

A labelled ocean SAR imagery dataset of ten geophysical phenomena from Sentinel-1 wave mode

Chen Wang, Alexis Mouche, Pierre Tandeo, Justin Stopa, Nicolas Longepe, Guillaume Erhard, Ralph Foster, Douglas Vandemark, Bertrand Chapron

▶ **To cite this version:**

Chen Wang, Alexis Mouche, Pierre Tandeo, Justin Stopa, Nicolas Longepe, et al.. A labelled ocean SAR imagery dataset of ten geophysical phenomena from Sentinel-1 wave mode. Geoscience Data Journal, 2019, 6 (2), pp.105-115. 10.1002/gdj3.73 . hal-02275886

HAL Id: hal-02275886

https://imt-atlantique.hal.science/hal-02275886

Submitted on 2 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Received: 28 September 2018 | Revised: 30 April 2109 | Accepted: 17 June 2019

DOI: 10.1002/gdj3.73

DATA PAPER

Geoscience Data Journal RMetS WILEY

A labelled ocean SAR imagery dataset of ten geophysical phenomena from Sentinel-1 wave mode

Chen Wang[1,2] | **Alexis Mouche**[1] | **Pierre Tandeo**[2] | **Justin E. Stopa**[3] | **Nicolas Longépé**[4] | **Guillaume Erhard**[4] | **Ralph C. Foster**[5] | **Douglas Vandemark**[6] | **Bertrand Chapron**[1]

[1]Laboratoire d'Océanographie Physique et Spatiale (LOPS), IFREMER, University in Brest, CNRS, IRD, Brest, France

[2]Lab-STICC, IMT Atlantique, UBL, Brest, France

[3]Department of Ocean Resources and Engineering, University of Hawaii at Manoa, Honolulu, Hawaii, USA

[4]Space and Ground Segment, Collecte Localisation Satellites (CLS), Plouzané, France

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

[6]Ocean Processes Analysis Laboratory, University of New Hampshire, Durham, New Hampshire, USA

Correspondence
Chen Wang, LOPS, IFREMER, Brest, France.
Email: chen.wang@ifremer.fr

Funding information
ESA Sentinel-1A Mission Performance Center, Grant/Award Number: 4000107360/12/I-LG; ESA S1-4SCI Ocean Study, Grant/Award Number: 4000115170/15/I-SBo; CNES TOSCA program; NASA Physical Oceanography, Grant/Award Number: NNX17AH17G; China Scholarship Council (CSC)

Abstract

The Sentinel-1 mission is part of the European Copernicus program aiming at providing observations for Land, Marine and Atmosphere Monitoring, Emergency Management, Security and Climate Change. It is a constellation of two (Sentinel-1 A and B) Synthetic Aperture Radar (SAR) satellites. The SAR wave mode (WV) routinely collects high-resolution SAR images of the ocean surface during day and night and through clouds. In this study, a subset of more than 37,000 SAR images is labelled corresponding to ten geophysical phenomena, including both oceanic and meteorologic features. These images cover the entire open ocean and are manually selected from Sentinel-1A WV acquisitions in 2016. For each image, only one prevalent geophysical phenomenon with its prescribed signature and texture is selected for labelling. The SAR images are processed into a quick-look image provided in the formats of PNG and GeoTIFF as well as the associated labels. They are convenient for both visual inspection and machine learning-based methods exploitation. The proposed dataset is the first one involving different oceanic or atmospheric phenomena over the open ocean. It seeks to foster the development of strategies or approaches for massive ocean SAR image analysis. A key objective was to allow exploiting the full potential of Sentinel-1 WV SAR acquisitions, which are about 60,000 images per satellite per month and freely available. Such a dataset may be of value to a wide range of users and communities in deep learning, remote sensing, oceanography and meteorology.

Dataset

Identifier: https://doi.org/10.17882/56796
Creator: Wang, C., Mouche, A., Tandeo, P., Stopa, J. E., Longépé, N., Erhard, G., Foster, R., Vandemark, D. and Chapron, B.
Title: Labelled SAR imagery dataset of ten geophysical phenomena from Sentinel-1 wave mode (TenGeoP-SARwv)
Publisher: Sea scientific open data publication (SEANOE)
Publication year: 2018
Resource type: Dataset
Version: 1.0

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2019 The Authors. *Geoscience Data Journal* published by Royal Meteorological Society and John Wiley & Sons Ltd.

KEYWORDS
manual labelling, ocean surface phenomena, Sentinel-1 wave mode, Synthetic aperture radar

1 | INTRODUCTION

The world's ocean covers more than 70% of the Earth's surface, playing a crucial role in influencing the climate system. Comprehensive measurements and observations of ocean surface are essential to better understand air–sea interactions as well as to develop high-resolution climate models (Topouzelis and Kitsiou, 2015; Schneider *et al.*, 2017). Among the various space-borne sensors, Synthetic Aperture Radars (SAR) met both high-resolution and all weather day-and-night imaging criteria. SAR backscattering is very sensitive to the sea surface roughness composed of centimetre-scale waves. When air–sea interactions are strong enough to modulate these short waves, SAR can capture signatures of geophysical processes such as ocean waves (Ardhuin *et al.*, 2009; Collard *et al.*, 2009), atmospheric processes (Atkinson and Wu Zhang, 1996; Young *et al.*, 2005; Alpers *et al.*, 2016) and oceanic processes (Espedal *et al.*, 1996; Jia *et al.*, 2018). Therefore, SAR is a unique tool for extensive observation of ocean–atmosphere interactions at sub-km scales (Brown, 2000; Jackson and Apel, 2004).

SAR sensors have a variety of acquisition modes. A common one is wide-swath which provides data over several hundred kilometres. More specifically for Sentinel-1, wide-swath acquisitions in TOPS mode (De Zan and Monti Guarnieri, 2006) (Extended wide swath EW and Interferometric wide swath IW) are mainly used for monitoring of sea ice areas and coastal regions over the ocean. Due to power and data constraints of contemporary systems, the wide-swath mode with high-resolution capability is not able to collect data continuously and globally. The 'WaVe mode' (WV or WM), by contrast, is dedicated to measuring ocean waves from the global open ocean. This mode was firstly introduced on Earth observation mission by the European Space Agency (ESA) for the European Remote Sensing (ERS-1/2) missions (1991–2003) (Kerbaol *et al.*, 1998). Since then, acquisitions in WV have been pursued on Envisat advanced SAR (ASAR) mission (2002–2012) (Stopa *et al.*, 2016) and now Sentinel-1 (Torres *et al.*, 2012), providing more than 25 years of high-resolution observations of the world's ocean. The recent launches of Sentinel-1 (S-1) A and B in April 2014 and 2016 for the European Copernicus Program enable routine SAR WV acquisitions to be available. These two sensors collect nearly 120,000 WV vignettes with ocean surface imprints of 20 × 20 km in each month. Their spatial resolution is about 5 m. The primary intent of the small-sized vignettes is to provide ocean swell directional spectra as an ESA Level-2 ocean product (Torres *et al.*, 2012). However, they also capture a much wider range of geophysical processes that are of significant interest in ocean–atmosphere interactions. Global coverage is combined with high resolution and routine acquisitions in all weather conditions during day and night; and at such high resolution (5 m) make S-1 WV a presence and unique data source for new geophysical applications.

