DISCOVERING CAUSALITIES FROM CARDIOTOCOGRAPHY SIGNALS USING IMPROVED CONVERGENT CROSS MAPPING WITH GAUSSIAN PROCESSES

Guanchao Feng[*], **J. Gerald Quirk**[†], **Petar M. Djurić** [*]

[*]Department of Electrical and Computer Engineering, Stony Brook University

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

Abstract

Convergent cross mapping (CCM) is designed for causal discovery in coupled time series, where Granger causality may not be applicable because of a separability assumption. However, CCM is not robust to observation noise which limits its applicability on signals that are known to be noisy. Moreover, the parameters for state space reconstruction need to be selected using grid search methods. In this paper, we propose a novel improved version of CCM using Gaussian processes for discovery of causality from noisy time series. Specifically, we adopt the concept of CCM and carry out the key steps using Gaussian processes within a non-parametric Bayesian probabilistic framework in a principled manner. The proposed approach is first validated on simulated data, and then used for understanding the interaction between fetal heart rate and uterine activity in the last two hours before delivery and of interest in obstetrics. Our results indicate that uterine activity affects the fetal heart rate, which agrees with recent clinical studies.

Index Terms—

Convergent cross mapping; state space reconstruction; Gaussian processes; fetal heart rate; uterine activity

1. INTRODUCTION

During labor, a fetus can be deprived of adequate levels of oxygen and can become hypoxic and acidotic. If the oxygen supply drops below a certain threshold, asphyxia occurs, and this can lead to permanent brain damage or even death of the fetus [1]. Cardiotocography (CTG) is the most widely used technology for monitoring the well-being of fetuses during labor. CTG comprises of the fetal heart rate (FHR) and uterine activity (UA) signals, which are both recorded and visually inspected by clinicians [2]. The purpose of CTG is to alert obstetricians of such alterations in blood flow and oxygen content that are reflected in the CTG patterns for appropriate and timely intervention. The interpretation of FHR recordings is a highly intricate and complex task with high inter- and intra-variable evaluations among obstetricians, notwithstanding the availability of various clinical guidelines from both the National Institute of Child Health and Human Development (NICHD) and the International Federation of Gynecology and Obstetrics (FIGO) [3, 4]. In fact, the current guidelines for FHR evaluation have been criticized for simplistic interpretation [5]. The reliable

interpretation of CTG tracings requires better understanding of the interaction between FHR and UA signals. In this paper, we address the problem of making inference about causality from CTG signals, which interestingly, has largely been overlooked in the literature.

The gold standard for identifying causal relationships is using controlled randomized experiments. In many situations, however, these experiments cannot be performed [6]. When we work with signals, their samples are usually stored in the form of time series. The samples are ordered in time, which helps in addressing the detection of causality because cause should occur before the effect [7, 8]. The Granger causality is the best known concept for discovering causalities [9]. The concept relies on two fundamental principles: (a) the effect does not precede its cause in time, and (b) the causing series contains unique information about the series being caused that is not available otherwise. However, when two time series are dynamically coupled, the second principle of the Granger causality is violated, and consequently, the causal discovery results may be unreliable [10].

Another approach for determining causality is known as convergent cross mapping (CCM), which was proposed in [11]. The method is designed for coupled time series and is based on state space reconstruction (SSR). We recall that in dynamical systems theory two time series are causally related if they are from the same dynamical system, or equivalently, share a common attractor manifold $\mathcal{M}$. Further, the signature of the causing series is embedded in the effect series [11]. Although CCM has been successfully applied to perform causal inference in many communities, e.g., social media [12] and neuroscience [13], it can be shown that CCM is sensitive to observation noise [14]. To improve the applicability of CCM, some variants of CCM have been proposed, e.g. [15, 16].

In this paper, we propose a fully Gaussian process (GP)-based version of CCM. The GP-based SSR step is capable of automatic learning an attractor manifold of better quality from noisy observations in a principled manner, unlike the original SSR method in CCM where the parameters for reconstruction are usually selected with grid search. For the cross mapping step, the GP-based method is able to provide cross mapping results under a probabilistic framework. More importantly, because of the Bayesian nature of the GP framework, the GPs are data efficient and robust to overfitting. As a result, our GP-based CCM can provide better convergence which is critical in distinguishing causation from correlation. We first validated the GP based approach using the well-studied Lorenz system with and without observation noise. Then we implemented the original CCM and the GP-based CCM, respectively, on a segment of real CTG recordings. The results of the original CCM are ambiguous, whereas the results of the GP-based CCM clearly indicate that the changes in the UA signal cause changes in the FHR signal, and not vice versa. Our finding is consistent with recent clinical studies [17].

2. BACKGROUND

2.1. Takens' Theorem

The CCM framework is built on Takens' theorem proposed by Floris Takens in [18], which is of great importance in attractor reconstruction, i.e., the reconstruction of the state space of a system. When the conditions of the theorem are satisfied, the theorem provides guarantees

that, generically, the information about hidden states of a dynamical system can be reconstructed from a single observation variable of the system. 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*

- $\phi: \mathcal{M} \to \mathcal{M}$ *is a C^2-diffeomorphism of $\mathcal{M}$ in itself,*

- $y: \mathcal{M} \to \mathbb{R}$ *is a C^2-differentiable function, the map $\Phi_{(\phi, y)}: \mathcal{M} \to \mathbb{R}^{2d+1}$ given by*

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

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 $E = 2d + 1$.

