Dynamical predictability and initialization: a statistical prediction perspective

Richard Kleeman

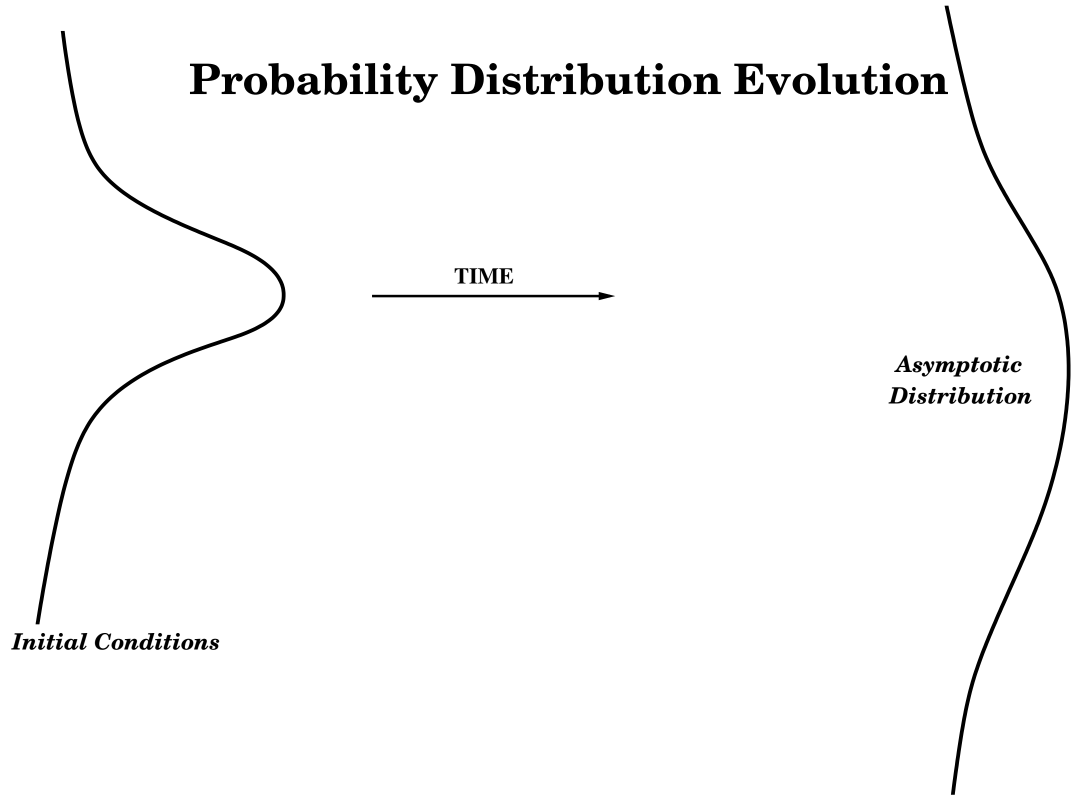
Courant Institute of Mathematical Sciences New York

