BAYESIAN ESTIMATION OF POSTERIOR DENSITIES FOR PHARMACOKINETIC MODELS HAVING CHANGING PARAMETER VALUES

David S. Bayard and Roger W. Jelliffe

Technical Report: 00-1

April 10, 2000

Laboratory of Applied Pharmacokinetics
University of Southern California School of Medicine
Los Angeles, California 90033

BAYESIAN ESTIMATION OF POSTERIOR DENSITIES FOR PHARMOCOKINETIC MODELS HAVING CHANGING PARAMETER VALUES

D.S. Bayard*, and R.W. Jelliffe†
Laboratory of Applied Pharmacokinetics
University of Southern California
School of Medicine
Los Angeles, CA 90033

Abstract

This paper considers the updating of posterior densities for pharmacokinetic models which have changing parameter values. For estimation purposes, the interactive multiple model (IMM) estimation algorithm is proposed, which is currently a popular algorithm in the aerospace community for tracking maneuvering targets. The IMM algorithm is described, and compared to the multiple model (MM) and MAP Bayesian estimation methods, which are presently popular for pharmacokinetic problems where parameter values do not change. The methods are evaluated using both simulated and real clinical data for the drug Tobramycin.

Keywords Health sciences, estimation, filtering, Markov chain, decision support systems, pharmocokinetics, therapeutic drug monitoring

*Senior Research Scientist, Jet Propulsion Laboratory, Pasadena CA, 91109
†Laboratory of Applied Pharmacokinetics, University of Southern California, Los Angeles CA, 90033

Contents

1 Introduction

This paper considers the updating of posterior densities for pharmacokinetic models which have changing parameter values. The main assumption is that the prior probability has the form of a discrete density and that parameter changes can be suitably modeled as "jumps" from one model support point to another within the same discrete set. The authors' main motivation for studying this estimation problem is for developing better controllers, i.e., to extend the present Multiple Model (MM) active control approach outlined in [1] to designing dosage regimens for unstable patients with varying parameter values.

The discrete prior assumption is consistent with recent results on maximum likelihood (ML) estimation which have shown that the nonparametric ML estimate of a population pharmacokinetic model is of this form (cf., Lindsay [2], Mallet [3]). Our laboratory has used a Nonparametric Expectation Maximization (NPEM) approach to determine such optimal discrete ML population models [4].

In the case of discrete priors, the multiple model (MM) estimation approach provides an exact analytical solution for updating Bayesian posteriors [5]. This has led to renewed interest in multiple model approaches of estimation and control for application to pharmacokinetic control problems [6][7][1]. The MM estimator works well where the patient's parameters are unknown but constant [1]. However, in the case where parameters can vary, the MM estimator is sub-optimal, and several modifications of the basic MM approach have been suggested in the literature (cf., [8][9] for survey). Recently, the Interactive Multiple Model (IMM) algorithm has emerged as a simple and effective suboptimal approach [10][11][9].

In this paper the IMM algorithm is applied to updating posterior densities in pharmacokinetic estimation problems which have changing parameter values. The IMM algorithm is described, and compared to the sequential MM and Maximum A-Posteriori (MAP) Bayesian estimation methods which are presently used for posterior updating when pharmacokinetic parameters do not change. The methods will be demonstrated using both simulated and real clinical data for a patient receiving the drug Tobramycin. The report has been presented in an earlier abridged form as a conference paper [13].

2 Multiple Model Linear Estimation

2.1 Multiple Model Formulation

Consider the s linear models,

$$\dot{x}(\theta_\ell, t) = A(\theta_\ell)x(\theta_\ell, t) + B(\theta_\ell)u(t) + W(\theta_\ell)w(t); \quad [\ell = 1, ..., s] \tag{1}$$

where $x(\theta_\ell, t) \in R^n$ is the state of the ℓth model at time t; $u \in R^1$ is an exogenous input (e.g., IV infusion rate), and $w \in R^n$ is a white Gaussian vector noise process with unit covariance matrix.

Define a set of event times $t_k, \quad k = 1, ..., N$. Let the input $u(t)$ be a *constant* u_k during each

interval $[t_k, t_{k+1})$ i.e., let

$$u(t) = \sum_{k=1}^{N-1} \chi[t_k, t_{k+1})(t) u_{k+1}, \quad u = [u_1, ..., u_N]^T \tag{2}$$

where χ denotes the standard indicator function (i.e., $\chi[t_k, t_{k+1})(t) = 1$ for $t \in [t_k, t_{k+1})$ and 0 elsewhere). Then each model (1) can be discretized *exactly* as,

$$x_{k+1}(\theta_\ell) = \Phi_k(\theta_\ell) x_k(\theta_\ell) + \Gamma_k(\theta_\ell) u_k + \mathcal{W}_k(\theta_\ell) w_k \tag{3}$$

where,

$$\Phi_k(\theta_\ell) = exp(A(\theta_\ell)\Delta(k)) \tag{4}$$

$$\Gamma_k(\theta_\ell) = \int_0^{\Delta(k)} exp(A(\theta_\ell)\sigma) d\sigma B(\theta_\ell) \tag{5}$$

$$\Delta(k) = t_{k+1} - t_k \tag{6}$$

and where the choice of $\mathcal{W}_k$ is used (see Section 3) to model various errors in the clinical environment as process noise.

3 Process Noise Models

The process noise $\mathcal{W}_k w_k$ in (3) is represented as the sum of three separate terms as follows,

$$\mathcal{W}_k w_k = \mathcal{M}_k w_k^1 + D_k w_k^2 + T_k w_k^3 \tag{7}$$

where $\mathcal{M}_k w_k^1$ represents the **modelling error**; $D_k w_k^2$ represents the **dosage preparation error**; and $T_k w_k^3$ represents the **dosage timing error**. These noise terms w_k^i, $i = 1, 2, 3$ are assumed to be mutually statistically independent Normal noise processes with unit covariance, and the total process noise covariance can be correspondingly calculated as,

$$\mathcal{W}_k \mathcal{W}_k^T = \mathcal{M}_k \mathcal{M}_k^T + D_k D_k^T + T_k T_k^T \tag{8}$$

Each of the contributing process noise sources will be treated separately below.

3.1 Modelling Error

It is assumed that errors in the structure of the continuous-time models (1) have been "absorbed" into the process noise term $W(\theta_\ell)$. This is a common way to represent additional compartments, nonlinearities, interactions, complexities, etc., which are not captured by the dynamics part of (1) alone. It is assumed that W in (1) has already been suitably chosen to reflect modelling error in the *continuous-time* model. As such, the transformation to *discrete-time* can be calculated as follows. Consider the linear system,

$$\dot{x} = Ax + Ww \tag{9}$$

Assuming that $x(t_k) = 0$, the propagation of the state from time t_k to time t_{k+1} can be calculated as,

$$x_{k+1} \stackrel{\Delta}{=} x(t_{k+1}) = \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau)} W w(\tau) d\tau \tag{10}$$

Consider the representation $x_{k+1} = \mathcal{M}_k w_k^1$ for the new discrete-time state. Then,

$$\mathcal{M}_k \mathcal{M}_k^T = Cov(x_{k+1}) = E[x_{k+1} x_{k+1}^T] \tag{11}$$

$$= E\Big[\int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau_1)} W w(\tau_1) d\tau_1 \Big(\int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau_2)} W w(\tau_2) d\tau_2\Big)^T\Big] \tag{12}$$

$$= \int_{t_k}^{t_{k+1}} \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau_1)} W E[w(\tau_1) w^T(\tau_2)] W^T (e^{A(t_{k+1}-\tau_2)})^T d\tau_1 d\tau_2 \tag{13}$$

$$= \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau_1)} W W^T (e^{A(t_{k+1}-\tau_1)})^T d\tau_1 \tag{14}$$

$$= \int_0^{\Delta(k)} e^{A(\theta)\sigma} W(\theta_\ell) W^T(\theta_\ell) (e^{A(\theta)\sigma})^T d\sigma \tag{15}$$

Here, (10) is substituted into (11) to get (12); the noise property $E[w(\tau_1) w^T(\tau_2)] = \delta(\tau_1 - \tau_2)$ is used in in (13) to get (14), and the change of variable $\sigma = t_{k+1} - \tau_1$ is used in (14) to get (15).

The expression (15) can be computed exactly using integration. In order to simplify the computation, the approximation,

$$e^{A\sigma} \sim I \tag{16}$$

can be used in (15) to give the simpler (but approximate) expression,

$$\mathcal{M}_k \mathcal{M}_k^T = W_k W_k^T (t_{k+1} - t_k) \tag{17}$$

which gives upon rearranging,

$$\mathcal{M}_k = W_k \sqrt{t_{k+1} - t_k} \tag{18}$$

3.2 Dosage Amount Error

Usually, there is some error in the preparation of the prescribed dose amount. This error is modelled using the extra term $D_k w_k^2$ in the overall process noise of (7). An analysis to model the effect of dose amount error is given here.

Let a desired dose at time $t = t_k$ be specified as u_k. Due to preparation error the amount is given as $(1 \pm \delta_D) u_k$ where δ_D is a constant percentage error. Then, (assuming no other error sources) the state can be calculated as,

$$x_{k+1} = \Phi_k x_k + \Gamma_k u_k \pm \delta_D \Gamma_k u_k \tag{19}$$

The deterministic perturbation $\delta_D \Gamma_k u_k$ will be treated *heuristically* as if it came from a random additive perturbation,

$$\delta_D \Gamma_k u_k \simeq D_k w_k^2 \tag{20}$$

where w_k^2 is white Gaussian noise. The noise scaling matrix is chosen to give 1-sigma values approximating to deterministic perturbation $\pm\delta_D\Gamma_k u_k$, i.e.,

$$T_k = diag\{\delta_D\Gamma_k u_k\} \tag{21}$$

Here the operator $V = diag(v)$ defines a diagonal matrix V using the entries of the vector v.

3.3 Dosage Timing Error

Due to practical constraints, doses are rarely given at exactly the time they are prescribed or recorded. In order to model the effect of dose timing errors, the extra term $T_k w_k^3$ has been added to the overall process noise in (7). Intuitively, the effects of dose timing errors are larger when doses are larger. A dose timing error model which characterizes this effect is given here.

