MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume **7**, Number **2**, April **2010**

doi:10.3934/mbe.2010.7.213

pp. **213–236**

A COMPARISON OF NONLINEAR FILTERING APPROACHES IN THE CONTEXT OF AN HIV MODEL

H. Thomas Banks, Shuhua Hu, Zackary R. Kenz
and Hien T. Tran

Center for Research in Scientific Computation
Raleigh, NC 27695-8205, USA

(Communicated by Azmy S. Ackleh)

Abstract. In this paper three different filtering methods, the Extended Kalman Filter (EKF), the Gauss-Hermite Filter (GHF), and the Unscented Kalman Filter (UKF), are compared for state-only and coupled state and parameter estimation when used with log state variables of a model of the immunologic response to the human immunodeficiency virus (HIV) in individuals. The filters are implemented to estimate model states as well as model parameters from simulated noisy data, and are compared in terms of estimation accuracy and computational time. Numerical experiments reveal that the GHF is the most computationally expensive algorithm, while the EKF is the least expensive one. In addition, computational experiments suggest that there is little difference in the estimation accuracy between the UKF and GHF. When measurements are taken as frequently as every week to two weeks, the EKF is the superior filter. When measurements are further apart, the UKF is the best choice in the problem under investigation.

1. **Introduction.** The modeling of the physiologic and immunologic response to HIV infection in humans is generating a substantial amount of research effort, and significant progress has been made in the treatment of HIV-infected patients. One of the most prevalent treatment strategies for acutely infected HIV patients is highly active anti-retroviral therapy (HAART) which utilizes two or more drugs. However, despite the success of HAART, patient-specific optimal schemes for its use need to be considered. Grave side effects of taking drugs, viral mutations and the high cost of drugs all motivate a substantial research effort in this area.

Open-loop control, a control that is pre-computed for a given dynamic model and initial conditions, is one technique that has been employed in a number of works (e.g., [2, 8]) to design treatment therapies for HIV patients. However, this technique may be inadequate for reasons such as poor patient adherence, increasing drug resistance due to virus mutations, and drug side effects as noted. Another approach that has been used to design dynamic HIV treatment therapies utilizes feedback controls such as those based on the state dependent Riccati equation (SDRE) approach used in [6] and on receding horizon control methodology in [9, 11]. A fundamental characteristic of feedback control is that it depends on the current state of the system.

2000 *Mathematics Subject Classification.* 65C30, 92D30, 93E10.

Key words and phrases. extended Kalman filter, unscented Kalman filter, Gauss-Hermite filter, HIV.

Hence, this method can be used to design *adaptive treatment schedules* for HIV patients based on the patient's current status (e.g., current CD4+ T cell count and viral load). Because HIV modeling generally involves partial observations and noisy measurements from combined compartments, the method by which the state is obtained at each sampling time is of special concern, and an efficient estimation technique is needed to develop a successful implementation of feedback control.

State and parameter estimation and the development of associated adaptive feedback control schemes in the setting investigated here offer challenges different from those in many engineering applications where high frequency uncensored observations are often the norm. In addition to low frequency sampling in longitudinal data sets (data points are often very expensive both in financial costs of associated assays as well as in emotional/physical costs to patients), the data itself is frequently censored from below due to limitations on assays in discriminating low values. Thus a number of challenges include development of filters for state and parameter estimation in the context of low frequency sampling and partial state observations that are censored. Successful efforts in these areas must be combined with feedback control of nonlinear dynamics which are often only approximate for patient response. Here we discuss a first step in development of a needed methodology by attempting to discover an appropriate filtering approach to use with partial state uncensored observations that are collected rather infrequently by usual engineering standards.

The model used in this paper was developed and validated as a predictive tool in [3], wherein two types of target cells, along with their corresponding infected states, free virus, and immune effector cells (CTL) are included as states in the model. Fitting with clinical data demonstrated that this model provides reasonable fits to numerous patient longitudinal data sets and has impressive predictive capability when comparing model simulations with parameters based on estimation using only half of the longitudinal observations. For computational ease in our presentation here, without loss of generality, we omit the noninfectious virus V_{NI} component from the model in [3]. This will not affect the dynamics of this model as it is completely decoupled from all the other compartments. The model we use is

$$\begin{aligned}
\dot{T}_1 &= \lambda_1 - d_1 T_1 - (1-\epsilon_1) k_1 V_I T_1 \\
\dot{T}_2 &= \lambda_2 - d_2 T_2 - (1-\mu\epsilon_1) k_2 V_I T_2 \\
\dot{T}_1^* &= (1-\epsilon_1) k_1 V_I T_1 - \delta T_1^* - m_1 E T_1^* \\
\dot{T}_2^* &= (1-\mu\epsilon_1) k_2 V_I T_2 - \delta T_2^* - m_2 E T_2^* \\
\dot{V}_I &= N_T 10^3 \delta (T_1^* + T_2^*) - [c + (1-\epsilon_1) 10^3 k_1 T_1 + (1-\mu\epsilon_1) 10^3 k_2 T_2] V_I \\
\dot{E} &= \lambda_E + b_E \frac{T_1^* + T_2^*}{T_1^* + T_2^* + K_b} E - d_E \frac{T_1^* + T_2^*}{T_1^* + T_2^* + K_d} E - \delta_E E,
\end{aligned} \tag{1}$$

with a state vector initial condition

$$(T_1(0), T_2(0), T_1^*(0), T_2^*(0), V_I(0), E(0))^T.$$

Here the state variables are T_1, the uninfected CD4+ T-cells; T_2, the uninfected target cells of a second kind; T_1^*, the infected CD4+ T-cells; T_2^*, the infected target cells of a second kind; V_I, the infectious virus; and E, the immune effectors. The units for T_1, T_2, T_1^*, T_2^* and E are cells/μl-blood, and the unit for V_I is RNA copies/ml-plasma. The factors 10^3 are introduced to convert between microliter (μl) and milliliter (ml) scales, preserving the units from some of the earlier published papers [1]. The particular target cells of a second kind are not specified, and

(according to [3, 5]) might be related to macrophages or brain cells or inactivated memory. The model also includes terms that model drug efficacy. The control term ϵ_1 represents the efficacy of a Reverse Transcriptase Inhibitor (RTI). For a more detailed description of model parameters and rationale for the model (1) we refer the reader to the article [3]. While there are somewhat improved and generalized versions of this model [5], the version we have chosen to use here is representative and more than adequate for demonstration of the behavior of the classes of filters we wish to compare.

We observe that our model (1) is nonlinear and hence we must explore nonlinear estimation methods. The Extended Kalman Filter (EKF) was used in [10] for state estimation for the HIV model of [3] (also without the V_{NI} component and without the scaling factor). Unlike the nonlinear least squares approach this technique does not require all the data at once. That is, this methodology only uses data as it is received and thus it can be used "online" to estimate the parameters in an adaptive approach. However, the EKF was found in [10] to have difficulty with state estimation even when the time span between measurements is only five days. This motivates a need to consider alternative filtering methods. Accordingly, we introduce the Gauss-Hermite Filter (GHF) and Unscented Kalman Filter (UKF) as possible alternatives. Even though all three of these methods are based on a Gaussian assumption (that is, the posterior distribution can be approximated by a Gaussian distribution), the process by which this approximation is obtained is different and therefore leads to different filter performance. It has been demonstrated (e.g., [12, 17]) that the performance of both the GHF and UKF are superior to the EKF in numerous other nonlinear problems.

Even though this paper might be considered an extension of the efforts in [10], we have made a significant modification here when applying these filters to our problem. Instead of using the model directly as in [10], we applied the filters to a *log-scaled version* of our model. This was done because log-transformation is a standard technique to render the transformed observation error more nearly normally distributed. This is in agreement with the experience that measurement errors in observations of CD4+ T-cells and viral loads can be assumed to be well approximated by log normal distributions. Moreover, from a numerical and computational point of view, by using a log-transformed system one can resolve the problem of states becoming unrealistically negative due to round-off errors. More importantly, all these filters are derived for systems where the states are defined on $\mathbb{R}^n$; this may lead to some difficulties when one apples the filters to a system such as (1) where the states are only defined in $\mathbb{R}^n_+$; log-scaling mitigates this potential difficulty.

The remainder of this paper is organized as follows. In Section 2 we give a brief introduction of the EKF, UKF and GHF and their application to a general nonlinear continuous system with discrete observations. In Section 3, the filters are implemented to estimate model states as well as model parameters from simulated noisy data, and compared in terms of estimation accuracy and computational time. We conclude the paper in Section 4 with some remarks and suggestions for future efforts.

2. **Filter descriptions.**

In this section we will use a capital italic letter to denote a random variable or random vector unless otherwise indicated, and use the corresponding small letter to denote its realization. A capital Roman letter is used to denote a non-random matrix. In addition, we may occasionally use the following

shorthand notations to ease the presentation: X_t for $X(t)$, $P_{t_{k+1}}^{t_k}$ for $\mathrm{P}^{t_k}(t_{k+1})$, etc. We will use $\mathcal{E}\{\cdot\}$ to denote the expectation of a random variable or vector.

Next we give a brief introduction of the three filters that we use, the EKF, GHF and UKF, and their application in the context of a general nonlinear continuous system

$$dX(t) = f(X(t),t)dt + \sigma(t)dB(t),\ t \geq t_0, \tag{2}$$

with discrete observations taken according to the measurement equation

$$Y_k = h(X(t_k),t_k) + V_k,\ t_{k+1} > t_k \geq t_0,\ k = 1,2,3,\ldots.$$

Here $X(t)$, assumed to satisfy an Ito stochastic differential equation, is a n dimensional column vector random variable for any fixed time t with X_{t_0} being normally distributed with mean $\hat{x}_{t_0}$ and covariance matrix P_{t_0}, i.e., $X_{t_0} \sim \mathcal{N}(\hat{x}_{t_0}, \mathrm{P}_{t_0})$, $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ is a non-random function of x and t, $B(t)$ is r-dimensional Brownian motion with $\mathcal{E}[(dB(t))(dB(t))^T] = \mathrm{Q}(t)dt$, where $\mathrm{Q}(t)$ is an $r \times r$ matrix, and $\sigma : \mathbb{R} \to \mathbb{R}^{n\times r}$ is a non-random function of time t. The random variable Y_k is a l-dimensional column vector, $h : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^l$ is a non-random function of x and t, and V_k are white Gaussian random sequence with covariance matrix R_k, i.e., $V_k \sim \mathcal{N}(0, \mathrm{R}_k)$, where R_k is a $l \times l$ matrix. In addition, we assume that X_{t_0}, $\{B(t)\}$ and $\{V_k\}$ are independent.

