A Comparison Between Interpolation and Assimilation as Cartography Methods for the SAGE II Aerosol Product

Ghislain Franssens, Dominique Fonteyn, Martine De Mazière, and Didier Fussen

BISA–Belgian Institute for Space Aeronomy, B-1180 Brussels, Belgium

We compare a new spherical interpolation algorithm and 4D Var transport assimilation as cartography methods for the SAGE II aerosol data. The new interpolation algorithm is based on a new concept called a Multipoint Taylor (MT) series. A MT series represents a function from its function and spatial derivative values at an irregular sample grid. Spatial derivatives are estimated by a local thin plate fit, capturing the local trend of the data. This is a fast and versatile spatial mapping tool, but it ignores the temporal variation of the data and thus results in an asynoptic map. The assimilation uses the Adjoint Model (AM) approach and is based on a simple transport model, using given wind fields. An iterative process minimizes a cost function of the squared error between measurements and model predictions, with respect to the initial state. The converged initial state is then propagated in time to produce synoptic maps. An introduction to and assessment of both methods is given. Some examples of maps for the SAGE II, validated, level-2, aerosol extinction coefficient at 1.02 μm, obtained by both methods, are presented.

1. INTRODUCTION

This paper presents two different tools for the production of aerosol optical extinction maps. The two approaches considered are:

i. a newly developed spherical interpolation method;

ii. 4D variational data assimilation, which produces a map that is consistent with the chosen atmospheric transport model and given wind fields.

The specific features of both methods are discussed and compared. Both methods are applied to NASA's Stratospheric Aerosol and Gas Experiment (SAGE) II, level-2, aerosol product.

Carefully prepared visualization of satellite data can significantly improve the comprehensibility of the data and make them more accessible to a wider user community. In particular, maps of the aerosol load of the atmosphere are helpful in interpreting dynamic transport features in the atmosphere. Further, various aerosol parameters influence the photochemical evolution and radiation equilibrium of the atmosphere. Maps are then an easy way to correlate the distribution of aerosol with other species in the atmosphere. Also, interpolating irregular scattered data to a regular grid is necessary if one wants to compare data obtained from different experiments for cross-validation purposes. Finally, the aerosol content of the atmosphere is a disturbing factor for many Earth remote sensing applications and often corrections have to be made for the aerosol extinction at specific locations.

Inverse Methods in Global Biogeochemical Cycles
Geophysical Monograph 114

Global mapping of aerosol data obtained from satellite occultation experiments (such as SAGE II) is not evident for the following reasons.

i. Sampling of the data is highly irregular and must first be transformed to a representation on a global regular grid.

ii. The raw data have a limited coverage of the globe, depending on the length of the time window considered. For example, for the SAGE II data, fairly global coverage is obtained only after about one month.

iii. Depending on the time of the year and inclination of the orbit, the data set obtained by solar occultation measurements may show uncovered areas, where data is completely missing.

iv. The data to be mapped are contaminated by both experimental errors and errors associated with the geometrical and spectral inversion schemes used to produce level-2 data. In addition, extinction data may be disturbed by the presence of high altitude cirrus clouds and/or Polar Stratospheric Clouds (PSC). In cases of high aerosol loading of the atmosphere (e.g. after a volcanic eruption), data in certain regions may be lacking because the atmosphere becomes too opaque for the instrument to function properly.

v. When assimilation is used as a mapping tool for aerosols, one needs particle number densities to propagate. For example in the case of the SAGE II data, the product consists of the optical extinction coefficient $\beta(1/\text{km})$, at the four wavelengths 0.385, 0.453, 0.525 and 1.02 μm [*Chu et al.*, 1989]. Therefore, one must make assumptions in deriving necessary additional properties. In practice, one usually assumes a simple physical aerosol model to make the conversion.

A new interpolation algorithm was developed for direct interpolation on the sphere. The choice of algorithms that interpolate irregularly scattered data on the sphere is very limited [*Dierckx*, 1995]. The most commonly used spherical algorithm, for Lagrangian interpolation (i.e. using function values only), is the one described in [*Renka*, 1982; *Renka and Cline*, 1984] and implemented in e.g. Interactive Data Language (IDL). If the samples have substantial error bars, as is the case for the SAGE II data, interpolation is even a more difficult task. It was found that the (IDL implementation) of the above algorithm, failed (due to sample points being too close) or produced totally unacceptable maps (due to noise on the samples).

Our algorithm addresses the more general Hermite-Birkhoff interpolation problem (using given function and derivative values), for data randomly scattered on the sphere. As far as we are aware of, no robust and practical algorithm is available today to solve this problem. Our solution to this problem consists in the construction of an interpolant that has the form of a truncated Multipoint Taylor (MT) representation. This representation uses both function and partial derivative values at the sample points and is a natural solution to the Hermite-Birkhoff interpolation problem. By solving this more general problem, the algorithm designer can control what values the derivatives of the interpolant at each sample point will have. In classical methods, these values are implicitly fixed by the choice of the representation, for example by using polynomials or spherical harmonics. We obtain the partial derivative values, required by the MT representation, from a thin plate fitting procedure, applied to the neighborhood of each sample point. This fit captures the local trend of the data and also reduces the effect of the measurement errors on the interpolation result. The interpolation step is optionally followed by a smoothing operation (low-pass spatial frequency filtering) in order to remove possible unphysical, fast variations in the resulting maps. The end product is then a two-dimensional (global) map, showing, for instance, the aerosol extinction coefficient at a certain altitude or the total column optical extinction of the whole atmosphere.

Our second approach aims at the production of maps that are consistent with the dynamical behavior of the atmosphere. As basis for the assimilation, it is assumed that aerosol particles are inert tracers. In the case of aerosols under normal conditions, this is a good approximation and we can take the atmosphere model to be a pure transport model. In perturbed conditions however, such as at PSC formation, volcanic eruptions, etc., the chemistry and microphysics of the aerosols should be considered also. In this article we restrict ourselves to normal conditions. More precisely, we assume that the aerosol particle number density is proportional to the extinction coefficient. The particle density field is then propagated through 3D space by numerically solving the initial value transport problem. The model dynamics is based on wind field analyses from the European Centre for Medium-range Weather Forecasts (ECMWF). The optimal initial number density field is sought that minimizes the total squared error

between model predicted and measurement values in those space-time points where observations are available. This is done, by minimizing in an iterative way, a cost function of the difference between observations and model predictions. Once the optimal initial field is found, propagating it in time yields the synoptic aerosol particle density map.

2. THE INTERPOLATION METHOD

The interpolation method has the following features.

i. The surface of the sphere is decomposed by spherical Delaunay triangulation.

ii. The interpolant is represented as a truncated MT series.

iii. The local behavior of the data is estimated from a thin plate fitting procedure.

iv. An optional, 2D spatial spectral filter can be applied.

The resulting interpolant is infinitely differentiable everywhere. The computational cost is linear with the number of sample points and the method was found to be always numerically stable.

2.1. Domain Decomposition

In order to have a computational cost that is linear in the number of sample points, it is necessary to have a local interpolation algorithm. This means that the value in an interpolation point should only depend on the nearest sample points. This is achieved by using compact support basis functions in the MT representation. To this end, we decompose the surface of the globe into spherical Delaunay triangles. As compact support area for the basis functions we use the interior of the Voronoi polygon around each sample point.

2.2. The Multipoint Taylor Series Interpolant

The interpolant $\tilde{f}(\mathbf{x})$, in a point $\mathbf{x} = (\theta, \varphi)$, is represented as a truncated MT series, having the form

$$\tilde{f}(\mathbf{x}) = \sum_{n=1}^{N} \sum_{k+l=0}^{M_n} f_n^{k,l} U_n^{k,l}(\mathbf{x}) \tag{1}$$

wherein the $f_n^{k,l}$ are the values of the partial derivatives at the sample points,

$$f_n^{k,l} = \left. \frac{\partial^{k+l} f(\theta, \varphi)}{\partial \theta^k \partial \varphi^l} \right|_{\mathbf{x}=\mathbf{x}_n} \tag{2}$$

and the $U_n^{k,l}(\mathbf{x})$ are basis functions. N is the number of sample points and M_n the number of partial derivatives retained in the approximation at the n^{th} point. Higher order basis functions $U_n^{k,l}(\mathbf{x}), k+l > 0$ are expressed in terms of zero order functions $U_n^{0,0}(\mathbf{x})$ as

$$U_n^{k,l}(\mathbf{x}) = \frac{(\theta - \theta_n)^k}{k!} \frac{(\varphi - \varphi_n)^l}{l!} U_n^{0,0}(\mathbf{x}) \tag{3}$$

The zero order basis functions $U_n^{0,0}(\mathbf{x})$ are constructed in such a way that they satisfy

$$U_n^{0,0}(\mathbf{x})\big|_{\mathbf{x}=\mathbf{x}_m} = \delta_{mn}, \quad 1 \le n, m \le N \tag{4a}$$

$$\left. \frac{\partial^{i+j} U_n^{0,0}(\theta, \varphi)}{\partial \theta^i \partial \varphi^j} \right|_{\mathbf{x}=\mathbf{x}_m} = 0, \quad \begin{matrix} i+j > 0 \\ 1 \le n, m \le N \end{matrix} \tag{4b}$$

In addition, to ensure convergence to the correct function, the zero order functions $U_n^{0,0}(\mathbf{x})$ must form a partition of unity over the sphere

$$\sum_{n=1}^{N} U_n^{0,0}(\mathbf{x}) \equiv 1 \tag{5}$$

It is possible to construct basis function with the required properties. In our implementation, each zero order basis function is constructed over its associated Voronoi polygon, in such a way that it equals one at the interior sample point and zero outside and on the boundary of the Voronoi polygon. In addition, it has all its partial derivatives zero at the interior point, as well as on the Voronoi boundary. For the explicit construction of the basis functions we refer to [*Franssens*, 1999a,b].

