

# Toward the Next Generation of Microwave Sounders: Benefits of a Low-Earth Orbit Hyperspectral Microwave Instrument in All-Weather Conditions Using AI

Eric S. Maddy , Flavio Iturbide-Sanchez , *Senior Member, IEEE*, and Sid Ahmed Boukabara 

**Abstract**—This study presents scientific results that serve as arguments for advocating the development of a hyperspectral microwave sensor (HyMS). Through simulation experiments, the results of this study demonstrate the major benefits of HyMS sensor observations in low-Earth orbit (LEO), including; 1) increased information content over the microwave region, 2) improved temperature and moisture sounding in all-weather conditions, resulting from higher signal-to-noise ratios, finer vertical resolution, and a reduced dependence on background information due to the increased spectral resolution around oxygen and water vapor absorption features between 23 and 183 GHz, 3) improved profiling of hydrometeors, and 4) improved resilience to radio frequency interference, demonstrated at 23 GHz, associated with the redundant information provided by the HyMS. The deployment of HyMS instruments in LEO orbit is expected to provide an improved knowledge of the state of the atmosphere, particularly if deployed in the form of a constellation, due to the enhanced temporal, spatial, and spectral resolution capabilities that those sensors can provide with respect to present meteorological microwave sounders. This work takes advantage of artificial intelligence (AI), particularly its capability to rapidly and simultaneously process hundreds of channels and retrieve large sets of geophysical parameters, to assess the impact of HyMS in geophysical space. The results presented in this manuscript are expected to contribute to the design of the next generation of microwave sounders, but also to consider the usage of AI to fully exploit the information content provided by these sensors, particularly if deployed in the form of a constellation of satellites.

**Index Terms**—Artificial intelligence (AI), data assimilation, hyperspectral instruments, machine learning, microwave (MW) instruments, remote sensing.

## I. INTRODUCTION

THE present generation of Polar-orbiting operational microwave (MW) sounders such as the advanced microwave

Manuscript received 9 November 2023; revised 5 January 2024; accepted 16 January 2024. Date of publication 22 January 2024; date of current version 8 February 2024. This work was supported by the U.S. Department of Commerce under NOAA/NESDIS the Systems Architecture and Engineering (SAE) Joint Venture Program, under Contract ST133017CQ0058/1332KP20FNEED0029. (Corresponding author: Eric S. Maddy.)

Eric S. Maddy is with the Riverside Technology, Inc., Silver Spring, MD 20910 USA (e-mail: eric.maddy@noaa.gov).

Flavio Iturbide-Sanchez is with the NOAA/NESDIS/STAR, College Park, MD 20740 USA (e-mail: flavio.iturbide@noaa.gov).

Sid Ahmed Boukabara is with the NASA, Washington, DC 20546 USA (e-mail: sid.boukabara@noaa.gov).

Digital Object Identifier 10.1109/JSTARS.2024.3356858

sounding unit (AMSU) onboard Aqua, the AMSU onboard the European organization for the exploitation of meteorological satellites meteorological operational series, and the advanced technology microwave sounder (ATMS) onboard the Suomi-National Polar-orbiting Partnership, NOAA-20 and NOAA-21, provide critical observations for improving global and regional data assimilation and nowcasting applications [1]. These sensors have demonstrated outstanding performance and when combined with conventional observations such as radiosondes, and aircraft measurements, these platforms serve as a backbone of the Global Observing System for numerical weather prediction (NWP) and other weather and climate applications. Nevertheless, these operational platforms and their suite of instruments have some associated constraints. They are expensive to build, launch, and operate; therefore, these platforms are traditionally flown one at a time; i.e., one satellite in a morning/afternoon polar orbit (9:30 AM/1:30 PM equatorial crossing time).

In the last few years, we have observed the demonstration of an innovative technology in space: the CubeSat MW sounder radiometers. This is the case of the NASA low-Earth orbit (LEO) temporal experiment for storms and tropical systems—demonstration and time-resolved observations of precipitation structure and storm intensity with a constellation of smallsats (TROPICS) pathfinder MW sensors launched on 5 September, 2018 and 30 June, 2021, respectively. NASA followed up with the launch of four more TROPICS CubeSats in May 2023. These technology demonstrations aim at achieving low-cost, high-performance sensor designs, facilitating the implementation of large sensor constellations, with associated deployment flexibility. Due to that, the performance requirements of the new generation of meteorological MW sounders need to be discussed to allow new concepts and capabilities.

Numerous previous studies have shown that sampling the MW spectrum at higher spectral resolution increases the information content of those measurements. For instance, the clear-sky information content of hyperspectral MW concept sensors was assessed in [2], [3], [4], [5], and [6] and found significant improvements in the temperature and moisture retrievals (e.g., 0.5–1 K in temperature, and up to 3× improvements in total water). Aires et al. [7] assessed the all-sky information content with a focus on the selection of channels for NWP data assimilation.

In this manuscript, the assessments of the information content of hypothetical hyperspectral microwave sensor (HyMS) instruments in simulated all-sky conditions using theoretical and empirical approaches are presented. Theoretical assessments utilize the traditional Bayesian analyses of vertical resolution (degrees of freedom for signal) and geophysical product error quantification. For empirical assessments of remote sensing skills, we use the multi-instrument inversion and data assimilation preprocessing system-artificial intelligence (MIIDAPS-AI) to perform an inversion of simulated MW radiances onto geophysical quantities. The performance of temperature, moisture, as well as precipitating and nonprecipitating retrievals are assessed relative to geophysical performance based on the program of record NOAA-20 ATMS instrument observations. In addition, using MIIDAPS-AI we provide assessments of the sensitivity (or resilience) of hyperspectral measurements to radio frequency interference (RFI) and provide quantitative assessments for the spectral stability requirements of said hyperspectral MW instruments.

The article is organized as follows. In Section II, we describe the simulation system used for empirical and theoretical assessments of information content. Section III details the benefits of a hyperspectral MW instrument over the program of record ATMS instrument. Those assessments include temperature, moisture, and hydrometeor sounding capabilities and resilience to RFI. Section IV summarizes the major findings and observations derived from the study reported in this manuscript.

## II. SIMULATION DESIGN

### A. Community Radiative Transfer Model

Developed by the U.S. Joint Center for Satellite Data Assimilation, the community radiative transfer model (CRTM) provides accurate and fast satellite radiance simulations and Jacobian calculations from the surface to the top of the atmosphere (TOA) under all weather and surface conditions [8]. The CRTM is a critical component of NOAA/NCEP's and NASA's data assimilation systems and physical retrieval algorithms such as the operational microwave integrated retrieval system (MiRS) [9]. It is applicable to a wide variety of both real and hypothetical instruments within the MW, infrared (IR), and UV/visible bands at various spectral resolutions and channel centers within those bands.

For the research presented in this article, we have utilized CRTM Release 2.3.0, herein referred to as CRTM, with coefficients spanning the MW spectrum between 1 and 330 GHz at boxcar channel spectral resolutions between 1 and 100 MHz. Table I lists the CRTM coefficients utilized in this study. When combined, the CRTM coefficients comprise 11 143 unique hyperspectral channels between 1 and 330 GHz. In the following, we will refer to this superset of 11 143 channels as the full-resolution HyMS configuration.

As argued in [3], it may be possible to infer cloud and hydrometeor particle effective radii from HyMS spectra, however, in this study, we prescribed particle effective radii as in the MiRS [9]. We also note that the parameterizations of cloud and particle scattering, which are based on Mie-Lorenz theory in the version

TABLE I  
CRTM COEFFICIENTS USED IN THIS STUDY

Frequency Start, GHz	Frequency End, GHz	Bandwidth, MHz
1	330	100
23	24	10
52	58	1
58	62	10
114	120	10
173	183.31	10

of the CRTM used in this study, are known to limit the accuracy of CRTM all-sky calculations [10], [11]. In the future, we plan on utilizing updated parameters and parameterizations as described in [10] and assess the ability of hyperspectral MW sounders to retrieve cloud microphysical properties such as effective radius and/or particle habit.

### B. ECMWF83 Dataset

In this study, we use the European Centre for Medium-Range Weather Forecasts (ECMWF) 83 profiles [12], [13] to estimate the information content of HyMS over a representative range of atmospheric conditions. Herein referred to as the ECMWF83 dataset, these profiles build upon the thermodynamic initial guess retrieval (TIGR) dataset experience and represent a diverse set of atmospheric temperature, water vapor, surface, and trace gas conditions at a high vertical resolution. Full-resolution HyMS configuration simulations for clear-sky and all-sky ECMWF83 geophysical profiles are shown in Fig. 1. The methodology to perform all-sky simulations using the ECMWF83 is described in the following section.

### C. NOAA Finite Volume Cubed-Sphere Global Forecasting System Dataset

In this study, we utilize 20 days in 2022, spanning a little over 1 year, of global GFS model output (<https://registry.opendata.aws/noaa-gfs-bdp-pds/>) colocated to satellite instrument footprints. Satellite instrument geolocation and pointing parameters were pulled from operational NOAA-20 ATMS observations for the corresponding dates in 2020 and 2021. To reduce the total number of cases, we pulled a random subset of observations (~5% of the full resolution data) for each of four 6-hour windows around 00, 06, 12, and 18 z. In total, a robust set of 2 529 788 global colocations covering all seasons were made between ATMS observation locations and GFS profiles.

