

Article

Towards a Data-Derived Observation Error Covariance Matrix for Satellite Measurements

Yan-An Liu ^{1,2} , Zhenglong Li ³ and Melin Huang ^{4,*}

¹ Key Laboratory of Geographic Information Science, Ministry of Education, East China Normal University, Shanghai 200241, China

² School of Geographic Sciences, East China Normal University, Shanghai 200241, China

³ Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin-Madison, Madison, WI 53706, USA

⁴ Space Science and Engineering Center, University of Wisconsin-Madison, Madison, WI 53706, USA

* Correspondence: phmelin@snolab.ca

Received: 15 May 2019; Accepted: 24 July 2019; Published: 27 July 2019



Abstract: The observation error covariance (R) matrix is a key component in the data assimilation (DA) process for retrieval of atmospheric state parameters (ASPs), also impacting the subsequent numerical weather forecast. However, one commonly used type of R matrix depends on instrument noise, which contravenes reality because the retrieved ASPs would depend on the instrument used. Other types of R matrix rely on the observation operator (H), analyzed state (x_a), background error covariance (B) matrix or the background state (x_b), and the selected forecast ensemble. All these dependences reduce the representativeness of the R matrix, since the correctness of H needs verification and no true values exist for x_a or x_b . As such, a better method to correctly specify the R matrix is needed. Through the physical mechanism occurring between incident radiation and particles in the atmosphere, which complies with the phenomena of energy absorption and emission, correlations among bands or channels in a detected atmospheric radiance spectrum occur. This paper thus proposes a data-derived R matrix based on a large number (N) of detected atmospheric radiance spectra constructed from N real-time measurements, where N real-time measurements can be acquired by staring at some observation location of interest during a short amount of time. This data-derived R matrix for satellite radiance observations does not rely on any assumed quantities and is unambiguous. Technically, recording N real-time measurements is achievable by modifying the trigger configuration of data recording from ground.

Keywords: atmospheric state parameters (ASPs); background matrix (B matrix); data assimilation (DA); inter-channel error correlations; observation error covariance matrix (R matrix)

1. Introduction

Modern numerical weather prediction (NWP) systems require accurate determination of the initial atmospheric state to produce a reliable weather forecast. This is achieved through data assimilation (DA). The DA attempts to blend two sources of information, the observations and the background fields, which are weighted by their respective error statistics. Although many types of observations can be assimilated in NWP systems, we focused on satellite observations, which have been found to have the largest impact of all the different observation types [1]. The observation error covariance matrix (R matrix) describes the deviation of observed radiances from true (or expected) radiances, and the background error covariance matrix (B matrix) describes the difference between the background state and the true atmospheric state. Ideally, R and B matrices should be independent of each other. Unfortunately, the design of most current DA systems forces the two to interact with each other to some

extent (i.e., through iterative optimization) [2]. As a result, analysis based on the DA process highly depends on R and B matrices. However, the R and B matrices are not perfectly known, contributing to the uncertainty of subsequent weather forecasts.

While forecast is affected by analysis, it is also affected by the forecast model. This paper focuses on DA, and has little to do with forecast. Beside R and B matrices, the analysis also depends on the DA methodology. For global NWP, DA methods started as simple horizontal interpolation methods [3]. Later, more advanced approaches were developed, such as three-dimensional (3D) or four-dimensional (4D) with a multivariate relationship [4]. For example, 3D and 4D variational methods were created to minimize the cost function to determine the optimal analyzed state based on a priori state, observations, and the prescribed Gaussian uncertainty statistics for the background and observations [5]; various forms of ensemble Kalman filters were introduced to consider the flow-dependent forecast (i.e., background) error covariance in the DA process [6], since background errors are expected to be flow-dependent on weather conditions. Later, hybrid approaches combining variational and ensemble DA methods were proposed [7,8], where an ensemble was used to build the initial background error covariances followed by evolving or updating the background error covariances by the NWP forecast model at the end of the DA process. Hybrid DA methods have become more popular and been employed at several NWP centers and groups [9–22]. More information on the formulation of various DA methods can be found in Gustafsson et al. [23] and Bannister [24].

The DA methods are processed under some assumptions when minimizing the cost function. In practical applications, however, these assumptions are difficult to justify [25]. One assumption that plays a critical role in DA process is about the B matrix (see review by Bannister [24] and references therein). If variational DA methods are used in operations, the B matrix is assumed to be static for a set period of time [26]. If ensemble-variational DA methods are employed in operations, the B matrix is assumed to be flow-dependent and to evolve with NWP forecast during the time window. In general, the B matrix is estimated from forecast ensemble. However, it is not clear how much the variation within the forecast ensemble can represent the background deviation from the true atmospheric state. The NWP model engaged for evolution of B matrix in the ensemble-variational DA process contains uncertainty or model error. As such, the derived B matrix is likely not representative.

Another assumption that plays an important role in DA process is about the R matrix. Until recently, the diagonal R matrix has been engaged in most operational DA systems at NWP centers [27]. Therein, instrument noise (plus forward model uncertainty and representativeness error) is placed in R 's diagonals by leaving zeros in R 's off-diagonals. However, from a physical viewpoint and as shown by many studies, inclusion of inter-channel error correlations in the R matrix is physically sound and may improve analysis accuracy [28–38]. Other studies have focused on diagnosing inter-channel error correlations in assimilation analysis for certain observation [39–46]. Currently the Met Office has already accounted for inter-channel error correlations in operational global and regional systems for Atmospheric Infrared Sounder (AIRS), Infrared Atmospheric Sounding Interferometer (IASI), and Cross-track Infrared Sounder (CrIS) instruments [23]. Note that the term “inter-channel error correlations” in this paper refers to correlations among channels or bands of a detected radiance spectrum in which each radiance channel or band corresponds to its specified physical quantity, such as temperature, water vapor, gas compositions, surface temperature, and emissivity, in the atmosphere.

Since 1963, many methods have been developed for quantifying the inter-channel error correlations in the R matrix [47–52]. Among these methods, the one proposed by Desroziers et al. [52] has been widely used to estimate the observation error variances and inter-channel error correlations in Météo-France [52] and ECMWF (European Center for Medium-Range Weather Forecasts) [37] 4D-Var systems, and has also been applied to IASI data to estimate the structure of R matrix in the Met Office 4D-Var system [39]. Desroziers et al.'s scheme [52] was combined with the maximum-likelihood approach to estimate the spatial correlation length-scale of observation errors [53]. Some other studies also applied this technique [35,39,52], for example.

As the R matrix based on Desroziers et al. [52] relies on the observation operator (H), analyzed state (x_a), B matrix or background state (x_b), and selected forecast ensemble, all of which contain uncertainties, the R matrix used in the DA process may cause some problems. For example, the convergence speed of the DA process may decrease [27,39], which might indicate a possible limitation of those inter-channel error correlations and problems in the methods used to estimate the R matrix in DA process [2]. The most common problem is that this R matrix is not symmetrical (note that symmetry is a necessity for R matrix), leading to the eigenvalues of this R matrix being not positive-definite (note that positive-definite eigenvalues are required in DA process). Thus, this R matrix must be artificially made to be symmetrical and the eigenvalues must be altered to be positive-definite prior to their use in assimilation systems [2,35,39]. Tabcart et al. [2] suggested that these problems are caused by the interaction between the R and B matrices. The interaction is a numerical interaction rather than a physical interaction, which is a well-known issue during the iterations of the DA process in an effort to produce optimal outcomes for succeeding weather forecasts.

Given the above limitations, a better method to characterize R matrix is needed, as stated in previous studies [2,23,39]. Since the assumptions used in estimating R and B matrices are not always valid, reduction or entire elimination of the dependency on the assumptions and complete separation of the R matrix from the B matrix should enable us to produce a more physically sound R matrix. This study introduces a potential method to derive the R matrix, which requires no additional assumptions. Following the path of radiation through the atmosphere up to the radiation received by sensors and transformed to radiance, the correlations among the channels or bands in a detected atmospheric radiance spectrum allows us to propose a data-derived R matrix. This R matrix is built fully based on a large number (N) of detected atmospheric radiance spectra constructed from N real-time measurements (Appendix A), where N real-time measurements can be acquired by observing some location of interest during a short amount of time. The only two requirements in the data derivation of the R matrix are the fabrication of instruments and the data calibration procedures, which are fundamental for any data analysis.

Our first motivation for creating a data-derived R matrix using a large number of real-time measurements is as follows. In traditional approaches, the derived R matrix based on statistical forecast ensemble cannot fully represent the observation errors because it is unclear how much the variation within the ensemble sample can represent the observation errors (i.e., variance and covariance) between observed and true radiance. This derived R matrix may not fully characterize the atmospheric state of an interested moment. Conversely, since real-time measured data contain representative information of the atmospheric state at the time when the data were recorded, the derived R matrix based on real-time data can appropriately characterize the observation errors for the atmospheric state at that time when data were recorded. One critical difference between our proposed method and traditional approaches is that our proposed data-derived R matrix is built without assumptions (i.e., no dependence on H , x_a , x_b , or B matrix), whereas the traditional R matrices are constructed upon some assumptions that may not be always valid. The other motivation for this study is the desire to inspire the community to consider completely using real-time data to construct R matrix for satellite radiance observations from a different perspective.

Three notes need to be explained. Firstly, the method proposed in this paper for construction of a data-derived R matrix is only applicable to satellite measurements. Secondly, the meaning of “a large number of measurements” used here does not refer to the multi-measurements that are used in current numerical weather prediction within a certain time window. For multi-measurements, one single measurement is taken from each observation location at a moment of interest and thus multi-measurements correspond to multiple different atmospheric states. We do not use multi-measurements to construct the R matrix because the R matrix constructed from multiple different atmospheric states may be good enough within the time period when the multi-measurements were recorded, but not necessary ideal for assimilating the measurements. Thirdly, this paper presents the theoretical basis of a data-derived R matrix based on N real-time measured radiances and suggests

a conceptual design of a trigger configuration for satellites to record N real-time measurements. To record N real-time measurements, practical experimentation is required.

This paper is organized as follows. Section 2 briefly describes the methods that have been used to estimate the R matrix in literature. Remarks are provided about these methods and the long-used diagonal-only R matrix. Section 3 explains (i) radiation transmission through atmosphere and the radiation received at the sensors of an instrument and transformed to radiance, and (ii) the accompanying correlations among bands or channels of a detected atmospheric radiance spectrum. These are used to form the theoretical basis of the data-derived R matrix using N real-time measured radiances. Section 4 lists the advantages of using a data-derived R matrix. Since no real data currently exist for constructing a data-derived R matrix, Section 5 presents a conceptual design of a trigger configuration for satellites to record such data, and Section 6 provides simple simulation data with simple toy examples to demonstrate our proposed method. Our summary and recommendations for future work are discussed in Sections 7 and 8, respectively.

2. Estimation of the R Matrix

In this section, the existing methods for constructing the R matrices are introduced. The common factor among these methods is that they are based on the use of a single measurement (Appendix A). For example, Rodgers [54] and Daley [55] described the developments in these methods, and Rodgers [56] reviewed these methods in 1976. However, no single measurement is exact [57] because each measurement of atmospheric state must have varying degrees of error.

By focusing on some observation location of interest, one potential solution is to record a large number of real-time measurements during a short amount of time (Appendix A). These data can be used to derive the R matrix without any assumption, and the derived R matrix is completely free from dependence on H , x_a , and x_b (or B matrix). This is the focus of Section 3.

2.1. Earlier Methods

Among methods for quantifying inter-channel error correlations, the simplest method [47,48] uses a least square fit of an isotropic correlation model to the correlation sample, which is statistically collected from observation-minus-background (OMB) among all pairs of stations. By extrapolating the obtained fit curve to the origin, the ratio between the observation and background (i.e., forecast) error standard deviation can be determined. The extracted error covariance parameters are approximately homogeneous and isotropic and vary slowly with season.

