

Multigrid solvers and multigrid preconditioners for the solution of variational data assimilation problems

Laurent Debreu, Emilie Neveu, Ehouarn Simon, François-Xavier Le Dimet, Arthur Vidard

▶ **To cite this version:**

Laurent Debreu, Emilie Neveu, Ehouarn Simon, François-Xavier Le Dimet, Arthur Vidard. Multigrid solvers and multigrid preconditioners for the solution of variational data assimilation problems. Quarterly Journal of the Royal Meteorological Society, 2016, 142 (694), pp.515-528. 10.1002/qj.2676 . hal-00874643

HAL Id: hal-00874643

https://inria.hal.science/hal-00874643

Submitted on 18 Oct 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multigrid solvers and multigrid preconditioners for the solution of variational data assimilation problems

Laurent Debreu [a]*Emilie Neveu[b] Ehouarn Simon[c] François-Xavier Le Dimet[a] Arthur Vidard[a]

[a]*INRIA Grenoble Rhône-Alpes and Laboratoire Jean Kuntzmann, Grenoble, France*

[b]*University of Savoy*

[c]*Nansen Environmental and Remote Sensing Center, Bergen, Norway*

*Correspondence to: Laboratoire Jean Kuntzmann, 51 rue des Mathematiques, 38400 Saint Martin d'Heres, France.

E-mail: Laurent.Debreu@inria.fr

In order to lower the computational cost of the variational data assimilation process, we investigate the use of multigrid methods to solve the associated optimal control system. On a linear advection equation, we study the impact of the regularization term of the optimal control and the impact of discretization errors on the efficiency of the coarse grid correction step. We show that even if the optimal control problem leads to the solution of an elliptic system, numerical errors introduced by the discretization can alter the success of the multigrid methods. The view of the multigrid iteration as a preconditioner for a Krylov optimization method leads to a more robust algorithm. A scale dependent weighting of the multigrid preconditioner and the usual background error covariance matrix based preconditioner is proposed and brings significant improvements. **Copyright © 0000 Royal Meteorological Society**

Key Words: Variational Data assimilation; Multigrid methods; Preconditioning; Transport equation

Received . . .

Citation: . . .

1. Introduction

Data assimilation methods are a way of combining different sources of information: a priori information (background), observations and numerical models according to error statistics on these sources. Data assimilation methods can be divided into two groups. First, sequential methods are based on the Kalman filtering or ensemble approach (Evensen 2006) and make the state vector evolve in time along with its error statistics. Then, variational methods (Le Dimet and Talagrand 1986) are based on optimal control techniques and minimize a cost function $J(\mathbf{x})$ that measures the distance between the model trajectory and observations. Both methods have huge computational costs and have to be simplified for operational purposes. The two approaches are well known to be equivalent in the linear case and in absence of model error. In this paper we focus on variational data assimilation methods in the context of geophysical fluids. Assuming $\mathbf{x}$ is the control vector, the necessary condition of optimality at $\mathbf{x}^\star$ is given by the Euler equation $\nabla_{\mathbf{x}} J(\mathbf{x}^\star) = 0$. This leads to the solution of a large unconstrained minimization problem. This paper is an attempt to use multigrid methods for solving the resulting system.

In the optimal control framework, several attempts have been made to apply multigrid methods, either for linear or non linear optimization (see Borzì and Schulz (2009) for a review). Lewis and Nash (2005) focus on the control of the initial condition for a linear advection equation with a specific cost function and discretization scheme that renders the problem fully elliptic (i.e. large scales components of the error are more efficiently reduced on a grid at coarser resolution) and thus well suited for multigrid methods.

In this paper, multigrid methods are applied to a simple 2D (1D in space, 1D in time) variational assimilation problem using a cost function $J(\mathbf{x})$ that mimics the ones used in more realistic applications in the context of geophysical fluids. The model is linear and leads to a quadratic cost function. Most of the actual implementations of variational data assimilation use a Gauss Newton (a.k.a. incremental) approach where a succession of minimizations of quadratic cost functions are performed.

In section 2, the variational data assimilation problem is described and its characteristics (in particular its ellipticity) are derived. Section 3 introduces the multigrid algorithm and the convergence criteria. Its application to the variational data assimilation is studied in section 4 where the main ingredients of the multigrid algorithms are discussed: smoothing property, design of transfer operators and approximation property. The effect of the background error correlation matrix on the ellipticity is studied as well as the impact of numerical model errors on the approximation property. Numerical experiments are performed in section 5 for 1D and 2D variational problems and illustrate the behavior of the multigrid algorithm for different parameters of the cost function and of the numerical model. The direct application of the multigrid method as a solver for the data assimilation problems may lead to non convergence in particular cases when the approximation property is strongly violated. For this reason, we also investigate the view of the multigrid method as a preconditioner for a conjugate gradient algorithm. This adds more robustness and leads to performance results that significantly outperformed traditional monogrid methods preconditioning based on the background error covariance matrix (or its square root). We show how to design a preconditioner that tries to take advantages of both preconditioners.

2. Variational data assimilation and associated linear system

We consider the time evolution of a system governed by the following equation:

$$\begin{aligned} \frac{\mathrm{d}X}{\mathrm{d}t} &= F(X) \\ X(t = t_0) &= \mathbf{x} \end{aligned} \tag{1}$$

$\mathbf{x}$ is the initial condition at time $t = t_0$ and will be our control parameter. The variational data assimilation

problem consists in finding the minimum of a cost function $J(\mathbf{x})$ that measures the distance from the numerical model to the observations and includes a background or regularization term associated to a first guess $\mathbf{x}_b$.

$$J(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}-\mathbf{x}_b\right)^T B^{-1}\left(\mathbf{x}-\mathbf{x}_b\right) + \frac{1}{2}\left(H\left(X(\mathbf{x},t)\right)-y\right) R^{-1}\left(H\left(X(\mathbf{x},t)\right)-y\right) \tag{2}$$

Here y are the observations. H is the observation operator from the model to the observations space, R and B are respectively the observations and background error covariances matrices. In the following R will be assumed to be a diagonal matrix with constant standard deviation equal to σ_{obs}. When observations are available at a number N_{obs} of different times t_i, the second term of the right hand side of (3) can be more precisely written as:

$$J(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}-\mathbf{x}_b\|^2_{B^{-1}} + \frac{1}{2}\sum_{i=0}^{N_{\text{obs}}-1}\left(H_i\left(X(\mathbf{x},t_i)\right)-y_i\right) R_i^{-1}\left(H_i\left(X(\mathbf{x},t_i)\right)-y_i\right) \tag{3}$$

At a minimum $\mathbf{x}^\star$ of J, the gradient is zero

$$\nabla_{\mathbf{x}} J(\mathbf{x}^\star) = 0 \tag{4}$$

When the model F and the observations operator H are linear, the cost function is quadratic and the solution of 4 is equivalent to the solution of

$$\mathbf{A}\mathbf{x}^\star = b \tag{5}$$

where $\mathbf{A}$ is the Hessian of the cost function:

$$\mathbf{A} = B^{-1} + \mathbf{H}^T R^{-1}\mathbf{H}$$

where $\mathbf{H}$ includes both the model and the observation operators and the right hand side b is given by

$$b = B^{-1}\mathbf{x}_b + \mathbf{H}^T R^{-1} y$$

The solution of (5) can thus be written:

$$\mathbf{x} = \left(B^{-1} + \mathbf{H}^T R^{-1}\mathbf{H}\right)^{-1}\left(B^{-1}\mathbf{x}_b + \mathbf{H}^T R^{-1} y\right)$$

or

$$\mathbf{x} = \mathbf{x}_b + \left(B^{-1} + \mathbf{H}^T R^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T R^{-1}(y - H\mathbf{x}_b)$$

Using the Shermann-Morrison-Woddbur formula:

$$\left[A^{-1} + B^T C^{-1} B\right]^{-1} B^T C^{-1} = AB^T\left[C + BAB^T\right]^{-1}$$

we get

$$\mathbf{x} = \mathbf{x}_b + B\mathbf{H}^T\left(R + \mathbf{H}B\mathbf{H}^T\right)^{-1}(y - H\mathbf{x}_b) \tag{6}$$

Eq. (6) shows that a correction is computed in the observation space weighted by the observations and background error covariance matrices, this correction is brought back to the model state and then spatially distributed according to the B matrix. In practice, the minimization is conducted using $\mathbf{x}_b$ as a starting point. For our problem, the scale of the final correction $(\mathbf{x} - \mathbf{x}_b)$ is thus prescribed by the correlation length as specified by the B matrix. The main idea is to solve the system (5) by multigrid methods and to evaluate the main characteristics of the problem that impacts their convergence and robustness. Among these characteristics, the ones of the background error covariance matrix are of prime importance.

