Moderated and Drifting Linear Dynamical Systems

Jinyan Guan [1]   JGUAN1@EMAIL.ARIZONA.EDU
Kyle Simek [1]   KSIMEK@EMAIL.ARIZONA.EDU
Ernesto Brau [2]   BRAUAVIL@BC.EDU
Clayton T. Morrison [3]   CLAYTONM@EMAIL.ARIZONA.EDU
Emily A. Butler [4]   EABUTLER@EMAIL.ARIZONA.EDU
Kobus Barnard [1]   KOBUS@CS.ARIZONA.EDU

[1] Department of Computer Science, University of Arizona [2] Department of Computer Science, Boston College
[3] School of Information: Science, Technology, and Arts, University of Arizona
[4] Norton School of Family and Consumer Sciences, University of Arizona

Abstract

We consider linear dynamical systems, particularly coupled linear oscillators, where the parameters represent meaningful values in a domain theory, and thus learning what affects them contributes to explanation. Rather than allow perturbations of latent states, we assume that temporal variation beyond noise is explained by parameter drift, and variation across coupled systems is a function of moderating variables. This change in model structure reduces opportunities for efficient inference, and we propose sampling procedures to learn and fit the models. We test our approach on a real dataset of self-recalled emotional experience measurements of heterosexual couples engaged in a conversation about a potentially emotional topic, with body mass index (BMI) being considered as a moderator. We evaluate several models on their ability to predict future conversation dynamics (the last 20% of the data for each test couple), with shared parameters being learned using held out data. We validate the hypothesis that BMI affects the conversation dynamic in the experimentally chosen topic.

1. Introduction

In many standard applications of linear dynamical systems (LDS), such as tracking an object moving under Newton's laws, the parameters are known, and the latent states (e.g., the true object position) are the target of inference. Further, it may make sense that the latent state is subject to

perturbations (e.g., the object is buffeted by air), in which case it is convenient to model the latent state with Gaussian perturbations at each time step. In other applications, the parameters need to be inferred for a predictive task. In this case, the trade-offs in parameter, latent state, and observation errors might be optimized for prediction, and the relation between them may not matter so much. By contrast, for the kind of applications considered here, the parameters correspond directly to quantities in a scientific model, and therefore understanding what affects those parameters across individual examples is important. Here it makes less sense to allow latent state noise, especially when observation noise is substantive. Further, given that we consider the model itself as indicative of individuals and their circumstances, this suggests that temporal variation beyond noise processes might be better explained by changes (drift) in the model. Hence, in this paper we force all variation beyond observation noise to be the result of deterministic process with stochastic parameters.

We take as a motivating example a class of social psychological models aimed at explaining the emotional dynamics of interacting couples. Growing evidence suggests that coupled linear oscillators (CLOs), a subset of LDS models, are well-suited to model the oscillatory dynamics of emotional interactions. However, most work to-date (Boker & Ghisletta, 2001; Boker et al., 2004; Chow et al., 2005; Butner et al., 2005; Boker & Laurenceau, 2006; Finan et al., 2010; Ferrer & Steele, 2011; Steele & Ferrer, 2011; Ferrer & Helm, 2013) has been limited to fitting the data using regression methods that rely on necessarily noisy estimates of first and second derivatives of the observations. While convenient to implement using common statistical packages, there are serious limitations to this regression-based approach underlying all CLO modeling work in the social sciences domain. These limitations include the following: 1) there is a fundamental disconnect between error

as explained by the model and error that is minimized during regression; 2) directly estimating derivatives from noisy data leads to poor estimates; LDS derivatives are better applied to estimated latent states; 3) these regression models do not fit initial states, which if used at all, are taken from the data points or their mean; 4) there is no natural extension to a multivariate observation case; and finally 5) these regression approaches are only evaluated based on model residual error reduction rather than on predictive power.

In this work, we develop a generative Bayesian model and inference algorithms for LDS models (instantiated in the case of CLO models) that avoids the above limitations. We then extend the models in two ways. First, we consider that the parameters are a function of moderating variables. Second, we allow the model parameters to drift under a Gaussian process. These extensions, together with the requirement of flexible models with multiple latent variables, rule out analytic solutions, so we develop sampling-based inference approaches for learning and fitting these models.

Experimental domain. In the particular application showcased here, the latent variables are the emotional states of romantically involved male-female partners during a conversation. Available temporal observations range from physiological measurements to self-reported ratings on emotional state (see Section 6); we chose the latter for our experiments. Potential moderators include body mass index (BMI) for each participant, which we used for the experiments reported here. One hypothesis driving the data collection supposes that BMI influences conversation dynamics when the topic of discussion is related to health.

We consider that all data is generated for each participant independently, conditioned on their latent emotion variable. As an LDS, the new states for both partners are a linear function of the previous states of both. As we are interested in the nature of the dynamic process, we choose to make the state variables deterministic (no latent state noise). The parameters for the LDS may be individualistic (which over-fits), shared, or a function of moderators such as BMI, reflected by the original experiment design. Each of these models are combined with parameter drift via a Gaussian process (e.g., increasing anger is reflected in changes in reactions to one's partner).

Contributions. **1).** We introduce Gaussian process parameter drift in coupled LDS models, making it possible for model variation within a system over time to be accounted for by model parameter changes rather than latent state perturbations. We develop a Gibbs sampler for inference in this model. **2).** We similarly introduce parameter moderation to explain coupled LDS data over multiple instances (e.g., a number of couples), for situations where parameters, as signatures of dynamic behavior, can be linked to other data such as health related outcomes (e.g., BMI). **3).**

We detail a multilevel cross validation scheme in which shared parameters are learned and then used to predict near future interactions in held-out couples. **4).** We introduce a new domain for predictive modeling, namely interpersonal emotional processes and similar sub-areas of social psychology. This domain has very challenging data and modern machine learning methods for building predictive and explanatory models has had limited impact so far.

2. Related work

Finding good parameter values for ordinary differential equations (ODE) from observables has significant history. For example, researchers (Bard, 1973; Varah, 1982; Anger, 1990; Li et al., 2005) have looked at iterating over parameter sets, integrating the ODE for each set. To reduce the computational cost, smoothing estimators of ODEs have been proposed (Varah, 1982; Ramsay et al., 2007; Liang & Wu, 2008; Dattner & Klaassen, 2013; Hall & Ma, 2013) that do not require solving the ODE numerically. However, they need reliable estimates of the derivatives of the states, which is problematic with noisy and sparse data.

In the domain of biomedical systems, Gelman *et al.* propose inferring parameters of the differential equation of pharmacokinetic (PK) models using Bayesian posterior simulation (1996). They apply priors based on scientific theories and estimate the distribution of individual characteristics. Reviews of statistical and computational work in PK models are provided by Davidian and Giltinan (2003) and Pillai *et al.* (2005). Li *et al.* introduce coupling between the variables for this domain (2004) and fit parameters using a two step iterative process. They also model parametric change over time but with simple linear or piecewise constant models. Subsequently, Li *et al.* and Yu *et al.* introduce an MCMC sampling approach for fitting the parameters (2007; 2008). Like our proposed method, all of these involve parameter fitting and hierarchical Bayesian modeling. However, the specifics of the domain allow for special relationships between variables and require only first derivatives, facts which are exploited for inference. Here we develop an approach for arbitrary LDS.

