PREDICTING THE PROGRESS AND THE PEAK OF AN EPIDEMIC

Branko Ristic — DSTO, Australia, branko.ristic@dsto.defence.gov.au

Alex Skvortsov — DSTO, Australia, alex.skvortsov@dsto.defence.gov.au

Mark Morelande — The University of Melbourne, Australia, m.morelande@ee.unimelb.edu.au

ABSTRACT

The problem is statistical prediction of the number of people that will be infected with a contagious illness in a closed population over time. The prediction is based on the Susceptible-Infectious-Recovered (SIR) model of epidemic dynamics with inhomogeneous population mixing. The paper presents a theoretical analysis of the predictive accuracy based on the Cramér-Rao lower bound (CRLB). The CRLB provides a tool that enables us to quantify the prediction accuracy of a scale of an epidemic as a function of the prior uncertainty of SIR model parameters, measurement accuracy of the number of infected people and the amount of data available for processing. A verification of the theoretical analysis is carried out by Monte Carlo simulations.

Index Terms— Epidemic model, epidemic prediction, mathematical biology, importance sampling, Cramér-Rao bound, nonlinear filtering.

1. INTRODUCTION

An epidemic is a chain reaction of disease spread within a population. From the early days of humanity diseases have been the source of fear and superstition, infecting some individuals and sparing the others. Over the past hundred years mathematics has been used to understand and predict the spread of diseases. The progress of an epidemic in many cases can be described by mathematical models that involve only a few parameters. One class of such models, called compartmental models, is based on a premise that the population can be subdivided into sets of distinct classes in relation to the disease. The common classes (or compartments) in the SIR dynamic model are: susceptible (S), infectious (I) and recovered (R). Susceptible individuals have never come into contact with the disease. They are able to catch the disease and thus to move to the I compartment. Eventually the infectious individuals recover and thus move into the R compartment. Assuming that (1) the population size is fixed (no births or deaths), and (2) the population mixing is homogeneous (each susceptible person is equally likely to become the next victim), the dynamics of the compartments can be simply expressed by the Kermack-McKendick model [1]. Looking at the number of infected people over time in a large population with no prior exposure to the disease, one can observe three main phases. First is the establishment phase, where the number in compartment I grows very slowly. Second is the exponential-like growth of the compartment I. Finally in the endemicity phase, when the significant proportion of population is immune, the incidence of new infections tends to decline. When the susceptible population is exhausted, the primary epidemic is over.

Recently we have witnessed an increased interest in the prediction of epidemic spreads (due to bioterrorism or emerging diseases) using the proposed mathematical models [2]. The benefits of epidemic prediction are manyfold, from the evaluation of control measures (e.g. vaccination) to forecasting the needs of affected population. Kao [3] used the spatial spread models to predict the epidemic of "foot and mouth diseases" in the UK in 2001. Anderson *et al* [4] analysed the the data from the SARS epidemic (Hong Kong, 2003), in order to estimate the basic reproductive ratio for this disease.

In this paper we adopt the power-law scaling for inhomogeneous population mixing in the temporal SIR epidemic model [5]. This approach has been found to result in better epidemic modelling in big cities. Using the counts of infected (and possibly recovered) people as measurements, the goal is to predict the temporal progress of an epidemic, that is how many people will be infected with a contagious illness over time. The most important aspect of the prediction is the epidemic peak: its timing and its size. There are five unknown parameters in the power-scaled SIR model. While common techniques for epidemic prediction require the model parameters to be known precisely, we estimate the parameters from the available data, assuming only some uncertain prior knowledge of their values. Our approach is based on the Cramér-Rao lower bound (CRLB), which enables us to quantify the prediction accuracy of a size of an epidemic as a function of: prior uncertainty of the scaled SIR model parameters; measurement accuracy of the number of infected/recovered people and the amount of data available for estimation.

