Support vector machine tropical wind speed retrieval in the presence of rain for Ku-band wind scatterometry

Xingou Xu[1], Ad Stoffelen[2]

[1] The CAS Key Laboratory of Microwave Remote Sensing, National Space Science Center, Chinese Academy of Sciences, Beijing, 100190, China

[2] Royal Netherlands Meteorological Institute, De Bilt, 3730 AE, The Netherlands

Correspondence to: Ad Stoffelen (ad.stoffelen@knmi.nl)

Abstract. Wind retrieval parameters, i.e., quality indicators and the 2DVAR analysis speeds, are explored with the aim to improve wind speed retrieval during rain for tropical regions. We apply the well-researched support vector machine (SVM) method in machine learning (ML) to solve this complex problem in a data-orientated regression. To guarantee the effectiveness of SVM, the inputs are extensively analysed to evaluate their appropriateness for this problem, before the results are produced. ~~Subsequently, triple collocation shows that the similarity of the resolved Ku-band (OSCAT-2) wind speed in rain is better than the 2DVAR speed, with respect to the collocated C-band (ASCAT) speed, which is much less affected by rain.~~ The comparisons between distributions and differences between data of rain-contaminated winds, corrected winds and good quality C-band winds, illustrate that the rain-distorted wind distributions become more nominal with SVM, hence ~~eliminating~~ reducing much the rain-induced biases and error variance. Further confirmation is obtained from a case with synchronous Himawari-8 observation indicating rain (clouds) in the scene. Furthermore, the ~~determination~~ estimation of simultaneous rain rate is attempted with some success to retrieve both wind and rain. Although, additional observations or higher resolution may be required to better assess the accuracy of the wind and rain retrievals, the Machine Learning (ML) results demonstrate benefits of such methodology in geophysical retrieval and nowcasting applications.

1 Introduction

It is well-known that the structure of the atmosphere and ocean depends on the motions driven by radiation affecting the redistribution of heat. The circulations imply wind convergence to elevate water vapour from the ocean surface that then forms clouds and rain, while rain, in turn, causes downdrafts. These interactions of the air and the ocean underneath are connected by the basic mass, momentum and energy equations involving winds, heat and moisture are vital for understanding the earth system (Gill, 1982). In the tropics, the resolution of moist convection is key for improving earth system simulations (Bony, et al., 2015).

Observed ocean surface wind-fields (OSW) are essential to investigate such processes and related applications. An efficient method of acquiring a large coverage and good quality OSW are the retrievals from scatterometers with an application history up to 40 years (Jones et al, 1979; Stoffelen et al., 2019). Scatterometers are real aperture radars providing stable and accurate

normalized radar cross-sections (NRCS) of the wind-roughened ocean surface in different azimuthal directions from oblique incidence angles. The winds are then obtained in a maximum likelihood estimation method (MLE) from the measured NRCSs within a wind vector cell (WVC) with reference to a Geophysical Model Function (GMF). Generally, a WVC is a square of the size 25 km × 25 km sized square, and GMFs are empirical models mapping NRCSs from scatterometers in different
35   frequencies, polarizations and observing geometries to winds.

    Rain products provide another important information for air-sea interaction. In the Global Precipitation Mission (GPM), one of the core instruments is the dual-frequency radar (DPR) working at Ku- and Ka-band in nadir-looking mode. Rain is then obtained by relating the radar cross-sections to a chosen distribution of precipitation particles. Meanwhile, rain products from infrared observations are also widely used, for example, rain rates from the Spinning Enhanced Visible and Infrared Imager
40   (SEVIRI) on-board the Meteosat Second Generation (MSG) Satellite, which is derived by considering retrieved cloud condensed water path (CWP), particle distribution, and cloud thermodynamic phase (E. L. A. Wolters, M. Van Den Hurk et al, 2011). Both rain products described above are good references for rain in Ku-band wind scatterometry (Xu et al., 2020a), though the high spatial and temporal variability of rain generally challenges small collocation errors and high correlation between instantaneous rain data sets (e.g., Liu et al., 2020).

45   Combined retrievals of wind and rain are generally applying synchronous passive measurements from radiometers for rain in the scatterometer case (Stiles et al., 2010) while and in addition to rains for winds are retrieved in GPM researches (Li et al., 2014). Radiometer winds are of coarser spatial resolution and are not adept for wind direction retrieval, which would require the third and fourth Stokes parameters that are now generally obtained in low signal to noise ratio (SNR). Scatterometers are not specifically designed for acquiring precipitation profiles. When rain clouds affect the observations, the winds obtained
50   from a wind GMF will deviate from the truth, resulting in biases in the retrieved wind and an increased retrieval residual, called MLE. Since rain is spatially more heterogeneous than winds are, rain can be captured and estimated in the NRCS set within a WVC. Considering the distances of NRCS observations to the wind GMF, the retrieved wind and with the reference to rain observations from the Tropical Rainfall Measuring Mission (TRMM) Precipitation Radar (PR), wind and rain may be segregated (Owen et al., 2011; Draper et al., 2004). Furthermore, the heterogeneous rain within a WVC can be depicted from
55   indicators applied in scatterometer Quality Control (QC) (Portabella et al., 2002; Lin et al., 2017). J_{oss} is a recent indicator developed for tropical regions for rain screening, which has been verified to correlate well with rain for Ku-band scatterometers (Xu et al., 2020a; Xu, et al., 2020b). From a conceptual point of view, the MLE identifies the WVC NRCS sets that do not follow the wind GMF. Two main reasons have been identified for such discrepancy, which are 1) enhanced wind variability and 2) rain. Besides Fortunately, collocated operational C and Ku-band observations are available, when due to the longer
60   wavelength at C band (about 5 cm) than Ku band (about 3 cm), standard QC, based on MLE, rejects ten times more Ku-band winds than atC band winds, i.e., about 5% of its observations. Specifically Hence, specifically, in tropical regions, the accepted C-band winds can be used to verify their Ku-band collocations, which helped to develop the Joss indicator and verify the performance of the other Ku-band QC indicators. In addition, extreme convergence and divergence in C-band winds has been related to tropical moist convection and rain, where convergence proceeds rain by about 30 minutes, while extreme divergence

65   occurs simultaneously with rain in convective downdrafts for C-band winds, hence illustrating the physical integrity of C-band winds in the presence of rain. C-band rejections correspond to the most extreme variability in WVCs, including wind gradients induced by heavy precipitation downdrafts (King et al., 2017). The different rain signatures in C- and Ku-band scatterometers can cast a light on developing methods for correction of the rain-affected winds in Ku-band scatterometer retrievals by referring to their C-band collocations. Particularly, the combination of *MLE* and *Joss* appears promising to segregate wind variability

70   and rain effects in Ku-band retrievals.

    To derive the complexly associated wind and rain information referred to above, machine learning (ML) may prove to be a powerful tool, which can be applied with knowledge of the validity of the underlying principles (Reichstein et al., 2019). In fact, ML methods have long been well-researched in wind scatterometry (Thiria et al., 1993; Stiles et al., 2010). For common roughness conditions, it cannot exceed the performance of GMF-based methods (Cornford et al., 1999), but ML may be

75   effective in rainy conditions. Among the ML methods, support vector machine (SVM) is one based on the Mercer theorem, by complement of empirical risk minimization with Vapnik-Chervonenkis (VC) confidence, infer statistical relations without a priori distributions and gives no regional minimum (Vapnik, 1998). It can establish an information space based on the training set and if the data applied in training is well representative for the problem, it also requires less samples than other ML methods. Aside, SVM already provides good results in rain rate estimates (Kumar et al, 2021).

80   In this research, SVM is applied for wind correction of rain-affected winds of Ku-band scatterometers, considering quantified rain and rain effect information captured in the QC indicators of Ku-band observations. The GPM rain products and collocated accepted winds from C-band products are used as references. When this SVM model has been established, without C-band collocations, the rain contaminated winds can be corrected with Ku-band winds and their QC indicators alone. First, in the method part, the underlying principles of the problem of rain signatures in scatterometry are addressed in detail with a

85   brief of error requirement for assimilation application before data description. Then, in the experimental part, results for the testing set, not applied in the training procedure, demonstrate a minimum mean difference of -0.12 m/s at about 8 m/s and a largest difference of -3.25 m/s at about 14 m/s ASCAT speed. The distribution~~s~~ of the corrected winds~~are obtained and,together with~~ the scatter plots against C-band winds ~~is~~are inspected, with ~~the~~ a check on wind differences in each wind speed bin of the original and corrected winds against ASCAT winds, ~~prove~~ proving the more unbiased and symmetric error of the corrected

90   set, ~~addressing~~ illustrating the advantage of applying SVM. ~~Together with the analysis on~~ The similarity of the corrected distribution~~s~~ with the references provided from collocated ASCAT winds and the reduced mutual differences, ~~it~~indicates ~~is indicated~~ that to a certain extent the local (WVC) wind scales are recovered ~~from analysis of the error distributions~~ by the SVM corrections. Result~~s~~ suggest~~s~~ that the method resolves the heterogeneity induced by rain clouds in *MLE* and *Joss* ~~from~~ with the setting~~s~~ of the proposed SVM. Furthermore, a case with~~out~~ ~~data that is not with~~ rain collocations, and thus not involved

95   in deriving the ~~method of this research~~ corrections, is provided as a case study for verification~~case study~~. Where simultaneous image~~s~~ from the Himawari-8 provide~~s~~ a concrete view of the rain clouds in the scene.

    In the discussion part, rain labelling and regression SVMs are established with the same inputs, attempting rain estimation from scatterometer winds by employing SVM. The rain identification accuracy is 72% for the independent test set not applied

in the training procedures. While for rain rate estimation, the correlation coefficient of SVM rain with GPM products achieves
100   0.47 for the independent testing set. An analysis of the uncertainties in the SVM model and possible improvements in the rain estimation procedure are also discussed. The corrected winds increase the global wind coverage and, in synergy with the rain information ~~research~~ provided, benefit now-casting applications (Majumdar et al, 2021). This research illustrates an example of ~~n a~~ complex ~~scene of~~ datadriven ML methods, complementary to traditional methods in complex problems, which motivates and demonstrates the adhibition of the ML method in meteorological applications.