2.3. Local Thin Plate Fit

The general formula (1) requires partial derivative values, which are seldom available. There are various ways to (numerically) estimate derivatives. The most obvious one is using finite difference formulas. However, in practice function values are often contaminated by measurement errors, which makes this is a very error prone method. We use a more robust technique that circumvents the explicit calculation of the derivatives.

We start with approximating the unknown function locally by a thin plate, bent in such a way that it passes through (or near) the sample values. By using the physical model of an elastic, thin plate and determining its shape from the minimization of its potential energy, one is assured that no wild oscillations or overshoots will occur in the resulting interpolant. So, we search

in the neighborhood of each sample point an optimal shape function $s_n(\mathbf{x})$, which minimizes the following functional

$$\xi(s_n) = \delta(s_n) + \lambda\eta(s_n) \tag{6}$$

The first term is a weighted, least square cost function,

$$\delta(s_n) \triangleq \sum_{i=1}^{I_n} w_i \left(f_i - s_n(\mathbf{x}_i)\right)^2 \tag{7}$$

The summation runs over the considered sample point $\mathbf{x}_n$ and its N_n neighbors on the Voronoi boundary, so $I_n = N_n + 1$. The numbers w_i are weights, which can take into account the error on the function values f_i. The second term in (6) is an approximation of the total curvature of the plate shape $s_n(\mathbf{x})$

$$\eta(s_n) \triangleq \int_{R^2} \left(\left(\frac{\partial^2 s_n}{\partial\theta^2}\right)^2 + 2\left(\frac{\partial^2 s_n}{\partial\theta\partial\varphi}\right)^2 + \left(\frac{\partial^2 s_n}{\partial\varphi^2}\right)^2 \right) d\Omega \tag{8}$$

and λ plays the role of fitting parameter. The larger this parameter, the larger the stiffness of the fitted shape becomes and rapid spatial variations are suppressed. The selected value usually depends on the quality of the data to be interpolated. Typical values are in the range $0 \leq \lambda \leq 1$ for high quality data, to $1 < \lambda < 5$ for data with larger errors (above 5%).

One part of the approximation in (8) lies in the planar, instead of spherical, measure $d\Omega = d\theta d\varphi$, used in the integral. Since we use local basis function, this is justified because we only need that portion of the shape function over the interior of the Voronoi polygon, i.e. in the immediate neighborhood of the considered sample point. As a consequence of this approximation, it becomes possible to obtain the optimal shape function $s_n(\mathbf{x})$ as a linear combination of simple analytical functions. The expansion coefficients are obtained by solving a small linear system (typically of size 10×10), at each sample point.

We now use the derivatives of the found thin plate shape in (1) and by using (3) obtain the modified interpolant

$$\begin{aligned} \tilde{f}(\mathbf{x}) &= \sum_{n=1}^{N} \sum_{k+l=0}^{+\infty} s_n^{k,l} U_n^{k,l}(\mathbf{x}) \\ &= \sum_{n=1}^{N} \sum_{k+l=0}^{+\infty} s_n^{k,l} \frac{(\theta-\theta_n)^k}{k!} \frac{(\varphi-\varphi_n)^l}{l!} U_n^{0,0}(\mathbf{x}) \\ &= \sum_{n=1}^{N} s_n(\mathbf{x})\, U_n^{0,0}(\mathbf{x}) \end{aligned} \tag{9}$$

This final result shows that only the shape function itself is needed.

Because the support of the basis functions is limited to its Voronoi polygon, only a local portion of $s_n(\mathbf{x})$ around its sample point is used. Moreover, any interpolated value is just a linear combination of only three thin plate shapes, associated with the three vertexes of the triangle in which the interpolation point is located. So, for any interpolation point $\mathbf{x}$, the sum in (9) is reduced to three terms. This is the implemented formula. A more detailed treatment can be found in [*Franssens*, 1999b].

2.4. *Final Filtering*

After interpolation, as computed by (9), an optional spectral filter can be applied. The main purpose is to remove unrealistic variations in the result, that have a higher spatial frequency than the Nyquist sampling frequency. The result is a smoothing effect. In the examples shown in section 4, a low-pass, spatial filter of Gaussian shape

$$g(f_x, f_y) = \exp\left(-\left((f_x/f_a)^2 + (f_y/f_b)^2\right)\right) \tag{10}$$

was used. Herein are f_a and f_b filter characteristics, determining the spectral width of the filter, and depending on the sampling characteristics of the data.

3. THE ASSIMILATION METHOD

Several methods have been developed for data assimilation. Sequential methods put a model in a state, which in general is not consistent with its dynamics and/or physics. This disturbs the model and it needs some time to reach consistency. To keep the disturbance as small as possible, data are prepared and only a correction of the model state "in the direction" of the data takes place. Sequential methods such as Nudging, Successive Correction, Optimal Interpolation, Kalman Filtering, etc., differ in the degree of consistency the correction has, with respect to the model dynamics and/or physics.

In contrast, the Adjoint Method (AM) always guarantees full consistency with the dynamics and/or physics [*Giering*, 1999]. By varying control variables, one adjusts model trajectories to be as close as possible to the measured data. In data assimilation, the control variables are typically the initial conditions or the forcing for the model [*Talagrand and Courtier*, 1987]. To quantify the misfit of the model prediction, one defines

a cost function, which is then minimized with an iterative algorithm. Starting with a first guess, each iteration step computes an improved vector of control variables. Thereby, the search direction is computed from the gradient of the cost function with respect to the control variables. The AM serves computing this gradient vector.

3.1. *The Adjoint Assimilation Method*

We give an introduction to the AM that parallels the one in [*Fisher and Lary*, 1995]. Consider a physical system and a model describing this system. Let $\mathbf{y}^o$ be an M-dimensional vector of observations of the physical system and $\mathbf{y}$ the M-dimensional vector computed by the model, at the position and time of these observations. To quantify the misfit, define a scalar product $(.,.) : R^M \times R^M \to R$, such that $\mathbf{x}, \mathbf{y} \mapsto (\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{M} x_i y_i$ and a cost function

$$j \triangleq \frac{1}{2}(\mathbf{y} - \mathbf{y}^o, \mathbf{y} - \mathbf{y}^o) \tag{11}$$

In order to manipulate the model, we introduce an N-dimensional unknown control (also called state) vector $\mathbf{x}$ (in practice, the initial values of the initial value problem). The dependence of $\mathbf{y}$ on $\mathbf{x}$ within the model will be given by a mapping $\mathbf{h} : R^N \to R^M$, such that $\mathbf{x} \mapsto \mathbf{y} = \mathbf{h}(\mathbf{x})$. In general, this mapping includes a model run to near the place and time of the observations and also contains an interpolation to the exact measurement location. Then the cost function can be regarded as a mapping $j : R^N \to R$, such that $\mathbf{x} \mapsto j(\mathbf{x})$ with

$$j(\mathbf{x}) = \frac{1}{2}(\mathbf{h}(\mathbf{x}) - \mathbf{y}^o, \mathbf{h}(\mathbf{x}) - \mathbf{y}^o) \tag{12}$$

The problem is to determine the (unique) value of $\mathbf{x}$ that minimizes j. When the number of observations is limited, there might not exist a unique global solution or uniqueness is only achieved in some neighborhood of the measurements points. Uniqueness of the global solution can be forced by the introduction of a background term (see next section). Effective minimization algorithms require the gradient $\nabla j(\mathbf{x})$ of j with respect to $\mathbf{x}$ at a given point $\mathbf{x}_0$. To this end, we compute the first order variation δj of the cost function with respect to the control vector at $\mathbf{x} = \mathbf{x}_0$. We have

$$\delta j = \left(\nabla j(\mathbf{y})|_{\mathbf{y}_0 = \mathbf{h}(\mathbf{x}_0)}, \delta \mathbf{y}\right) \tag{13}$$

