The problem that is considered by this paper is the estimation of static parameters. Given the observation set as $y_{1:T} = (y'_1, \ldots, y'_T)'$ and using Bayes's theorem yield the target density for the state, $p(x_\tau | y_{1:T}, \theta, \phi)$. This can be computed with sampling-based sequential Monte Carlo (MC) techniques (Künsch, 2005; Godsill *et al.*, 2004).

Unknown parameters θ and ϕ can be then be determined by maximizing the likelihood (see Kitagawa (1996)):

$$L(\theta, \phi | y_{1:T}) = \prod_{t=1}^{T} \int p(y_t | x_t, \phi)\, p(x_t | y_{1:t-1}, \theta)\, \mathrm{d}x_t$$
$$\approx \prod_{t=1}^{T} \left\{ N_t^{-1} \sum_{i=1}^{N_t} p(y_t | x_{t|t-1}^{(i)}, \theta, \phi) \right\}$$

where the latter approximation relies on $\{x_{t|t-1}^{(i)}\}, i = 1, \ldots, N_t$, which is a sample from the predictive density generated via sequential MC sampling. The resultant likelihood is affected by MC sampling variability and challenges optimizers; incorporation of kernel density estimation appears useful (de Valpine, 2004).

Computationally, application of these MC approaches to higher dimensional dynamic systems is a major challenge. Ideas based on dynamical analysis (Chorin and Krause, 2004) and effective approximations, e.g. the ensemble Kalman filter (Evensen, 2003), appear promising. I have compared some of these methods for non-linear dynamic systems (Dowd, 2007). It would be interesting to compare these further with the parameter estimation method of this paper, extended to the case of stochastic DEs.

David J. D. Earn (*McMaster University, Hamilton*)

Estimation of parameters of non-linear differential equations from noisy, observed time series is a problem that arises frequently in applied science. Unfortunately, anyone who has tried this is likely to be familiar with serious theoretical and computational challenges. The new method of Ramsay and colleagues is very welcome, and it will be interesting to see how it fares on a wide range of problems.

The method may prove particularly useful for the study of transmission dynamics of infectious diseases (Anderson and May, 1991). The state variables in the basic *susceptible–infectious–recovered* (*SIR*) *model* are the numbers of individuals who are susceptible (S), infectious (I) and recovered or immune (R), and the parameters are the rates of birth (ν), death (μ), transmission (β) and recovery (γ):

$$\dot{S} = \nu - (\beta I + \mu) S, \tag{41a}$$

$$\dot{I} = \beta I S - (\gamma + \mu) I, \tag{41b}$$

$$\dot{R} = \gamma I - \mu R. \tag{41c}$$

Note that I records the *prevalence* of the disease, i.e. the number of individuals who are currently infected. We typically observe *incidence*, i.e. $\int \beta S I\, \mathrm{d}t$, where the integral is over the reporting interval (typically weekly or monthly, but sometimes daily).

For human diseases that have been present in the population for years, we typically have estimates of all the parameters from data other than time series of reported cases or deaths. Moreover, the SIR model as formulated in equation (41) has a globally asympotically stable equilibrium, so we can easily compare the predicted equilibrium with the observed times series (without the aid of Ramsay and colleagues).

The catch is that the transmission rate β is rarely constant in practice. Instead, β often varies seasonally, either because of seasonally changing aggregation patterns (London and Yorke, 1973) or other seasonal factors that may be difficult to pin down (Dushoff *et al.*, 2004). Seasonal forcing drastically changes the dynamics of the SIR model, often leading to co-existing stable cycles (Schwartz and Smith, 1983) or chaos (Schaffer, 1985). Since the conclusions that we draw depend strongly on the estimated amplitude of seasonal forcing (Earn *et al.*, 2000; Bauch and Earn, 2003), we need a credible way of estimating time variation in β and we rarely have useful data to work with other than incidence time series. The method of Ramsay and colleagues is begging to be applied to this problem and I look forward to comparing the results that are inferred from it and from previous methods (Fine and Clarkson, 1982; Ellner *et al.*, 1998; Finkenstädt and Grenfell, 2000; Bjornstad *et al.*, 2002; Wallinga and Teunis, 2004).

Finally, it is worth mentioning a vexing issue that has the potential to undermine parameter estimation for differential equation models of disease spread. The process of infectious disease transmission is fundamentally stochastic. Solutions of the SIR model (41) can be thought of as ensemble means of the true stochastic process (Kurtz, 1980), but any incidence time series represents only one realization of that stochastic process and may not accurately reflect the mean. In the specific context that I have highlighted—estimating a seasonal forcing function—this problem may not be serious if we have data covering many seasons, but it is worth bearing in mind.