2.2. Convergent Cross Mapping

The CCM framework is composed of two steps. The first step is state space construction (SSR) using each time series. Given a time series $x(t)$, a shadow manifold $\mathcal{M}_x^{\mathrm{DE}}$ of dimension E, as an estimate or reconstruction of latent attractor manifold $\mathcal{M}_x$, is constructed using delay embedding. The point corresponding to time instant t on $\mathcal{M}_x^{\mathrm{DE}}$ is an E-dimensional vector $\mathbf{m}_x^{\mathrm{DE}}(t) = [x(t), x(t - \tau), \dots, x(t - (E - 1)\tau)]^{\top}$. If two time series, $x(t)$ and $y(t)$, are from the same dynamical system, $\mathcal{M}_x^{\mathrm{DE}}$ and $\mathcal{M}_y^{\mathrm{DE}}$ should be topologically similar, since they both are diffeomorphic to the true manifold $\mathcal{M}$.

In the second step, given two shadow manifolds $\mathcal{M}_x^{\mathrm{DE}}$ and $\mathcal{M}_y^{\mathrm{DE}}$ their correspondence is tested for causal discovery. Specifically, given a point in $\mathcal{M}_y^{\mathrm{DE}} \in \mathbb{R}^{N \times E}$, CCM will find its $E + 1$ nearest neighbors (the minimum number of points to bound a simplex in $\mathbb{R}^E$) in $\mathcal{M}_y^{\mathrm{DE}}$ and their corresponding time indices and then test whether these time indices can be used to identify nearby points on $\mathcal{M}_x^{\mathrm{DE}}$, and vice versa. This is implemented using cross mapping with simplex projection, i.e., estimating $x(t)$ with $\mathcal{M}_y^{\mathrm{DE}}$ denoted as $\hat{x}(t) | \mathcal{M}_y^{\mathrm{DE}}$, and vice versa. Essentially, the CCM test measures the extent to which the historical record of one time series can reliably estimate states of the other time series. If $x(t)$ is a stochastic driver or cause of $y(t)$ (e.g., x is an environmental driver and y is a population variable), information about the states of x can be recovered from the observations of y, but not vice versa.

Furthermore, the CCM takes convergence into consideration that not only distinguishes CCM from general cross prediction, but also allows for distinguishing causation from simple correlation. The convergence means that cross-mapped estimates improve in estimation capacity, which is usually measured by correlation, with the length of time-series, L (sample size used to construct the history). If the Pearson correlation coefficient (PCC), denoted as ρ,

is adopted as a capacity metric and the conditions of Takens' theorem are satisfied, it can be proved that the PCC as a function of L will converge to 1 as L approaches infinity. In practice, because of observation noise and finite number of observations, the correlation will only converge to a value less than 1 [11].

The choice of reconstruction parameters E and τ need to be carefully selected. In many situations, the true manifold $\mathcal{M}$ that is responsible for generating the data is usually unknown. As a result, we do not have any knowledge about the true dimension d of the manifold $\mathcal{M}$ (or $E = 2d + 1$). In practice, the common approach of choosing E is using false nearest neighbours [19] in a grid search manner, which is not principled. In theory, τ is a free parameter and can be arbitrarily selected. However, because in reality the number of observation is finite, the value of τ will actually affect the quality of $\mathcal{M}^{\mathrm{DE}}$ in term of capturing the underlying dynamics of the system. One common approach for choosing τ is based on mutual information and is also obtained by a grid search [20].

2.3. Gaussian Processes

In machine learning literature, GPs provide powerful and flexible Bayesian nonparametric framework for modeling functions and mappings, and they have been successfully applied in both supervised and unsupervised learning [21, 22]. By definition, a GP indexed by $\mathbf{x}$ is a stochastic process in which every finite collection of random variables has a multivariate normal distribution, and it 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 $k_f(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$. Conceptually, a real valued function $f(\mathbf{x})$ can be seen as a vector with infinite dimensionality. Therefore, GPs are suitable for specifying the prior distribution of a latent function $f(\mathbf{x})$, and our prior knowledge and assumption about $f(\mathbf{x})$ can be conveniently encoded in the design of the covariance function without assuming any analytical form of $f(\mathbf{x})$. To reduce the number of hyperparameters, a GP is often assumed to be zero mean, and we write $f(\mathbf{x}) \sim GP(0, k_f(\mathbf{x}, \mathbf{x}'))$. The covariance function $k_f(\mathbf{x}, \mathbf{x}')$ is of great importance because it maps the distance, or similarity, between the inputs $\mathbf{x}$ and $\mathbf{x}'$ to the covariance between the outputs $f(\mathbf{x})$ and $f(\mathbf{x}')$.

3. MODEL DESCRIPTION

3.1. State Space Reconstruction

Given a time series, we first construct a shadow manifold $\mathcal{M}^{\mathrm{init}}$ using delay embedding, with $\tau = 1$ (delay by one sample) and E that is relatively large, e.g., $E = 20$. 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, $\mathcal{M}^{\mathrm{init}}$ is of high dimension and the variables from the different dimensions are highly correlated. Therefore, $\mathcal{M}^{\mathrm{init}}$ is not suitable for simplex projection where $E+1$ nearest neighbours need to be identified because the Euclidean distance in high dimensions is large, and neighbours in high dimensions are sparse.

