On Generalized Cross Validation for Stable Parameter Selection in Disease Models[1]

A. Smirnova
E-mail: asmirnova@gsu.edu
Phone: (404) 413-6409
Dept of Mathematics and Statistics
Georgia State University
Atlanta, GA 30303, USA

M. Martcheva
E-mail: maia@ufl.edu
Phone: (352) 294-2319
Department of Mathematics
University of Florida
Gainesville, FL 32611

H. Liu
E-mail: hliu16@student.gsu.edu
Phone: (404) 413-6400
Dept of Mathematics and Statistics
Georgia State University
Atlanta, GA 30303, USA

Abstract. In this paper we study advantages and limitations of the Generalized Cross Validation (GCV) approach for selecting a regularization parameter in the case of a partially stochastic linear irregular operator equation. The research has been motivated by an inverse problem in epidemiology, where the goal was to reconstruct a time dependent treatment recovery rate for *Plasmodium falciparum*, the most dangerous form of malaria. Initial numerical simulations gave rise to a theoretical analysis of the expected value of the GCV function and the efficiency of the GCV method for different noise levels. It was shown that, as opposed to L-curve, the GCV does not necessarily generate a systematic error in the value of the regularization parameter for Tikhonov's stabilizing algorithm.

Keywords: generalized cross-validation, ill-posed problems, regularization parameter selection.
Mathematics Subject Classification: 47A52, 65F22.

1. The model of P. *falciparum* Malaria in India

Malaria is an infectious disease affecting humans and mosquitos. The virus cannot be transmitted directly between mosquitos, which means that a mosquito can only acquire the disease through an infected human's blood, and once a mosquito gets infected it never recovers. Therefore, the dynamic of the mosquito population is given by the following graph [MH10]

$$X \to Y,$$

indicating that mosquitos can only move from a susceptible group to the infected one. In the model, we assume that the susceptible mosquito population size, $X(t)$, along with infected mosquito population size, $Y(t)$, are described by the differential equations,

$$X'(t) = \Lambda - rX(t)I(t) - dX(t), \tag{1.1}$$

$$Y'(t) = rX(t)I(t) - dY(t). \tag{1.2}$$

[1]This work is supported by NSF under grant (DMS-1112897)

The dynamics of human population results in a more complicated model than (1.1)-(1.2) and can be outlined as

$$S \to C \to I \to R \to S.$$

It takes into consideration the fact that malaria cannot be transmitted from human to human, which means that humans can only receive the disease from infected mosquito's bites. The difference between humans and mosquitos is that humans can actually recover from the disease, and reinfection of recovered humans is common in malaria. Moreover, human infection has an additional stage, called symptomatic, which is $C(t)$ in the above graph. That is, instead of directly moving to the infected class, a susceptible human will move to the symptomatic class before getting infected. In the model, $S(t)$, $C(t)$, $I(t)$, and $R(t)$ are described by the equations below,

$$S'(t) = \lambda(t) - \beta S(t)Y(t) - \mu S(t) + \omega R(t), \tag{1.3}$$

$$C'(t) = \beta(S(t) + R(t))Y(t) - (\mu + \nu)C(t), \tag{1.4}$$

$$I'(t) = \nu C(t) - (\mu + \gamma(t))I(t), \tag{1.5}$$

$$R'(t) = \gamma(t)I(t) - \beta R(t)Y(t) - (\mu + \omega)R(t). \tag{1.6}$$

The information on parameters and variables for humans and mosquitos is given in Table 1.

Table 1. Parameters and variables

Parameters	**Interpretation**
Λ	mosquito birth rate
r	transmission coefficient human $\to$ mosquito
d	mosquito death rate
M	total mosquito population size
μ	human natural death rate
β	transmission coefficient mosquito $\to$ human
ν	symptomatic stage recovery rate
ω	recovered $\to$ susceptible recruitment rate
Variables	**Interpretation**
$X(t)$	susceptible mosquito population size
$Y(t)$	infected mosquito population size
$I(t)$	infected human population size
$P(t)$	total human population size
$S(t)$	susceptible human population size
$C(t)$	population size with clinical symptoms
$R(t)$	recovered human population size
$y(t)$	proportion of infected mosquitos
$\gamma(t)$	human treatment-recovery rate
$\lambda(t)$	human birth rate

The most difficult parameter to estimate, appears to be the treatment recovery rate $\gamma(t)$ [MH10]. Since it was not feasible to fit the data with a constant γ, even for the early period of 1984-1997, a variable recovery rate $\gamma(t)$ has been adopted in [MH10] and set to be proportional to the total expenditures for the malaria control, which was, in turn, assumed to be proportional to the total human population size, i.e., $\gamma(t) = \lambda P(t)$. With the use of NonlinearModelFit, the proportionality coefficient λ has been estimated as $\lambda = 0.024$ person/year. This value of the parameter results in a remarkable agreement between the reported number of clinical cases and the symptomatic curve

$C(t)$ generated by the model for the period from 1984 to 1997. However, after 1997 the model predicts a significantly larger number of cases than the number actually observed. A possible explanation for the discrepancy could be the 1998 launch of a Roll Back Malaria (RBM) initiative aimed at reducing global malaria cases by half from 2000 till 2010, and by three quoters in 2015. With new financial and logistics support in place, the treatment recovery rate $\gamma(t)$ has changed since 1998. So, with the initial date being 1984, the coefficient λ had to be increased 13.5 years into the model in order to account for the enhanced malaria treatment after the start of RBM. To that end, a step function in the form $cP(t)H(t-13.5)$ was added to $\lambda P(t)$. However, this did not seem to work and eventually the symptomatic data has been fitted with the addition of scaled step functions to the mosquito death rate, d, and to the scaled mosquito-human transmission coefficient, β, [MH10].

In our study we propose to use the inverse problem approach in order to reconstruct the time-dependent treatment recovery rate function $\gamma(t)$ and to compare our estimated $\gamma(t)$ with the one obtained in [MH10]. The algorithm presented below includes two linear inverse subproblems: differentiation of discrete noisy data as well as computation of the second derivative of the discrete noisy data. These subproblems are unstable in the sense that even very small inaccuracies in the reported data can result in large errors in the computed derivatives. For this reason, one has to incorporate the technique known as regularization, which allows to strike a proper balance between accuracy and stability and, therefore, to minimize the computational error. One of the main challenges in the implementation of any regularization scheme is to choose a "near optimal" value of the regularization parameter responsible for striking this balance.

