Maximizing the Volume of Collocated Data from Two Coordinated Suborbital Platforms

JOSEPH S. SCHLOSSER,[a,b] RYAN BENNETT,[c] BRIAN CAIRNS,[d] GAO CHEN,[a] BRIAN L. COLLISTER,[a] JOHNATHAN W. HAIR,[a] MICHAEL JONES,[a,e] MICHAEL A. SHOOK,[a] ARMIN SOROOSHIAN,[f,g,h] KENNETH L. THORNHILL,[a,e] LUKE D. ZIEMBA,[a] AND SNORRE STAMNES[a]

[a] *NASA Langley Research Center, Hampton, Virginia*
[b] *NASA Postdoctoral Program, NASA Langley Research Center, Hampton, Virginia*
[c] *Bay Area Environmental Research Institute, Ventura, California*
[d] *NASA Goddard Institute for Space Studies, New York, New York*
[e] *Science Systems and Applications, Inc., Lanham, Maryland*
[f] *Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, Arizona*
[g] *Department of Chemical and Environmental Engineering, The University of Arizona, Tucson, Arizona*
[h] *James C. Wyant College of Optical Sciences, The University of Arizona, Tucson, Arizona*

(Manuscript received 3 January 2023, in final form 21 December 2023, accepted 17 January 2024)

ABSTRACT: Suborbital (e.g., airborne) campaigns that carry advanced remote sensing and in situ payloads provide detailed observations of atmospheric processes, but can be challenging to use when it is necessary to geographically collocate data from multiple platforms that make repeated observations of a given geographic location at different altitudes. This study reports on a data collocation algorithm that maximizes the volume of collocated data from two coordinated suborbital platforms and demonstrates its value using data from the NASA Aerosol Cloud Meteorology Interactions Over the western Atlantic Experiment (ACTIVATE) suborbital mission. A robust data collocation algorithm is critical for the success of the ACTIVATE mission goal to develop new and improved remote sensing algorithms, and quantify their performance. We demonstrate the value of these collocated data to quantify the performance of a recently developed vertically resolved lidar + polarimeter–derived aerosol particle number concentration (N_a) product, resulting in a range-normalized mean absolute deviation (NMAD) of 9% compared to in situ measurements. We also show that this collocation algorithm increases the volume of collocated ACTIVATE data by 21% compared to using only nearest-neighbor finding algorithms alone. Additional to the benefits demonstrated within this study, the data files and routines produced by this algorithm have solved both the critical collocation and the collocation application steps for researchers who require collocated data for their own studies. This freely available and open-source collocation algorithm can be applied to future suborbital campaigns that, like ACTIVATE, use multiple platforms to conduct coordinated observations, e.g., a remote sensing aircraft together with in situ data collected from suborbital platforms.

SIGNIFICANCE STATEMENT: This study describes a data collocation (i.e., selection) process that aims to maximize the volume of data identified to be simultaneously collected in time and space from two coordinated measurement platforms. The functional utility of the resultant dataset is also demonstrated by extending the validation of aerosol particle number concentration derived from standard lidar and polarimeter data products from a suborbital mission that used two aircraft platforms.

KEYWORDS: Atmosphere; Filtering techniques; Statistics; Aerosol optical properties; Aerosols/particulates; Measurements

1. Introduction

Advanced suborbital remote sensing instruments such as multiangle, multispectral polarimeters and multispectral lidar are increasingly used for observations of meteorological, gas, aerosol particle, and cloud properties in the atmosphere. Over the past decade, National Aeronautics and Space Administration (NASA) remote sensing instruments have been deployed in numerous suborbital campaigns that feature multiple aircraft platforms encompassing in situ and remote sensing measurements. This increasing trend of using multiple aircraft is born of the need to both validate remote sensor measurements and retrievals and the need to leverage the strengths of both remote sensing instruments and in situ instruments in studies of the atmosphere.

To leverage data gathered from multiple coordinated aircraft platforms, the data must be collocated. Following the defined convention established by previous works, "collocation" is defined as the act of matching data gathered from separate measurement platforms using predetermined spatial distance and measurement time difference thresholds (Finlon et al. 2022). The exact spatiotemporal threshold that is appropriate for collocating data depends on the type of processes that are

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

Corresponding authors: Joseph S. Schlosser, joseph.s.schlosser@nasa.gov; Snorre Stamnes, snorre.a.stamnes@nasa.gov

DOI: 10.1175/JTECH-D-23-0001.1

of interest. Furthermore, instruments can have either single-pixel swath (i.e., along-track only) or multipixel swath, differing pixel resolutions, and collect data from one-dimensional (1D) point-like in situ data or single-pixel polarimeter data (e.g., x), to 2D image or lidar data (e.g., x, z or x, y), to 3D radar data (e.g., x, y, z), where x is latitude, y is longitude, and z is altitude or depth.

This study aims to add to the traditional methods used for nearest-neighbor finding problems by searching for multiple discrete time segments rather than finding the k nearest neighbors of the entire dataset. This study uses simple brute-force nearest-neighbor finding to serve as the method for nearest-neighbor finding, but there are two possible methods that can be used to optimize the nearest-neighbor finding method process. These methods are k–d nearest-neighbor (KNN) finding and balltree finding, which reduce the computational burden that would be required for the brute-force calculation of all the distances between different pixels in the radar swath and the in situ measurements taken with a second platform (Friedman et al. 1977; Omohundro 1989).

There have been several studies that have developed and implemented collocation algorithms specifically for multipixel-swath 3D radar-derived cloud products (Heymsfield et al. 2020; Duffy et al. 2021; Chase et al. 2018). These works have generally implemented KNN finding methods that optimize the nearest-neighbor problem by searching only neighbors that are closest to each other (Friedman et al. 1977). These methods are useful in these contexts; however, they do add complexity to the collocation algorithm and having accurate distances between the platforms is a feature of this process and are provided in an output file for future researchers working on each project.

Multiple studies on collocation have been conducted but are related to satellite data, whereas this study focuses on suborbital data (Nalli et al. 2018; Buehler et al. 2004). These studies do illustrate applications of nearest-neighbor finding; however, they do not consider specific discontinuities where multiple disparate sets of data points from one platform can be valid for collocation to a given data point from a second platform. To our knowledge, there have not been any published collocation algorithms that extend the traditional nearest-neighbor methods to maximize the volume of data that are collocated, which is especially important for missions where remote sensing instruments provide vertically resolved 2D or 3D data together with in situ point-like measurements. While taking advantage of all viable data is in of itself a valuable goal, several studies have called for an increase in the amount of collocated in situ and remote sensing data from multiple aircraft (e.g., Sawamura et al. 2017; Gao et al. 2019; Pistone et al. 2019; Sorooshian et al. 2019).

As more researchers need to study data combined from separate measurement platforms, developing an efficient solution to the challenge of maximizing the available comparable data that are gathered from multiple coordinated platforms is needed. There are a several suborbital campaigns that feature multiple coordinated aircraft platforms with either in situ instruments, remote sensing instruments, or both remote sensing and in situ measurements. There are also suborbital campaigns that feature aircraft and marine vessel platforms including ground stations in their coordinated operation. Table 1 lists several campaigns that feature two or more platforms that can benefit from the multiplatform collocation algorithm described in this work. While many of the missions on this list are mature enough that researchers have manually collocated their data, this collocation effort takes time and energy for each researcher. In addition, any mistakes made in implementing collocation can lead to incorrect results. This work aims to unify the collocation process and make it easier for researchers to focus on data analyses steps.

ACTIVATE serves as an ideal campaign to demonstrate the collocation algorithm as it is a suborbital research campaign dedicated to extensive coordination between two aircraft (Sorooshian et al. 2019). A King Air collected remote sensing data (i.e., lidar and polarimetry) and released dropsondes, while flying at high altitudes usually between 8 and 10 km. Simultaneously, the second aircraft, a Falcon, collected in situ data while operating between approximately 150 m above the ocean surface and just above the top of the planetary boundary layer (PBL) at approximately 1–4 km. Throughout ACTIVATE these two aircraft kept their flight tracks in close spatiotemporal proximity, often within 5 min and 6 km.

In this work, we propose a solution for the selection of data from two spatially and temporally coordinated platforms, which maximizes the potential number of collocation of data samples. Furthermore, the method presented here is instrument agnostic by nature and requires only that there are two platforms with remote sensing or in situ measurements. The collocation algorithm outlined in this work can be used for past and future multiplatform campaigns with coordinated remote sensing and in situ aircraft measurements. While the measurements on the ACTIVATE platforms have fine spatial resolution, i.e., single-pixel-swath 1D passive polarimetric observations, single-pixel-swath but vertically resolved 2D lidar observations, and point-like in situ measurements, the algorithm presented here can also be applied to platforms with multipixel-swath 2D image and 3D radar measurements. We discuss how additional measurement-specific collocation steps can be readily applied for refined collocation of such measurements using this method. Freely available and open-source Python and MATLAB codes are provided to apply this collocation process to other campaigns.

