School of Mathematical
and Physical Sciences

Department of Mathematics and Statistics

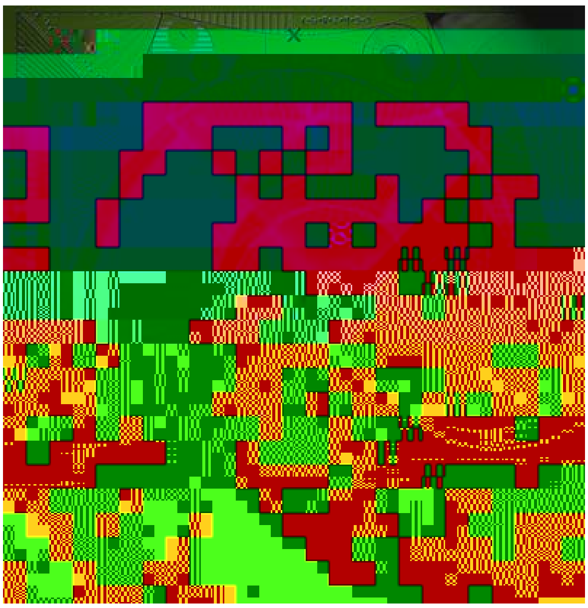
Preprint MPS_2010-28

10 September 2010

Initial Distribution Spread: A density forecasting approach

by

R.L. Machete and I.M. Moroz



Initial Distribution Spread: A density forecasting approach

R. L. Machete [1] and I. M. Moroz [a]

[a]

Ideally, the initial ensemble should be drawn from the underlying invariant measure, in which case we have a **perfect initial ensemble**. A perfect initial ensemble is especially useful in the scenario when our forecasting model is isomorphic to the model that generated the data, which scenario is called the **perfect model scenario** (PMS) [2, 5]. When there is no isomorphism between the forecasting model and the model that generated the data, then we are in the **imperfect model scenario** (IMS). A perfect model with a perfect initial ensemble would give us a **perfect forecast** [6]. If either our model or initial ensemble is not perfect, then we have no reason to expect perfect forecasts.

In all realistic situations, we have neither a perfect model nor a perfect initial ensemble, yet we may be required to issue a meaningful forecast probability density function (pdf). Roulston and Smith [7] proposed a methodology for making forecast distributions that are consistent with historical observations from ensembles. This is necessary because the forecast ensembles are not drawn from the underlying invariant measure due to either imperfect initial ensembles or model error. Their methodology was extended by Broecker and Smith [8] to employ continuous density estimation techniques [9, 10] and blend the ensemble pdfs with the empirical distribution of historical data, which is referred to as **climatology**. The resulting pdf is what will be taken as the forecast pdf in this paper.

The quality of the forecast pdfs can be assessed using the logarithmic scoring rule proposed by Good [11] and termed **ignorance** by Roulston and Smith [12], borrowed from information theory [13, 14]. Here, we discuss a way of choosing the initial distribution spread (IDS) to enhance the quality of the forecast pdfs. The point is that if the spread is too small our forecasts may be over confident and if it is too large our forecasts may have low information content. Our goal is to choose an IDS that yields the most informative forecast pdfs and determine, for instance, if this varies with the lead time of interest. As is commonly done in data assimilation and ensemble forecasting (e. g. see [1, 15]), we only consider Gaussian initial distributions. In traditional data assimilation and ensemble forecasting techniques, estimation of the initial distribution is divorced from forecasting: this is the main point of departure in our approach. We revisit this later in the discussion of the results in § 5.

Our numerical forecasting experiments were performed on the Moore-Spiegel (M-S) [16] system and an electronic circuit motivated by the M-S system. Indeed electronic circuits have been studied to enhance our understanding of chaotic systems and Chua circuits [17] are among famous examples. Recently, Gorlov and Strogonov [18] applied ARIMA models to forecast the time to failure of Integrated Circuits. Hence, electronic circuits have not only been studied to enhance our understanding of chaotic systems and the forecasting of real systems, but also to understand the circuits themselves and to address practical design questions.