In this study, we define ten categories of different oceanic or atmospheric phenomena. These are the most common phenomena that can be observed in S-1 WV vignettes. The categories are pure ocean waves (POW), wind streaks (WS), micro convective cells (MCC), rain cells (RC), biological slicks (BS), sea ice (SI), icebergs (IB), low wind areas (LWA), atmospheric fronts (AF) and oceanic fronts (OF). Details on these definitions are introduced in Section 2. A labelled SAR WV dataset containing 37,553 images is then established. Within each image, only the prevalent geophysical phenomenon with clear signature and/or pattern is presented. The images are derived from the Single Look Complex (SLC) product of S-1 WV (Torres *et al.*, 2012), and provided in formats of Portable Network Graphics (PNG) and Georeferenced Tagged Image File Format (GeoTIFF). The proposed dataset, called TenGeoP-SARwv for 'Ten Geophysical Phenomena from SAR wave mode', is provided by IFREMER and publicly available at sea scientific open data publication (SEANOE): http://www.seanoe.org/data/00456/56796/. The methodology used to create the dataset is described in section 3. Such a labelled dataset could benefit the strategic development of massive ocean SAR data analysis. Deep learning signal processing algorithms that are now a common form of supervised learning may be exploited (LeCun *et al.*, 2015; Cheng *et al.*, 2017). In addition, this unique dataset is also significant to the communities of remote sensing, oceanography and meteorology. Discussion and perspectives regarding to potential applications and dataset refinement are given in section 4.

2 | GEOPHYSICAL PHENOMENA BY SENTINEL-1 WAVE MODE

The ESA S-1 mission is a constellation of two polar orbiting, sun-synchronous satellites (S-1 A and S-1 B) launched in April 2014 and 2016 respectively (Torres *et al.*, 2012). These two satellites both have a 12-day repeat cycle at the equator, and are phased at 180° to provide an effective 6-day repeat cycle. For each satellite, the expected

life time is 7 years. Both carry a C-band SAR instrument with centre frequency of 5.405 GHz (5.5 cm wavelength). There are four exclusive imaging modes (Interferometric Wide swath, Extra Wide swath mode, Strip Map and Wave Mode[1]) for the S-1 SAR sensors. WV is the default operational mode over open ocean unless wide-swath SAR images are requested for particular applications. Note that, at present, there is no WV acquisition in the Arctic Ocean and closed seas such as Red, Black, Mediterranean and Caribbean seas. And in this study, we only use WV data acquired by Sentinel-1A in 2016. However, extensive validations have confirmed that SAR images acquired by S-1B are characteristic essentially equivalent to that from S-1A.

2.1 | Acquisitions

The vignettes of S-1 WV are collected in 20 × 20 km scenes at two alternate incidence angles of 23.8° (WV1) and 36.8° (WV2). They are acquired over the global open ocean with an along-track sampling separation of 100 km and an across-track distribution of roughly 200 km. Pixels within each vignette have 5 m ground resolution and can be obtained in VV (default) or HH polarization. For the polarization configurations, the first letter stands for the polarization of the emitting transmitter (as SAR is an active radar), whereas the second one is for the receiver polarization configuration. This study only relies on vignettes acquired in VV polarization as they account for more than 99% of all S-1 WV acquisitions. The S-1 WV backscatter consists of intensity and phase history and can be potentially processed into the SLC products for wave applications (Torres *et al.*, 2012). Using the digital number (DN) of these complex products and Look-Up-Table (*sigmaNaught*) annotated in the product, we compute the normalized radar cross section σ_0. This is the common radar parameter used to describe radar return backscattered by the ocean surface to the SAR sensors.

$$\sigma_0 = \frac{|\mathrm{DN}|^2}{sigmaNaught^2}. \quad (1)$$

2.2 | Ten geophysical phenomena

In this subsection, the ten defined oceanic or atmospheric phenomena are presented. Due to the 20-km WV image size, scales of observed geophysical phenomena are limited to about between 0.1 and 5 km. We focus on the prescribed ten geophysical phenomena because they are commonly observed by the S-1 WV SAR vignettes. It is worth noting that the WV can also capture signatures of other geophysical phenomena like internal waves (Jia *et al.*, 2018) and atmospheric gravity waves (Li *et al.*, 2013), and signatures of ships or platforms. While such phenomenaS are seldom seen in the open ocean, we may include those categories in this dataset in the near future.

2.2.1 | Pure ocean waves (POW)

Ocean waves including ocean swell and wind waves are the most prevalent feature in all SAR images (Fu and Holt, 1982; Jackson and Apel, 2004). Signatures of ocean waves often coexist with other oceanic and/or atmospheric phenomena. The short wind waves (centimetre to metre scale) are produced by local surface winds while ocean swell are longer (hundreds of metres) surface waves that are generated by distant weather systems such as storms or cyclones. These mechanical waves are propagating without any wind forcing after the wind blows for a period of time over a fetch of water (Ardhuin *et al.*, 2009). Ocean waves can be observed in all ocean basins, and their measurement with SAR relies on the theories of microwave scattering from rough sea surface (Hasselmann *et al.*, 1985; Collard *et al.*, 2009; Stopa *et al.*, 2016). SAR imaging of swell waves is typically influenced and distorted by different geophysical phenomena. This makes wave interpretation of SAR imagery difficult. Our definition of pure ocean waves (POW) is a SAR vignette that contains boundless ripples throughout the image, as displayed in Figure 1a. The following criteria are adopted for this category:

1. Periodic signatures of ocean waves dominate the whole image
2. Wavelengths are between 0.1 and 0.8 km
3. Intensity modulation within the scene is homogeneous
4. There is no other competing geophysical feature or pattern

2.2.2 | Wind streaks (WS)

