Physics Letters A 374 (2010) 4496–4505

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

Parameter estimation based synchronization for an epidemic model with application to tuberculosis in Cameroon

Samuel Bowong [a,b,*,1], Jurgen Kurths [b,c]

[a] Laboratory of Applied Mathematics, Department of Mathematics and Computer Science, Faculty of Science, University of Douala, P.O. Box 24157 Douala, Cameroon
[b] Postdam Institute for Climate Impact Research (PIK), Telegraphenberg A 31, 14412 Potsdam, Germany
[c] Department of Physics, Humboldt Universitat zu Berlin, 12489 Berlin, Germany

ARTICLE INFO

Article history:
Received 6 May 2010
Received in revised form 14 June 2010
Accepted 2 September 2010
Available online 6 September 2010
Communicated by A.R. Bishop

Keywords:
Epidemiological models
Parameter identification
Adaptive synchronization
Tuberculosis

ABSTRACT

We propose a method based on synchronization to identify the parameters and to estimate the underlying variables for an epidemic model from real data. We suggest an adaptive synchronization method based on observer approach with an effective guidance parameter to update rule design only from real data. In order, to validate the identifiability and estimation results, numerical simulations of a tuberculosis (TB) model using real data of the region of Center in Cameroon are performed to estimate the parameters and variables. This study shows that some tools of synchronization of nonlinear systems can help to deal with the parameter and state estimation problem in the field of epidemiology. We exploit the close link between mathematical modelling, structural identifiability analysis, synchronization, and parameter estimation to obtain biological insights into the system modelled.

1. Introduction

The spread of communicable disease is a dynamical process, as such, the understanding and control of infectious disease outbreaks and epidemics is pertinent to the temporal evolution of disease propagation. Historically, this aspect of disease transmission has been studied using coarse-grained dynamical representation of populations, known as compartmental models [1–5]. In these models, a population is divided into a number of epidemiological states (or classes) and the time evolution of each state is described by a differential equation. Deterministic nonlinear ordinary differential equation models of infection dynamics have proved to be very useful in terms of providing insight into infection processes and their control [4]. A lot of basic results have been obtained. The basic and important research subjects for these systems are the existence of a threshold value which distinguishes whether an infectious disease will die out, the local and global stability of the disease-free equilibrium and endemic equilibria, the existence of periodic solutions, the persistence and extinction of the disease, etc. Although these approaches, and their more complex variants, have been instrumental in understanding several features of infectious diseases over the past three decades, a systematic construction of an epidemiological model that best matches parameter estimation of the model equations from real data is needed. However, relating such models to observed data is not a straightforward matter due to the complexity of the model and the noise in the data. A brief survey on previous studies provides the context of this Letter. Some parameter estimation methods have been developed for longitudinal clinical data [6]. Huang and Wu [7], and Huang et al. [8] have recently proposed a Bayesian approach to estimate the parameters in HIV dynamics models for a population of patients. Their method is highly dependent on informative priors for most of the parameters and is computationally intensive, which may limit its use. Xia [9], Xia and Moog [10], and Jeffrey and Xia [11] investigated the identifiability of HIV dynamics models, but the solution of the identifiability problem was not provided for the case where only measurements of viral load are available. Hulin et al. [12] used the concept of identification function and identification equation to study the parameter identifiability of a three-dimensional HIV/AIDS dynamics, but two parameters are not identifiable (i.e., are indistinguishable). Unfortunately, nothing has been done in term of estimating parameters of tuberculosis. The approach taken in this Letter is based on synchronization. Synchronization is a well-known phenomenon in biology, physics, engineering and many other scientific disciplines [13–19].

* Corresponding author at: Laboratory of Applied Mathematics, Department of Mathematics and Computer Science, Faculty of Science, University of Douala, P.O. Box 24157 Douala, Cameroon. Tel.: +237 99 96 41 64; fax: +237 22 31 02 90.
E-mail address: sbowong@gmail.com (S. Bowong).
[1] Samuel Bowong is also with UMI 209 IRD/UPMC UMMISCO, Bondy, Projet MASAIE INRIA Grand Est, France and Projet GRIMCAPE LIRIMA, Cameroun.

0375-9601/$ – see front matter
doi:10.1016/j.physleta.2010.09.008

Adaptive synchronization based parameter identification methods, among others, are widely used and involve constructing a response system (called "computational model") to synchronize with the drive system (i.e., the "real" system with some or all unknown parameters to be estimated) through the unidirectional coupling signal. Due to information acquisition restrictions (or cost), we often have to face a challenging scenario in which multiple parameters need to be estimate but only a scalar (physical) variable is measurable. Because one usually cannot find proper parameter update rules by exploiting information obtained from only a scalar time series, parameter identification with adaptive synchronization from a scalar time series is not well understood and still remains challenging until now. Therefore, it is not only of theoretical interest but also of practical value to synchronize nonlinear dynamical systems with (multiple) unknown parameters from real data. Although, the parameter and state estimation problem for nonlinear systems, due to its difficulty and its practical interest, is one of the most challenging in control theory, it has never been addressed for estimating state and unknown parameters of epidemiological models.

In this Letter, we present a synchronization based parameter estimation approach to matching dynamical variables of an SEI epidemiological model to real data. We show that the model is algebraically observable and identifiable with respect to the output data, hence, a differential parametrization of the output and its time derivatives can be obtained. Based on these facts, we propose a synchronization method based on an adaptive observer to parameter estimation by exploiting information obtained only from real data, which is crucial for many practical applications such as the prediction and the control of an infectious disease. We analyze the convergence of the adaptive observer. The proposed synchronization method includes three steps: (i) finding some proper control signals such that the system's output synchronizes with the real data; (ii) designing parameter update rules in terms of a necessary condition for ensuring local synchronization; and (iii) determining the value for each parameter update rate for ensuring the local stability of the synchronization manifold. Simulation studies are performed to estimate all parameters of a tuberculosis model (TB) from noisy data of the region of Center in Cameroon. Compared to existing results [6–12], our method is less computationally intensive and easier to implement. In addition, all parameters are identifiable which is not the case of many works in the literature and the proposed adaptive observer has fast convergence property. It is our view that this study represents the first work that provides an in-depth parameter estimation based on synchronization. Although we have applied the method to TB, the proposed methodology is generally applicable to any infectious disease modelled by ordinary differential equations.

2. The model

The aim of this section is to describe the model and state the parameter estimation problem.