Calderhead *et al.* propose a Bayesian statistical approach to treat the hidden state variables as random variables with Gaussian process (GP) priors (2009). This approach replaces expensive ODE integration with closed-form marginalization over the derivative states. Chkrebtii *et al.* extended this idea to a general methodology for general systems of differential equations (2013). This is further improved by Dondelinger *et al.* who sample from the joint distribution of the ODE parameters and GP hyperparameters (2013). Wang and Barber improved the GP-based ODE models by directly linking the state derivative information with the system observations (2014). In general, this body

of work focuses on using GPs for hidden state variation instead of parameter variation as we propose.

Gaussian processes have also been used to model the latent parameters of differential equations (Lawrence et al., 2006; Titsias et al., 2008; Gao et al., 2008) with particular forms. In system biology, Lawrence *et al.* treat the concentration of a transcription factor as a latent function with a Gaussian process prior (2006). When the production rate depends linearly on the transcription factor, parameter estimation is exact and tractable. When the dependence is not linear, they use the functional gradient of the likelihood and prior to learn the *maximum a posteriori* (MAP) solution for the transcription factors and other hyper-parameters by MAP-Laplace approximation. In a later work, Titsias *et al.* developed an efficient Markov Chain Monte Carlo (MCMC) sampling approach with novel proposal distribution to find the MAP estimate of the model parameters (2008). In these models, time-varying parameters are confined to the constant coefficient of a linear ODE, whereas our model allows all coefficients to drift. These works inspired the Latent Force Models (LFMs) for differential equations (Alvarez et al., 2009) but coupling is not considered. Significant work has been done on modeling phase shift, characterized by inferring discontinuous changes in model parameters (e.g., (S. M. Oh et al., 2008; A. S. Willsky et al., 2009)). In contrast, we study progressively and smoothly changing parameters.

In summary, many elements of our work are shared with prior efforts. However, the proposed parameter drift model for a general LDS and the associated Gibbs sampler are new, as is the particular approach to integrating moderation.

3. Model

We represent each person's hidden emotional state as a real function of time. The observables are self-reported ratings of the person's hidden emotional state over time. Let $(x_t^a, x_t^b) \in \mathbb{R}^2$ be a couple's hidden emotional state at time $t \geq 1$. Higher values of x_t^a and x_t^b represent positive emotion, while lower values represent negative emotion. Because interactions tend to be asymmetric, the ordering of couples matters. For example, in the case of heterosexual couples used in our experiments, x_t^a and x_t^b denote the male and female partner, respectively. Similarly, parent/child relationships would also be identifiable and asymmetric (our models are easily extended to more than two people).

For an interacting couple, we model their emotional dynamics as a coupled oscillator system, in which an individual's future emotional state is influenced by her current state and the emotional state of her partner. The parameters of this model are different for each couple, but allowing them to be completely free (baseline model CLO-

MLE) leads to over-fitting. Instead, for most models considered, the parameters come from a common distribution, with the mean potentially being a function of moderators (here, BMI). Finally, we explore a model that allows parameters to drift smoothly over time.

3.1. Coupled Linear Oscillator

Damped linear oscillator models have been used to represent the regulatory processes of emotion oscillations of an individual (Chow et al., 2005). A coupled linear oscillator (CLO) is an extension of the damped linear model to model interactions between two oscillators. It is useful to model both intra- and inter-personal characteristics, including frequency, damping, and coupling (Boker & Laurenceau, 2007). In a CLO system, the accelerations of a pair of oscillators are a function of their velocities and positions. Specifically, if we let x_t^a and x_t^b be the positions of the oscillators at time t, their dynamics are defined by the second-order differential equation

$$\ddot{x}_t^a = f^a x_t^a + d^a \dot{x}_t^a + c^a(x_t^b - x_t^a) \tag{1}$$
$$\ddot{x}_t^b = f^b x_t^b + d^b \dot{x}_t^b + c^b(x_t^b - x_t^a). \tag{2}$$

Here, $\ddot{x}_t = \frac{d^2 x}{dt^2}$ denotes acceleration and $\dot{x}_t = \frac{dx}{dt}$ denotes velocity at time t, f^a and f^b are the oscillation frequencies, d^a and d^b are the linear damping coefficients, and c^a and c^b are the coupling terms between the oscillators.

3.2. State space representation

A system that is defined by a set of differential equations can be represented in state space (Nise, 2010). Typically, a deterministic continuous-time dynamic system is represented by the following equations:

$$\dot{\mathbf{x}}_t = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t \tag{3}$$
$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t, \tag{4}$$

where $\mathbf{x}$ is the vector containing the system's state variables. Given a state $\mathbf{x}_{t_1}$ at time t_1, the future state $\mathbf{x}_t$ at any time $t > t_1$ is completely determined by the system matrix $\mathbf{A}$ and the optional input matrix $\mathbf{B}$ and input/control vector $\mathbf{u}_t$. The output vector $\mathbf{y}$ is determined by the output matrix $\mathbf{C}$ and optional feed-forward matrix $\mathbf{D}$ and $\mathbf{u}_t$. In our case, both $\mathbf{B}$ and $\mathbf{D}$ are zero matrices.

A stochastic discrete-time dynamic system is defined by a stochastic process in which the value of $\mathbf{x_i}$ at time t_i is stochastically determined by the value of $\mathbf{x_{i-1}}$ at time t_{i-1}:

$$\mathbf{x}_i = g(\mathbf{u}_i, \mathbf{x}_{i-1}, \boldsymbol{\epsilon}_i) \tag{5}$$
$$\mathbf{y}_i = h(\mathbf{x}_i, \mathbf{u}_i, \boldsymbol{\delta}_i), \tag{6}$$

where g is the transition model, h is the observation model, $\boldsymbol{\epsilon}_i$ is the system noise at time t_i and $\boldsymbol{\delta}_i$ is the observation

noise at time t_i (Murphy, 2012). State $\mathbf{x}_i$ can be obtained from $\mathbf{x}_{i-1}$ by numerical integration, using the derivative of $\mathbf{x}$ defined in equation (3) with a certain step size Δt. We use zero system noise ($\boldsymbol{\epsilon}_t$), so that variance is explained either by observation noise or parameter drift, in support of domain theory.

We let the state vector $\mathbf{x}_t = (x_t^{\mathrm{a}}, x_t^{\mathrm{b}}, \dot{x}_t^{\mathrm{a}}, \dot{x}_t^{\mathrm{b}})^T$ represent the emotion states and their derivatives with respect to time for both partners, where $(x_t^{\mathrm{a}}, x_t^{\mathrm{b}}) \in \mathbb{R}^2$. The derivative of the state vector is $\dot{\mathbf{x}}_t = (\dot{x}_t^{\mathrm{a}}, \dot{x}_t^{\mathrm{b}}, \ddot{x}_t^{\mathrm{a}}, \ddot{x}_t^{\mathrm{b}})$. Using (3) to represent equations (1) and (2) results in the state equation:

$$\begin{pmatrix} \dot{x}_t^{\mathrm{a}} \\ \dot{x}_t^{\mathrm{b}} \\ \ddot{x}_t^{\mathrm{a}} \\ \ddot{x}_t^{\mathrm{b}} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ f^{\mathrm{a}} - c^{\mathrm{a}} & c^{\mathrm{a}} & d^{\mathrm{a}} & 0 \\ c^{\mathrm{b}} & f^{\mathrm{b}} - c^{\mathrm{b}} & 0 & d^{\mathrm{b}} \end{pmatrix} \begin{pmatrix} x_t^{\mathrm{a}} \\ x_t^{\mathrm{b}} \\ \dot{x}_t^{\mathrm{a}} \\ \dot{x}_t^{\mathrm{b}} \end{pmatrix}. \quad (7)$$

Given the initial states $(x_{t_1}^{\mathrm{a}}, x_{t_1}^{\mathrm{b}}, \dot{x}_{t_1}^{\mathrm{a}}, \dot{x}_{t_1}^{\mathrm{b}})$ we can compute the state at any time $t > t_1$ by evolving the differential equation forward in time.