Wind streaks are known to be the sea surface imprint of atmospheric boundary layer (ABL) rolls (Vandemark *et al.*, 2001). They usually occur in near-neutral to moderately unstable stratification conditions and span the whole depth of ABL. The approximately wind aligned wind streaks are the result of an embedded overturning coherent secondary circulation in the boundary layer that is induced by the vertical shear of the mean horizontal wind that can be further modified by the mean vertical stratification profile (Brown, 1980; Etling and Brown, 1993; Young *et al.*, 2002). The enhanced upward and downward wind perturbations near the surface between roll circulations are usually strong enough to modulate centimetre-scale waves and therefore induce organized imprints on the sea surface roughness. Consequently, wind streaks are frequently observed by SAR images as periodic, quasi two-dimensional and roll-shaped patterns (Alpers and Brümmer, 1994; Young *et al.*, 2002), as displayed in Figure 1b. It shows that the periodic pattern of wind streaks is superimposed at top signatures of ocean waves. In addition, some vignettes of wind streaks contain cell-shaped patterns (Micro

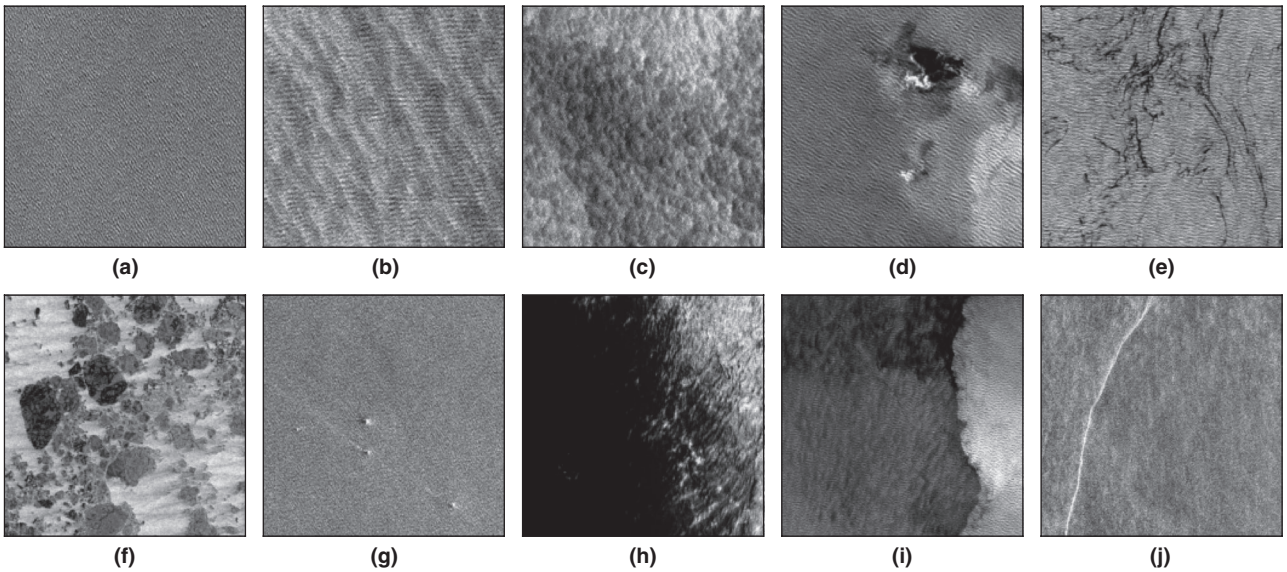


FIGURE 1 From (a) to (g) are image examples of pure ocean waves, wind streaks, micro convective cells, rain cells, biological slicks, sea ice, icebergs, low wind area, atmospheric front and oceanic front

convective cells). This indicates the transition between two different regimes in the marine ABL (Atkinson and Wu Zhang, 1996; Jackson *et al.*, 2004). This transient stage can sometimes be tricky to decipher between the two classes. The criteria of wind streaks (WS) are:

1. Periodic linear features dominate the whole image
2. Wavelengths are between 0.8 and 5 km
3. Intensity modulation within the scene is homogeneous
4. Periodic signatures of ocean waves can coexist

2.2.3 | Micro convective cells (MCC)

Atmospheric convective cells are coherent structures of updrafts and downdrafts in the ABL (Khalsa and Greenhut, 1985; Atkinson and Wu Zhang, 1996). The local temperature difference between air and sea produces strong vertical exchange of heat. It creates cell-shaped rising/descending air, which leads to horizontal variability of sea surface wind speed. This wind variability modulates the centimetre-scale waves and thus the sea surface roughness. Therefore, coherent, periodic and cell-shaped patterns are normally visible on SAR images (Babin *et al.*, 2003). Note that the scale (radius) of atmospheric convective cells captured within these 20-by-20 km WV vignettes is about 1 km. It indicates that the cells here are mainly associated with shallow dry convection, where latent heat from condensation plays no role in the dynamics (Atkinson and Wu Zhang, 1996). This results in a category that we define as micro convective cells (MCC). However, roll-shaped pattern caused by wind streaks are also often visible in such vignettes. The key to distinguish between WS and MCC categories is based on which pattern dominates the image. An example of MCC vignette is displayed in Figure 1c. The criteria of this category are:

1. Coherent, periodic and cell-shaped features dominate the whole image
2. Scales are about 1 km
3. Intensity modulation within the scene is homogeneous
4. Periodic signatures of ocean waves can coexist, but they can be strongly distorted

2.2.4 | Rain cells (RC)

Rain can occur in many forms, such as downdraft, stratified rain, rain bands, squall lines and so on. Although the scattering mechanisms of C-band SAR for rain signatures are not fully understood, they can be generally characterized by high and low contrasts in backscatter (Alpers *et al.*, 2016). Here we only focus on the rain cells that are typically associated with downdraft patterns. Their signatures can be clearly captured by WV SAR vignettes. Our definition of rain cells (RC) largely concentrate on the vignettes containing circular- or semi-circular–shaped areas. This is typical signatures of wind gust fronts caused by the downdraft. Besides, bright and/or dark patches usually appear inside the circular areas. Dark patches are usually explained by signal attenuation due to rain droplets in the atmosphere. Bright areas are generally associated with splash due to the heavy rain impacts sea surface roughness. An example is given in Figure 1d. Note that the circular shape of RC is expected to be larger than that of MCC and may be sometimes larger than the vignette size. Our criteria for this category are:

1. Circular- or semi-circular–shaped areas are visible on SAR image
2. There are bright and/or dark patches inside the circular areas
3. Intensity modulation within the scene is heterogeneous
4. Periodic signatures of ocean waves can coexist with RC

2.2.5 | Biological slicks (BS)

