MULTI-TASK GAUSSIAN PROCESS REGRESSION FOR THE DETECTION OF SLEEP CYCLES IN PREMATURE INFANTS

Nele Sophie Brügge[1], *Jan Grasshoff*[2], *Arne Weigenand*[3], *Philipp Rostalski*[2,4]

[1]Institute of Medical Informatics, University of Lübeck
[2]Fraunhofer Research Institution for Individualized and Cell-Based Medical Engineering, Lübeck, Germany
[3]Drägerwerk AG & Co. KGaA, Lübeck, Germany
[4]Institute for Electrical Engineering in Medicine, University of Lübeck

ABSTRACT

Preterm birth is the leading cause of infant mortality. Consequently, preterm infants require special attention and medical care, with sleep being a central element for the development of cognitive functions. Studies on neonatal sleep suggest that the pattern of their sleep stages is determined by an endogenous ultradian rhythm, superimposed by other rhythms and external influences. In this article, we propose the use of multi-task Gaussian process regression as a flexible non-parametric approach to analyze this kind of sleep data while incorporating prior knowledge, such as of correlations between signals, signal periodicity, information from manual annotations and certain other signal properties. As a result of the regression of heart and respiratory rate data of preterm infants, ultradian rhythms with a period of $58{\pm}5\,\mathrm{min}$ could be extracted. Together with other model parameters, knowledge about the characteristics of ultradian rhythms potentially provides insights into the maturational and health status of the preterm infants. These, in turn, could be used to optimize the care of critically ill patients.

Index Terms— Gaussian process, Bayesian modeling, regression, preterm infant, sleep cycling

1. INTRODUCTION

With an increasing trend, about 15 million babies are born prematurely each year, putting the global incidence of preterm birth at about $10\,\%$. Extremely preterm infants with gestational ages as low as 23 to 24 weeks continue to contribute disproportionately to infant mortality, even in developed countries [17].

Improvement in medical care of premature infants is thus of high importance, with sleeping behavior being a central factor. Whereas sleeping patterns of preterm infants may appear chaotic, particularly in term neonates a distinct periodicity of sleep stages exists [12]. Also in preterm infants, the cyclicity of sleep stages has been the subject of prior research [3], [15] and has already been demonstrated in [3] documenting a cycle duration of $60\,\mathrm{min}$. Scher et al. have further shown that sleep cycles of $68\,\mathrm{min}$ exist in premature infants with gestational ages between 25 and 30 weeks [19]. Awareness of ultradian sleep patterns in preterm infants would not only allow adaptation of nursing interventions and treatments to the infant's sleep pattern [14], [16], but could also be of clinical relevance to assess the maturity of the preterm infant's brain [6] as well as to detect specific diseases at an early stage [13].

Using Gaussian process regression, this work investigates the ultradian sleep rhythm of preterm infants under the assumption that it is superimposed by other rhythms and external influences.

According to the results of [12], the cyclicity of sleep stages is particularly pronounced in the absence of strong external influences on the infant's sleep, while the routine of the neonatal intensive care unit (NICU) is not synchronized with the sleep pattern of the neonates. Therefore, the first superimposing rhythm is that of scheduled nursing interventions which usually cause the infants to awaken. This potentially leads to an impediment of the development of a stable sleep rhythm and interference with early sensory development processes [12]. Correspondingly, a masking effect of the NICU was described in [5] and [10] and dominant rhythms induced by feeding and nursing interventions could be detected.

A second rhythm considered as superimposing the ultradian rhythm is the $24\,\mathrm{h}$ circadian rhythm which could be identified in infant skin and rectal temperature data in [5] and [10] albeit with low amplitude in some cases.

We propose a multi-task Gaussian process model for preterm infant vital data regression and simultaneous estimation of those rhythms, focusing in particular on the ultradian rhythm. These models offer the advantages of allowing an integration of prior knowledge about sub-processes, take into account correlations between different data series and can be applied to non-equidistantly sampled data without prior interpolation. The mentioned circadian, nursing and ultradian rhythms are approximated using different periodic kernels. While fitting the Gaussian process to the data, the individual ultradian period of each preterm infant is learned automatically. In addition, information about unplanned nursing or parental interventions is integrated using a manually created, categorical signal.