2 Method

Research on observation errors, i.e., the deviations from the truth, together with the monitoring information obtained from differences between scatterometer winds and models, supports NWP. ~~before obtaining monitoring information of differences between scatterometer winds and models.~~ Among the errors ~~which~~, undetermined geophysical dependencies including rain effects are to be corrected to better understand model biases (Stoffelen et al., 2021), while it cannot be achieved by a first order

110   correction. Apart from this, the control variables, define multivariate background errors and correlated errors between variables are modeled by linear regression (Descombes et al, 2015). Also, the 3DVAR and Kalman Filter assumes linear or quasi-linear and Gaussian features in observation operator and error distributions respectively, when 4DVAR considers additional dynamical constraints in the time dimension (Parrish et al., 1992; Courtier et al., 1994). Hence linearized Gaussian or quasi-Gaussian errors are vital for the assimilation of observations. We seek to address and correct biases in Ku-band scatterometer

115   wind retrievals due to rain. In the following part, first, the complex rain signatures in wind scatterometer observations are analyzed, demonstrating non-Gaussian error features before the principles of SVM are introduced.

2.1 Rain characteristics in MLE, J_{oss} and the fractal parameter α

When compared to the C-band winds that are of good quality (accepted), collocated Ku-band QC-rejected WVCs in tropical regions are affected by rain due to the shorter observing wavelength (Xu et al., 2020a). The wind Quality Control (QC) is

120   determined by QC indicators, and the indicator widely applied in operational wind products, is the MLE residual obtained through wind inversion. Using all N NRCS measurements obtained within a WVC Maximum Likelihood Estimation (MLE) procedures are applied for wind retrieval. The MLE residual is a normalized Euclidian distance to the cone determined by GMFs (Stoffelen et al., 1997):

$$MLE = \frac{1}{N} \sum_i^N \frac{\left(\sigma_i^o - \sigma_{sim_i}\right)^2}{\left(K_{pi} \bullet \sigma_{sim_i}\right)^2} \tag{1}$$

125   Where σ_i^o is the i th NRCS of the N NRCSs within a WVC, K_{pi} is a dimensionless constant determined by instrument noise ~~represents the variance of NRCS within this WVC~~, and σ_{sim_i} is from a wind GMF indexed by observing geometry and the local wind vector. Before wind inversion, NRCS are well-calibrated for instrumental as well as GMF uncertainties that are generally small (~2%), and are reproducible or systematic. NRCS calibration and GMF bias term uncertainties lead to wind

speed probability density function variations. Errors in the harmonic terms of the GMF may lead to wind direction errors, and

130   in systematic wind speed errors that have associated wind direction errors and vice versa (Portabella et al., 2002). During the 2DVAR procedure that optimizes wind vector selection (Vogelzang et al., 2011), essentially the WVC MLE associated with the selected direction is determined. At the same time, the 2DVAR low-pass filtered analysis winds, which are here referred to as 2DVAR winds, are calculated When rain affects the NRCS, the GMF does not represent the NRCS measurements well, as rain effects are not considered in the wind GMFs (Stoffelen, 1998). Therefore, this part of the GMF error due to missed or

135   incompletely modelled rain processes generates errors of a class that cannot be eliminated by calibration, and induces deviation of error distributions from the well-calibrated random Gaussian shape. Note that the KNMI QC flag is based on MLE values, and in the Ku-band rejections and C-band acceptances in tropical regions, the rejections are mainly caused by rain. Hence, MLE values of the 2DVAR selected Ku-band wind can be related to rain effects that alter the amplitudes of NRCSs.

      However, at the same time, the 2DVAR winds do not use QC-flagged WVCs and are hence not affected by local disturbances

140   introduced by rain. The wind speed correction procedure employed here, hence does not change the 2DVAR analysis field, nor the selected wind direction at the rain-affected WVCs obtained during the elaborated 2DVAR Multiple Solution Scheme (MSS) (Vogelzang et al, 2018). Then, the rain effect is estimated by the wind speed difference of the 2DVAR analysis wind speed f and the selected observational wind speed f_s, corresponding to the wind direction obtained by 2DVAR (Xu et al., 2020a):

145    $J_{\text{oss}} = f - f_s$ (2)

      As Note that the 2DVAR winds are low-pass filtered and of relatively coarse resolution, and ignoring rain-affected WVCs through MLE-based QC (J. Vogelzang, 2017). The Since the spatially heterogeneous tropical rain clouds are generally of smaller spatial scale than a WVC, rain effects in the 2DVAR analysis winds can be ignored and taken as the *true* winds (J. Vogelzang and A. Stoffelen, 2018). Hence, J_{oss} values can screen and eliminate false alarm rate (FAR) for MLE based QC

150   results for Ku-band wind products after 2DVAR processing, indicating rain information (Xu et al, 2020b).

      Usually rain clouds will cause negative Joss for wind speeds below 15 m/s. A WVC is usually partially heavy rain, and since Ku-band rain saturates around 18m/s, hereafter the parameter for area fraction α for Ku-band winds can be expressed as:

      $\alpha = \frac{Joss}{f-18}$ (3)

155   As 18 m/s winds cannot be distinguished from rain and to allow rain sensitivity, the rain effect correction set is limited to:

      $J_{\text{oss}} > 0.33\,f - 5$ (4)

      For retrieved 2DVAR speed smaller or equal to 11m/s. For 2DVAR wind speed larger than 11m/s, the FAR set is limited to Joss < -1.33 (Xu et al., 2021). Then the negative values of α corresponding to positive J_{oss} when wind speeds are smaller than 18 m/s can be owed to effects of local variance of the ocean surface. Larger wind speed than 18 m/s and positive J_{oss} may

160   happen when both rain and winds are large in the scene. For tropical rain this practically only occurs in hurricanes, but has not

yet been investigated with respect to J_{oss} in the criterion above. Thus, this parameter can provide relative information of rain within the WVC from 2DVAR residuals.

Enhanced wind variability enhances MLE, due to beam collocation errors. In particular, extreme wind convergence and divergence is associated with heavy rain (King et al., 2017). The wind variability associated with heavy precipitation may enhance the wind speed, just like rain does at Ku-band, but which has been investigated by comparing the 2DVAR winds with ASCAT winds. ASCAT winds are equally sensitive to wind speed variations at the surface, but much less sensitive to rain cloud scattering effects. Hence, the effect due to amplitude alternations for a single NRCS in a tropical scene with rain clouds can be obtained by the rain screening ability of J_{oss}.

From the above contents and equations, rain effects can be represented by MLE, J_{oss}, and the observational wind in the Ku-band retrieval, while the 2DVAR analysis wind provides information on rain sensitivity. In this research, for the C-band QC accepted and Ku-band rejected WVCs, after the FAR set is eliminated, the Ku-band WVCs are collocated with rain rates from GPM products. Then MLE, J_{oss} values and the 2DVAR winds and observational winds are applied as inputs to the SVM model, with the training destination set as the collocated C-band winds. In the established model, corrected winds closer to the observed C-band winds may ~~can~~ be obtained for rain-affected Ku-band WVCs, by eliminating non-Gaussian errors within a WVC caused by rain. Moreover, the SVM model, when established, could ~~can~~ be applied for Ku-band rejections.

1.2 The principle of SVM regression

The SVM regression procedures map input vectors to a space of higher dimension before the regression is conducted. When the mapping is obtained and thus described by kernel functions determined from the training sets, non-linear features are linearized. This provides a possibility for solving problems that are non-convex and difficult to solve in the original input space, as well as linearizing intricate relations. Specifically, during the training procedure, weights for the input vectors in the training set in the mapped space is determined, and the corresponding support vectors (SVs) can be identified by the values of corresponding weights. While the weights are applied to scale similarities with other vectors in the training set. On the other hand, they are obtained by minimizing distances with the targets of the training vectors. Moreover, the similarity is measured between the kernel function mapped inputs. In this way, it allows the data involved in training to embody the underlying model in a space that facilitates information extraction. Furthermore, L2-normdistance minimization is achieved by an objective function expressed as the distances between the vectors in the training sets to the plane fixed by the weighted support vectors in the mapped space (Vapnik, 1998).

The employed kernel functions are linear, generally polynomial or Gaussian radial basis functions (RBF). Among them, the RBF, or the Gaussian kernel, is superior in unlimited dimension mapping and easier in hidden parameter setting. For RBF, the similarity between a vector x and the selected support vector $l^{(1)}$ is expressed as (Vapnik, 1998; Smola et al., 1998):

$$f_1 = \exp(-\frac{||x - l^{(1)}||^2}{2\sigma^2}) \tag{4}$$

Where σ is the scale parameter weighting the similarity of x and $l^{(1)}$. And the larger the value of σ is, the more x and $l^{(1)}$ can be taken as similar. If the L_2 distance is applied:

$$f^1 = \exp(-\frac{\sum_{j=1}^{n}(x_j - l_j^{(1)})^2}{2\sigma^2})$$ (5)

195 When θ_i are weights, $y^{(i)}$ is the target value corresponding to x_i, the objective function can be expressed as:

$$min_\theta(C\sum_{i=1}^{m} y^{(i)} cost_1(\theta^T f^{(i)}) + (1 - y^{(i)})cost_0(\theta^T f^{(i)}) + \frac{1}{2}\sum_{j=1}^{m}\theta_j^2)$$ (6)

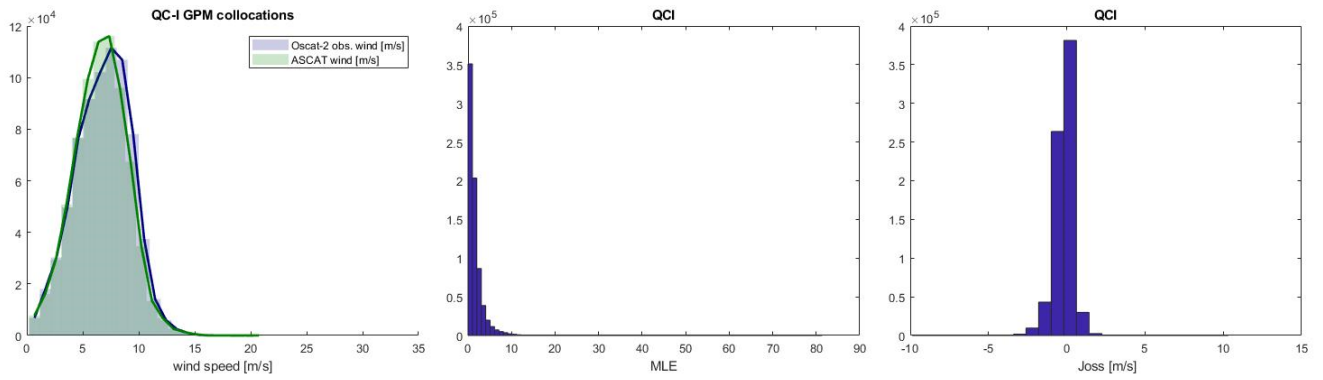
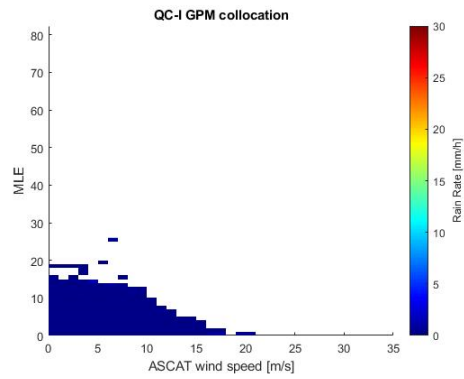
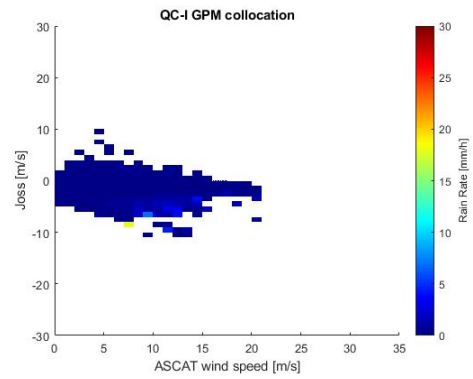
where C is the relaxation coefficient and the L2 distance is applied as the cost functions $cost_1$ and $cost_0$ (Smola et al., 1998; Chang et al., 2019).

