Set-up of PenWP for SeaWinds-derived coastal winds

OSI_VS22_02

Authors: **G. Grieco**[1], M. Portabella[2], J. Vogelzang[3], A. Verhoef[3], A. Stoffelen[3]

[1] Institute of Marine Sciences (ISMAR-CNR)

[2] Barcelona Expert Center (BEC ICM-CSIC)

[3] Koninklijk Nederlands Meteorologisch Instituut (KNMI)

Visiting Scientist Activity

Technical Report

Date: 7th November, 2023

Contents

Abstract **3**

1 Introduction **4**

2 Software improvements **6**
2.1 Implementation of the software for the computation of f in C 6
2.2 Update of `seawinds_l1b_bufr` 8
2.3 Processor handling 9

3 Methodology **11**
3.1 Theoretical background 11
3.2 Noise regularization 11
3.3 WVC Buffer area 16
3.4 Computation of f 17

4 Dataset **23**

5 Results and discussion **24**
5.1 Software performances 24
5.2 Noise regularization 26
5.3 Wind retrievals 27

6 Conclusions and future work **54**

Abstract

This study presents the implementation of a procedure to mitigate land contamination that affects the acquisition of the Seawinds pencil beam scatterometer. This procedure is named noise regularization because, after it is applied, the level of noise of the corrected acquisitions is constant. The effects of noise regularization are evaluated both at Normalized Radar Cross Section (σ_0), and wind retrieval levels. The results show that this methodology is effective in mitigating land contamination and that coastal sampling increases by 30% in the coastal band within 30 km to the coast, compared to the state-of-the-art retrieval procedure implemented at OSI-SAF. This figure increases to 300% in the first 10 km. The results also show that there is room for improvement if wind retrieval quality control is refined. The study also shows the implementation of a more efficient software for the calculation of the Land Contribution Ratio (f) index.

1 Introduction

Wind plays a major role in several coastal phenomena, such as the diffusion of momentum, heat, carbon, nutrients, and pollutants. All these phenomena affect life along the coasts, including humans. For this reason and others not listed here, accurate coastal winds have garnered the scientific community's attention. Nowadays, scatterometers represent the best and most mature technology to retrieve winds from space, capable of accuracies better than 1 ms^{-1} in speed and $\pm 20^o$ in direction, with spatial resolutions rarely better than 20 km. Due to the "coarse" spatial resolution, scatterometer-derived coastal winds may be heavily affected by land contamination [1]. In fact, land/ocean Normalized Radar Cross Section (σ_0) contrasts can lead to biased retrievals. Some efforts have been made to tackle this problem, but there is still much to do in this direction.

The authors of [1] show how to quantify land contamination by introducing the Land Contribution Ratio index (f), which represents the weighted average of a high resolution Land Sea Mask (LSM), the weights being the values of the Spatial Response Function (SRF) of the instrument. Then, the excessively contaminated σ_0s (typically $f > 2\%$) are discarded. The results show that the retrievals improve, even if the sampling along the coasts is highly reduced. The authors of [1] apply this procedure to the Seawinds scatterometer, which flew onboard the US polar orbiting satellite platform Quick Scatterometer (QuikSCAT) from 1999 to 2009. A similar approach has been followed also by the authors of [2, 3] for ASCAT.

Recently, the authors of [4] showed how to remove land contamination from Seawinds σ_0s, with the notable result of increasing the sampling towards the coast. They characterize the monthly climatology of Seawinds land σ_0 in coastal areas using the time series that spans the entire mission. This contribution is weighted by f and then subtracted from the measured σ_0. This product, namely QuikSCAT 4.1, is freely available from the NASA Physical Oceanography Distributed Active Archive Center (PO.DAAC) website [5]. The limit of this methodology is that the diurnal variability of σ_0, mainly due to variations in soil moisture, is not filtered out. As a consequence, wind retrievals may be biased. The authors of [6] show an alternative empirical method, applied to ASCAT measurements. They assumed that the dependence of σ_0 is linear with f and that the noise that affects σ_0 has negligible variation with σ_0. Then, after the land σ_0 is estimated, it is subtracted from the measured σ_0.

The same approach has also been applied to Seawinds acquisitions in [7]. In [7], the authors show that this approach is not successful for Seawinds because σ_0 noise is far from being constant. In fact, the results show that

the specific slice σ_0 noise (K_p) is very high in low wind regimes and that peripheral slices with respect to (w.r.t.) the footprint centroid are much noisier than those closer [8]. Therefore, the application of the empirical σ_0 correction methodology shown in [6] can lead to an excessive unrealistic number of negative σ_0 when applied to Seawinds.
In this study, a new correction methodology based on the so-called "noise regularization" is presented, together with the wind retrievals obtained with the corrected σ_0s. Furthermore, the implementation of the software for the computation of f in `C` is described, together with an analysis of its performance.
This document is organized as follows. Section 2 shows the software updates, both those concerning the computation of f and those of `seawinds_l1b_bufr`, used to produce the intermediate files to feed the processor `penwp`; Section 3 describes the noise regularization in detail. Furthermore, it describes how to solve the sampling problems related to the Seawinds acquisition geometry, already revealed in [7], and the sensitivity analysis of f w.r.t. to the method of computation (analytical or parameterized), and to some SRF and LSM parameters (SRF size, SRF and LSM grid spacing); Section 4 describes the dataset used; Section 5 show the results related to the software performance and to the effects of noise regularization on the retrievals; Section 6 reports on the preliminary conclusions and accounts for the future plan.

(a) Analytical (b) LUT-based

Figure 1: Flow charts of the software that computes f.

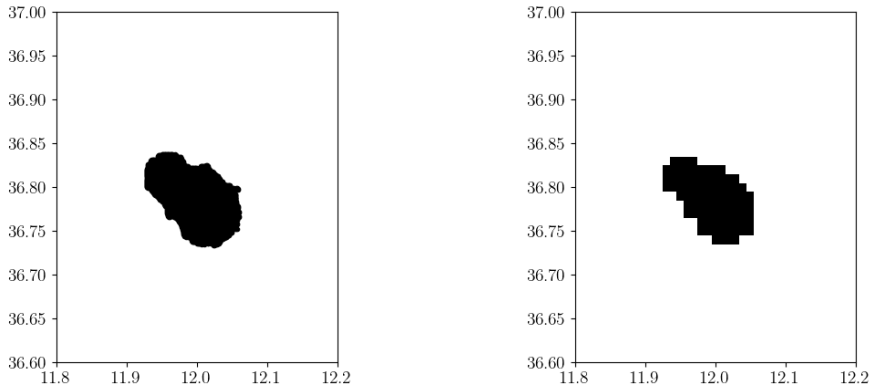
2 Software improvements

In this section, various software aspects related to the entire retrieval procedure are dealt with. Section 2.1 shows the implementation of the software for the computation of f; Section 2.2 shows the modifications implemented in `seawinds_l1b_bufr`, while Section 2.3 shows how all the software parts are integrated and managed.

2.1 Implementation of the software for the computation of f in C

In [9] two different methods for the computation of f have been presented and compared: one based on the analytical SRF model described in [10] and one based on a look-up table (LUT) of pre-computed parameterized SRFs. Both software were implemented in `python` with limited computational efficiency. One of the objectives of the EUMETSAT Visiting Scientist Activity (VSA) 22 02 was the implementation of both software in a more efficient programming language such as `FORTRAN` or `C`. Both have been implemented in `C`, as schematically indicated by the flow charts in Figure 1.

Figure 1a (1b) shows the flow chart of the implementation that is based on the analytical (LUT-based) SRF formulation. As can be seen, the two implementations are very similar, the only difference being the use of SRF LUT in the LUT-based implementation. Also, the software architecture is very similar, and many parts of the code are shared, even if the two implementations are completely independent. Module "SRF LUT" of Figure 1b consists of a LUT of pre-computed SRFs, which are parameterized according to the polarization of the signal, the orbit time and the antenna azimuth angle. SRF LUT was provided by Prof. Dave Long of Brigham Young University (BYU) together with the necessary software to query it.



(a) $\Delta x = 0.001^o$ (b) $\Delta x = 0.01^o$

Figure 2: LSM around the island of Pantelleria, in the Mediterranean basin. Black points represent land.

LSM	SPACING (^{o})	SIZE	# OF FILES
HR	0.001	≈60 Gb	64800
LR	0.01	618 Mb	1

Table 1: HR *vs* LR LSM

In addition to LUT, the software needs other ancillary information, namely, the distance of the slice centroid to the coastline and LSM. The distance of the slice centroid to the coastline is calculated using a `FOTRAN 90` program called `qs_slice_distance2coast.F90` (module `distance_to_coast` in the figures), which is based on a tool provided with the library `genscat` of `penwp`. This module queries an LSM database with a grid spacing of 1 km. This information is necessary to reduce the number of unuseful computations of f. In fact, if the slice centroid is farther than 80 km from the coast, f can be reasonably approximated to 0 (1) if it is offshore (inland). LSM is derived from the Global Self-consistent Hierarchical High-resolution Geography (GSHHG) database and was calculated with two grid spacings, namely 0.01^o and 0.001^o.

Figure 2a (2b) shows the high resolution (low resolution) LSM in the surroundings of the small island of Pantelleria, in the central Mediterranean basin. It gives an idea of the details resolved by the two resolutions. The effects of grid spacing on f are discussed in Section 5.

Table 1 resumes the characteristics of the two LSM databases: grid spacing, size and number of files. It is apparent that increasing the resolution by

HR	LR
High spatial resolution	Low spatial resolution
High number of FP operations	Low number of FP operations
High number of I/O operations	Low number of I/O operations
High memory load	Low memory load

Table 2: **Pros** and **Cons** of HR *vs* LR LSM

a factor 10 has a non-negligible impact on the management of the code. In fact, the size of the database increases by a factor ≈100, while the number of files increases from 1 to 64800. In fact, the HR LSM database is organized in "tiles" of 1^ox1^o, and only the necessary tiles are loaded. This fosters a more efficient use of RAM but increases the number of I/O operations. Table 2 resumes the pros and cons of using LR or HR versions of LSM. It is trivial that LR LSM improves code speed. The question is to assess whether it is worth it or not in terms of f accuracy. This aspect is dealt with in Section 5.

f is computed using the following formula [7]:

$$f = \frac{\sum_{ij} SRF_{ij} LSM_{ij}}{\sum_{ij} SRF_{ij}} \tag{2.1}$$

where $i \in [1, N]$, $j \in [1, M]$, N, M being the dimensions (in number of points) of SRF along range and cross-range directions, respectively. SRF has an elliptical symmetry due to range filtering [10]. SRF grid spacing is fixed for the LUT-based version of the software: $N = 251$ and $M = 81$, corresponding to a grid spacing of ≈0.81±0.04 x 1.97±0.12 km^2. In this way, the range direction, which is characterized by a more rapid decline in SRF values, is sampled with a finer grid. Instead, in the analytical version of the software, SRF grid spacing can be set-up by the user, even if there is not the possibility to discriminate between the two directions. Before SRF_{ij} is multiplied by LSM_{ij}, LSM is closest interpolated to the SRF grid point. In addition, in the analytical version of the software, the user can set the SRF size, which is recommended not to be less than 160 km, to keep errors in f less than 1e-1.

2.2 Update of `seawinds_l1b_bufr`

`seawinds_l1b_bufr` is a `FOTRAN 90` tool that was created to allow `penwp` to retrieve winds from Seawinds. This tool reads Full-Resolution (FR), also called Level-1 (L1) files and computes the integrated "views", which are then

stored in bufr files that are fed to `penwp`. This tool was modified to allow retrievals on a 12.5 km grid and to overcome an intrinsic sampling issue related to the Seawinds architecture. This problem was already pointed out in [7] and is described in Section 3.3 in detail.
The introduction of a buffer area surrounding the WVCs solved this problem. The size of the buffer area is hard-coded. The sensitivity analysis shown in Section 3.2 shows that a good option is 2.5 km.

2.3 Processor handling

The entire retrieval procedure is divided into two main stages: the first aims to provide all the necessary ancillary information needed by `seawinds_l1b_bufr`, while the second consists in running `seawinds_l1b_bufr` and `penwp`. The two stages are kept separated because handling is more flexible and efficient. The first stage consists of the following steps:

- Assessment of Seawinds slice σ_0 noise;
- Transformation of the original Seawinds hdf4 file (native FR format) into hdf5;
- Computation of the slice distance to the coastline;
- Computation of the WVC centroid coordinates;
- Computation of the WVC distance to the coastline;
- slice assignment (without any buffer area);
- Computation of f;
- Computation of the corrected slice σ_0s;

This stage is controlled by a unique `python` script that runs each step in the order listed above. This stage should be further improved and optimized. This is part of the future activities. In particular, all `python` scripts should be moved to `C` or `FORTRAN`. Note that the distance from the coast (both at the slice and at the WVC level) is necessary to limit the burden of computation. In fact, f is computed only for slices within $(-80, 80)$ km from the coast (the sign - indicates inland), and the slice correction is done only for WVCs that are less than 100 km from the coast. Furthermore, note that the WVC grid computed at this stage is before any optimization after retrieval occurs. Therefore, it is a regular WVC with a spacing of 12.5 km. The equations used for this purpose were provided by Dr. Alex Fore from NASA JPL. Finally,

note that the slice assignment at this stage does not need any buffer area. This step is again needed to optimize the burden; therefore, each slice must be assigned to a unique WVC. At the end of the last step, a new Seawinds FR file is produced. This file is identical to the original one, but the field `slice_sigma0`. In the new file, it contains the corrected slice σ_0s.
The proper retrieval stage consists of the following steps:

- Transform the modified Seawinds hdf4 file into hdf5;
- Launch of `seawinds_l1b_bufr`;
- Launch of `penwp`;
- Conversion of bufr into netcdf;

3 Methodology

3.1 Theoretical background

Let's consider a random variable X with probability distribution function (pdf) f_X and cumulative distribution function (CDF) F_X. What is the transformation $g : \mathbb{R} \to \mathbb{R}$ that allows $Y = g(X)$ to have a CDF equal to F_Y? This is possible with a two-step procedure if we assume that both F_X and F_Y^{-1} are analytically known. In fact, it can be demonstrated that $u = F_X(x)$ is uniformly distributed, and $y = F_Y^{-1}(F_X(x))$. If F_X^{-1} and F_Y are also analytically known, one can go back and forth between the two domains of X and Y. The demonstration that u is uniformly distributed is intrinsic in the definition of CDF. In fact,

$$F_U(u) = P(U \leq u) = u \tag{3.1}$$

$$f_U(u) = \frac{dF_U(u)}{du} = 1 \tag{3.2}$$

where P stands for probability. Furthermore,

$$u = F_X(x) = P(U \leq u) = P(U \leq F_X(x)) \tag{3.3}$$

$$u = F_Y(y) = P(U \leq u) = P(Y \leq y) = P(F_Y^{-1}(U) \leq y) \tag{3.4}$$

By equalizing equation 3.3 to 3.4 we have

$$y = F_Y^{-1}(F_X(x)) \tag{3.5}$$

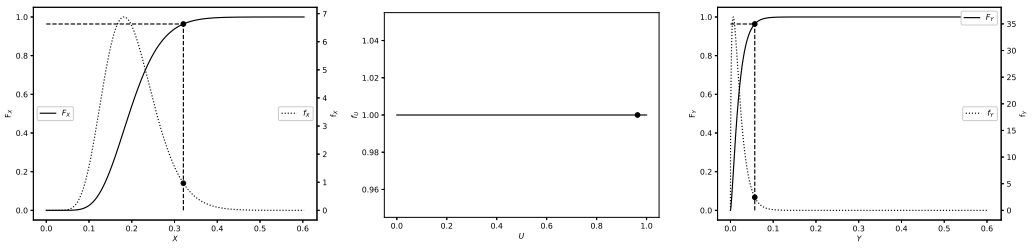
The inverse is also true:

$$x = F_X^{-1}(F_Y(y)) \tag{3.6}$$

Figure 3 may help to understand the problem from a graphycal point of view. The solid (dotted) curve in Figure 3a represents $F(X)$ ($f(X)$). The value of the F (f) associated to a given realization x of the random variable X is indicated with a black full circle. The distribution of the random variable $U = F_X(X)$ is shown in Figure 3b, with the corresponding value $u = F_X(x)$ with the same marker. Finally, the solid (dotted) curve in Figure 3c shows $F(Y)$ ($f(Y)$), together with the value $y = F_Y^{-1}(F_X(x))$ with the same marker.

