Optimal Estimation Retrievals and Their Uncertainties

What Every Atmospheric Scientist Should Know

Maximilian Maahn, David D. Turner, Ulrich Löhnert, Derek J. Posselt, Kerstin Ebell, Gerald G. Mace, and Jennifer M. Comstock

ABSTRACT: Remote sensing instruments are heavily used to provide observations for both the operational and research communities. These sensors do not provide direct observations of the desired atmospheric variables, but instead, retrieval algorithms are necessary to convert the indirect observations into the variable of interest. It is critical to be aware of the underlying assumptions made by many retrieval algorithms, including that the retrieval problem is often ill posed and that there are various sources of uncertainty that need to be treated properly. In short, the retrieval challenge is to invert a set of noisy observations to obtain estimates of atmospheric quantities. The problem is often complicated by imperfect forward models, by imperfect prior knowledge, and by the existence of nonunique solutions. Optimal estimation (OE) is a widely used physical retrieval method that combines measurements, prior information, and the corresponding uncertainties based on Bayes's theorem to find an optimal solution for the atmospheric state. Furthermore, OE also allows the relative contributions of the different sources of error to the uncertainty in the final retrieved atmospheric state to be understood. Here, we provide a novel Python library to illustrate the use of OE for inverse problems in the atmospheric sciences. We introduce two example problems: how to retrieve drop size distribution parameters from radar observations and how to retrieve the temperature profile from ground-based microwave sensors. Using these examples, we discuss common pitfalls, how the various error sources impact the retrieval, and how the quality of the retrieval results can be quantified.

AFFILIATIONS: **Maahn**—Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, and NOAA/Physical Sciences Lab, Boulder, Colorado; **Turner**—NOAA/Global Systems Lab, Boulder, Colorado; **Löhnert and Ebell**—Institute for Geophysics and Meteorology, University of Cologne, Cologne, Germany; **Posselt**—Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California; **Mace**—Department of Atmospheric Science, University of Utah, Salt Lake City, Utah; **Comstock**—Pacific Northwest National Laboratory, Richland, Washington

Science relies on observations to develop theories about nature, and to determine if those theories accurately approximate how nature works. A wide variety of in situ and remote sensing techniques are used to characterize, understand, and quantify properties and processes that occur in the atmosphere and the surface. Improving our understanding of these processes, and how they interact with each other and the environment is critically important for advancing numerical weather prediction and climate models.

In response to this recognized need, our field has seen an explosion in the number and diversity of remote sensing instrumentation. We are using advanced active remote sensors such as lidars, radars of various wavelengths, sodars, scintillometers, and global positioning systems. We are using passive remote sensors like infrared spectrometers, microwave radiometers, and imaging radiometers that operate at wavelengths from the visible to the infrared, and beyond. All of these instruments are taking advantage of various physical laws, many embodied in the principles of radiative transfer, to gain new insights into the processes in the atmosphere.

There is a common thread that holds in most of these observations: we are not actually observing what we want to know. These instruments are measuring a change in voltage, the number of photons passing into a detector over a certain period, a Doppler shift in radar or lidar frequency, and the intensity of the backscattered energy. However, we are generally interested in very specific atmospheric variables, e.g., the ice water content at a certain altitude or within a certain volume, temperature and water vapor at distinct heights, or the aerosol and cloud droplet number concentration. Thus, we are left with the problem of extracting very specific information from the remote sensing observations that are typically only partially related to the variable of interest. In atmospheric sciences, we often call this inverse process a retrieval.

Very often, if we have a measurement that has some sensitivity to the atmospheric variable we desire, we can compute the signal that we would observe with our remote sensor using a so-called forward model, i.e., the forward process. In other words, if we knew the atmospheric state, which includes here all variables affecting the observed signal, we could reproduce the measurement. These forward models ideally are based upon first principles, so that we have a fair degree of confidence in their fidelity. However, they are often nonlinear, which makes them difficult if not impossible to invert analytically. Thus, the retrieval problem is essentially the development of an algorithm that is used to invert the forward model (F) so that we can derive the atmospheric variable that we desire ($\mathbf{x}$; e.g., humidity, temperature, drop size distribution) from the observation that we have made from our remote sensor ($\mathbf{y}$; e.g., brightness temperature, radar reflectivity).

Stephens (1994) provided a classic illustration for a retrieval. Suppose that you desire a description of a dragon but you observe only the footprints that the dragon makes in the sand. Now, if you already know the dragon, you can pretty easily describe the tracks it might make in the sand; i.e., you can develop a forward model. But if you observe only the tracks in the sand, it will be much more difficult to describe the dragon in any detail. You will likely be able to tell that it was a dragon and not a deer, but there will be aspects that you will be unable to characterize: the dragon's color, if it has wings, etc. A retrieval can combine the observations

(large footprints) with prior information (most dragons have wings and the ones making large footprints are green) to get the most likely state (it was a green dragon with wings).

The retrieval process is complicated by many factors. First, there is the uniqueness problem. There is no guarantee that there is only a single $\mathbf{x}$ that maps to the observation $\mathbf{y}$; it is quite possible that $F(\mathbf{x}) = F(\mathbf{x}') = \mathbf{y}$, where $\mathbf{x} \neq \mathbf{x}'$. This could imply that there is a distribution of states around $\mathbf{x}$ that will map to $\mathbf{y}$, but may also mean that there are several very different states mapping to $\mathbf{y}$. Because of this, and because there are often more unknowns than there are measurements, inversion problems are more often than not ill posed. Second, there is no such thing as a perfect instrument implying that $\mathbf{y}$ has some noise component and can also be represented by a probability distribution. Third, there are still uncertainties within the models: either in the physics themselves (single-scattering properties, for instance) or in the ancillary input datasets that are needed to drive the model (along with the atmospheric variable we desire) to simulate the observation. Thus, we have to invert a set of noisy observations using an imperfect forward model where multiple discrete descriptions of the atmospheric state could reproduce the measurements!

Often, a prior dataset (i.e., a climatology) of the atmospheric observations is used together with a forward model to simulate the observations that would be associated with each $\mathbf{x}$. Using a prior constrains ill-posed problems where the most likely state is selected from the possible solutions based on prior information. The accuracy of a retrieval will depend strongly on the prior information the user has before taking the measurement (i.e., the prior dataset with mean $\mathbf{x}_a$ and covariance matrix $\mathbf{S}_a$) used in the construction of the retrieval relative to the particular case being retrieved. If the current case is well represented by the prior dataset (i.e., $\mathbf{x}_{\mathrm{truth}}$ is "close" to $\mathbf{x}_a$), then the retrieval will likely be reasonably accurate, but if the current case is poorly represented in the prior dataset (i.e., $\mathbf{x}_{\mathrm{truth}}$ is "far" from $\mathbf{x}_a$), then the accuracy of the retrieval will likely be poorer, especially if the forward model is nonlinear. This means that during retrieval development that the user must not only choose an adequate prior (e.g., using a tropical dataset as a prior for tropical measurements) but must also have a basic physical understanding about the variable the user wants to retrieve. However, the benefits of adding information from the prior to the retrieval comes at a price and the retrieval can perform poorly when applied to very unlikely cases such as extreme events.

