Atmospheric
Measurement
Techniques

Propagating information content: an example with advection

David D. Turner[1], Maria P. Cadeddu[2], Julia M. Simonson[1,3,4], and Timothy J. Wagner[5]

[1]NOAA/Global Systems Laboratory, Boulder, CO, USA
[2]Argonne National Laboratory, Argonne, IL, USA
[3]Cooperative Institute for Research in the Environmental Sciences, University of Colorado, Boulder, CO, USA
[4]Developmental Testbed Center, Boulder, CO, USA
[5]Space Science and Engineering Center, University of Wisconsin – Madison, Madison, WI, USA

Correspondence: David D. Turner (dave.turner@noaa.gov)

Abstract. The mathematical algorithm to derive geophysical information from remote sensing observations is called a retrieval. The mathematics of many retrieval problems are ill-posed, and thus a priori information is used to help constrain the derived geophysical variable to realistic values. One quantity of interest, therefore, is the information content of the observation. Perfect information content in the observation would be achieved if the retrieval were able to capture any perturbation in the desired geophysical variable with the proper magnitude.

Many new data products can be derived by combining geophysical variables retrieved from multiple different remote sensors. This paper explores, for the first time, how to derive the information content of these derived products. The approach uses traditional error propagation techniques to derive the uncertainty of the derived field twice, both when the observations are used in the retrieval and also when only the a priori information from each remote sensor is propagated. These two uncertainties are then used to provide an estimate of the information content of the derived geophysical variable.

This study demonstrates how to propagate the uncertainties from six different instruments to provide the information content for water vapor and temperature advection. A multi-month analysis demonstrates that, in a mean sense, the information content for temperature advection is nearly unity for all heights below 700 m while, the information content for water vapor advection is somewhat more variable but still larger than 0.6 in the convective boundary layer.

1 Introduction

Observations are absolutely essential for science and understanding nature. They can serve both as the source of ideas (e.g., "This is an interesting observation; I wonder what it means?") and means to evaluate hypotheses (e.g., "My model suggests this is true; can I make an observation that confirms that the model is correct?"). In both of these cases, it is critical to understand the uncertainty in the observation in order to correctly interpret the result.

Observations used in the natural sciences take advantage of many different physical principles. Some of these instruments are considered "in situ"; in other words, the instrument makes its measurement of the desired geophysical variable at the point of interest. Other instruments are remote sensors, where the instrument is placed some distance from the location of interest. Remote sensors come in two general types: (a) active remote sensors, wherein the instrument transmits some signal such as electromagnetic energy or sound towards the measurement volume and analyzes the portion of the energy that is scattered from the volume towards the detector, and (b) passive remote sensors, which observe scattered or emitted signals (typically electromagnetic radiation) from the measurement volume.

Seldom do we measure the actual geophysical variable that we desire; instead, virtually all instruments measure a signal that provides information that is related by some physical process to the variable we desire (Maahn et al., 2020). For example, a simple mercury-based thermometer provides a measure of temperature as the depth of the mercury in a vacated tube is directly proportional to the temperature due to

the thermal expansion of the mercury in the reservoir; characterizing the measurement uncertainty for instruments like this is reasonably straightforward.

Deriving geophysical variables from the observations made by remote sensors is more challenging. Generally, we have a physically based "forward model" (denoted as F) that relates the geophysical variable we want to observe with our actual measurement; thus, the retrieval problem is essentially deriving the inverse F to map from the observation to our desired geophysical variables. However, deriving the geophysical variables of interest from observations made by passive remote sensors is often a mathematically ill-posed problem; i.e., there are often many possible values for the geophysical variables that would map through F to our observation, especially given that there is always uncertainty in the observation. Thus, we use additional a priori information (i.e., information collected before the observation is made) to constrain the retrieval. This is not a new endeavor: scientists have been retrieving information from passive remote sensors for many decades (e.g., Smith et al., 1970), with the mathematical development of these "inverse methods" preceding it (e.g., Twomey, 1966). A good high-level overview of different retrieval methods is given by Maahn et al. (2020), but a large number of detailed texts exist that explore the retrieval, or inverse theory (e.g., Tarantola, 2005; Rodgers, 2000).

The challenge with retrievals from passive remote sensors is understanding not only the uncertainty in the retrieval, but also the information content that is offered by the observation itself given that there is also some contribution from the a priori constraint. Westwater and Strand (1968) provide a concise definition for information content vis à vis retrievals: "The information content... is defined as a reduction in the uncertainty in the (retrieval) after the (observations) are introduced." The information content of a remote sensing measurement is not a new concept, but is an important one as it allows the user of the retrieval to understand how many independent pieces of information are in the observation itself and how that information is distributed among the geophysical variables that are retrieved.

Often, geophysical variables retrieved from remote sensors are used to derive estimates of other geophysical variables. An example of this is deriving convective available potential energy (CAPE) from a ground-based passive remote sensor from which thermodynamic profiles are retrieved. Blumberg et al. (2017) demonstrated how to use Monte Carlo sampling of the posterior covariance matrix of the retrieved profile to estimate a large number of profiles that would technically satisfy the radiance observation, computed CAPE from each profile, and then estimated the uncertainty in CAPE by looking at the distribution of the CAPE values provided by the Monte Carlo sampling. Monte Carlo sampling is a computational approach to estimate uncertainties, but a more traditional error propagation approach could have also been adopted (e.g., the "error analysis" chapter in Bevington and Robinson, 2003).

However, there are times when geophysical variables are derived using multiple remote sensors, each with their own uncertainties and information contents. In this paper, we explore the idea of propagating uncertainties and information content from multiple instruments through the derivation equation to provide the uncertainties and information content of the derived quantity. To our knowledge, this is the first time this has been demonstrated for information content.

We chose the recent work by Wagner et al. (2022) that derives the profiles of horizontal water vapor and temperature advection using a network of ground-based instruments to illustrate this approach. This paper will first explain our method to propagate the information content, perform a detailed examination of a single case, and then provide a more statistical description of the information content in the derived advection products.

2 Observations used to compute advection

Horizontal advection occurs when there is a spatial gradient in a scalar variable over and upwind of desired region, which is then advected over the region. Wagner et al. (2022) used the network of profiling ground-based remote sensors at the Department of Energy's Atmospheric Radiation Measurement (ARM; Turner and Ellingson, 2016) Southern Great Plains (SGP; Sisterson et al., 2016) site in north-central Oklahoma. At each of the network sites, there are two instruments: the Atmospheric Emitted Radiance Interferometer (AERI; Knuteson et al., 2004) and a Doppler wind lidar (DL; Pearson et al., 2009). The AERI is a passive infrared spectrometer, and thus the "TROPoe" algorithm is used to retrieve thermodynamic profiles above the instrument (Turner and Löhnert, 2014; Turner and Blumberg, 2019). The DL is an active remote sensor that measures radial velocities along the direction of the outgoing laser beam, and by scanning the lidar in a velocity–azimuth display (VAD; Browning and Wexler, 1968) manner (i.e., making measurements at a number of different azimuth directions at a constant elevation angle), profiles of horizontal winds can be derived (e.g., Newsom et al., 2017). Key to this study is that we have a full error covariance matrix for all data used in the analysis. We will first describe the two datasets, then discuss how advection is derived from them.

2.1 Temperature, humidity, and wind retrievals

The thermodynamic and wind profiles were both retrieved using physical-iterative retrieval methods that are based upon Gaussian statistics; this is usually referred to colloquially as "optimal estimation". Optimal estimation approaches provide an error covariance matrix, denoted $\mathbf{S}_x$, which embodies the uncertainty for each retrieval. In this study, temperature and humidity retrievals use the TROPoe algorithm, which was explained in Turner and Löhnert (2014). These retrievals

were done at 5 min resolution. Two separate studies have reframed the derivation of the horizontal winds from VAD scans using a retrieval approach, thereby constraining the derived winds with a priori information (Baidar et al., 2023; Gebauer and Bell, 2024). We elected to use the Baidar et al. (2023) wind retrievals, which we will refer to as "DLoe." While the Gebauer and Bell algorithm allows for the inclusion of non-Doppler lidar observations in the wind profile retrieval, in the absence of those additional observations the approach essentially defaults to that of Baidar et al. (2023). The temporal resolution of the DLoe data was 15 min. Therefore, we do not expect the results shown here to have any significant dependence on which DL retrieval algorithm was used.

