Adjoint Systems for Models of Cell Signaling Pathways and Their Application to Parameter Fitting

Krzysztof Fujarewicz, Marek Kimmel, Tomasz Lipniacki, and Andrzej Świerniak

Abstract—The paper concerns the problem of fitting mathematical models of cell signaling pathways. Such models frequently take the form of sets of nonlinear ordinary differential equations. While the model is continuous in time, the performance index used in the fitting procedure involves measurements taken at discrete time moments. Adjoint sensitivity analysis is a tool which can be used for finding the gradient of a performance index in the space of parameters of the model. In the paper, a structural formulation of adjoint sensitivity analysis called the Generalized Backpropagation Through Time (GBPTT) is used. The method is especially suited for hybrid, continuous-discrete time systems. As an example, we use the mathematical model of the NF-κB regulatory module, which plays a major role in the innate immune response in animals.

Index Terms—Biology and genetics, modeling, ordinary differential equations, parameter learning.

1 Introduction

Cell signaling pathways are cascades and feedback loops of biochemical reactions which make possible transduction of signals within and between biological cells. There are many approaches to modeling interactions in cell and cell population on the molecular level [6], [13], [22]. They differ in precision and generality. The most accurate approach to modeling of the dynamics of cell signaling pathways takes into account the stochasticity that is important on the molecular level. Ordinary differential equations (ODEs) provide a deterministic description of cell signaling pathways [14], [17], [25]. Assuming a high rate of spatial diffusion processes involved, the description in the terms of ODEs is valid for a single cell. It is also frequently used to describe the averaged dynamics of cell population [1]. The system of ODEs involves concentrations of protein, protein complexes, and messenger RNAs (mRNAs) as time variables.

To compare different models and to test their ability to model processes for which experimental data are available, an efficient method of parameter estimation is needed. Unfortunately, while the model is continuous in time, all available measurement techniques, such as Western blot expression analysis, electrophoretic mobility shift assays, or gene expression microarrays, provide measurements at discrete time moments only. Moreover, these time moments frequently are nonuniformly distributed and may be different for different signals measured.

Estimation of parameters of continuous-time models based on discrete-time measurements may be accomplished in various ways. The simplest approach involves estimation of time derivatives of the model's variables at the measurements' times. The system of ODEs is then reduced to a set of nonlinear algebraic conditions, which are best approximated using regression. Such an approach may be used when the observational noise is small and the measurements give enough information about the dynamic variables. Unfortunately, in the case of the problem addressed in this paper, the observed measurements are very noisy and, moreover, not all variables are directly observed.

The *initial value approach* [20] may be considered more appropriate because it takes into account the dynamic nature of the problem. The approach involves minimization of a quadratic performance index involving differences between the measurements and the model's outputs with respect to the model's parameters and the initial values of the variables. Under the assumption that the measurement errors have a Gaussian distribution, this approach is equivalent to the maximum likelihood estimation. Recently, the initial value approach has been successfully used for identification of the JAK-STAT signaling pathway [24]. The initial value approach is an iterative optimization procedure where the information about the gradient of the objective function may be calculated via the sensitivity analysis. In this case, several sensitivity models (or one model simulated several times) have to be used in order to find all of the components of the gradient.

In the paper [24], still another method, the *multiple shooting* method proposed in [27] and developed by Bock [2], [3], was suggested for systems where local minima or chaotic behavior occurs.

- *K. Fujarewicz and A. Świerniak are with the Institute of Automatic Control, Silesian University of Technology, Akademicka 16, 44-101 Gliwice, Poland. E-mail: {krzysztof.fujarewicz, andrzej.swierniak}@polsl.pl.*
- *M. Kimmel is with the Institute of Automatic Control, Silesian University of Technology, Akademicka 16, 44-101 Gliwice, Poland, and the Department of Statistics, Rice University, PO Box 1892, Houston, TX 77251. E-mail: kimmel@stat.rice.edu.*
- *T. Lipniacki is with the Department of Statistics, Rice University, PO Box 1892, Houston, TX 77251 and the Institute of Fundamental Technological Research, Swietokrzyska 21, 00-049 Warsaw, Poland. E-mail: tlipnia@ippt.gov.pl.*

Manuscript received 28 Apr. 2005; revised 16 Jan. 2006; accepted 23 Aug. 2006; published online 10 Jan. 2007.
For information on obtaining reprints of this article, please send e-mail to: tcbb@computer.org, and reference IEEECS Log Number TCBB-0034-0405.

1545-5963/07/$25.00 © 2007 IEEE Published by the IEEE CS, CI, and EMB Societies & the ACM

Adjoint sensitivity analysis is a technique frequently used in practical optimization problems such as identification or optimal control design. It is based on efficient calculation of the gradient of a performance index and reduces the computational cost compared to straight (tangent linearized) sensitivity analysis because it gives the whole gradient in a single computation run. Adjoint systems are defined in the literature as either continuous or discrete in time and cannot be directly applied to solve the problem of parameter fitting presented above because of its hybrid, continuous-discrete nature.

Recently, we have developed a formulation of adjoint sensitivity analysis, the Generalized Backpropagation Through Time (GBPTT), which is suitable for solving hybrid problems. Our method, first published in [8], is an extension of Backpropagation Through Time (BPTT) [29], known in neural network theory. BPTT is an offline method used for learning in recurrent dynamic neural networks. It involves the "unfolding" in time of the currently trained recurrent neural network. As a result, a static neural network is obtained and the standard Backpropagation (BP) algorithm can be used. Such an approach is easy to apply to recurrent neural networks with simple structure. In the case of more complicated discrete-time dynamic systems, the subset of which is neural networks, unfolding in time is not easy and, in the case of continuous-time systems, it is not possible at all. The GBPTT method does not have these disadvantages. It does not require unfolding in time and it can be used for both discrete and continuous-time dynamic systems. In [10], the preliminary results of the application of adjoint sensitivity analysis to signaling pathways based on model-based "synthetic data" were presented.

The data employed in the present paper are biological measurements obtained using blotting techniques. Blot images (one image per time point per biochemical species) can be considered gray-scale densities which need to be integrated to produce concentrations of respective biochemical species. Because of the nonlinearity of the gray scale and other factors, conversion to concentrations is inaccurate. Overcoming this difficulty, e.g., by careful calibration, is outside the scope of this paper. Also, in most cases, concentration levels may be compared within a single blot series, but cannot be compared to concentrations estimated based on other blots. This latter difficulty we successfully address in this paper by assuming the existence of unknown multipliers (one multiplier per blot series) and estimating them within the framework of the GBPTT methodology.

The proposed approach is employed to fitting the model of the signaling pathway of NF-κB recently proposed in [17] to the data published in [14] and [16].

The paper is organized as follows: In Section 2, the mathematical model of the NF-κB regulatory module, an important example of a model of cell signaling pathway, is briefly presented. Then, in Section 3, we describe the data that, after quantification, have been used to fit the model. In Section 4, we formally state the problem of fitting of the continuous-time model to discrete-time measurements and the problem of finding the gradient of the performance index. Section 5 describes the GBPTT method, which is used in Section 6 to solve the problem of finding the gradient of the performance index. Sections 7 and 8 present numerical results of fitting parameters of the NF-κB regulatory module and results of model validation.

2 NF-κB Regulatory Module

Transcription factor NF-κB promotes expression of about 100 genes that play important roles in within and between-cell signaling. It is involved in cellular stress responses, cell growth, survival, and apoptosis. The model presented in this section was proposed by Lipniacki et al. [17]. Readers interested in biological and mathematical details are referred to this paper and [14], [16].

The mathematical model of the NF-κB regulatory module includes 15 first-order nonlinear differential equations:

$$\dot{x}_1 = k_{prod} - k_{deg}x_1 - k_1 u x_1, \tag{1}$$

$$\dot{x}_2 = k_1 u x_1 - k_3 x_2 - k_2 u x_2 x_8 - k_{deg}x_2 - a_2 x_2 x_{10} + t_1 x_4 - a_3 x_2 x_{13} + t_2 x_5, \tag{2}$$

$$\dot{x}_3 = k_3 x_3 + k_2 u x_2 x_8, \tag{3}$$

$$\dot{x}_4 = a_2 x_2 x_{10} - t_1 x_4, \tag{4}$$

$$\dot{x}_5 = a_3 x_2 x_{13} - t_2 x_5, \tag{5}$$

$$\dot{x}_6 = c_{6a} x_{13} - a_1 x_6 x_{10} + t_2 x_5 - i_1 x_6, \tag{6}$$

$$\dot{x}_7 = i_1 k_v x_6 - a_1 x_7 x_{11}, \tag{7}$$

$$\dot{x}_8 = c_4 x_9 - c_5 x_8, \tag{8}$$

$$\dot{x}_9 = c_2 + c_1 x_7 - c_3 x_9, \tag{9}$$

$$\dot{x}_{10} = -a_2 x_2 x_{10} - a_1 x_6 x_{10} + c_{4a} x_{12} - c_{5a} x_{10} - i_{1a} x_{10} + e_{1a} x_{11}, \tag{10}$$

$$\dot{x}_{11} = -a_1 x_7 x_{11} + i_{1a} k_v x_{10} - e_{1a} k_v x_{11}, \tag{11}$$

$$\dot{x}_{12} = c_{2a} + c_{1a} x_7 - c_{3a} x_{12}, \tag{12}$$

$$\dot{x}_{13} = a_1 x_6 x_{10} - c_{6a} x_{13} - a_3 x_2 x_{13} + e_{2a} x_{14}, \tag{13}$$

$$\dot{x}_{14} = a_1 x_7 x_{11} - e_{2a} k_v x_{14}, \tag{14}$$

$$\dot{x}_{15} = c_{2c} + c_{1c} x_7 - c_{3c} x_{15}. \tag{15}$$

In the model, state variables x_i, $i = 1, 2, \ldots, 15$, are concentrations of proteins, complexes of proteins, or their transcripts:

- x_1—IKK kinase in the neutral state,
- x_2—IKK kinase in the active state,
- x_3—IKK kinase in the inactive state,
- x_4—complexes of proteins (IKKa|IκBα),

- x_5—complexes of proteins (IKKa|IκBα|NFκB),
- x_6—protein NFκB,
- x_7—protein NFκB in the nucleus,
- x_8—protein A20,
- x_9—protein A20 transcript,
- x_{10}—free IκBα protein,
- x_{11}—free nuclear IκBα protein,
- x_{12}—IκBα transcript,
- x_{13}—cytoplasmic complexes of proteins (IκBα|NFκB),
- x_{14}—nuclear complexes of proteins (IκBα|NFκB), and
- x_{15}—control gene transcript.

The input signal u is a Boolean variable 1 or 0 and is equal to 1 when the signaling pathway is stimulated by an extracellular signal, tumor necrosis factor (TNF), or interleukin-1 (IL-1). State variables and the input signal are time-dependent, which is not indicated in the model to simplify the notation. The remaining quantities in the model (1)-(15) are parameters. In [17], a part of these parameters have been assumed as known, whereas 10 parameters have been heuristically fitted based on the data from [14] and [16].

3 Data

Concentrations of different proteins and their complexes can be measured using different techniques, such as blotting techniques (Western Blot, Northern Blot, etc.) or electrophoretic mobility shift assay (EMSA). These techniques are different, but, in most cases, the outcome of the measuring process is images. An example of such an image is presented in Fig. 1a. This picture was presented in reference [14] and represents concentrations of nuclear NF-κB at 20 different time moments before and during persistent TNF stimulation. This particular data has been assayed using EMSA technique. In most cases, times of observations are not uniformly distributed over the experiment period. Frequently, they are denser at the beginning of the experiment. In this particular experiment, measurement times were chosen at: 0, 2, 5, 10, 15, 30, and 45 minutes and 1, 1.25, 1.5, 1.75, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, and 6 hours. Moreover, different signals, such as molar concentrations of proteins or mRNA levels, can be measured at different time moments.

Since all measuring techniques used to generate our data are only semiquantitative, it means that, without deriving a standardizing curve, one can compare protein or mRNA levels only within one blot. Hence, quantification may be performed only for different blots separately. The result of such quantification, performed on the blot in Fig. 1a, is presented in Fig. 1b. Because the scale is unknown, the values are normalized so that the maximum value is equal to 1.

In this paper, we use data, images from blots, and other assays from two experiments published in reference [14] and two experiments published in reference [16]. In all four cases, experiments have been performed on mouse fibroblasts. The first two experiments, from [14], have been performed for two different stimulations with TNF, persistent, and one-hour pulse. The other two experiments, from [16], have been performed for wild type cells and for A20 deficient cells, both under persistent IL-1 stimulation.

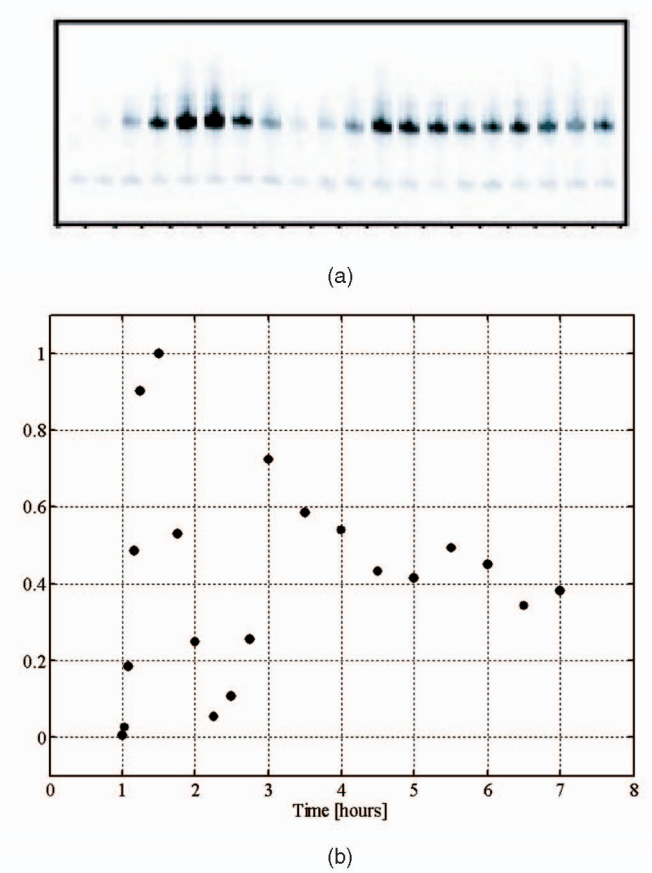


Fig. 1. Concentrations of nuclear NF-κB before and during persistent TNF stimulation: (a) the image from EMSA [14] and (b) the result of quantification based on the image. TNF stimulation starts at 1 hour.

These images are presented in Section 7 (Figs. 7 and 8), where they are compared to signals obtained using our fitting procedure.

4 Problem Formulation

Let us assume that I variables are measured at discrete-time moments. The variable numbered i is measured at N_i time moments, $t_{i,1}, t_{i,2}, \ldots, t_{i,N_i}$. Γ_i is the set of these time moments,

$$\Gamma_i = \{t_{i,1}, t_{i,2}, \ldots, t_{i,N_i}\}. \tag{16}$$

The ith measured signal takes, at time $t_{i,n}$, the real value $y_{exp,i}(t_{i,n})$. Let us denote by $v_{exp,i}(t_{i,n})$ the result of quantification (based on a single spot in the blot image) of this value. This result is a function of $y_{exp,i}(t_{i,n})$:

$$v_{exp,i}(t_{i,n}) = v_{exp,i}\big(y_{exp,i}(t_{i,n})\big). \tag{17}$$

In general, this function is nonlinear, involving, for example, saturation, dead zone, nonlinearities of blotting analysis, etc. However, in this paper, we assume that these nonlinearities may be neglected and the function (17) is linear,

$$v_{exp,i}(t_{i,n}) = w_{exp,i} \cdot y_{exp,i}(t_{i,n}), \tag{18}$$

where $w_{exp,i}$ is a multiplier to be estimated.

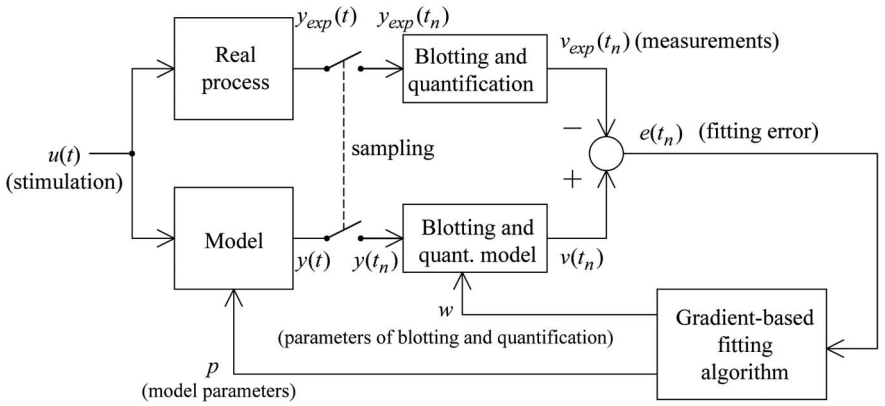


Fig. 2. The data flowchart and the proposed gradient-descent approach to parameter fitting.

We have $\sum_{i=1}^{I} N_i$ scalar measurements.

The output of the model includes I continuous-time signals, $y_i(t)$, $i = 1, 2, \ldots, I$. Quantities

$$v_i(t_{i,n}) = w_i \cdot y_i(t_{i,n}) \tag{19}$$

are compared to measurements (17). Multiplier w_i is an estimate of $w_{exp,i}$.

Now, we introduce the quadratic performance index to be minimized,

$$J = \frac{1}{2} \sum_{i=1}^{I} \sum_{n=1}^{N_i} q_i \, e_i^2(t_{i,n}), \tag{20}$$

where $e_i(t_{i,n})$ is the error of fitting,

$$e_i(t_{i,n}) = v_i(t_{i,n}) - v_{exp,i}(t_{i,n}), \tag{21}$$

and q_i, $i = 1, 2, \ldots, I$, are positive constants to be specified.

