4.2 ALGORITHM AND SENSITIVITY ANALYSIS OF INFORMATION-THEORETIC ENSEMBLE-BASED OBSERVATION TARGETING

Han-Lim Choi*, Jonathan P. How,
Massachusetts Institute of Technology, Cambridge, MA.
and James A. Hansen
Naval Research Laboratory, Monterey, CA.

1 INTRODUCTION

Complex characteristics of the weather dynamics such as being chaotic, uncertain, and of multiple time- and length-scales, leads to the necessity of a large sensor network. Expansion of the static observation network is limited by geographic aspects; thus, an adaptive sensor network incorporating mobile sensor platforms (e.g. UAVs) has become an attractive solution to construct effectively large networks. Palmer (1998) and Daescu (2004) located the potential error-growing sites based on the sensitivity information inherited in the dyanamics; meanwhile, Majumdar (2002) and Leutbecher (2003) quantified the future forecast covariance change within the framework of approximation of extended Kalman filter. However, due to the enormous size of the system – the network dimension of $\mathcal{O}(10^3)$ in the state dimension of $\mathcal{O}(10^6)$ (Majumdar, 2002), the selection strategy was very simple – for instance, two flight paths was *greedily* selected out of 49 predetermined paths in Majumdar (2002).

This work provides an information-theoretic way to perform the targeting of a sensor network in a large space when the goal is to reduce the uncertainty in a specified verification space/time. The dynamic sensor targeting problem is formulated as a static sensor selection problem associated with the ensemble-based filtering (Evensen, 1996; Whitaker, 2002). Mutual information, which can take into account the correlation amongst verification variables, is introduced as a measure of uncertainty reduction, and computed under the Gaussian assumption. To address the computational challenge resulting from the expense of determining the impact of each measurement choice on the uncertainty reduction in the verification site, the commutativity of mutual information is exploited. This enables the contribution of each measurement choice to be computed by propagating information *backwards* from the verification space/time to the search space/time. This backward computation significantly reduces the required number of ensemble updates that is computationally intensive. Numerical experiments using Lorenz-95 model validates the computational effectiveness of the proposed information-theoretic ensemble targeting algorithm.

In spite of limitation of ensemble size for a realistic weather model, there have no research on sensitivity of the targeting solution to the ensemble size. As an essential step toward implementation of the presented information-theoretic ensemble targeting algorithm to realistic cases, this paper performs sensitivity analysis of the ensemble-based targeting with respect to limited ensemble size. Monte-Carlo experiments with the Lorenz-95 targeting example characterizes three important aspects of sensitivity: discrepancy of the predicted and actual information gain value, performance degradation of the targeting solution, and inconsistency of the targeting solution with respect to the choice of ensemble. In addition, based on the statistical analysis of the entropy estimation, this work proposes new predictors of the degree of impact that limitation of ensemble size might have on the solution optimality: the range-to-noise ratio (RNR) and the probability of correct decision (PCD). Numerical investigation of the Lorenz-95 targeting example will validate the effectiveness of these predictors.

* *Corresponding Author Address:* Han-Lim Choi, Massachusetts Institute of Technology, Dept. of Aeronautics and Astronautics, Cambridge, MA 02139; e-mail:hanlimc@mit.edu

2 INFORMATION-THEORETIC SENSOR TARGETING

2.1 *Entropy and Mutual Information*

Entropy representing the amount of information hidden in a random variable A_1 is defined as

$$\mathcal{H}(A_1) = -E\left[\log\left(p_{A_1}(a_1)\right)\right]. \quad (1)$$

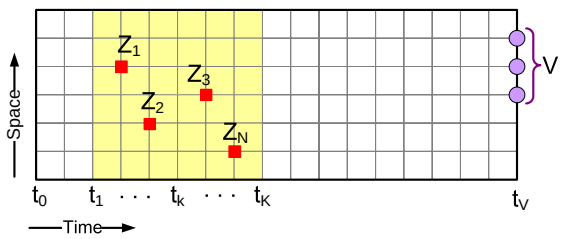

Figure 1: Multiple targeting as a sensor placement problem)

The joint entropy of two random variables $\mathcal{H}(A_1, A_2)$ is defined in a similar way; it is related to the conditional entropy $\mathcal{H}(A_2|A_1)$ as $\mathcal{H}(A_1, A_2) = \mathcal{H}(A_1) + \mathcal{H}(A_2|A_1)$.

The mutual information that is employed in this paper as a metric of the uncertainty reduction of the forecast by the measurement is defined as:

$$\mathcal{I}(A_1; A_2) = \mathcal{H}(A_1) - \mathcal{H}(A_1|A_2). \tag{2}$$

Note that the mutual information is commutative (Cover, 1991):

$$\mathcal{I}(A_1; A_2) \equiv \mathcal{I}(A_2; A_1) = \mathcal{H}(A_2) - \mathcal{H}(A_2|A_1). \tag{3}$$

Moreover, in case a random vector A is jointly Gaussian, its entropy is expressed as

$$\begin{aligned} \mathcal{H}(A) &\equiv \mathcal{H}(A_1, \cdots, A_k) \\ &= \tfrac{1}{2} \log\det(\mathrm{Cov}(A)) + \tfrac{k}{2} \log(2\pi e). \end{aligned} \tag{4}$$

This work computes the mutual information between the verification and measurement variables by exploiting the commutativity under the Gaussian assumption.

2.2 *Sensor Targeting Algorithm*

Figure 1 illustrates the sensor targeting problem in a spatial-temporal gridspace. The objective of sensor targeting is to deploy n_s sensors in the search space/time region (yellow) in order to reduce the uncertainty in the verification region (purple region) at the verification time t_V. Without loss of generality, it assumed that each grid point is associated with a single state variable that can be directly measured. Denote the state variable at location s as X_s, and the measurement of X_s as Z_s, both of which are random variables. Also, define $Z_\mathcal{S} \equiv \{Z_1, Z_2, \cdots, Z_N\}$ and $X_\mathcal{S} \equiv \{X_1, X_2, \cdots, X_N\}$ as the sets of all corresponding random variables over the entire search space of size N. Likewise, $\mathcal{V} \equiv \{V_1, V_2, \cdots, V_M\}$ denotes the set of random variables representing states in the verification region at t_V, with M being the size of verification space. With a slight abuse of notation, this paper does not distinguish a set of random variables from the random vector constituted by the corresponding random variables. Measurement is subject to Gaussian noise that is uncorrelated with any of the state variables

$$Z_s = X_s + N_s;\ N_s \sim \mathcal{N}(0, R_s),\ \forall\ s \in \mathcal{S} \triangleq \mathbb{Z}_+ \cap [1, N],$$
$$\mathrm{Cov}(N_s, Y_p) = 0,\ \ \forall\ Y_p \in X_\mathcal{S} \cup \mathcal{V}.$$