3. Data and experiments

200 **3.1 The expression of rain in wind retrieval parameters**

The representativeness of the data sets from which the featured SVs are obtained is vital in the SVM procedure. In this research, the C- and Ku-band collocations of scatterometer winds are from the Advanced scatterometer-A (ASCAT-A) and ASCAT-B onboard the Meteorological Operational Satellite Program of Europe (MetOp) series and the Scatterometer onboard the Scatsat-1 Satellite (OSCAT-2) respectively (Eumetsat, 2018), (E-Directory, 2019). Then the ASCAT-A, ASCAT-B and
205 OSCAT-2 L2 wind products are from the Ocean and Sea Ice Satellite Application Facility (OSI SAF) of the European Organization for the Exploitation of Metrological Satellites (EUMETSAT), over a period from October 2016 to January 2019. The WVC sizes are 25 km×25 km on the Earth surface. Where the OSCAT-2 Ku-band winds are sea surface temperature (SST) corrected sweet swath WVCs with ~~more optimal~~ better NRCS azimuth diversity than the nadir and edge swath (M. Portabella, 2002). The collocation time lag is within 30 minutes (min) with the spatial distances between ASCAT~~s~~ and OSCAT-2 WVC
210 centres less than 12.5 km. While ~~And~~ the background winds are from the European Center for Medium-range Weather Forecasts (ECMWF), the 10-m 3 hourly forecast 0.125° winds are used. GPM rain products used here are the version-5 0.1°-gridded Integrated Multi-satellitE Retrievals for GPM-F (IMERG-F) (Huffman et al, 2018) within a time difference to OSCAT-2 WVCs of 4.8 min. Furthermore, rain products are area-weighted over the OSCAT-2 WVCs to obtain WVC-representative rain rates (Xu et al., 2020c). Finally, for validation, the images of the 11th band (medium infrared, MI, 8.6μm)
215 with 2 km resolution in the tropics are also used for reference (Otemachi, C.-k. et al., 2015).

Figure 1 plots the 732,614 collocated wind speeds in ASCAT-A accepted and OSCAT-2 accepted set (QC-I set) in (a), corresponding MLE values of OSCAT-2 in (b), J_{oss} in (c), collocated rain rates in (d) and (e). Figure 2 shows the same plots for the ASCSAT-A accepted winds, but now for rejected OSCAT-2 collocations (QC-II), after that the false alarms in the KNMI OSCAT flags were eliminated by J_{oss} (FAE), with 9,339 WVCs (Xu et al, 2020b).

220

225 **Figure 1: Collocated wind speed distributions in the QC-I set (a), corresponding MLE distribution of OSCAT-2 (b), J_{oss} (c), collocated rain rates with reference to MLE (d) and J_{oss} (e).**

230   (d)   (e)

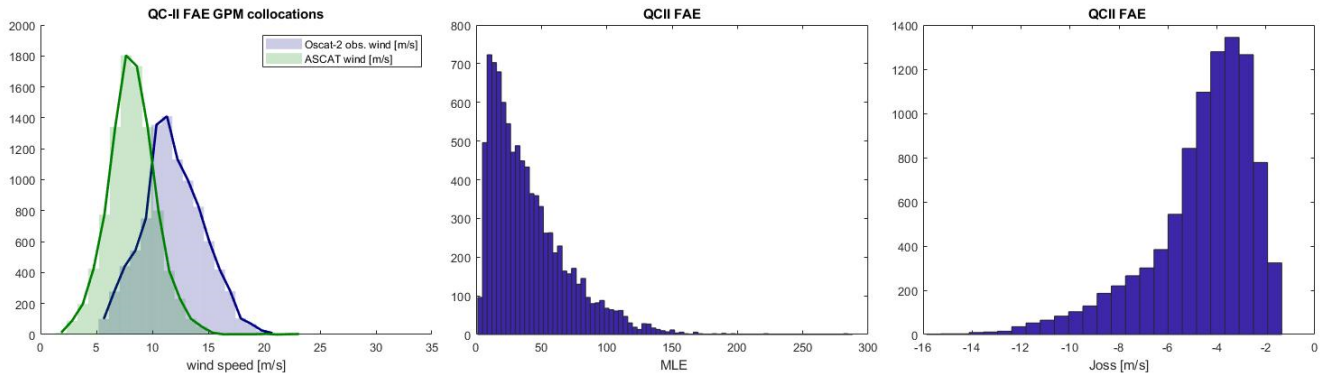
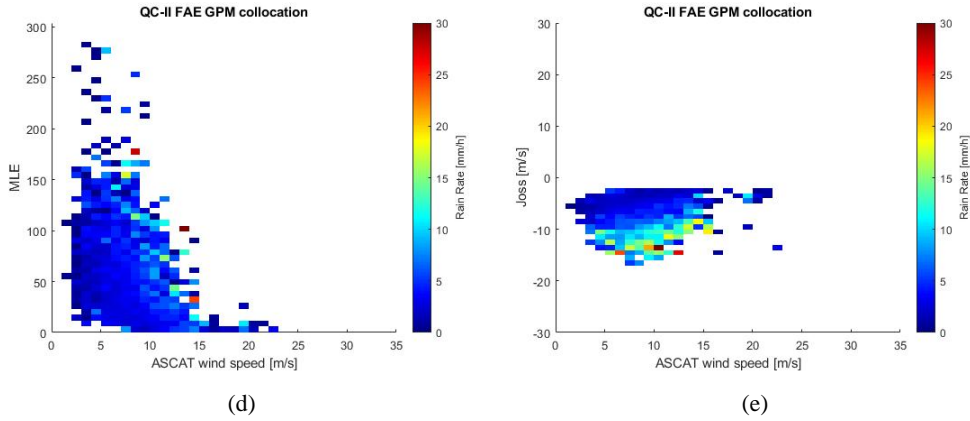



Figure 2: Collocated wind speed distributions in the QC-II FAE set (a), corresponding MLE distribution of OSCAT-2 (b), J_{oss} (c), collocated rain rates with reference to MLE (d) and J_{oss} (e).

In Figure 1 (a) we note that observed wind distributions from ASCAT and OSCAT-2 are similar, while in Figure 2 (a), the
235   Ku-band winds are much elevated with respect to ASCAT and clearly suspect, as the ASCAT wind ~~while latter~~ distribution

appears nominal and similar to that in Figure 1(a). The MLE values are mostly nominal and distributed over the bins under 10

in Figure 1 (b), while typical values are very large and around 50 in Figure 2 (b). ~~In~~ For comparison ~~By design~~, in (c) of Figure

1, J_{oss} values are small with values close to 0 where in Figure 2 values are typically 4 m/s. Comparing (d) and (e) in ~~the~~ both

Figures, there is little rain in QC-I, while rain is dominant in Figure 2, consistent with both the elevated MLE and J_{oss} values.
240   Also, in Figure 2 (e), the criterion of J_{oss} in the FAE set can be observed from its upper limit.

We note from Figure1 and Figure 2 that rain casts effects on OSCAT-2 data while collocated ASCAT winds remain of

acceptable quality. The winds distorted by rain (clouds) are clearly segregated by the FAE, resulting in a deformed speed

distribution, as well as much elevated MLE and J_{oss}, that all can be potentially related to WVC rain rate.

3.2 SVM for Ku-band wind correction in rain

245   For the correction of rain effects a SVM model is established, where the inputs are determined by the wind-rain related

parameters, as described in the previous sections. Specifically, the inputs and outputs are in Table 1.

Table 1 Inputs and outputs for the SVM of wind speed correction:

Inputs	Output	Values of output for training
OSCAT-2 MLE in dB	Corrected OSCAT-2 wind speed [m/s]	ASCAT wind speed [m/s]
α from OSCAT-2 WVC		
OSCAT-2 2DVAR speed		
OSCAT-2 observational speed		

The SVM tool from *sklearn* is applied, which is based on the *libsvm* to realize the procedure described in section 2 for SVM
250   (C-C Chang and C-J Lin, 2019). In total, there are 18,528 WVCs obtained from FAE in OSCAT-2 collocations for ASCAT-A and ASCAT-B together. Among them, 70% (12, 969 WVCs) are used in training and 30% (5,559 WVCs) for testing or validation. Note that the testing set is not applied in the training procedure.

4. Results and validations

4.1 Results

255   Starting from the large input biases illustrated in Figure 3 (a), typically 5 m/s, Figure 3 shows the corrected winds against the accepted winds from ASCAT-A and ASCAT-B for the training set in (a) and the validation set in (b). While in (c) and (d), the observational winds and 2DVAR winds of OSCAT-2 are also plot against ASCAT winds. With some of the corresponding statistics listed from (a) to (d), in Table 2.

