Improved Planetary Boundary Layer Sounding Using Hyperspectral Microwave and Backscatter Lidar Data Fusion

Antonia Gambacorta, Alexander Kotsakis, Dave Gershman, Narges Shahroudi, Robert Rosenberg, John Blaisdell, Edward Nowottnick, Kenneth Christian, Jordan A. Caraballo-Vega, James MacKinnon, Patrick Stegmann, Stephen D. Nicholls, Joseph Santanello and William G. Blumberg

***Abstract*—This study presents a first-of-its-kind comprehensive data fusion approach combining hyperspectral microwave (HMW) with backscatter lidar (BSL) measurements for improved atmospheric thermodynamic sounding, with particular emphasis on the Earth's Planetary Boundary Layer (PBL). This is a simulation-based trade study to demonstrate the enhancement of HMW over traditional microwave (MW) only measurements and the additional benefits of incorporating BSL with both approaches. This pioneering HMW+BSL fusion methodology represents a major advancement, achieving superior performance compared to traditional thermodynamic remote sensing approaches. Specifically, this configuration demonstrates significant enhancement in PBL temperature bias vertical stability and reduces standard deviation error (SDV) by 30% compared to traditional MW-only performance. Water vapor retrievals show similar improvements, with SDV reductions of 50% in the PBL and bias values consistently maintained below the 10% requirement threshold of the PBL DSI program, compared to PoR errors exceeding 30% bias in challenging cloudy regimes. Case studies across diverse oceanic regions reveal particular advantages of this data fusion approach in complex atmospheric conditions, especially in regions dominated by marine stratocumulus clouds and strong temperature inversions where conventional passive-only retrievals are challenging. Beyond thermodynamic profile improvements, our analysis demonstrates remarkable advances in the detection of PBL height (PBLH), with the HMW+BSL configuration achieving mean absolute errors within the 100 meter requirement threshold of the PBL DSI program, representing a step-change improvement over passive-only approaches. This work directly addresses observational gaps identified in the 2017 Earth Science Decadal Survey, positioning our integrated sensing approach as both a near-term enhancement to existing Earth observation capabilities and a pathfinder for future PBL mission architectures.**

***Index Terms*—Planetary Boundary Layer, Data Fusion, Hyperspectral Microwave, Backscatter Lidar, Artificial Intelligence and Machine Learning.**

I. Introduction

A. Gambacorta, Gershman D., Caraballo-Vega J., Nowottnick E., Santanello J. are with NASA Goddard Space Flight Center, Greenbelt, MD, USA.

A. Kotsakis, Shahroudi N., Christian, K., Stegmann P. are with Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA, NASA Goddard Space Flight Center, Greenbelt, MD, USA

J. Blaisdell and R. Rosenberg are with Science Applications International Corporation, Reston, VA, NASA Goddard Space Flight Center, Greenbelt, MD, USA

Nicholls S. is with Science Systems and Applications, Inc., Greenbelt, MD, USA

MacKinnon J. is with Voyager Technologies, Denver, CO, USA

William G. Blumberg is with Millersville University, Millersville, PA, USA

THE Earth's planetary boundary layer (PBL) serves as a critical component of Earth's atmosphere, functioning as a dynamic interface that regulates heat, moisture, and momentum exchanges between Earth surface and the free troposphere [1], [2]. Despite its fundamental importance, weather and climate models struggle to accurately resolve the complex, nonlinear processes driving PBL dynamics [3], [4]. These limitations significantly impact our predictive capabilities for weather and climate patterns, constraining our ability to protect lives, livelihoods, and socioeconomic stability [5], [6].

Satellite remote sensing from the current Program of Record (PoR) has significantly advanced PBL science but still suffers from challenging constraints. These predominantly include limited information content, inadequate spatial and temporal resolution, and deficiencies in radiative transfer and inversion techniques. Consequently, remote sensing techniques designed to process the PoR often yield incomplete and insufficiently accurate estimates of temperature (T) and water vapor (q) - from now on referred to as *thermodynamic soundings* -within the PBL, leaving key questions in PBL science unsolved. Recognizing these challenges, the 2017 National Academies of Sciences, Engineering and Medicine Earth Science Decadal Survey emphasizes the need for improved observational capabilities and recommends the Earth PBL as an Incubation-Targeted Observable for future exploration [7]. This recommendation underscores the necessity for advancing technology, data processing and sounding techniques to better resolve PBL thermodynamic properties.

In response to this call, a competed NASA PBL Study Team was established in January 2020. The team produced a comprehensive summary report assessing the current state of PBL technology and thermodynamic sounding capabilities within the existing PoR [8], [9]. The report identifies significant science gaps in representing PBL processes, highlighting three science focus areas such as surface and PBL interactions, PBL-cloud interactions, PBL convection and extreme weather. The report acknowledges that while technological advances are crucial, they alone are insufficient, stressing that no single sensor can achieve the accuracy and measurement scale required to fully resolve the PBL. Rather, the report advocates for a dedicated PBL New Observing Strategy, or a 'system of systems' aimed at integrating space-borne, airborne, and ground-based passive and active technologies, in a data fusion approach necessary to achieve a comprehensive observation of

the PBL multi-scale processes. In this strategy, the proposed architecture would harness the strengths of different sensors to overcome their individual limitations. Space-borne instruments could offer broad, global coverage, while airborne and ground-based technologies could provide high-resolution and localized data. Passive sensors could provide high horizontal coverage, while active sensors could augment the vertical resolution domain. Together, these integrated observations would enhance the overall horizontal, vertical, and temporal resolution of PBL measurements, reduce model uncertainties, and improve the overall accuracy of PBL estimations. This synergistic approach is essential to advance our understanding of the dynamics of PBL, thereby enhancing weather prediction and climate projections.

In 2021, NASA Earth Science and Technology Office (ESTO) established the Decadal Survey Incubation (DSI) Program and announced a research call for innovative data fusion approaches to improve PBL thermodynamic sounding capabilities from the existing PoR, identify unresolved gaps and inform the formulation of a future PBL sensor architecture via observation simulation sensitivity experiments (OSSEs). In light of these recommendations, we configured a novel passive-active multi-sensor data fusion architecture, the *P*assive-*A*ctive *N*ear-real *T*ime *AR*chit*E*cture for PBL Sound*I*ng (PANTA REI, Greek for *Everything Flows* [10]) to develop data fusion retrieval systems aimed at demonstrating improved sounding capability of PBL height, temperature and water vapor. The study presented in this paper demonstrates enhanced PBL thermodynamic profiling through a novel data fusion approach, combining hyperspectral microwave (HMW) and backscatter lidar (BSL) observations. The significance of this research aligns with key recommendations from the PBL Study Team Report, which specifically identifies HMW sensors and lidar systems as essential technologies for future PBL observations. Our comprehensive strategy incorporates additional passive (*e.g.*, infrared and visible) and active (*e.g.*, radars) PBL DSI essential components, which will be examined in detail in a subsequent publication.

Our study utilizes simulated observations that reflect the capabilities of both PoR and next-generation instruments. The PoR microwave (MW) component is modeled after the Advanced Technology Microwave Sounder (ATMS, [11]). The HMW component is modeled after the Hyperspectral Microwave Photonic Instrument (HyMPI) design described in [12], [13], [14], and directly supports the upcoming Advanced Ultra-high Resolution Optical and RAdiofrequency (AURORA) Pathfinder, a 2024 NASA ESTO project developing the first photonic integrated circuits based HMW sounder ([15]) as part of the space node of the future PBL sounding architecture. The BSL component of this study is modeled after the Atmospheric Lidar for Clouds and Aerosol Transport (ALICAT) system ([16]). The modularity of this data fusion architecture ensures that this system can be adapted to support any BSL sensor envisioned by the PBL DSI program, such as airborne BSLs (*e.g.*, Differential Absorption Lidars (DIAL, [17]) and the Cloud Physics Lidar (CPL, [18])) as well as space-borne BSL systems from both the PoR (*e.g.*, EarthCare ([19]), IceSat-2 ([20], and CALIPSO ([21]) and the next generation lidar missions (*e.g.*, the Concurrent Artificially-intelligent Spectrometry and Adaptive Lidar System (CASALS, [22]) and the Cloud Aerosol Lidar for Global Scale Observations of the Ocean-Land-Atmosphere System (CALIGOLA, [23])). Conceived as a modular design, we intend to allow for our data fusion architecture to support future and current active/passive inputs, as the instrument simulators allow for studying the impacts to the signal-to-noise ratio (SNR) induced by the design choices of the different sensors and the modularity of the system to enable transfer learning from real data. Previous research has established the individual contributions of these technology prototypes to atmospheric and, more specifically, PBL science ([12], [24] and references therein). The work presented in this paper breaks new ground by investigating their combined potential. This study presents the first comprehensive assessment of an integrated passive-active (and more specifically, HMW-BSL) observing strategy tailored to enhance PBL thermodynamic performance and vertical resolution, meeting the stringent requirements outlined in the PBL Study Team Report (Table I) and actively being developed by the DSI community in advance of the next Decadal Survey. Results from this study will be summarized and translated into a Science Applications Traceability Matrix (Section V). By demonstrating the synergistic benefits of these complementary technologies, our findings contribute valuable insights to address current observational gaps and support the development of a comprehensive PBL sounding mission architecture study.

II. Why is HMW-BSL data fusion key to PBL sounding?

Space-borne passive sensors measure vertically integrated radiation emitted by both the Earth's surface and atmospheric constituents, after transmission through the atmospheric column. The radiative signal originating from the PBL undergoes substantial absorption and scattering processes. In the MW domain, these are mostly caused by hydrometeors present in the overlying atmospheric regions, while gases and aerosols predominantly affect the infrared domain. This radiative transfer process significantly attenuates the distinctive spectral signatures that originate from the PBL, resulting in an attenuation of critical PBL thermodynamic information. As radiation propagates upward toward the satellite sensor, the unique spectral characteristics that would otherwise reveal fine vertical structure within the PBL become progressively obscured, limiting the ability of passive-only systems to resolve sharp vertical gradients or thin atmospheric features. This atmospheric attenuation, combined with sensor technology constraints (*e.g.* finite channel spectral resolution and measurement errors) creates a fundamental challenge in passive remote sensing known as *ill-conditioning*. In mathematical terms, the total number of degrees of freedom available in passive measurements is lower than the number of unknown parameters. For PBL observations specifically, this limitation becomes critical: the available degrees of freedom fall significantly below the number of temperature and water vapor vertical grid points needed to achieve the sub-kilometer vertical

TABLE I: PBL Observational Goals. Table adapted from Table 8.1 of the 2021 PBL Study Team Report, [8]. For clarity, the term accuracy in this table will be replaced from now on in the text by the term "bias", more commonly used in operational frameworks to specify and demonstrate retrieval requirements.

Variable	Horizontal Resolution	Vertical Resolution	Temporal Resolution	Accuracy
Water Vapor Temperature PBLH	0.1 - 100 Km	0.1 - 1 Km 0.1 - 1 Km -	Minutes to Monthly	10% 1K 100 m

resolution and accuracy requirements of the PBL DSI program, as specified in Table 8.1 of the PBL Study Team Report ([8]) and reproduced here in Table I. This limitation remains true also for hyperspectral sensors and poses a fundamental physical constraint that microwave-only or, more generally, passive-only sensing approaches struggle to overcome.

Clouds, or more generally hydrometeors, represent a fundamental source of ill-conditioning in passive sounding of the PBL. This challenge stems from inherent spectral correlation issues that complicate atmospheric state retrieval. Passive sensors inherently measure radiative signals integrated across their finite-resolution channels, simultaneously capturing signal from multiple atmospheric constituents, such as oxygen, water vapor, trace gases, aerosols. When hydrometeors are present within the instrument field of view, they introduce complex radiative processes through absorption, emission, and scattering that vary non-linearly with the hydrometeor microphysical properties and vertical distribution. This complexity creates a mathematically ill-posed inversion problem where multiple distinct atmospheric configurations can produce nearly identical radiance signatures. The resulting ambiguity significantly constrains our ability to isolate the thermodynamic signature of the PBL from the integrated effects of overlying cloud layers. The spectral correlation challenge, coupled with typically insufficient *a priori* hydrometeor information, fundamentally restricts the PBL information content retrievable from passive measurements alone in cloudy scenes.

Space-borne backscatter lidar (BSL) measurements on the contrary can detect the structure of the PBL with high vertical (less than 100 meters) and horizontal (about 300 meters) resolution. While passive sensors can provide large-scale information on atmospheric properties, BSLs can distinguish between cloud heights and types (*e.g.*, water droplets or ice crystals), as well as detect smoke, dust, and pollution. They can also estimate PBL depth (*i.e.*, mixed layer) by analyzing the vertical variance of the backscattered signal from aerosols and shallow clouds ([25]; [26]). This complementary relationship presents a compelling scientific opportunity: integrating space-borne MW and BSL sensing technologies through advanced data fusion approaches. Such integration promises to overcome the individual limitations of each sensing modality, simultaneously enhancing information content, vertical resolution, and spatial coverage of PBL observations to meet comprehensive sounding requirements.

Our data fusion strategy blends the unique strengths of each instrument and infers PBL height (PBLH) and vertical profiles of temperature and water vapor. Space-borne microwave passive sensors measure upwelling thermal radiation, while backscatter lidars emit laser radiation that scatters off atmospheric constituents (such as clouds and aerosols) and measure the intensity of the returned signal. The physical interdependence among temperature, water vapor and clouds (and aerosols) serves as the link between the two measurements. Lidar-inferred cloud top and bottom pressure are based on finding the height where the steepest backscattering gradients occur. As a radiative quantity, cloud top pressure is a function of the vertical distribution of water condensate throughout the atmospheric column as well as the thermodynamic properties of the column. Similarly, lidar-inferred PBLH is based on finding the height where the steepest backscattering gradient occurs by means of clouds and aerosol presence. PBLH is typically aligned with the base of the capping temperature inversion and, during daytime, an increase in potential temperature at the top of the mixed layer, as well as a decrease in water vapor abundance [1]. In this combined strategy, the monochromaticity of the laser in the nanometer spectral domain is the key element. The BSL acts as an independent high vertical resolution source of information on clouds and aerosols, which is *physically but not spectrally* correlated with temperature and water vapor signals. This distinctive capability provides a crucial geophysical constraint on clouds (and aerosols), enabling precise cloud flagging and differentiation between the microwave radiative signatures of clouds and those originating from temperature and water vapor. By accurately isolating and characterizing cloud properties — top and bottom altitude, optical depth, and phase — this data fusion approach effectively addresses a primary source of ill-conditioning in passive-only retrievals. Preliminary results of this data fusion approach were introduced in [27], [28]. This paper presents a comprehensive demonstration and quantification of the improvement in the PBL vertical thermodynamic retrieval structure and accuracy resulting from this data fusion technique.

