Parameter Estimation Through Ignorance

Hailiang Du[1] and Leonard A. Smith[1,2]

1 Centre for the Analysis of Time Series, London School of Economics, U.K.

2 Pembroke College, Oxford, U.K.

Abstract

Parameter estimation for nonlinear systems based on variations in the accuracy of probability forecasts is considered. Empirical results for the Logistic Map are presented at various noise levels and sampling rates. Selecting parameter values by minimizing Ignorance, a proper local skill score for continuous probability forecasts as a function of the parameter values is easier to implement in practice than alternative nonlinear methods. As expected, it is more effective when the forecast error distributions are non-Gaussian. Initial experiments suggest that our approach is also useful for identifying best parameter in an imperfect model as long as the notion of best is well defined. The information deficit, defined as the difference between the Empirical Ignorance and Implied Ignorance can be used to identify remaining forecast system inadequacy, in both perfect and imperfect model scenario.
TAKE HOME POINT: Probability forecasts provide better parameter estimation.

Technical problem statement

Assume the evolution of a system state $\mathbf{x}_i \in \mathbb{R}^m$ is governed by 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 simplicity forecasts are evaluated on a scalar observation below, even when $m > 1$. Assuming additive measurement noise δ_i yields observations $s_i = x_i + \delta_i$.
The least squares (LS) method estimates parameters by minimising 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 [1]. The LS method fails because the assumption of Independent Normal Distributed (IND) forecast errors does not hold, even with IND observational noise. Given model structure $F(x, \mathbf{a})$ and observations generated by 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.

Ignorance-Based parameter estimation

A point value based on an imperfectly observed initial state is incomplete as a forecast; given observational uncertainty, an ensemble of initial states of the system consistent with given observations is required to propagate this initial uncertainty, suggesting probabilistic forecasts via Monte Carlo ensembles. A probabilistic skill score is a function $S(p(y), Y)$, where Y is the outcome and $p(y)$ is a probability forecast. The Ignorance Score [2] is given by:

$$S(p(y), Y) = -log_2(p(Y)) \quad (2)$$

Ignorance is the only proper local score for continuous variables [3]. In practice, given N forecast-outcome pairs $(p_i(y), Y_i, i = 1, ..., N)$, the Empirical Ignorance is:

$$S_{EI}(p(y), Y) = \frac{1}{N}\sum_{i=1}^{N} -log_2(p_i(Y_i)) - S_{clim}, \quad (3)$$

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.

An ensemble forecast is based on a collection of simulations simultaneously. Perhaps the simplest method for forming an ensemble of initial states 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.With this Inverse Noise method, the initial states are unlikely to be consistent with the long term model dynamics.Continuous forecast distributions can be produced from an ensemble by kernel dressing its members. Standard kernel dressing is used below (see the appendix for more details).