Due to noise and frequent interventions with high impact on the data leading to many artefacts, the underlying data is of high complexity and can thus not easily be analyzed visually. However, Gaussian processes were shown to be able to detect periodicity even in the present case and additionally, to accurately regress it.

2. GAUSSIAN PROCESS MODEL

A Gaussian process represents a generalization of the Gaussian probability distribution and can further be understood as a probability distribution over functions. In the following, observations are given by the data points $\mathbf{x}_i$ with corresponding values y_i. Section 2.1 and 2.2 are intended to provide a brief introduction to Gaussian processes. For a more detailed explanation, see [18].

Funding information: The data acquisition was funded in part by Drägerwerk AG & Co. KGaA.
Acknowledgments: We thank Prof. Dr. Singer and the Section Neonatology and Pediatric Intensive Care Medicine of the University Medical Center Hamburg-Eppendorf for performing the data acquisition.

2.1. Single-task Gaussian process

A Gaussian process is fully characterized by its mean and covariance functions $m(\mathbf{x})$ and $k(\mathbf{x}, \mathbf{x}')$ and hereafter referred to as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{1}$$

where $m(\mathbf{x})$ can be assumed to be zero without restriction of generality and $k(\mathbf{x}, \mathbf{x}')$ is a kernel function.

Problem-specific covariance matrices can be selected which are typically paramatrized by a set of hyperparameters allowing a more precise fit to the data points. In the following, ℓ denotes the length scale and σ^2 the variance hyperparameter. A simple but widely used covariance function known as squared exponential (SE)-covariance function is given by

$$k_{\text{SE}}(\mathbf{x}, \mathbf{x}') = \sigma_{\text{SE}}^2 \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2\ell_{\text{SE}}}\right). \tag{2}$$

To adequately model periodic signals, periodic covariance functions can be used. One example used in this work is

$$k_{\text{per}}(\mathbf{x}, \mathbf{x}') = \sigma_{\text{per}}^2 \exp\left(-\frac{1}{2}\left(\frac{\sin\left(\frac{\pi}{T}(\mathbf{x} - \mathbf{x}')\right)}{\ell_{\text{per}}}\right)^2\right) \tag{3}$$

introducing an additional hyperparameter, the period T.

By evaluating the covariances of all n training data points $X = \{\mathbf{x}_i \mid i = 1, ..., n\}$, the corresponding covariance matrix $K(X, X)$ can be obtained. As an example, the entry $K_{i,j}$ holds the value of $k(\mathbf{x}_i, \mathbf{x}_j)$ and the covariance matrix $K(X, X_*)$ of X and the the set of n_* test data points $X_* = \{\mathbf{x}_k^* \mid k = 1, ..., n_*\}$ is of dimension $n \times n_*$.

Typically, only noisy versions $y = f(\mathbf{x}) + \varepsilon$ of the function values are available, leading to the vector $\mathbf{y} = [y_1, ..., y_n]^\top$ of all training data points. To fit the process to the given training data, the prior distribution must be constrained to only contain those functions that run through the observed data points, resulting in the posterior distribution. For this purpose, the conditional distribution $\mathbf{f}_*|X_*, X, \mathbf{y}$ is formed where $\mathbf{f}_* = [f(\mathbf{x}_1^*), ..., f(\mathbf{x}_n^*)]^\top$ denotes the vector of function values of the data points in X_*.

Assuming an additive, independent and identically distributed Gaussian noise ε of variance σ_{noise}^2 and zero mean, the prior on the observations becomes $\text{cov}(\mathbf{y}) = K(X, X) + \sigma_{\text{noise}}^2 I$ and the posterior distribution is given by

$$\begin{aligned}\mathbf{f}_*|X_*, X, \mathbf{y} \sim \mathcal{N}(&K(X_*, X)\,\text{cov}(\mathbf{y})^{-1}\mathbf{y},\\ &K(X_*, X_*) - K(X_*, X)\text{cov}(\mathbf{y})^{-1}K(X, X_*))\end{aligned} \tag{4}$$

yielding the central prediction equation of the process [18].