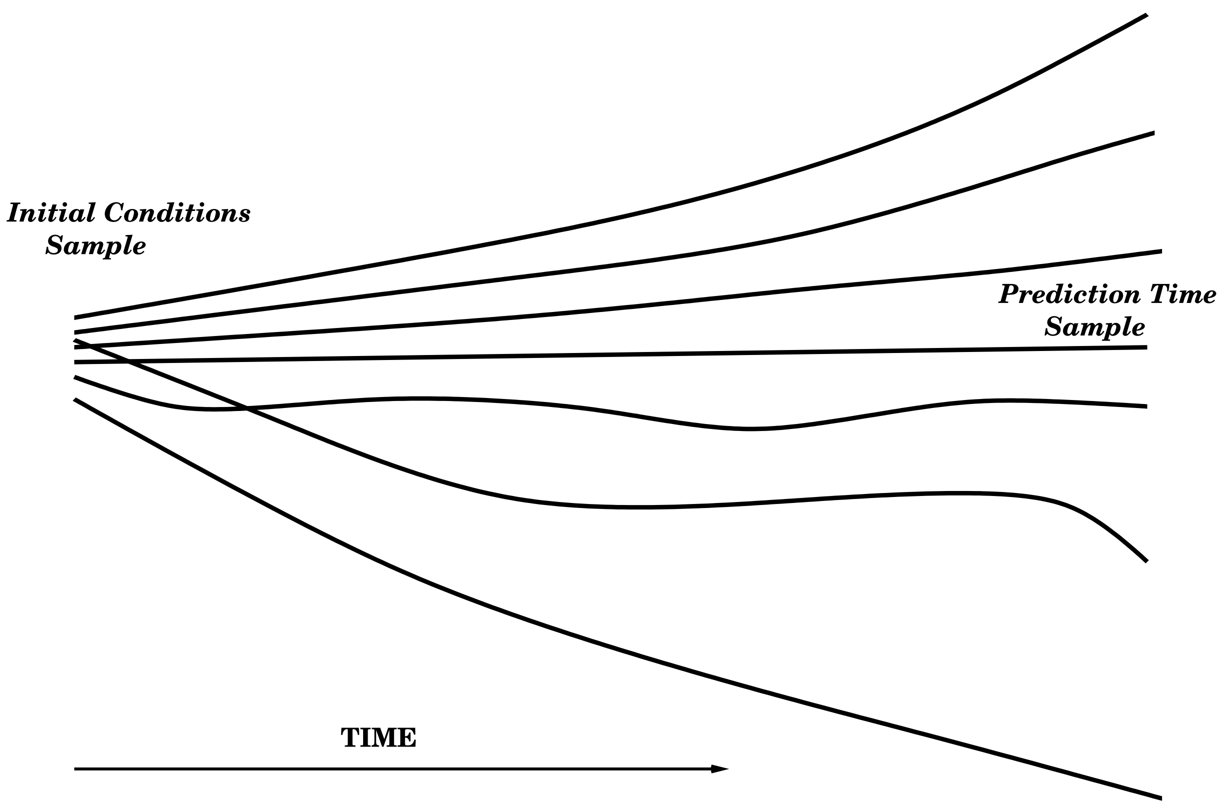
Relevant Publications

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

R. Kleeman and A. J. Majda, "Predictability in a model of geostrophic turbulence" J. Atmos Sci, 2004, in press.

Probability Distribution Evolution





Relative Entropy

Discrete Form:

$$D(p \parallel q) \equiv \sum_{x \in \mathcal{H}} p(x) \ln \left(\frac{p(x)}{q(x)} \right)$$

where $\mathcal{H}$ is some partitioning of our state space R^n.

Differential Form:

$$D(f \| g) \equiv \int_{all\ R^n} f(\overrightarrow{z}) \ln \left(\frac{f(\overrightarrow{z})}{g(\overrightarrow{z})} \right) d\overrightarrow{z}$$

The differential form can be considered as the limit of the discrete form (unlike absolute entropy).

Relative Entropy Properties

Non-negativity

$D(p(t), q(t)) \geq 0 \quad \forall t$

Equality holds only when $p(t) = q(t)$

Temporal Monotonicity

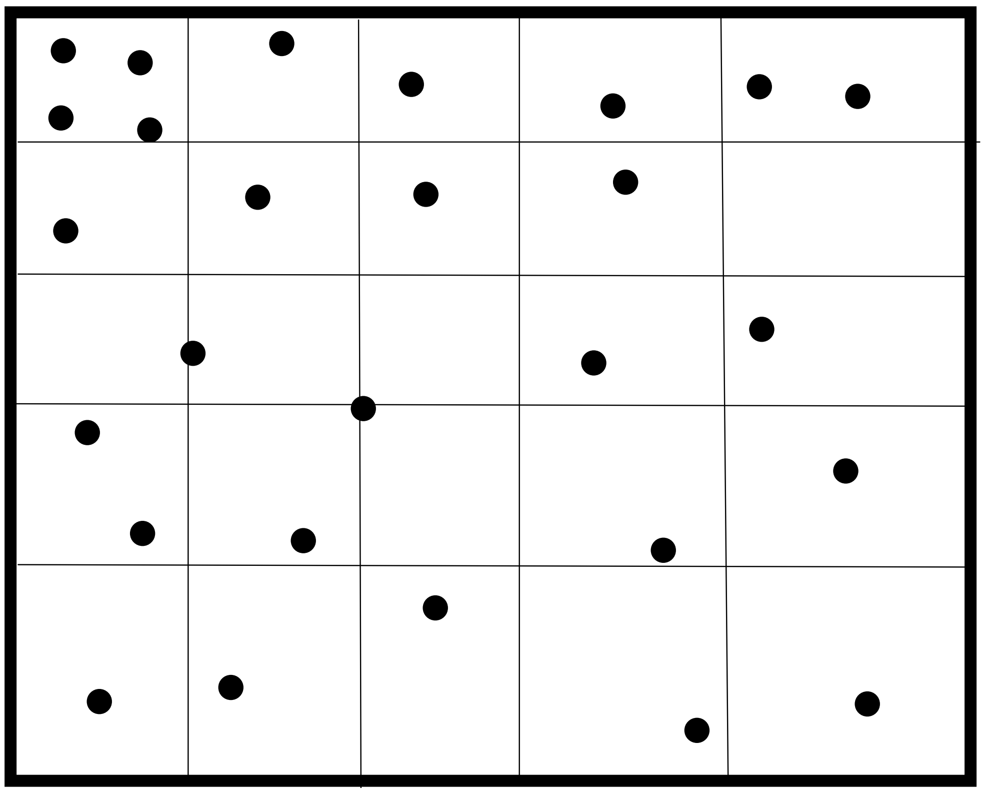
$D(p(t_1), q(t_1)) \geq D(p(t_2), q(t_2)) \quad t_2 > t_1$