2. SIR EPIDEMIC MODEL FOR INHOMOGENEOUS MIXING

An outbreak of an epidemic is usually far more rapid than the vital dynamics of a population. Hence we can neglect the birth-death process and migrations to state that

$$S + I + R = P \tag{1}$$

where P is the (constant) number of people in the population. For simplicity and without loss of generality we can consider a normalised system where $P = 1$, $s = S/P$, $i = I/P$ and $r = R/P$. The SIR dynamics for inhomogeneous population mixing can be expressed then by the following set of differential equations [5]:

$$\frac{ds}{dt} = -\frac{\rho}{\tau}\, i\, s^{\nu} \tag{2}$$

$$\frac{di}{dt} = \frac{\rho}{\tau}\, i\, s^{\nu} - \frac{i}{\tau} \tag{3}$$

where ρ, ν and τ are the parameters of the model, to be described in the sequel. Note that the differential equation for the number of recovered people is redundant: by differentiation of (1) we get $ds/dt + di/dt + dr/dt = 0$ and from (2) and (3) it follows that $dr/dt = i/\tau$.

978-1-4244-2354-5/09/$25.00

Parameter ρ is referred to as the *basic reproductive number*[1]. It represents the average number of people infected by a direct contact with a sick person, before his/her recovery. The epidemic spreads out only if $\rho > 1$, otherwise it gradually disappears. For most diseases parameter ρ is approximately known. For example, for smallpox $\rho \in [3, 5]$, while for the measles $\rho \in [16, 18]$. Parameter τ denotes the average number of days a sick person is infectious. For most diseases this number is also partially known. Parameter ν represents the power-law scaling and reflects the population mixing (which depends on the social structure). For homogeneous population, $\nu = 1$. In general it is very difficult (and costly) to measure ν, although it is known that in the cities this parameter typically ranges from 1.7 to 2.06. Let $p(\rho)$, $p(\tau)$ and $p(\nu)$ denote the prior probability density functions (PDFs) of ρ, τ and ν, respectively.

The prediction of epidemic dynamics needs to be carried out based on the measurements of the number of infected (and possibly the number of recovered) people. These measurements arrive typically at non-uniform sampling intervals. The initial values of i and s in (2) and (3) in general are not precisely known and will be specified based on the first measurements of infected and recovered people. Let the prior PDFs of s_0 and i_0 be $p(s_0)$ and $p(i_0)$, respectively.

Appendix A presents a verification of the SIR model against a set of experimental data.

3. PROBLEM FORMULATION

The problem of epidemic prediction will be formulated in the framework of nonlinear filtering [6]. This requires the specification of the state vector, its initial (prior) PDF, the state dynamic model and the measurement model.

Since parameters ρ, τ and ν are only partially known, we need to include them in the state vector. Thus we adopt the state of an epidemic to be defined as follows:

$$\mathbf{x} = \begin{bmatrix} i & s & \tau & \nu & \rho \end{bmatrix}^\intercal \tag{4}$$

where $^\intercal$ is the matrix transpose. Using eqs. (2) and (3), the evolution of epidemic state can be written as $\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x})$ where

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} (\rho s^\nu - 1)i/\tau & -\rho i s^\nu/\tau & 0 & 0 & 0 \end{bmatrix}^\intercal. \tag{5}$$

The nonlinear differential equation governing the evolution of the state cannot be solved in closed-form. The Euler method provides a simple approximation valid for small integration interval $T > 0$: $\mathbf{x}(t+T) \approx \mathbf{x}(t) + T\mathbf{g}(\mathbf{x}(t))$. The state-evolution in discrete-time can then be expressed as:

$$\mathbf{x}_{k+1} \approx \mathbf{f}_k(\mathbf{x}_k) \tag{6}$$

where k is the discrete-time index and transition function $\mathbf{f}_k(\mathbf{x}_k)$ is given by

$$\mathbf{f}_k(\mathbf{x}_k) = \begin{bmatrix} \mathbf{x}_k[1] + \frac{T\mathbf{x}_k[1]}{\mathbf{x}_k[3]}\left[\mathbf{x}_k[5]\mathbf{x}_k[2]^{\mathbf{x}_k[4]} - 1\right] \\ \mathbf{x}_k[2] - \frac{T\mathbf{x}_k[5]\cdot\mathbf{x}_k[1]}{\mathbf{x}_k[3]}\mathbf{x}_k[2]^{\mathbf{x}_k[4]} \\ \mathbf{x}_k[3] \\ \mathbf{x}_k[4] \\ \mathbf{x}_k[5] \end{bmatrix} \tag{7}$$