Let Γ denote the set of all measurement times,

$$\Gamma = \bigcup_{i=1}^{I} \Gamma_i = \{t_1, t_2, \ldots, t_N\}. \tag{22}$$

In practice, the cardinality of this set is less than the sum of the cardinalities of sets Γ_i, $i = 1, 2, \ldots, I$, because some of these times coincide for different measurements.

The performance index (20) may be expressed in the vector form

$$J = \frac{1}{2} \sum_{n=1}^{N} e^T(t_n) Q e(t_n), \tag{23}$$

where Q is a diagonal matrix $Q = diag(q_1, q_2, \ldots, q_I)$ and vector $e(t_n)$ is defined as follows:

$$e_i(t_n) = \begin{cases} e_i(t_{i,n}) & \text{if} \quad t_n \in \Gamma_i \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

Let w be a vector of unknown multipliers appearing in (19) and let p be the vector of all parameters of the model (1)-(15) that are to be fitted. The problem which is solved in this paper may be formally stated as follows:

Problem 1. *Find optimal vectors p and w minimizing the performance index (23) under constraints $w \geq 0$, $p \geq 0$.*

To solve the problem using a gradient descent approach, it is necessary to find the gradients of the performance index. Hence, the following problem has to be solved:

Problem 2. *Find gradients of the performance index (23),*

$$\nabla_p J, \ \nabla_w J. \tag{25}$$

The approach presented in this paper is schematically depicted in Fig. 2. We are about to formulate a gradient-based fitting algorithm, the primary objective of which is to solve Problem 2.

Fitting is complicated for several reasons. First, the problem has a dual nature—the model and the modeled system are continuous-time, whereas the performance index is discrete-time and takes the form of the sum (23). Second, times of measurements may be nonuniformly distributed and may be different for different outputs. Third, because it is difficult to compare different blot-series, it is necessary to find additional multipliers. In addition, data were generated from several experiments performed in the presence of different stimulations in different laboratories. In the next section, we will present the methodology which allows approaching the problem stated above.

5 Method: Generalized Backpropagation through Time for Continuous-Discrete Systems

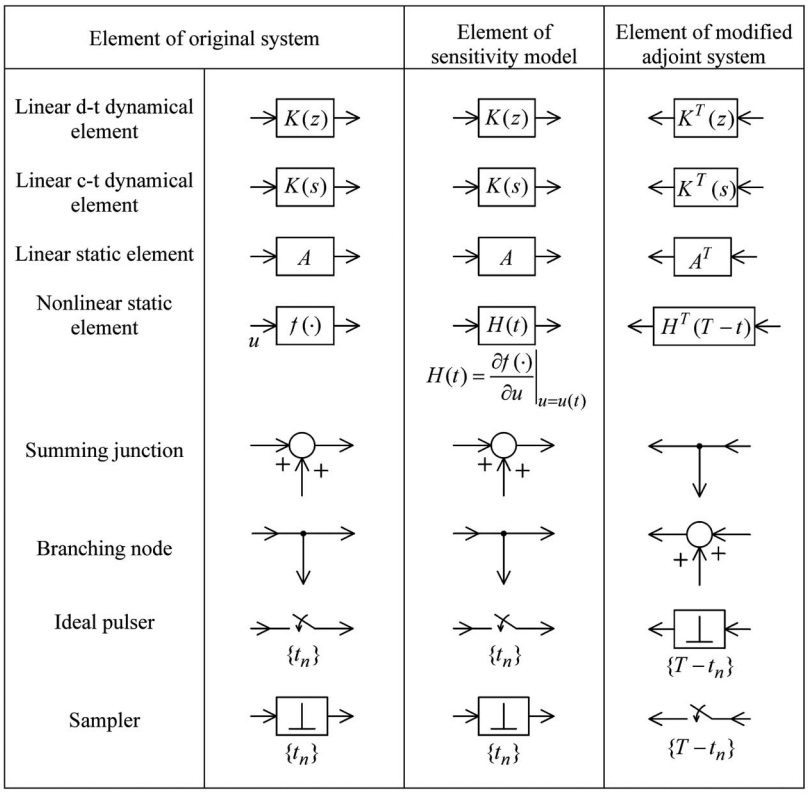
At the beginning of this section, the basic idea of GBPTT will be explained using discrete-time nonlinear dynamic systems as an introductory example. Then, it will be extended to continuous-discrete systems.

Let us consider a discrete-time system

$$\begin{cases} x(t_{n+1}) = f(x(t_n), u(t_n)) \\ y(t_n) = g(x(t_n), u(t_n)); \end{cases} \qquad n = 0, 1, \ldots, N, \tag{26}$$

where x, u, and y are vectors of the state, input, and output of the system, respectively. Vector functions f and g are differentiable and generally nonlinear. System (26) updates its state and output at time moments $t_0, t_1, \ldots, t_N$. We do not assume that times $t_0, t_1, \ldots, t_N$ are uniformly distributed.

TABLE 1
Rules for Construction of the Sensitivity Model and of the Modified Adjoint System



Let us introduce the *input-output sensitivity function*:

$$S_{u_j(t_m)}^{y_i(t_n)} \stackrel{\text{def}}{=} \frac{\partial y_i(t_n)}{\partial u_j(t_m)}. \tag{27}$$

Such a function describes the sensitivity of the ith component of the output signal y at time t_n with respect to the jth component of the input signal u at a previous time t_m. The input-output sensitivity function defined for vectors u and y is the Jacobi matrix

$$S_{u(t_m)}^{y(t_n)} \stackrel{\text{df}}{=} \begin{bmatrix} \frac{\partial y_1(t_n)}{\partial u_1(t_m)} & \cdots & \frac{\partial y_1(t_n)}{\partial u_r(t_m)} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m(t_n)}{\partial u_1(t_m)} & \cdots & \frac{\partial y_m(t_n)}{\partial u_r(t_m)} \end{bmatrix}. \tag{28}$$

The input-output sensitivity function has a universal character. One can show that a range of identification and optimal control problems for nonlinear dynamical systems can be reduced to the problem of finding such a function. Indeed, in practical problems, the performance index can be expressed as a value of an additional output signal of the system at a specified time. On the other hand, constant parameters to be optimized can be viewed as values, at a specified time, of additional input signals of the system. For example, in [8], the problems of optimal control in open and closed loop structures have been reduced to the problem of finding input-output sensitivity function. Moreover, the input-output sensitivity function allows finding the gradient of the performance index in the space of signals (control signals or time varying parameters). Once the value or values of the input-output sensitivity function are found, any gradient descent method can be utilized to optimize the performance index.

The GBPTT method assumes that the system analyzed is given in a graphical form. Any system (26) can be represented as a flowchart composed of elements listed in the first column of Table 1, i.e., the linear discrete-time dynamical element, the linear static element, the nonlinear static element, the summing junction, and the branching node. Other elements in the first column of Table 1 will be discussed later. Construction of the sensitivity model and of the so-called modified adjoint system [15] is mnemonic and depends on replacing all elements by corresponding elements depicted in the second and the third columns, respectively. In addition, in the modified adjoint system, the directions of all signals are reversed. The modified adjoint system differs from the "standard" adjoint system in that the time is reversed, which is suitable for simulation because the obtained system is causal. From now on, we will omit the word "modified."

In order to obtain the value of the input-output sensitivity function (27), the jth input of the sensitivity model has to be stimulated by a Kronecker pulse at time moment t_m. Then, the required value of the input-output sensitivity function (27) is equal to the ith output signal at time moment t_n. For many parameters to be optimized, the sensitivity model has to be solved (simulated) many times or many sensitivity models have to be simulated. To decrease the computational effort, the adjoint system is used. Now, due to the reversal of signal directions and the

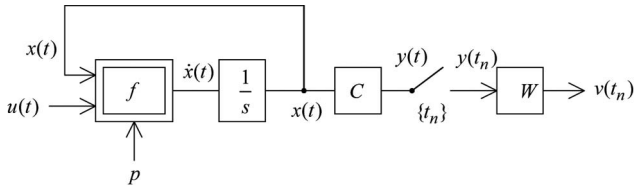


Fig. 3. The model with output channel generating a discrete time signal compared to measurements.

reversal of time, the ith input signal is stimulated at time $T - t_n$ by a Kronecker pulse. Then, the input-output sensitivity function (27) will be equal to the jth output signal at time moment $T - t_m$.

Similarly, for continuous-time systems, the input-output sensitivity function may be defined and rules for creation of the sensitivity model and the adjoint system with instructions how to stimulate them can be specified.

The approach presented above has been extended to hybrid, continuous-discrete systems [9]. In this paper, an approach has been proposed and applied to continuous-time neural networks based on discrete time measurements.

The continuous-discrete time system can be presented as a flowchart composed of the elements of the entire first column of Table 1. There are three additional elements not mentioned above, the continuous-time linear dynamical element, the ideal pulser, and the sampler. The ideal pulser connects an output of a discrete-time subsystem with an input of a continuous-time subsystem. The sampler does the same, but in the opposite direction. In the adjoint system (see the third column of Table 1), all pulsers are to be replaced by samplers and all samplers by pulsers. The rules for generating the desired input-output sensitivity model remain the same as in the case of discrete-time systems.

