Spectral discontinuity design: Interrupted time series with spectral mixture kernels

David Leeftink H.LEEFTINK@STUDENT.RU.NL and **Max Hinne** M.HINNE@DONDERS.RU.NL
Radboud University, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands

Editors: Emily Alsentzer[⊗], Matthew B. A. McDermott[⊗], Fabian Falck, Suproteem K. Sarkar, Subhrajit Roy[‡], Stephanie L. Hyland[‡]

Abstract

Quasi-experimental designs allow researchers to determine the effect of a treatment, even when randomized controlled trials are infeasible. A prominent example is interrupted time series (ITS) design, in which the effect of an intervention is determined by comparing the extrapolation of a model trained on data acquired up to moment of intervention, with the interpolation by a model trained on data up to the intervention. Typical approaches for ITS use (segmented) linear regression, and consequently ignore many of the spectral features of time series data. In this paper, we propose a Bayesian nonparametric approach to ITS, that uses Gaussian process regression and the spectral mixture kernel. This approach can capture more structure of the time series than traditional methods like linear regression or AR(I)MA models, which improves the extrapolation performance, and hence the accuracy of causal inference. We demonstrate our approach in simulations, and use it to determine the causal effect of Kundalini yoga meditation on heart rate oscillations. We show that our approach is able to detect the causal effect of interventions that alter the spectral features of these time series.

Keywords: Interrupted time series; Gaussian process regression; spectral mixture kernel; causal inference

1. Introduction

Much of scientific research is concerned with causal questions. A prime example is the establishment of a causal effect of a medical treatment, which is typically determined in randomized controlled trial (RCT) study designs. Here, the effect of confounding variables that may influence the outcome cancel out, so that the causal effect can (in principle) readily be determined (Pearl and Mackenzie, 2018). When the RCT is not possible, due to pragmatic or ethical considerations, *quasi-experimental designs* (QEDs) may provide an alternative. In these study designs randomness is replaced with additional constraints, that under certain assumptions approximately recreate the conditions of the RCT. Quasi-experimental designs have been used in many scientific domains, such as medicine (Penfold and Zhang, 2013), neuroscience (Marinescu et al., 2018), or epidemiology (Bhaskaran et al., 2013).

In this paper, we focus on a quasi-experimental design that is tailored to time series data: interrupted time series (ITS) design (McDowall et al., 1980; Hausman and Rapson, 2018). Here, the observations are split into two subsets: one containing all observations up to an intervention, and one containing all observations from the intervention on. ITS is especially relevant for medical applications, as time series are

prevalent in health care, and RCT may be prohibitive due to ethical, financial, or urgency constraints. For instance, one might use ITS to study the effect of a drug on heart rate measured via electrocariography (ECG) (Matowe et al., 2003), or the effect of public health care policies (Bernal et al., 2017; Vokó and Pitter, 2020). The latter is especially relevant in the COVID-19 pandemic, where this study design is used frequently (Vokó and Pitter, 2020; Figueiredo et al., 2020; Hamadani et al., 2020). ITS establishes a causal effect by comparing the extrapolation based on the pre-intervention data, based on the interpolation based on the post-intervention data. The majority of ITS designs model the observations pre- and post threshold using (segmented) polynomial regression (Jandoc et al., 2015), but these parametric models ignore many of the temporal properties of these data sets (Hausman and Rapson, 2018; Nason, 2013). This is an fundamental shortcoming, as the ITS causal effect estimation relies on having an adequate model for the extrapolation of the pre-intervention time series. If this extrapolation is poor, due to model misspecification, the findings of the ITS study become questionable.

We propose a novel method for ITS designs that replaces traditional parametric models with Gaussian process (GP) regression (Rasmussen and Williams, 2006), together with a spectral mixture kernel. This results in a flexible Bayesian nonparametric ITS design. It extends recently proposed work that applies a similar model to regression discontinuity (RD) design (Hinne et al., 2020; Branson et al., 2019). Our main contribution is that we use a spectral mixture kernel (Wilson and Adams, 2013) as the covariance function of the GP. This kernel allows for the identification of more temporal structure in the data than other covariance functions (Schulz et al., 2017), which improves extrapolation quality, and hence the estimation of the causal effect of the intervention.

The paper is structured as follows. In section 2, we describe the Bayesian model comparison approach for ITS that we use in our method. In section 3, we recap Gaussian process regression (GPR) and describe the spectral mixture kernel. In section 4, we discuss a simulation, followed by an applications on heart rate oscillations of people trying Kundalini yoga meditation in section 5. We provide a discussion of potential extensions of the model in section 6 and conclude in section 7.

2. Causal inference using interrupted time series

We provide a brief introduction of the background of causal inference using ITS designs. For a more in-depth discussion, we refer to e.g. Bernal et al. (2017); McDowall et al. (1980).