When entropy is adopted as the uncertainty metric, the sensor targeting problem can be written as

$$\begin{aligned} \mathbf{s}_F{}^\star &= \underset{\mathbf{s}\subset\mathcal{S}:|\mathbf{s}|=n_s}{\arg\max}\ \mathcal{I}(\mathcal{V}; Z_\mathbf{s}) \triangleq \mathcal{H}(\mathcal{V}) - \mathcal{H}(\mathcal{V}|Z_\mathbf{s}) \\ &\overset{G}{=} \underset{\mathbf{s}\subset\mathcal{S}:|\mathbf{s}|=n_s}{\arg\max}\ \tfrac{1}{2}\,\mathrm{ldet}\,(\mathrm{Cov}(\mathcal{V})) - \tfrac{1}{2}\,\mathrm{ldet}\,(\mathrm{Cov}(\mathcal{V}|Z_\mathbf{s})) \end{aligned} \tag{5}$$

where $\mathrm{ldet}()$ stands for the $\log\det()$ function, and $\overset{G}{=}$ denotes the equality under the Gaussian assumption. Calling the above formulation as *forward* selection, Choi (2007a) verified that the computation time for the selection decision can be dramatically reduced by reformulating it as a *backward* selection:

$$\begin{aligned} \mathbf{s}_B{}^\star &= \underset{\mathbf{s}\subset\mathcal{S}:|\mathbf{s}|=n_s}{\arg\max}\ \mathcal{I}(Z_\mathbf{s}; \mathcal{V}) \triangleq \mathcal{H}(Z_\mathbf{s}) - \mathcal{H}(Z_\mathbf{s}|\mathcal{V}) \\ &\overset{G}{=} \underset{\mathbf{s}\subset\mathcal{S}:|\mathbf{s}|=n_s}{\arg\max}\ \tfrac{1}{2}\,\mathrm{ldet}\,(\mathrm{Cov}(Z_\mathbf{s})) - \tfrac{1}{2}\,\mathrm{ldet}\,(\mathrm{Cov}(Z_\mathbf{s}|\mathcal{V})) \\ &= \underset{\mathbf{s}\subset\mathcal{S}:|\mathbf{s}|=n_s}{\arg\max}\ \tfrac{1}{2}\,\mathrm{ldet}\,(\mathrm{Cov}(X_\mathbf{s}) + R_\mathbf{s}) \\ &\qquad - \tfrac{1}{2}\,\mathrm{ldet}\,(\mathrm{Cov}(X_\mathbf{s}|\mathcal{V}) + R_\mathbf{s})\,. \end{aligned} \tag{6}$$

Note that $\mathbf{s_F}^\star \equiv \mathbf{s_B}^\star$, since mutual information is commutative.

In finding the optimal solution using explicit enumeration, which results in the worst-case computational complexity, the forward method computes the posterior entropy of $\mathcal{V}$ for every possible measurement candidate $\mathbf{s}$. This means that the evaluation and determinant calculation of the posterior covariance of $\mathcal{V}$ has to be done $\binom{N}{n_s}$ times. On the other hand, the covariance update can be done as a constant-time process for the backward method by updating the covariance of the entire search space $\mathcal{S}$ at once; then, a combinatorial search extracts the pair of $n_s \times n_s$ submatrices consisting of identical row/column indices from the prior and posterior covariance matrices of $\mathcal{S}$, that reveal the largest dif-

ference between their determinants. As the covariance update usually incurs high computational expense – especially for the ensemble-based targeting problems – the backward selection saves a significant amount of computation time. In addition, the backward formulation is preferable if $|\mathcal{V}| > n_s$, since it computes the determinants of smaller matrices.

3 ENSEMBLE-BASED TARGETING

3.1 *Ensemble Square Root Filter*

This work utilizes sequential ensemble square root filters (EnSRF) (Whitaker, 2002) for both data assimilation and adaptive targeting, as it represents better computational stability and effectiveness than other ensemble filters such as the ensemble Kalman filter (EnKF) and the ensemble transform Kalman filter (ETKF). In EnSRF, the state estimate and the estimation error covariance are represented by the ensemble mean and perturbation ensemble variance, respectively. The perturbation ensemble being defined as $\tilde{\mathbf{X}} \equiv \eta\left(\mathbf{X} - \bar{\mathbf{x}} \times \mathbf{1}^T\right) \in \mathbb{R}^{L_S \times M}$, the error covariance is approximated as

$$\mathbf{P} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^T/(M-1), \tag{7}$$

where $\mathbf{X}$ is the ensemble matrix, $\bar{\mathbf{x}}$ is the ensemble mean vector, L_S and M denote the number of state and the ensemble members, and η is the inflation factor to avoid underestimation of the covariance. The propagation step for the EnSRF corresponds to the integration

$$\mathbf{X}^f(t+\Delta t) = \int_t^{t+\Delta t} \dot{\mathbf{X}} dt \tag{8}$$

with the initial condition $\mathbf{X}(t) = \mathbf{X}^a(t)$, superscripts 'f' and 'a' denote the *forecast* and the *analysis*, respectively. The measurement update step for the EnSRF consists of the mean update and the perturbation update as:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{K}(\mathbf{z} - \mathbf{H}\bar{\mathbf{x}}^f),\ \tilde{\mathbf{X}}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\tilde{\mathbf{X}}^f \tag{9}$$

where $\mathbf{z}$ and $\mathbf{H}$ are the measurement vector and the observation matrix, while $\mathbf{K}$ represents the Kalman gain determined by a nonlinear matrix equation of $\mathbf{X}$. In the sequential framework devised for efficient implementation, the ensemble update by the m-th observation is written as $\tilde{\mathbf{X}}^{m+1} = \tilde{\mathbf{X}}^m - \alpha_m \beta_m \mathbf{p}_i^m \tilde{\xi}_i^m$ with $\alpha_m = (1+\sqrt{\beta_m R_i})^{-1}, \beta_m = (\mathbf{P}_{ii}^m + R_i)^{-1}$, when i-th state is measured. $\mathbf{p}_i^m$, $\tilde{\xi}_i^m$, and $\mathbf{P}_{ii}^m$ are the i-th column, the i-th row, and (i,i) element of the prior perturbation matrix $\mathbf{P}^m$, respectively. α_m is the factor for compensating the mismatch between the serial update and the batch update, while $\beta_m \mathbf{p}_l^m$ is equivalent to the Kalman gain.

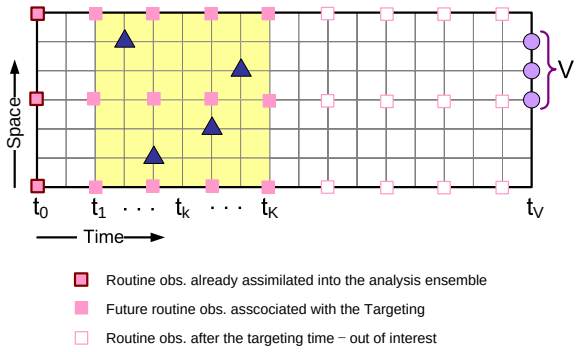

Figure 2: Observation structure over space-time

3.2 *Ensemble Augmentation for Prior Setup*

In the context of numerical weather prediction, the goal of observation targeting is to improve the weather forecast in the future time t_V by deploying observation networks over a short future time horizon $(t_1, \ldots, t_K)$, where the current time is t_0 at which the analysis ensemble is available. Also the routine observation network that periodically takes measurement until t_K should be taken into account for the targeting decision. Note that the routine observations from t_{K+1} through t_V are not considered, as what is of particular interest is the enhancement of $\Delta T(= t_V - t_K)$-interval forecast based on observations taken up until t_K. The observation structure will be similar to Fig. 2. Note that the targeting decision should not rely on the actual measurement values over $[t_1, t_K]$, since this information is not available at the decision time t_0. Thus, the ensemble-based targeting process tried to construct an observation network that is expected to give the greatest uncertainty reduction in the verification space-time in the *statistical* sense. EnSRF targeting determines where and when to make additional observations based only on the covariance information, not on the actual value of the future measurements.

The effect of the future routine observations is processed first and the output of this routine process is treated as a prior information for the sensor selection problem in which only the additional observations are involved. Processing the observation networks over the time amounts to a EnSRF ensemble update for the augmented forecast ensemble defined

as

$$\mathbf{X}^f_{aug} = \left[\left(\mathbf{X}^f_{t_1}\right)^T \quad \cdots \quad \left(\mathbf{X}^f_{t_K}\right)^T \quad \left(\mathbf{X}^f_{t_V}\right)^T \right]^T, \tag{10}$$

where $\mathbf{X}^f_{t_k} = \mathbf{X}^a_{t_0} + \int_{t_0}^{t_k} \dot{\mathbf{X}}(t)dt$, with $\mathbf{X}^a_{t_0} \in \mathbb{R}^{L_S \times M}$ being the analysis ensemble at t_0. The routine observation matrix for the augmented system is expressed as

$$\mathbf{H}^r_{aug} = [\, \mathrm{diag}(\mathbf{H}^r) \;\; \mathbf{0}_{Kn_r \times L_S}], \;\; \mathbf{H}^r \in \mathbb{R}^{n_r \times L_S} \tag{11}$$

where n_r is the number of routine measurement at each time step. If a state can be measured directly, $\mathbf{H}^r \in \{0,1\}^{n_r \times L_S}$ with every row having only one nonzero element.