Consider the propagation of the continuous-time system,

$$\dot{x} = Ax + Bu \tag{22}$$

starting with zero state $x_{k-1} = x(t_{k-1}) = 0$ at time $t = t_{k-1}$, where the input is (ideally) intended to change from $u = u_{k-1}$ to $u = u_k$ at time $t = t_k$. However, due to dose timing error, the input changes as early as $t^- = t_k - \delta_T$ or as late as $t^+ = t_k + \delta_T$. Then depending upon the actual dose time, one can correspondingly compute the new state as,

$$x_{k+1}^+ = \int_{t_{k-1}}^{t^-} e^{A(t_{k+1}-\tau)}Bd\tau u_{k-1} + \int_{t^-}^{t_{k+1}} e^{A(t_{k+1}-\tau)}Bd\tau u_k \tag{23}$$

or,

$$x_{k+1}^- = \int_{t_{k-1}}^{t^+} e^{A(t_{k+1}-\tau)}Bd\tau u_{k-1} + \int_{t^+}^{t_{k+1}} e^{A(t_{k+1}-\tau)}Bd\tau u_k \tag{24}$$

Subtracting (23) from (24) gives,

$$\begin{aligned}
\delta x_{k+1} &= x_{k+1}^+ - x_{k+1}^- && (25)\\
&= \int_{t^-}^{t^+} e^{A(t_{k+1}-\tau)}Bd\tau(u_k - u_{k-1}) && (26)\\
&= \int_{-\delta_T}^{\delta_T} e^{A(t_{k+1}-t_k+\sigma)}Bd\sigma(u_k - u_{k-1}) && (27)\\
&= \Phi_k \int_{-\delta_T}^{\delta_T} e^{A\sigma}Bd\sigma(u_k - u_{k-1}) && (28)
\end{aligned}$$

where (27) follows by the substituting $\sigma = t_k - \tau$ into (26), and (28) follows by rearranging (27) and using definition (51).

The expression (28) can be computed exactly using integration. In order to simplify the computation, the approximation,

$$e^{A\sigma} \sim I + A\sigma + \frac{1}{2}A^2\sigma^2 \tag{29}$$

can be used in (15) to give the simpler (but approximate) expression,

$$\delta x_{k+1} \simeq \Phi_k(2\delta_T I + A^2\delta_T^3/3)B(u_k - u_{k-1}) \tag{30}$$

The perturbation δx_{k+1} will be treated *heuristically* as if it came from a random additive perturbation,

$$\delta x_{k+1} \simeq T_k w_k^3 \tag{31}$$

where w_k^3 is white Gaussian noise. The noise scaling matrix is chosen to give 1-sigma values approximating the worst-case deterministic perturbation δx_{k+1}, i.e.,

$$T_k = diag\Big\{\Phi_k(2\delta_T I + A^2\delta_T^3/3)B(u_k - u_{k-1})\Big\} \tag{32}$$

4 Measurement Noise Models

4.1 Measurement Model

Events during which measurements are taken are indicated using a flag vector $\mathcal{H} = [\mathcal{H}_1, ..., \mathcal{H}_N]$ where,

$$\mathcal{H}_k = \begin{cases} 1 & \text{if a measurement is taken at time } t = t_k \\ 0 & \text{otherwise} \end{cases} \tag{33}$$

The actual measurements are generated by the truth model θ^* and are assumed to be of the form,

$$y_k = C_k(\theta^*)x_k(\theta^*) + \mathcal{V}_k(\theta^*)v_k \tag{34}$$

where θ^* is assumed to be in the model set (i.e., $\theta^* = \theta_\ell$ for some $\ell = 1, .., s$); v_k is a white Gaussian vector noise process.

Consider the multiple model system equations (1)(3). Each model has a measurement equation of the form,

$$y_k = C_k(\theta_\ell)x_k(\theta_\ell) + \mathcal{V}_k(\theta_\ell)v_k \tag{35}$$

4.2 Assay Errors

For the purposes of modelling measurement errors associated with laboratory assays, the measurement noise in (35) is decomposed into two terms,

$$\mathcal{V}_k(\theta_\ell)v_k = \mathcal{V}_k^f v_k^1 + \mathcal{V}_k^a v_k^2 \tag{36}$$

where v_k^1 and v_k^2 are independent white unit normal noise sequences; $\mathcal{V}_k^f(\theta_\ell)$ is a fixed quantity (assumed specified); and $\mathcal{V}_k^a$ is the assay error represented in terms of the polynomial,

$$\mathcal{V}_k^a = a_0 + a_1 y_k + a_2 y_k^2 + a_3 y_k^3 \tag{37}$$

This represents a slight abuse of the multiple model formulation since the covariance has measurement dependent terms. However, assay error polynomials are often available in practice and (37) provides a convenient method for incorporating this extra knowledge.

5 Multiple Model (MM) Update Formulas

Given a discrete prior, linear dynamics, and Gaussian process and measurement noises, the conditional probability density of the state is known to propagate as the following Gaussian sum [5],

$$p(x_k|I_k) = \sum_{\ell=1}^{s} p(\theta_\ell|I_k) \cdot N(x_k - \widehat{x}_k(\theta_\ell), P_k(\theta_\ell)) \tag{38}$$

where it has been convenient to use the notations,

$$N(x, X) \stackrel{\Delta}{=} \frac{1}{(2\pi)^{n/2}|X|^{1/2}} e^{-\frac{1}{2}x^T X^{-1} x}; \quad x \in R^n \tag{39}$$

$$I_k = [u_1, ..., u_{k-1}, y_1, ..., y_k] \tag{40}$$

At each new event time t_{k+1}, the means $\widehat{x}_k(\theta_\ell)$ and covariances $P_k(\theta_\ell)$ of the models $\theta_\ell, \ \ell = 1, ..., s$ are updated using a bank of Kalman filters (KFs) as follows [5],

A-Priori Update

$$\widehat{x}_{k+1|k}(\theta_\ell) = \Phi_k(\theta_\ell)\widehat{x}_k(\theta_\ell) + \Gamma_k(\theta_\ell)u_k \tag{41}$$

$$M_{k+1}(\theta_\ell) = \Phi_k(\theta_\ell)P_k\Phi_k^T(\theta_\ell) + \mathcal{W}_k(\theta_\ell) \tag{42}$$

A-Posteriori Update (State)

$$\widehat{x}_{k+1}(\theta_\ell) = \widehat{x}_{k+1|k}(\theta_\ell) + \mathcal{H}_{k+1}K_{k+1}(\theta_\ell)\Big(y_{k+1} - C_{k+1}(\theta_\ell)\widehat{x}_{k+1|k}(\theta_\ell)\Big) \tag{43}$$

$$K_{k+1}(\theta_\ell) = P_{k+1}(\theta_\ell)C_{k+1}^T(\theta_\ell)\mathcal{V}_{k+1}^{-1}(\theta_\ell) \tag{44}$$

$$P_{k+1}(\theta_\ell) = M_{k+1}(\theta_\ell) - \mathcal{H}_{k+1}M_{k+1}(\theta_\ell)C_{k+1}^T(\theta_\ell)\Omega_{k+1}^{-1}(\theta_\ell)C_{k+1}(\theta_\ell)M_{k+1}(\theta_\ell) \tag{45}$$

$$\Omega_{k+1}(\theta_\ell) \stackrel{\Delta}{=} C_{k+1}(\theta_\ell)M_{k+1}(\theta_\ell)C_{k+1}^T(\theta_\ell) + \mathcal{V}_{k+1}(\theta_\ell) \tag{46}$$

A-Posteriori Update (Model Probabilities)

$$p(\theta_\ell|\mathcal{I}_{k+1}) = \begin{cases} p(y_{k+1}|\mathcal{I}_k, u_k, \theta_\ell)p(\theta_\ell|\mathcal{I}_k)/d_k & \text{if } \mathcal{H}_{k+1} = 1 \\ p(\theta_\ell|\mathcal{I}_k) & \text{if } \mathcal{H}_{k+1} = 0 \end{cases} \tag{47}$$

$$d_k = \sum_{\ell=1}^{s} p(y_{k+1}|\mathcal{I}_k, u_k, \theta_\ell)p(\theta_\ell|\mathcal{I}_k) \tag{48}$$

$$p(y_{k+1}|\mathcal{I}_k, u_k, \theta_\ell) = N\Big(y_{k+1} - C_{k+1}(\theta_\ell)\widehat{x}_{k+1|k}(\theta_\ell), \ \Omega_{k+1}(\theta_\ell)\Big) \tag{49}$$

where the notation (39) is used, and the probabilities are initialized as,

$$p(\theta_\ell|\mathcal{I}_0) = p(\theta_\ell); \quad \ell = 1, ...s \tag{50}$$

It is noted that if no measurement is taken at time t_{k+1}, then the a-posteriori estimates are taken to be identical to the a-priori estimates, i.e.,

$$\widehat{x}_{k+1}(\theta_\ell) = \widehat{x}_{k+1|k}(\theta_\ell) \tag{51}$$

$$P_{k+1}(\theta_\ell) = M_{k+1}(\theta_\ell) \tag{52}$$

$$p(\theta_\ell|\mathcal{I}_{k+1}) = p(\theta_\ell|\mathcal{I}_k) \tag{53}$$

and it is not necessary to compute (44)(46)(48)(49).

Since the conditional density in (38) is updated after each separate measurement, this procedure defines a *sequential* algorithm for updating the Bayesian posteriors over a given horizon or collection of measurements.

The MM estimate is defined as the mean of the conditional density (38), i.e.,

$$\widehat{x}_k^{MM} = \sum_{\ell=1}^{s} p(\theta_\ell|I_k) \cdot \widehat{x}_k(\theta_\ell) \tag{54}$$

The sequential MM estimate is simply a weighted average of the model-conditioned state estimates at each time instant, weighted by their model probability. In contrast, the sequential MAP Bayesian estimate is defined by taking the state of the single model whose probability is the largest at any given time instant, i.e.,

$$\widehat{x}_k^{MAP} = \widehat{x}_k(\theta_{\ell_k^*}) \tag{55}$$

where,

$$\ell_k^* = arg \max_{\ell} \quad p(\theta_\ell|I_k) \tag{56}$$

6 Interactive Multiple Model (IMM) Update Formulas

In the interactive multiple model approach, the true subject's parameters are permitted to jump from the i'th model θ_i to the j'th model θ_j, at time k, where the indices i, j, k are arbitrary. For this purpose, a more general notation must be adopted. Since the models can jump as a function of time k, it is convenient to index the models by number and time as follows,

$$\theta_i(k), \quad i = 1, ..., s; \quad k = 1, ..., N-1 \tag{57}$$

The quantity $\theta_i(k)$ *is associated with the hypothesis that the true subject has parameters* θ_i *at time* k.

Consider the case where there are s models at time $k-1$. Let the probability of a jump from model i to model j at time k be denoted as,

$$\pi_{ji}(k) \stackrel{\Delta}{=} p(\theta_j(k)|\theta_i(k-1)) \tag{58}$$

The model probabilities are then propagated as the Markov Chain,

$$p(\theta_j(k)|I_{k-1}) = \sum_i \pi_{ji}(k) p(\theta_i(k-1)|I_{k-1}); \quad j = 1, ..., s \tag{59}$$

There is generally no restriction on the choice of π_{ji} (they can even be dependent on I_{k-1} if necessary). However, for simplicity they will be chosen in the present application as,

$$\pi_{ji}(k) = \begin{cases} \alpha & \text{if } i = j \\ (1-\alpha)/(s-1) & \text{otherwise} \end{cases} \tag{60}$$

This choice assigns probability α to all diagonals of the Markov Chain matrix (i.e., the probability of not jumping is the same for all models) and the remaining probability is equally assigned across the remaining models (jumps are equally probable between any two models).