2. **Numerical Procedure and Computational Preliminaries**

Introduce the notation, $T(t) := S(t) + R(t)$, for the class combining susceptible and recovered humans. The initial value for $Y(t)$, the number of infected mosquitos, can hardly be obtained. However, the initial value for the proportion of infected mosquitos has been estimated in [MH10]. Thus, we denote $y(t) := Y(t)/(X(t)+Y(t))$, $b := \beta(X(t)+Y(t))$, and arrive at the following system

$$\begin{cases} P = T + I + C \\ y' = r(1-y)I - dy \\ T' = \lambda(t) - bTy - \mu T + \gamma(t)I \\ C' = bTy - (\mu+\nu)C \end{cases} \tag{2.1}$$

with $P(t)$ being the total human population. Our goal is to find $\gamma(t)$, the human treatment-recovery rate, given estimated values of λ, r, b, μ, ν [MH10], as well as the data available for $C(t)$ and $P(t)$. From (2.1), one derives

$$\gamma(t) = \frac{by^2[-\lambda + C' + (\mu+\nu)C] + y(C'' + 2\mu C' + \nu C' + \mu^2 C + \mu\nu C) - y'[C' + (\mu+\nu)C]}{by^2(P-C) - y[C' + (\mu+\nu)C]}, \tag{2.2}$$

where $y(t)$ satisfies the differential equation

$$y' = E(t)y + B(t) + \frac{D(t)}{y}, \tag{2.3}$$

$$E := rC - rP - d, \quad B := r(P-C) + \frac{r}{b}[C' + (\mu+\nu)C], \quad D := -\frac{r}{b}[C' + (\mu+\nu)C]. \tag{2.4}$$

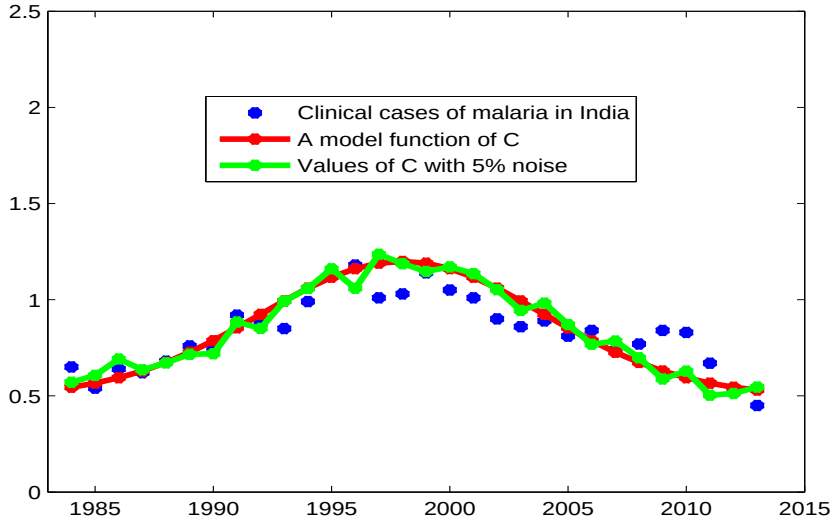

Figure 1. Clinical cases, $C(t)$, of P. falciparum malaria in India and the simulated data

This nonlinear ODE can be solved numerically in a stable manner. However, computation of $C'(t)$ from discrete experimental data for $C(t)$ is unstable and has to be regularized. Moreover, once $y(t)$ has been evaluated from (2.3), one will need to calculate $\gamma(t)$ using (2.2). To that end, one has to find $C''(t)$, given discrete experimental data for $C(t)$. This step is also unstable.

Table 2. Clinical cases, $C(t)$, of P. falciparum malaria in India in 1984-2013 [MS]

Year	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
Incidence (million)	0.65	0.54	0.64	0.62	0.68	0.76	0.75	0.92	0.88	0.85
Year	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Incidence (million)	0.99	1.14	1.18	1.01	1.03	1.14	1.05	1.01	0.90	0.86
Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Incidence (million)	0.89	0.81	0.84	0.74	0.77	0.84	0.78	0.67	0.53	0.45

The data on clinical cases, $C(t)$, of P. *falciparum* malaria in India over the period 1984-2013 [MS] is given in Table 2. In order to test our solver for stable numerical differentiation, we begin numerical analysis with experiments for simulated data using the following Gaussian function

$$C(t) = \tilde{a}e^{-\frac{(t-\tilde{b})^2}{2\tilde{c}^2}} + \tilde{d}, \tag{2.5}$$

with $\tilde{a} = 0.7, \tilde{b} = 15, \tilde{c} = 6$, and $\tilde{d}$ computed from the initial condition $C(t_0) = 0.53$. This function allows to mimic the real data on clinical cases listed in Table 2. Since the real data is reported yearly, we take discrete values of model function (2.5), add normally distributed random noise to these discrete values, and then spline. Noise in the data can be due to a sampling method, corrupt or incomplete reporting, or any other possible reason. The graphs of the real clinical data, simulated data (2.5), and the noisy simulated data are shown in Figure 1. The first and second

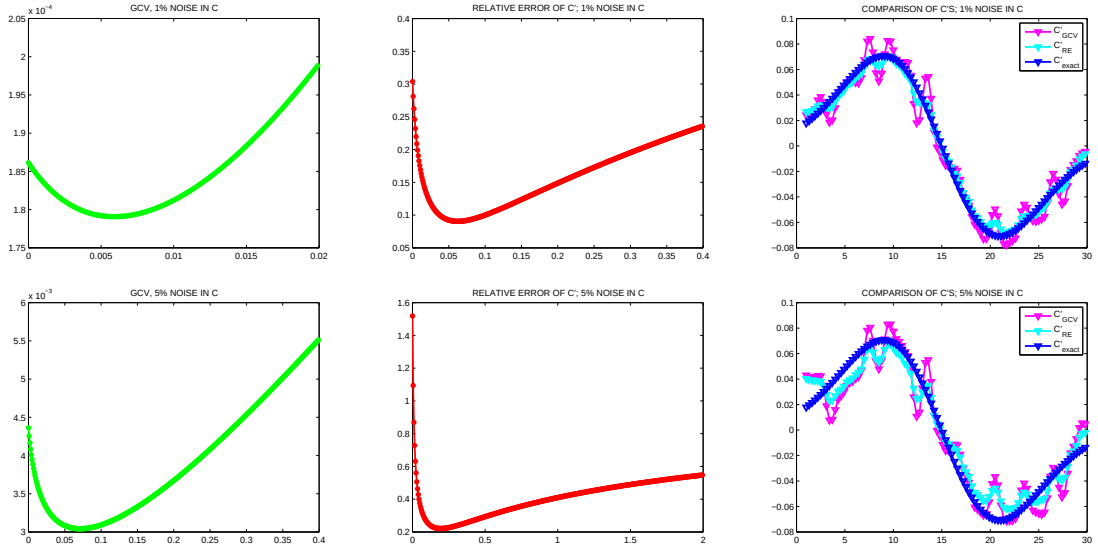

Figure 2. GCV curves, relative error curves, and comparison of the derivatives, $C'(t)$

derivatives of simulated data (2.5) can be evaluated analytically:

$$C'(t) = -\frac{\tilde{a}(t-\tilde{b})}{\tilde{c}^2} e^{-\frac{(t-\tilde{b})^2}{2\tilde{c}^2}}, \quad C''(t) = \frac{\tilde{a}}{\tilde{c}^2}\left[\frac{(t-\tilde{b})^2}{\tilde{c}^2} - 1\right] e^{-\frac{(t-\tilde{b})^2}{2\tilde{c}^2}}. \tag{2.6}$$

Given these analytic expressions, we can directly compare our numerical solutions with the exact derivatives to assess the efficiency of the algorithm for estimating C' and C''. While a regularizing operator, R_α, for the computation of C'_α from the discrete noisy data C_σ can be generated by a number of algorithms, the selection of a regularization parameter, α, constitutes a significant difficulty. Indeed, as it has been established by A. Bakushinsky in [B84], in order to construct a convergent stabilizing strategy in an infinite dimensional case one needs to use the information on the noise level in the data (apart from the data itself). However, for the data on clinical P. *falciparum* malaria cases, C_σ, the noise level cannot be measured. Therefore, one has no alternative but to resort to a heuristic parameter selection method, like L-curve, quasi-optimality and quotient criteria, or generalized cross validation (GCV) [BL11, H98, HO93, H92, LY97, RR13, V02, W90, L78]. Note that quasi-optimality and quotient criteria for the parameter selection *are* justified in case of finite-dimensional linear equations [L78].