2.1. Model construction

In our model, each individual is in one of three permitted states: susceptible (S), latently infected (E) or infectious (I). The state of the individuals evolves in time and depends on their previous state and the connections with other individuals. The probabilities of transitions between different states are described with the following parameters: Λ is the recruitment into the population; β, the probability that a susceptible individual will be infected by infectious (it is also denotes how contagious the disease is); α is the probability that a latently infected individual becomes infectious (this value is connected with the average time of incubations); γ is the probability that an infectious individual will recover or be isolated from the rest of the population (e.g., in a hospital); μ is the probability that an individual in the population died from reasons not related to the disease; and d is the probability that an infectious individual died of the disease. For the present case, the changes in time of individuals is one of the possible states S, E and I are described by the following system of differential equations:

$$\begin{cases} \dot{S} = \Lambda - \beta SI - \mu S, \\ \dot{E} = \beta SI + \gamma I - (\mu + \alpha)E, \\ \dot{I} = \alpha E - (\mu + d + \gamma)I. \end{cases} \tag{1}$$

2.2. Problem formulation

Setting $x_1 = S$, $x_2 = E$ and $x_3 = I$, the model (1) may be written as follows:

$$\begin{cases} \dot{x}_1 = \Lambda - \beta x_1 x_3 - \mu x_1, \\ \dot{x}_2 = \beta x_3 x_1 + \gamma x_3 - (\mu + \alpha)x_2, \\ \dot{x}_3 = \alpha x_2 - (\mu + d + \gamma)x_3. \end{cases} \tag{2}$$

Let $\theta = (\beta, \Lambda, \alpha, d, \gamma, \mu)$ denotes all the system parameters. We suppose that the number of infectious is available for measurement, that is, the number of infectious x_3 can be obtained from real data.

The problem addressed in this Letter consists in determining the unknown parameters Λ, β, α, γ, d and μ from a given recorded set

$$\{x_3(t_k)\}_0^{k=n}, \quad t_k \in \mathfrak{T},$$

where $\mathfrak{T}$ is a discrete set of observable times

$$\mathfrak{T} = (t_1, t_2, \ldots, t_n), \qquad t_{j+1} - t_j = T, \quad j = \{1, 2, \ldots, n-1\}.$$

To achieve this goal, we shall build an adaptive observer for system (1). This is the aim of the next section.

3. Results

3.1. Some algebraic properties

Before tackling the practical problem of estimating the parameters in our epidemiological model from real data, one must first consider a structural question relating to the model proposed. Assuming perfect, noise-free and continuous output data, it is necessary to check whether all of the model parameters are uniquely determined by a given output. Should any of the parameters (or combinations of parameters) have physical, or practical, significance then it is essential that the answer to this structural identifiability (see, for example [20–23]) be established. It is important to note that an identifiable model does not automatically guarantee a good fit to experimental data, or that a good fit is obtained only for a unique set of parameter values. However, for an unidentifiable model an uncountable number of parameter vectors will give the same fit to the experimental data.

We introduce two useful properties [24] that system (2) satisfies.

Definition 1. Consider an undetermined system of ordinary differential equations:

$$X(t, x, \dot{x}, \theta) = 0, \tag{3}$$

where $x = (x_i)_1^{i=n} \in \mathbb{R}^n$ is a state vector and $\theta \in \mathbb{R}^p$ is a constant parameter vector. Suppose that there exists a smooth, local and one to one correspondence between the solution $x(t)$ of system (3) and an arbitrary function $y(t) = h(t, x(t)) \in \mathbb{R}$; then, the state x_i is said to be algebraically observable with respect to $y(t)$ if it satisfies

$$x_i = \frac{f_i(y, \ldots, y^{(q)}, \theta)}{g_i(y, \ldots, y^{(s)}, \theta)},$$

where f_i, g_i and h are smooth maps, $y^{(k)}$ is the kth derivative of y, q and s are integers with $q \leqslant s$. The variable y is the output. If x_i is observable for every $i = 1, \ldots, n$, then we say that the system is completely observable.

Indeed, we show that system (2) satisfies the previous definition when the output vector is $y = x_3$. Clearly, the variables x_1 and x_2 can be written as

$$x_1 = \frac{\ddot{y} + (2\mu + \alpha + d + \gamma)\dot{y} + [\mu(\mu + d + \gamma) + \alpha(\mu + d)]y}{\beta\alpha y} \quad \text{and} \quad x_2 = \frac{\dot{y} + (\mu + d + \gamma)y}{\alpha}, \tag{4}$$

hence, system (2) is algebraically observable with respect to the selected output.

Identifiability is a basic system property and determines whether all parameters can be uniquely estimated based on measured outputs. For a formal definition of identifiability of dynamics systems, we refer the readers to Refs. [25,26]. The basic idea of algebraic identifiability is to allow one to identify parameters by solving algebraic equations based only on the initial values and the measurements of outputs variables.

Definition 2. System (3) is said to be algebraically identifiable if there exists a time t^*, a positive integer k, and a function $W : \mathbb{R}^p \times \mathbb{R}^{p(k+1)} \to \mathbb{R}^p$ such that

$$\det \frac{\partial W}{\partial \theta} \neq 0, \tag{5}$$

where det(.) stands for the determinant of $\frac{\partial W}{\partial \theta}$ and

$$W\big(y, \dot{y}, \ddot{y}, \ldots, y^{(k)}, \theta\big) = 0, \tag{6}$$

hold on $[0, t^*]$, where $\dot{y}$, $\ddot{y}$, $y^{(k)}$ are the derivatives of y with respect to t. Eq. (6) is called the identification equation and the function $W(.)$ the identification function.

To formulate an identification function, we may need to eliminate the unobservable (susceptible and latent) state variables from the original system (2) by taking higher-order derivatives of the output (observation) variables. For our model, based on the third equation of (2), we have for the second derivative:

$$\ddot{y} = \beta\alpha x_1 y - (2\mu + \alpha + d + \gamma)\dot{y} - \big[\mu(\mu + d + \gamma) + \alpha(\mu + d)\big]y. \tag{7}$$

Continuing to take the 3rd-order derivative and combining it with the first equation of system (2), we obtain

$$\begin{aligned} y^{(3)} = {} & \frac{\dot{y}\ddot{y}}{y} + (2\mu + \alpha + d + \gamma)\left(\frac{\dot{y}^2}{y} - \ddot{y}\right) + \big[\beta\alpha\Lambda - \mu\big[\mu(\mu + d + \gamma) + \alpha(\mu + d)\big]\big]y \\ & - \beta\big[\mu(\mu + d + \gamma) + \alpha(\mu + d)\big]y^2 - \beta y\ddot{y} - \beta(2\mu + \alpha + d + \gamma)y\dot{y} - \mu\ddot{y} - \mu(2\mu + \alpha + d + \gamma)\dot{y}, \end{aligned} \tag{8}$$

an equation that does not depend on the unobservable (susceptible and latent) state variables x_1 and x_2.