The detection of a causal effect is naturally formulated using the potential outcomes framework by Rubin (1974). Consider an observation i with independent variable $x_i \in \mathbb{R}^P$. Throughout this paper we consider only the univariate case and assume $P = 1$, but the extension to multidimensional input is straightforward. In addition, we observe an indicator variable z_i. Here, $z_i = 1$ indicates the intervention has been applied, and $z_i = 0$ indicates it has not. The outcome depends on treatment, that is

$$y_i = \begin{cases} y_i(0) & \text{if } z_i = 0, \\ y_i(1) & \text{otherwise.} \end{cases} \tag{1}$$

The individual causal effect is defined as the difference between these two potential outcomes, that is

$$d_i = y_i(1) - y_i(0) \ . \tag{2}$$

Instead of the individual causal effect, we often estimate the *average causal effect* (ACE)

instead, defined by the differences in the expectations of the population:

$$d_{\text{ACE}} = \mathbb{E}[y(1)] - \mathbb{E}[y(0)] \ . \tag{3}$$

In the randomized controlled trial, as individuals are randomly assigned to the intervention or control group, all differences other than treatment are integrated out in these expectations (Bloom, 2012). In ITS designs however, the allocation to intervention or control group is based on a threshold (O'Keeffe and Baio, 2016):

$$z_i = \begin{cases} 1 & \text{if } x_i \geq x_0 \text{ and} \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where x_0 is the threshold that defines at which point the intervention is applied. Here, the causal estimand can be obtained as the difference between observed post-intervention regression function, $f_1^{(I)}$, and the *extrapolated* pre-intervention regression function, $f_1^{(C)}$ (Kim and Steiner, 2016). This difference does not need to be constant, so the causal estimand d_{ITS} is itself a function of $x \geq x_0$, given by (Kim and Steiner, 2016)

$$d_{\text{ITS}}(x) = f_1^{(I)}(x) - f_1^{(C)}(x)\,, \quad x \geq x_0 \ . \tag{5}$$

2.1. Causal inference as model comparison

Most ITS implementations determine the presence of a causal effect by assuming a parametric form for $f_1^{(C)}$ and $f_1^{(I)}$, such as linear regression, and subsequently testing for $\beta \neq 0$, with β either the intercept or the slope of the regression. This has a few downsides. First, this approach allows one to reject the null hypothesis of no effect, but not to find evidence in favor of the null. Second, the effect size estimate is provided as a point estimate, ignoring the associated uncertainty. Both issues can be remedied by using Bayesian model comparison instead (Wagenmakers, 2007), which proceeds as follows.

We define two models, $\mathcal{M}_0$ and $\mathcal{M}_1$, that correspond to the null model and the alternative model, respectively. In $\mathcal{M}_0$, all observations $D = \{(x_i, y_i)\}_{i=1}^n$ are modelled by the same regression function f_0, so that

$$y_i \sim \mathcal{N}\left(f_0(x_i), \sigma_n^2\right) \ , \tag{6}$$

with observation noise σ_0^2. In $\mathcal{M}_1$, separate regression functions are used pre- and post intervention instead, resulting in

$$y_i \sim \begin{cases} \mathcal{N}\left(f_1^{(C)}(x_i), \sigma_C^2\right) & \text{if } x_i < x_0 \ , \\ \mathcal{N}\left(f_1^{(I)}(x_i), \sigma_I^2\right) & \text{if } x_i \geq x_0 \ , \end{cases} \tag{7}$$

where $f_1^{(C)}$ and $f_1^{(I)}$ correspond to the latent functions pre- and post intervention, respectively, and analogously σ_C^2 and σ_I^2 represent the corresponding observation noise levels.

The two models are compared via the Bayes factor, the ratio of marginal likelihoods

$$\begin{aligned} \text{BF}_{10} &= \frac{p(D \mid \mathcal{M}_1)}{p(D \mid \mathcal{M}_0)} \\ &= \frac{p(D^{(C)} \mid \mathcal{M}_1) p(D^{(I)} \mid \mathcal{M}_1)}{p(D \mid \mathcal{M}_0)} \ , \end{aligned} \tag{8}$$

where $D^{(C)} = \{(x_i, y_i) \mid x_i < x_0\}$ and $D^{(I)} = \{(x_i, y_i) \mid x_i \geq x_0\}$. This step follows because the observations prior to the intervention are independent of the observations after the intervention (conditioned on $\mathcal{M}_1$).

3. Gaussian process regression

The latent functions $F = \left(f_0, f_1^{(C)}, f_1^{(I)}\right)$ form the basis of our model comparison approach. We follow Branson et al. (2019); Hinne et al. (2020) and place a Gaussian process (GP) prior on each function $f \in F$, so that we have

$$f \sim \mathcal{GP}\left(\mu(x;\theta), k(x, x';\theta)\right) \ , \tag{9}$$

where $\mu(x;\theta)$ and $k(x,x';\theta)$ are the mean and covariance function of the GP, respectively, and θ is a set of hyperparameters that we will discuss later. We will drop the dependency on the hyperparameter θ from the notation when no confusion is likely to arise. The first GP parameter $\mu(x)$ is the expected mean of the function f, and is often assumed to be zero for standardized data. The second parameter $k(x,x')$ is known as the covariance function or *kernel*, and indicates how similar the expected responses $f(x)$ and $f(x')$ are, depending on the locations x, x', and the hyperparameters θ. By choosing the appropriate covariance function k, widely different latent processes can be modelled (Rasmussen and Williams, 2006). For instance, with a linear covariance function, we essentially recover linear regression again, while the commonly used squared-exponential kernel results in smooth functions that are infinitely differentiable.