The cross validation is based on successfully leaving out elements $C_\sigma^{(i)}$ of the vector C_σ; computing a regularized estimate of the unknown function C' from the reduced data set; and predicting the left out coordinate $C_\sigma^{(i)}$ with the estimated value of C'. The regularization parameter is then chosen as a minimizer of the overall prediction error. The GCV function is obtained from the regular cross validation by a substitution that makes it invariant with respect to orthogonal transformations of the response vector C_σ. In our initial experiments we write the numerical differentiation problem in the form of Volterra's integral equation of the first kind and use a finite-dimensional version of the Tikhonov stabilizing algorithm [TA77, TGSY95, TLY98] in L_2 combined with GCV to reconstruct C' from (2.5) for different levels of discrete white noise. The discrete approximation in the solution space, $\mathcal{X}$, is carried out by means of cubic B-spline functions, and the corresponding regularized

problem is solved for the unknown vector of the spline expansion coefficients. We take m, the number of the spline functions, to be 100, while n, the size of the data space, is 30 (the data is being collected over a 30-year period, see Figure 1 and Table 2). This gives us an undetermined $m \times n$ linear system to handle by Tikhonov's algorithm. Since this is the case of simulated data, we compare the GCV regularization parameter, α_{GCV}, with the optimal one, α_{RE}, which minimizes the relative error. The results are summarized in Figure 2 and Table 3. One can see that an overall performance of GCV is satisfactory despite its tendency to under-regularize the computed solution, which is typical for a small sample [LDA12]. One may also observe that GCV does not capture "little" noise and returns $\alpha = 0$, while in fact the regularization *is* needed. This behavior will be explained in our theoretical analysis below.

Table 3. Experimental results for various relative noise levels

Noise	α_{GCV}	α_{RE}	$\frac{\|C'_{\alpha_{GCV}} - C'\|}{\|\hat{C}'\|}$	$\frac{\|C'_{\alpha_{RE}} - C'\|}{\|\hat{C}'\|}$
0.5%	0	3.8×10^{-2}	0.7686	6.12×10^{-2}
1%	6.00×10^{-3}	6.20×10^{-2}	2.09×10^{-1}	9.05×10^{-2}
3%	3.90×10^{-2}	1.34×10^{-1}	2.65×10^{-1}	1.67×10^{-1}
5%	7.10×10^{-2}	1.94×10^{-1}	3.09×10^{-1}	2.21×10^{-1}
8%	1.18×10^{-1}	2.79×10^{-1}	3.63×10^{-1}	2.83×10^{-1}
12%	1.81×10^{-1}	3.88×10^{-1}	4.21×10^{-1}	3.47×10^{-1}
15%	2.30×10^{-1}	4.69×10^{-1}	4.56×10^{-1}	3.86×10^{-1}
20%	3.16×10^{-1}	6.04×10^{-1}	5.03×10^{-1}	4.41×10^{-1}

Once the first derivative, C'_α, has been computed, we can use it as the data and evaluate C''_α in the same manner. However, it turns out that in this particular situation the GCV method does not work for the second derivative. The discussion in Section 3 sheds some light on the cases when GCV is unable to detect noise in the data even if its level is substantial. Nevertheless, the experiments for C' do give us assurance that if GCV works, the value of α_{GCV} is rather reliable.

Encouraged by that observation, we proceed to the clinical data. Yet again, the GCV fails to reconstruct the derivative (by showing minimum at $\alpha = 0$), see Figure 3. To overcome the failure of generalized cross validation, we opt to consider two partial data sets. Specifically, denote $t_1 = 1984, t_2 = 1985, ..., t_{30} = 2013$. The original data set, as illustrated in Table 2, consists of 30 points: $\{C(t_1) = 0.65, C(t_2) = 0.54, ..., C(t_{30}) = 0.45\}$. Instead of using this full set, we conduct the experiments with "odd" and "even" partial data sets, $\{C(t_1), C(t_3), ..., C(t_{29})\}$ and $\{C(t_2), C(t_4), ..., C(t_{30})\}$, separately. We can see from the graphs that for both partial data sets, the GCV curves give clear minima. The regularization parameters for "odd" and "even" partial data sets, denoted as α_O and α_E, respectively, are very close: $\alpha_O = 4.0 \times 10^{-2}$, and $\alpha_E = 5.2 \times 10^{-2}$.

The two estimates of C' created from the partial data sets, α_O and α_E, are shown in Figure 3. In order to get an approximation of C' for the full data set, one possible approach is to take the average of two partial estimates. i.e., set $C'_{Full} = (C'_{\alpha_O} + C'_{\alpha_E})/2$. Alternatively, one can find the average of α_O and α_E, let $\alpha_{Full} = (\alpha_O + \alpha_E)/2 = 4.6 \times 10^{-2}$, and then generate C'_{Full} from the full data set using α_{Full} as the regularization parameter. The comparison of these two approaches is also shown in Figure 3. The first approach makes more sense because α is a function of noise and the average of two α's is not necessarily reflective of the total noise in the full data set.

The next step is to find C'' by using C'_{Full} as the new data set and evaluating α by the GCV method. Unable to succeed with the full set, we construct two approximations in the same manner