We use the Bayesian GP latent variable model (GPLVM) [23], to infer the low dimensional manifold $\mathcal{M}^{\mathrm{GP}}$ that is responsible for generating $\mathcal{M}^{\mathrm{init}}$. Let $\mathcal{M}^{\mathrm{init}} \in \mathbb{R}^{N \times Q}$ be a matrix whose rows lie in $\mathcal{M}^{\mathrm{init}}$ and similarly, $\mathcal{M}^{\mathrm{GP}} \in \mathbb{R}^{N \times E}$ is a matrix with rows that lie in $\mathcal{M}^{\mathrm{GP}}$. The generative process can then be expressed as follows:

$$M^{\mathrm{init}} = \mathbf{f}\left(M^{\mathrm{GP}}\right) + \epsilon, \tag{1}$$

where $\epsilon \in \mathbb{R}^{N \times Q}$ is a matrix whose rows are zero mean Gaussian with covariance $\sigma_\epsilon^2 I$. Since our purpose is not to predict how the system will evolve on the attractor manifold in the future, we initialize each dimension in $\mathbf{f}$ as an independent draw from a GP, i.e., $f \sim GP(0, k(\mathbf{x}, \mathbf{x}'))$, where the covariance function is a Q-dimensional radial basis function (RBF), which has the following form:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{q=1}^{Q} \frac{1}{\ell_q}(x_q - x_q')^2\right). \tag{2}$$

The learning requires maximizing the marginal likelihood given by:

$$p\left(M^{\mathrm{init}}\right) = \int p\left(M^{\mathrm{init}} | M^{\mathrm{GP}}\right) p\left(M^{\mathrm{GP}}\right) dM^{\mathrm{GP}}. \tag{3}$$

Unlike the GP regression framework, this marginal likelihood is intractable because M^{GP} and M^{init} are related by the covariance function in a highly nonlinear manner, and in general, nonlinear mapping will not preserve Gaussianity. This is handled in [23] by employing variational inference, and approximating the true posterior $p(M^{\mathrm{GP}} | M^{\mathrm{init}})$ by a Gaussian variational distribution $q(M^{\mathrm{GP}})$, from which a tractable lower bound on the marginal likelihood was obtained and then adopted for learning.

In our work, the mean of $q(M^{\mathrm{GP}})$ is used as reconstructed attractor manifold, and the covariance $q(M^{\mathrm{GP}})$ is adopted for measuring the uncertainty in the learning. In the covariance function, shown in (2), each dimension q is associated with $r_q = \frac{1}{\ell_q}$, which can be seen as an importance weight of dimension q in the modeling. Since ℓ_q as a hyperparameter will be automatically learned from the data, this is known as automatic relevance determination (ARD). We initialize $Q = E$, and use ARD for learning the dimension of $\mathcal{M}^{\mathrm{GP}}$. The values of the importance weights of irrelevant or redundant embedding dimensions will be close to zero, and the final learned Q will be much smaller than E.

3.2. Cross Mapping

Let the two reconstructed attractor manifolds be $\mathcal{M}_x^{\mathrm{GP}}$ and $\mathcal{M}_y^{\mathrm{GP}}$ and that they correspond to two time series $x(t)$ and $y(t)$, respectively. We will use them to discover causality by using GP-based cross mapping. For convenience, we only discuss the cross mapping of $\hat{x}(t) | \mathcal{M}_y^{\mathrm{GP}}$ using GP regression framework.

For each time instant t_0 in $L - E + 1 \quad t_0 \quad L$, we first find $\mathbf{m}_y^{\mathrm{GP}}(t_0)$ and its corresponding $Q + 1$ nearest neighbours for $t < t_0$ on $\mathscr{M}_y^{\mathrm{GP}}$, denoted as $\left\{ \mathbf{m}_y^{\mathrm{GP}}\!\left(t_{n_i}^0\right) \right\}_{i=1}^{Q+1}$, where $t_{n_i}^0$ is time index of the ith nearest neighbor of $\mathbf{m}_y^{\mathrm{GP}}(t_0)$. Then we train a GP regression model to learn the mapping from $\widetilde{\mathbf{y}}_{t_0} = \left[y\!\left(t_{n_1}^0\right), y\!\left(t_{n_2}^0\right), ..., y\!\left(t_{n_{Q+1}}^0\right) \right]^{\top}$ to $y(t_0)$, using the generative process as follows:

$$y(t_0) = g\!\left(\widetilde{\mathbf{y}}_{t_0}\right) + \eta, \tag{4}$$

where g is governed by a GP with zero mean and covariance function $k_g(\mathbf{x}, \mathbf{x}')$ which is a $Q + 1$ dimensional RBF, and $\eta \sim \mathcal{N}\!\left(0, \sigma_\eta^2\right)$ is a white Gaussian noise.