Last, this collocation algorithm is applied to ACTIVATE data and the data collocation mask resulting from this algorithm is also made publicly available for researchers that wish to use data within a specified distance and time interval from both ACTIVATE aircraft in atmospheric science studies. To demonstrate the utility of the collocation algorithm, this ACTIVATE data collocation mask is used to perform a quantitative comparison of ambient aerosol particle number concentration (N_a).

2. Methods

a. *Summary of ACTIVATE*

The ACTIVATE dataset features 162 coordinated science flights across six ACTIVATE deployments that occurred

TABLE 1. Acronym, years active, featured platform, and reference corresponding to several campaigns that feature multiple suborbital platforms focusing on remote sensing and in situ measurements of aerosol and cloud optical and microphysical properties.

Acronym	Years active	Featured platforms	Reference
The North Atlantic Marine Boundary Layer Experiment (NAMBLEX)	2002	One aircraft and one ground station	Heard et al. (2006)
California Nexus (CalNex)	2010	Two aircraft and multiple ground stations	Ryerson et al. (2013)
Carbonaceous Aerosol and Radiative Effects Study (CARES)	2010	Two aircraft and multiple ground stations	Zaveri et al. (2012)
Regional Experiments for Land-atmosphere Exchanges 2012 (REFLEX 2012)	2012	One aircraft and multiple ground stations	Timmermans et al. (2015)
Two-Column Aerosol Project (TCAP)	2012/13	Two aircraft and multiple ground stations	Berg (2016), Müller et al. (2014)
North Atlantic Aerosols and Marine Ecosystems Study (NAAMES)	2015–18	One aircraft and one marine vessel	Behrenfeld et al. (2019)
International Cooperative Air Quality Field Study in Korea (KORUS-AQ)	2016	Two aircraft	Crawford et al. (2021)
Observations of Aerosols above Clouds and their Interactions (ORACLES)	2016/17	Two aircraft and multiple ground stations	Redemann et al. (2021)
Cloud, Aerosol and Monsoon Processes Philippines Experiment ($CAMP^2Ex$)	2019	Two aircraft and multiple fixed ground stations	Reid et al. (2023)
Fire Influence on Regional to Global Environments Experiment-Air Quality (FIREX-AQ)	2019	Four aircraft, two mobile ground stations, multiple fixed ground stations, and one unmanned aircraft	Warneke et al. (2023)
Aerosol Cloud Meteorology Interactions Over the western Atlantic Experiment (ACTIVATE)	2019–22	Two aircraft and multiple ground stations	Sorooshian et al. (2019)
Investigation of Microphysics and Precipitation for Atlantic Coast-Threatening Snowstorms (IMPACTS)	2020–present	Two aircraft and multiple ground stations	McMurdie et al. (2022)
Arctic Radiation-Cloud-Aerosol-Surface-Interaction Experiment (ARCSIX)	Planned for 2024	Two aircraft and multiple ground stations	https://espo.nasa.gov/ARCSIX_White_Paper
Plankton, Aerosol, Cloud, ocean Ecosystem Postlaunch Airborne Experiment (PACE-PAX)	Planned for 2024	Two aircraft	https://espo.nasa.gov/pace-pax

between 14 February 2020 and 18 June 2022. The six ACTIVATE deployments occurred between the following dates:

1) 14 February–12 March 2020,
2) 13 August–30 September 2020,
3) 27 January–2 April 2021,
4) 13 May–30 June 2021,
5) 30 November 2021–29 March 2022, and
6) 3 May–18 June 2022.

ACTIVATE provides a rich dataset to investigate numerous atmospheric processes over the western North Atlantic Ocean, including aerosol–cloud interactions that represent the largest uncertainty in estimates of total anthropogenic radiative forcing (IPCC 2014). During the first 5.5 ACTIVATE deployments, these statistical surveys and process studies were carried out using NASA Langley Research Center in Virginia as a base of operations. The final half of the sixth ACTIVATE deployment featured Bermuda as the base of operations. Historically, there have only been a limited number of aerosol–cloud interaction studies that focused on the western North Atlantic (e.g., Quinn et al. 2019; Sorooshian et al. 2020; Dadashazar et al. 2021b,a). The ACTIVATE methodology and dataset are described in more detail in Sorooshian et al. (2023).

ACTIVATE fills this critical knowledge gap in the dynamic western North Atlantic environment with its wide range of meteorology and aerosol species (Corral et al. 2021; Painemal et al. 2021). For the majority of the year, the western North Atlantic's persistent cloud cover, only temporarily interspersed with clear-sky conditions amid broken cloud fields, makes passive remote sensing measurements of aerosol properties in this region very challenging (Feingold 2003; Braun et al. 2021; Painemal et al. 2021). The advanced passive and active remote sensing and in situ ACTIVATE dataset is important for understanding processes governed by aerosol particle and cloud drop number concentrations, but collocation with simultaneous in situ measurements is critical to assess and advance the capabilities of lidar and polarimetric remote sensing of aerosol and cloud properties. We demonstrate the usefulness of this collocation product by applying it to a novel 2D lidar and polarimeter–derived N_a retrieval that was developed specifically for ACTIVATE (Schlosser et al. 2022).

b. Data

The 1-Hz navigational data from each aircraft are used to create the data collocation mask. The navigational data include the time, latitude, and longitude of the aircraft, which are used to define a navigational point. Time and aircraft altitude, latitude, and longitude are provided by an Applanix POSAV 610 for the King Air and the Falcon. The latitude and longitude measurements used for collocation are both accurate to within 1.5 m for each aircraft. In addition to the navigational data used for collocation, this study also uses the measurements and methods outlined in Schlosser et al. (2022) to demonstrate an important practical application of the data collocation mask for ACTIVATE's science objective to improve remote sensing retrievals of N_a. For this application, Research Scanning Polarimeter (RSP) and multiwavelength High Spectral Resolution Lidar (HSRL-2) data are used from the King Air and Laser Aerosol Spectrometer (LAS) and Cloud Droplet Probe (CDP) data are used from the Falcon.

The RSP aerosol product is based on an optimal estimate using the Research Scanning Polarimeter Microphysical Aerosol Properties from Polarimetry (RSP-MAPP) algorithm (Stamnes et al. 2018). Fine- and coarse-mode aerosol optical and microphysical properties are directly retrieved using seven channels that measure the total and polarized radiance across the visible-shortwave spectrum (wavelength = 410–2260 nm) with over 100 viewing angles between ±55°. The RSP has a field of view of 14 mrad, which results in a 126-m footprint for an aircraft at 9 km altitude. The HSRL-2 products include ambient vertically resolved lidar backscattering and extinction coefficients and ambient linear depolarization ratio (LDR) at wavelengths of 355, 532, and 1064 nm (Fernald 1984; Hair et al. 2008; Burton et al. 2018). The HSRL-2 field of view is 1 mrad, which corresponds to a 9-m footprint for an aircraft at 9-km altitude.

The in situ N_a measurements are taken from the LAS (Model 3340, TSI, Inc.), which measures concentrations of particles with dry particle diameter (D) ranging in sizes from 94 to 7500 nm at a 1 Hz temporal resolution. The LAS samples were actively dried with a 6″ Perma Pure Monotube Dryer 700 for all but 30 flights. The 30 flights between 14 May and 30 June 2021 passively dried using ram heating. The N_a measurements provided by the LAS are provided at standard temperature and pressures (273.15 K and 1013 mb). While the LAS has a measurement range up to 7500 nm, the maximum cutoff D of the sample inlet prevents the measurement of particles with ambient D greater than 5000 nm (McNaughton et al. 2007; Chen et al. 2011). To take into account potential hygroscopic effects, we only include particles with dry optical D up to 3488 nm in this analysis. With the total N_a measured by the LAS referred to from this point forward as N_{LAS}.

Ambient liquid water content (LWC) and N_d are used to classify in situ data as cloud-free, ambiguous, or cloud. Ambient LWC and N_d are both derived from ambient particle size distribution measured by a CDP (Droplet Measurement Technologies; Sinclair et al. 2019). The CDP can measure particles in the ambient D size range of 2000–50 000 nm and the N_d derived by the CDP is noted by N_{CDP}. Measurements where LWC is between 0.001 and 0.02 g m^{-3} and N_d is between 5 and 50 cm^{-3} are classified as ambiguous, i.e., not entirely cloud-free. Thus, for this study, measurements are considered cloud-free where LWC and N_d are less than 0.001 g m^{-3} and 5 cm^{-3}, respectively.

The two ACTIVATE aircraft executed flights that can be broadly categorized into two mission types: "statistical surveys" and "process studies." The average research flight duration for all ACTIVATE flights is 3.3 h. During the statistical surveys that comprised 89% of missions, the King Air would fly at cruising altitude (8–10 km) while the Falcon would fly a vertical stair-stepping pattern in, and just above, the PBL (0.15–4 km). The utility of the vertical stair-stepping pattern is outlined in detail in a previous study (Dadashazar et al. 2022), and allows for the efficient in situ characterization of gas, cloud, aerosol, and meteorological quantities of the PBL across multiple flights and deployments.