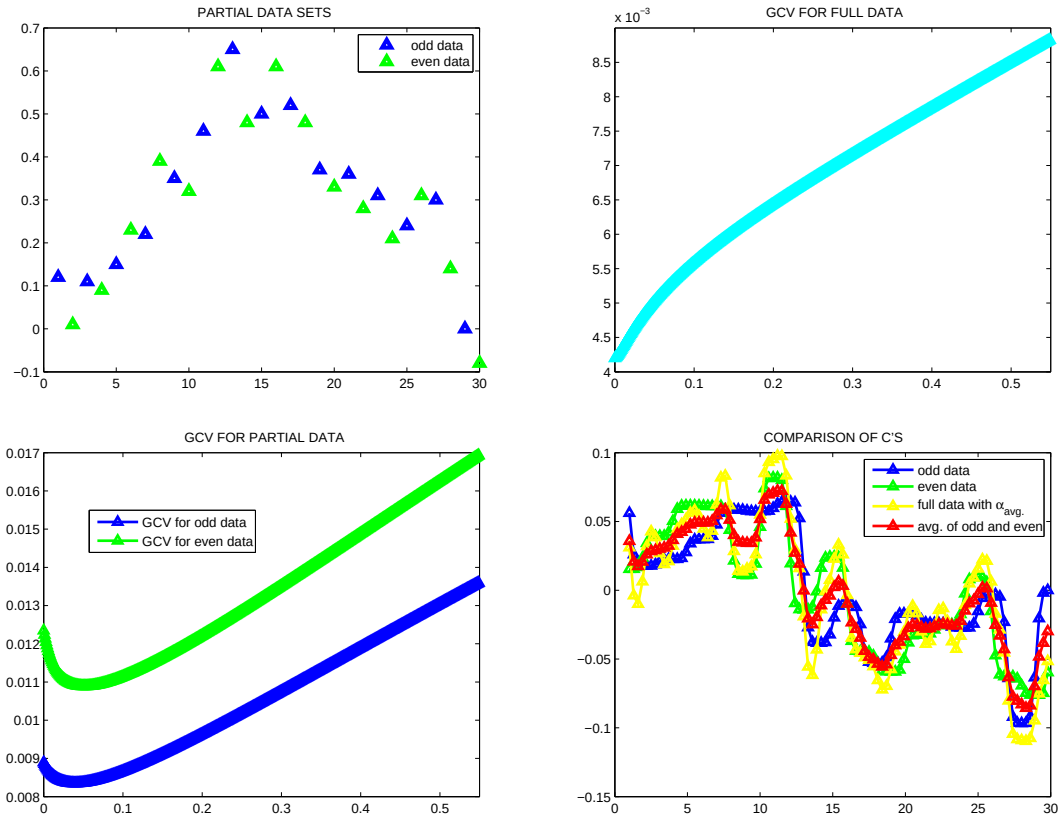

Figure 3. Reconstruction of the derivative, $C'(t)$, for real data

as above. One is the average of two partial estimates, that is, $C''_{Full} = (C''_{\alpha_O} + C''_{\alpha_E})/2$, where the two parameters $\alpha_O = 7.5 \times 10^{-2}$ and $\alpha_E = 1.3 \times 10^{-1}$. The other is the solution of the regularized problem with $\alpha_{Full} = (\alpha_O + \alpha_E)/2 = 1.025 \times 10^{-1}$. The results are shown in Figure 4.

3. **Theoretical Analysis of Generalized Cross Validation**

One can see that for this model, the GCV method is a fairly reliable indicator of the required regularization level despite of its tendency to under-smooth given a small data set. In what follows we present some theoretical analysis highlighting advantages and limitations of the GCV strategy in case of a partially stochastic linear operator equation of the first kind with a random noise in the response vector:

$$Ax + \eta = f_\sigma, \quad \mathcal{X} \to \mathbb{R}^n. \tag{3.1}$$

The true solution x in (3.1) lies in a deterministic Hilbert space $\mathcal{X}$, the operator A maps $\mathcal{X}$ into $\mathbb{R}^n$, and $f_\sigma \in \mathbb{R}^n$ is the noise contaminated right-hand side. We assume here that $\mathcal{X}$ is either infinite dimensional or its dimensionality is much larger than n. Hence, the problem is ill-posed due to instability and the lack of data. Let η be a discrete white noise, i.e., $E(\eta_i) = 0$ and $E(\eta_i \eta_j) = \sigma^2 \delta_{ij}$, for $i, j = 1, 2, ..., n$, where the variance σ^2 is fixed and independent of n, that is, $E(||\eta||^2/n) = \sigma^2$ for each n. Consider a singular system of the operator A, $\{u_i, s_i, v_i\}_{i=1}^n$, with singular values

$$s_1 \geq s_2 \geq ... \geq s_n > 0.$$

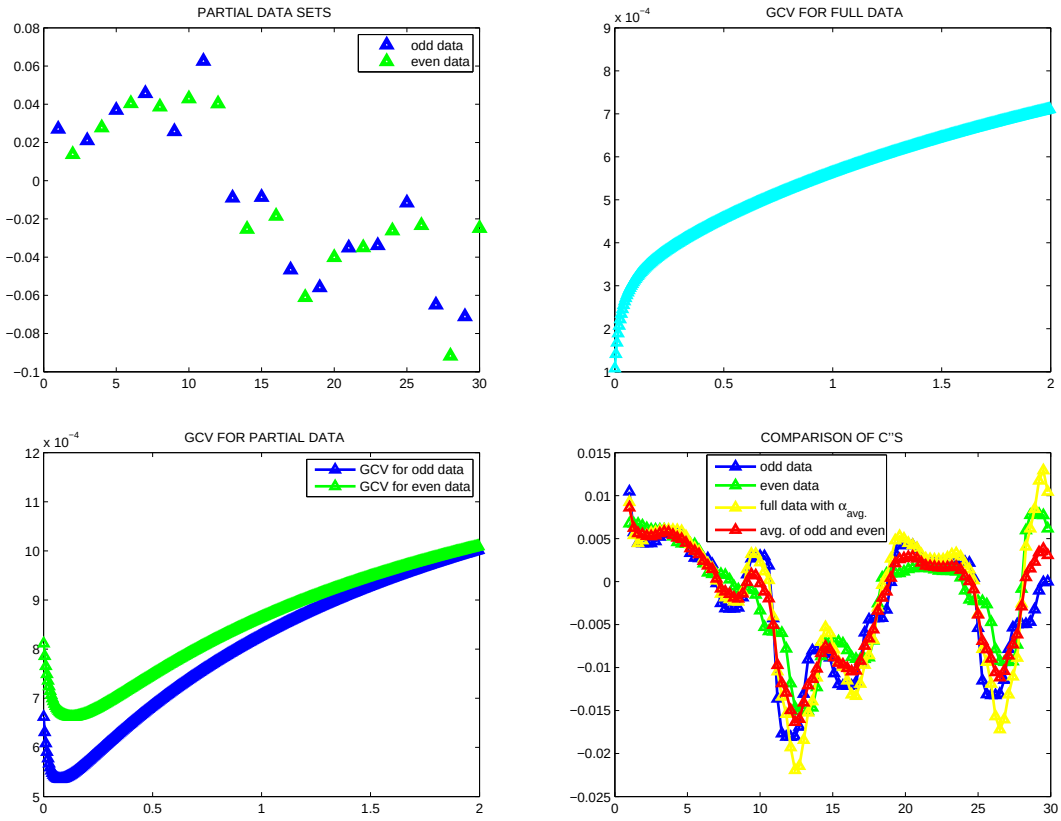

Figure 4. Reconstruction of the second derivative, $C''(t)$, for real data

Here $v_i \in \mathcal{X}$ with $(v_i, v_j)_{\mathcal{X}} = \delta_{ij}$, and $u_i \in \mathbb{R}^n$ with $(u_i, u_j)_{\mathbb{R}^n} = u_i^T u_j = \delta_{ij}$. Suppose that a regularized solution, x_α, has a linear filter representation

$$x_\alpha = \sum_{i=1}^{n} \omega_\alpha(s_i^2) \frac{(u_i, f_\sigma)}{s_i} v_i := R_\alpha f_\sigma, \tag{3.2}$$

which defines a stabilizing strategy $R_\alpha : \mathbb{R}^n \to \mathcal{X}$. Introduce the influence matrix $A_\alpha := AR_\alpha$. By (3.2), one gets [V02]

$$A_\alpha = U diag(\omega_\alpha(s_i^2)) U^T,$$

where U is the orthogonal matrix whose ith column is the left singular vector u_i of the operator A. From the above identity, it follows that

$$trace(A_\alpha) = \sum_{i=1}^{n} \omega_\alpha(s_i^2).$$

To proceed with our analysis, we rewrite

$$\begin{aligned} r_\alpha &:= Ax_\alpha - f_\sigma = AR_\alpha f_\sigma - f_\sigma = (AR_\alpha - I)(Ax + \eta) \\ &= (A_\alpha - I)Ax + (A_\alpha - I)\eta \\ &= \sum_{i=1}^{n}[\omega_\alpha(s_i^2) - 1]s_i(x, v_i)u_i + \sum_{i=1}^{n}[\omega_\alpha(s_i^2) - 1](\eta, u_i)u_i. \end{aligned}$$

