4.6 ACCURATE REFERENCING OF SATELLITE DATA TO RADIOSONDE PROFILES; THEORY AND CASE STUDY

N. S. Pougatchev*[1], D. J. Seidel[2], G. E. Bingham[1], S. V. Kireev[3], and D. C. Tobin[4]
[1]Space Dynamics Laboratory, North Logan, Utah
[2]National Oceanic and Atmospheric Administration, Air Resources Laboratory, Silver Spring, Maryland
[3]Hampton University, Hampton, Virginia
[4]Space Science and Engineering Center, University of Wisconsin-Madison, Madison, Wisconsin

1. ABSTRACT

A linear mathematical model for accurate referencing of satellite data –atmospheric profile retrievals – to radiosonde profiles is presented. The model provides a theoretical basis and a practical tool for the assessment of the accuracy and precision of the referencing for a particular satellite system and ground site.

The satellite measurement and the profile used for reference are generally taken at a different time and space; moreover, they sample the atmosphere differently, i. e. they have different vertical sensitivity and resolution. All these factors cause apparent differences between the compared entities. To make the comparison of the satellite data accurate, the model accounts for those factors allowing one to separate them from the possible bias (accuracy) and noise (precision) of the satellite system.

To account for time and space differences the model uses statistical characteristics (mean value, covariance and correlation) of the ensembles of the true atmospheric states on which the satellite system and the system used for comparison perform the measurements. To reconcile the differences in vertical sensitivity and resolution the averaging kernel formalism is implemented.

For the case study the model has been applied to a set of radiosonde temperature profiles taken over the ARM Southern Great Plain site and simulated AIRS retrievals. It has been demonstrated how unaccounted temperature differences/errors between compared profiles depend on the time interval separating them. In this particular study, for two sets of profiles (107 profiles each) separated by less than six hours, the mean unaccounted error is within 0.3±0.2 K.

The model can be used for referencing the satellite data from instruments such as CrIMS, IASI, and AIRS to other data sets for use as Earth System or Climate Data Records (ESDRs or CDRs) as well as for assessment and interpretation of validation results when the previously mentioned sources of discrepancies are significant.

2. INTRODUCTION

To be suitable for use in Earth System and Climate studies, satellite data must be unified and coherent. One way to unify and make coherent the satellite data sets is to reference accurately all of them to the same global continuous set of measurements of a known quality and relation to the true state of the atmosphere and with proven value for climate studies. Data from the global radiosonde network have long been used for climate studies as well as for calibration and validation of retrievals of atmospheric temperature and water vapor profiles from satellite radiance observations (Kelly et al., 1991; Durre et al., 2005; Reale and Thorne, 2004). While some consider radiosonde profiles a “gold standard” against which to judge satellite products (Clery, 2006), it is well known that the temperature and humidity soundings contain errors (Nash et al., 2006) that are difficult to remove (Wang et al., 2002; Miloshevich et al., 2000, 2006; Sherwood et al., 2005). In many cases bias errors are not constant over time (Gaffen, 1994; Lanzante et al., 2003; Free et al. 2002). All of these problems with radiosonde data diminish the ease with which they can be used for calibration of satellite retrievals.

The issue of time-varying biases presents a particular problem for Climate Data Recerds (CDRs), where long-term stability of observational error is needed to ensure long-term climate change is well monitored. Nevertheless, large-scale features of climate variability (such as ENSO, response to volcanic eruptions, and the stratospheric quasi biennial oscillation) are well captured by radiosonde observations (Seidel et

* *Corresponding Author Address*: Nikita S. Pougatchev, Space Dynamics Laboratory, 1695 N. Research Park Way, North Logan, UT 84341; nikita.pougatchev@sdl.usu.edu

al., 2004). Radiosonde data, therefore, provide a good basis for comparison with satellite profile observations if proper care is taken to minimize data problems.

Radiosonde data were used for validation of the AIRS instrument on EOS AQUA platform, temperature and water vapor retrievals (Tobin et al, 2006, Divakarla et al., 2006, Miloshevich et al. 2006) Two different approaches have been used for the comparison. Thus, in the work of Tobin et al. (2006) the "best estimate" radiosonde data set for the Southern Great Plain 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 (Divakarla et al., 2006) 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 assessing measurement uncertainties indicated as a function of time and location for all of the data.

For accurate comparison of any data sets, one needs to know the errors of the compared quantities. This uncontroversial general statement in the case of comparison of atmospheric profiles obtained by different techniques raises 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 an averaging kernel formalism in application to CO and ozone retrieval validation (Pougatchev et al., 1999; Rodgers and Connor, 2003; Migliorini et al., 2004; Meijer et al., 2003), but the error caused by non-collocation was beyond the scope of the studies. For accurate comparison of remote sensing atmospheric profiles, all kind of errors should be taken into account, which leads us to two major requirements:

First, all significant Environmental Data Record (EDR) sources of error must be identified along with error propagation paths through the Instrument-Retrieval chain; this constitutes application of an End-to-End Error Model (E2EM) to evaluate the expected level of error in the EDR. Second, actual errors of satellite EDRs must be assessed through comparison with some relevant data set of known accuracy and precision. The comparison is subject to additional error caused by the difference between the compared measurement systems as well as by non-collocation of the data sets. To estimate and reduce the error, the EDR Assessment Model (EDRAM) has been developed. This model accounts for specific characteristics of the considered satellite instrument as well as a particular reference data set. In other words, the EDRAM allows individual analysis of a particular geographical location and vertical area of the atmosphere. The use of the EDRAM will enable global coverage, high accuracy and specificity of comparison.

Graphically, the concept of the work is illustrated by Figure 1.

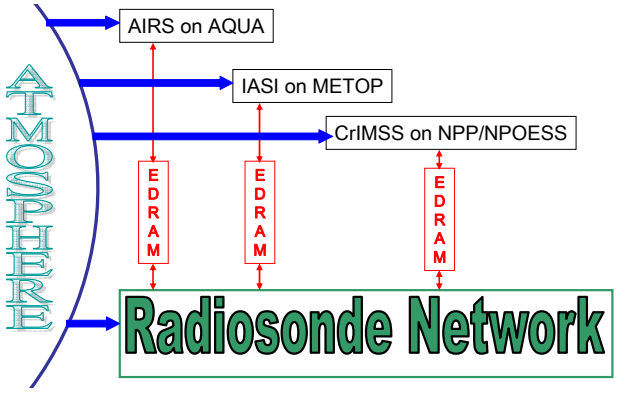

Figure 1. The Radiosonde Network serves as the common reference data set for three satellite systems. EDRAM is the unified tool for the referencing.

The EDRAM allows individual error analysis of a particular geographical location as well as a vertical area of the atmosphere.

