Information Theory and Predictability
Lecture 11: Information flow

1 Introduction

Within a general random dynamical system uncertainty can be thought of as flowing from one location, dimension or subspace to another. A good example of such a flow is provided by a stochastically forced system where the random forcing is constantly injecting uncertainty. The analysis of this flow is of significant interest in many practical applications. A particular application is provided by prediction where uncertainty flows from location to another as the prediction proceeds in time. It is clear if one was able to reduce uncertainty in one particular location then benefits would accrue in the locations to which the uncertainty flowed. There are many other possible applications of this study some of which are detailed in [1] and [9].

Information or uncertainty flow does not in general satisfy the usual intuitive notions derived from fluid flow. Thus for example we shall see below that uncertainty flow is not necessarily symmetric.

A more rigorous formulation of uncertainty flow is presently in the process of detailed development and we shall outline in this Lecture several approaches that have been proposed in the literature. Earlier work tended to be empirical in the sense that flow functionals were proposed and then justified intuitively by appeal to the well understood properties. Later work has focused on more fundamental formulations.

2 Empirical approaches

There have been two functionals proposed in the physics and atmospheric science literature (see [1], [9], [3] and[8]):

Suppose we have available[1] the joint probability distribution function $p(x(t), y(t_0))$ of a particular prediction (random) variable $X(t)$ and another variable $Y(t_0)$ at the initial condition time t_0. Now if we had perfect knowledge of $Y(t_0)$ then the (univariate) distribution for $X(t)$ would change to reflect this improved knowledge of our system. The resulting univariate distribution is the conditional distribution[2]

$$p(x(t)|y(t)) \equiv \frac{p(x(t), y(t_0))}{p(y(t_0))}$$

[1] In a practical situation such a distribution will only be available to us as a sample estimate at reasonably coarse resolution however for the present we shall ignore this technical difficulty. Later we shall discuss the practical implementation of these ideas.

[2] We use lower-case to denote particular numerical choices for the random variables

Consider now the conditional entropy introduced in Lecture 2

$$\begin{aligned} H(X(t)|Y(t_0)) &\equiv -\int\int dx'(t)dy'(t_0)p(x'(t),y'(t_0))\ln\left(p(x'(t)|y'(t_0))\right) \\ &= -\int dy'(t_0)p(y'(t_0))\int dx'(t)p(x'(t)|y'(t_0))\ln\left(p(x'(t)|y'(t_0))\right) \end{aligned}$$

The second line here shows that the conditional entropy is the expected entropy of the (univariate) conditional distribution. Since this entropy reflects the expected uncertainty in the distribution given perfect knowledge of $Y(t_0)$ it will intuitively be less than $H(X(t))$ which measures the unconditioned uncertainty of $X(t)$. In fact we have seen already in Lecture 2 that the difference of these two entropies is the so called mutual information I between $x(t)$ and $y(t_0)$:

$$I(X(t);Y(t_0)) = H(X(t)) - H(X(t)|Y(t_0)) \tag{1}$$

Intuitively then (see [3]) such a mutual information represents the expected reduction in uncertainty in $x(t)$ associated with perfect knowledge of the initial condition variable $Y(t_0)$.

We shall refer to $I(X(t);Y(t_0))$ henceforth as the "time lagged mutual information" or TLMI. The TLMI does not measure the information flow into X from other random variables: Consider a typical multivariate distribution for the initial time t_0. For spatial points that are close together there is often a high correlation between random variables $X(t_0)$ and $Y(t_0)$ and hence the TLMI between $Y(t_0)$ and $X(t)$ may measure simply the importance of persistence to the particular random variable $X(t)$. Schreiber (see [8]) has suggested another functional which in his view, better reflects information flow: Suppose there was on the contrary no information flowing into X from other random variables such as Y. In such a case one could model the sequence of random variables $X(t)$ as a univariate process meaning that

$$p(x(t)|x(t_0)) = p(x(t)|x(t_0),y(t_0)) \qquad \forall x(t),x(t_0),y(t_0) \tag{2}$$

In other words knowledge of the other random variables Y at the initial conditions makes no difference to the future of the random variable X if $x(t_0)$ is known. Schreiber suggests using the deviation from this property as a measure of the information flow from Y into X. We can use the relative entropy of the two distributions as such a measure. For particular choices of the conditioning variables $x(t_0)$ and $y(t_o)$ this is

$$D(p(x(t)|x(t_0),y(t_0)),p(x(t)|x(t_0))) \equiv \int dx(t)p(x(t)|x(t_0),y(t_0))\ln\left[\frac{p(x(t)|x(t_0),y(t_0))}{p(x(t)|x(t_0))}\right]$$