Following the nomenclature of Rodgers (2000), we will represent the covariance of the a priori information as $\mathbf{S}_\mathrm{a}$, the uncertainty in the observations as the covariance matrix $\mathbf{S}_\mathrm{m}$, and the sensitivity of the forward model F as $\boldsymbol{K} = \partial F / \partial \boldsymbol{x}$, where $\boldsymbol{x}$ represents the geophysical variables we are retrieving. The posterior covariance matrix of the retrieval of $\boldsymbol{x}$ is then

$$\mathbf{S}_x = \left(K^T \mathbf{S}_\mathrm{m}^{-1} K + \mathbf{S}_\mathrm{a}^{-1} \right)^{-1} \tag{1}$$

as given by Eq. (3.31) in Rodgers (2000), where the superscript T in this context denotes matrix transpose. Most practitioners only show the square root of the diagonal of $\mathbf{S}_x$, as this represents the 1σ uncertainties at that level for that variable; however, the off-diagonal elements of $\mathbf{S}_x$ represent the covariance in the uncertainties between levels and/or variables and will be important for this study.

The averaging kernel of the retrieval provides a wealth of information about the retrieved quantities. Again, following Eq. (3.28) in Rodgers (2000), the averaging kernel of the retrieval is computed as

$$A = \mathbf{S}_x K^T \mathbf{S}_\mathrm{m}^{-1} K. \tag{2}$$

The diagonal of the A is extremely important, as it provides a measure of the degrees of freedom for signal (DFS) that the observations provide to the retrieval for each variable (and as we are working with profiles here, each element of the diagonal is for a specific height above the ground), with the sum of the diagonal (i.e., the trace of A) being the total DFS for the entire retrieval. The DFS at a given height ranges from 0 (i.e., there is no information in the observations) to 1 (i.e., there is perfect information content in the observations). The latter implies that, if there was a perturbation to the state vector (i.e., the true atmospheric values of the variables we are retrieving), then the retrieval would perfectly capture the magnitude of that perturbation. In other words, the DFS quantifies the information content for each variable that is being retrieved in the vector $\boldsymbol{x}$. Other definitions of information content are possible that are related to the DFS. For example, the Shannon information content (Shannon, 1948) is related to the decrease in entropy between the a priori estimate

(S_a) and the estimate after the measurement (i.e., $\mathbf{S}_x$). In this work, however, we limit to the analysis of DFS, which we will discuss in more detail in Sect. 3.

The AERI instruments have diminishing information content on the thermodynamic profile above 1 km, with very little information above 3 km. However, as illustrated in Turner and Blumberg (2019), observations from other instruments can be added as part of the observation vector to improve the retrieved solution (and thus increase its information content). The TROPoe retrievals used for this work, which were processed by the ARM data center, included AERI radiance data, microwave radiometer brightness temperatures at 23.8 and 31.4 GHz, surface observations of temperature and relative humidity, and temporally interpolated profiles of temperature and water vapor mixing ratio from the radiosondes launched roughly every 6 h at the SGP C1 facility as part of the observation vector. However, to prevent overfitting to the C1 radiosondes, the uncertainties in the radiosonde temperature profiles were assumed to be 20 °C and 5 g kg^{-1} at the surface, decreasing linearly to 4.5 °C and 1 g kg^{-1} at 3 km, to allow the algorithm to place more emphasis on the high-time-resolution remote sensors as part of the retrieved profile. The uncertainties in the collocated microwave radiometer brightness temperatures were assumed to be 0.3 K for each frequency. Thus, virtually all of the information content in the TROPoe retrievals below 1.5 km is from the AERI instruments, but the microwave radiometer (MWR) and radiosondes start to have more influence above 2 km (especially for the retrieved water vapor profile).

2.2 Case study: 13 June 2019

For this study, we will only use the "down" triangle (Fig. 1) discussed in Wagner et al. (2022), namely, the triangle that is created by the site near Waukomis, OK (E37; located at 36.311° N, 97.928° W), the central facility (C1; located at 36.606° N, 97.485° W), and the site near Morrison, OK (E39; located at 36.374° N, 97.069° W). Note that the distances from E37 to C1, C1 to E39, and E39 to E37 are approximately 50, 45, and 78 km, respectively.

Figure 2 provides an example of the retrieved temperature ($\boldsymbol{T}$; panels a1, a2, a3) and water vapor ($\boldsymbol{q}$; panels b1, b2, and b3) profiles over the SGP site for 13 June 2019 at the E37, C1, and E39 sites. Similarly, the DLoe-retrieved winds are shown in Fig. 2, with $\boldsymbol{u}$ (panels c1, c2, and c3) and $\boldsymbol{v}$ (panels d1, d2, and d3) winds from the three sites, respectively. The thermodynamic and kinematic evolution looks qualitatively very similar across the three sites over this day; however, there are small variations in the retrieved profiles that affect the calculated advection (shown in the next subsection).

Note that for TROPoe, the algorithm simultaneously retrieves both $\boldsymbol{T}$ and $\boldsymbol{q}$ (i.e., $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{T}, \boldsymbol{q} \end{bmatrix}^T$) so that the posterior covariance matrix $\mathbf{S}_{tq}$ (computed using Eq. 1) includes the level-to-level covariances of temperature to temperature, water vapor to water vapor, and temperature to water vapor.
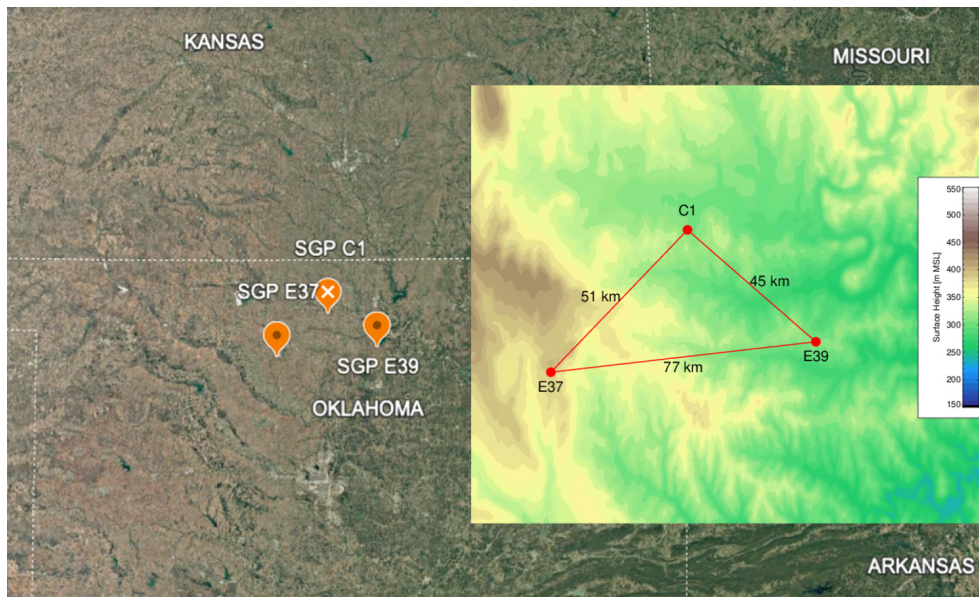
Figure 1. A satellite image of north-central Oklahoma, showing the locations of the three sites used in this analysis (©Google Earth). The inset shows the three sites, including the distances between them, where the background color indicates the elevation across the domain.

Similarly, the DLoe algorithm simultaneously retrieves the u and v wind components (i.e., $x = [u, v]^T$), and thus its posterior covariance matrix S_{uv} includes the cross-correlated errors between u and v. For TROPoe, S_m was approximated as a diagonal matrix from the AERI radiance uncertainties (Turner and Blumberg, 2019). For DLoe, S_m was specified as a diagonal matrix based upon the DL's signal-to-noise ratio at each height (Baidar et al., 2023). The uncertainties in the retrieved T and q as well as u and v winds were derived from the square root of the diagonals of S_{tq} and S_{uv}, respectively, and are shown in Fig. 3.

The temperature retrievals show very small uncertainties, less than 0.5 °C, below 500 m that increase to nearly 1.5 °C near 2 km above ground level (Fig. 3a1, a2, and a3). Note that differences in the noise characteristics among the AERIs will result in differences in the retrieval uncertainties; this is also true for the Doppler lidar systems. High-frequency variation is seen in the uncertainties for the C1 temperature retrieval (Fig. 3 panel a2), which is a result of temporal variation in the uncertainties in the downwelling infrared radiation observations in the temperature-sensitive spectral region observed by the AERI at that site. The water vapor uncertainty data are qualitatively very similar across the three sites, with again the lowest uncertainties below 500 m (Fig. 3b1, b2, and b3). The uncertainties in the u and v winds from the DLs are relatively low below about 1.5 km, but above that height the uncertainties increase drastically due to the poor signal-to-noise ratio in the DL radial velocity observations above 1.5 km due to a relative lack of aerosols. However, Fig. 3 (panels c1 and d1) also demonstrates that the DL at the E37 site has much poorer data quality (i.e., higher instrument noise levels) relative to

the DLs at the other two sites (Fig. 3c2, c3, d2, and d3), with the uncertainties in the retrieved winds from the E37 DL being much larger than for the other two DLs (Fig. 3c2, c3, d2, d3), especially above 1 km. It is important to note here that the a priori information used for TROPoe and DLoe was identical at the three sites; thus, the variability across the three sites in the derived uncertainties and information is due to the differences in the instrument uncertainties at the different locations.

