

INFERENCE ABOUT CAUSALITY FROM CARDIOTOCOGRAPHY SIGNALS USING GAUSSIAN PROCESSES

Guanchao Feng^{*}, J. Gerald Quirk[†], and Petar M. Djurić^{*}

^{*}Department of Electrical and Computer Engineering, Stony Brook University

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

ABSTRACT

In this paper, we propose a novel and simple method for discovery of Granger causality from noisy time series using Gaussian processes. More specifically, we adopt the concept of Granger causality, but instead of using autoregressive models for establishing it, we work with Gaussian processes. We show that information about the Granger causality is encoded in the hyper-parameters of the used Gaussian processes. 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 fetal heart rate, which agrees with recent clinical studies.

Index Terms— Gaussian processes, Granger causality, cardiotocography, fetal heart rate, uterine activity

1. INTRODUCTION

The main goals of science are to understand Nature and, based on this understanding, predict how the world around us evolves. The identification of causal relationships is an important part of scientific research, since it provides us with insights about consequences for actions [1]. The gold standard for identifying causal relationships is using controlled randomized experiments. However, in many situations, these experiments are either impractical, unethical, or simply impossible [2]. The problem of inferring causal interactions from data has challenged scientists and philosophers for centuries [3] and many efforts have been made to solve it. [4].

Causal inference from time series is one important area of research in this domain, where many concepts and methods have been proposed. They include intervention causality [1, 5, 6], structural causality [7] and Granger causality [8, 9] (see [10] for a detailed review). The Granger causality is probably the most prominent and most widely used concept although its usefulness is somewhat controversial [10].

In practice, when detecting Granger causality a (vector) autoregressive (AR) model is often used, and yet in reality, many causal relationships are likely to be nonlinear, this giving rise to doubts about the approach [11]. Further, it has been shown that non-linearity can be helpful in causal discovery [2]. Therefore, instead of using AR models, we adopt the use of Gaussian processes (GPs) for this purpose. They are more powerful for learning functions or mappings and moreover, they can accommodate prior knowledge and assumptions easily. A similar idea was proposed in [12], for testing Granger causality between time series using GPs. Unlike [12], instead of only relying on model evidence for casual discovery, we directly

learn the mappings of possible causes and use cross-validation and model evidence for model selection. Further, we look into the hyper-parameters for causal discovery, which is more natural in the modeling sense and more robust to over-fitting.

In this paper, we are also interested in making inference about causality from cardiotocography (CTG). 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. 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) [13–15]. In fact, the current guidelines for FHR evaluation have been criticized for simplistic interpretation [16]. The classification of CTG tracings by computerized systems remains a challenging problem [17]. For improved understanding of CTG recordings, the interactions between FHR tracings and UA is crucial, and especially establishing if there is causality between them. Interestingly, this issue has been largely overlooked in the literature of computerized analysis of CTG.

In the machine learning literature, Gaussian processes (GPs) provide data-efficient and flexible Bayesian machinery for learning functions or mappings from data. GPs have been successfully applied in both supervised and unsupervised learning tasks [18]. For example, in our previous work [19], we proposed a GP-based method that employs UA signals to recover missing samples of FHR recordings and had excellent results. This work also provided evidence that the UA signals contain information about fetal well-being. Also, there are many applications that exploit the hyper-parameters of GPs for making inference. For example, in [20, 21], the hyper-parameters are used for detecting change points, and in [22], for epilepsy detection from electroencephalograms.

In this paper, we propose a novel and simple method for discovery of Granger causality from noisy time series using GPs. Our approach to finding the possible causes and effects is based on the hyper-parameters of the GPs. Our hypothesis is that information about causality is encoded in the covariances of the GPs, and in particular in the characteristic length scales of the used features by the GPs. We use these length scales to define coefficients that reflect on the direction of causality. We tested the method based on these coefficients on simulated data and then applied it to CTG. The results indicate that the UA is a Granger cause of FHR. Our finding is consistent with recent clinical studies [23].

