Received: 10 August 2023 | Revised: 21 September 2024 | Accepted: 8 October 2024

DOI: 10.1002/gdj3.282

Geoscience Data Journal | RMetS | WILEY

DATA ARTICLE

A multi-tagged SAR ocean image dataset identifying atmospheric boundary layer structure in winter tradewind conditions

Chen Wang[1,2] | **Justin E. Stopa**[3] | **Doug Vandemark**[4] | **Ralph Foster**[5] | **Alex Ayet**[6] | **Alexis Mouche**[7] | **Bertrand Chapron**[7] | **Peter Sadowski**[8]

[1]School of Marine Sciences, Nanjing University of Information Science & Technology, Nanjing, China

[2]The Key Laboratory of Space Ocean Remote Sensing and Application, Ministry of Natural Resources, Beijing, China

[3]Department of Ocean Resources and Engineering, University of Hawai'i at Mānoa, Honolulu, Hawaii, USA

[4]Ocean Process Analysis Laboratory, University of New Hampshire, Durham, New Hampshire, USA

[5]Applied Physics Laboratory, University of Washington, Seattle, Washington, USA

[6]CNRS, Université Grenoble Alpes, Inria, Grenoble INP, GIPSA-Lab, Grenoble, France

[7]Laboratoire d'Océanographie Physique et Spatiale (LOPS), Univ. Brest, CNRS, IRD, IFREMER, Brest, France

[8]Department of Information and Computer Science, University of Hawai'i at Mānoa, Honolulu, Hawaii, USA

Correspondence
Justin E. Stopa, University of Hawai'i at Mānoa, Honolulu, HI, USA.
Email: stopa@hawaii.edu

Funding information
National Natural Science Foundation of China, Grant/Award Number: 42206179 and 42350003; NASA, Grant/Award Number: 80NSSC20k0822 and 80NSSC23K0978; ESA Sentinel-1 Mission Performance Center, Grant/Award Number: 4000107360/12/I-LG; MITI, ALESE; European Space Agency, Grant/Award Number: 4000135827/21/NL; CNES, Grant/Award Number: I-CASCADE; National Science Foundation, Grant/Award Number: 2132150

Abstract

A dataset of multi-tagged sea surface roughness synthetic aperture radar (SAR) satellite images was established near Barbados from January to June 2016 to 2019. It is an advancement of the Sentinel-1 Wave Mode TenGeoP-SARwv (a labelled SAR imagery dataset of 10 geophysical phenomena from Sentinel-1 wave mode) dataset that targets SAR marine atmospheric boundary layer (MABL) coherent structures. Twelve tags define roll vortices, convective cells, mixed rolls and convective cells, fronts, rain cells, cold pools and low winds. Examples are provided for each signature. The final dataset is comprised of 2100 Sentinel-1 wave mode SAR images acquired at 36 incidence angle over an 8°×8°region centered at 51° W, 15° N. Each image is tagged with one or multiple phenomena by five experts. This strategy extends the TenGeoP-SARwv by identifying coexisting phenomena within a single SAR image and by the addition of mixed roll/cell states and

Dataset details Identifier: https://doi.org/10.17882/93947

Creator: Chen Wang, Justin E. Stopa, Doug Vandemark, Ralph Foster, Alex Ayet, Alexis Mouche and Bertrand Chapron

Title: A multi-tagged Sentinel-1 wave mode SAR image dataset near Barbados (MulTags-SARwv)

Publisher: Sea scientific open data publication (SEANOE)

Publication year: 2023

Resource type: Dataset

Version: 2.0 (following on the TenGeoP-SARwv: https://doi.org/10.17882/56796)

cold pools. The dataset includes PNG-formatted SAR image files along with two text files containing the file name, the central latitude/longitude, expert tags for each image, and all dataset metadata. There is a high degree of consensus among expert tags. The dataset complements existing hand-labelled ocean SAR image datasets and offers the potential for new deep-learning SAR image classification model developments. Future use is also expected to yield new insights into the tradewind MABL processes such as structure transitions and their relation to the stratification.

KEYWORDS

Barbados, marine atmospheric boundary layer, rolls and cells, Sentinel-1 wave mode, synthetic aperture radar

1 | INTRODUCTION

Shallow clouds in the tropical marine atmospheric boundary layer (MABL) play a significant role in the earth's climate (Bony, 2005; Sherwood et al., 2014). The region windward of Barbados is a popular testbed to manifest cloud characteristics (Medeiros & Nuijens, 2016). Using 900 visually classified optical satellite images, Stevens et al. (2020) showed that the MABL near Barbados is dominated by four types of shallow convection patterns during the boreal winter. Bony et al. (2020) further found that the prevalent factors impacting the relative occurrence of these cloud morphologies are related to MABL wind speed and one measure of stability. While clouds are one indicator of the flow dynamics within the MABL, observed patterns have spatial variation at 20–2000 km (Bony et al., 2020; Brilouet et al., 2023; Schulz et al., 2020; Stevens et al., 2020), which is much larger than MABL turbulence scales that are typically <2 km. This dataset was developed to address this gap.

EUREC^{4}A was a recent large field campaign dedicated to improving our understanding of the shallow cumuli formation (Bony et al., 2017). It is a distant follow-on to the 1969 Barbados Oceanographic and Meteorological Experiment (BOMEX) (Holland, 1970) and the December 1972 PuertoRico Experiment (LeMone & Pennell, 1976). The principle motivation for this dataset was to examine the transitions between rolls and cells in the MABL documented during BOMEX (Grossman, 1982). We expand the image tagging to provide useful information about the MABL state that should help explain the connections between MABL processes and clouds. The SAR MABL imaging principle lies in detection of the wind-induced sea surface roughness variations, as most km-scale image variability can be attributed to the wind modulation due to organized roll vortices and convective cells spanning the MABL (Vandemark et al., 2001; Wang et al., 2020). In addition to these coherent structures, many localized convective events such as rain, fronts, cold pools, and regional winds too low to induce roughness patterns also leave distinct imprints on SAR images. The ability to monitor these atmospheric features means that SAR is a valuable resource for MABL investigation (Brown, 2000).

An important SAR advantage is its capability to monitor the MABL day and night and under nearly all weather conditions. There is no doubt that ~40 years of satellite SAR observations have greatly advanced our knowledge of sea surface winds, waves and currents (Johannessen et al., 1991; Li, 2016; Mourad et al., 2000) as well as the oceanic and atmospheric processes at km-scales (Alpers et al., 2016; Kudryavtsev et al., 2012; Li et al., 2013; Young et al., 2005). However, one limitation of using SAR data for MABL studies is that most of the previous studies focused on one particular process and were based on the analysis of a small number of selected images. SAR image interpretation can be complicated by the coexistence of multiple phenomena and/or pattern variability for a given single phenomenon within small images. For example, rain can induce various SAR signatures that depend on rain types and rates, wind speeds, and SAR sensor configurations (Alpers et al., 2016). In another example, temporal and spatial transitions between MABL rolls and cells (Atkinson & Wu Zhang, 1996; Salesky et al., 2017) often yield ambiguous features (linear vs. cellular) on SAR images. Similar difficulties are also noted when classifying shallow cloud types using optical satellite imagery, where a handcrafted image labelling approach was proposed and implemented (Schulz et al., 2020). Such an approach is consistent with the building of a 10 phenomena dataset (TenGeoP-SARwv, a labelled SAR imagery dataset of 10 geophysical phenomena from Sentinel-1 wave mode) using European's Sentinel-1 (S-1) wave mode (WV) SAR data (Wang, Mouche, et al., 2019).

This paper documents an extension of ocean SAR image tagging for the winter tradewind MABL near

Barbados. Twelve tags are defined to describe MABL features commonly observed in this region. Our preliminary motivation is to extend tagging beyond single classifications and expand the range of MABL coherent structures. A new MABL dataset that can complement recent surface layer and boundary layer profile measurements is thus established following the TenGeoP-SARwv. Two thousand one hundred representative S-1 WV SAR images were selected and labelled (tagged) by hand to illustrate the procedures and scientific potentials. Because we have found that the MABL state often changes at sub-20 km scales, one or multiple tags are assigned to each image. Five subject experts independently tagged each image in the dataset. Details about the Sentinel-1 satellite data and the study area are provided in Section 2. In Section 3, we define and illustrate MABL-related scene tags including their distinguishing features. The dataset creation procedure and organization are given in Section 4, followed by the summary and perspective in Section 5.

