

Estimation of Consecutively Missed Samples in Fetal Heart Rate Recordings

Guanchao Feng*, J. Gerald Quirk†, Cassandra Heiselman† and Petar M. Djurić*

* Department of Electrical and Computer Engineering

† Department of Obstetrics/Gynecology, Stony Brook University Hospital
Stony Brook University

Stony Brook, NY 11794, USA

Email: *{guanchao.feng, petar.djuric}@stonybrook.edu, †{Cassandra.Heiselman, J.Gerald.Quirk}@stonybrookmedicine.edu

Abstract—During labor, fetal heart rate (FHR) is monitored externally using Doppler ultrasound. This is done continuously, but for various reasons (e.g., fetal or maternal movements) the system does not record any samples for varying periods of time. In many settings, it would be quite beneficial to estimate the missing samples. In this paper, we propose a (deep) Gaussian process-based approach for estimation of consecutively missing samples in FHR recordings. The method relies on similarities in the state space and on exploiting the concept of attractor manifolds. The proposed approach was tested on a short segment of real FHR recordings. The experimental results indicate that the proposed approach is able to provide more reliable results in comparison to several interpolation methods that are commonly applied for processing of FHR signals.

Index Terms—Fetal heart rate, deep Gaussian processes, attractor, state space, consecutively missed samples

I. INTRODUCTION

The fetus depends on the mother for placental exchange of oxygen and carbon dioxide, which relies on adequate maternal blood gas concentrations, uterine blood supply, placental transfer and fetal gas transport. During labor, disruption of any of these can cause fetal hypoxia, which, despite compensatory mechanisms, may lead to acidosis which in turn may cause permanent brain damage or even death of the fetus [1]. The aim of fetal monitoring is to alert obstetricians of such challenges or risks for appropriate and timely intervention [2]. The most widely adopted approach of fetal monitoring during labor is by Cardiotocography (CTG) in which both fetal heart rate and uterine activity signals are recorded and then visually inspected by clinicians [3]. Despite the availability of clinical guidelines regarding FHR evaluation from both the National Institute of Child Health and Human Development (NICHD) and the International Federation of Gynecology and Obstetrics (FIGO) [4], [5], the interpretation of FHR demonstrates high intra- and inter-observer variability due to the subjectivity in visual inspections and complexity of CTG recordings [6]. Therefore, many efforts have been made in automated or computerized analysis of FHR recordings. Computerized methods, equipped with data-driven machine learning techniques, are capable of extracting features and discovering patterns that cannot be seen or interpreted by naked eyes.

This work has been supported by NIH under Award 1RO1HD097188-01.

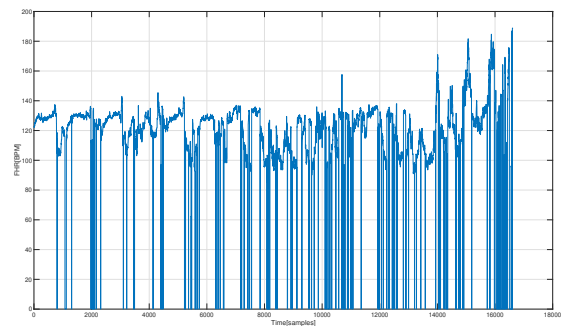


Fig. 1. A segment of un-preprocessed (raw) FHR recording.

One of the hallmarks of FHR recordings obtained externally using Doppler ultrasound is the common existence of missing samples, as shown in Fig. 1, which are often caused by, e.g., fetal or maternal movements and misplaced electrodes. Empirically, the percentage of missing samples varies from 0-40% and 0-10% for external Doppler ultrasound measurements and internal direct fetal electrocardiogram (FECG), respectively. Although FECG can provide more accurate measurements, in practice, the FHR recordings are usually obtained externally using Doppler ultrasound techniques because of their non-invasive nature. It is worth noting that there are still no guidelines on what percentage of missing samples will disqualify an FHR recording from visual inspection or from computerized analysis. Although the high percentage of missing samples can be tolerated by obstetricians because of the robustness of human visual perception, computerized analysis requires proper handling of such distortions. In [7], the authors showed that the values of various popular FHR features can change dramatically if the missing samples are not properly estimated or addressed.