The paper is organized as follows. In the next section, we provide a brief background on Granger causality and GPs. In Section 3, we present the details of our model. In Section 4, we describe our experiment results. Finally, we conclude with final remarks in

This work has been supported by NSF under Award CCF-1618999.

2. BACKGROUND

2.1. Granger causality

Two principles are fundamental in Granger causality: (i) the effect does not precede its cause in time; (ii) the causal series contains unique information about the series being caused that is not available otherwise. The Granger causality is usually tested using an AR model or a vector autoregressive (VAR) model. For the bivariate case, given $x_1(t)$ and $x_2(t)$, the AR models are

$$\begin{aligned} x_1(t) &= \sum_{j=1}^p a_{11,j} x_1(t-j) + \sum_{j=1}^p a_{12,j} x_2(t-j) + e_1(t), \\ x_2(t) &= \sum_{j=1}^p a_{21,j} x_1(t-j) + \sum_{j=1}^p a_{22,j} x_2(t-j) + e_2(t), \end{aligned} \quad (1)$$

where p is the order of the model, and $e_1(t)$ and $e_2(t)$ are perturbations. If the prediction error of $x_1(t)$ is reduced by using $x_2(t)$ or equivalently, the coefficients $\{a_{12,j}\}_{j=1}^p$ are jointly significantly different from zero, then $x_2(t)$ is a Granger cause of $x_1(t)$. A statistical test is often performed where the null hypothesis is that $\{a_{12,j}\}_{j=1}^p$ are jointly zero. Similarly, a test is performed on $\{a_{21,j}\}_{j=1}^p$.

2.2. Gaussian Processes

By definition, a GP is a collection of random variables such that every finite collection of those random variables has a joint Gaussian distribution. This makes GPs suitable to model distributions over functions, and infinite dimensionality becomes a blessing instead of a curse given their consistency property. To be more specific, if y denotes an output and \mathbf{x} signifies a vector of input variables, and if $y = f(\mathbf{x})$, where $f(\mathbf{x})$ is a real-valued function, then a GP can be seen as the distribution of the function $f(\mathbf{x})$. This GP is completely specified by its mean function $m(\mathbf{x})$ and covariance function $k_f(\mathbf{x}_i, \mathbf{x}_j)$, which are defined by $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$ and $k_f(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{E}[(f(\mathbf{x}_i) - m(\mathbf{x}_i))(f(\mathbf{x}_j) - m(\mathbf{x}_j))]$. In machine learning, a GP is usually assumed to be zero mean, that is, $m(\mathbf{x}) = 0$ for every \mathbf{x} . It is also practical and common to assume the presence of observation noise i.e.,

$$y = y(\mathbf{x}) = f(\mathbf{x}) + \epsilon, \quad (2)$$

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

In the GP regression framework, a prior distribution of the latent function can be directly placed in the function space. With the model and the Gaussian noise assumption, one can obtain tractable posteriors of the latent function and marginal likelihoods, which makes the GPs a powerful Bayesian non-parametric machinery for learning functions or mappings. Although the function $f(\mathbf{x})$ is non-parametric, the covariance function $k_f(\mathbf{x}_i, \mathbf{x}_j)$ is parameterized by its set of hyper-parameters θ .

Let $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ denote the collection of all input vectors, and \mathbf{K}_{ff} the covariance matrix obtained by evaluating the covariance function for \mathbf{X} , i.e., $\mathbf{K}_{ff} = \mathbf{k}_f(\mathbf{X}, \mathbf{X})$. Then the prior probability density function (pdf) of \mathbf{f} given \mathbf{X} is given by

$$p(\mathbf{f}|\mathbf{X}, \theta) = \mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{K}_{ff}). \quad (3)$$