Here $\mathbf{x}_k[j]$ represents the jth component of vector $\mathbf{x}_k$.

[1] In the population dynamics literature this parameter is by convention denoted R_0. We opt for ρ in order to avoid a possible confusion with the number of recovered people R at initial time.

For a fictitious disease and city population with the following parameters $\rho = 2.6$, $\tau = 10.2$ days, $\nu = 2.06$, and initial values $i_0 = 0.002$ and $s_0 = 0.998$, the time evolution of i_k, s_k and r_k, computed using dynamic equation (6), is plotted in Fig.1. The integration time step in the implementation was set to $T = 30$ minutes. The red line in Fig.1 indicated the number of infected people, $i_k = \mathbf{x}_k[1]$. In this case the epidemic peaks after 41.17 days when about 14.15% of population is infected.

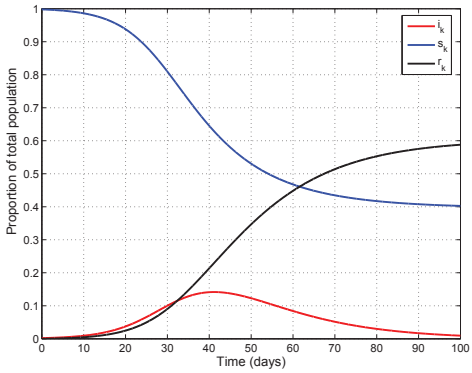

Fig. 1. SIR model: the number of infected i_k (red line), susceptible s_k (blue line) and recovered r_k people.

The measurements available for estimation/prediction are the number of infected and the number of recovered cases. The sources of these two types of measurements are assumed independent. The measurements arrive occasionally, at time instances $t_\ell > 0$, and at discrete-time ℓ are then modelled by:

$$\alpha_\ell = i_\ell + u_\ell \tag{8}$$
$$\beta'_\ell = r_\ell + v'_\ell \tag{9}$$

where u_ℓ and v'_ℓ are zero-mean white and mutually independent Gaussian, $u_\ell \sim \mathcal{N}(0, \sigma_u^2)$, $v'_\ell \sim \mathcal{N}(0, \sigma_v^2)$. Since $r_\ell = 1 - s_\ell - i_\ell$, we can introduce an equivalent measurement $\beta_\ell = (1 - \beta'_\ell)$, which according to (9) takes the form:

$$\beta_\ell = i_\ell + s_\ell + v_\ell \tag{10}$$

where $v_\ell \sim \mathcal{N}(0, \sigma_v^2)$.

The measurements vector $\mathbf{z}_\ell = \begin{bmatrix} \alpha_\ell & \beta_\ell \end{bmatrix}^\intercal$ can now be linearly related to the state vector as follows:

$$\mathbf{z}_\ell = \mathbf{H}\mathbf{x}_\ell + \mathbf{w}_\ell \tag{11}$$

with

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix}. \tag{12}$$

where $\mathbf{w}_\ell \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_\mathbf{w})$ and $\mathbf{R}_\mathbf{w} = \text{diag}[\sigma_u^2 \quad \sigma_v^2]$. The assumption here is that measurements α_ℓ and β_ℓ come from uncorrelated sources.