A wide range of techniques has been developed to perform retrievals (e.g., Twomey 1977; Tibshirani 1996; Rodgers 2000; Tarantola 2005; Stephens and Kummerow 2007; Aster et al. 2018), and many scientific disciplines use inversion theory as part of their normal activities. Retrieval algorithms generally fall into two distinct groups: "statistical" and "physical" retrievals[1] (Kidder and Vonder Haar 1995). Statistical retrieval algorithms develop empirical relationships between the atmospheric variable $\mathbf{x}$ and the observation $\mathbf{y}$ using techniques like linear, polynomial or lasso regression algorithms, empirical orthogonal functions, neural networks, and theoretical relationships between scattering, emission, and transmission with the atmospheric properties of interest (e.g., Nakajima and King 1990). These are essentially statistical inferences. While being computationally efficient, statistical retrievals can struggle when dealing with nonunique or nonlinear problems and including all uncertainty sources for estimating retrieval uncertainties can be challenging. In contrast, physical retrievals use and invert a forward model by an iterative process to converge to a solution considering the prior. By this, physical retrievals can deliver more insight into the nature of the inverse problem. This includes quantifying the information content of the measurement, but also consistency checks whether the measurement can be reproduced with the solution.

Physical retrievals exploit the idea, introduced earlier, that observations, forward models, and the assumptions necessary to drive them have inherent uncertainty and can be represented by probability distributions. These algorithms typically use the Bayes's theorem:

[1] For brevity, we omitted hybrid approaches combining both approaches.

$$P(\mathbf{x}|\mathbf{y}) = P(\mathbf{y}|\mathbf{x})P(\mathbf{x})/P(\mathbf{y}), \qquad (1)$$

which states how probable a certain $\mathbf{x}$ is given $\mathbf{y}$ [i.e., $P(\mathbf{x}|\mathbf{y})$]. This probability density function is already the solution for the inverse problem, which means that the solution includes an uncertainty estimate. Bayes's theorem tells us that it can be obtained by multiplying the probability of $\mathbf{y}$ given $\mathbf{x}$ [$P(\mathbf{y}|\mathbf{x})$; the likelihood] with the a priori probability density of the state $P(\mathbf{x})$ (e.g., based on a climatology, a model forecast, or a measurement at some distance away in space or time). The probability $P(\mathbf{y}|\mathbf{x})$ represents the uncertainty in the measurement process and also in the model that converts the state $\mathbf{x}$ into observations $\mathbf{y}$, while the division by $P(\mathbf{y})$ is effectively a normalization factor.

The optimal estimation (OE) is a widely used physical retrieval method and the equations (Rodgers 2000) can be derived from Eq. (1) if one assumes all of the probabilities have a Gaussian distribution. OE also implicitly assumes that there is no bias in the measurements and prior dataset; in other words, the average of their uncertainties is zero. In this study, we illustrate the use and typical challenges of OE, which naturally combines the uncertainties in both the observations and the a priori information. A complete literature overview is beyond the scope of this study, but important applications include atmospheric profiling based on microwave or infrared radiometers (Turner and Löhnert 2014), cloud and precipitation properties (Löhnert et al. 2004; Mace et al. 2016) trace gas concentrations retrievals (Crisp et al. 2004; Wunch et al. 2011), and data assimilation for numerical weather prediction models (Bauer et al. 2006; Janisková 2015). Since retrieved data are utilized so heavily in our field and we continually develop more advanced instrumentation and algorithms, it is important that the retrieval uncertainties are accurately quantified and made available with the retrieved information.

To illustrate how OE works, we discuss two simplified retrieval applications: retrieving temperature and humidity profiles from a multichannel ground-based microwave radiometer and retrieving drop size distribution parameters from ground-based cloud radar observations. We use the recently published Python pyOptimalEstimation library[2] and also provide the code to reproduce all results and figures in supplemental Jupyter notebooks that can be executed in a web browser[3] without installing any software. We strongly encourage the reader to explore the behavior of the example retrieval applications with the Jupyter notebooks on their own. The library, the examples, and the extensive documentation form a toolset that allows the reader to develop their own retrieval algorithms.

How OE works

OE is trying to find an optimal solution $\mathbf{x}_{op}$ given an observation and prior information using an iterative process (Fig. 1) where the prior is described by the mean $\mathbf{x}_a$ and the covariance matrix $\mathbf{S}_a$. We start the iteration with a first guess for the
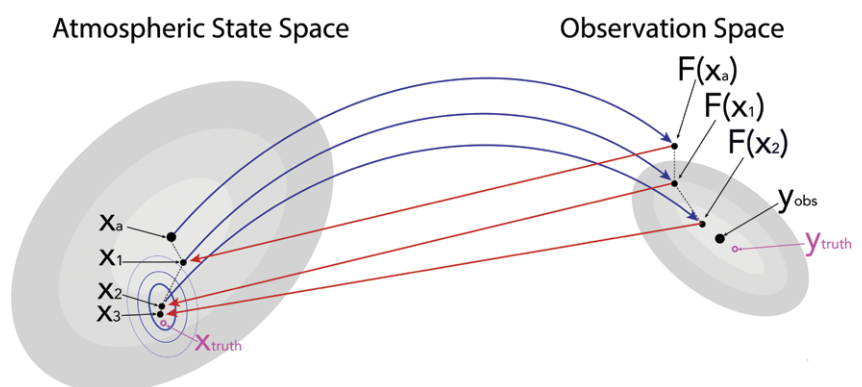

Fig. 1. Optimal estimation principle: The ellipses show the (left) prior state and (right) measurement uncertainty. The iterative process starts with applying the forward operator F to the first guess (here $x_0 = x_a$). Based on the difference of $F(x_a)$ to y_{obs} (which is close to but not equal to an ideal measurement y_{truth} representing x_{truth}), x_1 is obtained (which requires inverting F). This is repeated until the retrieval converges to a solution $x_3 = x_{op}$ that is close to the true state x_{truth}.