2.3 Advection

Michael (1994) demonstrated that the horizontal advection of a scalar φ ($\underset{\varphi}{\leftrightarrow}$) is computed as a line integral around the network of observations that outlines a polygon in space. Wagner et al. (2022) extended this to vertical remote sensors to get profiles of horizontal advection. For this work here, we used observations at three sites (the triangle shown in Fig. 1), and the advection was computed as

$$\underset{\varphi}{\leftrightarrow} = \frac{-\sum_1^3 \overline{\varphi}\,(\overline{u}\Delta y - \overline{v}\Delta x)}{A_{\text{triangle}}}, \tag{3}$$

where A_{triangle} is the area of the triangle (in m^2), the summation is over each leg of the triangle, the averaged quantities ($\overline{\varphi}$, $\overline{u}$, and $\overline{v}$) are computed from the two vertices that make up that leg, and Δx and Δy are the distances between the two vertices of that leg in the zonal and meridional directions (in m) (Wagner et al., 2022). Using Eq. (3) as the forward model F, we compute the profiles of advection of $\underset{T}{\leftrightarrow}$ and $\underset{q}{\leftrightarrow}$ simul-
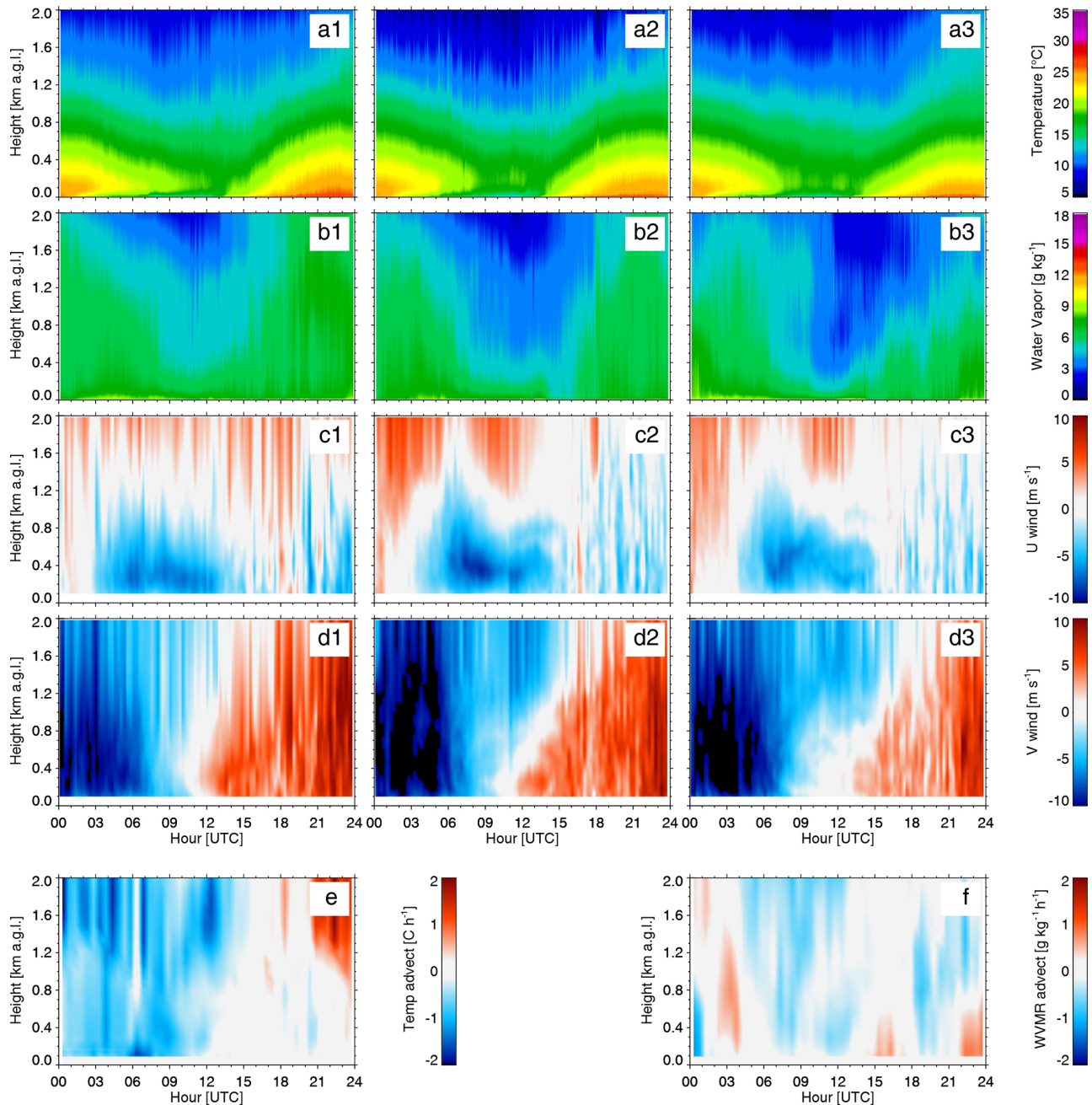
Figure 2. Time–height cross-sections of temperature and humidity retrieved from the AERI instruments (rows **a** and **b**), as well as u and v winds retrieved from the DLs (rows **c** and **d**), at the E37 (column 1), C1 (column 2), and E39 (column 3) sites for 13 June 2019. The derived temperature advection and water vapor advection fields are in panels (**e**) and (**f**), respectively. The time (x axis) is UTC; local time is UTC−5.

taneously as

$$\overset{\leftrightarrow}{x} = F\left(x_{\text{E37}}, x_{\text{C1}}, x_{\text{E39}}, w_{\text{E37}}, w_{\text{C1}}, w_{\text{E39}}\right), \qquad (4)$$

where $\overset{\leftrightarrow}{x} = \begin{pmatrix} \overset{\leftrightarrow}{T} \\ \overset{\leftrightarrow}{q} \end{pmatrix}$, $x = \begin{pmatrix} T \\ q \end{pmatrix}$, and $w = \begin{pmatrix} u \\ v \end{pmatrix}$, where T, q, u, and v are all profiles that have been interpolated to the same vertical grid (as defined by the TROPoe retrievals of T and q).