Although in situ sensors provide valuable data for analysis of climate variability and change, satellite observations are the best source of globally complete and spatially consistent observations for climate studies. Interest in the vertical profile of temperature stems from its singular importance in climate change detection and attribution studies (Hansen et al., 1997; Santer et al., 1996; CCSP, 2006)) due to the unique fingerprints that different

climate forcings make on the vertical temperature profile. Controversy over the interpretation of satellite upper air temperature datasets (NRC, 2000; Wentz and Schabel, 2000; Fu et al., 2004) has figured prominently in scientific and policy debates, in large part because the true nature of the observational errors is not completely understood (CCSP, 2006; Thorne et al., 2005).

3. TECHNICAL APPROACH AND METHODOLOGY

3.1 END-TO-END MODEL FOR ATMOSPHERIC RETRIEVALS

In the context of this work, the E2EM for atmospheric profile retrieval is a system of linear equations that relates statistical characteristics (mean value and covariance) of errors originated in the Instrument-Retrieval chain to the statistical characteristics of the error in the retrieved profile or EDR. The E2EM consists of two different parts: (i) error analysis and characterization; and (ii) assessment of the actual error of real/proxy EDRs with the aid of EDRAM. The overall concept is presented in Figure 2.

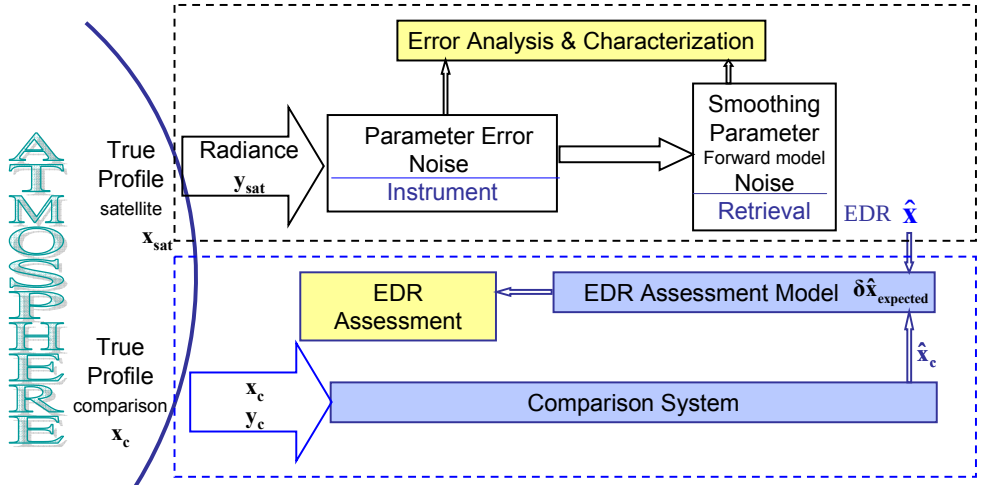

Figure 2. Overall concept of error modeling. The upper dashed block encompasses elements of error analysis and characterization; the lower one represents EDR assessment. For an explanation of the parameters, see sections 3.1.1 and 3.1.2

In this context, *characterization* shows how retrieval is related to the true state of the atmosphere, while *error analysis* reveals how various sources of errors propagate to the final product. This part of the E2EM is a mathematical representation of how uncertainties/errors of various origins starting at the entrance aperture (front end) propagate into the final products, in our case the vertical profile of atmospheric temperature and constituents or EDRs. The E2EM allows one to estimate the expected performance of the instrument. The EDRAM assesses actual performance of the measurement system while on orbit. The goal of both activities is the same: namely, to asses the system performance, just at different stages of development and operation. Therefore, both activities should be considered in the same methodological manner.

3.2 ERROR ANALYSIS AND CHARACTERIZATION

We will, for the most part, follow the methodology and notation developed by C. D. Rodgers (Rodgers, 1976, 1990, and 2000; Rodgers and Connor, 2003). Measured signal, states of the atmosphere, both true and retrieved, and instrument and model parameters will be presented by column vectors (bold lower case symbols) with corresponding covariance matrices (bold upper case symbols).

The equations and formulas we present below will be used in the practical error analysis, characterization, and assessment of the EDR retrievals.

The measured signal (in our case a set of spectrally and radiometrically calibrated radiances - a spectrum, $\{y_i\}$ i=1, 2,..., m) can be written in a form

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \mathbf{b}) + \boldsymbol{\varepsilon} \quad (1)$$

where *forward model* **F** describes the physics of the measurements; **x** is the unknown state (in our case, it can be the true state of the atmosphere) $\{x_i\}$ i=1, 2, ... , n; **b** is a vector of some other set of parameters that influence the measurements but are not included in **x (**e.g., instrument parameters etc., which we call forward model parameters).

In linear approximation the difference between the actually measured signal **y** and its expected value $\mathbf{y_a}$, in other words error in the measured signal, $\boldsymbol{\delta}\mathbf{y} \equiv \mathbf{y} - \mathbf{y_a}$ can be expressed as follows:

$$\begin{aligned} \boldsymbol{\delta}\mathbf{y} = {} & \mathbf{K_x}(\mathbf{x} - \mathbf{x_a}) && \text{state error} \\ & + \mathbf{K_b}(\mathbf{b} - \mathbf{b_a}) && \text{model parameter error} \\ & && (\text{instrument and atmosphere}) \quad (2) \\ & + \boldsymbol{\Delta}\mathbf{F}(\mathbf{x}, \mathbf{b}) && \text{forward model error} \\ & + \boldsymbol{\varepsilon} && \text{measurement noise} \end{aligned}$$

Rows of the **K**-matrices contain the weights with which variations (errors) of corresponding elements of state and forward model parameters contribute to the variation of signal (signal error). In atmospheric sounding, when **x** is the state of the atmosphere (vertical profile), the rows of the $\mathbf{K_x}$ matrix are called the *weighting function* or *Jacobians.*

Forward model error $\boldsymbol{\Delta}\mathbf{F}(\mathbf{x}, \mathbf{b})$ usually accounts for simplification of the radiative transfer for the sake of computational efficiency and is likely to be systematic.

In equation (2) the second and the fourth terms represent the contribution of errors associated with the instrument (instrument model parameter error and measurement noise) into the signal error. The instrument error model for the signal can be written as

$$\boldsymbol{\delta}\mathbf{y} = \mathbf{K_{bin}}\boldsymbol{\delta}\mathbf{b} + \boldsymbol{\varepsilon} \quad (3)$$

where $\boldsymbol{\delta}\mathbf{b} = (\mathbf{b_{in}} - \hat{\mathbf{b}}_{\mathbf{in}})$ and $\hat{\mathbf{b}}_{\mathbf{in}}$ is our best knowledge of the instrument parameters (their nominal value). In other words, error in the instrument parameter $\boldsymbol{\delta}\mathbf{b}$ propagates to $\boldsymbol{\delta}\mathbf{y}$ being amplified by weighting function matrix $\mathbf{K_{bin}}$. The covariance of the error associated with the instrument is

$$\mathbf{S}_{\boldsymbol{\delta}\mathbf{y}} = \mathbf{K}_{\boldsymbol{\delta}\mathbf{b}}\mathbf{S}_{\boldsymbol{\delta}\mathbf{b}}\mathbf{K}^{\mathbf{T}}_{\boldsymbol{\delta}\mathbf{b}} + \mathbf{S}_{\boldsymbol{\varepsilon}} \quad (4).$$

Actual values of **K**-matrices, $\mathbf{S}_{\boldsymbol{\varepsilon}}$, and $\mathbf{S}_{\boldsymbol{\delta}\mathbf{b}}$ are determined during the pre-launch error analysis and laboratory tests of the instrument. An example of real $\mathbf{S}_{\boldsymbol{\delta}\mathbf{y}}$ for the Short Wavelength IR (SWIR) band for the CrIS instrument is presented in Figure 3.

