SIAM/ASA J. Uncertainty Quantification
Vol. 7, No. 1, pp. 31–52

Modeling and Predicting Chaotic Circuit Data*

Wanting Xu†, Michael L. Stein†, and Ian Wisher‡

Abstract. Analyzing chaotic observations generated from some unknown nonlinear dynamics presents significant challenges for modeling the process and predicting future evolutions. We consider time series data measured from an electrical circuit that exhibits chaotic behavior. We compare and contrast the performance of Gaussian process (GP) and neural network (NN) models in short-term prediction and capturing the long-term dynamics. Major difficulties in modeling observations generated by such a physical process include distinguishing and characterizing the model and observational errors. We explore the effects of different types of model and observational errors on the likelihood function of the initial state using data generated under our fitted NN model. Our results show the distinctive effects of model and observational errors on inferring the initial state. In the absence of model error, we obtain exponentially growing information about the initial state for more observations even under temporally correlated observational errors. However, even with a tiny model error much smaller than what is measured physically and otherwise correct dynamics, the likelihood method no longer identifies the true initial state and hence cannot track the system for an indefinitely long time. Nonetheless, for this circuit system, our fitted predictors based on the GP and NN models appear to be promising in capturing the long-term dynamics in the presence of stochastic thermal noise.

Key words. chaotic time series, dynamical system, model error, neural network, Gaussian process

AMS subject classifications. 34C28, 37D45, 62M45, 60G15

DOI. 10.1137/17M1142387

1. Introduction.

The study of nonlinear dynamics and chaos has traditionally focused on developing and investigating mathematical invariants that describe and classify the asymptotic behavior of the iterates [26, 28]. This leads to important invariants such as the Lyapunov exponent and fractal dimension with which knowledge about the future evolution of the dynamical system can be learned, once its mathematical description is known. However, challenges still remain to analyze observations generated by a dynamical system whose mathematical formulation is at least partially unknown. These challenges include, but are not limited to, calculating geometric and dynamical invariants of an underlying strange attractor [13, 41, 50], modeling the deterministic portion of the dynamical evolution from the observations [17, 32], and constructing a predictive model directly from the observations [15, 22]. A comprehensive discussion of the problems arising from analyzing real observations from a chaotic system can be found in [1] and [2]. In this work, we focus on modeling the observed dynamical system and predicting its future evolution. The modeling and prediction task is made difficult in particular by the characterization of possible stochastic model error of the

*Received by the editors August 7, 2017; accepted for publication (in revised form) October 15, 2018; published electronically January 2, 2019.
http://www.siam.org/journals/juq/7-1/M114238.html

†Department of Statistics, University of Chicago, Chicago, IL 60637 (wxu@galton.uchicago.edu, stein@galton.uchicago.edu).
‡Department of Physics, University of Chicago, Chicago, IL 60637 (wisher@uchicago.edu).

underlying dynamics, observational error during measurement, and the separation of the two. The goal of this work is to provide some insights into modeling the dynamics and characterizing different types of errors for observations measured from some chaotic physical systems. We investigate these issues in the context of a chaotic circuit system and combine both real data analysis and simulation studies based on models fitted to the real data.

Actual observations from physical systems are typically measurements of a few observables, while the underlying state space of the system can be higher-dimensional. One common method used to reconstruct the state space from the observations s_t is time-delay embedding, which uses the time-lagged observation vectors $(s_{n\tau}, s_{(n+1)\tau}, \ldots, s_{(n+m-1)\tau})$ for some embedding dimension m, time lag τ, and $n \geq 1$ to form coordinates of the reconstructed state space [2, 15]. This method is, to some extent, supported by the Takens embedding theorem [48], which indicates that the information about the underlying state space is preserved by the time-lagged observation vectors for certain embedding dimensions. With the time-delay embedding and the reconstructed state space, the modeling and prediction of observed chaotic data have revolved around local techniques, global techniques, and the combination of the two [2]. A common local prediction technique discussed in various instances [15, 17, 22] is to construct the k-step ahead predictor at a time point in the form of a polynomial in m variables for some embedding dimension m. The parameters of the polynomial are fitted using the near neighbors of the m-dimensional time-lagged observation vector and their corresponding k-step ahead observations. For example, the simplest nonlinear method of local forecasting proposed by [36] is to find the nearest neighbor s_t to the current observation s_n, and use the value s_{t+1} as the prediction for s_{n+1}. This prediction-by-analogue method is essentially fitting a constant—a zeroth order polynomial. One improvement to this simple analogue method is to use more than one near neighbor and predict by taking the average [42] or distance-weighted average [47] of their outputs. Some studies [30] have also investigated selecting the hyperparameters of this method such as the number of near neighbors optimally. Further improvements can be made by fitting a first order polynomial using the near neighbors [15, 38], and [22] investigates higher order polynomials. The use of higher order polynomials increases the model complexity and is expected to produce better predictions. However, since the number of parameters increases exponentially in the order of the polynomial, more computational efforts are also encountered. In addition, the local models are discontinuous, which is undesirable if the goal is to obtain a description of the underlying continuous dynamics.

Global models, on the other hand, describe the whole set of observations by representing the model mapping as an expansion in some basis functions, e.g., as a polynomial or ratio of polynomials, and fitting the parameters using the entire data set [2, 9, 15]. The method is also subject to computational difficulties when the model is of high complexity. Modeling using radial basis functions [44] is an example of combining features from both the local and global techniques. The model is constructed and interpreted globally but also maintains good local properties through the locally centered radial basis functions [15]. In our work, we construct global models trained for one-step ahead prediction using Gaussian process (GP) and neural network (NN) models. GP provides a statistical basis for interpolation, model diagnostic, and uncertainty calibration, while NN has proven effective in modeling some chaotic time series data [6, 19, 23, 27, 51].

There has been a wide range of studies on using GP and NN to model and predict dynam-

ical systems arising in various areas [12, 14, 21, 24, 53]. Particularly for chaotic time series prediction, [6, 18, 43] have considered feed-forward NNs. Recurrent and nonlinear autoregressive with exogenous input (NARX) NNs have been utilized in [4, 5, 19, 37, 54]. However, the application areas of the aforementioned studies have mainly concentrated on classical chaotic systems such as the logistic map, the Lorenz system, and Mackey–Glass equations. Although studying the performance of modeling and prediction methods using the standard systems is insightful, we highlight the importance of investigating observations measured from real nonlinear and chaotic dynamical systems, which, in addition to the unknown dynamics, almost always also contain nonnegligible but unknown systematic and observational noise. The existence of these uncertainties associated with real systems complicates modeling and state estimation, as we will show later in this paper. Some works have been applied to benchmark nonlinear time series data sets, e.g., the sunspot series [4, 37]. For data such as the sunspot series, the focus is often on short-term prediction [4], perhaps because accurate long-term prediction is not feasible for such a complex system given the limited data. In contrast, this work explores real observations from a chaotic dynamical system whose state vector under a nominal model is fully and accurately observed at each time step, thus at least raising the possibility of accurate long-term prediction and identification of the initial state of the system from the observations. Specifically, our goal is to understand the performance and limitations of applying commonly used modeling techniques such as GPs and NNs to these data.

We consider analyzing and modeling voltage measurement data generated by a laboratory-built electrical circuit [38]. The observations are relatively smooth, concentrate on a low-dimensional attractor, and exhibit sensitive dependence on the initial condition. A nominal differential equations model, based on a simplistic description of the circuit components, has systematic deficiencies when fitted to the data. A more sophisticated attempt to build a physical model for the data using the circuit simulator SPICE [20, 40] did not fix these problems. We hence focus on modeling and prediction using GP and NN models. For both methods, we train a one-step ahead predictor based on the input-output pairs of m preceding observations and the current one for some embedding length $m > 0$. To investigate the capacities of the predictors in capturing the dynamics, we investigate the tradeoff between one-step prediction and long-term tracking. We find that both models perform similarly in one-step prediction, and the prediction error decreases as m increases. In contrast, the ability of the model propagations to track the observations improves at first but then degrades as m becomes larger for both models, which suggests a moderate value of m produces better balance between one-step prediction and long-term tracking.

One of our goals for analyzing observations generated from some unknown dynamics is to investigate the effects and characterizations of the model and observational errors. We consider this aspect by performing simulations with data generated by our fitted models. The fitted models capture the chaotic character of the observations, and the simulated data is qualitatively similar to the observations. We explore the effects of model and observational errors on the likelihood function and the identifiability of the initial state. We find that with independent and identically distributed (i.i.d.) observational error and no model error, the likelihood ratio between the true initial state and neighboring points increases exponentially in the number of observations. However, with even a tiny stochastic model error but otherwise correct dynamics, the true initial state no longer maximizes the likelihood function, and there

does not seem to be an initial state able to track the observations for an indefinitely long time. A temporally correlated observational error with no model error, on the other hand, preserves the identifiability of the true initial state as the maximizer of the likelihood function. The information provided by the likelihood about the true initial state also grows exponentially with more observations but at a lower rate than that for the i.i.d. observational error. To connect with our circuit system, we also perform simulations with a model error mimicking the thermal noise estimated from a reconstructed physical circuit and show the previously fitted GP and NN models seem to provide an accurate description of the long-term dynamics.