3.2 Noise regularization

Scatterometer acquisitions within several tens of kilometres from the coast may be contaminated by land. The degree of contamination also depends



(a) (b) (c)

Figure 3: pdf transformation. Figure 3a (3c): pdf is a $\chi^2_{Norm}(k_1, \mu_1)$ ($\chi^2_{Norm}(k_2, \mu_2)$), with $k_1 = 22$ ($k_2 = 3$) and $\mu_1 = 0.2$ ($\mu_2 = 0.02$), which represents a typical distribution of contaminated (non contaminated) Seawinds slice σ_0 values. See Equation 3.8 for the meaning of k and μ.

on the orientation of the SRF w.r.t. to the coastline. Equation 3.7 is a model of the land contaminated σ_0. SEA and $LAND$ contributions to σ_0 are weighted by $1-f$ and f, respectively, and so are the noises attributed to them, namely $\epsilon^{SEA}(\sigma_0^{SEA})$ and $\epsilon^{LAND}(\sigma_0^{LAND})$, respectively. The value σ_0^{SEA} (σ_0^{LAND}) in round brackets emphasizes the fact that the noise depends on σ_0 [11]. Finally, noise contributions are under-bracketed to emphasize that they sum into an indistinguishable noise component. Equation 3.7 represents an underdetermined problem. In fact, we have one measurement (σ_0) and four unknowns (σ_0^{SEA}, σ_0^{LAND}, ϵ^{SEA} and ϵ^{LAND}).

$$\sigma_0 = (1-f)\sigma_0^{SEA} + f\sigma_0^{LAND} + \underbrace{[(1-f)\epsilon^{SEA}(\sigma_0^{SEA}) + f\epsilon^{LAND}(\sigma_0^{LAND})]}_{\epsilon} \quad (3.7)$$

Noise regularization consists in "projecting" contaminated σ_0s onto the distribution of non-contaminated σ_0s, whose expected value is σ_0^{SEA} and its standard deviation is $\sigma_{\epsilon^{SEA}}$. Therefore, the pdf transformation described in the previous section (3.1) is the core of the noise regularization. The basic assumptions of noise regularization are the following:

- Land contamination has a linear trend with f;
- σ_0^{SEA} and σ_0^{LAND} have local negligible variations;
- $F(X)$ and $F^{-1}(Y)$ are analytically known;

Under these assumptions, noise regularization consists of the following steps:

- A matrix of 5x5 WVCs at 12.5 km is centered around the target WVC, as depicted in Figure 4;

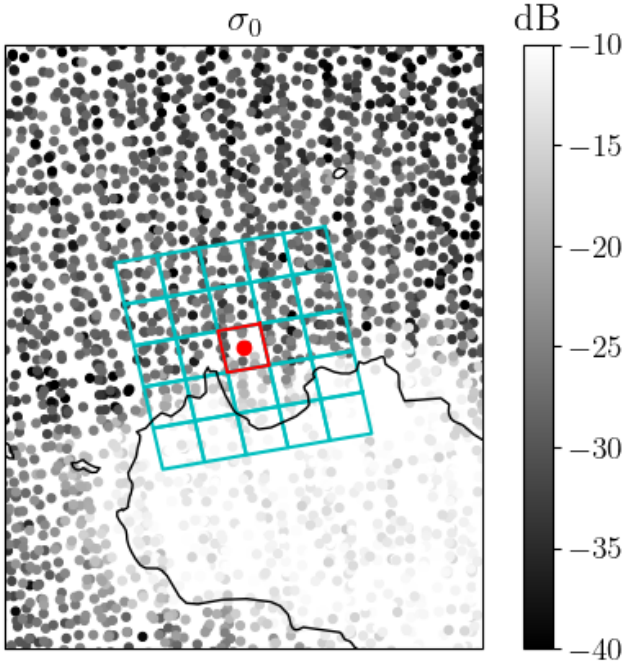


Figure 4: Cyan boxes: set of 24 WVCs surrounding the target WVC (red-framed). Full circles: slice σ_0 centroids. Colorbar indicates σ_0 level in logarithmic scale.

- All slice σ_0s in this matrix are used to regress σ_0^{SEA}, and the slope (a) with a least square method (see Figure 5);
- In case σ_0^{SEA} is negative, σ_0^{SEA} is set to the expected value of all slice σ_0s in the matrix with $f < 0.02$. In case this value is also negative, the WVC is flagged;
- σ_0^{SEA} and all $\sigma_0(f)$ values of the fitting curve are used to estimate ϵ^{SEA} and ϵ^{LAND} ($\epsilon(f)$). For this purpose, a LUT of pre-computed values of $\hat{K}_p$ is queried. The LUT of $\hat{K}_p$ is calculated using the methodology described in [8] for each day of the year. This LUT is a 4D table depending on the polarization (H-Pol and V-Pol), view (Aft and Fore), slice index (0-7), σ_0 level, which spans from a level equivalent to a wind speed of $\approx$1 ms^{-1} up to $\approx$50 ms^{-1};
- Contaminated slice σ_0s are "projected" onto the distribution of those non contaminated.

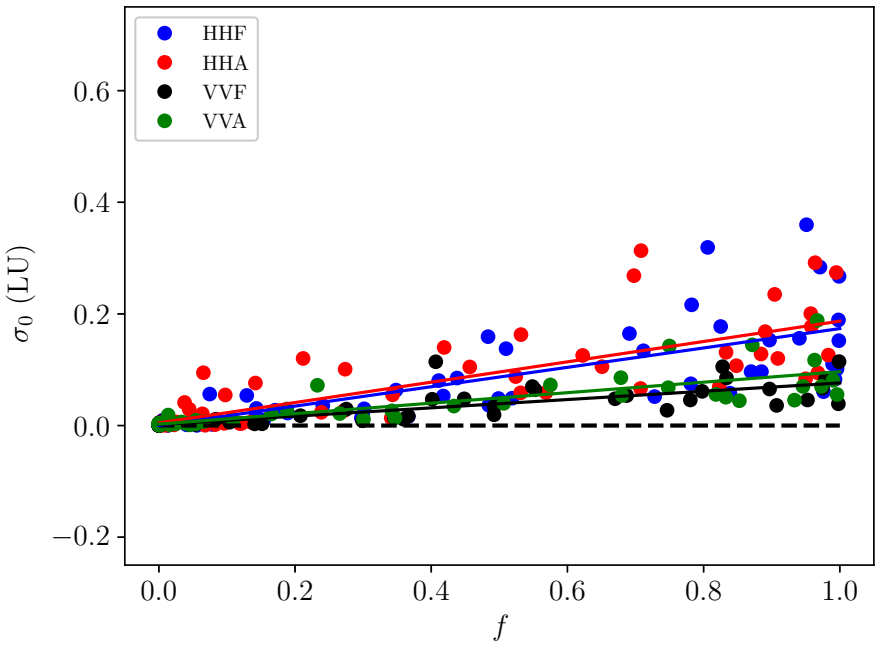


Figure 5: Scatter plot of slice σ_0 in linear units *vs* f. Slice σ_0s are those in the 5x5 WVC matrix in Figure 4 and are segregated according to the four flavors. HH (VV) stands for H-Pol (V-Pol), while A (F) stands for Aft (Fore). Regression curves follow the same color code as the markers.

From a physical point of view, σ_0 represents a normalized radar cross-section and, therefore, is positive definite. σ_0 is proportional to the power of the signal that reaches the receiver antenna, which is, in turn, equal to the sum of the squares of the real and imaginary parts of the signal voltage. Therefore, in the case of only one look, a noise-free σ_0 distribution is a χ^2 distribution with two degrees of freedom (dof), assuming that the real and imaginary parts of the signal are Gaussian distributed with identical standard deviation (equal to the amplitude of the voltage signal) and null mean [11]. In the case of a higher number of looks, σ_0 is distributed as normalized χ^2 (χ^2_{Norm}), as described in Equation 3.8

$$f_{\sigma_0} = \frac{1}{\alpha_{\sigma_0} 2^{\frac{k}{2}} \Gamma(\frac{k}{2})} \left(\frac{\sigma_0}{\alpha_{\sigma_0}}\right)^{\frac{k}{2}-1} \exp -\frac{\sigma_0}{2\alpha_{\sigma_0}} \tag{3.8}$$

$$F_{\sigma_0} = P(\frac{k}{2}, \frac{x}{2\alpha_x}) \tag{3.9}$$

$$\alpha_{\sigma_0} = \frac{\mu_{\sigma_0}}{k}$$

$$\sigma_{\sigma_0} = \mu_{\sigma_0}\sqrt{\frac{2}{k}}$$

$$k = \frac{2}{K_p^2}$$

where k stands for dof, μ stands for expected value, σ for standard deviation and $P(a, x)$ is the regularized lower incomplete gamma function, defined as

$$P(a, x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt \tag{3.10}$$

Once K_p and μ_{σ_0} are known, k and α_{σ_0} can be calculated, and so finally f_{σ_0} and F_{σ_0}. $P(a, x)$ has its analytical inverse and both are available in common math libraries such as the package `scipy.special` of `python`. Therefore, once σ_0^{SEA} and k^{SEA}, and $\sigma_0(f)$ and $k(f)$ are known, noise regularization is straightforward. Note that the detrending procedure described in [6] is a particular case of noise regularization, under the assumption that the distribution of σ_0 is Gaussian and the noise can be considered constant with varying σ_0 level. This assumption works well with ASCAT data because the equivalent number of looks of ASCAT is much higher than for Seawinds (K_p is much smaller). Noise regularization is a more general approach and can be used with all distributions with cross-dependent expected values and standard deviations. Finally, note that despite χ^2_{Norm} being adequate for

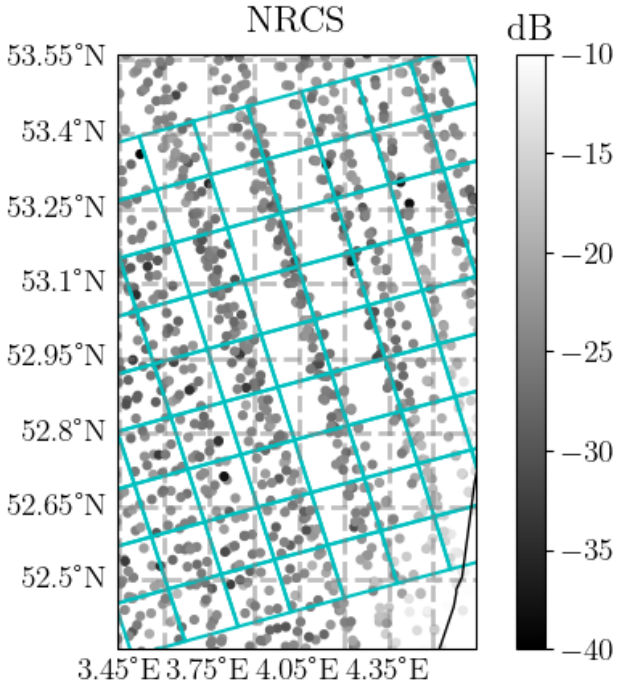


Figure 6: Sampling issues with grid spacing of 12.5 km offshore the Netherlands. Gray markers represent the slice centroids. Color represents the intensity of σ_0 in a logarithmic scale. The Cyan grid represents the grid of wind retrievals with spacing equal to 12.5 km before retrieval. Orbit ID is 40653.

Seawinds σ_0s ([8]), it cannot predict negative values. Therefore, noise regularization cannot deal with negative σ_0s. Fortunately, good-quality negative σ_0s are rather rare and this aspect does not seem an important limit to this methodology. The results shown in Section 5.2 will better clarify that.

3.3 WVC Buffer area

The Seawinds sampling issues at 12.5 km are well represented in Figure 6. This figure shows the slice centroids (grey markers) in an area offshore the Netherlands. The colorbar represents the intensity of σ_0 on a logarithmic scale, while the 12.5 km wind retrieval grid is depicted in cyan. In the central part of the picture, many cyan boxes are almost empty. In such cases, not all views are populated with valid raw acquisitions; therefore, some integrated σ_0 values are missing and the retrieval is not successful.

This aspect is further clarified by Figure 7. This figure represents an extract of a WVC matrix of 9 elements located in the Sicilian Channel, in the Mediterranean basin, where these sampling issues occur for the orbit ID 40653. Figures 7a to 7d represent the number of raw acquisitions available for each flavor. It is apparent that some WVCs have missing acquisitions (central column of Figure 7c and top left WVCs of Figure 7d). It comes out that the number of integrated views is equal to three for all these five WVCs and the retrieval is not possible (the flag `not_enough_good_sigma0_for_wind_retrieval` is raised and the retrievals are masked).

To overcome this problem, a buffer area surrounding each WVC is added, and all raw acquisitions belonging to this area receive multiple WVC assignments. Figure 8 shows the same matrix of 3x3 WVCs depicted in Figures 7; the original WVC grid is represented by solid black lines, while the new buffer area is colored in cyan. In this case, a buffer area of 2.5 km is represented on each side. Taking into account the target WVC with the cyan cross marker, all the slices assigned to it are represented by full-colored markers, where the color code represents the four flavors. It is clear that if the buffer area was not considered, no blue markers would be assigned to the central WVC with the consequent lack of an integrated view (HHA). Full markers receive multiple assignments if they are located in the buffer area.

Figure 11 shows how slices are assigned. If the target slice (red marker) lies in the inner core of the WVC (out of the buffer area), it is assigned only to that WVC (Figure 9a); If it is located in the lateral part of the buffer area, it is assigned to both contiguous WVCs (Figure 9b); Finally, if it is located in the corner, it is assigned to all four contiguous WVCs (Figure 9c).

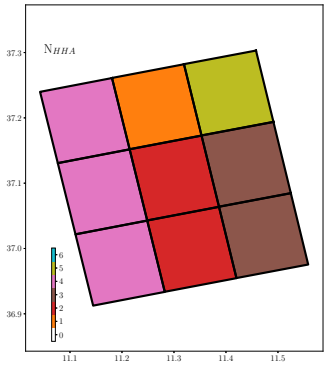
A sensitivity analysis with varying buffer area sizes was performed to select the optimal choice. This size was varied from 1.25 to 6.25 km with a step of 1.25 km. The results are shown in Section 5.

3.4 Computation of f

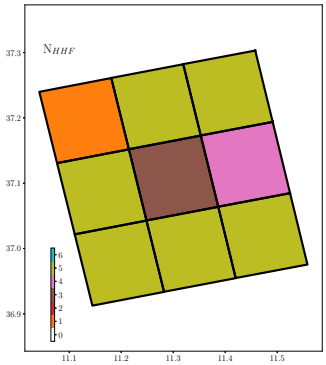
This section describes the experiment that has been set up to compare the softwares that compute f using the analytical and the LUT-based SRF models. It was designed to assess their performance in terms of both computational efficiency and accuracy. The target area is the island of Pantelleria, in the central Mediterranean basin. It is a small island (see Figure 2a), slightly oblong. f was computed with eight different orientations w.r.t. the coastline, as depicted in Figure 10.

Table 3 shows the configurations of the two softwares that were used to compute f.

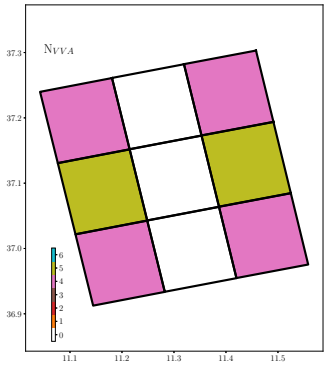
Note that the size of SRF is never lower than 160 km for the codes using



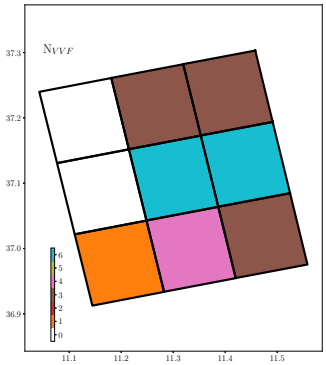
(a)



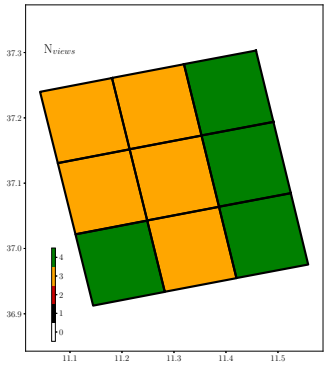
(b)



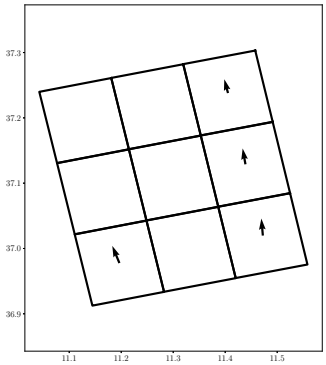
(c)



(d)



(e) Number of integrated views.



(f) Retrieved winds

Figure 7: Number of available raw acquisitions (Figures 7a to 7d) for each flavor. HH (VV) stands for H-Pol (V-Pol) and A (F) stands for Aft (Fore). The total number of integrated views is represented in Figure 7e, while the retrieved winds are depicted in Figure 7f.