Finally, we conclude that system (2) is identifiable with respect to the output y, because the above differential parametrization of the output y can be written as

$$W\big(y, \dot{y}, \ddot{y}, y^{(3)}, \theta\big) = 0.$$

For a dynamical system with unobservable (susceptible and latent) state variables such as the case of the dynamical model (2), when the initial values of the unobservable (susceptible and latent) state variables are unknown, we may not be able to directly use the original

dynamical system to estimate the identifiable but unknown parameters. In this case, we may need to employ the identification equation (8) to obtain parameter estimates. For example, we can used Eq. (8) to deal with this problem for the epidemiological dynamical system (1) when only the number of infectious y is observable. To do this, let $y_1 = y$, $y_2 = \dot{y}$ and $y_3 = \ddot{y}$ and rewrite Eq. (8) as

$$\begin{cases} \dot{y}_1 = y_2, \\ \dot{y}_2 = y_3, \\ \dot{y}_3 = \dfrac{y_2 y_3}{y_1} + (2\mu+\alpha+d+\gamma)\left(\dfrac{y_2^2}{y_1} - y_3 - \beta y_1 y_2 - \mu y_2\right) + \beta\alpha\Lambda y_1 \\ \qquad - \beta y_1 y_3 - \mu y_3 - \big[\mu(\mu+d+\gamma)+\alpha(\mu+d)\big]\big(\mu y_1 + \beta y_1^2\big). \end{cases} \tag{9}$$

This system is solvable if the initial values of the output (observable) variable y_1 and its derivatives y_2 and y_3 are available. Thus, the identifiable parameters can be estimated from this system of equations and the initial values of the unobservable (susceptible and latent) state variables (S and E) are not needed. Usually, the initial values $y_1(t_0)$, $y_2(t_0)$ and $y_3(t_0)$ are not known exactly and they can be treated as additional parameters to be estimated from the data. This will add the number of unknown parameters and, hence will increase the complexity of the parameter estimation problem. To summarize, the dynamical system (9) does not suffice to compute the value of the states nor to estimate the unknown parameters of the system, because one needs to know the value of the real initial condition, which is unavailable for measurement. To overcome this difficulty, we shall used a tool from control theory called adaptive observer. This is, we shall construct another dynamical system whose state will provide an estimate of the real unmeasured state and unknown parameters of the considered model, and this will be true regardless of the observer's initial condition: we need not to take care about the choice of the initial condition of the observer.

3.2. Model parameter estimation using an adaptive observer

One of the important problems in control theory is to reconcile the available data with the mathematical model. This problem is known as the observability problem, and it is related to the construction of "observer" (called some times software sensors) for dynamical systems. In this section, we show how to apply this theory to address the parameter estimation problem for the epidemiological model (1). More precisely, we present an adaptive observer, i.e., a dynamical system that is driven by the given time series data and that possesses additional ODEs governing (slow) variations of model parameters. Once the parameters have converged to the "right" values, the model synchronizes with the driving time-series data. The adaptive observer is based on model (9). To build an adaptive observer some extra terms are added to implement driving by real data s:

$$\begin{cases} \dot{y}_1 = y_2 - L_1\rho(s - y_1), \\ \dot{y}_2 = y_3 - L_2\rho^2(s - y_1), \\ \dot{y}_3 = \dfrac{y_2 y_3}{y_1} + (2\mu+\alpha+d+\gamma)\left(\dfrac{y_2^2}{s} - y_3 - \beta s y_2 - \mu y_2\right) + \beta\alpha\Lambda s - \beta s y_3 \\ \qquad - \mu y_3 - \big[\mu(\mu+d+\gamma)+\alpha(\mu+d)\big]\big(\mu s + \beta s^2\big) - L_3\rho^3(s - y_1), \end{cases} \tag{10}$$

where ρ is the so-called high-gain parameter, and parameters β, Λ, α, d, γ and μ are adjusted as follows

$$\begin{aligned} &\dot{\beta} = \delta_1 N_1(s, y_2, y_3)(s - y_1), \qquad \dot{\Lambda} = \delta_2 N_2(s, y_2, y_3)(s - y_1), \qquad \dot{\alpha} = \delta_3 N_3(s, y_2, y_3)(s - y_1), \\ &\dot{d} = \delta_4 N_4(s, y_2, y_3)(s - y_1), \qquad \dot{\gamma} = \delta_5 N_5(s, y_2, y_3)(s - y_1) \quad \text{and} \quad \dot{\mu} = \delta_6 N_6(s, y_2, y_3)(s - y_1), \end{aligned} \tag{11}$$

where δ_i are suitably positive constants, the mappings $N_i(s, y_2, y_3)$ (with $i = 1, 2, \ldots, 6$) need to be specified, and L_1, L_2 and L_3 are constants to be determined later.

Without loss of generality, we assume that higher-order derivatives of s can be estimated using local polynomials or other smoothing methods based on the real data s. Also, suppose that the true values $\bar{\Lambda}$, $\bar{\alpha}$, $\bar{d}$, $\bar{\gamma}$ and $\bar{\mu}$ of the parameters Λ, α, d, γ and μ, respectively are known as a "preknowledge". Notice that this assumption has been introduced only to show how to design the mappings $N_i(s, y_2, y_3)$. Therefore we assume this assumption in order to proceed with the discussion.

Let $e_1 = \rho^2(s - y_1)$, $e_2 = \rho(\bar{y}_2 - y_2)$, $e_3 = \bar{y}_3 - y_3$, $\tilde{\beta} = \bar{\beta} - \beta$, $\tilde{\Lambda} = \bar{\Lambda} - \Lambda$, $\tilde{\alpha} = \bar{\alpha} - \alpha$, $\tilde{d} = \bar{d} - d$, $\tilde{\gamma} = \bar{\gamma} - \gamma$ and $\tilde{\mu} = \bar{\mu} - \mu$ where $\bar{y}_2 = \dot{s}$ and $\bar{y}_3 = \ddot{s}$. Then, we obtain the following error system:

$$\begin{cases} \dot{e}_1 = \rho e_2 + L_1\rho e_1, \\ \dot{e}_2 = \rho e_3 + L_2\rho e_1, \\ \dot{e}_3 = \dot{\bar{y}}_3 - \dfrac{y_2 y_3}{y_1} - \big[2(\bar{\mu}-\tilde{\mu}) + \bar{\alpha} - \tilde{\alpha} + \bar{d} - \tilde{d} + \bar{\gamma} - \tilde{\gamma}\big]\left(\dfrac{y_2^2}{s} - y_3 - (\bar{\beta}-\tilde{\beta})s y_2 - (\bar{\mu}-\tilde{\mu})y_2\right) \\ \qquad - (\bar{\beta}-\tilde{\beta})(\bar{\alpha}-\tilde{\alpha})(\bar{\Lambda}-\tilde{\lambda})s + (\bar{\beta}-\tilde{\beta})s y_3 + (\bar{\mu}-\tilde{\mu})y_3 \\ \qquad + \big[(\bar{\mu}-\tilde{\mu})(\bar{\mu}-\tilde{\mu}+\bar{d}-\tilde{d}+\bar{\gamma}-\tilde{\gamma}) + (\bar{\alpha}-\tilde{\alpha})(\bar{\mu}-\tilde{\mu}+\bar{d}-\tilde{d})\big]\big[(\bar{\mu}-\tilde{\mu})s + (\bar{\beta}-\tilde{\beta})s^2\big] + L_3\rho e_1, \\ \dot{\tilde{\beta}} = -\dfrac{\delta_1}{\rho^2}N_1(z,\theta)e_1, \qquad \dot{\tilde{\Lambda}} = -\dfrac{\delta_2}{\rho^2}N_2(z,\theta)e_1, \qquad \dot{\tilde{\alpha}} = -\dfrac{\delta_3}{\rho^2}N_3(z,\theta)e_1, \\ \dot{\tilde{d}} = -\dfrac{\delta_4}{\rho^2}N_4(z,\theta)e_1, \qquad \dot{\tilde{\gamma}} = -\dfrac{\delta_5}{\rho^2}N_5(z,\theta)e_1, \qquad \dot{\tilde{\mu}} = -\dfrac{\delta_6}{\rho^2}N_6(z,\theta)e_1, \end{cases} \tag{12}$$

where $z = (s, y_2, y_3)^T$.