Temperature and water vapor advection fields calculated for 13 June 2019 are shown in Fig. 2 (panels e and f, respectively). Perhaps the most notable feature seen in this example is the deep layer of cold air advection that ends at 12:00 UTC (Fig. 2e) when the meridional wind direction changes from northerly to southerly (Fig. 2d1, d2, and 3), which corresponds nicely to the change in the synoptic pattern (Fig. 4). Also, there are small pulses of positive water vapor advection
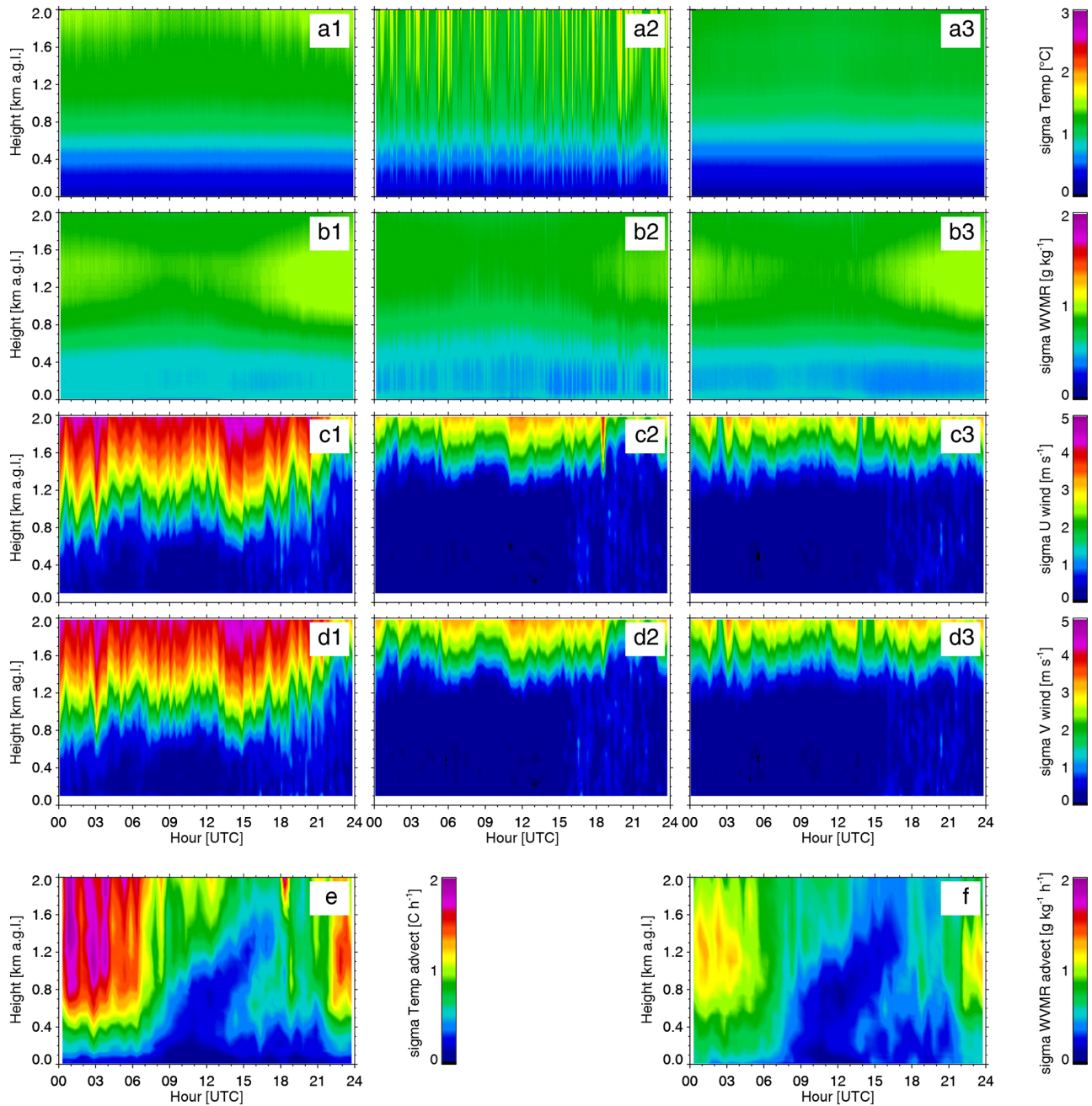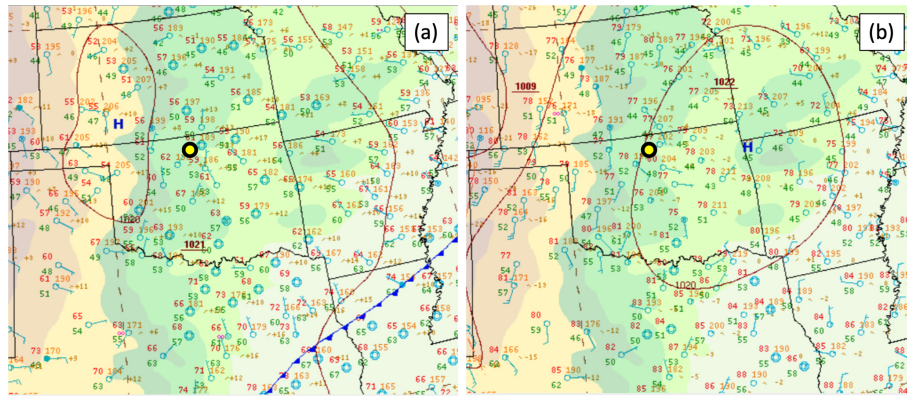
Figure 3. Same as Fig. 2, but showing the 1σ uncertainties of the various products, which were computed as the square root of the diagonal of the posterior covariance matrices.

at 03:00 and 15:00 UTC (Fig. 2f) that are associated with the onset and demise of the easterly component of the wind (seen in Fig. 2c1, c2, and c3).

The advantage of this formulation (Eq. 4) is that the uncertainties in the advection of $\boldsymbol{T}$ and $\boldsymbol{q}$ (i.e., $\sigma_{\overset{\leftrightarrow}{x}}$) can be easily estimated using standard error propagation techniques (e.g., Bevington and Robinson, 2003) as

$$
\sigma_{\overset{\leftrightarrow}{x}}^2 = \sigma_{x_{\mathrm{E37}}}^2 \left(\frac{\partial F}{\partial x_{\mathrm{E37}}} \right)^2 + \sigma_{x_{\mathrm{C1}}}^2 \left(\frac{\partial F}{\partial x_{\mathrm{C1}}} \right)^2 + \sigma_{x_{\mathrm{E39}}}^2
$$
$$
\left(\frac{\partial F}{\partial x_{\mathrm{E39}}} \right)^2 + \sigma_{w_{\mathrm{E37}}}^2 \left(\frac{\partial F}{\partial w_{\mathrm{E37}}} \right)^2 + \sigma_{w_{\mathrm{C1}}}^2 \left(\frac{\partial F}{\partial w_{\mathrm{c1}}} \right)^2
$$
$$
+ \sigma_{w_{\mathrm{E39}}}^2 \left(\frac{\partial F}{\partial w_{\mathrm{E39}}} \right)^2, \tag{5}
$$

Figure 4. Surface synoptic maps at 06:00 **(a)** and 18:00 UTC **(b)** on 13 June 2019. The yellow dot indicates the approximate location of the ARM C1 site. From the NOAA Weather Prediction Center map archive at https://www.wpc.ncep.noaa.gov/archives/web_pages/sfc/sfc_archive.php (last access: 13 May 2025).

since there are no correlated errors among any of the six different instruments. Writing this in terms of covariance matrices, we get

$$\mathbf{S}_{\underset{x}{\leftrightarrow}} = K_{x,\mathrm{E37}}^T \mathbf{S}_{x,\mathrm{E37}} K_{x,\mathrm{E37}} + K_{x,\mathrm{C1}}^T \mathbf{S}_{x,\mathrm{C1}} K_{x,\mathrm{C1}}$$
$$+ K_{x,\mathrm{E39}}^T \mathbf{S}_{x,\mathrm{E39}} K_{x,\mathrm{E39}} + K_{w,\mathrm{E37}}^T \mathbf{S}_{w,\mathrm{E37}} K_{w,\mathrm{E37}}$$
$$+ K_{w,\mathrm{C1}}^T \mathbf{S}_{w,\mathrm{C1}} K_{w,\mathrm{C1}} + K_{w,\mathrm{E39}}^T \mathbf{S}_{w,\mathrm{E39}} K_{w,\mathrm{E39}}, \quad (6)$$

where K is the Jacobian of F for both $\boldsymbol{x}$ and $\boldsymbol{w}$, computed at the three different sites, and the superscript T in this context represents matrix transpose. Note that the translation of Eqs. (5) to (6) uses the fact that $K_x = \partial F / \partial x$ and that the covariance of $\boldsymbol{x}$ can be written as σ_x^2 or as the matrix $\mathbf{S}_x$.

Using this approach, the uncertainties in the retrieved thermodynamic and wind profiles were propagated to provide the uncertainties in the temperature and moisture advection (i.e., $\mathbf{S}_{\underset{x}{\leftrightarrow}}$) for the observations on 13 June 2019 using Eq. (6). Time–height cross-sections of the 1σ uncertainties of the temperature and moisture advection are shown in Fig. 3e and f, respectively. The uncertainties in both the temperature and water vapor advection for this day are small near the surface (less than $0.3\,\mathrm{K\,h^{-1}}$ and $0.5\,\mathrm{g\,kg^{-1}\,h^{-1}}$, respectively) and generally increase with height. In particular, the uncertainty in the magnitude of the temperature advection above $1.2\,\mathrm{km}$ from 00:00 to 07:00 UTC is quite large (larger than $1.5\,\mathrm{K\,h^{-1}}$), suggesting that the cold air advection shown in this time period (Fig. 2e) has a lot of uncertainty. However, there are particularly low uncertainties in the derived temperature and water vapor advection from approximately 09:00 to 15:00 UTC from the surface to nearly 1500 m that seem associated with the change in the synoptic pattern (i.e., the change in direction of the low-level winds shown in Fig. 4).

3 Information content propagation

For retrievals, the concept of information content is used to express how much information we are gaining through the measurements over our prior knowledge. Information content heavily depends on the instrument's noise level and the sensitivity of the forward model. It is known that the information content in the AERI thermodynamic retrievals is not large and that it decreases rapidly with height above the surface (e.g., Turner and Löhnert, 2014, 2021); the sum of the total information content from the surface to 3 km (i.e., $\sum_{z=0\,\mathrm{km}}^{z=3\,\mathrm{km}} \mathrm{DFS}(z)$) is approximately 5 for $\boldsymbol{T}(z)$ and between 3 and 5 for $\boldsymbol{q}(z)$. The DL generally has high information content (near unity for each height level) where there is sufficient aerosol concentration to provide the backscattered signal (which is usually only within the atmospheric boundary layer). The question we want to address here is how does the information content in individual instruments translate to information content for the derived advection that requires using data from multiple instruments?

