https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

A Machine-learning Based Marine Planetary Boundary Layer (MPBL) Moisture Profile Retrieval Product from GNSS-RO Deep Refraction Signals

Jie Gong[1,*], Dong L. Wu[1], Michelle Badalov[2], Manisha Ganeshan[3,1], and Minghua Zheng[4]

[1]Climate and Radiation Lab, NASA Goddard Space Flight Center
[2]Dept. of Computer Science, Univ. of Maryland College Park
[3]GESTAR-II/Morgan State University
[4]University of California at San Diego

Correspondence: Jie Gong (Jie.Gong@nasa.gov)

Abstract. Marine planetary boundary layer (MPBL) water vapor amount and gradient impact the global energy transport through directly affecting the sensible and latent heat exchange between the ocean and atmosphere. Yet, it is a well-known challenge for satellite remote sensing to profile MABL water vapor, especially when cloud or sharp gradient of water vapor are present. Wu et al. (2022) identified good correlations between Global Navigation Satellite System (GNSS) deep refraction signals (SNR) and the global MPBL water vapor specific humidity when the radio occultation (RO) signal is ducted by the moist PBL layer, and they laid out the underlying physical mechanisms to explain such a correlation. In this work, we apply a machine-learning/artificial intelligence (ML/AI) technique to realize pixel-level MPBL water vapor profiling. A convolutional neural network (CNN) model is trained using 20 months of global collocated hourly ERA-5 reanalysis and COSMIC1 1 Hz SNR observations between 975 – 850 hPa with 25 hPa vertical resolution, and then the model is applied to both COSMIC1 and COSMIC2 in other time ranges for independent retrieval and validation. Monte Carlo Dropout method was employed for the uncertainty estimation. Comparison against multiple field campaign radiosonde/dropsonde observations globally suggests SNR-retrieved water vapor consistently outperforms ERA-5 reanalysis and the Level-2 standard retrieval product at all six pressure levels between 975 hPa and 850 hPa, indicating real and useful information is gained from the SNR signal albeit training was performed against the reanalysis. The only exception is in the deep tropics where the fundamental assumption for SNR-retrieval to work is invalidated frequently by interactions among ocean surface, MPBL and shallow convections. Climatology and diurnal cycle of MPBL structure constructed from the ML-SNR technique is studied and compared to the reanalysis. Disparities of climatology suggest ERA-5 may systematically produces dry biases at high-latitudes, and wet biases in marine stratocumulus regions. The diurnal cycle amplitudes are too weak and off-phase in ERA-5, especially in Arctic and stratocumulus regions. These areas are particularly prone to PBL processes where this GNSS-SNR water vapor product may contribute the most.

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

1 Introduction

As a key component of Earth's lower atmosphere, the planetary boundary layer (PBL) water vapor plays a pivotal role in Earth's energy budget, exerting profound influence on weather and climate processes. It is an essential factor of the Earth's energy budget, influencing radiative forcing and consequently climate variability and long-term change. Furthermore, PBL water vapor is instrumental in modulating local and regional weather patterns by affecting cloud formation, precipitation and temperature. Therefore, study of PBL water vapor stands as a vital element in advancing our comprehension of Earth's atmosphere and its broader implications for our planet's climate system.

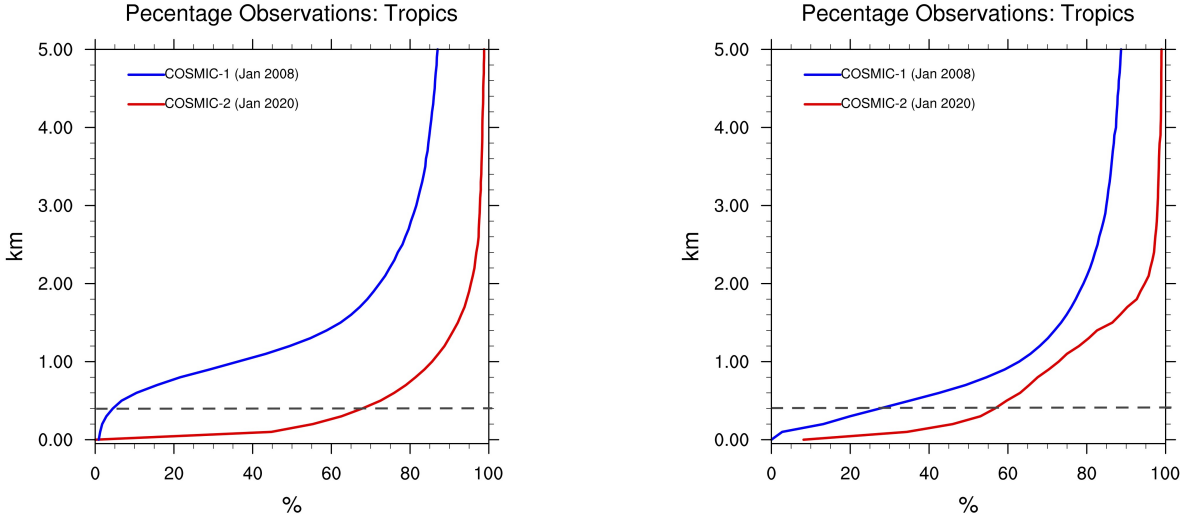
70% of the Earth's surface is covered by water. The sensible and latent heat exchange between ocean boundary and the marine PBL (MPBL) happens at different spatial and temporal scales, which is determined by not only ocean surface properties (e.g., wind speed, sea surface temperature) but also MPBL thermodynamic structures. For example, under the context of polar proneness to the climate change, Boisvert et al. (2015) found Arctic PBL humidity and temperature biases in the reanalysis are the major error sources for the evaporation estimation compared to satellite observations. Cloud-cimate feedback is another motivation highlighted by NASA's PBL incubation study (Teixeira et al. (2021)). As another example, Milan et al. (2019) found strong correlation between MPBL cloud top height and below-cloud water vapor amount using two joint satellite retrieval products.

Data sparsity is a critical problem for advancing MPBL science. Satellite remote sensing undoubtedly provides the best solution in terms of global coverage, but it is very difficult to retrieve MPBL WV and its vertical distribution when cloud or sea ice are present. When clouds are present in the scene, it often overwhelms the emission signal from the MPBL water vapor and prevents passive instruments sensing the below-cloud atmosphere. When sea ice is present in the scene, scattering or surface emission from the sea ice are often inseparable from water vapor emission signals and distort the retrieval result. Taking the aforementioned two research as examples, Boisvert et al. (2015) uses Level 2 AIRS water vapor and temperature retrieval products, which are only available for clear or partially-cloudy sky situations, so it inherently contains a sampling bias. Milan et al. (2019) derived MPBL total WV amount from subtracting MODIS above-cloud water vapor from AMSU-A total column water vapor, which couldn't be used to gain vertical information of WV in the MPBL.

Using low-frequeny MW L-band to transmit signals along the limb path, the Global Navigation Satellite System (GNSS) satellite overcomes the two above difficulties and provide superb vertical resolution of the MPBL water vapor under all-sky conditions. GNSS Radio Occultation (GNSS-RO) retrieves temperature and water vapor profiles routinely from the Level 2 bending angle product, the latter of which is used operationally in numerical weather data assimilation systems to improve weather forecasts (e.g.,Kuo et al. (2000)). Because of the rapid growth of SmallSat/CubeSat constellations from both the commercial and non-profit sectors, GNSS-RO technique provides a promising future for the needed global spatial-temporal sampling of MPBL WV and its variability. Like other limb sounders, the disadvantage of GNSS-RO is its relatively coarse horizontal resolution (several hundred kilometers) that smears out horizontally inhomgeneous signals. This is typically not a big conern in MPBL as vertical gradient is much sharper and harder to characterize if not using in-situ measurements (e.g., shipborne radiosonde).

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

55 However, GNSS-RO WV retrieval profiles have excessively high failure rate in the MPBL. That is because the GNSS-RO signal-to-noise ratio (SNR) decreases with height due to the atmospheric defocusing effect, and the Level-2 refractivity retrievals which require a high SNR drops subsequently in the MPBL as the signal becomes noisy. As a result, the GNSS-RO water vapor retrievals that rely on the refractivity profiles often fail to converge. Fig. 1 gives an example of the success statistics (%) as a function of height for temperature (Fig. 1a) and water vapor (Fig. 1b) over the tropical ocean. Using $0.5km$
60 and $1km$ above the ocean surface as the reference line, we can see although the COSMIC-2 (Constellation Observing System for Meteorology, Ionosphere, and Climate-2) has significantly improved its SNR compared to its predecessor COSMIC-1, the success rate is still less than 60% at $0.5km$ and slightly over 70% at $1km$ for the GNSS-RO WV retrieval, while this number is only 25% and 60% for COSMIC-1 at respective altitudes. The low SNR widely exists for commercial GNSS satellites especially in the lowest 500 m above the surface (Maneshan et al. (2024)).



(a) AtmPrf success rate

(b) WetPrf success rate

Figure 1. Level 2 atmPrf (temperature) and wetPrf (water vapor) successful retrieval rate (%) as a function of height from COSMIC-1 during January 2008 (blue) and COSMIC-2 during January 2020 (red). Success rate is calculated by dividing number of valid Level-2 retrieval files over number of Level-1B file at a certain height. The grey dashed lines mark the reference at 0.5 km from the tropical ocean surface.

65 Wu et al. (2022) found out that the Level-1B deep SNR from the straight-line height (H_{SL}) is statistically significantly correlated with the MPBL water vapor amount in the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) after averaging over a month and a $2.5°X2.5°$ grid box. The averaging is necessary to effectively beat down the random noise. This paper attributed such a positive correlation to the strong refraction from a horizontally

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

stratiform and dynamically quiet MPBL water vapor layer that acts to enhance the SNR amplitude at deep H_{SL} through ducting and diffraction/interference. Some caveats of this work limit its application to weather phenomena. First, it builds upon a single level regression statistics, which was found the highest at $H_{SL} = -100km$ in the tropics, and $H_{SL} = -80km$ at high latitudes. Hence, any simple linear regression-based retrieval algorithm will understandably suffer from arbitrary latitudinal discontinuities. As a matter of fact, SNR at different H_{SL} levels are found correlated with MPBL water vapor with different signs and magnitudes (e.g., Fig. 2), which should be used together to enhance the information content. Secondly, the robust relationship is only found for monthly averages in Wu et al. (2022), because the profile-by-profile noise is usually too high for yielding a meaningful retrieval from SNR, and only through averaging large amount of profiles can the noise be lowered down to the level where signal stands out. These are all caveats of traditional statistical approach. Machine learning approach, however, is suitable at picking up weak signals through large amount of training data.