The rest of the paper is organized as follows. In section 2, we introduce the data set, investigate the deficiencies of the nominal model in capturing the dynamics present in the data, and discuss the ultimately unsuccessful efforts to model the system using the circuit simulator SPICE [20, 40]. In section 3, we compare the prediction performances of GP and NN models and investigate their capacities in describing the long-term dynamics. Section 4 considers the effects of model and observational errors using simulation data qualitatively similar to the observations and compares the fitted models in section 3 with simulations under model errors estimated from a reconstructed physical circuit.

2. The circuit data. In this work, we consider a time series consisting of voltage measurements of a laboratory-built electrical circuit [38]. The circuit was built on a breadboard using capacitors, resistors, operational amplifiers, and multipliers. A diagram of the circuit is shown in Figure 1, where V_1, V_2, and V_3 are the nodes at which voltages are measured. In the rest of the section, we refer to the three voltages measurements V_1, V_2, and V_3 as coordinates x, y, and z, so that the observation at time step n is $s_n = (x_n, y_n, z_n)$. The circuit was allowed to run for several minutes before data collection, and the measurements were taken at a frequency of 10kHz for about 1.5 minutes, resulting in three time series each with length 1 million. There were nine sets of measurements taken under different conditions in [38, Chapter 2], and the data we consider here belong to "set7." Figure 2 shows the observations for the initial 100,000 time points, and Figure 3 shows the trajectories of the initial 1000 observations. The data points loop around and fall densely on a two-dimensional manifold, as shown in Figure 2. Moreover, Figure 3 shows the observed trajectory is fairly smooth, indicating relatively low observational noise. Figure 4 shows the difference in Euclidean norm at each subsequent time point between two trajectories that start close by. Specifically, we select from all time points t, other than the initial one, the t^* that minimizes $\|s_0 - s_t\|$, and we consider the differences $\|s_n - s_{t^*+n}\|$ for $n = 0, 1, \ldots, 500$. The exponentially growing divergence of the two trajectories suggests the observations exhibit sensitive dependence on the initial conditions.

Because of the simple structure of the circuit, applying Kirchhoff's law to the idealized behavior of the circuit components yields the following nominal model:

$$\begin{aligned} \frac{dx}{dt} &= a_1 y, \\ \frac{dy}{dt} &= -a_2 y + a_3 x - a_4(x + z) - a_5 x z^2, \\ \frac{dz}{dt} &= a_6 x. \end{aligned} \tag{1}$$

This set of ODEs (1) is isomorphic to the Moore–Spiegel system [39], which is a nonlinear

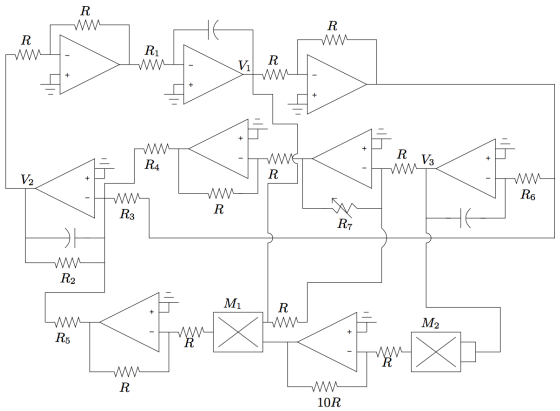

Figure 1. *Circuit diagram* [38, *Figure* 2.5]. V_1, V_2, *and* V_3 *indicate the nodes at which voltages are measured.*

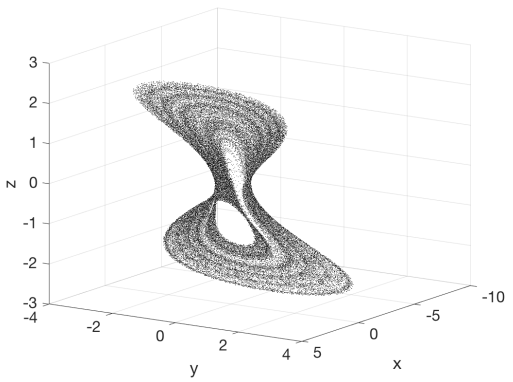

Figure 2. *The initial* 100,000 *observations.* x, y, *and* z *correspond to the voltage measurements* V_1, V_2, *and* V_3 *accordingly. The observations in the* 3D *space fall on a two-dimensional attractor.*

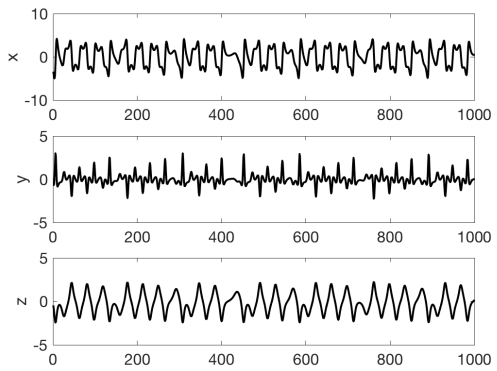

Figure 3. *Trajectories of the initial* 1000 *observations.*

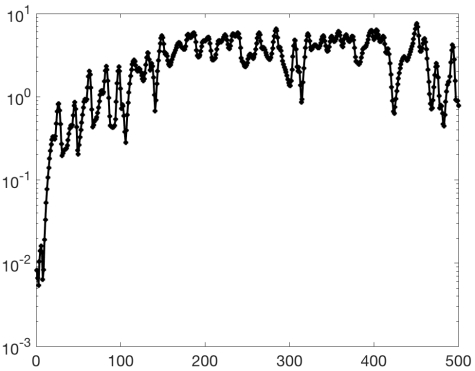

Figure 4. *The trajectory differences* $\|s_n - s_{t^*+n}\|$ *for* $n = 0, 1, \ldots, 500$. s_{t^*} *is the closest observation to the initial observation* s_0.

thermodynamical oscillator that has its physical origin in fluid dynamics. It models the displacement $z(t)$ of a small mass element attached at a fixed point to an elastic spring oscillating in a temperature stratified fluid. The element exchanges heat with the ambient fluid and its buoyancy depends on the temperature. The Moore–Spiegel system, like the Lorenz attractor, is one of the classical low order dynamical systems that exhibit chaotic behavior for certain choices of the parameter values [7, 39].

Since our data is measured from a real circuit, obtaining an accurate physical representation of the underlying dynamics would be helpful to model the system and the error terms. Therefore, the nominal model derived from the theoretical description of the circuit serves as our first choice of the system dynamics. However, [38] points out that significant discrepancies exist between the observations and the nominal model (1). For example, with the parame-

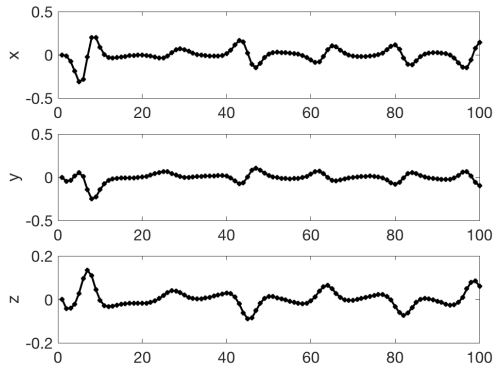

Figure 5. *The difference between observations and one-step ahead predictions at the initial* 100 *time points with the fitted nominal model* (1).

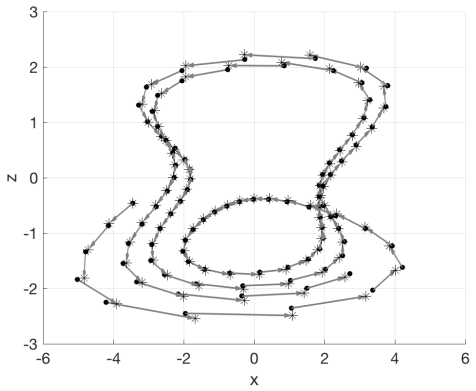

Figure 6. *The observations (dot) and the predictions (asterisk) for the initial* 100 *time points projected to the x-z plane, with the fitted nominal model* (1).

ter values used in the circuit, (1) settles down to a periodic orbit, whereas the observations manifest chaotic behavior. To investigate the deficiency of the nominal model, we estimate the parameters of (1) by minimizing the sum of squared one-step ahead prediction errors of the initial 2000 observations. Figure 5 shows the prediction errors of the fitted model (1) for the initial 100 time points. The trajectories for the prediction errors are fairly smooth and systematic, suggesting possible dynamics not captured by model (1). Figure 6 shows the observations and predictions projected to the x-z plane. Systematic departures of predictions from observations are also noticeable. For example, on the lower left portion of the plane where $x < 0$ and $z < 0$, the predictions tend to be ahead of the observations, while on the lower right portion where $x > 0$ and $z < 0$, the predictions tend to lag behind the observations. The systematic patterns in the prediction errors suggest the inadequacy of model (1) in describing the dynamics present in the observations.

