

Utah State University

DigitalCommons@USU

---

Space Dynamics Lab Publications

Space Dynamics Lab

---

1-1-2006

## Validation Assessment Model for Atmospheric Retrievals

Nikita Pougatchev

Gail Bingham

Joel Cardon

Karen St. Germain

Stephen Mango

Joe Tansock

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.usu.edu/sdl\\_pubs](https://digitalcommons.usu.edu/sdl_pubs)

---

### Recommended Citation

Pougatchev, Nikita; Bingham, Gail; Cardon, Joel; St. Germain, Karen; Mango, Stephen; Tansock, Joe; Zavyalov, Vladimir; Kireev, Stanislav; and Tobin, David, "Validation Assessment Model for Atmospheric Retrievals" (2006). *Space Dynamics Lab Publications*. Paper 107.

[https://digitalcommons.usu.edu/sdl\\_pubs/107](https://digitalcommons.usu.edu/sdl_pubs/107)

This Article is brought to you for free and open access by the Space Dynamics Lab at DigitalCommons@USU. It has been accepted for inclusion in Space Dynamics Lab Publications by an authorized administrator of DigitalCommons@USU. For more information, please contact [digitalcommons@usu.edu](mailto:digitalcommons@usu.edu).



---

**Authors**

Nikita Pougatchev, Gail Bingham, Joel Cardon, Karen St. Germain, Stephen Mango, Joe Tansock, Vladimir Zavyalov, Stanislav Kireev, and David Tobin

## Validation assessment model for atmospheric retrievals

Nikita Pougatchev<sup>a</sup>, Gail Bingham<sup>a</sup>, Joel Cardon<sup>a</sup>, Karen St. Germain<sup>b</sup>,  
Stephen Mango<sup>b</sup>, Joe Tansock<sup>a</sup>, Vladimir Zavyalov<sup>a</sup>, Stanislav Kireev<sup>c</sup>, and David Tobin<sup>d</sup>

<sup>a</sup>Space Dynamics Laboratory, 1695 North Research Park Way North Logan, UT 84341;

<sup>b</sup>E/IP Integrated Program Office, Centre Building, Ste. 2450, 8455 Colesville Road,  
Springfield, MD 20910

<sup>c</sup>Hampton University, Tyler St., Hampton VA 23668

<sup>d</sup>Space Science and Engineering Center, University of Wisconsin-Madison,  
1225 West Dayton St., Madison, WI 53706-1695

### ABSTRACT

A linear mathematical error model for the assessment of validation activity of atmospheric retrievals is presented. The purpose of the validation activity is to assess the actual performance of the remote sensing validated system while in orbit by comparing its measurements to some relevant—validating—data sets.

The validating system samples volumes of the atmosphere at times and locations that are different from the ones when and where the validated system makes its own observations. The location of the validating system can be either stationary, e.g. a ground ARM site, or movable, e.g. an aircraft or some other satellites. The true states may be correlated or not. The sampled volumes differ from each other by their location, timing, and size. The validated and validating systems have different vertical resolution and grid, absolute accuracy, and noise level. All the above factors cause apparent differences between the data to be compared.

The validation assessment model makes the comparison accurate by allowing for the differences. The model can be used for assessment and interpretation of the validation results when the above mentioned sources of discrepancies are significant, as well as for evaluation of a particular validating data source.

**Keywords:** atmospheric, validation, remote sensing, retrieval

### 1. INTRODUCTION

The purpose of validation is to assess actual performance of the validated system while in orbit, in other words, to establish how well the theoretical characterization and error analysis actually represent the properties of the real data.

The validated system performs a set of measurements  $\hat{\mathbf{x}}$  on an ensemble of true states  $\mathbf{x}$ ;

$$\hat{\mathbf{x}} = \mathbf{r}(\mathbf{x}) + \mathbf{e}$$

where  $\mathbf{r}(\mathbf{x})$  is a nominal retrieval with the absence of any errors in the measured signal and in the forward model, and  $\mathbf{e}$  represents retrieval errors caused by the factors mentioned above. The error term can be characterized by its mean value

$E\{\mathbf{e}\} = \mathbf{\Delta}$  (bias) and covariance  $\mathbf{S}_e$  (retrieval noise). From the error analysis we have a priori  ${}^a\mathbf{S}_e$ , and we presume that for a nominally performing system  $\mathbf{\Delta} = \mathbf{0}$ . In the process of validation we want to estimate the actual value of  $\mathbf{\Delta}$  and  $\mathbf{S}_e$ :

$$\hat{\mathbf{\Delta}} = E\{\hat{\mathbf{x}} - \mathbf{r}(\mathbf{x})\} \text{ and } \hat{\mathbf{S}}_e = E\{(\hat{\mathbf{\Delta}} - (\hat{\mathbf{x}} - \mathbf{r}(\mathbf{x})))(\hat{\mathbf{\Delta}} - (\hat{\mathbf{x}} - \mathbf{r}(\mathbf{x})))^T\}.$$

Validation is accomplished through a comparison with a set of profiles of a known quality and known relation to the nominal retrieval  $\mathbf{r}$  and/or the true state of the atmosphere  $\mathbf{x}$  at the moment of observation by the validated system. The validation data set can be based on real measurements by other (validating) systems or modeling.

For accurate comparison of any data sets, one needs to know the errors of the compared quantities. This inarguable general statement in the case of comparison of atmospheric profiles obtained by different techniques translates into some specific methodological issues to be resolved. Some of the problems are related to the physical principles of satellite measurement techniques; others reflect the spatial nonuniformity and temporal variations of the atmosphere. Satellite instrument and reference measurements sample the atmosphere with different vertical resolution, accuracy and noise level. The time and location of the compared data do not coincide. All the above mentioned factors cause inevitable differences between the compared profiles and their statistics.