In the literature, the estimation of missing FHR samples are often addressed using methods based on sparse representation and dictionary learning, where the sparse coding step and dictionary update step are applied in an alternating fashion until convergence [8], [9]. However, such methods usually exploit linear transformations while the underlying relationship may be better modeled by non-linear transformations. More

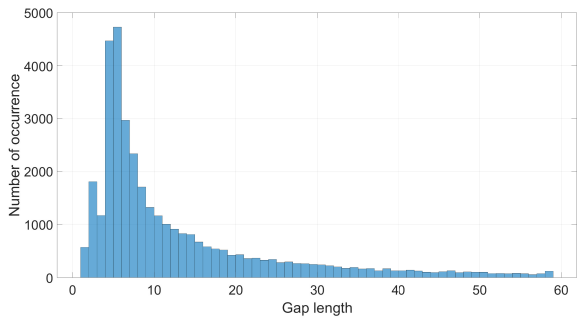


Fig. 2. A histogram of the length of consecutively missing samples, i.e., gap length, in the open access intrapartum cardiotocography database described in [11].

importantly, such methods are not able to provide estimation results within a probabilistic framework, instead, only point estimates of missing samples are provided. In our previous work [10], we proposed a Gaussian processes-based method for estimation of missing FHR samples, where the time instants of missing samples were assumed to have a uniform distribution. However, in reality, the missing samples are more likely to occur in a consecutive manner, i.e., in the form of bursts. In Fig. 2, we present a histogram of the length of consecutively missing FHR samples obtained from an open access intrapartum cardiotocography database [11]. We observe that the most frequently occurring gap length is 5.

In this paper, we propose a (deep) GP-based method that is capable of utilizing attractor manifolds for the estimation of consecutively missing samples in FHR recordings. The underlying idea is that by incorporating an attractor manifold, the proposed approach is able to utilize not only the correlation in time but also the similarity in state space, which can benefit the estimation of consecutively missing samples in FHR recordings. We first validated the ability of the GP-based model for attractor reconstruction on a Lorenz system, where the ground truth attractor manifold is accessible. Then the GP-based model was tested on a short segment of a real FHR signal. The results indicate that the GP-based approach perform better than benchmark models that are commonly applied for gap treatment in automated FHR analysis.

II. BACKGROUND

A. Open Access CTG Database

In this work, we utilized an open access intrapartum CTG database from the Czech Technical University in Prague and the University Hospital in Brno, which has 552 CTG recordings where both FHR and UA signals are sampled at 4Hz. The CTG recordings were carefully selected from 9164 recordings collected between year 2010 and 2012 using various criteria, including, singleton pregnancy, gestational age more than 36 weeks and no a priori known developmental defects, etc. The details about this database are available in [11].

B. State Space Reconstruction

The concept of state space or phase space, which is a space where all the possible states of a studied system are represented, is fundamental in dynamical system modeling. However, often the phase space and the mathematical description of a dynamical system are unknown. Many efforts have been made in developing attractor reconstruction methods in order to reconstruct the phase space of a system. To that end, Takens' theorem, proposed by Floris Takens in [12], is of great importance. It provides theoretical guarantee that, generically, the information about the hidden states of a dynamical system can be reconstructed from a single observation variable of the system, when the stated conditions in the theorem are satisfied. Next, we state the theorem:

Theorem 1 (Takens' theorem): Let \mathcal{M} be a compact manifold of (integer) dimension d . Then for generic pairs (ϕ, y) , where

- $\phi: \mathcal{M} \rightarrow \mathcal{M}$ is a C^2 -diffeomorphism of \mathcal{M} in itself,
- $y: \mathcal{M} \rightarrow \mathbb{R}$ is a C^2 -differentiable function,

the map $\Phi_{(\phi,y)}: \mathcal{M} \rightarrow \mathbb{R}^{2d+1}$ given by

$$\Phi_{(\phi,y)}(x) := (y(x), y(\phi(x)), y(\phi^2(x)), \dots, y(\phi^{2d}(x)))$$

is an embedding of \mathcal{M} in \mathbb{R}^{2d+1} .