Notice that after the parameter estimation ($\tilde{\beta}, \tilde{\Lambda}, \tilde{\alpha}, \tilde{d}, \tilde{\gamma}, \tilde{\mu} = 0$) is ensured, the error system becomes

$$\dot{e} = \rho A e + B f(z), \tag{13}$$

where $e = (e_1, e_2, e_3)^T$,

$$f(z) = \dot{\bar{y}}_3 - \frac{y_2 y_3}{y_1} - (2\bar{\mu} + \bar{\alpha} + \bar{d} + \bar{\gamma})\left(\frac{y_2^2}{s} - y_3 - \bar{\beta} s y_2 - \bar{\mu} y_2\right) - \bar{\beta}\bar{\alpha}\bar{\Lambda} s + \bar{\beta} s y_3 + \bar{\mu} y_3 + \left[\bar{\mu}(\bar{\mu} + \bar{d} + \bar{\gamma}) + \bar{\alpha}(\bar{\mu} + \bar{d})\right](\bar{\mu} s + \bar{\beta} s^2),$$

$$A = \begin{bmatrix} L_1 & 1 & 0 \\ L_2 & 0 & 1 \\ L_3 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Therefore, the necessary condition for a successful parameter estimation is that one can find a proper form for each L_i (with $i = 1, 2, 3$) such that the error system (13) is globally/locally asymptotically stable at the origin. The constants L_1, L_2 and L_3 are chosen such that the polynomial $P(\lambda) = \lambda^3 - L_1\lambda^2 - L_2\lambda - L_3$ satisfies the Routh–Hurwitz condition (i.e., all its roots are contained in the left-hand side of the complex plane). For the error system (13), we choose a typical form for the Lyapunov function

$$V(e) = e^T P e,$$

where P is a positive definite and symmetric matrix, solution of the Lyapunov equation $A^T P + PA = -Q$ with Q a positive definite and symmetric matrix. Since the trajectories of system (1) are bounded, we can assume that the trajectories of system (9) are also bounded. Hence, $\|z\| \leqslant r$ for a certain $r > 0$. Furthermore, since the parameters $\bar{\Lambda}$, $\bar{\alpha}$, $\bar{d}$, $\bar{\gamma}$ and $\bar{\mu}$ are assumed to be constant, one can suppose that $|f(z)| \leqslant k_f$ for some $k_f > 0$. Differentiating V with respect to time yields

$$\begin{aligned} \dot{V}(e) &= \dot{e}^T P e + e^T P \dot{e} = \rho e^T \left(A^T P + PA\right) e + 2 e^T P B f(z) \\ &\leqslant -\lambda_{min}(Q)\rho\|e\|^2 + 2\lambda_{max}(P)\|B\|\|e\|\left\|f(z)\right\| \\ &\leqslant -\rho\lambda_{min}(Q)\|e\|^2 + 2 r k_f \lambda_{max}(P), \end{aligned}$$

where $\lambda_{min}(Q)$ and $\lambda_{max}(P)$ are the minimum and maximum eigenvalues of the matrices Q and P, respectively. Therefore, the error e exponentially decays and is ultimately bounded. Note that as ρ increases, $e(t)$ will decrease, which also decrease the exponential estimation error bound. This argument shows that with the proposed method, ρ should be made as large as possible. Notice that one can also guarantee the local asymptotic stability according to the conditional Lyapunov exponents approach.

However, even when one can find the constants L_1, L_2 and L_3 such that system (13) is locally asymptotically stable, it is easy to see from the error system (12) that the difference between β, Λ, α, d, γ and μ and their trues values $\bar{\Lambda}$, $\bar{\alpha}$, $\bar{d}$, $\bar{\gamma}$ and $\bar{\mu}$ which are unknown for us in general will destroy the synchronization between the system's output and the real data. To preserve synchronization, one has to choose the functions $N_i(z, \theta)$.

To this purpose, by linearizing the error system (12) at the origin, the Jacobian matrix is given by

$$J = \begin{bmatrix} L_1\rho & \rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ L_2\rho & 0 & \rho & 0 & 0 & 0 & 0 & 0 & 0 \\ L_3\rho & 0 & 0 & \tau_1(z,\bar{\theta}) & \tau_2(z,\bar{\theta})s & \tau_3(z,\bar{\theta}) & \tau_4(z,\bar{\theta}) & \tau_5(z,\bar{\theta}) & \tau_6(z,\bar{\theta}) \\ -\frac{\delta_1}{\rho^2}N_1(z,\theta) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\delta_2}{\rho^2}N_2(z,\theta) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\delta_3}{\rho^2}N_3(z,\theta) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\delta_4}{\rho^2}N_4(z,\theta) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\delta_5}{\rho^2}N_5(z,\theta) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\delta_6}{\rho^2}N_6(z,\theta) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

where

$$\tau_1(z,\bar{\theta}) = -(2\bar{\mu} + \bar{\alpha} + \bar{d} + \bar{\gamma}) s y_2 + \bar{\alpha}\bar{\Lambda} s - \left[\bar{\mu}(\bar{\mu} + \bar{d} + \bar{\gamma}) + \bar{\alpha}(\bar{\mu} + \bar{d})\right] s^2 - s y_3, \qquad \tau_2(z,\bar{\theta}) = \bar{\beta}\bar{\alpha} s,$$

$$\tau_3(z,\bar{\theta}) = \frac{y_2^2}{s} - y_3 - \bar{\beta} s y_2 - \bar{\mu} y_2 + \bar{\beta}\bar{\Lambda} s - (\bar{\mu} + \bar{d})(\bar{\mu} s + \bar{\beta} s^2), \qquad \tau_4(z,\bar{\theta}) = \frac{y_2^2}{s} - y_3 - \bar{\beta} s y_2 - \bar{\mu} y_2 - \bar{\alpha}(\bar{\mu} s + \bar{\beta} s^2),$$

$$\tau_5(z,\bar{\theta}) = \frac{y_2^2}{s} - y_3 - \bar{\beta} s y_2 - \bar{\mu} y_2 - \bar{\mu}(\bar{\mu} s + \bar{\beta} s^2),$$

$$\tau_6(z,\bar{\theta}) = 2\left(\frac{y_2^2}{s} - y_3 - \bar{\beta} s y_2 - \bar{\mu} y_2\right) - (2\bar{\mu} + \bar{\alpha} + \bar{d} + \bar{\gamma})(y_2 + \bar{\mu} s + \bar{\beta} s^2) - \left[\bar{\mu}(\bar{\mu} + \bar{d} + \bar{\gamma}) + \bar{\gamma}(\bar{\mu} + \bar{d})\right] s - y_3.$$

It follows that the characteristic polynomial of the Jacobian is defined by

$$\phi(\lambda) = -\lambda^5\left[\lambda^4 - L_1\rho\lambda^3 - L_2\rho^2\lambda^2 - L_3\rho^3\lambda + \Theta(z,\bar{\theta})\right],$$

Table 1
Numbers of reported TB cases in the region of Center in Cameroon from 2003 to 2007.