An example of the statistical survey flight pattern is shown in Fig. 1. The Falcon starts by flying away from the coast in the stair-stepping pattern. Once the Falcon reaches 73°W, it reverses course and then performs a vertical spiral sounding before resuming the stair-stepping pattern along the return path. After takeoff, the King Air proceeds to reach a nominal altitude of 9 km and retains that altitude while flying out and back along the same path.

By contrast, process study missions are focused on understanding specific atmospheric processes such as cold air outbreaks (Corral et al. 2022; Li et al. 2022; Tornow et al. 2022). During some of these process study missions, the Falcon would execute a back and forth stair-stepping flight pattern in a single vertical column, i.e., a wall pattern, while the King Air uses remote sensing measurements and dropsondes to characterize the area from aloft, often by flying in a large circular pattern. An example of the flight paths the two aircraft take during this type of process study flight is depicted in Fig. 2.

By using Bermuda as a base for the statistical surveys and process studies, the final ACTIVATE deployment is a unique dataset. One reason this deployment is unique is that it features transect flights where the two aircraft fly between LaRC and Bermuda. The last of these transect flights occurred on 18 June 2022. The flight paths that the two aircraft follow on this transect flight from Bermuda to LaRC are illustrated in Fig. 3. The Falcon follows a statistical survey flight pattern and performs an extended vertical spiral sounding as it nears LaRC. The King Air flies at 9 km throughout the transect.

c. Collocation algorithm description

The spatiotemporal collocation algorithm presented in this work is a brute-force nearest-neighbor finding procedure that is applied from the perspective of each of ACTIVATE's platforms, rather than specific instruments. Furthermore, this algorithm extends the nearest-neighbor finding procedure to consider discrete time segments as their own subsets for collocation. Previous works have used k–d tree methods for efficiently finding the nearest neighbors between radar and in situ data for its computational efficiency (Finlon et al. 2022; Heymsfield et al. 2020; Duffy et al. 2021; Chase et al. 2018).

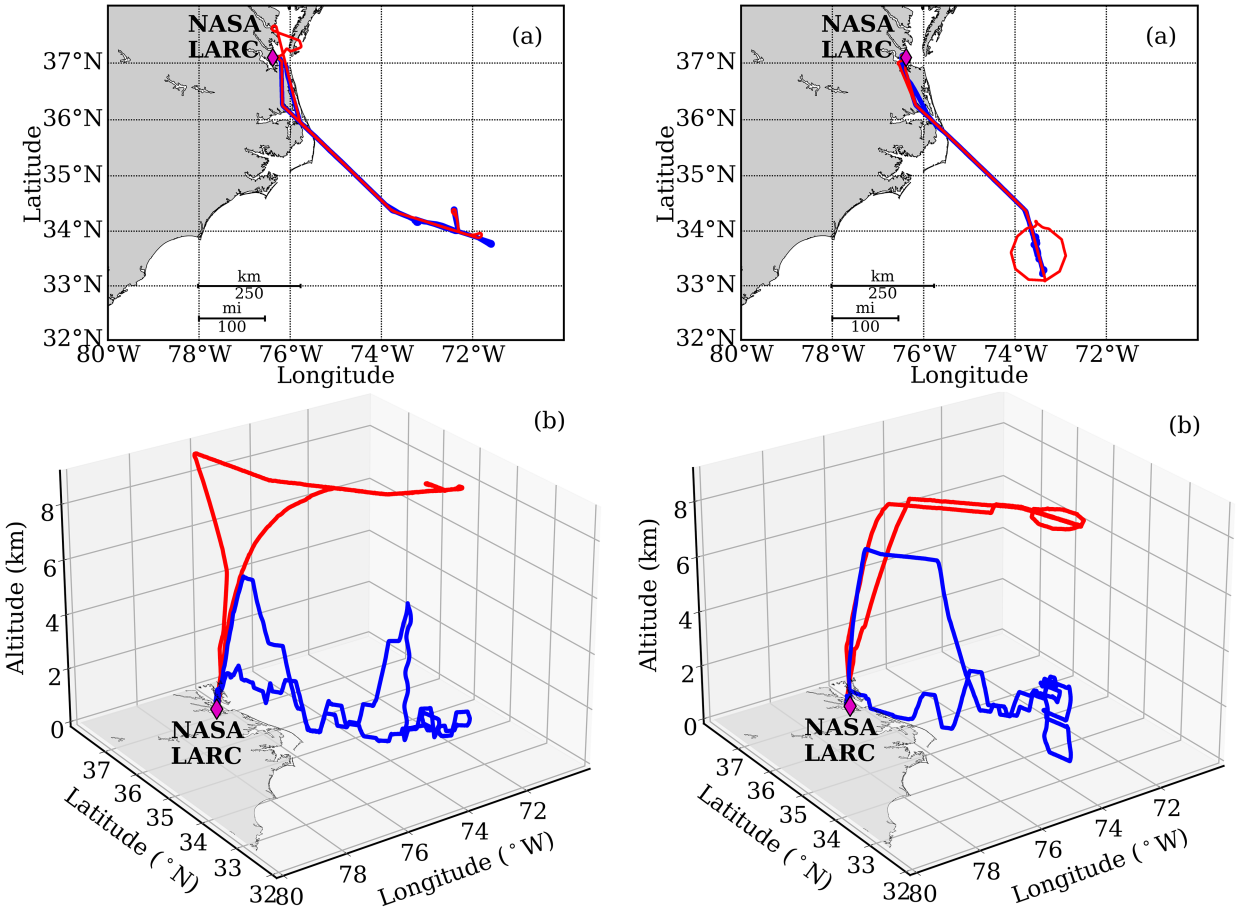

FIG. 1. Example (a) 2D and (b) 3D flight maps from 26 Aug 2020 that demonstrate the statistical survey flight pattern that was used for the majority of Aerosol Cloud Meteorology Interactions Over the western Atlantic Experiment (ACTIVATE) flights. The red and blue lines represent King Air and Falcon flight tracks, respectively.

FIG. 2. Example (a) 2D and (b) 3D flight maps from 28 Feb 2020 that demonstrate an example ACTIVATE process study flight pattern. The red and blue lines represent King Air and Falcon flight tracks, respectively.

The computational burden of the brute-force method is mitigated because of the single-pixel-swath (along-track) measurements of the HSRL-2, which result in 2D data in terms of its cross section through the atmosphere. The in situ data are point-like in that they are 1D (at the location of the aircraft). The collocation method and output files presented here have the additional benefit that measurement-specific collocation can be applied. An example of measurement-specific collocation would be in the event of 3D measurements from one platform, which would require one extra collocation step to match the closest point from the multipixel swath that occurs when the two aircraft are at their closest. In this example the extra collocation step can also be optimized by replacing the brute-force method with balltree nearest-neighbor finding methods that are as computationally efficient but more accurate than *k–d* tree methods because of its use of the haversine function to compute distances, for remote sensing instruments that have wider swath widths (Omohundro 1989).

We define two aircraft navigational points to have close spatiotemporal collocation if they are within 15 km and 30 min, which allows us to maximize the number of potential atmospheric processes that might be captured when using the collocation mask for analyzing ACTIVATE data. But the procedures that are provided in this work allow researchers to easily define their own spatiotemporal criteria depending on the specific datasets they are collocating. These criteria can be impacted by a number of variables, which include the number of platforms that are coordinating their efforts, platform speed(s), instrument resolution, and response time, and the parameter variability at the desired representative scale.

The spatial and temporal threshold criteria used here are generally suitable for ACTIVATE's goals to better quantify particle–droplet relationships and the lidar and polarimetric remote sensing retrieval capability of aerosol and cloud optical and microphysical properties. Additional spatiotemporal filtering can easily be applied to find points closer in space or time, and thereby reduce the number of valid points, as desired. The final criteria used will depend on the specific investigation that is being performed and the spatiotemporal variability of the phenomena being investigated.

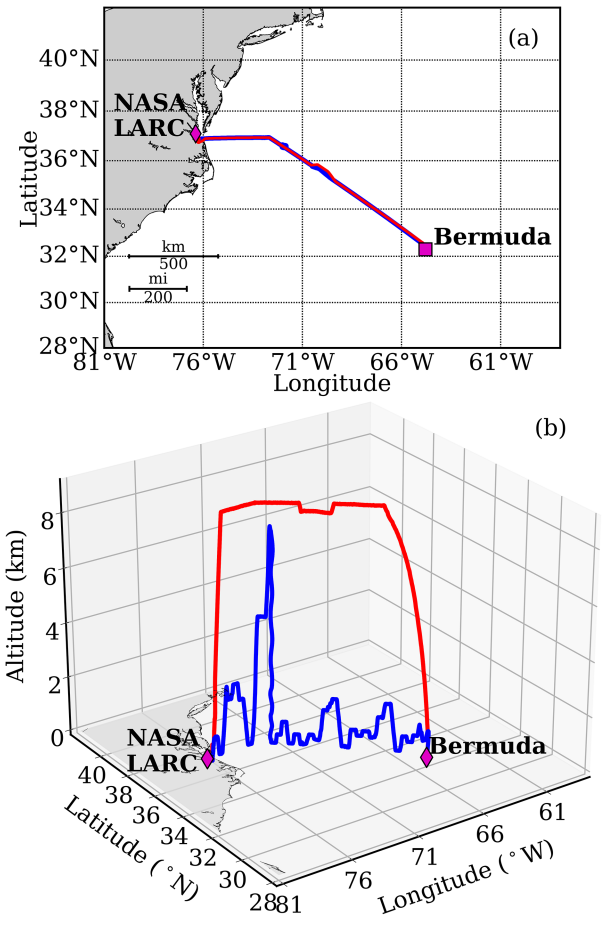

FIG. 3. Example (a) 2D and (b) 3D flight maps from 18 Jun 2022 that demonstrate a Bermuda transect flight from ACTIVATE. The red and blue lines represent King Air and Falcon flight tracks, respectively.