The most common choice of ϕ is a delay by a constant τ . A fundamental contribution of Takens' theorem is the claim that for a reliable reconstruction of a manifold \mathcal{M} of dimension d , it is sufficient to have a delay embedding of dimension $E = 2d + 1$. We point out that in reality the true manifold \mathcal{M} that is responsible for generating the data is usually latent. As a result, we do not have any knowledge about its true dimension d (or $E = 2d + 1$). Furthermore, theoretically speaking, τ is a free parameter that can be arbitrarily selected. Nevertheless, in practice, because the length of the time series, i.e., number of observations, is finite, the value of τ will actually affect the quality of attractor reconstruction. If τ is too small, each dimension will be strongly correlated. On the other hand if τ is too large, we will lose information about the underlying dynamical system. In practice, E and τ are often selected using false nearest neighbours [13] and mutual information-based method described in [14], respectively, which are of grid search nature and not principled.

C. Gaussian Processes

A GP is a stochastic process with every finite collection of random variables having a multivariate normal distribution [15]. Essentially it extends a multivariate Gaussian distribution to infinite dimensionality. Therefore, a GP can be seen as a distribution of a real-valued function $f(\mathbf{x})$ in which \mathbf{x} denotes the input, which is usually a vector. In the machine learning literature, GPs provide powerful and flexible Bayesian non-parametric framework for modeling functions and mappings, and they have been successfully applied in both supervised and unsupervised learning [16]. Similar to Gaussian distributions, a GP is completely specified by its mean function $m(\mathbf{x})$ and covariance function $k_f(\mathbf{x}, \mathbf{x}')$, which are defined by $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$, and $k_f(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$.

To reduce the number of hyperparameters, a GP is often assumed to be zero mean. The GP framework frees us from assuming specific analytical form of the latent function and provides proper uncertainties management within the Bayesian framework. With GPs, assumptions and prior knowledge can be conveniently encoded in their covariance functions. The generative process that is commonly used with GPs is as follows:

$$y = y(\mathbf{x}) = f(\mathbf{x}) + \epsilon, \quad (1)$$

where $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is additive white Gaussian noise.

Let $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ be the collection of all input vectors and $\mathbf{y} \in \mathbb{R}^{N \times 1}$ be the corresponding outputs. The covariance matrix is denoted as $\mathbf{K}_{ff} \in \mathbb{R}^{N \times N}$, which is constructed by evaluating the covariance function on \mathbf{X} . Then the prior distribution of the latent \mathbf{f} is given by $p(\mathbf{f}|\mathbf{X}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{K}_{ff})$, where $\boldsymbol{\theta}$ denotes the hyperparameters in GP. The learning requires maximizing the log-likelihood under the additive Gaussian noise assumption, which can be derived as

$$\log p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K}| - \frac{N}{2} \log 2\pi, \quad (2)$$

from which the hyperparameters $\boldsymbol{\theta}$ and noise variance σ_ϵ^2 can be estimated by using training data and maximizing the log-likelihood.

D. Deep Gaussian Processes

The joint Gaussianity enforced by the definition of GP ensured the tractability of the framework. However, as a trade-off, the joint Gaussianity also introduced limitation on the expressiveness of GP, e.g., the simple step function cannot be well modeled by a simple GP prior. It can be shown that a GP is closely related to a neural network [17]. With the development of deep neural networks, deep Gaussian processes (DGPs) emerged naturally. The concept of a DGP was first proposed in [18], where the main idea is that nonlinear mappings will not preserve Gaussianity. Therefore, by adopting a function composition, the expressiveness of the model is improved. As a trade-off, the marginal log-likelihood that is required for training is no longer tractable under a deep GP setting. In [18], the intractability was addressed by using variational inference. Many efforts have been made for inference of DGP, e.g., in [19], a doubly stochastic variational inference algorithm was proposed. In our work, we adopted the inference method from [18].

III. MODEL DESCRIPTION

Given a FHR recording $x(t)$, we first construct an initial attractor manifold reconstruction \mathbf{M}^{init} using delay embedding, with $\tau = 1$ (delay by one sample) and E that is relatively large, e.g., $E = 10$. The intuition of it is that, a relatively large E ensures the conditions in Takens' theorem to be satisfied, and $\tau = 1$ is the minimum delay, which suggests that there is no information loss regarding the underlying dynamics. Consequently, \mathbf{M}^{init} is of high dimension, E , and the variables from the different dimensions are highly correlated. The point

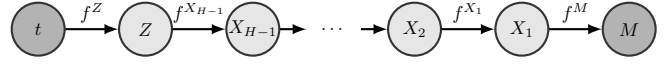


Fig. 3. A general form of dynamically constrained DGPs

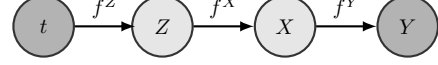


Fig. 4. A dynamically constrained DGP with three layers of mappings.

corresponding to time instant t on $\mathcal{M}_x^{\text{init}}$ is an E -dimensional vector $\mathbf{m}_x^{\text{init}}(t) = [x(t), x(t - \tau), \dots, x(t - (E - 1)\tau)]^T$.