2 | S-1 WV SAR DATA NEAR BARBADOS

S-1 WV is the default one of SAR image acquisition mode over the global open ocean. Since 2016, the S-1A (launched in 2014) and S-1B (launched in 2016) satellites, respectively, collect more than 60k images per month at two incidence angles of WV1 (~23.5°) and WV2 (~36.5°). These two satellites are designed to fly the same orbital plane but with a 180° phase difference, which results in an effective 6-day repeat cycle. Unfortunately, the S-1B mission ended permanently on 23 December 2021. A twin S-1C satellite is scheduled to launch soon as a complement. The WV images are 20 km by 20 km with a 5 m pixel resolution. Compared to WV1, WV2 at the larger incidence angle is often more favorable for MABL observation because it has higher detectability of the MABL coherent structure sea surface roughness changes (Wang et al., 2020). Thus, only WV2 images are included in this study. There are ~12,000 images collected over the Barbados region (10°–20° N, 45°–64° W) by both satellites between 2016 and 2019 as displayed in Figure 1 (a). Not all images were acquired at VV polarization and those over land or of low quality (the mean Normalized Radar Cross Section is below the noise level quantified by the Noise Equivalent Sigma Zero of −22 dB) are filtered out using the same criteria as in Wang, Tandeo, et al. (2019).

A preliminary synopsis of SAR-observed MABL conditions in this region can be made with the previously developed CMwv classification tool that is designed to detect 10 ocean and atmospheric phenomena (Wang, Tandeo, et al., 2019). Of the 10, rolls, cells, rains, low winds, and atmospheric fronts belong to MABL

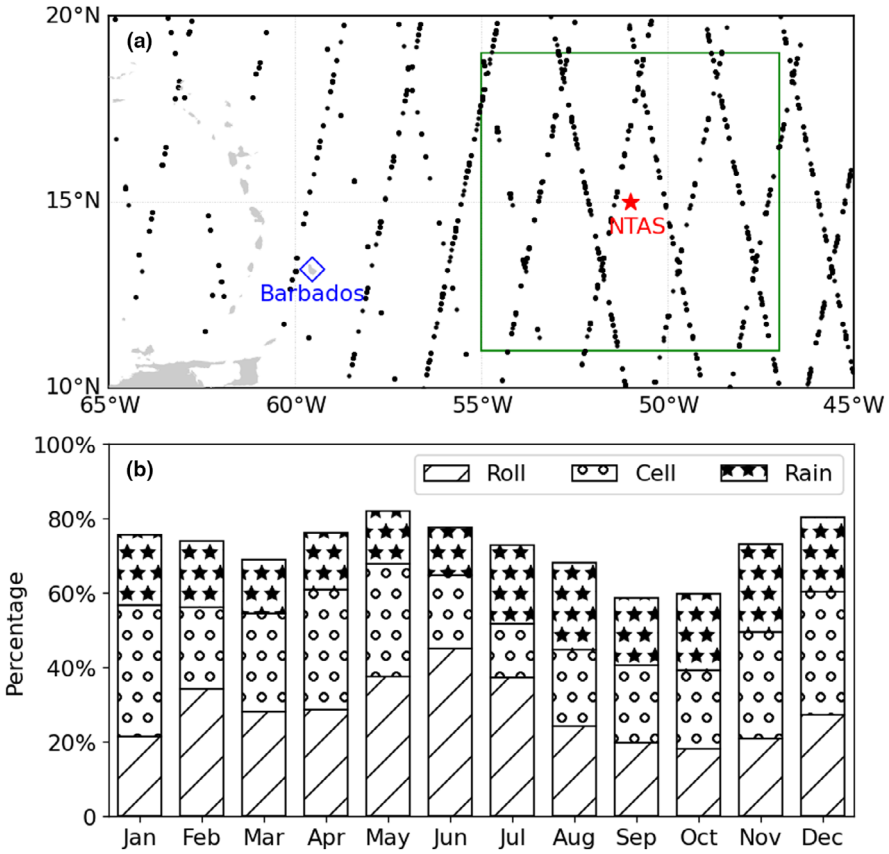

FIGURE 1 (a) WV2 VV SAR images acquired by S-1A and S-1B from 2016 to 2019 offshore of Barbados. The red star gives the location of the Northwest Tropical Atlantic Station (NTAS) buoy and the green box shows the region of interest. (b) Monthly percentages of MABL rolls, cells, and rain based on automated classification with the deep-learning model described in Wang, Tandeo, et al. (2019).

signatures. This tool performs well on the first four classes with an *F*-score larger than 0.85, while relatively poor for atmospheric fronts. Note that the *F*-score is calculated from precision and recall to evaluate the CMwv predictive performance (Wang, Tandeo, et al., 2019). It ranges from 0 to 1, with a score of 1 representing perfect precision and recall. Figure 1b displays the rates of detected MABL rolls, cells, and rain for each month. They account for ~70% of the total images. In comparison, the occurrence of rolls and cells is somewhat complementary, with more cells observed in winter and more rolls in summer. CMwv was designed to identify the single MABL process that dominates an entire 20 km scale SAR WV image. Rolls and cells were also loosely defined to represent only their mature stages. To study evolving MABL morphologies or when multiple MABL signatures occur in one image, a more detailed classification approach is necessary. The study area and temporal period from January to June 2016 to 2019 were chosen to align with ongoing field and modeling studies of the dry-convective winter tradewinds and observed low-level cloud variation in this environment (Bony et al., 2017, 2020; Stevens et al., 2020).

As a follow-on and extension dataset of TenGeoP-SARwv, we hand labelled 2100 randomly selected images near the Northwest Tropical Atlantic Station (NTAS) buoy (see Figure 1 and Section 4.1). These phenomenological tags defined in the next section cover all the MABL-relevant patterns and local features for this region of interest. Indeed, these 2100 SAR images might be too few to be directly applied in training a deep machine learning method. This dataset is built to illustrate the tagging rules and the interpretations of MABL-relevant phenomena on SAR images. These categories and tagging procedures shall help guide the related SAR image classification projects.

3 | THE MABL-RELATED PHENOMENA AND TAGS

After inspection of numerous SAR images and several iterations within a team of five subject experts, 12 tags (listed in Table 1) were determined to encompass all phenomena within the region. They can be divided into three groups. The first group contains coherent structures of the MABL that typically prevail over the whole 20 km by 20 km WV image. The second group contains more localized features (scale < 5 km) associated with boundary layer convective events. The third group includes two additional tags for the appearance and uniformity of MABL coherent structures on SAR images. Definitions of these classes are built from the TenGeoP-SARwv tags (Wang, Mouche, et al., 2019), but now with more explicit differentiation among the MABL morphology. We provide three supporting image examples for each tag, where the key features are annotated with white lines and/or curves.

TABLE 1 List of the short and long names of these 12 defined tags.

ID	Short name	Long name
01	WS	Wind streaks or MABL rolls
02	WS > MC	Dominated more by WS than MC
03	WS ~ MC	Dominated nearly equal by WS and MC
04	WS < MC	Dominated more by MC than WS
05	MC	Micro convective cells
06	AB	Air-mass boundaries
07	GF	Gust fronts
08	CP	Microscale cold pools
09	RC	Rain cells
10	LW	Low winds
11	NV	Negligible atmospheric variability
12	UN	Uniformity

3.1 | Coherent structures

The coherent structures visible on SAR images are quasi-periodic features induced by wind perturbations associated with MABL rolls/cells. Normally, their length scales are distinct from the obvious ocean swell, which has much smaller wavelengths. We focus on coherent structures with wavelengths between 0.8 and 5 km. Ocean waves and swell are present in nearly all SAR images and generally coexist with the MABL phenomena. First coherent MABL features are identified in each SAR image. Thereafter, five tags are proposed to characterize the dominance of rolls and cells and their transition, which is inspired by the rolls/cells categorizing reported by Grossman (1982).

3.1.1 | 01—WS (Wind streaks or MABL rolls)

This tag is applied if an area of at least >30% of the entire SAR image is dominated by linear parallel streaks. The streaks can be faint and multi-scale within 0.8–5 km. This tag is usually observed over the whole image or on one side of AB or GF. The streaking has a clear prevalent direction as annotated by the thin white straight lines in Figure 2. Streaks can be also corrugated or varicose. However, there must be no cellular signatures (i.e., like Micro convective cells (MC) below) visible. We do not tag for WS within the cold pools (CP, see below).