Then according to Proposition 8.15 in [V02],

$$E\left(\frac{1}{n}||r_\alpha||^2\right) = \frac{1}{n}||(A_\alpha - I)Ax||^2 + \frac{\sigma^2}{n}trace[(A_\alpha - I)^2] \tag{3.3}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(\omega_\alpha(s_i^2) - 1)^2 s_i^2(x, v_i)^2 + \frac{\sigma^2}{n}\sum_{i=1}^{n}(\omega_\alpha(s_i^2) - 1)^2. \tag{3.4}$$

The expected value of the GCV function takes the form

$$EG(\alpha) := E\left(\frac{\frac{1}{n}||r_\alpha||^2}{[\frac{1}{n}trace(I - A_\alpha)]^2}\right) = \frac{E(\frac{1}{n}||r_\alpha||^2)}{[1 - \frac{1}{n}traceA_\alpha]^2} := \frac{T(\alpha)}{V(\alpha)},$$

where

$$T(\alpha) := E\left(\frac{1}{n}||r_\alpha||^2\right), \qquad V(\alpha) := \left[1 - \frac{1}{n}traceA_\alpha\right]^2. \tag{3.5}$$

For Tikhonov's regularization, the filter is given by the formula

$$\omega_\alpha^{Tikh}(s^2) = \frac{s^2}{s^2 + \alpha}. \tag{3.6}$$

Therefore from (3.3)-(3.6), one derives

$$T(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\frac{\alpha^2 s_i^2 \nu_i^2}{(s_i^2 + \alpha)^2} + \frac{\sigma^2}{n}\sum_{i=1}^{n}\frac{\alpha^2}{(s_i^2 + \alpha)^2} = \frac{\alpha^2}{n}\sum_{i=1}^{n}\frac{s_i^2\nu_i^2 + \sigma^2}{(s_i^2 + \alpha)^2}, \quad \nu_i := (x, v_i),$$

$$V(\alpha) = \left[\frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{s_i^2}{s_i^2 + \alpha}\right)\right]^2 = \frac{\alpha^2}{n^2}\left[\sum_{i=1}^{n}\frac{1}{s_i^2 + \alpha}\right]^2.$$

Combining these two expressions, one concludes that

$$EG(\alpha) = \frac{n\sum_{i=1}^{n}\frac{s_i^2\nu_i^2 + \sigma^2}{(s_i^2 + \alpha)^2}}{\left[\sum_{i=1}^{n}\frac{1}{s_i^2 + \alpha}\right]^2} := \frac{Q(\alpha)}{Z(\alpha)}. \tag{3.7}$$

Since $n < \infty$, the functions $Q(\alpha)$ and $Z(\alpha)$ are differentiable for all $\alpha > 0$. From (3.7), one obtains

$$Q'(\alpha) = -2n\sum_{i=1}^{n}\frac{s_i^2\nu_i^2 + \sigma^2}{(s_i^2 + \alpha)^3}, \quad Z'(\alpha) = -2\sum_{i=1}^{n}\frac{1}{s_i^2 + \alpha}\sum_{i=1}^{n}\frac{1}{(s_i^2 + \alpha)^2}.$$

A necessary condition of an unconstrained minimum is

$$EG'(\alpha) = \frac{Q'(\alpha)Z(\alpha) - Q(\alpha)Z'(\alpha)}{Z^2(\alpha)}$$
$$= \frac{-2n\sum_{i=1}^n \frac{s_i^2\nu_i^2+\sigma^2}{(s_i^2+\alpha)^3}\sum_{i=1}^n \frac{1}{s_i^2+\alpha} + 2n\sum_{i=1}^n \frac{s_i^2\nu_i^2+\sigma^2}{(s_i^2+\alpha)^2}\sum_{i=1}^n \frac{1}{(s_i^2+\alpha)^2}}{\left[\sum_{i=1}^n \frac{1}{s_i^2+\alpha}\right]^3} = 0.$$

Multiplication by $\frac{1}{2n}\left[\sum_{i=1}^n \frac{1}{s_i^2+\alpha}\right]^3 > 0$ yields

$$f(\alpha) := \frac{EG'(\alpha)}{2n}\left[\sum_{i=1}^n \frac{1}{s_i^2+\alpha}\right]^3 \tag{3.8}$$
$$= -\sum_{i=1}^n \frac{s_i^2\nu_i^2+\sigma^2}{(s_i^2+\alpha)^3}\sum_{j=1}^n \frac{1}{s_j^2+\alpha} + \sum_{i=1}^n \frac{s_i^2\nu_i^2+\sigma^2}{(s_i^2+\alpha)^2}\sum_{j=1}^n \frac{1}{(s_j^2+\alpha)^2} = 0. \tag{3.9}$$

Below we will look more closely at the function $f(\alpha)$ in order to locate the global minimum of $EG(\alpha)$. Introduce the notations $z_i := \frac{s_i^2\nu_i^2+\sigma^2}{(s_i^2+\alpha)^2}$ and $p_i := \frac{1}{s_i^2+\alpha}$. Then (3.8) implies

$$f(\alpha) = \sum_{i=1}^n z_i \sum_{j=1}^n p_j^2 - \sum_{i=1}^n z_i p_i \sum_{j=1}^n p_j.$$

Notice that in the above summation, we can cancel $2n$ terms, and then we have $2n^2-2n = 2n(n-1)$ (divisible by 4) terms left. Let us group together $z_i p_j^2, z_j p_i^2, -z_i p_i p_j$, and $-z_j p_j p_i$, for $1 \le i < j \le n$. There will be $n(n-1)/2$ such groups. One has

$$z_i p_j^2 + z_j p_i^2 - z_i p_i p_j - z_j p_j p_i = z_i p_j(p_j - p_i) - z_j p_i(p_j - p_i) = (z_i p_j - z_j p_i)(p_j - p_i),$$

and $f(\alpha)$ takes the form

$$f(\alpha) = \sum_{i=1}^n \sum_{j=i+1}^n (z_i p_j - z_j p_i)(p_j - p_i).$$

Clearly, since $s_1 \ge s_2 \ge ... \ge s_n > 0$, one concludes that $p_1 \le p_2 \le ... \le p_n$. Hence, $p_j - p_i \ge 0$ for $j > i$. Then the sign of $f(\alpha)$ will be determined by the terms

$$z_i p_j - z_j p_i = \frac{1}{(s_i^2+\alpha)(s_j^2+\alpha)}\left[\frac{s_i^2\nu_i^2+\sigma^2}{s_i^2+\alpha} - \frac{s_j^2\nu_j^2+\sigma^2}{s_j^2+\alpha}\right], \quad 1 \le i < j \le n.$$