The NOAA GFS dataset serves two purposes in this study. First, the dataset is used to assess the information content of temperature, moisture, cloud, and hydrometeor retrievals in spatially and temporally representative all-sky and all-surface conditions. Second, because the ECMWF83 dataset includes only clear-sky profiles (i.e., no clouds or hydrometeors are reported), we utilized the colocated GFS dataset to develop a look-up table of temperature, moisture, cloud liquid water (CLW), graupel water, and rainwater profiles in order to account for all-sky conditions. To find cloud and hydrometeor profiles most consistent with the ECMWF83 temperature and moisture

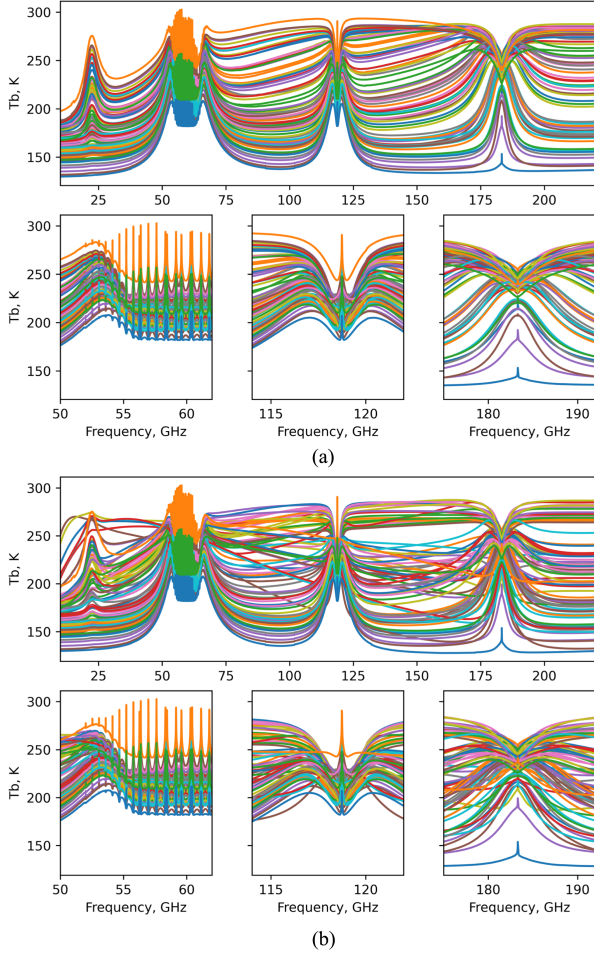


Fig. 1. Example (a) clear-sky and (b) all-sky high spectral resolution simulations using the CRTM and the ECMWF83 dataset. Bottom panels show the features at the 50, 118, and 183 GHz microwave absorption lines. Details regarding the all-sky simulations are presented in Section II-C.

profiles, we first remove any profiles where the integrated cloud and hydrometeor amounts are less than 0.025 mm. We then define a figure of merit based on the Euclidean distance between each of the ECMWF83 temperature and moisture profiles and GFS temperature and moisture profiles as in (1)

$$\Delta x^{i,j} = \overline{\left(T_{\text{ECMWF}}^i - T_{\text{GFS}}^j\right)^2 + \left(Q_{\text{ECMWF}}^i - Q_{\text{GFS}}^j\right)^2}. \quad (1)$$

In (2),  $T_{\text{ECMWF}}^i$  and  $Q_{\text{ECMWF}}^i$  are the  $i$ th ECMWF83 temperature and moisture profiles,  $T_{\text{GFS}}^j$  and  $Q_{\text{GFS}}^j$  are the  $j$ th GFS temperature and moisture profiles, and the overbar represents the average overall pressure levels from the TOA down to the surface level. For each of the ECMWF83 profiles, we append the GFS profile clouds and hydrometeors from the collocation dataset where the Euclidean distance between GFS and ECMWF83 is a minimum, i.e., for each  $(T_{\text{ECMWF}}^i, Q_{\text{ECMWF}}^i)$  we select the  $j'$ -th GFS profiles of cloud and hydrometeors, where  $j' = \arg \min_j \Delta x^{i,j}$ . We note that we obtained similar profiles using the L1 or Manhattan distance between profiles. Simulations resulting from the Euclidean distance minimization method described above are shown in the right column of Fig. 1.

### III. BENEFITS OF A HYMS INSTRUMENT

In the following, Section III-A is dedicated to presenting our methodologies to assess and optimize the information content of hyperspectral MW observations from a theoretical perspective. In particular, the information content with respect to MW spectral absorption features observed between 10 and 220 GHz, spectral resolution in those bands, instrument noise, and retrieved state vector elements are presented. In Section III-B, we apply AI to our optimized HyMS sensor observations to assess; 1) temperature, moisture, and hydrometeor sounding in simulated all-weather, all-surface conditions; 2) the resilience/sensitivity to RFI interference; and 3) the sensitivity to spectral channel centroid uncertainty. All assessments are performed using the program-of-record ATMS instrument as a benchmark.

#### A. Information Content Analysis and HyMS Channel Selection

In order to reduce the number of channels used for the assessment, we utilized the channel selection methodology described in [14], [15], [16], [17], and [18] and using the degrees of freedom for the signal to assign incremental importance to each HyMS channel to the inversion process. The degrees of freedom for signal, or  $dfs$ , is a dimensionless quantity and a measure of the number of independent pieces of information that can be estimated from a measurement. The  $dfs$  is defined as the trace of the averaging kernel matrix,  $A$ , as

$$dfs = \text{trace}(A) = \text{trace}\left(\left[K^T S_\epsilon^{-1} K + S_a^{-1}\right]^{-1} K^T S_\epsilon^{-1} K\right). \quad (2)$$

In (2),  $K$  is the Jacobian of the observation operator produced by the CRTM;  $S_\epsilon = S_e + S_f$  is the observation covariance matrix, which includes both the observation noise (NEDT) ( $S_e$ ) and the forward model uncertainty ( $S_f$ );  $S_a$  is the background atmospheric covariance and superscripts  $T$  and  $^{-1}$  are, respectively, the matrix transpose and inverse.

For the purpose of selecting channels, we compose the atmospheric state vector, background covariance,  $S_a$ , and Jacobian,  $K$ , using temperature and moisture profiles from TOA to the surface, cloud and hydrometeor profiles, and surface temperature. Since the magnitude of the  $dfs$  is a function of the interparameter correlation and standard deviation of the background covariance, we utilized two background covariances for our experiments to account for different applications. The first, denoted  $S_a^{(1)}$ , was defined using the operational MiRS background covariance. The MiRS background constraint is computed relative to a global, all-conditions mean value so the variances and correlations in  $S_a^{(1)}$  are quite large and, therefore, the total estimated  $dfs$  are also quite large. To assess the information content corresponding to more subtle errors in the background constraint, i.e., retrievals that might utilize short-term forecast information from an NWP model, the second background covariance assessed, denoted  $S_a^{(2)}$ , used the interparameter correlation from  $S_a^{(1)}$ , rescaled by the standard deviation of differences between 12 and 6 h forecasts valid at the same in the NOAA GFS model for 1 month of data. This procedure of using forecast model differences to compute background perturbations is sometimes referred to as the National Meteorological Center method [18].

TABLE II  
CHANNEL SELECTION BANDS CONSIDERED, BANDWIDTH CONFIGURATIONS  
ASSESSED, AND NOMINAL NEDT PER SQUARE ROOT OF BANDWIDTH

Frequency Start, GHz	Frequency End, GHz	Bandwidth Configurations, MHz	NEDT/sqrt(MHz), K
50	62	1, 10, 20, 50	5.4
114	121	10, 20, 50	10.0
172	192	10, 20, 50, 100	15.0
10	220	200	Variable

For our study, the instrument covariance,  $S_e$ , was prescribed as a diagonal matrix with elements proportional to the square root of the bandwidth in megahertz. Baseline noise levels in the Ku/K/Ka [4.2 K/sqrt(MHz)], V [5.2 K/sqrt(MHz)], W [10 K/sqrt(MHz)], G [5 K/sqrt(MHz)], and D [15 K/sqrt(MHz)] bands follow [4] and/or are consistent with current on-orbit MW sounder sensors such as ATMS.  $S_f$  was assumed diagonal and constant with a 1 K<sup>2</sup> magnitude across the full MW spectrum. We note that since we expect that many HyMS future technologies utilize a digital backend to process the measured signals, we might expect some noise correlation between channels (e.g., due to apodization in the channelization of hyperspectral channels) and future research should address this in the selection of channels and assessment of the information content of HyMS sensors.

Computing the incremental  $df_s$  over tens of thousands of full-resolution HyMS configuration channels requires significant computational resources and time to loop over all channel combinations. To reduce the number of calculations required, we performed an incremental selection of channels focusing on four sounding and surface-sensitive bands and at bandwidths spanning the full-resolution HyMS spectrum configurations. Table II lists the different experiments performed for each of the HyMS instrument bands and bandwidths assessed. In addition, to further reduce the number of channels selected in any one band, we restricted our channel selection to select the first 200 most informative channels in each of the bands. In this channel selection process, it was found that 200 channels were sufficient to explain most of the total information content in each of the bands and bandwidths assessed; that is, the  $df_s$  asymptotes to a nearly constant value when additional channels were added. The information content was computed at the highest spectral resolution and then subsequently computed by averaging the high spectral resolution channels over the spectral bandwidths shown in the third column of Table II. An assessment of the information content of HyMS at the various bandwidth configurations was then performed.

The channel selection methodology described above was performed for each profile in the ECMWF83 database under clear-sky and all-sky simulated conditions, for land and ocean-like prescribed emissivities ( $e = 0.9$ ,  $e = 0.6$ ), at nadir and off-angle (zenith = 45°), and using both  $S_a^{(1)}$  and  $S_a^{(2)}$ .