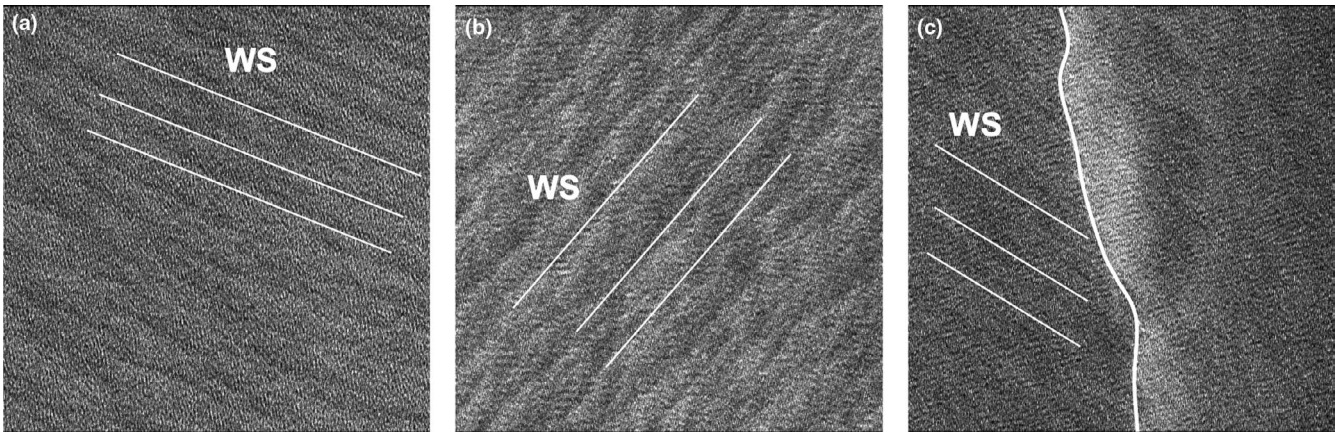

FIGURE 2 Examples of WS tags: (a) and (b) are across the entire image and (c) with GF. The thin and white straight lines indicate the dominating direction of WS. The thick and white curved line in (c) indicates the GF that gives the WS area boundary. See details in Section 3.2 on depicting GF.

3.1.2 | 02—WS > MC

This tag indicates that both WS and MC signatures are observed. The important characteristic is that the WS signatures dominate. A region (>30%) or the entire image is dominated by the combined features. The WS directionalities must be clear but weak 2D-cellular features are also evident as annotated by the thin white curved lines in Figure 3. The 2D cells are often observed billowing along the streaks as annotated by the thin white dashed arcs. This scene is usually observed over the whole image or one side of AB or GF, yet not for small regions like the inside of CPs.

3.1.3 | 03—WS~MC

A region (>30%) or the entire image is continuous for both WS and MC features. This is depicted with the thin white curved lines and cell collisions in Figure 4. The streak directions are modulated by the 2D-cells, showing more circular features than Figure 3, tag 02. The MC signature has a popcorn-like texture with round but irregular shapes and multiple scales. The key distinction is that either WS or MC is dominant. It is usually over the whole image or one side of AB or GF, yet not for small regions like the inside of CPs.

3.1.4 | 04—WS < MC

A region (>30%) or the entire image is dominated by cellular patterns. They are visible as regular ellipses with a consistent directional orientation as annotated in Figure 5. Their scales can be multiple and are typically within 0.8–3 km. WS signature is still distinguishable despite being weaker and somewhat blurred by the MC features. It must be possible to identify the WS direction as annotated by the thin white dashed and curved lines. This tag implies MC signatures dominate compared to WS. This is usually seen over the whole image or one side of AB or GF, yet not for small regions like the inside of CPs.

3.1.5 | 05—MC (Micro convective cells)

A region (>30%) or the entire image is dominated by the cellular modulations that resemble a honeycomb. The dark-bright patterns highlighted are quite regular without a discernible orientation as annotated in Figure 6. Their scale can be multiple from 0.8 to 3 km but is nearly consistent within one image. There are no evident WS features and no directionality in the MC patterns being observed. This tag is usually over the whole image or one side of AB or GF, yet not for small regions like the inside of CPs.

3.2 | Localized features

The study-defined localized signatures of MABL phenomena are mostly associated with convective events. They are frequently observed along with MABL coherent structures in a given SAR scene, and thus multiple tags are normally required for relevant images. Compared to the TenGeoP-SARwv, we distinguish further the linear feature events with tags of AB and GF. This is meaningful because they potentially indicate the same or changing MABL stratification across a frontal line. Similarly, we distinguish cold pools from rain events, as cold pools are typically associated with rain events but reflect a stage where it is not raining. Tagging of rain events and low wind conditions are similar to the definitions in TenGeoP-SARwv.

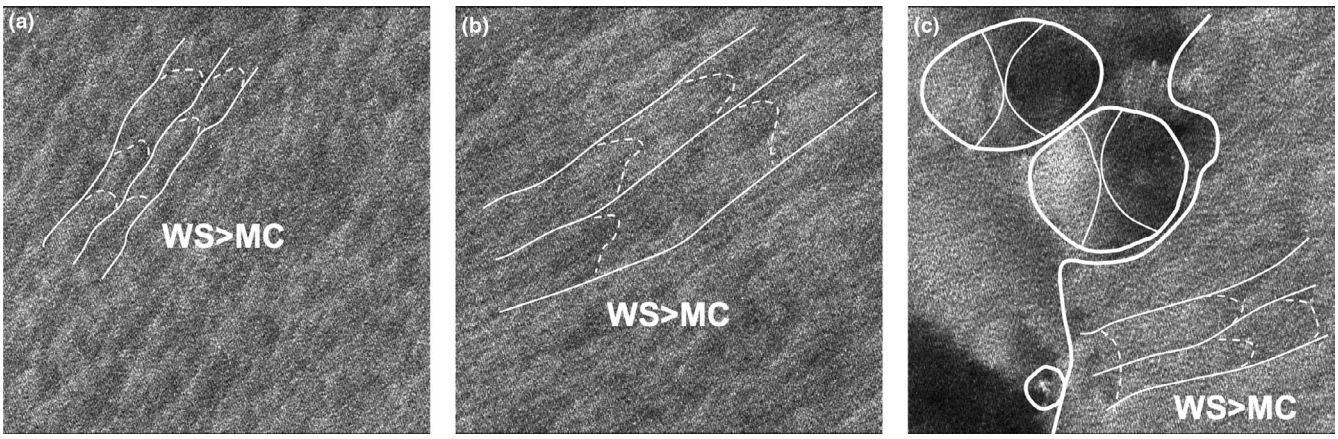

FIGURE 3 Examples of WS > MC (02) tags: (a) and (b) the whole image and (c) with GF, RC, and CP. The thin curved white lines indicate the WS directions and the 2D-celulares are annotated by the thin white dashed arcs. The thick white curved line in (c) indicates the GF that gives the WS area boundary. See details in Section 3.2 on depicting GF, Rain cells (RC), and CP.

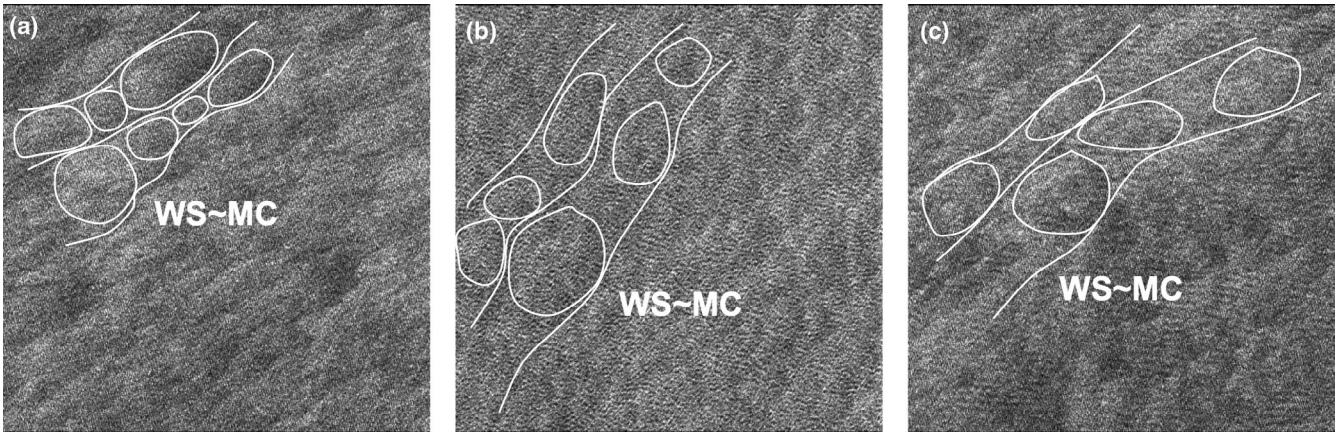

FIGURE 4 Examples of WS ~ MC tags: (a), (b), and (c) the whole image. The thin white curved lines indicate the WS directions and the thin white ovals annotate the MCs.

