P6.12 UNIFIED TREATMENT OF MEASUREMENT BIAS AND CORRELATION IN VARIATIONAL ANALYSIS WITH CONSIDERATION OF THE PRECONDITIONING PROBLEM

R. James Purser

General Sciences Corporation
Beltsville, MD

John C. Derber

Environmental Modeling Center, NCEP
Camp Springs, MD

1. INTRODUCTION

As remotely sensed data become an increasingly dominant source of the information provided to operational forecast models through assimilation procedures, the related problems of correlated or biased measurement error and the poor numerical conditioning of the formal inversion by the assimilation become increasingly severe. This paper will address a unified approach to the treatment of measurement bias and correlation through the use of ancillary variables. The proposed treatment further inflates the already large condition number intrinsic to the analysis inversion problem but, by employing an observation-space form of the analysis and adopting an extension to the averaged block-matrix preconditioners recently advocated by Daley and Barker (2000) based on grouping the data into overlapping small clusters, we expect to be able to achieve a dramatic reduction in the condition number.

2. PRIMAL AND DUAL FORMS OF 3D-VAR

We adopt some notational suggestions of Ide et al. (1993). The primal (e.g., grid or spectral space) variational principle for 3D-VAR seeks the particular model state, $\boldsymbol{x} = \boldsymbol{x}^a$ that minimizes the cost function $\mathcal{L}_1(\boldsymbol{x})$ defined by

$$\mathcal{L}_1(\boldsymbol{x}) = \mathcal{L}_a(\boldsymbol{x}) + \mathcal{L}_y(\boldsymbol{x}), \tag{1}$$

with

$$2\mathcal{L}_a(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{x}^b)^T \boldsymbol{B}_a^{-1}(\boldsymbol{x} - \boldsymbol{x}^b), \tag{2}$$

and

$$2\mathcal{L}_y(\boldsymbol{x}) = (\boldsymbol{y} - \mathcal{H}(\boldsymbol{x}))^T \boldsymbol{R}^{-1}(\boldsymbol{y} - \mathcal{H}(\boldsymbol{x})). \tag{3}$$

$\mathcal{H}(\boldsymbol{x})$ denotes the nonlinear measurement operator of true state $\boldsymbol{x}$, to which we add error terms $\boldsymbol{r}$ to get actual observations $\boldsymbol{y}$. The covariances of errors $\boldsymbol{r}$ form the matrix,

$$\boldsymbol{R} = \langle \boldsymbol{r}\boldsymbol{r}^T \rangle. \tag{4}$$

$\boldsymbol{x}_b$ denotes the background field and $\boldsymbol{B}_a$ is the background error covariance, which we assume can be expressed:

$$\boldsymbol{B}_a = \boldsymbol{C}_a \boldsymbol{C}_a^T, \tag{5}$$

where $\boldsymbol{C}_a$ possibly possesses more columns than rows.

Expressing

$$\boldsymbol{x}^b - \boldsymbol{x} = \boldsymbol{C}_a \boldsymbol{v}, \tag{6}$$

we obtain a basic preconditioning of the primal problem by solving it for $\boldsymbol{v}$. Then, from the minimizer, $\boldsymbol{v}^a$, the proper analysis increment is simply,

$$\boldsymbol{x}^a - \boldsymbol{x}^b = -\boldsymbol{C}_a \boldsymbol{v}^a. \tag{7}$$

The dual (observation space) variational problem seeks the solution to vector which is usually of smaller dimensionality than the analysis increment vector. It can be shown that an implicit representation of the dual problem, valid even in the presence of significant nonlinearity of the operators involved, is the minimization of:

$$\mathcal{L}_2(\boldsymbol{f}) = \boldsymbol{f}^T(\boldsymbol{R} + \boldsymbol{H}_a \boldsymbol{B}_a \boldsymbol{H}_a^T)\boldsymbol{f} - 2\boldsymbol{f}^T \hat{\boldsymbol{d}}, \tag{8}$$

where

$$\boldsymbol{x}^a - \boldsymbol{x}^b = \boldsymbol{B}_a \boldsymbol{H}_a^T \boldsymbol{f}, \tag{9}$$

