6.9 Estimation of analysis error variances with the Physical-space Statistical Analysis System*

Ricardo Todling†, Runhua Yang†, Jing Guo†, and Stephen E. Cohn
Data Assimilation Office, NASA/GSFC, Greenbelt, MD, USA

1 Introduction

A few years ago the National Center for Environmental Prediction (NCEP, former National Meteorological Center; Parrish and Derber 1992) and the European Centre for Medium-Range Weather Forecasts (ECMWF, Courtier et al. 1998) converted their analysis systems from optimal interpolation (OI) to 3-dimensional variational spectral statistical interpolation (spectral 3D-Var). In the same spirit, the NASA/Data Assimilation Office has replaced its OI system with the physical-space statistical analysis system (PSAS), which is a variant of the spectral 3D-Var schemes operating directly in physical space instead of spectral space (Cohn et al. 1998). One of the purposes of these new schemes is to have a data assimilation system capable of performing global analysis, thus avoiding the questionable local approximations required in OI. Both the spectral and physical-space approaches represent improvements over OI, but do not readily provide computationally feasible means of estimating analysis error standard deviations due to their implementation formulation.

The importance of determining analysis errors is manifold. The most common motivation is to facilitate development of advanced data assimilation schemes. Any attempt to evolve error covariances, in the direction of Kalman filter-like methods, requires knowledge of analysis errors and, more generally, of the complete analysis error covariance. These errors are also needed to provide the correct norm for singular-vector-based forecast systems (Barkmeijer et al. 1998), and can also be used in conjunction with the breeding method (Toth and Kalnay 1997). But if for nothing else, these errors provide a rough measure of analysis quality. Knowledge of analysis accuracy is fundamental for development and validation of new observing instruments, and can also be used as background errors in retrieval algorithms.

In this article we discuss results of analysis error estimates calculated using two distinct methods implemented in PSAS. The first method resembles Fisher's procedure but is designed in the proper PSAS context and with the PSAS philosophy in mind, that is, perform operations in observation space. The second method is that proposed by Riishøjgaard (2000). In section 2 we present these methods. Section 3 shows some results, and remarks are made in section 4.

2 Technicalities

In analysis systems such as PSAS an estimate of the analysis errors can be obtained from the diagonal of the $n \times n$ analysis error covariance matrix, $\mathbf{P}^a$, given by

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{P}^f \mathbf{H}^T \mathbf{\Gamma}^{-1} \mathbf{H} \mathbf{P}^f \,, \tag{1}$$

[e.g., Todling and Cohn (1994), eq. (3.19d)], where $\mathbf{\Gamma}$ is the innovation covariance matrix given by

$$\mathbf{\Gamma} \equiv \mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R} \,, \tag{2}$$

and $\mathbf{P}^f$ and $\mathbf{R}$ are the $n \times n$ forecast and $p \times p$ observation error covariance matrices, respectively. Equation (1) is not directly implemented in any operational assimilation system, particularly due to the typical size of the analysis state vector, i.e., $n \approx 10^6$. If anything, advanced data assimilation systems have $\mathbf{P}^a$, or its inverse as for analysis systems such as ECMWF's, implemented as an operator, thus allowing for matrix-vector operations. This feature has been exploited in Barkmeijer et al. (1998) and Klinker et al. (1998) to obtain meaningful singular vectors for ensemble forecasting purposes, and to obtain estimates of analysis errors in a specific direction determined by the singular vectors, respectively. These works were not aimed at estimating analysis errors as complete fields representing the analysis accuracy. Fisher (1996) presented an attempt to derive complete analysis error variance fields, in the context of spectral variational analysis. Fisher's algorithm is based on performing

*Corresponding author address: Dr. Ricardo Todling, Data Assimilation Office, NASA/GSFC, Code 910.3, Greenbelt, MD 20771. *e-mail:* todling@dao.gsfc.nasa.gov.

†Affiliation: Science Applications International Corporation.

a truncated eigendecomposition of the spectral 3D-Var Hessian. The algorithm is computationally expensive since it operates in the analysis state space.

In PSAS, (1) is available in operator form, particularly the so called *error reduction term* $\Delta\mathbf{P}^a \equiv \mathbf{P}^f\mathbf{H}^T\mathbf{\Gamma}^{-1}\mathbf{H}\mathbf{P}^f$. Therefore, we can in principle use a Lanczos-type algorithm to calculate the dominant eigenmodes of this matrix and generate conservative estimates of analysis error variances in a correspondingly equivalent procedure as that developed by Fisher for the spectral 3D-Var Hessian. However, this would suffer from similar computational demands as it would operate in the analysis space. Alternatively, we can follow the same rational used behind transforming the analysis equations from OI into its PSAS equivalent and have the problem in hand be solved in observation space.