To continue we need the linearization of the (in general non-linear) mapping $\mathbf{h}$ around $\mathbf{x}_0$. Supposing $\mathbf{h}$ is regular, it has a Taylor series expansion around $\mathbf{x}_0$. The first order variation of $\mathbf{y}$ can be related to the first order variation in $\mathbf{x}$ by

$$\delta \mathbf{y} = \mathbf{H}(\mathbf{x}_0)\delta \mathbf{x} \tag{14}$$

wherein the linear operator (matrix) $\mathbf{H}(\mathbf{x}) : R^N \to R^M$, such that $\mathbf{x} \mapsto \mathbf{H}(\mathbf{x}) \triangleq \nabla \mathbf{h}(\mathbf{x})$ denotes the linearization or Jacobian of $\mathbf{h}$ at $\mathbf{x}$. Performing the differentiation in (13), using (11) and (14), we get for (13)

$$\delta j = (\mathbf{h}(\mathbf{x}_0) - \mathbf{y}^o, \mathbf{H}(\mathbf{x}_0)\delta \mathbf{x}) \tag{15}$$

Using the definition of the adjoint operator $\mathbf{H}^*(\mathbf{x}) : R^M \to R^N$, satisfying for any $\mathbf{u} \in R^N$, $\mathbf{v} \in R^M$,

$$(\mathbf{v}, \mathbf{H}\mathbf{u}) = (\mathbf{H}^*\mathbf{v}, \mathbf{u}) \tag{16}$$

equation (15) can be written as follows

$$\delta j = (\mathbf{H}^*(\mathbf{x}_0)(\mathbf{h}(\mathbf{x}_0) - \mathbf{y}^o), \delta \mathbf{x}) \tag{17}$$

Consequently, according to (13), the gradient of the cost function with respect to the control variables must be

$$\nabla j(\mathbf{x}) = \mathbf{H}^*(\mathbf{x})(\mathbf{h}(\mathbf{x}) - \mathbf{y}^o) \tag{18}$$

at any point $\mathbf{x}$ where the mapping $\mathbf{h}$ is regular. The linear operator $\mathbf{H}(\mathbf{x})$ represents the tangent linear model at $\mathbf{x}$. Its adjoint linear operator $\mathbf{H}^*(\mathbf{x})$ represents the adjoint linear model. Both operators depend on the point $\mathbf{x}$ at which the model is linearized. According to (18), the difference $\mathbf{h}(\mathbf{x}) - \mathbf{y}^o$ can be interpreted as the forcing of the adjoint model.

A detailed analysis of the required basic numerical operations yields that the computation of the cost function and its gradient, for a given vector of control variables, takes only 2-5 times the computation of the cost function. Alternatively, the gradient vector $\nabla j(\mathbf{x})$ could be approximated by finite differences, which needs at least $N+1$ computations of the cost function. The use of the AM approach thus has two advantages. First, it saves a lot of run time, especially for large N, and secondly, the computed gradient, needed to minimize the cost function, is exact. Notice that, from the definition of the inner product and the fact that the matrix $\mathbf{H}$ is a Jacobian matrix, always hold that $\mathbf{H}^* = \mathbf{H}^T$.

3.2. *Formulation of the 4D Var Algorithm*

Based on the previous theoretical framework, four-dimensional, variational data assimilation (4D Var) can now be formulated. 4D Var seeks to produce an analysis, which fits a set of observations taken over a period

of time, subject to the strong constraint that the evolution of the analyzed quantities is governed by a deterministic model. By imposing the equations of the model as strong constraints, the analysis problem is reduced to that of determining initial values for the model, such that the subsequent evolution minimizes a cost measure of the fit to the observations.

The 4D Var uses the following cost function

$$j \triangleq j_b + j_{m=0} \tag{19a}$$

$$j_b \triangleq \frac{1}{2}(\mathbf{x}_b - \mathbf{x}_0)^T \mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x}_0) \tag{19b}$$

$$j_m \triangleq \sum_{k=m}^{K} \frac{1}{2}(\mathbf{y}_k - \mathbf{y}_k^o)^T (\mathbf{R}_k)^{-1}(\mathbf{y}_k - \mathbf{y}_k^o) \tag{19c}$$

Herein is $\mathbf{x}_0$ the vector of initial concentrations, $\mathbf{x}_b$ an independent estimate of the initial concentrations, and $\mathbf{B}$ the covariance matrix of expected errors in $\mathbf{x}_b$. The expression j_b is usually called the 'background term' of the cost function, and $\mathbf{x}_b$ 'the background'. The vector $\mathbf{y}_k^o$ in j_0, given by (19c) for $m = 0$, consists of all observations which are considered valid at the timestep k, and $\mathbf{y}_k$ is the vector of model predictions for these observations. In the analysis presented here, $\mathbf{y}_k$ is a linear function of concentrations, i.e.

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k \tag{20}$$

Herein is $\mathbf{x}_k$ the vector of concentrations at timestep k, and $\mathbf{H}_k$, which maps $\mathbf{x}_k$ to model equivalents of observational values, is sometimes called the 'observation operator'. As minimum aerosol model we assume that extinction is proportional to concentration, so the $\mathbf{H}_k$ are constant diagonal matrices. The matrix $\mathbf{R}_k$ is the covariance matrix for the random errors in the term $\mathbf{y}_k - \mathbf{y}_k^o$, which would be expected given a perfect analysis. With the random errors we mean the 'observation errors' and 'representativeness errors' introduced in simulating the observations.

The strong constraints of the model equations are incorporated into the analysis by regarding j as a function of the initial concentrations $\mathbf{x}_0$. Concentrations at subsequent times are determined by integrating the model equations forward in time. This procedure results in two major simplifications. First, it eliminates the need to use a minimization algorithm with constrains. Numerical algorithms for unconstrained minimization are considerably more efficient and less prone to problems of ill-conditioning, than are algorithms for constrained minimization. Secondly, the number of independent variables is reduced by a factor of $K + 1$.

We use a descent algorithm to produce a convergent sequence of estimates of the vector $\mathbf{x}_0$ that minimizes the cost function. It requires the gradient of the cost function with respect to $\mathbf{x}_0$. This gradient is evaluated by integrating the adjoint model. This is done in the following way.

Consider an infinitesimal variation, $\delta\mathbf{x}_0$, in the initial concentrations in the partial cost function j_m, given by (19c). At each timestep $m \leq k \leq K$, there will be corresponding infinitesimal variations, $\delta\mathbf{x}_m$ and δj_m in the concentrations and in the function j_m. This j_m depends only on concentrations at steps $m \leq k \leq K$. Since these concentrations are uniquely determined by the equations of the model and those for any step $0 \leq k' < m$, it is legitimate to regard j_m as a function of the densities at step k' only. By definition, the gradient of j_m with respect to the concentrations at step k' satisfies the equation

$$\delta j_m = (\nabla_{\mathbf{x}_{k'}} j_m)^T \delta\mathbf{x}_{k'} \tag{21}$$

Suppose that $\nabla_{\mathbf{x}_m} j_m$ is known. We now want to calculate $\nabla_{\mathbf{x}_{m-1}} j_{m-1}$ by induction. First, from (19a) we have

$$\nabla_{\mathbf{x}_0} j = \nabla_{\mathbf{x}_0} j_b + \nabla_{\mathbf{x}_0} j_0 \tag{22}$$

From (19c), we can write

$$\nabla_{\mathbf{x}_{m-1}} j_{m-1} = \nabla_{\mathbf{x}_{m-1}} t_{m-1} + \nabla_{\mathbf{x}_{m-1}} j_m \tag{23a}$$

where we defined,

$$t_m \triangleq \frac{1}{2}(\mathbf{y}_m - \mathbf{y}_m^o)^T (\mathbf{R}_m)^{-1}(\mathbf{y}_m - \mathbf{y}_m^o) \tag{23b}$$

The ease with which the first term in (23a) may be evaluated depends on the complexity of the model. To evaluate the second term in (23a), notice that for an arbitrary infinitesimal variation in $\mathbf{x}_{m-1}$, we have by (21)

$$\delta j_m = (\nabla_{\mathbf{x}_{m-1}} j_m)^T \delta\mathbf{x}_{m-1} \tag{24}$$

and by definition

$$\delta j_m = (\nabla_{\mathbf{x}_m} j_m)^T \delta\mathbf{x}_m \tag{25}$$

Hence,

$$(\nabla_{\mathbf{x}_{m-1}} j_m)^T \delta\mathbf{x}_{m-1} = (\nabla_{\mathbf{x}_m} j_m)^T \delta\mathbf{x}_m \tag{26}$$

If the equations of the (discretized) model are written in the form

$$\mathbf{x}_m = \mathbf{h}_{m-1}(\mathbf{x}_{m-1}) \tag{27}$$

then, for infinitesimal variations,

$$\delta\mathbf{x}_m = \mathbf{H}_{m-1}(\mathbf{x}_{m-1})\delta\mathbf{x}_{m-1} \tag{28}$$

where $\mathbf{H}_{m-1}(\mathbf{x}_{m-1})$ is the Jacobian containing the partial derivatives of the elements of $\mathbf{h}_{m-1}(\mathbf{x}_{m-1})$ with respect to the elements of $\mathbf{x}_{m-1}$. Subsituting (28) into (26) gives, after transposing

$$\nabla_{\mathbf{x}_{m-1}} j_m = (\mathbf{H}_{m-1})^T(\mathbf{x}_{m-1})\nabla_{\mathbf{x}_m} j_m \tag{29}$$

Hence, substituting (29) in (23a) yields

$$\nabla_{\mathbf{x}_{m-1}} j_{m-1} = \nabla_{\mathbf{x}_{m-1}} t_{m-1} + (\mathbf{H}_{m-1})^T(\mathbf{x}_{m-1})\nabla_{\mathbf{x}_m} j_m \tag{30}$$

This is the adjoint tangent linear (ATL) equation. Given $\nabla_{\mathbf{x}_m} j_m$, equation (30) allows $\nabla_{\mathbf{x}_{m-1}} j_{m-1}$ and, by induction, $\nabla_{\mathbf{x}_0} j_0$ to be calculated. Once $\nabla_{\mathbf{x}_0} j_0$ is known, the required gradient of the cost function with respect to the initial conditions is given by (22). Starting the induction at step K we have

$$\nabla_{\mathbf{x}_K} j_K = \nabla_{\mathbf{x}_K} t_K \tag{31}$$

The algorithm used to minimize j is then as follows.

i Start with an initial guess for $\mathbf{x}_0$.

ii Integrate the model to give $\mathbf{x}_k$ for $k = 1,...,K$.

iii Evaluate j. If it is small enough, STOP.

iv Iterate the ATL equations to calculate $\nabla j(\mathbf{x}_0)$.

v Use descent algorithm to find an $\mathbf{x}_0$, reducing j.

vi GOTO (ii).