The four panels in Figure 1 illustrate the interconnections between the key atmospheric variables at play in our HMW+BSL data fusion system. A comprehensive atmospheric cross-section along 20°W longitude, spanning from 5°N to 45°N latitude. The top-left panel displays the vertical temperature profile (K), revealing thermal stratification from the surface to 200 hPa. The bottom-left panel shows water vapor concentration (g/kg), with typically higher moisture content (blue-green) in the boundary layer and drier conditions (brown) at higher altitudes. The top-right panel overlays cloud liquid and ice content (dark and clear blue contour lines, respectively) with dust mixing ratio (orange-red shading, 10^{-4} kg/kg). This panel highlights four distinctive atmospheric scenarios labeled a-d: (a) high-level ice water cloud with a cloud fraction of 0.950, (b) low-level liquid water cloud with a cloud fraction of 0.958, (c) clear sky conditions, and (d) significant dust loading. The black curve in all the panels, delineates the PBLH. The dataset used for the first three panels was derived from the

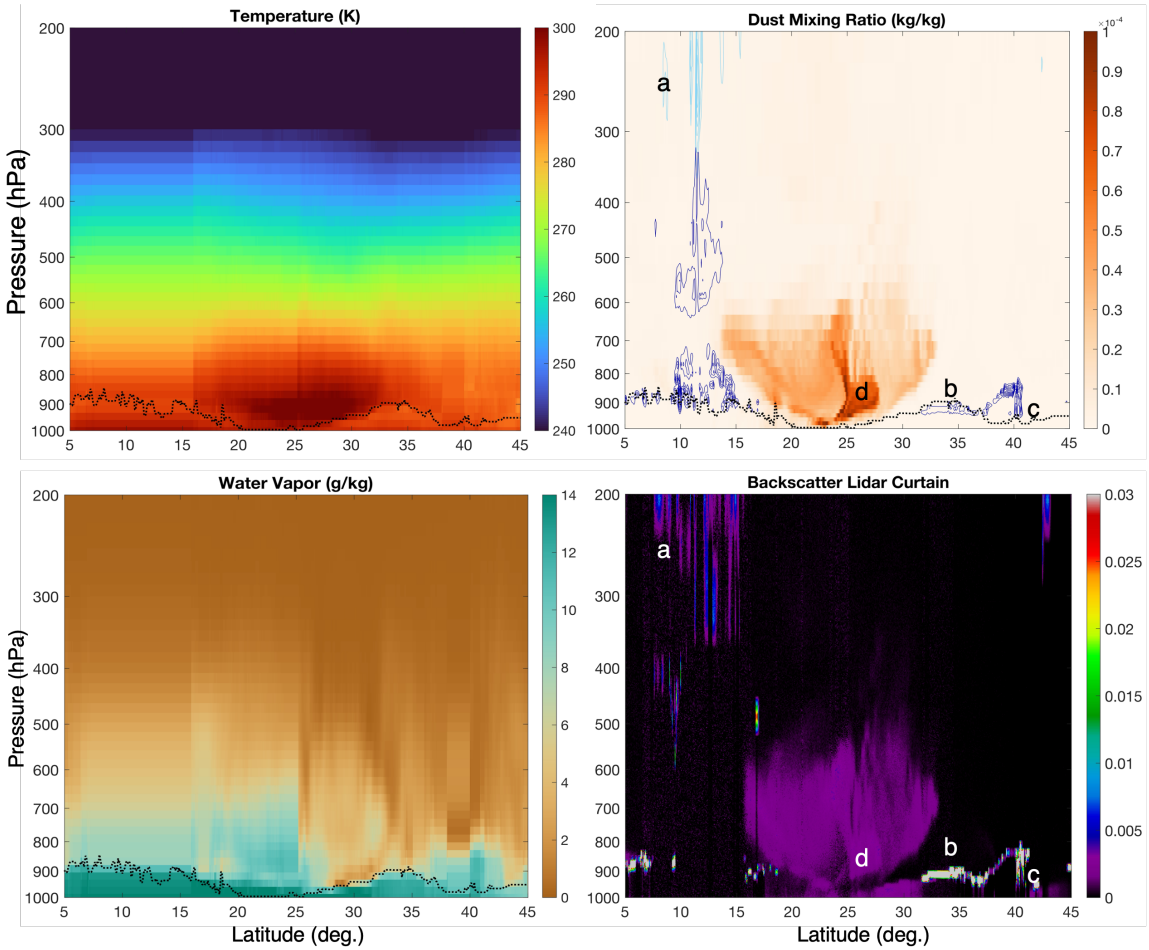


Fig. 1: Atmospheric vertical profile analysis along a fixed longitude transect from a single day, ocean scene, in the NASA GEOS-5 Nature Run (G5NR). **Left top**: Temperature; **Left bottom**: Water Vapor; **Top right**: ice (cyan contours) and liquid water (blue contours) cloud content and aerosol mixing ratio (orange-red shading); **Bottom right**: BSL curtain. **Labels a-d**: (a) high-level ice water cloud with a cloud fraction of 0.950, (b) low-level liquid water cloud with a cloud fraction of 0.958, (c) clear sky conditions, and (d) significant dust loading. The black curve in all the panels, locates the PBL height.

Global Modeling and Assimilation Office (GMAO) GEOS-5 Nature Run (G5NR) [53]. Physical correlations between temperature, water vapor, and cloud/aerosol properties are evident throughout the three panels. The bottom-right panel presents the simulated BSL profiles along the same transect. The intensity scale clearly distinguishes between the four highlighted atmospheric scenarios, showing the distinct signatures of clouds and aerosols and their physical interdependence with temperature and water vapor.

Figure 2 illustrates the simulated brightness temperature spectra across a frequency range of 0-250 GHz as would be measured by a hyperspectral microwave sensor for the four atmospheric scenarios identified in Figure 1. These spectra display notable features including the oxygen absorption lines centered at 60 GHz and 118 GHz, and the water vapor absorption lines centered at 23 GHz and 183 GHz from which atmospheric temperature and water vapor are inferred, respectively. The plot reveals distinct spectral signatures for ice cloud (case a, blue line), water cloud (case b, red line), clear sky (case c, yellow line), and dust (case d, purple line) conditions observed in the interstitial *window* regions between the absorption lines and from which hydrometeor and surface properties (emissivity and temperature) are typically inferred [74]. Compared to clear sky conditions, ice clouds (blue line) show distinctively higher (lower) brightness temperatures in the window region before (past) the left wing of the 183 GHz water vapor absorption line, opposite to the weaker but still noticeable effect of water clouds (red line). The dust scenario (purple line) shows the distinctive Saharian Air Layer (SAL) radiative heating feature associated to dust and concurrent dry anomalies [75], [73], [78], [79]. As clearly demonstrated in the figure, hyperspectral coverage in the MW spectrum of Earth's radiation (as opposed to a few dozen sparse channels in the current MW PoR) fully captures the diverse signatures of clouds and dust and can help distinguish them through their unique spectral responses across the contiguous frequency range. However, while these spectral signatures are present, they remain relatively subtle and challenging to distinguish without prior knowledge of the atmospheric state and surface properties. These simulations were conducted

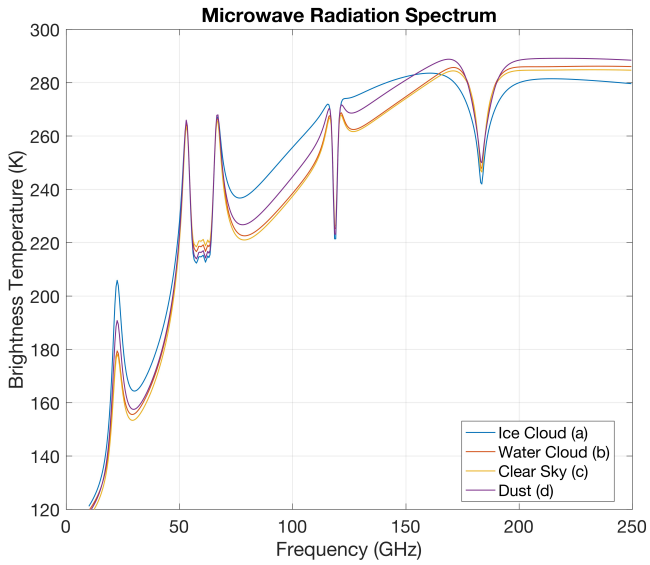


Fig. 2: Simulated hyperspectral microwave brightness temperature spectra corresponding to the atmospheric, ocean scene, truth profiles a), b), c) and d) highlighted in Figure 1.

exclusively over ocean surfaces. Spectral correlation becomes significantly more complex on land, where emissivity features are prominent and highly variable depending on the type of surface. Uncertainties in surface emissivity modeling represent a crucial issue in the land microwave retrieval problem [29], [30], [31]. As the scope of this study focuses on the impacts of data fusion on atmospheric thermodynamic retrieval performance, considerations of land surface emissivity will be deferred to a subsequent investigation (see also Section VI).

III. State of-the-art in Combined Passive + Active techniques

Past research has demonstrated that retrievals using a combination of ground-based lidars and passive (IR, MW) instruments sensitive to the PBL thermodynamics blend the individual strengths of the instruments. These retrievals take on the high vertical resolution properties from lidar, while the passive observations may provide information at places the ground-based lidar does not measure, such as near the surface and off-nadir view angles ([34], [35], [36], [37], [38] and references therein). The work presented here extends these lessons to downward looking space-borne sensors. Methods combining passive and active measurements from space have historically focused on constraining or validating passive cloud product retrievals through comparisons with active observations (see [39], [40], [41], [42], to cite just a few examples). To our best knowledge, no previous study has systematically investigated the passive-active synergistic retrieval of atmospheric temperature and water vapor profiles, particularly within the PBL.

Previous research has identified fundamental discrepancies between these two measurement types that pose significant challenges for data fusion [39], [42]. These discrepancies stem from inherent differences in sensing physics: passive sensors exhibit peak sensitivity within cloud layers due to their integrated radiative transfer properties, while active sensors generate maximum return signals at cloud tops due to backscattering physics. This fundamental mismatch can create a significant inconsistency when attempting to utilize active cloud products as first-guess constraints in passive physical retrieval algorithms. Our data fusion methodology overcomes this complex challenge through an innovative machine learning approach that explicitly addresses these physical discontinuities. Rather than attempting to force compatibility between inherently different measurement sensitivities, this approach specifically learns these discontinuities as part of its training process. This enables the system to identify and leverage meaningful physical correlations between the different measurement types, even in the presence of varying response characteristics. The result is a unified retrieval framework that systematically strengthens the accuracy of the temperature and water vapor profile, particularly in cloudy conditions where traditional retrievals perform poorly. In this vein, our approach establishes a new paradigm for multi-sensor, passive-active atmospheric profiling that maximizes information content while maintaining physical consistency across measurement domains.

The remaining part of this paper is structured as follows. Section IV presents the comprehensive technical framework and methodological innovations that underlie our data fusion architecture. This section details our novel approach to integrating passive and active remote sensing data streams, including the statistical learning techniques that address fundamental sensing discontinuities and the algorithmic structures that enable robust thermodynamic profile retrievals under diverse atmospheric conditions. Experimental results are analyzed in Section V through dual perspectives: comprehensive global and regional statistical evaluations that quantify performance across diverse atmospheric regimes, and detailed case studies specifically selected to demonstrate capabilities directly relevant to NASA's PBL mission science objectives. These include challenging scenarios involving complex cloud structures, strong temperature inversions, and moisture gradients, which are critical to understanding boundary layer dynamics but traditionally difficult to observe accurately from space. These experimental findings are finally synthesized into a Science Application Traceability Matrix (SATM) that explicitly maps our measurement capabilities against the specific science requirements of the PBL DSI program, providing a clear framework for assessing mission readiness and application potential. This structured evaluation approach enables direct comparison with the threshold and baseline requirements established for next-generation PBL observation systems. The paper concludes (Section VI) by contextualizing our results within the broader priorities articulated in the 2017 Decadal Survey, demonstrating how this work directly addresses identified observational gaps in PBL science. This discussion highlights the significance of our integrated sensing approach as both a near-term enhancement to existing Earth observation capabilities and a pathfinder for future PBL mission architectures.

IV. Technical approach and methodology

Our approach leverages the benefits of Artificial Intelligence (AI)/Machine Learning (ML) in the development of a statistical algorithm to obtain temperature, water vapor and PBLH. The growing utilization of AI/ML methods in remote sensing inverse methods rests, among other factors, in the possibility of efficiently processing large volumes of data and modeling the non-linear relationship between sensor observations and atmospheric truth [43]. This offers a complementary approach to physical retrieval methods confronted with the practical need for data thinning and linearization approximations of the radiative transfer equation ([44], [45]).

The AI/ML model in this study is built upon simulated MW and BSL measurements. This approach represents a critical advancement for PBL-focused retrieval algorithms, diverging from most of the legacy passive sounding statistical algorithms used for the PoR (*e.g.*, [46], [47] to cite a few) that rely on real satellite observations often matched to Numerical Weather Prediction (NWP) model profiles to satisfy the need for readily available, global training datasets. Using simulated data offers two significant advantages. First, it enables precise control over the atmospheric training dataset, ensuring comprehensive PBL variability representation and precise co-registration with satellite measurements. Traditional approaches suffer from temporal mismatches (up to 1.5 hours) and spatial discrepancies (up to 0.25 degrees) between satellite data and NWP gridded profiles. These misalignments introduce substantial errors due to the shorter timescales of PBL feedback to forcing ($\sim$1 hour, as explained in [1]). Furthermore, rapidly moving clouds can create miscollocation issues in gridded NWP models that have been observed to produce radiative transfer simulation errors approaching 50 Kelvin [48]. These challenges highlight the necessity of truth-consistent simulated measurements for developing accurate PBL statistical models. Secondly, simulated data enable pre-launch OSSEs to demonstrate improved product performance from novel hyperspectral microwave technology and data fusion solutions while allowing consistent comparisons to the PoR. These studies are instrumental in identifying pathways to address technology and uncertainty gaps in current systems, thereby informing future advancements in both technology and sounding techniques.

The AI/ML data fusion model presented in this study estimates the non-linear relationship, f, expressing the mathematical mapping between the MW brightness temperatures and BSL profiles (the input data) and the atmospheric temperature, water vapor and PBLH (the output data). This relationship is expressed in Equation 1:

$$\begin{aligned}&[T(z), q(z), \mathrm{PBLH}] = \\ &f([\mathrm{BT}_1, \mathrm{BT}_2, ..., \mathrm{BT}_n], [\mathrm{BS}_1, \mathrm{BS}_2, ..., \mathrm{BS}_m])\end{aligned} \tag{1}$$

where $[BT_1, BT_2, ..., BT_n]$ represents the set of microwave brightness temperatures across *n* spectral channels, $[BS_1, BS_2, ..., BS_m]$ denotes the BSL profile discretized into m vertical layers, $T(z)$ is the vertical temperature profile, $q(z)$ is the vertical water vapor profile, $PBLH$ is the planetary boundary layer height and z is the vertical coordinate variable of the training data set. Our analysis will demonstrate enhanced performance in retrieving temperature profiles, water vapor distributions, and PBLH through this data fusion approach. By integrating complementary information from passive microwave and active lidar observations, the model achieves superior vertical profiling capability and accuracy, particularly within the PBL. Performance will be compared to an equivalent model using passive-only observations. Further, both models - to which we will refer to as *HMW + BSL* and *HMW-only*, respectively - will be evaluated against models employing simulated MW measurements from the PoR — specifically ATMS ([11]), referred to as *ATMS-only* and *ATMS + BSL* — to quantify improvements over existing operational capabilities.

The current demonstration is conducted exclusively over oceanic regions, where surface emissivity modeling is relatively well-established and is constrained to nadir-viewing geometries due to the operational limitations of current spaceborne lidar sensors. Parallel development efforts on hyperspectral microwave land surface emissivity and emerging technological advances in off-nadir lidar capabilities will be discussed in Section VI.

A. Sensor measurements

1) HMW Measurements: The HMW component is modeled after the Hyperspectral Microwave Photonic Instrument (HyMPI) design ([13], [14]) and fully described in [12]). For this study, we use a subset of the full microwave thermal domain, focusing on the oxygen and water vapor bands in the 50 - 58 GHz and 183 GHz absorption lines, along with twenty-two channels across all the window regions which carry information on PBL water vapor, surface and hydrometeor radiative signals. Specifically, the oxygen band is sampled between 52.6–57.3 GHz with a spectral resolution of 10 MHz, totaling 470 channels. The water vapor band is sampled between 173.3–193.3 GHz, with a spectral resolution of 40 MHz, totaling 510 channels. This configuration has been studied in [12] and has been shown to contribute up to $\sim$ 50% improvements in temperature and water vapor sounding compared to the performance from the PoR. Spectral resolution characteristics and Instrument Noise Equivalent Delta Temperature (NEDT) values are summarized in Table II derived from [12]. Interchannel correlation has not been explicitly modeled in this study as this feature is learned and encoded in a Fully Connected Neural Network (FCNN) approach developed for this experiment and fully described in Section IV-D. Methods to explicitly model and mitigate HMW spectral correlations in the context of physical retrieval and data assimilation frameworks are underway. Those refer to spectral thinning strategies such as channel selection, spectral grouping, and principal component compression (as discussed in [32] and references therein) and will be the subject of a separate publication.

2) PoR MW Measurements: We will use ATMS as the PoR baseline against which to compare the performance improvement harnessed in hyperspectral microwave measurements.

TABLE II: HMW and ATMS Specifications

Channel Frequency (GHz)	Spectral Resolution (MHz)	Instrument Noise (K)
ATMS		
23.80	270	0.25
31.40	180	0.31
50.3	180	0.37
51.76	400	0.28
52.80	400	0.28
53.596±0.115	170	0.29
54.40	400	0.27
54.94	400	0.27
55.50	330	0.29
57.290344	330	0.43
57.290344±0.217	78	0.56
57.290344±0.3222±0.048	36	0.59
57.290344±0.3222±0.022	16	0.86
57.290344±0.3222±0.010	8	1.23
57.290344±0.3222±0.0045	3	1.95
88.2	2000	0.29
165.5	3000	0.46
183.31±7	2000	0.38
183.31±4.5	2000	0.46
183.31±3	2000	0.54
183.31±1.8	2000	0.59
183.31±1	2000	0.73
HMW		
52.6-57.3	10	2.1
173.3 - 193.3	40	1.0
6.92	350	0.3
10.65	100	0.7
15.37	150	0.5
18.70	200	0.5
21.3	200	0.5
22.35	290	0.4
23.80	400	0.3
31.65	300	0.4
36.50	1000	0.2
40.25	500	0.3
50.30	200	0.5
89.0	3000	0.1
101.0	2000	0.1
110.65	1300	0.2
150.00	3000	0.1
157.00	3000	0.1
165.5	3000	0.1
202.00	3000	0.1
207.00	3000	0.1
229.00	3000	0.1
237.00	3000	0.1
251.00	3000	0.1

ATMS measures the same atmospheric oxygen and water vapor emitted radiation as in our HMW configuration using only eleven and five channels, respectively, along with four additional window channels. Detailed spectral characteristics and NEDT are also outlined in Table II, derived from [11].

3) BSL Measurements: BSL measurements are modeled after ALICAT [16]. ALICAT was designed as a SmallSat elastic backscatter lidar for deployment in an inclined orbit as part of the Atmosphere Observing System (AOS), similar to the Time-varying Optical Measurements of Clouds and Aerosol Transport (TOMCAT) concept [33]. ALICAT was designed to provide backscatter profiles at both 1064 nm and 532 nm wavelengths, with linear depolarization measurements available at each wavelength (see Table III). The instrument would offer spatial resolution of 100 meters horizontally and 30 meters vertically. The 532 nm wavelength is employed in this study. Future work will also analyze the utilization of additional wavelengths.

