# Information Flow/Transfer
# Review of Theory and Applications

Richard Kleeman

Courant Institute of Mathematical Sciences

With help from X. San Liang, Andy Majda and John Harlim and support from the NSF CMG and ATM programs.

# Work on information theory applications

R. Kleeman. Measuring dynamical prediction utility using relative entropy. J. Atmos Sci, 59:2057-2072, 2002.

A. J. Majda, R. Kleeman, and D. Cai. A framework of predictability through relative entropy. Meth. Appl. Anal., 9:425-444, 2002.

R. Kleeman, A. J. Majda, and I. Timofeyev. Quantifying predictability in a model with statistical features of the atmosphere. Proc. Natl. Acad. Sci., 99:15291-15296, 2002.

M.K. Tippett, R. Kleeman, and Y. Tang. Measuring the potential utility of seasonal climate predictions. Geophys. Res. Lett., 31:L22201, 2004. doi 10.1029/2004GL020673.

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

R. Kleeman and A. J. Majda. Predictability in a model of geostrophic turbulence. J. Atmos Sci, 62:2864-2879, 2005.

R. Abramov, A.J. Majda, and R. Kleeman. Information theory and predictability for low frequency variability. J. Atmos Sci, 62:65-87, 2005.

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.

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.

R. Kleeman. Statistical predictability in the atmosphere and other dynamical systems. Physica D, 230:65-71, 2007.

R. Kleeman. Information flow in ensemble weather predictions. J. Atmos Sci., 64(3):1005-1016, 2007.

Y. Tang, R. Kleeman, and A.M. Moore. Comparison of Information-based Measures of Forecast Uncertainty in Ensemble ENSO Prediction. J. Clim., 21:230-247, 2008.

R. Kleeman. Limits, variability and general behaviour of statistical predictability of the mid-latitude atmosphere. J. Atmos Sci, 65:263-275, 2008.

# The problem of information flow

Suppose we have a set of random variables $X_m^k$ where the upper index refers to space and the lower index refers to time.

In general the uncertainty or entropy of a variable for a given spatial location will be be time dependent  but may also depend on the time evolution of the random variables at other spatial locations.

Conceptually we shall consider the information flow between two spatial locations to be the time rate of change in entropy or uncertainty  at a particular location due to the influence of another spatial location. If all spatial locations evolve independently then the information flow within the system is zero. In most interesting dynamical systems however this will not be the case.

The motivation for studying this is that uncertainty created at one location may influence uncertainty at another location at a later time. This has clear implications for the problem of prediction.

# Intuitively defined measures of transfer

Problem studied first by physicists in the 1980s. Kaneko (1986) studied the propagation of perturbations in simple non-linear dynamical systems using a "moving frame" or co-moving Lyupanov exponent. Maximization of this showed the prefered velocity of growing perturbations. Since regular Lyupanov exponents are often related to entropy production it was natural to try to find an information theoretic counterpart for co-moving exponents. This turned out empirically and in the systems studied, to be the time lagged mutual information of random variables:

$$I(X_n^j; X_m^k) \equiv \sum_{x \in \mathcal{X}_n^j} \sum_{y \in \mathcal{X}_m^k} p(x,y) \ln \frac{p(x,y)}{p(x)p(y)}$$

Which has the natural velocity scale

$$v \equiv \Delta x(n-m)/\Delta t(j-k)$$

The lagged mutual entropy turned out to be maximized when this velocity scale matched that which maximized the co-moving Lyupanov exponent.

# Intuitively defined measures of transfer

In addition to this match of physical propagation scales, mutual information has an appealing intepretation as the reduction in uncertainty of $X_m^k$ due to perfect knowledge of $X_n^j$ i.e. roughly speaking, the contribution of uncertainty in the former due to uncertainty in the latter.

This measure of information flow was further verified as physically plausible in more complex and realistic dynamical systems by Vastano and Swinney (1988). It was however shown to give misleading results in certain pathological situations by Schreiber (1990). In particular when both $X_m^k$ and $X_n^j$ are subjected to a synchronized source of uncertainty then unphysical transfers are possibly indicated by the lagged mutual information. This is somewhat analogous to over interpreting a correlation as causitive. Note that the mutual information reduces to a simple function of correlation in the case that the distributions are Gaussian.