Only the covariance information is updated for the purpose of the targeting; therefore, the ensemble update step is solely associated with the perturbation ensemble update:

$$\tilde{\mathbf{X}}^a_{aug} = (I - \mathbf{K}^r_{aug}\mathbf{H}^r_{aug})\tilde{\mathbf{X}}^f_{aug} \tag{12}$$

without consideration of ensemble mean update. In the EnSRF scheme, this process can obviously be performed one observation by one observation. $\tilde{\mathbf{X}}^a_{aug}$ will be utilized to construct the prior covariance matrix for the sensor selection problem. Since the selection problem only concerns the search space $\mathcal{S}$ and the verification region $\mathcal{V}$, it deals with the following submatrix of the $\tilde{\mathbf{X}}^a_{aug}$:

$$\tilde{\mathbf{X}}_{\mathcal{S}\cup\mathcal{V}} = \begin{bmatrix} \tilde{\xi}^T_{X_1} & \cdots & \tilde{\xi}^T_{X_N} & \tilde{\xi}^T_{V_1} & \cdots & \tilde{\xi}^T_{V_M} \end{bmatrix}^T \tag{13}$$

where $\tilde{\xi}_{(\cdot)}$ represents the row vector of $\tilde{\mathbf{X}}^a_{aug}$ corresponding to the subscribed variable. The covariance matrix can be evaluated as

$$\mathrm{Cov}(X_{\mathcal{S}} \cup \mathcal{V}) = \tilde{\mathbf{X}}_{\mathcal{S}\cup\mathcal{V}}\tilde{\mathbf{X}}^T_{\mathcal{S}\cup\mathcal{V}}/(M-1). \tag{14}$$

Note that $X_{\mathcal{S}} \cup \mathcal{V}$ might not be jointly Gaussian in spite of measurements being linear and Gaussian, if the dynamics is nonlinear. EnSRF is based on the assumption that at least for a given time instance, the ensemble distribution is sufficiently close to Gaussian distribution determined by the ensemble mean and covariance. This work assumes more that the ensemble distribution is sufficiently close to Gaussian within the time window $[t_0, t_V]$.

3.3 *Ensemble Targeting Algorithm*

After obtaining the *prior* ensemble $\tilde{\mathbf{X}}_{\mathcal{S}\cup\mathcal{V}}$, the forward ensemble targeting algorithms are written as

$$\mathbf{s}^\star_{F,En} = \arg\max_{\mathbf{s}} \; \mathrm{ldet}\tilde{\mathbf{X}}_{\mathcal{V}}\tilde{\mathbf{X}}^T_{\mathcal{V}} - \mathrm{ldet}\tilde{\mathbf{X}}_{\mathcal{V}|\mathbf{s}}\tilde{\mathbf{X}}^T_{\mathcal{V}|\mathbf{s}}, \tag{15}$$

and the backward ensemble targeting is expressed as

$$\mathbf{s}^\star_{B,En} = \arg\max_{\mathbf{s}} \; \mathrm{ldet}\left(\tilde{\mathbf{X}}_{\mathbf{s}}\tilde{\mathbf{X}}^T_{\mathbf{s}} + (M-1)R_{\mathbf{s}}\right) - \mathrm{ldet}\left(\tilde{\mathbf{X}}_{\mathbf{s}|\mathcal{V}}\tilde{\mathbf{X}}^T_{\mathbf{s}|\mathcal{V}} + (M-1)R_{\mathbf{s}}\right). \tag{16}$$

$\tilde{\mathbf{X}}_{\mathcal{V}} \in \mathbb{R}^{|\mathcal{V}|\times M}$ and $\tilde{\mathbf{X}}_{\mathcal{V}|\mathbf{s}} \in \mathbb{R}^{|\mathcal{V}|\times M}$ are the prior and posterior (perturbation) ensemble associated with the verification region, while $\tilde{\mathbf{X}}_{\mathbf{s}} \in \mathbb{R}^{n_s \times M}$ and $\tilde{\mathbf{X}}_{\mathbf{s}|\mathcal{V}} \in \mathbb{R}^{n_s \times M}$ are those for each measurement candidate.

3.4 *Numerical Results*

This section will briefly summarize the numerical results of the proposed ensemble-based targeting given in Choi (2007a). The two-dimensional Lorenz-95 model (Lorenz, 1998) with the following governing equations are considered:

$$\dot{y}_{ij} = (y_{i+1,j} - y_{i-2,j})\, y_{i-1,j} + \mu\,(y_{i,j+1} - y_{i,j-2})\, y_{i,j-1} - y_{ij} + F, \; (i = 1, \ldots, L_{on}, \; j = 1, \ldots, L_{at}). \tag{17}$$

where the subscript i denotes the west-to-eastern grid index, while j denotes the south-to-north grid index. There are $L_{on} = 36$ longitudinal and $L_{at} = 9$ latitudinal grid points, and the external forcing term $F = 8.0$. The dynamics in (17) are subject to cyclic boundary conditions in longitudinal direction ($y_{i+L_{on},j} = y_{i-L_{on},j} = y_{i,j}$) and to the constant advection condition($y_{i,0} = y_{i,-1} = y_{i,L_{at}+1} = 4$ in advection terms) in the latitudinal direction, to model the mid-latitude area as an annulus. Regarding the time scale, 0.05 time units are are equivalent to real 6 hours.

The routine network with size 93 is assumed to be already deployed over the grid space (blue 'o' in fig. 3). The static network is dense in two portions of the grid space called lands, while it is sparse the other two portions of the space called oceans. The routine network takes measurements every 0.05 time units. The leftmost part of the right land mass (consisting of 10 grid points depicted with red '□' in the figures) is the verification region. The goal is to reduce the forecast uncertainty in the verification region 0.5 time units after the targeting time. The targeting time $t_K = 0.05$ with $K = 1$. The analysis ensemble at t_0 with size 1024 is obtained by running the EnSRF with routine observations for 500 time units. All the grid points in the left ocean are considered as a candidate point to locate additional mea-

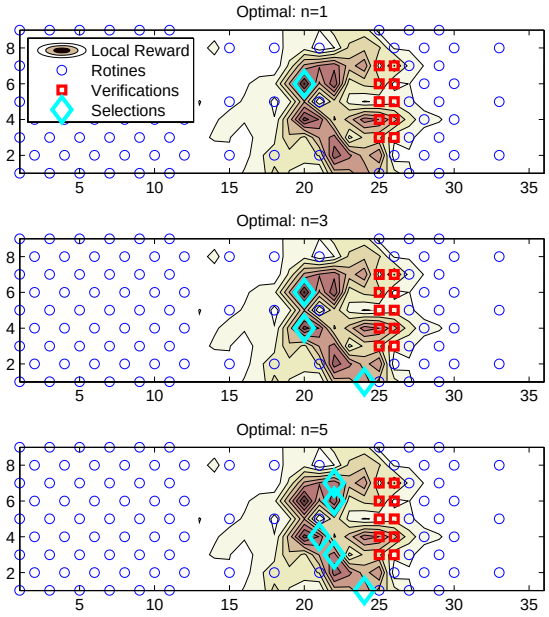

Figure 3: Optimal solution for ensemble-based targeting

surement at t_1; therefore, $N = 108$. The measurement noise variance R_s is 0.2^2 for routines and 0.02^2 for additional observations. With this setting, the targeting results, as the number of targeting points n_s being increased, are obtained.