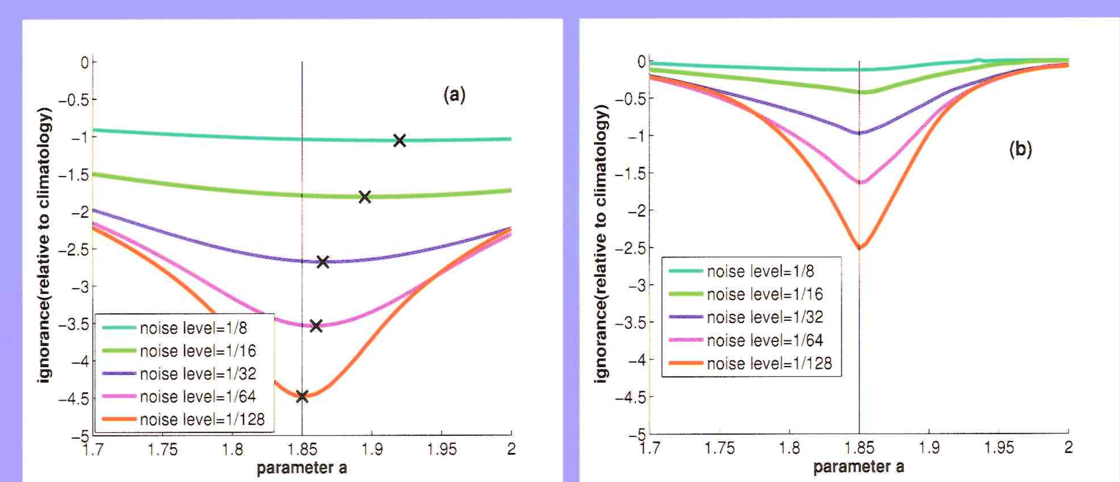

Fig 1: Ignorance-Based parameter estimation for Logistic Map with $\mathbf{a} = 1.85$; initial condition ensembles are formed by Inverse Noise. Five different noise levels are tested, each given 1024 forecasts; (a) Ignorance as a function of $\mathbf{a}$ for $\tau = 1$, the minima are marked with an "x"; (b) for $\tau = 4$.

Evaluation and Results

Figures 1 shows the Empirical Ignorance scores as a function of lead time, τ, and parameter value a for the Logistic Map, five different noise levels σ and two lead times are considered. In panel (a) $\tau = 1$ and panel (b) $\tau = 4$. The vertical line marks the true parameter value of 1.85. Note the bias away from the True value. IB estimates at longer τ (Figure 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. 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. Figure 2 shows using a more dynamically consistent ensemble of initial conditions (in this case merely consistent with the current observation) produces less biased results at short τ and also improves larger τ.

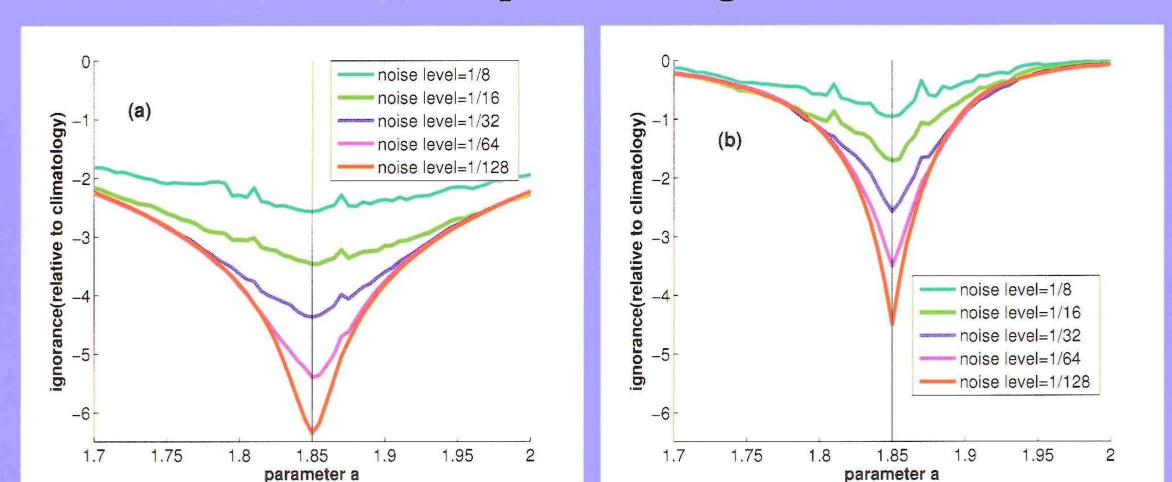

Fig 2: Parameter estimation for Logistic Map with $\mathbf{a} = 1.85$ using dynamically consistent ensembles. Contrast the (improved) Ignorance value relative to Figure 1 where the same lead times and noise level are used.

Imperfect Model Scenario

When the model class does not admit an empirically adequate model, the notion of a "True" parameter value is lost. The IB approach remains useful for identifying 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

$$\tilde{G}(x) = a((1-\iota)x(1-x) + \frac{4\iota}{5}x(1-2x^2+x^3)). \quad (4)$$

The model in this case is

$$G(x) = ax(1-x). \quad (5)$$

$\iota = 0.1$ is considered here. Figure 3(a) shows the Empirical Ignorance scores as a function of parameter value for Logistic model at lead time 1. Figure 3(b) the noise level is fixed while 5 different lead times are examined. Note the dashed black line reflects the true parameter value used in the Quartic system only therefore is no longer the target of the estimate. Results for three independent experiments are shown, indicating that the bias away from the system parameter value is robust. The notation of "best" here is defined in terms of forecast skill given an Inverse Noise initial condition ensemble.

