Special Topics

3DVAR Analysis/Retrieval of 3D water vapor from GPS slant water data

Liu, H. and M. Xue, 2006: 3DVAR retrieval of 3D moisture field from slant-path water vapor observations of a high-resolution hypothetical GPS Network. Mon. Wea. Rev., 134, 933-949.

Main results of the paper:

- A 3DVAR method is developed to retrieve three-dimensional moisture field of atmosphere from a ground-based GPS slant-path water vapor (SWV) observation network.

- The inclusion of an analysis background makes the retrieval feasible. An explicit Gaussian spatial filter is used to model the background error covariances.

- Anisotropic spatial filter that is based on flow-dependent background error structures is implemented and tested.

- The use of flow-dependent background error covariance modeled by the anisotropic spatial filter improves the moisture retrieval.

- Surface moisture observations are important for the analysis near ground, especially when flow-dependent background error covariances are not used.

- Vertical filtering is necessary for obtaining accurate analysis increment at the low levels. The role of flow-dependent background error covariance is more prominent when the density of ground-based GPS receiver stations decreases.

Background of GPS data

- The microwave radio signals transmitted by GPS satellites are delayed by the atmosphere as they propagate to the ground-based GPS receivers.

- The total delay along the slant path is composed of three parts: ionospheric delay, hydrostatic delay and wet delay.

- Ionospheric delay observed by a dual-frequency GPS receiver can be calculated to millimeter accuracy.

- The hydrostatic delay can be estimated with known knowledge of pressure and temperature.

- So the slant-path wet delay (SWD) is obtained by subtracting the ionospheric and hydrostatic delays from the total delay.

- Then zenith wet delay (ZWD) can be obtained by projecting SWD observations onto the zenith and averaging them over a certain time period.

- Further, the precipitable water (PW, defined as the vertically integrated water vapor) can be calculated from ZWD.

- In recent years, researchers have demonstrated that assimilating observations of ZWD and PW data into mesoscale numerical models provides beneficial impacts on short-range precipitation forecast of convective weather.

- ZWD and PW data are products derived from the original total delay along the slant paths of GPS observations. The derivation involves additional assumptions.

- With variational method available as a powerful analysis tool, it is better to directly use the GPS slant-path observations to obtain a three-dimensional moisture distribution.

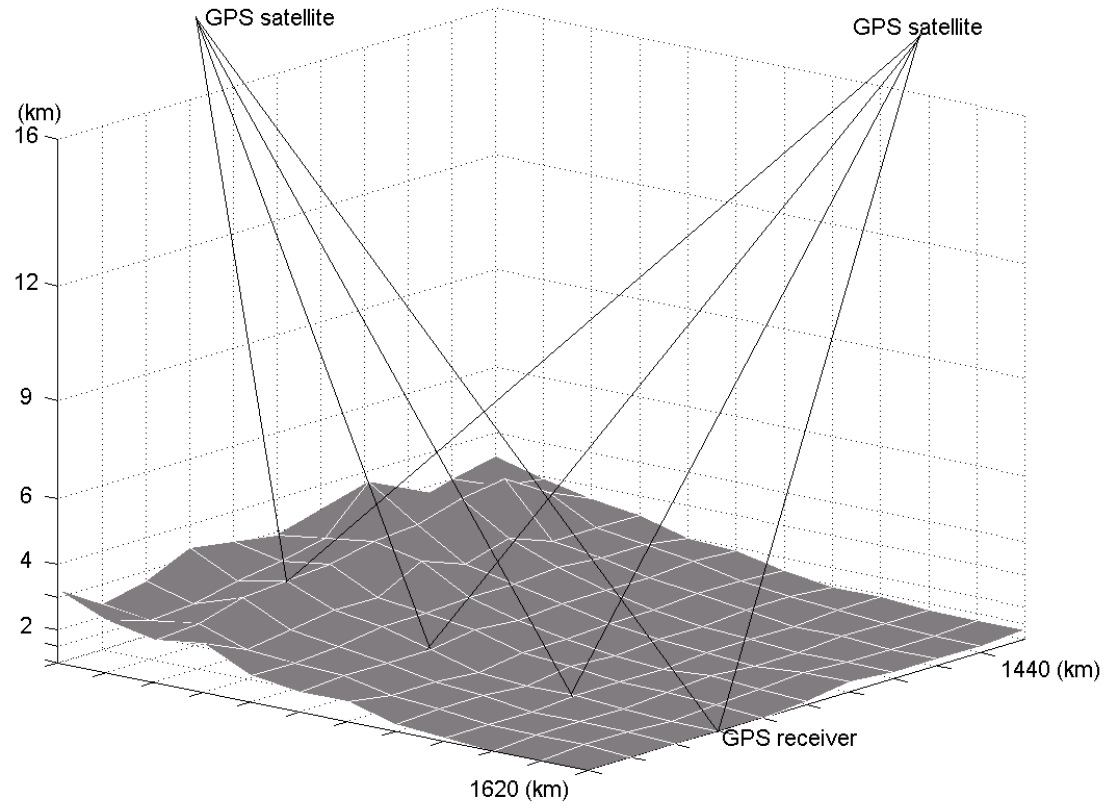
- Therefore, the slant-path water vapor observation is the focus of this paper.

- SWD is nearly proportional to the quantity of water vapor integrated along the slant path, the relationship between the SWD and SWV is

$$SWV = \Pi \bullet SWD,$$

  where SWV and SWD are given in units of length, and Π is a dimensionless constant, a function of weighted mean temperature of the atmosphere.

- SWV can provide vertical structure information of atmospheric moisture distribution through intercepting paths.



A schematic of hypothetic ground-based GPS observation network whose data are analyzed using 3DVAR. Shaded surface represents terrain. Dark solid lines are slant paths between ground-based GPS receivers and GPS satellites. Dotted lines give a sense of the vertically stretched grid although the actual grid levels are in terrain-following coordinate.

3DVAR Retrieval Method

The retrieval method used in this paper is based on 3DVAR method (Lorenc 1981; Daley 1991) of data assimilation which minimizes the following cost function,

$$
\begin{aligned}
J(x) &= J_b(x) + J_{swv}(x) + J_{sfc}(x) + J_c(x) \\
&= \frac{1}{2}(x - x_b)^T \mathbf{B}^{-1}(x - x_b) + \frac{1}{2}\left[H_{swv}(x) - SWV\right]^T \mathbf{R}_{swv}{}^{-1}\left[H_{swv}(x) - SWV\right]. \qquad (2) \\
&+ \frac{1}{2}\left[H_{sfc}(x) - q_{v_{sfc}}\right]^T \mathbf{R}_{sfc}{}^{-1}\left[H_{sfc}(x) - q_{v_{sfc}}\right] + \frac{1}{2}\left(\frac{|x| - x}{2}\right)^2
\end{aligned}
$$

- J is composed of four terms: background constraint term, GPS SWV observation term, the term for conventional surface moisture observations and the non-negative weak constraint. Standard notations are used here.

- The second term, J_{swv}, represents the departure of the analysis, calculated from control variable q_v through the observation operator H_{swv}, from the observations SWV that is measured by the ground-based GPS network.

- The matrix $\mathbf{R}_{\text{swv}}$ is the observation error covariance matrix, which is often simplified to be diagonal under the assumption that observation errors are not correlated.

- The magnitude of variances, the diagonal elements of matrix $\mathbf{R}_{\text{swv}}$, compared to the background error variances, determines the relative weight of observation and background for the analysis.

- Observation error variances for SWV and surface observations are specified

- J_{sfc}, is added in the cost function to better recover the moisture structure near surface by analyzing surface station measurements of water vapor

- In order to avoid the negative water vapor at high levels in the minimization process, a non-negative-moisture weak constraint term, J_c, is also included in this cost function.

- The inclusion of background term is significant for the 3DVAR analysis.

- It not only can eliminate the under-determined problem associated with the number of control variables exceeding the number of observations, but also provide more accurate analysis through the background error covariance matrix.

