

基于 Unscented 卡尔曼滤波器的信号转导通路 参数与不可观测状态估计

刘太元¹, 贾建芳^{1,2}, 王宏¹, 岳红¹

(1. 中国科学院自动化研究所, 北京 100080; 2. 中北大学自动化系, 太原 030051)

摘要: 系统生物学的一个主要目的是建立细胞信号转导通路的数学模型。然而细胞信号转导通路多具有较强非线性、参与生化反应物质多等特点, 同时测量数据往往不完备, 并且混有来自实验各个阶段的各种噪声, 这些都给模型参数估计带来很大困难。该文采用 Unscented 卡尔曼滤波器估计信号转导通路未知参数与模型不可观测状态。以肿瘤坏死因子诱导的核转录因子 κB 信号转导通路为例进行了仿真, 结果表明, 采用该方法可以在噪声干扰下较准确地估计系统未知参数和不能观测状态。

关键词: 系统生物学; Unscented 卡尔曼滤波器; 参数估计; 信号转导通路

中图分类号: Q332, N945.14

0 引言

今天, 系统生物学^[1-3]已经吸引了越来越多的关注, 成为生命科学领域研究的热点。系统生物学的一个主要目标是建立细胞生化网络, 特别是细胞信号转导通路和代谢网络的数学模型, 以用来分析细胞的动态属性, 预测未来输出。先进技术和实验方法的不断发展使得现在可以获得关于蛋白质组、基因组和代谢组学等高通量的数据, 这些数据包含有关于生化网络拓扑结构和动态特性的丰富信息。为了提取这些有用的信息, 需要建立生化网络的数学模型。通过建立数学模型, 可以将散乱的信息整理成一个连续、有序的整体; 可以从逻辑上更严密地研究和计算生化网络中关键的组分和相互作用, 对通常的想法和观点做出重要的校正, 并且获得关于生化网络的一些本质属性^[4]。一个好数学模型应当建立在实验数据的基础上, 它不仅能够复现实验取得的数据, 同时还要能够在尽可能小的误差水平下, 预测系统未来输出。质量作用定律和生化系统理论给生化网络的建模提供了一个框架。在这些框架下, 人们已经建立了一部分细胞生化网络的数学模型^[5-9], 并应用于科学研究中, 取得了大量成果。

数学模型的结构确定以后, 随之而来的便是模型中未知参数的辨识问题。参数辨识, 工程上也称为逆向建模问题, 它的目的是通过测量数据来确定模型中未知的参数。模型中的参数一部分可以从文献中查到, 一部分可以通过实验测量得到, 还有一

部分则必须根据观测数据, 采用辨识的方法来获得。系统的动态特性往往对某些参数十分敏感, 能否准确地估计系统的未知参数直接决定着一个数学模型的好坏。细胞生化网络多具有以下特点: 1) 存在强的非线性特性, 有些系统甚至存在混沌现象^[10]。2) 参加生化系统反应的组分众多, 这直接导致系统的数学模型的维数较高, 同时, 随着系统维数的增加, 参数数量更是急剧增加。如由 Hoffman 等^[5]建立的肿瘤坏死因子 α (tumor necrosis factor- α , TNF α) 诱导的核转录因子 (nuclear factor κB , NF- κB) 信号转导通路数学模型包括 24 个状态变量和 64 个反应, Cho 等^[11]建立的信号转导数学模型有 31 个状态变量和 94 个反应, 而 Schoeberl 等^[6]建立的表皮生长因子 (epidermal growth factor, EGF) 信号通路更是包括 94 个状态变量和 125 个反应, 复杂的模型结构不仅使得参数难以精确辨识, 同时使得辨识参数的计算量大大增加。3) 从实验得到的数据往往混杂着来自实验各阶段的种种噪声, 且对于很多系统往往只能得到一部分状态的测量数据, 其它的参加生化反应的各种组分的时间响应序列往往无法得到。所有这些都给生化系统的参数辨识工作带来极大的挑战。

收稿日期: 2006-07-28

基金项目: 中国科学院海外杰出学者基金项目 (2004-1-4)

通讯作者: 刘太元, 电话: (010)82617643,

E-mail: taiyuan.liu@ia.ac.cn

Moles等比较了各种参数优化的方法, 结果发现, 广泛使用的基于梯度的优化方法往往不能够得到参数的全局最优值^[12]。建立有效的优化方法来估计网络模型的未知参数仍然是复杂生化网络参数辨识领域的具大挑战^[13]。

除了基于优化的思想来估计参数以外, 另外一种方法是采用基于迭代预测误差的方法来估计参数。同优化方法不同, 基于迭代预测误差的方法首先给定参数的初始估计值, 在该初始估计值下预测系统的输出, 然后计算预测输出与实际测量之间的误差, 并根据这一误差值来调整参数估计值, 得到新的参数估计值, 在新的参数估计值下, 产生新的预测输出, 这一过程一直进行下去, 直到满足某种条件为止。

本文采用 Unscented 卡尔曼滤波器 (unscented kalman filter, UKF) 来估计信号转导通路的未知参数和不可观测状态。UKF 作为卡尔曼滤波器的一种扩展形式, 同典型的扩展卡尔曼滤波器相比, 它不需要计算模型的雅可比矩阵或海森矩阵, 它的实现只需要矩阵的代数运算即可完成, 计算复杂度大大降低。同时由于卡尔曼滤波器的迭代结构, 参数估计值停止在局部最小值处的概率大大降低。采用这种方法来估计系统的参数值, 未知参数被看作原系统的增广状态变量, 同原系统状态变量一起组成一新的系统, 将 UKF 应用到新生成的系统, 便可以得到对未知参数和不可观测状态的估计。

1 方 法

1.1 状态估计与卡尔曼滤波器

考虑如下的一般动态系统:

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \boldsymbol{\theta}, \boldsymbol{\omega}_k) \\ \mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{v}_k) \end{cases} \quad (1)$$

式中 $\mathbf{x}_k \in \mathbf{R}^n$ 表示系统状态向量, $\mathbf{y}_k \in \mathbf{R}^m$ 表示系统输出向量, $\boldsymbol{\theta} \in \mathbf{R}^p$ 为系统未知参数向量, $\boldsymbol{\omega}_k$ 为均值为 0, 方差为 \mathbf{R}_k 的过程噪声, \mathbf{v}_k 为均值为 0, 方差为 \mathbf{Q}_k 的输出噪声, \mathbf{f} 、 \mathbf{h} 为已知线性或非线性函数。由于随机噪声的影响, 系统状态和输出同样成为随机变量。在采样时刻 t_1, t_2, \dots, t_N , 可以得到输出的一组采样值 $Y = \{Y_1, Y_2, \dots, Y_N\}$, 通过滤波器的方法来进行状态估计实质上是求取模型状态在已知测量输出条件下的概率密度 $\rho(\mathbf{x}_N | Y_1, Y_2, \dots, Y_N)$ 。当系

统为线性时, 系统可以如下表述:

$$\begin{cases} \mathbf{x}_k = \mathbf{A} \mathbf{x}_{k-1} + \boldsymbol{\omega}_{k-1} \\ \mathbf{y}_k = \mathbf{C} \mathbf{x}_k + \mathbf{v}_k \end{cases} \quad (2)$$

对方程(2)所示的动态系统, 典型的卡尔曼滤波器来估计系统状态的算法如下:

$$\begin{cases} \hat{\mathbf{x}}_k = \mathbf{A}_k \hat{\mathbf{x}}_{k-1} + \mathbf{K}_k (\mathbf{y}_k - \mathbf{C}_k \mathbf{A}_k \hat{\mathbf{x}}_{k-1}) \\ \mathbf{K}_k = \mathbf{P}_k \mathbf{C}_k^T (\mathbf{C}_k \mathbf{P}_k \mathbf{C}_k^T + \mathbf{R}_k)^{-1} \\ \mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{C}_k) \mathbf{P}_k' \\ \mathbf{P}_k' = \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^T + \mathbf{Q}_{k-1} \end{cases} \quad (3)$$

这里 $\hat{\mathbf{x}}_k$ 为在 k 时刻状态的估计值, \mathbf{K}_k 为卡尔曼增益矩阵, \mathbf{P}_k 为估计误差的协方差阵, \mathbf{P}_k' 为预测误差的协方差阵。