2.1. Definition of the background error covariance matrix B, preconditioning and implicit solve

The background error covariance matrix B is a positive definite matrix that can be factored as $B = \Sigma C \Sigma$ where Σ is a diagonal matrix of background error standard deviations σ_b and C a symmetric matrix of background error correlations. The correlations are usually modeled by applying a diffusion operator (Weaver and Courtier

2001). Indeed time integration over the interval $[0, T]$ of the diffusion equation

$$\frac{\partial \eta}{\partial t} = \nu \Delta \eta \tag{7}$$

approximates a Gaussian correlation function of length L_{corr} if $L_{\text{corr}}^2 = 2\nu T$. If (7) is integrated with an explicit Euler scheme, the discrete form writes

$$\eta(T) = \mathcal{L}_+^{M_{\text{laplac}}} \eta(0), \qquad \mathcal{L}_+ = [I + \nu \Delta t \, \Delta] \tag{8}$$

where $M_{\text{laplac}} \Delta t = T$ and $\nu \Delta t = \frac{L_{\text{corr}}^2}{2 M_{\text{laplac}}}$. When associated with a second order centered approximation of the laplacian, the operator L_+ is constrained by the stability condition $\nu \Delta t / \Delta x^2 \leq 1/2$ which leads to

$$M_{\text{laplac}} \geq 2 \left(\frac{L_{\text{corr}}}{\Delta x} \right)^2 \tag{9}$$

A diagonal normalization matrix Λ is then computed so that C written under the form

$$C = \Lambda \mathcal{L}_+^{M_{\text{laplac}}} \Lambda$$

has ones along its diagonal. The coefficients of Λ can be computed using the relation

$$e_i^T C e_i = 1 \qquad \Leftrightarrow (\mathcal{L}_+^{M_{\text{laplac}}/2} \Lambda e_i)^T (\mathcal{L}_+^{M_{\text{laplac}}/2} \Lambda e_i) = 1, \quad \forall i$$

So that

$$\Lambda_{ii} = \sqrt{(\mathcal{L}_+^{M_{\text{laplac}}/2} e_i)^T (\mathcal{L}_+^{M_{\text{laplac}}/2} e_i)}$$

In the following we will assume that the coefficients of the diagonal matrix Σ are equal to σ_b.

2.1.1. Preconditioning

The performance of an iterative minimization method for the solution of (5) is linked to the condition number of this system: $\mathcal{K}(\mathbf{A}) = \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})}$ where $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ are the smallest and largest eigenvalues of $\mathbf{A}$. The smaller is the condition number the smaller will be the rate of convergence. When the matrix B and its square root are available as in the preceding derivation (where $B^{1/2} = \Sigma \Lambda \mathcal{L}_+^{M_{\text{laplac}}/2}$), an alternative, symmetrically preconditioned, form of (5) can be deduced (see Courtier (1997)):

$$\mathbf{A}^P \mathbf{y}^\star = B^{T/2} \mathbf{A} B^{1/2} \mathbf{y}^\star = B^{1/2} b \tag{10}$$

where $\mathbf{x}^\star = B^{1/2} \mathbf{y}^\star$.

We get

$$\mathbf{A}^P = B^{T/2} \mathbf{A} B^{1/2} = I + (H B^{1/2})^T R^{-1} (H B^{1/2})$$

$\mathbf{A}^P$ has a smallest eigenvalue superior to 1 and its condition number $\mathcal{K}(\mathbf{A}^P)$ is bounded.

2.1.2. Implicit solve

At very high resolution, the stability constraint (9) of the diffusion operator can lead to a large increase of the corresponding cost due to the required number of time steps. In order to alleviate this cost, the diffusion equation can be integrated in time using an implicit scheme (Carrier and Ngodock (2010); Mirouze and Weaver (2010)). In that case (8) is rewritten as

$$\eta(T) = \mathcal{L}_-^{-M_{\text{laplac}}} \eta(0), \qquad \mathcal{L}_- = [I - \nu \Delta t \, \Delta] \tag{11}$$

This has now been adopted in practical application. Note that each step now requires the inversion of the Laplacian operator that is also potentially computationally extensive. However Gratton *et al.* (2013) have shown that it can be efficiently solved by a multigrid scheme. In the following, we will use this formulation since its leads to a very fast and explicit computation of B^{-1} used in the non preconditioned system:

$$B^{-1} = \Sigma^{-1} \Lambda^{-1} \mathcal{L}_-^{M_{\text{laplac}}} \Lambda^{-1} \Sigma^{-1}$$

It will allow us to easily compare the non preconditionned version to the preconditioned version.

2.2. *Design of the numerical experiments*

In order to study the eigenpairs of our data assimilation problem, we introduce some of the parameters of our idealized test case. The domain is a one dimensional periodic domain of size L. The uniform grid is composed of N cells of stepsize Δx.

The distance between two observations points is denoted by $L_{\rm obs}$ so that the observations are eventually located every $L_{\rm obs}/\Delta x$ grid points. The observation operator H is a simple projection operator at corresponding grid points.

The background error correlation matrix is based on the implicit form of the diffusion operator and the number $M_{\rm laplac}$ of iterations is equal to 4 which is here sufficient to accurately approximate the explicit form of the correlation matrix. The numerical model approximates the solution of a linear advection equation:

$$\frac{\partial u}{\partial t} + c\,\frac{\partial u}{\partial x} = 0, \text{ with } c > 0, x \in [0, L], t \in [0, T] \quad (12)$$

using the initial condition $u(x, t = 0) = u^0(x)$ which represents our control parameter. The numerical values of the following parameters are fixed:

$$c = 1\mathrm{m.s}^{-1}, L = 100\mathrm{m}, T = 78.125\mathrm{s}$$

The discretization of (12) is achieved using a Lax Wendroff scheme (Lax and Wendroff 1960) which is second order accurate both in time and space. An important parameter of the numerical model is the Courant number $\mu = c\,\dfrac{\Delta t}{\Delta x}$ where Δt is the time step. This parameter controls the accuracy of the discretization. The discretization is stable if $\mu \leq 1$ and for $\mu < 1$ implicit diffusion is introduced by the numerical scheme. The special value of $\mu = 1$ leads to a numerical solution that coincides to the exact solution.

2.3. *Eigenstructure of the data assimilation problem*

The convergence of the optimization problem or equivalently the solution of the linear system $\mathbf{A}x = b$ is known to be dependent on two main points: the conditioning of the matrix and the spectrum of its eigenvalues. We recall that the matrix $\mathbf{A}$ is given by:

$$\mathbf{A} = \mathbf{A}^{NP} = B^{-1} + \mathbf{H}^T R^{-1} \mathbf{H}$$

in non preconditioned mode and by

$$\mathbf{A} = \mathbf{A}^{P} = I + (\mathbf{H}B^{1/2})^T R^{-1} (\mathbf{H}B^{1/2})$$

in preconditioned mode.

In the case of a fully observed system $H = I_d$, and with a linear advection equation, a detailed study on the eigenvalues of theses operators have been presented in Neveu *et al.* (2011). Several results on the condition number of these matrices have also been given in Haben *et al.* (2011). We summarize here the main characteristics:

1. If we focus only on the background term ($\mathbf{A}^{NP} = B^{-1}$) on our periodic domain, simple computations lead to the following condition number:

   $$\mathcal{K}(B^{-1}) = \left[1 + \frac{2}{M_{\rm laplac}} \left(\frac{L_{\rm corr}}{\Delta x}\right)^2\right]^{M_{\rm laplac}}$$

   Its condition number can be relatively high and is increased along with the resolution ($\Delta x \searrow$) and with the correlation length $L_{\rm corr}$.
2. The preconditioning ensures that the smallest eigenvalues is bounded by below by 1. Obviously the preconditioning is effective when the relative weight of the background term is not negligible in comparison with the weight on the observation error. This weight is dependent of the ratio of the standard error deviations $\sigma_b/\sigma_{\rm obs}$ and on the ratio $L_{\rm obs}/\Delta x$. For time dependent problems, the weight of the observations term is also increased accordingly

to the size of the assimilation windows. At some point, the preconditioned version behaves worse than the non preconditioned version. Indeed for a fixed background error correlation matrix, the condition number of the non preconditioned system decreases when the condition number of the preconditioned system increases. This can be easily proved in the case when the model is fully observed ($H = I_d$) since in that case it can be shown that:

$$\mathcal{K}(A^{NP})\mathcal{K}(A^{P}) = \mathcal{K}(B)$$

3. If the system is not fully observed ($H^T H \neq Id$), but when there is one observation point every $L_{\rm obs}/\Delta x$ grid points, the eigenvalues may become clustered if the background error correlation length is relatively small. Indeed in this case, the eigenvalue problem can be approximatively in $p = L/L_{\rm obs}$ similar eigenvalues problems between two observation points.

