

## CMEMS Service Evolution 21-SE-CALL1



---

# QA best practices and protocols on QC for radial and total HF radar data (D3.1)



---

**March 6th, 2017**



## FOREWORD

This document is the deliverable D.3.1 from INCREASE WP3. In the last months, there has been intense exchange within the European HFR experts for the definition of standards QA and QC procedures. This progress has been achieved in collaboration with JERICO-Next project and has been shared and supported by the HF radar community through INCREASE project efforts (INCREASE HFR Experts Workshop, September 2016). In this context, WP3 will make a step forward, focusing on the practical aspects linked to their implementation to real data. The main goal of this deliverable is to provide a demonstration on the application of these procedures, which will be the basis for the development of INCREASE HFR basic products (M3.1).



## Abbreviations list

APM: Antenna Pattern Measurement  
BF: Beam Forming  
BoB: Bay of Biscay  
CMEMS: Copernicus Marine Environment Monitoring Service  
CODAR: Coastal Ocean Dynamics Application Radar  
DF: Direction Finding  
GDOP: Geometric Dilution Of Precision  
GoM: Gulf of Manfredonia  
HFR: High Frequency Radar  
JERICO-NEXT: Joint European Research Infrastructure network for Coastal Observatory – Novel European eXpertise for coastal observaTories  
NaN: Not a Number  
OMA: Open-boundary Modal Analysis  
QA/QC: Quality Assessment/Quality Control  
QARTOD: Quality Assurance/Quality Control of Real-Time Oceanographic Data  
SDN: SeaDataNet  
SNR: Signal to Noise Ratios  
TAC: Thematic Assembly Centres  
WERA: WavE Radar  
WP: Work Package



## Index

<b>Abbreviations list</b> .....	<b>3</b>
<b>1 List of Figures</b> .....	<b>6</b>
<b>2 Contributing Experts</b> .....	<b>9</b>
<b>3 Introduction</b> .....	<b>10</b>
3.1 Objectives of INCREASE project .....	10
3.2 Objectives of this report.....	10
<b>4 QA recommendations</b> .....	<b>11</b>
<b>5 Data, metadata and QC standards</b> .....	<b>17</b>
5.1.1 QC tests for radial and total data .....	18
5.1.2 Demonstration on QC tests application to two real cases.....	19
<b>6 Implementation of QC tests to radial data</b> .....	<b>23</b>
6.1 Syntax Test .....	23
6.1.1 Definition and computation.....	23
6.2 Over-water .....	23
6.2.1 Definition and computation.....	23
6.2.2 Application .....	24
6.3 Velocity threshold .....	25
6.3.1 Definition and computation.....	25
6.3.2 Application .....	25
6.4 Variance threshold .....	29
6.4.1 Definition and computation.....	29
6.4.2 Application .....	30
6.5 Average radial bearing .....	32
6.5.1 Definition and computation.....	32
6.5.2 Application .....	33
6.6 Median filter .....	34
6.6.1 Definition and computation.....	34
6.6.2 Application .....	35



<b>7</b>	<b>Implementation of QC tests to total data .....</b>	<b>39</b>
7.1	Syntax Test .....	39
7.1.1	Definition and computation.....	39
7.2	Data Density threshold.....	39
7.2.1	Definition and computation.....	39
7.2.2	Application .....	40
7.3	Velocity threshold .....	41
7.3.1	Definition and computation.....	41
7.3.2	Application .....	41
7.4	Variance threshold .....	44
7.4.1	Definition and computation.....	44
7.4.2	Application .....	44
7.5	Balance of contributing radials.....	47
7.5.1	Definition and computation.....	47
7.5.2	Application .....	47
7.6	GDOP Threshold .....	49
7.6.1	Definition and computation.....	49
7.6.2	Application .....	49
<b>8</b>	<b>Conclusions.....</b>	<b>54</b>
<b>9</b>	<b>Bibliography .....</b>	<b>55</b>



## 1 List of Figures

Figure 1 – Trajectory (in green) of the boat carrying the transponder used to calibrate the antenna of Matxitxako (Basque County HFR network in the SE Bay of Biscay) in September 2014. Yellow marker shows the position of the receiving antenna (courtesy of QUALITAS).....	12
Figure 2 – Example of the antenna patterns measured for Gulf of Manfredonia HFR network in one APM campaign. Crossed Loop antenna patterns are shown in the upper 4 panels, monopole antenna patterns are shown in the 4 lower panels.....	13
Figure 3 – Example of the scheduled maintenance procedures for the 2-site HFR network in the SE Bay of Biscay. Minimum recommended frequency for APM is one every two years. ....	14
Figure 4 – Example of a total velocity map generated by the GoM network. The network was located in the Southern Adriatic Sea, in the Gulf of Manfredonia and it is composed by four sites: Vieste (VIES), Pugnochiuso (PUGN), Mattinatella (MATT) and Manfredonia (MANF). The four HFRs have been operating at a frequency of 25 MHz.....	20
Figure 5 – Example of a total velocity map generated by the BoB network. The network is located in the South-eastern BoB and it is composed by two sites: Higer (HIGE) and Matxitxako (MATX). The two HFRs operate at a frequency of 4.5 MHz. ....	21
Figure 6 – Example of radial velocity data (a) and related over-water QC variable (b). This data has been measured by the CNR-ISMAR HFR network in the Ligurian Sea (TINO station) on 7 <sup>th</sup> November 2016. ....	24
Figure 7 – Evolution in time of the maximum radial velocity during year 2014 for MATX station (a), belonging to the BoB HFR network, and for the MATT station (b), belonging to the GoM HFR network.....	26
Figure 8 – Sensitivity to the maximum radial velocity threshold. The number of times the maximum radial velocity threshold labelled as “bad data” the original velocity field is given in the vertical (z) axis, normalized on the number of times the specific cell has content (i.e. is not a NaN value). The horizontal plane (axes x, y) provide the range and bearing of the radial data cells. The colour patches of the surfaces report the average differences between the radial velocities and the threshold for each spatial cell. Panels (a) and (c) refer to MATX site using 0.6 m/s and 1 m/s as threshold respectively. Panels (b) and (d) refer to MATT site using 0.6 m/s and 1 m/s as threshold respectively. ....	27
Figure 9 – Evolution in time (December 2014) of the average radial velocity module with and without the application of different velocity thresholds for MATX (a) and MATT (b) sites. ....	29
Figure 10 – Evolution in time of the maximum temporal variance during year 2014 for MATX station (a), belonging to the BoB HFR network, and for the MATT station (b), belonging to the GoM HFR network.....	30
Figure 11 – Sensitivity to the maximum temporal variance threshold. The number of times the maximum temporal variance threshold labelled as “bad data” the original velocity field is given in the vertical (z) axis, normalized on the number of times the specific cell has content (i.e. is not a NaN value). The horizontal plane (axes x, y) provide the range and bearing of the radial data cells. The color patches of the surfaces report the average differences between the original temporal variances and the threshold for each spatial cell. Panels (a) and (c) refer to MATX site using 0.15 m <sup>2</sup> /s <sup>2</sup> and 0.35 m <sup>2</sup> /s <sup>2</sup> as threshold respectively. Panels (b) and (d) refer to MATT site using 0.15 m <sup>2</sup> /s <sup>2</sup> and 0.35 m <sup>2</sup> /s <sup>2</sup> as threshold respectively.....	31



Figure 12 – Evolution in time (December 2014) of the average radial velocity with and without the application of different variance thresholds for MATX (a) and MATT (b) sites. .... 32

Figure 13 – Evolution in time of the average bearing of radial velocity vectors during year 2014 for MATX station (a), belonging to the BoB HFR network, and for the MATT station (b), belonging to the GoM HFR network. .... 33

Figure 14 – (a) Time series of the total number of valid radial data, maximum and mean range and mean bearing angle for Higer and Matxixako antennas. The three arrows in the lower panel show the dates of the antenna reparation, and the two APMs, respectively. (b) Total and radial fields for June 1, 23:00 during the period of malfunctioning (before the antenna reparation). (c) Total and radial fields for June 2, 14:00 during the period of normal operation (after the antenna reparation) ..... 34

Figure 15 – Sensitivity to the median filter thresholds. The number of times the median filter labelled as “bad data” the original velocity field is given in the vertical (z) axis, normalized on the number of times the specific cell has content (i.e. is not a NaN value). The horizontal plane (axes x, y) provides the range and bearing of the radial data cells. The color patches of the surfaces report the average differences between the original velocities and the median values for each spatial cell. Panels (a) and (c) refer to MATX site using  $RCLim = 10\text{ km}$ ,  $AngLim = 30^\circ$ ,  $CurLim = 0.5\text{ m/s}$  and  $RCLim = 5\text{ km}$ ,  $AngLim = 10^\circ$ ,  $CurLim = 0.02\text{ m/s}$  respectively. Panels (b) and (d) refer to MATT site using  $RCLim = 10\text{ km}$ ,  $AngLim = 30^\circ$ ,  $CurLim = 0.5\text{ m/s}$  and  $RCLim = 5\text{ km}$ ,  $AngLim = 10^\circ$ ,  $CurLim = 0.02\text{ m/s}$  respectively. .... 36

Figure 16 – Radial velocity maps without and with applying the median filter for MATX and MATT sites. Panels (a), (c) and (e) refer to MATX site using no median filter,  $RCLim = 10\text{ km}$ ,  $AngLim = 30^\circ$ ,  $CurLim = 0.5\text{ m/s}$  and  $RCLim = 5\text{ km}$ ,  $AngLim = 10^\circ$ ,  $CurLim = 0.02\text{ m/s}$  respectively. Panels (b), (d) and (f) refer to MATT site using no median filter,  $RCLim = 10\text{ km}$ ,  $AngLim = 30^\circ$ ,  $CurLim = 0.5\text{ m/s}$  and  $RCLim = 5\text{ km}$ ,  $AngLim = 10^\circ$ ,  $CurLim = 0.02\text{ m/s}$  respectively. .... 38

Figure 17 – Evolution in time (December 2014) of the average radial velocity with and without the application of different median filter thresholds for MATX (a) and MATT (b) sites. . 39

Figure 18 – Histogram of the number of radials used for total combination on the BoB HFR system in 2014. .... 40

Figure 19 – (a) Spatial distribution of the mean number of radials used for total combination. (b) Standard deviation with respect to the mean values. .... 41

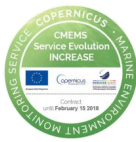
Figure 20 – Evolution in time of the maximum total velocity during year 2014 for BoB network (a) and for GoM network (b). .... 42

Figure 21 – Sensitivity to the maximum total velocity threshold. The number of times the maximum total velocity threshold labelled as “bad data” the original velocity field is given in the vertical (z) axis, normalized on the number of times the specific cell has content (i.e. is not a NaN value). The horizontal plane (axes x, y) provides the latitude and longitude of the total data cells. The colour patches of the surfaces report the average differences between the total velocities and the threshold for each spatial cell. Panels (a) and (c) refer to BoB site using 0.6 m/s and 1 m/s as threshold respectively. Panels (b) and (d) refer to GoM site using 0.6 m/s and 1 m/s as threshold respectively. .... 43

Figure 22 – Evolution in time (December 2014) of the average total velocity with and without the application of different velocity thresholds for BoB (a) and GoM (b) networks. .... 44

Figure 23 – Evolution in time of the maximum variance during year 2014 for BoB (a) and GoM (b) networks. .... 45

Figure 24 – Sensitivity to the maximum variance threshold. The number of times the maximum variance threshold labelled as “bad data” the original velocity field is given in the vertical (z) axis, normalized on the number of times the specific cell has content (i.e. is not a



NaN value). The horizontal plane (axes  $x, y$ ) provides the latitude and longitude of the total data cells. The colour patches of the surfaces report the average differences between the original variances and the threshold for each spatial cell. Panels (a) and (c) refer to BoB network using  $0.15 \text{ m}^2/\text{s}^2$  and  $0.35 \text{ m}^2/\text{s}^2$  as threshold respectively. Panels (b) and (d) refer to GoM network using  $0.15 \text{ m}^2/\text{s}^2$  and  $0.35 \text{ m}^2/\text{s}^2$  as threshold respectively..... 46

Figure 25 – Evolution in time (December 2014) of the average total velocity with and without the application of different variance thresholds for BoB (a) and GoM (b) networks. .... 47

Figure 26 – Histogram of the percent of radials from MATX and HIGE sites respectively, used for total combination ..... 48

Figure 27 – (a) Spatial distribution of the percent of number of radials coming from the different contributing sites. (b) Standard deviation with respect to the mean values. .... 48

Figure 28 – Evolution in time of the maximum GDOP and related maximum and minimum incidence angles during year 2014 for BoB (a) and GoM (b) networks. .... 50

Figure 29 – Sensitivity to the maximum GDOP threshold. The number of times the maximum GDOP threshold labelled as “bad data” the original velocity field is given in the vertical ( $z$ ) axis, normalized on the number of times the specific cell has content (i.e. is not a NaN value). The horizontal plane (axes  $x, y$ ) provide the latitude and longitude of the total data cells. The colour patches of the surfaces report the average differences between the original GDOP and the threshold for each spatial cell. Panels (a) and (c) refer to BoB network using 2.20 and 1.41 as threshold respectively. Panels (b) and (d) refer to GoM network using 2.20 and 1.41 as threshold respectively. .... 51

Figure 30 – Total velocity maps without and with applying the GDOP thresholds for BoB and GoM networks. Panels (a), (c) and (e) refer to BoB network using no GDOP thresholding, 2.20 and 1.41 as threshold respectively. Panels (b), (d) and (f) refer to GoM network using no GDOP thresholding, 2.20 and 1.41 as threshold respectively. . 52

Figure 31 – Evolution in time of the number of velocity vectors present in the total velocity field with and without applying GDOP thresholds for BoB (a) and GoM (b) networks. .... 53

Figure 32 – Evolution in time (December 2014) of the average total velocity with and without the application of different GDOP thresholds for BoB (a) and GoM (b) networks. .... 53





## 2 Contributing Experts

- L. Corgnati (CNR-ISMAR, EuroGOOS HFR Task Team)
- A. Rubio (AZTI, INCREASE TEAM)
- C. Mantovani (CNR-ISMAR, EuroGOOS HFR Task Team)
- J.L. Asensio (AZTI, INCREASE TEAM)
- J. Mader (AZTI, INCREASE TEAM)
- A. Novellino (ETT, INCREASE TEAM)
- M. Alba (ETT, INCREASE TEAM)
- A. Griffa (CNR-ISMAR, EuroGOOS HFR Task Team)

In addition, C. Quentin (CNRS, Mediterranean Institute of Oceanography) was asked to review this document and contribute to the contents of this deliverable in relation to the application of the QC to WERA systems.

## 3 Introduction

The accurate monitoring of ocean surface transport, which is inherently chaotic and depends on the details of the surface velocity field at several scales, is key for the effective integrated management of coastal areas, where many human activities concentrate. This has been the main driver for the growth of coastal observatories along the global ocean coasts. Among the different measuring systems, coastal High Frequency Radar (HFR) is the unique technology that offers the means to map ocean surface currents over wide areas and temporal resolution. The European HFR systems are playing an increasing role in the overall operational oceanography marine services. Their inclusion into CMEMS is crucial to ensure the improved management of several related key issues as Marine Safety, Marine Resources, Coastal & Marine Environment, Weather, Climate & Seasonal Forecast.

### 3.1 Objectives of INCREASE project

**INCREASE** will set the necessary developments towards the integration of the existing European HFR operational systems into the CMEMS, following **four main objectives**:

- (i) Provide HFR quality controlled real-time surface currents and key derived products.
- (ii) Set the basis for the management of historical data and methodologies for advanced delayed mode quality-control techniques.
- (iii) Boost the use of HFR data for improving CMEMS numerical modelling systems
- (iv) Enable an HFR European operational node to ensure the link with operational CMEMS.

After the fulfilment of WP1 (definition of data and metadata formats), in WP3 the main objective will be to set the recommendations and methodologies to help the adoption of these standards, including Quality Assessment/Quality Control (hereinafter QA/QC) procedures, by the HFR community.

### 3.2 Objectives of this report

This report is the main deliverable for INCREASE WP3 and focuses on the development of QA/QC operational basic data, which will be in turn the basis for the development of additional data products.

**QA is an important component to ensure good quality of the operational data and the main QA procedures are today quite well defined. They will not be further investigated here, just quickly described and enriched with recommendations on existing protocols to follow. Concerning QC tests, and based on the discussion engaged in the INCREASE HFR Experts Workshop, the focus of this document is to provide basic recommendations for the correct application of the QC tests identified by the community to be necessary for ensuring the good quality of real-time data. Most of these tests are ranked as mandatory for the future European HF radar metadata and data model, while some are still under discussion in order to be tuned according to the community needs.**



## 4 QA recommendations

QA means ensuring that the measurement process is taking place in the best available conditions, knowing exactly the state of the system and all the parameters that can affect the quality of the measurements time by time and trying to optimize them. QA involves processes that are mostly employed with hardware. HFRs are land-based remote sensing platforms, which, in comparison with other ocean observational techniques, allow the operations of maintenance and restitution of hardware components be relatively easy and less costly.

QA procedures strictly depend on the type of antenna processing performed by the device. The HFR systems belong to two categories, characterized by different antenna processing: Direction Finding (DF) and Phased Array (PA) antennas. DF refers to receiving antennas and can be implemented with a compact n-element directional antenna mounted on a single post, or with an array of antennas arranged in a specific geometry. PA can refer either to transmitting or receiving antennas, and use beamforming to maximize the transmitted energy in a specific direction or for the determination of the bearing angle. The simplest configuration of a Phased Array is a linear phased array whips with  $\lambda/2$  spacing, where  $\lambda$  is the transmit wavelength (Stewart and Joy, 1974). CODAR systems are the most used DF systems (Barrick et al., 1977; Fernandez and Paduan, 1996), WERA devices are the most common PA systems (Gurgel et al., 1999), even if WERA receiving antennas can be used also in DF mode.

Several aspects at different stages have to be taken into account from the beginning of the antenna installation to ensure the best performances of the equipment. Of course, a basic knowledge of radio equipment and best practices related to their usage is required. For instance, the attenuation of the HFR signal depends on the working frequency and the range of the measurement can also be reduced by high sea states, because in that case the noise floor rises up and weaker echoes become invisible. On the other hands, “long range” systems working at lower frequencies (4-5MHz) may not work properly in specific basins (e.g. Mediterranean sea) because ocean waves needed for the Bragg scattering are missing in that areas, and consequently no echo is sent back to the receiving antenna.

Also, since HFR signals propagates better through the conductive sea surface than through the ground, placing the transmitting antenna far from the sea will result in a power loss and in a smaller SNR in the received signal, again.

