que foi classificada em um estado durante o treinamento ocorresse em um outra época em um registro experimental, a rede tenderia a classificar essa época no mesmo estado do treinamento. Ou seja, a distribuição em *z-score* adicionaria um viés do valor individual para as classificações nesses modelos. Porém, no nosso trabalho, nós implementamos o GMM ao invés de algoritmos de redes neurais, e mais importante, o nosso método utiliza um algoritmo (o EM, descrito em detalhes na seção 3.6) que ajusta as distribuições de cada estado, baseado inicialmente no registro treinamento, mas personalizando essas distribuições em cada registro. Sendo assim, foi possível manter as distribuições em *z-score* e atingir o elevado acurácia que obtivemos. Curiosamente, em uma revisão sobre métodos de classificação, Katsageorgiou et al, 2015 discutiram que a escolha dos processos de padronização dos dados pode ser mais importante para a classificação do que o próprio algoritmo escolhido. Neste trabalho demonstramos que o algoritmo teve um papel tão importante quanto a padronização do dado, haja visto que a



combinação entre o uso das distribuições em *z-score* com o GMM teve contribuições indispensáveis para a alta acurácia da *Toolbox*.

Por outro lado, obter registros eletrofisiológicos com um alto nível de qualidade é um grande desafio (Souza et al. 2018), uma vez que os dados gerados em padrões mais baixos de qualidade são uma das principais causas do baixo acurácia dos classificadores. Masato Yamabe et al 2019 publicaram um trabalho no qual dois algoritmos de rede neurais foram combinados em um único método de classificação para atingir elevado acurácia da classificação até mesmo em registros com níveis elevados de ruído de rede. Porém, registros com qualidade fora do ideal, mesmo que aparentemente bem classificados, quando utilizados para responder perguntas biológicas, podem trazer informações contraditórias e acabar gerando conclusões precipitadas. A *Toolbox* proposta apresenta diversas figuras ao usuário que possibilitam ter um controle maior sobre a qualidade do dado, por meio de figuras de épocas representativas e de análises das distribuições do *z-scores* das variáveis. Além disso, um preditor da qualidade do sinal registrado é o valor da área sob a curva ROC (AUC ROC), que indica o grau de dificuldade que o classificador teve para o registro em questão. Valores baixos para a AUC da taxa de verdadeiros positivos (Figura 9-E) e altos para a AUC da taxa de falsos positivos (Figura 9-F) para os diferentes limiares indicam que os grupos formados pelo GMM treinado diferem muito dos grupos do registro de treinamento, sugerindo ser consequência da má qualidade do dado (Figura 28). Portanto, a definição de uma qualidade mínima no registro é essencial não só para a etapa da classificação, mas também para todas as análises.

## **5.5 - Aplicações da *Toolbox* para diversos modelos murinos e da doença de Parkinson**

O estudo sobre o impacto de diferentes doenças no sono tem atraído atenção devido ao alto potencial debilitante que distúrbios no sono podem gerar. Os pacientes com distúrbios no sono apresentam diversos quadros associados a prejuízos de memória, com sonolência excessiva e até agravamento de outros sintomas (Brownlow, Miller, and Gehrman 2020; Dauvilliers et al. 2018). A ideia de que modelos animais dessas doenças possam replicar esses distúrbios do sono tem motivado uma série de estudos realizados com modelos da doença de Alzheimer (O’Riordan, Hu,

and Rowan 2018; Schreiner et al. 2021; Zhang et al. 2016), de Parkinson (De Castro Medeiros et al. 2019; Stefani and Högl 2019), epilepsia (Mendes et al. 2019), entre outros. O desenvolvimento das técnicas de manipulação genética tem criado modelos animais de doenças cada vez mais confiáveis e carregados de possibilidades para investigações futuras que requerem um método de classificação de alta reprodutibilidade que esteja à altura desses trabalhos. A nossa *Toolbox* apresentou um acurácia elevado em registros de camundongos modelo da doença de Parkinson, o que a coloca como ferramenta importante para a expansão de pesquisas nesse campo. Os resultados preliminares (Figura 25) indicaram que a classificação feita pela *Toolbox* foi capaz de reproduzir os prejuízos de sono já descritos em outros trabalhos (Claassen 2011; Vo et al. 2014), e abre a possibilidade de investigações mais profundas sobre possíveis alterações na arquitetura do ciclo circadiano desses animais. Os resultados preliminares do registro classificado em rato sugerem que a nossa *Toolbox* possa ser utilizada também para experimentos neste outro modelo murino, que tem sido amplamente utilizado como modelo experimental de estudos do sono.

## 5.6 - Próximos passos

A *Toolbox* desenvolvida neste trabalho apresenta um acurácia satisfatório e oferece ao usuário um relatório completo da classificação, contendo representações gráficas e numéricas de todas as etapas. Além disso, desenvolvemos uma interface gráfica para facilitar a sua implementação nos diferentes laboratórios, e a disponibilizamos livremente. Acreditamos que o nosso método de classificação pode ser utilizado para diferentes modelos animais de doenças e é uma ferramenta poderosa para maior compreensão do ciclo circadiano. Ainda vamos comparar classificações de um mesmo registro feitas por pelo menos 3 usuários cegos para os grupos experimentais e/ou animais modelo para excluir a possibilidade de viés da classificação manual. Nas próximas edições da *Toolbox*, vamos oferecer análises que vão facilitar ainda mais as investigações, com explorações (1) sobre diferentes parâmetros da arquitetura do ciclo circadiano, como tempo gasto em cada estado e tamanho dos episódios (sequência de épocas) de cada estado, (2) sobre as características e ocorrência de eventos oscilatórios como ripples e spindles e (3) sobre marcadores eletrofisiológicos do ajuste homeostático (potência em delta normalizada

ao longo do tempo). Dessa forma, a *Toolbox* ampliará sua utilidade e poderá ser uma ferramenta central no campo, colaborando intensamente para a reprodutibilidade dos achados e o avanço do conhecimento sobre o sono.

## 6. Referências

- Adamantidis, Antoine, and Anita Lüthi. 2018. "Optogenetic Dissection of Sleep-Wake States In Vitro and In Vivo."
- Adamantidis, Antoine R. 2019. "Oscillating Circuitries in the Sleeping Brain." *Nature Reviews Neuroscience*. <http://dx.doi.org/10.1038/s41583-019-0223-4>.
- Alex, B Y, and J O H N Lon. 2019. "Artificial Intelligence and Black-Box Medical Decisions : Accuracy versus Explainability." : 15–21.
- Anafi, Ron C., Matthew S. Kayser, and David M. Raizen. 2019. "Exploring Phylogeny to Find the Function of Sleep." *Nature Reviews Neuroscience* 20(2): 109–16. <http://dx.doi.org/10.1038/s41583-018-0098-9>.
- Barger, Zeke et al. 2019. "Robust, Automated Sleep Scoring by a Compact Neural Network with Distributional Shift Correction." *PLoS ONE* 14(12): 1–18. <http://dx.doi.org/10.1371/journal.pone.0224642>.
- Bastide, Matthieu F. et al. 2015. "Pathophysiology of L-Dopa-Induced Motor and Non-Motor Complications in Parkinson's Disease." *Progress in Neurobiology* 132: 96–168.
- Benchenane, Karim et al. 2010. "Coherent Teta Oscillations and Reorganization of Spike Timing in the Hippocampal- Prefrontal Network upon Learning." *Neuron* 66(6): 921–36. <http://dx.doi.org/10.1016/j.neuron.2010.05.013>.
- Bjorness, Theresa E. et al. 2016. "An Adenosine-Mediated Glial-Neuronal Circuit for Homeostatic Sleep." *Journal of Neuroscience* 36(13): 3709–21.
- Boyce, Richard, Stephen D Glasgow, and Sylvain Williams. 2016. "Causal Evidence for the Role of REM Sleep Teta Rhythm in Contextual Memory Consolidation." 23(2013).
- Braak, Heiko et al. 2003. "Staging of Brain Pathology Related to Sporadic Parkinson's Disease.Pdf." 24: 197–211.
- Brankačk, Jurij, Valeriy I. Kukushka, Alexei L. Vyssotski, and Andreas Draguhn. 2010. "EEG Gamma Frequency and Sleep-Wake Scoring in Mice: Comparing Two Types of Supervised Classifiers." *Brain Research* 1322: 59–71.

- Brownlow, Janeese A., Katherine E. Miller, and Philip R. Gehrman. 2020. "Insomnia and Cognitive Acurácia." *Sleep Medicine Clinics* 15(1): 71–76. <https://doi.org/10.1016/j.jsmc.2019.10.002>.
- Bruijn, Berry De. 2011. "Revisiting the Area Under the ROC." : 532–36.
- Burdakov, Denis. 2020. "How Orexin Signals Bias Action: Hypothalamic and Accumbal Circuits." *Brain Research* 1731(September). <https://doi.org/10.1016/j.brainres.2018.09.011>.
- Buzsáki, György. 2010. "Neural Syntax: Cell Assemblies, Synapsembles, and Readers." *Neuron* 68(3): 362–85.
- . 2015. "Hippocampal Sharp Wave-Ripple: A Cognitive Biomarker for Episodic Memory and Planning." *Hippocampus* 25(10): 1073–1188.
- Buzsáki, György, Costas A. Anastassiou, and Christof Koch. 2012. "The Origin of Extracellular Fields and Currents-EEG, ECoG, LFP and Spikes." *Nature Reviews Neuroscience* 13(6): 407–20.
- Buzsáki, György, and David Tingley. 2018. "Space and Time: The Hippocampus as a Sequence Generator." *Trends in Cognitive Sciences* 22(10): 853–69. <https://doi.org/10.1016/j.tics.2018.07.006>.
- Buzsáki, György, and Xiao-Jing Wang. 2012. "Mechanisms of Gamma Oscillations." *Annual Review of Neuroscience* 35(1): 203–25.
- Cabral, Henrique O. et al. 2014. "Oscillatory Dynamics and Place Field Maps Reflect Hippocampal Ensemble Processing of Sequence and Place Memory under NMDA Receptor Control." *Neuron* 81(2): 402–15. <http://dx.doi.org/10.1016/j.neuron.2013.11.010>.
- Carvalho, Miguel M et al. 2017. "Effect of Levodopa on Reward and Impulsivity in a Rat Model of Parkinson ' s Disease." 11(August): 1–13.
- De Castro Medeiros, Daniel, Cleiton Lopes Aguiar, Márcio Flávio Dutra Moraes, and Gilberto Fisone. 2019. "Sleep Disorders in Rodent Models of Parkinson's Disease." *Frontiers in Pharmacology* 10(November): 1–12.
- Cho, Jounhong Ryan et al. 2017. "Dorsal Raphe Dopamine Neurons Modulate Arousal and Promote Wakefulness by Salient Stimuli." *Neuron* 94(6): 1205-1219.e8. <http://dx.doi.org/10.1016/j.neuron.2017.05.020>.
- Cirelli, Chiara, and Giulio Tononi. 2015. "Cortical Development, Electroencephalogram Rhythms, and the Sleep/Wake Cycle." *Biological Psychiatry* 77(12): 1071–78. <http://dx.doi.org/10.1016/j.biopsych.2014.12.017>.

- Ciric, Jelena, Slobodan Kapur, Milka Perovic, and Jasna Saponjic. 2019. "Alterations of Sleep and Sleep Oscillations in the Hemiparkinsonian Rat." *Frontiers in Neuroscience* 13(FEB).
- Citri, Ami, and Robert C. Malenka. 2008. "Synaptic Plasticity: Multiple Forms, Functions, and Mechanisms." *Neuropsychopharmacology* 33(1): 18–41.
- Claassen, Daniel. 2011. "Sleep Disturbances in Parkinson's Disease Patients and Management Options." *Nature and Science of Sleep*: 125.
- Cohen, Mike X. 2014. "Analyzing Neural Time Series Data." : 32 ، ص 117.
- Colgin, Laura Lee. 2016. "Rhythms of the Hippocampal Network." *Nature Reviews Neuroscience* 17(4): 239–49. <http://dx.doi.org/10.1038/nrn.2016.21>.
- . 2020. "Five Decades of Hippocampal Place Cells and EEG Rhythms in Behaving Rats." *Journal of Neuroscience* 40(1): 54–60.
- Crisler, Shelly, Michael J. Morrissey, A. Michael Anch, and David W. Barnett. 2008. "Sleep-Stage Scoring in the Rat Using a Support Vector Machine." *Journal of Neuroscience Methods* 168(2): 524–34.
- Crunelli, Vincenzo et al. 2018. "Dual Function of Thalamic Low-Vigilance State Oscillations: Rhythm-Regulation and Plasticity." *Nature Reviews Neuroscience* 19(2): 107–18.
- D. L. Commins, R. A. Shaughnessy, K. J. Axt, G. Vosmer, and L. S. Seiden. 1990. "Variability among Brain Regions in the Specificity of 6-Hydroxydopamine (6-OHDA)-Induced Lesions D." *Journal of Neural Transmission* 1874: 145–50.
- Dauvilliers, Yves et al. 2018. "REM Sleep Behaviour Disorder." *Nature Reviews Disease Primers* 4(1): 19. <http://www.nature.com/articles/s41572-018-0016-5>.
- Delorme, Arnaud, and Scott Makeig. 2004. "EEGLAB: An Código fechado *Toolbox* for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis." 134: 9–21.
- Diba, Kamran, Asohan Amarasingham, Kenji Mizuseki, and György Buzsáki. 2014. "Millisecond Timescale Synchrony among Hippocampal Neurons." *Journal of Neuroscience* 34(45): 14984–94.
- Drieu, Céline, Ralitsa Todorova, and Michaël Zugaro. 2018. "Nested Sequences of Hippocampal Assemblies During Behavior Support Subsequent Sleep Replay." *Science* 679(November): 675–79.
- Drieu, Céline, and Michaël Zugaro. 2019. "Hippocampal Sequences during Exploration: Mechanisms and Functions." *Frontiers in Cellular Neuroscience*

- 13(June): 1–22.
- Durán, Ernesto et al. 2018. “Sleep Stage Dynamics in Neocortex and Hippocampus.” *Sleep* 41(6): 1–11.
- Eban-Rothschild, Ada et al. 2016. “VTA Dopaminergic Neurons Regulate Ethologically Relevant Sleep-Wake Behaviors.” *Nature Neuroscience* 19(10): 1356–66.
- Fantini, Maria Livia et al. 2018. “Sleep and REM Sleep Behaviour Disorder in Parkinson’s Disease with Impulse Control Disorder.” : 305–10.
- Faust, Oliver et al. 2019. “A Review of Automated Sleep Stage Scoring Based on Physiological Signals for the New Millennia.” *Computer Methods and Programs in Biomedicine* 176: 81–91.
- Frommlet, Florian. 2020. “Improving Reproducibility in Animal Research.” *Scientific Reports*: 3–4. <https://doi.org/10.1038/s41598-020-76398-3>.
- Gao, Vance, Fred Turek, and Martha Vitaterna. 2016. “Multiple Classifier Systems for Automatic Sleep Scoring in Mice.” *Journal of Neuroscience Methods* 264: 33–39. <http://dx.doi.org/10.1016/j.jneumeth.2016.02.016>.
- Gaston, Kevin J., Marcel E. Visser, and Franz Hölker. 2015. “The Biological Impacts of Artificial Light at Night: The Research Challenge.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 370(1667).
- Geoffrey McLachlan, David Peel. 2000. “Finite Mixture Models.”
- Girardeau, Gabrielle et al. 2009. “Selective Suppression of Hippocampal Ripples Impairs Spatial Memory.” *Nature Neuroscience* 12(10): 1222–23.
- Girardeau, Gabrielle, and Michaël Zugaro. 2011. “Hippocampal Ripples and Memory Consolidation.” *Current Opinion in Neurobiology* 21(3): 452–59.
- Gros, Priti, and Aleksandar Videnovic. 2017. “Sleep and Circadian Rhythm Disorders in Parkinson’s Disease.” *Current Sleep Medicine Reports* 3(3): 222–34.
- Grosmark, Andres D. et al. 2012. “REM Sleep Reorganizes Hippocampal Excitability.” *Neuron* 75(6): 1001–7.
- Gross, Brooks A. et al. 2009. “Código fechado Logic-Based Automated Sleep Scoring Software Using Electrophysiological Recordings in Rats.” *Journal of Neuroscience Methods* 184(1): 10–18.
- Guyon, Nicolas et al. 2020. “Network Asynchrony Underlying Increased Broadband Gamma Power Authors.”
- Hainmueller, Thomas, and Marlene Bartos. 2020. “Dentate Gyrus Circuits for Encoding, Retrieval and Discrimination of Episodic Memories.” *Nature Reviews*

