

Parameter estimation through ignorance

Hailiang Du¹ and Leonard A. Smith^{1,2}

¹*Centre for the Analysis of Time Series, London School of Economics, London WC2A 2AE, England, United Kingdom*

²*Pembroke College, Oxford OX1 1DW, England, United Kingdom*

(Received 25 October 2010; revised manuscript received 10 February 2012; published 16 July 2012)

Dynamical modeling lies at the heart of our understanding of physical systems. Its role in science is deeper than mere operational forecasting, in that it allows us to evaluate the adequacy of the mathematical structure of our models. Despite the importance of model parameters, there is no general method of parameter estimation outside linear systems. A relatively simple method of parameter estimation for nonlinear systems is introduced, based on variations in the accuracy of probability forecasts. It is illustrated on the logistic map, the Henon map, and the 12-dimensional Lorenz96 flow, and its ability to outperform linear least squares in these systems is explored at various noise levels and sampling rates. As expected, it is more effective when the forecast error distributions are non-Gaussian. The method selects parameter values by minimizing a proper, local skill score for continuous probability forecasts as a function of the parameter values. This approach is easier to implement in practice than alternative nonlinear methods based on the geometry of attractors or the ability of the model to shadow the observations. Direct measures of inadequacy in the model, the “implied ignorance,” and the information deficit are introduced.

DOI: [10.1103/PhysRevE.86.016213](https://doi.org/10.1103/PhysRevE.86.016213)

PACS number(s): 05.45.Tp, 02.60.-x, 05.10.-a

I. INTRODUCTION

The estimation of physical constants (parameters) plays a central role in the physical sciences. Yet there is no general method of parameter estimation for nonlinear dynamical systems [1]. In what may have been one early use of least squares (LS) (see [2,3]), Gauss [4] predicted where the newly discovered Ceres would appear as it emerged from behind the sun. The prediction involved both a parameter (Newton’s gravitational constant) and initial conditions (brief observations of Ceres before occultation). This success, where other methods failed, supported both the least squares approach and the mathematical form of Newton’s laws. A minimum ignorance (MI) approach to parameter estimation for use in dynamical systems is presented and illustrated in nonlinear cases where the common “least squares” approaches can be systematically biased. While the focus is on cases in which the data archive is relatively large and the model structure is correct,¹ the approach may prove useful outside this perfect model scenario [7]. In the MI approach, many large sets of probability forecasts are made, each using different parameter values; the quality of those parameter values is determined by the quality of the corresponding set of probability forecasts. A measure of the internal consistency of probability forecasts is also introduced, providing quantitative insight into modeling inadequacy.

Parameter estimation for deterministic nonlinear models poses several challenges, as nonlinear processes can be sensitive to initial conditions and parameter specifications.

¹Specifically, we refer to cases where there is a parameter value for which the mathematical model is empirically adequate. The motion of Mercury is inconsistent with the mathematical form of Newton’s laws; an internally consistent description requires general relativity. It is not that the value of a parameter in Newton’s laws is uncertain, but rather the value is indeterminate [5]: no value will yield results consistent with observed planetary motion. In the “perfect model scenario,” the “true” parameter is unknown but does exist [6].

Traditional methods, like least squares, are suboptimal when forecast errors are non-Gaussian, even if the observational uncertainties are normally distributed. One aim of this paper is to stress that fact given the common, and often unguarded, use of least squares. Several methods have been proposed to address the shortcomings of traditional methods: McSharry and Smith [8] estimate model parameters by incorporating the global behavior of the model into the selection criteria; Creveling *et al.* [9] and Maybhate and Amritkar [10] have exploited synchronization for parameter estimation; Smith *et al.* [11] focused on the geometric properties of trajectories; Heald and Stark [12] include estimation of the noise model. Recently Quinn and Abarbanel [13] demonstrated that parameter and state can be estimated via evaluation of a discrete time path integral in model state space. They also note applications in a number of fields including neurobiology, atmospheric and oceanic sciences, cell biology, chemical engineering, wastewater treatment, and biochemistry. There are also variational approaches [14], multiple shooting methods [15], and sequential methods based loosely on the Kalman filter [15–18]. Several of these alternative approaches are contrasted with the MI approach in the conclusion section.

Results of MI parameter estimation are presented and critically examined for three chaotic systems: the logistic map, the Henon map, and the 12-dimensional (12-D) Lorenz96 flow. MI is shown to outperform linear least squares in these systems when the nonlinearities are relevant; at small lead times and low noise levels MI and LS are comparable. The MI method does not solve the problem of nonlinear parameter estimation completely, but it does highlight the failure of common linear methods and allow significant progress in some nonlinear cases, progress which may generalize beyond the perfect model scenario.

II. TECHNICAL PROBLEM STATEMENT

Parameter estimation is a ubiquitous problem in scientific modeling [1,8,9,11–13,19,20]. While well understood in linear

systems [21–23], challenges remain in nonlinear systems [24]. Discussions of parameter estimation typically assume the following: dynamical systems are linear (or can be linearized), the mathematical structure of the model is perfect (thus “true” parameter values exist), and the statistics of observational uncertainty are known (the “noise model” is perfect). A more complete discussion is provided by Tarantola [24], who in Fig. 3.2 sketches six schematic examples, four that are linear or linearizable, one requiring fully nonlinear methods, and one too complex for his methods to be used. Problems of the fifth category in the context of prediction, the so-called “forward problem,” are approached here.