It should be first pointed out that the backward algorithm gives the same solution with the forward algorithm for all the cases, while the it works much faster than the forward algorithm (Table 1). In the table, $T_{(\cdot)}$ and $\hat{T}_{(\cdot)}$ represent the actual and estimated computation times. Figure 3 illustrates a representative optimal sensor locations(cyan $\diamond$) for n_s =1, 3, and 5. The shaded contour in both figures represents the local reward value for each grid point, that is the entropy reduction of the verification site by a single measurement at the corresponding grid point. It is noted that the optimal solution for $n = 5$ looks very different from the solution for $n = 1$ and $n = 3$. This implies that coupled correlation structure amongst the search space should be taken into account to find the optimal targeting solution, and it could be arbitrarily suboptimal just picking the points with largest uncertainty. Choi (2007b) incorporated mobility of the multiple sensor platforms into the targeting decision, but this paper will not present those results because the sensitivity analysis with respect to the ensemble size is done for the simple sensor targeting problem described in this section.

4 EFFECTS OF LIMITED ENSEMBLE SIZE

As the quality of state representation by statistical ensembles is enhanced the ensemble size increases, the ratio M/L_S can be used to predict the quality of state representation. The numerical experiments in the previous section considered $M = 1024$ and $L_S = 324$, which leads to $M/L_S = 3.2$. It is noted that M/L_S is much smaller than this for a realistic weather model: a typical ensemble size ($\sim \mathcal{O}(10^2)$) is much smaller than the state dimension ($\sim \mathcal{O}(10^6)$), which results in M/L_S being $\mathcal{O}(10^{-4})$. This limitation in ensemble size is caused by the computational expense of storing and processing additional ensembles. Several researches have been performed to figure out and mitigate the effect of small M/L_S on the quality of forecast and data assimilation. Buizza (1998) showed that the impact of ensemble size on the forecast performance is highly dependent on the choice of metric. The ensemble mean is less sensitive to the ensemble size, while the spread-skill relation and outlier statistic is sensitive to the ensemble size. Hamill (2001) suggested distance-dependent EnKF data assimilation to mitigate the effect of error in estimation of correlation caused by small ensemble size.

On the other hand, since the quality of estimation of mutual information value for each measurement candidate is important for the targeting problem, a more reasonable index that represents the quality of the targeting solution is the ratio, the number of ensembles to the size of the covariance matrix whose determinant is computed; that is, n/M for the backward targeting algorithm and $|\mathcal{V}|/M$ for the forward targeting algorithm. Thus, the forward and the backward algorithms will represent different characteristics in terms of the sensitivity to the ensemble size.

Table 1: Actual and estimated computation times for ensemble-based targeting algorithms

N	n_s	T_F(s)	T_B(s)	$\hat{T}_F$ (s)	$\hat{T}_B$(s)
108	1	0.13	0.13	0.10	0.14
108	2	8.17	0.16	10.4	0.15
108	3	344.5	0.72	561.4	0.61
108	4	–	18.8	5.7 hr	14.6
108	5	–	461.3	6.5 day	349.3

4.1 *Well-Posedness*

The first concern regarding the limited ensemble size is *well-posedness* of the algorithm, as the ldet is not even defined for a rank-deficient matrix. Recall that the forward and backward targeting algorithms compute the determinants of $\left[\tilde{\mathbf{X}}_{\mathcal{V}|\cdot}\tilde{\mathbf{X}}_{\mathcal{V}|\cdot}^T\right]$ and $\left[\tilde{\mathbf{X}}_{\mathbf{s}|\cdot}\tilde{\mathbf{X}}_{\mathbf{s}|\cdot}^T + (M-1)R_{\mathbf{s}}\right]$, respectively. Then note that

1. Forward algorithm might suffer from rank-deficiency unless $|\mathcal{V}| \gg M$.

2. Backward algorithm will not be subject to rank-deficiency for nontrivial measurement noise with $R > 0$.

Since nontrivial sensing noise usually exists in practice, the backward algorithm is preferred to the forward algorithm in terms of algorithm well-posedness. In addition, for a given ensemble size M, handling a smaller covariance matrix is desirable for less sensitivity to the ensemble size (details discussed later). Recalling that n_s is usually smaller than $|\mathcal{V}|$, the backward algorithm is again preferred in this respect.

However, *well-posedness* of the backward algorithm regardless of the ensemble size does not mean that it provides *the* optimal solution insensitive to the ensemble size. Instead, the covariance matrix computed from the small number of ensembles can be inaccurate for small M, and results in a suboptimal targeting solution.

4.1 *Impact on Targeting Solution*

Monte-Carlo experiments are performed to quantify the degree of performance degradation caused by limited ensemble size. The scenario is the same as the previous section with a Lorenz-95 model, and $n = 2$ additional measurement points are selected from the search space of size $N = 108$ to reduce the forecast uncertainty of the verification site of size $|\mathcal{V}| = 10$. It is first assumed that true covariance field is well represented by the full set of ensembles with size $M_0 = 1024$; only the backward ensemble targeting is considered for this experiment.

For a given ensemble size M, 50 different combinations are chosen from the full set. Figure 4 compares the predicted optimal information gain $\hat{\mathcal{I}}(\mathbf{s}^\star_{B,M})$ to the actual information gain $\mathcal{I}(\mathbf{s}^\star_{B,M})$. For the backward targeting solution $\mathbf{s}^\star_{B,M}$ is based on the prior and posterior ensemble matrices $\mathbf{Y}_{\mathcal{S}}$ and $\mathbf{Y}_{\mathcal{S}|\mathcal{V}}$ of size $N \times M$, the above mentioned quantities are defined as:

$$\begin{aligned}\hat{\mathcal{I}}(\mathbf{s}^\star_{B,M}) = \tfrac{1}{2}\mathrm{ldet}\big(\mathbf{Y}_{\mathbf{s}^\star_{B,M}}\mathbf{Y}^T_{\mathbf{s}^\star_{B,M}}/(M-1) + RI_{n_s}\big) \\ - \tfrac{1}{2}\mathrm{ldet}\big(\mathbf{Y}_{\mathbf{s}^\star_{B,M}|\mathcal{V}}\mathbf{Y}^T_{\mathbf{s}^\star_{B,M}|\mathcal{V}}/(M-1) + RI_{n_s}\big) \quad (18)\end{aligned}$$

and

$$\begin{aligned}\mathcal{I}(\mathbf{s}^\star_{B,M}) = \tfrac{1}{2}\mathrm{ldet}\big(\mathbf{X}_{\mathbf{s}^\star_{B,M}}\mathbf{X}^T_{\mathbf{s}^\star_{B,M}}/(M_0-1) + RI_{n_s}\big) \\ - \tfrac{1}{2}\mathrm{ldet}\big(\mathbf{X}_{\mathbf{s}^\star_{B,M}|\mathcal{V}}\mathbf{X}^T_{\mathbf{s}^\star_{B,M}|\mathcal{V}}/(M_0-1) + RI_{n_s}\big) \quad (19)\end{aligned}$$

where $\mathbf{X}_{(\cdot)} \in \mathbb{R}^{N\times M_0}$ denotes the full ensemble matrix.

Figure 4 plots the average value taken over the different ensemble combinations with the error representing the 5th and 95th percentiles. As a reference, the average predicted and actual information gain for the random selection strategy is also plotted, while the red dotted line represents the mutual information value for the optimal solution computed with M_0 ensembles.

Note that there is a *discrepancy* between the predicted and actual information gain values (Δ_1 in the figure), which increases as ensemble size is reduced. This discrepancy indicates the magnitude of the estimation error by limitation of ensemble size. It is found that the average magnitude of discrepancy increases to 0.29 for $M = 100$. In addition to discrepancy, small ensemble size causes *performance degradation* (Δ_2 in the figure). The average value of performance degradation grows to 13% of the true optimal solution value as M decreases to 100. Another aspect that should be noted is *inconsistency* of the targeting solution for small ensemble size, which is represented by the size of errorbar. For $M = 100$, the size of errorbar increases to 32% of the true optimal solution value.