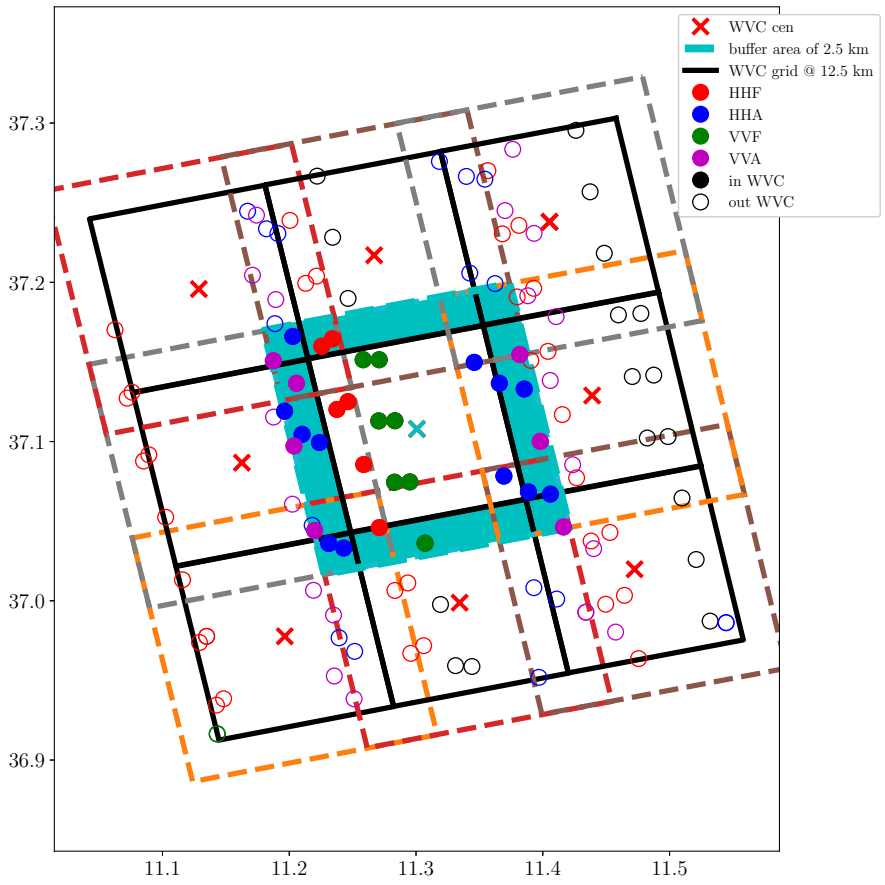
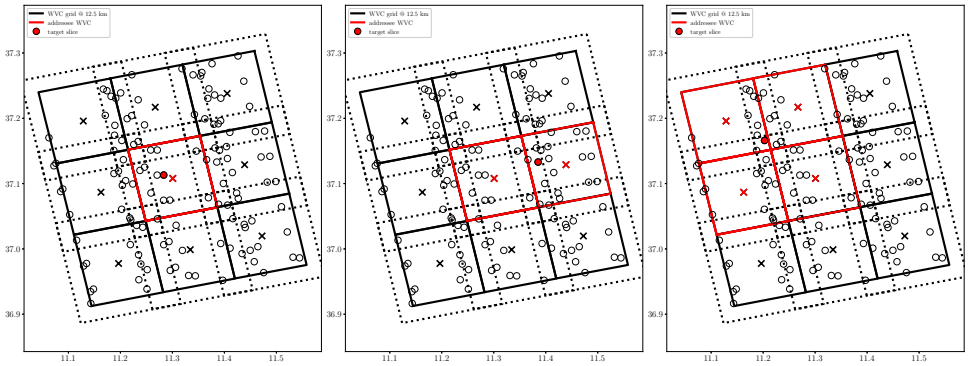


Figure 8: 3x3 WVC matrix extracted from the WVC grid at 12.5 km in the Sicilian Channel (same as Figure 7). The target WVC centroid is represented by the cyan cross. Full-colored markers represent the slices that are assigned to the target WVC. Color code represents the four flavors. Black solid lines: WVC grid at 12.5 km. Dotted lines: buffer boundaries.



(a) (b) (c)

Figure 9: Scheme of slice assignment. Red marker: centroid of the target slice. Red frame: WVC(s) of assignment. Solid black frames: WVC grid at 12.5 km. Dotted frames: WVC buffer areas (in this case, the size is 2.5 km).

Code	SIZE (km)	SRF Δ (km)	MASK Δ (km)
A LR	160	1.1	1.1
A LR	200	1.1	1.1
A FR	160	0.5	0.1
A FR	160	0.1	0.1
A FR	200	0.1	0.1
LUT LR	NA	$\approx$2	1.1
LUT FR	NA	$\approx$2	0.1

Table 3: Table of sensitivity tests. A (LUT) stands for Analytical (LUT-based); LR (FR) stands for Low Resolution (Full Resolution) LSM. Note that the software using LR and FR LSM are two distinct versions. NA stands for Not Applicable.

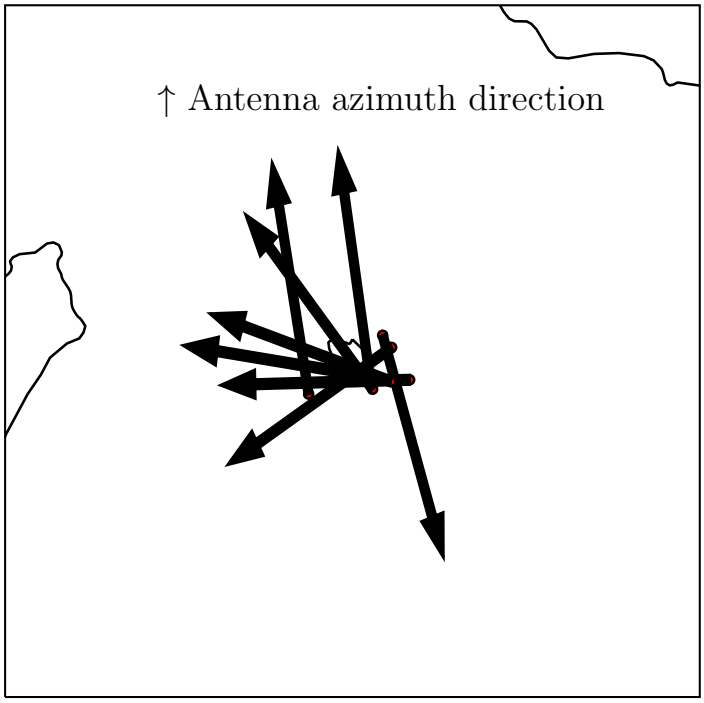
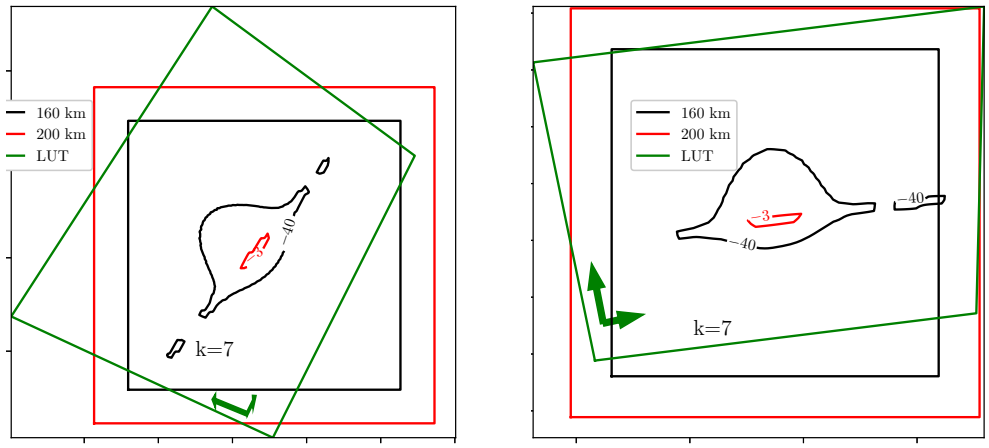


Figure 10: Map of the island of Pantelleria, with the arrows indicating the antenna azimuth angle of the eight acquisitions. Red markers: nadir satellite platform position.



(a) (b)

Figure 11: Left (right): LUT-derived SRF isolines for slice number 7 (0-based numbering) at high (low) latitudes. Short (long) green arrow: range (cross-range) direction. Green box: LUT-based SRF area.

analytical SRF. This threshold value was selected because it guarantees that the error in f is lower than 1e-3.
Figure 11a (11b) shows the isoline contours at -3 dB (red) and -40 dB (black) of the LUT-derived SRF at high (low) latitudes for a peripheral slice (number 7, in 0-based numbering). The green box represents the area of definition of the LUT-based SRF, while the black (red) box represents an area of 160x160 km^2 (200x200 km^2), which represents the area of definition of the analytical SRF. This figure shows that if we consider a box with linear size of 160 km, only some negligible parts of the SRF (below -40 dB) are neglected. This causes an error lower than 7e-4 in f. Note that a peripheral slice was considered because of the presence of a bump in the SRF shape.

4 Dataset

In this study, all the fourteen orbits acquired during the 10^{th} of April 2007 were used. They are identified by orbit IDs from 40651 to 40664. They are available free of charge from the PO.DAAC web site [5]. The files used here are exactly the same as those used in [7]. However, the purpose of this study is different, and therefore the quality control (QC) applied is slightly different.

Two QuikSCAT quality flags are used for QC, namely `sigma0_qual_flag` and `slice_qual_flag`. Only the first of the 16 bits of `sigma0_qual_flag` is used, which accounts for the usability of the entire egg. If the egg is usable, the remaining QC is done by `slice_qual_flag`. Only 0.65% of the eggs is usable.

`slice_qual_flag` consists of 32 bits, 4 per each of the 8 slices acquired in the same egg. Bit 0 of each quadruplet is raised if the gain does not exceed the peak gain threshold value; Bit 1 accounts for the sign of the slice σ_0; Bit 2 is raised if the Signal to Noise Ratio (SNR) is under an acceptable threshold value; Bit 3 is raised if the slice center is not located. In case one among bits 0-2-3 is raised, the QC implemented in `seawinds_l1b_bufr` sets the slice σ_0 value to `missing_values`. Note that negative values are not discarded, but almost half of them are flagged because bits 0, 2 and 3 are raised. Almost 3% of the negative slices are good to use.

5 Results and discussion

5.1 Software performances

The results of the sensitivity tests show that if a linear SRF size of 200 km is used instead of 160 km, differences in f are less than 2e-4. This result suggests that 160 km represents a good compromise between accuracy and computational efficiency.
Using a LR LSM instead of HR LSM (1.1 km *vs* 0.1 km at the equator), the differences in f are lower than 7e-4 for the analytical code. Instead, they are less than 0.015 for the LUT-based code. These results suggest that HR LSM is not worth it; therefore, LR LSM is appropriate for our needs.
Finally, considering the fastest configurations of both analytical and LUT-based codes (A 1.1 km *vs* LUT 1.1 km), the differences can increase to 0.03. Figure 12b shows the case for which the highest differences occur. Red (black) curves represent the -3 dB and -30 dB LUT-based (analytical) SRF isoline contours, while the green contour represents the island of Pantelleria. For the sake of repeatability, this case is extracted from orbit ID 40653, with frame index 3717, pulse index 15 and slice index 3 (all indices are 0-based). -3 dB contours of all the slices of the footprint are reported in Figure 12a with the same color code. Note that the wavy pattern of the analytical -30 dB contour is not present in the LUT-based contour due to its coarser spatial resolution. However, due to the very low SRF values, they are not responsible for the differences. Instead, it is apparent that the LUT-based SRF -3 dB contour is narrower than the corresponding analytical one. However, these differences are not expected to play an important role in wind retrievals.

All code configurations were timed on a desktop machine equipped with the following hardware:

- Processor: Intel Xeon Gold 5218;
- clock: 2.3 GHz;
- # of cores: 32;
- # of CPUs: 64;
- RAM: 66 Gb;

Furthermore, all codes were checked with `valgrind`, which is a tool to check memory errors and code efficiency, among other things.
The benchmark of this performance evaluation is the orbit with ID 40653, and all codes were compiled with the only option `-O2`. Only four of the eight

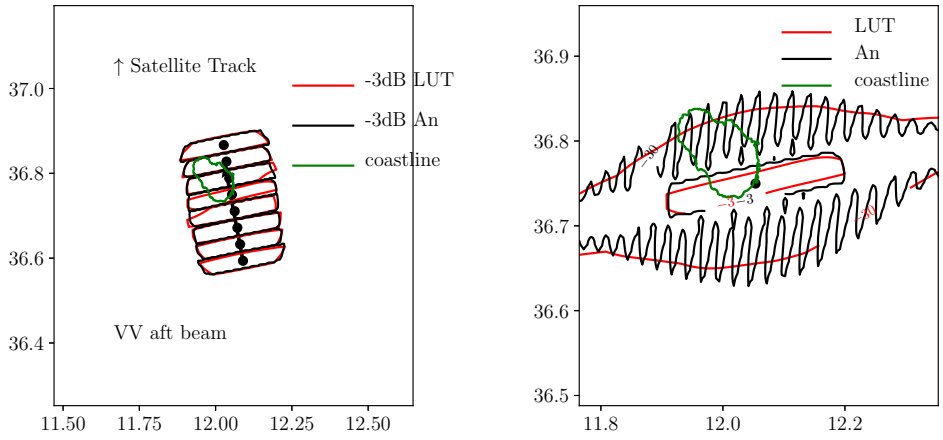


Figure 12: Left plot, red (black) curves: -3 dB SRF contour isolines obtained with LUT-based (analytical) SRF for the case with highest differences between LUT 1.1 km and A 1.1 km (see Table 3). The green contour represents the island of Pantelleria. Right plot: zoom in on slice number 3 from the left plot, for which the highest differences occur. -30 dB contours are also present. This case corresponds to orbit ID 40653, frame index 3717, pulse index 15 and slice index 3. All indices are 0-based. $f^A = 0.1518$ and $f^{LUT} = 0.1221$.

Code	SIZE (km)	SRF Δ (km)	LSM Δ (km)	TIME HH:MM:SS	RAM (Gb)
A LR	160	1.1	1.1	07:02:05	1.5
A LR	200	1.1	1.1	11:02:51	1.5
A FR	160	0.5	0.1	37:11:43	3.1
LUT LR	NA	$\approx$2	1.1	01:04:42	3.5

Table 4: Table of timing.

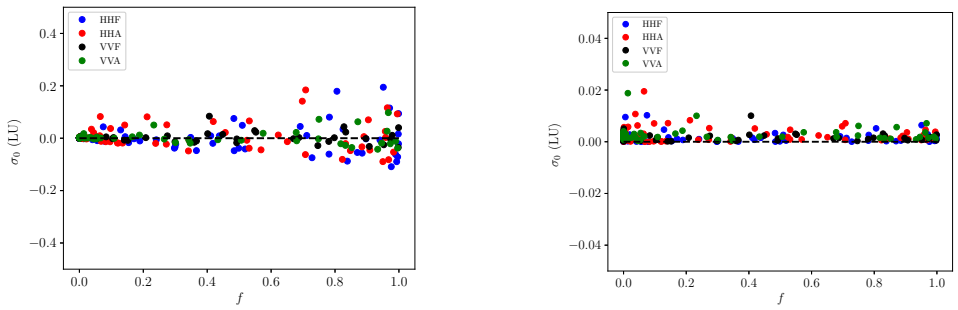
configurations described in Table 3 were timed because the others were too time consuming. Table 4 resumes the results.

The results show that LUT LR is the most efficient code. It takes "only" around 1 H to process an entire orbit. The most efficient analytical code takes 7 times this time. The results suggest that LUT LR should be used for operational purposes.
Furthermore, the leading factor in the accuracy of f is the method used to model the SRF. Other aspects, namely grid spacing, grid size, and LSM resolution, play a secondary role.
The best computational rate is $\approx$1 orbit/H. This is not very low, but with the aid of a dedicated cluster, parallel computation can help scaling this rate, depending on the available number of CPUs. The results shown later in this section were obtained with f computed with the LUT LR code version.

5.2 Noise regularization

Figure 13b shows the corrected σ_0s for the WVC shown in Figure 5 after the noise regularization procedure was applied. This scatter plot is compared to that obtained with a simple linear de-trend methodology, used in [7] for Seawinds and inspired by [6]. Note that the negative σ_0s disappeared and that y-axis of Figure 13b are ten times smaller than in Figure 13a. Note that while the scatter in Figure 13a increases with f, it is quite constant in Figure 13b. In fact, the level of noise after applying the noise regularization is quite constant with f, whatever the original level of noise of the contaminated values was. This is the reason why this technique was given this name.

