Comparison of Optical Flow Derivation Techniques for Retrieving Tropospheric Winds from Satellite Image Sequences

JASON M. APKE,[a] YOO-JEONG NOH,[a] AND KRISTOPHER BEDKA[b]

[a] *Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado*
[b] *NASA Langley Research Center, Hampton, Virginia*

(Manuscript received 3 June 2022, in final form 8 September 2022)

ABSTRACT: This study introduces a validation technique for quantitative comparison of algorithms that retrieve winds from passive detection of cloud- and water vapor–drift motions, also known as atmospheric motion vectors (AMVs). The technique leverages airborne wind-profiling lidar data collected in tandem with 1-min refresh-rate geostationary satellite imagery. AMVs derived with different approaches are used with accompanying numerical weather prediction model data to estimate the full profiles of lidar-sampled winds, which enables ranking of feature tracking, quality control, and height-assignment accuracy and encourages mesoscale, multilayer, multiband wind retrieval solutions. The technique is used to compare the performance of two brightness motion, or "optical flow," retrieval algorithms used within AMVs, 1) patch matching (PM; used within operational AMVs) and 2) an advanced variational optical flow (VOF) method enabled for most atmospheric motions by new-generation imagers. The VOF AMVs produce more accurate wind retrievals than the PM method within the benchmark in all imager bands explored. It is further shown that image regions with low texture and multilayer-cloud scenes in visible and infrared bands are tracked significantly better with the VOF approach, implying VOF produces representative AMVs where PM typically breaks down. It is also demonstrated that VOF AMVs have reduced accuracy where the brightness texture does not advect with the mean wind (e.g., gravity waves), where the image temporal noise exceeds the natural variability, and when the height assignment is poor. Finally, it is found that VOF AMVs have improved performance when using fine-temporal refresh-rate imagery, such as 1- versus 10-min data.

KEYWORDS: Cloud tracking/cloud motion winds; Lidars/Lidar observations; Remote sensing; Satellite observations; Wind profilers

1. Introduction

It is widely understood that accurate tropospheric wind observations are critical to understanding cloud and weather processes, aerosol and pollutant transport, and Earth's global circulation (Zeng et al. 2016). Wind observations are also routinely used to improve the initial state inputs for numerical weather prediction (NWP) models (e.g., Le Marshall et al. 2008; Wu et al. 2014). A key source of inferring these observations since the 1960s are atmospheric motion vectors (AMVs; Menzel 2001), which are retrieved from coherent cloud- and water vapor (WV)-drift motions within sequences of satellite visible (VIS), infrared (IR), and WV imagery. When coupled with precise height-assignment methods, AMVs yield tropospheric wind derivations over large domains and fill in data gaps in regions with sparse in situ information, such as the open oceans or near the poles (Velden et al. 2005).

Cloud and WV drifts used in AMVs are quantified by retrieval and navigation of the apparent brightness motion in image sequences, or the optical flow (OF; Horn and Schunck 1981, hereafter HS81). The underlying OF retrieval techniques in today's operational AMV algorithms (e.g., Santek et al. 2019, and references within) retain many traits from initial automation efforts in the 1990s, employing assumptions that require strict pruning via quality control (QC) in target selection, tracking, and postprocessing, to derive accurate winds. The resulting AMVs are thus sparse, meaning that winds are derived for only a small fraction of pixels from an image sequence, with particularly reduced spatial density in regions containing low texture, multilayer motions, and strong deformations. These sparse AMVs commonly miss mesoscale motions important for regional modeling, emerging machine learning research, and decision-making algorithms (Bedka and Mecikalski 2005; Bedka et al. 2009; Apke et al. 2016; Stettner et al. 2019).

The new generation of geostationary satellite imagers, such as the Geostationary Operational Environmental Satellite (GOES)-R series Advanced Baseline Imager (ABI; Schmit et al. 2017) and the Himawari Advanced Himawari Imager (AHI; Bessho et al. 2016), have temporal, radiometric, and spatial resolutions coupled with state-of-the-art image navigation and registration that enable use of advanced OF retrieval techniques for most atmospheric motions (e.g., Apke et al. 2020; Apke and Mecikalski 2021). These techniques relax common assumptions made in target-based processing and are even designed to return "dense" OF (DOF), where motion is derived at every image pixel. DOF algorithms function by intuitively using regional texture and information combined with assumptions on motion behavior to derive OF where brightness features are ordinarily difficult to track. If performed accurately, DOF derivations for AMV retrieval could significantly increase the density of retrieved assimilation-ready winds.

Denotes content that is immediately available upon publication as open access.

Corresponding author: Jason Apke, jason.apke@colostate.edu

DOI: 10.1175/JTECH-D-22-0057.1

Validation of DOF-based AMVs is hence critical to automation efforts as it guides technique development and QC prior to products being sent to end users.

While extensive research exists in OF retrieval validation and benchmarking (e.g., Baker et al. 2011; Butler et al. 2012; Geiger et al. 2012), little work has been done to compare the inherent accuracy of AMVs derived from various OF retrieval methods. The lack of work is in part related to challenges in establishing reliable truth datasets that fairly compare one OF retrieval approach to another. Sparse OF retrieval-based AMV comparisons (Santek et al. 2019) have been based on validation statistics for only the pixels where each algorithm produces a solution, and separates overall performance and spatial density into different metrics, making retrieval ranking ambiguous. Moreover, it is not known a priori what altitude a motion from an image pixel or region represents for any given OF retrieval approach, so even pixels with corresponding winds from different AMV approaches may represent different points in space. This height correspondence problem is further complicated when considering that different imager channels used for OF retrieval contain motions from different heights in the atmosphere, or even different layers as implied by Velden and Bedka (2009). These uncertainties are compounded with the sparse nature of truth information (i.e., rawinsondes and wind profilers), which often do not directly sample the tracked features in question.

This study aims to address these challenges with a novel profile validation approach performed using an advanced NASA airborne lidar wind-profiling dataset. These Doppler Aerosol Wind (DAWN) lidar data simultaneously capture both cloud height and very precise vector winds and were collected within domains imaged at 1-min intervals by the *GOES-17* ABI. The approach is used to demonstrate how a specific DOF retrieval algorithm based on penalty integral minimization, known as a variational optical flow (VOF) retrieval technique, can produce accurate winds in regions where traditional AMV OF retrieval methods break down. A brief background on OF retrieval methods within AMV derivations, as well as methods in both AMV and OF retrieval validation, is provided within section 2. The AMV OF retrieval methods used and validation technique are described in section 3. Section 4 details the validation results, and VOF technique modification to address retrieval uncertainties. Section 5 overviews the primary conclusions and plans for future work in DOF-based AMVs.

2. Background

a. *OF techniques and operational AMVs*

The first automated AMV techniques emulate the human process of "picture-pair" correlations where operators would manually track features across three successive satellite images (Merrill et al. 1991; Nieman et al. 1997; Schmetz et al. 1993). This automation involves a five-step process: 1) identify "targets" of interest, or small, box-shaped subsets (e.g., ~100 km^2) of the image, centered over gradients, minima, or maxima in brightness; 2) height assign using IR imagery and properties such as cloud type and phase in combination with NWP thermodynamic profile guidance (e.g., CO_2 slicing, H_2O intercept, and optimal estimation methods; Nieman et al. 1993; Heidinger 2013); 3) extract winds from the NWP at the assigned height for a forecast displacement estimate; 4) correspond the target with a matching candidate scene within a permissible search region around the forecast displacement in the subsequent image, yielding the apparent motion; and 5) repeat step 4 for the image at a previous point in time and average the motions to produce an AMV. Steps 4 and 5 perform target correspondence with either cross-correlation or least squares methods that are OF retrieval techniques often referred to as "patch matching" (PM; Fortun et al. 2015, and references within). This fundamental process is essentially still used for operational AMVs produced by satellite data processing centers around the world (e.g., Santek et al. 2019, and references within).

The accuracy of any OF retrieval approach is reduced where assumptions made within are flawed, and PM implements several assumptions that are often incorrect in practice. First, PM assumes that the feature brightness only changes as a function of motion between scans, also known as "brightness constancy" (HS81). Satellite imagery sequences often violate this assumption as clouds and WV features condense, evaporate, and occlude between scans. Brightness constancy also breaks down with rapid illumination changes in VIS, or vertical motion and cooling in IR. Second, PM assumes that all the motions within the target scene flow with the same speed and direction, or a "neighborhood consistency," dating back to early OF retrieval work by Lucas and Kanade (1981). This assumption is violated with any type of deformation, such as rotation or divergence, or motion discontinuities (Shi and Tomasi 1994). These first two assumptions both break down in the case of motions common within transparent scenes containing optically thin upper-level cirrus or WV features. Finally, PM assumes there is sufficient texture available to track. In the case of near textureless scenes from any imagery type, a corresponding match would be largely dictated by image noise versus brightness motion related to cloud and WV drifts, presuming a singular match even exists in the first place.

PM's sensitivity to such assumptions is why automated AMV algorithms have long required QC pruning procedures (e.g., Nieman et al. 1997; Daniels et al. 2010). In preprocessing, the National Oceanic and Atmospheric Administration (NOAA) National Environmental Satellite Data Information Service (NESDIS) AMV algorithm (Bresky et al. 2012) quantifies AMV target quality with tests on cloud layer number, brightness gradient magnitude, contrast, and coherence. Within the tracking process, checks on acceleration between the forward and backward computed displacements penalize erroneous motions caused by occlusions, multilayer scenes, and image noise. In postprocessing, NESDIS AMVs check for height-assignment related biases and compare remaining motions to nearby AMVs and numerical models, penalizing any unrealistic deviations. The QC process outputs so-called quality indicator (Holmlund 1998; Holmlund et al. 2001) and expected error (Le Marshall et al. 2004) values such that

users can prune retrieved AMVs to their needs. For example, the pruning methods are very strict for extracting very accurate data assimilation (DA) ready winds and will often only yield synoptic-scale flows (Bedka and Mecikalski 2005). Despite their limitations, AMVs commonly rank within the top five most important assimilated observations within observing system experiments for global NWP models (e.g., Sato and Riishojgaard 2016; Auligne et al. 2016; Kim et al. 2017).

While operational AMVs are suitably accurate for DA, AMVs with reduced QC pruning can retrieve useful motions at finer spatial scales, or in regions that operational AMVs are simply not designed to provide (so-called mesoscale AMVs; Bedka and Mecikalski 2005). Rabin et al. (2004), Velden et al. (2005), and Bedka et al. (2009) demonstrate that mesoscale AMVs resolve mid- and upper-level WV deformations within preconvective environments, low-level confluence, vertical wind shear, hurricane eyewall rotation, and convective outflows. Apke et al. (2016, 2018) and Sandmæl et al. (2019) also show how 1-min mesoscale AMV-derived flow fields exhibit convective cloud-top divergence and rotational signatures unique to long lived and severe weather–producing thunderstorms. Oyama et al. (2018) even correlates mesoscale AMV-derived radial velocity atop tropical cyclones to tropical cyclone intensification and internal structure changes.

Many alternative OF approaches were developed over the last four decades to support technology such as automated surveillance, self-driving cars, and biomedical imaging, which relax PM assumptions (e.g., Fortun et al. 2015, and references within) and could be used for retrieving better mesoscale AMVs. VOF is one such technique designed to return DOF originating from the seminal work by HS81. The HS81 method works by minimizing a "penalty integral," also referred to as an "energy function," which is function of the images used and the OF field motions and increases with violations in quantified assumptions on brightness behavior and flow smoothness over time. The HS81 method is unique in that it is solves over the entire image at once, makes no assumption on flow pattern like PM, and intuitively uses regional information to derive motion when little context on pixelwise texture is available.

Development in VOF algorithms since HS81 centered around redesigning penalties and the optimization process to better handle >1-pixel displacements (Anandan 1989; Brox and Malik 2011), preserve motion discontinuities (Black and Anandan 1996), manage motions caused by illumination changes (Brox et al. 2004), relax unrealistic flow smoothness assumptions (Corpetti et al. 2006; Zimmer et al. 2011), and use nonlocal image information to enhance flow retrieval accuracy (Sun et al. 2014). Modern VOF algorithms utilize GPUs and are computationally efficient enough to keep pace with the temporal resolution and size of new-generation satellite images like those from ABI and AHI instruments (Sundaram et al. 2010).