The observations consist of self-rated subjective feelings, which we model as the latent emotional state perturbed by Gaussian noise. Let the output vector $\mathbf{y} = (y_t^{\mathrm{a}}, y_t^{\mathrm{b}})$ be the observed state; in the case of univariate output per individual, the observation model (eq. 6) then becomes

$$\begin{pmatrix} y_t^{\mathrm{a}} \\ y_t^{\mathrm{b}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_t^{\mathrm{a}} \\ x_t^{\mathrm{b}} \\ \dot{x}_t^{\mathrm{a}} \\ \dot{x}_t^{\mathrm{b}} \end{pmatrix} + \begin{pmatrix} \delta_t^{\mathrm{a}} \\ \delta_t^{\mathrm{b}} \end{pmatrix}, \quad (8)$$

where $\delta_t^{\mathrm{a}}, \delta_t^{\mathrm{b}} \sim \mathcal{N}(0, \sigma_o^2)$. We can easily extend the output vector to be multivariate by modifying the output matrix $\mathbf{C}$ and utilizing a non-zero feed-forward matrix $\mathbf{D}$ to include other components of observed emotion states, such as heart rate and facial expressions.

3.3. Bayesian hierarchical model

We develop a Bayesian hierarchical model which allows us to estimate the group variability to provide a prior distribution for individuals. While we expect different couples to have different emotional dynamics, in similar social settings couples might exhibit common patterns, such as the frequencies of the oscillations. Furthermore, it has been shown that health indicators such as body weight can influence emotion interaction patterns in couples (Reed et al., 2015). Therefore, we define each couple's CLO parameters as arising from a linear function of both partners' body mass index (BMI) with Gaussian noise. This introduces dependence between couples by sharing linear coefficients. Specifically, let $\boldsymbol{\theta}_i = (f_i^{\mathrm{a}}, f_i^{\mathrm{b}}, d_i^{\mathrm{a}}, d_i^{\mathrm{b}}, c_i^{\mathrm{a}}, c_i^{\mathrm{b}})$ be the parameters for couple i, and let $(\omega_i^{\mathrm{a}}, \omega_i^{\mathrm{b}})$ be their BMI values, then each element θ_{ij} is given by:

$$\theta_{ij} = \mu_{ij} + \epsilon_j,$$

where

$$\mu_{ij} = \alpha_j + \beta_j \cdot \overline{\omega}_i + \gamma_j \cdot \hat{\omega}_i \quad (9)$$

and $\overline{\omega}_i = (\omega_i^{\mathrm{a}} + \omega_i^{\mathrm{b}})/2$; $\hat{\omega}_i = \omega_i^{\mathrm{b}} - \omega_i^{\mathrm{a}}$; $\epsilon_j \sim \mathcal{N}(0, \sigma_j^2)$. Here, α_j, β_j, and γ_j are the shared coefficients of the BMI values and σ_j^2 is the variance for the j-th parameter among the couples. We define a group-shared CLO model as at least one of the CLO parameter having a non-zero α coefficient. By using different combinations of non-zero β and γ on different parameters, we can explore various CLO models with group-shared and BMI-dependent parameters. For example, we can allow the same gender of all the couples to share the same prior distribution for the damping value and let each person to have different value for frequency and coupling parameters. We will later show that the prior on $\boldsymbol{\theta}_j$ of all the couples forms a Bayesian linear regression model and can be sampled by using Gibbs-sampling (section 4). We let the joint prior distribution of $\mathbf{q}_j = (\alpha_j, \beta_j, \gamma_j)$ and σ_j^2 be the Normal-Inverse-Gamma (NIG), i.e., $(\mathbf{q}_j, \sigma_j^2) \sim \mathcal{NIG}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, a_0, b_0)$, and treat each $(\mathbf{q}_j, \sigma_j^2)$ as independent. Figure 1 shows the graphical model for group-shared CLO model with BMI-dependent means and couple-shared variance.
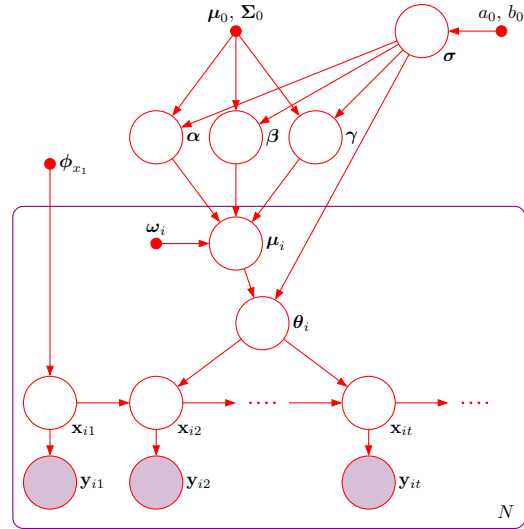

Figure 1. A graphical model of the state space model for the jth parameter (index j is omitted for clarity) with BMI-dependent Gaussian prior. Here, the jth CLO parameter θ_{ij} for couple i is a random variable generated from distribution $\mathcal{N}(\mu_{ij}, \sigma_j^2)$. Group-shared BMI-coefficients α_j, β_j and γ_j and shared variance σ_j^2 are random variables generated from the joint prior distributions $(\alpha_j, \beta_j, \gamma_j, \sigma_j^2) \sim \mathcal{NIG}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, a_0, b_0)$. ϕ_{x_1} represents a multivariate Gaussian prior for the initial state.

3.4. Time-varying CLO parameters

By letting the CLO parameters change over time, we can model emotion dynamics in longer period of social interactions where the interaction dynamics vary over time. Re-

call that the elements of A are the CLO parameters. In a system with fixed dynamics, the state matrix A in equation 3 are constants. By letting A change over time, we can model a system with changing dynamics. To ensure smooth changes of the dynamics, we define a Gaussian process prior over sequences of CLO parameters.