$$\hat{\boldsymbol{d}} = \boldsymbol{y} - \left(\mathcal{H}(\boldsymbol{x}^a) + \boldsymbol{H}_a(\boldsymbol{x}^b - \boldsymbol{x}^a)\right), \tag{10}$$

and $\boldsymbol{H}_a$ is the linearization of $\mathcal{H}_a$, provided the $\boldsymbol{H}_a$, $\boldsymbol{x}^a$ and $\hat{\boldsymbol{d}}$ are frozen during the minimization. The dependence of $\hat{\boldsymbol{d}}$ on $\boldsymbol{x}$ is weak in the vicinity of the analysis $\boldsymbol{x}_a$, so it can be re-evaluated periodically in an outer iteration of the solution procedure and, approximately,

$$\hat{\boldsymbol{d}} \approx \boldsymbol{d} = \boldsymbol{y} - \mathcal{H}(\boldsymbol{x}^b). \tag{11}$$

Corresponding author address: R. J. Purser, W/NP2 RM 207, WWBG, 5200 Auth Road, Camp Springs, MD 20746

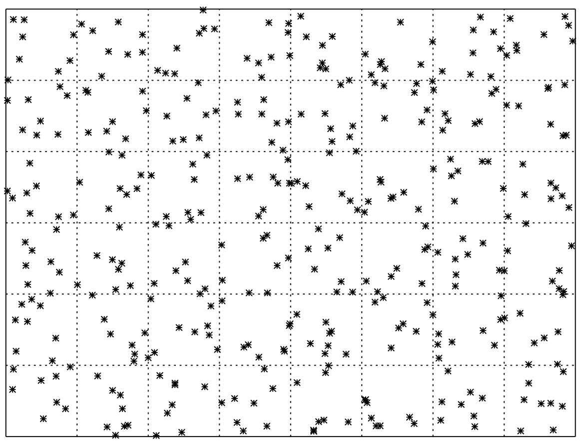

Figure 1. Doubly periodic domain subdivided into boxes and with data locations plotted

The common assumption that $\boldsymbol{R}$ is diagonal, and that it can therefore be inverted trivially, is an especially useful one to make for the dual problem because it allows the $\boldsymbol{R}$ to be used as a basic preconditioning operator, giving the same condition number for the dual problem as obtained for the basic preconditioned primal one (see Courtier 1997 and next section). Unfortunately, the diagonal assumption for $\boldsymbol{R}$ is often wrong for various reasons.

3. DATA WITH CORRELATED ERRORS

Instead of representing correlated data using a non-diagonal $\boldsymbol{R}$, we recommend an approach by which the sources of correlation in the errors of the data are, to the extent possible, identified and incorporated as additional analysis variables. This leaves a diagonal $\boldsymbol{R}$ that can be inverted and used for basic preconditioning. The added correlated terms are modelled as the product of white noise operated upon by a filter which shapes the final correlation structure.

$$\boldsymbol{y} = \mathcal{H}(\boldsymbol{x}) + \boldsymbol{H}_\tau \boldsymbol{C}_\tau \boldsymbol{\tau} + \boldsymbol{r}. \tag{12}$$

Here, $\boldsymbol{\tau}$ represents a white-noise random vector ($\langle \boldsymbol{\tau}\boldsymbol{\tau}^T \rangle = \boldsymbol{I}$), $\boldsymbol{C}_\tau$ the filter simulating the correlation of error in time and/or space, and $\boldsymbol{H}_\tau$ samples the resulting correlated field of error at times and locations that correspond to the actual observations used. Components of residual error vector $\boldsymbol{r}$ now become truly independent. Near the true solution, the effective innovation now takes the form

$$\hat{\boldsymbol{d}} = \mathcal{H}(\boldsymbol{x}) - \mathcal{H}(\boldsymbol{x}^a) + \boldsymbol{H}_a \boldsymbol{C}_a \boldsymbol{v}^a + \boldsymbol{H}_\tau \boldsymbol{C}_\tau \boldsymbol{\tau} + \boldsymbol{r} \tag{13}$$