# Intuitively defined measures of transfer

Schreiber (2000) suggested a new information theoretic measure of flow which overcame the problem of lack of causality identified in his 1990 study. He considered the situation where each spatial location was a Markov process of order q. Thus the probability function at any particular point depends only on the values for the previous q random variables. In such a situation there can, by construction, be no information flow between spatial points. He then tested the deviation from this null hypothesis using a (conditional) relative entropy functional. For an order 1 Markov process this transfer entropy is defined as

$$T(j \to i, n) \equiv \sum p(i_{n+1}, i_n, j_n) \ln \frac{p(i_{n+1}|i_n, j_n)}{p(i_{n+1}|i_n)}$$

Here the i and j indices refer to different spatial random variable values at various times given by the subscripts. This formula can be extended in an obvious manner to a Markov process of order q.

Notice that from the practical computational viewpoint (to be visited shortly) the transfer entropy (TE) is defined with respect to trivariate distributions whereas the lagged mutual information functional is defined with respect to bivariate distributions. The order q TE is defined on distributions of dimension q+2.

# Somewhat more formal approaches

There is something a little unsatisfying about the intuitive approaches proposed above in that they do not stem from a governing global flow equation for uncertainty. Motivated by this Liang and Kleeman (2005) explored a different somewhat more formal approach. Suppose we have a 2D autonomous system given by

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{F}(\boldsymbol{x})$$

From the associated Liouville equation we can easily derive an equation for the evolution of the total entropy of the system

$$\frac{dH}{dt} = E\left(\nabla \bullet \boldsymbol{F}\right)$$

Where

$$E\left(G\right) \equiv \iint p(x_1, x_2) G dx_1 dx_2$$

Thus the evolution of entropy is controlled by the expected (globally averaged) contraction or expansion of phase space namely $\nabla \bullet \boldsymbol{F}$

# Somewhat more formal approaches

The evolution of the marginal distribution $p(x_1)$ can be derived easily by integrating the original Liouville equation in the $x_2$ direction and from that it is easy to derive the evolution equation for the corresponding marginal entropy

$$\frac{dH_1}{dt} = -\iint p(x_1, x_2) \left[ \frac{F_1}{p(x_1)} \frac{\partial p(x_1)}{\partial x_1} \right] dx_1 dx_2$$

Now if the random variable $X_1$ were to evolve in isolation then its entropy should satisfy the modified evolution equation

$$\frac{dH_1^*}{dt} = E\left(\frac{\partial F_1}{\partial x_1}\right) = \iint p(x_1, x_2) \frac{\partial F_1}{\partial x_1} dx_1 dx_2$$

Thus it seems reasonable to identify the difference $H_1 - H_1^* \equiv T_{2\to 1}$

with the flow of uncertainty between component 2 and component 1 of the system.

# Somewhat more formal approaches

The above approach has been tested in a number of simple systems and gives qualitatively (but not quantitatively) similar results to the transfer entropy of Schreiber. It has been extended to the case of n dimensions by Liang and Kleeman (2007a) and (2007b) using a similar entropy evolution philosophy. The resulting functionals become complex and the issue of practical computation is still being considered.

Another potentially more practical extension has been made by Majda and Harlim (2007) to a two component dynamical system. The two components of the system are allowed to have arbitrary finite dimension. In addition they consider the case of stochastic forcing in the underlying dynamical system:

$$\begin{aligned} \frac{dx}{dt} &= F_1(x) + F_{12}(x,y) \\ \frac{dy}{dt} &= F_2(x,y) + \sigma \dot{W} \end{aligned}$$

# Somewhat more formal approaches

Associated with this system is evidently a Fokker Planck rather than the simpler Liouville equation. The general idea of Liang and Kleeman (2005) above still goes through in a similar manner. Importantly however practical results are obtained for cases when the marginal distribution

$p_1(x,t)$ has the form of an exponential family distribution and in particular when it is Gaussian they obtain

$$T_{y \to x} = -E_p(\nabla_x \bullet F_{12}) + E_p\left(F_{12} \bullet \Gamma^{-1}\left(x - E_p(x)\right)\right)$$

where $\Gamma$ is the covariance matrix of the Gaussian marginal. This expression is simply a series of moments with respect to the multivariate probability distribution and therefore amenable to practical calculation using ensembles. Many practical ensembles are quasi-Gaussian.

