Gaussian Conditional Random Fields for Modeling Patients' Response to Acute Inflammation Treatment

Vladan Radosavljevic[a*] VLADAN@TEMPLE.EDU
Kosta Ristovski[a*] KOSTA.RISTOVSKI@TEMPLE.EDU
Zoran Obradovic[a] ZORAN.OBRADOVIC@TEMPLE.EDU
[a]Temple University, 1805 N. Broad St., Philadelphia, PA 19122 [*]These authors contributed equally

Abstract

Acute inflammation, a medical condition characterized by a systemic inflammatory response to an infection, is the leading cause of death in non-surgical hospitalized patients. Early diagnosis and appropriate treatment can significantly improve the odds of survival. However, this is a challenging objective due to a very fast progression that requires quick and continuous patient-specific decisions. To aid clinicians in assessing the benefits of different treatment strategies, we propose the Gaussian Conditional Random Field (GCRF) approach for predicting a patient's response to therapy. GCRF is a data-driven method that learns patient behavior from historical data. Benefits of GCRF include modeling correlation among outputs, quantifying the uncertainty of predictions, and integrating conventional and statistical methods. Here we provide evidence that application of GCRF results in better predictive power than alternatives when applied to modeling patients' responses to therapy for acute inflammation.

1. Introduction

Planning effective personalized therapeutic strategies for life-threatening conditions is one of the major challenges in medical practice. It is especially critical in rapid progression medical conditions like acute inflammation, a systematic inflammatory response syndrome triggered by infection. A widely used treatment for acute inflammation consists of broad-spectrum antibiotics and/or intravenous fluids where doses are adjusted manually based on clinicians' experience of the benefit of treatment. However, such a therapeutic regimen is often inadequate, as more personalized therapy would be far more effective. Inadequate treatment results in a mortality rate of 30-35% and for every hour that the administration of appropriate therapy is delayed, the mortality rate increases by about 7% (Thiel et al., 2010). Therefore, there is a critical need for tools that can aid clinicians in assessing the benefits of various patient-specific treatment strategies that will maximize the probability of treatment success.

Predictive models developed to predict a patient's response to the treatment can be used to assess the benefits of various treatment strategies (Figure 1). To construct an accurate predictive model, practitioners often rely on domain-based assumptions about patients' behavior. Such a domain-driven predictive model was previously used in acute inflammation treatment (Clermont et al., 2010) where patient's response to two medications was modeled by a set of ordinary differential equations. An alternative would be to learn the predictive model directly from historical data, without making any domain-based assumptions. Such data-driven models have been utilized as predictive models for medical applications, including regulation of glucose supply (Wang et al., 2009), an exploration of optimal dosing of anticancer agents (Noble et al., 2010), and defining an optimal anesthesia (Yelneedi et al., 2009).

To successfully implement a data-driven model in applications related to medical decisions, it is necessary to consider some specific requirements that were not addressed by any of the above-mentioned approaches. Before a data-driven model is accepted by the medical community, it is imperative to quantify the uncertainty of its predictions. The uncertainty in predictions gives clinicians sufficient confidence to put the new method into practice. Also, the predictive model has to be able

Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013. JMLR: W&CP volume 28.

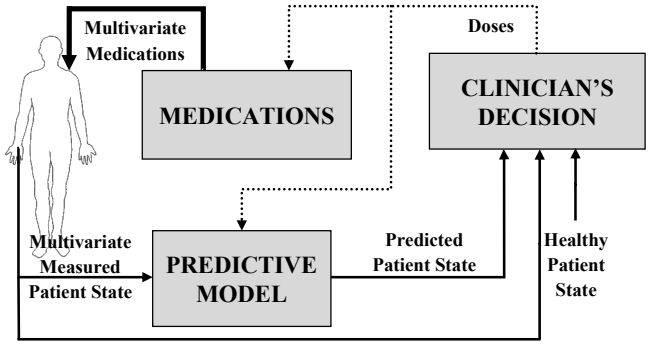

Figure 1. Predictive model for assisting clinician decisions.

to make accurate predictions of the future patient's behavior not only just one step ahead but also several steps ahead in order to give clinicians opportunity to fully assess the therapy benefit.

To meet clinician's requirements, we propose the Gaussian Conditional Random Field (GCRF) approach for predicting a patient's response to therapy. As in (Clermont et al., 2010), all evaluations in this work are performed on virtual patients generated by a mathematical model that emulates inflammatory response. Our results show significant improvement in prediction accuracy and uncertainty estimations comparing to Gaussian process (GP) and linear autoregressive exogenous model (ARX).