Let us write the eigendecomposition of the error reduction term $\Delta\mathbf{P}^a$ as

$$\Delta\mathbf{P}^a\mathbf{Y} = \mathbf{Y}\mathbf{\Lambda}\,, \tag{3}$$

where $\mathbf{\Lambda}$ is the $p\times p$ diagonal eigenvalue matrix and $\mathbf{Y}$ is the $n \times p$ matrix of eigenvectors; the size of these matrices is determined by the rank of $\Delta\mathbf{P}^a$. Defining

$$\mathbf{X} \equiv \mathbf{\Gamma}^{-1}\mathbf{H}\mathbf{P}^f\mathbf{Y}\,, \tag{4}$$

we arrive at the following generalized eigenvalue problem involving $\mathbf{X}$:

$$\mathbf{H}\mathbf{P}^f\mathbf{P}^f\mathbf{H}^T\mathbf{X} = \mathbf{\Gamma}\mathbf{X}\mathbf{\Lambda}\,, \tag{5}$$

where $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues, and we notice that this eigendecomposition operates in observation space. Assuming that only a certain number $\tilde{p}$ of dominant eigenmodes are needed to obtain a reliable approximation of the error reduction matrix, and representing these truncated eigendecomposition by the pair of eigenvalue-eigenvector matrices $\tilde{\mathbf{\Lambda}}$ and $\tilde{\mathbf{X}}$ of dimensions $\tilde{p} \times \tilde{p}$ and $p \times \tilde{p}$, respectively, the approximate error reduction matrix, denoted by $\Delta\mathbf{S}^a$, can be obtained as

$$\Delta\mathbf{S}^a \equiv \mathbf{P}^f\mathbf{H}^T\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T\mathbf{H}\mathbf{P}^f\,. \tag{6}$$

In writing the expression above we notice that $\tilde{\mathbf{X}}$ is $\mathbf{\Gamma}$-normalized, that is, $\tilde{\mathbf{X}}^T\mathbf{\Gamma}\tilde{\mathbf{X}} = \mathbf{I}_{\tilde{p}\times\tilde{p}}$. The amount of error reduction variance explained by this approximation follows immediately from the eigenvalue matrix through the ratio $trace(\Delta\mathbf{S}^a)/trace(\Delta\mathbf{P}^a) = trace(\tilde{\mathbf{\Lambda}})/trace(\Delta\mathbf{P}^a)$. In practice, an estimate of the trace of $\Delta\mathbf{P}^a$ is needed and can be obtained, as in our implementation in PSAS, by using randomized trace estimation (Girard 1991).

Riishøjgaard (2000) proposed another algorithm to derive estimates of analysis error variances. Instead of using a spectral truncation of the error reduction matrix, Riishøjgaard's procedure is based on using a spectral truncation of the innovation covariance matrix. In this case, we can write the partial eigendecomposition of $\mathbf{\Gamma}$ as

$$\mathbf{\Gamma}\tilde{\mathbf{Z}} = \tilde{\mathbf{Z}}\tilde{\mathbf{\Omega}}\,, \tag{7}$$

where $\tilde{\mathbf{\Omega}}$ is the $p' \times p'$ diagonal matrix the p' dominant eigenvalues, $\tilde{\mathbf{Z}}$ is the $p \times p'$ matrix whose columns correspond to the p' dominant eigenvectors of $\mathbf{\Gamma}$. Therefore, similarly to (6), Riishøjgaard's approximation for the error reduction term in (1) is given by

$$\Delta\mathbf{S}^a \equiv \mathbf{P}^f\mathbf{H}^T\tilde{\mathbf{Z}}\tilde{\mathbf{\Omega}}^{-1}\tilde{\mathbf{Z}}^T\mathbf{H}\mathbf{P}^f\,. \tag{8}$$

Here, the trace of $\Delta\mathbf{S}^a$ cannot be obtained directly from the eigenvalue matrix $\tilde{\mathbf{\Omega}}$ as this relates to the trace of the innovation covariance. This is one different between approximations (6) and (8), that is, the amount of explained error reduction variance in Riishøjgaard's approximation cannot be assessed directly from the eigenvalues. Correspondingly, the eigenvector matrix used in (8) contains structures that are not necessarily the dominant structures in the error reduction matrix. This subtlety has been recognized by Riishøjgaard (2000) and may be of relevance depending on the observing system configuration and error covariances involved in the problem. Similar remarks are made in Daley and Barker (2000) when comparing two procedures for analysis error variance estimation referred to by these authors as global and local approximations, which essentially correspond to variants of approximations (6) and (8), respectively.