Let $\widetilde{\mathbf{Y}} = \left\{ \widetilde{\mathbf{y}}_{t_i} \right\}_{i=1}^{N}$ denote the collection of all input vectors, and $\mathbf{K}_{gg}$ the covariance matrix obtained by evaluating the covariance function for $\widetilde{\mathbf{Y}}$, i.e., $\mathbf{K}_{gg} = \mathbf{k}_g(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{Y}})$. Then the prior probability density function (pdf) of $\mathbf{g}$ given $\widetilde{\mathbf{Y}}$ is given by

$$p(\mathbf{g} | \widetilde{\mathbf{Y}}, \boldsymbol{\theta}) = \mathcal{N}\!\left(\mathbf{g} | 0, \mathbf{K}_{gg}\right). \tag{5}$$

The hyperparameters $\boldsymbol{\theta}$ in the covariance function and the noise variance σ_η^2 are learned in the training stage by maximizing the (tractable) model evidence,

$$\begin{aligned} \log p(\mathbf{y} | \widetilde{\mathbf{Y}}, \boldsymbol{\theta}) &= \log \mathcal{N}\!\left(\mathbf{y} | 0, \mathbf{K}_{gg} + \sigma_\eta^2 \mathbf{I}\right) \\ &= -\frac{1}{2}\mathbf{y}^T \mathbf{K}^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K}| - \frac{N}{2}\log 2\pi, \end{aligned} \tag{6}$$

where $\mathbf{K} = \mathbf{K}_{gg} + \sigma_\eta^2 \mathbf{I}$. If we have test inputs $\widetilde{\mathbf{y}}_*$, the predictive pdf $p\!\left(g_* | \widetilde{\mathbf{y}}_*, \widetilde{\mathbf{Y}}, \mathbf{y}, \boldsymbol{\theta}\right)$ will be Gaussian with a mean given by:

$$E(g_*) = \left[\mathbf{k}_g\!\left(\widetilde{\mathbf{y}}_*, \widetilde{\mathbf{Y}}\right) \right] \mathbf{K}^{-1}\mathbf{y}. \tag{7}$$

Since we need to perform cross mapping, we modify (7) as

$$\hat{x}(t_0) = \left[\mathbf{k}_g\!\left(\widetilde{\mathbf{x}}_{t_0}, \widetilde{\mathbf{Y}}\right) \right] \mathbf{K}^{-1}\mathbf{y}, \tag{8}$$

where $\widetilde{\mathbf{x}}_{t_0}$ is formed from $x(t)$ and corresponding to the ordered time indices of the neighbours of $\mathbf{m}_y^{\mathrm{GP}}(t_0)$ on $\mathscr{M}_x^{\mathrm{GP}}$. The variance of the cross mapping can also be computed readily [21]. If x drives y, the estimation capacity of $\hat{x}(t) | \mathscr{M}_y^{\mathrm{GP}}$, which is usually measured by the correlation coefficient ρ between $\hat{x}(t)$ and $x(t)$, will be improved with the increase of L and will converge.

4. EXPERIMENTS AND RESULTS

4.1. Synthetic Data

We first validate our GP-based method on the well-known Lorenz system defined by

$$dx/dt = a(y - x),$$
$$dy/dt = x(c - z) - y,$$ \hfill (9)
$$dz/dt = xy - bz.$$

The system is nonlinear, non-periodic, three-dimensional and deterministic.

We generated a Lorenz attractor (a set of chaotic solutions of the Lorenz system) $\mathcal{M}$ of length 365 with (9) and a classic set of parameter values $a = 10$, $b = \frac{8}{3}$, and $c = 28$, along with the three time series obtained by projecting points in $\mathcal{M}$ to the X, Y and Z axes shown in Fig. 1.

To make our experiments more realistic, we added white Gaussian noise $e \sim \mathcal{N}(0, \sigma_e^2)$, and we set $\sigma_e^2 = 0$, 1 and 9, respectively. Then we implemented the original CCM and our GP-based method to discover causality between $X(t)$ and $Y(t)$. Recall that in the original CCM, the SSR parameters are selected using grid search [19, 20]. From (9), we know that the true causal relationship is that $X(t)$ is a cause of $Y(t)$, and $Y(t)$ is also a cause of $X(t)$.

The comparison of SSR results for $Y(t)$ are shown in Fig. 2, where we see that the GP-based SSR consistently provided better reconstructions. This was especially the case for $\sigma_e^2 = 9$, where the $\mathcal{M}_Y^{\text{DE}}$ is essentially unrecognizable, whereas $\mathcal{M}_Y^{\text{GP}}$ is still topologically similar to $\mathcal{M}$. It is worth noting that when $\sigma_e^2 = 0$, i.e., noise-free, the $\mathcal{M}_Y^{\text{DE}}$ is distorted, especially for the curves within the left lobe, whereas, in $\mathcal{M}_Y^{\text{GP}}$ there is no such distortion. Moreover, the true dimensionality of $\mathcal{M}$ is correctly estimated with and without observation noise.

Finally, the test results of both methods are summarized in Fig. 3. We can see that both methods are able to identify the correct causal relationship when $\sigma_e^2 = 0$. However, for $\sigma_e^2 = 1$ and 9, the GP-based method demonstrated better convergence, which is crucial for distinguishing causation from correlation.

4.2. Real CTG Segment

In our experiments with real CTG data, we selected data records from an open access database that were acquired at the obstetrics ward of the University Hospital in Brno, Czech Republic. A detailed description of the database can be found in [24]. We applied our GP-based method on a real CTG segment of length 480 samples, which corresponds to a duration of 2 minutes (the sampling rate for both FHR and UA signals was 4 Hz). The FHR and UA signals and their corresponding attractor manifolds obtained by the GP-based method are shown in Fig. 4. Although the FHR and UA recordings are very different, their attactor manifolds are similar.