2. Virtual Patient Model

To significantly reduce the chance of a clinical failure and to save on the costs of clinical trials, biomedical researchers use computer simulations of body processes (often called virtual patients) to perform preliminary tests of hypotheses before they prove them in real patient studies. Virtual patients are generated using a carefully determined mathematical model to simulate the process of interest. A significant advantage of having a virtual patient model for experiments is the possibility of testing different approaches on the same virtual patient and comparing the outcomes. We will use the mathematical model for inflammatory response recently proposed in (Clermont et al., 2010) that is capable of simulating:

- an evolution of a bacterial pathogen population (P) that initiates the cascade of inflammation,
- dynamics of pro-inflammatory mediators (N),
- markers of tissue damage/dysfunction (D),
- anti-inflammatory mediators evolution (CA),

which are controlled by doses of pro-inflammatory ($PIDOSE$) and anti-inflammatory ($AIDOSE$) therapies. This mathematical model is based on the system of ordinary differential equations (ODE)

$$\begin{aligned}
\frac{dP}{dt} &= k_{pg}\left(1-\frac{P}{P_\infty}\right)-\frac{k_{pm}s_mP}{\mu_m+k_{mp}P}- \\
&\quad -k_{pn}f(N)P, \qquad (1)\\
\frac{dN}{dt} &= \frac{s_{nr}R}{\mu_{nr}+R}-\mu N+PIDOSE(t), \qquad (2)\\
\frac{dD}{dt} &= \frac{k_{dn}f(N)^6}{x_{dn}^6+f(N)^6}-\mu_dD, \qquad (3)\\
\frac{dCA}{dt} &= s_c+\frac{k_{cn}f(N+k_{cnd}D)}{1+f(N+k_{cnd}D)}-\mu_cCA+ \\
&\quad +AIDOSE(t), \qquad (4)
\end{aligned}$$

where

$$R=f(k_{np}P+k_{nn}N+k_{nd}D), f(x)=\frac{x}{1+\left(\frac{CA}{c_\infty}\right)^2}. \quad (5)$$

All variables used in the mathematical model except patient state $[P \quad N \quad D \quad CA]$ are parameters with valid ranges specified in (Clermont et al., 2010). Although conceptual, ODE is capable of modeling the complex effect of pathogen (P) on the patient. An increase of pathogen level P leads to the series of positive and negative feedback reactions that are all successfully modeled by ODE. In particular, an increase of P leads to the development of a pro-inflammatory response, which causes an increase of N in (2) and to the development of tissue damage, which causes an increase of D in (3). Equation (1) simulates a positive effect of inflammation where an increase of N reduces level of pathogen P. However, (3) simulates a negative effect of inflammation where an increase of N further damages tissue causing rapid increase of D. An increase of D mobilizes a negative feedback in (4), or anti-inflammatory response (CA), which lowers level of N and inhibits damage to tissue (decrease of D) (Clermont et al., 2010). The strength of positive and negative feedbacks depends on the parameter values in ODE. By varying parameter values we can simulate variability among patients.

Variability in the population of virtual patients is obtained by random initialization of three parameters in ODE k_{pg}, k_{cn}, and k_{nd} and by random initialization of the initial conditions P_0 and CA_0 from uniform distribution on valid ranges ($k_{pg} \in [0.3, 0.6], k_{cn} \in [0.03, 0.05], k_{nd} \in [0.015, 0.025], P_0 \in [0, 1], CA_0 \in [0.0938, 0.1563]$). All other parameters were fixed to referent values as in (Clermont et al., 2010) except k_{cnd} that covaries with k_{cn} and k_{np} that covaries with

k_{nd} (Clermont et al., 2010). In all of the simulations, t is an hourly step that starts from $t = 0$ when patient state and parameters are initialized. Then, patient state evolves according to ODE through the simulation time of 168 hours (one week).

3. Predictive Model

Virtual patient state is represented by the four outputs P, N, D, and CA. To predict four-dimensional state using ARX model that we proposed in (Radosavljevic et al., 2012) we split the predictive model into four submodels. Each of submodels was responsible for prediction of one of the outputs P, N, D, and CA, keeping the same set of inputs for each of the sub-models. If we denote $\mathbf{s}_t$ and $\mathbf{u}_t$ to represent patient's state and control signal (medication doses) at time point t respectively, it can be written

$$\mathbf{s}_t = (P_t, N_t, D_t, CA_t)^T, \quad (6)$$

$$\mathbf{u}_t = (AIDOSE_t, PIDOSE_t)^T. \quad (7)$$

In the ARX representation, the output of one submodel $y_t \in \mathbf{s}_t$ at time step t depends on the delayed outputs and the exogenous control inputs $\mathbf{u}$ as

