Tellus SERIES A DYNAMIC METEOROLOGY AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Tellus (2009), 61A, 144–153
Printed in Singapore.

TELLUS

Measuring information content from observations for data assimilations: connection between different measures and application to radar scan design

By QIN XU[1*], LI WEI[2] and SEAN HEALY[3], [1]*NOAA/National Severe Storms Laboratory, Norman, OK, USA;* [2]*Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, OK, USA;* [3]*European Centre for Medium-Range Weather Forecasts, UK*

(Manuscript received 13 April 2008; in final form 2 October 2008)

ABSTRACT

The previously derived formulations for using the relative entropy and Shannon entropy difference (*SD*) to measure information content from observations are revisited in connection with another known information measure—degrees of freedom for signal, which is defined as the statistical average of the signal part of the relative entropy. For a linear assimilation system, the statistical average of the relative entropy reduces to the *SD*. The formulations are extended for four-dimensional variational data assimilation (4DVar). The extended formulations reveal that the information content increases (or decreases) as the model error increase (or decrease) and/or become strongly (or weakly) correlated in 4-D space. These properties are also highlighted by illustrative examples, and the extended formulations are shown to be potential useful for designing optimum phased-array radar scan configurations to maximize the extractable information contents from radar observations by a 4DVar analysis system.

1. Introduction

Properly measuring information content from observations can be very important for data assimilation and many related applications (such as designing observation systems, remote sensing strategies, targeted observations, and data compression strategies). The subject concerning how to measure the information content extracted from observations (or compressed observations) by a Bayesian optimal analysis is not new. In many early studies of satellite observations of the atmosphere (Peckham, 1974; Eyre, 1990; Huang and Purser, 1996; Purser et al., 2000; Rodgers, 2000), the Shannon entropy (Shannon, 1949) of probability density function (pdf) was used as a measure of information content. In particular, the Shannon entropy difference (*SD*) between the analysis pdf and background pdf was used to measure the information content extracted from observations by the analysis for designing optimum satellite radiometer configurations and data compression (superobbing) strategies.

Recently, Kleeman (2002) and Majda et al. (2002) advocated the use of the relative entropy of the prediction and climatological pdfs as a measure of the utility of a particular statistical prediction, and demonstrated advantages of using the relative entropy for predictability studies. Inspired by these studies, Xu (2007) examined the use of the relative entropy, versus the *SD*, as a measure of the amount of information extracted from observations by a Bayesian optimal analysis. The observations and analysis considered in Xu (2007), however, confined in spatial dimensions, although the analysis can be either a three-dimensional variational (3DVar) analysis (Lorenc, 1981; Daley, 1991) or Kalman Filter (KF) analysis (Jazwinski, 1970; Cohn, 1997). To measure information content from four-dimensional (4-D) observations such as those from phased-array radar rapid scans, the previous formulations need to be extended by considering the temporal dimension in addition to the spatial dimensions. The extended formulations can be then used, at least formally, to measure the information content extracted from 4-D observations by a four-dimensional variational (4DVar) analysis (Lewis and Derber, 1985; LeDimet and Talagrand, 1986; Bennett, 1992; Courtier, 1997). This extension is important for designing optimal phased-array radar scan configurations at the newly established the National Weather Radar Testbed (NWRT) in Norman, Oklahoma (Zrnic et al., 2007). This is the motivation of this study.

The paper is organized as follows. In the next section, the previously derived formulations for the two information measures (that is, the relative entropy and *SD*) are briefly reviewed and remarked in connection with another information measure—degrees of freedom for signal (Wahba et al., 1995; Rodgers,

*Corresponding author.
e-mail: qin.xu@noaa.gov
DOI: 10.1111/j.1600-0870.2008.00373.x

2000). The formulations are extended in Section 3 to measure information contents from observations extracted by a 4DVar analysis system. Illustrative examples are presented in Section 4 to highlight the potential usefulness of the extended formulations and related properties for designing optimum phased-array radar scan configurations to maximize the information contents extractable from radar observations by a 4DVar analysis system. Conclusions follow in Section 5.

2. Information contents from observations extracted by 3DVar analyses

2.1. Review of the previous results

When observations are assimilated into a numerical weather prediction (NWP) model by a Bayesian optimal analysis, the background state is provided by the NWP model prediction valid at the analysis time. As shown in Xu (2007), the information extracted from radar observations by an analysis can be measured by the relative entropy defined by $R(p, q) = \langle \ln (p/q) \rangle_{\mathrm{p}}$, where $q = q(\mathbf{x})$ is the background pdf, $p = p(\mathbf{x})$ is the analysis pdf, $\mathbf{x} \in R^n$ is the state vector, and $\langle () \rangle_{\mathrm{p}} \equiv \int d\mathbf{x} p(\mathbf{x})()$ denotes the expectation of () based on $p(\mathbf{x})$. For Gaussian pdfs, this gives

$$\begin{aligned} R(p,q) &= \langle \ln(p/q) \rangle_{\mathrm{p}} \\ &= J_b + [\ln \mathrm{Det}(\mathbf{B}\mathbf{A}^{-1}) + \mathrm{Tr}(\mathbf{A}\mathbf{B}^{-1} - \mathbf{I}_{\mathrm{n}})]/2, \end{aligned} \tag{1}$$

where $J_{\mathrm{b}} = (\mathbf{a} - \mathbf{b})^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{a} - \mathbf{b})/2$ is the signal part of the information content which has the same form as the background term in the costfunction, $\mathbf{b}$ is the background mean, $\mathbf{a}$ is the analysis mean, $\mathbf{B}$ is the background covariance matrix, $\mathbf{A}$ is the analysis covariance matrix, $\mathbf{I}_{\mathrm{n}}$ is the $n \times n$ identity matrix, $()^{\mathrm{T}}$ denotes the transpose of (), Det() denotes the determinant of (), Tr() denotes the trace of (). The second term on the right-hand side of (1) is the dispersion part of the information content. The Shannon entropy of $q(\mathbf{x})$ is defined by $S(q) = -\langle \ln q \rangle_{\mathrm{q}}$, and the first term in the dispersion part of relative entropy in (1) is given by the SD:

$$\begin{aligned} SD \equiv S(q) - S(p) &\equiv \langle \ln p \rangle_{\mathrm{p}} - \langle \ln q \rangle_{\mathrm{q}} = \ln \mathrm{Det}(\mathbf{B}\mathbf{A}^{-1})/2 \\ &= Tr[-\ln(\mathbf{A}\mathbf{B}^{-1})]/2. \end{aligned} \tag{2}$$

For a Bayesian optimal analysis (Jazwinski, 1970),

$$\mathbf{a} = \mathbf{b} + \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{d} \tag{3}$$

and $\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$ or, equivalently,

$$\mathbf{A} = \mathbf{B} - \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}, \tag{4}$$

where $\mathbf{R}$ is the observation error covariance matrix, $\mathbf{H}$ is the tangent-linearization of the observation operator $H()$ at $\mathbf{x} = \mathbf{b}$, $\mathbf{d} = \mathbf{y} - H(\mathbf{b})$ is the innovation vector; and $\mathbf{y} \in R^m$ is the observation vector. Substituting these relationships into (1) and (2) gives

$$R(p,q) = \sum [d_i'^2 \lambda_i^2/(1+\lambda_i^2)^2 + \ln(1+\lambda_i^2) - \lambda_i^2/(1+\lambda_i^2)]/2, \tag{5}$$

$$SD = S(q) - S(p) = \sum [\ln(1+\lambda_i^2)]/2. \tag{6}$$

Here, d_i' is the ith element of $\mathbf{d}' = \mathbf{U}^{\mathrm{T}}\mathbf{R}^{-1/2}\mathbf{d}$, $\mathbf{U}$ is the left orthogonal matrix given by the singular value decomposition (SVD) of the scaled observation operator:

$$\mathbf{M} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}, \tag{7}$$

where λ_i is the ith diagonal element of the diagonal matrix $\mathbf{\Lambda}$, and the summation is over i from 1 to $r = \mathrm{rank}(\mathbf{M})$. Note that the observation space is transformed by $\mathbf{U}^{\mathrm{T}}\mathbf{R}^{-1/2}$ in (5) and (6). In this transformed space, the information content becomes separable between components associated with different singular values of $\mathbf{M}$.