Let $\mathcal{Y}_\tau = \{y_k : t_k \leq \tau\}$ denote the information available by observing the process up to time τ, where y_k is a realization of Y_k. The filtering problem is to find the "best" estimate $\hat{x}^\tau(t)$ of $X(t)$ based on $\mathcal{Y}_\tau$. The "best" is understood in the sense of minimum mean-squared error (MMSE) for each fixed t

$$\hat{x}^\tau(t) = \arg\min_{\xi} \mathcal{E}\left\{(X(t)-\xi)^T(X(t)-\xi)|\mathcal{Y}_\tau\right\},$$

where the superscript in $\hat{x}^\tau$ is determined by the subscript in $\mathcal{Y}_\tau$. The MMSE estimate $\hat{x}^\tau(t)$ of $X(t)$ based on $\mathcal{Y}_\tau$ is the conditional mean

$$\hat{x}^\tau(t) = \mathcal{E}\left\{X(t)|\mathcal{Y}_\tau\right\},$$

and the covariance matrix of this estimate, denoted by $\mathrm{P}^\tau(t)$, is given by

$$\mathrm{P}^\tau(t) = \mathcal{E}\{(X(t)-\hat{x}^\tau(t))(X(t)-\hat{x}^\tau(t))^T|\mathcal{Y}_\tau\}.$$

Hence, to determine the best estimate $\hat{x}^\tau(t)$ of $X(t)$ based on $\mathcal{Y}_\tau$ as well as the conditional covariance matrix, we need to find the conditional probability density function $p(x,t|\mathcal{Y}_\tau)$ of $X(t)$ at any time t. However, this density $p(x,t|\mathcal{Y}_\tau)$ satisfies the Fokker-Planck equation

$$\frac{\partial p(x,t|\mathcal{Y}_\tau)}{\partial t} + \sum_{i=1}^{n} \frac{\partial (p(x,t|\mathcal{Y}_\tau) f_i(x,t))}{\partial x_i} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2[(\sigma(t)\mathrm{Q}(t)\sigma^T(t))_{ij} p(x,t|\mathcal{Y}_\tau)]}{\partial x_i \partial x_j},$$

between any observation period $t_k < t < t_{k+1}$, and then is updated using Bayes formula with a measurement due to a new observation at t_{k+1}. Thus in general, it can not be obtained in closed form. Therefore in many applications it is conventionally assumed that the distribution is Gaussian so that the distribution is completely parameterized by just the mean and covariance; this is precisely the assumption made in the EKF, GHF and UKF.

All three filters, the EKF, UKF and GHF, utilize a "predictor-corrector" implementation. Given current state estimate $\hat{x}_{t_k}^{t_k} = \hat{x}^{t_k}(t_k)$ and covariance matrix estimate $\mathrm{P}_{t_k}^{t_k} = \mathrm{P}^{t_k}(t_k)$ at time t_k, the filter first predicts the states at time t_{k+1} by using only model dynamics to obtain the predicted quantities $\hat{x}_{t_{k+1}}^{t_k} = \hat{x}^{t_k}(t_{k+1})$

and $\mathrm{P}_{t_{k+1}}^{t_k} = \mathrm{P}^{t_k}(t_{k+1})$. When new data y_{k+1} is available at time t_{k+1}, a linear update rule is specified to obtain the updated quantities $\hat{x}_{t_{k+1}}^{t_{k+1}}$ and $\mathrm{P}_{t_{k+1}}^{t_{k+1}}$, where the weights are chosen to minimize the mean squared error of the estimate.

2.1. **The extended Kalman filter.** In the EKF, the expected values occurring in time and measurement updates are computed by linearization of the system function f and measurement function h at the current state estimate.

To start the filter algorithm, we set $k = 0$ initially and set $\hat{x}_{t_0}^{t_0} = \hat{x}_{t_0}$ and $\mathrm{P}_{t_0}^{t_0} = \mathrm{P}_{t_0}$ for given initial conditions $\hat{x}_{t_0}$, P_{t_0}. We then compute the predicted state $\hat{x}_{t_{k+1}}^{t_k} = \hat{x}^{t_k}(t_{k+1})$, by solving the following ordinary differential equations

$$\frac{d\hat{x}^{t_k}(t)}{dt} = f(\hat{x}^{t_k}(t), t), \quad t_k < t < t_{k+1},$$
$$\hat{x}^{t_k}(t_k) = \hat{x}_{t_k}^{t_k}.$$

We simultaneously compute the predicted error covariance matrix $\mathrm{P}_{t_{k+1}}^{t_k} = \mathrm{P}^{t_k}(t_{k+1})$, by solving the ordinary differential equations in the time interval (t_k, t_{k+1})

$$\frac{d\mathrm{P}^{t_k}(t)}{dt} = \mathrm{P}^{t_k}(t)\left[\nabla f(\hat{x}^{t_k}(t), t)\right]^T + \nabla f(\hat{x}^{t_k}(t), t)\mathrm{P}^{t_k}(t) + \sigma(t)\mathrm{Q}(t)\sigma^T(t),$$
$$\mathrm{P}^{t_k}(t_k) = \mathrm{P}_{t_k}^{t_k}.$$

Here Q and σ are the parameters of the noise in the stochastic process $X(t)$ as described above. Once this prediction step is complete, we can incorporate the new data information at time t_{k+1}. We compute the *updated state* and *updated error covariance matrix* with the observation y_{k+1} by computing solutions to the equations

$$\hat{x}_{t_{k+1}}^{t_{k+1}} = \hat{x}_{t_{k+1}}^{t_k} + \mathrm{G}_{k+1}(y_{k+1} - h(\hat{x}_{t_{k+1}}^{t_k}, t_{k+1})),$$

and

$$\mathrm{P}_{t_{k+1}}^{t_{k+1}} = \left[\mathrm{I} - \mathrm{G}_{k+1}\nabla h(\hat{x}_{t_{k+1}}^{t_k}, t_{k+1})\right]\mathrm{P}_{t_{k+1}}^{t_k},$$

respectively, where G_{k+1} is defined by

$$\mathrm{G}_{k+1} = \mathrm{P}_{t_{k+1}}^{t_k}\left(\nabla h(\hat{x}_{t_{k+1}}^{t_k}, t_{k+1})\right)^T\left[\nabla h(\hat{x}_{t_{k+1}}^{t_k}, t_{k+1})\mathrm{P}_{t_{k+1}}^{t_k}\left(\nabla h(\hat{x}_{t_{k+1}}^{t_k}, t_{k+1})\right)^T + \mathrm{R}_{k+1}\right]^{-1}.$$

Recall that R_k is the covariance matrix in the noise for the observation process. We thus have the updated state values and covariance matrix at time t_{k+1}. We then move to the next time step, so we increment k by 1 and return to the predictor step.

The EKF has been successfully applied to numerous nonlinear filtering problems in the engineering and physical sciences. However, its performance can be extremely poor when nonlinear dynamics are significant between observations. Moreover, it is not a good choice for some application problems where Jacobian matrices are difficult to calculate (or may not even exist). For more discussions on the EKF, the interested reader can consult [13] among numerous other texts.

2.2. **The unscented Kalman filter.** Instead of linearization of the system function f and measurement function h at current state estimates as required by the EKF, in order to compute the expected values occurring in time and measurement updates, the UKF is based on the principle that a discrete distribution composed of a set of deterministically chosen sampled points with the corresponding weights can be used to approximate the standard normal distribution. It is founded on

intuition: "it is easier to approximate a probability density function than it is to approximate an arbitrary nonlinear function". Given a function $\tilde{f}(\tau)$, the UKF approximates integrals by

$$\int_{\mathbb{R}^n} \tilde{f}(\tau)\frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}|\tau|^2}d\tau \approx \sum_{i=1}^{N}\tilde{f}(q_i)w_i. \tag{3}$$

Here $N = 2n+1$. The values for the sampling points q_i and the weights w_i (see [14, 15, 16] for more details) are defined by

$$q_i = \begin{cases} \sqrt{n+\kappa}e_i, & 1 \le i \le n, \\ -q_{i-n}, & n+1 \le i \le 2n, \\ 0, & i = 2n+1, \end{cases} \quad \text{and} \quad w_i = \begin{cases} \dfrac{1}{2(n+\kappa)}, & 1 \le i \le 2n, \\ \dfrac{2\kappa}{2(n+\kappa)}, & i = 2n+1, \end{cases}$$

where e_i is the ith unit vector in $\mathbb{R}^n$, and $\kappa \in \mathbb{R}$ can be any number providing $n+\kappa \neq 0$. The variable κ provides an extra degree of freedom to "fine tune" the higher order moments of the approximation and can be used to reduce the overall predication error. This discrete distribution has the same first, second, and higher odd moments as the standard normal distribution.

The UKF was originally designed for a discrete system with discrete observations. To apply to the continuous model (2) we need to discretize the model over each time frame (t_k, t_{k+1}). We first subdivide the interval (t_k, t_{k+1}) into the M subintervals $(t_k+(j-1)\delta_t, t_k+j\delta_t)$ where $1 \le j \le M$ and $\delta_t = (t_{k+1}-t_k)/M$. Then we use an Euler method approximation

$$X^{k,j} = X^{k,j-1} + \delta_t f(X^{k,j-1}, t_k+(j-1)\delta_t) + \sigma(t_k+(j-1)\delta_t)W^{k,j},\ 1 \le j \le M \tag{4}$$

with $W^{k,j} = B(t_k+j\delta_t) - B(t_k+(j-1)\delta_t)$, where $X^{k,j}$ denotes the finite difference approximation of $X(t_k+j\delta_t)$.