260   (a)   (b)

                               (c)   (d)

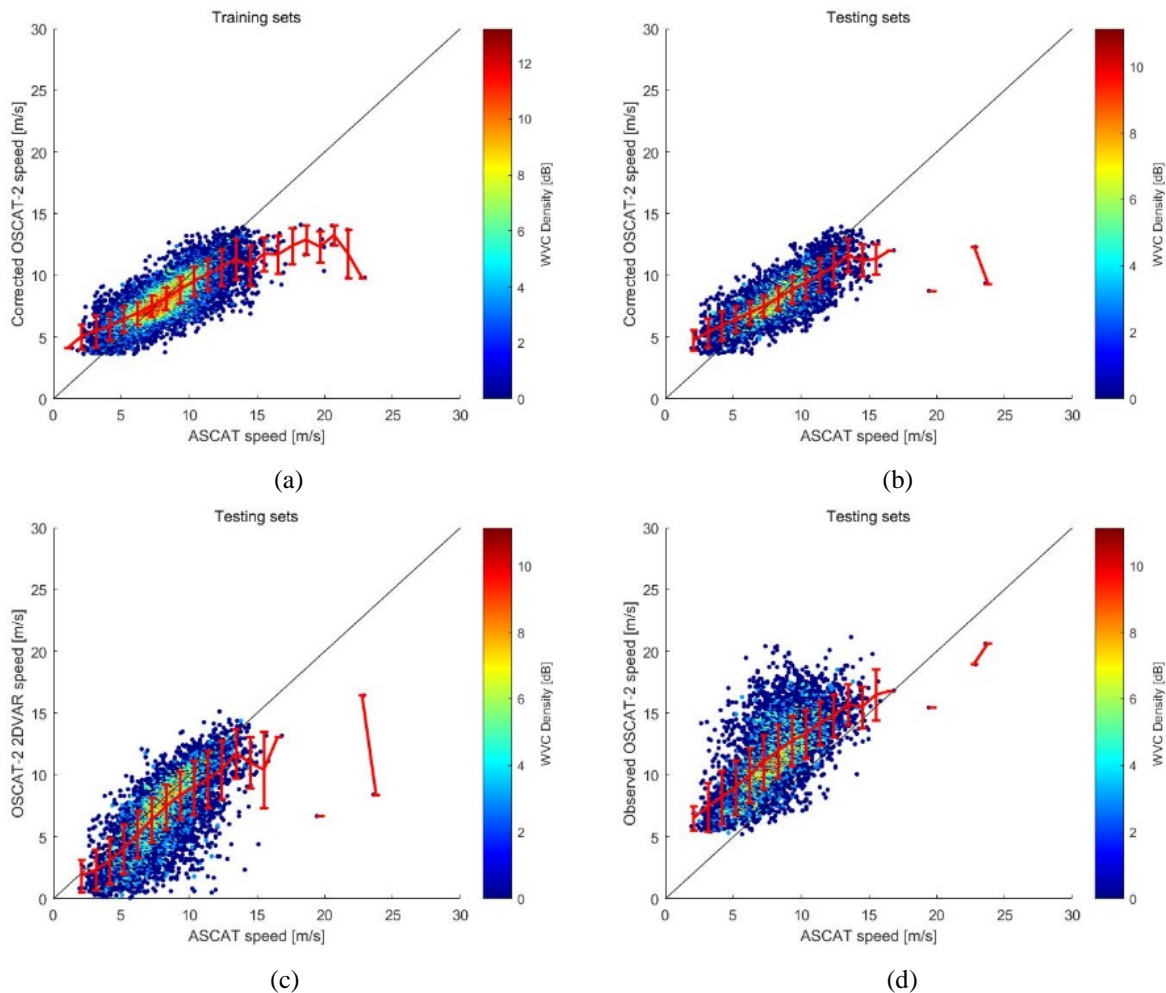



Figure 3: The corrected winds against accepted winds from ASCAT-A and ASCAT-B for the training set (a) and validation set (b)-(d), where (b) the corrected, (c) the 2DVAR and (d) observational OSCAT-2 wind speed against ASCAT wind speed are depicted.

265

Table 2 Corresponding mean and standard deviation of difference (SDD) statistics to Figure 3 (a-d).

(a) Corrected winds in the training set

ASCAT-A & ASCAT-B wind speeds [m/s]	Mean values of the corrected winds [m/s]	SDD corrected - ASCAT winds [m/s]
4.14	5.86	1.11
6.21	6.95	0.91
8.28	8.13	0.93
10.34	9.54	1.21
12.41	10.66	1.50
14.48	10.79	1.59

(b) Corrected winds in the validation set

ASCAT-A & ASCAT-B wind speeds [m/s]	Mean values of the corrected winds [m/s]	SDD corrected – ASCAT winds [m/s]
4.14	5.93	1.14
6.21	6.97	0.92
8.28	8.16	0.96
10.34	9.39	1.26
12.41	10.70	1.49
14.48	11.23	1.30

(c) OSCAT-2 DVAR winds (testing set)

ASCAT-A & ASCAT-B wind speeds [m/s]	Mean values of the 2DVAR~~observational~~ winds [m/s]	SDD 2DVAR – ASCAT winds [m/s]
4.14	3.06	1.86
6.21	5.22	1.95
8.28	7.43	1.84
10.34	9.01	2.19
12.41	10.46	2.37
14.48	11.04	2.04

(d) OSCAT-2 observational winds (testing set)

ASCAT-A & ASCAT-B wind speeds [m/s]	Mean values of the observational ~~2DVAR~~ winds [m/s]	SDD observational – ASCAT winds [m/s]
4.14	8.22	2.17
6.21	10.02	2.21
8.28	11.96	2.15
10.34	13.32	1.99
12.41	14.82	1.70
14.48	15.50	1.67

270

As can be seen from Figure 3 (a) and (b), and from the corresponding values in Table 2 (a) and (b), the testing set exhibits similar statistics to the training set for wind speed correction established by SVM. Note that most of the QC-II FAE wind speeds are distributed from about 4 m/s to 14 m/s, which is typical for rain clouds in moist convection (Xu et al., 2020a). For speeds in this range, the largest differences of mean values with the bin centre values are -3.69 m/s and -3.25 m/s at about 14

275   m/s ASCAT speed for the training and testing set, respectively. Then the bias value decreases as wind speed decreases. And for both sets, reach a minimum at about 8 m/s of -0.15 m/s and -0.12 m/s. Then the bias increases with decreasing wind speeds,

to 1.72 m/s and 1.79 m/s at about 4 m/s. This trend is consistent with the SDD, with the smallest SDD of 0.87 m/s for both sets at about 7m/s. The consistency of the training set and testing set indicates the stability of the SVM model established. Besides, it is noteworthy that there is a sign change for these speed differences, suggesting an excessive speed range suppression for

280 wind speeds both lower and higher than around 8 m/s respectively. This trend also exists in Figures 3(c) and 3(d) of the observational and 2DVAR wind against ASCAT winds, as seen from the curvature of the red lines representing mean bin values, though they are generally smaller and larger than the ASCAT wind speed for the 2DVAR and observational speed respectively, while the distances are larger in absolute values for the observational winds. This is consistent with the fact that the OSCAT 2DVAR wind filters the details of the local wind changes, ignoring wind variability due to rain that is captured by

285 the C-band observations of good quality at finer resolutions. We further note that Figure 3 and Table 2 are based on a conditional binning of ASCAT winds, while ASCAT winds are not perfect and OSCAT is not perfectly collocated with ASCAT. Such uncertainty in ASCAT also has the tendency to flatten the red curves in Figure 3.

In Figure 4, the distributions of wind speed of the OSCAT-2 observational wind speed, OSCAT-2 2DVAR speed, collocated ASCAT speed and that of the SVM-corrected speed are displayed for the testing set.

290

(a)   (b)

(c)

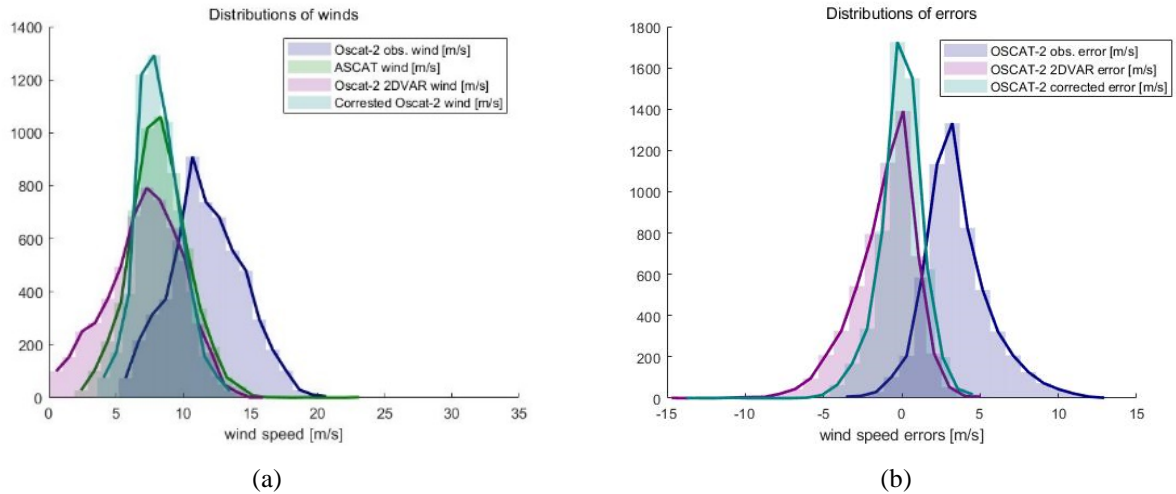
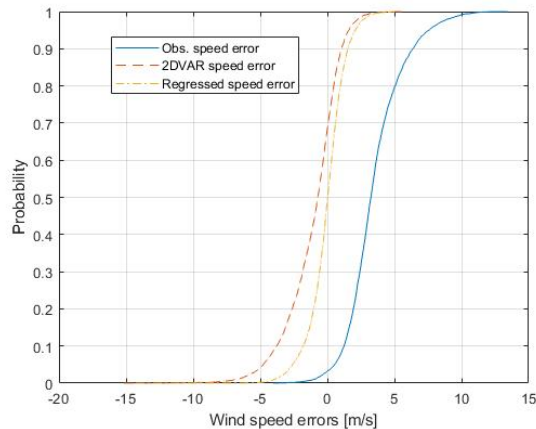



Figure 4: Distribution of the different wind speeds (a), speed errors with the reference from ASCAT (b) and the cumulative distribution function (CDF) of speed errors (c) corresponding to the testing set.

From Figure 4 (a), the blue curve indicates rain affected OSCAT-2 winds are elevated and skewed to higher speeds, peaking at around 12 m/s. They also deviate from the corresponding 2DVAR speeds (purple) as well as the collocated ASCAT winds (green). Similar to the latter two, the SVM-corrected winds (lighter blue) peak at a similar speed around 8 m/s. This is also consistent with Figure 1(a). Moreover, note that the 2DVAR wind distribution extends to the lowest speeds and deviates more than the corrected winds from ASCAT observations.. Anyway, the corrected winds show a very similar shape to the ASCAT distribution, proving the effectiveness of the SVM. Figure 4 (b) demonstrates the speed errors defined as the differences with respect to the ASCAT observations. Consistent with (a), the errors distribute more symmetrically and over the smallest range for the corrected winds. The more Gaussian-like features of this speed error as compared to the other groups can be more easily observed from (c) where the cumulative distribution function (CDF) is obtained. In the figure, the blue, red and black lines are the CDFs of observational, 2DVAR and regressed speed error respectively. Except for the most symmetric feature of the black curve in bias, about 90% of the values lay between -2.0 and 2.0, which indicates again that the corrected winds are close to the ASCAT observations. In addition to Figure 4, Figure 5 demonstrates in detail and directly from the data that the statistics has been improved after SVM corrections.