The effect of different vertical resolutions was addressed by using averaging kernel formalism in application to CO and ozone retrieval validation<sup>1, 2, 3, 4</sup>, but the error caused by non-collocation was left out of the scope of the studies. In practice two approaches are used to handle the issue of non-collocation<sup>5</sup>. Thus, in the work of Tobin et al.<sup>6</sup> the “best estimate” radiosonde data set for the Southern Great Plains ARM site has been created. In this work the ground-based remote sensing and GOES satellite data were used to interpolate the original radiosonde measurements to correct for time and space differences between sondes and AIRS overpasses. In another study<sup>7</sup> no particular site adjustments were made, but global averages were compared. The first approach gives high accuracy data for comparison but is limited to the sites equipped by the ground-based remote sensors and requires dedicated sonde launches. The second approach provides global coverage but is not capable of providing measurement uncertainties indicated as a function of time and location for all of the data.

The goal of this work is to develop a linear mathematical error model for the assessment of validation activity of atmospheric profile retrievals. The validation assessment model makes the comparison accurate by estimating the difference. The model can be used for assessment and interpretation of the validation results when the above mentioned sources of discrepancy are significant, as well as for evaluation of a particular validating data source.

In Section 2 we present theoretical basis of the VAM mostly following the methodology, terminology, and notations developed by C. D. Rodgers<sup>2, 8, 9, 10</sup>. In Section 3 we substantiate it by presenting the results of a case study of the application of the VAM to a set of radiosonde profiles taken at a Southern Great Plains ARM site and simulated AIRS retrievals.

## 2. MODEL FOR VALIDATION OF PROFILE RETRIEVAL – THEORETICAL BASIS

Consider a validation scenario that consists of some elementary validation acts as shown in Figure 1. The validating system samples the volume of the atmosphere characterized by its true state  $\mathbf{x}_2$  at the time and location that are different from when and where the validated system makes its own observation of  $\mathbf{x}_1$ . Location of the validating system can be either stationary, e.g. a ground ARM site, or it can be movable, e.g. an aircraft or other satellites. The true states may be correlated or not, e.g. validation against historical records. The sampled volumes differ from each other by their location, timing, and size. The validated and validating systems have different characteristics, including vertical resolution and grid, absolute accuracy, and noise level. All the above factors cause an apparent difference between the data to be compared.

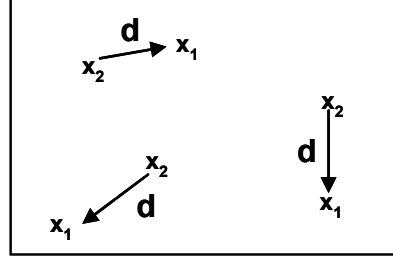


Fig. 1. Validation scenario: The rectangular encompasses the validation area;  $x_1$  is the state of the atmosphere sampled by the validated system,  $x_2$  is the one sampled by the validating system; states  $x_1$  and  $x_2$  are separated by time interval  $\tau$  and space vector  $d$ .

In linear approximation the retrieved profiles can be presented as follows<sup>10</sup>:

$$\hat{x}_i = x_{ai} + A_i(x_i - x_{ai}) + \Delta\hat{x}_i + \varepsilon_i \quad i = 1, 2 \quad (1)$$

where  $x_i$  and  $\hat{x}_i$  are true and retrieved profiles at the validation and observation sites and  $A_i$ ,  $x_{ai}$ ,  $\Delta\hat{x}_i$ , and  $\varepsilon_i$  are corresponding averaging kernel, a priori profile, measurement bias, and retrieval noise, respectively. Index 1 is assigned to the terms related to the validated system and index 2 to the validating one. Different  $A_i$ ,  $x_{ai}$ ,  $\Delta\hat{x}_i$ , and  $\varepsilon_i$  account for differences in the characteristics of the measurement systems. Equation (1) can represent both remote sensing and direct measurements; in the latter case  $A_2 = I$ , where  $I$  is an identity matrix.

For the proper comparison of the measurements, we must ensure that we compare like with like, in other words that profiles  $\hat{x}_1$  and  $\hat{x}_2$  are on the same grid. In further theoretical consideration we assume that the compared profiles and their characteristics are on the same vertical grid. We will address the issue of different grids in Section 3.

The true states  $x_1$  and  $x_2$  are functions of coordinate  $z$  and time  $t$  where  $x_1 = x(z_1, t_1)$   $x_2 = x(z_2, t_2)$ . The separation between the measurements in space is  $d = z_2 - z_1$  and in time is  $\tau = t_2 - t_1$ . The true mean value of the ensemble of states is  $\bar{x}_1 = E\{x_1(z_1, t_1)\}$  and  $\bar{x}_2 = E\{x_2(z_2, t_2)\}$ . The variations of the states about their means  $\bar{x}_1$  and  $\bar{x}_2$  are characterized by their covariances  $S_{x_1} = E\{(x_1 - \bar{x}_1)(x_1 - \bar{x}_1)^T\}$  and  $S_{x_2} = E\{(x_2 - \bar{x}_2)(x_2 - \bar{x}_2)^T\}$ .

Correlation between the true states  $x_1$  and  $x_2$  can be characterized by covariances  $S_{12} = E\{(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)^T\}$  and  $S_{21} = E\{(x_2 - \bar{x}_2)(x_1 - \bar{x}_1)^T\}$  for the covariances  $S_{12} = S_{21}^T$  and  $S_{12} = S_{21} = S_{x_1} = S_{x_2}$  when  $\tau = 0$  and  $d = 0$ .