To begin the filter algorithm, we set $k=0$ and set $\hat{x}_{t_0}^{t_0} = \hat{x}_{t_0}$ and $\mathrm{P}_{t_0}^{t_0} = \mathrm{P}_{t_0}$. We then compute the predicted state $\hat{x}_{t_{k+1}}^{t_k} = \hat{x}^{t_k}(t_{k+1})$, and the predicted covariance matrix $\mathrm{P}_{t_{k+1}}^{t_k} = \mathrm{P}^{t_k}(t_{k+1})$ through the time frame $t_k < t < t_{k+1}$. To begin the predictor step, we set $j=0$, pick $M > 0$, and let $\delta_t = (t_{k+1}-t_k)/M$. Then compute the factorization $\mathrm{P}_{t_k+(j-1)\delta_t}^{t_k} = \mathrm{S}^T\mathrm{S}$ using the Cholesky decomposition. Set $\tilde{x}_i = \mathrm{S}^T q_i + \hat{x}_{t_k+(j-1)\delta_t}^{t_k}$. We then compute the following sums to obtain the states and the covariance matrix

$$\begin{aligned} \hat{x}_{t_k+j\delta_t}^{t_k} &= \sum_{i=1}^{N}(\tilde{x}_i + \delta_t f(\tilde{x}_i, t_k+(j-1)\delta_t))w_i, \\ \mathrm{P}_{t_k+j\delta_t}^{t_k} &= \mathrm{P}_{xx} + \sigma(t_k+(j-1)\delta_t)\mathrm{Q}(t_k+(j-1)\delta_t)\sigma(t_k+(j-1)\delta_t)^T. \end{aligned}$$

where P_{xx} is defined by

$$\mathrm{P}_{xx} = \sum_{i=1}^{N}(\tilde{x}_i+\delta_t f(\tilde{x}_i, t_k+(j-1)\delta_t) - \hat{x}_{t_k+j\delta_t}^{t_k})(\tilde{x}_i+\delta_t f(\tilde{x}_i, t_k+(j-1)\delta_t) - \hat{x}_{t_k+j\delta_t}^{t_k})^T w_i.$$

After the sums have been computed, we increment j and repeat the preceding steps beginning with the factorization of $\mathrm{P}_{t_k+(j-1)\delta_t}^{t_k}$. We continue iterating until $j=M$, which will yield $\hat{x}_{t_k+1}^{t_k}$ and $\mathrm{P}_{t_{k+1}}^{t_k}$.

Once this prediction step is complete, we can incorporate the new data information at time t_{k+1} to update the state and covariance matrix. First, we compute the

factorization $\mathrm{P}_{t_{k+1}}^{t_k} = \tilde{\mathrm{S}}^T\tilde{\mathrm{S}}$ as before and set $\tilde{x}_i = \tilde{\mathrm{S}}^T q_i + \hat{x}_{t_{k+1}}^{t_k}$. Then compute

$$\begin{aligned} \hat{x}_{t_{k+1}}^{t_{k+1}} &= \hat{x}_{t_{k+1}}^{t_k} + \mathrm{L}_{k+1}(y_{k+1} - z_{k+1}), \\ \mathrm{P}_{t_{k+1}}^{t_{k+1}} &= \mathrm{P}_{t_{k+1}}^{t_k} - \mathrm{L}_{k+1}\mathrm{P}_{xz}^T, \end{aligned}$$

where

$$\begin{aligned} z_{k+1} &= \sum_{i=1}^{N} h(\tilde{x}_i, t_{k+1})w_i, \\ \mathrm{P}_{xz} &= \sum_{i=1}^{N} (\tilde{x}_i - \hat{x}_{t_{k+1}}^{t_k})(h(\tilde{x}_i, t_{k+1}) - z_{k+1})^T w_i, \\ \mathrm{P}_{zz} &= \sum_{i=1}^{N} (h(\tilde{x}_i, t_{k+1}) - z_{k+1})(h(\tilde{x}_i, t_{k+1}) - z_{k+1})^T w_i, \\ \mathrm{L}_{k+1} &= \mathrm{P}_{xz}(\mathrm{R}_{k+1} + \mathrm{P}_{zz})^{-1}. \end{aligned}$$

Following the update step, we have the updated state values and covariance matrix at time t_{k+1}. We then move to the next time step, so we increment k by 1 and return to the predictor step above.

We note that the UKF is easier to implement than the EKF as it does not require the calculation of Jacobian matrices. However, the UKF may require some "fine tuning" in order to prevent the propagation of a non-positive definite covariance matrix for a state vector dimension higher than three [4]. For a more extensive treatment of the UKF, see [14, 15, 16].

2.3. **The Gauss-Hermite filter.** Similar to the UKF approach, the GHF does not linearize the system function and measurement function at the current state estimate to obtain the expectation values occurring in time and measurement updates. However, the expectation values in the GHF are calculated by using a Gaussian-Hermite quadrature rule instead of by approximating the standard normal distribution as used in the UKF. Given a function $\tilde{f}(\tau)$, the quadrature rule for expected values is expressed by

$$\int_{\mathbb{R}^n} \tilde{f}(\tau)\frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}|\tau|^2} d\tau \approx \sum_{i_1=1}^{m} \cdots \sum_{i_n=1}^{m} \tilde{f}(\varrho_{i_1}, \varrho_{i_2}, \ldots, \varrho_{i_n})\omega_{i_1}\omega_{i_2}\cdots\omega_{i_n}, \qquad (5)$$

where m is the number of quadrature points used in a one-dimensional quadrature rule. The quadrature points and their corresponding weights are calculated as follows. Let $\mathrm{A} \in \mathbb{R}^{m\times m}$ be a symmetric tridiagonal matrix with zero diagonal elements and its $(j, j+1)$th element defined by $\mathrm{A}_{j,j+1} = \sqrt{\frac{j}{2}}$, $j = 1, 2, \ldots, m-1$. Then A has m eigenvalues, which are denoted by ρ_j, $j = 1, 2, \ldots, m$. Let v_j be the normalized eigenvector of A corresponding to the eigenvalue ρ_j, $j = 1, 2, \ldots, m$. Then the quadrature points and its corresponding weights are given by

$$\varrho_j = \sqrt{2}\rho_j, \quad \omega_j = (v_{j1})^2, \quad j = 1, 2, \ldots, m,$$

where v_{j1} is the first element of normalized eigenvector v_j. Hence, in order to evaluate the integral in (5) we need m^n-point function evaluations. To have a notation consistent with (3), we may rewrite the right side of (5) into one sum

notation as in (3), and express it as

$$\int_{\mathbb{R}^n} \tilde{f}(\tau)\frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}|\tau|^2}d\tau \approx \sum_{i=1}^{N}\tilde{f}(q_i)w_i,$$

where $N = m^n$, $q_i = (\varrho_{i_1}, \varrho_{i_2}, \ldots, \varrho_{i_n})$ and $w_i = \omega_{i_1}\omega_{i_2}\cdots\omega_{i_n}$, $i = 1, 2, \ldots, N$.

The GHF was also originally designed for a discrete system with discrete observations. Hence, in order to apply the GHF to a continuous system with discrete observations, we again need to discretize the continuous model, and we will use the same approximation scheme as we used in the UKF. The algorithm for the GHF is exactly the same as that for the UKF except for the choice of q_i and its corresponding weight w_i in evaluating the integral.

Like the UHF, the GHF does not require the calculation of Jacobian matrices. However, the obvious disadvantage of the GHF is that the required number of points to evaluate the integral scales geometrically with the number of dimensions. For more detailed information on the GHF, the interested reader is referred to [12].

3. **Numerical simulations.** As mentioned in the introduction, we will apply the EKF, GHF and UKF to the log-scaled HIV system instead of the original system. We first rewrite the HIV model (1) as vector system

$$\dot{\bar{x}} = g(\bar{x}; \bar{\theta}), \tag{6}$$

where $\bar{x} = (T_1, T_2, T_1^*, T_2^*, V_I, E)^T$, and $\bar{\theta}$ is the vector for model parameters given by

$$\bar{\theta} = (\lambda_1, d_1, \epsilon_1, k_1, \lambda_2, d_2, \mu, k_2, \delta, m_1, m_2, N_T, c, \lambda_E, b_E, K_b, d_E, K_d, \delta_E).$$

The available measurements (based on our current collaborative efforts with clinical researchers) are assumed to include total CD4+ T-cell count, number of viral load copies, and immune effector T-cell count. In this model (6), the total CD4+ T-cell counts are represented by $\bar{x}_1(t;\bar{\theta}) + \bar{x}_3(t;\bar{\theta})$ (i.e., $T_1(t;\bar{\theta}) + T_1^*(t;\bar{\theta})$), viral load copies are represented by $\bar{x}_5(t;\bar{\theta})$ (i.e., $V_I(t;\bar{\theta})$), and immune effector T-cell counts are represented by $\bar{x}_6(t;\bar{\theta})$ (i.e., $E(t;\bar{\theta})$).

Because the values of model parameters are in dramatically different scales (varied from 10^{-7} to 10^2), we also transform all the model parameters into their log-scaled counterparts except parameters ϵ_1 and μ. The values of these two parameters can be zero and they are already on the scale of 10^{-1}. Let $x_i = \log_{10}\bar{x}_i$, $i = 1, 2, \ldots, 6$, $\theta_i = \log_{10}\bar{\theta}_i$ for $i = 1, 2, 4, 5, 6, 8, 9, 10, \ldots, 19$, and $\theta_i = \bar{\theta}_i$ for $i = 3, 7$. Then we have

$$\dot{x} = f(x; \theta), \tag{7}$$

where

$$f_i(x) = \frac{10^{-x_i}}{\ln(10)}g_i(\bar{x}, \tilde{\theta}), \quad i = 1, 2, \ldots, 6$$

with $\tilde{\theta} = (10^{\theta_1}, 10^{\theta_2}, \theta_3, 10^{\theta_4}, 10^{\theta_5}, 10^{\theta_6}, \theta_7, 10^{\theta_8}, \ldots, 10^{\theta_{19}})$. The observation process is given by

$$Y_k = h(x(t_k)) + V_k,$$

where h is defined by

$$h(x) = \begin{bmatrix} \log_{10}(10^{x_1} + 10^{x_3}) \\ x_5 \\ x_6 \end{bmatrix},$$

and $\{V_k\}$ is a white Gaussian sequence with covariance matrix R_k, i.e., $V_k \sim \mathcal{N}(0, \mathrm{R}_k)$. However, in our simulations (for data generation and filter computation), we use the following stochastic model instead of (7)

$$dX(t) = f(X(t); \theta)dt + dB(t), \tag{8}$$

that is, we append white noise to (7) with $\mathcal{E}[(dB(t))(dB(t))^T] = \mathrm{Q}dt$. The observation process is then given by

$$Y_k = h(X(t_k)) + V_k. \tag{9}$$

The additional model noise is included because it can reduce the chance of the covariance matrix being non-positive definite in both the UKF and GHF. To ensure this new model has similar dynamics as (7) (i.e., to provide only small perturbations to the states) we choose $\mathrm{Q} = 10^{-6}\mathrm{I}$.