Figure 5 verifies that the discrepancy, the performance degradation, and the inconsistency due to limited ensemble size is not just an issue solely for the information-theoretic targeting. The figure depicts the similar error bar plots in case the uncertainty reduction of the verification site is measured by absolute decrease in the trace of covariance matrix of the verification site. In this case the predicted and actual uncertainty reductions are expressed as

$$\begin{aligned}\hat{\mathcal{T}}(\mathbf{s}^\star_{T,M}) =& \mathtt{tr}\big(\mathbf{Y}_{\mathcal{V}}\mathbf{Y}^T_{\mathcal{V}}/(M-1)\big) \\ & - \mathtt{tr}\big(\mathbf{Y}_{\mathcal{V}|\mathbf{s}^\star_{T,M}}\mathbf{Y}^T_{\mathcal{V}|\mathbf{s}^\star_{T,M}}/(M-1)\big) \quad (20)\end{aligned}$$

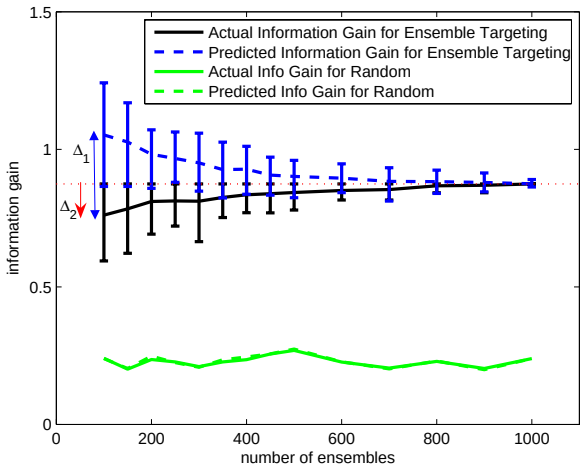

Figure 4: Error bar plots of predicted/actual objective value maximizing information gain

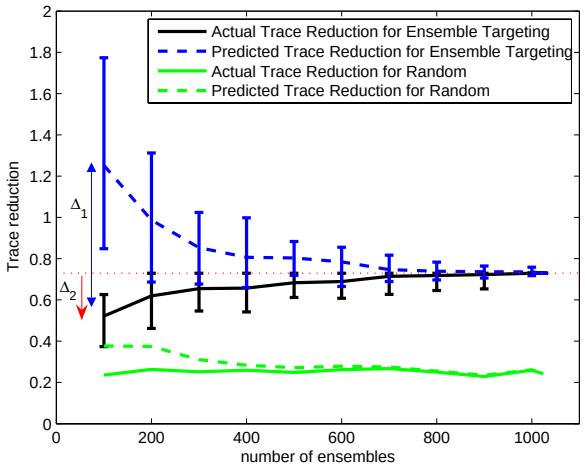

Figure 5: Error bar plots of predicted/actual objective value maximizing trace reduction

and

$$\mathcal{T}(\mathbf{s}^{\star}_{T,M}) = \mathtt{tr}\big(\mathbf{X}_{\mathcal{V}}\mathbf{X}_{\mathcal{V}}^{T}/(M_0-1)\big) - \mathtt{tr}\big(\mathbf{X}_{\mathcal{V}|\mathbf{s}^{\star}_{T,M}}\mathbf{X}_{\mathcal{V}|\mathbf{s}^{\star}_{T,M}}^{T}/(M_0-1)\big), \quad (21)$$

which should be computed in a *forward* manner because the commutativity does not hold for the trace reduction. A qualitatively similar trend in discrepancy, performance degradation, and inconsistency is found for $\hat{\mathcal{T}}$ and $\mathcal{T}$ as the case of information-theoretic targeting. Quantitatively, it can be seen that the trace-based targeting suffers from more significant discrepancy and performance degradation in this particular experiment.

5 ANALYSIS OF EFFECTS OF LIMITED ENSEMBLE SIZE

Given the experimental findings of the optimality degradation and the solution inconsistency due to limited ensemble size, this section investigates how the ensemble size is related to the solution optimality of the targeting algorithm and discusses why such large inconsistencies can occur, even for cases where M is two orders of magnitude larger than n_s.

5.1 *Sample Entropy Estimation*

As the simplest analysis, this section starts with the estimation of population entropy $h \equiv \log p$ (the factor 2 and the additive term are neglected for notational simplicity; but they can be incorporated without any difficulty) and the population variance p for comparison for a scalar Gaussian random variable X. Estimation is assumed to be based on $M = m+1$ randomly drawn samples (ensembles). *Unbiased* estimators for these two quantities exist (Misra, 2005):

$$\hat{p} = \frac{1}{m}\sum_{i=1}^{m+1}(x_i - \bar{x})^2 \quad (22)$$

$$\hat{h} = \log\frac{m\hat{p}}{2} - \psi\left(\frac{m}{2}\right) \quad (23)$$

where $\psi(\cdot)$ is a digamma function defined as $\psi(x) = \frac{d}{dx}\left(\log\Gamma(x)\right) = \frac{\Gamma'(x)}{\Gamma(x)}$.

Since $\frac{m\hat{p}}{p} \sim \chi^2_m$, the pdfs of the estimation errors $\tilde{p} \equiv \hat{p} - p$ and $\tilde{h} \equiv \hat{h} - h$ are expressed as

$$f_{\tilde{p}}(x) = \frac{m}{p} f_{\chi^2_m}\left(\frac{m}{p}(x+p)\right), \quad (24)$$

$$f_{\tilde{h}}(x) = 2e^{x+\psi(m/2)} f_{\chi^2_m}\left(2e^{x+\psi(m/2)}\right) \quad (25)$$

where $f_{\chi^2_m}(x) = \frac{1}{2}\left(\frac{x}{2}\right)^{m/2-1} e^{-x/2}$, $x \geq 0$. Also, the second and third central moments are computed as:

$$\mathbb{E}\left[(\hat{p}-p)^2\right] = \frac{2p^2}{m}, \quad \mathbb{E}\left[(\hat{p}-p)^3\right] = \frac{8p^3}{m^2}, \quad (26)$$

and

$$\mathbb{E}\left[(\hat{h}-h)^k\right] = \psi^{(k-1)}\left(\frac{m}{2}\right), \quad k = 2,3 \quad (27)$$

where $\psi^{(k-1)}$ is the $k-1$-th derivative of the digamma function. It is noted that the pdf of $\tilde{h}$ does not depend on the true value h, while that of $\tilde{p}$ depends on p. Figure 6 depicts the pdfs for both estimation errors for the case $m = 99$ and $p = 1$; it can be found that

for this value of p the second moment of $\tilde{p}$ is almost same as that of $\tilde{h}$ for any m (figure 7). Since the variance of $\tilde{p}$ is proportional to p, the shape of a pdf with a larger p will be more dispersed. In figure 6 the shapes of Gaussian pdfs with same variances, which are $\mathbb{E}[\tilde{p}^2]_{p=1} = 0.1421^2$ and $\mathbb{E}[\tilde{h}^2] = 0.1429^2$, are also plotted for comparison. Also, it can be found that pdf of $\tilde{p}$ is positively skewed (right long tail) while that of $\tilde{h}$ is slightly negatively skewed. Since the skewness is relatively small for $\tilde{h}$ and it is reduce further for large m, the pdf of $\tilde{h}$ is well approximated as a Gaussian.

As figure 7 shows the standard deviation of $\tilde{h}$ with respect to m that almost coincides that of $\tilde{p}$ in case $p = 1$, it is conceived that the estimation error of entropy estimation decreases on the order of $1/\sqrt{m}$ as m increases. For instance, in order to estimate p with estimation error standard deviation being less than 10% of the true value, more than 200 ensembles are needed. When regarding p itself as a *signal* not the variance of another signal, this can be interpreted as more than 200 samples are needed to have bigger than 20 dB SNR.