For this, we need to look at the averaging kernel in more detail. The averaging kernel A (Eq. 2) can also be expressed as

$$A = \frac{K^T \mathbf{S}_m^{-1} K}{K^T \mathbf{S}_m^{-1} K + \mathbf{S}_a^{-1}}$$
$$= \frac{K^T \mathbf{S}_m^{-1} K + \mathbf{S}_a^{-1} - \mathbf{S}_a^{-1}}{K^T \mathbf{S}_m^{-1} K + \mathbf{S}_a^{-1}}$$
$$= I - \frac{\mathbf{S}_a^{-1}}{K^T \mathbf{S}_m^{-1} K + \mathbf{S}_a^{-1}}$$
$$= I - \frac{\mathbf{S}_x}{\mathbf{S}_a}. \quad (7)$$

This was also shown by Cadeddu et al. (2017), where $\boldsymbol{I}$ is the identity matrix. The prior covariance matrix $\mathbf{S}_a$ illustrates

the climatological "volume" of state vector, and the posterior covariance matrix S_x is the "volume" of that space that results after the retrieval is performed and the information from the observations is included. If there is no information about the state vector from the observations, then S_x will be approximately S_a, and thus A is approximately 0. If there is a lot of information in the observations, then S_x will be markedly smaller than S_a and thus A will be approximately 1. This matches the high-level definition of information content provided by Westwater and Strand (1968) well.

From Eq. (7), we see that if we can obtain $S_{a,\underset{x}{\leftrightarrow}}$ (the advection prior covariance) and $S_{\underset{x}{\leftrightarrow}}$ (the advection posterior covariance) then we can estimate the information content of the derived advection that we are gaining through the measurements. We have already demonstrated that $S_{\underset{x}{\leftrightarrow}}$ can be obtained through error propagation using Eq. (6), since the AERI and DL retrievals provide posterior covariance matrices. To get $S_{a,\underset{x}{\leftrightarrow}}$, we again use Eq. (6), but instead replace the posterior covariance matrices S_x and S_w with the prior covariance matrices used in the TROPoe ($S_{a,x}$) and DLoe ($S_{a,w}$) retrievals, respectively. Note that the three facilities (E37, C1, and E39) use identically the same prior covariance matrix in TROPoe, and the same is true for DLoe; these have been derived using 20 years of summertime radiosonde data launched at the SGP central facility.

4 Example

To compute the information content (DFS) for the temperature and water vapor advection, the prior and posterior covariance matrices from the observations were propagated through the forward model using Eq. (6), and then the averaging kernel A was computed using Eq. (7). The diagonal of A provides a profile of DFS for both temperature and water vapor advection, which are shown in Fig. 5 (panels e and f, respectively). There is a striking similarity to the spatial patterns of the 1σ uncertainties (Fig. 3e and f) and the DFS (Fig. 5e and f) time–height cross-sections. Generally speaking, the DFS and 1σ uncertainties are anticorrelated, with higher DFS values being associated with lower 1σ uncertainties. In the region of cold air advection above 1.2 km from 00:00 to 07:00 UTC, the DFS is very small for temperature advection (Fig. 5e), suggesting that there is no information in that region and thus those advection values should not be trusted. However, the DFS figures also suggest that the information content on advection can often be near unity at heights approaching 2 km (Fig. 5e and f), even though the AERI information content is very limited with DFS < 0.05 at any height above 50 m for T and DFS < 0.3 for q (Fig. 5a1–a3 and b1–b3); this is because the advection is essentially an evaluation of spatial gradients, which the AERI is able to determine even with its limited information content in the vertical. The DL at the E37 site also is clearly the outlier of the three DLs from an information content perspective, which can be seen by comparing the c1 and d1 panels with the c2, c3, d2, and d3 panels in Fig. 5, due to the larger instrument noise level in the E37 DL.

There are several natural questions that could be asked to better understand the information content results. For example, are the site-to-site differences in the posterior covariances (as evidenced by the changes in the 1σ uncertainty profiles shown in Fig. 5 for a given variable like T or u) impactful on the advection DFS profiles? To test this, we performed two tests: use the posterior covariance matrix from the C1 retrieval as the posterior for the E37 and E39 retrievals for (a) T and q as well as (b) u and v. In both of these tests, there was relatively little change to the resulting time–height profile of DFS for temperature and moisture advection (DFS differences were less than 0.1; not shown). We found this surprising, especially since the E37 DL has much larger uncertainties in u and v above 1 km than the other two sites; however, when we used the E37 DL posterior for all three sites, the information content decreased by nearly 0.1 uniformly above 1.2 km (not shown). Another sensitivity test performed was to inflate the TROPoe posterior covariances by a factor of 2. The new DFS time–height cross-section for the temperature and advection data is shown in Fig. 6 (panels a and b, respectively). Comparing these DFS results with the baseline (Fig. 5e and f) shows a qualitatively similar evolution of the DFS profiles but also a decrease in the DFS for temperature and moisture advection by approximately 0.2 to 0.4 along the top contour (Fig. 6c and d). Interestingly, inflating the DLoe posterior covariance matrices by a factor of 2 had little effect (DFS differences less than 0.1) below 2 km, with some DFS differences approaching 0.3 around 2.5 km (not shown). Presumably, this is because advection is a spatial calculation, and the uncertainties at the vertices have relatively little impact on the derived advection. However, this result would likely depend on the size of the polygon used for the calculation; Wagner et al. (2022) demonstrated that the current spacing of the C1, E37, and E39 facilities is close to optimal in minimizing both the random error and sampling error in the calculation.

So far, we have focused on the diagonal elements of the covariance matrices, as the square root of the diagonal provides the 1σ profile of uncertainties. Figure 7 shows the 1σ profiles derived from the advection's prior covariance (i.e., $S_{a,\underset{x}{\leftrightarrow}}$) and the mean 1σ profile from the advection's posterior covariance (i.e., from $S_{\underset{x}{\leftrightarrow}}$) for the 13 June 2019 case. Clearly, the addition of the observations adds information, as the posterior uncertainty profile is smaller. However, it is important to realize that the information content profile is not 1 minus the ratio of these two profiles; instead, as illustrated in Eq. (8), the off-diagonal elements also play a role.

The off-diagonal elements from both $S_{a,\underset{x}{\leftrightarrow}}$ and $S_{\underset{x}{\leftrightarrow}}$ are shown in Fig. 8, where the covariance matrices were converted to correlation matrices for display purposes. Since
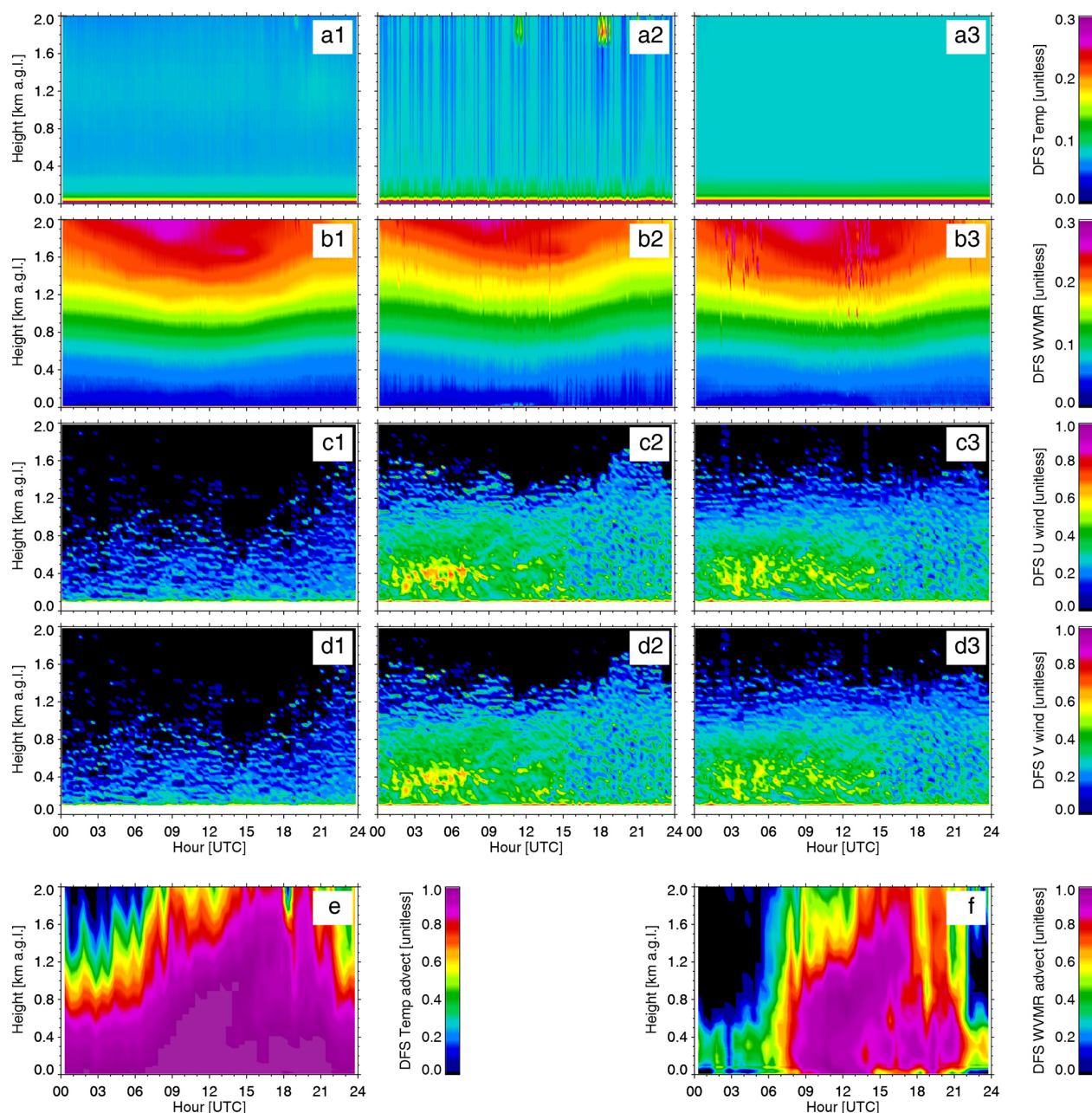
Figure 5. Same as Fig. 2 with the three columns denoting E37, C1, and E39 (for panels **a**, **b**, **c**, and **d**), but showing the DFS of the various products.