atmospheric state $\mathbf{x}_0$. If available, measurements from another instrument, model results, or the retrieval results at an earlier sample time can be used for $\mathbf{x}_0$, but in the absence of other information we follow the common approach to use $\mathbf{x}_a$ for $\mathbf{x}_0$. The forward operator F is used to transform $\mathbf{x}_a$ from $\mathbf{x}$ to $\mathbf{y}$ space, resulting in $F(\mathbf{x}_a)$. In the next step, the difference between $F(\mathbf{x}_a)$ and $\mathbf{y}_{\mathrm{obs}}$ is used to make the next guess $\mathbf{x}_1$ considering the distance from $\mathbf{x}_1$ to $\mathbf{x}_a$. This is done by weighting these differences by the measurement uncertainty covariance matrix $\mathbf{S}_y$ and, respectively, the prior state covariance matrix $\mathbf{S}_a$. The fact that $\mathbf{S}_a$ and $\mathbf{S}_y$ are covariance matrices implies that the errors of $\mathbf{x}_a$ and $\mathbf{y}_{\mathrm{obs}}$ can be described with normal distributions. Obtaining $\mathbf{x}_1$ from $\mathbf{x}_a$ and $\mathbf{y}_{\mathrm{obs}}$ requires inverting F, which is typically not possible analytically. Therefore, the Jacobian matrix $\mathbf{K} = \partial \mathbf{y}/\partial \mathbf{x}$ is used to linearize and invert F, which results in the requirement that F is *moderately linear*. This means that for OE to work, F must behave in approximately a linear manner for small perturbations used to obtain $\mathbf{K}$. In atmospheric science, most problems are moderately linear due to the propensity of exponentials and power laws used within them and the absence of discontinuities. Because $\mathbf{K}$ is only an approximation for F around $\mathbf{x}_0$, $\mathbf{x}_1$ obtained with this process is typically not the solution yet and the process must be repeated until $\mathbf{x}$ converges to a stable, optimal solution $\mathbf{x}_{\mathrm{op}}$. Note that $\mathbf{K}$ depends on $\mathbf{x}$ used for estimating the Jacobian, which means that $\mathbf{K}$ must be recalculated for every iteration step, which might be computationally expensive depending on the lengths of $\mathbf{x}$ and $\mathbf{y}$.

Mathematically, the iterative process for iteration $i + 1$ can be summarized with

$$\mathbf{x}_{i+1} = \mathbf{x}_a + (\mathbf{S}_a^{-1} + \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_e^{-1}\mathbf{K}_i)^{-1}\,\mathbf{K}_i^{\mathrm{T}}\mathbf{S}_e^{-1}[\mathbf{y} - F(\mathbf{x}_i,\mathbf{b}) + \mathbf{K}_i(\mathbf{x}_i - \mathbf{x}_a)], \qquad (2)$$

where $\mathbf{S}_e$ is the *effective* measurement uncertainty. In fact, F depends not only on $\mathbf{x}_i$, but also on the parameter vector $\mathbf{b}$, which includes additional input variables for the forward operator that are not retrieved by the algorithm and are considered fixed. For example, radiative transfer models used for microwave radiometers require not only temperature and humidity profiles (which compose typically $\mathbf{x}$), but also a pressure profile. Additionally, there are almost always other fixed parameters in the forward model that encompass sources of uncertainty; e.g., for microwave radiometers, there are uncertainties in the line parameters that are used in the radiative transfer model (Cimini et al. 2018). Even though often ignored, the uncertainties of $\mathbf{b}$ can be also considered with the corresponding covariance matrix $\mathbf{S}_b'$ (in $\mathbf{y}$ space[4]),which is why $\mathbf{S}_y$ is replaced by the more general $\mathbf{S}_e = \mathbf{S}_y + \mathbf{S}_b'$ in Eq. (2).

The iteration stops when the convergence criterion[5]

$$(\mathbf{x}_i - \mathbf{x}_{i+1})^{\mathrm{T}}\,\mathbf{S}_i^{-1}(\mathbf{x}_i - \mathbf{x}_{i+1}) \ll \mathrm{length}(\mathbf{x}) \qquad (3)$$

is fulfilled. Here, $\mathbf{S}_i$ describes the uncertainty of the retrieved $\mathbf{x}_i$ and can be obtained by combining the uncertainties related to $\mathbf{x}$ and $\mathbf{y}$ space:

$$\mathbf{S}_i = (\mathbf{S}_i^{-1} + \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_e^{-1}\mathbf{K}_i)^{-1}. \qquad (4)$$

Thus, the convergence criterion in Eq. (3) is really stating "the retrieval is considered converged when the change in $\mathbf{x}$ is smaller than the derived uncertainty of $\mathbf{x}$." After convergence is reached at iteration i, $\mathbf{x}_i$ and $\mathbf{S}_i$ are assumed to be the optimal solution $\mathbf{x}_{\mathrm{op}}$ and uncertainty $\mathbf{S}_{\mathrm{op}}$. The fact that $\mathbf{S}_{\mathrm{op}}$ can be easily obtained is an important advantage of OE and other physical retrieval methods in comparison to statistical retrievals where additional steps are required to estimate the uncertainties (e.g., Cadeddu et al. 2009).

Sources of retrieval uncertainty

Every retrieval algorithm suffers from the same fundamental sources of uncertainty related to the three covariance matrices

[4] The matrix $\mathbf{S}_b'$ in $\mathbf{y}$ space can be easily obtained from the covariance of the elements of $\mathbf{b}$ (call this $\mathbf{S}_b$) and then computing the Jacobian of F with respect to $\mathbf{b}$, denoted as $\mathbf{K}_b$ with $\mathbf{S}_b' = \mathbf{K}_b\mathbf{S}_b\mathbf{K}_b^{T}$.

[5] There is also a criterion to test for convergence in $\mathbf{y}$ space instead of $\mathbf{x}$ space [Rodgers 2000, their Eq. (5.33)], but we found that our sample retrievals obtain convergence faster in $\mathbf{x}$ space without any significant difference in the quality of the solution (see supplement A).

$\mathbf{S}_y$, $\mathbf{S}_a$, and $\mathbf{S}_y$. To illustrate this, we designed a simplified retrieval (see supplement A) for atmospheric temperature and humidity profiles based on brightness temperature measurements of a common ground-based microwave radiometer [humidity and temperature profiler (HATPRO); Rose et al. 2005].