2.2. Multi-task Gaussian process

In this work, several related data series, in this context also referred to as tasks, are given. Instead of training a single-task Gaussian process for each task, multi-task Gaussian processes [2] were applied. By their definition, not only the covariance between values of one modality, but also between values across different signals is of relevance. Input signals can have different sampling frequencies, differ in sampling intervals or, in extreme cases, even have different input spaces [1]. For task $d = 1, ..., D$, the training data set can thus be expressed as $S_d = (X_d, \mathbf{y}_d)$ where the number of data points N_d assigned to the respective task d can vary [1]. One way to also consider the correlation between tasks is to assume two independent covariance functions $k_{\text{t}}(\mathbf{x}, \mathbf{x}')$ and $k_{\text{c}}(d, d')$ and to combine them via multiplication, as also suggested in [7]. Using matrix notation and the Kronecker product $\otimes$, the resulting covariance matrix is

$$K_{\text{MT}}(X, X) = B \otimes K_{\text{t}}(X, X) \tag{5}$$

where B is the cross-task covariance matrix of dimension $D \times D$ and index t indicates that covariance within one task is addressed. Here we assume that different data points of one task represent different points in time, which is why the latter is subsequently referred to as temporal covariance. For greater flexibility, Q kernels, each with a separate covariance matrix B_q, can be added as in

$$K_{\text{LMC}}(X, X) = \sum_{q=1}^{Q} B_q \otimes K_{\text{t},q}(X, X). \tag{6}$$

In this case, the resulting model is often termed a linear model of coregionalization (LMC) [1].

The matrices B_q are indirectly given by $B_q = WW^\top$ through the lower triangular matrix W. Consequently, they each add $\frac{1}{2}D(D+1)$ hyperparameters θ_{c} to the hyperparameters θ_{t} of the covariance functions $K_{\text{t,q}}$. By the parametrization of W, it is ensured that $WW^\top$ always yields a positive semidefinite matrix. A scaling of B_q through row- and column-wise division by the root of the diagonal element of the respective column and row can be performed to obtain more interpretable correlation matrices. Hyperparameter tuning can be done automatically by minimizing the negative log marginal likelihood $p(\mathbf{y}|X, \boldsymbol{\theta})$ using gradient-based optimization techniques [18].

2.3. Modeling of the assumed signal components

The following section explains the modeling of the assumed signal components. In this context, the assumption of an additive superposition of rhythms and external factors is made.

2.3.1. Circadian rhythm

One modeled periodic signal component is the circadian rhythm. It should be noted that factors such as lighting, noise or frequency of nursing interventions as a function of daytime are expected to strongly influence the emergence of the circadian rhythm and that it thus cannot necessarily be attributed to endogenous processes.

For modeling, the simple periodic kernel from (3) is used where the period duration T is set to $24\,\text{h}$ and the length scale ℓ to $1\,\text{h}$, resulting in the kernel k_{CD}. The comparably short length scale allows the modeled circadian signal to deviate from a simple sinusoidal form which is based on the observation that in many subjects the circadian component shows a dip at night and a plateau during the day. Since the variance is both diurnally and subject dependent, σ_{CD}^2 is one hyperparameter to be learned.

2.3.2. Interventional rhythm

Another rhythm that affects the infant's sleep is the rhythm of nursing interventions in the NICU. The basic kernel

$$k_{\text{M}}(\mathbf{x}, \mathbf{x}') = \sigma_{\text{M}}^2 \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|}{1.5\,\text{h}}\right), \tag{7}$$

used to model this subsignal is a Matérn kernel with the associated parameter $\nu = 1/2$ and length scale $\ell_{\text{M}} = 1.5\,\text{h}$, transformed into a periodic kernel $k_{\text{M,per}}$ by harmonic analysis [8]. This kernel allows a better approximation of non-smooth functions, which is particularly relevant for modeling the impact of nursing interventions, as, for instance, heart rate can increase rapidly in response to painful procedures.

Since the amplitude of the subsignal can vary and the function does not repeat itself exactly over time, the kernel was multiplied by an SE kernel defined in (2) to produce a quasiperiodic kernel. The length scale of the SE kernel was set to $5\,\mathrm{h}$ and determines in which time intervals strict periodicity of the Matérn kernel is to be sustained. Overall, the final quasiperiodic kernel for modeling the interventional rhythm can be described as

$$k_{\text{int}}(\mathbf{x},\mathbf{x}') = k_{\text{M,per}}(\mathbf{x},\mathbf{x}') \exp\left(-\frac{|\mathbf{x}-\mathbf{x}'|^2}{10\,\mathrm{h}}\right). \quad (8)$$

The hyperparameters of k_{int} to be learned are the variance and the period length, which is initialized with $2\,\mathrm{h}$ or $3\,\mathrm{h}$ based on nursing protocols.