Consider a special case $\nu_i^2 = \nu = const(\nu > 0)$. Denote $y := s_i^2$. Then $g(y) := \frac{y\nu+\sigma^2}{y+\alpha} = \nu\frac{y+\frac{\sigma^2}{\nu}}{y+\alpha}$ is decreasing for $\alpha < \frac{\sigma^2}{\nu}$ and increasing otherwise. If we additionally assume that for at least one value of i, $1 \le i < n$, $s_{i+1} < s_i$, then

$$f(\alpha) < 0 \quad \text{for} \quad \alpha < \frac{\sigma^2}{\nu}, \quad f(\alpha) > 0 \quad \text{for} \quad \alpha > \frac{\sigma^2}{\nu}, \quad \text{and} \quad f(\alpha) = 0 \quad \text{for} \quad \alpha = \frac{\sigma^2}{\nu}, \tag{3.10}$$

which means the expected value of the GCV function has global minimum at $\alpha(\sigma) = \frac{\sigma^2}{\nu}$. This simple example gives some idea as to how (and why) the GCV parameter selection method works. One can see that, even though the information on the noise level is not being used, the choice of the regularization parameter *is* done based on the noise level σ^2. That is, of course, "on average" since the analysis has been carried out for the expected value of the generalized cross validation.

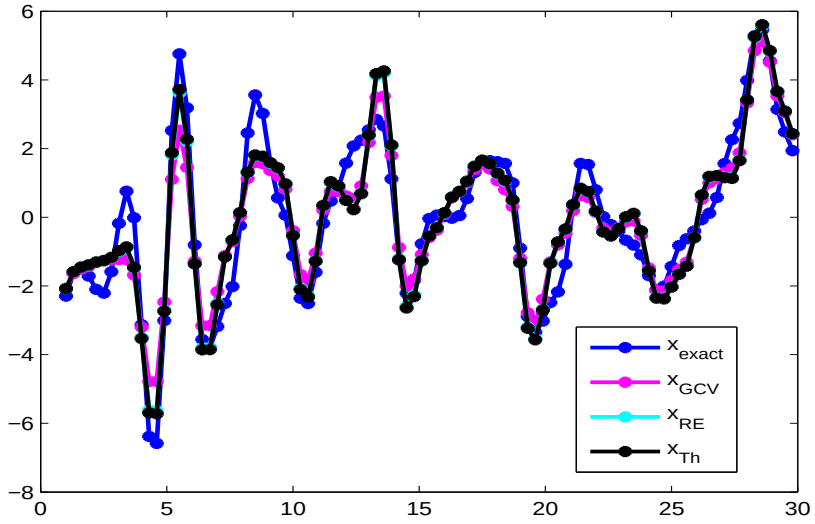

Figure 5. Comparison of solutions, $x(t)$, for 5% noise in the data ($\nu_i^2 = \nu = 100$)

The above observation also illustrates that, as opposed to L-curve [LY97], the GCV approach does not result in a systematic error because, at least for this type of exact solution, $\alpha(\sigma) = \frac{\sigma^2}{\nu} \to 0$ as $\sigma \to 0$.

Figure 5 and Table 4 illustrate numerical simulations for this special case with $\nu = 100$. In the table, α_{Th} stands for the exact value of $\frac{\sigma^2}{\nu}$, while α_{GCV} is the actual value obtained as the minimizer of the GCV function. One can see that α_{Th} is remarkably close to the optimal value, α_{RE}, and α_{GCV} provides a rather accurate estimate for both α_{RE} and α_{Th}. All three approximate solutions turn out to be similar.

Table 4. Relative errors for theoretical, cross validated, and optimal values of α

Noise	α_{GCV}	α_{RE}	α_{Th}	$\frac{\|\|x_{\alpha_{GCV}} - \hat{x}\|\|}{\|\|\hat{x}\|\|}$	$\frac{\|\|x_{\alpha_{RE}} - \hat{x}\|\|}{\|\|\hat{x}\|\|}$	$\frac{\|\|x_{\alpha_{Th}} - \hat{x}\|\|}{\|\|\hat{x}\|\|}$
0.5%	4.68×10^{-4}	4.00×10^{-6}	2.31×10^{-5}	$4.586189e-02$	$3.688925e-02$	$3.692068e-02$
1%	9.80×10^{-4}	6.00×10^{-5}	9.25×10^{-5}	$8.981639e-02$	$7.369934e-02$	$7.373231e-02$
3%	3.46×10^{-3}	8.80×10^{-4}	8.33×10^{-4}	$2.448693e-01$	$2.151304e-01$	$2.151459e-01$
5%	6.58×10^{-3}	2.60×10^{-3}	2.31×10^{-3}	$3.670473e-01$	$3.389240e-01$	$3.392022e-01$
8%	1.26×10^{-2}	6.92×10^{-3}	5.92×10^{-3}	$5.019018e-01$	$4.833143e-01$	$4.844241e-01$
12%	2.41×10^{-2}	1.61×10^{-2}	1.33×10^{-2}	$6.204218e-01$	$6.103246e-01$	$6.126281e-01$
15%	3.58×10^{-2}	2.54×10^{-2}	2.08×10^{-2}	$6.801905e-01$	$6.727331e-01$	$6.754755e-01$
20%	6.22×10^{-2}	4.48×10^{-2}	3.70×10^{-2}	$7.481651e-01$	$7.418337e-01$	$7.443643e-01$

Now consider the general case: $m \le \nu_i^2 \le M$, $i = 1, 2, ..., n$, for some $M > m \ge 0$. Intuitively, one would think that for $m \le \nu_i^2 \le M$, instead of one point $\alpha(\sigma) = \frac{\sigma^2}{\nu}$, there will be an interval $[\alpha_{min}, \alpha_{max}]$ such that $f(\alpha) < 0$ when $\alpha < \alpha_{min}$ and $f(\alpha) > 0$ when $\alpha > \alpha_{max}$. Let us see if this

is always the case. The sign of $z_i p_j - z_j p_i$ is determined by the sign of the following function

$$\begin{aligned} h(\alpha) :=& \frac{s_i^2\nu_i^2 + \sigma^2}{s_i^2 + \alpha} - \frac{s_j^2\nu_j^2 + \sigma^2}{s_j^2 + \alpha} \\ =& \frac{s_i^2 s_j^2(\nu_i^2 - \nu_j^2) + \alpha(s_i^2\nu_i^2 - s_j^2\nu_j^2) - \sigma^2(s_i^2 - s_j^2)}{(s_i^2 + \alpha)(s_j^2 + \alpha)}, \quad i < j. \end{aligned}$$

Notice that whenever

$$\nu_1^2 > \nu_2^2 > ... > \nu_n^2, \tag{3.11}$$

the numerator is positive for any $\alpha \geq 0$ if $s_i = s_j$. If $s_i > s_j$, then it is still positive as long as $\frac{s_i^2 s_j^2(\nu_i^2 - \nu_j^2)}{s_i^2 - s_j^2} > \sigma^2$. Hence, when

$$\sigma^2 < \min_{1 \leq i < j \leq n,\ s_i < s_j} \frac{s_i^2 s_j^2(\nu_i^2 - \nu_j^2)}{s_i^2 - s_j^2}, \tag{3.12}$$

the expected value of the GCV function has a global minimum at $\alpha = 0$. In other words, for "small" noise satisfying (3.12) the GCV function suggests not to regularize. If the coefficients ν_i^2, $i = 1, 2, ..., n$, are not ordered according to (3.11), then one has

$$h(\alpha) \leq \frac{s_i^2 s_j^2(M - m) + \alpha(s_i^2 M - s_j^2 m) - \sigma^2(s_i^2 - s_j^2)}{(s_i^2 + \alpha)(s_j^2 + \alpha)}.$$

