Targeted observation strategies for prediction improvement in non-linear and/or non-Gaussian regimes.

R. Kleeman[1]

[1]Courant Institute of Mathematical Sciences, New York University

LMS Durham Symposium on Mathematics of Data Assimilation, August 2011

Talk Outline

Talk Outline

Statement of problem

- Uncertainty in important predictions is always unavoidable but potentially reducible.
- The origin of this uncertainty can be due to initial condition uncertainty.
- The uncertainty of initial conditions at a particular location induces via dynamical effects an uncertainty in a prediction at another location at a later time.
- Detailed knowledge of this "flow" enables one to determine strategies to reduce **appropriately** initial condition errors.

Statement of problem

- Uncertainty in important predictions is always unavoidable but potentially reducible.
- The origin of this uncertainty can be due to initial condition uncertainty.
- The uncertainty of initial conditions at a particular location induces via dynamical effects an uncertainty in a prediction at another location at a later time.
- Detailed knowledge of this "flow" enables one to determine strategies to reduce **appropriately** initial condition errors.

Statement of problem

- Uncertainty in important predictions is always unavoidable but potentially reducible.
- The origin of this uncertainty can be due to initial condition uncertainty.
- The uncertainty of initial conditions at a particular location induces via dynamical effects an uncertainty in a prediction at another location at a later time.
- Detailed knowledge of this "flow" enables one to determine strategies to reduce **appropriately** initial condition errors.

Statement of problem

- Uncertainty in important predictions is always unavoidable but potentially reducible.
- The origin of this uncertainty can be due to initial condition uncertainty.
- The uncertainty of initial conditions at a particular location induces via dynamical effects an uncertainty in a prediction at another location at a later time.
- Detailed knowledge of this "flow" enables one to determine strategies to reduce **appropriately** initial condition errors.

Approaches to problem

- Historically perhaps the first approach to this in an atmospheric context was due to the ECMWF group in the late 1990s who linearized predictions about a given "best" trajectory. The assumption then of Gaussian prediction distributions allows for an exact analytical solution which we review later.
- A different approach was later suggested by Bishop and co-workers from NCEP. They used an ensemble Kalman filter to assimilate potential additional observations in order to determine the best approach to reduce a particular prediction variance. The nature of a Kalman filter assumes both Gaussian prediction distributions as well as a linear dynamical system. The first is assumed in the Bayesian incorporation of observations while both are assumed in translating the error covariance forward in time.

Approaches to problem

- Historically perhaps the first approach to this in an atmospheric context was due to the ECMWF group in the late 1990s who linearized predictions about a given "best" trajectory. The assumption then of Gaussian prediction distributions allows for an exact analytical solution which we review later.
- A different approach was later suggested by Bishop and co-workers from NCEP. They used an ensemble Kalman filter to assimilate potential additional observations in order to determine the best approach to reduce a particular prediction variance. The nature of a Kalman filter assumes both Gaussian prediction distributions as well as a linear dynamical system. The first is assumed in the Bayesian incorporation of observations while both are assumed in translating the error covariance forward in time.

Approaches to problem

- Many generalizations of the Kalman filter have been proposed in the last decade or two to overcome the two limiting assumptions mentioned and are the subject of this workshop and participants and co-workers have made important contributions here. Important related work by Eyink, Anderson, Bocquet, Snyder and Hunt on particle filters, Gaussian mixtures and local filters should also be mentioned.
- In this talk we consider instead a direct approach to the targeting problem which relies on information theory and large ensembles.
- The key point conceptually is to define the flow of uncertainty from initial condition to prediction random variables. This needs to be done in such a way as to ensure that it can be calculated with practical ensembles. Given the computational expense of large ensembles it is important then to intercompare results with existing restrictive methods.