We point out that there are some model parameters that are patient specific in that the values of these parameters may vary from patient to patient. Hence, we do not know the values of these parameters in advance. In this effort we also wish to test the performance of these filters in adaptively estimating the model parameters as well as model states. To do this, we append the parameters to the model (8) as additional states. We obtain

$$\begin{aligned} dX(t) &= f(X(t); \theta)dt + dB(t), \\ \dot{\theta_e} &= 0, \end{aligned} \tag{10}$$

where θ_e are the model parameters that are to be estimated.

Simulated data sets were generated in the following manner. We used (8) or (10) with θ set to the "true" values in Table 1 below to generate N realizations (we used $N = 20$ for the results reported here) of the state vector $X(t)$. We used these realizations in (9) along with N different realizations for V_k to generate realizations of Y_k. This provides N longitudinal data sets for the state vector and corresponding observations. We used each of these in the filter algorithms to generate estimates or realizations of the filters. These were then compared using the N realizations by taking the average (over the N differences) root mean square (RMS) of the difference between data and estimated states. In order to demonstrate the differences in filter state estimation based on frequency of observations, we used the simulated data with measurements taken 1, 7, 14 and 28 days apart, respectively. The filter results are then examined and compared using the average RMS error for each observed state and each model compartment. The average RMS error for the jth model compartment $\mathrm{RMS}_{m,j}$ at time t_k is defined by

$$\mathrm{RMS}^k_{m,j} = \Big(\frac{1}{N}\sum_{i=1}^{N}\big[x^i_j(t_k) - \hat{x}^i_j(t_k)\big]^2\Big)^{1/2}, \tag{11}$$

which is based on the N different simulation runs. Here the subscript j denotes the jth component of the "true" noisy state vector $x(t_k)$ (the data) and the corresponding state estimate $\hat{x}(t_k)$, and the superscript i denotes the ith simulation run. The average RMS error for the jth observed state $\mathrm{RMS}_{o,j}$ at time t_k is defined by

$$\mathrm{RMS}^k_{o,j} = \Big(\frac{1}{N}\sum_{i=1}^{N}\big[h_j(x^i(t_k)) - h_j(\hat{x}^i(t_k))\big]^2\Big)^{1/2}, \tag{12}$$

where the subscript j denotes the jth component of the observation function $h(x(t_k))$, $j = 1, 2, 3$. Since real time feedback is also important, we examine the computational expense of each filter.

3.1. **Generation of simulated data.** In order to test our three filters, we needed to create simulated data based on a "true" state value. To do this, we used Euler method approximation (the method used in the predictor step of the UKF), which guarantees nontrivial system noise in the simulated data, to numerically solve model equations (8) through a specified time span (0 to 364 days in our simulation runs) with time mesh size chosen to be 0.001, and then used (9) to obtain simulated data with additional observation noise at the relevant observation times (every 1, 7, 14, or 28 days) with constant noise covariance matrix $\mathrm{R}_k \equiv \mathrm{diag}([0.1^2, 0.25^2, 0.07^2])$. In general, the numbers in R_k are motivated by our experience with experimental data [1, 3, 5]. Here they are chosen based on the values of model states.

The initial states were set to be $x_0 = (\log_{10}(800), \log_{10}(3.198), -4, -6, 1, -2)^T$. The values of model parameters ("true values") used to generate the simulated data are given in Table 1. We simulated a treatment schedule for 364 days, with

parameter	value	unit	parameter	value	unit
λ_1	10	$\frac{\text{cells}}{\mu\text{l-blood} \cdot \text{day}}$	λ_2	0.03198	$\frac{\text{cells}}{\mu\text{l-blood} \cdot \text{day}}$
d_1	0.01	$\frac{1}{\text{day}}$	d_2	0.01	$\frac{1}{\text{day}}$
k_1	8e-7	$\frac{\text{ml-plasma}}{\text{copies} \cdot \text{day}}$	k_2	1e-4	$\frac{\text{ml-plasma}}{\text{copies} \cdot \text{day}}$
m_1	0.01	$\frac{\mu\text{l-blood}}{\text{cells} \cdot \text{day}}$	m_2	0.01	$\frac{\mu\text{l-blood}}{\text{cells} \cdot \text{day}}$
δ	0.7	$\frac{1}{\text{day}}$	c	13	$\frac{1}{\text{day}}$
μ	0.34	-	N_T	100	$\frac{\text{copies} \cdot \text{ml-blood}}{\text{cells} \cdot \text{ml-plasma}}$
λ_E	1e-3	$\frac{\text{cells}}{\mu\text{l-blood} \cdot \text{day}}$	δ_E	0.1	$\frac{1}{\text{day}}$
b_E	0.3	$\frac{1}{\text{day}}$	d_E	0.25	$\frac{1}{\text{day}}$
K_b	0.1	$\frac{\text{cells}}{\mu\text{l-blood}}$	K_d	0.5	$\frac{\text{cells}}{\mu\text{l-blood}}$
ϵ_1	0.8	-			

TABLE 1. Values of parameters in the HIV model.

treatment off for the first 30 days, then alternating on for 15 days and off for 45 days, and then off treatment for the last 34 days of the year. This is shown in Figure 1, and also appears at the bottom of every graph of states or parameters that are

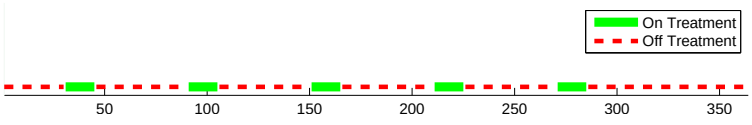

FIGURE 1. The treatment schedule used in the experiments.

run on data with treatment. The data is thus fairly dynamic, i.e., the values of states are within a fairly wide range, for example, the value of V_I varies from 10^{-1} to 10^7.

3.2. **Examination of simulation results.** We will examine the observed states average RMS error defined in (12) as well as the model compartments average RMS error defined in (11), each when measurement data is taken 1 day, 7 days, 14 days or 28 days apart, respectively. Each of the simulated data sets for a given sampling frequency was taken from the same N "true" data sets, but taken at the specified measurement frequency.

In order to test the filters under "poor" starting conditions with relative uncertainty, we set $\hat{x}_{t_0}^{t_0} = 0.6x_0$ and $\mathrm{P}_{t_0}^{t_0} = 0.01\mathrm{I}$ for all the simulation results presented in this section. In both the UKF and GHF, we set $\delta_t = 0.005$ in the discretization scheme (4). In some of our efforts, we also included parameter estimation as part of the filtering process; the results are illustrated in Section 3.2.2. The parameters that are estimated were given initial values of 90% of the value in Table 1, and the variance of each of these parameters was set to be 0.005 multiplied by the value of the parameter squared, and then these additional values were added into an extended P matrix.

3.2.1. *State Estimation.* In this section, numerical results are obtained by applying the filtering algorithms to model (8) for state-only estimation with measurements frequency varying from 1 day, 7 days, 14 days to 28 days.

Figure 2 depicts the results for the average RMS error of observed states defined by (12) (i.e., scaled CD4+ T-cells, scaled viral load V_I, and scaled immune effector T-cells E), where the plots in the left column, middle column and right column are the results obtained by using the EKF, UKF and GHF, respectively. Even though this figure reveals that the UKF and GHF appear to adjust to the data slightly faster than the EKF, the average RMS errors obtained by the EKF are much smaller than those obtained by the UKF and GHF after the filters adjust to the data (around day 100) for all the measurement frequencies that we have investigated. In addition, we observe that there is not much difference in terms of estimation accuracy between the UKF and GHF as the average RMS errors obtained by them are almost on the same level (this observation can be further confirmed in Tables 2 and 3 shown below). Also note that the UKF and GHF appear to oscillate, whereas the EKF is much smoother. We believe that this is due to the discretization of the continuous model during the predictor step of the UKF and GHF. In addition, Figure 2 indicates that the measurement frequency has no significant effect on the performance of the filters after day 100. This is probably because for this case the nonlinearity in the dynamics is not so important.

We also examine and compare the filtering results obtained for the first observed state (scaled CD4+ T-cells) and each model compartment by averaging its average RMS errors over time (this obtained error will be termed as ARMS error in the following presentation). The ARMS errors for the observed state (scaled CD4+ T-cells) and the jth model compartment are defined by

$$\mathrm{ARMS}_{o,1} = \frac{1}{n_t}\sum_{k=1}^{n_t}\mathrm{RMS}_{o,1}^{k}, \text{ and } \mathrm{ARMS}_{m,j} = \frac{1}{n_t}\sum_{k=1}^{n_t}\mathrm{RMS}_{m,j}^{k}, j = 1, 2, \ldots, 6 \quad (13)$$

respectively, where n_t denotes the total number of observation time points. The

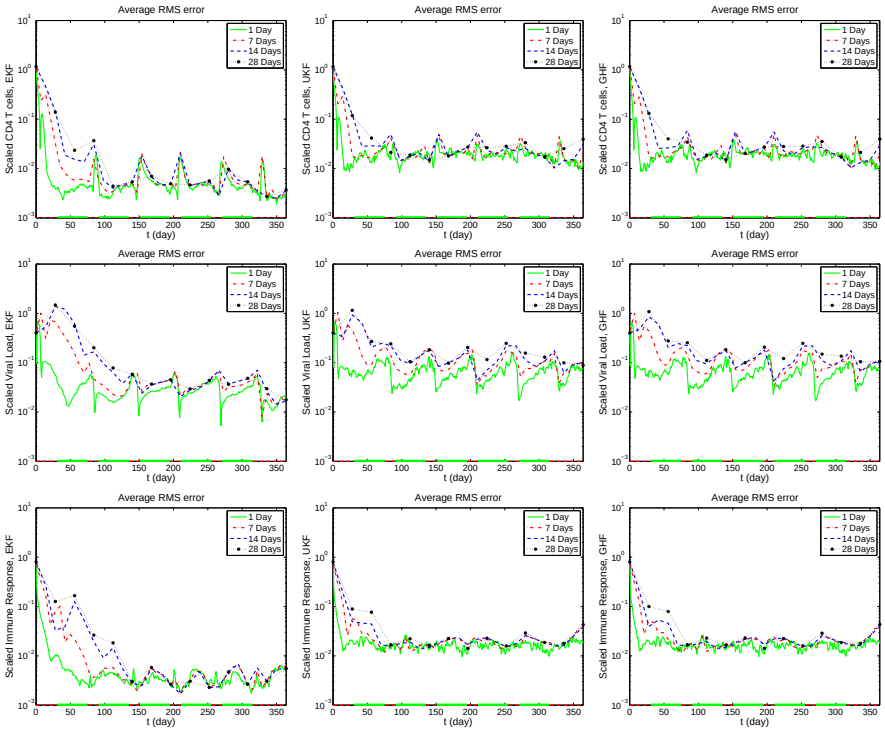

FIGURE 2. State estimation with different measurement frequencies for the observed states: scaled CD4+ T-cells (top row), scaled viral load V_I (middle row), and scaled immune effectors E (bottom row); average RMS errors for each observed state as obtained using the EKF (left column), UKF (middle column) and GHF (right column).