VAN: Variational analysis with no inversion of B

- Since **B** is very large for typical meteorological problems, its direct inversion as required by (2) is therefore never attempted.

- Huang (2000) presents a method named variational analysis using a filter (VAF), in which the control variable is redefined as,

$$v = \mathbf{B}^{-1} (x - x_b). \qquad (3)$$

- It is not the full analysis field itself but the increment field relative to the background, multiplied by the inverse of **B**.

The cost function is therefore redefined as,

$$
\begin{aligned}
J(v) &= J_b(v) + J_{swv}(v) + J_{sfc}(v) + J_c(v) \\
&= \frac{1}{2} v^{\mathrm{T}} \mathbf{B}^{\mathrm{T}} v + \frac{1}{2}\left[H_{swv}\left(\mathbf{B}v + x^b\right) - SWV \right]^{\mathrm{T}} \mathbf{R}_{swv}{}^{-1} \left[H_{swv}\left(\mathbf{B}v + x^b\right) - SWV \right] \\
&+ \frac{1}{2}\left[H_{sfc}\left(\mathbf{B}v + x^b\right) - q_{v_{sfc}} \right]^{T} \mathbf{R}_{sfc}{}^{-1} \left[H_{sfc}\left(\mathbf{B}v + x^b\right) - q_{v_{sfc}} \right] \qquad (4) \\
&+ \frac{1}{2}\left(\frac{|\mathbf{B}v + x^b| - \left(\mathbf{B}v + x^b\right)}{2} \right)^2 .
\end{aligned}
$$

- This modification avoids the inversion of **B** in the new cost function.

- To start the minimization iterations, the first guess of $v^0 = 0$ because $x^0 = x_b$. Therefore $\mathbf{B}^{-1}$ is not needed in specifying the first guess of v either.

- Moreover, VAF method uses a spatial filter to model the effect of **B** matrix instead of calculating and storing the matrix.

- The new variational analysis scheme is simpler and more flexible in practical implementations.

$$
\begin{aligned}
\mathbf{v}^0 &= 0, \\
\nabla_{\mathbf{v}} J &= \mathbf{B}^{\mathrm{T}}\{\mathbf{v} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}[H(\mathbf{B}\mathbf{v} + \mathbf{x}^b) - \mathbf{y}]\}, \\
\mathbf{v}^{n+1} &= \mathbf{v}^n - \alpha \nabla_{\mathbf{v}} J, \quad \text{and} \\
\mathbf{x}^{\mathrm{VAN}} &= \mathbf{x}^b + \mathbf{B}\mathbf{v}^{\infty}. \qquad (5)
\end{aligned}
$$

VAF: Variational analysis using a filter

- Matrix B ($N \times N \sim 1014$) is too large to be stored in memory by present-day computers and therefore has to be calculated every time it is needed.
- In VAN, B is required during each iteration loop. This makes an operational implementation of VAN difficult.
- However, a close inspection of VAN, in particular of the gradient of the cost function $\nabla_v J$, reveals that B may be modeled by a spatial filter.
- The relation between the multiplication of B and an application of a filter was discussed by Lorenc (1992) .
- Linear transforms are necessary to construct a filter from B when a non-uniform grid is used. In Huang (2000), a uniform grid is used for simplicity.

Multiplying a vector **x** by the **B** matrix, the result vector **x*** has its elements

$$x_i^* = \sum_{j=1}^{N} b_{ij} x_j, \qquad (6)$$

where b_{ij} are the elements of **B** ($i = 1, N; j = 1, N$).

In the case of a uniform grid, (6) could be considered as filtering **x** with a spatial filter with the filter coefficients b_{ij} and the filter span covering the whole model domain.

Therefore the VAN scheme can be written by using a filter to model **B**. This new scheme is called VAF, variational analysis using a filter, and can be summarized as

$$\begin{aligned} \mathbf{v}^0 &= 0, \\ \nabla_{\mathbf{v}} J &= \mathcal{G}\{\mathbf{v} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}[H(\mathcal{G}\mathbf{v} + \mathbf{x}^b) - \mathbf{y}]\}, \\ \mathbf{v}^{n+1} &= \mathbf{v}^n - \alpha \nabla_{\mathbf{v}} J, \quad \text{and} \\ \mathbf{x}^{\mathrm{VAF}} &= \mathbf{x}^b + \mathcal{G}\mathbf{v}^{\infty}, \end{aligned} \qquad (7)$$

where $\mathcal{G}$ is a spatial filter that should be designed based on the a priori knowledge of the covariance matrix **B.**

For example, consider a two-dimensional univariate problem with homogeneous and isotropic background error correlations. The following Gaussian function could be used to calculate **B** (Daley 1991)

$$b_{ij} = \sigma_b^2 \exp\left[-\left(\frac{r_{ij}}{L}\right)^2\right], \qquad (8)$$

where σ_b is the standard deviation of the background error, r_{ij} is the distance between model grid points i and j, and L is a length scale that can be determined theoretically or by observations.

- If the B matrix is constructed by (8), then the multiplication in (6) is equivalent to applying a Gaussian filter to the vector.

- The motivation for using a filter to model B is to reduce the memory consumption of storing B and the computational burden caused by multiplying B.

- Lorenc (1992) and Hayden and Purser (1995) have demonstrated that a multiple iteration recursive filter (no extra memory is required) asymptotically approaches to a Gaussian filter as the iteration tends to infinity.

- In this paper, a truncated Gaussian filter is used. Although some extra memory is needed to store the filter coefficients, this approach is more flexible.

- Other forms of B (Daley 1991) may easily be used to determine the filter coefficients.

- By using a truncated Gaussian filter, we in fact set many elements in B to zero. There is no guarantee of positiveness of the modified B. An additional filter windowing is necessary for the convergence of iterations.

Isotropic Gaussian filter

- The choice of spatial filter should be based on a priori knowledge of the covariance matrix **B**.
- For instance, the following Gaussian filter function can be used to represent **B** for homogeneous and isotropic background error field (Daley 1991) for a three-dimensional univariate problem,

$$b_{ij} = \sigma_b^2 \exp\left[-\left(\frac{r_{ij}}{L_r} \right)^2 \right], \qquad (5)$$

- σ_b^2 is the variance of background error,
- r_{ij} is the spatial distance between grid point *i* and grid point *j*,
- L_r is the de-correlation length scale decided by the background error correlation and is in practical use sometimes tied to the observation station density.
- The above model represents an isotropic background error covariance.

Flow-dependent anisotropic filter

- The use of isotropic background error covariance is based on the assumption that background errors at nearby points are similar (Riishojgaard 1998).
- But the background error covariance should be flow-dependent and such covariance should improve the analysis, especially when data are sparse.
- An anisotropic filter is therefore considered to model the flow-dependent **B** matrix.
- Simply, the following expression can provide such an anisotropic filter,