Therefore, $f(\alpha) < 0$ for $\alpha < \alpha_{min}$, where

$$\alpha_{min} := \min_{1 \leq i < j \leq n} \frac{\sigma^2(s_i^2 - s_j^2) - s_i^2 s_j^2(M - m)}{s_i^2 M - s_j^2 m}, \tag{3.13}$$

provided that

$$\sigma^2 > \max_{1 \leq i < j \leq n,\ s_i > s_j} \frac{s_i^2 s_j^2(M - m)}{s_i^2 - s_j^2} = \max_{1 \leq i < n,\ s_i > s_{i+1}} \frac{s_i^2 s_{i+1}^2(M - m)}{s_i^2 - s_{i+1}^2}. \tag{3.14}$$

At the same time, for any $1 \leq i < j \leq n$,

$$h(\alpha) \geq \frac{-s_i^2 s_j^2(M - m) + \alpha(s_i^2 m - s_j^2 M) - \sigma^2(s_i^2 - s_j^2)}{(s_i^2 + \alpha)(s_j^2 + \alpha)}.$$

That means, $f(\alpha) > 0$ for $\alpha > \alpha_{max}$, where

$$\alpha_{max} := \max_{1 \leq i < j \leq n} \frac{\sigma^2(s_i^2 - s_j^2) + s_i^2 s_j^2(M - m)}{s_i^2 m - s_j^2 M}, \tag{3.15}$$

$$\text{if} \quad s_i^2 m - s_j^2 M > 0 \quad \text{for all} \quad 1 \leq i < j \leq n. \tag{3.16}$$

So, the expected value of the GCV function will reach its global minimum inside the interval $[\alpha_{min}, \alpha_{max}]$ for α_{min} and α_{max} satisfying (3.13) and (3.15), respectively, if inequalities (3.14) and (3.16) hold. Once again, we observe that for small perturbations (not satisfying (3.14)) the success of the GCV scheme is not guaranteed. That may be the case illustrated in Table 3, the first row, when the GCV fails to capture 0.5% relative noise in the simulated data.

4. **Reconstruction of Human Treatment-Recovery Rate**

Here we use algorithm (2.2)-(2.4) and the approximate values of $C'(t)$ and $C''(t)$ computed in Section 2 to evaluate $y(t)$, the proportion of infected mosquitos and $\gamma(t)$, the treatment-recovery rate for humans. At the first step, we solve an auxiliary inverse problem in order to compute the human birth rate, $\lambda(t)$. Adding together equations in (1.3)-(1.6), one concludes that $P'(t) = \lambda(t) - \mu P(t)$. On the other hand, following [MH10], we assume that Indian population satisfies the logistic human population model:

$$P' = aP\left(1 - \frac{P}{K}\right) - \mu P, \tag{4.1}$$

where a is the growth rate for human population, and K is the carrying capacity of the environment. This implies $\lambda(t) = aP\left(1 - \frac{P}{K}\right)$. Using the census data for $P(t)$ listed in Table 5 and considering the average life expectancy in India over the 30 year period to be 65 (i.e., $\mu = 1/65$), we formulate the auxiliary inverse problem to estimate the remaining parameters, a and K.

Table 5. The real data set for Indian human population [FAO]

Year	1984	1985	1986	1987	1988	1989
Population (million)	764.749	781.737	798.942	816.329	833.834	851.375
Year	1990	1991	1992	1993	1994	1995
Population (million)	868.891	886.349	903.750	921.108	938.453	955.804
Year	1996	1997	1998	1999	2000	2001
Population (million)	973.148	990.460	1007.747	1025.015	1042.262	1059.501
Year	2002	2003	2004	2005	2006	2007
Population (million)	1076.706	1093.787	1110.626	1127.144	1143.289	1159.095
Year	2008	2009	2010	2011	2012	2013
Population (million)	1174.662	1190.138	1205.625	1221.156	1236.687	1252.140

To that end, we spline the above data, partition our time interval $[1, 30]$ with grid points $\{t_j\}_1^J$ by taking a small step size h, and apply the central difference formula to estimate $P'(t)$. Let $P_j := P(t_j)$, $j = 1, 2, ..., J$. Then

$$(2hP_j)(a - \mu) + (-2hP_j^2)\frac{a}{K} = P_{j+1} - P_{j-1}, \quad j = 2, 3, ..., J - 1.$$

Thus, one arrives at a linear system with $J-2$ equations and 2 unknowns, $\mathcal{A}x = \mathcal{F}$, which is solved in the sense of least squares. As the solution to this auxiliary problem, we get $a = 5.410770 \times 10^{-2}$ and $K = 2.525319 \times 10^3$. With two unknowns only, the above inverse problem is rather stable and does not need to be regularized.

Table 6. Values of Parameters

Parameter	a (computed)	K (computed)	r_0 (fitted)	b_0 [MH10]	d_0 [MH10]
Value	0.05410770	2525.319	0.00037477	16.9923	12.55
Units	years^{-1}	numbers (million)	years^{-1}	years^{-1}	years^{-1}
Parameter	η [MH10]	ξ [MH10]	y_0 [MH10]	μ [MH10]	ν [MH10]
Value	0.0391625	0.00626284	0.02	1/65	365/7
Units	years^{-1}	unitless	unitless	years^{-1}	years^{-1}

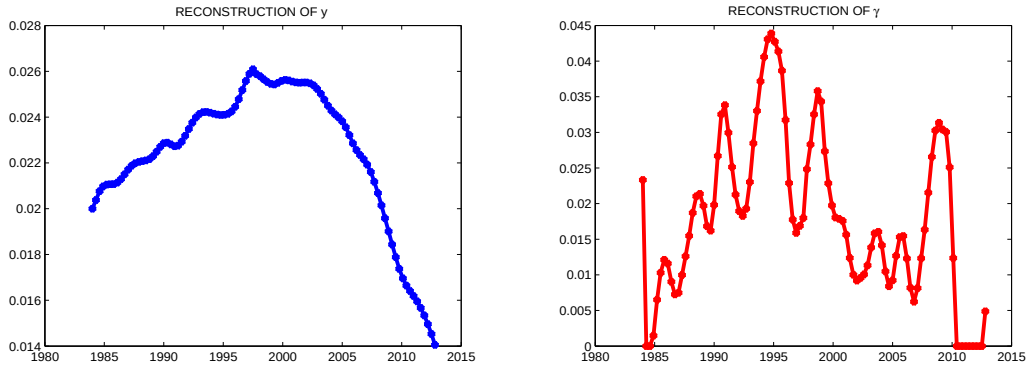

Figure 6. Reconstruction of $y(t)$ and $\gamma(t)$ for real data

Given $C'(t)$ and $C''(t)$ approximated in Section 2 as well as the expression for $\lambda(t)$, the unknown functions $y(t)$ and $\gamma(t)$ can now be reconstructed by (2.2)-(2.4). In [MH10], the parameters $b(t)$, $r(t)$, and $d(t)$ are considered to be step functions due to the start of the Enhanced Malaria Control Project (EMCP) in 1997, i.e.,