The above results lead to the following important points: (1) By applying the truncated transformation $\mathbf{I}_{\mathrm{s}}\mathbf{U}^{\mathrm{T}}\mathbf{R}^{-1/2}$ to the observation vector $\mathbf{y}$, the observations can be compressed into superobservations given by $\mathbf{y}_{\mathrm{s}} = \mathbf{I}_{\mathrm{s}}\mathbf{U}^{\mathrm{T}}\mathbf{R}^{-1/2}\mathbf{y}$, where $\mathbf{I}_{\mathrm{s}}$ is the $s \times s$ identity matrix that projects R^{m} onto R^{s} ($s < m$) and m is the dimension of $\mathbf{y}$. (2) This SVD-based compression causes no information loss unless $s < r = \mathrm{rank}(\mathbf{M}) \le \min(n, m)$ where n is the dimension of $\mathbf{x}$. (3) For a given truncation number of $s < r$, the information loss caused by the SVD-based compression is minimum in the dispersion part, while the signal part of the information loss depends on not only the truncated non-zero singular values but also the associated components of $\mathbf{d}$'.

2.2. Remarks

(1) We note that if the Bayesian optimal-estimation problem is linear (or nearly linear), then the relative entropy in (1) can be written into

$$R(p,q) = J_{\mathrm{b}} + SD - DFS/2, \tag{8}$$

where $DFS \equiv \langle 2J_{\mathrm{b}} \rangle$ is the degrees of freedom for signal as defined in (2.46) of Rodgers (2000), and $\langle () \rangle = \int \int d\mathbf{x}\, d\mathbf{y}\, p(\mathbf{x}, \mathbf{y})()$ denotes the expectation of () based on the joint pdf $p(\mathbf{x}, \mathbf{y}) = q(\mathbf{x})p(\mathbf{y}|\mathbf{x})$ with $q(\mathbf{x}) = [(2\pi)^n \mathrm{Det}(\mathbf{B})]^{-1/2} \exp[-(\mathbf{x} - \mathbf{b})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{b})/2]$ and $p(\mathbf{y}|\mathbf{x}) = [(2\pi)^m \mathrm{Det}(\mathbf{R})]^{-1/2} \exp\{-[\mathbf{y} - H(\mathbf{x})]^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y} - H(\mathbf{x})]/2\}$. As explained in section 2.4.2 of Rodgers (2000), DFS measures how many of degrees of freedom of an observation are related to signal (versus noise). For a linear (or linearized) observation operator $H() = \mathbf{H}$, we have $\langle \mathbf{d}\mathbf{d}^{\mathrm{T}} \rangle = \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R}$. Substituting this result with (3)–(4) and (7) into the definition of DFS gives [also see eqs (2.46)–(2.62) of Rodgers (2000)]

$$\begin{aligned} DFS &\equiv \langle 2J_b \rangle = \langle (\mathbf{a}-\mathbf{b})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{a}-\mathbf{b})\rangle \\ &= \mathrm{Tr}[\langle(\mathbf{a}-\mathbf{b})(\mathbf{a}-\mathbf{b})^{\mathrm{T}}\rangle\mathbf{B}^{-1}] \\ &= \mathrm{Tr}[\mathbf{BH}^{\mathbf{T}}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\langle\mathbf{dd}^{\mathrm{T}}\rangle(\mathbf{HBH}^{\mathrm{T}}+\mathrm{R})^{-1}\mathbf{H}] \\ &= \mathrm{Tr}[\mathbf{BH}^{\mathbf{T}}(\mathbf{HBH}^{\mathrm{T}}+\mathbf{R})^{-1}\mathbf{H}] \\ &= \mathrm{Tr}(\mathbf{I}_n - \mathbf{AB}^{-1}) \\ &= \sum \lambda_i^2/(1+\lambda_i^2). \end{aligned} \tag{9}$$

This derives (8) from (1), and it also leads to

$$\langle R(p,q)\rangle \;=\; \langle SD\rangle = SD. \tag{10}$$

Note that the relative entropy is the expectation of $\ln(p/q)$ based on $p(\mathbf{x})$ while $p(\mathbf{x}) = p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x}, \mathbf{y})/[\int d\mathbf{x}p(\mathbf{x}, \mathbf{y})]$ is the posterior analysis pdf conditioned by $\mathbf{y}$. This implies that the relative entropy depends on the realization of $\mathbf{y}$. In particular, the factor $\mathbf{a} - \mathbf{b}$ in the signal part J_b depends on the realization of $\mathbf{d} = \mathbf{y} - H(\mathbf{b})$ according to (3). For an individual realization of $\mathbf{d}$, the relative entropy can measure both the signal and dispersion parts of the information content, and these two parts can be related to the improvements made by the analysis to the mean and covariance, respectively, as shown in Xu (2007). By considering all possible realizations of $\mathbf{d}$, the averaged relative entropy reduces to the *SD* if the Bayesian optimal-estimation problem is linear (or nearly linear), as shown in (10).

(2) The number of degrees of freedom for signal was defined by $DFS = \mathrm{Tr}(\mathbf{R}^{-1/2}\mathbf{HAH}^{\mathrm{T}}\mathbf{R}^{-1/2})$ in Wahba et al. (1995). This definition is equivalent to the definition in (9) [see (2.46) of Rodgers, 2000]. A proof of this equivalence was given in appendix A of Fisher (2003). The equivalence can be also verified by using the SVD of $\mathbf{R}^{-1/2}\mathbf{HA}^{1/2}$ and (A12) of Xu (2007). In this case, since cyclic permutation of matrix factors does not change the trace of the product, it is easy to see that $\mathrm{Tr}(\mathbf{R}^{-1/2}\mathbf{HAH}^{\mathrm{T}}\mathbf{R}^{-1/2}) = \mathrm{Tr}(\mathbf{AH}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}) = \mathrm{Tr}(\mathbf{I}_n - \mathbf{AB}^{-1})$, where (4) is used in the last step.

(3) According to (1) and (8), $SD - DFS/2$ is the dispersion part of $R(p, q)$. This implies that *DFS* measures only the dispersion part of the information content, and this is a similarity between *DFS* and *SD*. One can also verify that *DFS* is invariant only to a linear transformation, and this is another similarity between *DFS* and *SD*. The relative entropy measures both the signal and dispersion parts and is strictly invariant with respect to any smooth invertible transformation of variables (see section 6 of Xu, 2007, and section 2.4 of Majda et al., 2002). As pointed out by James Purser (personal communication, 2006), the *SD* is additive for successive inclusions of observations into the analyses, but the relative entropy is not additive. Like the relative entropy, *DFS* is not additive and thus is not convenient for computing cumulative information content from successive groups of observations. This is an important difference between *DFS* and *SD*.