对于非线性系统, 扩展卡尔曼滤波器 (extended kalman filter, EKF) 是使用最广泛的方法。EKF 通过泰勒级数展开的方法来一阶或二阶近似非线性函数 \mathbf{f} 和 \mathbf{h} , 对于一阶近似, 可以得到如下的雅可比矩阵:

$$\mathbf{F}_k = \left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^T} \right|_{\mathbf{x}=\hat{\mathbf{x}}_k}; \quad \mathbf{H}_k = \left. \frac{\partial \mathbf{h}(\mathbf{x})}{\partial \mathbf{x}^T} \right|_{\mathbf{x}=\hat{\mathbf{x}}_k} \quad (4)$$

大多数一阶 EKF 的估计都是有偏或不一致的, 为了解决这个问题, 出现了卡尔曼滤波器的多种扩展形式, UKF 便是其中的一种。

1.2 Unscented 变换

UKF 首先由 Julier 和 Uhlman^[14]提出, 并且被推广到包括非线性系统辨识、神经网络训练等非线性估计领域^[15-18]。UKF 的提出基于如下的思想: 即对随机变量概率密度函数的近似要比对非线性函数的近似来的更加容易^[19], UKF 采用逼近概率密度函数的方法, 而不像扩展卡尔曼滤波器那样逼近非线性函数。UKF 通过一组确定的采样点, 称为 sigma 点, 来近似概率密度函数, 并且它无须计算雅可比矩阵或海森矩阵。UKF 建立在 unscented 变换 (unscented transformation, UT) 的基础上。

假设 \mathbf{x} 是 n 维随机变量, 且它的均值为 $\bar{\mathbf{x}}$, 它的方差为 \mathbf{P}_{xx} , \mathbf{y} 为另外一个随机变量, 并且

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) \quad (5)$$

UT 变换是一种计算随机变量经过非线性变换以后的统计特性的方法。如图 1 所示, UT 变换是在保证采样均值和方差分别为 $\bar{\mathbf{x}}$ 和 \mathbf{P}_{xx} 的前提下, 构造一组确定性点集, 被称为 sigma 点。将非线性变换作用于这组采样点, 则 \mathbf{y} 的统计特征便可以由变换后的点集得到, 对于 n 维的向量, 采样点为

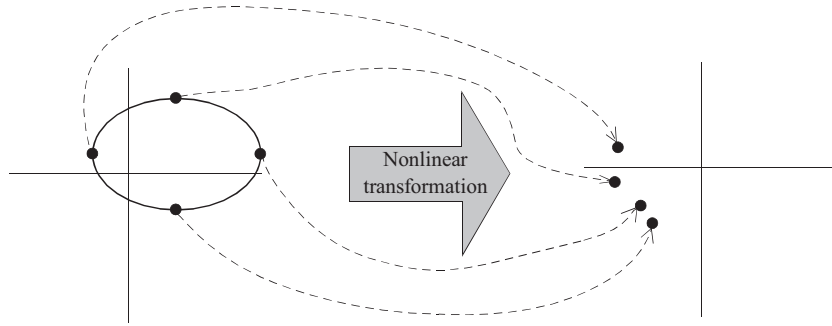


Fig.1 The unscented transformation. A set of weight sigma points is selected deterministically in order that the sample mean and sample variance match those of the priori distribution. The nonlinear function is applied to the data set and the predicted mean and covariance can be calculated from the transformed data set

2n+1 维。sigma 样点的构造如下:

$$\begin{cases} \chi_0 = \bar{x} \\ \chi_i = \bar{x} + [\sqrt{(n+\lambda)P_x}]_i, & i=1, \dots, n \\ \chi_i = \bar{x} - [\sqrt{(n+\lambda)P_x}]_{i-n}, & i=n+1, \dots, 2n \\ W_0^{(m)} = \lambda / (n+\lambda) \\ W_0^{(c)} = \lambda / (n+\lambda) + (1-\alpha^2 + \beta) \\ W_i^{(m)} = W_i^{(c)} = 1 / [2(n+\lambda)], & i=n+1, \dots, 2n \end{cases} \quad (6)$$

这里 $\lambda = \alpha^2(n + \kappa) - n$ 为尺度参数, α 决定了 sigma 点的分散程度并且通常设为一个小的正数如 $1e-3$ 。 κ 是另外一个尺度调节参数, 通常设置为 0, β 用来包含关于 x 概率分布的先验知识, 对于服从高斯分布的随机变量来说, $\beta = 2$ 是最优的^[20]。 $[\sqrt{(n+\lambda)P_x}]_i$ 是矩阵平方根的第 i 行。 $W_i^{(m)}$, $W_i^{(c)}$ 分别为对应均值和协方差的加权值。这些 sigma 点非线性函数变换后得到:

$$y_i = f(x_i) \quad i=0, \dots, 2n \quad (7)$$

而 y 的均值和方差通过对变换后的 sigma 样点的加权获得:

$$\begin{cases} \bar{y} \approx \sum_{i=0}^{2n} W_i^{(m)} y_i \\ P_y \approx \sum_{i=0}^{2n} W_i^{(c)} (y_i - \bar{y})(y_i - \bar{y})^T \end{cases} \quad (8)$$

从以上的构造过程可以看到, sigma 样点的构造方法同一般的采样方法有着显著差别, 它所需的采样点的个数是确定性的 $2n+1$, 而不像 Monte-Carlo 方法那样需要大量的采样点。同时在这些确定性的采样点下, UT 可以保证对高斯分布的三阶精度近似, 对非高斯分布的也至少保证二阶精度的近似, 通过合理的选择 α 和 β , 还可以达到更高阶的近似精度^[21,22]。

1.3 unscented 卡尔曼滤波器

UKF 建立在 UT 变换的基础上, 将 UT 变换用于非线性的卡尔曼滤波上, 便得到了 UKF。当系统噪声为非加性时, 为了应用 UKF 算法, 必须对原系统状态扩维, 使得 $x_k^- = [x_k^T w_k^T v_k^T]^T$, 以便将噪声项也包括进来。而当噪声为 0 均值的加性噪声时则可以不做扩维处理, 以减少运算量。UKF 算法可以概括如下:

1) 初始化

$$\bar{x}_0 = E[x_0], \quad P_0 = E[(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T] \quad (9)$$

2) 在采样时刻 k , 计算 sigma 点集:

$$\chi_{k-1} = [\hat{x}_{k-1}, \hat{x}_{k-1} + \eta \sqrt{P_{k-1}}, \hat{x}_{k-1} - \eta \sqrt{P_{k-1}}] \quad (10)$$

3) 预测

$$\begin{cases} \chi_{k|k-1} = f(\chi_{k-1}) \\ \hat{x}_k^- = \sum_{i=0}^{2n} W_i^{(m)} \chi_{i,k|k-1} \\ P_k^- = \sum_{i=0}^{2n} W_i^{(c)} [\chi_{i,k|k-1} - \hat{x}_k^-][\chi_{i,k|k-1} - \hat{x}_k^-]^T + R^v \\ Y_{k|k-1} = h(\chi_{k|k-1}) \\ \hat{y}_k^- = \sum_{i=0}^{2n} W_i^{(m)} Y_{i,k|k-1} \end{cases} \quad (11)$$

4) 用于非线性卡尔曼滤波, 估计状态

$$\begin{cases} P_{y\bar{y}_k} = \sum_{i=0}^{2n} W_i^{(c)} [Y_{i,k|k-1} - \hat{y}_k^-][Y_{i,k|k-1} - \hat{y}_k^-]^T + R^n \\ P_{x\bar{y}_k} = \sum_{i=0}^{2n} W_i^{(c)} [\chi_{i,k|k-1} - \hat{x}_k^-][Y_{i,k|k-1} - \hat{y}_k^-]^T \\ K_k = P_{x\bar{y}_k} P_{y\bar{y}_k}^{-1} \\ \hat{x}_k = \hat{x}_k^- + K_k (y_k - \hat{y}_k^-) \\ P_k = P_k^- - K_k P_{y\bar{y}_k} K_k^T \end{cases} \quad (12)$$