Approaches to problem

- Many generalizations of the Kalman filter have been proposed in the last decade or two to overcome the two limiting assumptions mentioned and are the subject of this workshop and participants and co-workers have made important contributions here. Important related work by Eyink, Anderson, Bocquet, Snyder and Hunt on particle filters, Gaussian mixtures and local filters should also be mentioned.
- In this talk we consider instead a direct approach to the targeting problem which relies on information theory and large ensembles.
- The key point conceptually is to define the flow of uncertainty from initial condition to prediction random variables. This needs to be done in such a way as to ensure that it can be calculated with practical ensembles. Given the computational expense of large ensembles it is important then to intercompare results with existing restrictive methods.

Approaches to problem

- Many generalizations of the Kalman filter have been proposed in the last decade or two to overcome the two limiting assumptions mentioned and are the subject of this workshop and participants and co-workers have made important contributions here. Important related work by Eyink, Anderson, Bocquet, Snyder and Hunt on particle filters, Gaussian mixtures and local filters should also be mentioned.
- In this talk we consider instead a direct approach to the targeting problem which relies on information theory and large ensembles.
- The key point conceptually is to define the flow of uncertainty from initial condition to prediction random variables. This needs to be done in such a way as to ensure that it can be calculated with practical ensembles. Given the computational expense of large ensembles it is important then to intercompare results with existing restrictive methods.

Talk Outline

Flow measures

- Three different measures have been proposed in the mathematical physics literature. The Time Lagged Mutual Information of Kaneko; the Transfer Entropy of Schreiber and the information flow of Liang. The first two derive from direct probabilistic arguments concerning random variables while the third is derived from a fundamental entropy evolution equation for the dynamical system.
- Presently the first method is the most practical in large dynamical systems. Moreover it has a direct and transparent interpretation in terms of the targeting problem.....

Flow measures

- Three different measures have been proposed in the mathematical physics literature. The Time Lagged Mutual Information of Kaneko; the Transfer Entropy of Schreiber and the information flow of Liang. The first two derive from direct probabilistic arguments concerning random variables while the third is derived from a fundamental entropy evolution equation for the dynamical system.
- Presently the first method is the most practical in large dynamical systems. Moreover it has a direct and transparent interpretation in terms of the targeting problem.....

Flow Measures

Time lagged mutual information

$$
\begin{aligned}
TLMI &= I(X(t); Y(t+\tau)) \\
&= \int\int p(x(t), y(t+\tau)) \log\left(\frac{p(x(t), y(t+\tau))}{p(x(t))p(y(t+\tau))}\right) dxdy \\
&= h(Y(t+\tau)) - h(Y(t+\tau)|X(t))
\end{aligned}
$$

- where h is the entropy functional and also its conditional cousin. Thus the TLMI is the reduction of uncertainty in prediction random variable $Y(t+\tau)$ due to perfect knowledge of initial condition variable $X(t)$.

Flow Measures

Time lagged mutual information

$$
\begin{aligned}
TLMI &= I(X(t); Y(t+\tau)) \\
&= \int\int p(x(t), y(t+\tau)) \log\left(\frac{p(x(t), y(t+\tau))}{p(x(t))p(y(t+\tau))}\right) dxdy \\
&= h(Y(t+\tau)) - h(Y(t+\tau)|X(t))
\end{aligned}
$$

- where h is the entropy functional and also its conditional cousin. Thus the TLMI is the reduction of uncertainty in prediction random variable $Y(t+\tau)$ due to perfect knowledge of initial condition variable $X(t)$.

Flow measures (practicalities)

Note that the practical calculation of the TLMI measure is feasible with ensembles since it is only a bivariate functional rather than a higher order multivariate functional subject to the curse of dimensionality. This latter issue affects the other two measures of uncertainty flow. TLMI is also invariant under non-linear non-degenerate state variable transformations since it is the relative entropy of two distributions (the full joint distribution and the hypothetical independent joint distribution).