The IMM estimate is computed identically to the MM estimate discussed in Section 5 above, except for an additional mixing step which makes use of the Markov chain (59) [11][9]. The main idea behind the mixing step is to write the probability of the state x_{k-1} *conditioned on the hypothesis that the truth model is* θ_j *at the next instant* k, i.e.,

$$p(x_{k-1}|I_{k-1}) = \sum_{i=1}^{s} p(x_{k-1}|\theta_j(k), I_{k-1}) p(\theta_j(k)|I_{k-1}) \tag{61}$$

With this conditioning, each term in (61) can be calculated separately. The second term on the RHS of (61) is given by the Markov chain (59). The first term on the RHS of (61) is calculated as,

$$p(x_{k-1}|\theta_j(k), I_{k-1}) = \sum_{i=1}^{s} p(x_{k-1}|\theta_j(k-1), \theta_j(k), I_{k-1}) \cdot p(\theta_i(k-1)|\theta_j(k), I_{k-1}) \tag{62}$$

$$= \sum_{i=1}^{s} p(x_{k-1}|\theta_j(k-1), I_{k-1}) \cdot \mu_{i|j}(k) \tag{63}$$

where the *mixing probabilities* have been conveniently defined as,

$$\mu_{i|j}(k) \stackrel{\Delta}{=} p(\theta_i(k-1)|\theta_j(k), I_{k-1}) \tag{64}$$

At this point it is noted that the expression in (63) is a Gaussian sum, i.e., a weighted sum of s different Gaussian densities. In principle, the propagation from this point forward can be calculated "exactly" by splitting the Kalman filter bank (i.e., by conditioning on the enumerated jump possibilities) into s banks of s filters each to track the propagation of each Gaussian pulse separately. This gives s^2 Kalman filters. If only a single jump at time k were permitted over the entire horizon, this bank of s^2 KFs would give an exact solution to the problem. However, jumps are generally allowed at other times as well. The enumeration of jumps in this manner will split the bank at each time instant, giving a total of s^N required Kalman filters over a horizon of N steps. Unfortunately, such computation quickly becomes unwieldy. For example, a discrete density with $s = 30$ points over a horizon of $N = 10$ gives $30^{11} = 18 \times 10^{15}$. It is at this point that the IMM algorithm affords a useful approximation.

The key approximation in the IMM algorithm is to replace the Gaussian sum expression (61) by a *single* Gaussian density whose mean and covariance are chosen equal to that of the entire sum, i.e.,

$$\sum_{i=1}^{s} p(x_{k-1}|\theta_j(k-1), I_{k-1}) \cdot \mu_{i|j}(k) \simeq N(\widehat{x}_{k-1}(\theta_j(k), P_{k-1}(\theta_j(k)) \tag{65}$$

where,

$$\widehat{x}_{k-1}(\theta_j(k)) \triangleq \sum_{i=1}^{s} \widehat{x}_{k-1}(\theta_i(k-1))\mu_{i|j}(k) \tag{66}$$

$$P_{k-1}(\theta_j(k)) \triangleq \sum_{i=1}^{s}[P_{k-1}(\theta_i(k-1))+(\widehat{x}_{k-1}(\theta_i(k-1))-\widehat{x}_{k-1}(\theta_j(k))(\widehat{x}_{k-1}(\theta_i(k-1))-\widehat{x}_{k-1}(\theta_j(k))^T]\mu_{i|j}(k) \tag{67}$$

This collapses the number of Kalman filters back to s, and the recursive procedure can be reproduced at the next iteration to give an algorithm which propagates using only s Kalman filters. All that remains is to find a formula for the mixing probabilities defined in (64). This is done using Bayes' rule as follows,

$$\mu_{i|j}(k) \triangleq p(\theta_i(k-1)|\theta_j(k), I_{k-1}) \tag{68}$$

$$= \frac{p(\theta_i(k-1), \theta_j(k)|I_{k-1})}{p(\theta_j(k)|I_{k-1})} \tag{69}$$

$$= \frac{\pi_{ji}(k)p(\theta_i(k-1)|I_{k-1})}{p(\theta_j(k)|I_{k-1})} \tag{70}$$

The IMM mixing step is applied directly after each a-posteriori update in Section 5 in the standard MM equations. It is mechanized by first computing the Markov chain update (59), then computing the mixing probabilities using (70), and then computing the equivalent mean and covariance in (66)(66) for each model. The computation then continues with the standard MM a-priori update in Section 5, and then a-posteriori update (after the next measurement) where the "after jump" probabilities $p(\theta_\ell(k+1)|I_k)$ from (59) now play the role of the prior probabilities $p(\theta_\ell|I_k)$ in (47)(48).

7 Numerical Example (Simulation)

7.1 Tobramycin Infusion Model

In this section, the above estimation approaches are applied to a simulated example. For the purposes of this paper, the Tobramycin pharmacokinetic model will be represented as the 1-compartment model,

$$\dot{x}(\theta_\ell, t) = A(\theta_\ell)x(\theta_\ell, t) + B(\theta_\ell)u \tag{71}$$

$$A(\theta_\ell) = [-K_e(\ell)]; \quad B(\theta_\ell) = [1] \tag{72}$$

The measured variable y is the concentration in the central compartment, i.e.,

$$y(\theta_\ell, t) = C(\theta_\ell)x(\theta_\ell, t) \tag{73}$$

$$C(\theta_\ell) = [1/V(\ell)] \tag{74}$$

Here, the vector $\theta_\ell = [K_e(\ell), V(\ell)]$ attains discrete values θ_ℓ, $\ell = 1, ..., s$ in a set Θ, each with probability $p(\theta_\ell)$. Specifically, the elimination rate constant $K_e(\ell)$ is defined by,

$$K_e(\ell) = K_i(\ell) + K_s(\ell) \cdot Ccr \tag{75}$$

where Ccr is the measured or estimated creatinine clearance of the patient, and the volume $V(\ell)$ is given by,

$$V(\ell) = V_s(\ell) \cdot BW \tag{76}$$

where BW denotes the patient's body weight (in kilograms). For this simulation, the parameters $K_i(\ell), K_s(\ell), V_s(\ell)$ and their prior probabilities $p(\theta_\ell)$ are based on a small NPEM population model of Tobramycin. The discrete joint density parameters are summarized in Table 1.

ℓ	$p(\theta_\ell)$ probability	K_s hr^{-1}/unit CCr	V_s L/kg	K_i hr^{-1}
1	.5002753e-01	.944339e-03	.617568	.693150e-02
2	.5147634e-01	.980314e-03	.783951	.693150e-02
3	.3221748e-01	.156341e-02	.355751	.693150e-02
4	.6929264e-01	.169681e-02	.347757	.693150e-02
5	.4872111e-01	.170880e-02	.403218	.693150e-02
6	.1022884	.173279e-02	.514140	.693150e-02
7	.5445284e-01	.256770e-02	.250625	.693150e-02
8	.1190895	.279554e-02	.304387	.693150e-02
9	.8498735e-01	.330219e-02	.397622	.693150e-02
10	.6120562e-01	.339662e-02	.459378	.693150e-02
11	.4238511e-01	.371740e-02	.192965	.693150e-02
12	.7461239e-01	.447886e-02	.214750	.693150e-02
13	.1703741	.449086e-02	.270211	.693150e-02
14	.3829396e-01	.484011e-02	.260518	.693150e-02

Table 1: Pharmacokinetic model parameters and probabilities

The true subject's body weight and creatinine clearance are chosen as piecewise contant in time according to the values given in Table 2.

Time (hrs)	BW (kg)	Ccr (units Ccr)
34.00	80	109.8376
64.00	80	95.4720
186.50	80	23.9322
213.00	80	12.6395
228.58	80	19.9181
250.25	80	18.2493
301.50	80	18.7030

Table 2: Body weight and creatinine clearance values

7.2 Jump Parameter Simulation

Tobramycin is administered as a train of short .5 hour infusions whose infusion rates are plotted in Figure 1. The motivation for using this infusion regimen, is that it was given to a particular patient whose responses will be studied in a separate clinical data section of this report (cf., Section **??** below).

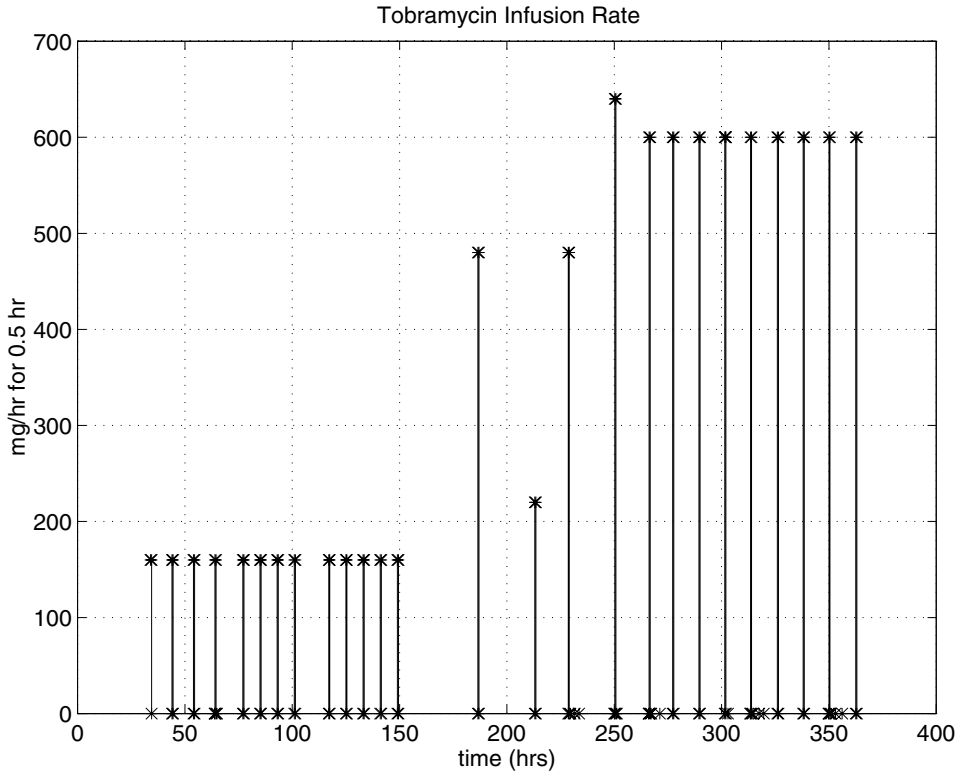

Figure 1: Tobramycin dosage inputs

The system response is simulated by assuming that the true subject's parameters jump from model $\ell = 1$ to that of model $\ell = 7$ at time $t = 255$ hours. The true subject's response (concentration y of the central compartment) is shown plotted in Figure 2. The "*" symbol denotes measurements of the form y_k in (34) taken at the specific times t_k shown. While the concentrations shown after the jump at $t = 255$ hours may appear to be somewhat unrealistic at first, they are not outside the limits of the clinical experience of the authors. In any event, the simulated clinical situation shown here presents a significant challenge to any pharmacokinetic analysis, which is the main point to be made.