The hyper-parameters θ are learned in the training stage by maximizing the marginal likelihood or model evidence,

$$\begin{aligned} \log p(\mathbf{y}|\mathbf{X}, \theta) &= \log \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}_{ff} + \sigma_\epsilon^2 \mathbf{I}) \\ &= \log \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}) \\ &= -\frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K}| - \frac{N}{2} \log 2\pi. \end{aligned} \quad (4)$$

If we have test inputs \mathbf{X}_* , the mean predictive distribution $p(\mathbf{f}_*|\mathbf{X}_*, \mathbf{X}, \theta)$ will be Gaussian with a mean and covariance given by

$$E(\mathbf{f}_*) = [\mathbf{K}_f(\mathbf{X}_*, \mathbf{X})] \mathbf{K}^{-1} \mathbf{y}, \quad (5)$$

$$\text{cov}(\mathbf{f}_*) = \mathbf{K}_f(\mathbf{X}_*, \mathbf{X}_*) - [\mathbf{K}_f(\mathbf{X}_*, \mathbf{X})] \mathbf{K}^{-1} [\mathbf{K}_f(\mathbf{X}_*, \mathbf{X})]^T. \quad (6)$$

In this work, we use the time variable as an input and we do not specify lengths of histories in the model. In other words, we model the time series simply as functions of time and do not make Markovian assumptions (as in AR processes). From (5), we see that the mean of the predictive distribution is a linear combination of all the previous observations.

2.3. Automatic Relevance Determination

The design of the covariance is of great importance, since it quantifies the distance or similarity between the inputs to the covariance between outputs. Thus, it encodes our prior knowledge or assumptions about the latent function, e.g., smoothness, periodicity, and stationarity. One of the most popular covariance functions is the radial basis function (RBF), which for the 1-D case has the form

$$k_{RBF}(x_i, x_j) = \sigma_f^2 \exp(-\frac{1}{l}(x_i - x_j)^2), \quad (7)$$

where the characteristic length-scale $l > 0$ and the signal variance σ_f^2 are its hyper-parameters. They also have interesting meanings; σ_f^2 represents the variability of the function, and l affects the model complexity in that dimension. If l is small, a small change in the input distance will cause a large change in the covariance of the outputs and vice versa. This leads to another important interpretation of the characteristic length-scale, which is a measure of the importance or relevance of that dimension in the modeling. If l is small, the corresponding dimension is relevant. When \mathbf{x} is a vector, we can compute the $r = \frac{1}{l}$ values for each dimension, where, e.g., each dimension could be a different feature, and then use them for feature selection. This is known as automatic relevance determination (ARD) [18], and is used in supervised learning and automatic dimensionality reduction in unsupervised learning [24–26].

3. MODEL DESCRIPTION

Given two time series x_t and y_t , we would like to determine their Granger causality. For each time series, as shown in Fig. 2, we can model it as a function of time and the history values of the other time series up to a certain length w , similar to the order of an AR model, i.e.,

$$x_t = f_x(t, y_{t-w:t-1}) + \epsilon_{x,t}, \quad (8)$$

$$y_t = f_y(t, x_{t-w:t-1}) + \epsilon_{y,t}, \quad (9)$$

where $\epsilon_{x,t} \sim \mathcal{N}(0, \sigma_x^2)$ and $\epsilon_{y,t} \sim \mathcal{N}(0, \sigma_y^2)$ are independent and additive white Gaussian noises, and the latent functions f_x and f_y are governed by two GPs, respectively, i.e.,

$$\begin{aligned} f_x &\sim \mathcal{GP}(0, k_x([t, y_{t-w:t-1}], [t', y'_{t-w:t-1}])), \\ f_y &\sim \mathcal{GP}(0, k_y([t, x_{t-w:t-1}], [t', x'_{t-w:t-1}])), \end{aligned} \quad (10)$$

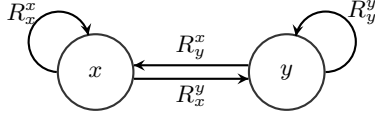


Fig. 1. Illustration of our model, where the arrows stand for flow of information, and the normalized relevance weights indicate importance.

where the covariance functions k_x and k_y have the same form but each has a different set of hyper-parameters, and $[\cdot, \cdot]$ denotes concatenation.