Holds for a discrete temporal process when the forward temporal conditional probability
$p(\overrightarrow{x}(t+1)|\overrightarrow{x}(t))$ is time invariant where $\overrightarrow{x}$ is the **full** state vector.

Invariance

The relative entropy is invariant under general non-linear transformations G of state vector variables $G : \overrightarrow{x} \rightarrow \overrightarrow{y}$ providing the corresponding Jacobian does not vanish.

Coarse Graining and Monte Carlo simulations



Information Content Tradeoff:

Coarser partitions imply less knowledge and so lower information content

Finer partitions imply uncertainty over partition probability (sampling error) hence lost information

This can be all made precise mathematically.

Related to Boltzmann's paradox and irreversibility.

Ensembles and PDFs

- **Partitioning of state space enables an ensemble to define a sample estimate $\widehat{p_i}$ for the (true/population) probability p_i of a particular partition element i.**

- **A particular partition sample probability estimate has a a (meta) probability associated with it $P(\widehat{p_i})$ which can be calculated using Bayes theorem.**

- **The information loss in assuming $\widehat{p_i}$ when p_i holds is the relative entropy $D(p, \widehat{p})$.The total expected information loss can then be obtained using $P(p)$ and this relative entropy.**

- **Bottom line: Coarse partition implies tight $P(p)$ and so small expected information loss. However it also reveals less about state space and hence has a smaller information content.**

- **Cautionary tale: Assuming a Gaussian distribution can be shown to give a higher apparent information content than that obtained by partitioning. However, what is the uncertainty associated with assuming this Gaussian distribution?? There must be considerable "sampling" information loss associated with the assumed Gaussian model which cannot be easily quantified.**

Marginal Entropies and State Space Partitions

Suppose we have a state-space of dimension N which we denote by $\{x_i\}$. Suppose further we calculate the relative entropy on such a space with respect to a particular partition Γ. Denote this by $D_\Gamma(x_1, \ldots, x_N)$. Consider now a general marginal distribution $p(x_{j_1}, \ldots, x_{j_K})$. Define the following heirarchy of "marginal relative entropies":

$$\begin{array}{rcc} D_\Gamma^{(1)}(\overrightarrow{x}) & \equiv & \frac{1}{N}\sum_{i=1}^{N} D_\Gamma(x_i) \\ D_\Gamma^{(2)}(\overrightarrow{x}) & \equiv & \frac{1}{C_2^N}\sum_{i,j=1}^{N} D_\Gamma(x_i, x_j) \\ \vdots & \vdots & \vdots \\ D_\Gamma^{(N)}(\overrightarrow{x}) & \equiv & D_\Gamma(x_1, \ldots, x_N) \end{array}$$

We call these the univariate, bivariate, trivariate etc relative entropies. Using the chainrule of relative entropy one can show that

$$D_\Gamma^{(1)}(\overrightarrow{x}) \leq D_\Gamma^{(2)}(\overrightarrow{x}) \leq \ldots \leq D_\Gamma^{(N)}(\overrightarrow{x}) = D_\Gamma$$

showing that information increases for a particular partition as we consider greater multivariate behaviour.

HOWEVER one can also prove

$$D_{\Gamma}^{(i)}(\overrightarrow{x}) \leq D_{\Lambda}^{(i)}(\overrightarrow{x}) \; for \; \Lambda \sqsubseteq \Gamma$$

and finer partitionings are possible without significant sampling information loss for lower order $D_{\Gamma}^{(i)}(\overrightarrow{x})$

This paradoxical fact simply reflects the fact that there is an instrinsically FINITE amount of information in an ensemble and this is often well short of that in the full pdf.

Atmospheric Predictability Experiments

- T42 global dynamical core. 5 vertical levels (Puma model from U. Hamburg).

- Physics replaced by Rayleigh friction at surface and Newtonian cooling for radiation and convection.

- Model produces realistic mid-latitude synoptic variability. Northern winter hemisphere analyzed.

- Ensembles of initial conditions produced assuming a Gaussian distribution with uniform decorrelation of around $1000km$. Point variances in prognostic variables (temperature, divergence, vorticity and surface pressure) of roughly one order of magnitude less than climatology assumed.