Assume the evolution of a system state $\mathbf{x}_i \in \mathbb{R}^m$ is governed by a finite dimensional, discrete time, deterministic nonlinear dynamical system:

$$\mathbf{x}_{i+1} = F(\mathbf{x}_i, \mathbf{a}), \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^m$ and the model’s parameters are contained in the vector $\mathbf{a} \in \mathbb{R}^l$. For $m = 1$, the state x_i is a scalar. For simplicity, forecasts are evaluated on a scalar observation below, even when $m > 1$. Assuming additive measurement noise δ_i yields observations $s_i = x_i + \delta_i$. A set of $l + 1$ sequential measurements $s_i, s_{i+1}, \dots, s_{i+l}$ would, in general, be sufficient to determine \mathbf{a} in a noise free setting (i.e., $\delta_i = 0 \forall i$) [8]. With noise, the task is somewhat harder.

Given model structure $F(x, \mathbf{a})$ and observations generated by a particular parameter \mathbf{a}_0 (the “true” parameter value), one can identify values for \mathbf{a} consistent with the available information. Parameter estimates are made on the basis of the skill of the probability forecast. To ease comparison with previous work [8,11], the approach is illustrated using three nonlinear models: the one-dimensional logistic map,

$$F(x, a) = 1 - ax^2, \quad (2)$$

the two-dimensional Henon map,

$$\begin{aligned} x_{i+1} &= 1 - ax_i^2 + y_i, \\ y_{i+1} &= bx_i, \end{aligned} \quad (3)$$

and a 12-D Lorenz96 flow [25].

III. MINIMUM IGNORANCE PARAMETER ESTIMATION

The least squares (LS) method estimates parameters by minimizing the root mean square error of a point forecast. Even given infinite data, the optimal LS solution is biased when applied to the logistic map [8]. The LS method fails because the assumption of independent normal distributed (IND) forecast errors does not hold, even with IND observational noise. This is to be expected in nonlinear models.

A point value based on an imperfectly observed initial state is incomplete as a forecast [26]; given observational uncertainty, an ensemble of initial states of the system consistent with given observations [27] is required to propagate this initial uncertainty, suggesting probabilistic forecasts via Monte Carlo ensembles.

A. Scoring probabilistic forecasts

A probabilistic skill score is a function $S(p(y), Y)$, where Y is the outcome and $p(y)$ is a probability forecast [28]. The

ignorance score [29,30] is given by

$$S[p(y), Y] = -\log_2[p(Y)]. \quad (4)$$

Ignorance is the only proper local score for continuous variables [31,32]. In practice, given N forecast-outcome pairs $(p_i(y), Y_i, i = 1, \dots, N)$, the empirical ignorance is

$$S_{\text{EI}}[p(y), Y] = \frac{1}{N} \sum_{i=1}^N -\log_2[p_i(Y_i)] - S_{\text{clim}}, \quad (5)$$

where S_{clim} is defined using the unconditional probability or “climatology” of y , denoted $p_c(y)$; this is simply the natural measure projected onto the forecast variable. The zero skill of ignorance is then

$$S_{\text{clim}} = \int -p_c(y) \log_2[p_c(y)] dy. \quad (6)$$

While other proper skill scores might be used in this context, ignorance is the only proper local skill score for continuous variables; it is invariant under smooth changes of coordinates. Ensuring these properties is desirable in parameter estimation.

B. Ensembles and probability forecasting

An ensemble forecast is based on a collection of simulations simultaneously. There are many methods for forming an ensemble of initial states [33–35]. Perhaps the simplest method is to add draws from the inverse of the observational noise to the observation to define ensemble members. In that case ensemble members are equally weighted, as each ensemble member is an independent draw. With this inverse noise method, the initial states are unlikely to be consistent with the long term model dynamics (e.g., they are not “on the attractor” should one exist).

Continuous forecast distributions can be produced from an ensemble by kernel dressing its members. Standard kernel dressing is used below (see [28,30] for more details and [36] for a Bayesian approach). Define an N_e member ensemble at time i to be $X_i = [x_i^1, \dots, x_i^{N_e}]$ and treat all ensemble members as exchangeable: the ensemble interpretation methods used do not depend on the ordering of the ensemble members [28]. Standard kernel dressing transforms the ensemble members into a probability density function p_m where

$$p_m(y : X, \kappa) = \frac{1}{N_e \kappa} \sum_{j=1}^{N_e} K\left(\frac{y - x^j}{\kappa}\right). \quad (7)$$

In this case the forecast distribution is a sum of Gaussian kernels $K(\cdot)$, the j th ensemble member being replaced by a kernel centered at x^j . For each value of \mathbf{a} , the kernel width, κ , is chosen to minimize the empirical ignorance defined in Eq. (5) above [28]. There remains the chance that the verification lies outside the range of any finite ensemble, even if the verification is selected from the same distribution as the ensemble itself; the probability of this happening is $\gtrsim \frac{2}{N_e}$. Given the nonlinearity of the model, these points may be very far from the ensemble and appear as “outliers” or “bad busts.”