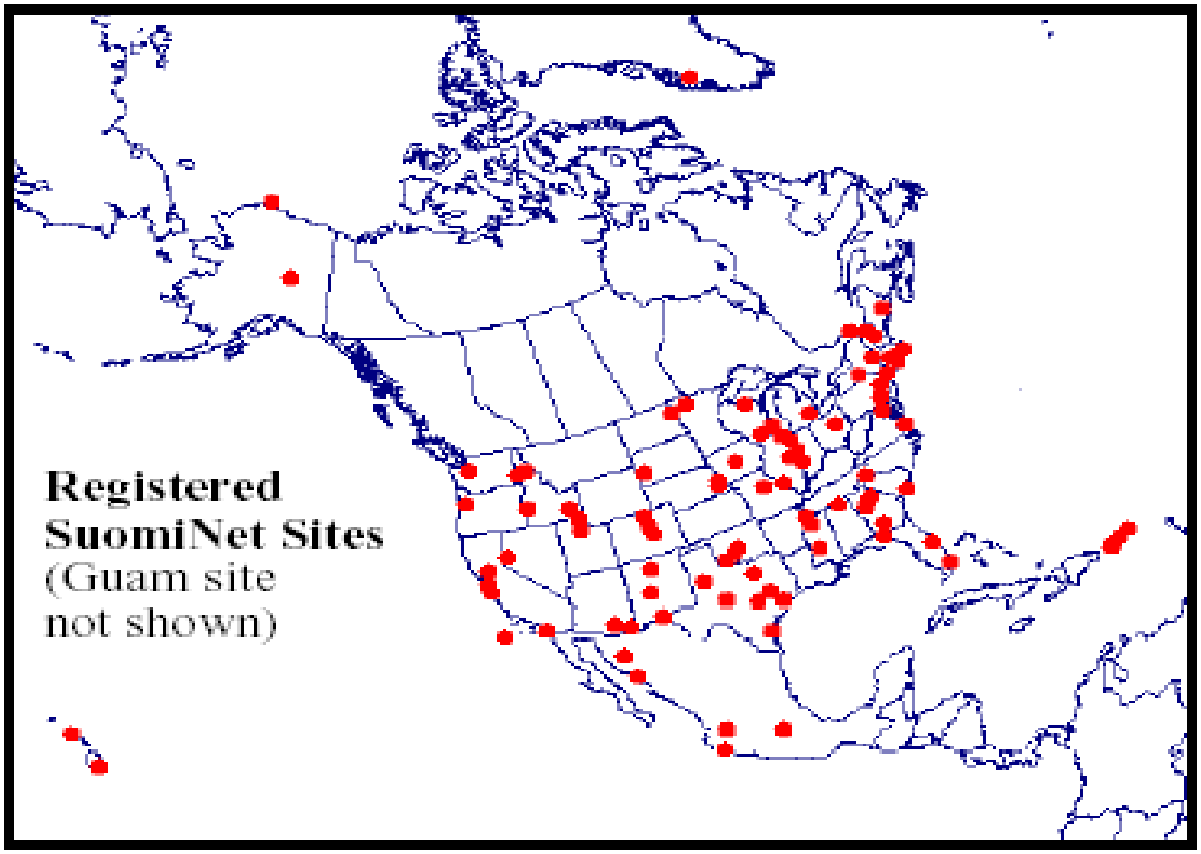
$$b_{ij} = \sigma_b^2 \exp\left[-\left(\frac{r_{ij}}{L_r}\right)^2\right] \exp\left[-\left(\frac{f_i - f_j}{L_f}\right)^2\right] \qquad (6)$$

- f is a field whose pattern represents that of the background error and we will call it the error field.
- In this study, f is either the true error of the background or an estimate of it.
- L_f is the length scale in error field space, compared to the length scale L_r in physical space, and is decided by the correlation of background error.
- The new background error covariance between two points defined by the Eq. (6) will follow the shape of the error field and fall off rapidly in the directions where the error field gradient is strong while isotropic covariance will dominate in directions where the error field changes slowly.
- Eq. (6) shows that in the case L_f goes to infinity, the anisotropic covariance reduces to the isotropic form in Eq. (5).

Observing System Simulation Experiment (OSSE)

Currently, high-resolution GPS observation network with large spatial coverage does not exist in the United States. Numerical experiments can be done using simulated data, hence Observing System Simulation Experiments (OSSE).

There does exist a not so dense GPS network, called the SuomiNet:



The 'nature' run or Truth

- High-resolution observations from hypothetical GPS networks are created from ARPS forecast fields for a dryline case that occurred on June19, 2002 over the Southern Great Plains.

- The ARPS model is initialized using analysis of the ARPS Data Analysis System at 1200 UTC June 19, 2002, and integrated for 8 hours.

- The computational domain is over the Southern Great Plains with 9 km grid spacing and 43 layers in the vertical.

- Stretched vertical grid coordinate is used with a minimum vertical grid spacing of 100 meters in boundary layer.

- The 9-km 8-hour forecast field is thinned by sampling specific humidity every 4 grid points, yielding a resolution of 36 km and a horizontal grid size of 46×41.

- This gridded field is defined as the ‘nature’ and used to generate the hypothetical GPS slant water vapor observation data.

- The specific humidity field from the ‘nature’ is presented in Fig. 1.

Simulation of GPS SW Data

For our OSS experiments, the slant-path water vapor is obtained by formula,

$$SWV_{ij} = \int_{i^{th}\,receiver}^{j^{th}\,satellite} q_v \, ds, \qquad (7)$$

- ds is the length of elements along slant path
- SWV_{ij} is the integrated water vapor along the slant path between the i^{th} ground-based GPS receiver and the j^{th} GPS satellite,
- q_v is the specific humidity along the path elements.
- The hypothetical GPS network is composed of nine irregularly distributed satellites simultaneously in view, and of 132 ground-based receivers which are evenly distributed in the analysis domain.
- The horizontal resolution of GPS receivers is 144 km.
- Both sampling and analysis grids are on the ARPS terrain-following coordinate.
- A schematic is given in Fig. 2 to illustrate the hypothetical GPS observation network.
- Surface moisture observations are available at GPS receiver sites.

Table 1. List of moisture retrieval experiments

Experiment	background	Flow-dependent B	Obs, obs error	Obs resolution	Filter	Correlation coefficient
CNTL	smoothed truth	Yes, on true background error	SWV+sfc, no	1 ob/4 grid intervals	3D	0.926
SNF	smoothed truth	No	SWV+sfc, no	1 ob/4 grid intervals	3D	0.830
SUF	smoothed truth	Yes, on updated background	SWV+sfc, no	1 ob/4 grid intervals	3D	0.832
LTF	Logarithmic	Yes, on true background error	SWV+sfc, no	1 ob/4 grid intervals	3D	0.827
LNF	logarithmic	No	SWV+sfc, no	1 ob/4 grid intervals	3D	0.821
STFNSFC	smoothed truth	Yes, on true background error	SWV, no	1 ob/4 grid intervals	3D	0.894
SNFNSFC	smoothed truth	No	SWV, no	1 ob/4 grid intervals	3D	0.668
STFNVF	smoothed truth	Yes, on true background error	SWV+sfc, no	1 ob/4 grid intervals	2D	0.801
STF_ER	smoothed truth	Yes, on true background error	SWV+sfc, yes	1 ob/4 grid intervals	3D	0.790
SNF_LR	smoothed truth	No	SWV+sfc, no	1 ob/8 grid intervals	3D	0.679
STF_LR	smoothed truth	Yes, on true background error	SWV+sfc, no	1 ob/8 grid intervals	3D	0.870