Hollingsworth and Lönnberg [49] analyzed the statistical structure of errors from OMB and partitioned the errors into background and observation error components, which is called the Hollingsworth and Lönnberg method. The background here means the calculated radiance from $H(x_b)$. This method was built with statistical ensemble and was established based on two assumptions: Firstly, the background and observation errors are independent of each other, and, secondly, the background errors are homogeneously correlated in the horizontal plane while the observation errors are horizontally uncorrelated.

Due to the assumptions made in Hollingsworth and Lönnberg method, Dee and Da Silva [50] raised questions about this method's robustness and argued that background errors and observation errors could be correlated with each other. Dee and Da Silva [50] thus proposed a method to evaluate the background error and observation error covariance parameters based on maximum-likelihood covariance parameter estimation as described in Dee [58]. This method produced statistical error covariance parameters that varied with time. However, the assumption of correlations between background and observation error covariance parameters results in interaction or trade-offs between background error and observation error covariance parameters during iterations of the DA process, and ultimately, the acquired error covariance parameters may deviate from the true content. This is the fundamental limitation of this method if multiple parameters are entangled during simultaneous estimation.

Inspired by the generalized cross-validation approach [59,60], Desroziers and Ivanov [51] proposed a method to tune the observation error and background error covariance parameters based on observation-minus-analysis (OMA) from statistics of analysis innovations. The tuning was achieved using a randomization technique with a perturbation of either observations or background fields when computing the statistical average of the cost function's values at minima.

A method based on the combinations of OMB, OMA, and background-minus-analysis (BMA) for statistical estimation of the background errors (for the B matrix), observation errors (for the R matrix), and analysis errors was proposed by Desroziers et al. [52]. This method has become the most popular for estimation of the observation errors (i.e., variance and covariance) in the R matrix. The next section presents the formation of the R matrix based on this method.

2.2. Desroziers' R Matrix

Below, we only focus on the construction of the R matrix. Desroziers et al. [52] and Stewart et al. [35] provide more details. Denote d_b^o as the background innovation vector, which is the difference between the observations y , and their background counterparts $H(x_b)$, where the superscript o denotes observation. Similarly, d_a^o is the analysis innovation, which is the difference between the observations and their analysis counterparts, $H(x_a)$. Upon the first assumption of no correlation between background and observation errors, Desroziers et al. [52] showed that the statistical expectation of the product of d_b^o and d_a^o approximates the R matrix. According to Desroziers et al. [52], the covariance of innovation d_b^o is $E[d_b^o(d_b^o)^T] = R + HBH$; the covariance of d_a^o and d_b^o is $E[d_a^o(d_b^o)^T] = R$. Therefore, the latter covariance representing R matrix is presented in Equation (1).

$$R^D(i, j) = \frac{1}{N} \sum_{k=1}^N d_a^o(i) d_b^o(j) = \frac{1}{N} \sum_{k=1}^N [y_k(i) - H(x_a)(i)][y_k(j) - H(x_b)(j)], \quad (1)$$

where the superscript D denotes the method by Desroziers et al., indices i and j are the band or channel numbers of a radiance spectrum, and index k runs over all forecast ensemble from 1 to N . The second assumption made by Desroziers et al. is that H , x_a , and B matrix are accurate when the R^D matrix constructed using Equation (1) is applied in analyses. Unfortunately, uncertainties always exist in H , x_a , and B matrix. As a consequence, problems arise.

The first fundamental problem is that this R^D matrix is not uniformly symmetrical [30,35,50]. Since $y_k(i) - H(x_a)(i)$ is not necessarily equal to $y_k(i) - H(x_b)(i)$ for component i , and $y_k(j) - H(x_a)(j)$ is not necessarily equal to $y_k(j) - H(x_b)(j)$ for component j :

$$\frac{1}{N} \sum_{k=1}^N d_a^o(i) d_b^o(j) \neq \frac{1}{N} \sum_{k=1}^N d_a^o(j) d_b^o(i) \implies R^D(i, j) \neq R^D(j, i), \quad (2)$$

which is a result of the violation of the second assumption introduced by Desroziers et al. This is also a violation of the physical requisite for data analysis. The issue caused by the asymmetric R^D matrix is that the eigenvalues of this R^D matrix are not all positive (i.e., not positive definite). Therefore, prior to using this R^D matrix in assimilation systems, artificial treatments are required to symmetrize the asymmetrical R^D matrix (by taking the mean of the original R^D matrix and its transpose matrix) and to modify any negative eigenvalue to be positive [39]. Some other treatments, as proposed in Desroziers et al., involve applying an iterative procedure for updating the contents of R^D matrix until R^D matrix is closer to being symmetrical [35,52]. Regardless of which treatment is employed, the resulting R^D matrix is influenced by the specification of H , x_a , and B matrix, which are not perfectly known.

The second fundamental problem is experienced by all methods proposed in literature, which is the use of statistical forecast ensemble in the derivation of the R matrix [47–52] as well as of the B matrix. Strictly, the variation among the forecast ensemble may not fully reflect the departure of real observations from the true atmospheric state at that time when data were recorded, and thus the extent to which this derived R matrix can represent the true observation error covariance matrix for the atmospheric state at that time when data were recorded cannot be ascertained.

2.3. Diagonal-Only R Matrix

The diagonal-only observation error matrix (R^d matrix), where instrument noise plus forward model uncertainty and representativeness error are placed in R^d 's diagonals by leaving zeros in R^d 's off-diagonals, is still used in many atmospheric research analyses and operational weather forecast (ARWF). For this case, the observation error in R^d 's diagonals is given by [39,56]:

$$\varepsilon = y - y^t, \quad (3)$$

where ε is the observation error between the observed radiance y and the true radiance y^t . Hence, the diagonal R^d matrix is:

$$R^d(i, i) = [\varepsilon\varepsilon^T](i, i), \quad (4)$$

with ε^T being the transpose of ε . According to previous studies [39,56], ε is interpreted as instrument noise (plus forward model uncertainty and representativeness error).

The first remark on this diagonal R^d matrix is that the retrieved atmospheric state parameters (ASPs; referring to temperature, water vapor, gas compositions, surface temperature, emissivity, etc.) rely on which instrument is being used because the instrument noise is included in the R^d matrix. This contravenes reality because real ASPs are independent of any instrument and thus retrieved ASPs should be independent of any instrument. Given a measured system (e.g., an atmospheric state), suppose that we use M different instruments to measure the same atmospheric state, each of which instrument has a different level of noise. Regardless of the instrument used, any known level of instrument noise needs to be subtracted and not included in the data analysis so that the output of the analysis is not be affected by the instrument used. Rather than observation noise (or named error or variation) induced by measurements within a measured system, note that here we are discussing instrument noise due to the instrument itself that generally can be obtained from laboratory experiments.

In terms of measurements, the observed radiance y at a given time is one of many varying radiances around the true radiance y^t . Thus, the second remark is that ε in Equation (3) should be interpreted as a pure deviation between the observed y and the true y^t rather than instrument noise. Unfortunately, since single measurement (Appendix A) has been used in ARWF, neither the true y^t nor this variance is obtainable. The observed y and true y^t vary with scenes (i.e., atmospheric states); thus, the observation errors in an R matrix vary with atmospheric states. So, pre-launch measurements from the laboratory are unable to provide information about variances for an R matrix. It is crucial that this variance is not equal to instrument noise in general. Recent studies [35,39] have demonstrated that diagonal elements in an observation error covariance matrix have visible differences from instrument noise. Therefore, use of the diagonal-only R^d matrix imposes some limitations in analyses.

3. Data-Derived R Matrix

Physically, correlations exist among bands or channels of a detected atmospheric radiance spectrum. This section explains these correlations based on the phenomena of energy absorption and emission caused by electron transition, which have been observed in the study of electromagnetic radiation absorbed and emitted by atoms [61] and have been applied in the study of radiative properties occurred in atmosphere [62]. This is followed by the derivation of an R matrix using N detected radiance spectra constructed from N real-time measurements acquired by observing some location of interest for a short amount of time.

3.1. Physical Mechanism of Radiation Through Atmosphere

An atmospheric state in thermodynamic equilibrium (hereafter, "atmospheric state in equilibrium") is viewed from macroscopic aspects, whereas any fluctuation state (around equilibrium) caused by physical processes between radiation and atmosphere is viewed from microscopic aspects.

One fluctuation state is acquired from one single measurement, and a macroscopic equilibrium is composed of many microscopic fluctuation states, among which fluctuation states are almost or exactly balanced. Therefore, given an observation location over a short amount of detection time, the atmospheric state in equilibrium may not be reflected by any single measurement, but by the average of a large number of measurements.

Figure 1 depicts the radiation spectrum and which parts of that spectrum are detected by the instrument sensor. The region between two adjacent black lines in the radiation spectrum (called a segment radiation) corresponds to the measurement between the two black lines in the instrument sensor. In the following, when a radiation of a given wavenumber (or energy) is termed, this means the named segment radiation as denoted above; when a radiation without any specified wavenumber (or energy) is termed, this indicates the radiation covering all possible wavenumbers (or energies) within. Below, we present the interactions between radiation and atmosphere from microscopic aspects.

When the sun's radiation enters the earth's atmosphere ("atmosphere" hereafter), part of the sun's radiation in visible (VIS) and the near infrared (NIR) spectrum is reflected back to space by the atmosphere and clouds before reaching the earth's surface. However, part of the sun's radiation passes through the atmosphere and subsequently is reflected by the earth's surface back to space or absorbed by the earth's surface; simultaneously, emissions from radiation of longer wavelengths, such as infrared, are radiated out back to space from the earth's surface and atmosphere. All these occurrences depend on the atmospheric states.

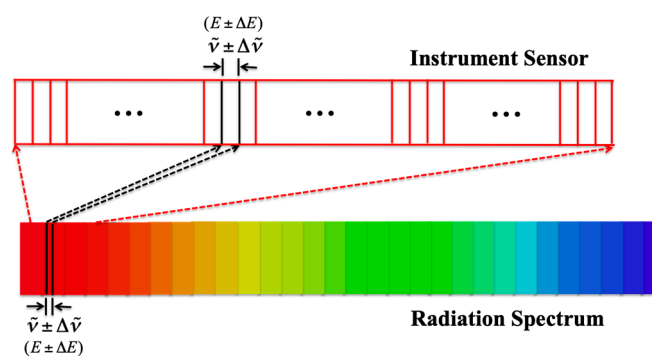


Figure 1. Detection of a portion of radiation (e.g., infrared; lower part) by an instrument sensor (upper part), where $\tilde{\nu} \pm \Delta\tilde{\nu}$ ($E \pm \Delta E$) represents the width of wavenumber (energy) of a segment radiation (see the segment bordered by two black lines in the radiation spectrum) that is detected by the channel or band of $\tilde{\nu} \pm \Delta\tilde{\nu}$ ($E \pm \Delta E$) in an instrument sensor (see the channel or band bordered by two black lines in the instrument sensor).

After, when radiation travels upward through the atmosphere, interactions occur between radiation and molecules and particles in the atmosphere, caused by particle targets such as CO₂, ozone, aerosol, water vapor, precipitation, and cloud [61,62]. These interaction processes include absorption, emission, elastic scattering, and inelastic scattering. Generally, the interaction processes of absorption, emission, and inelastic scattering involve energy exchange between incident radiation and particles in the atmosphere. In quantum mechanics, given the energy (E) of radiation (or photons; i.e., a segment radiation), the Planck's energy-frequency relation states that E is proportional to its frequency (ν): $E = h\nu = hc/\lambda = hc\tilde{\nu}$, where h is the Planck constant, c is the speed of light, and λ and $\tilde{\nu}$ are wavelength and wavenumber, respectively. When energy changes, its corresponding frequency (wavelength and wavenumber as well) changes as well. As a consequence, given an incident radiation of a specified energy E (or specified wavenumber $\tilde{\nu}$), radiation of different energy E' (i.e., different wavenumber $\tilde{\nu}'$) may be emitted after the interaction. In this case, a couple of basic formations for the energy of outgoing radiation are possible. Firstly, the energy of the incident radiation can be completely absorbed by particles in the atmosphere, and no radiation is emitted as a result. Secondly, when particle targets

in the atmosphere are thin, incident radiation may only lose some portion of its energy, and radiation with less energy is emitted. Thirdly, one special formation can occur. Sometimes, the particles in atmosphere absorb part or all of the energy of the incident radiation and store such energy (denoted E_p) inside the particles. Then, the next time another new incident radiation of energy E_{in} interacts with such particles, the incident radiation may release some energy (denoted E_r) to the particles, but gain the energy E_p freed from the particles. The resulting energy of the outgoing radiation, $E_{in} - E_r + E_p$, can be greater or less than the original incident energy E_{in} , depending on the positive or negative balance of $E_r - E_p$.