Artificial Intelligence/Machine learning (AI/ML) technique is trending in the past decade. It is more and more appreciated by the satellite remote sensing field in recent years. Traditionally physics-based radiative transfer (RT) theories and modelings are used to link the remote sensing measurements (e.g., GNSS radio occultation signal) to the physical quantities (e.g., temperature and water vapor profiles). They are often highly non-linear and involve many assumptions/simplifications, which propagate into part of the retrieval errors subsequently. Given the fact that satellite measurements usually contain large amount of data, and the association is highly non-linear between the measurement space and the physical space, the retrieval process becomes an ideal testbed for ML capabilities. Some pioneer works had attempted this approach to retrieve PBL atmosphere profiles and achieved notable success. For example, Ye et al. (2021) used the routine radiosonde measurement at a Atmospheric Radiation Measurement (ARM) site as the ground truth to train a ground-based infrared spectrometer to predict the PBL height. The capability is limited to only the stations where both observations are routinely available. Milestein and Blackwell (2016) employed a neural network (NN) framework on retrieving the temperature and water vapor profiles from the spaceborne Atmospheric Infrared Sounder (AIRS) observations (AIRS Version 7 product). The training "truth" was from the ECMWF analysis fields. It is worth mentioning that Milstein (2022), as a follow-up work, pointed out the ML-only retrieval framework tends to smooth out sharp gradient features in proximity to the PBL top. To mitigate this caveat, Milstein et al. (2023) employs the 3D deep neural network training on the AIRS granule image against ERA-5 reanalysis that helps PBL height recognition from passive imagers.

In this paper, we will thoroughly explore the ML capability at retrieving the MPBL WV information from the deep SNR signal at profile-by-profile basis (i.e., Level-2 standard). Section 2 introduce the training and validation datasets as well as the model structure; Section 3 presents the retrieval results and independent validation; Section 4 will expand the discussion to the usage of this product in studying MPBL water vapor climatology and diurnal variabilities; Section 5 summarizes the major findings and shortcomings of the current work that may be improved in the future.

2 Data and Model

This section introduces the training, validation and independent validation datasets, as well as the ML model architecture.

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

2.1 Training and Validation Datasets

The definition of SNR follows Wu et al. (2022) which uses the normalized SNR (S_{RO}):

$$S_{RO} = (SNR - \sigma)/(SNR_0 - \sigma) \quad (1)$$

$$\sigma_S^2 = VAR(S_{RO} - \overline{S_{RO}}) \quad (2)$$

SNR_0 is the free atmosphere SNR. In practice, we use averaged SNR between 35 and 65 km altitude range as the SNR_0. σ is the instrument-specific noise determined for each individual instrument from very deep H_{SL}. The value for σ used in this work is an updated version from Table A1 in Wu et al. (2022) and shown in the Appendix A. Wu et al. (2022) also found an instrument-dependent shift of the mean S_{RO} profile as a function H_{SL}. Luckily, such an issue can be resolved to use the excessive phase at L1 as the vertical coordinate. In practice, the raw calculated S_{RO} and σ_S^2 are mapped to a fixed 52-level $Log_{10}(\phi_{L1})$ vertical grid. It is roughly linearly correlated with H_{SL}. The value for the vertical grid is listed in Table A2 in the Appendix A.

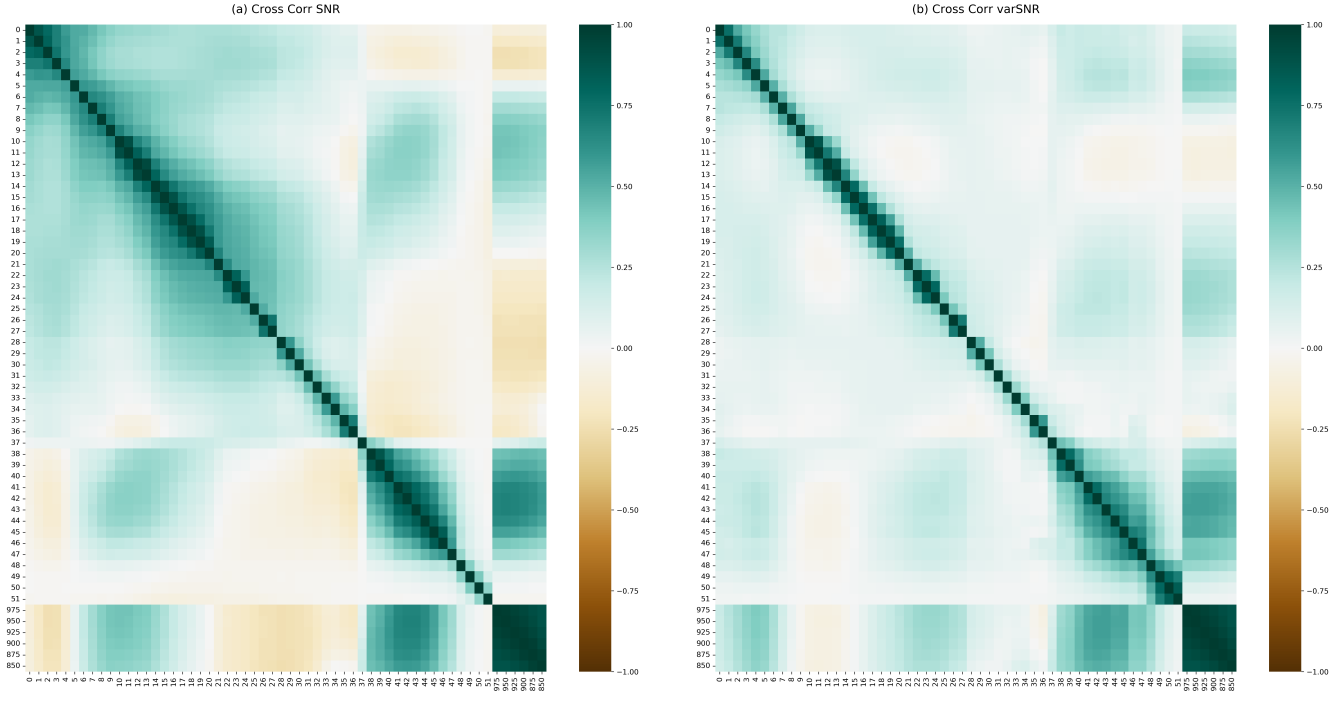
The ERA-5 reanalysis is so far the best global reanalysis dataset in terms of PBL water vapor amount and distribution. Johnston et al. (2021) compared specific humidity from ERA-5 and MERRA-2 reanalysis against collocated and coincident GNSS-RO Level-2 wetPrf specific humidity retrieval profiles, and found ERA-5 outperforms MERRA-2 everywhere in the PBL. They both exhibit consistent dry biases with larger bias from mid-high latitudes. However, ERA-5 percentage bias is roughly 1/2 of that from the MERRA-2 reanalysis. Given that many previous works used ERA-5 reanalysis or ECMWF analysis for training or validating the satellite retrievals for water vapors (e.g., Milestein and Blackwell (2016), Milstein et al. (2023)), especially some recent ones using it as the standard to evaluate recent GNSS-RO missions (e.g., Chang et al. (2022), Zhran (2023), Maneshan et al. (2024)) , it is well justified to use ERA-5 hourly reanalysis as the "training" dataset to create a large sample globally.

In this work, we created a collocated and coincident ERA-5 - SNR training and validation dataset. The SNR records are from both COSMIC-1 and COSMIC-2, and the periods for training, independent testing, and prediction are listed in Table 1. Note that the validation period is independent from training period to avoid potential self-correlation using standard random splitting procedure. The prediction period however covers both training and validation periods.

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

Table 1. Training and validation periods

Training (90% and 10% random-splitting)	COSMIC-1	2012.01-2012.12, 2016.01-2016.03, 2017.01-2017.03
	COSMIC-2	2020.01-2020.05
Testing	COSMIC-1	2018.01-2018.03
Prediction	COSMIC-1	2012.01-2012.12, 2013.01-2013.12, 2016.01-2016.03, 2017.01-2017.03, 2018.01-2018.03
	COSMIC-2	2020.01-2020.12



(a) S_{RO} & MPBL specific humidity

(b) σ^2_{SNR} & MPBL specific humidity

Figure 2. Cross-correlation among S_{RO} (left) and σ^2_{SNR} (right) at various SLH levels from one random day for COSMIC1 and their cross-correlation with collocated ERA5 specific humidity at 975 - 850 hPa (bottom-right corner of each panel). Only grid indices are shown in the axis titles, and the corresponding $Log_{10}(\phi_{L1})$ values can be found in Table A2.

125 Fig. 2 elucidates the linear correlation between S_{RO} at each of the 52 levels and ERA-5 specific humidity at 975, 950, 925, 900, 875 and 850 hPa over global ocean. The largest positive correlations are found around Level #11 to Level #15, which

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

roughly correspond to $H_{SL} = -100km$ to $-80km$ (Table A2). Based on the monthly averages, Wu et al. (2022) found the highest correlation at $H_{SL} = -100km$ in the tropics and at $H_{SL} = -80km$ for the polar regions, which is consistent with our profile-by-profile correlation as well. But Fig. 2 also shows positive or negative correlations at different $Log_{10}(\phi_{L1})$ levels, which impede methods like multi-variable regression from working. σ_S^2 also exhibits non-linear patterns with generally weaker correlations with MPBL water vapor that are opposite in sign compared to that of S_{RO}. It is worth noting that these relationships are also instrument dependent, as can be clearly seen in the S_{RO} cross-correlation for Metop-A and Metop-B in the Appendix Fig. A1. In this work, the target variables are specific humidity at the aforementioned 6 pressure levels ($975hPa$, $950hPa$, $925hPa$, $900hPa$,$875hPa$ and $850hPa$). The input parameters are 52 levels of S_{RO}, 52 levels of σ_S^2, latitude, longitude, month and ascending/descending flag.