The test results of the CTG segment using the original CCM and our GP-based method are shown in Fig. 5, where X denotes UA signal and Y denotes FHR signal. The results of the original CCM is ambiguous, since for both directions, the correlation coefficient shows decreasing trend, after $L = 400$. Moreover, there was no convergence for both directions. The test results of the GP-based method clearly showed that changes in the UA signal is a cause

of changes in the FHR signal and not vice versa. For cross mapping from UA to FHR $(\hat{Y}(t)|\mathscr{M}_X^{GP})$, the correlation coefficient showed high variance and no convergence, which indicates that FHR is not a cause of UA. Meanwhile, in the cross mapping from FHR to UA $(\hat{X}(t)|\mathscr{M}_Y^{GP})$, the correlation coefficient showed small variance, and was gradually improved with L, and finally converged around 0.99. This suggests that changes in the UA signal cause changes in the FHR signal. This finding is consistent with clinical studies [17].

5. CONCLUSION

In this paper, we proposed a fully GP-based version of CCM for casual discovery that is robust to observation noise. Both the SSR and cross mapping steps are carried out using GPs within the Bayesian nonparametric probabilistic framework, which is data efficient and robust to overfitting. The method is also capable of properly handling uncertainties. We first validated our approach on synthetic data with different levels of noise variance and found that the GP-based CCM demonstrated better convergence, which is critical for distinguishing causation from correlation. Then we applied the GP-based CCM on a segment of real CTG recordings, and the results indicate that uterine activity affects fetal heart rates. The proposed method can readily be adopted for causal discovery in other areas, e.g., neuroscience.

Acknowledgments

6. REFERENCES

[1]. Blackman James A, "The relationship between inadequate oxygenation of the brain at birth and developmental outcome," Topics in Early Childhood Special Education, vol. 9, no. 1, pp. 1–13, 1989.

[2]. Georgieva Antoniya, Abry Patrice, Chudá ek Václav, Djuri Petar M, Frasch Martin G, Kok René, Lear Christopher A, Lemmens Sebastiaan N, Nunes Inês, Papageorghiou Aris T, et al., "Computer-based intrapartum fetal monitoring and beyond: A review of the 2nd workshop on signal processing and monitoring in labor (oct 2017, oxford uk)," Acta Obstetricia et Gynecologica Scandinavica, 2019.

[3]. Ayres-de Campos Diogo, Spong Catherine Y, Chandraharan Edwin, and FIGO Intrapartum Fetal Monitoring Expert Consensus Panel, "FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography," International Journal of Gynecology & Obstetrics, vol. 131, no. 1, pp. 13–24, 2015. [PubMed: 26433401]

[4]. Macones George A, Hankins Gary DV, Spong Catherine Y, Hauth John, and Moore Thomas, "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.

[5]. Ugwumadu A, "Are we (mis) guided by current guidelines on intrapartum fetal heart rate monitoring? case for a more physiological approach to interpretation," BJOG: An International Journal of Obstetrics & Gynaecology, vol. 121, no. 9, pp. 1063–1070, 2014. [PubMed: 24920154]

[6]. Hoyer Patrik O, Janzing Dominik, Mooij Joris M, Peters Jonas, and Scholkopf Bernhard, "Nonlinear causal discovery with additive noise models," in Advances in Neural Information Processing Systems, 2009, pp. 689–696.

[7]. Cheng Patricia W and Buehner Marc J, "Causal learning," The Oxford Handbook of Thinking and Reasoning, pp. 210–233, 2012.

[8]. Feng Guanchao, Quirk J Gerald, and Djurić Petar M, "Inference about causality from cardiotocography signals using Gaussian processes," in ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) IEEE, 2019, pp. 2852–2856.

[9]. Eichler Michael, "Causal inference with multiple time series: principles and problems," Phil. Trans. R. Soc. A, vol. 371, no. 1997, pp. 20110613, 2013. [PubMed: 23858481]

[10]. Tsonis Anastasios A, Deyle Ethan R, Ye Hao, and Sugihara George, "Convergent cross mapping: theory and an example," in Advances in Nonlinear Geosciences, pp. 587–600. Springer, 2018.

[11]. Sugihara George, May Robert, Ye Hao, Hsieh Chih-hao, Deyle Ethan, Fogarty Michael, and Munch Stephan, "Detecting causality in complex ecosystems," Science, vol. 338, no. 6106, pp. 496–500, 2012. [PubMed: 22997134]

[12]. Luo Chuan, Zheng Xiaolong, and Zeng Daniel, "Causal inference in social media using convergent cross mapping," in 2014 IEEE Joint Intelligence and Security Informatics Conference IEEE, 2014, pp. 260–263.

[13]. Schiecke Karin, Pester Britta, Feucht Martha, Leistritz Lutz, and Witte Herbert, "Convergent cross mapping: Basic concept, influence of estimation parameters and practical application," in 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) IEEE, 2015, pp. 7418–7421.

[14]. Mønster Dan, Fusaroli Riccardo, Tylén Kristian, Roepstorff Andreas, and Sherson Jacob F, "Causal inference from noisy time-series data—testing the convergent cross-mapping algorithm in the presence of noise and external influence," Future Generation Computer Systems, vol. 73, pp. 52–62, 2017.