Ensemble Design

- 9,600 member ensemble constructed with 90 day integration each.

- Initial condition means drawn from an extended climatological integration (1200 years) which was also used to obtain a 9,600 member climatological ensemble.

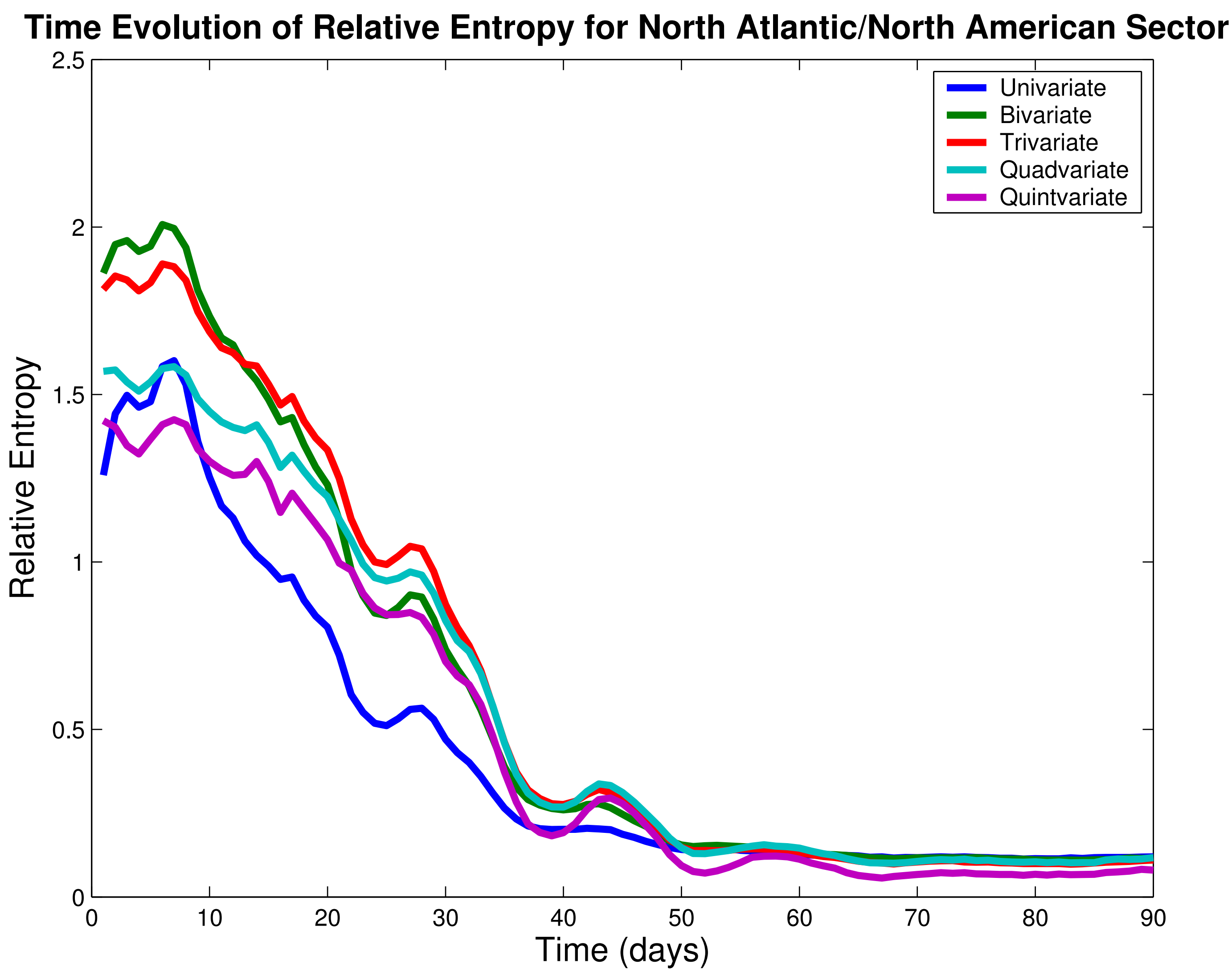
- Predictability examined in the North American-Atlantic Sector (90W-0W and 20N-65N).

- Multivariate/Marginal relative entropy calculated using first ten stream function EOFs which explain more than 90% of variance.