The GP posterior predictive distribution is available in closed form (MacKay, 2002; Rasmussen and Williams, 2006). This makes it straightforward to estimate d_{ITS} by taking the difference in the predictive distributions of $f_1^{(C)}(x)$ and $f_1^{(I)}(x)$, evaluated on $x \geq x_0$. Predicting in the GP framework proceeds as follows. The matrix K has its entries defined as $K_{ij} = k(x_i, x_j)$ and contains all (cross-)covariances. For a single new point x_* to predict, we write $\boldsymbol{k}_* = (k(x_1, x_*), \ldots, k(x_n, x_*))$ to reflect the covariance of the new response with all training responses. The latter is denoted as $\boldsymbol{y} = (y_1, \ldots, y_n)$. This allows us to write the posterior predictive mean and variance at location x_* using the following key equations (Rasmussen and Williams, 2006):

$$\mathbb{E}[f_*] = \boldsymbol{k}_*^T (K + \sigma_n^2 I)^{-1} \boldsymbol{y}, \tag{10}$$

$$\mathbb{V}[f_*] = k(x_*, x_*) - \boldsymbol{k}_*^T (K + \sigma_n^2 I)^{-1} \boldsymbol{k}_* \ , \tag{11}$$

where I is the $n \times n$ identity matrix. From these equations, it follows that the predictive distribution for a particular test point is itself Gaussian, with mean $\mathbb{E}[f_*]$ and variance $\mathbb{V}[f_*]$ (Bishop, 2006).

The hyperparameters θ of the covariance function are typically unknown and are usually optimized by maximizing the log marginal likelihood $\log[p(\boldsymbol{y} \mid D, \theta)]$. This term is also available in closed form, provided the observation model in (6) (and (7)) is Gaussian (Bishop, 2006; Rasmussen and Williams, 2006):

$$\begin{aligned} \log p(\boldsymbol{y} \mid D, \theta) = &-\frac{1}{2} \boldsymbol{y}^T (K + \sigma_{\text{obs}}^2 I)^{-1} \boldsymbol{y} \\ &- \frac{1}{2} \log |K + \sigma_n^2 I| \\ &- \frac{n}{2} \log 2\pi \ . \end{aligned} \tag{12}$$

3.1. Spectral mixture kernel

In (Branson et al., 2019) and (Hinne et al., 2020), QEDs are implemented using either the exponential, the squared-exponential, or the Matérn covariance functions. Both of these approaches are more flexible than classical polynomial regression or ARMA models, but they still capture few of the interesting characteristics of time series data. Here, we discuss the *spectral mixture kernel* introduced by Wilson and Adams (2013), which we propose to use in the ITS context to address this shortcoming.

An important property of kernels is *stationarity*. A kernel is stationary if the covariance between two responses $f(x)$ and $f(x')$ only depends on the distance $\tau = |x - x'|$ between the inputs, so that it is invariant to translations of the inputs (Wilson, 2014; Rasmussen and Williams, 2006; Duvenaud, 2014). According to Bochner's theorem (Bochner, 1959), any stationary kernel has a corresponding spectral density $S(\omega)$ which is its Fourier dual (Wilson and Adams,

2013), that is

$$k(\tau) = \int S(\omega) e^{2\pi i \omega^T \tau} \, \mathrm{d}\omega$$
$$S(\omega) = \int k(\tau) e^{-2\pi i \omega^T \tau} \, \mathrm{d}\tau \quad . \tag{13}$$

Spectral densities for many popular kernels can be obtained by plugging the kernel into equation (13). One can also work the other way around however, starting with a spectral density and transforming it into a time-domain covariance function. This provides a powerful method for constructing new kernels, where one starts by modeling a spectral density, and subsequently converts this to a time-domain covariance function. This can lead to highly expressive covariance functions that are still interpretable in their spectral representation.

Consider the following spectral density $S(\omega)$, modelled as a univariate Gaussian distribution

$$\begin{aligned} S(\omega) &= \mathcal{N}\left(\omega; \mu, \sigma^2\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\omega-\mu)^2}{2\sigma^2}\right) \quad . \end{aligned} \tag{14}$$

Via (13), $S(\omega)$ is transformed into a stationary covariance function in the temporal domain, which then has the following form:

$$k(\tau) = \exp\left(-2\pi^2\tau^2\sigma^2\right)\cos\left(2\pi\tau\mu\right) \quad . \tag{15}$$

The flexibility of the covariance function is greatly increased by instead defining $S(\omega)$ as a Gaussian mixture model (Bishop, 2006) with Q components. This results in

$$S(\omega) = \sum_{q=1}^{Q} w_q \mathcal{N}\left(\omega; \mu_q, \sigma_q^2\right) \quad , \tag{16}$$

with μ_q and σ_q^2 the mean and variance of mixture component q, and w_q is weight. By once more transforming this spectral representation into a covariance function via (13), we obtain the spectral mixture kernel (Wilson and Adams, 2013)

$$k(\tau) = \sum_{q=1}^{Q} w_q \cos\left(2\pi\tau\mu_q\right) \exp\left(-2\pi^2\tau^2\sigma_q^2\right) \quad . \tag{17}$$

This covariance function has hyperparameters $\theta = (Q, \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\sigma})$, corresponding to the weight, mean, and covariance of each Gaussian component. These have a clear intuitive interpretation: μ_q indicates the frequency of component q, the inverse of the variance $1/\sigma_q$ can be interpreted as the length scale of the component, that reflects how quickly that frequency contribution changes with the input x, and lastly the weights w_q indicate the relative contribution of each mixture component (Wilson and Adams, 2013)

3.2. Model estimation

To compute the Bayes factor that quantifies whether there is a causal effect (using (8)), we require the model marginal likelihoods

$$p(D \mid \mathcal{M}_m) = \int p(D \mid \mathcal{M}_m, \theta) p(\theta \mid \mathcal{M}_m) \, \mathrm{d}\theta \quad , \tag{18}$$

where $m \in \{0, 1\}$. Unfortunately, the model marginal likelihood is not analytically available, so we instead adopt the common approach in GPR and approximate this quantity with the Bayesian information criterion (BIC) instead (Schwarz, 1978; Kim and Teh, 2018):

$$\log p(D \mid \mathcal{M}_m) \approx \log p(D \mid \mathcal{M}_m, \hat{\theta}) - \frac{k}{2} \log n \quad , \tag{19}$$

where $\hat{\theta}$ are the hyperparameters optimized via (12), and k are the number of hyperparameters of the model.