[15]. Ye Hao, Deyle Ethan R, Gilarranz Luis J, and Sugihara George, "Distinguishing time-delayed causal interactions using convergent cross mapping," Scientific reports, vol. 5, pp. 14750, 2015. [PubMed: 26435402]

[16]. Ma Huanfei, Aihara Kazuyuki, and Chen Luonan, "Detecting causality from nonlinear dynamics with short-term time series," Scientific reports, vol. 4, pp. 7464, 2014. [PubMed: 25501646]

[17]. Sletten Julie, Kiserud Torvid, and Kessler Jorg, "Effect of uterine contractions on fetal heart rate in pregnancy: a prospective observational study," Acta obstetricia et gynecologica Scandinavica, vol. 95, no. 10, pp. 1129–1135, 2016. [PubMed: 27460848]

[18]. Takens Floris, "Detecting strange attractors in turbulence," in Dynamical Systems and Turbulence, Warwick 1980, pp. 366–381. Springer, 1981.

[19]. Kennel Matthew B, Brown Reggie, and Abarbanel Henry DI, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," Physical Review A, vol. 45, no. 6, pp. 3403, 1992.

[20]. Fraser Andrew M and Swinney Harry L, "Independent coordinates for strange attractors from mutual information," Physical Review A, vol. 33, no. 2, pp. 1134, 1986.

[21]. Rasmussen Carl Edward, Gaussian Processes for Machine Learning, Citeseer, 2006.

[22]. Feng Guanchao, Quirk J Gerald, and Djurić Petar M, "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.

[23]. Titsias Michalis and Lawrence Neil D, "Bayesian Gaussian process latent variable model," in Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 2010, pp. 844–851.

[24]. Chudáček Václav, Spilka Jiří, Burša Miroslav, Janků Petr, Hruban Lukáš, Huptych Michal, and Lhotská Lenka, "Open access intrapartum CTG database," BMC pregnancy and child-birth, vol. 14, no. 1, pp. 16, 2014.
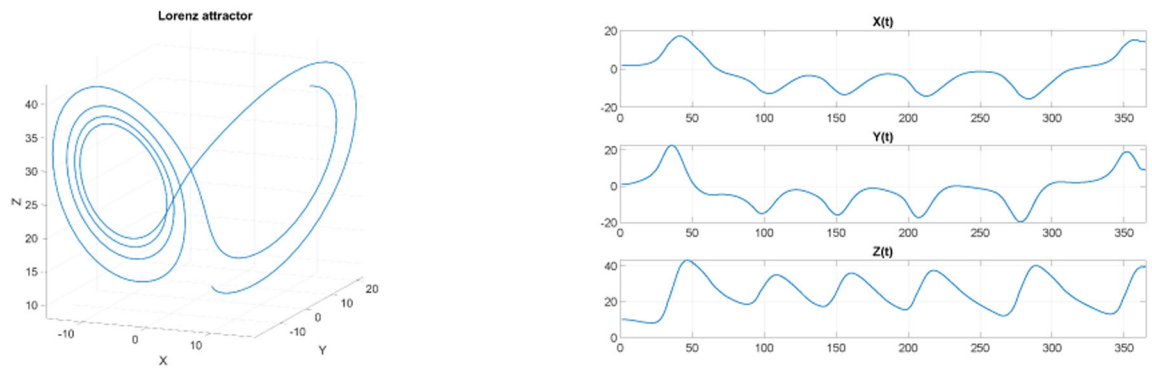
Fig. 1.
Synthetic data, the Lorenz attactor $\mathcal{M}$ (left) along with the three time series obtained by projecting $\mathcal{M}$ to X, Y and Z axis (right).
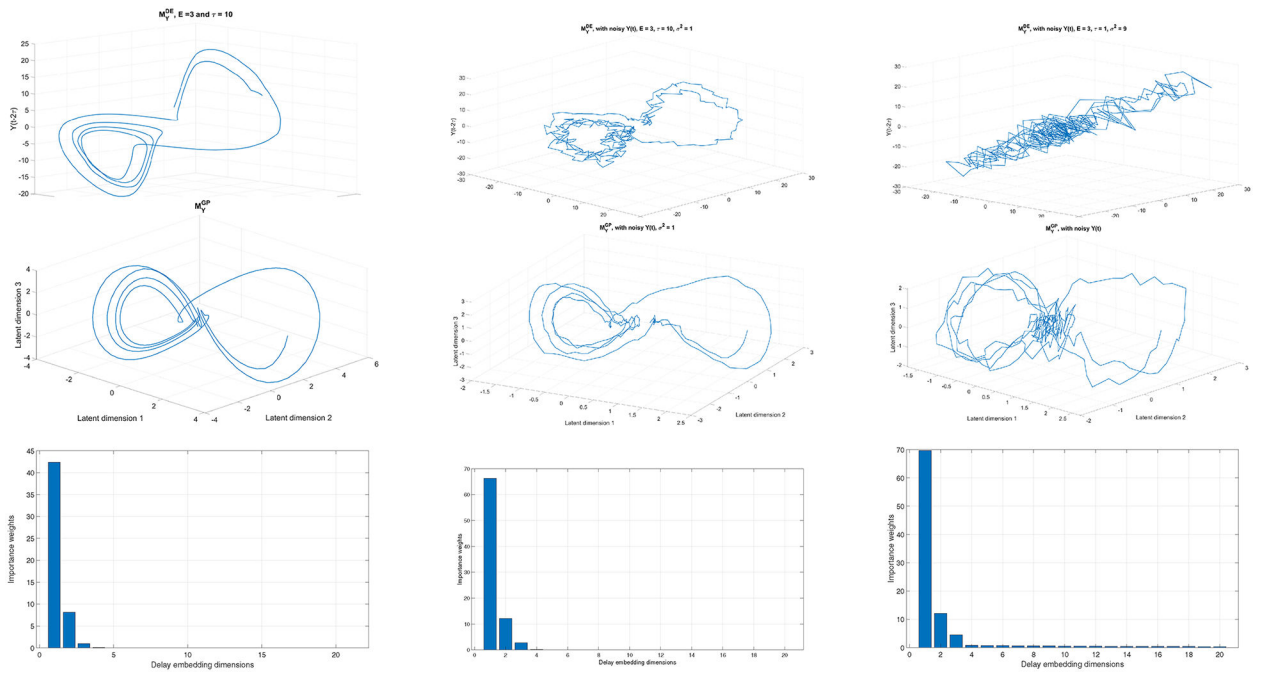
Fig. 2.