For the temperature and moisture sounding bands at 50, 118, and 183 GHz, the maximum  $df_s$  averaged over all 83 profiles was inspected for each of the bandwidth resolutions assessed. At 50 GHz, with a maximum of 200 channels within the spectral

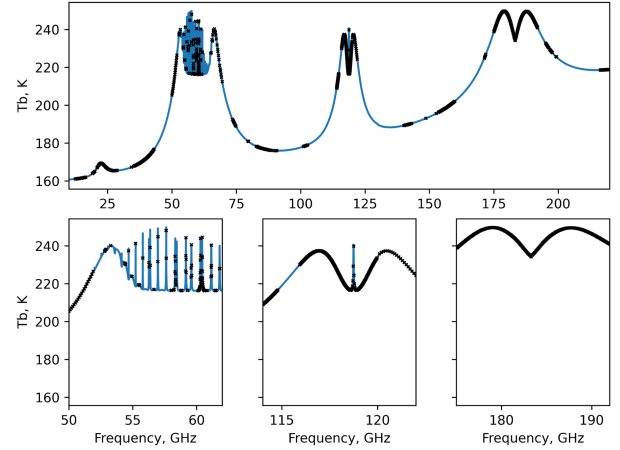


Fig. 2. Full resolution HyMS configuration simulation using the concatenation of all CRTM coefficients (blue) and selected channels using the degrees of freedom methodology (black x's).

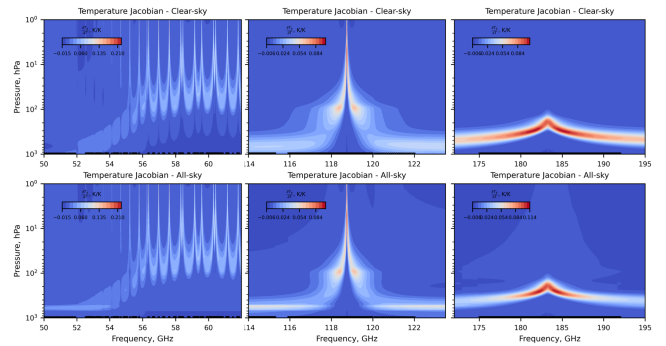


Fig. 3. Full resolution HyMS temperature profile Jacobians for an ECMWF83 Profile 1 at a nadir-viewing angle in the 50, 118, and 183 GHz sounding regions. The top and bottom row correspond to clear-sky and all-sky simulations, respectively. Location of selected channels shown in the figure by the black x's near to the surface.

interval between 50 and 62 GHz, small improvements in  $df_s$  were noted between 1 and 10 MHz experiments, indicating that increasing the spectral resolution finer than 10 MHz would yield diminishing returns in total information content. This is not surprising given that with the exception of the channels peaking within the upper stratosphere, current state-of-the-art sensors such as ATMS use channels with bandwidths much greater than 5 MHz to sound the single oxygen line at 57.29 GHz. At 118 and 183 GHz, a similar behavior was observed between 10 and 20 MHz for the 118 GHz oxygen line and between 50 and 100 MHz for the 183 GHz water vapor line. Channels within the atmospheric window were then selected using the broader band 200 MHz simulations excluding the major sounding bands in the channel selection. The results were then concatenated over all atmospheric and surface conditions and unique channels were chosen over all channel selection experiments resulting in a total of approximately 1100 channels spanning between 1 and 220 GHz.

Fig. 2 shows the full-resolution HyMS spectrum of 11 143 channels and the locations of the final selection of 1100 channels. Figs. 3 and 4 show the temperature and moisture

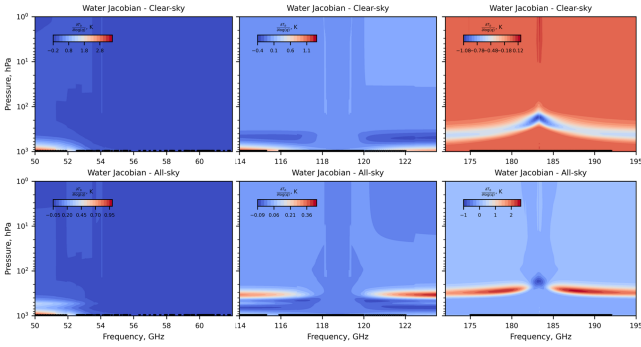


Fig. 4. Full resolution HyMS water profile Jacobians for an ECMWF83 Profile 1 at a nadir-viewing angle in the 50, 118, and 183 GHz sounding regions. The top and bottom rows correspond to clear-sky and all-sky simulations, respectively. Location of selected channels shown in the figure by the black x's near to the surface.

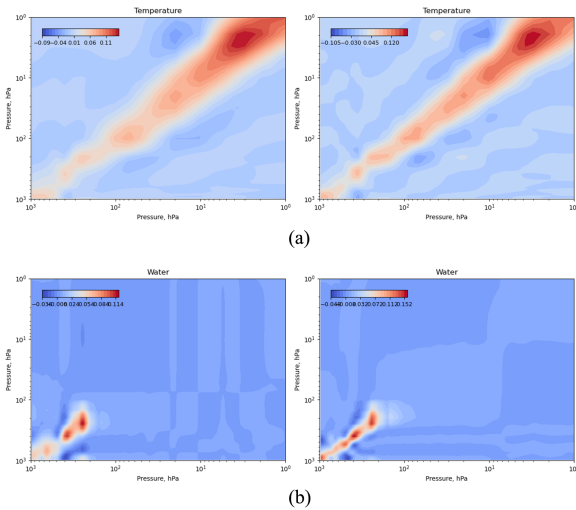


Fig. 5. Comparison of ATMS and HyMS temperature and moisture averaging kernels. Panel (a) shows the ATMS temperature averaging kernel (left) and the HyMS selected channel temperature averaging kernel (right). Panel (b) shows ATMS water vapor averaging kernel (left) and the HyMS selected channel water vapor averaging kernel (right). All computations use ECMWF profile 1 and  $S_a^{(1)}$ .

Jacobians for ECMWF83 profile 1 in each of the temperature and moisture sounding bands at 50, 118, and 183.1 GHz. In this case, the temperature Jacobian is defined as a derivative of the simulated channel brightness temperature per unit change in a layer temperature, and the water vapor Jacobian is defined as the derivative of the simulated channel brightness temperature to a unit logarithmic change in a layer water vapor amount. In general, the optimal channel selection shows good coverage at line centers and in between lines, which sense higher and lower in the atmosphere, respectively, as well as within the interstitial window regions between the three main sounding bands. In addition, the weak water vapor line at 22.23 GHz is sampled almost completely by our channel selection.

Example temperature and water vapor averaging kernels for ATMS and the HyMS selected channels and using  $S_a^{(1)}$  are shown in Fig. 5. As in (2),  $A$  is related to the  $dfs$  and describes the sensitivity of a retrieval to true changes in the atmospheric state

vector. Because the numerator and denominator of the averaging kernel are of the same units,  $A$  is a unitless measure of the information content of retrieval for each state vector element. Ideally,  $A$  would be the identity matrix indicating that a unit change in the true atmospheric state would yield an identical change in the retrieval for every layer and furthermore indicating that the vertical resolution of the retrieval is equal to the vertical resolution of the state vector elements. While falling short from being equal to the identity, it is clear from Fig. 5 that the HyMS averaging kernels for temperature are tighter (less correlation between state vector elements) around the diagonal as compared to ATMS, and the tighter correlation is reflected in the larger  $dfs$  for HyMS. Over the ECMWF83 dataset, HyMS temperature  $dfs$  ranges between 7.4 and 8.8 units, while for moisture HyMS  $dfs$  ranges between 2 and 5.3 units. Comparatively ATMS,  $dfs$  ranges between 5.4 and 6.3 units for temperature and 1.1 and 3.3 for moisture. Over all cases, HyMS temperature  $dfs$  is usually two units larger than ATMS and HyMS water vapor  $dfs$  is 1.7 units larger than ATMS. In addition to the increased  $dfs$ , the predicted uncertainty of the HyMS temperature and moisture retrievals, based on the square root of the diagonal of the predicted retrieval covariance matrix (not shown) are reduced by roughly 0.25–0.5 K and 5%–10% throughout the entire atmospheric column (and troposphere for water vapor), as is reported in Section III-B. Therefore, the finer spectral resolution and increased number of channels of the HyMS instrument significantly increases the information content relative to ATMS for sensing atmospheric temperature and moisture. Although no graphical results of the averaging kernels are presented, similar results are obtained using the rescaled MiRS background covariance  $S_a^{(2)}$  and the same HyMS selected channel for temperature and water vapor. In this case, the computed HyMS  $dfs$  are 1.6 and 1.8 units greater than ATMS, which is comparable to the results presented in Fig. 5. Similar to the results using  $S_a^{(1)}$ , this increase in the information of HyMS retrievals results in temperature and moisture retrievals with reduced predicted uncertainties of approximately 0.25–0.75 K and 5%–10%. Details of the impact of HyMS in the thermodynamic products are reported in Section III-B.