The transfer entropy (which we abbreviate as TE) is then the expected value of this deviation when all possible values of the conditioning variables are considered i.e.

$$\begin{aligned} T(Y\to X,t,t_0) &\equiv \int\int dx(t_0)dy(t_0)p(x(t_0),y(t_0))D(p(x(t)|x(t_0),y(t_0)),p(x(t)|x(t_0))) \\ &= \int\int dx(t_0)dy(t_0)p(x(t),x(t_0),y(t_0))\ln\left[\frac{p(x(t)|x(t_0),y(t_0))}{p(x(t)|x(t_0))}\right] \end{aligned} \tag{3}$$

If we consider now the spatially close $x(t_0)$ and $y(t_0)$ it is clear then if they are highly correlated then $p(x(t)|x(t_0))$ will be close to $p(x(t)|x(t_0), y(t_0))$ since the addition of $y(t_0)$ will not provide much further influence on $x(t)$ beyond that already provided by $x(t_0)$. We note that due to its definition in terms of the relative entropy functional D, transfer entropy is non-negative but also generally non-symmetric since usually

$$T(Y \to X, t, t_0) \neq T(X \to Y, t, t_0)$$

which allows a direction to be associated with information interchange between random variables.

One can also write the transfer entropy in a form analogous to equation (1):

$$T(Y \to X, t, t_0) = H(X(t)|X(t_0)) - H(X(t)|X(t_0), Y(t_0))$$

which shows that it has the interpretation of the expected reduction in uncertainty in the target random variable $X(t)$ due to knowledge of the initial time variable $Y(t_0)$ *beyond that which knowledge of the target variable at the initial condition time would give.*

In summary TLMI allows one to identify which initial condition variables require better observation in order to reduce prediction error. TE excludes the influence of persistence in such a calculation and might be thought of as better reflecting the intuitive meaning of information flow. As we shall see TLMI has the practical advantage over TE of requiring smaller ensemble sizes for its calculation. This is a result of the bivariate nature of TLMI as opposed to the trivariate nature of TE.

3 Application to atmospheric prediction

The functionals discussed in the previous section have been used in several applications which are detailed in the citations given at the start of that section. Here we consider a semi-realistic atmospheric ensemble prediction setup (it is the same as that discussed in the last section of the previous lecture). The concrete application we have in mind is to decide how exactly the initial conditions need to be improved in order to reduce uncertainty in predictions at other locations and with respect to other dynamical variables. Results shown are taken from published work by the lecturer [2].

We focus on a particular target point where we wish to improve predictions (i.e. reduce uncertainty). In particular we choose a target point in the middle of the North Atlantic. In Figure 1 and Figure 2 we consider a surface target point and consider how to reduce uncertainty in one day predictions of temperature and zonal (east west) wind. Figure 1 show a plot of TLMI and TE for temperature. These show that in order to improve temperature predictions at one day the best strategy is to reduce uncertainty in the vicinity of the target point. Uncertainty flow from other areas is not all that important relative to the persistence of uncertainty at the target region (compare TLMI with TE).