The formal derivation of the GBPTT method for continuous-discrete time systems was presented in [11]. Moreover, the sketch of the proof of the method correctness can be found as supplemental material which can be found on the Computer Society Digital Library at http://computer.org/tcbb/archives.htm.

6 Solution of the Problem

The model (1)-(15) with unknown multipliers describing the quantification of the data can be presented in the form of a flowchart shown in Fig. 3.

Vector function f represents the right sides of (1)-(15). Vectors $x(t)$ and $\dot{x}(t)$ consist of state variables $x_1, \ldots, x_{15}$ and their derivatives, respectively. Input signal $u(t)$ is one-dimensional and it represents stimulation by TNF or IL-1. Vector p consists of all 10 fitted parameters

$$p = [\, k_1 \;\; k_2 \;\; k_3 \;\; i_1 \;\; e_{2a} \;\; i_{1a} \;\; k_{prod} \;\; k_{deg} \;\; c_5 \;\; c_{4a} \,]^T. \tag{29}$$

Matrix W is diagonal $W = diag(w_1, w_2, \ldots, w_I)$ and the problem of finding the gradient $\nabla_w J$ is reduced to the problem of finding the gradient of the performance index with respect to the matrix W: $\nabla_W J$. Matrix C describes how the observed I outputs depend on state variables, so the dimension of C is $I \times 15$. In most cases, output variables are simply a subset of state variables.

Now, in order to calculate gradients (25), we replace the original problem by the problem of finding input-output sensitivity functions. To accomplish this, let us extend the flowchart from Fig. 3. The result is presented in Fig. 4.

The vector of parameters p is now a constant "signal" generated by channeling an additional discrete-time signal $\tilde{p}(t_n) = p \cdot \delta_k(t)$ through an ideal pulser and an integrator. Symbol $\delta_k(t)$ represents the Kronecker pulse at $t = 0$ and, so, the signal $\tilde{p}(t_n)$ is nonzero only at this time moment. The same is true for the ideal pulser, which generates, at the same time, a single Dirac pulse. Constant matrix W is generated in a similar way. The discrete-time part of the system functions at time moments $t_1, t_2, \ldots, t_N$. Additional signal $\tilde{W}(t_n) = W \cdot \delta_k(t - t_1)$ is different from 0 only at time moment t_1. This signal passes through a summation operator represented in the flowchart by the discrete transfer function $z/(z-1)$.

The output signal $\tilde{J}(t_n)$ is generated in such a way that, at the final discrete time moment, it is equal to the performance index,

$$\tilde{J}(t_N) = J. \tag{30}$$

Because of these changes, the problem of finding gradients (25) is now equivalent to the problem of finding the following input-output sensitivity functions:

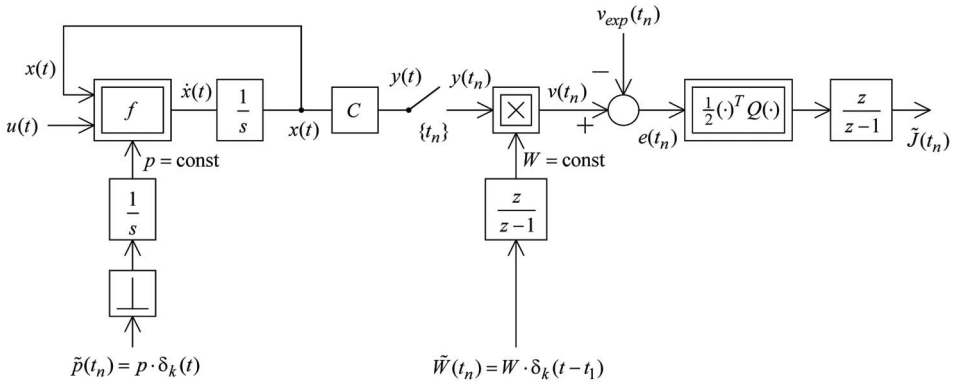


Fig. 4. Extended flowchart of the system.

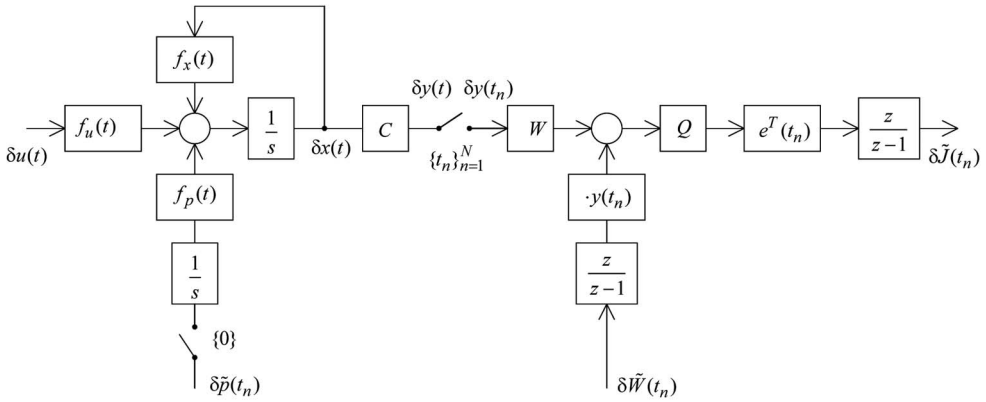


Fig. 5. Sensitivity model.

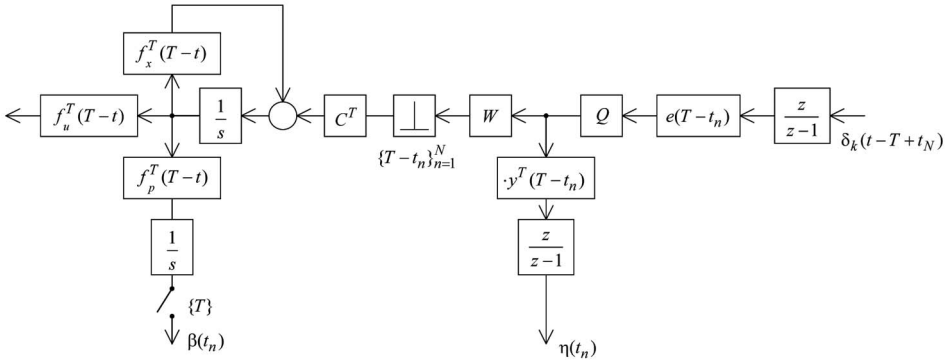


Fig. 6. Modified adjoint system.

$$S_{\tilde{p}(0)}^{\tilde{J}(t_N)} = [\nabla_p J]^T, \quad S_{\tilde{W}(t_1)}^{\tilde{J}(t_N)} = [\nabla_W J]^T. \tag{31}$$

Applying the rules of Table 1 to the flowchart, one obtains a sensitivity model, which is presented in Fig. 5. The multiplication block $v(t_n) = W \cdot y(t_n)$ in the model in Fig. 4 is a special case of the nonlinear static element (Table 1) with two inputs and one output. The variation of the output equals $\delta v(t_n) = W \cdot \delta y(t_n) + \delta W \cdot y(t_n)$, which corresponds to the part of the sensitivity model. The modified adjoint system is presented in Fig. 6.

Both systems were built under the assumption that matrices W and Q were diagonal and, therefore, it was not necessary to transpose them.

The modified adjoint system of Fig. 6 is used to obtain the input-output sensitivity functions (31). The discrete-time part of the modified adjoint system is updated at time moments $T - t_N, T - t_{N-1}, \ldots, T - t_1$. This system, stimulated by a Kronecker pulse at time moment $T - t_N$, produces discrete-time output signals $\beta(t_n)$ and $\eta(t_n)$. At the final time moment, they are equal to the required input-output sensitivity functions (31):

$$\beta^T(T) = S_{\tilde{p}(0)}^{\tilde{J}(t_N)}, \quad \eta^T(T - t_1) = S_{\tilde{W}(t_1)}^{\tilde{J}(t_N)}. \tag{32}$$

The adjoint system can be considered a continuous-time system stimulated by a series of Dirac pulses generated by the ideal pulser,

$$\sum_{n=1}^{N} WQe(T - t_n)\delta(T - t_n). \tag{33}$$

It is impossible to simulate ideal Dirac pulses. This difficulty may be overcome by simulating the continuous-time part of the adjoint system with jumps of state in discrete time moments. Another approach used by us consists of replacing ideal Dirac pulses by short-duration quasi-ideal pulses. This approximation of the ideal pulse gives satisfactory results.

The next section describes the results of application of the proposed technique of gradient calculation.

7 Numerical Results

The GBPTT method and the resulting modified adjoint system have been used to fit parameters (29) of the model (1)-(15) to data described in Section 2. The data consist of results of four different experiments. This makes calculation of the gradient of the performance index, defined as a sum of four separate performance indexes (23), more difficult. In these experiments, different input signals have been applied (persistent and 1-hour pulses) [14] and measurements were taken with and without knocking-out the protein A-20 [16]. This latter effect can be modeled by an additional input signal of the model.