This statistical analysis can be extended to a multivariate case. The same sort of unbiased estimator of $\log\det P$ for $P \succ 0 \in \mathbb{R}^{n\times n}$ is written as

$$\hat{\mathcal{H}} = \mathrm{ldet}\left(\frac{m}{2^n}\hat{P}\right) - \prod_{i=0}^{n-1}\psi\left(\frac{m-i}{2}\right) \qquad (28)$$

where $\hat{P} \equiv \tilde{\mathbf{X}}\tilde{\mathbf{X}}^T/m$ is the minimum-variance unbiased estimator of P (Misra, 2003). Then, the estimation error variance becomes

$$\mathbb{E}\left[\left(\hat{\mathcal{H}} - \mathcal{H}\right)^2\right] = \sum_{i=0}^{n-1}\psi^{(1)}\left(\frac{m-i}{2}\right), \qquad (29)$$

which depends on the dimension of the random vector n, the sample size $m+1$, but not the true value of $\mathcal{H}$. Figure 8 depicts the error standard deviation for various values of n and m. The plots shows that large n and small m lead to large estimation error (on the order of $\sqrt{n/m}$). This dependency of error standard deviation on the order of $\sqrt{n/m}$ will be utilized to figure out the impact of limited ensemble size on the performance of the ensemble-based targeting, in the following section.

5.2 *Range-to-Noise Ratio*

The statistical analysis in the previous section is a generic result that does not take into account

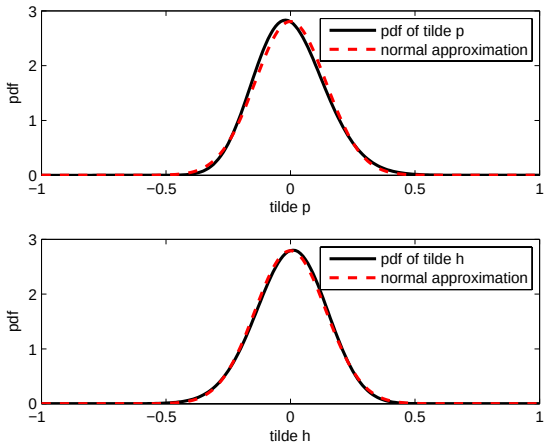

Figure 6: Probability density of estimation error ($m = 99$)

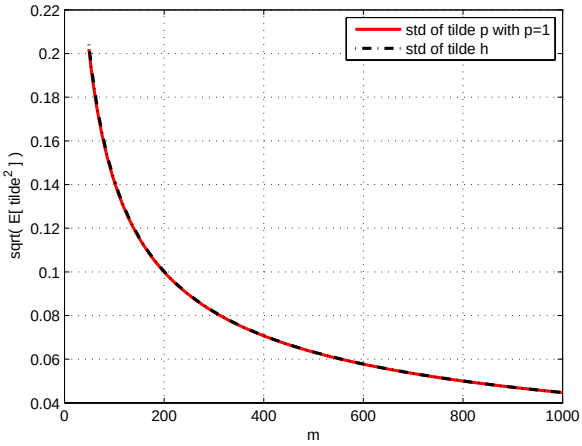

Figure 7: Standard deviation of $\tilde{p}$ and $\tilde{h}$

the specific aspects of the problem. Whether or not the limited sample (or ensemble) size has a big impact depends on the objective of the problem. For instance, when tracking the mean of the underlying random variable X, the accuracy in the estimation of p might not be important. Since $\mathbb{E}\left[(\mu_X - \bar{X})^2\right] = p/(m+1)$, the relative accuracy $\mathbb{E}\left[(\mu_X - \bar{X})^2\right]/\mu_X^2 = [(m+1)\mathrm{SNR}]^{-1}$, which is small for a large SNR signal.

The ensemble-based targeting problem must determine the best measurement candidate from other suboptimal measurement candidates. An important predictor of the degree of impact that limitation of sample size might have on the solution optimality is the *range-to-noise ratio* of the mutual information val-

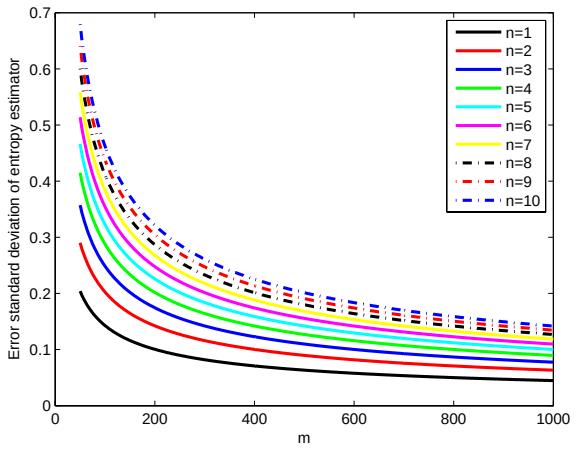

Figure 8: Standard deviation of $\tilde{\mathcal{H}}$

ues:

$$\mathrm{RNR} = \frac{\sup_{\mathbf{s}} \mathcal{I}(\mathbf{s}) - \inf_{\mathbf{s}} \mathcal{I}(\mathbf{s})}{\sup_{\mathbf{s}} \sqrt{\mathbb{E}\big[\big(\hat{\mathcal{I}}(\mathbf{s}) - \mathcal{I}(\mathbf{s})\big)^2\big]}}. \tag{30}$$

Utilizing the statistics of the entropy estimation described in the previous section, the $\sup$ value in the denominator can be obtained without regard to the true values of $\mathcal{I}(\mathbf{s})$. Estimation of mutual information can be treated as estimation of the two – prior and posterior – entropies and substraction of those two. Since the bias term $\prod_{i=0}^{n-1} \psi\left(\frac{m-i}{2}\right)$ is the same for the prior and posterior entropy estimation, the estimation error of the mutual information can be expressed as

$$\begin{aligned}
&\mathbb{E}\left\{\left[\tfrac{1}{2}(\hat{\mathcal{H}}^- - \hat{\mathcal{H}}^+) - \tfrac{1}{2}(\mathcal{H}^- - \mathcal{H}^+)\right]^2\right\} \\
&= \tfrac{1}{4}\mathbb{E}\big[(\hat{\mathcal{H}}^- - \mathcal{H}^-)^2\big] + \tfrac{1}{4}\mathbb{E}\big[(\hat{\mathcal{H}}^+ - \mathcal{H}^+)^2\big] \\
&\quad - \tfrac{1}{2}\mathbb{E}\big[(\hat{\mathcal{H}}^- - \mathcal{H}^-)(\hat{\mathcal{H}}^+ - \mathcal{H}^+)\big] \\
&= \tfrac{1}{2}\sum_{i=0}^{n-1} \psi^{(1)}\left(\tfrac{m-i}{2}\right) - \tfrac{1}{2}\mathbb{E}\big[(\hat{\mathcal{H}}^- - \mathcal{H}^-)(\hat{\mathcal{H}}^+ - \mathcal{H}^+)\big].
\end{aligned} \tag{31}$$

where superscripts '-' and '+' denote the prior and posterior. It can be shown that the cross correlation term in the final expression is always non-negative; so the estimation error of the mutual information is upper-bounded by

$$\mathbb{E}\big[\big(\hat{\mathcal{I}}(\mathbf{s}) - \mathcal{I}(\mathbf{s})\big)^2\big] \le \tfrac{1}{2}\sum_{i=0}^{n-1} \psi^{(1)}\left(\tfrac{m-i}{2}\right) \triangleq \sigma_{m,n}^2 \tag{32}$$

where equality holds if the prior and posterior entropy estimators are uncorrelated, which corresponds to infinite mutual information. With this upper bound, the RNR can be approximated as

$$\mathrm{RNR} \approx \frac{\sup_{\mathbf{s}} \mathcal{I}(\mathbf{s}) - \inf_{\mathbf{s}} \mathcal{I}(\mathbf{s})}{\sigma_{m,n}}. \tag{33}$$