- Neuroscience* 21(3): 153–68. <http://dx.doi.org/10.1038/s41583-019-0260-z>.
- Harris, Kenneth D. 2005. “Neural Signatures of Cell Assembly Organization.” *Nature Reviews Neuroscience* 6(5): 399–407.
- Hastings, Michael H., Elizabeth S. Maywood, and Marco Brancaccio. 2018. “Generation of Circadian Rhythms in the Suprachiasmatic Nucleus.” *Nature Reviews Neuroscience* 19(8): 453–69. <http://dx.doi.org/10.1038/s41583-018-0026-z>.
- Herreras, Oscar. 2016. “Local Field Potentials: Myths and Misunderstandings.” 10(December): 1–16.
- Hoo, Zhe Hui, Jane Candlish, and Dawn Teare. 2017. “What Is an ROC Curve ?” : 357–59.
- Iyer, Vishakh et al. 2019. “Acute Levodopa Dosing Around-the-Clock Ameliorates REM Sleep without Atonia in Hemiparkinsonian Rats.” *npj Parkinson’s Disease* 5(1): 1–11. <http://dx.doi.org/10.1038/s41531-019-0096-2>.
- Jain, Anil K. 2010. “Data Clustering: 50 Years beyond K-Means Q.” *Pattern Recognition Letters* 31(8): 651–66. <http://dx.doi.org/10.1016/j.patrec.2009.09.011>.
- JIANG, Dihong, Yu MA, and Yuanyuan WANG. 2019. “Sleep Stage Classification Using Covariance Features of Multi-Channel Physiological Signals on Riemannian Manifolds.” *Computer Methods and Programs in Biomedicine* 178: 19–30.
- K. Pavlova, Milena, and Véronique Latreille. 2019. “Sleep Disorders.” *American Journal of Medicine* 132(3): 292–99.
- Kaiser, Walter, and Jana Steiner-Kaiser. 1983. “Neuronal Correlates of Sleep, Wakefulness and Arousal in a Diurnal Insect.” *Nature* 301(5902): 707–9.
- Kajikawa, Yoshinao, and Charles E. Schroeder. 2011. “How Local Is the Local Field Potential?” *Neuron* 72(5): 847–58. <http://dx.doi.org/10.1016/j.neuron.2011.09.029>.
- Katsageorgiou, Vasiliki Maria et al. 2015. “Sleep-Stage Scoring in Mice: The Influence of Data Pre-Processing on a System’s Acurácia.” *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2015-Novem*: 598–601.
- . 2018. “A Novel Unsupervised Analysis of Electrophysiological Signals Reveals New Sleep Substages in Mice.” *PLoS Biology* 16(5): 1–23.
- Kirik, Deniz, Carl Rosenblad, and Anders Björklund. 1998. “Characterization of

- Behavioral and Neurodegenerative Changes Following Partial Lesions of the Nigrostriatal Dopamine System Induced by Intrastratial 6-Hydroxydopamine in the Rat." *Experimental Neurology* 152(2): 259–77.
- Klinzing, Jens G, Niels Niethard, and Jan Born. 2019. "Mechanisms of Systems Memory Consolidation during Sleep." *Nature Neuroscience*. <http://dx.doi.org/10.1038/s41593-019-0467-3>.
- Koike, Bruna Del Vecchio et al. 2017. "Electrophysiological Evidence That the Retrosplenial Cortex Displays a Strong and Specific Activation Phased with Hippocampal Theta during Paradoxical (REM) Sleep." *Journal of Neuroscience* 37(33): 8003–13.
- Koronowski, Kevin B., and Paolo Sassone-Corsi. 2021. "Communicating Clocks Shape Circadian Homeostasis." *Science* 371(6530): eabd0951.
- Kroeger, Daniel et al. 2017. "Cholinergic, Glutamatergic, and GABAergic Neurons of the Pedunculopontine Tegmental Nucleus Have Distinct Effects on Sleep/Wake Behavior in Mice." *Journal of Neuroscience* 37(5): 1352–66.
- De La Prida, Liset M. 2020. "Potential Factors Influencing Replay across CA1 during Sharp-Wave Ripples." *Philosophical Transactions of the Royal Society B: Biological Sciences* 375(1799).
- Lampert, Thomas, Andrea Plano, Jim Austin, and Bettina Platt. 2015. "On the Identification of Sleep Stages in Mouse Electroencephalography Time-Series." *Journal of Neuroscience Methods* 246: 52–64. <http://dx.doi.org/10.1016/j.jneumeth.2015.03.007>.
- Lamprecht, Raphael, and Joseph LeDoux. 2004. "Structural Plasticity and Memory." *Nature Reviews Neuroscience* 5(1): 45–54.
- Latchoumane, Charles Francois V., Hong Viet V. Ngo, Jan Born, and Hee Sup Shin. 2017. "Thalamic Spindles Promote Memory Formation during Sleep through Triple Phase-Locking of Cortical, Thalamic, and Hippocampal Rhythms." *Neuron* 95(2): 424-435.e6. <http://dx.doi.org/10.1016/j.neuron.2017.06.025>.
- Li, Hongyang, and Yuanfang Guan. 2021. "DeepSleep Convolutional Neural Network Allows Accurate and Fast Detection of Sleep Arousal." *Communications Biology* 4(1): 1–11. <http://dx.doi.org/10.1038/s42003-020-01542-8>.
- Liu, Jiabei et al. 2016. "MT1 and MT2 Melatonin Receptors: A Therapeutic Perspective." *Annual Review of Pharmacology and Toxicology* 56(October 2015): 361–83.



- Lopes-aguiar, Cleiton et al. 2020. “Long-Term Potentiation Prevents Ketamine-Induced Aberrant Neurophysiological Dynamics in the Hippocampus-Prefrontal Cortex Pathway in Vivo.” : 1–15.
- Luo, Yan Jia et al. 2018. “Nucleus Accumbens Controls Wakefulness by a Subpopulation of Neurons Expressing Dopamine D1 Receptors.” *Nature Communications* 9(1). <http://dx.doi.org/10.1038/s41467-018-03889-3>.
- Maingret, Nicolas et al. 2016. “Hippocampo-Cortical Coupling Mediates Memory Consolidation during Sleep.” *Nature Neuroscience* 19(7): 959–64. <http://www.nature.com/doi/10.1038/nn.4304>.
- Mander, Bryce A, Joseph R Winer, and Matthew P Walker. 2017. “Review Sleep and Human Aging.” *Neuron* 94(1): 19–36. <http://dx.doi.org/10.1016/j.neuron.2017.02.004>.
- Marshall, Lisa, Nathan Cross, Sonja Binder, and Thien Thanh Dang-Vu. 2020. “Brain Rhythms during Sleep and Memory Consolidation: Neurobiological Insights.” *Physiology* 35(1): 4–15.
- Maurice Abou Jaoude, Haoqi Sun, Kyle R. Pellerin, Milena Pavlova, Rani A. Sarkis, Sydney S. Cash, M. Brandon Westover\*, Alice D. Lam\*. 2018. “Expert-Level Automated Sleep Staging of Long-Term Scalp EEG Recordings Using Deep Learning Maurice.” 40: 1–30.
- Mendes, Renan Augusto Viana et al. 2019. “Hijacking of Hippocampal–Cortical Oscillatory Coupling during Sleep in Temporal Lobe Epilepsy.” *Epilepsy and Behavior* (xxxx): 106608. <https://doi.org/10.1016/j.yebeh.2019.106608>.
- Miyawaki, Hiroyuki, and Kamran Diba. 2016. “Regulation of Hippocampal Firing by Network Oscillations during Sleep.” *Current Biology* 26(7): 893–902. <http://dx.doi.org/10.1016/j.cub.2016.02.024>.
- Miyawaki, Hiroyuki, Brendon O. Watson, and Kamran Diba. 2019. “Neuronal Firing Rates Diverge during REM and Homogenize during Non-REM.” *Scientific Reports* 9(1): 1–14.
- Nasraoui, Olfa. 2019. *Clustering Methods for Big Data Analytics*.
- Nath, Tanmay et al. 2019. “Using DeepLabCut for 3D Markerless Pose Estimation across Species and Behaviors.” *Nature Protocols* 14(July). <http://dx.doi.org/10.1038/s41596-019-0176-0>.
- Neckelmann, D., O. E. Olsen, S. Fagerland, and R. Ursin. 1994. “The Reliability and Functional Validity of Visual and Semiautomatic Sleep/Wake Scoring in the Moll-

- Wistar Rat." *Sleep* 17(2): 120–31.
- Nicoll, Roger A. 2017. "A Brief History of Long-Term Potentiation." *Neuron* 93(2): 281–90. <http://dx.doi.org/10.1016/j.neuron.2016.12.015>.
- Norimoto, Hiroaki et al. 2018. "Hippocampal Ripples Down-Regulate Synapses." *Science* 359(6383): 1524–27.
- O’Riordan, Kenneth J., Neng Wei Hu, and Michael J. Rowan. 2018. "A $\beta$  Facilitates LTD at Schaffer Collateral Synapses Preferentially in the Left Hippocampus." *Cell Reports* 22(8): 2053–65. <https://doi.org/10.1016/j.celrep.2018.01.085>.
- Ohayon, Maurice M. 2002. "Epidemiology of Insomnia: What We Know and What We Still Need to Learn." *Sleep Medicine Reviews* 6(2): 97–111.
- Oishi, Yo et al. 2017. "Slow-Wave Sleep Is Controlled by a Subset of Nucleus Accumbens Core Neurons in Mice." *Nature Communications* 8(1): 1–12.
- Pinnell, R C, R K Almajidy, and U G Hofmann. 2016. "Versatile 3D-Printed Headstage Implant for Group Housing of Rodents." *Journal of Neuroscience Methods* 257: 134–38. <http://dx.doi.org/10.1016/j.jneumeth.2015.09.027>.
- Raadt, Alexandra De, Matthijs J Warrens, Roel J Bosker, and Henk A L Kiers. 2019. "Kappa Coefficients for Missing Data."
- Rytkönen, Kirsi-marja, Jukka Zitting, and Tarja Porkka-heiskanen. 2011. "Automated Sleep Scoring in Rats and Mice Using the Naive Bayes Classifier." 202: 60–64.
- Saito, Yuki C. et al. 2018. "Serotonergic Input to Orexin Neurons Plays a Role in Maintaining Wakefulness and REM Sleep Architecture." *Frontiers in Neuroscience* 12(NOV): 1–9.
- Sataloff, Robert T, Michael M Johns, and Karen M Kost. 2008. "The Elements of Statistical Learning." *Springer Series in Statistics*.
- Satterlie, Richard A. 2011. "Do Jellyfish Have Central Nervous Systems?" *Journal of Experimental Biology* 214(8): 1215–23.
- Scammell, Thomas E., Elda Arrigoni, and Jonathan O. Lipton. 2017. "Neural Circuitry of Wakefulness and Sleep." *Neuron* 93(4): 747–65. <http://dx.doi.org/10.1016/j.neuron.2017.01.014>.
- Schapira, Anthony H.V., K. Ray Chaudhuri, and Peter Jenner. 2017. "Non-Motor Features of Parkinson Disease." *Nature Reviews Neuroscience* 18(7): 435–50. <http://dx.doi.org/10.1038/nrn.2017.62>.
- Schreiner, Simon J et al. 2021. "Reduced Regional NREM Sleep Slow-Wave Activity Is Associated With Cognitive Impairment in Parkinson Disease." 12(February): 1–

8.

- Siegel, Jerome M. 2008. "Do All Animals Sleep?" *Trends in Neurosciences* 31(4): 208–13.
- Silva-Pérez, Manuel, Alvaro Sánchez-López, Natalia Pompa-del-Toro, and Miguel Escudero. 2020. "Identification of the Sleep–Wake States in Rats Using the High-Frequency Activity of the Electroencephalogram." *Journal of Sleep Research* (October): 1–10.
- Silvani, Alessandro et al. 2017. "Muscle Activity During Sleep in Human Subjects , Rats , and Mice : Towards Translational Models of REM Sleep Without Atonia."
- Sirota, Anton et al. 2008. "Entrainment of Neocortical Neurons and Gamma Oscillations by the Hippocampal Teta Rhythm." *Neuron* 60(4): 683–97.
- Soltani, Sara et al. 2019. "Sleep–Wake Cycle in Young and Older Mice." *Frontiers in Systems Neuroscience* 13(September): 1–14.
- Souza, Annie C. et al. 2018. "Recording Day and Night: Advice for New Investigators in the Sleep and Memory Field." *Handbook of Behavioral Neuroscience* 28: 43–62.
- Stefani, Ambra, and Birgit Högl. 2019. "Sleep in Parkinson's Disease." *Neuropsychopharmacology* (March): 1–8.
- Stott, Simon R W, and Roger A. Barker. 2014. "Time Course of Dopamine Neuron Loss and Glial Response in the 6-OHDA Striatal Mouse Model of Parkinson's Disease." *European Journal of Neuroscience* 39(6): 1042–56.
- Takahashi, Muneyoshi, Hiroshi Nishida, A. David Redish, and Johan Lauwereyns. 2014. "Teta Phase Shift in Spike Timing and Modulation of Gamma Oscillation: A Dynamic Code for Spatial Alternation during Fixation in Rat Hippocampal Area CA1." *Journal of Neurophysiology* 111(8): 1601–14.
- Tononi, Giulio, and Chiara Cirelli. 2014. "Sleep and the Price of Plasticity: From Synaptic and Cellular Homeostasis to Memory Consolidation and Integration." *Neuron* 81(1): 12–34. <http://dx.doi.org/10.1016/j.neuron.2013.12.025>.
- . 2020. "Sleep and Synaptic Down-Selection." *European Journal of Neuroscience* 51(1): 413–21.
- Del Tredici, K., and H. Braak. 2016. "Sporadic Parkinson's Disease: Development and Distribution of  $\alpha$ -Synuclein Pathology." *Neuropathology and Applied Neurobiology* 42(1): 33–50.
- Unal, Ilker. 2017. "Defining an Optimal Cut-Point Value in ROC Analysis: An

- Alternative Approach.” 2017.
- Varga, Viktor et al. 2008. “The Presence of Pacemaker HCN Channels Identifies Teta Rhythmic GABAergic Neurons in the Medial Septum.” *Journal of Physiology* 586(16): 3893–3915.
- Vladimir, Svetnik et al. 2020. “A Deep Learning Approach for Automated Sleep-Wake Scoring in Pre-Clinical Animal Models.” *Journal of Neuroscience Methods* 337(February): 108668. <https://doi.org/10.1016/j.jneumeth.2020.108668>.
- Vo, Quynh et al. 2014. “Polysomnographic Features of Sleep Disturbances and Rem Sleep Behavior Disorder in the Unilateral 6-OHDA Lesioned Hemiparkinsonian Rat.” *Parkinson’s Disease* 2014.
- de Vries, Ingmar E.J., Heleen A. Slagter, and Christian N.L. Olivers. 2020. “Oscillatory Control over Representational States in Working Memory.” *Trends in Cognitive Sciences* 24(2): 150–62. <https://doi.org/10.1016/j.tics.2019.11.006>.
- Vyazovskiy, Vladyslav V. et al. 2009. “Cortical Firing and Sleep Homeostasis.” *Neuron* 63(6): 865–78. <http://dx.doi.org/10.1016/j.neuron.2009.08.024>.
- Wickwire, Emerson M., Fadia T. Shaya, and Steven M. Scharf. 2016. “Health Economics of Insomnia Treatments: The Return on Investment for a Good Night’s Sleep.” *Sleep Medicine Reviews* 30: 72–82. <http://dx.doi.org/10.1016/j.smr.2015.11.004>.
- Yamabe, Masato et al. 2019. “MC-SleepNet: Large-Scale Sleep Stage Scoring in Mice by Deep Neural Networks.” *Scientific Reports* 9(1): 1–12.
- Yamamoto, Jun, and Susumu Tonegawa. 2017. “Direct Medial Entorhinal Cortex Input to Hippocampal CA1 Is Crucial for Extended Quiet Awake Replay.” *Neuron* 96(1): 217–227.e4. <https://doi.org/10.1016/j.neuron.2017.09.017>.
- Yildirim, Ozal, Ulas Baran Baloglu, and U. Rajendra Acharya. 2019. “A Deep Learning Model for Automated Sleep Stages Classification Using PSG Signals.” *International Journal of Environmental Research and Public Health* 16(4).
- Yu, Xiao et al. 2019. “GABA and Glutamate Neurons in the VTA Regulate Sleep and Wakefulness.” *Nature Neuroscience* 22(1): 106–19. <http://dx.doi.org/10.1038/s41593-018-0288-9>.
- Zant, Janneke C. et al. 2016. “Cholinergic Neurons in the Basal Forebrain Promote Wakefulness by Actions on Neighboring Non-Cholinergic Neurons: An Opto-Dialysis Study.” *Journal of Neuroscience* 36(6): 2057–67.
- Zeng, Tao, Christopher Mott, Daniel Mollicone, and Larry D Sanford. 2012. “Automated

Determination of Wakefulness and Sleep in Rats Based on Non-Invasively Acquired Measures of Movement and Respiratory Activity.” *Journal of Neuroscience Methods* 204(2): 276–87.  
<http://dx.doi.org/10.1016/j.jneumeth.2011.12.001>.