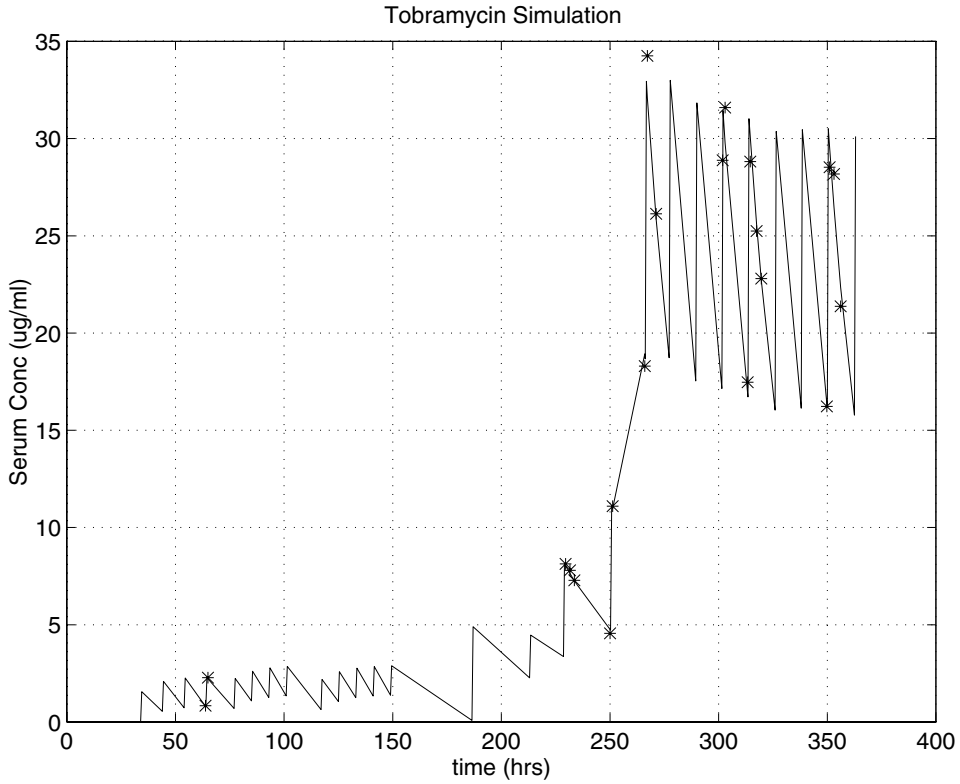

Figure 2: Tobramycin simulation response: True (solid line), measurements (*)

7.3 MM and MAP Bayesian Estimation Results

Given the input and output data from the simulated response, the MM estimation approach discussed in Section 5 is applied to updating posteriors. The sequential MM estimate of the output y is shown as the solid line in Figure 3. The sequential MAP Bayesian estimate is shown as the dashed line in Figure 3. The MM and MAP estimates are hard to distinguish since they are nearly identical. The noisy measurements are superimposed using the symbol "*" to show the correspondence.

It is seen in Figure 3 that the estimated output agrees well with the first 7 measurements, which all occur before the jump at time $t = 255$ hours. After the jump, it takes some time for the MM estimator to track the change. Specifically, both the sequential MM and MAP Bayesian estimates appear to track correctly by the 13th measurement (at time $t = 313.33$ hours).

The propagation of the sequential posterior model probabilities with each measurement update is shown in Figure 4. Before the jump at $t = 255$ hours, the estimator settles (correctly) on model $\ell = 1$ with nearly probability one. At measurement 9 (the second measurement after the jump occurs) the probability of model $\ell = 1$ drops sharply, and the estimator converges momentarily on model $\ell = 3$, until the 17th measurement (at time $t = 349.83$ hours) when it (correctly) settles on $\ell = 7$. This 7th model is maintained with nearly probability one for the remainder of the updates.

Estimates of the elimination constant K_e, volume V, and clearance $K_e * V$ corresponding to the mode of the sequential MM estimator are plotted in Figure 5. A drop in the true and estimated values is seen after the jump time of $t = 255$ hours.

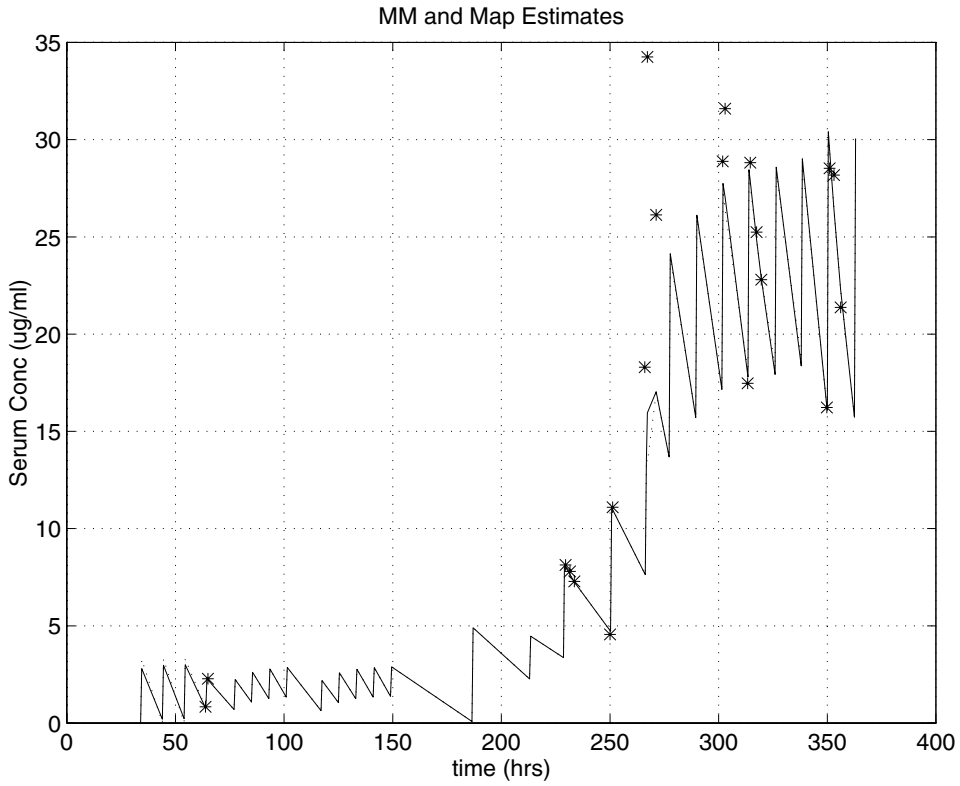

Figure 3: MM (solid) and MAP Bayesian (dotted) estimates of serum concentration measurements (*)