An important consideration in developing approximations to estimate analysis error variances is computational cost. In both of the approximation above there are two main parts contributing to computation burden: (i) the corresponding eigendecompositions; and (ii) the operations of $\mathbf{P}^f\mathbf{H}^T$ to as many (scaled) eigenvectors are required. The latter cost is equally incurred in both schemes above, therefore, comparatively the final cost is determined by the cost of solving the different types of eigenvalue problems in these two approximations. The generalized eigenvalue problem in (5) requires application of the operator $\mathbf{H}\mathbf{P}^f\mathbf{P}^f\mathbf{H}^T$ to a vector at each iteration of the Lanczos-type procedure. Furthermore, (5) also requires a conjugate gradient solution at each iteration of the Lanczos-type procedure (e.g., Lehoucq et al. 1997). And finally, after the eigenvalues have converged to the requested accuracy, the converged eigenvectors need to be

rescaled by the "mass" matrix $\mathbf{\Gamma}$. Conversely, in Riishøjgaard's procedure the only operation needed is the application of $\mathbf{\Gamma}$ to a vector within the Lanczos iterations. This renders Riishøjgaard's procedure considerably more attractive from a computational standpoint.

3 Results

In this section we show results obtained with the two procedures described in the previous section for deriving estimates of analysis error variances from PSAS. We seek for estimates of geopotential height analysis error variances at levels 700 mb, 500 mb, 400 mb, and 300 mb induced by the assimilation of radiosondes geopotential heights and DAOTOVS geopotential height retrievals, between these four levels, for 0Z, 1 February 1998. Fig. 1 displays the distribution of observations at 500 mb: radiosondes (top); DAOTOVS (bottom). Notice the almost inexistent overlap between the two types of observing systems due to shadowing of retrievals in the neighborhood of the radiosonde data. Also, the retrievals are thinned to a resolution of roughly 5^o longitude by 4^o latitude and therefore become comparatively less dense than the radiosonde network over regions such as North America, Europe and Asia.

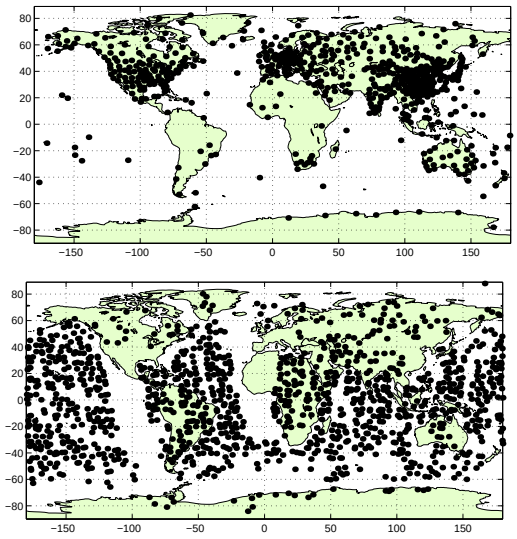

Figure 1: Map of 500 hPa geopotential height observations from the radiosonde network (top) and TOVS retrievals (bottom) on 0Z, 1 February 1998.

The structure of analysis errors is dependent on the observing system as well as in the forecast and observation covariances. Particularly, these errors are largely dependent on the forecast error standard deviations. The top panel of Fig. 2 shows the 500 mb geopotential height error standard deviation for the date and time considered here. It is relevant to mention that currently, in PSAS, these errors are representative of a monthly mean error rather than an actual snap shot, however, this should not affect the qualitative conclusions drawn from the results discussed below.

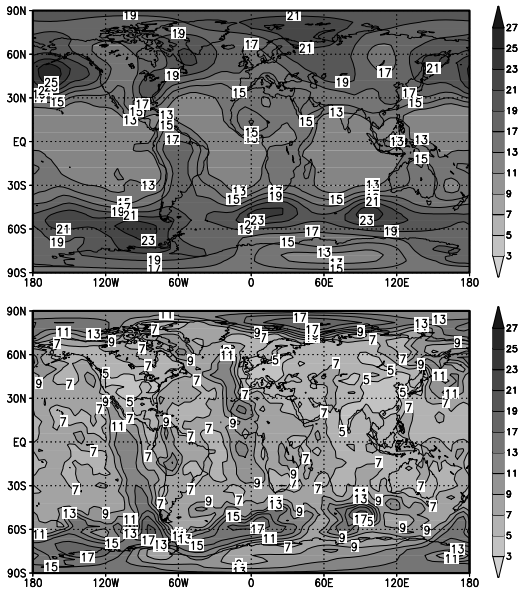

Figure 2: Map of 500 hPa geopotential height forecast (top) and analysis (bottom) error standard deviations; analysis errors result correspond to exact result.