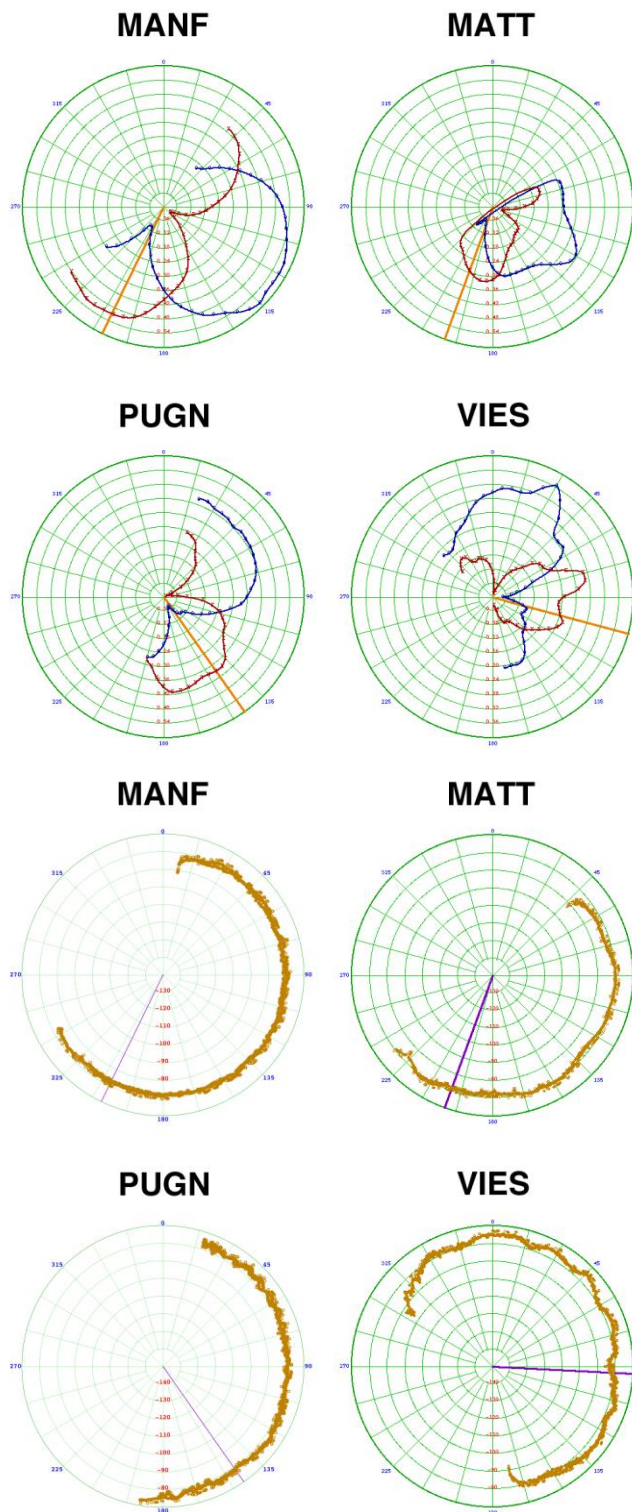
At the time of installation, a series of hardware tests should be performed. They include basic hardware tests to identify possible malfunctioning at the signal generation or amplification stage (self-generated signal noise or distortion), cable loss (due to excessive length, damages, improper connector installation), antenna performances (damages, wrong cable connection, improper grounding). Once the quality is assessed, operational parameters will be associated with such setup and then monitored during the life of the system. Any changes on these parameters will be an indicator that most likely something has changed in the hardware setup, and it shall be investigated.

Specific guidelines are provided by manufacturers regarding maximum distances from the sea and position of the antennas versus each other, as well as environmental requirements. For instance, in a common setup of a Phased Array (PA) HFR system, the separation between receiving and transmitting antennas must be at least 100 m; for long range systems a distance of 250 m is recommended; nearby metallic objects should be avoided as well as big buildings or cliff faces, because they produce distortions in the antenna pattern or unwanted reflections.

Antenna pattern measurement (APM), especially for direction finding configurations, is a key step on Quality Assurance. Direction Finding (DF) antennas need regular calibration of the receive antenna, since its response pattern can be affected by changes in the near environment, especially when these changes involve metallic objects or structures. If differences between the current beam pattern and the one adopted (either ideal or previously measured) are not taken into account, angular bias may occur as large as 35 degrees (Barrick and Lipa, 1986). Kohut and Glenn (2003) also showed that system accuracy depends on the amount of distortion of the measured pattern. It is recommended to calibrate direction finding antennas each year or when changes that could distort the pattern are suspected or detected (Kohut and Glenn, 2003). For CODAR receive antennas, the ideal or theoretical patterns of the two crossed-loop elements are cosine-dependent and oriented orthogonal to each other. Their measured or real patterns can be obtained through a calibration process in which a transponder is carried on a boat along a circular trajectory centred on the shore site. A typical APM (Antenna Pattern Measurement) scenario is represented in Figure 1, in the SE Bay of Biscay. An example of measured patterns from one APM campaign in the Gulf of Manfredonia is shown in Figure 2, where it is visible that MANF antenna pattern is close to an ideal pattern with a bearing of 200 degrees, while VIES antenna pattern is more distorted and so radial velocities maps probably have greater uncertainty (Laws et al., 2010).

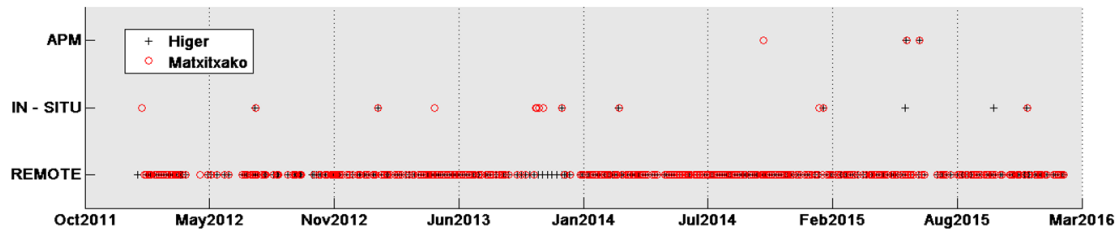


*Figure 1 – Trajectory (in green) of the boat carrying the transponder used to calibrate the antenna of Matxitxako (Basque County HFR network in the SE Bay of Biscay) in September 2014. Yellow marker shows the position of the receiving antenna (courtesy of QUALITAS).*



*Figure 2 – Example of the antenna patterns measured for Gulf of Manfredonia HFR network in one APM campaign. Crossed Loop antenna patterns are shown in the upper 4 panels, monopole antenna patterns are shown in the 4 lower panels.*

To ensure QA, operators in Europe follow the main recommendations of the HFRs manufacturers, but do not apply common procedure sets (see INCREASE D1.1, for more detail). While most part of the systems (74%) are controlled in situ periodically (every 3-6 months or yearly), other systems only benefit from sporadic visits. For several of the systems additional remote check is performed in a monthly basis or even daily. Figure 3 provides an example of site maintenance, including the frequency of remote, in situ and APM operations.



*Figure 3 – Example of the scheduled maintenance procedures for the 2-site HFR network in the SE Bay of Biscay. Minimum recommended frequency for APM is one every two years.*

Assessment of environmental noise or interferences at specific frequencies is another key step on QA. The presence of noise in the frequency bands allocated for HFRs can cause significant signal degradation and consequently measurement inaccuracy or even data loss. Noise in those frequency bands is mostly originated by human activities; however, the influence of the ionosphere is also frequent, mostly near noon or in the early afternoon, with a peak during periods of greatest sunspot activity. Atmospheric intense events, like heavy rain, and high sea status, can both increase the noise level of the signal. From the recent inventory of European HFRs, the occurrence of interferences has been reported as a significant issue. Indeed, around 30% of the 50 European HFR systems are experiencing interferences at some level (Mader et al. 2016). In the cases where a persistent interference has been reported, it has been observed mostly in the 13.5 MHz band, during the afternoon. Such interferences are avoided in some cases by shifting the system operational bandwidth, however when this workaround is not possible they represent a critical issue. Before selecting a site, the frequency spectrum in that location should be scanned for few days for detecting any persistent interference in advance.

Other guidelines for site planning take into account aspects, like energy supply, internet communications, accessibility, that are not directly related to data quality, but that are important for the overall quality service in case it is provided in operational way. Besides, risk prevention should be taken into account as well. Several recommendations can be provided to reduce the risk of damages due to severe weather conditions (wind, humidity, heat, lightning), or the action of animals (for instance rodents that can damage the cables). Site maintenance is capital to ensure a high-quality data flow so frequent remote diagnostics of hardware performance are recommended. Then, quarterly (or when any problems arises) visits to the sites are needed to check connections, cables and antenna status.



Another remarkable aspect in QA is the proper setup of the processing software. Although data can be reprocessed later, and mistakes can be amended, a careful check of the processing parameters should be performed at the installation time, and periodically repeated. For instance, crucial settings in CODAR systems have impact on the automatic selection of the first order settings, i.e. the part of the spectrum which is supposed to contain all the information on currents velocity. Such settings depend on the site, but also on the sea state, so optimal settings can slightly change for the same site within different periods of time.

As the network grows, there will be a growing need for a defined set of QA protocols. During the INCREASE HFR Experts Workshop (La Spezia, September 2016) this issue was discussed, and as a first step, it was agreed to refer to the protocols set by the experts from the Southern California Coastal Ocean Observing System ([www.sccoos.org](http://www.sccoos.org)). Their document entitled: “The Deployment & Maintenance of a High-Frequency Radar (HFR) for Ocean Surface Current Mapping: Best Practices” (U.S. IOOS, 2016a), defines a set of best practices developed from the collective experience of the operators usage of compact direction finding radar systems, specifically, the SeaSonde® family of direction-finding HFRs manufactured by CODAR Ocean Sensors, Ltd. This is also the reference document used by the US Integrated Ocean Observing System (IOOS), cited in the last manual of the “Quality Assurance of Real-Time Oceanographic Data” (QARTOD) program (see U.S. IOOS, 2016b).

The best practices discussed in this document focus mainly on the installation, operation, and maintenance of CODAR SeaSonde. Further revisions of this document are expected to include other manufacturers of HFR equipment. This is an important aspect for the European community, where 20% of the existing systems provide from other manufacturers.

The Radiowave Operators Working Group (ROWG) is also pushing towards this direction through the organization of periodic workshops since 2005, which bring together the technical and managerial personnel involved in all aspects of operating an HFR system or network within the IOOS context. The three main pillars of the ROWG Charter are:

1. To foster collaboration between new and experienced HFR operators.
2. To develop procedures governing HFR operations including:
  - Site logistics
  - Processing to component vectors
  - Total vector products and data management
3. To provide recommendations to users, developers, manufacturers, and program managers.

The participation of European experts involved in HFR operations in this forum is especially encouraged.

Finally, it is worth mentioning the ongoing work in JERICO-Next project. Indeed, Task 2.3 of JERICO-Next deals specifically with the harmonization of the different observing elements within the JERICO infrastructure network, whose specific Task 2.3.1 deals



with HFRs. Subtask 2.3.1 aims to review and discuss the state-of-the art methodologies utilized for processing and analysing the data that such systems generate, including Quality Assurance procedures, computation of error maps, etc., in order to define Best Practices. The first deliverable (D2.1 report on the status of HF-radar systems and cabled coastal observatories and Cabled Observatories, October 2016) provided an overview of the state-of-the-art methodologies utilized during the planning and installation phase of HF-radar sites, and reviews the main relevant operational aspects, applications, and quality assessment and data management issues. A devoted meeting held in Mars 2016 in San Sebastian contributed to set the basis of this work (see <http://www.jerico-ri.eu/project-information/meeting-reports/>).





## 5 Data, metadata and QC standards

While all the radars share the same principles of operation, differences in signal transmission, reception and processing yield variations in metadata, quality control metrics and spatial registration. Even within the same type, HFRs may have different spatial ranges and resolutions, depending typically on the working frequency and available bandwidth. Quality Control (QC) procedures depend on the antenna processing (Direction Finding for CODAR and WERA, or Beam Forming for phased arrays as WERA) of the specific system.

During the fruitful days of the INCREASE HFR Experts Workshop, the main issues concerning data formats and QC procedures have been discussed in order to meet the needs of both the HFR community and CMEMS operational services. In particular, the main points that have been analysed for achieving a common consensus are data format, metadata structure, QC flagging scheme and QC tests. The final goal of this task, together with the continuous discussion in the framework of JERICO-Next project, is the implementation of the European common model for data and metadata for real time HFR data. The roadmap for achieving this goal was presented in INCREASE deliverable D1.1 and the reader is referred to this document for further information. The main elements of this model are summarized as follows:

- Data Format: the HFR community agreed in the decision of producing data using netCDF-4 classic model format, in order to apply the state of the art version. For the specific purposes of the INCREASE project, the netCDF-4 classic data will be then converted in netCDF-3.6.1 by the central HFR node to be developed by WP4. This double data production will meet both HFR community needs and CMEMS IN-SITU TAC needs.

- Metadata Structure: the CMEMS IN-SITU TAC reference conventions for the metadata attributes are CF-1.6 and OceanSITES. Thus Mandatory Attributes have been defined and include attributes necessary to comply with CF-1.6 and OceanSITES conventions. For the data variables the SDN P09 vocabulary is used.

- Standards on dimension, naming, definition and syntax of coordinate variables and data variables, including QC variables derived from the established QC tests, have been defined and implemented.

- A preliminary set of mandatory and recommended QC tests which are manufacturer-independent, i.e. they do not rely on particular variables or information provided only by a specific device, have been defined for total and radial data. These sets of QC tests are the required ones for labelling the data as Level 2B (for radial velocity) and Level 3B (for total velocity) data. Please refer to Table 8 of D1.1 for the processing level definition.

- Reference QC flagging scheme to be used is the one from CMEMS IN-SITU TAC, i.e. the OceanSITES one, that is reported in Table 1. As also described in the comments of Table 1, the CMEMS IN-SITU TAC strategy for quality flagging

states that a “good data” flag is to be assigned to a data if and only if all QC tests are passed by the data. This strategy will obviously be adopted as a standard by the HFR community and it has been decided to include in each file a gridded variable reporting QC flags for each velocity vector (for radial files and for total files). In order to complement this QC flagging strategy, it has also been decided to include in each file a QC variable for each performed QC test, in order to have an overall file quality flagging and all the specific flagging for the QC tests.

*Table 1 – OceanSITES QC Flagging scheme.*

Code	Meaning	Comment
0	unknown	No QC was performed
1	good data	All QC tests passed
2	probably good data	
3	potentially correctable bad data	These data are not to be used without scientific correction or re-calibration
4	bad data	Data have failed one or more QC tests
5	-	Not used
6	-	Not used
7	nominal value	Data were not observed but reported (e.g. instrument target depth)
8	interpolated value	
9	missing value	

### 5.1.1 QC tests for radial and total data

The selected QC tests have been chosen among the ones listed in the QARTOD manual, reported in INCREASE D1.1 and after the discussion with the European experts engaged at the INCREASE HFR Experts Workshop (La Spezia, September 2016). The choice of the tests was made in order to ensure that they are manufacturer-independent, i.e. they do not rely on particular variables or information provided only by a specific device. From the work on the application of these tests, through the development of the scripts and identification of the needed variables, we realize that this is not always the case. Indeed as we will discuss later on, some of the tests are not needed for WERA PA systems. Consequently, we will need to work further on these definitions, probably changing some of the “mandatory tests” into “recommended” tests, which could be avoided by the users from WERA PA systems, when necessary.

In any case, the sets of standard tests that are further developed in the next sections have been defined both for radial and total velocity data. The final goal is to select and define the QC battery of tests required for labelling the data as Level 2B (for radial velocity) and Level 3B (for total velocity) data.



The tests chosen for radial data are:

- ✓ Syntax test
- ✓ Over-water test
- ✓ Velocity Threshold test
- ✓ Variance Threshold test
- ✓ Median Filter test
- ✓ Average Radial Bearing test

The tests chosen for total data are:

- ✓ Syntax test
- ✓ Data Density Threshold
- ✓ Balance of contributing radials
- ✓ Velocity Threshold
- ✓ Variance Threshold
- ✓ GDOP Threshold

The definition and the application guidelines of each of these tests are provided in the next sections. For some of them, HFR operators will need to select the best thresholds. Since a successful QC effort is highly dependent upon selection of the proper thresholds, this choice is not straightforward, and may require trial and error before final selections are made. These thresholds should not be determined arbitrarily, but based on historical knowledge or statistics derived from historical data. In the next sections we provide recommendations on how to choose these thresholds through the application of the listed battery of tests to two different HFR systems.

#### 5.1.2 Demonstration on QC tests application to two real cases

The processing from raw data to the final products, as well as the data formats, are different, depending on the system configuration and/or manufacturer, but also on the software used. Apart from the manufacturer's software (CODAR's Radial Suite and WERA software, SeaView software), an extended tool for HFR data processing is the HFR\_Progs toolbox, developed by the University of California and NPS (see <https://cencalarchive.org/~cocmpmb/COCMP-wiki>). HFR\_Progs toolbox, developed in Matlab®, allows different methods and error quantification, starting from radial files generated by a WERA or a CODAR system. Files are converted into Matlab® binary files and all the initial information contained in radial data is kept in the Matlab® file structure.

For this demonstration our starting point will be the CODAR radial and total data files provided by the manufacturer. This is done for practical reasons since the partners of INCREASE project operate CODAR systems and this is also the most extended technology in Europe at the moment. Please note that, although for this demonstration we are focusing in two CODAR HFRs, efforts are done to orient WERA PA and other types of HFR systems operators through the recommendations, following QARTOD indications (U.S. IOOS, 2016b).

CODAR files contain data in ASCII format and a header with information on the system configuration and processing information, which are needed to build the metadata and feed the mandatory QC tests.