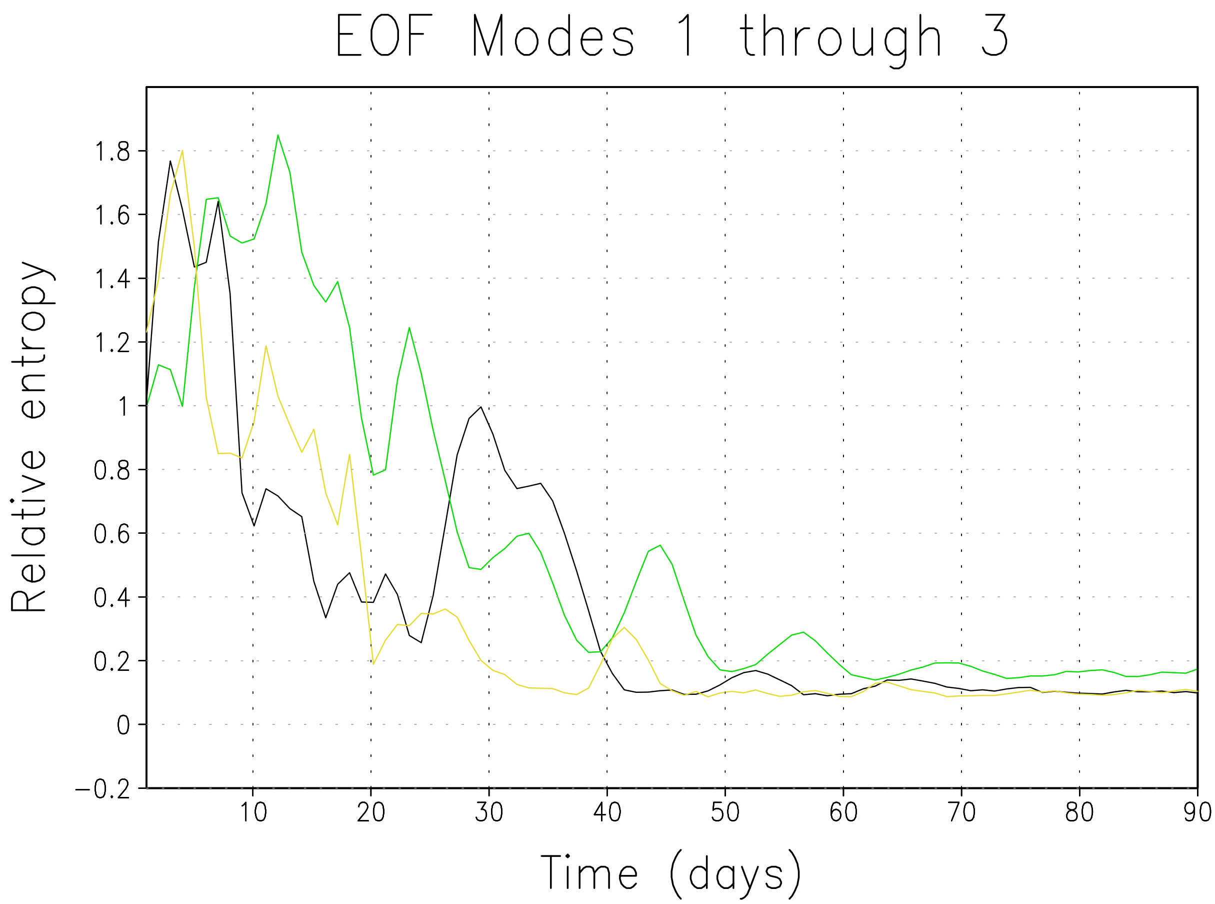
- Number of partitions used per variate: Univariate 1024. Bivariate 32. Trivariate 10. Quadravariate 6 and Quintvariate 4. This implies that there were a total of 1024, 1024, 1000, 1296 and 1024 partitions respectively.