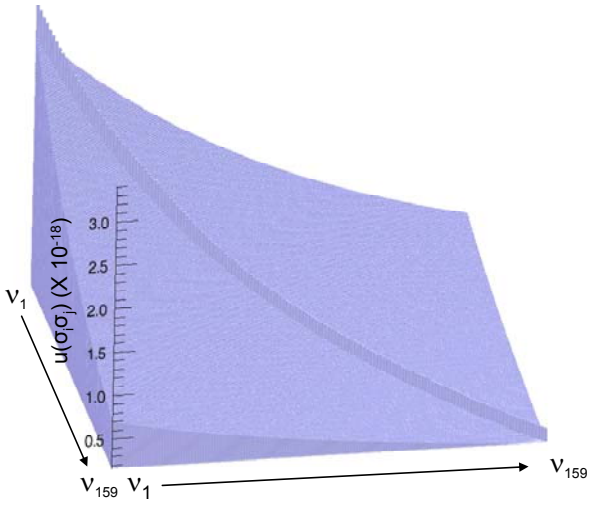

Figure 3. Covariance matrix for the CrIS SWIR signal. Detector shot noise and radiometric calibration error are taken into account. Non-zero off-diagonal elements are due to spectrally correlated radiometric calibration error.

A retrieval **r**(**y**) is a way of finding a state $\hat{\mathbf{x}}$ such that **F**($\hat{\mathbf{x}}$) is consistent with **y**. This leads us to a transfer function **r**

$$\hat{\mathbf{x}} = \mathbf{r}(\mathbf{F}(\mathbf{x}, \mathbf{b}) + \boldsymbol{\varepsilon}, \hat{\mathbf{b}}, \mathbf{x_a}, \mathbf{c}). \quad (5)$$

Equation (5) describes the operation of the entire observing system, including both the instrument and retrievals. The a priori state $\mathbf{x_a}$ and the set of parameters **c** are not included in the forward function but affect the retrieval and are subjects of uncertainty. Linearization with respect to **y** allows us to write a linear model for the error in retrieval $\hat{\mathbf{x}}$, i.e. the difference between the retrieved and true states (simplified Equation (3.16) in Rodgers (2000)).

$$\begin{aligned}\hat{\mathbf{x}}-\mathbf{x} &= (\mathbf{A}-\mathbf{I})(\mathbf{x}-\mathbf{x}_a) && \text{smoothing error}\\ &+\mathbf{G}_y\mathbf{K}_b\delta\mathbf{b} && \text{instrument error}\\ &+\mathbf{G}_y\Delta\mathbf{F}(\mathbf{x},\mathbf{b}) && \text{forward model error}\\ &+\mathbf{G}_y\boldsymbol{\varepsilon} && \text{retrieval noise}\end{aligned} \quad (6)$$

where $\mathbf{G}_y = \dfrac{\partial \mathbf{r}}{\partial \mathbf{y}}$ is the so-called *gain matrix* that shows sensitivity of the retrieval to measurement or, equivalently,what is the same, to measurement error. The *averaging kernel* matrix **A** characterizes the sensitivity of the retrieval to the true state

$$\mathbf{A} = \mathbf{G}_y\mathbf{K}_x = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{x}} \quad (7)$$

For illustration, Figure 4 presents averaging kernels for temperature retrievals as they would be made by the AIRS instrument .The width of the peaks can be used as a measure of the vertical resolution.

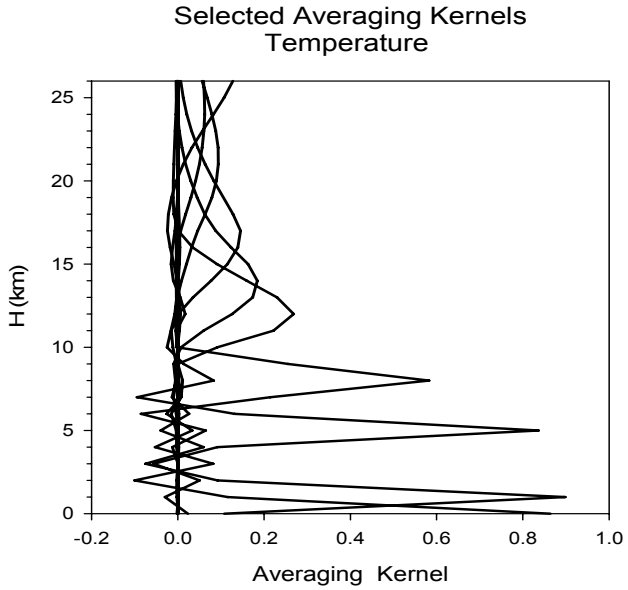

Figure 4. Selected averaging kernels for temperature retrievals in 1 km thick layers.

The practical value of Equation (6) is that it is applicable to any retrieval technique, not just optimal estimation as it is often misperceived.

The first term in (6) allows us to estimate the smoothing error of the retrieval, in other words error caused by finite *vertical resolution* of the observing system. For retrieval assessment, averaging kernels are absolutely necessary because they provide the means for comparison of measurements performed by systems with different vertical resolution.

Comparing (2) and (6) one can see that the error in the signal is "amplified" by the *gain matrix* $\mathbf{G}_y$ in the retrieval. Hence, the corresponding covariance matrices are:

$$\hat{\mathbf{S}}_b = \mathbf{G}_y\mathbf{K}_b\mathbf{S}_b\mathbf{K}_b^T\mathbf{G}_y^T \quad (8)$$

for instrument parameters error and

$$\hat{\mathbf{S}}_\varepsilon = \mathbf{G}_y\mathbf{S}_\varepsilon\mathbf{G}_y^T \quad (9)$$

for retrieval noise.

The model described above constitutes the basis for practical error analysis and characterization of a satellite-based measurement system. Equations (2) and (6) and associated formulas allow one to estimate expected errors in the retrieved profiles and their statistical characteristics.