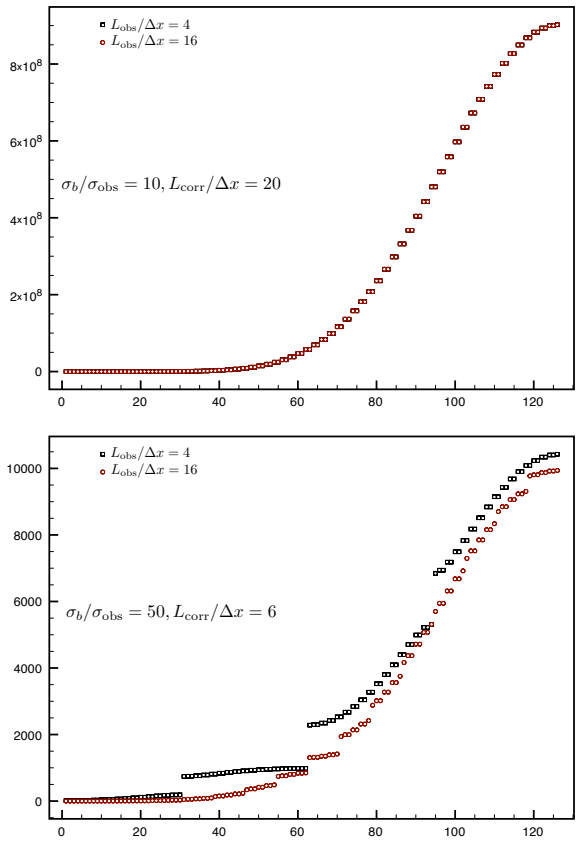


Figure 1. Eigenvalues spectrum of the non preconditioned version for the 1DVAR case ($N_{\rm obs} = 1$)

The characteristics of the systems can thus be studied as a function of the following parameters: $\sigma_b/\sigma_{\rm obs}, L_{\rm corr}/\Delta x, L_{\rm obs}/\Delta x, N_{\rm obs}$. Figures (1,2) represent the eigenvalue spectrum of the unpreconditioned and the preconditioned version of the algorithm for the 1DVAR case ($N_{\rm obs} = 1$, the observations being at initial time t_0). Table (1) shows the corresponding condition numbers.

$\sigma_b/\sigma_{\rm obs}$	$L_{\rm corr}/\Delta x$	$L_{\rm obs}/\Delta x$	NP	P
10	20	4	1441483	1131
10	20	16	14427643	284
50	6	4	766	8439
50	6	16	66608	2660

Table 1. 1DVAR Condition numbers for the non preconditioned (NP) version and for the preconditioned (P) version

In the rest of the paper, $L_{\rm obs}/\Delta x$ is fixed and equal to 16.

We now suppose that observations are located at several times t_i with $t_{i+1} = t_i + T_{\rm obs}$ where $T_{\rm obs}$ is fixed to $7.8125s$. For a 2DVAR algorithm, extending the length of the assimilation windows increases the weight on the observation term and thus render the preconditioned version less efficient. Figure (3) represents the eigenvalues for different number of observation in times ($N_{\rm obs} = 1$ (i.e. 1DVAR), $N_{\rm obs} = 5$ and $N_{\rm obs} = 10$).

The corresponding condition numbers are indicated in table (2).

$N_{\rm obs}$	Non preconditioned	Preconditioned
1	14427643	284
5	1029078	1697
10	530060	3110

Table 2. 2DVAR Condition numbers, $\sigma_b/\sigma_{\rm obs} = 10, L_{\rm corr}/\Delta x = 20$

For the non preconditioned case, the extension of the length of the assimilation window is associated with an increase of the smallest eigenvalues due to the larger weight of the observation term, the largest eigenvalues still coincide with the ones of the B^{-1} matrix. Thus globally the condition number decreases.

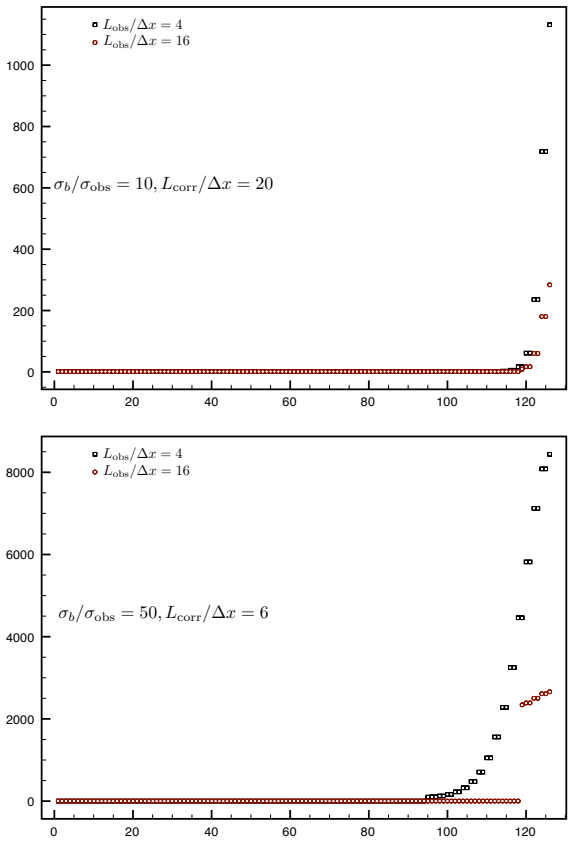


Figure 2. Eigenvalues spectrum of the preconditioned version for the 1DVAR case ($N_{\rm obs} = 1$)

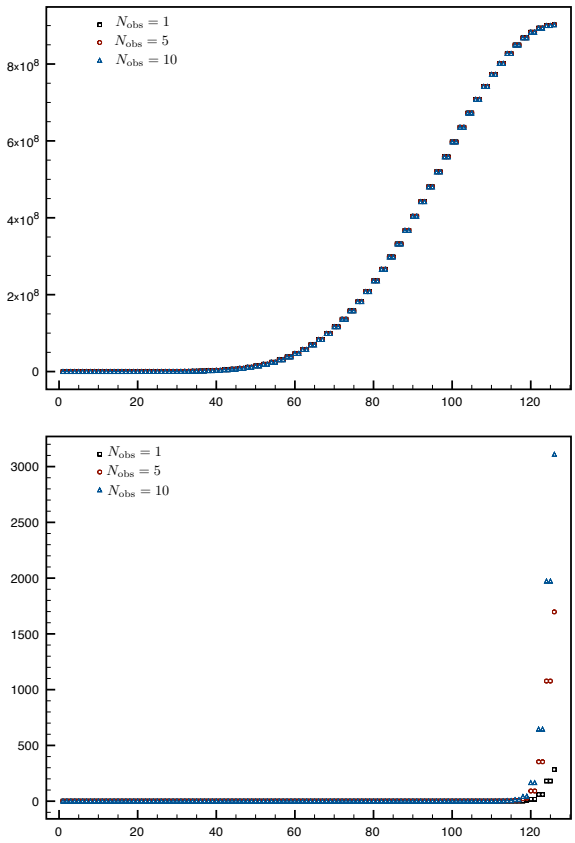


Figure 3. Eigenvalues spectrum for the 2DVAR case for $\sigma_b/\sigma_{\rm obs} = 10, L_{\rm corr}/\Delta x = 20$. Non preconditioned (top), preconditioned (bottom)

For the preconditioned case, the smallest eigenvalue is still one while the largest eigenvalues increase with the weight of the observation term. Globally the condition number increases.

2.3.1. Ellipticity

In a typical optimization method, the error relative to large eigenvalues will be faster removed than the one relative to the small eigenvalues. This is because a matrix-vector product based iterative method will use as a basic ingredient the residual ($\mathbf{A}\mathbf{x} - b = \mathbf{A}(\mathbf{x} - \mathbf{x}^\star)$) which will be small for small eigenvalues even if the error ($\mathbf{x} - \mathbf{x}^\star$) itself has a large amplitude. So it is important to identify the scales relatives to the eigenvectors. These scales are here defined using a filtering based on interpolation and restriction operators. If v is an eigenvector, then the measured quantity is

$$S(v) = \frac{\|(I - I_c^f I_f^c)v\|_2}{\|v\|_2}$$

where I_f^c is a restriction operator from the current grid Ω_f to a 2 times coarser resolution grid Ω_c and I_c^f is a an interpolation operator from Ω_c to Ω_f. By construction $S(v)$ is small at large scales and close to one at small scales. On figure (4), $S(v_i)$ is plotted for the N eigenvectors ($N = 128$) of the non preconditioned and the preconditioned version. The eigenvectors v_i have been ordered according to increasing eigenvalue.

The characteristics of the non preconditioned version with small eigenvalues (seen on figure (3)) corresponding to large scales (seen on figure (4)) (and vice versa) is characteristic of an *elliptic* system. In our data assimilation problem, this characteristic mainly comes from the background (or regularization) term which enables the control of the small scales of the initial condition, a control that could not be performed with the observation term only, in particular in the case of a partially observed system. This ellipticity is at the base of the use of multigrid methods that