We propose to use the hyper-parameters of the GPs to determine Granger causality. In particular, we use the concept of relevance weights, which we define by $r = 1/l$. We measure the relevance of the history of y_t when modeling x_t with the maximum value of the relevance weight of the input $y_{t-w:t-1}$, i.e., $r_y^x = \max(r_{y_{t-1}}, r_{y_{t-2}}, \dots, r_{y_{t-w}})$, where w is the history length, and $r_{y_{t-k}}$ is the relevance weight of the y sample with lag k . We denote the relevance weight of time, or equivalently x_t , by r_x^x . It represents the overall relevance of the history of x_t when modelling x_t . Finally, we normalize r_x^x and r_y^x by their sum for proper comparison, i.e.,

$$R_x^x = \frac{r_x^x}{r_x^x + r_y^x}, \quad (11)$$

$$R_y^x = \frac{r_y^x}{r_x^x + r_y^x}. \quad (12)$$

Similarly, we can define r_y^y and r_x^y , and their normalized versions R_y^y and R_x^y . The former metric measures the percentage of relevance of the history of y_t on y_t and the latter, the percentage of relevance of a window of history of x_t on y_t . An illustration of our model is shown in Fig 1.

If the true direction of Granger causality is from x_t to y_t , then the relevance of the past values of x_t in modeling of y_t will be greater than the past values of y_t in modeling x_t . The reason is simple: the cause occurs before the effect and the change in cause entails a change in effect, and not the other way around. We note that from a machine learning perspective, the cause and effect are often viewed as a correlation only. Our interpretation of the values of R_x^x , R_y^x , R_x^y , and R_y^y is similar to that of the VAR model, which is that the information of Granger causality is encoded in the relevance weights of the GP models. Therefore, with our approach, if the true Granger causality is from x_t to y_t , this will be indicated as $R_x^x > R_y^x$, and vice versa. Obviously, it is possible that the above analysis of the two time series can show that there is cause and effect in both directions. In principle, the determination of causality would require comparisons of the obtained normalized relevances with thresholds. Setting thresholds, however, is out of the scope of this paper.

If there is no Granger causality between the two time series, the information from the history of the other time-series will not benefit the modeling of the modeled time series. This will be encoded with small values of both R_x^y and R_y^x . We note that since our method is based on GP regression, if we have more prior knowledge about the interaction or relationship between the time series, e.g., in forms of superpositions and/or compositions, this can be included in the used model. In that case, the hyper-parameters will encode even richer information about the possible interactions. However, then the comparisons should be made more carefully, i.e., the relevance weights should also be normalized by the importance of the corresponding explanatory variables of the model.

The above discussion is based on the assumption that we have properly selected the covariance function. If an inappropriate covariance function is used, the modeling and predictive performance will severely be deteriorated, which will most likely lead to unreliable conclusions. The choice of covariance function and window length of history w are model selection problems. Often this choice depends on how much we know about the addressed problem and the nature of the problem itself. One may use cross validation and exploit model evidence and predictive performance to select good covariance functions and window lengths. A tutorial on designing of covariance functions and model selection in GPs can be found in [18]. The window size w should not be very long, since this will increase the number of hyper-parameters in the model. Empirically, for determining Granger causality, we found that the model is robust with different choices of w .

4. EXPERIMENTS AND RESULTS

4.1. Simulations

In this section, we first provide a description of a test for our method. We simulated three pairs of time series (three cases), all shown in Fig. 2. The complexity of relationships between the time series was gradually increased as we moved from the first case to the third case. For each case, we wanted to discover the Granger causality. We used the RBF covariance function with ARD between two observed noisy time series x_t and y_t , each of length $T = 300$. The ground truth in the three simulations was that x_t was a Granger cause of y_t . For Case 2 and Case 3, we included nonlinear mappings and function composition to increase the complexity of the relationships. To remove bias, rather than using deterministic functions, we generated the mappings by GPs.