Recent research is already demonstrating the capabilities of DOF-based AMV and motion derivations. Wu et al. (2016) shows with rawinsonde and reanalysis data that a DOF retrieval approach by Farnebäck (2001) could produce comparably accurate dense AMVs in AHI imagery relative to PM-based approaches. Stettner et al. (2019) and Apke et al. (2020) outline how Brox et al. (2004) DOF could be used to retrieve hurricane vortex-scale AMVs in low-texture central dense overcast regions and accurately track outflow boundary motion in VIS and IR imagery. Apke and Mecikalski (2021) use a modified Sun et al. (2014) approach with 1-min VIS imagery to demonstrate how DOF-based AMVs can produce convective cloud-top flow fields with reduced motion-discontinuity artifacts that hindered previous methods. These early results indicate a growing consensus that DOF derivation methods are valuable within myriad satellite remote sensing applications and imply that the community would benefit from benchmark datasets that inform development toward more accurate motion retrievals.

b. *Validation research for OF and AMV retrieval*

A common approach for AMV validation involves first deriving AMVs, and then matching with corresponding truth winds to establish AMV error characteristics. The truth datasets are typically rawinsondes (Schmetz et al. 1993; Bresky et al. 2012; Velden et al. 1997) or stationary wind profilers (e.g., Bedka et al. 2009). This approach is challenged by the scarcity of both the truth data and AMVs, so most studies use some spatiotemporal threshold (e.g., validate AMVs within ~150 km, ~100 hPa, and ~30 min of a rawinsonde observation). Large thresholds imply mesoscale fluctuations are not adequately represented in the truth dataset; hence, mesoscale wind variations correctly captured by the AMVs would be considered errors. An alternative is validation with NWP and reanalysis product winds, which can be resampled to all AMV locations (Wu et al. 2016). Of course, NWP-based validation rewards AMVs that derive motions at a similar spatial scale as the NWP model used, which is often invalid, especially so for mesoscale features.

Deficiencies aside, these validations reveal common uncertainties from operational AMVs. Schmetz and Holmlund (1992) and Nieman et al. (1993) first identified the height-assignment step as a primary source of AMV errors. AMVs would often be placed too high in the troposphere, leading to a well-documented slow-wind bias that later approaches aim to correct (Nieman et al. 1997; Velden et al. 1998; Bresky et al. 2012). Using a large multispectral AMV dataset collocated with frequent rawinsonde profile collections, Velden and Bedka (2009) estimate that height-assignment uncertainty contributes up to 70% of the observation error, even when accounting for errors caused by temporal and spatial wind variability. Of course, an important caveat is that these uncertainty contributions are quantified specifically on QC pruned AMVs, which implies, aside from height assignment, established pruning techniques are exceptionally good at identifying where the tracking from the underlying PM OF technique does not perform well.

The common validation approach is quite difficult to use for AMV retrieval ranking, comparison, and technique development, as different AMV methods produce winds at different locations with varying degrees of accuracy and density.

An alternative validation approach first establishes a ground-truth dataset, and then attempts to derive an AMV for those specific images or targets. This approach is more common in computer vision literature, where so-called benchmarks such as Middlebury (Baker et al. 2011), MPI-Sintel (Butler et al. 2012), and KITTI (Geiger et al. 2012) contain image sequences with known motions that challenge OF algorithms using stop-motion laboratory sequences, synthetic computer-generated movies, and even laser scanners collocated with stereo-vision cameras. New OF approaches can then quantitatively evaluate their performance relative to other algorithms and varying image characteristics such as texture, blur, and motion discontinuities.

More recent AMV validation studies follow the alternative validation approach using NWP simulations that contain information on winds and cloud-top height (CTH) for truth. For example, Bormann et al. (2014), Hernandez-Carrascal and Bormann (2014), and Lean et al. (2015) demonstrate that AMVs derived from simulated IR and WV imagery produce similar error statistics and characteristics to rawinsonde-based studies. Posselt et al. (2019) expands previous simulations, computing 3D AMVs from WV images at multiple pressure levels. Their analysis framework enables comparison of the performance and yield of AMVs with varying instrument and PM approach characteristics, such as target size and temporal frequency. They determine that moisture patterns and flow regimes impact WV AMV error characteristics.

Simulated imagery studies like Posselt et al. (2019) are limited to IR emissions, however, as VIS reflectance is difficult to accurately simulate, even though a vast majority of high-quality AMVs are produced from VIS imagery (e.g., Daniels et al. 2019). Also, the simulated IR often underrepresents natural brightness variability present in real data, which can be vital for accurate tracking. Further, these simulation studies focus on identifying the uncertainty and error characteristics of operational AMV techniques, again utilizing the QC pruning, and not yet taking advantage of new DOF techniques. Hence, the common height-assignment-related problems still dominate reported uncertainties, and pixels that may be trackable with better OF approaches are disregarded. This study aims to provide a real dataset and validation technique that allows for AMV retrieval method comparison following the alternative validation approach used by the computer-vision community. The method will be used here to demonstrate how new OF retrieval techniques can successfully track regions that operational AMVs routinely ignore.

3. Data and methods

a. Data

This study used winds collected from the DAWN (Kavaya et al. 2014) over a 2-week period in April 2019 during the NASA Aeolus Calibration and Validation Test Flight Campaign (Bedka et al. 2021) as a truth dataset. DAWN is a 2-μm coherent-detection Doppler wind lidar that flew aboard the NASA DC-8 aircraft, which measures the Doppler shift of laser shots below the aircraft at multiple azimuths with a 30.12° tilt angle along the flight path. Line of sight wind velocities are retrieved at a ~33-m vertical resolution where sufficient aerosol concentrations or clouds exist to produce backscatter at each azimuth. For most of the campaign, DAWN collected data at five azimuth angles, equally spaced at 22.5° intervals covering ±45° centered on the aircraft nose. When the multiple azimuth data are combined, horizontal wind vectors are retrieved at ~5-km intervals along the flight track. While DAWN itself is a spatially and temporally localized sample of winds, the DC-8 frequently sampled clouds within the *GOES-17* field of view, which can be used for retrieving AMVs, which is optimal for the validation strategy described in section 3c. The campaign included five research flights (see Fig. 1 in Bedka et al. 2021), four of which are used in this study (Table 1). We did not analyze one case study on 17–18 April 2019 because DAWN was operating in a test mode for most of the flight, where only a single azimuth was used; hence, vertical profiles of horizontal wind vectors could not be derived. The DAWN winds were also omitted when the aircraft roll was ≥2.5°. Based on comparisons with 61 dropsondes, DAWN demonstrated ideal high precision with a wind speed root-mean-squared error (bias) of 1.2 m s^{-1} (0.13 m s^{-1}) and wind direction root-mean-squared error (bias) of 7.6° (−1°).

TABLE 1. Aeolus calibration and validation field campaign case studies and mesosectors used for comparison of DAWN winds to AMV retrievals.

Date(s) used	Time range	*GOES-17* mesosector
23 Apr 2019	0045–0700 UTC	1
25–26 Apr 2019	1930–0520 UTC	2
27–28 Apr 2019	1800–0400 UTC	1
29–30 Apr 2019	2000–0430 UTC	1

The flights were supported by coverage from a ~1000 km × 1000 km regional *GOES-17* ABI "mesoscale domain sector," refreshed every 1 min. This provided updates from all 16 ABI channels, which can be used to compute AMVs. The 0.64-μm VIS imagery band (3.9-μm IR band), which has a nominal ~500-m (~2-km) resolution at nadir, is used for cloud-top OF derivation where the solar zenith angle is < (≥) 85°. The 3.9-μm imagery is used here instead of longwave IR bands as it provides a signal with less WV attenuation of low-level cloud textures (e.g., Daniels et al. 2010), which are prominent in the DAWN validation dataset. In addition, we collect the 6.9-μm "midlevel" WV band (nominally ~2-km resolution) for motions at higher altitudes and over clear air. Each channel is then normalized to floating-point values from 0 to 255 following the settings in Table 2. IR bands were impacted by the *GOES-17* loop heat pipe problem during this study (NESDIS 2018), and thus contain some noise through each research flight. The temporal noise was severe enough in the 6.9-μm band that OF could only be computed with sufficient accuracy using 10-min separation between images.

The final data collected are the NOAA Global Forecast System (GFS) analysis and +3-h forecast fields. These fields provide winds for initial derivations in OF approaches at any

TABLE 2. Settings used in the VOF algorithm tested.

Parameter	Value
Coarse-to-fine pyramid levels (K)	4
Inner iterations (L)	3
Conjugate gradient iterations	30
Pyramid scale factor (SF)	0.5
γ (0.64/3.9/6.9 μm)	1/0.5/0.25
α (0.64/3.9/6.9 μm)	5/3/4
$\rho_d(x)$ and $\rho_s(x)$	$\sqrt{x+\epsilon}$
ϵ	0.001
Normalization max brightness (0.64/3.9/6.9 μm)	628.987/2/44.998
Normalization min brightness (0.64/3.9/6.9 μm)	−20.290/0/−0.2472

given time and height. The GFS fields also provide the background environmental data within the cloud height–assignment algorithm. Linear (bilinear) interpolation in time (space) is used between 3-h outputs (discrete pixel locations) to the times (locations) in question.

b. *Optical flow techniques*

The PM approach for all channels and the VOF approach for the 6.9-μm 10-min steps require first-guess motions to produce accurate solutions. For pixel CTH retrieval, we use the NOAA Algorithm Working Group Cloud Height Algorithm (ACHA; Heidinger and Pavolonis 2009; Heidinger 2013), part of the NOAA Enterprise Cloud Algorithms. The algorithms are operated in the Clouds from the Advanced Very High Resolution Radiometer (AVHRR)-Extended (CLAVR-x) system, which is the NOAA development framework for operational cloud retrieval products (Pavolonis et al. 2005; Heidinger et al. 2010, 2012; Walther and Heidinger 2012). ACHA CTH is produced at the same resolution as the IR and WV bands and is upscaled for VIS imagery with nearest neighbor interpolation in this study. With both the PM and VOF approaches used on the 0.64- and 3.9-μm bands, we initially assume the height of the wind is equivalent to the CTH of the central pixel. This assumption is tested further in section 4b and compared to the actual CTHs retrieved from DAWN, inferred by the highest occurrence of a signal-to-noise ratio returned value > 10 dB. For the 6.9-μm OF, height is assigned following Daniels et al. (2010) by matching pixel brightness temperature to the nearest occurrence in the GFS, using the lowest match in the cases of thermal inversions and linear interpolation to achieve subgrid precision.

PM OF is replicated here following the approach used in the nested targets in the NESDIS AMV algorithm (Daniels et al. 2010). A 5 × 5 pixel target is centered over every image pixel and tracked forwards in time to the next image with Neumann boundary conditions for the image edges. Note, in the NESDIS AMV algorithm, these 5 × 5 pixel targets are typically clustered to find dominant motion in 15 × 15 pixel regions, which is a form of QC pruning of erroneous motions that is not done here. For the sake of comparison with the DOF algorithm, we track all targets without any pruning. The CTH assignment from ACHA is used to pinpoint the corresponding GFS winds data to generate a first-guess displacement. The algorithm then searches for a corresponding location in the subsequent frame within a 5 × 5, 3 × 3, and 4 × 4 pixel search region centered over the first-guess displacement in the 0.64-, 3.9-, and 6.9-μm imagery, respectively. These search regions constrain PM displacements to be within ~25 m s^{-1} in the x and y directions of the first guess for the 1-min 0.64-μm and 10-min 6.9-μm bands, and ~50 m s^{-1} for the 3.9-μm band at the locations sampled. This constraint is larger for the 1-min 3.9-μm band to allow for >1 pixel of uncertainty from the GFS in any given direction. The algorithm seeks a matching candidate with least squares, returning the first guess in the unlikely event where a singular match does not exist.

As the candidate locations are returned with integer precision, we avoid artificial binning of winds along pixel locations and retrieve subpixel precision with quadratic interpolation of the sum-of-squared error using the match-adjacent pixels. In the cases of minima existing outside the designated search region, the algorithm outputs the integer displacement along the region edge. This subpixel interpolation is important with 1-min image refresh rates, as cloud drifts between frames are often very small, especially for the 3.9-μm band in the case studies here. With interpolation, PM tracking can still be achieved on these small displacements without imposing an artificial 1-pixel displacement speed limit.