The gradient-based fitting procedure has been implemented in Matlab. The systems (original and adjoint) have been modeled in Simulink and expressions for nonstationary multidimensional functions $f_x^T(T - t)$ and $f_p^T(T - t)$, appearing in the adjoint system, have been automatically derived using Matlab Symbolic Toolbox. To minimize the

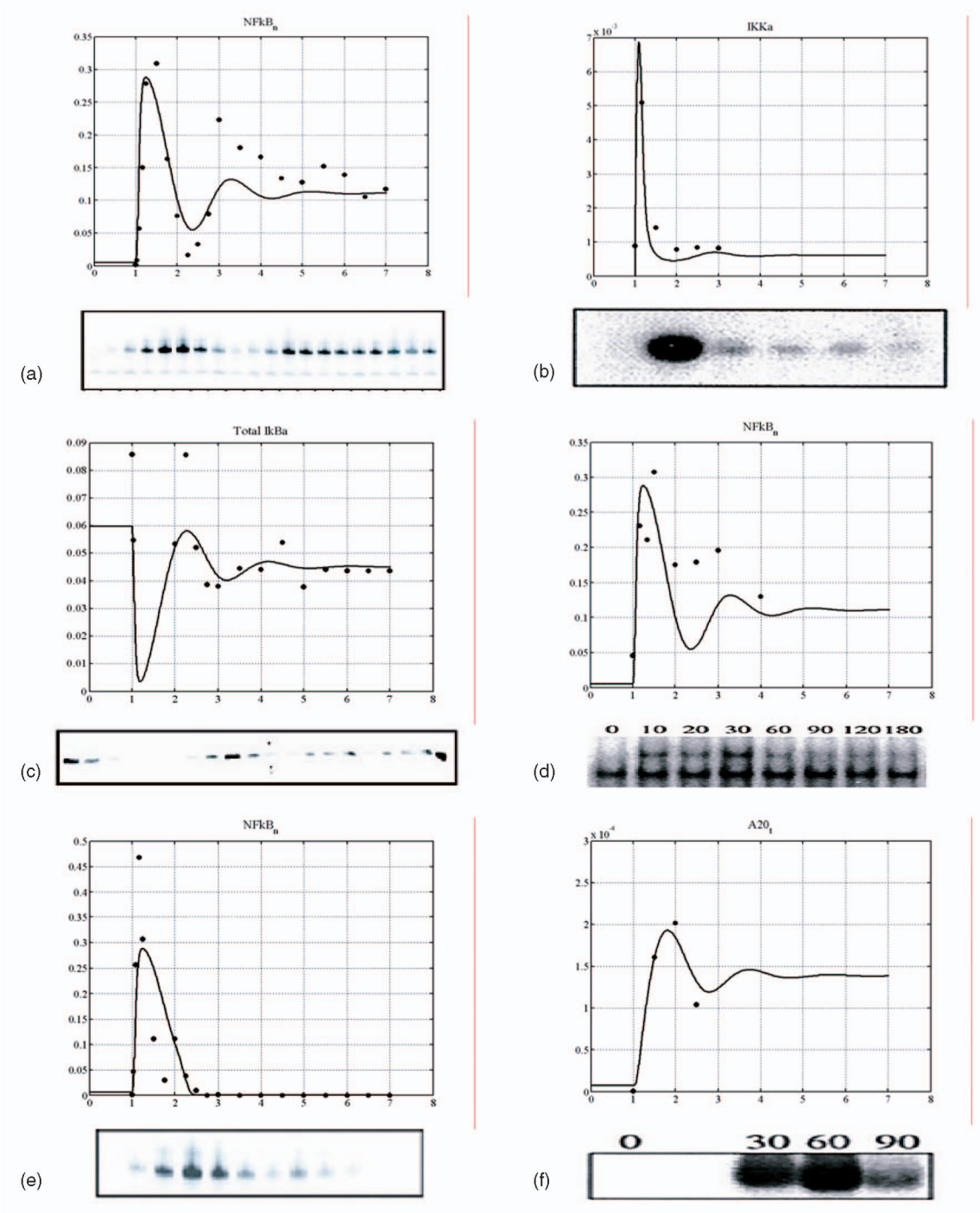


Fig. 7. Results of the fitting procedure. Each plot depicts the output signal of the model with fitted parameters (solid line) and rescaled measurements (dots). Corresponding blot images are also presented. (a), (b) Experiment 1 from [14], persistent TNF stimulation. (c) Experiment 2 from [14], 1-hour TNF stimulation. (d), (e), (f) Experiment 3 from [16] wild type cells, persistent TNF stimulation.

performance index (23), the conjugate gradient algorithm has been applied and 1,000 iterations have been performed. Each iteration includes a simulation of the original system followed by saving the trajectories needed for the adjoint system, and a simulation of the adjoint system followed by updating optimized vectors of parameters and multipliers. Before each iteration, initial conditions for the original system with updated parameters have to be found. These initial conditions are provided by steady state solutions of the system with no excitation. Ideally, they might be found by solving a set of nonlinear algebraic equations describing the steady state. Unfortunately, this problem is numerically ill-conditioned. Therefore, we simulated the model without excitation until the steady state has been reached. Because the measurements (obtained from quantification of the blots) were normalized, the weights q_i in (20) have been chosen equal to 1. The procedure started with parameters fitted in [17] and varied by a factor 1.5 up and down.

Results are presented in Figs. 7 and 8. Each plot presents signals obtained using the model $y_i(t)$ with fitted parameters

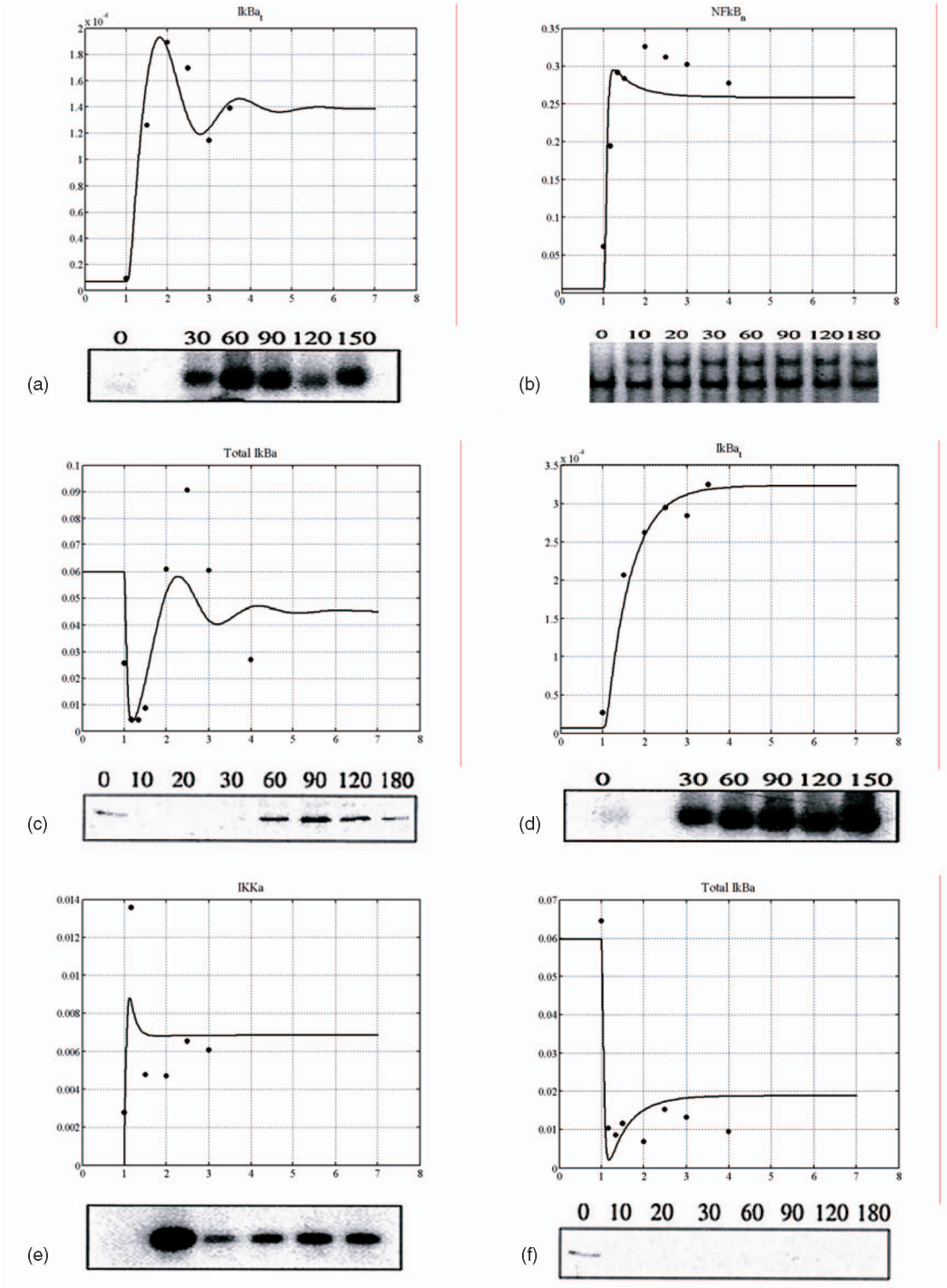


Fig. 8. Results of the fitting procedure (continued). (a), (b) Experiment 3 from [16], wild type cells, persistent TNF stimulation. (c), (d), (e), (f) Experiment 4 from [16], A20 deficient cells, persistent TNF stimulation.