Biological slicks (BS) in the ocean are natural films that accumulate at the water–air boundary (Jackson *et al.*, 2004). These surface slicks are typically only one molecular layer thick (approximately 3 nm) and consist of sufficiently hydrophobic substances. This thin film influences air–sea fluxes of momentum, heat and gas (Espedal *et al.*, 1996). Under low wind speeds, sea surface capillary and short gravity waves can be damped by the natural films. Thus, their signatures are usually visible as dark filaments on SAR images. The slicks captured by S-1 WV are generally randomly distributed over sea surface, see Figure 1e. Due to the coverage limitation of WV vignettes, the scale of slicks is hard to be quantified. However in some cases, they can be tracers of the ocean circulation such as surface currents, ocean fronts and eddies (Johannessen *et al.*, 1996). The following criteria are used to define this category:

1. Dark filaments are visible on SAR image
2. Intensity modulation within the scene is heterogeneous
3. Periodic signatures of ocean waves can coexist with BS, but they can be distorted

2.2.6 | Sea ice (SI)

Sea ice is defined as frozen ocean water which could be growing or melting. It is typically sorted according to whether or not it is attached to the shoreline, or described based on its development stages, such as new ice, nilas, young ice, first-year and old (Jackson *et al.*, 2004). SAR backscattering of sea ice essentially depends on the ice type, and therefore can be quite diverse due to the wide range of ice types. The textures of sea ice on SAR images are fairly complex. They can be roughly characterized by web shapes, three-dimensional structure, wiggly fractures, and high contrast (dark and bright patches) (Soh and Tsatsoulis, 1999). Our aim here was not to identify different sea ice types, but rather distinguish sea ice from open ocean water. Therefore, this category contains SAR vignettes of all ice types in the Southern Ocean near Antarctica. One sea ice example is shown in Figure 1f. The criteria of this category are:

1. Textural contexts are complex, which can be web-shaped, wiggly fractures, pebble-like, fractal and so on
2. Patches with sharp boundaries are usually visible on SAR image
3. There are strong intensity contrasts between different patches
4. Periodic signatures of ocean waves can coexist, but they can be severely distorted
5. Vignettes are mainly collected from the Southern Ocean near Antarctica

2.2.7 | Iceberg (IB)

Icebergs are large pieces of frozen freshwater that have broken off a glacier or an ice shelf and are floating freely in open water or sea ice area. They are categorized according to the size including growler (0–5 m), bergy bit (5–15 m), small berg (15–60 m), medium berg (60–120 m), large berg (120–220 m) and very large berg (>220 m), and/or with respect to their shape such as tabular, non-tabular, blocky, wedge, dry dock and pinnacle (Jackson *et al.*, 2004). In SAR images, icebergs appear as a cluster of pixels that have the uniformly high/low backscatter signals compared to the surroundings (sea water and sea ice). In our definition, the iceberg vignette contains one or several icebergs that are visible as bright targets. Possibly, there is a relatively dark shadow associated with the small bright cluster. This category focuses on the icebergs in the open sea water, as displayed in Figure 1g. Thus, the criteria of this category are:

1. Bright or Dark targets associated with dark shadows are visible on SAR image
2. Intensity modulation of the surroundings is homogeneous
3. Periodic signatures of ocean waves can coexist with IB
4. They are mainly distributed in the Southern Ocean near Antarctica

2.2.8 | Low wind areas (LWA)

When the local surface winds are too weak, sea state normally remains stationary for hours. Generally, there is no signature of ocean swell propagation and the small cm-scale roughness is absent too. Consequently, SAR backscatter from such sea surface is weak, resulting in dark areas on SAR images (Topouzelis and Kitsiou, 2015). Note that low wind condition is also necessary for the presence of biological slicks on SAR image. Thus, signatures of BS may exist at the boundaries of dark areas. In addition, LWA can also occur in areas where wind speed and/or direction suddenly change. Appearance of such LWA typically has a very large dark area accompanying by an atmospheric front. To distinguish from the definition of atmospheric front, the LWA category focuses on the vignettes that are dominated by a unique dark patch. An example is shown in Figure 1h. The criteria of LWA are:

1. A unique dark patch dominates the SAR image
2. Intensity modulation within the scene is heterogeneous
3. Periodic signatures of ocean waves are absent

2.2.9 | Atmospheric front (AF)

sAtmospheric fronts are associated with air mass boundaries and thus strong near-surface horizontal gradients of wind, temperature and/or humidity (Johannessen *et al.*, 1996). Unstable atmosphere conditions generally lead to occurrence of rain and low and high wind areas along the fronts. Therefore, signatures of atmospheric fronts observed by SAR are largely complex and have been called different names, including lobe, cleft, vortex, front and secluded front, based on their pattern (Young *et al.*, 2005). Figure 1i presents a vignette example of a typical atmospheric front observed by S-1 WV. This category is defined by the following criteria:

1. The edge of the front is typically not sharp, but rather a bit mottled or occluded
2. Besides the front, there are obvious intensity gradients
3. Intensity modulation of the surroundings is homogeneous
4. Periodic signatures of ocean waves can coexist with AF

2.2.10 | Oceanic front (OF)

Oceanic fronts are the boundaries between two distinct water masses that can be caused by a difference in oceanic temperature, salinity and/or density. The water masses near an oceanic front usually move in different directions, leading to downwelling or upwelling along the front and hence create a sea surface roughness anomaly (Rascle *et al.*, 2017). Enhanced or reduced sea surface roughness anomalies are visible as the bright or dark lines in SAR vignettes, as displayed in Figure 1g. Beside such lines, there are no obvious intensity gradients on the SAR image. This is the main distinction between OF and AF. The criteria of this category are:

1. A thin bright or dark mono-filament like linear feature is visible on SAR image
2. There is no obvious intensity gradient across the linear feature
3. Intensity modulation of the surroundings is homogeneous
4. Periodic signatures of ocean waves can coexist with OF

3 | DATASET CREATION

3.1 | SAR image processing

The 20×20 km image with 5 m resolution provides an image with more than 4,000 pixels in range and azimuth directions. This full-scale WV intensity image is not necessary for visual interpretation of oceanic or atmospheric phenomena. For instance, an example of a full resolution σ_0 image with features of wind streaks is shown in Figure 2a. It is clear that the wind streaks are concealed due to the low intensity contrast. In addition, the subplot of σ_0 mean values along range is displayed in Figure 2a*. It indicates that values of σ_0 slightly vary with different incidence angles within images. Therefore, three processing steps are applied to σ_0 images to enhance broad-scale features of oceanic and atmospheric phenomena.