Collocation is determined by comparing each of the time stamps from the primary platform to that of the secondary platform. Any secondary platform time stamps that have a difference of less than 30 min are then checked for the horizontal separation of the platforms, which is limited to 15 km for this work. The following equation defines the spatiotemporal criteria used for this work:

$$\text{spatio-temporal criteria} = \begin{cases} \Delta t < 30 \text{ min} \\ \text{and} \\ \Delta x < 15 \text{ km} \end{cases}, \tag{1}$$

where Δt and Δx are the temporal separation and the horizontal separation, respectively. The temporal separation is then calculated at each navigational point of the primary platform. This calculation is performed by taking the difference between the primary platform's time stamp and the secondary platform's set of time stamps:

$$\Delta \mathbf{t} = \left\{ \bigcup_{n=0}^{\text{L1}} \mathbf{t}_{\text{primary},n} - \mathbf{t}_{\text{secondary}}, \right. \tag{2}$$

where n corresponds to the index of the primary time stamp, L1 is the length of the primary platform's dataset, and $\mathbf{t}_{\text{primary}}$ and $\mathbf{t}_{\text{secondary}}$ are the sets of primary and secondary time stamps, respectively. Once the $\Delta \mathbf{t}$ set is calculated for a given primary platform time stamp, the set of $\Delta \mathbf{x}$ is then calculated. As discussed above, this algorithm uses the haversine formula for calculating distance between the platforms (Van Brummelen 2013; de Mendoza y Ríos 1795). The haversine formula is expected to have an accuracy of between 0.3% and 0.55% in the distance calculations, a systematic error related to the eccentricity of Earth. The following three equations apply the haversine formula for calculating the Δx of the two aircraft at a given primary platform time stamp:

$$\mathbf{a} = \left\{ \bigcup_{n=0}^{L1} \sin^2\left(\frac{\phi_{\text{primary},n} - \phi_{\text{secondary}}}{2}\right) + \cos\phi_{\text{primary},n} \cos\phi_{\text{secondary}} \sin^2\left(\frac{\boldsymbol{\lambda}_{\text{primary},n} - \boldsymbol{\lambda}_{\text{secondary}}}{2}\right), \right. \tag{3}$$

$$\mathbf{c} = 2 \arctan\left(\frac{\sqrt{\mathbf{a}}}{\sqrt{1 - \mathbf{a}}}\right), \tag{4}$$

$$\Delta \mathbf{x} = R_E \cdot \mathbf{c}, \tag{5}$$

where ϕ and λ are latitude and longitude in radians, respectively, and R_E is Earth's mean radius (6371 km). After calculating the sets of $\Delta \mathbf{t}$ and $\Delta \mathbf{x}$, the spatiotemporal criteria from Eq. (1) are applied to create $\mathbf{s}$, which is a vector of ones and zeros with length equal to the number of data points in the primary platform's dataset. All points in the set $\mathbf{s}$ with corresponding spatiotemporal separation that meet the criteria are given a value of 1 and all points in the set that do not meet the criteria are given a value of 0. To determine the number of valid time segments the set of indices is checked for discontinuities. Discontinuities are found by first calculating the differential of $\mathbf{s}$ using the following equation:

$$\Delta \mathbf{s} = \left\{ 0, \bigcup_{n=1}^{\text{L2}-1} \mathbf{s}_{n+1} - \mathbf{s}_n, \right. \tag{6}$$

where L2 is the length of the secondary platform's dataset. Any points where the differential is +1 (−1) mark the beginning (end) of a time segment. Starting with the closest in time, each segment is checked and the nearest (spatial) time stamp is stored as an output along with the corresponding secondary platform's time stamp and horizontal separation Δx in kilometers between the two aircraft.

If more than one time segment is found, then the additional time segments represent situations where the two aircraft started less than 15 km apart, became separated by more than that, and then achieved less than 15-km separation again, all within 30 min. This discontinuity occurs at aircraft turning points or during special flight patterns such as vertical spiral soundings. The algorithm returns one value (or set of values) per segment within the 15-km and 30-min window for each

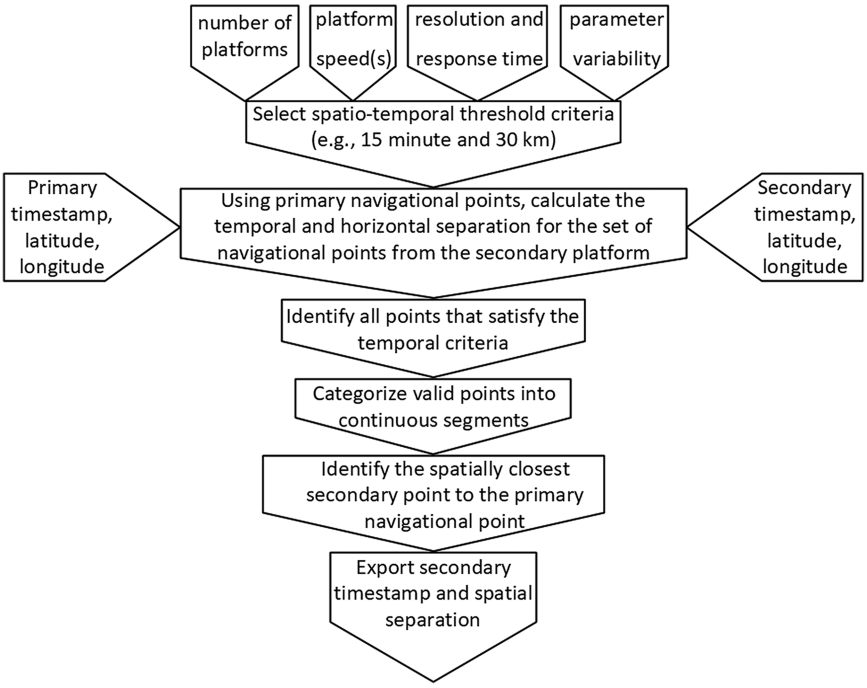

FIG. 4. Flowchart of the collocation procedure.

primary platform's navigational point, up to a maximum of 10 segments per navigational point. The collocation process is performed twice: once for the King Air as the primary platform, and once for the Falcon as the primary platform. Thus, two data collocation mask files are produced for each research flight to allow maximum flexibility in analyzing remote sensing or in situ data for different research objectives. A flowchart illustrating this procedure is provided in Fig. 4. The method discussed here implicitly assumes the measurement times for the in situ and remote sensors are synchronized with the navigational time of their respective platforms but the measurements are not required to have the same sampling resolution as the

TABLE 2. Example of data collocation mask file structure using navigational points. The time stamp and aircraft separation for each segment are in units of seconds after midnight (UTC) and meters, respectively. In this example the Falcon is the primary aircraft, and the King Air is the secondary aircraft.

Column No.	Column definition	Example values		
1	Time_Start	59 824	59 825	59 826
2	KingAir_Time_Start_Segment_1	61 302	61 303	61 304
3	KingAir_Time_Start_Segment_2	58 298	58 298	58 298
4	KingAir_Time_Start_Segment_3	58 105	58 104	58 104
5	KingAir_Time_Start_Segment_4	−999 999	−999 999	−999 999
6	KingAir_Time_Start_Segment_5	−999 999	−999 999	−999 999
7	KingAir_Time_Start_Segment_6	−999 999	−999 999	−999 999
8	KingAir_Time_Start_Segment_7	−999 999	−999 999	−999 999
9	KingAir_Time_Start_Segment_8	−999 999	−999 999	−999 999
10	KingAir_Time_Start_Segment_9	−999 999	−999 999	−999 999
11	KingAir_Time_Start_Segment_10	−999 999	−999 999	−999 999
12	Aircraft_Separation_Segment_1	190	190	200
13	Aircraft_Separation_Segment_2	13 610	13 560	13 510
14	Aircraft_Separation_Segment_3	1060	1180	1300
15	Aircraft_Separation_Segment_4	−999 999	−999 999	−999 999
16	Aircraft_Separation_Segment_5	−999 999	−999 999	−999 999
17	Aircraft_Separation_Segment_6	−999 999	−999 999	−999 999
18	Aircraft_Separation_Segment_7	−999 999	−999 999	−999 999
19	Aircraft_Separation_Segment_8	−999 999	−999 999	−999 999
20	Aircraft_Separation_Segment_9	−999 999	−999 999	−999 999
21	Aircraft_Separation_Segment_10	−999 999	−999 999	−999 999

