UNIVERSITÀ DEGLI STUDI DI GENOVA
SCUOLA POLITECNICA
DICCA
Department of Civil, Chemical and Environmental Engineering

MASTER THESIS
IN
ENVIRONMENTAL ENGINEERING

Development and validation of methods for quality control and gap filling in high-frequency coastal radar datasets from the northwestern Mediterranean

Supervisors:
Prof. Marcello G. Magaldi
Prof. Giovanni Besio
Co-supervisors:
Dott. Roberta Sciascia
Dott. Alberto Baudena

Candidate:
Mattia Codrino

ACADEMIC YEAR 2025-2026

Development and validation of methods for quality control and gap filling in high-frequency coastal radar datasets from the northwestern Mediterranean

Abstract

High-frequency coastal radars (HFR) deliver continuous, high-resolution measurements of the ocean surface currents and are useful for applications ranging from Search and Rescue operations to wider coastal monitoring. However the operational value of HFR data relies on two critical issues: the presence of physically inconsistent vectors that have to be filtered out by means of quality control (QC), and the discontinuities in spatial and temporal coverage caused by bad sea conditions and radio-frequency interference. This thesis addresses both issues for the HFR-TirLig network, the monitoring of surface currents in the north-western Mediterranean Sea. First, the existing QC pipeline was revised, tightening the thresholds on the CSPD and VART filters, and adding a new spatial-derivative filter, MEDF, calibrated against drifter-derived velocity statistics. The Progressive Rim Gap-Filling Method was reformulated into separate temporal and spatial phases and evaluated with an artificial-hole system. Realistic gaps were imposed on complete data and the reconstructed values were compared to the actual observations using the coefficient of determination (R^2). The results show that the spatial stage has a better reconstruction ability for a wider range of gap sizes than the temporal stage, which rapidly degrades beyond its 6-hour persistence limit, providing a practical criterion for selecting the appropriate interpolation mode.

Sviluppo e validazione di metodi per il controllo di qualità e gap filling dei dati radar costieri ad alta frequenza nel Mediterraneo nord-occidentale

Sommario

I radar costieri ad alta frequenza (High-Frequency Radars, HFR) forniscono misure continue e ad alta risoluzione delle correnti superficiali marine e rappresentano uno strumento fondamentale per applicazioni che spaziano dalle operazioni di Ricerca e Soccorso (Search and Rescue) al monitoraggio costiero su larga scala. Tuttavia, il valore operativo dei dati HFR dipende da due aspetti critici: la presenza di vettori fisicamente inconsistenti, che devono essere rimossi mediante procedure di controllo qualità (Quality Control, QC), e le discontinuità nella copertura spaziale e temporale causate da condizioni del mare avverse e da interferenze a radiofrequenza. Questa tesi affronta entrambe le problematiche per la rete HFR-TirLig, dedicata al monitoraggio delle correnti superficiali nel Mediterraneo nord-occidentale. In primo luogo, è stata rivista la pipeline di controllo qualità esistente, rendendo più restrittive le soglie dei filtri CSPD e VART e introducendo un nuovo filtro basato sulle derivate spaziali, denominato MEDF, calibrato utilizzando statistiche di velocità derivate da misure di drifter. Successivamente, il metodo Progressive Rim Gap-Filling è stato riformulato separando le fasi di interpolazione temporale e spaziale, ed è stato valutato mediante un sistema di buchi artificiali. Sono state infatti imposti buchi realistici su dati completi e i valori ricostruiti sono stati confrontati con le osservazioni originali utilizzando il coefficiente di determinazione (R^2). I risultati mostrano che la fase di interpolazione spaziale offre una migliore capacità di ricostruzione per un intervallo più ampio di dimensioni dei buchi, mentre la fase temporale degrada rapidamente oltre il limite di persistenza di 6 ore, fornendo così un criterio pratico per la scelta della modalità di interpolazione più appropriata.

Contents

List of Figures

List of Tables

1. Introduction

Coastal seas are among the most active and important areas of the ocean. Monitoring the surface circulation in these regions with proper spatial coverage and time resolution is essential for both scientific understanding and managing the coastal ocean (Lorente, Aguiar, et al. 2022; Rubio et al. 2017). Of the remote-sensing technologies available, High-Frequency Radar (HFR) has become one of the most widely used systems because it can provide near-real-time maps of surface current velocity over areas ranging from tens to hundreds of kilometers (Paduan and Washburn 2011; Roarty, Cook, et al. 2019).

The benefits of HFR observations go beyond serving as a scientific dataset for studying coastal dynamics (Rubio et al. 2017). HFR-derived current fields are frequently used in various environmental and societal applications. These include reconstructing the transport of pollutants to support fishery management and protect ecosystems, helping to mitigate oil spill impacts by modeling how pollutants spread and settle, assessing how port and coastal activities affect marine protected areas, monitoring extreme events, detecting meteotsunamis, and assisting maritime security operations, such as vessel navigation and Search and Rescue (SAR) (Corgnati, Maristella Berta, et al. 2024; Lorente, Aguiar, et al. 2022; Rubio et al. 2017).

Beyond the safety and pollution applications, for ecological research HFR derived current fields can be used to study marine larval dispersal. Larval dispersal is a key process affecting the connectivity of populations, recruitment success, genetic exchange and long-term viability of marine species. Since many marine organisms spend an early stage of their life cycle as planktonic larvae, the physical transport occurring during this stage is a key factor in the dispersal between geographically isolated populations, and as such in the efficacy of Marine Protected Area (MPA) networks and the general conservation strategies that accompany them. Larval dispersal experiments have thus become a principal tool for the study of ecological connectivity and for the development of ecosystem-based management approaches (Jones, Srinivasan, et al. 2007; Cowen et al. 2009; Planes et al. 2009; Jones, Almany, et al. 2009). Accurate data on ocean currents are vital for this type of study, as larval transport is significantly dictated by ocean circulation. Surface currents determine the routes that larvae take, the habitats they occupy and the sites where they ultimately settle, so realistic dispersal simulations require detailed hydrodynamic data. Thus high-resolution current fields are routinely coupled with particle-tracking models to simulate connectivity patterns and recruitment dynamics. Results of the larval dispersal simulation are very sensitive to the quality of the hydrodynamic forcing used, so that uncertainties or deficiencies of the underlying current data can be directly translated into the ecological assessment and management decisions derived from the simulations (Staaterman et al. 2014; Swearer et al. 2019; Roberta Sciascia, Guizien, et al. 2022; Choukroun et al. 2024).

The SAR application of HFR derived currents is particularly important. It highlights why the reliability and completeness of such fields are critical, as they can directly impact human lives (Roarty, Updyke, et al. 2024). Maritime search planning typically relies on Lagrangian trajectory models. These models predict where a missing vessel, object, or person in water might be based on surface currents, wind drift (leeway), and, when applicable, wave-induced drift (Breivik et al. 2012; Di Maio et al. 2016). The precision of these predictions depends on the quality of the surface current field used to drive the drift model. As the search area expands over time and with the uncertainty of transport estimates, a reliable, continuous current field can significantly reduce the area rescue teams need to search, increasing the chances and speed of recovery (Reyes et al. 2022; Roarty, Cook, et al. 2019). In 2019, the Italian Coast Guard was involved in 1,875 SAR missions (Reyes et al. 2022). This illustrates the scale of this application along the Italian coastline, including the Ligurian and northern Tyrrhenian coasts monitored by the HFR-TirLig network discussed in this work.

The completeness and reliability of the HFR-derived current field drive the two methodological goals of this thesis. On one side, HFR total velocity products are affected by erroneous or non-physical vectors that arise from frequency interference or extreme weather events (Gauci et al. 2016). If these errors are not corrected, they may be mistaken for real oceanographic signals in any downstream application, including drift and trajectory models for SAR planning (Roarty, Updyke, et al. 2024). The impact of these errors is not limited to their origin; a single flawed vector can spread its error throughout the entire reconstructed field when used as input for spatial or temporal interpolation. Therefore, a rigorous Quality Control (QC) process is essential to identify and remove these artifacts while preserving as much genuine physical signal as possible before any subsequent use of the data (Roarty, Updyke, et al. 2024; Mantovani et al. 2020).

On the other side, even after quality control, HFR coverage is inevitably discontinuous in space and time (Hernández-Carrasco et al. 2018; Fredj et al. 2016). Gaps can arise for various environmental and technical reasons, like adverse sea conditions or radio-frequency interference (Fredj et al. 2016; Gauci et al. 2016). Since such gaps are unavoidable over long times and extensive areas, an operational HFR network should provide, along with raw and quality-controlled data, a gap-filled product suitable for practical and scientific use (Hernández-Carrasco et al. 2018; Fredj et al. 2016; Lorente, Aguiar, et al. 2022). This is crucial for Lagrangian applications, such as trajectory reconstruction for SAR or oil spill response, which need continuous velocity fields rather than isolated, possibly incomplete snapshots (Hernández-Carrasco et al. 2018; Reyes et al. 2022). Thus, developing and validating a reliable, efficient gap-filling method, along with a solid framework to assess its performance under controlled and realistic gap conditions, is the key objective of this work.

This thesis addresses both aspects of the problem for the HFR-TirLig network, which monitors surface currents along the Ligurian and northern Tyrrhenian coasts of the northwestern Mediterranean Sea. Chapter 2 describes the area of interest and the data used in this work, and includes the operating principles of HFR systems and the radar and drifter data used for validation. In Chapter 3 the quality control filters applied to the radar dataset in its original and updated form are discussed. In Chapter 4 the gap-filling methodology tested in this work and the procedure adopted to generate controlled artificial data gaps for its validation are presented. Chapter 5 presents the results obtained on the robustness of the gap-filling methodology and their implications for the operational use of the HFR-TirLig network.

2. Area of interest and Data

2.1 Area of Interest

The study area of interest is located in the Northwestern Mediterranean (NW Med), between 8.13°E–10.33°E longitudes and 43.47°N–44.43°N latitudes. This area comprises a transitional coastal zone at the boundary between the eastern Ligurian sea (LS) and the northern Tyrrhenian sea (TYS) connected by the Corsica Channel, one of the most dynamically relevant straits in the western Mediterranean basin (Angeletti et al. 2020).

The Ligurian Sea is a marginal sea of the Mediterranean Sea that is semi-enclosed, bordered on the north and east by the Italian regions of Liguria and Tuscany and on the south by the island of Corsica. It is one of the most oceanographically active sub-basins of the NW Mediterranean, characterized by a cyclonic large scale circulation active in the entire year (Millot 1999). The basin has a distinct morphological asymmetry, with the western part being different from the eastern part (Enrichetti et al. 2019). The western zone is characterized by a northeast-southwest orientation of the coastline, a narrow continental shelf and high bathymetric depths (>2,500 m) (Corgnati, Maristella Berta, et al. 2024). The eastern zone, characterized by a wide Tuscan shelf, is relatively shallow and has an axial southeast-northwest direction (Enrichetti et al. 2019). The Corsica Channel, with a sill depth of about 450 m, is located at the southern limit of the LS and plays a key role in the transfer of water masses between the Tyrrhenian and Ligurian basins (M. Astraldi, Gasparini, et al. 1990).

The Northern Current (NC), or Ligurian-Provençal Current or Ligurian Current, is the main large-scale circulation feature of the region. It is the northern branch of the large cyclonic gyre typical of the NW Mediterranean and flows westward along the continental slope (M. e. a. Astraldi 1995). The NC is formed north of the Corsica Channel by the convergence of two different currents: the Eastern Corsica Current (ECC) or Tyrrhenian Current (TC) flowing northward along the eastern coast of Corsica and through the Corsica Channel and the Western Corsica Current (WCC) flowing along the western coast of Corsica (Roberta Sciascia, Magaldi, et al. 2018). Both currents have a seasonal behavior, the TC being the most variable one. The transport through the Corsica Channel is maximum in winter when the northward current is significantly enhanced and it almost vanishes in summer (M. Astraldi, Gasparini, et al. 1990; M. Astraldi and Gasparini 1992). This seasonality is caused by a steric sea-surface gradient driven by density differences between the Ligurian basin and the Tyrrhenian basin (M. Astraldi and Gasparini 1992).

The NC has considerable mesoscale variability in the form of meanders, eddy generation and lateral shifts of its core, features especially prominent during the more energetic cold season (Barth et al. 2005; Sammari et al. 1995). This variability is not limited to occasional meandering episodes but reflects a persistent dynamical regime of the basin. Numerical modeling based on a two-way nested model showed that the NC undergoes continuous meandering and frequently interacts with the Winter Intermediate Water lenses from the open basin; this interaction, together with the effect of the basin irregular topography, was determined as a major factor responsible for the large spatial variability of the current (Barth et al. 2005). An additional description of this variability on seasonal and mesoscale time scales is presented through direct current-meter and hydrographic observations from the field experiments PROLIG-2 and PROS-6, carried out in the western Mediterranean from May to December 1985 (Sammari et al. 1995). This highly energetic fluctuating boundary current has a direct impact on the reconstruction of missing HFR data discussed later in this thesis.

The dynamics of the area of interest are driven by the combination of large scale boundary current forcing and local coastal processes. The Arno River is a major river of central Italy, 241 km long, with a drainage basin of roughly 8,200 km^2 in the Tuscan Apennines; its discharge is a major source for the density gradients and coastal frontal structures in the near-shore waters of the northern Tyrrhenian shelf (Ludwig et al. 2009). In addition, wind-induced phenomena may drive the NC core and its meanders toward the Italian coast, increasing the cross-shore transport of momentum and tracers (Alberola C 1995).

2.2 HF Radars

High-frequency radar has evolved from a niche research instrument into one of the most widely deployed technologies for continuous, remote observations of ocean surface currents during the past few decades. These instruments are fixed at the coastline and remotely sense surface current fields over broad coastal domains, generally extending 30 to 200 km offshore, at spatial resolutions of about 0.2 to 6 km and temporal resolutions ranging from 15 minutes to 1 hour, depending mainly on their operating frequency (Corgnati, Maristella Berta, et al. 2024). Their capability to supply synoptic, near-real-time maps of surface circulation over such large areas has made them an increasingly vital component of coastal ocean observing systems around the world (Corgnati, Maristella Berta, et al. 2024; Roarty, Cook, et al. 2019).