3.3 EDR ASSESSMENT MODEL - EDRAM

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

$$\hat{\mathbf{x}} = \mathbf{r}(\mathbf{x}) + \mathbf{e} \quad (10)$$

where **r(x)** is a nominal retrieval without any errors in the measured signal or in the forward model and **e** represents retrieval errors caused by these factors. The error term can be characterized by its mean value $E\{\mathbf{e}\} = \boldsymbol{\Delta}$ (bias) and covariance $\mathbf{S}_e$ (retrieval noise). From the error analysis and characterization, we know the nominal retrieval **r** and a priori retrieval noise ${}^a\mathbf{S}_e$, and we presume that for a nominally performing system $\boldsymbol{\Delta} = \mathbf{0}$. The assessment activity yields an estimate of the actual values of $\boldsymbol{\Delta}$ and $\mathbf{S}_e$: $\hat{\boldsymbol{\Delta}} = E\{\hat{\mathbf{x}} - \mathbf{r}(\mathbf{x})\}$ and $\hat{\mathbf{S}}_e = E\{(\hat{\boldsymbol{\Delta}} - (\hat{\mathbf{x}} - \mathbf{r}(\mathbf{x}))(\hat{\boldsymbol{\Delta}} - (\hat{\mathbf{x}} - \mathbf{r}(\mathbf{x}))^T\}$.

Schematically, the EDR assessment process is illustrated by Figure 5.

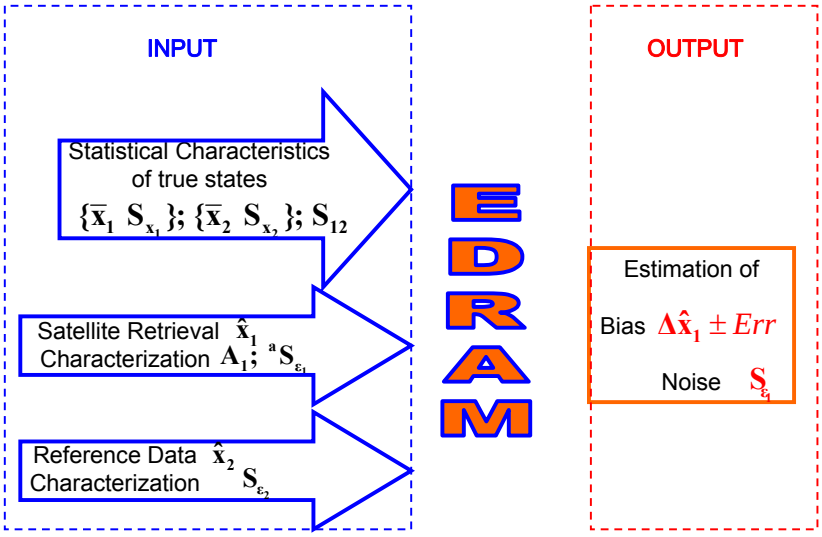

Figure 5. EDRAM Input and Output data.

The input of an assessment activity is a set of profiles of a known quality and known relation to the nominal retrieval **r** and the true state of the atmosphere **x**; in our case it is radiosonde data. The reference system samples the true state $\mathbf{x}_2$ of the atmosphere characterized by its mean value and covariance $\{\bar{\mathbf{x}}_2\ \mathbf{S}_{x_2}\}$ at a time and location different from when and where the satellite system makes its own observation of $\mathbf{x}_1$ with $\{\bar{\mathbf{x}}_1\ \mathbf{S}_{x_1}\}$, respectively. The true states are correlated with cross-covariance $\mathbf{S}_{12}$. The satellite and reference systems have different characteristics, i.e., vertical resolution (averaging kernel **A**) and noise level ($\mathbf{S}_{\varepsilon_2}$ and $\mathbf{S}_{\varepsilon_2}$, respectively). All these factors cause an apparent difference between the data to be compared. The EDRAM makes the comparison accurate by estimating this difference. In its general form, the EDRAM provides the tool for accurate comparison of atmospheric profiles of any origin. Here we consider a particular case when satellite retrievals are referenced to radiosonde measurements. Because of the novelty of the approach we will derive the basic formulas that we will use for the practical data analysis.

In linear approximation the retrieved and radiosonde profiles can be presented as follows:

$$\begin{aligned}\hat{\mathbf{x}}_1 &= \mathbf{x}_a + \mathbf{A}(\mathbf{x}_1 - \mathbf{x}_a) + \Delta\hat{\mathbf{x}} + \boldsymbol{\varepsilon}_1 \\ \hat{\mathbf{x}}_2 &= \mathbf{x}_2 + \boldsymbol{\varepsilon}_2\end{aligned} \qquad (11)$$

where $\hat{\mathbf{x}}_1$, $\hat{\mathbf{x}}_2$, $\mathbf{x}_1$, and $\mathbf{x}_2$ are retrieved, radiosonde, and true profiles; and $\mathbf{A}$, $\mathbf{x}_a$, $\Delta\hat{\mathbf{x}}$, $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ are the corresponding averaging kernel, a priori profile, retrieval bias, and retrieval and measurement noise, respectively. Index 1 is assigned to the terms related to the satellite system and index 2 to the radiosondes.

The true states $\mathbf{x}_1$ and $\mathbf{x}_2$ are functions of coordinate **z** and time *t* $\mathbf{x}_1 = \mathbf{x}(\mathbf{z}_1, t_1)$ $\mathbf{x}_2 = \mathbf{x}(\mathbf{z}_2, t_2)$. True mean value of the ensemble of states is $\bar{\mathbf{x}}_1 = E\{\mathbf{x}_1(\mathbf{z}_1, t_1)\}$ and $\bar{\mathbf{x}}_2 = E\{\mathbf{x}_2(\mathbf{z}_2, t_2)\}$. The variations of the states about their means $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ are characterized by their auto-covariances $\mathbf{S}_{x_1} = E\{(\mathbf{x}_1 - \bar{\mathbf{x}}_1)(\mathbf{x}_1 - \bar{\mathbf{x}}_1)^T\}$ and $\mathbf{S}_{x_2} = E\{(\mathbf{x}_2 - \bar{\mathbf{x}}_2)(\mathbf{x}_2 - \bar{\mathbf{x}}_2)^T\}$. Correlation between the true states $\mathbf{x}_1$ and $\mathbf{x}_2$ can be characterized by cross-covariances $\mathbf{S}_{12} = E\{(\mathbf{x}_1 - \bar{\mathbf{x}}_1)(\mathbf{x}_2 - \bar{\mathbf{x}}_2)^T\}$ and $\mathbf{S}_{21} = E\{(\mathbf{x}_2 - \bar{\mathbf{x}}_2)(\mathbf{x}_1 - \bar{\mathbf{x}}_1)^T\}$ for the covariances $\mathbf{S}_{12} = \mathbf{S}_{21}^{\mathrm{T}}$ and $\mathbf{S}_{12} = \mathbf{S}_{21} = \mathbf{S}_{x_1} = \mathbf{S}_{x_2}$ when $t_1 \equiv t_2$ and $\mathbf{z}_1 \equiv \mathbf{z}_2$.