Then we model \mathbf{M} as the output of a dynamically constrained DGP, as illustrated in Fig. 3, where the input \mathbf{Z} is a function of t , i.e., \mathbf{Z} is dynamically constrained. Essentially, we model $\mathbf{M} = f^M(f^{X_1}(\dots f^{X_{H-1}}(f^Z(\mathbf{t}))))$. Since the FHR signal $x(t)$ contains missing samples, there are three types of E dimensional vectors (rows) or points in \mathbf{M} : fully observed vectors collected in \mathbf{M}^{fo} , partially observed vectors collected in \mathbf{M}^{po} , and entirely unobserved vectors collected in \mathbf{M}^{eu} . We train the DGP using the pairs of fully observed vectors and their corresponding time instants, i.e., we set \mathbf{Y} and \mathbf{t} in Fig. 3 to \mathbf{M}^{fo} and \mathbf{t}^{fo} , respectively for training.

Without loss of generality, we continue the discussion with a simpler example of DGP, shown in Fig. 4.

The generative process according to this DGP takes the form

$$\begin{aligned} y_{ni} &= f_i^Y(\mathbf{x}_n) + \epsilon_{ni}^Y, & i &= 1, \dots, d_y, & \mathbf{x}_n &\in \mathbb{R}^{d_x}, \\ x_{nj} &= f_j^X(\mathbf{z}_n) + \epsilon_{nj}^X, & j &= 1, \dots, d_x, & \mathbf{z}_n &\in \mathbb{R}^{d_z}, \\ z_{nl} &= f_l^Z(t_n) + \epsilon_{nl}^Z, & l &= 1, \dots, d_z, & t &\in \mathbb{R}, \end{aligned} \quad (3)$$

where ϵ_{ni}^Y , ϵ_{nj}^X and ϵ_{nl}^Z are additive white Gaussian noises, f^Y , f^X and f^Z are the latent mappings governed by three different GPs, i.e., $f^Y \sim GP(\mathbf{0}, k^Y(\mathbf{X}, \mathbf{X}'))$, $f^X \sim GP(\mathbf{0}, k^X(\mathbf{Z}, \mathbf{Z}'))$, and $f^Z \sim GP(\mathbf{0}, k^Z(t, t'))$. In this work, the covariance function of the GP for the dynamic layer, i.e., the first layer, is a Matérn class covariance function with $\nu = \frac{3}{2}$, and the covariance functions of the GPs for the remaining layers are RBFs with automatic relevance determination. The dimension for all layers, except for the dynamic layers, are set to E , although different configurations can be explored.

The learning requires maximization of the log-marginal likelihood,

$$\log p(\mathbf{Y}|\mathbf{t}) = \log \int_{\mathbf{X}, \mathbf{Z}} p(\mathbf{Y}|\mathbf{X})p(\mathbf{X}|\mathbf{Z})p(\mathbf{Z}|\mathbf{t})d\mathbf{X}d\mathbf{Z}. \quad (4)$$

Because of the nonlinear mappings, Gaussianity is not preserved. Consequently, the log-likelihood in Eq. 4 is intractable. This difficulty can be addressed by introducing $N_p \leq N$ inducing or pseudo input-output pairs for each intermediate or latent layers, such that, in each layer, the inducing input-output pairs are mapped by the same latent mapping in that layer. Specifically, $\tilde{\mathbf{Z}} \in \mathbb{R}^{N_p \times d_z}$ and $\mathbf{U}^X = f^X(\tilde{\mathbf{Z}})$ where $\mathbf{U}^X \in \mathbb{R}^{N_p \times d_x}$ are introduced for the first intermediate layer. Similarly, $\tilde{\mathbf{X}} \in \mathbb{R}^{N_p \times d_x}$ and $\mathbf{U}^Y = f^Y(\tilde{\mathbf{X}})$ where $\mathbf{U}^Y \in \mathbb{R}^{N_p \times d_y}$ are introduced for the second intermediate