(a)

(b)

(c)   (d)

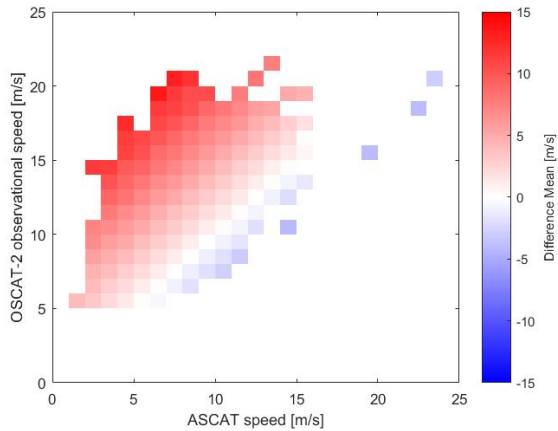
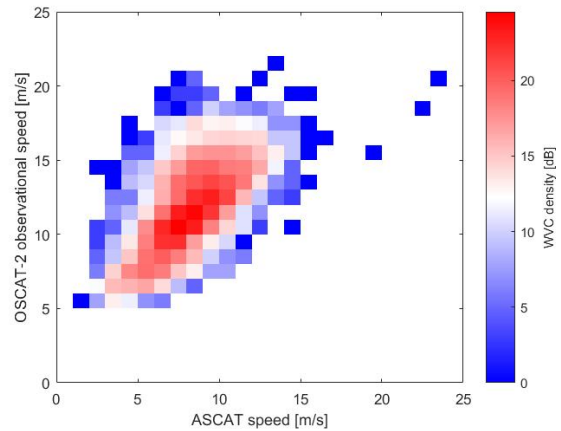
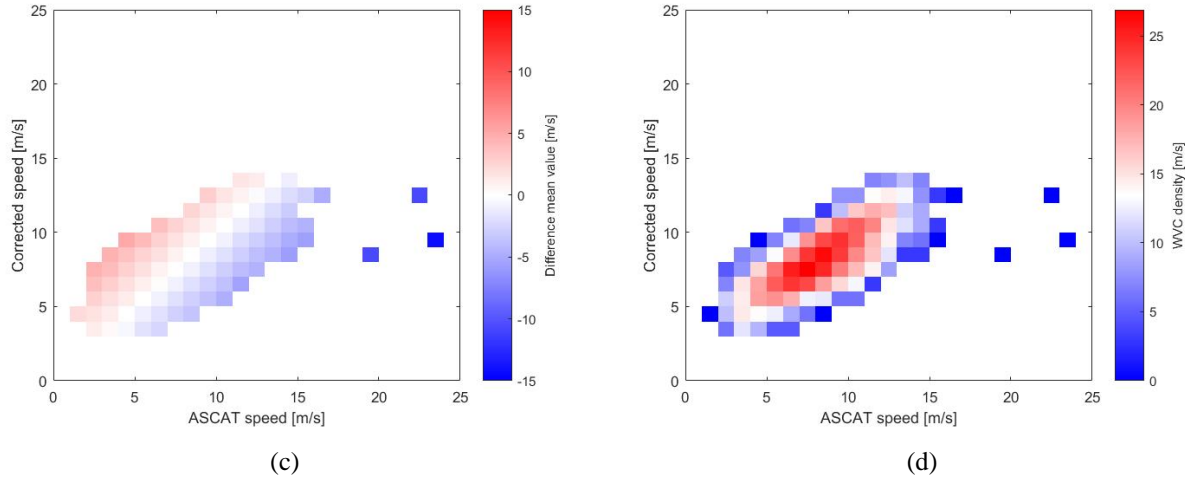



Figure 5: Mean values in 1 m/s bin of winds against mean difference values in the same bin of vertical minus horizontal values (colour) (a) and (c), and sample WVC density (b) and (d)

315   Figure 5 are plot from the testing set, where the horizontal and vertical axis are wind speed of ASCAT and that of observational and corrected OSCAT-2 speed in m/s for (a), (b) and (c), (d) respectively. Moreover, in (a) and (c) depicted in the colour bar ~~dimension~~, as functions of the horizontal and vertical speeds, are the average values of differences of speed from the vertical minus horizontal axis in corresponding bins. In (b) and (d) the colour represents WVC density in a bin. In (a), it can be observed that deviations from the C-band accepted collocations are due to rains vary with the reference wind speeds in

320   a similar linear way while for each wind speed there are multiple differences induced by rain. This is consistent with the quasi-linear relationship between J_{oss} and rain rates in Figure 2, and explains that such second-order (speed difference v.s. speed) relations involving multiple parameters (rain, wind and wind-rain correlations) cannot be corrected by ~~more~~ simple linear methods. While in (b), the corresponding density of samples indicates non-uniform characteristics of the distribution of the differences for each reference speed (horizontal axis), implying skewed error distributions. At the same time, in (c) and (d), it

325   can be seen that by SVM corrections, most of the differences are corrected, while (d) shows more evenly distributed difference patterns for the moderate wind speeds, where rain contamination effects appear better resolved, implying more uniform and normal difference values. This goes along with the distribution of corrected OSCAT winds slightly skewed away from the diagonal, this may due to the lack of samples in higher wind speeds.

4.2 Spatial consistency of corrected winds

330   In this section, to obtain a spatial view of the results, figures of the collocated data on a randomly selected date (22nd, May 2017) are provided in Figure 6, where (a) shows the wind speed of OSCAT-2 in both QC-I and QC-II collocations, and that of the rest FAE set. The same set is displayed in (b), but where the FAE OSCAT-2 wind speeds are from the SVM corrections,. In (c) the regressed wind is replaced by the ASCAT accepted winds. Furthermore, data in Figure 4 is without ~~GMP~~GPM collocations, and the SVM winds are retrieved directly from the model established in section 3.2.

335

(a)        (b)

(c)        (d)

340      (e)        (f)

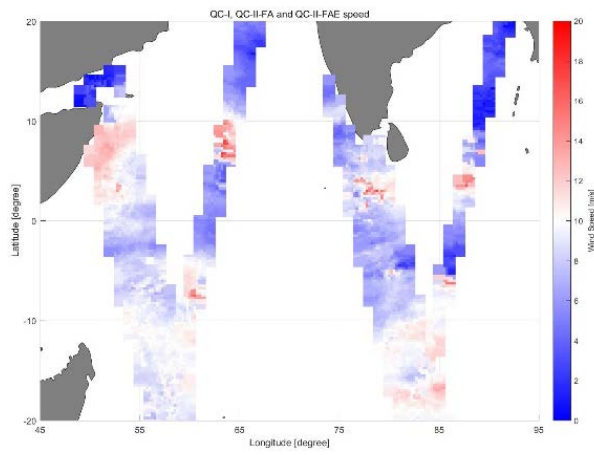
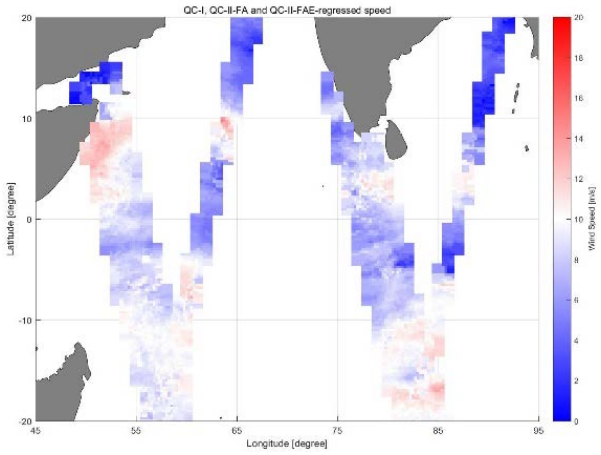
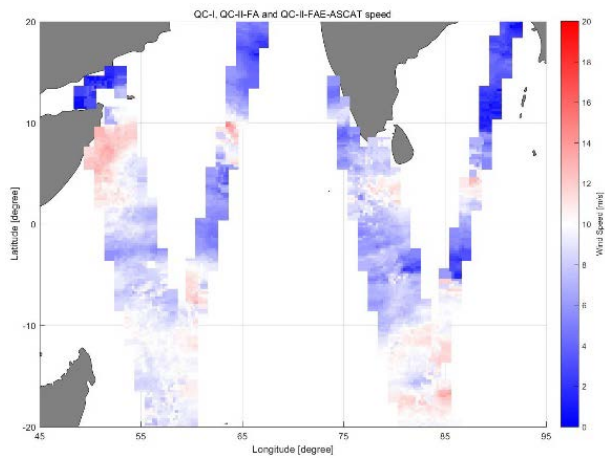
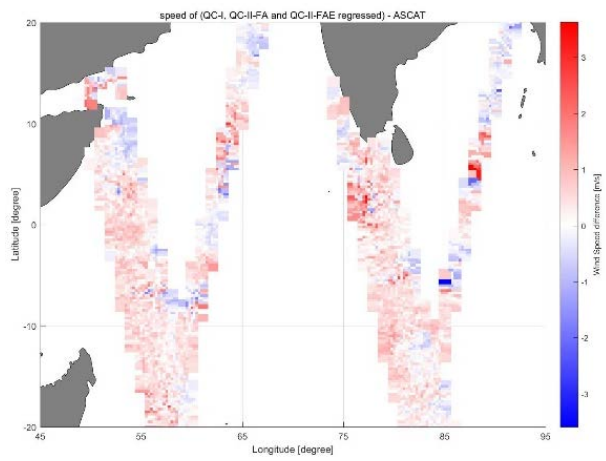
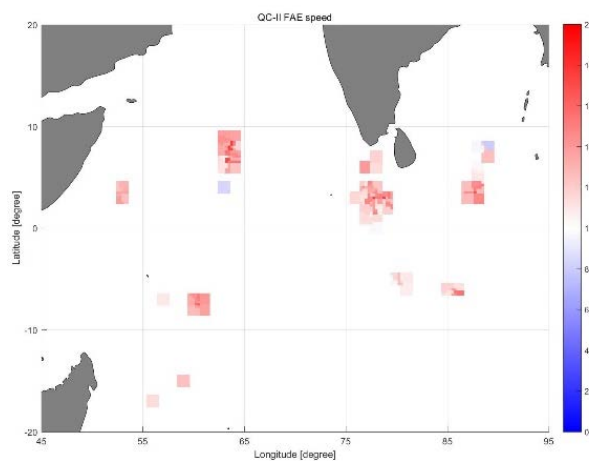
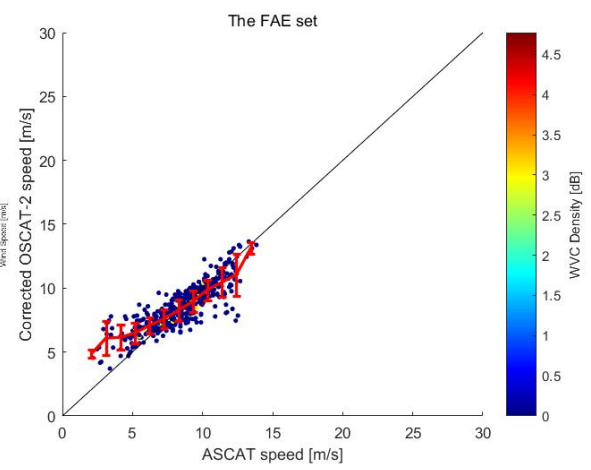