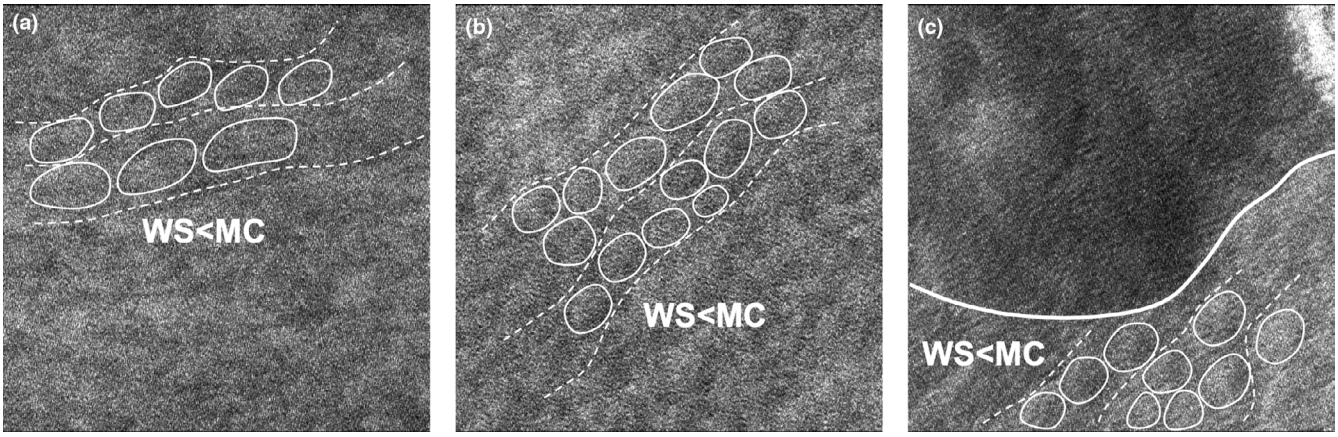

FIGURE 5 Examples of WS < MC tags: (a) and (b) the whole image and (c) with GF. The thin white ellipses indicate the 2D-cellular MC and the thin white dashed and curved lines denote the WS directions. See details in Section 3.2 on depicting GF.

3.2.1 | 06—AB (Air-mass boundaries)

They appear as sharp/ragged boundaries with separate bright and dark intensity regions. The two side areas of AB show no or the same type of coherent structure. It means that there are no clear structure changes across the AB. This tag is normally used in conjunction with one of NV or 01-05 WS/MC. Such examples are given in Figure 7.

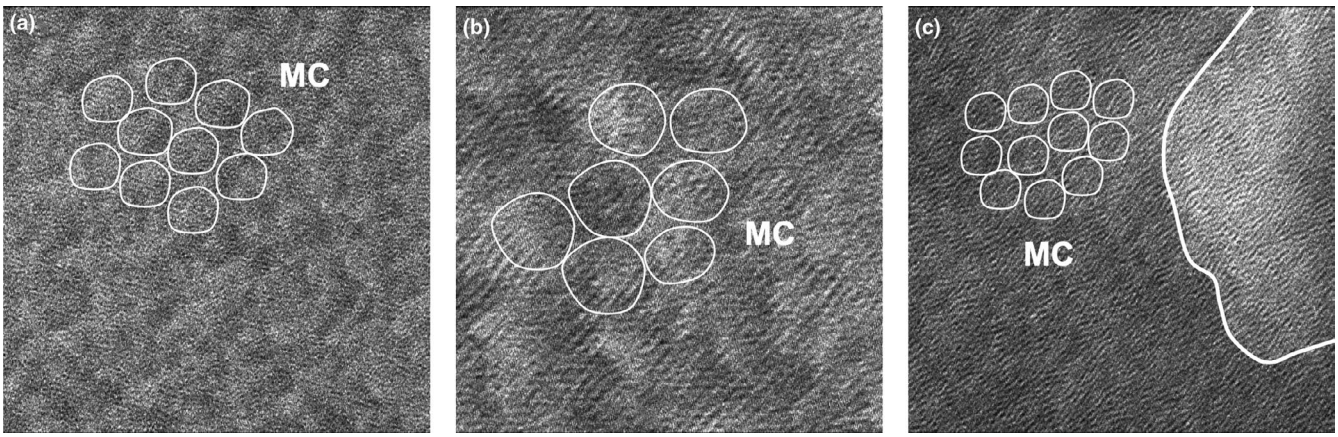

FIGURE 6 Examples of MC tags: (a) and (b) the whole image and (c) with GF (thick solid line). The thin white rounds highlight the 2D-cellular MC. See details in Section 3.2 depicting GF.

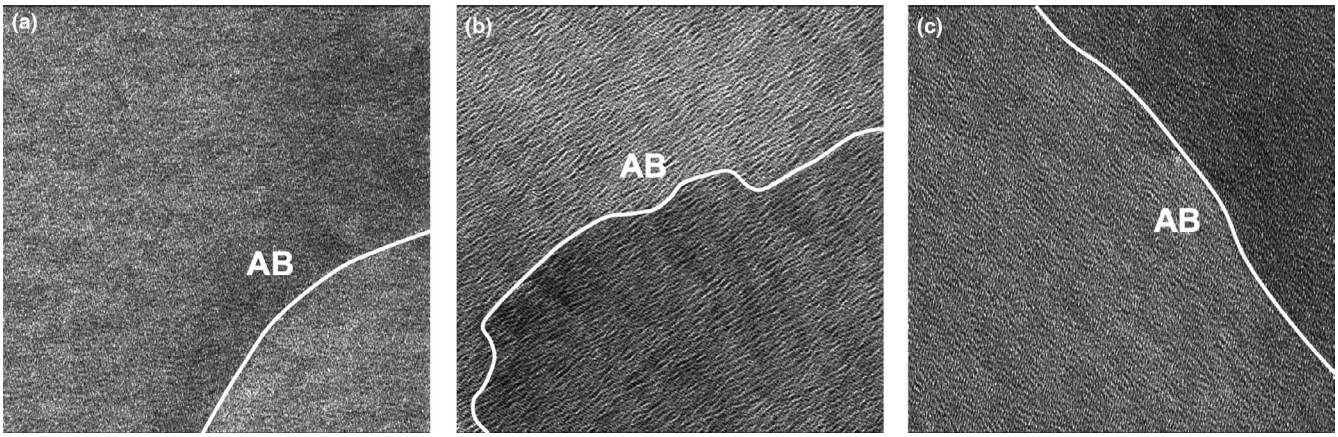

FIGURE 7 Examples of AB tags: (a) with WS > MC, (b) MC, and (c) WS > MC.

3.2.2 | 07—GF (Gust fronts)

They appear as sharp/ragged boundaries with separate bright and dark intensity regions. The two side areas of GF show changes in coherent structures (e.g., the type, scale and/or direction) as shown in Figure 8. At least one coherent structure among 01-05 WS/MC (but not NV or LW) should be tagged in conjunction with GF. It is required to tag even small regions (<30%) of NV or coherent structures behind the gust front. If tagging a GF, LW cannot be selected as a class.

3.2.3 | 08—CP (microscale cold pools)

They appear as nearly closed round areas with distinct bright and dark regions insides (i.e., the divergent zone). At least 75% of the CP outer boundary should be within the image. The combined bright and dark region looks like a 'tennis ball' as illustrated in Figure 9. The two seams of the tennis ball, which are inside the outer edge of the cold pool, have the following characteristics: (i) between the edge and the first seam, it is a bright area, almost uniform, (ii) between the two seams, the backscatter decays progressively and the image gets darker, (iii) between the second seam and the edge, the image is uniformly dark. It is common to see RC and coherent structures near the object and the CP boundaries are inherently GFs.

3.2.4 | 09—RC (Rain cells)

They appear as bright and dark patches with irregular shapes at the scales of a hundred meters to kilometers. This tag is often used with AB, GF, and CP. Typical examples are given in Figure 10.

3.2.5 | 10—LW (Low winds)

They appear as very dark areas (>70%) without any recognizable features. Typical examples are given in Figure 11. This tag is used mostly without any other tags of MABL-related phenomena.