In addition to the independent testing which is standard procedure for ML/AI training, a handful of shipborne radiosonde campaigns and airborne dropsonde campaigns data are collected for further independent assessment. The campaign names, location and total number of valid profiles are presented in Fig. 3 and Table 2. We can see from the summary of weather scenarios during each campaign that this independent validation dataset comprehensively covers major marine weather regimes from extremely dry Southern Ocean (MARCUS), mid-latitude stratocumulus region (MAGIC), tropical trade cumulus region (EUREC4A, ATOMIC), to episodically wet atmospheric river events (ARRecon). This exercise is critical for assessing the quality of ERA-5, Level 2 retrieval, and Level 1 SNR-based retrieval under different weather scenarios.

Table 2. Campaign Information

Campaign Name	Period used for validation	Location	Weather Regime	Type	Reference
MARCUS	2017.11-2018.03	Southern Ocean	Mixed-Phase PBL cloud	Radiosonde	Evan et al. (2022)
ATOMIC	2020.01-2020.02	Tropical North Atlantic	Tropical trade wind zone	Radiosonde and Dropsonde	George et al. (2021)
EUREC4A	2020.01-2020.02	Tropical North Atlantic	Tropical trade wind zone	Radiosonde	Stephan et al. (2021)
MAGIC	2012.10-2013.09	Eastern North Pacific Ocean	Subtropical MPBL	Radiosonde	Evan et al. (2022)
ARRecon	2018.02; 2020.01-2020.02	Northeast Pacific off the coast of California	Atmospheric River	Dropsonde	Zheng et al. (2024)

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

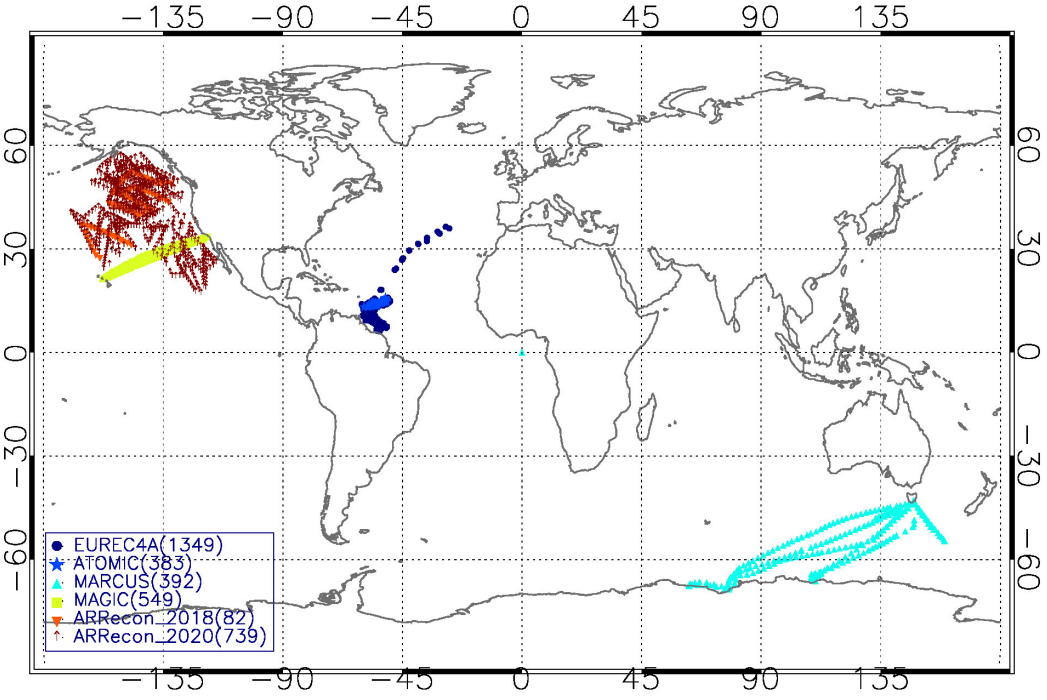


Figure 3. Map for radiosonde/dropsonde locations from different shipborne or airborne campaigns.

2.1.1 Machine Learning Model Selection

The Convolutional Neural Network (CNN) model is chosen as our regression ML model. Compared to some old-fashioned ML models (e.g., random forest, gradient boosting), CNN learns also the vertical cross-correlation within the 52-layer input SNR profiles, as well as within the targeted 6-layers of specific humidity profiles. We conducted a comprehensive search of best hyperparameters using the root-mean-square-error (RMSE) as the loss function. The final selected architecture contains one 1-D convolutional block consisting of a Conv1d layer, a MaxPool1d layer, a BatchNorm1d, two dense layers, and a ReLU activation function. The Kernel size is selected to be 5, and a dropout layer of rate = 0.1 is added between layers. 100 epochs suffice the needs for quick convergence for both training and validation.

In the prediction step, 30 predictions were carried out given each input set of variables, the mean and standard deviation of which were used as the final prediction and errorbar. Note that the dropout layers were designed in ML as a standard technique to regularize model over-fitting (Srivastava et al. (2013)), but were also employed widely as a Bayesian-approximation to quantify model uncertainties(Gal and Ghahramani (2016)). Admittedly the current method only provides a quantification for

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

ML model errors. There is no consideration of SNR measurement errors nor propagation of the error to the final retrievals at this moment, although this is certainly some procedure to be in place in the future works.

We also tried some old-fashioned ML models, e.g., random forest (RF), gradient boosting (GB), logistic regression (LR), support vector machine (SVM) and one deep learning model multilayer perceptron (MLP). The model performances are actually very close in terms of evaluating the RMSE except for the LR and SVM, the latter of which performed discernibly worse than the rest ML models. It is not a surprise finding as this is a relatively simple and straightforward task that ML models should handle easily, but not the case for multi-variable linear regression type of logistic models (hence, it explains the poor performance of LR and SVM). As the main focus of this paper is science and new information content embedded in SNR signals, we will not deviate the attention to spend more time discussing these model results. The semi-transparency of RF and GB models is appreciated by us though. We compared the feature importance rankings with Wu et al. (2022) findings, and find high consistencies (e.g., high ranking of SNR at $H_{SL} = -100\ km$ in the tropics, and SNR at $H_{SL} = -80\ km$ ranks the top in the polar region).

3 Results

3.1 Retrieval Performance Evaluation

As the first comparison, Fig. 4 and Fig. 5 showcase the statistical closeness to the $1:1$ line and the resemblance of geographical distributions for the three independent testing months: January - March, 2018, for COSMIC-1. All 6 pressure levels are compiled together to make Fig. 4, but were otherwise look extremely similar if plotting layer-by-layer. The only deviation from $1:1$ line occurs at very small specific humidity values (ERA-5 specific humidity $< 1g/kg$), i.e., very dry conditions, normally occurs at high-latitudes.

Such a discrepancy reveals itself more clearly when we map out the percentage difference (Fig. 5b). The largest percentage differences indeed are shown at polar regions as well as near the coastal lines with SNR-retrieved humidity tends to be larger than ERA-5. Note that to satisfy ducting conditions in order to use SNR signal at deep H_{SL}, the surface is required to be flat. Therefore, the discrepancies around the coastal line are believed to be related to issues with SNR-retrievals when topography starts to play a role. However, as we will show later in Fig. 8, ERA-5 indeed shows consistent dry-bias at high-latitudes compared to independent radiosonde measurements. So SNR-retrieval might produce a closer-to-truth results as will be seen later as well. Moreover, one can visually discern discrepancies in the tropical deep convection/ITCZ zones with ERA-5 in general wetter than SNR-retrieved values. Such a discrepancy is not conspicuous in Fig. 5b simply because of the large value in the denominator. We will also show later that none of the three datasets we will evaluate (SNR-retrieval, GPS-RO Level 2 retrieval and ERA-5 reanalysis) capture well the tropical MPBL structures. For the SNR method, it is probably because the ducting assumption is easily and frequently violated in the tropical MPBL.

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

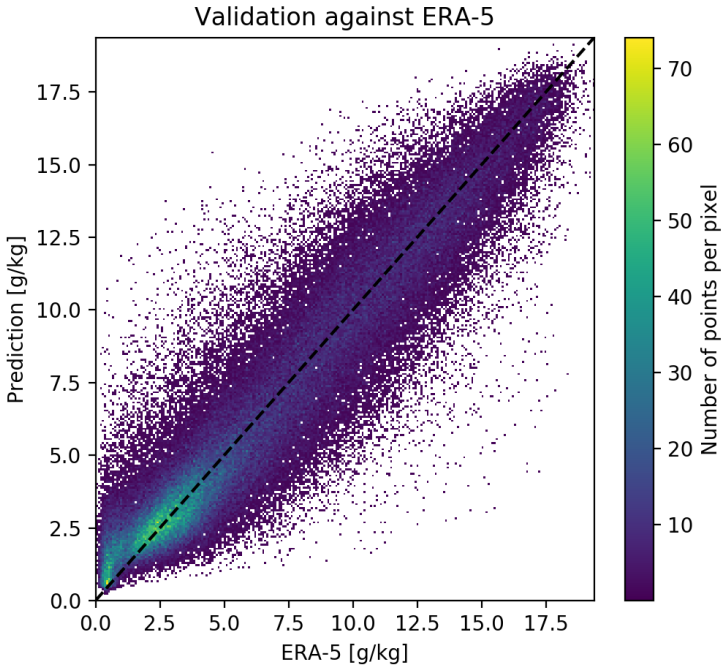


Figure 4. Heatmap for independent validation from January - March 2018 for all 6 levels together.