Here, the prior temperature and humidity profiles consist of a 16-yr spring average (March–May) of radiosonde observations at the U.S. Department of Energy Atmospheric Radiation Measurement (ARM) program at the North Slope of Alaska (NSA) site in Utqiaġvik (formerly known as Barrow), Alaska (Verlinde et al. 2016). As a forward operator, we use a fast and simple radiative transfer model that does not account for scattering by hydrometeors (Löhnert et al. 2004). We do not use real radiometer observations but obtain a synthetic measurement based on a radiosonde and the forward operator. This has two main advantages: first, systematic forward model errors cancel each other out and—more importantly—we know the true atmospheric state $\mathbf{x}_{\text{truth}}$, which we can compare to the retrieval's result $\mathbf{x}_{\text{op}}$. Therefore, we always recommend using synthetic observations during retrieval development before applying it to real observations. Standard values for the uncertainties in the observed microwave brightness temperatures are used for the diagonal elements of $\mathbf{S}_y$; other error sources are neglected in the beginning. We want to stress, however, that the retrieval accuracy can be significantly reduced when using real observations. This is associated to "real world" challenges such as biases in the measurements, the representatives of the prior data, or the quantification of forward model errors. An extensive discussion of OE retrievals based on real observations can be found, e.g., in Ebell et al. (2017) or Martinet et al. (2017).

Uncertainties related to the prior. The prior dataset serves as a tremendous constraint for the entire retrieval; in fact, the uncertainty in the retrieval must be smaller than the uncertainty in the prior. Otherwise, the retrieval does not add any information from the observations. However, if we do not have a good measure of the variability and covariance of the true atmospheric state, then the retrieval could easily give a biased result. In some cases, the prior does not serve as a serious constraint because the information content in the observations is so high, such as retrieving liquid water path from a microwave radiometer (Turner et al. 2007). In other cases, it is difficult to derive the $\mathbf{S}_a$ matrix from existing observations (e.g., the vertical level-to-level covariance of liquid water content), and thus model simulations (Löhnert et al. 2001) or educated guesses (Maahn et al. 2015) are used to provide this information to serve as a constraint in the retrieval. However, for retrieving temperature and humidity profiles from microwave radiometer (MWR) observations, the prior is critically important. In supplement A, we are trying to obtain 120 atmospheric quantities (60 height levels for each temperature and humidity from 0 to 22,500 m) from only 14 radiometer measurements, which makes the retrieval problem underconstrained. The discrepancy between the lengths of $\mathbf{x}$ and $\mathbf{y}$ underlines why a prior is required to get useful information out of the radiometer.[6]

Here, we modify the MWR retrieval to assess the impact of using a suboptimal $\mathbf{S}_a$ that is not consistent with the prior dataset. In the atmosphere, temperature and humidity are correlated to each other and to themselves at various height levels. This makes the nondiagonal elements of $\mathbf{S}_a$ crucial (Fig. 2): The retrieval considering nondiagonal elements (reference run) converges to a $\mathbf{x}_{\text{op}}$ with a reasonable agreement to $\mathbf{x}_{\text{true}}$ and a significant reduction of the uncertainties (Fig. 2a). If we assume that there are no level-to-level correlations in the prior (i.e., by setting nondiagonal elements to zero), the humidity solution stays close to the prior, the uncertainties are hardly reduced with respect to the prior, and there is even a nonphysical artifact in the humidity profile (Fig. 2b).

The result $\mathbf{x}_{\text{op}}$ is also biased when a wrong prior dataset is used (Fig. 2c): When using summer radiosonde observations

[6] However, if we were trying to retrieve only 14 atmospheric quantities, the prior would still be critical because passive radiometer measurements are typically correlated.
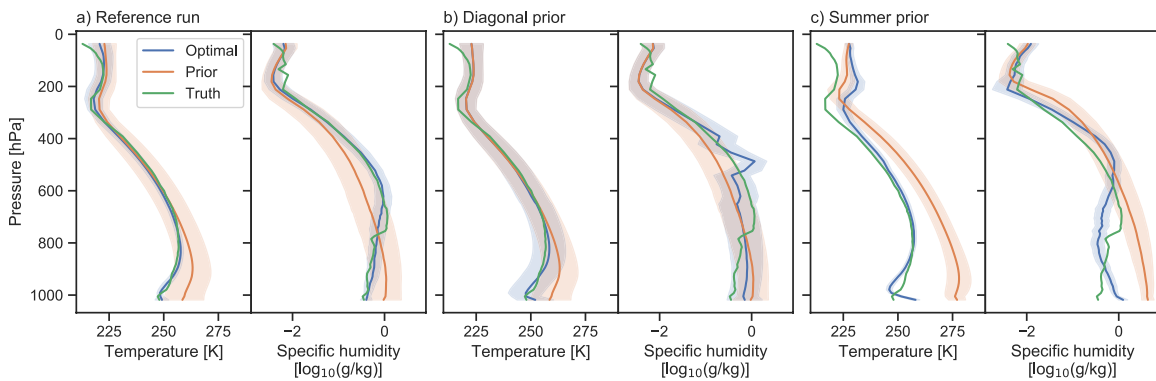
Fig. 2. True (green), prior (orange), and optimal (blue) profile for (left) temperature and (right) specific humidity. (a) The correct spring $\mathbf{S}_a$ (reference run) was used, (b) the nondiagonal entries of $\mathbf{S}_a$ were neglected (set to zero), and (c) the wrong $\mathbf{S}_a$ (summer instead of the correct spring) was used. The shading shows the uncertainties (1 sigma) related to the prior and optimal.

(June–August) to obtain $\mathbf{x}_a$ and $\mathbf{S}_a$, $\mathbf{x}_{\text{truth}}$ is mostly outside of the 1-sigma range of $\mathbf{S}_a$. The result $\mathbf{x}_{\text{op}}$ differs from $\mathbf{x}_{\text{truth}}$ mostly in the upper atmosphere (200–300 hPa) and close to the surface where $\mathbf{x}_{\text{op}}$ bends to $\mathbf{x}_a$. But how can the reader detect such a situation for a real-world application when $\mathbf{x}_{\text{truth}}$ is unknown? To identify issues with $\mathbf{S}_a$, we strongly advise that the user apply statistical χ^2 tests (see supplement A) to test whether the retrieved measurement $\mathbf{y}_{\text{op}}$ agrees with $\mathbf{y}_{\text{obs}}$ and whether $\mathbf{y}_{\text{op}}$, $\mathbf{y}_{\text{obs}}$, and $\mathbf{x}_{\text{op}}$ are in agreement with the prior state considering the assumed uncertainties. These tests fail for the retrieval ignoring nondiagonal elements or using a wrong prior.