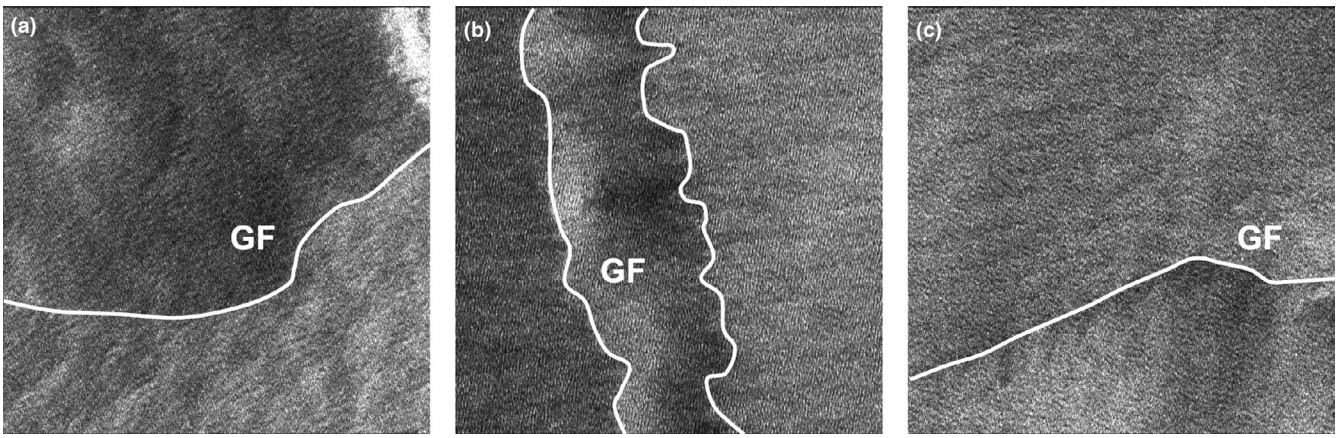

FIGURE 8 Examples of GF tags: (a) with CP, NV, and WS < MC, (b) with CPs, NV, and WS, and (c) with WS and WS > MC.

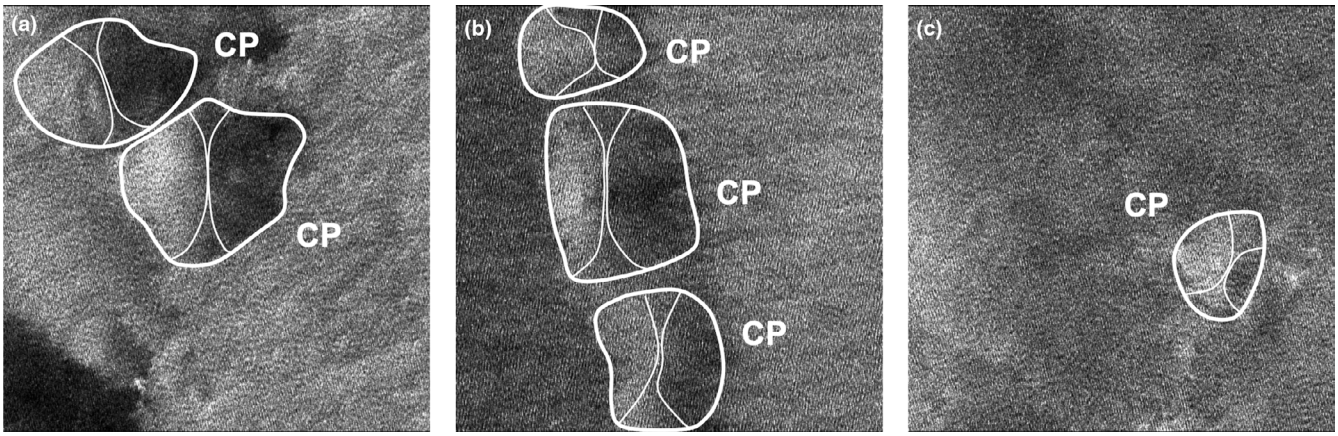

FIGURE 9 Examples of CP tags: (a) with WS ~ MC, GF, and RC, (b) with GFs, and (c) with subtle MC.

3.3 | Additional tags

3.3.1 | 11—NV (Negligible atmospheric variability)

A portion of the image (>30%), or the entire image, is dominated by ocean wave/swell ripples without any evident MABL-relevant features at scales of 0.8-5 km. The image at a low resolution looks like 'frosted glass' as shown in Figure 12a. This tag can represent an entire image or be isolated on one side of air mass boundaries (AB) or gust fronts (GF, see below) like in Figure 12b,c. It is not tagged if it covers <30% of the image.

3.3.2 | 12—UN (Uniformity)

This is not an independent tag, but an adjective tag to indicate homogeneity in the coherent structures observed across an entire WV image. The entire image must be solely dominated by NV or one of 01-05 WS/MC with an overall consistency in the grayscale intensity. There is no presence of other oceanic and/or atmospheric phenomena aside from ocean swell The UN tag is applied for the above image examples of Figures 2a,b, 3a,b, 4a-c, 5a,b, 6a,b, and 12a.

4 | DATASET CREATION

Eight experts created the dataset, among which Alexis Mouche (AM) and Bertrand Chapron (BC) provided the SAR images and fruitful comments regarding the tag definition and potential applications. Peter Sadowski (PS) suggested the tagging tool of Labelbox and helped define rules to ensure consistency. Doug Vandemark (DV), Ralph Foster (RF), Justin E. Stopa (JS), Alex Ayet (AA), and Chen Wang (CW) tagged all 2100 images independently. These were divided into three main steps: SAR image preparation and sampling strategy performed by JS, expert tagging performed by DV, RF, JS, AA, and CW, and dataset organization performed by JS. The tagging rules and class definitions were developed through conversations between JS, DV, RF, AA, and CW during preliminary tagging exercises.

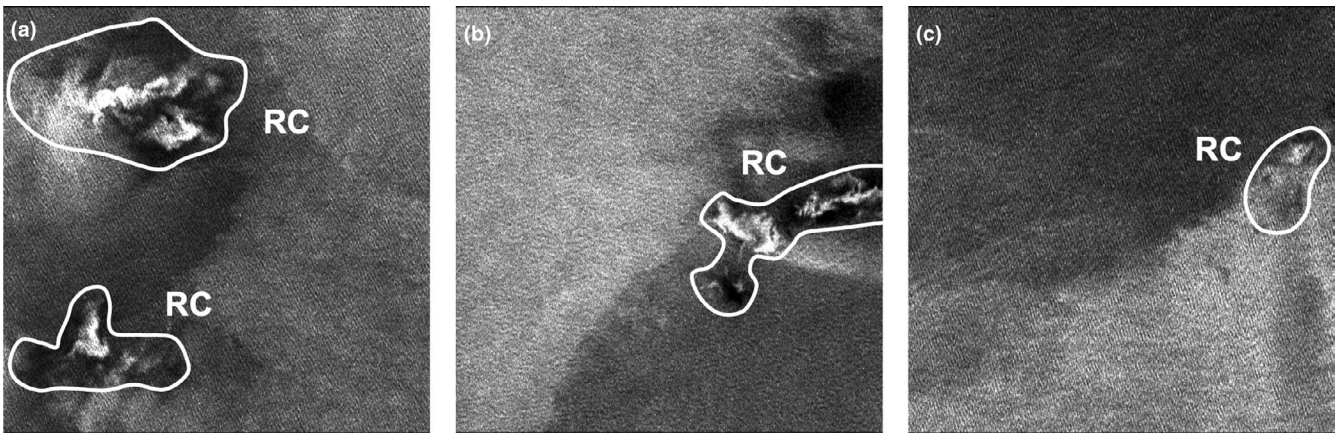

FIGURE 10 Examples of RC tags: (a) with WS ~ MC and GF, (b) with GF, WS < MC and NV, and (c) with GF, WS < MC and NV.

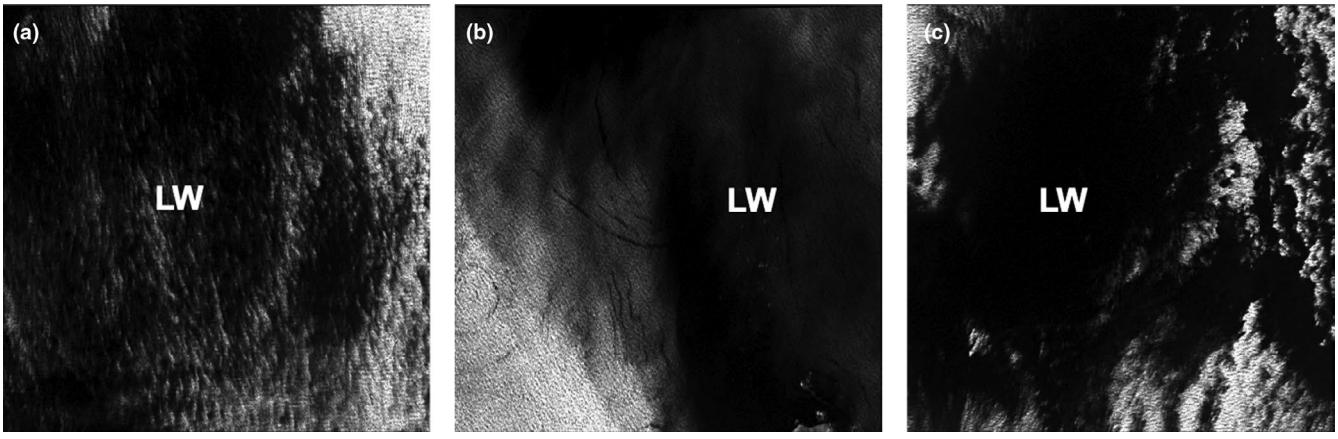

FIGURE 11 Examples of LW tags.