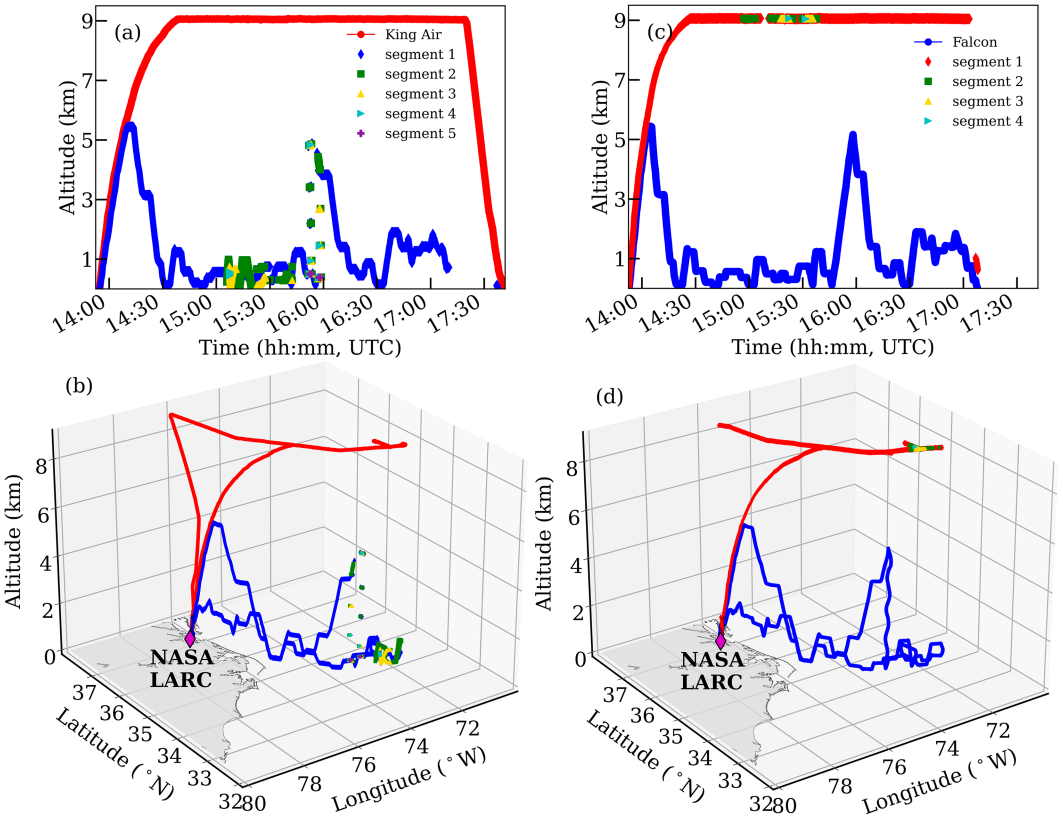

FIG. 5. Time series and 3D flight maps of the altitude of the primary aircraft with the altitude of the secondary aircraft at each of the collocated segments from the ACTIVATE research flight that occurred on 26 Aug 2020. The primary aircraft are (a),(b) the King Air and (c),(d) the Falcon. The collocated segments from the secondary aircraft are sorted by the time difference Δt, with segment 1 closest in time and each segment afterward having a higher Δt.

navigational time. The in situ data were combined using the NASA online merging tool (https://www-air.larc.nasa.gov/missions/etc/onlinemergedoc.pdf).

d. *Aerosol number concentration validation process*

As noted above, this work aims to use the output of the collocation algorithm to explore how well the ambient N_a derived from the polarimeter + lidar method (i.e., $N_{\mathrm{HSRL+RSP}}$) agrees with the N_a derived from the in situ measurements (i.e., N_{LAS}). To accomplish this validation, the King Air (primary aircraft) data collocation mask output is applied to 1-Hz N_{LAS} and further constrained to a spatiotemporal threshold of 6 min and 15 km. The masked N_{LAS} data are then filtered for all ambiguous and cloudy data points. It is also necessary to align the temporal resolutions of the King Air data to use the RSP and HSRL-2 data from the King Air in conjunction with the data collocation mask.

The RSP and HSRL-2 are aligned to the same spatial resolution using the steps outlined in Schlosser et al. (2022). The data collocation mask is then aligned with the RSP aerosol products by using the RSP scan duration and the scan window, which are 60/72 s per scan and 10 scans per sample, respectively. All of the primary platform time stamps within the scan time of each RSP time stamp are gathered with the secondary platform's segmented time stamps. The associated N_{LAS}, LWC, and N_{CDP} are also gathered. For each time segment, if any ambiguous or cloudy data exist in the time window, the data from that window and segment are removed from the set. If there are only cloud-free data present the average of the N_{LAS} data are taken. Finally, each of the time window averaged N_{LAS} points is compared with $N_{\mathrm{HSRL+RSP}}$ by using the $N_{\mathrm{HSRL+RSP}}$ bin that contains the secondary platform's altitude at a given time stamp. Cases with large concentrations of coarse-mode particles are avoided by removing points where N_{CDP} is greater than 0.2 cm^{-3} (Gonzalez et al. 2022).

3. Results

The collocation mask files are formatted in a standard International Consortium for Atmospheric Research on Transport and Transformation (ICARTT) format and stored on the ACTIVATE data repository at https://doi.org/10.5067/SUBORBITAL/ACTIVATE/DATA001 (Northup et al. 2017). Within the contents of each file are the primary platform's 1-Hz time series, and the collocated secondary platform times including the corresponding horizontal distance (in meters) between each aircraft at each collocated point. Table 2 illustrates the organization of the ICARTT data collocation mask files that are

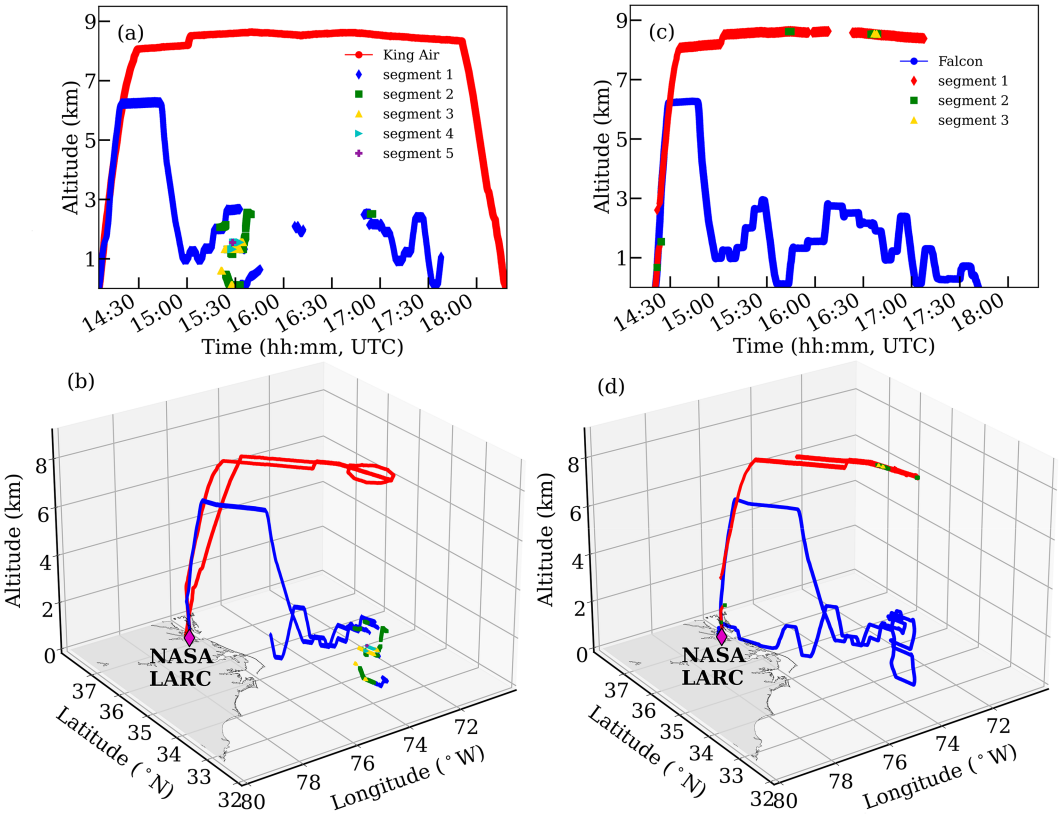

FIG. 6. Time series and 3D flight maps of the altitude of the primary aircraft with the altitude of the secondary aircraft at each of the collocated segments from the ACTIVATE research flight that occurred on 28 Feb 2020. The primary aircraft are the (a),(b) King Air and (c),(d) the Falcon. The collocated segments from the secondary aircraft are sorted by the time difference Δt, with segment 1 closest in time and each segment afterward having a higher Δt.

generated using the method outlined above. Using propagation of uncertainty, the resulting spatial separation provided in the data collocation mask has an uncertainty of 3 m.