Figure 14b (14a) shows a map of corrected (non corrected) σ_0s offshore Sicily, in the Mediterranean basin. It is apparent that noise regularization is effective in correcting coastal contaminated acquisitions. There are some dark red markers left, very few actually, but they are rather close to the coastline. In particular, the distribution of coastal σ_0s is similar to that of



(a) Linear de-trend

(b) Noise regularization

Figure 13: Right (left): scatter plot of corrected σ_0 with noise regularization (linear de-trend) *vs* f for the WVC off-shore Sicily shown in Figure 4. σ_0s are segregated according to flavor: HH (VV) stands for H-Pol (V-Pol) and A (F) for Aft (Fore). Note that y-axis limits of Figure 13b are 10 times smaller than those of Figure 13a. Finally, note that no negative values are present in Figure 13b.

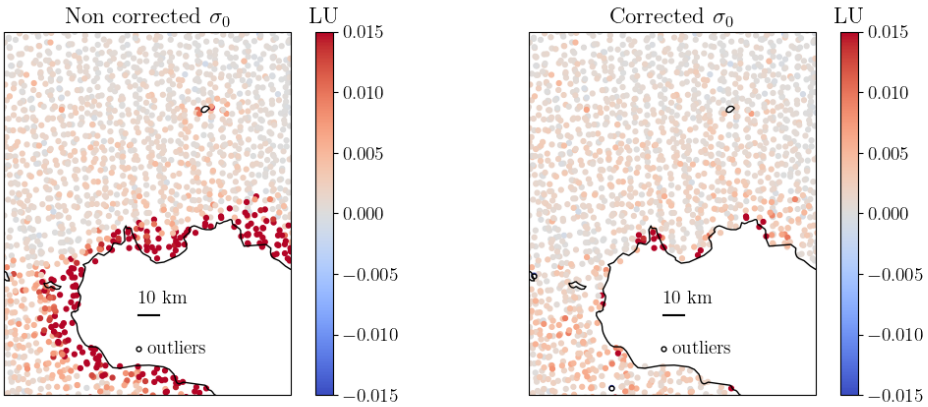
open sea σ_0s. This aspect is in line with expectations, even if one should be careful when comparing coastal distributions to open-sea distributions. In fact, coastal dynamics can heavily characterize the distribution of some geophysical parameters, and some meaningful differences should be expected. Note that the color bar used is of the kind "cool-warm" to emphasize the presence of negative σ_0s. In addition, slices with missing values after correction are reported as empty black circles. Note that their number is very low.

Figure 15b shows the same for the area test offshore Netherlands. The improvement is apparent here as well. Note the presence of dark spots in the surroundings of the Rotterdam harbor. This is likely caused by the massive presence of ships, due to the intense traffic. The same likely happens in the internal sea IJsselmeer.

Finally, noise regularization appears to significantly improve the correction of σ_0 compared to the linear detrend implemented in [7]. Both maps obtained with the old method are reported in Figure 16 for completeness. Note that all colorbars are identical to make comparison easier.

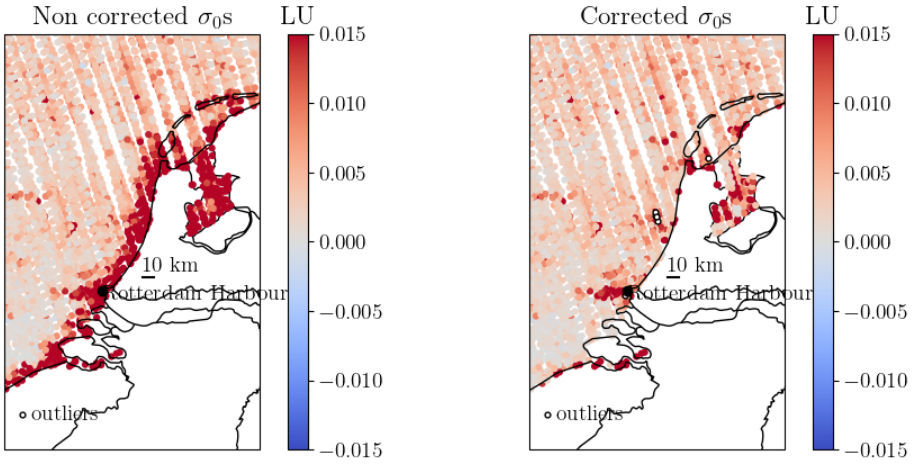
5.3 Wind retrievals

Figure 17 shows the Seawinds-derived wind field for six different buffer areas surrounding the WVCs. The target area is in the central Mediterranean basin, where orbit ID 40653 has the sampling issues described previously.



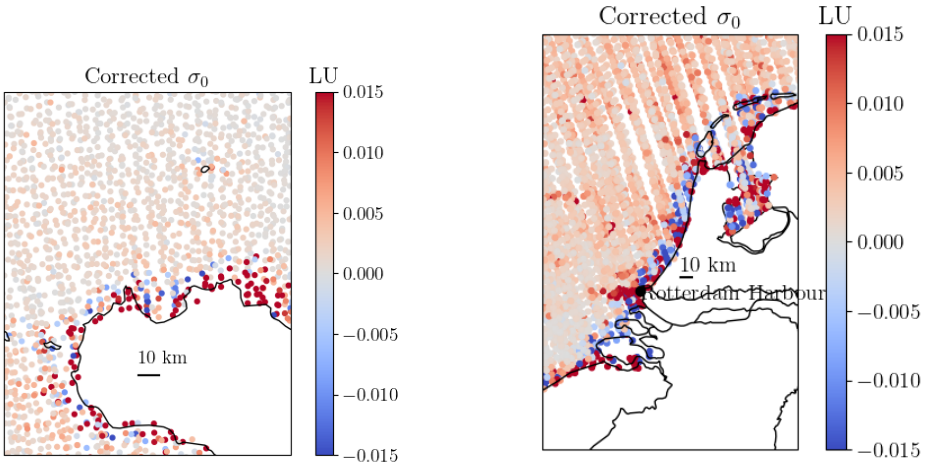
(a) (b)

Figure 14: Right (left): map of corrected (non corrected) σ_0s (linear units) in an area test offshore Sicily in the Mediterranean basin. Note that the color bar is of the kind "cool-warm" with equal absolute extreme values to emphasize the presence of any negative σ_0 (blue). Note that only two slices have missing values (outliers).



(a) (b)

Figure 15: Right (left): map of corrected (non corrected) σ_0s (linear units) in an area test offshore Netherlands. Note that the color bar is of the kind "cool-warm" with equal absolute extreme values to emphasize the presence of any negative σ_0 (blue). Note that only five slices have missing values (outliers).

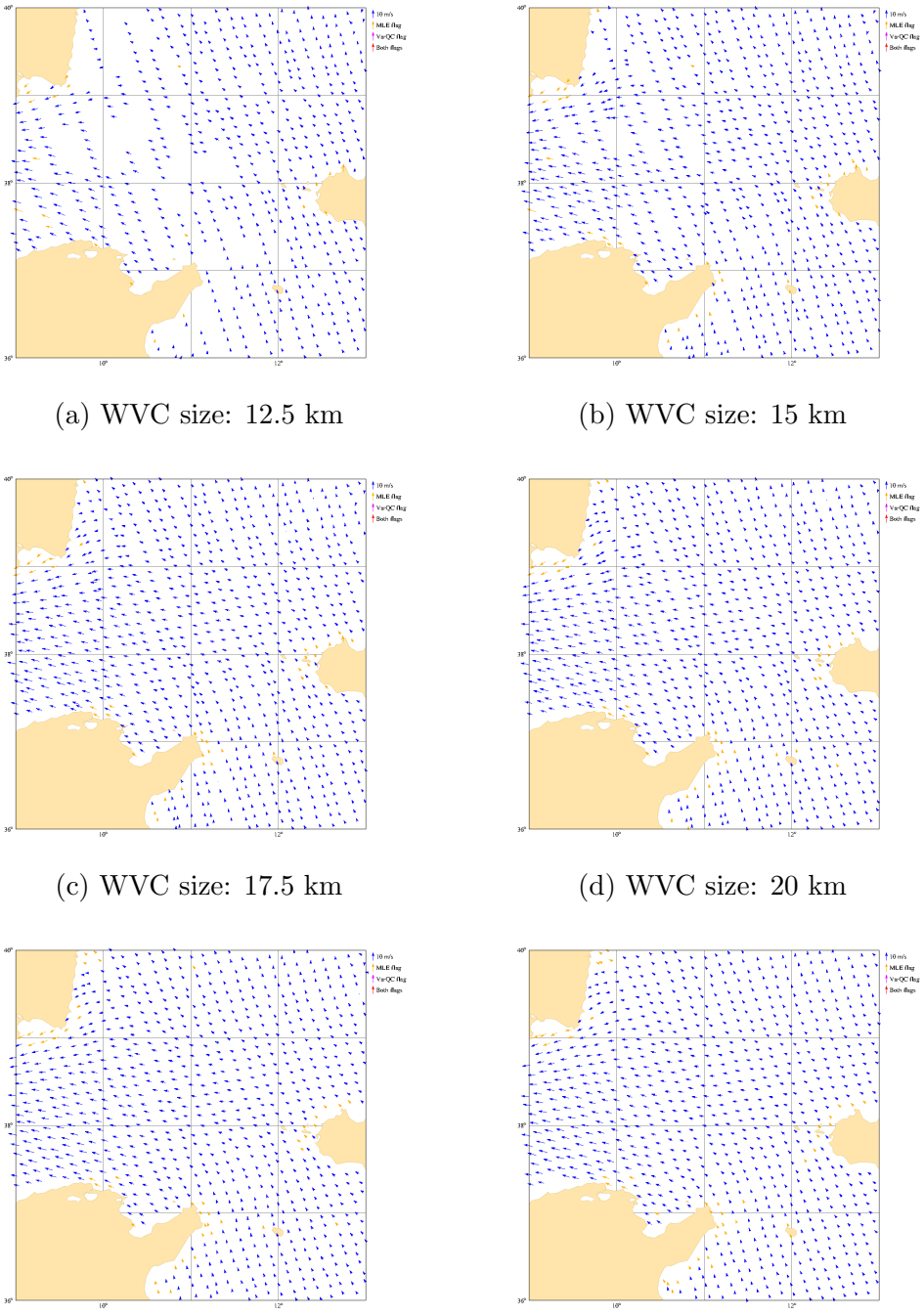


(a) (b)

Figure 16: Left (right): map of corrected σ_0s (linear units) offshore Sicily (Netherlands) obtained with the method described in [7]. Note that the color bar is identical to that of Figures 14 and 15, to make comparison easier.

The buffer area varies from 0 (WVC size equal to 12.5 km in Figure 17a) to 6.25 km (WVC size equal to 25 km in Figure 17f). Note that a buffer size of 2.5 km is equivalent to a WVC size of 17.5 km (2.5 km per each side). Furthermore, note that the WVC spacing is always set to 12.5 km.
It can be seen that the larger is the buffer area, the lower the number of empty WVCs is. In particular, when the buffer area is set to 0, the number of missing WVCs is quite large. A buffer area of 2.5 km seems the best compromise between the capability to "fill" all WVCs and the spatial resolution of the retrieved winds.
Note that the larger is the WVC buffer area, the more regular the WVC grid appears. This is due to the fact that, enlarging the WVC size, the number of slices increases, and the after-retrieval WVC centroid is closer to the pre-retrieval centroid.

Three different retrieval experiments were set up, as resumed in Table 5: CTRL stands for "control" and represents the experiment with the state-of-the-art configuration at OSI-SAF: all slices with $f > 0.02$ are discarded and no correction was applied; NC stands for "No-Correction": all slices with $f > 0.5$ are discarded and no correction was applied. This experiment is run to assess the impact of noise regularization; Finally, NR stands for "Noise Regularization": all slices with $f > 0.5$ are discarded and noise regularization was applied.



(a) WVC size: 12.5 km

(b) WVC size: 15 km

(c) WVC size: 17.5 km

(d) WVC size: 20 km

(e) WVC size: 22.5 km

(f) WVC size: 25 km

Figure 17: Wind vector field derived from Seawinds with varying buffer area sizes in the central Mediterranean basin. The orbit ID is 40653. Blue arrows: regular winds. The scale of the arrows is reported in the legend. Yellow (magenta) arrows: the `knmi_quality_control_fails` (`variational_quality_control_fails`) flag is raised. Red arrows: both flags are raised.

Name	f^{th}	Buffer size (km)	Noise reg	Orbit IDs
CTRL	0.02	2.5	NO	40651-40664
NC	0.5	2.5	NO	40651-40664
NR	0.5	2.5	YES	40651-40664

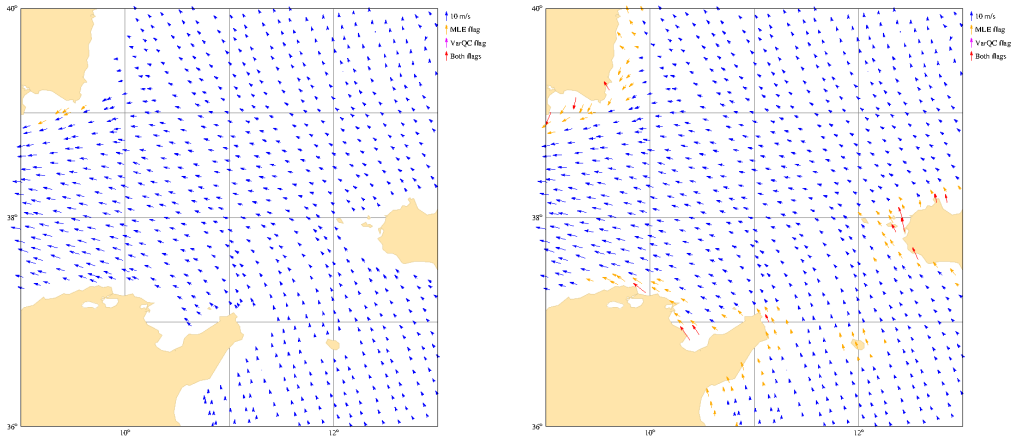
Table 5: Table of experiments. "reg" stands for "regularization". Note that a buffer size of 2.5 km corresponds to a WVC size of 17.5 km, and that WVC centroids are on a 12.5 km grid.

Figure 18a shows the Seawinds-derived wind field in the central Mediterranean area with CTRL. As expected, the coastal band within ≈15-20 km is scarcely sampled. Figure 18b shows the retrieval obtained with NC. It is apparent that the sampling improves but the quality of the retrievals dramatically degrades. In fact, the number of flagged WVCs is very high in the coastal band, both those with MLE (`knmi_quality_control_fails`) raised and those with MLE and VarQC (`variational_quality_control_fails`) raised together. The situation improves considerably in Figure 18c, where the sampling with good quality retrievals improves and the occurrences of flagged WVCs are quite reduced compared to NC. Note that flagged winds are consistent with the surrounding good winds, a sign that land contamination is reduced and that a proper tuning of MLE threshold values could further reduce flagging. This aspect is left for the future.

Figures 19 and 20 show the wind field offshore Sicily, in the Mediterranean basin, and offshore Netherlands, respectively. The same comments expressed for Figure 18 also apply in this case.

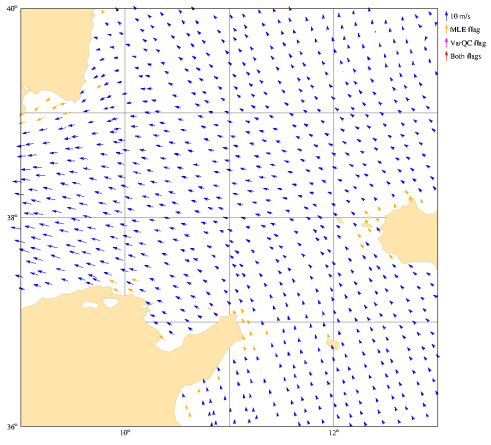
Finally, Figure 21 shows the comparison among the three experiments in the almost closed northern Adriatic basin. In a narrow basin like this, the sampling of coastal areas almost doubles if noise regularization is applied. Note that the Venetian lagoon, which is located in the northwestern part of the Adriatic, is sadly known for the frequent "Acqua Alta" (literally "high water") events, which are essentially driven by the simultaneous presence of strong tides, cyclonic atmospheric patterns, and Scirocco. Acqua Alta events are not easily predictable because numerical weather prediction (NWP) models are not accurate along the coast, and scatterometer-derived winds are often flagged due to land contamination. Figure 21 shows that NR sampling is quite satisfactory close to the lagoon, therefore this methodology offers new perspectives in this area. Of course, the quality of the retrieved winds should be properly validated in several wind regimes.