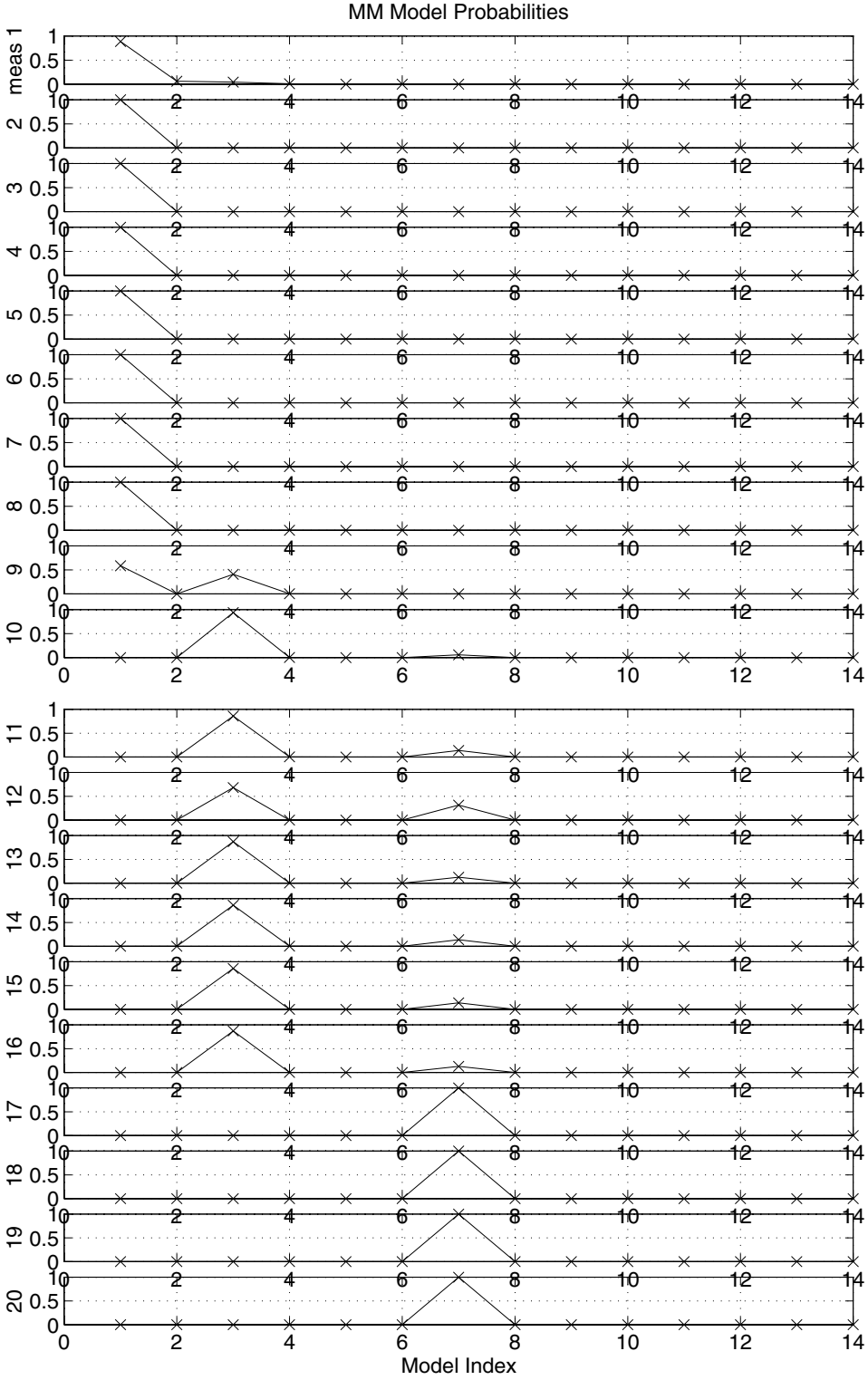

Figure 4: MM and MAP Bayesian model probabilities

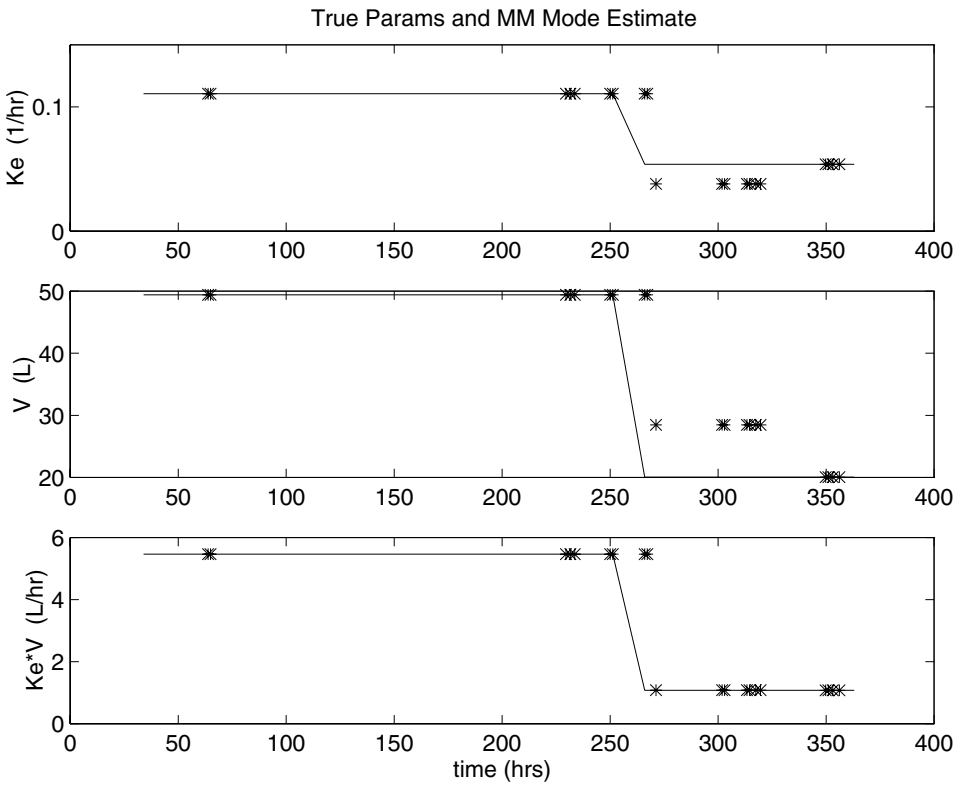

Figure 5: PK Parameters: True (solid) and MM mode estimates (*)

7.4 IMM Estimation Results

Here, the IMM estimation approach discussed in Section 6, is applied to updating posteriors. For this purpose, a value of $\alpha = .99999$ is used to define the Markov Chain probabilities in (60). The IMM estimate of the output y is shown as the solid line in Figure 6. The estimated output agrees well with the first 7 measurements, which all occur before the jump at time $t = 255$ hours. After the jump, it only takes until the 12th measurement for the IMM estimator to track correctly (at time $t = 303$ hours).

The propagation of the posterior model probabilities is shown in Figure 7. Before the jump at $t = 255$ hours, the estimator settles (correctly) on model $\ell = 1$ with nearly probability one. At measurement 8 (the first measurement after the jump occurs) the probability of model $\ell = 1$ drops sharply, and the estimator converges immediately (and correctly) to model $\ell = 7$. This 7th model is maintained with nearly probability one for the remainder of the updates.

Sequential MM mode estimates of K_e, V, and $K_e * V$ are plotted in Figure 8. A drop in true and estimated values is seen after the jump time $t = 255$ hours.

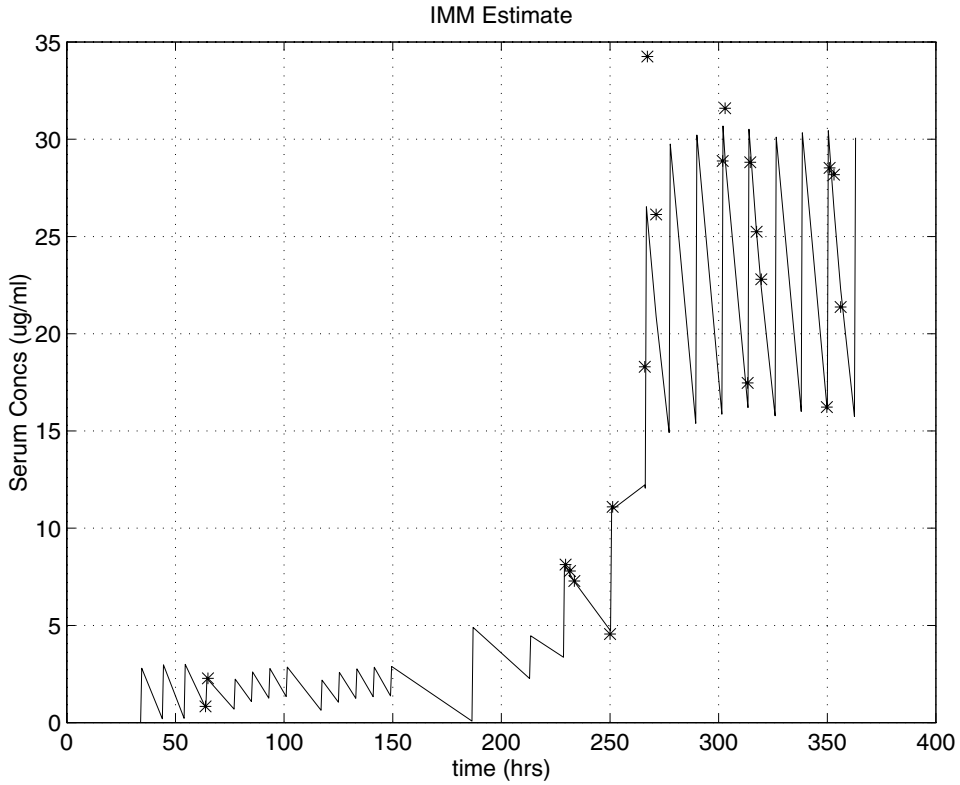

Figure 6: IMM estimate of serum concentration (solid), and measurements (*)