TABLE III: ALICAT Specifications

Parameter	Value
Laser Type	Nd:YVO_4
Laser Wavelengths	1064 and 532 nm
Depolarization	1064 and 532 nm
Laser Repetition Rate	4 kHz
Laser Pulse Energy	3 mJ (1064 nm); 2 mJ (532 nm)
Laser Pulse Length	$\sim$10 ns
Telescope Diameter	60 cm
Telescope Field of View	115 μrad (1064 nm); 85 μrad (532 nm)
Vertical Resolution	30 m
Horizontal Resolution	70 Hz or 100 m along track

B. Instrument Simulators

1) CRTM Simulator: For MW instrument simulations, we use the Community Radiative Transfer Model (CRTM, [60]). The CRTM is an all-sky radiative transfer model for the efficient computation of satellite instrument observations from atmospheric input profiles, together with their associated Jacobians. This means that the CRTM can also compute the sensitivity of the satellite instrument observations with respect to variations in the atmospheric input state, such as gas concentration and cloud optical variables. The CRTM uses the Optical Depth in Pressure Space (ODPS) parameterization for atmospheric transmittance in the MW band and the Advanced Doubling Adding method for light scattering by aerosol and hydrometeor clouds. The spectra illustrated in Figure 2 are derived from CRTM. The reader is referred to Figure 2, 3 and 4 of [12] for an illustration of CRTM-derived ATMS and HMW temperature and water vapor Jacobians. The CRTM is free open-source software (FOSS) written in objected-oriented Fortran and maintained by the the Joint Center for Satellite Data Assimilation (JCSDA). A detailed description of the process that is necessary to compute CRTM instrument coefficients for the passive instruments used in this study is given in [61], and a description of the aerosol scattering properties used in the CRTM is published in [62], [63]. At the time of writing, the simulator software for passive MW observations relies on CRTM version 2.4.0.

*2) BSL Simulator: T*he Goddard Space Flight Center (GSFC) BSL simulator was developed for the Cloud-Aerosol Transport System (CATS) ([49], [50], [51]) and refined during the development of ALICAT. This tool simulates raw photon counts that are a function of the instrument design (*e.g.* transmitted laser energy, receiver aft optic efficiencies), solar viewing conditions, surface reflectance, and wavelength. This tool simulates raw photon counts that are a function of the instrument design (e.g. transmitted laser energy, receiver optic efficiencies), solar viewing conditions, surface reflectance, and wavelength. For this study, simulations were computed using the 532 nm wavelength. Since the design includes single-photon counting module (SPCM) detectors, the BSL simulator prescribes random uncertainty on the measured photon counts assuming a Poisson distribution. This procedure has been documented for photon-counting detectors in [52]. Systematic uncertainty, such as calibration uncertainty, was prescribed in

the simulator in the conversion from photon counts to total attenuated backscatter, used as the primary observable input from the BSL. Conceived as a modular design, we intend to allow for our algorithm to support other future and current active/passive inputs, as the instrument simulators allow for studying the impacts to the signal-to-noise ratio (SNR) induced by the design choices of the different sensors.

C. Training and Testing Dataset

Our training, validation and testing truth dataset draws from G5NR ([53]), which is a 2-year global, non-hydrostatic mesoscale simulation from June 2005 to May 2007 and publicly accessible by the scientific community. This gridded dataset has a 7 km horizontal resolution and a vertical spacing of 72 hybrid sigma-pressure levels. It includes various meteorological variables (*e.g.*, temperature, surface pressure, wind, cloud liquid and ice water profiles), aerosol tracers (*e.g.*, dust, sea salt, black carbon), and atmospheric constituents. As a free-running model without assimilated data, all of the G5NR geophysical variables do not have any associated uncertainty. Speciated aerosol extinction and backscatter properties are provided using look-up tables of optical properties, as described in [49]. Assumptions of cloud extinction-to-backscatter ratio were made for liquid (20 Sr) and ice (40 Sr) cloud types to yield profiles of backscatter along with extinction. The training dataset includes twelve days, the fifteenth of each month in the G5NR year 2006, sampled globally with a three-hour cadence. The validation dataset is represented by a 20%, globally sampled portion from the G5NR August 15th 2006 day. Testing was performed on a whole separate day, specifically August 3rd, 2006. All figures presented in this study (global, regional and test cases) were obtained from this independent testing dataset.

The 72 hybrid sigma-pressure spacing represents a sophisticated hybrid vertical coordinate system in numerical weather prediction and climate models. This coordinate system combines terrain-following logarytmically based sigma coordinates in the lower atmosphere with pure pressure coordinates aloft, enabling a smooth transition that is particularly well-suited for resolving complex PBL thermodynamics. These 72 layers are strategically concentrated near the surface, providing the vertical detail needed to capture fine-scale temperature and moisture gradients, surface-driven convection, and the evolution of temperature inversions that regulate pollution dispersion and local climate. This high-resolution structure allows for accurate simulation of key PBL processes, including diurnal cycles, land-atmosphere exchanges, and cloud formation. These characteristics form the backbone of our training dataset, ensuring that the algorithm can robustly support the PBL DSI mission objectives: quantifying surface-atmosphere exchange, capturing PBL–cloud feedbacks, and tracking convective and extreme weather evolution.

For this study, we deliberately selected scenes with high SNRs to establish a robust, foundational proof of concept for our data fusion methodology. Specifically, this initial demonstration focuses on daytime conditions and scenarios where the simulated ground return from the lidar exceeds

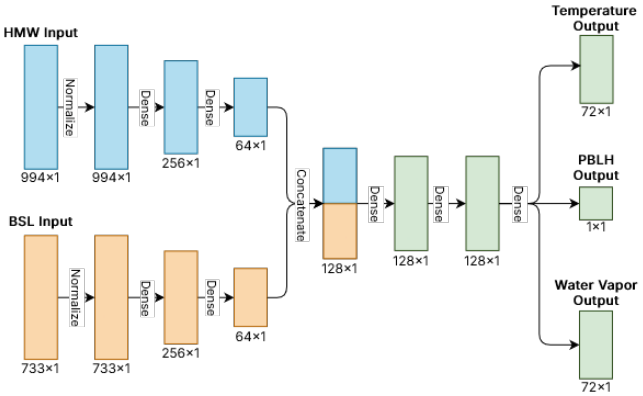


Fig. 3: An outline of the HMW + BSL data-fusion model structure. The diagram displays the paths for each of two distinct inputs (HMW, using a one dimensional vector of 994 channels and BSL, using an input vector of 733 vertical levels). The flow illustrates how the inputs are reduced in dimensionality and combined to form a final prediction of temperature, water vapor, and PBL height. Predicted profiles are displayed on the same vertical grid of the training truth profiles.

$10^{-2}km^{-1}sr^{-1}$, confirming that the lidar signal successfully penetrates to the surface. This strategic approach capitalizes on sharp PBL vertical gradients and ensures sufficiently strong, complete atmospheric BSL returns. The BSL provides a unique advantage for PBL studies compared to active radar sensors, as radars typically exhibit minimal sensitivity below 1 km above ground level due to interference from surface interactions known as *ground clutter effects* [54], [55]. Lidars on the contrary, can effectively observe these low-level features unless fully attenuated by higher cloud layers. By prioritizing these optimal sensing conditions, we maximize the effectiveness of the augmentation of the BSL signal within our MW + BSL data fusion framework, allowing us to clearly evaluate the core methodological advantages before addressing more challenging PBL scenarios (*e.g.*, shallower nighttime PBL) in subsequent investigations.

D. AI/ML Architecture

Expanding on an initial data-fusion development [28], the model developed for this study explores novel large-scale solutions [56] by utilizing a fully connected neural network (FCNN). FCNN's are a special kind of supervised neural network where every output in one layer influences every input in the following one. These layers are considered fully connected or *dense* and, as a result, can work with any kind of input data, regardless of its size or structure. This versatility makes a FCNN a clear choice considering that the inputs come in different sizes based on the number of brightness temperature channels and BSL vertical levels.

The specific structure of our model is illustrated in Figure 3. This makes use of two paths of dense layers, one for each input dataset (HMW and BSL). As a result, this model can leverage the performance improvements that come from combining the different modalities of each dataset. The novelty behind this

model is its ability to find non-linear solutions within a large range of input features, providing an extraordinary flexibility during the learning phase that has yet to be exploited on these datasets. Given the large volume of data, this framework prioritizes model stability. Each path makes use of a normalization layer that constrains the range of each of the input values, thus making the model more resilient to outliers and noise in the data. For the HMW and BSL paths, normalization is followed by a set of dense layers of decreasing size, enabling our model to prioritize the selection of only the most important features from each path. This ultimately improves model stability and convergence potential. After all paths are complete, they are concatenated together into a larger layer that includes all learned features. Then, a set of dense layers leads to the final output layers: a scalar PBL height prediction and two 72-sigma pressure level predictions for temperature and water vapor. The model leverages approximate predictions and grid-based approaches to fine-tune the exact number and size of the dense layers in each path ideal for data fusion. This aspect is instrumental to satellite data missions attempting to fuse large-scale number of features. As such, this work has the potential to expand to multiple sensors in the future (*e.g.*, hyperspectral infrared, visible, radar, etc.). Regardless, the general structure of concatenating these two dataset paths together towards the output retains the same flexibility.

For this experiment, we trained four distinct models to evaluate sensor performance and data fusion capabilities. First, we developed standalone ATMS-only and HMW-only models to assess the enhanced performance expected from a hyperspectral MW sensor compared to the PoR. Subsequently, we created two data fusion models by incorporating BSL augmentation: ATMS + BSL and HMW + BSL, respectively. A comprehensive comparative analysis of all four models is presented in Section V. An important methodological consideration is that all simulations in this study assume identical spatial footprint conditions based on the G5NR 7 km spatial grid resolution. This creates a controlled experimental environment in which no distinctions exist between ATMS, HMW, and BSL surface conditions or vertical atmospheric variability within their respective fields of view, an idealization that differs from real-world observational conditions. In operational settings, significant scale differences exist between these sensors. BSL footprints operate at fine scales (of the order of 100 meters), while passive microwave sensors function at much coarser resolutions. While future HMW space sensors (e.g., AURORA Pathfinder) are expected to have footprints comparable to the G5NR spatial grid at nadir, ATMS footprints range from approximately 16 km to 75 km, depending on frequency [49]. This footprint disparity would likely degrade actual ATMS retrieval performance compared to our simulation results, as increased surface emissivity heterogeneity and atmospheric variability within larger fields of view become more difficult to constrain during retrieval processing. Despite these recognized limitations, our study design specifically isolates and examines the augmented information content and vertical resolution improvements resulting from enhanced spectral coverage and MW-BSL data fusion. To minimize the impact of surface heterogeneity, we focus our analysis on ocean regimes, which provide more homogeneous surface and atmospheric conditions compared to land scenes, allowing our simulations to more reasonably approximate real atmospheric observations. Section VI will address these limitations through our presentation of ongoing activities to improve passive-active co-registration techniques and real data field campaign validation plans.

V. Experimental results

A. Global statistics overview

Figure 4 and Figure 5 present a global temperature and water vapor retrieval statistics derived from a daily ensemble of G5NR temperature and water vapor fields (specifically, August 3, 2006) used as truth and fully independent from the training and validation datasets. This testing dataset meets the same filtering thresholds applied during the training phase (daytime, BSL signal greater than $10^{-2}km^{-1}sr^{-1}$). This consistency is crucial since our primary goal at this stage is proving enhanced PBL structure resulting from the data fusion technique, without introducing additional variables or biases that might arise from differences between training and evaluation. Future generalizations will be discussed in Section VI. Results are shown for four models: ATMS-only (blue), HMW-only (red), ATMS + BSL (green), and HMW + BSL (yellow). The DSI PBL Study Team identifies accuracy as a metric for PBL thermodynamic performance. For clarity, this document replaces the term ”accuracy” with ”statistical bias” or simply ”bias”, which is more commonly used in operational frameworks to specify and demonstrate retrieval requirements. Bias and standard deviation error (SDV) are displayed on the left and right sides, respectively. The top right panels in each figure provide an enlarged view of statistics between the surface and approximately 650 hPa. Key observations can be summarized as follows.

1) Figure 4, Temperature results: The ATMS-only model bias statistics (blue, left panel) exhibit the characteristic tropopause negative to positive oscillation typically found in ATMS operational algorithms ([58] and references therein). This feature is due to the low signal to noise, isothermal nature of the tropopause. A generally positive bias is found in the mid-troposphere (200-600 hPa) and in the PBL. The HMW-only model (red) demonstrates significantly improved bias characteristics compared to ATMS-only, particularly in the upper troposphere/lower stratosphere (UT/LS) region where the HMW-only reduces this to near zero. HMW-only appears to reduce this bias also in the middle troposphere and the PBL, showing a generally improved vertical structure delineation. Both BSL-augmented models (green and yellow) show significant improvements in the PBL (900-1000 hPa), with bias values closer to zero. The HMW+BSL combination (yellow) appears to provide the most balanced bias profile throughout the atmosphere, with values consistently closer to zero, especially in the PBL.

The ATMS-only model shows the highest SDV values overall, particularly pronounced in the PBL (850-950 hPa) where values exceed 1.5K. The inset panel (650-1000 hPa) highlights substantial error reduction in the HMW+BSL model compared

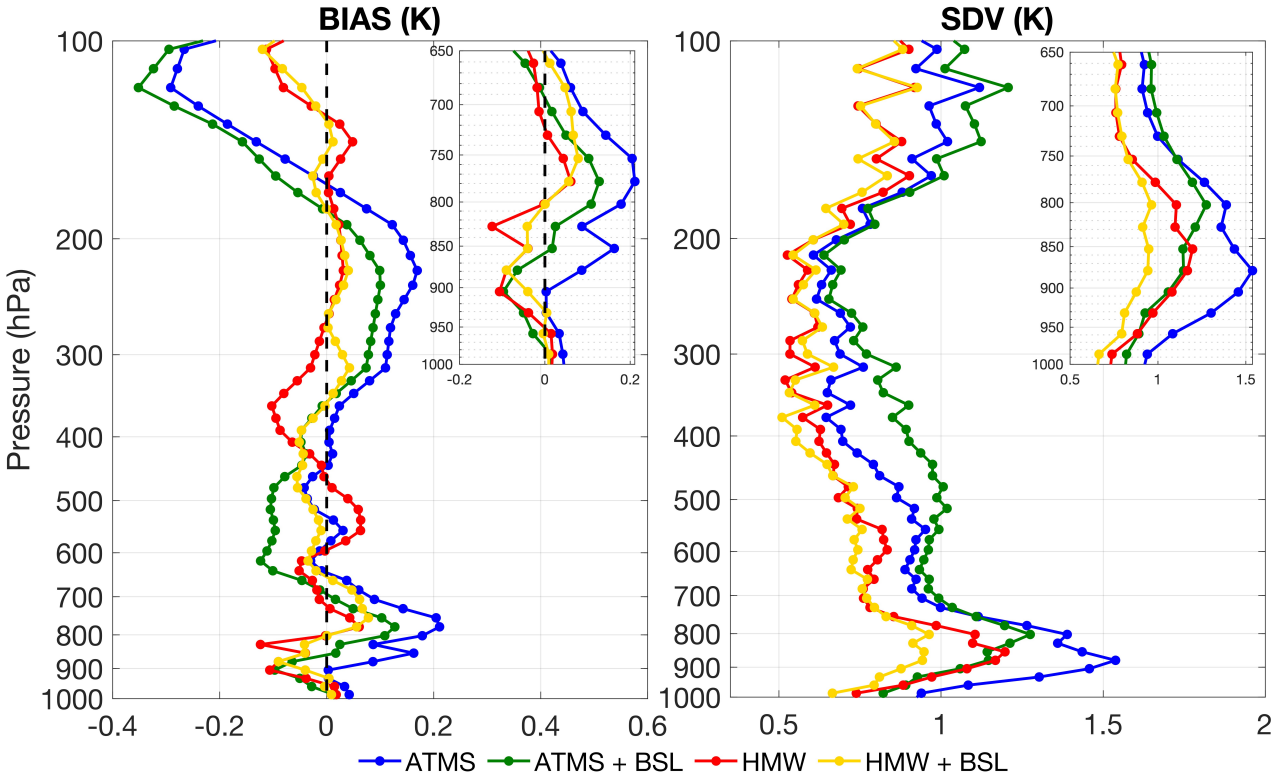


Fig. 4: Temperature bias (left) and standard deviation error (right) statistics for the ATMS-only (blue), ATMS+BSL (green), HMW-only (red), and HMW+BSL (yellow). The top right panels provide an enlarged view of statistics between the surface and approximately 650 hPa.

to others in this region. All models show increased SDV around the PBL-free troposphere transition (800-900 hPa) level, but the magnitude differs significantly. The HMW+BSL fusion model consistently achieves the lowest SDV values throughout most of the profile, demonstrating dramatic improvements in the PBL, with values exceeding 50% compared to the ATMS-only near 900 hPa and an overall improvement of ∼30% throughout the entire PBL. PBL temperature statistics are summarized in Table-IV.

