Kernel-based Characterization of Dynamics in a Heterogeneous Population of Septic Patients Under Therapy

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

Abstract

Sepsis is a medical condition characterized as a systemic inflammatory response to an infection. The high level of heterogeneity among sepsis patients is one of the main reasons of unsuccessful clinical trials. A more careful targeting of specific therapeutic strategies to more biologically homogeneous groups of patients is essential to developing effective sepsis treatment. We propose a kernel-based approach to characterize dynamics of inflammatory response in a heterogeneous population of septic patients. Our method utilizes Linear State Space Control (LSSC) models to take into account dynamics of inflammatory response over time as well as the effect of therapy applied to the patient. We use a similarity measure defined on kernels of LSSC models to find homogeneous groups of patients. An application of the proposed method to analysis of dynamics of inflammatory response to sepsis therapy in 64 virtual patients identified four biologically relevant homogeneous groups providing the initial evidence that patient-specific sepsis treatment based on several treatment protocols is feasible.

1. Introduction

One of the most challenging problems in clinical practice is planning of personalized therapy regimens and durations. Personalized therapy is especially critical in rapid progression medical conditions like sepsis, a systematic inflammatory response syndrome triggered by infection. Sepsis is the leading cause of death in critically ill patients in the United States (Hotchkiss & Karl, 2003). It is often diagnosed too late and the patient is then treated with broad-spectrum antibiotics and/or intravenous fluids with dosages adjusted manually even though more specific therapy would be far more effective. The high level of heterogeneity among sepsis patients is one of the main causes of failed clinical trials (Hotchkiss & Karl, 2003). A more careful targeting of specific therapeutic strategies to more biologically homogeneous groups of patients is essential to developing effective sepsis treatment (Marshall, 2005).

Systems that cluster heterogeneous patient populations have proven to be extremely useful to the development of multimodal treatment for cancer (Marshall, 2005). In this study our main objective is to cluster a heterogeneous population of septic patients into groups with biologically similar responses to medications and similar systemic reactions to an infection (homogenous groups with similar dynamics). To achieve this objective we will represent the dynamic of each patient by a linear state space control (LSSC) model where sequence of medication dosages is considered as a control signal. Then, similarity between dynamics of patients inflammatory responses can be defined through the similarity of LSSC models. Therefore, our main focus becomes: (1) to define an appropriate measure of similarity between two LSSC models with medications as control signals; (2) to cluster LSSC models according to a similarity measure defined in (1); and (3) to validate biological relevancy of obtained clusters.

The problem of comparisons of dynamical models has been studied in the past decade. Different metrics are utilized for comparing Autoregressive Models (AR). In (Martin, 2000) AR models are compared based on cepstrum coefficients. This concept is extended to state-space models by considering the subspace angles between subspaces that are derived from AR mod-

Appearing in the *ICML 2012 Workshop on Clinical Data Analysis*, Edinburgh, Scotland, UK, 2012.

els (Cock & Moor, 2002). The unifying framework for comparisons of state-space models is presented in (Vishwanathan et al., 2006). On the other hand, these approaches do not take into account control signals and thus they are not directly applicable to LSSC models.

In this paper we propose a metric for measuring similarity between two LSSC models. Our approach is based on kernels designed for state-space models which are initially considered in (Vishwanathan et al., 2006). We expanded work proposed in (Vishwanathan et al., 2006) to obtain kernels for LSSC models and we defined similarity between models based on calculated kernels. Finally, we showed how the proposed similarity measure can be deployed in the spectral clustering algorithm to find homogenous clusters. Besides being used for clustering patients with respect to their expected response to a therapy, the proposed similarity measure can also be used to detect unusual patients dynamics.