3.1.1 | Re-calibration of σ_0

The σ_0 as measured by SAR over the ocean is highly dependent on the local ocean surface wind and viewing angles of the radar (incidence and azimuth angles). For a given wind speed, the overall σ_0 decreases along the range direction, as displayed in the inset plot of Figure 2a*. This decreasing trend in range is mainly associated with the increasing incidence angle, which is common to all C-band VV SAR imagery. The empirical geophysical model function, such as CMOD5.N for VV C-band SAR (Hersbach, 2010), models the σ_0 dependence on wind vector and radar incidence angle. To reduce the incidence angle effect, we use CMOD5.N to construct a reference factor by assuming a constant wind of 10 m/s at 45° relative to the antenna look angle. The σ_0 of each vignette can then be re-calibrated by dividing the reference factor. Note that the σ_0 values are in linear scale. Such re-calibrated σ_0 is referred as sea surface roughness (ssr) and is shown in Figure 2b. Specifically the ssr can be written:

$$\mathrm{ssr} = \frac{\sigma_0}{\mathrm{CMOD5.N}(10,45,\mathrm{inc})} \tag{2}$$

where inc is the radar incidence angle for each pixel. The difference of intensity in ssr image between near (left) and far (right) field is significantly reduced (Figure 2b*.

3.1.2 | Downsampling

The fine-resolution SAR vignettes are not favourable for visual interpretation of larger scale geophysical features, especially since our category definitions focus on phenomena with scale of tens to thousands of metres. The expected length scales of larger phenomena corresponding to the category definition are from 100 m to 5 km. Therefore to better highlight the larger features, a moving averaging window of 10-by-10 pixels is applied to the ssr images. This averaging also reduces the speckle noise of SAR vignettes (Lee *et al.*, 1994). The ssr intensity images (Figure 2b) are then downsampled by 1/10 yielding a resolution of 50 m, as shown in Figure 2c. As shown, the pattern of wind streaks overlapping on ocean swell is appreciably highlighted. It is worth noticing that the spatial filtering applied in this study

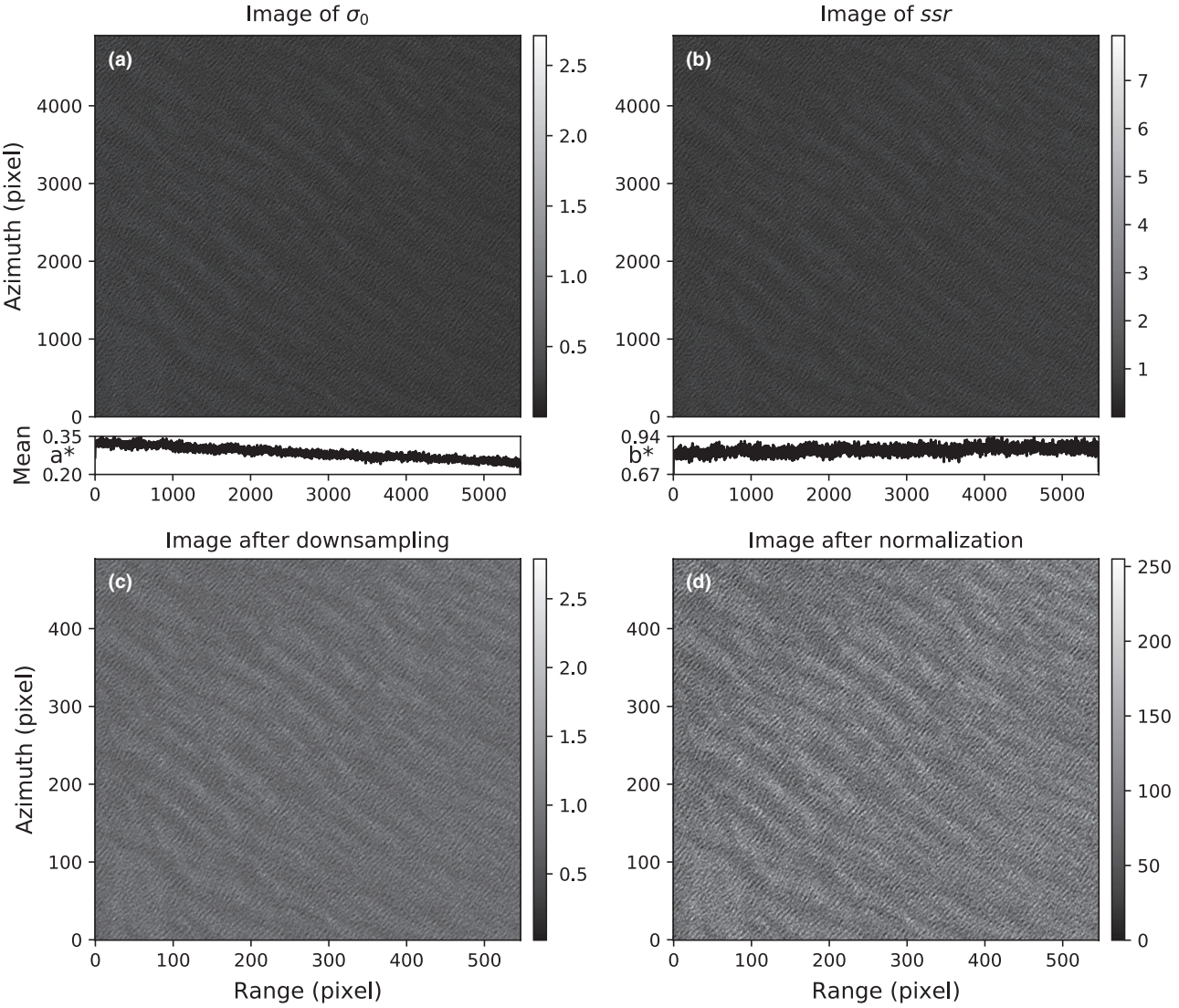


FIGURE 2 SAR image processing of the case with wind streaks. (a) σ_0 image (b) image of sea surface roughness derived from σ_0 (c) intensity image down sampled from roughness (d) intensity image with enhanced contrast. Subplots of * in (a) and (b) are profiles of σ_0 and roughness images along the range direction

achieves similar results as for the classical SAR multi-look technique. But the later performs the filtering in the image spectral domain, which is relatively time-consuming.

3.1.3 | Normalization

1. **For Shuman visual inspection** To enhance intensity contrast of the downsampled images, a statistical method of percentile is used to normalize each vignette. By sorting intensity values of an image, the proportion of principal data that falls between two given percentiles can be estimated (Natrella, 2013). For each image, this method split the ordered intensity values into hundredths, and pixel values between the 1st (minimum) and 99th (maximum) are normalized into a 8 bits grey scale ([0, 255]). With this processing, potential remaining anomaly (speckle noise) in the ssr values are effectively filtered. In Figure 2d, the image of wind streaks in grey scale after normalization is presented. It shows that the normalized image with enhanced contrast is better suited for visual interpretation of our identified oceanic and atmospheric phenomena.
2. **For machine learning-based exploitation** For the sake of machine learning approaches, another normalization process is also implemented. As opposed to the dataset used for human visualization where the retained minimum and maximum values are specific for each image, fixed values of 0 and 3 common to the entire database are applied to all downsampled ssr images. In between these two values, the quantization process is instead performed on 16 bits ([0, 65,535]) ensuring all texture and radiometric information are numerically maintained.