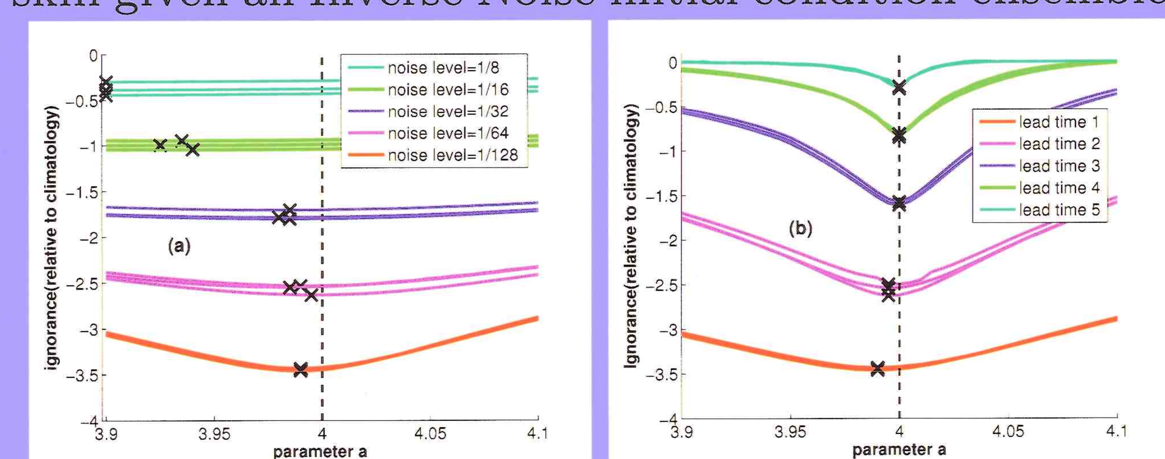

Fig 3: Parameter estimation for Logistic model in the Imperfect Model Scenario.Results from three independent realizations are shown; note consistency in locating the minimum (×). a) Empirical Ignorance scores as a function of the parameter value for lead time 1 forecast at several noise level; b) at 5 different lead times given Noise Level= 1/128.

Implied Ignorance

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

$$\int -p_m(y)log_2(p_m(y))dy \quad (6)$$

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 corresponds to an *information deficit* (in bits), and will expose shortcomings anywhere in the forecast methodology.

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 PDF is an incomplete reflection of the expected uncertainties. In the perfect model case (Figure 4(a) and (c)), the Empirical Ignorance and Implied Ignorance should converge to within sampling error at the True parameter when a (many step) dynamically consistent ensemble is employed. On the other hand, in the imperfect model case (Figure 4(b) and (d)) the information deficit will remain nonzero no matter what one may do due to the model inadequacy.