Uncertainties related to the forward operator. The second source of uncertainty is associated with the forward model. While forward models are usually built on first principle understanding, many approximations still exist either due to limitations in our understanding of the physics or for computational efficiency reasons. The use of a 1D radiative transfer model instead of a 3D radiative transfer model is one example of an approximation that is frequently used, even though the 3D effects are important in most cloud scenes (e.g., Liang and Girolamo 2013). Furthermore, the inputs and assumptions needed by F often consist of more than just the elements of $\mathbf{x}$ that are represented by the parameter vector $\mathbf{b}$ and its covariance matrix $\mathbf{S}_b$. These so-called model parameters need to be known as their uncertainties also contribute to the uncertainty of F.

The forward model uncertainties are typically ignored in most current retrieval algorithms (i.e., $\mathbf{S}_e = \mathbf{S}_y$) assuming that the forward model is perfect. However, for some cases, ignoring the uncertainties associated with $\mathbf{b}$ is the Achilles heel of the retrieval itself, which can result in a drastic underestimation of the error in the retrieved $\mathbf{x}$ (e.g., Mace et al. 2016). This can be also seen for our MWR retrieval. Because the uncertainties of our forward operator have not been specifically quantified yet, we rely on the results of Cimini et al. (2018), who analyzed the uncertainties in the gas spectroscopy used in our forward microwave radiative transfer model. They found significantly correlated errors for some channels, particularly around the oxygen line. When adding their $\mathbf{S}_b'$ (their Fig. 9) to our $\mathbf{S}_e$, the balance between $\mathbf{S}_e$ and $\mathbf{S}_a$ shifts and more emphasis is put on $\mathbf{S}_a$. This leads to larger uncertainties for $\mathbf{x}_{\text{op}}$ so that the optimal to prior uncertainty ratio[7] is increased by up to 11 percentage points (Fig. 3a).

This can also be seen as a reduction in the information content that is retrieved from the measurements that can be quantified using the degrees of freedom for signal d.[8] These

[7] Defined as sqrt[diag($\mathbf{S}_{\text{op}}$)/diag($\mathbf{S}_a$)].

[8] We obtain d from the averaging kernel $\mathbf{A} = \mathbf{I} - \mathbf{S}_{\text{op}}\mathbf{S}_a^{-1}$ (Turner et al. 2016). The elements of the diagonal of $\mathbf{A}$ show the degrees of freedom for a signal that can be obtained for the individual elements of $\mathbf{x}$. We define d as the trace of $\mathbf{A}$. In case diag($\mathbf{S}_{\text{op}}\mathbf{S}_a^{-1}$) $\rightarrow$ 0, d reaches the maximum value of length of $\mathbf{x}$.

describe the number of independent information pieces that can be obtained from the measurement, given the prior state. For a retrieval where the observations have perfect information content, d would be equal to the number of measurements or state variables, whichever is smaller. The reference retrieval can retrieve a total of $4.67d$ (of a maximum value of length of $\mathbf{y} = 14$) for the humidity and temperature profiles combined. Also d can be separated by state variables (Fig. 3b), which shows that the highest information contents can be obtained for temperature close to the surface and for humidity around 500 hPa. By including the forward model uncertainty $\mathbf{S}_b'$ in the retrieval, the total d is reduced to 4.01, with the primary loss of information occurring in the middle atmosphere (Fig. 3b).

Fig. 3. (a) The ratio of the uncertainty in the optimal solution (i.e., the square root of the diagonal of $\mathbf{S}_{op}$) to the uncertainty in the prior (i.e., the square root of the diagonal of $\mathbf{S}_a$) and (b) individual degrees of freedom for signal. The statistics for temperature are in green and specific humidity are in red. The solid line is the reference run with $\mathbf{S}_e = \mathbf{S}_y$ and the dashed–dotted line is for the configuration where the forward operator uncertainty is added $\mathbf{S}_e = \mathbf{S}_y + \mathbf{S}_b'$

Uncertainties related to the measurement.
The perhaps most obvious source of uncertainty in a retrieval is the uncertainty in the observations themselves. $\mathbf{y}$ is typically multidimensional, i.e., there are observations from multiple channels from a single instrument, or multiple instruments used in the retrieval. There will be random uncorrelated errors for each particular element of $\mathbf{y}$, as well as correlated errors between the different elements of $\mathbf{y}$. This can be concisely specified in a covariance matrix $\mathbf{S}_y$. Similar to $\mathbf{S}_a$, making $\mathbf{S}_y$ larger (smaller) changes the weights in the retrieval and puts more (less) weight on the prior at the expense of the observations. In supplement A, we detail how quadrupling (i.e., doubling the measurement uncertainty deviation) $\mathbf{S}_e$ reduces the retrieved d by 0.75 and leads to $\mathbf{x}_{op}$ being closer to $\mathbf{x}_a$.

For many instruments, it is easier to estimate the uncorrelated error in each particular element of the observation. For example, the basic measurement of microwave radiometers is typically a voltage output by the receiver where the random uncertainty is related to the noise characteristics of the receiver. But errors can be also correlated among measurements, e.g., if multiple measurements are obtained by the same receiver subject to common amplifier elements. Determining this correlated error is typically much harder but can be done (Tobin et al. 2007; Maahn and Löhnert 2017; Turner and Blumberg 2019). In our MWR retrieval, we also neglected nondiagonal elements in $\mathbf{S}_y$ initially. To assess the impact, we included the nondiagonal elements to the $\mathbf{S}_y$ used in $\mathbf{S}_e$ (not accounting for $\mathbf{S}_b$). Both the random noise and the correlation of the random noise between the HATPRO frequencies have been determined by continuously measuring the brightness temperatures of a blackbody with a known temperature.

Even though the changes in retrieval uncertainty and the element-wise d are small in comparison to the reference run that assumed $\mathbf{S}_y$ to be diagonal, the resulting total d is increased by 0.19. Therefore, accounting for nondiagonal elements is one of the few occasions where considering "additional" error sources leads to an enhancement instead of a decrease of retrieval quality because correlations between the different elements of $\mathbf{y}$ (i.e., the different channels in the microwave radiometer) can reduce the effective noise of the instrument. Referring to Fig. 1, correlated uncertainties would result in smaller shaded areas around $\mathbf{y}_{obs}$ compared to uncorrelated uncertainties having the same variances.
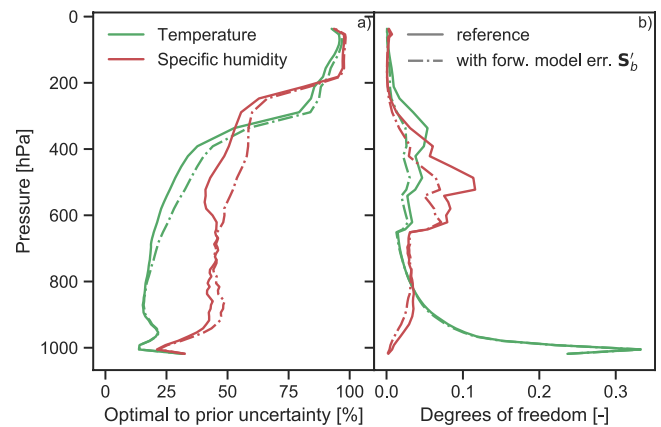
Impact of retrieval assumptions

So far, we have investigated error sources related to the individual components of OE. But what happens when the core assumptions about OE are violated? OE requires that the forward operator be moderately linear[9] (i.e., linear when perturbing $\mathbf{x}$) and that the measurement as well as the prior error can be described by normal (Gaussian) distributions. We have assumed that the long-term averaged radiosonde profile prior dataset represents the un-biased population truth. Thus, the distribution of prior data itself must follow a normal distribution.