Talk Outline

Model and initialization method

- A T42 L5 primitive equation dry dynamical core model with realistic depiction of mid-latitude jets and storms was utilized. It has realistic orography and relaxes temperatures toward a meridionally varying but zonally uniform state meant to represent radiative equilibrium. The model was configured for the northern winter
- The initial conditions were taken to follow a Gaussian distribution. It had uniform variances two orders of magnitude smaller than climatology. Covariances with a decay scale of $1000km$ horizontally were assumed and no vertical covariance was assumed. This was intended to crudely mimic a uniformly coarse observing network.
- Ensembles were constucted using this Gaussian distribution and a filter due to Lynch used to reduce gravity wave imbalances. The prediction variable analyzed was located in the mid-Atlantic in mid latitudes. The TLMI with respect to the full set of initial condition variables was then evaluated. Results stabilized statistically at between 250 and 500 ensemble members.

Model and initialization method

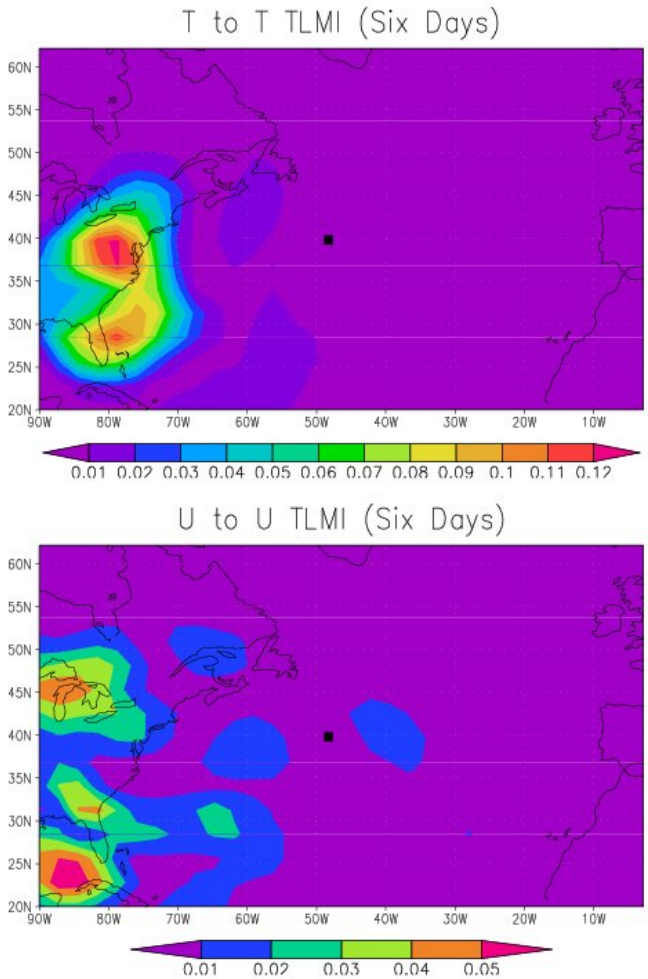
- A T42 L5 primitive equation dry dynamical core model with realistic depiction of mid-latitude jets and storms was utilized. It has realistic orography and relaxes temperatures toward a meridionally varying but zonally uniform state meant to represent radiative equilibrium. The model was configured for the northern winter
- The initial conditions were taken to follow a Gaussian distribution. It had uniform variances two orders of magnitude smaller than climatology. Covariances with a decay scale of $1000km$ horizontally were assumed and no vertical covariance was assumed. This was intended to crudely mimic a uniformly coarse observing network.
- Ensembles were constucted using this Gaussian distribution and a filter due to Lynch used to reduce gravity wave imbalances. The prediction variable analyzed was located in the mid-Atlantic in mid latitudes. The TLMI with respect to the full set of initial condition variables was then evaluated. Results stabilized statistically at between 250 and 500 ensemble members.