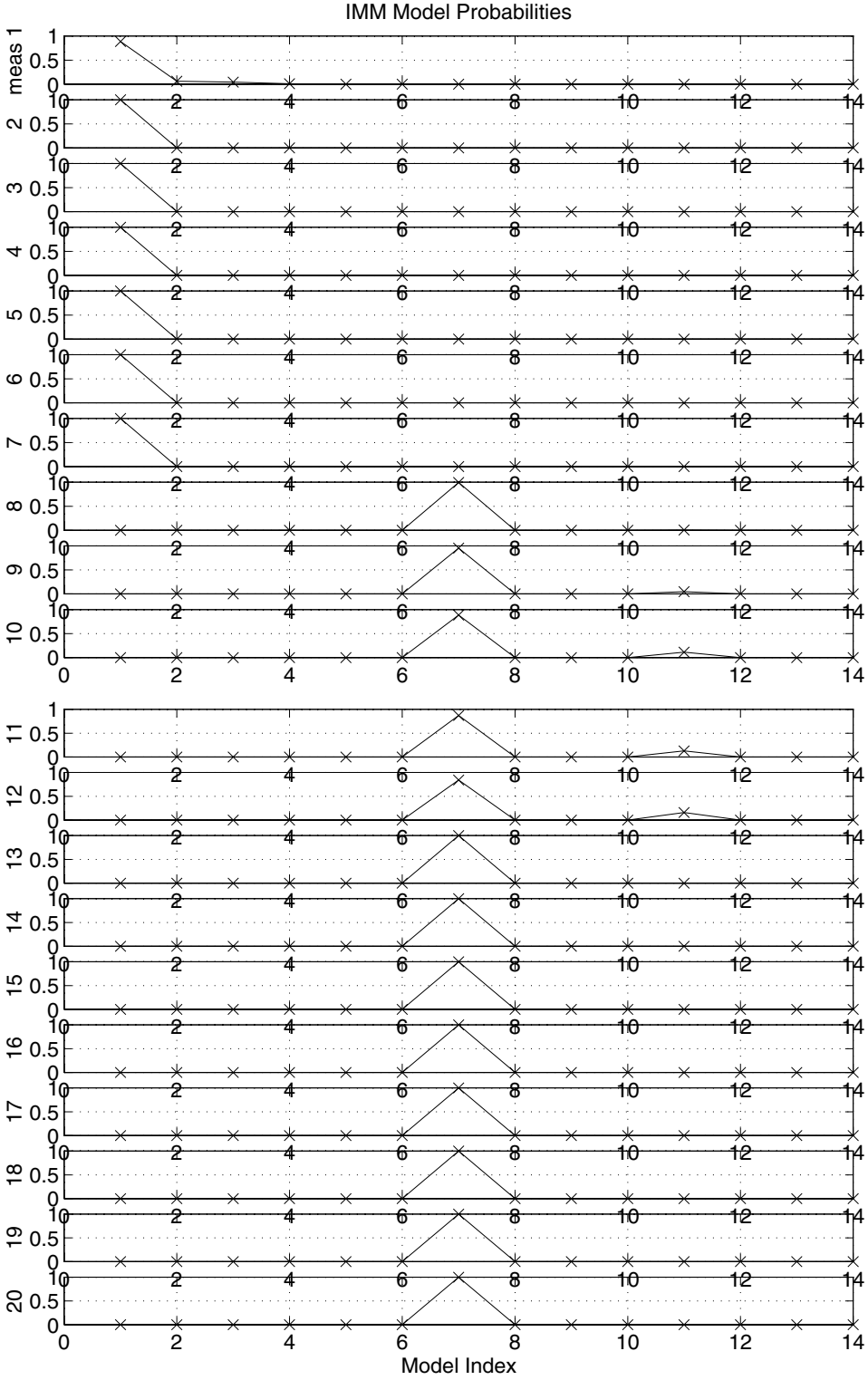

Figure 7: IMM model probabilities

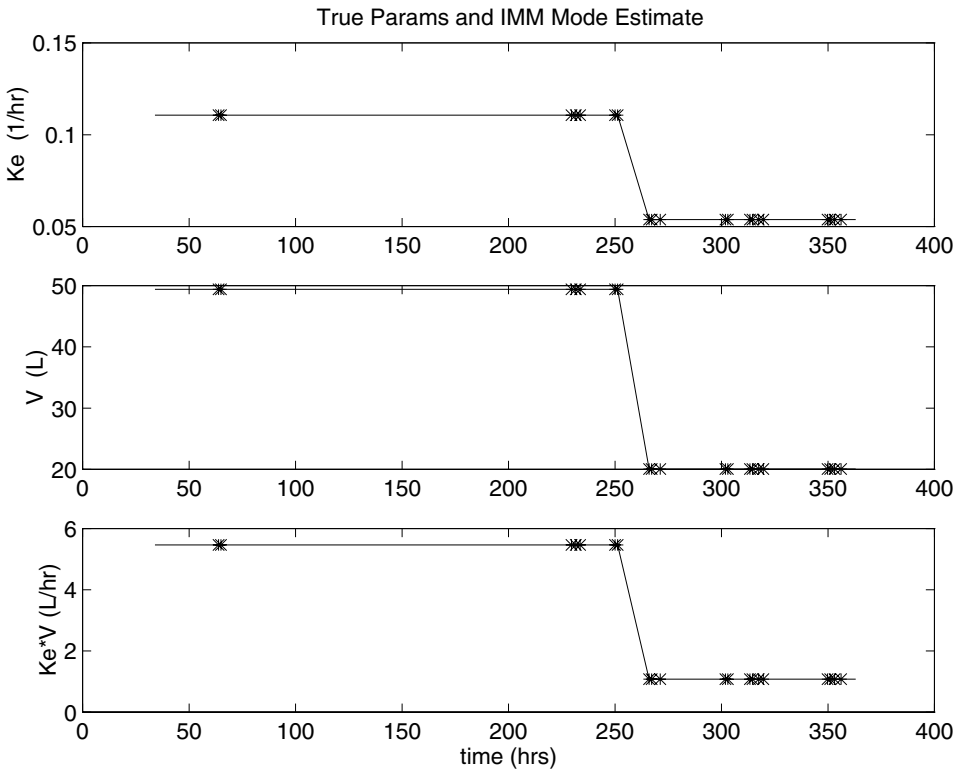

Figure 8: PK Parameters: True (solid) and IMM mode estimates (*)

7.5 Comparison of MM, MAP Bayesian and IMM

Comparing model probabilities in Figure 4 with Figure 7, the IMM algorithm detects the jump immediately, while the MM and MAP Bayesian estimates do not react to the jump until the second measurement afterwards, and in addition are seen to take a brief (incorrect) excursion to model $\ell = 3$. This same relative detection performance can be seen by comparing the sequential MM and IMM parameter estimates in Figure 5 and Figure 8, respectively.

As an additional means of comparison, the true estimation errors (known from simulation) are plotted in Figure 9. The IMM (solid line) has a much smaller estimation error and approximately half the total integrated error compared to MM (dashed) and MAP Bayesian (dotted).

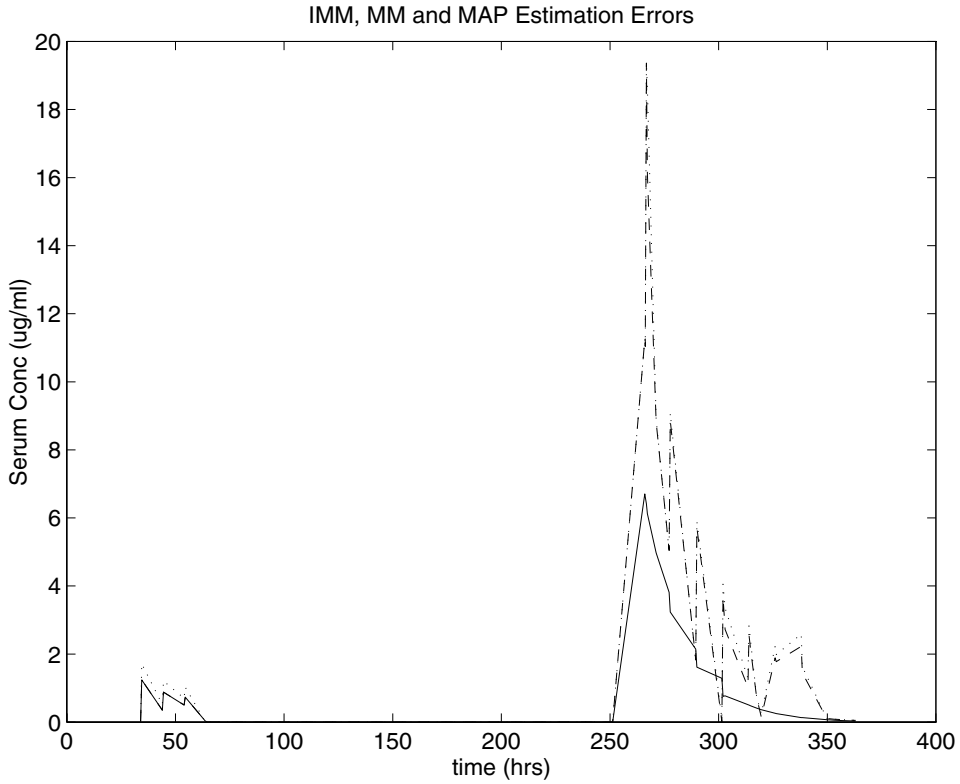

Figure 9: Comparison of estimation errors for IMM (solid), MM (dashed), and MAP Bayesian (dotted), for patient with jump in parameter values at time $t = 255$ hours

8 Numerical Example (Clinical Data)

8.1 IMM Estimation Results (Nominal)

The IMM estimator was next applied to a jump parameter scenario using real clinical data. The clinical data set uses the Tobramycin dose regimen shown Figure 1 and has measured body weight and creatinine clearance values given in Table 2, (this motivated using this same data set in the earlier simulation study). The actual measured clinical serum Tobramycin concentrations found in the patient are shown with "*" in Figure 10. The sharp increase in measured levels after 200 hours shows the change in the patient's serum concentrations as the result of having to give much higher doses to reach desired concentrations.

Background on this clinical data set is briefly given. In the fall of 1991, through the courtesy of Dr. Evan Begg, Division of Clinical Pharmacology, Christchurch Hospital, Christchurch, New Zealand, one of the authors (R.J.) had the opportunity to analyze data of a patient receiving Tobramycin therapy for pyelonephritis. For the first 150 hours, he appeared to be responding adequately to his therapy. When the two serum concentrations during this time (the first two data points in Figure 10) were fitted using the nonsequential MAP Bayesian procedure in the USC*PACK software [12], his V_s was $0.183 L/kg$, his K_s was $0.0028\ hr^{-1}$/unit of creatinine clearance (CCr), and K_i was $0.0069\ hr^{-1}$. Between 150 and 180 hours, he had a surprising relapse and went into clearcut septic shock. Using the same USC*PACK software to analyze only his subsequent data (the remaining data points in Figure 10), his V_s had increased over threefold to $0.608\ L/kg$, his K_s to $0.0062\ hr^{-1}$/unit of CCr, and his K_i to $0.025\ hr^{-1}$.

The IMM estimate of the serum concentration is shown (solid line) in Figure 10. The tracking of the measured serum concentrations is quite acceptable.

The sequential IMM propagation of the posterior model probabilities with each measurement update is shown in Figure 11. After the second measurement, model $\ell = 11$ (having parameters $V_s = 0.193\ L/kg$, $K_s = 0.0037\ hr^{-1}$/unit of CCr, and $K_i = 0.0069\ hr^{-1}$), was found with nearly probability one. However, after the third measurement, a jump is indicated by the sharp drop in probability of model $\ell = 11$ to nearly zero, and the associated increase in probability of the model $\ell = 2$ (having parameters $V_s = 0.784\ L/kg$, $K_s = 0.00098\ hr^{-1}$/unit of CCr, and $K_i = 0.0069\ hr^{-1}$), to nearly one. Model $\ell = 2$ then dominates in probability over the remainder of the horizon, although there are brief flirtations (some of them strong) with models $\ell = 6, 10, 12, 14$ and 9 along the way. Intuitively, IMM is explaining measured data by finding appropriate combinations of responses in the model set.

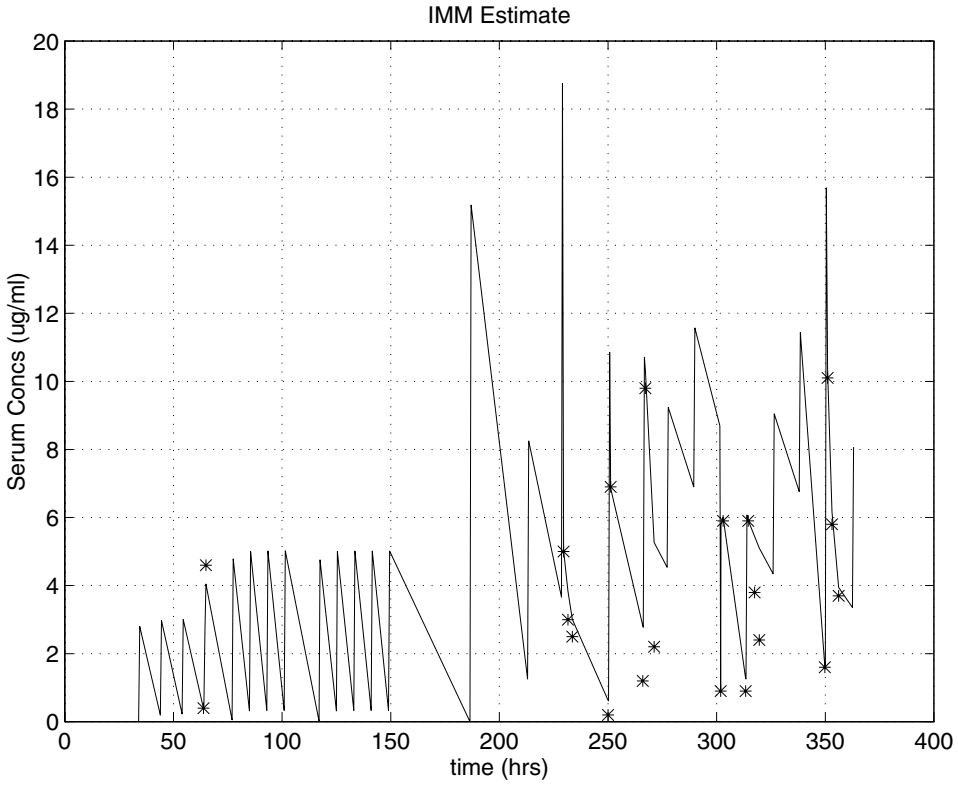

Figure 10: IMM estimate of serum concentration (solid), and measurements (*) for clinical data set. At about $t = 180$ hours the patient condition changed significantly, and septic shock occurred.

