Technical Report - Ensemble Prediction System (EPS)

E. Ramona Stefanescu

Department of Mechanical Engineering, University of Buffalo,

ers32@buffalo.edu, http://www.acsu.buffalo.edu/~ers32/

1 Review

1.1 Singular vectors

[2] indicated that perturbation growth in realistic models is related to the eigenvalues and eigenvectors of the operator product of the tangent forward linear and adjoint operators, i.e. to the *singular values* and *singular vectors* of the tangent forward operator. [3] (ECMWF implementation) identified singular vectors in a primitive equation model with a large number of degrees of freedom.

1.2 Linearized model equations

Let χ be the state vector of a generic autonomous system, whose evolution equations can be written as:

$$\frac{\partial \chi}{\partial t} = A(\chi) \tag{1.1}$$

$\chi(t)$ represents an integration of the above eq. from t_0 to t. The time evolution of a small perturbation x around the time evolving trajectory $\chi(t)$ can be described by the linearized model equations:

$$\frac{\partial x}{\partial t} = A_l(x) \tag{1.2}$$

where

$$A_l = \left.\frac{\partial A(x)}{\partial x}\right|_{\chi(t)} \tag{1.3}$$

is the tangent operator computed at the trajectory point $\chi(t)$. Let $L(t, t_0)$ be the integral forward propagator of the dynamical equation 1.2 linearized about a non-linear trajectory $\chi(t)$:

$$x(t) = L(t, t_0)x(t_0), \tag{1.4}$$

that maps a perturbation x at initial time t_0 to the optimization time t.

1.3 Inner product

Consider two perturbations x and y, and a positive Hermitian matrix E. Denote the inner product the $(\dots;\dots)_E$ the inner product between the two vectors

$$(x;y)_E = \langle x; Ey\rangle, \tag{1.5}$$

where $\langle \dots;\dots\rangle$ identifies the canonical Euclidean scalar product,

$$\langle x;y\rangle = \sum_{i=1}^{N} x_i y_i. \tag{1.6}$$

Denote by $\parallel \cdots \parallel_E$ the norm associated withe the inner product $(\dots;\dots)_E$,

$$\parallel x \parallel_E^2 = (x;x)_E = \langle x; Ex\rangle \tag{1.7}$$

1.4 Adjoint operator

Denote by L^{*E} the adjoint by L with respect to the inner product $(\dots;\dots)_E$,

$$(L^{*E}x;y)_E = (x;Ly)_E \tag{1.8}$$

The adjoint of L with respect to the inner product defined by E can be written in terms of the adjoint L^* defined with respect to the canonical Euclidean scalar product,

$$L^{*E} = E^{-1}L^*E. \tag{1.9}$$

1.5 Singular values and singular vectors

From the definitions above, it follows that the squared norm of a perturbation x at time t can be computed as

$$\parallel x(t) \parallel_E^2 = (x(t_0); L^{*E}Lx(t_0))_E. \tag{1.10}$$

From Eq. 1.5 it follows that the problem of finding the phase space directions x for which $\parallel x(t) \parallel_E^2 / \parallel x(t_0) \parallel_E^2$ is maximum can be reduced to the computation of the eigenvectors $v_I(t_0)$ with the largest eigenvalues σ_i^2, i.e. to the solution of the eigenvalue problem:

$$L^{*E}Lv_i(t_0) = \sigma_i^2 v_i(t_0). \tag{1.11}$$

The square roots of the eigenvalues, σ_i, are called the *singular value* and the eigenvectors $v_i(t_0)$ the (right) *singular vectors* of L with respect to the inner product E. The singular vectors with largest singular values identify the directions characterized by maximum growth. The time interval $t - t_0$ is called the optimization time interval.

At optimization time t, the i-th singular vector evolves into

$$v_i(t) = L(t,t_0)v_i(t_0), \tag{1.12}$$

a vector with the total norm equal to

$$\| v_i(t) \|_E^2 = \sigma_i^2. \tag{1.13}$$

Since any perturbation $x(t)/\| x(t)) \|_E$ can be written as a linear combination of the singular vectors $v_i(t)$, it follow that

$$max_{\| x(t_0) \|_E \neq 0} = \left(\frac{\| x(t) \|_E}{\| x(t_0) \|_E} \right) = \sigma_1, \tag{1.14}$$

which implies that maximum growth as measured by the norm $\| \cdots \|_E$ is associated with the dominant singular vector v_1.