3.3. Implementation of the 4D Var Algorithm

The model behind the assimilation consists of two parts:

i the pure advection transport model;

ii the physical aerosol model.

In the transport model we used a T21 horizontal resolution, 5 vertical stratospheric levels and a 6-hour time step. This 3-D grid uses the same coordinates as the low-resolution version of the ECMWF Numerical Weather Prediction (NWP) model. The advection is implemented using the semi Lagrangian scheme of [*Smolarkiewicz and Rasch*, 1991]. The driving dynamical fields are either analyzed ECMWF fields or output from a T63 resolution run from ECWMF's NWP. Any aerosol parameter information, contained in the SAGE II data, is necessarily restricted due to the limited spectral coverage of the measurements. Since the optical extinction at 1.02 μm is proportional to the specific surface area density [*Thomasson et al.*, 1997], we have extrapolated this to aerosol total number density. Thus our aerosol model is very simple: total number density is taken to be proportional to the optical extinction coefficient and aerosols are considered basically inert tracers. Events as thin cirrus clouds and PSC are not taken into account. In order to avoid contamination of the assimilation results by cirrus cloud or PSC, observations below the isentropic temperature level of 195 K were rejected.

The background term (19b) in the cost function ensures the uniqueness of the solution of the minimization. In the presented case, the background term was constructed by the interpolation of the SAGE II 1.02 μm optical extinction data of December 1988 onto the model grid. The covariance matrix of the expected errors in the background term was derived from the variability at each model grid point in the interpolated fields for the months December, January and February for 2 consecutive winters, 1987-1988 and 1988-1989.

The minimization process requires the explicit knowledge of the gradient vector of the cost function with respect to the model initial state. This gradient vector was determined by backward integration of the adjoint model. The adjoint model was constructed with the aid of an automatic differentiation tool [*Giering*, 1997]. The quasi-Newton algorithm M1QN3 from INRIA [*Gilbert and Lemaréchal*, 1989] has been used in the minimization of the cost function.

4. EXAMPLES

4.1. Interpolation

A test was devised to measure the self-consistency of the interpolant. It consisted of removing a single sample point from the set of samples and to let the interpolation algorithm reproduce the removed sample value from the remaining ones. This was done in succession for each sample point and the resulting errors were averaged.

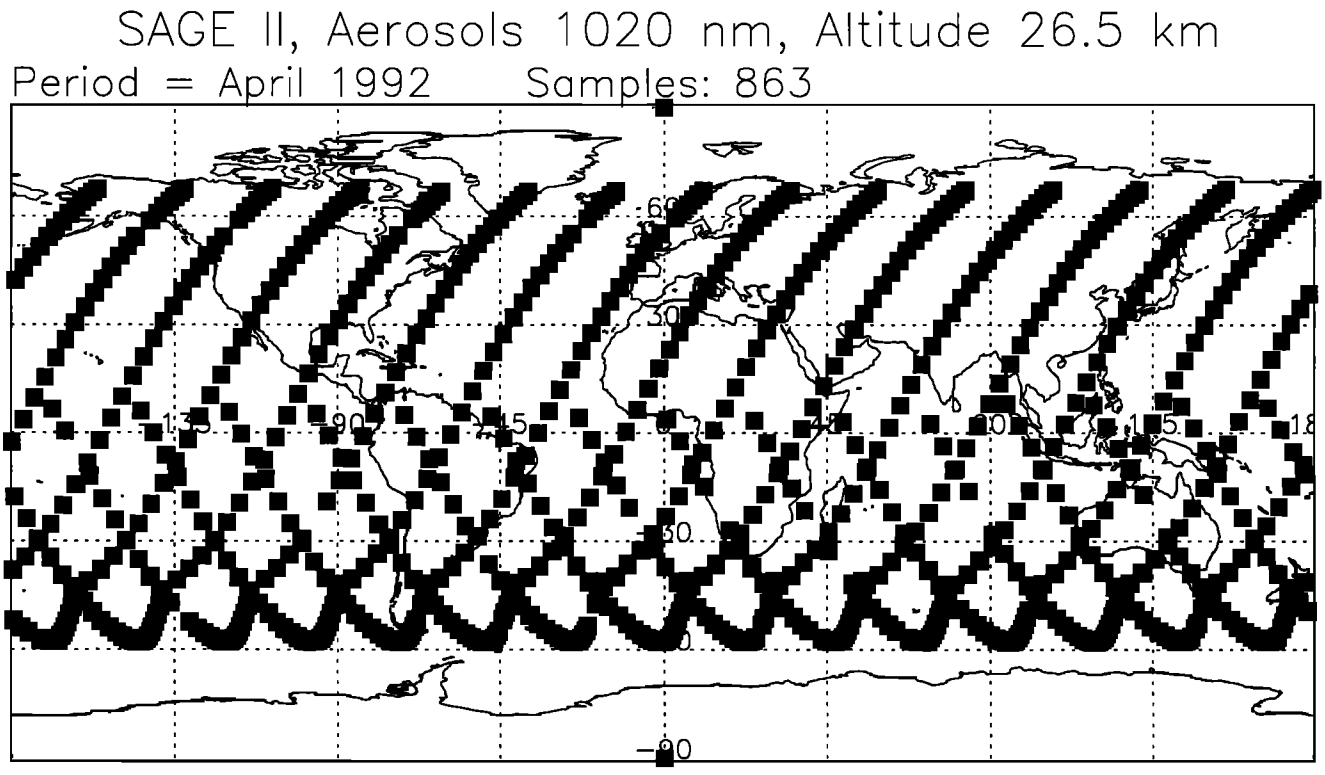

Figure 1. SAGE II observation points for April 1992.

The root mean square error between interpolated values $\tilde{f}_n$ and true values f_n is defined as

$$\varepsilon_{RMS} = \sqrt{\frac{\sum_{n=1}^{N} \left(\tilde{f}_n - f_n\right)^2}{\sum_{n=1}^{N} (f_n)^2}} \tag{32}$$

We applied this to a test function, being Gaussian in latitude and constant in longitude, sampled in 128 points, and found the following errors:

Maximum relative error	:	0.043
Mean relative error	:	0.016
Root mean square error	:	0.019

This consistency error was found to be of the same order as the interpolation error, described below. The value of the fitting parameter λ, introduced in (6), was found to have no major influence on the error values in this case.

Figure 1 shows the SAGE II, aerosol extinction coefficient, observations for April 1992, at an altitude of 26.5 km, and for the 1.02 μm wavelength channel. From this data we obtain the interpolated map shown in Figure 2, with 1° resolution in both directions. This map shows that aerosols are mainly found in an equatorial belt, which shows only little longitudinal variation. The SAGE II samples (measurements) in this example had the following errors:

Maximum relative error	:	0.989
Mean relative error	:	0.033
Root mean square error	:	0.081

Since the interpolation is a linear operation on the samples, it is easy to compute the resulting error on the interpolation values. We obtained the following:

Maximum relative error	:	0.958
Mean relative error	:	0.036
Root mean square error	:	0.070

We see that the mean relative error of the interpolation values, (3.6%), is of the same order as the mean relative errors on the given samples, (3.3%). This indicates that the interpolation algorithm does not significantly increases the uncertainty of the interpolated data.