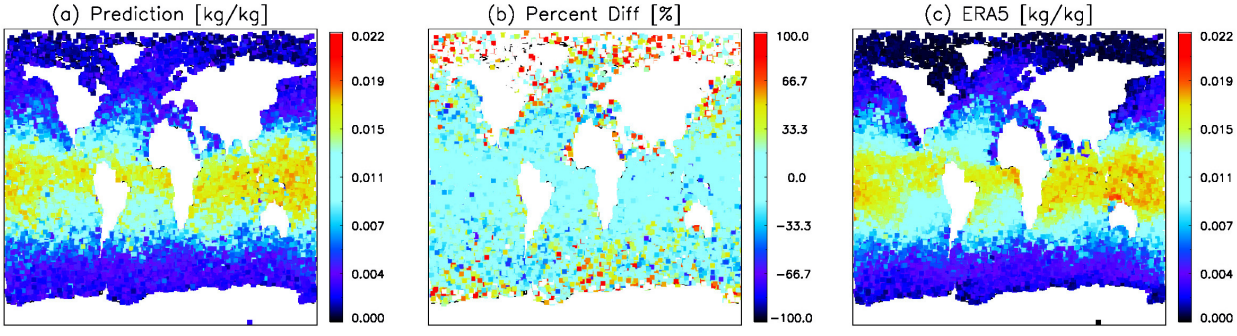


Figure 5. Geographic distribution of the (a) predicted values versus (c) ERA-5 validation values at 950 hPa for January - March, 2018. The middle panel is the percentage difference between (a) and (c).Only ERA-5 samples that collocate and coincident with COSMIC-1 SNR-retrievals are selected for this comparison

3.2 Uncertainty Quantification

Unfortunately, for the very dry conditions, SNR-retrieved specific humidity also inherently comes along with large uncertainties, as can be clearly seen in Fig. 6. The SNR signal is too weak in this situation to yield any robust retrievals even with

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

powerful ML models. Although we still believe the SNR-retrievals might be "more correct" than ERA-5 for very dry conditions, in practice we mark any retrieval with greater than 50% uncertainty with a quality flag in the published product, and those data do not pass the the quality control to be used later in this paper for independent validation nor constructing the climatologies. We can also see from Fig. 6 that almost all SNR-retrievals greater than $2g/kg$ passes the quality control. Readers should keep in mind that our current uncertainty estimation approach under-estimates the real uncertainty because it does not take into account SNR errors.

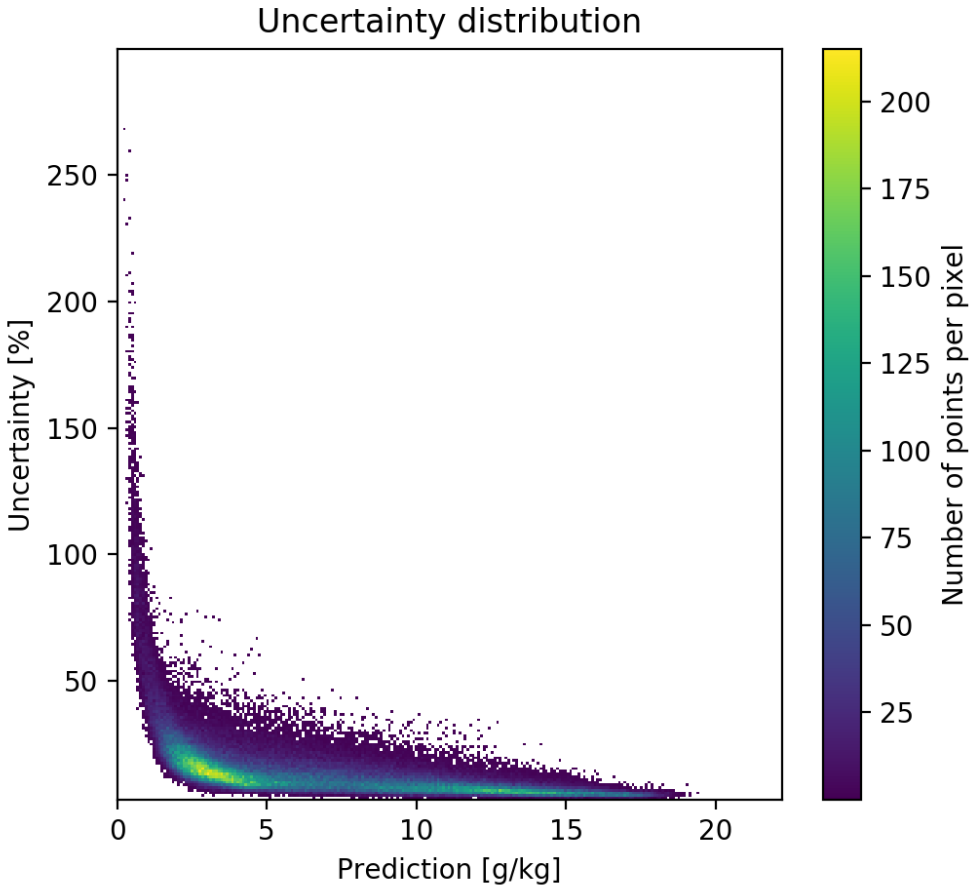


Figure 6. Percentage uncertainty distribution as a function of predicted value.

3.3 Comparison to Independent Radiosondes

In order to find collocation samples in every campaign, the collocation criteria are slightly different given the consideration of (1) the abundance of radiosonde/dropsonde profiles; (2) the typical spatial and temporal homogeneity of the local weather regime; (3) the availability of daily COSMIC-1 or COSMIC-2 profiles. In practice, for EUREC4A and ATOMIC, collocation is

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

defined as longitude difference within $2°$, latitude difference within $1.5°$, and time difference within $1\ hr$. For the two Southern Ocean campaign, the thresholds become $4°$,$2.5°$ and $2\ hr$ correspondingly. For ARRecon and MAGIC, the thresholds are $4°$,$1.5°$ and $2\ hr$.

Fig. 7 shows the layer-by-layer comparison for all collocated samples from all campaigns. SNR-retrieval results are shown in filled colors while Level-2 retrievals are shown in open symbols. In addition, the averages from each campaign collocation subsets are connected together for better visual comparison against the $1:1$ line. We can see both SNR-retrieval and Level-2 retrieval demonstrate close agreement with ground "truth" for different weather regimes. In general, better correlation are found when the MPBL is relatively dry or moderately-wet. For the ARRecon campaign specifically, SNR-retrieval has an overall better agreement compared to Level-2 retrievals (black lines versus black dash-dotted lines), and the few extremely large specific humidity values ($> 12 g/kg$) are well captured by the SNR-retrieval but not the Level-2 retrievals. The mean of all ARRecon collocated samples also suggest SNR-retrieval is the only one that doesn't produce a bias, while both ERA-5 and GNSS Level-2 retrievals are slightly dry-biased in atmospheric rivers. Such a close agreement appears to become noisier at $850\ hPa$, again demonstrated that signals at sharp boundaries (i.e., PBL top) are hard to retrieve. ERA-5 from each campaign (only considering samples that SNR-retrieval collocation is found) exhibit good agreement to the ground truth too if only mean values of each subset is considered (grey dotted lines). However, the scatter plots are much noisier if all collocations are plotted (Fig. 8). For the two deep tropics campaigns ATOMIC and EUREC4A, we can clearly see none of the three datasets capture the humidity conditions in the MPBL. They are all dry-biased, and means from ERA-5 reanalysis are less dry-biased than GPS retrieved values at $975\ hPa$ and $950\ hPa$.

For the convenience of pinpointing ERA-5 MPBL issues, we also make Fig. 8 as each valid radiosonde/dropsonde profile from all 6 campaigns can always collocate an ERA-5 reanalysis data sample within $1.5°$ longitude, $1°$ latitude and $1\ hr$ difference. Now we can clearly see ERA-5 didn't capture the MPBL humidity change in the majority time during the EUREC4A campaign with large wet-biases. Another discernible bias happens in the Southern Ocean during the MARCUS campaign, where ERA-5 is consistently dry-biased. Overall ERA-5 shows a small dry-bias globally at all levels, which agrees with early findings by Johnston et al. (2021) who used Level-2 GPS-RO retrievals to identify such a dry bias. Note that some of the campaign profiles (e.g., ARRecon dropsondes) are actually assimilated in the ERA-5 data, so it is not a completely independent validation strictly speaking.

The violin plots in Fig. 9 reveals more detailed difference in comparison statistics with respect to the radiosonde/dropsonde "truth", which more comprehensively demonstrate the values (and cons) of the SNR-retrieval. Only correlation coefficients of all collocated samples collected from each campaign are displayed. The ARRecon-2018 and ARRecon-2020 samples are further combined. From Fig. 9a, we can see both SNR-retrieval and Level 2 retrieval agree much better with the ground truth across all extra-tropical campaigns compared to ERA-5. The SNR-retrieval exhibits more robust correlation while Level 2 retrieval are really poor (negative correlations) at some levels. However, for the two deep-tropics campaigns, above statements do not work anymore. All three work really poor for the ATOMIC campaign with barely any correlation with the ground truth or even negative correlation for the Level 2 retrieval. For the EUREC4A campaign, things remain similar for RO retrievals no matter using Level 1 or Level 2 data, but ERA-5 works much better at capturing the humid MPBL structure in this case albeit