Physically, the mechanism for the energy exchange between incident radiation and particles in the atmosphere follows the phenomena of energy absorption and emission that occur between incident radiation and atoms (or molecules) [61]. Therefore, the wavenumber (or energy) of the outgoing radiation can be zero (i.e., no outgoing radiation), less, or greater than that of the incident radiation. This means that with respect to the wavenumber ($\tilde{\nu}_x$) of the incident radiation, the wavenumber ($\tilde{\nu}_y$) of the outgoing radiation can be different from that of the incident radiation and range from zero to infinity, though we know an infinite wavenumber is impossible. This suggests that radiation of wavenumber $\tilde{\nu}_x$ is correlated with the radiation of wavenumber $\tilde{\nu}_y$.

The same physical mechanism is applied to all segments in a radiation beam. After radiation beams pass through the atmosphere, the detected radiation at the sensors installed on an instrument is transformed to radiance in terms of per steradian per square meter during a short detection time frame, forming an atmospheric radiance. Recall the meaning of inter-channel error correlations among bands or channels provided in Section 1. The statement “radiation of wavenumber $\tilde{\nu}_x$ has correlations to radiation of wavenumber $\tilde{\nu}_y$ ” equivalently suggests that a channel or band of wavenumber $\tilde{\nu}_x$ has physical correlations with the channel or band of wavenumber $\tilde{\nu}_y$ in a detected atmospheric radiance spectrum. From microscopic aspects, the notion of atmospheric fluctuation states induced by various extents of interactions between radiations and particles in the atmosphere allow us to examine correlations among those atmospheric physical variables (e.g., temperature, water vapor, various gas compositions, surface temperature, emissivity, etc.) via studying correlations among the channels or bands of a detected atmospheric radiance spectrum. The extent of the correlations depends on the atmospheric state at the time when data are recorded. All these can be explored through the observation error covariance (R) matrix that is presented in the next section using a large number of real-time measurements of radiance spectra.

3.2. Construction of Data-Derived R Matrix

Before proceeding, some notes about the differences among instrument noise, instrument random error, and observation error are provided. Instrument noise (as noise equivalent delta temperature) is produced by the instrument itself, independent of measurements of a measured system (e.g., atmospheric state here), and is usually obtainable from the laboratory. Instrument random error, produced by instrument itself as well, may occur during measurements and is usually uncontrollable and random. Observation error, independent of the instrument, is caused by atmospheric variations. Recall the macroscopic versus microscopic aspects mentioned earlier. Each atmospheric variation state corresponds to one fluctuation state (viewed from a microscopic aspect). An atmospheric state in equilibrium (viewed from macroscopic aspects) can be determined from averaging many atmospheric variation (or fluctuation) states.

Construction of an R matrix directly based on statistical measured data has been commonly used in many experiments, such as fundamental sciences in physics and chemistry. Here, we introduce this approach to the construction of an R matrix in atmospheric sciences. By observing some location of interest within a solid view angle $\Delta\Omega$ during a short amount of detection time $\Delta\tau^N$ around a limited area Δa , suppose that N measurements of radiance spectra (Appendix A) are collected with a given instrument. The N measured radiance spectra are used to construct one R matrix for this observation location of interest. Each measurement contains n_{ch} spectral radiance bands or channels. Denote

indices i and j as band or channel numbers and index k runs from 1 to N measurements. For the k th measurement, the measured radiance $y_k(i)$ for the i th band or channel can be expressed as:

$$y_k(i) = \lambda(i) + \varepsilon_k(i), \tag{5}$$

where $\lambda(i)$ is the true value of radiance $y(i)$ and $\varepsilon(i)$ is induced by atmospheric variation (and instrument random error, if any). The variation $\varepsilon(i)$ in each radiance band or channel is assumed to be a Gaussian or Gaussian-like distribution, which is widely accepted in real world practices. Any known instrument noise and any known non-signal events (if any) are subtracted in Equation (5) since they are irrelevant to the physical quantity that describes the measured system. Note that physical quantity and measured system can be referred to as atmospheric state parameters and atmospheric state, respectively, in our case.

From N measured radiance spectra:

$$\sigma_y^2(i) = \frac{1}{N} \sum_{k=1}^N [y_k(i) - \mu_y(i)]^2, \mu_y(i) = \frac{1}{N} \sum_{k=1}^N y_k(i), \tag{6}$$

where $\mu_y(i)$ is the mean value of measured radiance $y(i)$ for the i th band or channel, $\sigma_y(i)$ is the uncertainty of $\mu_y(i)$, and $\sigma_y^2(i)$ is the variance of $y(i)$. Given the band or channel i , the N measurements of $y(i)$ can be equivalently regarded as a single measurement of N -dimensional vectors. Each band or channel, hence, has its own mean value $\mu_y(i)$ and associated uncertainty $\sigma_y(i)$ that are obtained from the N -dimensional vectors. For a radiance spectrum of n_{ch} bands or channels, the joint probability density function (p.d.f) is the product of n_{ch} Gaussians [63,64]:

$$g(\mu_1, \dots, \mu_{n_{ch}}; \lambda_1, \dots, \lambda_{n_{ch}}; \sigma_1, \dots, \sigma_{n_{ch}}) = \prod_{i=1}^{n_{ch}} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(\mu_i - \lambda_i)^2}{2\sigma_i^2}\right], \tag{7}$$

where σ_i , μ_i , and λ_i are abbreviations of $\sigma_y(i)$, $\mu_y(i)$, and $\lambda(i)$, respectively. The vector quantity λ is the true value of vector μ (i.e., the radiance spectrum describing an atmospheric state in equilibrium) and is given as a function of vector p and vector x , $\lambda(p;x)$, where p denotes known parameters (such as atmospheric pressures) and vector x represents unknown parameters (such as atmospheric state parameters of an atmospheric state in equilibrium). The least square method is used to extract the quantity x . Here, the physical meaning of $\lambda(p;x)$ is the same as the observation operator $H(x)$; accordingly, $\lambda(p;x)$ is replaced by $H(x)$ hereafter.

Taking the logarithm of the joint p.d.f. and dropping those terms unrelated to the vector x gives the log-likelihood function:

$$\log L(x) = -\frac{1}{2} \sum_{i=1}^{n_{ch}} \frac{[\mu_i - H(x)]^2}{\sigma_i^2}. \tag{8}$$

The quantity x can be found by minimizing the χ^2 quantity:

$$\chi^2(x) = \sum_{i=1}^{n_{ch}} \frac{[\mu_i - H(x)]^2}{\sigma_i^2}. \tag{9}$$

Equations (8) and (9) are used to describe the situation among which bands or channels of a radiance spectrum are independent. However, due to the effects of correlations among bands or channels, the term $\frac{1}{\sigma^2}$ should be replaced by a covariance matrix R , and Equations (8) and (9) accordingly become:

$$\log L(x) = -\frac{1}{2} \sum_{i,j=1}^{n_{ch}} [\mu_i - H(x)(i)](R_{ij})^{-1}[\mu_j - H(x)(j)], \tag{10}$$

$$\chi^2(x) = \sum_{i,j=1}^{n_{ch}} [\mu_i - H(x)(i)](R_{ij})^{-1}[\mu_j - H(x)(j)], \tag{11}$$

where $R_{ij} \equiv R(i, j)$. Modified from the variance $\sigma_y^2(i)$ in Equation (6), the covariance between the i th and j th bands or channels is expressed as:

$$\sigma_y(i, j) \equiv R(i, j) = \frac{1}{N} \sum_{k=1}^N [y_k(i) - \mu_k(i)][y_k(j) - \mu_k(j)], \quad (12)$$

and should be placed in R 's off-diagonal elements. This data-derived R matrix is guaranteed to be symmetrical, which meets the DA process requirement. Moreover, the correlation, $cor_y(i, j)$ between the i th and j th bands or channels is then given by:

$$cor_y(i, j) = \frac{\sigma_y(i, j)}{\sigma_y(i)\sigma_y(j)}. \quad (13)$$

Using standard error propagation, the covariance matrix U for estimated x is given by:

$$U = [H(x_{best})^T R^{-1} H(x_{best})]^{-1}, \quad (14)$$

where x_{best} means the found x that is closest to the true x . The diagonals in U are the variances of the corresponding variances of vector x .

Equations (6)–(14) are well known and can be found in textbooks of probability theory and statistics or related [63,64]. The reason why we introduce them here is to remind people that the R matrix used in the DA system of ARWF can be obtained directly from a large number of real-time measured radiance spectra, rather than using diagonal-only R^d or constructing R^D matrix based on theoretical H , human-assumed B matrix, and x_a , which cause problems in the DA process as mentioned previously.

Then, how does the data-derived R matrix work on the cost function? The role of the χ^2 in Equation (11) is the same as the second term (inside the braces) in the cost function $J^{current}$ used in the current DA system:

$$J^{current}(x) = (x - x_b)^T B^{-1} (x - x_b) + \sum_m \left\{ \sum_{i,j=1}^{n_{ch}} [y(i) - H(x)(i)] (R_{ij}^*)^{-1} [y(j) - H(x)(j)] \right\}, \quad (15)$$

except that R_{ij}^* is either R_{ij}^D or R_{ij}^d , and single measurement y at each observation location is used in $J^{current}$. Conversely, we propose using the average μ from N measurements recorded at each observation location for the cost function $J^{proposed}$ in the DA system:

$$J^{proposed}(x) = (x - x_b)^T B^{-1} (x - x_b) + \sum_m \left\{ \sum_{i,j=1}^{n_{ch}} [\mu_i - H(x)(i)] (R_{ij})^{-1} [\mu_j - H(x)(j)] \right\}, \quad (16)$$

where R_{ij} is given in Equation (12). The notation m runs over all included observation locations from various instruments during a certain time window.

4. Advantages of Using Data-Derived R Matrix

There are many advantages of using data-derived R matrix, such as independence on forward model, x_a , or x_b for y^t , no forward model uncertainty, no representative error, no entanglement with B matrix, and symmetric matrix etc. Several other advantages are explained below.

1. View from Macroscopic and Microscopic Aspects

An atmospheric state in equilibrium can be viewed from a macroscopic viewpoint that is composed of many microscopic aspects, where each microscopic aspect represents one fluctuation state around the macroscopic state in equilibrium. That is, by taking repeated measurements at an observation location during a short amount of time, each measurement represents one fluctuation state, whereas the average of those repeated measurements is regarded as the expectation of the truth of the macroscopic

state in equilibrium. The deviation between each measurement and the average is the observation error and is used for constructing a data-derived R matrix.

2. Accurate Expectation of True Radiance y^t

The R^D matrix in Equation (1) is constructed based on Desroziers et al. [52], where $H(x_a)$ and $H(x_b)$ are taken as the true radiance y^t . However, $H(x_a)$ and $H(x_b)$ are the radiance calculated from a radiative transfer model or algorithm (i.e., forward model) upon assumptions such that R^D relies on H , x_a , and x_b (or B matrix).