In particular, throughout this document, examples of application and impact of the QC tests will be reported with reference to two different case studies:

- the Adriatic deployment of the CNR-ISMAR HFR network (GoM), deployed in the Gulf of Manfredonia in the Southern Adriatic Sea and composed by the four sites of Vieste (VIES), Pugnochiuso (PUGN), Mattinatella (MATT) and Manfredonia (MANF), which are equipped with four 25 MHz CODAR SeaSonde Systems (Corgnati et al., 2015). This network (Figure 4) has been operational in the period May 2013 – June 2015, and it was then moved in the Ligurian Sea, where is now operational.
- the Euskalmet Basque Country HFR network (BoB), deployed in the Bay of Biscay and composed by the two sites of Higer (HIGE) and Matxitxako (MATX), which operates two 4.5 MHz CODAR SeaSonde systems (Rubio et al., 2011, Figure 5);

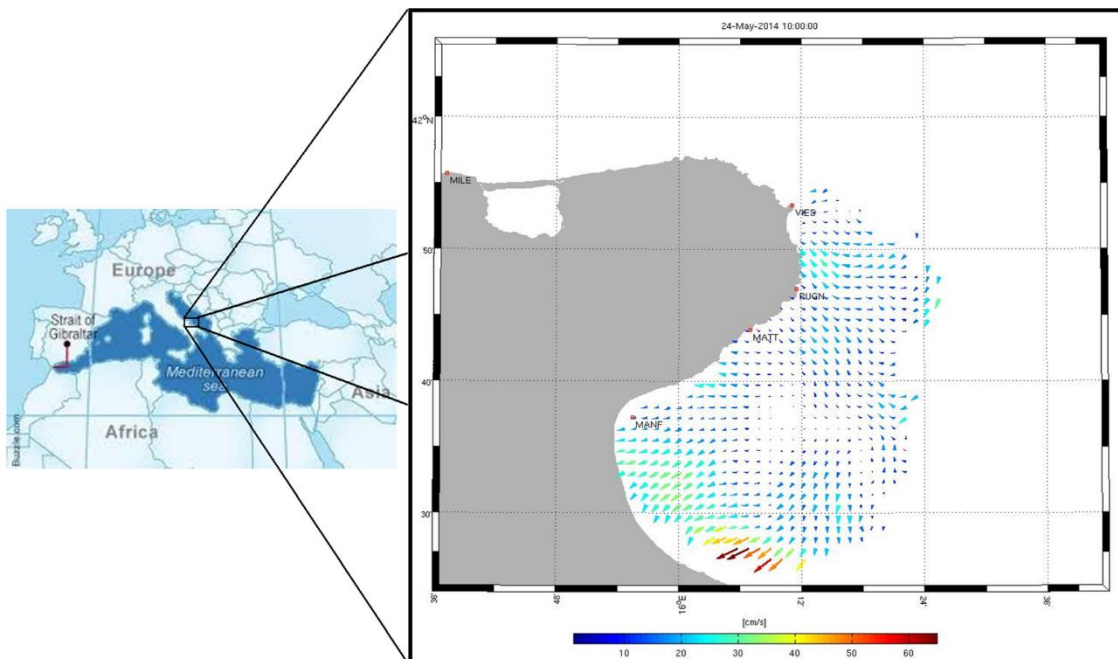


Figure 4 – Example of a total velocity map generated by the GoM network. The network was located in the Southern Adriatic Sea, in the Gulf of Manfredonia and it is composed by four sites: Vieste (VIES), Pugnochiuso (PUGN), Mattinatella (MATT) and Manfredonia (MANF). The four HFRs have been operating at a frequency of 25 MHz.

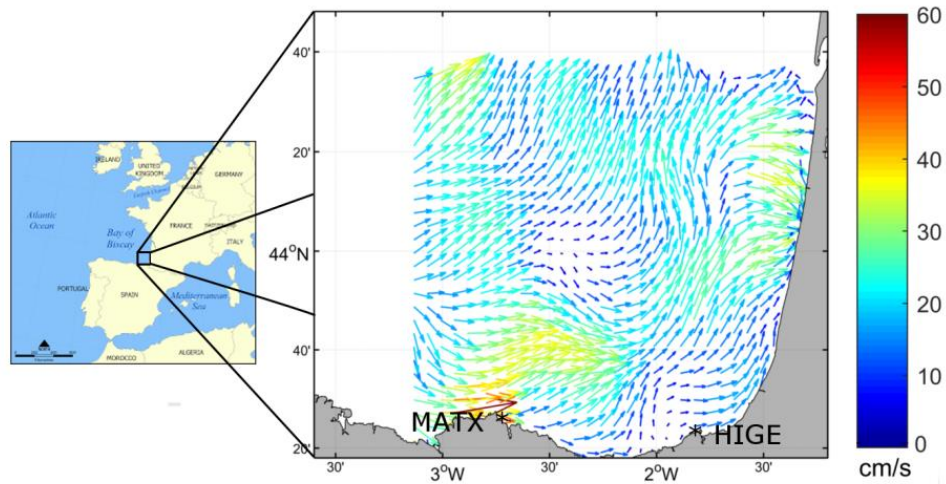


Figure 5 – Example of a total velocity map generated by the BoB network. The network is located in the South-eastern BoB and it is composed by two sites: Higer (HIGE) and Matxitxako (MATX). The two HFRs operate at a frequency of 4.5 MHz.

These two networks are characterized by different operating frequencies and provide data with different spatial resolution and coverage. The detailed technical aspects of these two networks are provided in Table 2.

Table 2 – Main characteristics of the systems used for the demonstration.

HFR NETWORK	Gulf of Manfredonia				Basque Country	
Number of SITES	4				2	
Name of sites	VIES	PUGN	MATT	MANF	MATX	HIGE
Sites lat , lon coordinates	41,89	41,78	41,73	41,62	43,45	43,38
	16,18	16,19	16,12	15,93	-2,75	-1,78
Manufacturer	CODAR				CODAR	
Type of radar	Direction Finding				Direction Finding	
Temporal resolution (minutes)	60				60	
Spatial resolution of total velocity grid (m)	1500				5000	
Transmit Frequency (MHz)	24,525	26,275			4,525	
Transmit Bandwidth (KHz)	150				40	



In addition to the differences in sampling and coverages, they are located in areas which are affected by different ocean processes.

The circulation in the area the Southern Adriatic Sea is part of the general cyclonic circulation of the Adriatic (Poulain, 2001), forced by buoyancy input and wind. A well-defined boundary current, the Western Adriatic Current (WAC), flows to the southeast on the shelf with a core centred on the 100 m isobaths. The WAC is energetic (surface speeds up to 0.5 m s<sup>-1</sup>) and with significant fluctuations especially close to the tip of the Gargano Cape, where the current tends to detach rejoining the coast south of the Gulf (Taillandier et al., 2008). The flow in the Gulf of Manfredonia (at the lee of the Cape) is characterised by high variability, with both cyclonic and anticyclonic recirculations and typical root means square (RMS) velocities of the order of 0.20 m/sec.

The primary surface circulation pattern in the BoB area presents a marked seasonal variability. A key component of this variability is the slope Iberian Poleward Current (IPC). In winter, the IPC flows eastward along the Spanish coast and northward along the French coast, affecting the upper water column and presenting velocities up to 0.7 m/s. In summer, the flow is reversed and three times weaker than in winter [-Solabarrieta et al. 2014]. Overlaid to the density-driven slope circulation, wind-induced currents are the main drivers of the surface ocean circulation in the area and can be at the origin of intense surface velocities. Mesoscale coherent structures have also been observed within the HFR footprint area (Rubio et al, 2013).

For both networks, the time period considered for the examples of application of the QC tests is the year 2014.



## 6 Implementation of QC tests to radial data

Six QC tests are being tested for radial data. All of them have been defined as mandatory for the European common data and metadata model for real-time HFR data.

### 6.1 *Syntax Test*

#### 6.1.1 Definition and computation

This test will ensure the proper formatting and the existence of all the necessary fields within the radial netCDF file. This test is performed on the netCDF files and it assesses the presence and correctness of all data and attribute fields and the correct syntax throughout the file. The reference schemes for data and metadata structures and syntax are described in the deliverable 5.13 of the JERICO-Next project (<http://www.jerico-ri.eu/project-information/deliverables/>).

Other complementary test to verify some of the fields contained in the metadata can be also applied, and the correct values should be established by the operators. This can include verification of the timestamp value (the timestamp should be coherent with the actual time for operational data and coincide with the filename and within a consistent range in any case), existence and consistent data concerning site names, site codes, time zones, site locations. For instance for EU systems the site location should be in a range defined by  $[-10 +35]^{\circ}\text{E}$  Latitude and  $[-25 70]^{\circ}\text{N}$  Longitude. The suit of verifications made by the IOOS National network on radial data can be found at QUARTOD manual (U.S. IOOS, 2016b).

The result of the test is a scalar QC variable that is inserted into the netCDF file after the test completion.

### 6.2 *Over-water*

#### 6.2.1 Definition and computation

This test labels radial vectors that lie on land with a “bad data” flag and radial vectors that lie on water with a “good data” flag. The output is a gridded QC variable, with the same dimensions of the radial velocity data variable, containing, for each cell, the flag related to the vector lying in that cell.

In CODAR data, the variable `VFLG` is used for the implementation of the test, since this variable takes values of 0 to indicate that the vector lies on water and takes the value of 128 indicating that the vector lies on land.

Although it is always present in the CODAR data format, the convenience of this variable depends on the SeaSonde Radial Site software release. It is only properly implemented for Release 7 or higher and for previous software versions it can be always hosting values equal to zero (or other values) and does not allow discriminating between vectors in and out the angular filter area.

For WERA PA systems, the angle of view of the radar is limited to 120° and on land the SNR is too low and doesn't provide any solution. Thus, the Seaview Sensing software removes all cells on land before any processing is done. In this case, the QC flag for the over-water test should be automatically produced containing all "good data" values.

It is left to the HFR operator to apply an additional masking if none of these solutions is working.

### 6.2.2 Application

Figure 6 shows an example of application of the over-water test on a radial velocity map measured by the CNR-ISMAR HFR network in the Ligurian Sea (TINO station). We have chosen this specific station (not in the GoM, nor in the BoB networks) system since it provides an example of how this test can be crucial for some installations which produce a high number of vectors on land.

Panel (a) shows a map of radial velocities, including a number of velocity vectors lying on land. The colour palette refers to velocity values. The panel (b) shows the map of over-water flags related to the velocity field of panel (a). In panel (b) the colour palette is referred to the values of the flag: 1 if the vector lies on water (blue cells), 4 if the vector lies on land (red cells).

This structure allows for having the velocity field and the mask within the same data file.

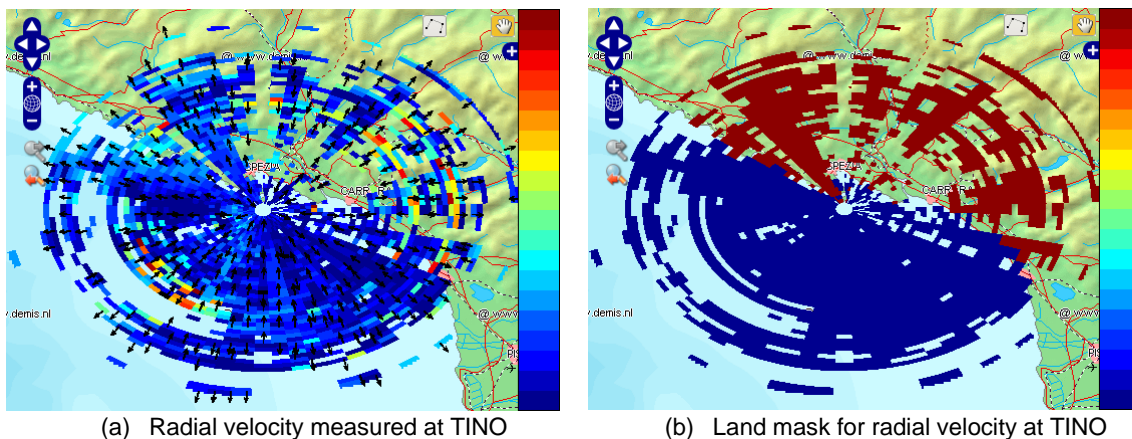


Figure 6 – Example of radial velocity data (a) and related over-water QC variable (b). This data has been measured by the CNR-ISMAR HFR network in the Ligurian Sea (TINO station) on 7<sup>th</sup> November 2016.

It has to be noticed that the presence of velocity vectors on land can be significant in particular situations, like for instance the one reported in Figure 6, where the radar antenna is installed on top of a little island surrounded by the sea but with land separated by the island just by a small portion of water.





## 6.3 *Velocity threshold*

### 6.3.1 Definition and computation

This test labels radial velocity vectors whose module is bigger than a maximum velocity threshold with a “bad data” flag and radial vectors whose module is smaller than the threshold with a “good data” flag. The output is a gridded QC variable with the same dimensions of the radial velocity data variable, containing, for each cell, the flag related to the vector lying in that cell.

The threshold for the implementation of this test has to be applied to the absolute value of the variable `VELO`, present in both CODAR and WERA radial LLUV formats, containing the amplitude of the radial current vector velocity (negative velocity implies 180° direction change).

The choice of the maximum radial velocity threshold has to be made carefully, since using a tight threshold can lead to the elimination of accurate data. The recommendation in this case is to examine the time-series of the maximum radial velocity of the data (during at least one complete seasonal cycle for which the proper functioning of the device is assessed) and consult the literature.

Furthermore, it is also recommended to analyse the impact of the selected threshold (or of the candidate thresholds) on the velocity fields of the dataset containing at least one complete seasonal cycle.

The aim of this filter is to identify unexpected extremely high velocities for further investigation. It reinforces the median filter (see paragraph 6.6), which also contributes to tag unexpected high values, and gives further information on the nature of the suspect values.

Examples of application of this test are provided in the next subsection for the two case study areas.

### 6.3.2 Application

With reference to BoB and GoM HFR networks, Figure 7 shows the evolution in time of the maximum radial velocity for MATX (a) and MATT (b) stations during the year 2014.

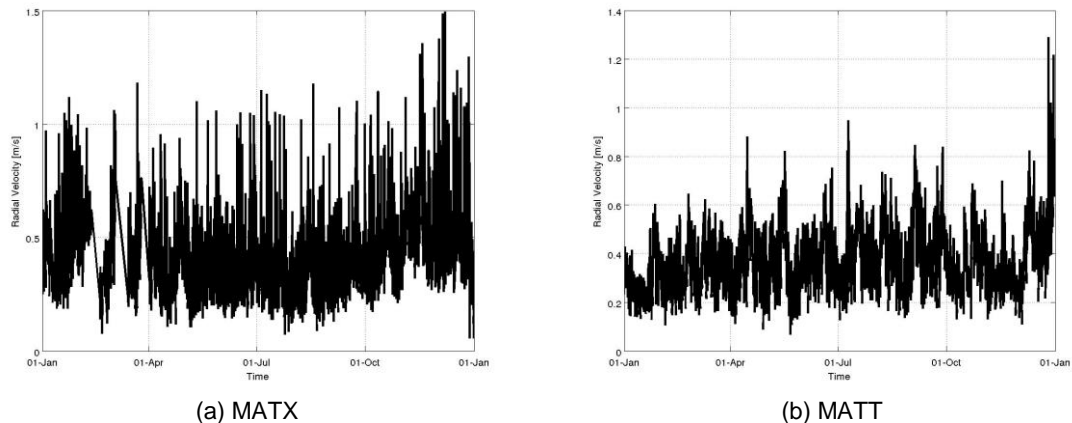


Figure 7 – Evolution in time of the maximum radial velocity during year 2014 for MATX station (a), belonging to the BoB HFR network, and for the MATT station (b), belonging to the GoM HFR network.

The two graphs are a precious aid in determining the range where maximum velocities lie: Figure 7 shows that (for year 2014) in MATX station the maximum velocity mainly ranges in [0.4 – 0.7] m/s, while in MATT station the maximum velocity normally lies in the range [0.2 – 0.6] m/s. On these bases, it is now possible to define a maximum velocity threshold compatible with the velocity values measured by the stations of the network. For this example let us consider two different thresholds, 0.6 m/s and 1 m/s, in order to verify the impact they have on the velocity fields. This approach is recommended, since it allows for comparing the effects of different thresholding strategies.

Figure 8 shows (on z axis) the number of times the maximum velocity threshold labelled as “bad data” the original velocity fields in MATX and MATT stations using the two thresholds listed above. The z-axis values are normalized on the number of times the specific cell has content (i.e. is not a NaN value). The colour patches of the surfaces report the average differences between the radial velocity and the threshold for each spatial cell. The X-Y axes grid is the exact range-bearing (which provides the radial direction in degrees clockwise from true North) of the original data. In this way it is possible to localize each radial velocity cell.

Figure 8(a) shows that in MATX site the velocity vectors exceeding the 0.6 m/s thresholds very often lie at the farthest border of the coverage. Indeed, for this installation values over 0.6 m/s are quite frequent and realistic in this area, linked to the intensification of the shelf-slope current (see Figure 5).

When increasing the threshold to 1 m/s (Figure 8(c)), the velocity vectors are more rarely cut by the threshold and, as expected, the values exceeding the threshold lie at the farthest border of the coverage.

Figure 8(b) shows that in MATT site the threshold is active along all ranges for bearing in [100 – 130]° but very often for small differences between velocities and thresholds. If the threshold is set to 1 m/s, the test acts very rarely, as depicted in Figure 8(d).

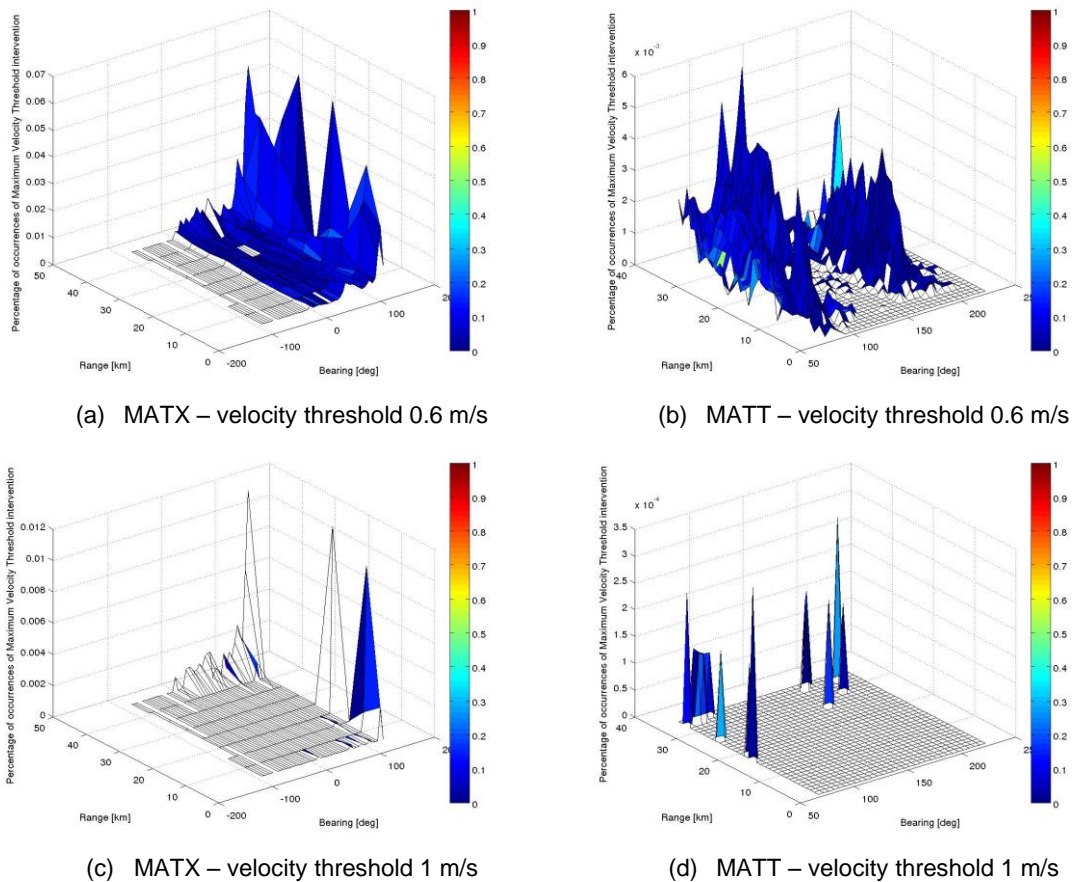


Figure 8 – Sensitivity to the maximum radial velocity threshold. The number of times the maximum radial velocity threshold labelled as “bad data” the original velocity field is given in the vertical (z) axis, normalized on the number of times the specific cell has content (i.e. is not a NaN value). The horizontal plane (axes x, y) provide the range and bearing of the radial data cells. The colour patches of the surfaces report the average differences between the radial velocities and the threshold for each spatial cell. Panels (a) and (c) refer to MATX site using 0.6 m/s and 1 m/s as threshold respectively. Panels (b) and (d) refer to MATT site using 0.6 m/s and 1 m/s as threshold respectively.