4.1.1. Case 1: A delayed function with additive noise

This simple simulation represents the case where we have a time series x_t , which represents a sinusoid in noise, and another time series y_t , which is generated by the same, but delayed, sinusoid. The time series y_t also contained additive noise. More precisely, we generated the time series according to the following model:

$$x_t = \sin(t + 0.5) + \epsilon_{x,t}, \quad (13)$$

$$y_t = 0.5 \sin(t) + \epsilon_{y,t}, \quad (14)$$

where t was equally spaced in $[0, 2\pi]$, and where the additive Gaussian noises in x_t and y_t was white, and $\epsilon_{x,t} \sim \mathcal{N}(0, \sigma_x^2)$ and $\epsilon_{y,t} \sim \mathcal{N}(0, \sigma_y^2)$, respectively, with $\sigma_x^2 = 1$ and $\sigma_y^2 = 4$. The noises $\epsilon_{x,t}$ and $\epsilon_{y,t}$ were independent of each other.

4.1.2. Case 2: A noisy function with a nonlinear mapping

With the second simulation we generated more complicated relationship between the time series x_t and y_t . The time series x_t was a superposition of two deterministic signals in additive Gaussian noise. The time series y_t was composed of two parts, a function sampled from a GP that used x_t as an input and a deterministic function of time. This time series also had zero mean additive Gaussian noise which was independent from the noise of x_t . The exact generative model of the data is given by

$$x_t = \sin(t/10) + (t/50)^2 + \epsilon_{x,t}, \quad (15)$$

$$y_t = f(x) + \text{sinc}((t/40)^2) + \epsilon_{y,t}, \quad (16)$$

$$f(x) \sim \mathcal{GP}(0, k(x, x')), \quad (17)$$

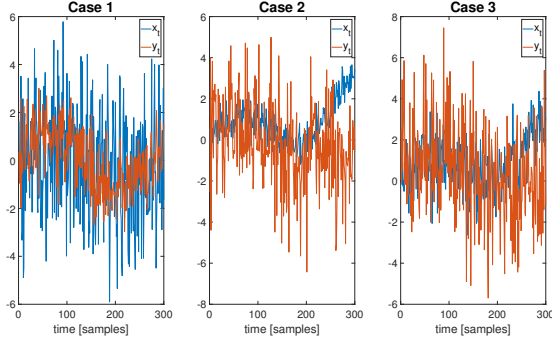


Fig. 2. Simulated time series pairs in each simulation cases.

where the noise variances were $\sigma_x^2 = 0.25$ and $\sigma_y^2 = 4$.

4.1.3. Case 3: A delayed noisy function with two layers of function composition

In this simulation, we further complicated the previous case by introducing another layer of function composition and delay. We first generated a time series x_t as in Case 2, and we used its delayed version as the input to a GP with an RBF covariance function. A function was then sampled from this GP and used as the input to another GP. This generative model is described by

$$x_t = \sin(t/10) + (t/50)^2 + \epsilon_{x,t}, \quad (18)$$

$$y_t = f_2(f_1(x_{t-2.5})) + \text{sinc}((t/40)^2) + \epsilon_{y,t}, \quad (19)$$

$$f_1(x) \sim \mathcal{GP}(0, k_1(x, x')), \quad (20)$$

$$f_2(x) \sim \mathcal{GP}(0, k_2(x, x')), \quad (21)$$

where k_1 and k_2 are both RBF covariance functions but with different sets of hyper-parameters. The noises, as before, were independent and zero-mean Gaussian with variances $\sigma_x^2 = 1$ and $\sigma_y^2 = 4$.

4.1.4. Results

The results of our method are summarized in Table. 1. The method correctly determined the causality in all three cases. We repeated the experiment five times, and correct decisions were made in each of them.