Zhang, Xiaomin et al. 2016. “Impaired Teta-Gamma Coupling in APP-Deficient Mice.” *Scientific Reports* 6: 1–10. <http://dx.doi.org/10.1038/srep21948>.

Zhang, Xiaomin, Alois Schlögl, and Peter Jonas. 2020. “Selective Routing of Spatial Information Flow from Input to Output in Hippocampal Granule Cells.” *Neuron* 107(6): 1212-1225.e7.

## Apêndice 1 - Interface Gráfica

O objetivo da criação da interface gráfica foi facilitar o uso da *Toolbox*, limitando a interação do usuário com painéis simples que reflitam a ordem de execução dos passos até o resultado final. A interface gráfica do usuário (GUI) é iniciada após o usuário executar o arquivo de inicialização de extensão MLAPP.

A primeira janela aberta é a interface primária (Figura Suplementar 1), que consiste em painéis separados para cada uma das funções principais do software. A caixa de texto *Status* (Figura Suplementar 1-D, último painel) contém informações sobre o andamento dos processos inicializados a partir da interface principal. Os textos *Busy* e *Inactive* serão apresentados na caixa caso algum processo esteja ou não sendo executado, respectivamente.

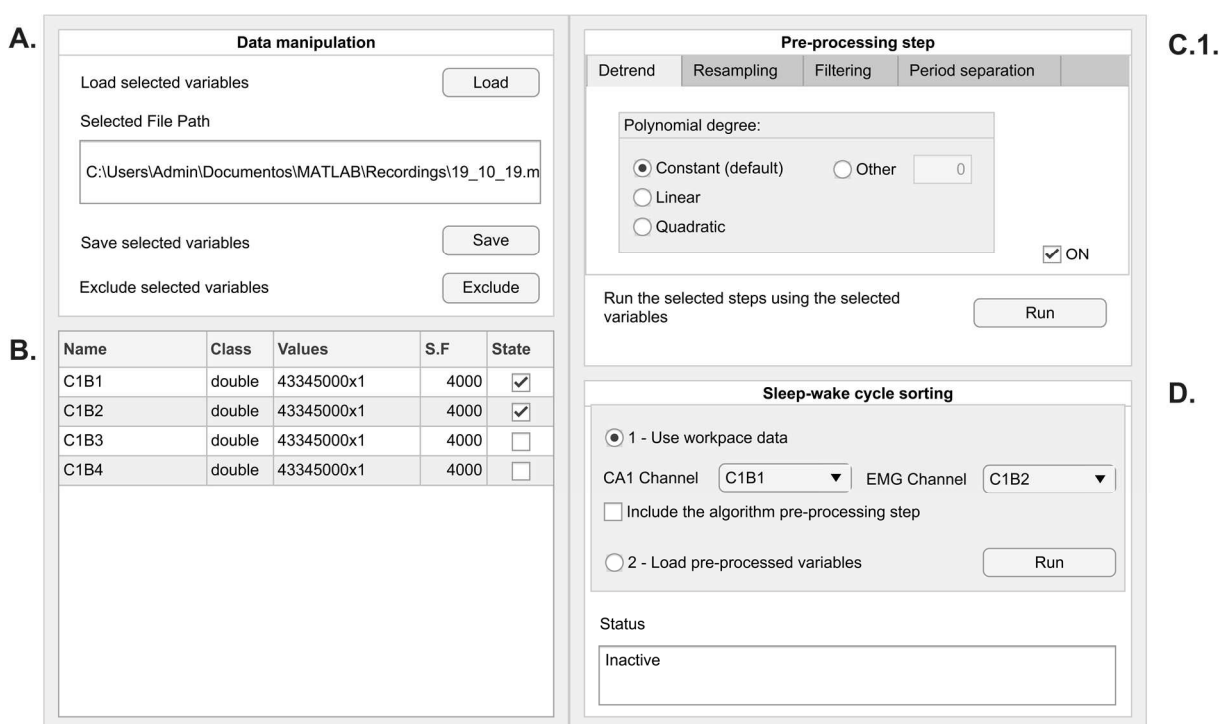


Figura Suplementar 1. Interface principal. A. Painel Data Manipulation. Opções para carregar os arquivos que serão utilizados na classificação; B. Painel Workspace. Demonstra as variáveis já carregadas que serão utilizadas; C. Painel Pre-processing Steps. Configurações das etapas de pré-processamento; D. Painel Sleep-wake cycle sorting. Designa quais os arquivos para a classificação.

O painel *Data manipulation* (Figura Suplementar 1-A.) agrupa funções que permitem que os dados sejam carregados, salvos ou excluídos. Ao clicar no botão

*Load*, uma janela de seleção de arquivos é aberta, permitindo que o usuário selecione arquivos com a extensão MAT. Após a seleção do arquivo, uma nova janela é aberta (Figura Suplementar 2), permitindo que o usuário visualize todas as variáveis armazenadas no arquivo selecionado.

Check the variables which should be loaded:

Name	Size	Class	Load State	<input checked="" type="checkbox"/> All	Sampling Frequency	<input checked="" type="checkbox"/> All
C1B1	43345000x1	double	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	
C1B2	43345000x1	double	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	
C1B3	43345000x1	double	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	
C1B4	43345000x1	double	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	

Selected File: C:\Users\Admin\Documents\MATLAB\Recordings\19\_10\_19.mat

Sampling Frequency

Variables 
 Manual 
 None

Figura Suplementar 2. Painel Load. Seleção dos arquivos .mat a serem abertos.

Informações básicas sobre essas variáveis, como nome (*Name*), dimensões (*Size*) e classe (*Class*) são exibidas na janela (Figura Suplementar 2), aceitando que o usuário selecione quais variáveis serão carregadas ao marcar/desmarcar as caixas de seleção presentes na coluna *Load State*. Por fim, a coluna *Sampling Frequency* contém caixas de seleção que indicam quais variáveis possuem uma taxa de amostragem associada. O conjunto de botões de opção (*Sampling Frequency*) definem a forma pela qual o valor de taxa de amostragem será associado às variáveis, sendo que:

- 1 - *Variables* indica que uma das variáveis da lista contém o valor de taxa de amostragem e a caixa de listagem ao lado permite a seleção da variável em questão;

2 - *Manual* indica que o valor da taxa de amostragem será inserido manualmente na caixa de edição ao lado;

3 - *None* indica que nenhum valor de taxa de amostragem será associado às variáveis.

Posteriormente, ao clicar em *Load*, as variáveis selecionadas serão carregadas, contendo ou não a taxa de amostragem associada. Em contrapartida, ao clicar em *Cancel*, o carregamento é cancelado e o usuário será direcionado à interface principal.

Na interface principal (Figura Suplementar 1), abaixo do painel *Data Manipulation*, há um painel (Figura 30-B) com a finalidade de simular a *Área de trabalho* do MATLAB, sendo um espaço para exibir, em colunas, informações sobre as variáveis carregadas, como nome (*Name*), classe (*Class*), valores (*Values*) ou dimensão caso seja um vetor, taxa de amostragem associada (*S.F.*) e o estado de seleção (*State*). A última coluna define quais variáveis serão consideradas (caixa de seleção marcada) ou não (caixa de seleção desmarcada) nos processos de exclusão, salvar ou de pré-processamento.

Ainda no painel *Data manipulation*, ao clicar no botão *Save* (Figura Suplementar 1-A), uma janela de seleção de diretório será aberta, permitindo que o usuário escolha o local em que as variáveis serão salvas. Contudo, apenas as variáveis cujo estado de seleção esteja marcado (variáveis marcadas) serão salvas. De forma semelhante, ao clicar no botão *Exclude*, as variáveis marcadas serão excluídas do painel *Workspace*, deixando também de ocupar espaço na memória.

O painel *Pre-processing step* (Figura Suplementar 1-B) consiste em 4 abas, sendo que cada uma delas agrupa um conjunto de informações relacionadas a uma das possíveis funções constituintes da etapa de pré-processamento. A aba *Detrend* (Figura Suplementar 1-B) permite ao usuário controlar a função de remoção de tendências polinomiais do sinal. Nela, o painel *Polynomial degree* apresenta botões de seleção que modificam o grau polinomial utilizado pela função. Caso a opção *Constant* seja selecionada, o grau polinomial é igual a 0; *Linear*, igual a 1; *Quadratic*, igual a 2; caso *Other*, permite que o usuário insira o grau na caixa de texto ao lado.



**c.2** Detrend Resampling Filtering Period separation

Output sampling frequency (Hz)  Input/Output frequency ratio

Change input sampling frequency (Hz)

ON

**c.2** Detrend Resampling Filtering Period separation

Filter type

High-pass (Hz)   Band-pass

Low-pass (Hz)   Notch (Hz)

Sampling frequency

ON

**c.3** Detrend Resampling Filtering Period separation

Period Length (sec)   Sampling frequency

Period Length (samples)

Number of Periods

Number of discarded samples:

ON

Figura Suplementar 3. Abas com outras das configurações para o pré-processamento. C.1. Aba Resampling. Redefinição da taxa de amostragem; C.2. Aba Filtering. Configurações da filtragem do sinal; C.3. Aba Period separation. Determinar tamanho das épocas.

A aba *Resampling* (Figura Suplementar 3-C.1) permite ao usuário controlar a função de modificar a taxa de amostragem do sinal. Na caixa de texto *Output sampling frequency* o usuário pode inserir a taxa de amostragem final que o dado terá após ser processado. Contudo, o usuário também pode inserir a razão entre a taxa de amostragem inicial e final na caixa de texto *Input/Output frequency ratio*. Independentemente da caixa de texto editada, o valor da outra será automaticamente atualizado. Por definição padrão, a taxa de amostragem inicial é o valor encontrado associado à variável na coluna *S.F* do painel *Workspace*. Contudo, a caixa de seleção *Change input sampling frequency* permite que o usuário insira um outro valor para a taxa de amostragem inicial que não aquela presente na área de trabalho.

A aba *Filtering* (Figura Suplementar 3-C.2) permite que o usuário controle a função de filtragem do sinal. No painel *Filter type*, há botões de opção que definem o tipo de filtro que será executado: *High-pass*, indica ao algoritmo que será um filtro passa-alta; *Low-pass*, um filtro passa-baixa; *Band-pass*, um filtro passa-banda; *Notch*, um filtro notch. Caso as opções *High-pass*, *Low-pass* ou *Notch* sejam selecionadas, apenas a caixa de texto referente a cada uma delas será habilitada, permitindo que o usuário informe a frequência de corte para o filtro. No caso da opção *Band-pass*, as caixas de texto referentes ao *High-pass* e *Low-pass* serão habilitadas, contendo as frequências de limite inferior e superior, respectivamente. Ainda, a taxa de

amostragem de cada variável é definida pela coluna *S.F.* da área de trabalho, porém, pode ser redefinida ao marcar a caixa de seleção *Sampling Frequency* e adicionar um novo valor na caixa de texto ao lado.

Por último, a aba *Period separation* (Figura Suplementar 3-C.3) controla a separação do dado em épocas de duração definida. A caixa de texto *Period Length* é a única que pode ser editada, definindo a duração das épocas em segundos. A caixa de texto *Period Length* indica a duração das épocas em número de amostras, a caixa de texto *Number of Periods* indica o número total de épocas, e a caixa de texto *Number of discarded samples* indica o número de amostras descartadas.

Em cada uma das abas há um caixa de seleção *ON*, que define a função presente na aba está ativa e será executada quando o botão *Run* for pressionado. Ao desmarcar a caixa de seleção, o texto *OFF* aparecerá como resposta. Ainda, apenas as variáveis cuja caixa de seleção da coluna *State* estejam marcadas serão pré-processadas quando o botão *Run* for pressionado. A ordem do pré-processamento segue a ordem das abas, e assim que cada um dos passos for finalizado, um aviso será apresentado ao usuário.

No painel *Sleep-wake cycle sorting* (Figura Suplementar 1-C) há um grupo de botões de opção que definem a origem dos dados que serão utilizados pelo algoritmo de classificação de estágios do ciclo sono-vigília.

- A opção 1 - *Use workspace data* define que os dados serão provenientes da área de trabalho (Figura Suplementar 1-B), sendo necessário que o usuário selecione os registros de LFP de CA1 e do eletromiograma através das caixas de listagem *CA1 Channel* e *EMG Channel*, respectivamente. Ambas as caixas de listagem são atualizadas automaticamente com os nomes das variáveis presentes na área de trabalho. Ademais, a caixa de seleção *Include the algorithm pre-processing step* define se o pré-processamento será (marcada) ou não (desmarcada) realizado automaticamente de acordo com os padrões que melhor se ajustam ao algoritmo de classificação. Nesse caso, não é necessário que o pré-processamento do painel *Pre-processing step* seja realizado.

- A opção 2 - *Load pre-processed variables* define que o algoritmo de classificação utilizará um arquivo de nome *data\_variables.mat*, que contém os dados originados pelo pré-processamento realizado pelo algoritmo de classificação. Portanto, essa opção só é válida caso o algoritmo de classificação já tenha gerado o arquivo para o conjunto de dados em questão. Por fim, ao clicar no botão *Run*, caso a opção 1 tenha sido selecionada, as variáveis selecionadas nas caixas de listagem serão transferidas ao algoritmo de classificação. Caso a opção 2 tenha sido selecionada, uma janela de seleção de arquivos será aberta, requisitando que o usuário selecione o arquivo *data\_variables.mat* correspondente.

Posteriormente, após finalizado o pré-processamento dos dados, uma nova janela será aberta (Figura Suplementar 4). Assim como a interface principal, há também uma caixa de texto *Status*, que contém textos indicativos dos processos que estão sendo executados pelo algoritmo.

The screenshot shows a software window with the following elements:

- Output Path:** A button labeled "Output Path" and a text box containing the path "C:\Users\Documents\MATLAB\Recordings".
- Recording Parameters:** A panel containing:
  - Animal Group:** A dropdown menu currently set to "Control".
  - Noise Range (Hz) - Only for figures:** Two input fields: "Inferior" with the value "45" and "Superior" with the value "55".
  - Recording Time (Hours):** Two input fields: "Beginning" with the value "20" and "End" with the value "19".
- Run visual inspection:** A panel with two checked checkboxes:
  - Save some representative epochs
  - Run visual inspection (Mandatory for the first time)  
It might take a bit long because the epochs are going to be loadedAn unchecked checkbox is also present:
  - Continue an unfinished visual inspection
- OK:** A button at the bottom right of the window.
- Status:** A label and a text box with a scroll bar at the bottom left.

Figura Suplementar 4. Janela com informações das figuras de interação com o usuário.

Além disso, há o botão *Output Path*, que ao ser pressionado, abre uma janela de seleção de diretório. Nesse caso, o diretório selecionado será o destino para todos os dados e figuras produzidas pelo algoritmo de classificação, sendo apresentado na caixa de texto ao lado do botão. O painel *Recording Parameters* inclui informações

sobre os registros fornecidos pelo usuário: a caixa de listagem *Animal Group* apresenta as opções *Control* e *Experimental*, definindo se os registros foram realizados em animais controle ou experimentais, respectivamente; as caixas de texto do grupo *Noise Range* definem o limite mínimo (*Inferior*) e máximo (*Superior*), em Hertz, para o possível ruído de rede presente no sinal; e o conjunto de caixas *Recording Time* definem o horário de início (*Beginning*) e término (*End*), em horas, dos registros.

A caixa de seleção *Save some representative epochs* define se após finalizada a classificação, o algoritmo vai salvar (marcada) ou não (desmarcada) figuras com períodos representativos de cada um dos estados. Já no painel *Run visual inspection*, a caixa de seleção *Run visual inspection* define se uma etapa de inspeção visual pelo usuário será (marcada) ou não (desmarcada) executada. Por último, a caixa de seleção: *Continue an unfinished visual inspection*, define se o usuário quer (marcada) ou não (desmarcada) continuar uma inspeção visual não finalizada.