In order to further characterize the information content of HyMS and its impact on the different atmospheric layers relative to ATMS, Fig. 6 shows eigenvectors,  $u_i$ , scaled by corresponding eigenvalues,  $\lambda_i$ , of the mean temperature and moisture averaging kernels computed over all 83 cases and using  $S_a^{(1)}$ . For each eigenvector, the eigenvalues and corresponding fraction of total variance,  $\frac{\lambda_i}{\sum_{i=1}^n \lambda_i}$ , are also shown in the legend. As compared to ATMS, the HyMS-scaled eigenvectors are generally larger near the surface layers and the associated variance (representing the level information content in the form of the variability in the observations associated with changes in the state of the atmosphere) is spread in more eigenvectors. For instance, for temperature, the 7th and 8th HyMS eigenvalues represent 7% and 4% of the total variance while the corresponding ATMS eigenvectors represent 4% and 2%, respectively. Turning to moisture, the 5th and 6th HyMS eigenvectors represent 5% and 2% of the total variance, while the corresponding ATMS

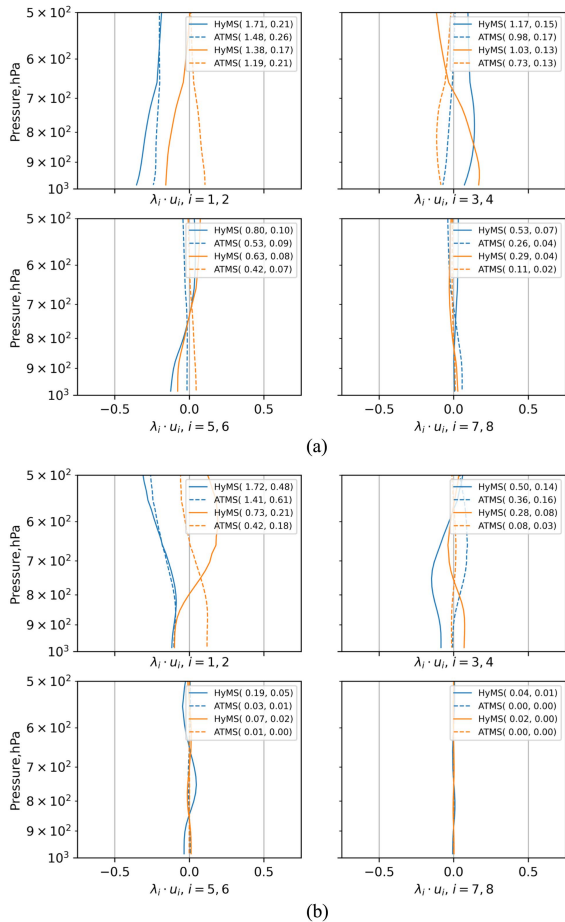


Fig. 6. First eight HyMS and nominal ATMS eigenvectors,  $u_i$ , scaled by corresponding eigenvalues,  $\lambda_i$ , of the mean temperature (a) and moisture (b) averaging kernels over all 83 profiles and using  $S_a^{(1)}$ . Eigenvalue and corresponding fractional total variance are shown in the legend for each eigenvalue pair.

eigenvectors represent  $<1\%$  of the total variance. This characteristic suggests the improved capability to resolve finer vertical resolution changes in the atmosphere throughout the troposphere and in the layers closest to the surface for both temperature and moisture. The shape of the eigenvectors also suggests that HyMS retrieval sensitivity is highly correlated with the rest of the tropospheric column, since no discontinuities or sharp transitions are observed. The 2nd eigenvector of the water vapor  $A$  also reveals the ability of HyMS to provide more information in the mid-troposphere; albeit strongly coupled with moisture between 800 hPa surface.

### B. MIIDAPS-AI: HyMS Geophysical Assessment in All-Sky Conditions

Fundamentally, the multi-instrument inversion and data assimilation preprocessing system, AI version, or MIIDAPS-AI for short, is a deep fully connected neural network that defines a nonlinear mapping between instrument measurements (IR radiances and/or MW brightness temperatures) and geophysical parameters such as temperature and moisture profiles, cloud and hydrometeor profiles and integrated amounts (e.g., CLW

and ice water path), and spectral surface emissivities [19]. As an extension of traditional 1-D variational approaches such as the MiRS [9], (and its extension to IR satellite observations, MIIDAPS, [20]), MIIDAPS-AI has been successfully applied to IR and MW polar and GEO sounders and imagers and is valid for any sensor with valid CRTM coefficients [21]. MIIDAPS-AI can additionally serve as a preprocessor for data assimilation and data fusion [22].

For training, testing, and independent validation of MIIDAPS-AI models for HyMS and ATMS, we follow the procedures described in [19]. All-sky, all-surface simulations using collocations between ATMS orbital parameters and GFS dataset are split into three groups: 1) training, which consists of 52 random 6-hour windows during 2020-01-01 and 2020-12-21 for a total of 1 575 943 samples; 2) testing, which consists of 19 random 6-hour windows over the same period for a total of 661 756 samples; and 3) validation, which consists of 2020-12-21 15z - 23z, 2021-11-11, 2021-11-21 for a total of 292 089 samples. The robust training dataset is used for optimization of neural network weights, the testing dataset serves to select the best model during optimization, and the validation dataset is used for independent assessment of results in the following sections.

In this study, the MIIDAPS-AI output retrieval state vector includes surface temperature as well as temperature, moisture at 33 pressure layers between 0.1 hPa and the surface, and path-integrated cloud, rain, and graupel profiles in 16 pressure layers between 100 hPa and the surface in order to account for all-sky conditions. The inputs to MIIDAPS-AI include instrument brightness temperatures, viewing angle geometry, one-hot encoded land-sea mask, forecast surface pressure, and sine and cosine of the UTC hour of the day. Models are trained using a mean-absolute-error loss between predicted and target collocated GFS profiles and the best weights are selected using the minimum loss as computed using the testing dataset over a series of iterations. As shown in [19], partly because MIIDAPS-AI is a statistical method, based on a large number of training and testing profiles, the algorithm may predict a solution with performance better than that estimated using a Bayesian inversion and using a static background covariance.

In the following section, we present the results of fully trained MIIDAPS-AI models for the ATMS and the notional HyMS instrument (using the channels selection methodology described in Section III-A) on the completely independent validation dataset.

1) *Temperature and Humidity Sounding in All-Sky Conditions*: As mentioned in the introduction, previous studies have shown that HyMS-like instruments show an improved ability to perform temperature and moisture sounding throughout the atmospheric column. Most of those studies have focused on clear-sky and ocean-only conditions. In the following, we perform a comparative assessment of the gold-standard ATMS sounder versus the notional HyMS instrument using all-sky, all-surface simulations over the entire independent validation dataset.

One such assessment is shown in Fig. 7, where we have shown profiles of temperature and moisture bias and standard deviation (MIIDAPS-AI minus true GFS) statistics for HyMS and ATMS over the entire independent validation dataset. We note that both

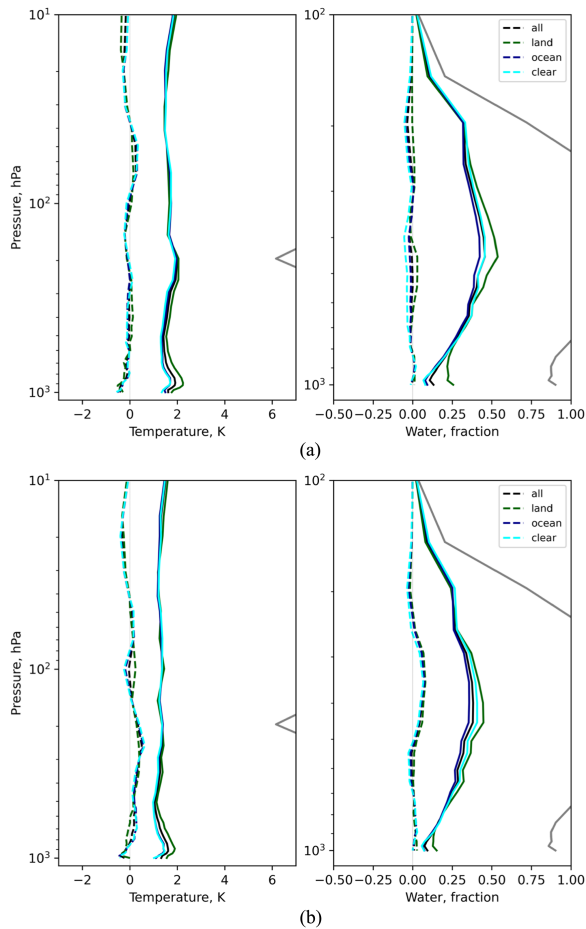


Fig. 7. MIIDAPS-AI temperature (left) and water vapor (right) profile statistics for ATMS (a) and notional HyMS (b) with optimal channel selection for all (black), land (green), ocean (blue), and clear-sky (cyan) cases in the independent validation dataset compared against the GFS model. The mean bias and standard deviation between MIIDAPS-AI and the true simulated profiles are shown as the dotted and solid lines, respectively.

the HyMS and the ATMS retrievals are reported for 100% of all-surface/all-weather cases (i.e., the yield of both MIIDAPS-AI based retrievals is 100% and includes all cases with conditions spanning clear to cloudy and precipitating as well as ocean, land, sea-ice, and snow surfaces).

Compared to ATMS, HyMS shows improved sounding near the surface to TOA by 0.25–0.5 K. HyMS water vapor sounding is improved over the entire vertical column with 5%–10% reduced RMSE near the surface and throughout the troposphere compared to ATMS. This reduction in random and systematic errors between HyMS and ATMS is similar to the theoretical results obtained in Section III-A using the Bayesian/variational approach (not shown). In Fig. 7, we also separate statistics for land, ocean, and clear sky conditions. The results for ocean cases are similar to the all cases results that are described above; however, over land, the HyMS assessment shows a reduction of temperature errors relative to ATMS, and more dramatically for water vapor and near the surface.