Given a sample climatology of the system from historical data, probabilistic forecasts may be improved out of sample by blending the dressed ensemble with the sample climatology [28], thereby allowing both narrower kernels and fewer

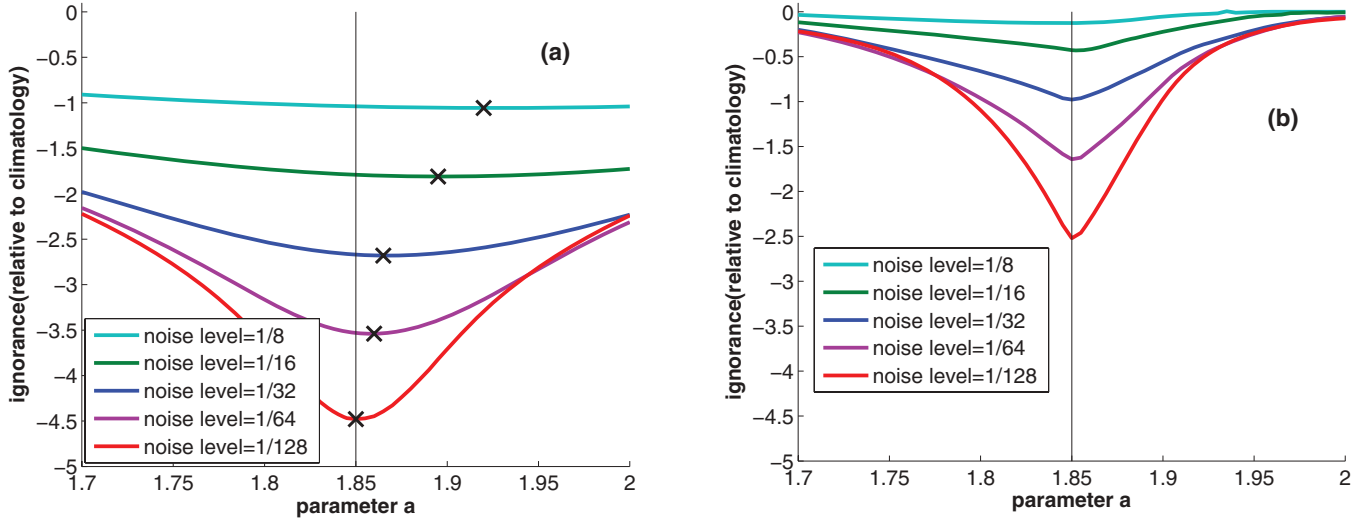


FIG. 1. (Color) Minimum ignorance parameter estimation for logistic map with $a = 1.85$; initial condition ensembles are formed by inverse noise. Five different noise levels are tested, each given 1024 forecasts. (a) Ignorance as a function of a for $\tau = 1$; the minima are marked with an “x.” (b) Ignorance as a function of a for $\tau = 4$.

bad busts. Blending with climatology yields the forecast distribution

$$p(y) = \alpha p_m(y) + (1 - \alpha)p_c(y), \quad (8)$$

where p_m reflects the ensemble and p_c the climatology. The probability forecast obtained will be a function of a . Values of a with small empirical ignorance are deemed better.

Comparing forecast performance of different models is not a *fair* comparison without blending climatology. It might be the case that without blending climatology model A outscores model B, while after blending climatology model B scores higher than model A. Since the sample climatology is available to any model, the comparison should include this information.

IV. EVALUATION AND RESULTS

Figures 1 and 2 show the empirical ignorance scores as a function of lead time τ and parameter value a for the logistic map and for the Henon map. Figure 1 shows five different noise levels σ , for two lead times. In panel (a) $\tau = 1$ and in panel (b)

$\tau = 4$. The vertical line marks the “true” parameter value of 1.85; Fig. 2 reports results from the Henon map showing how a MI approach outperforms a LS method. The “+” in each panel reflects the “true” values of a and b . Panel (a) shows the inferiority of the LS error.

Returning to Fig. 1(a), note the bias away from the “true” value. MI estimates at longer τ [Fig. 1(b)] tend to provide less biased estimates. The small τ bias is due to imperfections in the initial ensemble: neither the observation itself nor the initial ensemble formed by inverse noise are consistent with the long time dynamics. The natural measure of the logistic map is not uniform; for some parameter values it may be fractal. A dynamically consistent ensemble is an ensemble of initial conditions which are not only consistent with the observational noise but also consistent with the natural measure.² Figure 3

²Note that for a structurally perfect model the dynamically consistent ensemble will approach a perfect ensemble [27] at the “true” parameter value when a long window of observations is considered.

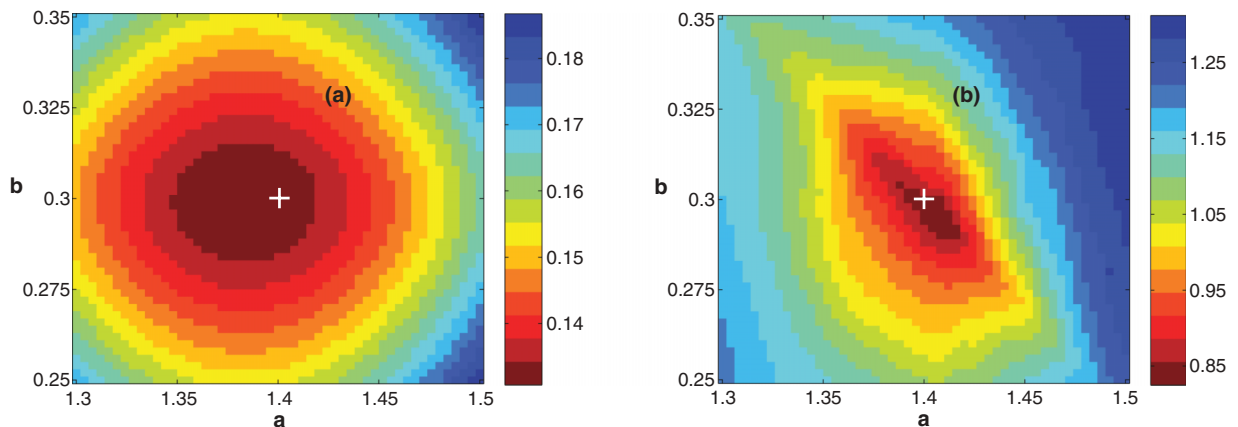


FIG. 2. (Color) Parameter estimation for Henon map ($a = 1.4$; $b = 0.3$), noise level equals 0.05, given 1024 forecasts at lead time 4. (a) Cost function based on LS. (b) Cost function based on forecast ignorance.