measurement frequency	filtering algorithm	scaled T_1	scaled T_2	scaled T_1^*	scaled T_2^*	scaled V_I	scaled E
1 day	EKF	0.0178	0.0263	0.0464	0.0356	0.0396	0.0091
	UKF	0.0312	0.0404	0.0814	0.0594	0.0732	0.0206
	GHF	0.0311	0.0403	0.0813	0.0600	0.0731	0.0206
7 days	EKF	0.0487	0.1241	0.1492	0.1360	0.1136	0.0336
	UKF	0.0642	0.1264	0.1842	0.1644	0.1516	0.0427
	GHF	0.0643	0.1289	0.1840	0.1680	0.1505	0.0428
14 days	EKF	0.0761	0.1355	0.2531	0.1980	0.1964	0.0529
	UKF	0.0910	0.1253	0.2546	0.2253	0.1996	0.0583
	GHF	0.0916	0.1342	0.2563	0.2253	0.2005	0.0588
28 days	EKF	0.1050	0.1493	0.3117	0.2403	0.2180	0.0836
	UKF	0.1184	0.1449	0.3440	0.2780	0.2495	0.0861
	GHF	0.1196	0.1425	0.3434	0.2793	0.2488	0.0872

TABLE 2. The ARMS error for each model compartment with state estimation.

ARMS errors for the model compartments obtained at different measurement frequency are given in Table 2, while those for the first observed state (scaled CD4+ T-cells) are listed in Table 3. From these two tables, we see that for all the measure-

filtering algorithm	measurement frequency			
	1 day	7 days	14 days	28 days
EKF	0.0145	0.0407	0.0730	0.1010
UKF	0.0271	0.0533	0.0855	0.1136
GHF	0.0270	0.0532	0.0863	0.1155

TABLE 3. The ARMS error for the observed scaled CD4+ T-cells with state estimation.

ment frequencies that we have investigated there is little difference in the estimation accuracy between the UKF and GHF as the ARMS errors obtained by them are pretty much similar to each other. In addition, we observe that the ARMS errors obtained by EKF are in general smaller than those obtained by the UKF and GHF. This means that the EKF performs better than both of the UKF and GHF in terms of estimation accuracy.

To have a good idea of the dynamics of the average RMS errors for all the target cell model compartments (scaled version of T_1, T_1^*, T_2 and T_2^*) with the state-only estimation, we present in Figure 3 the filtering results obtained for measurements taken 1 day apart with time plots of the average RMS errors for these model compartments. Interested readers can refer to Section 3.2.1 and the Appendix in [7] for similar plots obtained for other measurement frequencies (7, 14, 28 days).

In order to also have a good idea of the dynamics of the data, observed states and model compartments, as well as the performance of each filter in estimation of the observed states and model compartments, we present the results with the time plots for the estimated values versus the true values obtained by applying the filters to a typical data set in Figures 4 and 5. For demonstration purposes, here we only present the results obtained for measurements taken 1 day and 14 days apart. Interested readers can refer to Section 3.2.1 and the Appendix in [7] for similar plots obtained for measurements taken 7 days and 28 days apart. The plots in the left column of Figure 4 are for the observed states (CD4+ T-cells, viral load level V_I and immune effector T-cells E) with measurements taken 1 day apart, and those in the right column are for these observed states with measurements 14 days apart. Figure 5 contains the results for model compartments (T_1, T_2, T_1^* and T_2^*), where the plots in the left column are obtained with measurements 1 day apart and those in the right column for measurements 14 days apart. From these two figures, we can see that all three filters give pretty good estimates for both observed states and model compartments. Moreover this conclusion is also true when measurements are taken 7 days and 28 days apart, as has been presented in [7].

3.2.2. *State and Parameter Estimation.* In this section, numerical results obtained by applying the filtering algorithms to model (10) for state estimation as well as parameter estimation of λ_1 and N_T with different measurement frequencies (1 day, 7 days, 14 days and 28 days) are given. The results presented in this section are typical for different combinations of parameters that we estimated in a number of simulation studies.

In Figure 6 we depict the results for the average RMS error of observed states (scaled CD4+ T-cells, scaled viral load V_I, and scaled immune effector T-cells E),

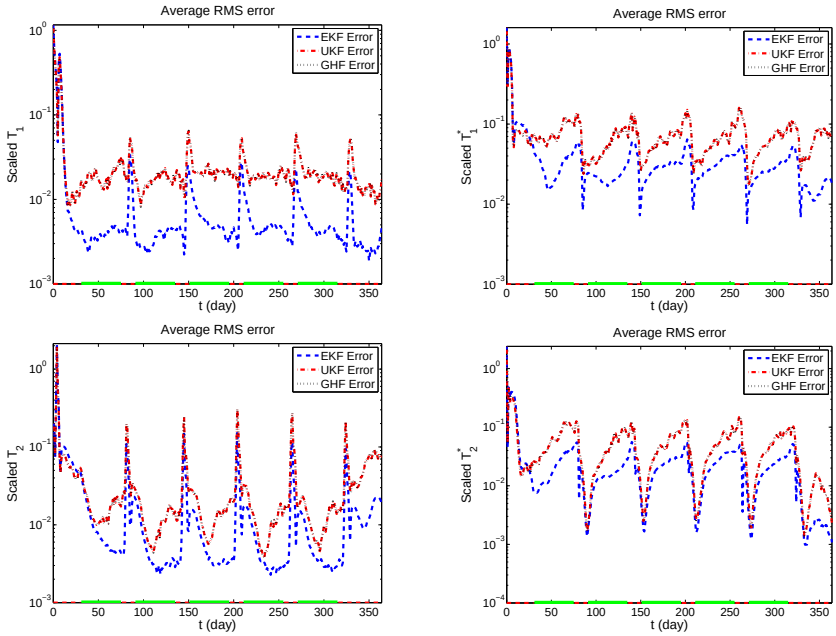

FIGURE 3. Average RMS errors for model compartments (scaled version of T_1, T_1^*, T_2 and T_2^*) with measurements 1 day apart for state-only estimation.

where the plots in the left column, middle column and right column are the results obtained by using the EKF, UKF and GHF, respectively. From this figure, we see that with measurements taken at 1 day apart the average RMS errors for the observed states obtained by using the EKF are still much smaller than those obtained by using the UKF and GHF. However, as the measurements are taken less frequently, the performances of the EKF deteriorate much more rapidly than do those of either the UKF or GHF. Hence, the measurement frequency has a significant effect on the estimation accuracy of the EKF, but has less effect on the GHF and UKF. Thus with parameter estimation included in the filtering process, the behavior of the filters is significantly different from that we observed in Section 3.2.1. This is likely because the nonlinearities in this new problem (state plus parameter estimation) become more significant as the measurement frequency decreases, which renders the EKF less effective than either the GHF and UKF. This type of decline in EKF performance is observed in numerous other nonlinear problems.

We also examine and compare the filtering results for each model compartment, each estimated parameter and observed state (scaled CD4+ T-cells) by averaging its average RMS errors over time (defined the same as that in (13)). The ARMS errors for the observed state (scaled CD4+ T-cells) with different measurement frequencies are summarized in Table 4, and those for model compartments (scaled version of T_1, T_2, T_1^*, T_2^*, V_I and E) and estimated parameters (scaled version of λ_1 and N_T) are listed in Table 5. From these two tables, we can see that for all the measurement frequencies that we have investigated the performance of the UKF is still quite similar to that of the GHF as the difference in the ARMS errors obtained by them is small. We also observed that with measurements taken 1 day apart,

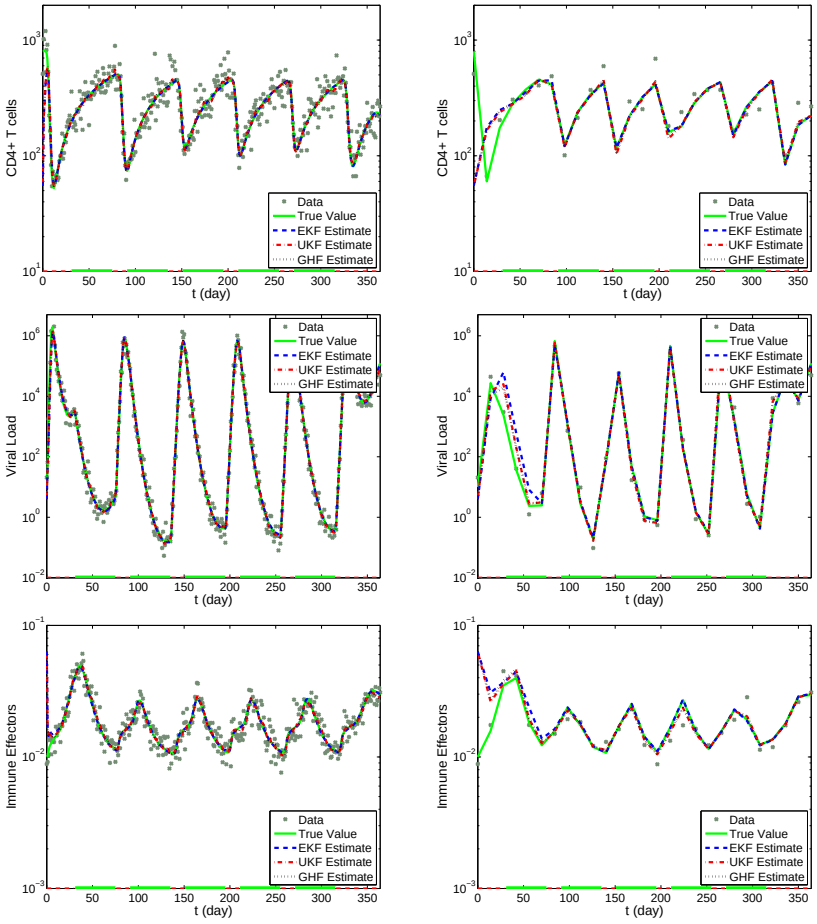

FIGURE 4. The estimated values versus the true values for the total CD4+ T-cell count, viral load level (V_I) and immune effector T-cell count (E) obtained by applying the filters to a typical data set for the state estimation. (left): results are obtained with measurements 1 day apart; (right): results are obtained with measurements 14 days apart.