The hyperparameters $\boldsymbol{w}$, $\boldsymbol{\mu}$, and $\boldsymbol{\sigma}$ are obtained by optimizing the log marginal likelihood (12) summed with their respective log prior density. The number of mixture components Q is optimized by selecting the

value of Q that maximizes (12), out of a range of components. The initial values of the hyperparameters are determined by fitting a Gaussian mixture model with Q components to the empirical spectrum obtained with the Lomb–Scargle periodogram, a commonly used method for detecting periodic features in unequally sampled data (Vanderplas, 2018). The parameters of the mixture model are then used to initialize the parameters of the spectral mixture kernel.

Subsequently, we compute the causal effect size given $\mathcal{M}_1$ by evaluating the posterior predictive expectations of $f_1^{(I)}$ and $f_1^{(C)}$ over a range of values $x \geq x_0$. We then summarize this effect size function by its mean and its largest absolute value.

After a causal effect has been determined via (8) and quantified via (5), inspecting the optimized hyperparameters of $f_1^{(I)}$ and $f_1^{(C)}$ allows for a further interpretation of the detected effect. For instance, a particular frequency component may have received a lower mixture weight after intervention, which indicates which property of the time series was most affected by the change.

Our pipeline is implemented using GPflow 2.0 (Matthews et al., 2017), and is available at https://github.com/DavidLeeftink/Spectral-Discontinuity-Design.

4. Simulation study

Arguably one of the most straightforward discontinuities in periodic features of a signal is a change in frequency. Here, we simulate data from f_{sim} defined as a sum of two cosines, with frequencies 12Hz and 25Hz (see Appendix A). For each cosine, from x_0 onward, the frequency is increased by $\alpha = \{0, \ldots, 8\}$. We apply our analysis to these generated data, using Gamma(1.0,1.0) priors for $\boldsymbol{w}$ and Gamma(8.0, 2.0) priors for $\boldsymbol{\sigma}$, and improper uniform priors on the means $\boldsymbol{\mu}$. Results are averaged over 20 runs.

An example data set and ITS analysis using $\alpha = 4$ are shown in 1A. The model correctly recovers the true frequencies underlying the data, as well as the decreased amplitude of the second harmonic component, and finds barely worth mentioning evidence in favor of an effect ($\log \text{BF}_{10} = 0.15$) (Wagenmakers, 2007). The estimated spectral mixture of the continuous model is centered between the true frequencies of the control and intervention group. This faithfully represents the continuous hypothesis that the observations can be explained without any changes in spectral content. As the discontinuity size grows larger, the standard deviation of the components of the continuous model increases as well, since it has to account for a larger difference. The discontinuous model instead correctly identifies the true mixture components.

We quantify the performance of our GPR ITS approach using the RMSE between the true effect d_{ITS}, and the estimates obtained by taking samples from the posterior for $f_1^{(C)}$ and $f_1^{(I)}$, as well as the estimate obtained using an ARMA model (Prado and West, 2010). The ARMA parameters are determined using a grid search and evaluating the model with its BIC score. The result is shown in 1. The figure clearly shows the consistent lower RMSE for the GPR approach. In addition, Table 1 shows the log Bayes factors for the GPR analysis and the baseline. Note that the GPR model detects an effect after an increase of $\alpha = 4$Hz.

5. Heart rate oscillations during meditation

Peng et al. (1999) studied the effect of Kundalini Yoga mediation techniques on heart rate oscillations. The authors hypothesize that such mediation results in quiescent heart rate oscillations. However, they find the opposite effect; the meditation seems to