Because correlation measures only linear relationships, in the following consideration we assume that the variation of the true states about their means  $\bar{x}_1$  and  $\bar{x}_2$  -  $\delta x_1$  and  $\delta x_2$  are correlated so that

$$\delta x_1 = B\delta x_2 + \xi \quad (2)$$

where  $\xi$  is random with  $\bar{\xi} = 0$  and covariance  $S_\xi$ . The relation (2) can be interpreted in terms of conditional probability density function (*pdf*)  $P(\delta x_1 | \delta x_2)$  of  $\delta x_1$  given  $\delta x_2$ , i.e.  $B\delta x_2$  is the expected value of  $\delta x_1$  given  $\delta x_2$

$$B\delta x_2 = \int \delta x_1 P(\delta x_1 | \delta x_2) d(\delta x_1) \quad (3)$$

For covariances  $S_{x_1}$ ,  $S_{x_2}$ , and  $S_\xi$  the following relations are true:

$$\text{cov}(x_2, \xi) = 0 \quad (4)$$

$$S_{x_1} = BS_{x_2}B^T + S_\xi \quad (5)$$

$$S_{12} = BS_{x_2} \quad (6)$$

$$S_{21} = S_{x_2}B^T \quad (7)$$

The function  $B = I$  when  $\tau = 0$  and  $d = 0$  and  $B = 0$  when  $S_{12} = S_{21} = 0$ .

We assume that the validating system is free from bias  ${}^a\Delta\hat{x}_2 = 0$  to the accuracy characterized by  ${}^aS_{\Delta\hat{x}_2}$  and that its retrieval noise covariance is  $S_{\varepsilon_2}$ ; the a priori bias of the validated system is  ${}^a\Delta\hat{x}_1 = 0$  with uncertainty represented by  ${}^aS_{\Delta\hat{x}_1}$  and retrieval noise covariance  ${}^aS_{\varepsilon_1}$ . The purpose of validation is to estimate the actual values of  $\Delta\hat{x}_1$  and  $S_{\varepsilon_1}$ .

Following the approach from Rodgers and Connor<sup>2</sup> and Equations (2) and (3) we simulate retrieval  $\hat{x}_1$  with  $\hat{x}_2$ :

$$\hat{x}_{12} = A_1B\hat{x}_2 = A_1B(I - A_2)x_{a2} + A_1BA_2x_2 + A_1B\varepsilon_2 \quad (8)$$

For estimation of the bias  $\Delta\hat{x}_1$ , consider the difference  $\delta\hat{x}$ :

$$\begin{aligned} \delta\hat{x} &\equiv \hat{x}_1 - \hat{x}_{12} = [(I - A_1)x_{a1} - A_1B(I - A_2)x_{a2}] + A_1x_1 - A_1BA_2x_2 + \Delta\hat{x}_1 + (\varepsilon_1 - A_1B\varepsilon_2) = \\ &= [(I - A_1)x_{a1} - A_1B(I - A_2)x_{a2}] + A_1\bar{x}_1 - A_1BA_2\bar{x}_2 + A_1\delta x_1 - A_1BA_2\delta x_2 + \Delta\hat{x}_1 + (\varepsilon_1 - A_1B\varepsilon_2) \\ &= [(I - A_1)x_{a1} - A_1B(I - A_2)x_{a2}] + A_1\bar{x}_1 - A_1BA_2\bar{x}_2 + A_1(B\delta x_2 + \xi) - A_1BA_2\delta x_2 + \Delta\hat{x}_1 + (\varepsilon_1 - A_1B\varepsilon_2) \end{aligned}$$

and its mean  $\bar{\delta\hat{x}}$

$$\bar{\delta\hat{x}} \equiv \bar{\hat{x}}_1 - \bar{\hat{x}}_{12} = [(I - A_1)x_{a1} - A_1B(I - A_2)x_{a2}] + A_1\bar{x}_1 - A_1BA_2\bar{x}_2 + \Delta\hat{x}_1 \quad (9)$$

Then,

$$\Delta\hat{x}_1 = \bar{\delta\hat{x}} - {}^e\delta\hat{x} \quad (10)$$

where

$${}^e\bar{\delta\hat{\mathbf{x}}} \equiv [(\mathbf{I} - \mathbf{A}_1)\mathbf{x}_{a1} - \mathbf{A}_1\mathbf{B}(\mathbf{I} - \mathbf{A}_2)\mathbf{x}_{a2}] + \mathbf{A}_1\bar{\mathbf{x}}_1 - \mathbf{A}_1\mathbf{B}\mathbf{A}_2\bar{\mathbf{x}}_2 \quad (11)$$

Covariance of  $\delta\hat{\mathbf{x}}$  about the mean -  $\mathbf{S}_{\delta\hat{\mathbf{x}}}$  is:

$$\begin{aligned} \mathbf{S}_{\delta\hat{\mathbf{x}}} &\equiv E\{(\delta\hat{\mathbf{x}} - \bar{\delta\hat{\mathbf{x}}})(\delta\hat{\mathbf{x}} - \bar{\delta\hat{\mathbf{x}}})^T\} = \\ &= E\{[(\mathbf{A}_1\mathbf{B} - \mathbf{A}_1\mathbf{B}\mathbf{A}_2)\delta\mathbf{x}_2 - \mathbf{A}_1\xi + (\boldsymbol{\varepsilon}_1 - \mathbf{A}_1\mathbf{B}\boldsymbol{\varepsilon}_2)] [(\mathbf{A}_1\mathbf{B} - \mathbf{A}_1\mathbf{B}\mathbf{A}_2)\delta\mathbf{x}_2 - \mathbf{A}_1\xi + (\boldsymbol{\varepsilon}_1 - \mathbf{A}_1\mathbf{B}\boldsymbol{\varepsilon}_2)]^T\} = \\ &= (\mathbf{A}_1\mathbf{B}(\mathbf{I} - \mathbf{A}_2))\mathbf{S}_{x_2}(\mathbf{A}_1\mathbf{B}(\mathbf{I} - \mathbf{A}_2))^T + \mathbf{A}_1\mathbf{S}_{\xi}\mathbf{A}_1^T + \mathbf{S}_{\boldsymbol{\varepsilon}_1} + (\mathbf{A}_1\mathbf{B})\mathbf{S}_{\boldsymbol{\varepsilon}_2}(\mathbf{A}_1\mathbf{B})^T \end{aligned} \quad (12)$$

where  $\mathbf{S}_{\boldsymbol{\varepsilon}_1}$  and  $\mathbf{S}_{\boldsymbol{\varepsilon}_2}$  are characteristics of the retrieval noise.