Quarter	2003	2004	2005	2006	2007
First quarter	525	618	719	748	707
Second quarter	515	594	729	770	644
Third quarter	524	557	716	675	691
Four quarter	569	731	731	657	772

where

$$\Theta(z,\bar{\theta}) = \delta_1 N_1\tau_1 + \delta_2 N_2\tau_2 + \delta_3 N_3\tau_3 + \delta_4 N_4\tau_4 + \delta_5 N_5\tau_5 + \delta_6 N_6\tau_6.$$

Prediction regarding the local stability of the error system (12) requires at least a detailed analysis of the eigenvalues of the characteristic equation $\phi(\lambda) = 0$. The characteristic equation $\phi(\lambda) = 0$ has five zero roots and the local stability of the error system (12) is determined by the roots of the following polynomial:

$$\Gamma(\lambda) = \lambda^4 - L_1\rho\lambda^3 - L_2\rho^2\lambda^2 - L_3\rho^3\lambda + \Theta(z,\bar{\theta}) = 0.$$

In terms of the Routh–Hurwitz criterion, the necessary conditions for ensuring stability of the error system (12) is that $L_1, L_2, L_3 < 0$ and Θ is nonnegative. Since we do not have any preknowledge about the values of the estimated parameters $\bar{\beta}$, $\bar{\Lambda}$, $\bar{\alpha}$, $\bar{d}$, $\bar{\gamma}$ and $\bar{\mu}$ of Λ, α, d, γ and μ, we choose

$$N_1(z,\theta) = \frac{\tau_1(z,\theta)}{1+|\tau_1(z,\theta)|}, \qquad N_2(z,\theta) = \tau_2(z,\theta), \qquad N_3(z,\theta) = \frac{\tau_3(z,\theta)}{1+|\tau_3(z,\theta)|},$$
$$N_4(z,\theta) = \frac{\tau_4(z,\theta)}{1+|\tau_4(z,\theta)|}, \qquad N_5(z,\theta) = \frac{\tau_5(z,\theta)}{1+|\tau_5(z,\theta)|} \quad \text{and} \quad N_6(z,\theta) = \frac{\tau_6(z,\theta)}{1+|\tau_6(z,\theta)|}, \tag{14}$$

which indicates

$$\Theta(z,\theta) = \frac{\delta_1\tau_1^2(z,\theta)}{1+|\tau_1(z,\theta)|} + \delta_2\tau_2^2(z,\theta) + \frac{\delta_3\tau_3^2(z,\theta)}{1+|\tau_3(z,\theta)|} + \frac{\delta_4\tau_4^2(z,\theta)}{1+|\tau_4(z,\theta)|} + \frac{\delta_5\tau_5^2(z,\theta)}{1+|\tau_5(z,\theta)|} + \frac{\delta_6\tau_6^2(z,\theta)}{1+|\tau_6(z,\theta)|} > 0.$$

Note that the right-hand sides of all parameters ODEs vanish in the case of perfect synchronization $s = y_1$.

4. Application to tuberculosis and simulation validation

In this section, we address the problem of parameter estimation and reconstruction of tuberculosis from data of TB in the region of Center in Cameroon. Tuberculosis according to the WHO [27] is now the world's leading killer of adults; 30 million adults are expected to die from TB in the next 10 years. With the spread of HIV, coupled with deterioration of conditions in many cities, not just only in developing countries, but throughout the developed world as well, and the explosion in international travel, a resurgence of tuberculosis has occurred in Tokyo, New York, London, and other major cities [28]. In Cameroon, despite government's efforts towards providing free treatment and other cost-cutting measures, statistics from the National Comittee of Fight against Tuberculosis [29] show that the rate of tuberculosis infection is still not dropping. Cameroon has about 33 197 of estimated TB cases recorded each year, which result in 2000 deaths annually. Moreover, the total of the notified cases for TB of all forms increased from 2434 in 1980 to 24 062 in 2007. With the introduction of the direct observation therapy strategy (DOTS) in Cameroon, the situation has not improved significantly as the detection rate remains at 69%, while the treatment rate was 74% in 2006.

In the following, we perform computer simulation studies to validate the identifiability results from the theoretical analysis in the previous section. Now, in the model (1), the state variables and parameters are as follows: S, E and I stand for susceptible, latently infected (exposed to TB but not infectious) and infectious (has active TB), respectively. Susceptible individuals acquire TB infection following contact with infectious. Latently infected individuals are assumed to acquire some immunity as a result of infection, which reduces the risk of subsequent infection but does not fully prevent it. β is the effective contact rate of infectious that is sufficient to transmit infection to susceptible; α is the progression rate from latent state to infectious state; we assume that the initiation of therapeutics immediately remove individuals from active status and place them into a latent state at a rate γ; μ is the natural mortality and d the TB-induced death rate.

In the following, we perform computer simulation studies to validate the identifiability and parameter estimation results from the theoretical analysis in the previous section.

Numerical simulations have been conducted using the estimation of the TB notification rate (based on the quarterly number of TB notifications conducted in 2003, 2004, 2005, 2006 and 2007) as shown in Table 1.

We select a spline-based interpolating polynomial to approximate the selected set of data windows, since a lower-order polynomial can be more accurate than higher-order polynomials. This is because, above all if the data set $\{t_k, y(t_k)\}$, includes local abrupt changes in the values of $y(t)$ for a steady change in the value of t, then higher-order interpolating polynomial produces more oscillations around the abrupt changes.

Note that only in the ideal case with no measurement error, a rather small number of measurements is sufficient to determine the identifiable parameters uniquely. In practice, all measurements are measured with some error. Consequently, more data are needed to ensure that despite the measurement noise, we can get the estimation of the parameters with a desired accuracy. To minimize this, we add a Gaussian noise to the real data. The noise level is set to $\sigma = 1$.

The comparison of the data with the curve of the interpolation spline polynomial is shown in Fig. 1.