$$y_t = f(\mathbf{s}_{t-1}, ..., \mathbf{s}_{t-n_y}, \mathbf{u}_{t-1}, ..., \mathbf{u}_{t-n_u}) + \epsilon \quad (8)$$

where f denotes a function and ϵ is noise term. As we can see from (8), output y_t depends on the input vector $\mathbf{x}_t = [\mathbf{s}_{t-1}, ..., \mathbf{s}_{t-n_y}, \mathbf{u}_{t-1}, ..., \mathbf{u}_{t-n_u}]^T$, where n_y and n_u are delays for state and control signals respectively. Assuming that control signal is known up to time t, the goal is to predict the output of the system T steps ahead, i.e., we need to find the predictive distribution of y_{t+p}, $p = 1 \dots T$. Multiple-step-ahead predictions of a system modeled by (8) can be achieved by iteratively making repeated one-step-ahead predictions, up to the desired horizon T. Disadvantage of this method is that it predicts outputs independently, not taking into account temporal correlation among outputs, which decreases accuracy due to error propagation. Thus, we developed Gaussian Conditional Random Fields as predictive model, which is capable of modeling correlations among outputs over the horizon T.

3.1. Gaussian Conditional Random Fields

Gaussian conditional random fields are developed based on idea of Continuous conditional random fields(CCRF) that are used to model conditional distribution $P(\mathbf{y}|X)$ over all outputs $\mathbf{y}$ given all inputs X (Qin et al., 2008), as

$$P(\mathbf{y}|X) = \frac{1}{Z(X, \boldsymbol{\alpha}, \boldsymbol{\beta})} \exp\left(\phi(\mathbf{y}, X, \boldsymbol{\alpha}, \boldsymbol{\beta})\right), \quad (9)$$

where the term in the exponent $\phi(\mathbf{y}, X, \boldsymbol{\alpha}, \boldsymbol{\beta})$, and normalization constant $Z(X, \boldsymbol{\alpha}, \boldsymbol{\beta})$ are defined as

$$\begin{aligned} \phi(\mathbf{y}, X, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \sum_{i=1}^{N} A(\boldsymbol{\alpha}, y_i, X) + \sum_{j \sim i} I(\boldsymbol{\beta}, y_i, y_j, x), \\ Z(X, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \int_{\mathbf{y}} \exp\left(\phi(\mathbf{y}, X, \boldsymbol{\alpha}, \boldsymbol{\beta})\right) d\mathbf{y}. \end{aligned} \quad (10)$$

The output y_i is associated with inputs X by a real-valued function called the association potential $A(\boldsymbol{\alpha}, y_i, X)$, where $\boldsymbol{\alpha}$ is a K-dimensional set of parameters. In general, A takes as input X, which could be any useful combination of explanatory variables from data set D. To model interactions between outputs $y_i \sim y_j$, a real valued function called the interaction potential $I(\boldsymbol{\beta}, y_i, y_j, X)$ is used, where $\boldsymbol{\beta}$ is an L dimensional set of parameters.

In CCRF applications, A and I are often conveniently defined as linear combinations of a set of feature functions f and g in terms of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (Lafferty & Pereira, 2001),

$$\begin{aligned} A(\boldsymbol{\alpha}, y_i, X) &= \sum_{k=1}^{K} \boldsymbol{\alpha}_k f_k(y_i, X), \\ I(\boldsymbol{\beta}, y_i, y_j, x) &= \sum_{l=1}^{L} \boldsymbol{\beta}_l g_l(y_i, y_j, X). \end{aligned} \quad (11)$$

The use of feature functions is convenient because it allows modeling of arbitrary relationships between inputs and outputs. In this way, any potentially relevant feature function could be included to the model and its degree of relevance would be determined automatically by the learning algorithm.