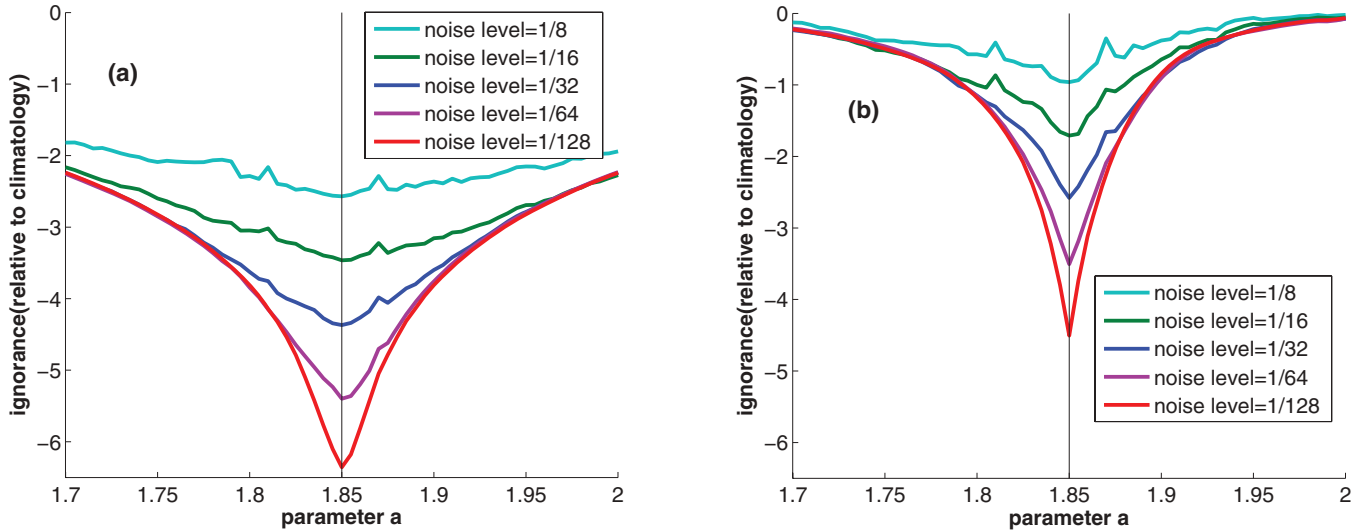


FIG. 3. (Color online) Parameter estimation for logistic map with $a = 1.85$ using dynamically consistent ensembles. Contrast the (improved) ignorance value relative to Fig. 1 where the same lead times and noise level are used.

shows that using a more dynamically consistent ensemble of initial conditions (in this case merely consistent with the current observation³) produces less biased results at short τ and also improves larger τ .

To demonstrate that the advantage of MI parameter estimation need not vanish in higher dimensional systems, the single parameter Lorenz96 system [25] is considered with $m = 12$ and the parameter $F = 17$, using inverse noise ensembles. Nonlinear effects are reduced at smaller lead times τ and lower noise levels σ . In Lorenz96, the estimation errors of LS and MI are roughly the same with $\tau = 0.5$ and $\sigma = 0.1$ (this is $\sim 0.2\%$ of the range of the data). Increasing the noise level to $\sigma = 1$, the estimation error from LS is approximately eight times that of MI. Alternatively keeping $\sigma = 0.1$ and increasing to $\tau = 1$ yields an error in the LS estimate approximately three times larger. For any smooth $F(x)$ the linear approximation will hold in the limit of infinitesimally small observational noise; even in this limit MI estimation will outperform linear methods which, like variants of the Kalman filter, fail to respect the natural measure of F .

V. IMPERFECT MODEL SCENARIO

In the statistics literature, parameters within the perfect model scenario, where a “true” value is thought to exist but is unknown, are sometimes referred to as “quantities with a well defined physical meaning” (see, for instance, [19]). Here the distinction is made between fitting parameters in a “physical model” and a “curve fitting model,” where in the second case parameters are defined only relative to some goal. As will be demonstrated below, if the mathematical structure of the model

is imperfect there is no unique value of the parameter that is “optimal”; the “best” parameter may vary with application (the lead time of the forecast, for example). Within the perfect model scenario the “true” or optimal parameter value exists but is unknown; outside the perfect model scenario this value is not unknown but undefined, and one is dealing not with uncertainty but with ambiguity [5].

All analysis techniques including LS are limited to exploring the information contained in the data; large forecast-outcome archives and lower observational noise levels contain more information and thus allow better parameter estimates when the model structure is perfect. When the model class does not admit an empirically adequate model, the notion of a “true” parameter value is lost. The MI approach remains useful for identifying the best parameter in an imperfect model if a notation of “best” is defined in terms of forecast performance.