Model and initialization method

- A T42 L5 primitive equation dry dynamical core model with realistic depiction of mid-latitude jets and storms was utilized. It has realistic orography and relaxes temperatures toward a meridionally varying but zonally uniform state meant to represent radiative equilibrium. The model was configured for the northern winter
- The initial conditions were taken to follow a Gaussian distribution. It had uniform variances two orders of magnitude smaller than climatology. Covariances with a decay scale of $1000km$ horizontally were assumed and no vertical covariance was assumed. This was intended to crudely mimic a uniformly coarse observing network.
- Ensembles were constucted using this Gaussian distribution and a filter due to Lynch used to reduce gravity wave imbalances. The prediction variable analyzed was located in the mid-Atlantic in mid latitudes. The TLMI with respect to the full set of initial condition variables was then evaluated. Results stabilized statistically at between 250 and 500 ensemble members.

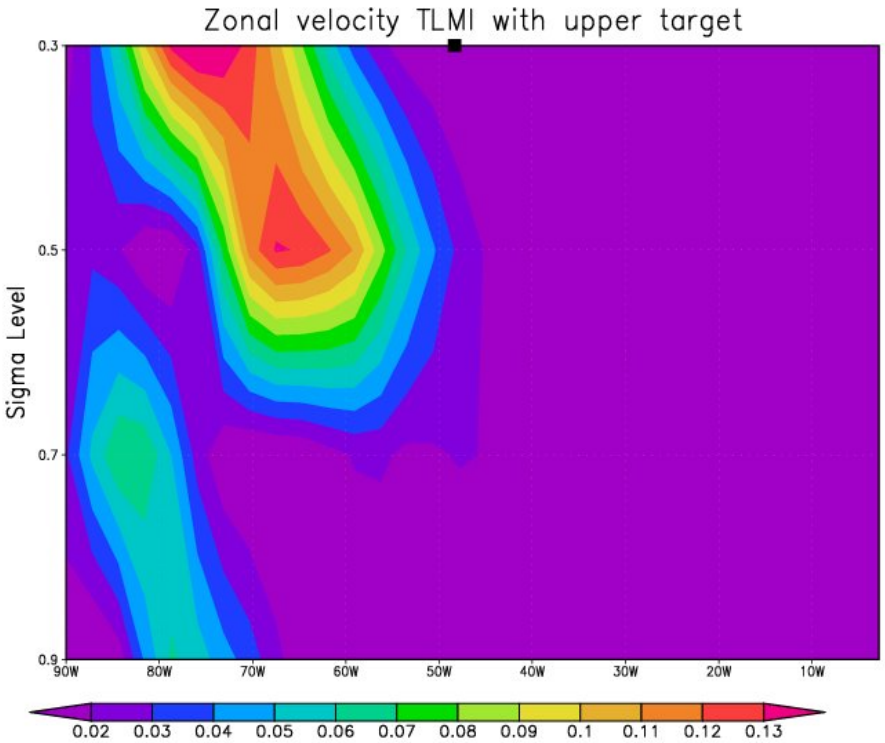
Talk Outline

Model Results Horizontal



Results are shown for a 6 day surface prediction which was well into the non-linear regime of predictions. The prediction and target initial conditions were taken to be both temperature (top) and zonal wind (bottom). Note the limited compact region to the west of the prediction optimization region.

Model Results Vertical



Results are shown for a 1 day tropopause zonal velocity prediction and zonal velocity initial conditions at various heights. Shown is a vertical section along a latitude of 40°N. Note the strong upward vertical propagation of uncertainty.

Talk Outline

Linearization formalism

- We follow the tensor formalism of Palmer (1998). Here the linear propagator matrix/tensor L takes random variables forward in time:

Propagator equation

$$e^i(t) = L^i_j e^j(0).$$

- The sensitivity vector measures the sensitivity of a prediction variable scalar cost function J to changes in the initial condition variables and is given by.....