3.1.1. Feature Functions

Construction of appropriate feature functions in CRF is a manual process that depends on prior beliefs of a practitioner about what features could be useful. The choice of features is often constrained to reduce the complexity of learning and inference from CRF. In general, to evaluate $P(\mathbf{y}|X)$ during learning and inference, one would need to use time consuming sampling methods (Xin et al., 2009). However, if A and I are defined as quadratic functions of $\mathbf{y}$, learning and inference can be accomplished in a computationally efficient manner (Qin et al., 2008). Let us assume we are given K unstructured models, $R_k(X), k = 1, \dots K$, that predict single output y_i taking into account X (as a special case, only x_i can be used as X). The quadratic feature functions for the association potential can be written as

$$f_k(y_i, X) = -(y_i - R_k(X))^2, \quad k = 1, \dots K. \quad (12)$$

These feature functions follow the basic principle for association potentials in that their values are large when predictions and outputs are similar. To model the correlation among outputs, we introduce the quadratic feature functions for the interaction potential as

$$g_l(y_i, y_j, X) = -S_{ij}^{(l)}(y_i - y_j)^2, \quad l = 1, \dots L, \quad (13)$$

that imposes that outputs y_i and y_j have similar values if they are similar in user defined measure $S_{ij}^{(l)}$.

3.1.2. Gaussian Canonical Form

In this section, we show that $P(\mathbf{y}|X)$ for CRF model, which uses quadratic feature functions, can be represented as a multivariate Gaussian distribution. The resulting CRF model can be written as

$$\begin{aligned} P(\mathbf{y}|X) = &\frac{1}{Z}\exp\Big(-\sum_{i=1}^{N}\sum_{k=1}^{K}\alpha_k(y_i - R_k(X))^2 - \\ &- \sum_{i\sim j}\sum_{l=1}^{L}\beta_l S_{ij}^{(l)}(y_i - y_j)^2\Big). \end{aligned} \quad (14)$$

The exponent in (14), which we denote as E, is a quadratic function in terms of $\mathbf{y}$. Therefore, $P(\mathbf{y}|X)$ can be transformed to form of Gaussian distribution by representing E in the information form

$$E = -\frac{1}{2}(\mathbf{y}^T Q \mathbf{y}) + \mathbf{y}^T \mathbf{b} + const, \quad (15)$$

where Q and $\mathbf{b}$ are canonical parameters of Gaussian distribution. We found by matching (14) and (15)

$$Q = 2(Q_1 + Q_2), \quad (16)$$

$$Q_{1ij} = \begin{cases} \sum_{k=1}^{K}\alpha_k, & i = j \\ 0, & i \neq j \end{cases}, \quad (17)$$

$$Q_{2ij} = \begin{cases} \sum_k \sum_{l=1}^{L}\beta_l S_{ik}^{(l)}, & i = j \\ -\sum_{l=1}^{L}\beta_l S_{ik}^{(l)}, & i \neq j \end{cases}, \quad (18)$$

$$b_i = 2\sum_{k=1}^{K}\alpha_k R_k(x). \quad (19)$$

As the resulting conditional distribution is Gaussian, we call resulting CRF *the Gaussian* CRF (GCRF).

3.1.3. Learning and inference

The learning task is to choose α and β to maximize the conditional log-likelihood,

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \underbrace{\arg\max}_{\boldsymbol{\alpha},\boldsymbol{\beta}} L(\boldsymbol{\alpha}, \boldsymbol{\beta}), \;\; L = \log P(\mathbf{y}|X). \quad (20)$$

To have a feasible model with real valued outputs, Z must be integrable, which is ensured by the constraint that all elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are greater than 0. In this setting, learning is a constrained optimization problem. To convert it to the unconstrained optimization, we adopt a technique used in (Qin et al., 2008) that applies the exponential transformation of the parameters to guarantee that they are positive. All parameters are learned by the gradient-based optimization. To apply it, we need to find the gradient of the conditional log-likelihood.

$$\begin{aligned} \frac{\partial P}{\partial \alpha_k} = &-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \frac{\partial Q}{\partial \alpha_k}(\mathbf{y} - \boldsymbol{\mu}) + \\ &(\frac{\partial b}{\partial \alpha_k} - \boldsymbol{\mu}^T \frac{\partial Q}{\partial \alpha_k})(\mathbf{y} - \boldsymbol{\mu}) + \frac{1}{2}Tr(Q^{-1}\frac{\partial Q}{\partial \alpha_k}), \end{aligned} \quad (21)$$

$$\frac{\partial P}{\partial \beta_k} = -\frac{1}{2}(\mathbf{y} + \boldsymbol{\mu})^T \frac{\partial Q}{\partial \beta_k}(\mathbf{y} - \boldsymbol{\mu}) + \frac{1}{2}Tr(Q^{-1}\frac{\partial Q}{\partial \beta_k}), \quad (22)$$

$$\boldsymbol{\mu} = Q^{-1}\mathbf{b}. \quad (23)$$

The inference task is to find the outputs $\mathbf{y}$ for a given inputs X, such that the conditional probability $P(\mathbf{y}|X)$ is maximized. The GCRF model is Gaussian and, therefore, the maximum a posteriori estimate of $\mathbf{y}$ is obtained as the expected value $\boldsymbol{\mu}$ of the GCRF distribution,

$$\mathbf{y}_* = \underbrace{\arg\max}_{y} P(y|X) = \boldsymbol{\mu}. \quad (24)$$

Uncertainty for each output can be taken as corresponding element from the diagonal of covariance matrix in $P(\mathbf{y}|X)$.