1.4 基于 UKF 的参数估计

UKF 被提出以后, 广泛应用于非线性估计领域包括非线性系统辨识, 神经网络学习及双估计问题等领域得到广泛应用。通过卡尔曼滤波器来估计系统未知参数有两种方法: 联合状态法和双估计法。联合状态法将待估计的未知参数视为一个新的状态变量, 追加到系统原始状态, 得到增广状态, 然后对增广状态进行卡尔曼滤波从而得到参数的估计值。令 $\bar{\mathbf{x}}(t)$ 为增广未知状态, 则

$$\bar{\mathbf{x}}(t) = \begin{bmatrix} \mathbf{x}(t) \\ \boldsymbol{\theta}(t) \end{bmatrix} \quad (13)$$

并且参数的动态过程表示为

$$\dot{\boldsymbol{\theta}} = \mathbf{0} \quad (14)$$

方程(14)是将未知参数看作一个数值保持不变的常量状态增加到原模型中。采用这种方法, 增广状态的方差阵在滤波过程中将作为一个整体更新。而双估计的方法则包括分别针对原始状态和参数的两个滤波器, 一个估计真实的状态, 一个估计未知参数, 两个滤波器是相互分离的。本文将采用第一种方法来进行参数估计, 关于第二种方法, 读者可以参考文献[23]及其文后参考文献, 这里不在赘述。

2 仿真结果

目前, 建立细胞信号转导通路的数学模型一般是在了解清楚信号转导通路的机理以后建立起来的。在生物系统中, 信号的转导大多通过两种方式进行: 一是蛋白质之间的相互作用和酶促反应, 如蛋白质的磷酸化和脱磷酸化; 二是细胞内信使蛋白的合成与分解。可以使用如下的基本化学反应来概括这些反应:



其中 A、B、AB、C、D 为反应物, k_f 、 k_b 为速率常数。反应物之间的相互作用就由速率常数 k_f 、 k_b , 以及各反应的初始浓度所决定。对于反应物 A 浓度变化, 可以用如下的微分方程来表示:

$$d[A]/dt = k_b[C][D] - k_f[A][B] \quad (16)$$

其中 $[\]$ 表示反应物的浓度。其它反应物的动态变化方程同方程(16)基本相似。如果假设最后一步是不可逆的, 标准的酶促反应则可以看作是按顺序

发生的上述两个反应, 如下所示。



其中 E 为酶, S 为底物, P 为产物, k_1 、 k_2 、 k_3 为速率常数。整个反应完全由 3 个速率常数 k_1 、 k_2 、 k_3 , 以及底物 S、酶 E 和产物 P 的初始浓度所决定。

综上所述, 基于质量作用定律建立的信号转导通路的数学模型可以用一组非线性常微分方程如下表示:

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{g}(\mathbf{x}, \mathbf{p}) \\ \mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{p}) \end{cases} \quad (18)$$

其中 $\mathbf{x} \in R^n$ 为参与信号转导通路的各蛋白质的浓度向量, $\mathbf{y} \in R^m$ 为信号转导通路的测量输出向量, $\mathbf{p} \in R^l$ 为信号转导通路中涉及的参数向量, 包括酶促反应力学常数, 蛋白质绑定速率系数, 细胞质与细胞核之间的传输速率系数等。 $\mathbf{f}(\cdot)$ 、 $\mathbf{g}(\cdot)$ 为关于系统状态的已知函数。

从模型(18)可以看出, 信号转导通路为一非线性动态系统。采用 UKF 来估计信号转导通路参数和状态, 同 EKF 相比, 不需要进行线性近似, 也就不需要计算模型的雅可比矩阵或海森矩阵, 大大降低了运算的复杂程度和计算量。基于 UKF 的信号转导通路参数与不可观测状态估计步骤总结如下:

- 1) 程序初始化, 包括 UKF 参数选择等;
- 2) 按照方程(13)、(14)对模型(18)中未知参数做相应处理, 形成新的增广状态和增广模型;
- 3) 按照方程(6), 根据模型输出的测量数据构造 sigma 点集;
- 4) 对新的增广模型使用 UKF 进行滤波, 直到满足一定条件为止。

上述算法步骤 4) 中, 算法停止的条件可以是滤波的迭代次数, 也可以为参数或状态估计的误差限度。

为了检查 UKF 算法的有效性, 以 TNF α 诱导 NF- κ B 信号转导通路模型为例对其进行了研究。NF- κ B 信号转导通路是在生物进化中细胞对内外界应急做出迅速反应的一条十分保守而重要的通路, NF- κ B 是一种控制着各种基因转录的重要转录调节因子, 它调节大量与细胞应急反应 (如免疫应答、炎症反应和细胞抗凋亡) 作用相关的基因的转录。NF- κ B 功能的失调与许多人类病症如类风湿性关节炎、癌症等直接相关, 许多重要疾病如肿

瘤或免疫性疾病的发生,都是由细胞信号转导失误所致,因此自1986年发现以来,NF- κ B一直是细胞生物学及免疫学领域的研究热点。

NF- κ B通常以同源或异源二聚体非活性形式存在于几乎所有类型细胞的细胞质中,家族的五位成员分别是: NF- κ B1 (p50), NF- κ B2 (p52/p100), Rel A(p65), c-rel, Rel B。很多物质能够激活

NF- κ B,其中肿瘤坏死因子 TNF α 对NF- κ B的活化是研究最广泛而深入的。如图2所示,在未接受外来刺激的细胞中,NF- κ B同一种称为 I κ B 的蛋白质绑定在一起,位于细胞质中,I κ B 家族的主要成员有 I κ B α 、I κ B β 和 I κ B ϵ 。细胞在 TNF α 的刺激下,激活 I κ B 蛋白激酶 (I κ B kinase, IKK), IKK 使 NF- κ B 抑制蛋白 I κ B 氨基端磷酸化,磷酸化的

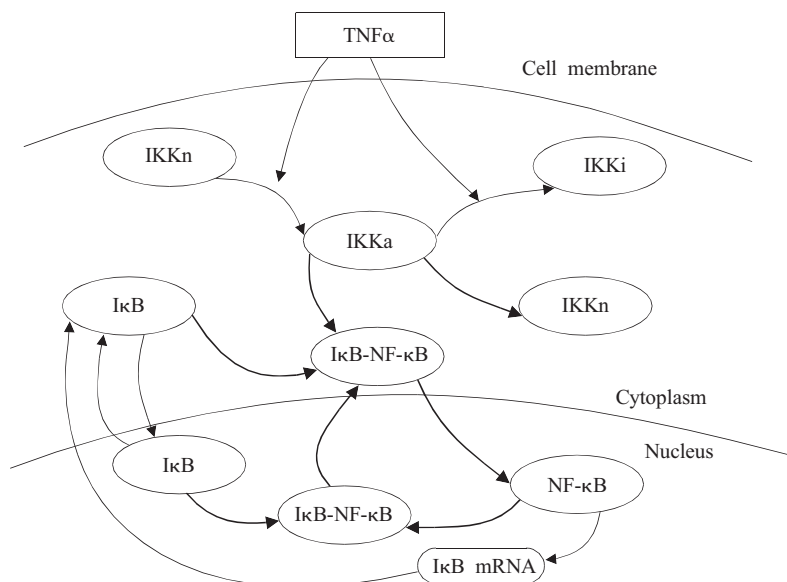


Fig.2 Schematic depiction of TNF α mediated NF- κ B signal transduction pathway. The NF- κ B bound with three I κ B isoforms lies in cytoplasm in normal state. Upon TNF α stimulation, IKK is activated and leads to phosphorylation and degradation of I κ B proteins. Free NF- κ B enters the nucleus, activating genes, including I κ B α , I κ B β and I κ B ϵ

I κ B 进一步被蛋白酶降解,使 NF- κ B 从 NF- κ B-I κ B 复合体中解离, NF- κ B 被活化,其 DNA 结合位点暴露,并转位于细胞核,调节 I κ B 基因的转录。