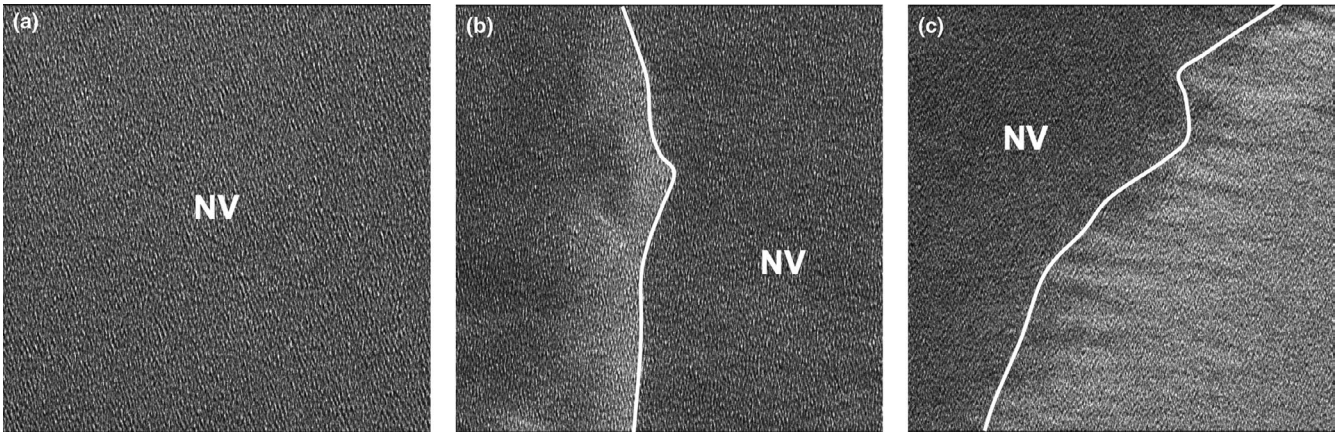

FIGURE 12 Examples of NV tags: (a) the whole image, (b) with AB, and (c) with GF. The white lines give the boundaries of NV areas. See details in Section 3.2 for the definition of AB and GF.

4.1 | Selection and processing of the WV images

A region spanning 8° × 8° centered over the Northwest Tropical Atlantic Station (NTAS) buoy (51° W, 15° N) was selected. Two thousand and one hundred S-1 WV2 SAR images were selected within this region from January to June in 2016–2019 as given in Figure 13. A fixed number of 350 images were collected within each month to minimize the potential impact of uneven temporal sampling

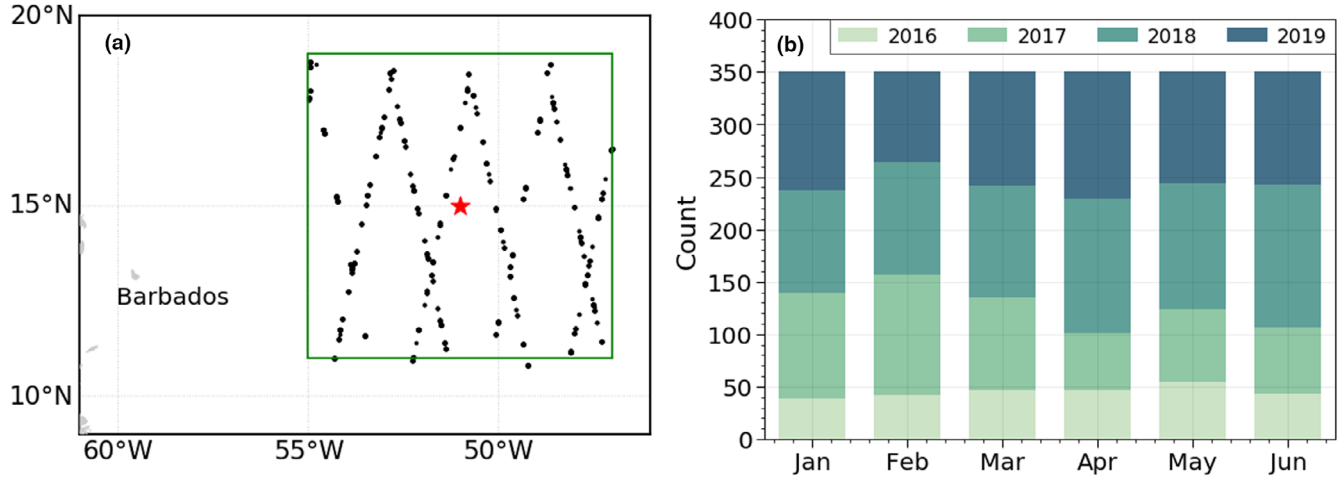

FIGURE 13 Spatial (a) and temporal (b) distributions of the 2100 S-1 WV SAR images. The red star in (a) gives the location of the Northwest Tropical Atlantic Station (NTAS) buoy.

on geophysical interpretation. An equal number of images were selected for incidence angles of 23° and 37°.

All SAR images were generated using SLC products of S-1 WV as described in Wang, Mouche, et al. (2019).

The original 5 m resolution was downsampled to 50 m per pixel, and the incidence angle effect was eliminated using CMOD5.N with a constant wind vector input. These processed images represent the sea surface roughness. To enhance the visibility of SAR images and facilitate their use in label tools, each image is normalized between the 1st and 99th percentile and then mapped to the range of 0–255 at grayscale and saved in Portable Network Graphics (PNG) format. The taggers were not provided with any information other than the PNG images.

4.2 | Expert tagging

We used Labelbox (https://labelbox.com), an online platform that allows several people to label the same database simultaneously, to tag our SAR image dataset. The tagging record of each expert is kept and in the end, five sets of tags were collected for this particular dataset. As discussed in the label rules above, each image can be associated with multiple tags.

In defining the 12 tags and creating the final dataset, three rounds of tagging were carried out. The first was an experimental tagging using 100 SAR images. Each expert provided their tags based on the visual inspection of the SAR images, with only limited discussion on tag definition. Low consensus on image tags across these 100 images led to a review and refinement of the labelling guide. A second round provided a double check and slight further refinements on the tagging and label guide details. The third and final round provided the independent tagging of the 2100 image dataset without further discussion.

4.3 | Dataset organization

This dataset is organized into a folder containing all the processed SAR images in PNG format, as well as two accompanying text files. One text file provides additional details for each image, including file name, center latitude/longitude,T and experts' tags. All relevant details about the dataset organization are documented in another text file. SAR images are provided in PNG format to facilitate visual inspection. The original Sentinel-1 SLC products are publicly available on ESA's data hub at https://sentinel.esa.int/web/sentinel/sentinel-data-access. All auxiliary files related to this dataset are available upon request from the authors.

4.4 | Tag analysis

Figure 14 provides data population across the labelled tags of the 2100 SAR images. Note that multiple tags are allowed for each image, therefore the total tag number of each expert is larger than 2100. Overall, a high consensus is observed despite some divergence for the classes WS, WS > MC, NV, UN, and AB. DV and RF were in agreementfor almost all of the categories. These are deemed as the most stable and/or trustworthy. By comparison, CW sorts more WS images into WS > MC and labels more MC images. Both JS and AA have fewer instances NV and AB, while tagging more UN cases. Taking the RC (rain cell) and CP (cold pools) as other examples, all

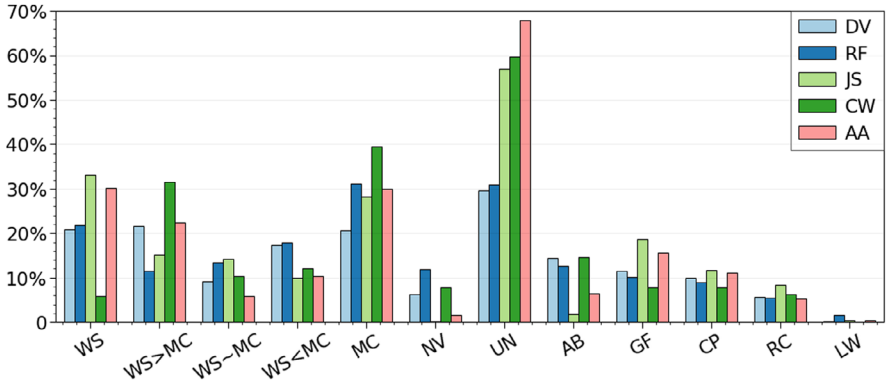

FIGURE 14 A histogram of the 12 tags labelled by five experts for the 2100 S-1 WV SAR images. Expert initials are provided in the legend and relate to those listed in Section 4.