3.2. Temporal Form of GCRF

Multiple-step-ahead predictions over horizon T can be achieved by GCRF defined as

$$\begin{aligned} P(\mathbf{y}_T|X) = &\frac{1}{Z}\exp\Big(-\sum_{p=1}^{T}\sum_{k=1}^{K}\alpha_{k,p}(y_{t+p} - R_{k,p}(X))^2 - \\ &- \sum_{p=2}^{T}\beta_l(y_{t+p} - y_{t+p-1})^2\Big), \end{aligned} \quad (25)$$

where we use separate parameters $\alpha_{k,p}$ for each unstructured predictor $R_{k,p}(X)$ at each time step over the horizon. Interaction potential from (25) is defined to impose fair assumption that patient state is not going to dramatically change between two consecutive time steps. Model parameters are learned on N horizons each of length T. To make model feasible and to

provide good estimate of uncertainty we defined parameters $\boldsymbol{\alpha}$ and β as

$$\alpha_{k,p} = \frac{e^{a_{k,p}}}{\sigma^2_{k,1}}, \qquad \beta = e^b, \tag{26}$$

where $\sigma^2_{k,1}$ is uncertainty of corresponding unstructured predictor for the first-step prediction. To have model completely defined we need to specify which unstructured predictors will be used in GCRF.

3.3. Unstructured Predictors

Unstructured predictors that will be used in GCRF are Gaussian process (non-linear) model and linear regression. Both models are intended to learn function (8). They iteratively provide predictions over horizon T and uncertainty of the first-step prediction.

3.3.1. Gaussian Process regression

A Gaussian process(GP) is generalization of a multivariate Gaussian distribution over finite vector space to a function space of infinite dimension. In usual regression setting we have following input-output relationship

$$y_i = f(\mathbf{x_i}) + \epsilon, \;\; \epsilon \sim N(0, \sigma^2_y). \tag{27}$$

Gaussian process defines prior over functions denoted by

$$f(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \tag{28}$$

where $m(\mathbf{x})$ is the mean function and $k(\mathbf{x}, \mathbf{x}')$ is the kernel (covariance function) that is required to be positive definite. For finite set of data points X and the mean function set to zero, this process defines joint distribution

$$p(\mathbf{f}|X) \sim N(0, \mathbf{K}), \tag{29}$$

where $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. Marginal distribution of outputs is

$$p(\mathbf{y}|\mathbf{X}) = \int p(\mathbf{y}|\mathbf{f}) p(\mathbf{f}|\mathbf{x}) d\mathbf{f} = N(\mathbf{y}|0, \mathbf{C}), \tag{30}$$

where $C = K + \sigma^2_y I_N$. The joint density of the observed outputs $\mathbf{y}$ and test output y_* is given by

$$\begin{pmatrix} \mathbf{y} \\ y_* \end{pmatrix} = N\left(0, \begin{bmatrix} C & \mathbf{k}_* \\ \mathbf{k}_*^T & c_* \end{bmatrix}\right). \tag{31}$$

Therefore, posterior predictive density is calculated as

$$\begin{aligned} p(y_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}) &= N(\mu_*, \sigma^2_*), & (32)\\ \mu_* &= \mathbf{k}_*^T C^{-1} \mathbf{y}, & (33)\\ \sigma^2_* &= c_* - \mathbf{k}_*^T C^{-1} \mathbf{k}_*. & (34) \end{aligned}$$

In our implementation, we use squared exponential kernel defined as

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp(-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^T M (\mathbf{x} - \mathbf{x}')), \tag{35}$$

where σ_f^2 controls vertical scale of the function changes. Diagonal matrix M includes scales for each dimension of $\mathbf{x}$, $M = diag(\mathbf{m}^{-2})$. All kernel parameters as well as noise variance are estimated using maximum likelihood approach (Rasmussen, 2006).