We illustrate the importance of assuming normal distributions by introducing a second example (see supplement B for details) using radar reflectivity Z_e and mean Doppler velocity V_d of a vertically pointing cloud radar to retrieve a raindrop size distribution (DSD). We parameterize the DSD using a scaling parameter N_w, the raindrop mass-spectrum mean diameter D_m and the normalized mass spectrum standard deviation σ_m' (Williams et al. 2014). These three quantities form the state vector $\mathbf{x}$ while Z_e and V_d are combined into the measurement vector $\mathbf{y}$. The a priori information ($\mathbf{x}_a$ and $\mathbf{S}_a$) is obtained from the observations from three disdrometers deployed near Huntsville, Alabama, from December 2001 to October 2011 (Williams et al. 2014). Because violating retrieval assumptions will likely impact the robustness of the retrieval, we analyze the retrieval for 100 randomly chosen profiles. Similar to the first example, we use synthetic observations $\mathbf{y}_{obs}$ derived with a radar simulator (Mech et al. 2020) that are not included in the prior.

One option to deal with nonnormally distributed prior data are to transform the data so that it follows a normal distribution more closely. In supplement B, we use quantile–quantile (QQ) plots to show that the prior $\mathbf{x}$ follows a normal distribution closer when applying a logarithm to $\mathbf{x}$. Note, this is a rather simple solution for making the distributions more normal; however, some retrieval problems require more complex transformations (e.g., Anscombe 1973; Mason et al. 2017). When comparing the retrieval quality for linear and logarithmic $\mathbf{x}$, a clear improvement in retrieval quality can be seen for the latter: the percentage of converged retrievals raises from 86% to 100%. This also applies to the number of profiles passing the statistical χ^2 tests (see "Uncertainties related to the prior" section) and a linearity test that tests the assumption that the forward operator is moderately linear. Further, we show with another QQ plot in supplement B that the distribution of "real" retrieval errors of the optimal solution ($\mathbf{x}_{truth} - \mathbf{x}_{op}$) is in better agreement with the estimated $\mathbf{S}_{op}$ when using a logarithmic $\mathbf{x}$.

Another way to show the enhanced quality is to look into the distributions of the optimal to prior uncertainty ratio and the individual d for all profiles (Fig. 4). The clear reduction in spread when using the logarithm of the state variables is obvious indicating that the retrieval produces more consistent results that do not depend as much on the individual profile. The median values of the optimal to prior uncertainty ratio distribution are lower for the linear state variables indicating that the linear retrieval version is in general underestimating retrieval uncertainties. At the same time, d can be greatly overestimated for individual profiles.

[9] If the uncertainties in either the observation space or the state space are highly non-Gaussian, then other techniques such as the Markov chain Monte Carlo (MCMC; Posselt et al. 2008) method should be used.
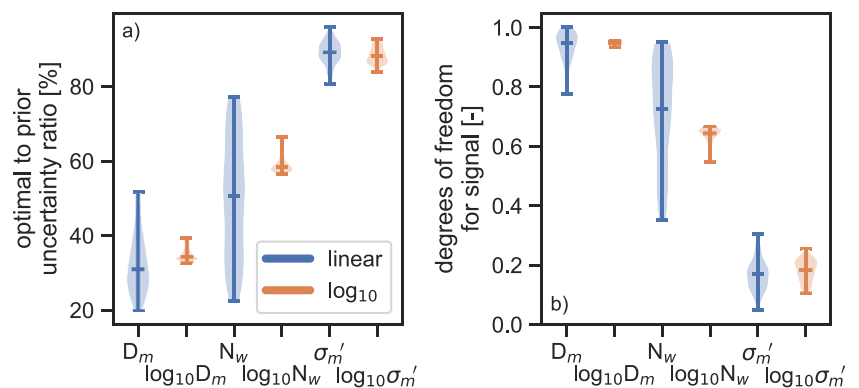

Fig. 4. Violin plots showing (a) the optimal to prior uncertainty ratio and (b) the individual d using the linear (blue) and logarithmic state variables (orange).

Summary and concluding remarks

Most of the observations of the Earth–atmosphere system are from remote sensors (e.g., satellite observations, weather radar observations) that are used in a wide range of operational and research applications. For all of these sensors, atmospheric variables of interest are never directly observed but must be derived from the observations within an inversion algorithm, which often requires the use of prior constraints. The uncertainties in these retrieved quantities arise from three sources: the prior dataset used to either construct or constrain the retrieval (characterized by the covariance matrix $\mathbf{S}_a$), the uncertainties in the forward model assumptions and input ancillary datasets used in the forward model ($\mathbf{S}_b$), and the uncertainties in the observations themselves ($\mathbf{S}_y$). Within the inversion algorithm, these uncertainties can be propagated to provide uncertainties in the retrieved atmospheric state $\mathbf{x}_{op}$ given by the covariance matrix $\mathbf{S}_{op}$. In a series of examples, we have utilized the optimal estimation (OE) retrieval method to show the impact of how changes in these uncertainties result in changes in the retrieved state vector $\mathbf{x}_{op}$ and $\mathbf{S}_{op}$. We also show that non-Gaussian uncertainty distribution can lead to a degradation of the OE performance. Nonnormally distributed state variables should be normalized to avoid negative impacts on retrieval quality and robustness. Using forward operators that are grossly nonlinear (i.e., are not moderately nonlinear) will also lead to a decrease of accuracy.

The supplemental Jupyter Notebooks, which can be run online in a web browser,[10] provide the complete code for all examples. We strongly encourage the readers to experiment with the examples by themselves. Together with the pyOptimalEstimation Python library,[11] this gives the readers all information to get started with their own OE retrieval projects.