It is important to notice that in this context,  ${}^e\bar{\delta\hat{\mathbf{x}}}$  is not the error but represents the expected difference between nominally performing measurement systems instead. The purpose of the validation is to determine the deviation from the expected difference and the statistical significance of the deviation.

Attainable accuracy of  $\Delta\hat{\mathbf{x}}_1$  in Equation (10) is limited by the accuracy of our a priori knowledge and the bias of the reference system. In practice we know the mean of the states with some uncertainties characterized by the covariances. The uncertainty of the assumption characterized by the covariance results in an additional error with covariance  $\tilde{\mathbf{S}}$ :

$$\tilde{\mathbf{S}} = \mathbf{A}_1\tilde{\mathbf{S}}_{\bar{\mathbf{x}}_1}\mathbf{A}_1^T + (\mathbf{A}_1\mathbf{B}\mathbf{A}_2)\tilde{\mathbf{S}}_{\bar{\mathbf{x}}_2}(\mathbf{A}_1\mathbf{B}\mathbf{A}_2)^T \quad (13)$$

Finite accuracy of the assumption  ${}^a\Delta\hat{\mathbf{x}}_2 = \mathbf{0}$  is characterized by covariance  ${}^a\mathbf{S}_{\Delta\hat{\mathbf{x}}_2}$  and results in additional error in the bias of validated system

$$\tilde{\mathbf{S}}_{\Delta\hat{\mathbf{x}}_2} = (\mathbf{A}_1\mathbf{B}\mathbf{A}_2){}^a\mathbf{S}_{\Delta\hat{\mathbf{x}}_2}(\mathbf{A}_1\mathbf{B}\mathbf{A}_2)^T \quad (14)$$

In the process of validation a set of  $\{\delta_i\hat{\mathbf{x}}\} i = 1, 2, \dots, N$  is measured. We assume that the measurements are made such that all  $\delta\hat{\mathbf{x}}_i$  are statistically independent. The mean of the sample  ${}^s\bar{\delta\hat{\mathbf{x}}}$  relates to the mean of the ensemble  $\bar{\delta\hat{\mathbf{x}}}$  as  ${}^s\bar{\delta\hat{\mathbf{x}}} = \bar{\delta\hat{\mathbf{x}}} + \boldsymbol{\varepsilon}_s$  where  $\boldsymbol{\varepsilon}_s$  represents the error due to the difference between the mean of the sample and the mean of the ensemble. It is a random vector with covariance  $\mathbf{S}_{\boldsymbol{\varepsilon}_s} = N^{-1}\mathbf{S}_{\delta\hat{\mathbf{x}}}$

Then the estimation of the bias of the validated system is

$${}^s\Delta\hat{\mathbf{x}}_1 = {}^s\bar{\delta\hat{\mathbf{x}}} - {}^e\bar{\delta\hat{\mathbf{x}}} \quad (15)$$

The covariance of the estimate is

$$\mathbf{S}_{{}^s\Delta\hat{\mathbf{x}}} = \mathbf{S}_{\boldsymbol{\varepsilon}_s} + \tilde{\mathbf{S}} + \tilde{\mathbf{S}}_{\Delta\hat{\mathbf{x}}_2} \quad (16)$$

The retrieval noise  $\mathbf{S}_{\varepsilon_1}$  of the validated system can be estimated based on analysis of the measurements by the validated system on the ensemble of states  $\mathbf{x}_1$  with  $\mathbf{S}_{\mathbf{x}_1}$  such that  $\mathbf{A}_1 \mathbf{S}_{\mathbf{x}_1} \mathbf{A}_1^T \ll \mathbf{S}_{\varepsilon_1}$ .

For validation planning and evaluating of a particular validation data source one can use the following relations:

$${}^a \mathbf{S}_{\delta \hat{\mathbf{x}}} = (\mathbf{A}_1 \mathbf{B} (\mathbf{I} - \mathbf{A}_2)) \mathbf{S}_{\mathbf{x}_2} (\mathbf{A}_1 \mathbf{B} (\mathbf{I} - \mathbf{A}_2))^T + \mathbf{A}_1 \mathbf{S}_{\xi} \mathbf{A}_1^T + \mathbf{S}_{\varepsilon_1} + (\mathbf{A}_1 \mathbf{B}) \mathbf{S}_{\varepsilon_2} (\mathbf{A}_1 \mathbf{B})^T \quad (17)$$

$$\mathbf{S}_{\varepsilon_1} \approx {}^a \mathbf{S}_{\delta \hat{\mathbf{x}}} / N \quad (18)$$

$$\mathbf{S}_{\mathbf{s}_{\Delta \hat{\mathbf{x}}}} = {}^a \mathbf{S}_{\delta \hat{\mathbf{x}}} / N + \tilde{\mathbf{S}}_{\delta \hat{\mathbf{x}}} + \tilde{\mathbf{S}}_{\Delta \hat{\mathbf{x}}_2} \quad (19)$$

Equation (19) gives us an estimate of the validation accuracy attainable given the accuracy of our knowledge of the characteristics of the measurement systems and the true states, and the size of data sample used for the validation.