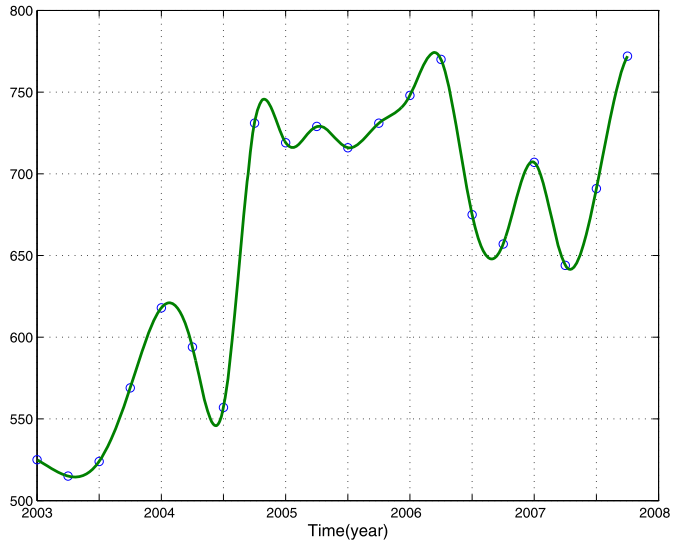

Fig. 1. The quarterly numbers of new reported TB cases and its fitted curve.

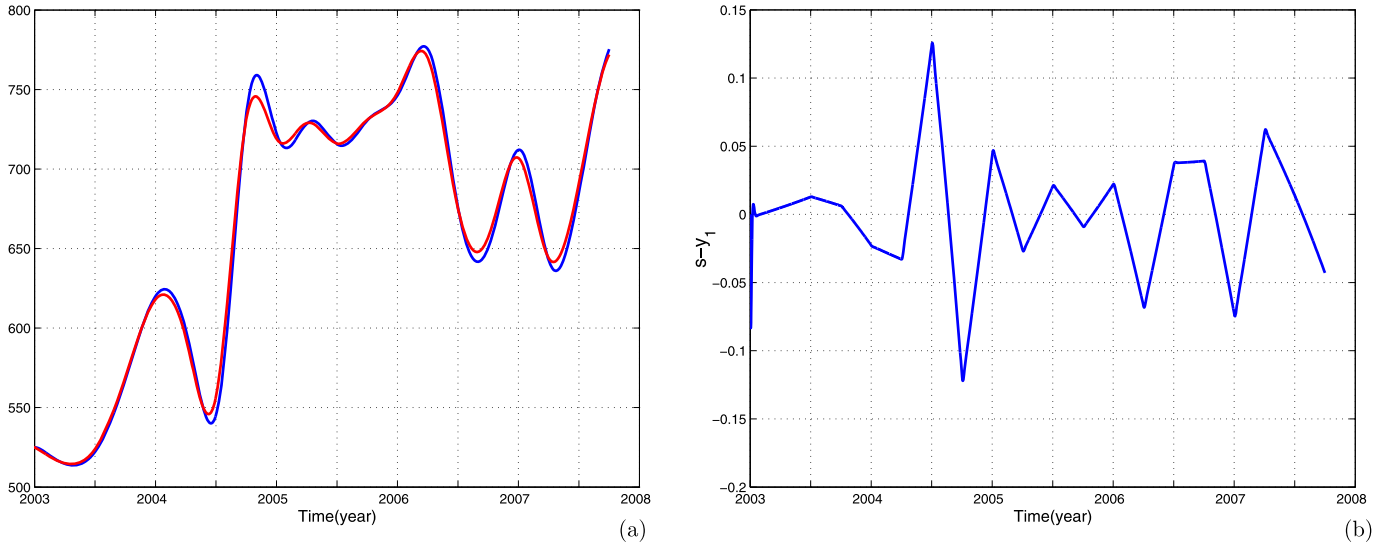

Fig. 2. (a) Synchronization of the measured signal s (red line) and the corresponding variable y_1 (blue line); (b) Synchronization error $s - y_1$ when $\sigma = 1$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this Letter.)

The values of the estimator parameters were chosen to be $L_1 = -3$, $L_2 = -3$ and $L_3 = -1$. Then, the eigenvalues of the polynomial $\lambda^3 - L_1\lambda^2 - L_2\lambda - L_3 = 0$ are located at -1. The free parameter ρ was taken as $\rho = 5$.

Fig. 2 shows the temporal evolution of the measured signal data s (red line) and the corresponding variable y_1 (blue line) of the adaptive observer (10), (11) and (14). The underlying parameter estimation process is shown in Fig. 3. As can be seen in Fig. 2, the model variable y_1 converges to the driving signal s with a small synchronization error $s - y_1$ (Fig. 2(a) and Fig. 2(b)). The fact that some of the estimated parameters do not converge to fixed values but oscillate with a small amplitude may be interpreted as an indication for a limited number of data that we used.

The parameter estimation results are reported in Table 2.

From Table 2, $\mu = 0.0170$ per year gives an average life expectancy of 59 years which is approximately the average life expectancy of Cameroonian people which was 52 year in 2008 [30]. Also, the value of $\gamma = 0.64408$ per year is not far from 0.74 per year which was the TB treatment rate in Cameroon in 2008. In addition, the TB-induced mortality $d = 0.0012$ per year is also approximately the same than the TB-induced mortality in Cameroon which is 0.008 per year.

Substituting those values of parameters into system (1), we obtain the following TB transmission model to predict the evolution of TB in the region of Center in Cameroon:

$$\begin{cases} \dot{S} = 1000.2 - 2.04SI - 0.017S, \\ \dot{E} = 2.04SI + 0.6408I - (0.017 + 0.0024)E, \\ \dot{I} = 0.0024E - (0.0170 + 0.0012 + 0.6408)I. \end{cases} \tag{15}$$

We take the first quarter of 2003 as the start time of simulation. The statistics show that the total population of the Center region of Cameroon population in 2003 is $N(0) = 2\,981\,000$ [30]. According to the National Committee of Fight against Tuberculosis in