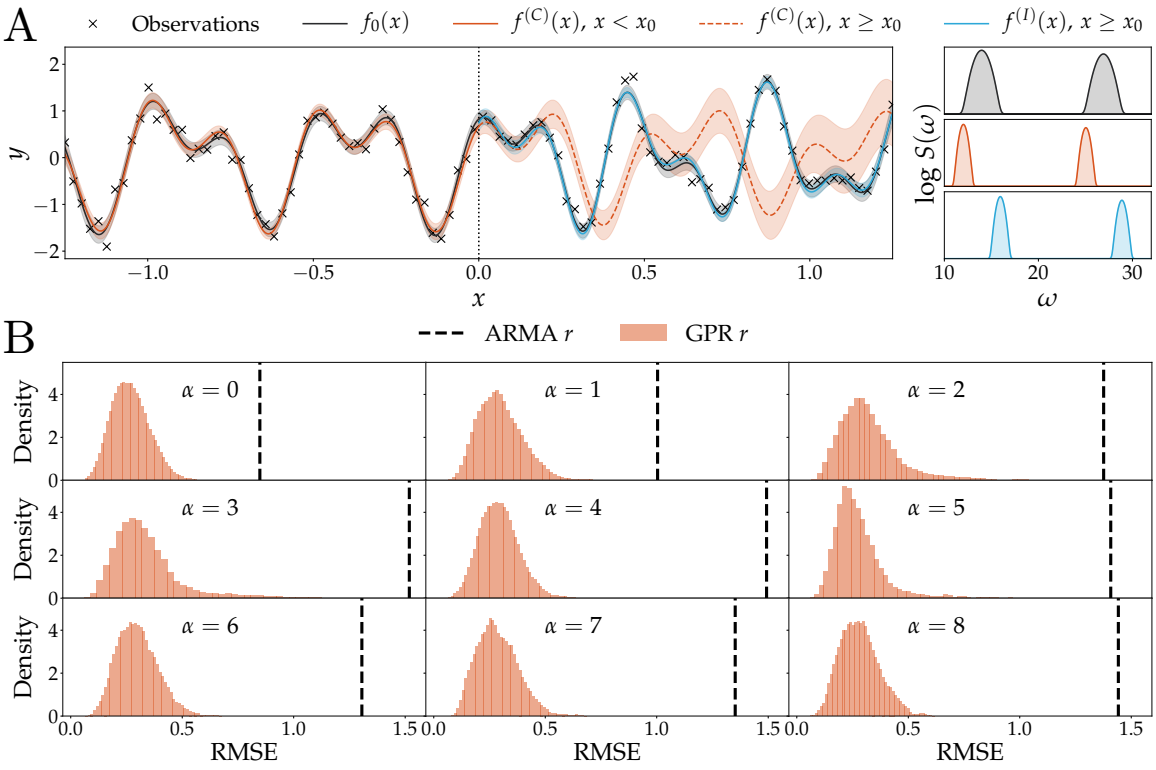

Figure 1: Simulation study. A. Model fit and extrapolation of $\mathcal{M}_0$ and $\mathcal{M}_1$. The data were generated with a frequency shift of $\alpha = 4$. We find $\log \text{BF}_{10} = 0.15$. The shaded interval represents two standard deviations around the mean. The colors of the power density spectrum correspond to the legend of the regression. B. The RMSE between the estimated and true d_{ITS} (see Eq. (5)), using posterior samples of our approach and an ARMA baseline. Due to the extrapolation capabilities of the spectral mixture kernel, the RMSE is consistently lower for the GPR model than for the ARMA model.

induce higher frequency oscillations instead. The experiment is well-suited for ITS design, but is difficult to perform with standard techniques such as the fast Fourier transform or the periodogram, as the data is not uniformly sampled. Gaussian process regression does not have this prerequisite, and is applicable here.

We apply our methodology to the same data set to see if we can determine a causal effect of the meditation on the measured heart rate. The data is obtained from the PhysioNet database and consists of heart rates of two women and two men, of ages 20–52 (mean 33) (Goldberger et al., 2000). To ensure the model comparison reflects a discontinuity in spectral density rather than mean value, the mean function $\mu(x)$ of the continuous model $\mathcal{M}_0$ is extended with a changepoint kernel (Saatçi et al., 2010), which transitions between two constant kernels at the intervention point. Similarly, the mean functions of both submodels in $\mathcal{M}_1$ are extended with a constant kernel to ensure the

Table 1: The log Bayes factors for the GPR analysis and the ARMA baseline. Shown are the mean and standard error over 20 runs.

α	0	1	2	3	4	5	6	7	8
GPR	-27.4 ± 0.3	-16.8 ± 0.6	-9.1 ± 0.6	-1.8 ± 0.5	4.0 ± 0.5	7.0 ± 0.9	9.8 ± 0.6	10.9 ± 0.6	14.2 ± 0.8
ARMA	-10.4 ± 0.4	-10.3 ± 0.4	-10.0 ± 0.5	-9.1 ± 0.3	-8.5 ± 0.6	-7.3 ± 0.5	-7.0 ± 0.4	-4.5 ± 1.4	-4.3 ± 0.6

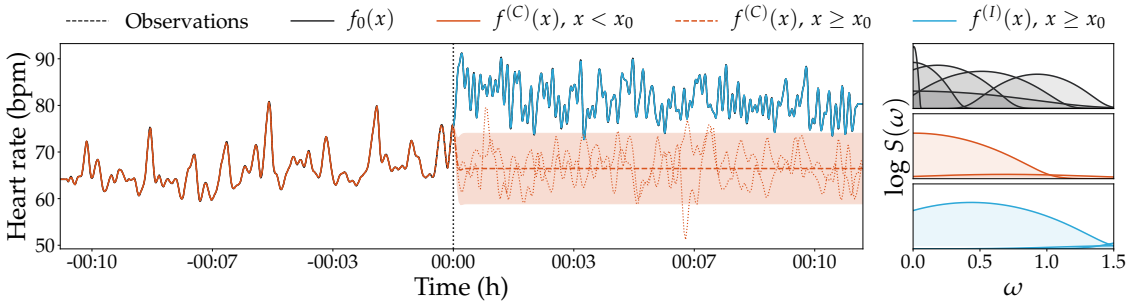

Figure 2: Kundalini Yoga. Model fit and extrapolation for subject 3 of the Kundalini meditation data set. The onset of the actual meditation is shown by the vertical dashed line. Note that the continuous regression f_0 is difficult to see due to its overlap with $f_1^{(C)}$ and $f_1^{(I)}$. To ensure the model comparison does not reflect the mean increase of heart rate, f_0 is extended with a changepoints kernel that transitions between two constant kernels at the intervention points, while $f_1^{(C)}$ and $f_1^{(I)}$ are extended with a constant kernel. The extrapolation, indicated by the dashed (mean) and dotted (posterior samples) red lines, is poor in comparison to the actual observations, which is corroborated by the large log Bayes factor in favor of the causal effect ($\log \mathrm{BF}_{10} = 281.2$).

mean value is not captured by low frequency components.