The VOF method used here is an approach available within the Optical flow Code for Tracking, AMV, and Nowcasting Experiments (OCTANE; Apke 2021; https://github.com/JasonApke/OCTANE) that shares properties with Zimmer et al. (2011). This approach resolves DOF $\mathbf{U} = [u, v]^{\mathrm{T}}$ at every image I pixel Ω located in space at $\mathbf{x} = [x, y]^{\mathrm{T}}$ and in time t and $t + \Delta t$ by minimizing a penalty function E, where

$$E(\mathbf{U}) = \sum_{i,j\in\Omega} [\rho_d(\mathrm{BC}_{i,j}) + \gamma\rho_g(\mathrm{GC}_{i,j}) + \alpha\rho_s(\mathrm{SC}_{i,j})]. \quad (1)$$

Equation (1) variationally penalizes OF guess deviations from assumptions of brightness constancy (BC), gradient constancy (GC), and the smoothness constraint (SC) with constants γ and α controlling the importance of each penalty. These constraints are encapsulated in Charbonnier robust functions to preserve discontinuities, where

$$\rho_d(r) = \rho_g(r) = \rho_s(r) = \sqrt{r^2 + \epsilon^2} \quad (2)$$

represents a differentiable approximation of the absolute value function with $\epsilon = 0.001$. The penalizations are defined as

$$\begin{aligned} \mathrm{BC}_{i,j} &= |C_1 [I(\mathbf{x}_{i,j} + \mathbf{U}_{i,j}, t + \Delta t) - I(\mathbf{x}_{i,j}, t)]|^2, \\ \mathrm{GC}_{i,j} &= |C_2 [I_x(\mathbf{x}_{i,j} + \mathbf{U}_{i,j}, t + \Delta t) + I_x(\mathbf{x}_{i,j}, t)] \\ &\quad + C_3 [I_y(\mathbf{x}_{i,j} + \mathbf{U}_{i,j}, t + \Delta t) + I_y(\mathbf{x}_{i,j}, t)]|^2, \\ \mathrm{SC}_{i,j} &= |\nabla u_{i,j}|^2 + |\nabla v_{i,j}|^2, \end{aligned} \quad (3)$$

where ∇ is the horizontal gradient operator, I_x and I_y are the x and y derivatives of the brightness, and C_1, C_2, and C_3 are defined as

$$C_1 = \frac{1}{|\nabla I|^2 + \epsilon}, C_2 = \frac{1}{|\nabla I_x|^2 + \epsilon}, C_3 = \frac{1}{|\nabla I_y|^2 + \epsilon}. \quad (4)$$

OCTANE follows Brox et al. (2004) to set up the minimization of (1) as a system of linear equations to be iteratively solved, using the discretization from Brox (2005) for the divergence of $\rho_s(\mathrm{SC}_{i,j})$.

The linear system of equations is solved with a coarse-to-fine optimization scheme using specific settings outlined in Table 2. OCTANE uses a preconditioned conjugate gradient scheme (Sundaram et al. 2010) for fast and efficient iterative processing at every level, rendering 2000 × 2000 pixel image files in < 10 s on an Nvidia RTX 1070. The resulting pixel displacements are then navigated to zonal and meridional motions for comparison to DAWN.

c. *Validation strategy*

The goal of any AMV retrieval is to provide accurate winds from any point that is passively sampled by a satellite imager within a column. To account for the unknowns of where a given AMV retrieval method actually samples drifts in the atmosphere, the validation strategy here is to use the entire column of DAWN winds for every scan and provide a best guess for what the wind should be at every ~33-m grid cell. The DAWN ~33-m grid cells in a scan contain wind information at the DC-8 location from altitudes ranging from the surface to ~13 km and contain missing values where DAWN did not sample for any given point in time. For the first-guess fields, we use the flow interpolated from the GFS datasets. We then fill in grid cells that are believed to be sampled by AMVs with their respective motion derivations. VOF and PM AMVs as well as CTH and WV height information are collected from the nearest satellite pixels in space–time to the DC-8 location for any given profile. The AMV and height information is used regardless of regional characteristics to include ordinarily pruned AMV targets within the validation. We presume the AMVs derive the motions within a ~200-m column (±3 DAWN pixels) centered on the initial height assignment for all cloud/WV pixels, so each pixel within the column is filled with the same wind derivation. The quantified uncertainties are compared here between PM and VOF-based AMVs using the same height assignment method to highlight differences specifically related to the tracking methodology. The uncertainties are quantified by the root-mean-square vector difference (RMSVD) of the guess and the truth at all the DAWN pixels N, or

$$\mathrm{RMSVD} = \sqrt{\frac{\sum_{i \in N} (u_{\mathrm{Guess}} - u_{\mathrm{DAWN}})_i^2 + (v_{\mathrm{Guess}} - v_{\mathrm{DAWN}})_i^2}{N}}. \quad (5)$$

When analyzing DAWN pixels individually, we also use the analogous measure of vector difference magnitude (VDM), defined as

$$\mathrm{VDM} = \sqrt{(u_{\mathrm{Guess}} - u_{\mathrm{DAWN}})^2 + (v_{\mathrm{Guess}} - v_{\mathrm{DAWN}})^2}. \quad (6)$$

There are many advantages to validating with this strategy. First, it is possible to rank AMV approaches with both differing tracking and height-assignment methods based on the singular RMSVD values. Second, this validation strategy allows for multiple-altitude derivations of motion at any given location, encouraging multilayer winds derivations or even 3D winds algorithms, which may soon be possible with advances in instrumentation, computer vision, and machine learning. Third, this approach does not set any restriction on which channel(s) is (are) used to derive the flow, encouraging exploration of the best imagery to use for AMVs.

This strategy is not without limitations. For example, not all winds are sampled in each column from DAWN. If a region does not have sufficient aerosols for backscatter, or if the DC-8 flies underneath a sampled target, DAWN may contain missing values where the truth winds must be inferred from some other data source. AMVs with missing truth are disregarded here, though NWP or temporal smoothing of DAWN measurements can serve as a simple solution for these occurrences when height-assignment benchmarking is in question. Another limitation is that it is not always clear that the DAWN lidar and GOES ABI are sampling the winds from the same region. It is assumed here that the winds measured from each should be similar, though measurement differences may be present when the OF is determined by textures from different surrounding areas.

4. Results and discussion

a. *AMV approach comparison*

The DC-8 flight tracks coupled with DAWN observational capabilities provided a large sample size of winds below 10 km (Fig. 1a). For the AMV derivations, the 0.64- and 3.9-μm imagery are dominated by boundary layer marine-stratocumulus clouds (Figs. 1b,c), while the mid- to upper levels are sampled mostly by the 6.9-μm imagery (Fig. 1d). Even with the column filling approach used, the total AMVs with collocated DAWN observations are a small fraction of the truth samples available from DAWN at all heights.

The statistics from the DAWN observations with collocated AMVs show clear overall reduction in RMSVD when using VOF-based approaches over PM for every case and every channel observed (Fig. 2). The largest improvement was seen on the low-texture 3.9-μm band of ~9 m s^{-1}, while the smallest change, still a respectable ~1.7 m s^{-1} improvement, was seen on the high-texture, fine-spatial-resolution 0.64-μm imagery. The PM AMV RMSVDs are also much higher than reported ranges for NESDIS AMVs (~4–6 m s^{-1}; Daniels et al. 2019), indicating our samples contain a wealth of targets which are ordinarily pruned by QC as intended. Overall, the best performance was found with the 0.64-μm VOF winds. Despite improvement over PM, VOF overall performs poorly on the infrared bands here. There are obvious tracking issues for both OF approaches shown in the case-by-case breakdown, especially within the 3.9-μm imagery on 27 April 2019 and the 6.9-μm imagery on 23 April 2019. These issues are explored further in section 4b.

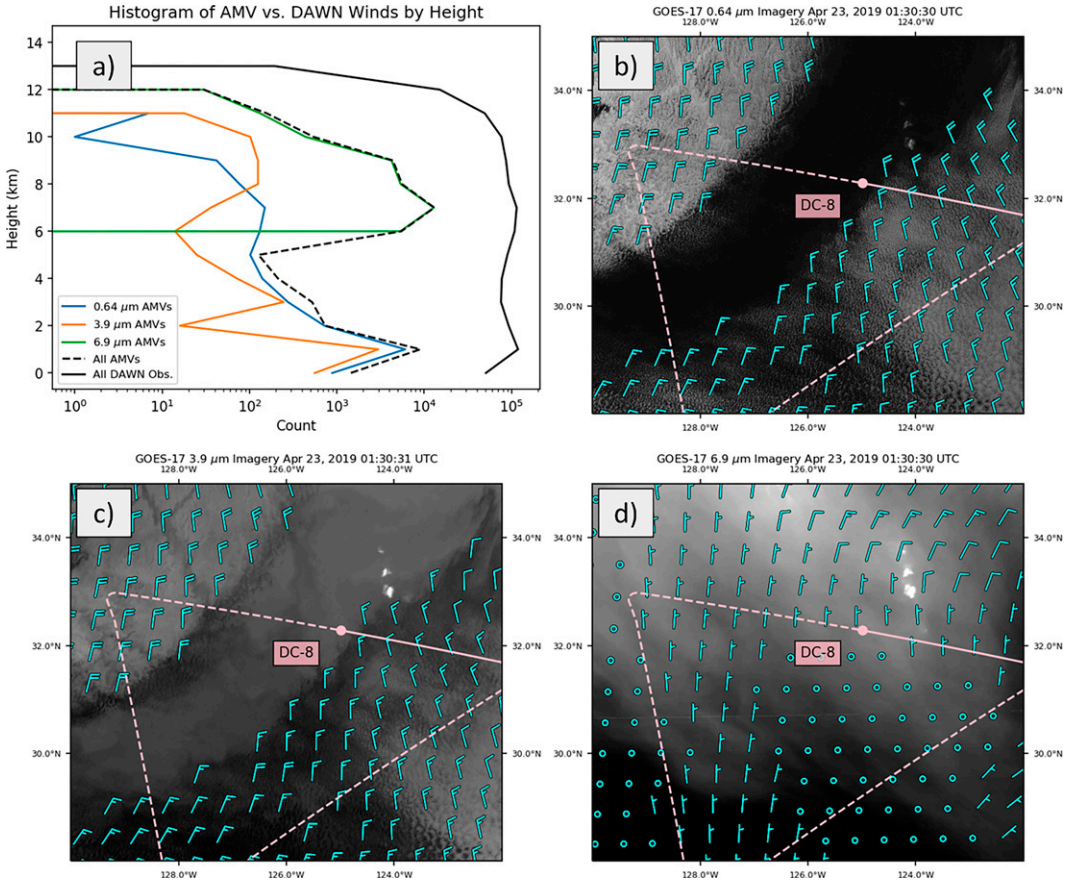

FIG. 1. (a) Histogram of AMV and DAWN observations by height along with the 0130 UTC 23 Apr 2019 (b) 0.64-, (c) 3.9-, and (d) 6.9-μm imagery and subset of cloud-top and water vapor VOF-derived AMVs (cyan wind barbs), shown with the previous and future flight paths of the NASA DC-8 in pink and pink dashed lines, respectively.

Within the DAWN pixels containing AMV derivations, it is possible to break down performance as a function of target characteristics. For example, NESDIS AMVs routinely remove low-texture (LT) targets to improve PM performance.

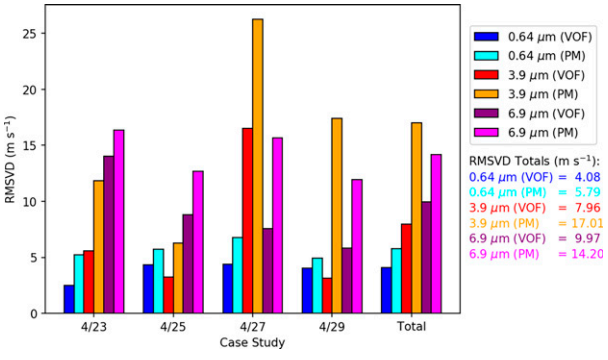

FIG. 2. Bar chart of RMSVD values of the winds derived by VOF and PM when compared to DAWN observations, broken down by case study, channel, and OF algorithm used. The total RMSVD values for all pixels with corresponding AMV derivations are shown on the right.