O algoritmo de classificação fica ocioso até que o botão *OK* seja pressionado, fazendo com que todos os botões, caixas de texto, caixas de seleção e caixas de listagem sejam bloqueados, impedindo que o usuário modifique essas informações. Posteriormente, o algoritmo continuará sua execução, atualizando a caixa de texto *Status* conforme necessário. Todas as figuras geradas pelo algoritmo são criadas no plano de fundo e salvas automaticamente no diretório informado como destino ao pressionar o botão *Output Path* (Figura Suplementar 4).

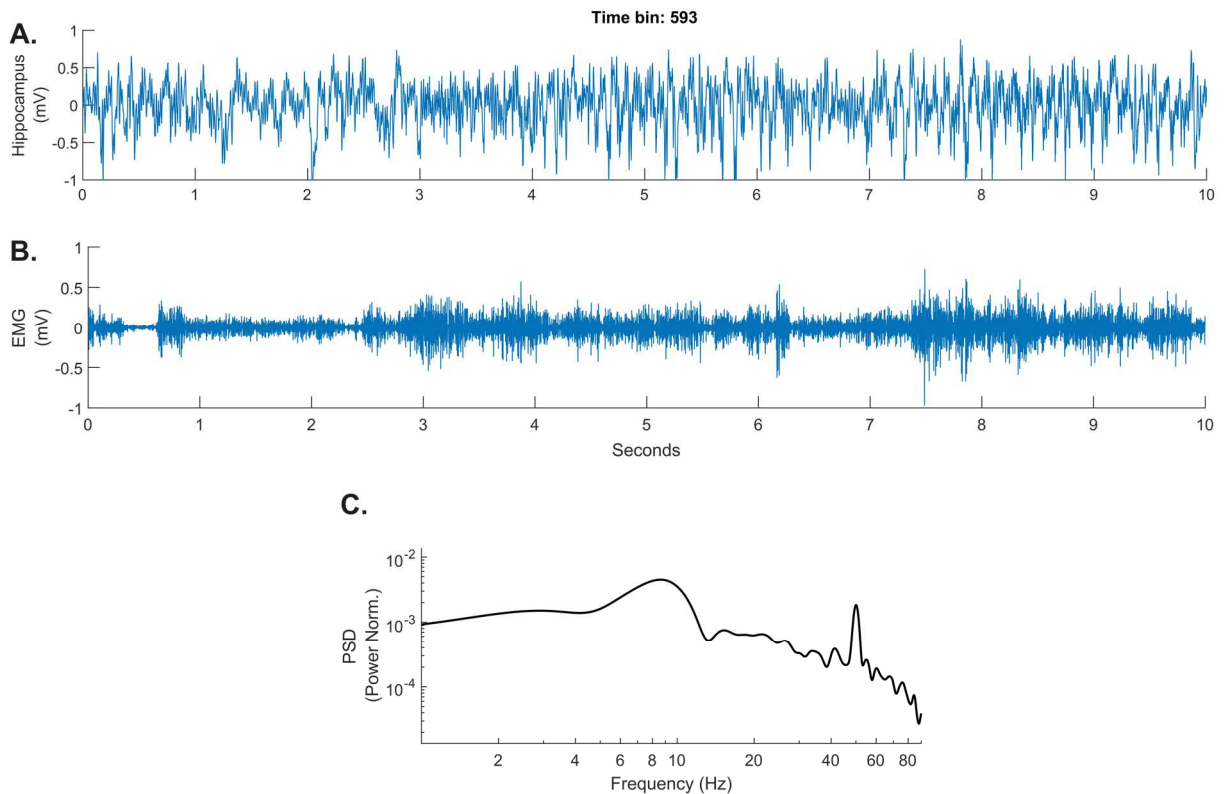
Após o início da classificação, uma outra janela é aberta (Figura Suplementar 5). Nela, o usuário vai indicar qual banda de frequência será adicionada à razão Teta/Delta durante a classificação. Contudo, é recomendado que o usuário avalie a distribuição e os gráficos de dispersão presentes nas figuras *Frequency bands distribution*, *Frequency bands distribution over time* e *Frequency bands combined*. É possível clicar no botão *Open PDF files*, que abrirá os arquivos das figuras utilizando o leitor de arquivos PDF padrão do sistema. Posteriormente, ao decidir qual banda de frequência será adicionada, no painel *New frequency bands*, o usuário poderá selecionar o botão de opções referente à sua escolha. Por último, o botão *OK* fechará

a janela e salvará a seleção do usuário, adicionando essa informação à execução do algoritmo, que é retomada.

The image shows a software dialog box with a light gray background. At the top, a text box contains the message: "Our default mode for the classification is using hippocampal Theta/Delta ratio. Would you like to add other frequency bands for the classification? Which ones?". Below this is a section titled "New frequency bands" containing a list of radio button options: "None" (selected), "Add Beta", "Add Low Gamma", "Add High Gamma", "Add all Gamma range", and "Use 6 to 90 Hz/ Delta". At the bottom of the dialog, there is a text prompt: "Open PDF files to evaluate any possible additions of frequency bands according to the presented distributions:". Below this prompt are two buttons: "Open PDF files" on the left and "OK" on the right.

Figura Suplementar 5. Escolha das bandas de frequência para a classificação.

Caso a caixa de seleção *Run visual inspection* tenha sido selecionada, assim que a inspeção visual for necessária (como demonstrado pelo texto na caixa *Status*), uma nova janela será aberta (Figura Suplementar 6). Essa janela é a interface que permite que o usuário classifique um conjunto de épocas de 10 segundos aleatórias do registro, que serão utilizadas, posteriormente, para definir os limiares de probabilidade posterior para cada um dos estágios do ciclo. Em um dado instante todas as informações e gráficos presentes nessa janela são referentes a uma única época de 10 segundos do registro, porém, conforme a ação do usuário, essa época pode ser alterada e todas as informações e gráficos são atualizados. A janela possui um título com a informação do número da época apresentada (*Time bin*), um gráfico do sinal de LFP do hipocampo (Figura Suplementar 6-1.) e um gráfico do sinal do EMG (Figura Suplementar 6-2.) referente à época de 10 segundos. Há também um gráfico de PSD (Figura Suplementar 6-3.). Por último, há um gráfico de dispersão (Figura Suplementar 6-4.), indicando a distribuição de todas as épocas de 10 segundos que constituem o registro, além do destaque da época de referência em vermelho.



Stop and save inspection      Previous      Next

I am sure!      Transition

1 - AWAKE    2 - NREM    3 - REM      4 - AWAKE ↔ NREM    5 - NREM ↔ REM    6 - REM ↔ AWAKE    0 - None of those

7 - Epoch counter

Figura Suplementar 6. Inspeção Visual. A. LFP do hipocampo da época escolhida; B. Sinal do EMG da mesma época; C. PSD do LFP da época selecionada;

Quando uma época de 10 segundos é apresentada, é possível que o usuário a classifique, indicando a qual estágio do ciclo sono vigília ela pertence. Pressionando o botão:

- 1 - *AWAKE*, é classificada como estado acordado;
- 2 - *NREM*, como sono NREM;
- 3 - *REM*, como sono REM.

Caso julgue ser uma época de transição, o usuário pode pressionar:

- 4 - *AWAKE* ↔ *NREM*, indicando que ser uma transição entre o estado acordado e o sono NREM;
- 5 - *NREM* ↔ *REM*, entre o sono NREM e o sono REM;
- 6 - *REM* ↔ *AWAKE*, entre o sono REM e o estado acordado.

Caso julgue que a época em questão não pertence a nenhuma dessas classificações, o usuário pode pressionar o botão:

0 - *None of those*.

Após pressionar qualquer um desses botões, a próxima época é carregada e todos os gráficos são atualizados. Ademais, o usuário pode retroceder à época anterior pressionando o botão *Previous* ou avançar para a próxima, apertando o botão *Next*. Há também o botão:

7 - *Epoch counter*, que exibe uma mensagem que informa o número de épocas já apresentadas.

É possível que o usuário pressione os números associados aos botões, ou as setas direcionais para a esquerda e direita em seu teclado, sem haver a necessidade de que os botões sejam pressionados. Por fim, o botão *Stop and save inspection* permite que a inspeção visual seja interrompida e armazenada para que o usuário possa continuá-la posteriormente. Caso esse botão seja pressionado, todas as janelas são fechadas, exceto a interface principal.

Quando o usuário completar 0,5% do total de épocas classificadas, um aviso indicando o término da inspeção é apresentado, os dados referentes à inspeção visual são salvos, a interface da inspeção visual é fechada e a execução do algoritmo de classificação é retomada.

A interface de entrada de informações para o algoritmo de classificação (Figura 30) é mantida após o término da inspeção visual, visto que ela ainda mantém a função de indicar os processos que estão sendo executados através da caixa de texto *Status*. Contudo, nessa etapa, não é necessário que o usuário interaja com a interface, visto que o algoritmo está sendo executado em segundo plano. Por fim, quando a classificação for completamente finalizada, a caixa de texto *Status* apresentará a mensagem *Done*, e após 5 segundos a janela será fechada, mantendo apenas a interface principal aberta.

## Apêndice 2 - Instruções para utilização do classificador de estágios do ciclo sono-vigília



Renan Augusto Viana Mendes ([renanvadico@gmail.com](mailto:renanvadico@gmail.com))

Ikaro J. S. Beraldo ([ikaroberaldo@ufmg.br](mailto:ikaroberaldo@ufmg.br))



## Requisitos e informações básicas

A interface gráfica foi implementada utilizando a versão 2020b do MATLAB e testes foram feitos utilizando a versão 2020a do Windows 10. Portanto, é possível que a execução dos códigos apresente erros em versões mais antigas ou em sistemas operacionais diferentes.

Pacotes oficiais do MATLAB necessários:

- MATLAB 9.9
- Signal Processing *Toolbox* 8.5
- Statistics and Machine Learning *Toolbox* 12.0
- Curve Fitting *Toolbox* 3.5.12

As figuras presentes neste documento são referentes a um conjunto de dados que é utilizado como treinamento para o Modelo de Mistura de Gaussianas (GMM). Utilize-as, portanto, como referência para uma boa classificação de qualquer outro conjunto de dados que seja utilizado.

## Primeiros passos

Baixe os *scripts*, funções, classes e arquivos de variáveis necessários para a execução da interface e algoritmo de classificação.

- Link:

[https://drive.google.com/drive/folders/1eds\\_7Ba7fY4W7ojL8htVF9VaPQWrWotH?usp=sharing](https://drive.google.com/drive/folders/1eds_7Ba7fY4W7ojL8htVF9VaPQWrWotH?usp=sharing)

Inclua todos os *scripts*, funções, classes e arquivos de variáveis no conjunto de pastas de procura (Search Path) do MATLAB.

- Instruções:

[https://www.mathworks.com/help/matlab/matlab\\_env/add-remove-or-reorder-folderson-the-search-path.html](https://www.mathworks.com/help/matlab/matlab_env/add-remove-or-reorder-folderson-the-search-path.html)

Para iniciar a interface gráfica, abra e execute a classe *RMS\_pwelch\_integrate*. Posteriormente, não será necessária a execução de nenhum outro arquivo. Após a

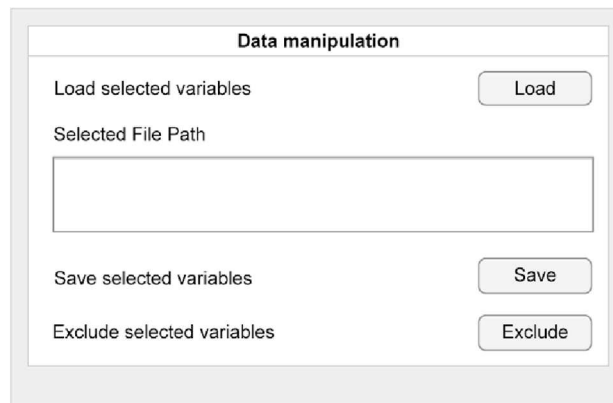
inicialização da janela principal, espere alguns instantes até que todos os elementos da mesma sejam carregados.

## Interface Principal

The screenshot displays the main interface of the software, divided into two main panels. The left panel is titled 'Data manipulation' and contains a 'Load selected variables' button with a 'Load' button next to it, a 'Selected File Path' text input field, a 'Save selected variables' button with a 'Save' button next to it, and an 'Exclude selected variables' button with an 'Exclude' button next to it. Below these buttons is a table with the following columns: Name, Class, Values, S.F, and State. The table is currently empty. The right panel is titled 'Pre-processing step' and has four sub-sections: 'Detrend', 'Resampling', 'Filtering', and 'Period separation'. The 'Filtering' sub-section is active and shows 'Polynomial degree:' with three radio button options: 'Constant (default)' (selected), 'Linear', and 'Quadratic'. There is also an 'Other' radio button with a text input field containing '0'. A checkbox labeled 'ON' is checked. Below this is a 'Run the selected steps using the selected variables' button with a 'Run' button next to it. The 'Sleep-wake cycle sorting' sub-section has two radio button options: '1 - Use workspace data' (selected) and '2 - Load pre-processed variables'. It also has two dropdown menus for 'CA1 Channel' and 'EMG Channel', a checkbox for 'Include the algorithm pre-processing step', and a 'Run' button. At the bottom of the right panel is a 'Status' section with a text input field containing the word 'Inactive'.

A primeira janela aberta é a interface primária, que consiste em painéis separados para cada uma das funções principais do software. A caixa de texto *Status* (canto inferior direito) contém informações sobre o andamento dos processos inicializados a partir da interface principal. Os termos *Busy* e *Inactive* serão apresentados na caixa caso algum processo esteja ou não sendo executado, respectivamente.

## Manipulação de dados

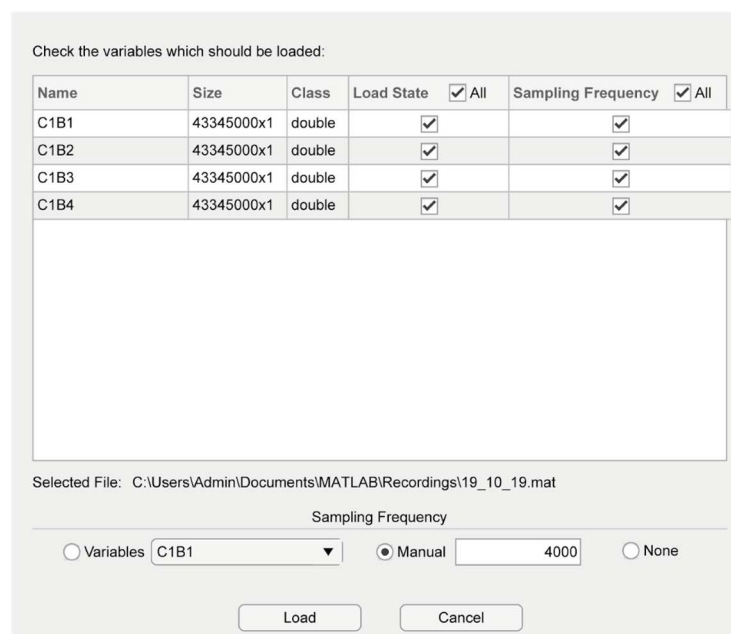


The image shows a dialog box titled "Data manipulation". It contains three main sections: "Load selected variables" with a "Load" button, "Selected File Path" with a text input field, and "Save selected variables" with a "Save" button. Below the "Save" button is "Exclude selected variables" with an "Exclude" button.

O painel *Data manipulation* agrupa funções que permitem que os dados sejam carregados, salvos ou excluídos.

### Carregar variáveis

Para carregar arquivos de extensão MAT, clique no botão *Load*. Uma janela de seleção de arquivos será aberta, onde o arquivo poderá ser selecionado.



The image shows a dialog box titled "Check the variables which should be loaded:". It contains a table with columns: Name, Size, Class, Load State,  All, Sampling Frequency, and  All. Below the table is a text input field for "Selected File:" with the value "C:\Users\Admin\Documents\MATLAB\Recordings\19\_10\_19.mat". Below that is a "Sampling Frequency" section with radio buttons for "Variables", "Manual", and "None". The "Variables" radio button is selected, and a dropdown menu shows "C1B1". The "Manual" radio button is also selected, and a text input field shows "4000". At the bottom are "Load" and "Cancel" buttons.