Figure 2 shows for one subject the corresponding regression and extrapolation. We see that both the continuous and the discontinuous models capture the data well, due to the flexibility of the spectral mixture kernel. However, the continuous model requires more spectral mixture components to do so; $Q = 6$ for f_0 compared to $Q = 2$ for $f_1^{(C)}$ and $Q = 3$ for $f_1^{(I)}$. For each of the participants, the effect is clearly visible, as can be seen from the large positive log Bayes factors (see Table 2). Note that these large Bayes factors follow from the consistency property of Bayes factors, which implies that as more and more evidence comes in, $\log \mathrm{BF}_{10} \to \infty$ (Ly et al., 2016; Stefan et al., 2019).

6. Discussion

There are a number of ways in which the approach could be improved or extended. First of all, we determine the number of latent mixture components Q by simply optimizing the (approximated) log marginal likelihood. Preferably, we would extend the approach by learning Q in a data-driven Bayesian nonparametric way, for instance by replacing the spectral mixture kernel with a spectral

Table 2: The result of the ITS analysis for the meditation data set for each subject j. Shown are the log Bayes factor, the maximum and mean effect size (in bpm increase), and the number of spectral mixture components in the continuous model (Q_0), in the pre-meditation phase (Q_C) and in the meditation phase (Q_I).

j	$\log \mathrm{BF}_{10}$	$d_{\max}$	d_{mean}	Q_0	Q_C	Q_I
1	370.6	50.7	33.6	3	2	2
2	626.2	47.9	31.6	3	3	2
3	281.2	25.0	14.8	6	2	3
4	508.1	63.6	41.7	4	2	2

Dirichlet process mixture kernel (Tian et al., 2019), or a Levy process (Jang et al., 2017).

Furthermore, our contribution adopts standard practices in GPR by replacing the exact log model marginal likelihood in (18) with the BIC approximation (Kim and Teh, 2018; Duvenaud, 2014). This largely ignores the uncertainty in the hyperparameters, and assumes their distribution is unimodal. In practice, we noticed that this procedure is sensitive to its initialization, which suggests that alternatives, such Hamiltonian Monte Carlo (HMC) approaches may be preferred. However, HMC is computationally demanding, and even the computation of the log marginal likelihood in (12) becomes difficult due to its $O(n^3)$ complexity. Sparse GPR may provide a solution (Quiñonero-Candela and Rasmussen, 2005; Lázaro-Gredilla et al., 2010) to scale the approach to data sets with particularly large n.

Our approach assumes, by its dependency on the spectral mixture kernel, that the latent processes f_0, $f_1^{(C)}$, and $f_1^{(I)}$ are stationary. However, using recent work on kernels for non-stationary processes (Remes et al., 2017), this assumption may be lifted, although the substantial increase in flexibility that follows from a non-stationary kernel will make it more difficult to detect an effect in the ITS framework.

In several realistic applications of ITS, such as studies of the effect of novel medication, it is to be expected that an effect is not instantaneously visible. Instead, the observed response will follow an impulse response function (IRF) such as typically estimated in nonlinear multivariate systems (Koop et al., 1996). The IRF can be estimated using GPR (Ambrogioni et al., 2017), and this extension can be incorporated into $\mathcal{M}_1$, so that the transition into a new stationary regime is estimated separately from the spectral mixture of the new stable state itself. This increases the interpretability of the spectral mixture, as it does no longer have to capture the spectral components to reflect the response function.

7. Conclusion

We have presented a novel method for interrupted time series design that uses GPR with a spectral mixture kernel (Wilson and Adams, 2013). The approach describes the similarity in the observed responses as a function of the spectral properties of the latent function. This allows both the null model (in which pre- and post intervention time series have the same spectral mixture) and the alternative model to be able to describe the data adequately. This is in contrast with commonly adopted methods in ITS designs, such as segmented linear regression and ARMA models (Jandoc et al., 2015), that extrapolate poorly on data sets with non-trivial spectral distributions. Our approach is particularly suited for applications in the medical domain, such as detecting the effect of treatment on heart rhythm.

References

L Ambrogioni, M Hinne, M A J van Gerven, and E Maris. GP CaKe: Effective brain connectivity with causal kernels. In I Guyon, U V Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, and R Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 950–959. Curran Associates, Inc., 2017.

J L Bernal, S Cummins, and A Gasparrini. Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology*, 46(1):348–355, 2017.

K Bhaskaran, A Gasparrini, S Hajat, L Smeeth, and B Armstrong. Time series regression studies in environmental epidemiology. *International Journal of Epidemiology*, 42(4):1187–1195, 2013.

C M Bishop. *Pattern recognition and Machine Learning*. Springer, 2006.

H S Bloom. Modern regression discontinuity analysis. *Journal of Research on Educational Effectiveness*, 5(1):43–82, 2012.

S Bochner. *Lectures on Fourier integrals*, volume 42. Princeton University Press, 1959.

Z Branson, M Rischard, L Bornn, and L W Miratrix. A nonparametric Bayesian methodology for regression discontinuity designs. *Journal of Statistical Planning and Inference*, 202:14–30, 2019.

D Duvenaud. *Automatic model construction with Gaussian processes*. PhD thesis, Computational and biological learning laboratory, University of Cambridge, 2014.

