Variational Atmospheric Retrieval Scheme (VARS)

for GPS Radio Occultation Data

Version 1.1
October 2005

COSMIC Project Office

University Corporation for Atmospheric Research

1. Overview of retrieval method

Measurements from GPS Radio Occultation (RO) technique (i.e. bending angle or refractive index) can be used to retrieve atmospheric parameters such as temperature, pressure (or altitude), and humidity profiles. One-dimensional Variational (1D-Var) method is an effective way to combine information provided by GPS RO and a given priori atmospheric state in a statistically optimal way. The method consists of finding the most probable atmospheric state (i.e. analysis) by minimizing a cost function (J)

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

As shown above the cost function is defined as the summed distance of the current state ($\mathbf{x}$) to the background state ($\mathbf{x}_b$) and to the observations ($\mathbf{y}_o$) in which the misfits are weighted by the inverse of error covariance matrices. The modeled observations ($\mathbf{y}$) are given by

$$\mathbf{y} = H(\mathbf{x}), \tag{2}$$

where H is the (forward) observation operator mapping model state to observation space. The covariance matrices of background error ($\mathbf{B}$) and observation error ($\mathbf{R}$) are assumed Gaussian. These errors are also assumed unbiased and uncorrelated each other. The gradient of J with respect to $\mathbf{x}$ is given by:

$$\nabla_{\mathbf{x}}J = \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\left(H(\mathbf{x}) - \mathbf{y}_o\right), \tag{3}$$

where $\mathbf{H} = \dfrac{\partial H}{\partial \mathbf{x}}$ is the tangent linear approximation of the (non-linear) observation operator H and $\mathbf{H}^{\mathbf{T}}$ is the adjoint operator of $\mathbf{H}$. Since the gradient of J tells us only the direction toward the minimum location of J at current state, the practical solution needs to be sought iteratively using an algorithm of gradient descent as:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{M}_k\nabla_{\mathbf{x}_k}J, \tag{4}$$

where the matrix $\mathbf{M}_k$ contains information on the optimal step size and the optimal descent direction at the number of iteration k provided by the descent algorithm. Descent algorithms differ how to estimate $\mathbf{M}_k$ and the limited memory algorithm for bound constrained optimization (L-BFGS-B) is used for the implementation of Variational Atmospheric Retrieval Scheme (VARS).

VARS also supports an alternative method, so called incremental formulation. In the incremental approach, the cost function is defined as

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^{\mathbf{T}}\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}\left(\mathbf{H}\delta\mathbf{x} - \mathbf{d}\right)^{\mathbf{T}}\mathbf{R}^{-1}\left(\mathbf{H}\delta\mathbf{x} - \mathbf{d}\right) \tag{5}$$

where $\delta\mathbf{x}$ is the increment, and the optimum value $\delta\mathbf{x}_a$ is added to the background $\mathbf{x}_b$ to provide the analysis $\mathbf{x}_a$.

$$\mathbf{x}_a = \mathbf{x}_b + \delta\mathbf{x}_a \qquad (6)$$

and $\mathbf{d}$ is the innovation vector

$$\mathbf{d} = \mathbf{y}_o - \mathbf{H}\mathbf{x}_b. \qquad (7)$$

The potential benefits of using incremental formulation are the reduction of computational cost due to the use of a simplified linear observation operator and the uniqueness of the minimum of J. In this formulation, however, the coefficients for the tangent linear model and its adjoint model are determined only by the background and thus strictly speaking the coefficients are not valid at an analysis state when a cycle of minimization is completed. Therefore new J needs to be minimized with updating the coefficients reflecting the analysis increment yielded from previous minimization, which is so called outer-loop. On the contrary the formulation which uses a full state to provide the coefficients (i.e. the method expressed by Eq. 1) is uninhibited but that can lead to the termination of minimization at a local value of multiple minima of J due to the non-linearity of observation operator. Both formulations produce a similar result, in most cases, if the quality of background is reasonable.

Meanwhile a fast convergence of minimization can be achieved by introducing new vector $\mathbf{v} = \sqrt{\mathbf{B}^{-1}}\delta\mathbf{x}$ and thus Eq. 5 can be expressed as

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{v} + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}) \qquad (8)$$

and the gradient of J is obtained differencing Eq. 8 with respect to $\mathbf{v}$

$$\nabla_{\mathbf{v}}J = \mathbf{v} + \sqrt{\mathbf{B}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}). \qquad (9)$$

The change of variable also can be applied to Eqs (1) and (3) but not presented here.

2. Observation operators

For neutral atmospheric condition, after the removal of ionospheric contribution, the atmospheric refractivity (N) can be expressed by

$$N = 10^6(n-1) = 77.6\frac{P}{T} + 3.73\times10^5\frac{e}{T^2}, \qquad (10)$$

where n is atmospheric refractive index, P is pressure in hPa, T is temperature in K, and e is water vapor pressure in hPa. Because of the local property of atmospheric refractivity (or refractive index), the computation of refractivity from temperature, pressure, and moisture is

straightforward.

