Will it rain tomorrow? Improving probabilistic forecasts.

E. Wheatcroft[1] and L. A. Smith[1,2]

[1]Centre for the Analysis of Time Series, London School of Economics
[2]Pembroke College, Oxford University, U.K.
Email: *e.d.wheatcroft@lse.ac.uk*

1 Introduction

Chaos is the phenomenen of small differences in the initial conditions of a process causing large differences later in time, often colloquially referred to as the 'butterfly effect'. Perhaps the most well known example though is in meteorology where small differences in the current conditions can have large effects later on. This effect is famously summed up by the notion that 'when a butterfly flutters its wings in one part of the world, it can eventually cause a hurricane in another'. Of course this is only a fictional example but let's suppose that we know this is true but we don't know whether the butterfly has flapped its wings or not. Do we accept that we can't predict what's going to happen? Or can we gain some insight? Now suppose we know from experience that the probability of the butterfly flapping its wings is 0.05, i.e. 5 percent. With this information we might conclude that the probability of a hurricane occuring is 0.05 also. This is of course an oversimplified and unrealistic example, but it illustrates the concept of *ensemble forecasting* in that a degree of belief about uncertainty of the initial conditions can give us a better idea of the probability of a future event.

Figure 1: a rainy day?

2 Ensemble Forecasting

Typically, weather forecasts work by feeding some initial conditions into a model that attempts to simulate the natural processes involved. However, since weather has chaotic properties, if we don't know the initial conditions precisely, our forecast could end up predicting an outcome that is a long way from reality. Suppose a weather model needs, among other things, the total rainfall for a particular day in a particular area. Can we rely on the accuracy of instruments that measure such things? If it is a particularly windy day, will our rain gauge definately collect the exact same amount of water as with the same amount of rain on a still day? The answer is no, measuring instruments can only take measurements to finite precision. This small error could cause our forecast to be very inaccurate. So if we know that we can only observe current conditions approximately and our models are chaotic what can we do? Fortunately, we have an approach that can deal with this problem. Suppose we run our model many times with slightly different initial conditions, each resulting in a different simulation of the weather. If the variety of initial conditions reflects the uncertainty in the measurements, this should give us an idea of what *could* happen. This forms the basis of *ensemble forecasting*. Figure 2 is an example of an ensemble forecast. Each green line represents an *ensemble member* and the black line represents the 'truth'. Note how the ensemble stays very close to the truth for a long time, but starts to diverge very quickly later on.

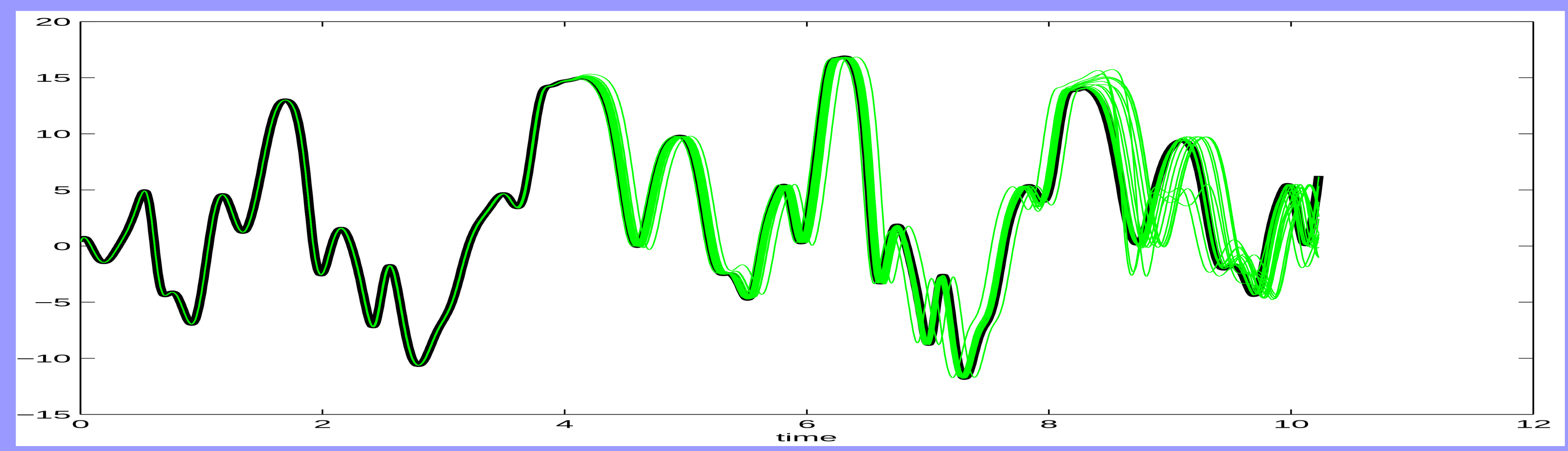

Figure 2: Evolution of an ensemble forecast (green) of the black line over time in a simple theoretical example