Gaussian process prior A Gaussian process (GP) is a generalization of the Gaussian distribution that defines a probability distribution over the space of of continuous functions (Rasmussen & Williams, 2006). A Gaussian process is specified by a mean function $m(t)$ and a covariance function $k(t, t')$. If a random function $f : \mathbb{R} \to \mathbb{R}$ is GP-distributed, then every finite subset $(f(t_1), \ldots, f(t_J)) \sim \mathcal{N}((m(t_1), \ldots, m(t_J)), \mathbf{K}_{ij} = k(t_i, t_j)))$. Let $\boldsymbol{\theta}(t) = (f^{\mathrm{b}}(t), f^{\mathrm{a}}(t), d^{\mathrm{b}}(t), d^{\mathrm{a}}(t), c^{\mathrm{b}}(t), c^{\mathrm{a}}(t))$ be the CLO parameters for some couple as a function of time; for simplicity we omit the couple's subscript. Each parameter $\theta_j(t)$ is an independent random function with distribution

$$\theta_j(t) \sim \mathcal{GP}(m_j(t), k_j(t, t')). \qquad (10)$$

We define the mean $m_j(t) = \mu_j$ as defined in equation (9). The covariance is the squared-exponential function $k_j(t, t') = \sigma_j^2 \exp(-\frac{(t-t')^2}{2s_j^2})$, which generates smooth functions. It is defined by two parameters that are shared across couples: a variance parameter σ_j^2 that controls the overall scale of variations, and a scale parameter s_j, which controls the smoothness. The graphical model with Gaussian process prior on the parameters $\boldsymbol{\theta}_i$ is shown in Fig. 2.
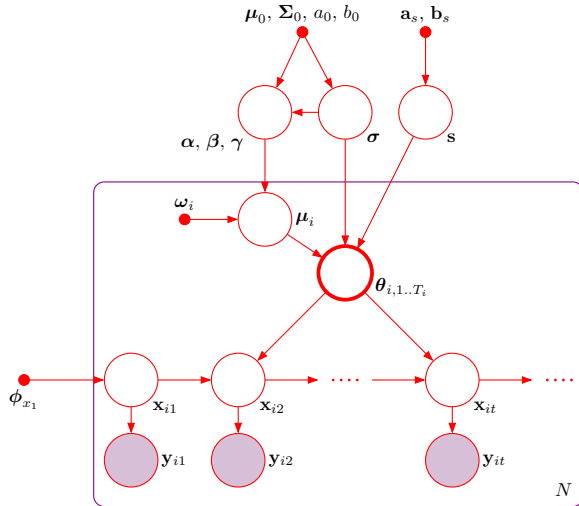

Figure 2. A graphical model of the state space model for the jth parameter (index j is omitted for clarity) with Gaussian process prior on the CLO parameters $\boldsymbol{\theta}_{i,1..T_i}$, where T_i is the time-length of couple i. The mean function $\mathbf{m}$ is defined by $\boldsymbol{\mu}_i$ and covariance function $\mathbf{k}(\mathbf{t}, \mathbf{t}')$ is defined by the shared signal variance $\boldsymbol{\sigma}$ and scale parameters $\mathbf{s}$. Here, $s_j \sim \mathcal{IG}(a_{sj}, b_{sj})$ where s_j is the length-scale of the jth CLO parameter.

3.5. Likelihood

Given the hidden emotional states of a couple and the values of its CLO parameters, the observed states at each discrete time t are conditionally independent. Since the hidden emotional states are deterministically determined by the CLO parameters and the initial values, for each couple i, the likelihood of the observed data $\mathbf{y}_{i,1:T_i}$ conditioned on the CLO parameters $\boldsymbol{\theta}_i$ and the initial state $\mathbf{x}_{i,1}$ is given by:

$$p(\mathbf{y}_{i,1:T_i}|\mathbf{x}_{i,1}, \boldsymbol{\theta}_{i,1:T_i}) = p(\mathbf{y}_{i,1}|\mathbf{x}_{i,1}) \prod_{t=2}^{T_i} p(\mathbf{y}_{i,t}|\mathbf{x}_{i,1}, \boldsymbol{\theta}_{i,t}),$$
(11)

where $\mathbf{y}_{i,1}|\mathbf{x}_{i,1} \sim \mathcal{N}(\mathbf{x}_{i,1}, \sigma_o^2)$, and $\mathbf{y}_{i,t}|\mathbf{x}_{i,1}, \boldsymbol{\theta}_{i,t} \sim \mathcal{N}(f(\mathbf{x}_{i,1}, \boldsymbol{\theta}_{i,t}, t), \sigma_o^2)$, where σ_o^2 is the variance in the observations. The value of function f is the numerical solution of the ODE defined by the state equation 7. Since all the couples are conditionally independent given the model parameters, the likelihood of all the couples is:

$$p(\mathbf{Y}_{1:T}|\mathbf{X}_1, \boldsymbol{\Theta}) = \prod_{i=1}^{N} \left(p(\mathbf{y}_{i,t=1}|\mathbf{x}_{i,1}) \prod_{t=2}^{T_i} p(\mathbf{y}_{i,t}|\mathbf{x}_{i,1}, \boldsymbol{\theta}_{i,t}) \right).$$
(12)

4. Parameter learning

Let $\mathbf{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_6\}$ be the set of BMI coefficients for all six model parameters. When the CLO parameters $\boldsymbol{\Theta}$ do not change over time, we learn the group-shared $\mathbf{Q}$ and $\boldsymbol{\sigma}$ by drawing samples from the joint distribution $p(\mathbf{Q}, \boldsymbol{\sigma}, \boldsymbol{\Theta}, \mathbf{X}_1|\mathbf{Y}_{1:T})$ under the hierarchical model using MH-Gibbs sampling (Alg. 1). For the time-varying CLO model, we also learn the length-scale parameter $\mathbf{s}$ by drawing samples from $p(\mathbf{Q}, \boldsymbol{\sigma}, \mathbf{s}, \boldsymbol{\Theta}, \mathbf{X}_1|\mathbf{Y}_{1:T})$ (Alg. 2).

Algorithm 1 Sample procedure from $p(\mathbf{Q}, \boldsymbol{\sigma}, \boldsymbol{\Theta}, \mathbf{X}_1|\mathbf{Y})$

Initialize $\boldsymbol{\Theta}^{(1)}, \mathbf{X}_1^{(1)}, \mathbf{Q}^{(1)}$ and $\boldsymbol{\sigma}^{(1)}$.
repeat
$\quad \mathbf{X}_1^{(m)} \sim p(\mathbf{X}_1|\boldsymbol{\Theta}^{(m-1)}, \mathbf{Q}^{(m-1)}, \boldsymbol{\sigma}^{(m-1)}, \mathbf{Y})$
$\quad \boldsymbol{\Theta}^{(m)} \sim p(\boldsymbol{\Theta}|\mathbf{Q}^{(m-1)}, \sigma^{(m-1)}, \mathbf{X}_1^{(m)}, \mathbf{Y})$
$\quad \{\mathbf{Q}^{(m)}, \boldsymbol{\sigma}^{(m)}\} \sim p(\mathbf{Q}, \boldsymbol{\sigma}|\boldsymbol{\Theta}^{(m)}, \mathbf{X}_1^{(m)}, \mathbf{Y})$
until $p(\boldsymbol{\Theta}, \mathbf{Q}, \boldsymbol{\sigma}, \mathbf{X}_1|\mathbf{Y})$ converges