The oscillatory patterns in the bias and SDV profiles suggest varying sensitivity to atmospheric thermal structure across the models. The HMW-based models (red and yellow) appear better at capturing temperature inversions and fine vertical structure compared to ATMS-based models. The upper troposphere (100-300 hPa) shows less impact from the BSL augmentation, as expected given the threshold filter adopted in this study, which focuses primarily on scenarios with a strong signal from the PBL. The passive-active synergistic effect is particularly evident in the yellow profile (HMW+BSL), which shows optimal performance throughout most of the atmospheric column. The improvement from ATMS-only to HMW+BSL represents a substantial advancement in temperature profile retrieval capability.

This figure effectively demonstrates that while hyperspectral microwave sensing (HMW) provides notable improvements over traditional microwave sounding (ATMS), the integration with BSL offers the most comprehensive temperature profiling solution, with particular benefits in the meteorologically critical PBL. This represents a transformative improvement in temperature profiling capability. Such substantial improvements would significantly enhance the accuracy of numerical weather prediction models, particularly for phenomena strongly influenced by PBL thermodynamics, including convective initiation, fog formation, and air quality forecasting.

2) Figure 5, Water vapor: The ATMS-only model (blue) shows the highest positive biases (left panel), particularly in the PBL and around 700 hPa, with values exceeding 30% in some regions, hence well above the 10% threshold of the PBL DSI program (Table I). The HMW-only model (red) already demonstrates significantly reduced bias compared to ATMS-only, confirming the better bias expected from hyperspectral coverage. Both BSL-augmented models (green and yellow) show significant bias reduction compared with their counterparts. This is particularly noticeable around 700-900 hPa, where sharp PBL-free troposphere transitions normally occur, indicating the effectiveness of the data fusion. The HMW+BSL combination (yellow) appears to produce the most consistently low bias across multiple pressure levels, reducing this to near zero in the PBL, now far exceeding the 10% PBL DSI bias requirement. As for temperature, the oscillatory patterns in the water vapor bias profiles suggest the models have different vertical resolution capabilities. The hyperspectral models (red and yellow) appear to capture more vertical structure than the ATMS-only model. The addition of BSL data to both ATMS

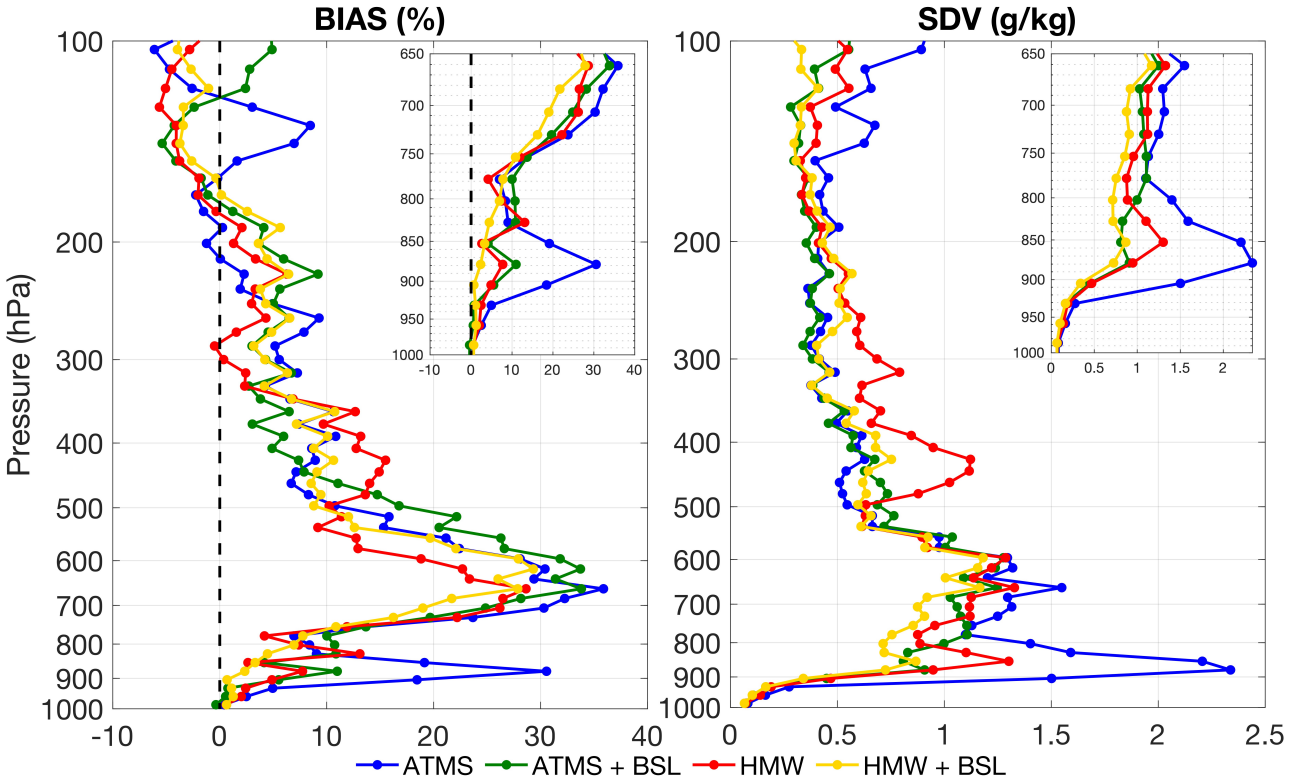


Fig. 5: Water Vapor bias (left) and standard deviation error (right) statistics for the ATMS-only (blue), ATMS+BSL (green), HMW-only (red), and HMW+BSL (yellow). The top right panels provide an enlarged view of statistics between the surface and approximately 650 hPa.

and HMW improves retrieval bias, with particularly significant improvements in the PBL. The HMW+BSL combination appears to provide optimal performance across most of the atmospheric profile.

Similar patterns emerge in the SDV analysis, with the ATMS-only model showing the highest errors overall. The inset panels (650-1000 hPa) reveal important differences in the PBL. Consistent with the bias statistics, all models show increased standard deviation in the PBL-free troposphere transition region (900 hPa level). The fusion models, however, demonstrate consistently superior performance compared to their single-sensor counterparts. The HMW+BSL configuration delivers the best results, with an overall improvement of $\sim$50% throughout the PBL. PBL water vapor statistics are summarized in Table-IV.

These figures effectively demonstrate the progressive improvement in temperature and water vapor retrieval performance when moving from traditional microwave sounding (ATMS-only) to hyperspectral microwave (HMW). Further incorporation of BSL capabilities in the sensor fusion approaches is observed to dramatically improve performance along the entire atmospheric column, particularly in the PBL where DSI requirements are fully met.

B. Regional statistics overview

The model performance is further evaluated by analyzing the statistics across four distinct ocean basins, revealing important advantages of different sensing approaches and providing more insight on the performance observed on the global scale. East Pacific, North Atlantic, South Atlantic, and South Pacific domains were selected to intercompare regions with different surface, aerosol and meteorological variability (Figure 6). The differences observed in the bias between each domain demonstrate the influence of the aforementioned factors.

In Region A (East Pacific), all retrieval methods demonstrate relatively strong performance, particularly in the free troposphere. The ATMS-only configuration (blue) exhibits noticeable degradation in the boundary layer around 850-950 hPa, where temperature inversions commonly occur caused by marine stratocumulus clouds. The progression to HMW-only (red) yields modest improvements throughout the profile, with more pronounced enhancements in the PBL where the additional spectral information helps resolve vertical structure. The integration of BSL capabilities (green and yellow) provides further refinement, particularly below 900 hPa, where the BSL penetrates inversion layers, providing added information content to passive-only approaches.

Region B (North Atlantic) displays similar patterns but with generally better performance across all configurations, likely reflecting less persistent cloud cover and weaker inversions in this region. The incremental improvements from ATMS-only to HMW-only to fusion approaches are consistent but less dramatic than in other regions, suggesting this environment presents fewer fundamental challenges to passive sounding

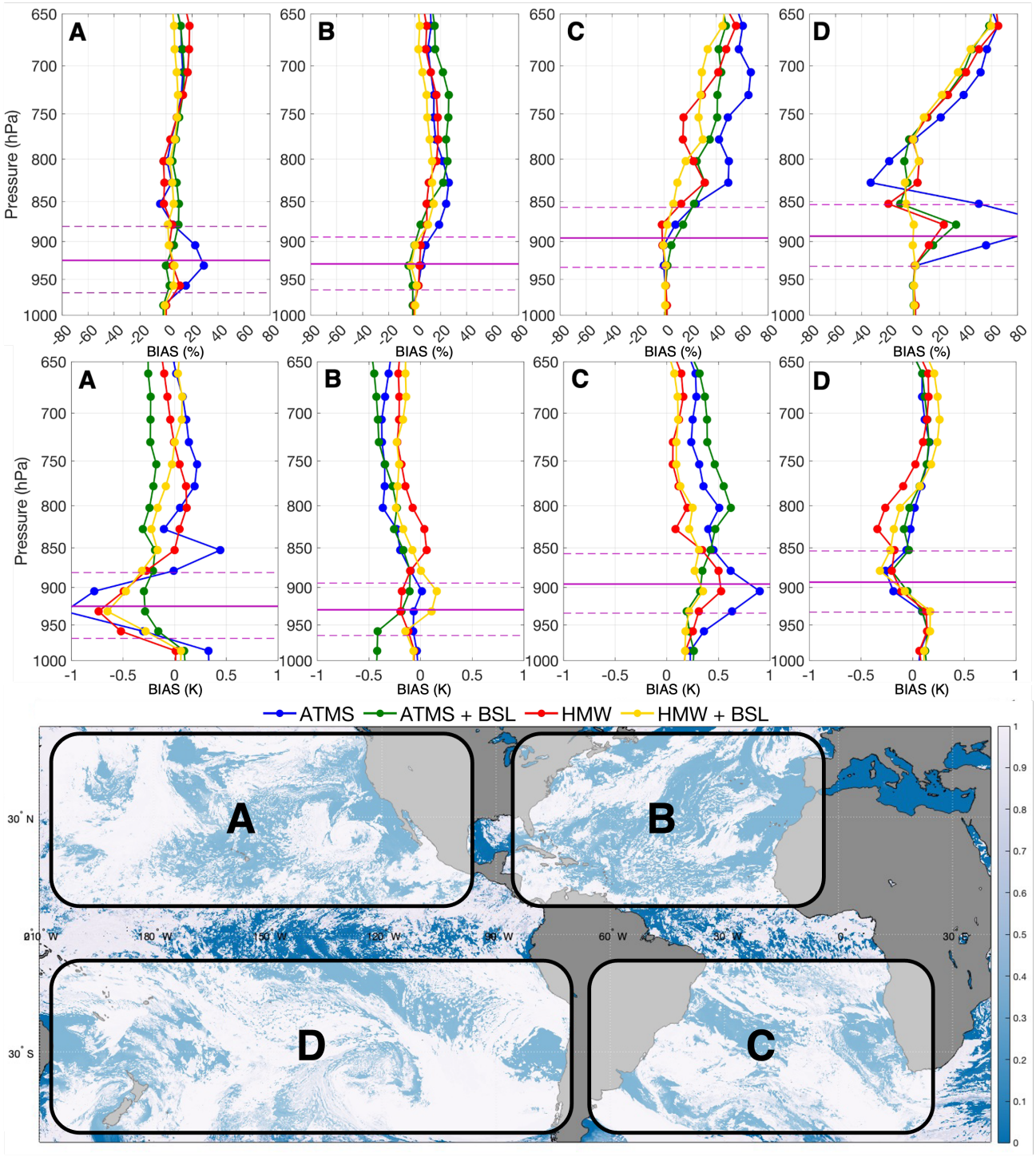


Fig. 6: Bias for the ATMS-only (blue), ATMS+BSL (green), HMW-only (red), and HMW+BSL (yellow) models for water vapor (top row) and temperature (middle row) predicted profiles for the East Pacific (A), North Atlantic (B), South Atlantic (C), and South Pacific (D). The solid purple line indicates the mean G5NR PBLH for each region and the dashed lines indicate $\pm$ one standard deviation.

techniques.

Region C (South Atlantic) presents significantly greater challenges, particularly between 700-850 hPa, where all retrieval approaches show degraded performance. This region is characterized by extensive marine stratocumulus cloud decks that form beneath strong temperature inversions. These cloud layers introduce significant liquid water emission that masks the underlying atmospheric thermal and moisture signals. The hyperspectral capability (red) provides tangible improvements over ATMS-only (blue), as the enhanced spectral resolution helps distinguish between cloud and atmospheric contributions. The addition of BSL (yellow and green) yields substantial gains in the boundary layer, as the active sounding component can penetrate through cloud layers and directly sample the inversion structure that passive systems struggle to resolve.

Region D (South Pacific) exhibits the most severe retrieval difficulties, particularly in the 800-900 hPa layer where persistent, thick stratocumulus clouds dominate. The pronounced bias degradation in the ATMS-only configuration stems from the challenging sharp temperature and humidity gradients associated with the strong inversions at stratocumulus cloud tops. The progression to HMW-only shows notable improvement confirming the enhanced vertical resolution obtained from hyperspectral coverage. The BSL-augmented approaches deliver dramatic enhancements, particularly for the HMW+BSL within the PBL where the bias reaches nearly zero for water vapor and less than one Kelvin for temperature.

Across all regions, the consistent pattern of incremental improvement from ATMS-only to HMW-only to BSL-fusion approaches illustrates the fundamental limitations of passive-only sounding in complex atmospheric environments. Passive microwave observations struggle with vertical resolution in the presence of temperature inversions. The hyperspectral capability however is observed to provide enhanced vertical resolution, allowing better discrimination of atmospheric layers. The addition of BSL measurements provides direct vertical structure information that fundamentally changes the retrieval problem, constraining solutions in ways that passive measurements alone cannot achieve. The HMW+BSL configuration in particular, consistently demonstrates superior performance across all regions and atmospheric levels, highlighting how the synergistic combination of HMW and BSL information creates an unprecedented PBL thermodynamic retrieval capability, particularly valuable in meteorologically complex regions dominated by marine boundary layer clouds and inversions.

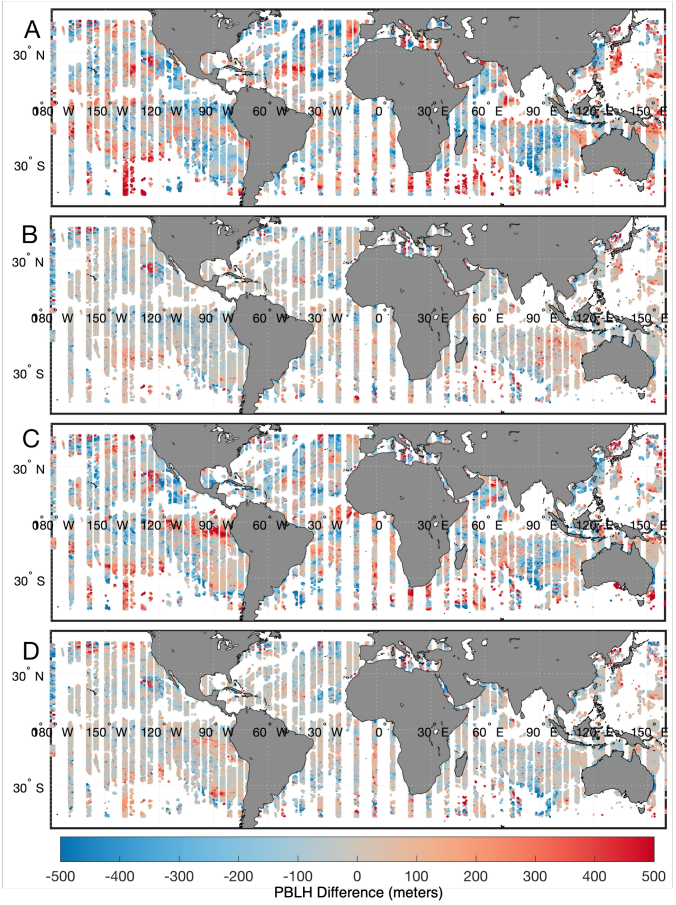


Fig. 7: Global maps illustrating the differences between predicted and reference values for planetary boundary layer height across four model configurations: ATMS-only (A), ATMS+BSL (B), HMW-only (C), and HMW+BSL (D). The longitudinal gaps visible in these maps result from the necessary subsetting of training and testing datasets to optimize computational efficiency. Variance tests (not shown for brevity) confirmed that this subsetting approach caused no significant information loss in the training and analysis.

C. Planetary Boundary Layer Height

The PBLH defines the vertical extent of effective turbulent mixing in the lower atmosphere. Within this layer, vertical diffusivity is substantially enhanced compared to the free troposphere above it. In the G5NR, PBLH is specifically defined as the height of the full level below the edge where the diffusivity coefficient first decreases below 2 m^2/s. Although temperature and humidity profiles obtained through passive remote sensing can indirectly estimate PBLH based on their relationship with diffusivity, these passive soundings lack the necessary vertical resolution to determine PBLH within the strict accuracy requirements (100 meters) established by the PBL DSI program. This limitation is clearly demonstrated in Figure 7 and Figure 8. Here, using the same model architecture and training data, PBLH was predicted globally across a wide range of atmospheric states. Similarly to the temperature and water vapor profile results, ATMS-only, ATMS + BSL, HMW-only, and HMW + BSL model configurations are compared to assess the model predicted PBLH performance on the same independent dataset used in Figure 4 and Figure 5.