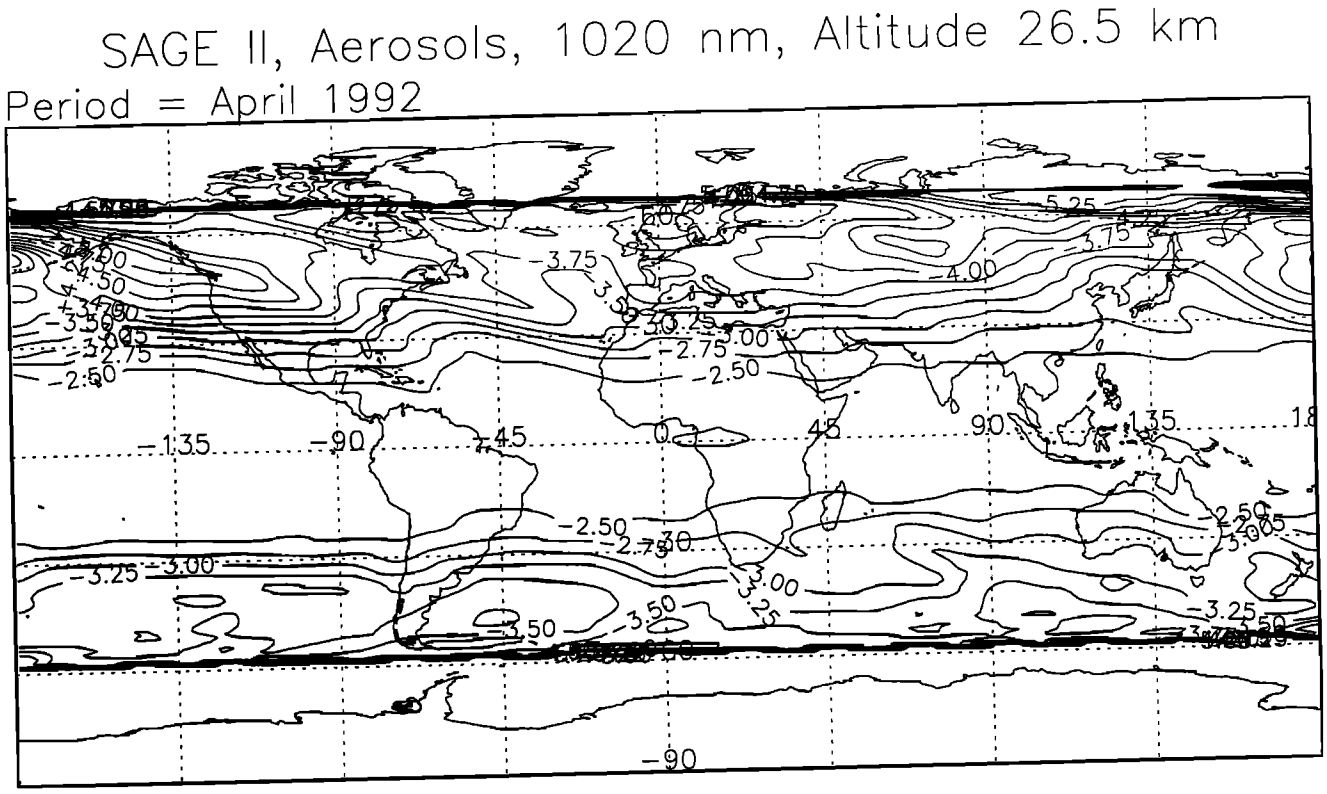

Figure 2. Contour map of aerosol extinction for April 1992. Plotted quantity is $\log_{10}(\beta)$.

For certain periods (e.g. February, August) and in equatorial regions, the SAGE II instrument, being a solar occultation limb scanning device, shows a lack of latitudinal coverage, due to the inclination of the satellite orbit. An example of this situation is given in Figure 3, showing the observation points for February 1992. The equatorial aerosol reservoir is not adequately sampled during this period. Applying the interpolation algorithm to this data produces the map shown in Figure 4. Despite the large under-sampled equatorial area, we still obtain a reasonable map. The algorithm extends the data trend, found at the edges of the data gap, and reproduces the expected aerosol reservoir at the tropics. This test shows that the new algorithm is capable of handling highly irregularly samplings and can fill in empty areas. However, we are aware that the resulting map in Figure 4 is less trustworthy in those regions where data points are missing. The here considered month, February 1992, is an example of a data set that needs extra information (e.g. from assimilation), to fill in the gaps, in order to be mapped properly. To aid in interpreting a direct map like Figure 4, one could introduce a flag to display only interpolated values within some given proximity to the sample points.

Finally, we show in Figures 5 and 6 another example of SAGE II measurement locations and interpolated aerosol map, respectively, for January 1989, at a pressure altitude of 50 mbar, and for the 1.02 μm wavelength channel. The map in Figure 6 is to be compared with the one shown in Figure 9, obtained by assimilation.

4.2. Assimilation

Some cases were selected to test the implementation of the aerosol model and the behavior of the assimilation process. The assimilation parameters were determined for a period with a low aerosol load for two assimilation periods, 10 days and 20 days, starting on January 10, 1989.

Figure 7 shows a comparison between model calculated optical extinction and observations, as a function of altitude for one particular event on January 22, 1989, 06:00 UTC at 47°N and 22°W. The SAGE II observations are represented by the symbols with the (horizontal) error estimates. The solid line shows the assimilation result at this position with all observations taken into account. The dashed line is the result when every

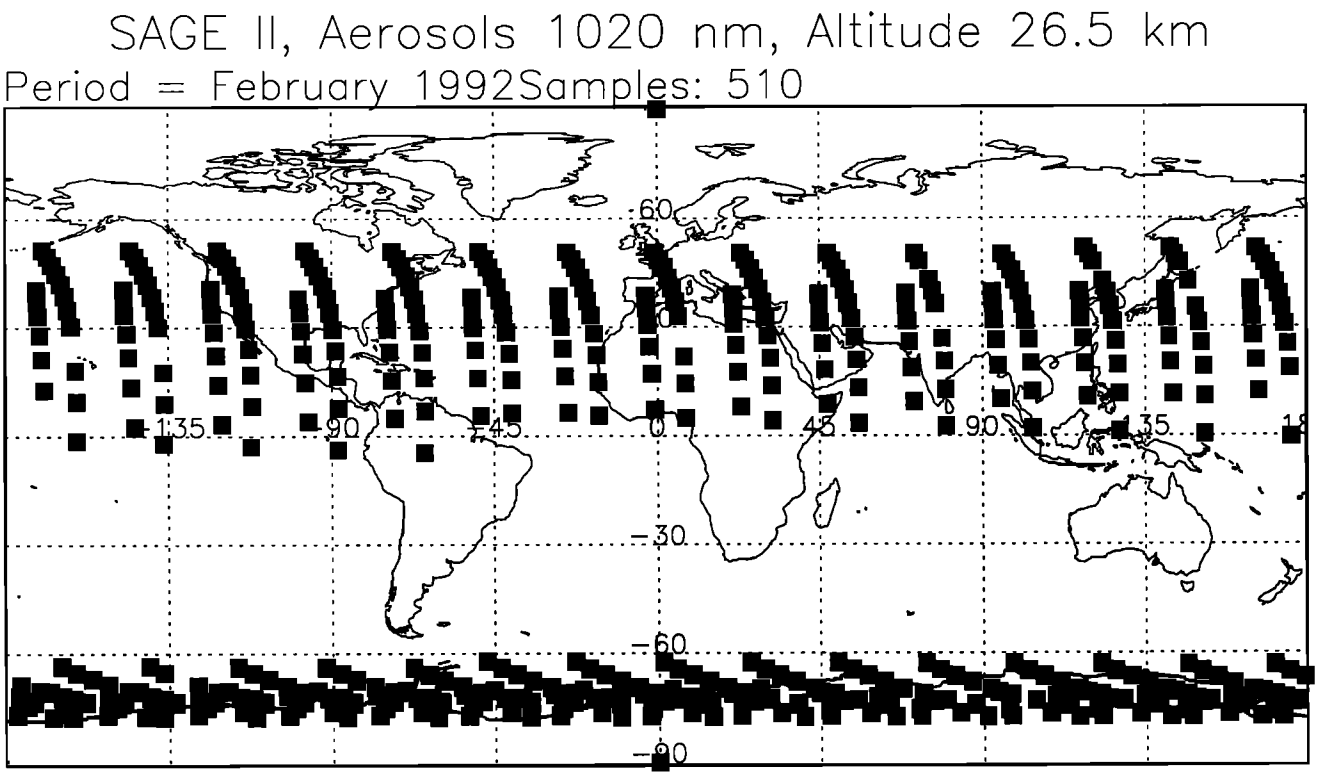

Figure 3. SAGE II observation points for February 1992, showing lack of coverage in the equatorial region.