Stephen P. Ellner (*Cornell University, Ithaca*)
I congratulate the authors for two important contributions: the profiling method and for highlighting to the statistical community the problems of fitting dynamical systems models. Non-linear differential equations are core models in many sciences, including my own discipline of ecology, but are sorely neglected in statistical research (Ellner and Guckenheimer, 2006). In ecology, low dimensional non-linear dynamic models that would have been called a caricature or metaphor 20 years ago have proved remarkably successful in confrontations with real data (e.g. Zimmer (1999) and Turchin (2003)). These models necessarily leave out many 'inessentials' (rare species, spatial variability, etc.) and are often deterministic even though we know better. Omitted inessentials are problematic for non-linear systems because they can have a large effect on long-term model trajectories even if their effect at any instant is small. Wanting $f\{\hat{x}(t),\theta\}$ to be near $\mathrm{d}\hat{x}(t)/\mathrm{d}t$, where $\hat{x}(t)$ is near the data, is a more reasonable hope for a model with the right 'essentials'.

But for practical acceptance I believe that selection of the smoothing parameter λ must be automated on a defensible basis. The profiling criterion immediately suggests cross-validation. Straight leave-one-out methods are computationally infeasible for end-users (though computer and algorithmic improvements may change this situation), but we can still use the principle of predicting something that was not used in fitting. Dynamic models predict the future, so we can evaluate them on the basis of forecasting accuracy. Let $\phi_t(x_0;\theta)$ be the model solution at time t starting from $x(0)=x_0$. A measure of prediction error at time interval τ is

$$\mathrm{PE}(\lambda;\tau)=\sum_j \|y(t_j)-\varphi_\tau\{\hat{x}_\lambda(t_j-\tau);\hat{\theta}_\lambda\}\|^2. \tag{42}$$

PE should be large if $\hat{x}_\lambda$ undersmooths or oversmooths the data, either way throwing off parameter estimates. I tried this criterion on the FitzHugh–Nagumo system (modifying MATLAB code that was provided by Hooker), with the omitted inessential being an additive perturbation to $\mathrm{d}V/\mathrm{d}t$ (Fig. 14(a)) that changes the period of the oscillations (Fig. 14(b) *versus* Fig. 14(c)). Fitting five artificial data sets by profiling with a range of λ-values, PE selects a range of λ-values that is good for parameter estimation (Figs 14(d) and 14(e)). Profiling with a 'good' λ performs comparably with two-step methods in which the data are smoothed without regard to the model, and the ordinary differential equation is then fitted to the smooth or its time derivative; with a 'bad' λ profiling is less successful. Profiling's big advantage over two-step methods is that it does not need data on all state variables but, as this small example indicates, success may depend on choosing λ well.

Chong Gu (*Purdue University, West Lafayette*)
The authors are to be congratulated for a fine paper on a challenging problem. As shown in the paper, fitting data to models derived from ordinary differential equations (ODEs) involves numerous issues such as the numerical strategies and the methodological framework, and it is the methodological aspects that we shall comment on.

First let us attempt a crude parallel between the setting of the paper and a standard cubic spline as the minimizer of

$$\sum_{i=1}^{n}\{y_i-x(u_i)\}^2+\lambda\int\left(\frac{\mathrm{d}^2x}{\mathrm{d}u^2}\right)^2\mathrm{d}u. \tag{43}$$

Setting $\lambda=\infty$ in expression (43) forces $\mathrm{d}^2x/\mathrm{d}u^2=0$ that characterizes a *static system*, with the solution of the form $x(u)=c_1+c_2u$, where (c_1,c_2) are to be determined by the data (y_i,u_i) through the least squares; if precise readings of (u,x) are available from $x(u)=c_1+c_2u$, we need only two pairs of 'initial values' to pin down (c_1,c_2). Likewise, replacing $\mathrm{pen}(x)=\int(\mathrm{d}^2x/\mathrm{d}u^2)^2\,\mathrm{d}u$ in expression (43) by $\mathrm{pen}(x)=\int(\mathrm{d}^2x/\mathrm{d}u^2+\omega^2x)^2\,\mathrm{d}u$ yields an L-spline, and setting $\lambda=\infty$ then forces $\mathrm{d}^2x/\mathrm{d}u^2+\omega^2=0$ with the solution of the form $x(u)=c_1\sin(\omega u)+c_2\cos(\omega u)$. Compare these with

$$\sum_{i=1}^{n}\{y_i-x(t_i)\}^2+\lambda\int\left\{\frac{\mathrm{d}x}{\mathrm{d}t}-f(x,\mathbf{u},t|\boldsymbol{\theta})\right\}^2\mathrm{d}t, \tag{44}$$

where for simplicity we consider only a single ODE. The main difference between expressions (43) and (44) is the time variable t in expression (44) and the implicit dependence of x on u. The system parameter