*Remark*

Two sources of different natures contribute to the total covariance in Equation (12), namely, natural variation of the true states and retrieval noise. Most probably, the variations have different correlation times. Therefore, it is important to ensure that the sampling preserves the statistical independence to make Equation (18) applicable. For example, assume that the noise is independent between single measurements, that we can take  $K$  measurements during a correlation time of the natural variation -  $t_c$ , and that the validation lasted  $M t_c$ ; Equation (14) then takes the form

$$\mathbf{S}_{\mathbf{s}_{\delta \hat{\mathbf{x}}}} = [(\mathbf{A}_1 \mathbf{B} (\mathbf{I} - \mathbf{A}_2)) \mathbf{S}_{\mathbf{x}_2} (\mathbf{A}_1 \mathbf{B} (\mathbf{I} - \mathbf{A}_2))^T + \mathbf{A}_1 \mathbf{S}_{\xi} \mathbf{A}_1^T + (\mathbf{S}_{\varepsilon_1} + (\mathbf{A}_1 \mathbf{B}) \mathbf{S}_{\varepsilon_2} (\mathbf{A}_1 \mathbf{B})^T) / K] / M \quad (20)$$

## 2.1. Model for validation of profile retrieval – particular cases

(i) Consider the case when the vertical resolution of the validating system is high enough to assume that  $\mathbf{A}_2 = \mathbf{I}$ , for example, in situ air-borne sensors, lidars, etc. Then Equations (11) and (12) take the forms

$${}^e \delta \hat{\mathbf{x}} = (\mathbf{I} - \mathbf{A}_1) \mathbf{x}_{a1} + \mathbf{A}_1 \bar{\mathbf{x}}_1 - \mathbf{A}_1 \mathbf{B} \bar{\mathbf{x}}_2 \quad (21)$$

and

$$\mathbf{S}_{\delta \hat{\mathbf{x}}} = \mathbf{A}_1 \mathbf{S}_{\xi} \mathbf{A}_1^T + \mathbf{S}_{\varepsilon_1} + (\mathbf{A}_1 \mathbf{B}) \mathbf{S}_{\varepsilon_2} (\mathbf{A}_1 \mathbf{B})^T \quad (22)$$

The term  $\mathbf{A}_1 \mathbf{S}_{\xi} \mathbf{A}_1^T$  in Equation (22) represents the residual variations of  $\mathbf{x}_1$  that cannot be inferred from the correlative validation data set.

(ii) The validation scenario discussed above is the generalization of the scenario that was analyzed by Rodgers and Connor<sup>2</sup>. They examined the case when two remote sounders with different characteristics sample the same true state. That corresponds to  $\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}$ ;  $\mathbf{S}_{\mathbf{x}_1} = \mathbf{S}_{\mathbf{x}_2} = \mathbf{S}_{\mathbf{x}}$ ,  $\mathbf{S}_{\xi} = \mathbf{0}$  and  $\mathbf{B} = \mathbf{I}$ , and the equation (12) takes form

$$\begin{aligned} \mathbf{S}_{\delta \hat{\mathbf{x}}} &\equiv E\{(\delta \hat{\mathbf{x}} - \delta \hat{\mathbf{x}})(\delta \hat{\mathbf{x}} - \delta \hat{\mathbf{x}})^T\} = \\ &= (\mathbf{A}_1 - \mathbf{A}_1 \mathbf{A}_2) \mathbf{S}_{\mathbf{x}} (\mathbf{A}_1 - \mathbf{A}_1 \mathbf{A}_2)^T + \mathbf{S}_{\varepsilon_1} + \mathbf{A}_1 \mathbf{S}_{\varepsilon_2} \mathbf{A}_1^T \end{aligned} \quad (23)$$

which is identical to Equation (30) in their work. The estimated bias is

$${}^s \Delta \hat{\mathbf{x}}_1 = {}^s \delta \hat{\mathbf{x}} - [(\mathbf{I} - \mathbf{A}_1) \mathbf{x}_{a1} - \mathbf{A}_1 (\mathbf{I} - \mathbf{A}_2) \mathbf{x}_{a2}] + (\mathbf{A}_1 - \mathbf{A}_1 \mathbf{A}_2) \bar{\mathbf{x}} \quad (24)$$



with covariance

$$\mathbf{S}_{s_{\Delta\hat{\mathbf{x}}}} = \mathbf{S}_{\varepsilon_s} + \tilde{\mathbf{S}}_{\Delta\hat{\mathbf{x}}_2} \quad (25)$$

(iii) One of the approaches to the validation is validation against a priori/historical data records<sup>11</sup>. In this case the only information about  $\mathbf{x}_1$  is  $\bar{\mathbf{x}}_1$ ,  $\tilde{\mathbf{S}}_{\bar{\mathbf{x}}_1}$ , and  $\mathbf{S}_{\mathbf{x}_1}$ , i.e.  $\mathbf{B} = \mathbf{0}$ ,  $\delta\hat{\mathbf{x}} = \hat{\mathbf{x}}_1$ , and Equations (11), (12), and (16) take the forms

$${}^e\delta\hat{\mathbf{x}} = (\mathbf{I} - \mathbf{A}_1)\mathbf{x}_{a1} + \mathbf{A}_1\bar{\mathbf{x}}_1 \quad (26)$$

$$\mathbf{S}_{\delta\hat{\mathbf{x}}} \equiv E\{(\delta\hat{\mathbf{x}} - \delta\hat{\mathbf{x}})(\delta\hat{\mathbf{x}} - \delta\hat{\mathbf{x}})^T\} = \mathbf{A}_1\mathbf{S}_{\mathbf{x}_1}\mathbf{A}_1^T + \mathbf{S}_{\varepsilon_s} \quad (27)$$

$$\mathbf{S}_{s_{\Delta\hat{\mathbf{x}}}} = \mathbf{S}_{\varepsilon_s} + \mathbf{A}_1\tilde{\mathbf{S}}_{\bar{\mathbf{x}}_1}\mathbf{A}_1^T \quad (28)$$

### 3. RESULTS OF CASE STUDY

This section demonstrates the practicability of the theoretical basis presented in Section 2. We applied the VAM to a set of radiosonde profiles taken at the ARM Southern Great Plains site from July to December of 2002. These are the same data that have been used to build the “best estimate data set” in Tobin et al.<sup>6</sup>. In particular, we analyzed the impact of the time difference between satellite and radiosonde measurements on the assessment of accuracy of the AVTP retrieval. In the following case study we take the radiosonde profiles for accurate representation of the true states of the atmosphere.

