Chapter 2
Examples as a Guide to the Issues

Prologue to the Chapter

Before we delve into the general structure of using information from measurements to complete models of those measurements, we will illustrate many of the questions involved by taking a look at some well-trodden ground. Completing a model means that we have estimated all the unknown parameters in the model, allowing us to predict the development of the model in its state space given a set of initial conditions and a statement of the forces acting to drive it.

We will review an example from simple fluid dynamics, the Malkus/Howard waterwheel, and then we discuss an example of a nonlinear electronic circuit, the Colpitts oscillator, widely used in commercial products. A third example will be a consideration of a Hodgkin–Huxley model of a neuron.

The first and third examples are done in computer simulation in what we call "twin experiments". In these we generate "data" from a model with known fixed parameters. We show how we can estimate these parameters as well as the values of the unobserved and observed state variables of the dynamics at the end of an observation period, $t = T$. We then predict for times beyond T using the estimated parameters and full set of estimated state variables. In later chapters we return to the neurobiological example and use what we have learned to develop models for experimental observations of neurons from a functional neuronal network. In the second example, the Colpitts circuit, we have both numerical and experimental ingredients, yet retain substantial control over the proposed equations modeling the experiment.

The examples are quite revealing of the much harder problems we aim to address. Their formulation requires some undergraduate physics while their solution requires more demanding ideas. It is our hope that in the context of these quite simple settings, the demanding ideas stand out. The material also introduces one of the ways, a quite widely used method actually, for answering the questions of interest. We will see after the next two chapters that the approach is an approximation to the full answer to questions about estimating fixed parameters and state variables.

H.D.I. Abarbanel, *Predicting the Future: Completing Models of Observed Complex Systems*, Understanding Complex Systems, DOI 10.1007/978-1-4614-7218-6_2,

Though the measurements are recognized as noisy, this approach takes the models as having no errors. In that sense it is appropriate for the twin experiments covered here, but it does not provide error bars or any statistical context to the overall problem. Those important aspects will be available via the discussion upcoming in future chapters.

2.1 The Malkus Waterwheel

We begin our discussion of completing models of complex systems with examples that are only a "little bit complex". The first example is from a fluid mechanical system. This is the waterwheel devised by Malkus in the 1970s (Malkus 1972). There are other detailed descriptions of this waterwheel primarily concentrating on the lovely idea that a simple mechanical apparatus can exhibit chaotic motions as described by Lorenz in 1963 (Strogatz 1994). An experimental study of the waterwheel is available from Illing et al. (2012a,b) who also proceed in the spirit discussed here. Their laboratory realization of the Malkus waterwheel is shown in Fig. 2.1.

The waterwheel comprises a series of leaky buckets arrayed around the circumference of a wheel of radius R. The wheel is tilted with respect to the vertical along which gravity acts. Looking at the wheel from above we associate a spatial location on the wheel by an angle $\theta;\ 0 \leq \theta \leq 2\pi$. We describe the water in the buckets at various θ by a mass density $m(\theta, t)$, and we bring water into the buckets at a rate $Q(\theta)$. The forces on the water in the buckets drive the wheel through the torque associated with the weight of the water $g' R \int_0^{2\pi} \mathrm{d}\theta\, m(\theta, t) \sin\theta$. $g' = g \sin(\alpha)$ is the effective gravitational acceleration. We select the inflow of the water to be symmetric around the top of the wheel at $\theta = 0$:

$$Q(\theta) = \sum_{n=0}^{\infty} q_n \cos(n\theta). \tag{2.1}$$

We also account for leakage from the buckets via a factor $-km(\theta, t)$ in the continuity equation expressing conservation of water. With the water source $Q(\theta)$ and the loss term, conservation of water is expressed as

$$\frac{\partial m(\theta, t)}{\partial t} + \omega(t) \frac{\partial m(\theta, t)}{\partial \theta} = Q(\theta) - km(\theta, t), \tag{2.2}$$

where the angular velocity of the wheel is $\omega(t)$. The second term on the left-hand side of this equation represents advective transport of water from one angle θ to another as the wheel rotates at angular velocity $\omega(t)$.

An equation for $\omega(t)$ comes from equating the rate of change of the angular momentum $I\omega(t)$, with I the moment of inertia of the wheel with water in the buckets, to the frictional losses in the pivot of the wheel $-\nu\omega(t)$ plus the torque on

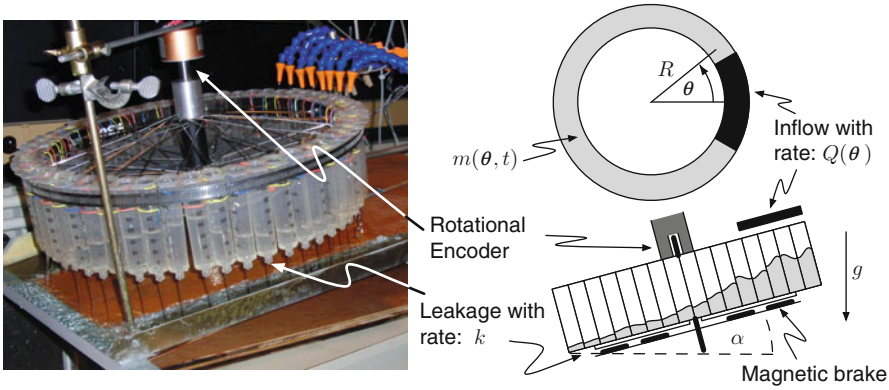

Fig. 2.1 An experimental setup of the Malkus waterwheel from Lucas Illing and his students at Reed college (Illing et al. 2012a,b). I am grateful to professor Illing for permission to use their excellent graphics. *Left* The experimental setup. *Right* Illustration of the quantities used in the Malkus dynamical equations. The **effective acceleration of gravity** in this experiment is $g\sin(\alpha)$ where α is the angle of tilt of the rotating wheel. Friction opposing the rotation is provided by a magnetic brake

the wheel arising from water in the buckets. This leads to

$$I\frac{\mathrm{d}\omega(t)}{\mathrm{d}t} = -\nu\omega(t) + g'R\int_0^{2\pi}\mathrm{d}\theta\, m(\theta,t)\sin\theta. \tag{2.3}$$

Both $Q(\theta)$ and $m(\theta,t)$ are periodic in θ, so the partial differential equation for $m(\theta,t)$ may be simplified if we write

$$m(\theta,t) = \sum_{n=0}^{\infty}[a_n(t)\cos(n\theta) + b_n(t)\sin(n\theta)]. \tag{2.4}$$

Comparing coefficients of $\cos(n\theta)$ and $\sin(n\theta)$, we have

$$\begin{aligned}
\frac{\mathrm{d}a_n(t)}{\mathrm{d}t} &= -n\omega(t)b_n(t) - ka_n(t) + q_n \\
\frac{\mathrm{d}b_n(t)}{\mathrm{d}t} &= n\omega(t)a_n(t) - kb_n(t) \\
\frac{\mathrm{d}\omega(t)}{\mathrm{d}t} &= -\frac{\nu}{I}\omega(t) + \frac{\pi g'R}{I}b_1(t).
\end{aligned} \tag{2.5}$$

These equations show that knowing $\omega(t), a_n(0), b_n(0)$, and q_n determines $a_n(t)$ and $b_n(t)$ for all n. Also we see that the three quantities $\{a_1(t), b_1(t), \omega(t)\}$ form a separate subset of the whole infinite-dimensional space labeled by the location θ on the wheel. Solving the set of three ordinary differential equations for $n = 1$ yields $\omega(t)$ which, along with q_n, enables the determination of $m(\theta,t)$.

For $n = 0$, we have

$$\frac{\mathrm{d}a_0(t)}{\mathrm{d}t} = -ka_0(t) + q_0$$
$$\frac{\mathrm{d}b_0(t)}{\mathrm{d}t} = -kb_0(t), \tag{2.6}$$

indicating that $a_0(t) \to \frac{q_0}{k}$ and $b_0(t) \to 0$ exponentially rapidly in time. These then drop out of our consideration.

For $n = 1$, we have

$$\frac{\mathrm{d}a_1(t)}{\mathrm{d}t} = -\omega(t)b_1(t) - ka_1(t) + q_1$$
$$\frac{\mathrm{d}b_1(t)}{\mathrm{d}t} = \omega(t)a_1(t) - kb_1(t)$$
$$\frac{\mathrm{d}\omega(t)}{\mathrm{d}t} = -\frac{\nu}{I}\omega(t) + \frac{\pi g' R}{I}b_1(t). \tag{2.7}$$

If we scale by constants $\{\alpha, \beta, \gamma, \delta\}$ and translate the dynamical variables as $a_1(t) = \alpha x(t) + \frac{q_1}{k}$, $b_1(t) = \beta y(t)$, $\omega(t) = \gamma x(t)$, and $t = \delta\tau$, we may choose the constants to arrive at

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \sigma(y(t) - x(t))$$
$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = -y(t) + Ax(t) - z(t)x(t)$$
$$\frac{\mathrm{d}z(t)}{\mathrm{d}t} = -Bz(t) + x(t)y(t) \qquad B = 1. \tag{2.8}$$

2.1.1 A Physics Question About the Waterwheel

Most discussions of the Malkus waterwheel (Strogatz 1994) are concerned with properties of the solutions for $a_1(t), b_1(t), \omega(t)$ and illustrating ideas about bifurcations and chaos in nonlinear systems. Those are important issues.

We ask here the following, somewhat different, question:

- If we observe `only` the rotation rate of the wheel $x_{\mathrm{data}}(t)$ (or $\omega_{\mathrm{data}}(t) = \frac{x_{\mathrm{data}}(t)}{\gamma}$) over a time segment $[0, T]$, can we use the information in that time series along with a model for the waterwheel flow to determine the constants σ, A, B and the values of the unobserved state variables $y(t)$ and $z(t)$ over that period of time? Further, if we know σ, A, B and $\{x(T), y(T), z(T)\}$, can we predict the flow for $t > T$?

This is the essential kind of question we pose about complex systems where the inability to measure or observe all of the relevant variables is standard. Also present in larger and more complex problems is the chaotic motion we encounter in the first

two examples discussed here. These matters are easier to identify and diagnose in our "only slightly complex" examples, so we will attend to them for a bit. Then with the lessons in mind, we'll turn to a general formulation of the data assimilation problem. The observation of a subset of the variables, here we observe only $\omega_{\text{data}}(t)$, is meant to represent the realistic situation we face in larger complex dynamical settings.

Returning to the simple waterwheel flow, we call estimating σ, A, B `completing the model`. We denote evaluating $\{x(t > T), y(t > T), z(t > T)\}$ `predicting the future` of our model of waterwheel flow:

- The first part completes the formulation of our model by specifying the unknown constants in the physical description of the flow.
- The second part, equally important, validates (or invalidates) our model of the flow, by utilizing our estimates of $\{x(T), y(T), z(T)\}$ as initial conditions for the model to predict the state of the flow for $t > T$.

We will demonstrate explicitly that these two tasks can be accomplished, and along the way, we will identify issues that arise even in this simple flow that make it just a "little complex". The overall task is called `data assimilation` as information from the data is assimilated or incorporated into the model allowing prediction (Evensen 2009).

To address the question, we generated a "data set" by solving Eq. (2.8) with $\sigma_{\text{data}} = 16.0, A_{\text{data}} = 40.2, B_{\text{data}} = 1.0$ using some randomly selected initial conditions $x(0), y(0), z(0)$ within a cube in (x, y, z) space 20 "units" on a side. We integrated these equations using a standard fourth-order Runge Kutta method with a time step $\Delta t = 0.005$. The first 90,000 points of the three time series were discarded, and the remaining 2^{16} data points were recorded and stored.

We call the observed data $x_{\text{data}}(t)$, and we also know $y_{\text{data}}(t)$ and $z_{\text{data}}(t)$, we store them away for later comparison with our estimates; we treat them as `unobserved` variables. We inform our waterwheel flow model only of the values of $x_{\text{data}}(t)$.

This kind of exploration of data assimilation in which we generate the data from a known dynamical source and then investigate how well our methods work to produce accurate estimates of the known (unobserved) orbits and known parameters of the dynamical source is called a "twin experiment", (Faragó et al. 2005). It is always important to exercise a twin experiment on any model of observed data.

Our method for estimating the parameters and the state variables within and at the end an data acquisition window $[0 \leq t \leq T]$ is to create a model of the data

$$\begin{aligned} \frac{\mathrm{d}X(t)}{\mathrm{d}t} &= \sigma(Y(t) - X(t)) \\ \frac{\mathrm{d}Y(t)}{\mathrm{d}t} &= -Y(t) + AX(t) - Z(t)X(t) \\ \frac{\mathrm{d}Z(t)}{\mathrm{d}t} &= -BZ(t) + X(t)Y(t) \end{aligned} \tag{2.9}$$

and for a selection of initial conditions $X(0), Y(0), Z(0)$ and parameters σ, A, B generate a solution $X(t; X(0), Y(0), Z(0), \sigma, A, B)$ to compare to our data set $x_{\text{data}}(t)$. The comparison is made through a least-squares evaluation of the difference

$$\frac{1}{N}\sum_{n=1}^{N}(x_{\text{data}}(t_n) - X(t_n))^2, \tag{2.10}$$

$t_n = n\Delta t$. We then seek a minimum of this comparison metric as we vary $X(0), Y(0), Z(0)$ and σ, A, B. This choice of comparison metric assumes implicitly that the errors in the model output $X(t)$ relative to the known data $x_{\text{data}}(t)$ are Gaussian distributed.

When we try this comparison using the model we have in Eq. (2.9), this does not work well, even though we know the actual model in this twin experiment. To see the barrier to success, we fixed A, B to the values used in generating the data, fixed $X(0), Y(0), Z(0)$ to be different from that used in generating the data, and evaluated Eq. (2.10) for various values of σ. In Fig. 2.2 the green curve labeled with $K = 0$ is Eq. (2.10) with $X(t)$ directly from the model. It shows many local minima in the σ dependence of our comparison function. A search over the value of σ in producing the model output $X(t)$ is seen by eye to be very unlikely to yield the correct value ($\sigma = 16.0$) as among the many local minima; the region near $\sigma = 16.0$ would probably not be reached by a search algorithm unless we know where to start it beforehand. Further, a search algorithm is quite likely to fall into one of the numerous local minima, and usually it does.

The origin of the many local minima acting as an impediment to the estimation of σ is the incoherence between the model system output $X(t)$ and the data $x_{\text{data}}(t)$: both of them are chaotic and have different initial conditions. The instabilities in the nonlinear dynamical system that give rise to the chaos of these orbits, also lead these two orbits $X(t)$ and $x_{\text{data}}(t)$, to move differently around the same attractor of the model dynamical system with no correlation between each other as they emanate from different initial conditions, even if σ is set equal to the correct value. The two orbits $X(t)$ and $x_{\text{data}}(t)$ are not synchronized with each other.