3 From Ensembles to Probability Distributions

Suppose we have an ensemble of forecasts of the amount of precipitation for tomorrow. On its own, this information is of limited use. What decision makers often require is a probability distribution so that they can get an idea about the likelihood of an event. For example, a farmer might be interested in the likelihood of a certain volume of rain so that they can choose the right time to plant their seeds. To get a probability distribution from an ensemble forecast, we use a well known method in statistics called *kernel density estimation*. The principle here is that we replace each ensemble member with a bell shaped distribution called a kernel which can overlap with those of other ensemble members. Summing the height of all of the curves at each point produces an overall probability distribution which describes the data. We can alter the width of the kernels which will change the shape of the overall distribution. Figure 3 illustrates how this works. The plot on the left represents some data as a histogram and the plot on the right shows the individual kernels (red dotted) which are summed to produce the probability density (blue).

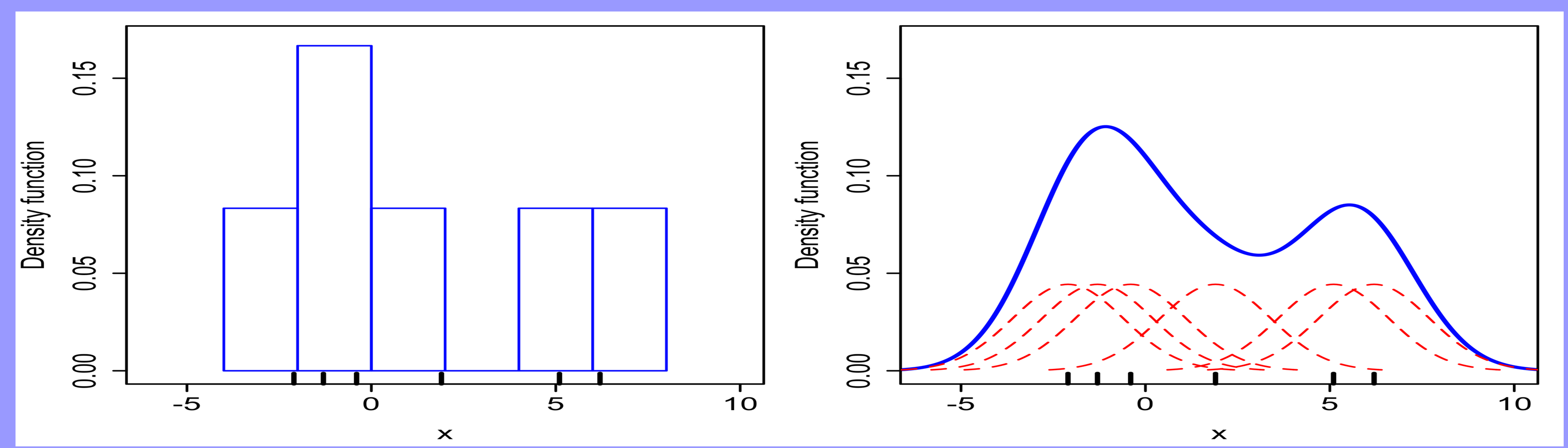

Figure 3: An illustration of kernel density estimation

4 Choosing the kernel width

Kernel density estimation requires a parameter that decides the width of the kernels. This parameter is very important, a poor choice of kernel width will result in a distribution that tells us very little about what we're trying to predict. An approach to doing this is to use previous experience of ensembles and to choose a kernel width that optimises a skill score that tells us the quality of our forecast (having compared it with what really happened). We can then use this width in our future probability distributions. In 1952 I.J Good devised a score that can be used to describe the difference in value between forecasts, we call this the *ignorance* and it is given by $ign = -log_2(p)$, where p is the height of the distribution at the verification (what actually happened).