Attactor manifold reconstruction results for the synthetic dataset using delay embedding with optimal setting (top), and GP-based method (bottom) along with the importance weight of each dimension in $\mathcal{M}_Y^{\text{init}}$ that reveals the dimensionality of $\mathcal{M}_Y^{\text{GP}}$ (middle). From left to right, the variance of observation noise is $\sigma_e^2 = 0$, 1, and 9, respectively.
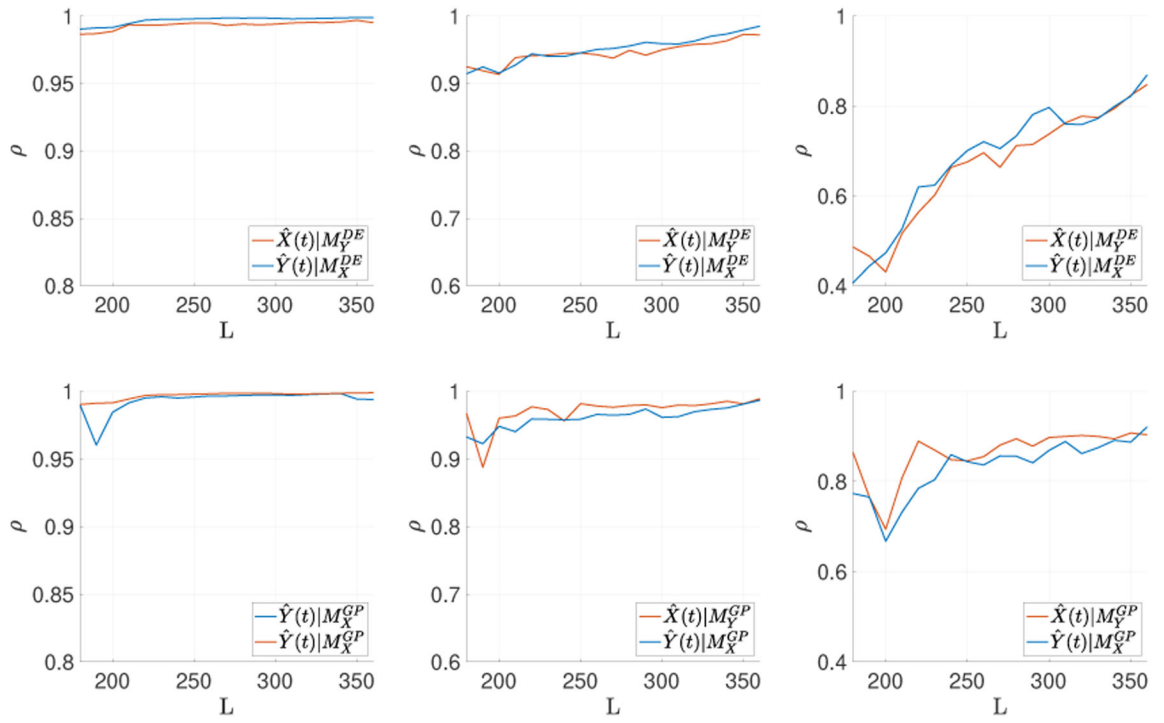
Fig. 3.
The test results on synthetic data using the orginal CCM (top), and the GP-based method (bottom). From left to right, the variance of