Retrieval experiments and results

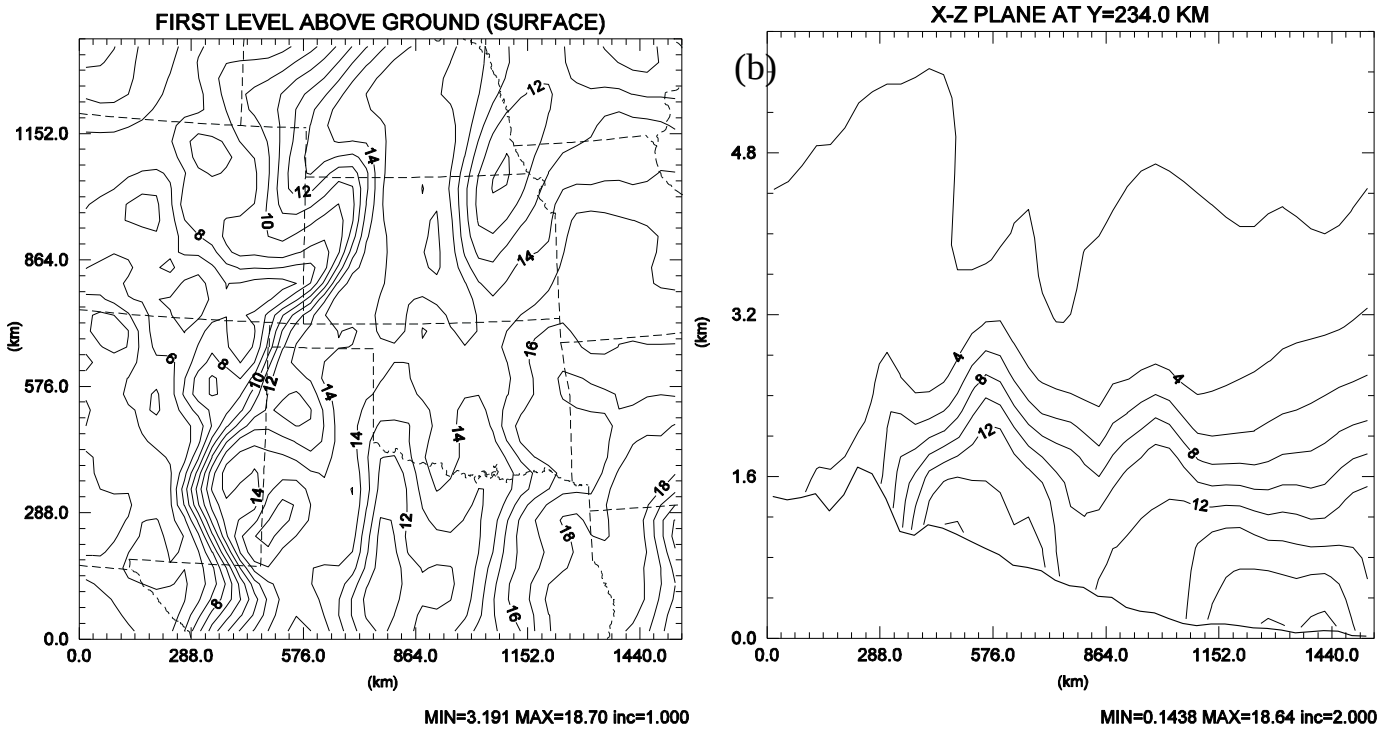


Figure 1. Specific humidity field (g kg^{-1}) from the ARPS simulated 'nature' for an IHOP case at 20 UTC, 19 June, 2002 (a) at the surface and (b) in the east-west vertical cross-section at y = 234 km. A roughly north-south zone of strong horizontal moisture gradient is located to the west of Kansas, Oklahoma and Texas, representing the dryline. In vertical cross-section, a boundary between the dry and moist air is oriented nearly vertically in the lowest 1.5 km then turns horizontal to the east.

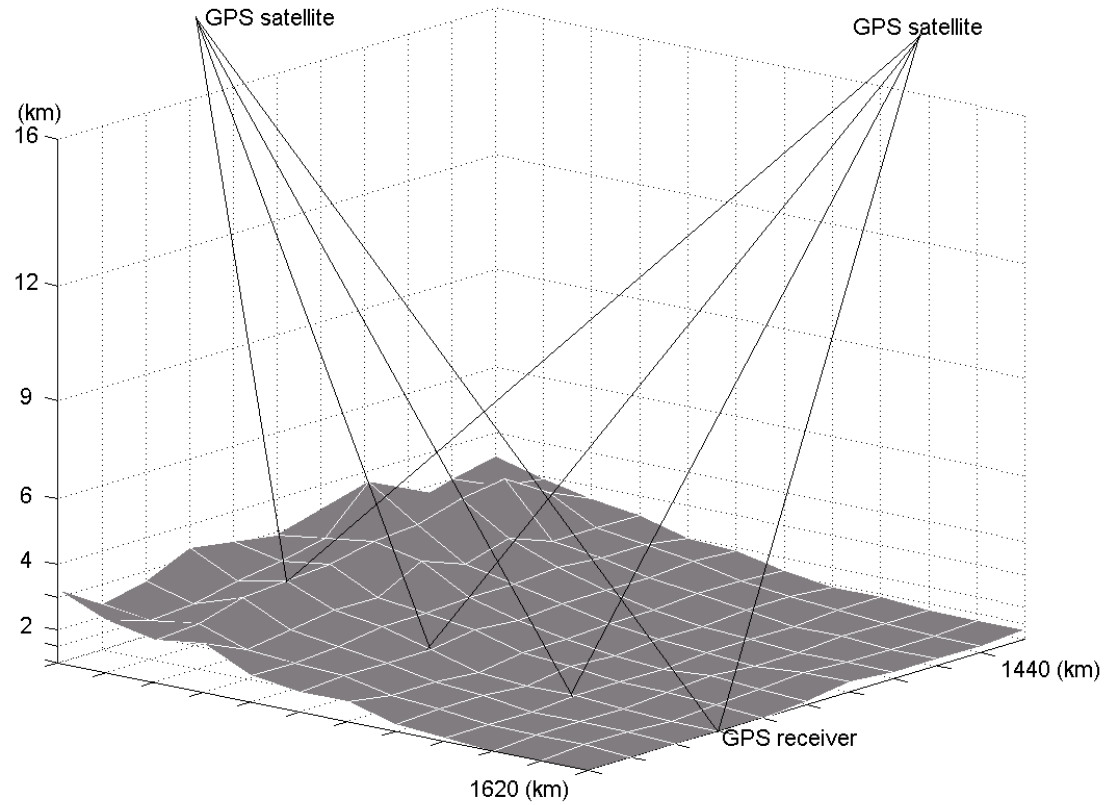


Figure 2. A schematic of hypothetic ground-based GPS observation network whose data are analyzed using 3DVAR. Shaded surface represents terrain. Dark solid lines are slant paths between ground-based GPS receivers and GPS satellites. Dotted lines give a sense of the vertically stretched grid although the actual grid levels are in terrain-following coordinate.

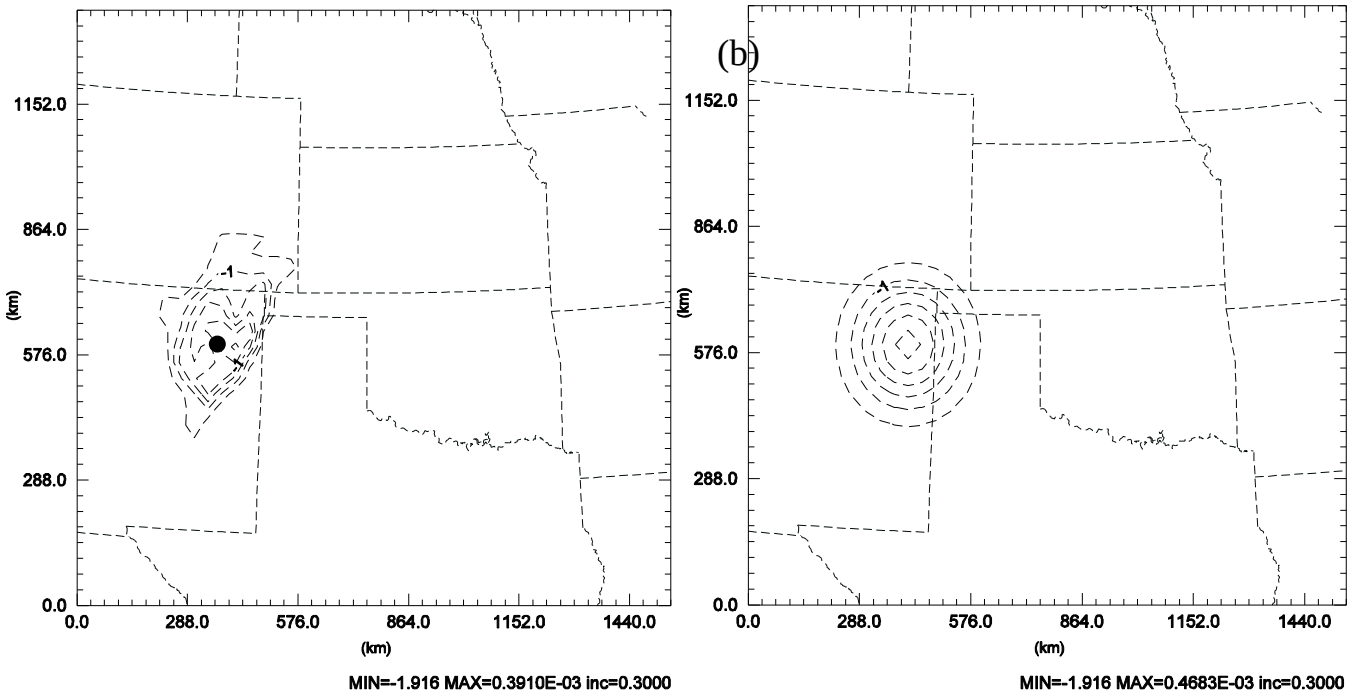


Figure 3. Specific humidity increment field in g kg^{-1} at the surface from single moisture observation tests, (a) for 3DVAR analysis with flow-dependent background error covariance and (b) for 3DVAR analysis with isotropic covariance. Only one specific humidity observation is present whose location is marked by the black dot in the figure.