To remove seasonal cycle variation from the analyzed data, we de-seasonalized the whole set; for each month we calculated the monthly mean profile and extracted it from each particular profile pertaining to the month. Then out of the full set of 424 de-seasonalized profiles, we constructed two ensembles so that each sonde in the first ensemble ( $\mathbf{x}_1$ ) had at least one reciprocal sonde in the second ensemble ( $\mathbf{x}_2$ ) with a launch time difference less than or equal to  $\tau$  hours, where  $\tau = 3, 6, 12, 24, 48, \text{ and } 72$  hours. For every  $\tau \geq 6$  hours the size of the ensemble was greater than 100. Then we calculated auto-covariances  $\mathbf{S}_{\mathbf{x}_1}, \mathbf{S}_{\mathbf{x}_2}$  and cross-covariance  $\mathbf{S}_{12}$ . The results are presented in Figure 2.

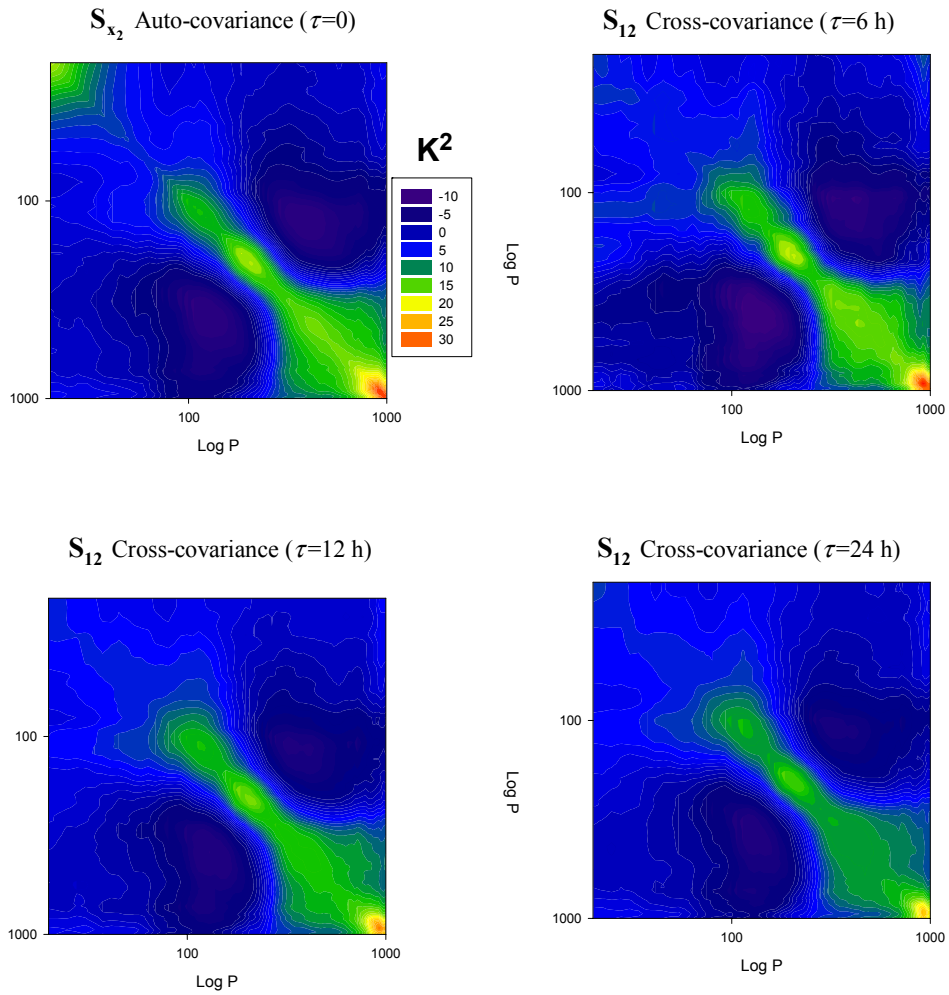


Fig. 2. Auto-covariance and selected cross-covariance matrices.

As one can see, there is noticeable decreasing with the  $\tau$  correlation between the ensembles. Given  $S_{x_2}$  and  $S_{12}$  using Equations (13) and (14) we calculated  $S_{\xi}$ , which characterizes the uncorrelated temperature difference between  $x_1$  and  $x_2$ . For the comparison of two radiosonde profiles, the square root of the diagonal elements of  $S_{\xi}$  can be interpreted as *rms* error caused by non-coincidence of launch times. Plots for error are displayed in Figure 3.

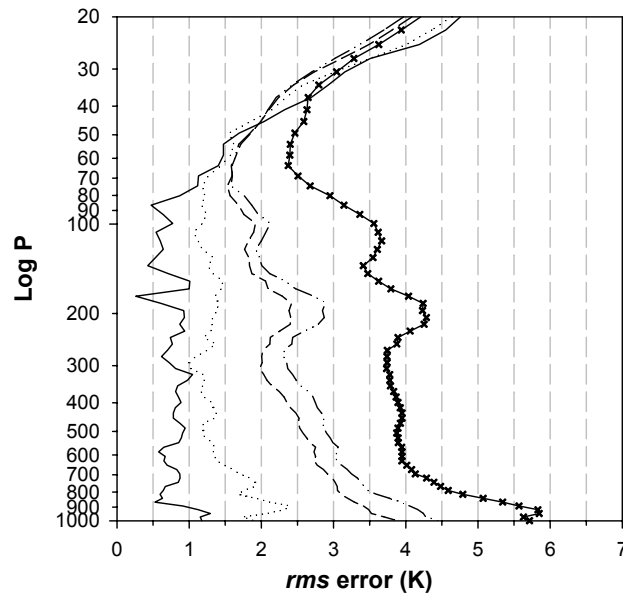


Fig. 3. The square root of  $\mathbf{S}_z$  diagonal elements (*rms* error for comparison of a single pair of radiosondes) for different  $\tau$ . The solid line is for  $\tau = 3$  h; the dotted line is for  $\tau = 6$  h; the dashed line is for  $\tau = 12$  h; the dash-dot-dot line is for  $\tau = 24$  h; and the solid-with-crosses is for  $\tau = \infty$ , i.e. no correlation at all, validation against historical records.