HF radars rely on resonant backscatter resulting from coherent reflection of the transmitted wave by the ocean waves whose wavelength is $1/2$ of that of the transmitted wave (Rubio et al. 2017). The energy reflected at one wave crest is in phase with the emitted wave (Paduan and Graber 1997). This is the Bragg scattering phenomena and it results in the 1st order peak of the received spectrum (Paduan and Graber 1997). In the absence of currents, the frequency of the 1st order peak has a Doppler shift caused by the phase velocity of the waves in the radial direction of the transmitting antenna (Rubio et al. 2017). This speed (V_g) is known since it is the wave propagation speed and is given by the dispersion relation in deep waters.

The Doppler shift is presented in the two peaks of the spectrum of the received signal, symmetrically situated on either sides of the frequency (Rubio et al. 2017). If gravitational waves are propagated over a current field, the peaks are shifted in the frequency domain and an asymmetric spectrum is obtained (Rubio et al. 2017). The radial component of the current, which is the current in the same direction of the signal, can be directly calculated from the speed difference between the theoretical speed of the waves, for given wavelength, and the velocity observed through the spectrum; the difference is the speed of the current (Rubio et al. 2017).

The Doppler frequency shift of the two first-order Bragg peaks, ω_B, follows the general relation:

$$\omega_B = \kappa_B V, \tag{2.1}$$

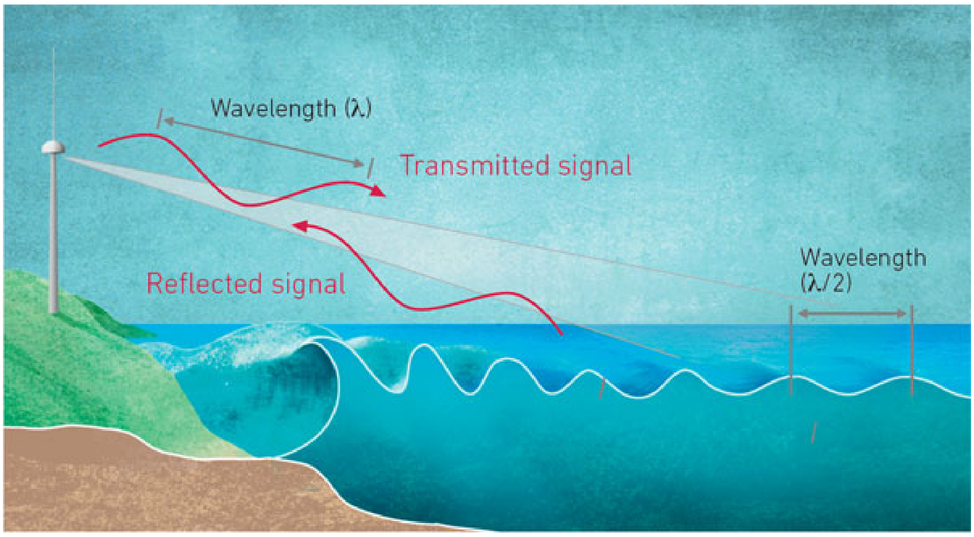

Figure 2.1: Conceptual diagram of Bragg backscatter. A coastal HF radar antenna transmits an electromagnetic wave (red arrow, "Transmitted signal") toward the sea surface, where it resonates with ocean waves of half its wavelength - brackets mark "Wavelength (λ)" for the radar signal and "Wavelength (λ/2)" for the resonant surface waves - before returning to the antenna as the "Reflected signal" (second red arrow). (European HFR Node 2026c)

where κ_B is the Bragg wavenumber and V is the phase speed of the Bragg-resonant wave relative to the antenna (Laws 2001). This phase speed can be decomposed as follows:

$$V = V_g + \Delta V, \tag{2.2}$$

where V_g is the still-water phase speed of the Bragg wave, given by the deep-water dispersion relation:

$$V_g = \sqrt{\frac{g\lambda}{2\pi}} = \sqrt{\frac{g}{k}} \tag{2.3}$$

where λ is the wavelength, g the gravitational acceleration (Paduan and Graber 1997), and ΔV is a shift due to the near-surface current field (Laws 2001). Equation (2.2) formalizes the principle described above: the still-water phase speed V_g is known a priori from the dispersion relation, so comparing it with the phase speed V obtained from the observed Doppler shift yields directly ΔV, i.e. the radial component of the near-surface current.

The shift in phase speed ΔV corresponds to a depth-weighted average of the horizontal current profile $u(z)$, with z taken positive upward from the mean sea surface (Stewart et al.

1974):

$$\Delta V = 2\kappa_B \int_{-\infty}^{0} u(z)\, e^{2\kappa_B z}\, dz. \tag{2.4}$$

Equation (2.4) shows that the radial current retrieved from the Doppler shift is not simply the current at the sea surface, but an exponentially weighted average of the current over the water column, with the weighting determined by the Bragg wavenumber κ_B.

A single HFR station can only estimate the radial component of the current. To recover the full two-dimensional current vector, one needs to combine radial measurements from at least two stations observing the same patch of ocean from sufficiently different viewing angles (Mantovani et al. 2020). At the region where coverage areas of two or more stations overlap, the radial velocity contributions are projected on a common Cartesian grid and combined by means of an unweighted least-squares fit, providing at each grid point the two horizontal components of the total surface current vector (Mantovani et al. 2020).

The surface current maps obtained from HFR systems are not free from spurious or physically implausible measurements. Knowing their origin is a necessary premise for the quality control procedures discussed in chapter 3.

A first source of error is related to the direction-finding technique itself: systems such as the Codar SeaSonde estimate the direction of arrival of the backscattered signal through the MUlti-ple SIgnal Classification (MUSIC) algorithm, whose accuracy depends critically on an accurate characterization of the receiving antenna pattern (Corgnati, Maristella Berta, et al. 2024).

There is another source of anomalous vectors: geometric. Because the total current vectors are assembled from radial measurements of the same patch of ocean from at least two stations viewing from different angles, the accuracy of the reconstructed vector is highly sensitive to the angle of intersection of the contributing radials: as this angle approaches 0° or 180°, the geometry becomes ill-conditioned, and small errors in the radial velocities are amplified into large, often non-physical errors in the total vector, a sensitivity commonly expressed by the Geometric Dilution of Precision (GDOP) (Corgnati, Maristella Berta, et al. 2024).

Finally, external environment and instrumental effects, which are outside the processing algorithm, can degrade or corrupt the underlying radial measurements before they even reach the combina-tion stage. Erroneous or missing vectors in the raw data stream can be caused by radio-frequency interference, adverse sea-state conditions and intermittent hardware and communication failures, ranging from antenna malfunctions to network connectivity issues between individual stations and the processing center (Corgnati, Maristella Berta, et al. 2024).

Taken together, these three categories explain why an HFR total-velocity dataset can occasionally contain vectors with unrealistic magnitudes or directions, and motivate the systematic, multi-test quality control pipeline described in the following chapter.

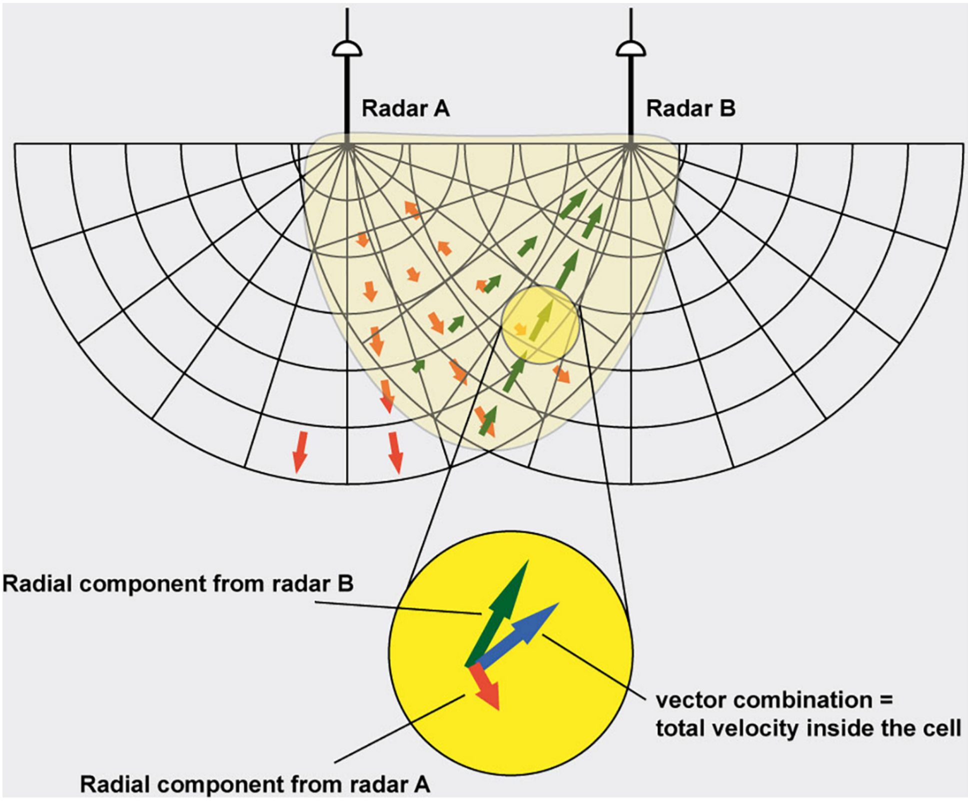

Figure 2.2: Schematic of the radial-to-total current reconstruction for a two-station HFR network. Each station (Radar A, Radar B) measures the radial velocity component ; only within the shaded overlap region, where the two grids intersect, do both stations sample the same patch of ocean and can a total current vector be reconstructed. Orange arrows represent the radial component measured by Radar A and green arrows the radial component measured by Radar B. The detail of one grid cell (highlighted in yellow) shows how the two radial components are combined through the least-squares fit into the total surface current vector for that cell (blue arrow). ((Mantovani et al. 2020))

2.3 HFR System

The HFR-TirLig network is part of the CNR-ISMAR network and is built of 6 operational CODAR Seasonde HF Radar systems, which are the following: LIGW in Celle Ligure, MONT in Monterosso al Mare, PCOR in Punta Corone, PFIN in Portofino, TINO in Isola del Tino, VIAR in Viareggio (CNR-ISMAR 2016). All the operational stations are collecting hourly surface currents and waves measurements along the coast of the North-Western Tyrrhenian Sea and Ligurian Sea. The operational network is composed by some stations emitting at central frequency of 13.5 MHz and some at 26.275 MHz, providing hourly radial current data with respectively 1.5 km and 1 km range resolution and radial coverage close respectively to 80 km and 40 km (Mantovani et al. 2020). The combination of radial current measurement provides hourly total current data over a regular grid with 2 km spatial resolution, over a total coverage of approximately 9,000 km^2 (European HFR Node 2026a).

Total velocities are derived using least square fit that maps radial velocities measured from individual sites onto a cartesian grid. The final product is a map of the horizontal components of ocean currents on a regular grid in the area of overlap of two or more radar stations. This results in a dataset consisting of maps of total velocity of the surface current in the Northwestern Tyrrhenian and Ligurian Sea averaged over a time interval of 1 hour around the cardinal hour. The surface ocean velocities estimated by the HF Radar are representative of the upper 0.3-2.5 meters of the ocean (Rubio et al. 2017).

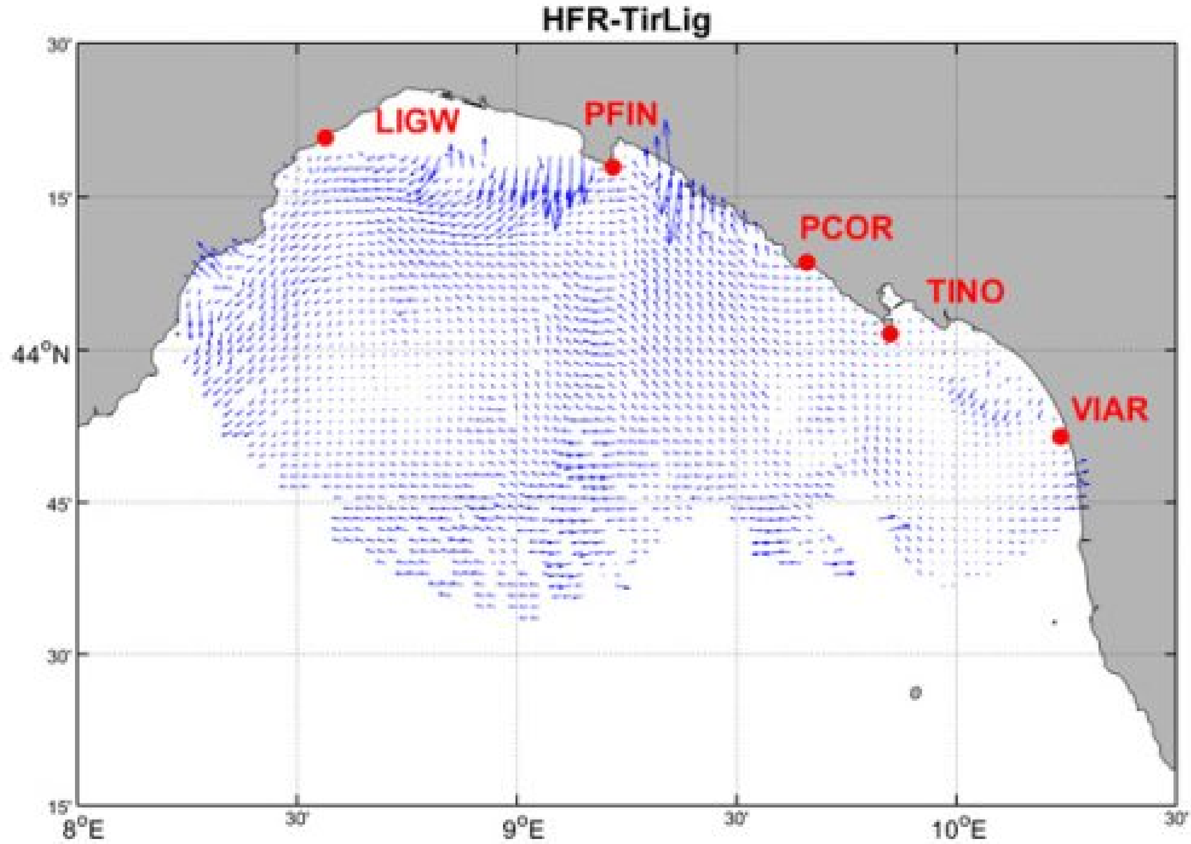

Figure 2.3: Map of the HFR-TirLig network domain along the Ligurian/northern Tyrrhenian coast (8°–10°E, 43.75°–44.5°N). The five red dots mark the radar stations: LIGW (Celle Ligure), PFIN (Portofino), PCOR (Punta Corone), TINO (Isola del Tino) and VIAR (Viareggio). (European HFR Node 2026b)

2.4 Radar Data

The dataset used in this study comprises the near-real-time total surface current velocity fields generated by the HFR-TirLig network and distributed by the European HFR Node infrastructure (Lorente, Aguiar, et al. 2022; Corgnati, Maristella Berta, et al. 2024). The data were downloaded from the ERDDAP data server of the European HFR Node, a web-based interface that allows standardized access to gridded scientific datasets through the OPeNDAP/griddap protocol, using the dataset EUHFR_NRTcurrent_HFR-TirLig-Total_v3 (CNR-ISMAR 2016). ERDDAP (Environmental Research Division's Data Access Protocol) is a data server that provides a uniform interface to access scientific datasets in a variety of formats.