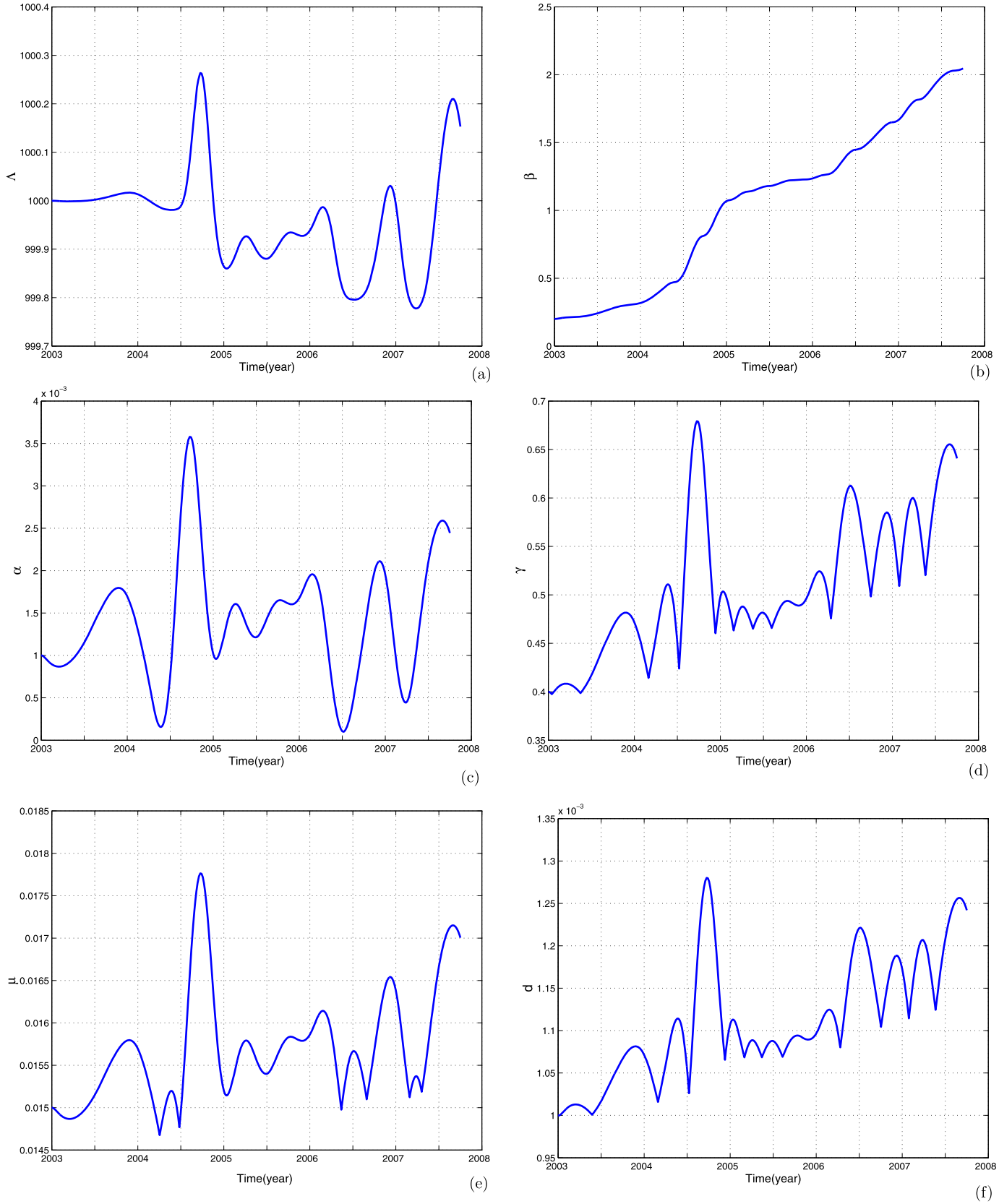

Fig. 3. Model parameter estimation when $\sigma = 1$. (a) Λ; (b) β; (c) α; (d) γ; (e) μ and (f) d.

Table 2
Estimation results for $\sigma = 1$.

Parameter	β	Λ	α	d	γ	μ
Estimated value	2.04	1000.2	0.0024	0.0012	0.6408	0.0170

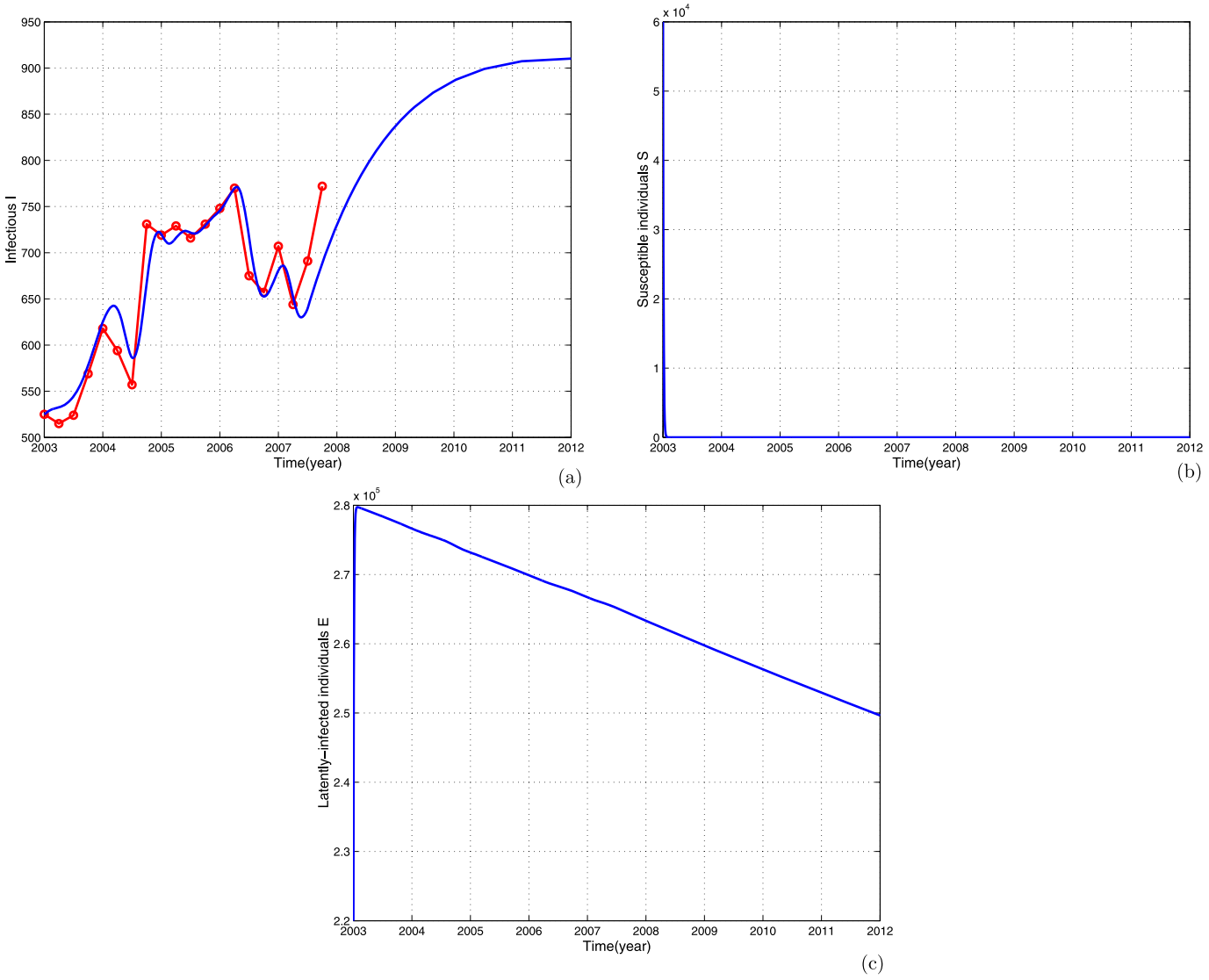

Fig. 4. The trend of the three epidemiological populations when $\sigma = 1$. (a) New TB cases reported number (red line) and infectious (blue line); (b) Susceptible individuals and (c) latently-infected individuals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this Letter.)