In a manner we discuss in detail below, we can potentially tame the lack of synchronization by adding to the model equations a term that induces the model output $X(t)$ to move along with the data $x_{\text{data}}(t)$. We achieve this by changing the model equations to

$$\begin{aligned}
\frac{\mathrm{d}X(t)}{\mathrm{d}t} &= \sigma(Y(t) - X(t)) + K(x_{\text{data}}(t) - X(t)) \\
\frac{\mathrm{d}Y(t)}{\mathrm{d}t} &= -Y(t) + AX(t) + Z(t)X(t) \\
\frac{\mathrm{d}Z(t)}{\mathrm{d}t} &= -BZ(t) + X(t)Y(t);
\end{aligned} \tag{2.11}$$

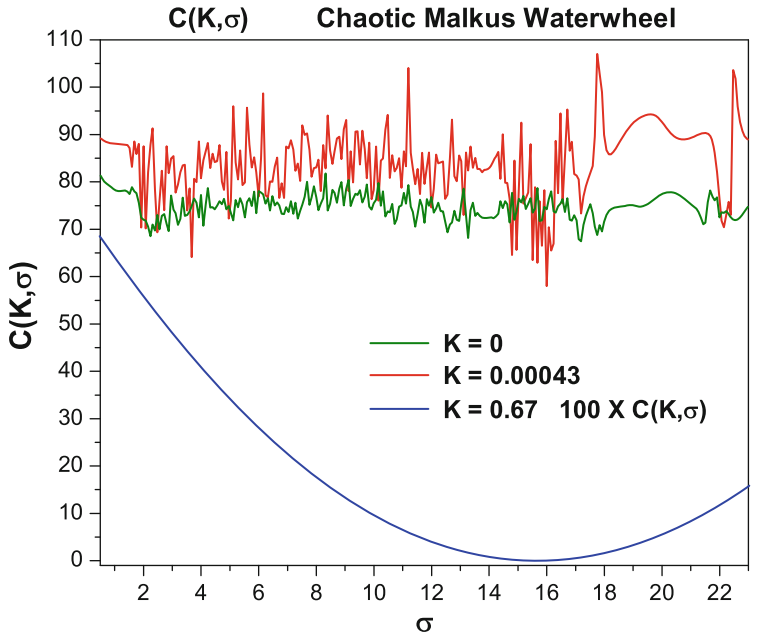

Fig. 2.2 We display the cost function $C(K,\sigma)$ Eq. (2.10) as a function of the parameter σ in the dynamical equations (2.11) for various values of K, the coupling strength of the data into the model. Other parameters in Eq. (2.11) are held fixed. The minimization of $C(K,\sigma)$ is expected to allow the estimation of σ, however, we see for small K; $K = 0, K = 0.00043$ the incoherent interference of the ingredients of $C(K,\sigma)$, $x_{\text{data}}(t_n)$, and $X(t_n)$ yields many local minima with none apparently at the known value $\sigma = 16$. When the coupling of the data into the model is larger, here $K = 0.67$, the data and the model output synchronize, and the σ variation of $C(K,\sigma)$ is now smooth, allowing an accurate estimation of this parameter. The cost function $C(K,\sigma)$ decreases as $\approx K^{-2}$, so the last curve here (in *blue*) is multiplied by 100 so it can be displayed on an equal footing for the smaller values of K

$K \geq 0$. This connects information in $x_{\text{data}}(t)$ directly into the model equations. In the red and blue curves in Fig. 2.2 we see that when K is very small, the incoherence between $X(t)$ and $x_{\text{data}}(t)$ remains, while when we select a larger value of $K = 0.67$, the picture changes completely and the dependence of Eq. (2.10) on σ is smooth with a clear, easy to reach minimum at the correct value of σ. The key here is that the two chaotic signals $X(t)$ and $x_{\text{data}}(t)$ have synchronized for this value of K, and the previous incoherence has transmuted into coherence.

This device of adding a term to drive the two time series together is known in geophysics as "nudging" and has been used in nonlinear dynamics to analyze synchronization between chaotic systems (So et al. 1994; Parlitz 1996; Parlitz et al. 1996; Maybhate and Amritkar 1999; Sakaguchi 2002; Dochain 2003; Voss et al. 2004; Parlitz and Yu 2011; Huang 2004; Konnur 2003). It is important to note here that had we used this approach on the "$z(t)$" components, it would not have worked (Pecora and Carroll 1990), a feature not recognized in the geophysics

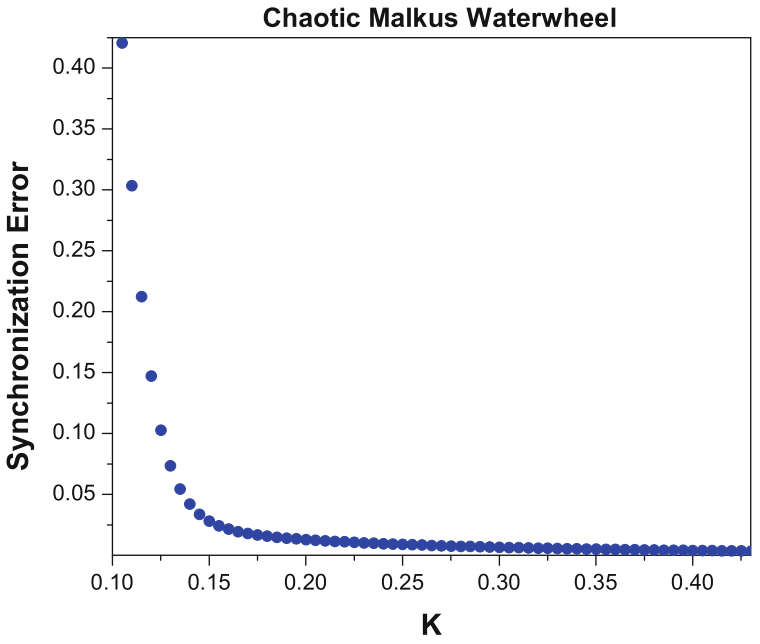

Fig. 2.3 The dependence of the synchronization error $C(K, \sigma)$ between two solutions of the dynamical equations (2.11) where all parameters are the same in the two solutions, but they differ in their initial conditions. For $K = 0$ the two chaotic solutions develop independently in time. They do not synchronize because of the instability in the "communications channel". When the solutions are coupled as in Eq. (2.11) strongly enough to reduce the positive conditional Lyapunov exponent (*CLE*) at $K = 0$ below zero, the solutions synchronize, and as seen in Fig. 2.2, estimation based on minimizing the synchronization error is possible

literature, but quite relevant to our discussion as we move through this book. The new constant K we have introduced has no physical meaning, it is only a device to force the model output to the observed $x_{\text{data}}(t)$.

To orient ourselves with respect to the idea of synchronization of two chaotic time series, we plot in Fig. 2.3 a different characteristic of the function $C(K, \sigma)$:

$$C(K, \sigma) = \frac{1}{N} \sum_{n=1}^{N} (\mathbf{x}_{\text{data}}(t_n) - X(t_n))^2, \tag{2.12}$$

and we ask when this metric goes to zero as a function of K with σ fixed along with A and B while the initial conditions for the data and the model remain different. So, we solve the Eq. (2.11) with $\sigma = 16$ for various K and plot this least-squares metric of similarity of the model $X(t)$ time series and the data $x_{\text{data}}(t)$ times series. The outcome of this is in Fig. 2.3 where $C(K, \sigma)$ is displayed versus K; we see a region without synchronization, $C(K, \sigma) \neq 0$, followed by persistent synchronization for larger K, $C(K, \sigma) \approx 0$.

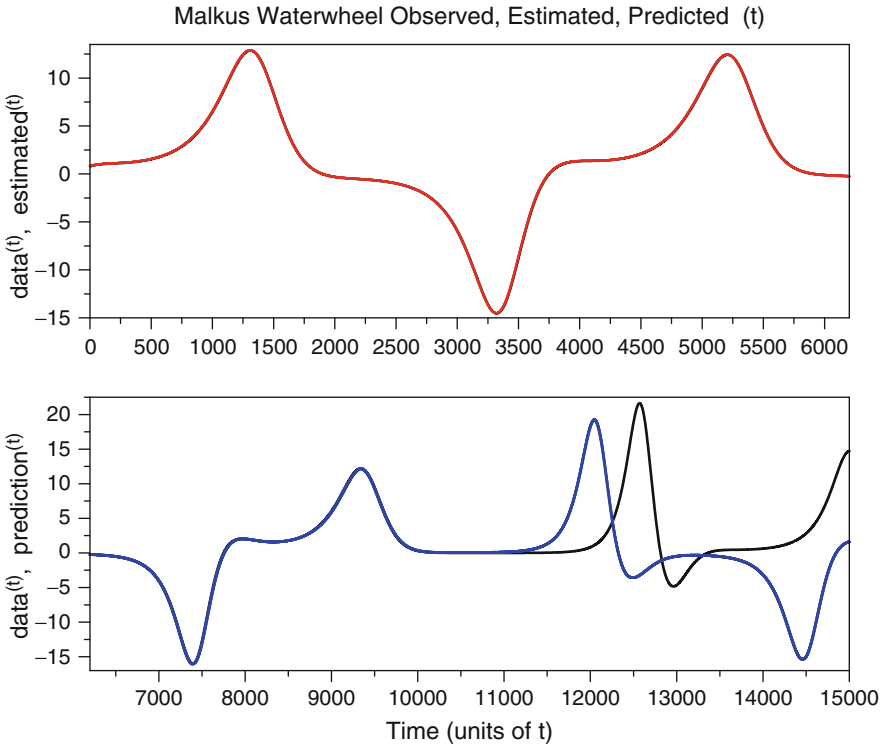

Fig. 2.4 *Top Panel* Estimation of the parameters and states for the chaotic Malkus waterwheel. Using 6,000 data points separated by $\Delta t = 0.01$ we minimized Eq. (2.13) subject to the dynamical equations (2.14). The parameters were estimated very accurately. We display the observed angular velocity comprising the data in *black* and the estimated angular velocity in *red*. The other two, unobserved, dynamical variables were also estimated throughout the observation window, but they are not displayed. *Bottom Panel* Using the estimates of the parameters and the observed and unobserved state variables at the end of the observation window, we integrated the dynamical waterwheel equations forward in time to predict the behavior of the waterwheel. The known (*black*) and predicted solutions (*blue*) agree well until about time 11,500. The estimated states and the known data at the beginning of prediction differ by small amounts, and this difference grows because of the positive Lyapunov exponent in this system. This leads to an exponential separation of the prediction and the data consistent with the value of the largest positive Lyapunov exponent

When we have synchronized the data with the appropriate function of the model output, here simply $x_{\text{data}}(t) \approx X(t)$, we expect to be more successful in our estimation of the parameters σ, A, B as well as the unobserved states $y(t), z(t)$. There is one more idea that is helpful to introduce into the discussion. In nonlinear dynamical systems such as our waterwheel equations, the instability associated with chaotic behavior of the orbits is inhomogeneous in the state space $\{X(t), Y(t), Z(t)\}$ location of the orbit (Abarbanel 1996; Kantz and Schreiber 2004). The magnitude of the instability curing factor we call K may need to be larger or smaller in various parts of the orbit. To accomplish this in a natural manner, we replace the constant

K by a function of time $K \rightarrow u(t) \geq 0$ and add a cost for the regularization accomplished by $u(t)$ to the comparison function. $u(t)$ utilizes t as an indicator where one is located in $\{X, Y, Z\}$ space and allows the strength of the push on the model toward the data to vary locally.

We then seek to minimize

$$C = \frac{1}{N} \sum_{n=1}^{N} \left\{ (\mathbf{x}_{\text{data}}(t_n) - X(t_n))^2 + u(t_n)^2 \right\}, \tag{2.13}$$

subject to our model equations as a constraint:

$$\begin{aligned} \frac{\mathrm{d}X(t)}{\mathrm{d}t} &= \sigma(Y(t) - X(t)) + u(t)(x_{\text{data}}(t) - X(t)) \\ \frac{\mathrm{d}Y(t)}{\mathrm{d}t} &= -Y(t) + AX(t) + Z(t)X(t) \\ \frac{\mathrm{d}Z(t)}{\mathrm{d}t} &= -BZ(t) + X(t)Y(t), \end{aligned} \tag{2.14}$$

We have carried out this numerical optimization task, and the resulting estimates of the fixed parameters are very accurate: $\{\sigma_{\text{est}} = 16.00002, A_{\text{est}} = 40.20018, B_{\text{est}} = 0.999993\}$. In the Top Panel of Fig. 2.4 we display the estimated $\omega_{\text{est}}(t)$ in red and the known $\omega_{\text{data}}(t)$ as a black line. In addition to the estimated parameters, we estimate the unobserved variables $Y(t), Z(t)$ over the observation window $[0, T]$. This completes the model (2.9) as we now have estimates of the fixed parameters $\{\sigma, A, B\}$.

Using the estimates of σ, A, B and $X(T), Y(T), Z(T)$ we integrated the dynamical equations (2.9) with $u(t) = 0$ forward for $t > T$. In the Bottom Panel of Fig. 2.4 the predicted $\omega_{\text{pred}}(t)$ is shown in blue along with the known data still in black.

This dynamical system is chaotic, and the estimates of σ, A, B, $X(T)$, $Y(T)$, and $Z(T)$ inevitably have errors, small though they may be in this example. This leads to two solutions to the waterwheel equations with slightly different initial conditions: one is the "data" here, and the other is the predicted $\omega(t)$ using the values estimated at $t = T$. As in all chaotic systems these orbits with different initial conditions diverge due to the intrinsic instability in the dynamics. This too we see in the departure of $\omega_{\text{data}}(t)$ and $\omega_{\text{pred}}(t)$ from one another (Abarbanel 1996; Kantz and Schreiber 2004).

We do not show the fact that the estimates of $\{Y(t), Z(t)\}$ are also quite accurate when we compare them to the known values in the "data" set. We could infer this by the accuracy of the predictions. In an experimental setting, not twin experiment simulations, we would not be able to make this comparison, and we would have to rely on the quality of predictions for observed state variables as our sole metric for success in estimations.

We turn now to a discussion of another "slightly complex" example drawn from nonlinear electronic circuits. We will repeat many of the steps presented in the waterwheel example and give some elaboration on the tools we utilized until now.

We did not discuss the numerical methods used to perform the minimization of the comparison function nor will we in the circuit example. We save this for the chapters following the general formulation of the data assimilation problem. In this chapter our goal is to provide a tempting sense of the flavor of the arguments and the issues and save the full meal for later.

2.2 The Colpitts Oscillator

The mechanical example of the waterwheel is not isolated or special. The issues raised in addressing data assimilation within it appear in numerous examples of nonlinear model dynamics describing observed data. To illustrate many of the same questions, we turn to an example from an electrical circuit which has a nonlinear circuit element: a bipolar transistor. In this example, we have a quite different kind of twin experiment as we generate the data from a laboratory analog circuit we built for the purpose as well as the opportunity for a fully numerical simulation of the circuit. The issues encountered in the mechanical example, which also has been realized in experiments, are now seen to be associated with the actual response of a physical system and not just a numerical simulation as before.

Our circuit is a nonlinear oscillator invented by Colpitts in the 1920s (Colpitts 1927). It has linear RLC elements, familiar from textbooks and undergraduate laboratory sessions, as well as a bipolar transistor (not part of Colpitts' original design). There are three state variables and their associated nonlinear ordinary differential equations describe this circuit; it expresses chaotic behavior (Abarbanel 1996; Kantz and Schreiber 2004; Kennedy 1994) when the forcing of the circuit becomes strong enough.

The question we propose about the circuit is whether a measured time series of one of its three independent dynamical variables can be passed to a model of the circuit and used within the model to estimate unknown parameters in the circuit model along with estimates of the remaining two `unobserved` model state variables. In the case of the bipolar junction transistor, we have (at least) two competing models of the nonlinear element in the circuit, and we ask if we can differentiate between them in the same setting: one measured dynamical variable passed to the model.

In practice, we are able to measure everything about the Colpitts circuit, so it provides a welcome testbed where we can examine the efficacy of any data assimilation method, and the key question is only to distinguish between two relatively well-tested models of the nonlinear bipolar transistor (Ebers and Moll 1954; Gummel and Poon 1970).