Figure 22a (22b) shows the sampling ratio of NR w.r.t. CTRL (NC) at various distances from the coastline. Data are binned in 5 km width bins.



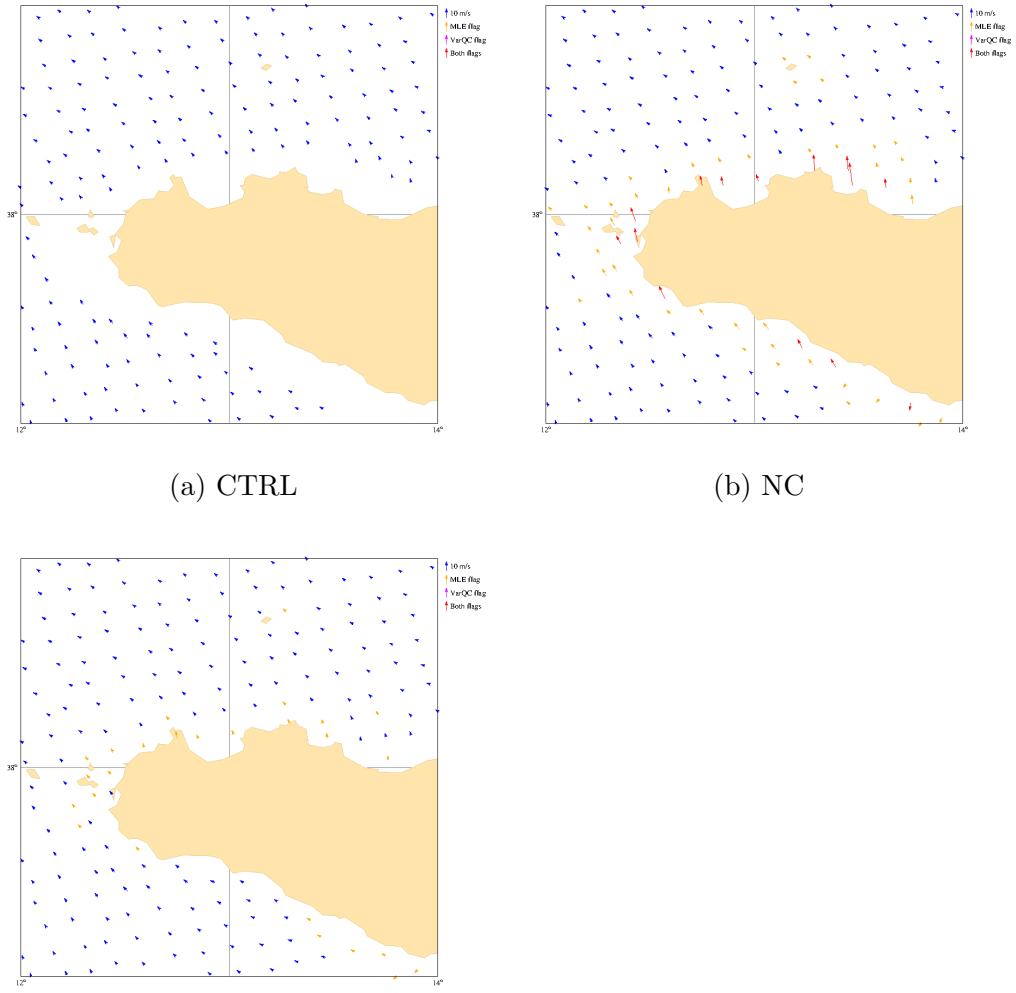
(a) CTRL

(b) NC



(c) NR

Figure 18: Left: wind vector field derived from Seawinds in the central Mediterranean basin with CTRL. Center: same as left with NC. Right: same as left with NR. The orbit ID is 40653. Blue arrows: regular winds. The scale of the arrows is reported in the legend. Yellow (magenta) arrows: the `knmi_quality_control_fails` (`variational_quality_control_fails`) flag is raised. Red arrows: both flags are raised.

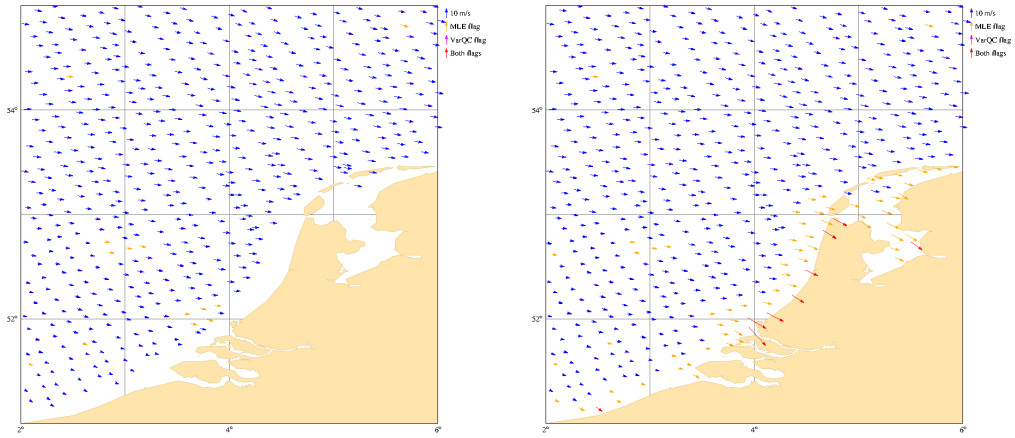


(a) CTRL

(b) NC

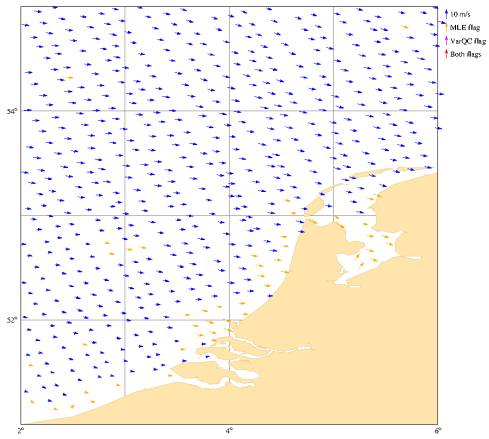
(c) NR

Figure 19: Left: wind vector field derived from Seawinds offshore Sicilty, in the Mediterranean basin, with CTRL. Center: same as left with NC. Right: same as left with NR. The orbit ID is 40653. Blue arrows: regular winds. The scale of the arrows is reported in the legend. Yellow (magenta) arrows: the `knmi_quality_control_fails` (`variational_quality_control_fails`) flag is raised. Red arrows: both flags are raised.



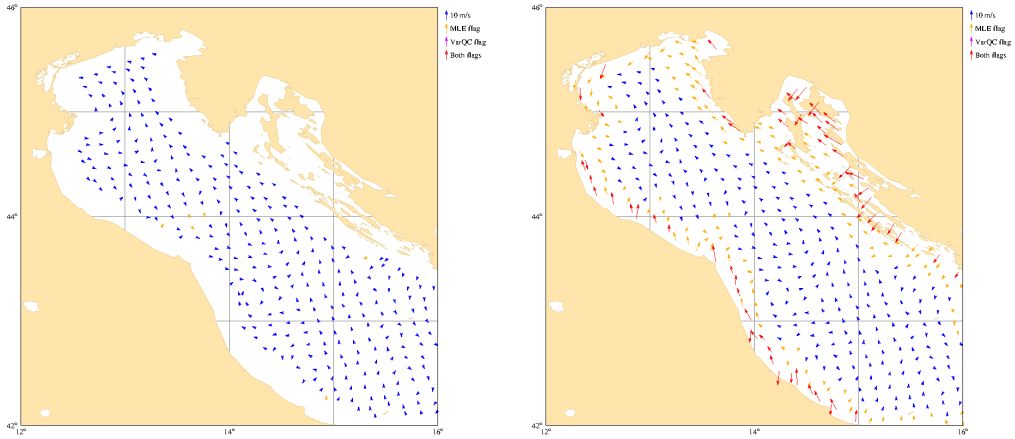
(a) CTRL

(b) NC



(c) NR

Figure 20: Left: wind vector field derived from Seawinds offshore Netherlands with CTRL. Center: same as left with NC. Right: same as left with NR. The orbit ID is 40653. Blue arrows: regular winds. The scale of the arrows is reported in the legend. Yellow (magenta) arrows: the `knmi_quality_control_fails` (`variational_quality_control_fails`) flag is raised. Red arrows: both flags are raised.

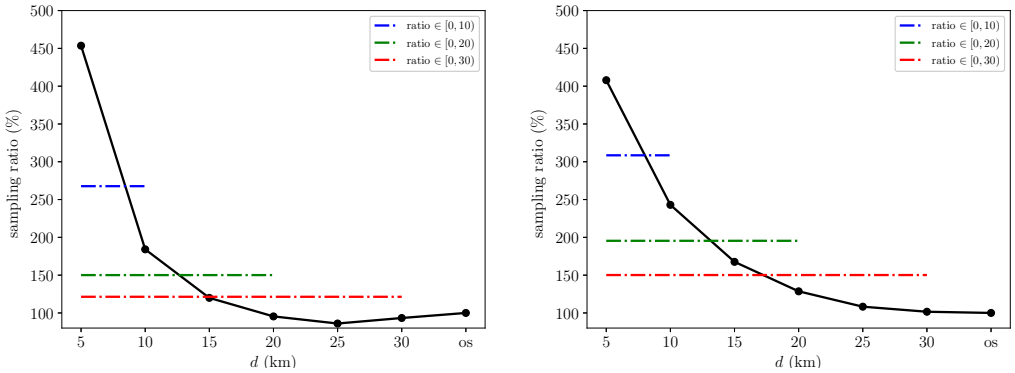


(a) CTRL

(b) NC

(c) NR

Figure 21: Left: wind vector field derived from Seawinds in the northern Adriatic sea, in the Mediterranean basin, with CTRL. Center: same as left with NC. Right: same as left with NR. The orbit ID is 40653. Blue arrows: regular winds. The scale of the arrows is reported in the legend. Yellow (magenta) arrows: the `knmi_quality_control_fails` (`variational_quality_control_fails`) flag is raised. Red arrows: both flags are raised.

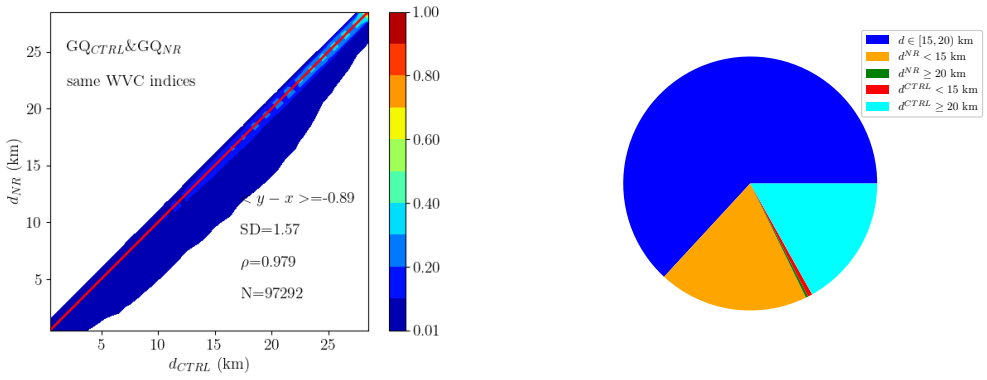


(a) CTRL (b) NC

Figure 22: Left (right): sampling rate of NR w.r.t. CTRL (NC) at several distance to the coastline (bins of 5 km). "os" stands for offshore. Blue dash-dotted curve represents the sampling ratio in the first 10 km to the coastline. Red and green lines represent the same ratio in the first 20 and 30 km, respectively.

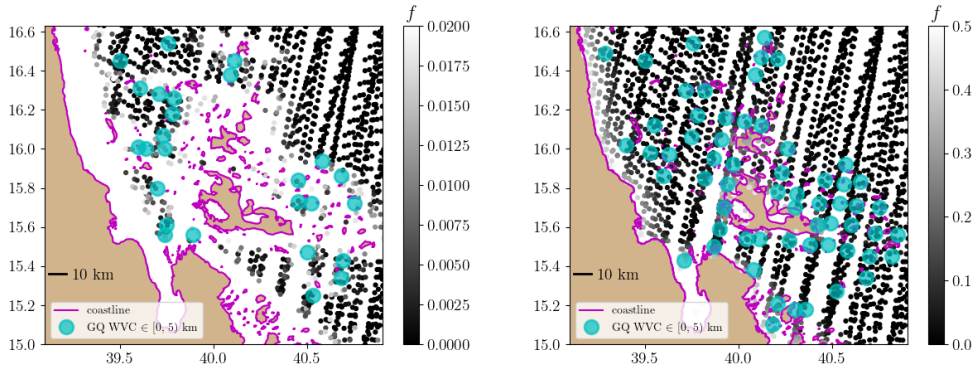
The dashed-dotted blue line accounts for the sampling rate in the first 10 km from the coast, while the red and green lines account for the same ratio in the first 20 and 30 km, respectively. It is apparent that the sampling improves by 30% (50%) w.r.t. CTRL (NC) in the coastal band within 30 km. This number increases to 450% (400%+) in the first 5 km.

Note that the sampling ratio w.r.t. CTRL falls below 100% between 20 and 30 km. The reason is that the WVC centroids are on average farther from the coast in CTRL compared to NR. This causes a different binning, especially at these distances. To clarify this, Figure 23a shows a 2D histogram of the distance from the coast of the WVC centroids obtained with NR w.r.t. the WVC distance obtained with CTRL. Some statistical information is reported in the panel, among which the bias, equal to slightly less than 1 km toward the coast. Furthermore, the pie chart of Figure 23b shows the distribution of the good quality (GQ) WVC centroid ditstances related to the bin $[15, 20)$ km. In particular, the blue slice represents the share of WVCs with the same indices for which d^{CTRL} and d^{NR} are in the same bin. The yellow (red) slice represents the share of WVCs with the same indices for which $d^{NR} < 15$ km ($d^{CTRL} < 15$ km) and $d^{CTRL} \in [15 - 20)$ km ($d^{NR} \in [15 - 20)$). The cyan (green) slice represents the share of WVCs with the same indices for which $d^{CTRL} \geq 20$ km ($d^{NR} \geq 20$ km) and $d^{NR} \in [15 - 20)$ km ($d^{CTRL} \in [15 - 20)$ km). Cyan and yellow slices are much larger than green and red, which testifies that CTRL WVCs are generally farther away from the coast than NR WVCs.



(a) (b)

Figure 23: Left: 2D histogram of the WVC distance from the coast for NR *vs* CTRL with same WVC indices. The histogram is normalized to the maximum value. The bias is reported with $< y - x >$, GQ stands for good quality, SD for standard deviation, ρ for correlation coefficient and N is the total number of samples. Right: The blue slice represents the share of GQ WVCs with the same indices belonging to the same bin with $d \in [15 - 20)$ km. The yellow (red) slice represents the share of GQ WVCs with the same indices for which $d^{NR} < 15$ km ($d^{CTRL} < 15$ km) and $d^{CTRL} \in [15 - 20)$ km ($d^{NR} \in [15 - 20)$ km). The cyan (green) slice represents the share of GQ WVCs with the same indices for which $d^{CTRL} \geq 20$ km ($d^{NR} \geq 20$ km) and $d^{NR} \in [15 - 20)$ km ($d^{CTRL} \in [15 - 20)$ km).



(a) (b)

Figure 24: Left (right): map of f in the area of the Dahlak archipelago, in the Red Sea. Slice centroids with $f > 0.02$ ($f > 0.5$) are discarded. Cyan markers represent the after-retrieval WVC centroids within 5 km for the CTRL (NR) experiment. Note that the two colorbars saturate at different levels.

Finally, note that the number of WVCs within 5 km from the coast should be almost absent for CTRL. Actually, there are a few of them. Figure 24 can help clarify this aspect. Figure 24a (24b) shows a map of the slice centroids with $f < 0.02$ ($f < 0.5$) in the archipelago of Dahlak, in the Red Sea. Gray tones relate to f. Cyan markers represent the WVC centroids within 5 km to the coast. Note that all CTRL retrieval occurrences are close to small islands, and never to the continental coastline. In fact, in the cases of small f, slices are not discarded and wind retrieval can occur.