these covariance matrices are symmetric, only half of each is shown with the prior shown below the diagonal and the posterior above the diagonal. Note the large magnitude of the level-to-level correlation in the advection of temperature with itself (Fig. 8a) and water vapor with itself (Fig. 8b), as well as the cross-correlation of temperature and water vapor advection (Fig. 8c). However, after the retrieval, both the magnitude of the diagonal (Fig. 7) and the magnitude of the off-diagonal terms are markedly reduced. There are some negative correlations seen between two different levels (e.g., the

correlation of temperature advection with itself at 300 and 900 m is approximately −0.3 – see Fig. 8a); these features are similar to the structure of the correlation matrices in the TROPoe covariance matrices but much weaker (see Turner and Blumberg, 2019, Fig. 10 for an example).
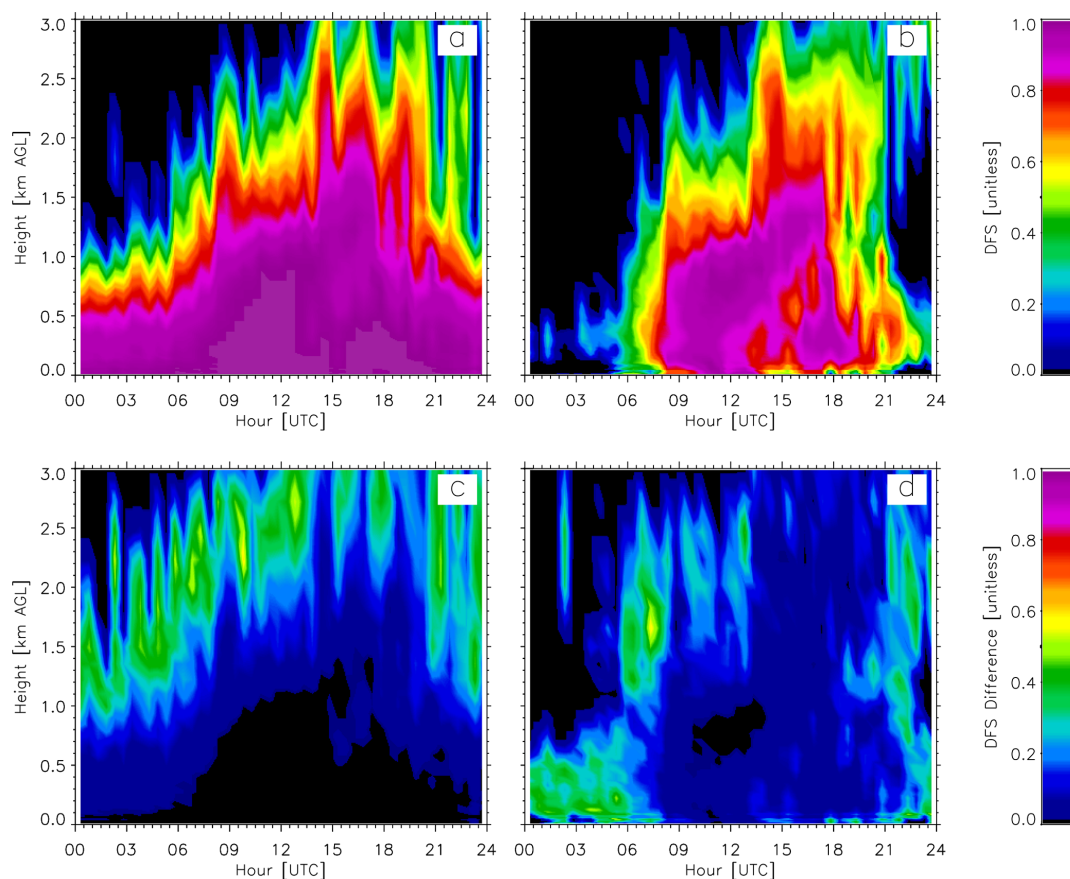
Figure 6. The time–height cross-section of DFS for temperature and moisture advection (panels **a** and **b**, respectively) for 13 June 2019, where the posterior covariance matrices from the TROPoe retrievals of T and q were inflated by a factor of 2. Panels **(c)** and **(d)** show the time–height cross-section of the differences of the DFS from Fig. 5 (panels **e** and **f**) with panels **(a)** and **(b)** here.
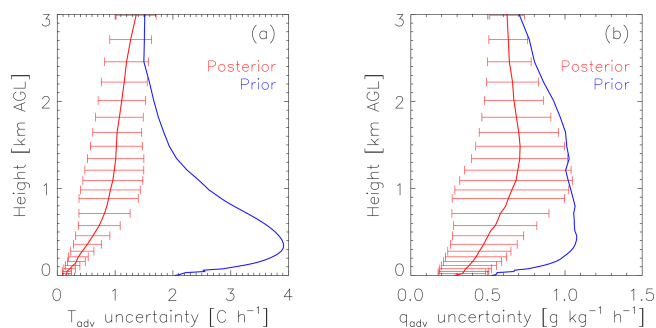

Figure 7. The 1σ profiles for temperature **(a)** and water vapor **(b)** advection, derived from the prior (blue) and posterior (red) covariance matrices for 13 June 2019. The posterior profile is the mean at each height over the day, with the error bars showing the standard deviation at each height over that day.

5 Statistical summary

The example on 13 June 2019 shown in Figs. 2, 3, and 5 provides an illustration of the derived advection profiles, its uncertainties, and its information content, which uses retrievals

from six independent instruments in the derivation. This particular example was chosen because there was marked temporal variability in the 24 h period. However, we are interested in more general statements about the uncertainty and information content in the derived advection. Recently, nearly 2 years of advection data were derived from the SGP observations (January 2018 to September 2019) using the Wagner et al. (2022) approach; however, that analysis did not include a description of the information content. Here, we analyze that same dataset to provide a sense of the average magnitude of the information content in both the temperature and water vapor advection and how that information content varies both over the diurnal cycle and as a function of height over two 4-month periods. Table 1 indicates the number of days of data that were available for each month in the 2-year record and the number of cases in the "cool" and "warm" seasons.