Smoothed background

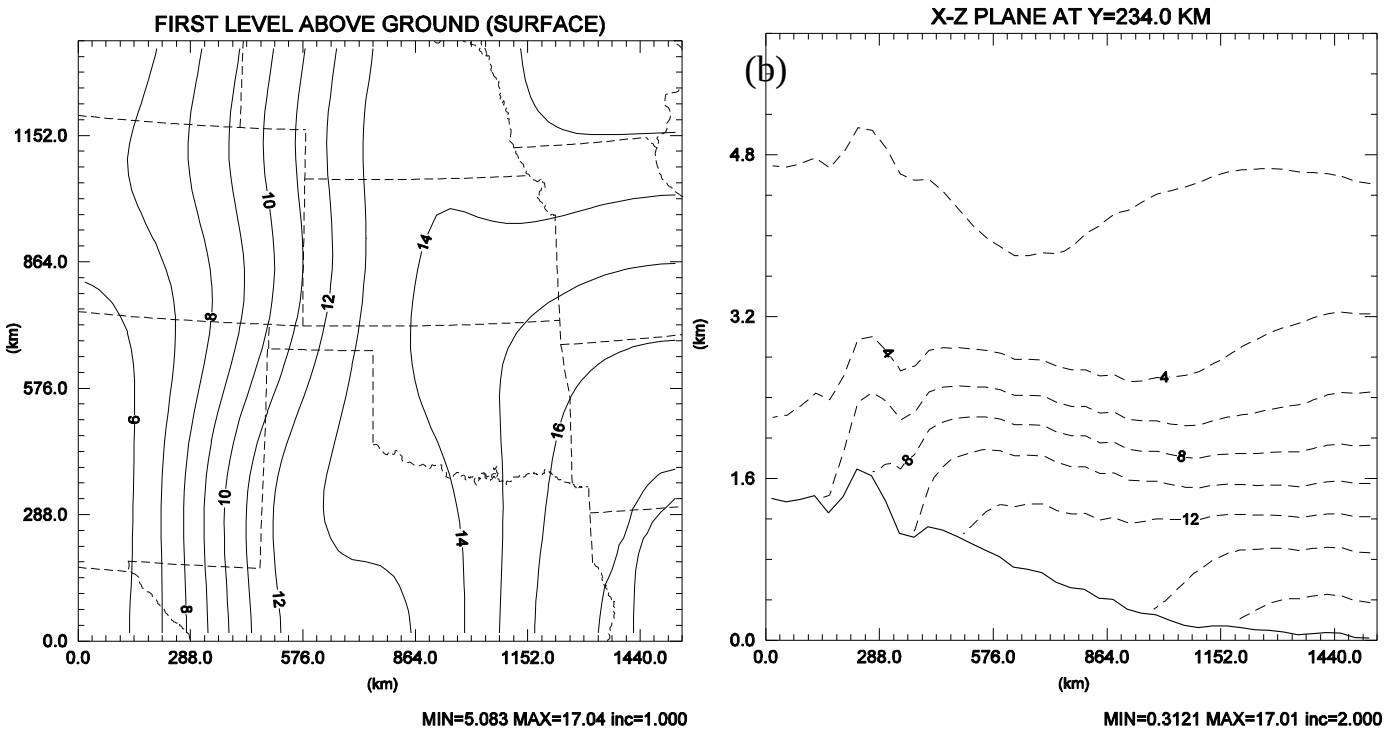


Figure 4. Background specific humidity field in g kg^{-1}, obtained by smoothing 'nature' 50 times using a 9-point filter in the horizontal, (a) at the surface and (b) in the east-west vertical cross-section at y = 234 km.

Analysis from CNTL Experiment, using flow-dependent B based on true error

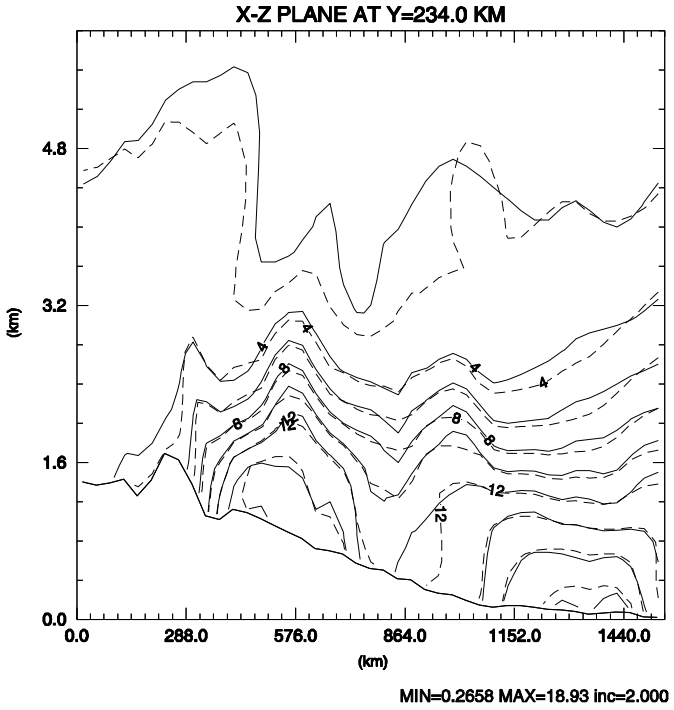


Figure 5. East-west vertical cross-section of specific humidity field (g kg^{-1}) at y = 234 km. Solid line is for 'nature' and dotted line from CNTL.