2. Methodology

2.1. Linear state-space control models (LSSC)

One important class of dynamical systems are discrete linear state-space control models known as Kalman filters. We use them to represent dynamics of each patient in the form

$$x_{t+1} = Ax_t + Bu_t + w_t, \qquad w_t \sim N(0, Q), \tag{1}$$
$$y_{t+1} = Cx_{t+1} + v_t, \qquad v_t \sim N(0, R), \tag{2}$$
$$x_1 = x_0 + \eta, \qquad \eta \sim N(0, Q_0), \tag{3}$$

where the noise terms w_t, v_t, η are assumed to be zero mean *i.i.d.* normal random variables, y_t is observed vector at time t ($t = 1, 2, \ldots, T$), x_t is the hidden state vector and u_t is control signal. T is the length of the observed sequence. Vector x_0 (initial state) and matrices A, B, C, Q, R, Q_0 are parameters of the model, which we will refer to as Θ. As the model has hidden vectors, learning parameters will be achieved by the expectation-maximization (EM) algorithm. Parameter learning is derived similarly to the technique shown in (Bishop, 2006) and only final equations are presented here. We denote

$$x_{t|t} = E[x_t|y_1, \ldots, y_t], \tag{4}$$
$$P_{t|t} = E[(x_t - \hat{x}_{t|t})(x_t - \hat{x}_{t|t})^T|y_1, \ldots, y_t]. \tag{5}$$

Having that all distributions in the model are multivariate Gaussians we can write forward propagation

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} + Bu_t, \tag{6}$$
$$P_{t+1|t} = AP_{t|t}A^t. \tag{7}$$

When observation y_{t+1} becomes available measurement update equations are

$$\hat{x}_{t+1|t+1} = \hat{x}_{t|t} + K(y_{t+1} - C\hat{x}_{t+1|t}), \tag{8}$$
$$P_{t+1|t+1} = (I - KC)P_{t+1|t}, \tag{9}$$
$$K = P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}, \tag{10}$$
$$\hat{x}_{1|1} = \hat{x}_0 + K_1(y_1 - Cx_0), \tag{11}$$
$$P_{1|1} = (I - K_1C)Q_0, \tag{12}$$
$$K_1 = Q_0C^T(CQ_0C^T + R)^{-1}. \tag{13}$$

Backward recursion known as smoothing is obtained considering that the entire sequence of length T is available and is in the form

$$\hat{x}_{t|T} = \hat{x}_{t|t} + J_t(\hat{x}_{t+1|T} - \hat{x}_{t+1|t}), \tag{14}$$
$$P_{t|T} = P_{t|t} + J_t(P_{t+1|T} - P_{t+1|t})J_k^T, \tag{15}$$
$$J_k = P_{t|t}A^TP_{t+1|t}^{-1}. \tag{16}$$

EM algorithm is applied on logarithm of complete likelihood

$$p(y, x|\Theta) = p(x_1)\prod_{t=2}^{T} p(x_t|x_{t-1}, A, B, Q, u_{t-1}) \times \prod_{t=1}^{T} p(y_t|x_t, C, R). \tag{17}$$

Forward, update and smoothing steps are executed at first in the E-step. Afterwards, following expectations are calculated

$$E[x_{t|T}] = \hat{x}_{t|T}, \tag{18}$$
$$E[x_{t|T}x_{t-1|T}^T] = J_{t-1}P_{t|T} + \hat{x}_{t|T}\hat{x}_{t-1|T}^T, \tag{19}$$
$$E[x_{t|T}x_{t|T}^T] = P_{t|T} + \hat{x}_{t|T}\hat{x}_{t|T}^T. \tag{20}$$

The M-step is defined by minimizing expectation over hidden state vectors of complete log-likelihood with respect to Θ as

$$\Theta = \underset{\Theta}{\operatorname{argmax}} E_{x_{1:T}}[\log p(y, x|\Theta)]. \tag{21}$$

2.2. Binet-Cauchy kernel on LSSC models

In order to find patients with similar inflammatory responses, we utilize an algebraic approach to compare two LSSC models. The novelty of our work is that our models include control signal u, which was not addressed in related approaches for time-series modeling. As a first step in defining similarity, we introduce kernel between two LSSC models M_1 and M_2. Kernel is computed as a discounted dot-product of coresponding trajectories y_1 and y_2 generated using models M_1

and M_2 that have the same initial conditions and use the same control sequences, similar to (Vishwanathan et al., 2006)

$$K(M_1, M_2) = E_{x_0,u} \sum_{t=1}^{\infty} e^{-\lambda t} y_1^T(t) W y_2(t), \quad (22)$$

where λ is discounting factor that down-weights samples as t increases, and W is a weight matrix that puts different emphasis on different components of trajectories. Trajectory y is found from (7) following forward propagation process

$$y_t = C(A^{t-1}x_0 + \sum_{i=0}^{t-2} A^i(Bu_{t-1-i} + v_{t-1-i})) + w_t. \quad (23)$$