TABLE 2 The percentage of the total five WS/MC MABL coherent structure and RC/CP expert tags obtained by visual inspection of the 2100 SAR images.

	WS/MC coherent structures (%)	RC/CP (%)	LW (%)
DV	61.57	15.57	0.24
RF	68.00	14.62	1.52
JS	70.71	20.05	0.48
CW	71.48	14.15	0.05
AA	72.19	16.38	0.33
CMwv	61.62	14.57	0.05

Note: In terms of the five WS/MC tags, only images without any other coexisting phenomena are used to allow comparison with the results of the automated classification tool (CMwv) developed in Wang, Tandeo, et al. (2019). SAR images for MABL studies.

experts tagged ~7% and ~10% occurrence percentages that are consistent with the previous reports using independent observational approaches (Brilouet et al., 2023; Vial et al., 2021).

We also conducted an objective comparison with SAR image results predicted using the automated classification tool (CMwv) from Wang, Tandeo, et al. (2019). This was performed considering the five WS/MC coherent structures and RC/CP. Note that the tool is only designed to classify entire images having WS, MC, or RC. That is to say, the joint and relative contribution of WS and MC are not a model objective, and CP was not considered as a class. Thus, only the visual inspection tags of WS/MC without any other coexisting local events can be evaluated for comparison with the model results. As listed in Table 2, the percentage of WS/MC classified by CMwv is 61.62%, which roughly equals that of DV and is a bit lower than that of RF, JS, and CW. The CMwv classified RC population is consistent with the visual inspection of RC plus CP by all experts and CMwv is within 5% of all hand labeled tags. Such a comparison provides one independent validation of this new labeled dataset and further validation of the classification tool. And it also illustrates the necessity to develop an advanced or alternative tool that permits multiple labelling of the SAR images for geophysical studies.

5 | DISCUSSION AND PERSPECTIVES

In this work, a detailed definition of 12 SAR-observed MABL phenomena observed over the ocean using Sentinel-1 WV mode imagery is presented. Using these SAR image texture labels, a dataset consisting of the labelling tags of 2100 SAR images collected near Barbados is created. The image labelling rules permitted multiple tagging within each image, as visually performed by five subject experts. The dataset includes pre-processed SAR images in PNG format, two text files that document the metadata, the expert tags, and auxiliary information for each image. Preliminary analysis of the tags illustrates a high level of population consensus among experts. All expert confirmed that 70% of the images contain rolls and cells or coherent structures within the MABL. ~400 (~27%) of them are depicted along with the local features of AB, GF, CP, or RC. This implies that even within this 20 km by 20 km observational window, the MABL coherent structures occur simultaneously with the convective events. It is consistent with the previous reports that WS/MC coincidence is typically seen near the edge of rain cells, gust fronts, or atmospheric fronts. Fine tagging of transitional states or related phenomena as presented in this dataset is challenging yet beneficial to further geophysical studies. The transitional states between roll dominance and cell dominance are grouped into five subcategories, depending on their relative dominance on a SAR image. As a matter of fact, such detailed classification is challenging for visual inspection during the labelling process, which might lead to slightly different interpretations among the five experts. However, this dataset marks an invaluable resource for the investigation of roll-cell transition concerning their structures and occurrence conditions. This is also true of AB and GF, which is also a kind of

AB if taking a rough classification. Here, we take the definition that AB corresponds to the air-mass boundaries where no clear changes are observed crossing the front, while changes in coherent structures are associated with GF. These fine-tagged SAR images shall facilitate studies focusing on the MABL state transitions states. As a follow-on, quantitative means perhaps can be developed to estimate the WS/MC structure changes and AB against GF.

The new hand-labeled dataset near Barbados could be readily applied for further geophysical analysis in several aspects. These tagged SAR images allow us to depict the MABL morphology at high spatial resolution (Young et al., 2000). Together with measurements of the field campaigns and/or the NTAS buoy, our primary focus is the environmental conditions under which these coherent structures form or their transitions occur (O'Driscoll et al., 2023; Stopa et al., 2022). This shall validate the results based on numerical simulations (Salesky et al., 2017) and extend the land-based radar observations collected in central Oklahoma (Banghoff et al., 2020). These accurately tagged SAR images are located over the open ocean, allowing assessment of ABL roll and cell formation characteristics compared to that on land (Santellanes et al., 2021). In addition, the fine tagging of local convection events, like rains and cold pools shall support investigation of the low-level cloud dynamics (Bony et al., 2015). Whichever process of interest, SAR-measured events are in the so-called 'grey zone' scale of 1-10 km, which shall bridge the gap toward sub-mesoscale representations in MABL (Bony et al., 2017; Honnert et al., 2020; Kealy et al., 2019). From the technical point of view, this dataset may serve as a prototype, providing training input data for more sophisticated ocean SAR classification models that subcategorize multiple MABL processes. This could lead to new applications for the huge S-1 SAR database that target the enhanced use of MABL information in global atmosphere-ocean studies.

ACKNOWLEDGEMENTS

CW was supported by the National Natural Science Foundation of China under grants 42206179 and 42350003. DV, RF and JS were supported by NASA Physical Oceanography and Ocean Vector Wind Science Team through grants 80NSSC20K0822 and 80NSSC23K0978 respectively. JS was also supported by the National Science Foundation grant number 2132150. This work is referenced by the School of Ocean and Earth Science and Technology (SOEST) contribution number 11857. AM and BC were supported by ESA Contract No. 4000135827/21/NL-Harmony Science Data Utilization and Impact Study for Ocean. AM was also supported by ESA Sentinel-1 Mission Performance Center (4000107360/12/I-LG). AA, BC, and AM were supported by the MITI Grant ALESE and AA and BC were supported by the CNES grant I-CASCADE. The authors thank ESA for providing the data and IFREMER for the computing resources used in this study.

OPEN RESEARCH BADGES

This article has been awarded Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. Data is available at https://osf.io/tvyxz/wiki.

DATA AVAILABILITY STATEMENT

Data available at https://doi.org/10.17882/93947.

ORCID

Chen Wang https://orcid.org/0000-0002-0575-742X
Justin E. Stopa https://orcid.org/0000-0002-7477-8224
Doug Vandemark https://orcid.org/0000-0003-4367-5457
Alex Ayet https://orcid.org/0000-0002-4044-4488
Alexis Mouche https://orcid.org/0000-0003-1250-4436
Betrand Chapron https://orcid.org/0000-0001-6088-8775
Peter Sadowski https://orcid.org/0000-0002-7354-5461

REFERENCES

Alpers, W., Zhang, B., Mouche, A., Zeng, K. & Chan, P.W. (2016) Rain footprints on C-band synthetic aperture radar images of the ocean - revisited. *Remote Sensing of Environment*, 187, 169–185. Available from: https://doi.org/10.1016/j.rse.2016.10.015

Atkinson, B.W. & Wu Zhang, J. (1996) Mesoscale shallow convection in the atmosphere. *Reviews of Geophysics*, 34(4), 403–431. Available from: https://doi.org/10.1029/96RG02623

Banghoff, J., Sorber, J., Stensrud, D., Young, G. & Kumjian, M. (2020) A 10-year warm-season climatology of horizontal convective rolls and cellular convection in Central Oklahoma. *Monthly Weather Review*, 148(1), 21–42. Available from: https://doi.org/10.1175/MWR-D-19-0136.1

Bony, S. (2005) Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, 32(20), L20806. Available from: https://doi.org/10.1029/2005GL023851

Bony, S., Schulz, H., Vial, J. & Stevens, B. (2020) Sugar, gravel, fish, and flowers: dependence of mesoscale patterns of trade-wind clouds on environmental conditions. *Geophysical Research Letters*, 47(7), e2019GL085988. Available from: https://doi.org/10.1029/2019GL085988

Bony, S., Stevens, B., Ament, F., Bigorre, S., Chazette, P., Crewell, S. et al. (2017) EUREC4A: a field campaign to elucidate the couplings between clouds, convection and circulation. *Surveys in*

Geophysics, 38(6), 1529–1568. Available from: https://doi.org/10.1007/s10712-017-9428-0

Bony, S., Stevens, B., Frierson, D.M.W., Jakob, C., Kageyama, M., Pincus, R. et al. (2015) Clouds, circulation and climate sensitivity. *Nature Geoscience*, 8(4), 261–268. Available from: https://doi.org/10.1038/ngeo2398; http://www.nature.com/articles/

Brilouet, P.-E., Bouniol, D., Couvreux, F., Ayet, A., Granero-Belinchon, C., Lothon, M. et al. (2023) Trade wind boundary layer turbulence and shallow precipitating convection: new insights combining SAR images, satellite brightness temperature, and airborne in situ measurements. *Geophysical Research Letters*, 50(2), e2022GL102180. Available from: https://doi.org/10.1029/2022GL102180

Brown, R.A. (2000) Serendipity in the use of satellite scatterometer, SAR, and other sensor data. *Johns Hopkins APL Technical Digest (Applied Physics Laboratory)*, 21(1), 21–26.