Figure 25 shows the distributions of wind speed, separated according to the distance to the coastline. For comparison, the distribution of the offshore wind speeds and those of the entire coastal band ($[0, 30)$ km) are reported (dashed black and dashed cyan curves, respectively).
A few things are apparent:

- Offshore wind distributions are similar for all the experiments. This is expected since land contamination is negligible at 30+ km from the coast;
- NR coastal distributions are generally smoother than NC and CTRL, because they are more populated (see following figures);
- Offshore distribution is quite different from those coastal. This is also expected because coastal winds are generally lower than offshore. How-

ever, note that NC distributions are quite different than CTRL and NR, which are more similar to each other. In particular, coastal winds from NC are higher than those from CTRL and NR. This is likely due to residual land contamination;

- Finally, NR winds $\in [0, 10)$ km have a plateau within 0 and $\approx$3 ms^{-1} when compared to CTRL and NC. The reasons are not clear and should be further investigated in the future. In particular, the possibility that noise regularization can under correct land contamination will be examined. In fact, in this case, low winds would be less represented in the distribution, while the presence of medium winds would be more represented. However, as stressed in Figure 24, CTRL and NR can refer to different coastal areas in that range.

Figure 26 (27) shows the 2D histograms of the wind speeds (zonal and meridional wind components) compared to ECMWF for all the experiments. Histograms are normalized to the maximum value. Some statistical information is reported in the panels: bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N). This is an additional consistency check that confirms what is already revealed by Figure 25. In fact, all plots are very similar, which confirms that land contamination is negligible far from the coast.

Figure 28 (29) shows the same plots as Figure 26 (27) but for the coastal band within 30 km to the coast. Differences are more marked in this case. In fact, first of all, the total number of samples is quite different among the three experiments, being the highest for NR and the lowest for NC. In fact, even if NC sampling improves compared to CTRL, wind quality degrades, and many WVCs are flagged. Furthermore, the bias increases from CTRL to NC and NR. From the plots, it seems that the highest biases are located at low wind regimes, whatever the experiment is. This result seems to suggest that these biases are not related to σ_0 correction, but rather to model biases in coastal areas. They are well known in the literature, as extensively discussed in [12]. Finally, note that coastal sampling is not homogeneous for the three experiments (Figure 22), therefore, biases in the coastal band within 10 km weight differently in NR than in CTRL and NC. This aspect will be further discussed later in this section.

Figures 30, 32 and 34 (31, 33 and 35) show the same information as Figure 28 (29) segregated according to different parts of the swath, namely sweet, nadir and outer. The nadir part of the swath is characterized by a reduced azimuth diversity, which has an impact on the retrieval quality. The outer part is characterized by a lack of inner (H-Pol) acquisitions, with consequent further reduced azimuth and polarization diversity. Statistical

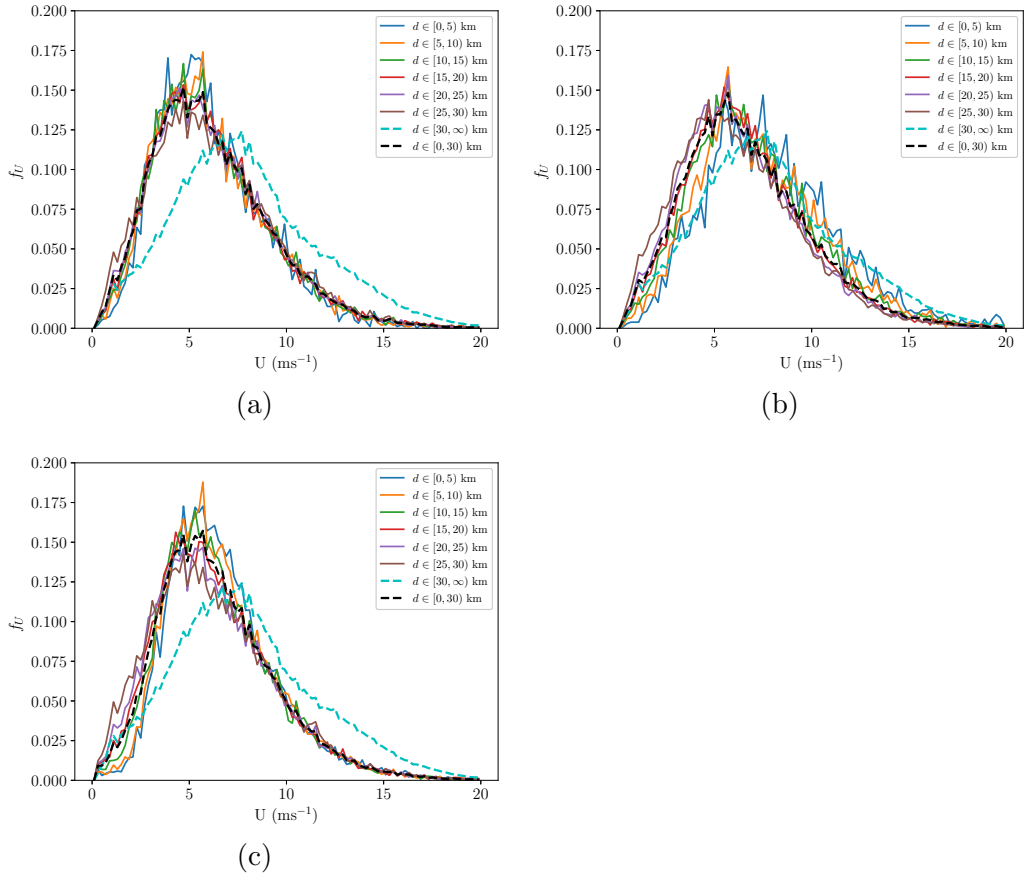
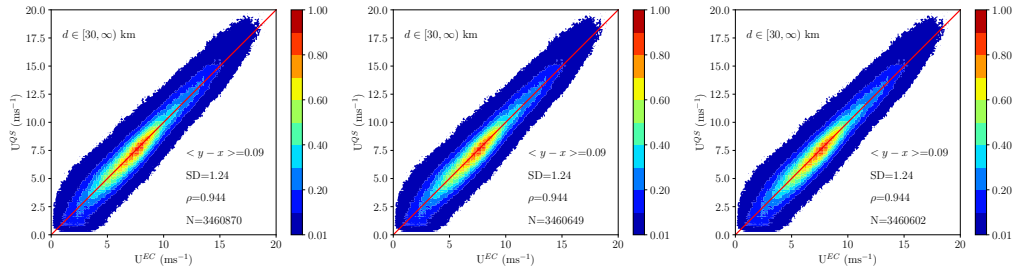
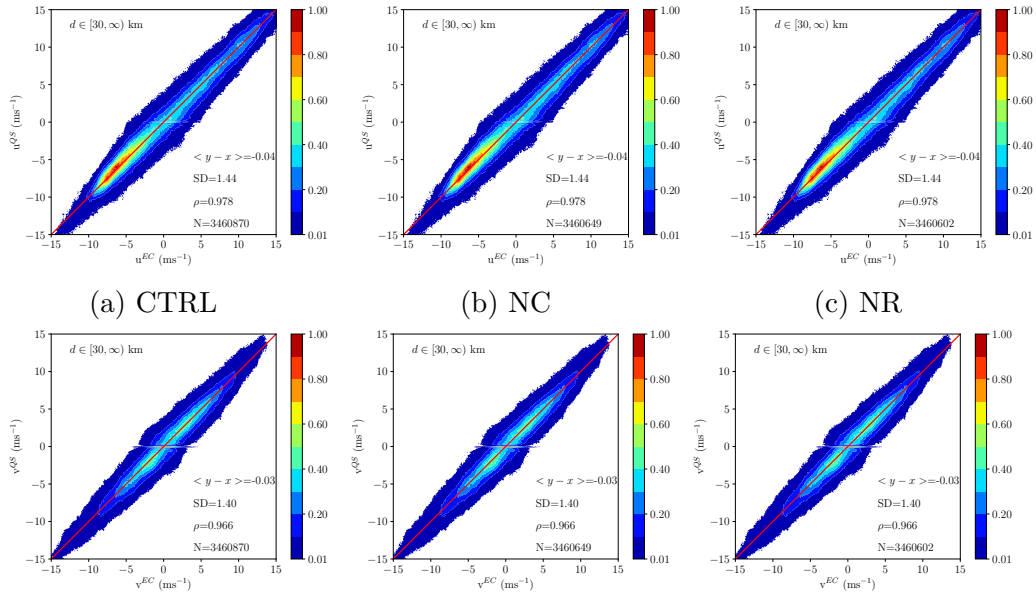


Figure 25: Distribution of the wind speeds at various distance to the coastline. Top left, top right and bottom left relate to CTRL, NC and NR, respectively. Black (cyan) dashed curve represents the pdf of all coastal (offshore) winds ($d \in [0, 30)$ km).



(a) CTRL (b) NC (c) NR

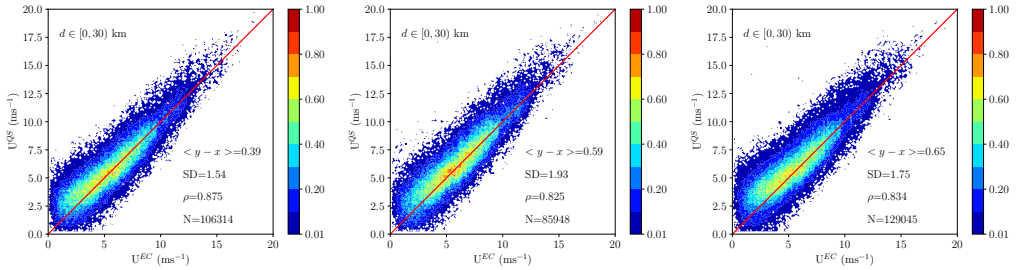
Figure 26: 2D histogram of Seawinds-derived offshore wind speeds (U^{QS}) *vs* ECMWF (U^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

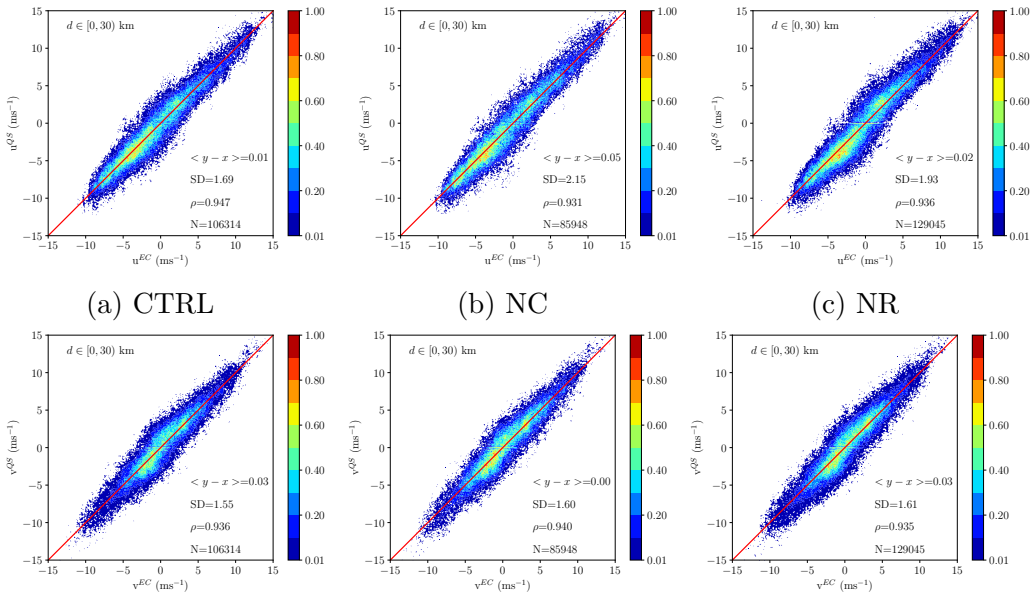
(d) CTRL (e) NC (f) NR

Figure 27: Top (bottom): 2D histogram of Seawinds-derived offshore zonal (meridional) wind component u^{QS} (v^{QS}) *vs* ECMWF u^{EC} (v^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

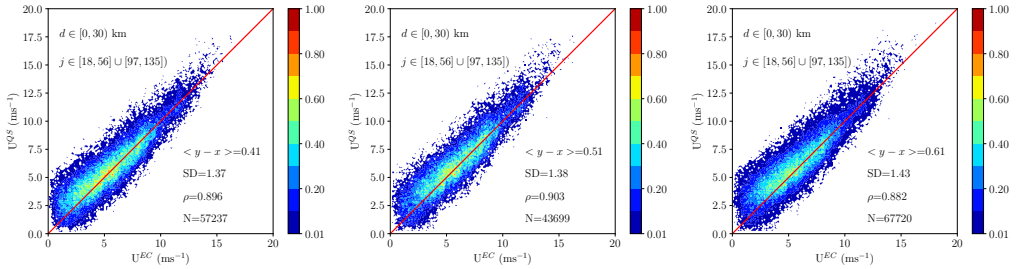
Figure 28: 2D histogram of Seawinds-derived coastal ($d \in [0, 30)$ km) wind speeds (U^{QS}) *vs* ECMWF (U^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

(d) CTRL (e) NC (f) NR

Figure 29: Top (bottom): 2D histogram of Seawinds-derived coastal ($d \in [0, 30)$ km) zonal (meridional) wind component u^{QS} (v^{QS}) *vs* ECMWF u^{EC} (v^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

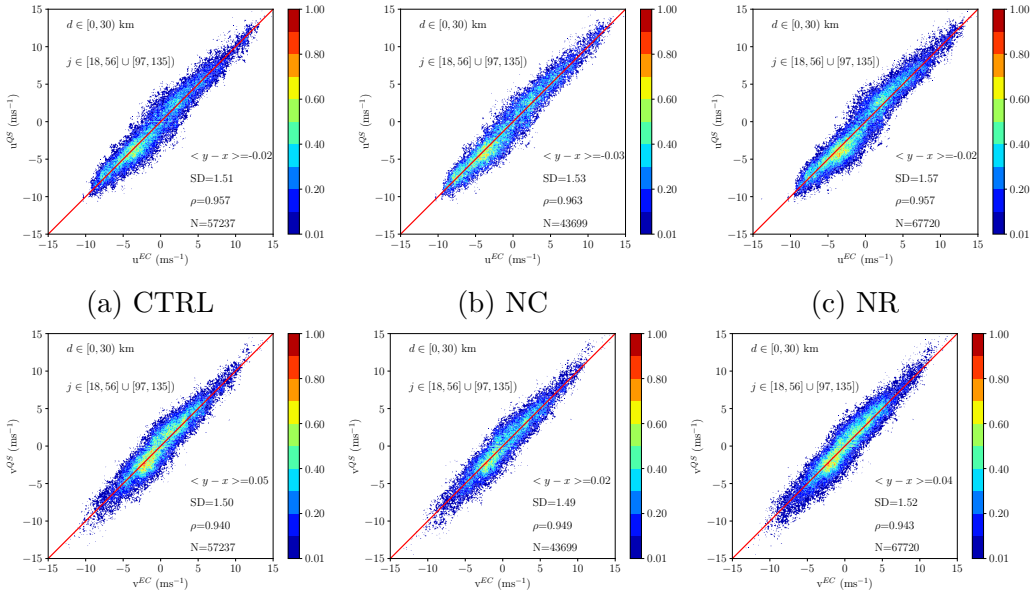
Figure 30: 2D histogram of Seawinds-derived coastal ($d \in [0, 30)$ km) wind speeds (U^{QS}) *vs* ECMWF (U^{EC}) in the sweet part of the swath. The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.

parameters confirm these facts. In fact, SD and biases are larger for outer and nadir w.r.t. sweet, and the differences among the experiments also increase. However, note that the sampling rate does not increase homogeneously in the three zones when noise regularization is applied. In fact, it increases by ≈33% in the outer beam, while it increases only by ≈ 18% in the sweet and nadir portions. This aspect is not clear and deserves a thorough analysis, which is left for the future.

Figures 36, 38, 40, 42, 44 and 46 (37, 39, 41, 43, 45 and 47) show the same information as Figure 28 (29) in different 5 km wide coastal strips. It can be seen that:

- The biases between retrieved winds and model winds reduce with distance to the coast, confirming the previous results;
- The differences among the experiments also reduce;
- NR biases are always lower than NC biases, confirming that noise regularization is effective in reducing land contamination;
- The biases in the coastal band within 5 km are much larger than farther, confirming that the overall coastal bias could be driven by these biases.