Let us denote the set of measurements collected up to time t_L by $\mathbf{z}_{1:L} = \{\mathbf{z}_\ell\}_{\ell=1}^L$, and the times when these measurements are collected by $\mathcal{T}_L = \{t_\ell\}_{\ell=1}^L$. In the Bayesian paradigm, the entity of interest is the PDF of the state at time t_k given $\mathbf{z}_{1:L}$, that is $p(\mathbf{x}_k|\mathbf{z}_{1:L})$. If $t_k > t_L$, this is a predictive density. We are interested in this predictive density well before the peak of $i_k = \mathbf{x}_k[1]$.

4. THEORETICAL ERROR ANALYSIS

Let us denote by $\hat{\mathbf{x}}_{k|L}$ any unbiased estimator of the state vector, based on the measurement set $\mathbf{z}_{1:L}$. The error covariance of $\hat{\mathbf{x}}_{k|L}$ is limited from below as follows [7]:

$$\mathrm{E}\left\{\left(\hat{\mathbf{x}}_{k|L}-\mathbf{x}_k\right)\left(\hat{\mathbf{x}}_{k|L}-\mathbf{x}_k\right)^{\intercal}\right\} \geq \mathbf{J}_{k|L}^{-1} \tag{13}$$

where $\mathbf{J}_{k|L}^{-1}$ is the CRLB, that is $\mathbf{J}_{k|L}$ is the information matrix defined as:

$$\mathbf{J}_{k|L} = \mathrm{E}\left\{\left[\nabla_{\mathbf{x}_k}\log p(\mathbf{x}_k,\mathbf{z}_{1:L})\right]\left[\nabla_{\mathbf{x}_k}\log p(\mathbf{x}_k,\mathbf{z}_{1:L})\right]^{\intercal}\right\}. \tag{14}$$

Here $\mathrm{E}\{\ \}$ is the expectation operator; $p(\mathbf{x}_k,\mathbf{z}_{1:L})$ is the joint probability distribution of the state $\mathbf{x}_k$ and the measurement set $\mathbf{z}_{1:L}$; $\nabla_{\mathbf{x}}$ denotes the gradient operator with respect to vector $\mathbf{x}$.

Let us introduce matrix $\mathbf{F}_k$, as the Jacobian of nonlinear function $\mathbf{f}_k$ defined in (7),

$$\mathbf{F}_k = \left[\nabla_{\mathbf{x}_k}\left[\mathbf{f}_k(\mathbf{x}_k)\right]^{\intercal}\right]^{\intercal} \tag{15}$$

The Jacobian $\mathbf{F}_k$ is state dependent and we assume here evaluated at the true state $\mathbf{x}_k$. Since the epidemic state $\mathbf{x}_k$ depends on the initial *random* state $\mathbf{x}_0$, characterised by initial PDF $p(\mathbf{x}_0) = p(i_0)p(s_0)p(\tau)p(\nu)p(\rho)$, the Jacobian is also random.

For the linear measurement equation as in (11), the information matrix $\mathbf{J}_{k|L}$, where $t_k > t_L$, can be computed using the following recursion [8, 9]:

$$\mathbf{J}_{m+1|L} = \begin{cases} (\mathbf{F}_m^{-1})^{\intercal}\,\mathbf{J}_{m|L}\,\mathbf{F}_m^{-1} + \mathbf{H}^{\intercal}\mathbf{R}_{\mathbf{w}}^{-1}\mathbf{H}, & \text{if } t_{m+1}\in\mathcal{T}_L \\ (\mathbf{F}_m^{-1})^{\intercal}\,\mathbf{J}_{m|L}\,\mathbf{F}_m^{-1} & \text{otherwise} \end{cases} \tag{16}$$

for $m = 0, 1, \ldots, k-1$. The recursion (16) starts with the initial information matrix $\mathbf{J}_{0|L}$, calculated from the prior density $p(\mathbf{x}_0)$ as [8]:

$$\mathbf{J}_{0|L} = \mathrm{E}\{-\nabla_{\mathbf{x}_k}\nabla_{\mathbf{x}_k}^{\intercal}\log p(\mathbf{x}_0)\} \tag{17}$$

Assuming Gaussian initial PDF: $p(\mathbf{x}_0) = \mathcal{N}(\bar{\mathbf{x}}_0, \mathbf{P}_0)$, it follows that $\mathbf{J}_{0|L} = \mathbf{P}_0^{-1}$.