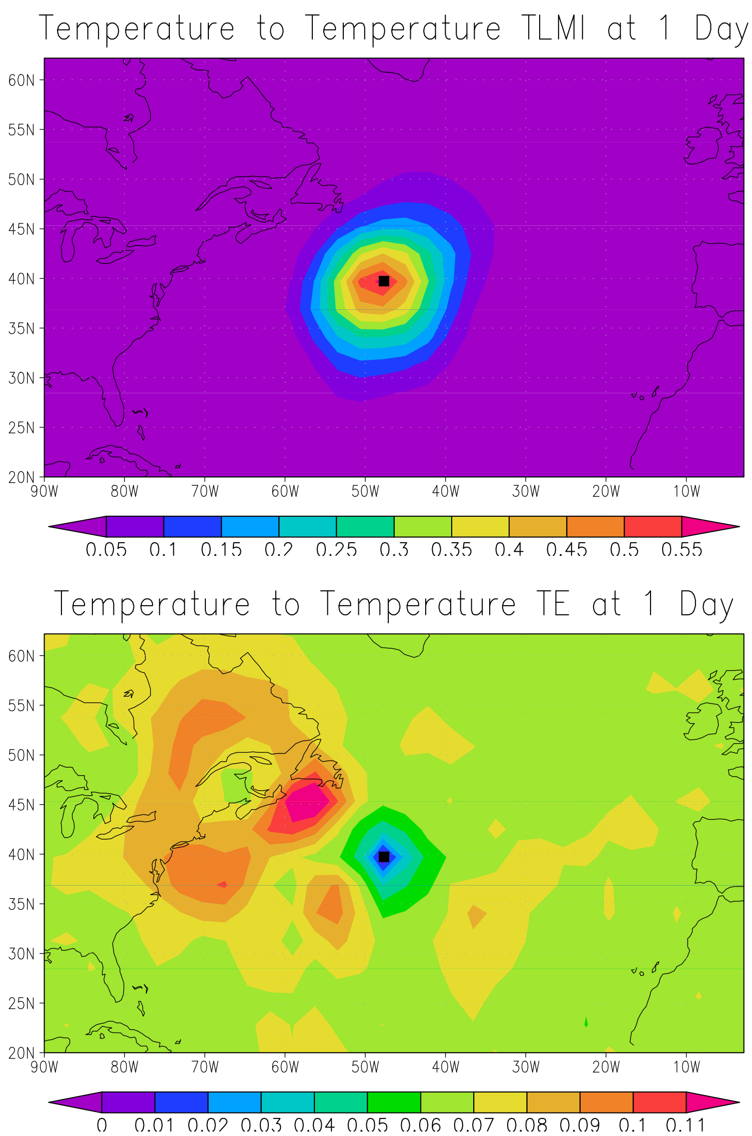

Figure 1: Uncertainty transfer to the target point (the solid square). The target and source variables are temperature and TLMI and TE (see text) are shown in nats. Note the reduced values for TE compared with TLMI.