最近, Hoffmann 等^[5]基于质量作用定律建立了由 TNF- α 诱导的 NF- κ B 信号转导通路模型,很好地描述了 I κ B 家族不同成员在 NF- κ B 信号转导通路中的不同控制作用,这个模型已经被广泛用到与 NF- κ B 信号转导通路相关的各个领域,并取得大量的研究成果^[24-28]。该模型包括 24 个状态变量和 64 个反应,具体的数学模型及参数真值见附录。

仿真在 Matlab 环境中进行。我们选择细胞质中的 IKK-I κ B α 复合物及细胞核中的 NF- κ B、I κ B α 、I κ B β 、I κ B ϵ 为系统输出,其余蛋白质及其复合物作为模型不可观测状态。选择 IKK-I κ B α 结合系数 a_1 、IKK-I κ B α -NF- κ B 结合系数 a_4 、IKK-

I κ B α -NF- κ B 结合系数 a_7 为未知参数,选择这 3 个参数是基于以下原因:研究证明,细胞核内 NF- κ B 蛋白在信号转导过程中所表现出的振荡行为调控着目标基因的表达^[29],同其它参数相比, a_1 、 a_4 和 a_7 对于选定的模型输出核内 NF- κ B 的敏感性系数高,模型对这些参数更为敏感,这些参数的估计结果将直接决定模型质量的优劣^[30,31]。仿真时系统状态的初始值选取系统达到稳态时的状态值,初始状态的获得见附录。采用 Matlab 中的 ODE23s 函数求取模型各状态的数值解,获得系统各状态动态过程的 200 个数据采样点。在此基础上,对各状态数据采样点分别施加均值为 0、方差为各状态最小数值 10% 的高斯噪声,所得到的数据集合作为系统输出测量数据和不可观测状态的实际数据。为了比较 UKF 与 EKF 在信号转导通路的参数和不可观测状态估计中的性能,同时采用扩展

卡尔曼滤波器进行参数和状态估计。未知参数的初始值均选为： $a_1=0.05$ ， $a_4=1.0$ ， $a_7=0.5$ 。得到的参数

估计结果如图 3 所示。

从图 3 可以看出，在经过 100 次叠代（200 分

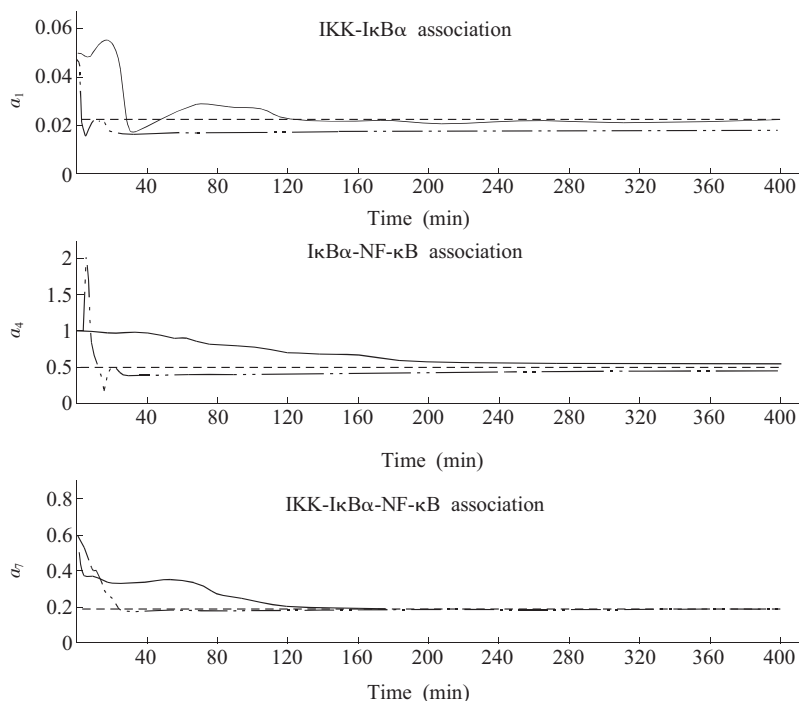


Fig.3 Parameter estimation of TNF α mediated NF- κ B signal transduction pathway. Three parameters including IKK-I κ B α association a_1 , IKKI κ B α -NF- κ B association a_4 and IKK-I κ B α NF- κ B association a_7 are estimated with unscented kalman filter. The dashed line represents the true value, the solid line represents the estimation of UKF and the dotted line represents the estimation of EKF. ---: true; —: UKF; -·-·-: EKF

钟)以后， a_1 、 a_4 、 a_7 的估计值均趋于稳定，估计值接近于真值，而UKF方法得到的估计值精度要高于EKF方法。最后的估计结果见表1。

高于EKF，且均取得了较满意的结果。

3 结 论

Table 1 Simulation results of parameter estimation with UKF and EKF

| Parameters | UKF estimation | EKF estimation | True values |
|------------|----------------|----------------|-------------|
| a_1 | 0.0226 | 0.0178 | 0.0225 |
| a_4 | 0.5217 | 0.4431 | 0.5 |
| a_7 | 0.1852 | 0.1916 | 0.185 |

图4给出了对系统输出的预测仿真结果。从图中可以看出UKF方法对输出的预测结果令人满意。预测值同实际测量值之间的误差很小。

建立生化系统的数学模型，精确辨识模型中未知参数是系统生物学的一个基本任务，然而参加生化反应的组分多，生化反应本身所表现出的强非线性及无法获得完备的实验数据等原因都使得对生化系统模型的辨识问题变得更加困难。本文采用UKF的联合状态方法，对模型参数加以估计，在得到参数估计值的同时，能够得到生化系统无法观测的状态估计。对TNF α 诱导NF- κ B信号转导模型的仿真结果表明，该方法比EKF具有更高的精度，能够同时给出参数和不可观测状态的良好估计。这种方法尤其适合在不能获得充足的测量数据的情况下估计模型参数。

图5, 6, 7给出了对系统不可观测状态的估计。系统的各个不可观测状态都得到了较好的估计，估计值与实际值误差较小。总体来说，在对系统参数和不可观测状态的估计中，UKF的精度要

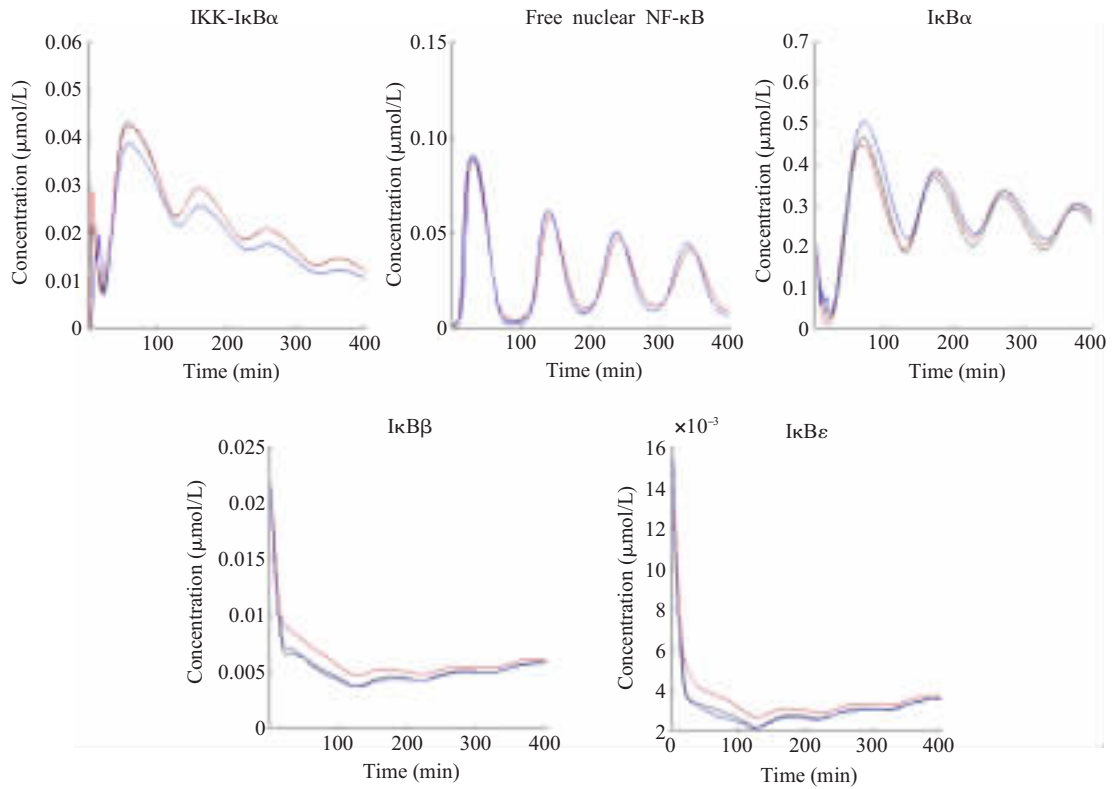


Fig.4 Estimation of system output of TNF α mediated NF- κ B signal transduction pathway. The black solid line is the true value, the red solid line is the estimation with UKF, and the blue line is the estimation with EKF