Linearization formalism

- We follow the tensor formalism of Palmer (1998). Here the linear propagator matrix/tensor L takes random variables forward in time:

Propagator equation

$$e^i(t) = L^i_j e^j(0).$$

- The sensitivity vector measures the sensitivity of a prediction variable scalar cost function J to changes in the initial condition variables and is given by.....

Linearization formalism

- We follow the tensor formalism of Palmer (1998). Here the linear propagator matrix/tensor L takes random variables forward in time:

Propagator equation

$$e^i(t) = L^i_j e^j(0).$$

- The sensitivity vector measures the sensitivity of a prediction variable scalar cost function J to changes in the initial condition variables and is given by.....

Linearization formalism

Sensitivity vector

$$s^i = g^{ij} (L^*)^k_j \nabla_k J$$

- Where g is the metric tensor used to define the problem inner product and raise and lower tensor indices; L^* is the adjoint tensor of L with respect to g and $\nabla_k J \equiv D_k$ is the gradient vector of the cost function in the direction specified by k. Singular vectors and values are the eigenvectors and eigenvalues of the matrix

Singular vector matrix

$$S^i_j = (L^*)^i_k L^k_j$$

Linearization formalism

Sensitivity vector

$$s^i = g^{ij} \left(L^*\right)_j^k \nabla_k J$$

- Where g is the metric tensor used to define the problem inner product and raise and lower tensor indices; L^* is the adjoint tensor of L with respect to g and $\nabla_k J \equiv D_k$ is the gradient vector of the cost function in the direction specified by k. Singular vectors and values are the eigenvectors and eigenvalues of the matrix

Singular vector matrix

$$S_j^i = \left(L^*\right)_k^i L_j^k$$

Linearization formalism

Sensitivity vector

$$s^i = g^{ij} \left(L^*\right)_j^k \nabla_k J$$

- Where g is the metric tensor used to define the problem inner product and raise and lower tensor indices; L^* is the adjoint tensor of L with respect to g and $\nabla_k J \equiv D_k$ is the gradient vector of the cost function in the direction specified by k. Singular vectors and values are the eigenvectors and eigenvalues of the matrix

Singular vector matrix

$$S_j^i = \left(L^*\right)_k^i L_j^k$$

Linearization formalism

- If we write the gradient vector as a linear combination of the final times singular vectors (the singular vectors with L applied) it is easy to prove that

Sensitivity vector in terms of singular vectors

$$s^i = \lambda(m) u^{im} f_m$$

- where the $\lambda(m)$ are the singular values; u^{im} the components of the singular vectors and f_m the coefficients for D_k in terms of the final singular vectors. It is clear that if the singular value spectrum is strongly peaked as it often is then the sensitivity vector to initial conditions resembles the dominant singular vectors.

Linearization formalism

- If we write the gradient vector as a linear combination of the final times singular vectors (the singular vectors with L applied) it is easy to prove that

Sensitivity vector in terms of singular vectors

$$s^i = \lambda(m) u^{im} f_m$$

- where the $\lambda(m)$ are the singular values; u^{im} the components of the singular vectors and f_m the coefficients for D_k in terms of the final singular vectors. It is clear that if the singular value spectrum is strongly peaked as it often is then the sensitivity vector to initial conditions resembles the dominant singular vectors.

Linearization formalism

- If we write the gradient vector as a linear combination of the final times singular vectors (the singular vectors with L applied) it is easy to prove that

Sensitivity vector in terms of singular vectors

$$s^i = \lambda(m) u^{im} f_m$$

- where the $\lambda(m)$ are the singular values; u^{im} the components of the singular vectors and f_m the coefficients for D_k in terms of the final singular vectors. It is clear that if the singular value spectrum is strongly peaked as it often is then the sensitivity vector to initial conditions resembles the dominant singular vectors.