In reality, the circuit components do not behave as simply as the nominal model (1) suggests. For example, the nominal model does not take into consideration any parasitic elements of the circuit, which are unavoidable and include stray inductance, capacitance, and resistance [3, 29]. These parasitics can alter the behavior of the circuit depending on the frequency of the signal, which in turn depends on the circuit components through (1). This intricate interaction is not captured by the nominal model, which assumes constant values for the circuit components at all frequencies. In addition, (1) models the behavior of the circuit on a macro level, but does not take into account effects on a micro level such as thermal noise, which is stochastic in nature. Therefore, our next attempt to obtain a more realistic physical description of the system by taking into account the complex behaviors of the circuit components is to use the circuit simulator SPICE [20, 40].

SPICE (Simulation Program with Integrated Circuit Emphasis) is a general-purpose electrical circuit simulator used in circuit design to verify the circuit operation at transistor level and predict the behavior of the designed circuit [40]. It enables designers to simulate the circuit even before prototyping. We use LTspice [20], a version of the SPICE simulator developed by

Linear Technology Corporation, to simulate the circuit in [38] in an attempt to obtain a good representation of the underlying dynamics. However, we were not able to obtain simulation results close to the observations. In particular, for all SPICE simulations with various initial values and component values the same as or slightly perturbed from their nominal values in [38], the resulting attractors in some cases do not qualitatively resemble that in Figure 2 and the simulations always have systematically much larger output ranges for the three voltages than the observations.

The main obstacle in carrying out a SPICE simulation of the circuit used in [38] is finding accurate component models and correctly estimating the parasitics involved. Unfortunately, no SPICE model for the original multiplier is available, so a drop-in replacement must be used as a substitute. The models for the amplifiers and multipliers also have unrealistic internal parasitic parameters that lead to numerical instabilities in the simulation. To account for these issues, simulations were run with different parasitic inductances and capacitors added to attempt to recover the original measurements. However, though some simulations qualitatively displayed the expected chaotic behavior of the system, the results were numerically unstable and unable to reliably match the measured data.

Given the ultimately unsuccessful efforts to obtain a physically inspired dynamics, in the next section, we will explore alternative data-driven models for the same one-step ahead prediction problem and investigate their capacities in capturing the long-term dynamics.

3. Gaussian process and neural network models. In this section, we consider a Gaussian process (GP) and a neural network (NN) model for prediction based on m previous observations with $m \geq 1$. Specifically, denote the observation at time step k as $s_k = (s_k(1), s_k(2), s_k(3))$, $k = 1, \ldots, N$. For each component $\nu \in \{1, 2, 3\}$, we model the outputs $s_k(\nu)$ as a GP (or NN) with inputs $(s_{k-m}, \ldots, s_{k-1})$ for $k = m+1, \ldots, N$. Note that each input is $3m$-dimensional. Due to the time series nature of the data, each observation can serve as either input or output in different training examples. For example, $s_{m+1}(\nu)$ is the target output in the first training example with input $(s_1, \ldots, s_m)$. Moreover, $s_{m+1}(\nu)$ is also part of the input $(s_2, \ldots, s_{m+1})$ in the second training example with $s_{m+2}(\nu)$ as its desired output. However, for each training example, the output and input share no common observations. A similar autoregressive structure is also considered in [4, 14, 53].

3.1. Model details and fitting. For the GP model [45, 46], we consider the following squared-exponential covariance function:

$$\text{Cov}\left(f(q), f(r)\right) = \sigma^2 e^{-\sum_{i=1}^{\beta} \left| \frac{q_i - r_i}{\theta_i} \right|^2} \tag{2}$$

for some $q, r \in \mathbb{R}^\beta$ and $\beta = 3m$. Furthermore, we include a nugget effect $\gamma > 0$ so the covariance matrix is $\Sigma = \sigma^2(R + \gamma I)$, where the (i, j)th element of R is the correlation between the ith and jth outputs. Note that the GP with the inputs and outputs specified above is not an internally consistent model for the observations, in the sense that the observations can not follow a joint normal distribution since they appear both as the outputs and inputs of the GP. A similar situation appears in [16], which considers emulation of dynamic computer codes using GP. We nonetheless use GP as a tool to fit a predictor by estimating parameters through maximizing the ostensible Gaussian likelihood and making predictions through interpolation.

Denote the outputs from all training examples as $\mathbf{s} = (s_{m+1}(\nu), \ldots, s_N(\nu))$ and the range parameters as $\theta = (\theta_1, \ldots, \theta_{3m})$; then the log-likelihood function by profiling over the scale parameter σ^2 satisfies

$$2l(\theta, \gamma) = -\log|R(\theta) + \gamma I| - n \log \mathbf{s}^T (R(\theta) + \gamma I)^{-1} \mathbf{s} - n \log(2\pi) + n \log n - n, \tag{3}$$

and

$$\hat{\sigma}^2 = \frac{\mathbf{s}^T (R(\theta) + \gamma I)^{-1} \mathbf{s}}{n},$$

where $n = N - m$. Then (3) is maximized to obtain maximum likelihood estimates (MLEs) $\hat{\theta}$ and $\hat{\gamma}$. At time step t with input $(s_{t-m}, \ldots, s_{t-1})$, the prediction of $s_t(\nu)$ is made by the empirical best linear predictor (EBLP),

$$\hat{s}_t(\nu) = \Sigma^T_{\mathbf{s}s_t(\nu)}(\hat{\theta}, \hat{\gamma}) \Sigma^{-1}_{\mathbf{ss}}(\hat{\theta}, \hat{\gamma}) \mathbf{s},$$

and is calibrated by the empirical mean squared error (EMSE),

$$\mathrm{EMSE}(\hat{s}_t(\nu)) = \Sigma_{s_t(\nu)s_t(\nu)}(\hat{\theta}, \hat{\gamma}) - \Sigma^T_{\mathbf{s}s_t(\nu)}(\hat{\theta}, \hat{\gamma}) \Sigma^{-1}_{\mathbf{ss}}(\hat{\theta}, \hat{\gamma}) \Sigma_{\mathbf{s}s_t(\nu)}(\hat{\theta}, \hat{\gamma}).$$

To address the concern that (3) may possess multiple maximizers, we experimented with different starting points for θ and γ. Although the final estimates obtained from multiple starting points differ, the relative differences of the maximized profile log-likelihoods (3) and the prediction errors are all less than 10^{-4} and 10^{-3}, respectively. Therefore, we do not see any evidence that the choice of initial estimates materially affects our results. A second model we experiment with is a feed-forward NN, a detailed account of which can be found in [25]. We fit an NN model consisting of one hidden layer with 20 neurons using the same set of inputs and outputs as those for fitting the GP model. The estimation and prediction are carried out using MATLAB Neural Network Toolbox 9.1 [10]. Note that for both GP and NN, three separate models are fitted for predicting the three components of the observation.

3.2. Prediction results. In this subsection, we investigate the one-step prediction and long-term tracking performances of the GP and NN models discussed in section 3.1. For each component ν, we fit GP and NN models based on the initial N observations and evaluate the root mean squared errors (RMSEs) for predicting in-sample and out-of-sample. We consider two sets of observations when evaluating out-of-sample: one is the next N observations after the initial N observations used to fit the model, and the other is the N observations at the tail end of the data set. Since the data are collected in a chronicled order and the properties of the electrical components may change with environmental conditions such as temperature as the circuit operates, we expect the tail part of the data set to be the most disparate to the in-sample data, and hence may provide information about possibly time-varying parameters.

Figure 7 shows the in-sample and the two types of out-of-sample RMSEs for the GP model fitted with the initial $N = 2000$ observations. Prediction errors decrease and gradually stabilize as the embedding length increases. The out-of-sample error evaluated with the tail part of the data set is only slightly larger than that evaluated with the second N observations, suggesting little variation of the parameters across time. The results for the NN model are

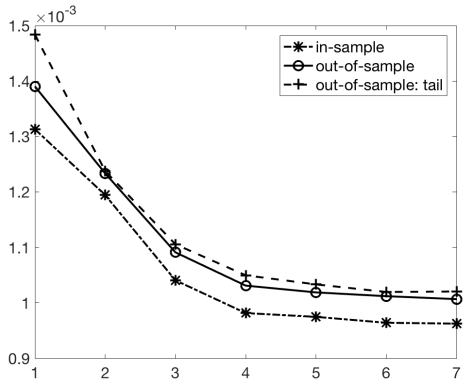

Figure 7. *RMSEs at each embedding length m for predicting the initial N observations (in-sample), the next N observations (out-of-sample), and the tail N observations (out-of-sample: tail) using the GP model. $N = 2000$, $m = 1, \ldots, 7$.*

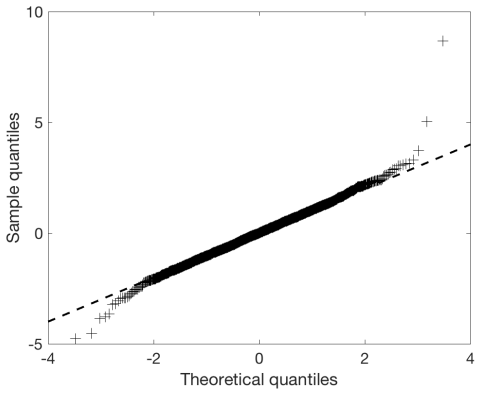

Figure 8. *QQ-plot of the out-of-sample standardized prediction errors for predicting the 3rd component when $m = 7$. Dashed reference line has slope 1 and intercept 0.*

only slightly worse; the differences of RMSEs between the two models are on the order of 10^{-5}, which is two orders of magnitude smaller than the prediction error itself. Figures SM1 and SM2 in Supplement SM1 show the RMSEs for the NN model and the comparison between the two models.