The dataset provides gridded maps of the horizontal ocean surface current velocity with an hourly temporal resolution (Roarty, Cook, et al. 2019; Corgnati, Mantovani, et al. 2018) for the north-western Tyrrhenian Sea and eastern Ligurian Sea (Corgnati, Maristella Berta, et al. 2024). The total velocity fields, described in section 2.3, are produced by integrating radial measurements from individual CODAR SeaSonde stations (Corgnati, Maristella Berta, et al. 2024) over a regular Cartesian grid with a spatial resolution of 2 km (Roberta Sciascia, Magaldi, et al. 2018). The grid covers a longitudinal extent from 8.13°E to 10.33°E and a latitudinal extent from 43.47°N to 44.43°N, covering an area of approximately 9,000 km^2. Each grid point provides an estimate of the horizontal velocity vector of the ocean surface current which is representative of the upper 0.3–2.5 meters of the water column (CNR-ISMAR 2016).

The main physical variables of interest in the dataset are the eastward (EWCT) and northward (NSCT) (Roarty, Cook, et al. 2019) components of the total current velocity vector (m/s). In addition to these two velocity components, the dataset contains a set of ancillary variables required for the quality control (QC) process (Mantovani et al. 2020; Roarty, Cook, et al. 2019) performed in this paper. These are the current QC flag variables applied to each individual test, GDOP_QC, CSPD_QC, VART_QC, DDNS_QC, POSITION_QC, and the global QC flag filter. The objective and rationale for these filters are described in more detail in chapter 3.

The dataset contains a temporal dimension in the form of timestamps rounded to the nearest hour (Corgnati, Mantovani, et al. 2018; Roarty, Cook, et al. 2019). Slight differences on the sub-minute level can sometimes be observed in the near real-time processing pipeline, therefore the timestamps were carefully checked and rounded to ensure temporal consistency across all time steps within the data retrieval workflow implemented here. The spatial dimensions are a one-dimensional latitude vector and a one-dimensional longitude vector, whose Cartesian product defines the two-dimensional horizontal grid. The dataset includes a depth dimension, but since the depth has only one level, the surface, it is not loaded (CNR-ISMAR 2016) to reduce the dimensionality without any loss of information.

One year was used as the time frame for this study, starting from 10 November 2025 and covering the previous 365 days, thus representing a full annual cycle of oceanographic variability (Roberta Sciascia, Magaldi, et al. 2018). The chosen time span allows for a statistically significant evaluation of the QC filters and the gap-filling procedure over the whole range of seasonal forcing conditions typical of the northwestern Mediterranean (Roberta Sciascia, Magaldi, et al. 2018).

The dataset is stored in the NetCDF-4 file format , following the CF (Climate and Forecast) metadata conventions and the European standard data model for HFR products (Roarty, Cook, et al. 2019; Corgnati, Mantovani, et al. 2018) with standardized variable names, units and coding schemes for quality flags (Roarty, Cook, et al. 2019). This interoperability allows programmatic access and integration (Mantovani et al. 2020; Lorente, Aguiar, et al. 2022) with standard scientific python libraries.

2.5 Drifter Data

Surface drifters are autonomous oceanographic instruments used to track the movement of water masses at or near the surface of the ocean. Position fixes are obtained at regular intervals, and the velocity of the drifter and the adjacent water parcel is calculated as the finite difference of consecutive position estimates divided by the elapsed time. This leads to a Lagrangian description of the flow field rather than the Eulerian, spatially fixed description provided by HFR systems (Röhrs et al. 2015). The complementary nature of drifter observations makes them a useful benchmark for the validation of radar-derived surface current products (Doronzo et al. 2025).

In this study, we use two drifter datasets collected during two different oceanographic campaigns in the Ligurian–Tyrrhenian region in 2019 and 2020. The dataset was provided in a pre-processed tabular file (vel_drifters_2019_2020_an58.parquet) containing the drifter identifier, geographic coordinates (longitude and latitude), timestamp, total speed magnitude, and eastward and northward components of velocity for each observation. The dataset consists of 272,450 observations from 60 unique drifter trajectories.

The first campaign took place between May and June 2019 with 40 drifters released simultaneously on 2 May 2019 (M. Berta, R. Sciascia, et al. 2020). The second campaign covers October 2020 to January 2021 with 20 drifters deployed on 8 October 2020 (M. Berta, Poulain, et al. 2021). The two campaigns together provide Lagrangian observations in contrasting seasonal conditions, late spring and autumn/early winter, including different circulation regimes in the north-western Mediterranean (Roberta Sciascia, Magaldi, et al. 2018; Roberta Sciascia, Guizien, et al. 2022).

Drifter trajectories cover a significant area of the Ligurian and northern Tyrrhenian Seas, with longitudes stretching from about 3.1°E to 10.5°E and latitudes between 39.5°N and 44.4°N. The drifter area of observation is quite similar to the area of the HFR-TirLig network, making it easy to visually overlay drifter information with radar observation locations. The positions of the drifters are fixed every 5 minutes, thereby allowing finer resolution when using the drifters data, since the HFR reports total currents on an hourly basis (Paduan and Washburn 2011; Doronzo et al. 2025). In order to compare with the radar data, drifter velocities are therefore temporally averaged to the HFR time grid.

3. QC Filters

3.1 Old QC Filters

The quality control (QC) pipeline used for the total velocity data from the HFR-TirLig network consists of a series of automated tests, each targeting a specific type of measurement error or observation artifact. The filters discussed in this section reflect the existing configuration and thresholds before any of the updates that will be later discussed.

The first test is the Geometric Dilution of Precision (GDOP) threshold filter. A low GDOP indicates a well-structured arrangement of radar stations, leading to accurate surface current data (Lorente, Piedracoba, et al. 2019; Corgnati, Mantovani, et al. 2018). Setting a threshold on GDOP for total combinations prevents the mixing of radials with intersection angles below a certain limit. GDOP effectively filters unreliable velocities due to poor geometry (Corgnati, Mantovani, et al. 2018). The QC filter marks all total vectors with a GDOP below a set threshold as "good data" (Mantovani et al. 2020). Vectors with a GDOP above that threshold are marked as bad data and excluded from further processing. In the initial configuration, the maximum GDOP threshold is set at 2.0 (Corgnati, Mantovani, et al. 2018; Corgnati 2024).

The second test is the Velocity Threshold filter (CSPD). This filter works on the current speed of the total velocity vector. Any total vector exceeding a threshold speed is labeled as non-physical and marked as bad data (Mantovani et al. 2020). In the original configuration, the threshold is 1.2 m/s. (Corgnati, Mantovani, et al. 2018; Corgnati 2024)

The third test is the Temporal Derivative filter (VART). For each total vector at a specific spatial location, this test calculates the absolute difference in velocity magnitude between the current measurement and the one from the same grid point in the previous time step. If this difference goes beyond a set threshold, the vector is flagged as bad data; otherwise, it receives a good data flag (Mantovani et al. 2020). In the original configuration, the threshold is 1.2 m/s (Corgnati 2024).

The fourth test is the Data Density Threshold filter (DDNS). The reliability of a total vector heavily depends on the number of individual radial velocity contributions used to calculate it (Fang et al. 2015). The DDNS test marks a total vector as good data only if it is supported by a minimum number of contributing radials, in line with the threshold established for normal operations; those below this threshold are flagged as bad data (Mantovani et al. 2020). In the original setup, the threshold is set at 3 contributing radial velocities (Roarty, Updyke, et al. 2024; Corgnati 2024).

The fifth test is the POSITION check. It ensures that the geographical coordinates linked to each vector are valid (Roarty, Updyke, et al. 2024).

Finally, the overall QC flag is formed by combining the results of all the previous tests. A vector receives a passing overall flag only if all individual tests return a "good data" result. If any test flags "bad data", the overall flag marks a failure (Corgnati 2024).

Table 3.1: Old Quality Control pipeline configuration

Short name	**Long name**	**Old threshold**
GDOP	GDOP Threshold	> 2.0
CSPD	Velocity Threshold	> 1.2 m/s
VART	Temporal Derivative	> 1.2 m/s
DDNS	Data Density Threshold	< 3
POSITION	Position	-
$QCflag$	Overall	-

3.2 New QC Filters

3.2 .1 Importance of QC

Applying quality control procedures to the HFR total velocity data is essential for any subsequent analysis or operational use of the dataset. This is especially important when the data is used for gap-filling methods. The underlying idea is that the observations provided as input should be physically consistent and represent actual ocean surface dynamics. Non-physical or erroneous vectors in the input can skew the fitting solution, corrupting the interpolated fields throughout the domain and possibly spreading errors beyond the location of the initial incorrect measurement (Corgnati, Maristella Berta, et al. 2024). Thus, the integrity of the gap-filled product directly relies on the quality of the observations it is built upon.

Figure 3.1 shows the vectors filtered out by the old overall QC flag. This filter inherits the limitations of its components and is not conservative enough. As shown in the figure, a significant number of questionable vectors pass through the full QC pipeline. These include vectors along the western Ligurian coastline that show magnitudes and directions inconsistent

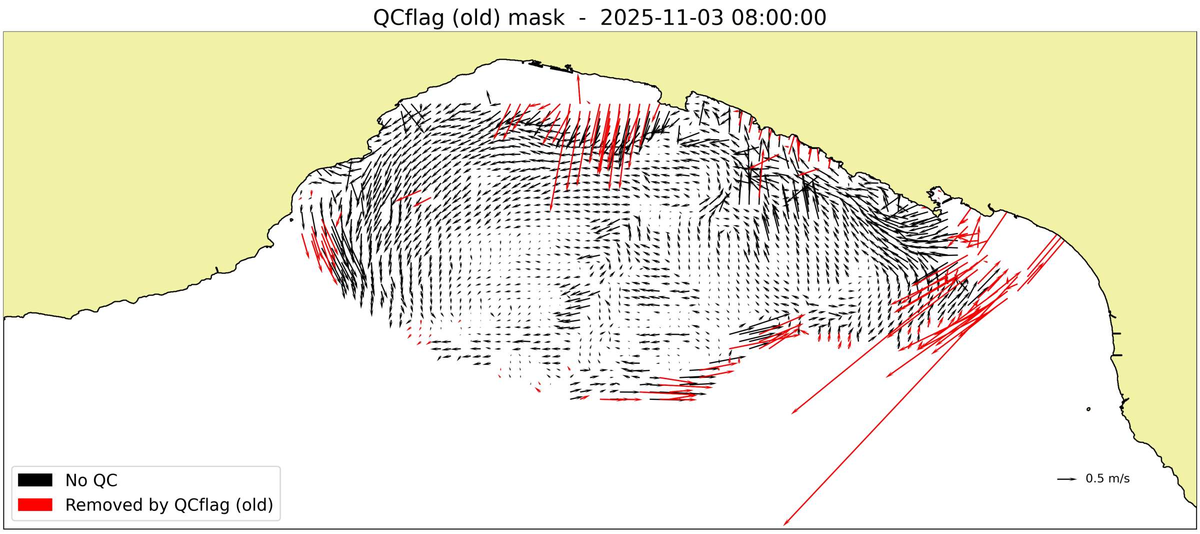

Figure 3.1: Old QCflag mask - Velocity field on 03-11-2025 at 8:00: the length of each arrow is proportional to the velocity field magnitude in that point, while the arrow direction shows the sum of the two components of the velocity field. Red arrows show points that didn't pass at least one of the tests included in the old QCflag, while black arrows show points that passed every test. The old-configuration removals are concentrated near the coast west of Portofino and in the southeastern portion.

with the surrounding flow field, along with persistent incorrect signals in the offshore south-eastern area. If these leftover artifacts remain uncorrected, they could undermine any further gap-filling analysis by introducing biases into the fitted solution.

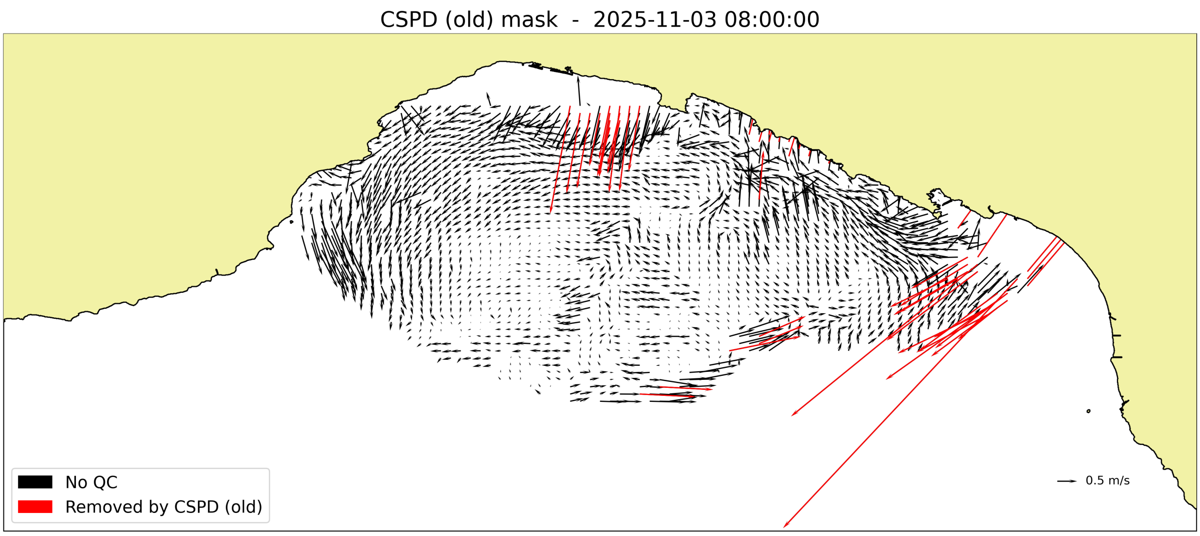

Figure 3.2: Old CSPD mask - Velocity field on 03-11-2025 at 8:00: the length of each arrow is proportional to the velocity field magnitude in that point, while the arrow direction shows the sum of the two components of the velocity field. Red arrows show points whose velocity field magnitude is larger than 1.2 m/s (the old threshold of the CSPD quality control), while black arrows with values lower than 1.2 m/s. The old-threshold removals are concentrated near the coast west of Portofino and in the southeastern portion.