High-texture (HT) targets are separated from LT targets here using the minimum eigenvalue of their respective structure matrix (Shi and Tomasi 1994), defined as

$$\mathbf{S} = \begin{pmatrix} \sum_{i\in n} I_{xi}^2 & \sum_{i\in n} I_x I_{yi} \\ \sum_{i\in n} I_x I_{yi} & \sum_{i\in n} I_{yi}^2 \end{pmatrix}, \tag{7}$$

where n represents every pixel in the 5×5 pixel target. When the minimum eigenvalue of $\mathbf{S}$ is large, the target contains either corners or salt and pepper texture that yields robust multidirectional information ideal for PM methods. Lower minimum eigenvalues contain either low texture or unidirectional gradients that are difficult to track with local OF techniques. LT targets are defined as those within the lowest 25th percentile of the minimum eigenvalues in the dataset collected for each channel.

For the 0.64- and 3.9-μm bands, the VDMs are significantly larger for LT targets than HT in PM (Figs. 3a,b). Within these same LT targets, VOF tracking has significantly lower VDMs than both HT and LT targets in PM. Further, there is a slight reduction in VDMs when tracking LT targets versus HT targets using VOF, implying performance loss is more a function

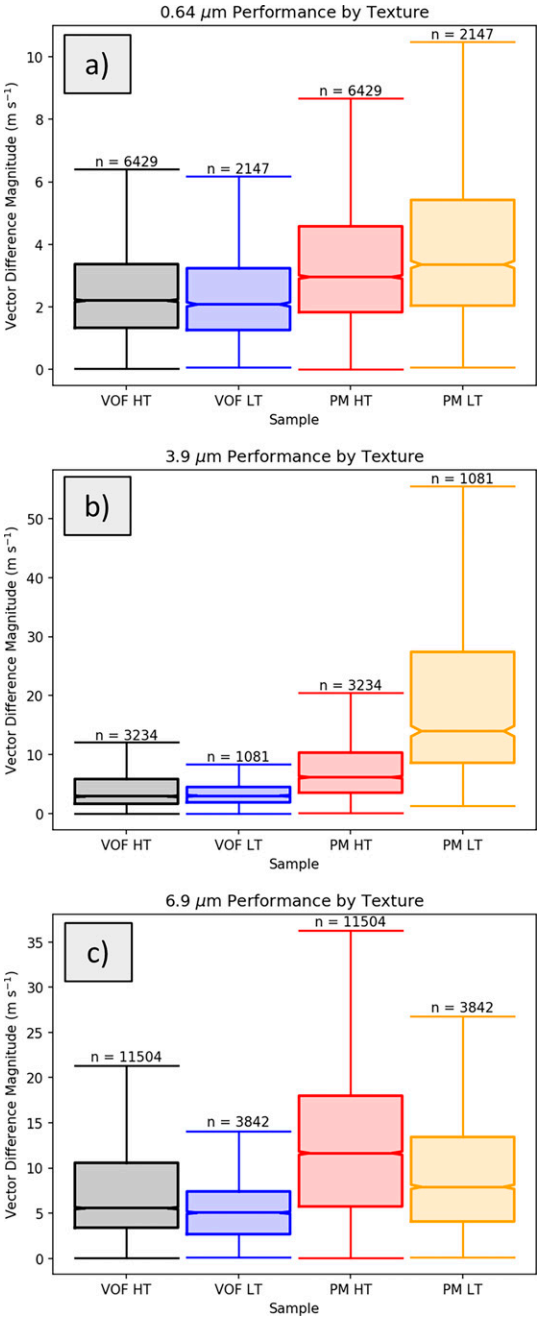

FIG. 3. Boxplots with sample sizes showing VOF and PM performance by high (HT) and low (LT) texture for the (a) 0.64-, (b) 3.9-, and (c) 6.9-μm bands with an interquartile range defined by the 25th and 75th percentiles, and whiskers defined by the points within 1.5 times the interquartile range. Notches to help imply significant differences are also shown around the medians defined by the interquartile range times 1.57 divided by the square root of the sample size.

of violations in OF constraints rather than lack of texture within this sample. The 6.9-μm motion (Fig. 3c) has a different relationship where HT targets have greater uncertainties than LT targets for both VOF and PM techniques, again, implying an issue within brightness motion assumptions made in each retrieval method (explored further in section 4b).

Another common error source for AMV derivation is the presence of multilayer (ML) cloud scenes, which are difficult to both properly track and height assign. We define ML cloud scenes as those which have ≥40% of target pixels outside 250 m of the derived height assignment, and otherwise classify the scene as single layer (SL). ML scenes contain larger VDMs for both VOF and PM in both 0.64 and 3.9 μm (Fig. 4), though again, VOF has significantly lower errors for the ML scenes than PM. Investigating a scene from 25 April shows how PM tends to yield noisy and inconsistent solutions in ML regions versus SL, compared with the smoother VOF (Fig. 5). While smoother, ML scenes still have significantly larger errors in VOF than SL scenes, likely related to challenges in height assignment and tracking with transparent motions, suggesting the ML issue is not fully solved with the VOF approach here. However, Figs. 3 and 4 both imply that, when using a VOF approach instead of PM, QC restrictions on lower textures and multilayer cloud scenes can be relaxed to significantly improve accurate, DA-ready AMV yields.

b. *VOF technique performance*

Now considering the entire DAWN observation sample with AMV and GFS first-guess winds, the final ranking scores of the benchmark demonstrate that VOF is a large improvement over PM tracking (Table 3). The overall RMSVD values are ~0.3 m s^{-1} lower when using VOF methods. The VOF scores are worse, however, than the GFS first-guess fields alone from the analysis and 3-h forecasts, which perform surprisingly well with a ~0.5 m s^{-1} lower RMSVD than PM. Taking cues from section 4a, addressing the low performance of the IR bands would significantly improve the overall scores of the VOF techniques.

Starting with the 6.9-μm wind retrievals, an investigation into the poor performance on 23 April reveals that the motions are sensitive to the gravity waves in the scene generated by orographic features the DC-8 flies over (e.g., see Fig. 8 in Bedka et al. 2021). These stationary waves produce a strong signal in the VOF BC and GC terms, which are heavily smoothed to nearby, lower texture pixels, yet the signal does not move with the ambient winds. Analyzing the energy with various DOF derivations show the 6.9-μm imagery may contain local minima with the wind motion when sufficient texture exists within the WV (e.g., Figs. 6a,b).

A desirable property for a VOF algorithm designed for AMV retrieval would be to converge on these minima and ignore the stronger propagating signal where possible. Following works such as Héas et al. (2007), we create a modified energy equation E_{wv} with a "hinting" term that adds weight to $\mathbf{U}$ solutions in the VOF retrieval near the GFS first-guess motion $\hat{\mathbf{U}}$, where

$$E_{\mathrm{wv}}(\mathbf{U}) = \sum_{i,j\in\Omega} [\rho_d(\mathrm{BC}_{i,j}) + \gamma\rho_g(\mathrm{GC}_{i,j}) + \beta|\mathbf{U}_{i,j} - \hat{\mathbf{U}}_{i,j}|^2 + \alpha\rho_s(\mathrm{SC}_{i,j})]. \quad (8)$$

The hinting term is weighted by constant β such that the first-guess motion is strong enough to filter outliers while not

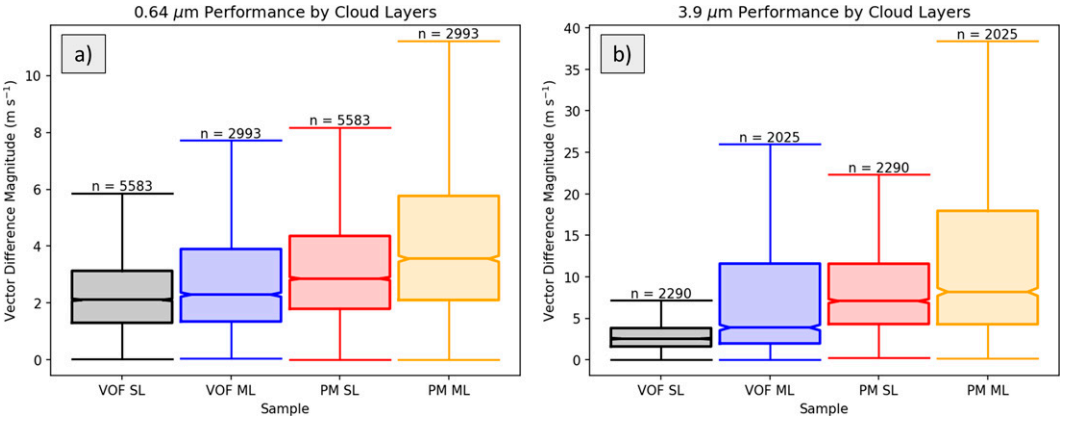

FIG. 4. As in Fig. 3, now showing VOF and PM performance samples with targets containing single-layer (SL) and multilayer (ML) cloud scenes for the (a) 0.64- and (b) 3.9-μm bands.

overwhelming the actual winds signal present in the imagery. Using $\beta_c = 0.5$, and

$$\beta = \beta_c \times 0.5^k. \tag{9}$$

With k set to the outer iteration index ranging from 0 to 4 here for coarsest to finest levels, respectively, we gradually reduce the impact of the hinting term as (8) is optimized. Setting $\hat{\mathbf{U}}$ to the GFS first-guess field (Fig. 6c), the resulting AMVs are much closer to the true winds measured by DAWN (Fig. 6d). GFS hinting terms come with the trade-off that the solutions will become biased toward inaccuracies from the GFS first guess. For example, the VOF AMVs here are less accurate if the model diagnostic wind field is wrong, if the model grid resolution cannot resolve the scales of the flow retrieved, or if the spatial location used to extract the flow from the model is incorrect.

Another common issue in quality is accurate height assignment. There are several instances in this benchmark where the ACHA algorithm identified clouds, which the VOF algorithm retrieved the motions below. In these instances, VOF is simply tracking the dominant texture in the scene, while ACHA was retrieving properties from the pixelwise IR information (e.g., Fig. 7). We take a few steps to filter these points, swapping targets containing slow motions < 5 m s^{-1} and excessive differences from the GFS first guess > 6 m s^{-1} with their respective GFS first-guess motion. Finally, as noted in section 4a, VOF errors are significantly larger for multilayer scenes, so targets with the ML designation are also swapped for the GFS first guess. There was no need for similar filtering of targets with the LT designation, as the VOF errors indicated in Fig. 3 were not significantly higher than HT targets. The resulting scores for the new benchmark, so-called VOF-QC, are much more competitive (also shown in Table 3). Note that similar QC techniques could be used to improve the benchmark scores of any AMV algorithm, including the PM technique used here.

As quality could also be impacted by the height-assignment method, we tested various methods in comparison to the actual heights from DAWN (Table 4). The center pixel height-assignment method described in section 3a, which generally performs well versus the actual heights from DAWN (e.g., Fig. 8), is compared to techniques where we take the mean CTH from ACHA of cloud pixels within a 9×9 pixel region centered over the DAWN sampled pixel (cloud pixel mean), and the mean of the highest 25% of pixels within the 9×9 pixel region (highest 25% mean).

The cloud pixel mean method produced the lowest overall RMSVDs using our benchmarking method here (and had the lowest root-mean-squared error compared to DAWN CTH) and is used for the wind height assignments in the 0.64- and 3.9-μm derived AMVs in the "VOF-QC" scores in Table 3. Notably, there is still room for RMSVD reduction with more improved height-assignment methods as demonstrated by the benchmark RMSVDs using the CTH from DAWN.