This paper is organised as follows: § 2 introduces the technical framewdir53heon

2 Forecasting

Consider a deterministic dynamical system,

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}(t)\ \boldsymbol{\lambda}) \tag{1}$$

with the initial condition $\boldsymbol{x}(0) = \boldsymbol{x}_0$, where $\boldsymbol{x}\ \boldsymbol{F} \in \mathbb{R}$, $\boldsymbol{\lambda} \in \mathbb{R}^d$ is a vector of parameters, $\boldsymbol{F}$ is a Lipschitz continuous (in $\boldsymbol{x}$), nonlinear vector field and t is time. By Picard's theorem [19], (1) will have a unique solution, say $\boldsymbol{\varphi}\ (\boldsymbol{x}_0; \boldsymbol{\lambda})$. If $\nabla.\boldsymbol{F}\quad 0$, this system might have an attractor [20], which if it exists we denote by A. In particular, we are interested in the case when the flow on this attractor is chaotic.

2.1 Forecast Density

For any point in state space, $\boldsymbol{x}$, and positive real number , let $B_{\boldsymbol{x}}(\ )$ denote an -ball centred at $\boldsymbol{x}$. Suppose that is some invariant measure (see appendix A) associated with the attractor A. For any $\boldsymbol{x}_0 \in A$, we define a new probability measure associated with $B_{\boldsymbol{x}_0}(\ )$ by

$$_0(E) = \lim_{T \to \infty} \frac{1}{T\ (B_{\boldsymbol{x}_0}(\ ))} \int_0^T \mathbf{1}_{\ \cap B_{\boldsymbol{x}_0}(\ )}(\boldsymbol{x}(t))\mathrm{d}t \tag{2}$$

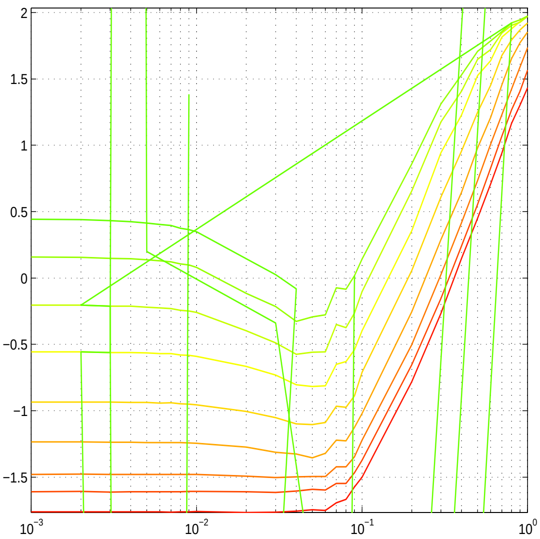
where $\mathbf{1}$ is an indicator function. This measure induces some probability density function, $p_0(\boldsymbol{x}\ \boldsymbol{x}_0\quad)$. We will call a set of points drawn from p_0

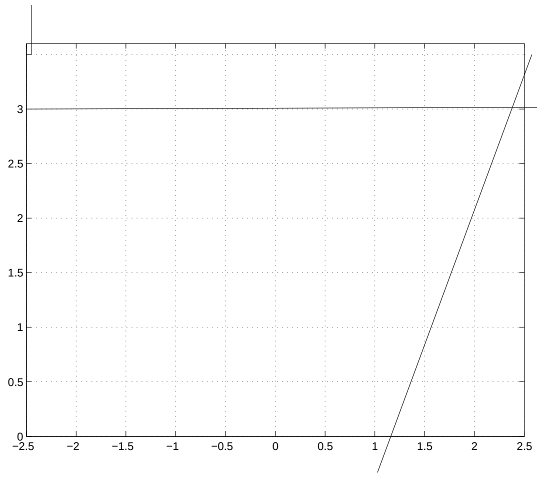
where $\boldsymbol{X}^{( )}$ is the random variable being forecast from the initial distribution corresponding to $\boldsymbol{x}$. Provided the underlying attractor is ergodic, we can rewrite (4) as

$$\mathbb{E}[S(t)] = \lim_{T \to \infty} \frac{1}{T} \int_0^T S(f (\boldsymbol{x}; \boldsymbol{x}) \ \boldsymbol{X}^{( )}) \mathrm{d}\tau. \tag{5}$$