---

For the model with fixed CLO parameters, we use random-walk Metropolis-Hastings (MH) to sample the initial states of each couple from $p(\mathbf{X}_1|\boldsymbol{\Theta}, \mathbf{Y})$ and $p(\boldsymbol{\Theta}|\mathbf{Q}, \boldsymbol{\sigma}, \mathbf{X}_1, \mathbf{Y})$. For the model with time-varying parameters, we use MH to sample $\boldsymbol{\Theta}$ with a proposal distribution constructed from a set of control points of the $\mathcal{GP}$ process as proposed by (Titsias et al., 2008). For updating group-shared parameters $\mathbf{Q}$ and $\boldsymbol{\sigma}$, we use Gibbs to sample from joint posterior distribution $p(\mathbf{Q}, \boldsymbol{\sigma}|\boldsymbol{\Theta})$ and $p(\mathbf{Q}, \boldsymbol{\sigma}|\boldsymbol{\Theta}, \mathbf{s})$ (in the time-varying case) by using a conjugate NIG prior dis-

tribution on $p(\mathbf{Q}, \boldsymbol{\sigma})$. Furthermore, we use Hamiltonian Monte Carlo (also called Hybrid Monte Carlo) (HMC) sampling technique (Duane et al., 1987; Neal, 1997) to sample the scale parameters $\mathbf{s}$ from its conditional posterior $p(\mathbf{s}|\mathbf{Q}, \boldsymbol{\sigma}, \boldsymbol{\Theta}) \propto p(\mathbf{s})p(\boldsymbol{\Theta}|\mathbf{s}, \mathbf{Q}, \boldsymbol{\sigma})$ where $p(\mathbf{s}) = \mathcal{IG}(\mathbf{a}_s, \mathbf{b}_s)$. Since we have minimal prior knowledge about how fast the parameters changes over time, we use a vague Inverse Gamma prior for $\mathbf{s}$ by setting both the shape a_s and the scale b_s of all the six CLO parameters to be 0.001.

Algorithm 2 Sample procedure from $p(\mathbf{Q}, \boldsymbol{\sigma}, \mathbf{s}, \boldsymbol{\Theta}, \mathbf{X}_1|\mathbf{Y})$

Initialize $\boldsymbol{\Theta}^{(1)}$, $\mathbf{X}_1^{(1)}$, $\mathbf{Q}^{(1)}$, $\boldsymbol{\sigma}^{(1)}$, and $\mathbf{s}^{(1)}$.
repeat
$\quad \mathbf{X}_1^{(m)} \sim p(\mathbf{X}_1|\boldsymbol{\Theta}^{(m-1)}, \mathbf{Q}^{(m-1)}, \boldsymbol{\sigma}^{(m-1)}, \mathbf{s}^{(m-1)}, \mathbf{Y})$

$\quad \boldsymbol{\Theta}^{(m)} \sim p(\boldsymbol{\Theta}|\mathbf{Q}^{(m-1)}, \boldsymbol{\sigma}^{(m-1)}, \mathbf{s}^{(m-1)}, \mathbf{X}_1^{(m)}, \mathbf{Y})$
$\quad \{\mathbf{Q}^{(m)}, \boldsymbol{\sigma}^{(m)}\} \sim p(\mathbf{Q}, \boldsymbol{\sigma}|\boldsymbol{\Theta}^{(m)}, \mathbf{s}^{(m-1)}, \mathbf{X}_1^{(m)}, \mathbf{Y})$
$\quad \mathbf{s}^{(m)} \sim p(\mathbf{s}|\mathbf{Q}^{(m)}, \boldsymbol{\sigma}^{(m)}, \boldsymbol{\Theta}^{(m)}, \mathbf{X}_1^{(m)}, \mathbf{Y})$
until $p(\boldsymbol{\Theta}, \mathbf{Q}, \boldsymbol{\sigma}, \mathbf{s}, \mathbf{X}_1|\mathbf{Y})$ converges

---

4.1. Gibbs sampling for Q and σ when θ are constant

In the model with constant CLO parameters, the jth CLO parameter for the ith couple has Gaussian prior distribution

$$\theta_{ij}|\mathbf{q}_j \sim \mathcal{N}(\boldsymbol{\omega}_i^\top \mathbf{q}_j, \sigma_j^2), \qquad (13)$$

where $\boldsymbol{\omega}_i = (1, \bar{\omega}_i, \hat{\omega}_i)^\top$. Let $\boldsymbol{\theta}_j = (\theta_{1j}, \theta_{2j}, \ldots, \theta_{Nj})^\top$ be the jth CLO parameters for all couples; its joint prior is

$$\boldsymbol{\theta}_j|\mathbf{q}_j \sim \mathcal{N}(\boldsymbol{\Omega}\mathbf{q}_j, \sigma_j^2\mathbf{I}), \qquad (14)$$

where $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_N)^\top$.

The full conditional distribution over BMI weights $\mathbf{q}_j$ and noise variance σ_j^2 is given by standard Bayesian regression, with the likelihood in (14) and conjugate prior $(\mathbf{q}_j, \sigma_j^2) \sim \mathcal{NIG}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, a_0, b_0)$:

$$\mathbf{q}_j, \sigma_j^2|\boldsymbol{\theta}_j \sim \mathcal{NIG}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*, a_*, b_*), \qquad (15)$$

where

$$\boldsymbol{\mu}_* = \left(\boldsymbol{\Omega}^\top\boldsymbol{\Omega} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\Omega}^\top\boldsymbol{\theta}\right), \qquad (16)$$

$$\boldsymbol{\Sigma}_* = \left(\boldsymbol{\Omega}^\top\boldsymbol{\Omega} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}, \qquad (17)$$

$$a_* = a_0 + N/2, \qquad (18)$$

$$b_* = b_0 + \left(\boldsymbol{\theta}^\top\boldsymbol{\theta} + \boldsymbol{\mu}_0\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}_*\boldsymbol{\Sigma}_*^{-1}\boldsymbol{\mu}_*\right)/2. \qquad (19)$$

4.2. Gibbs sampling Q and σ when θ has GP prior

In the model with time-varying CLO parameters, we model the changing parameters with the GP prior in equation (10). The covariance function $k()$ introduces time-correlations,

while the mean equation and the signal variance introduces dependence between couples via shared α, β, γ, and σ^2. This requires us to model parameters jointly over times and couples. For notational simplicity, we omit the subscript j over θ and $\mathbf{q}$ for the remainder of this section; all CLO parameters follow the same derivation.

Let $\boldsymbol{\theta} = (\theta_{1,1}, \theta_{1,2}, \cdots, \theta_{i,t}, \cdots, \theta_{N,T_N})^\top$ be the concatenation of all couples' parameters at all time-steps into a column vector. Let $\mathbf{k}_i$ be the covariance matrix for couple i obtained by evaluating the covariance function k at every pair of time inputs. Then,

$$\boldsymbol{\theta}|\mathbf{q}, \sigma, s \sim \mathcal{N}(\boldsymbol{\Omega}^\dagger\mathbf{q}, \sigma^2\mathbf{K}), \qquad (20)$$

where

$$\boldsymbol{\Omega}^\dagger = \begin{pmatrix} \boldsymbol{\omega}_1^\top \otimes \mathbf{1}_{T_1} \\ \boldsymbol{\omega}_2^\top \otimes \mathbf{1}_{T_2} \\ \cdots \\ \boldsymbol{\omega}_N^\top \otimes \mathbf{1}_{T_N} \end{pmatrix}, \qquad \mathbf{K} = \begin{pmatrix} \mathbf{k}_1 & & \\ & \ddots & \\ & & \mathbf{k}_N \end{pmatrix}.$$

Here, $\mathbf{1}_k$ is a column vector with k ones, T_i is the time length of couple i, and $\otimes$ is the Kronecker product.