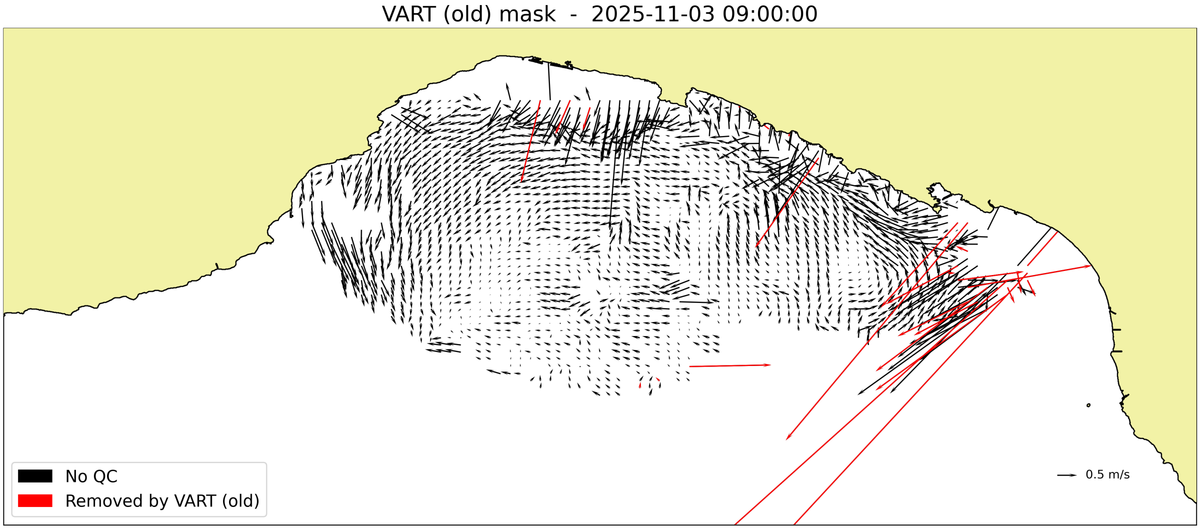

Figure 3.3: Old VART mask - Velocity field on 03-11-2025 at 9:00: the length of each arrow is proportional to the velocity field magnitude in that point, while the arrow direction shows the sum of the two components of the velocity field. Red arrows show points whose velocity field magnitude is at least 1.2 m/s (the old threshold of the VART quality control) larger than their magnitude at the previous timestep, while black arrows with differences lower than 1.2 m/s. The old-threshold removals are concentrated near the coast west of Portofino and in the southeastern portion.

3.2 .2 CSPD

The first test that was decided to be reworked was the CSPD test. In particular, it was decided to change the threshold used for this test, originally equal to 1.2 m/s. To do so, first, using drifter data, the PDF of the velocity was plotted, as shown in Figure 3.4. Since the drifter data doesn't have the same temporal scale as the radar data being used, first it was averaged every hour and then the velocity V was calculated with this formula:

$$V_i = \sqrt{u_i^2 + v_i^2} \tag{3.1}$$

where u is the EWCT velocity component and v is the NSCT velocity component.

In addition, to better study the data, three percentiles were plotted, corresponding to the three-sigma-rule values. The three-sigma-rule states that for a normal distribution, about 68.27% of values drawn from a normal distribution are within one standard deviation away from the mean; about 95.45% of the values lie within two standard deviations; nearly all values (99.73%) lie within 3 standard deviations of the mean. The values that exceed the three devi-

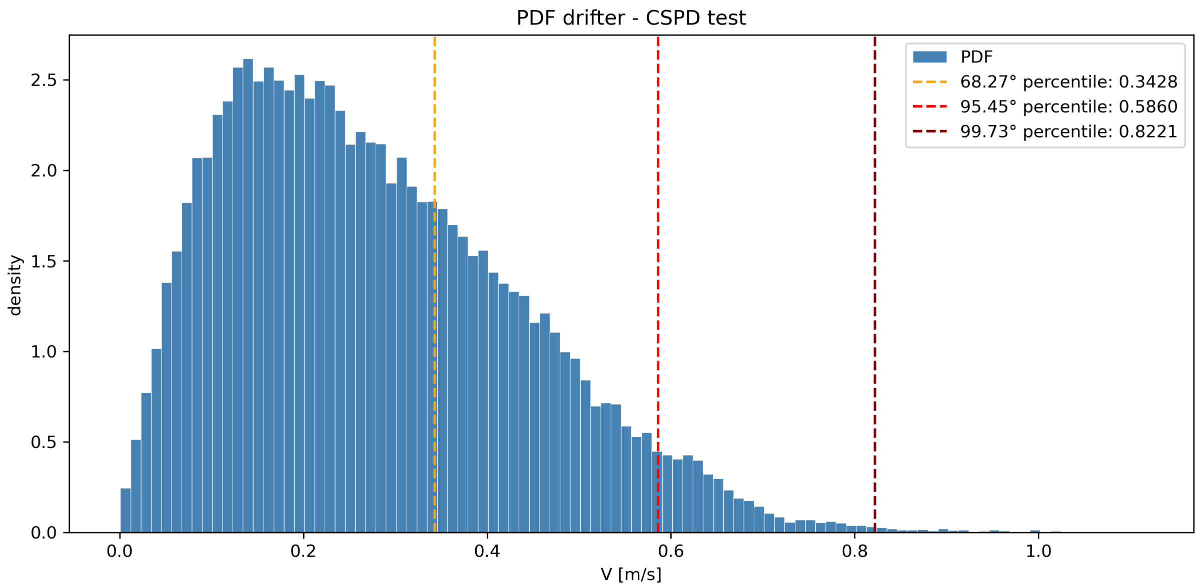

Figure 3.4: PDF of the drifter data for the CSPD test: in blue the distribution of the velocity magnitude of the drifters, the three dashed percentile lines mark the three percentiles corresponding to 1, 2, 3 standard deviations respectively - 68.27th (0.3428 m/s, orange), 95.45th (0.5860 m/s, red), 99.73rd (0.8221 m/s, dark red). Right-skewed distribution peaking near V = 0.2 m/s.

ations are abnormal errors, which need to be identified and removed from consideration (Zhao et al. 2013). So, in our case, assuming the distribution to be gaussian, the values exceeding the 99.73° percentile are considered non-physical and must be removed.

The Figure 3.4 shows that the majority of vectors' mean velocity is around 0.2 m/s,and then the density progressively decrease, until around 0.8 m/s or the 99.73° percentile which corresponds to the value of 0.8221 m/s.

Then the PDF of the velocity was plotted again, but this time using radar data, as shown in Figure 3.5. Again, the velocity was calculated using the formula used before:

$$V_i = \sqrt{u_i^2 + v_i^2} \tag{3.2}$$

In addition, the percentile corresponding to 0.8221 m/s, the value associated with the 99.73° in the drifter data PDF, was plotted. This was done to estimate what percentage of data would be removed in case the threshold of the new CSPD filter was changed to that value. The result shows that the percentile that corresponds to 0.8221 m/s is 95.26%.

While removing only 4.74% of the data might be reasonable, it was decided to test the new

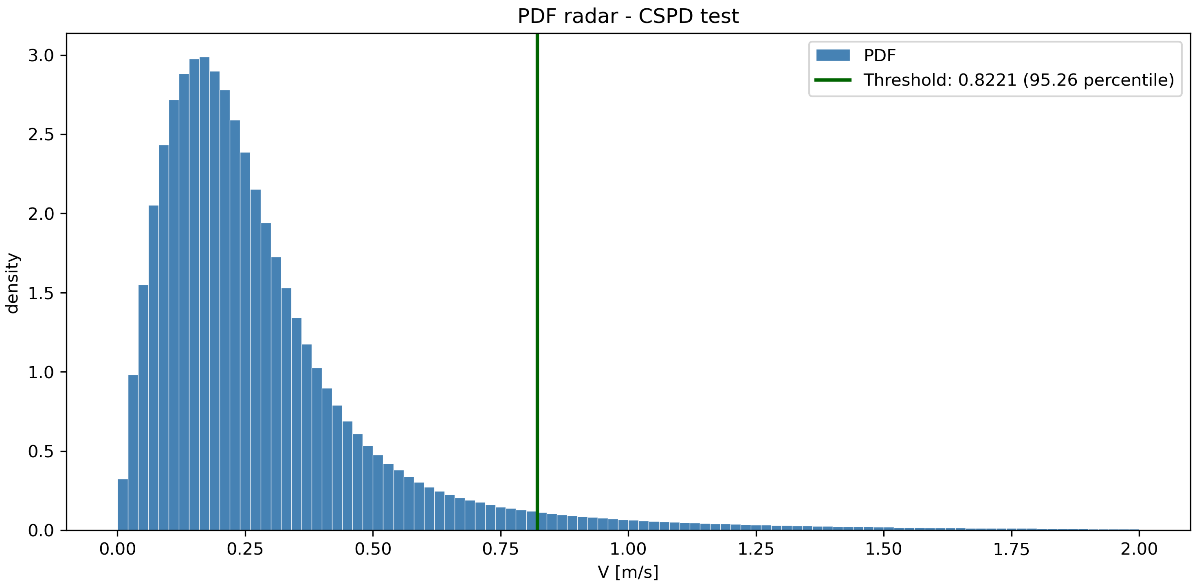

Figure 3.5: PDF of the radar data for the CSPD test: in blue the distribution of the velocity field magnitude of the cells, the green line marks 99.73rd-percentile value (0.8221 m/s) of the distribution of velocities measured by the observed drifters. This value corresponds to the 95.26 percentile of the velocities observed by the radar. Right-skewed distribution peaking near V = 0.2 m/s

threshold. A value of 0.8 m/s was used as the new threshold for simplicity, which corresponds to 95.00° percentile, which is still a reasonable result. As Figure 3.6 shows, with this threshold most of the spikes are removed, both the clear errors like the ones in the southeast portion of the domain and the likely non-physical vectors east of Portofino.

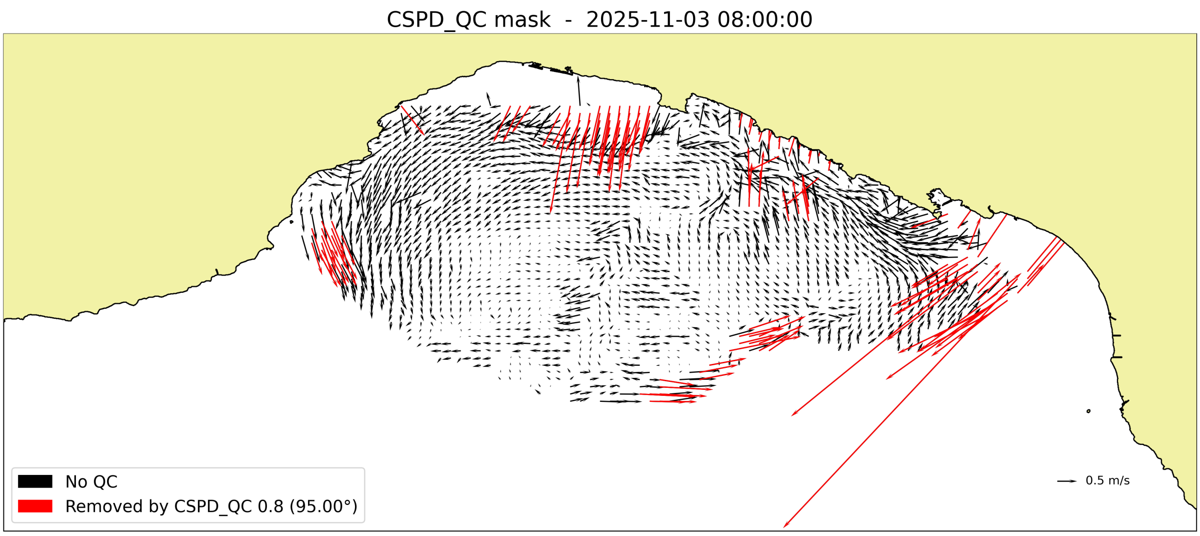

Figure 3.6: New CSPD mask - Velocity field on 03-11-2025 at 8:00: the length of each arrow is proportional to the velocity field magnitude in that point, while the arrow direction shows the sum of the two components of the velocity field. Red arrows show points whose velocity field magnitude is larger than 0.8 m/s (the new threshold of the CSPD quality control), while black arrows with values lower than 0.8 m/s. The new-threshold removals are concentrated near the coast west of Portofino and in the southeastern portion.

3.2 .3 VART

The second test that was decided to be reworked was the VART test. In particular, it was decided to both change the threshold used for this test, originally equal to 1.2 m/s, and the way the change in velocity dV used in this filter is calculated. To do so, first, using drifter data, the PDF of the change in velocity dV was plotted, as shown in Figure 3.7. Since the drifter data doesn't have the same temporal scale as the radar data being used, first it was averaged every hour and then the value dV was calculated with this formula:

$$dV_{1h}[i] = \sqrt{(u_i - u_{i-1})^2 + (v_i - v_{i-1})^2} \tag{3.3}$$

where u_i and v_i are the EWCT and NSCT velocity components at the timestep i and u_{i-1} and v_{i-1} are the EWCT and NSCT velocity components at the previous timestep i-1.

This is the first major difference of the new version of the VART filter and the old version of it, since the old one computes the absolute difference in velocity magnitude.