Cameroon [29] the number of new and relapse cases of TB was 532, then we take $I(0) = 532$. We assume that 70% of the Cameroonian population is infected with *Mycobacterium tuberculosis*. In this case $S(0) = 894\,300$. Then, one can find that $E(0) = 2\,086\,168$. Numerical simulations obtained are reported in Fig. 4. Fig. 4(a) illustrates the comparison of the quarterly reported data and the simulation of infectious in the region of Center in Cameroon. These round in the curve demonstrate the reported number of infectious, from January 2003 to December 2007. Figs. 4(b) and 4(c) give the trends of the susceptible and latently-infected individuals in the future several years, respectively.

In most epidemiological models discussed in the literature, the question of estimating unknown parameters has not been played a central role. Although the estimation of unknown parameters in an epidemiological model has been established in HIV/AIDS models using higher-order derivatives [10,11], multiple time point method and least square principle [12] to the author's knowledge, this is the first time the estimation of unknown parameters in an epidemiological model is based on tools of synchronization and has been applied to a TB model.

5. Discussions

The development of a more complex model in the context of its simpler ancestor is shown as the latest step in an on-going modelling process. This process of model validation and improvement includes structural identifiability and parameter estimation as essential components. The benefit of the larger amount of biological information contained in a more complex mechanistic model must be weighed

against its reduced tractability and appropriateness for application to data. The application of a model with a large number of state variables and parameters to data becomes more difficult both in terms of structural identifiability and, for parameter estimation, due to the contribution of noise.

In this Letter, we have used a technique from engineering [24] to investigate the algebraic identifiability of a popular three-dimensional SEI epidemiological dynamics model. We have exploited the algebraic properties of observability and identifiability that the SEI epidemiological model fills with respect to the output. We have found that all parameters in the model can be identified if only the number of infectious is measured. This fact permits us to obtain a differential parametrization to the output and its time derivative (from first to third). The differential parametrization to the output contains the information necessary for determining the remaining states and the unknown parameters.

We have used the concept of synchronization which is a promising alternative to the previously developed higher-order derivative (HOD) method [10,11] and multiple time point method (MTP) [12]. We have proposed an adaptive synchronization approach to the parameter estimation by exploiting information obtained only from real data, which is crucial for practical applications such as the prediction and control of an infectious disease. Although the parameter estimation problem has been established in HIV/AIDS models [9–12] using HOD and MTP methods, to the author's knowledge, this is the first time that such a problem has been studied using tools of synchronization. The proposed method has some advantages over the HOD and MTP methods in practical implementations. The proposed method is less computationally intensive and easier to implement and permits to identify all parameters of the model which is rather not the case of many works in the literature. In addition, the proposed adaptive observer has fast convergence properties.

Finally, to validate the identifiability analysis results, we have taken the TB as an example and we have performed simulation studies to estimate all parameters from noisy data of TB in the region of Center in Cameroon. The simulation results have confirmed our theoretical identifiability analysis. Although our analysis has been applied on a 3-dimensional TB dynamics model, the basic ideas and the proposed methodologies are generally applicable to any nonlinear epidemiological model. Our results, if complete, are however only a first step in this direction.

Acknowledgement

The authors would like to acknowledge the support of the Alexander von Humboldt Foundation, Germany.

References

[1] R.M. Anderson, R.M. May, Infectious Diseases of Humans: Dynamics and Control, Oxford University Press, Oxford, 1991.
[2] F. Brauer, C. Castillo-Chavez, Mathematical Models in Population Biology and Epidemiology, Springer, New York, 2001.
[3] O. Diekman, J.A.P. Heesterbeek, Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation, John Wiley & Sons, London, 2000.
[4] D.J. Earn, P. Rohani, B.M. Bolker, B.T. Grenfell, Science 237 (2000) 667.
[5] H.S. Hethcote, SIAM Rev. 42 (2000) 599.
[6] H. Wu, A.A. Ding, Biometrics 55 (1999) 410.
[7] Y. Huang, H. Wu, J. Appl. Stat. 33 (2006) 155.
[8] Y. Huang, D. Liu, H. Wu, Biometrics 62 (2006) 413.
[9] X. Xia, Automatica 39 (1999) 1983.
[10] X. Xia, C.H. Moog, IEEE Trans. Autom. Contr. 48 (2003) 330.
[11] A.M. Jeffrey, X. Xia, Identifiability of HIV/AIDS model, in: W.Y. Tan, H. Wu (Eds.), Deterministics and Stochastic Models of AIDS Epidemics and HIV Infections with Intervention, World Scientific, Singapore, 2005, p. 255.
[12] H. Wu, H. Zhu, H. Miao, A.S. Perelson, Bul. Math. Biol. 70 (2008) 7855.
[13] L.M. Pecora, T.L. Caroll, Phys. Rev. Lett. 64 (1990) 821.
[14] S. Bocalleti, J. Kurths, G. Osipov, D.L. Valladares, C.S. Zhou, Phys. Rep. 366 (2002) 1.
[15] A. Pikovsky, M. Rosemblum, J. Kurths, Synchronization: A Universal Concept in Nonlinear Sciences, Cambridge University Press, Cambridge, 2001.
[16] H. Nijmeijer, M.Y. Mareels, IEEE Trans. Circuits Syst. I: Fundam. Theory Appl. 44 (1997) 882.
[17] D. Yu, U. Parlitz, Phys. Rev. E 77 (2008) 066221.
[18] U. Parlitz, Phys. Rev. Lett. 76 (1996) 1232.
[19] S. Bowong, F.M. Moukam, C. Tchawoua, Phys. Rev. E 70 (2004) 066217.
[20] V. Hutson, K. Schmitt, Math. Biosci. 111 (1992) 1.
[21] R. Bellman, K.J. Astrom, 7 (1970) 329.
[22] E. Walter, Identifiability of State Space Models, Springer, Berlin, 1982.
[23] K.R. Godfrey, J.J. DiStefano III, Identifiability of model parameters, in: E. Walter (Ed.), Identifiability of Parametric Models, Pergamon, Oxford, 1987, p. 1.
[24] M. Fliess, H. Sira-Ramirez, ESAIM 9 (2003) 151.
[25] G. Conte, C.H. Moog, A.M. Perdon, Nonlinear Control Systems: An Algebraic Setting, Springer, London, 1999.
[26] S.T. Glad, in: European Control Conference, ECC97, Brussels, 1997.
[27] World Health Organization, Global Tuberculosis Control: Surveillance, Planning, Financing, World Health Organization, Geneva, Switzerland, 2009.
[28] T. Frieden, R.C. Driver, Tuberculosis 83 (2003) 82.
[29] National Committee of Fight Against Tuberculosis, Guide de personnel de la santé, Cameroon, 2009.
[30] National Institute of Statistics, Evolution des systèmes statistiques nationaux, Cameroon, 2009.