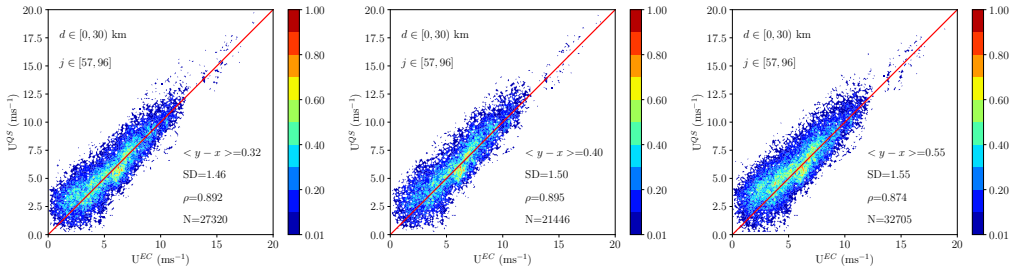
Figure 48 shows the same information as 36 but for MLE-flagged winds. It is interesting to note that the NR biases are much closer to the CTRL biases than the NC biases, confirming that the noise regularization does its job. Furthermore, many samples lie close to the bisector (red line), suggesting that a finer tuning of the MLE thresholds could reduce the number of flagged



(a) CTRL (b) NC (c) NR

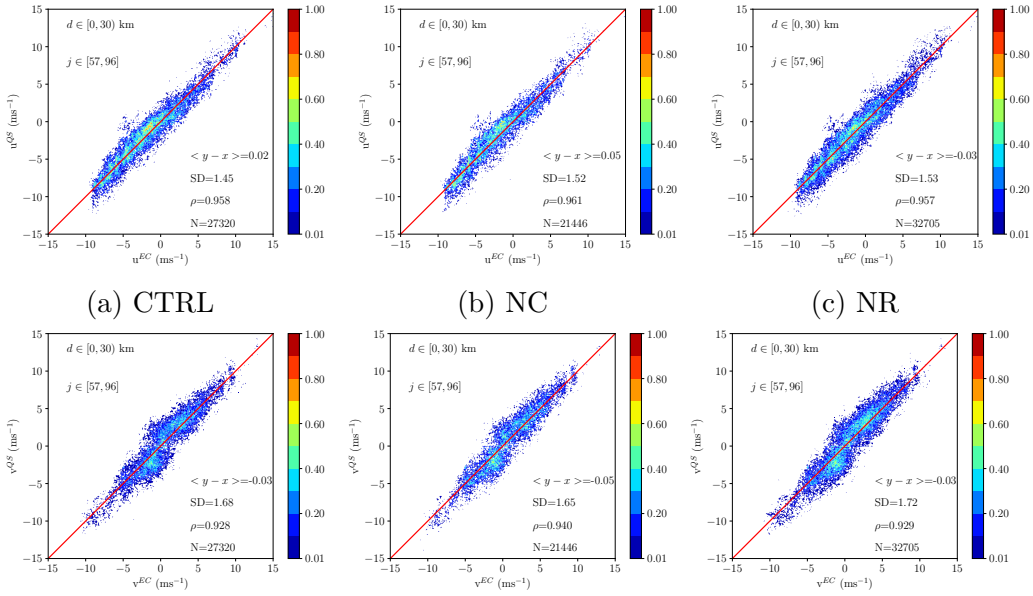
(d) CTRL (e) NC (f) NR

Figure 31: Top (bottom): 2D histogram of Seawinds-derived coastal ($d \in [0, 30)$ km) zonal (meridional) wind component u^{QS} (v^{QS}) *vs* ECMWF u^{EC} (v^{EC}) in the sweet part of the swath. The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

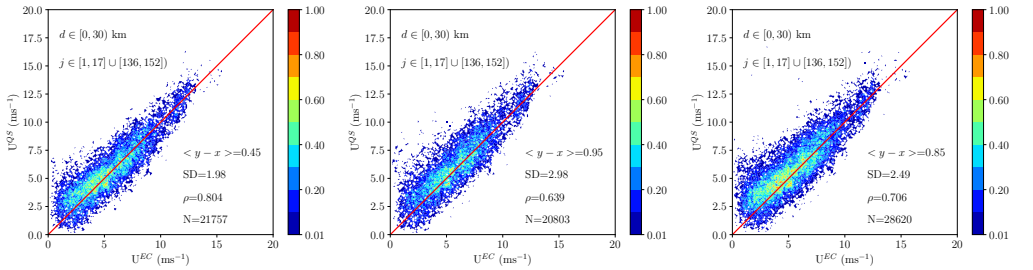
Figure 32: 2D histogram of Seawinds-derived coastal ($d \in [0, 30)$ km) wind speeds (U^{QS}) *vs* ECMWF (U^{EC}) in the nadir part of the swath. The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

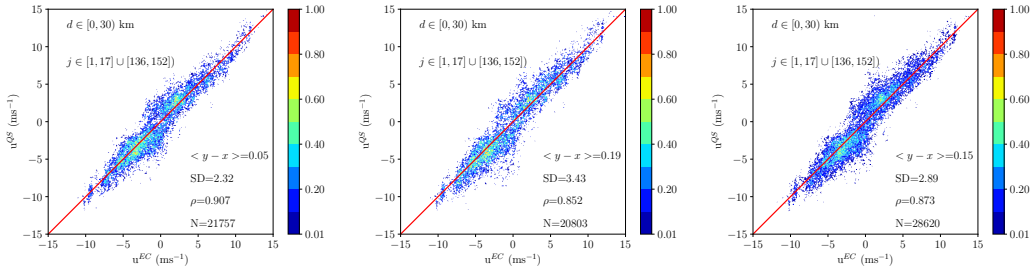
(d) CTRL (e) NC (f) NR

Figure 33: Top (bottom): 2D histogram of Seawinds-derived coastal ($d \in [0, 30)$ km) zonal (meridional) wind component u^{QS} (v^{QS}) *vs* ECMWF u^{EC} (v^{EC}) in the nadir part of the swath. The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.

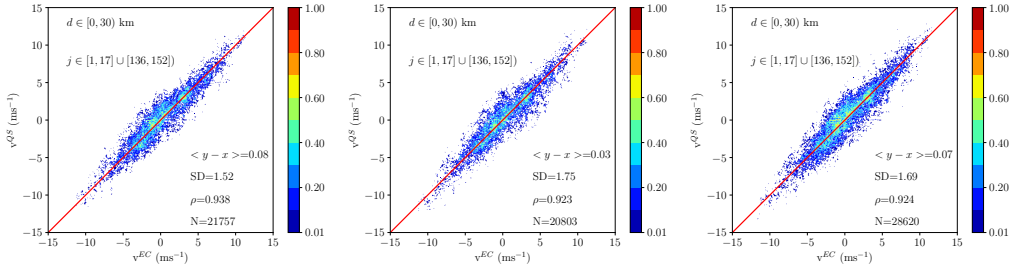


(a) CTRL (b) NC (c) NR

Figure 34: 2D histogram of Seawinds-derived coastal ($d \in [0, 30)$ km) wind speeds (U^{QS}) *vs* ECMWF (U^{EC}) in the outer part of the swath. The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.

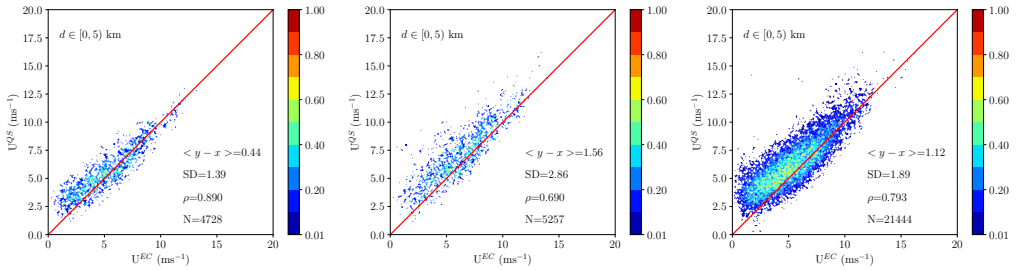


(a) CTRL (b) NC (c) NR



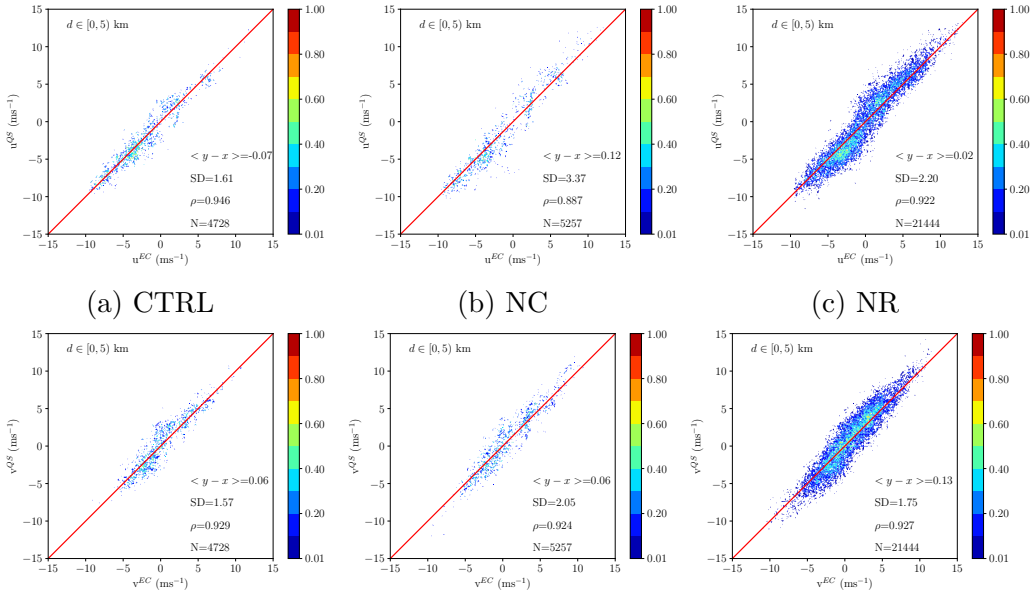
(d) CTRL (e) NC (f) NR

Figure 35: Top (bottom): 2D histogram of Seawinds-derived coastal ($d \in [0, 30)$ km) zonal (meridional) wind component u^{QS} (v^{QS}) *vs* ECMWF u^{EC} (v^{EC}) in the outer part of the swath. The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

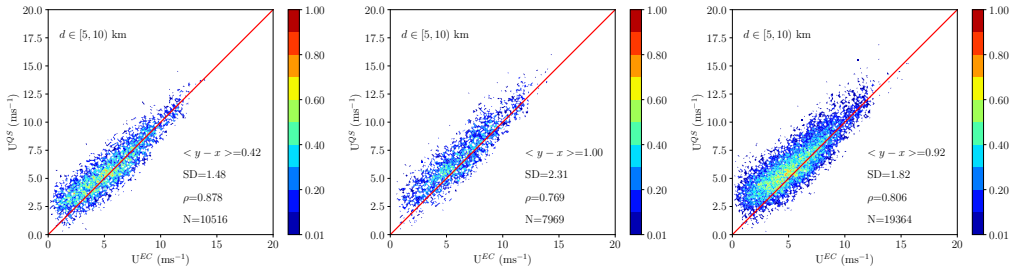
Figure 36: 2D histogram of Seawinds-derived coastal ($d \in [0, 5)$ km) wind speeds (U^{QS}) *vs* ECMWF (U^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

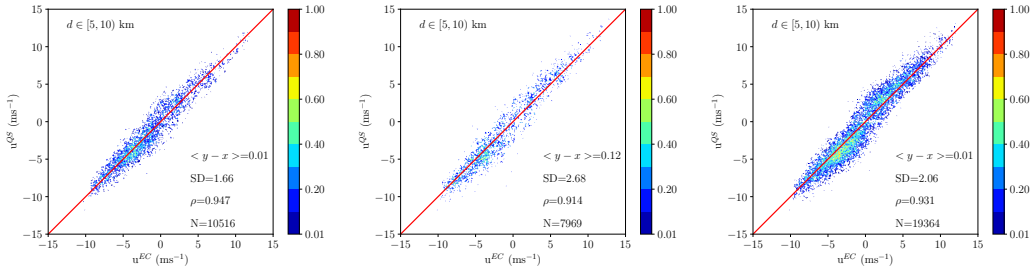
(d) CTRL (e) NC (f) NR

Figure 37: Top (bottom): 2D histogram of Seawinds-derived coastal ($d \in [0, 5)$ km) zonal (meridional) wind component u^{QS} (v^{QS}) *vs* ECMWF u^{EC} (v^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

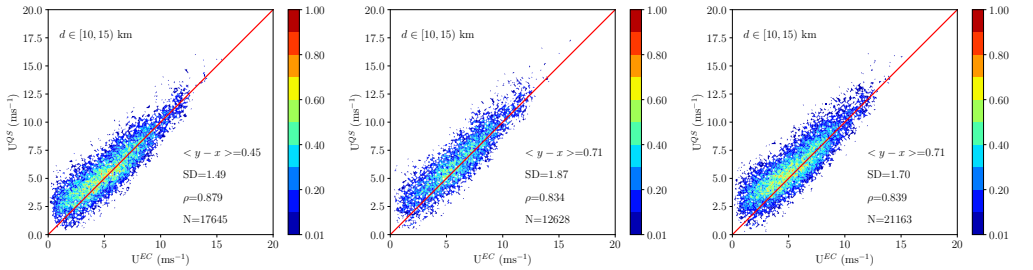
Figure 38: 2D histogram of Seawinds-derived coastal ($d \in [5, 10)$ km) wind speeds (U^{QS}) *vs* ECMWF (U^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

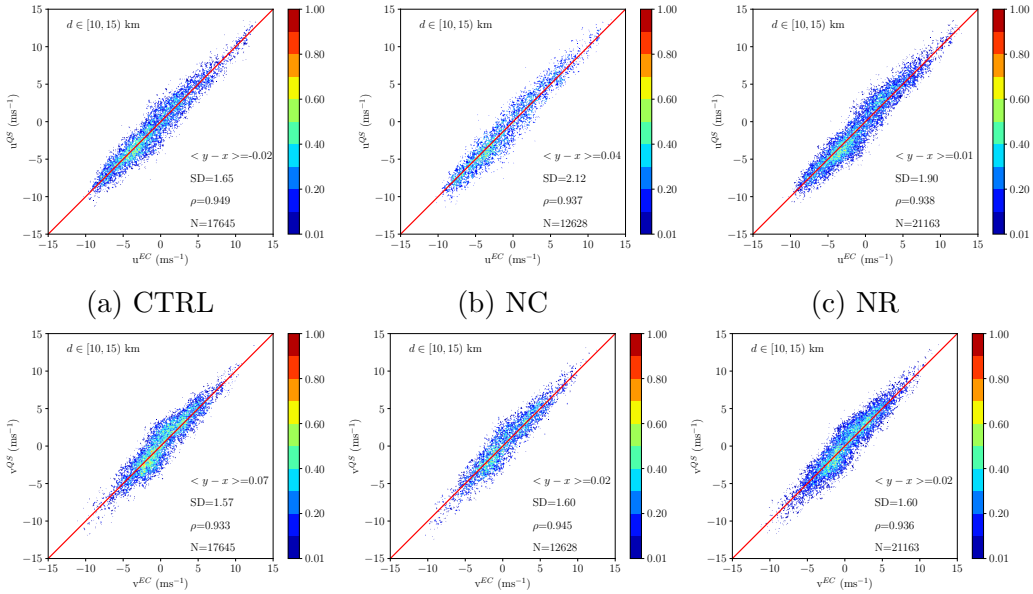
(d) CTRL (e) NC (f) NR

Figure 39: Top (bottom): 2D histogram of Seawinds-derived coastal ($d \in [5, 10)$ km) zonal (meridional) wind component u^{QS} (v^{QS}) *vs* u^{EC} (v^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

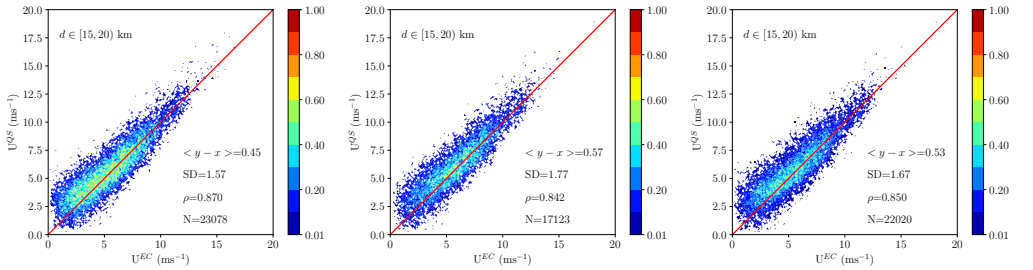
Figure 40: 2D histogram of Seawinds-derived coastal ($d \in [10, 15)$ km) wind speeds (U^{QS}) *vs* ECMWF (U^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