Figure 6: OSCAT-2 speed (m/s, in colour bars) for QC-I collocation set FA and FAE in the QCII set (a), and that of the QC-I, QC-II FA set when the FAE in QC-II are replaced SVM regressed speeds (b), then the FAE wind speeds are substituted by collocated ASCAT-A and ASCAT-B speeds (c). (d) is the differences of speeds in (c) with their corresponding ASCAT speed, and (e) indicate the FAE location while (f) shows the statistics of the corrected wind with ASCAT wind

345 In Figure 6, the abscissas are longitudes, while the ordinate represents latitudes, and both are in degree. Then the colour bars indicate wind speeds in m/s. Where the ascending and descending tracks are displayed together, with latter observations obtained replacing the former ones. It can be observed that the colour red in (a) is suppressed in (b), while (b) is also in better consistency with (c) than (a) is. This can be directly observed from (d), with the corrected wind locations from (e). (f) shows a generally accepted correction in this region with speed higher than 12 m/s overestimated. Similar trends can also be noted in

350 regions becoming much bluer, especially, in cases that can be found near the red regions. *Nota bene*, the higher wind regions with speed larger than 15 m/s are with fewer samples while also limited by the FA rule limiting J_{oss} to -1.33m/s, above which, the wind-rain tangling at higher speed cannot be well resolved. Moreover, a region with no GPM GMP collocation, thus not

(a)   (b)

(c)   (d)

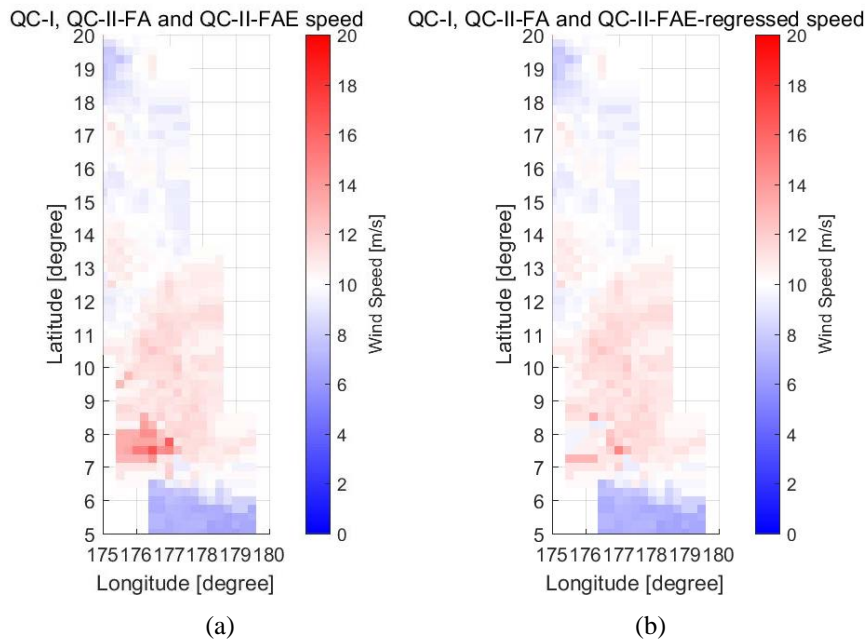
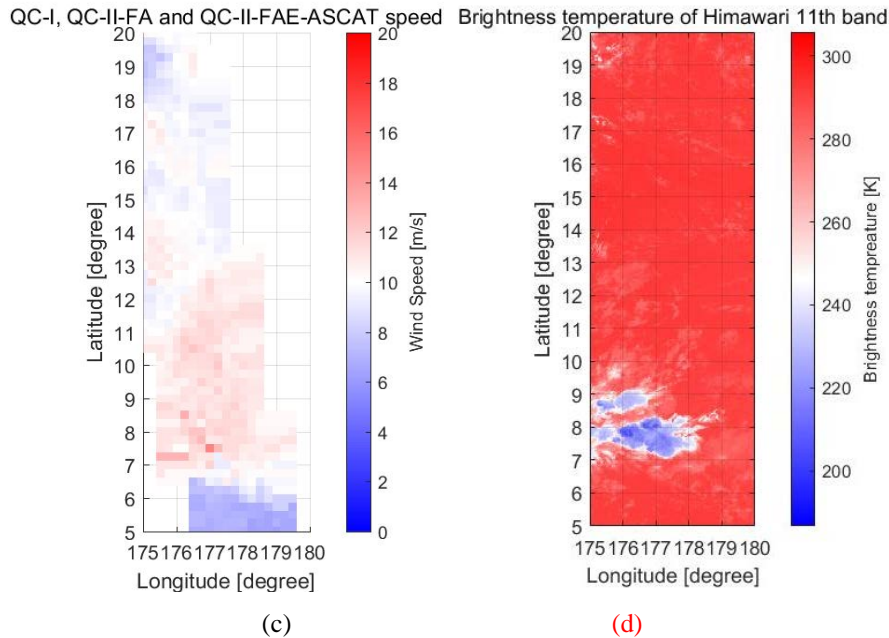



Figure 7: Wind speed of the QC-I, QC-II FA and QC-II FAE (a), the FAE set replaced by the SVM regressed speed (b) and by speeds from their ASCAT collocations (c), with the synchronous MIR (e) images from Himawari-8, where the green rectangle indicates the region in (a), (b) and (c).

360 involved in training procedures, is selected from the data set generating Figure 7, and is shown in Figure 8 as a case to validate the SVM regression method proposed. Wind speeds from the collocation set in QC-I, QC-II FA and QC-II FAE OSCAT-2 speeds are shown (a), that of QC-II FAE substituted by the SVM regressed speed for rain (cloud) correction (b) and that from the ASCAT collocations in C-band (c). There are 674 WVCs in Figure 7, with 13 FAE, and the observation time ranges from 09:19 to 09:24. Furthermore, the simultaneous image obtained around 09:20 is applied as references from the band 11 of
365 Himawari-8 satellite at medium infrared (MR, wavelength of 8.6μm) from Japan Aerospace Exploration Agency (JAXA).

  In Figure 7, the FAE set is distributed in the lower half in (a), where the colour is darker in red with lighter in white implying the existence of a wind front. After the correction, a more consistent set of wind speeds north of the front is obtained. In addition, rain clouds can be seen from (d) between 7-9 degrees N, with blue regions representing lower brightness temperatures (BT) and high probability of rain, where rain correction effects can be observed as well considering (a-c). This further confirms
370 the necessity of inclusion of 2DVAR J_{oss} for wind correction in case of rain. Although slightly overcorrected wind speeds occur in (b) around about 8 degrees N, it can be observed that (b) and (c) are more similar than (a) and (c), demonstrating the consistency between the SVM-regressed OSCAT and accepted ASCAT wind speeds. This can be further observed from WVCs between 9 and 10 degrees N 175 and 176 degree E, where (d) shows somewhat elevated BT of clouds, illustrating the effectiveness of the method proposed for such regions. More detailed statistics are shown in Figure 8.

(a)   (b)

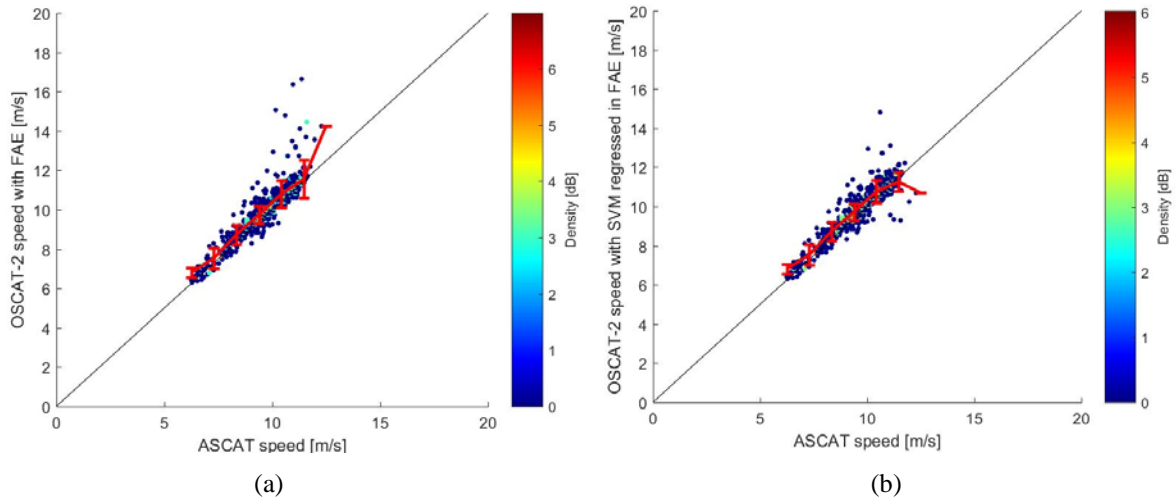



Figure 8: The FAE wind speed (a) and the corrected ones (b) against ASCAT wind speed in the data set of Figure 6.

It can be seen from Figure 8 that higher wind due to rain is suppressed by the method proposed, while for higher wind speed around 12 m/s, the SVM-regressed winds become somewhat less consistent with ASCAT *truth*, as discussed in the former section. The effectiveness of the SVM-regressed winds are further confirmed with the data in Figure 8, as they have not been applied in the derivation of the SVM.