The use of GP provides a statistical basis for a model diagnostic. Under the GP model assumption for $\{s_k(\nu)|k = m+1, \ldots, N\}$, for $\nu \in \{1, 2, 3\}$, we have that the standardized prediction errors are normally distributed. In other words, we expect that

$$\frac{\hat{s}_k(\nu) - s_k(\nu)}{\sqrt{\mathrm{EMSE}(\hat{s}_k(\nu))}} \sim \mathcal{N}(0, 1)$$

under the GP model. Figure 8 shows the QQ-plot of the out-of-sample standardized prediction errors for predicting the third component ($\nu = 3$) when $m = 7$. The standardized errors are generally well approximated by a standard normal distribution except for a small fraction of outliers.

Note that both the GP and NN models are constructed and fitted for one-step prediction based on m previous observations, while our aim is to understand the underlying long-term dynamics based on the observations. As a result, we investigate the capabilities of the previously developed models in long-term prediction. Specifically, we propagate the fitted models out-of-sample from multiple starting points and compare the trajectories of the model propagation and observations. One quantitative measure of long-term predictability is hence the area enclosed by the two trajectories. For $\nu \in \{1, 2, 3\}$, denote the model propagation and observations from the ith starting point as $\{p_k^{(i)}(\nu)\}_{k=1:T}$ and $\{s_k^{(i)}(\nu)\}_{k=1:T}$, respectively, for a tracking length of T. We define the area $A^{(i)}(\nu)$ between the two trajectories as the sum of the areas of the trapezoids whose vertices are consecutive observation and model propagation

points, i.e.,

$$A^{(i)}(\nu) = \sum_{k=1}^{T-1} \left[\frac{1}{2} \left(\max(s_{k+1}^{(i)}(\nu), p_{k+1}^{(i)}(\nu)) + \max(s_k^{(i)}(\nu), p_k^{(i)}(\nu)) \right) - \frac{1}{2} \left(\min(s_{k+1}^{(i)}(\nu), p_{k+1}^{(i)}(\nu)) + \min(s_k^{(i)}(\nu), p_k^{(i)}(\nu)) \right) \right]. \tag{4}$$

An illustration of this definition of area is shown in Figure SM3 in Supplement SM1. We calculate the average area over all components and all starting points defined by

$$\overline{A} = \frac{1}{3L} \sum_{i=1}^{L} \sum_{\nu=1}^{3} A^{(i)}(\nu) \tag{5}$$

and use that as a measure of the long-term predictability of a model. The area between two trajectories, instead of standard pointwise discrepancy metrics such as MSE, is used because we view a trajectory that is largely correct except for a small phase shift as being an accurate prediction, which would be reflected in a small value for our metric, whereas MSE could be quite large for even small phase shifts.

Figure 9 shows the average area between the model propagation and observations over $L = 500$ out-of-sample starting points each tracking out a length of $T = 2000$. We choose the starting points every 200 steps to make the starting points fairly evenly distributed over the attractor. For both models, when increasing the embedding length, the long-term predictability improves for small m and then deteriorates. This pattern suggests that an embedding length that is too large ($m > 2$ for GP and $m > 3$ for NN) may produce a model well suited for one-step prediction while performing worse for long-term tracking with our parameter estimates. Note that smaller one-step prediction error does not necessarily imply better long-term tracking performance if the model is misspecified, e.g., if the system contains stochastic model error. So in the following, we focus on the GP and NN models with $m = 2$ and $m = 3$, respectively.

The fact that the embedding length leading to the best performance in both one-step prediction and long-term tracking is larger than one for both GP and NN models is an indication that the state of the circuit system may not solely consist of the three voltage measurements at one time point. In fact, the circuit components have different bandwidths, and hence they respond to signal with different rates. Though the multipliers and amplifiers were selected for their flat frequency response, in practice there are small variations in the response on the order of a percent over the bandwidth of the circuit. As a result, a single measurement frequency will not perfectly capture the response of the components and subsequently the circuit. In this case, using an embedding length larger than one, i.e., incorporating previous observations, in the prediction can be seen roughly as a way to at least partially recover the frequency information, which presumably is also part of the state. The need to include past states in the prediction could also be due in part to observational errors, and it is difficult to distinguish between these possibilities.

Another interesting feature of the long-term prediction is the region on the state space where model propagations tend to lose track of the observations. To measure such a divergence

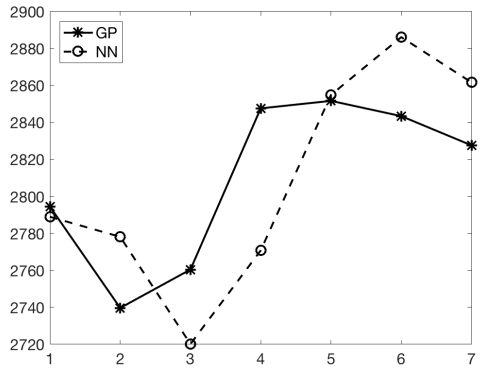

Figure 9. *Average area* $\overline{A}$*, defined in* (5)*, between observations and model propagations in* 2000 *steps for* 500 *out-of-sample starting points when* $m = 1, \ldots, 7$.

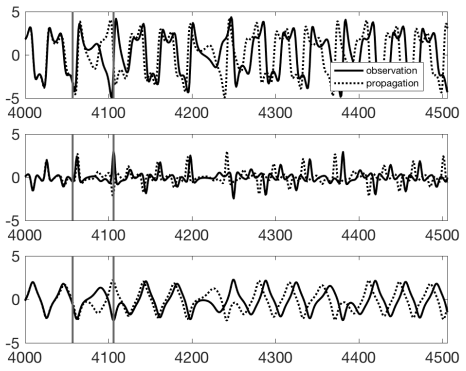

Figure 10. *Trajectories for observations and NN model propagations when* $m = 3$*. Starting time point is* 3801*. The interval* $[4057, 4106]$ *surrounded by vertical lines is the first window whose average area over the three components exceeds* 65.

of the model propagation from the observations, we use a moving window of length 50 and calculate the average area over the three components between the two trajectories inside the window. An average window area exceeding 65, by empirical experiments, seems to be a good indicator of a divergence between the model propagations and observations. Figure 10 shows one example of such a divergence pattern. Note that following the divergence, although the model propagations sometimes get back on track with the observations briefly, the tracking becomes noticeably worse. Figures 11 and 12 show, for GP and NN models, the regions on the state space where a divergence first occurs for the 500 pairs of trajectories. Note that both models suggest that the divergence largely concentrates in the region with the most significant nonlinearity, which is approximately $\{(x, y, z)||z| < 0.5, 0 < x < 1\}$. This phenomenon agrees with some existing results in the numerical analysis literature [35, 49, 52] that report amplification of numerical errors near the saddle points of chaotic attractors.

4. Identifiability of initial state under model and observational errors. In section 3, GP and NN predictors are constructed based on the observed data, and their performances in both one-step ahead and long-term predictions are investigated with different embedding lengths. In this section, we build on the fitted models in section 3 and the insights drawn from their long-term tracking performances to construct simulation systems that largely capture the chaotic character of the observed data. With such simulations, our goal is to understand the implications of applying statistical methods to infer important components of the system, particularly the initial state, when the data manifests chaotic behaviors and the system contains stochastic model and observational errors.