Analysis from CNTL Experiment, using flow-dependent B based on true error

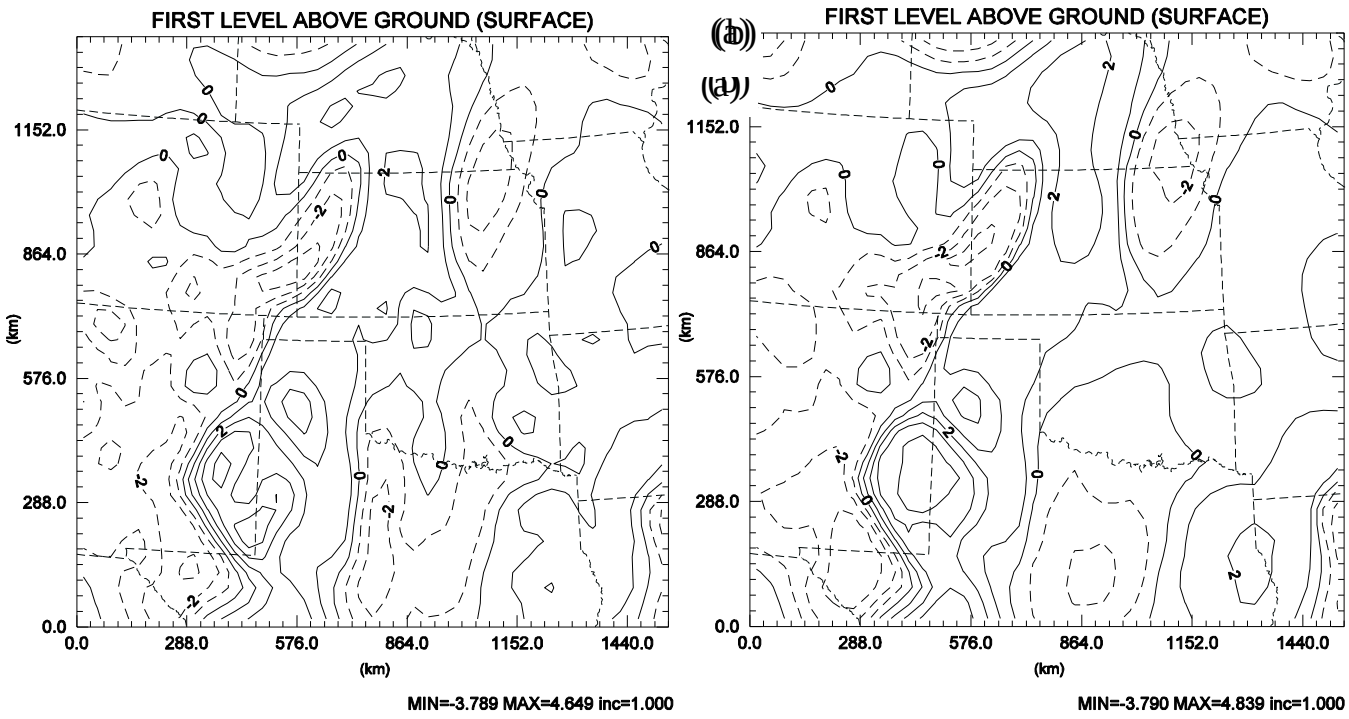


Figure 6. Specific humidity increment field in g kg^{-1} at the surface (a) from 'nature' and (b) from CNTL. Dashed lines represent negative values and solid lines positive values.

Analysis from SNF, without flow-dependent B

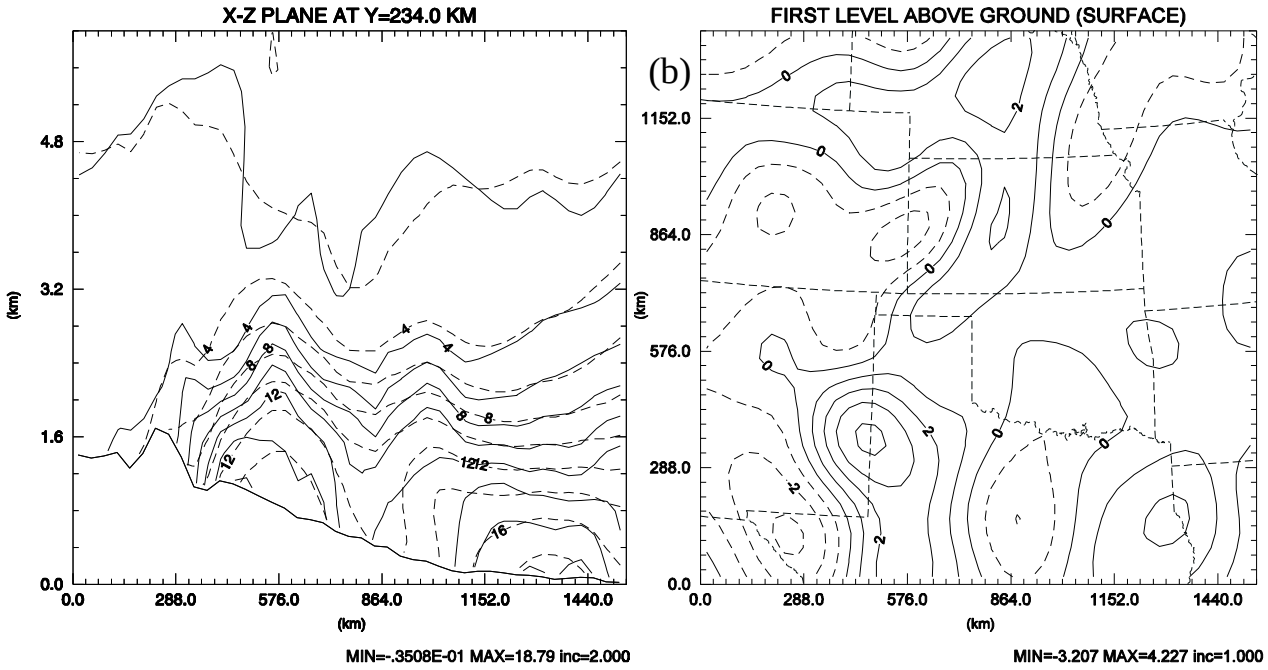


Figure 7. (a) East-west vertical cross-section of specific humidity field (g kg^{-1}) at y = 234 km where solid lines are for 'nature' and dotted lines for experiment SNF. (b) Analysis increment of specific humidity (g kg^{-1}) at the surface from experiment SNF, where dashed lines are for negative values and solid line for positive values.

Analysis from SUF, using flow-dependent B based on updated smoothed background

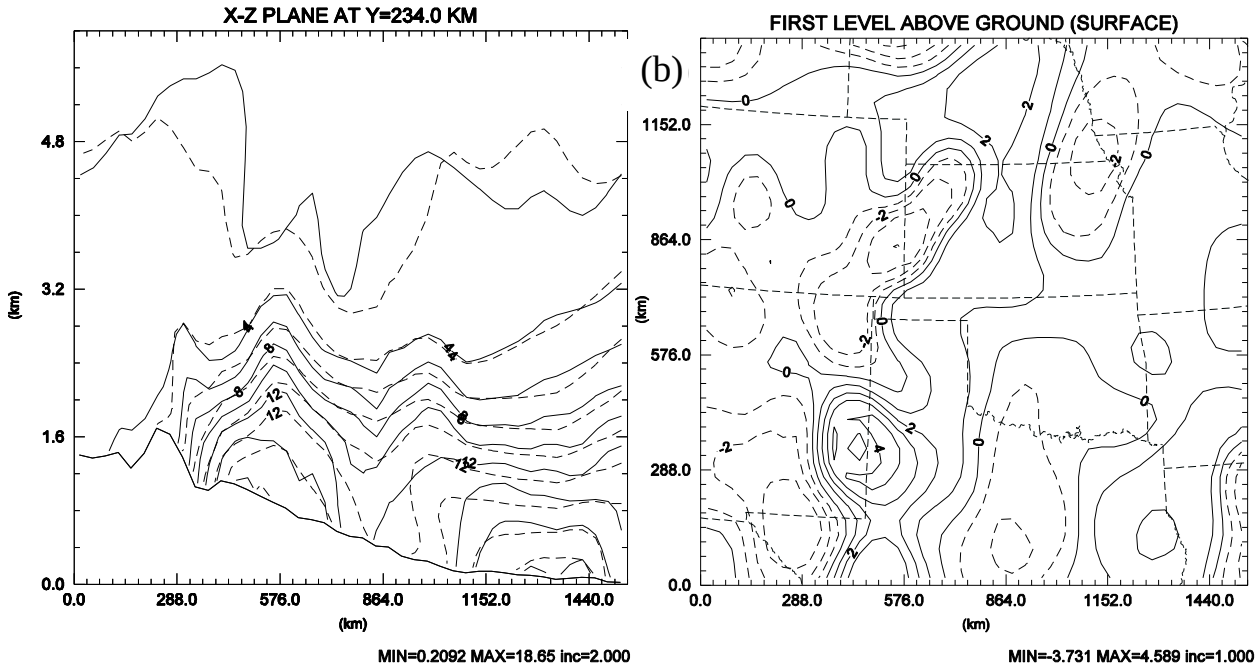


Figure 8. As Fig. 7 but for experiment SUF.