5. Discussion

Air-sea interaction in the vicinity of rain is complex and difficult to observe. In this research, the effect of rain in Ku-band wind scatterometery is explored for correction of retrieved wind under rainy conditions. The method employed is as follows: on the basis of the analysis of signatures induced by rain from parameters obtained during wind retrieval from scatterometers, rain effects are corrected as a function of these signatures. Specifically, for quantifying the heterogeneity induced by rain and its effect on the wind speed, the quality indicator MLE and J_{oss} are analyzed, with reference to the low-pass filtered 2DVAR winds and collocated ASCAT winds (Xu et al., 2020a). Accepted C-band ASCAT winds (Vogelzang et al., 2011) are used as reference to identify the rain effects and form the basis of a correction after establishing a SVM. Results show that the correction is adequate, especially at speeds with abundant information ~~for~~ in the Ku-band to segregate wind and rain (under 12m/s). The spatial consistency of the corrected winds with the ASCAT observational winds is identified as more similar than to the 2DVAR winds. Subsequently, a case is provided with comparison to MIR images to check for rain occurrence. This confirms that the SVM method proposed is effective. Hereafter, rain information extraction from scatterometers is established. Following which, further analysis and discussion on the remaining uncertainties are given, with a view to improve in our future work.

5.1 SVM for rain identification and regression

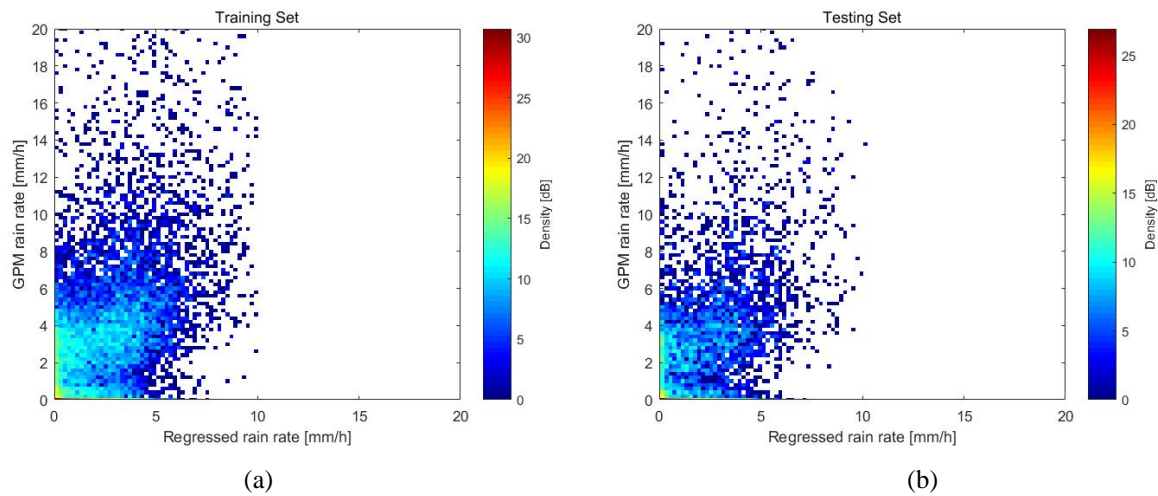
For a view of uncertainties unresolved with wind-rain tangling in Ku-band wind scatterometry, SVMs in the same input for rain identification and regression in the following Table:

400   Table 3 Inputs and outputs for the SVM of rain classification and regression:

Inputs	Output		Values of output for training	
	Rain Classification SVM	Rain Regression SVM	Rain Classification SVM	Rain Regression SVM
OSCAT-2 MLE in dB space	Rain or no-rain class	Rain rates [mm/h]	GPM rain rate, 0 mm/h in no-rain WVC	GPM rain rates [mm/h]
α from OSCAT-2 WVC				
OSCAT-2 2DVAR speed				
OSCAT-2 observational speed				

The data set is the same as for the wind correction, while the training target changed to GPM rain. The classification accuracies for both the training and testing set of rain identification SVM are the same and 72%. The results for rain regression are shown in the following figure, where the correlation coefficient of the SVM-regressed and GPM rain rates for the training

405   set and the testing set are both 0.47. Little skill for rain rate appears below 5 mm/hr, while GPM produces more extreme rain rates > 10 mm/hr. The corresponding scatter plots of the regressed rain rates in the training set and testing set are depicted in Figure 9.

(a)   (b)

410   **Figure 9: SVM regressed rains for training set (a) and validation set not involved in training (b).**

(a)   (b)

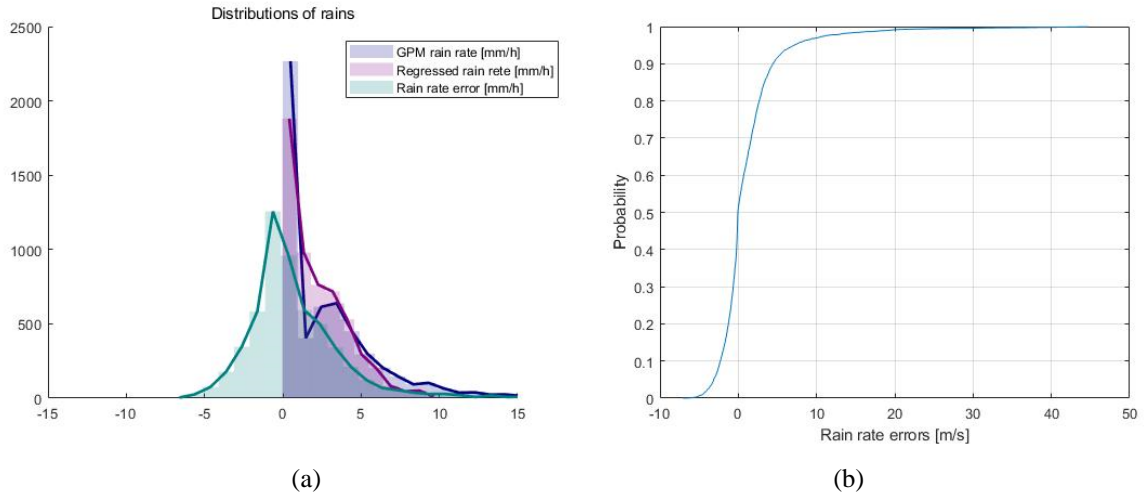



Figure 10: Distribution of GPM and the SVM regressed rains, with that of the error (a), the corresponding CDF of the error (b)

From visualization of the classification results (details not shown), non-rainy WVCs are less often incorrectly classified than rainy WVCs. Where larger 2DVAR speed are well-crowded and can be better discriminated in *MLE*, J_{oss} and α to the correct class, while this is more difficult for smaller 2DVAR speed WVCs. Light rain clouds have small effects to the wind observations. Correspondingly, Figure 10 (a) shows the distribution of rain rates from GPM (blue), SVM regression (purple) and that of the error defined as the GPM rain rate minus the regressed values (green). The corresponding CDF of error is shown in (b). In addition to Figure 9, Figure 10 (a) shows in detail that SVM regressed rain fails in capturing the non-convex feature in lower rain rate, and in prediction of higher rains. This may due to the L2 distance norm applied and lack of information as well as samples. For GPM rain above 10 mm/h, OSCAT rain rates are rather randomly distributed and presumably lack skill. However, from (b) in Figure 10, it can be observed that the error displays a feature of symmetry and steady increasing feature. And those within the range of [-2, 2] mm/h takes 34%, within [-5, 5] mm/h takes about 80%, consistent with the correlation coefficient value of 0.47. L1 distance, at the same time, including other sources of observation, with increasing number of samples may help improve the results. Xu et al. (2020b) find similar spread in rain products at the scatterometer spatial resolution, hence illustrating the applicability of the SVM rain product derived here.

5.2 Conclusions and further research

Rain features in wind scatterometry in the Ku-band can trigger QC rejections. These effects also provide opportunities to identify rain and perform wind corrections. The SVM method proposed acts well for medium and lower wind speeds, while the wind-rain tangling remains severe for higher wind speed. This can also be noted from the rain identification and regression SVMs in 5.1. While for lower speeds, the change of values of parameters considered may be caused by different wind-rain interaction with the ocean surface that alters the sea-state rather than only elevating the speed of wind due to rain cloud scattering that may be similar for C- and Ku-band and hence missed here.

On the other hand, from the rain features in *MLE* and *J$_{oss}$*, as well as the uncorrected speed, it can be seen that uncertainties can be introduced from the training parameters; the normalized *MLE* is designed to characterize errors that result in large deviations from the GMF for QC, but its accuracy depends on relative wind vector and azimuthal diversity of the NRCS views. While the 2DVAR speed is derived balancing errors in the observation space of a grid of WVCs and the NWP background, representing larger spatial scales, thus they can be considered as lower-bound estimates of the *true* values and uncertainties in the wind speeds can be different due to spatial heterogeneity. This may hamper the effectiveness of the rain screening ability of *J$_{oss}$*. In order to bound those uncertainties for better results in SVM, extra observations for rain (clouds) can help, while higher spatial resolution is obtained in the next generation of scatterometers for simultaneous ocean surface wind and current measurements, for example, Chelton et al. (2018) and Du et al. (2021). OSCAT-2 and ASCAT collocations provided a unique opportunity to study rain effects in Ku-band scatterometers. Rain effects are rather transient in nature, where the moist convection time scale is about 30 minutes. This implies updrafts, downdrafts and rain patterns in a WVC change very fast and rather strict collocation criteria would be needed to resolve rain effects well. With WindRad on FY3E a combined C- and Ku-band scatterometer has been launched on the 5th, July, 2021, which will provide parts of the swath with excellent azimuth diversity and both C- and Ku-band retrieval capability. Hence this mission will be useful to further elaborate this research.

Above all, the SVM can effectively represent the increasing effect of rain in elevating wind speeds as the true wind speed decreases showing the advantage of the ML method for such complex problems involving multiple interrelated variables. The method provides correction of deviations that are non-uniform and skew- to Gaussian-like features. This demonstrates the effectiveness of a ML method when used with representative parameters for addressing more complex problems. The corrected winds provide information previously lacking, while vital for nowcasting winds in the presence of moist convection and to improve initialization of NWP models in dynamic conditions. The rain regression in SVM indicates the potential of additional rain information observations for further exploration, as well as the promise of improved hybrid wind and rain optimization estimation methods based on ML using physically meaningful with parameters chosen considering their physical features for the problem at hand.

6 Code availability

None. But for the experiments, they can be reproduced upon request.