The CRLB computed as:

$$\mathbf{C}_{m|L} = \mathrm{E}\{\mathbf{J}_{m|L}^{-1}\}, \qquad m = 0,\ldots,k \tag{18}$$

where expectation is with respect to $\mathbf{x}_0$, is referred to as the expected conditional CRLB [10].

From (15) and (7) by differentiation we get the elements of the Jacobian $\mathbf{F}_k$:

$$\begin{aligned} F_k[1,1] &= \frac{\partial \mathbf{f}_k[1]}{\partial \mathbf{x}_k[1]} = 1 + T\frac{\mathbf{x}_k[5]}{\mathbf{x}_k[3]}\mathbf{x}_k[2]^{\mathbf{x}_k[4]} - \frac{T}{\mathbf{x}_k[3]} \\ F_k[1,2] &= \frac{\partial \mathbf{f}_k[1]}{\partial \mathbf{x}_k[2]} = T\frac{\mathbf{x}_k[1]\mathbf{x}_k[4]\mathbf{x}_k[5]}{\mathbf{x}_k[3]}\mathbf{x}_k[2]^{(\mathbf{x}_k[4]-1)} \end{aligned}$$

and likewise for the remaining terms.

Let us illustrate the predictive CRLB using the same parameters as in Fig.1. The measurements $\mathbf{z}_\ell$ were taken once per day (hence $\mathcal{T} = \{1\text{day}, 2\text{days}, \ldots\}$), with $\sigma_u = \sigma_v = 0.001$. The last available measurement is taken at $t_L = 10, 20$ and 30 days, and we compute for each case the CRLB for the period from zero time to $t_k = 100$ days. The initial covariance was set to $\mathbf{P}_0 = \mathrm{diag}[\sigma_u^2,\ \sigma_u^2+\sigma_v^2,\ \sigma_\rho^2,\ \sigma_\tau^2,\ \sigma_\nu^2]$, with $\sigma_\rho = 0.5$, $\sigma_\tau = 1.0$ and $\sigma_\nu = 0.5$. Fig.2 demonstrates the achievable predictive accuracy for the number of infected people i_k for: (a) $t_L = 10$ days, (b) $t_L = 20$ days and (c) $t_L = 30$ days. The red line is in all cases identical with the red line in Fig.1, it shows the true i_k. The blue dashed lines show the upper and lower 1σ limits, that is $i_k \pm \sqrt{\mathbf{C}_{k|L}[1,1]}$, for the case where both α_ℓ and β_ℓ measurements are available, see (8) and (10). Finally, the green dashed lines are indicating the same limits when only α_ℓ measurements are available.

