Observing and Predicting Chaotic Signals: Is 2% Noise Too Much?

Thomas Schreiber and Holger Kantz

Abstract

We discuss the influence of noise on the analysis of complex time series data. How harmful it is depends on the nature of the noise, the complexity of the signal and on the application in mind. We will give generally valid upper bounds on the feasible noise level for dimension, entropy and Lyapunov estimates and lower bounds for the optimal achievable prediction error. We illustrate in a number of examples why it is hard to reach these bounds in practice. We briefly sketch methods to detect, analyze and reduce measurement noise.

1 Introduction

All experimental data are to some extent contaminated by noise. That this is an undesirable feature is commonplace (by definition, noise is the unwanted part of the data.) But how bad is noise really? The answer is as usual: it depends. The nature of the system emitting the signal and the nature of the noise, determine if the noise can be separated from the clean signal at least to some extent. This done, the amount of noise introduces limits on how well a given analyzing task (prediction, etc.) can be carried out.

In order to focus the discussion in this chapter on the influence of noise, we will assume throughout that the data are otherwise largely well-behaved. By this we mean that the signal would be to some extent predictable by exploiting an underlying deterministic rule — were it not for the noise. This is the case for data sets which can be embedded in a low dimensional phase space, which are stationary and which are not too short. Violation of one of these requirements leads to further complications which will not be addressed here.

The signals we will consider in the following are typically observations of dynamical phenomena. By this we mean that they are based on a (generally unknown) low dimensional deterministic system. Thus the evolution can be expressed either as a set of ordinary differential equations or as a discrete time mapping. Although most physical systems evolve continuously in time, data are always sampled discretely. Since, furthermore, the results we will present assume a simpler form, without loss of generality, we will concentrate
on maps in what follows. Thus, a trajectory of the underlying system can be obtained by iterating

$$y_{n+1} = F(y_n),$$

starting from some initial condition $y_0$. Reality deviates from this description in several respects. First, we usually cannot obtain $y_n$ directly, instead we measure some function $g(y)$ of it, in most cases yielding only scalar values. Second, this measurement is always subject to some measurement error, be it due to random fluctuations or due to the discretization. Finally, at each moment the state $y_n$ of the system may be perturbed by a stochastic process.

For all kinds of data processing described below we have to reconstruct a higher dimensional trajectory $\{x_n\}$ which is in some sense equivalent to $\{y_n\}$, knowing only noisy scalar measurements $\{x_n\}$. Without noise, this is successfully achieved using a time-delay embedding [Sauer, Yorke, and Casdagli, 1991]. In the presence of noise state space reconstruction becomes a very intricate problem. However, we will cut the discussion short by assuming that a delay embedding of sufficiently high dimension forms a proper reconstruction and our data yield trajectories deterministic up to the noise. For the reconstruction problem we refer the reader to the literature, in particular to a paper by [Casdagli et al., 1991]. A review of nonlinear time series analysis can be found in [Grassberger, Schreiber, and Schaffrath, 1991].

2 Measurement Error and Dynamical Noise

Measurement noise refers to the corruption of observations by errors which are independent of the dynamics. The dynamics satisfies $y_i = f(y_{i-1})$, but we measure scalars $x_i = g(y_i) + \eta_i$, where $g$ is a smooth function that maps points on the attractor to real numbers, and the $\eta_i$ are independent and identically distributed (IID) random variables. (Even in multi-channel measurements, generally not $y_n + \eta_n$ is recorded, but different scalar variables corresponding to different measurement functions $g_j$.)

Dynamical noise, in contrast, is a feedback process wherein the system is perturbed by a small random amount at each time step:

$$x_i = f(x_{i-1} + \eta_{i-1}).$$

Dynamical and measurement noise are two notions of the error that may not be distinguishable a posteriori based on the data only. Both descriptions can be consistent to some extent with the same signal. \(^9\)

\(^9\) For strongly chaotic systems which are everywhere expanding (more precisely, for Axiom A systems), measurement and dynamical noise can be mapped onto each other [Eckmann and Ruelle, 1986].