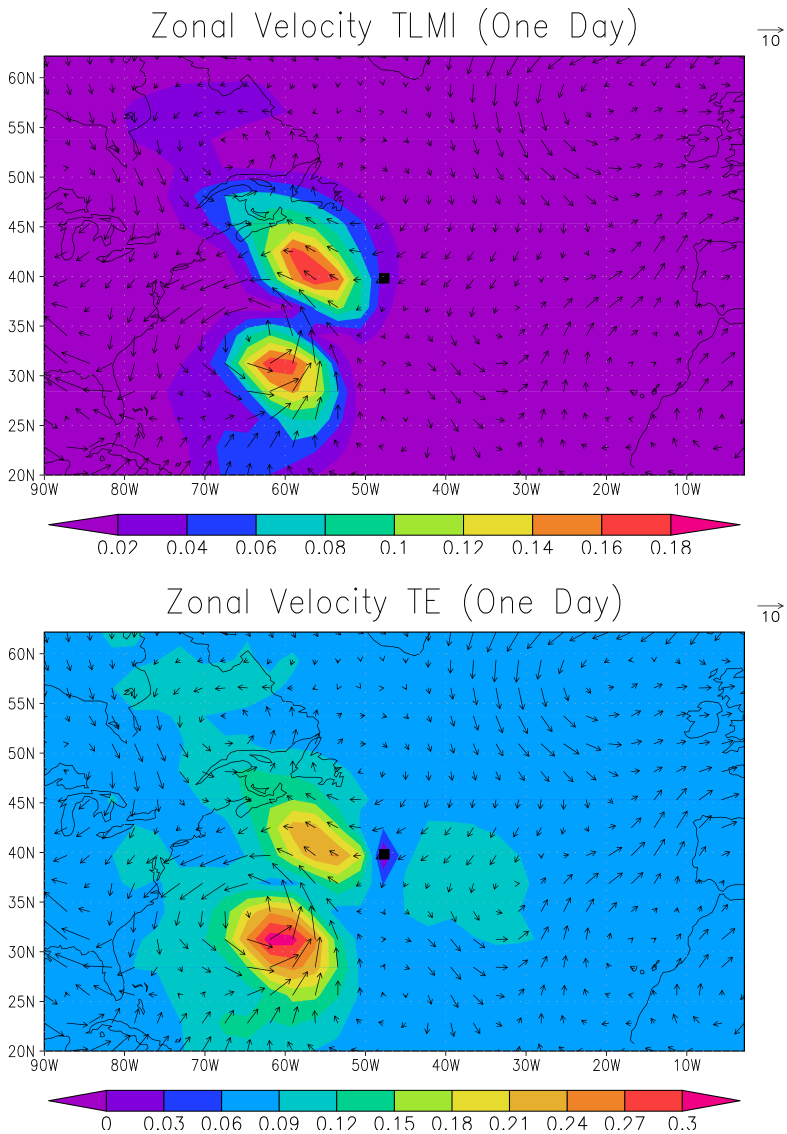

Figure 2: As for Figure 1 but for zonal wind. The mean winds for the ensemble at the initial conditions are shown as arrows. The TE exceeds the TLMI due to a sampling error offset of approximately 0.1 nats in TE.

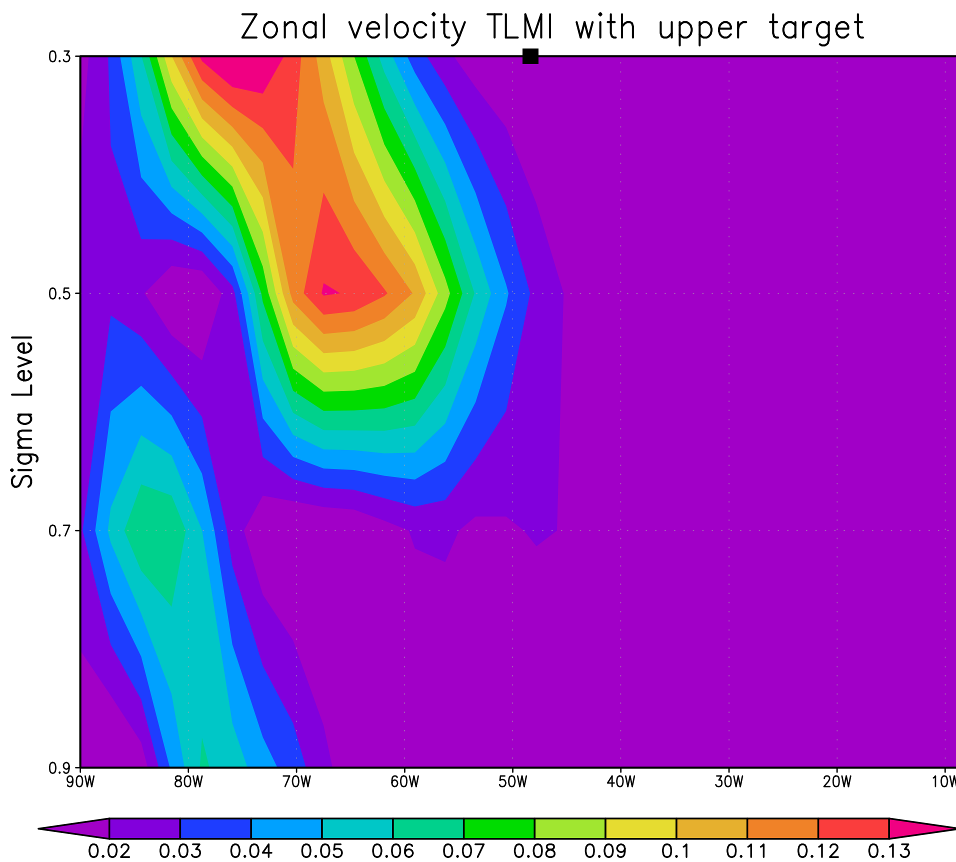

Figure 3: Longitude vertical plot of TLMI along 40 degrees north. The target here is in the upper troposphere (solid square) and is zonal wind like Figure 2. The vertical units can be converted approximately to pressure in millibars by multiplying by 1000.

The situation for zonal velocity is quite different as Figure 2 shows. Most of the uncertainty here flows in from remote locations (note the similarity between TE and TLMI). The important regions for the zonal wind target correspond very closely with a large surface cyclone in the mean initial conditions. This suggests that the dynamics underlying this regional cyclone need better definition to reduce uncertainty in the zonal wind field in the more general North Atlantic region.