3.3.2. Linear ARX

Linear regression form of ARX representation from (8) is

$$\mathbf{y_i} = \mathbf{w}^T \mathbf{x_i} + \epsilon, \;\; \epsilon \sim N(0, \sigma^2_y), \tag{36}$$

where $\mathbf{w}$ is unknown set of weights. The weight and noise variance are estimated by

$$\begin{aligned} \hat{\mathbf{w}} &= X^T (X^T X)^{-1} X \mathbf{y}, & (37)\\ \sigma^2_y &= \frac{(\mathbf{y} - \hat{\mathbf{w}}^T X)^T (\mathbf{y} - \hat{\mathbf{w}}^T X)}{N - k}, & (38) \end{aligned}$$

where X is matrix representation of all data available for training; N is number of training examples; k is length of the training example vector. Prediction and uncertainty of the test data are found using

$$\begin{aligned} \mathbf{y}_* &= \hat{\mathbf{w}}^T \mathbf{x}_*, & (39)\\ \sigma^2_* &= \sigma^2_y (1 + \mathbf{x}_* (X^T X)^{-1} * \mathbf{x}_*^T). & (40) \end{aligned}$$

4. Experimental Results

4.1. Dataset

The critical aspect of the predictive model design is the availability of representative training data to learn unknown parameters. Our objective is to address a real-life scenario in which data available for training of the predictive model come from clinical trials done on a small group of diverse patients observed in time. During the data collection process, we assume that the virtual patients are monitored and the treatment strategy revised hourly. The monitoring period for each patient is assumed to last for a week (168 hours). To generate a sequence of observations for a virtual patient we need to know model parameters, initial conditions, and a sequence of medication doses. Initial conditions and parameters are randomly generated following allowable ranges, while dosages are chosen as follows. For each of the virtual patients we used its own mathematical model as a predictive model to find the best strategy that will lead patient to healthy state. As such strategy is unrealistic in clinical practice, we use

Table 1. MSE at each time stamp (up to 10 steps ahead) when predicting pathogen level P.

Model	T=1	T=2	T=3	T=4	T=5	T=6	T=7	T=8	T=9	T=10
GCRF	0.0003	0.0019	0.0054	0.0107	0.0173	0.0260	0.0359	0.0464	0.0571	0.0672
ARX lag 0	0.0031	0.0119	0.0251	0.0426	0.0649	0.0915	0.1223	0.1570	0.1951	0.2365
ARX lag 1	0.0003	0.0017	0.0053	0.0123	0.0241	0.0424	0.0686	0.1038	0.1489	0.2041
ARX lag 2	0.0002	0.0016	0.0055	0.0132	0.0262	0.0454	0.0715	0.1043	0.1427	0.1848
GP lag 0	0.2143	0.1964	0.3416	0.4507	0.6237	0.7379	0.6312	0.4672	0.4017	0.4097
GP lag 1	0.0014	0.0099	0.0307	0.0677	0.1238	0.1854	0.2265	0.2427	0.2458	0.2510
GP lag 2	0.0025	0.0118	0.0328	0.0662	0.1163	0.1894	0.2946	0.4303	0.5741	0.6955

Table 2. MSE at each time stamp (up to 10 steps ahead) when predicting level of tissue damage D.

Model	T=1	T=2	T=3	T=4	T=5	T=6	T=7	T=8	T=9	T=10
GCRF	0.0001	0.0009	0.0025	0.0059	0.0112	0.0179	0.0258	0.0347	0.0443	0.0535
ARX lag 0	0.0022	0.0086	0.0192	0.0340	0.0530	0.0764	0.1041	0.1362	0.1729	0.2142
ARX lag 1	0.0002	0.0009	0.0027	0.0060	0.0113	0.0189	0.0293	0.0425	0.0587	0.0780
ARX lag 2	0.0002	0.0011	0.0035	0.0081	0.0153	0.0254	0.0384	0.0542	0.0725	0.0931
GP lag 0	0.0020	0.0060	0.0108	0.0173	0.0343	0.1124	0.3595	0.7622	1.1729	1.5570
GP lag 1	0.0025	0.0244	0.1283	1.7231	10.5825	16.7864	14.8040	10.9814	15.1810	17.9929
GP lag 2	0.0042	0.0219	0.0654	0.2395	0.8997	2.6021	4.8395	5.7687	5.5865	5.2554

a more clinically realistic scenario in which for an observed patient's state at time point k, the doses at k are reasonably close to ideal. This is modeled with random Gaussian noise added to $AIDOSE$ and $PIDOSE$ values found by the ideal strategy. We generated a small, well-balanced set (equal number of septic, aseptic and healthy patients) of 18 virtual patients (Radosavljevic et al., 2012) with week-long hourly observations and treatment adjustments. To evaluate prediction models we generated a test population of 35 patients with hourly observations and treatment adjustments for one week.