The four-level analysis case-study here includes a total of 7203 observations. In this case, it is still possible to calculate the exact analysis error variances without using either one of the approximations in the previous section. The bottom panel in Fig. 2 shows the exact geopotential height analysis error standard deviation at 500 mb. Comparing both panels in the figure we see the considerable error reduction due to the observations; a comparison with Fig. 1 shows the not surprising result that most of the error reduction occurs over the observing network — even the signature of the satellite tracks are visible in the structure of the analysis error standard deviations. The largest error reduction occurs over the radiosonde network. This is attributed to essentially three factors: the higher density of the radiosonde network over that of the retrievals; the slightly more accurate nature of the radiosonde data reflected in their observation error standard deviations; and the horizontally correlated nature of the retrieval errors, implying in

the retrievals having smaller information content than the horizontally uncorrelated radiosonde observations.

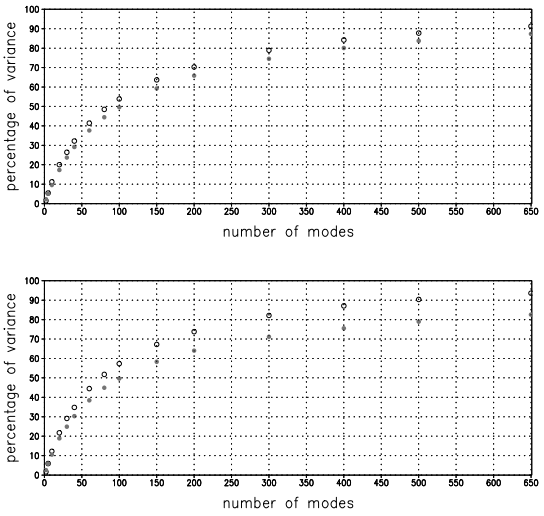

Figure 3: Percentage of total analysis error reduction variance captured when using (6; open circles) and (8; closed circles).

The significance of the approximations in (6) and (8) can be captured by looking at how much of the total error reduction variance is represented when using a certain number of eigenmodes in each approximation. That is, we can plot the quotient $100 \times trace(\Delta \mathbf{S}^a)/trace(\Delta \mathbf{P}^a)$, for each $\Delta \mathbf{S}^a$ calculated for a given number of modes, and for a each procedure. The top panel of Fig. 3 displays this quotient for the approximation in (6; open circles), and that in (8; closed circles). We see that the approximation solving the generalized eigenvalue problem always captures more variance than Riishøjgaard's approximation, but the difference does not seem to be very significant. As a matter of fact, it is difficult to see the benefits of (6) over (8) when looking at a map of their corresponding analysis error standard deviations (not shown). This conclusion is certainly dependent on the present configuration of PSAS as determined by its forecast and observation error covariances. We illustrate this point by showing in the bottom panel of Fig. 3 how the percentage variance quotient changes when the observation errors for the retrievals are artificially increased, that is, when the retrievals are made more inaccurate. In this case, the distinction between the two approximations becomes more relevant and considerably more variance is captured by basing the approximation in the solution of the generalized eigenvalue problem.

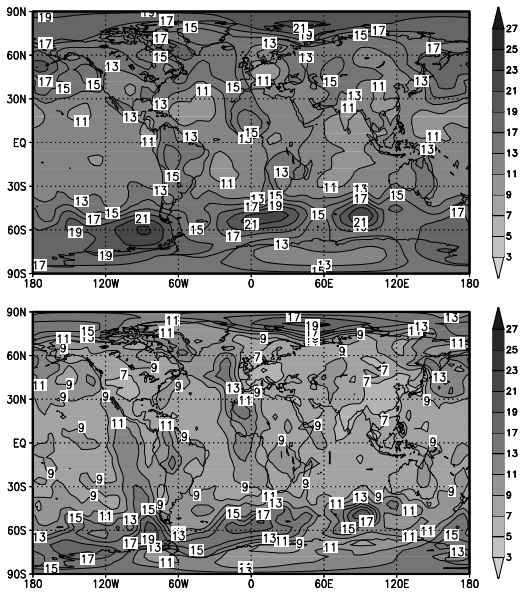

Figure 4: Maps of approximate analysis errors calculated using Riishøjgaard's procedure: top panel obtained when 60 modes are used in (8); bottom panel obtained when 650 modes are used (8). error standard deviations; analysis errors result correspond to exact result.