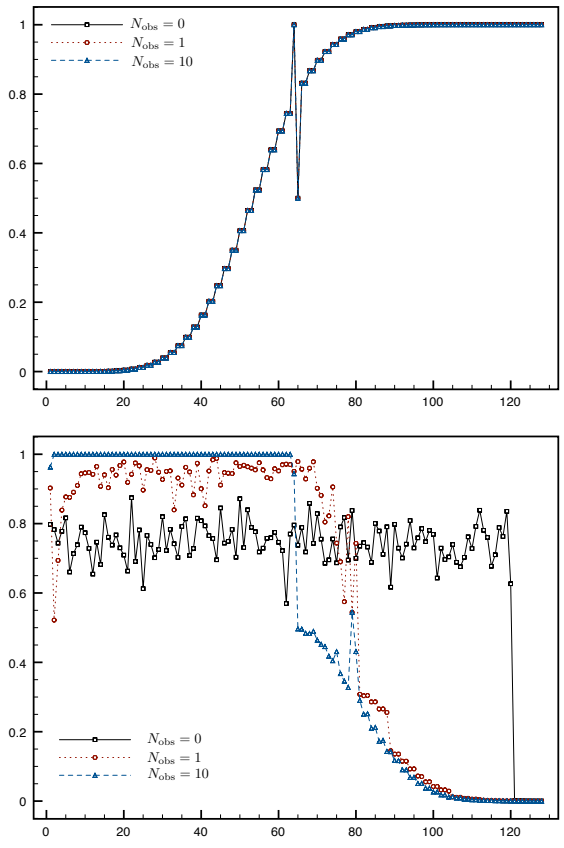


Figure 4. Scales of eigenvectors for $\sigma_b/\sigma_{\rm obs} = 10$, $L_{\rm corr}/\Delta x = 20$ and for different number of observation times. $N_{\rm obs} = 0$ corresponds to the background term only, $N_{\rm obs} = 1$ to observations only at initial time (1DVAR). Non preconditioned (top), preconditioned (bottom)

are introduced in the new section for general linear systems. On figure (4), it can be seen that the preconditioning removes the ellipticity of the original matrix. The large scales corresponds to large eigenvalues so that the large scales component of the error will be first reduced. Note that when the length of the assimilation windows is extended, it has almost no effect on the eigenvectors of the non preconditioned version while it has a strong impact on those of the preconditioned version. In particular, it can be seen (Figure (4, right) that the number of large scale modes in increased. For $N_{\rm obs} = 10$ (blue dashed curve), eigenvectors from approximatively 65 to 100 now corresponds to medium to large scales. Because they are still associated with small eigenvalues (cf figure (3)), the reduction of error components along these eigenvectors will be slow and can benefit from the use of a coarser resolution grid.

The main idea of the multigrid method is that if they are some large scale components that are slow to converge on the high resolution grid, they will reduced faster and at a smaller cost on a coarser resolution grid.

3. Multigrid methods

3.1. Multigrid methods: algorithm

Readers can refer to Briggs *et al.* (2000) for an excellent introduction to the subject. The general idea is to begin by reducing the small scale components of the error on the current (high resolution) grid Ω_f. This is called the pre-smoothing step and should be done in a few iterations according to the ellipticity of the system (large eigenvalues at small scales). The error is then smooth and can be appropriately computed on a coarse resolution grid Ω_c during the coarse grid correction step. The correction is then interpolated back to the fine grid. Since the interpolation operator can in turn produce small scale error components, a post-smoothing step is finally applied.

The basic algorithm with two grid levels writes: MULTIGRID(ν_1,ν_2,Ω_f,$\mathbf{x}_f$,b_f)

1. Pre-smoothing: Apply ν_1 steps of an iterative method S_1 on a fine grid

$$\mathbf{A}_f\mathbf{x}_f = b_f, \quad \mathbf{x}_f = S_1^{\nu_1}(\mathbf{x}_f, b_f)$$

2. Coarse grid correction

   - Transfer the residual on a coarser grid

$$r_f = b_f - \mathbf{A}_f\mathbf{x}_f, \quad r_c = I_f^c r_f, \quad I_f^c : \text{ restriction operator}$$

   - Solve the problem on the coarse grid

$$\mathbf{A}_c\delta\mathbf{x}_c = r_c \tag{13}$$

   - Transfer the correction on the fine grid

$$\mathbf{x}_f = \mathbf{x}_f + I_c^f\delta\mathbf{x}_c, \quad I_c^f : \text{ interpolation operator}$$

3. Post-smoothing: Apply ν_2 steps of an iterative method S_2 (most of the time identical to S_1) on a fine grid

$$\mathbf{A}_f \mathbf{x}_f = b_f, \quad \mathbf{x}_f = S_2^{\nu_2}(\mathbf{x}_f, b_f)$$

The extension of this two grid algorithm to a multi grid algorithm is recursively done by solving eq. 13 by a multigrid algorithm. Eq. 13 is replaced by

$$\text{For } n = 1 \ldots \gamma, MULTIGRID(\nu_1, \nu_2, \Omega_c, \delta\mathbf{x}_c, r_c)$$

The number γ of recursive calls determines the kind of multigrid algorithms: the best known been the V-cycle ($\gamma = 1$) and W-cycles ($\gamma = 2$) and are depicted on figure (5).

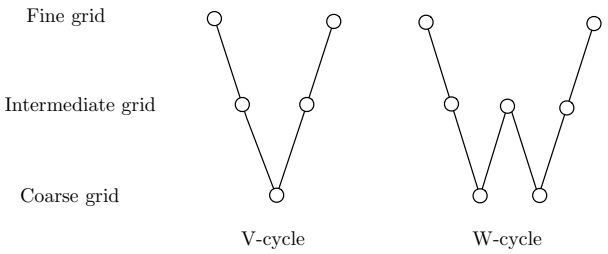


Figure 5. V-cycle and W-cycle algorithms in the case of three grid levels

Full Approximation Scheme variant In the multigrid algorithm introduced above, the coarse grid correction seeks a solution for the error $\delta\mathbf{x}_c$ (eq. 13). It is possible to rewrite the algorithm in an equivalent form where a coarse solution is instead searched. This is the Full Approximation Scheme (FAS) (Brandt 1982) which will have a natural extension to non linear operators and also permits the recursive use of multigrid acceleration techniques as will be introduced in (5.4). The coarse grid correction step is obtained by expressing the coarse grid correction $\delta\mathbf{x}_c$ as $\mathbf{x}_c - I_f^c \mathbf{x}_f$ and the problem is solved for the full approximation $\mathbf{x}_c$ instead of the correction $\delta\mathbf{x}_c$:

- Solve the problem on the coarse grid

$$\mathbf{A}_c \mathbf{x}_c = b_c \tag{14}$$

where

$$b_c = r_c + \mathbf{A}_c I_f^c \mathbf{x}_f = I_f^c b_f + \left(\mathbf{A}_c I_f^c - I_f^c \mathbf{A}_f\right) \mathbf{x}_f$$

- Transfer the correction on the fine grid

$$\mathbf{x}_f = \mathbf{x}_f + I_c^f \left(\mathbf{x}_c - I_f^c \mathbf{x}_f\right)$$

3.2. Ingredients of convergence: the smoothing and approximation properties

We refer to Hackbusch (2003) for a detailed explanation of the different ingredients of the convergence proof.
If $e_f = \mathbf{x}_f - \mathbf{x}^*$ is the error, it can be shown that

$$e_f^{\text{aftercoarsegridcorrection}} = \left(I - I_c^f \mathbf{A}_c^{-1} I_f^c \mathbf{A}_f\right) e_f^{\text{beforecoarsegridcorrection}}$$

Including the smoothing steps (and assuming only pre-smoothing is applied ($\nu_1 = \nu, \nu_2 = 0$)).

$$e_f^{k+1} \quad = \quad \left(I - I_c^f \mathbf{A}_c^{-1} I_f^c \mathbf{A}_f\right) S^\nu e_f^k$$

The smoothing steps S^ν should removed most of the error at small scales when the coarse grid correction step should remove large scales of the error. These two properties enable to show that the spectral radius of the multigrid iteration matrix $M = \left(I - I_c^f \mathbf{A}_c^{-1} I_f^c \mathbf{A}_f\right) S^\nu$ can be made less than one if a sufficient number of smoothing steps ν are applied.

Order of transfer operators and the approximation property The coarse grid correction matrix is defined by $\left(I - I_c^f \mathbf{A}_c^{-1} I_f^c \mathbf{A}_f\right)$. Let order($\mathbf{A}$) be the order of the differential operator that leads to the matrix $\mathbf{A}$. The following conditions on the order of the restriction (I_f^c) and prolongation (I_c^f) operators must hold (see Hemker (1990)):

$$order(I_f^c) \geq order(\mathbf{A}), \quad order(I_c^f) \geq order(\mathbf{A}) \tag{15}$$

Relations (15) corresponds to strong conditions that induce that the norm of the error amplification matrix (the

coarse grid correction matrix) but also the norm of the residual amplification matrix are bounded. These conditions are more restrictive than the usual ones $order(I_f^c) + order(I_c^f) \geq order(\mathbf{A})$ that induce only that the small scales component of the error are not amplified during one coarse grid correction step.

In addition, the approximation property that states that the coarse grid matrix $\mathbf{A}_c$ is close to the fine grid matrix $\mathbf{A}_f$ and induces a reduction of large scale error components is maintained by the underlying discretization order of the differential operator. When possible, another approach is to use the so called Galerkin approximation to deduce the coarse grid matrix $\mathbf{A}_c$. In this approximation, the restriction and interpolation operators are adjoints:

$$I_f^c = (I_c^f)^* \tag{16}$$

and the coarse grid matrix is defined by:

$$\mathbf{A}_c = I_f^c \mathbf{A}_f I_c^f \tag{17}$$

It can be shown that if these conditions are fullfiled, the components of the error that lie in the range of the interpolation operator before the coarse grid correction step are completely removed by the coarse grid correction step.

4. Application to data assimilation problems

We look at the main characteristics of the data assimilation problem in the light of the notions introduced previously: order of transfer operators, approximation property, ellipticity. In all our experiments, the refinement factor between the different grids is equal to 2 both in space and time. In this section, only the non preconditioned version of the data assimilation problem is addressed since we have seen that the preconditioning breaks the ellipticity of the original system. The study of the preconditioned version will be reintroduced in section (5) when the multigrid method will be used as a preconditioner instead of a solver.