When each channel from VOF-QC is analyzed individually (last three columns in Table 3), the high-performing 0.64-μm imagery provides the lowest overall RMSVD values of any approach tested. The 3.9- and 6.9-μm imagery slightly increased the errors over the GFS first guess alone, implying room in these channels for further improvement than what was done here. Additional common QC methods such as acceleration checks or coherency analyses may further improve the RMSVD values over what is reported. It is also worth noting that the GFS assimilates the NESDIS AMVs, which likely improves the accuracy of the GFS first-guess fields used here.

c. *Imaging strategy uncertainties*

The benchmarking technique can also be used for identifying uncertainties related to imaging strategies such as bands used, spatial resolutions, and image bit depth. As an example, one such strategy that has a complex relationship with AMV performance is the temporal resolution. Generally, finer temporal resolution leads to improved tracking performance for most AMV retrieval techniques (e.g., Posselt et al. 2019). From the VOF standpoint, finer temporal resolution decreases large displacements and influence from unpredictable brightness changes such as cloud evaporation and condensation. Both

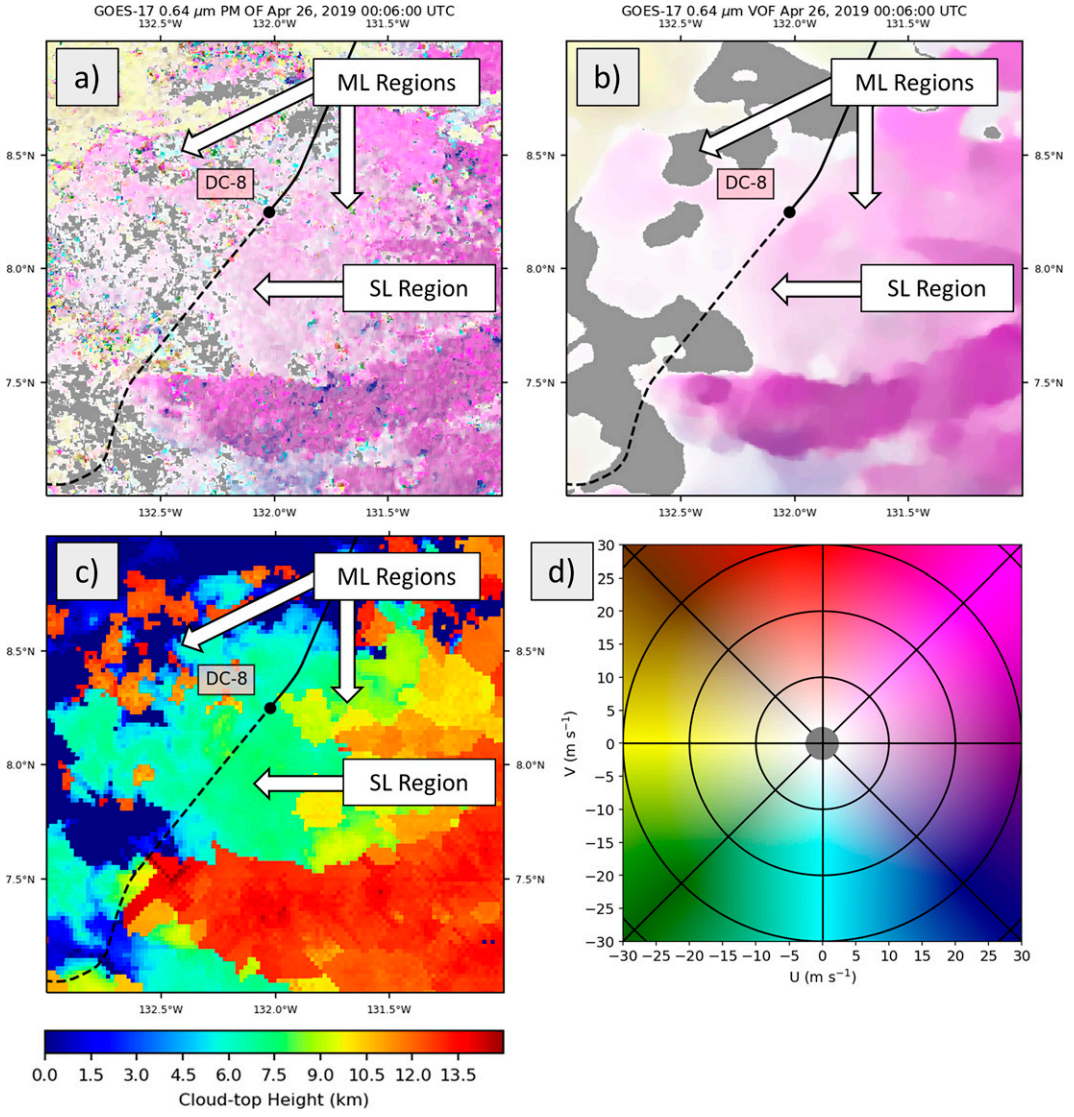

FIG. 5. Example color-shaded (a) PM- and (b) VOF-derived motions, shaded following the scale in (d), shown with the (c) ACHA cloud-top heights and the previous and future flight paths of the DC-8 in black solid and black dashed lines, respectively from 0006 UTC 26 Apr 2019.

Velden et al. (2005) and Daniels et al. (2010), however, point out that increasing temporal resolution also amplifies navigation and registration errors in the conversion of OF to winds. Further, brightness naturally fluctuates less with shorter time steps, which means that tracking errors associated with systematic image noise become amplified.

Within the benchmark here, degrading the resolution in 0.64 and 3.9 μm to 5 min does not significantly increase the

TABLE 3. Benchmark performance RMSVD values (m s^{-1}) for the PM, VOF, and VOF-QC models broken down by channels used and compared with the GFS first guess only. The "% filled" row indicates how many of the pixels out of a sample of 1 036 210 DAWN winds were estimated with AMVs vs estimated with a GFS first guess to derive the RMSVDs.

Samples	PM	VOF	GFS	VOF-QC	QC 0.64 μm	QC 3.9 μm	QC 6.9 μm
23 Apr	4.2931	4.0642	3.6506	3.6530	**3.6476**	3.6522	3.6545
25 Apr	4.4445	4.2991	4.1976	4.2057	**4.1968**	4.1979	4.2061
27 Apr	4.3286	3.9320	3.7649	3.7720	**3.7635**	3.7651	3.7732
29 Apr	3.7723	3.1888	**3.0940**	3.1310	3.0952	3.0949	3.1290
Total	4.2180	3.8984	3.6903	3.7024	**3.6891**	3.6911	3.7028
% filled	2.7250	2.7250	—	1.7749	0.4523	0.2473	1.0753

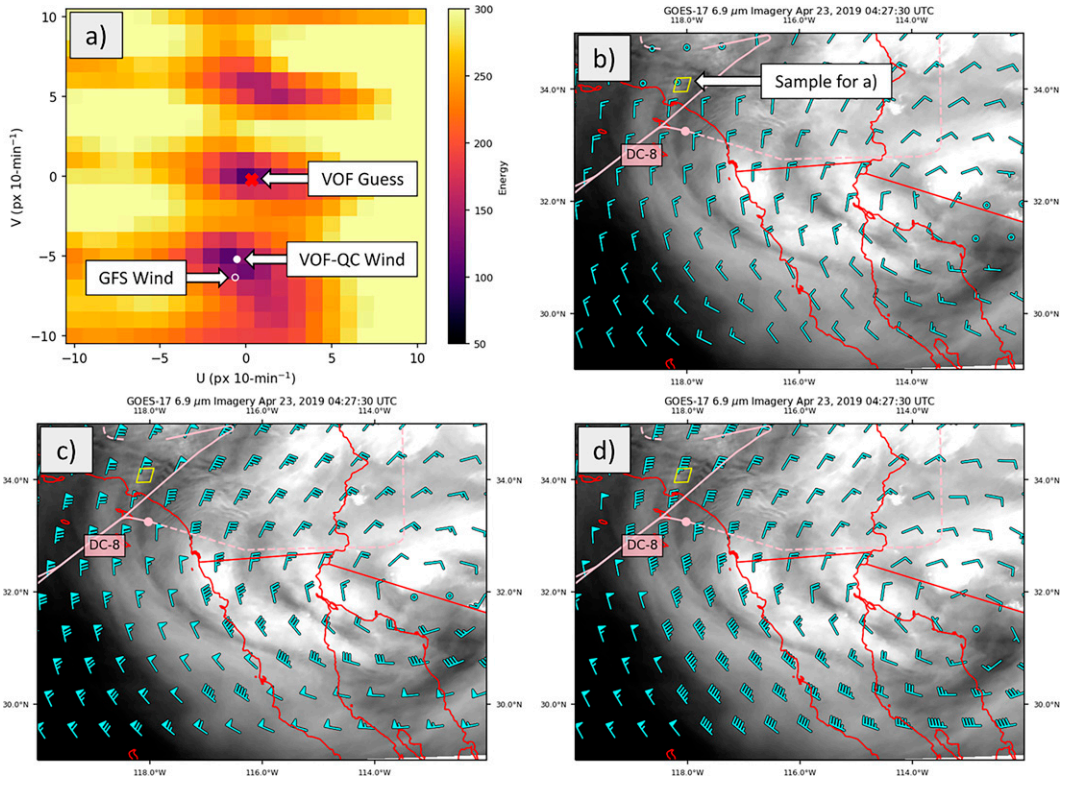

FIG. 6. (a) Color-shaded plot of the texture energy from BC and GC within a region with multiple OF motions in water vapor imagery at 0427 UTC, defined by the yellow box in (b)–(d), shown with the corresponding water vapor imagery and a subset of the (b) initial AMV winds derived from the VOF approach, (c) first-guess winds from the GFS model, and (d) VOF output using the GFS first guess as a hinting term (cyan barbs) with the previous and future flight paths of the NASA DC-8 overlayed in pink solid and pink dashed lines, respectively.

median VDMs, though 10-min AMVs perform significantly worse (Fig. 9a). The 5-min VIS and 10-min VIS and IR mean VDMs are higher than 1 min, indicating larger error outliers than 1 min. Of course, as shown in Fig. 1a, these samples are dominated by slow-moving boundary layer cumulus (Fig. 9b), and not the faster moving mid- and upper-level clouds typically observed in satellite imagery, which contain more nonlinearities related to large displacements that break down the

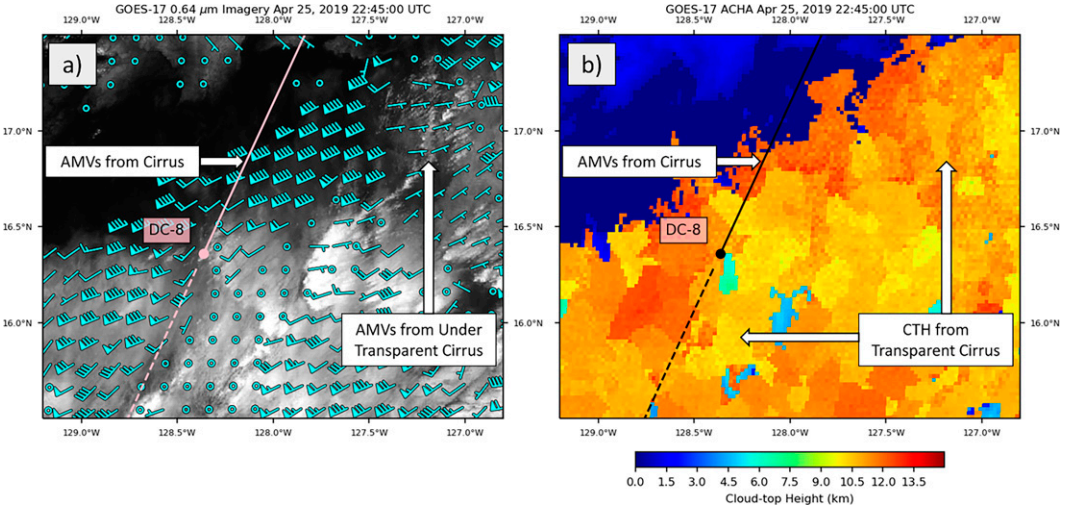

FIG. 7. Example of (a) 0.64-μm imagery with a subsample of VOF-derived AMVs shown with (b) ACHA-derived cloud-top heights at 2245 UTC 25 Apr 2019 and the previous and future flight paths of the NASA DC-8 in pink/black solid and pink/black dashed lines, respectively.

TABLE 4. Benchmark performance RMSVD values for varying height-assignment algorithms, including the root-mean-squared error of the approach when compared to the CTHs retrieved from DAWN (in km). Numbers in boldface highlight the best-performing height-assignment method, excluding the DAWN column.