Table 1. Simulation Results

	R_x^x	R_y^x	R_y^y	R_x^y	Granger causality
Case 1	0.9931	0.0069	0.1566	0.8434	x_t Granger causes y_t
Case 2	0.7914	0.2086	0.0008	0.9992	x_t Granger causes y_t
Case 3	0.4029	0.5971	0.3986	0.6014	x_t Granger causes y_t

4.2. Real data: CTG segment

In our experiments with real CTG data, we used 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 [27].

We applied the method on a real CTG segment of length 491 samples, as shown in Fig. 3, which corresponds to a duration of 2.04 minutes (the sampling rate for both FHR and UA signals was 4 Hz).

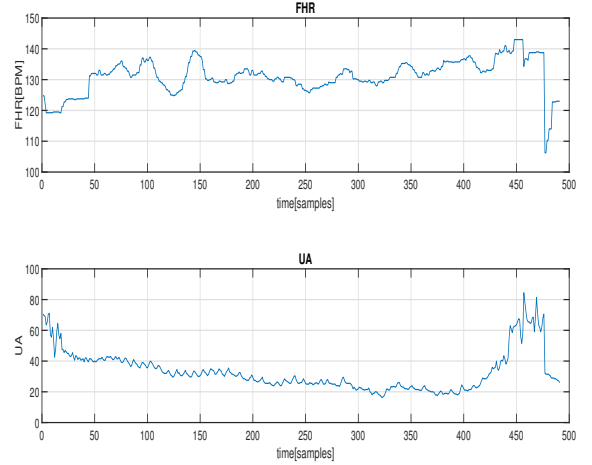


Fig. 3. A segment of not-preprocessed (raw) FHR tracing and the corresponding UA signals.

Table 2. Granger causality in CTG

	R_{FHR}^{FHR}	R_{UA}^{FHR}	R_{UA}^{UA}	R_{FHR}^{UA}
CTG signal	0.5090	0.4910	0.7917	0.2083

We used the RBF covariance function with a window size of $w = 4$. The results are shown in Table 2. The values of the normalized relevance weights indicate that the UA is a Granger cause of FHR. This agrees with a recent clinical study [23].

Another interpretation of the second principle of Granger causality is that if the UA Granger-causes the FHR, then past values of UA should contain information that helps in predicting future values of the FHR. With this information, the predictive performance of the method should be better than that of using information from past values of FHR alone. This is consistent with our observation in [19]. There we showed that adopting information from UA signals helps in recovering missing samples of FHR tracings.

We also used deep Gaussian processes [28] to see if FHR and UA signals have a common manifold. The results indicate that they cannot be generated from a common manifold, i.e., the FHR and UA encode different information about fetal well-being. This is also consistent with our observations in [29], where the performance of classification of FHR can be improved using features from UA.

5. CONCLUSIONS

In this paper, we proposed a Gaussian processes-based method for detecting Granger causality. We showed that the interaction or causal information can be extracted from the hyperparameters of the Gaussian processes. We tested our method on both simulated data and real CTG recordings and found the results very promising. Although we used bivariate time series, this methodology can be easily extended to more than two time series. Furthermore, if we have additional prior knowledge about the time series and their interactions, it can be easily injected in the GP framework. After applying our method on real CTG recording, we found that uterine activity is a Granger cause of fetal heart rate, which agrees with recent clinical studies.