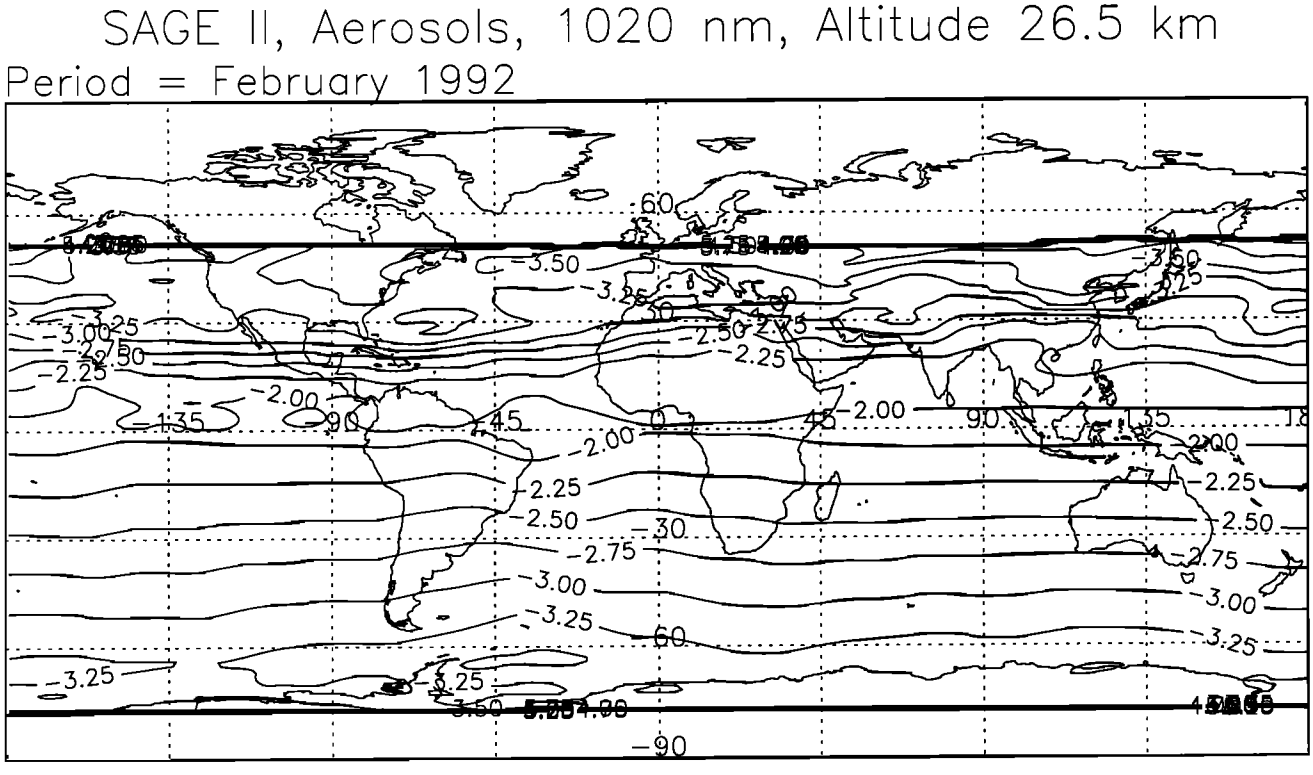

Figure 4. Contour map of aerosol extinction for February 1992, with interpolation over the equatorial region, based on the data trend at mid-latitudes. Plotted quantity is $\log_{10}(\beta)$.

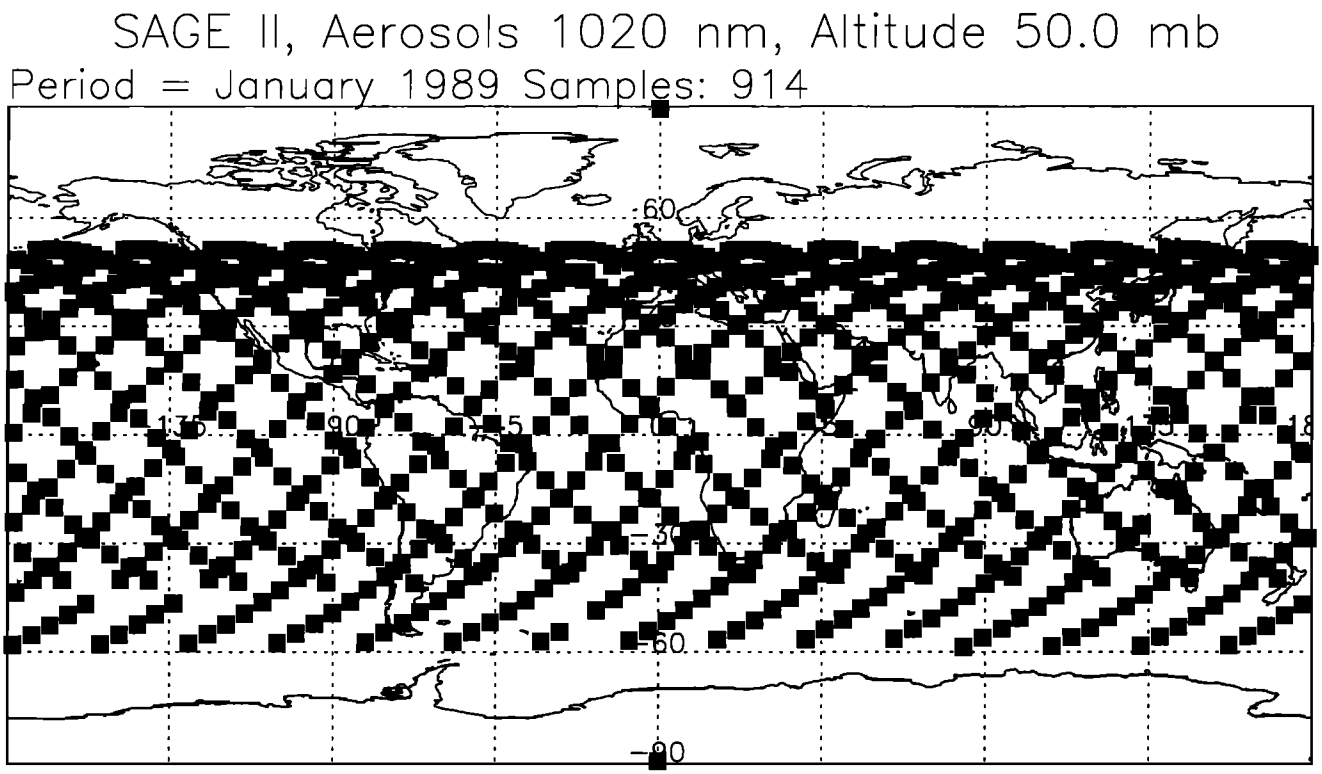

Figure 5. SAGE II observation points for January 1989, used for interpolation and assimilation.

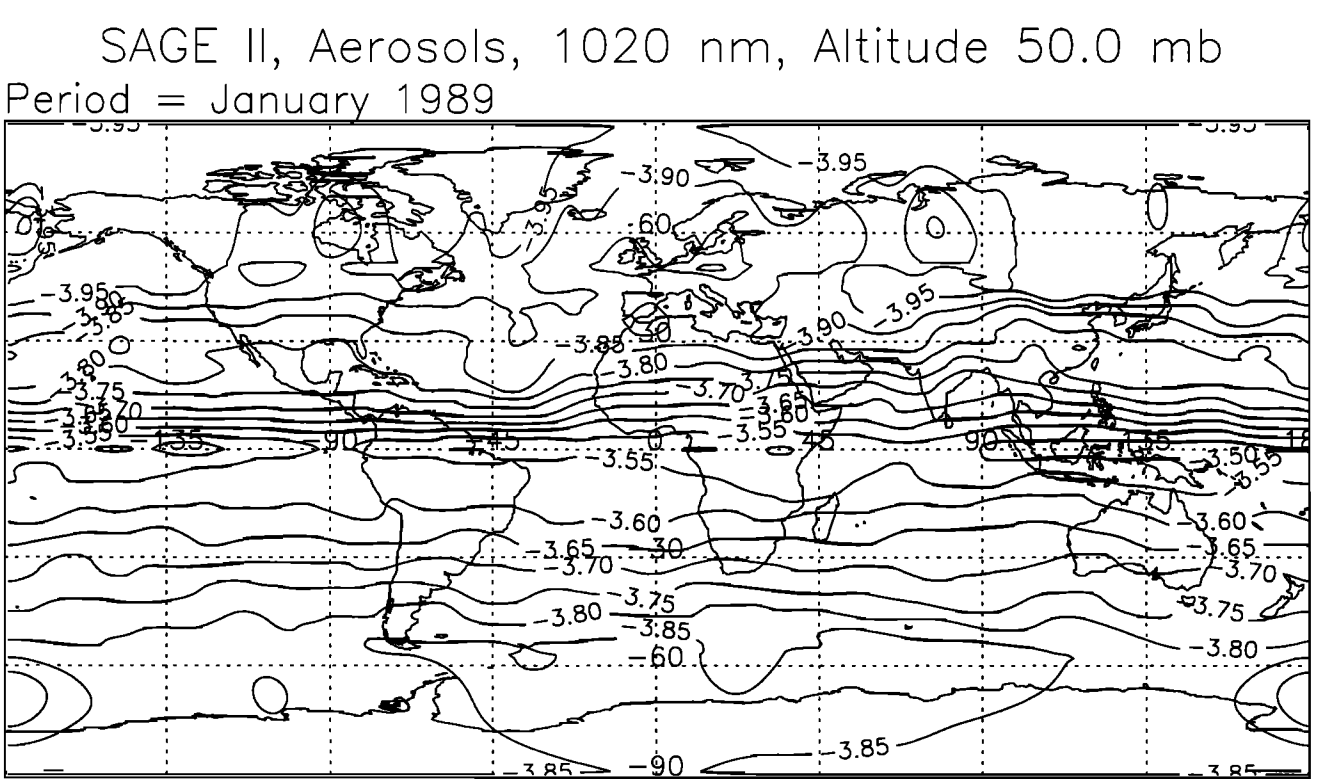

Figure 6. Interpolated aerosol extinction at 1.02 μm and 50 mbar, for January 1989. Plotted quantity is $\log_{10}(\beta)$.