4.2. ARX and GP Models

We trained three linear ARX and three GP models with lags 0, 1, and 2 for each of four GCRF submodels in which they are used as unstructured predictors. Each GCRF submodel was trained to predict $T = 10$ steps ahead in the future.

4.3. Evaluation Measures

To compare models with respect to accuracy we use the standard mean squared error (MSE) measure. In addition, we also use a the average negative log-predictive density (NLPD) (Quiñonero Candela et al., 2006) to estimate quality of uncertainty predictions. The average NLPD is defined as

$$NLPD = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{(y_i - y_{i*})^2}{2\sigma_{i*}^2} + \log \sigma_{i*}\right). \qquad (41)$$

NLPD is sensitive to the quality of prediction and uncertainty estimation and it penalizes both over and under confident predictions. Smaller values of NLPD correspond to better quality of the estimates. For given y_{i*}, NLPD reaches minimum for $\sigma_{i*}^2 = (y_i - y_{i*})^2$.

4.4. Results

We compared MSE of each of the unstructured predictors and GCRF at each time point $p = 1, \ldots 10$. Predictions of unstructured predictors are obtained recursively. Although we predict all four states, we present results when predicting pathogen level P and tissue damage D as these two states are important for determining treatment strategy. MSEs when predicting P are presented at Table 1, while MSEs when predicting D are presented at Table 2. For short horizons, GCRF predictions are either more accurate or comparable to unstructured predictions. For long horizons, GCRF predictions of P are much more accurate than unstructured predictions, while predictions of D are comparable to unstructured ones.

In addition, we compared NLPD values of GCRF and GP with lag 2 (GP naturally provides an uncertainty estimate). NLPD results are presented in Tables 3

Table 3. Average NLPD for each time stamp (up to 10 steps ahead) when predicting pathogen level P.

Model	T=1	T=2	T=3	T=4	T=5	T=6	T=7	T=8	T=9	T=10
GCRF	-4.6	-3.5	-2.8	-2.4	-2.1	-1.8	-1.6	-1.4	-1.3	-1.3
GP lag 2	-3.9	0.0	9.6	33.3	58.6	81.5	119.9	172.3	235.5	307.7

Table 4. Average NLPD for each time stamp (up to 10 steps ahead) when predicting level of tissue damage D.

Model	T=1	T=2	T=3	T=4	T=5	T=6	T=7	T=8	T=9	T=10
GCRF	-4.0	-1.8	-2.8	-2.1	-1.5	-1.0	-0.4	0.2	0.8	1.2
GP lag 2	0.5	25.0	173.2	289.0	637.7	950.2	1549.3	2187.6	3146.6	4351.3

and 4. NLPD of GCRF is much better than NLPD of GP. The large values of NLPD for GP are due to over-confident estimation of uncertainty (the proper propagation of uncertainty will be addressed in future work). This shows one more strength of GCRF as it achieves good quality of predictions even when combining low quality predictors. To get better insight to the quality of GCRF predictions, we compare GCRF NLPD to the minimal NLPD for given GCRF prediction. Minimal NLPD when predicting P ranges from -5.9 to -4, while GCRF NLPD ranges from -4.6 to -1.3 (Table 3). The larger deviation of NLPD from minimal value for longer prediction horizons may be explained by conservative (under-confident) estimate of uncertainty (Figure 2a). Minimal NLPD when predicting D ranges from -6.4 to -2.7. At the same time, GCRF NLPD ranges from -4 to 1.2. The NLPD deviation from minimal values might be described in over-confident predictions (Figure 2b). In Figure 2 we presented unstructured and GCRF predictions for a virtual patient observed over three hours. GCRF predictions are accompanied with an uncertainty estimate.

5. Conclusion and Future Work

We presented the GCRF model for predicting patient's response in acute inflammation treatment. We showed that, along with accurate prediction, GCRF also provides a reasonable measure of uncertainty. In future work we plan to incorporate GP models with propagation of uncertainty (Kocijan et al., 2004) and to expand the GCRF model to deal with partially observable data that are commonly found in clinical practice. Finally, our work in progress is aimed to use GCRF in model predictive control setup in order to suggest optimal therapy.

Acknowledgments

This work was funded in part by DARPA grant [DARPAN66001-11-1-4183] negotiated by SSC Pacific grant.

References

Clermont, Gilles, Rubin, Jonathan, and Day, Judy. Using nonlinear model predictive control to find optimal therapeutic strategies to modulate inflammation. *Mathematical Biosciences and Engineering*, 7 (4):739–763, 2010.