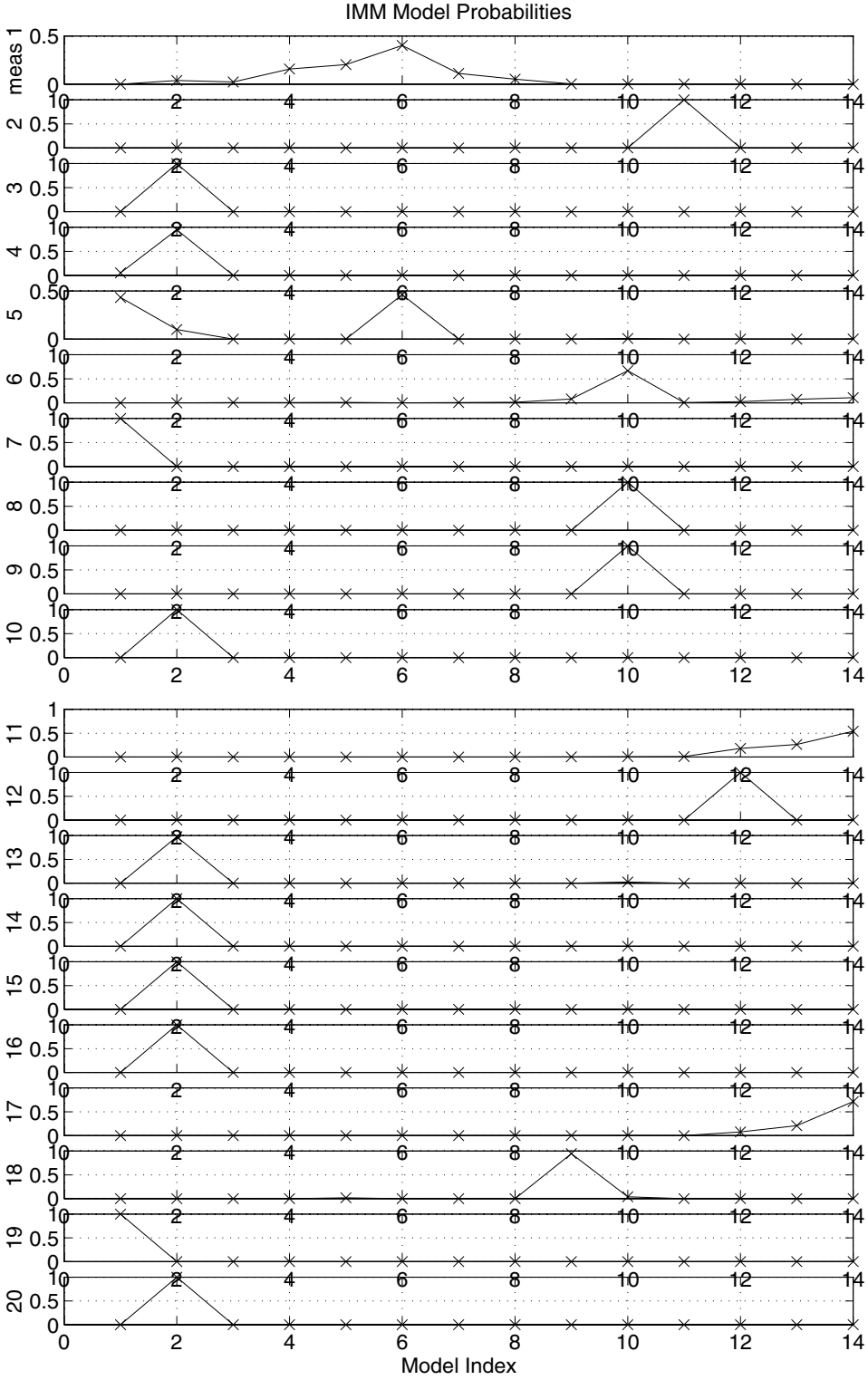

Figure 11: IMM model probabilities for clinical data set

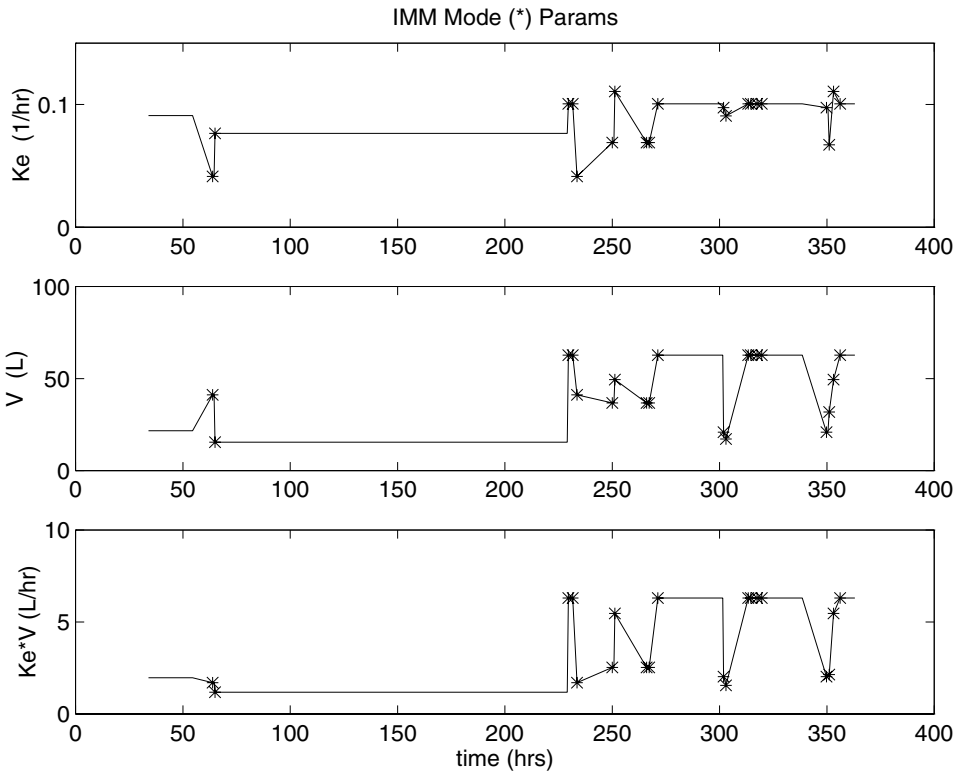

Figure 12: PK Parameters: IMM mode estimate (*) for clinical data set

8.2 IMM Estimation Results (with Extra Model)

The various fluctuations seen in the IMM model probabilities of Figure 11 indicate that there may not be a single model in the discrete population model set which adequately captures the response *after* the jump has occurred.

For this reason, an additional model (fitted to the third and subsequent measurements using the nonsequential MAP Bayesian procedure in the USC*PACK software [12]), was used to augment the original NPEM model. The additional model $\ell = 15$ has the parameters,

$$K_s(15) = 0.0062,\ V_s(15) = 0.608,\ K_i(15) = 0.025 \tag{77}$$

The IMM estimator was run again on the clinical Tobramycin data set using the augmented population model. Results are shown in Figure 13. An improvement in tracking of the measured data is seen compared to the earlier results in Figure 10 using the population model without augmentation.

The posterior model probabilities are shown in Figure 11, There is a clear indication that IMM is trying to converge on the new model $\ell = 15$ after the jump has occurred, although as before, there are a few fluctuations along the way.

It is worth mentioning that the fluctuations in the IMM estimate seen at measurements 9 and 18 in Figure 11 correspond to abnormally high peaks seen in Figure 13, and hence may actually be indicative of additional true patient parameter changes at these brief moments.

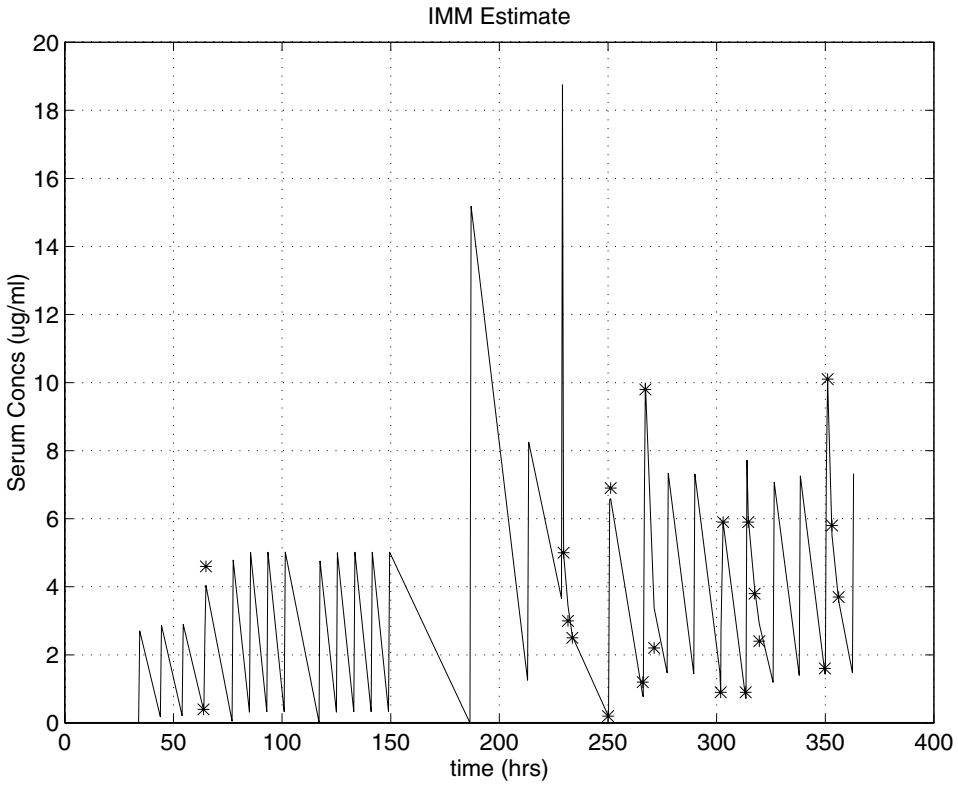

Figure 13: IMM estimate of serum concentration (solid), measurements (*): with extra model, clinical data set