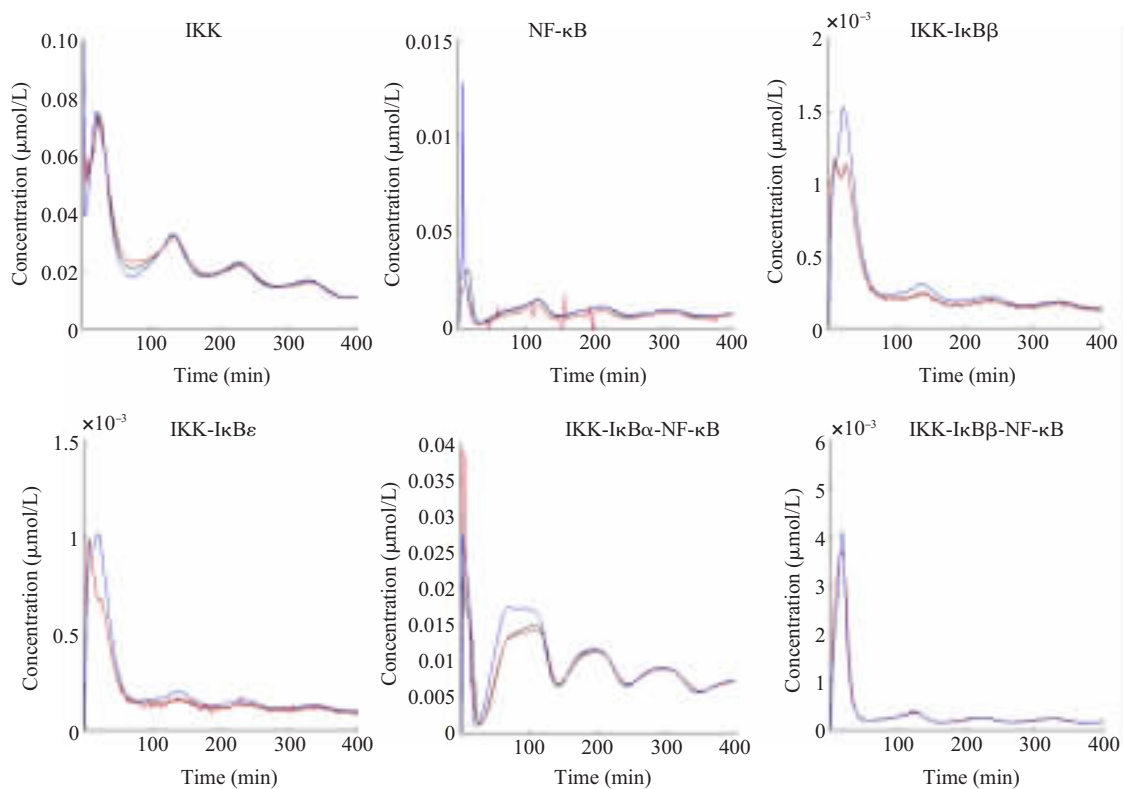


Fig.5 Estimation of unobservable states of TNF α mediated NF- κ B signal transduction pathway. The black solid line represents true values of states and the red solid line is the estimation with UKF, the blue solid line is the estimation with EKF. The unobservable states include IKK, NF- κ B in the nucleus, IKK-I κ B β , IKK-I κ B ϵ , IKK-I κ B α -NF- κ B, IKK-I κ B β -NF- κ B

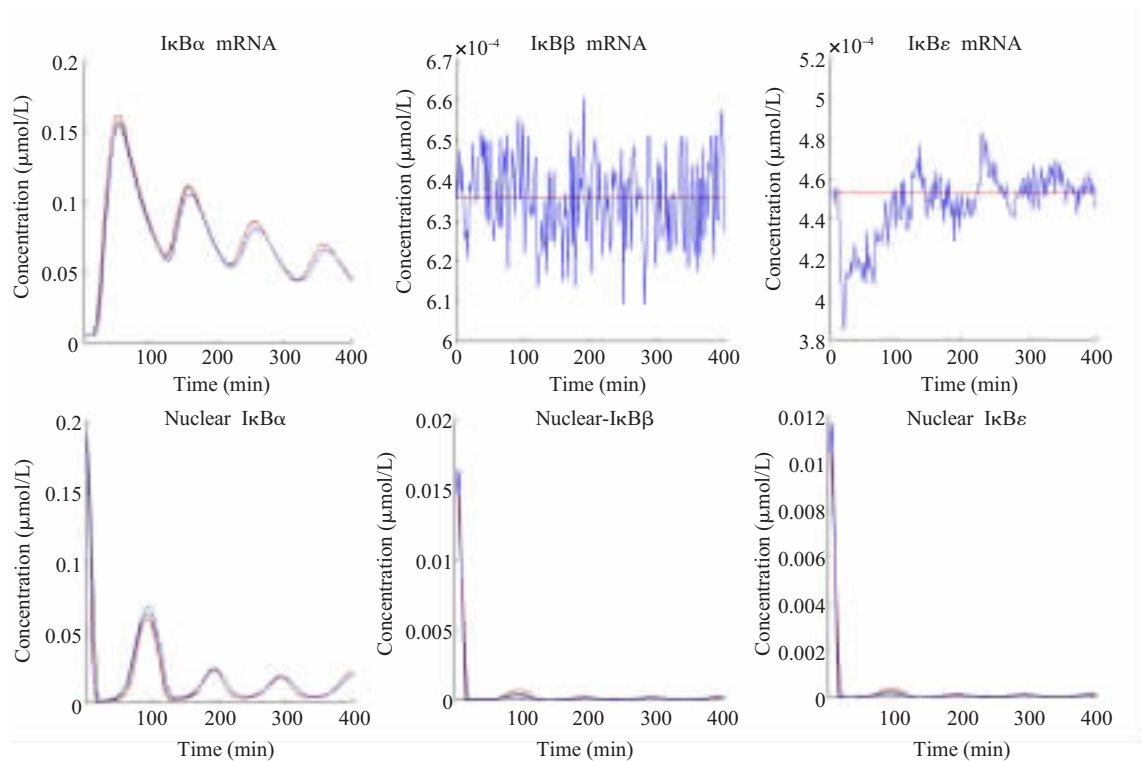


Fig.6 Estimation of unobservable states of TNF α mediated NF- κ B signal transduction pathway. The black solid line represents true values of states and the red solid line is the estimation with UKF, the blue solid line is the estimation with EKF. The unobservable states include: I κ B α mRNA, I κ B β mRNA, I κ B ϵ mRNA, I κ B α in the nucleus, I κ B β in the nucleus, I κ B ϵ in the nucleus