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

$$\delta\mathbf{x}_1 = \mathbf{B}\delta\mathbf{x}_2 + \xi \tag{12}$$

where $\boldsymbol{\xi}$ is random with $\bar{\xi} = 0$, auto-covariance $\mathbf{S}_\xi$, and cross-covariance $\mathbf{S}_{\mathbf{x}_1\xi} = E\{(\mathbf{x}_1 - \xi)(\mathbf{x}_1 - \xi)^{\mathrm{T}}\} = 0.$ In other words, the variation $\delta\mathbf{x}_1$ at the satellite site can be decomposed in two parts: correlated $\mathbf{B}\delta\mathbf{x}_2$ that can be derived from the measurement of $\mathbf{x_2}$, and uncorrelated - error - ξ. Given matrices $\mathbf{S}_{\mathbf{x}_1}$, $\mathbf{S}_{\mathbf{x}_2}$, and $\mathbf{S}_{12}$ we can calculate matrices $\mathbf{B}$ and $\mathbf{S}_\xi$

$$\mathbf{B} = \mathbf{S}_{12}\mathbf{S}_{\mathbf{x}_2}^{-1} \tag{13}$$

$$\mathbf{S}_\xi = \mathbf{B}\mathbf{S}_{\mathbf{x}_2}\mathbf{B}^{\mathrm{T}} - \mathbf{S}_{\mathbf{x}_1} \tag{14}$$

Because of the finite vertical resolution of the satellite retrievals and their nonuniform vertical sensitivity it makes sense to degrade vertical resolution of the radiosonde profiles to the satellite level and perform comparison only in those vertical areas of the atmosphere where the satellite instrument has sensitivity. Following the approach from Rodgers and Connor (2003) and Equations (11) and (12), we simulate retrieval $\hat{\mathbf{x}}_1$ with $\hat{\mathbf{x}}_2$

$$\hat{\mathbf{x}}_{12} = \mathbf{A}\mathbf{B}\hat{\mathbf{x}}_2 = \mathbf{A}\mathbf{B}\mathbf{x}_2 + \mathbf{A}\mathbf{B}\boldsymbol{\varepsilon}_2 \tag{15}$$

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

$$\begin{aligned}\delta\hat{\mathbf{x}} &\equiv \hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_{12} \\ &= (\mathbf{I} - \mathbf{A})\mathbf{x}_a + \mathbf{A}\bar{\mathbf{x}}_1 - \mathbf{A}\mathbf{B}\bar{\mathbf{x}}_2 \\ &+ \mathbf{A}(\mathbf{B}\delta\mathbf{x}_2 + \xi) - \mathbf{A}\mathbf{B}\delta\mathbf{x}_2 \\ &+ (\boldsymbol{\varepsilon}_1 - \mathbf{A}\mathbf{B}\boldsymbol{\varepsilon}_2) \\ &+ \Delta\hat{\mathbf{x}}_1\end{aligned} \tag{16}$$

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

$$\begin{aligned}\delta\bar{\hat{\mathbf{x}}} &\equiv \bar{\hat{\mathbf{x}}}_1 - \bar{\hat{\mathbf{x}}}_{12} \\ &= (\mathbf{I} - \mathbf{A})\mathbf{x}_a + \mathbf{A}\bar{\mathbf{x}}_1 - \mathbf{A}\mathbf{B}\bar{\mathbf{x}}_2 \\ &+ \Delta\hat{\mathbf{x}}_1\end{aligned} \tag{17}$$

Then consider

$$\Delta\hat{\mathbf{x}}_1 = \delta\bar{\hat{\mathbf{x}}} - {}^{\mathrm{e}}\delta\bar{\hat{\mathbf{x}}} \tag{18}$$

where ${}^{\mathrm{e}}\delta\bar{\hat{\mathbf{x}}}$ is the *expected difference.*

$${}^{\mathrm{e}}\delta\bar{\hat{\mathbf{x}}} = (\mathbf{I} - \mathbf{A})\mathbf{x}_a + \mathbf{A}\bar{\mathbf{x}}_1 - \mathbf{A}\mathbf{B}\bar{\mathbf{x}}_2 \tag{19}$$

The covariance of $\delta\hat{\mathbf{x}}$ about its mean - $\mathbf{S}_{\delta\hat{\mathbf{x}}}$ is

$$\mathbf{S}_{\delta\hat{\mathbf{x}}} = \mathbf{A}\mathbf{S}_\xi\mathbf{A}^{\mathrm{T}} + \mathbf{S}_{\varepsilon_1} + (\mathbf{A}\mathbf{B})\mathbf{S}_{\varepsilon_2}(\mathbf{A}\mathbf{B})^{\mathrm{T}} \tag{20}$$

where $\mathbf{S}_{\varepsilon_1}$ and $\mathbf{S}_{\varepsilon_2}$ are characteristics of the retrieval and radiosonde noise.

It is important to notice in this context that ${}^{\mathrm{e}}\delta\bar{\hat{\mathbf{x}}}$ *is not the error but instead represents the expected difference between nominally performing measurement systems.* The purpose of the EDR assessment 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 (18) is limited by the accuracy of our a priori knowledge of ${}^{\mathrm{e}}\delta\bar{\hat{\mathbf{x}}}$ and the bias of reference system ${}^{\mathrm{a}}\Delta\hat{\mathbf{x}}_2$. In practice, we know the mean of the states $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ with some uncertainties characterized by the covariances $\tilde{\mathbf{S}}_{\bar{\mathbf{x}}_1}$, $\tilde{\mathbf{S}}_{\bar{\mathbf{x}}_2}$ and $\tilde{\mathbf{S}}_{\bar{\mathbf{x}}_1\bar{\mathbf{x}}_2} = 0$. The uncertainty associated with the assumption that ${}^{\mathrm{a}}\Delta\hat{\mathbf{x}}_2 = \mathbf{0}$ is characterized by covariance $\mathbf{S}_{\Delta\hat{\mathbf{x}}_2}$. That results in additional error of ${}^{\mathrm{e}}\delta\bar{\hat{\mathbf{x}}}$ with covariance $\tilde{\mathbf{S}}$

$$\tilde{\mathbf{S}} = \mathbf{A}\tilde{\mathbf{S}}_{\bar{\mathbf{x}}_1}\mathbf{A}^{\mathrm{T}} + (\mathbf{A}\mathbf{B})(\tilde{\mathbf{S}}_{\bar{\mathbf{x}}_2} + \mathbf{S}_{\Delta\hat{\mathbf{x}}_2})(\mathbf{A}\mathbf{B})^{\mathrm{T}} \tag{21}$$

In the assessment process, a set of $\{\delta_i\hat{\mathbf{x}}\}$ $i = 1, 2, ..., N$ 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 ${}^{\mathrm{s}}\delta\bar{\hat{\mathbf{x}}}$ relates to the mean of the ensemble $\delta\bar{\hat{\mathbf{x}}}$ as ${}^{\mathrm{s}}\delta\bar{\hat{\mathbf{x}}} = \delta\bar{\hat{\mathbf{x}}} + \boldsymbol{\varepsilon}_{\mathbf{s}}$ where $\boldsymbol{\varepsilon}_{\mathbf{s}}$ 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 $\mathbf{S}_{\varepsilon_s} = N^{-1}\mathbf{S}_{\delta\hat{\mathbf{x}}}$.