Kocijan, J., Murray-Smith, R., Rasmussen, C.E., and Girard, A. Gaussian process model based predictive control. In *American Control Conference, 2004. Proceedings of the 2004*, pp. 2214–2219. IEEE, 2004.

Lafferty, J, McCallum A and Pereira, F. Conditional random fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings International Conference on Machine Learning*, 2001.

Noble, Sarah L, Sherer, Eric, Hannemann, Robert E, Ramkrishna, Doraiswami, Vik, Terry, and Rundell, Ann E. Using adaptive model predictive control to customize maintenance therapy chemotherapeutic dosing for childhood acute lymphoblastic leukemia. *Journal of theoretical biology*, 264(3):990–1002, 2010.

Qin, T, Liu, T, Zhang, X, Wang, D, and Li, H. Global Ranking Using Continuous Conditional Random Fields. *Neural Information Processing Systems*, 2008.

Quiñonero Candela, Joaquin, Rasmussen, Carl Edward, Sinz, Fabian, Bousquet, Olivier, and Schölkopf, Bernhard. Evaluating predictive uncertainty challenge. In *Proceedings of the First international conference on Machine Learning Chal-*

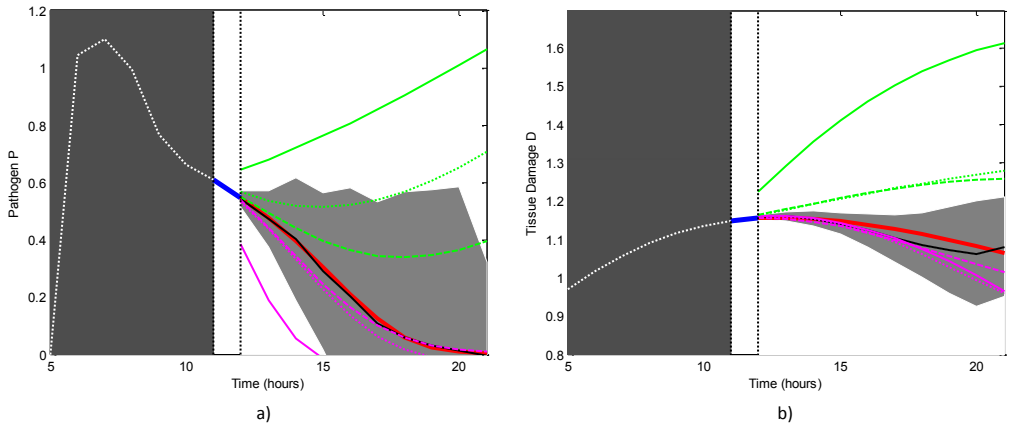

Figure 2. GCRF and unstructured predictions and GCRF confidence intervals of a) pathogen P and b) tissue damage D on a virtual patient; dotted white line unseen patient states; solid blue line observations; solid black line mean of GCRF prediction 10 steps ahead; gray area in around GCRF prediction GCRF 95% confidence interval; solid red line true patient future behavior; magenta - GP, green - ARX, solid - lag 0, dotted - lag 1, dashed - lag 2.

lenges, MLCW'05, pp. 1–27, Berlin, Heidelberg, 2006. Springer-Verlag.

Radosavljevic, Vladan, Ristovski, Kosta, and Obradovic, Zoran. A data mining approach for optimization of acute inflammation therapy. In *BIBM*, pp. 1–6. IEEE Computer Society, 2012.

Rasmussen, Carl Edward. Gaussian processes for machine learning. MIT Press, 2006.

Thiel, Steven W, Rosini, Jamie M, Shannon, William, Doherty, Joshua A, Micek, Scott T, and Kollef, Marin H. Early prediction of septic shock in hospitalized patients. *Journal of hospital medicine : an official publication of the Society of Hospital Medicine*, 5(1):19–25, 2010.

Wang, Youqing, Zisser, Howard, Dassau, Eyal, Jovanovic, Lois, and Doyle, Francis J. Model predictive control with learning-type set-point: Application to artificial pancreatic β-cell. *AIChE Journal*, 56(6):1510–1518, 2009.

Xin, Xin, King, Irwin, Deng, Hongbo, and Lyu, Michael R. A social recommendation framework based on multi-scale continuous conditional random fields. In *Proceedings of the 18th ACM conference on Information and knowledge management*, CIKM '09, pp. 1247–1256, New York, NY, USA, 2009. ACM.

Yelneedi, S., Lakshminarayanan, S., and Rangaiah, GP. A comparative study of three advanced controllers for the regulation of hypnosis. *Journal of Process Control*, 19(9):1458–1469, 2009.