By choosing various targets and dynamical variables we find that in general most of the uncertainty flows horizontally in the atmosphere. Occasionally however we do see some vertical propagation. Figure 3 shows just such a case. Here our target is in the upper troposphere near the jet-stream and we are considering again how to improve one day zonal wind forecasts. The jet-stream is primarily zonal in nature so we are really asking how to reduce uncertainty in the prediction of the jet-stream at a particular location. The plot of TLMI is a zonal section along 40 degrees north and shows quite a bit of uncertainty coming up from the mid-troposphere.

It is hoped that analyses like those shown above will have a future positive impact on deciding where to put additional elements of the observing platform which is used to initialize weather predictions.

4 More fundamental approaches

The arguments given in Section 2 regarding the use of TLMI and TE as measures of uncertainty flow are essentially heuristic. We now consider more fundamental approaches. For clarity we limit out approach here (see [4]) to a two dimensional system however the approach we outline has since been extended to n dimensional systems (see [5] and [6]) and also two components of arbitrary finite dimension (see [7]).

Recall from Lecture 5 that a general dynamical system satisfying

$$\frac{\partial x_i}{\partial t} = A_i(\mathbf{x}, t) \tag{4}$$

has an associated Fokker-Planck (or for this special case a Liouville) equation given by

$$\partial_t p = -\sum_{i=1}^{N} \partial_i \left[A_i(\mathbf{x}, t) p \right] \tag{5}$$

In Lecture 5 we derived an entropy evolution equation for this system:

$$\partial_t h = \int p \left[\sum_{i=1}^{N} \frac{\partial A_i}{\partial x_i} \right] d\overrightarrow{x} \equiv E(\nabla \bullet \overrightarrow{A}) \tag{6}$$

where E denotes expectation with respect to the probability density applying at a particular time. This has a straightforward interpretation in terms of the evolution of volume elements within the dynamical system: If one takes a small region within the original state space and evolves points within this ball according to (5) then the rate of change of the volume of the region will be $\nabla \bullet \overrightarrow{A}$ thus the rate of change of entropy is the average rate of change of all volume weighted by the relevant probability distribution.

Suppose we have a two dimensional system $p(x_1, x_2)$ and consider the time evolution of the marginal entropy with respect to the first component[3]. If the first component were not interacting with the second then this marginal entropy would evolve according to the appropriate one dimensional version of equation (6) i.e.

$$\partial_t h_1^* \equiv E(\frac{\partial A_1}{\partial x_1})$$

[3] i.e. $h_1 \equiv \int p(x_1) \log p(x_1) dx_1$ where $p(x_1) = \int p(x_1, x_2) dx_2$

where the star superscript is indicating this as a hypothetical isolated evolution of entropy. The true evolution of h_1 can be derived in a straightforward way from equation (5):

$$\partial_t h_1 = -\int\int \frac{p(x_1,x_2)}{p(x_1)} A_1 \frac{\partial p(x_1)}{\partial x_1} dx_1 dx_2$$

the difference between these two expressions (i.e. $\partial_t(h_1 - h_1^*)$) then has the natural interpretation as the flow of entropy from component 2 into component 1. The reverse flow can obviously be derived in a similar fashion and can be shown to not necessarily be of the opposite sign. The interested reader can find more details on this measure as well as a comparison with Schreiber's measure of transfer entropy (TE) in Liang and Kleeman [4].

References

[1] K. Kaneko. Lyapunov analysis and information flow in coupled map lattices. *Physica D*, 23:436, 1986.

[2] R. Kleeman. Information flow in ensemble weather predictions. *J. Atmos Sci.*, 64(3):1005–1016, 2007.

[3] L.-Y. Leung and G. R. North. Information Theory and Climate Prediction. *J. Clim.*, 3:5–14, 1990.

[4] X. S. Liang and R. Kleeman. Information transfer between dynamical system components. *Phys. Rev. Lett.*, 95(24):244101, 2005.

[5] X. S. Liang and R. Kleeman. A rigorous formalism of information transfer between dynamical system components I. Discrete maps. *Physica D.*, 231:1–9, 2007.

[6] X. S. Liang and R. Kleeman. A rigorous formalism of information transfer between dynamical system components II. Continuous flow. *Physica D.*, 227:173–182, 2007.

[7] A. J. Majda and J. Harlim. Information flow between subspaces of complex dynamical systems. *Proceedings National Academy of Science*, 104(23):9558–9563, 2007.

[8] T. Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, 2000.

[9] J. A. Vastano and H. L. Swinney. Information transport in spatiotemporal systems. *Phys. Rev. Lett.*, 60:1773, 1988.