Next consider a system-model pair in the imperfect model scenario. The Quartic system is defined as

$$\tilde{G}(\tilde{x}) = \tilde{a} \left[(1 - \tilde{\tau})\tilde{x}(1 - \tilde{x}) + \frac{4\tilde{\tau}}{5}\tilde{x}(1 - 2\tilde{x}^2 + \tilde{x}^3) \right]. \quad (9)$$

The model in this case is

$$G(x) = ax(1 - x), \quad (10)$$

which is just the logistic map in another form [37]. At $\tilde{\tau} = 0$ the model is perfect (as $\tilde{\tau} \rightarrow 0$ the model has structural error); $\tilde{\tau} = 0.1$ is considered here. Given the observations generated by the system with additive noise, the goal is to estimate the parameter of the (imperfect) model. Figure 4(a) shows the empirical ignorance scores as a function of parameter value for the logistic model at lead time 1; following Fig. 1 five different noise levels are examined. In Fig. 4(b) the noise level is fixed while five different lead times are examined. Note that the dashed black line reflects the parameter \tilde{a} used in the Quartic system which need no longer be the target of the model parameter value a . Results for three independent experiments are shown, indicating that the bias away from the

³Here the dynamically consistent ensemble of initial conditions is only consistent with the current observations; requiring consistency with a series of observations would result in more informative ensembles. To locate states also consistent with more past observations is much more costly.

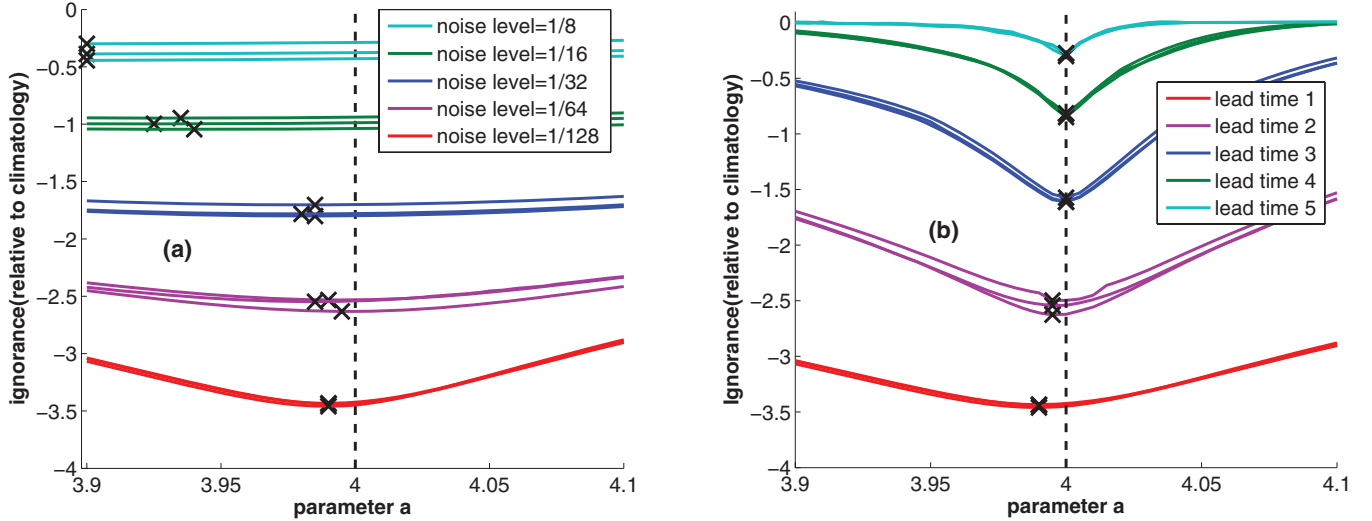


FIG. 4. (Color online) Parameter estimation for logistic model in the imperfect model scenario, with parameter $\tilde{a} = 4$ of the Quartic system, using inverse noise ensembles. Results from three independent realizations are shown, each given 1024 forecasts; note the consistency in locating the minimum (\times). The similarity of these three lines indicates the result is robust. (a) Empirical ignorance scores as a function of the parameter value for lead time 1 forecast at several noise levels. (b) Empirical ignorance scores as a function of the parameter value and lead time given at noise level 1/128.

system parameter value is robust. Figure 4 shows that the MI estimate varies with lead time and noise level. In both cases the notation of “best” is defined in terms of forecast skill given an inverse noise initial condition ensemble. The system parameter value \tilde{a} is not equal to the best model parameter value a .

VI. FURTHER DISCUSSION

Minimum ignorance parameter estimation considers the entire forecasting scenario; once the notion of “best” is defined any alteration of the forecasting scheme may alter the best parameter value. In this section, effects of ensemble formation and kernel dressing are discussed, and an alternative to “potential predictability” is suggested.

The variance of the standard kernel dressed ensemble is of course always larger than the variance of the raw ensemble, no matter how the kernel width is actually determined [38]. More complicated dressing methods exist (Brocker and Smith [28] for example introduced an improved kernel dressing, called “affine kernel dressing,” that is more flexible and robust). Standard kernel dressing is used here as it is straightforward to understand, easier to implement, and fit for our purpose. More advanced data assimilation methods may yield more informative ensembles (for example indistinguishable states [34] and Monte Carlo methods [13,33]). If it is costly to run the model (as with weather or climate models), inverse noise provides a much faster and cheaper first-pass estimate. There are also alternative low cost distributions one can use to blend with the dressed ensemble forecast other than the unconditional climatology, for example a dynamical climatology ensemble based on analogs to the current state (see the discussion of eRAP in [27]). The MI approach generalizes beyond estimating “physical” parameters as it can be used for structural parameters as, for example, in delay space reconstructions (see Farmer and Sidorowich [39] and citations

thereof) and model reduction [40]. Finally, note that it is also possible to estimate the parameters of the noise model(s) [12] within the MI framework.