This representation allows for identifying areas of the coverage where the test is acting frequently or rarely and for small or big differences with respect to the original velocity field.

If the threshold is put to zero, we expect to see a flat graph with value 1 (filter is acting 100% of times), with colours reflecting the average velocity in each cell over the period of time considered. While the threshold increases, a non-uniform modification of the plateau is expected, indeed in some cells velocities could remain all the time below the new threshold, in some others they could remain below the new threshold many times, in some others few times. In general then rising up the threshold will result in a non-proportional modification of every cell’s peak, until a uniform zero is reached (threshold higher than all the velocities). At every step, the colour is reflecting the average difference between the exceeding velocities and the threshold. Consider two cases a) and b) for the same cell in which we have five exceeding values giving the following

five differences  $a=(1 \text{ cm/s}; 2 \text{ cm/s}; 1 \text{ cm/s}; 1 \text{ cm/s}; 20 \text{ cm/s})$  or  $b=(5 \text{ cm/s}; 5 \text{ cm/s}; 5 \text{ cm/s}; 4 \text{ cm/s}; 6 \text{ cm/s})$ . Peak is of course the same in both cases. Since the average is always  $5 \text{ cm/s}$ , also the colour will be the same in both cases, and will be hiding some information. Imagine now to increase the threshold of  $2 \text{ cm/s}$ . Now differences will be  $a=(18 \text{ cm/s})$  and  $b=(3 \text{ cm/s}; 3 \text{ cm/s}; 3 \text{ cm/s}; 2 \text{ cm/s}; 4 \text{ cm/s})$ , with average value of  $a\text{-average}=18 \text{ cm/s}$ ,  $b\text{-average}=3 \text{ cm/s}$ . The  $a$ -case graph will result in smaller peak with much higher average value (colour turning very much toward red), while the  $b$ -case graph will result in the same peak as before with a colour turning slightly toward blue. The  $a$ -case shows a more suspect value of velocity than the  $b$ -case, thus the way the peaks and the colours change in function of the threshold's values should be taken into account.

In a general way, if the test is acting often and on a wide area for small amounts (plateau with high  $z$ -values and with colours associated to small differences between threshold and original velocity), it means that a) we have unexpected extremely high velocities everywhere and very often, which is a suspect circumstance, or b) the threshold is too low and should be increased, in order to avoid these frequent small corrections.

A plateau with colours associated to high differences between threshold and original velocity means that the filter is active very frequently on wide areas and for huge amounts. This situation could indicate that the threshold is too low, thus being always active.

The presence of peaks with colours associated to small differences between threshold and original velocity indicates frequent corrections for small amounts in specific cells, or, in other words, very often we have slightly exceeding velocities. If peaks are close each other, we could suppose a) critical area of the coverage, or b) very energetic area. If peaks are isolated and randomly spaced, we could suppose spikes. In both cases the threshold could be increased, because the little difference between the velocity value and the threshold suggests that the exceeding velocities are very close to the threshold. Increasing the threshold in this case could reveal more information.

If the test is acting very frequently just in some random cells (high peaks in separate cells), or specific zones where high velocities are not expected, for huge amounts (peaks with colours associated to big differences between threshold and original velocity), the threshold could be adequate, since it corrects recurring extremely high velocity values, without affecting at all the values in the rest of the domain.

Anyway, all these situations have to be investigated by the operator before choosing the threshold, according to the local history, to his experience and to literature.

Another useful and simple check to do when choosing the threshold for the QC test, is to examine the evolution in time of the average radial velocity module with and without the application of the candidate thresholds. In this way, it is possible to have a quick overview of the effects of the test (that keeps only velocities whose absolute value is lower than the threshold) on the velocity field. Figure 9 reports this comparison.

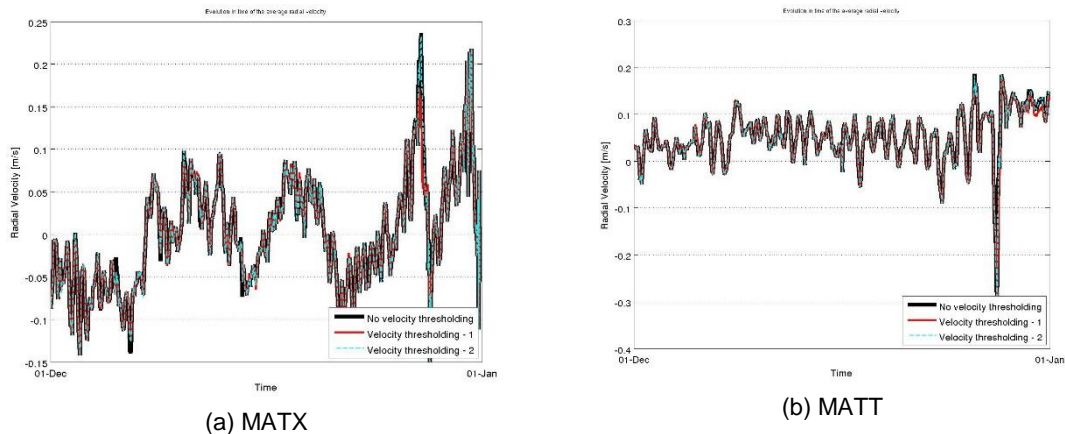


Figure 9 – Evolution in time (December 2014) of the average radial velocity module with and without the application of different velocity thresholds for MATX (a) and MATT (b) sites.

## 6.4 Variance threshold

### 6.4.1 Definition and computation

This test labels radial vectors whose temporal variance is bigger than a maximum threshold with a “bad data” flag and radial vectors whose temporal variance is smaller than the threshold with a “good data” flag. The output is a gridded QC variable with the same dimensions of the radial velocity data variable, containing, for each cell, the flag related to the vector lying in that cell.

High values of variance are mostly due to radio interference (low SNR) that are translated in Doppler lines and abnormal value in radial velocity. In DF systems, the origin of a big variance when SNR is low is also due to the wrong attribution of the direction of arrival of the radar echo. In this case, the algorithm assigns to a given range cell velocities that come from other cells, thus generating high uncertainty.

Another source of high variance could be the occasional or persistent high variability of the currents over the time, in a specific part of the coverage area.

To implement this test on CODAR data, the variance field has to be evaluated from the variable `STDV`, which contains the radial temporal standard deviation of current velocity over the coverage period (referred to the integration period necessary for the production of the hourly radial velocity data). In previous releases of the CODAR LLUV file format, the variable `STDV` was named `ETMP`.

For WERA systems, the variance threshold has to be applied to the variable `EVAR`, containing the radial variance of current velocity over coverage period (referred to the integration period necessary for the production of the hourly radial velocity data).

The choice of the maximum variance threshold has to be made carefully, since using a tight threshold can lead to the elimination of accurate data. The recommendation in this case is to examine the time-series of the maximum variance of the data (during at least

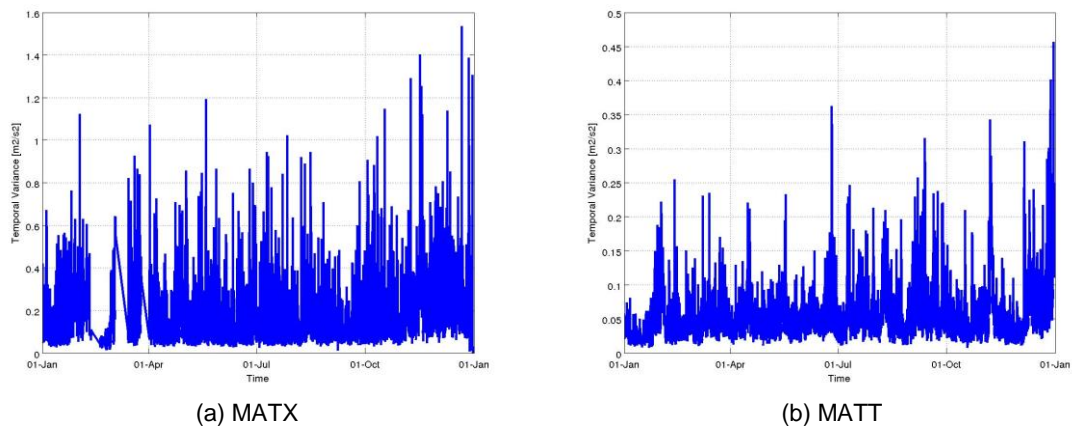
one complete seasonal cycle for which the proper functioning of the device is assessed) and consult the literature.

Furthermore, it is also recommended to analyse the impact of the selected threshold (or of the candidate thresholds) on the velocity fields of the dataset containing at least one complete seasonal cycle.

Examples of application of this test are provided in the next subsection for the two case study areas.

#### 6.4.2 Application

Figure 10 shows the evolution in time of the maximum temporal variance for the sites MATX (a) and MATT (b), belonging to the two case studies presented in Section 5.



*Figure 10 – Evolution in time of the maximum temporal variance during year 2014 for MATX station (a), belonging to the BoB HFR network, and for the MATT station (b), belonging to the GoM HFR network.*

Based on the variance trends in the two stations (the maximum temporal variance mainly ranges in  $[0 - 1] \text{ m}^2/\text{s}^2$  for MATX site and in the range  $[0 - 0.25] \text{ m}^2/\text{s}^2$  for MATT site) it is possible to define an adequate maximum temporal variance threshold. For this example let us consider two different thresholds,  $0.15 \text{ m}^2/\text{s}^2$  and  $0.35 \text{ m}^2/\text{s}^2$ , in order to verify the impact they have on the velocity fields. This approach is recommended for comparing the effects of different thresholding strategies.

Figure 11 shows (on z axis) the number of times the maximum temporal variance threshold labelled as “bad data” the original velocity fields in MATX and MATT stations using the two thresholds listed above. The z-axis values are normalized on the number of times the specific cell has content (i.e. is not a NaN value). The colour patches of the surfaces report the average difference between the original temporal variances and the threshold for each spatial cell. The X-Y axes grid is the exact range-bearing grid of the original data. In this way, it is possible to localize each radial velocity cell.

As explained in Section 6.3.2, Figure 11 shows the frequency and the amount of the intervention of the test and the precise localization of the cell affected by the test, thus being a valuable aid in the definition of the threshold to apply to data.

Figure 11(a) and Figure 11(b) show that the  $0.15 \text{ m}^2/\text{s}^2$  threshold is acting on the whole grid very often but with small amounts for both sites. Figure 11(c) and Figure 11(d) show that a higher threshold ( $0.35 \text{ m}^2/\text{s}^2$ ) acts less frequently in both sites: in MATX the action of the test is still spread, while in MATT the filter is active very rarely. As expected, in MATX site the vectors with higher variance lie on the farthest border of the coverage. The  $0.35 \text{ m}^2/\text{s}^2$  seems to be still too low for MATX, since the number of correction over the whole domain is still high.

Again, the recommendation is to investigate any particular situation emerging from these graphics before to define the operational threshold.

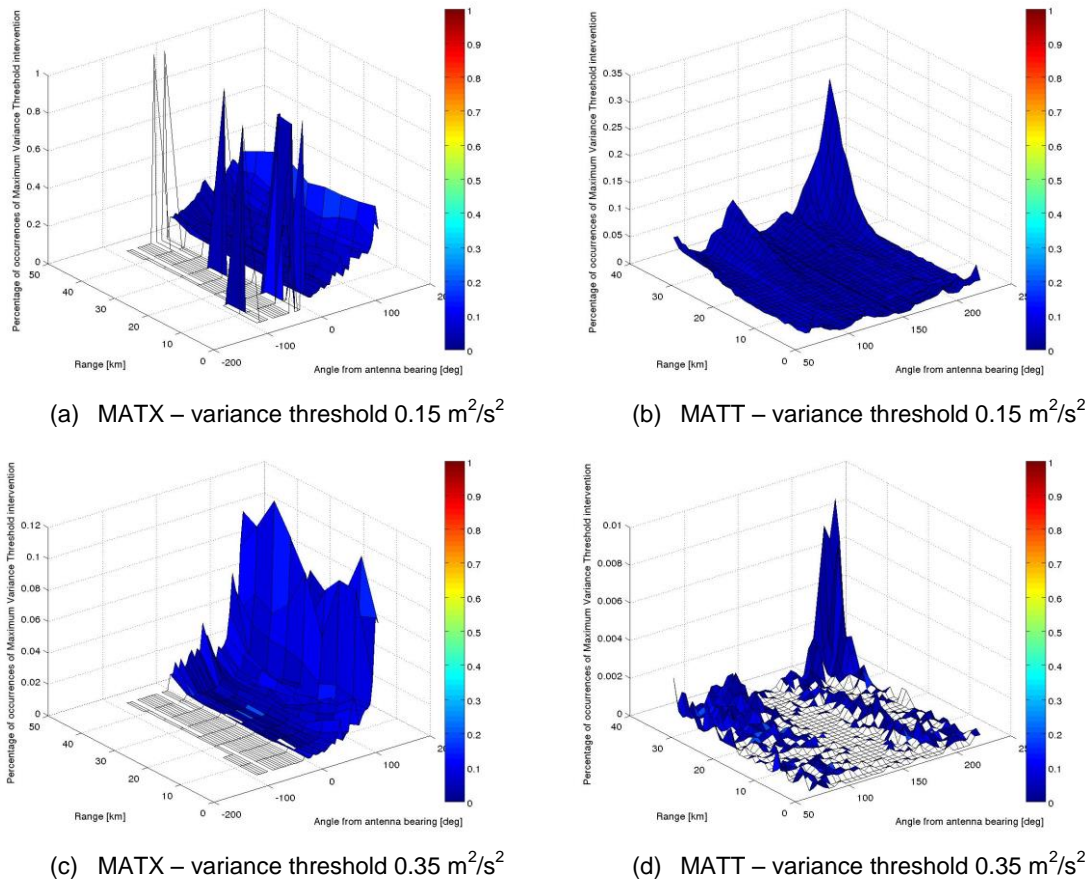


Figure 11 – Sensitivity to the maximum temporal variance threshold. The number of times the maximum temporal variance threshold labelled as “bad data” the original velocity field is given in the vertical (z) axis, normalized on the number of times the specific cell has content (i.e. is not a NaN value). The horizontal plane (axes x, y) provide the range and bearing of the radial data cells. The colour patches of the surfaces report the average differences between the original temporal variances and the threshold for each spatial cell. Panels (a) and (c) refer to MATX site using  $0.15 \text{ m}^2/\text{s}^2$  and  $0.35 \text{ m}^2/\text{s}^2$  as threshold respectively. Panels (b) and (d) refer to MATT site using  $0.15 \text{ m}^2/\text{s}^2$  and  $0.35 \text{ m}^2/\text{s}^2$  as threshold respectively.

As for the previous test, a useful and simple check to do when choosing the threshold is to examine the evolution in time of the average velocity module with and without the application of the candidate thresholds. In this way it is possible to have a quick overview of the effects of the test (that keeps only vectors whose variance is lower than the threshold) on the velocity field. Figure 12 reports this comparison.

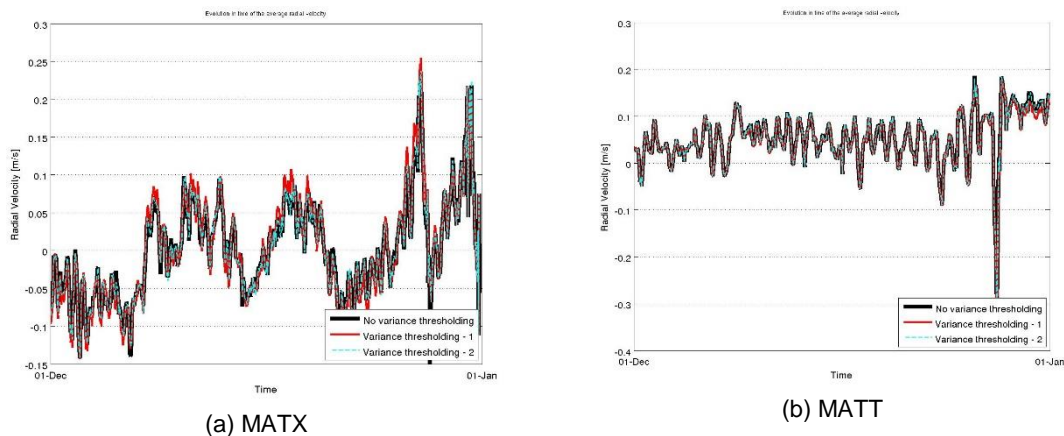


Figure 12 – Evolution in time (December 2014) of the average radial velocity with and without the application of different variance thresholds for MATX (a) and MATT (b) sites.

## 6.5 Average radial bearing

### 6.5.1 Definition and computation

Test determining that the average bearing of the radial velocity vectors lies within a specified interval centred on the expected value for proper operation. The value of proper operation has to be defined within a time interval when the proper functioning of the device is assessed. The interval has to be set according to the site-specific properties. The output is a scalar QC variable assuming “good data” value if the average radial bearing lies in the proper range and the “bad data” value if not.

This test is linked to the antenna pattern evolution, and it is very important for a compact antenna, i.e. for DF systems. It has not a significant meaning for PA, even if it can be evaluated for both kind of devices.