and we assume cancellation of the first two terms on the right. Effectively, preconditioned analysis increment, $\boldsymbol{v}$, is augmented with components $\boldsymbol{\tau}$ and, with substituted combinations,

$$\boldsymbol{q}^T \equiv \left[\boldsymbol{v}^T ; \boldsymbol{\tau}^T\right], \tag{14}$$

$$\boldsymbol{H} = [\boldsymbol{H}_a ; \boldsymbol{H}_\tau], \tag{15}$$

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_a & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_\tau \end{pmatrix}, \tag{16}$$

we regain standard variational forms, with correctly diagonal $\boldsymbol{R}$. $\boldsymbol{q}$ is the new augmented preconditioned analysis increment.

If we define the standard deviations matrix,

$$\boldsymbol{S} = \boldsymbol{R}^{1/2}, \tag{17}$$

and use it to rescale the innovations:

$$\boldsymbol{\eta} = \boldsymbol{S}^{-1} \hat{\boldsymbol{d}}, \tag{18}$$

then the equations that connect the basic preconditioned and augmented analysis increments $\boldsymbol{q}$ with $\boldsymbol{\eta}$ are:

$$\boldsymbol{q} = (\boldsymbol{I}_{(q)} + \boldsymbol{M}^T \boldsymbol{M})^{-1} \boldsymbol{M}^T \boldsymbol{\eta} \tag{19}$$

in the primal case, and:

$$\boldsymbol{q} = \boldsymbol{M}^T (\boldsymbol{M}\boldsymbol{M}^T + \boldsymbol{I}_{(y)})^{-1} \boldsymbol{\eta} \tag{20}$$

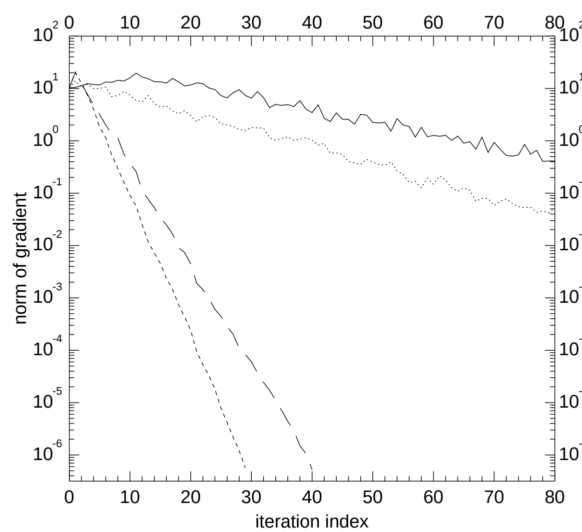

Figure 2. Convergence traces for an idealized iterative analysis using the conjugate gradient method under various preconditioning strategies: without special preconditioning (solid); with a single block matrix preconditioner based on clustering of data into groups each of four of the boxes of Fig. 1 arranged 2×2 (dotted); with the averaged effect of two block matrices formed by clusters as before but such that a maximum overlap between partitionings is obtained (long-dashed); with four optimally-overlapping patitionings contributing their block matrix preconditioners to a four-way average (short-dashed)

in the dual problem, where,

$$\boldsymbol{M} = \boldsymbol{S}^{-1}\boldsymbol{H}\boldsymbol{C}, \quad (21)$$

and $\boldsymbol{I}_{(q)}$ and $\boldsymbol{I}_{(\eta)}$ are identity operators in the respective spaces of $\boldsymbol{q}$ and $\boldsymbol{\eta}$. The equivalence of condition numbers of the two forms is now plain to see (Courtier 1997). However, the condition numbers typically obtained ($> 10^3$) are still much larger than we would wish.

Our proposed formalism can be thought of as a generalization, with possibly nontrivial correlations, of the method of treating observational bias implemented for satellite data used by Derber and Wu (1999). By allowing the covariance $\boldsymbol{R}$ of the effective random observation error $\boldsymbol{r}$ to take a diagonal form even when true data errors are correlated, a significant simplification in the analysis procedure is obtained. However, a hidden cost of this procedure is the invariably larger condition number of the resulting problem, a consequence of the effective errors $\boldsymbol{r}$ now being much smaller in magnitude than they would have been in the conventional 3D-VAR implementation. We are therefore obliged to pay greater attention to the preconditioning aspects of the analysis when the proposed modifications to the 3D-VAR formulation are adopted.