Then the estimation of the bias of the satellite system is

$$ {}^{s}\Delta\hat{\mathbf{x}}_1 = {}^{s}\delta\bar{\hat{\mathbf{x}}} - {}^{e}\delta\bar{\hat{\mathbf{x}}} \qquad (22), $$

and the covariance of the estimate is

$$ \mathbf{S}_{s_{\Delta\hat{\mathbf{x}}}} = \mathbf{S}_{\varepsilon_s} + \tilde{\mathbf{S}} \qquad (23). $$

The retrieval noise $\mathbf{S}_{\varepsilon_1}$ of the satellite system can be estimated based on analysis of the measurements by the satellite system on the ensemble of states $\mathbf{x_1}$ with $\mathbf{S}_{\mathbf{x}_1}$ such that $\mathbf{A}\mathbf{S}_{\mathbf{x}_1}\mathbf{A}^{\mathbf{T}} \approx {}^{a}\mathbf{S}_{\varepsilon_1}$.

For evaluating a particular reference data source/site, one can use the following relations:

$$ {}^{a}\mathbf{S}_{\delta\hat{\mathbf{x}}} = \mathbf{A}\mathbf{S}_{\xi}\mathbf{A}^{\mathbf{T}} + {}^{a}\mathbf{S}_{\varepsilon_1} + (\mathbf{AB})\mathbf{S}_{\varepsilon_2}(\mathbf{AB})^{\mathbf{T}} \qquad (24) $$

$$ \mathbf{S}_{\varepsilon_s} \approx {}^{a}\mathbf{S}_{\delta\hat{\mathbf{x}}} / N \qquad (25) $$

$$ \mathbf{S}_{s_{\Delta\hat{\mathbf{x}}}} = {}^{a}\mathbf{S}_{\delta\hat{\mathbf{x}}} / N + \tilde{\mathbf{S}} \qquad (26) $$

Equations (24) to (26) give us an estimate of the attainable assessment accuracy 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 assessment.

4. RESULTS OF CASE STUDY

This section demonstrates the practicability of the theoretical basis presented in previous section. 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. 2006). 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 τ hours, where τ =3, 6, 12, 24, 48, 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 6. As one can see, there is noticeable decreasing with the τ correlation between the ensembles. Given $\mathbf{S}_{\mathbf{x}_2}$ and $\mathbf{S_{12}}$ using Equations (13) and (14) we calculated $\mathbf{S}_{\xi}$, which characterizes the uncorrelated temperature difference between $\mathbf{x_1}$ and $\mathbf{x_2}$. For the comparison of two radiosonde profiles, the square root of the diagonal elements of $\mathbf{S}_{\xi}$ can be interpreted as *rms* error caused by non-coincidence of launch times. Plots for error are displayed in Figure 7.

One of the interpretations of the results in Figure 7 is that given a particular temperature profile $\mathbf{x_2}$ with associated statistical characteristics $\mathbf{S}_{\mathbf{x}_2}, \mathbf{S_{12}}, \bar{\mathbf{x}}_2$, and $\bar{\mathbf{x}}_1$, we can estimate profile $\mathbf{x_1}$ separated in time by less than τ with *rms* error indicated by the corresponding curve in Figure 7.

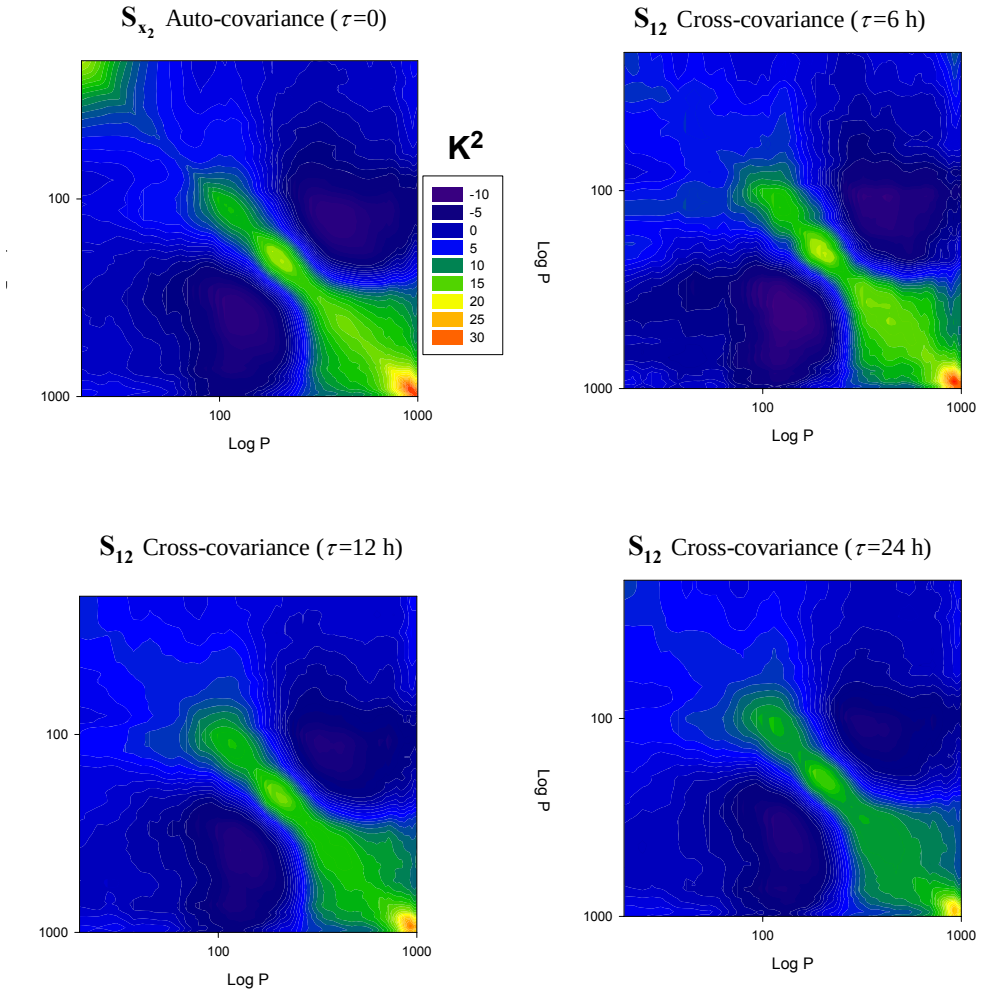

Figure. 6. Auto-covariance and selected cross-covariance matrices.