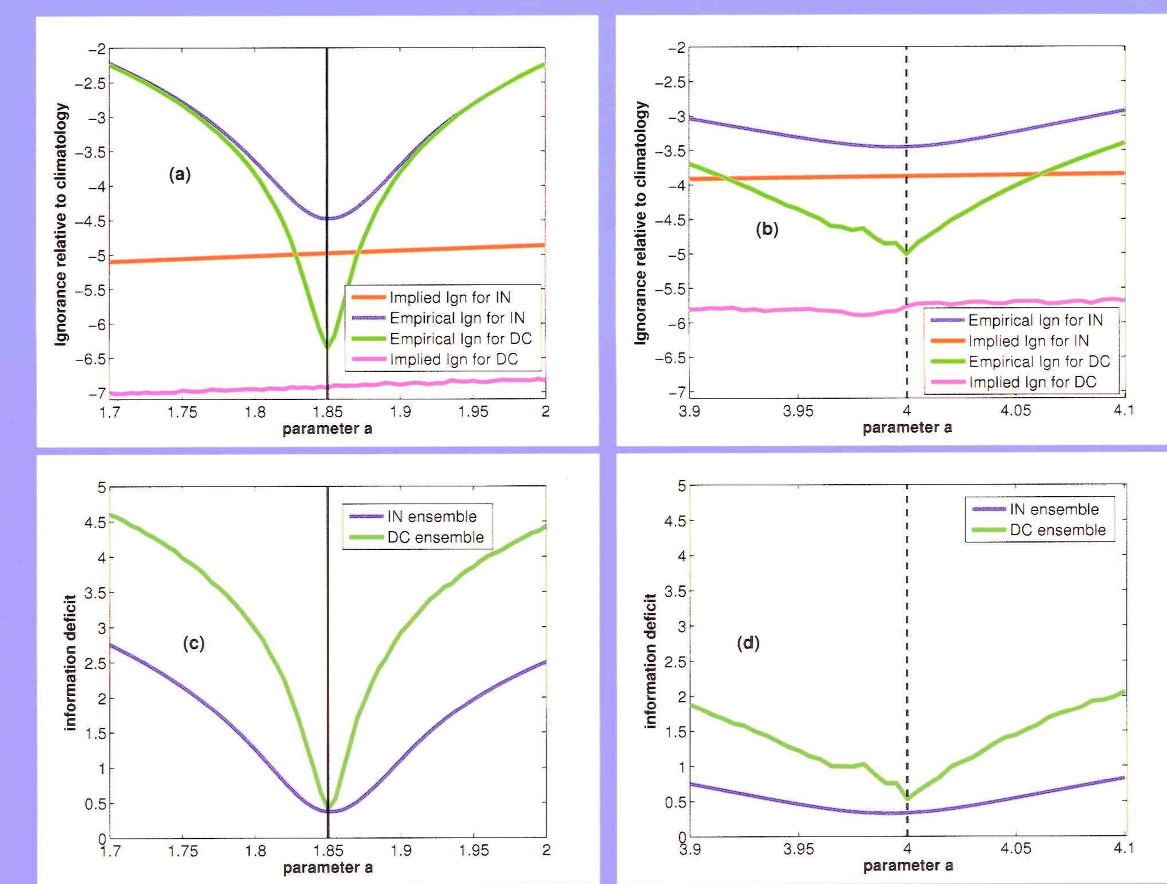

Fig 4: Empirical Ignorance and Implied Ignorance as a function of parameter value with noise level $\sigma = 1/128$ for lead time 1. (a) Perfect Model Scenario (b) Imperfect Model Scenario, (c) information deficit in the Perfect Model Scenario (d) Imperfect Model Scenario.

Higher dimensional case

For higher dimensional systems, the single parameter Lorenz96 system [?] 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. The estimation error of LS and IB are roughly the same with $\tau = 0.5$ and $\sigma = 0.1$. Increasing the noise level to $\sigma = 1$, the estimation error from LS is ~ 8 times that of IB. Alternatively keeping $\sigma = 0.1$ and increasing to $\tau = 1$ yields an error in the LS estimate ~ 3 times larger.

Summary

IB parameter estimation by ensemble prediction provides a useful tool, avoiding the shortcomings of other approaches. The information deficit revealed by the Implied Ignorance can reveal forecast system inadequacies and quantify the predictability in a more informative manner than "potential predictability" does.

Acknowledgements

This research was supported both by the the LSE's Grantham Institute and the ESRC Centre for Climate Change Economics and Policy, funded by the Economic and Social Research Council and Munich Re.

References

[1] P.E. McSharry and L.A. Smith, Better nonlinear models from noisy data: Attractors with maximum likelihood. *Phys. Rev. Lett.* **83**, 4285, (1999).

[2] D.J. Good, Rational decisions. *Journal of the Royal Statistical Society*, XIV(1) 107 (1952).

[3] J. Brocker, L.A Smith, Scoring Probabilistic Forecasts, *Weather and Forecasting*, 22, 382, (2006).