Sample outputs of the collocation mask are demonstrated in Figs. 5–7. These figures illustrate how multiple time stamps from the secondary platform can be associated with the primary platform. As discussed previously, this situation occurs where and when there are discontinuities in the spatial collocation of the two aircraft within each 30-min window. In Fig. 5, the discontinuity occurs at both the turnaround point and at the spiral maneuver portions of the statistical survey. In Fig. 6, the discontinuity occurs after the Falcon executes the wall pattern portion of the process study. In Fig. 7, the discontinuity occurs as the Falcon executes a spiral maneuver toward the end of the transit flight.

For ACTIVATE, regardless of which aircraft is chosen as the primary one, a majority of the collocated points are associated with only a single time segment found within the spatial threshold. Flights that contain a large number of segments are often process study flights that feature the Falcon flying wall patterns. Using the Falcon as the primary platform results in 1 787 223 navigational points that have at least one time segment within the spatiotemporal collocation threshold of 30 min and 15 km. The number of valid time segments corresponds to 95% of all the 1-Hz ACTIVATE data. Of these 1 787 223 valid navigational points, 26.9% have more than one valid time segment. Furthermore, 56.1% of the collocated time segments are within 5 min and 6 km. In total there are 111% more valid data points selected using this method, relative to using standard methods that only allow for one valid nearest neighbor in the selection process.

Using the King Air as the primary platform results in 1 836 032 navigational points (83.8% of all 1-Hz ACTIVATE data) that have at least one time segment within the spatiotemporal collocation threshold of 30 min and 15 km. Of these 1 836 032 navigational points, 15.2% have more than one collocated time segment. Finally, 73.0% of the valid time segments are within 5 min and 6 km. In total there are 21% more valid data points selected using the collocation method described in this paper relative to using standard methods that only allow for one valid nearest neighbor in the selection process. This 21% increase in data volume represents the increased amount of data that are viable for comparison in the $N_{\mathrm{HSRL+RSP}}$ validation, relative to a standard nonsegmented collocation algorithm.

The data collocation method presented here allows one to further filter the data based on the separation distance Δx or time difference Δt. For example, to find navigational

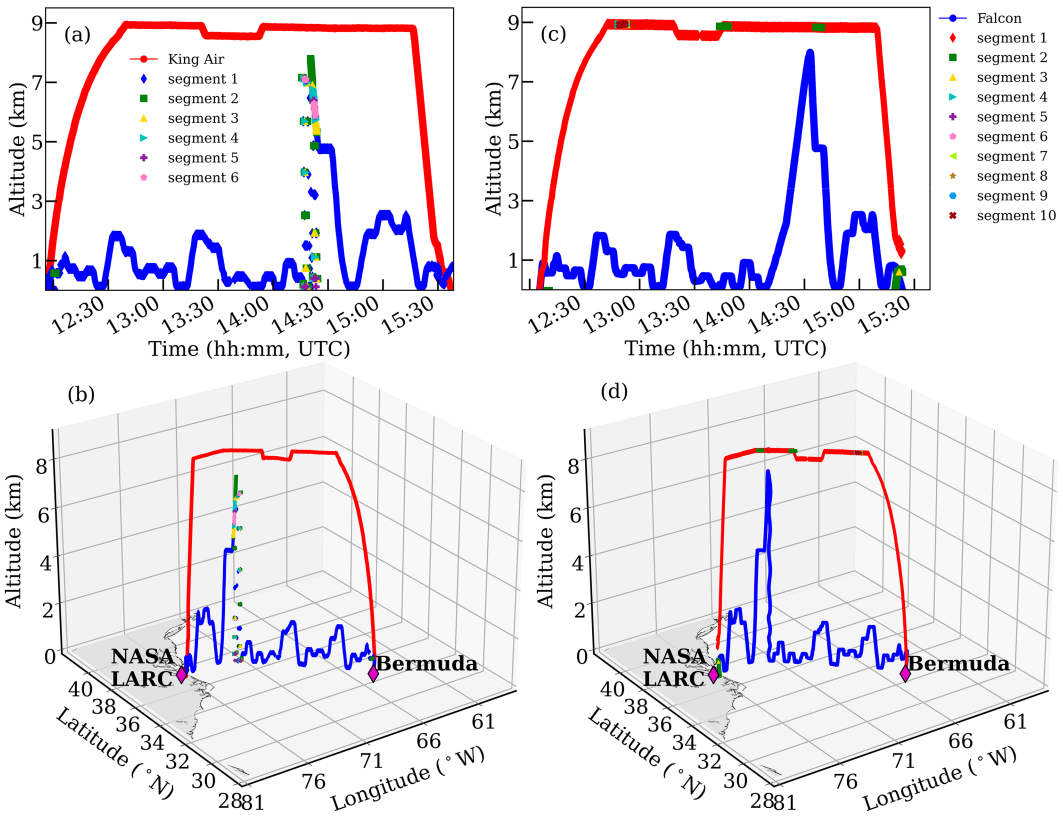

FIG. 7. Time series and 3D flight maps of the altitude of the primary aircraft with the altitude of the secondary aircraft at each of the collocated segments from the ACTIVATE research flight that occurred on 18 Jun 2022. The primary aircraft are (a),(b) the King Air and (c),(d) the Falcon. The collocated segments from the secondary aircraft are sorted by the time difference Δt, with segment 1 closest in time and each segment afterward having a higher Δt.

points that are within 15 km and 6 min, users can adjust the collocation mask columns to further screen the data based on the time separation for the two aircraft and the separation distance. This data collocation mask with the additional 15-km and 6-min spatiotemporal constraint used to validate ambient N_a is illustrated by Figs. 8 and 9.

Figure 8 allows us to explore how well $N_{\text{HSRL+RSP}}$ agrees with N_{LAS} for a research flight where the environmental and operational conditions are optimal, which occurred on 26 August 2020 (Schlosser et al. 2022). Figure 8a shows the time series of both $N_{\text{HSRL+RSP}}$ and N_{LAS} and Fig. 8b shows the one-to-one comparison of $N_{\text{HSRL+RSP}}$ and N_{LAS} resulting from this research flight. For this optimal case, there are a total of 21 valid points that remained after the additional spatiotemporal filtering. Of the 21 navigational points, segments 1, 2, and 3 have 17, 3, and 1 points, respectively.

This work reports both the 90th percentile (P_{90}) of the absolute relative bias as well as the range-normalized mean absolute deviation (NMAD) to quantify the reliability of this new $N_{\text{HSRL+RSP}}$ product. Lower values for the 90th percentile (P_{90}) of the absolute relative bias, and the NMAD, point to better agreement between the $N_{\text{HSRL+RSP}}$ remote sensing product and in situ measurements. Higher values of NMAD are likely due to simplifying assumptions made in the derivation of the $N_{\text{HSRL+RSP}}$, but can also be due to differences between in situ and remote sensing measurement methods, and collocation issues. The NMAD is defined as

$$\text{NMAD} = \frac{100\%}{\max(\mathbf{X}) - \min(\mathbf{X})} \times \frac{\sum_{j=1}^{n_p} |\mathbf{Y}_j - \mathbf{X}_j|}{n_p}, \tag{7}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are the set of in situ–derived N_a and remote sensing–derived N_a, respectively, and n_p is the total number of valid points.

For the 21 valid points $N_{\text{HSRL+RSP}}$ agrees to within a NMAD of 27% and 90% of these points have an absolute relative bias below 131% (e.g., P_{90} = 131%). The NMAD reported here is similar to the NMAD of 29% that was reported for this research flight (Schlosser et al. 2022). The correlation coefficient (r) and p value for these 21 valid points are 0.65 and 0.0015, respectively. The ICARTT files containing time series of $N_{\text{HSRL+RSP}}$ for each ACTIVATE research flight are provided along with the data collocation mask ICARTT files in the ACTIVATE data repository.