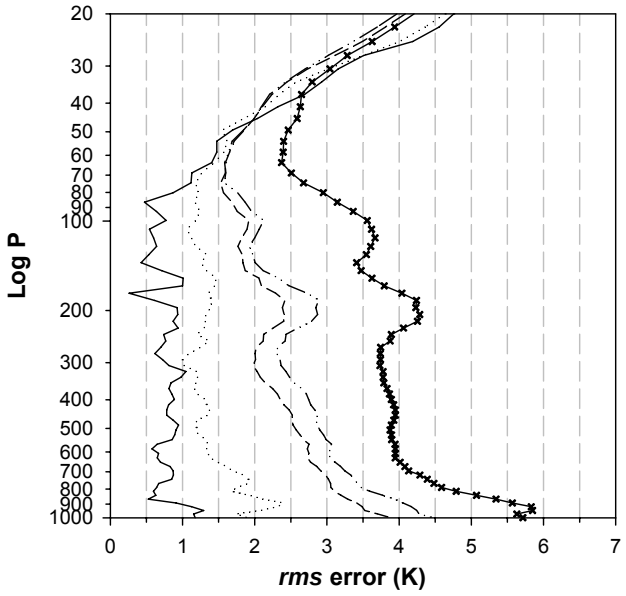

Figure. 7. The square root of $\mathbf{S}_{\xi}$ diagonal elements (rms error for comparison of a single pair of radiosondes) for different τ . The solid line is for $\tau = 3$ h; the dotted line is for $\tau = 6$ h; the dashed line is for $\tau = 12$ h; the dash-dot-dot line is for $\tau = 24$ h; and the solid-with-crosses is for $\tau = \infty$, i.e. no correlation at all, validation against historical records.

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 $\mathbf{x_a=0}$, we estimated the mean expected difference ${}^{e}\overline{\delta\hat{\mathbf{x}}}$ with the associated error (see Equation (25)). The results are presented in Figure 8.

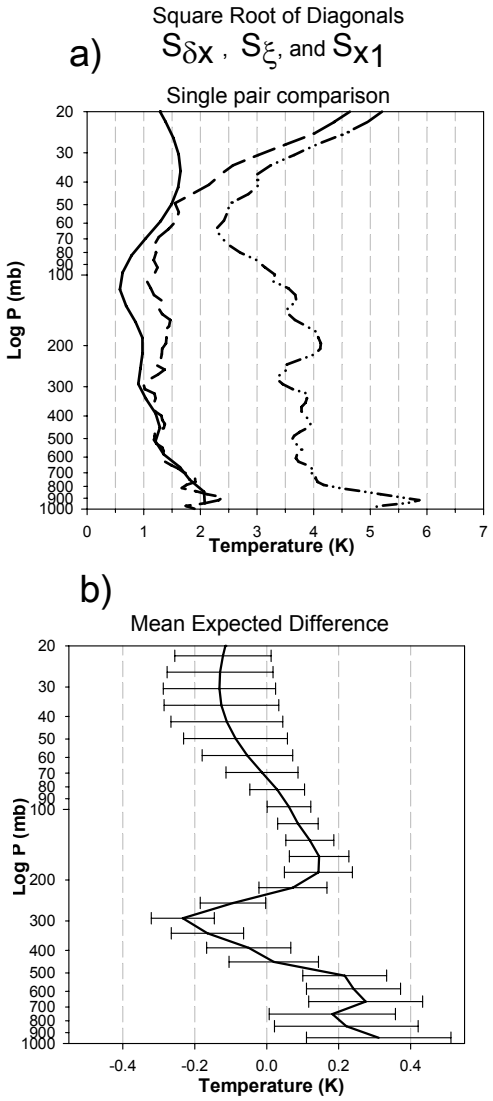

Figure. 8. a) Square root of diagonal elements of the covariance matrices: the solid line is a single pair satellite-radiosonde comparison error ($\mathbf{S}_{\delta\hat{\mathbf{x}}}$ matrix); the dashed line is a single pair radiosonde-radiosonde comparison error ($\mathbf{S}_{\xi}$ matrix); the dash-dot-dot line is the rms temperature variation of the analyzed ensembles ($\mathbf{S}_{\mathbf{x}_2}$ 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 8 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 difference $\delta\hat{\mathbf{x}}$.

The plot of non-coincidence error averaged between 800 and 800 mb is presented in Figure 9.

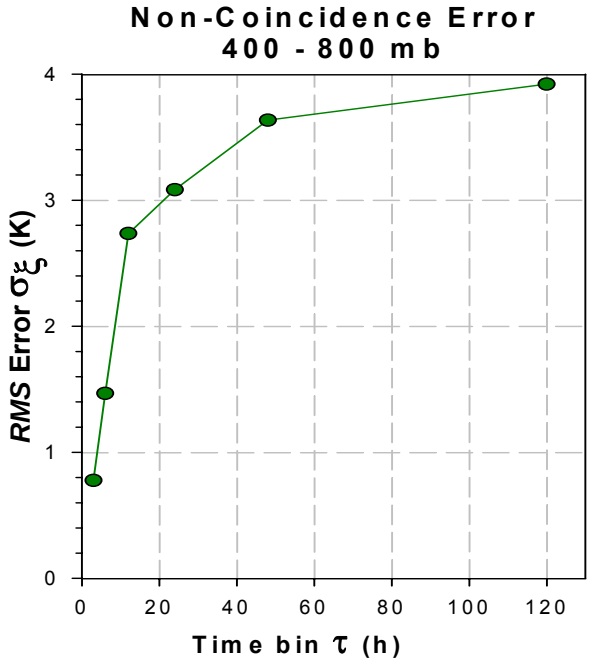

Figure 9. Non-coincidence temperature rms error averaged between 800 – 400 mb.

Initial part of the curve ($\tau \le 6h$) demonstrates good linear behavior with the slope approximately 0.2 K h^{-1}. Chahine et al, 2006 estimated global non-coincidence surface air temperature error based on analysis of the AIRS retrievals for the $\pm 3h$ and ± 100 km vicinity as 0.8 K. As one can see that the estimates obtained by completely different techniques are in a good agreement.

5. 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, 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 (Anderson et al., 2004, Kirk-Davidoff et al., 2005) will be addressed in future studies.

The presented EDR Assessment Model is a useful tool 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. It also can be used for validation planning and the interpretation of the results.

6. REFERENCES

Anderson, J. G., 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, 2004, J. Quant. Specrtr. Rad. Transf., 85, 367 - 383.

Chahine, M. T., et al, (2006), AIRS Improving Weather Forecasting and Proving New Data on Greenhouse Gases, Bulletin of the American Meteorological Society, 87, 911 – 926.

Clery, D. 2006, Technique From Outer Space Takes On Earth Observation, Science, 312, 48 – 49, DOI: 10.1126/science.312.5770.48

CCSP, 2006: *Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences*. Thomas R. Karl, Susan J. Hassol, Christopher D. Miller, and William L. Murray, editors, 2006. A Report by the Climate Change Science Program and the Subcommittee on Global Change Research, Washington, DC.

Divakarla, M., C. Barnet, M. D. Goldberg, L. McMillin, E. S. Maddy, W. W. Wolf, L. Zhou, and X. Liu (2006), 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.

Durre, I., T.Reale, D. Carlson, J. Christy, M. Uddstrom, M. Gelman, P.Thorne, 2005: Improving the Usefulness of Operational Radiosonde Data, Bull. Amer. Meteorol. Soc., 86, 411-416.

Durre, I., R.S. Vose and D.B. Wuertz. 2006: Overview of the Integrated Global Radiosonde Archive. J. Climate, 19, 53-68.