Linearization formalism

- If we assume that the initial condition variables are Gaussian random variables then, as argued plausibly by Palmer et. al., the natural metric tensor/matrix is the initial condition covariance matrix. Furthermore suppose we select the cost function J to be the squared magnitude of a particular prediction variable then one can also easily show that the sensitivity vector is then just the time lagged covariance matrix:

Sensitivity vector as a lagged covariance

$$s^i(I) = \left\langle e^i(0), e^I(t) \right\rangle$$

- where the index I refers to the cost function selection of a particular prediction variable.

Linearization formalism

- If we assume that the initial condition variables are Gaussian random variables then, as argued plausibly by Palmer et. al., the natural metric tensor/matrix is the initial condition covariance matrix. Furthermore suppose we select the cost function J to be the squared magnitude of a particular prediction variable then one can also easily show that the sensitivity vector is then just the time lagged covariance matrix:

Sensitivity vector as a lagged covariance

$$s^i(I) = \left\langle e^i(0), e^I(t) \right\rangle$$

- where the index I refers to the cost function selection of a particular prediction variable.

Linearization formalism

- If we assume that the initial condition variables are Gaussian random variables then, as argued plausibly by Palmer et. al., the natural metric tensor/matrix is the initial condition covariance matrix. Furthermore suppose we select the cost function J to be the squared magnitude of a particular prediction variable then one can also easily show that the sensitivity vector is then just the time lagged covariance matrix:

Sensitivity vector as a lagged covariance

$$s^i(I) = \left\langle e^i(0), e^I(t) \right\rangle$$

- where the index I refers to the cost function selection of a particular prediction variable.

Relationship to TLMI

- In the case of Gaussian distributions the mutual information can be calculated exactly as

Gaussian TLMI

$$I(e^i(0); e^l(t)) = \log\left(\frac{1}{1-\left(r^{il}(0,t)\right)^2}\right)$$

- where r^{il} is the correlation matrix. This is simply a rescaled covariance matrix so the functional connection to the usual linear sensitivity vector is clear. Note also that if we are dealing with a linear dynamical system with Gaussian initial conditions then the prediction variables are also Gaussian thus the equation above is exactly the TLMI in the linear Gaussian formalism. We therefore use this expression to intercompare methods.

Relationship to TLMI

- In the case of Gaussian distributions the mutual information can be calculated exactly as

Gaussian TLMI

$$I(e^i(0); e^l(t)) = \log\left(\frac{1}{1-\left(r^{il}(0,t)\right)^2}\right)$$

- where r^{il} is the correlation matrix. This is simply a rescaled covariance matrix so the functional connection to the usual linear sensitivity vector is clear. Note also that if we are dealing with a linear dynamical system with Gaussian initial conditions then the prediction variables are also Gaussian thus the equation above is exactly the TLMI in the linear Gaussian formalism. We therefore use this expression to intercompare methods.

Relationship to TLMI

- In the case of Gaussian distributions the mutual information can be calculated exactly as

Gaussian TLMI

$$I(e^i(0); e^l(t)) = \log\left(\frac{1}{1-\left(r^{il}(0,t)\right)^2}\right)$$

- where r^{il} is the correlation matrix. This is simply a rescaled covariance matrix so the functional connection to the usual linear sensitivity vector is clear. Note also that if we are dealing with a linear dynamical system with Gaussian initial conditions then the prediction variables are also Gaussian thus the equation above is exactly the TLMI in the linear Gaussian formalism. We therefore use this expression to intercompare methods.

Talk Outline

Methods intercomparison

- Using the standard version of this model we compared the linearized sensitivity analysis just discussed with the fully non-linear ensemble method discussed earlier. Note that the linear Gaussian TLMI does not compute the lagged correlation of initial condition and prediction ensembles. Instead it uses the linearized propagator matrix to project forward in time the initial condition covariance matrix and then takes the lagged correlation. These two things can be very different.
- Initial condition ensembles were constructed using a Gaussian distribution aligned within the local attractor with a homogenous variance of 0.1 (the attractor has dimensions of order 10.0 in all directions). Ensembles of size 10^5 were utilized and a tangent linear version of the model used for projecting forward in time the initial covariance matrix for the linear method. The full non-linear ensemble was used for the non-Gaussian TLMI calculation.