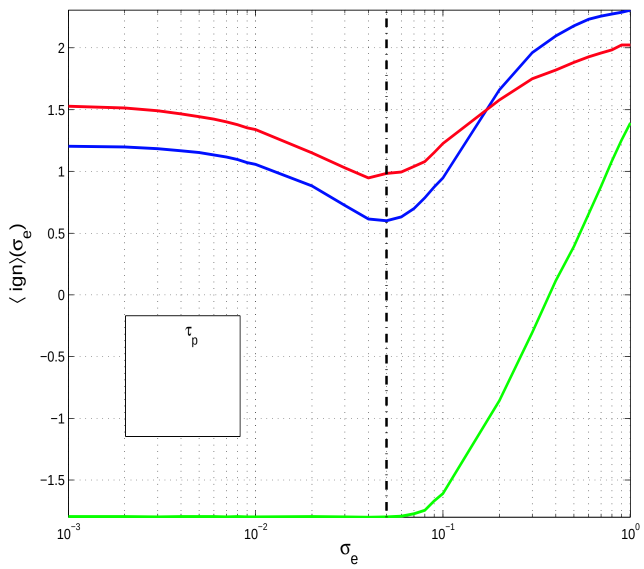
For each forecast, the underlying system can only furnish one verification of $\boldsymbol{X}$ and not the distribution $p (\boldsymbol{x}; \boldsymbol{x}_0)$. Therefore, we use (5) to score forecasts rather than (4). Discretise time according to $\tau = (i-1)\tau$, for $i = 1 \ 2 \ .. \ N$, where τ is the sampling time. This gives a sequence of forecast pdfs, $\{f (\boldsymbol{x}; \boldsymbol{x})\}_{=1}^{N}$, corresponding to verifications $\{\boldsymbol{X}^{( )}\}_{=1}^{N}$ and score S. We can thus discretise (5) to obtain the following empirical score to value the t-ahead forecast system:

$$\langle S \rangle (t) = \frac{1}{N} \sum_{=1}^{N} S(f (\boldsymbol{x}; \boldsymbol{x}) \ \boldsymbol{X}^{( )}$$