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

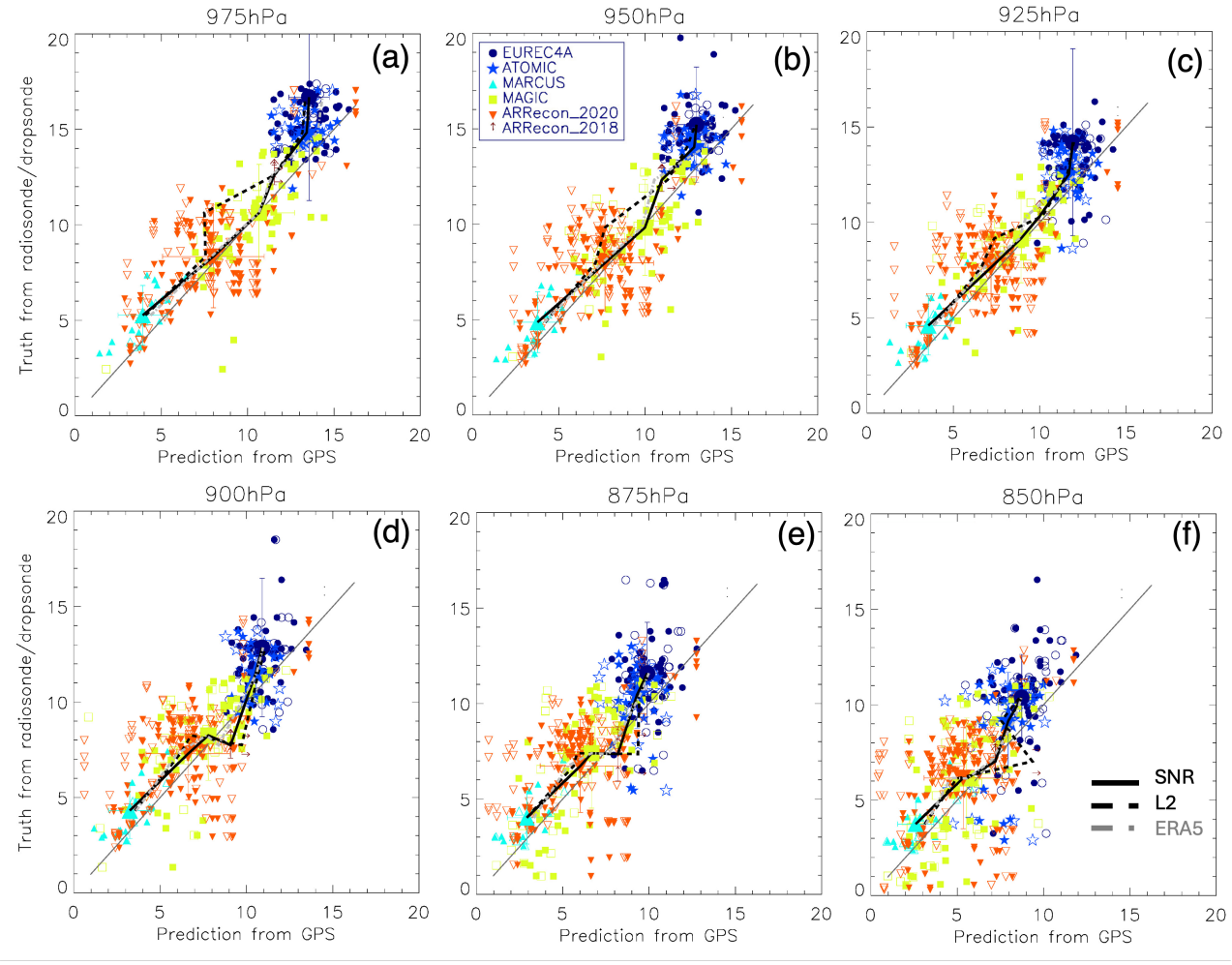


Figure 7. Scatter plots of collocated specific humidity comparison between radiosonde "truth" and retrievals from SNR (closed symbols) and Level-2 standard retrieval (open symbols) for each pressure level. The mean and standard deviation from the SNR retrieval are shown as bigger symbols. In addition, the mean retrieved values from each campaign as opposed to the mean from radiosonde "truth" are shown as the bold black line for SNR retrieval, bold black dash-dotted line for Level-2 retrievals, and grey dashed line for ERA-5.

it's still dry-biased (not shown). We achieved 40-50% across-board more collocation samples using the SNR-retrieval versus the Level 2 wetPrf retrieval, which is consistent with the general success rate shown in Fig. 1.

In Fig. 9b, the correlation statistics are binned by different pressure levels. We can see the SNR-retrieval and Level-2 retrieval are in general comparable in robustness in capturing the entire MPBL specific humidity vertical structure, both of which outperform the ERA-5 significantly. The correlation coefficients are consistently slightly higher than that from the Level-2 retrievals. 975 hPa is the only exception, where ERA-5 seems to outperform the Level-2 retrieval slightly, but SNR-retrieval is significantly better across-board than ERA-5.

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

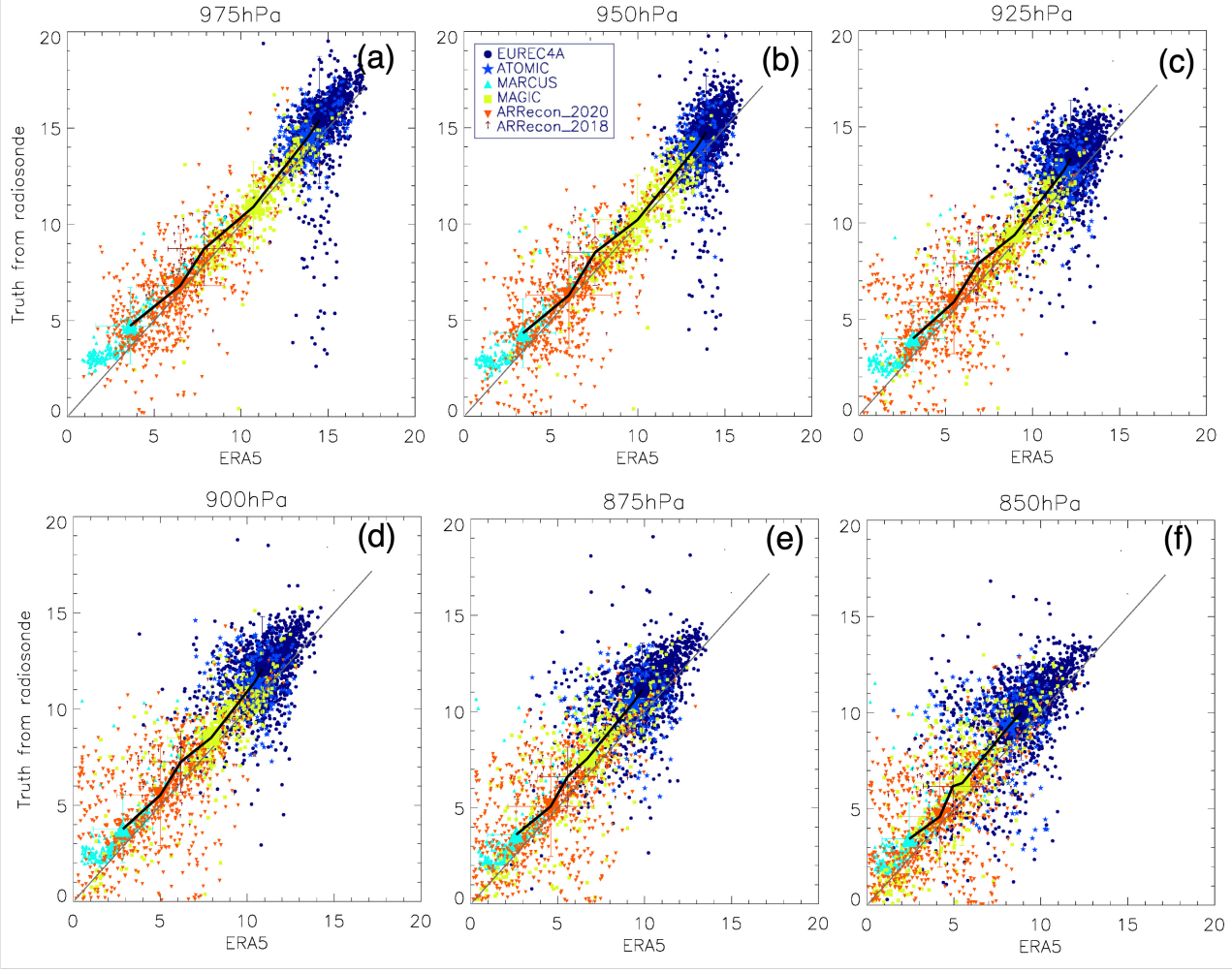


Figure 8. Same with Fig. 7, except for all available radiosonde/dropsonde samples in all these campaigns with collocated ERA-5 specific humidity. The bold black solid line connects the mean values from each campaign.

In conclusion, based on the independent comparison against limited radiosonde/dropsonde data available over the open ocean, we can draw the conclusions that our SNR-retrieval are the best at capturing the MPBL vertical structure from the surface (975 hPa) up to 850 hPa globally compared to the ERA-5 reanalysis except at the deep tropics. Compared to the operational Level-2 retrieval, the SNR method can achieve 50% more samples in the MPBL, and more stable performance globally. The reason for the poor performance of the SNR retrieval in the deep tropics is probably due to the breakdown of the original assumption: turbulent and mixing in the tropical PBL by frequent shallow convections constantly disrupt the ducting condition, causing SNR reemerging at the deep H_{SL} blending other information and hence are not useful for MPBL water vapor retrieval. The fact that SNR-retrieval statistics outperform ER-5 at all 6 pressure levels prove that real physical information from SNR observations is learnt and kept by the ML model for prediction.

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

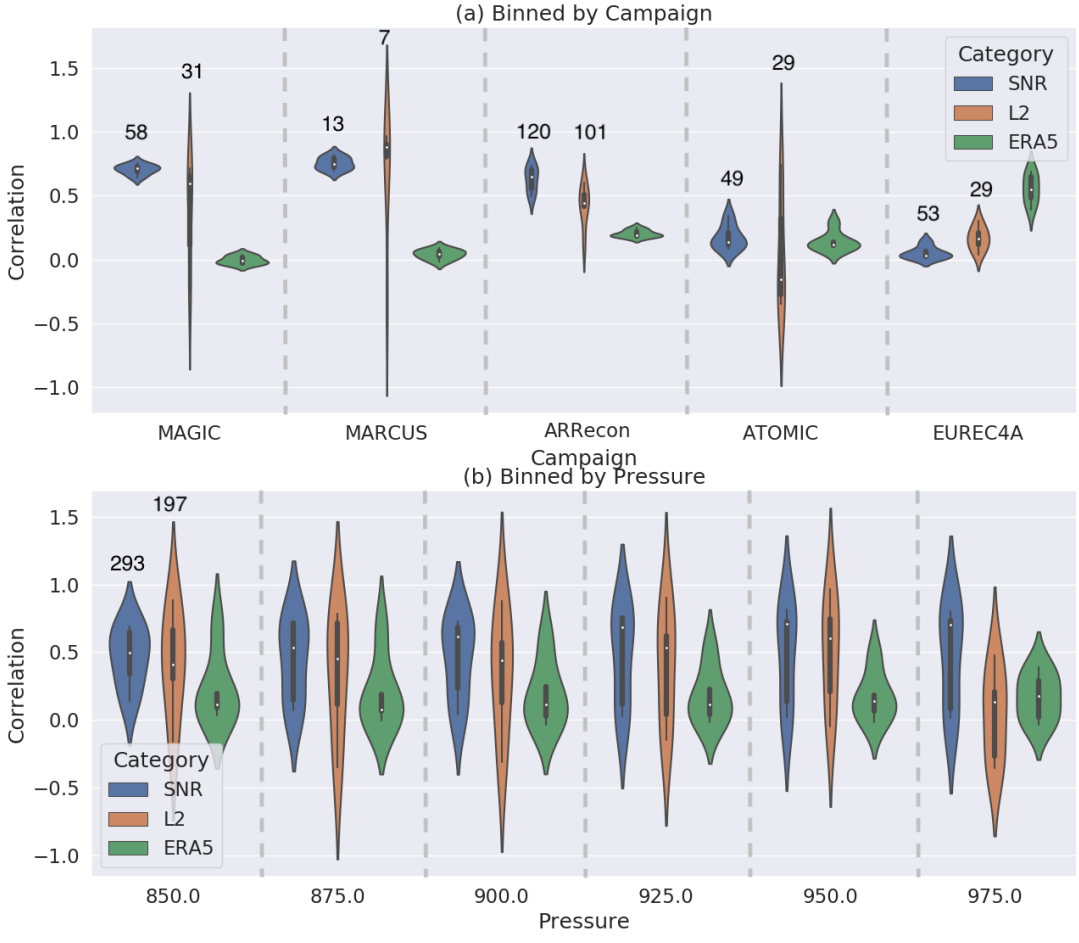


Figure 9. Violin plots of the correlation coefficients calculated from collocated samples for SNR-retrievals (blue), Level 2 retrievals (orange) and ERA-5 (green). (a) is all-level statistics for each campaign; and (b) is all campaign but binned by different pressure levels. Medium, standard deviation and minimum/maximum values are shown as the white dots, black box and extended vertical thin lines in each violin, respectively. The number of total samples are listed on top of each violin. For ERA-5, only the subset of samples that SNR-retrieval collocations available are selected to calculate the statistics.