Samples	DAWN	Center pixel	Cloud pixel mean	Highest 25% mean
23 Apr	3.6512	3.6533	**3.6530**	3.6549
25 Apr	4.2060	**4.2057**	4.2057	4.2069
27 Apr	3.7712	3.7728	**3.7720**	3.7728
29 Apr	3.1300	**3.1301**	3.1310	3.1315
Total	3.7015	3.7025	**3.7024**	3.7035
Height RMSE	0.0000	2.4450	**2.3486**	2.4498

assumptions when minimizing (1). It is also worth noting that the 5-min 3.9-μm imagery performed significantly better than 1-min tracking here. As there is not a similar performance improvement in the 0.64-μm imagery, it is unlikely this improvement was related to navigation and registration errors. This finding is consistent with studies that have demonstrated very low navigation and registration errors from *GOES-17* ABI products using known land features (e.g., Tan et al. 2020). The 3.9-μm imagery does, however, have larger temporal noise relative to both the natural background variability and the 0.64-μm band, which was likely the issue here.

To demonstrate the noise issue, a sample of 3.9-μm brightness temperatures from 15 pixels centered over the ocean were collected on the 23 April case for a 30-min period (0130–0200 UTC) in regions without cloud cover (Fig. 10a). Assuming the actual brightness temperatures at these pixels are constant throughout the period, deviations from the 30-min means at each pixel represent the system noise. The brightness temperatures here have an approximately Gaussian distribution, with a standard deviation of ~0.083 K (Fig. 11a), which is just below the documented ABI specifications for noise equivalent differential temperature (e.g., Table 2 in Schmit et al. 2017). In that same 30-min period, only 56% of pixels in the entire image had 1-min changes in brightness temperature with time that were larger in magnitude. This is an issue for the VOF approach that requires accurate measurements for the brightness time derivative (I_t) to properly minimize (1), as noise violates the brightness constancy assumption and must be overcome by regularization. Without regional information, PM is even more susceptible to noise-related brightness constancy violations that can be inferred from the poor performance in Fig. 2 and significantly larger errors in LT targets in Fig. 3b. Doing the same analysis with 0.64 μm, the ocean pixels have a reflectance factor standard deviation of ~0.0014, which ~75% of I_t magnitudes exceed at 1-min cadence in the 30-min period (Fig. 10b). With the time difference increased to 5-min, 73% of 3.9-μm pixels exceed the 3.9-μm standard deviation, increasing the contrast of I_t within cloud features (Fig. 11b), and resulting in the increased accuracy when compared to truth winds. Increasing the cadence beyond 5 min increases the impact of nonlinearities on the optimization of (1), reducing the VOF-based AMV accuracy.

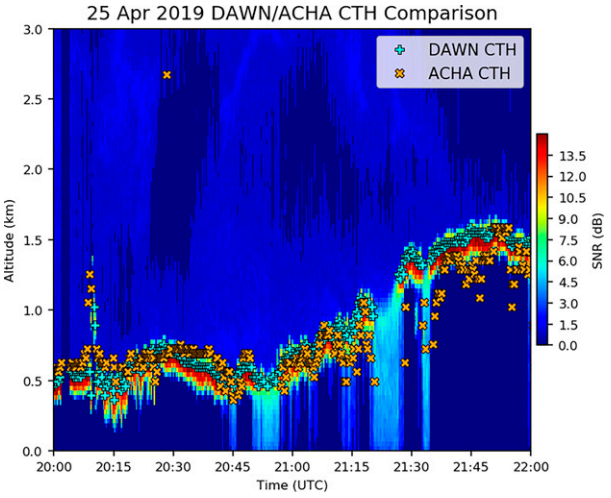

FIG. 8. Sample time series of DAWN signal-to-noise ratio (SNR; dB) on 25 Apr 2019 shown with CTH inferred from the height of the highest 10-dB return (DAWN CTH; cyan plus) compared with the CTH from ACHA (orange ×) at the nearest pixel.

5. Conclusions

A quantitative validation technique to compare OF approaches for retrieving winds from clouds and WV gradients in satellite imagery sequences was developed. The technique takes advantage of a NASA airborne Doppler wind lidar dataset collected in conjunction with 1-min geostationary satellite imagery and rewards approaches that return winds with both accurate tracking and height assignment. This comparison technique is essential toward identifying best practices in OF derivation for AMV retrieval and guiding development of new methods that can track features where AMVs break down. The technique was used to compare wind retrieval from the common PM OF technique used by operational NESDIS AMVs with an advanced VOF approach. The primary findings of this study are as follows:

- AMV winds from VOF tracking consistently outperformed PM techniques for all channels and case studies explored. The best VOF winds retrievals come from the low-noise, high-temporal-resolution, and fine-spatial-resolution 0.64-μm imagery, while VOF struggles more with winds derivations in lower texture scenes for 3.9- and 6.9-μm channel data.
- Partitioning samples by high- and low-texture regions, VOF produces significantly more accurate AMVs in low-texture targets than PM for all channels explored. The VOF winds performance is also not significantly reduced in this sample between high- and low-texture targets, implying less sensitivity to absence of local features. Partitioning samples by number of cloud layers, both VOF and PM have reduced accuracy within multilayer scenes, likely caused by either proximity to motion discontinuities or issues in motion height assignment. Despite the performance reduction, VOF winds are still significantly more accurate in multilayer scenes versus PM winds. The performance improvements in these samples demonstrate that VOF successfully tracks

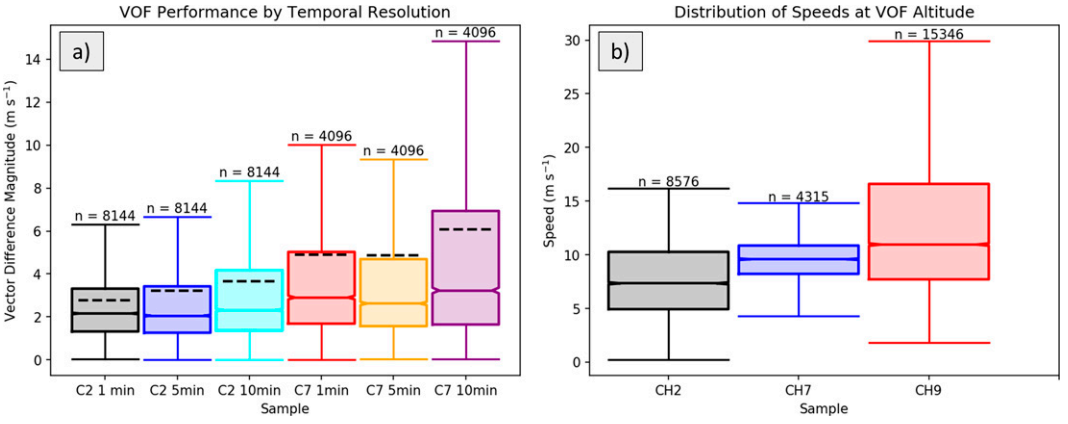

FIG. 9. Boxplots with the same whiskers and interquartile range as Fig. 3 showing (a) the VDMs of tracked AMVs with different channels and different temporal resolutions and the mean VDMs highlighted with black dashed lines and (b) the speed distributions of the VOF datasets sampled by channel. Note that samples for (a) using 1- and 5-min resolution do not include winds where 10-min resolution cannot be computed (when the mesoscale sectors move) for direct comparison, hence the smaller sample size.

features where routines used within operational AMVs commonly break down and require quality control pruning for accuracy.

- The best validation benchmark score produced here was with a GFS first guess augmented with quality-controlled observations from the 0.64-μm VOF AMVs alone. Supplementing quality controlled 3.9- or 6.9-μm derived winds slightly reduced the performance of the benchmark over the accurate GFS first guess in all cases explored, due to errors associated with propagating versus advecting targets and inaccuracies in motion height assignment.
- The benchmark was used to demonstrate a slight reduction in wind accuracy when using coarser time steps for 0.64- and 3.9-μm images, especially between 1- and 10-min cadences. This implies that the common GOES-R ABI full-disk cadence is less than optimal even for tracking slow-moving boundary layer cumulus with modern DOF techniques. The 3.9-μm imagery did show an improvement when the temporal resolution was reduced to 5 min, which was found to coincide with an increase in the strength of the natural signal relative to the ABI noise.

The PM approach used here is the same technique within the first steps of the operational NESDIS AMVs; however, NESDIS AMVs use a significant amount of pre- and postprocessing to improve the quality of the retrieved winds that was not tested here (for full details, see Daniels et al. 2010). For example, NESDIS AMVs cluster the 5 × 5 pixel targets within a larger 15 × 15 pixel region to filter out the noise that this benchmark highlights. The same clustering is used for height assignment, which was not tested here. Also, the benchmark by design encourages AMV techniques to retrieve winds for any pixels that were sampled by DAWN, which means that techniques that ordinarily prune any given targets before or after computation must produce solutions via alternative methods such as objective analysis (e.g., Apke et al. 2016, 2018) or hybrid VOF and PM approaches (e.g., Brox and Malik 2011; Weinzaepfel et al. 2013). This study highlights

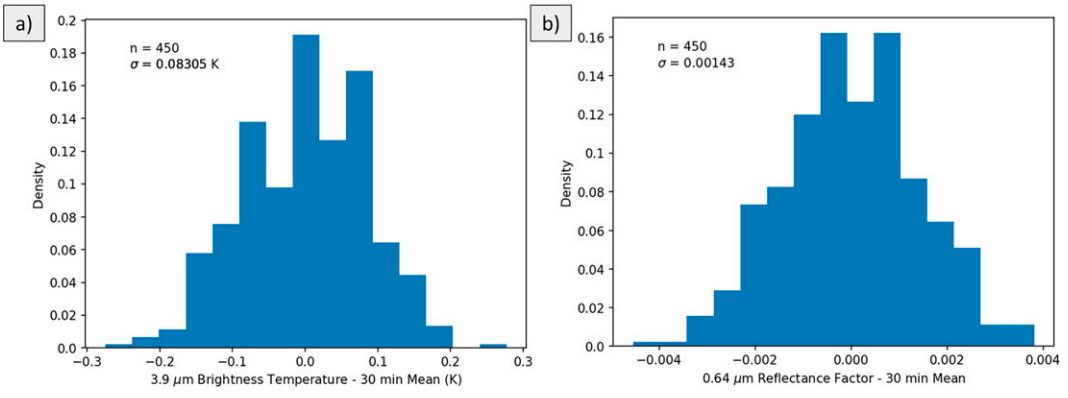

FIG. 10. Histograms of the (a) 3.9-μm channel brightness temperature and (b) 0.64-μm reflectance factors from the points highlighted in Fig. 8, normalized by the mean values from 0130 to 0200 UTC 23 Apr 2019. Also shown are the sample size (n) and standard deviation (σ) for each distribution over the 30-min period.

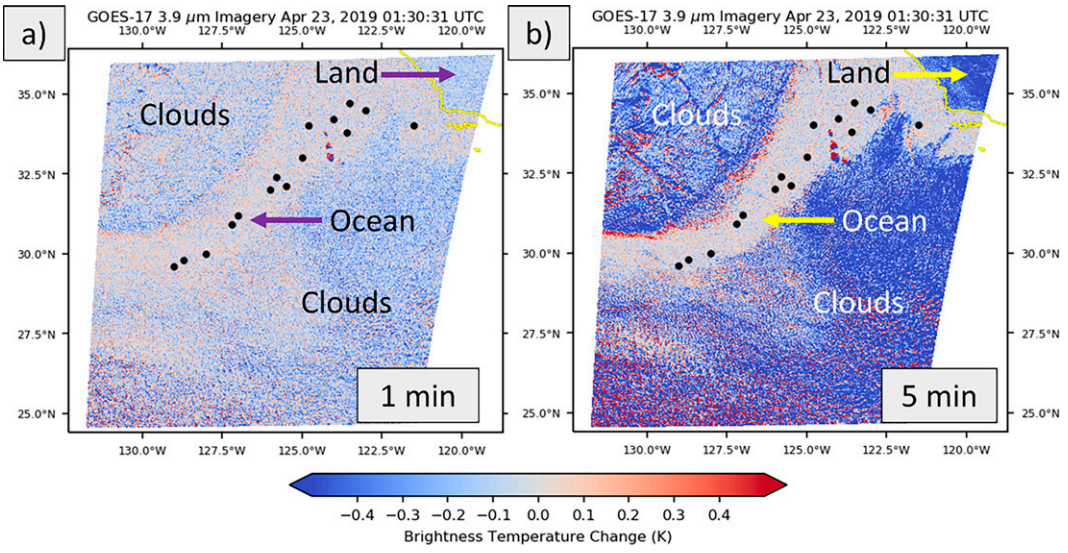

FIG. 11. Sample 3.9-μm I_t values (K) at 0130 UTC 23 Apr 2019 computed with (a) 1- and (b) 5-min image separation, shown with the pixel locations used for estimating channel noise (black points).

the weaknesses of PM by restricting pruning here, though it is plausible these alternative methods with stricter QC could reduce the RMSVD values observed in Table 3. The benchmark created here hence does not discourage the overall use of PM or different height assignment techniques; rather, it encourages research into finding the best methods for deriving any given wind at any given pixel and, by extension, identifying weaknesses in certain AMV retrieval practices.