2.3.3. Ultradian rhythm

For modeling the internal ultradian rhythm of the preterm infant, again the simple periodic kernel, given by (3), was used. Similar to k_{int}, periodicity of the kernel is constrained to $4\,\mathrm{h}$ signal intervals through multiplication by an SE kernel with $\ell = 4\,\mathrm{h}$, resulting in the quasiperiodic kernel k_{UD}.

Since T_{UD} and σ^2_{UD} are highly dependent on the respective subject, they are defined to be learned during training. The search space of T_{UD} was limited to a range of values from $0.4\,\mathrm{h}$ to $2\,\mathrm{h}$ using the projected gradient method as described in [20]. Based on the results from [19], an initial period duration of one hour was assumed.

2.3.4. Non-periodic signal components

To include interventions occurring outside the specified $2\,\mathrm{h}$ or $3\,\mathrm{h}$ rhythm, covered by k_{int}, into modeling, the manually created annotations of nursing and parental contact were incorporated. However, these also contain scheduled interventions that occur periodically. In order to prevent the modeling of the NICU rhythm from being distorted, all time points of planned interventions were removed from the signal by referring to nursing protocols. This non-periodic categorical annotation signal is approximated using another SE-kernel k_{ext} with $\ell = 0.5\,\mathrm{h}$.

Furthermore, additive Gaussian noise was added to the process by a noise kernel. The variances of the noise as well as of the kernel k_{ext} were defined as hyperparameters.

2.3.5. Combination into a Gaussian process kernel

Based on the assumption of additive subprocesses, the final Gaussian process can be expressed by the sum of the individual processes. This allows to generate interpretable models through the fact that the posterior can be decomposed into the individual kernels' contributions. To capture the correlation between the signals, the kernel-based covariance matrices $K_{\text{CD}}, K_{\text{int}}, K_{\text{UD}}$ and K_{ext} based on the respective kernels $k_{\text{CD}}, k_{\text{int}}, k_{\text{UD}}$ and k_{ext} are each multiplied by covariance matrix B_q with $q = 1, ..., 4$. As in (6), the covariance matrix of the entire Gaussian process can be written

$$\begin{aligned} K_{\text{total}}(X,X) &= B_{\text{CD}} \otimes K_{\text{CD}}(X,X) + B_{\text{int}} \otimes K_{\text{int}}(X,X) \\ &\quad + B_{\text{UD}} \otimes K_{\text{UD}}(X,X) \\ &\quad + B_{\text{ext}} \otimes K_{\text{ext}}(X,X) + \sigma^2_{\text{noise}} I. \end{aligned} \quad (9)$$

To avoid modeling the non-periodic annotations with periodic kernels and thereby falsifying the detected rhythms, the column and row of the matrices $B_{\text{CD}}, B_{\text{int}}$ and B_{UD} associated with the annotations are set to zero. The matrix B_{ext} is dense allowing the annotations to influence the modeling of the vital data. Table 1 gives an overview of the different subkernel's parameters.

Table 1: Overview of multi-task Gaussian process subkernel specifications. Column "qp" lists the length scales of the multiplied SE kernels, which define the interval of strict periodicity of kernel k_q. The entry "opt" means that the parameter is variable, not constrained and optimized throughout training. Parameters are given in hours.

k_q	kernel type	ℓ_q	σ^2_q	T_q	qp	dense B_q
k_{CD}	periodic (3)	1.0	opt	24.0	✗	✗
k_{int}	per. Matérn	1.5	opt	$[1.8, 3.1]$	5.0	✗
k_{UD}	periodic (3)	0.6	opt	$[0.4, 1.8]$	4.0	✗
k_{ext}	SE (2)	0.5	opt	–	✗	✓

3. EXPERIMENT AND RESULTS

3.1. Data

Vital data of preterm neonates were collected from Dec. 15, 2015 to Jan. 19, 2018 at the Paediatric Intensive Care Unit of the University Medical Center Hamburg-Eppendorf in cooperation with Drägerwerk AG & Co. KGaA. The experimental protocol was approved by the ethics committee of the Hamburg Chamber of Physicians and all parents gave their written consent prior to participation. Heart rate and respiratory rate of 9 neonates were used for analysis. Gestational ages ranged from 24 to 31 weeks and birth weights from $498\,\mathrm{g}$ to $1246\,\mathrm{g}$.