6. REFERENCES

- [1] M. Eichler, "Causal inference in time series analysis," *Causality: Statistical perspectives and applications*, pp. 327–354, 2012.
- [2] P. O. Hoyer, D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf, "Nonlinear causal discovery with additive noise models," in *Advances in neural information processing systems*, 2009, pp. 689–696.
- [3] L. Barnett, A. B. Barrett, and A. K. Seth, "Granger causality and transfer entropy are equivalent for Gaussian variables," *Physical review letters*, vol. 103, no. 23, p. 238701, 2009.
- [4] P. Spirtes and K. Zhang, "Causal discovery and inference: concepts and recent methodological advances," in *Applied informatics*, vol. 3, no. 1. SpringerOpen, 2016, p. 3.
- [5] M. Eichler and V. Didelez, "On Granger causality and the effect of interventions in time series," *Lifetime data analysis*, vol. 16, no. 1, pp. 3–32, 2010.
- [6] C. Berzuini, P. Dawid, and L. Bernardinell, *Causality: Statistical perspectives and applications*. John Wiley & Sons, 2012.
- [7] H. White and X. Lu, "Granger causality and dynamic structural systems," *Journal of Financial Econometrics*, vol. 8, no. 2, pp. 193–243, 2010.
- [8] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: Journal of the Econometric Society*, pp. 424–438, 1969.
- [9] —, "Testing for causality: a personal viewpoint," *Journal of Economic Dynamics and control*, vol. 2, pp. 329–352, 1980.
- [10] M. Eichler, "Causal inference with multiple time series: principles and problems," *Phil. Trans. R. Soc. A*, vol. 371, no. 1997, p. 20110613, 2013.
- [11] Y. Chikahara and A. Fujino, "Causal inference in time series via supervised learning," in *IJCAI*, 2018, pp. 2042–2048.
- [12] P.-O. Amblard, O. J. Michel, C. Richard, and P. Honeine, "A Gaussian process regression approach for testing Granger causality between time series data," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3357–3360.
- [13] D. Ayres-de Campos, C. Y. Spong, E. Chandrachan, 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>
- [14] 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.
- [15] S. Dash, J. G. Quirk, and P. M. Djurić, "Learning dependencies among fetal heart rate features using Bayesian networks," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 6204–6207.
- [16] A. Ugwumadu, "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.
- [17] P. Steer, "Commentary on antenatal cardiotocogram quality and interpretation using computers," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 121, no. s7, pp. 9–13, 2014.
- [18] C. E. Rasmussen, *Gaussian Processes for Machine Learning*. Citeseer, 2006.
- [19] 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.
- [20] S. Roberts, M. Osborne, M. Ebdon, S. Reece, N. Gibson, and S. Aigrain, "Gaussian processes for time-series modelling," *Phil. Trans. R. Soc. A*, vol. 371, no. 1984, p. 20110550, 2013.
- [21] Y. Saatçi, R. D. Turner, and C. E. Rasmussen, "Gaussian process change point models," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. Citeseer, 2010, pp. 927–934.
- [22] S. Faul, G. Gregoric, G. Boylan, W. Marnane, G. Lightbody, and S. Connolly, "Gaussian process modeling of EEG for the detection of neonatal seizures," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 12, pp. 2151–2162, 2007.
- [23] J. Sletten, T. Kiserud, and J. Kessler, "Effect of uterine contractions on fetal heart rate in pregnancy: a prospective observational study," *Acta obstetrica et gynecologica Scandinavica*, vol. 95, no. 10, pp. 1129–1135, 2016.
- [24] N. D. Lawrence, "Gaussian process latent variable models for visualisation of high dimensional data," in *Advances in neural information processing systems*, 2004, pp. 329–336.
- [25] A. Damianou, C. Ek, M. Titsias, and N. Lawrence, "Manifold relevance determination," *arXiv preprint arXiv:1206.4610*, 2012.
- [26] 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.
- [27] V. Chudáček, J. Spilka, M. Burša, P. Janků, L. Hruban, M. Hupčych, and L. Lhotská, "Open access intrapartum CTG database," *BMC pregnancy and childbirth*, vol. 14, no. 1, p. 16, 2014.
- [28] A. Damianou and N. Lawrence, "Deep Gaussian processes," in *Artificial Intelligence and Statistics*, 2013, pp. 207–215.
- [29] G. Feng, J. G. Quirk, and P. M. Djurić, "Supervised and unsupervised learning of fetal heart rate tracings with deep Gaussian processes," in *14th Symposium on Neural Networks and Applications*, 2018.