The results here motivate further exploration into novel DOF derivation approaches as imaging capabilities are improved. The VOF approach is just one of many DOF approaches that this comparison technique can be used to evaluate. Other techniques using machine learning (e.g., Vandal and Nemani 2022) offer an exciting alternative to VOF where traditional constraints and assumptions break down, such as WV regions with gravity wave propagation. In addition to other techniques, the benchmark can be used to evaluate the impacts of modifications to input images such as spectral filters and image spatial and temporal smoothing, quality control techniques such as acceleration checks, and imaging strategies such as channels used and spatial and temporal resolutions. Such modifications to the inputs may have different impacts on AMV accuracy depending on the DOF retrieval approach used. Moreover, best practices learned for retrieving atmospheric winds from drifting features in imagery here can be extended to other remote sensing datasets, such as DOF derivation from radar (e.g., Bechini and Chandrasekar 2017) or lidar (e.g., Dérian et al. 2015) scan sequences.

Aside from new DOF techniques, there are many future research opportunities in DOF-based AMV development. The height-assignment and profile filling techniques could be refined depending on the approach used. This study employed a pixelwise height and simple regional averaging-based height assignment for VOF techniques, which does not always effectively capture OF solution smoothing across a multitude of cloud and WV textures. Using conventional height assignment within the OF derivation to prevent OF solution smoothing across multiple cloud layers (e.g., Héas et al. 2007) or departing from pixelwise height estimation altogether to infer motion heights following the same regularization and texture constraints as the OF algorithm used may improve benchmark scores. Analogous to the height-assignment problems, the QC algorithms that provide guidance for which AMVs to keep or disregard prior to NWP DA must also be updated when using new DOF techniques, especially to prevent overpruning of newly trackable regions. Moreover, the conventional AMV methods only fill ~3% of the total wind observations in the benchmark dataset. Better benchmark results may be achieved with new products and instruments designed to retrieve more of the wind profile versus only the winds at the CTH and WV heights sampled by ABI. Finally, the validation benchmark method here can easily be extended to other observations made by DAWN or similar high-precision and high-resolution wind profilers. Increasing the frequency and geographic coverage of truth winds datasets will help solidify conclusions made for different instruments, and challenge new AMV algorithms especially when samples contain results that are not well represented by OF assumptions or background NWP models.

Acknowledgments. This research was supported by a NASA New Investigators Program Award 80NSSC21K0919. The authors thank Steve Wanzong and Dr. Yue Li at the Cooperative Institute for Meteorological Satellite Studies at the University of Wisconsin–Madison and Dr. Andrew Heidinger at NOAA/NESDIS for their assistance with using the ACHA algorithm. We also thank Zhaoyan Liu and Sharon Rodier

for their work in producing the DAWN wind retrieval data, and the NASA Langley Research Center and Armstrong Flight Research Center for successfully executing the 2019 Aeolus Cal/Val Test Flight campaign.

Data availability statement. GOES-R satellite imagery used for optical flow computation is publicly available via the National Oceanic and Atmospheric Administration Comprehensive Large Array-Data Stewardship System (NOAA CLASS; http://www.class.noaa.gov). Code to reproduce both the VOF and PM approaches used here is provided at https://github.com/JasonApke/OCTANE. In addition, netCDFs containing data needed to reproduce Table 3, including DAWN lidar data, VOF and GFS wind estimates, times, locations, and height assignments for testing new algorithms, are provided at https://www.doi.org/10.5061/dryad.05qfttf67 and a supplementary Python script to read the files provided at https://doi.org/10.5281/zenodo.7217550.

REFERENCES

Anandan, P., 1989: A computational framework and an algorithm for the measurement of visual motion. *Int. J. Comput. Vision*, **2**, 283–310, https://doi.org/10.1007/BF00158167.

Apke, J. M., 2021: Optical flow Code for Tracking, Atmospheric motion vector, and Nowcasting Experiments (OCTANE). GitHub, https://github.com/JasonApke/OCTANE.

——, and J. R. Mecikalski, 2021: On the origin of rotation derived from super rapid scan satellite imagery at the cloud tops of severe deep convection. *Mon. Wea. Rev.*, **149**, 1827–1851, https://doi.org/10.1175/MWR-D-20-0209.1.

——, ——, and C. P. Jewett, 2016: Analysis of mesoscale atmospheric flows above mature deep convection using super rapid scan geostationary satellite data. *J. Appl. Meteor. Climatol.*, **55**, 1859–1887, https://doi.org/10.1175/JAMC-D-15-0253.1.

——, ——, K. M. Bedka, E. W. McCaul Jr., C. R. Homeyer, and C. P. Jewett, 2018: Relationships between deep convection updraft characteristics and satellite based super rapid scan mesoscale atmospheric motion vector derived flow. *Mon. Wea. Rev.*, **146**, 3461–3480, https://doi.org/10.1175/MWR-D-18-0119.1.

——, K. A. Hilburn, S. D. Miller, and D. A. Peterson, 2020: Towards objective identification and tracking of convective outflow boundaries in next-generation geostationary satellite imagery. *Atmos. Meas. Tech.*, **13**, 1593–1608, https://doi.org/10.5194/amt-13-1593-2020.

Auligne, T., and Coauthors, 2016: Forecast Sensitivity and Observation Impact (FSOI) inter-comparison experiment. *Sixth WMO Workshop on the Impact of Various Observation Systems on NWP*, Shanghai, China, Amer. Geophys. Union.

Baker, S., D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, 2011: A database and evaluation methodology for optical flow. *Int. J. Comput. Vision*, **92**, 1–31, https://doi.org/10.1007/s11263-010-0390-2.

Bechini, R., and V. Chandrasekar, 2017: An enhanced optical flow technique for radar nowcasting of precipitation and winds. *J. Atmos. Oceanic Technol.*, **34**, 2637–2658, https://doi.org/10.1175/JTECH-D-17-0110.1.

Bedka, K. M., and J. R. Mecikalski, 2005: Application of satellite-derived atmospheric motion vectors for estimating mesoscale flows. *J. Appl. Meteor.*, **44**, 1761–1772, https://doi.org/10.1175/JAM2264.1.

——, C. S. Velden, R. A. Petersen, W. F. Feltz, and J. R. Mecikalski, 2009: Comparisons of satellite-derived atmospheric motion vectors, rawinsondes, and NOAA Wind Profiler observations. *J. Appl. Meteor. Climatol.*, **48**, 1542–1561, https://doi.org/10.1175/2009JAMC1867.1.

——, and Coauthors, 2021: Airborne lidar observations of wind, water vapor, and aerosol profiles during the NASA Aeolus Calibration and Validation (Cal/Val) Test Flight Campaign. *Atmos. Meas. Tech.*, **14**, 4305–4334, https://doi.org/10.5194/amt-14-4305-2021.

Bessho, K., and Coauthors, 2016: An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites. *J. Meteor. Soc. Japan*, **94**, 151–183, https://doi.org/10.2151/jmsj.2016-009.

Black, M. J., and P. Anandan, 1996: The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Comput. Vision Image Understanding*, **63**, 75–104, https://doi.org/10.1006/cviu.1996.0006.

Bormann, N., A. Hernandez-Carrascal, R. Borde, H. J. Lutz, J. A. Otkin, and S. Wanzong, 2014: Atmospheric motion vectors from model simulations. Part I: Methods and characterization as single-level estimates of wind. *J. Appl. Meteor. Climatol.*, **53**, 47–64, https://doi.org/10.1175/JAMC-D-12-0336.1.

Bresky, W. C., J. M. Daniels, A. A. Bailey, and S. T. Wanzong, 2012: New methods toward minimizing the slow speed bias associated with atmospheric motion vectors. *J. Appl. Meteor. Climatol.*, **51**, 2137–2151, https://doi.org/10.1175/JAMC-D-11-0234.1.

Brox, T., 2005: From pixels to regions: Partial differential equations in image analysis. Ph.D. thesis, Saarland University, 202 pp.

——, and J. Malik, 2011: Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, **33**, 500–513, https://doi.org/10.1109/TPAMI.2010.143.

——, A. Bruhn, N. Papenberg, and J. Weickert, 2004: High accuracy optical flow estimation based on a theory for warping. *2004 European Conf. on Computer Vision*, Prague, Czech Republic, ECCV, 25–36.

Butler, D. J., J. Wulff, G. B. Stanley, and M. J. Black, 2012: A naturalistic open source movie for optical flow evaluation. *2012 European Conf. on Computer Vision*, Florence, Italy, ECCV, 611–625, https://doi.org/10.1007/978-3-642-33783-3_44.

Corpetti, T., D. Heitz, G. Arroyo, E. Mémin, and A. Santa-Cruz, 2006: Fluid experimental flow estimation based on an optical-flow scheme. *Exp. Fluids*, **40**, 80–97, https://doi.org/10.1007/s00348-005-0048-y.

Daniels, J. M., W. C. Bresky, S. T. Wanzong, C. S. Velden, and H. Berger, 2010: GOES-R Advanced Baseline Imager (ABI) algorithm theoretical basis document for derived motion winds. NOAA/NESDIS/STAR Algorithm Theoretical Basis Doc., version 2.5, 98 pp., https://www.goes-r.gov/products/baseline-derived-motion-winds.html.

——, ——, A. A. Bailey, A. Allegrino, S. T. Wanzong, and C. S. Velden, 2019: A closer look at *GOES-16/17* atmospheric motion vectors derived from the Advanced Baseline Imager (ABI). *2019 Joint Satellite Conf.*, Boston, MA, Amer. Meteor. Soc., 15B.1, https://ams.confex.com/ams/JOINTSATMET/meetingapp.cgi/Paper/359039.

Dérian, P., C. F. Mauzey, and S. D. Mayor, 2015: Wavelet-based optical flow for two-component wind field estimation from

single aerosol lidar data. *J. Atmos. Oceanic Technol.*, **32**, 1759–1778, https://doi.org/10.1175/JTECH-D-15-0010.1.

Farnebäck, G., 2001: Two-frame motion estimation based on polynomial expansion. *Proc. Eighth IEEE Int. Conf. on Computer Vision*, Vancouver, BC, Canada, IEEE, 171–177.

Fortun, D., P. Bouthemy, and C. Kervrann, 2015: Optical flow modeling and computation: A survey. *Comput. Vision Image Understanding*, **134**, 1–21, https://doi.org/10.1016/j.cviu.2015.02.008.

Geiger, A., P. Lenz, and R. Urtasun, 2012: Are we ready for autonomous driving? The KITTI Vision Benchmark Suite. *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Providence, RI, IEEE, 3354–3361, https://doi.org/10.1109/CVPR.2012.6248074.

Héas, P., E. Mémin, N. Papadakis, and A. Szantai, 2007: Layered estimation of atmospheric mesoscale dynamics from satellite imagery. *IEEE Trans. Geosci. Remote Sens.*, **45**, 4087–4104, https://doi.org/10.1109/TGRS.2007.906156.

Heidinger, A. K., 2013: ABI cloud height. NOAA/NESDIS/STAR Algorithm Theoretical Basis Doc., version 3.0, 79 pp., https://www.star.nesdis.noaa.gov/goesr/docs/ATBD/Cloud_Height.pdf.

——, and M. J. Pavolonis, 2009: Gazing at cirrus clouds for 25 years through a split window. Part I: Methodology. *J. Appl. Meteor. Climatol.*, **48**, 1100–1116, https://doi.org/10.1175/2008JAMC1882.1.