1.6 Local projector operator

Denote by x_g the grid point representation of the state vector x, by S the spectral-to-grid point transformation operator, $x_g = Sx$, and by Gx_g the multiplication of the vector x_g, defined in grid point space, by the function $g(s) = \begin{cases} 1 & \forall s \in \Sigma, \\ 0 & \forall s \in \Sigma. \end{cases}$

where s defines the coordinate of a gird point, and Σ is a geographical region. Consider a generic vector x. The application of the local projection operator T defined as

$$T = S^{-1}GS, \tag{1.15}$$

to the vector x sets the vector x to zero for all grid points outside the geographical region Σ. These operators can be used to solve the following problem: find the perturbations with a) the fastest growth during the time interval $t - t_0$, (b) unitary E_0-norm at initial time and (c) maximum E-norm inside the geographical region Σ at optimization time. The solution to this problem is given by the singular values of the operator:

$$K = E^{-1/2}TLE_0^{-1/2} \tag{1.16}$$

2 Methodology

Each ensemble member evolution is given by integrating the following equation:

$$e_j(T) = e_0(d,0) + \underbrace{\frac{\partial e_j(d,0)}{\partial t}}_{\text{Initial uncertainty}} + \underbrace{\int_{t=0}^{T} \left[P_j(e_j,t) + \frac{\partial P_j(e_j,t)}{\partial t} + A_j(e_j,t) \right] dt}_{\text{Model uncertainty}} \tag{2.1}$$

where $e_j(d,0)$ is the initial condition, $P_j(e_j,t)$ represents the model tendency component due to parameterized physical process (turbulence, moist processes, orographic effect) as used for model j, $\frac{\partial P_j(e_j,t)}{\partial t}$ represents the random model errors (e.g. due to parametrized physical processes or sub-grid scale process - stochastic perturbation), and $A_j(e_j, t)$ is the r remaining tendency component (different physical parametrization or multi-model).

In the Meteorological Service of Canada (MSC) Monte Carlo approach, initial perturbations are generating by running separate data assimilation cycles:

$$e_j(0) = \Xi[e_j(\tau_1), o(\tau_1, \tau_2) + \delta o_j, P_j, A_j], \tag{2.2}$$

where (τ_1, τ_2) is the time spanned during each assimilation cycle, $o(\tau_1, \tau_2)$ and δo_j denote the vector of observations and corresponding random perturbations, and $\Xi[\ldots, \ldots, \ldots]$ denotes the data assimilation process. Note that each assimilation cycle depends on the model used in the assimilation.

In contrast, NCEP and ECMWF initial ensemble states are created by adding either bred or singular vectors $de_j(0)$ to the best estimate of the atmosphere at an initial time $e_0(0)$ that is produced by a high-resolution three- or four-dimensional data assimilation procedure:

$$\begin{aligned} e_0(0) &= \Xi[e_0(\tau_1), o(\tau_1, \tau_2), P_0, A_0], \\ e_j(0) &= e_0(0) + de_j(0). \end{aligned} \tag{2.3}$$

2.1 ECMWF

For each grid point $\mathbf{x} = (\lambda, \phi, \sigma)$ (identified by its latitude, longitude and vertical hybrid coordinate), the perturbed parametrized tendency (for each state vector component) is defined as:

$$\frac{\partial P_j(e_j, t)}{\partial t} = [1 + \langle r_j(\lambda, \phi, t)\rangle_{10,6}] P(e_j, t), \tag{2.4}$$

where P is the unperturbed diabetic tendency, and $\langle \ldots \rangle_{10,6}$ indicates that the same random number r_j is used inside a 10° box with a 6-h time window [1].

The initial perturbations $de_j(0)$ are defined as:

$$de_j(0) = \underline{\underline{A}} \cdot SV_{NH} + \underline{\underline{B}} \cdot SV_{SH} + \underline{\underline{A}} \cdot SV_{TC} \tag{2.5}$$

where for each geographical region (Northern and Southern Hemisphere (NH and SH), and Tropics (TC)) the coefficients of the linear combination matrices are set by comparing the singular vectors with analysis error estimates given by the ECMWF four-dimensional data assimilation (4DVAR).

2.2 NCEP

Formally, each member of the NCEP-EPS is defined by Eq. 2.1 with the same model version P being used for all members and with initial perturbations $de_j(0)$ defined as:

$$de_j(0) = \underline{\underline{RR}} \cdot BV_j. \tag{2.6}$$

The coefficients $\underline{\underline{RR}}$ of the linear combination matrices are defined by the regional rescaling algorithm. [4, 5]

NCEP produces 10 perturbed ensemble members both at 0000 and 1200 UTC every day out to 16-days lead time. For both cycles, the generation of the initial perturbations is done in five independently run breeding cycles, using the regional rescaling algorithm.

References

[1] R. Buizza, *Potential forecast skill of ensemble prediction and spread and skill distributions of the ecmwf ensemble prediction system*, Monthly Weather Review, 125 (1997), pp. 99–119.

[2] E. N. Lorenz, *A study of the predictability of a 28-variable atmospheric model*, Tellus, 17 (1965), pp. 321–333.

[3] F. Molteni, R. Buizza, T. N. Palmer, and T. Petroliagis, *The ecmwf ensemble prediction system: Methodology and validation*, Quarterly Journal of the Royal Meteorological Society, 122 (1996), pp. 73–119.

[4] Z. Toth and E. Kalnay, *Ensemble forecasting at nmc: The generation of perturbations*, Bulletin of the American Meteorological Society, 74 (1993), pp. 2317–2330.

[5] ——, *Ensemble forecasting at ncep and the breeding method*, Monthly Weather Review, 125 (1997), pp. 3297–3319.