Figure 7 demonstrates the consistent improvement in PBLH prediction when transitioning from ATMS (A) to HMW (C) and incorporating BSL into both configurations (B and D, respectively). This enhancement in PBLH prediction remains uniform across all latitudes and longitudes when comparing the lowest vertical resolution model (ATMS) to the highest vertical resolution model (HMW + BSL). The passive-only

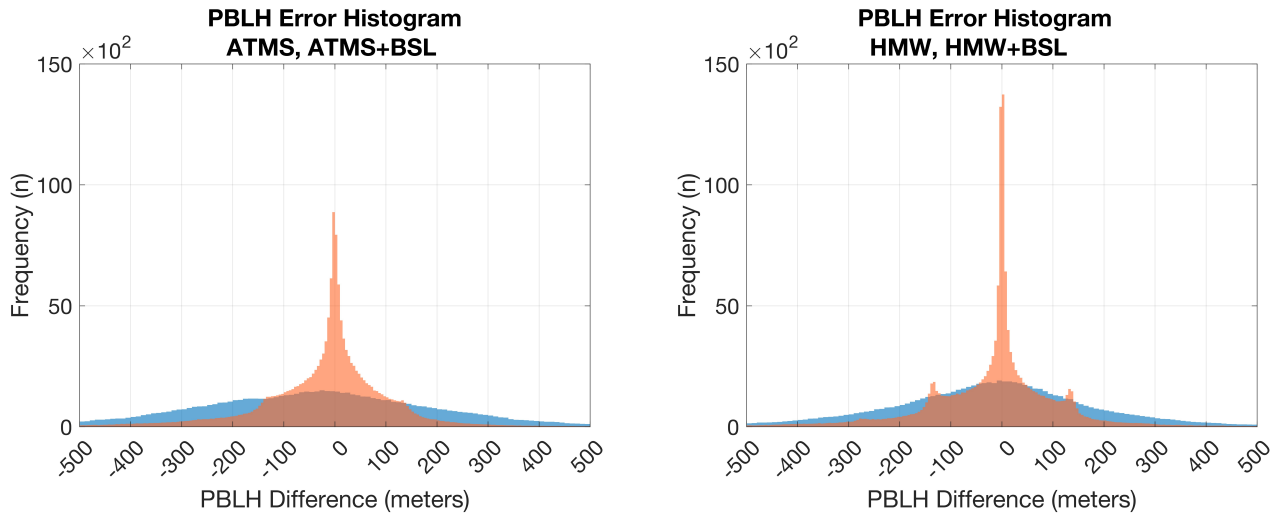


Fig. 8: Probability density functions for the difference between the predicted and truth values for ATMS-only (left,blue), ATMS+BSL (left,orange), HMW-only (right, blue), and HMW+BSL (right,orange) models for planetary boundary layer height.

models (ATMS and HMW) typically underperform in regions traditionally challenging for passive remote sensing. These include high latitudes, characterized by low-contrasted vertical temperature profiles, and areas off the western continental coasts, distinguished by low-level temperature inversions. Additional discrepancies stem from limitations in the G5NR itself. The absence of a shallow convection scheme in this G5NR version causes eddy diffusivity profiles to extend deeper into the atmosphere than they should. This limitation results in exaggerated mixing depths and inflated PBLH values across extensive portions of tropical low-pressure systems [76]. The integration of BSL measurements into the model yields substantial improvements in performance. The superior vertical resolution provided by BSL significantly mitigates these regional challenges, resulting in more accurate PBLH predictions across diverse geographical and meteorological conditions. These improvements are summarized in Figure 8 which provides a probability density function of PBLH differences between prediction models and G5NR truth for ATMS-only (left, blue), ATMS+BSL (left, orange), HMW-only (right, blue), and HMW+BSL (right, orange). In the left histogram (ATMS, ATMS+BSL), the orange distribution representing the ATMS+BSL model exhibits a much taller and narrower peak, reaching a maximum height of approximately 85-90 × 10² counts in zero-centered bin and a median value of -3.14 meters. This contrasts sharply with the blue distribution for ATMS-only, which displays a flatter, broader profile with a maximum height of only about 10-15 × 10² counts in the zero-centered bin and a degraded median value of -35.8 meters. The right histogram (HMW, HMW+BSL) shows even more pronounced differences. The orange HMW+BSL distribution demonstrates an extremely narrow distribution, with a maximum height of approximately 135 × 10² counts and a median of -1.53 meters. The blue HMW-only distribution remains similar to the ATMS-only case, showing a much flatter and broader distribution with a maximum height around 10-15 × 10² counts and a median of -15.5 meters. The quantitative difference is striking, with the maximum of the HMW+BSL distribution one order of magnitude greater than the ATMS+BSL case. The data fusion results represents a significant milestone in PBL height measurement capability that neither passive-only model could achieve on its own. Bias, Mean Absolute Error (MAE) and SDV values also show significant improvements and are summarized in Table-IV. A notable feature in the BSL-augmented distributions is the presence of secondary modes. These secondary modes stem from vertical gridding inconsistencies between the G5NR and the BSL measurements. Lidar PBLH detection fundamentally relies on identifying the top of the aerosol layer, which can differ from the actual turbulent mixing height. This aspect parallels the cloud-related passive-active discrepancies discussed previously. Our machine learning approach mitigates this issue through a data fusion methodology, learning the relationship between lidar backscatter profiles and passive thermodynamic soundings to derive PBLH values. However, since no interpolation is applied between grid points, the resulting PBLH values become discretized according to the respective vertical resolution of both the G5NR and BSL systems. This discretization directly explains the bimodal peaks observed at approximately ±150m in the BSL data fusion results. The G5NR vertical grid spacing is approximately 150m within the PBL, and with a diffusivity coefficient of 2 m^2/s not being entirely negligible, some atmospheric mixing may occur through the layer immediately above. Significantly, this discretization artifact appears more prominently in the HMW+BSL model compared to the ATMS+BSL model. This difference confirms the higher intrinsic vertical resolution capabilities of the HMW instrument compared to ATMS, as the HMW's enhanced resolution makes it more sensitive to these discretization effects when combined with BSL measurements. The need for optimized vertical gridding in PBL sounding has been discussed in [69]. Future developments will utilize an enhanced version of the G5NR that will expand to a 137 vertical layer grid as in [70], providing finer sampling in the

PBL domain (less than 100 meters). Concurrent developments are also underway to consistently incorporate this enhancement into the intrinsic vertical gridding of the CRTM, which is now configured with standard 100-level pressure spacing [71]. Both improvements should help mitigate the issue illustrated in Figure 8 and guide future vertical gridding optimization efforts. Those will be part of a follow-up publication.

D. Individual Test Cases

Global and regional statistics clearly demonstrated the improvement in the retrieval of temperature and water vapor profiles by combining the information content of HMW and BSL. To further demonstrate the capabilities and differences between the models, two case studies were extracted from the G5NR test dataset to demonstrate the value of the HMW + BSL model under different cloud and aerosol conditions that are important to the PBL DSI science objectives. One case was extracted from the North Atlantic (Figure 6, B) and one from the South Atlantic (Figure 6, C) basins.

The first case study comes from a location in the North Atlantic that is within a Saharan Dust plume (Figure 9, right). This location was chosen for two reasons. First, it poses a significant challenge to remote sounding by attempting to retrieve temperature over a location with a boundary layer cloud and high aerosol loading due to the presence of a Saharan Dust plume. Secondly, this geographical region holds exceptional scientific importance for PBL research as it serves as the genesis area for tropical easterly waves emerging from the west coast of Africa. These atmospheric disturbances contribute to a remarkable 60% of all Atlantic tropical cyclones, including 85% of major hurricanes [72]. Demonstrating the ability of the HMW+BSL combination to accurately retrieve PBL temperature and water vapor profiles in this region carries profound implications for meteorological prediction given the inverse relationship between Saharan dust transport and tropical Atlantic hurricane formation [68]. The G5NR atmospheric truth of this test case indicates the presence of a low-level cloud (blue curve in Figure 9, left) above which the elevated concentrations of dust is found (Figure 8 left, green). Both features are detected by the simulated BSL profile (red curve). All model configurations were tested on this location to evaluate the performance in this complex environment (Figure 9, middle). The HMW + BSL model (yellow curve) stands out as the most accurate compared to the truth profile (black), outperforming the ATMS (blue), the ATMS+BSL (green) and HMW (red) models. While the ATMS and ATMS+BSL appear to miss large part of the PBL vertical structure, both the HMW and HMW+BSL models are able to capture the strong temperature inversion, with the HMW+BSL also able to fully capture the free tropospheric warm layer above, which is likely influenced by the transport of warm air within the SAL and radiative warming near the top of the main plume. The bias of the HMW+BSL model in this type of scene further demonstrates how the combination of passive hyperspectral microwave and backscatter lidar can significantly improve the sounding capability in these complex scenes to further improve high-impact tropical cyclone modeling and forecasting.

The second case study examines a particularly challenging scene for passive remote sensing, characterized by a prominent low-level cloud fraction, combined with an abrupt water vapor gradient at the PBL top. The BSL attenuated backscatter return shows a pronounced increase coincident with the cloud feature, effectively detecting this atmospheric structure. Figures 10c and 10d provide broader context to the analysis by showing two critical inputs to the MW and BSL measurement simulations: cloud fraction and cloud optical depth respectively, representing how the same cloud feature is observed by the two instruments providing complementary information during the MW+BSL data fusion approach. The red circle identifies the specific location from which the vertical profiles in panels A and B were extracted, allowing direct connection between the cloud properties and resulting retrieval performance.

While the ATMS-only (blue) and HMW-only (red) retrievals struggle to accurately capture the moisture structure near the cloud top (850 hPa) and within the boundary layer, the BSL-augmented retrievals (green and yellow) closely track the truth profile (black). This improvement occurs because the BSL provides direct constraints on cloud location and boundaries, while the passive sensors contribute information about the integrated water vapor amount. The HMW+BSL combination (yellow) performs best because it leverages both the enhanced spectral information from hyperspectral microwave channels and the precise vertical structure from the lidar.

From a meteorological perspective, this scene represents a critical boundary layer configuration frequently associated with important weather transitions and air quality events. The sharp water vapor gradient at the PBL top indicates a strong capping inversion — a meteorological feature that plays a decisive role in convective inhibition, cloud formation processes, and the vertical mixing of pollutants. Accurately capturing this structure is essential for numerical weather prediction models, particularly for forecasting convective initiation, precipitation onset, and the potential for severe weather development. The presence of low-level clouds further complicates this scenario, as they both modify the radiative balance of the boundary layer and indicate specific thermodynamic conditions that influence subsequent atmospheric evolution. Uniformly cloudy footprints represent a significant opacity barrier for passive sounding instruments, as clearly demonstrated by the limited performance of both ATMS-only and HMW-only water vapor predictions (blue and red curves, Figure 10b). The integration of BSL data, with its strong return signal determined by cloud optical depth, dramatically enhances the vertical definition in the ATMS+BSL model (green curve), effectively ”sharpening” the profile at the PBL top. This improvement stems from the active sensor’s ability to precisely locate the top of the cloud. Most notably, the HMW+BSL configuration (yellow curve) emerges as the superior performer, comprehensively capturing both the internal structure of water vapor within the PBL and the sharp transition into the dry free troposphere. This exceptional performance results from the synergistic combination of HMW’s inherently higher vertical resolution capabilities with BSL’s precise cloud detection abilities. The integration of these complementary measurement techniques enables accurate atmospheric profiling even in meteorological

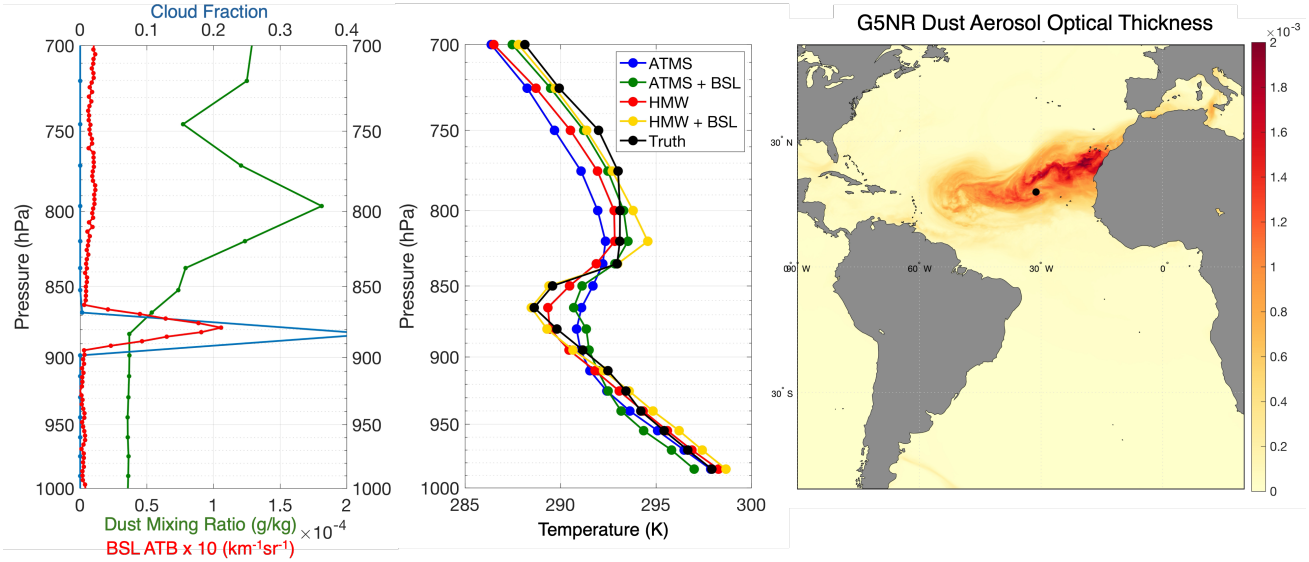


Fig. 9: Profiles of dust mixing ratio (left, green), cloud fraction (left, blue) and backscatter lidar total attenuated backscatter (left, red) for a location within the Atlantic (right, black marker) with G5NR dust optical thickness overplotted. Predicted temperature profiles for ATMS-only (middle,blue), ATMS+BSL (middle, green), HMW-only (middle, red), and HMW+BSL (middle, yellow). Truth profile shown in black.

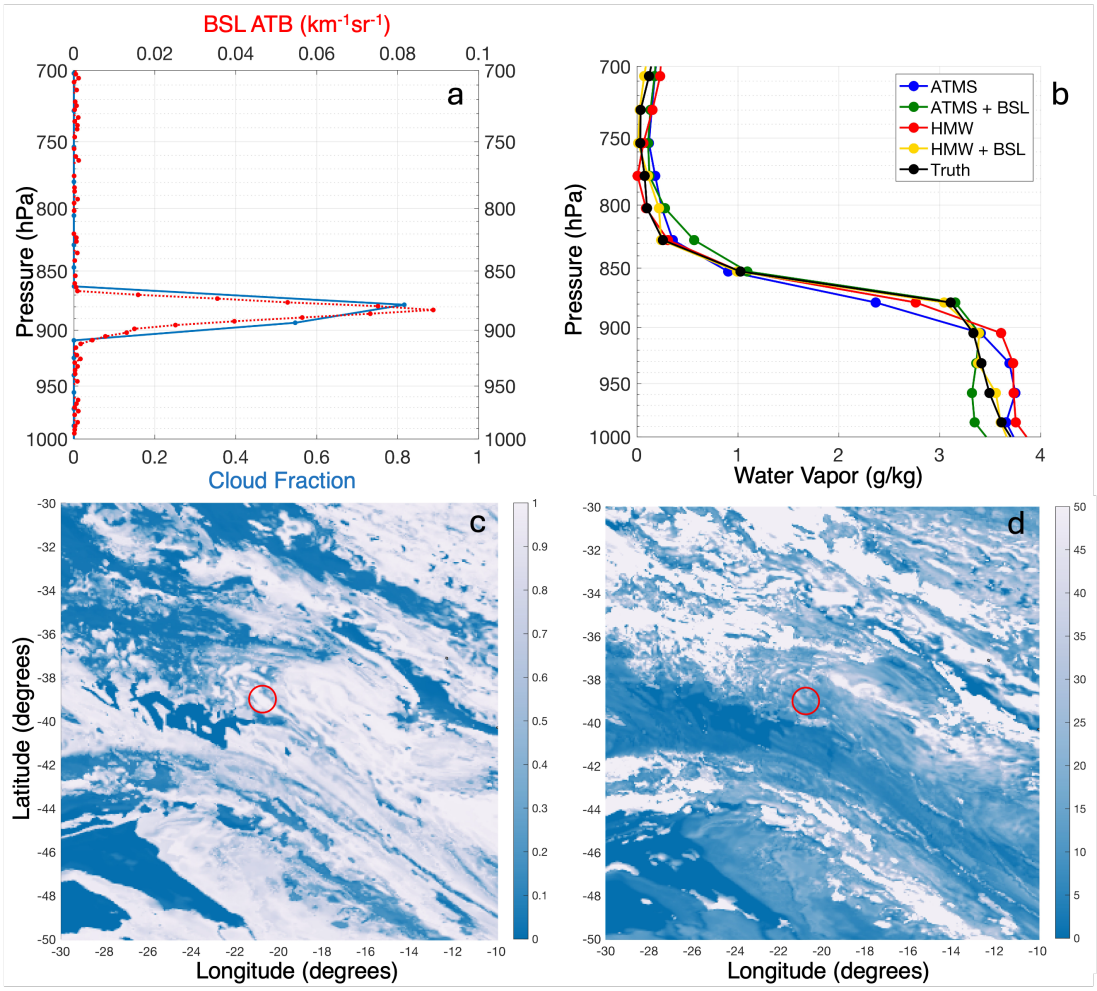


Fig. 10: Profiles of cloud fraction (upper left, blue) and backscatter lidar total attenuated backscatter (upper left, red) for a cloudy region in the South Atlantic. Low level cloud fraction (lower left) and cloud optical depth (lower right) plotted spatially with the location of the extracted profiles (red circle). Predicted water vapor profiles for ATMS-only (middle,blue), ATMS+BSL (middle, green), HMW-only (middle, red), and HMW+BSL (middle, yellow). Truth profile shown in black.

conditions that would severely challenge conventional passive-only observation systems.