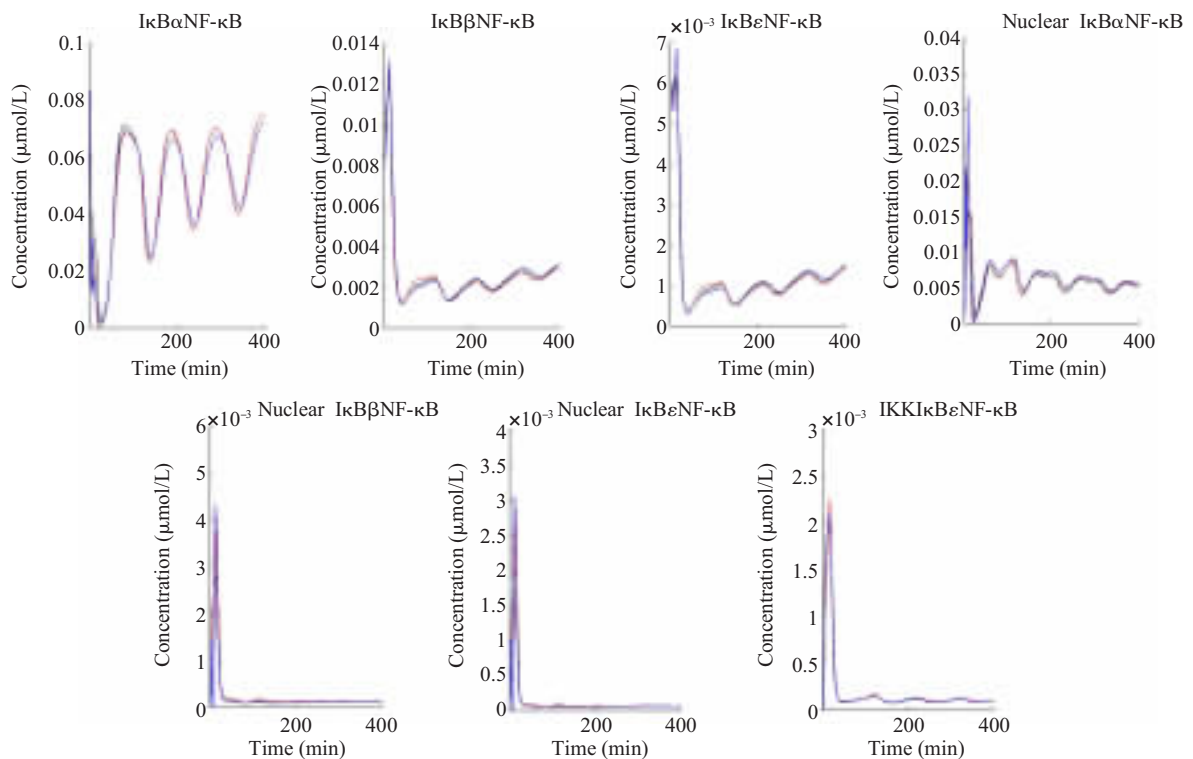


Fig.7 Estimation of unobservable states of TNF α mediated NF- κ B signal transduction pathway. The black solid line represents true values of states and the red solid line is the estimation with UKF, the blue solid line is the estimation with EKF. The unobservable states include I κ B α -NF- κ B, I κ B β -NF- κ B, I κ B ϵ -NF- κ B, I κ B α -NF- κ B in the nucleus, I κ B β -NF- κ B in the nucleus, I κ B ϵ -NF- κ B in the nucleus and IKK I κ B ϵ -NF- κ B

参考文献:

- [1] Hood L. Systems biology: integrating technology, biology, and computation. *Mechanisms of Ageing and Development*, 2003, 124(1):9~16
- [2] Kitano H. Systems biology: A brief overview. *Science*, 2002, 295(5560):1662~1664
- [3] Kitano H. Computational systems biology. *Nature*, 2002, 420(6912):206~210
- [4] Bailey JE. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. *Biotechnol Prog*, 1998,14(1):8~20
- [5] Hoffmann A, Levchenko A, Scott ML, Baltimore D. The I κ B-NF- κ B signaling module: temporal control and selective gene activation. *Science*, 2002,298(5596):1241~1245
- [6] Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nature Biotechnology*, 2002,20(4):370~375
- [7] Curto R, Voit EO, Sorribas A, Cascante M. Mathematical models of purine metabolism in man. *Math Biosci*, 1998, 151(1):1~49
- [8] Heinrich R, Neel BG, Rapoport TA. Mathematical models of protein kinase signal transduction. *Mol Cell*, 2002,9 (5): 957~970
- [9] Cho KH, Shin SY, Kim HW, Wolkenhauer O, McFerran B, Kolch W. Mathematical modeling of the influence of RKIP on the ERK signaling pathway. *Lect Notes Comput Sc*, 2003, 2602:127~141
- [10] Bhalla US, Iyengar R. Emergent properties of networks of biological signaling pathways. *Science*, 1999,283 (5400): 381~387
- [11] Cho KH, Shin SY, Kolch W, Wolkenhauer O. Experimental design in systems biology, based on parameter sensitivity analysis using a Monte Carlo method: a case study for the TNF α -Mediated NF- κ B signal transduction pathway. *SIMULATION*, 2003,79(12):726~739
- [12] Moles CG, Mendes P, Banga JR. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res*, 2003,13(11):2467~2474
- [13] Voit EO, Marino S, Lall R. Challenges for the identification of biological systems from *in vivo* time series data. *In Silico Biol*, 2005,5(2):83~92
- [14] Julier SJ, Uhlmann JK. A new extension of the Kalman filter to nonlinear systems. In: Ivan K. Proceedings of aerosense: the 11th international symposium on aerospace/defense sensing, simul and controls. Orlando, FL: SPIE, 1997. 182~193
- [15] Wan EA, van der Merwe R. The unscented Kalman filter for nonlinear estimation. In: proceedings of symposium 2000 on adaptive systems for signal processing, communication and control. Lake Louise, Alberta, Canada: IEEE press, 2000. 1255~1260
- [16] van der Merwe R, Wan EA. The square-root unscented Kalman filter for state and parameter-estimation. In: International conference on acoustics, speech, and signal processing. SaltLake City, Utah: IEEE press, 2001. 3461~3464
- [17] Sitz AaS U, Kurths J, Voss HU. Estimation of parameters and unobserved components for nonlinear systems from noisy time series. *Phys Rev E*, 2002,66(1):016210
- [18] Romanenko A, Castro JAAM. The unscented filter as an alternative to the EKF for nonlinear state estimation: a simulation case study. *Computers & Chemical Engineering*, 2004, 28(3):347~355
- [19] Julier SJ. The scaled unscented transformation. In: Proceeding of the 2002 American control conference. Anchorage, AK, USA: IEEE press, 2002. 4555~4559
- [20] Julier SJ, Uhlmann JK. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 2004,92(3):401~422
- [21] Julier SJ, Uhlmann JK. Consistent debiased method for converting between polar and Cartesian coordinate systems. In: Proceedings of SPIE: acquisition, tracking, and pointing 11. Orlando, FL. 1997. 110~121
- [22] Julier SJ. Skewed approach to filtering. Proceedings of the Signal and Data Processing of Small Targets 1998; Orlando, FL, USA: SPIE; 1998
- [23] Wan EA, Merwe Rvd, Nelson AT. Dual estimation and the unscented transformation. In: Solla SA, Lean TK, Muller KR. Advances in neural information processing systems 12. MIT press, 2000. 666~672
- [24] Cheong R, Bergmann A, Werner SL, Regal J, Hoffmann A, Levchenko A. Transient I κ B kinase activity mediates temporal NF- κ B dynamics in response to a wide range of tumor necrosis factor- α doses. *J Biol Chem*, 2006, 281(5):2945~2950
- [25] Lipniacki T, Paszek P, Brasier AR, Luxon BA, Kimmel M. Stochastic regulation in early immune response. *Biophys J*, 2006,90(3):725~742
- [26] Schilling M, Maiwald T, Bohl S, Kollmann M, Kreutz C, Timmer J, Klingmuller U. Computational processing and error reduction strategies for standardized quantitative data in biological networks. *FEBS J*, 2005,272(24):6400~6411
- [27] Slikker WJ, Young JF, Corley RA, Dorman DC, Conolly RB, Knudsen TB, Erstad BL, Luecke RH, Faustman EM, Timchalk C, Mattison DR. Improving predictive modeling in pediatric drug development: pharmacokinetics, pharmacodynamics, and mechanistic modeling. *Ann NY Acad Sci*, 2005,1053 (1):505~518
- [28] Feng XJ, Rabitz H. Optimal identification of biochemical reaction networks. *Biophys J*, 2004,86(3):1270~1281
- [29] Nelson DE, Ihekwaba AE, Elliott M, Johnson JR, Gibney CA, Foreman BE, Nelson G, See V, Horton CA, Spiller DG, Edwards SW, McDowell HP, Linitt JF, Sullivan E, Grimley R, Benson N, Broomhead D, Kell DB, White MR. Oscillations in NF- κ B signaling control the dynamics of gene expression. *Science*, 2004,306(5696):704~708