4.1. Order of transfer operators

Now let's have a look on the conditions that have to fulfilled the transfer operators according to our definition of the background error covariance matrix B. For a number of iterations M_{laplac} equal to 4 in (11), the order of the differential operator corresponding to B^{-1} is equal to 8. So that the orders of the transfer operators must be superior to 8. High order transfer operators have to be used. Here we will used generalized Shapiro low pass filters (Purser 1987) based on the following formula for a restriction operator of order $2p$:

$$I_f^c \mathbf{x} = R_f^c \left[I_d + \sum_{k=p}^{k=2p-1} \alpha_k \Delta^k \right] \mathbf{x}$$

where R_f^c is the trivial injection operator (see details in appendix). The interpolations operators are taken to the adjoint, w.r.t. the Euclidean scalar product, of the restriction operators : $I_c^f = \rho (I_f^c)^T$, ρ being the mesh refinement factors between 2 grid levels ($\rho = 2$ in all our experiments).

4.2. Approximation property

For our data assimilation problem, we will use the Galerkin condition (17) for the definition of the background error covariance matrix at coarse resolution.

$$B_c^{-1} = I_f^c B_f^{-1} I_c^f$$

Application of the Galerkin condition for the observation term is not affordable since it would require to run the model at high resolution on each grid on the hierarchy. We begin by studying only the term corresponding to the regularization operator B^{-1}. Figure (6) represents the application of the coarse grid correction matrix $\left(I - I_c^f \mathbf{A}_c^{-1} I_f^c \mathbf{A}_f \right)$ (with $\mathbf{A} = B^{-1}$) to the discrete Fourier modes (here computed with $N = 64$).

The evidence of the use of high order transfer operators to correctly reduce the large scale components of the error is clearly seen.

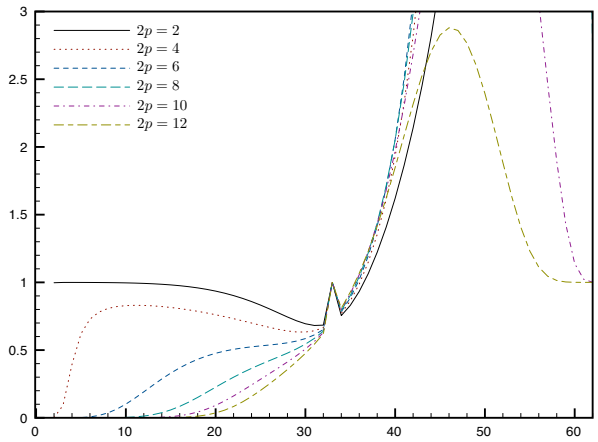


Figure 6. Approximation property for B, $M^{\text{imp}}_{\text{laplac}} = 4$, $N = 64$. The left part (resp. right part) of the picture corresponds to the large scale (resp. small scale) eigenvectors.

We now evaluate the approximation property when the observation term is present. We take a number of $N_{\text{obs}} = 10$ observations in time. An important parameter of the numerical discretization is the Courant number. For a Courant number equal to one, the numerical model is exact. For smaller value of the Courant number, numerical viscosity is added as can be seen on figure (7) and it greatly impacts the correctness of the coarse grid correction step.

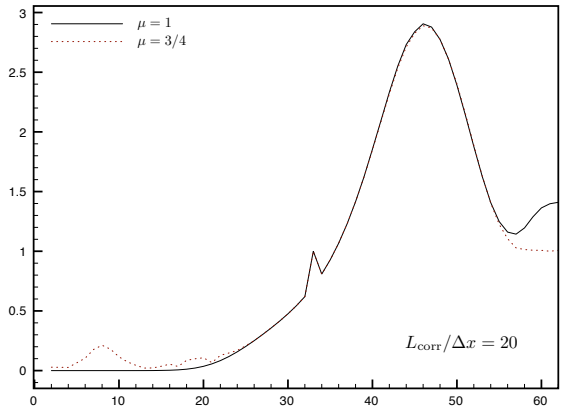


Figure 7. Coarse grid correction for a Courant number equal to 1 (exact numerical model) and for a Courant number equal to 3/4 for $N = 128$

The dissipative properties of the numerical model at coarse resolution breaks the approximation property and some of the large scale components of the error are not well reduced by the coarse grid correction step. The problem is amplified when the mesh is refined ($N = 256$) and/or when the background error correlation length is reduced (Figure (8)).

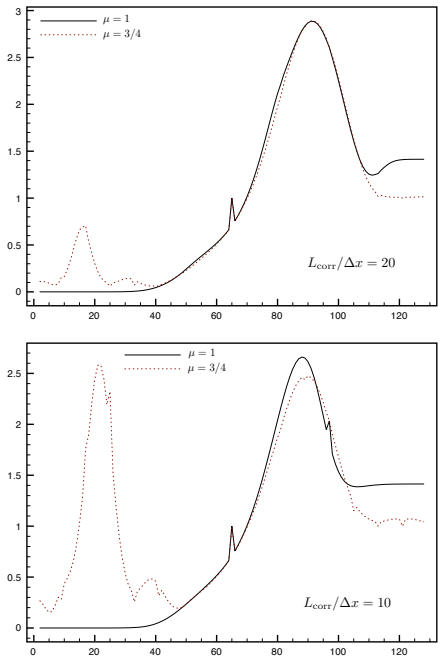


Figure 8. Coarse grid correction for a Courant number equal to 1 (exact numerical model) and for a Courant number equal to 3/4 for $N = 256$, for $L_{\text{corr}}/\Delta_x = 20$ (top) and $L_{\text{corr}}/\Delta_x = 10$ (bottom)

This will of course affect the performance of the multigrid method as we will see in section (5).

4.3. Smoothers

We now look at the different possibilities for the choice of a smoother for our data assimilation problem. We are exclusively concerned with black box smoothers where the only available information is in the matrix-vector product. For the solution of the symmetric positive definite system (5), the conjugate gradient (CG) method is often the method of choice. It is however well known that this algorithm cannot be considered as a smoother. Indeed the CG minimizes the $\mathbf{A}^{-1}$ norm of the residual $r(= \mathbf{A}x - b)$: $< r, \mathbf{A}^{-1}r >$ and thus for elliptic system can produce after a few iterations residuals at small scales with large amplitudes. Even if these residuals actually correspond to errors with small amplitudes, the restriction of these residuals to the coarse resolution grid can be aliased and accordingly produces wrong large scale correction of the control itself. As the CG algorithm, the minimal residual method (MINRES) is also based on a Lanczos iteration but produces iterates that minimize the Euclidian norm of the

residual instead of the $\mathbf{A}^{-1}$ norm for CG. In general this will lead to better smoothing properties than the conjugate gradient method.

More complex smoothers, that may require a lot of computational effort to be derived, exist and could be effective in particular when the same (or similar) matrix is used for successive minimizations, like it would be the case if the multigrid algorithm was used inside an incremental assimilation approach. As an example, sparse approximate smoothers based on the minimization of the Frobenius norm could be derived (Tang and Wan (2000)).

The smoother used in the next experiments is the MINRES algorithm.

5. Numerical experiments

5.1. Design of the assimilation experiments

The background x_b is taken to be

$$\mathbf{x}_b = 0.3 + \sin 2\pi \frac{x}{L}$$

while the true state is obtained by adding a Gaussian white noise corresponding to the B matrix:

$$\mathbf{x}_T = \mathbf{x}_b + B^{1/2}U, \qquad U \sim \mathcal{N}(0,1)$$

The standard deviation of the diagonal observation error covariance matrix is equal to $\sigma_{\rm obs} = 0.02 \approx 1.5\%$. The stopping criterion of all the experiments will be

$$\frac{\|\mathbf{x} - \mathbf{x}^\star\|_2}{\|\mathbf{x}^\star\|_2} \leq 10^{-4}$$

where $\mathbf{x}^\star$ is the solution of the assimilation problem (previously computed with a monogrid optimization).

5.2. 1DVAR experiments

For 1DVAR experiments, the numerical model is replaced by the identity and only one set of observations at time t_0 is used. We then run the monogrid and multigrid algorithm for different values of the ratio $\sigma_{\rm b}/\sigma_{\rm obs}$ and $L_{\rm corr}/\Delta x$. For this first series of experiments, only two grid levels are used. We focus here mainly on the number of fine grid iterations. The required number of iterations on the coarse grid is shown for information but will be discussed later. Here, on the coarse grid level, the resolution is done almost exactly using a B preconditioned minimization.

Table (3) shows the number of iterations to attain convergence for the monogrid preconditioned (P) and non preconditioned (NP) cases and for the multigrid non preconditioned (MGNP) case. The experiments are run using $W(1,1)$ cycles and the number of fine grid cells is here equal to $N = 128$. On table (3), the numbers between parentheses indicate the number of iterations on the coarse grid.

$\sigma_{\rm b}/\sigma_{\rm obs}$ \ $L_{\rm corr}/\Delta x$	6			20		
	P	NP	MGNP	P	NP	MGNP
5	2	167	2 (4)	5	330	2 (5)
20	2	179	6 (34)	5	473	2 (5)

Table 3. Number of iterations for $N = 128$ using $W(1,1)$ cycles, 2 grid levels

We can see that, since the correction is in the range of the B matrix, if the correlation length is large ($L_{\rm corr}/\Delta x = 20$), then the optimal solution can be captured by only one multigrid cycle (2 fine grid iterations).