The two test cases collectively illustrate why passive-only retrievals face inherent limitations in cloudy conditions, particularly in marine stratocumulus regions, and how the fusion of active and passive sensing technologies creates a measurement system capable of overcoming these challenges. The spatial context of these test cases further emphasizes the widespread nature of these cloud features across the globe, highlighting the broader applicability of these fusion techniques to global atmospheric monitoring.

E. Science Application Traceability Matrices

Science Applications Traceability Matrices (SATMs) represent a fundamental tool in the design and development of new technologies, serving as the essential bridge between scientific objectives and technical implementation. The SATM shown in Table IV exemplifies this critical connection by systematically evaluating how the four distinct measurement approaches — ATMS-only, HMW-only, ATMS+BSL, and HMW+BSL — perform against the PBL DSI mission requirements, clearly demonstrating the enhanced capabilities offered by the novel sensing technologies, especially when combined.

Each configuration is evaluated based on its ability to retrieve three critical atmospheric parameters under all-sky conditions. For temperature retrieval performance (All-sky T), the requirement specifies 1K bias for Goals G1 and G2. ATMS-only achieves 0.2K bias performance with 1.23K SDV, while HMW-only improves this to 0.1K bias with 1.1K SDV. The ATMS+BSL combination matches HMW-only at 0.1K bias with 0.96K SDV, and the HMW+BSL configuration provides the best performance at less than 0.1K bias with 0.85K SDV (corresponding to an overall SDV improvement of $\sim$30% over ATMS). While all configurations meet the 1K DSI PBL bias requirement, the fusion approaches demonstrate clearly superior performance in terms of both bias and SDV metrics. Regarding water vapor retrieval performance (All-sky q), the requirement calls for 10% bias for Goals G1 and G2. Here, ATMS-only falls significantly short at 30% bias performance with 1.19 SDV, whereas HMW-only just meets the threshold at 10% bias with 0.72K SDV. The ATMS+BSL configuration achieves 10% bias with comparable SDV value (0.76K), while HMW+BSL substantially exceeds requirements with less than 5% bias and SDV of 0.59K, corresponding to $\sim$50% improvement over ATMS-only. This parameter vividly demonstrates the limitations of single-instrument approaches compared to the advantages of sensor fusion strategies, particularly for water vapor retrievals, which are notoriously challenging. For PBL height determination (All-sky PBLH), the PBL DSI requirement specifies 100m bias for Goals G1 and G2. In this critical measurement, we specify both bias and mean absolute error (MAE) as the two values provide a more comprehensive evaluation. ATMS-only shows a MAE value of 192.13m, a bias of -32.89m and an SDV of 244.43m. HMW-only shows improvement, with a MAE of 164.03m, a bias of -17.60m and an SDV value of 220.45m. The ATMS+BSL combination becomes generally more stable with 94.11m MAE, -19.13m bias and 152.64m SDV, while HMW+BSL provides the best performance with 93.0m MAE, -17.87m bias, and 153.32m SDV, corresponding to $\sim$37% improvement over ATMS. While accuracy has been primarily defined in terms of bias, MAE serves as an additional useful accuracy metric. Notably, only BSL-enhanced models achieve MAE values below 100 meters (94.11 and 93.00 meters, respectively), successfully meeting the 100-meter accuracy requirement established by the PBL DSI program when MAE is used as the accuracy measure. This represents a significant milestone in PBL height measurement capability that neither passive-only model could achieve on its own. Overall, both PBLH histograms and statistical metrics clearly demonstrate that adding BSL to either ATMS or HMW models significantly improves PBLH performmance, with the HMW+BSL configuration emerging as the best-performing model. The progressive improvement across all parameters demonstrates the substantial value of sensor fusion strategies, particularly for challenging retrievals like PBLH. The SATM clearly indicates that while individual instruments such as ATMS and HMW have certain strengths, they face non-negligible limitations when deployed alone. The combined approaches, especially the HMW+BSL, consistently meet or exceed all scientific requirements for comprehensive PBL characterization.

VI. Conclusions

The study presented in this paper demonstrates a significant advancement in atmospheric thermodynamic sounding retrieval capabilities through the strategic fusion of hyperspectral microwave (HMW) and backscatter lidar (BSL) measurements, with particularly significant enhancements in the Earth's planetary boundary layer (PBL).

Our results demonstrate that fusing HMW and BSL observations yields superior PBL thermodynamic and PBLH retrievals compared to either technique used in isolation. The HMW+BSL configuration consistently outperforms all other methodologies tested across diverse atmospheric conditions, effectively mitigating the inherent limitations of passive-only sensing approaches. This synergistic integration captures fine-scale vertical structure that would otherwise remain unresolved, particularly at the critical PBL top where sharp transitions often occur. Most significantly, our analysis of challenging cloud-impacted scenes demonstrates that the HMW+BSL fusion approach effectively overcomes fundamental limitations that have historically hindered passive-only sounding performance. In regions dominated by marine stratocumulus and strong temperature inversions - where conventional passive-only retrievals frequently fail - the combined approach leverages complementary information, with the HMW providing enhanced sensitivity and the BSL directly constraining cloud boundaries and boundary layer structure.

A. Comparison with the Program of Record (PoR)

The BSL augmentation presented in this study has been shown to provide significant enhancement in the PBL temperature bias vertical stability and a PBL standard deviation error (SDV) reduction of $\sim$30%, compared to traditional microwave

PBL DSI GOAL	Geophysical Variables and Requirements	ATMS	HMW	ATMS + BSL	HMW + BSL
	All-sky T				
G1, G2	1K Bias	0.2 (1.23)	0.1 (1.10)	0.1 (0.96)	< 0.1 (0.85)
	All-sky q				
G1, G2	10% Bias	30% (1.19)	10% (0.72)	10% (0.76)	$< 5\%$ (0.59)
	All-sky PBLH				
G1, G2	100 m Bias	-32.89 (192.13) (154.64)	-17.60 (164.03) (148.32)	-19.13 (94.11) (121.69)	-17.87 (93.0) (123.19)

TABLE IV: Science Applications Traceability Matrix (SATM) for PBL DSI Goals 1 ("G1": PBL, Convection and Extreme Weather) and 2 ("G2": Cloudy PBL). Reported values correspond to Bias values derived from Figure 4 for temperature, "T", and water vapor, "q" and Figure 6 for PBLH. For T and q, values in parenthesis in the second row are corresponding standard deviation errors (SDV). For PBLH, first row indicates bias values, second row is for MAE and third row is for SDV values. For clarity, the term accuracy used in the PBL DSI Study Team has been replaced here by the more commonly used term "bias" to align with the definitions adopted in operational frameworks. Point source simulations have been used in this study across all sensors. See text for considerations on sensor footprint size.

only PoR and novel HMW-only performances. Water vapor retrievals show similar significant enhancements, with the HMW+BSL data fusion configuration reducing water vapor SDV by $\sim$50% in the PBL and maintaining bias values consistently below the 10% bias requirement threshold of the PBL DSI program, compared to PoR errors well exceeding 30% bias in challenging cloud regimes. Beyond thermodynamic profile improvements, our analysis demonstrates remarkable advances in PBL height (PBLH) determination. Notably, only the sensor fusion approaches involving BSL successfully meet this critical PBLH requirement, underscoring the value of complementary active measurement techniques.

B. Significance to the Decadal Survey Incubation Program

As clearly demonstrated in our Science Application Traceability Matrix (SATM), the HMW+BSL fusion approach represents a crucial step forward in meeting critical PBL science goals outlined in the Earth Science Decadal Survey. While standalone systems partially satisfy requirements, only the integrated approach consistently achieves the stringent thresholds necessary for comprehensive characterization of the PBL, reducing the temperature SDV to less than 1K and the bias of water vapor to better than 10% throughout the column while improving the determination of PBLH to within 100 meters.

The advances demonstrated in this study directly address all three critical PBL science objectives identified in the PBL Study Team report. First, our improved characterization of PBL structure enables better understanding of energy, moisture, and momentum exchanges that drive PBL development. Second, enhanced detection of cloud-topped boundary layers and sharp moisture gradients significantly advances our ability to observe PBL-cloud interactions, a critical component in improving climate sensitivity estimates. Third, precise thermodynamic profiling in the lower troposphere directly supports improved prediction of PBL convection and extreme weather by better characterizing atmospheric stability and moisture availability. This data fusion methodology directly supports the upcoming AURORA Pathfinder, a 2024 NASA ESTO project to develop the first photonic integrated circuits based HMW sounder ([15]) as part of the space node of the future PBL sounding architecture. The BSL component of this study can be adapted to support airborne BSLs (*e.g.* Differential Absorption Lidars (DIAL, [17]) and CPL, ([18])) as well as space-borne BSL systems from both the current PoR (*e.g.* EarthCare ([19]), IceSat-2 ([20], and CALIPSO ([21]) and the next generation lidar missions (*e.g.* CASALS ([22]) and CALIGOLA ([23])).

C. Implications for weather forecasting and climate monitoring

The capability of accurately resolving complex PBL structures has profound implications for operational meteorology. It enables more precise identification of atmospheric stability parameters, improves initialization of mesoscale weather models, and enhances prediction of boundary layer evolution - all critical factors for forecasting convective development, fog formation, and pollution dispersion events. Additionally, these improvements provide valuable insights into cloud-topped boundary layer dynamics, which remain one of the most significant sources of uncertainty in climate models [82], [83]. This represents a substantial advance in our ability to monitor and predict key weather phenomena that directly impact public safety, transportation, agriculture, and air quality management.

D. Future research directions

Looking forward, several research directions merit further investigation. A promising area for expansion is the incorporation of additional sensor measurements beyond microwave and lidar, such as infrared, visible, radio occultation and radar

observations, to further mitigate the ill-conditioning problem inherent in space-based atmospheric retrievals. Visible and radar measurements could provide complementary information on cloud microphysics. Infrared sounders could augment information content in thermodynamics, aerosols and trace gases. Additionally, this improved model shows significant potential for implementation within hybrid statistical-physical retrieval frameworks, where the AI-based data fusion approach could provide an agile, computationally efficient first guess to fill the null space of passive-only physical retrieval algorithms, and the physical algorithm enables error estimates and quantification of effective vertical resolution via averaging kernels. This integration would ensure robust quality control for the final product that an AI-only method cannot provide, while maintaining the computational efficiency and enhanced vertical resolution demonstrated in this study.

Future efforts should also address radiative transfer modeling improvements. With the rise of novel technologies and data fusion techniques enabling improved vertical resolution in the PBL, there is a need for optimization of the vertical gridding adopted in operational radiative transfer models like the CRTM. Also, this demonstration was conducted exclusively over oceanic regions, where salinity, wind and surface emissivity modeling are relatively well-established. Extending this analysis to land surfaces would provide valuable insights into the impact of data fusion over heterogeneous terrains. Parallel development efforts on hyperspectral microwave land surface emissivity models are actively underway, which will enable future extension of this retrieval methodology to terrestrial environments. This expansion will require incorporation of additional geophysical parameters into Equation 1, including surface stratification characteristics and terrain elevation data, to account for the complex radiative properties of varied land cover types.

A critical next step in validating this methodology will be the transition from simulated to real observations. The recently completed West-coast & Heartland Hyperspectral Microwave Sensor Intensive Experiment (WH²yMSIE, [85], [86]) provides an exceptional opportunity to verify our data fusion approach using real HMW and BSL observations from the Conical Scanning MIllimeter-Wave Radiometer - Hyperspectral (CoSMIR-H, [87]), CPL and DIAL sensors. This comprehensive field campaign collected coordinated HMW and BSL observations across diverse atmospheric and surface conditions, offering an ideal testbed to conduct spatial HMW and BSL co-registration trade studies and evaluate the real-world performance of the HMW+BSL fusion technique. By applying our methodology to this rich dataset, we will further refine the approach and demonstrate its practical utility in operational contexts. It is important to note that this study represents a first demonstration of the HMW+BSL fusion concept. To that end, it mainly focused on daily PBL cloudy scenes to encourage strong signals from BSL under highly contrasted PBL structure. Future work will expand the analysis to include more complex scenarios, including high clouds, additional aerosol types, nighttime conditions, and land scenes. These extensions will provide a more comprehensive evaluation of the methodology's performance across the full range of atmospheric scenarios encountered in operational environmental monitoring. The present analysis is constrained to nadir-viewing geometries due to the operational limitations of current real space-borne lidar measurement systems. However, emerging technological advances in deployable antennas and onboard processing are expected to increase the spatial resolution of microwave sensors such as AURORA Pathfinder [15], while current developments promise to improve off-nadir lidar capabilities such as those offered by CASALS ([22]). The off-nadir lidar capability will enhance the applicability of the proposed data fusion method to real data scenarios. Additionally, AI methods show significant promise for improving the co-registration of passive and active footprints. These developments are currently under investigation and are expected to substantially increase the spatial coverage and sampling frequency of the proposed data fusion technique, addressing current co-registration limitations in footprint size, swath width, and revisit time that are inherent to passive-active data fusion of real nadir-only lidar observations.

This work provides compelling evidence that the combined HMW+BSL approach represents not merely an incremental improvement but a step-change in atmospheric thermodynamic sounding capability. As we move toward the next generation of Earth observing systems, this fusion methodology offers a promising pathway to address longstanding challenges in PBL thermodynamic sounding, thereby advancing our capability to monitor, understand, and predict weather and climate processes so as to expand Earth science knowledge for the benefit of humanity.

VII. Acknowledgments

This work was supported by NASA ROSES NNH21ZDA001N-DSI research funding. The authors wish to thank the reviewers for the valuable suggestions offered to improve the manuscript.

References

[1] Stull, R.B., "An Introduction to Boundary Layer Meteorology". *Kluwer Academic Publishers*, 1988.

[2] Garratt, J.R., "The Atmospheric Boundary Layer", *Cambridge University Press*, 1994.

[3] Holtslag, A.A.M., "Stable atmospheric boundary layers and diurnal cycles: Challenges for weather and climate models", *Bulletin of the American Meteorological Society*, 94(11), 1691-1706, 2013.

[4] Sandu, I., "Why is it so difficult to represent stably stratified conditions in numerical weather prediction models?", *Journal of Advances in Modeling Earth Systems*, 5(2), 117-133, 2013.

[5] National Research Council, "Weather Services for the Nation: Becoming Second to None", *The National Academies Press*, 2012.

[6] World Meteorological Organization, "Valuing Weather and Climate: Economic Assessment of Meteorological and Hydrological Services", *World Meteorological Organization*, 2015.

[7] National Academies of Sciences, Engineering, and Medicine, 2018a: "Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space", *The National Academies Press: Washington, DC*, 2018.

[8] Teixeira, J, J.R. Piepmeier, A.R. Nehrir, C.O. Ao, S.S. Chen, C.A. Clayson, A.M. Fridlind, M. Lebsock, W. McCarty, H. Salmun, J.A. Santanello, D.D. Turner, Z. Wang, and X. Zeng, *Toward a Global Planetary Boundary Layer Observing System: The NASA PBL Incubation Study Team Report*, 2021.

[9] Teixeira, Joao, Jeffrey R. Piepmeier, Amin R. Nehrir, Chi O. Ao, Shuyi S. Chen, Carol A. Clayson, Ann M. Fridlind, Matthew Lebsock, Will McCarty, Haydee Salmun, Joseph A. Santanello, David D. Turner, Zhien Wang, and Xubin Zeng, "Toward a Global Planetary Boundary Layer Observing System: A Summary". *Bulletin of the American Meteorological Society*, BAMS-D-23-0228.1. https://doi.org/10.1175/BAMS-D-23-0228.1, 2025.

[10] Heraclitus, "Fragments", Floruit Circa 500 B.C.