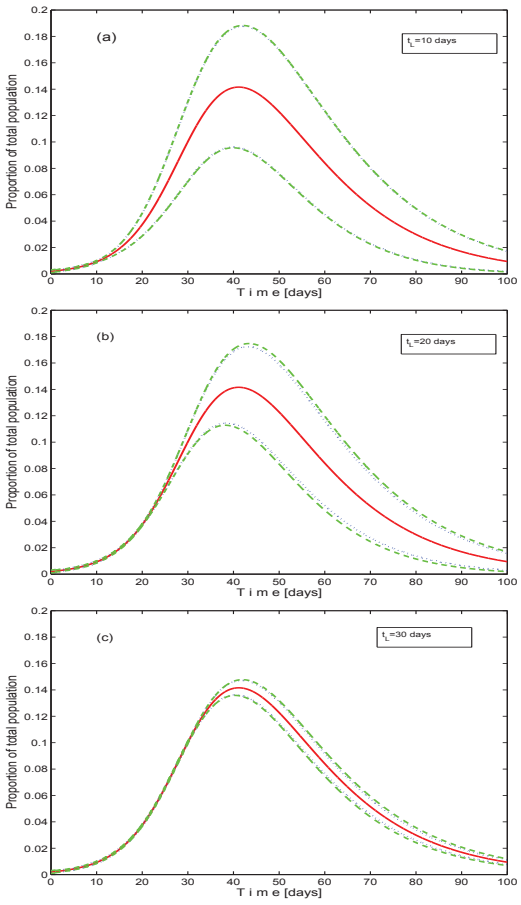

Fig. 2. The predictive accuracy based on the CRLB for: (a) $t_L = 10$ days; (b) $t_L = 20$ days; (c) $t_L = 30$ days

Observe from Fig.2 how the prediction uncertainty reduces as we increase t_L. If we adopt the differences in the timing and the size (maximum value) of the peaks of $i_k \pm \sqrt{\mathbf{C}_{k|L}[1,1]}$ curves to represent the measure of the timing and size 1σ uncertainty of the epidemic prediction, then from Fig.2 we can make the following observations. First, the lack of measurements β_ℓ has a very small impact on the accuracy in the prediction of the number of infected people (increases uncertainty only slightly). Second, the prediction of the timing of the peak of an epidemic appears to be more accurate than the prediction of the size (the maximum value of the number of infected) of the epidemic.

Note that based on the CRLB we can (similarly to Fig.2) quantify the epidemic prediction performance as a function of prior SIR parameter uncertainty (represented here by matrix $\mathbf{P}_0$) or the measurement accuracy (variances σ_u^2 and σ_v^2). Due to the limited space these results are not shown here.

In order to verify the CRLB based analysis we have performed a set of Monte Carlo runs and compared the obtained averaged error performance with the theoretical performance of the CRLB.

The SIR model parameters, the prior PDF and the characteristics of measurements α_ℓ, were the same as we described earlier (measurements β_ℓ not used). The initial values of i_0 and s_0 were based on the first measurement, assuming (correctly) that $r_0 = 0$, i.e. $\hat{i}_0 = \alpha_1$ and $\hat{s}_0 = 1 - \alpha_1$. The estimation of parameters τ, ν and ρ was carried out using a Bayesian importance sampling technique known as *progressive correction* (for details see [11]). Using only $t_L = 20$ days of measurements (i.e. a set $\alpha_{1:20}$), we obtain estimates $\hat{\tau}$, $\hat{\nu}$ and $\hat{\rho}$ which are then used with $\hat{i}_0$ and $\hat{s}_0$ in (6) to compute $\hat{i}_{m|L}$ for $m = 1, \ldots, 100$.

Let $\hat{i}^{(c)}_{m|L}$ denote a result obtained in Monte Carlo run $c = 1, \ldots, C$, then the RMS error is $\sqrt{\frac{1}{C}\sum_{c=1}^{C}(\hat{i}^{(c)}_{m|L} - i_k)^2}$, where i_k is the truth. Fig.3 shows the empirical RMS error obtained from $C = 100$ runs (blue solid line) and the theoretical square-root of the CRLB, i.e. $\sqrt{C_{m|L}[1,1]}$ (dashed red line). The agreement is fairly remarkable and confirms the theoretical analysis presented above.

5. SUMMARY

The paper presented a theoretical analysis of temporal predictability of epidemic evolution using the power-law scaled SIR dynamic model and the Cramér-Rao bound as a an error performance tool. For a given uncertainty of the SIR model parameters, measurement accuracy of infected people and the number of measurements available for estimation, we can determine theoretically the temporal spreading, the size and the timing of a peak of an epidemic. Monte Carlo simulations, using the progressive correction algorithm for SIR model parameter estimation, confirm the validity of theoretical analysis. The implication of this work goes beyond epidemiology as the framework is applicable to various population models, models for resource competition, terrorist cell spreading, etc.