In addition, to better study the data, three percentiles were plotted, corresponding to the

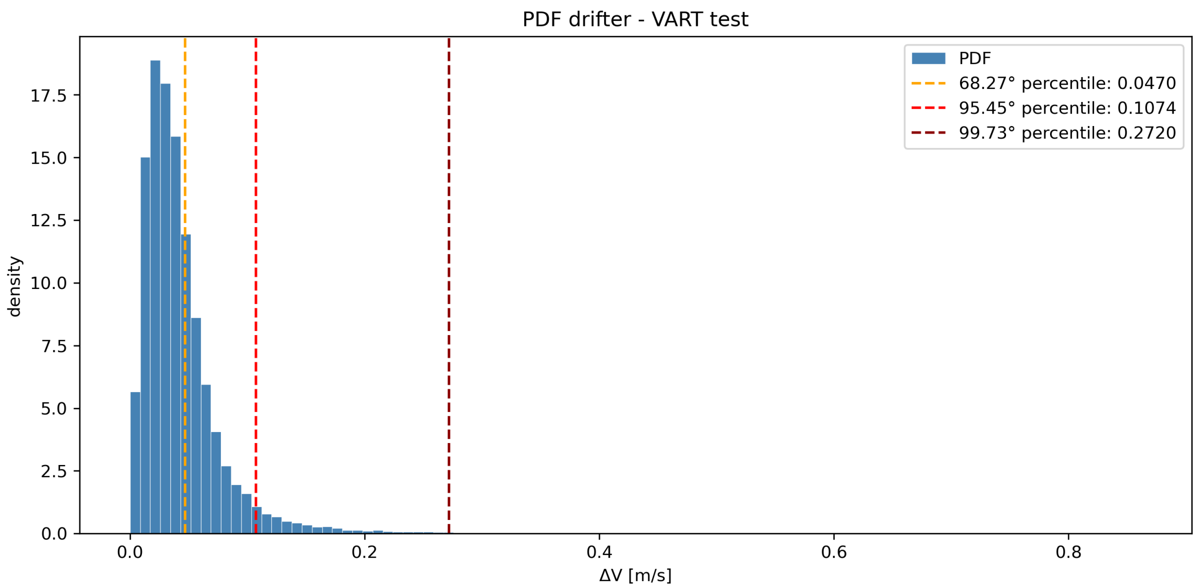

Figure 3.7: PDF of the drifter data for the VART test: in blue the distribution of the difference between the velocity magnitude of a drifter and their magnitude one hour before, the three dashed percentile lines mark the three percentiles corresponding to 1, 2, 3 standard deviations respectively - 68.27th (0.0470 m/s, orange), 95.45th (0.1074 m/s, red), 99.73rd (0.2720 m/s, dark red). Right-skewed distribution with long tail peaking near ΔV = 0.03 m/s.

three-sigma-rule values.

The Figure 3.7 shows that the majority of vectors' mean dV is around 0.03 m/s, and then the density progressively decrease, until around 0.3 m/s or the 99.73° percentile which corresponds to the value of 0.2720 m/s.

Then the PDF of the change in velocity was plotted again, but this time using radar data, as shown in Figure 3.8. Again, the dV was calculated using the formula used before:

$$dV_{1h}[i] = \sqrt{(u_i - u_{i-1})^2 + (v_i - v_{i-1})^2} \tag{3.4}$$

In addition, the percentile corresponding to 0.2720 m/s, the value associated with the 99.73° in the drifter data PDF, was plotted. This was done to estimate what percentage of data would be removed in case the threshold of the new VART filter was changed to that value. The result shows that the percentile that corresponds to 0.2720 m/s is 80.71%.

Since removing 19.29% of the data might be excessive, it was decided to test the new threshold. A value of 0.4 m/s was used as the new threshold for simplicity, which corresponds

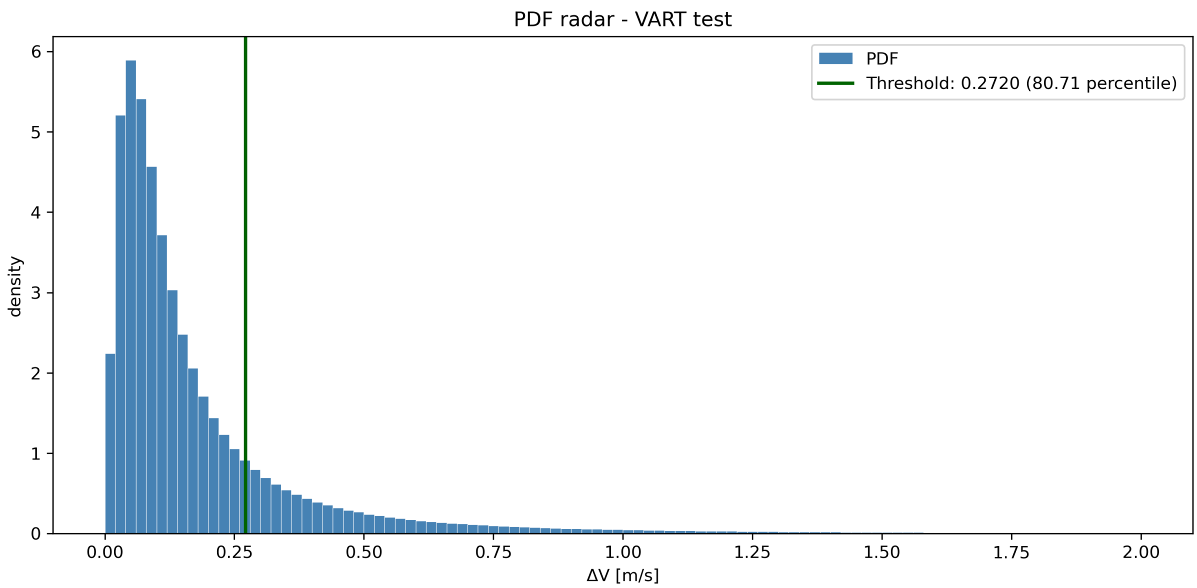

Figure 3.8: PDF of the radar data for the VART test: in blue the distribution of the difference between the velocity field magnitude of a cell and their magnitude at the previous timestep, the green line marks 99.73rd-percentile value (0.2720 m/s) of the distribution of velocities measured by the observed drifters. This value corresponds to the 80.71 percentile of the velocities observed by the radar. Right-skewed distribution with long tail peaking near ΔV = 0.05 m/s.

to 88.51° percentile, which is a more reasonable result. As Figure 3.9 shows, comparing it with Figure 3.6 that shows the domain at the previous timestep, with this threshold most of the suspect vectors are removed, both the spikes in the southeast portion of the domain and the abrupt changes in direction in the south portion of the domain.

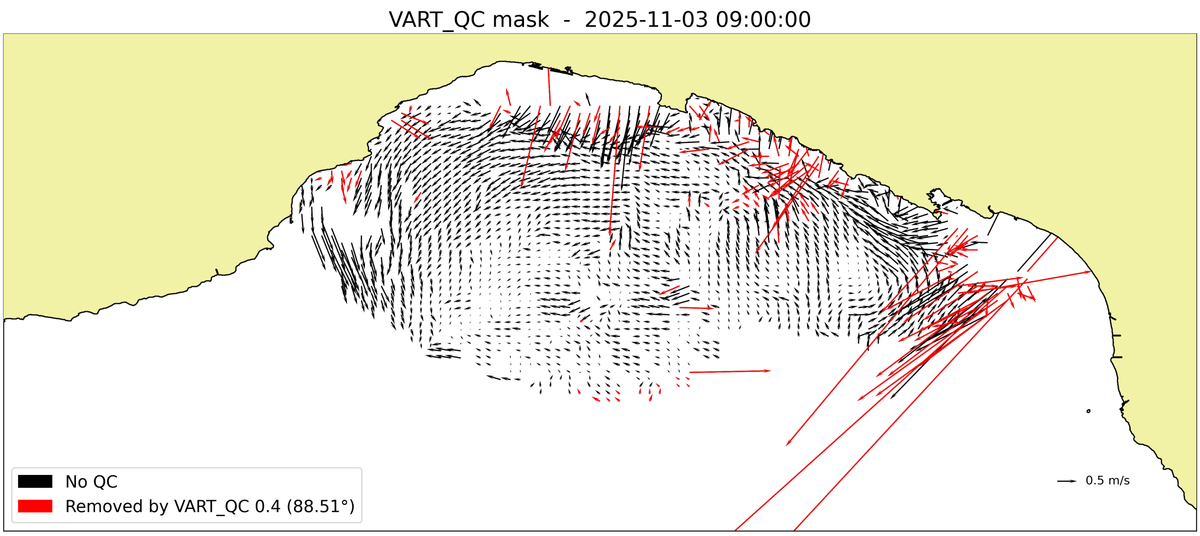

Figure 3.9: New VART mask - Velocity field on 03-11-2025 at 9:00: the length of each arrow is proportional to the velocity field magnitude in that point, while the arrow direction shows the sum of the two components of the velocity field. Red arrows show points whose velocity field magnitude is at least 0.4 m/s (the new threshold of the VART quality control) larger than their magnitude at the previous timestep, while black arrows with differences lower than 0.4 m/s. The new-threshold removals are concentrated near the coast west of Portofino and in the southeastern portion.

3.2 .4 MEDF

The third test that is introduced wasn't actually reworked since it was not present in the original QC pipeline. The method is referred to as the Spatial Derivative filter (MEDF) and is based on the Spatial Median Comparison test introduced in the QARTOD HFR Quality Control manual (Bushnell et al. 2022). This new test was added to the QC pipeline since even with the adoption of the updated thresholds for the CSPD and VART test, the QCflag filter wasn't yet able to remove all the questionable vectors present in the original setup. The test identifies anomalous velocity vectors by comparing their eastward and northward components with the median components of neighboring vectors. For each vector, the median u and v components are calculated within a user-defined spatial window expressed in grid steps. The deviation from the local median vector is then computed as the Euclidean distance between the component differences. A vector for which this deviation exceeds the specified threshold is classified with a "bad data" flag; otherwise the vector receives a "good data" flag. To find the best threshold to use in the test, first, using drifter data, the PDF of the change in velocity over 2 hours dV_{2h}

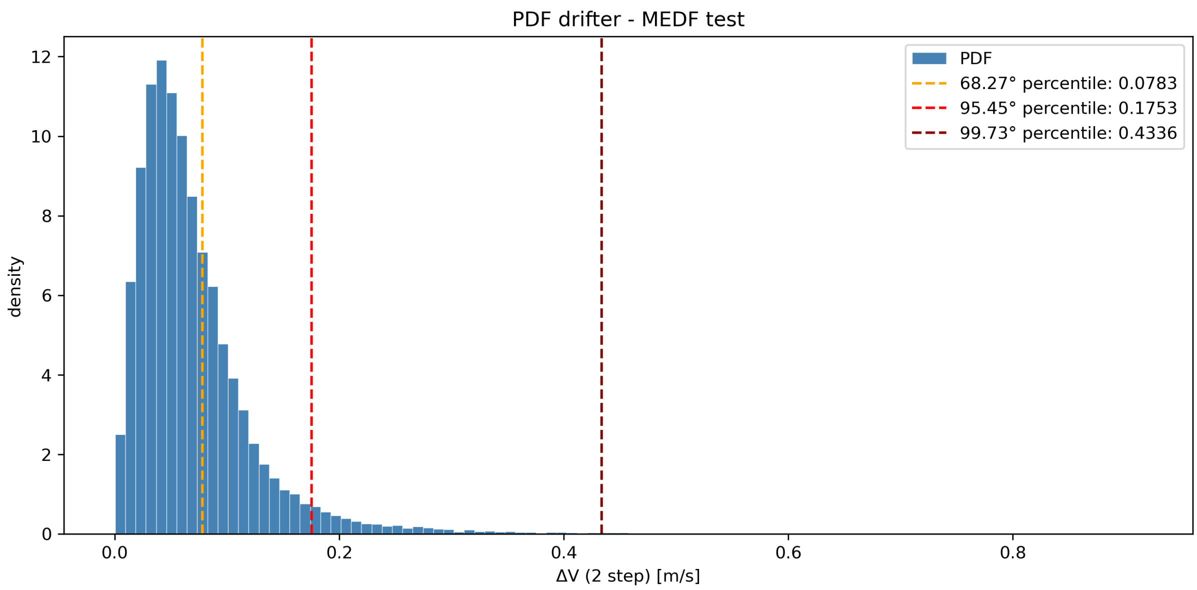

Figure 3.10: PDF of the drifter data for the MEDF test: in blue the distribution of the difference between the velocity magnitude of a drifter and their magnitude two hours before, the three dashed percentile lines mark the three percentiles corresponding to 1, 2, 3 standard deviations respectively - 68.27th (0.0783 m/s, orange), 95.45th (0.1753 m/s, red), 99.73rd (0.4336 m/s, dark red). Right-skewed distribution with long tail peaking near ΔV = 0.04 m/s.

was plotted, as shown in Figure 3.7. It was decided to study the change in velocity over 2 hours since the drifter data is only a time series, with no information regarding the spatial variability, so it was assumed the drifter would travel in 2 hours the same distance there is between two grid cells.

Since the drifter data doesn't have the same temporal scale as the radar data being used, first it was averaged every hour and then the dV_{2h} was calculated with this formula:

$$dV_{2h}[i] = \sqrt{(u_i - u_{i-2})^2 + (v_i - v_{i-2})^2} \tag{3.5}$$

where u_i and v_i are the EWCT and NSCT velocity components at the timestep i and u_{i-2} and v_{i-2} are the EWCT and NSCT velocity components at the 2 hours prior timestep.

In addition, to better study the data, three percentiles were plotted, corresponding to the three-sigma-rule values.