The circuit is easy enough to build (Quinn et al. 2009; De Feo and Maggio 2003), so we have the opportunity to examine it both experimentally and theoretically, and find that standard methods of `data assimilation` in the geophysical literature (Evensen 2009) cannot succeed when the oscillations of the Colpitts

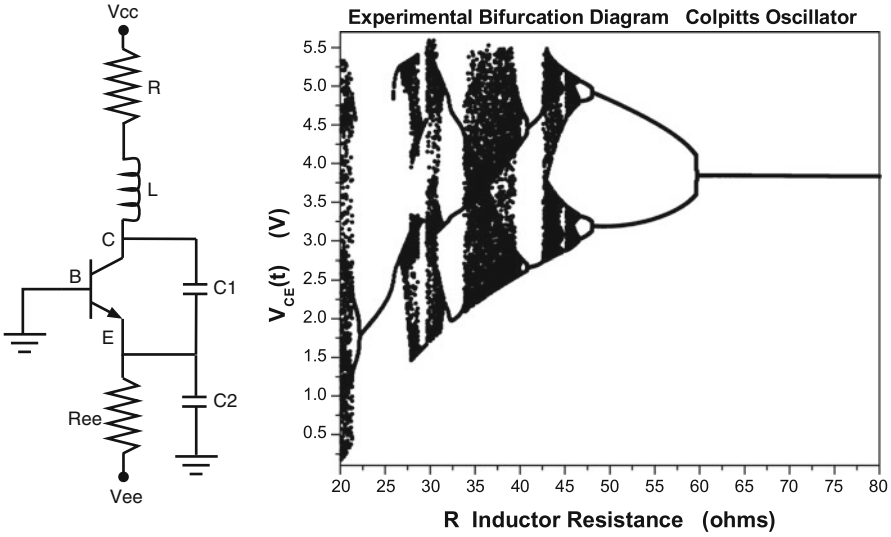

Fig. 2.5 *Left* The circuit diagram for the Colpitts circuit. The RLC circuit elements are well known from linear circuit theory (Purcell 1965), so the model of the circuit is essentially a choice of model for the bipolar transistor. *Right* The bifurcation diagram for the Colpitts circuit. As the resistance R is decreased the circuit is driven more strongly and its behavior goes from a fixed point for $R \geq 60$ ohms to periodic behavior and then chaotic oscillations for smaller values of R (Kennedy 1994). We operated this circuit with $R \approx 35$ ohms

circuit are chaotic. The techniques used to resolve this difficulty illuminate issues encountered in more complicated settings of substantial interest.

The Colpitts oscillator circuit diagram is shown in Fig. 2.5, (**Left**). Using the standard rules for analyzing electrical circuits (Purcell 1965), we determine there are three independent dynamical variables for the Colpitts circuit, and we choose them to be $V_{\rm CE}(t)$, the voltage at the collector relative to the emitter; $V_{\rm E}(t)$, the voltage at the emitter relative to ground; and $I_{\rm L}(t)$, the current through the inductor.

Our goal in this data assimilation example is to use measurements of $V_{\rm E}(t_n)$ taken at the times $t_n = n\Delta t;\ \Delta t = 10\,\mu$s for $n = 0, 1, \ldots, N = 1,000$ to estimate the fixed parameters in the model of the circuit and to estimate the `unobserved` $V_{\rm CE}(t_n)$ and the `unobserved` $I_{\rm L}(t_n)$ over the same observation period. We then wish to employ the estimated fixed parameters and the estimated values of the state variables $\{V_{\rm E}(T), V_{\rm CE}(T), I_{\rm L}(T)\}$ at $T = t_N = N\Delta t$ to predict the state of the system for $t > T$ using the model dynamics.

We know our estimates of the parameters and the state variables will not be perfectly accurate; therefore, if there is chaos in the system, it will exhibit itself both as inaccuracies in our prediction beyond T and as an impediment to our estimating the parameters and states of the circuit. In this section we will examine the circuit in an operating region where the driving forces are large enough to produce chaotic oscillations.

In thinking of the use of information in our data as a communications problem we would naturally seek to **synchronize** the observations with $V_E(t)$ from our model of the circuit. As the only nonlinear element of the circuit is the bipolar transistor, we can say our effort to estimate parameters and states, and predict with that information, is in essence testing our choice of a model for the physics of the transistor. In this regard, we will find that one standard model for bipolar transistors does very well in allowing estimates for states and parameters using data from this circuit, and another model does even better.

2.2.1 Colpitts Circuit Equations

The driving force in the circuit operation is the bias voltage across the circuit (Fig. 2.5 (**Left**)). When that driving force is small, the circuit dynamics yields a fixed point in state space. As that driving voltage is increased we encounter bifurcations (Fig. 2.5 (**Right**)) to limit cycle periodic oscillations of the circuit. When that voltage increases further, the circuit expresses chaotic oscillations.

The dynamics is described by three coupled first-order differential equations, which can be obtained directly using Kirchoff's laws. These are

$$\begin{aligned} C_1 \frac{\mathrm{d}V_{CE}(t)}{\mathrm{d}t} &= I_L(t) - I_C(V_E(t)), \\ C_2 \frac{\mathrm{d}V_E(t)}{\mathrm{d}t} &= I_L(t) - \frac{V_E(t) - V_{ee}}{R_{ee}} + I_B(V_E(t)), \\ L \frac{\mathrm{d}I_L(t)}{\mathrm{d}t} &= V_{CC} - V_E(t) - V_{CE}(t) - RI_L(t), \end{aligned} \tag{2.15}$$

where $I_C(V_E(t))$ and $I_B(V_E(t))$ are the currents going into the collector and base of the transistor.

A key ingredient is the model specifying the transistor currents. We begin with a simplified version of the Ebers–Moll equations (Ebers and Moll 1954):

$$I_C(V_E) = I_s \exp(\frac{-V_E}{V_T}), \tag{2.16}$$

$$I_B(V_E) = \frac{I_C(V_E)}{\beta_F}, \tag{2.17}$$

where β_F is the forward current gain and I_s is the reverse saturation current; both are properties of the particular transistor. $V_T = kT/e$ is the thermal voltage. These equations are nonlinear, and this is what leads to interesting behavior of the circuit.

The RLC circuit, without the transistor, has a natural period of $T_0 = 2\pi\sqrt{\frac{LC_1C_2}{C_1+C_2}}$, and has a damping time of $\tau = 2L/R$. Changing the quality factor, $Q = \pi\tau/T_0 = 2\pi L/RT_0$, can change the circuit behavior. The other parameters also matter, but

here we focus on only varying Q by changing the value of R and keeping all the other parameters fixed. As R decreases, Q increases as does the current flowing into the circuit from the terminal at V_{ee}.

The circuit was built out of common components. We used parts with the following values: $C_1 = 7.44\,\mu$F, $C_2 = 7.23\,\mu$F, $L = 11.74$ mH, and $R_{ee} = 0.392$ K ohms, and in the power supply, we took $V_{CC} = 5.03$ V, $V_{ee} = -5.10$ V. We also chose a 100 ohm potentiometer in the circuit to adjust the parameter R. R includes the resistance of the potentiometer plus the resistance of the inductor, which was not negligible. For the bipolar transistor we selected a 2N2222 BJT small-signal transistor. The fundamental frequency of this oscillator is around 700 Hz.

In Fig. 2.5 (**Right**) we display the bifurcation diagram for the Colpitts circuit. We show the asymptotic (in time) values taken by the voltage $V_{CE}(t)$ as a function of the resistance across the inductor. From a fixed point near $V_{CE} \approx 4$ V above $R \approx 60$ ohm, we see a transition to a periodic orbit and then to windows of chaos. The chaotic oscillations used in the analysis were measured with R set at $R = 35$ ohm. To construct this bifurcation diagram, we integrated the circuit equations using a range of values for R, and plotted the value of V_{CE} only at the times t where both $V_E(t) = V_{th}$ and $\frac{dV_E(t)}{dt} > 0$. This defines a Poincare' section (Abarbanel 1996; Kantz and Schreiber 2004; Strogatz 1994) of the three dimensional state space. The threshold value of $V_{th} = -0.6$ V was chosen because that is about where the transistor switches smoothly from "on" to "off".

We recorded all three dynamical variables $V_C(t), V_{CE}(t)$, and $I_L(t)$ from the circuit when it was operating in a chaotic regime. These state variables were measured every $\Delta t = 10\,\mu$s. The orbit traces out a strange attractor in the three-dimensional state space. A two-dimensional projection of the experimental attractor is shown in Fig. 2.6 (**Top**). This is compared to the model projected phase space which is calculated by numerically integrating the circuit equations using parameters from the actual circuit (Fig. 2.6 (**Bottom**)).

2.2.2 *Estimation with Chaotic Signals*

We return to the experimental circuit in a moment. By rescaling the circuit variables, we can cast Eq. (2.15) into the dimensionless form

$$\begin{aligned}
\frac{dy_1(t)}{dt} &= \alpha_D y_2(t) \\
\frac{dy_2(t)}{dt} &= -\gamma_D(y_1(t) + y_3(t)) - q_D y_2(t) \\
\frac{dy_3(t)}{dt} &= \eta_D(y_2(t) + 1 - \exp(-y_1(t))),
\end{aligned} \tag{2.18}$$

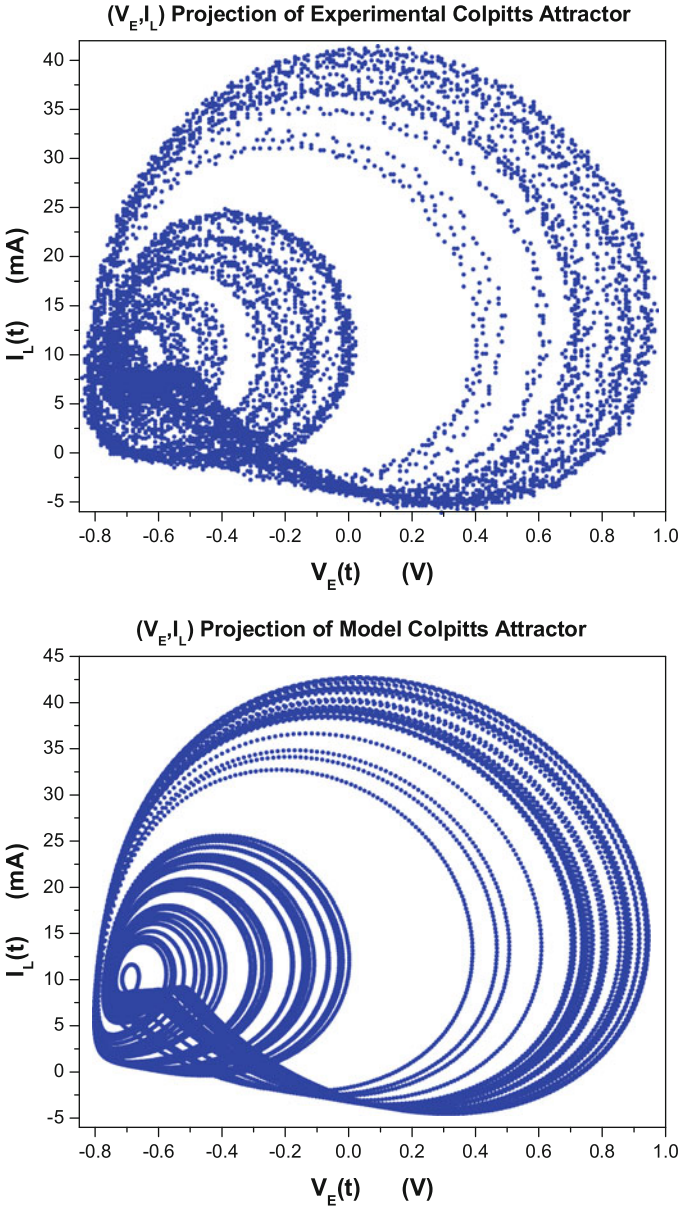

Fig. 2.6 Two-dimensional projection of the attractor: *Top Panel* Observed $V_{\mathrm{E}}(t)$ versus $I_{\mathrm{L}}(t)$. *Bottom Panel* $V_{\mathrm{E}}(t)$ versus $I_{\mathrm{L}}(t)$ from integrating the model equations. The signals were sampled every $\Delta t = 10\,\mu$s

in which the control parameter α_D plays the role of the driving force. For small α_D we have a fixed point for the trajectories of the circuit. For $\alpha_D \approx 5$ we have chaotic solutions.

Now we generate data $\{y_1(t), y_2(t), y_3(t)\}$ using the parameters ($\alpha_D = 5.0$, $\gamma_D = 0.0797, q = 0.6898, \eta_D = 6.2723$) and a selected set of initial conditions $\{y_1(0), y_2(0), y_3(0)\}$. Then we ask when we can accurately estimate the parameters $\{\alpha, \gamma, q, \eta\}$ and the unobserved state variables $\{x_2(t), x_3(t)\}$ in a model of the circuit given by

$$\begin{aligned}
\frac{\mathrm{d}x_1(t)}{\mathrm{d}t} &= \alpha x_2(t) \\
\frac{\mathrm{d}x_2(t)}{\mathrm{d}t} &= -\gamma(x_1(t) + x_3(t)) - q x_2(t) \\
\frac{\mathrm{d}x_3(t)}{\mathrm{d}t} &= \eta(x_2(t) + 1 - \exp(-x_1(t))).
\end{aligned} \tag{2.19}$$

We want to perform this estimation when only the data from the state variable $y_1(t_n); t_n = n\Delta t$ ($\Delta t = 10\,\mu$s) is presented to the model.

If we observe $y_1(t_n) = y_1(n)$ at N points in time, we would expect that minimizing a distance between the data $y_1(n)$ and the model output $x_1(t_n) = x_1(n)$ would serve to allow estimation of the parameters $\{\alpha, \gamma, q, \eta\}$ and the initial conditions $\{x_1(0), x_2(0), x_3(0)\}$ that produce a sequence of $x_1(t_n; \alpha, \gamma, q, \eta, x_1(0), x_2(0), x_3(0))$ best matched to the data $y_1(t_n)$. If the model is correct, the cost function should become quite small as the model is "tuned" to the data seen as a transmitter. In this situation, one would expect that predictions using the model equations with estimated values of the parameters and the full complement of state variables would be accurate until the chaotic nature of the orbits destabilized the predictions. If the model is incorrect, we may see a small cost function, but predictions are certain to be inaccurate. This emphasizes the role of predictions in establishing some validity for a model developed to describe experiments.

To this end, we employ a least-squares comparison distance between the data and the model output. This means we anticipate that minimizing the "cost function"

$$\frac{1}{2N} \sum_{l=0}^{N-1} (y_1(l) - x_1(l))^2, \tag{2.20}$$

subject to Eq. (2.19) should yield good estimates for parameters and state variables.

We examine a slightly different equation than Eq. (2.19) by injecting information from the data directly into the model using

$$\begin{aligned}
\frac{\mathrm{d}x_1(t)}{\mathrm{d}t} &= \alpha x_2(t) + k(y_1(t) - x_1(t)); \; k \geq 0 \\
\frac{\mathrm{d}x_2(t)}{\mathrm{d}t} &= -\gamma(x_1(t) + x_3(t)) - q x_2(t) \\
\frac{\mathrm{d}x_3(t)}{\mathrm{d}t} &= \eta(x_2(t) + 1 - \exp(-x_1(t))),
\end{aligned} \tag{2.21}$$

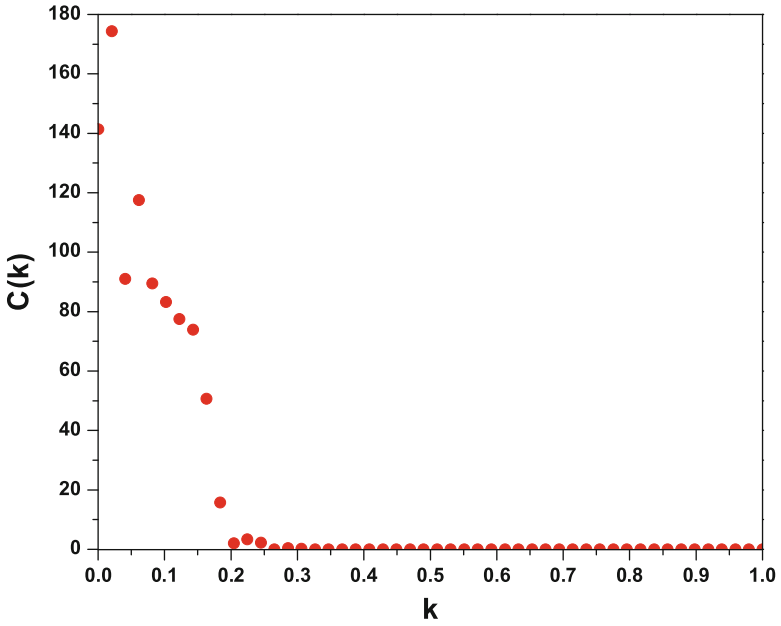

Fig. 2.7 We expect that when the data $y_1(n)$ presented to a model of the Colpitts circuit synchronizes with the corresponding output of the model $x_1(n)$, the ability to accurately estimate parameters and all states in the model will be maximized. Here we display the synchronization error $C(\eta, k) = \frac{1}{2N}\sum_{n=0}^{N-1}(y_1(n) - x_1(n))^2$ as a function of the coupling k between data and a model of the Colpitts circuit (Eq. (2.21))

which becomes Eq. (2.19) when $k = 0$. If we fix the three parameters $\{\alpha, \gamma, q\}$ and seek to estimate only η, then

$$C(\eta, k) = \frac{1}{2N}\sum_{n=0}^{N-1}(y_1(n) - x_1(n))^2 \tag{2.22}$$

should allow us to do that. The information on η and the unknown initial conditions $\{x_1(0), x_2(0), x_3(0)\}$ is contained in $x_1(t_n)$ from the solution to the model equations. The role of the coupling $k \geq 0$ is to convey information about the data $y_1(n)$ to the model equations. The term $k(y_1(t) - x_1(t))$ drives the model variable $x_1(n)$ to the data, and as $k \to \infty$, $|x_1(n) - y_1(n)| \to \frac{1}{k}$. $C(\eta, k) = \frac{1}{2N}\sum_{n=0}^{N-1}(y_1(n) - x_1(n))^2$ correspondingly decreases in magnitude as k^{-2}.