4. PRECONDITIONING BY CLUSTERING

The relatively trivial preconditioning, described in the previous section, that results from the non-dimensionalization of the problem using the operators at hand, is inadequate in practice, but an improvement would require the utilization of knowledge about the particular structural regularities typical of forecast background errors. One obvious regularity, to which the purely algebraic manipulations of the previous section are necessarily blind, is the tendency for correlations in the error to decay away to insignificance at a sufficient geographical separation. This feature was routinely exploited in traditional Optimum Interpolation methods of analysis (Gandin 1963) through the selection of only nearby data in the analysis of each gridded value. This structural regularity can be exploited also for preconditioning the dual form of 3D-VAR, as shown by Cohn et al. (1998). This involves grouping the observational data into geographical clusters small enough to permit the matrices of the restricted analysis inversion problems associated with each cluster to be inverted directly. Daley and Barker (2000) extend this technique by averaging the preconditioning operations associated with two completely different data groupings.

The value of these stategies, and a systematic extension to larger numbers of overlapping groupings, can be nicely illustrated in a simple idealized context of a scalar field with simple Gaussian forms

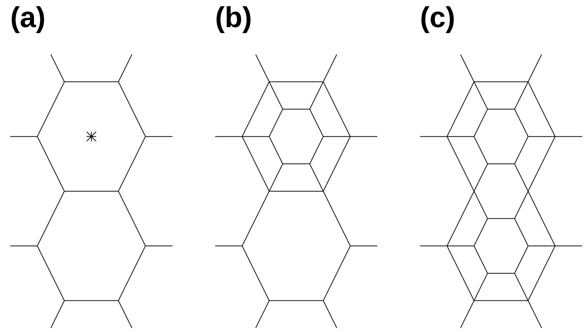

Figure 3. (a) Schematic example of a portion of a horizontal tesselation using even-sides polygons. (b) Result of the refinement rules applied to just one of the hexagons. (c) Result of the refinement rules when applied to both adjacent hexagons.

for the homogeneous covariances of background error on a two-dimensional doubly-periodic domain. Fig. 1 shows the domain divided into boxes in each of which is placed (randomly) an equal number of observations. The background standard deviation is taken to be ten times that of the data and the characteristic size of the Gaussian covariance profile is 1.5 times the box dimension. Fig. 2 shows sample convergence plots with the conjugate gradients technique under four different preconditioning strategies. The slowest convergence (solid line) is obtained using the trivial non-dimensional scaling preconditioning described in the previous section. Constructing the data clusters from 2×2 arrangements of boxes, the dotted line shows the small improvement in convergence when a single grouping of this kind is used to prescribe the observation-space preconditioner. As shown by Daley and Barker, a substantial improvement is obtained by averaging the preconditioning from two overlapping clus-

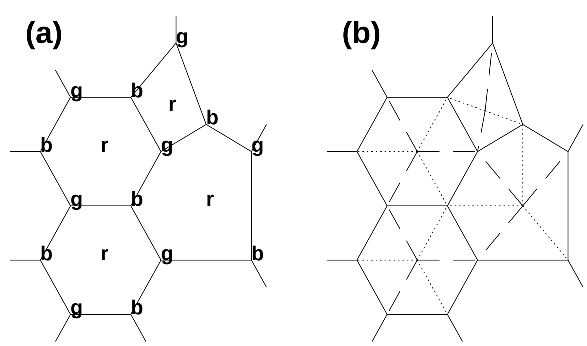

Figure 4. (a) Coloring rule applied to a tesselation by even-sided polygons. Each polygon has a red nucleus ('r') and alternating blue ('b') and green ('g') vertices. (b) The corresponding triagulation with lines distinguished to show which of the three possible partitionings by even-sided polygons they belong to.

ters; here we have optimally overlapping of clusters formed from 2×2 arrangements of boxes mutually staggered in both directions by one box-size. The convergence is shown for this case by the long-dashed line. Finally, we take the average of the four possible clusterings based on 2×2 squares tiling the domain, and show the resulting convergence by the short-dashed line, which is the best result. The condition numbers displayed in Fig. 2 range from about 1000 without clustering to about 12 with the four-cluster averaged preconditioner.