Time–height cross-sections of the mean information content for temperature advection and water vapor advection for the cool season are shown in the top row of Fig. 9. Note that the TROPoe algorithm has a minimum boundary layer height of 300 m; thus, the nocturnal boundary layer heights
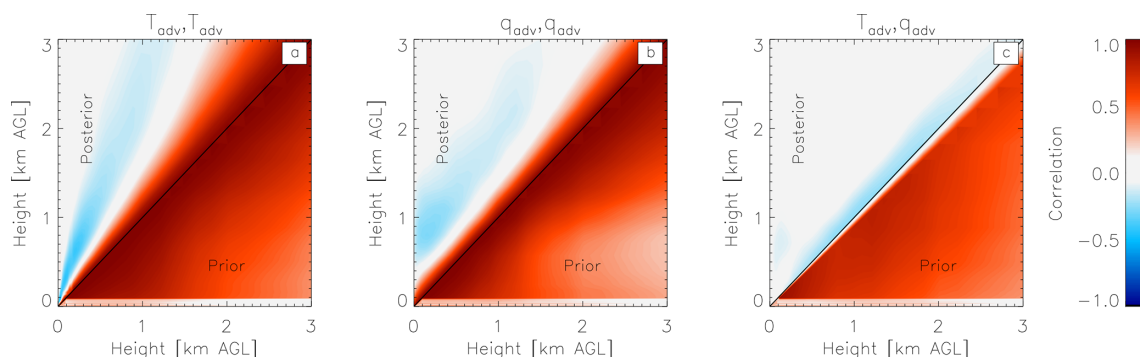
Figure 8. The level-to-level correlations between the advection of temperature/temperature **(a)**, water vapor/water vapor **(b)**, and temperature/water vapor **(c)** for 13 June 2019. The prior correlations are shown below the diagonal, and the posterior correlations are above the diagonal.
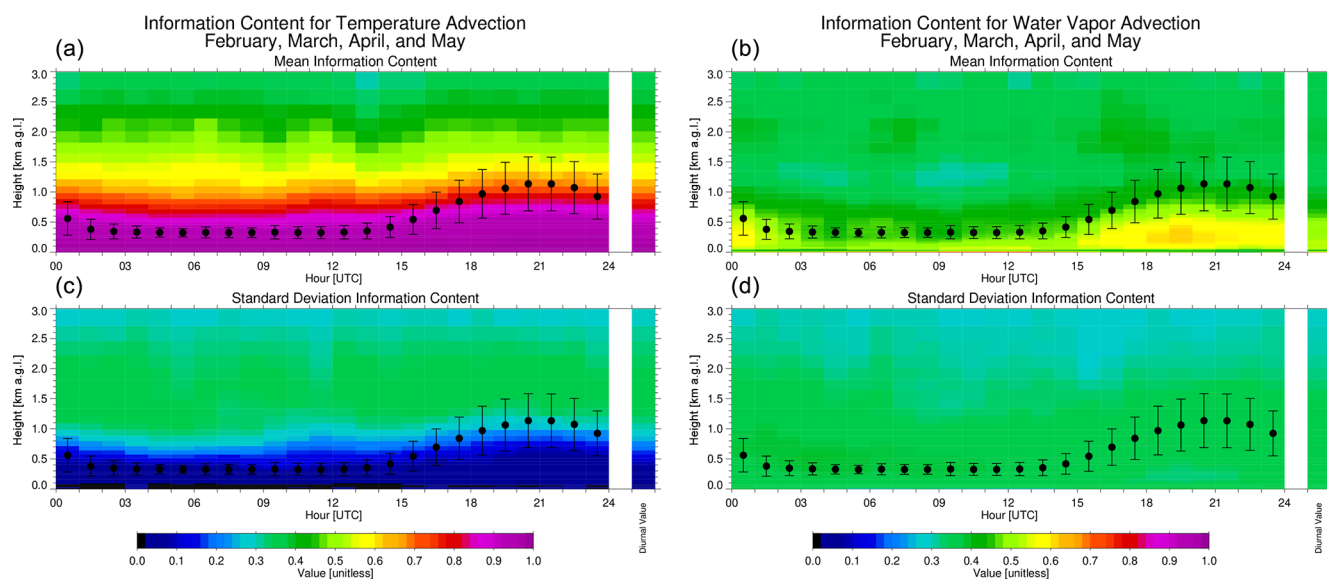

Figure 9. The mean diurnal information content **(a, b)** for temperature advection **(a, c)** and water vapor advection **(b, d)** for the "cool season" (Table 1). The variability of the information content, represented as the standard deviation, over this 4-month period is shown in the bottom row. The black dots illustrate the mean depth of the atmospheric boundary layer, which is derived from the TROPoe retrievals using a parcel method (e.g., similar to that used in Nielsen-Gammon et al., 2008), with the error bars indicating the variability of this height for each time over the days included in the analysis. The single bar on the right side of all four panels shows the value averaged over the entire diurnal cycle.

are largely this value. The standard deviations of the DFS data for each time and height are shown in the bottom row of Fig. 9 and provide a measure of the variability in the DFS across the nearly 150 d in this cool season analysis.

As can be seen, the information content for temperature advection (Fig. 9, left column) is above 0.9 (with a standard deviation less than 0.15) for most heights below 700 m, decreasing to a mean value of 0.8 by 1000 m. The standard deviation of the temperature DFS increases to about 0.4 at altitudes above 900 m, implying that there is marked variability in the DFS above this height from case to case. The mean DFS in the water vapor (Fig. 9, right column) has smaller values, with mean values between 0.5 and 0.7 in the lower

to middle part of the daytime boundary layer (i.e., between 14:00 and 24:00 UTC), with the mean DFS decreasing to 0.4 to 0.5 at and above the top of the boundary layer. The standard deviation of the water vapor advection DFS is approximately 0.3 to 0.4, regardless of time of day or height.

Figure 10 shows the same statistics for the warm season. There is little difference in the mean DFS or its standard deviation for the temperature advection over the diurnal cycle or vertically between the cool season (Fig. 9, left column) and warm season (Fig. 10, left column). However, there is a marked difference in the water vapor information content between the two seasons. The mean information content is much larger in the daytime boundary layer during the warm
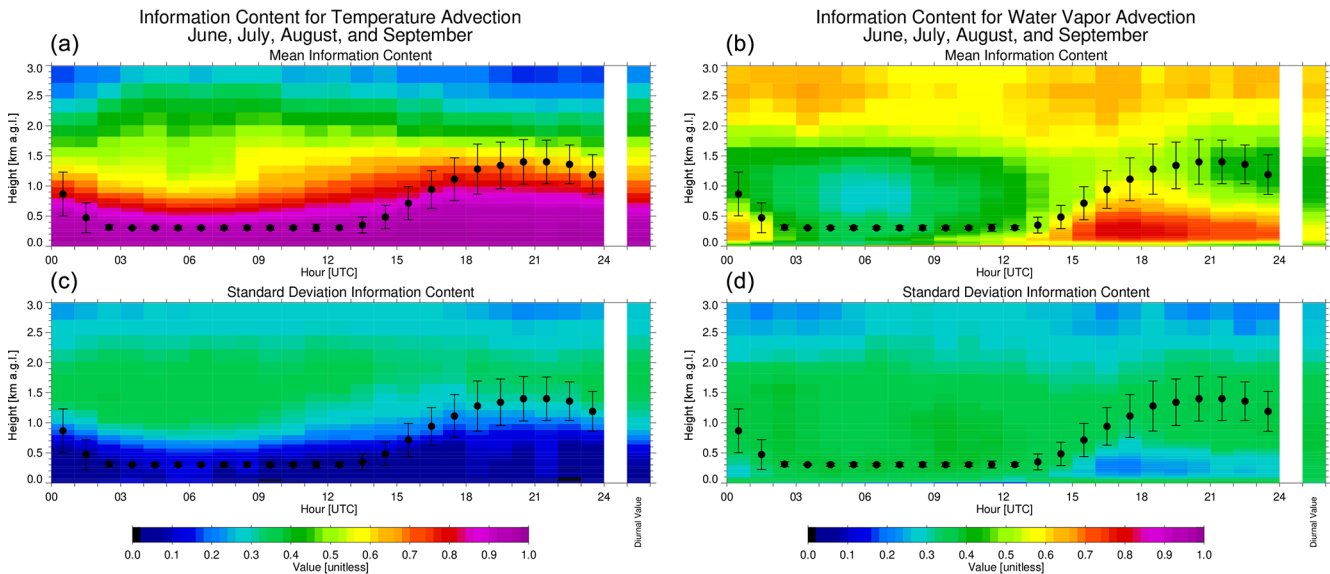
Figure 10. Same as Fig. 9, but for the "warm season" (Table 1).

Table 1. Number of days per month between 14 January 2018 and 17 September 2019 used in the analysis for the two seasons.

"Cool season"		"Warm season"	
Month	Number of days	Month	Number of days
February	27	June	27
March	54	July	31
April	28	August	25
May	43	September	15
Total	152	Total	98

season, with mean values of 0.7 to 0.8 in the warm season vs. 0.5 to 0.6 in the cool season. The variability in the DFS for water vapor advection is also about 50 % smaller in the daytime boundary layer in the warm season vs. the cool season. There is also an increase in the mean water vapor advection DFS above 2 km in the warm season, which is primarily contributed by the use of radiosonde data in the TROPoe retrievals with a smaller contribution from microwave radiometer brightness temperature observations, both of which have a larger impact due to the overall wetter conditions in the warm season.

6 Conclusions

To properly interpret remote sensing observations, which are often constrained using a priori information, it is important to understand the uncertainties and information content. Some geophysical variables are derived from remote sensing observations, and thus the uncertainties and information con-

tent need to be propagated through the derivation equation. This work demonstrates how to propagate information content from multiple remote sensors through a derivation equation for a new quantity. The key is to propagate the uncertainties from each individual retrieval through the derivation equation simultaneously, then propagate the uncertainties from the a priori constraints through the same equation in the same manner and look at the ratio of the covariance matrices derived from the two datasets.