[11] Edward Kim, Cheng-Hsuan J. Lyu, Kent Anderson, R. Vincent Leslie, William J. Blackwell, "S-NPP ATMS instrument prelaunch and on-orbit performance evaluation", *Journal of Geophysical Research: Atmospheres*, https://doi.org/10.1002/2013JD020483, 2014.

[12] A. Gambacorta et al., "Advancing Atmospheric Thermodynamic Sounding From Space Using Hyperspectral Microwave Measurements," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 5204-5218, 2023, doi: 10.1109/JSTARS.2023.3269697.

[13] A. Gambacorta et al., "The Hyperspectral Microwave Photonic Instrument (HYMPI)," IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 2022, pp. 7206-7209, doi: 10.1109/IGARSS46834.2022.9884548.

[14] F. Gambini, R. Moreira, D. Robles, A. Gambacorta and M. Stephen, "An Ultra-Compact, Narrow-Bandwidth, and High-Density Channel Photonic Integrated Channelizer Based on Serial Arrayed Waveguide Grating Architecture," in *Journal of Lightwave Technology*, vol. 42, no. 8, pp. 2908-2916, 15 April15, 2024, doi: 10.1109/JLT.2024.3349932.

[15] Gambacorta et al., 2024, "The Advanced Ultra-high Resolution Optical RAdiometer (AURORA) Pathfinder", https://ams.confex.com/ams/105ANNUAL/meetingapp.cgi/Paper/454398.

[16] Yorks, J. et al. (2024). A SmallSat Lidar Concept for Measurements of Aerosol and Cloud Spatiotemporal Variability. In: Singh, U.N., Tzeremes, G., Refaat, T.F., Ribes Pleguezuelo, P. (eds) Space-based Lidar Remote Sensing Techniques and Emerging Technologies. LIDAR 2023. Springer Aerospace Technology. Springer, Cham.

[17] Carroll, B. J., Nehrir, A. R., Kooi, S. A., Collins, J. E., Barton-Grimley, R. A., Notari, A., Harper, D. B., and Lee, J.: Differential absorption lidar measurements of water vapor by the High Altitude Lidar Observatory (HALO): retrieval framework and first results, Atmos. Meas. Tech., 15, 605–626, https://doi.org/10.5194/amt-15-605-2022, 2022.

[18] Matthew McGill, Dennis Hlavka, William Hart, V. Stanley Scott, James Spinhirne, and Beat Schmid, "Cloud Physics Lidar: instrument description and initial measurement results," Appl. Opt. 41, 3725-3734 (2002).

[19] Wehr, T., Kubota, T., Tzeremes, G., Wallace, K., Nakatsuka, H., Ohno, Y., Koopman, R., Rusli, S., Kikuchi, M., Eisinger, M., Tanaka, T., Taga, M., Deghaye, P., Tomita, E., and Bernaerts, D.: The EarthCARE mission – science and system overview, Atmos. Meas. Tech., 16, 3581–3608, https://doi.org/10.5194/amt-16-3581-2023, 2023.

[20] Abdalati, W., Zwally, H. J., Bindschadler, R., Csatho, B., Farrell, S. L., Fricker, H. A., et al. (2010). The ICESat-2 Laser Altimetry Mission. Proc. IEEE 98 (5), 735–751. doi:10.1109/JPROC.2009.2034765.

[21] Winker, D. M., and Coauthors, 2010: The CALIPSO Mission. Bull. Amer. Meteor. Soc., 91, 1211–1230, https://doi.org/10.1175/2010BAMS3009.1.

[22] Yang et al., 2024, "Development of Concurrent Artificially-Intelligent Spectrometry and Adaptive Lidar System (CASALS) for Swath Mapping From Space", https://ntrs.nasa.gov/citations/20240001267. Document ID 20240001267.

[23] Di Girolamo, P., Cosentino, A., Longo, F., Franco, N., Dionisi, D., Summa, D., Lolli, S., Suetta, E., Perna, A., and Zoffoli, S.: Cloud Aerosol Lidar for Global Scale Observations of the Ocean-Land-Atmosphere System – CALIGOLA, EGU General Assembly 2022, Vienna, Austria, 23–27 May 2022, EGU22-5644, https://doi.org/10.5194/egusphere-egu22-5644, 2022.

[24] Roldán-Henao, N., Yorks, J.E., Su, T., Selmer, P.A., Li, Z., "Statistically Resolved Planetary Boundary Layer Height Diurnal Variability Using Spaceborne Lidar Data", *Remote Sensing*, 2024, 16, 3252. https://doi.org/10.3390/rs16173252

[25] Palm, S. P., P.A. Selmer, J.E. Yorks, S. Nicholls, E.P. Nowottnick (2021), Planetary Boundary Layer Height Estimates from ICESat-2 and CATS Backscatter Measurements, *Frontiers*, 2, 29, https://doi.org/10.3389/frsen.2021.716951.

[26] McGrath-Spangler, E. L., and Denning, A. S. (2013). Global Seasonal Variations of. Midday Planetary Boundary Layer Depth from CALIPSO Space-Borne Lidar. *J. Geophys. Res. Atmos.* 118, 1226–1233. doi:10.1002/jgrd.50198.

[27] A. Kotsakis et al., "Hyperspectral Microwave Measurement Demonstrations of Improved Thermodynamic Sounding from Space," IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 2023, pp. 1448-1449, doi: 10.1109/IGARSS52108.2023.10282289.

[28] J. MacKinnon et al., "Multi-Path Fusion: A Hierarchical Machine Learning Approach for Combining Diverse Data Sets for a Forest Monitoring New Observing System," IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 2023, pp. 1708-1711, doi: 10.1109/IGARSS52108.2023.10282678.

[29] Prigent, C., Jimenez, C., Aires, F. (2016). Toward "all weather," long record, and real-time land surface temperature retrievals from microwave satellite observations. *Journal of Geophysical Research: Atmospheres*, 121(10), 5699-5717.

[30] Norouzi, H., Temimi, M., Khanbilvardi, R., Prigent, C., Turk, F. J., Gonzalez, R., Vega, M., Li, R. (2015). Assessment of the consistency among global microwave land surface emissivity products. *Atmospheric Measurement Techniques*, 8(3), 1197-1205.

[31] Ferraro, R. R., Peters-Lidard, C. D., Hernandez, C., Turk, F. J., Aires, F., Prigent, C., Lin, X., Boukabara, S. A., Furuzawa, F. A., Gopalan, K., Harrison, K. W., Karbou, F., Li, L., Liu, C., Masunaga, H., Moy, L., Ringerud, S., Skofronick-Jackson, G. M., Tian, Y., Wang, N.-Y. (2013). An evaluation of microwave land surface emissivities over the continental United States to benefit GPM-era precipitation algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 51(1), 378-398.

[32] A. Gambacorta and C. D. Barnet, "Methodology and Information Content of the NOAA NESDIS Operational Channel Selection for the Cross-Track Infrared Sounder (CrIS)," in IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 6, pp. 3207-3216, June 2013, doi: 10.1109/TGRS.2012.2220369.

[33] Yorks, J. E., and Coauthors, 2023: A SmallSat Concept to Resolve Diurnal and Vertical Variations of Aerosols, Clouds, and Boundary Layer Height. *Bull. Amer. Meteor. Soc.*, 104, E815–E836, https://doi.org/10.1175/BAMS-D-21-0179.1.

[34] Turner, D. D. and W. G. Blumberg, (2019), Improvements to the AERIoe Thermodynamic Profile Retrieval Algorithm, *IEEE Sel. Top. Appl. Earth Obs. Remote Sens.*, doi:10.1109/JSTARS.2018.2874968.

[35] Saeed U and Rocadenbosch F., Adaptive Estimation of the Stable Boundary Layer Height Using Combined Lidar and Microwave Radiometer Observations, *IEEE TGRS*, Vol. 54, No. 12, 2016, 10.1109/TGRS.2016.2586298.

[36] Cimini, D., Haeffelin, M., Kotthaus, S. et al. Towards the profiling of the atmospheric boundary layer at European scale—introducing the COST Action PROBE. *Bull. of Atmos. Sci. Technol.* 1, 23–42 (2020). https://doi.org/10.1007/s42865-020-00003-8.

[37] Smith, E. N., et al. (2021). Evaluation and application of multi-instrument boundary-layer thermodynamic retrievals. *Boundary-Layer Met.* 181, 95-123.

[38] Barrera-Verdejo, M., S. Crewell, U. Löhnert, E. Orlandio, and P. D. Girolamo, 2016: Ground-based lidar and microwave radiometry synergy for high vertical resolution absolute humidity profiling. *Atmos. Meas. Tech.* 9, 4013-4028.

[39] Wang, T., E. J. Fetzer, S. Wong, B. H. Kahn, and Q. Yue (2016), Validation of MODIS cloud mask and multilayer flag using CloudSat-CALIPSO cloud profiles and a cross-reference of their cloud classifications, *J. Geophys. Res. Atmos.*, 121, 11,620–11,635, doi:10.1002/2016JD025239.

[40] Yao, Z., J. Li, E. Weisz, A. Heidinger, and C.-Y. Liu (2013), Evaluation of single field-of-view cloud top height retrievals from hyperspectral infrared sounder radiances with CloudSat and CALIPSO measurements, *J. Geophys. Res. Atmos.*, 118, 9182–9190, doi:10.1002/jgrd.50681.

[41] Weisz, E., J. Li, W. P. Menzel, A. K. Heidinger, B. H. Kahn, and C.-Y. Liu (2007), Comparison of AIRS, MODIS, CloudSat and CALIPSO cloud top height retrievals, *Geophys. Res. Lett.*, 34, L17811, doi:10.1029/2007GL030676.

[42] Oreopoulos, L., N. Cho, and D. Lee (2017), New insights about cloud vertical structure from CloudSat and CALIPSO observations, *J. Geophys. Res. Atmos.*, 122, 9280–9300, doi:10.1002/2017JD026629.

[43] Advancing geoscience with AI. *Nat. Geosci.* 17, 947 (2024). https://doi.org/10.1038/s41561-024-01572-5.

[44] Rodgers, C.D., Inverse Methods for Atmospheric Sounding: Theory and Practice. 2000: World Scientific Publishing Company, Incorporated.

[45] Antonia Gambacorta, Atmospheric Soundings From Space: Bridging Theory and Practice, Reference Module in Earth Systems and Environmental Sciences, Elsevier, 2025, ISBN 9780124095489, https://doi.org/10.1016/B978-0-443-13220-9.00025-1.

[46] Milstein, A. B., J. A. Santanello, and W. J. Blackwell, 2023: Detail Enhancement of AIRS/AMSU Temperature and Moisture Profiles Using a 3D Deep Neural Network. *Artif. Intell. Earth Syst.*, 2, 220037, https://doi.org/10.1175/AIES-D-22-0037.1.

[47] J. Susskind, J. Blaisdell, L. Iredell, and F. Keita, "Improved Temperature Sounding and Quality Control Methodology Using AIRS/AMSU Data: The AIRS Science Team Version-5 Retrieval Algorithm", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, 3, doi:10.1109/TGRS.2010.2070508, 2011.

[48] DeSouza-Machado S., L. Larrabee Strow, Andrew Tangborn, Xianglei Huang, Xiuhong Chen, Xu Liu, Wan Wu, and Qiguang Yang, Single-footprint retrievals for AIRS using a fast TwoSlab cloud-representation model and the SARTA all-sky infrared radiative transfer algorithm, *Atmos. Meas. Tech.*, 11, 529–550, 2018 https://doi.org/10.5194/amt-11-529-2018.

[49] E.P. Nowottnick, J.E. Yorks, M.J. McGill, P.A. Selmer, and K.E. Christian, "A Simulation Capability Developed for NASA GSFC's Spaceborne Backscatter Lidars: Overview and Projected Performance for the Upcoming AOS Mission," in *Proceedings of the 30th International Laser Radar Conference*, J.T. Sullivan et al., Eds. Cham: Springer International Publishing, 2023, pp. 675–681. doi: 10.1007/978-3-031-37818-8_87.

[50] Pauly, R., J.E. Yorks, D.L. Hlavka, M.J. McGill, V. Amiridis, S.P. Palm, S.D. Rodier, M.A. Vaughan, P. Selmer, A.W. Kupchock, H. Baars, A. Gialitaki (2019), CATS 1064 nm Calibration and Validation, *Atmos. Meas. Tech.*, 12, 6241–6258, https://doi.org/10.5194/amt-12-6241-2019.

[51] Yorks, J. E., et al. (2016), "An overview of the CATS level 1 processing algorithms and data products," *Geophysical Research Letters*, 43(9), 4632-4639.

[52] Welton, E.J. and Campbell, J.R., 2002. Micropulse lidar signals: Uncertainty analysis. Journal of Atmospheric and Oceanic Technology, 19(12), pp.2089-2094.

[53] Putman, W., A.M. da Silva, L.E. Ott and A. Darmenov, (2014), Model Configuration for the 7-km GEOS-5.12 Nature Run, Ganymed Release (Non-hydrostatic 7 km Global Mesoscale Simulation). GMAO Office Note No. 5 (Version 1.0), 18 pp, available from http://gmao.gsfc.nasa.gov/pubs/.

[54] Marchand, R., G. G. Mace, T. Ackerman, and G. Stephens, 2008, Hydrometeor detection using Cloudsat—An Earth-orbiting 94-GHz cloud radar, *Journal of Atmospheric and Oceanic Technology*, 25(4), 519-533. https://doi.org/10.1175/2007JTECHA1006.1

[55] Mace, G. G., and N. L. Wrenn, 2013: Evaluation of the Hydrometeor Layers in the East and West Pacific within ISCCP Cloud-Top Pressure–Optical Depth Bins Using Merged CloudSat and CALIPSO Data. *Journal of Climate*, 26(23), 9429-9444. https://doi.org/10.1175/JCLI-D-12-00207.1

[56] C. Spradlin, J.A. Caraballo-Vega, J. Li, M.L. Carroll, J. Gong, and P.M. Montesano, "SatVision-TOA: A Geospatial Foundation Model for Coarse-Resolution All-Sky Remote Sensing Imagery," *arXiv preprint arXiv:2411.17000*, 2024. doi: 10.48550/ARXIV.2411.17000.

[57] C. -H. Lyu, E. J. Kim, L. M. Mccormick, R. V. Leslie and I. A. Osaretin, "JPSS-1 ATMS Postlaunch Active Geolocation Analysis," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 11, pp. 9462-9471, Nov. 2021, doi: 10.1109/TGRS.2020.3047339.

[58] N. R. Nalli et al., "Validation of Atmospheric Profile Retrievals From the SNPP NOAA-Unique Combined Atmospheric Processing System. Part 1: Temperature and Moisture," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 180-190, Jan. 2018, doi: 10.1109/TGRS.2017.2744558.

[59] W. J. Blackwell, L. J. Bickmeier, R. V. Leslie, M. L. Pieper, J. E. Samra, C. Surussavadee, and C. A. Upham, "Hyperspectral microwave atmospheric sounding," *IEEE Trans. Geosci. Remote Sens.*, vol. 49(1), pp. 128-–142, doi:10.1109/TGRS.2010.2052260, 2011.

[60] B. T. Johnson, C. Deng, P. Stegmann, Q. Liu, I. Moradi, and T. Auligne, "The Community Radiative Transfer Model (CRTM): Community-Focused Collaborative Model Development Accelerating Research to Operations," *Bulletin of the American Meteorological Society*, vol. 104(10), pp. E1817–E1830, doi:10.1175/BAMS-D-22-0015.1, 2023.

[61] P. Stegmann, B. Johnson, I. Moradi, B. Karpowicz, W. McCarty, H. Liu, S. Dutta, T. Auligne, "The CRTM transmittance coefficient package," *JQSRT*, vol. 336, pp. 109380, doi:10.1016/j.jqsrt.2025.109380, 2025.

[62] S. Lu, Q. Liu, S. Wei, B. T. Johnson, et al., "The Aerosol Module in the Community Radiative Transfer Model (v2.2 and v2.3): accounting for aerosol transmittance effects on the radiance observation operator," *Geoscientific Model Development*, vol. 15(3), pp. 1317–1329, doi:10.5194/gmd-15-1317-2022, 2022.

[63] S. Wei, S. Lu, B. T. Johnson, C. Deng, et al., "The Influence of Aerosols on Satellite Infrared Radiance Simulations and Jacobians: Numerical Experiments of CRTM and GSI", *MDPI Remote Sensing*, vol. 14(3), pp. 683, doi:10.3390/rs14030683, 2022.

[64] I. Moradi, B. Johnson, P. Stegmann, D. Holdaway, G. Heymsfield, and R. Gelaro, "Developing a Radar Signal Simulator for the Community Radiative Transfer Model," *IEEE Trans. on Geosci. and Rem. Sens.*, vol. 61., pp. 5111113, doi:0.1109/TGRS.2023.3330067, 2023.

[65] I. Moradi, P. G. Stegmann, B. T. Johnson, V. Barlakas, et al., "Implementation of a Discrete Dipole Approximation Scattering Database Into Community Radiative Transfer Model," *JGR: Atmospheres*, vol. 127(24), pp. e2022JD036957, doi:10.1029/2022JD036957, 2022.

[66] B. Karpowicz, P. G. Stegmann, B. T. Johnson, H. Christophersen, et al., "pyCRTM: A Python Interface for the Community Radiative Transfer Model," *JQSRT*, vol. 288(11), pp. 108263, doi:10.1016/j.jqsrt.2022.108263, 2022.