(4) The real atmosphere provides only a single true realization of $\mathbf{x}$. As the analysis uses all available observations in each assimilation cycle, there is only a single realization of $\mathbf{y}$ for each analysis and thus only a single realization of J_b is produced in each assimilation cycle. In this case, although *DFS* can be computed from $\mathrm{Tr}(\mathbf{I}_n - \mathbf{AB}^{-1}) = \sum\lambda_i{}^2/(1 + \lambda_i{}^2)$ in each cycle according to (9), estimating $DFS \equiv <2J_b>$ directly from $2J_b$ requires certain assumptions and approximations. In 3DVar, $\mathbf{B}$ and $\mathbf{R}$ are pre-estimated and fixed in time. If the observation operator $\mathbf{H}$ is also fixed in time, then the ergodicity assumption can be applied to $\mathbf{d}$ and thus *DFS* can be estimated by averaging $2J_b$ over a large number of assimilation cycles. This can be a simple and efficient way to estimate *DFS* for 3DVar and approximately for 4DVar since J_b is free and routinely available in 3DVar and 4DVar (Healy and Thépaut, 2006).

3. Extended formulations for 4DVar analyses

3.1. Extended formulations

The results reviewed in Section 2, along with the illustrative examples with real radar observations presented in Xu (2007), suggest that the information content extracted by an optimal analysis from dense observations (such as those from Doppler weather radars) depends on how well the observations are resolved by the analysis. If high-resolution radar observations are poorly resolved by the analysis (due to lack of fine structures in the background covariance and/or insufficient resolution in the analysis), then there can be a significant degree of information redundancy or observation resolution redundancy. The formulations reviewed in Section 2 were derived for spatial analyses only. To measure information content gained from phased-array radar rapid scans, we need to consider the temporal dimension in addition to the spatial dimensions and extend the formulations for 4DVar analyses in this section.

The time window for a 4DVar analysis can cover, say, $J + 1$ observation time levels (at t_j for $j = 0, 1, 2, \ldots J$) between the initial time level (at t_0) and ending time level (at t_J). The time evolution of background error is assumed to obey the tangent-linearized prediction model of the following form:

$$\delta\mathbf{x}_j = \mathbf{F}_j\delta\mathbf{x}_{j-1} + \mathbf{q}_j \quad \text{for} \quad j = 1, 2, \ldots J+1, \tag{11}$$

where $\delta\mathbf{x}_j = \mathbf{x}_j - \mathbf{b}_j$, $\mathbf{x}_j$ denotes the true state vector at t_j, $\mathbf{b}_j$ is the background estimation of $\mathbf{x}_j$, $\mathbf{F}_j$ is the tangent-linearized forward operator from t_{j-1} to t_j, and $\mathbf{q}_i$ is the model error integrated from t_{j-1} to t_j. The initial background error and model error are assumed to be unbiased (or bias-removed), that is, $\langle\delta\mathbf{x}_0\rangle = 0$ and $\langle\mathbf{q}_j\rangle = 0$ for $j = 1, 2, \ldots J$. The linearity of (11) ensures that the background error remains unbiased after the initial time, that is, $\langle\delta\mathbf{x}_j\rangle = 0$ for $j = 1, 2, \ldots J$. The tangent-linearized observation model is given by

$$\mathbf{d}_j = \mathbf{H}_j\delta\mathbf{x}_j + \mathbf{e}_j \quad \text{for} \quad j = 0, 1, 2, \ldots J, \tag{12}$$

where $\mathbf{d}_j = \mathbf{y}_j - H_j(\mathbf{b}_j)$, $\mathbf{y}_j$ is the observation vector at t_j, $\mathbf{H}_j$ is the tangent-linearization of $H_j()$ with respect to $\mathbf{b}_j$, $H_j()$ is the observation operator at t_j, and $\mathbf{e}_j$ denotes the observation error at t_j. The observation error is assumed to be unbiased or bias-removed, so $\langle \mathbf{e}_j \rangle = 0$ for $j = 0, 1, \ldots J$.

The 4-D state increment fields over the analysis time window can be represented by the extended vector: $\delta\mathbf{x} = (\delta\mathbf{x}_0{}^{T}, \delta\mathbf{x}_1{}^{T}, \ldots \delta\mathbf{x}_J{}^{T})^{T}$. With this extension, (11) can be written into

$$\delta\mathbf{x} = \mathbf{G}\mathbf{q}, \tag{13}$$

where $\mathbf{q} = (\delta\mathbf{x}_0{}^{T}, \mathbf{q}_1{}^{T}, \ldots \mathbf{q}_J{}^{T})^{T}$ and $\mathbf{G}$ is a $(J+1)\times(J+1)$ lower-triangle block matrix operator with the jth diagonal block given by $\mathbf{I}$ (the identity matrix in R^n) and the jkth block (at the jth row and kth column) given by $\mathbf{F}_{j-k}\ldots\mathbf{F}_2\mathbf{F}_1$ for $k < j$ (below the block-diagonal) and by zero for $k > j$ (above the block-diagonal). The covariance for the extended vector $\mathbf{q}$ is

$$\mathbf{Q} = \langle \mathbf{q}\mathbf{q}^{T} \rangle. \tag{14a}$$

If the vector components of $\mathbf{q}$ are not cross-correlated, then $\mathbf{Q}$ reduces to the following block diagonal matrix:

$$\mathbf{Q} = \mathrm{diag}(\mathbf{B}_0, \mathbf{Q}_1, \ldots \mathbf{Q}_J), \tag{14b}$$

where $\mathbf{B}_0 = \langle \delta\mathbf{x}_0 \delta\mathbf{x}_0{}^{T} \rangle$ is the background covariance at the initial time, which is the same as $\mathbf{B}$ in (1), and $\mathbf{Q}_j = \langle \mathbf{q}_j \mathbf{q}_j{}^{T} \rangle$ is the covariance of $\mathbf{q}_j$. The background covariance for the extended state vector in (13) is then given by $\mathbf{GQG}^{T}$, according to (14a) or (14b).

Similarly, the background and analysis mean vectors can be extended and denoted by $\mathbf{b} = (\mathbf{b}_0{}^{T}, \mathbf{b}_1{}^{T}, \ldots \mathbf{b}_J{}^{T})^{T}$ and $\mathbf{a} = (\mathbf{a}_0{}^{T}, \mathbf{a}_1{}^{T}, \ldots \mathbf{a}_J{}^{T})^{T}$, respectively. The extended innovation vector is $\mathbf{d} = (\mathbf{d}_0{}^{T}, \mathbf{d}_1{}^{T}, \ldots \mathbf{d}_J{}^{T})^{T}$. The extended observation error covariance matrix is $\mathbf{R} = \mathrm{diag}(\mathbf{R}_0, \mathbf{R}_1, \ldots \mathbf{R}_J)$. With the above extensions, the costfunction and related solutions for a 4DVar analysis can be cast into the same matrix forms as in (3)–(4) (Bennett, 1992; Courtier, 1997; Lorenc, 2003) and hence the information entropy formulations can be obtained in the same forms as in (1)–(2) for an optimal 3-D analysis except that the original background covariance matrix $\mathbf{B}$ in (1)–(4) is replaced by $\mathbf{GQG}^{T}$ and the original observation operator is replaced by $\mathbf{H} = \mathrm{diag}(\mathbf{H}_1, \mathbf{H}_2, \ldots \mathbf{H}_J)$. In this case, the scaled observation operator $\mathbf{M}$ and associated SVD in (7) should be replaced by

$$\mathbf{M} = \mathbf{R}^{-1/2}\mathbf{HGQ}^{1/2} = \mathbf{U}\Lambda\mathbf{V}^{T}. \tag{15}$$

By using the SVD of the extended $\mathbf{M}$ in (15), the same forms of formulations can be derived in terms of the singular values as in (5)–(6) and used to measure the information content extracted from 4-D observations by a 4DVar analysis.