In the case that the underlying dynamical system satisfies further conditions which are reasonable for certain geophysical flows they are also able to derive a very interesting relation between information flow; energy flow and the rate of change of relative entropy of the marginal distribution with respect to its equilibrium distribution.

# Application to observing networks for prediction

Consider a situation where one wishes to reduce the uncertainty of a prediction at a very particular spatial location. An example might be the prediction (or otherwise) of a very intense storm in a significant location.

How might the uncertainty of this prediction be reduced? One strategy is evidently to improve the observing system which defines the initial conditions for the prediction. There are however usually time constraints involved in such an improvement. For example one might choose to send out an aircraft to improve observations in a "critical" region. But where might such a region be located? For hurricane prediction this is usually fairly obvious but for mid-latitude storms much less so.

This has led to the concept of targeted observations and much sensitivity analysis of the complex dynamical models underlying the atmosphere has been undertaken. This has often had the limitation however that it is linear in nature and beyond a certain prediction time frame, this is a highly suspect assumption (3-4 days typically).

# Application to observing networks for prediction

Information flow offers a natural way of analyzing this problem: If one is able to calculate the uncertainty flow from initial time spatial points to prediction time spatial points then it becomes clear how the uncertainty at the latter times may be improved by reducing the uncertainty at the initial time.

Moreover if we have information on how this flow connects all spatial points then we can identify how the uncertainty may be optimally reduced at the initial time in order to reduce errors in predictions at particular spatial locations of interest.

Information flow calculations make no assumptions about linearity but do require a statistically significant number of "ensemble" predictions.

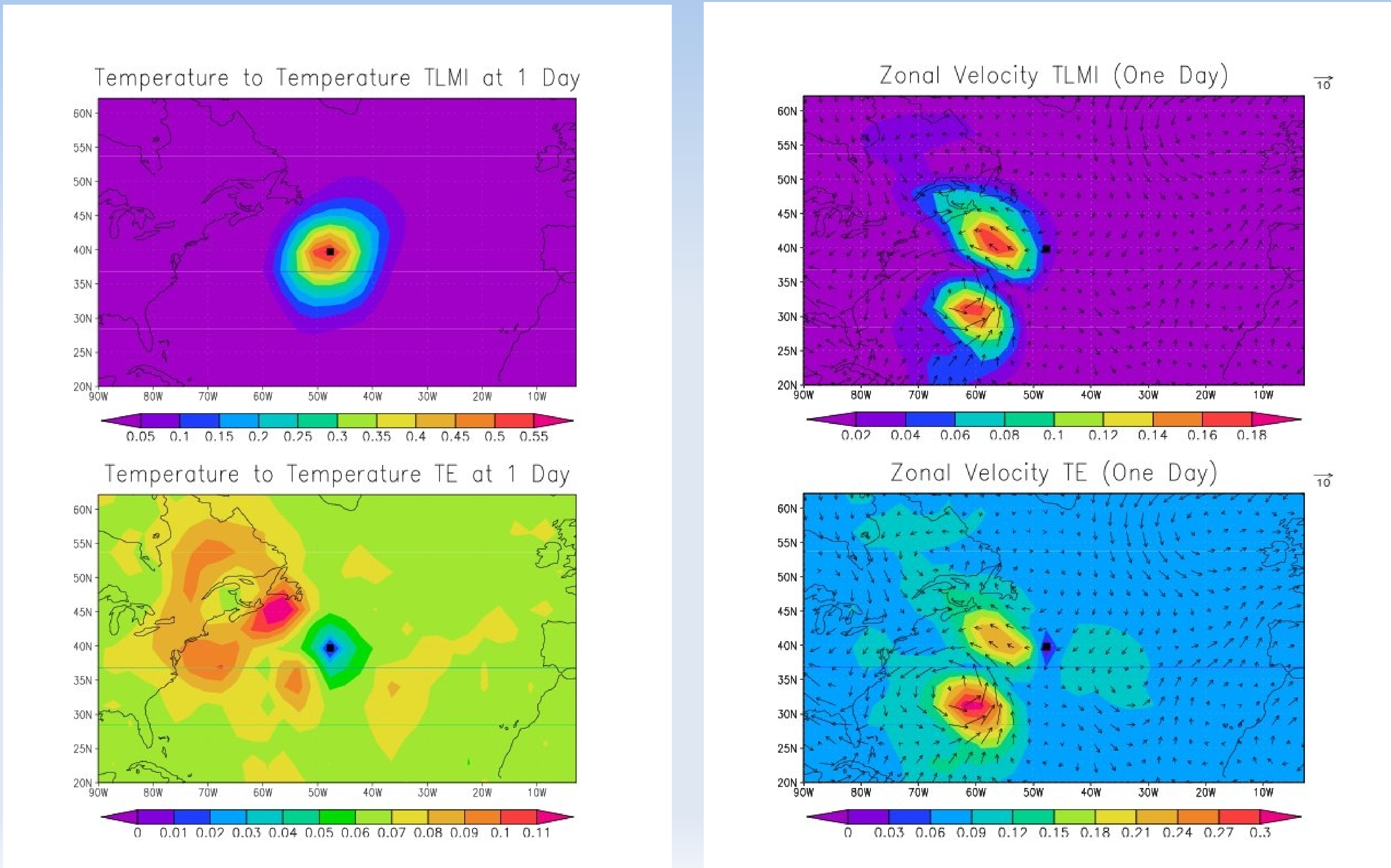
The latter restriction is important practically and requires the use of functionals of a low degree of multivariateness.