Name	Size	Class	Load State	<input checked="" type="checkbox"/> All	Sampling Frequency	<input checked="" type="checkbox"/> All
C1B1	43345000x1	double	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
C1B2	43345000x1	double	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
C1B3	43345000x1	double	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
C1B4	43345000x1	double	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Uma nova janela será aberta permitindo a seleção das variáveis que serão carregadas. Essa tabela apresenta informações básicas sobre essas variáveis salvas no arquivo selecionado: nome (*Name*), dimensões (*Size*) e classe (*Class*).

Ações:

1. Para definir quais variáveis serão carregadas, marque/desmarque as caixas de seleção presentes na coluna *Load State* e que sejam correspondentes às variáveis em questão.
2. Para definir quais variáveis possuem uma taxa de amostragem associada, marque/desmarque as caixas de seleção presentes na coluna *Sampling Frequency*.
3. Para definir a como a taxa de amostragem será associada às variáveis, selecione:
  - a. *Variables*: para selecionar na lista ao lado a variável que contém o valor de taxa de amostragem;
  - b. *Manual*: para inserir o valor da taxa de amostragem manualmente não campo de texto ao lado;
  - c. *None*: para não associar nenhum valor de taxa de amostragem às variáveis;

Para finalizar o carregamento das variáveis, clique no botão *Load*. Contudo, para cancelar o carregamento e retornar à interface principal, clique no botão *Cancel*.

Importante: os dados essenciais para a classificação são um registro de LFP de CA1 e um registro de eletromiografia. Ambos devem estar no formato MAT e na forma de vetores de mesmo comprimento e taxa de amostragem.

### Área de trabalho

Name	Class	Values	S.F	State
C1B1	double	43345000x1	4000	<input checked="" type="checkbox"/>
C1B2	double	43345000x1	4000	<input checked="" type="checkbox"/>
C1B3	double	43345000x1	4000	<input type="checkbox"/>
C1B4	double	43345000x1	4000	<input type="checkbox"/>

Abaixo do painel *Data Manipulation*, há o painel *Workspace*, com a finalidade de simular a *Área de trabalho* do MATLAB, sendo um espaço para exibir, em colunas, informações sobre as variáveis carregadas, como nome (*Name*), classe (*Class*), valores (*Values*), taxa de amostragem associada (*S.F.*) e o estado de seleção (*State*). A última coluna define quais variáveis serão consideradas (caixa de seleção marcada) ou não (caixa de seleção desmarcada) nos processos de exclusão, salvar ou de pré-processamento.

### Salvar variáveis

No painel *Workspace*, marque as caixas de seleção da coluna *State* que sejam referentes às variáveis que serão salvas. No painel *Data manipulation*, clique no botão *Save*. Uma janela de seleção de diretório será aberta, permitindo a seleção do local onde as variáveis serão salvas.

### Excluir variáveis

No painel *Workspace*, marque as caixas de seleção da coluna *State* que sejam referentes às variáveis que serão excluídas. No painel *Data manipulation*, clique no botão *Exclude*.

### Pré-processamento

The image shows a 'Pre-processing step' dialog box with the following elements:

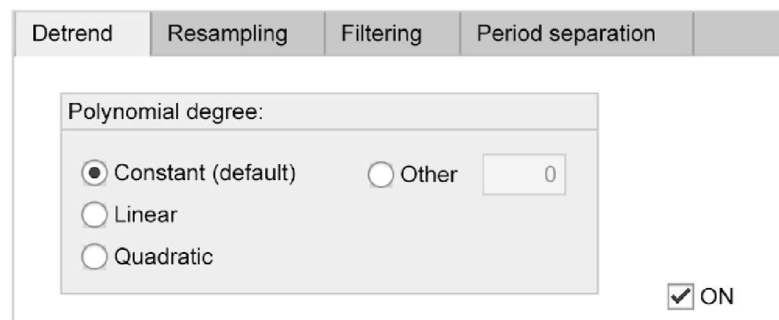
- Tabbed interface with 'Filtering' selected.
- 'Polynomial degree' section with radio buttons for 'Constant (default)', 'Linear', and 'Quadratic', and an 'Other' option with a text input field containing '0'.
- A checked checkbox labeled 'ON'.
- Text label: 'Run the selected steps using the selected variables'.
- 'Run' button.

O painel *Pre-processing step* consiste em 4 abas, sendo que cada uma delas agrupa um conjunto de informações relacionadas a uma das possíveis funções constituintes da etapa de pré-processamento.

Importante: esse passo é dispensável, uma vez que o algoritmo de classificação pode realizar o pré-processamento padrão (recomendado), caso essa opção seja selecionada (Ver: **Seleção de variáveis**)

Em cada uma das abas há uma caixa de seleção *ON/OFF* . Marque ou desmarque a caixa de seleção para ativar ou desativar a execução do passo de pré-processamento referente à aba selecionada. Posteriormente, após modificar as configurações de cada uma das etapas, clique no botão *Run* . A barra de *Status* apresentará o andamento do pré-processamento e indicará o fim com a mensagem “Done!”.

## Detrend



A aba *Detrend* permite ao usuário controlar a função de remoção de tendências polinomiais do sinal.

Selecione no painel *Polynomial degree* o botão referente ao grau polinomial utilizado pela função *detrend* :

- *Constant* (padrão): o grau polinomial é igual a 0;
- *Linear*: o grau polinomial é igual a 1;
- *Quadratic*: o grau polinomial é igual a 2
- *Other*: permite que o usuário insira o grau polinomial na caixa de texto ao lado (números naturais apenas);

## Reamostragem

The screenshot shows a control panel with four tabs: Detrend, Resampling, Filtering, and Period separation. The Resampling tab is active. It contains the following controls:

- Output sampling frequency (Hz):
- Input/Output frequency ratio:
- Change input sampling frequency (Hz):
- ON

A aba *Resampling* permite ao usuário controlar a função de modificar a taxa de amostragem do sinal. Por padrão, a taxa de amostragem inicial é aquela associada à variável (presente na coluna *S.F.* do *Workspace*).

- Na caixa de texto *Output sampling frequency* adicione a taxa de amostragem final que o dado terá após ser processado
- ou
- Adicione a razão entre a taxa de amostragem inicial e final na caixa de texto *Input/Output frequency ratio*.
  - Marque a caixa de seleção *Change input sampling frequency* e insira um outro valor para a taxa de amostragem inicial que não seja aquele já associado à variável (opcional).

## Filtragem

The screenshot shows a control panel with four tabs: Detrend, Resampling, Filtering, and Period separation. The Filtering tab is active. It contains the following controls:

- Filter type:
  - High-pass (Hz)
  - Band-pass
  - Low-pass (Hz)
  - Notch (Hz)
- Sampling frequency:
- ON

A aba *Filtering* permite que o usuário controle a função de filtragem do sinal. Por padrão, a taxa de amostragem inicial é aquela associada à variável (presente na coluna *S.F.* do *Workspace*).

No painel *Filter type*, selecione o botão de opção referente ao tipo de filtragem que será executada:

- *High-pass*: filtro passa-alta. Indique a frequência de corte na caixa de texto ao lado
- *Low-pass*: filtro passa-baixa. Indique a frequência de corte na caixa de texto ao lado.
- *Band-pass*: filtro passa-banda. Indique as frequências de corte inferior e superior nas caixas de texto *High-pass* e *Low-pass*, respectivamente;
- *Notch*: filtro rejeita-faixa. Indique a frequência que será atenuada na caixa de texto ao lado.
- Marque a caixa de seleção *Change input sampling frequency* e insira um outro valor para a taxa de amostragem inicial que não seja aquele já associado à variável (opcional).

## Separação em épocas

Detrend	Resampling	Filtering	Period separation
Period Length (sec)	<input type="text" value="10"/>		<input type="checkbox"/> Sampling frequency
Period Length (samples)	<input type="text" value="0"/>		<input type="text" value="0"/>
Number of Periods	<input type="text" value="0"/>		
Number of discarded samples:	<input type="text" value="0"/>		<input checked="" type="checkbox"/> ON

A aba *Period separation* controla a separação do dado em épocas de duração definida. Por padrão, a taxa de amostragem inicial é aquela associada à variável (presente na coluna *S.F.* do *Workspace*).

- Insira a duração das épocas em segundos na caixa de texto *Period Length*, visto que é a única que pode ser editada.
- Marque a caixa de seleção *Change input sampling frequency* e insira um outro valor para a taxa de amostragem inicial que não seja aquele já associado à variável (opcional).



## Classificação dos estágios do ciclo sono-vigília

### Seleção de variáveis

The screenshot shows a MATLAB-style dialog box titled "Sleep-wake cycle sorting". It features two radio buttons for data source selection: "1 - Use workspace data" (selected) and "2 - Load pre-processed variables". Under the first option, there are two dropdown menus for "CA1 Channel" (set to "C1B1") and "EMG Channel" (set to "C1B2"). An unchecked checkbox labeled "Include the algorithm pre-processing step" is also present. A "Run" button is positioned to the right of the second radio button. At the bottom, a "Status" label is above a text box displaying "Inactive".

O painel *Sleep-wake cycle sorting* é o inicializador do algoritmo de classificação dos estágios do ciclo sono-vigília. Primeiro, selecione a fonte dos dados utilizados como entrada para o algoritmo através das opções:

1 - *Use workspace data*: define que os dados serão provenientes da área de trabalho. Selecione os registros de LFP de CA1 e do eletromiograma através das caixas de listagem *CA1 Channel* e *EMG Channel*, respectivamente.

- Selecione a caixa de seleção *Include the algorithm pre-processing step* para incluir o pré-processamento padrão do algoritmo de classificação caso o pré-processamento das variáveis da área de trabalho ainda não tenha sido realizado (opcional).
- Clique no botão *Run* para que o algoritmo de classificação seja iniciado. Posteriormente, uma janela de seleção de diretório será aberta. Selecione o local onde os arquivos gerados pelo pré-processamento serão salvos.

Importante: o algoritmo de classificação sempre irá gerar o arquivo `data_variables.mat` e `ALL_DATA.mat` caso o usuário opte por utilizar dados provenientes da Área de Trabalho. Dessa forma, depois que os arquivos são salvos, é possível utilizar a segunda opção (*Load pre-processed variables*) para obter os dados de entrada para o algoritmo.

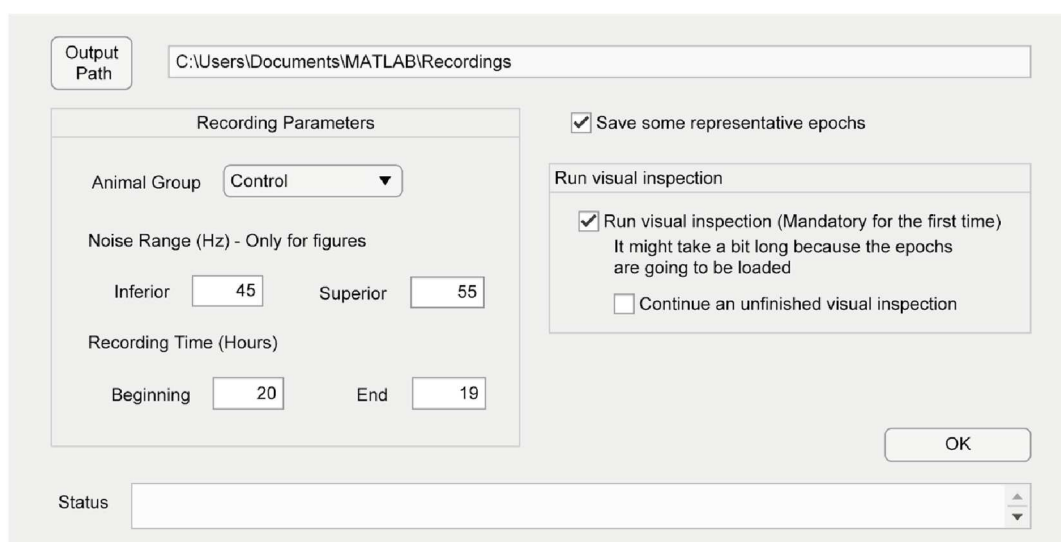
2 - *Load pre-processed variables*: define que o algoritmo de classificação utilizará um arquivo de nome *data\_variables.mat*, que contém os dados gerados pelo pré-processamento

realizado pelo algoritmo de classificação.

- Clique no botão *Run* para que o algoritmo de classificação seja iniciado. Posteriormente, uma janela de seleção de arquivos será aberta. Selecione o arquivo *data\_variables.mat* correspondente.

*Importante: essa opção só é válida caso o algoritmo de classificação já tenha gerado o arquivo para o conjunto de dados em questão (data\_variables e ALL\_DATA).*

### Informações sobre o registro e inputs iniciais



The screenshot shows a software window with the following elements:

- Output Path:** A text field containing the path `C:\Users\Documents\MATLAB\Recordings`.
- Recording Parameters:** A section containing:
  - Animal Group:** A dropdown menu set to "Control".
  - Noise Range (Hz) - Only for figures:** Two input fields: "Inferior" with value 45 and "Superior" with value 55.
  - Recording Time (Hours):** Two input fields: "Beginning" with value 20 and "End" with value 19.
- Save some representative epochs:** A checked checkbox.
- Run visual inspection:** A section containing:
  - Run visual inspection (Mandatory for the first time):** A checked checkbox with a note: "It might take a bit long because the epochs are going to be loaded".
  - Continue an unfinished visual inspection:** An unchecked checkbox.
- OK:** A button at the bottom right.
- Status:** A text field at the bottom left, currently empty.

Após a inicialização do algoritmo de classificação, uma nova janela será aberta. Ela é a principal forma de comunicação entre o algoritmo de classificação e o usuário. A barra de *Status* contém textos indicativos dos processos que estão sendo executados pelo algoritmo.

#### Ações:

- Para definir o diretório de saída para os dados e figuras gerados pelo algoritmo de classificação, clique no botão *Output Path*. Uma janela de seleção de diretório será aberta.

Importante: é recomendado que o diretório selecionado seja aquele no qual os arquivos *data\_variables* e *ALL\_DATA* tenham sido salvos.

- Através do painel *Recording Parameter*, defina as informações sobre o registro fornecido
- *Animal Group*: defina se o registro é referente a um animal controle ou experimental.
- *Noise Range*: defina o limite mínimo (*Inferior*) e máximo (*Superior*), em Hertz, para o possível ruído de rede presente no sinal.
- *Recording Time*: defina o horário de início (*Beginning*) e término (*End*), em horas, dos registros. Caso os horários não sejam exatos, faça uma aproximação e insira um número inteiro.
- *Save some representative epochs*: marque a caixa para salvar épocas representativas de cada um dos estágios do ciclo sono-vigília após a classificação ter sido finalizada.
  - *Run visual inspection*: defina se a etapa de inspeção visual será executada.

Importante: caso a inspeção visual ainda não tenha sido iniciada ou finalizada para este conjunto de dados, é obrigatório que esta opção esteja marcada.

- *Continue an unfinished visual inspection*: defina se uma inspeção visual não finalizada será retomada (Ver: 4.4.1. Interromper e salvar a inspeção; 4.4.2. Continuar uma inspeção visual não finalizada).

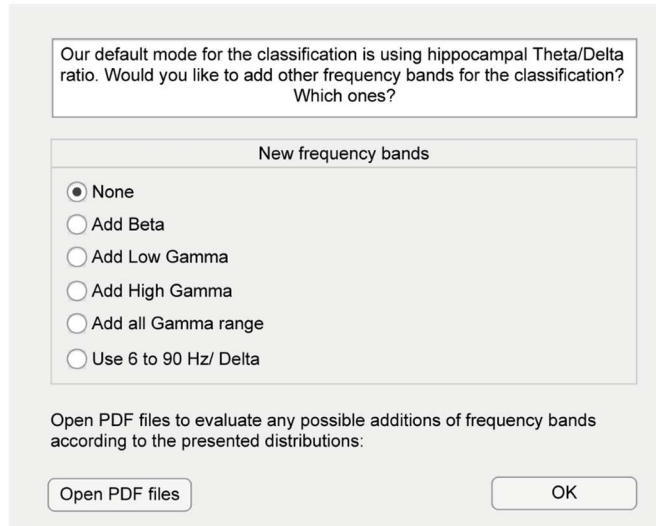
- Pressione o botão *OK* para que a execução do algoritmo seja retomada.

### **Figuras geradas:**

- *Frequency bands distribution*:  
Correlaciona os valores potência nas diferentes em todas as bandas de frequência que podem ser utilizadas na *Toolbox*;
- *Frequency bands distribution over time*:  
Apresenta a distribuição ao longo do tempo dos valores potência nas diferentes em todas as bandas de frequência que podem ser utilizadas na *Toolbox*;
- *Frequency bands combined*:

Demonstra gráficos de dispersão com as distribuições assumidas quando o usuário escolher cada combinação das bandas de frequências;

### Adição de novas bandas de frequência



The dialog box contains the following text and elements:

Our default mode for the classification is using hippocampal Theta/Delta ratio. Would you like to add other frequency bands for the classification?  
Which ones?