In contrast to the denominator, the numerator of (33) is problem-dependent. Moreover, the $\sup$ and $\inf$ values cannot be known unless the true covariances (and equivalently true entropies) are known. Regarding the $\inf$ value, note that $\mathcal{I}(\mathbf{s})$ is lower-bounded by zero; therefore, it is reasonable to say that $\inf_{\mathbf{s}} \mathcal{I}(\mathbf{s})$ is a very small positive quantity. This suggests that we can approximate the $\inf$ value in the numerator of (33) as zero. With regard to the $\sup$ value, since the 95%-confident interval estimator of $\mathcal{I}(\mathbf{s})$ is $\hat{\mathcal{I}}(\mathbf{s}) \pm 2\sigma_{m,n}$, the interval estimate for RNR is

$$\widehat{\mathrm{RNR}} = \left[\max\left\{0, \frac{\sup_{\mathbf{s}} \hat{\mathcal{I}}(\mathbf{s})}{\sigma_{m,n}} - 2\right\}, \frac{\sup_{\mathbf{s}} \hat{\mathcal{I}}(\mathbf{s})}{\sigma_{m,n}} + 2\right] \tag{34}$$

with confidence level 95%. The $\max$ function that RNR is positive. If the objective of computing RNR is to predict whether or not a *small* RNR would cause significant performance degradation of the targeting, the following one-sided interval estimator can also be used:

$$\overline{\mathrm{RNR}} = \left[0, \frac{\sup_{\mathbf{s}} \hat{\mathcal{I}}(\mathbf{s})}{\sigma_{m,n}} + 1.7\right] \tag{35}$$

with 95% confidence level.

5.3 *Probability of Correct Decision*

This section considers the probability that the ensemble-based targeting provides the true optimal(or $(1-\epsilon)$-optimal) solution, which is referred to as *probability of correct decision* (PCD) hereafter, for given values of RNR, m, n, and the total number of candidates q. To do this, the following are assumed:

1. There are a total of q measurement candidates denoted as $\mathbf{s}_1, \ldots, \mathbf{s}_q$. Without loss of generality, $\mathbf{s}_i$ corresponds to the i-th best targeting solution.

2. The true mutual information values are uniformly distributed over the corresponding range $\mathcal{I}(\mathbf{s}_1) - \mathcal{I}(\mathbf{s}_q)$. In other words,

$$\mathcal{I}(\mathbf{s}_i) = \mathcal{I}(\mathbf{s}_1) - (i-1)\delta \tag{36}$$

where

$$\delta = \frac{\mathcal{I}(\mathbf{s}_1) - \mathcal{I}(\mathbf{s}_q)}{q-1} = \mathrm{RNR} \cdot \frac{\sigma_{m,n}}{q-1}. \tag{37}$$

3. The estimation error of each mutual information value is distributed with $\mathcal{N}(0, \sigma^2_{m,n})$.

4. The estimation errors of the mutual information for each measurement candidate are uncorrelated each other. In other words,

$$\mathbb{E}\left[(\hat{\mathcal{I}}(\mathbf{s}_i) - \mathcal{I}(\mathbf{s}_i))(\hat{\mathcal{I}}(\mathbf{s}_j) - \mathcal{I}(\mathbf{s}_j))\right] = 0, \ \forall i \neq j. \tag{38}$$

Under these assumptions, it can be shown that for $i \leq q-1$

$$D_i \triangleq \hat{\mathcal{I}}(\mathbf{s}_1) - \hat{\mathcal{I}}(\mathbf{s}_i) \sim \mathcal{N}\left((i-1)\delta, 2\sigma^2_{m,n}\right) \tag{39}$$

and

$$\mathrm{Cov}(D_i, D_j) = \sigma^2_{m,n}, \ \ \forall i \neq j. \tag{40}$$

Given that the PCD can de interpreted as the probability that the ensemble-based targeting declares $\mathbf{s}_1$ to be the best candidate, the PCD can be written in terms of D_i's as

$$\mathrm{PCD} = \mathtt{Prob}\left[D_i > 0, \ \forall i\right]. \tag{41}$$

Using (39) and (40), the PCD can be computed as

$$\mathrm{PCD} = \int_{-\infty}^{\mathrm{RNR}} \cdots \int_{-\infty}^{\frac{i}{q-1}\mathrm{RNR}} \cdots \int_{-\infty}^{\frac{\mathrm{RNR}}{q-1}} f_{\mathcal{N}(0,\Sigma)}(x_1, \cdots, x_{q-1}) dx_1 \cdots dx_{q-1} \tag{42}$$

where $f_{\mathcal{N}(0,\Sigma)}$ is the pdf of the zero-mean multivariate Gaussian distribution with the covariance matrix of

$$\Sigma = I_{q-1} + \mathbf{1}_{q-1} \otimes \mathbf{1}^T_{q-1}. \tag{43}$$

where I_{q-1} denotes the $(q-1) \times (q-1)$ identity matrix, $\mathbf{1}_{q-1}$ is the $(q-1)$-dimension column vector with every element being unity, and $\otimes$ denotes the Kronecker product. That is, all the diagonal elements of Σ_q are 2, while all the off-diagonal elements are one.

Note that the PCD is expressed as a cdf of a $(q-1)$-dimensional normal distribution. In the special case of $q = 2$,

$$\mathrm{PCD}_{q=2} = \Phi\left(\frac{\mathrm{RNR}}{\sqrt{2}}\right) \tag{44}$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution. For the case with $q > 2$, eigenvalue decomposition of the inverse of Σ in (43) leads to PCD being a product of univariate normal cdfs, and there exists an efficient numerical algorithm based on Cholesky factorization (Genz, 1992).

Figure 9 shows how PCD changes with q and RNR. the plots shows that PCD is monotonically increasing with respect to RNR, while it is monotonically decreasing with respect to q. The dependency of PCD on q is crucial, since q is a very large number in practice – recall that $q = \binom{N}{n}$. Thus, PCD can be meaninglessly small for a large-scale selection problem. In addition, to calculate PCD for such large q is computationally very expensive, because it requires a cdf evaluation of a large-dimensional normal distribution.

For this reason, for a large q case, this work suggests to utilize the probability of ϵ-correct decision (ϵ-PCD) defined as

$$\epsilon\text{-PCD} = \mathtt{Prob}\left[\bigcup_{i=1}^{\lfloor \epsilon q \rfloor} \left(\hat{\mathcal{I}}(\mathbf{s}_i) > \hat{\mathcal{I}}(\mathbf{s}_j), \ \forall j \neq i\right)\right], \tag{45}$$

since it can still be used as an indicator of the impact of limited sample size on the degree of optimality, and also it can be computed can be computed tractably.

By the symmetry of the distribution of the true mutual information values, the lower bound of this ϵ-PCD can be computed by

$$\mathrm{PCD}_{\lfloor 1/\epsilon \rfloor} \leq \epsilon\text{-PCD}, \tag{46}$$

where equality holds if $\lfloor \epsilon q \rfloor$ and $\lfloor 1/\epsilon \rfloor$ are integers. In other words, if dividing q candidates into $1/\epsilon$ groups such that the i-th group consists of $\mathbf{s}_{(i-1)\epsilon q+1}$ through $\mathbf{s}_{i\epsilon q}$, then the decision of declaring one of the candidates in the first group to be the optimal solution is equivalent to the decision of distinguishing the best candidate out of $1/\epsilon$ candidates.

With this, figure 9 can be used to interpret the relation between RNR and $\lceil q \rceil$-PCD. For instance, the graph of PCD for $q = 10$ represents the relation between RNR and 10%-PCD for any size of targeting problem. In the picture, the dotted line indicates the RNR value above which 10%-PCD is greater than 90%, which is 16.5. This is interpreted as *in order to have a 90%-optimal targeting solution for 90% of the cases, RNR should be greater than 16.5*. In terms of the one-sided interval estimator of RNR in (35), this implies that

$$\sup_{\mathbf{s}} \hat{\mathcal{I}}(\mathbf{s}) > 18.2\sigma_{m,n} \approx 18.2\sqrt{n/m} \tag{47}$$

for the same qualification with 95% confidence level. The last approximate expression comes from $\sigma_{m,n} \approx \sqrt{n/m}$ for a small n.