We need to find expectation E over x_0 and u as we would like to define similarity that is independent of the initial conditions and the specific control sequence. We are assuming that (1) noises between models are uncorrelated; (2) two LSSC models are compared on the same random control sequences with zero-mean Gaussian distribution with covariance U. In order to have kernel independent on initial conditions we will assume that both models have the same initial values sampled from the zero-mean Gaussian distribution with covariance X_0. Using relevant term from (22) expectation over x_0 is calculated

$$\begin{aligned} &E_{x_0}(\sum_{t=1}^{\infty} e^{-\lambda t} y_1^T(t) W y_2(t)) = \\ &= E_{x_0}(\sum_{t=1}^{\infty} e^{-t} (C_1 A_1^{t-1} x_0)^T W (C_2 A_2^{t-1} x_0))) \\ &= E_{x_0}(e^{-\lambda} x_0^T (\sum_{t=0}^{\infty} e^{-\lambda t} (A_1^t)^T \hat{W} A_2^t) x_0) \\ &= E_{x_0}(e^{-\lambda} x_0^T M x_0) = e^{-\lambda} Tr(MX_0), \end{aligned} \quad (24)$$

where M is a solution of Sylvester's equation $M = e^{-\lambda} A_1^T M A_2 + \hat{W}$ and $\hat{W} = C_1^T W C_2$. Convergence of the infinite sum is achived for $e^{-\lambda}||A_1||||A_2|| < 1$ (Vishwanathan et al., 2006). Similarly, relevant term to calculate expectation over u is obtained from (22). After some rearrangement and taking expectation it is obtained

$$\begin{aligned} &E_u(\sum_{t=1}^{\infty} e^{-\lambda t} y_1^T(t) W y_2(t)) = \\ &= Tr(B_2 U B_1^T \sum_{t=2}^{\infty} \sum_{j=0}^{t-2} e^{-\lambda t} (A_1^j)^T \hat{W} A_2^j) \\ &= Tr(B_2 U B_1^T \sum_{j=0}^{\infty} \sum_{t=j+2}^{\infty} e^{-\lambda t} (A_1^j)^T \hat{W} A_2^j) \\ &= \frac{e^{-2\lambda}}{1 - e^{-\lambda}} Tr(B_2 U B_1^T M). \end{aligned} \quad (25)$$

Finally, kernel can be written as

$$\begin{aligned} K(M_1, M_2) =& e^{-\lambda} Tr(MX_0) + \\ &+ \frac{e^{-2\lambda}}{1 - e^{-\lambda}} Tr(B_2 U B_1^T M). \end{aligned} \quad (26)$$

2.3. Spectral clustering for state-space models

We utilize spectral clustering to group the LSSC models based on similarity between them. Clustering methods which require calculating cluster representatives are not applicable on this problem as it is not trivial to find the representatives in the case of LSSC models (Vishwanathan et al., 2006). Spectral clustering overcomes this by permitting the illustration of complex clusters of arbitrary shapes without calculating their representatives.

To determine clusters via spectral clustering, we need to (1) define a proper similarity measure based on kernels and (2) determine the appropriate number of clusters for a set of LSSC models.

2.3.1. Similarity between state-space models

For the spectral clustering task we need to define similarity between two models M_i and M_j which has to be in the 0-1 range. In order to define any similarity measure, kernels between models have to be normalized. Normalized kernels (Ah-Pine, 2010) are usually obtained using cosine measure, which does not carry information about norms of trajectories. The norms cannot be neglected in the sepsis application so cosine measure has to be modified for a use in the clustering of patients dynamics. Very useful modification to cosine normalization is proposed in (Ah-Pine, 2010) where a kernel is normalized using

$$\begin{aligned} k(M_i, M_j) &= \frac{K(M_i, M_j)}{F_m}, \\ F_t &= (\frac{1}{2}(K^m(M_i, M_i) + K^m(M_j, M_j)))^{1/m}, \end{aligned} \quad (27)$$

and K denotes un-normalized kernels obtained from (26) and $m > 0$ is the parameter which has to be set. It is shown in (Ah-Pine, 2010) that by increasing parameter m, the impact of norms to similarity is increasing. Therefore, we will use this way of normalization in our experiments. Similarity between two models is defined as