3.2 | TenGeoP-SARwv dataset

The TenGeoP-SARwv dataset is established based on the acquisitions of S-1A WV in VV polarization. This dataset consists of 37,553 SAR vignettes divided into our ten geophysical categories. For each category, the selection of SAR images covers the full year of 2016, and is manually labelled through visual inspection following the criteria documented in Section 2.2. Two screening standards are adopted. One is that only one individual geophysical phenomenon dominates across the whole vignette. The other one is that pattern structure of this phenomenon is clearly visible by human eye. Table 1 presents the counts of each category for each month. We attended to have 400 images labelled for each class in each month. However, we could not reach this number for the OF category. Moreover, only a few IB images were found in May to October due to the iceberg seasonality.

3.3 | Data format

The image patches are provided in formats of Portable Network Graphics (PNG) and Georeferenced Tagged Image File Format (GeoTIFF). PNG files are processed with floating normalization for better visualization of human eyes. While in GeoTIFF files, high precision values (16 bits) as well as the geographical information are kept for exploitation of machine learning-based approaches and geophysical application. In addition, text files containing description of categories and information for the file name, labelling, swath, capture time, and centre latitude and longitude of each image are also provided. These SLC products of S-1 WV are freely available at 'Sentinel open access hub' of ESA https://sentinel.esa.int/web/sentinel/sentinel-data-access. Notice that the GeoTIFF of TenGeoP-SARwv dataset is completely different from that of ESA SLC product. The latter contains original data and much more image processing information than this labelled dataset.

4 | DISCUSSION AND PERSPECTIVES

SAR images capture signatures of various geophysical phenomena that are associated with air–sea interactions. Most of them have been previously discussed to provide a comprehensive understanding of their imprints on SAR imagery (Alpers and Brümmer, 1994; Young *et al.*, 2005; Li *et al.*, 2013; Alpers *et al.*, 2016). Several of these phenomena factor significantly in the vertical transport of heat, moisture and momentum, and play key roles in the climate system (Khalsa and Greenhut, 1985; Ufermann and Romeiser, 1999; Vandemark *et al.*, 2001; Schneider *et al.*, 2017). Although understood, these manifestations of key geophysical phenomena are not systematically analysed or ingested in numerical models. In particular, automated detection and classification of these phenomena from the numerous SAR images is still challenging. The proposed SAR imagery dataset with individual annotations of oceanic or atmospheric phenomena should allow new efforts to test, validate and benchmark different methods for the identification of key geophysical processes. The annotations will allow massive classification of the data and open new perspectives for global or seasonal analysis of these phenomena. This work is a step towards broadening the scientific value of 25 years WV data acquired by ERS-1/2, Envisat ASAR and Sentinel-1 (Kerbaol *et al.*, 1998; Torres *et al.*, 2012; Stopa *et al.*, 2016). In addition, this labelled dataset can be directly used to statistically investigate the geophysical properties of the ten defined phenomena and characteristics of the imaged features. Therefore, such a dataset of labelled ocean SAR imagery is put forward for both scientific and engineering applications for different communities such as deep learning, remote sensing, oceanography and meteorology.

TABLE 1 SAR image numbers of the labelled dataset for each class in every month of 2016

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
POW	406	407	408	409	406	408	410	409	409	409	408	411	4,900
WS	406	409	403	407	404	391	396	398	397	398	384	404	4,797
MCC	396	384	370	384	385	365	386	388	380	391	384	385	4,598
RC	398	399	398	395	398	391	395	393	393	396	394	390	4,740
BS	398	394	395	398	397	339	397	398	397	400	398	398	4,709
SI	387	150	282	396	396	392	393	393	396	396	396	393	4,370
IB	399	417	308	146	58	29	10	14	12	29	159	399	1,980
LWA	137	137	138	220	201	95	214	144	207	207	241	219	2,160
AF	360	282	301	348	363	234	361	377	378	367	364	365	4,100
OF	61	85	64	102	131	60	116	135	96	108	132	109	1,199
Total	3,348	3,064	3,067	3,205	3,139	2,704	3,078	3,049	3,065	3,101	3,260	3,473	37,553

4.1 | SAR image classification with deep learning

In recent years, state-of-the-art deep learning based techniques, for example Convolutional Neural Network (CNN), have been developed rapidly for image processing applications (LeCun *et al.*, 2015). CNN model is a deep multilayer architecture that can be trained to automatically extract the optimal image feature representations and amplify the discrimination between different classes (LeCun *et al.*, 2015). This new approach has been widely introduced to remote sensing (Zhang *et al.*, 2016; Chen *et al.*, 2016; Cheng *et al.*, 2017). However, a lack of high-quality labelled datasets limits further application and development of CNN models for ocean SAR images. The proposed TenGeoP-SARwv dataset could be used as a training dataset for the identification and classification of different key geophysical phenomena occurring over open ocean. It can be used to directly fine-tune existing CNN models for straightforward geophysical applications, or explore new CNN architectures to improve feature representations. In fact, one relevant work has been conducted and the preliminary result was presented at the International Geoscience and Remote Sensing Symposium (IGARSS) in 2018. We believe that the large number of images in each of these ten classes satisfy the requirement to train a deep CNN model. Moreover, unsupervised learning algorithms by combining deep learning and reinforcement learning will become far more important (LeCun *et al.*, 2015) and can also benefit from this dataset. However, there is still room in the proposed dataset for improvement. In the near future, more geophysical categories corresponding to other oceanic or atmospheric phenomena should be included. In addition, some of the vignettes contain multiple geophysical phenomena within the same scene even they are small-sized SAR images. Multi-labelling of those images is more of interest and important to the deep learning community for methodologies exploitation.