[67] von Engeln, A., and J. Teixeira, "A Planetary Boundary Layer Height Climatology Derived from ECMWF Reanalysis Data." *J. Climate*, 26, 6575–6590, https://doi.org/10.1175/JCLI-D-12-00385.1, 2013.

[68] Wang, C., S. Dong, A. T. Evan, G. R. Foltz, and S. Lee, "Multidecadal Covariability of North Atlantic Sea Surface Temperature, African Dust, Sahel Rainfall, and Atlantic Hurricanes". *J. Climate*, 25, 5404–5415, https://doi.org/10.1175/JCLI-D-11-00413.1, 2012.

[69] Loveless, D., M., R. O. Knuteson, T. J. Wagner, M. L. Loveless, D. D. Turner, and R. B. Pierce, 2025: Importance of the vertical grid choice for passive infrared sounding from space-based and ground-based platforms. J. Atmos. Oceanic Technol. Accepted Manuscript.

[70] Hersbach H, Bell B, Berrisford P, et al. The ERA5 global reanalysis. Q J R Meteorol Soc. 2020; 146: 1999–2049. https://doi.org/10.1002/qj.3803

[71] L. L. Strow, S. E. Hannon, S. De Souza-Machado, H. E. Motteler and D. Tobin, An overview of the AIRS radiative transfer model, in IEEE Transactions on Geoscience and Remote Sensing, vol. 41, no. 2, pp. 303-313, Feb. 2003, doi: 10.1109/TGRS.2002.808244.

[72] Agudelo, P.A., Hoyos, C.D., Curry, J.A. et al., "Probabilistic discrimination between large-scale environments of intensifying and decaying African Easterly Waves," *Clim Dyn* 36, 1379–1401, https://doi.org/10.1007/s00382-010-0851-x, 2011.

[73] Wong, Sun, Andrew E. Dessler, Natalie M. Mahowald, Ping Yang, and Qian Feng. "Maintenance of Lower Tropospheric Temperature Inversion in the Saharan Air Layer by Dust and Dry Anomaly". *Journal of Climate*, 22.19, 5149-5162. https://doi.org/10.1175/2009JCLI2847.1, 2009.

[74] S. -A. Boukabara et al., "MiRS: An All-Weather 1DVAR Satellite Data Assimilation and Retrieval System," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 9, pp. 3249-3272, Sept. 2011, doi: 10.1109/TGRS.2011.2158438.

[75] Dunion, Jason P., and Christopher S. Velden. "The Impact of the Saharan Air Layer on Atlantic Tropical Cyclone Activity". *Bulletin of the American Meteorological Society*, 85.3, 353-366. https://doi.org/10.1175/BAMS-85-3-353, 2004.

[76] Arnold, Nathan. "Re: GEOS PBL." Received by Joseph Santanello, 30 Nov. 2022.

[77] Arnold, Nathan. "Re: GEOS PBL." Received by Joseph Santanello, 7 Dec. 2022.

[78] Rizza, Umberto, et al. "Evaluating the Direct Radiative Forcing of a Giant Saharan Dust Storm". *Atmospheric Research*, vol. 315, p. 107875, https://doi.org/10.1016/j.atmosres.2024.107875, 2025.

[79] Satheesh, S. K., and K. Krishna Moorthy. "Radiative Effects of Natural Aerosols: A Review". *Atmospheric Environment*, vol. 39, no. 11, pp. 2089–2110, https://doi.org/10.1016/j.atmosenv.2004.12.029, 2005.

[80] Ding, F., Iredell, L., Theobald, M., Wei, J., and Meyer, D., "PBL height from AIRS, GPS RO, and MERRA-2 products in NASA GES DISC and their 10-year seasonal mean intercomparison." *Earth and Space Science*, 8, e2021EA001859. https://doi.org/10.1029/2021EA001859, 2021.

[81] Fetzer, E. J., J. Teixeira, E. T. Olsen, and E. F. Fishbein, "Satellite remote sounding of atmospheric boundary layer temperature inversions over the subtropical eastern Pacific", *Geophys. Res. Lett.*, 31, L17102, doi:10.1029/2004GL020174, 2004.

[82] Su, T., Zhang, Y., Tian, J. (2025). Boundary-layer-coupled and decoupled clouds in global storm-resolving models: Comparisons with the ARM observations. *Journal of Geophysical Research: Atmospheres*, 130, e2024JD041915. https://doi.org/10.1029/2024JD041915

[83] Ahlgrimm, M., Forbes, R., Hogan, R. J., Sandu, I. (2018). Understanding global model systematic shortwave radiation errors in subtropical marine boundary layer cloud regimes. *Journal of Advances in Modeling Earth Systems*, 10, 2042–2060. https://doi.org/10.1029/2018MS001346

[84] https://www.nesdis.noaa.gov/our-satellites/future-programs/near-earth-orbit-network-neon

[85] https://earth.gsfc.nasa.gov/climate/campaigns/WHyMSIE

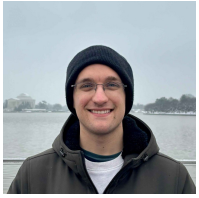
[86] https://ntrs.nasa.gov/citations/20240006548

[87] Kroodsma R., I. Adams, T. Ames, M. Fritts, A. Gambacorta, A. Kotsakis1, P. Mohammed, J. Piepmeier, "COSMIR-H: An Airborne Hyperspectral Microwave Sounder For Thermodynamic Sensing Of The

Planetary Boundary Layer", proceedings of IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 2024.

Antonia Gambacorta is a Research Physical Scientist with the Climate and Radiation Laboratory at the NASA Goddard Space Flight Center (GSFC), Greenbelt, MD, USA. She received the Laurea and M.S. degree in physics (2001) from the Universita' degli Studi A. Moro of Bari, Italy, and the M.S. (2005) and Ph.D. (2008) degrees in atmospheric physics from the University of Maryland Baltimore County, Baltimore, MD, USA. She currently serves as the Principal Investigator of the Advanced Ultra-high Resolution Optical RAdiofrequency (AURORA) Pathfinder and the Principal Investigator of the Hyperspectral Microwave Photonic Instrument (HyMPI). She is the Principal Investigator of the Westcoast and Heartland Hyperspectral Microwave Sensor Intensive Experiment (WH2yMSIE). She also serves as the Joint Polar Satellite System (JPSS) and Near Earth Orbit Network (NEON) Project Scientist. Prior to joining GSFC, she served as the Science Lead of the SNPP, Metop, and JPSS NOAA Unique Combined Atmospheric Processing System (NUCAPS) for NOAA/NESDIS/STAR and as a Subject Matter Expert in satellite remote sensing for NOAA JPSS. Her research interests include passive (hyperspectral microwave and infrared) and active (backscatter lidar) remote sensing techniques and inverse methods for the retrieval of atmospheric temperature and constituents, with a focus on Earth's planetary boundary layer science applications.

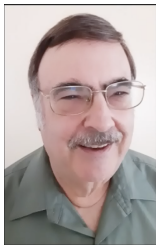
Alexander Kotsakis received the B.S. degree in meteorology with Valparaiso University, Valparaiso, IN, USA, in 2013, and the Ph.D. degree in atmospheric science with the University of Houston, Houston, TX, USA, in 2017. He is currently an Assistant Research Scientist with the Earth System Science Interdisciplinary Center-University of Maryland (contractor to NASA GSFC). He serves as the Demonstration and Application Lead for the Advanced Ultra-high Resolution Optical and RAdiometer (AURORA) Pathfinder science team. He recently served as deputy principal investigator and mission scientist for the Westcoast and Heartland Hyperspectral Microwave Sensor Intensive Experiment (WH2yMSIE), a field campaign focused on demonstrating the capability of a multisensor payload for a future PBL mission. He has extensive experience utilizing remote sensing measurements and meteorological data together to better understand atmospheric processes. He is currently supporting the development and demonstration of an AI/ML planetary boundary layer (PBL) retrieval algorithm using his extensive knowledge of the boundary layer meteorology and assists in trade studies for future hyperspectral microwave sensors.



Dave Gershman received the B.S. degree in Computer Science with the University of Central FLorida in 2024. He served as a Computer Engineer at NASA Goddard Space Flight Center. His work was centered on the Machine Learning (ML) and Artificial Intelligence (AI) efforts specifically working with the data-fusion of multi-modal data from various instruments. This included large amounts of data processing used to train and tune complex model architecture. He has proficiency in various programming languages and strives for creating code infrastructure that is both clean and efficient.

Narges Shahroudi received the B.S. and Ph.D. degrees in Electrical Engineering from the City University of New York in 2008 and 2014, respectively. She is currently an Assistant Research Scientist with the Earth System Science Interdisciplinary Center at the University of Maryland, supporting NASA Goddard Space Flight Center. Prior to joining NASA GSFC, she worked at NOAA/NESDIS/STAR, where she led data assimilation projects focused on integrating satellite observations into numerical weather prediction systems.She currently serves as the Trade Study and Physical Retrieval Lead for the Advanced Ultra-high Resolution Optical and RAdiometer (AURORA) Pathfinder science team. Her research spans a range of topics, including remote sensing, data assimilation, and machine learning. Her current efforts focus on the development and demonstration of PBL retrieval algorithms and conducting trade studies to support the design of future hyperspectral microwave instruments.

Robert Rosenberg Mr. Robert Rosenberg earned a BS in Earth and Space Science (1976), and an MS in Mechanical Engineering (1977) at SUNY Stony Brook. He has 46 years continuous service to NASA projects as a scientific programmer/analyst, meteorologist, team leader, and program manager. He currently manages two NASA Earth science projects: the Atmospheric Modeling and Data Assimilation and the NASA Land Information Systems. Mr. Rosenberg has also used his programming and meteorological expertise to support science research initiatives, starting with the First GARP Global Experiment (FGGE), and recently including the AIRS Sounder Science Team, for which he designed and ran observing system experiments to test the impact of AIRS/AMSU temperature retrievals on forecast skill using the NASA GEOS-5 Data Assimilation and modeling system. Mr. Rosenberg works with the NASA GSFC AURORA Pathfinder team assisting in observing system simulation experiments.



John Blaisdell received the B.A. degree in mathematics and the B.S. degree in physics from the University of Rochester, Rochester, NY, both in 1978. He received the M.S. and Ph.D. degrees in 1980 and 1982, respectively, from the University of Illinois, Urbana, both in physics. His dissertation research covered electronic structure of metal oxide semiconductors in the vicinity of surface defects. From 1982 to 1983, he was with the Chemistry Department, Johns Hopkins University, Baltimore, MD, where he studied the electronic structure of molecular crystals. He has been with Science Applications International Corp./General Sciences Corp., Beltsville, MD, since 1983, working on space and atmospheric applications. Since 1993, he has been implementing scientific algorithms for atmospheric sounding applications.

Ed Nowottnick received a B.S. in Physical Sciences in 2005, a M.S. in Atmospheric and Oceanic Science in 2007, and a Ph.D. in Atmospheric and Oceanic Science in 2011. As part of his dissertation and post-doctoral work at NASA GSFC, Dr. Nowottnick worked on the NASA GEOS model to improve dust parameterizations, before joining the GSFC Cloud Physics Lidar (CPL) and Cloud-Aerosol Tranport System (CATS) group in 2015, where has served as the aerosol typing lead. Dr. Nowottnick has led and supported numerous NASA field campaigns, including WH^{2}yMSIE. Dr. Nowottnick also served as Instrument Scientist for ALICAT, originally planned to be included in the AOS-Storm mission.

Kenneth Christian received a B.S. in Geography from the University of Idaho in 2010 and a Ph.D. in Meteorology from Pennsylvania State University in 2017. His dissertation work focused on conducting a global sensitivity study of modeled tropospheric oxidants (ozone, hydroxyl radical, and hydroperoxyl radical) for time periods corresponding to NASA airborne field campaigns. During his postdoctoral work at the University of Iowa and NASA GSFC, he utilized space-based lidar to constrain the altitude of pyrocumulonimbus aerosol emissions in the GEOS-Chem model and simulated the evolution and lifetime of the pyrocumulonimbus aerosol plumes. Dr. Christian is currently an Assistant Research Scientist at the University of Maryland's Earth System Science Interdisciplinary Center, where he works with the lidar atmospheric remote sensing group at NASA GSFC. He has supported airborne lidars in numerous NASA airborne field campaigns, including WH^{2}yMSIE, and serves as a member of the ICESat-2 science team, where he focuses on improving its atmospheric data products.

Jordan A. Caraballo-Vega received his B.S. degree in Computational Math and Computer Science from the University of Puerto Rico, Humacao, in 2016. Following his graduation, he contributed to research and software development efforts focused on Earth and Space science applications at NASA Goddard Space Flight Center. He is currently a Ph.D. candidate in Geographical Sciences at the University of Maryland, College Park. His interdisciplinary research focuses on developing high-performance computing and artificial intelligence software to support large-scale Earth and Space observation studies. His work includes the creation of GPU-accelerated software for satellite data processing, deep learning models for GeoAI and atmospheric problems, large-scale foundation model development, and open-source tools that enable scalable, reproducible science in cloud and on-premises environments. He is particularly interested in bridging the gap between Earth and Space science and AI through efficient and transparent data-driven and physics-informed methods.

James MacKinnon is currently working in the civil space field as a senior software engineer at Voyager Technologies specializing in high performance computing for in-space applications. Previously, he was at NASA Goddard Space Flight Center where he was a software lead working on many cutting-edge applications of Artificial Intelligence and Machine Learning for both the Earth and Planetary Sciences. His background covers a wide range of areas including AI/ML, algorithm development, embedded systems, and high reliability systems. He is the recipient of several NASA service awards including the Robert H. Goddard award and the NASA Exceptional Engineering Achievement Medal.

Patrick Stegmann received a Dr.-Ing. degree in mechanical and process engineering from the TU Darmstadt (Germany) in 2015. Dr. Stegmann was a postdoctoral research associate at Texas A&M University from 2016 to 2018, and from 2018 onward a software engineer and project scientist at the Joint Center for Satellite Data Assimilation (JCSDA), as well as a senior scientist at VIAVI Solutions. He is currently an Associate Research Scientist with the Earth System Science Interdisciplinary Center-University of Maryland and a cooperative scientist at NASA GSFC. He serves as the Radiative Transfer Lead for the Advanced Ultra-high Resolution Optical and RAdiometer (AURORA) Pathfinder science team. His research spans a diverse range of topics, including radiative transfer modeling, electromagnetic scattering, satellite data assimilation, software engineering in support of the CRTM and JEDI frameworks, and Position, Navigation, and Timing (PNT) technology.

Stephen D. Nicholls is a senior research scientist with Science Systems and Applications Inc. (SSAI) and a contractor at NASA Goddard Space Flight Center (GSFC). He received a B.S. degree in meteorology from the Pennsylvania State University, State College, PA, USA (2005), an M.S. degree in atmospheric science the from the State University of New York, University at Albany, Albany, NY, USA (2008), and a Ph.D. in atmospheric science from Rutgers University, New Brunswick, NJ, USA (2012). Dr. Nicholls presently serves as software architecture lead for *P*assive-*A*ctive *N*ear-real *T*ime *AR*chit*E*cture for PBL Sound*I*ng (PANTA-REI), which will be integral to the upcoming Advanced Ultra-high Resolution Optical RAdiometer (AURORA) Pathfinder mission. He also served as a supporting mission scientist for the recent West Coast and Heartland Hyperspectral Microwave Sensor Intensive Experiment (WH^{2}yMSIE) field campaign and as a lead forecaster for several NASA airborne field campaigns (IMPACTS, WH^{2}yMSIE, and GLOVE) and for the upcoming NURTURE campaign. His research interests include data fusion, numerical weather and climate prediction, machine learning and artificial intelligence, and active (lidar and radar) and passive (microwave and infrared radiometer) sensors.

Joseph Santanello received the M.S. degree in meteorology from Pennsylvania State University, State College, PA, USA, in 1999, and the Ph.D. degree in geography from Boston University, Boston, MA, USA, in 2005. He leads the GEWEX LoCo working group focused on coordinating international efforts to diagnose and improve land–atmosphere interactions in models from local to global scales, over the last 15 years. He serves as Land Team lead of the NASA Unified WRF (NU-WRF) modeling system, which includes coupling to NASA's Land Information System and facilitates model-data fusion research in the realm of land data assimilation (e.g., soil moisture) and land surface model calibration (e.g., surface fluxes). He served as the Lead of GSFC working group supporting NASA's Decadal Survey Incubation efforts for the PBL, assessing current capabilities and limitations of satellite remote sensing to routinely monitor the PBL for land surface and land–atmosphere studies and modeling, including engagement of GSFC and mission science teams, as well as the PBL community in response to the 2017 Decadal Survey.

William G. Blumberg is an Assistant Professor of Meteorology at Millersville University in Millersville, PA. Prior to joining their faculty, he was at the NASA Goddard Space Flight Center as a fellow of the NASA Postdoctoral Program. He obtained his meteorology degrees from the University of Oklahoma and helped obtain the funding for this research. His expertise ranges across many areas, including the atmospheric boundary layer and instrumentation, and he enjoys working on applied meteorology problems.