Reference [11] shows that the likelihood function for the initial state of a chaotic logistic map exhibits complex and irregular behaviors. In this section, we explore, through simulations, the extent to which similar behaviors of the likelihood functions arise for systems whose realizations are qualitatively similar to our observations. In addition, we compare the long-

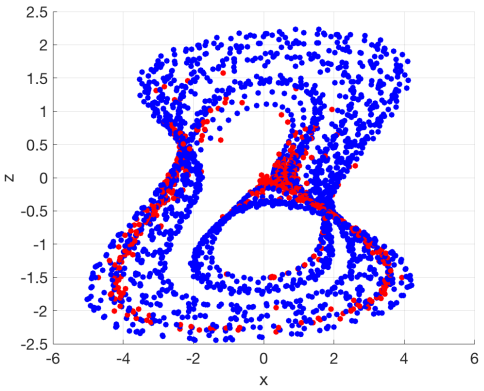

Figure 11. *Blue is a subset of the observations displayed to represent the attractor; red points show where the average area in a window of length* 50 *between the observations and GP model propagations with* $m = 2$ *first exceeds* 65 *for the* 500 *pairs of trajectories.*

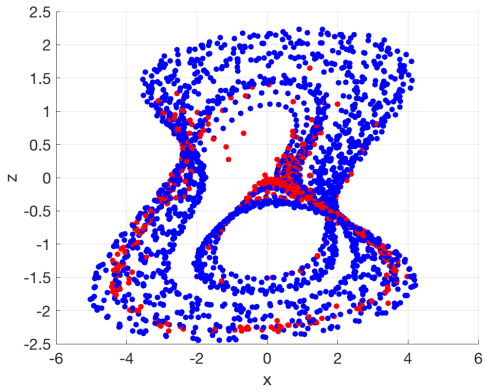

Figure 12. *Same as Figure* 11 *but using NN model propagations with* $m = 3$.

term predictability of the fitted GP and NN models in section 3 with simulations generated with an electronic noise term based on measurements from a reconstructed physical circuit. We found that the "chaotic likelihoods" shown in [11] for the logistic map, an ideal chaotic system, also appear for our system, which is constructed based on real chaotic observations. The fact that this NN approximation to the observed circuit data reproduces the extreme sensitivity to initial conditions found in simple simulated chaotic dynamical systems is a noteworthy finding given that the NN used is generic and was not developed to be able to mimic chaotic dynamics. However, even a tiny stochastic model error in the system changes the identifiability of the initial state significantly, suggesting that likelihood functions for initial conditions based on real data will never show the extreme behavior seen in [11]. Due to the computational difficulties of fitting a GP to a large training set, we focus on NN models in sections 4.1 and 4.2. However, we comment on the simulation results using a GP model fitted with a moderate number of observations at the end of section 4.1.

We consider the following state space formulation:

$$u_{t+1} = M(u_t) + \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, Q), \tag{6a}$$

$$w_t = H u_t + \eta_t, \tag{6b}$$

where (6a) models the evolution of the underlying states u_t with a stochastic model error ε_t, and (6b) models the noisy observations w_t with an observational error η_t. In this section, we investigate the effects of both i.i.d. and temporally correlated observational errors. Using a much larger training set here, the initial 400,000 observations, we explored multiple NN configurations. We found that, for an embedding length $m = 3$, a feed-forward NN model with two hidden layers and 10 neurons for each layer gave the smallest prediction errors, in contrast to the model with one hidden layer used earlier. Thus, we mainly report results for this two-layer model, although other models give similar results (see the end of section

4.1). We used a large training set for this exercise because we are interested in obtaining the best approximation we can to the circuit's behavior as a basis for further simulations. In contrast, in section 3, where we used a training set of only 2000 observations, our goal was to show that it is possible to fit an accurate model with a modest sample size. Furthermore, in section 3, we wanted to make a fair comparison to the GP approach for which maximizing (3) for larger sample sizes is computationally difficult. The results in section 3 indicate that $m = 3$ appears to perform well in long-term predictions for NN models fitted by minimizing the one-step prediction errors, and by using a large training set, we expect this model to be able to capture most of the dynamics in the data. Note that with this larger training set, we indeed obtain a better fit to the data. For example, the RMSE evaluated using the tail 2000 observations decreases by 8% compared to that of the NN model fitted with only the initial 2000 observations considered in section 3. Since the NN model predicts based on three previous observations, in this case, we have an augmented state vector $u_t^T = (\widetilde{u}_{t-2}^T, \widetilde{u}_{t-1}^T, \widetilde{u}_t^T)$, a model mapping M_N, and an observational mapping H_N defined as

$$\begin{aligned} &\begin{bmatrix} \widetilde{u}_{t-1} \\ \widetilde{u}_t \\ \widetilde{u}_{t+1} \end{bmatrix} = M_N \left(\begin{bmatrix} \widetilde{u}_{t-2} \\ \widetilde{u}_{t-1} \\ \widetilde{u}_t \end{bmatrix} \right) + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \widetilde{\varepsilon}_t \end{bmatrix}, \qquad \widetilde{\varepsilon}_t \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2 I_3), \\ &\widetilde{u}_{t+1} = f_{NN}(\widetilde{u}_{t-2}, \widetilde{u}_{t-1}, \widetilde{u}_t) + \widetilde{\varepsilon}_t, \qquad H_N = \begin{bmatrix} \mathbf{0} & \mathbf{0} & I_3 \end{bmatrix}, \end{aligned} \tag{7}$$

where $f_{NN}(\cdot)$ is the fitted NN predictor.

4.1. Independent and identically distributed observational error. In this subsection, we consider the observational error $\eta_t \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \sigma_\eta^2 I_3)$ and investigate the effects of model error on the likelihood functions. When there is no model error, the log-likelihood function of the initial state u_0 satisfies

$$2l_n(u_0) = -\sum_{k=0}^{n-1} \left\| w_k - HM^k(u_0) \right\|^2 / \sigma_\eta^2 - 3n \log(\sigma_\eta^2) - 3n \log(2\pi), \tag{8}$$

and the profile log-likelihood function obtained by profiling over σ_η^2 satisfies

$$\begin{aligned} 2\widetilde{l}_n(u_0) &= -3n \log(\hat{\sigma}_\eta^2) - 3n \log(2\pi) - 3n, \\ \hat{\sigma}_\eta^2 &= \frac{1}{3n} \sum_{k=0}^{n-1} \left\| w_k - HM^k(u_0) \right\|^2. \end{aligned} \tag{9}$$

In our experiments, the standard deviation for the observational error is $\sigma_\eta = 10^{-3}$, and that for the model error is chosen as small as $\sigma_\varepsilon = 10^{-12}$. We are interested in comparing the likelihoods (8) and (9) for observations generated with and without model errors. Specifically, we investigate the likelihood functions on a range of the last element of the initial state while fixing the other elements at their true values. Note that although the dimension of the state vector is nine for the NN model, the last element of a state corresponds to the z coordinate of the most recent observation. For the NN model, the simulated observations manifest a qualitatively similar attractor as the real data. Figures SM4–SM5 in Supplement SM1 show realizations of the simulation models.

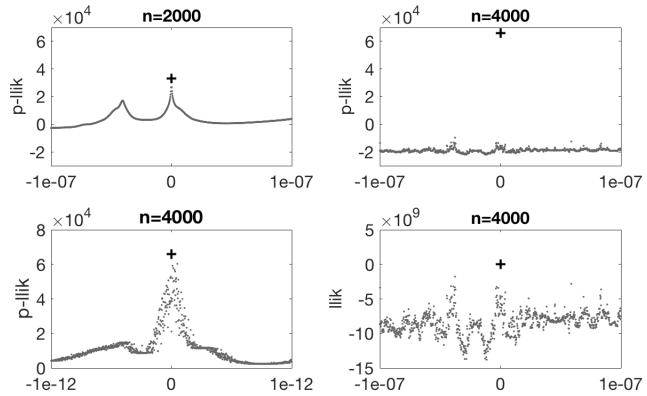

Figure 13. *Profile (i.e., maximized over σ_η^2) log-likelihoods (9) (top row and bottom left panel) and log-likelihood (8) (bottom right panel) for a range of the ninth element of the initial state with the other elements fixed at their true values. The simulated NN model M_N has no model error and has observational error $\eta_t \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \sigma_\eta^2 I_3)$, $\sigma_\eta = 10^{-3}$. Labels on the horizontal axes are differences of the simulation points minus the true initial state. The plus sign indicates true initial state.*

For the NN model, Figure 13 shows, with no model error, the likelihoods on a range of the ninth element of the initial state with the other elements fixed at their true values. We observe the same general pattern as in [11]: the likelihood functions on a fixed interval of initial states become more jagged and irregular as we increase the number of observations, and only show some smoothness when focusing on a narrower interval (see, for example, the top right and bottom left panels of Figure 13). In addition, much more information about the true initial state can be learned with more observations, since the relative difference of the profile log-likelihoods between the true initial state and the neighboring simulation points becomes more pronounced. Note that for a fixed number of observations, the profile log-likelihood is dramatically sharper than the log-likelihood at the true initial state with σ_η^2 fixed at its true value, and hence contains more information about the truth.