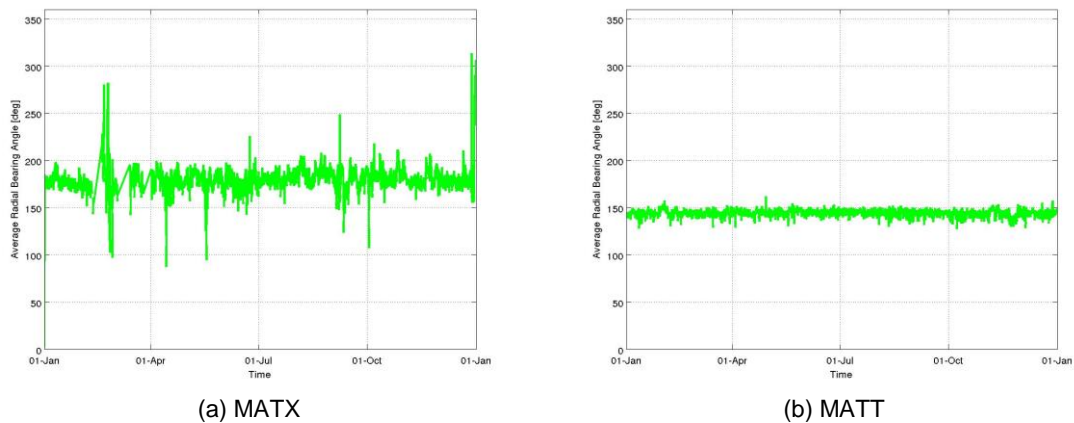
For both CODAR and WERA systems, the average radial bearing of a velocity field has to be evaluated by averaging the variable `BEAR`, which contains the current vector bearing in degrees clockwise from true North from the site origin.

The choice of the valid bearing range has to be made on the basis of the installation geometry and on the trend in historical series. The recommendation in this case is to examine the time-series of the average bearing of radial velocities (during at least one complete seasonal cycle for which the proper functioning of the device is assessed).



Figure 13 shows the evolution in time of the average bearing of radial velocities for the sites MATX (a) and MATT (b), belonging to the two case studies presented in Section 5.

An example of application of this test is provided in the next subsection.



*Figure 13 – Evolution in time of the average bearing of radial velocity vectors during year 2014 for MATX station (a), belonging to the BoB HFR network, and for the MATT station (b), belonging to the GoM HFR network.*

### 6.5.2 Application

To illustrate the efficacy and application of this test, we will use a specific situation occurred in the BoB HFR in May-June 2015 (Figure 14). At the end of May the antenna at HIGE site was seriously damaged. For several days, the antenna laid over the floor while continuing to emit. During this period of malfunctioning, the maximum and medium total range covered by the antenna remained unchanged. However there was a drastic change in the angular coverage and thus in the mean bearing of the area covered by the antenna (see map in Figure 14(b)), with the consequent decrease in the total number of valid radial and total data (Figure 14.a).

Due to the location of the two antennas in the BoB HFR and the change in the coast direction, the mean bearing angle of the oceanic area covered by the two antennas is different. The mean bearing angle for HIGE antenna is higher than that of Matxitxako antenna.

In figure 14, it can be observed how the mean bearing angle for HIGE antenna was lower than that for MATX antenna during the period of hardware problems (May 30 – June 2). The system was fixed on June 2 (see first arrow in Figure 14(a), lower panel) and the following days the APM of both antennas were performed (second and third arrows in Figure 14(a), lower panel). From the date of the last APM on the system restarted a normal operation (See Figure 14(c)).

As it can be observed in the temporal series of the mean bearing angles, small changes in the mean bearing angle during normal operation can occur. However any major change of the mean bearing angle is a good indicator of bad quality data. The

threshold has to be chosen in order to allow the small variations and detect any significant drift from the normal values. For instance, for the BoB HFR the optimal thresholds are [160 200] for MATX and [140 160] for HIGE.

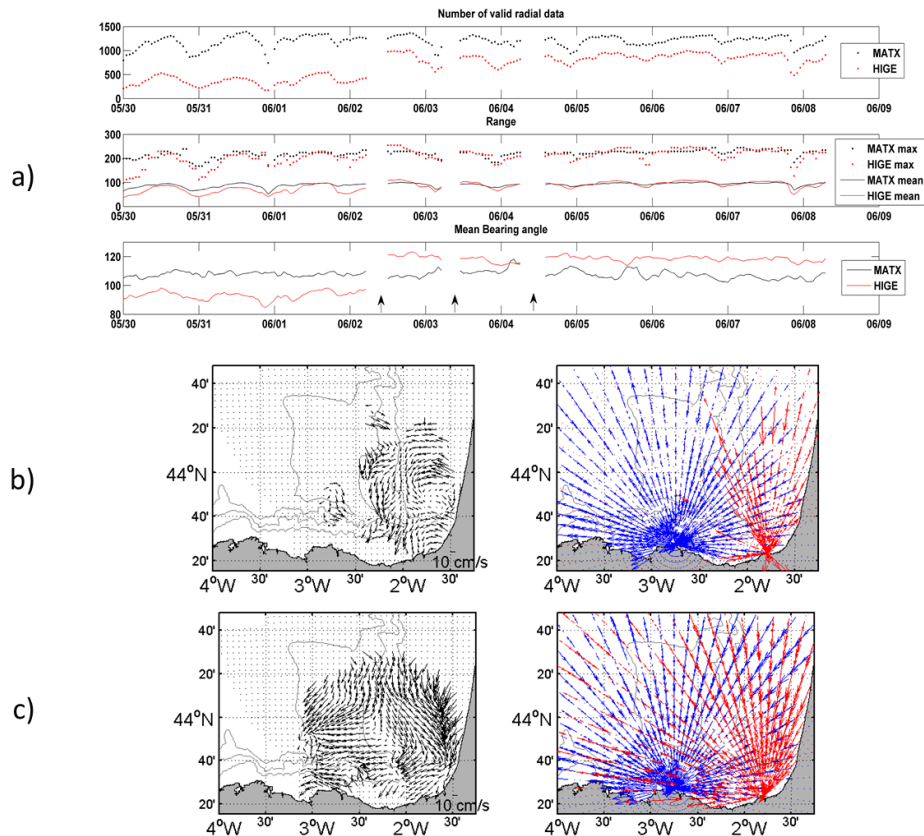


Figure 14 – (a) Time series of the total number of valid radial data, maximum and mean range and mean bearing angle for HIGE and MATX antennas. The three arrows in the lower panel show the dates of the antenna repair, and the two APMs, respectively. (b) Total and radial fields for June 1, 23:00 during the period of malfunctioning (before the antenna repair). (c) Total and radial fields for June 2, 14:00 during the period of normal operation (after the antenna repair)

## 6.6 Median filter

### 6.6.1 Definition and computation

This QC test aims at assessing the smoothness of the distribution of the velocity vectors in the radial field. For each source radial vector (i.e. radial bin), the median of all velocities within a radius of  $\langle RCLim \rangle$  and whose vector bearing (angle of arrival at site) is also within an angular distance of  $\langle AngLim \rangle$  degrees from the source vector's bearing is evaluated. If the difference between the vector's velocity and the median velocity is greater than a threshold  $\langle CurLim \rangle$ , then the vector is labelled as “bad data”. Otherwise the vector is labelled as “good data”. The output is a gridded QC variable with the same dimensions of the radial velocity data variable, containing the flag values for each cell.



In the application of this test there is also the possibility, if the difference between the vector's velocity and the median velocity is greater than the  $\langle \text{CurLim} \rangle$  threshold, of substituting the velocity vector with the median evaluated in the surrounding window and label it as “interpolated data”. The choice is up to the operator. Anyway, the recommendation is to keep the original values and to flag velocity vectors as “good data” or “bad data”, in order not to lose the measured information.

It has to be noticed that this test could be tricky when applied in situations with high heterogeneity due to islands or small structures present in the HFR coverage. In these cases some information can be lost.

To apply this test to both CODAR and WERA data, the window for the computation of the median value has to be built using the variables `VELO`, containing the current vector velocity, and `BEAR`, containing the current vector bearing.

The choice of the three thresholds ( $\text{RCLim}$ ,  $\text{AngLim}$  and  $\text{CurLim}$ ) has to be made carefully, after having analysed the behaviour of the current fields in the monitored area and according to literature indications. Also it must be taken into account the bearing and range resolution of the original data.

Furthermore, it is also recommended to analyse the impact of the selected threshold (or of the candidate thresholds) on the velocity fields of the dataset containing at least one complete seasonal cycle.

Examples of application of this test are provided in the next subsection for the two case study areas.

### 6.6.2 Application

For this example, two sets of thresholds are defined, in order to evaluate the impact on velocity fields of two different approaches.

The first set of thresholds is defined as:

- $\text{RCLim} = 10 \text{ km}$
- $\text{AngLim} = 30^\circ$
- $\text{CurLim} = 0.5 \text{ m/s}$

It is characterized by high values of all three parameters, thus acting only if the difference between the measured current vector under investigation and the median is big, and also by large range and bearing limits, thus evaluating the median velocity over wide areas around the analysed vector. It is expected that evaluating the median velocity over a wide area likely brings to bigger differences between the median and the original value. This setup keeps smooth velocities over large areas.

The second set of thresholds is defined as:

- $\text{RCLim} = 5 \text{ km}$
- $\text{AngLim} = 10^\circ$
- $\text{CurLim} = 0.02 \text{ m/s}$

This setup acts more frequently on the velocity field, since the velocity difference required for triggering the filter is much lower (0.02 m/s). In this case the median velocity is evaluated on a smaller area than in the previous case.

Figure 15 shows (on z axis) the number of times the median filter labelled as “bad data” the original radial velocity fields in MATX and MATT station using the two sets of thresholds listed above. The z-axis values are normalized on the number of times the specific cell has content (i.e. is not a NaN value).

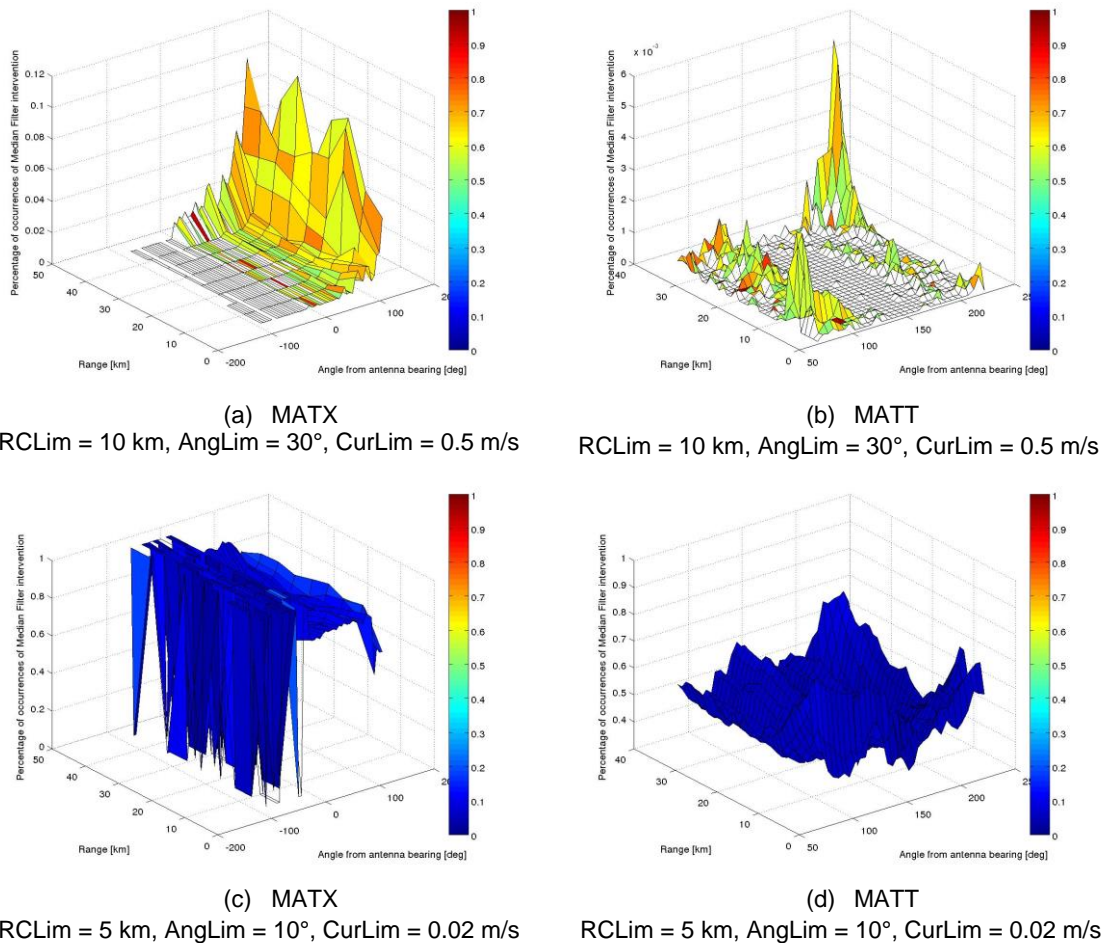


Figure 15 – Sensitivity to the median filter thresholds. The number of times the median filter labelled as “bad data” the original velocity field is given in the vertical (z) axis, normalized on the number of times the specific cell has content (i.e. is not a NaN value). The horizontal plane (axes x, y) provides the range and bearing of the radial data cells. The colour patches of the surfaces report the average differences between the original velocities and the median values for each spatial cell. Panels (a) and (c) refer to MATX site using RCLim = 10 km, AngLim = 30°, CurLim = 0.5 m/s and RCLim = 5 km, AngLim = 10°, CurLim = 0.02 m/s respectively. Panels (b) and (d) refer to MATT site using RCLim = 10 km, AngLim = 30°, CurLim = 0.5 m/s and RCLim = 5 km, AngLim = 10°, CurLim = 0.02 m/s respectively.

The colour patches of the surfaces report the average differences between the original velocities and the median velocities for each spatial cell. The X-Y axes grid is the exact range-bearing grid of the original data. In this way it is possible to localize each radial velocity cell.

As explained in Section 6.3.2, Figure 15 shows the frequency and the amount of the intervention of the test and the precise localization of the cell affected by the test, thus being a valuable aid in the definition of the thresholds to apply to data.

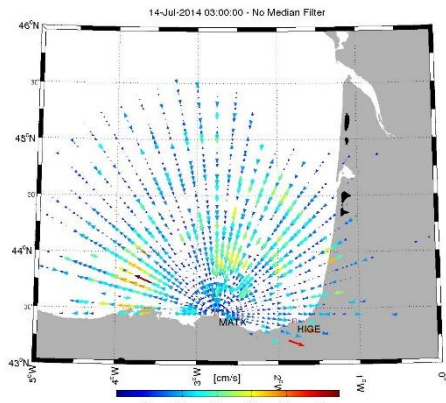
Figure 15(a) and Figure 15(b) show that the first set of thresholds is acting widely but not so frequently on both MATX and MATT data. In particular, it is acting rarely on MATT data and a little more frequently on MATX data. In MATX data, the more corrected values are the ones belonging to the farthest border of the coverage, as usual.

The second set of thresholds is acting very widely and very frequently on both data (Figure 15(c) and Figure 15(d)). This situation seems to suggest to increase the CurLim value (in order to let the filter act less) or to low down the RCLim and AngLim values, in order to reduce the area where to evaluate the median velocity (the biggest the area the most diverse the median and original velocity).

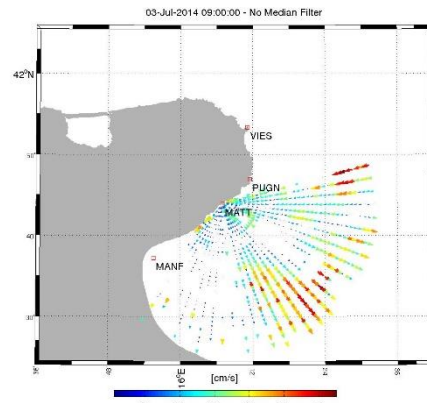
As for the previous tests, the recommendation is to investigate any particular situation emerging from these graphics before to define the operational thresholds.

Figure 16 shows an example of how the median filter would modify the different velocity fields. As expected, the second set of thresholds smooths the fields in a significant way, thus the filtered fields are very different from the original one. This visualization is very useful to manage the trade off in selecting the thresholds. Indeed, for the BoB system the results suggest that even the second threshold is too restrictive, since it is smoothing significantly the data along the radial directions. Since the range resolution of BoB is lower than that of GoM system the RCLim and CurLim should be probably higher in this case. This illustrates again the importance of choosing the correct thresholds taking into account the system and data characteristics.

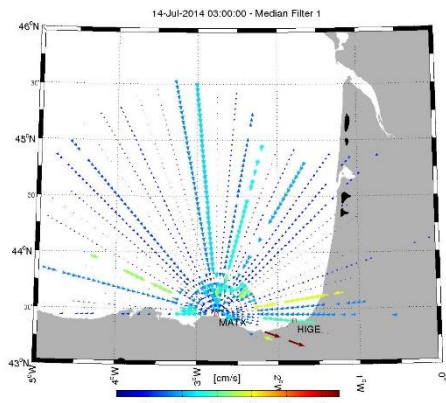
A useful and simple check to do when choosing the thresholds for the test, is to examine the evolution in time of the average velocity with and without the application of the candidate threshold sets. In this way it is possible to have a quick overview of the effects of the test (that keeps only velocities fulfilling the median filter constraint) on the velocity field. Figure 17 reports this comparison.



(a) MATX  
Original radial velocity field

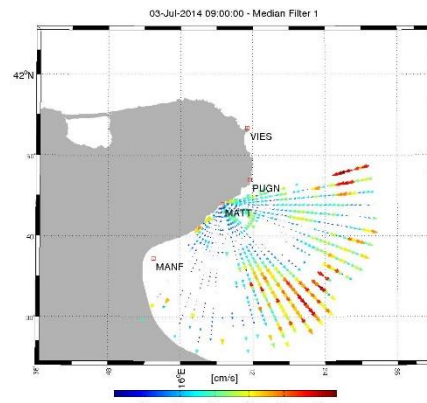


(b) MATT  
Original radial velocity field



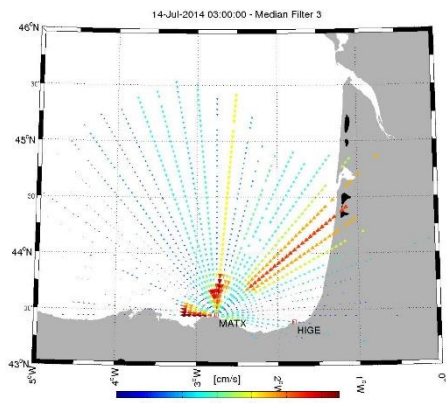
(c) MATX

RCLim = 10 km, AngLim = 30°, CurLim = 0.5 m/s



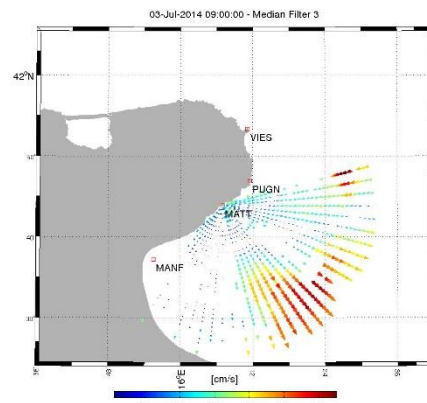
(d) MATT

RCLim = 10 km, AngLim = 30°, CurLim = 0.5 m/s



(e) MATX

RCLim = 5 km, AngLim = 10°, CurLim = 0.02 m/s



(f) MATT

RCLim = 5 km, AngLim = 10°, CurLim = 0.02 m/s

Figure 16 – Radial velocity maps without and with applying the median filter for MATX and MATT sites. Panels (a), (c) and (e) refer to MATX site using no median filter,  $RCLim = 10\text{ km}$ ,  $AngLim = 30^\circ$ ,  $CurLim = 0.5\text{ m/s}$  and  $RCLim = 5\text{ km}$ ,  $AngLim = 10^\circ$ ,  $CurLim = 0.02\text{ m/s}$  respectively. Panels (b), (d) and (f) refer to MATT site using no median filter,  $RCLim = 10\text{ km}$ ,  $AngLim = 30^\circ$ ,  $CurLim = 0.5\text{ m/s}$  and  $RCLim = 5\text{ km}$ ,  $AngLim = 10^\circ$ ,  $CurLim = 0.02\text{ m/s}$  respectively.