In a spherically symmetric atmosphere, the bending angle of GPS signal (α) relates to refractive index via Abel integral transform

$$\alpha(a) = -2a \int_{a}^{\infty} \left(\frac{\partial \ln n(x)}{\partial x} \right) \frac{dx}{\sqrt{x^2 - a^2}} , \tag{11}$$

where a is the impact parameter or the refraction radius (i.e. $a = nr$). The singularity at the lower bound can be resolved by either a change of integral variable or a proper choice of numerical integration method. For the upper bound, the integration needs to be started from the altitudes of satellites (GPS or LEO) or a sufficiently high altitude (e.g. top of atmosphere) to the tangent point of ray. Since the interval of integration should be small enough to reduce the numerical error, the integration can be costly in computation. The incremental approach in which a lower-resolution grid is used for the linear operators can significantly reduce the computational cost.

For a given impact parameter, geometric altitude at the tangent point of the ray (z) can be obtained from the refractive index and the local radius of the Earth (R_E):

$$z = \frac{a}{n} - R_E . \tag{12}$$

Similarly refractive index can be obtained from the impact parameter and the local radius of the Earth once the altitude is known. Since the bending angle is known only for a given impact parameter, however, the ambiguity between refractive index and geometric altitude is inevitable. Consequently, the location of bending angle observation in the computational grid of retrieval scheme should be recalculated if altitude or pressure is used as vertical coordinate. The problem can be easily remedied by taking refraction radius as the vertical coordinate for analysis. After 1D-Var retrieval, the analyses of temperature, pressure, and moisture are used to construct the optimal refractive index. Thus the altitude of the tangent point in turn can be determined. For the sake of convenience, the altitude profile of the tangent points might be used to interpolate analysis onto a regular altitude levels.

3. Overview of retrieval procedure

a. Read in input data
The input data for VARS are a namelist, COSMIC/CDAAC level2 observation file, background profile, observation error, and background error. The namelist provides some parameters for the run-time configuration of VARS package such as the choice of 1D-Var

formulation, observation type, resolution of output grid, and operation mode. The CDAAC level2 product contains both bending angle profile and refractivity profile and user can specify which data will be treated as observation. The required variables in the background profile are temperature, water vapor pressure, (total) pressure, and altitude. The observation error file contains systematic bias as well as random component which are latitude-dependent and estimated from the statistics of innovation vector for a reasonably long period (e.g. 1-month). In order to yield optimal performance, the information should be updated regularly from the statistics for a recent period. The contents in the background error file are similar with those in the observation error file except for the type of variables.

b. Generate synthetic observations for testing purposes (optional)

The lack of reliable, collocated, and independent observations is a practical obstacle to validate VARS and so the use of simulated observation is an easy way to test the retrieval procedure. For this purpose, synthetic observations are generated by applying observation operator to the first guess profile. Here we assume that the original input profile, which could have been used as background, perfectly describes the true atmospheric condition. Consequently the synthetic observation is the perfect (error free) observation. In this case the real CDAAC RO data are not used in the retrieval procedure except the information on the location of observations. We then generate a new background profile by adding unbiased random noises to the original input profile. The rest of retrieval procedure is the same with the real data case. When applied to an ensemble of soundings, this approach allows us to roughly estimate the effectiveness of the retrieval method. We also find this testing is useful for debugging and/or development of the retrieval code.

c. Set-up computational grid

If bending angle is taken as observation, impact parameter will be used as the vertical coordinate of the computational (intermediate) grid. Otherwise, altitude will be used as the vertical coordinate. Additionally, different computational grids can be used for bending angle observations with the incremental formulation: a higher resolution for the calculations of innovation vector and reference state and a lower resolution for minimization. As the computational cost is not a critical issue in the 1D-Var problems, the default setting is to use dense grids (e.g. 100 – 200 m) everywhere. This step is also involved with many interpolations to place information on the computational grid. First, the background profile is interpolated and then the pressure on the computational grid is adjusted to be hydrostatic. This is convenient to impose hydrostatic constraint on the increment. The hydrostatic equation is integrated both upward and downward starting from the level corresponding to 500 hPa. Moreover the profile of the standard

deviation of background error at the mean latitude of RO is evaluated and then it is vertically interpolated to the computational grid. The covariance matrix of background error is modeled with the interpolated standard deviation and a fifth-order correlation function, which is similar to a Gaussian function in shape but compactly supported, that is, the correlation decreased to zero at a finite distance. A component of the covariance matrix can be expressed by

$$b_{ij} = \sigma_i \sigma_j r_{ij} \tag{13}$$

where σ_i and σ_j are the standard deviations at two levels i and j, and r_{ij} and b_{ij} are correlation and covariance between the levels, respectively. By providing a different length scale for the error correlation, the shape of r_{ij} can be easily modified. The moisture correlation between stratosphere and troposphere is suppressed to prevent unrealistic moistening at stratospheric levels or drying at tropospheric levels. A series of matrix operation is performed to deduce the square root of the covariance matrix $\sqrt{\mathbf{B}}$, which is required for preconditioning. $\sqrt{\mathbf{B}}$ is the symmetric matrix which has the same eigenvectors as $\mathbf{B}$ and which has eigenvalues the square root of those of $\mathbf{B}$.