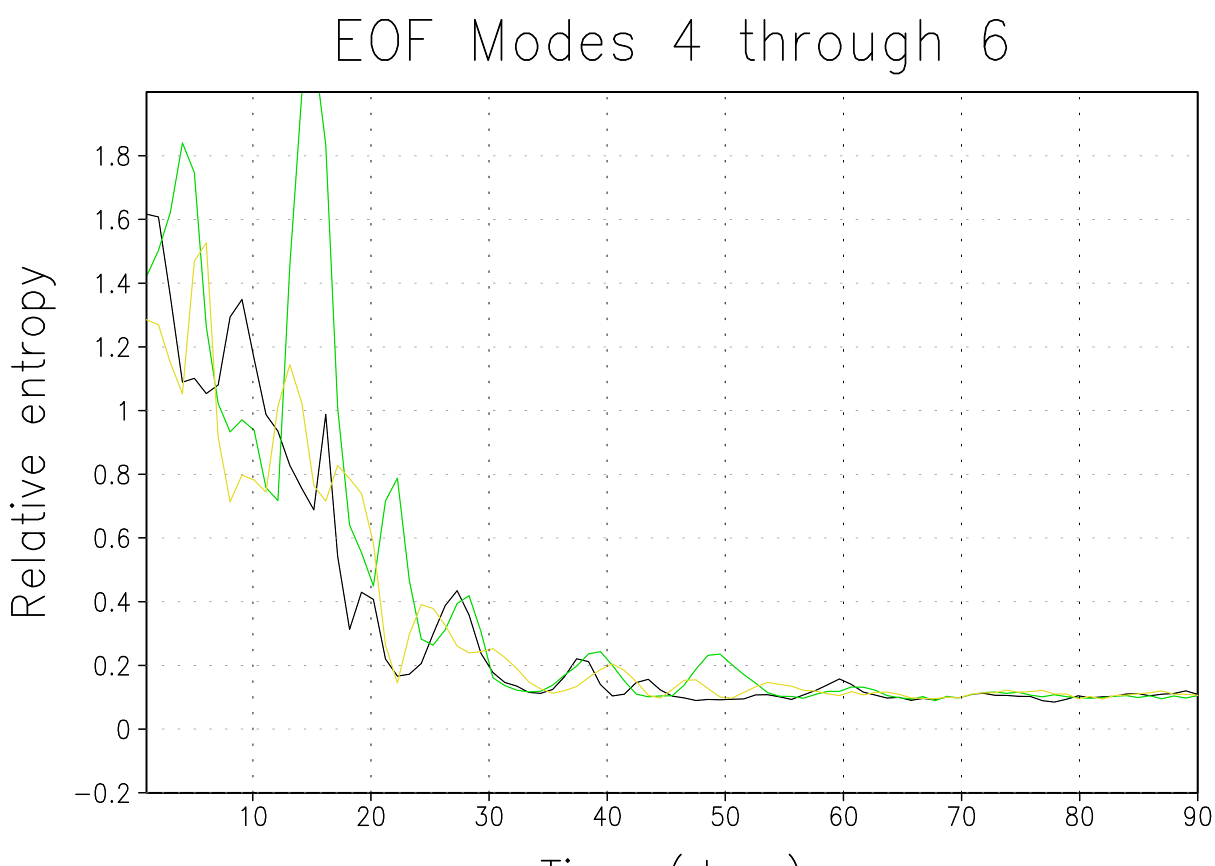
- Partitions were chosen so there were equal numbers of prediction ensemble members within each seperate EOF partitioning.

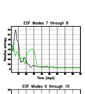
- All possible combinations of the ten EOFs were considered in multivariate calculations and the result shown is the average per combination.