In order to enable modeling of external influences on the infant's sleep, the time points at which there is direct nursing related or parental contact to the child were annotated by analyzing infrared video data. Depending on the intensity of the contact, a value from 0 to 2 was assigned to each time point. This results in an additional categorical signal, providing information about external influences due to human interaction.

3.2. Preprocessing and experimental setup

Prior to training, sampling rates of the input data were set to $1\,\mathrm{Hz}$. Furthermore, unrealistic values caused by measurement errors were removed from the data. These include, for example, heart rates of less than $100\,\mathrm{bpm}$, which can be caused by electrode detachment. Subsequently, standardization of the data was performed to obtain mean-free data with unit standard deviation. Input signals are then low-pass filtered and downsampled by a factor of 100 to reduce computation time. Note that this prevents high-frequency rhythms from being captured. The data was then split into $24\,\mathrm{h}$ periods overlapping by $6\,\mathrm{h}$, some of which contained gaps. During training, the optimization of the hyperparameters is done according to the Broyden-Fletcher-Goldfarb-Shanno algorithm [4], [9], [11], [21].

3.3. Regression and ultradian period durations

For validation of the regression, the correlation and root mean square error (RMSE) of the signals predicted by the Gaussian process with

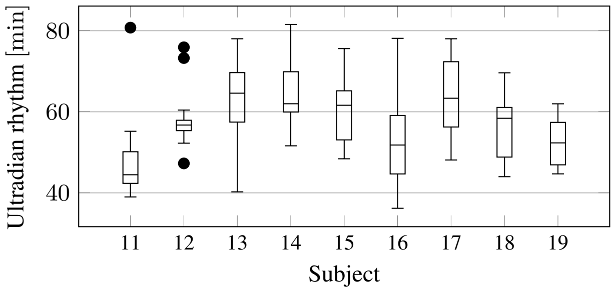

Fig. 1: Boxplots for the ultradian period durations of each subject. Depending on the length of the subject's recording, 17 to 41 data segments were considered, respectively.

the standardized original signals were calculated. If available, annotations of external interventions were used in addition to vital data. With an average RMSE of 0.17 ± 0.02 and 0.19 ± 0.03 and a correlation of 0.97 ± 0.01 and 0.96 ± 0.01 for standardized heart and respiratory rates, respectively, it is possible to regress the data using the proposed parameters with a high degree of accuracy.

The resulting ultradian periods and variances vary across subjects as expected. The mean value of the ultradian period duration is $58\,\mathrm{min}$ with a mean standard deviation of $9\,\mathrm{min}$, irrespective of initial values used in the optimization procedure. The minimum and maximum period durations are $49\,\mathrm{min}$ and $65\,\mathrm{min}$, respectively.

To illustrate the distribution of period durations over the course of one week of data collection, a boxplot diagram was created for each subject. This is shown in Fig. 1. For subjects 13, 16, and 17, the variance of the identified periods is increased due to a highly variable period duration over the time of data acquisition. In these cases, the period duration of the vital data varies to a large extent over time.

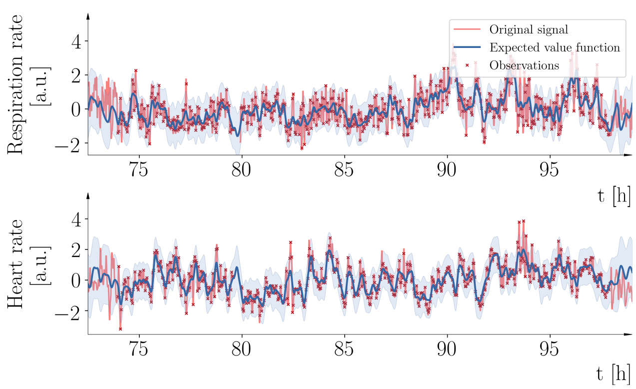

Fig. 2: Gaussian process regression on an example $24\,\mathrm{h}$ signal excerpt of data from subject 18. The confidence interval is highlighted in light blue. Respiration and heart rate are standardized and thus given in arbitrary units.