$$w_{i,j} = e^{-d^2(M_1, M_2)}, \quad (28)$$

where d is commonly used distance measure defined by

$$d(M_1, M_2) = \sqrt{2(1 - k(M_1, M_2))}. \quad (29)$$

Algorithm 1 Spectral Clustering

1: Define a set of LSSC models (vertices) $M = \{M_1, ..., M_n\}$ and specify number of clusters c
2: Compute affinity matrix G
3: Make diagonal matrix D whose (i,i) element is the sum of G's i-th row
4: Compute normalized version of G $L = D^{-\frac{1}{2}} G D^{-\frac{1}{2}}$
5: Find $v_1, v_2, ..., v_c$ the c largest eigenvectors of L and create matrix $V = [v_1\ v_2\ ...\ v_c]$ where v_i is i-th column in V
6: Find a matrix Y from V such that $Y_{ij} = \frac{V_{ij}}{\sqrt{\sum_j V_{ij}^2}}$
7: Treating each row of Y as a point, cluster all rows into c clusters via K-means clustering algorithm which attempts to minimize distortion
8: Assign LSSC model M_i to particular cluster if and only if row i of the matrix Y is assigned to that cluster

2.3.2. Spectral clustering algorithm

Spectral clustering is solving graph cut problem. We represent the set of LSSC models $M = \{M_1, \dots M_n\}$ as vertices, while similarities between models represent edge weights. The adjacency matrix G associated with this weighted undirected graph used in the spectral clustering algorithm is defined as

$$G_{ij} = \begin{cases} w_{ij} & \textit{if } i \neq j \textit{ and } (i,j) \textit{ is in the edge set} \\ 0 & \textit{otherwise} \end{cases}$$

We follow algorithm proposed in (Ng et al., 2001) to cluster models, see Algorithm 1.

2.3.3. Number of clusters

Our objective is to find stable clusters effectively. In the method proposed in (Ben-Hur et al., 2002), stable clusters are discovered based on sub-samples, following the intuition that cluster partitions are stable when similar partitions are found on different subsamples obtained by random sampling without replacement of 70-90% of the original dataset. In short, we firstly decide on a possible range for a number of clusters (c) and then apply the following algorithm. For each c from the possible range we: (1) generate and cluster pair of subsamples of the data; (2) compute the similarity between clusters obtained in (1) using the points common to both subsamples; and (3) repeat steps (1) and (2) multiple times to obtain distribution of similarities.

The number of stable clusters is determined by observing distributions of similarity for different values of k. We stop the process when distribution is changes from being skewed to 1. For more details see (Ben-Hur et al., 2002) and Section 4.2.

3. Experiments

3.1. Virtual patient

To significantly reduce chance of a clinical failure and to save 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 (Day, 2010). An important advantage of having a virtual patient model for experiments is the possibility of testing large number of hypotheses on same virtual patients, comparing the outcomes and filtering out possibly unnecessary tests on animal models and later on real patients. Virtual patients are generated using a carefully determined parametric mathematical model to simulate the process of interest. To follow a real-life scenario, virtual patient models are accompanied with well-defined constraints on parameter ranges that are specified in accordance with clinical practice.

3.1.1. Patient model

We will use a mathematical model recently proposed in (Day, 2010) that is capable of simulating: (1) an evolution of a bacterial pathogen population (P) that initiates the cascade of inflammation; (2) dynamics of pro-inflammatory mediators (N); (3) markers of tissue damage/dysfunction (D); and (4) evolution of anti-inflammatory mediators (CA) which are all 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):

$$\frac{dP}{dt} = k_{pg}\left(1 - \frac{P}{P_\infty}\right) - \frac{k_{pm} s_m P}{\mu_m + k_{mp} P} - k_{pn} f(N) P,$$

$$\frac{dN}{dt} = \frac{s_{nr} R}{\mu_{nr} + R} - \mu N + PIDOSE(t),$$

$$\frac{dD}{dt} = \frac{k_{dn} f(N)^6}{x_{dn}^6 + f(N)^6} - \mu_d D,$$

$$\frac{dCA}{dt} = \frac{k_{cn} f(N + k_{cnd} D)}{1 + f(N + k_{cnd} D)} - \mu_c CA + AIDOSE(t),$$