In the spirit of our overview discussion about synchronizing the data and the model, we ask for what values of k are the data $y_1(n)$ and the model output $x_1(n)$ synchronized, namely, $x_1(n) \approx y_1(n)$. For such k the cost function (2.22) should be very small. In Fig. 2.7 we show

$$C(\eta, k) = \frac{1}{2N}\sum_{n=0}^{N-1}(y_1(n) - x_1(n))^2, \tag{2.23}$$

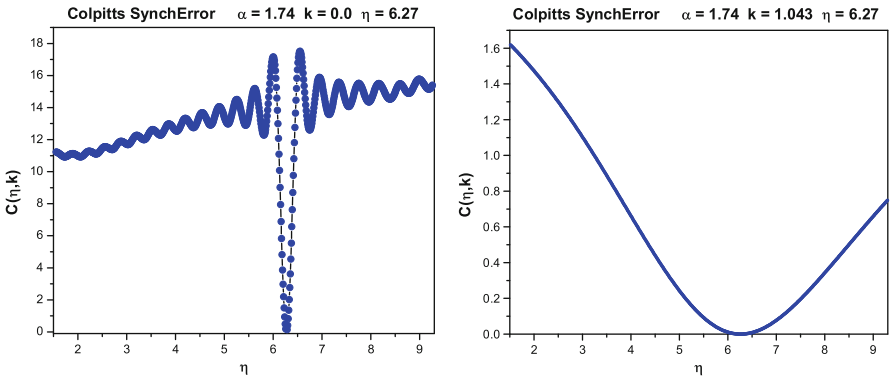

Fig. 2.8 The cost function $C(\eta, k)$ Eq. (2.23) for the Colpitts oscillator, Eq. (2.21), when the driving parameter is $\alpha = 1.74$. The circuit expresses limit cycle oscillations for this value of α. *Left* The coupling is $k = 0$. There is a global minimum at the value of the circuit parameter $\eta = 6.27$ which was used in the generation of the data. There are many local minima as well suggesting that knowledge of the correct value of η might be required to search the cost function $C(\eta, k = 0)$ for an accurate estimate. *Right* The coupling is $k = 1.043$ where the data signal $y_1(n)$ and the model output $x_1(n)$ are synchronized. Now there is a clear minimum for the cost function $C(\eta, k = 1.043)$ allowing accurate estimation of η

as a function of k. Note that at $k = 0$ the model output and the data are `not` synchronized, suggesting one might not be able to estimate η or other parameters in that situation. The model and the data do synchronize for $k \geq 0.25$, and we might anticipate that when the coupling between the data stream $y_1(n)$ and the model is large enough, good estimations will occur.

To further examine this, we now look at $C(\eta, k)$ for two different values of α and two values of k: $k = 0$ and $k = 1.043$. This is one value, $k = 0$, where synchronization does not occur, and one value, $k = 1.043$, for which synchronization is present.

We first select $\alpha = 1.74$ which is in a region where a limit cycle (a periodic orbit) is present. In Fig. 2.8 (**Left**) we see that at $k = 0$ there is a distinct minimum at $\eta \approx 6.27$, though searching for it starting from an initial guess well away from the correct value might be impeded by the local minima that surrounded the correct minimum. If we increase k to $k = 1.043$, the local minima are absent (Fig. 2.8 (**Right**)), and the correct minimum where $\eta = \eta_D$ is visible and easily reached via any standard search algorithm.

When $\alpha = 5.0$ and the orbits are chaotic, we see the outcome shown in Fig. 2.9. In the **Left Panel** $k = 0$ and the two signals $y_1(n)$, the data, and $x_1(n)$ are not synchronized. This produces a ragged interference pattern where multiple local minima are present without any minimum indicating a value where $\eta = \eta_D$ can be located. However, in the right panel when $k = 1.043$, and the orbits of the data and the model output are synchronized, we see that the function $C(\eta, k)$ has become smooth, and the correct minimum is easy to locate by eye or with any sensible search algorithm.

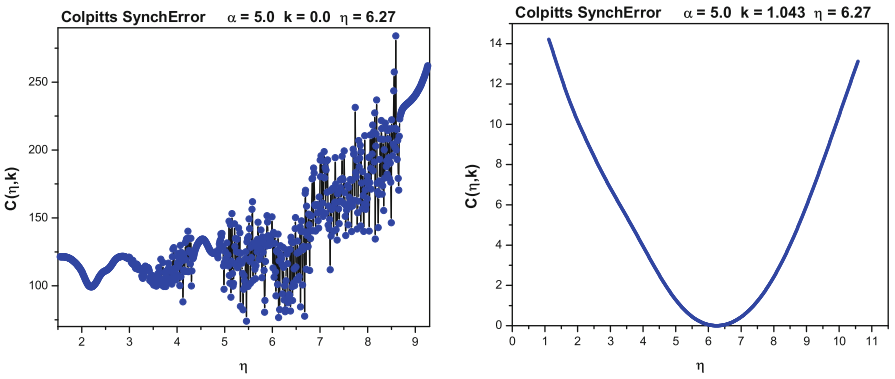

Fig. 2.9 The cost function $C(\eta, k)$, Eq. (2.23), for the Colpitts oscillator, Eq. (2.21), when the driving parameter is $\alpha = 5.0$. The circuit expresses chaotic oscillations for this value of α. *Left* The coupling is $k = 0$ where the data signal $y_1(n)$ and the model output $x_1(n)$ are not synchronized. There are many local minima of $C(\eta, k = 0)$ none of which appear to be at the value of the circuit parameter $\eta = 6.27$ which was used in the generation of the data. *Right* The coupling is $k = 1.043$ where the data signal $y_1(n)$ and the model output $x_1(n)$ are synchronized. Now there is a clear minimum for the cost function $C(\eta, k = 1.043)$ allowing accurate estimation of η

2.2.3 *Instability of the Synchronization Manifold*

Autonomous Systems: Lyapunov Exponents

How are we to understand the transition from ragged multiple local minima in $C(\eta, k)$ when $k = 0$ to the regularized, smooth $C(\eta, k)$ for larger k? The key lies in the *conditional* Lyapunov exponents (CLEs) of the model dynamics (Pecora and Carroll 1990). These are a generalization of the usual Lyapunov exponents for a nonlinear dynamical system (Abarbanel 1996; Kantz and Schreiber 2004).

Lyapunov exponents are found by considering a discrete time $t_n = t_0 + n\Delta t$, $\mathbf{x}(t_n) = \{x_1(t_n), x_2(t_n), x_3(t_n)\}$ version of the D-dimensional dynamics, $\mathrm{d}\mathbf{x}(t)/\mathrm{d}t = \mathbf{F}(\mathbf{x}(t), \mathbf{p})$,

$$\mathbf{x}(t_{n+1}) = \mathbf{x}(n+1) = \mathbf{f}(\mathbf{x}(t_n), \mathbf{p}) = \mathbf{f}(\mathbf{x}(n), \mathbf{p}), \tag{2.24}$$

and asking how $\mathbf{x}(n+1)$ varies with small changes in $\mathbf{x}(0)$. The $\mathbf{p}$ are parameters in the dynamics, and $\mathbf{f}(\mathbf{x}, \mathbf{p})$ is the discrete time version of the model differential equation. Lyapunov exponents are usually defined for autonomous dynamics where $\mathbf{f}(\mathbf{x}, \mathbf{p})$ is not explicitly dependent on time. In the present case this means $k(y_1(t) - x_1(t))$ is absent, so $k = 0$.

We are interested in the response of an orbit $\{\mathbf{x}(0), \mathbf{x}(1), \ldots, \mathbf{x}(M), \ldots\}$ of the discrete time dynamics to a perturbation at t_0. The quantity we want to analyze is the D-by-D-dimensional Jacobian matrix $(a, b = 1, 2, 3, \ldots, D)$

$$\frac{\partial x_a(M+1)}{\partial x_b(0)}, \tag{2.25}$$

as this tells us quantitatively how a small change in the state of a system at t_0 leads to a change in the orbit some time later. This Jacobian matrix satisfies the recursion relation

$$\begin{aligned}\frac{\partial x_a(M+1)}{\partial x_b(0)} &= \frac{\partial f_a(\mathbf{x}(M),\mathbf{p})}{\partial x_c(M)}\frac{\partial x_c(M)}{\partial x_b(0)} \\ &= \mathbf{Df}(\mathbf{x}(M),\mathbf{p})_{ac}\frac{\partial x_c(M)}{\partial x_b(0)}.\end{aligned} \tag{2.26}$$

Repeated indices are summed over. The solution to this time-iterated map for the $D \times D$ matrix $\frac{\partial x_a(M+1)}{\partial x_b(0)}$ is

$$\begin{aligned}\frac{\partial x_a(M+1)}{\partial x_c(0)} &= \mathbf{Df}(\mathbf{x}(M),\mathbf{p})_{aa'}\mathbf{Df}(\mathbf{x}(M-1),\mathbf{p})_{a'b'}\cdots\mathbf{Df}(\mathbf{x}(1),\mathbf{p})_{b'c} \\ &= \mathbf{Df}^M(\mathbf{x}(0),\mathbf{p})_{ac},\end{aligned} \tag{2.27}$$

noting that $\frac{\partial x_a(0)}{\partial x_b(0)} = \delta_{ab}$.

The Lyapunov exponents $\lambda_a;\ a = 1,2,\ldots,D$ associated with the dynamics $\mathbf{x}(n) \to \mathbf{x}(n+1)$ are given via the eigenvalues $\mathrm{e}^{\lambda_a};\mathbf{a} = 1,2,\ldots,D$ of the Oseledec matrix (Oseledec 1968)

$$\left(\mathbf{Df}^M(\mathbf{x}(0),\mathbf{p})^T \cdot \mathbf{Df}^M(\mathbf{x}(0),\mathbf{p})\right)^{1/2M}, \tag{2.28}$$

as $M \to \infty$. Oseledec proved that as $M \to \infty$, the λ_a

- Exist
- Are independent of $\mathbf{x}(0)$
- Are independent of the coordinate system in which the matrix is evaluated

If any of the $\lambda_a > 0$, so the eigenvalue is outside the unit circle, the system orbits are chaotic, and small perturbations to an orbit grow exponentially rapidly away from the orbit. When one or more of the $\lambda_a > 0$, orbits are unstable to small perturbations and move around the same system attractor but in a completely different order than in the unperturbed system. No orbits go off to infinity as in linear systems. Stability of the dynamical system requires $\sum_{a=1}^{D}\lambda_a < 0$ (Abarbanel 1996; Kantz and Schreiber 2004).

Driven Systems: Conditional Lyapunov Exponents

When $k \neq 0$, the time-iterated map is not autonomous because of the presence of $y_1(t)$ in the dynamical equations. However, if we consider the unidirectionally coupled data source and the model together as a whole system, the total dynamics

is autonomous. The Lyapunov exponents of the receiver system are called CLEs as they are conditional on the receipt of the driving signal $y_1(n)$ (Pecora and Carroll 1990). The model system is now

$$x_a(n+1) = f_a(\mathbf{x}(n), \mathbf{p}) + \delta_{a1} k(y_1(n) - x_1(n)),\ a = 1, 2, 3, \ldots, D \qquad (2.29)$$

and has a Jacobian $\frac{\partial x_a(n+1)}{\partial x_b(n)} = \mathbf{Df}(\mathbf{x}(n), \mathbf{p})_{ab} - \delta_{11} k$. This Jacobian is to be used in the Oseledec formula, just as in the autonomous case, and for large k can lead to positive Lyapunov exponents becoming negative.

If any of the CLEs is positive, the synchronization manifold $\mathbf{x}(n) = \mathbf{y}(n)$ is unstable. This leads to the multiple local minima in a comparison cost function such as $C(\eta, k)$ (Eq. (2.23)). The ingredients of this cost function $x_1(n)$ and $y_1(n)$ are both chaotic but incoherent with respect to one another when unsynchronized. We see that their interference pattern, captured in the cost function, expresses multiple local minima, none of which may lie at the correct value, when a positive CLE is associated with this lack of synchronization. If one has a single unstable direction, then the role of k is to move the positive CLE to a negative value and stabilize the perturbations to the synchronization manifold introduced by the search procedure in the minimization of the cost function.

The sensitivity to small changes in initial conditions $\mathbf{x}(0)$ is also true for small changes in the parameters $\mathbf{p}$, and when one is searching about in $(\mathbf{p}, \mathbf{x}(0))$ space for a minimum of a cost function, these small perturbations lead, through the instability, to large excursions in the value of the cost function. This is precisely what we see in Figs. 2.8 and 2.9. As k is increased, the CLEs decrease and become negative allowing synchronization and a smooth surface in the parameter or initial condition dependence of a cost function. The fact that we can regularize the behavior of the cost function with the coupling of a single data stream $y_1(n)$ demonstrates that there is only one unstable direction on the synchronization manifold when $y_1(n)$ alone is presented to the model.

2.2.4 Regularized Cost Function

We have just seen that when the data and the model output are chaotic, synchronization does not happen when the two signals are simply compared directly. The sensitivity to initial conditions means that searching for a set of $\mathbf{x}(0)$ for the model intrinsically destabilizes the search.

The remedy presented in the previous sections adds the global coupling $k(y_1(n) - x_1(n))$ to the model dynamical equations. This fails to recognize that the instability on the synchronization manifold is not uniform over the orbits of the dynamics. Also, when the estimation procedure is completed for a fixed value of k, there is a nonzero term $k(y_1(n) - x_1(n))$ remaining in the equations. This term has no physics in it. It was introduced to rid the search procedure of sensitivity to intrinsic instabilities in the dynamics.

Both issues can be addressed by making the coupling k dependent on time, $k \rightarrow u(t)$, and by adding a cost for nonzero values of the coupling to the cost function itself. The first action recognizes that nonzero $u(t)$ maybe be needed in some parts of the system attractor, while $u(t) = 0$ may be adequate in others. The instability on the synchronization manifold is not homogeneous across the manifold. The time label in $u(t)$ serves to indicate one's location on the attractor. The second action allows us, as we achieve synchronization, to remove the time-dependent coupling as part of the variational procedure. We see this as follows:

When we take $k \rightarrow u(t)$, the equations of motion become

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{F}(\mathbf{x}(t), \mathbf{p}) + \mathbf{u}(t)(\mathbf{y}(t) - \mathbf{x}(t)) \; ; \; u(t) \geq 0, \tag{2.30}$$

while we alter the cost function through

$$C(\mathbf{p}) = \frac{1}{2N} \sum_{n=0}^{N-1} (y_1(n) - x_1(n))^2 \rightarrow \frac{1}{2N} \sum_{n=0}^{N-1} [(y_1(n) - x_1(n))^2 + u(n)^2]. \tag{2.31}$$

When we minimize this cost function subject to the equations of motion, we see that as synchronization is achieved, $x_1(n) \approx y_1(n)$, the optimization procedure also sends $u(n) \rightarrow 0$.