Figs. 8 and 9 show the consistency between empirical temperature and water vapor retrieval assessments to the theoretical

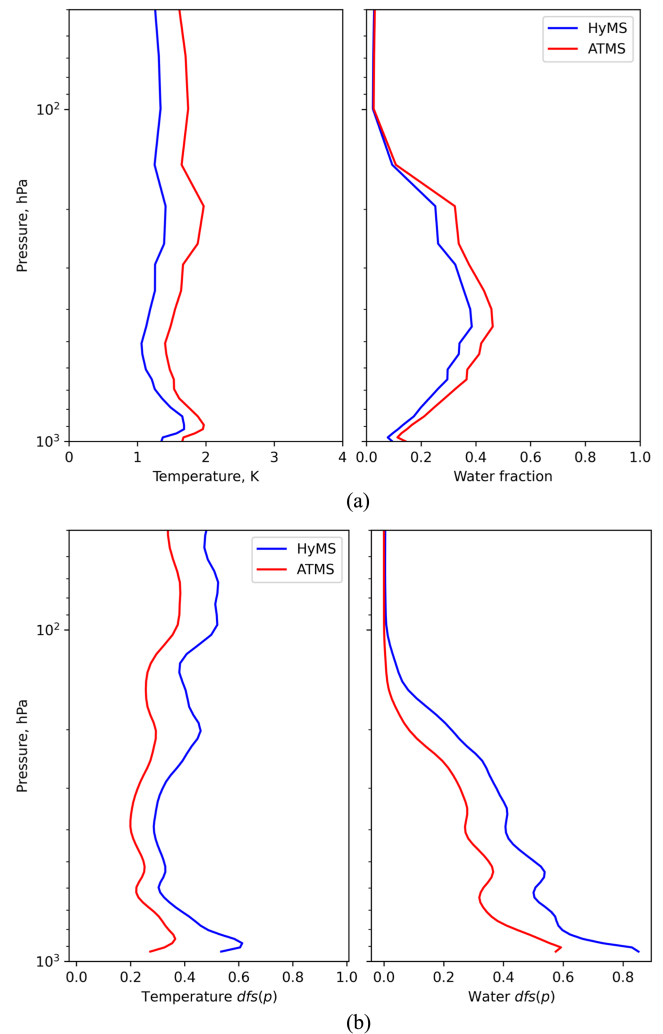


Fig. 8. (a) Comparison of summary ATMS and HyMS all-weather vertical temperature (left) and moisture (right) RMSE statistics (reproduced from Fig. 7). (b) Profiles of the average degrees of freedom for signal for temperature (left) and water vapor (right). Profiles of the degrees of freedom for signal are the diagonal elements of the averaging kernel matrix and are computed as described in Section III-A.

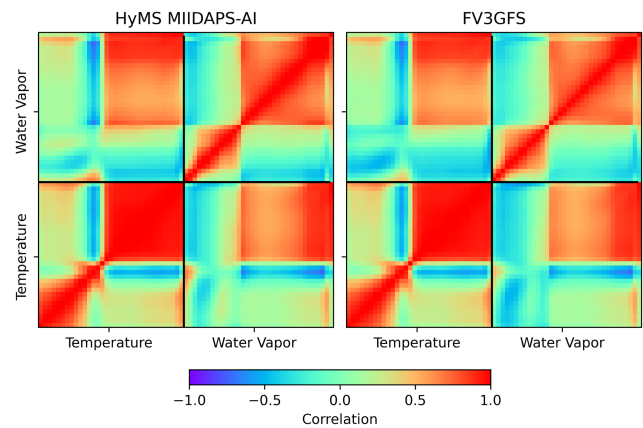


Fig. 9. Interparameter correlation between MIIDAPS-AI (left) and FV3GFS (right) temperature and water vapor retrieval states computed over the entire validation dataset. The correlation matrices are ordered such that the pressures toward the TOA are nearest to the left of the figure.

assessments shown in Section III-A. In the top panels of Fig. 8, we show a summary MIIDAPS-AI temperature and water vapor profile RMSE statistics for all cases, for HyMS and ATMS (reproduced from Fig. 7), and in the bottom panels we show the average degrees of freedom for signal profiles (diagonal elements of the averaging kernels) as computed over all 83 profiles and using  $S_a^{(1)}$ . As is evident from Fig. 8, the observed improvements in the statistical performance of HyMS as compared to ATMS over the entire tropospheric column are consistent with the theoretical information content analysis based on  $df$ s, which shows HyMS  $df$ s greater in magnitude than ATMS  $df$ s. Said another way, larger  $df$ s results in finer vertical resolution temperature and moisture retrievals and that increased  $df$ s results in an improved ability to sound the atmosphere, thus reducing the RMSE between MIIDAPS-AI and the true atmospheric state. This behavior indicates that the MIIDAPS-AI algorithm, while statistical in nature, provides retrievals that are consistent with physics-based information content. Fig. 9 shows that the atmospheric temperature and water vapor profiles are correlated to each other in roughly the same manner as these parameters are found to be correlated in the true correlative NWP fields of finite volume cubed-sphere global forecasting system (FV3GFS). As would be expected from an inversion of atmospheric profiles from an MW sounder with finite vertical resolution, the structure of the vertical correlation of MIIDAPS-AI temperature and moisture is similar, but shows some differences as compared to FV3GFS (e.g., in the temperature/moisture correlation block in Fig. 9). This behavior is similar to the correlations presented for MIIDAPS-AI as applied to NOAA-20 ATMS and CrIS in [19].

These results demonstrate the increased information content of HyMS over ATMS for both temperature and moisture resulting from the finer vertical resolution of the hyperspectral instrument in all-sky conditions. In addition, these results demonstrate that the machine-learning based retrieval algorithm used in this study, namely MIIDAPS-AI, produces retrieval states that are physically consistent with physics-based theoretical information content analyses.

2) *Cloud and Hydrometeor Sounding*: Along with temperature and moisture profiles, cloud (liquid) and hydrometeor (rain, graupel/ice) profiles at 16 levels between 100 hPa and the surface were included in the MIIDAPS-AI state vector, and neural network models were trained for both the notional HyMS and ATMS. Cloud and hydrometeor profiles are not continuous in the vertical, and we found that it was difficult to train networks with profiles of cloud and hydrometeor layer amounts. To address this difficulty, we modified the MIIDAPS-AI state vector to be composed using path-integrated (from TOA to layer) cloud and hydrometeor profiles,  $\hat{Q}_p$ , as

$$\hat{Q}_p = \sum_{i=1}^p Q_p, \quad p = 1, 2, \dots, p_{\text{surface}} \quad (3)$$

where  $Q_p$  is the cloud or hydrometeor column (in mm) at pressure layer  $p$ .

The top panels in Figs. 10 and 11 show vertical profile statistics of the MIIDAPS-AI cloud and hydrometeor retrievals for ATMS and HyMS, respectively, over land, ocean, and all

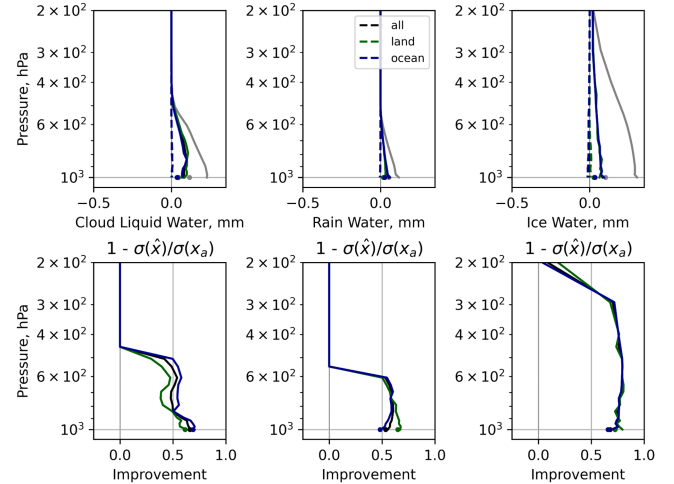


Fig. 10. MIIDAPS-AI path integrated liquid water cloud, rainwater, and graupel+ice water statistics for ATMS for all (black), land (green), ocean (blue) cases in the independent validation dataset. The mean bias and standard deviation between MIIDAPS-AI and the true simulated profiles are shown as the dotted and solid lines, respectively. The standard deviation of validation profiles is shown in grey. Top panels show the absolute bias and standard deviation. Bottom panels show the improvement in the MIIDAPS-AI estimation of cloud and hydrometeor relative to the standard deviation of the validation profiles.

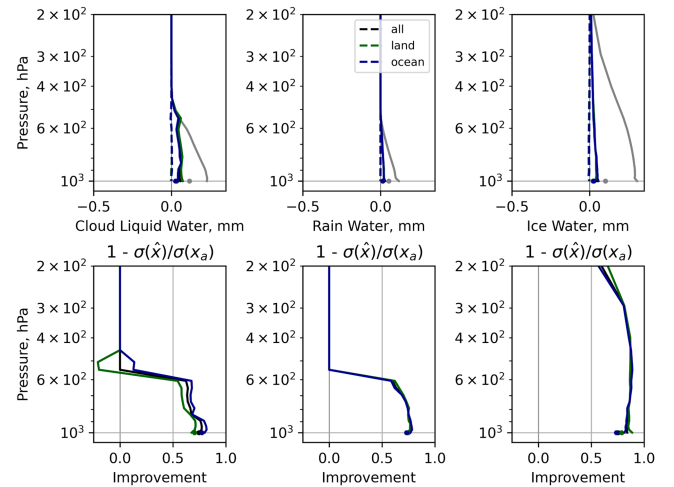


Fig. 11. MIIDAPS-AI path integrated liquid water cloud, rain water, and graupel+ice water statistics for the notional HyMS sensor for all (black), land (green), ocean (blue) cases in the independent validation dataset. The mean bias and standard deviation between MIIDAPS-AI and the true simulated profiles are shown as the dotted and solid lines respectively. The standard deviation of validation profiles is shown in grey. Top panels show the absolute bias and standard deviation. Bottom panels show the improvement in the MIIDAPS-AI estimation of cloud and hydrometeor relative to the standard deviation of the validation profiles.

surfaces. For reference, the standard deviation of the validation dataset is also shown in the top panels. The bottom panels of the two figures show the relative improvement in the MIIDAPS-AI retrievals over the validation dataset. The relative improvement (or retrieval error reduction) in this case is defined as one minus the ratio of the MIIDAPS-AI retrieval standard deviation relative to the truth (retrieval error), denoted by  $\sigma(\hat{x})$ , to the standard deviation of the background (maximum expected error), denoted by  $\sigma(x_a)$ . In this case, values closer to 1 indicate superior



performance, since the retrieval standard deviation (error) tends to be smaller than the maximum expected variance, given by the standard deviation of the validation dataset.