In the rest of the paper, the ratio $\sigma_b/\sigma_{\rm obs}$ is taken equal to 20.

5.3. Numerical experiments: 2DVAR assimilation

We now evaluate the behavior of the multigrid algorithm in a 2DVAR context. The model is integrated over a period f $T = 78.125$s and ten observations are taken each 7.8125s so that a number of 10 instants of observations are used. The fine grid resolution is increased by choosing $N = 256$ leading to a more difficult test in the case of model errors. Two experiments are done: the first one with a model free of numerical errors (the Courant number μ is unity) and the other with a Courant number of $\mu = 3/4$ which leads to both dissipative and dispersive errors that affect the approximation property. The numbers of iterations required

for convergence are indicated on table (4) below.

μ \ $L_{\rm corr}/\Delta x$	20			10		
	P	NP	MGNP	P	NP	MGNP
1	32	345	4 (54)	40	93	6 (109)
3/4	44	439	26 (411)	68	130	∞

Table 4. Number of fine grid iterations for $N_{\rm obs} = 10$, $N = 256$ and a perfect model ($\mu = 1$)

In the case of a perfect model ($\mu = 1$), the performance of the multigrid method is clearly superior to the one of the preconditioned version with only a few fine grid iterations to converge. When numerical errors are introduced ($\mu = 3/4$) all the algorithms are affected and the number of iterations increase. The introduction of diffusive error impacts the controllability of the initial state. As it was shown in section (4.2), the multigrid algorithm also suffers from a less accurate approximation property and its behavior drastically degrades. In the case of a relatively small correlation length ($L_{\rm corr}/\Delta x = 10$), the algorithm diverges. As previously mentioned, results presented here have been done using the MINRES algorithm as a smoother but they are worse with the CG method. Figure (9) represents the true solution and the evolution of the control after the first multigrid cycles. It is clear that at the scales where the approximation property is not valid, the error increases from one cycle to another.

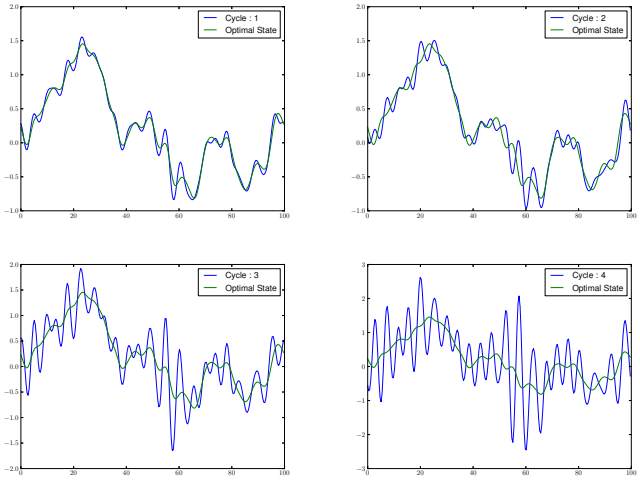


Figure 9. Evolution of the multigrid solution during the four first cycles and comparison with the true solution of the optimal control problem.

It is of course possible to make the process converge by increasing the number of pre- and post-smoothing steps, i.e. using $W(\nu, \nu)$ cycles with $\nu > 1$ to maintain convergence. The main difficult relies in the fact that, during these fine grid smoothing steps not only the small scales of the error have to be reduced but also the scales where the approximation property is not valid and these are difficult to predict so that ν can become quite large. This scheme clearly lacks of robustness. In addition, we would like to be able to use more than 2 levels and the approximation property still degrades as the number of levels is increased. The way to make the process more robust is to use the multigrid iteration as a preconditioner for an optimization method instead of a direct solver.

5.4. Multigrid as a preconditioner

As seen previously, the multigrid method applied to our toy data assimilation problem suffers from robustness due to several problems: the ellipticty of the non preconditioned version is not ensured particularly when background error correlation length is small, some of the errors of the numerical model (typically implicit diffusion) renders the approximation property less accurate. These problems are also common to several fields of application of multigrid methods (e.g. convection-diffusion problems with upwind discretizations see (Trottenberg *et al.* 2000, chap. 7)) and can sometimes be solved using specialized smoothers. Other methods have been developed to improve the behavior in this context. Multigrid acceleration techniques like the minimal residual method (Zhang 1998) or more general recombination of iterates of the multigrid process have been developed (Trottenberg *et al.* 2000, chap. 7). Another possibility is to integrate one multigrid cycle as a preconditioner for a Krylov minimization method. The original system $\mathbf{A}\mathbf{x} = b$ can be left preconditioned as

$$\mathbf{K}^{-1}\mathbf{A}\mathbf{x} = \mathbf{K}^{-1}b$$

where $\mathbf{K}^{-1}$ is an approximation of the inverse of $\mathbf{A}$ that will be given by one multigrid iteration. Following Tatebe (1993), the idea is here to choose $\mathbf{K}^{-1}$ as an operator which corresponds to one multigrid cycle. Remind that in one

multigrid iteration the evolution of the error $\mathbf{x}-\mathbf{x}^\star$ is given by :

$$e^{\text{after}} = \mathbf{M}e^{\text{before}}, \quad \text{where } \mathbf{M} = S_2^{\nu_2}\left(I - I_c^f \mathbf{A}_c^{-1} I_f^c \mathbf{A}_f\right) S_1^{\nu_1} \tag{18}$$

Using (18) it can be shown that the application of one multigrid cycle to a vector b and starting from a null initial guess leads to the following $\mathbf{K}^{-1}$ operator :

$$\mathbf{K}^{-1} = (I - \mathbf{M})\,\mathbf{A}^{-1} \tag{19}$$

where we have used the fact that $e^{\text{before}} = 0 - \mathbf{x}^\star = -\mathbf{A}^{-1}b$ and $e^{\text{after}} = \mathbf{K}^{-1}b - \mathbf{A}^{-1}b$. Ideally the error is completely removed in one multigrid cycle ($\mathbf{M}=0$) so that $\mathbf{K}^{-1}$ exactly corresponds to the inverse of $\mathbf{A}$. In the genereal case, the matrix $\mathbf{K^{-1}A}$ is given by $I-\mathbf{M}$. The application of the conjugate gradient on this matrix requires $I-\mathbf{M}$ to be symmetric positive definite which in turn requires ν_1 to be strictly positive for the definiteness while the symmetry can be ensured if $\nu_2=\nu_1$, $S_2=S_1$ being linear relaxation methods (Tatebe (1993)). These conditions are quite restrictive but the algorithm has been successfully applied with relaxed conditions in particular with $\nu_2=0$ (no post-smoothing) and ν_1 small (typically one) Bouwmeester *et al.* (2012)). An alternative is to use the multigrid preconditoning for a GMRES method (Oostelee and Washio 1998) that allows the use of non symmetric preconditioners. However several evaluations of the matrix vector product on the fine grid that we would like to avoid are still needed.

When neither pre- or post-smoothing steps are applied ($S_1^{\nu_1}=S_2^{\nu_2}=I$), the multigrid iteration matrix writes $\mathbf{K}^{-1} = I_c^f \mathbf{A}_c^{-1} I_f^c$, a matrix which is clearly not definite. Looking for alternatives to define a symmetric definite positive preconditioner, we propose a weighting of the multigrid iteration with the standard B preconditioning as follow:

$$\mathbf{K}^{-1} = I_c^f \mathbf{A}_c^{-1} I_f^c + (I - P_c^f I_f^c)^T B_f (I - P_c^f I_f^c) \tag{20}$$

where P_c^f is an interpolation operator. The symmetric positive definiteness of $\mathbf{A}_c$ and B_f implies the same properties for $\mathbf{K}^{-1}$. Indeed, $\mathbf{K}^{-1}$ is trivially positive and symmetric and the non singularity of $\mathbf{K}^{-1}$ is demonstrated by the fact that if $<\mathbf{K}^{-1}X_f, X_f>=0$ then $I_f^c\mathbf{x}_f=0$ and $(I-P_c^f I_f^c)X_f$=0 which implies $X_f=0$. An obvious choice for the operator P_c^f is the one which leads to the multigrid iteration matrix when the approximation property is valid and to the B_f preconditioning when it is not. This leads to the following choice of P_c^f:

$$P_c^f = \mathbf{A}_f I_c^f \mathbf{A}_c^{-1} \tag{21}$$

and we recognize in $(I - P_c^f I_f^c) = I - \mathbf{A}_f I_c^f \mathbf{A}_c^{-1} I_f^c$ nothing else than the coarse grid amplification matrix for the residual. However this definition of P_c^f introduces two applications of $\mathbf{A}_f$ and so has an important cost. In (21) the full matrix $\mathbf{A}$ can be approximated by the part which corresponds to the background term (B) and a still simpler expression is the one based only on the interpolation operator for which we have:

$$P_c^f = I_c^f \tag{22}$$

The objective of the preconditioner defined by (20,22) is obvious: the large scales components of the residual (for which $P_c^f I_f^c \approx I_d$) are preconditioned by the multigrid iteration matrix while its small scales components (for which $P_c^f I_f^c \approx 0$) are preconditioned by B_f (remarking that at small scales the background term is the dominant one in the $\mathbf{A}_f$ matrix). This preconditioner is symmetric positive definite and does not require one step of relaxation at high resolution.