For a data-derived R matrix, in probability and statistics theory [63,64], the expected value (known as expectation) of a variable is the average value of repeated measurements of the same experiments. Taking the average of repeated measurements as the expectation of the true value of a variable is generally used in experimental measurements [57]. We here apply such a concept to the data-derived R matrix by taking the average of the observed radiance (i.e., μ , see Equation (6)) as the expectation of the true radiance y^t , which is consistent with what is stated in note 1).

3. Correlations Accounted Among All Physical Quantities

The data-derived R matrix does not only address one particular atmospheric physical quantity, but also contains information of variances and correlations for all atmospheric physical quantities, where each radiance band or channel corresponds to its own representative atmospheric physical quantity. $R(i, i)$ simply characterizes the variance of atmospheric physical quantity in the i th radiance band or channel, and the correlation between any two atmospheric physical quantities in the i th and j th bands or channels can be readily obtained by Equation (13).

4. Less Flow Dependency

The observation error in an observation error covariance matrix is defined as the difference between the observed radiance y and the true radiance y^t , and attained y and y^t depends on the weather conditions at the time when data are recorded. The messages in the data-derived R matrix thus only depend on the weather conditions at a given observation location at a given time. Therefore, given an observation location at a time of interest, the data-derived R built in Equation (12) is only computed once. Hence, there is no need to update the data-derived R matrix within a DA process period. This considerably reduces computation time as compared with the traditional approach. The current DA system and the proposed DA system are compared in Figure 2.

In the current DA system (Figure 2a), if the output x_a from the DA process is not converged, adjustments to both R^* (i.e., either R^D or R^d) and B matrices are needed before processing the next DA cycle. This leads to complexity in the DA process and a less accurate x_a used for next-step weather forecast because the greater the number of factors (i.e., R^* and B matrices here) influencing x_a , the greater the uncertainty associated with x_a . Conversely, in the proposed DA system (Figure 2b), at each given observation location and time, the content in the data-derived R matrix is only calculated once and fixed during iterations of the DA process. Thus, only the B matrix needs adjustment for next cycle of the DA process if x_a has not converged. This approach simplifies the DA process as compared with the traditional approach and is expected to produce a more accurate x_a because only one factor (i.e., the B matrix) needs adjustment in the DA process. Note that $H(x)$ must be computed in each iteration of both DA systems.

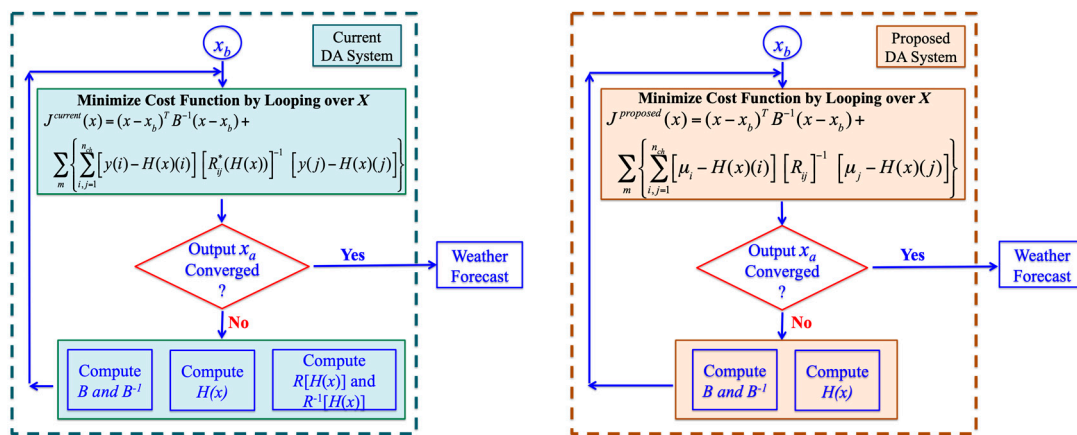


Figure 2. (a) Current data assimilation (DA) system (left) versus (b) proposed DA system (right), where the DA systems are indicated inside dashed squares. Note that R^* can be either R^D or R^d in the current DA system.

5. Temporal Variations of Inter-Channel Error Correlations

In the current DA system, the R^D or R^d matrix is usually assumed to be static for the day, a few days, or even for months. The messages in an observation error covariance matrix depend on weather conditions. Therefore, the data-derived R matrix allows us to quantify the change in inter-channel error correlations over time, where the change in the magnitude of R 's messages relies on how quickly the weather conditions change over time. Currently, there are no such real-time data; thus, we are unable to present the magnitude of change in this paper.

6. Gaussian Distribution of Variation

Here provides a supplementary explanation to the cost functions. The expression $[y(i) - H(x)(i)] R^{-1} [y(j) - H(x)(j)]$ in both Equations (15) and (16) is basically derived from the least square method [63,65] upon the assumption of Gaussian or Gaussian-like distribution of variation in each radiance band or channel [63], which is widely accepted in real world practices as mentioned previously. This means that the variation in each radiance band or channel follows a Gaussian or Gaussian-like distribution. A large enough sample can be used to verify if the distribution is a Gaussian or Gaussian-like distribution. However, recording only a few measurements does not prevent the variation in each radiance band or channel from behaving as a Gaussian or Gaussian-like distribution if the variation does follow a Gaussian or Gaussian-like distribution.

5. Conceptual Design of Proposed Trigger Configuration

Technically, taking a large number of measurements within a short detection time $\Delta\tau^N$ (Appendix A) is achievable by modifying the trigger configuration of data recording from ground, based on a private communication with specialists in hyper-spectral hardware instruments. As for the number of measurements within $\Delta\tau^N$ needed, this may require simulation and/or practical experimentation by considering the desired extent of uncertainty in the analyses. This uncertainty includes that induced by atmospheric variation and instrument systematic variation, and the uncertainty level, which can be obtained by the covariance matrix U in Equation (14), would be revealed in the retrieved atmospheric state parameters.

The traditional trigger configuration continuously takes single measurements, with each measurement taken within detection time $\Delta\tau^S$ (Appendix A), where not all data are yet used in practice. To fully use the collected data, the trigger configuration is proposed to take N measurements within $\Delta\tau^N$ with some time separation (ΔT) between two consecutive N measurements, regardless of the low earth orbit (LEO) detection sources or geosynchronous satellites. Figure 3 illustrates the

traditional trigger configuration (TTC) versus the proposed trigger configuration (PTC). The load of the data recorded with the PTC can thus be reduced and would be less expensive.

The detection time $\Delta\tau^N$ may be longer than $\Delta\tau^S$. Both $\Delta\tau^N$ and ΔT can be studied using simulation and/or practical experimentation. The fluctuation in the measured radiance, possibly due to recording a large number of measurements within a short time span, does not have any significant impact on the retrieved ASPs (see the study presented in Section 6.1). Therefore, if we use data recorded with the TTC in the future, the long-term issues and costs due to inaccurate weather forecasts that we have experienced at least since the first satellite was launched for taking atmospheric data would continue, which would outweigh the cost of data captured with PTC.

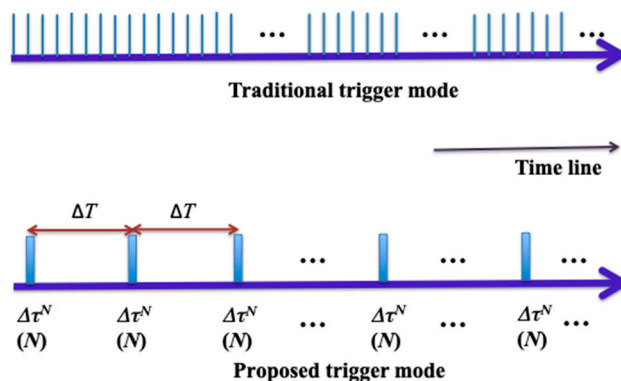


Figure 3. Traditional trigger configuration (TTC) versus proposed trigger configuration (PTC), where each blue bar represents detection time when scanning one location. Symbol $\Delta\tau^N(N)$ denotes taking N measurements within $\Delta\tau^N$.

6. Simple Simulation Studies

Since no such real data exist, we instead used simulation data with simple toy examples to demonstrate how the data-derived R matrix works and how the retrieved outputs perform compared with the methods employed in current DA systems. This was inspired by Desroziers et al. [52], where simulation data with toy examples were used to demonstrate their methods. Figure 4 illustrates the simple simulation example, where three bands or channels were considered for simplification. The atmospheric state parameters are represented as $X \equiv (x_1, x_2, x_3)^T$ and the measured radiance at the instrument sensor is described as $Y \equiv (y_1, y_2, y_3)^T$. The relationship between X and Y is through matrix A , which is $Y = AX \equiv H(x)$. The first term in the cost function, $(x - x_b)^T B^{-1} (x - x_b)$, was unable to be simulated in this study and thus is ignored here, which would not affect the purpose of our demonstration. Four different observation operators were adopted in this study as follows:

$$H_{cr}^{corr1}(x) \equiv A_{cr}^{corr1} X = \begin{pmatrix} 0.20 & 0.27 & 0.53 \\ 0.46 & 0.37 & 0.17 \\ 0.11 & 0.56 & 0.33 \end{pmatrix} X, \tag{17}$$

$$H^{corr2}(x) \equiv A^{corr2} X = \begin{pmatrix} 0.950 & 0.020 & 0.030 \\ 0.025 & 0.960 & 0.015 \\ 0.017 & 0.043 & 0.940 \end{pmatrix} X, \tag{18}$$

$$H_{wr1}(x) \equiv A_{wr1} X = \begin{pmatrix} 0.18 & 0.32 & 0.50 \\ 0.39 & 0.40 & 0.21 \\ 0.14 & 0.50 & 0.36 \end{pmatrix} X, \tag{19}$$

$$H_{wr2}(x) \equiv A_{wr2} X = \begin{pmatrix} 0.16 & 0.36 & 0.48 \\ 0.37 & 0.44 & 0.19 \\ 0.18 & 0.47 & 0.35 \end{pmatrix} X, \tag{20}$$

and

$$X^{true} = (6.0, 17.0, 9.0), \mu = (10.6, 10.7, 13.2), \quad (21)$$

$$Y_1^{single} = (9.9, 11.3, 13.8) \quad (22)$$

In the following, notations *corr*, *cr*, and *wr* denote correlation, correct, and wrong, respectively; $H(A)$ means H and/or A are/is used rather than indicating that H is a function of A . Hence, $H_{cr}^{corr1}(A_{cr}^{corr1})$ is taken as a correct $H(A)$, both H_{wr1} and H_{wr2} (both A_{wr1} and A_{wr2}) are taken as two incorrect H 's (two incorrect A 's), and both H_{cr}^{corr1} and H_{cr}^{corr2} (both A_{cr}^{corr1} and A_{cr}^{corr2}) are used for illustration of correlations between any two channels or bands. X^{true} is assigned as the true values of ASPs in the equilibrium state; radiances Y_1^{single} and Y_2^{single} are regarded as representing two fluctuation states around an atmospheric state in equilibrium and are obtained according to $Y = A_{cr}^{corr1} X^{true} [1 + G]$ when single measurement is considered, where G is a Gaussian distribution centered at zero with a deviation of 0.05 if no other deviation values are applied elsewhere; and μ is the averaged radiance from N measured radiances. Similar to what is usually performed in DA analysis, when looking for the best X_{best} that is closest to X^{true} , X is restricted within $X^{true} \pm 5$ during minimization of the cost function. Note that equation $Y = AX[1 + G]$ used for radiance generation and for construction of R matrix is abbreviated as $Y = AX$ hereafter. For a large number of measurements, $N = 200$ is used in our study.