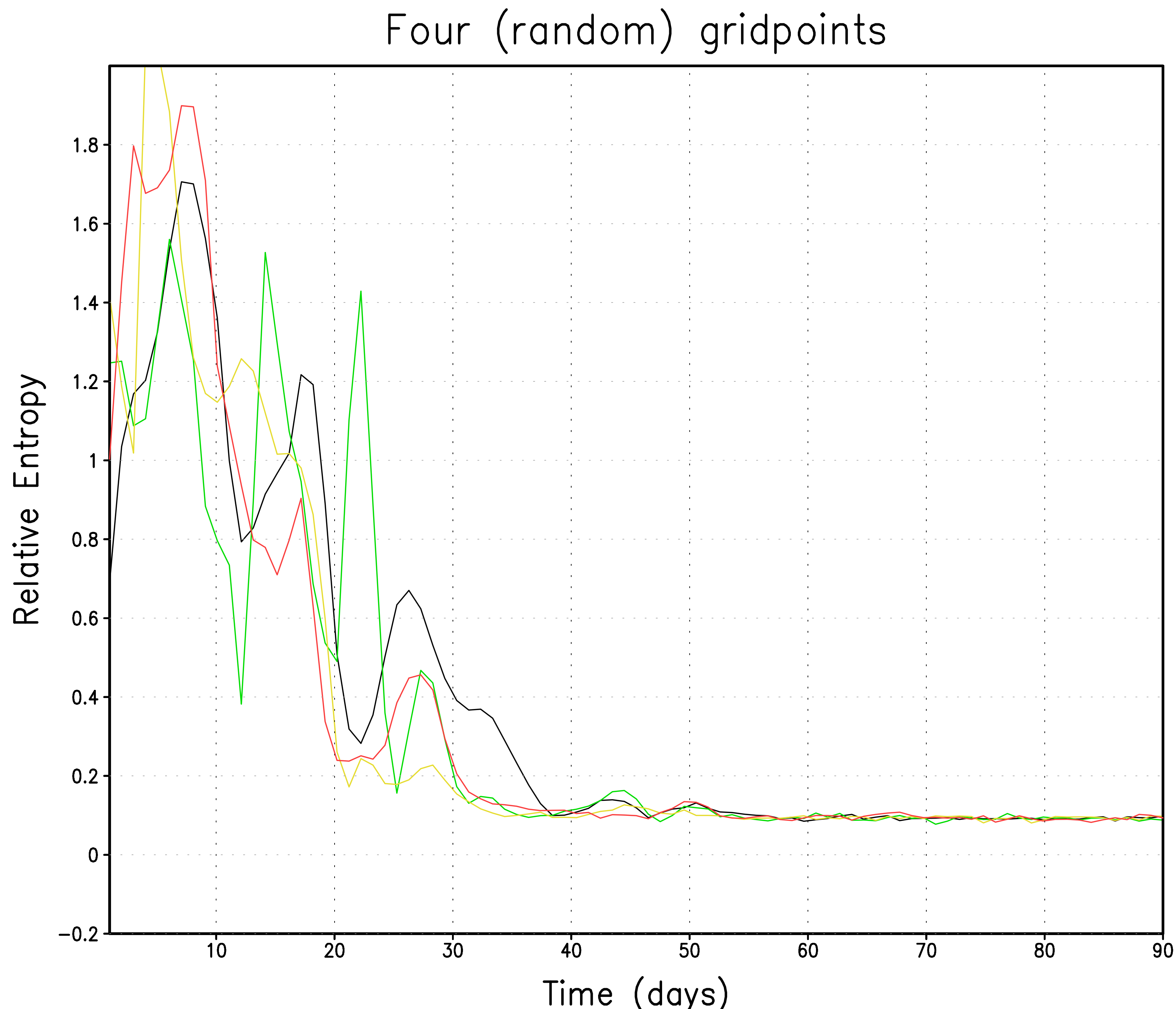


Time Evolution of Relative Entropy for North Atlantic/North American Sector
Relative Entropy
Time (days)
Univariate
Bivariate
Trivariate
Quadvariate
Quintvariate
0
0.5
1
1.5
2
2.5
0
10
20
30
40
50
60
70
80
90


EOF Modes 1 through 3
Relative entropy
Time (days)
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
10
20
30
40
50
60
70
80
90


EOF Modes 4 through 6
Relative entropy
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
10
20
30
40
50
60
70
80
90





Four (random) gridpoints
Relative Entropy
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
10
20
30
40
50
60
70
80
90
Time (days)

Information Flow

Undeveloped area still requiring considerable theoretical development. A complete formalism analogous to fluid flow would be desirable.

Two possible measures suggested in the literature:

Time Delayed Mutual Information

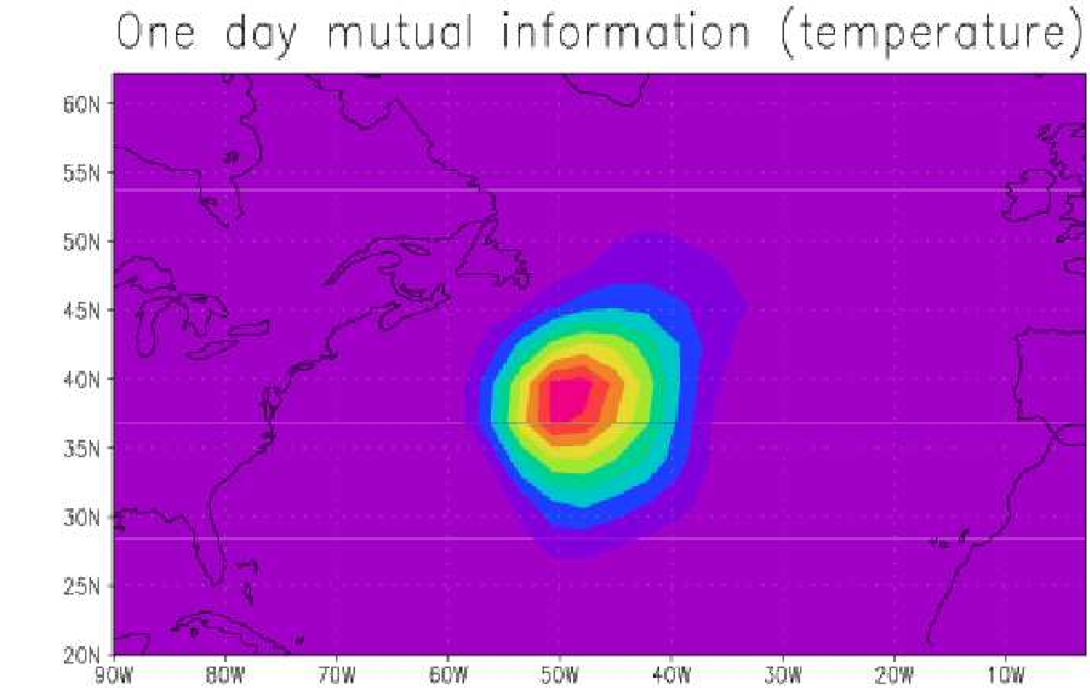
$$I(x(t), y(t_0))$$

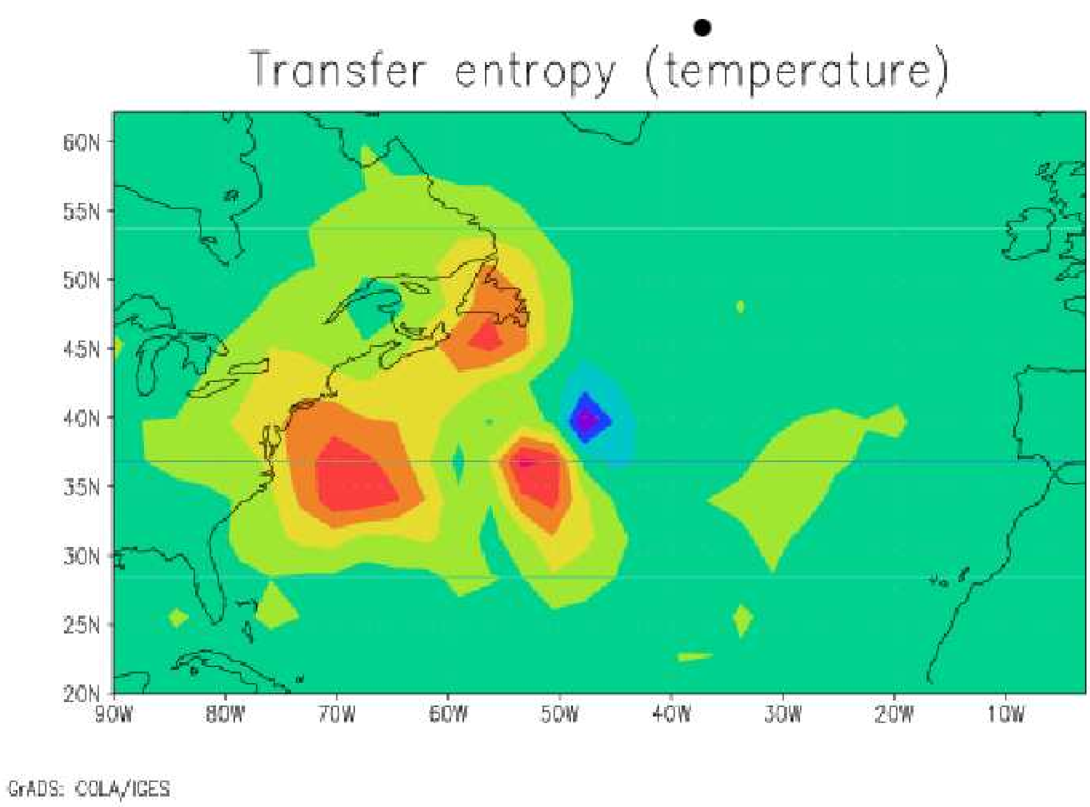
$$I(A, B) = H(A) - H(A|B)$$

Transfer Entropy

$$T(y \to x, t, t_0) \equiv$$

$$\sum_{y(t_0)} p(y(t_0)) D(p(x(t)|x(t_0), y(t_0)), p(x(t)|x(t_0)))$$


One day mutual information (temperature)
60N
55N
50N
45N
40N
35N
30N
25N
20N
90W
80W
70W
60W
50W
40W
30W
20W
10W


Transfer entropy (temperature)
60N
55N
50N
45N
40N
35N
30N
25N
20N
90W
80W
70W
60W
50W
40W
30W
20W
10W
GrADS: COLA/IGES

Conclusions

- The degree of "difference" between the prediction and climatological distributions/ensembles can be viewed as measuring the usefulness of a prediction. It can be quantified in a universal manner using information theory.

- In a realistic atmospheric model, experiments with large ensembles show that the overall mid-latitude predictability declines more or less linearly to be close to zero around 40-50 days. This result is reasonably robust to the definition of predictability. There is evidence of significant useful predictibility in the 10-20 day range but this may be resolution dependent.

- Predictability of individual EOFs or gridpoints does not show this approximately linear decline. This is due to the phenomenon of information flow.

- Information flow measures suggested in the physics literature were examined. One can interpret these as measuring the usefulness of earlier observations to prediction. Results show that information flows to particular locations from quite well defined upstream locations suggesting that this methodology could be used to improve the observational network.