4 Discussions

In the section, we present and discuss some use case examples in order to demonstrate how to use this SNR-retrieved MBPL specific humidity to identify model or reanalysis issues.

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

4.1 Climatology

As suggested in Johnston et al. (2021), MERRA-2 reanalysis has larger dry-biases especially in the polar regions compared to ERA-5. Here we map out the climatological distribution from our SNR-retrieved water vapor to track down geographical discrepancies in the Arctic (Fig. 10 and Antarctic (Fig. 11). The coldest months were not selected because of the concern of sea ice. Overall again we can see the SNR-retrieved polar MPBL is much more humid than that from MERRA-2 ($> 100\%$

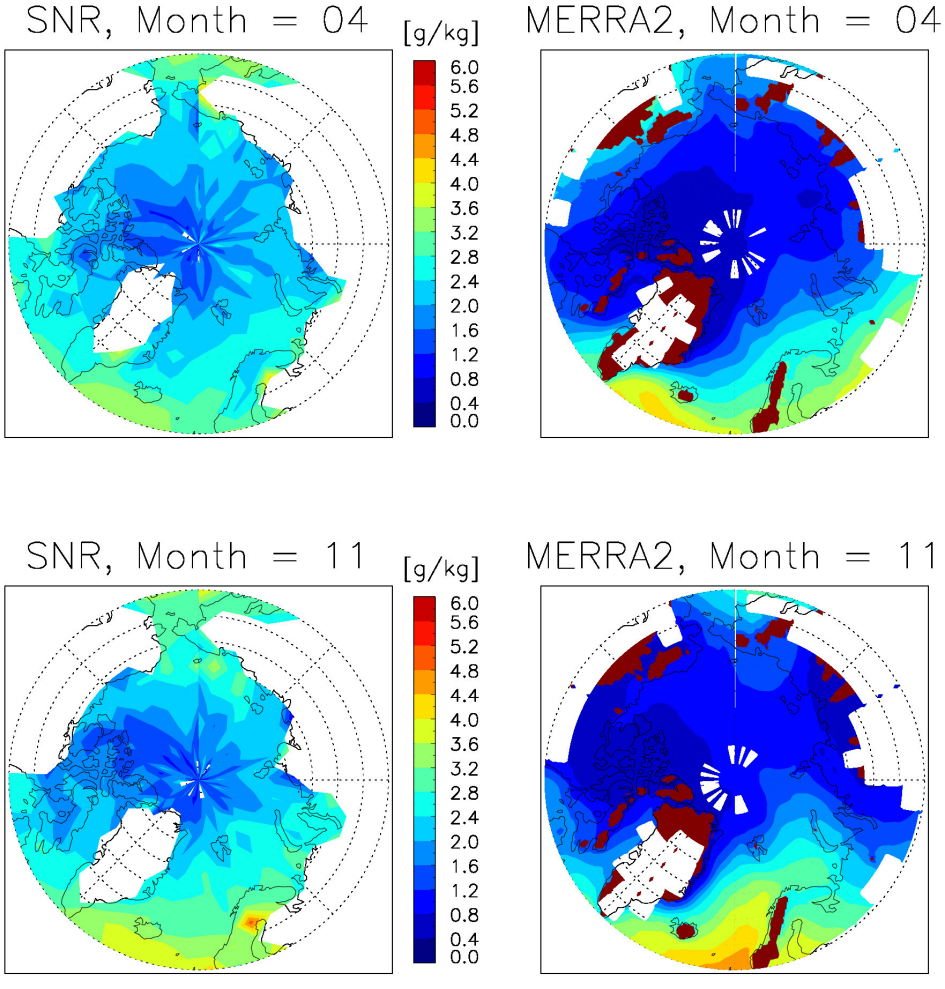


Figure 10. Monthly averages of 950 hPa specific humidity from COSMIC-1 SNR retrieval (left) compared to MERRA-2 reanalysis (right) for Arctic during April (top) and November (bottom), 2012 and 2013.

in most areas). If we neglect sampling induced geographical inhomogeneities in the SNR-retrievals, we can actually see in

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

Fig. 10 that the general patterns and gradients are in general agreeable. The largest differences are that the wet intrusion along the Bering strait seems to be too weak during both April and November in MERRA-2, and the wet intrusion associated with the North Atlantic overturning circulation seems to be too strong during November in MERRA-2. These discrepancies connect possible root causes down to the ocean circulation, and up to the Arctic front, and should be further investigated in a whole Earth-system point of view.

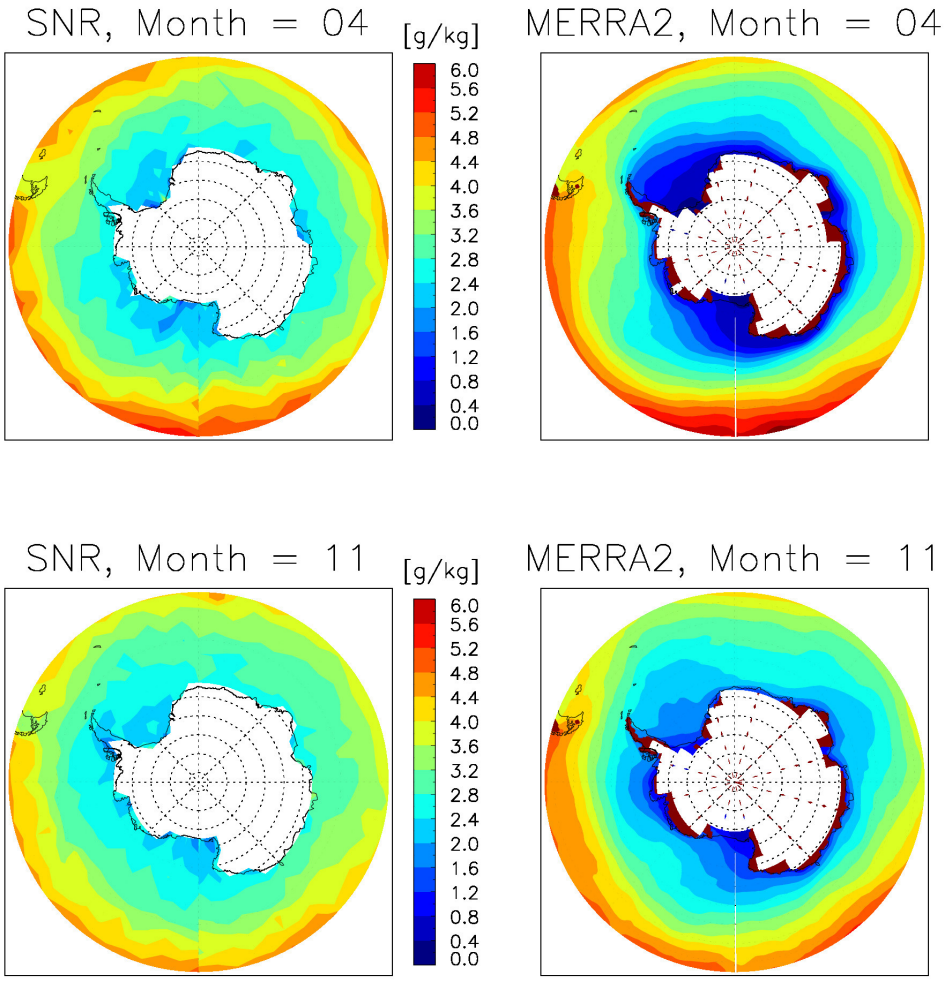


Figure 11. Same with Fig. 10, except for the Antarctic/Southern Ocean.

Although Southern Ocean and South Pole seem more boring (Fig. 11), we can actually observe some interesting potential issues related to topographies. For example, the tip of the Andes mountain effectively blocks MPBL water transport across the mountains, but such a local effect on humidity is further downstream in MERRA-2. The gradient of water vapor amount from

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

north to south is apparently much weaker compared to MERRA-2, which impacts the latent heat and sensible heat computation when considering global energy transport.

4.2 Diurnal Variation

It is well-known that global climate models (GCMs) have serious issues at reproducing the cloud, precipitation and convection diurnal cycles. Although such a notorious problem is mostly attributed to the issues with cumulus parameterization schemes, we argue that the diurnal cycle of MPBL water vapor also plays a nontrivial role as it ties closely to the shallow cumulus and stratocumulus, the latter, for example, is also closely related to the MPBL height diurnal variation (e.g.,Liu and Liang (2010),Teixeira et al. (2021)). Ground truth of the diurnal variation of MPBL water vapor structures is extremely rare, probably because of the high cost associated with long-duration shipborne campaign that often only launches radiosondes twice daily and hence cannot capture the diurnal variabilities. Therefore, here we only aim at showing the discrepancies between ERA-5 and our SNR-retrieval generated diurnal cycle rather than determining which is right.

Several campaign regions and corresponding months are selected motivated by the observed diurnal variations of the MPBL height established in Liu and Liang (2010). These 4 regions include tropical east Pacific (EPIC campaign), stratocumulus region off the coast of Chile (VOCALS campaign), South Indian ocean (INDOEX campaign), and the Arctic open ocean. The last one was added for the sole interest to check if there is any diurnal cycle in the coldest season.