3.2. Remarks

(1) For an imperfect model (with $\mathbf{Q}_j \neq 0$ for $j = 1, 2, \ldots J$), $\mathbf{M}$ in (15) is a $(J+1)m\times n(J+1)$ matrix. This matrix is much larger than the $m\times n$ matrix $\mathbf{M}$ in the original (7). The rank of $\mathbf{M}$ in (15) can as large as $(J+1)\min(m, n)$. As this rank can be $J+1$ times as large as the rank of $\mathbf{M}$ in (7), the number of nonzero singular values can be increased by $J+1$ times. Thus, according to (5) or (6), the information content extracted from a set of 4-D observations by a 4DVar analysis with an imperfect model can be much larger than that from any subset of 3-D observations extracted by a 3-D analysis.

(2) If the model is perfect [with $\mathbf{Q}_j = 0$ in (14b) for $j = 1, 2, \ldots J$], then $\mathbf{M}$ reduces to its first block column which is a $(J+1)m\times n$ matrix, while all the remaining J block columns become zero (as $\mathbf{Q}_j = 0$). This first block column contains $J+1$ submatrices with the $(j+1)$th submatrix given by $\mathbf{R}_j{}^{-1/2}\mathbf{H}_j\mathbf{F}_j\ldots\mathbf{F}_1\mathbf{B}_0{}^{1/2}$. Thus, if the model error reduces to zero, the upper bound of rank($\mathbf{M}$) will reduce from $(J+1)\min(m, n)$ to $\min[(J+1)m, n] \leq n$ and the information content extracted by the 4DVar analysis will reduce significantly. Intuitively, we may envision that the perfect model operator maps the 4-D observations (forward in time along the model solution trajectories) into the 3-D space at the ending time ($t = t_J$) of the 4-D analysis time window. As the mapped observations are densely packed into this 3-D space, their resolution redundancy can be readily revealed. If the observations are already dense in the 3-D space (with $m \geq n$ or nearly so) and mapped more densely from the 4-D space to the 3-D space by the perfect-model operator, then the number of non-zero singular values will not increase [as it is still bounded by rank($\mathbf{M}$) $\leq n$] but the magnitudes of these non-zero singular values will increase mostly above or far above 1. Note that when $\lambda_i >$ or $>> 1$, the *DFS*-component term $\lambda_i{}^2/(1 + \lambda_i{}^2)$ in (9) will increase with λ_i much more slowly than the *SD*-component term $[\ln(1 + \lambda_i{}^2)]/2$ in (6). Thus, although the *SD* and *DFS* will both reduce significantly in the perfect-model 4-D case, the reduced *SD* will remain to be significantly larger than the *SD* computed from 3-D observations (at a single time level) and the reduced *DFS* can be very close to the *DFS* computed from 3-D observations, especially if the observations are sufficiently dense in the 3-D space.

(3) If the model errors are non-correlated in time and space, then each $\mathbf{Q}_j$ in (14b) is a diagonal matrix and the block diagonal matrix diag$\{\mathbf{Q}_1, \ldots \mathbf{Q}_J\}$ is completely diagonal. If the model errors become less (or more) spatially correlated, then $\mathbf{Q}$ in (14b) will become more (or less) diagonal, and this will increase (or decrease) the singular values of $\mathbf{M}$ in (15) and thus increase (or decrease) the information content extracted by the 4DVar analysis. Similarly, if the model errors become temporarily less (or more) correlated, then the matrix $\mathbf{Q}$ in (14a) becomes more (or less) bock-diagonal, and this will increase (or decrease) the information content extracted by the 4DVar analysis.

4. Illustrative examples

In this section, illustratve examples will be presented to highlight the potential usefulness of the above extended formulations and remarked properties for designing optimum phased-array radar scan configurations (SCs) to maximize the information contents that can be extracted by a 4DVar analysis system from radar observations. For this purpose, we need to consider and compare a variety of different SCs designed over wide ranges of measurement accuracies and resolutions, but all these SCs are constrained by the same limited measurement capability of a given radar. Under this constraint, a simple rule will be used to regulate the trade-offs between measurement accuracy and resolutions. The detailed considerations and related selections of different SCs are described in the following subsection.

4.1. *Radar scan configurations and trade-offs between accuracy and resolutions*

It is well known in Doppler weather radar signal sampling and processing (Doviak and Zrnic, 1993) that a large number of uncorrelated samples need to be acquired to obtain an estimate of radial velocity (along the radar beam direction) in a resolution volume (typically 250 m $\times$ 1° $\times$ 1°) and to reduce the uncertainty (caused by random meteorological scatterers) of the estimated velocity. The total number of samples for all the radar radial-velocity estimates (level-II data) of a volume scan is the number of the estimates multiplying the number of samples needed for each estimate (in each resolution volume). The number of acquirable samples per unit time, however, always has a limit for a given phased-array radar (or any other radar), especially if the radar is designed for multifunction tasks. Constrained by this limit, if the measurement resolution is increased in one of the (spatial and temporal) dimensions without reducing the measurement accuracy, then there must be a proportional decrease of resolution at least in one of the remaining dimensions. Likewise, if the resolution is increased, say, by μ times in a spatial or temporal dimension without changing the resolutions in the remaining dimensions, then the measurement error variance will increase by μ times (because the error variance of an estimate is inversely proportional to the number of independent samples used to yield the estimate according to the well-known central limit theorem in statistics).

Note that each level-II radial-velocity observation is an averaged value (weighted by radar antenna-gain function and reflectivity) over the radar resolution volume, while the true state represented by the model at each grid point is an averaged value over the volume represented by that grid point. When each observation innovation is computed, there can be an error in the observation operator due to the limited model resolution. This error can be attributed to the observation as a representativeness error. The significance of this error depends on the intensity of subgrid turbulence not resolvable by the model grid but sampled by the observations. When the model resolution is close to or higher than the radar resolution, this error should be very small. This type of error is not considered in this paper, so the measurement error considered in the above simple rule is the total observation error in each illustrative example.

By using the above simple rule, nine different SCs are considered based on the radial-velocities scanned by the NWRT phased-array radar on 2 June 2004 when a squall line moved southeastward through the central Oklahoma area. This data set was also used for the illustrative examples in Xu (2007) except that the data are now selected from three consecutive scans (rather than a single volume scan) over 36 km radial range (between $2.8 \le r \le 38.4$ km) along azimuth 130° and elevation 0.75° at 21:0606, 21:0740 and 21:0916 UTC. The volume scan rate is $\Delta\tau = 30$ s per volume. The original radar data have a spatial resolution of 240 m in the radial direction. The data are processed through quality control and then thinned to 720 m resolution (and thus de-correlated) along the radar beam. The estimated observation error standard deviation is $\sigma_{\rm ob} = 2.5$ m s^{-1}. Thus, we have $\sigma_{\rm ob} = 2.5$ m s^{-1}, $\Delta\tau = 30$ s and $\Delta r = 0.72$ km and denote this SC as SC$_{11}$, as listed in the first row and first column of Table 1.