In addition to this case study validation, the collocation process allows us to perform a statistical validation for every flight of the six ACTIVATE deployments. There are a total of 6933 valid N_{LAS}–$N_{\text{HSRL+RSP}}$ points. Figure 9 shows a

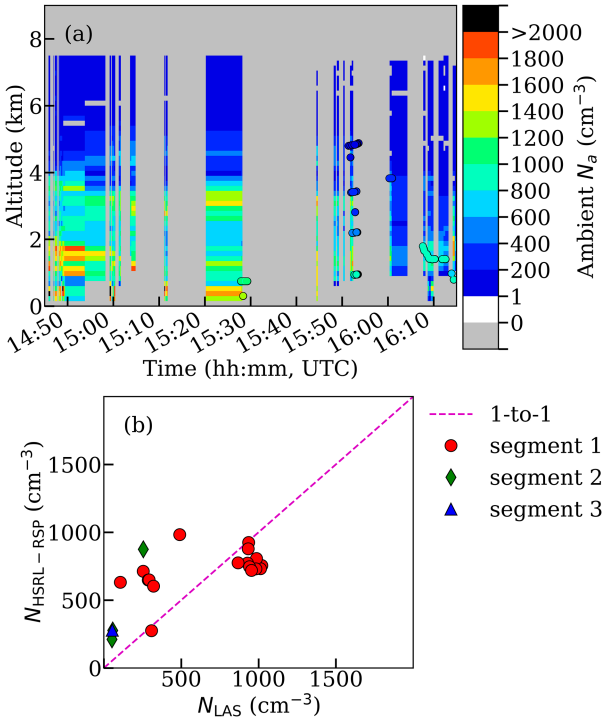

FIG. 8. (a) The Falcon's flight track points colored by N_{LAS} and a surface plot colored by $N_{HSRL+RSP}$ derived from the King Air, and (b) a one-to-one scatterplot of N_{LAS} and $N_{HSRL+RSP}$. These observations are from the ACTIVATE research flight that occurred on 26 Aug 2020.

one-to-one heat map resulting from these valid N_{LAS}–$N_{HSRL+RSP}$ points. The NMAD that results from the comparison of this dataset is 9% and the P_{90} of absolute relative bias is 138%. The r and p value associated with this comparison are 0.46 and $<10^{-4}$, respectively.

4. Conclusions

This work proposes a solution for maximizing the amount of data available from multiple coordinated platforms, which can be used for future and past multiplatform campaigns with coordinated remote sensing and in situ aircraft measurements. Using ACTIVATE data, we show a 21% increase in the volume of collocated data relative to a nonsegmented nearest-neighbor finding method. We also demonstrate the value of the collocation algorithm by performing a quantitative comparison between in situ and point-like remote sensing–derived ambient aerosol particle number concentration (N_a). While this algorithm is independent of the spatial dimensionality of the measurements made with each platform (i.e., measurement agnostic), additional steps would need to be applied after this initial platform collocation for 3D data such as from wide-swath 3D radar systems. Future work involves optimizing the algorithm to use balltree methods and to address nonspherical pixel shapes. The Python and MATLAB procedures that are associated with this work are freely available and open source to enable researchers to apply the collocation

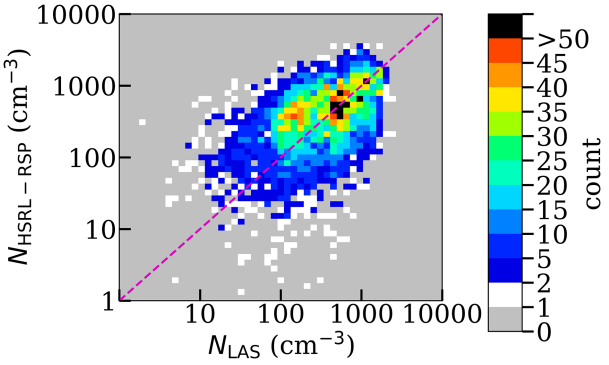

FIG. 9. One-to-one heat map of all 6933 valid N_{LAS}–$N_{HSRL+RSP}$ points that were taken from the six ACTIVATE deployments.

algorithm to other field campaigns for multiplatform comparisons and studies. These procedures can be leveraged for a variety of different mission types that feature independent datasets with at least one remote sensing instrument and one in situ platform. This collocation algorithm can also be applied to a single aircraft system as long as the in situ and remote sensing sampling periods are separated prior to collocation.

Acknowledgments. Funding for this research was provided by the NASA ACTIVATE mission, a NASA Earth Venture Suborbital-3 (EVS-3) investigation funded by NASA's Earth Science Division and managed through the Earth System Science Pathfinder Program Office. A.S. was partially supported by ONR Grant N00014-22-1-2733. J.S.S. was supported by the NASA Postdoctoral Program at NASA Langley Research Center, administered by Oak Ridge Associated Universities under contract with NASA. We wish to thank the pilots and aircraft maintenance personnel of NASA Langley Research Services Directorate for their work in conducting the ACTIVATE flights.

Data availability statement. Publicly available datasets were used for this product. These data can be found at https://doi.org/10.5067/SUBORBITAL/ACTIVATE/DATA001. The data selection source code and a code along with the example application code are available at the following repository: https://doi.org/10.6084/m9.figshare.20489442.v2.

REFERENCES

Behrenfeld, M. J., and Coauthors, 2019: The North Atlantic Aerosol and Marine Ecosystem Study (NAAMES): Science motive and mission overview. *Front. Mar. Sci.*, **6**, 122, https://doi.org/10.3389/fmars.2019.00122.

Berg, L. K., 2016: Two-Column Aerosol Project (TCAP) field campaign report. ARM Tech. Rep. DOE/SC-ARM-16-032, 16 pp., https://doi.org/10.2172/1254831.

Braun, R. A., A. McComiskey, G. Tselioudis, D. Tropf, and A. Sorooshian, 2021: Cloud, aerosol, and radiative properties over the western North Atlantic Ocean. *J. Geophys. Res. Atmos.*, **126**, e2020JD034113, https://doi.org/10.1029/2020JD034113.

Buehler, S. A., M. Kuvatov, V. O. John, U. Leiterer, and H. Dier, 2004: Comparison of microwave satellite humidity data and radiosonde profiles: A case study. *J. Geophys. Res.*, **109**, D13103, https://doi.org/10.1029/2004JD004605.

Burton, S. P., and Coauthors, 2018: Calibration of a high spectral resolution lidar using a Michelson interferometer, with data examples from ORACLES. *Appl. Opt.*, **57**, 6061–6075, https://doi.org/10.1364/AO.57.006061.

Chase, R. J., and Coauthors, 2018: Evaluation of triple-frequency radar retrieval of snowfall properties using coincident airborne in situ observations during OLYMPEX. *Geophys. Res. Lett.*, **45**, 5752–5760, https://doi.org/10.1029/2018GL077997.

Chen, G., and Coauthors, 2011: Observations of Saharan dust microphysical and optical properties from the eastern Atlantic during NAMMA airborne field campaign. *Atmos. Chem. Phys.*, **11**, 723–740, https://doi.org/10.5194/acp-11-723-2011.

Corral, A. F., and Coauthors, 2021: An overview of atmospheric features over the western North Atlantic Ocean and North American East Coast—Part 1: Analysis of aerosols, gases, and wet deposition chemistry. *J. Geophys. Res. Atmos.*, **126**, e2020JD032592, https://doi.org/10.1029/2020JD032592.

——, and Coauthors, 2022: Cold air outbreaks promote new particle formation off the U.S. East Coast. *Geophys. Res. Lett.*, **49**, e2021GL096073, https://doi.org/10.1029/2021GL096073.

Crawford, J. H., and Coauthors, 2021: The Korea–United States Air Quality (KORUS-AQ) field study. *Elementa*, **9**, 00163, https://doi.org/10.1525/elementa.2020.00163.

Dadashazar, H., and Coauthors, 2021a: Aerosol responses to precipitation along North American air trajectories arriving at Bermuda. *Atmos. Chem. Phys.*, **21**, 16 121–16 141, https://doi.org/10.5194/acp-21-16121-2021.

——, and Coauthors, 2021b: Cloud drop number concentrations over the western North Atlantic Ocean: Seasonal cycle, aerosol interrelationships, and other influential factors. *Atmos. Chem. Phys.*, **21**, 10 499–10 526, https://doi.org/10.5194/acp-21-10499-2021.

——, and Coauthors, 2022: Analysis of MONARC and ACTIVATE airborne aerosol data for aerosol-cloud interaction investigations: Efficacy of stairstepping flight legs for airborne in situ sampling. *Atmosphere*, **13**, 1242, https://doi.org/10.3390/atmos13081242.

de Mendoza y Ríos, J., 1795: *Memoria sobre algunos metodos nuevos de calcular la longitud por las distancias lunares y explicaciones prácticas de una teoría para la solución de otros problemas de navegación*. Imprenta Real, 13 pp.

Duffy, G., G. Mcfarquhar, S. W. Nesbitt, and R. Bennartz, 2021: Demonstration of a consistent relationship between dual-frequency reflectivity and the mass-weighted mean diameter in measurements of frozen precipitation from GCPEX, OLYMPEX, and MC3E. *J. Atmos. Sci.*, **78**, 2533–2547, https://doi.org/10.1175/JAS-D-20-0174.1.

Feingold, G., 2003: Modeling of the first indirect effect: Analysis of measurement requirements. *Geophys. Res. Lett.*, **30**, 1997, https://doi.org/10.1029/2003GL017967.

Fernald, F. G., 1984: Analysis of atmospheric lidar observations: Some comments. *Appl. Opt.*, **23**, 652–653, https://doi.org/10.1364/AO.23.000652.

Finlon, J. A., L. A. McMurdie, and R. J. Chase, 2022: Investigation of microphysical properties within regions of enhanced dual-frequency ratio during the IMPACTS field campaign. *J. Atmos. Sci.*, **79**, 2773–2795, https://doi.org/10.1175/JAS-D-21-0311.1.