where

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

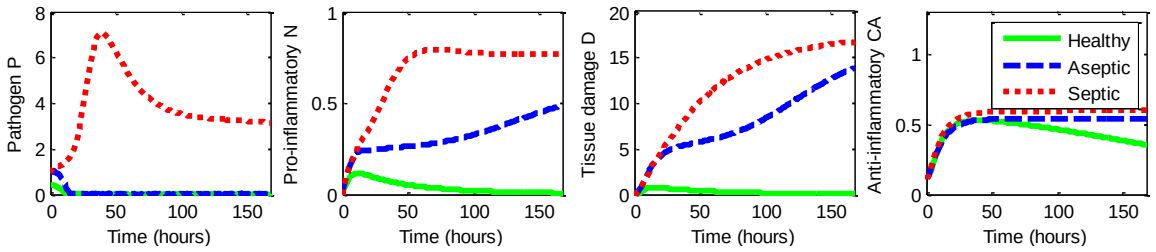

Figure 1. Evolution of pathogen population (P), pro-inflammatory mediators (N), tissue damage (D), and anti-inflammatory mediators (CA) of three virtual patients with healthy (blue), aseptic (green), and septic (red) outcomes in the absence of therapy.

All variables used in the mathematical model except output signals P, N, D, CA are parameters with valid ranges specified in (Day, 2010). Population of virtual patients was generated by random sampling from uniform distribution on valid ranges of the initial conditions for P and CA (while N and D were set to 0) and three parameters in ODE: k_{pg} - growth rate of pathogen P; k_{cn} - maximum production rate of anti-inflammatory mediator CA; and k_{nd} - activation of phagocytes N by tissue damage D. Since many molecules are produced by the same biological processes, parameters k_{cnd} - relative effectiveness of damaged tissue D in inducing production of the anti-inflammatory mediator CA, k_{nn} - activation of resting phagocytes N by already activated phagocytes N, and k_{np} - activation of phagocytes N by pathogen P were chosen to co-vary with k_{cn} and k_{nd} (Day, 2010). If two parameters k_1 and k_2 co-vary that means if k_1 is randomly generated and it is $r\%$ relatively different from its reference value k_{1ref}, then the value for k_2 should also be $r\%$ relatively different from its reference value k_{2ref}

In (Day, 2010) the variability in k_{cnd} was specified to co-vary with k_{cn}, and k_{nn} and k_{np} were specified to co-vary with k_{nd} so that the rates at which body were producing CA and N were balanced. Variability of immune response in the population of virtual patients is obtained by changing parameter values and initial conditions. All other parameters were fixed to referent values as in (Day, 2010).

Although conceptual, ODE is capable of modeling different complex dynamics of a patients immune system response to infection induced by pathogen P. From ODE, large P leads to the development of a pro-inflammatory response N. When the growth rate kpg of the pathogen P is low, activated phagocytes N are capable of clearing the pathogen in normal individuals. However, if the growth rate k_{pg} is high, then a sufficiently large level of pathogen can induce a persistent infection despite the immune system response

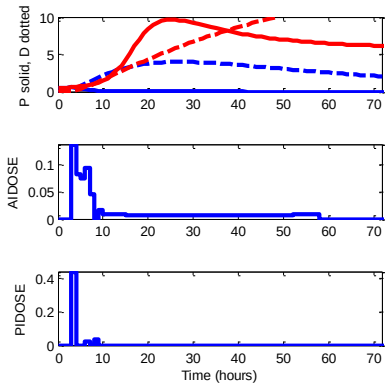

Figure 2. A patient response (P and D at top panel) on non-ideal treatment sequence ($AIDOSE$ at middle panel; $PIDOSE$ at bottom panel) with healthy outcome (blue at top panel) that would be septic otherwise (red at top panel).

by activated phagocytes N. Large N indicates faster elimination of pathogen P. However, large N damages tissue D and therefore mobilizes a negative feedback, or anti-inflammatory response CA, which lowers N (Day, 2010). Also, CA inhibits damage to tissue D that may be caused by N. The rate and dynamics of immune system response are completely governed by parameters of the mathematical model which vary among virtual patient population representing the heterogeneity among real patients.