By simultaneously reducing the observation error variance $\sigma_{\rm ob}{}^2$ and the spatial resolution $1/\Delta r$ by 5 (or 25) times, SC$_{11}$ is modified into SC$_{21}$ (or SC$_{31}$) in the second (or third) row in the first column of Table 1. By simultaneously reducing the temporal resolution $1/\Delta\tau$ and increasing the spatial resolution $1/\Delta r$ by two (or three) times, SC$_{i1}$ (for $i = 1, 2, 3$) in first column of Table 1 is further modified into SC$_{i2}$ (or SC$_{i3}$) in the second (or third) column of Table 1. All the nine SCs are constrained by the same radar observation capability (measured by the number of samples acquired per unit time). Since the model domain size is $D = 36$ km in the radial range, the number of observation points within the model domain is $m = D/\Delta r$ as listed for each SC in Table 1. The assimilation time window is set to $\tau = 90$ s, so the number of observation time levels used by the analysis is $J + 1 = \tau/\Delta\tau$ as listed on the top of each column in Table 1. The total number of observations used by the analysis in each assimilation cycle is $m\times(J + 1) = 150$ (30 or 6) for any of the three SCs listed in the first (second or third) row of Table 1. This total number $m\times(J + 1)$ is inversely proportional to the observation error variance $\sigma_{\rm ob}{}^2$ according to the aforementioned simple rule for the trade-offs between observation accuracy and resolutions.

4.2. *Simple prediction model and analysis system used for illustrative examples*

Simple 1-D and 2-D nonlinear advection equations were used in Qiu and Xu (1992) to demonstrate the merits of the adjoint technique (similar to that used in 4DVar) for vector wind retrievals from single-Doppler radar observations. Since then, simplified 2-D and 3-D advection equations have been successfully used with the adjoint technique to retrieve storm winds from single-Doppler radar radial-velocity observations (Xu et al., 1995, 2001a,b). Inspired by these previous studies, a simple

Table 1. Nine scan configurations (SCs) with different settings of σ_{ob}, $\Delta\tau$ and Δr constrained by the same radar observation capability. As described in Sections 4.1 and 4.3, $J + 1 = \tau/\Delta\tau$ is the number of observation time levels used by the analysis, $m = D/\Delta r$ is the number of observation points within the model domain (with $D = 36$ km in the radial range), *DR* is the dispersion part of the relative entropy, *SD* is the Shannon entropy difference, and *DFS* is the number of degrees of freedom for signal. Note that $J + 1 = 1$ for the SCs listed in the third column, so (15) reduces to (7) and the analysis reduces to a 3DVar-type

	$\Delta\tau = 30$s $J+1=\tau/\Delta\tau=3$	$\Delta\tau = 45$s $J+1=\tau/\Delta\tau=2$	$\Delta\tau = 90$s $J+1=\tau/\Delta\tau=1$
$\sigma_{\mathrm{ob}} = 2.5$ m s^{-1}	SC_{11}: $\Delta r = 0.72$ km $m = 50$ $DR = 8.51$ $SD = 14.93$ $DFS = 12.84$	SC_{12}: $\Delta r = 0.48$ km $m = 75$ $DR = 9.06$ $SD = 14.86$ $DFS = 11.61$	SC_{13}: $\Delta r = 0.24$ km $m = 150$ $DR = 8.73$ $SD = 11.70$ $DFS = 5.94$
$\sigma_{\mathrm{ob}} = 2.5/5^{1/2} = 1.1$ m s^{-1}	SC_{21}: $\Delta r = 3.6$ km $m = 10$ $DR = 8.78$ $SD = 14.92$ $DFS = 12.28$	SC_{22}: $\Delta r = 2.4$ km $m = 15$ $DR = 9.87$ $SD = 16.45$ $DFS = 13.17$	SC_{23}: $\Delta r = 1.2$ km $m = 30$ $DR = 8.75$ $SD = 11.74$ $DFS = 5.98$
$\sigma_{\mathrm{ob}} = 0.5$ m s^{-1}	SC_{31}: $\Delta r = 18$ km $m = 2$ $DR = 5.54$ $SD = 7.98$ $DFS = 4.88$	SC_{32}: $\Delta r = 12$ km $m = 3$ $DR = 7.32$ $SD = 10.01$ $DFS = 5.38$	SC_{33}: $\Delta r = 6$ km $m = 6$ $DR = 8.51$ $SD = 11.25$ $DFS = 5.48$

1-D advection equation for the radial velocity is used as the prediction model for the illustrative purpose in this section. The model equation is tangent-linearized and then discretized by using the Lax-Wendroff finite-difference scheme (Strikwerda, 2004). The discretized tangent-linear model equation has following form:

$$\delta v_{i,j+1} = (\delta v_{i+1,j} + \delta v_{i-1,j})/2 - [v_{i,j}(\delta v_{i+1,j} - \delta v_{i-1,j}) + \delta v_{i,j}(v_{i+1,j} - v_{i-1,j})]\Delta t/(2\Delta x) + q_{i,j+1}, \tag{16}$$

where $v_{\mathrm{i,j}}$ is the background velocity at the ith grid point and t_{j}, $\delta v_{\mathrm{i,j}}$ is the error perturbation at the ith grid point and t_{j}, and $q_{\mathrm{i,j+1}}$ is the model error integrated from t_{j} to $t_{\mathrm{j+1}}$ at the ith grid point. The grid spacing is $\Delta x = 2.4$ km and the number of grid points is $n = 16$, so the model domain covers a radial range of 36 km. The initial background field is given simply by the observed radial-velocity, although it can be given by the analysis produced by the Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS, Hodur, 1997) as shown by the illustrative examples in Xu (2007). The qualitative aspect of the computed information contents (in Table 1) is independent of the choice of the background field.

As mentioned earlier, the assimilation time window is $\tau =$ 90 s. For the radar SCs in the first (or second) column of Table 1, the observation time interval is $\Delta\tau = 30$ (or 45) s and $J + 1 = \tau/\Delta\tau = 3$ (or 2), so each analysis time window contains three (or two) time levels of observations. In this case, the model integration time step is selected to match the observation time interval, that is, $\Delta t = \Delta\tau = 30$ (or 45) s, and this selection also satisfies the Courant–Friedrichs–Lewy condition for computational stability. For the radar SCs in the third column of Table 1, the observation time interval is $\Delta\tau = 90$ s and $J + 1 = 1$, so there is only a single time level of observations in each analysis time window. In this case, (15) reduces to (7) for the SVD of $\mathbf{M}$ and the subsequent computations of the information contents.

The initial background error covariance is modelled by a Gaussian correlation function with the de-correlation length set to $L = 3\Delta x = 7.2$ km and the error standard deviation set to $\sigma = 8.3$ m s^{-1} (the same value as used for the illustrative examples in Xu, 2007). We assume that the model error is Gaussian white in time in the original differential equation and is independent of the initial condition. Note that the time-integration of a Gaussian white-noise process is a Brownian motion process, so the variance of the time-integrated model error in (16) is proportional to Δt (see sections 3.5 and 3.8 of Jazwinski, 1970). As $\Delta t = \Delta\tau$, the model error variance is set to $q^2(3\Delta\tau/\tau) = 3q^2/(J + 1)$ with $q = 1$ m s^{-2}. Furthermore, the time-integrated model errors are not correlated between different time intervals and not correlated with the initial background error. In this case, $\mathbf{Q}$ is block-diagonal as given in (14b). If the model errors are also not correlated between different grid points, then $\mathbf{Q}_j = q^2(3\Delta\tau/\tau)\mathbf{I}$ in (14b), where $\mathbf{I}$ is the identity matrix in R^n.