layer. The joint probability density function (pdf) can be formulated accordingly:

$$\begin{aligned} p(\mathbf{Y}, \mathbf{F}^Y, \mathbf{F}^X, \mathbf{X}, \mathbf{Z}, \mathbf{U}^Y, \mathbf{U}^X, \tilde{\mathbf{X}}, \tilde{\mathbf{Z}}|\mathbf{t}) \\ = p(\mathbf{Y}|\mathbf{F}^Y)p(\mathbf{F}^Y|\mathbf{U}^Y, \mathbf{X})p(\mathbf{U}^Y|\tilde{\mathbf{X}}) \\ \times p(\mathbf{X}|\mathbf{F}^X)p(\mathbf{F}^X|\mathbf{U}^X, \mathbf{Z})p(\mathbf{U}^X|\tilde{\mathbf{Z}})p(\mathbf{Z}|\mathbf{t}). \end{aligned} \quad (5)$$

We note that the intractability is caused by the marginalization of the latent variables, which corresponds to the terms $p(\mathbf{F}^Y|\mathbf{U}^Y, \mathbf{X})$ and $p(\mathbf{F}^X|\mathbf{U}^X, \mathbf{Z})$. With the inference method proposed in [18], these two terms can be canceled out and a variational lower bound on Eq. 4 can be obtained as

$$\begin{aligned} \mathcal{G}(\mathbf{Y}, \mathbf{F}^Y, \mathbf{X}, \mathbf{F}^X, \mathbf{Z}, \mathbf{U}^X, \mathbf{U}^Y|\mathbf{t}) \\ = \frac{p(\mathbf{Y}|\mathbf{F}^Y)p(\mathbf{U}^Y)p(\mathbf{X}|\mathbf{F}^X)p(\mathbf{U}^X)p(\mathbf{Z}|\mathbf{t})}{q(\mathbf{U}^Y)q(\mathbf{X})q(\mathbf{U}^X)q(\mathbf{Z})}. \end{aligned} \quad (6)$$

The detailed derivation of Eq. 6 can be found in [20] and [21]. This tractable lower bound not only can be used as the objective function for training, but also for providing guidance for setting the number of layers. Some recently analytical analysis on the setting the number of inducing points can be found in [22]. In our experiments, we adopted the DGP shown in Fig. 4, and the number of inducing points was $N_p = N$.

For prediction, the true posterior distribution $p(\mathbf{M}^{po}|\mathbf{M}^{fo})$ or $p(\mathbf{M}^{eu}|\mathbf{M}^{fo})$ are approximated using the Gaussian variational distribution $q(\mathbf{M}^{eu})$ or $q(\mathbf{M}^{fo})$ under the variational framework. In our work, we use the mean of $q(\mathbf{M}^{eu})$ and $q(\mathbf{M}^{fo})$ as corresponding point estimates, which can be regarded as MAP estimates since the mean of a Gaussian distribution is also its mode. The uncertainty in learning and prediction is embedded in the covariance, which can provide additional guidance in model selection. The specific derivation and computation can be found in [20]. Finally, for the missing samples of the time series $x(t)$, their final estimates are averages over all available estimates, provided in both recovered \mathbf{M}^{eu} and \mathbf{M}^{po} .

IV. EXPERIMENTS AND RESULTS

A. Synthetic Data

We first show that the GP-based method can provide a better attractor reconstruction comparing with direct delay embedding on the well known Lorenz system defined by Eq. 7, which was first studied by Edward Lorenz in modeling atmospheric convection. We recall that the Lorenz system is nonlinear, non-periodic, three-dimensional and deterministic,

$$\begin{aligned} dx/dt &= a(y - x), \\ dy/dt &= x(c - z) - y, \\ dz/dt &= xy - bz. \end{aligned} \quad (7)$$

We generated the ground truth attractor (a set of solutions for Eq. 7) \mathcal{M} of length 369 with a classic set of parameter values $a = 10$, $b = \frac{8}{3}$, and $c = 28$. Then $x(t)$ was used to reconstruct the attractor using dynamically constrained DGP and direct delay embedding, respectively. For direct delay embedding, E and τ were optimally set using the grid