For each of the three cloud and hydrometeor types, the HyMS retrievals generally show reduced standard deviation compared to ATMS. For instance, averaged over the entire retrieved profile from 100 hPa to the surface, HyMS cloud, rain, and ice retrieval root-mean-squared errors (RMSE) relative to the true GFS profiles are reduced by 18.2%, 37.1%, and 36.6% relative to ATMS. Between 700 hPa and the surface, the HyMS cloud, rain, and ice retrieval RMSE are reduced by 28.7%, 39.7%, and 30.8% relative to ATMS. These results show that the new information provided by HyMS constrains the retrieved solution to produce smaller errors compared to ATMS.

We would expect this behavior relative to ATMS, since HyMS observations are able to resolve finer vertical structure for temperature and moisture, including a continuous sampling of higher frequencies which capture more of the scattering signal of frozen water and also include more channels within the window regions which enables the retrieval to distinguish the spectral signatures of cloud and hydrometeors.

3) *Resilience to 5G RFI*: RFI is expected to increase with the roll-out of 5G and future telecommunication technologies. RFI is currently a known error source in the water vapor absorption band around the 23 GHz spectral region and is also expected in the 50 GHz region [4]. Current MW sensors such as ATMS and AMSU are affected to some degree by RFI and these sounders, in particular, are limited by channel frequencies and bandwidths fixed at launch [23]. In this section, we seek to answer whether HyMS technology can potentially increase the resilience of MW sounder remote sensing algorithms to RFI via redundancy in the information content within and outside of spectral ranges affected by RFI.

In this study, for every observation in our validation dataset, our RFI model is implemented using the following:

$$\widehat{T_b}(\nu_1, \nu_2) = T_b(\nu_1, \nu_2) + \text{mask} * r * \Delta T_b^{r,fi}(\nu_1, \nu_2) \quad (4)$$

where  $\widehat{T_b}(\nu_1, \nu_2)$  are the full-resolution HyMS RFI-affected brightness temperatures between frequencies;  $(\nu_1, \nu_2)$  and  $T_b(\nu_1, \nu_2)$  are the unaffected brightness temperatures within that spectral interval; mask is a binary flag determined from a spatially averaged ( $0.5^\circ \times 0.5^\circ$ ) RFI mask built using SMAP L1b (<https://urs.earthdata.nasa.gov>) flags over the month of May 2023,  $r$  is a random uniform deviate  $U[0, 1)$  or constant equal to 1, and  $\Delta T_b^{r,fi}(\nu_1, \nu_2)$  is the magnitude of RFI over the spectral interval. In this study, we set  $\nu_1 = 23.75$  GHz and  $\nu_2 = 24.75$  GHz and vary  $\Delta T_b^{r,fi}(\nu_1, \nu_2)$  between 0 and 50 K. For each observation in the validation dataset, we interpolate the RFI mask to the observation location and if the interpolated mask value is  $> 1$ , we set mask = 1; otherwise, we set mask = 0. Because the strength of RFI at a given location can vary considerably as a function of the time of day, satellite viewing geometry, and many other factors, we also implemented our model in two ways. The first adds a constant RFI level to each location where mask = 1, i.e.,  $r = 1$ . The second model scales the RFI level at each location where the mask = 1 to be between zero and one using a random uniform deviation for each sample; i.e.,

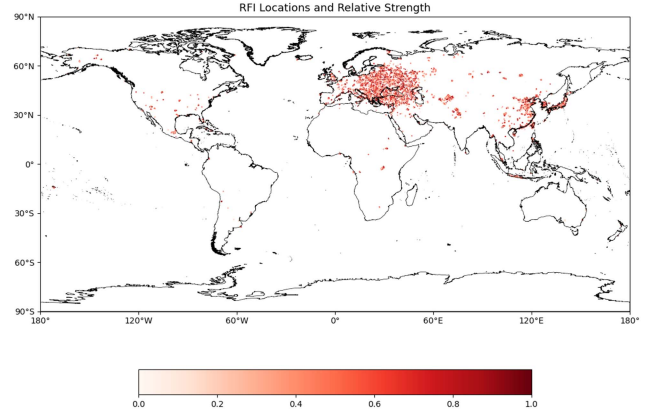


Fig. 12. RFI locations and relative strength using a random uniform deviate to scale the magnitude of the RFI ( $\Delta T_b^{r,fi}$ ).

$r \sim U[0, 1)$ . Fig. 12 shows the RFI mask and the normalized relative strength using the second model.

Fig. 13 shows an assessment of the effect of RFI on temperature and moisture retrievals using MIIDAPS-AI and ATMS and HyMS observations over the validation dataset. In Fig. 13, we show the RMSE in ATMS and HyMS MIIDAPS-AI coarse tropospheric layer temperature and water vapor retrievals as a function of RFI level,  $\Delta T_b^{r,fi}(\nu_1, \nu_2)$ , at  $24.25 \pm 0.5$  GHz spectral region. For each RFI level, we used the MIIDAPS-AI model that was trained without any RFI contamination and applied that MIIDAPS-AI model to RFI affected observations with various strengths.

As expected, as the strength of the RFI level increases, MIIDAPS-AI retrievals show increasing RMSE for both temperature and moisture and ATMS is affected significantly more by  $\Delta T_b^{r,fi}(\nu_1, \nu_2) > 5$  K. ATMS observations include a single broad channel at 23.8 GHz with 270 MHz bandwidth and that channel has significant overlap with our RFI model spectral range. HyMS on the other hand, includes many channels around the weak water vapor line at 22.23 GHz (see Fig. 2), and additionally, multiple channels throughout the MW spectrum interstitial window regions. The results presented in Fig. 13, clearly show that spectral redundancy makes the HyMS MIIDAPS-AI retrievals more resilient to RFI as compared to ATMS, even when RFI levels approach 50 K.

4) *Radiometric Sensitivity to Channel Centroid Spectral Shifts*: High spectral resolution observations are more sensitive to imperfect knowledge of the spectral response, especially in absorption bands with sharp spectral contrast between the line center and wings of lines. In this section, we perform an assessment of the sensitivity of HyMS temperature and moisture retrievals to shifts in the centroids of spectral response functions with an aim to answer the question: What is the maximum spectral shift we should tolerate in order to keep certain quality on HyMS geophysical products?

Fig. 14 shows the maximum channel-dependent change in brightness temperature due to a spectral shift, denoted  $\Delta\nu$ . For each spectrum, the HyMS instrument channel centroids were shifted by magnitude  $\Delta\nu$ , and the shifted brightness temperatures were computed by convolving the shifted spectral response over the high-resolution simulation. The difference between the

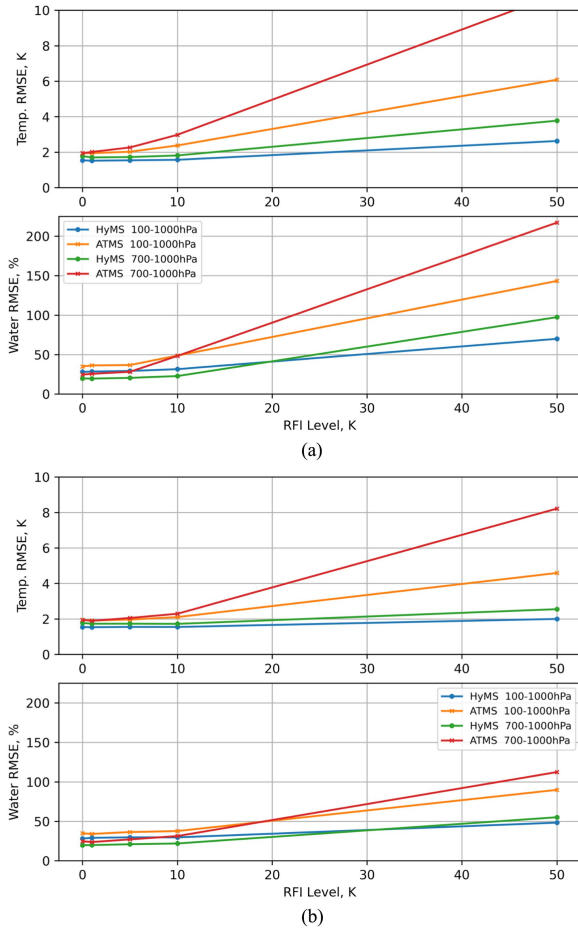


Fig. 13. Assessment of the RMSE in ATMS and HyMS MIIDAPS-AI coarse layer temperature and water vapor retrievals as a function of RFI level,  $\Delta T_b^{rfi}(\nu_1, \nu_2)$ , at  $24.25 \pm 0.5$  GHz spectral region. Panel (a) corresponds to a constant RFI level at each masked observation location. Panel (b) corresponds to a notional RFI level [random uniform distribution with the same maximum amplitude as Panel (a)] at each masked observation location.