(solid line). In addition, black dots depict quantified data obtained from corresponding blots. Values of quantified data have been divided by fitted multipliers to present them in the same scale as outputs of the model. The corresponding blot image is attached at the bottom of each plot. One can observe the mismatch between the data and the signals of the model. This is caused by observational noise. If a model with more parameters were employed, the mismatch would be reduced, which would result in lower values of the performance index. However, this might lead to overparameterization, i.e., the situation in which the model fits the identification data extremely well but fits any other data rater poorly. The next section includes validation of the model, which allows determining if overparameterization occurred.

TABLE 2
Values of Fitted Parameters

Parameters	Numerical fit	Values fitted heuristically in [17]
k_1	0.0074	0.0025
k_2	0.89	0.10
k_3	0.00072	0.00150
i_1	0.0134	0.0025
e_{2a}	0.018	0.010
i_{1a}	0.0019	0.0010
k_{prod}	0.000017	0.000025
k_{deg}	0.00139	0.000125
c_5	0.0028	0.0003
c_{4a}	0.18	0.50

The values of 10 fitted parameters are collected in Table 2. In the same table, heuristically fitted values from [17] are presented. One can observe differences of up to one order of magnitude between values of parameters that differ. Surprisingly, these differences do not cause large differences in model behavior and the performance index. To obtain the performance index (23) for parameters fitted heuristically, we performed optimization with respect to multipliers under fixed parameters of the model. The result was 2.96, while our numerical fit of course gave a better value of 2.02.

The preliminary conclusion is that there is a large set in the parameter space resulting in similar behavior of the model and similar value of the performance index. This phenomenon may be related to the identifiability of our model and it will be also discussed in the next section of the paper. It might be possible that not all parameters can be estimated from the data or only some functions of parameters can be estimated [18], [26]. Unfortunately, the system of differential equations analyzed in this paper is highly complicated and the so-called similarity transformation cannot be performed easily. We checked the possibility of the simplest type of nonidentifiability in which one or part of the parameters does not influence the performance index and we observed that the performance index varies with variations of all parameters. The identifiability related properties may be also inferred ex post based on the eigenvalue analysis of the covariance matrix of the parameter estimates. The covariance matrix may be calculated as the inverse Hesse matrix of the performance index.

8 Model Validation

Because we did not have access to any additional experimental data, we performed the so-called cross validation approach frequently used in statistics and machine learning. The data analyzed in this paper was produced in four independent experiments described in Section 3. We performed four iterations of cross validation. In each iteration, the results of one experiment have been removed from the data set and the remaining three experiments were used to develop the model. Then, observations from the removed experiment were compared to the output of the model under the same stimulation. Results of the procedure described above are presented in Figs. 9 and 10. Solid lines represent outputs of the model developed based on all data, while dashed lines present predictions of signals not taken into account in model developing. There is no major discrepancy between models developed based on complete versus incomplete data. Sometimes the scale of signals is different, see, for example, Fig. 9d, but the profiles of all signals are conserved. The validation results look very accurate, especially with Experiment 4 data removed. The model obtained based on Experiments 1, 2, and 3 (wild types) was able to predict the behavior of the cell population with A20 knock-out. See Fig. 10c, Fig. 10d, Fig. 10e, and Fig. 10f.

We also examined values of fitted parameters obtained in different cross-validation steps. Results are presented in Fig. 11. The values of parameters differ considerably, some of them by about 100 times. This result is similar to that obtained in the previous section. Different estimates of parameters may yield very similar outputs of the model, even these nonobservable. We also check the influence of the initial parameter values and the influence of additional observational noise. Again, the fitting results (parameters) were sensitive to these factors, while the signals generated by the models (both observed and nonobserved signals) were less sensitive.

9 Conclusions

Models of molecular signaling pathways in cells depend on a large number of parameters, such as reaction rates and degradation rates of proteins, rates of transcription and translation, rates of transport, and other. Some of these parameters, such as transcription rates, are known quite accurately, while others are known up to the order of magnitude. However, there is a group of parameters which are not known and which might vary over a very wide range. This group includes, for example, the parameters of active degradation of proteins and mRNAs. The difficulty is compounded by the fact that human genetic polymorphisms may cause individual variability of some of these parameters. Also, usually it is impossible to measure these parameters in independent, specially designed experiments.

For these reasons, estimation of parameters of a model can only be accomplished by fitting models of pathways involving these parameters to available data. Frequently, the number of independently measured variables is not much greater than the number of coefficients. As an effect, the estimates of parameters are not very reliable. Even if the comparison of model prediction to data leads to a visually acceptable fit, it remains unclear if the fit can be improved. Therefore, it is important to develop procedures aimed at optimization of the fit.

This paper considers an estimation procedure which accomplishes this task. The procedure minimizes the mean square error by using the equivalence between a dynamical system and its reverse-time adjoint. The methodology, Generalized Backpropagation Through Time, was developed in [8] and extended to continuous-discrete systems in [9]. The method originally was developed for systems

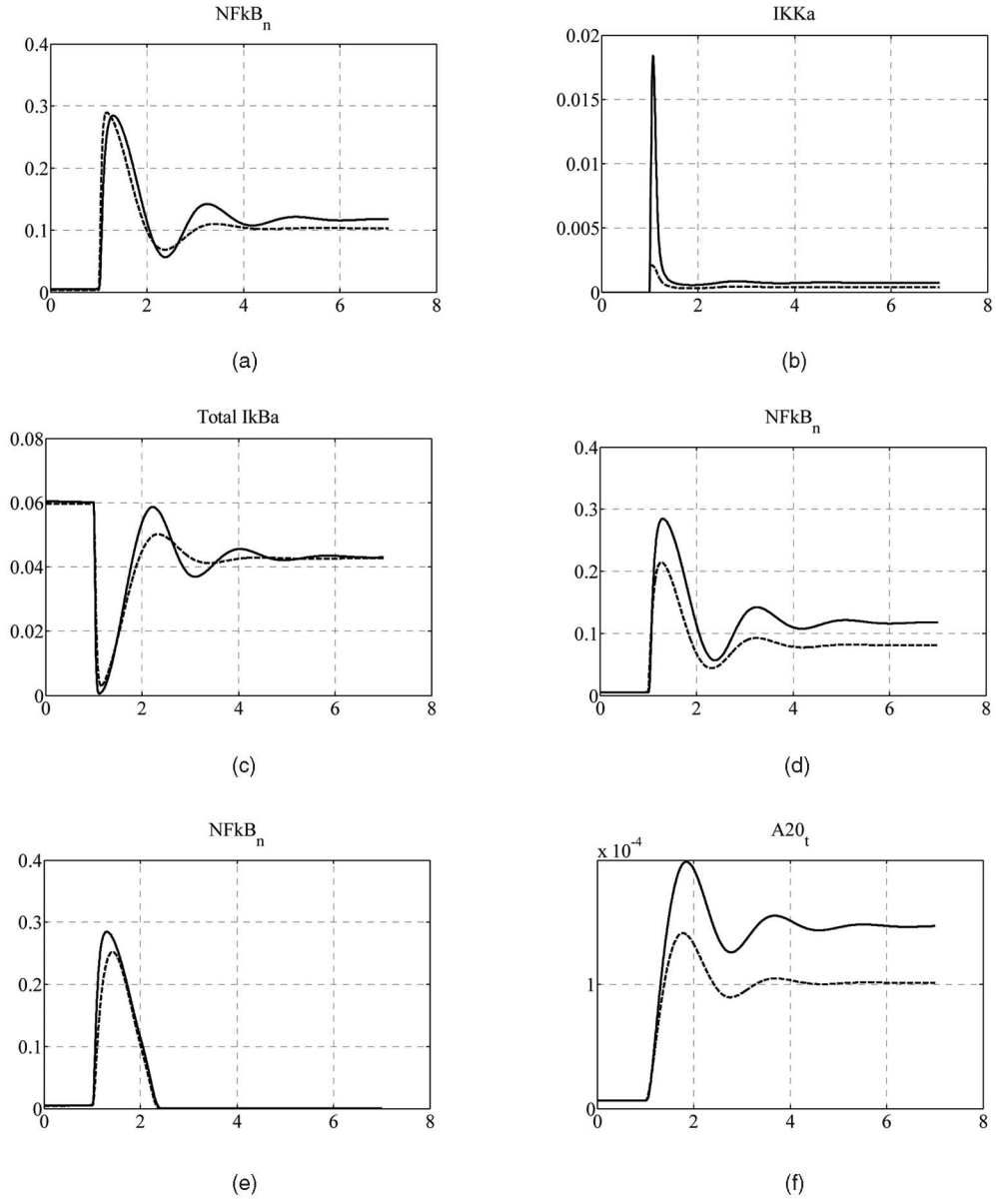


Fig. 9. Validation results. Each plot depicts output signals of the model based on complete data (solid line) versus those obtained by the model developed with data from one experiment missing (dashed line). (a), (b) Experiment 1 from [14], persistent TNF stimulation. (c) Experiment 2 from [14], one hour TNF stimulation. (d), (e), (f) Experiment 3 from [16] wild type cells, persistent TNF stimulation.

defined by a flowchart and has been implemented in the Matlab-Simulink environment. In the current application, since mathematical models of cell signaling pathways assume the form of systems of ODEs, the software allows description of the model. The equations of the adjoint model are created automatically by using Matlab's Symbolic-Toolbox. Then, the conjugate gradient algorithm is applied. Using a standard PC (Pentium IV, 2.4 GHz), the optimization process takes about 4 hours.