In simulations t is an hourly step that starts from 0. At $t = 0$, outputs and parameters are set to initial conditions. Then, all four outputs evolve according to ODE through the simulation time t. According to (Day, 2010) there are three possible outcomes at the end of simulation time, which are shown in Figure 1. A patient is in healthy state if $P = 0$, $N = 0$, $D = 0$, and $CA > 0$ at the end of simulation. The aseptic death state of the patient is defined as $P = 0$, $N > 0$, $D > 0$, and $CA > 0$. The third possible outcome is septic death, where all outputs are non-zero. Leading the

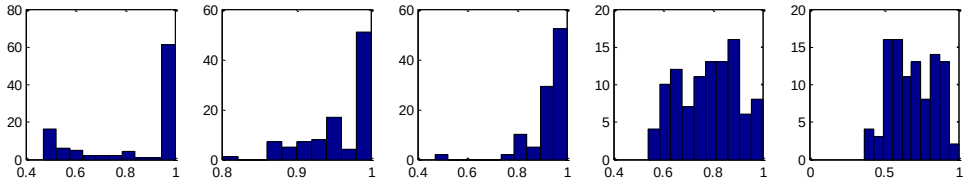

Figure 3. Histograms of similarity distributions for different #clusters (#clusters=2 most left up to #clusters=6 most right)

virtual patient to healthy state is a difficult challenge of maintaining a balance between objectives $P = 0$ and $D = 0$ (if these conditions are satisfied Nwill eventually be 0 so there is no need to have $N = 0$ as an objective).

3.1.2. Heterogenous virtual patients' data

The critical aspect of the virtual patient analysis is the availability of representative data. Our objective was to address a real-life scenario in which data come from clinical trials done on a small group of heterogeneous patients observed in time. Accordingly, a set of $N_{patients}$ virtual patients with hourly observations over one week (168 hours) was generated from ODE equations. To generate a sequence of observations for a virtual patient we need to know parameters of the, initial conditions and a therapy sequence. Initial conditions and parameters are randomly generated following allowable ranges while a therapy sequence was carefully chosen.

In industrial applications, control sequences are usually generated randomly so that they span the whole operational range and adequately characterize the response of the system. Random generation of treatments is not a clinically relevant scenario. Instead we propose the following approach. For each of the $N_patients$ we used its own mathematical model as a predictive model in model predictive control (MPC) scenario (Day, 2010). Such predictive models give perfect prediction of the patients future states, as their predictions are identical to future observations at every time point (ideal predictive models). The ideal control sequences for each virtual patient would be obtained by minimizing the objective function as in (Day, 2010). The ideal control sequences are not realistic in clinical practice and are also unsuitable for learning data-driven models because they do not contain enough dynamics to sufficiently represent the system response. Therefore, we used a more clinically realistic scenario, such that for an observed patients state at time point t, therapy dosages at t may not bethe ideal but it should be reasonably close to the ideal. This is modeled such that at each time point t random Gaussian noise is added to $AIDOSE$ and $PIDOSE$ values found by the MPC with the ideal predictive control. Then, instead of treating patients with the ideal therapy sequences we treated patients with non-ideal ones, which gave a wider range of system response (Fig 2). From the clinical aspect, where data are limited and expensive, it is important to follow a real life scenario and perform analysis on a relatively small number of patients. We generated a population of 100 virtual patients by randomly choosing parameters of the mathematical model. The total simulation time was 168 hours (1 week) with hourly measurements of outputs P, N, D, and CA. State of the patient was determined based on the values of the outputs at the end of the simulation time of 168 hours. $N_{patients} = 64$ were selected to receive therapy according to the criterion that their value of N exceeded 0.05 at certain point in time (Day, 2010). Non-ideal control sequences were generated for each of $N_{patients}$. Among $N_{patients} = 64$ patients there were 8 with septic, 14 with aseptic, and 42 with healthy outcomes after application of non-ideal control.