# Application to observing networks for prediction

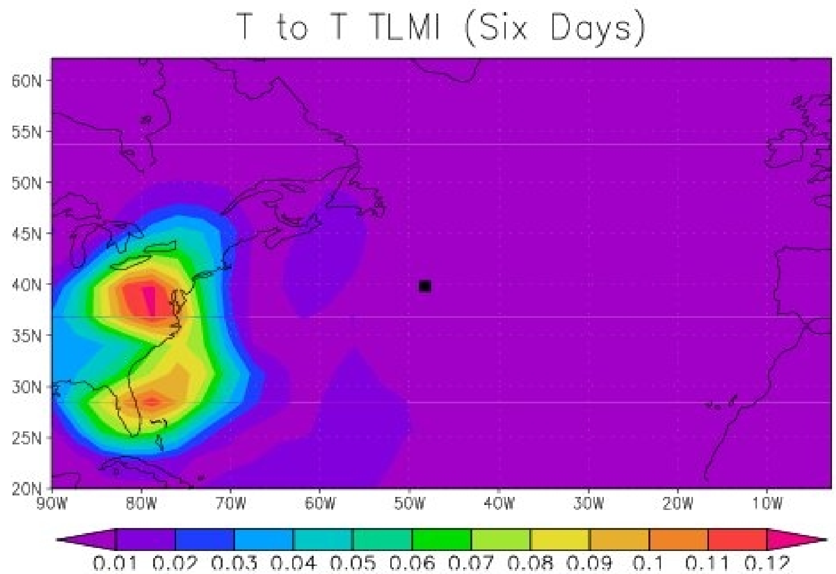
Kleeman (2007) carried out this program in a (somewhat) realistic model of the atmosphere using both intuitive measures (lagged mutual information and transfer entropy) of information flow discussed above. There are plans to extend the computation to the more formal measures later once some technical issues are resolved.