- [30] Ihekwa AEC, Broomhead DS, Grimley RL, Benson N, Kell DB. Sensitivity analysis of parameters controlling oscillatory signalling in the NF- κ B pathway: the roles of IKK and I κ B. *Systems Biology, IEE Proceedings*, 2004, 1(1):93~103
- [31] Yue H, Brown M, Knowles J, Wang H, Broomhead SD, Kell DB. Insights into the behaviour of systems biology models from dynamic sensitivity and identifiability analysis: a case study of an NF- κ B signalling pathway. *Mol BioSyst*, 2006,2:640~649

ESTIMATE PARAMETERS AND UNOBSERVABLE STATES OF SIGNALLING PATHWAYS WITH UNSCENTED KALMAN FILTER

LIU Tai-yuan¹, JIA Jian-fang^{1,2}, WANG Hong¹, YUE Hong¹

(1. Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China;

2. Department of Automation, North University of China, Taiyuan 030051, China)

Abstract: One object of systems biology is to develop the mathematical models of biochemical networks in cell and to analysis the system dynamic properties based on these models as well as to predict the system output. However, strong nonlinearity, complexity, noisy and incomplete measurements make the parameter estimation more difficult. In this paper, Unscented Kalman filter was proposed to estimate the unknown parameters and the unobservable state variables simultaneously. TNF α mediated NF- κ B signal transduction pathway was taken as an example to illustrate the effectiveness of the method. Simulation results are encouraging and show that both parameters and unobservable state variables can be estimated well.

Key Words: Systems biology; Unscented Kalman filter; Parameter estimation; Signal transduction pathways

This work was supported by a grant from The Outstanding Overseas Chinese Scholars Fund of Chinese Academy of Sciences (2004-1-4)

Received: July 28, 2006

Corresponding author: LIU Tai-yuan, Tel: +86(10)82617643, E-mail: taiyuan.liu@ia.ac.cn

附录:

1 TNF α 诱导的 NF- κ B 信号转导数学模型^[5]

$$\begin{aligned} \frac{dy_1}{dt} &= -k_{02} \times y_1 - a_1 \times y_1 \times y_{10} + (d_1 + r_1) \times y_2 - a_2 \times y_1 \times y_{11} \\ &\quad + (d_2 + r_2) \times y_3 - a_3 \times y_1 \times y_{12} + (d_3 + r_3) \times y_4 - a_7 \times y_1 \times y_{19} \\ &\quad + (d_1 + r_1) \times y_5 - a_8 \times y_1 \times y_{20} + (d_2 + r_5) \times y_6 - a_9 \times y_1 \times y_{21} + (d_3 + r_6) \times y_7 \\ \frac{dy_2}{dt} &= a_1 \times y_1 \times y_{10} - (d_1 + r_1) \times y_2 - a_4 \times y_2 \times y_8 + d_4 \times y_5 \\ \frac{dy_3}{dt} &= a_2 \times y_1 \times y_{11} - (d_2 + r_2) \times y_3 - a_5 \times y_3 \times y_8 + d_5 \times y_6 \\ \frac{dy_4}{dt} &= a_3 \times y_1 \times y_{12} - (d_3 + r_3) \times y_4 - a_6 \times y_4 \times y_8 + d_6 \times y_7 \\ \frac{dy_5}{dt} &= a_7 \times y_1 \times y_{19} + a_4 \times y_2 \times y_8 - (d_1 + d_4 + r_4) \times y_5 \\ \frac{dy_6}{dt} &= a_8 \times y_1 \times y_{20} + a_5 \times y_3 \times y_8 - (d_2 + d_5 + r_5) \times y_6 \end{aligned}$$

$$\frac{dy_7}{dt} = a_9 \times y_1 \times y_{21} + a_6 \times y_4 \times y_8 - (d_3 + d_6 + r_6) \times y_7$$

$$\begin{aligned} \frac{dy_8}{dt} = & -a_4 \times y_{10} \times y_8 + d_4 \times y_{19} - a_5 \times y_{11} \times y_8 + d_5 \times y_{20} \\ & - a_6 \times y_{12} \times y_8 + d_6 \times y_{21} - a_4 \times y_2 \times y_8 + (r_4 + d_4) \times y_5 \\ & - a_5 \times y_3 \times y_8 + (r_5 + d_5) \times y_6 - a_6 \times y_4 \times y_8 + (r_6 + d_6) \times y_7 \\ & + deg_4 \times y_{19} + deg_4 \times y_{20} + deg_4 \times y_{21} - k_1 \times y_8 + k_{01} \times y_9 \end{aligned}$$

$$\frac{dy_9}{dt} = k_1 \times y_8 - a_4 \times y_{16} \times y_9 + d_4 \times y_{22} - a_5 \times y_{17} \times y_9 + d_5 \times y_{23} - a_6 \times y_{18} \times y_9 + d_6 \times y_{24} - k_{01} \times y_9$$

$$\frac{dy_{10}}{dt} = -a_1 \times y_1 \times y_{10} + d_1 \times y_2 - a_4 \times y_{10} \times y_8 + d_4 \times y_{19} + tr_1 \times y_{13} - deg_1 \times y_{10} - tp_1 \times y_{10} + tp_2 \times y_{16}$$

$$\frac{dy_{11}}{dt} = -a_2 \times y_1 \times y_{11} + d_2 \times y_3 - a_5 \times y_{11} \times y_8 + d_5 \times y_{20} + tr_1 \times y_{14} - deg_1 \times y_{11} - 0.5 \times tp_1 \times y_{11} + 0.5 \times tp_2 \times y_{17}$$

$$\frac{dy_{12}}{dt} = -a_3 \times y_1 \times y_{12} + d_3 \times y_4 - a_6 \times y_{12} \times y_8 + d_6 \times y_{21} + tr_1 \times y_{15} - deg_1 \times y_{12} - 0.5 \times tp_1 \times y_{12} + 0.5 \times tp_2 \times y_{18}$$

$$\frac{dy_{13}}{dt} = tr_{2a} + tr_2 \times y_9^2 - tr_3 \times y_{13}$$

$$\frac{dy_{14}}{dt} = tr_{2b} - tr_3 \times y_{14}$$

$$\frac{dy_{15}}{dt} = tr_{2c} - tr_3 \times y_{15}$$

$$\frac{dy_{16}}{dt} = tp_1 \times y_{10} - tp_2 \times y_{16} - a_4 \times y_{16} \times y_9 + d_4 \times y_{22}$$

$$\frac{dy_{17}}{dt} = 0.5 \times tp_1 \times y_{11} - 0.5 \times tp_2 \times y_{17} - a_5 \times y_{17} \times y_9 + d_5 \times y_{23}$$

$$\frac{dy_{18}}{dt} = 0.5 \times tp_1 \times y_{12} - 0.5 \times tp_2 \times y_{18} - a_6 \times y_{18} \times y_9 + d_6 \times y_{24}$$

$$\frac{dy_{19}}{dt} = a_4 \times y_{10} \times y_8 - d_4 \times y_{19} - a_7 \times y_1 \times y_{19} + d_1 \times y_5 + k_2 \times y_{22} - deg_4 \times y_{19}$$

$$\frac{dy_{20}}{dt} = a_5 \times y_{11} \times y_8 - d_5 \times y_{20} - a_8 \times y_1 \times y_{20} + d_2 \times y_6 + 0.5 \times k_2 \times y_{23} - deg_4 \times y_{20}$$

$$\frac{dy_{21}}{dt} = a_6 \times y_{12} \times y_8 - d_6 \times y_{21} - a_9 \times y_1 \times y_{21} + d_3 \times y_7 + 0.5 \times k_2 \times y_{24} - deg_4 \times y_{21}$$

$$\frac{dy_{22}}{dt} = a_4 \times y_{17} \times y_9 - d_4 \times y_{22} - k_2 \times y_{22}$$

$$\frac{dy_{23}}{dt} = a_5 \times y_{17} \times y_9 - d_5 \times y_{23} - 0.5 \times k_2 \times y_{23}$$

$$\frac{dy_{24}}{dt} = a_6 \times y_{18} \times y_9 - d_6 \times y_{24} - 0.5 \times k_2 \times y_{24}$$