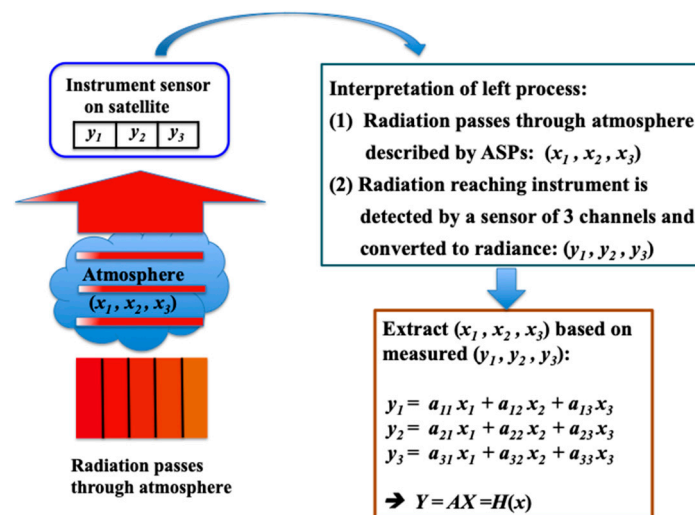


Figure 4. Simple toy examples employed to demonstrate the performance of the data-derived R matrix.

Below, we present four cases (Cases A–D). The purpose of Case A (Section 6.1) was to demonstrate the importance of verification for correctness of H and to investigate the impact of fluctuation in measured radiances on the retrieved X_{best} . The questions addressed in Case B (Section 6.2) were: Is diagonal-only R^d matrix appropriate? Is single measurement enough? Does H need verification? Similar to Case B, the purpose of Case C (Section 6.3) was to answer: Is R^D matrix appropriate? Is single measurement enough? Does H need verification? Finally, the purpose of Case D (Section 6.4) was to determine if correlations could be automatically manifested from the R matrix. For each case, we described its analysis and provided results and discussions.

6.1. Case A

Firstly, N sets of measured radiances Y were obtained from N real-time measurements, which were generated according to $Y = A_{cr}^{corr1} X^{true}$. The data-derived R matrix was then constructed using the N sets of radiances based on Equations (6) and (12). Subsequently, the second term of the cost function in Equation (16) was minimized by looping over X to look for the best X_{best} using various $H(A)$. Lastly, the uncertainty of each retrieved x_i in X_{best} could be obtained from the covariance U matrix (Equation (14)).

The fluctuation in measured radiance could be magnified as a result of taking N measurements over a short time span. To visualize the effect of taking N measurements on the retrieved X_{best} , the fluctuation was simulated by the deviation in the Gaussian function. Deviation values of 0.05, 0.10, 0.15, 0.20, 0.30, and 0.40 were investigated, and the corresponding generated radiances were denoted by Y_d^N or simply Y_d hereafter, with subscript d representing the deviation values. The procedure of this analysis is depicted in Figure 5.

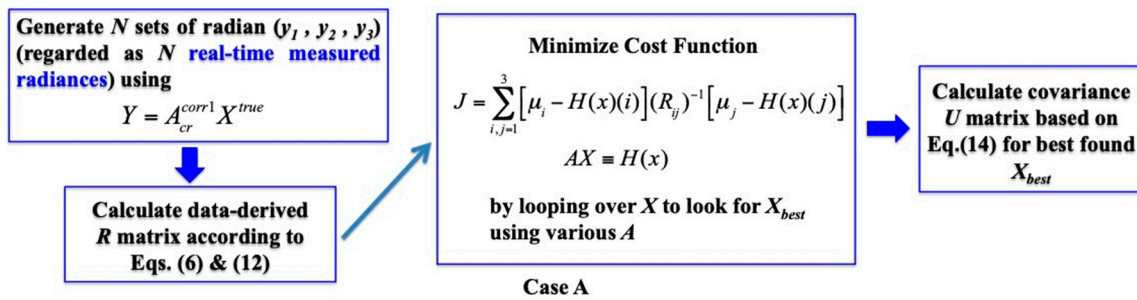


Figure 5. Analysis procedures for Cases A.

The results are listed in Table 1. In the cost function (Figure 5), μ is the average of all N sets of measured radiances, considered as the expectation of true radiance; and the R matrix was fully built from real-time data, providing good representativeness of the true R matrix. As seen in Table 1, only a correct H_{cr}^{corr1} (A_{cr}^{corr1}) could produce a correct X_{best} that was consistent with the true X^{true} , where both the cost function value J and the uncertainty for each retrieved x_i were the smallest. Note that the best X_{best} would fall outside $X^{true} \pm 5$ with a bigger J value if any component of the retrieved X_{best} reaches the boundary of $X^{true} \pm 5$. This suggests that verification of the correctness of H (or the forward model) is necessary for ASP retrieval in a DA system.

Under the same conditions except for deviation values, the results show that the extracted values of X_{best} using various deviations (i.e., various extents of fluctuation) were almost the same. Hence, only the results using deviations of 0.05, 0.15, and 0.40 are listed in Table 1. These results indicate that fluctuation in measured radiances did not have any significant impact on the retrieved X_{best} even when the deviation was up to 40%.

We noted three features in Table 1. The first one was that the bigger the deviation, the smaller the value of J . This occurred because the R matrix was in the denominator of the cost function J and a larger fluctuation produced bigger errors and bigger error correlations in the R matrix, resulting in smaller J value. The second feature was that the J value for the case using A_{wr1} was smaller than when using A_{wr2} because A_{wr1} was chosen to be closer to A_{cr}^{corr1} than A_{wr2} . The third feature was that the larger the deviation (i.e., fluctuation) in the measured radiances, the greater the uncertainties in the retrieved X_{best} , which was reasonable.

Table 1. Results for Case A, where Y_d^N or Y_d represent data from N measured radiances. The second column means the R matrix is based on what for construction and the third column represents what H (A) is used in minimization.

Y	R Matrix Construction	$H(A)$ in Minimization	$X^{true} = (6.0, 17.0, 9.0)$ Best found X_{best}	J
$Y_{0.05}^N$	$Y_{0.05} = A_{cr}^{corr1} X^{true}$	$H_{cr}^{corr1} (A_{cr}^{corr1})$	$(5.98 \pm 0.34, 17.03 \pm 0.86, 8.98 \pm 0.47)$	0.0005
$Y_{0.05}^N$	$Y_{0.05} = A_{cr}^{corr1} X^{true}$	$H_{wr1} (A_{wr1})$	$(2.28 \pm 0.51, 20.55 \pm 1.18, 7.20 \pm 0.51)$	0.0006
$Y_{0.05}^Y$	$Y_{0.05} = A_{cr}^{corr1} X^{true}$	$H_{wr2} (A_{wr2})$	$(1.03 \pm 1.23, 17.20 \pm 1.92, 6.50 \pm 0.80)$	38.4205
$Y_{0.15}^N$	$Y_{0.15} = A_{cr}^{corr1} X^{true}$	$H_{cr}^{corr1} (A_{cr}^{corr1})$	$(6.08 \pm 0.86, 16.97 \pm 2.47, 9.00 \pm 1.36)$	0.0002
$Y_{0.15}^N$	$Y_{0.15} = A_{cr}^{corr1} X^{true}$	$H_{wr1} (A_{wr1})$	$(2.20 \pm 1.57, 20.53 \pm 3.65, 7.17 \pm 1.54)$	0.0002
$Y_{0.15}^N$	$Y_{0.15} = A_{cr}^{corr1} X^{true}$	$H_{wr2} (A_{wr2})$	$(1.03 \pm 3.43, 17.38 \pm 5.29, 6.47 \pm 2.42)$	4.6355

Table 1. Cont.

Y	R Matrix Construction	$H(A)$ in Minimization	$X^{true} = (6.0, 17.0, 9.0)$ Best found X_{best}	J
$Y_{0.40}^N$	$Y_{0.40} = A_{cr}^{corr1} X^{true}$	$H_{cr}^{corr1} (A_{cr}^{corr1})$	$(5.87 \pm 2.63, 17.13 \pm 6.84, 9.08 \pm 3.59)$	0.00001
$Y_{0.40}^N$	$Y_{0.40} = A_{cr}^{corr1} X^{true}$	$H_{wr1} (A_{wr1})$	$(2.40 \pm 4.39, 20.30 \pm 9.61, 7.15 \pm 4.64)$	0.00002
$Y_{0.40}^N$	$Y_{0.40} = A_{cr}^{corr1} X^{true}$	$H_{wr2} (A_{wr2})$	$(1.02 \pm 9.25, 17.20 \pm 14.5, 6.85 \pm 6.68)$	0.5600

6.2. Case B

Single measurement for radiance Y_1^{single} (or Y_2^{single}) and diagonal-only R^d matrix were considered in this case. Firstly, Y_1^{single} (or Y_2^{single}) was adopted as the measured radiance with the existence of deviation from the average of radiance μ . Secondly, the variances in the R^d matrix were taken from the diagonal variances of the ensemble-derived R matrix, where the ensemble-derived R matrix in Case B was built using N sets of ensemble radiances that were generated according to $Y = AX^{true}$ with $A = A_{cr}^{corr1}$, A_{wr1} , or A_{wr2} . Unlike the current DA process in which the R matrix varies with X when it needs re-computation during minimization iterations, X^{true} was used to compute the ensemble-derived R matrix for all minimization iterations in Case B (and Case C as well). The purpose for doing so was to imitate the consideration of forward model uncertainty and representativeness error in the ensemble-derived R matrix and R^d 's diagonals so that both the ensemble-derived R matrix and R^d 's diagonals were as close as the true R matrix. Thirdly, the second term of the cost function in Equation (15) was minimized by looping over X to look for the best X_{best} using various $H(A)$, where R^d and Y_1^{single} (or Y_2^{single}) were used here. Note that the $H(A)$ matrix used in building the ensemble-derived R matrix (i.e., the second step) and minimization of the cost function (i.e., the third step) were the same. Finally, the uncertainty of each retrieved x_i in X_{best} could be obtained from the covariance U matrix. The procedure of this analysis is displayed in Figure 6.

The results are presented in Table 2. No $H(A)$ was able to produce an X_{best} consistent with X^{true} because single measured radiance Y_1^{single} (or Y_2^{single}) only represents one fluctuation state around an atmospheric state in equilibrium, introducing deviation from the equilibrium state, and because R^d does not contain full contents. Notwithstanding, the use of a correct H_{cr}^{corr1} (A_{cr}^{corr1}) in minimizing the cost function could produce an X_{best} closest to X^{true} with the smallest J values and the least uncertainty for each retrieved x_i , which again reveals the importance of verifying the correctness of H . With respect to Case A, this study inspired us to think that a single measurement for the radiance spectrum may not be sufficient and R^d may not be appropriate either.

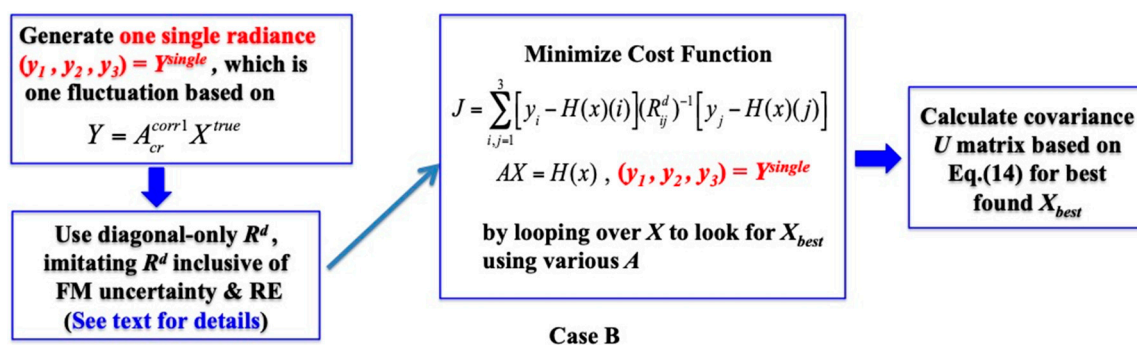


Figure 6. Analysis procedures for Cases B, where FM uncertainty and RE stand for forward model uncertainty and representativeness error, respectively.