New frequency bands

- None
- Add Beta
- Add Low Gamma
- Add High Gamma
- Add all Gamma range
- Use 6 to 90 Hz/ Delta

Open PDF files to evaluate any possible additions of frequency bands according to the presented distributions:

Open PDF files      OK

Uma nova janela será aberta, permitindo que uma nova banda de frequência seja adicionada à razão Teta/Delta durante a classificação.

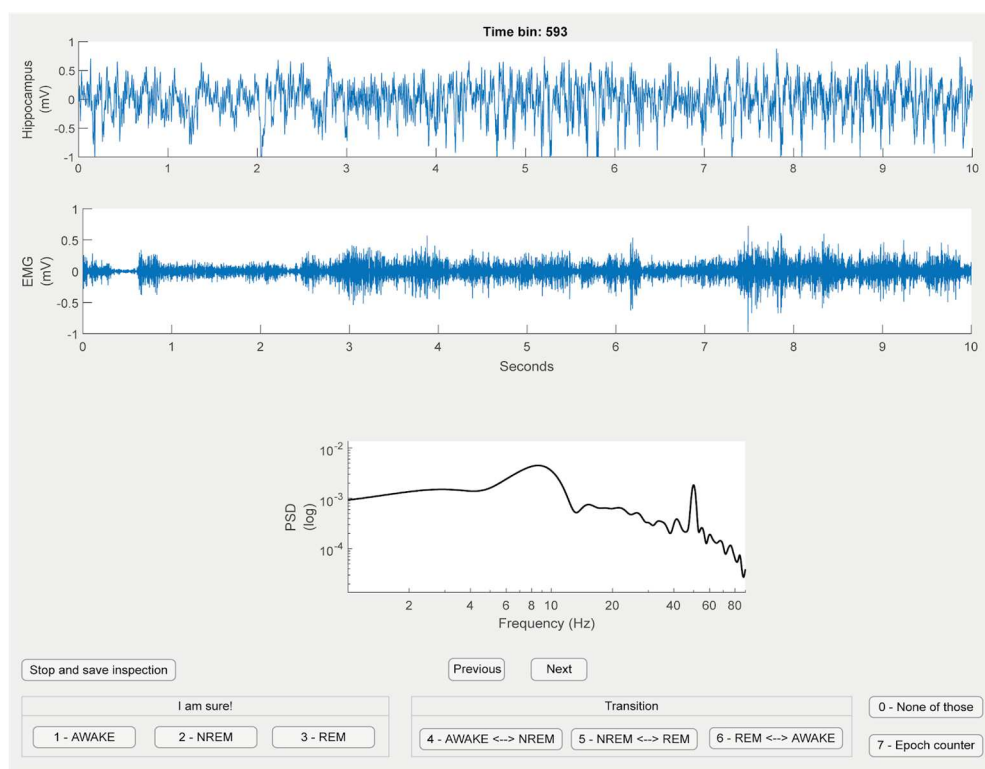
Importante: é recomendado que os gráficos de distribuição e dispersão presentes nas figuras 'Frequency bands distribution', 'Frequency bands distribution over time' e 'Frequency bands combined' sejam avaliados.

- Clique no botão *Open PDF files* para abrir os arquivos das figuras utilizando o leitor de arquivos PDF padrão do sistema;
- No painel *New frequency bands*, selecione o botão de opções referente à banda de interesse.
- Clique no botão *OK* para fechar a janela e retomar a execução do algoritmo.

### Figuras geradas:

- *Data distribution*
- *GMM clusters*

## Inspeção visual



Caso a caixa de seleção *Run visual inspection* tenha sido selecionada, assim que a inspeção visual for necessária, uma nova janela será aberta. Essa janela é a interface que permite que um conjunto aleatório de épocas do registro seja classificado manualmente.

Em um dado instante todas as informações e gráficos presentes nessa janela são referentes a uma única época de 10 segundos do registro, porém, conforme a ação do usuário, essa época pode ser alterada e todas as informações e gráficos são atualizados. A janela possui um título com a informação do número da época apresentada (*Time bin*), um gráfico do sinal de LFP (Figura 3.a) do hipocampo (*Hippocampus*) e um gráfico do sinal (Figura 3.b) do eletromiograma (*EMG*) referente à época de 10 segundos.

É necessário que um total de 40 épocas seja classificado para que esta etapa seja finalizada. Contudo, é possível interromper e salvar a inspeção visual para que esta seja retomada futuramente (Ver: **Interromper e salvar a inspeção; continuar uma inspeção visual não finalizada**).

Ações:

A cada época apresentada selecione:

- 1 - *AWAKE*: para classificar a época como parte estado acordado;
- 2 - *NREM*: para classificar a época como parte do sono NREM;
- 3 - *REM*: para classificar a época como parte do sono NREM;
- 4 - *AWAKE*  $\longleftrightarrow$  *NREM*: para classificar a época como uma transição entre o estado acordado e o sono NREM;
- 5 - *NREM*  $\longleftrightarrow$  *REM*: para classificar a época como uma transição entre o sono NREM o sono REM;
- 6 - *REM*  $\longleftrightarrow$  *AWAKE*: para classificar a época como uma transição entre o sono REM e o estado acordado;
- 7 - *Epoch counter*: para exibir uma mensagem que informa o número de épocas já apresentadas;
- 0 - *None of those*: caso julgue que a época em questão não pertence a nenhum destes estados;
- *Previous*: para retroceder para a época anterior;
- *Next*: para avançar para a próxima época;

Detalhe: É possível pressionar os números associados aos botões, ou as setas direcionais para a esquerda e direita em seu teclado, sem haver a necessidade de que os botões sejam pressionados.

### **Interromper e salvar a inspeção**

Pressione o botão *Stop and save inspection* para interromper e salvar a inspeção visual atual. Posteriormente, as janelas de inspeção visual e inputs iniciais serão fechadas, retornando à interface principal.

### **Continuar uma inspeção visual não finalizada**

Caso a opção *Continue an unfinished visual inspection* tenha sido marcada, durante a inicialização da inspeção visual, uma janela de seleção de arquivo será aberta. Selecione o arquivo *IDX\_Visual\_Inspection* correspondente ao conjunto de dados utilizados.

Posteriormente, a inspeção visual será retomada a partir da última época não classificada.

### **Após inspeção visual**

Quando 40 épocas forem classificadas para cada um dos 3 estágios do ciclo sono-vigília, um aviso indicando o término da inspeção será apresentado. Posteriormente, a execução do algoritmo de classificação é retomada.

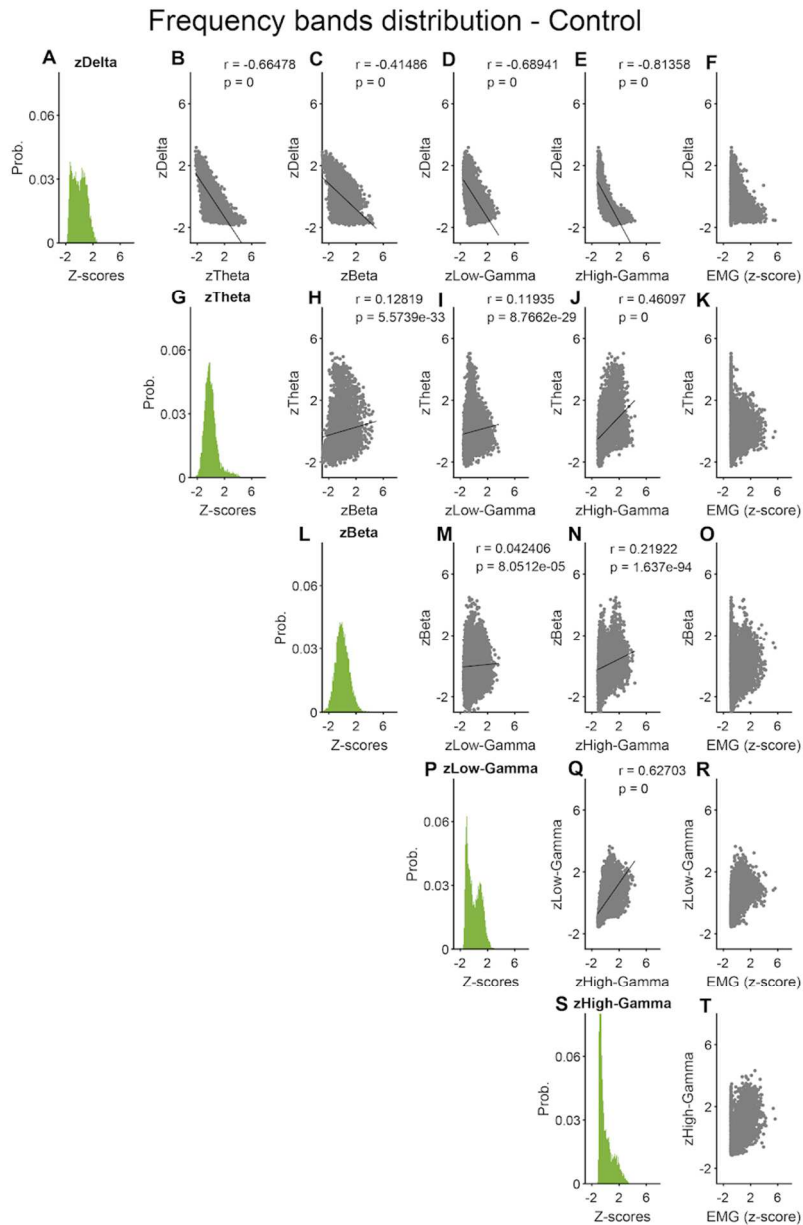
Figuras geradas:

- *Visual Inspected data distribution*
- *Clusters Formation*
- *Final Clusters*
- *Final Classification over time*
- *Final Acurácia*
- *State distribution for frequency bands*
- *Sorted 6 to 90 Hz*
- *Figuras representativas*

Nesta etapa, não será necessário interagir com a interface, visto que o algoritmo estará sendo executado em segundo plano. Após alguns minutos, a caixa de texto *Status* apresentará a mensagem *Done* , e após 5 segundos a janela será fechada, mantendo apenas a interface principal aberta.

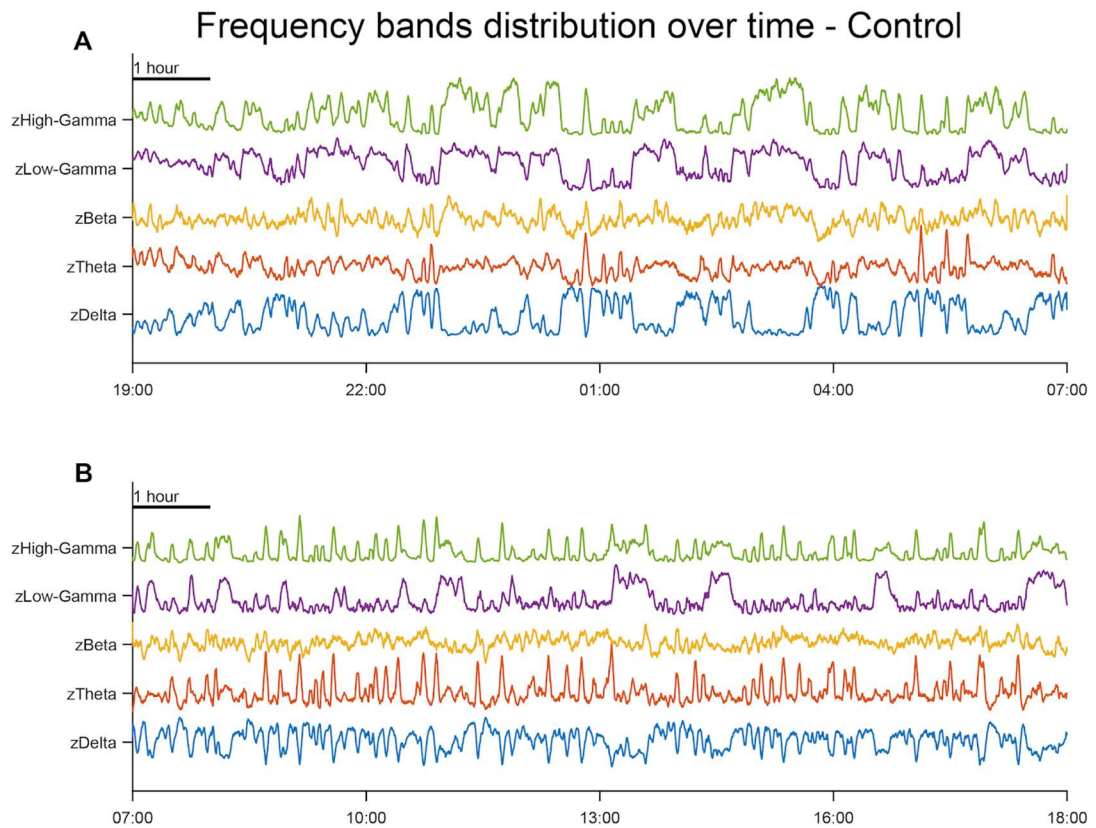
Verifique a pasta de saída informada, visto que as figuras e dados gerados pelo algoritmo de classificação estarão organizados lá.

Todas as figuras geradas pela *Toolbox*



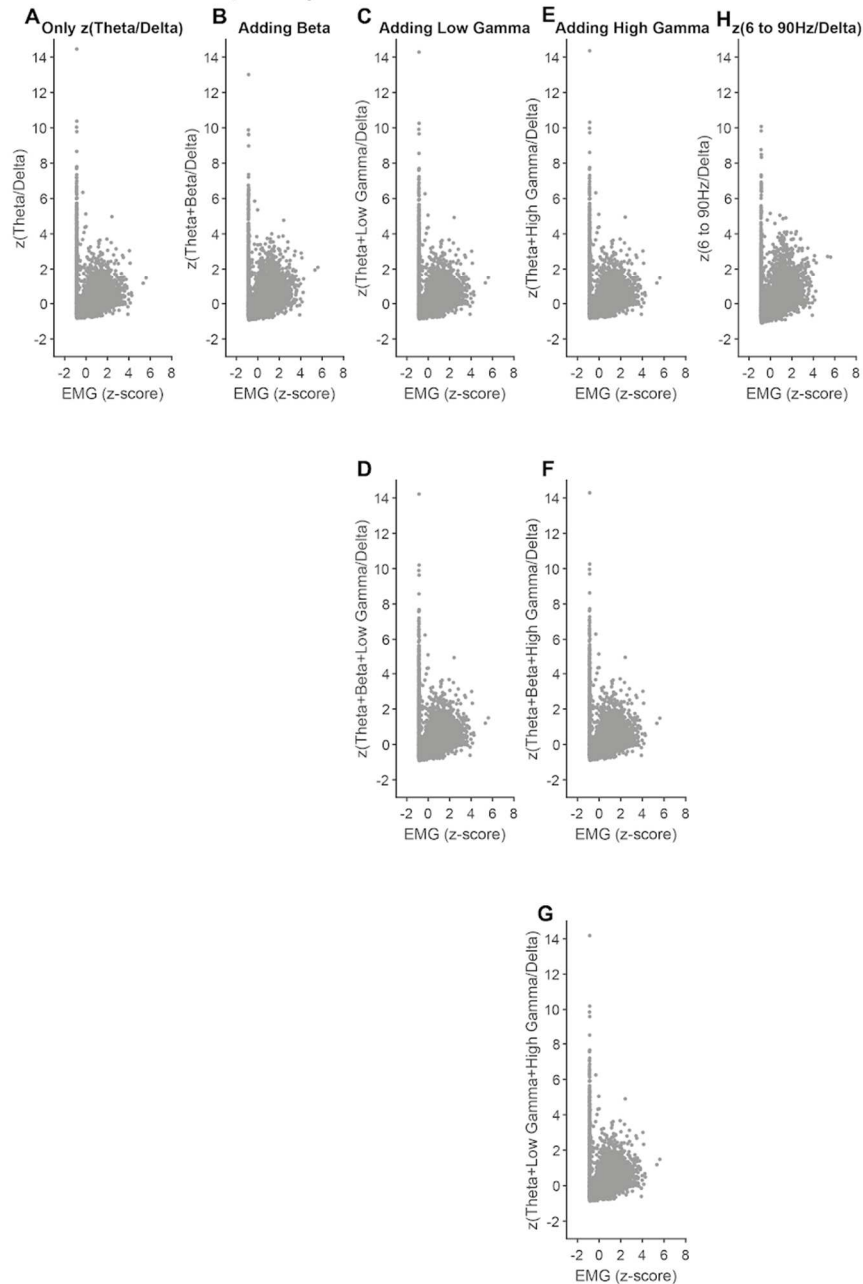
**Figura “Frequency bands distribution”:** A: Histograma da com os valores de potência em delta; B a E: Combinação de delta com as outras bandas de frequência; F: Delta vs EMG; G: Histograma da com os valores de potência em Teta; H a J: Combinação de Teta com as outras bandas de frequência; K: Teta vs EMG; L: Histograma da com os valores de potência em beta; M a N: Combinação de beta com as outras bandas de frequência; O: Beta vs EMG; P: Histograma da com os valores de potência em low gamma; Q: Combinação de low gamma com high gamma; F: Low gamma vs EMG; S: Histograma da com os valores de potência em high gamma; Q: High gamma vs EMG;