The averaged specific humidity at $950\ hPa$ agrees well between two dataset in the two Pacific campaigns, but ERA-5 is wetter in the South Indian ocean, and significantly drier in the Arctic ocean. Compared to the SNR-retrieved diurnal cycle, MPBL water vapor diurnal cycle in ERA-5 is too weak in 3 areas but not the INDOEX campaign region. They seem to be slightly off-phase as well. To put into context of the diurnal cycle of PBL height (Liu and Liang (2010)), in the EPIC campaign region, the SNR-retrieved water vapor show a dip in the morning (8-9 am local time), coinciding with the dip of PBL height in this region. Then the water vapor continues to build up throughout the day, while the PBL height reaches maximum at local noon and then drops down. In the VOCALS campaign stratocumulus region, the water vapor diurnal cycle derived from the SNR method echos the PBL height change very well (Liu and Liang (2010), Fig. 10), indicating the important role water vapor plays in driving the diurnal cycle of stratocumulus, but ERA-5 barely shows any diurnal signature. In the INDOEX campaign region, the diurnal cycle from ERA-5 and SNR method agrees reasonably well, both anti-correlated with the diurnal cycle of PBL height change observed during that campaign. In the Arctic ocean, ERA-5 apparently has set some arbitrary threshold to keep the water vapor at a constant low level, while SNR-retrievals suggest a weak diurnal variation.

Overall from this exercise, we can see the diurnal coupling between MPBL water vapor, PBL height and clouds are vastly different from area to area. However, ERA-5 likely under-produced the diurnal cycle of MPBL water vapor globally. For SNR-retrievals, day-to-day variability is overwhelmingly large, making it harder to interpret or trust a diurnal cycle. Ultimately, the lack of MPBL water vapor ground "truth" measurements will keep this topic foggy. Other shipborne measurements, e.g., upward pointing radiometers, might be helpful to dientangle this mystery in the future.

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

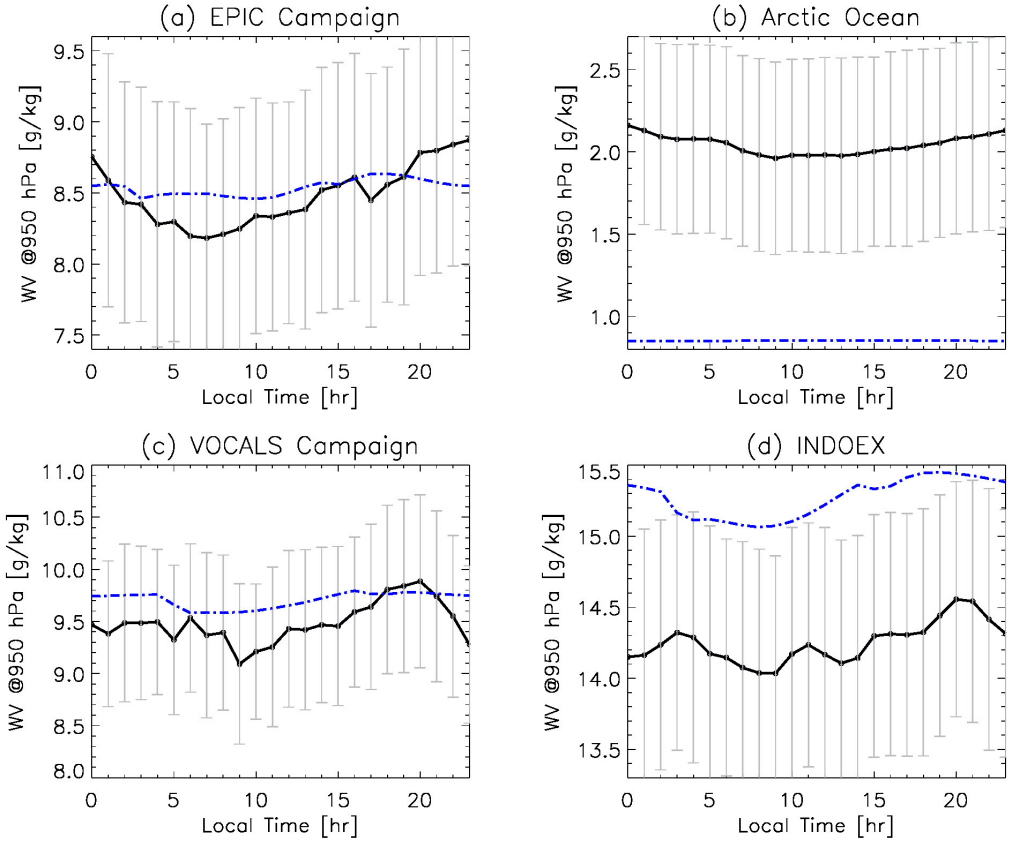


Figure 12. Multi-year mean diurnal variation of 950 hPa specific humidity retrieved from COSMIC-1 and COSMIC-2SNR (black with errorbar in grey) and from ERA-5 hourly reanalysis (dash-dotted blue) during December - February for (a) EPIC campaign region, $130°W - 120°W$, $15°N - 30°N$; (b) Arctic ocean, $180°W - 180°E$, $70°N - 90°N$; (c) VOCALS campaign region, $90°W - 74°W$, $28°S - 24°S$; (d) INDOEX campaign region, $55°E - 75°E$, $25°S - 15°S$.

5 Conclusions

Marine planetary boundary layer (MPBL) water vapor amount and gradient are among the key factors to couple the ocean and atmosphere cloud, precipitation and convection together, but meanwhile it is also among the hardest object to retrieve from satellite remote sensing perspective. Given the penetration capability of GNSS signal through clouds, we proposed a novel way in Wu et al. (2022) to utilize the GNSS signal-to-noise (SNR) ratio in the deep H_{SL} to retrieve MPBL water vapor profiles. In this paper, we demonstrated it is workable at profile-by-profile level, leveraging the power of machine learning (ML) in capturing weak and non-linear signals. The surprising and novel findings in this paper, is that the ML-trained model can make better predictions that outperforms the training dataset (i.e., ERA-5), which demonstrates the real information content in the SNR signal which would otherwise not be harnessed using traditional statistical methods. The new SNR-retrievals has more

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

stable performance against the operational Level-2 GNSS-RO retrievals, and it can produce 50% more successful retrievals in the lowest $1\ km$ where observations are critical to understand ocean-atmosphere coupling.

We then showed two use cases to demonstrate possible ways to use this dataset. There is no conclusive results because of lack of ground "truth" to validate, but we do find both reanalysis tend to systematically produces dry biases at high-latitudes, and too weak diurnal cycles over global oceans. This SNR-retrieval dataset also has its own caveats. Whenever the "ducting" condition is violated (e.g., coastal topography, convective tower, mixing and turbulence in the MPBL), the fundamental assumption breaks down, resulting in poor performances. More extensive comparisons and validations against other high-quality ground measurements are needed in the future.

Data availability. The Level 2 SNR-retrieval product for the prediction period (see Table 1) has been published on zenodo (Gong et al. (2024)). We welcome use and feedbacks.

COSMIC1, COSMIC2, Metop-A and Metop-B Level 1 and Level 2 atmPrf and wetPrf data are downloaded from https://data.cosmic.ucar.edu/gnss-ro/. ATOMIC data are downloaded from https://psl.noaa.gov/atomic/data/. EUREC4A data are downloaded from https://doi.org/10.25326/137. SOCRATES data are downloaded from https://data.eol.ucar.edu/master_lists/generated/socrates/. MARCUS data are downloaded from ARM data request portal. MAGIC data are downloaded from ARM data request portal. ARRecon data are downloaded from https://ARRecon.ucsd.edu/arrecon_data/ specially processed to fit the needs of this research. Interested users are encouraged to contact the last author for assistants of post-processed data.

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

Appendix A: A

Table A1. Summary of GNSS-RO instrument noise (σ) used in this work

Instrument Name	Orbit	Noise (σ)
COSMIC-1/C1	Ascending	10.1
	Descending	10.9
COSMIC-1/C2	Ascending	10.2
	Descending	10.9
COSMIC-1/C3	Ascending	9.6
	Descending	10.4
COSMIC-1/C4	Ascending	10.6
	Descending	11.2
COSMIC-1/C5	Ascending	10.1
	Descending	11.1
COSMIC-1/C6	Ascending	9.2
	Descending	10.7
COSMIC-2/E1	Ascending	17.0
	Descending	17.5
COSMIC-2/E2	Ascending	17.5
	Descending	17.8
COSMIC-2/E3	Ascending	17.2
	Descending	17.9
COSMIC-2/E4	Ascending	17.5
	Descending	17.7
COSMIC-2/E5	Ascending	17.4
	Descending	17.8
COSMIC-2/E6	Ascending	17.5
	Descending	17.8

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

Table B1. Excessive Phase L1 grid for this work

Parmeter	Grid values
$Log_{10}(\phi_{L1})$	1.26245, 1.33846, 1.41162, 1.48144, 1.54777, 1.62428, 1.69679, 1.76530, 1.82995, 1.89098, 1.94866, 1.97000, 2.00325, 2.02000, 2.05500, 2.08000, 2.10415, 2.13000, 2.15091, 2.17000, 2.19548, 2.23805, 2.27875, 2.30103, 2.32222, 2.37000, 2.41497, 2.44000, 2.55630, 2.59000, 2.63000, 2.69020, 2.75000, 2.81291, 2.86000, 2.92428, 2.95000, 3.02531, 3.10000, 3.11727, 3.15000, 3.20140, 3.22000, 3.25000, 3.27875, 3.30000, 3.32000, 3.35025, 3.41664, 3.47857, 3.53656, 3.59106
Rough corresponding H_{SL} [km]	-150, -140, -130, -120, -110, -108, -106, -104, -102, -100, -98, -96, -94, -92, -90, -80, -70, -60, -50, -40, -37, -33, -30, -26, -23, -20, -19, -17, -15, -13, -11, -9, -7, -5, -3, -2, -1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 15, 17, 19

Author contributions.