Regression results of each $24\,\mathrm{h}$ segment were displayed graphically, as in Fig. 2, and each signal segment was additionally analyzed with respect to the plausibility of the detected ultradian rhythms. However, due to the superposition of the signal components, they are hard to distinguish visually. For this reason, the signal was decomposed into the predictions made with the different kernels presented in section 2.3. The result is depicted in Fig. 3. It is demonstrated that each kernel contributes to the overall result and reasonable rhythms could be found in most cases.

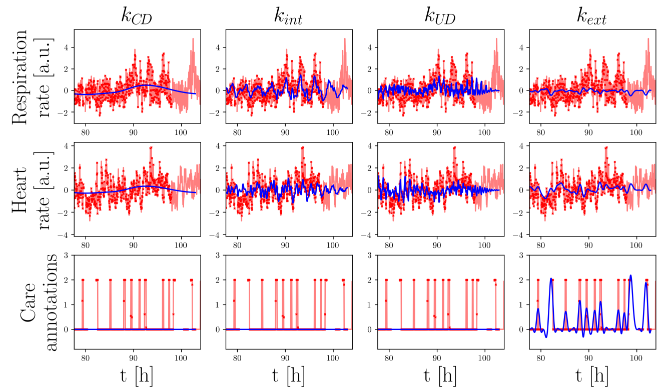

Fig. 3: Individual signal contributions obtained by decomposing the kernel into its components represented by k_{CD}, k_{int}, k_{UD} and k_{ext} for an exemplary time interval of $24\,\mathrm{h}$. The original signal is shown in red and the predicted expected value function in blue. Observations are indicated by red markers. Due to standardization, respiration and cardiac rates are given in arbitrary units.

Further, it can be seen that non-zero function values of respiratory and cardiac rates after the last observations are derived from nursing annotations which are available another $5\,\mathrm{h}$ in the future, suggesting learned correlations between the data series. Otherwise, in the absence of any correlations, the expectation value function would drop to zero in the course of a length scale $\ell_{\mathrm{ext}} = 0.5\,\mathrm{h}$ for $t > 98\,\mathrm{h}$. This is confirmed by the correlation matrices created from B_{ext} as described in section 2.3.5 taking values of $0.63, 0.79, 0.74,$ and 0.81 for correlation between interventions and heart rate, being strongly positive for all four subjects. The correlation between annotations and respiratory rate is mainly negative with values of $-0.46, -0.50, 0.04,$ and -0.55. In practice, these findings imply that during external intervention the heart rate of the considered preterm infants increases, while the respiratory rate decreases.

4. CONCLUSION

In this paper, a multi-task Gaussian process model was proposed to analyze sleep rhythms of preterm infants. Concurrently, it was investigated to what extent Gaussian processes can also be applied to complex data series with several influencing factors and with the integration of annotation data series. Based on assumptions about endogenous rhythms and external influences, a regression of the vital data was performed to this end. Prior knowledge was incorporated through the choice of suitable kernels and their hyperparameters.

The regression of the data was possible with high accuracy. In most of the data sets, a low-amplitude circadian rhythm and in all data sets, a prominent rhythm of nursing interventions could be detected. Considering the absence of ground truth, the ultradian rhythms found were compared with the results of previous research and are consistent with the findings in [19]. The mean period duration was $58\,\mathrm{min} \pm 9\,\mathrm{min}$. Although it can not be clearly clarified to what extent these are affected by caregiver or parental interventions, the results therefore support the hypothesis that even premature infants develop ultradian sleep rhythms. The fitted parameters of the Gaussian process, especially entries of the correlation matrices as well as period lengths and variances of detected rhythms, may provide valuable information about the maturational or health status of the preterm infant. A useful extension of the model would be to include body temperature or electroencephalography data.

5. REFERENCES

References

[1] M. Álvarez, L. Rosasco, and N. Lawrence, "Kernels for vector-valued functions: A review," *Foundations and Trends in Machine Learning*, vol. 4, Jun. 30, 2011.

[2] E. V. Bonilla, K. M. Chai, and C. Williams, "Multi-task gaussian process prediction," in *Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems*, Vancouver, British Columbia, Canada, Jan. 1, 2007, p. 8.

[3] I. Borghese, K. L. Minard, and E. Thoman, "Sleep rhythmicity in premature infants: Implications for development status.," *Sleep*, vol. 18 7, pp. 523–30, 1995.