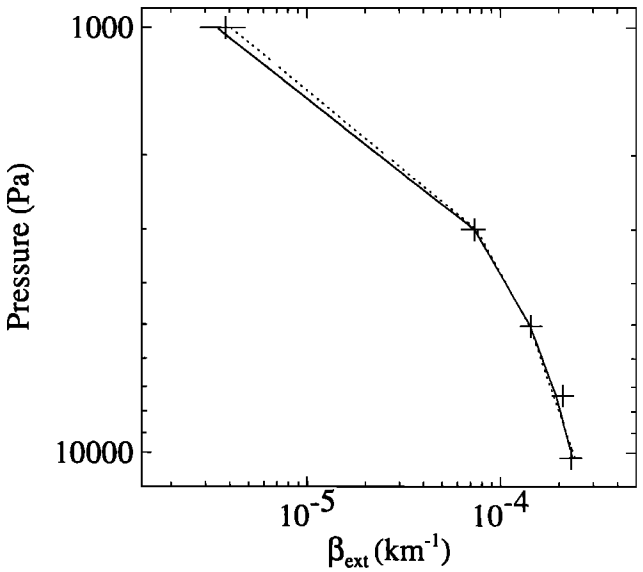

Figure 7. Comparison of model calculated extinction and observations as a function of altitude on January 22, 1989, 06:00 UTC at 47°N and 22°W (see text for details).

eighth observation for each hemisphere was not taken into account in the assimilation (the here considered position being one of them). Both cases are for the same period and use the same first guess. The model is able to reliably reproduce the observations left out, as is shown by the dashed line.

As already mentioned, the SAGE II data can be contaminated by thin cirrus clouds in the tropical regions. The aerosol model is not able to take into account these events when occurring. Several observations in the tropics are found difficult to reproduce. Figure 8 shows an example of an aerosol extinction profile, for the event on January 11, 1989, 06:00 UTC at 2.7°N and 28.1°E., where the assimilation is not able to reproduce the observations at pressure heights 70 and 100 mbar. We applied the statistical model of [*Kent et al.*, 1993] to these observations. The extinction ratio at 0.525 and 1.02 μm for these heights is 1.33 and 0.67, respectively. This indicates that the extinction must be attributed to a mixture of cirrus clouds and aerosols. In this way, assimilation could aid in the detection of non-aerosol events. On the other hand, the presence of cirrus cloud disturbs the accuracy of the assimilation results. This highlights the shortcomings of the simple aerosol model used. It should also be clear that the assimilation, as implemented here, will not be able to reproduce transient effects like a major volcanic eruption.

Unlike the Kalman filter technique, 4D Var assimilation does not provide an error estimate of the analysis. Due to the limited latitude coverage of the SAGE II data, the assimilation affects only a limited space region. With the aid of a second model aerosol variable, this limited region can be indicated. This artificial model aerosol variable is assimilated independently from the original model variable, but with a first guess increased by 50%. The difference between both model variables is indicative of the influenced regions by the observations. The regions, where the two model variables differ by 50%, are solely determined by the first guess of the assimilation.

Figure 9 shows the assimilated aerosol extinction for January 20, 1989, 00:00 UTC, at 1.02 μm and pressure altitude of 50 mbar. It is to be compared with Figure 6, showing a monthly mean for January 1989. Both figures use the same contour scale. It is found that the assimilation method can bring out details and features, which can not be resolved by interpolation alone.

5. CONCLUSIONS

5.1. Interpolation

The new algorithm has been applied to numerous tests, including sparse data situations, uniform background aerosol mode and high volcanic aerosol mode. The algorithm proved to be numerically very stable and robust. Tests were done with the number of sample points varying from a few tens up to 10,000 and highly

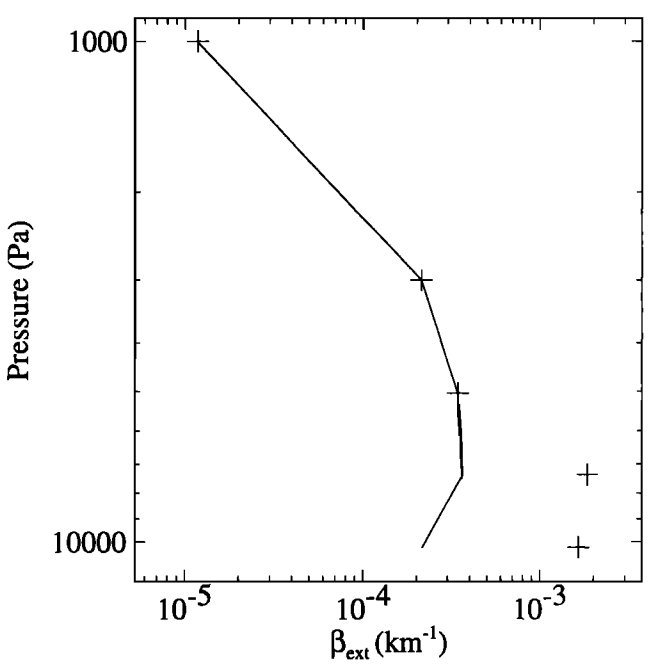

Figure 8. Comparison of model calculated extinction and observations as a function of altitude on January 11, 1989, 06:00 UTC at 3°N and 28°E (see text for details).

irregular samplings never caused the algorithm to fail. Also the local trend estimation, based on a local thin plate shape modeling of the variation of the data, never showed signs of ill-conditioning. As compared to assimilation, it is a very fast numerical method.

The limitations of a cartography, based on interpolation alone, are rather due to the lack of coverage of the data, than to the algorithm itself. Even in cases where no data is available over large areas of the globe, an estimate is produced by the algorithm in these areas, based on the trend of the data at the boundary of these areas. Data sets showing limited coverage however, are better not mapped by interpolation alone and require assimilation to more accurately fill in the gaps.

An obvious limitation of interpolating time sequential data, is the time skew error. The interpolation algorithm as it stands, does not take the distance-in-time into account of neighbor samples, when computing the thin plate fit.

5.2. *Assimilation*

4D Var assimilation has proven to be a valuable tool to reconstruct locally, missing spatial and temporal information of stratospheric aerosol density, from sufficiently close observation points. The presented implementation is sufficiently general that it allows easy extensions. One could use multi-wavelength information, as well as a more complex microphysics model for aerosols, including the assimilation of chemical species and their chemical transformations. This way the assimilation method would also be useful for mapping aerosols under non-normal conditions (e.g. volcanic eruptions).

The spatial validity of the assimilation is determined by the coverage of the measurements and the winds in the stratosphere, which back trace the influence. If the first guess initial condition is not the solution, the gradient of the cost function, with respect to the first guess, determines the domain that will be covered by the assimilation. It was found that the satellite coverage strongly limits the spatial domain in which the analyzed fields are influenced by the observations. In addition, one finds that little exchange takes place between the Northern Hemisphere, equatorial region, and the Southern Hemisphere, within the time window of the assimilation. This zonal separation reflects the dynamic behavior of the stratosphere. This means that assimilation is also limited to fill in equatorial gaps, unless the assimilation can be run over a sufficiently long time period. Two factors limit however the length of this assimilation period: computational resources and the physical model used for the aerosol evolution. Because of the limited spectral content of the available data, only a very simple physical aerosol model could be used. It was found that the temporal validity, obtained with the current model, is limited to less than 10 days when ran on a Cray parallel computer.

5.3. *Comparison of Both Methods*

Cartography by interpolation alone is fast and justified when the temporal variation of the species to be mapped is low over the considered time frame. In the case of the SAGE II data, it would be the preferred tool to produce a climatology or studies of long-term evolution (order of a season). It is an isotropic method and the dynamics of the atmosphere are not taken into account.

Assimilation enhanced cartography does takes the detailed dynamical evolution of the species into account. Due to present limitations in computer resources and in the physical model used, it is limited to short-time evolution (typical 10 days). Because wind dynamics in the stratosphere is largely anisotropic (transport takes place mainly parallel with the equator), also this method has difficulties filling in observational gaps in latitude. Both methods thus fail to produce a reliable map for sparse sampled periods (e.g. February). For data with a better observational coverage (e.g. January), assimilation produces a more detailed map and can bring out short-term temporal variations.

As cross checking between both methods, temporal mean values were calculated at each model grid point from the assimilation analysis. These mean values agreed very well with the interpolated field for the concerning period. This originates from the fact that the assimilated aerosol fields exhibit the dynamical variability superimposed on the interpolated fields.

When one has the necessary computational resources available, assimilation is without doubt the preferred cartography method.

NOTATION

$\mathbf{x} = (\theta, \varphi)$	An arbitrary point on the sphere (interpolation).
$\mathbf{x}_n$	The location of sample point n.
$\tilde{f}(\mathbf{x})$	Truncated MT series interpolant.
$f_n^{k,l}$	Partial derivatives of order k,l at sample point $\mathbf{x}_n$.
$U_n^{k,l}(\mathbf{x})$	General MT basis functions at sample point $\mathbf{x}_n$.
$U_n^{0,0}(\mathbf{x})$	Zero order MT basis function at sample point $\mathbf{x}_n$.