Using the same conjugate $\mathcal{NIG}$ prior as in section 4.1, we have $\mathbf{q}, \sigma^2|\boldsymbol{\theta}, s \sim \mathcal{NIG}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*, a_*, b_*)$, where

$$\boldsymbol{\mu}_* = \left(\boldsymbol{\Omega}^{\dagger\top}\mathbf{K}^{-1}\boldsymbol{\Omega}^\dagger + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\Omega}^{\dagger\top}\mathbf{K}^{-1}\boldsymbol{\theta}\right), \tag{21}$$

$$\boldsymbol{\Sigma}_* = \left(\boldsymbol{\Omega}^{\dagger\top}\mathbf{K}^{-1}\boldsymbol{\Omega}^\dagger + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}, \qquad (22)$$

$$a_* = \left(2a_0 + \sum_{i=1}^N T_i\right)/2, \qquad (23)$$

$$b_* = b_0 + \left(\boldsymbol{\theta}^\top\mathbf{K}^{-1}\boldsymbol{\theta} + \boldsymbol{\mu}_0\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 - \boldsymbol{\mu}_*\boldsymbol{\Sigma}_*^{-1}\boldsymbol{\mu}_*\right)/2. \quad (24)$$

5. Model evaluation

We develop a multi-stage evaluation procedure to learn the shared model parameters $\mathbf{Q}$, $\boldsymbol{\sigma}$ and $\mathbf{s}$ and evaluate the predictive power of the learned models using 9-fold cross validation. First, we randomly divide the couples into nine groups. We get the MAP estimates of the shared parameters using the couples in eight out of the nine groups by algorithms developed in section 4. We then estimate the predictive errors using the couples in the nineth group by integrating out couple-specific parameters $\boldsymbol{\Theta}$ and the initial values $\mathbf{X}_1$. In particular, we use 100 samples from the posterior for the first 80% of time to compute 100 estimates for each time point in the next 20% (held out) of time for each testing couple. We use these estimates to provide a Monte Carlo estimate of the expected squared error with respect to the posterior distribution, and report the square root as RMSE. In the non-drift case, the prediction into the last 20% for each estimate is simply from the time evolution of the ODE and initial values. In the drift case, we

use the samples of the CLO parameters for the first 80% to compute the mean of the GP predictive distribution for the CLO parameters. We then evolve that ODE into the last 20% for prediction.

Specifically, RMSE of the predictions of the last 20% time points (t' to T) for couple i with observations $\mathbf{y}_{i,t':T}$ is:

$$\text{RMSE}(\mathbf{y}_{i,t':T}) = \sqrt{\sum_{t=t'}^{T} (\text{E}[f(t; \mathbf{x}_{i1}, \boldsymbol{\theta}_{it})] - \mathbf{y}_{it})^2 / T},$$

where $\text{E}[f(t; \mathbf{x}_{i1}, \boldsymbol{\theta}_{it})]$ is the expected value of $\mathbf{y}_{it}$ by involving ODE from time 1 forward to time t by equation 3 with samples of initial state $\mathbf{x}_{i1}$ and CLO parameters $\boldsymbol{\theta}_{it}$. We then repeat this step nine times by using each group as a testing data in turn and learn the shared parameters in the rest corresponding eight groups. We record the fitting and predicting RMSE of each couple in the testing group of each fold and compute the mean of the errors and the standard error of the mean.

6. Experiments

We tested our models and inference processes both on synthetic data generated from the model and real data that collected in a social psychology experiment.

Experiments with synthetic data. The synthetic data is composed of 50 couples with observations at 50 time points for each. The observations of each couple were generated by using ancestral sampling from the graphical model defined in Fig. 2. Specifically, the CLO parameters of all the couples share the same prior distribution defined by $\boldsymbol{\alpha}$, $\boldsymbol{\sigma}$, which were generated from $\mathcal{NIG}(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0, a_0, b_0)$, and $\mathbf{s}$, which were generated from $\mathcal{IG}(\mathbf{a}_s, \mathbf{b}_s)$. The noise sigma in both data generation and fitting was 0.5. The number of sampling iterations was 30,000 for parameter learning and 100,000 for parameter fitting. For model comparison, we provide the fitting and prediction errors for three base line models: predicting the last 20% of each couple by **1)** a line that fitted to each partner's first 80% observations; **2)** the average of the first 80% observations; and **3)** a fitted CLO model by maximum likelihood estimation (MLE) of $p(\mathbf{Y}|\boldsymbol{\Theta})$ over the first 80% time points.

The results show that the hierarchical model with learned shared parameters $\boldsymbol{\alpha}$, $\boldsymbol{\sigma}$, and $\mathbf{s}$ can predict much better than MLE estimation of the parameters from observations.

Experiments with real data. The real data is composed of recalled self-rating emotion experience of 38 heterosexual couples with different joint weight status during an emotional conversation in a social experiment lab setting as reported by Reed *et al.* (2015). During that experiment, couples were engaged in a video-taped conversation of up to 20 minutes on the importance of living a healthy lifestyle

Table 1. RMSE of fitting and predicting on synthetic data. The average of both partner's RMSE is reported $\pm$ the error in the estimate. The first three rows are results of three base line models: average, line, and maximum likelihood estimation CLO model. Results in row with $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\alpha}, \boldsymbol{\sigma})$ is obtained by treating $\boldsymbol{\theta}$ constant and learning the couple-shared $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}$ from sampling $p(\boldsymbol{\alpha}, \boldsymbol{\sigma}, \boldsymbol{\theta}, \mathbf{x}_1 | \mathbf{y})$. Results in row with $\boldsymbol{\theta} \sim \mathcal{GP}(m_\alpha, k_{\boldsymbol{\sigma}, \mathbf{s}})$ is obtained by treating $\boldsymbol{\theta} \sim \mathcal{GP}(m(\mathbf{t}), k(\mathbf{t}, \mathbf{t}'))$, in which $m(\mathbf{t}) = \boldsymbol{\alpha}$, and $k(t, t') = \sigma^2 \exp(-\frac{(t-t')^2}{2s^2})$. The values of $\boldsymbol{\alpha}$, $\boldsymbol{\sigma}$, and $\mathbf{s}$ are learned by sampling from the joint distribution $p(\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{s}, \mathbf{x}_1 | \mathbf{y})$. Finally, the last row is obtained by using the optimal scale parameter $\mathbf{s}^*$ in data generation and learning the values of $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}$ from the MAP of $p(\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{x}_1 | \mathbf{s}^*, \mathbf{y})$.

MODEL	FITTING	PREDICTING
AVERAGE	0.72 ± 0.04	0.90 ± 0.09
LINE	0.62 ± 0.02	1.12 ± 0.14
MLE-CLO	0.51 ± 0.01	1.05 ± 0.15
$\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\alpha}, \boldsymbol{\sigma})$	0.57 ± 0.02	0.86 ± 0.12
$\boldsymbol{\theta} \sim \mathcal{GP}(m_\alpha, k_{\boldsymbol{\sigma}, \mathbf{s}})$	0.52 ± 0.08	0.80 ± 0.10
$\boldsymbol{\theta} \sim \mathcal{GP}(m_\alpha, k_{\boldsymbol{\sigma}, \mathbf{s}^*})$	0.48 ± 0.01	0.76 ± 0.07

and the positive and negative impacts they have on each other's health behaviors. Right after the conversation, both partners were asked to watch their own video recordings separately and rate how they remembered feeling during the conversation using a rating dial. In particular, the couples were asked to rate their emotions continuously second by second through the whole conversation. Following Reed *et al.*, we averaged this data into 10 second intervals to reduce noise. The number of resulting time points ranges from 5 to 99, with 48 being the average. The rating dial measurements ranged from -2.0 (very negative experience) to 2.0 (very positive experience). We somewhat arbitrarily set 0.5 as the sigma of the observation noise during both parameter learning and fitting.