The top right panel of Figure 13 seems to indicate that the true initial state maximizes the profile log-likelihood. Admittedly, the simulation is based on finitely many discrete points in some neighborhood of the true initial state, and hence does not fully represent the entire likelihood function. However, to add evidence to the remark that the true initial state can be identified as the maximizer of the likelihood function under no model error, in the bottom left panel of Figure 13, we evaluate the profile log-likelihood for 1000 points of the ninth element of the initial state in the interval $(u_0(9) - 10^{-12}, u_0(9) + 10^{-12})$ when $n = 4000$. In this case, the distance between the true initial state and the closest simulation point is on the order of 10^{-15}, and the true initial state still has the largest likelihood in the simulation interval. Note that due to the finite precision of computers, the model being simulated essentially has a discrete state space, and a difference of 10^{-15} is fairly close to machine precision. As a result, the true initial state may maximize the likelihood for a large enough number of observations given the effectively discrete state space. In section 4.2, we will look more closely into the comparison of the likelihoods of the true initial state and its neighbors for an increasing

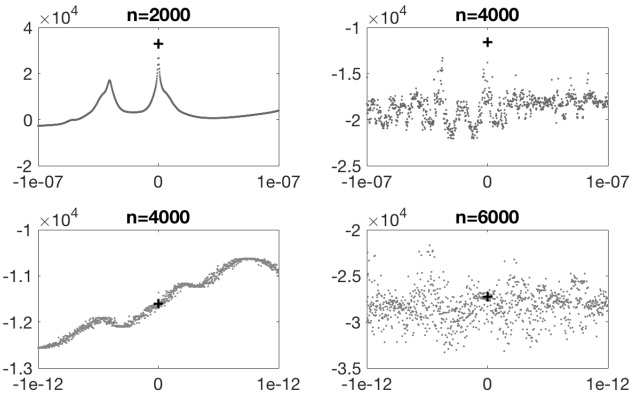

Figure 14. *Profile (i.e., maximized over σ_η^2) log-likelihoods (9) (all four panels) for a range of the ninth element of the initial state with the other elements fixed at their true values. The simulated NN model M_N has model error $\widetilde{\varepsilon}_t \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2 I_3)$ as defined in (7), $\sigma_\varepsilon = 10^{-12}$, and has observational error $\eta_t \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \sigma_\eta^2 I_3)$, $\sigma_\eta = 10^{-3}$. Labels on the horizontal axes are differences of the simulation points minus the true initial state. The plus sign indicates true initial state.*

number of observations. Finally, we point out that the simulation model considered here is an NN which by no means is designed to produce chaotic dynamics. However, it does yield a system that, when there is no model error, produces results for the likelihood with respect to the initial state like those shown in [11].

Figure 14 shows the profile log-likelihoods on a range of the ninth element of the initial state with the other elements fixed at their true values under a small model error with standard deviation $\sigma_\varepsilon = 10^{-12}$. Similar to the case without model error, we observe wilder behavior of the likelihood functions with more observations. However, increasing the number of observations from 2000 to 4000 no longer helps much in identifying the true initial state. In the top row of Figure 14, although more observations make the profile log-likelihood more concentrated around the true initial state, the difference in log-likelihoods between the true initial state and the neighboring points only increases slightly, compared with the sharp increase under no model error (top row of Figure 13). The pattern is more obvious on a smaller scale: see the bottom row of Figure 14, where the true initial state clearly no longer maximizes the likelihood. When the number of observations increases, we do not gain more information about the true initial state. An interesting feature to note in this case is that in the bottom row of Figure 14, the profile log-likelihood attains local maxima at different points when $n = 4000$ and $n = 6000$. It shows that, even with the correct dynamics for the deterministic part of the model, there does not exist a starting point that can track this simulated NN system with even a tiny stochastic model error for an indefinitely long time.

We highlight that the models considered here are derived from observed data and are not simulations under some idealized chaotic system like the logistic map. Therefore, Figures 13 and 14 show that the similarly irregular behavior of the likelihoods as those in [11] can happen under a dynamical system estimated from data when there is measurement error but no model error. However, Figure 14 shows that even a tiny model error can dramatically affect the

identifiability of the initial state when there is a sufficiently large number of observations. This result suggests that real systems, which we maintain should essentially always contain some stochastic model error, will not behave like idealized deterministic chaos plus measurement error when inferring the initial state from the observations. Moreover, the irregular behaviors of the log-likelihoods shown in Figures 13 and 14 also suggest that the estimation of the initial state would be an exceedingly difficult optimization problem in practice, especially if the unknown initial state were vector-valued.

Finally, we note that the patterns observed in Figures 13 and 14 do not depend on the structure of the NN. For example, we fit a one-layer NN with 20 neurons using the same training set and embedding length $m = 3$. Figures SM6–SM7 in Supplement SM1 show the similar behaviors of the log-likelihoods of the initial state under increasing observations. Furthermore, the "chaotic likelihood" is not particular to NN models and can also be reproduced by GP models fitted to the circuit data. For instance, we fit a GP model by maximizing (3) using the initial 4000 observations with an embedding length $m = 2$. Figure SM8 in Supplement SM1 shows the profile log-likelihoods (9) of the last element of the initial state under no model error. The likelihood functions become increasingly jagged and concentrated at the true initial state with more observations.

4.2. Temporally correlated observational error. In this subsection, we consider observational errors that are temporally correlated with an AR(1) structure. That is, we have $\eta_t = \phi\eta_{t-1} + v_t$ for some $\phi \in \mathbb{R}$ and $v_t \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \sigma_v^2 I)$. A temporally correlated observational error can be used to model dependencies in the measurement process resulting from, for example, a limited instantaneous rate of response to changes in signal for the measurement device. When there is no model error, the profile log-likelihood function of the initial state obtained by profiling over σ_η^2 satisfies

$$2\widetilde{l}_n(u_0) = -3n\log(\hat{\sigma}_\eta^2) - \log(|R|) - 3n\log(2\pi) - 3n, \tag{10}$$

where

$$\hat{\sigma}_\eta^2 = \frac{\mathbf{e}^T R^{-1}\mathbf{e}}{3n}, \quad R = \text{toeplitz}(1, \phi, \ldots, \phi^{n-1}) \otimes I_3, \quad \mathbf{e} = \left[w_k - HM^k(u_0)\right]_{k=0}^{n-1}.$$

We take $\phi = 0.9$ and $\sigma_v = 10^{-3}$ as the standard deviation for the innovation. Realizations of the simulated model under this temporally correlated observational error again form an attractor similar to that of the real data and are shown in Figure SM9 in Supplement SM1.

Figure 13 showed that more information about the true initial state can be learned with more observations under i.i.d. observational error. We now investigate this finding in more detail and compare the effects of different observational error structures. We use the difference in the profile log-likelihood function between the true initial state and the neighboring simulation points as a measure of identifiability of the true initial state. The neighboring simulation points are taken on a regular grid centered at the last three elements of the true initial state for the NN model. With this specification, we have a total of 26 simulation points excluding the true initial state. Denote the shortest distance from the simulation points to the true initial state as d. An illustration of this setup of the simulation points is shown in Figure SM10 in Supplement SM1.

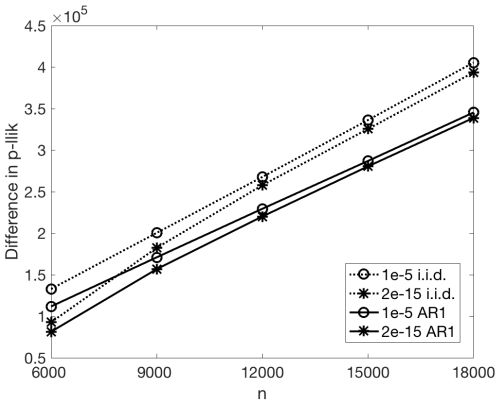

Figure 15. *Difference in the profile log-likelihood function between the true initial state and the neighboring simulation points for the NN model with no model error for various n. Numbers in the legend are the distances d between the simulation points and the true initial state.*

Figure 15 shows, for the NN model, the differences in the profile log-likelihood functions (9) and (10) under i.i.d. and AR(1) observational errors, respectively. For each observational error type, we experiment with simulation points closer to the true initial state with $d = 2 \times 10^{-15}$ and those further away with $d = 10^{-5}$. In accordance with the feature shown in the top row of Figure 13, as the number of observations increases, the difference of the profile log-likelihood between the true initial state and the neighboring points increases under both types of observational errors. Note that the linear rate of the increase in Figure 15 implies an exponential rate of increase in the likelihood ratio. In addition, as indicated by the larger slope of the dashed lines, the information about the true initial state grows faster in the number of observations for the i.i.d. observations than for the AR(1) observational errors measured with the same set of simulation points.

For both types of observational errors, the difference in the profile log-likelihood is larger between the true initial state and the simulation points further away than those closer. However, this discrepancy between points at different distances to the true initial state diminishes as the number of observations increases. This pattern shows that the neighboring points, regardless of their distance to the true initial state, become similarly worse in tracking the observations relative to the true initial state. This is due to the chaotic feature of the system so that even a tiny departure from the true initial state can lead to significant divergence in the propagation with a long enough horizon. In other words, an initial state with a tiny but nonzero departure from the truth, in the long run, does not have much advantage in tracking the observations over an initial state with a large departure.