[4] C. G. Broyden, "The convergence of a class of double-rank minimization algorithms," *IMA Journal of Applied Mathematics*, vol. 6, no. 1, pp. 76–90, 1970.

[5] C. Bueno and L. Menna-Barreto, "Environmental factors influencing biological rhythms in newborns: From neonatal intensive care units to home," *Sleep Science (Sao Paulo, Brazil)*, vol. 9, no. 4, pp. 295–300, Dec. 2016.

[6] S. Cabon, F. Porée, A. Simon, B. Met-Montot, P. Pladys, O. Rosec, N. Nardi, and G. Carrault, "Audio- and video-based estimation of the sleep stages of newborns in neonatal intensive care unit," *Biomedical Signal Processing and Control*, vol. 52, pp. 362–370, Jul. 1, 2019.

[7] R. Dürichen, M. Pimentel, L. Clifton, A. Schweikard, and D. Clifton, "Multi-task gaussian process models for biomedical applications," in *Proceedings of the International Conference on Biomedical and Health Informatics (IEEE BHI)*, Valencia, Spanien, Jun. 1, 2014.

[8] N. Durrande, J. Hensman, M. Rattray, and N. Lawrence, "Detecting periodicities with gaussian processes," *PeerJ Computer Science*, vol. 2, e50, Apr. 13, 2016.

[9] R. Fletcher, "A new approach to variable metric algorithms," *The Computer Journal*, vol. 13, no. 3, pp. 317–322, Jan. 1, 1970.

[10] S. Glotzbach, D. Edgar, and R. Ariagno, "Biological rhythmicity in preterm infants prior to discharge from neonatal intensive care.," *Pediatrics*, vol. 95 2, pp. 231–7, 1995.

[11] D. Goldfarb, "A family of variable-metric methods derived by variational means," *Mathematics of Computation*, vol. 24, no. 109, pp. 23–23, Jan. 1, 1970.

[12] S. N. Graven and J. V. Browne, "Sleep and brain development: The critical role of sleep in fetal and early neonatal brain development," *Newborn and Infant Nursing Reviews*, Brain Development of the Neonate, vol. 8, no. 4, pp. 173–179, Dec. 1, 2008.

[13] R. M. Harper, B. Leake, H. Hoffman, D. O. Walter, T. Hoppenbrouwers, J. Hodgman, and M. B. Sterman, "Periodicity of sleep states is altered in infants at risk for the sudden infant death syndrome," *Science (New York, N.Y.)*, vol. 213, no. 4511, pp. 1030–1032, Aug. 28, 1981.

[14] R. Kaneko, "Ultradian rhythm and its individual differences in self-demand bottle feeding : Suggestions for feeding schedule in group infant care.," *Research and clinical center for child Development*, no. 19, pp. 23–31, 1996.

[15] G. Koch, K. Jost, S. Schulzke, R. Koch, M. Pfister, and A. Datta, "The rhythm of a preterm neonate's life: Ultradian oscillations of heart rate, body temperature and sleep cycles," *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 48, Jun. 2021.

[16] J. Park, "Sleep promotion for preterm infants in the nicu," *Nursing for Women's Health*, vol. 24, pp. 24–35, Feb. 2020.

[17] R. M. Patel, "Short- and long-term outcomes for extremely preterm infants," *American Journal of Perinatology*, vol. 33, no. 3, pp. 318–328, Feb. 2016.

[18] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*, ser. Adaptive computation and machine learning. Cambridge, Mass: MIT Press, 2006, 248 pp.

[19] M. S. Scher, M. W. Johnson, and D. Holditch-Davis, "Cyclicity of neonatal sleep behaviors at 25 to 30 weeks' postconceptional age," *Pediatric Research*, vol. 57, no. 6, pp. 879–882, Jun. 2005.

[20] M. Schmidt, E. Berg, M. Friedlander, and K. Murphy, "Optimizing costly functions with simple constraints: A limited-memory projected quasi-newton algorithm," in *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, vol. 5, Clearwater Beach, Florida USA: PMLR, Apr. 2009, pp. 456–463.

[21] D. F. Shanno, "Conditioning of quasi-newton methods for function minimization," *Mathematics of Computation*, vol. 24, no. 111, pp. 647–647, Sep. 1, 1970.