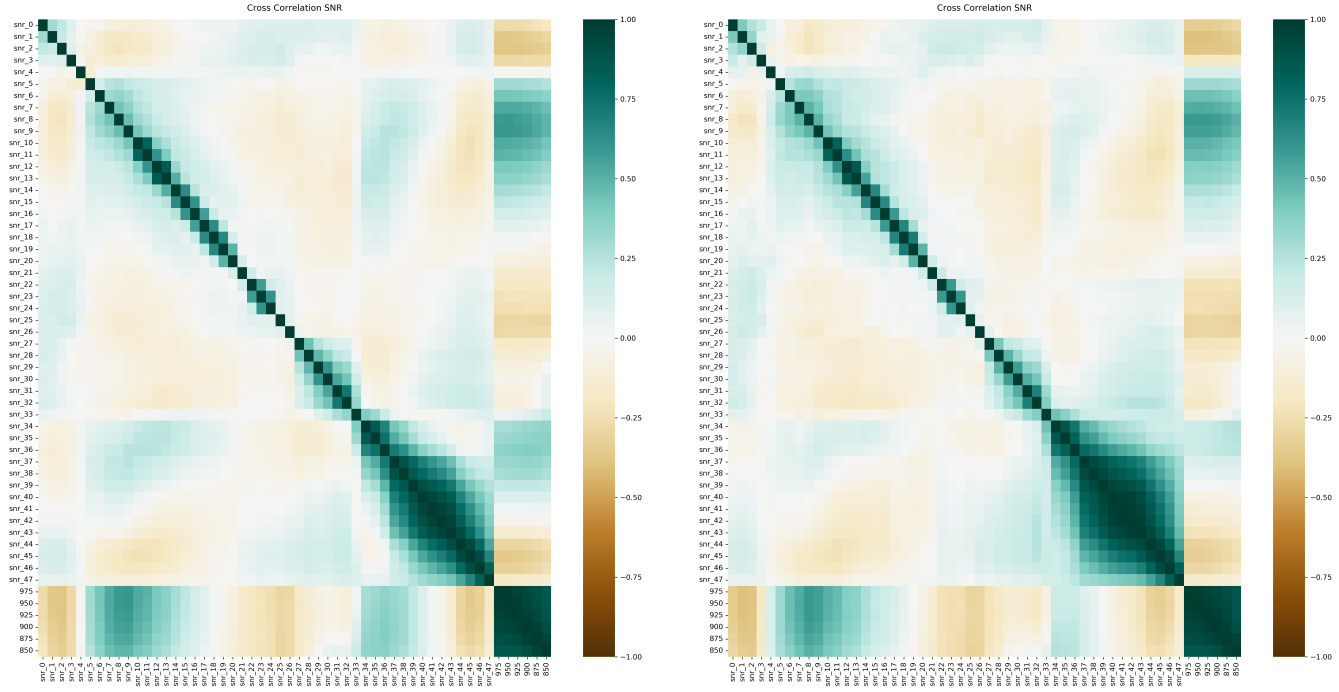
D.L.W. came up with the initial idea. J. G. designed the methodology, executed the plan, built he model, and conducted the validation and data analysis. M.B. helped conducted the hyperparameter tuning. M.G. provided Fig.1. M. Z. provided the high verical resolution AR-Recon data. All authors participated in result discussion and interpretation.

Competing interests.

The authors have no competing interests.

Acknowledgements. J.G. is grateful to Dr. Mariel Friburg at NASA Goddard in providing financial support of M.B. We thank the potential editor and reviewers to provide insightful comments.

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024



(a) Metop-A S_{RO} & MPBL specific humidity

(b) Metop-B S_{RO} & MPBL specific humidity

Figure A1. Cross-correlation among S_{RO} (left) and σ^2_{SNR} (right) at various SLH levels from one random day for COSMIC1 and their cross-correlation with collocated ERA5 specific humidity at 975 - 850 hPa (bottom-right corner of each panel). Only grid indices are shown in the axis titles, and the corresponding $Log_{10}(\phi_{L1})$ values can be found in Table A2.

References

Boisvert, L. N., Wu, D. L., Vihma,T., Susskind, J.: Verification of air/surface humidity differences from AIRS and ERA-Interim in support of turbulent flux estimation in the Arctic, J. Geophys. Res., doi:10.1002/2014JD021666, 2015.

Gong, J., Wu, D. L.: Level-2 marine planetary boundary layer (MPBL) specific humidity GNSS signal-to-noise (SNR) ratio retrieval product, Zenodo, doi:10.5281/zenodo.10899848, 2024.

Kuo, Y., Sokolovskiy,S., Anthes,R.A., and Vandenberghe, F.: Assimilation of GPS Radio Occultation Data for Numerical Weather Prediction, Space Sci., doi:10.3319/TAO.2000.11.1.157(COSMIC), 2000.

Liu, S.,Liang, X.:Observed Diurnal Cycle Climatology of Planetary Boundary Layer Height,J. Clim., doi:10.1175/2010JCLI3552.1, 2010.

Millán, L. F., Lebsock, M. D., and Teixeira, J.: Variability of bulk water vapor content in the marine cloudy boundary layers from microwave and near-infrared imagery, Atmos. Chem. Phys., 19, 8491–8502, doi:10.5194/acp-19-8491-2019, 2019.

Wu, D.L.,Gong, J., and Ganeshan, M.: GNSS-RO Deep Refraction Signals from Moist Marine Atmospheric Boundary Layer (MABL), Atmosphere, doi:10.3390/atmos13060953,2022.

Ye, J., Liu, L., Wang, Q., Hu, S. and Li, S.: A Novel Machine Learning Algorithm for Planetary Boundary Layer Height Estimation Using AERI Measurement Data, IEEE Geosci., doi:10.1109/LGRS.2021.3073048, 2021.

https://doi.org/10.5194/egusphere-2024-973
Preprint. Discussion started: 20 June 2024

Milestein, A., Blackwell, B.: Neural network temperature and moisture retrieval algorithm validation for AIRS/AMSU and CrIS/ATMS, J. Geophys. Res.-Atm., doi:10.1002/2015JD024008, 2016.

Milestein, A.: Planetary Boundary layer (PBL) Final Report, Lincoln Lab, MIT, https://apps.dtic.mil/sti/pdfs/AD1166514.pdf

Teixeira, J., Piepmeier, J. R., Nehrir, A. R., Ao, C. O., Chen, S. S., Clayson, C. A., Fridlind, A. M., Lebsock, M., McCarty, W., Salmun, M., Santanello, J. A., Turner, D. A., Wang, Z., Zeng, X.: TOWARD A GLOBAL PLANETARY BOUNDARY LAYER OBSERVING SYSTEM: THE NASA PBL INCUBATION STUDY TEAM REPORT, NASA, https://ntrs.nasa.gov/api/citations/20230001633/downloads/AFridlindPBLTowardsReport.pdf, 2021.

Xu, X., Zou, X.: COSMIC-2 RO Profile Ending at PBL Top with Strong Vertical Gradient of Refractivity, doi:10.3390/rs14092189, 2022.

Ganeshan, M., D. L. Wu, J. Gong: Lower atmospheric sounding capability from Spire and operational GNSS RO products for Arctic PBL studies (submitted).

Johnston, B.R.; Randel, W.J.; Sjoberg, J.P.: Evaluation of Tropospheric Moisture Characteristics Among COSMIC-2, ERA5 and MERRA-2 in the Tropics and Subtropics. Remote Sens., doi:10.3390/rs13050880, 2021.

Milstein, A., Santanello, J. A., Blackwell, B.: Detail Enhancement of AIRS/AMSU Temperature and Moisture Profiles Using a 3D Deep Neural Network, Artificial Intelligence in Earth Science, doi:10.1175/AIES-D-22-0037.1, 2023.

Chang, H., Lee, J., Yoon, H., Morton, Y. J., and Saltman, A.: Performance assessment of radio occultation data from GeoOptics by comparing with COSMIC data, Earth, Planets and Space, doi:10.1186/s40623-022-01667-6, 2022.

Zhran, M.: An evaluation of GNSS radio occultation atmospheric profiles from Sentinel-6, The Egypsian J. of Remote Sensing and Space Sci., doi:10.1016/j.ejrs.2023.07.004, 2023.

Vömel, H., Brown, W.: SOCRATES -2018 Radiosonde Data Quality Report, UCAR/NCAR - Earth Observing Laboratory, University Corporation for Atmospheric Research, doi:10.5065/D69P30HG, 2018.

Keeler, Evan; Burk, Ken; Kyrouac, Jenni, ARM - Balloon-borne sounding system (BBSS) WNPN output data, sondewnpn.b1, URL=https://doi.org/10.5439/1595321

George, G., Stevens, B., Bony, S., Pincus, R., Fairall, C., Schulz, H., Kölling, T., Kalen, Q. T., Klingebiel, M., Konow, H., Lundry, A., Prange, M., and Radtke, J.: JOANNE: Joint dropsonde Observations of the Atmosphere in tropical North atlaNtic meso-scale Environments, Earth Syst. Sci. Data, 13, 5253–5272, doi:10.5194/essd-13-5253-2021, 2021.

Stevens, B., and co-authors: EUREC4A, Earth Syst. Sci. Data, doi:10.5194/essd-13-4067-2021, 2021.

Stephan, C. C., Schnitt, S., Schulz, H., Bellenger, H., de Szoeke, S. P., Acquistapace, C., Baier, K., Dauhut, T., Laxenaire, R., Morfa-Avalos, Y., Person, R., Quiñones Meléndez, E., Bagheri, G., Böck, T., Daley, A., Güttler, J., Helfer, K. C., Los, S. A., Neuberger, A., Röttenbacher, J., Raeke, A., Ringel, M., Ritschel, M., Sadoulet, P., Schirmacher, I., Stolla, M. K., Wright, E., Charpentier, B., Doerenbecher, A., Wilson, R., Jansen, F., Kinne, S., Reverdin, G., Speich, S., Bony, S., and Stevens, B.: Ship- and island-based atmospheric soundings from the 2020 EUREC4A field campaign, Earth Syst. Sci. Data, 13, 491–514, doi:10.5194/essd-13-491-2021, 2021.

Zheng, M., Torn, R., Delle Monache, L., Doyle, J., Ralph, F. M., Tallapragada, V., Davis, C., Steinhoff, D., Wu, X., Wilson, A. M., Papadopoulos, C., Mulrooney, P. (2024). An Assessment of Dropsonde Sampling Strategies for Atmospheric River Reconnaissance. Monthly Weather Review, doi:10.1175/MWR-D-23-0111.1, 2024.

Nitish Srivastava and Geoffrey Hinton and Alex Krizhevsky and Ilya Sutskever and Ruslan Salakhutdinov: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, Journal of Machine Learning Research, http://jmlr.org/papers/v15/srivastava14a.html, 2014.

Yarin Gal and Zoubin Ghahramani: Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, Proceedings of the 33 rd International Conference on Machine Learning, https://proceedings.mlr.press/v48/gal16.pdf