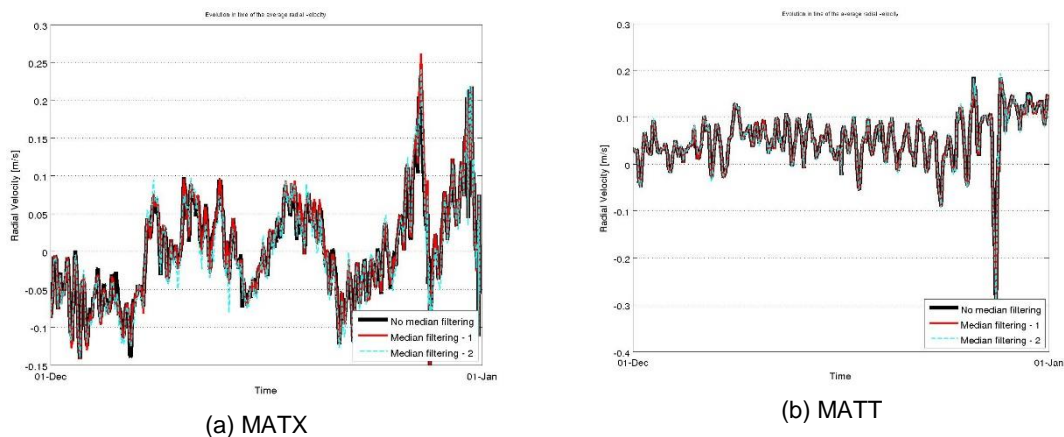


Figure 17 – Evolution in time (December 2014) of the average radial velocity with and without the application of different median filter thresholds for MATX (a) and MATT (b) sites.

## 7 Implementation of QC tests to total data

Six QC tests are being tested for total data. Five of them have been defined as mandatory for the European Common data and metadata model for real-time HFR data. The “Balance Of Contributing Radials” test has been defined as recommended for CODAR operators.

### 7.1 Syntax Test

#### 7.1.1 Definition and computation

This test will ensure the proper formatting and the existence of all the necessary fields within the total netCDF file. This test is performed on the netCDF files and it assesses the presence and correctness of all data and attribute fields and the correct syntax throughout the file. The reference schemes for data and metadata structures and syntax are described in the deliverable 5.13 of the JERICO-Next project.

As for radials (see section 6.1.1) complementary test to verify some of the fields contained in the metadata (timestamp, site names, site codes, time zones, site locations) can be also applied, and the correct values should be established by the operators.

The result of the test is a scalar QC variable that is entered into the netCDF file after the test completion.

### 7.2 Data Density threshold

#### 7.2.1 Definition and computation

This test labels total velocity vectors with a number of contributing radials bigger than the threshold with a “good data” flag and total velocity vectors with a number of

contributing radials smaller than the threshold with a “bad data” flag. The output is a gridded QC variable with the same dimensions of the total velocity components data variable, containing, for each cell, the flag related to the vector lying in that cell.

For data produced by CODAR systems, the information about the number of contributing radials has to be inferred from the variables  $S1CN$ ,  $S2CN$ , ...,  $SnCN$  ( $n$  is the total number of contributing sites), which contain the number of contributing radial vectors per site to each total.

This test aims at assessing the robustness of the total velocity vectors. The higher is the number of the contributing radials, the more the resulting total vector can be considered reliable, provided that the variance of the values is not too high.

### 7.2.2 Application

As for other QC tests, the correct threshold for the application of this test has to be chosen in view of the historical data set. Two plots are useful to establish an adequate threshold. On one hand, we show in Figure 18 the histogram of the number of radials that are contributing to the total vectors for the BoB installation. The distribution corresponds to one year of data (2014) and considers all the total grid cells. We can observe that the most frequent number of radials available for total combination is between 30 and 60, being the number of totals whose combination is done from less than 10 radials relatively small. If we look at the distribution of the number of radials used for total combination in space (Figure 19), the areas where the number of radial is small are mostly located at the periphery of the radar coverage. So for this case using a threshold of 10 radial vectors seems adequate.

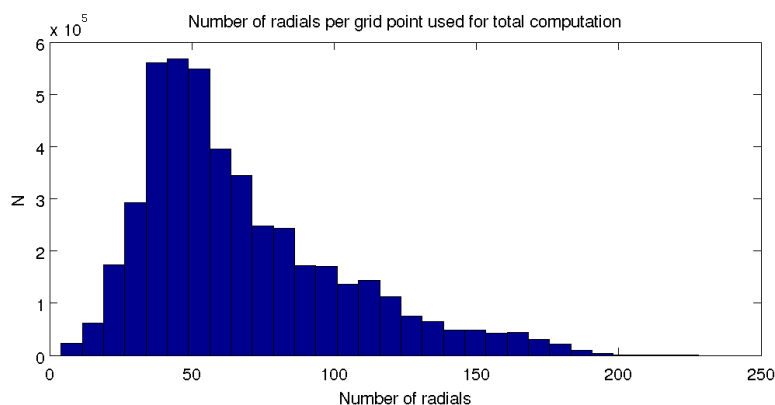


Figure 18 – Histogram of the number of radials used for total combination on the BoB HFR system in 2014.



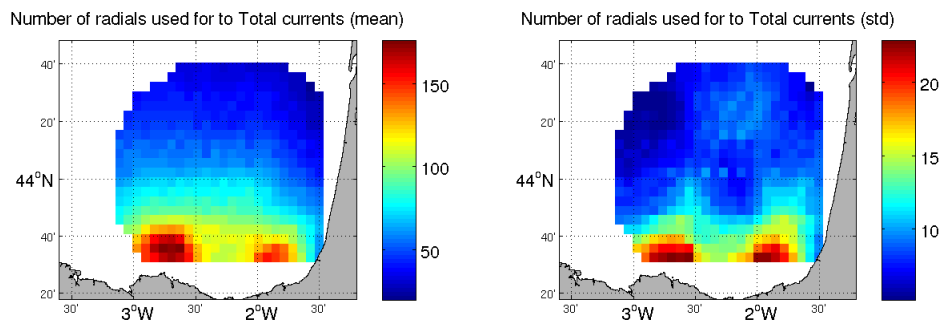


Figure 19 – (a) Spatial distribution of the mean number of radials used for total combination. (b) Standard deviation with respect to the mean values.

### 7.3 Velocity threshold

#### 7.3.1 Definition and computation

This test labels total velocity vectors whose module is bigger than a maximum velocity threshold with a “bad data” flag and total vectors whose module is smaller than the threshold with a “good data” flag. The output is a gridded QC variable with the same dimensions of the total velocity data variable, containing, for each cell, the flag related to the vector lying in that cell.

It has to be noticed that both CODAR and WERA LLUV data formats contain the zonal (U) and meridional (V) components of the total velocity vectors. Thus, the module of the total velocity, that is the quantity to be compared with the threshold, has to be evaluated from the two components.

For both data type, the module of the total velocity vectors have to be computed using the variables `VELU` and `VELV` present in CODAR and WERA data structure.

The choice of the maximum total velocity threshold has to be made carefully, since using a tight threshold can lead to the elimination of accurate data. The recommendation in this case is to examine the time-series of the maximum total velocity of the data (during at least one complete seasonal cycle for which the proper functioning of the device is assessed) and consult the literature.

Furthermore, it is also recommended to analyse the impact of the selected threshold (or of the candidate thresholds) on the velocity fields of the dataset containing at least one complete seasonal cycle.

Examples of application of this test are provided in the next subsection for the two case study areas.

#### 7.3.2 Application

With reference to BoB and GoM HFR networks, Figure 20 shows the evolution in time of the maximum total velocity for both the networks during the year 2014.

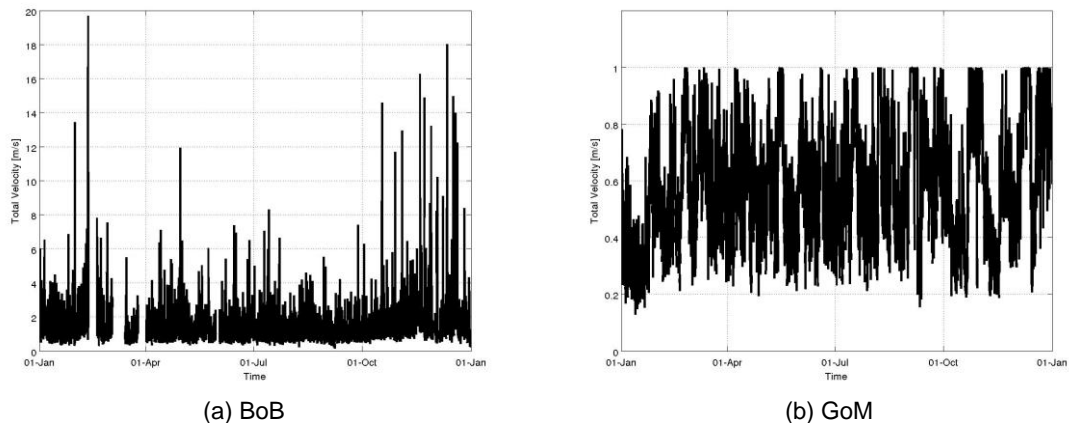


Figure 20 – Evolution in time of the maximum total velocity during year 2014 for BoB network (a) and for GoM network (b).

The two graphs show that (for year 2014) the maximum total velocity mainly ranges in  $[0 - 4]$  m/s for BoB network and in  $[0.2 - 0.8]$  m/s for GoM network. Based on these trends, it is possible to define a maximum velocity threshold compatible with the velocity values measured by the networks. For this example let us consider two different thresholds, 0.6 m/s and 1 m/s, in order to verify the impact they have on the velocity fields. This approach is recommended, since it allows for comparing the effects of different thresholding strategies.

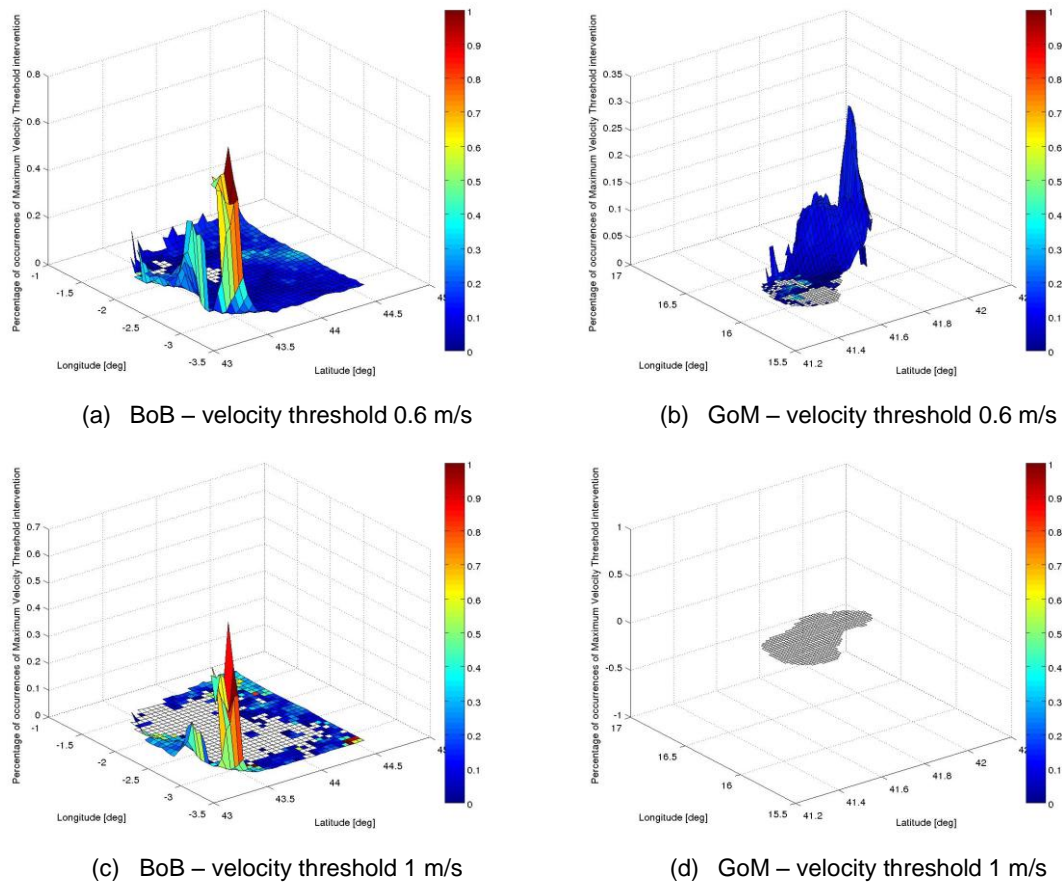
Figure 21 shows (on z axis) the number of times the maximum total velocity threshold labelled as “bad data” the original velocity fields in BoB and GoM data using the two thresholds listed above. The z-axis values are normalized on the number of times the specific cell has content (i.e. is not a NaN value). The colour patches of the surfaces report the average differences between the original total velocities and the threshold for each spatial cell. The X-Y axes grid is the exact latitude-longitude grid of the original data. In this way, it is possible to localize each total velocity cell.

As explained in Section 6.3.2, Figure 21 shows the frequency and the amount of the intervention of the test and the precise localization of the cell affected by the test, thus being a valuable aid in the definition of the threshold to apply to data.

Figure 21(a) shows that in BoB data the 0.6 m/s threshold is acting on the whole grid very often but with small amounts except for the area bounded by latitude in  $[-3 -2]^\circ$  E and longitude in  $[43.4 - 43.6]^\circ$  N, where big differences between velocity and threshold are present. Figure 21(c) highlights that a bigger threshold (1 m/s) acts less frequently all over the coverage, keeping big differences in the area described above. Unless a scientific reason is provided, those data look suspect.

Concerning GoM data, Figure 21(b) shows that the test acts mainly in the central part of the coverage. As expected from the trend depicted in Figure 20, in Figure 21(d) it is evident that the 1 m/s threshold is not acting at all on data.

As mentioned before, the recommendation is to investigate any particular situation emerging from these graphics before to define the operational threshold.



*Figure 21 – Sensitivity to the maximum total velocity threshold. The number of times the maximum total velocity threshold labelled as “bad data” the original velocity field is given in the vertical (z) axis, normalized on the number of times the specific cell has content (i.e. is not a NaN value). The horizontal plane (axes x, y) provides the latitude and longitude of the total data cells. The colour patches of the surfaces report the average differences between the total velocities and the threshold for each spatial cell. Panels (a) and (c) refer to BoB site using 0.6 m/s and 1 m/s as threshold respectively. Panels (b) and (d) refer to GoM site using 0.6 m/s and 1 m/s as threshold respectively.*

As for the tests on radial data, a useful and simple check to do when choosing the threshold is to examine the evolution in time of the average total velocity with and without the application of the candidate thresholds. In this way it is possible to have a quick overview of the effects of the test (that keeps only velocities whose amplitude is smaller than the threshold) on the velocity field. Figure 22 reports this comparison.

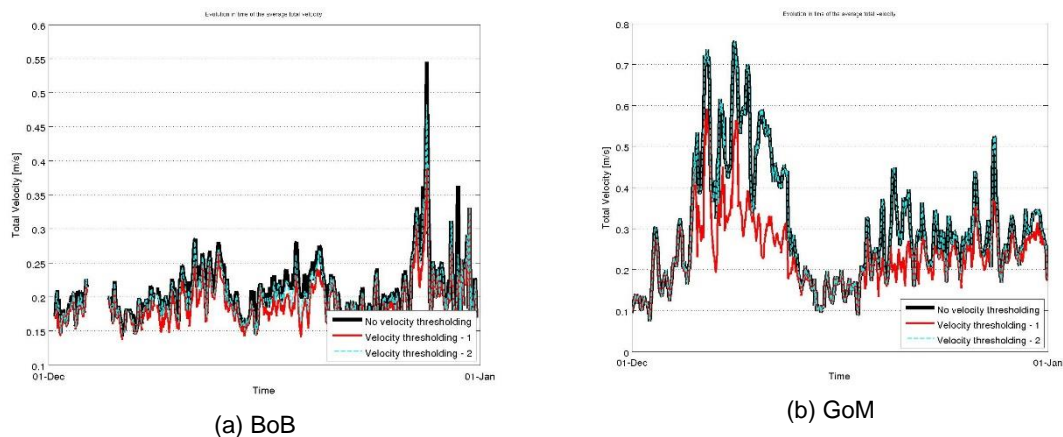


Figure 22 – Evolution in time (December 2014) of the average total velocity with and without the application of different velocity thresholds for BoB (a) and GoM (b) networks.

## 7.4 Variance threshold

### 7.4.1 Definition and computation

This test labels total velocity vectors whose variance is bigger than a maximum threshold with a “bad data” flag and total vectors whose variance is smaller than the threshold with a “good data” flag. The output is a gridded QC variable with the same dimensions of the total velocity component data variables, containing, for each cell, the flag related to the vector lying in that cell.

To implement this test on CODAR data, the total velocity variance field has to be evaluated from the standard deviations of the U and V components of the total vectors. These standard deviations are contained in the variables  $U_{QAL}$  and  $V_{QAL}$  present in the total data structure.

The choice of the maximum variance threshold has to be made carefully, since using a tight threshold can lead to the elimination of accurate data. The recommendation in this case is to examine the time-series of the maximum variance of the data (during at least one complete seasonal cycle for which the proper functioning of the device is assessed) and consult the literature.

Furthermore, it is also recommended to analyse the impact of the selected threshold (or of the candidate thresholds) on the velocity fields of the dataset containing at least one complete seasonal cycle.

Examples of application of this test are provided in the next subsection for the two case study areas.