4.3. *Information contents computed for illustrative examples*

With the above prediction model and analysis system, the singular values are computed for the extended $\mathbf{M}$ in (15) and then used to compute the information contents measured (i) by the dispersion part of the relative entropy, that is,

$$DR \equiv \sum [\ln(1+\lambda_i^2) - \lambda_i^2/(1+\lambda_i^2)]/2 = SD - DFS/2 \quad (17)$$

according to (5) and (8), (ii) by the Shannon entropy difference, *SD* according to (6), and (iii) by the degrees of freedom for signal *DFS* according to (9). The singular values used in (17), (6) and (9) are obtained by the SVD of the extended $\mathbf{M}$ in (20) with $\mathbf{Q}$ given by (14b), $\mathbf{Q}_j = q^2(3\Delta\tau/\tau)\mathbf{I}$ and $q = 1$ m s^{-2}. $\mathbf{B}_0$ is modelled by a Gaussian correlation function with the de-correlation length set to $L = 3\Delta x = 7.2$ km and the error standard deviation set to $\sigma = 8.3$ m s^{-1} (the same value as used the illustrative examples in Xu 2007). The results are listed for each SC in Table 1. As shown, SC_{22} provides the highest information content (measured by *DR*, *SD* or *DFS*), so SC_{22} is the optimal SC among the nine SCs in Table 1. Further changing the trade-offs between the observation accuracy and resolutions beyond the ranges covered by the nines SCs in Table 1 can only further reduce the information contents (not shown). This further confirms the optimality of SC_{22} in the parameter space of (σ_{ob}, $\Delta\tau$, Δr) constrained by the simple rule described in Section 3.1, although this optimality is only coarsely resolved in the parameter space by the nine SCs in Table 1. The existence of optimal SC should be independent of the choice of the information measure (*DR*, *SD* or *DFS*), but different information measures may lead to slightly different selections of optimal SC in the parameter space if the parameter values of (σ_{ob}, $\Delta\tau$, Δr) are fine-tuned for the optimal SC at much higher resolutions than those represented by the nine SCs in Table 1.

The *j*th truncated values of *DR*, *SD* and *DFS* can be defined and denoted by

$$DR_j \equiv \sum\nolimits_j [\ln(1+\lambda_i^2) - \lambda_i^2/(1+\lambda_i^2)]/2, \quad (18)$$

$$SD_j \equiv \sum\nolimits_j [\ln(1+\lambda_i^2)]/2 \quad (19)$$

and

$$DFS_j \equiv \sum\nolimits_j \lambda_i^2/(1+\lambda_i^2), \quad (20)$$

respectively. Here, $\sum_j$ denotes the summation over *i* from 1 to *j*. These truncated information measures are plotted as functions of the truncation number *j* for five different SCs in Figs. 1a–c to show how the information contents accumulate as *j* increases. As explained in Section 2.1, the truncation number is bounded by the rank of $\mathbf{M}$, that is, $r = \text{rank}(\mathbf{M})$. Note that $\mathbf{M}$ defined in (15)

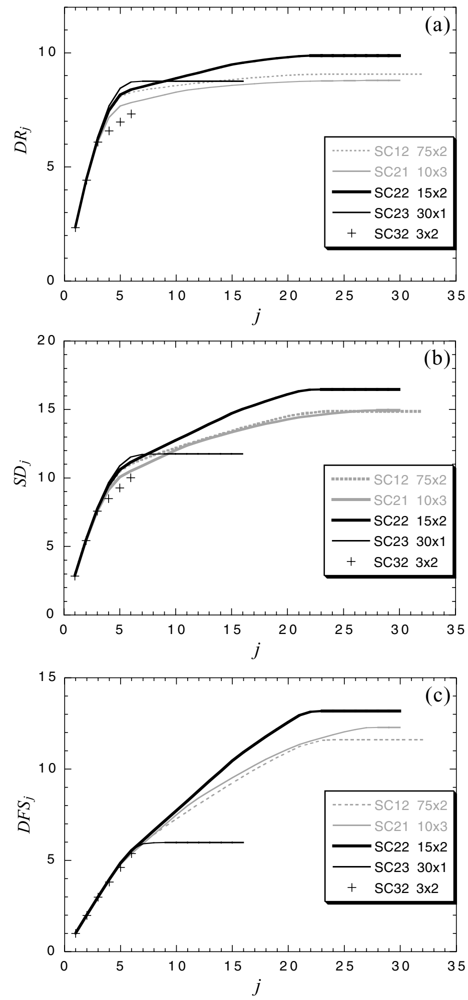

Fig. 1. (a) DR_j, (b) SD_j and (b) DFS_j plotted as functions of the truncation number *j* for SC_{12}, SC_{21}, SC_{22}, SC_{23} and SC_{32}. The parameter settings for $\mathbf{Q}_j$ and $\mathbf{B}_0$ are the same as in Table 1. The details are described in the second paragraph of Section 4.3.

is a $(J+1)m \times n(J+1)$ matrix, so $r \le r_{mn} \equiv \min(m, n) \times (J+1)$. For SC_{12}, SC_{21}, SC_{22}, SC_{23} and SC_{32}, we have $r_{mn} = 32$, 30, 30, 16 and 6, respectively. According to our computations, the singular values can be as small as 10^{-6} but they are always above zero. Since all the computed singular values (down to 10^{-6}) are considered to be 'non-zero' in the plotting, each curve is plotted over its full range from $j = 1$ to r_{mn}. We may consider that a singular value is effectively zero if it is smaller than 1% of the leading singular value λ_1. The effective rank can be then defined as the number of effective non-zero singular value, and denoted by r_e. In this case, we have $r_e = 24, 28, 23, 8$ and 6 for SC_{12}, SC_{21}, SC_{22}, SC_{23} and SC_{32}, respectively, according to the computed singular values. As we can see from each panel of Fig. 1, each plotted curve becomes flat as *j* reaches r_e except for SC_{32}. This suggests that the information accumulation becomes effectively saturated as the truncation number reaches r_e (except for SC_{32}).

This is a common feature for the three information measures, although the information saturation (at $j = r_e$) is revealed most distinctly by the DFS_j curve and least distinctly by the RD_j curve for each given SC, as shown in Figs. 1a–c.

The above information saturation implies information redundancy (that is, redundancy in the total number of pieces of information provided by the observations) or observation resolution redundancy. As in Xu (2007), the redundancy can be quantified by $(N_m - r_e)/N_m$ in terms of percentage of reducible number of observations, where $N_m \equiv m \times (J + 1)$ is the total number of observations and $N_m \geq r_{mn} \geq r_e$. For SC_{12}, we have $N_m = 150$ $(> r_{mn} = 32)$ and $(N_m - r_e)/N_m = (150 - 24)/150 = 84\%$, so the redundancy is very high. For SC_{21}, we have $N_m = 30$ $(= r_{mn})$ and $(N_m - r_e)/N_m = (30 - 28)/30 = 6.7\%$, so the redundancy is quite small. For SC_{22}—the optimal SC, we have $N_m = 30$ $(= r_{mn})$ and $(N_m - r_e)/N_m = (30 - 23)/30 = 23.3\%$, so the redundancy is higher than that for SC_{21}. For SC_{23}, we have $N_m = 30$ $(> r_{mn} = n \times 1 = 16)$ and $(N_m - r_e)/N_m = (30 - 8)/30 = 73.3\%$, so the redundancy is very high. For SC_{32}, we have $N_m = 6$ and $(N_m - r_e)/N_m = 0$, so there is no resolution redundancy, but the information content provided by SC_{32} is the smallest among the five SCs plotted in Fig. 1. Thus, as the information content is maximized by the optimal SC for a give analysis system, the information redundancy is not necessarily very small.