Table 2 shows the mean and the standard deviation of RMSE of the three base line models on this data set. Table 3 shows the results of the stationary CLO with priors defined by group-shared parameters with different combinations of moderators. Finally, table 4 shows the results for the time-varying CLO models.

Table 2. RMSE of fitting and predicting on real emotion data of three baseline models $\pm$ the error in the estimate. Not surprisingly, the fitting error of MLE-CLO model has the smallest fitting error and the largest predicting error due to over-fitting.

MODEL	FITTING	PREDICTING
AVERAGE	0.52 ± 0.03	0.70 ± 0.05
LINE	0.47 ± 0.03	0.76 ± 0.08
MLE-CLO	0.39 ± 0.03	1.10 ± 0.19

Table 3. RMSE of fitting and predicting of stationary CLO with group-shared priors $\pm$ the error in the estimate. The model with $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\alpha}, \boldsymbol{\sigma})$ has all couples share the same mean of the CLO parameters. For subsequent models, the prior distribution of $\boldsymbol{\theta}$ for the CLO model depends on the couples' average BMI (second row), difference in BMI (third row), and both average and difference (last row).

MODEL	FITTING	PREDICTING
$\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\alpha}, \boldsymbol{\sigma})$	0.44 ± 0.03	0.68 ± 0.06
$\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\alpha} + \bar{\omega}\boldsymbol{\beta}, \boldsymbol{\sigma})$	0.42 ± 0.03	0.64 ± 0.06
$\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\alpha} + \hat{\omega}\boldsymbol{\gamma}, \boldsymbol{\sigma})$	0.43 ± 0.03	0.61 ± 0.05
$\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\alpha} + \bar{\omega}\boldsymbol{\beta} + \hat{\omega}\boldsymbol{\gamma}, \boldsymbol{\sigma})$	0.41 ± 0.02	0.59 ± 0.05

Table 4. RMSE of fitting and predicting of time-varying CLO parameters with group-shared GP priors $\pm$ the error in the estimate. The mean function of the GP is shared across all couples for the model in the first row. For subsequent models, the mean function depends on couples' average BMI (second row), difference in BMI (third row), and both average and difference (last row).

MODEL	FITTING	PREDICTING
$\boldsymbol{\theta} \sim \mathcal{GP}(m_{\boldsymbol{\alpha}}, k_{\boldsymbol{\sigma}, \mathbf{s}})$	0.46 ± 0.03	0.62 ± 0.06
$\boldsymbol{\theta} \sim \mathcal{GP}(m_{\boldsymbol{\alpha} + \bar{\omega}\boldsymbol{\beta}}, k_{\boldsymbol{\sigma}, \mathbf{s}})$	0.46 ± 0.03	0.61 ± 0.06
$\boldsymbol{\theta} \sim \mathcal{GP}(m_{\boldsymbol{\alpha} + \hat{\omega}\boldsymbol{\gamma}}, k_{\boldsymbol{\sigma}, \mathbf{s}})$	0.45 ± 0.03	0.61 ± 0.06
$\boldsymbol{\theta} \sim \mathcal{GP}(m_{\boldsymbol{\alpha} + \bar{\omega}\boldsymbol{\beta} + \hat{\omega}\boldsymbol{\gamma}}, k_{\boldsymbol{\sigma}, \mathbf{s}})$	0.46 ± 0.03	0.62 ± 0.06

The results in Table 3 show that the group-shared CLO model, where emotion patterns come from a distribution learned from the training couples, better predicts the emotion states of unseen couples than individual CLO models fit to the observations. Furthermore, the BMI moderator improved prediction accuracy to some extent, which supports the findings suggested in (Reed et al., 2015).

To see how couples with different BMI values exhibit different emotional dynamic patterns, we plot the predicted emotional state under different BMI moderators (Figure 3). We can see that couples in which the woman is heavier than the man may be emotionally disconnected and the woman may show increasingly volatile emotional responses. This could contribute to poorly modulated emotion responses when discussing health and lifestyle choice.

When we let the CLO parameters of the real data drift over time, it improves prediction accuracy, reducing the RMSE from 0.68 to 0.62 when moderators are not present. However, when the moderators are used, we did not observe much difference in prediction accuracy. One possible explanation is that the participants did not have significant changes in their emotion dynamics during the short conversation of the experiment, at least to the extent that can be extracted from this data which is very noisy. Alternatively, the effect of drift could be correlated with the moderator.
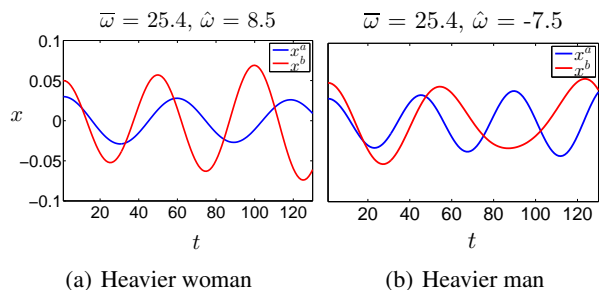

(a) Heavier woman (b) Heavier man

Figure 3. Visualizing the predicted emotional states over time with the same initial state but differing BMI moderator values. x^{a} is the emotional state for the male partner, and x^{b} is for the female partner. Figure (a) shows the emotional states for couples in which the woman is heavier than the man. Figure (b) shows the emotional states for couples in which the man is heavier than the woman.

7. Discussion and Conclusion

We have developed models and inference methods to endow linear dynamical systems with parameter drift, and parameters that are moderated by other variables. We focus on the case where it is assumed that the parameters are scientifically interesting, and thus we purposely exclude latent state perturbations as is more common. While including latent state perturbation in the models is not difficult, doing so would relax the pressure to fit the parameters, which might be troublesome with data sets where there is significant unexplained variance such as ours. We have applied our modeling and inference approach to a challenging social interaction data set, and were able to confirm a previous result which was based on regression methods. By contrast to the approach taken in that work, which is standard in the social sciences, we developed a cross-validation approach around the ability of the models to predict conversation patterns into the near future.

There is much left to do. Currently our learning algorithms are slow — on the order of hours for the 36 couple data set due to the large number of samples that are needed for learning. Further, we have only explored coupled oscillators in detail. We are planning to consider other LDS models on other data such, as well as modeling triads (e.g., mom, dad, child). However, our methods are now at the state where we can provide them to the social science community via web services (Predoehl et al., 2015).

Acknowledgments

References

A. S. Willsky, E. B. Sudderth, M. I. Jordan, and E. B. Fox. Nonparametric Bayesian learning of switching linear dynamical systems. In *Advances in Neural Information Processing Systems*, pp. 457–464, 2009.

Alvarez, M. A., Luengo, D., and Lawrence, N. D. Latent force models. In *International Conference on Artificial Intelligence and Statistics*, pp. 9–16, 2009.