2.2.5 *Experimental Colpitts Oscillator Redux*

Regularized Cost Function and Parameter Estimation

We return to the experimental setup for the Colpitts oscillator recognizing we must regulate the search procedure by using the cost function

$$C(R', u) = \frac{1}{2N} \sum_{l=1}^{N-1} [(y_1(n) - x_1(n))^2 + u(n)^2], \tag{2.32}$$

subject to the model dynamical equations

$$\begin{aligned}
C_1 \frac{\mathrm{d}V_{\mathrm{CE}}(t)}{\mathrm{d}t} &= I_{\mathrm{L}}(t) - I_{\mathrm{C}}(V_{\mathrm{E}}(t)), \\
C_2 \frac{\mathrm{d}V_{\mathrm{E}}(t)}{\mathrm{d}t} &= I_{\mathrm{L}}(t) - \frac{V_{\mathrm{E}}(t) - V_{\mathrm{ee}}}{R_{\mathrm{ee}}} + I_{\mathrm{B}}(V_{\mathrm{E}}(t)) + u(t)(V_{\mathrm{E-data}}(t) - V_{\mathrm{E}}(t)), \\
L \frac{\mathrm{d}I_{\mathrm{L}}(t)}{\mathrm{d}t} &= V_{\mathrm{CC}} - V_{\mathrm{E}}(t) - V_{\mathrm{CE}}(t) - R' I_{\mathrm{L}}(t),
\end{aligned} \tag{2.33}$$

and $y_1(t) = V_{\mathrm{E-data}}(t)$, $x_1(t) = V_{\mathrm{E}}(t)$.

First we examine the dependence of the cost function on the resistance R' to see if the regularization associated with stabilizing motion on the synchronization manifold is present experimentally. We coupled our experimental Colpitts oscillator to another "model" circuit, and we evaluated the cost function $C(R', u)$ as we varied the resistance R' across the inductor for four values of the coupling $u(t) = u$ in Eq. (2.33), where $u \geq 0$ is constant. We see in Fig. 2.10 the same ragged set of local minima observed in our numerical simulation of the model oscillator. As u is increased, these local minima are smoothed out, allowing for the accurate estimation of R' whose value in the data-generating circuit was 35 Ω.

We estimated the unobserved state variable $V_{\mathrm{CE}}(t)$ and $I_{\mathrm{L}}(t)$ as well as the circuit parameters, when the emitter to base voltage $V_{\mathrm{E-data}}(t) = y_1(t)$ was presented to a model of the circuit. We minimized

$$\sum_{n=1}^{1{,}000} [(V_{\mathrm{E-data}}(t_n) - V_{\mathrm{E}}(t_n))^2 + u(n)^2], \tag{2.34}$$

subject to the Colpitts model equations of motion, Eq. (2.33), where $t_n = n\Delta t$ and $\Delta t = 10\,\mu$s to estimate the parameters in the circuit and the values of $\{V_{\mathrm{E}}(n), V_{\mathrm{CE}}(n), I_{\mathrm{L}}(n)\}$ during the observation window $0 \leq t \leq T = 10$ ms.

We also measured the parameters in the model using standard methods (Quinn et al. 2009). In Table 2.1 we show these parameters along with the estimated values when we assumed the Ebers–Moll (Ebers and Moll 1954) model for the bipolar transistor in the circuit. Except for the parameter β_{F} the estimates and the measured values are accurately in agreement with each other. In the estimation of the parameters shown in Table 2.2 we used the Gummel–Poon (Gummel and Poon 1970) model instead, and now we see that all the parameters are much more accurately estimated.

What should we conclude from these results of data assimilation? Well, both the Ebers–Moll and Gummel–Poon models estimate most of the parameters in the Colpitts circuit quite well. The latter repairs the one flaw in the Ebers–Moll representation of the bipolar transistor, the value of β_{F}, yet that model yields very good predictions. With the set of experiments we have performed, we cannot tell from the analysis of the time series alone which model to select. The Gummel–Poon model does give a slightly smaller numerical value of the cost function (Quinn et al. 2009). However, if we want to distinguish these models, we should design another set of experiments which probe the nonlinear device in other ways. One idea could be to allow the driving voltage to be time dependent in a rich manner. We will see in other examples, especially with neuron data and models, that the design of the experiment may be a critical factor in distinguishing models. Prediction will continue to be the key metric.

State Variable Estimation and Prediction

The regularized numerical optimization method yields estimates of the unobserved state variables $V_{\mathrm{CE}}(t_n)$ and $I_{\mathrm{L}}(t_n)$ in addition to estimates of the circuit parameters.

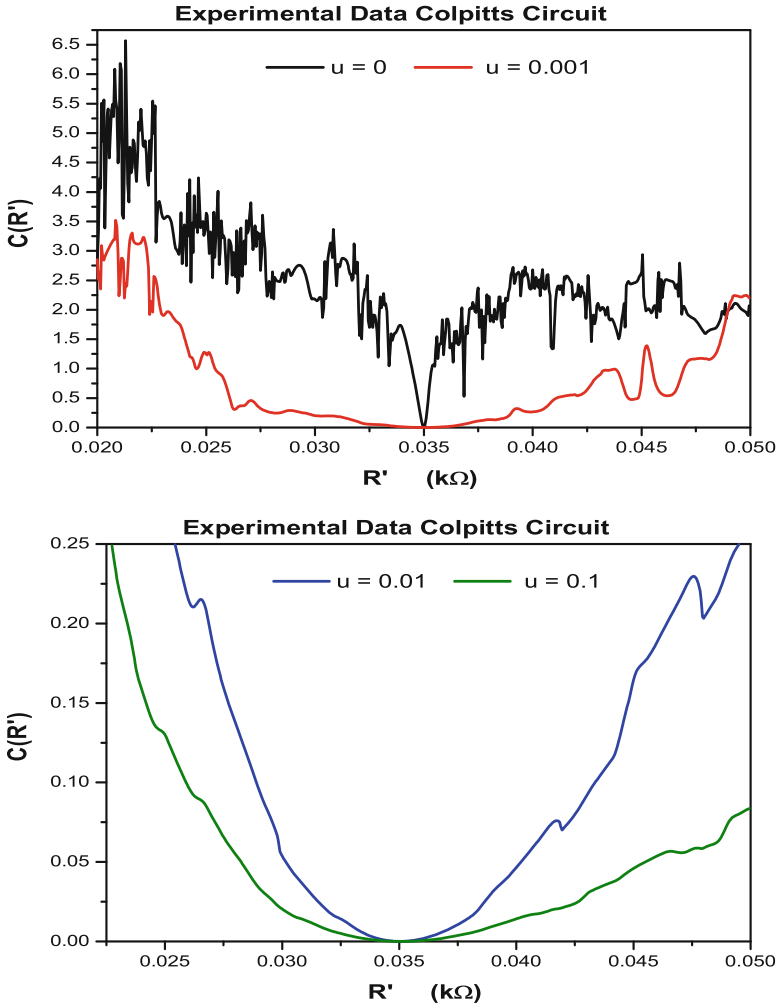

Fig. 2.10 Colpitts oscillator circuit cost function $C(R')$, Eq. (2.32), as a function of the coupling strength in the dynamical equations for the circuit Eq. (2.33). *Upper Panel* The many local minima for $u = 0$ and $u = 0.001$ arise because the voltage in the data $y_1(t_n)$ is not synchronized with the voltage $x_1(t_n)$ in the model and each is chaotic. This lack of synchronization is characterized by the CLEs: one is positive here. *Lower Panel* As we increase u to $u = 0.01$ and $u = 0.1$ the signals synchronize as the largest CLE on the synchronization manifold is moved to negative values. For the larger values of u, $C(R')$ is now smooth and the search for its minimum becomes easy. $C(R')$ decreases as $\frac{1}{u^2}$ for large u, and this effect is seen in the changed vertical scale between the upper and lower panels in the figure

In Fig. 2.11 we display the three state variables in the observation window $[0, T = 10\,ms]$. The observed values of all three state variables are displayed in red, while the values estimated using the regularized cost function are shown in black. Only $V_E(t)$ was presented to the circuit model.

Table 2.1 Estimated parameters and measured values

Name	Variational estimation	Measured	Units
C_2	7.02	7.23	μF
L	12.2	11.74	mH
R	40.0	39.3	ohms
V_0	0.661	0.63	V
V_T	25.0	27	mV
β_F	72.0	180	1

The Ebers–Moll (Ebers and Moll 1954) model of a bipolar transistor was used for the model in the minimization of the regularized cost function $C(R')$. All parameters except β_F are estimated with high accuracy

Table 2.2 Parameters estimated with the Gummel–Poon model (Gummel and Poon 1970) which includes the emitter resistance R_E along with the measured values of the parameters

Name	Variational estimation	Measured	Units
C_2	7.08	7.23	μF
L	12.00	11.74	mH
R	39.71	39.3	ohms
V_0	0.637	0.63	V
V_T	26.0	27	mV
β_F	179.0	180	1
R_E	0.23	Fixed value	ohms

Now all parameters are accurately estimated by our regularized procedure. R_E was fixed during the estimation

We may use the estimates for $\{V_E(T), V_{CE}(T), I_L(T)\}$ at the end of the observation window and the estimated circuit fixed parameters to predict the behavior of the Colpitts oscillator for $t > T$ utilizing the model differential equations, setting $u(t) = 0$. The result of this calculation is displayed in Fig. 2.12 with known values of the state variables in red and predicted values in black. Because the oscillations of the Colpitts circuit are chaotic, we expect that errors in the estimated state variables $\{V_E(T), V_{CE}(T), I_L(T)\}$ will grow in a manner associated with the positive Lyapunov exponent of the dynamics. We see that this limits the prediction horizon to about 8 ms here. This is consistent with the estimation of the positive Lyapunov exponent for the Colpitts circuit.

We set $u(t) = 0$ in making predictions with the model circuit equations as $u(t)$ is not a physical quantity but has been introduced to regularize the numerical optimization procedure. When the estimation is completed, we expect that $u(t)$ has been driven to zero by the optimization procedure and that the estimates required for prediction are independent of $u(t)$. This would tell us that the model we have used as a nonlinear filter to pass information from the observations of $V_E(t)$ to the estimates of the parameters and the unobserved state variables $\{V_{CE}(t), I_L(t)\}$ is consistent with the data. The validity of the model is examined in its ability to predict beyond the observation window.

In the estimation of states and parameters for the Colpitts model we have the values of $u(t)$ to determine whether this expectation is realized. In Fig. 2.13 we display the values of $u(t)$ in the observation window $[0, T]$, and we see the values

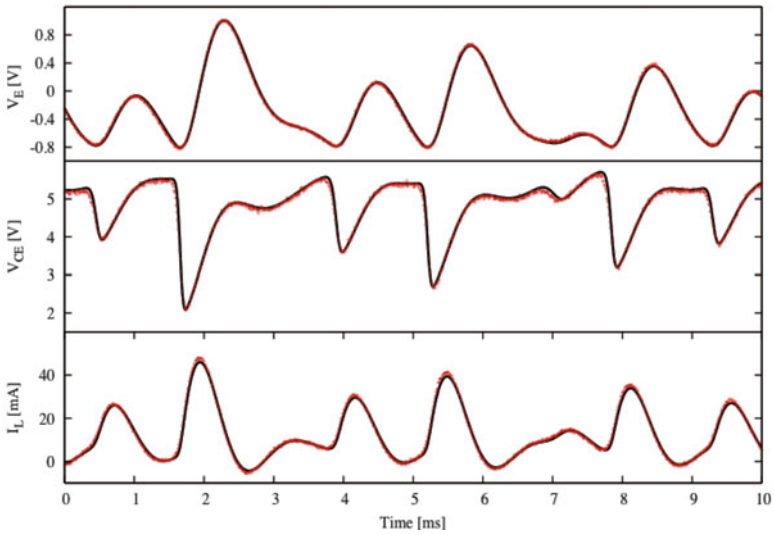

Fig. 2.11 The three Colpitts circuit dynamical variables in the observation window $0 \leq t \leq T = 10$ ms. In the *Top Panel* we see the observed $V_{\mathrm{E}}(t)$ in *red* and the estimated values of $V_{\mathrm{E}}(t)$ in *black*. In the *Middle Panel* we display $V_{\mathrm{CE}}(t)$: the known value is in *red*, and using the regularized estimation method, we calculate the estimates of $V_{\mathrm{CE}}(t)$ shown in *black*. $V_{\mathrm{CE}}(t)$ is not presented to the model for the Colpitts oscillator. In the *Middle Panel* we display the same for $I_{\mathrm{L}}(t)$: the known value is in *red*, and using the regularized estimation method, we calculate the estimates of $I_{\mathrm{L}}(t)$ shown in *black*. $I_{\mathrm{L}}(t)$ is not presented to the model for the Colpitts oscillator. The role of $u(t)$ is to locally regularize the surface over which we minimize the cost function (2.34). At the end of the estimation procedure, we expect $u(t) \approx 0$

appear "small". We can also examine the smallness of the control or regulatory variable $u(t)$ in more detail by evaluating the following dimensionless ratio in

$$R(t) = \frac{F_E(t)^2}{F_E(t)^2 + [u(t)(V_{\mathrm{E-data}}(t) - V_{\mathrm{E}}(t))]^2}, \tag{2.35}$$

through the observation window. $F_E(t)$ is the portion of the vector field for the Colpitts oscillator in the equation for $\frac{\mathrm{d}V_{\mathrm{E}}(t)}{\mathrm{d}t}$, namely,

$$F_E(t) = I_{\mathrm{L}}(t) - \frac{V_{\mathrm{E}}(t) - V_{\mathrm{ee}}}{R_{\mathrm{ee}}} + I_{\mathrm{B}}(V_{\mathrm{E}}(t)). \tag{2.36}$$

This tells us the relative importance of the control term $u(t)(V_{\mathrm{E-data}}(t) - V_{\mathrm{E}}(t))$ compared to the dynamics element in the equations of motion. If $R(t)$ is near unity, the role of the regularization by $u(t)$ is unimportant. We display $R(t)$ in Fig. 2.14 and note that the minimum value of $R(t)$ over the observation window is about 0.999995 in this experiment. This demonstrates that the data and the model are consistent, and the role of the control $u(t)$ is not essential. Again, the validity of the model is tested in its ability to predict.

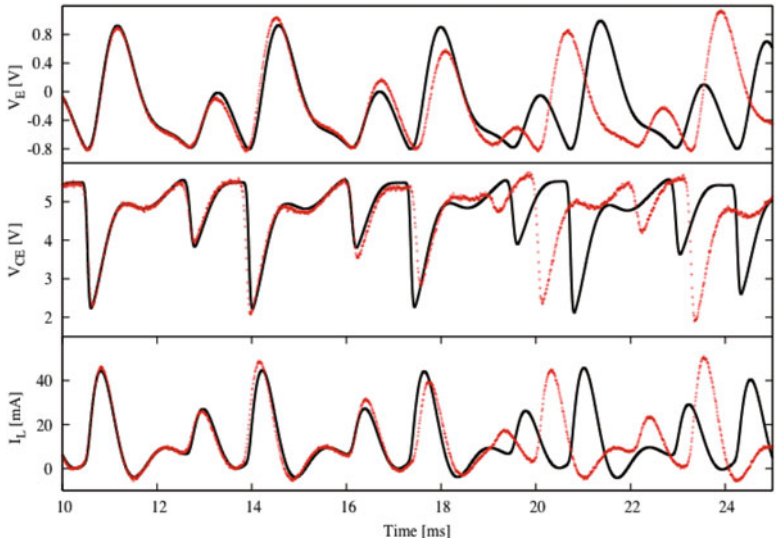

Fig. 2.12 Using the estimated values of the Colpitts state variables $\{V_{\mathrm{E}}(T), V_{\mathrm{CE}}(T), I_{\mathrm{L}}(T)\}$ at $T = 10\,\mathrm{ms}$, along with the estimated parameters of the experimental circuit, we forecast the behavior of the circuit into $t > T$. The measured values of each state variable is displayed in *red*, while the predicted values are in *black*. These deviate from one another as expected because of the chaotic behavior of the circuit and the model

This ends our introduction via an example chaotic circuit realized in a laboratory setting. The estimation procedure has been to minimize a regularized cost function subject to deterministic equations of motion for the nonlinear circuit. Developing the model using Kirchoff's rules and performing the estimation of the unknown parameters and unobserved state variables are independent elements of the data assimilation task. The methods show the consistency of the model with the data and then demonstrate the ability of the completed model to predict beyond the observation window.