Friedman, J. H., J. L. Bentley, and R. A. Finkel, 1977: An algorithm for finding best matches in logarithmic expected time. *ACM Trans. Math. Software*, **3**, 209–226, https://doi.org/10.1145/355744.355745.

Gao, M., and Coauthors, 2019: Inversion of multiangular polarimetric measurements over open and coastal ocean waters: A joint retrieval algorithm for aerosol and water-leaving radiance properties. *Atmos. Meas. Tech.*, **12**, 3921–3941, https://doi.org/10.5194/amt-12-3921-2019.

Gonzalez, M. E., and Coauthors, 2022: Relationships between supermicrometer particle concentrations and cloud water sea salt and dust concentrations: Analysis of MONARC and ACTIVATE data. *Environ. Sci. Atmos.*, **2**, 738–752, https://doi.org/10.1039/D2EA00049K.

Hair, J., and Coauthors, 2008: Airborne high spectral resolution lidar for profiling aerosol optical properties. *Appl. Opt.*, **47**, 6734–6752, https://doi.org/10.1364/AO.47.006734.

Heard, D. E., and Coauthors, 2006: The North Atlantic Marine Boundary Layer Experiment (NAMBLEX). Overview of the campaign held at Mace Head, Ireland, in summer 2002. *Atmos. Chem. Phys.*, **6**, 2241–2272, https://doi.org/10.5194/acp-6-2241-2006.

Heymsfield, A. J., C. Schmitt, C.-C.-J. Chen, A. Bansemer, A. Gettelman, P. R. Field, and C. Liu, 2020: Contributions of the liquid and ice phases to global surface precipitation: Observations and global climate modeling. *J. Atmos. Sci.*, **77**, 2629–2648, https://doi.org/10.1175/JAS-D-19-0352.1.

IPCC, 2014: Summary for policymakers. *Climate Change 2014: Impacts, Adaptation, and Vulnerability*, C. Field et al., Eds., Cambridge University Press, 1–32.

Li, X.-Y., and Coauthors, 2022: Large-eddy simulations of marine boundary layer clouds associated with cold-air outbreaks during the ACTIVATE campaign. Part I: Case setup and sensitivities to large-scale forcings. *J. Atmos. Sci.*, **79**, 73–100, https://doi.org/10.1175/JAS-D-21-0123.1.

McMurdie, L. A., and Coauthors, 2022: Chasing snowstorms: The Investigation of Microphysics and Precipitation for Atlantic Coast-Threatening Snowstorms (IMPACTS) campaign. *Bull. Amer. Meteor. Soc.*, **103**, E1243–E1269, https://doi.org/10.1175/BAMS-D-20-0246.1.

McNaughton, C. S., and Coauthors, 2007: Results from the DC-8 Inlet Characterization Experiment (DICE): Airborne versus surface sampling of mineral dust and sea salt aerosols. *Aerosol Sci. Technol.*, **41**, 136–159, https://doi.org/10.1080/02786820601118406.

Müller, D., and Coauthors, 2014: Airborne multiwavelength High Spectral Resolution Lidar (HSRL-2) observations during TCAP 2012: Vertical profiles of optical and microphysical properties of a smoke/urban haze plume over the northeastern coast of the US. *Atmos. Meas. Tech.*, **7**, 3487–3496, https://doi.org/10.5194/amt-7-3487-2014.

Nalli, N. R., and Coauthors, 2018: Validation of atmospheric profile retrievals from the SNPP NOAA-unique combined atmospheric processing system. Part 1: Temperature and moisture. *IEEE Trans. Geosci. Remote Sens.*, **56**, 180–190, https://doi.org/10.1109/TGRS.2017.2744558.

Northup, E., G. Chen, and K. Aikin, 2017: ICARTT file format standards, version 2.0. NASA, https://www.earthdata.nasa.gov/esdis/esco/standards-and-references/icartt-file-format.

Omohundro, S. M., 1989: Five balltree construction algorithms. International Computer Science Institute Tech. Rep. 562, 23 pp., https://omohundro.files.wordpress.com/2009/03/omohundro89_five_balltree_construction_algorithms.pdf.

Painemal, D., and Coauthors, 2021: An overview of atmospheric features over the western North Atlantic Ocean and North American East Coast—Part 2: Circulation, boundary layer, and clouds. *J. Geophys. Res. Atmos.*, **126**, e2020JD033423, https://doi.org/10.1029/2020JD033423.

Pistone, K., and Coauthors, 2019: Intercomparison of biomass burning aerosol optical properties from in situ and remote-sensing instruments in ORACLES-2016. *Atmos. Chem. Phys.*, **19**, 9181–9208, https://doi.org/10.5194/acp-19-9181-2019.

Quinn, P., and Coauthors, 2019: Seasonal variations in western North Atlantic remote marine aerosol properties. *J. Geophys. Res. Atmos.*, **124**, 14 240–14 261, https://doi.org/10.1029/2019JD031740.

Redemann, J., and Coauthors, 2021: An overview of the ORACLES (Observations of Aerosols above Clouds and their Interactions) project: Aerosol–cloud–radiation interactions in the southeast Atlantic basin. *Atmos. Chem. Phys.*, **21**, 1507–1563, https://doi.org/10.5194/acp-21-1507-2021.

Reid, J. S., and Coauthors, 2023: The coupling between tropical meteorology, aerosol lifecycle, convection, and radiation, during the Cloud, Aerosol and Monsoon Processes Philippines Experiment (CAMP2Ex). *Bull. Amer. Meteor. Soc.*, **104**, E1179–E1205, https://doi.org/10.1175/BAMS-D-21-0285.1.

Ryerson, T. B., and Coauthors, 2013: The 2010 California Research at the Nexus of air quality and climate change (CalNex) field study. *J. Geophys. Res. Atmos.*, **118**, 5830–5866, https://doi.org/10.1002/jgrd.50331.

Sawamura, P., and Coauthors, 2017: HSRL-2 aerosol optical measurements and microphysical retrievals vs. airborne in situ measurements during DISCOVER-AQ 2013: An intercomparison study. *Atmos. Chem. Phys.*, **17**, 7229–7243, https://doi.org/10.5194/acp-17-7229-2017.

Schlosser, J. S., and Coauthors, 2022: Polarimeter + lidar–derived aerosol particle number concentration. *Front. Remote Sens.*, **3**, 885332, https://doi.org/10.3389/frsen.2022.885332.

Sinclair, K., B. van Diedenhoven, B. Cairns, M. Alexandrov, R. Moore, E. Crosbie, and L. Ziemba, 2019: Polarimetric retrievals of cloud droplet number concentrations. *Remote Sens. Environ.*, **228**, 227–240, https://doi.org/10.1016/j.rse.2019.04.008.

Sorooshian, A., and Coauthors, 2019: Aerosol–cloud–meteorology interaction airborne field investigations: Using lessons learned from the U.S. West Coast in the design of ACTIVATE off the U.S. East Coast. *Bull. Amer. Meteor. Soc.*, **100**, 1511–1528, https://doi.org/10.1175/BAMS-D-18-0100.1.

——, and Coauthors, 2020: Atmospheric research over the western North Atlantic Ocean region and North American East Coast: A review of past work and challenges ahead. *J. Geophys. Res. Atmos.*, **125**, e2019JD031626, https://doi.org/10.1029/2019JD031626.

——, and Coauthors, 2023: Spatially coordinated airborne data and complementary products for aerosol, gas, cloud, and meteorological studies: The NASA ACTIVATE dataset. *Earth Syst. Sci. Data*, **15**, 3419–3472, https://doi.org/10.5194/essd-15-3419-2023.

Stamnes, S., and Coauthors, 2018: Simultaneous polarimeter retrievals of microphysical aerosol and ocean color parameters from the "MAPP" algorithm with comparison to high-spectral-resolution lidar aerosol and ocean products. *Appl. Opt.*, **57**, 2394–2413, https://doi.org/10.1364/AO.57.002394.

Timmermans, W. J., and Coauthors, 2015: An overview of the Regional Experiments For Land-atmosphere Exchanges 2012 (REFLEX 2012) campaign. *Acta Geophys.*, **63**, 1465–1484, https://doi.org/10.2478/s11600-014-0254-1.

Tornow, F., and Coauthors, 2022: Dilution of boundary layer cloud condensation nucleus concentrations by free tropospheric entrainment during marine cold air outbreaks. *Geophys. Res. Lett.*, **49**, e2022GL098444, https://doi.org/10.1029/2022GL098444.

Van Brummelen, G., 2013: *Heavenly Mathematics: The Forgotten Art of Spherical Trigonometry.* Princeton University Press, 192 pp.

Warneke, C., and Coauthors, 2023: Fire Influence on Regional to Global Environments and Air Quality (FIREX-AQ). *J. Geophys. Res. Atmos.*, **128**, e2022JD037758, https://doi.org/10.1029/2022JD037758.

Zaveri, R. A., and Coauthors, 2012: Overview of the 2010 Carbonaceous Aerosols and Radiative Effects Study (CARES). *Atmos. Chem. Phys.*, **12**, 7647–7687, https://doi.org/10.5194/acp-12-7647-2012.