One of the interpretations of the results in Figure 3 is that given a particular temperature profile  $\mathbf{x}_2$  with associated statistical characteristics  $\mathbf{S}_{x_2}$ ,  $\mathbf{S}_{12}$ ,  $\bar{\mathbf{x}}_2$ , and  $\bar{\mathbf{x}}_1$ , we can estimate profile  $\mathbf{x}_1$  separated in time by less than  $\tau$  with *rms* error indicated by the corresponding curve in Figure 3.

To simulate the smoothing error of the satellite retrieval, we applied AIRS-like averaging kernels (see Figure 4) to Equation (20). Each averaging kernel is for the temperature profile retrieval in a 1-km thick layer; spectral resolution and noise level are those for the AIRS instrument. Certainly, averaging kernels depend on the state of the atmosphere for a particular retrieval, but they are not critical to the error analysis. Thus, we calculated  $\mathbf{S}_{\delta\hat{\mathbf{x}}}$ , the covariances of the single pair comparison error caused by the time difference in sonde launch and satellite overpass only. Finally, using Equation (19) with the assumption  $\mathbf{x}_a = \mathbf{0}$ , we estimated the mean expected difference  ${}^e\delta\hat{\mathbf{x}}$  with the associated error (see Equation (25)). The results are presented in Figure 4.

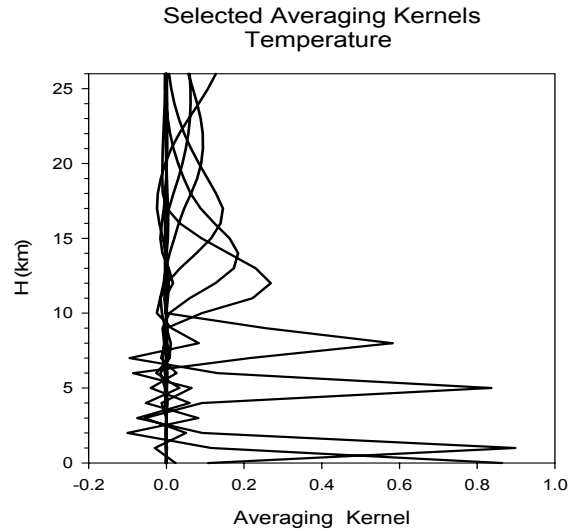


Fig. 4. Selected averaging kernels for temperature profile retrieval.

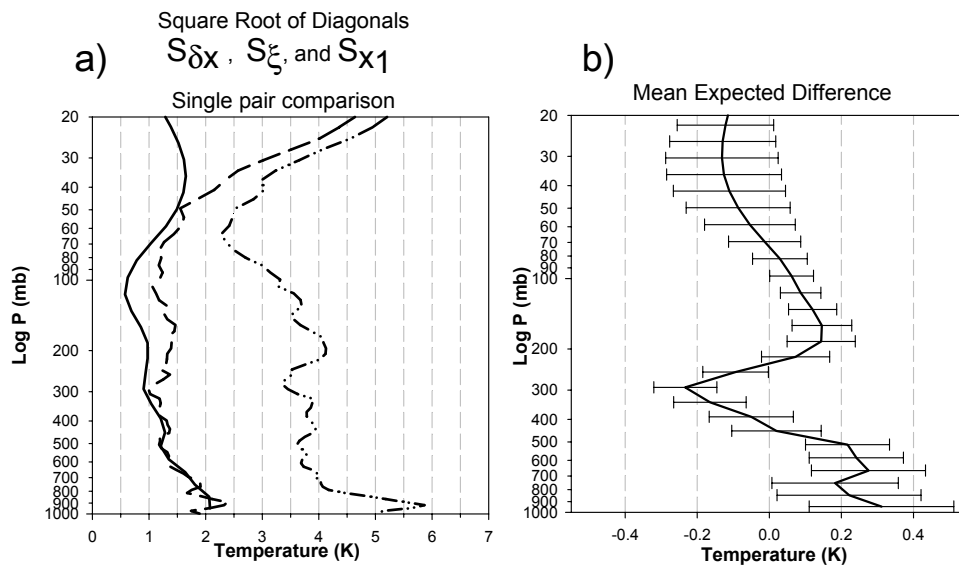


Fig. 5. a) Square root of diagonal elements of the covariance matrices: the solid line is a single pair satellite-radiosonde comparison error ( $\mathbf{S}_{\delta\hat{x}}$  matrix); the dashed line is a single pair radiosonde-radiosonde comparison error ( $\mathbf{S}_{\xi}$  matrix); the dash-dot-dot line is the rms temperature variation of the analyzed ensembles ( $\mathbf{S}_{x_2}$  matrix). b) The solid line is the estimation of the mean expected difference between the ensembles of satellite and radiosonde observations; the length of the error bars is the solid curve from a) divided by the square root of the ensemble sample size (107 profiles).

Looking at the plots in Figure 5 a), we see that in the presented case the effect of the averaging kernels is two fold: (i) smoothing per se removes structures of high frequency but small amplitude from the error pattern; (ii) above approximately 300 mb the estimated error of the satellite-radiosonde comparison is smaller than the error for the radiosonde-radiosonde. This is because we compare the satellite retrieval  $\hat{\mathbf{x}}_1$  with its simulation  $\hat{\mathbf{x}}_{12}$  (see Equation (8)).