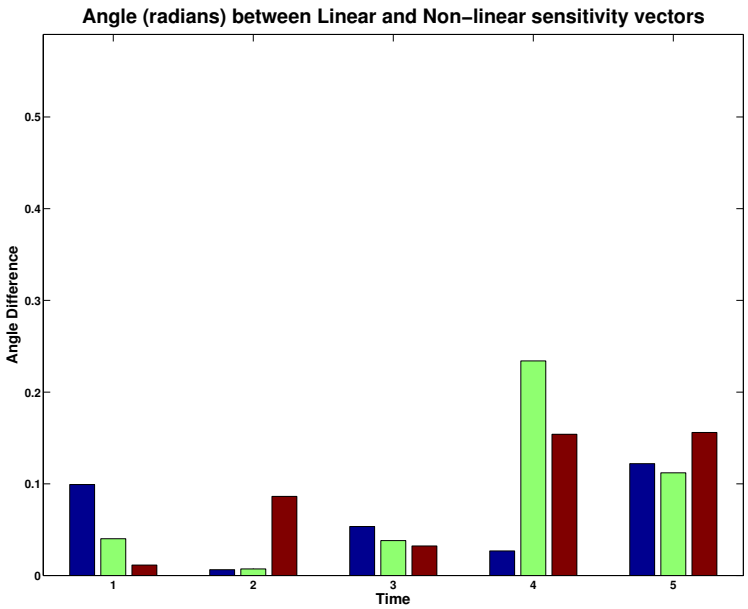
Methods intercomparison

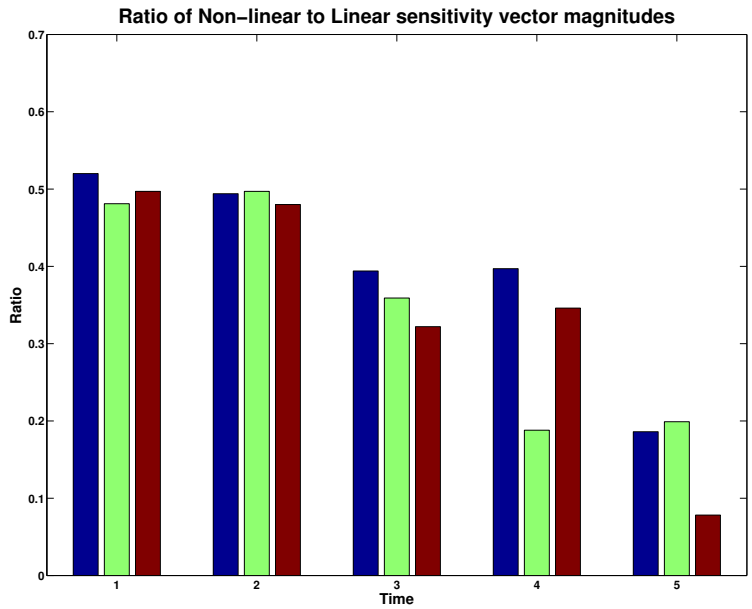
- Using the standard version of this model we compared the linearized sensitivity analysis just discussed with the fully non-linear ensemble method discussed earlier. Note that the linear Gaussian TLMI does not compute the lagged correlation of initial condition and prediction ensembles. Instead it uses the linearized propagator matrix to project forward in time the initial condition covariance matrix and then takes the lagged correlation. These two things can be very different.
- Initial condition ensembles were constructed using a Gaussian distribution aligned within the local attractor with a homogenous variance of 0.1 (the attractor has dimensions of order 10.0 in all directions). Ensembles of size 10^5 were utilized and a tangent linear version of the model used for projecting forward in time the initial covariance matrix for the linear method. The full non-linear ensemble was used for the non-Gaussian TLMI calculation.