4.2 | Ocean SAR remote sensing

Space-borne SAR provides a unique means to observe the ocean surface. Despite the multi-scale nature of ocean surface waves, C-band SAR mainly responds to cm-scale sea surface roughness through Bragg resonant scattering (Alpers and Brümmer, 1994; Jackson *et al.*, 2004; Li *et al.*, 2013). While radar signal depends on radar properties (wavelength, polarization, and incidence and azimuth angles), ocean SAR imagery can be generally interpreted as variability in sea surface roughness. Different oceanic or atmospheric phenomena are frequently captured by SAR images owing to their modulations on near-surface wind stress as well as on cm-scale ocean waves (Vandemark *et al.*, 2001; Alpers *et al.*, 2016). How strong the modulations should be to make these phenomena visible on SAR images is still an open question from a statistical point of view. With the TenGeoP-SARwv dataset, one can potentially investigate the environmental conditions under which these phenomena occur. This would help to better understand their impact on sea surface roughness and therefore how they are imaged by SAR. In addition to that, ocean swell is of first interest as they are a fundamental phenomenon over the open ocean. The swell spectra inversion from SAR measurements are still distorted by the presence of various oceanic or atmospheric phenomena. The dataset of TenGeoP-SARwv can help us quantify the impact of these phenomena on SAR forward mapping from ocean wave spectrum to SAR image spectra. Therefore, it may be possible to recover more ocean swell estimation by taking their impact into account. This will benefit the study of global and local wave climate. Given the relatively small size of the WV vignette (20 km), the imaged area can be roughly assumed as homogeneous and often-time, only one phenomenon dominates. However, it should also be mentioned that the small and sparse vignette coverage restricts imaging of large-scale phenomena, such as upwelling, internal waves, hurricanes, among others.

4.3 | Geophysical investigation

SAR imagery yields high-resolution imprints of ocean surface and provides significant geophysical parameters for global weather and climate analysis, demonstrating its indispensable contributions to the Earth monitoring system (Brown, 2000). Investigation of key geophysical phenomena by utilizing SAR data, for example wind streaks and micro convective cells, has been performed for many years, mostly in the stage of case and field studies (Alpers and Brümmer, 1994; Vandemark *et al.*, 2001; Levy, 2001; Babin *et al.*, 2003; Li *et al.*, 2013). Statistical analysis of key geophysical phenomenon based on SAR data are barely attempted due to lack of reliable dataset. This proposed TenGeoP-SARwv dataset opens perspective to use S-1 WV acquisitions for global geophysical phenomena analysis. Combined with other environmental parameters, these labelled SAR vignettes can be used directly to address geophysical characteristics of the ten defined specific phenomena. The occurrence and atmospheric conditions of one specific phenomenon can be of particular interest (Levy, 2001). Furthermore, classification of the whole acquisitions of S-1 WV vignette based on the automated methodologies inspires us to map the monthly variations and seasonal changes of these geophysical phenomena in the context of climate modelling. However, the small footprint of S-1 WV limits the observation of larger scale geophysical phenomena. Some of the vignettes only capture part of the phenomena signatures, for instance, a corner of

large-sized rain cell and a portion of atmospheric front or oceanic front.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their helpful comments and suggestions regarding this manuscript. This image dataset is created thanks to Sentinel-1A data access via ESA and through Sentinel-1A Mission Performance Center (Contract No. 4000107360/12/I-LG). Copernicus data (2016) are used. A. Mouche, B. Chapron, N. LongÉpÉ and C. Wang were supported by S1-4SCI Ocean Study (Contract No. 4000115170/15/I-SBo) and CNES (TOSCA program, COWS project). R. Foster and D. Vandemark were supported by NASA Physical Oceanography grant NNX17AH17G. C. Wang also would like to acknowledge the financial support of China Scholarship Council (CSC) for his PhD study.

ORCID

Chen Wang https://orcid.org/0000-0002-0575-742X

REFERENCES

Alpers, W., Zhang, B., Mouche, A., Zeng, K. and Chan, P.W. (2016) Rain footprints on C-band synthetic aperture radar images of the ocean - Revisited. *Remote Sensing of Environment*, 187, 169–185. ISSN 00344257. Available at: https://doi.org/10.1016/j.rse.2016.10.015.

Alpers, W. and Brümmer, B. (1994) Atmospheric boundary layer rolls observed by the synthetic aperture radar aboard the ers-1 satellite. *Journal of Geophysical Research: Oceans*, 99(C6), 12613–12621. ISSN 2156–2202. Available at: https://doi.org/10.1029/94JC00421.

Ardhuin, F., Chapron, B. and Collard, F. (2009) Observation of swell dissipation across oceans. *Geophysical Research Letters* ISSN: 00948276. Available at: https://doi.org/10.1029/2008GL037030.

Atkinson, B.W. and Wu Zhang, J. (1996) Mesoscale shallow convection in the atmosphere. *Reviews of Geophysics*, 34(4), 403. ISSN 8755–1209. Available at: https://doi.org/10.1029/96RG02623.

Babin, S.M., Sikora, T.D. and Winstead, N.S. (2003) A case study of satellite synthetic aperture radar signatures of spatially evolving atmospheric convection over the Western Atlantic Ocean. *Boundary-Layer Meteorology*, 106, 527–546. Available at: https://doi.org/10.1023/A:1021236600569.

Brown, R.A. (1980) Longitudinal instabilities and secondary flows in the planetary boundary layer: a review. *Reviews of Geophysics*, 18(3): 683–697. ISSN 19449208. Available at: https://doi.org/10.1029/RG018i003p00683.

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. ISSN 02705214.

Chen, S., Wang, H., Xu, F. and Jin, Y.Q. (2016) Target classification using the deep convolutional networks for SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8), 4806–4817. ISSN 0196–2892. Available at: https://doi.org/10.1109/TGRS.2016.2551720.

Cheng, G., Han, J. and Lu, X. (2017) Remote sensing image scene classification: benchmark and state of the art. *Proceedings of the IEEE*, 105(10), 1865–1883. ISSN 00189219. Available at: https://doi.org/10.1109/JPROC.2017.2675998.

Collard, F., Ardhuin, F. and Chapron, B. (2009) Monitoring and analysis of ocean swell fields from space: New methods for routine observations. *Journal of Geophysical Research: Oceans*. ISSN 21699291. Available at: https://doi.org/10.1029/2008JC005215.

Espedal, H.A., Johannessen, O.M. and Knulst, J. (1996) Satellite detection of natural films on the ocean surface. *Geophysical Research Letters*, 23(22), 3151–3154. ISSN 00948276. Available at: https://doi.org/10.1029/96GL03009.

Etling, D. and Brown, R.A. (1993) Roll vortices in the planetary boundary layer: a review. *Boundary-Layer Meteorology*, 65(3), 215–248. ISSN 1573–1472. Available at: https://doi.org/10.1007/BF00705527.

Hasselmann, K., Raney, R.K., Plant, W.J., Alpers, W., Shuchman, R.A., Lyzenga, D.R *et al*. (1985) Theory of synthetic aperture radar ocean imaging: a MARSEN view. *Journal of Geophysical Research: Oceans*, 90(C3), 4659–4686. ISSN 2156–2202. Available at: https://doi.org/10.1029/JC090iC03p04659.