filtering algorithm	measurement frequency			
	1 day	7 days	14 days	28 days
EKF	0.0229	0.0773	0.1177	0.1541
UKF	0.0307	0.0739	0.1028	0.1350
GHF	0.0308	0.0723	0.1167	0.1373

TABLE 4. The ARMS error for the observed state, scaled CD4+ T-cells, with state-and-parameter estimation, where the measurement frequency varies from 1 day, 7 days, 14 days to 28 days.

the ARMS errors obtained by EKF are much smaller than those obtained by the UKF and GHF. However, when measurements are taken 7 and 14 days apart, we

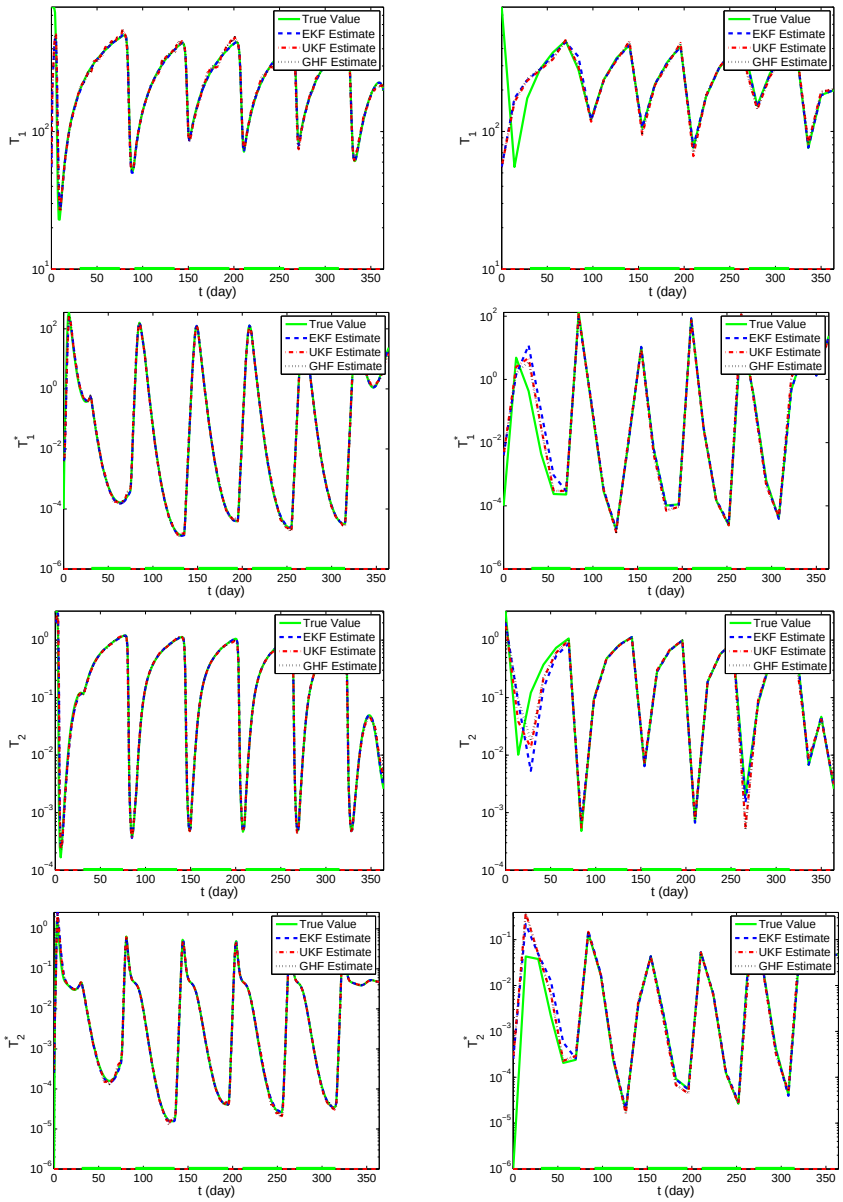

FIGURE 5. The estimated values versus the true values for model compartments T_1, T_1^*, T_2 and T_2^* obtained by applying the filters to a typical data set with state estimation. (left): results for measurements 1 day apart; (right): results for measurements 14 days apart.

see that all three filters have similar performance as the ARMS errors obtained by them are similar to each other. When measurements are taken 28 days apart, we observe that the EKF begins to perform worse than do either the UKF or GHF; i.e., the ARMS errors obtained by the EKF are much larger than those obtained

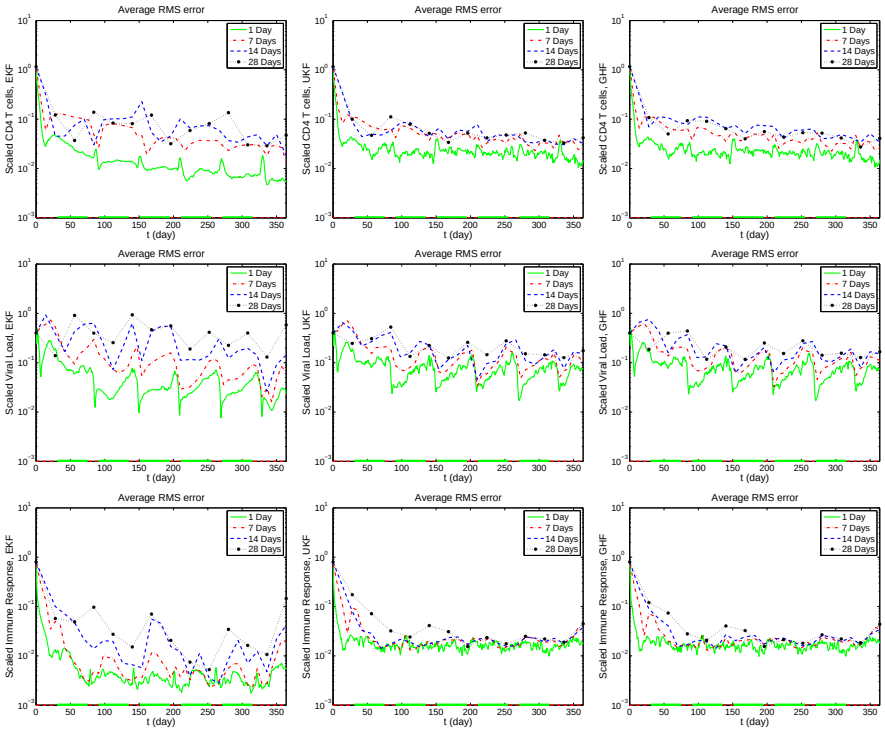

FIGURE 6. State estimation as well as parameter estimation with different measurement frequency for the observed states: scaled CD4+ T-cells (top row), scaled viral load level V_I (middle row), and scaled immune effector T-cells E (bottom row); average RMS errors for each observed state as obtained using the EKF (left column), UKF (middle column) and GHF (right column).

by the UKF and GHF. This further confirms that when parameter estimation is included in the filtering process, the behavior of the filters is different from what we observed in Section 3.2.1.

To investigate the dynamics of the average RMS errors for all the other model compartments (scaled version of T_1, T_2, T_1^* and T_2^*) and estimated parameters (scaled version of λ_1 and N_T) with the state-and-parameter estimation, we present in Figure 7 the results obtained for measurements taken 1 day apart with time plots of the average RMS errors for these model compartments and estimated parameters. Interested readers can refer to Section 3.2.2 in [7] for similar plots obtained for other measurement frequencies.

To also consider the dynamics of the data, observed states and model compartments, as well as the performance of each filter in estimation of the observed states, model compartments and parameters, we present in Figures 8-10 the filtering results with the time plots for the estimated values versus the true values obtained by applying the filters to a typical data set. Here we only present representative results obtained for measurements taken 1 day and 14 days apart. Interested readers can refer to Section 3.2.2 in [7] for similar plots obtained for measurements taken 7 days and 28 days apart. The plots in the left column of Figure 8 are for the

meas. freq.	filter alg.	scaled T_1	scaled T_2	scaled T_1^*	scaled T_2^*	scaled V_I	scaled E	scaled λ_1	scaled N_T
1 day	EKF	0.0234	0.0271	0.0537	0.0380	0.0537	0.0095	0.0148	0.0186
	UKF	0.0323	0.0418	0.0848	0.0599	0.0832	0.0208	0.0181	0.0194
	GHF	0.0323	0.0412	0.0843	0.0589	0.0823	0.0206	0.0191	0.0191
7 days	EKF	0.0852	0.1453	0.1616	0.1676	0.1423	0.0308	0.0689	0.0513
	UKF	0.0858	0.1406	0.1878	0.1868	0.1666	0.0435	0.0760	0.0444
	GHF	0.0837	0.1456	0.1854	0.1891	0.1601	0.0428	0.0713	0.0406
14 days	EKF	0.1214	0.1820	0.3598	0.3100	0.2966	0.0613	0.1043	0.0423
	UKF	0.1098	0.1446	0.2666	0.2309	0.2150	0.0611	0.1061	0.0339
	GHF	0.1224	0.1353	0.2743	0.2192	0.2255	0.0626	0.1323	0.0435
28 days	EKF	0.1592	0.4311	0.5250	0.6826	0.4286	0.0969	0.1030	0.0774
	UKF	0.1433	0.1480	0.3126	0.3479	0.2314	0.0958	0.0824	0.0526
	GHF	0.1449	0.1823	0.3086	0.3931	0.2243	0.0916	0.0875	0.0628

TABLE 5. The ARMS error for each model compartment and estimated parameter with state-and-parameter estimation, where the measurement frequency varies from 1 day, 7 days, 14 days to 28 days.

observed states (CD4+ T-cells, viral load level V_I and immune effector T-cells E) with measurements taken 1 day apart, and those in the right column are for these observed states with measurements taken 14 days apart. Figures 9 and 10 are for model compartments T_1, T_2, T_1^*, T_2^* and estimated parameters λ_1 and N_T with measurements taken at 1 day and 14 days apart, respectively. These three figures reveal that all three filters produce excellent estimates for both the observed states and model compartments with measurements taken 1 day apart, and provide good estimates with measurements 14 days apart. Actually the filters still provide reasonable estimates with measurements 28 days apart (this is detailed in [7]). This conclusion was not obtained in [10], where the EKF was reportedly unable to accurately estimate the states even when the measurements were taken five days apart. We suspect that this may be because in our effort here the filters were applied to the log scaled version of the HIV model instead of the original system as was used in [10].