d. Calculate innovation vector

Applying the observation operator to the background and comparing with observations, the innovation vector $\mathbf{d}$ is prepared. $\mathbf{d}$ belongs to the observation space, which means the vector has the same size with the observation vector $\mathbf{y}_o$ and defined at every level of $\mathbf{y}_o$. Furthermore, the coefficients needed for the tangent linear and the adjoint computations are calculated and stored.

e. Perform a quality check

Prior to a quality check, a latitude-dependent estimate of background error, systematic bias as well as standard deviation, are interpolated to the each level of observations. These errors are stored in the derived data type which has been used for the innovation vector. Some portion of the estimated bias is subtracted from the innovation vector to remove potential observational bias such as the negative refractivity bias at a lower altitude levels. While the bias corrected innovation vector is used for the quality check and retrieval, the uncorrected vector needs to be collected to update the observation errors for operational use. Similar with the observation error, the profile of a combined error at the mean latitude of RO data is evaluated and then it is vertically interpolated to the locations of observations. The combined error (σ_c) includes both observation and background errors (i.e. $\sigma_c = \sqrt{\sigma_o^2 + \sigma_b^2}$) and that represents the expected size of the components of the innovation vector. Wherever the component of the bias corrected $\mathbf{d}$ exceeds a specified multiple of σ_c, the datum is assumed suspicious in quality and hence it is discarded.

f. Minimization of the cost function

Since the preconditioning is conceptually a normalization of state vector with the square root of covariance matrix of background error, the control variables are limited not to exceed a specified value (default is 3.0) in size. This helps the optimization algorithm to find a physically realistic solution. For moisture, additional constraints, positive and sub-saturated in total amount (i.e. background plus increment), are imposed.

(1) Compute the increment in physical space from the control variables: $\delta\mathbf{x} = \sqrt{\mathbf{B}}\mathbf{v}$. At the beginning of iteration, $\mathbf{v} = 0$.

(2) Perform the tangent linear operator and store $\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})$.

(3) Evaluate the cost function: $J = \dfrac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{v} + \dfrac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})$.

(4) Perform the adjoint operator by adding the forcing terms and evaluate the gradient of the observational cost function in the physical space: $\nabla_{\mathbf{x}}J_o = \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})$.

(5) Perform the adjoint of the change of variable: $\nabla_{\mathbf{v}}J_o = \sqrt{\mathbf{B}}\nabla_{\mathbf{x}}J_o$

(6) Evaluate the gradient: $\nabla J = \mathbf{v} + \nabla_{\mathbf{v}}J_o$.

(7) Test for convergence. If not converged, go to step (2) and repeat.

(8) To apply additional outer-loop, update the innovation vector and the coefficients for the tangent linear and adjoint calculations with the resulting analysis. Go to step (1) and repeat the whole cycle of iteration.

g. Generate output files

If bending angle is used for the retrieval, the analysis is transformed from the impact parameter space to altitude levels. On completion of the retrieval, an output file, 'Run/anal.out', is created.

4. Namelist variables

mtop: [integer] Top of output grid. Unit is geometric altitude in meter.

msfc: [integer] Bottom of output grid. Unit is geometric altitude in meter.

mdz: [integer] Interval of output grid. Unit is geometric altitude in meter.

mtop_q: [integer] Top of moisture analysis. Unit is geometric altitude in meter and above the altitude moisture increment is set to zero.

mode_run: [integer] Operation mode.

        1 = production

        2 = testing with synthetic observation

        -1 = test for the tangent linear part

        -2 = test for the adjoint part

incremental: [logical] If set to .TRUE., the incremental formulation will be used

otype: [integer] Type of observation operator

        1 = refractivity

        2 = bending angle

fixed_poe: [real]: Percentage error of observation

remove_ob_bias: [logical] If set to .TRUE., observational bias will be removed prior to 1D-Var retrieval

frm_ob_bias: [real] Fraction of observational bias to be removed

amp_fac_be: [real] Amplification factor for background error. Default value is 1.0 which is for NCEP/AVN 12-h forecast. A smaller value might be used for better background profiles

dsmax: [real] A threshold for quality check. Any observation deviates more than the multiple of background error from background is discarded

itmax: [integer] Maximum number of iterations allowed for an inner-loop

nloop: [integer] Number of outer-loops

kind_be: [integer] Type of covariance matrix of background error

        1 = diagonal

        2 = includes correlation

hydro_bg: [logical] If set to .TRUE., background pressure will be reconstructed via hydrostatic equation

ihydro: [integer] Direction of hydrostatic integration. Upward (+1) or downward (-1).

sl_of_t: [real] Scale-length of background temperature error correlation in meter.

sl_of_p: [real] Scale-length of background pressure error correlation in meter.

sl_of_q: [real] Scale-length of background moisture error correlation in meter.