“Potential predictability” reflects the utility an existing forecast system would have *if* it were perfect [41]. Interpreting this as utility carries some risk, of course, as the actual system may be much more predictable (or much less) than the dynamics of the current generation of models. An alternative approach which can quantify the (historical) impact of model inadequacy is to contrast the empirical ignorance with the implied ignorance, defined as

$$\int -p_m(y) \log_2[p_m(y)] dy. \quad (11)$$

The implied ignorance is the ignorance one would expect to observe if in fact the probability forecast was perfect. The difference between empirical ignorance and implied ignorance reveals an *information deficit* (in bits), which exposes shortcomings anywhere in the forecast methodology. In contrast with the so called “estimate” of skill from “potential predictability” experiments which assumes the model is perfect, the information deficit quantifies just how far the predictability of the current model is from (its internal) perfection. Within the perfect model scenario, the implied ignorance can approach the empirical ignorance for the “true” parameter values (if and only if the entire ensemble forecasting package is perfect). Even when the model structure is mathematically correct, the empirical ignorance may be greater than the implied ignorance, indicating that the model probability density function is an incomplete reflection of the expected uncertainties.² Indeed the information deficit provides quantitative information on second order uncertainty. Figure 5 illustrates this in both the perfect model case [Figs. 5(a) and 5(c)] and the imperfect model case [Figs. 5(b) and 5(d)]. In the perfect model case, the empirical ignorance and implied ignorance should converge