Grossman, R.L. (1982) An analysis of vertical velocity spectra obtained in the bomex fair-weather, trade-wind boundary layer. *Boundary-Layer Meteorology*, 23(3), 323–357. Available from: https://doi.org/10.1007/BF00121120

Holland, J.Z. (1970) Preliminary report on the BOMEX Sea-air interaction program. *Bulletin of the American Meteorological Society*, 51(9), 809–820. Available from: https://doi.org/10.1175/1520-0477(1970)051<0809:PROTBS>2.0.CO;2

Honnert, R., Efstathiou, G.A., Beare, R.J., Ito, J., Lock, A., Neggers, R. et al. (2020) The atmospheric boundary layer and the "gray zone" of turbulence: a critical review. *Journal of Geophysical Research: Atmospheres*, 125(13), e2019JD030317. Available from: https://doi.org/10.1029/2019JD030317

Johannessen, J.A., Shuchman, R.A., Johannessen, O.M., Davidson, K.L. & Lyzenga, D.R. (1991) Synthetic aperture radar imaging of upper ocean circulation features and wind fronts. *Journal of Geophysical Research*, 96(C6), 10411. Available from: https://doi.org/10.1029/91JC00301

Kealy, J.C., Efstathiou, G.A. & Beare, R.J. (2019) The onset of resolved boundary-layer turbulence at Grey-zone resolutions. *Boundary-Layer Meteorology*, 171(1), 31–52. Available from: https://doi.org/10.1007/s10546-018-0420-0

Kudryavtsev, V., Myasoedov, A., Chapron, B., Johannessen, J.A. & Collard, F. (2012) Imaging mesoscale upper ocean dynamics using synthetic aperture radar and optical data. *Journal of Geophysical Research: Oceans*, 117, C04029. Available from: https://doi.org/10.1029/2011JC007492

LeMone, M.A. & Pennell, W.T. (1976) The relationship of trade wind cumulus distribution to subcloud layer fluxes and structure. *Monthly Weather Review*, 104(5), 524–539. Available from: https://doi.org/10.1175/1520-0493(1976)104<0524:TROTWC>2.0.CO;2

Li, X., Zheng, W., Yang, X., Zhang, J.A., Pichel, W.G. & Li, Z. (2013. ISSN 0022-4928) Coexistence of atmospheric gravity waves and boundary layer rolls observed by SAR*. *Journal of the Atmospheric Sciences*, 70(11), 3448–3459. Available from: https://doi.org/10.1175/JAS-D-12-0347.1

Li, X.-M. (2016) A new insight from space into swell propagation and crossing in the global oceans. *Geophysical Research Letters*, 43(10), 5202–5209. Available from: https://doi.org/10.1002/2016GL068702

Medeiros, B. & Nuijens, L. (2016) Clouds at Barbados are representative of clouds across the trade wind regions in observations and climate models. *Proceedings of the National Academy of Sciences of the United States of America*, 113(22), E3062–E3070. Available from: https://doi.org/10.1073/pnas.1521494113

Mourad, P.D., Thompson, D.R. & Vandemark, D.C. (2000) Extracting fine-scale wind fields from synthetic aperture radar images of the ocean surface. *Johns Hopkins APL Technical Digest*, 21, 108–115.

O'Driscoll, O., Mouche, A., Chapron, B., Kleinherenbrink, M. & López-Dekker, P. (2023) Obukhov length estimation from spaceborne radars. *Geophysical Research Letters*, 50(15), e2023GL104228. Available from: https://doi.org/10.1029/2023GL104228

Salesky, S.T., Chamecki, M. & Bou-Zeid, E. (2017) On the nature of the transition between roll and cellular Organization in the Convective Boundary Layer. *Boundary-Layer Meteorology*, 163(1), 41–68. Available from: https://doi.org/10.1007/s10546-0160220-3

Santellanes, S.R., Young, G.S., Stensrud, D.J., Kumjian, M.R. & Pan, Y. (2021) Environmental conditions associated with horizontal convective rolls, cellular convection, and No organized circulations. *Monthly Weather Review*, 149(5), 1305–1316. Available from: https://doi.org/10.1175/MWR-D-20-0207.1

Schulz, H., Bony, S. & Stevens, B. (2020) Combining crowdsourcing and deep learning to explore the mesoscale Organization of Shallow Convection. *Bulletin of the American Meteorological Society*, 101(11), E1980–E1995. Available from: https://doi.org/10.1175/BAMS-D-19-0324.1

Sherwood, S.C., Bony, S. & Dufresne, J.-L. (2014) Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, 505(7481), 37–42. Available from: https://doi.org/10.1038/nature12829

Stevens, B., Bony, S., Brogniez, H., Hentgen, L., Hohenegger, C., Kiemle, C. et al. (2020) Sugar, gravel, fish and flowers: mesoscale cloud patterns in the trade winds. *Quarterly Journal of the Royal Meteorological Society*, 146(726), 141–152. Available from: https://doi.org/10.1002/qj.3662

Stopa, J.E., Wang, C., Vandemark, D., Foster, R., Mouche, A., & Chapron, B. (2022). Automated global classification of surface layer stratification using high-resolution sea surface roughness measurements by satellite synthetic aperture radar. *Geophysical Research Letters*, 49, e2022GL098686. https://doi.org/10.1029/2022GL098686

Vandemark, D., Mourad, P.D., Bailey, S.A., Crawford, T.L., Vogel, C.A., Sun, J. et al. (2001) Measured changes in ocean surface roughness due to atmospheric boundary layer rolls. *Journal of Geophysical Research: Oceans*, 106(C3), 4639–4654. Available from: https://doi.org/10.1029/1999JC000051

Vial, J., Vogel, R. & Schulz, H. (2021) On the daily cycle of mesoscale cloud organization in the winter trades. *Quarterly Journal of the Royal Meteorological Society*, 147(738), 2850–2873. Available from: https://doi.org/10.1002/qj.4103

Wang, C., Mouche, A., Tandeo, P., Stopa, J.E., Longépé, N., Erhard, G. et al. (2019) A labelled ocean SAR imagery dataset of ten geophysical phenomena from Sentinel-1 wave mode. *Geoscience Data Journal*, 6(2), 105–115. Available from: https://doi.org/10.1002/gdj3.73

Wang, C., Tandeo, P., Mouche, A., Stopa, J.E., Gressani, V., Longepe, N. et al. (2019) Classification of the global Sentinel-1 SAR vignettes for ocean surface process studies. *Remote Sensing of Environment*, 234, 111457. Available from: https://doi.org/10.1016/j.rse.2019.111457

Chen Wang, Douglas Vandemark, Alexis Mouche, Bertrand Chapron, Huimin Li, and Ralph C Foster. An assessment of marine atmospheric boundary layer roll detection using Sentinel-1 SAR data. *Remote Sensing of Environment*, 250:112031, 2020. https://doi.org/10.1016/j.rse.2020.112031

Young, G.S., Sikora, T.N. & Winstead, N.S. (2005) Use of synthetic aperture radar in Finescale surface analysis of synoptic-scale fronts at sea. *Weather and Forecasting*, 20(3), 311–327. Available from: https://doi.org/10.1175/WAF853.1

Young, G.S., Sikora, T.D. & Winstead, N.S. (2000) Inferring marine atmospheric boundary layer properties from spectral characteristics of satellite-borne SAR imagery. *Monthly Weather Review*, 128, 1506–1520. Available from: https://doi.org/10.1175/1520-0493(2000)128<1506:IMABLP>2.0.CO;2

How to cite this article: Wang, C., Stopa, J.E., Vandemark, D., Foster, R., Ayet, A., Mouche, A. et al. (2025) A multi-tagged SAR ocean image dataset identifying atmospheric boundary layer structure in winter tradewind conditions. *Geoscience Data Journal*, 12, 1–14. Available from: https://doi.org/10.1002/gdj3.282