Table 2. Results for Case B, where Y_1^{single} (or Y_2^{single}) means data from single measured radiance, the second column means R^d matrix is based on what for construction, and the third column represents what $H(A)$ is used in minimization.

Y	R^d Construction	$H(A)$ in Minimization	$X^{true} = (6.0, 17.0, 9.0)$ Best Found X_{best}	J
Y_1^{single}	$Y_{0.05} = A_{cr}^{corr1} X^{true}$	$H_{cr}^{corr1} (A_{cr}^{corr1})$	$(6.40 \pm 1.20, 19.73 \pm 1.30, 6.22 \pm 0.95)$	0.0007
Y_1^{single}	$Y_{0.05} = A_{wr1} X^{true}$	$H_{wr1} (A_{wr1})$	$(3.88 \pm 2.06, 22.00 \pm 2.12, 4.77 \pm 1.47)$	1.5889
Y_1^{single}	$Y_{0.05} = A_{wr2} X^{true}$	$H_{wr2} (A_{wr2})$	$(2.67 \pm 3.66, 22.00 \pm 3.50, 4.60 \pm 2.12)$	13.0121
Y_2^{single}	$Y_{0.05} = A_{cr}^{corr1} X^{true}$	$H_{cr}^{corr1} (A_{cr}^{corr1})$	$(6.88 \pm 1.28, 16.80 \pm 1.40, 8.27 \pm 0.98)$	0.0007
Y_2^{single}	$Y_{0.05} = A_{wr1} X^{true}$	$H_{wr1} (A_{wr1})$	$(3.51 \pm 1.92, 20.21 \pm 2.02, 6.40 \pm 1.44)$	0.0007
Y_2^{single}	$Y_{0.05} = A_{wr2} X^{true}$	$H_{wr2} (A_{wr2})$	$(1.04 \pm 3.69, 22.02 \pm 3.51, 5.09 \pm 2.12)$	2.4603

6.3. Case C

The analysis procedure in Case C was the same as that in Case B except R^D taken from the ensemble-derived R matrix was adopted. See Figure 7 for the analysis procedure.

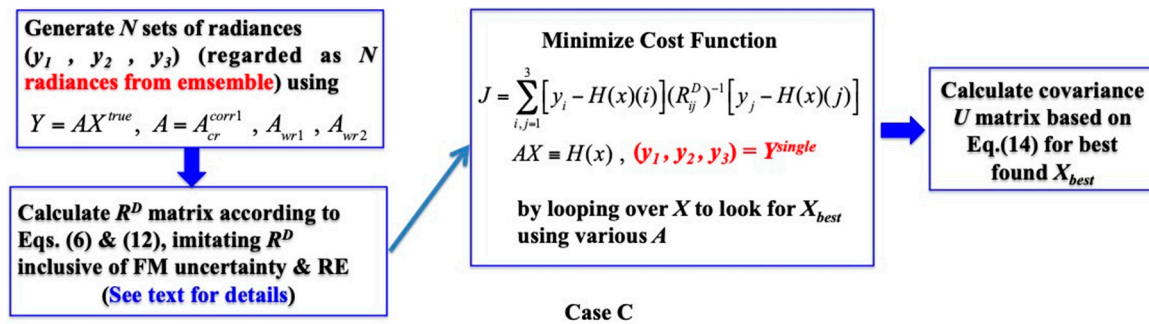


Figure 7. Analysis procedures for Cases C, where FM uncertainty and RE stand for forward model uncertainty and representativeness error, respectively.

The results are shown in Table 3. Another study was conducted using $Y = A_{cr}^{corr1} X^{true}$ for all R^D constructions, the results of which are provided in Table 4. No $H(A)$ was able to produce an X_{best} consistent with X^{true} . Similar to Case B, this occurred because Y_1^{single} (or Y_2^{single}) was used and R^D might not possess good representativeness of the true R matrix. Once again, using a correct $H_{cr}^{corr1} (A_{cr}^{corr1})$ to minimize the cost function could produce an X_{best} closest to X^{true} with the smallest J values and least uncertainty for each retrieved x_i . As such, verification of the correctness of H is important.

Table 3. Results for Case C, where Y_1^{single} (or Y_2^{single}) denotes data from single measured radiance, the second column means the R^D matrix is based on what for construction, and the third column represents what $H(A)$ is used in minimization.

Y	R^D Construction	$H(A)$ in Minimization	$X^{true} = (6.0, 17.0, 9.0)$ Best Found X_{best}	J
Y_1^{single}	$Y_{0.05} = A_{cr}^{corr1} X^{true}$	$H_{cr}^{corr1} (A_{cr}^{corr1})$	$(6.40 \pm 0.29, 19.73 \pm 0.84, 6.23 \pm 0.43)$	0.0008
Y_1^{single}	$Y_{0.05} = A_{wr1} X^{true}$	$H_{wr1} (A_{wr1})$	$(1.83 \pm 0.31, 22.00 \pm 0.87, 4.02 \pm 0.45)$	12.2964
Y_1^{single}	$Y_{0.05} = A_{wr2} X^{true}$	$H_{wr2} (A_{wr2})$	$(1.03 \pm 0.32, 22.00 \pm 0.87, 4.02 \pm 0.45)$	1450.378
Y_2^{single}	$Y_{0.05} = A_{cr}^{corr1} X^{true}$	$H_{cr}^{corr1} (A_{cr}^{corr1})$	$(6.90 \pm 0.27, 16.80 \pm 0.88, 8.27 \pm 0.46)$	0.0007
Y_2^{single}	$Y_{0.05} = A_{wr1} X^{true}$	$H_{wr1} (A_{wr1})$	$(3.52 \pm 0.31, 20.19 \pm 0.78, 6.41 \pm 0.41)$	0.0019
Y_2^{single}	$Y_{0.05} = A_{wr2} X^{true}$	$H_{wr2} (A_{wr2})$	$(1.03 \pm 0.31, 22.01 \pm 0.89, 4.04 \pm 0.46)$	406.2181

Table 4. Results for Case C, where Y_1^{single} (or Y_2^{single}) denotes data from single measured radiance, the second column means the R^D matrix is based on what for construction, and the third column represents what $H(A)$ is used in minimization.

Y	R^D Construction	$H(A)$ in Minimization	$X^{true} = (6.0, 17.0, 9.0)$ Best Found X_{best}	J
Y_1^{single}	$Y_{0.05} = A_{cr}^{corr1} X^{true}$	$H_{cr}^{corr1} (A_{cr}^{corr1})$	$(6.40 \pm 0.29, 19.73 \pm 0.84, 6.23 \pm 0.43)$	0.0008
Y_1^{single}	$Y_{0.05} = A_{cr}^{corr1} X^{true}$	$H_{wr1} (A_{wr1})$	$(2.60 \pm 0.51, 22.00 \pm 1.25, 4.03 \pm 0.56)$	4.6674
Y_1^{single}	$Y_{0.05} = A_{cr}^{corr1} X^{true}$	$H_{wr2} (A_{wr2})$	$(1.02 \pm 1.21, 19.72 \pm 1.82, 4.02 \pm 0.80)$	79.7171
Y_2^{single}	$Y_{0.05} = A_{cr}^{corr1} X^{true}$	$H_{cr}^{corr1} (A_{cr}^{corr1})$	$(6.90 \pm 0.30, 16.80 \pm 0.78, 8.28 \pm 0.45)$	0.0007
Y_2^{single}	$Y_{0.05} = A_{cr}^{corr1} X^{true}$	$H_{wr1} (A_{wr1})$	$(3.52 \pm 0.43, 20.19 \pm 0.09, 6.41 \pm 0.58)$	0.0010
Y_2^{single}	$Y_{0.05} = A_{cr}^{corr1} X^{true}$	$H_{wr2} (A_{wr2})$	$(1.04 \pm 1.15, 19.07 \pm 1.83, 5.26 \pm 0.76)$	24.3265

In summary, the findings from Cases A, B, and C provided some important learnings. Firstly, with the adoption of N real-time measured radiances, constructing the data-derived R matrix from N measured radiances, and the correct observation operator H_{cr}^{corr1} could produce an X_{best} consistent with X^{true} with the smallest J value and almost the least uncertainty for each retrieved x_i (see the first result in each case of using different deviations in Table 1). Secondly, conversely, if R^d or R^D together with single measured Y_1^{single} (or Y_2^{single}) was employed, discrepancy appeared in the retrieved X_{best} compared to X^{true} . Thirdly, as long as a correct H_{cr}^{corr1} was used in the minimization of the cost function, the retrieved X_{best} was closest to X^{true} with the smallest J value and the least uncertainty for each retrieved x_i . This reveals the importance of verifying the correctness of H (or the forward model) used in the DA process. One important result was that fluctuation in the measured radiances did not have any significant impact on the retrieved X_{best} , even when the deviation was up to 40%. This suggests that fluctuation possibly caused by taking N measurements (if any) is not an issue.

Although linear H was employed in our toy simulation studies, the conclusions presented here should be the same even if the real H was non-linear. That is because the nonlinear forward problem could be linearized around the background.

6.4. Case D

The generation of the data-derived R matrix in Case D was the same as that in Case A, where H_{cr}^{corr1} (A_{cr}^{corr1}) and H_{cr}^{corr2} (A_{cr}^{corr2}) were employed for illustration of correlations between any two channels or bands. Once the data-derived R matrix was built, the correlation matrix was acquired based on Equation (13). Figure 8 outlines the analysis procedure.

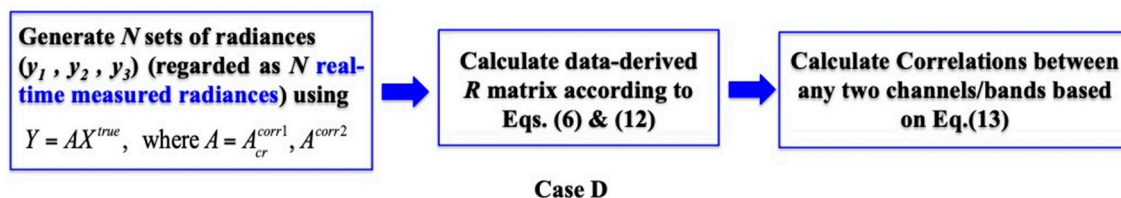


Figure 8. Analysis procedures for Cases D.

The results of the correlations are listed in Table 5. When using H_{cr}^{corr2} (A_{cr}^{corr2}), the off-diagonals in the data-derived R matrix were much smaller than those in the diagonals, resulting in much smaller correlations in the off-diagonals than those in the diagonals. Conversely, if H_{cr}^{corr1} (A_{cr}^{corr1}) was used, the magnitudes in the diagonals and off-diagonals in the data-derived R matrix were comparable, and the correlations in both the diagonals and off-diagonals were also comparable. This study demonstrates that the correlations between any two channels or bands could be manifested from the R matrix itself. Therefore, despite significant or insignificant correlations between any two channels or

bands, the data-derived R matrix proposed in this paper provides a method to acquire correlations between any two channels or bands (i.e., correlations between any two atmospheric physical quantities).

Table 5. Results for Case D, where Y_d^N or Y_d indicate data from N measured radiances, and the second column means the R matrix was based on what for construction.