In general, we never have such a uniformly dense distribution of the data. It is possible to define a systematic strategy by which general data can be grouped into a small number of overlapping tilings of the domain, so that each tile contains roughly the same amount of data and no datum lies near the tile boundary in all of the groupings using only three overlapping groupings. This scheme should provide the analogue, for irregular data, of the best of the methods illustrated in Fig. 2. The construction can occur as the end state of a progression of refinements, at each stage of which a primary decomposition of the domain involves even-sided polygons. Fig. 3(a) shows such an arrangement of hexagons. Initially, the domain is coarsely decomposed. If warranted by a sufficient quantity of the data it contains, a polygon may be subdivided according to rules that maintain the even-sided property. From a designated central point of the original polygon (the star in Fig. 3(a)) take the midpoints of the segments that link it with the polygon's vertices and create a similar but half-size polygon together with peripheral quadrilaterals, as in Fig. 3(b). Should two or more of the neighboring original polygons be subdivided in this way, merge the pairs of quadrilaterals from the adjacent parent polygons, as shown in Fig. 3(c), to form a single new hexagon instead.

These rules of refinement can be continued for as many generations as required to equitably partition the data. Finally, we may 'color' the central points of each polygonal tile 'red' and their vertices alternately 'blue' and 'green', as shown in Fig 4(a). But now the 'blue' points may be regarded as the central points of a secondary tiling of the domain, whose vertices are alternately 'red' and 'green'. Likewise, the 'green' points define center for a third set of tiles. The boundaries of the three co-existing groupings are shown in Fig 4(b) for this example. This three-way overlapping tiling construction offers a way of generalizing the precondition strategy proposed by Daley and Barker (2000) that ensures each datum lies well inside at least one of the polygonal tiles.

5. DISCUSSION AND CONCLUSION

Having set out a unified formalism for the treatment in 3D-VAR of biased or correlated measurement errors, we have also pointed out that a consequence of excising the correlated error components and effectively assigning them to an augmented set of 'analysis' variables is that the condition number of the analysis problem increases substantially. In order to achieve a satisfactory preconditioning, it becomes necessary to exploit the structural regularity inherent in a background error field, for example, by grouping the data into manageable clusters from which preconditioning operators can be calculated by direct inversion. To this end, we propose a generalization of the method of Daley and Barker (2000), which is demonstrably effective in reducing the condition number. At present this preconditioning strategy seems restricted to observation space forms of the analysis. It remains to be seen whether analogous methods might apply in model state space.

6. ACKNOWLEDGMENTS

This work was partially supported by the NSF/NOAA Joint Grants Program of the US Weather Research Program. This research is also in response to requirements and funding by the Federal Aviation Administration (FAA). The views expressed are those of the authors and do not necessarily represent the official policy or position of the FAA.

REFERENCES

Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Mon. Wea. Rev.*, **126,** 2737–3052.

Courtier, P., 1997: Dual formulations of four-dimensional variational assimilation. *Quart. J. Roy. Meteor. Soc.*, **123,** 2449–2461.

Daley, R., and E. Barker, 2000: *NAVDAS Source Book 2000.* NRL Publication NRL/PU/7530–00-418. 153 pp.

Derber, J. C., and W.-S. Wu, 1998: The use of TOVS cloud-cleared radiances in the NCEP's SSI analysis system. *Mon. Wea. Rev.*, **126,** 2287–2299.

Gandin, L. S., 1963: *Objective Analysis of Meteorological Fields.* Gidrometeorologicheskoe Izdatel'stvo, 242 pp. Translated by Isreal Program for Scientific Translations.

Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: operational sequential and variational. *J. Meteor. Soc. Japan*, **75,** 181–189.