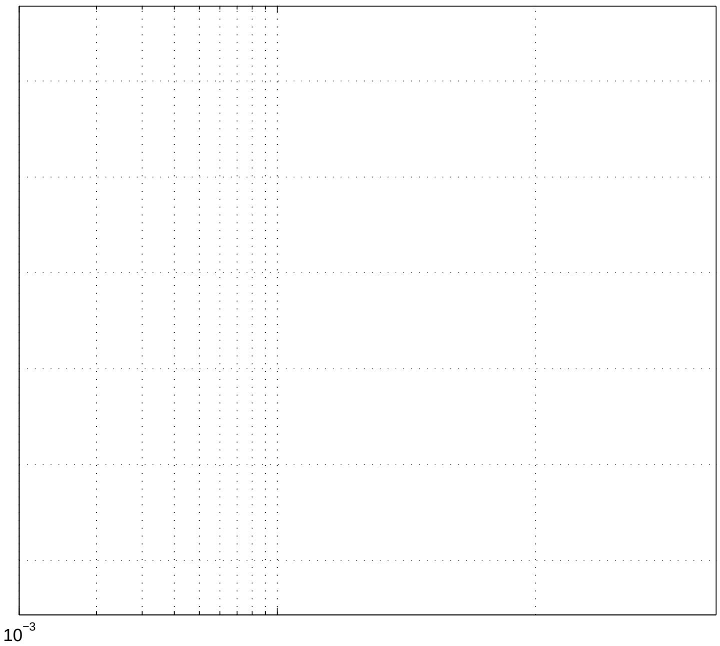

2
1.5
1
0.5
0
−0.5
−1
−1.5
10^{-3}
10^{-2}
10^{-1}
10^{0}

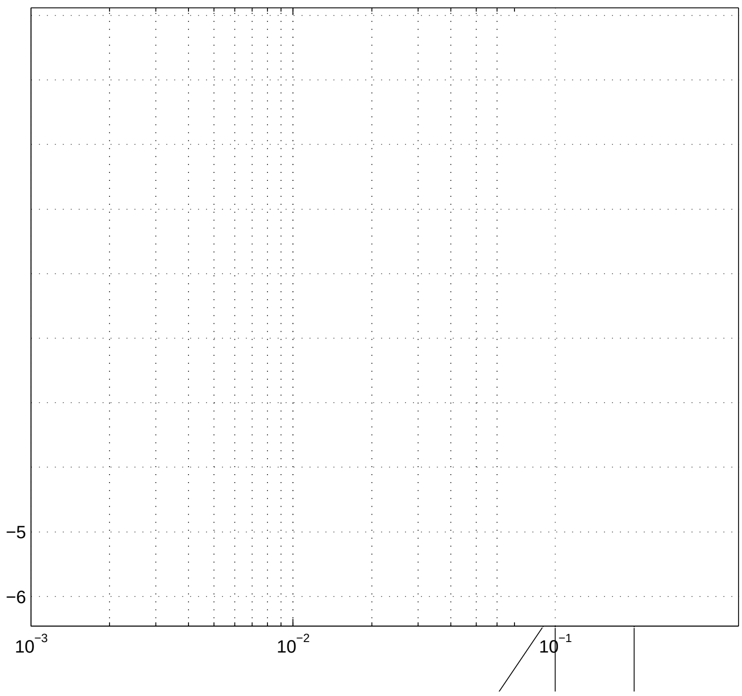


3
2.5
2
1.5
1
0.5
0
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5

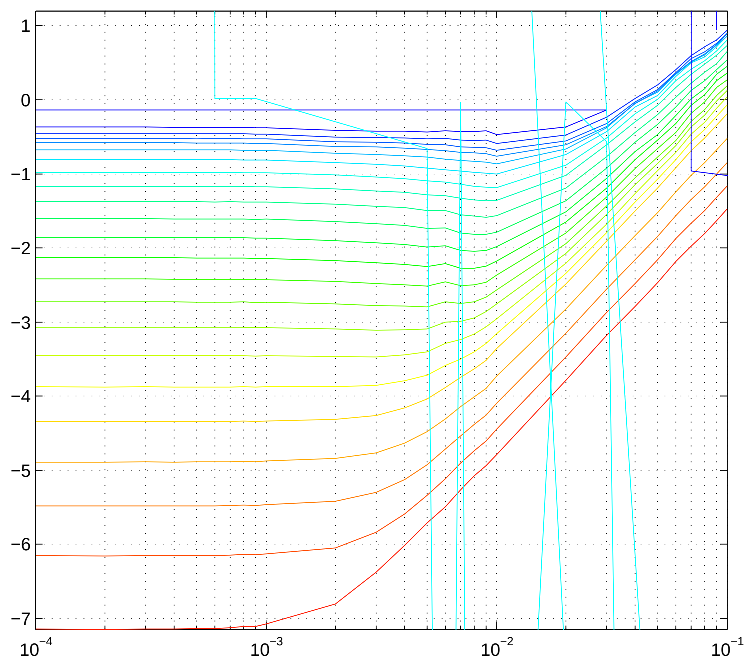


2
1.5
1
0.5
0
−0.5
−1
−1.5
⟨ ign⟩(σ_e)
τ_p
10^-3
10^-2
10^-1
10^0
σ_e


10−3



−5
−6
10^{-3}
10^{-2}
10^{-1}



1
0
−1
−2
−3
−4
−5
−6
−7
10^{-4}
10^{-3}
10^{-2}
10^{-1}

the method presented here could be used to determine the spread of this distribution, regardless of the data assimilation technique. For a given structure of the correlation matrix, we would seek the scalar multiple that yields the most informative forecast distributions.

Other techniques for producing the initial ensemble aim at selectively sampling those points that are dynamically the most relevant. In particular, the ECMWF ensemble prediction system seeks perturbations of the initial state based on the leading singular

[15] J. Broecker, U. Parlitz, M. Ogorzalek, Nonlinear noise reduction, Proc. of the IEEE 90.

[16] W. D. Moore, E. A. Spiegel, A thermally excited nonlinear oscillator, The Astrophysical Journal 143 (1966) 871–887.

[17] L. O. Chua, Experimental chaos synchronisation in Chua's circuit, International Journal of Bifurcation and Chaos 2 (1992) 704.

[18] M. I. Gorlov, A. V. Strogonov, ARIMA Models Used to Predict the Time to Degradation Failure of TTL IC's, Russian Microelectronics 36 (2007) 261–270.

[19] E. A. Coddington, N. Levinson, Theory of Ordinary Di erential Equations, New York:McGraw-Hill, 1955.

[20] E. Ott, T. Sauer, J. A. Yorke (Eds.), Coping With Chaos: Analysis of chaotic data and the exploitation of chaotic systems, John Wiley and Sons Inc., 1994.

[21] J. Broecker, L. A. Smith, Scoring Probabilistic Forecasts: The importance of being
4234(v)44122(r5.558(()3.887o)-38.04773[T(e)]4Ff4234(v)44122(r356.)-3e e] Ff c