Y	R Matrix Construction	Data-Derived R Matrix	Correlation Matrix
$Y_{0.05}^N$	$Y_{0.05} = A_{cr}^{corr1} X^{true}$	$\begin{pmatrix} 0.108 & 0.092 & 0.137 \\ 0.092 & 0.115 & 0.152 \\ 0.137 & 0.152 & 0.227 \end{pmatrix}$	$\begin{pmatrix} 1.000 & 0.824 & 0.871 \\ 0.824 & 1.000 & 0.943 \\ 0.871 & 0.943 & 1.000 \end{pmatrix}$
$Y_{0.05}^N$	$Y_{0.05} = A_{cr}^{corr2} X^{true}$	$\begin{pmatrix} 0.085 & 0.025 & 0.004 \\ 0.025 & 0.662 & 0.002 \\ 0.004 & 0.002 & 0.192 \end{pmatrix}$	$\begin{pmatrix} 1.000 & 0.106 & 0.030 \\ 0.106 & 1.000 & 0.006 \\ 0.030 & 0.006 & 1.000 \end{pmatrix}$

7. Summary

Atmospheric state parameter profiles (atmospheric profiles) are key inputs for numerical weather prediction. The observation error covariance (R) matrix factor plays a pivotal role in the retrievals of atmospheric profiles. Since a single measurement is used in atmospheric research analyses and operational weather forecast (ARWF), correct variance and covariance cannot be obtained for the R matrix. As a consequence, the diagonal R^d matrix arranged with instrument noise (plus forward model uncertainty and representativeness error) in the diagonal elements to serve as variances by leaving zeros in off-diagonal has been used in ARWF for years. Recently, a new R^D matrix structure based on the method of Desroziers et al. [52] was introduced and has been adopted in, for example Météo-France and ECMWF 4D-Var systems, and has also been applied to AIRS, IASI, and CrIS data in Met Office 4D-Var systems.

However, the instrument noise acting as variance in R^d 's diagonal led to the dependence of the retrieved ASPs on the instrument being employed. This conflicts with reality in that ASPs should not depend on any instrument. The R^D matrix relies on H , x_a , and the B matrix or x_b , yet the correctness of H requires validation and no true value exists for x_a or x_b . This leads to the R^D matrix being not uniformly symmetrical, resulting in eigenvalues not fully positive-definite, causing less representativeness because R^D 's structure depends on users' assignment to x_a and x_b , users' choices of H , and the selected forecast ensemble that generally cannot describe the atmospheric state at the time when data were captured. Even after diagnosis of the R^D matrix through iterations in the DA process, the diagnosed R^D matrix still has problems as reported in the literature. The problems are caused by the use of single measurement, resulting in an R matrix entangled with B matrix in the DA process.

To unentangle the R and B matrices, the complete isolation of the R matrix from the B matrix is required. To acquire correct variances and covariances for the R matrix, a potential solution is to use a large number (N) of real-time measurements within a short detection time $\Delta\tau^N$ to replace single measurement. An atmospheric state in equilibrium (viewed from macroscopic aspects) can be determined from averaging over many atmospheric variation (or fluctuation) states (viewed from a microscopic aspect). Thus, the R matrix used in the DA system for retrieval of parameters x of an atmospheric state in equilibrium can be achieved by taking many fluctuation states. From a physical viewpoint based on microscopic aspects, each fluctuation state involves interactions between radiation and the particles in the atmosphere, and the interactions comply with the phenomenon of the energy absorption and emission occurring between incident radiation and atoms (or molecules). The wavenumber (i.e., energy) of the outgoing radiation may change after passing through all the layers in the atmosphere, subsequently being received at sensors installed in instruments and converted to radiance to form a radiance spectrum. This suggests correlations exist among bands or channels of a detected atmospheric radiance spectrum. Following probability theory and statistics, a correct R matrix can thus be derived based on N real-time measurements, independent of any assumed quantities, and hence having no ambiguity.

Apart from constructing the R matrix, using N real-time measurements has several other advantages. Unlike single measurement, which is merely a variation around an equilibrium state, the mean of the N measurements can adequately represent the equilibrium state during the detection time $\Delta\tau^N$. Another advantage is that an N real-time measured sample is capable of describing the atmospheric state at the time when the data were captured, whereas forecast ensemble is generally incapable of doing so. A further advantage is that the uncertainty of the N measured radiance spectra conveys some messages. Firstly, finer structures of H would not signify if they were smaller than the uncertainty, equivalently meaning that a more precise H does not mean more accurate retrieval. Secondly, this uncertainty provides us with some information about the uncertainty level that can be tolerated on H when H is linearized, whereas linearization of H may potentially be used to verify the correctness of H using the N measured radiance spectra.

Technically, taking N measurements during a short amount of detection duration $\Delta\tau^N$ is achievable by modifying the trigger configuration of data recording from ground. The traditional trigger configuration captures single measurements successively, but many data are not used in practice. Therefore, to fully use detected data and to reduce the load of the captured data, the proposed trigger configuration involves recording N measurements within $\Delta\tau^N$ with some time between two consecutive N -measurement events. Our study revealed that fluctuation in the measured radiances that could be inferred by taking N measurements does not have any significant impact on the retrieved ASPs.

Several advantages of taking N real-time measurements of radiances for constructing a data-derived R matrix were discussed in this paper. Studies using simulation data with simple toy examples were also presented in this paper. These studies demonstrate that the adoption of N real-time measured radiances, a data-derived R matrix constructed from the N measured radiances, and a correct H can produce accurate ASPs with smallest cost function J value and the least uncertainty for each retrieved ASP. These studies also revealed that verification for correctness of H (or forward model) is necessary.

8. Recommendations for Future Work

As mentioned in Section 1, the retrieval results of ASPs from the DA system for successive weather forecast depend on H , the design of the DA methods, and the information in R and B matrices. In traditional ARWF, entanglement occurs in the DA system among (1) H , which needs verification; (2) DA methods; (3) the R matrix, which relies on instrument noise or H , x_a , and x_b ; and (4) the B matrix whose evolution depends on the NWP (or forecast) model. Furthermore, R and/or B matrices are usually constructed from forecast ensemble rather than real-time data. Thus, even if the R matrix is acquired correctly, the DA system still cannot guarantee producing accurate retrieved ASPs without proper H , correct B , and an appropriate DA method. Therefore, to ensure the whole DA system functions properly for correct production of retrieved ASPs and for next-step weather forecasting, it is a step-by-step course. The first and easiest step would be to fully isolate the R matrix from H , x_a , and x_b (or B matrix) and to independently and directly acquire the R matrix from N real-time measurements as explained in this paper. After that, the following tasks would be verification of H and building the B matrix with the measured N real-time data, with the aim of being free from reliance on the NWP (or forecast) model or historical forecast ensemble. Once R , H , and B are attained without entanglement with one another, the goal would be the determination of the DA method that performs the best.

Lastly, we hope to inspire people to think from a new perspective about which one, single measurement or a large number of measurements, should be used for the R matrix as well as ARWF, and what can be accomplished with a large number of measurements to improve the accuracy of the retrieval of atmospheric state parameters and of numerical weather prediction.

Author Contributions: Conceptualization, Y.-A.L., Z.L., and M.H.; Investigation, M.H.; Methodology, Y.-A.L., Z.L., and M.H.; Formal analysis, M.H.; Writing—original draft preparation, M.H.; Writing—review & editing, Y.-A.L., Z.L., and M.H.; Funding acquisition, Y.-A.L.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 41601469).

Acknowledgments: Since 2013, M.H. has been investigating the issues in atmospheric research and weather forecast. M.H. is grateful to the Space Science and Engineering Center (SSEC) at the University of Wisconsin-Madison (Madison, WI, USA) for providing a good working environment. M.H. is indebted to SSEC's financial support during 2016–2017 for this work. Furthermore, we would like to express our gratitude to the four reviewers for reviewing our manuscript so carefully and providing constructive comments and suggestions. We are also thankful to the academic editor for handling our manuscript and going over our manuscript and responses to the reviewers.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Nomenclature

1. Single measurement refers to the acquisition of one single measurement inside a view angle $\Delta\Omega$ within a short amount of detection time $\Delta\tau^S$ around a limited area Δa by staring at some observation location of interest. Here, one single measurement means to obtain one radiance spectrum, which is adopted for retrieval of atmospheric state parameter profiles. If an interferometer is used, one single measurement indicates the acquisition of one single interferogram used to attain one radiance spectrum.
2. A large number of measurements refer to collecting N (e.g., $N \geq 200$) measurements inside the same $\Delta\Omega$ within a short amount of detection time $\Delta\tau^N$ around the same Δa by staring at some observation location of interest, where each measurement is used to construct one radiance spectrum, and the average of the N radiance spectra along with the variance and covariance is employed for retrieval of atmospheric state parameter profiles.

The superscripts S on $\Delta\tau^S$ and N on $\Delta\tau^N$ represent single measurement and N measurements, respectively. The detection time $\Delta\tau^N$ may be longer than $\Delta\tau^S$ or could be about the same, which could be studied by using real data.

References

1. Thépaut, J.-N. Satellite Data Assimilation in Numerical Weather Prediction: An Overview. Available online: <https://www.ecmwf.int/en/elibrary/12657-satellite-data-assimilation-numerical-weather-prediction-overview> (accessed on 8 September 2003).
2. Taboart, J.M.; Dance, S.L.; Haben, S.A.; Lawless, A.S.; Nichols, N.K.; Waller, J.A. The conditioning of least-squares problems in variational data assimilation. *Numer. Linear Algebr. Appl.* **2018**, *25*, e2165. [[CrossRef](#)]
3. Eliassen, A. *Provisional Report on Calculation of Spatial Covariance and Autocorrelation of the Pressure Field*; Report No. 5; Videnskaps-Akademiets Institutt for Vaer-Og Klimaforskning: Oslo, Norway, 1954; 12p.
4. Lorenc, A.C. A Global Three-Dimensional Multivariate Statistical Interpolation Scheme. *Mon. Weather Rev.* **1981**, *109*, 701–721. [[CrossRef](#)]
5. Le Dimet, F.X.; Talagrand, O. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A* **1986**, *38A*, 97–110. [[CrossRef](#)]
6. Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **1994**, *99*, 10143–10162. [[CrossRef](#)]
7. Liu, C.; Xiao, Q.; Wang, B. An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part I: Technical Formulation and Preliminary Test. *Mon. Weather Rev.* **2008**, *136*, 3363–3373. [[CrossRef](#)]
8. Lorenc, A.C. The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var. *Q. J. R. Meteorol. Soc.* **2003**, *129*, 3183–3203. [[CrossRef](#)]
9. Buehner, M.; McTaggart-Cowan, R.; Beaulne, A.; Charette, C.; Garand, L.; Heilliette, S.; Lapalme, E.; Laroche, S.; Macpherson, S.R.; Morneau, J.; et al. Implementation of Deterministic Weather Forecasting Systems Based on Ensemble-Variational Data Assimilation at Environment Canada. Part I: The Global System. *Mon. Weather Rev.* **2015**, *143*, 2532–2559. [[CrossRef](#)]
10. Gustafsson, N.; Bojarova, J. Four-dimensional ensemble variational (4D-En-Var) data assimilation for the High Resolution Limited Area Model (HIRLAM). *Nonlinear Process. Geophys.* **2014**, *21*, 745–762. [[CrossRef](#)]
11. Wang, X. Incorporating Ensemble Covariance in the Gridpoint Statistical Interpolation Variational Minimization: A Mathematical Framework. *Mon. Weather Rev.* **2010**, *138*, 2990–2995. [[CrossRef](#)]