Note that the advantage of the multigrid preconditioned algorithm is that it can also be applied on the preconditioned version of the system ($\mathbf{A}^P$). This can be important for problems where the matrix B^{-1} is not easily obtainable. When the multigrid algorithm is applied as a solver for $\mathbf{A}^P$, the non ellipticity of the problem makes that the

small scales of the errors are not quickly reduced by the minimization method and this leads, through aliasing on the coarse grid, to a divergence of the multigrid cycles. When the multigrid algorithm is used inside a Krylov method as a preconditioner, this amplification of the small scales by the multigrid cycle will not lead to a globally divergent algorithm and the large scales components of the error will be reduce at a lower cost on coarser resolution grids. In that case, the multigrid preconditioner (20) should be replaced by

$$\mathbf{K}^{-1} = I_c^f \mathbf{A}_c^{-1} I_f^c + (I - P_c^f I_f^c)^T (I - P_c^f I_f^c) \qquad (23)$$

where the obvious difference is that the term corresponding to small scales has already been preconditioned so that the B matrix has to be removed. The second term of the right hand side is only present to make the matrix $\mathbf{K}^{-1}$ non singular.

In the following, the usual multigrid preconditioner given by (19) will be denoted $\mathbf{K}_{MG}^{-1}$ while the weighting of the multigrid preconditioner with the B term will be written $\mathbf{K}_{BMG}^{-1}$. Two versions of both preconditioners are tested depend if if the original system was preconditioned (using $\mathbf{A}^P$) or not (using $\mathbf{A}^{NP}$).

Eigenvalues of the multigrid preconditioned matrix

The eigenvalues of the multigrid preconditioned matrix $\mathbf{K}^{-1}\mathbf{A}$ are plotted on figure (10) for the different expressions of the multigrid preconditioner $\mathbf{K}^{-1}$.

The hybrid preconditioners ($\mathbf{K}_{BMG}^{-1}$) have clearly lower condition numbers than the multigrid only preconditioners ($\mathbf{K}_{MG}^{-1}$). In particular, these last ones have very small eigenvalues and do not benefit from the B preconditioning. For the $\mathbf{K}_{MG}^{-1}$ preconditioners, All the preconditioners performs well in the sense that most of the eigenvalues are clustered around 1. For the usual multigrid preconditioned matrix $\mathbf{K}_{MGNP}^{-1}$, increasing the number of pre smoothing steps from one to two leads to increase of the smallest eigenvalues. But this is more efficiently satisfied by the

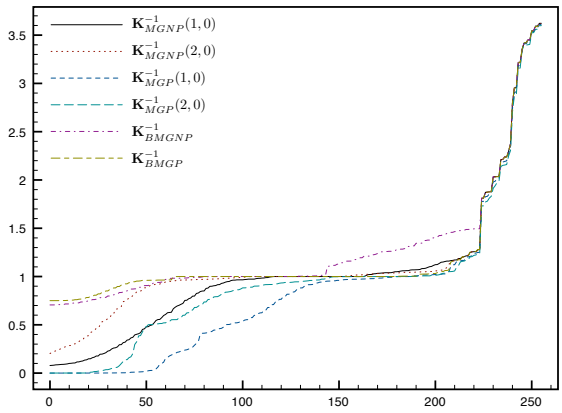


Figure 10. Eigenvalues spectrum for the multigrid preconditioner ($\mathbf{K}_{MG}^{-1}A$) for one and two presmoothing steps and for the hybrid multigrid / B preconditioned system ($\mathbf{K}_{BMG}^{-1}A$). $L_{corr}/\Delta x = 10$, $N = 256$, $N_{\text{obs}} = 10$

new preconditioner that use the B preconditioner at small scales. This preconditioner results in a very small condition number of $\mathcal{K}(\mathbf{K}_{BMGNP}^{-1}\mathbf{A}^{NP}) = 5.1, \mathcal{K}(\mathbf{K}_{BMGP}^{-1}\mathbf{A}^{P}) = 4.8$ while the monogrid non preconditioned and preconditioned have a condition number equal to $\mathcal{K}(\mathbf{A}^{NP}) = 2187$ and $\mathcal{K}(\mathbf{A}^{P}) = 6213$ respectively.

Choice of the number of iterations on the coarse grid level(s)

In practice, for an efficient algorithm, the coarse grid problem is not solved exactly. Using the FAS formulation (cf 3.1), the multigrid preconditioned conjugate gradient can be recursively extended to the case of more than two grid levels. At each grid level, the multigrid preconditioned CG is used to (approximately) solve the corresponding system. In the following, we will limit the number of iterations of the CG algorithm at each coarse grid level(s). When applied to an originally non preconditioned system ($\mathbf{A}^{NP}$), the numbers of iterations at each level can be very low and similar to the typical ones used when the multigrid method is used as a solver. This comes from the ellipticty of the $\mathbf{A}^{NP}$ matrix. For the originally preconditioned system ($\mathbf{A}^{P}$), the coarse grid level(s) are mainly present to solve the large scale of the errors at a lower computation cost. But on the coarser grids the rate of convergence is not smaller, so that more CG iterations are required. Note however that in general the cost of these coarse grid iterations are really

much cheaper and are often fully negligible in comparison with the cost of the finer grid integrations.

The use of approximate computations of $\mathbf{A}_c^{-1}$ in the expression of $\mathbf{K}^{-1}$ renders this preconditioner not constant over the conjugate gradient iterations. For this reason, we use the flexible form of the conjugate gradient (Notay 2000).

Numerical results of the multigrid preconditioned experiments

Tables (5,6) show the results of the multigrid preconditioned experiments for 2 and 3 grid levels when applied to $\mathbf{A}^{NP}$ (table 5) and $\mathbf{A}^{P}$ (table 6). The number of CG iterations on the coarse grids is limited to 30 in the 2 grid levels case. In the three grid levels case, it is limited to 3 when working on $\mathbf{A}^{NP}$ and limited to 5 on the second level and to 10 on the third level when working on $\mathbf{A}^{P}$. These limitations have been hand tuned and can probably be improved/adjusted using advanced control of the error between two successive grid levels (e.g. Thekale *et al.* 2010). However the chosen values reflects the fact mentioned above that the application of the multigrid preconditioned to an originally non preconditioned system ($\mathbf{A}^{NP}$ benefit from the ellipticity of this system, while the application of $\mathbf{A}^{P}$ requires more coarse grid(s) iterations.

$L_{\text{corr}}/\Delta x$	20	10
Monogrid	353	98
Multigrid Precon 2-levels	5 (150)	6 (180)
Multigrid Precon 3-levels	3 (9, 27)	7 (21, 126)

Table 5. Multigrid as a preconditioner on $\mathbf{A}^{NP}$ (original system under non preconditioned form): Number of fine and coarse grids iterations for $N_{\text{obs}} = 10$, $N = 256$ and $\mu = 3/4$

$L_{\text{corr}}/\Delta x$	20	10
Monogrid	27	51
Multigrid Precon 2-levels	3 (89)	6 (180)
Multigrid Precon 3-levels	6 (30, 300)	6 (30, 300)

Table 6. Multigrid as a preconditioner on $\mathbf{A}^{P}$ (original system under $B^{1/2}$ preconditioned form): Number of fine and coarse grids iterations for $N_{\text{obs}} = 10$, $N = 256$ and $\mu = 3/4$

The good behavior of the multigrid preconditioned algorithms is confirmed. The number of required fine grid iterations is much less than the monogrid experiments. This means that in a more realistic 3D application with a higher ratio between the computation cost of coarse and fine grids, the performance in term of computation time will be much less for the multigrid algorithms. While the use of 3 grid levels is clearly an advantage for the application of the multigrid preconditioning on $\mathbf{A}^{NP}$, it is less clear on $\mathbf{A}^{P}$ probably due to fundamental difference in their elliptic characteristics.

Increase of the resolution and the length of the assimilation windows: In these last experiments, the number of grid points is increased to $N = 4096$ while the length of the assimilation windows is extended to T=234s so that there are 30 times of observations. Figure (11) shows the evolution of the error $\frac{\|\mathbf{x}-\mathbf{x}^\star\|_2}{\|\mathbf{x}^\star\|_2}$ as a function of the number of iterations on the fine grid and as a function of the total computational time that includes the time spent on coarse grid levels and the time spent during grid interactions.

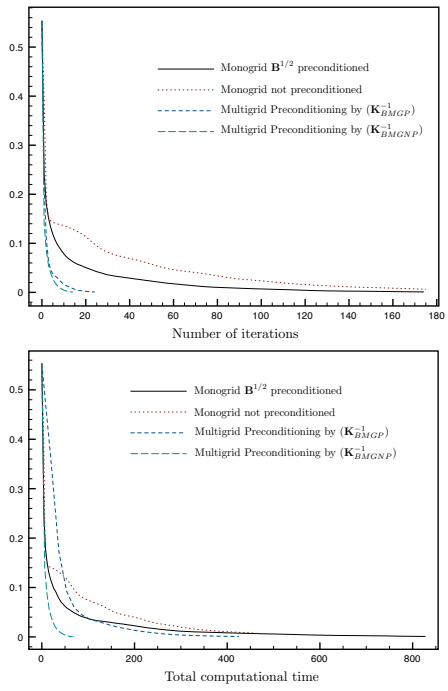


Figure 11. Evolution of the error as a function of the number of fine grid iterations (left) and of the total computational time in seconds (right) for $L_{\text{coor}}/\Delta x = 10$ and with three grid levels.