The GBPTT method is similar in spirit to other approaches to system analysis proposed in a range of areas such as signal flow graphs theory [23], [7], electrical circuits [4], digital filters [5], [19], automatic differentiation [12], and, finally, neural networks [21], [28]. The GBPTT method is most similar to the approach proposed by Wan and Beaufays [28], but it is more universal, first because not only discrete-time neural networks can be trained. Second, the GBPTT produces a causal adjoint system, generating the gradient of the performance index, while the method proposed in [28] gives a noncausal system containing z operators. Moreover, using the GBPTT method, one finds not only the gradient of a performance index in the space of optimized parameters but also in the space of signals that influence the system. The GBPTT method has been extended to hybrid continuous-discrete systems. In [5], the approach has been applied to learn about continuous-time neural networks based on discrete time measurements.

In this paper, the method is applied to a biological system for the first time. This system, the signaling pathway of the $\mathrm{NF}\kappa\mathrm{B}$ transcription factor, was already modeled by us [17], but parameter estimates were selected using a heuristic iteration. The current application has resulted in estimates which are improved in the sense of reducing the mean square error. The performance index, which is

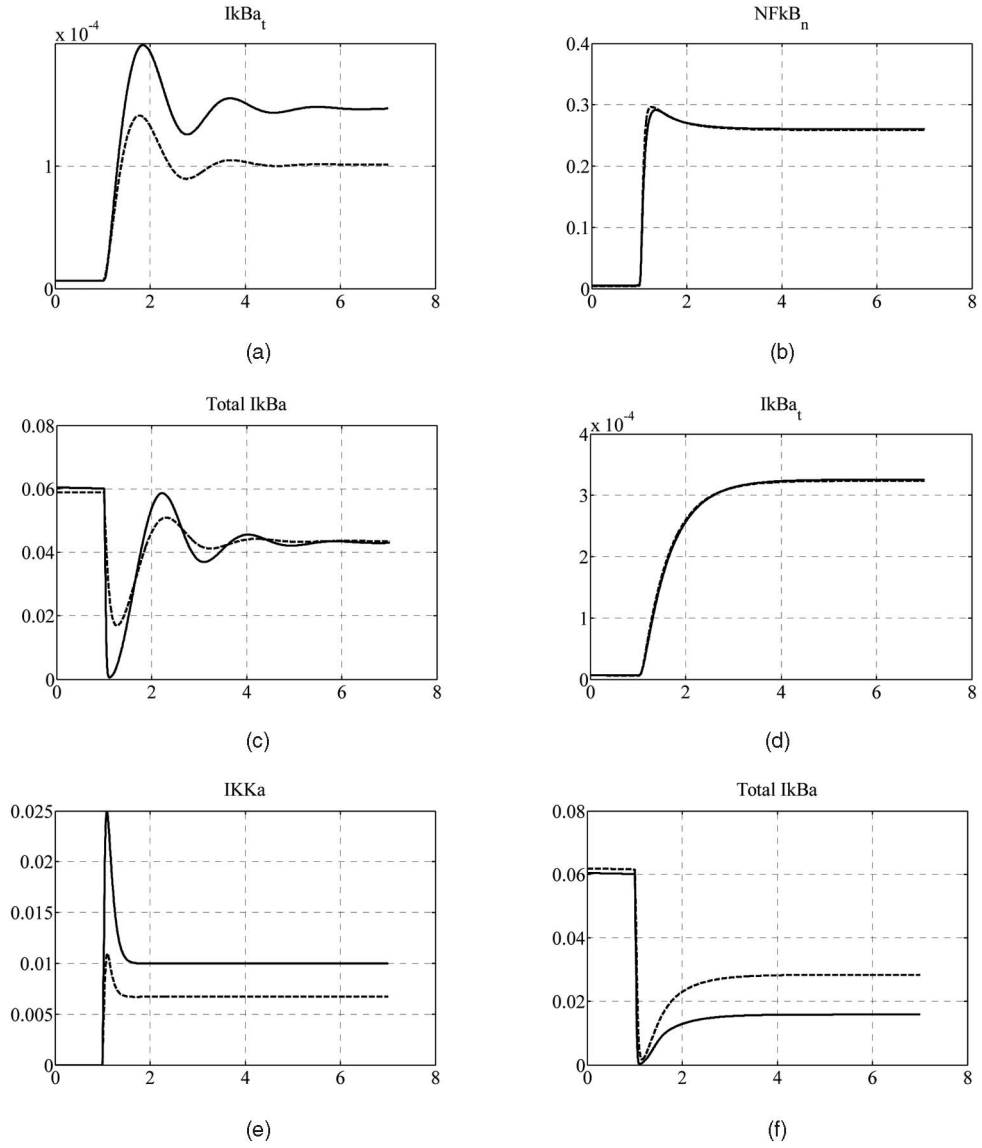


Fig. 10. Validation results (continued). (a), (b) Experiment 3 from [16], wild type cells, persistent TNF stimulation. (c), (d), (e), (f) Experiment 4 from [16], A20 deficient cells, persistent TNF stimulation.

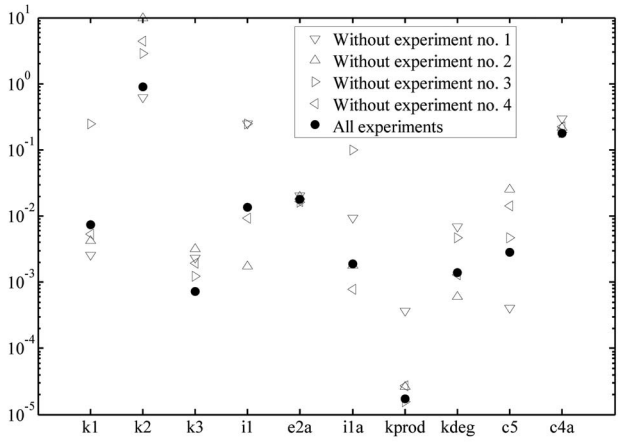


Fig. 11. Values of 10 fitted parameters obtained in four iterations of cross-validation compared to the numerical fit based on all data available.

minimized, is defined as a weighted sum of discrete-time differences between outputs of the model and the measurements. Comparison between heuristic and least-squares estimates reveals their similarity in the sense of minimization of the performance index as well as in the sense of estimated parameter values. The properties of solutions obtained using the method proposed, for example, their uniqueness and sensitivity to noise, are not very well known and will be the subject of future research.

ACKNOWLEDGMENTS

This work has been supported in part by the NHLBI contract N01-HV-28184 (K. Fujarewicz, M. Kimmel, T. Lipniacki), in part by the Polish Ministry of Science and Higher Education grant 3T11A 019 29 (K. Fujarewicz, M. Kimmel, T. Lipniacki) and in part by the EC program contract MRTN-CT-2004-503661 (M. Kimmel, A. Świerniak).

REFERENCES

[1] A. Arkin, J. Ross, and H.A. McAdams, "Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage Lambda-Infected Escherichia Coli Cells," *Genetics,* vol. 149, pp. 1633-1648, 1998.

[2] H. Bock, "Numerical Treatment of Inverse Problems in Chemical Reaction Kinetics," *Modeling of Chemical Reaction Systems,* K. Ebert, P. Deuflhard, and W. Jager, eds., vol. 18, pp. 102-125, 1981.

[3] H. Bock, "Recent Advances in Parameter Identification for Ordinary Differential Equations," *Progress in Scientific Computing,* P. Deuflhard and E. Hairer, eds., vol. 2, pp. 95-121, 1983.

[4] J. Bodrewijk, "Inter-Reciprocity Applied to Electrical Networks," *Applied Scientific Research,* vol. 6B, pp. 1-74, 1956.

[5] R.E. Crochiere and A.V. Oppenheim, "Analysis of Linear Digital Networks," *Proc. IEEE,* vol. 63, no. 4, pp. 581-595, 1975.

[6] H. de Jong, "Modeling and Simulation of Genetic Regulatory Systems: A Literature Review," *J. Computational Biology,* vol. 9, no. 1, pp. 67-103, 2002.