We note a theoretical result [33, Theorem 3] regarding signal extraction that is relevant to what Figure 15 shows. That result proves that, for an invertible chaotic system Ψ under no model error and Gaussian observational noise, it is impossible to consistently infer any single state from the infinite two-sided observation time series. The result is established based on the existence of homoclinic points. Two points $x_0 \neq x_0'$ constitute a homoclinic pair [28, 33, 34] if $\lim_{|n|\to\infty}(1+\alpha)^{|n|}|\Psi^n(x_0) - \Psi^n(x_0')| = 0$ for some $\alpha > 0$. In words, the trajectories of the

two distinct points approach each other exponentially quickly both forward and backward in time. Consequently, if the observational noise is unbounded, for example Gaussian, there is a positive probability that no matter how many observations one has, the true state x_0 that generates the observations has a lower likelihood than its homoclinic point x_0', and hence cannot be inferred from the data. However, for all realizations of the ODE and NN models we have experimented with, we find that the true initial state has the highest likelihood among the simulation points, and we appear to gain exponentially growing information about the truth with more observations. One possible resolution for this apparent discrepancy between the theory and our empirical results lies in the relatively small observational noise used in the simulations, which leads to a perhaps tiny, though positive, probability that the homoclinic point gives a higher likelihood than the true initial state. As a result, we do not observe the effects of homoclinic points in our limited number of simulations.

4.3. Simulations with electronic noise. In this section, we consider the simulated NN model M_N with no observational error and a model error estimated from a reconstructed physical circuit. The circuit was built using the same components as those in [38] and was fully covered to be thermally controlled at 20 °C. We expect this reconstructed circuit to mimic the circuit system in [38], and thus provide an estimate of the model error encountered in reality. The model error in this case is taken to be the electronic noise in the circuit, which consists mostly of the Johnson (thermal) noise [31]. The Johnson noise is approximately white and has very nearly a Normal distribution [8]. The standard deviation for the Johnson noise was measured for one component in the circuit expected to contribute the most to the entire noise level, and was estimated to be 100μV with a 3dB bandwidth of 1kHz set by the integrator of the circuit. Thus, we maintain that the noise level for the entire circuit in [38] should fall in the range 100μV to 400μV by taking into account the noise contributed from other components and that the experiment in [38] was conducted with a higher room temperature. The simulated NN model with a model error implied by the physical circuit can be regarded as at least a fair approximation to the underlying dynamics of the circuit system. As a result, the performance of two realizations generated from this simulation model in tracking each other sets a benchmark on how well any fitted model can possibly track the observations. In the following, we revisit the fitted GP and NN models in section 3 and evaluate their tracking abilities by comparing with the simulation model under the estimated range of the model error.

Figure 16 shows the histograms of the average areas between the fitted GP model propagations and the real observations, and those between two realizations of the simulated model (7) for the same 500 starting points considered in section 3. The results for the fitted NN model in section 3 are very similar, and a comparison of both models with the simulations is shown in Figure SM11 in Supplement SM1. For a smaller estimate of the model error $\sigma_\varepsilon = 10^{-4}$, the tracking performance of the GP model is clearly inferior to the simulation model. It indicates that the fitted models do not seem to fully capture the dynamics if the model error is indeed this small. However, as the model error increases, we see much better agreement between the fitted models and the simulations. As discussed previously, the smaller end of the model error estimate $\sigma_\varepsilon = 10^{-4}$ is measured with one component of the reconstructed circuit, while the noise level of the entire circuit in [38] is expected to be larger due to contributions

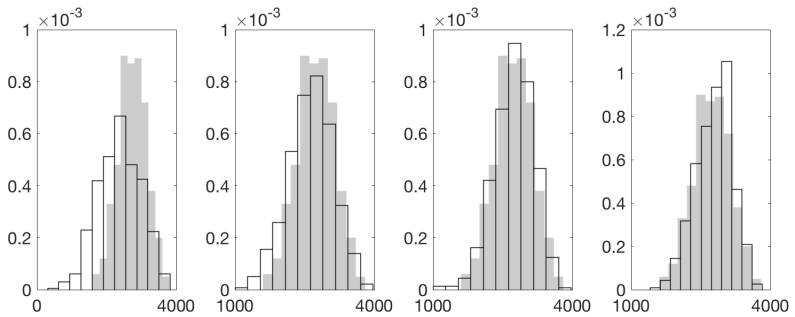

Figure 16. *Histograms (gray) of average areas between real observations and GP model propagations (embedding length $m = 2$) in* 2000 *steps for the* 500 *starting points considered in section* 3*; histograms (white with solid bar outlines) of average areas between two realizations in* 2000 *steps from* (7) *under model error with standard deviation σ_ε and no observational error for the same* 500 *starting points. From left to right: $\sigma_\varepsilon = 10^{-4}, 2 \times 10^{-4}, 3 \times 10^{-4}, 4 \times 10^{-4}$. Histograms are normalized so that the areas of bars in each histogram sum to one.*

from other components and a higher temperature. Consequently, the fitted GP model may approach the limit of how well one can track the system in the presence of a stochastic model error in the circuit. Table 1 additionally supports this point by showing the proportions of starting points from which the simulation model produces a smaller tracking area than the fitted models tracking the observations. When the model error $\sigma_\varepsilon = 10^{-4}$ is small, the tracking performance of the fitted models is worse than that of the simulations for around 65% of the starting points. However, for larger model errors, for example when $\sigma_\varepsilon = 3 \times 10^{-4}$, the fitted models track better compared with the simulations for about half of the starting points. Thus, if the true model error has this plausible standard deviation of 3×10^{-4}, then the GP and NN models track the system as well as possible.

Table 1

Proportions of starting points for which the average area between two trajectories is smaller for the simulations than for the GP and NN model propagations tracking the real observations. The simulation trajectories for each starting point are two realizations from (7) *under model error with standard deviation σ_ε and no observational error.*

σ_ε	10^{-4}	2×10^{-4}	3×10^{-4}	4×10^{-4}
GP	68.8%	56.2%	49.4%	43.2%
NN	65.4%	54.8%	46.8%	46.2%

5. Discussion. The analysis of observations generated from unknown nonlinear and chaotic dynamical systems poses significant challenges to modeling the underlying dynamics and characterizing stochastic model and observational errors. We looked into some aspects of these problems with voltage measurement data generated by a laboratory-built electrical circuit. The nominal model of the measurements and more sophisticated physical models using a circuit simulator show notable deficiencies in capturing the dynamics in the observations, so we turned to GP and NN models that are trained for one-step prediction. With the squared exponential covariance for GP and the feed-forward NN configurations we considered, both models perform similarly in one-step prediction. The prediction error decreases as the em-

bedding length increases. However, good short-term prediction does not necessarily lead to a good representation of the underlying dynamics. As we show for both GP and NN models, the best performance in tracking the observations long-term occurs at a moderately valued embedding length, while a larger embedding length results in quick deterioration in the tracking abilities. A natural alternative to fitting the circuit data is to use a state space model. We experimented with fitting one assuming no model error and a Normal observational error using a likelihood method. In particular, we used a GP for the dynamics and estimated the initial states, parameters of the GP, and observational error covariance by maximizing the likelihood function. However, we were not able to obtain results better than the fitted GP and NN predictors. The underperformance of the state space model in this case may be caused by the nonnegligible model error present in the circuit system that contains thermal noise. It is very difficult to estimate parameters for the deterministic nonlinear dynamics, model error covariances, and observational error covariances simultaneously, and we did not experience success when we tried such an approach.

To learn more about the effects of model and observational errors on the likelihood function of the initial state, we turned to simulations. The simulated data were generated by our fitted NN model, which manifests a low-dimensional attractor and chaotic features similar to the observations. We found that the effects of model and observational errors on inferring the initial state are quite different. In the absence of model error, when there are enough observations, the true initial state appears to maximize the likelihood function, and the likelihood ratio between the true initial state and neighboring points grows exponentially in the number of observations. Temporally correlated observational errors result in a slower rate of increase for the likelihood ratio but preserve the exponential growth. However, in the presence of a model error, even with a standard deviation as small as 10^{-12}, an increasing number of observations no longer provides unboundedly increasing information about the true initial state through the likelihood function. Moreover, in this case, there does not seem to exist an initial state capable of tracking the observations indefinitely even with the correct dynamics.

Note that the standard deviation of the model error considered in section 4.1 is significantly smaller than the measured electronic noise in the circuit, yet it is still impossible, through the likelihood function, to identify the initial state and hence track the observations in the long term even with the correct deterministic part of the dynamics. This points out that chaotic systems in reality, which almost always contain stochastic model errors, do not behave as ideally as toy systems for deterministic chaos in terms of the identifiability of the initial state under an increasing number of observations. More optimistically, our comparisons in section 4.3 show that it may be possible, at least in some cases, to construct predictors that generate long-term predictions approaching the tracking limit of the system given the existence of a stochastic model error in the dynamics.