### 7.4.2 Application

Figure 23 shows the evolution in time of the maximum variance for the BoB (a) and GoM (b) networks, the two case studies presented in Section 5.

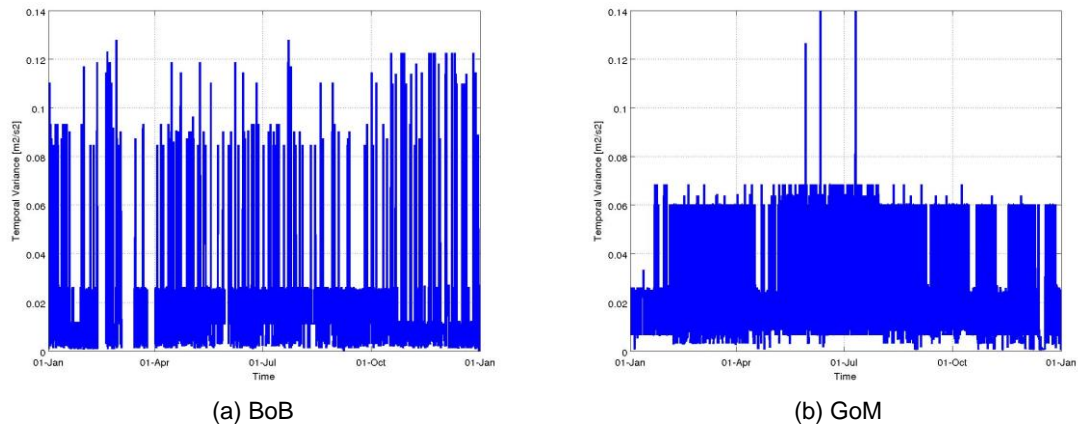


Figure 23 – Evolution in time of the maximum variance during year 2014 for BoB (a) and GoM (b) networks.

Based on the variance trends in the two networks (the maximum variance mainly ranges in  $[0 - 0.12] \text{ m}^2/\text{s}^2$  for BoB data and in the range  $[0 - 0.07] \text{ m}^2/\text{s}^2$  for GoM data) it is possible to define an adequate maximum variance threshold. For this example let us consider two different thresholds,  $0.15 \text{ m}^2/\text{s}^2$  and  $0.35 \text{ m}^2/\text{s}^2$ , in order to verify the impact they have on the velocity fields. This approach is recommended for comparing the effects of different thresholding strategies.

Figure 24 shows (on z axis) the number of times the maximum variance threshold labelled as “bad data” the original velocity fields in BoB and GoM networks using the two thresholds listed above. The z-axis values are normalized on the number of times the specific cell has content (i.e. is not a NaN value). The colour patches of the surfaces reports the average difference between the original variances and the threshold for each spatial cell. The X-Y axes grid is the exact latitude-longitude grid of the original data. In this way, it is possible to localize each radial velocity cell.

As explained in Section 6.3.2, Figure 24 shows the frequency and amount of the intervention of the test and the precise localization of the cell affected by the test, thus being a valuable aid in the definition of the threshold to apply to data.

As expected from the trends depicted in Figure 23, Figure 24 shows that the selected thresholds do not act on BoB and GoM data.

Again, the recommendation is to investigate any particular situation emerging from these graphics before to define the operational threshold.

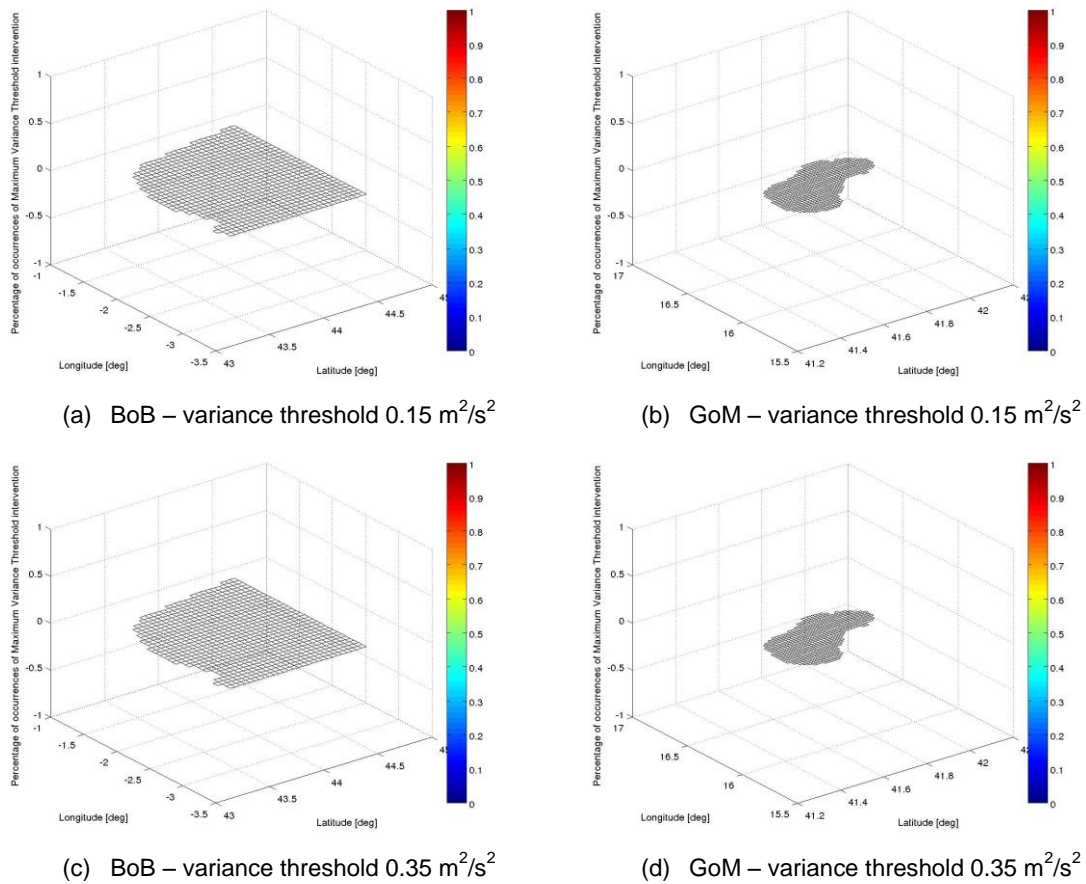


Figure 24 – Sensitivity to the maximum variance threshold. The number of times the maximum variance threshold labelled as “bad data” the original velocity field is given in the vertical ( $z$ ) axis, normalized on the number of times the specific cell has content (i.e. is not a NaN value). The horizontal plane (axes  $x, y$ ) provides the latitude and longitude of the total data cells. The colour patches of the surfaces report the average differences between the original variances and the threshold for each spatial cell. Panels (a) and (c) refer to BoB network using  $0.15 \text{ m}^2/\text{s}^2$  and  $0.35 \text{ m}^2/\text{s}^2$  as threshold respectively. Panels (b) and (d) refer to GoM network using  $0.15 \text{ m}^2/\text{s}^2$  and  $0.35 \text{ m}^2/\text{s}^2$  as threshold respectively.

As for the previous test, a useful and simple check to do when choosing the threshold is to examine the evolution in time of the average velocity with and without the application of the candidate thresholds. In this way it is possible to have a quick overview of the effects of the test (that keeps only velocities whose variance is lower than the threshold) on the velocity field. Figure 25 reports this comparison.

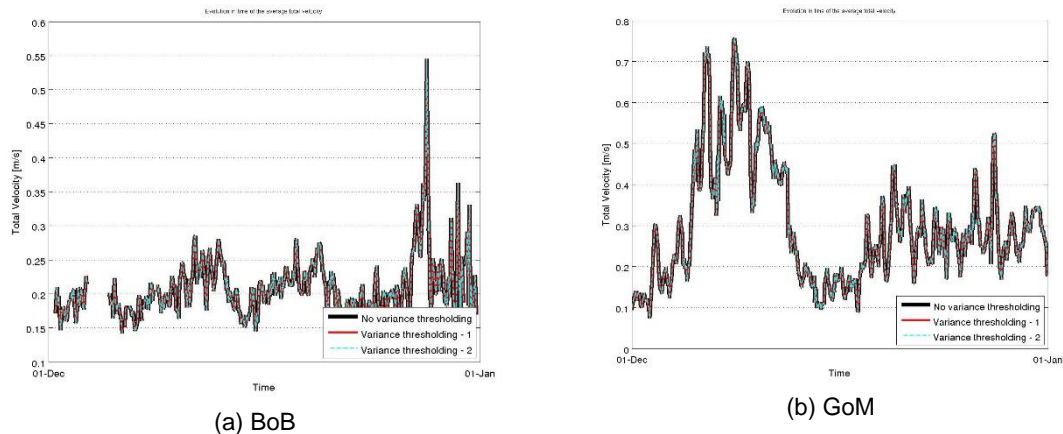


Figure 25 – Evolution in time (December 2014) of the average total velocity with and without the application of different variance thresholds for BoB (a) and GoM (b) networks.

## 7.5 Balance of contributing radials

### 7.5.1 Definition and computation

This test checks if the number of radials coming from the different contributing sites are balanced for the combination into the total velocity vectors. Each data cell is labelled with a “bad data” flag if the requested balance ratio is not achieved and with a “good data” flag if the balance ratio is achieved. The output is a gridded QC variable with the same dimensions of the total velocity component data variables, containing, for each cell, the flag related to the vector lying in that cell.

In CODAR data, the information about the number of contributing radial vectors per site to each total vector is contained in the variables  $S1CN$ ,  $S2CN$ , ...,  $SnCN$ , where  $n$  is the total number of contributing sites.

### 7.5.2 Application

In Figure 26 we can observe that the most part of the total data produced in the BoB HFR come from a balanced number of radials (up to 60% coming from one site). However the number of total data produced with less balanced contribution (with 70 – 80 % coming from one site) is quite high and choosing a threshold too tight would imply to lose a significant quantity of data. The spatial distribution of the percent of radials from the different sites is shown in Figure 27. For the BoB installation the areas where there is less balance are those located near each antenna, where indeed each antenna produces its highest radial data density. This suggests that the application of this QC test is not straightforward, at least not using a unique threshold for the whole coverage area.

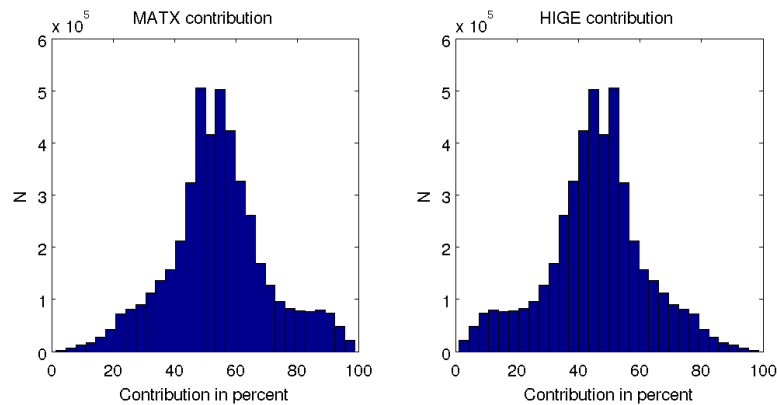


Figure 26 – Histogram of the percent of radials from MATX and HIGE sites respectively, used for total combination

It has also to be noticed that, if the Data Density test is passed, velocity vectors could be good even if the balance is poor. The suggested approach is not to choose a threshold too low, but to keep a non-extremely-unbalanced ratio between the radials coming from the contributing sites. For instance, a reasonable threshold would be to keep a limit of not more than 90% from one radial site since a lower one 70-80% would imply too much data loss. Another strategy could be applying the test only when the number of radials from one site is very low, in order not to lose much info near the antennas.

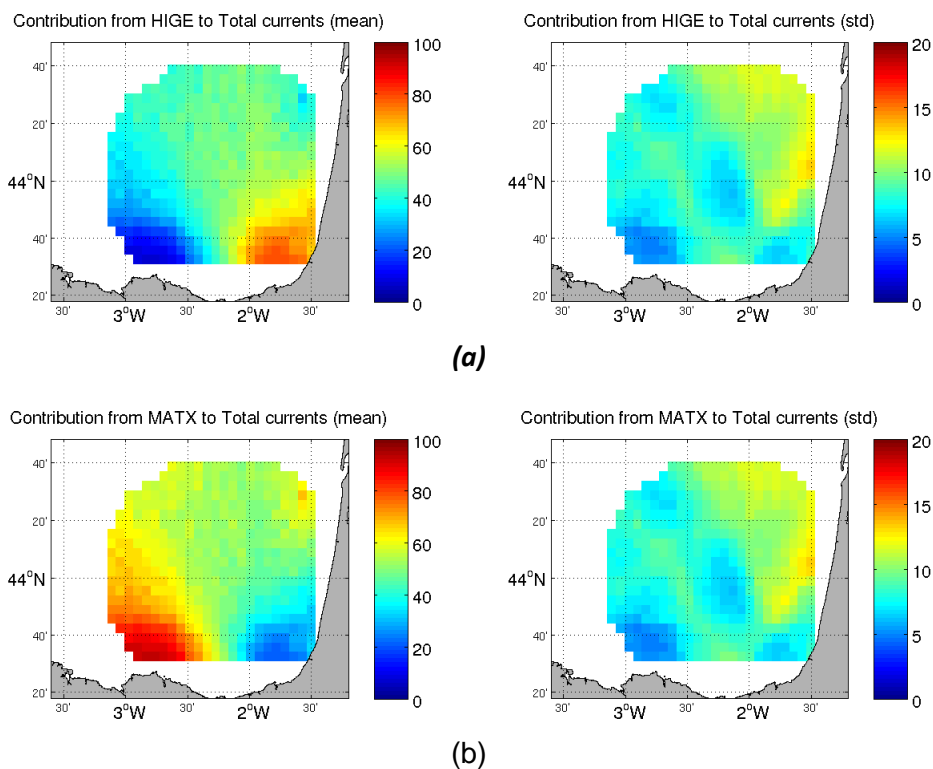


Figure 27 – (a) Spatial distribution of the percent of number of radials coming from the different contributing sites. (b) Standard deviation with respect to the mean values.





## 7.6 *GDOP Threshold*

### 7.6.1 Definition and computation

This test labels total velocity vectors whose GDOP is bigger than a maximum threshold with a “bad data” flag and the vectors whose GDOP is smaller than the threshold with a “good data” flag. The output is a gridded QC variable with the same dimensions of the total velocity component data variables, containing, for each cell, the flag related to the vector lying in that cell.

To implement this test on CODAR data, the GDOP threshold has to be applied to the variable `CQUAL` present in the CODAR total files. This variable contains the covariance matrix of the angle of incidence of radial vectors. The GDOP error is evaluated as the square root of the trace of this matrix. This means that the GDOP value constraints the range of incidence angles allowed for combining radial vectors into total vectors.

For WERA systems, the GDOP information is contained in the file `GDOP.dat`. To enable WERA systems to generate this file, it is necessary to insert the line “`WRITE_GDO=.TRUE.`” in the configuration file in the file `params.cfg`.

The choice of the maximum GDOP threshold has to be made carefully, since using a tight threshold can lead to the elimination of accurate data and to a dramatic reduction of the velocity field coverage. The recommendation in this case is to examine the time-series of the maximum GDOP and the related incidence angle ranges (during at least one complete seasonal cycle for which the proper functioning of the device is assessed) and consult the literature.

Furthermore, it is also recommended to analyse the impact of the selected threshold (or of the candidate thresholds) on the velocity fields of the dataset containing at least one complete seasonal cycle.

Examples of application of this test are provided in the next subsection for the two case study areas.

### 7.6.2 Application

Figure 28 shows the evolution in time of the maximum GDOP and related incidence angle ranges for the BoB (a) and GoM (b) networks, the two case studies presented in Section 5.

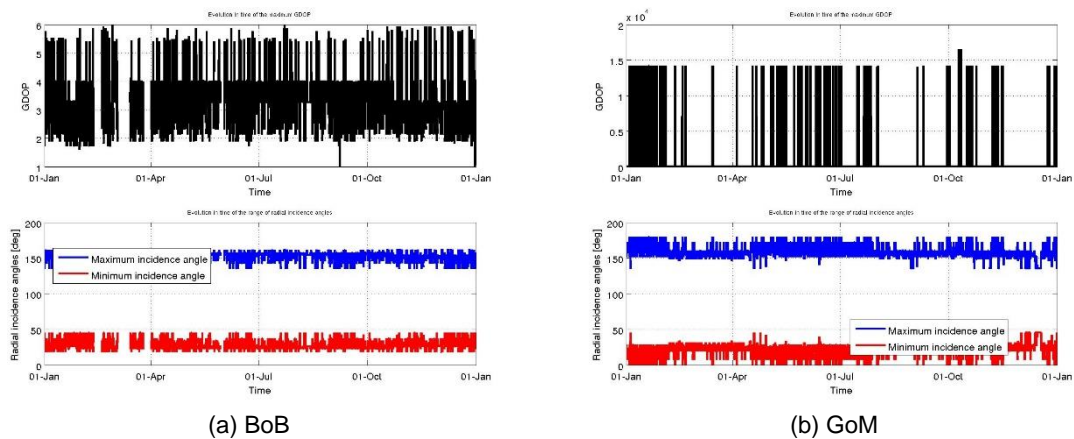


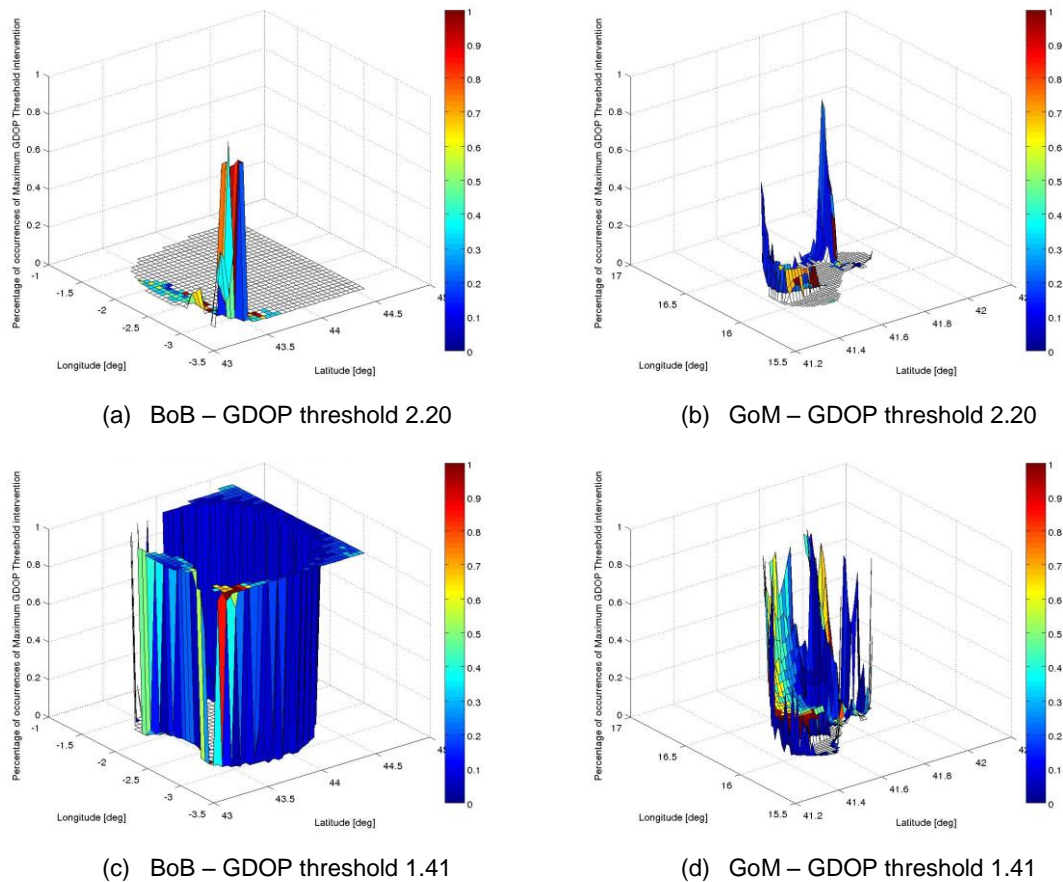
Figure 28 – Evolution in time of the maximum GDOP and related maximum and minimum incidence angles during year 2014 for BoB (a) and GoM (b) networks.