其中 y_i ($i=1, \dots, 24$)为信号转导过程中的中间产物, 各变量所代表的物质见表1。参数真值见表2。

Table 1 Representation of state variables of TNF α mediated NF- κ B signal transduction pathway

| State variable | Component | State variable | Component | State variable | Component | State variable | Component |
|----------------|---|----------------|--------------------------------|----------------|---------------------------|----------------|--|
| y1 | IKK | y7 | IKKI κ BcNF- κ B | y13 | I κ B α t | y19 | I κ B α NF- κ B |
| y2 | IKKI κ B α | y8 | NF- κ B | y14 | I κ B β t | y20 | I κ B β NF- κ B |
| y3 | IKKI κ B β | y9 | NF- κ Bn | y15 | I κ B ϵ t | y21 | I κ B ϵ NF- κ B |
| y4 | IKKI κ B ϵ | y10 | I κ B α | y16 | I κ B α n | y22 | I κ B α nNF- κ Bn |
| y5 | IKKI κ B α NF- κ B | y11 | I κ B β | y17 | I κ B β n | y23 | I κ B β nNF- κ Bn |
| y6 | IKKI κ B β NF- κ B | y12 | I κ B ϵ | y18 | I κ B ϵ n | y24 | I κ B ϵ nNF- κ Bn |

The subscript "n" means the protein lies in the nucleus and "t" means the mRNA

Table 2 Summary of the parameter values of TNF α mediated NF- κ B signal transduction pathway

| Reaction coefficients | Symbol | Values | Units |
|---|--------|-----------------------|---------------------------------|
| IKK-I κ B α association | a1 | 22.5×10^{-3} | $\mu\text{M}^{-1}\text{s}^{-1}$ |
| IKK-I κ B β association | a2 | 6.0×10^{-3} | $\mu\text{M}^{-1}\text{s}^{-1}$ |
| IKK-I κ B ϵ association | a3 | 9.0×10^{-3} | $\mu\text{M}^{-1}\text{s}^{-1}$ |
| I κ B α -NF- κ B association | a4 | 0.5×100 | $\mu\text{M}^{-1}\text{s}^{-1}$ |
| I κ B β -NF- κ B association | a5 | 0.5×100 | $\mu\text{M}^{-1}\text{s}^{-1}$ |
| I κ B ϵ -NF- κ B association | a6 | 0.5×100 | $\mu\text{M}^{-1}\text{s}^{-1}$ |
| IKK-I κ B α NF- κ B association | a7 | 1.85×10^{-1} | $\mu\text{M}^{-1}\text{s}^{-1}$ |
| IKK-I κ B β NF- κ B association | a8 | 4.8×10^{-2} | $\mu\text{M}^{-1}\text{s}^{-1}$ |
| IKK-I κ B ϵ NF- κ B association | a9 | 7.0×10^{-2} | $\mu\text{M}^{-1}\text{s}^{-1}$ |
| IKK-I κ B α NF- κ B dissociation | d1 | 1.25×10^{-3} | s^{-1} |
| IKK-I κ B β NF- κ B dissociation | d2 | 1.75×10^{-3} | s^{-1} |
| IKK-I κ B ϵ NF- κ B dissociation | d3 | 1.75×10^{-3} | s^{-1} |
| IKKI κ B α -NF- κ B dissociation | d4 | 0.5×10^{-3} | s^{-1} |
| IKKI κ B β -NF- κ B dissociation | d5 | 0.5×10^{-3} | s^{-1} |
| IKKI κ B ϵ -NF- κ B dissociation | d6 | 0.5×10^{-3} | s^{-1} |
| IKK-I κ B α catalysis | r1 | 4.07×10^{-3} | s^{-1} |
| IKK-I κ B β catalysis | r2 | 1.5×10^{-3} | s^{-1} |
| IKK-I κ B ϵ catalysis | r3 | 2.2×10^{-3} | s^{-1} |
| IKKI κ B α NF- κ B catalysis | r4 | 2.04×10^{-2} | s^{-1} |
| IKKI κ B β NF- κ B catalysis | r5 | 7.5×10^{-3} | s^{-1} |
| IKKI κ B ϵ NF- κ B catalysis | r6 | 1.1×10^{-2} | s^{-1} |
| I κ B α inducible mRNA synthesis | tr2 | 1.65×10^{-2} | μMs^{-1} |
| I κ B α constitutive mRNA synthesis | tr2a | 1.54×10^{-6} | μMs^{-1} |
| I κ B β constitutive mRNA synthesis | tr2b | 1.78×10^{-7} | μMs^{-1} |
| I κ B ϵ constitutive mRNA synthesis | tr2e | 1.27×10^{-7} | μMs^{-1} |
| I κ B mRNA degradatioin | tr3 | 2.8×10^{-4} | s^{-1} |
| constitutive I κ B translation rate | tr1 | 4.08×10^{-3} | s^{-1} |
| constitutive I κ B degradation(free) | deg1 | 1.13×10^{-4} | s^{-1} |
| constitutive I κ B degradation (complexed to NF- κ B) | deg4 | 2.25×10^{-5} | s^{-1} |
| I κ B α nuclear import | tp1 | 3×10^{-4} | s^{-1} |
| I κ B α nuclear export | tp2 | 2×10^{-4} | s^{-1} |
| I κ B β nuclear import | 0.5tp1 | 1.5×10^{-4} | s^{-1} |
| I κ B β nuclear export | 0.5tp2 | 1×10^{-4} | s^{-1} |
| I κ B ϵ nuclear import | 0.5tp1 | 1.5×10^{-4} | s^{-1} |
| I κ B ϵ nuclear export | 0.5tp2 | 1×10^{-4} | s^{-1} |
| NF- κ B nuclear import | k1 | 0.9×10^{-1} | s^{-1} |
| NF- κ B nuclear export | k01 | 0.8×10^{-4} | s^{-1} |
| I κ B α -NF- κ B nuclear export | k2 | 1.38×10^{-2} | s^{-1} |
| I κ B β -NF- κ B nuclear export | 0.5k2 | 5.2×10^{-3} | s^{-1} |
| I κ B ϵ -NF- κ B nuclear export | 0.5k2 | 5.2×10^{-3} | s^{-1} |
| IKK slow adaption | k02 | 1.2×10^{-4} | s^{-1} |

2 NF- κ B 信号转导通路模型状态稳态和动态响应的获取

在表 3 所给出的状态初值下, 采用 Matlab 中 ODE23s 函数求取模型的数值解。仿真运行 2000 分钟后, 系统达到稳态。保存各状态此时的稳态值作为下次仿真新的初始值, 对 IKK 施加幅值为 $0.1 \mu\text{mol/L}$ 的扰动, 作为外来刺激引起的 IKK 浓度的变化, 同样采用 Matlab 中 ODE23s 函数求取模型数值解, 仿真运行 400 分钟, 得到的状态轨迹即视为系统在受到 $\text{TNF}\alpha$ 刺激时, NF- κ B 信号转导通路模型各状态的动态变化过程。

Table 3 Summary of initial values for 24 state variables in simulation

| State variable | Initial values(μM) | State variable | Initial values(μM) | State variable | Initial values (μM) | State variable | Initial values (μM) |
|----------------|---------------------------------|----------------|---------------------------------|----------------|----------------------------------|----------------|----------------------------------|
| y1 | 0 | y7 | 0 | y13 | 0 | y19 | 0 |
| y2 | 0 | y8 | 0.1 | y14 | 0 | y20 | 0 |
| y3 | 0 | y9 | 0 | y15 | 0 | y21 | 0 |
| y4 | 0 | y10 | 0 | y16 | 0 | y22 | 0 |
| y5 | 0 | y11 | 0 | y17 | 0 | y23 | 0 |
| y6 | 0 | y12 | 0 | y18 | 0 | y24 | 0 |