Analysis from LTF, using flow-dependent B based on the truth but starting from logarithmic background/first guess

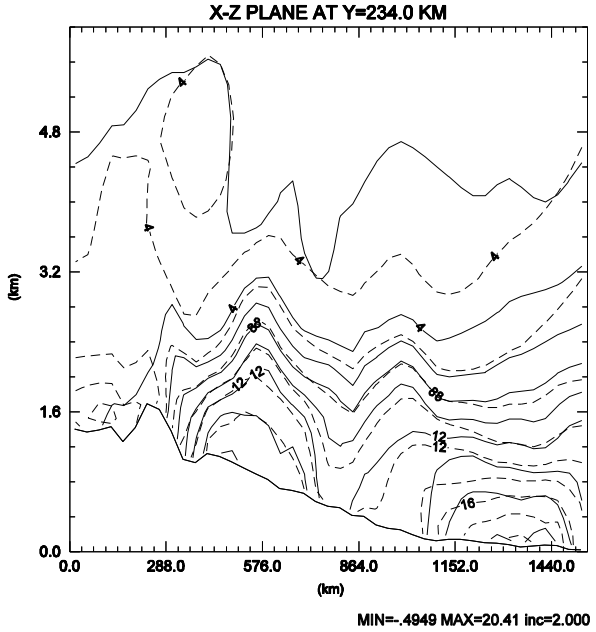


Figure 9. As Fig. 5 but dotted lines are from experiment LTF.

Analysis from LNF, without flow-dependent B and starting from logarithmic background/first guess

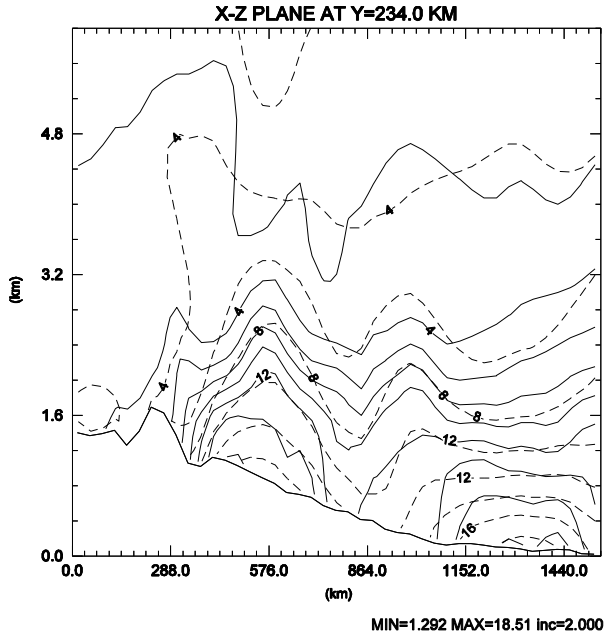


Figure 10. As Fig. 5 but dotted lines are from experiment LNF.

Analyses from STFNSFC (b) is for experiment SNFNSFC - No surface data

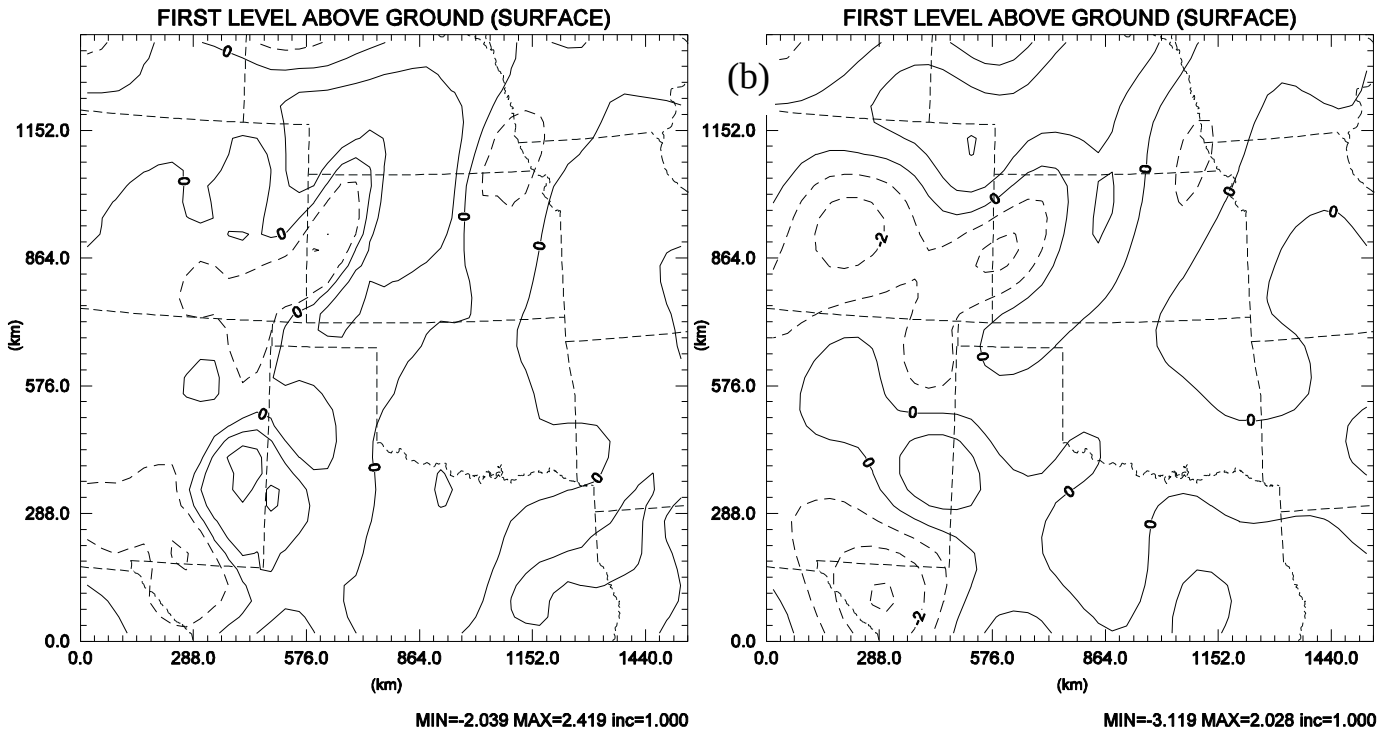


Figure 11. As Fig. 6 but (a) is for experiment STFNSFC (b) is for experiment SNFNSFC.

STFNSFC	smoothed truth	Yes, on true background error	SWV, no	1 ob/4 grid intervals	3D	0.894
SNFNSFC	smoothed truth	No	SWV, no	1 ob/4 grid intervals	3D	0.668

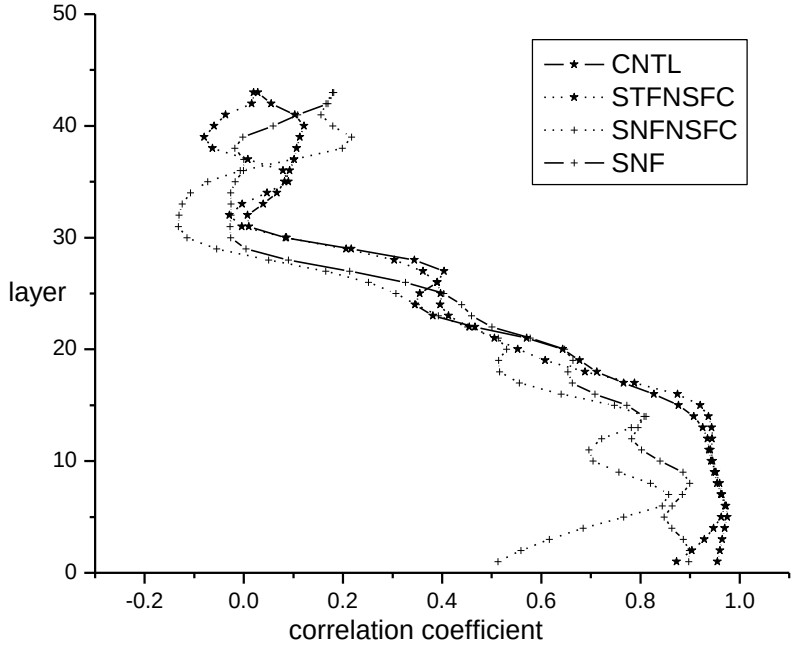


Figure 12. Profiles of correlation coefficient of specific humidity increment (difference from background, in g kg^{-1}) between those of 'nature' and 3DVAR analysis from experiments CNTL, STFNSFC, SNFNSFC, and SNF, plotted for different model levels.