As illustration of analysis error standard deviations obtained with Riishøjgaad's procedure Fig. 4 displays results when 60 eigenmodes (top panel) and 650 eigenmodes are included in the approximation. With 60 eigenmodes a lot of the forecast error signature is still present, particularly over the Southern Hemisphere high latitudes where observations are not very dense (see Fig. 1). However, with 650 modes, which is less than 10% percent of the total number of observations, most of the structure in the exact analysis error standard deviations displayed in the lower panel of Fig. 2 is capture.

4 Closing remarks

We have compared two methods to calculate approximate analysis error variances in the context of PSAS. The first method is based on a generalized eigenvalue problem involving the square of the forecast error covariance matrix and the innovation covariance matrix. The second method is that of Riishøjgaard (2000) and involves the eigendecomposition of the innovation covariance matrix. The first procedure has the advantage of providing direct information on the amount of analysis error variance captured for a given number of eigenmodes, while the same is not so for the second method. However,

when computational requirements are taken in to account, Riishøjgaard (2000) procedure is far more attractive.

Furthermore, with the current configuration of PSAS, that is, current forecast and observation error covariances, as well as observational coverage, our experiments show that for a given number of modes Riishøjgaard (2000) procedure captures only slightly less variance than the procedure solving the generalized eigenvalue problem. Combining this with computational requirements it seems more adequate to use Riishøjgaard's procedure to estimate analysis error variances with PSAS.

Some of the conclusions from our work are very much in accordance with the findings in Daley and Barker (2000), particularly in what refers to computational issues. Although an algorithm based on Riishøjaard's approximation seems to be the more promising for practical applications than other alternatives, our current implementation is nowhere near prime time: to get a reasonable estimate of the analysis error variances using all analysis levels and the typical number of observations in a synoptic time the computational cost would be of about one order of magnitude larger than it takes to perform a single PSAS analysis. The use of local approximations of the innovation covariance $\mathbf{\Gamma}$, and correspondingly of the "projection matrix" $\mathbf{P}^f\mathbf{H}^T$, as used in Daley and Barker (2000), is a potential alternative. Work is in progress to merge the development of the so called reduced PSAS (Lyster et al. 2001, personal communication), which localizes these matrices, with the development of the analysis error algorithm. This combination is expected to bring the cost of producing analysis error variance estimates down to an acceptable level.

Acknowledgments. It is a pleasure to thank Arlindo da Silva, Dick Dee, Gregory Gaspari and Lars Peter Riishøjgaard for helpful discussions throughout the course of this work. This research is supported by the NASA EOS Interdisciplinary Project on Data Assimilation.

References

Lehoucq, R. B., D. C. Sorensen, C. Yang, 1997: *ARPACK Users' Guide: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods.* [Available online: http://www.caam.rice.edu/software/ARPACK/].

Barkmeijer, J., M. Van Gijzen, and F. Bouttier, 1998: Singular vectors and estimates of the analysis-error covariance metric. *Quart. J. Roy. Meteor. Soc.*, **124**, pp. 1695-1713.

Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz, D. Lamich, 1998: Assessing the effects of data selection with the DAO Physical-space Statistical Analysis System. *Mon. Wea. Rev.*, **126**, pp. 2913-2926.

Courtier, P., and Coauthors, 1998: The ECMWF implementation of three dimensional variational assimilation (3D-VAR). Part I: Formulation. *Quart. J. Roy. Meteor. Soc.*, **124**, 1783-1808.

Daley, R., and E. Barker, 2000: The NRL 3DVAR Source Book. NRL, Monterey, Ca.

Fisher, M., 1996: The specification of background error variances in the ECMWF variational analysis system. *Proc. ECMWF Seminar on Data Assimilation*, 645-652.

Girard, D. A., 1991: A fast 'Monte-Carlo cross-validation' procedure of large least squares problems with noisy data. *Numer. Math.*, **56**, 1-23.

Klinker, E., F. Rabier, and R. Gelaro, 1998: Estimation of key analysis errors using the adjoint technique. *Quart. J. Roy. Meteor. Soc.*, **124**, 1909-1933.

Parrish, D. F., and J. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747-1763.

Riishøjgaard, L.P., 2000: A method for estimating the analysis error variance in a physical space statistical data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **126**, 1367-1386.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.