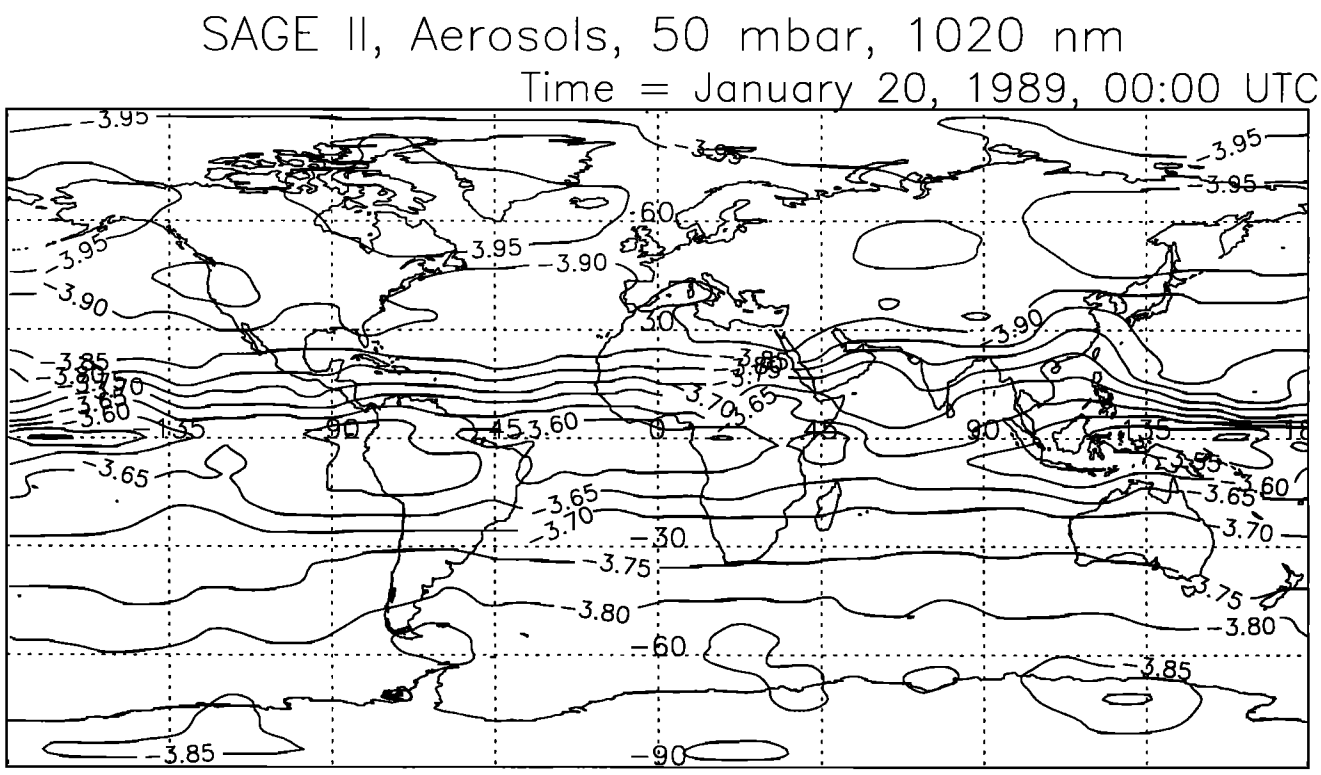

Figure 9. Assimilated aerosol extinction at 1.02 μm and 50 mbar, on January 20 1989, 00:00 UTC. Plotted quantity is $\log_{10}(\beta)$.

$s_n(\mathbf{x})$	Shape function in the neighborhood of sample point $\mathbf{x}_n$.
λ	Lagrangian fitting parameter.
f_x, f_y	Longitudinal and latitudinal spatial frequencies.
f_a, f_b	Longitudinal and latitudinal spatial filter cut-off frequencies.
$\mathbf{y}^o$	M-dimensional vector of observations of the physical system.
$\mathbf{y}$	M-dimensional vector of predictions computed by the model.
$\mathbf{x}$	N-dimensional unknown control (state) vector of the model (assimilation).
$(.,.)$	Scalar (inner) product defined on R^M.
$\mathbf{h}$	Mapping giving the dépendence of $\mathbf{y}$ on $\mathbf{x}$.
j	Cost function, a functional depending on $\mathbf{x}$.
$\mathbf{H}(\mathbf{x})$	Linearization of the mapping $\mathbf{h}$ around $\mathbf{x}$, (Jacobian).
$\mathbf{H}^*(\mathbf{x})$	Adjoint linearization of $\mathbf{h}$ around $\mathbf{x}$, with respect to the scalar product.
$\mathbf{x}_0$	Initial control (state) vector, (initial concentrations).
$\mathbf{x}_b$	Background control (state) vector.
$\mathbf{B}$	Covariance matrix of the expected errors in $\mathbf{x}_b$.
$\mathbf{y}_k^o$	Vector of all observations considered valid at the timestep k.
$\mathbf{y}_k$	Vector of model predictions for the $\mathbf{y}_k^o$.
$\mathbf{R}_k$	Covariance matrix of the expected errors in $\mathbf{y}_k - \mathbf{y}_k^o$.
$\mathbf{H}_k$	In our application, the linear map from $\mathbf{x}_k$ to $\mathbf{y}_k$ (observation operator).
$\mathbf{h}_{m-1}$	Discretized model operator mapping $\mathbf{x}_{m-1}$ into $\mathbf{x}_m$.
ε_{RMS}	Root mean square error.

Acknowledgments. This research was funded by the European Space Agency (ESA) in the framework of the Data User Programme (DUP), contract DAMS2P (Development of Global Aerosol Mapping from Satellites level-2 Products, contract no. 12160/96/I-HE). We recognize the SAGE II

team for the use of the SAGE II data, R. Giering for making TAMC available and J-C Gilbert from INRIA for the use of M1QN3.

REFERENCES

Chu, W., M. McCormick, J. Lenoble, C. Brogniez and P. Pruvost, SAGE II Inversion Algorithm, *J. Geophys. Res., 94*, 8339-8351, 1989.

Dierckx, P. *Curve and Surface Fitting with Splines*, Clarendon Press, Oxford, 1995.

Fisher, M. and D. Lary, Lagrangian four-dimensional variational data assimilation of chemical species, *Q. J. R. Meteorol. Soc, 121*, 1681-1704, 1995.

Franssens, G., A new non-polynomial univariate interpolation formula of Hermite type, *Adv. in Comp. Math., 10*, 367-388, 1999.

Franssens, G., Multivariate Hermite interpolation based on a Multipoint Taylor series representation, *in preparation*, 1999.

Giering, R., Tangent linear and adjoint biogeochemical models, this volume, 1999, American Geophysical Union.

Giering, R., *Tangent linear and Adjoint Model Compiler: User manual*, Center for Global Change Sciences, MIT, Cambridge, USA, 1997.

Gilbert, J.-C. and C. Lemaréchal, Some numerical experiments with variable storage quasi-Newton algorithms, *Math. Prog., B25*, 407-435, 1989.

Kent, G., D. Winker, M. Osborn and K. Skeens, A model for the separation of cloud and aerosol in SAGE II occultation data, *J. Geophys. Res., 98*, 20725-20735, 1993.

Renka, R. and A. Cline, A triangle based interpolation method, *Rocky Mountain Journal of Mathematics, 14*, 223-237, 1984.

Renka R., *Interpolation of data on the surface of a sphere*, Oak Ridge National Laboratory, Report ORNLICSD-108, 1982.

Smolarkiewicz, P. and P. Rasch, Monotone advection on the sphere: an eulerian versus semi-lagrangian approach, *J. Atm. Sci., 48*, 793-810, 1991.

Talagrand, O. and P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Q. J. R. Meteorol. Soc, 113*, 1311-1328, 1987.

Thomason, L., L. Poole and T. Deshler, A global climatology of stratospheric aerosol surface area density deduced from Stratospheric Aerosol and Gas Experiment II measurements: 1984-1994, *J. Geophys. Res., 102*, 8967-8976, 1997.

Ghislain Franssens, BISA - Belgian Institute for Space Aeronomy, Ringlaan 3, B-1180 Brussels,BELGIUM. (e-mail: Ghislain.Franssens@oma.be)

Dominique Fonteyn, BISA - Belgian Institute for Space Aeronomy, Ringlaan 3, B-1180 Brussels, BELGIUM. (e-mail: Dominique.Fonteyn@oma.be)

Martine De Mazière, BISA - Belgian Institute for Space Aeronomy, Ringlaan 3, B-1180 Brussels, BELGIUM. (e-mail: Martine.DeMaziere@oma.be)

Didier Fussen, BISA - Belgian Institute for Space Aeronomy, Ringlaan 3, B-1180 Brussels, BELGIUM.(e-mail: Didier.Fussen@oma.be)