Anger, G. *Inverse Problems in Differential Equations*. Springer, 1990.

Bard, Y. *Nonlinear Parameter Estimation*. Academic Press, 4th edition, 1973.

Boker, S., Neale, M., and Rausch, J. Latent differential equation modeling with multivariate multi-occasion indicators. In *Recent Developments on Structural Equation Models*, volume 19 of *Mathematical Modelling: Theory and Applications*, pp. 151–174. Springer Netherlands, 2004.

Boker, S. M. and Ghisletta, P. Random coefficients models for control parameters in dynamical systems. *Multilevel Modelling Newsletter*, 13(1):10–17, 2001.

Boker, S. M. and Laurenceau, J. P. Dynamical systems modeling: An application to the regulation of intimacy and disclosure in marriage. In *Models for intensive longitudinal data*, pp. 195–218. Oxford University Press, New York, 2006.

Boker, S. M. and Laurenceau, J. P. Coupled dynamics and mutually adaptive context. *Lawrence Erlbaum Associates Publishers*, pp. 299–324, 2007.

Butner, J., Amazeen, P. G., and Mulvey, G. M. Multilevel modeling of two cyclical processes: Extending differential structural equation modeling to nonlinear coupled systems. *Psychological Methods*, 10(2):159–177, 2005.

Calderhead, B., Girolami, M., and Lawrence, N. D. Accelerating Bayesian inference over nonlinear differential equations with Gaussian Processes. In *Advances in Neural Information Processing Systems 21*, pp. 217–224. Curran Associates, Inc., 2009.

Chkrebtii, O. A., Campbell, D. A., Girolami, M. A., and Calderhead, B. Bayesian uncertainty quantification for differential equations. *arXiv:1306.2365v2*, 2013.

Chow, S. M., Ram, N., Boker, S. M., Fujita, F., and Clore, G. Emotion as a thermostat: Representing emotion regulation using a damped oscillator model. *Emotion*, 5(2): 208–225, 2005.

Dattner, I. and Klaassen, C. AJ. Estimation in systems of ordinary differential equations linear in the parameters. *arXiv:1305.4126*, 2013.

Davidian, M. and Giltinan, D. M. Nonlinear models for repeated measurement data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics*, 8(4):387–419, 2003.

Dondelinger, F., Husmeier, D., Rogers, S., and Filippone, M. ODE parameter inference using adaptive gradient matching with Gaussian processes. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, pp. 216–228, 2013.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. Hybrid monte carlo. *Physics letters B*, 195(2):216–222, 1987.

Ferrer, E. and Helm, J. L. Dynamical systems modeling of physiological coregulation in dyadic interactions. *International Journal of Psychophysiology*, 88(3):296–308, 2013.

Ferrer, E. and Steele, J. S. Dynamic systems analysis of affective processes in dyadic interactions using differential equations. In *Advances in longitudinal methods in the social and behavioral sciences*. Information Age Publishing, 2011.

Finan, P. H., Hessler, E. E., Amazeen, P. G., Butner, J., Zautra, A. J., and Tennen, H. Nonlinear oscillations in pain prediction accuracy: A dynamical systems approach to understanding daily pain prediction. *Nonlinear dynamics, psychology, and life sciences*, 14(1):27, 2010.

Gao, P., Honkela, A., Rattray, M., and Lawrence, N. D. Gaussian process modelling of latent chemical species: applications to inferring transcription factor activities. *Bioinformatics*, 24(16):70–75, 2008.

Gelman, A., Bois, F., and Jiang, J. Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association*, 91:1400–1412, 1996.

Hall, P. and Ma, Y. Quick and easy one-step parameter estimation in differential equations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2013.

Lawrence, N. D, Sanguinetti, G., and Rattray, M. Modelling transcriptional regulation using Gaussian Processes. In *Advances in Neural Information Processing Systems*, pp. 785–792, 2006.

Li, L., Lin, X., Brown, M. B., Gupta, S., and Lee, K. A population pharmacokinetic model with time-dependent

covariates measured with errors. *Biometrics*, 60(2):451–460, 2004.

Li, L., Yu, M., Chin, R., Lucksiri, A., Flockhart, D. A., and Hall, S. D. Drug–drug interaction prediction: a Bayesian meta-analysis approach. *Statistics in Medicine*, 26(20): 3700–3721, 2007.

Li, Z., Osborne, M. R., and Trvan, T. Parameter estimation of ordinary differential equations. *IMA Journal of Numerical Analysis*, 25(2):264–285, 2005.

Liang, H. and Wu, H. Parameter estimation for differential equation models using a framework of measurement error in regression models. *Journal of the American Statistical Association*, 103(484):1570–1583, 2008.

Murphy, K. P. 18. state space models. In *Machine Learning: a Probabilistic Perspective*, Adaptive computation and machine learning series, pp. 631–660. The MIT Press, 2012. ISBN 0262018020.

Neal, R. M. Monte Carlo implementation of Gaussian Process models for Bayesian regression and classification. Technical Report 9702, Department of Statistics, University of Toronto, 1997.

Nise, N. S. The general state-space representation. In *Control Systems Engineering*, chapter 3.3, pp. 123–124. John Wiley & Sons, Inc., 6 edition, 2010.

Pillai, G., Mentré, F., and Steimer, J. Non-linear mixed effects modeling – from methodology and software development to driving implementation in drug development science. *Journal of Pharmacokinetics and Pharmacodynamics*, 32(2):161–183, 2005.

Predoehl, Andrew, Guan, Jinyan, Butler, Emily, and Barnard, Kobus. CompTies web app, 2015. URL http://www.compties.org/COM.html.

Ramsay, J. O., Hooker, G., Campbell, D., and Cao, J. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5): 741–796, 2007.

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

Reed, Rebecca, Barnard, Kobus, and Butler, Emily. Distinguishing emotional co-regulation from co-dysregulation: An investigation of emotional dynamics and body-weight in romantic couples. *Emotion*, 15(1):45–60, 2015.

S. M. Oh, J. M. Rehg, T. Balch, and F. Dellaert. Learning and inferring motion patterns using parametric segmental switching linear dynamic systems. *International Journal of Computer Vision*, 77(1-3):103–124, 2008.

Steele, J. S. and Ferrer, E. Latent differential equation modeling of self-regulatory and coregulatory affective processes. *Multivariate Behavioral Research*, 46(6):956–984, 2011.

Titsias, M. K., Lawrence, N. D., and Rattray, M. Efficient sampling for Gaussian Process inference using control variables. In *Advances in Neural Information Processing Systems*, volume 21, pp. 1681–1688, Vancouver, British Columbia, Canada, 2008. Curran Associates, Inc.

Varah, J. M. A spline least squares method for numerical parameter estimation in differential equations. *SIAM Journal on Scientific and Statistical Computing*, 3(1): 28–46, 1982.

Wang, Y. and Barber, D. Gaussian Processes for Bayesian estimation in ordinary differential equations. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, Beijing, 2014. JMLR Workshop and Conference Proceedings.

Yu, M., Kim, S., Wang, Z., Hall, S., and Li, L. A Bayesian meta-analysis on published sample mean and variance pharmacokinetic data with application to drug–drug interaction prediction. *Journal of Biopharmaceutical Statistics*, 18(6):1063–1083, 2008.