3.2. Design of LSSC models

Each of 64 patients is modeled with the same model structure. For our experiments we consider LSSC models having time lags p and q for both hidden states and control sequence respectfully. Model is designed as

$$\begin{aligned} x_{t+1} &= A_1 x_t + \cdots + A_p x_{t-p+1} + \\ &\quad + B_1 u_t + \cdots + B_q u_{t-q+1} + v_t, \\ y_{t+1} &= I x_{t+1} + w_t, \end{aligned} \tag{30}$$

where $y = [PNDCA]$ and $u = [AIDOSEPIDOSE]$. Defined this way models could be easily transformed in augmented equations in the form from (1) and trained accordingly. Lag values are experimentally selected using the BIC criterion which provided $p = 2$ and $q = 1$ as the best choice. Further experiments are performed using the model with the lag operators set to these values.

Table 1. Cluster statistics.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
# Healthy	13	12	0	17
# Aseptic	2	4	4	3
# Septic	0	0	8	0
P_0	0.75±0.18	0.65±0.21	0.74±0.19	0.65±0.16
CA_0	0.120±0.020	0.121±0.016	0.129±0.150	0.119±0.016
kpg	0.42±0.09	0.42±0.07	0.53±0.06	0.45±0.09
knd	0.021±0.003	0.022±0.003	0.020±0.003	0.020±0.002
kcn	0.043±0.004	0.038±0.007	0.039±0.005	0.040±0.006
kcnd	51.4±5.4	46.0±8.6	47.1±6.1	47.4±8.0
knn	0.0104±0.0015	0.0109±0.0017	0.0098±0.0014	0.0098±0.0013
knp	0.104±0.015	0.109±0.017	0.098±0.014	0.098±0.013

3.3. Kernels

Once the LSSC model is set, we need to determine parameters for model similarity calculation. The important factors that influence the value of the kernel are the initial state of LSSC given by x_0, the sequence of control signals given by u, and the value of parameter λ. As we focus on similarities between dynamics we will use identical initial states sampled from a multivariate Gaussian distribution covariance matrix which is equal to identity matrix I in our application. We assume that control signals u are sampled from a multivariate Gaussian distribution $N(0, U_0)$ with mean 0 and covariance matrix equal to σI where σ is set to 0.25 in accordance to levels of $AIDOSE$ and $PIDOSE$ from the previous section. Another parameter that we need to set is λ. The larger the values of λ are the heavier attenuation of output y when time increases. On the other hand, small values of λ allow for analysis of long term interactions. Since we would like to find similar inflammatory responses over the long range we used small value of λ. The smallest possible value of λ is limited by existence of kernel. Therefore, we set λ to 0.15. Parameter m is set to 100 because we want to take into account both cosine distance and norms with strict penalties on both (see (Ah-Pine, 2010)). Also, similarities between models with kernel distances d greater than 1.4 are set to 0. (normalized kernel is not positive).

4. Results

4.1. Outlier detection

We used kernels to find unusual dynamics. From the top panel of Fig 4, top panel, we see that dynamics of the patient #48 is different from dynamics of remaining 63 patients according to kernel similarity. Therefore, we exclude the model for the patient #48 proceed with analysis of the remaining 63 models.

4.2. Number of clusters

In the experiments conducted to discover the number of stable clusters we changed the possible number of clusters c from 2 to 15. For each c we generated 100 pairs of model subsamples with sampling rate 0.9 and measured similarity between each pair of models. Histograms of the resulting similarity distributions are shown in Fig 3 (for $c > 6$ histograms are omitted because they are following a similar pattern as that for $c = 6$). Following procedure suggested at (Ben-Hur et al., 2002) we noticed enormous change in the similarity distribution between $c = 4$ and $c = 5$. Therefore our selection is $c = 4$ clusters.

4.3. Cluster analysis

Properties of four identified clusters are summarized in Table 1. Clusters 1, 2 and 4 contain a number of healthy patients along with aseptic ones. Cluster 3 contains 12 patients, among whom 8 were with septic outcome while 4 were with aseptic outcome. All septic patients from the virtual population were grouped into this cluster. In order to characterize each cluster, we provide average and standard deviation of patients parameters within the cluster. To identify statistically significant differences among clusters, we ran a paired one-tailed t-test of the null hypothesis that patients parameters have equal means with unequal variances, against the alternative that the mean of one patient is greater than the mean of other patient with significance $\alpha = 0.05$. In the case of P_0 the null hypothesis is rejected between cluster 1 and cluster 4, meaning that cluster 1 has significantly larger initial value of P_0 than cluster 4. In terms of k_{pg}, cluster 3 (all septic and aseptic patients) has significantly larger value than other clusters. As k_{pg} corresponds to the rate of