A M Figueiredo, A Daponte Codina, M Figueiredo, M Saez, and A Cabrera León. Impact of lockdown on COVID-19 incidence and mortality in China: an interrupted time series study. *Bull World Health Organ*, 2020.

A Goldberger, L Amaral, L Glass, J Hausdorff, P C Ivanov, R Mark, and H E Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation [Online]*, 101(23):e215–e220, 2000.

J D Hamadani, M I Hasan, A J Baldi, S J Hossain, S Shiraji, M S A Bhuiyan, S F Mehrin, J Fisher, F Tofail, S M M U Tipu, S Grantham-McGregor, B-A Biggs, S Braat, and S-R Pasricha. Immediate impact of stay-at-home orders to control COVID-19 transmission on socioeconomic conditions, food insecurity, mental health, and intimate partner violence in Bangladeshi women and their families: an interrupted time series. *The Lancet Global Health*, 2020.

C Hausman and D S Rapson. Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics*, 10(1):533–552, 2018.

M Hinne, Q F Gronau, D van den Bergh, and E-J Wagenmakers. A conceptual introduction to bayesian model averaging. *Advances in Methods and Practices in Psychological Science*, 3(2):200–215, 2020.

R Jandoc, A M Burden, M Mamdani, E L Linda, and S M Cadarette. Interrupted time series analysis in drug utilization research is increasing: systematic review and recommendations. *Journal of Clinical Epidemiology*, 68:950–956, 2015.

P A Jang, A Loeb, M Davidow, and A G Wilson. Scalable levy process priors for spectral kernel learning. In *Advances in Neural Information Processing Systems*, pages 3940–3949, 2017.

H Kim and Y W Teh. Scaling up the automatic statistician: Scalable structure discovery using Gaussian processes. volume 84 of *Proceedings of Machine Learning Research*, pages 575–584, Playa Blanca, Lanzarote, Canary Islands, 09–11 Apr 2018. PMLR.

Y Kim and P Steiner. Quasi-experimental designs for causal inference. *Educational Psychologist*, 51(3-4):395–405, 2016.

G Koop, M H Pesaran, and S M Potter. Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1):119 – 147, 1996.

M Lázaro-Gredilla, J Quiñonero-Candela, C E Rasmussen, and A R Figueiras-Vidal. Sparse spectrum Gaussian process regression. *Journal of Machine Learning Research*, 11:1865–1881, 2010.

A Ly, J Verhagen, and E-J Wagenmakers. Harold jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, 72:19 – 32, 2016.

D J C MacKay. *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, USA, 2002.

I E Marinescu, P N Lawlor, and K P Kording. Quasi-experimental causality in neuroscience and behavioural research. *Nature Human Behaviour*, pages 1–11, 2018.

L K Matowe, C A Leister, C Crivera, and J M Korth-Bradley. Interrupted time series analysis in clinical research. *Ann Pharmacother*, 37(7-8):1110–1116, 2003.

A G Matthews, M van der Wilk, T Nickson, K. Fujii, A Boukouvalas, P León-Villagrá, Z Ghahramani, and J Hensman. GPflow: A Gaussian process library using TensorFlow. *Journal of Machine Learning Research*, 18(40):1–6, apr 2017.

D McDowall, R McCleary, E Meidinger, and R Hay. *Interrupted time series analysis*. Thousand Oaks, CA, USA, 1980.

G P Nason. Stationary and non-stationary time series. *Introductory Econometrics: A Practical Approach*, (1994):273–285, 2013.

A G O'Keeffe and G Baio. Approaches to the estimation of the local average treatment effect in a regression discontinuity design. *Scandinavian Journal of Statistics*, 43(4): 978–995, 2016.

J Pearl and D Mackenzie. *The book of Why: The new science of cause and effect*. Basic Books, 2018.

R B Penfold and F Zhang. Use of interrupted time series analysis in evaluating health care quality improvements, 2013.

C Peng, Joseph E Mietus, Yanhui Liu, Gurucharan Khalsa, Pamela S Douglas, Herbert Benson, and Ary L Goldberger. Exaggerated heart rate oscillations during two meditation techniques. *International Journal of Cardiology*, 70:101–107, 1999.

R Prado and M West. *Time Series: Modeling, Computation, and Inference*. Chapman &Hall/CRC, 1st edition, 2010. ISBN 1420093363.

J Quiñonero-Candela and C E Rasmussen. A unifying view of sparse approximate Gaussian process regression. *Journal of Machine Learning Research*, 6:1939–1959, 2005.

C E Rasmussen and C K I Williams. *Gaussian Processes for Machine Learning*. 2006.

S Remes, M Heinonen, and S Kaski. Nonstationary spectral kernels. *Advances in Neural Information Processing Systems*, 2017-December(Nips):4643–4652, 2017.

D B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

Y Saatçi, R Turner, and C E Rasmussen. Gaussian process change point models. In *Proceedings of the 27th international conference on machine learning*, 2010.

E Schulz, J B Tenenbaum, D Duvenaud, M Speekenbrink, and S J Gershman. Compositional inductive biases in function learning. *Cognitive Psychology*, 99:44 – 79, 2017.

G Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.

A M Stefan, Q F Gronau, F D Schönbrodth, and E-J Wagenmakers. A tutorial on Bayes Factor Design Analysis using an informed prior. *Behavior Research Methods*, 51:1042–1058, 2019.

J Tian, P Yan, and D Huang. Kernel analysis based on Dirichlet processes mixture models. *Entropy*, 21(9):1–18, 2019.