The number of required iterations for the monogrid cases has greatly increased and is now close to 200. One order of magnitude less iterations are required for the multigrid

preconditioned experiments. In term of total computational cost, the minimum is found to be for the multigrid preconditioned applied to $\mathbf{A}^{NP}$ and is less than 1/10 the cost of the monogrid cases.

When applied to $\mathbf{A}^{P}$, the multigrid method suffers from the number of required coarse grid iterations and from the fact that, in our model problem, these coarse grid iterations do not have a negligible cost. Indeed, with the chosen limitations on the number of iterations at each grid level, the number of total iterations during one fine grid step is given by 1 (fine grid) + 5 (second grid level) + 5×10 (third grid level). With a refinement ratio of 2 in space and time, this means that the computational cost associated with one fine grid iteration of the multigrid preconditioned algorithm is $1 + \frac{5}{2\times 2} + \frac{5\times 10}{(2\times 2)\times(2\times 2)} = 5.375$ times the cost of a monogrid iteration. As previously mentioned, this cost will be much less in a more realistic application in higher dimensions and potentially with higher mesh refinement factors.

6. Conclusion and perspectives

This paper has focused on the application of multigrid methods to the solution of variational data assimilation problem. Using a cost function with characteristics similar to the ones typically used in geophysical applications, we showed the importance role played by the background error covariance matrix and an eventual preconditioning on the problem ellipticity. Then on the model problem of a transport equation, the quality of the approximation of the fine grid system by a coarser resolution one has been investigated. Numerical errors, in particular of dissipative nature, alter the quality of this approximation property. Robustness and efficiency have been achieved by using multigrid iteration inside the preconditioning step of a conjugate gradient algorithm. The final results are very encouraging and show that, on our simple test case, the gain can be of one order of magnitude both in term of iterations and in term of total computational time.

The next step is obviously to experiment in a more complex model setting. The behavior of the approximation property is interesting to study when the problem has a less advective nature than our transport equation. In addition, when the multigrid method is used as a solver, it is worth to study the design of more complex smoothers specifically adapted. In the context of data assimilation for non linear problems, the incremental approach leads to a series of similar systems that may benefit from these advanced smoothers. When the multigrid method is used as a preconditioner, the choice of the underlying minimization method, that was here chosen to be a conjugate gradient algorithm may also be important. Finally, the multigrid preconditioner has been here coupled with a traditional preconditioning based on the square root of the background error covariance matrix. Hybridization of the multigrid preconditioner with more advanced preconditioner is also of prime interest.

References

Borzì A, Schulz V. 2009. Multigrid methods for pde optimization. *SIAM Rev.* **51**(2): 361–395, doi:10.1137/060671590, URL http://dx.doi.org/10.1137/060671590.

Bouwmeester H, Dougherty A, Knyazev AV. 2012. Nonsymmetric multigrid preconditioning for conjugate gradient methods. *CoRR* **abs/1212.6680**.

Brandt A. 1982. Guide to multigrid development. In: *Multigrid Methods*, *Lecture Notes in Mathematics*, vol. 960, Hackbusch W, Trottenberg U (eds), Springer Berlin Heidelberg, ISBN 978-3-540-11955-5, pp. 220–312, doi:10.1007/BFb0069930, URL http://dx.doi.org/10.1007/BFb0069930.

Briggs WL, Henson VE, McCormick SF. 2000. *A multigrid tutorial, second edition*. Society for Industrial and Applied Mathematics, second edn, doi:10.1137/1.9780898719505, URL http://epubs.siam.org/doi/abs/10.1137/1.9780898719505.

Carrier MJ, Ngodock H. 2010. Background-error correlation model based on the implicit solution of a diffusion equation. *Ocean Modelling* **35**(1–2): 45 – 53, doi:10.1016/j.ocemod.2010.06.003, URL http://www.sciencedirect.com/science/article/pii/S1463500310

Courtier P. 1997. Dual formulation of four-dimensional variational assimilation. *Quarterly Journal of the Royal Meteorological Society* **123**(544): 2449–2461, doi:10.1002/qj.49712354414, URL http://dx.doi.org/10.1002/qj.49712354414.

Evensen G. 2006. *Data assimilation: The ensemble kalman filter*. Springer-Verlag New York, Inc.: Secaucus, NJ, USA, ISBN 354038300X.

Gratton S, Toint PL, Tshimanga J. 2013. Conjugate gradients versus multigrid solvers for diffusion-based correlation models in data assimilation. *Quarterly Journal of the Royal Meteorological Society* : n/a–n/adoi:10.1002/qj.2050, URL http://dx.doi.org/10.1002/qj.2050.

Haben S, Lawless A, Nichols N. 2011. Conditioning and preconditioning of the variational data assimilation problem. *Computers & Fluids* **46**(1): 252 – 256, doi:10.1016/j.compfluid.2010.11.025, URL http://www.sciencedirect.com/science/article/pii/S0045793010003732 ¡ce:title¿10th ICFD Conference Series on Numerical Methods for Fluid Dynamics (ICFD 2010)¡/ce:title¿.

Hackbusch W. 2003. *Multi-grid methods and applications*. Springer Series in Computational Mathematics, Springer, ISBN 9783540127611, URL http://books.google.fr/books?id=ObBmg6ab-ukC.

Hemker P. 1990. On the order of prolongations and restrictions in multigrid procedures. *Journal of Computational and Applied Mathematics* **32**(3): 423 – 429, doi:10.1016/0377-0427(90)90047-4, URL http://www.sciencedirect.com/science/article/pii/0377042790900474.

Lax P, Wendroff B. 1960. Systems of conservation laws. *Communications on Pure and Applied Mathematics* **13**(2): 217–237, doi:10.1002/cpa.3160130205, URL http://dx.doi.org/10.1002/cpa.3160130205.

Le Dimet FX, Talagrand O. 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A* **38A**(2): 97–110, doi:10.1111/j.1600-0870.1986.tb00459.x, URL http://dx.doi.org/10.1111/j.1600-0870.1986.tb00459.x.

Lewis RM, Nash SG. 2005. Model problems for the multigrid optimization of systems governed by differential equations. *SIAM J. Sci. Comput.* **26**(6): 1811–1837, doi:10.1137/S1064827502407792, URL http://dx.doi.org/10.1137/S1064827502407792.

Mirouze I, Weaver AT. 2010. Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Quarterly Journal of the Royal Meteorological Society* **136**(651): 1421–1443, doi:10.1002/qj.643, URL http://dx.doi.org/10.1002/qj.643.

Neveu E, Debreu L, Le Dimet FX. 2011. Multigrid methods and data assimilation, convergence study and first experiments on non-linear equations. *ARIMA* **14**: 63–80.

Notay Y. 2000. Flexible conjugate gradients. *SIAM J. Sci. Comput.* **22**(4): 1444–1460, doi:10.1137/S1064827599362314, URL http://dx.doi.org/10.1137/S1064827599362314.

Oostelee C, Washio T. 1998. On the use of multigrid as a preconditioner. In: *Ninth International Conference on Domain Decomposition Methods*. Bergen.

Purser RJ. 1987. The filtering of meteorological fields. *Journal of Climate and Applied Meteorology* **26**(12): 1764–1769, doi:10.1175/1520-0450(1987)026¡1764:TFOMF¿2.0.CO;2, URL http://dx.doi.org/10.1175/1520-0450(1987)026<1764:TFOMF>2.0.C

Tang W, Wan W. 2000. Sparse approximate inverse smoother for multigrid. *SIAM Journal on Matrix Analysis and Applications* **21**(4): 1236–1252, doi:10.1137/S0895479899339342, URL http://epubs.siam.org/doi/abs/10.1137/S0895479899339342.

Tatebe O. 1993. The multigrid preconditioned conjugate gradient method. In: *Proc. of Sixth Copper Moutain Conference on Multigrid Methods*. NASA Conference Publication 3224, pp. 612–634.

Thekale A, Gradl T, Klamroth K, Rüde U. 2010. Optimizing the number of multigrid cycles in the full multigrid algorithm. *Numerical Linear Algebra with Applications* **17**(2-3): 199–210, doi:10.1002/nla.697, URL http://dx.doi.org/10.1002/nla.697.

Trottenberg U, Oosterlee C, Schuller A. 2000. *Multigrid*. Elsevier Science, ISBN 9780080479569, URL http://books.google.fr/books?id=9ysyNPZoR24C.

Weaver A, Courtier P. 2001. Correlation modelling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society* **127**(575): 1815–1846, doi:10.1002/qj.49712757518, URL http://dx.doi.org/10.1002/qj.49712757518.

Zhang J. 1998. Multi-level minimal residual smoothing: a family of general purpose multigrid acceleration techniques. *Journal of Computational and Applied Mathematics* **100**(1): 41 – 51, doi:10.1016/S0377-0427(98)00133-2, URL http://www.sciencedirect.com/science/article/pii/S0377042798(