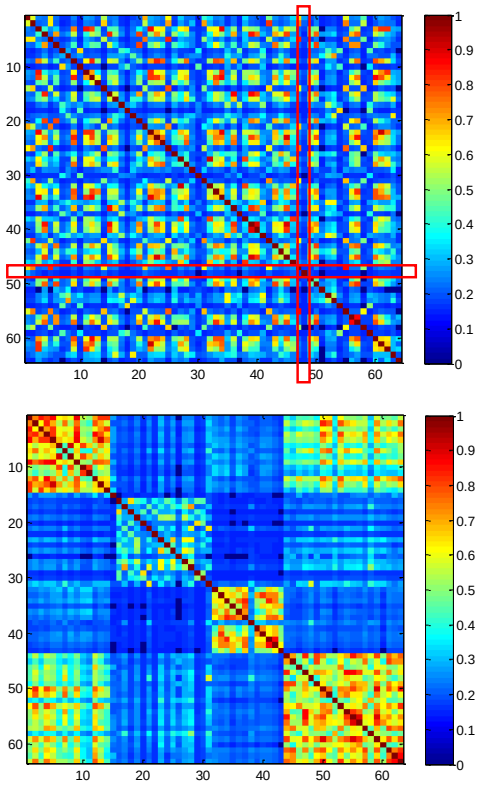

Figure 4. Similarity matrices obtained using kernel with $Lambda = 0.15$. The top plot shows similarity among 64 non clustered patients, red box designates an outlier (patient #48) not similar to others. The bottom plot shows similarity among 63 patients (without an outlier) grouped into four clusters.

pathogen development this means that cluster 3 contains patients who had more serious pathogen infections than patients in other clusters. When looking at k_{nd} (and related k_{nn} and k_{np}) we found that significantly higher values are in cluster 2 when comparing to cluster 3 and 4. Therefore, the difference between cluster 2 and cluster (mostly healthy patient) is in different dynamics of production of phagocytes N. Cluster 1 is significantly different from the other cluster in the rate of production of anti-inflammatory mediator CA (k_{cn} and related k_{cnd}).

5. Conclusion

We have developed a kernel-based approach to characterize dynamics of inflammatory response in a heterogeneous population of septic patients. Our method utilizes Linear State Space Control (LSSC) models to take into account the dynamic of inflammatory response over time as well as the effect of therapy applied to the patient. An application of the proposed method to analysis of dynamics of inflammatory response to sepsis therapy in 64 virtual patients identified four biologically relevant homogeneous groups. Our research in progress is aimed at developing prototypes trained separately on homogeneous groups of patients in order utilize them in model predictive control strategy for defining patient-specific sepsis treatments.

Acknowledgments

This work was funded in part by DARPA grant DARPA-N66001-11-1-4183 negotiated by SSC Pacific (to ZO).

References

Ah-Pine, J. Normalized kernels as similarity indices. In *Advances in Knowledge Discovery and Data Mining*, pp. 362–373, 2010.

Ben-Hur, A, Elisseeff, A, and Guyon, I. A stability based method for discovering structure in clustered data. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 17:6–17, 2002.

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

Cock, K D and Moor, B D. Subspace angles between ARMA models. *Systems Control Letters*, 46:265–270, 2002.

Day, J. Using nonlinear model predictive control to find optimal therapy strategies to modulate inflammation. *Electrical Engineering*, 2010.

Hotchkiss, R S and Karl, I E. The pathophysiology and treatment of sepsis. *The New England journal of medicine*, 348:138–50, 2003.

Marshall, J C. The staging of sepsis: understanding heterogeneity in treatment efficacy. *Critical Care (London, England)*, 9(6):626–8, January 2005.

Martin, R J. A metric for ARMA processes, 2000.

Ng, A Y, Jordan, M I, and Weiss, Y. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems 14*, Advances in Neural Information Processing Systems, pp. 849–856. MIT Press, 2001.

Vishwanathan, S V N, Smola, A J, and Vidal, R. Binet-Cauchy kernels on dynamical systems and its application to the analysis of dynamic scenes. *International Journal of Computer Vision*, 73:95–119, 2006.