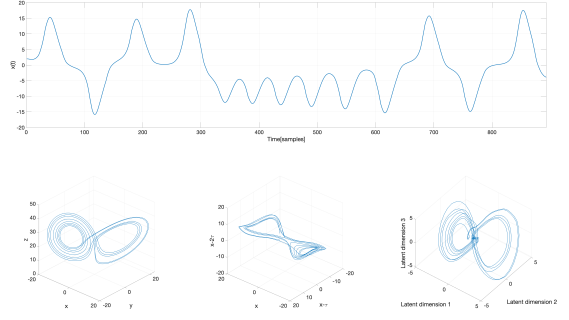


Fig. 5. The time series $x(t)$ generated with Lorenz system (upper), the ground truth attractor (bottom left), reconstructed attractor using direct delay embedding (bottom middle) and reconstructed attractor using DGP (bottom right).

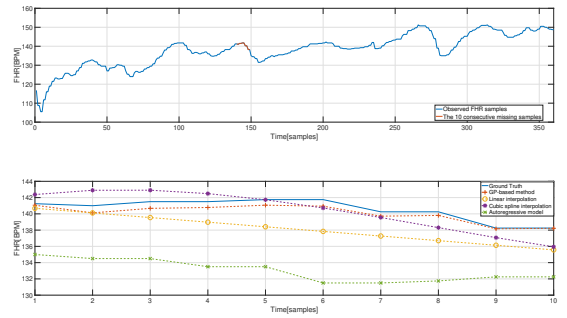


Fig. 6. The segment of FHR recording adopted for experiments (upper), and the corresponding estimation results for different methods (bottom).

search method described in [13] and [14], respectively. The reconstruction results, ground truth attractor \mathcal{M} , and $x(t)$ are shown in Fig. 5, where we can see that the estimation results of the DGP are more topologically similar to the ground truth attractor than the results obtained by delayed embedding.

B. Real Segment of FHR Recording

We selected a short FHR segment which contains 360 samples, then 10 consecutive samples were randomly selected to be missing, by setting their values to zero. We then implemented the GP-based approach to recover the values of consecutively missing samples. The estimation results are shown in Fig. 6, where the estimation results provided by linear interpolation, cubic spline interpolation, and an autoregressive model are also included for benchmark purpose. The order of the autoregressive model was optimally selected with Bayesian information criterion (BIC) from model orders from 2 to 10.

The performance of each estimation method is summarized in Table I, where the mean squared error (MSE) was adopted as a performance metric. The results, clearly show that the proposed method outperformed its competitors.

TABLE I
COMPARISONS OF PERFORMANCE

estimation method	MSE [BPM]
GP-based method	0.36
Linear interpolation	7.08
Cubic spline interpolation	1.98
Autoregressive model	58.83

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a Bayesian approach for estimation of consecutive missing samples in FHR recordings based on deep Gaussian processes. By way of attractor manifold learning, our approach has the capacity of utilizing the similarity in state space instead of only considering the correlation in time. The experimental results on a real segment of FHR recording showed that the estimates provided by this GP-based approach are closer to the ground truth and that the proposed method achieved a much better MSE compared to that of the benchmark methods. We point out that we modeled the dynamic of attractor manifold as a function of time, which may not be suitable, depending on the shape of the attractor manifold, e.g., involving sharp changes. In such cases, the resulting posteriors will have large variance and the GPs will have low signal-to-noise ratio (SNR). In future work, we plan to explore the direction of modeling the dynamic of attractor manifold as functions of previously observed samples on the attractor manifold, where the length of history to be included must be determined carefully.