3.3. **Filtering summary.** Note that in order to apply the GHF and UKF we need first to discretize the model. To investigate the effect of the discretization time step δ_t on the performance of the UKF and GHF, we used several different values for δ_t. We found that if we choose δ_t too small, then we see more oscillation, which is especially obvious for the case when the solution is in the equilibrium state. In addition, a smaller δ_t will dramatically increase the computational time, which is undesirable. On the other hand, if δ_t is taken too large, then we may obtain a non-positive definite covariance matrix P which is not factorable, and the worst scenario is that it does not capture the true dynamics. As a result, when using the GHF and UKF it is important to take care when choosing δ_t to obtain the best performance while still being able to actually carry out the filter computations.

To examine the computational expense of the filters, we list in Table 6 the computational times for the EKF, UKF and GHF when they are applied to a typical data set with measurement frequency varying from 1 day, 7 days, 14 days to 28 days. All the simulations are run on the same regular Dell computer with 3 GB of

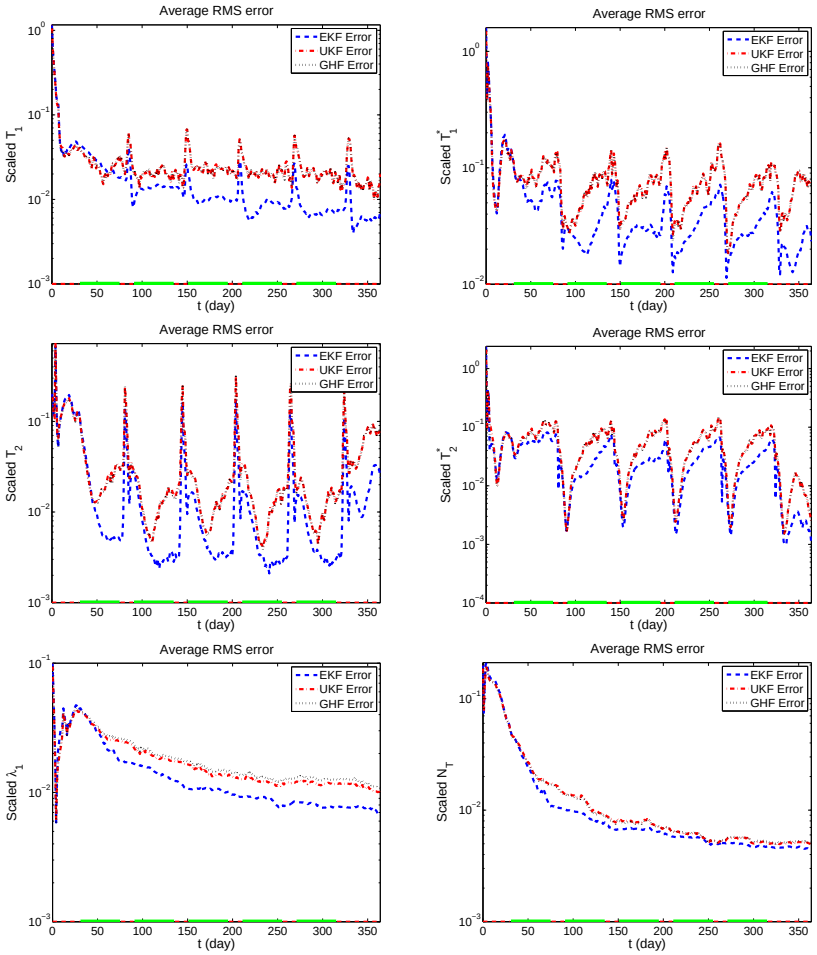

FIGURE 7. Average RMS errors for model compartments (scaled version of T_1, T_2, T_1^* and T_2^*) and estimated parameters (scaled version of λ_1 and N_T) with measurements 1 day apart for state-and-parameter estimation.

Measurement frequency	Computational times (seconds)		
	EKF	UKF	GHF
1 Day	44.748693	231.815131	10746.288109
7 Days	15.874674	226.688416	10724.935096
14 Days	13.207722	230.456907	10801.656850
28 Days	15.074135	233.797818	10826.238094

TABLE 6. The computational times (in seconds) for the EKF, UKF and GHF when they are applied to a typical data set with different measurement frequency.

RAM total and Intel Core 2 Quad Q6600 at 2.40GHz processor. From Table 6 we

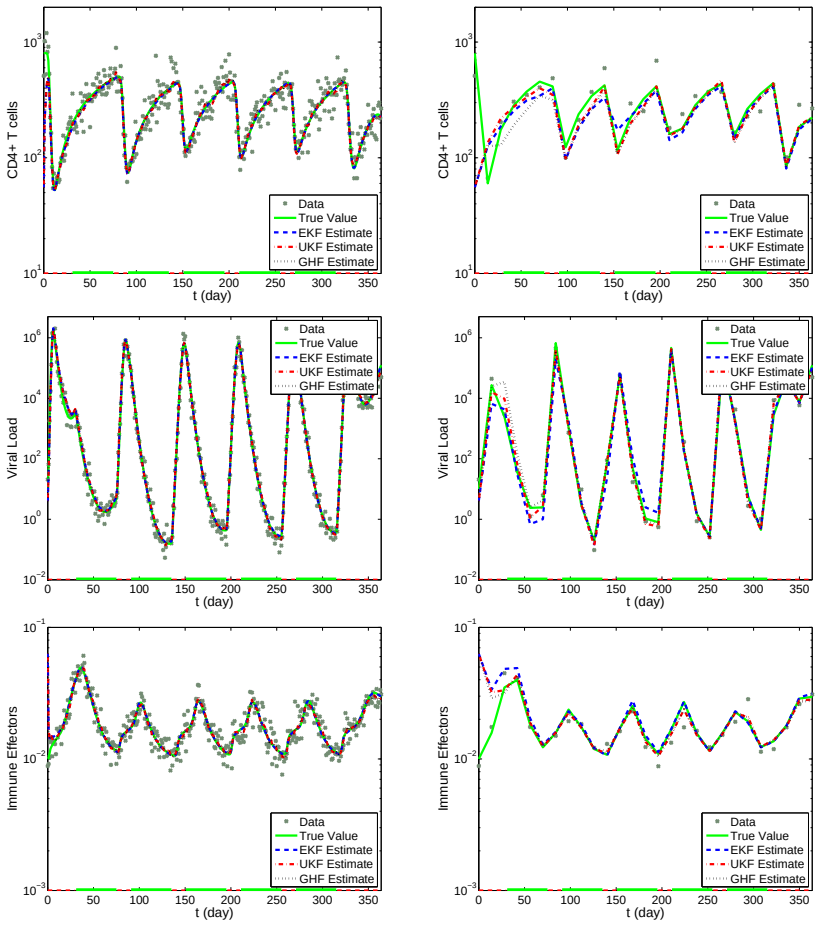

FIGURE 8. The estimated values versus the true values for the total CD4+ T-cell count, viral load level (V_I) and immune effector T-cell count (E) obtained by applying the filters to a typical data set for the state-and-parameter estimation. (left): results are obtained with measurements 1 day apart; (right): results are obtained with measurements 14 days apart.

see that in all cases the GHF is the most computational expensive algorithm, and EKF is the least expensive one.

The simulation results in Sections 3.2.1 and 3.2.2 suggest there is little difference in the estimation accuracy between the UKF and GHF. Since the GHF is the most computationally expensive algorithm and the required number of points to evaluate integrals scales geometrically with the number of dimensions, it makes the GHF computationally infeasible when dealing with estimation of a large number of parameters. As a result, the GHF is not a preferred choice for this particular problem. As for the EKF and UKF, the simulation results in Section 3.2.1 (when only estimating model states) suggest that the EKF performs better than the UKF in terms of estimation accuracy in all the measurement frequencies that we investigated. When estimating parameters as well as model states, the results in Section