The Figure 3.10 shows that the majority of vectors' mean dV_{2h} is around 0.04 m/s, and then the density progressively decrease, until around 0.4 m/s or the 99.73° percentile which

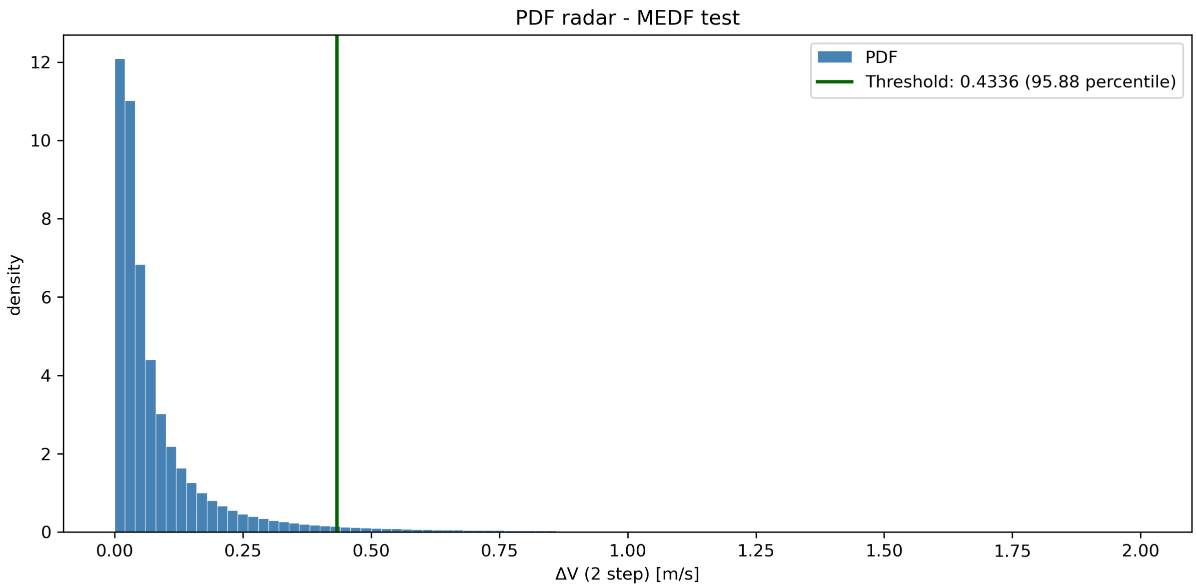

Figure 3.11: PDF of the radar data for the MEDF test: in blue the distribution of the difference between the velocity field magnitude of a cell and the median magnitude of its surrounding cells, the green line marks 99.73rd-percentile value (0.4336 m/s) of the distribution of velocities measured by the observed drifters. This value corresponds to the 95.88 percentile of the velocities observed by the radar. Right-skewed distribution with long tail peaking near ΔV = 0.05 m/s.

corresponds to the value of 0.4336 m/s.

Then the PDF of the change in velocity over 2 hours was plotted again, but this time using radar data, as shown in Figure 3.8. This time, having the neighboring data available, the dV_{2h} was calculated using the test formula:

$$dV_i = \sqrt{(u_i - u_{\mathrm{med_i}})^2 + (v_i - v_{\mathrm{med_i}})^2} \tag{3.6}$$

where u_i and v_i are the EWCT and NSCT velocity components at the timestep i and u_{med_i} and v_{med_i} are the local median eastward and northward velocity components computed from all valid neighboring velocity vectors within the selected spatial window.

In addition, the percentile corresponding to 0.4336 m/s, the value associated with the 99.73° in the drifter data PDF, was plotted. This was done to estimate what percentage of data would be removed in case the threshold of the brand new MEDF filter was set to that value. The result shows that the percentile that corresponds to 0.4336 m/s is 95.88%.

While removing only 4.12% of the data might be reasonable, it was decided to test the

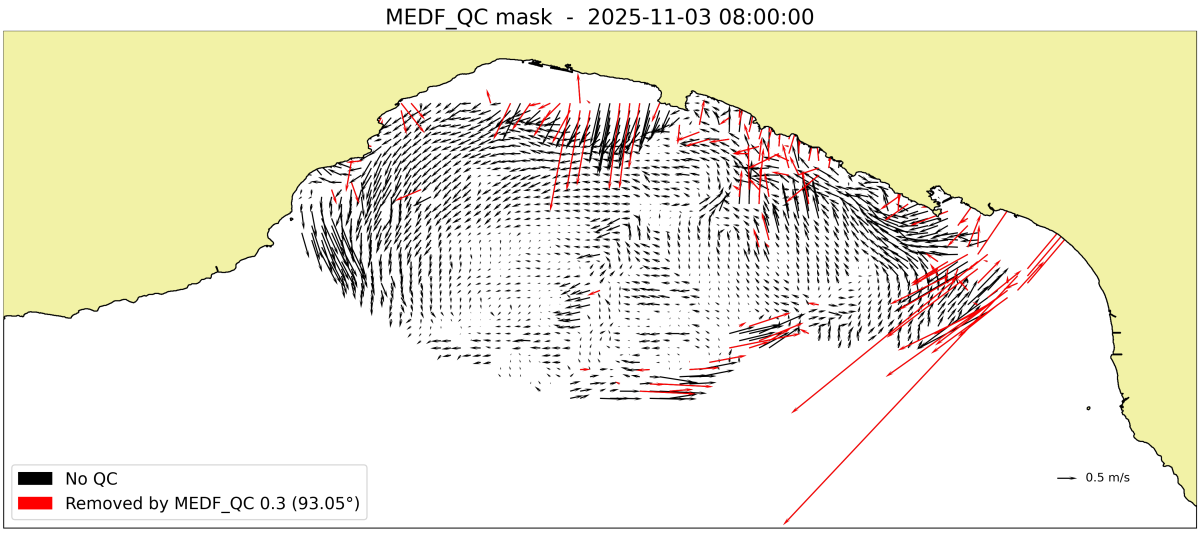

Figure 3.12: New MEDF mask - Velocity field on 03-11-2025 at 8:00: the length of each arrow is proportional to the velocity field magnitude in that point, while the arrow direction shows the sum of the two components of the velocity field. Red arrows show points whose velocity field magnitude is at least 0.3 m/s (the new threshold of the MEDF quality control) larger than their surrounding vectors' magnitude, while black arrows with differences lower than 0.3 m/s. The new-threshold removals are concentrated near the coast west of Portofino and in the southeastern portion, plus an additional denser group in the northern part of the domain.

new threshold. First, a value of 0.4 m/s was used as the new threshold for simplicity, but since the result wasn't as pronounced as hoped for and some of the suspect vectors weren't getting removed, it was chosen to lower the threshold at 0.3 m/s, which corresponds to 93.05° percentile, which is still a reasonable result. As Figure 3.12 shows, with this threshold most of the suspect vectors are removed, both the spikes in magnitude like the ones in the southeast portion of the domain and the abrupt changes in direction like those present in the north portion of the domain.

3.2 .5 QCflag

After the changes applied to the CSPD filter and the VART filter and the introduction of the new MEDF filter, the new version of the overall QCflag filter was then calculated. This new version of the overall QCflag filter consists of the logical sum of the all filters described in Table 3.2. The newer threshold was preferred over the older one if available.

As already mentioned at the beginning of the chapter, the overall QCflag filter aggregates the results of all individual tests: a vector receives a passing overall flag only if every individual test has returned a "good data" result; if any of the tests raises a "bad data" flag, the overall flag reflects the failure. Since lower thresholds have being used for two filters and an additional filter was implemented, this would result in the overall QCflag filter having a larger effect and an higher percentage of data being removed. This can be seen in Figure 3.14, that shows where and how much the new overall QCflag filter (below) is working compared to its old version (above), by counting the percentage of occurrence of "bad data" flags. The figure shows that the new version has a similar behavior to the old one while being more conservative.

The final result of the Quality Control (QC) filters part of this work can be seen in Figure 3.13. It can be seen, comparing it with Figure 3.1 that shows the vectors removed by the old version of the overall QCflag filter, how this new configuration is able to clean the area off the coast between Portofino and La Spezia and also to remove most of the suspect vectors off the western Ligurian coast.

Table 3.2: **Updated Quality Control pipeline configuration**

Short name	Long name	Old threshold	New threshold
GDOP	GDOP Threshold	> 2.0	-
CSPD	Velocity Threshold	> 1.2 m/s	> 0.8 m/s
VART	Temporal Derivative	> 1.2 m/s	> 0.4 m/s
DDNS	Data Density Threshold	< 3	-
POSITION	Position	-	-
MEDF	Spatial Derivative	-	> 0.3 m/s

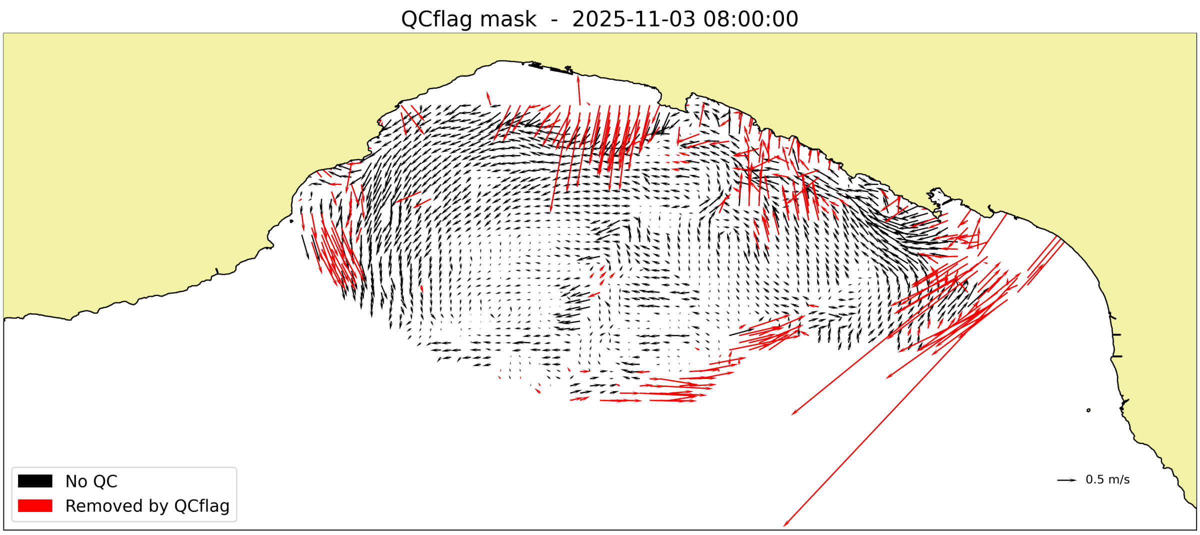

Figure 3.13: New QCflag mask - Velocity field on 03-11-2025 at 8:00: the length of each arrow is proportional to the velocity field magnitude in that point, while the arrow direction shows the sum of the two components of the velocity field. Red arrows show points that didn't pass at least one of the tests included in the new QCflag, while black arrows show points that passed every test. The new-configuration removals are concentrated near the coast of Portofino and in the southeastern portion, but also present in other parts of the domain. The area between Portofino and La Spezia is almost entirely "cleaned".

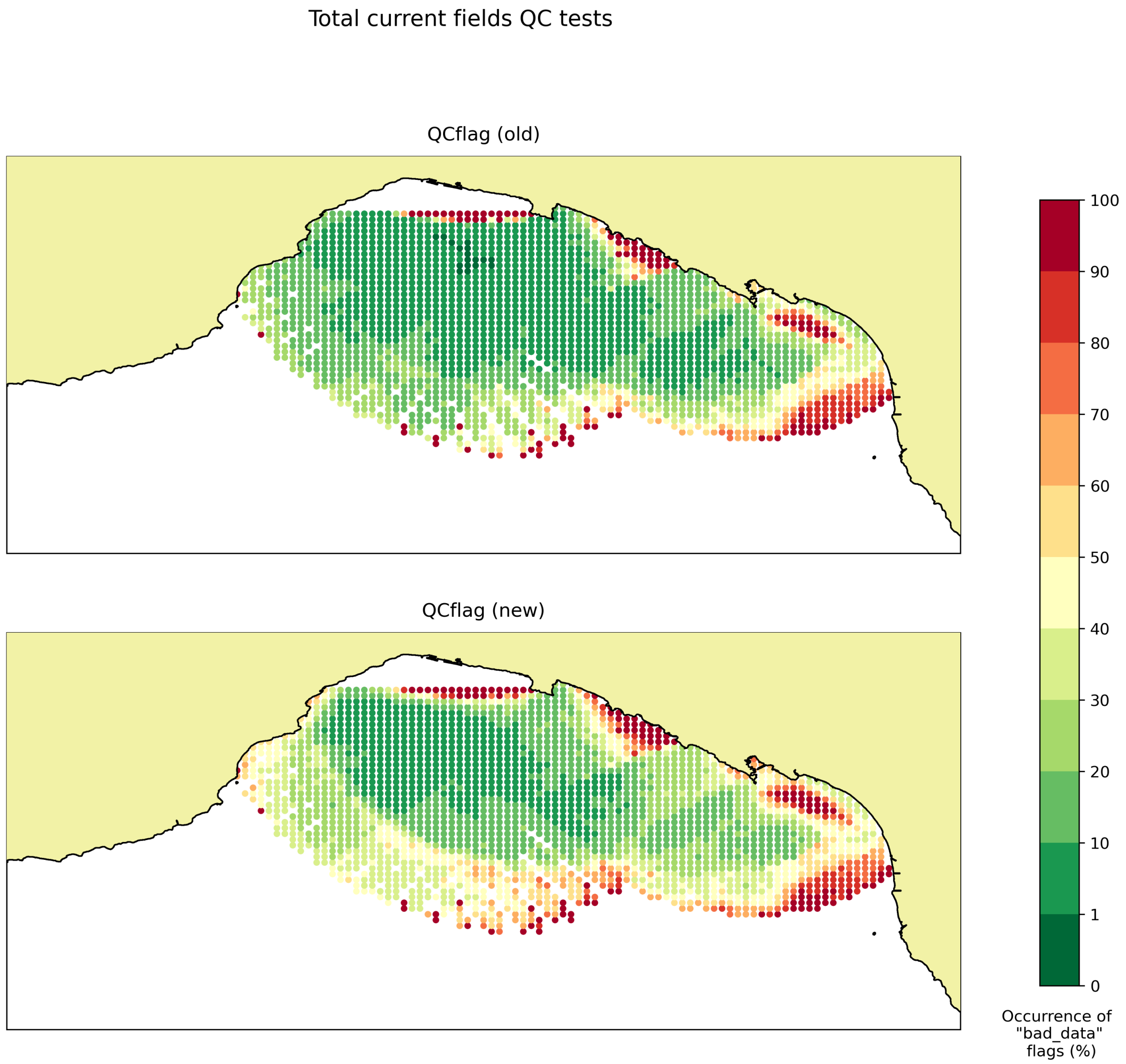

Figure 3.14: Spatial heatmap of the QCflag test. Top panel "QCflag (old)" and bottom panel "QCflag (new)". It shows where the filter removed more data in the year analyzed (10 Nov 2024 - 10 Nov 2025), in particular how many times a vector didn't pass the test in proportion of the number of its valid values. The more red an area is, the more has been targeted by the filter. The new version shows a visibly larger proportion of "bad" cells in the same hotspot areas.