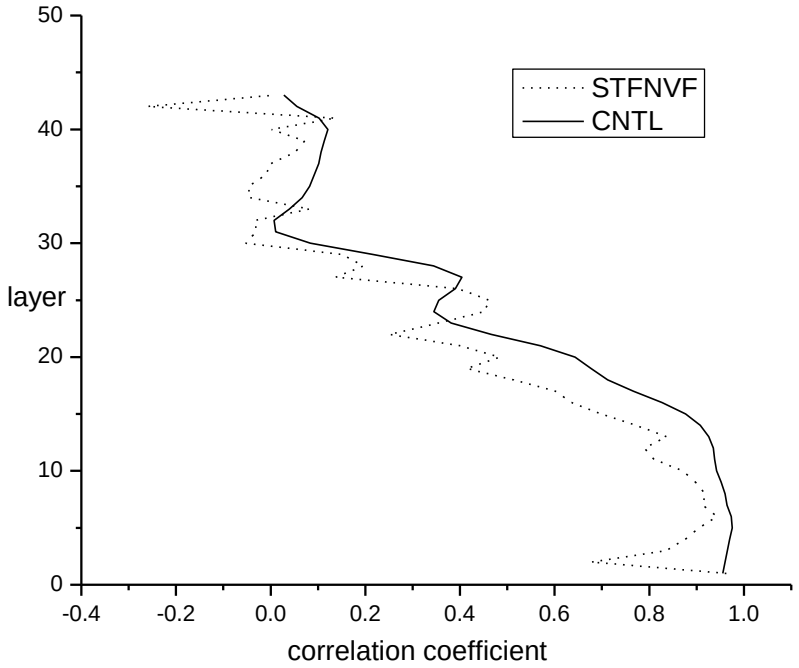


Figure 13. Profiles of correlation coefficient of specific humidity increment (g kg^{-1}) between those of 'nature' and 3DVAR analysis from CNTL (solid line) and experiment STFNVF (dotted line). Mean height of each level is given in the caption of Figure 12.

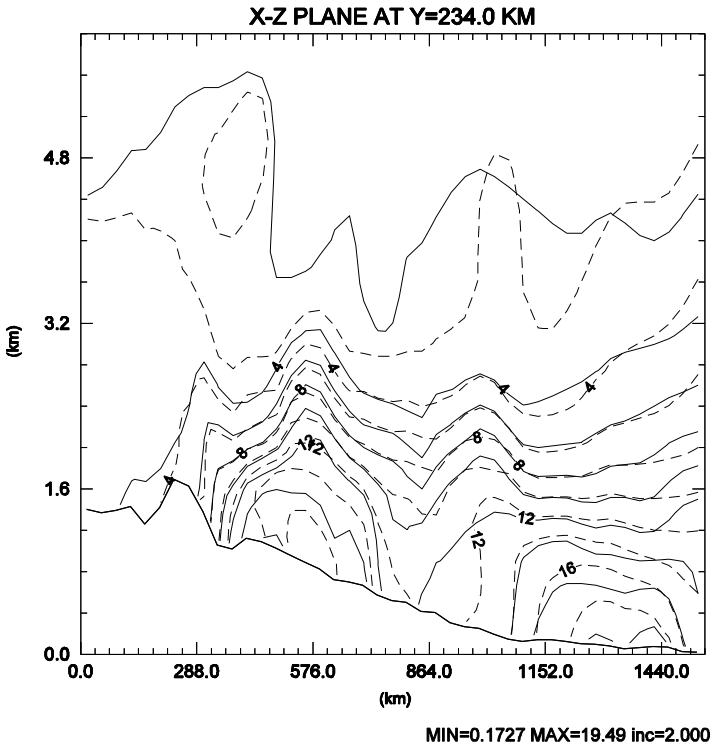


Figure 14. As Fig. 5 but dotted lines are for experiment STF_ER.

STF_ER	smoothed truth	Yes, on true background error	SWV+sfc, yes	1 ob/4 grid intervals	3D	0.790

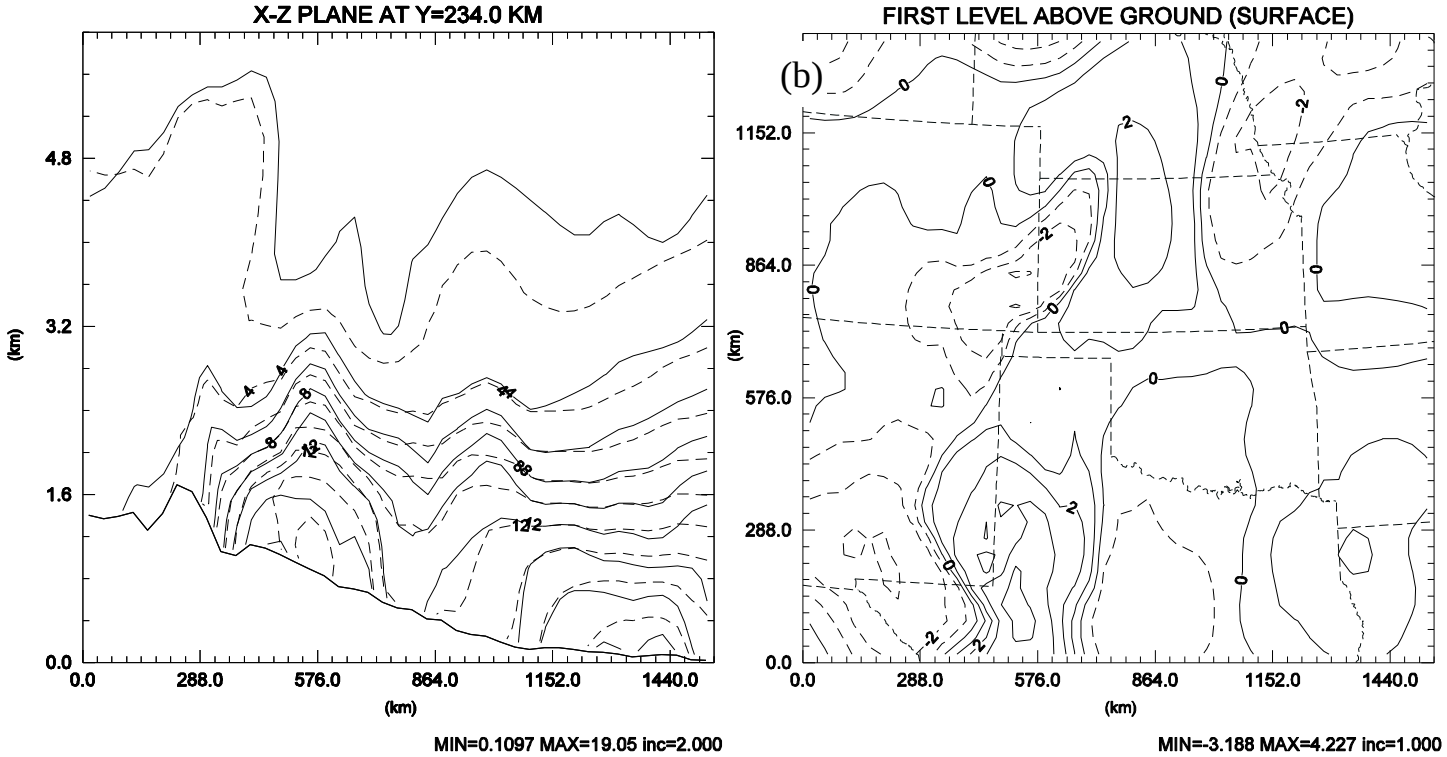


Figure 15. As Fig. 7 but for experiment STF_LR.

STF_LR	smoothed truth	Yes, on true background error	SWV+sfc, no	1 ob/8 grid intervals	3D	0.870