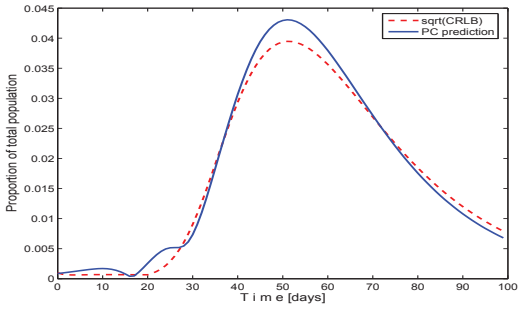

Fig. 3. Theoretical CRLB versus the empirical RMS error for the epidemic prediction based on $L = 20$ measurements

A. SIR MODEL AGAINST EXPERIMENTAL DATA

Experimental data were obtained using CROWD, an agent based population model [12] of a virtual town of 5000 inhabitants, created in accordance with the Australian Census Bureau data. The CROWD includes typical age/sex breakdown and family-household-workplace habits with realistic day-to-day contacts for a disease spread. The blue line in Fig.4 shows the number of people of this town infected by a fictitious disease, reported once per day during a period of 154 days. The dashed red line represents the power-law scaled SIR model fit with parameters $\hat{\nu} = 1.2042$, $\hat{\tau} = 9.3825$ and $\hat{\rho} = 2.2932$. These estimates were obtained using the progressive correction algorithm, applied to all data points.

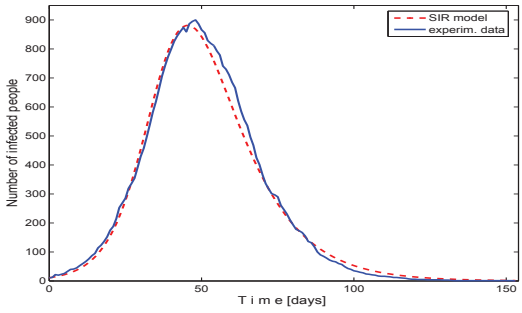

Fig. 4. Measured number of infected I_k (experimental data obtained using CROWD) versus the power-scaled SIR model fit

B. REFERENCES

[1] R. M. Anderson and R. M. May, "Population biology of infectious diseases: Part 1," *Nature*, vol. 280, pp. 361–367, 1979.

[2] B. D. Elderd, V. M. Dukic, and G. Dwyer, "Uncertainty in predictions of disease spread and public health responses to bioterrorism and emerging diseases," *PNAS*, vol. 103, no. 42, pp. 15693–15697, Oct. 2006.

[3] R. R. Kao, "The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK," *Trends Microbiol*, vol. 10, pp. 279–286, 2002.

[4] R. M. Anderson, C. Fraser, and A. C. Ghani, "Epidemiology, transmission dynamics and control of SARS: the 2002-2003 epidemic," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 359, pp. 1091–1105, 2004.

[5] P. D. Stroud, S. J. Sydoriak, J. M. Riese, J. P. Smith, S. M. Mniszewski, and P. R. Romero, "Semi-empirical power-law scaling of new infection rate to model epidemic dynamics with inhomogeneous mixing," *Mathematical biosciences*, vol. 203, pp. 301–318, 2006.

[6] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, 1979.

[7] H. L. Van Trees, *Detection, Estimation and Modulation Theory (Part I)*, John Wiley & Sons, 1968.

[8] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramér-Rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Proc.*, vol. 46, no. 5, pp. 1386–1396, 1998.

[9] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter*, Artech House, 2004.

[10] H. Van Trees and K. L. Bell, Eds., *Bayesian bounds*, IEEE Press, 2007.

[11] N. Oudjane and C. Musso, "Progressive correction for regularized particle filters," in *Proc. 3rd Int. Conf. Information Fusion*, Paris, France, 2000.

[12] A. Skvortsov, R. Connell, P. Dawson, and R. Gailis, "Epidemic spread modeling: Alignment of agent-based simulation with a simple mathematical model," in *Proc. Int. Conf. Bioinformatics & Comput. Biology*, Las Vegas, USA, June 2007, pp. 487–890, CSREA Press.