Note that, with $J + 1 = 1$, SC_{23} provides only a single time level of observations for the analysis. In this case, as explained in Section 4.2, (15) reduces to (7) and the analysis reduces to a 3DVar-type. As shown in Table 1 and Fig. 1, when the SC changes from SC_{22} to SC_{23} and thus the analysis reduces from a 4DVar-type (with an imperfect model) to a 3DVar-type, the information content reduces significantly (by about 10% in *DR*, 30% in *SD*, or 50% in *DFS*) and the resolution redundancy increases drastically (from 6.7 to 73.3%). This exemplifies the property remarked in Section 3.2(1).

In the above examples, the model error parameter is set to $q = 1$ m s^{-2} (see Section 4.2). If the model is assumed to be perfect, then $q = 0$. By setting $q = 0$, 1, and 2 m s^{-2}, respectively, three SD_j curves are plotted for SC_{22} in Fig. 2. By comparing the ending *SD* values of these curves, we can see that the *SD*-measured information content decreases from 16.45 to 11.19 (or increases from 16.45 to 23.31) as q decreases from 1 to 0 m s^{-2} (or increases from 1 to 2 m s^{-2}). Furthermore, as the model becomes perfect ($q = 0$), the upper bound of rank(**M**) reduces from $(J + 1)\min(m, n) = 30$ to $\min[(J + 1)m, n] = n = 16$, which is very close to the number of observations ($m = 15$) acquired by SC_{22} at a single time level. In this case, the reduced *SD* (= 11.19) is still significantly larger than the *SD* (= 9.84) computed by (6)–(7) from a single time level of observations acquired by SC_{22}. The *DFS*-measured information content is reduced more significantly (from 13.17 to 5.73) and becomes very close to the *DFS* (= 5.58) computed by (9) with (7) from a single time level of observations acquired by SC_{22}. Also, as q reduces to 0, the effective rank reduces to $r_e = 8$ and

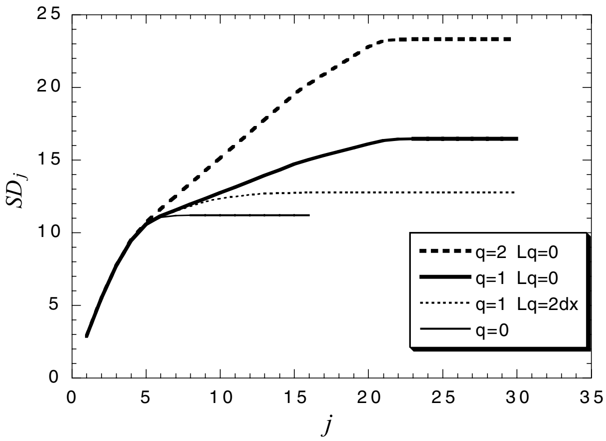

Fig. 2. SD_j plotted as functions of the truncation number j for SC_{22} with four different settings for the model error covariance matrix: (i) $q = 2$ m s^{-2} and $L_q = 0$, (ii) $q = 1$ m s^{-2} and $L_q = 0$, (iii) $q = 1$ m s^{-2} and $L_q = 2\Delta x$, and (iv) $q = 0$. The details are described in the last two paragraphs of Section 4.3.

the resolution redundancy becomes very high (73.3%). These results confirm the remarks in Section 3.2(2),

In all the previous examples, the model errors are assumed to be spatially uncorrelated and the model error covariance matrix is given by $\mathbf{Q}_j = q^2(3\Delta\tau/\tau)^2\mathbf{I}$ [see (14b)] where **I** is the identity matrix in R^n. This assumption can be relaxed if **I** is replaced by a non-diagonal matrix constructed by a Gaussian correlation function with a given de-correlation length L_q. When $L_q = 0$, $\mathbf{Q}_j$ reduces to $q^2(3\Delta\tau/\tau)^2\mathbf{I}$ as in all the previous examples. By setting L_q to $2\Delta x$ (= 4.8 km) with $q = 1$ m s^{-2}, the SD_j curve is computed for SC_{22} and plotted in Fig. 2. By comparing this curve with that for $L_q = 0$, we can see that as L_q increases from 0 to $2\Delta x$, the information content decreases from 16.45 to 12.77, the effective rank reduces from $r_e = 23$ to 16, and the resolution redundancy increases from 23.3 to 46.6%. Thus, as the model errors become more correlated, the analysis extracts less information from the observations, and the observations become more redundant (in resolution or total number of pieces). This exemplifies and extends the property remarked in Section 3.2.

5. Conclusions

In this paper, the previously derived formulations for the relative entropy and *SD* are reviewed in connection with another information measure – degrees of freedom for signal (*DFS*). As the *SD*, the *DFS* measures only the dispersion part of the information content. However, unlike the *SD*, the *DFS* is not additive for successive inclusions of observations into the analyses, so it is not as convenient as the *SD* for computing cumulative information content from successive groups of observations. The relative entropy is also not additive (as pointed out by James Purser, personal communication, 2006), but it can measure both the signal and dispersion parts of the information content, and these two parts can be related to the improvements made by the analysis to the mean and covariance, respectively, as shown in Xu (2007). As remarked in Section 2.2, the dispersion part of the relative entropy (*DR*) can be related to *SD* and *DFS* by $DR = SD - DFS/2$ if the Bayesian optimal-estimation problem is linear or nearly linear. The signal part of the relative entropy has the same form as the background term J_b in the costfunction used by the Bayesian optimal analysis. As J_b is constructed by the analysis increment, the signal part of the relative entropy depends on the observations used by the analysis. By considering all possible realizations of observation, the statistical average (expectation) of J_b gives *DFS*/2 [according to the definition of *DFS* in (2.46) of Rodgers, 2000]. Thus, for a linear (or nearly linear) Bayesian optimal-estimation problem, the statistical average of the relative entropy reduces to *SD* [as shown in (10)].

The previous formulations are extended to measure information contents from observations extracted by a 4DVar analysis. The extended formulations reveal the following properties (as remarked in Section 3.2): (i) The information content extracted from a set of 4-D observations by a 4DVar analysis with an imperfect model can be much higher than that from any subset of 3-D observations extracted by a 3DVar analysis. (ii) The information content extracted by a 4DVar analysis from a given set of 4-D observations increases (or decreases) as the model error increases (or decreases). (iii) The information content extracted by a 4DVar analysis from a given set of 4-D observations will increase (or decrease) as the model errors become less (or more) correlated in 4-D space. The above properties are exemplified by illustrative examples in Section 4. In addition, the illustrative examples show that the observations become less (or more) redundant as the information content increases (or decreases) in each of the above-described scenarios.

The illustrative examples also show that the extended formulations and above properties can be useful for designing optimum phased-array radar scan configurations (SCs) to maximize the extractable information contents from radar observations by a 4DVar analysis system. In these illustrative examples, different SCs are designed over wide ranges of observation accuracies and resolutions (without considering the observation operator error as explained in Section 4.1). As all these SCs are constrained by the same limited measurement capability of a given phased-array radar, there are trade-offs between measurement accuracy and resolutions. Under this constraint, the existence of optimal SC is demonstrated by the illustrative examples. The optimal SC is shown to be not sensitive to the choice of the information measure (*DR*, *SD* or *DFS*), although different information measures may lead to slightly different selections of optimal SC (not yet resolved by the coarsely selected SCs in Table 1). On the other hand, the computed information contents and their determined optimal SC are quite sensitive to the background and model error covariances and to the spatial and temporal resolutions of the analysis system.