$$b(t) = b_0(1 - \xi H(t - 13.5)), \quad r(t) = r_0(1 - \xi H(t - 13.5)), \quad d(t) = d_0 + \eta H(t - 13.5),$$

where $H(t) = 0$ for $t \leq 0$ and $H(t) = 1$ for $t > 0$ (the time $t = 0$ corresponds to 1983), and the values of the parameters are listed in Table 6. However, even though a variety of mosquito control measures aiming to reduce the population of mosquitos were incorporated with the EMCP in 1997, one probably cannot expect their immediate effect on the disease transmission rates. Therefore in this paper, we modify $b(t)$ and $r(t)$

$$b(t) = b_0 \left(1 - \xi H(t - 14.5)(t - 13.5)^{1.6}\right), \quad r(t) = r_0 \left(1 - \xi H(t - 14.5)(t - 13.5)^{1.6}\right),$$

(with re-fitted value of r_0) so that the two transmission rates decrease gradually after the implementation of the EMCP in 1997. With these $b(t)$ and $r(t)$, the estimated values of $y(t)$ and $\gamma(t)$ are shown in Figure 6.

One can see that introduction of the Roll Back Malaria program results in the decrease of infected mosquito population and positively affects the treatment recovery rate. Treatment-recovery rate $\gamma(t)$ decreased in the period 2000-2006, possibly due to increasing resistance to sulphalene/ sulphadoxine-pyrimethamine (SP) used in chloroquine resistant areas in India at that time. In 2005, Artemisinin combination therapy (ACT) (artesunate (AS) plus SP) was approved for use in chloroquine-resistant areas and the spread of application of ACT continuously grew until 2010 when AS plus SP became the standard of treatment in India [AA14]. Our treatment and recovery rate captures the better treatment strategies in the period 2006-2010 by showing significant spike. After 2010, our treatment rate shows decline, possibly due to increasing resistance to SP, one of the drugs in ACT. India responded by replacing the combination by artemether lumefantrine in 2013 in seven areas [AA14].

5. **Conclusions**

The theoretical analysis carried out in Section 3 provides some insight into what can and cannot be expected in the framework of the GCV approach. The absence of the systematic error [LY97] is a nice feature. At the same time, "little" noise may be hard to capture, and the overall tendency to under-regularize clearly exhibits itself for a small sample. In [TY00], the authors show a number

of counterexamples illustrating a possible lack of success in the implementation of GCV method for 2×2 linear systems. And even for a relatively "large" noise, there are cases when GCV fails.

Nevertheless, for the lack of a better alternative, when the noise level is unknown, the GCV method can, in many cases, be applied to determine a "near optimal" regularization parameter. At times, its implementation needs to be combined with the use of partial data sets as shown in Section 2. Our numerical simulations for *Plasmodium falciparum* malaria disease, for the most part, confirm the theoretical findings for both synthetic and real data.

From practical perspective the estimated proportion of infected mosquitos $y(t)$ captures the success of the Roll Back Malaria program in controlling infected mosquitos whose decline follows the decline in human cases. At the same time our estimated treatment-recovery rate $\gamma(t)$ illustrates the complex nature of the interplay between treatment of malaria and resistance to major drugs. Article [AA14] outlines the almost continuous changes in treatment policies made in India, and our treatment-recovery rate shows that such changes are necessary for the continued success of malaria treatment in the face of resistance.

References

[AA14] Anupkumar R. Anvikar, Usha Arora, G.S. Sonal, Neelima Mishra, Bharatendu Shahi, Deepali Savargaonkar, Navin Kumar, Naman K. Shah, and Neena Valecha [2014], *Antimalarial drug policy in India: Past, present & future*, Indian J Med Res. **139(2)**, 205–215.

[B84] Bakushinskii, A.B. [1984] *Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion.* USSR Comput. Math. Math. Phys. **24**, 181–182.

[BL11] Bauer, F. and Lukas, M. A., [2011] *Comparing parameter choice methods for regularization of ill-posed problems* . Mathematics and Computers in Simulation, **81**, 1795–1841.

[FAO] FAOSTAT: http://faostat.fao.org/site/550/DesktopDefault.aspx?PageID=550#ancor

[H98] Hansen, P.C. [1998] *Rank-deficient and discrete ill-posed problems. Numerical aspects of linear inversion.* SIAM Monographs on Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

[HO93] Hansen, P.C. and O'Leary, D.P. [1993] *The use of the L-curve in the regularization of discrete ill-posed problems.* SIAM J. Sci. Comput. **14** N6, 1487-1503.

[H92] Hansen, P.C. [1992] *Analysis of discrete ill-posed problems by means of the L-curve.* SIAM Rev. **34** N4, 561-580.

[LY97] Leonov, A. S. and Yagola, A. G. [1997] *The L-curve method always produces an unremovable systematic error.* (Russian) Vestnik Moskov. Univ. Ser. III Fiz. Astronom. **6**, 17-19.

[LDA12] Lukas, M.A., De Hoog, F.R. and Anderssen, R.S. [2012] *Performance of robust GCV and modified GCV for spline smoothing.* Scand. J. Stat. **39** N1, 97-115.

[L78] Leonov, A. S. [1978] *Justification of the choice of regularization parameter according to quasi-optimality and quotient criteria.* USSR Comp. Math. and Math. Phys. **18**, N6, 115.

[MS] Malaria Site: http://www.malariasite.com/malaria/MalariaInIndia.htm

[MH10] Martcheva, M. and Hoppensteadt, F. [2010] *India's approach to eliminating plasmodium falciparum malaria: a modeling perspective*, J. Biological Systems. **18**, N4, 867–891.

[NM] http://www.nationmaster.com/graph/peo_pop-people-population

[RR13] Reichel, L. and Rodriguez, G. [2013] *Old and new parameter choice rules for discrete ill-posed problems.* Numer. Algorithms **63** N1, 65-87.

[TA77] Tikhonov, A.N. and Arsenin, V.Y. [1977] *Solutions of ill-posed problems.* Translated from the Russian. Preface by translation editor Fritz John. Scripta Series in Mathematics. V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York-Toronto, Ont.-London.

[TGSY95] Tikhonov, A.N., Goncharsky, A.V., Stepanov, V.V., and Yagola, A.G. [1995] *Numerical methods for the solution of ill-posed problems*, Kluwer Academic Publishers, Dordrecht.

[TLY98] Tikhonov, A.N., Leonov, A.S. and Yagola, A.G. [1998] *Nonlinear ill-posed problems,* V.1, 2, Chapman and Hall, London.

[TY00] Titarenko, V.N. and Yagola, A.G. [2000] *Application of the GCV method for well-posed and ill-posed problems.* (Russian) Vestnik Moskov. Univ. Ser. III Fiz. Astronom., 2000, **4**, 15-18.

[V02] Vogel, C.R. [2002] *Computational Methods for Inverse Problems, research monograph*, Frontiers in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

[W90] Wahba, G. [1990] *Spline models for observational data.* CBMS-NSF Regional Conference Series in Applied Mathematics, 59. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.