4. Gap-filling

4.1 Methods

4.1 .1 Original Method

The presence of spatial and temporal gaps is a key limitation of HFR total velocity products. These gaps occur for several environmental and technical reasons, such as radio-frequency interference, poor sea conditions that reduce the backscatter signal, and occasional problems with the antenna or communication system (Corgnati, Maristella Berta, et al. 2024). Because these gaps are nearly unavoidable over long time periods and large areas, an operational HFR network should provide, alongside raw and quality-controlled data, a gap-filled product that is useful for practical and scientific applications. This is especially important for applications that rely on continuous velocity fields, like Lagrangian trajectory reconstruction. Various gap-filling methods have been suggested in the literature. These range from optimal interpolation to artificial neural networks, self-organizing maps, variational methods, empirical orthogonal functions, and open-boundary modal analysis (OMA) (Corgnati, Maristella Berta, et al. 2024).

In this work, we implement and assess a different gap-filling approach: an iterative, data-driven algorithm developed at CNR-ISMAR (Magaldi M. G., personal communication), referred to here as the Progressive Rim Gap-Filling Method. This method does not depend on any dynamic or geometric constraints related to the coastline or on a pre-computed modal basis. Instead, it works directly on the gridded velocity fields, combining both a temporal and a spatial filling stage. Its main advantages are its simplicity and low computational cost, which make it a very strong candidate for a general-purpose gap-filling baseline to test on the HFR-TirLig dataset.

The procedure is applied separately to the eastward (EWCT) and northward (NSCT) velocity components. It consists of two sequential stages: a temporal interpolation stage and a spatial interpolation stage.

Temporal interpolation stage. The quality-controlled dataset is first interpolated along the time axis using a linear interpolation method with a maximum persistence gap of 6 hours.

For each grid cell, a missing value is replaced with a value linearly interpolated between the two closest valid observations, but only if they are no more than 6 hours apart. Gaps longer than this are intentionally left unfilled. Extending beyond this threshold could mean extrapolating the current field over a period when the true ocean state might have changed, leading to unrealistic and misleading estimates. This choice is based on the traditional persistence assumption in short-term gap filling of geophysical time series, which suggests that the ocean state remains coherent only over limited hours. This is also supported by the temporal autocorrelation study done on the data (Magaldi M. G., personal communication) that showed a big drop in the results after the 6-hour mark (Figure 4.1). A boolean mask (time_interp_mask) is created at this stage. This mask identifies every grid point and time step where a value was reconstructed solely through temporal interpolation. This mask is included in the final dataset, allowing later analyses to differentiate between temporally filled and originally observed points.

Spatial interpolation stage. The remaining gaps after the temporal stage, either due to exceeding the 6-hour limit or lacking valid neighboring time steps, are filled in space at each time step through the fill2DradarGaps routine. This routine uses an iterative nearest-neighbor averaging scheme. In each iteration, the velocity field is forward and backward filled along the longitude dimension, then forward and backward filled along the latitude dimension, with a fill limit of one grid cell. This creates four candidate fields, each representing the propagation of the nearest valid value by one grid cell in one of four directions: east, west, north, and south. The four candidates are stacked along an extra dimension and averaged, using only the directions where a valid neighboring value exists. As a result, any grid point still missing after one iteration but adjacent to valid cells will receive the average from its valid neighbors. This filling and averaging process repeats iteratively, expanding the filled region outward by one grid cell from the edges of each data patch. The process continues until no missing values are left anywhere in the domain, which is confirmed by counting the number of NaN values at the first time step of the processed field. Conceptually, this iterative process acts like a discrete diffusion: it does not fit any physical model to the flow. Instead, it creates a smooth field through local averaging, spreading the last available observations inward from the edges of gaps until the entire grid is filled. Since this spatial filling procedure does not account for coastline geometry, it treats ocean and land grid cells the same. If not restricted, it could extend velocity values over land cells on the Cartesian grid. To fix this issue, once the iterative filling has finished, the resulting fields are masked using an independently defined ocean mask (oceanMsk). For each grid cell, the fraction of valid observations throughout the entire time series is calculated. Only cells with more than 20% valid data are included in the ocean mask for analysis. This process removes all grid points outside the sea domain, discarding meaningless values generated during filling. A second boolean mask (space_interp_mask) is then created, comparing the field from the temporal stage with the field from the spatial stage to identify the points filled exclusively by the spatial method.

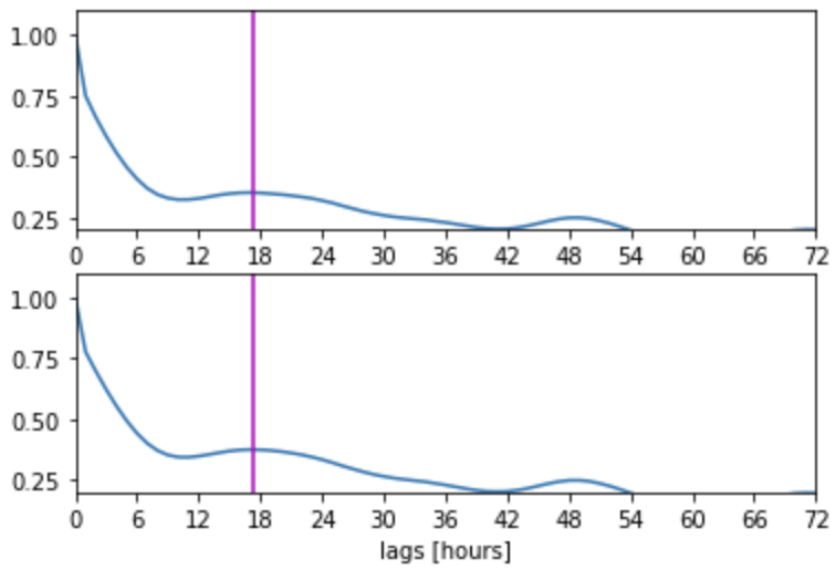

Figure 4.1: Temporal autocorrelation function of the eastward (EWCT, top panel) and northward (NSCT, bottom panel) surface velocity components, averaged over all grid points of the HFR-TirLig domain, as a function of the time lag (0–72 hours). Both components show a similar behavior: a steep initial decorrelation, with the ACF dropping from 1 to about 0.35 within the first 9–10 hours, followed by a shallow local maximum located close to the inertial period, a further decline toward a minimum around 40 hours, and a weak secondary oscillation near 48 hours.

4.1 .2 Updated Method

In the original Progressive Rim Gap-Filling Method, the temporal and spatial filling stages run sequentially within a single routine. This setup makes it challenging to isolate the specific contribution of each stage to the overall velocity field reconstruction. Since the two stages rely on different assumptions, evaluating their combined output doesn't allow for independent examination of each stage's behavior and limitations. Therefore, the method was reorganized into two separate implementations, one for temporal interpolation only and one for spatial interpolation only. This way, each stage can be studied and validated separately.

Temporal-only implementation. In this version, the spatial filling routine is completely removed. After the temporal linear interpolation, the resulting field is simply masked with the ocean mask, with no further spatial processing of the remaining gaps. A significant change involves the temporal interpolation step itself: the maximum persistence limit of 6 hours is removed. In this updated version, linear interpolation along the time axis occurs without a gap duration limit. For any grid point, missing values between two valid observations are linearly interpolated, regardless of the time between those observations. The only gaps that remain unfilled are at the very beginning or end of the time series, where no valid observation exists on one side, as linear interpolation cannot extrapolate beyond available data. This adjustment aims to allow for a full characterization of the temporal interpolation stage without restricting it a priori to short gaps.

Spatial-only implementation. In this second version, the temporal interpolation stage is completely skipped. The spatial filling routine fill2DradarGaps is applied directly to the input velocity field without any prior temporal processing. The iterative nearest-neighbor averaging scheme remains unchanged from the earlier description. Since no temporal interpolation happens beforehand, the boolean mask identifying the points filled by the spatial method is now determined directly against the filtered field (the raw field to which quality control has been applied), rather than against a field already partially filled in time, as in the combined version of the method. This ensures that the mask accurately identifies every point whose reconstructed value comes solely from the spatial filling technique.

Separating the method in this way allows examination of the temporal and spatial components of the Progressive Rim Gap-Filling Method independently, each driven solely by its own underlying assumption (temporal or spatial continuity).

4.1 .3 Artificial Hole - dimension

The validation of any gap-filling method needs a standard to compare the reconstructed velocity field with the true ocean state. When the gap-filling method is applied to the quality-controlled HFR-TirLig dataset's natural gaps, this comparison isn't straightforward since we don't know the value of a missing observation by definition. To address this, an artificial gap is created in a part of the dataset that is otherwise complete with valid observations. By masking known values as missing, we can apply the gap-filling method to this altered field and compare the reconstructed values to the originals. This approach gives a quantitative measure of the method's reconstruction skill rather than just qualitative assessments.

For the artificial gap to be meaningful, its size and duration need to reflect real gaps. If it's much smaller than typical gaps seen in practical situations, the test won't be representative. On the other hand, a gap that's too large might explore a type of missing data that rarely occurs. Therefore, it is important first to analyze the statistical distribution of natural gaps in the quality-controlled dataset regarding their duration and spatial extent.

We examine two separate aspects of the natural gaps, being their duration and size.

Temporal duration. For each grid cell in the ocean mask, we review the time series of a year of the eastward velocity component to find instances of consecutive missing values (using the northward velocity component instead does not change the result of the test, since if the value of a component is missing, the value of the other is also missing). We use a run-length encoding method to achieve this. We add a valid value at both ends of the boolean series indicating missing data. Then, we identify where the series changes from valid to missing (marking the start of a gap) and from missing to valid (marking its end). The difference between these two points reveals the duration of each gap in hours, based on the dataset's hourly sampling frequency. We repeat this for each grid cell, combining the durations into a single distribution that represents the temporal behavior of gaps across the entire domain. Figure 4.2 shows this distribution, along with the 50th, 75th, 90th, 95th, and 99th percentiles. The distribution is strongly right-skewed, indicating that most gaps are small, while a smaller number of significantly larger gaps forms a long tail. The 50th percentile corresponds to 3 hours, while the 75th percentile is 7 hours, meaning that at least three-quarters of the temporal gaps last a few hours at most. The

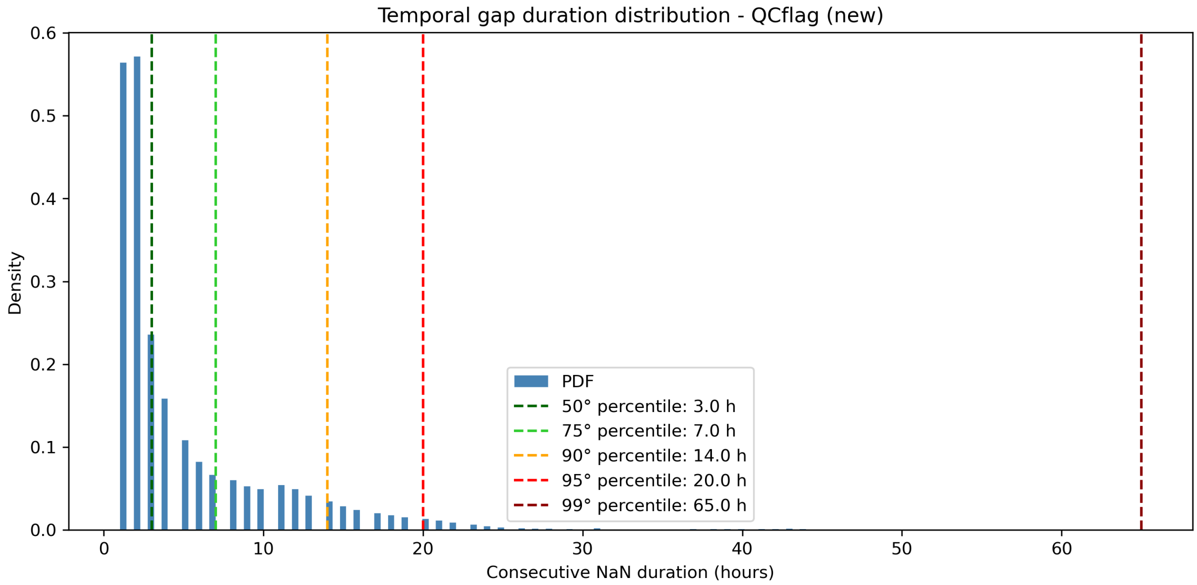

Figure 4.2: PDF of the temporal gap duration: in blue the distribution of the temporal gaps which were measured in HFR data domain during the period from 10/10/2024 to 10/10/2025 (section 2.4), the five dashed percentile lines mark the five percentiles chosen for the following analyses - 50th (3.0 h, dark green), 75th (7.0 h, green), 90th (14.0 h, orange), 95th (20.0 h, red), 99th (65.0 h, dark red). Right-skewed distribution with a long tail peaking near 0–3 h and rapid decay.

90th and 95th percentiles are 14 and 20 hours respectively, and the 99th percentile is 65 hours, showing that some gaps can last nearly three days.

Spatial size. For every time step, we identify the grid cells that are missing and located within the ocean mask (again, by ocean mask, we mean the ensemble of grid points for which at least 20% of the values are valid). We isolate the connected groups of these cells using an 8-connected labeling algorithm, which links neighboring missing cells (including diagonal ones) into distinct gap features. We count the number of cells in each labeled feature to determine the spatial size of each gap, expressed as the number of grid cells. This process is repeated for every time step, resulting in a single distribution that describes the spatial extent of gaps over the entire year long time series. Figure 4.3 shows this distribution along with the 50th, 75th, 90th, 95th, and 99th percentiles. This distribution is also strongly right-skewed and has a pronounced tail. The 50th percentile corresponds to only 2 cells, and the 75th percentile is 6 cells, confirming that most spatial gaps are very localized, often involving isolated or nearly isolated cells. The 90th percentile reaches 43 cells, the 95th percentile 144 cells, and the 99th percentile goes up to 1412 cells, indicating that extended spatial gaps, likely due to prolonged issues with radar stations, can occur despite being rare.