Free, M., I. Durre, E. Aguilar, D. Seidel, T.C. Peterson, R.E. Eskridge, J.K. Luers, D. Parker, M. Gordon, J. Lanzante, S. Klein, J. Christy, S. Schroeder, B. Soden, and L.M. McMillin, 2002: Creating climate reference datasets: CARDS workshop on adjusting radiosonde temperature data for climate monitoring. Bull. Amer. Meteor. Soc., 83, 891-899.

Fu, Q., C.M. Johanson, S.G. Warren, and D.J. Seidel, 2004: Contribution of stratospheric cooling to satellite-inferred tropospheric temperature trends, Nature, 429, 55-58: doi: 10.103/nature02524.

Gaffen, D.J., 1994: Temporal inhomogeneities in radiosonde temperature records. J. Geophys. Res., 99, 3667-3676.

Hansen, J., M. Sato, R. Reudy, A. Lacis, K. Asamoah, K. Beckford, S. Borenstein, E. Brown, B. Cairns, B. Carlson, B. Curran, S. de Castro, L. Druyan, P. Etwarrow, T. Ferede, M. Fox, D. Gaffen, J. Glascoe, H. Gordon, S. Hollandsworth, X. Jiang, C. Johnson, N. Lawrence, J. Lean, J. Lerner, K. Lo., J. Logan, A. Luckett, M.P. McCormick, R. McPeters, R. Miller, P. Minnis, I. Ramberran, G. Russell, P. Russell, P. Stone, I. Tegen, S. Thomas, L. Thomason, A. Thompson, J. Wilder, R. Willson, J. Zawodny, 1997: Forcings and chaos in interannual to decadal climate change. J. Geophys. Res., 102, 25679-25720.

Kelly, G., E. Andersson, A. Hollingsworth, P. Lönnberg, J. Pailleux, and Z. Zhang, 1991: Quality control of operational physical retrievals of satellite sounding data. *Mon. Wea. Rev.,* **119,** 1866–1880.

Kirk-Davidoff, D. B., R. M. Goddy, and J. G. Anderson, 2005, Analysis of sampling errors for climate monitoring, J. Climate, 18, 810 - 822.

Lanzante, J.R., S.A. Klein, and D.J. Seidel, 2003: Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology. J. Climate, 16, 224-240.

D. J. Lary and L. Lait, (2006), "Using Probability Distribution Functions for Satellite Validation, Geoscience and Remote Sensing," IEEE Transactions, 44(5), 1359 - 1366.

Meijer, Y. J., R. J. van der A, R. F. van Oss, D. P. J. Swart, H. M. Kelder, P. V. Johnston, Global Ozone Monitoring Experiment ozone profile characterization using interpretation tools and lidar measurements for intercomparison, J. Geophys. Res., 108, NO. D23, 4723, doi:10.1029/2003JD003498, 2003.

Migliorini, S., C. Piccolo, 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.

Miloshevich, L.M., H. Voemel, A. Paukkunen, A.J. Heymsfield, and S.J. Oltmans, Characterization and correction of relative humidity measurements from Vaisala RS80-A radiosondes at cold temperatures. *J. Atmos. Oceanic Technol.,* 18, 135-156, 2000.

Miloshevich, L. M., H. Voemel, D. Whiteman, B. Lesht, F. J. Schmidlin, and F. Russo (2006), Absolute accuracy of water vapor measurements from six operational radiosonde types launched during AWEX-G and implications for AIRS validation, *J. Geophys. Res.*, *111*, doi:10.1029/2005JD006083, 2006.

Nash, J., R. Smout, T. Oakley, B. Pathack and S. Kurnosenko, 2005: WMO intercomparison of high quality radiosonde systems- Mauritius, 2-25 February 2005, Final Report.

NRC, Reconciling Observations of Global Temperature Change, Panel on Reconciling Temperature Observations, National Research Council, 104 pages, 2000.

Pougatchev, N. S., 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 -26,207, 1999.

Reale, A.L and Peter Thorne, 2004: Satellite upper air network (SUAN). Proceedings of SPIE Volume: 5548, Atmospheric and Environmental Remote Sensing Data Processing and Utilization, p 128-140.

Reale, A.L Satellite Upper Air Network and the Climate Retrieval Problem. Presented at Workshop on Long-term Temperature Trends, UKMO, Exeter, 13-19 September, 2004.

Rodgers, C. D., Retrieval of Atmospheric Temperature and Composition From Remote Measurements of Thermal Radiation, Rev. Geophys. and Space Phys., 14, p609-624, 1976.

Rodgers, C. D., Characterization and error analysis of profiles retrieved from remote sounding measurements, J. Geoph. Res., 95, 5587-5595, 1990.

Rodgers, C.D., Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific Publishing Co. Ltd., 2000.

Rodgers, C. D., and B. J. Connor, Intercomparison of remote sounding instruments, *J. Geophys. Res.*, *108*(D3), 4116, doi:10.1029/2002JD002299, 2003.

Santer, B.D., T.M.L. Wigley, T.P. Barnett, and E. Anyamba, 1996: Detection of climate change, and attribution of causes. In *Climate Change 1995: The Science of Climate Change* , edited by J.T. Houghton, L.G. Meira Filho, B.A. Callander, N. Harris, A. Kattenberg and K. Maskell, Cambridge University Press, Cambridge, 407-443.

Seidel, D.J., J.K. Angell, J. Christy, M. Free, S.A. Klein, J.R. Lanzante, C. Mears, D. Parker, M. Schabel, R. Spencer, A. Sterin, P. Thorne, and F. Wentz, 2004: Uncertainty in signals of large-scale climate variations in radiosonde and satellite upper-air temperature datasets, J. Climate, 17, 2225-2240.

Sherwood, S. C., J. R. Lanzante and C. L. Meyer (2005), Radiosonde daytime biases and late 20th century warming, Science, 309, 1556-1559.

Snell Hilary E., Richard Lynch, and Jean-Luc Moncet, Retrieval Algorithm for NPOESS-CrIMSS and future operational sounders, 13th Conference on Satellite Meteorology and Oceanography. Retrieved July 11, 2006, from http://ams.confex.com/ams/13SATMET/techprogram/paper_78940.htm

Thorne, P.W., Parker, D.E., Christy, J.R. and Mears, C.A., 2005c: Uncertainties in climate trends: lessons from upper-air temperature records. Bull. Amer. Meteor. Soc., 86, 1437-1442.

Tobin, D.C., 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 (2006), 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.

Wang, J., H. L. Cole, D. J. Carlson, E. R. Miller, K. Beierle, A. Paukkunen, and T. K. Laine, 2002: Corrections of humidity measurement errors from the Vaisala RS80 radiosonde - Application to TOGA_COARE data. J. Atmos. Oceanic Technol., Vol. 19, 981-1002.

Wentz, F.J., and M. Schabel, 2000: Precise climate monitoring using complementary satellite data sets, Nature, 403, 414 - 416.