Methods intercomparison

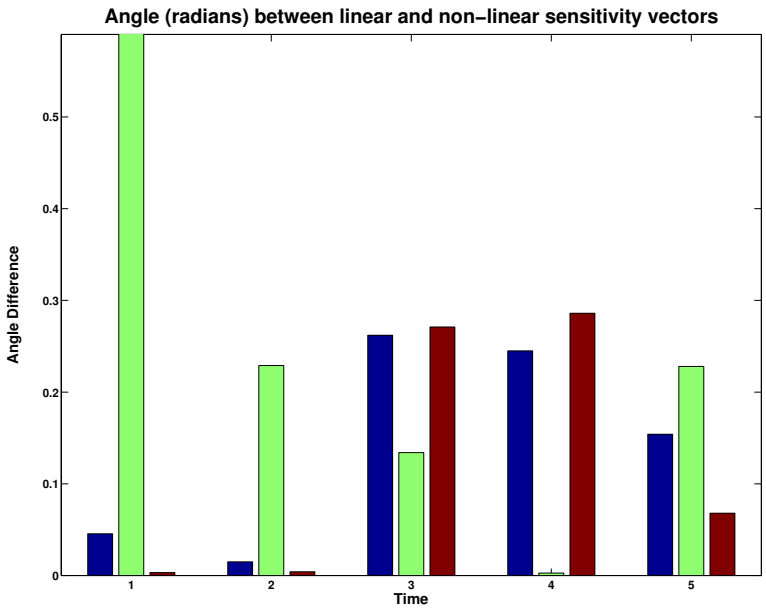
- The TLMI is a 3×3 matrix of non-negative entries for both linear and non-linear cases with each column corresponding to a different prediction optimization variable. The angle between vectors from the two methods was calculated as well as the ratio of their magnitudes. Two random vectors with non-negative entries have an average angle of 0.59 radians (assuming entries drawn uniformly and independently) so we use that level as a benchmark to evaluate differences. As a first look two initial conditions were taken at random from the attractor and results are shown. Prediction times were taken well into the non-linear regime out to around half way toward relaxation of the ensemble to equilibrium.

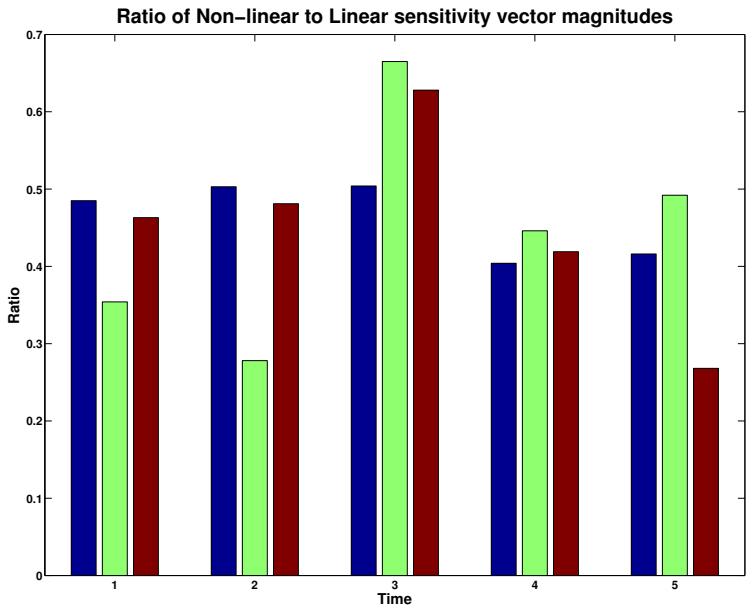
Methods intercomparison





Methods intercomparison





For Further Reading I

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

R. Kleeman.
Information theory and dynamical system predictability.
Entropy. 13:612–649, 2011.