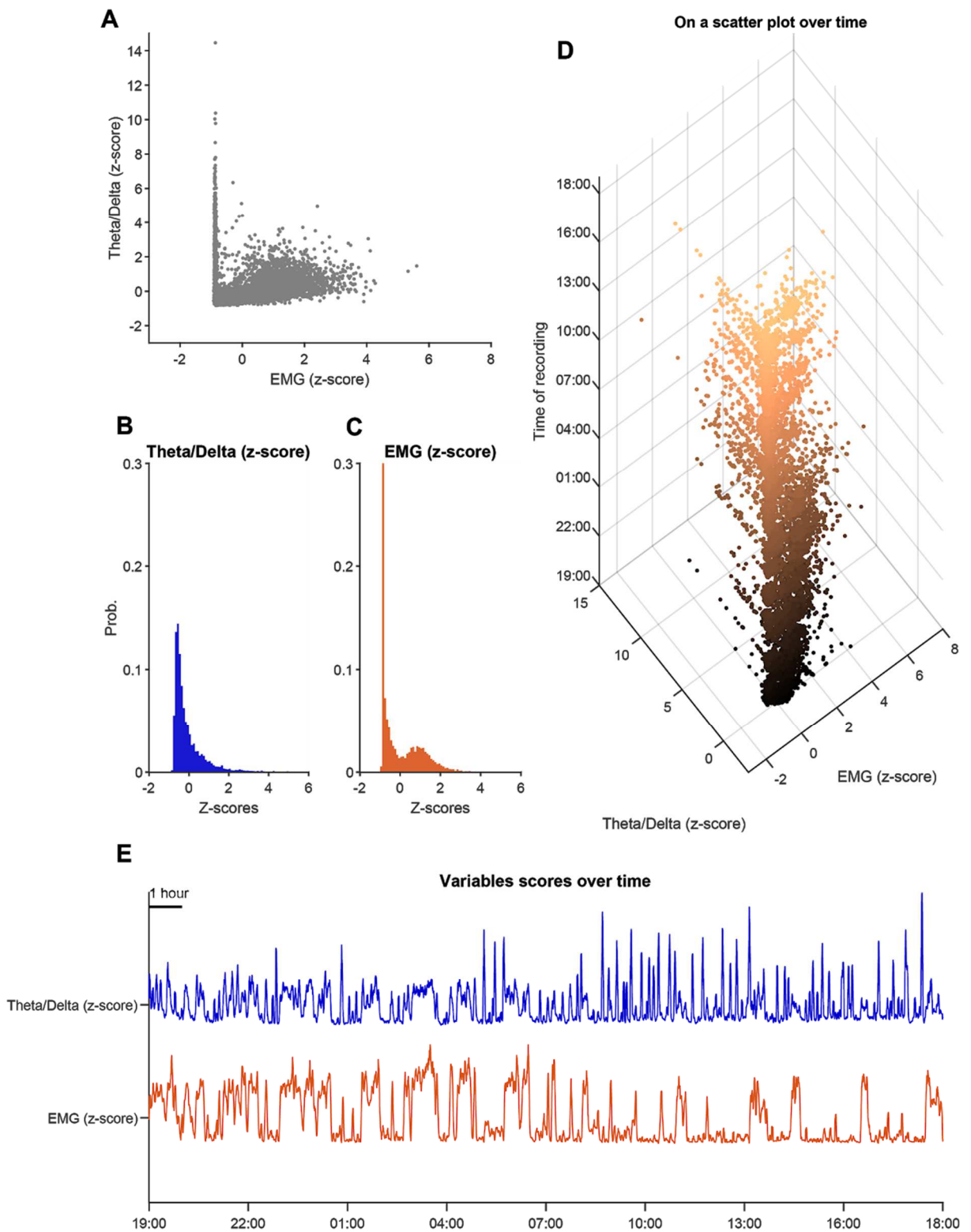
**Figura “Frequency bands distribution over time”:** Distribuição dos valores de potência em cada banda de frequência ao longo do tempo, na fase escura (A) e clara (B).

### Frequency bands combined - Control

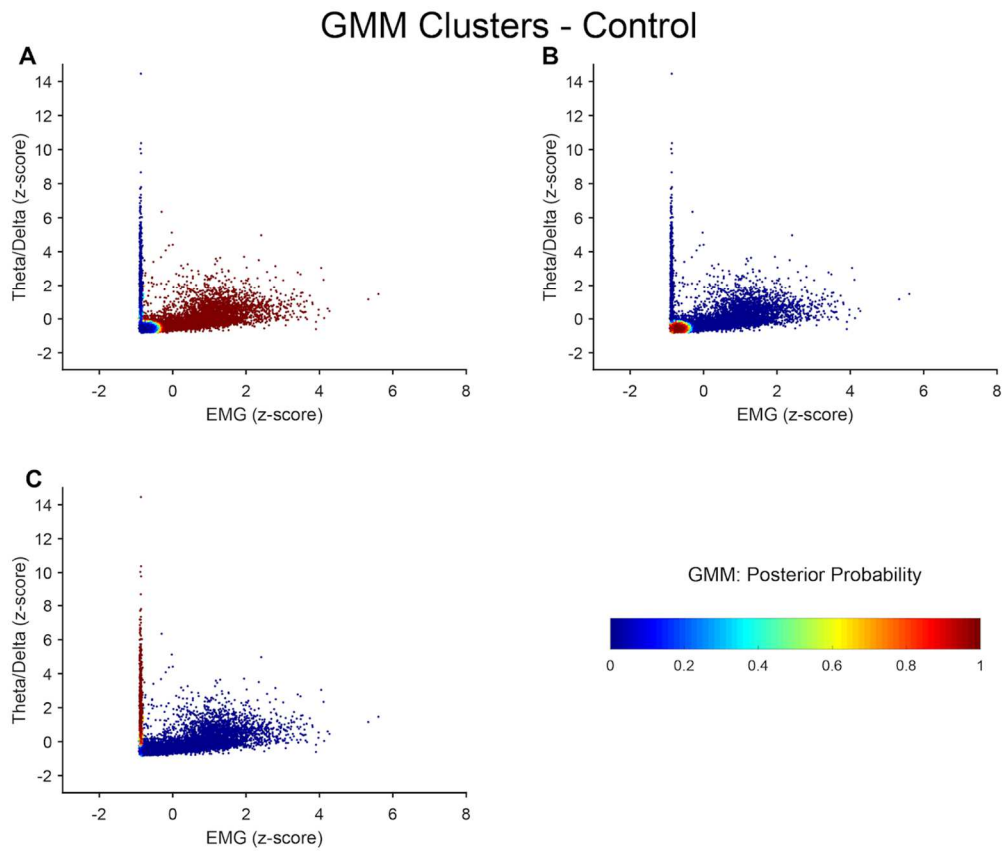


**Figura “Frequency bands combined”:** A: Distribuição dos valores da relação Teta/delta e RMS do EMG; B: Distribuição dos valores da relação Teta + beta/delta e RMS do EMG; C: Distribuição dos valores da relação Teta + low gamma/delta e RMS do EMG; D: Distribuição dos valores da relação Teta + beta + low gamma/delta e RMS do EMG; E: Distribuição dos valores da relação Teta + high gamma/delta e RMS do EMG; F: Distribuição dos valores da relação Teta + beta + high gamma/delta e RMS do EMG; G: Distribuição dos valores da relação Teta + beta + low gamma + high gamma/delta e RMS do EMG; H: Distribuição dos valores de potência de 6 a 90 Hz / a delta e RMS do EMG;

## Data distribution - Control

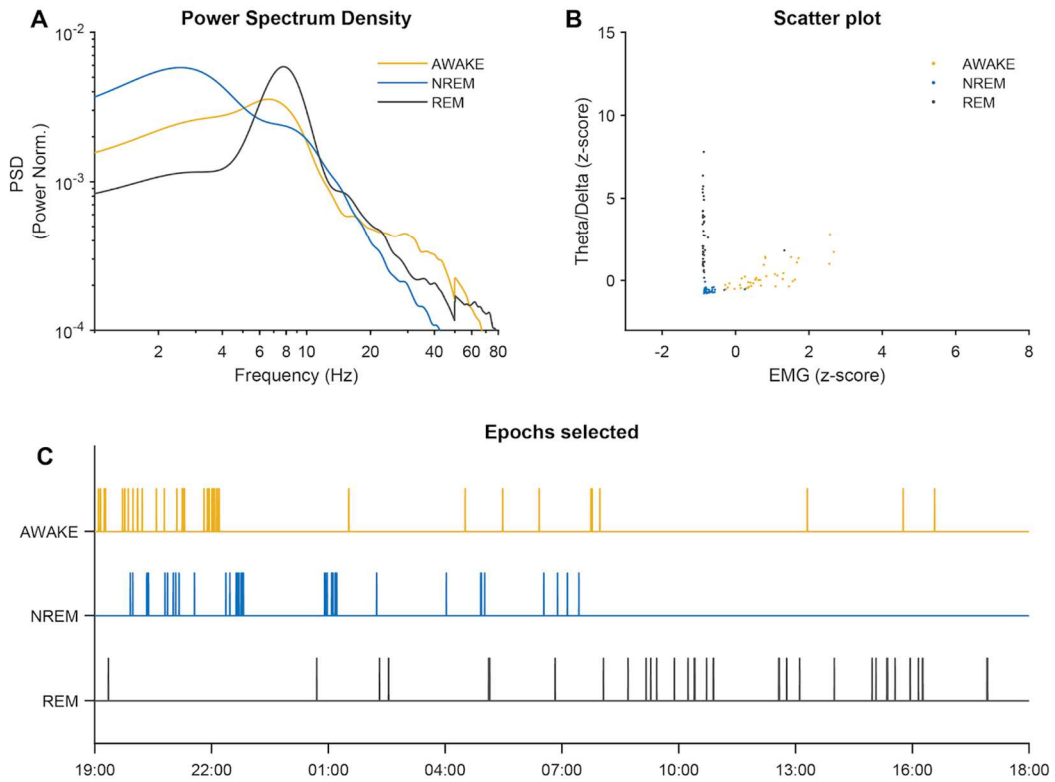


**Figura "Data Distribution":** A: Gráfico de dispersão com os valores de cada época da relação Teta/delta (que pode variar de acordo com a escolha do usuário) e do RMS do EMG; B: Histograma da distribuição dos valores da relação Teta/delta; C: Histograma da distribuição dos valores de RMS do EMG;



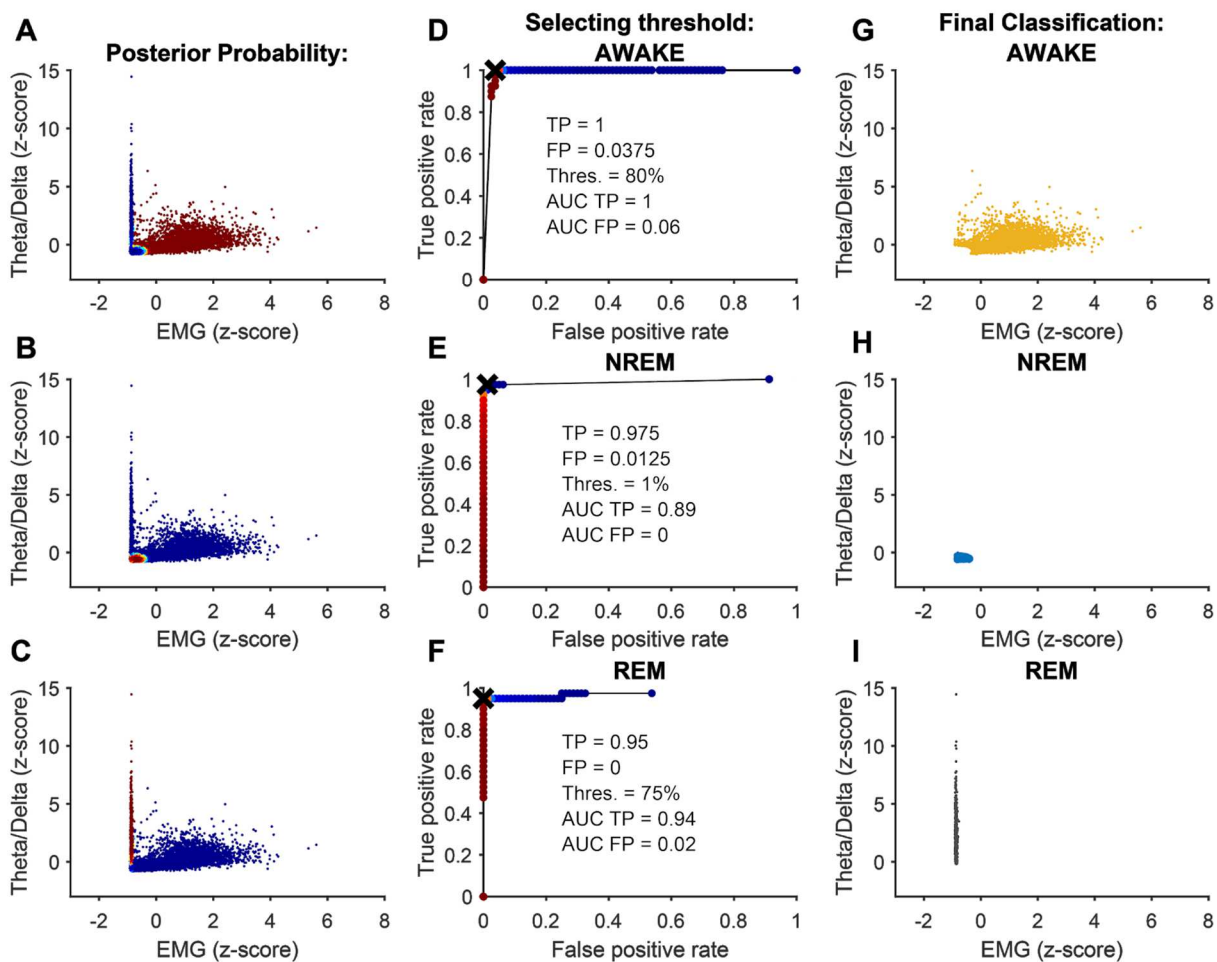
**Figura “GMM clusters”:** A: Gráfico de dispersão com o valor de probabilidade de cada época pertencer a cada grupo que será designado como acordado; B: Gráfico de dispersão com o valor de probabilidade de cada época pertencer a cada grupo que será designado como sono NREM; C: Gráfico de dispersão com o valor de probabilidade de cada época pertencer a cada grupo que será designado como sono REM.