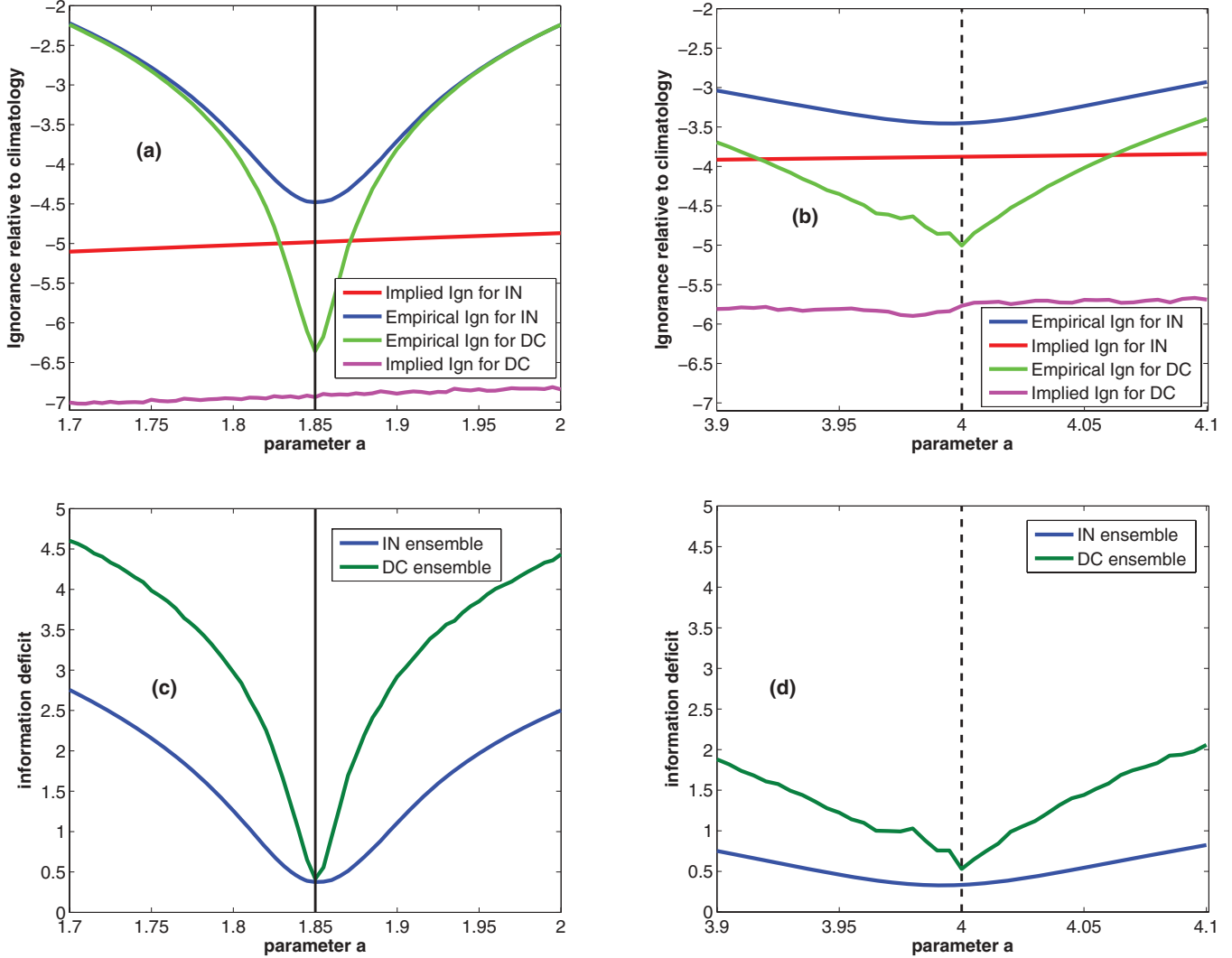


FIG. 5. (Color) Empirical ignorance and implied ignorance as a function of parameter value with noise level $\sigma = 1/128$ for lead time 1. Curves for both inverse noise ensemble and dynamically consistent ensemble. 1024 forecasts are considered in each case. (a) Perfect model scenario with the logistic map: $F(x, a) = 1 - ax^2$. (b) Imperfect model scenario with system-model pair of Eqs. (9) and (10). (c) Information deficit in the perfect model scenario. (d) Information deficit in the imperfect model scenario.

to within sampling error at the “true” parameter when a (many step) dynamically consistent ensemble is employed. In Figs. 5(a) and 5(b), the upper blue line shows the empirical ignorance for an inverse noise ensemble; the lower green line shows the empirical ignorance for a (one step) dynamically consistent ensemble. The upper red line and lower purple line correspond to the implied ignorance for each ensemble formation strategy, respectively. Figures 5(c) and 5(d) show that the information deficits correspond to Figs. 5(a) and 5(b). The information deficit will remain nonzero as long as inverse noise is used. In Figs. 5(a) and 5(c) the information deficit for dynamically consistent ensembles remains nonzero because these dynamically consistent ensembles are only consistent with one observation; as the window of dynamical consistency increases and the ensemble size increases the information deficit will approach zero. On the other hand, in the imperfect model case [Figs. 5(b) and 5(d)] the information deficit will remain nonzero no matter what one may do due to the model inadequacy.

Also note in Fig. 5 that the information deficit of the inverse noise ensemble is smaller than that of the dynamically consistent ensemble. This is somewhat misleading, in the same way that potential predictability is consistently misleading. Confusion can be avoided by noting that forecasts using the (one step) dynamically consistent ensemble provide almost two bits more information beyond those from the inverse noise ensembles.

VII. CONCLUSION

Although widely popular, the method of LS is optimal only in a narrow context, a fact stressed by Kalman [42]; LS is often applied well outside its mathematical remit. While a general account of parameter estimation remains lacking, the straightforward minimum ignorance approach introduced here is shown to yield good parameter estimation in several chaotic systems. Initial experiments suggest that the MI approach is

also useful for identifying the best parameter in an imperfect model as long as the notion of “best” is well defined.

MI is expected to perform well against the myriad of modern alternatives. The initial value approach [14] is reminiscent of four-dimensional variational assimilation [43] (4DVAR), minimizing the cost function not only for the initial condition but also for the parameter values; like 4DVAR it is computationally expensive and suffers from local minima. These variational methods differ from the LS method, inasmuch as LS fails fundamentally while the initial value approach fails numerically. Simply put, the root of this failure lies in chaotic likelihoods [44]. Voss [15] applied a multiple shooting method to address the local minima problem in the initial value approach, an initial value approach in short windows resembling a similar spin up procedure applied to 4DVAR [45]; the approach remains expensive and Voss’s examples show varying success. MI might be considered as a useful prefilter for the method of [15]; even then that method requires *ad hoc* continuity constraints.

Sequential (recursive) methods provide an alternative approach. Kalman filter methods are most often applied to state estimation; to estimate the parameter one may simply add the parameter vector to the state vector [15]. For weakly nonlinear systems the extended Kalman filter [18] can be used. For strong nonlinear systems, Voss [15] introduced the unscented Kalman filter. Notice that for such sequential methods the parameter vector is allowed to evolve in time; expectation maximization algorithms [16] account for this. Each of these methods perform better where the Gaussianity assumption holds more strongly. MI does *not* require any Gaussian constraints whatsoever.

MI parameter estimation also has the advantage that it is easy to use. Methods which contrast the natural measure of the model with the observations [8] are significantly more complicated and grow more so as the dimensionality

of the model increases. Alternative methods which contrast shadowing times of the model as a function of parameter values [11] are significantly more computationally expensive. MI estimation using inverse noise ensembles is straightforward to implement and relatively inexpensive computationally. It will fail to indicate the “true” parameter value when the ensemble is not distributed consistently with respect to the model’s long term dynamics (natural measure), but the parameter value MI suggests will give better probabilistic forecasts than the “true” parameter value *as long as* the flawed ensemble formation scheme is used. Investing more in data assimilation is shown to improve parameter estimates. When the mathematical structure of the model is incommensurate with the structure of the system generating the observations, the ultimate goal of parameter estimation is unclear. As illustrated above, the “optimal” parameter value may, for example, vary with lead time. In such cases, MI can still provide useful parameter estimation as long as the goal (“optimal”) is well defined.

MI parameter estimation by ensemble prediction provides a useful tool, avoiding the shortcomings of other approaches. The information deficit reflected in the implied ignorance can reveal forecast system inadequacies and quantify the predictability in a more informative manner than “potential predictability” does. We are optimistic that this framework will allow some progress outside the perfect model scenario.

ACKNOWLEDGMENTS

This research was supported both by the London School of Economics Grantham Research Institute and the Economic and Social Research Council Centre for Climate Change Economics and Policy, funded by the Economic and Social Research Council and Munich Re. L. A. S. gratefully acknowledges support from Pembroke College, Oxford.