In other words, the corresponding true profiles  $\mathbf{x}_1$  and  $\mathbf{x}_2$  contribute to the comparison with weights determined by the averaging kernels. Since the sensitivity of the retrievals (the peak amplitude of the averaging kernels) drops with altitude (see Figure 4), so does the  $\delta\hat{\mathbf{x}}$ .

#### 4. CONCLUSIONS AND DISCUSSION

From the results of the presented case study, we deduce the following conclusions: (i) A six-hour maximum time difference between satellite and radiosonde measurements corresponds to twelve-hour periods in radiosonde launches, a realistic scenario for many stations. (ii) The matrix  $\mathbf{S}_\xi$  can be stably inferred from real radiosonde profiles. (iii) For a single comparison, rms error caused by non-coincidence in time varies from 0.5 K at the 100 mb level to 2 K at the surface. By analyzing a sample of size  $N$ , the error can be reduced by a factor of  $1/\sqrt{N}$ ; thus we need  $N > 4$  to make the error less than 1 K at all practicable altitudes. For Earth system and climate studies, extended time intervals (a season and longer) present the most interest; hence, in the proposed work we can accumulate samples large enough to attain the required accuracy of the reference. In practice, the assumption of  $1/\sqrt{N}$  may not be always valid. Possible errors caused by a diurnal cycle and periodicity of the overpasses<sup>12,13</sup> will be addressed in future studies.

The presented Validation Assessment Model is a useful tool for validation planning and the interpretation of the results. It also can be used for the evaluation of consistency between the data from different sources that is necessary for building coherent and uniform data sets for Earth system and climate studies.

#### ACKNOWLEDGMENTS

This research was performed as part of an Internal Government Study awarded to the Space Dynamics Laboratory by the National Polar-orbiting Operational Environmental Satellite System / Integrated Program Office (NPOESS/IPO) (contract number 50-SPNA-1-00037).

#### REFERENCES

1. N. S. Pougatchev, G. W. Sachse, H. E. Fuelberg, C. P. Rinsland, R. B. Chatfield, V. S. Connors, N. B. Jones, J. Notholt, P. C. Novelli and H. G. Reichle, "Pacific Exploratory Mission-Tropics carbon monoxide measurements in historical context," *J. Geophys. Res.*, 104(26), 195- 207 (1999).
2. C. D. Rodgers and B. J. Connor, "Intercomparison of remote sounding instruments," *J. Geophys. Res.*, 108(D3), 4116, doi:10.1029/2002JD002299 (2003).
3. S. Migliorini, C. Piccolo and C. D. Rodgers, "Intercomparison of direct and indirect measurements: Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) versus sonde ozone profiles," *J. Geophys. Res.*, 109(D19316), doi:10.1029/2004JD004988 (2004).
4. Y. J. Meijer, R. J. van der A, R. F. van Oss, D. P. J. Swart, H. M. Kelder and P. V. Johnston, "Global Ozone Monitoring Experiment ozone profile characterization using interpretation tools and lidar measurements for intercomparison," *J. Geophys. Res.*, 108(D23), 4723, doi:10.1029/2003JD003498 (2003).
5. M. T. Chahine, T. S. Pagano, H. H. Auman, R. Atlas, C. Barnet, J. Blaisdell, L. Chen, M. Divakarla, E. J. Fetzer, M. Goldberg, C. Gautier, S. Granger, S. Hannon, F. W. Irion, R. Kakar, E. Kalnay, B. H. Lambrigsten, S-Y. Lee, J. Le Marshall, W. W. McMillan, L. McMillin, E. T. Olsen, H. Rivercomb, P. Rosenkranz, W. L. Smith, D. Staeln, L. L. Strow, J. Susskind, D. Tobin, W. Wolf, and L. Zhou, "AIRS Improving Weather Forecasting and Providing New Data on Greenhouse Gases," *Bulletin of American Meteorological Society*, v. 87, NO. 7, pp. 911 - 926, 2006.
6. D. C. Tobin, H. E. Revercomb, R. O. Knuteson, B. Lesht, L. L. Strow, S. E. Hannon, W. F. Feltz, L. Moy, E. J. Fetzer and T. Cress, "Atmospheric Radiation Measurement site atmospheric state best estimates for Atmospheric

- Infrared Sounder temperature and water vapor retrieval validation,” *J. Geophys. Res.*, 111(D09S14), doi:10.1029/2005JD006103 (2006).
7. M. Divakarla, C. Barnet, M. D. Goldberg, L. McMillin, E. S. Maddy, W. W. Wolf, L. Zhou and X. Liu, “Validation of Atmospheric Infrared Sounder temperature and water vapor retrievals with matched radiosonde measurements and forecasts,” *J. Geophys. Res.*, 111, doi:10.1029/2005JD006116 (2006).
  8. C. D. Rodgers, “Retrieval of Atmospheric Temperature and Composition From Remote Measurements of Thermal Radiation,” *Rev. Geophysics. and Space Phys.*, 14, 609-624 (1976).
  9. C. D. Rodgers, “Characterization and error analysis of profiles retrieved from remote sounding measurements,” *J. Geophys. Res.*, 95, 5587-5595 (1990).
  10. C. D. Rodgers, *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Scientific Publishing Co. Ltd., Singapore, 2000.
  11. D. J. Lary and L. Lait, “Using Probability Distribution Functions for Satellite Validation, Geoscience and Remote Sensing,” *IEEE Transactions*, 44(5), 1359 - 1366 (2006).
  12. J. G. Anderson, J. A. Dykema, R. M. Goody, H. Hu and D. B. Kirk-Davidoff, “Absolute, spectrally-resolved, thermal radiance: a benchmark for climate monitoring from space,” *J. Quant. Spectr. Rad. Transf.*, 85, 367 – 383 (2004).
  13. D. B. Kirk-Davidoff, R. M. Goody and J. G. Anderson, “Analysis of sampling errors for climate monitoring,” *J. Climate*, 18, 810 – 822 (2005).