We contend that more work is needed to accurately characterize these three input covariance matrices. Too often, our community assumes that the measurement covariance matrix $\mathbf{S}_y$ is diagonal and that there is no correlation between different measurements within the observational vector $\mathbf{y}$. We challenge instrument developers to devise methods to test this assumption, and for scientists that develop retrievals to include the results from these tests in their algorithms. However, it must be stressed that such a detailed error characterization may not lead to a retrieval improvement if other error sources such as measurement biases are not correctly determined and accounted for before the retrieval application.

[10] https://github.com/maahn/pyOptimalEstimation_examples

[11] https://github.com/maahn/pyOptimalEstimation

The prior dataset is critical as a constraint for physical retrievals. Often, however, the data needed to build a well-characterized prior dataset are inadequate or simply unavailable. For example, there are very few observational datasets available that allow us to determine the level-to-level covariance of cloud microphysical properties, which is critical information that is needed for cloud property retrievals, and therefore the community is using other sources such as model simulations.

Also, we believe that the uncertainties and assumptions in the forward model parameters have been neglected for too long. Forward models may be fundamentally incorrect (e.g., applying 1D radiative transfer approaches to situations that are inherently 3D), or may have uncertainties in the model parameters that affect retrieval results. In some applications, the uncertainty in $\mathbf{b}$ can dominate the uncertainty in the retrieval. Characterizing the uncertainties in F may require additional supporting data so that closure studies can be performed. Thus, a well-constructed set of field campaigns may be able to improve the characterization of both $\mathbf{S}_a$ and $\mathbf{S}_b$ simultaneously.

It is important to recognize that OE is just one tool, albeit a powerful one, that can be used to retrieve atmospheric information from remote sensing observations. If the limitations of OE make it inapplicable to a problem, other physical retrieval approaches such as the computationally expensive Markov chain Monte Carlo method (Posselt and Mace 2014; Posselt

et al. 2017) are more appropriate for highly non-Gaussian cases or where the forward model is highly nonlinear.

Last, even the best retrievals can be only as good as the underlying observations stressing the need for enhanced instruments that can constrain retrievals better; this could be achieved using new instrument concepts, improved designs, or smarter sensor synergies.

References

Anscombe, F. J., 1973: Graphs in statistical analysis. *Amer. Stat.*, **27**, 17–21, https://doi.org/10.1080/00031305.1973.10478966.

Aster, R. C., B. Borchers, and C. H. Thurber, 2018: *Parameter Estimation and Inverse Problems.* Elsevier, 406 pp.

Bauer, P., P. Lopez, A. Benedetti, D. Salmond, and E. Moreau, 2006: Implementation of 1D+4D-Var assimilation of precipitation-affected microwave radiances at ECMWF. I: 1D-Var. *Quart. J. Roy. Meteor. Soc.*, **132**, 2277–2306, https://doi .org/10.1256/qj.05.189.

Cadeddu, M. P., D. D. Turner, and J. C. Liljegren, 2009: A neural network for real-time retrievals of PWV and LWP from Arctic millimeter-wave ground-based observations. *IEEE Trans. Geosci. Remote Sens.*, **47**, 1887–1900, https://doi .org/10.1109/TGRS.2009.2013205.

Cimini, D., P. W. Rosenkranz, M. Y. Tretyakov, M. A. Koshelev, and F. Romano, 2018: Uncertainty of atmospheric microwave absorption model: Impact on ground-based radiometer simulations and retrievals. *Atmos. Chem. Phys.*, **18**, 15 231–15 259, https://doi.org/10.5194/acp-18-15231-2018.

Crisp, D., and Coauthors, 2004: The Orbiting Carbon Observatory (OCO) mission. *Adv. Space Res.*, **34**, 700–709, https://doi.org/10.1016/j.asr.2003.08.062.

Ebell, K., U. Löhnert, E. Päschke, E. Orlandi, J. H. Schween, and S. Crewell, 2017: A 1-D variational retrieval of temperature, humidity, and liquid cloud properties: Performance under idealized and real conditions. *J. Geophys. Res. Atmos.*, **122**, 1746–1766, https://doi.org/10.1002/2016JD025945.

Giangrande, S., and T. Toto, 2016: Interpolated sonde (interpolatedsonde). ARM user facility, accessed 11 January 2019, https://doi.org/10.5439/1095316.

Janisková, M., 2015: Assimilation of cloud information from space-borne radar and lidar: Experimental study using a 1D+4D-Var technique. *Quart. J. Roy. Meteor. Soc.*, **141**, 2708–2725, https://doi.org/10.1002/qj.2558.

Kidder, S. Q., and T. H. Vonder Haar, 1995: *Satellite Meteorology: An Introduction.* 1st ed. Academic Press, 466 pp.

Liang, L., and L. D. Girolamo, 2013: A global analysis on the view-angle dependence of plane-parallel oceanic liquid water cloud optical thickness using data synergy from MISR and MODIS. *J. Geophys. Res. Atmos.*, **118**, 2389–2403, https://doi.org/10.1029/2012JD018201.

Löhnert, U., S. Crewell, C. Simmer, and A. Macke, 2001: Profiling cloud liquid water by combining active and passive microwave measurements with cloud model statistics. *J. Atmos. Oceanic Technol.*, **18**, 1354–1366, https://doi.org/10.1175/1520-0426(2001)018<1354:PCLWBC>2.0.CO;2.

——, ——, and ——, 2004: An integrated approach toward retrieving physically consistent profiles of temperature, humidity, and cloud liquid water. *J. Appl. Meteor.*, **43**, 1295–1307, https://doi.org/10.1175/1520-0450(2004)043<1295:AIATRP>2.0.CO;2.

Maahn, M., and U. Löhnert, 2017: Potential of higher-order moments and slopes of the radar Doppler spectrum for retrieving microphysical and kinematic properties of Arctic ice clouds. *J. Appl. Meteor. Climatol.*, **56**, 263–282, https://doi.org/10.1175/JAMC-D-16-0020.1.

——, ——, P. Kollias, R. C. Jackson, and G. M. McFarquhar, 2015: Developing and evaluating ice cloud parameterizations for forward modeling of radar moments using in situ aircraft observations. *J. Atmos. Oceanic Technol.*, **32**, 880–903, https://doi.org/10.1175/JTECH-D-14-00112.1.

Mace, G. G., S. Avey, S. Cooper, M. Lebsock, S. Tanelli, and G. Dobrowalski, 2016: Retrieving co-occurring cloud and precipitation properties of warm marine boundary layer clouds with A-Train data. *J. Geophys. Res. Atmos.*, **121**, 4008–4033, https://doi.org/10.1002/2015JD023681.