Hersbach, H. (2010) Comparison of C-Band scatterometer CMOD5.N equivalent neutral winds with ECMWF. *Journal of Atmospheric and Oceanic Technology*, 27(4), 721–736. ISSN 07390572. Available at: https://doi.org/10.1175/2009JTECHO698.1.

Jackson, C.R. and Apel, J.R.eds (2004). Synthetic Aperture Radar Marine User's Manual. US Department of Commerce, National Oceanic and Atmospheric', Administration, National Environmental Satellite, Data, and Information Serve, Office of Research and Applications, Washington, DC.

Jia, T., Liang, J.J., Li, X.M. and Sha, J. (2018) SAR observation and numerical simulation of internal solitary wave refraction and reconnection behind the Dongsha Atoll. *Journal of Geophysical Research: Oceans*. ISSN: 21699291. Available at: https://doi.org/10.1002/2017JC013389.

Johannessen, J.A., Shuchman, R.A., Digranes, G., Lyzenga, D.R., Wackerman, C., Johannessen, O.M *et al*. (1996) Coastal ocean fronts and eddies imaged with ERS 1 synthetic aperture radar. *Journal of Geophysical Research*, 101(C3), 6651. ISSN 0148–0227. Available at: https://doi.org/10.1029/95JC02962.

Kerbaol, V., Chapron, B. and Vachon, P.W. (1998) Analysis of ERS-1/2 synthetic aperture radar wave mode imagettes. *Journal of Geophysical Research: Oceans*, 103(C4), 7833–7846. ISSN 2156–2202. Available at: https://doi.org/10.1029/97JC01579.

Khalsa, S.J.S. and Greenhut, G.K. (1985) Conditional sampling of updrafts and downdrafts in the marine atmospheric boundary layer. *Journal of the Atmospheric Sciences*, 42(23), 2550–2562. ISSN 0022–4928. Available at: https://doi.org/10.1175/1520-0469(1985)042<2550:CSOUAD>2.0.CO;2.

Soh, L.K. and Tsatsoulis, C. (1999) Texture analysis of sar sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2 I), 780–795. ISSN 01962892. Available at: https://doi.org/10.1109/36.752194.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep learning. *Nature*, 521(7553), 436–444. ISSN 0028–0836. Available at: https://doi.org/10.1038/nature14539.

Lee, J.S., Jurkevich, L., Dewaele, P., Wambacq, P. and Oosterlinck, A. (1994) Speckle filtering of synthetic aperture radar images: a review. *Remote Sensing Reviews*, 8(4), 313–340. ISSN 02757257. Available at: https://doi.org/10.1080/02757259409532206.

Fu, L.-L. and Holt, B. (1982) Seasat views oceans and sea ice with synthetic-aperture radar, Vol 81. Pasadena, CA: California Institute of Technology, Jet Propulsion Laboratory.

Levy, G. (2001) Boundary layer roll statistics from SAR, *Geophysical Research Letters*, 28, 1993–1995. Available at: https://doi.org/10.1s029/2000GL012667.

Li, X., Zheng, W., Yang, X., Zhang, J.A., Pichel, W.G. and Ziwei, L. (2013) Coexistence of atmospheric gravity waves and boundary layer rolls observed by SAR*. *Journal of the Atmospheric Sciences*, 70(11), 3448–3459. ISSN0022–4928. Available at: https://doi.org/10.1175/JAS-D-12-0347.1.

Natrella, M. (2013)NIST/SEMATECH e-handbook of statistical methods. NIST/SEMATECH.

Rascle, N., Molemaker, J., Marié, L., Frédéric, N., Chapron, B., Lund, B *et al*. (2017) Intense deformation field at oceanic front inferred from directional sea surface roughness observations. *Geophysical Research Letters*, 44(11), 5599–5608. ISSN 19448007. Available at: https://doi.org/10.1002/2017GL073473.

Schneider, T., Teixeira, J., Bretherton, C.S., Brient, F., Pressel, K.G., Schar, C *et al*. (2017) Climate goals and computing the future of clouds. *Nature Climate Change*, 7(1), 3–5. ISSN 1758–678X. Available at: https://doi.org/10.1038/nclimate3190.

Stopa, J.E., Ardhuin, F., Husson, R., Jiang, H., Chapron, B. and Collard, F. (2016) Swell dissipation from 10 years of Envisat advanced synthetic aperture radar in wave mode, *Geophysical Research Letters*. ISSN: 19448007. Available at: https://doi.org/10.1002/2015GL067566.

Topouzelis, K. and Kitsiou, D. (2015) Detection and classification of mesoscale atmospheric phenomena above sea in SAR imagery. *Remote Sensing of Environment*, 160, 263–272. ISSN 00344257. Available at: https://doi.org/10.1016/j.rse.2015.02.006.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D. and Davidson, M., Attema, E *et al*. (2012) GMES Sentinel-1 mission. *Remote Sensing of Environment*, 120, 9–24. ISSN 00344257. Available at: https://doi.org/10.1016/j.rse.2011.05.028.

Ufermann, S. and Romeiser, R. (1999) Numerical study on signatures of atmospheric convective cells in radar images of the ocean. *Journal of Geophysical Research: Oceans*, 104, 25707–25719. Available at: https://doi.org/10.1029/1999JC900224.

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. ISSN 01480227. Available at: https://doi.org/10.1029/1999JC000051.

Young, G.S., Kristovich, D.A.R., Hjelmfelt, M.R., Foster, R.C. (2002) Rolls, streets, waves, and more: a review of quasi-two-dimensional structures in the atmospheric boundary layer. *Bulletin of the American Meteorological Society*, 83(7), 997–1001. ISSN 00030007.

Young, G.S., Sikora, T.N. and 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. ISSN 0882–8156. Available at: https://doi.org/10.1175/WAF853.1.

De Zan, F. and Monti Guarnieri, A. (2006) TOPSAR: terrain observation by progressive scans. *IEEE Transactions on Geoscience and Remote Sensing*, 44(9), 2352–2360. Available at: https://doi.org/10.1109/tgrs.2006.873853.

Zhang, L., Zhang, L. and Bo, D. (2016) Deep learning for remote sensing data: a technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2), 22–40. ISSN 21686831. Available at: https://doi.org/10.1109/MGRS.2016.2540798.

How to cite this article: Wang C, Mouche A, Tandeo P, et al. A labelled ocean SAR imagery dataset of ten geophysical phenomena from Sentinel-1 wave mode. *Geosci Data J*. 2019;00:1–11. https://doi.org/10.1002/gdj3.73