We illustrated this approach using a network of remote sensors to derive the horizonal advection. To derive advection, we needed to have (at least) three non-colinear sites that measure profiles of the quantity of interest (in this case, temperature and humidity) with wind profiling at the same locations. In our case, we had six separate instruments, with three providing thermodynamic profiles (using the same a priori information) and three providing kinematic profiles (again, using the same a priori information for all three). However, because of differences in the noise characteristics of the different instruments, the uncertainties and information content derived from each individual instrument varied. The posterior covariance matrices (i.e., the individual retrieval uncertainties) were used to derive uncertainties and information content profiles for the derived temperature advection and water vapor advection profiles.

A statistical analysis of the information content profiles demonstrates that there is nearly perfect information content (i.e., DFS close to 1) for temperature advection below 700 m. This suggests that if there is a true change in the temperature advection below that level, then the observed temperature advection would capture the magnitude of that change at the right level. The associated information content for water vapor advection is different though; it is a strong function

of height, time of day, and season. Nonetheless, the daytime mean information content for water vapor advection in the boundary layer in the warm season is above 0.7, suggesting that the magnitude of any true perturbation in the water vapor advection would be largely captured by this instrument suite.

This work demonstrates how to derive the information content of an observation that is derived from multiple remote sensing datasets. The key aspect is to frame the individual derivations as retrievals so that both prior and posterior covariance matrices are available. Propagating information content, as was illustrated here, can inform the user of the derived data on where the signal-to-noise ratio is the best and potentially reduce the risk of misusing the data. However, it has been shown that a single microwave radiometer making azimuth scans can identify spatial gradients of water vapor (Schween et al., 2011). If this was paired with an instrument measuring horizontal wind profiles, water vapor advection could potentially be derived, but the propagation of uncertainties and information content could be performed the same way as shown here.

References

Baidar, S., Wagner, T. J., Turner, D. D., and Brewer, W. A.: Using optimal estimation to retrieve winds from velocity-azimuth display (VAD) scans by a Doppler lidar, Atmos. Meas. Tech., 16, 3715–3726, https://doi.org/10.5194/amt-16-3715-2023, 2023.

Bevington, P. R. and Robinson, D. K.: Data Reduction and Error Analysis for the Physical Sciences, third edition, McGraw-Hill, 320 pp., ISBN 0-07-911243-9, 2003.

Blumberg, W. G., Wagner, T. J., Turner, D. D., and Correia Jr., J.: Quantifying the accuracy and uncertainty of diurnal thermodynamic profiles and convection indices derived from the Atmospheric Emitted Radiance Interferometer, J. Appl. Meteor. Clim., 56, 2747–2766, https://doi.org/10.1175/JAMC-D-17-0036.1, 2017.

Browning, K. and Wexler, R.: The determination of kinematic properties of a wind field using Doppler radar, J. Appl. Meteor., 7, 105–113, 1968.

Cadeddu, M. P., Marchand, R., Orlandi, E., Turner, D. D., and Mech, M.: Microwave passive ground-based retrievals of cloud and rain liquid water path in drizzling clouds: Challenges and possibilities, IEEE T. Geosci. Remote Sens., 55, 6468–6481, https://doi.org/10.1109/TGRS.2017.2728699, 2017.

Gebauer, J. G. and Bell, T. M.: A flexible, multi-instrument optimal estimation retrieval for wind profiles, J. Atmos. Ocean. Technol., 41, 605–620, https://doi.org/10.1175/JTECH-D-23-0134.1, 2024.

Knuteson, R. O., Revercomb, H. E., Best, F. A., Ciganovich, N. C., Dedecker, R. G., Dirkx, T. P., Ellington, S. C., Feltz, W. F., Garcia, R. K., Howell, H. B., Smith, W. L., Short, J. F., and Tobin, D. C.: Atmospheric Emitted Radiance Interferometer. Part II: Instrument performance, J. Atmos. Ocean. Technol., 21, 1777–1789, 2004.

Maahn, M., Turner, D. D., Loehnert, U., Posselt, D. J., Ebell, K., Mace, G. G., and Comstock, J. M.: Optimal estimation retrievals and their uncertainties: What every atmospheric scientist should know, B. Am. Meteorol. Soc., 101, E1512–E1523, https://doi.org/10.1175/BAMS-D-19-0027.1, 2020.

Michael, P.: Estimating advective tendencies from field measurements, Mon. Weather Rev., 122, 2202–2209, 1994.

Newsom, R., Shi, Y., and Krishnamurthy, R.: Doppler Lidar (DLPPI), 2018-01-02 to 2019-11-01, Southern Great Plains

(SGP) Central Facility (C1), Waukomis Facility (E37), and Morrison Facility (E39), Atmospheric Radiation Measurement (ARM) user facility, ARM Data Center [data set], https://doi.org/10.5439/1025186, 2020.

Newsom, R. K., Brewer, W. A., Wilczak, J. M., Wolfe, D. E., Oncley, S. P., and Lundquist, J. K.: Validating precision estimates in horizontal wind measurements from a Doppler lidar, Atmos. Meas. Tech., 10, 1229–1240, https://doi.org/10.5194/amt-10-1229-2017, 2017.

Nielsen-Gammon, J. W., Powell, C. L., Mahoney, M. J., Angevine, W. M., Senff, C., White, A., Berkowitz, C., Doran, C., and Knupp, K.: Multisensor estimation of mixing heights over a coastal city, J. Appl. Meteor. Clim., 47, 27–43, https://doi.org/10.1175/2007JAMC1503.1, 2008.

Pearson, G., Davies, F., and Collier, C.: An analysis of the performance of the UFAM pulsed Doppler lidar for observing the boundary layer, J. Atmos. Ocean. Technol., 26, 240–250, https://doi.org/10.1175/2008JTECH1128.1, 2009.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, Series on Atmospheric, Oceanic, and Planetary Physics, Vol. 2, World Scientific, 238 pp., ISBN 981-02-2740-X, 2000

Schween, J., Crewell, S., and Löhnert, U.: Horizontal-humidity gradient from one single-scanning microwave radiometer, IEEE Geosci. Remote Sens. Lett., 8, 336–340, https://doi.org/10.1109/LGRS.2010.2072981, 2011.

Shannon, C. E.: A mathematical theory of communication, Bell. Sys. Tech. J., 27, 379–423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x, 1948.

Sisterson, D. L, Peppler, R. A., Cress, T. S., Lamb, P. J., and Turner, D. D.: The ARM Southern Great Plains (SGP) site. The Atmospheric Radiation Measurement Program: The First 20 Years, Meteor. Monograph. Am. Meteor. Soc. 57, 6.1–6.14, https://doi.org/10.1175/AMSMONOGRAPHS-D-16-0004.1, 2016.

Smith, W. L., Woolf, H. M., and Jacob, W. J.: A regression method for obtaining real-time temperature and geopotential height profiles from satellite spectrometer measurements and its application to NIMBUS 2 "SIRS" observations, Mon. Weather Rev., 98, 582–603, 1970.

Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, Society for Industrial and Applied Mathematics, 342 pp., ISBN 0-89871-572-5, 2005.

Turner, D. D.: Tropospheric Optimal Estimation Retrieval (TROPOE), 2018-01-02 to 2019-11-01, Southern Great Plains (SGP) Central Facility (C1), Waukomis Facility (E37), and Morrison Facility (E39), Atmospheric Radiation Measurement (ARM) user facility, ARM Data Center [data set], https://doi.org/10.5439/1996977, 2022.

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

Turner, D. D. and Ellingson, R. G. (Eds.): The Atmospheric Radiation Measurement (ARM) Program: The First 20 Years, Meteor. Monograph, 57, American Meteorological Society, https://doi.org/10.1175/AMSMONOGRAPHS-D-16-0001.1, 2016.

Turner, D. D. and Löhnert, U.: 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, 2014.

Turner, D. D. and Löhnert, U.: Ground-based temperature and humidity profiling: combining active and passive remote sensors, Atmos. Meas. Tech., 14, 3033–3048, https://doi.org/10.5194/amt-14-3033-2021, 2021.

Twomey, S.: Indirect measurements of atmospheric temperature profiles from satellites: II. Mathematical aspects of the inversion problem, Mon. Weather Rev., 94, 363–366, 1966.

Wagner, T. J., Turner, D. D., Heus, T., and Blumberg, W. G.: Observing profiles of derived kinematic field quantities using a network of profiling sites, J. Atmos. Ocean. Tech., 39, 335–351, https://doi.org/10.1175/JTECH-D-21-0061.1, 2022.

Westwater, E. R. and Strand, O. N.: Statistical information content of radiation measurements used in indirect sensing, J. Atmos. Sci., 25, 750–758, 1968.