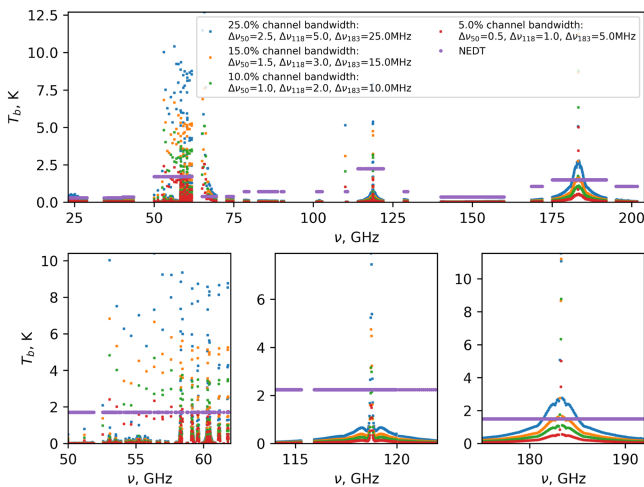


Fig. 14. Magnitude of maximum brightness temperature difference,  $T_b$ , due to a spectral shift,  $\Delta\nu$ , over the entire validation dataset. Shifts computed for constant  $\Delta\nu$  across the HyMS spectrum, and for constant  $\Delta\nu$  in the three sounding bands at (50, 118, and 183 GHz). For reference, spectral NEDT for the variable bandwidths of the HyMS instrument concept is also shown.

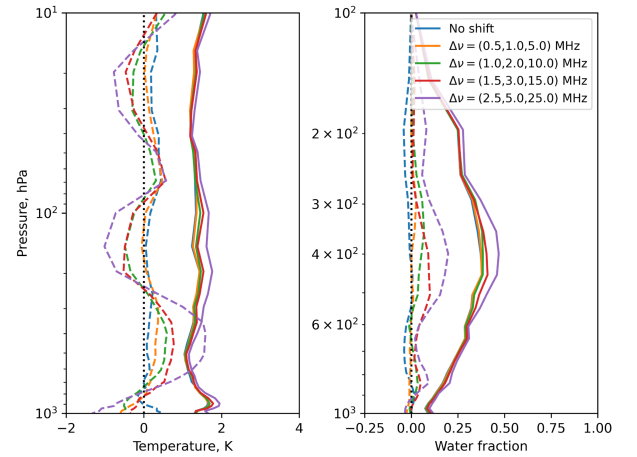


Fig. 15. Assessment of the mean bias and standard deviation of HyMS MIIDAPS-AI temperature and moisture retrievals as a function of spectral shift,  $\Delta\nu$ .  $\Delta\nu$  correspond to roughly 5%, 10%, 15%, and 25% of the bandwidth of the HyMS instrument channels in each of the three sounding bands at 50, 118, and 183 GHz.

shifted spectrum and unshifted spectrum is shown in Fig. 14. For comparison, HyMS instrument NEDT is also shown as we would expect that any remote sensing algorithm (unless trained to be insensitive to spectral shifts) would be susceptible to additional uncertainties in geophysical parameter retrievals due to shifts greater than or roughly equal to the instrument uncertainty. To provide a partial answer to our question, constant spectral shifts across the full HyMS spectrum as well as band-dependent shifts are shown. For the band-dependent shifts, the values in parenthesis correspond to shifts within the sounding bands at 50, 118, and 183 GHz. These band-dependent shifts correspond to 5%, 10%, 15%, and 25% of the bandwidth of the HyMS instrument channels.

Fig. 15 shows the MIIDAPS-AI HyMS temperature and moisture retrieval bias and standard deviation as compared to the true profiles for several spectral shifts. As the magnitude of  $\Delta\nu$  increases, the temperature retrievals show an increasing oscillatory bias in the profile statistics. This oscillatory bias is due to the fact that a spectral shift moves the apparent temperature response of the HyMS channels along the absorption line wings, essentially shifting the retrieved profile along the negative/positive temperature lapse rate in the troposphere/stratosphere. For the largest spectral shifts of  $\sim 25\%$  of the HyMS channel bandwidths, the temperature retrieval bias and standard deviation are both degraded relative to the “no shift” and lesser magnitude shifts. These largest spectral shifts induce positive and negative biases of almost 2 K in the troposphere and -1 K in the stratosphere.

Turning to the moisture results, spectral shifts of less than 15% show minimal impact on the retrieval results. For spectral shifts larger than 15%, both the bias and standard deviation show increasing degradation. For instance, a 15% spectral shift in each of the three bands results in a 10%–15% bias in tropospheric water vapor and a 6% increase in standard deviation relative to the lesser magnitude spectral shifts and “no shift” results. Based on these results, we recommend that spectral stability

should be a requirement for any HyMS-like sounders and on the order of 10% spectral bandwidth, which for our notional HyMS instrument corresponds to 1, 2, and 10 MHz in the 50, 118, and 183 GHz bands.

As mentioned in [19], MIIDAPS-AI Jacobians, which are derivatives of the retrieved geophysical parameters due to changes in the radiometric inputs, exhibit physically consistent spectral responses. These results, which utilize also MIIDAPS-AI for this retrieval sensitivity assessment, also demonstrate that the ML/AI-based algorithm learns not just a statistical projection of radiometric signals into geophysical signals, but MIIDAPS-AI also behaves physically in its temperature and moisture retrieval response to radiometric errors.

#### IV. SUMMARY OF RESULTS AND DISCUSSION

In this manuscript, we performed an assessment of a hypothetical hyperspectral MW sounder, an instrument with 1000s of channels between 1 and 220 GHz, with respect to temperature, moisture, and hydrometeor sounding. Using the incremental theoretical information content of the simulated HyMS sensor observations, we ranked and selected the most informative candidate channels in the MW at bandwidths ranging from 1 to 200 MHz. Moving beyond theoretical information content analyses which require prescriptions of background error correlations and variances, this study demonstrates the value of ML/AI to efficiently exploit the information content of HyMS instruments and quantify its impact. For instance, using MIIDAPS-AI, an efficient and state-of-the-art ML/AI-based remote sensing emulator, our notional HyMS configuration, which possesses 1100 channels, shows improvements in temperature, moisture, and cloud and hydrometeors sounding in all-sky conditions as compared to operational ATMS. This result is expected, especially given that higher spectral resolution and more channels correspond to 1) increased purity of channel information, i.e., the ability to separate sources of errors among channels; 2) redundancy of information, and therefore a noise reduction across spectral lines; 3) observation of new resonances, especially in the 50–60 GHz region; and 4) increased ability to sample far wings/continuum, which corresponds to improved sounding near the surface. Higher quality of retrieved hydrometeor profiles is mostly associated with the measurement of more scattering signals at higher frequencies as well as additional information within the window regions. As we have shown in Section III-A, all of those factors corresponded to increased information content and vertical resolution from the TOA to the surface as compared to state-of-the-art sounders currently flying. We, therefore, expect those near-surface improvements to enhance our understanding of atmospheric processes in all-weather conditions; especially with respect to severe weather and PBL studies.

In addition to the sounding performance assessments in nominal conditions, we have performed analyses of the sensitivity (and resilience) of our candidate HyMS sensor to RFI contamination in the 23 GHz spectral region. The spectral redundancy of our notional HyMS sensor around channels likely to be affected by RFI and improves the resilience of MIIDAPS-AI as applied to the HyMS over MIIDAPS-AI as applied to the program of record ATMS to these types of signals. Since RFI is

expected to increase with the roll-out of future telecommunication technologies (e.g., 6G) and also expected in the 50 GHz region, we plan on extending our analyses to assess those types of signals in a forthcoming manuscript.

Finally, we assess the sensitivity of MIIDAPS-AI as applied to our candidate HyMS sensor to spectral shifts ranging from 5% to 25% of the channel bandwidths. We have shown that HyMS spectral uncertainties greater than 10% bandwidth can introduce significant vertically correlated biases in temperature and water vapor retrievals. Those analyses suggest that the HyMS sensor spectral stability requirement should be <10% of the bandwidth of the measured channels. This corresponds to 1 and 2 MHz in the 50 and 118 GHz oxygen temperature sounding bands and 10 MHz in the 183 GHz water vapor sounding band.

Although not explicitly addressed in this work, we expect HyMS sensor measurements to improve our understanding of fundamental physics/spectroscopy of the MW spectrum. For instance, high spectral MW observations represent a more direct measurement of line widths/strengths and pressure-induced shift of absorption lines. In addition, a more continuous coverage of the MW spectrum, especially in the window regions, should improve our understanding of the water vapor continuum as well as parameterizations of cloud and hydrometeor particle scattering in the MW. Those characteristics of a HyMS sensor should, therefore, improve radiative transfer algorithms (both line-by-line and fast model such as CRTM) accuracy as a result.

#### ACKNOWLEDGMENT

The authors wish to thank Kevin Garrett, Igor Polonsky, and Yingtao Ma for providing the high spectral resolution microwave CRTM coefficients used in this study and the two anonymous reviewers for their helpful comments on the initial versions of this manuscript.

The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author(s) and do not necessarily reflect those of NOAA or the Department of Commerce.

#### REFERENCES

- [1] S. A. Boukabara, K. Garrett, and V. K. Kumar, "Potential gaps in the satellite observing system coverage: Assessment of impact on NOAA's numerical weather prediction overall skills," *Monthly Weather Rev.*, vol. 144, no. 7, pp. 2547–2563, Jul. 2016.
- [2] W. J. Blackwell et al., "Hyperspectral microwave atmospheric sounding," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 128–142, Jan. 2011.
- [3] S. Boukabara and K. Garrett, "Benefits of a hyperspectral microwave sensor," in *Proc. SENSORS*, 2011, pp. 1881–1884.
- [4] C. D. Kummerow et al., "Hyperspectral microwave sensors—Advantages and limitations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 764–775, Dec. 2022, doi: [10.1109/JSTARS.2021.3133382](https://doi.org/10.1109/JSTARS.2021.3133382).
- [5] A. Gambacorta et al., "The hyperspectral microwave photonic instrument (HYMPI) —Advancing our understanding of the earth's planetary boundary layer from space," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 4468–4471, doi: [10.1109/IGARSS46834.2022.9883151](https://doi.org/10.1109/IGARSS46834.2022.9883151).
- [6] J.-F. Mahfouf, C. Birman, F. Aires, C. Prigent, E. Orlandi, and M. Milz, "Information content on temperature and water vapour from a hyper-spectral microwave sensor," *Quart. J. Roy. Meteorological Soc.*, vol. 141, no. 693, pp. 3268–3284, Oct. 2015.
- [7] F. Aires, C. Prigent, S. A. Buehler, P. Eriksson, M. Milz, and S. Crewell, "Towards more realistic hypotheses for the information content analysis of cloudy/precipitating situations—Application to a hyperspectral instrument in the microwave," *Quart. J. Roy. Meteorological Soc.*, vol. 145, no. 718, pp. 1–14, 2019.