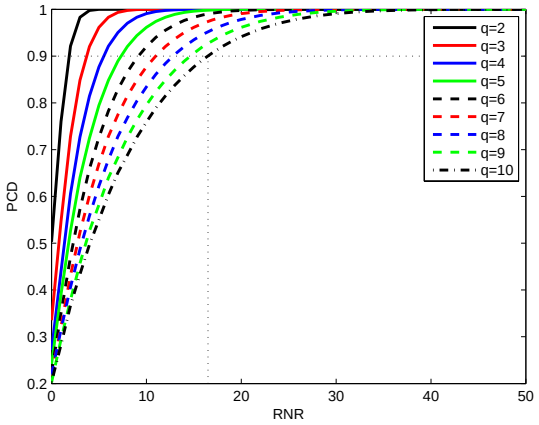

Figure 9: Probability of Correct Decision for targeting with q candidates

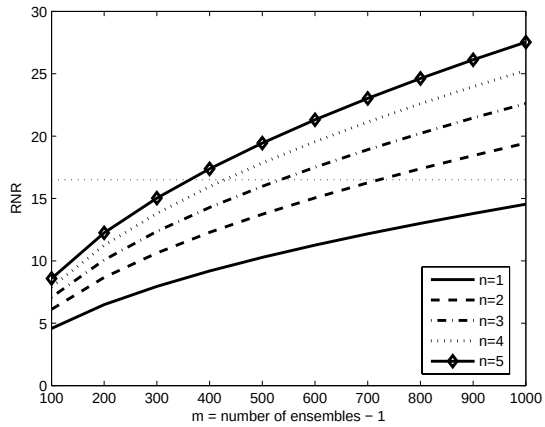

Figure 10: Range-to-noise ratio for Lorenz-95 example for various m and n

5.4 *Lorenz-95 Targeting Example*

The results of figure 4 showed a significant performance degradation for the targeting example using the Lorenz-95 model. This section will verify that that performance degradation can be predicted in terms of RNR and PCD. The RNR expression in (33) is considered under the assumption that $M_0 = 1024$ ensembles are sufficiently large to correctly estimate $\sup_{\mathbf{s}} \mathcal{I}(\mathbf{s})$ and $\inf_{\mathbf{s}} \mathcal{I}(\mathbf{s})$.

Figure 10 depicts the RNR values for $m \in [100, 1000]$ and for $n \in [1, 5]$. Note that for a given n, RNR decreases as m increases, while for a given m, it increases as n increases. For $n = 1$, the requirement of RNR>16.5 that achieves 90%-optimality for 90% of cases is not satisfied even with $m = 1000$, while $m = 400$ meets the same requirement for $n = 5$. Dependency of RNR on m is simply reflecting the fact that $\sigma_{m,n}$ is an increasing function of m for fixed n. The increasing tendency of RNR with respect to n is caused by the fact that the optimal mutual information value grows faster than $\mathcal{O}(\sqrt{n})$. Since

$$\mathrm{RNR} \approx \frac{\sup_{\mathbf{s}} \mathcal{I}(\mathbf{s})}{\sqrt{n/m}} \tag{48}$$

for a small n, RNR becomes an increasing function of n if the $\sup$ value in the numerator grows faster than $\mathcal{O}(\sqrt{n})$, which is the case for the Lorenz-95 example in this work. Also, seeing that the marginal increment of RNR diminishes as n increases, it is conceivable that there exists a threshold $\bar{n}$ over which increasing n no more improves RNR.

The Monte-Carlo simulation for figure 4 considers $n = 2$; it can be seen that the size of error bar becomes smaller than 10% of the optimal solution value when the ensemble size is larger than 700. This result is consistent with figure 10 which indicates that more than 750 ensembles are needed for 90%-optimality. Also, if $m = 100$ and $n = 2$, figure 10 shows that RNR=6.1; in figure 9, it can be found that the graph that crosses the dotted line of PCD=0.9 at RNR=6 is the one for $q = 4$. Since $\mathrm{PCD}_{q=4} = 25\%$-PCD, it is conceived that the ensemble targeting with $m = 100$ will provide 75%-optimal solutions for 90% of the cases, which is consistent with the result in figure 4.

6 CONCLUSIONS

This paper presented a computationally efficient algorithm for targeting additional observation networks to achieve improvement of ensemble forecast in terms of information-theoretic uncertainty metric, and performed sensitivity analysis of the proposed algorithm with respect to limitation of ensemble size that could be faced in practical implementation. While the computational effectiveness of the suggested algorithm is verified by numerical experiments, small ensemble size incurred substantial performance degradation of the targeting solution. To quantify the degree of impact that ensemble size might cause on the performance of the targeting solution, range-to-noise ratio and probability of correct decision are introduced and derived from the statistical analysis of entropy estimation. The presented statistical analysis was verified to be consistent with the numerical experiments.

ACKNOWLEDGMENT

This work is funded by NSF CNS-0540331 as part of the DDDAS program with Dr. Frederica Darema as the overall program manager.

REFERENCES

Bishop, C.H., B.J. Etherton, and S.J. Majumdar, 2001: Adaptive Sampling with the Ensemble Transform Kalman Filter: Part I Theoretical Aspects, *Monthly Weather Review*, **129**, 420-436.

Buizza, R., and T.N. Palmer, 1998: Impact of Ensemble Size on Ensemble Prediction, *Monthly Weather Review*, **126**, 2503-2518.

Choi, H.-L., J.P. How, and J.A. Hansen, 2007a: Ensemble-Based Adaptive Targeting of Mobile Sensor Networks, *American Control Conference*.

Choi, H.-L., and J.P. How, 2007b: A Multi-UAV Targeting Algorithm for Ensemble Forecast Improvement, *AIAA Guidance, Navigation, and Control Conference*.

Cover, T.M., and J.A. Thomas, 1991: *Elements of Information Theory*, Wiley Interscience.

Daescu, D.N., and I.M. Navon, 2004: Adaptive Observations in the Context of 4D-Var Data Assimilation, *Meteorology and Atmospheric Physics*, **85**, 205-226.

Evensen, G., and P.J. van Leeuwen, 1996: Assimilation of Altimeter Data for the Agulhas Current Using the Ensemble Kalman Filter with a Quasigeostrophic Model, *Monthly Weather Review*, **123**, 85-96.

Genz, A., 1992: Numerical Computation of Multivariate Normal Probabilities, *J. Comp. Graph. Stat.*.

Hamill, T.M., J.S. Whitaker, and C. Snyder, 2001: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter,*Monthly Weather Review*, **129**, 2776-2790.

Leutbecher, M., 2003: Adaptive Observations, the Hessian Metric and Singular Vectors," *ESMWF Seminar 2003*.

Lorenz, E.N., and K.A. Emanuel, 1998: Optimal Sites for Supplementary Weather Observations: Simulation with a Small Model, *Journal of the Atmospheric Sciences*, **55**, 399-414.

Majumdar, S.J., C.H. Bishop, B.J. Etherton, and Z. Toth, 2002: Adaptive Sampling with the Ensemble Transform Kalman Filter: Part II Field Programming Implementation, *Monthly Weather Review*, **130**,1356-1369.

Misra, N., H. Singh, and E. Demchuck, 2005: Estimation of the Entropy of a Multivariate Normal Distribution, *Journal of Multivariate Analysis*, **92**, 324-342.

Palmer, T.N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular Vectors, Metrics, and Adaptive Observations, *Journal of the Atmospheric Sciences*, **55**, 633-653.

Whitaker, J.S. and H.M. Hamill, 2002: Ensemble Data Assimilation without Perturbed Observations, *Monthly Weather Review*, **130**, 1913-1924.