Figure 28 shows that maximum GDOP ranges in [2 - 5] in BoB network. This means that radial vectors are mainly combined when lying in the range [45 - 130]°. For GoM network, maximum GDOP ranges in [0 - 1.4], thus meaning that radial vectors are mainly combined when lying in the range [30 - 150]°. Based on these trends, it is possible to define an adequate maximum GDOP threshold. For this example let us consider two different thresholds, 2.20 and 1.41, in order to verify the impact they have on the velocity fields. This approach is recommended for comparing the effects of different thresholding strategies.

Figure 29 shows (on z axis) the number of times the maximum GDOP threshold labelled as “bad data” the original velocity fields in BoB and GoM networks using the two thresholds listed above. The z-axis values are normalized on the number of times the specific cell has content (i.e. is not a NaN value). The colour patches of the surfaces report the average difference between the original GDOP and the threshold for each spatial cell. The X-Y axes grid is the exact latitude-longitude grid of the original data. In this way, it is possible to localize each radial velocity cell.

As explained in Section 6.3.2, Figure 29 shows the frequency and amount of the intervention of the test and the precise localization of the cell affected by the test, thus being a valuable aid in the definition of the threshold to apply to data.

Figure 29 (a) and Figure 29 (b) show that the first threshold is acting very rarely and always on the border of the coverage for both networks. This is a consistent result, since the coverage borders are usually affected by higher GDOP. As expected, Figure 29 (c) and Figure 29 (d) show that the second threshold is more rigid, thus acting more frequently. Anyway, the threshold mainly operates on the coverage borders.



*Figure 29 – Sensitivity to the maximum GDOP threshold. The number of times the maximum GDOP threshold labelled as “bad data” the original velocity field is given in the vertical (z) axis, normalized on the number of times the specific cell has content (i.e. is not a NaN value). The horizontal plane (axes x, y) provide the latitude and longitude of the total data cells. The colour patches of the surfaces report the average differences between the original GDOP and the threshold for each spatial cell. Panels (a) and (c) refer to BoB network using 2.20 and 1.41 as threshold respectively. Panels (b) and (d) refer to GoM network using 2.20 and 1.41 as threshold respectively.*

As for the previous tests, the recommendation is to investigate any particular situation emerging from these graphics before to define the operational thresholds.

It has to be noticed that limiting the GDOP, i.e. limiting the incidence angle range for radial vector combination, means reducing the coverage of the velocity field, since only the total vectors whose contributing radials have an allowed incidence angle are kept in the field. Figure 30 shows the different coverages of the velocity fields for the two case studies applying the two thresholds used above, with respect to the original data. As expected, the second threshold reduces the coverage in a significant way. This visualization is very useful to manage the trade off in selecting the thresholds.

Another way to assess the impact of GDOP thresholding on the spatial coverage is to analyse the evolution in time of the number of velocity vectors present in the field, without any GDOP thresholding and with different GDOP thresholds, as shown in Figure 31.

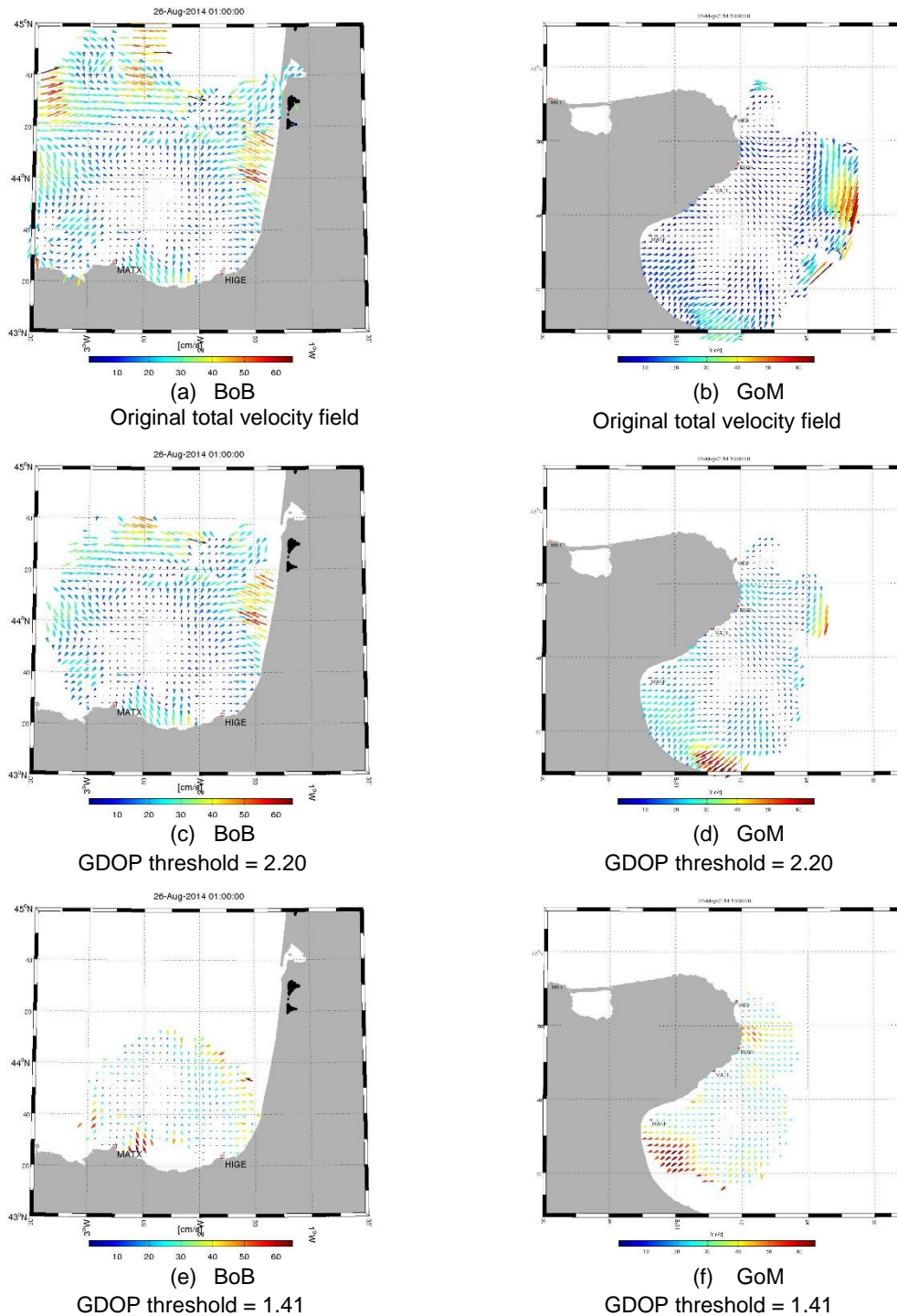


Figure 30 – Total velocity maps without and with applying the GDOP thresholds for BoB and GoM networks. Panels (a), (c) and (e) refer to BoB network using no GDOP thresholding, 2.20 and 1.41 as threshold respectively. Panels (b), (d) and (f) refer to GoM network using no GDOP thresholding, 2.20 and 1.41 as threshold respectively.

As expected, the lower the GDOP threshold, the smaller the spatial coverage and the fewer the velocity vectors.

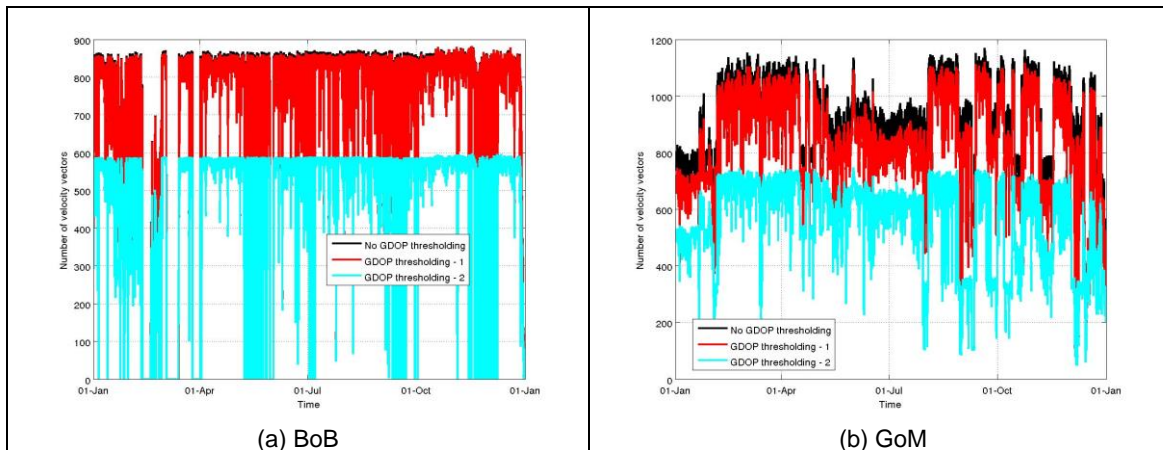


Figure 31 – Evolution in time of the number of velocity vectors present in the total velocity field with and without applying GDOP thresholds for BoB (a) and GoM (b) networks.

As for the previous test, a useful and simple check to do when choosing the threshold is to examine the evolution in time of the average velocity with and without the application of the candidate thresholds. In this way it is possible to have a quick overview of the effects of the test (that keeps only velocities whose GDOP is lower than the threshold) on the velocity field. Figure 32 reports this comparison and shows that GDOP threshold has a significant effect on the average total velocity since it is a very effective method to exclude areas with high current uncertainty and high inaccurate speeds.

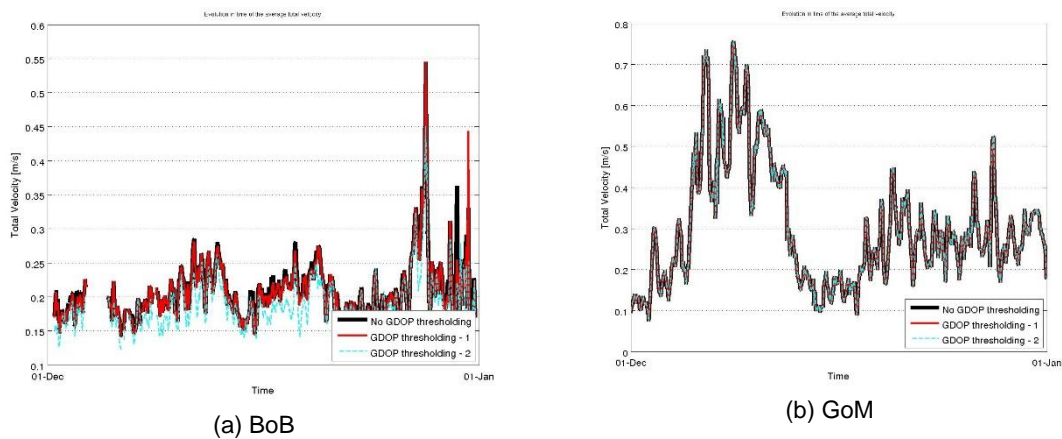


Figure 32 – Evolution in time (December 2014) of the average total velocity with and without the application of different GDOP thresholds for BoB (a) and GoM (b) networks.



## 8 Conclusions

HFRs offer the means to map ocean surface currents over wide areas and with a good temporal resolution. To make HFR information beneficial for the integrated management of coastal areas and ensure the HFRs play an increasing role in the overall operational oceanography marine services, the quality of the data has to be controlled operationally.

Two major steps towards optimized HFR data quality are the Quality Assessment of the measuring processes and the operational Quality Control of the data.

QA involves processes that are mostly employed with hardware. Since, HFRs are land-based remote sensing platforms, the operations of maintenance and restitution of hardware components can be relatively easy and less costly. However several aspects at different stages have to be taken into account from the beginning of the antenna installation to ensure the best performances of the equipment. The main QA procedures for HFR operation are today quite well defined. On this base, this document has provided an enriched description of the existing protocols and basic recommendations to follow.

QC involves follow-on steps that support the delivery of high quality data involving processes employed with software. While all the radars share the same principles of operation, differences in signal transmission, reception and processing yield variations in metadata, quality control metrics and spatial registration. Thus, in addition to the efforts made by the HFR community during the fruitful days of the INCREASE HFR Experts Workshop on defining the main QC procedures necessary to the integration of HFRs into CMEMS operational services, the aim of this document is to showcase the application of this procedures to real data in order to provide practical recommendations on their implementation.

Although the two chosen systems for the exercises are CODAR DF HFRs, a special effort has been done to consider the particularities of WERA HFRs. A detailed definition and a practical description of the different tests and how should they be computed for radial and total data is provided. The election of the required thresholds has proven not to be straightforward for most of the cases. The general recommendation is to examine the historical data time-series during at least one complete seasonal cycle for which the proper functioning of the device is assessed, and consult the literature. In addition, different useful diagnostics and approaches to guide the operators in the choice of adequate thresholds and for comparing the effects of different thresholding strategies are provided.

From the ensemble of tests checked for radial and total data in this work, the “Balance of contributing radial test” (QC test on total data) is the only one that has shown major technical drawbacks. Our analysis suggests that the application of this QC test is not straightforward, at least not using a unique threshold for the whole coverage area. Accordingly, our recommendation would be not to consider it as a mandatory test.



## 9 Bibliography

- D. E. Barrick, M. Evans, and B. Weber, 1977. Ocean surface currents mapped by radar, *Science*, vol. 198, no. 4313, pp. 138–144.
- Barrick, D.E. and B.J. Lipa, 1986: Correcting for distorted antenna patterns in CODAR ocean surface measurements. *IEEE J. Ocean. Eng.*, OE-11, 304-309.
- Corgnati, L.; Mantovani, C.; Griffa, A.; Bellomo, L.; Carlson, D.F.; Magaldi, M.G.; Berta, M.; Paziienza, G.; D'Adamo, R., 2015. "The ISMAR high frequency coastal radar network: Monitoring surface currents for management of marine resources," in *OCEANS 2015 - Genova*, vol., no., pp.1-8, 18-21 May 2015. doi: 10.1109/OCEANS-Genova.2015.7271524. WOS:000380485500206
- Fernandez D.M. and Paduan, J.D., 1996. Simultaneous CODAR and OSCAR measurements of ocean surface currents in Monterey Bay," in *Geoscience and Remote Sensing Symposium, 1996. IGARSS'96. Remote Sensing for a Sustainable Future. International*, vol. 3. IEEE, pp. 1749–1752.
- Gurgel, K-W., Antonischki, G., Essen, H-H. and Schlick, T., 1999. Wellen radar (WERA): a new ground-wave HF radar for ocean remote sensing, *Coastal Engineering*, vol. 37, no. 3, pp. 219–234.
- Kohut, J. T., Glenn, S. M., 2003. Improving HF radar surface current measurements with measured antenna beam patterns. *J. Atmos. Ocean. Tech.*, 20, 1303–1316, 2003.
- Laws K, Paduan J., Vescky J., 2010. Estimation and Assessment of Errors Related to Antenna Pattern Distortion in CODAR SeaSonde High-Frequency Radar Ocean Current Measurements, *Journal of atmospheric and oceanic Technology*, June 2010, DOI: <http://dx.doi.org/10.1175/2009JTECHO658.1>
- Poulain, P.-M., 2001. Adriatic Sea surface circulation as derived from the drifter data between 1990 and 1999, *J. Mar. Syst.*, 29, 3–32.
- Rubio A, Solabarrieta L, González M, Mader J, Castanedo S, Medina R, Charria G & Aranda JA, 2013. Surface circulation and Lagrangian transport in the SE Bay of Biscay from HF radar data. *OCEANS - Bergen*, 2013
- Solabarrieta L, Rubio A, Castanedo S, Medina R, Fontán A, González M, Fernández V, Charria G & Hernández C, 2014. Surface water circulation patterns in the southeastern Bay of Biscay: New evidences from HF radar data. *Cont. Shelf. Res.*, 74: 60-76.
- Stewart, R.H. and Joy, J.W., 1974. HF radio measurements of surface currents," in *Deep Sea Research and Oceanographic Abstracts*, vol. 21, no. 12. Elsevier, pp. 1039–1049.
- Taillandier, V., A. Griffa, P. M. Poulain, R. Signell, J. Chiggiato, and S. Carniel, 2008. Variational analysis of drifter positions and model outputs for the reconstruction of surface currents in the central Adriatic during fall 2002, *J. Geophys. Res.*, 113, C04004, doi:10.1029/2007JC004148.
- U.S. IOOS - Integrated Ocean Observing System, 2016a. Manual for Real-Time Quality Control of High Frequency Radar Surface Currents Data: A Guide to Quality Control and Quality Assurance of High Frequency Radar Surface



- Currents Data Observations. 57 pp.  
(<https://cordc.ucsd.edu/projects/mapping/documents/SCCOOS-BestPractices.pdf>)
- U.S. IOOS - Integrated Ocean Observing System, 2016b. Manual for Real-Time Quality Control of High Frequency Radar Surface Currents Data: A Guide to Quality Control and Quality Assurance of High Frequency Radar Surface Currents Data Observations. 58 pp. ([https://ioos.noaa.gov/wp-content/uploads/2016/06/HFR\\_QARTOD\\_Manual\\_05\\_26\\_16.pdf](https://ioos.noaa.gov/wp-content/uploads/2016/06/HFR_QARTOD_Manual_05_26_16.pdf))