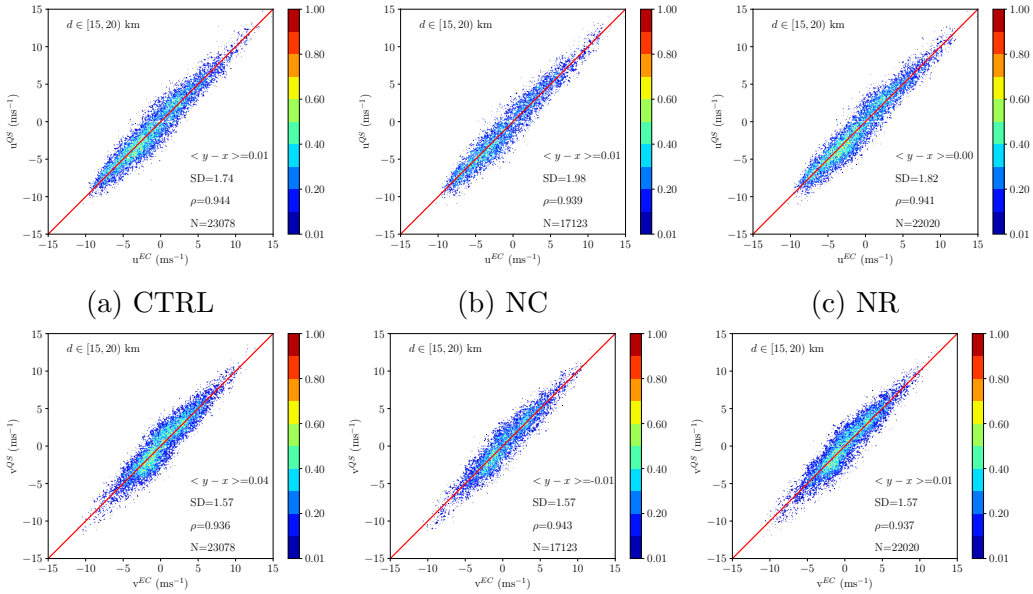
(d) CTRL (e) NC (f) NR

Figure 41: Top (bottom): 2D histogram of Seawinds-derived coastal ($d \in [10, 15)$ km) zonal (meridional) wind component u^{QS} (v^{QS}) *vs* ECMWF u^{EC} (v^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

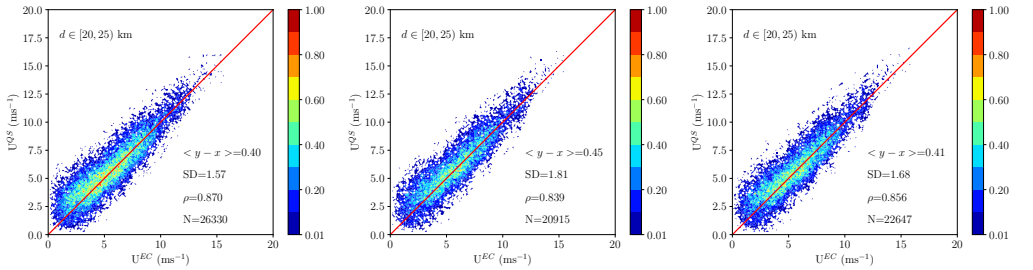
Figure 42: 2D histogram of Seawinds-derived coastal ($d \in [15, 20)$ km) wind speeds (U^{QS}) *vs* ECMWF (U^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

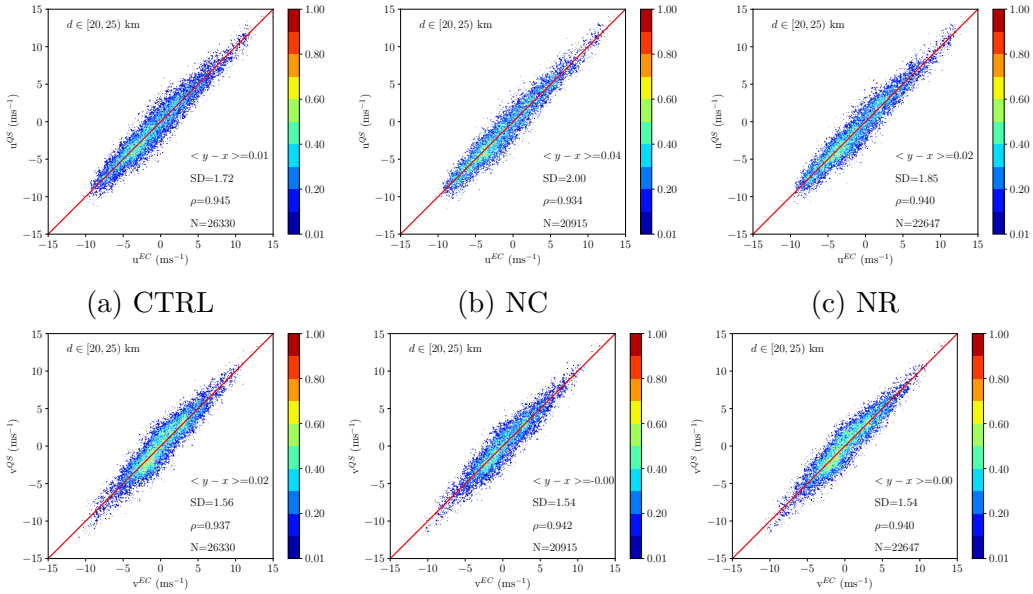
(d) CTRL (e) NC (f) NR

Figure 43: Top (bottom): 2D histogram of Seawinds-derived coastal ($d \in [15, 20)$ km) zonal (meridional) wind component u^{QS} (v^{QS}) *vs* ECMWF u^{EC} (v^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

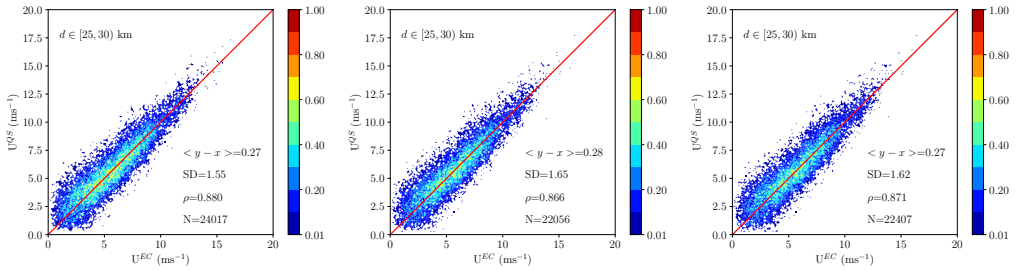
Figure 44: 2D histogram of Seawinds-derived coastal ($d \in [20, 25)$ km) wind speeds (U^{QS}) *vs* ECMWF (U^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

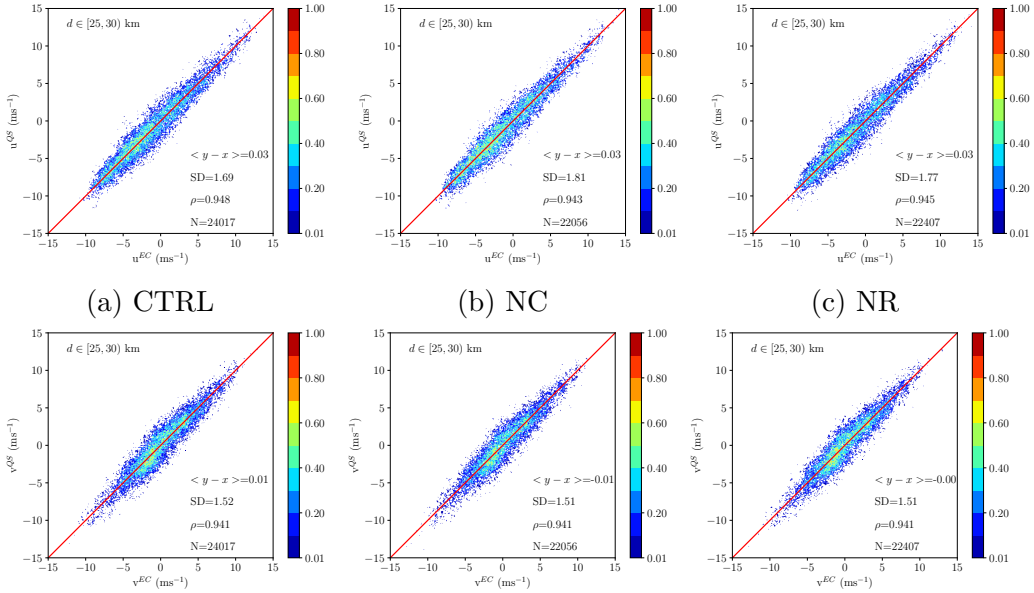
(d) CTRL (e) NC (f) NR

Figure 45: Top (bottom): 2D histogram of Seawinds-derived coastal ($d \in [20, 25)$ km) zonal (meridional) wind component u^{QS} (v^{QS}) *vs* ECMWF u^{EC} (v^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

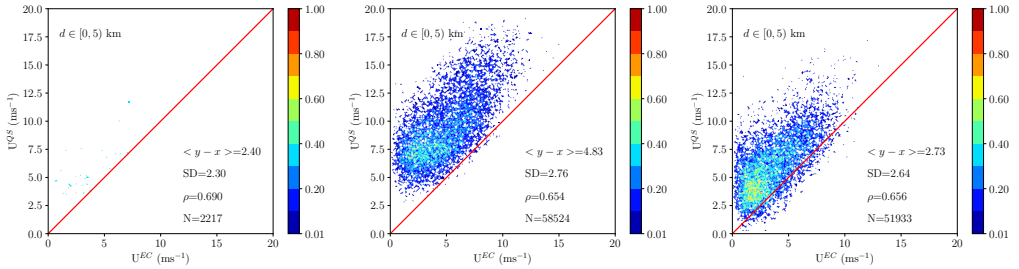
Figure 46: 2D histogram of Seawinds-derived coastal ($d \in [25, 30)$ km) wind speeds (U^{QS}) *vs* ECMWF (U^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

(d) CTRL (e) NC (f) NR

Figure 47: Top (bottom): 2D histogram of Seawinds-derived coastal ($d \in [25, 30)$ km) zonal (meridional) wind component u^{QS} (v^{QS}) *vs* ECMWF u^{EC} (v^{EC}). The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.



(a) CTRL (b) NC (c) NR

Figure 48: 2D histogram of Seawinds-derived coastal ($d \in [0, 5)$ km) wind speeds (U^{QS}) *vs* ECMWF (U^{EC}) for which the flag MLE is raised. The experiment is indicated in the caption. The histogram is normalized to the maximum value. The bias ($< y - x >$), standard deviation (SD), correlation coefficient (ρ) and total number of samples (N) are reported in the panel.

winds, with a consequent further increase in coastal sampling. This work is left for future studies.

Finally, there is no evidence of important biases in the retrieved wind directions. The overall bias of the NR coastal winds is less than 1^o compared to ECMWF. No figure is shown for the sake of brevity. It is in line with the CTRL biases, which amount to $\approx 0.5^o$.

6 Conclusions and future work

This document presents the results of the EUMETSAT OSI-SAF Visiting Scientist Activity (VSA) 22 02. The objectives of this activity were:

- Implement a noise regularization procedure to mitigate land contamination of SeaWinds σ_0s;
- Implement the necessary software in a more efficient programming language such as `FORTRAN` or `C` and integrate it with the processor `penwp`.
- Update `penwp` to retrieve winds from Seawinds on a grid with a spacing of 12.5 km;
- Retrieve winds in coastal areas.
- Validate the retrievals with coastal buoys.

The software used to compute f was implemented in `C` in four different flavors, depending on the method to compute SRF and the resolution of LSM. The four versions were compared both in terms of accuracy and computational efficiency. For what concerns accuracy, the main differences are due to the SRF model, whether it is analytical or derived from the LUT of parameterized SRFs provided by Prof. Dave Long of BYU. However, these differences never exceed 3%, which is tolerable for the purposes of this study. For what concerns computational efficiency, the best rate obtained is of 1 orbit/hour, and was obtained with the LUT-based code, with an LSM spacing of 1.1 km at the equator.
The software tool `seawinds_l1b_bufr` was updated to handle a WVC grid spacing of 12.5 km. Furthermore, a buffer area was implemented to solve the Seawinds sampling problems at 12.5 km. A sensitivity analysis was performed to select the most convenient buffer area size, which was set to 2.5 km (WVC size of 17.5 km).
A σ_0 correction scheme based on noise regularization was successfully implemented. The effects of this methodology were evaluated at both σ_0 and wind levels.
Three retrieval experiments were set up to assess the impact of noise regularization on the coastal winds:

- A control experiment (CTRL), which represents the state-of-the-art at OSI-SAF. In particular, all acquisitions with $f > 0.02$ are discarded and no σ_0 correction is applied

- An experiment with no σ_0 corrections where all acquisitions with $f > 0.5$ are discarded (NC);
- An experiment with noise regularization where all acquisitions with $f > 0.5$ are discarded (NR);

The comparison between NR and CTRL was mainly aimed at assessing the improvement of coastal sampling, while the comparison between NR and NC helped to understand the effectiveness of noise regularization. The results show that

- The sampling rate improves when noise regularization is applied, both compared to CTRL (+30%) and NC (+50%), especially within 10 km to the coast (≈300%). From the analysis shown here, it emerges that the sampling increases more in the outer than in the sweet and nadir portions of the swath. This aspect is not yet uderstood and is left future studies;
- The overall bias of retrieved coastal winds w.r.t. ECMWF increases when noise regularization is applied. This is mainly due to the increased sampling within 5 km, whose biases w.r.t. ECMWF are likely due to model biases. However, the possibility that noise regularization under corrects σ_0s must be considered. This is left for the future;
- Noise regularization reduces the biases of MLE flagged winds;

Validation against coastal buoys and winds derived from Synthetic Aperture Radars (SARs) is left for the future. However, the validation of coastal winds is a complex problem. In fact, buoys measure punctual winds on a time interval (typically 10 min), while scatterometers measure instantaneous winds that are representative of a much larger area. Problems related to representativeness and the role of orography must be carefully taken into account. Validation against SAR-derived winds raises other questions. The first one concerns the validation of SAR-derived winds themselves; in fact, wind retrieval from SAR is an underdetermined problem, and most of the concerns are related to the retrieval of wind direction; furthermore, submesoscale SAR wind patterns have not yet been extensively studied, and many features could be artificial.
A deeper analysis of the wind distributions w.r.t. to the direction is also suitable. For this purpose, a much larger dataset is recommended.
Part of the software used in the entire retrieval process is still written in an inefficient programming language (`python`). It is recommended that they be implemented in a more efficient language such as `FORTRAN` or `C`.

The results presented here suggest that a finer adjustment of the MLE threshold values could further improve coastal sampling. This aspect is also left for the future.
Finally, a comparison with the NASA JPL Seawinds-derived product 4.1 could be helpful to highlight the limits and advantages of noise regularization.

References

[1] M. P. Owen and D. G. Long, "Land-contamination compensation for quikscat near-coastal wind retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 839–850, 2009.

[2] R. D. Lindsley, C. Anderson, J. Figa-Saldaña, and D. G. Long, "A parameterized ascat measurement spatial response function," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4570–4579, 2016.

[3] R. D. Lindsley, J. R. Blodgett, and D. G. Long, "Analysis and validation of high-resolution wind from ascat," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 5699–5711, 2016.

[4] A. G. Fore, B. W. Stiles, P. T. Strub, and R. D. West, "Quikscat climatological data record: Land contamination flagging and correction," *Remote Sensing*, vol. 14, no. 10, 2022.

[5] "https://podaac.jpl.nasa.gov/dataset/qscat_level_1b_v2."

[6] J. Vogelzang and A. Stoffelen, "Ascat land correction, report for the eumetsat ocean and sea ice saf," tech. rep., Koninklijk Nederlands Meteorologisch Instituut, 2022. SAF/OSI/CDOP3/KNMI/TEC/TN/384.

[7] G. Grieco, M. Portabella, J. Vogelzang, V. A., and S. A., "Quikscat normalized radar cross section noise characterization for coastal wind field retrieval," tech. rep., Istituto di Scienze Marine (ISMAR-CNR), 2021. OSI-SAF VS Technical Report # OSI-SAF 20-03.

[8] G. Grieco, A. Stoffelen, A. Verhoef, J. Vogelzang, and M. Portabella, "Analysis of data-derived seawinds normalized radar cross-section noise," *Remote Sensing*, vol. 14, no. 21, 2022.

[9] G. Grieco, M. Portabella, J. Vogelzang, V. A., and S. A., "Initial development of pencil-beam scatterometer coastal processing," tech. rep., Barcelona Expert Center (BEC ICM-CSIC), 2020. OSI-SAF VS Technical Report # OSI-SAF 20-01.

[10] M. W. Spencer, C. Wu, and D. G. Long, "Improved resolution backscatter measurements with the seawinds pencil-beam scatterometer," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 1, pp. 89–104, 2000.

[11] R. E. Fischer, "Standard deviation of scatterometer measurements from space," *IEEE Transactions on Geoscience Electronics*, vol. 10, no. 2, pp. 106–113, 1972.

[12] M. Belmonte Rivas and A. Stoffelen, "Characterizing era-interim and era5 surface wind biases using ascat," *Ocean Science*, vol. 15, no. 3, pp. 831–852, 2019.