This skill score rewards a high probability placed on the verification, a lower ignorance indicating a better forecast. This research is concerned with exactly how we choose this kernel width. We consider 4 different ways of doing this:

1. Method 1 - A constant kernel width for all ensembles optimised over previous forecasts.
2. Method 2 - A kernel width that depends on the standard deviation (the spread) of the ensemble in the form $\alpha + \beta\sigma_{ens}$ where σ_{ens} is the standard deviation of the ensemble. The α and β that optimise the ignorance are found from previous forecasts and verifications.
3. Method 3 - A kernel width that depends on the mean nearest neighbour in the ensemble, that is the average distance each ensemble member is away from the ensemble member that is closest to it. This is of the form $\alpha + \beta\phi_{ens}$ where ϕ_{ens} where ϕ_{ens} is the mean nearest neighbour in the ensemble. Again, we find α and β from previous forecasts and verifications.
4. Method 4 - Optimal method for bell shaped data that doesn't rely on previous experience.

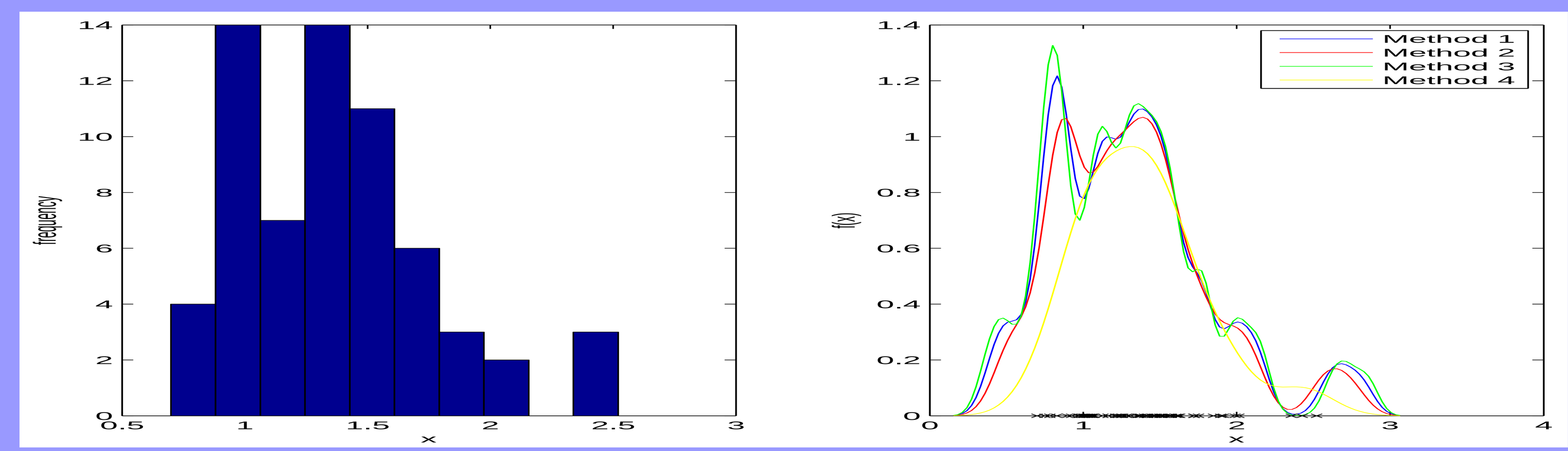

Figure 4: Kernel density estimates of an ensemble with kernel widths chosen in different ways.

4.1 Which method is best?

How do we decide which method is best? Ideally, we want our kernel width to forecast as high a probability as possible on the event that actually occurs. To test this, we generate a trajectory which we assume to be truth, then make an ensemble forecast and see how well each method does. We do this a total of 2048 times for 2 different systems known as the Lorenz '96 system and the Moore-Spiegel system. The table below gives the mean ignorance score for each method (recall, lower is better), in both cases, method 2 does better than method 1. Method 3 also does better than method 1 but not as well as method 2. Method 4 does poorly for Lorenz '96 and better for Moore-Spiegel, but it is unreliable since it is optimal for bell shaped data, whereas ensembles tend to be more irregularly shaped.

Method	Lorenz '96	Moore Speigel
1	0.1625	-2.2255
2	0.1294	-2.6975
3	0.1562	-2.5820
4	0.1928	-2.6461

5 Conclusion

Ensemble forecasting can tell us a lot about the likelihood of future events in chaotic systems. Kernel density estimation can convert an ensemble into a probabilty forecast. Choosing a kernel width that depends on the individual attributes of each forecast can improve estimation and these methods can be applied to real life chaotic systems such as weather forecasts.

Acknowledgements: The support of NERC is gratefully acknowledged.