Acknowledgments. The authors would like to thank Leonard A. Smith and Hailiang Du for providing the data and valuable feedback during this project.

REFERENCES

[1] H. Abarbanel, *Analysis of Observed Chaotic Data*, Springer, New York, 2012.

[2] H. D. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, *The analysis of observed chaotic data in physical systems*, Rev. Mod. Phys., 65 (1993), 1331.
[3] A. Agarwal and J. Lang, *Foundations of Analog and Digital Electronic Circuits*, Morgan Kaufmann, San Francisco, CA, 2005.
[4] M. Ardalani-Farsa and S. Zolfaghari, *Chaotic time series prediction with residual analysis method using hybrid Elman–NARX neural networks*, Neurocomput., 73 (2010), pp. 2540–2553.
[5] M. Assaad, R. Boné, and H. Cardot, *Predicting chaotic time series by boosted recurrent neural networks*, in International Conference on Neural Information Processing, Springer, New York, 2006, pp. 831–840.
[6] G. Atsalakis and K. Tsakalaki, *Simulating annealing and neural networks for chaotic time series forecasting*, Chaotic Model. Simul., 1 (2012), pp. 81–90.
[7] N. Balmforth and R. Craster, *Synchronizing Moore and Spiegel*, Chaos, 7 (1997), pp. 738–752.
[8] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communication*, Springer, New York, 2012.
[9] B. Bayly, I. Goldhirsch, and S. A. Orszag, *Independent degrees of freedom of dynamical systems*, J. Sci. Comput., 2 (1987), pp. 111–121.
[10] M. H. Beale, M. T. Hagan, and H. B. Demuth, *Neural Network Toolbox^{TM} User's Guide*, The MathWorks, Natick, MA, 1992.
[11] L. M. Berliner, *Likelihood and Bayesian prediction of chaotic systems*, J. Amer. Statist. Assoc., 86 (1991), pp. 938–952.
[12] S. A. Billings, *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-temporal Domains*, John Wiley & Sons, New York, 2013.
[13] P. Bryant, R. Brown, and H. D. Abarbanel, *Lyapunov exponents from observed time series*, Phys. Rev. Lett., 65 (1990), 1523.
[14] J. Q. Candela, A. Girard, J. Larsen, and C. E. Rasmussen, *Propagation of uncertainty in Bayesian kernel models—application to multiple-step ahead forecasting*, in Proceedings of the ICASSP, vol. 2, IEEE, Washington, DC, 2003, pp. 701–704.
[15] M. Casdagli, *Nonlinear prediction of chaotic time series*, Phys. D, 35 (1989), pp. 335–356.
[16] S. Conti, J. P. Gosling, J. E. Oakley, and A. O'Hagan, *Gaussian process emulation of dynamic computer codes*, Biometrika, 96 (2009), pp. 663–676.
[17] J. P. Crutchfield and B. S. McNamara, *Equations of motion from a data series*, Complex Syst., 1 (1987), pp. 417–452.
[18] J. Deppisch, H.-U. Bauer, and T. Geisel, *Hierarchical training of neural networks and prediction of chaotic time series*, Phys. Lett. A, 158 (1991), pp. 57–62.
[19] E. Diaconescu, *The use of NARX neural networks to predict chaotic time series*, WSEAS Trans. Comput. Res., 3 (2008), pp. 182–191.
[20] M. Engelhardt, *LTSpice/SwitcherCAD IV*, Linear Technology Corporation, Milpitas, CA, 2011.
[21] J. Faraway and C. Chatfield, *Time series forecasting with neural networks: A comparative study using the air line data*, J. Roy. Statist. Soc. C Appl. Statist., 47 (1998), pp. 231–250.
[22] J. D. Farmer and J. J. Sidorowich, *Predicting chaotic time series*, Phys. Rev. Lett., 59 (1987), 845.
[23] A. Gholipour, B. N. Araabi, and C. Lucas, *Predicting chaotic time series using neural and neurofuzzy models: A comparative study*, Neural Process. Lett., 24 (2006), pp. 217–239.
[24] I. Ginzburg and D. Horn, *Combined neural networks for time series analysis*, in Advances in Neural Information Processing Systems, Denver, CO, 1994, pp. 224–231.
[25] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, MA, 2016.
[26] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Appl. Math. Sci. 42, Springer, New York, 2013.
[27] G. Heydari, M. Vali, and A. A. Gharaveisi, *Chaotic time series prediction via artificial neural square fuzzy inference system*, Expert Syst. Appl., 55 (2016), pp. 461–468.
[28] M. W. Hirsch, S. Smale, and R. L. Devaney, *Differential Equations, Dynamical Systems, and an Introduction to Chaos*, Academic Press, San Diego, CA, 2012.
[29] P. Horowitz and W. Hill, *The Art of Electronics*, Cambridge University Press, New York, 1989.
[30] A. Jayawardena, W. Li, and P. Xu, *Neighbourhood selection for local modelling and prediction of hydrological time series*, J. Hydrol., 258 (2002), pp. 40–57.

[31] J. B. Johnson, *Thermal agitation of electricity in conductors*, Phys. Rev., 32 (1928), 97.
[32] K. Judd and A. Mees, *Modeling chaotic motions of a string from experimental data*, Phys. D, 92 (1996), pp. 221–236.
[33] S. P. Lalley, *Beneath the noise, chaos*, Ann. Statist., 27 (1999), pp. 461–479.
[34] S. P. Lalley, *Removing the noise from chaos plus noise*, in Nonlinear Dynamics and Statistics, Springer, New York, 2001, pp. 233–244.
[35] S. Liao, *On the reliability of computed chaotic solutions of non-linear differential equations*, Tellus A, 61 (2009), pp. 550–564.
[36] E. N. Lorenz, *Atmospheric predictability as revealed by naturally occurring analogues*, J. Atmos. Sci., 26 (1969), pp. 636–646.
[37] Q.-L. Ma, Q.-L. Zheng, H. Peng, T.-W. Zhong, and L.-Q. Xu, *Chaotic time series prediction based on evolving recurrent neural networks*, in International Conference on Machine Learning and Cybernetics, Vol. 6, IEEE, Washington, DC, 2007, pp. 3496–3500.
[38] R. L. Machete, *Modeling a Moore-Spiegel Electronic Circuit: The Imperfect Model Scenario*, Ph.D. thesis, University of Oxford, Oxford, UK, 2007.
[39] D. W. Moore and E. A. Spiegel, *A thermally excited non-linear oscillator*, Astrophys. J., 143 (1966), p. 871.
[40] L. W. Nagel and D. Pederson, *SPICE (Simulation Program with Integrated Circuit Emphasis)*, Tech. Report UCB/ERL M382, EECS Department, University of California Berkeley, Berkeley, CA, 1973.
[41] U. Parlitz, *Estimating Lyapunov exponents from time series*, in Chaos Detection and Predictability, Springer, Berlin, Heidelberg, Germany, 2016, pp. 1–34.
[42] A. Pikovsky, *Discrete-time dynamic noise filtering*, Sov. J. Commun. Technol. Electron., 31 (1986), p. 81.
[43] I. Poli and R. Jones, *A neural net model for prediction*, J. Amer. Statist. Assoc., 89 (1994), pp. 117–121.
[44] M. J. Powell, *Radial basis functions for multivariable interpolation: A review*, in Algorithms for Approximation, Clarendon Press, New York, 1987, pp. 143–167.
[45] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, 2006.
[46] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York, 1999.
[47] G. Sugihara and R. M. May, *Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series*, Nature, 344 (1990), pp. 734–741.
[48] F. Takens, *Detecting strange attractors in turbulence*, in Dynamical Systems and Turbulence, Springer, New York, 1981, pp. 366–381.
[49] J. Teixeira, C. A. Reynolds, and K. Judd, *Time step sensitivity of nonlinear atmospheric models: Numerical convergence, truncation error growth, and ensemble design*, J. Atmos. Sci., 64 (2007), pp. 175–189.
[50] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, *Determining Lyapunov exponents from a time series*, Phys. D, 16 (1985), pp. 285–317.
[51] F. Z. Xing, E. Cambria, and X. Zou, *Predicting evolving chaotic time series with fuzzy neural networks*, in International Joint Conference on Neural Networks, International Neural Network Society, Hoffman Estates, IL, 2017, pp. 3176–3183.
[52] L.-S. Yao, *Computed chaos or numerical errors*, Nonlinear Anal. Model. Control, 15 (2010), pp. 109–126.
[53] G. Zhang, B. E. Patuwo, and M. Y. Hu, *Forecasting with artificial neural networks: The state of the art*, Internat. J. Forecast., 14 (1998), pp. 35–62.
[54] J.-S. Zhang and X.-C. Xiao, *Predicting chaotic time series using recurrent neural network*, Chinese Phys. Lett., 17 (2000), 88.