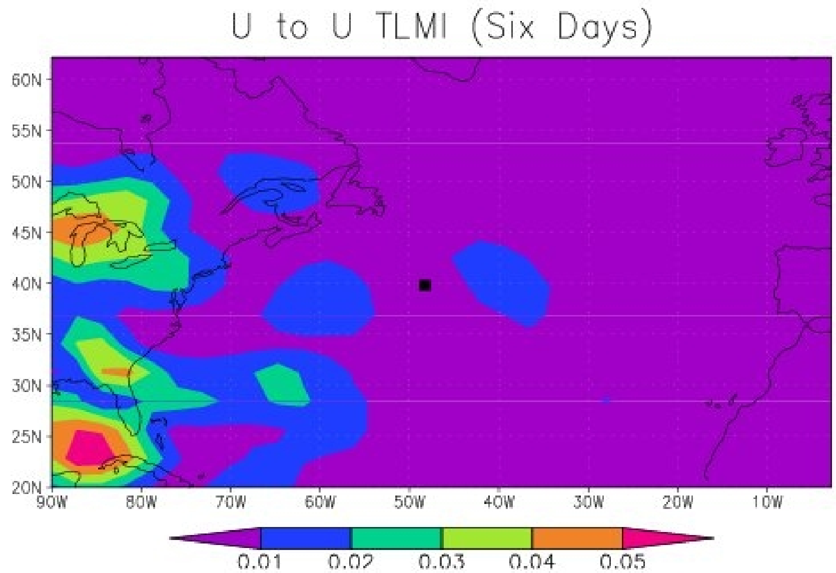
A T42 primitive equation dry dynamical core model with realistic depiction of mid-latitude jets and storms was utilized. A simple multivariate Gaussian distribution was assumed for initial conditions and sample trajectories of length 10 days generated using this distribution with means drawn from a very long model run (ergodicity is assumed). The ensembles were of size 10,000 which enabled highly statistically significant (non-noisy) estimates of the information flow functionals.

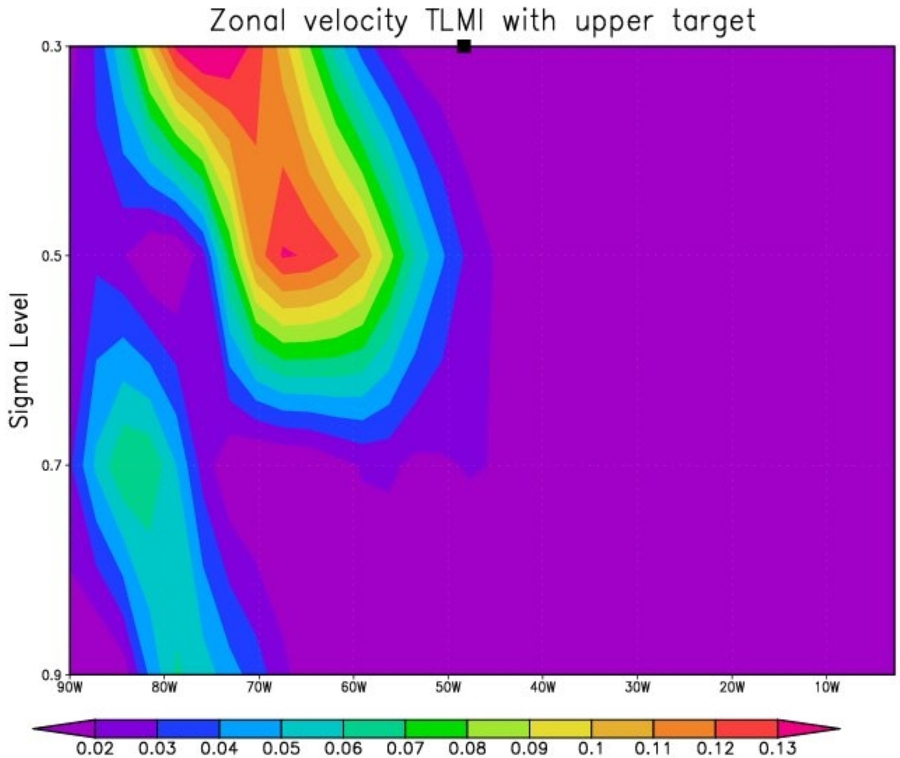
# Application to observing networks for prediction



# Application to observing networks for prediction



# Application to observing networks for prediction. Practical issues

The ensemble sizes used above are not yet practical for a real world application. The ensembles are however quasi-Gaussian meaning that analytical expressions for the relevant functionals can be easily derived and used as good approximations. Since these involve only the first two moments of the samples/ensembles much smaller ensembles are required (order 200 rather than 10,000).

In all practical cases there was little qualitative difference between the two intuitive measures except for the persistence effect. This suggests that the theoretical debate between different flow functionals may not be important for this particular application. Of course in other applications that may not be the case. This is under active current investigation.

# Conclusions

- Significant theoretical progress has been made in defining information flow. More work is still required.
- It has been shown that these concepts can be applied to the practical prediction problem of targetted observing networks. Unlike all other proposed techniques the present methods do not make any assumptions of linearity.
- Many other applications of this theory are potentially possible in other areas.