In an experimental setting more complex than this one we expect higher noise levels in the observations and errors in the models. Further, we anticipate that there will be uncertainty in the values of the state of the model $\mathbf{x}(t_0 = 0)$ at the start of the measurements. The use of deterministic dynamical equation constraints will not be accurate generally, and we will not turn to the development of a data assimilation framework that can account for all of these features.

2.2.6 Numerical Optimization Methods

There is no shortage whatsoever of well-developed, even very well-tested, numerical methods for solving the variational problems associated with data assimilation.

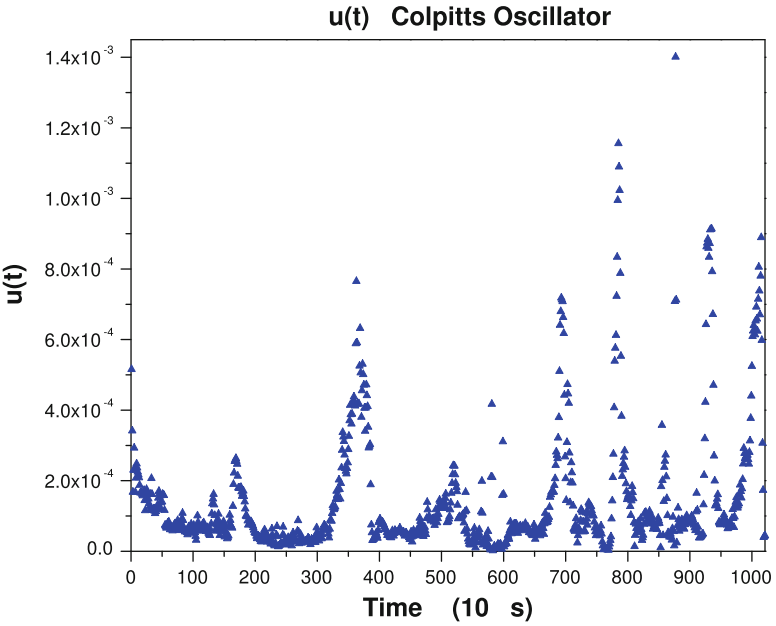

Fig. 2.13 The value of the control or regulation variable $u(t)$ throughout the observation window when the estimation procedure is completed. The dimensionless evaluation, $R(t)$, of the importance of $u(t)$ is in Fig. 2.14

Many methods use the explicit equations to find $\mathbf{x}(n > 0)$ from an initial condition $\mathbf{x}(0)$ leading to an action, often called an objective function or a cost function, which varies in a complex manner on the initial conditions $\mathbf{x}(0)$ and the parameters $\mathbf{p}$. In principle, one can use the same equations of motion and the estimated parameters $\mathbf{p}$ to integrate forward from t_0 at the beginning of the observation window, where we have an estimate of $\mathbf{x}(0)$ to the end of the observation window at $t_m = T$ and beyond, $t > T$. If the model is nonlinear and has chaotic orbits in the basin of attraction of $\mathbf{x}(0)$ for the estimated parameters $\mathbf{p}$, the sensitivity to $\mathbf{x}(0)$ and $\mathbf{p}$ is likely to produce quite inaccurate estimates for $\mathbf{x}(T)$ rendering prediction problematic. Indeed, estimates of $\mathbf{x}(t)$ within the observation window may suffer from this inaccuracy as well.

In the opening chapter of this book we discussed two kinds of data assimilation tasks. The first is comprised of "twin experiments" wherein one has generated data $\{\mathbf{y}(0), \mathbf{y}(1), \ldots, \mathbf{y}(m)\}$ from a known model and known class of measurement functions, and the purpose of exercising the data assimilation path integral, in variational approximation or otherwise, is to test one's numerical methods. This is always important and is certainly necessary when one has new methods that estimate only $\mathbf{x}(0)$ and $\mathbf{p}$ but might be failing. The failure is likely to come from requiring too large a time step $t_{n+1} - t_n$ or estimating states with uncertainties as to where

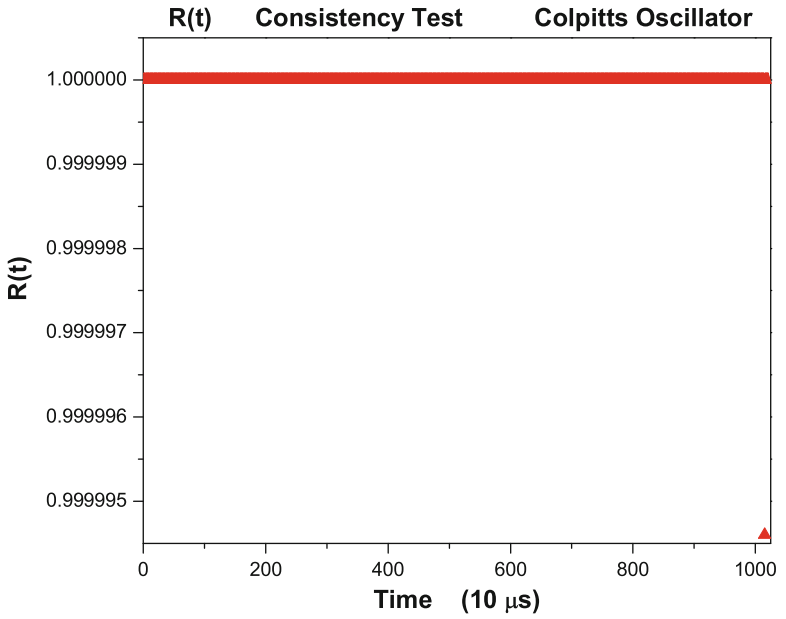

Fig. 2.14 $R(t)$, Eq. (2.35), over the observation window after the estimation procedure has completed. When $R(t)$ is near unity, the data and the model are consistent, and the role of $u(t)$ has been to regularize the search. We see here, in this twin experiment, that our model of the bipolar junction transistor is consistent with the observed experimental data

they begin (errors in $\mathbf{x}(0)$), and one can systematically correct the method for these flaws.

The other task, clearly the goal of all the methodology, is to proceed accurately when only a subset of state variables (L out of D) are known in the observation window, and estimating all of them accurately at $t_m = T$ is required. In this situation the metric of the quality of the dynamical model and its consistency with the data comes from prediction for $t > T$. If the full state $\mathbf{x}(T)$ is badly estimated, we may not know it from the data in the observation window as $D - L$ of the states are unobserved, and the L observed states may be well estimated. We will only recognize the diminished quality of the model as we predict beyond T.

This reasoning may not be important if the dynamical model has only regular behavior in state space, but nonlinear models may well have chaotic oscillations in regions of parameter and state space encountered in the search procedures. This cautions one to use the type of optimization approaches in SNOPT (Gill and Wright 1982; Gill et al. 2005) and IPOPT (Wächter and Biegler 2006).

The idea then is to use the measurements, their relation to the model state variables, and the dynamical model to establish an objective function. This is the objective function for use in the numerical packages. If the dynamics is to be used as equality constraints, namely no model errors, we seek to minimize the objective

function over the $(m+1)D$ values of the $\mathbf{x}(n)$ plus the NP fixed parameters $\mathbf{p}$, subject to the mD equality constraints

$$g_a(\mathbf{x}(n+1), \mathbf{x}(n), \mathbf{p}) = 0; \quad a = 1, 2, \ldots, D; \quad n = 0, 1, \ldots, m-1. \tag{2.37}$$

This takes a problem with $D + NP$ quantities to determine and enlarges it into a problem with $(m+1)D + NP$ quantities to vary. The space in which the numerical optimization operates can be and usually is much larger than the $D + NP$-dimensional subspace. As explained in Gill et al. (2005) in the larger space the searches involve sparse matrices allowing for speed and accuracy. From the point of view of applying the outcome of the variational data assimilation to the understanding and prediction of the properties of the model, indeed, the testing of the model itself for consistency with the data, the apparent additional work pays off. This is called the "direct method."

In our formulations of this strong variational problem for use with SNOPT or IPOPT, we use (Abarbanel et al. 2008) a Simpson rule integration method along with a Hermite interpolation method (Strang 1986) for implementing the equality constraints. The Simpson rule requires a midpoint evaluation, and the interpolation to the midpoint uses the Hermite method.

2.3 A Hodgkin–Huxley Neuron Model

2.3.1 Biophysics of the Hodgkin–Huxley Model

Networks of neurons exhibit rich dynamical behaviors, including rhythmic bursting and patterned sequence generation (Stein et al. 1997; Laurent et al. 2001; Johnston and Wu 1995; Koch 1999). These dynamics derive from the intrinsic properties of individual neurons and from the connections among them within the network. The membrane voltage of a neuron depends on currents from a diverse collection of ion channels, many of which have nonlinear voltage-dependent dynamics (Johnston and Wu 1995). General forms for the dynamics of many of the major families of ion channels have been characterized (ModelDB 2012; Graham 2002), but the kinetic parameters vary according to the neuron where the ion channels are located.

Neurons, from the perspective of this discussion, are nonlinear oscillators with competing feedback mechanisms associated with the flow of ions into and from a cell's interior across channels through the cell membrane. Such a neuron model was described by Hodgkin and Huxley (Johnston and Wu 1995) some decades ago and comprises a competition between inflow of Na^+ which has a higher concentration outside than inside the cell and outflow of K^+ which has a higher concentration inside the cell. Unstimulated, a neuron maintains a resting potential relative to its exterior of about -65 mV representing a balance between diffusion of differing concentrations of ions and electrostatic forces on the charged ions. When the neuron

is forced by a positive current, its voltage rises as Na^+ ions enter the cell through a channel whose voltage-dependent permeability to Na^+ increases. As the potential passes a threshold the neuron produces a spike in voltage, called an action potential, which is tempered and turned off as the voltage-dependent permeability to K^+ ions leaving the cell rises. Action potentials travel down axons from neuron to neuron, one neuron's output turning into another neuron's input, providing the intercellular communication as the basis of network operation.

We look at a neuron with Na^+ and K^+ channels as well as a "leak" channel that is an adjustment factor representing leakage of charge through the cell membrane. The equations of motion comprise conservation of charge in the form

$$C\frac{\mathrm{d}V(t)}{\mathrm{d}t} = \text{intrinsic, voltage-dependent ion currents} + \text{external stimulation currents.} \tag{2.38}$$

The intrinsic currents have the standard Hodgkin–Huxley form

$$g\,m(t)^{n_1}h(t)^{n_2}(E_{\text{rev}} - V(t)), \tag{2.39}$$

where n_1, n_2 are integers, E_{rev} is the Nernst potential (Johnston and Wu 1995) coming from the balance of diffusion and electrostatic forcing, $m(t)$ and $h(t)$ satisfy linear kinetic equations with voltage-dependent coefficients, and g is the maximal conductance of the ion channel. The dimensionless voltage-dependent gating variables $\{m(t), h(t)\}$ lie between 0 and 1, representing the percentage of ion channels of a particular type open or closed.

The neuron dynamics is determined with this current conservation equation along with kinetic equations for each of the $a_i(t) = \{m(t), h(t)\}$ associated with each channel:

$$\frac{\mathrm{d}a_i(t)}{\mathrm{d}t} = \alpha_{a_i}(V(t))(1 - a_i(t)) - \beta_{a_i}(V(t))a_i(t). \tag{2.40}$$

The original HH model has four degrees of freedom or four state variables: voltage and three kinetic coefficients governing the permeability of a Na^+ channel and a K^+ channel.

Stimulating currents of different signs can reveal much of a neuron's response properties, but due to the large variety of channels expressed in most neurons, determining the contribution of specific channel types usually requires intracellular or extracellular pharmacological manipulation to block all but the channel of direct interest. This allows the determination of the parameters in the ion current Eq. (2.39).

The efficacy and specificity of the pharmacological agents is often a concern, and data typically have to be pooled from many neurons, possibly obscuring differences among individual cells. Furthermore, such pharmacological manipulations require numerous experimental steps and lengthy recording durations, which often has

the practical implication of limiting analysis to exceptional recordings in which the neuron is fortuitously maintained in a healthy condition for an extended period of time. Further, there is no certainty that the cell membrane properties are not affected by the pharmacological agents employed. Based on these results, biophysical Hodgkin–Huxley models of the neurons can be constructed, with the parameters fixed by experimental results or chosen based on simulation and iterative search (ModelDB 2012).

Our approach to the question of estimating the parameters and states of an HH neuron works directly with the measured $V(t)$ and the selected stimulating current. Pharmacology is bypassed.

We consider the most common case, where the only measurement is the membrane voltage when a selected stimulus current is applied to the cell. Measuring a time series for $a_i(t)$ directly is not presently possible. A stimulus current $I_{\text{app}}(t)$ is applied through the same electrode recording the membrane voltage to probe the responses of voltage-gated channels. A data set usually consists of the applied current and the observed voltage over some observation window $t = [t_0, t_m = T]$ with measurements made at times $t_n = \{t_0, t_1, \ldots, t_m\}$ within that window.

Working solely with observations of the membrane voltage and a known stimulus or driving function of the neuron, the goal is to estimate all of the fixed parameters of the model, including maximal conductance, intracellular ion concentrations, and channel-gating kinetics as well as the unobserved gating dynamics $a_i(t)$ throughout the observation window. Knowledge of the quantities $V(T)$ and all of the $a_i(T)$ along with estimates of all fixed parameters in the full model allow us to use the model to predict the response of the neuron for $t > T$ given $I_{\text{app}}(t > T)$. If these predictions are accurate for a broad range of biologically plausible stimuli, then the estimates of the model parameters provide a parsimonious, biophysically interpretable description of the neuron's behavior. Furthermore, such results motivate the hypothesis that the model neurons will respond accurately to a diverse set of stimuli when used in models of interesting functional networks. This suggests that networks of such neurons could be constructed in a two-stage process: (1) using stimulus, response data to analyze the classes of individual neurons in the network using the dynamical methods outlined here, and (2) determine the connections among such neurons recognizing the connection strengths as parameters in a dynamical system using data on the network as a whole.

These requirements raise a number of practical and theoretical questions:

1. How many measurements $L \leq D$ are required? Typically for single neurons $L = 1$ (voltage), and given that most of the states of the system will remain unmeasurable, $L << D$. Equivalently, given that observations are generally limited to $V(t)$, is there an upper limit to the complexity of the models that can be completed from the data?
2. How often is it necessary to make measurements? Namely, given a fixed observation window $[t_0, \ldots, t_m = T]$, how frequently should the measurable variables be sampled? How large can one choose the intervals between observations, $t_{n+1} - t_n$?

3. Given that measurements are noisy and models are always wrong in some respect, how robust are the estimates of parameters and unobserved state variables to these errors?
4. What kinds of stimuli lead to adequate exploration of the state space of the model so that all the parameters and unobserved state variables can be estimated with a similar degree of confidence?
5. What metrics are most appropriate for testing the validity of a neuronal model, especially in light of the intrinsic variability of real neurons?