Generally speaking, the optimal radar SC should have relatively high spatial and temporal resolutions for a high-resolution assimilation system. In this case, the observation space and analysis space can both become very large and the scaled observation operator matrix [see (15)] will be too large to be used for computing the singular values needed in the formulations derived in this paper. This limits their practical applications, and the information contents need to be estimated by other efficient methods (Fisher, 2003; Healy and Thepaut, 2006). Furthermore, the accuracy of the computed information content depends on the accuracies of the estimated error covariances. When the formulations are extended for 4DVar in this paper, it is assumed that the model error covariance is given accurately. In practice, however, the model error covariance is often unknown but assumed to be zero in 4DVar. When the perfect-model assumption is used for an imperfect model, the analysis error covariance is underestimated in the perfect-model 4DVar and the true analysis error covariance contains an additional term due to the neglected model error [see eq. (9) of Healy and White (2005)]. This complication is not considered in this paper, and the related problems require further detailed studies. Accurately estimating the required error statistics should be very important for the concerned design of optimal SC for 4DVar radar data assimilation. As shown recently by Xu et al. (2007), the conventional innovation method can be extended to estimate radar radial-velocity observation and background error covariances for a 3DVar system, but how to estimate the error statistics (especially the model error statistics) for a 4DVar radar data assimilation system still confronts many challenging issues and requires great research endeavors beyond the scope of this paper.

6. Acknowledgments

The authors are thankful to Dr. James Purser and Prof. Richard Kleeman for their insightful comments and suggestions that improved the presentation of the results. The computational resource was provided by the OU Supercomputing Center for Education & Research at the University of Oklahoma. The research work was supported by the ONR Grant N000140410312 to CIMMS, the University of Oklahoma. Funding was also provided to CIMMS by NOAA under NOAA-University of Oklahoma Cooperative Agreement #NA17RJ1227, U.S. Department of Commerce.

References

Bennett, A. F. 1992. *Inverse Method in Physical Oceanography*. Cambridge University Press, New York, 346 pp.

Cohn, S. 1997. An introduction to estimation theory. *J. Metrorol. Soc. Japan* **75**, 257–288.

Courtier, P. 1997. Dual formulation of four dimensional variational assimilation. *Quart. J. Roy. Meteorol. Soc.* **123**, 2449–2461.

Daley, R. and Barker, E. 2001. NAVDAS: formulation and diagnostics. *Mon. Wea. Rev.* **129**, 869–883.

Doviak, J. D. and Zrnic, D. S. 1993. *Doppler Radar and Weather Observations*. 2nd Edition. Academic Press, New York , 562 pp.

Eyre, J. R. 1990. The information content of data from satellite sounding systems. A simulation study. *Quart. J. Roy. Meteorol. Soc.* **116**, 401–434.

Fisher, M. 2003. Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems. *Technical Memorandum No. 397*. ECMWF, Shinfield Park, Reading . [available on line http://www.ecmwf.int/publications/library/do/references/show?id=83951]

Healy, S. B. and White, A. A. 2005. Use of discrete Fourier transforms in the 1D-Var retrieval problem. *Quart. J. Roy. Meteorol. Soc.* **131**, 63–72.

Healy, S. B. and Thépaut, J.-N. 2006. Assimilation experiments with CHAMP GPS radio occultation measurements. *Quart. J. Roy. Meteorol. Soc.* **132**, 605–623.

Hodur, R. M. 1997. The Naval Research Laboratory's coupled ocean/atmosphere mesoscale prediction system (COAMPS). *Mon. Wea. Rev.* **125**, 1414–1430.

Huang, H.-L. and Purser, R. J. 1996. Objective measures of the information density of satellite data. *Meteorol. Atmos. Phys.* **60**, 105–117.

Jazwinski, A. H. 1970. *Stochastic Processes and Filtering Theory*. Academic Press, San Diego, 376 pp.

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

LeDimet, F. X. and Talagrand, O. 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus* **38**A, 97–111.

Lewis, J. and Derber, J. 1985. The use of adjoint equations to solve a variational adjustment problem with advective constraint. *Tellus* **37**A, 309–322.

Lorenc, A. 1981. A global three-dimensional multivariate statistical analysis system. *Mon. Wea. Rev.* **109**, 701–721.

Lorenc, A. C. 2003. Modeling of error covariances by 4D-Var data assimilation. *Quart. J. Roy. Meteorol. Soc.* **129**, 3167–3182.

Majda, A. J., Kleeman, R. and Cai, D. 2002. A mathematical framework for quantifying predictability through relative entropy. *Methods Appl. Anal.* **9**, 425–444.

Peckham, G. 1974. The information content of remote measurements of atmospheric temperature by satellite infra-red radiometry and optimum radiometer configurations. *Quart. J. Roy. Meteorol. Soc.* **100**, 406–419.

Purser, R. J., Parrish, D. F. and Masutani, M. 2000. Meteorological observational data compression; An alternative to conventional "super-Obbing". *Office Note 430*, National Centers for Environmental Prediction, Camp Springs, MD, 12 pp. [available on line http://www.emc.ncep.noaa.gov/officenotes/FullTOC.html#2000]

Qiu, C. and Xu, Q. 1992. A simple adjoint method of wind analysis for single-Doppler data. *J. Atmos. Oceanic Technol.* **9**, 588–598.

Rodgers, C. D. 2000. *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific Publishing Co. Ltd., London, 256 pp.

Shannon, C. E. 1949. Communication in the presence of noise. *Proc. I.C.E.* **37**, 10–21.

Strikwerda, J. C. 2004. *Finite Difference Schemes and Partial Differential Equations*. Chapman & Hall, Philadelphia, 450 pp.

Wahba, G., Johnson, D. R., Gao, F. and Gong, J. 1995. Adaptive tuning of numerical weather prediction models: randomized GCV in three- and four-dimensional data assimilation. *Mon. Wea. Rev.* **123**, 3358–3370.

Xu, Q. 2007. Measuring information content from observations for data assimilation: relative entropy versus Shannon entropy difference. *Tellus* 59A, 198–209.

Xu, Q., Gu, H. and Qiu, C. J. 2001a. Simple adjoint retrievals of wet-microburst winds and gust-front winds from single-Doppler radar data. *J. Appl. Meteorol.* **40**, 1485–1499.

Xu, Q., Gu, H. and Yang, S. 2001b. Simple adjoint method for three-dimensional wind retrievals from single-Doppler radar. *Quart. J. Roy. Meteorol. Soc.* **127**, 1053–1067.

Xu, Q., Nai, K. and Wei, L. 2007. An innovation method for estimating radar radial-velocity observation error and background wind error covariances. *Quart. J. Roy. Meteorol. Soc.* **133**, 407–415.

Xu, Q., Qiu, C., Gu, H. D. and Yu, J. 1995. Simple adjoint retrievals of microburst winds from single-Doppler radar data. *Mon. Wea. Rev.* **123**, 1822–1833.

Zrnic, D. S., Kimpel, J. F., Forsyth, D. F., Shapiro, A., Crain, G. and co-authors. 2007. Agile-beam phased array radar for weather observations. *Bull. Amer. Meteorol. Soc.* **88**, 1753–1766.