the observation noise is $\sigma_e^2 = 0$, 1, and 9, respectively. The time series length L is increased from 180 to 360 with a step size of 10.
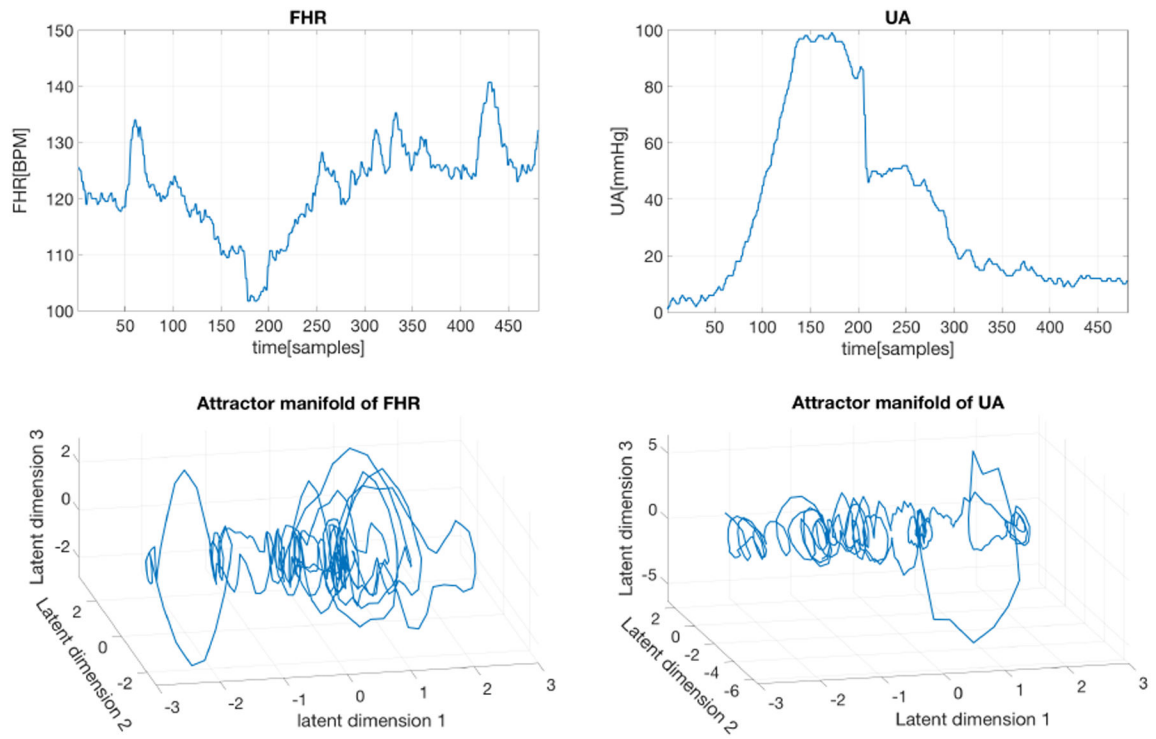
Fig. 4.
Raw FHR and UA segments (top), and their corresponding reconstructed attractor manifolds with the GP-based method (bottom).
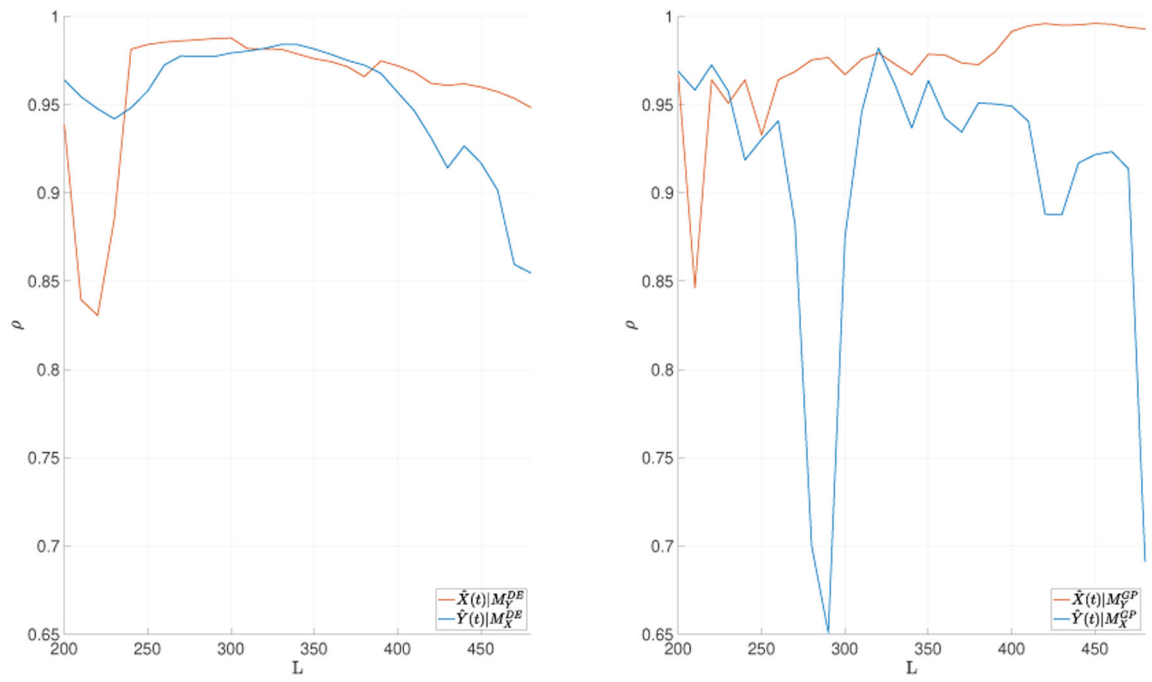
Fig. 5.
The test results on CTG segment using original CCM (left), and the GP-based method (right), where X denotes UA signal, and Y denotes FHR signal. The time series length L is increased from 200 to 480 with a step size of 10.