——, W. C. Straka, C. C. Molling, J. T. Sullivan, and X. Wu, 2010: Deriving an inter-sensor consistent calibration for the AVHRR solar reflectance data record. *Int. J. Remote Sens.*, **31**, 6493–6517, https://doi.org/10.1080/01431161.2010.496472.

——, A. T. Evan, M. J. Foster, and A. Walther, 2012: A naive Bayesian cloud-detection scheme derived from *CALIPSO* and applied within PATMOS-x. *J. Appl. Meteor. Climatol.*, **51**, 1129–1144, https://doi.org/10.1175/JAMC-D-11-02.1.

Hernandez-Carrascal, A., and N. Bormann, 2014: Atmospheric motion vectors from model simulations. Part II: Interpretation as spatial and vertical averages of wind and role of clouds. *J. Appl. Meteor. Climatol.*, **53**, 65–82, https://doi.org/10.1175/JAMC-D-12-0337.1.

Holmlund, K., 1998: The utilization of statistical properties of satellite-derived atmospheric motion vectors to derive quality indicators. *Wea. Forecasting*, **13**, 1093–1104, https://doi.org/10.1175/1520-0434(1998)013<1093:TUOSPO>2.0.CO;2.

——, C. S. Velden, and M. Rohn, 2001: Enhanced automated quality control applied to high-density satellite-derived winds. *Mon. Wea. Rev.*, **129**, 517–529, https://doi.org/10.1175/1520-0493(2001)129<0517:EAQCAT>2.0.CO;2.

Horn, B. K. P., and B. G. Schunck, 1981: Determining optical flow. *Artif. Intell.*, **17**, 185–203, https://doi.org/10.1016/0004-3702(81)90024-2.

Kavaya, M. J., J. Y. Beyon, G. J. Koch, M. Petros, P. J. Petzar, U. N. Singh, B. C. Trieu, and J. Yu, 2014: The Doppler Aerosol Wind (DAWN) airborne, wind-profiling coherent-detection lidar system: Overview and preliminary flight results. *J. Atmos. Oceanic Technol.*, **31**, 826–842, https://doi.org/10.1175/JTECH-D-12-00274.1.

Kim, M., H. M. Kim, J. Kim, S.-M. Kim, C. Velden, and B. Hoover, 2017: Effect of enhanced satellite-derived atmospheric motion vectors on numerical weather prediction in East Asia using an adjoint-based observation impact method. *Wea. Forecasting*, **32**, 579–594, https://doi.org/10.1175/WAF-D-16-0061.1.

Lean, P., S. Migliorini, and G. Kelly, 2015: Understanding atmospheric motion vector vertical representativity using a simulation study and first-guess departure statistics. *J. Appl. Meteor. Climatol.*, **54**, 2479–2500, https://doi.org/10.1175/JAMC-D-15-0030.1.

Le Marshall, J., A. Rea, L. M. Leslie, R. Seecamp, and M. Dunn, 2004: Error characterization of atmospheric motion vectors. *Aust. Meteor. Mag.*, **53**, 123–131.

——, J. Jung, T. Zapotocny, C. Redder, M. Dunn, J. Daniels, and L. P. Riishojgaard, 2008: Impact of MODIS atmospheric motion vectors on a global NWP system. *Aust. Meteor. Mag.*, **57**, 45–51.

Lucas, B. D., and T. Kanade, 1981: An iterative image registration technique with an application to stereo vision. *Proc. Int. Joint Conf. on Artificial Intelligence*, Vancouver, BC, Canada, IJCAI, 674–679.

Menzel, W. P., 2001: Cloud tracking with satellite imagery: From the pioneering work of Ted Fujita to the present. *Bull. Amer. Meteor. Soc.*, **82**, 33–48, https://doi.org/10.1175/1520-0477(2001)082<0033:CTWSIF>2.3.CO;2.

Merrill, R. T., W. P. Menzel, W. Baker, J. Lynch, and E. Legg, 1991: A report on the recent demonstration of NOAA's upgraded capability to derive cloud motion satellite winds. *Bull. Amer. Meteor. Soc.*, **72**, 372–376, https://doi.org/10.1175/1520-0477-72.3.372.

NESDIS, 2018: GOES-17 loop heat pipe fact sheet. NOAA Satellite and Informative Service Doc., 1 p., https://www.goes-r.gov/imagesContent/newsItems/NESDISfactsheetLoopheatpipe.pdf.

Nieman, S. J., J. Schmetz, and W. P. Menzel, 1993: A comparison of several techniques to assign heights to cloud tracers. *J. Appl. Meteor.*, **32**, 1559–1568, https://doi.org/10.1175/1520-0450(1993)032<1559:ACOSTT>2.0.CO;2.

——, W. P. Menzel, C. M. Hayden, D. Gray, S. T. Wanzong, C. S. Velden, and J. Daniels, 1997: Fully automatic cloud drift winds in NESDIS operations. *Bull. Amer. Meteor. Soc.*, **78**, 1121–1134, https://doi.org/10.1175/1520-0477(1997)078<1121:FACDWI>2.0.CO;2.

Oyama, R., M. Sawada, and K. Shimoji, 2018: Diagnosis of tropical cyclone intensity and structure using upper tropospheric atmospheric motion vectors. *J. Meteor. Soc. Japan*, **96B**, 3–26, https://doi.org/10.2151/jmsj.2017-024.

Pavolonis, M. J., A. K. Heidinger, and T. Uttal, 2005: Daytime global cloud typing from AVHRR and VIIRS: Algorithm description, validation, and comparisons. *J. Appl. Meteor.*, **44**, 804–826, https://doi.org/10.1175/JAM2236.1.

Posselt, D. J., and Coauthors, 2019: Quantitative assessment of state-dependent atmospheric motion vector uncertainties. *J. Appl. Meteor. Climatol.*, **58**, 2479–2495, https://doi.org/10.1175/JAMC-D-19-0166.1.

Rabin, R. M., S. F. Corfidi, J. C. Brunner, and C. E. Hane, 2004: Detecting winds aloft from water vapour satellite imagery in the vicinity of storms. *Weather*, **59**, 251–257, https://doi.org/10.1256/wea.182.03.

Sandmæl, T. N., C. R. Homeyer, K. M. Bedka, J. M. Apke, J. R. Mecikalski, and K. Khlopenkov, 2019: Evaluating the ability of remote sensing observations to identify significantly severe and potentially tornadic storms. *J. Appl. Meteor. Climatol.*, **58**, 2569–2590, https://doi.org/10.1175/JAMC-D-18-0241.1.

Santek, D., and Coauthors, 2019: 2018 Atmospheric Motion Vector (AMV) intercomparison study. *Remote Sens.*, **11**, 2240, https://doi.org/10.3390/rs11192240.

Sato, Y., and L. P. Riishojgaard, 2016: Report from the Sixth WMO Workshop on the Impact of Various Observing System on NWP. WMO Tech. Rep. CGMS-45, 21 pp.

Schmetz, J., and K. Holmlund, 1992: Operational cloud motion winds from Meteosat and the use of cirrus clouds as tracers. *Adv. Space Res.*, **12**, 95–104, https://doi.org/10.1016/0273-1177(92)90202-9.

——, ——, J. Hoffman, B. Strauss, B. Mason, V. Gaertner, A. Koch, and L. Van De Berg, 1993: Operational cloud-motion winds from Meteosat infrared images. *J. Appl. Meteor.*, **32**, 1206–1225, https://doi.org/10.1175/1520-0450(1993)032<1206:OCMWFM>2.0.CO;2.

Schmit, T. J., P. Griffith, M. M. Gunshor, J. M. Daniels, S. J. Goodman, and W. J. Lebair, 2017: A closer look at the ABI on the GOES-R series. *Bull. Amer. Meteor. Soc.*, **98**, 681–698, https://doi.org/10.1175/BAMS-D-15-00230.1.

Shi, J., and Tomasi, 1994: Good features to track. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Seattle, WA, IEEE, https://doi.org/10.1109/CVPR.1994.323794.

Stettner, D., C. Velden, R. Rabin, S. Wanzong, J. Daniels, and W. Bresky, 2019: Development of enhanced vortex-scale atmospheric motion vectors for hurricane applications. *Remote Sens.*, **11**, 1981, https://doi.org/10.3390/rs11171981.

Sun, D., S. Roth, and M. J. Black, 2014: A quantitative analysis of current practices in optical flow estimation and the principles behind them. *Int. J. Comput. Vision*, **106**, 115–137, https://doi.org/10.1007/s11263-013-0644-x.

Sundaram, N., T. Brox, and K. Keutzer, 2010: Dense point trajectories by GPU-accelerated large displacement optical flow. *11th European Conf. on Computer Vision*, Crete, Greece, ECCV, 438–451, https://doi.org/10.1007/978-3-642-15549-9_32.

Tan, B., and Coauthors, 2020: GOES-R series image navigation and registration performance assessment tool set. *J. Appl. Remote Sens.*, **14**, 032405, https://doi.org/10.1117/1.JRS.14.032405.

Vandal, T. J., and R. R. Nemani, 2022: Temporal interpolation of geostationary satellite imagery with optical flow. *IEEE Trans. Neural Networks Learn. Syst.*, https://doi.org/10.1109/TNNLS.2021.3101742, in press.

Velden, C. S., and K. M. Bedka, 2009: Identifying the uncertainty in determining satellite-derived atmospheric motion vector height attribution. *J. Appl. Meteor. Climatol.*, **48**, 450–463, https://doi.org/10.1175/2008JAMC1957.1.

——, C. M. Hayden, S. J. W. Nieman, W. P. Menzel, S. Wanzong, and J. S. Goerss, 1997: Upper-tropospheric winds derived from geostationary satellite water vapor observations. *Bull. Amer. Meteor. Soc.*, **78**, 173–196, https://doi.org/10.1175/1520-0477(1997)078<0173:UTWDFG>2.0.CO;2.

——, T. L. Olander, and S. Wanzong, 1998: The impact of multispectral *GOES-8* wind information on Atlantic tropical cyclone track forecasts in 1995. Part I: Dataset methodology, description, and case analysis. *Mon. Wea. Rev.*, **126**, 1202–1218, https://doi.org/10.1175/1520-0493(1998)126<1202:TIOMGW>2.0.CO;2.

——, and Coauthors, 2005: Recent innovations in deriving tropospheric winds from meteorological satellites. *Bull. Amer. Meteor. Soc.*, **86**, 205–224, https://doi.org/10.1175/BAMS-86-2-205.

Walther, A., and A. K. Heidinger, 2012: Implementation of the daytime cloud optical and microphysical properties algorithm (DCOMP) in PATMOS-x. *J. Appl. Meteor. Climatol.*, **51**, 1371–1390, https://doi.org/10.1175/JAMC-D-11-0108.1.

Weinzaepfel, P., J. Revaud, Z. Harchaoui, and C. Schmid, 2013: DeepFlow: Large displacement optical flow with deep matching. *IEEE Int. Conf. on Computer Vision*, Sydney, NSW, Australia, IEEE, 1385–1392, https://doi.org/10.1109/ICCV.2013.175.

Wu, Q., H.-Q. Wang, Y.-J. Lin, Y.-Z. Zhuang, and Y. Zhang, 2016: Deriving AMVs from geostationary satellite images using optical flow algorithm based on polynomial expansion. *J. Atmos. Oceanic Technol.*, **33**, 1727–1747, https://doi.org/10.1175/JTECH-D-16-0013.1.

Wu, T. C., H. Liu, S. J. Majumdar, C. S. Velden, and J. L. Anderson, 2014: Influence of assimilating satellite-derived atmospheric motion vector observations on numerical analyses and forecasts of tropical cyclone track and intensity. *Mon. Wea. Rev.*, **142**, 49–71, https://doi.org/10.1175/MWR-D-13-00023.1.

Zeng, X., S. Ackerman, R. D. Ferraro, T. J. Lee, J. J. Murray, S. Pawson, C. Reynolds, and J. Teixeira, 2016: Challenges and opportunities in NASA weather research. *Bull. Amer. Meteor. Soc.*, **97**, ES137–ES140, https://doi.org/10.1175/BAMS-D-15-00195.1.

Zimmer, H., A. Bruhn, and J. Weickert, 2011: Optic flow in harmony. *Int. J. Comput. Vision*, **93**, 368–388, https://doi.org/10.1007/s11263-011-0422-6.