The model we address (Abarbanel et al. 2011; Toth et al. 2011) consists of the dynamics of membrane voltage $V(t)$ driven by two voltage-gated ion channels, Na and K, and a "leak" current and an external, applied current we call $I_{\text{app}}(t)$. The dynamics of the model comprise the equations for voltage

$$\begin{aligned}\frac{\mathrm{d}V(t)}{\mathrm{d}t} &= \frac{1}{C}\left\{g_{\text{Na}}m(t)^3h(t)(E_{\text{Na}}-V(t)) + g_{\text{K}}n(t)^4(E_{\text{K}}-V(t))\right.\\ &\quad \left. + g_{\text{L}}(E_{\text{L}}-V(t)) + I_{\text{DC}} + I_{\text{app}}(t)\right\},\\ &= F_V(V(t),m(t),h(t),n(t)) \end{aligned} \tag{2.41}$$

where the g_X's indicate maximal conductances and the E_X's reversal potentials, for each of the Na, K, and leak channels. I_{DC} is a DC current, and $I_{\text{app}}(t)$ is an applied time-dependent external current selected in an experiment. We refer to this as the `NaKL` HH model.

The gating variables $a_i(t) = \{m(t), h(t), n(t)\}$ are discussed in many textbooks and reviews (Johnston and Wu 1995; ModelDB 2012; Graham 2002), and each satisfies a first-order kinetic equation of the form

$$\frac{\mathrm{d}a_i(t)}{\mathrm{d}t} = \frac{a_{i0}(V(t)) - a_i(t)}{\tau_i(V(t))}. \tag{2.42}$$

The kinetic terms $a_{i0}(V)$ and $\tau_i(V)$ are taken here in the form

$$\begin{aligned} a_{i0}(V) &= \frac{1}{2}\left(1+\tanh\left[\frac{(V-\text{va})}{\text{dva}}\right]\right)\\ \tau_i(V) &= t_{a0} + t_{a1}\left(1-\tanh^2\left[\frac{(V-\text{vat})}{\text{dvat}}\right]\right) \text{ or}\\ \tau_i(V) &= t_{a0} + t_{a1}\tanh\left[\frac{(V-\text{vat})}{\text{dvat}}\right]. \end{aligned} \tag{2.43}$$

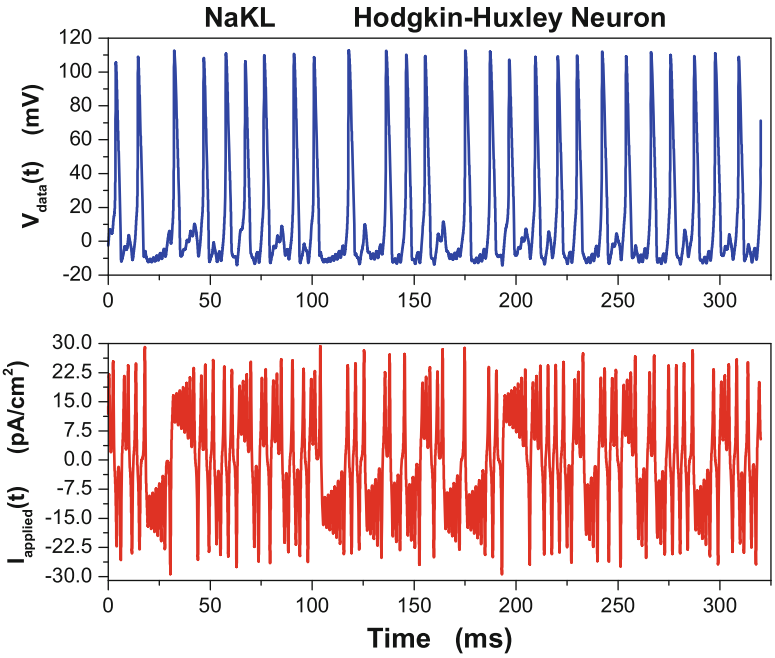

Fig. 2.15 NaKL Hodgkin–Huxley neuron model. *Top Panel* The "data" in this twin experiment. The response membrane voltage. *Bottom Panel* The stimulating current. This is an **observed** state variable

In the language used in our general formulation, the model state variables are $\mathbf{x}(t) = \{V(t), m(t), h(t), n(t)\}$, and the parameters are $\mathbf{p} = \{C, g_{\mathrm{Na}}, E_{\mathrm{Na}}, g_{\mathrm{K}}, E_{\mathrm{K}}, \ldots, \mathrm{dvat}\}$. In a twin experiment, the data $\{V_{\mathrm{data}}(t), m_{\mathrm{data}}(t), h_{\mathrm{data}}(t), n_{\mathrm{data}}(t)\}$ are generated by solving these HH equations for some initial condition $\mathbf{x}(0) = \{V(0), m(0), h(0), n(0)\}$ and some choice for the parameters, the DC current and $I_{\mathrm{app}}(t)$, and the stimulating current. The data $y(t)$ presented to the model consist of only the voltage $V_{\mathrm{data}}(t)$ along with additive noise: $y(t) = V_{\mathrm{data}}(t) + \mathrm{noise}(t)$.

2.3.2 *Estimating Parameters and Unobserved States of the HH Model*

We solved the HH equations (2.41) and (2.42) using the stimulating current shown in the Bottom Panel of Fig. 2.15. The Top Panel of this figure displays the time series of $V(t)$ resulting from the calculation. A fourth-order Runge Kutta method was employed using a time step of 0.01 ms. This current was selected to probe the dynamical range of the currents in the NaKL model including the spiking region for high voltage and the subthreshold region. The current wave form was taken from

Table 2.3 Parameters used in the NaKL HH model to generate data for a twin experiment

Name	Value	Name	Value
C	$1.0\,\mu$F/cm^2	vh	-60.0 mV
g_{Na}	120.0 mS/cm^2	dvh	-15.0 mV
E_{Na}	50.0 mV	t_{h0}	1.0 ms
g_{K}	20.0 mS/cm^2	t_{h1}	7.0 ms
E_{K}	-77.0 mV	vht	-60.0 mV
g_{L}	0.3 mS/cm^2	dvht	-15.0 mV
E_{L}	-54.4 mV	vn	-55.0 mV
vm	-40.0 mV	dvn	30.0 mV
dvm	15.0 mV	t_{n0}	1.0 ms
t_{m0}	0.1 ms	t_{n1}	5.0 ms
t_{m1}	0.4 ms	$v$nt	-55.0 mV
vmt	-40.0 mV	dvnt	30.0 mV
dvmt	15.0 mV	I_{DC}	7.3 pA/cm^2

The model includes Na, K, and leak currents

the output of the Malkus waterwheel, but its essential ingredients are that it changes slowly compared to the response of the neuron and causes the neuron response to visit significant parts of the $\{V(t), m(t), h(t), n(t)\}$ state space. To the equations of motion for $V(t)$, we added a regulating term $u(t)(y(t) - V(t))$.

When the cost function

$$\frac{1}{N}\sum_{n=0}^{N-1}\Big[(y(t_n) - V(t_n))^2 + u(t_n)^2\Big] \tag{2.44}$$

was minimized subject to Eqs. (2.41) and (2.42), we determined the 18 fixed parameters in the model along with the values of $\{V(t), m(t), h(t), n(t)\}$ at each point in the observation window $\{t_0 = 0, \ldots, t_n = n\Delta t, \ldots, t_m = T = 90\,\mathrm{ms}\}$. This involves a numerical minimization in 36,018 dimensional space. This was accomplished in order of ten minutes on a commodity PC.

The parameters selected and the outcome of the numerical optimization are shown in Tables 2.3 and 2.4. In Fig. 2.16, we show in the **Left Panel** the known, "data", membrane voltage along with the estimated $V(t)$ resulting from the optimization procedure. In the **Right Panel** we display the known Na^+ activation gating variable $m(t)$ along with the estimation of this resulting from the numerical optimization. $m(t)$ is an **unobserved** state variable. Its estimation is achieved by the nonlinear filtering performed by the dynamical equations of motion used as constraints on the cost function as it is minimized. In treating experimental data, $m(t)$ and any other gating variables are not observed, and we have no way to examine the precision with which they are estimated. One of the important uses of twin experiments is that it permits us to assess how well unobserved quantities are being estimated and, in that manner, build confidence in our overall state and parameter estimation effort.

Table 2.4 Parameters in generating the "Data" and the estimated parameters in NaKL HH model; I_{DC} was fixed in all calculations

Name	Value in "data"	Estimated value
g_{Na} (mS/cm^2)	120	118.79
E_{Na} (mV)	50.0	49.99
g_K (mS/cm^2)	20	20.35
E_K (mV)	−77.0	−76.97
g_L (mS/cm^2)	0.3	0.2955
E_L (mV)	−54.4	−54.11
vm = vmt (mV)	−40.0	−40.08
dvm = dvmt (mV)	15.0	14.90
t_{m0} (ms)	0.1	0.1009
t_{m1} (ms)	0.4	0.3982
vh = vht (mV)	−60.0	−59.91
dvh = dvht (mV)	−15.0	−14.88
t_{h0} (ms)	1.0	1.004
t_{h1} (ms)	7.0	7.047
vn = vnt (mV)	−55.0	−54.91
dvn = dvnt (mV)	30.0	29.28
t_{n0} (ms)	1.0	1.012
t_{n1} (ms)	5.0	4.99

As the term $u(t)(y(t) - V(t))$ introduced the nonphysical terms $u(t_n)$ into the cost function, we would like to see that at the end of the numerical procedures it is absent. To assess this we evaluated the dimensionless quantity

$$R(t) = \frac{F_V(V(t), m(t), h(t), n(t))^2}{F_V(V(t), m(t), h(t), n(t))^2 + (u(t)(y(t) - V(t)))^2} \tag{2.45}$$

throughout the observation window using the final values for all quantities at the termination of the estimation procedure. This is shown in Fig. 2.17. We see that $R(t)$ is effectively united for all times in the observation window, showing the unimportance of the unphysical control term at the final stage of the numerical process.

2.3.3 Predicting the Response of the HH Model

A numerical optimization method solving the constrained optimization problem "minimize this cost function subject to these equations as constraints" will always produce numbers. In the case of twin experiments we can test the accuracy of those numbers as we know them a priori. In the more interesting case of experiments where only the stimulating currents and the response voltages are known to us, we need to ask more of the numerical output.

We take as the primary goal of the data assimilation effort we describe through this monograph the prediction of new response voltages when the model neuron

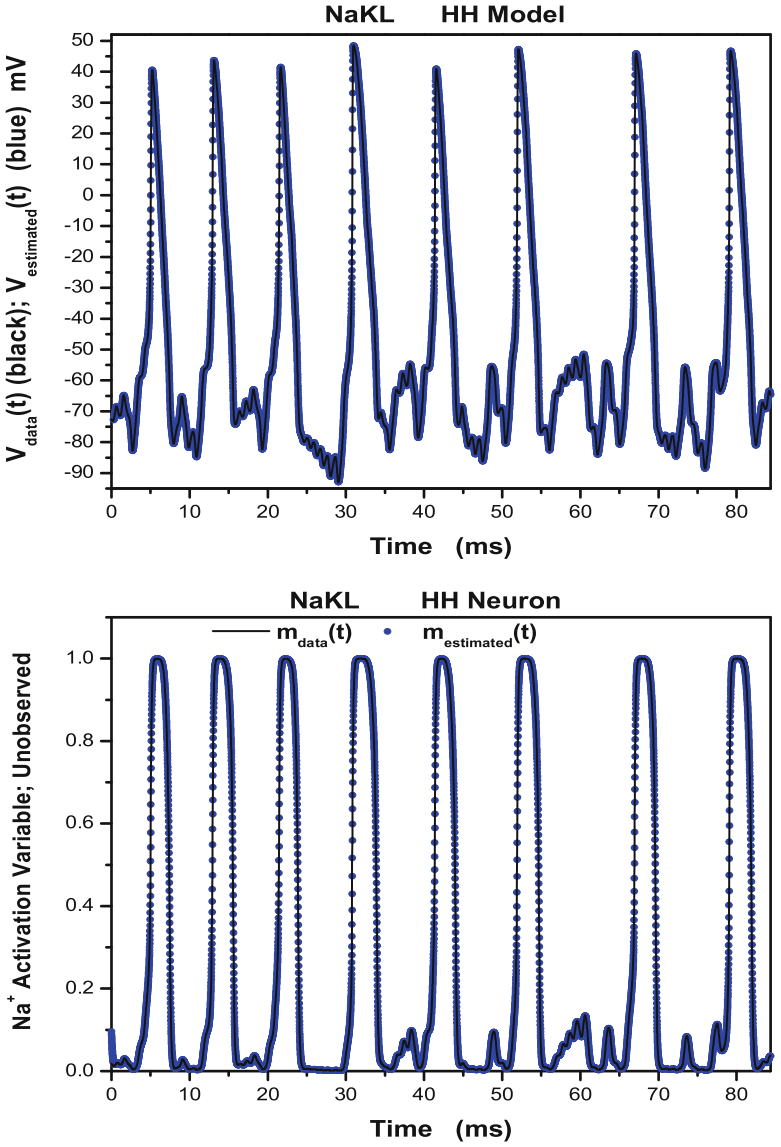

Fig. 2.16 NaKL model. *Top* We display the known and estimated membrane voltage $V_{\text{data}}(t), V_{\text{estimated}}(t)$ from "data" generated in a twin experiment using the NaKL model as a data source. *Bottom* We display the known and estimated Na^+ activation gating variable $m_{\text{data}}(t), m_{\text{estimated}}(t)$ from "data" generated in a twin experiment using the NaKL model as a data source. In an experiment $m(t)$ would be an `unobserved` state variable. Here the unobserved variables are known to us and comparison is possible

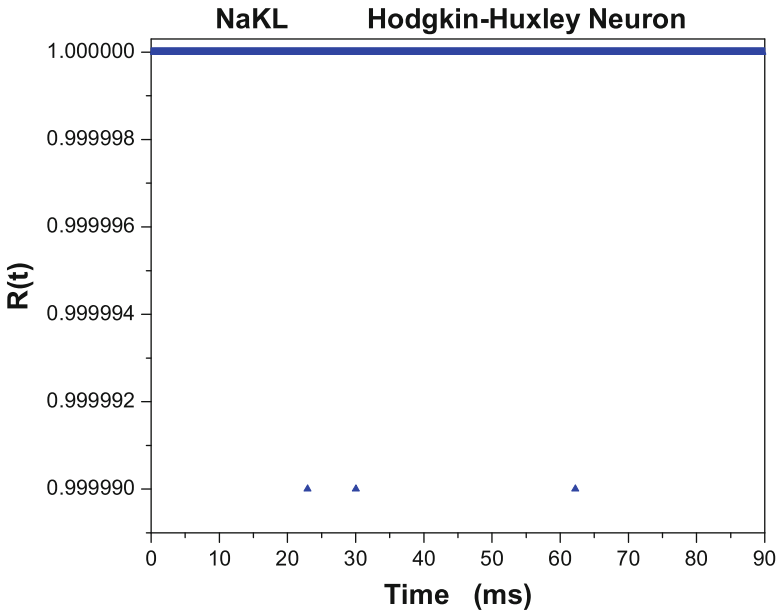

Fig. 2.17 $R(t)$, Eq. (2.45), over the observation window after the estimation procedure has completed. When $R(t)$ is near unity, the data and the model are consistent, and the role of $u(t)$ has been to regularize the search

encounters a different forcing function $I_{\text{app}}(t)$. This is not universally agreed on Wunsch and Heimbach (2007) using the methods described in this book or other instantiations of the ideas here. It is argued that the goal is to establish that the state of the system is well approximated by the data assimilation procedure. However, as we will see as we proceed, that can be quite misleading.

Our metric for success in the data assimilation will be the ability of the completed model to predict future response to different forcing. In order to accomplish this we require all the parameters $\mathbf{p}$ in the model, thus completing it, and all of the state variables at the end of the observation window so the dynamical equations, with $u(t) = 0$, may be used to predict forward in time. This gives a rationale to the minimization algorithms we use, and which we will discuss in another section of the book. These estimate the state at each point in the observation window, including the endpoint $t = T$.