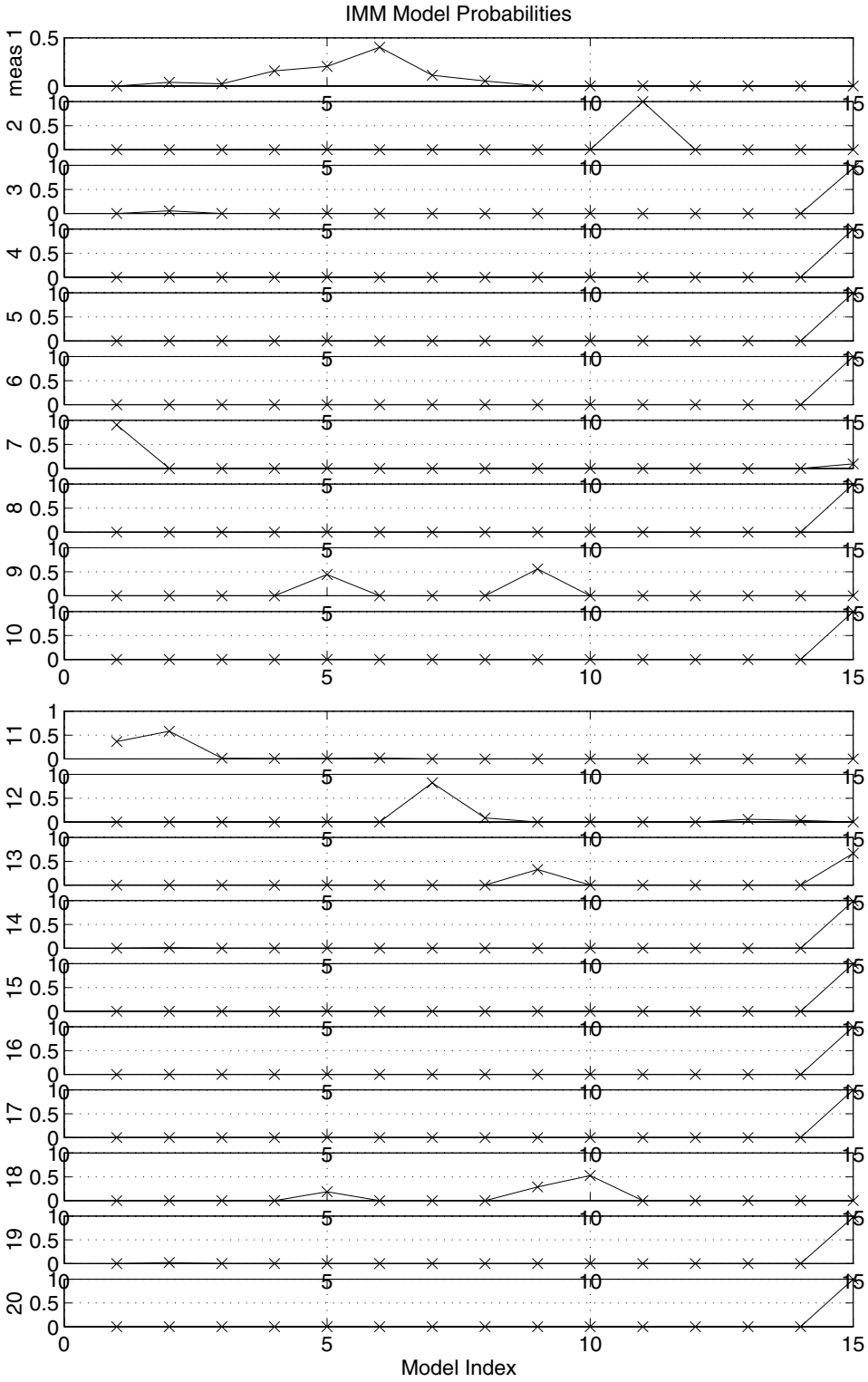

Figure 14: IMM model probabilities: extra model, clinical data set

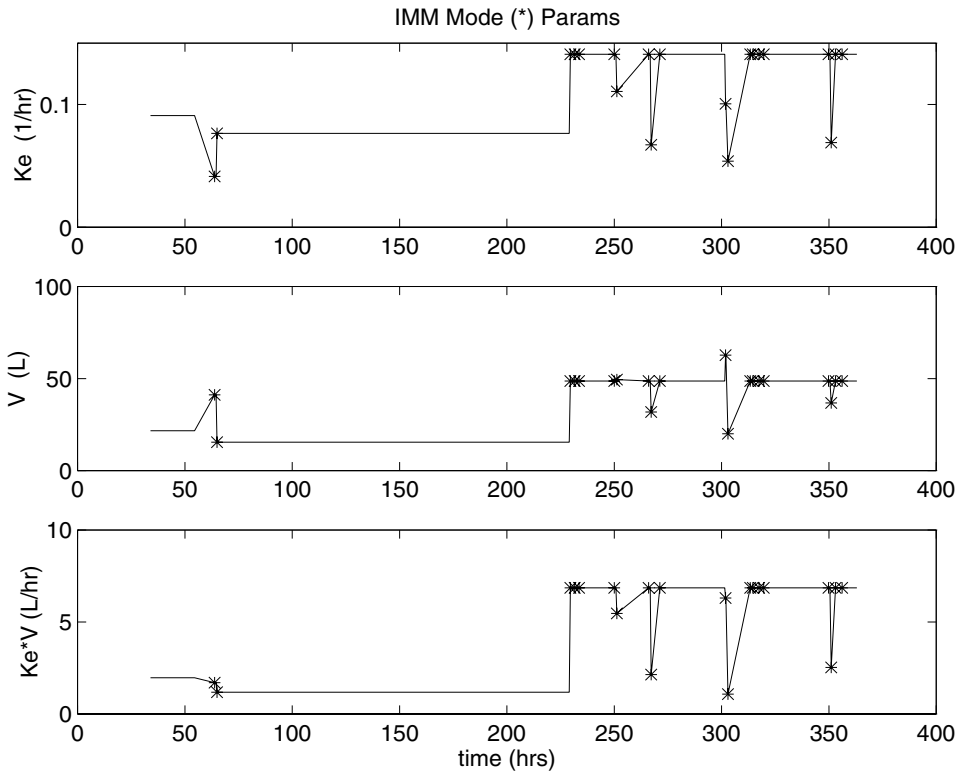

Figure 15: PK Parameters: IMM mode estimate (*): extra model, for clinical data set

9 Conclusions

We have examined the use of the interacting multiple model (IMM) estimation algorithm for clinical tracking of time-varying pharmacokinetic parameters. The performance of the IMM estimator was compared with the sequential MM and MAP Bayesian estimators using a simulated example involving the drug Tobramycin. The IMM algorithm was able to detect a simulated jump cleanly and immediately without delay, compared to the MM and MAP Bayesian estimates which did not react immediately to the jump, and which had more fluctuations before converging to the correct model. The simulation example also showed that the IMM algorithm had only about half the total integrated error relative to both the MM and MAP estimates over the course of the treatment.

The IMM algorithm was then applied to a real clinical data set. The IMM algorithm was not as decisive as it was on simulated data, but it was able to track the response of the changing patient quite successfully. An apparent difficulty was that the truth model after the jump was not well represented within the IMM's discrete set of models. To overcome this, an additional model was added to the model set. The IMM was re-run and was able to catch the jump cleanly, and with improved tracking performance. The good tracking behavior of the IMM estimation algorithm suggests that it could be used beneficially to extend the present Multiple Model (MM) active control approach in Bayard (1995) to computing dosage regimens for changing patients. This conjecture remains to be examined in more detail in the future.

The results indicate that the IMM algorithm is a good method for tracking time-varying pharmacokinetic parameter values, especially when the discrete prior contains models which are representative of the truth model before and after any jumps. Ideally, existing population models should be augmented with models that support the consideration of such jump parameter scenarios. In practice, this might be done by,

1. using population models based on more patients studied, not just the 14 subjects used in the present example
2. detecting jumps in past records and refitting models to the data separately before and after the jumps,
3. developing new population modeling techniques which systematically account for jumps in the data.

The best method for augmenting existing population models for the purpose of tracking such jumps remains as an interesting area for further investigation.

References

[1] Bayard, D.S., Jelliffe, R.W., Schumitzky, A., Milman, M.H., and Van Guilder, M., *Precision drug dosage regimens using multiple model adaptive control: Theory, and application to simulated vancomycin therapy*, in Selected Topics in Mathematical Physics, Prof. R. Vasudevan Memorial Volume, Ed. by R. Sridhar, K. S. Rao, V. Lakshminarayanan, World Scientific Publishing Co., Madras, pp. 407-426, 1995.

[2] Lindsay, B.G., "The geometry of mixture likelihoods: A general theory", Ann. Statist., vol. 11, pp. 86-94, 1983.

[3] Mallet A., "A maximum likelihood estimation method for random coefficient regression models", Biometrika, vol. 73, pp. 645, 1986.

[4] Schumitzky A., "Nonparametric EM Algorithms for Estimating Prior Distributions", App. Math. and Computation, vol. 45, pp. 143-158, 1991.

[5] Lainiotis, D.G.,"Partitioning: A unifying framework for adaptive systems, Part 1: Estimation", Proc. IEEE, vol. 64, no. 8, pp. 1126-1143, 1976.

[6] Schumitzky A., Bayard, D.S., Milman, M.H., and Jelliffe, R.W.,"Multiple model linear quadratic control of pharmacokinetic Systems," Proc. 1994 Western Multiconference: Simulations in the Health Sciences, pp. 99-104, 1994.

[7] Bayard, D.S., Milman M.H., and Schumitzky, A., "Design of dosage regimens: A multiple model stochastic control approach", International Journal of Bio-Medical Computing, vol. 36, pp. 103-115, 1994.

[8] Tugnait, J.K., "Detection and estimation for abruptly changing systems," Automatica, vol. 15, no. 5, pp. 607-615, 1982.

[9] Mazor, E., Averbuch, A., Bar-Shalom, Y., Dayan, J., "Interacting multiple model methods in target tracking: A survey", IEEE Trans. Aerospace and Electronic Systems, vol. 34, no. 1, pp. 103-123, January 1998.

[10] Blom, H.A.P., "An efficient filter for abruptly changing systems," Proc. 23rd Conf. on Decision and Control, Las Vegas, NV, December 1984.

[11] Blom, H.A.P., Bar-Shalom, Y., "The Interacting Multiple Model algorithm for systems with Markovian switching coefficients," IEEE Trans. Automatic Control, vol. 33, pp. 780-783, no. 8, August 1988.

[12] Jelliffe, R.W., Schumitzky, A., Van Guilder, M. *et. al.*, User Manual for Version 10.7 of the USC*PACK Collection of PC Programs. Los Angeles: Laboratory of Applied Pharmacokinetics, University of Southern California School of Medicine, 1995.

[13] Bayard, D.S., Jelliffe, R.W., "Bayesian estimation of posterior densities for pharmacokinetic models having changing parameter values," Proc. SCS 2000 Western Multiconference, Int. Conf. on Health Sciences Simulation, San Diego, CA, January 23-27, 2000.