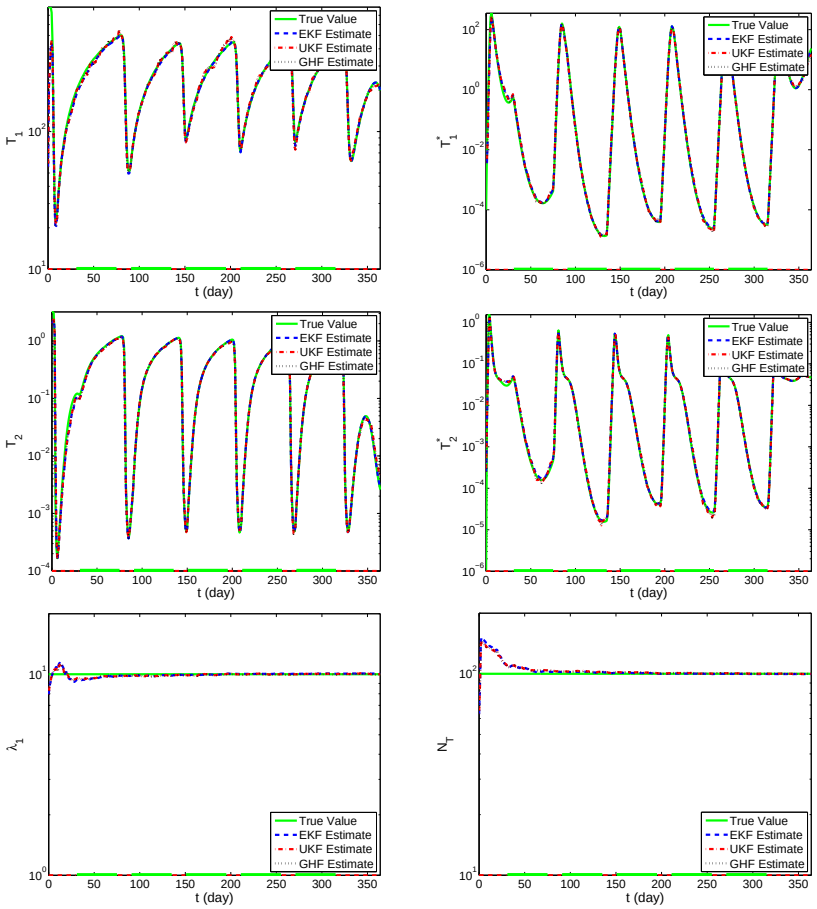

FIGURE 9. State estimation as well as parameter estimation with measurements 1 day apart. The estimated values versus the true values for model compartments T_1, T_2, T_1^*, T_2^* and estimated parameters λ_1 and N_T obtained by applying the filters to a typical data set.

3.2.2 reveal that the EKF still performs better than the UKF with measurements 1 day apart, has comparable performance as the UKF with measurements 7 days and 14 days apart, but performs worse than the UKF with measurements 28 days apart. Note that parameter estimation is essential to this particular problem as the values of a number of the parameters cannot be obtained from laboratory or clinical standards, in part because values of some of the parameters are very much patient-specific. Hence, the decision on which of these two filters should be used relies on the conclusions in Section 3.2.2, that is, on the measurement frequency. Another consideration is that the UKF requires "fine tuning" in order to prevent the propagation of non-positive definite covariance matrices, while there is no such problem with the EKF. Based on all this, we conclude that the EKF is our best choice when the measurement frequency is sufficiently high (such as sampling every

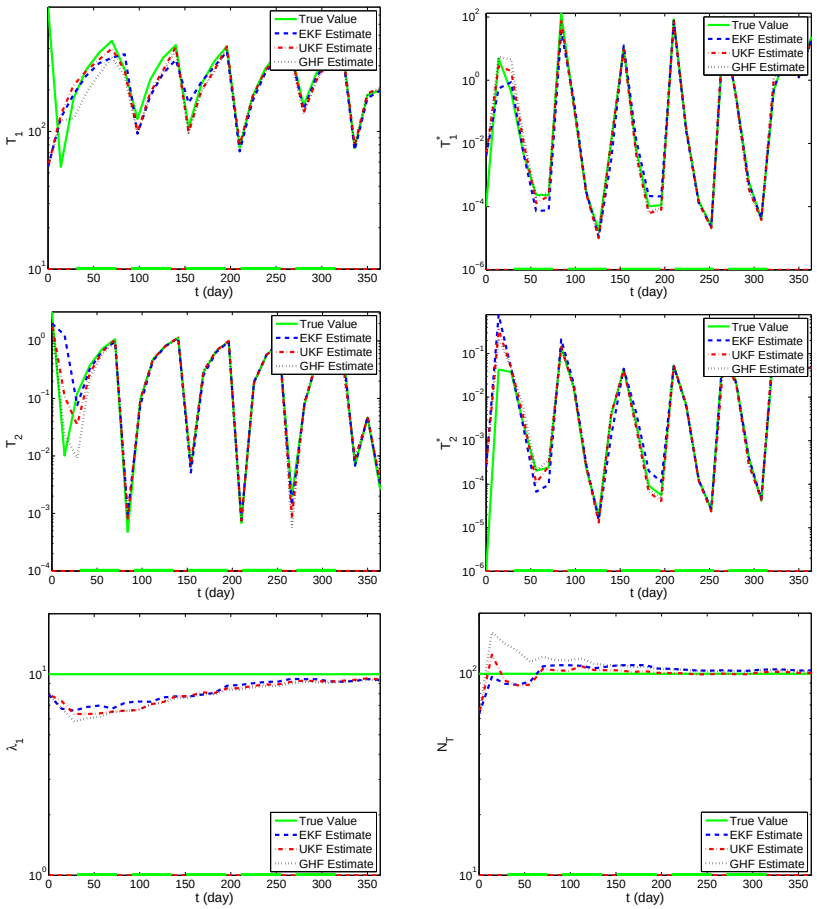

FIGURE 10. State estimation as well as parameter estimation with measurements 14 days apart. The estimated values versus the true values for model compartments T_1, T_2, T_1^* and T_2^* and estimated parameters λ_1 and N_T obtained by applying the filters to a typical data set.

one or two weeks), but when the measurement frequency is as low as data points 4 weeks apart then the UKF is the better choice.

4. **Concluding remarks.** As a first step in designing adaptive treatment therapies for HIV patients, in this paper we have applied the Extended Kalman Filter, the Unscented Kalman Filter and the Gauss-Hermite Filter to the log-scaled version of an HIV model, and then compared the performance of these filters in terms of estimation accuracy and computational time for estimation of both states and parameters. Numerical results suggest that the EKF is our best choice when the measurements are taken as frequently as every week or two, but when the measurement frequency is as low as every 4 weeks then the UKF is most likely the best choice. Because monthly sampling in clinical longitudinal settings is a likely scenario, our future efforts on development of adaptive schemes will be focused on the UKF.

In the clinical data available to us at present, the measurements of the viral load data are censored because the assay can accurately detect only down to some lower limit (400 copies/ml-plasma for a standard assay and 50 copies/ml-plasma for an ultra-sensitive assay). Hence, one of our immediate future efforts is to investigate how effectively one can apply these filters in examples with censored data.

Acknowledgments. This research was supported in part by Grant Number R01A I071915-07 from the National Institute of Allergy and Infectious Diseases and in part by the Air Force Office of Scientific Research under grant number FA9550-09-1-0226. The efforts of ZRK were supported in part by the Department of Education with a GAANN Fellowship under grant number P200A070386.

REFERENCES

[1] B. M. Adams, H. T. Banks, M. Davidian, H.-D. Kwon, H. T. Tran, S. N. Wynne and E. S. Rosenberg, *HIV dynamics: modeling, data analysis, and optimal treatment protocols*, J. Computational and Applied Mathematics, **184** (2005), 10–49.

[2] B. M. Adams, H. T. Banks, H.-D. Kwon and H. T. Tran, *Dynamic multidrug therapies for HIV: optimal and STI control approaches*, Mathematical Biosciences and Engineering, **1** (2004), 223–241.

[3] B. M. Adams, H. T. Banks, M. Davidian and E. S. Rosenberg, *Model fitting and prediction with HIV treatment interruption data*, CRSC-TR05-40, N.C. State University, October, 2005; Bulletin of Mathematical Biology, **69** (2007), 563–584.

[4] I. Arasaratnam and S. Haykin, *Discrete-time nonlinear filtering algorithms using Gaussian-Hermite quadrature*, Proceedings of the IEEE, **95** (2007), 953–977.

[5] H. T. Banks, M. Davidian, Shuhua Hu, G. M. Kepler and E. S. Rosenberg, *Modelling HIV immune response and validation with clinical data*, CRSC-TR07-09, N.C. State University, March, 2007; J. Biological Dynamics, **2** (2008), 357–385.

[6] H. T. Banks, H.-D. Kwon, J. A. Toivanen and H. T. Tran, *A state-dependent Riccati equation-based estimator approach for HIV feedback control*, Optimal Control Appl. Methods, **27** (2006), 93–121.

[7] H. T. Banks, S. Hu, Z. R. Kenz and H. T. Tran, "A Comparison of Nonlinear Filtering Approaches in the Context of an Hiv Model," CRSC-TR09-16, N.C. State University, July, 2009.

[8] R. V. Culshaw, S. Ruan and R. J. Spiteri, *Optimal HIV treatment by maximising immune response*, Journal of Mathematical Biology, **48** (2004), 545–562.

[9] J. David, "Optimal Control, Estimation, and Shape Design: Analysis and Applications," Ph.D. Thesis, North Carolina State University in Raleigh, 2007.

[10] J. David, H. T. Tran and H. T. Banks, *HIV model analysis and estimation implementation under optimal control based treatment strategies*, CRSC-TR08-07, N.C. State University, April, 2008; Intl. J. Pure Applied Math., **57** (2009), 357–392.

[11] J. David, H. T. Tran and H. T. Banks, *Receding horizon control of HIV*, CRSC-TR09-21, N.C. State University, December, 2009; Optimal Control, Applications and Methods, submitted.

[12] K. Ito and K. Xiong, *Gaussian filters for nonlinear filtering problems*, IEEE Transactions on Automatic Control, **45** (2000), 910–927.

[13] A. Jazwinski, "Stochastic Processes and Filtering Theory," Dover Publication, New York, 2007.

[14] S. J. Julier and J. K. Uhlmann, *A general method for approximating nonlinear transformations of probability distributions*, 1996. [Online]. Can be downloaded from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.46.6718&rep=rep1&type=pdf.

[15] S. J. Julier, J. K. Uhlmann and H. F. Durrant-Whyte, *A new method for the nonlinear transformation of means and covariances in filters and estimators*, IEEE Transactions on Automatic Control, **45** (2000), 477–482.

[16] S. J. Julier and J. K. Uhlmann, *Unscented filtering and nonlinear estimation*, Proceedings of the IEEE, **92** (2004), 401–422.

[17] E. A. Wan and R. V. D. Merwe, *The unscented Kalman filter for nonlinear estimation*, Proc. IEEE Adaptive Syst. Signal Process., Commun., Control Symp., (2000), 153–158.

Received July 18, 2009; Accepted January 17, 2010.

E-mail address: htbanks@ncsu.edu
E-mail address: shu3@ncsu.edu
E-mail address: zrkenz@ncsu.edu
E-mail address: tran@ncsu.edu