## Visually Inspected Data distribution - Control



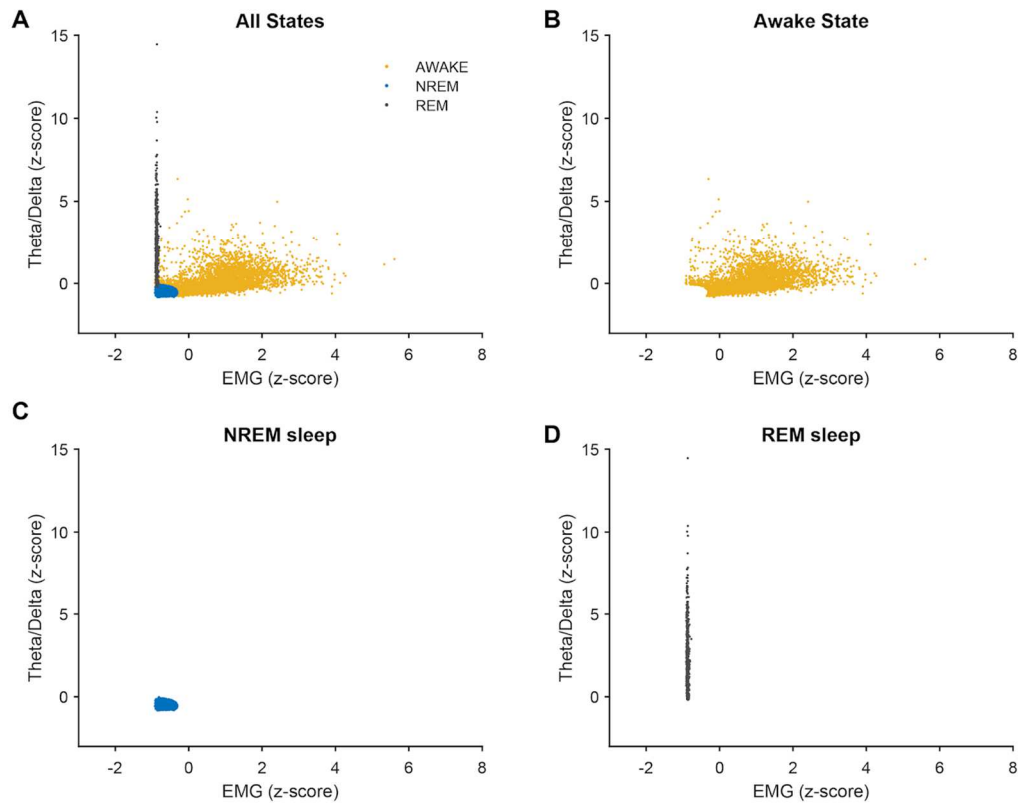
**Figura “Visually Inspected Data Distribution”:** A: PSD médio das épocas classificadas em cada estado; B: Gráfico de dispersão com os valores individuais de cada época classificada; C: Classificação ao longo do tempo.

## Clusters Formation - Control



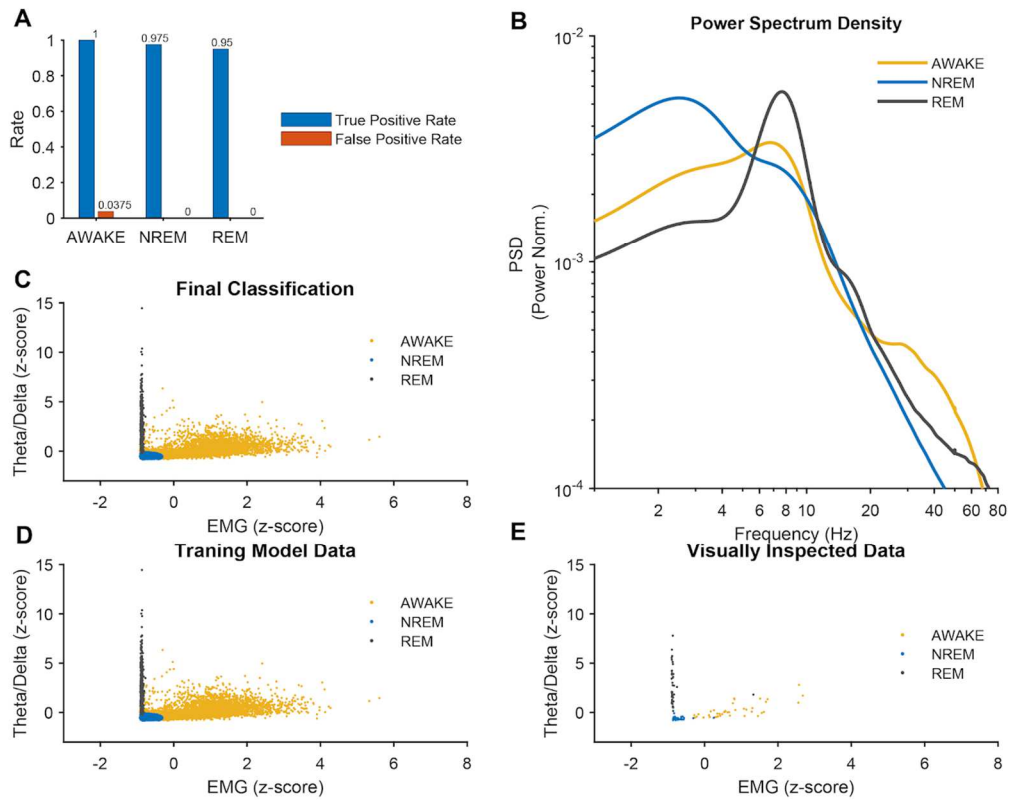
**Figura "Clusters Formation":** A ao C: Gráficos de dispersão com os grupos formados; D ao F: Curva ROC calculada para definir cada estado. Limiar escolhido é marcado no gráfico (como "X") e sua respectiva taxa de verdadeiros positivos (TP) e falsos positivos (FP) descritas ao meio, junto do valor da área sob a curva da taxa de TP e FP por cada limiar; G ao I: Classificação final.

## Final Clusters - Control



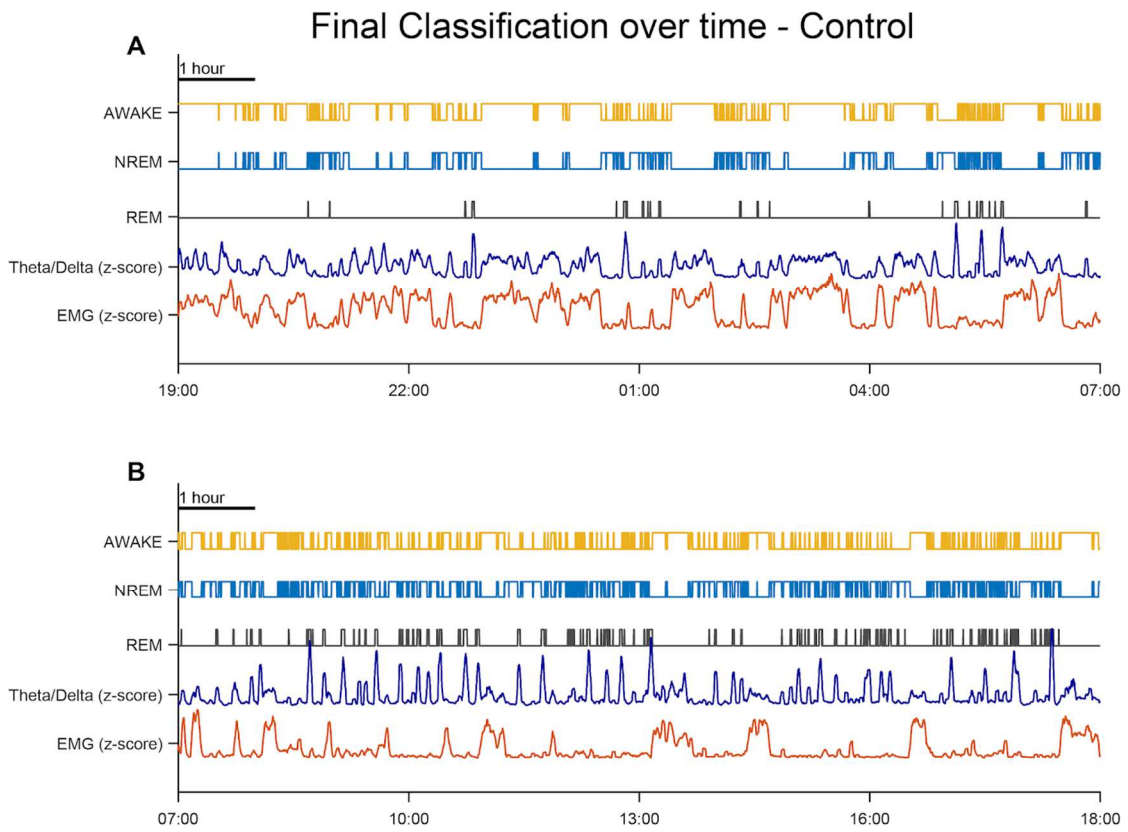
**Figura "Final Cluster":** A: Gráfico de dispersão com todas as épocas classificadas; B: Somente épocas classificadas como acordado; C: Somente épocas classificadas como sono NREM; D: Somente épocas classificadas como sono REM.

## Final Performance - Control



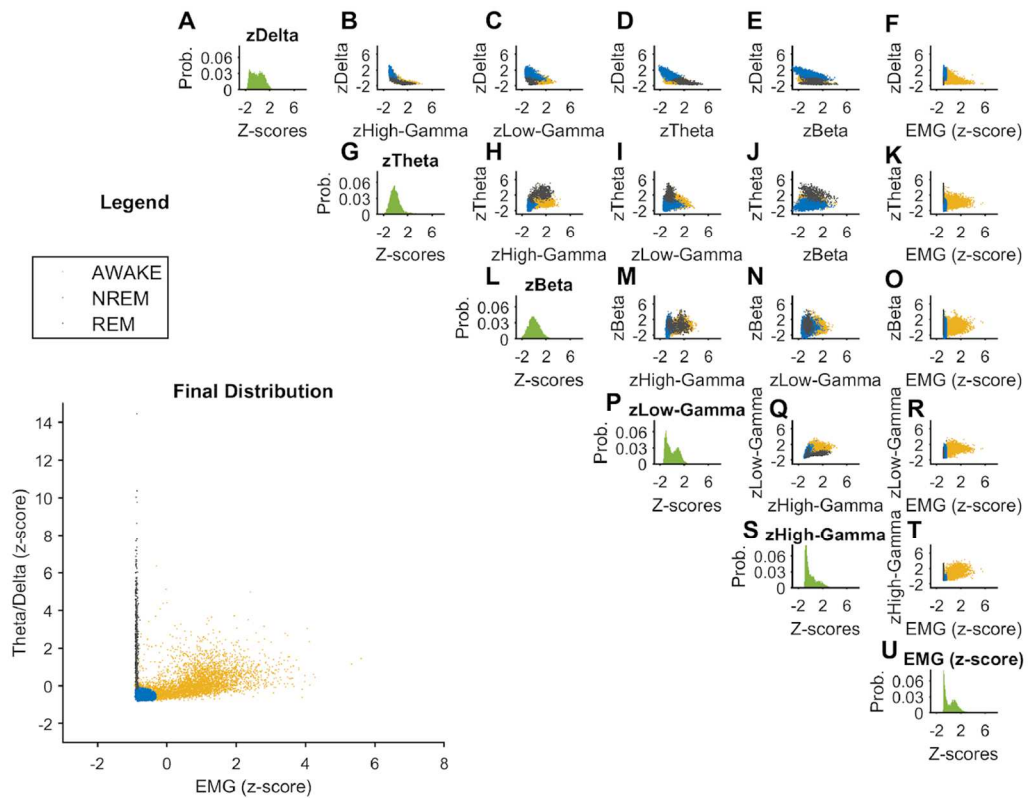
**Figura “Final Acurácia”:** A: Taxa de verdadeiros positivos (TP) e falsos positivos (FP) calculadas; B: PSD média das épocas classificadas em cada estado; C: Gráfico de dispersão com todas as épocas classificadas; D: Gráfico de dispersão com todas as épocas utilizadas como treinamento; E: Gráfico de dispersão com todas as épocas classificadas manualmente.



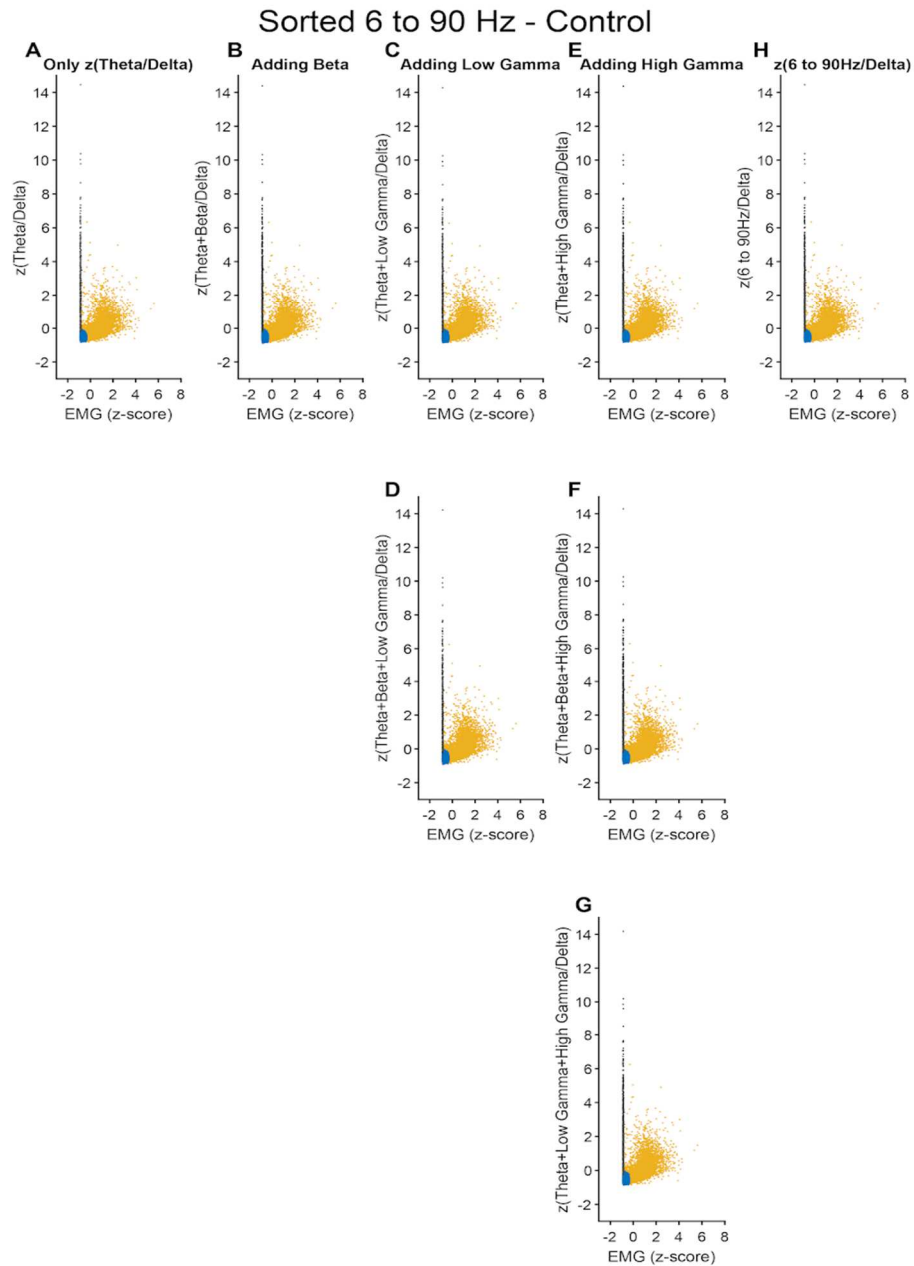


**Figura “Final Acurácia”:** Classificação ao longo do tempo com distribuição dos valores da relação Teta/delta e do RMS do EMG para a fase escura (A) e clara (B).

## State distribution for frequency bands - Control



**Figura “State distribution for frequency bands”:** A: Histograma da com os valores de potência em delta; B a E: Combinação de delta com as outras bandas de frequência com épocas classificadas; F: Delta vs EMG; G: Histograma da com os valores de potência em Teta; H a J: Combinação de Teta com as outras bandas de frequência com épocas classificadas; K: Teta vs EMG; L: Histograma da com os valores de potência em beta; M a N: Combinação de beta com as outras bandas de frequência com épocas classificadas; O: Beta vs EMG; P: Histograma da com os valores de potência em low gamma; Q: Combinação de low gamma com high gamma com épocas classificadas; F: Low gamma vs EMG; S: Histograma da com os valores de potência em high gamma; T: High gamma vs EMG; U: Histograma do EMG.



**Figura “Sorted 6 to 90 Hz”:** A: Distribuição dos valores da relação Teta/delta e RMS do EMG em cada época e sua respectiva classificação (em todos os gráficos da figura); B: Distribuição dos valores da relação Teta + beta/delta e RMS do EMG; C: Distribuição dos valores da relação Teta + low gamma/delta e RMS do EMG; D: Distribuição dos valores da relação Teta + beta + low gamma/delta e RMS do EMG; E: Distribuição dos valores da relação Teta + high gamma/delta e RMS do EMG; F: Distribuição dos valores da relação Teta + beta + high gamma/delta e RMS do EMG; G: Distribuição dos valores da relação Teta + beta + low gamma + high gamma/delta e RMS do EMG; H: Distribuição dos valores de potência de 6 a 90 Hz / a delta e RMS do EMG;