Martinet, P., D. Cimini, F. D. Angelis, G. Canut, V. Unger, R. Guillot, D. Tzanos, and A. Paci, 2017: Combining ground-based microwave radiometer and the AROME convective scale model through 1DVAR retrievals in complex terrain: An Alpine valley case study. *Atmos. Meas. Tech.*, **10**, 3385–3402, https://doi.org/10.5194/amt-10-3385-2017.

Mason, S. L., J. C. Chiu, R. J. Hogan, and L. Tian, 2017: Improved rain rate and drop size retrievals from airborne Doppler radar. *Atmos. Chem. Phys.*, **17**, 11 567–11 589, https://doi.org/10.5194/acp-17-11567-2017.

Mech, M., M. Maahn, S. Kneifel, D. Ori, E. Orlandi, P. Kollias, V. Schemann, and S. Crewell, 2020: PAMTRA 1.0: A Passive and Active Microwave Radiative Transfer tool for simulating radiometer and radar measurements of the cloudy atmosphere. *Geosci. Model Dev. Discuss.*, https://doi.org/10.5194/gmd-2019-356.

Nakajima, T., and M. D. King, 1990: Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. Part I: Theory. *J. Atmos. Sci.*, **47**, 1878–1893, https://doi.org/10.1175/1520-0469(1990)047<1878:DOTOTA>2.0.CO;2.

Petersen, W. A., A. Tokay, P. N. Gatlin, and M. T. Wingo, 2010: GPM ground validation two-dimensional video disdrometer (2DVD) NSSTC. Global Hydrology Resource Center, accessed 11 January 2019, https://doi.org/10.5067/GPMGV/NSSTC/2DVD/DATA201.

Posselt, D. J., and G. G. Mace, 2014: MCMC-based assessment of the error characteristics of a surface-based combined radar-passive microwave cloud property retrieval. *J. Appl. Meteor. Climatol.*, **53**, 2034–2057, https://doi.org/10.1175/JAMC-D-13-0237.1.

——, T. S. L'Ecuyer, and G. L. Stephens, 2008: Exploring the error characteristics of thin ice cloud property retrievals using a Markov chain Monte Carlo algorithm. *J. Geophys. Res.*, **113**, D24206, https://doi.org/10.1029/2008JD010832.

——, J. Kessler, and G. G. Mace, 2017: Bayesian retrievals of vertically resolved cloud particle size distribution properties. *J. Appl. Meteor. Climatol.*, **56**, 745–765, https://doi.org/10.1175/JAMC-D-16-0276.1.

Project Jupyter and Coauthors, 2018: Binder 2.0—Reproducible, interactive, sharable environments for science at scale. *Proc. 17th Python Sci. Conf.*, Austin, TX, Enthought, 113–120, https://doi.org/10.25080/Majora-4af1f417-011.

Rodgers, C. D., 2000: *Inverse Methods for Atmospheric Sounding: Theory and Practice.* World Scientific, 240 pp.

Rose, T., S. Crewell, U. Lohnert, and C. Simmer, 2005: A network suitable microwave radiometer for operational monitoring of the cloudy atmosphere. *Atmos. Res.*, **75**, 183–200, https://doi.org/10.1016/j.atmosres.2004.12.005.

Stephens, G. L., 1994: *Remote Sensing of the Lower Atmosphere: An Introduction.* Oxford University Press, 562 pp.

——, and C. D. Kummerow, 2007: The remote sensing of clouds and precipitation from space: A review. *J. Atmos. Sci.*, **64**, 3742–3765, https://doi.org/10.1175/2006JAS2375.1.

Tarantola, A., 2005: *Inverse Problem Theory and Methods for Model Parameter Estimation.* Society for Industrial and Applied Mathematics, 349 pp.

Tibshirani, R., 1996: Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc.*, **58B**, 267–288, https://doi.org/10.1111/J.2517-6161.1996.TB02080.X.

Tobin, D. C., P. B. Antonelli, H. E. Revercomb, S. T. Dutcher, D. D. Turner, J. K. Taylor, R. O. Knuteson, and K. H. Vinson, 2007: Hyperspectral data noise characterization using principle component analysis: Application to the atmospheric infrared sounder. *J. Appl. Remote Sens.*, **1**, 013515, https://doi.org/10.1117/1.2757707.

Turner, D. D., and U. Löhnert, 2014: Information content and uncertainties in thermodynamic profiles and liquid cloud properties retrieved from the ground-based Atmospheric Emitted Radiance Interferometer (AERI). *J. Appl. Meteor. Climatol.*, **53**, 752–771, https://doi.org/10.1175/JAMC-D-13-0126.1.

——, and W. G. Blumberg, 2019: Improvements to the AERIoe thermodynamic profile retrieval algorithm. *IEEE J. Sel. Top. Appl. Remote Sens. Earth Obs.*, **12**, 1339–1354, https://doi.org/10.1109/JSTARS.2018.2874968.

——, S. A. Clough, J. C. Liljegren, E. E. Clothiaux, K. E. Cady-Pereira, and K. L. Gaustad, 2007: Retrieving liquid water path and precipitable water vapor from the Atmospheric Radiation Measurement (ARM) microwave radiometers. *IEEE Trans. Geosci. Remote Sens.*, **45**, 3680–3690, https://doi.org/10.1109/TGRS.2007.903703.

——, S. Kneifel, and M. P. Cadeddu, 2016: An improved liquid water absorption model at microwave frequencies for supercooled liquid water clouds. *J. Atmos. Oceanic Technol.*, **33**, 33–44, https://doi.org/10.1175/JTECH-D-15-0074.1.

Twomey, S., 1977: The influence of pollution on the shortwave albedo of clouds. *J. Atmos. Sci.*, **34**, 1149–1152, https://doi.org/10.1175/1520-0469(1977)034<1149:TIOPOT>2.0.CO;2.

Verlinde, J., B. D. Zak, M. D. Shupe, M. D. Ivey, and K. Stamnes, 2016: The ARM North Slope of Alaska (NSA) sites. *The Atmospheric Radiation Measurement (ARM) Program: The First 20 Years*, Meteor. Monogr., No. 57, Amer. Meteor. Soc., 8.1–8.13, https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0023.1.

Williams, C. R., and Coauthors, 2014: Describing the shape of raindrop size distributions using uncorrelated raindrop mass spectrum parameters. *J. Appl. Meteor. Climatol.*, **53**, 1282–1296, https://doi.org/10.1175/JAMC-D-13-076.1.

Wunch, D., and Coauthors, 2011: The total carbon column observing network. *Philos. Trans. Roy. Soc.*, **369A**, 2087–2112, https://doi.org/10.1098/RSTA.2010.0240.