[7] A. Fettweiss, "A General Theorem for Signal-Flow Networks," *Archiv für Elektronik un d Übertragungstechnik,* vol. 25, no. 12, pp. 557-561, Dec. 1971.

[8] K. Fujarewicz, "Optimal Control of Nonlinear Systems Using Generalized Back Propagation through Time," *Proc. 18th IASTED Conf. Modelling, Identification, and Control,* pp. 468-471, Feb. 1999.

[9] K. Fujarewicz and A. Galuszka, "Generalized Backpropagation through Time for Continuous Time Neural Networks and Discrete Time Measurements," *Lecture Notes in Computer Science,* pp. 190-196, 2004.

[10] K. Fujarewicz, M. Kimmel, and A. Swierniak, "On Fitting of Mathematical Models of Cell Signaling Pathways Using Adjoint Systems," *Math. Biosciences and Eng.,* vol. 2, no. 3, pp. 527-534, 2005.

[11] K. Fujarewicz, "Generalized Backpropagation through Time for Continuous-Time Neural Models Identification," *Proc. Industrial Simulation Conf.,* pp. 178-181, June 2006.

[12] "Automatic Differentiation of Algorithms: Theory, Implementation and Application," *Proc. First SIAM Workshop Automatic Differentiation,* A. Griewank and G. Coliss, eds., 1991.

[13] J. Hasty, D. McMillen, F. Isaacs, and J.J. Collins, "Computational Studies of Gene Regulatory Networks: In Numero Molecular Biology," *Nature Rev. Genetics,* vol. 2, pp. 268-279, 2001.

[14] A. Hoffman, A. Levchenko, M.L. Scott, and D. Baltimore, "The $I\kappa B - NF - \kappa B$ Signaling Module: Temporal Control and Selective Gene Activation," *Science,* vol. 298, pp. 1241-1245, 2002.

[15] T. Kailath, *Linear Systems.* Prentice Hall, 1980.

[16] E.G. Lee, D.L. Boone, S. Chai, S.L. Libby, M. Chien, J.P. Lodolce, and A. Ma, "Failure to Regulate TNF-Induced NF-κB and Cell Death Responses in A20-Deficient Mice," *Science,* vol. 289, pp. 2350-2354, 2000.

[17] T. Lipniacki, P. Paszek, A.R. Brasier, B. Luxon, and M. Kimmel, "Mathematical Model of NF-κB Regulatory Module," *J. Theoretical Biology,* vol. 228, pp. 195-215, 2004.

[18] L. Ljung and T. Glad, "On Global Identifiability for Arbitrary Model Parametrizations," *Automatica,* vol. 30, pp. 265-276, 1994.

[19] A. Oppenheim and R. Zoppoli, *Digital Signal Processing.* Prentice Hall, 1989.

[20] K. Schittkowski, "Parameter Estimation in Systems of Nonlinear Equations," *Numerical Math.,* vol. 68, pp. 129-142, 1994.

[21] B. Srinivasan, U.R. Prasad, and N.J. Rao, "Backpropagation through Adjoints for the Identification of Non Linear Dynamic Systems Using Recurrent Neural Models," *IEEE Trans. Neural Networks,* pp. 213-228, Mar. 1994.

[22] P. Smolen, D.A. Baxter, and J.H. Byrne, "Modeling Transcriptional Control in Gene Networks: Methods, Recent Results, and Future Directions," *Bull. Math. Biology,* vol. 62, pp. 247-292, 2000.

[23] D. Tellegen, "A General Network Theorem, with Applications," *Philips Research Report,* vol. 7, pp. 259-269, 1952.

[24] J. Timmer, T.G. Müller, I. Swameye, O. Sandra, and U. Klingmüller, "Modeling the Nonlinear Dynamics of Cellular Signal Transduction," *Int'l J. Bifurcation and Chaos,* vol. 14, no. 6, pp. 2069-2079, 2004.

[25] J.J. Tyson, K.C. Chen, and B. Novak, "Sniffers, Buzzers, Toggles and Blinkers: Dynamics of Regulatory and Signaling Pathway in the Cell," *Current Opinions in Cell Biology,* vol. 15, pp. 221-231, 2003.

[26] S. Vajda, K. Godfrey, and H. Rabitz, "Similarity Transformation Approach to Identifiability of Nonlinear Compartmental Models," *Math. Bioscience,* vol. 93, pp. 217-248, 1989.

[27] B. van Domselaar and P. Hemker, "Nonlinear Parameter Estimation in Initial Value Problems," Technical Report NW 18/75, Math. Centre Amsterdam, 1975.

[28] E. Wan and F. Beaufays, "Diagrammatic Derivation of Gradient Algorithms for Neural Networks," *Neural Computation,* vol. 8, no. 1, pp. 182-201, Jan. 1996.

[29] P.J. Werbos, "Backpropagation through Time: What It Does and How to Do It," *Proc. IEEE,* vol. 78, pp. 1550-1560, 1990.

Krzysztof Fujarewicz received the MS degree in automatic control from the Silesian University of Technology in Gliwice in 1992 and the PhD degree in automatic control and robotics from the Silesian University of Technology in 1999. In 1999, he became an assistant professor at the Institute of Automatic Control, Silesian University of Technology, Gliwice. In 2003-2006, he was a visiting researcher at Rice University, Houston, several times. His main research interests are various applications of neural networks, identification and optimal control of nonlinear systems, classification, clustering, and feature selection for biomedical data.

Marek Kimmel received the doctoral degree from the Silesian Technical University in Gliwice, Poland, in 1980 and, in 1996, the habilitation in mathematical sciences from the Jagiellonian University, Krakow, Poland. He is a professor of statistics at Rice University in Houston, Texas, a professor of biostatistics and applied mathematics (adjunct) at the M.D. Anderson Cancer Center in Houston, and a professor of biometry (adjunct) at the School of Public Health of the University of Texas in Houston as well as a professor in the Systems Engineering Group at the Silesian Technical University in Gliwice, Poland. His postdoctoral training took place at Memorial Sloan-Kettering Cancer Center in New York. Dr. Kimmel is a fellow of the American Statistical Association. His principal interests are stochastic modeling of human disease (in particular lung cancer progression and screening), statistical and population genetics, biostatistics, and bioinformatics. He is a member of the editorial boards of the *Journal of the National Cancer Institute*, *Journal of Theoretical Biology*, *Mathematical Biosciences*, and *Journal of Biological Systems*. He published a monograph (jointly with David Axelrod) *Branching Processes in Biology* (Springer 2002) and coedited several volumes of papers in mathematical biology and biostatistics. He coauthored two statistical genetics software packages, GENOCHECK and CASPAR, available on the NCBI Web site. Dr. Kimmel has advised approximately 30 PhD and MS students in the United States, Poland, and France. He has organized numerous international meetings, schools, and conferences, including the conferences on Mathematical Population Dynamics, started in 1987. His research has been supported by grants from the NIH (NCI), NSF, NATO, and KBN (the Polish Committee for Scientific Research). He has wide experience in data analysis and statistical computation and he taught graduate-level courses on survival analysis, modeling with stochastic processes, statistical process control, and bioinformatics.

Tomasz Lipniacki received the MS degree in physics from the University of Warsaw in 1991 and the PhD degree in mechanics from the Institute of Fundamental Technological Research (IFTR) in 1998. From 2003-2006, he was a visiting researcher at Rice University, Houston, Texas. He is currently an assistant professor in the Department of Mechanics and Physics of Fluids at IFTR. He has published approximately 20 papers in internationally refereed journals. His main research activity is in the mechanics of superfluids, DNA dynamics, regulatory pathways of innate immunity, and stochastic processes in intracellular dynamics.

Andrzej Świerniak received the MA degree in mathematics from the University of Silesia, Katowice, Poland, in 1975, and the MS, PhD, and DS (habilitation) degrees, all in control engineering, from the Silesian University of Technology, Gliwice, Poland, in 1972, 1978, and 1988, respectively. He is presently a full professor of automatic control and bioinformatics and the head of the Department of Automatic Control, Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology. He was a visiting researcher at Oxford University, United Kingdom, Rice University, Houston, Texas, and the University of Montreal, Canada, in 1982, 1993, 1994, respectively, a visiting associate professor at the University of Mississippi in 1989, and a visiting scholar at Ohio State University, Columbus, in 2003. Dr. Swierniak is an author or coauthor of more than 250 journal articles, book chapters, and conference papers, guest editor of special issues of *Mathematical Bioscience and Engineering*, the *Archives of Control Sciences*, and the *International Journal of Applied Mathematics and Computer Science*, and an advisor of the *Journal of Biological Systems*. He was elected a member of the Polish Committee for Scientific Research, Committee of Automatic Control and Robotics, and Committee of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences. His main research interests include the theory of optimal and robust control and its applications in molecular biology, biotechnology, bioinformatics, and medicine. He is member of the IEEE, American Mathematical Society, Society of Mathematical Biology, Polish Society of Theoretical and Applied Electrotechnics, Polish Mathematical Society, and European Mathematical Society.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.