We have taken from our numerical optimization the estimation of the four quantities $\{V(T), m(T), h(T), n(T)\}$ and the many fixed parameters and used the dynamical equations to predict forward for $t > T = 90$ ms. In Fig. 2.18 we show in the Top Panel a repeat of the known voltage (black) and the estimated voltage (red) in the observation window, and in the Bottom Panel, we display the known voltage in black and the predicted voltage in blue. This is a twin experiment, $R(t) \approx 1$, and $V(t)$ is an observed variable appearing explicitly in the cost function, so we may not be surprised at this juncture at the accuracy of this result.

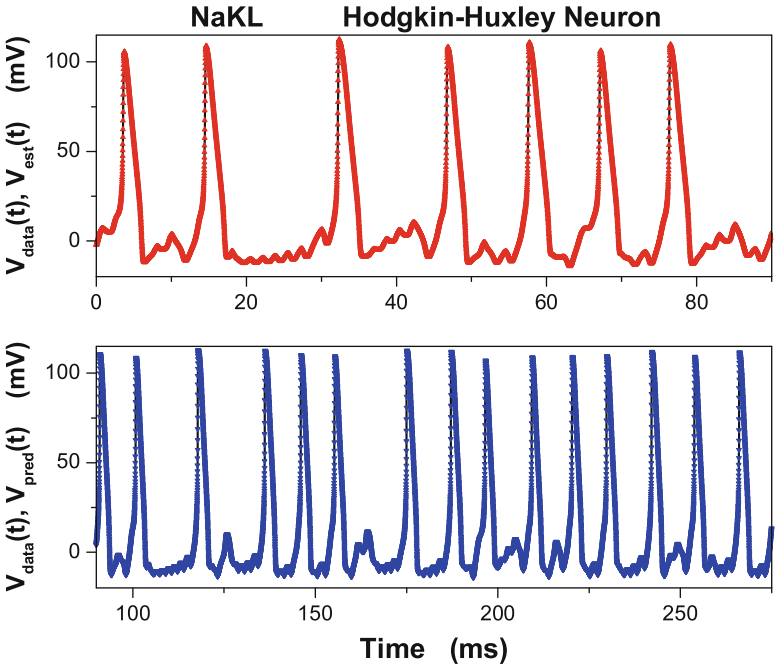

Fig. 2.18 NaKL model. *Top Panel* The "data" voltage was presented to the NaKL model. This shows those "data" and the estimated output voltage from the model. *Bottom Panel* This displays "data" voltage presented to the NaKL model and the predicted voltage after the observations are completed. This is an observed state variable

To further inquire into the method, we display in Fig. 2.19 the same type of result for an **unobserved** state variable. Here we show the K^+ gating variable $n(t)$. In black is the known data generated using the model equations with known parameters. Also displayed is the estimation (in red) of $n(t)$ in the observation window $[0, 90\,ms]$ and the prediction for $t > 90\,ms$ in blue. The accuracy of the results is consistent with the $R(t)$ test and speaks for itself.

2.3.4 Consequences of the Wrong Model

We address one last question of an introductory nature. If our procedure is so clever, can it give us signals of the model being `wrong`? To make this inquiry, we presented our NaKL model with membrane voltage "data" from a different neuron model which we call the NaKLh model. The biophysical difference was that the "wrong" data came from a neuron model with an additional voltage-gated current active primarily in the subthreshold voltage region. Using this wrong data, we carried out the same estimation procedure as above using 76.3 ms of "data".

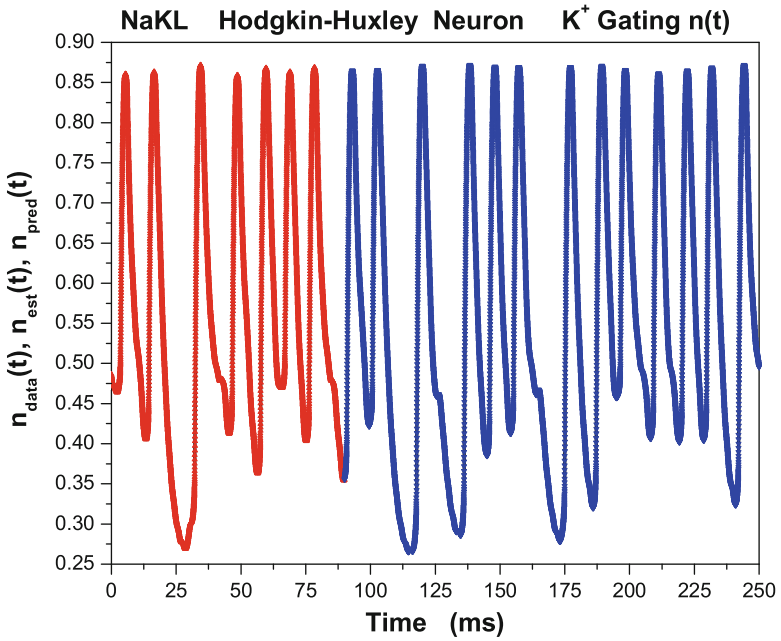

Fig. 2.19 NaKL model. Estimating and predicting the K^+ gating variable $n(t)$. The "data" $n_{\text{data}}(t)$ are in *black*; the estimated values $n_{\text{est}}(t)$ are in *red*; the predicted values $n_{\text{pred}}(t)$ using the 18 estimated parameters and the values of $\{V(T), m(T), h(T), n(T)\}$ at $T = 90$ ms and the dynamical equations with $u(t) = 0$. This is an unobserved state variable

In Fig. 2.20 we see the estimated voltage (blue) as well as the known voltage (black) presented to the model. The current stimulating the neuron model is essentially the same as before. While this estimation appears rather accurate, we take away a quite different conclusion when we examine the values of $R(t)$ in the observation window. These are displayed in Fig. 2.21 in which we observe significant deviations from $R(t) \approx 1$. In this twin experiment, further evidence that something is awry comes from examining the estimation of the Na^+ gating variable $m(t)$ in the observation window (Fig. 2.22). The significant disagreement between the known $m(t)$ and the estimated $m(t)$ is associated with the fact that the nonlinear filter of voltage information, namely, the model neuron, is incorrect. Of course, this indicator is not available for experimental data. Finally, though we do not display it, the predictions of the model are also incorrect in this situation.

2.4 Synopsis and Perspectives: "Slightly Complex" Examples

The examination of three low-dimensional example problems from fluid dynamics, electronic circuit,s and neurobiology has illustrated many of the issues addressed in this monograph. We have been presented in each case with one time series of

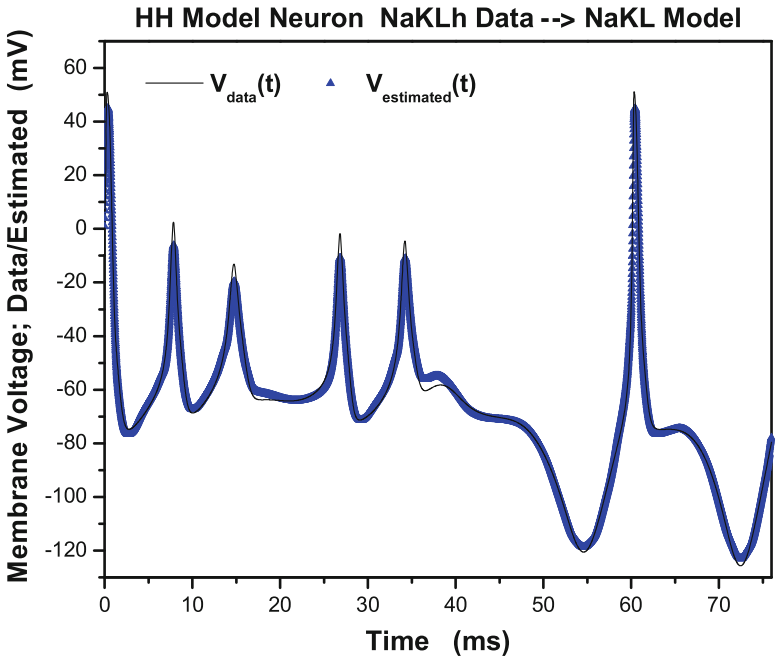

Fig. 2.20 The voltage data, *black line*, and the estimated voltage, *blue triangles*, when we present the "wrong" voltage data to an NaKL neuron model. There is an additional current in the data voltage and it is not represented in the simpler NaKL model. Nonetheless, the estimation procedure appears to produce a good representation of the presented data

measured data $y(t)$ from a physical system, and recognizing that the state of the system is represented by a model with a D-dimensional state variable $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_D(t)]$ $D \geq 1$, we sought to determine the fixed parameters $\mathbf{p}$ of the model as well as the unobserved states of the model at some time T when observations were completed. Using the information in the data we showed in each case that selecting a metric to compare the data $y(t)$ with a physically motivated function of the state $h(\mathbf{x}(t))$, we could try to minimize the distance associated with the metric, constrained by the dynamical equations of the model, and successfully determine $\mathbf{p}$ and the unobserved states. We also demonstrated that for $t > T$, we could predict the state of the model system.

When one has a measurement function $h_l(\mathbf{x})$ expressing the relation between the observed quantity and the model state, the cost function contains terms of the form $[y_l(t) - h_l(\mathbf{x}(t))]$, and in the use of variational principles as we used in this chapter, one requires the derivatives of the measurement function itself. This can be an issue when the measurement function has thresholds or very sharp variations. However, when we arrive at the discussion of the direct evaluation of the path integral representing the full answer to statistical data assimilation questions, these derivatives of measurement functions are avoided, but more about that as we proceed.

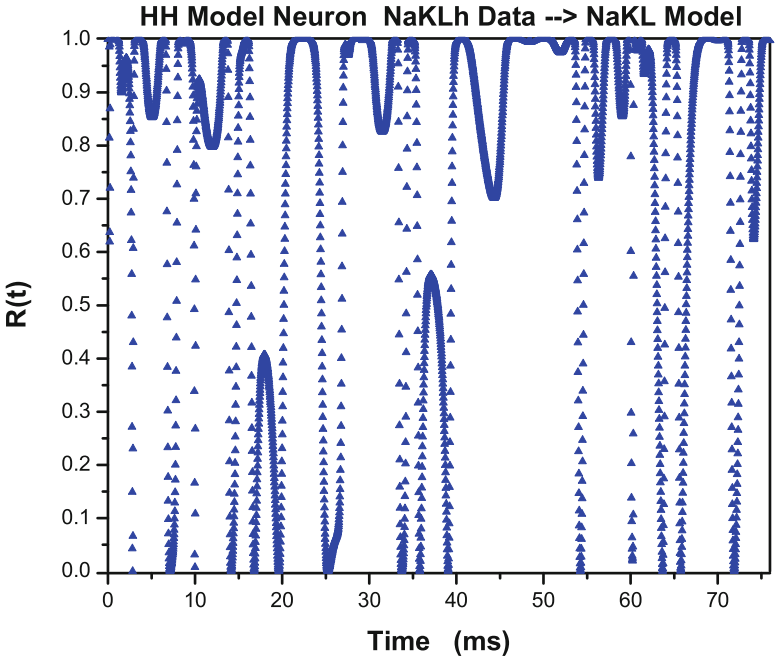

Fig. 2.21 The dimensionless ratio $R(t)$ resulting when an NaKL model is presented with data from a neuron model with an additional current, the NaKLh model. This shows that the NaKL model is inconsistent with the data

In our discussion in this chapter we made the simplifying assumption that $h(\mathbf{x}(t)) = x_1(t)$ and that the metric to use is a least-squares distance between the data and the corresponding model output

$$C(\mathbf{p}) = \frac{1}{T}\int_0^T \mathrm{d}t\,(y(t) - x_1(t))^2$$
$$= \frac{1}{m+1}\sum_{n=0}^{m}(y(t_n) - x_1(t_n))^2. \tag{2.46}$$

In the second form we recognized that measurements were made at the $m + 1$ discrete times $t_n = n\Delta t$; $T = m\Delta t$. In addition to this standard formulation of the problem of estimations of parameters and state variables, we encountered a difficulty when the data time series and the model output were chaotic, and a regularization procedure was suggested that smoothed out the irregular plethora of local minima in $C(\mathbf{p})$ during the estimation period and then disposed of the regulation when the estimation was completed.

A number of questions arise about the procedures we have employed in this introductory chapter:

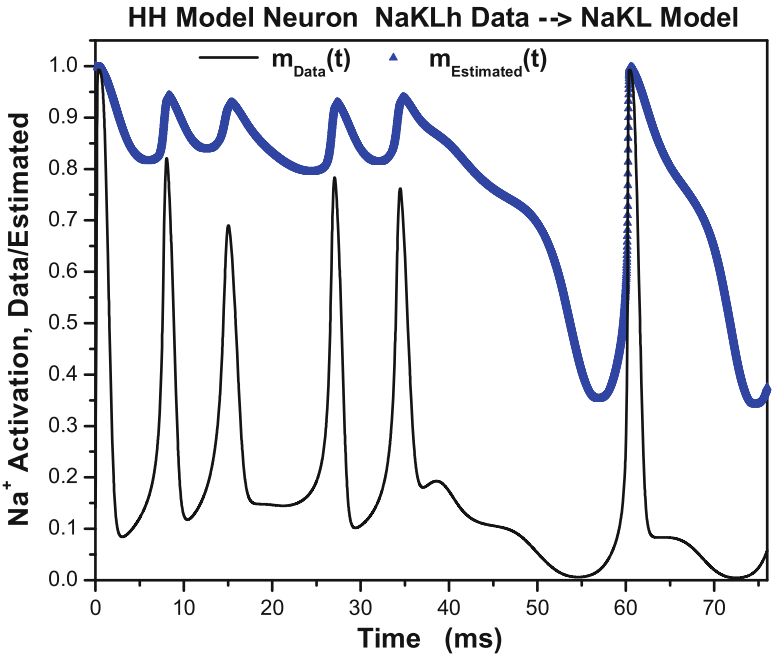

Fig. 2.22 The known Na^+ gating variable $m(t)$, *black line*, and the estimated $m(t)$, *blue triangles*, when data from the "wrong" model (NaKLh) is presented to an NaKL model. Since this is a twin experiment, we are able to compare the actual $m(t)$ in the data with the $m(t)$ estimated by the variational principle. This type of comparison is absent in a real experiment where only membrane voltage is observed

- We assumed that a least-squares metric to compare our data $y(t)$ to the model output $x_1(t)$ should transmit enough information to enable estimation of the parameters and unobserved states in the models. What formulation replaces that when the statistics of the data are not Gaussian?
- We more or less assumed that passing measurements from one of the system state variables to the model would be enough to enable accurate estimation of the model parameters and the model's unobserved states. When does that remain true for more complex systems? When are more observations at each measurement time required?
- We estimated unobserved model states and model parameters using an optimization principle with the model dynamical equations as equality constraints. What is the correct rule when the model, as well as the observations, has errors?
- We more or less assumed that when we estimated the state values, in particular the initial conditions $\mathbf{x}(t_0)$ when observations begin, those could be determined precisely. Is this always so? What role does uncertainty in the initial conditions play in the estimation task?

- We regularized the instabilities on the synchronization manifold with the addition of a control $u(t)$ which drives the model output to the data, then is itself driven to zero by the optimization procedure. This appears ad hoc. What other approaches can one use?
- We took **prediction** for times beyond the observation or assimilation window $t > T$ as the metric for a successful analysis of a model provided with data from observations. This goes beyond a "good fit" of model output to observed data as it requires the full state of the model to be accurately estimated at the completion of observations, and it uses the nonlinear model as a filter of information in the data to be utilized in determining the unobserved state variables. Are there better metrics for model completion and correctness? Can one develop criteria for testing the consistency of the model with the available data?

We do not have complete answers to these issues and many others that arise as we proceed. Nonetheless, we now turn to the general formulation of data assimilation questions and methods for answering them in a practical fashion.

http://www.springer.com/978-1-4614-7217-9

Predicting the Future
Completing Models of Observed Complex Systems
Abarbanel, H.
2013, XVI, 238 p. 97 illus., 91 illus. in color., Hardcover
ISBN: 978-1-4614-7217-9