J T Vanderplas. Understanding the Lomb–Scargle periodogram. *The Astrophysical Journal Supplement Series*, 236(1):16, 2018.

Z Vokó and J G Pitter. The effect of social distance measures on COVID-19 epidemics in Europe: an interrupted time series analysis. *GeroScience*, 42(4):1075–1082, 2020.

E-J Wagenmakers. A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5):779–804, 2007.

A G Wilson. Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes. *thesis, U. of Cambridge*, (January):226, 2014.

A G Wilson and R Adams. Gaussian process kernels for pattern discovery and extrapolation. volume 28 of *Proceedings of Machine Learning Research*, pages 1067–1075, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

Appendix A. Simulation study: generative model

The data used for in section 4 is generated the following way:

$$y_i \sim \mathcal{N}(f(x_i), \sigma^2), \tag{20}$$

where $f(x_i)$ is defined as

$$f(x_i) = \sin(12x_i) + \tfrac{2}{3}\cos(25x_i) \tag{21}$$

for $x < x_0$ and

$$f(x_i) = \sin((12+\alpha)x_i) + \frac{2}{3}\cos((25+\alpha)x_i) \tag{22}$$

otherwise, where $\sigma = 0.2$ and $x_0 = 0$. For each data set, $n = 200$ equally spaced observations are generated such that the number of observations is the same for the control and intervention model. We simulate 20 data sets for each value of $\alpha \in \{0, 1, \ldots, 8\}$. At the intervention threshold, the frequency of both functions is increased by the discontinuity size α.

Appendix B. Kundalini yoga: Other participants

The results of our analysis for the other participants of the data discussed by Peng et al. (1999) is shown in Figure 3.

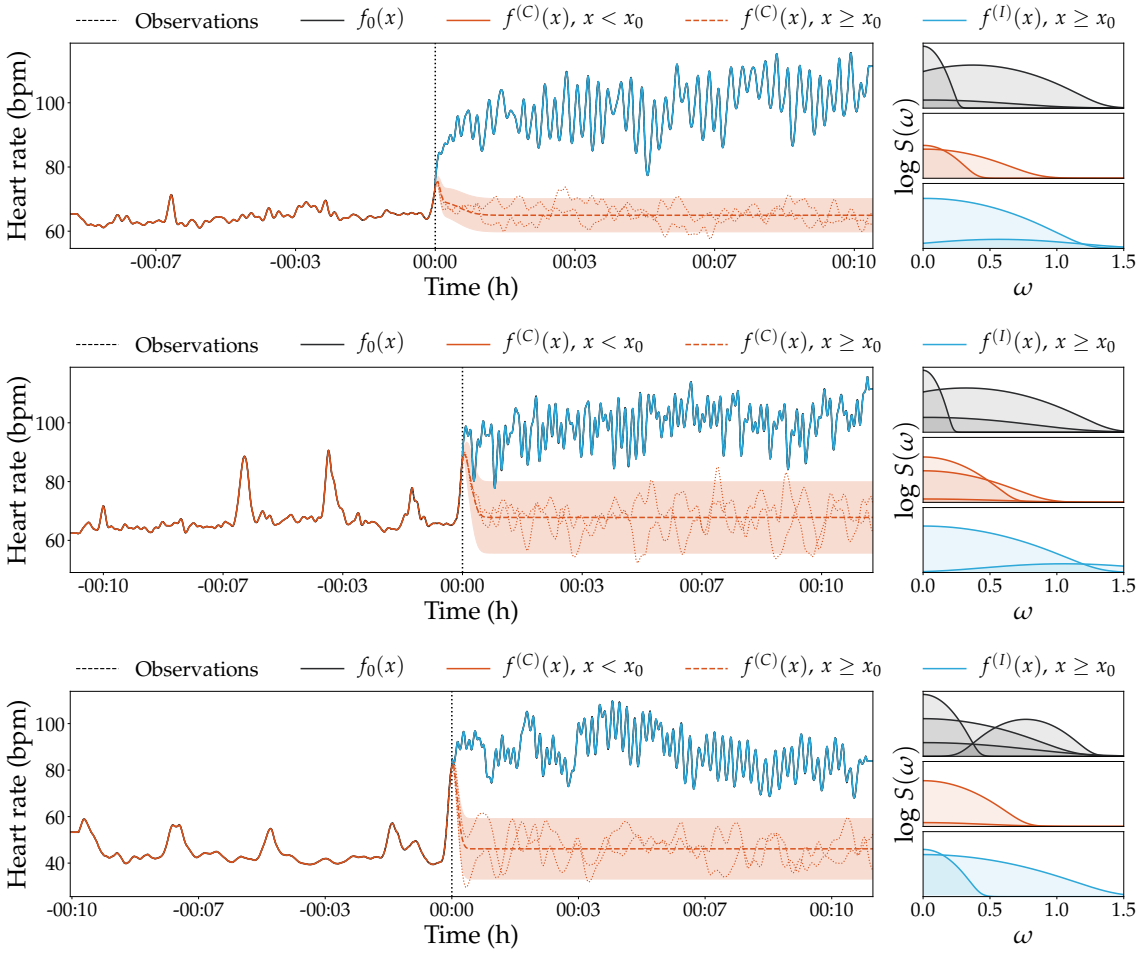

Figure 3: Model fit and extrapolation for the remaining Kundalini meditation data sets. The onset of the actual meditation is shown by the vertical dashed line. Note that the continuous regression f_0 is difficult to see due to its overlap with $f_1^{(C)}$ and $f_1^{(I)}$.