-
- [1] V. F. Pisarenko and D. Sornette, *Phys. Rev. E* **69**, 036122 (2004).
 - [2] S. M. Stigler, *Ann. Stat.* **9**, 465 (1981).
 - [3] D. Teets and K. Whitehead, *Mathematics Magazine* **72**, 83 (1999).
 - [4] C. F. Gauss, *Theoria Motus Corporeum Coelestium*. English Translation by C. H. Davis (Little, Brown, Boston, 1857).
 - [5] L. A. Smith and N. Stern, *Phil. Trans. R. Soc. A* **369**, 1 (2011).
 - [6] L. A. Smith, *A Very Short Introduction to Chaos* (Oxford University Press, Oxford, 2007).
 - [7] L. A. Smith, *Proc. Nat. Acad. Sci.* **4**, 2487 (2002).
 - [8] P. E. McSharry and L. A. Smith, *Phys. Rev. Lett.* **83**, 4285 (1999).
 - [9] D. R. Creveling, P. E. Gill, and H. D. I. Abarbanel, *Phys. Lett. A* **372**, 2640 (2008).
 - [10] A. Maybhate and R. E. Amritkar, *Phys. Rev. E* **59**, 284 (1999).
 - [11] L. A. Smith, M. C. Cuellar, H. Du, and K. Judd, *Phys. Lett. A* **374**, 2618 (2010).
 - [12] J. Heald and J. Stark, *Phys. Rev. Lett.* **84**, 2366 (2000).
 - [13] J. C. Quinn and H. D. I. Abarbanel, *Quarterly Journal Royal Meteorological Society* **136**, 1855 (2010).
 - [14] K. Schittkowski, *Numer. Math.* **68**, 129 (1994).
 - [15] H. U. Voss, J. Timmer, and J. Kurths, *Int. J. Bifurcation and Chaos* **14**, 1905 (2004).
 - [16] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Journal Royal Statistical Society* **39**, 1C38 (1977).
 - [17] Z. Ghahramani and S. Roweis, *Learning Nonlinear Dynamical Systems Using an EM Algorithm, Advances in Neural Information Processing Systems 11* (MIT Press, Cambridge, 1999), pp. 431–437.
 - [18] M. Ghil and P. Malanotte-Rizzoli, *Adv. Geophys.* **33**, 141 (1991).
 - [19] A. Van Den Bos, *Parameter Estimation for Scientists and Engineers* (Wiley-Interscience, New York, 2007).
 - [20] C. Yao and E. M. Bollt, *Physica D* **227**, 78 (2007).
 - [21] A. Bjorck, *Numerical Methods for Least Squares Problems* (SIAM, Philadelphia, 1996).
 - [22] G. Casella and R. L. Berger, *Statistical Inference* (Wadsworth and Brooks/Cole, Pacific Grove, California, 1990).
 - [23] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem* (SIAM, Philadelphia, 1991).

- [24] A. Tarantola, *Inverse Problem Theory and Model Parameter Estimation* (SIAM, Philadelphia, 2004).
- [25] E. N. Lorenz, *Predictability: A Problem Partly Solved*. In Proc. Seminar on Predictability, Vol. 1 (ECMWF, Reading, UK, 1995).
- [26] E. N. Lorenz, *Tellus* **17**, 321 (1965).
- [27] L. A. Smith, in *Proc. International School of Physics "Enrico Fermi"*, Course CXXXIII (Societ'a Italiana di Fisica, Bologna, Italy 1997), p. 177.
- [28] J. Brocker and L. A. Smith, *Tellus A* **60**, 663 (2008).
- [29] I. J. Good, *J. R. Stat. Soc., Series B* **14**, 107 (1952).
- [30] M. S. Roulston and L. A. Smith, *Mon. Weather Rev.* **130**, 1653 (2002).
- [31] J. M. Bernardo, *Ann. Stat.* **7**, 686 (1979).
- [32] J. Brocker and L. A. Smith, *Weather Forecasting* **22**, 382 (2006).
- [33] G. Evensen, *J. Geophys. Res.* **99**, 10143 (1994).
- [34] K. Judd and L. A. Smith, *Physica D* **151**, 125 (2001).
- [35] M. Leutbecher and T. N. Palmer, *J. Comput. Phys.* **227**, 3515 (2008).
- [36] A. E. Raftery *et al.*, *Mon. Weather Rev.* **131**, 1155 (2005).
- [37] R. M. May, *Nature (London)* 261 (1976).
- [38] D. S. Wilks, *Meteorol. Appl.* **13**, 243 (2006).
- [39] J. D. Farmer and J. J. Sidorowich, *Phys. Rev. Lett.* **59**, 845 (1987).
- [40] E. M. Bollt, *I. J. Bifurcation Chaos* **17**, 1199 (2007).
- [41] *AMS (American Meteorological Society), Glossary of Meteorology*, edited by T. S. Glickman, 2nd ed. (American Meteorological Society, Boston, 2000).
- [42] R. E. Kalman, *Trans AMSE Ser. D J. Basi. Eng.* **82**, 35 (1960).
- [43] A. C. Lorenc, *Quart. J. Roy. Meteor. Soc.* **112**, 1177 (1986).
- [44] M. L. Berliner, *J. Am. Stat. Assoc.* **86**(416), 938 (1991).
- [45] C. Pires, R. Vautard, and O. Talagrand, *Tellus A* **48**, 96 (1996).