- [8] S. Ding et al., "Validation of the community radiative transfer model," *J. Quantitative Spectrosc. Radiative Transfer*, vol. 112, pp. 1050–1064, 2011, doi: [10.1016/j.jqsrt.2010.11.009](https://doi.org/10.1016/j.jqsrt.2010.11.009), 2011.
- [9] S.-A. Boukabara et al., "MiRS: An all-weather IDVAR satellite data assimilation and retrieval system," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 9, pp. 3249–3272, Sep. 2011.
- [10] I. Moradi et al., "Implementation of a discrete dipole approximation scattering database into community radiative transfer model," *J. Geophysical Res., Atmospheres*, vol. 127, 2022, Art. no. e2022JD036957, doi: [10.1029/2022JD036957](https://doi.org/10.1029/2022JD036957).
- [11] P. G. Stegmann, G. Tang, P. Yang, and B. T. Johnson, "A stochastic model for density-dependent microwave snow- and graupel scattering coefficients of the NOAA JCSDA community radiative transfer model," *J. Quantitative Spectrosc. Radiative Transfer*, vol. 211, pp. 9–24, 2018, doi: [10.1016/j.jqsrt.2018.02.026](https://doi.org/10.1016/j.jqsrt.2018.02.026).
- [12] F. Chevallier, S. D. Michele, and A. P. McNally, "Diverse profile datasets from the ECMWF 91-level short-range forecast," Tech. Rep. NWP SAF-EC-TR-010, NWP SAF, Exeter, U.K., 2006.
- [13] A. Chédin, N. A. Scott, C. Wahiche, and P. Moulinier, "The improved initialization inversion method: A high resolution physical method for temperature retrievals from satellites of the TIROS-N series," *J. Appl. Meteorol.*, vol. 24, pp. 128–143, 1985.
- [14] F. Rabier, N. Fourrié, D. Chafaï, and P. Prunet, "Channel selection methods for infrared atmospheric sounding interferometer radiances," *Quart. J. Roy. Meteorological Soc.*, vol. 128, pp. 1011–1027, 2002.
- [15] N. Fourrié and J. Thépaut, "Evaluation of the AIRS near-real-time channel selection for application to numerical weather prediction," *Quart. J. Roy. Meteorological Soc.*, vol. 129, pp. 2425–2439, 2003, doi: [10.1256/qj.02.210](https://doi.org/10.1256/qj.02.210).
- [16] S. D. Michele and P. Bauer, "Passive microwave radiometer channel selection based on cloud and precipitation information content estimation," *Quart. J. Roy. Meteorological Soc.*, vol. 132, pp. 1299–1323, 2005.
- [17] A. D. Collard, "Selection of IASI channels for use in numerical weather prediction," *Quart. J. Roy. Meteorological Soc.*, vol. 133, pp. 1977–1991, 2007.
- [18] D. F. Parrish and J. C. Derber, "The national meteorological center's spectral statistical interpolation analysis system," *Monthly Weather Rev.*, vol. 120, pp. 1747–1763, 1992.
- [19] E. S. Maddy and S. A. Boukabara, "MIIDAPS-AI: An explainable machine-learning algorithm for infrared and microwave remote sensing and data assimilation preprocessing—Application to LEO and GEO sensors," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8566–8576, Aug. 2021, doi: [10.1109/JSTARS.2021.3104389](https://doi.org/10.1109/JSTARS.2021.3104389).
- [20] E. Jones, E. Maddy, K. Garrett, and S.-A. Boukabara, "The MIIDAPS algorithm for retrieval and quality control for microwave and infrared observations," in *Proc. 23rd Conf. Integr. Observing Assimilation Syst. Atmos. Oceans Land Surf. Appl. Data Assimilation*, 2019, pp. 1–6.
- [21] Y. Han et al., "JCSDA community radiative transfer model (CRTM)—Version 1," NOAA, Tech. Rep. NESDIS 122, 2006.
- [22] S. A. Boukabara, E. S. Maddy, and F. Iturbide-Sanchez, "Can we design NWP data assimilation based entirely on AI techniques? assessing advantages and challenges," Presented at the ITSC-24: Tromsø, Norway, Mar. 16–22, 2023. [Online]. Available: [https://itwg.ssec.wisc.edu/wordpress/wpcontent/uploads/2023/05/oral.11.01.iturbide\\_itsc24.pdf](https://itwg.ssec.wisc.edu/wordpress/wpcontent/uploads/2023/05/oral.11.01.iturbide_itsc24.pdf)
- [23] Q. Liu, C. Cao, C. Grassotti, and Y.-K. Lee, "How can microwave observations at 23.8 GHz help in acquiring water vapor in the atmosphere over land?," *Remote Sens.*, vol. 13, 2021, Art. no. 489, doi: [10.3390/rs13030489](https://doi.org/10.3390/rs13030489).



**Eric S. Maddy** received the B.S. degree in physics and B.A. degree in mathematics from the Frostburg State University, Frostburg, MD, USA, in 2001, and the M.S. and Ph.D. degrees in atmospheric physics from the University of Maryland Baltimore County, Baltimore, MD, USA, in 2003 and 2007, respectively.

He is currently an expert on traditional and exploratory (artificial intelligence) polar and geostationary IR/MW remote sounding, data assimilation, and data fusion. He has more than 19 years of experience working with National Oceanic and Atmospheric Administration center for SaTellite Applications and Research leading projects on the development of carbon trace gases retrievals from operational hyperspectral sounders, fusing algorithms and products from multiple instrument platforms (CrIS/ATMS/VIIRS, IASI/AMSU-A MHS/AVHRR), and leading the development of artificial intelligence versions of those algorithms.



**Flavio Iturbide-Sanchez** (Senior Member, IEEE) received the B.S.E.E. degree in electronics engineering from the Autonomous Metropolitan University, Mexico City, Mexico, in 1999, the M.S.E.E. degree in electrical engineering from the Advanced Studies and Research Center, National Polytechnic Institute, Mexico City, Mexico, in 2001, and the Ph.D. degree in electrical and computer engineering from the University of Massachusetts, Amherst, MA, USA, in 2007.

His Ph.D. research focused on the miniaturization, development, calibration, and performance assessment of low-cost and power-efficient microwave radiometers for remote sensing applications. From 2001 to 2005, he was a Research Assistant with the Microwave Remote Sensing Laboratory, University of Massachusetts, where he was involved in the design, development, and characterization of highly integrated multichip modules and microwave circuits for low-noise, low-power consumption, high-gain, and high-stability microwave radiometers. From 2005 to 2007, he was with the Microwave Systems Laboratory, Colorado State University, Fort Collins, CO, USA, focusing on the demonstration of a low-cost and power-efficient compact microwave radiometer for humidity profiling. From 2008 to 2018, he supported the development of operational physical retrieval systems that employ hyperspectral-infrared and microwave observations implemented for the Polar Operational Environmental Satellites Project and the Joint Polar Satellite System (JPSS). He has been a Physical Scientist with NOAA/NESDIS/Center for Satellite Applications and Research, College Park, MD, USA, since 2018, where he has led the calibration and validation of the JPSS Cross-track Infrared Sounder and Advanced Technology Microwave Sounder instruments. He is also leading exploitation, evaluation, and value assessment activities on emerging environmental remote sensing technologies and technology concepts in support of the planning of the space architecture of NOAA's next generation of Earth observing systems. His research interests include satellite remote sensing, satellite data assimilation, inverse theory applied to geoscience fields, weather forecasting, earth system science, small satellites, and the design of remote sensing systems for earth observations based on emerging technologies and concepts.

Dr. Iturbide-Sanchez was recipient of the 2022 NOAA Administrator Award and the 2023 NASA's Robert H. Goddard Honor Award in the category of Science Teams. He is currently the Cochair of the World Meteorological Organization Global Space-based Inter-Calibration System Research Working Group Microwave Subgroup.



**Sid Ahmed Boukabara** received the Ingénieur degree in electronics and signal processing from the National School of Civil Aviation, Toulouse, France, in 1994, the M.S. degree in signal processing from the National Polytechnic Institute of Toulouse, Toulouse, France, in 1994, and the Ph.D. degree in remote sensing from the Denis Diderot University, Paris, France, in 1997.

He has 28 years of cumulated experience in satellite remote sensing design, research, and related applications. Eighteen of those years were spent with the

National Oceanic and Atmospheric Administration where he was the Principal Scientist with the Office of Systems Architecture and Advanced Planning, supporting the effort to plan, design, and evolve the nation's operational space and ground architectures of the future. In addition, he has held the position of Deputy Director and acting Director for the multiagency U.S. Joint Center for Satellite Data Assimilation as well as acting deputy director of the Center for Satellite Applications and Research.

He currently serves as the *senior Program Scientist for strategy* of the Earth Science Division (ESD) within NASA's Science Mission Directorate. He has been a member of the ESD leadership team since April 2023, supporting the Director and Deputy Director in framing and executing NASA Earth Science strategy and the multiagency collaboration efforts. As senior program scientist, he is also responsible for facilitating strategic science initiatives that require coordination across multiple elements of ESD encompassing technology, flights, data systems, research and earth action.