6 Appendix:

The mean values and standard deviations of differences

For the comparison of two collocated groups of data, one of which is set as reference group. Then figures and values are obtained by grouping the reference data (depicted as horizontal axis) and the other data set to compare (vertical axis) into *i* bins of the same sample number *j*. For the mean values of the reference data, *Ref$_i$*, (in tables they are put in the first column), there is corresponding *Ave$_i$* (in tables as the second column) and standard deviation values (third column) calculated for the

data to compare (in figures as the vertical axis). Specifically, the following equations describe the calculation of the mean value Ave_i and standard deviation of difference (SDD) Std_i:

$$Ave_i = \frac{\sum_{j=1}^{N_i} Obv\_Value_j}{N_i} \tag{A.1}$$

$$Std_i = \frac{1}{N_i}\sqrt{\sum_{j=1}^{N_i}\left(Obv\_Value_j - Ref_i\right)^2} \tag{A.2}$$

470   Where the values of the group to compare is Obv_Value.

7 Code availability

None. But for the experiments, they can be reproduced upon request.

8 Author contribution

475   Xingou Xu: Methodology, experiment, analysis, original draft writing.

Ad Stoffelen: Conceptualization, methodology, analysis, review and implementation of draft.

9 Competing interests

Ad Stoffelen is associate editor of AMT.

Acknowledgements

480   The authors would like to thank the Royal Netherlands Meteorology Institute (KNMI), the European Organization for the Exploitation of Meteorological Satellites (EUMET-SAT), the European Centre for Medium-Range Weather Forecasts (ECMWF), the National Aeronautics and Space Administration (NASA), and the Japan Aerospace Exploration Agency (JAXA) for the provision of the data products applied.

References

485   Bony, S., Stevens, B., Frierson, D., Jakob, C., Kageyama, M., Pincus, R., Shepherd, T. G., Sherwood, S. C., Siebesma, A. P., and Sobel, A. H.: Clouds, circulation and climate sensitivity, Nature Geoscience, 8, 261-268, https://doi.org/10.1038/ngeo2398, 2015.

Chang, C.-C. and Lin, C.-J.: LIBSVM: A library for support vector machines, ACM Trans. Intell. Syst. Technol., 2, Article 27, 10.1145/1961189.1961199, 2011.

490 Chelton, D. B., Schlax, M. G., Samelson, R. M., Farrar, J. T., Molemaker, M. J., McWilliams, J. C., and Gula, J.: Prospects for future satellite estimation of small-scale variability of ocean surface velocity and vorticity, Progress in Oceanography, 173, 256-350, 10.1016/j.pocean.2018.10.012, 2019.

Cornford, D., Nabney, I. T., and Bishop, C. M.: Neural network-based wind vector retrieval from satellite scatterometer data, Neural computing & applications, 8, 206-217, 10.1007/s005210050023, 1999.

495 Courtier, P., Thépaut, J. N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, Quarterly Journal of the Royal Meteorological Society, 120, 1367-1387, 10.1002/qj.49712051912, 1994.

Descombes, G., Auligné, T., Vandenberghe, F., Barker, D., and Barré, J.: Generalized background error covariance matrix model (GEN_BE v2. 0), Geoscientific Model Development, 8, 669-696, 10.5194/gmd-8-669-2015, 2015.

SCATSat-1—Satellite Missions: https://directory.eoportal.org/ web/eoportal/satellite-missions/s/scatsat-1 last access, 2021

500 Draper, D. W. and Long, D. G.: Simultaneous wind and rain retrieval using SeaWinds data, IEEE Transactions on Geoscience and Remote Sensing, 42, 1411-1423, 10.1109/tgrs.2004.830169, 2004.

Du, Y., Dong, X., Jiang, X., Zhang, Y., Zhu, D., Sun, Q., Wang, Z., Niu, X., Chen, W., and Zhu, C.: Ocean Surface Current multiscale Observation Mission (OSCOM): Simultaneous measurement of ocean surface current, vector wind, and temperature, Progress in Oceanography, 102531, https://doi.org/10.1016/j.pocean.2021.102531, 2021.

505 Gill, A. E.: Atmosphere-Ocean Dynamic, Int. Geophys. Ser, 1982.

King, G. P., Portabella, M., Lin, W., and Stoffelen, A.: Correlating extremes in wind and stress divergence with extremes in rain over the Tropical Atlantic, 2017.

Kumar, A., Ramsankaran, R., Brocca, L., and Muñoz-Arriola, F.: A simple machine learning approach to model real-time streamflow using satellite inputs: Demonstration in a data scarce catchment, Journal of Hydrology, 595, 126046, 510 10.1016/j.jhydrol.2021.126046, 2021.

Li, L., Im, E., Connor, L. N., and Chang, P. S.: Retrieving ocean surface wind speed from the TRMM precipitation radar measurements, IEEE Transactions on Geoscience and Remote Sensing, 42, 1271-1282, 10.1109/TGRS.2004.828924, 2004.

Lin, W. and Portabella, M.: Toward an improved wind quality control for RapidScat, IEEE Transactions on Geoscience and Remote Sensing, 55, 3922-3930, 10.1109/TGRS.2017.2683720, 2017.

515 Linwood Jones, W. B., P. G.; Boggs, D. M.;Bracalente, E. M.; Brown, R. A. ;Dome, G.; Ernst,J. A.; Halberstam, I. M.; Overland, J. E.; Peteherych, S.; Pierson, W. J.; Wents, F. J.; Woiceshyn, P. M.;Wurtele, M. G.: Seasat scanning multichannel microwave radiometer: results of the Gulf of Alaska Workshop, Science, 204, 817, 10.1126/science.204.4400.1415, 1979.

Liu, Chian-Yi, Putu Aryastana, Gin-Rong Liu and Wan-Ru Huang, Assessment of satellite precipitation product estimates over Bali Island, Atmospheric Research 244, 105032, https://doi.org/10.1016/j.atmosres.2020.105032, 2020.

520   Majumdar, S. J., Sun, J., Golding, B., Joe, P., Dudhia, J., Caumont, O., Chandra Gouda, K., Steinle, P., Vincendon, B., and Wang, J.: Multiscale Forecasting of High-Impact Weather: Current Status and Future Challenges, Bulletin of the American Meteorological Society, 102, E635-E659, https://doi.org/10.1175/BAMS-D-20-0111.1, 2021.

Otemachi, C.-k.: Himawari-8/9 Himawari Standard Data User's Guide, Japan Meteorological Agency, Tokoy, Japan, 2015.

Owen, M. P. and Long, D. G.: M-ary Bayes estimator selection for QuikSCAT simultaneous wind and rain retrieval, IEEE
525   transactions on geoscience and remote sensing, 49, 4431-4444, 10.1109/TGRS.2011.2143721, 2011.

Parrish, D. F. and Derber, J. C.: The National Meteorological Center's spectral statistical-interpolation analysis system, Monthly Weather Review, 120, 1747-1763, 10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2, 1992.

Portabella, M.: Wind field retrieval from satellite radar systems, Astron. Meteorol. Dept., Universitat de Barcelona, Domplein, The Netherlands, 2002.

530   Portabella, M. and Stoffelen, A.: Characterization of residual information for SeaWinds quality control, IEEE transactions on geoscience and remote sensing, 40, 2747-2759, 10.1109/TGRS.2002.807750, 2002.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., and Carvalhais, N.: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195-204, 10.1038/s41586-019-0912-1, 2019.

Smola, A. and Schölkopf, B.: A Tutorial on Support Vector Regression Royal Holloway College, London, 1998.

535   Stiles, B. W. and Dunbar, R. S.: A neural network technique for improving the accuracy of scatterometer winds in rainy conditions, IEEE Transactions on Geoscience and Remote Sensing, 48, 3114-3122, 10.1109/TGRS.2010.2049362, 2010.

Stoffelen, A. C. M.: Scatterometry, Utrecht University, 1996.

Stoffelen, A. and Anderson, D.: Scatterometer data interpretation: Measurement space and inversion, Journal of atmospheric and oceanic technology, 14, 1298-1313, 10.1175/1520-0426(1997)0142.0.CO;2, 1997.

540   Stoffelen, A. and Vogelzang, J.: Wind bias correction guide, EUMETSAT, Darmstadt, Germany, 2018.

Stoffelen, A., Kumar, R., Zou, J., Karaev, V., Chang, P. S., and Rodriguez, E.: Ocean surface vector wind observations, in: Remote Sensing of the Asian Seas, Springer, 429-447, 2019.

Thiria, S., Mejia, C., Badran, F., and Crepon, M.: A neural network approach for modeling nonlinear transfer functions: Application for wind retrieval from spaceborne scatterometer data, Journal of Geophysical Research: Oceans, 98, 22827-
545   22841, 10.1029/93JC01815, 1993.

Vapnik, V.: Statistical learning theory 624, Wiley, New York, 2 pp.1998.

Vogelzang, J.: Two dimensional variational ambiguity removal (2DVAR), 2007.

Vogelzang, J, Stoffelen, A. Improvements in Ku-band scatterometer wind ambiguity removal using ASCAT-based empirical background error correlations. Q J R Meteorol Soc. 2018; 144: 2245– 2259. https://doi.org/10.1002/qj.3349.

550   Vogelzang, J. and Stoffelen, A.: NWP model error structure functions obtained from scatterometer winds, IEEE transactions on geoscience and remote sensing, 50, 2525-2533, 10.1109/TGRS.2011.2168407, 2011.

Vogelzang, J., Stoffelen, A., Verhoef, A., and Figa-Saldaña, J.: On the quality of high-resolution scatterometer winds, Journal of Geophysical Research: Oceans, 116, 10.1029/2010JC006640, 2011.

Wolters, E., Van Den Hurk, B., and Roebeling, R.: Evaluation of rainfall retrievals from SEVIRI reflectances over West Africa using TRMM-PR and CMORPH, Hydrology and Earth System Sciences, 15, 437-451, 10.5194/hess-15-437-2011, 2011.

Xu, X. and Stoffelen, A.: Improved rain screening for ku-band wind scatterometry, IEEE Transactions on Geoscience and Remote sensing, 58, 2494-2503, 10.1109/TGRS.2019.2951726, 2020a.

Xu, X., Stoffelen, A., and Meirink, J. F.: Comparison of ocean surface rain rates from the global precipitation mission and the Meteosat second-generation satellite for wind scatterometer quality control, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 2173-2182, 10.1109/JSTARS.2020.2995178, 2020b.

Xu, X., Stoffelen, A., Lin, W., and Dong, X.: Rain False-Alarm-Rate Reduction for CSCAT, IEEE Geoscience and Remote Sensing Letters, 10.1109/LGRS.2020.3039622, 2020c.

Xu, X. Stoffelen, A.: A Further evaluation of the quality indicator Joss for Ku-band wind scatterometry in tropical regions, IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (accepted),2021.

555

560

565