REFERENCES

- [1] C. S. Bobrow and P. W. Sothill, "Causes and consequences of fetal acidosis," *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 80, no. 3, pp. F246–F249, 1999.
- [2] K. Afors and E. Chandrharan, "Use of continuous electronic fetal monitoring in a preterm fetus: clinical dilemmas and recommendations for practice," *Journal of pregnancy*, vol. 2011, 2011.
- [3] G. Georgoulas, P. Karvelis, J. Spilka, V. Chudáček, C. D. Stylios, and L. Lhotská, "Investigating ph based evaluation of fetal heart rate (FHR) recordings," *Health and technology*, vol. 7, no. 2-3, pp. 241–254, 2017.
- [4] D. Ayres-de Campos, C. Y. Spong, E. Chandrharan, and F. I. F. M. E. C. Panel, "FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography," *International Journal of Gynecology & Obstetrics*, vol. 131, no. 1, pp. 13–24, 2015. [Online]. Available: <http://dx.doi.org/10.1016/j.ijgo.2015.06.020>
- [5] G. A. Macones, G. D. Hankins, C. Y. Spong, J. Hauth, and T. Moore, "The 2008 National Institute of Child Health and Human Development workshop report on electronic fetal monitoring: Update on definitions, interpretation, and research guidelines," *Journal of Obstetric, Gynecologic, & Neonatal Nursing*, vol. 37, no. 5, pp. 510–515, 2008.
- [6] A. Georgieva, P. Abry, V. Chudáček, P. M. Djurić, M. G. Frasch, R. Kok, C. A. Lear, S. N. Lemmens, I. Nunes, A. T. Papageorghiou *et al.*, "Computer-based intrapartum fetal monitoring and beyond: A review of the 2nd workshop on signal processing and monitoring in labor (october 2017, oxford, uk)," *Acta obstetrica et gynecologica Scandinavica*, vol. 98, no. 9, pp. 1207–1217, 2019.
- [7] J. Spilka, V. Chudáček, M. Burša, L. Zach, M. Huptych, L. Lhotská, P. Janků, and L. Hruban, "Stability of variability features computed from fetal heart rate with artificially infused missing data," in *Computing in Cardiology (CinC)*, 2012. IEEE, 2012, pp. 917–920.
- [8] V. P. Oikonomou, J. Spilka, C. D. Stylios, and L. Lhotská, "An adaptive method for the recovery of missing samples from FHR time series." in *CBMS*, 2013, pp. 337–342.

- [9] F. Barzideh, J. Urdal, K. Hussein, K. Engan, K. Skretting, P. Mdoe, B. Kamala, and S. Brunner, "Estimation of missing data in fetal heart rate signals using shift-invariant dictionary," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 762–766.
- [10] G. Feng, J. G. Quirk, and P. M. Djurić, "Recovery of missing samples in fetal heart rate recordings with Gaussian processes," in *Signal Processing Conference (EUSIPCO), 2017 25th European*. IEEE, 2017, pp. 261–265.
- [11] V. Chudáček, J. Spilka, M. Burša, P. Janků, L. Hruban, M. Huptych, and L. Lhotská, "Open access intrapartum CTG database," *BMC pregnancy and childbirth*, vol. 14, no. 1, p. 16, 2014.
- [12] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*. Springer, 1981, pp. 366–381.
- [13] M. B. Kennel, R. Brown, and H. D. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Physical Review A*, vol. 45, no. 6, p. 3403, 1992.
- [14] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, p. 1134, 1986.
- [15] C. E. Rasmussen, *Gaussian Processes for Machine Learning*. Citeseer, 2006.
- [16] G. Feng, J. G. Quirk, and P. M. Djurić, "Supervised and unsupervised learning of fetal heart rate tracings with deep gaussian processes," in *2018 14th Symposium on Neural Networks and Applications (NEUREL)*. IEEE, 2018, pp. 1–6.
- [17] D. Duvenaud, O. Rippel, R. Adams, and Z. Ghahramani, "Avoiding pathologies in very deep networks," in *Artificial Intelligence and Statistics*, 2014, pp. 202–210.
- [18] A. Damianou and N. Lawrence, "Deep gaussian processes," in *Artificial Intelligence and Statistics*, 2013, pp. 207–215.
- [19] H. Salimbeni and M. Deisenroth, "Doubly stochastic variational inference for deep Gaussian processes," in *Advances in Neural Information Processing Systems*, 2017, pp. 4591–4602.
- [20] A. Damianou, "Deep gaussian processes and variational propagation of uncertainty," Ph.D. dissertation, University of Sheffield, 2015.
- [21] A. C. Damianou, M. K. Titsias, and N. D. Lawrence, "Variational inference for latent variables and uncertain inputs in Gaussian processes," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1425–1486, 2016.
- [22] D. R. Burt, C. E. Rasmussen, and M. Van Der Wilk, "Rates of convergence for sparse variational gaussian process regression," *arXiv preprint arXiv:1903.03571*, 2019.