12. Schwartz, C.S.; Liu, Z.; Lin, H.-C.; Cetola, J.D. Assimilating aerosol observations with a “hybrid” variational-ensemble data assimilation system. *J. Geophys. Res. Atmos.* **2014**, *119*, 4043–4069. [[CrossRef](#)]
13. Schwartz, C.S.; Liu, Z. Convection-Permitting Forecasts Initialized with Continuously Cycling Limited-Area 3DVAR, Ensemble Kalman Filter, and “Hybrid” Variational–Ensemble Data Assimilation Systems. *Mon. Weather Rev.* **2014**, *142*, 716–738. [[CrossRef](#)]
14. Zhang, M.; Zhang, F. E4DVar: Coupling an Ensemble Kalman Filter with Four-Dimensional Variational Data Assimilation in a Limited-Area Weather Prediction Model. *Mon. Weather Rev.* **2012**, *140*, 587–600. [[CrossRef](#)]
15. Gustafsson, N.; Bojarova, J.; Vignes, O. A hybrid variational ensemble data assimilation for the High Resolution Limited Area Model (HIRLAM). *Nonlinear Process. Geophys.* **2014**, *21*, 303–323. [[CrossRef](#)]
16. Clayton, A.M.; Lorenc, A.C.; Barker, D.M. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Q. J. R. Meteorol. Soc.* **2013**, *139*, 1445–1461. [[CrossRef](#)]
17. Kuhl, D.D.; Rosmond, T.E.; Bishop, C.H.; McLay, J.; Baker, N.L. Comparison of Hybrid Ensemble/4DVar and 4DVar within the NAVDAS-AR Data Assimilation Framework. *Mon. Weather Rev.* **2013**, *141*, 2740–2758. [[CrossRef](#)]
18. Bishop, C.H.; Hodyss, D. Adaptive Ensemble Covariance Localization in Ensemble 4D-VAR State Estimation. *Mon. Weather Rev.* **2011**, *139*, 1241–1255. [[CrossRef](#)]
19. Wang, X.; Lei, T. GSI-Based Four-Dimensional Ensemble–Variational (4DEnsVar) Data Assimilation: Formulation and Single-Resolution Experiments with Real Data for NCEP Global Forecast System. *Mon. Weather Rev.* **2014**, *142*, 3303–3325. [[CrossRef](#)]
20. Kleist, D.T.; Ide, K. An OSSE-Based Evaluation of Hybrid Variational–Ensemble Data Assimilation for the NCEP GFS. Part I: System Description and 3D-Hybrid Results. *Mon. Weather Rev.* **2015**, *143*, 433–451. [[CrossRef](#)]
21. Kleist, D.T.; Ide, K. An OSSE-Based Evaluation of Hybrid Variational–Ensemble Data Assimilation for the NCEP GFS. Part II: 4DEnVar and Hybrid Variants. *Mon. Weather Rev.* **2015**, *143*, 452–470. [[CrossRef](#)]
22. Liu, C.; Xiao, Q.; Wang, B. An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part II: Observing System Simulation Experiments with Advanced Research WRF (ARW). *Mon. Weather Rev.* **2009**, *137*, 1687–1704. [[CrossRef](#)]
23. Gustafsson, N.; Janjić, T.; Schraff, C.; Leuenberger, D.; Weissmann, M.; Reich, H.; Brousseau, P.; Montmerle, T.; Wattrelot, E.; Bučánek, A.; et al. Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Q. J. R. Meteorol. Soc.* **2018**, *144*, 1218–1256. [[CrossRef](#)]
24. Bannister, R.N. A review of operational methods of variational and ensemble-variational data assimilation. *Q. J. R. Meteorol. Soc.* **2017**, *143*, 607–633. [[CrossRef](#)]
25. Liu, Y.-A.; Huang, H.-L.A.; Gao, W.; Lim, A.H.N.; Liu, C.; Shi, R. Tuning of background error statistics through sensitivity experiments and its impact on typhoon forecast. *J. Appl. Remote Sens.* **2015**, *9*, 096051. [[CrossRef](#)]
26. Kleist, D.T.; Parrish, D.F.; Derber, J.C.; Treadon, R.; Wu, W.-S.; Lord, S. Introduction of the GSI into the NCEP Global Data Assimilation System. *Weather Forecast.* **2009**, *24*, 1691–1705. [[CrossRef](#)]
27. Weston, P. Progress towards the implementation of correlated observation errors in 4D-Var. In *Forecasting Research Technical Report 560*; Met Office: Exeter, UK, 2011.
28. Rabier, F.; Fourrie, N.; Chafai, D.; Prunet, P. Channel selection methods for Infrared Atmospheric Sounding Interferometer radiances. *Q. J. R. Meteorol. Soc.* **2002**, *128*, 1011–1027. [[CrossRef](#)]
29. Liu, Z.-Q.; Rabier, F. The potential of high-density observations for numerical weather prediction: A study with simulated observations. *Q. J. R. Meteorol. Soc.* **2003**, *129*, 3013–3035. [[CrossRef](#)]
30. Rainwater, S.; Bishop, C.H.; Campbell, W.F. The benefits of correlated observation errors for small scales. *Q. J. R. Meteorol. Soc.* **2015**, *141*, 3439–3445. [[CrossRef](#)]
31. Collard, A.D. On the choice of observation errors for the assimilation of AIRS brightness temperatures: A theoretical study. In *Technical Memorandum AC/90*; ECMWF: Reading, UK, 2004.
32. Stewart, L.M.; Dance, S.L.; Nichols, N.K. Correlated observation errors in data assimilation. *Int. J. Numer. Methods Fluids* **2008**, *56*, 1521–1527. [[CrossRef](#)]
33. Stewart, L.M.; Dance, S.L.; English, S.J.; Eyre, J.R.; Nichols, N.K. *Observation Error Correlations in IASI Radiance Data*; Mathematical Report Series; University of Reading: Reading, UK, 2009; Available online: https://www.researchgate.net/publication/41571544_Observation_error_correlations_in_IASI_radiance_data (accessed on January 2009).

34. Stewart, L.M.; Dance, S.L.; Nichols, N.K. Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. *Tellus A Dyn. Meteorol. Oceanogr.* **2013**, *65*, 19546. [[CrossRef](#)]
35. Stewart, L.M.; Dance, S.L.; Nichols, N.K.; Eyre, J.R.; Cameron, J. Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Q. J. R. Meteorol. Soc.* **2014**, *140*, 1236–1244. [[CrossRef](#)]
36. Bormann, N.; Bauer, P. Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data. *Q. J. R. Meteorol. Soc.* **2010**, *136*, 1036–1050. [[CrossRef](#)]
37. Bormann, N.; Collard, A.; Bauer, P. Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. II: Application to AIRS and IASI data. *Q. J. R. Meteorol. Soc.* **2010**, *136*, 1051–1063. [[CrossRef](#)]
38. Waller, J.A.; Dance, S.L.; Lawless, A.S.; Nichols, N.K. Estimating correlated observation error statistics using an ensemble transform Kalman filter. *Tellus A Dyn. Meteorol. Oceanogr.* **2014**, *66*, 23294. [[CrossRef](#)]
39. Weston, P.P.; Bell, W.; Eyre, J.R. Accounting for correlated error in the assimilation of high-resolution sounder data. *Q. J. R. Meteorol. Soc.* **2014**, *140*, 2420–2429. [[CrossRef](#)]
40. Bormann, N.; Saarinen, S.; Kelly, G.; Thépaut, J.-N. The Spatial Structure of Observation Errors in Atmospheric Motion Vectors from Geostationary Satellite Data. *Mon. Weather Rev.* **2003**, *131*, 706–718. [[CrossRef](#)]
41. Bormann, N.; Geer, A.J.; Bauer, P. Estimates of observation-error characteristics in clear and cloudy regions for microwave imager radiances from numerical weather prediction. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 2014–2023. [[CrossRef](#)]
42. Bormann, N.; Bonavita, M.; Dragani, R.; Eresmaa, R.; Matricardi, M.; McNally, A. Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Q. J. R. Meteorol. Soc.* **2016**, *142*, 1767–1780. [[CrossRef](#)]
43. Campbell, W.F.; Satterfield, E.A.; Ruston, B.; Baker, N.L. Accounting for Correlated Observation Error in a Dual-Formulation 4D Variational Data Assimilation System. *Mon. Weather Rev.* **2017**, *145*, 1019–1032. [[CrossRef](#)]
44. Waller, J.; Ballard, S.; Dance, S.; Kelly, G.; Nichols, N.; Simonin, D. Diagnosing Horizontal and Inter-Channel Observation Error Correlations for SEVIRI Observations Using Observation-Minus-Background and Observation-Minus-Analysis Statistics. *Remote Sens.* **2016**, *8*, 581. [[CrossRef](#)]
45. Waller, J.A.; Simonin, D.; Dance, S.L.; Nichols, N.K.; Ballard, S.P. Diagnosing Observation Error Correlations for Doppler Radar Radial Winds in the Met Office UKV Model Using Observation-Minus-Background and Observation-Minus-Analysis Statistics. *Mon. Weather Rev.* **2016**, *144*, 3533–3551. [[CrossRef](#)]
46. Cordoba, M.; Dance, S.L.; Kelly, G.A.; Nichols, N.K.; Waller, J.A. Diagnosing atmospheric motion vector observation errors for an operational high-resolution data assimilation system. *Q. J. R. Meteorol. Soc.* **2017**, *143*, 333–341. [[CrossRef](#)]
47. Gandin, L. *Objective Analysis of Meteorological Fields*; Translated from the Russian. Jerusalem (Israel Program for Scientific Translations). *Q. J. R. Meteorol. Soc.* **1965**, *92*, 242p.
48. Rutherford, I.D. Data Assimilation by Statistical Interpolation of Forecast Error Fields. *J. Atmos. Sci.* **1972**, *29*, 809–815. [[CrossRef](#)]
49. Hollingsworth, A.; Lönnberg, P. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus A* **1986**, *38A*, 111–136. [[CrossRef](#)]
50. Dee, D.P.; da Silva, A.M. Maximum-Likelihood Estimation of Forecast and Observation Error Covariance Parameters. Part I: Methodology. *Mon. Weather Rev.* **1999**, *127*, 1822–1834. [[CrossRef](#)]
51. Desroziers, G.; Ivanov, S. Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Q. J. R. Meteorol. Soc.* **2001**, *127*, 1433–1452. [[CrossRef](#)]
52. Desroziers, G.; Berre, L.; Chapnik, B.; Poli, P. Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.* **2005**, *131*, 3385–3396. [[CrossRef](#)]
53. Ménard, R. Error covariance estimation methods based on analysis residuals: Theoretical foundation and convergence properties derived from simplified observation networks. *Q. J. R. Meteorol. Soc.* **2016**, *142*, 257–273. [[CrossRef](#)]
54. Rodgers, C.D. *Inverse Methods for Atmospheric Sounding: Theory and Practice*; World Scientific Publishing Pte. Ltd.: Singapore, 2000; 256p.
55. Daley, R. *Atmospheric Data Analysis*; Cambridge University Press: Cambridge, UK, 1993.

56. Rodgers, C.D. Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev. Geophys.* **1976**, *14*, 609–624. [CrossRef]
57. JCGM. Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement. Available online: https://www.bipm.org/utis/common/documents/jcgm/JCGM_100_2008_E.pdf (accessed on 26 July 2019).
58. Dee, D.P. On-line Estimation of Error Covariance Parameters for Atmospheric Data Assimilation. *Mon. Weather Rev.* **1995**, *123*, 1128–1145. [CrossRef]
59. Craven, P.; Wahba, G. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* **1978**, *31*, 377–403. [CrossRef]
60. Golub, G.H.; Heath, M.; Wahba, G. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics* **1979**, *21*, 215–223. [CrossRef]
61. Martin, W.C.; Wiese, W.L. Atomic Spectroscopy—A Compendium of Basic Ideas, Notation, Data, and Formulas. Available online: <https://www.nist.gov/pml/atomic-spectroscopy-compendium-basic-ideas-notation-data-and-formulas> (accessed on 1 November 2017).
62. Yang, P.; Liou, K.-N.; Bi, L.; Liu, C.; Yi, B.; Baum, B.A. On the radiative properties of ice clouds: Light scattering, remote sensing, and radiation parameterization. *Adv. Atmos. Sci.* **2015**, *32*, 32–63. [CrossRef]
63. Cowan, G. *Statistical Data Analysis*; Oxford University Press: Oxford, UK, 1998; 212p.
64. Taylor, J.R. *An Introduction to Error Analysis*, 2nd ed.; University Science Books: Sausalito, CA, USA, 1996; 327p.
65. Bouttier, F.; Courtier, P. Data Assimilation Concepts and Methods. Available online: <https://www.ecmwf.int/node/16928> (accessed on March 1999).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).