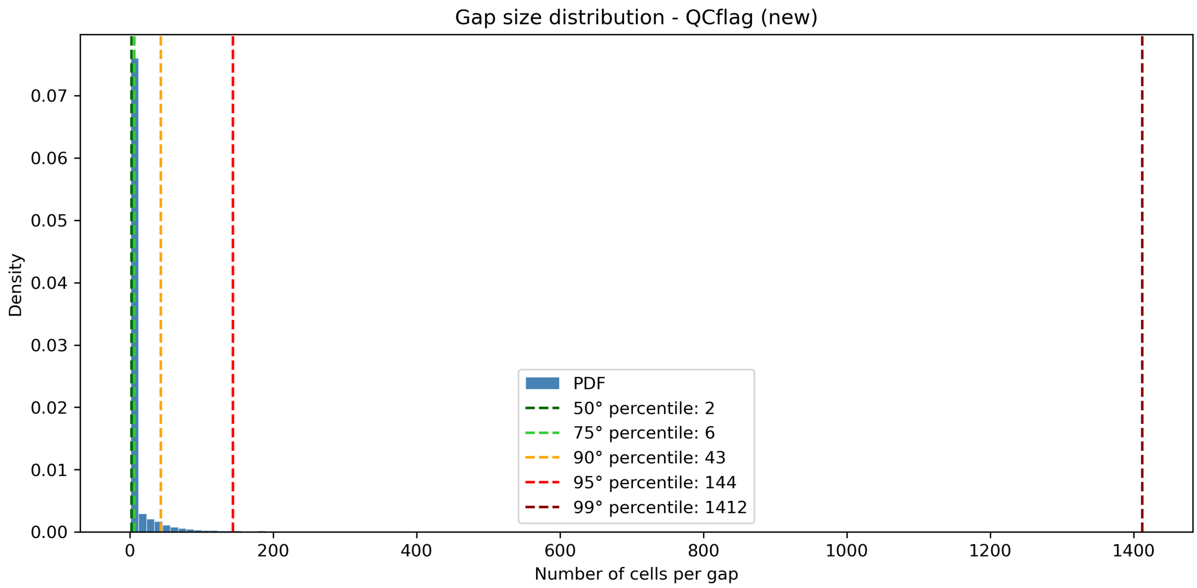

Figure 4.3: PDF of the spatial gap size: in blue the distribution of the spatial gaps which were measured in HFR data domain during the period from 10/10/2024 to 10/10/2025 (section 2.4), the five dashed percentile lines mark the five percentiles chosen for the following analyses - 50th (2 cells, dark green), 75th (6 cells, green), 90th (43 cells, orange), 95th (144 cells, red), 99th (1412 cells, dark red). Each spatial gap is identified as a series of adjacent non valid points using an 8-connected labeling algorithm. Extremely right-skewed distribution with a very long tail peaking near 0-10 cells and rapid decay.

4.1 .4 Artificial Hole - position

After deriving a set of potential duration and size values , we implemented two independent procedures to insert an artificial gap into the quality-controlled dataset, one for each of the stages of the Progressive Rim Gap-Filling Method.

Instead of inserting a simple rectangular block of missing values, we use a random region-growing algorithm to create an irregularly shaped, 8-connected group of exactly n grid cells. Starting from a single seed cell, the algorithm randomly selects a cell which is already part of the patch, tests its eight neighboring cells in random order, and adds to the patch the first available neighbor which is not part of it yet. This continues until the mask contains exactly n cells that are in the patch, after which the mask is cropped to its bounding box and padded with an inactive border of one cell. The result is a compact but irregularly shaped patch, which better represents the fragmented nature of real gaps compared to regular shapes.

For the temporal test, the spatial extent of the gap is fixed to a single grid cell so that reconstruction relies solely on temporal interpolation, avoiding any confusion from spatial filling. For the spatial test, the gap size is set approximately to each of the candidate values from the percentile analysis. We repeat the entire placement and reconstruction process 10 times for each size, using independent random selections to ensure that the influence of any single shape or location of the gaps on the resulting final statistics is averaged out.

Temporal valid position. In the temporal test, a valid combination of spatial location and time window must meet two conditions. First, the selected grid cell for the gap must contain a valid measurement at the reference time step, ensuring we hide a genuine observation rather than a pre-existing gap. Second, we search for a time window of the required duration around the reference time step. We test various candidate windows, offset by different hours prior to the reference time, and accept a window only if the two grid points immediately surrounding it are both fully valid at that location. This second condition is crucial as the temporal interpolation reconstructs missing values based on linear interpolation between the valid observations bracketing the gap. We perform this search exhaustively across the entire spatial grid and all candidate temporal offsets, producing a list of all valid (position, time-window) combinations for that reference time step. We then randomly select one combination from this list, avoiding systematic bias toward specific areas or time periods. This procedure is repeated independently for each time step in the dataset used as the reference. If no valid combination exists (usually because local data is too sparse), we skip that reference time step.

Spatial valid position. For the spatial test, we perform the search independently at each time snapshot in the dataset, moving the gap shape across all possible grid positions. We accept a candidate position only if it meets two conditions. The first is an internal-validity requirement: at least 75% of the cells within the gap must contain valid data. This threshold is intentionally relaxed from the strict 100% used for the temporal test, as an irregular patch, especially the largest candidate sizes from the spatial-size distribution's tail, is more likely to overlap with naturally missing cells. The second requirement is a border-validity condition: we obtain a one-cell-wide ring of cells immediately around the gap by applying morphological dilation to the hole mask using an 8-connected structuring element. All cells in this ring must be fully valid. This condition ensures that the artificial gap is completely surrounded by genuine, valid observations, which is crucial for a meaningful test of the spatial interpolation stage of the Progressive Rim Gap-Filling Method. This stage reconstructs missing values by spreading information inward from the valid neighboring cells. A hole bordered by more missing data would not provide a representative test of an isolated gap. As in the temporal case, one of the acceptable positions is chosen at random for each time step and also for each of the 10 repetitions performed per candidate size.

After selecting a position and, for the temporal test, a temporal window, the corresponding cells of both velocity components are marked as missing in a duplicate of the quality-controlled dataset. The original values are kept separately for later comparison. A three-dimensional boolean array, matching the full dimensions of the dataset, records whether each specific cell belongs to the artificial hole being tested at every time step and grid point. This array serves as the reference mask used to isolate the values that were artificially hidden from the reconstructed field.

The artificially degraded dataset is processed using only the interpolation stage relevant to the test being performed, consistent with the split structure of the Progressive Rim Gap-Filling Method. In the temporal test, a linear interpolation along the time dimension is applied at each grid point, without a maximum persistence limit. In the spatial test, only the spatial interpolation stage is applied, using the iterative nearest-neighbor propagation scheme along the longitude and latitude dimensions. In both cases, the reconstructed field is restricted to the operational ocean mask defined earlier in this section. It is then combined with the original filtered field and the hole-location mask into a single output dataset. This forms the basis for the quantitative comparison presented in the following section, between the values hidden by the artificial hole and those reconstructed by the gap-filling method.

4.1 .5 R^2 Estimation

After applying the artificial-hole procedure described in subsection 4.1 .4, individually for the temporal and spatial cases and for each hole dimension from the corresponding percentile-based set, a quantitative measure of the reconstruction skill of the Progressive Rim Gap-Filling Method is achieved by comparing the original (true) velocity values that were masked out with the values produced by the gap-filling procedure at the same locations. This comparison is done separately for the eastward (EWCT) and northward (NSCT) velocity components, as well as for the temporal and spatial artificial-hole experiments. The two configurations test different limiting behaviors of the algorithm, namely the temporal interpolation stage and the nearest-neighbor spatial filling stage.

For each hole dimension, the true and reconstructed values are gathered at the grid points and time steps affected by the artificial hole, using the boolean hole mask created during the hole-insertion step. In the spatial case, this extraction takes place over ten independent runs for each hole size, since the shape and location of the artificial hole within the domain are randomly generated or selected at each time step. Running the experiment multiple times for a specific hole size and combining the results reduces sampling variability from this randomness. This approach creates a larger and more representative pool of paired values, particularly for larger hole sizes, which are less common and thus more sensitive to the exact shape or placement of the random hole. For each hole dimension, the paired true and reconstructed values collected across all time steps (and across all repetitions) are combined into two one-dimensional arrays, one for each velocity component, after discarding any pair where either the true or the reconstructed value is missing (NaN).

The quality of the reconstruction is indicated by the coefficient of determination, R^2, calculated between the true values and the reconstructed values using the linregress function from the scipy.stats module. This function fits a linear regression between the two variables and provides, among other outputs, the Pearson correlation coefficient r (Virtanen et al. 2020), from which R^2 is derived as the square of r. The linregress estimate was uniformly applied to both the temporal and spatial artificial-hole experiments to ensure that the results from both sets are directly comparable. Figure 4.4 shows an example of this application.

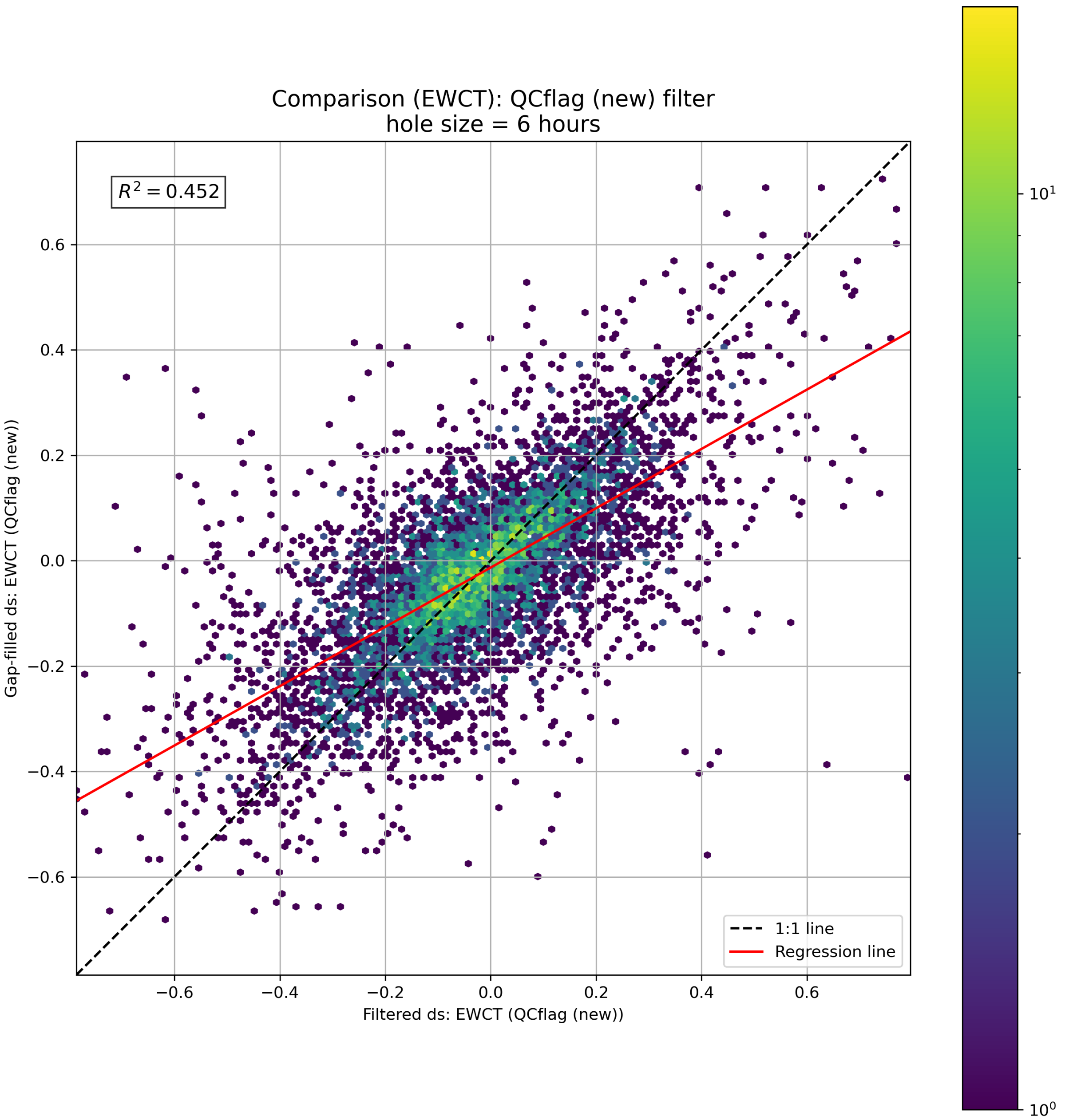

Figure 4.4: . Example comparison between the true (filtered) and reconstructed values of the eastward velocity component (EWCT) for the temporal artificial-hole experiment at a hole duration of 6 hours. Each point corresponds to one true/reconstructed pair extracted at a single grid point and time step by the artificial-hole procedure. The horizontal axis gives the true, quality-controlled value and the vertical axis gives the value produced by the gap-filling method at the same location; point color shows the local density of overlapping points on a logarithmic scale. The black dashed line marks the 1:1 relationship that a perfect reconstruction would follow, while the solid red line is the linear regression fitted between true and reconstructed values, whose squared correlation coefficient defines the R^2 statistic. It can also be noted that the fitted regression line is more inclined than the ideal 1:1 line, a consequence of the smoothing inherent to the interpolation process, which results in a systematic underestimation of the reconstructed velocity values relative to the true ones.

4.2 Results

Temporal case. Figure 4.5 shows the resulting R^2 values based on the duration of the artificial hole, measured in hours, for the eight durations selected, inspired by the results of subsection 4.1 .3 (2, 4, 6, 9, 12, 16, 20, and 24 hours) for both the EWCT and NSCT components. The reconstruction skill decreases consistently and rather sharply as the gap duration increases. For the shortest duration (2 hours), R^2 reaches 0.704 for EWCT and 0.796 for NSCT. This shows that the linear, persistence-based temporal interpolation stage can reconstruct short data gaps with reasonable accuracy. As the gap duration increases, performance drops quickly. After 6 hours, R^2 falls to 0.452 for EWCT and 0.583 for NSCT. It continues to decline for longer gaps, hitting values below 0.25 for EWCT and 0.30 for NSCT after 12 hours. For the longest duration tested (24 hours), R^2 drops to 0.165 for EWCT and 0.253 for NSCT. This pattern matches the persistence assumption in the temporal stage of the Progressive Rim Gap-Filling

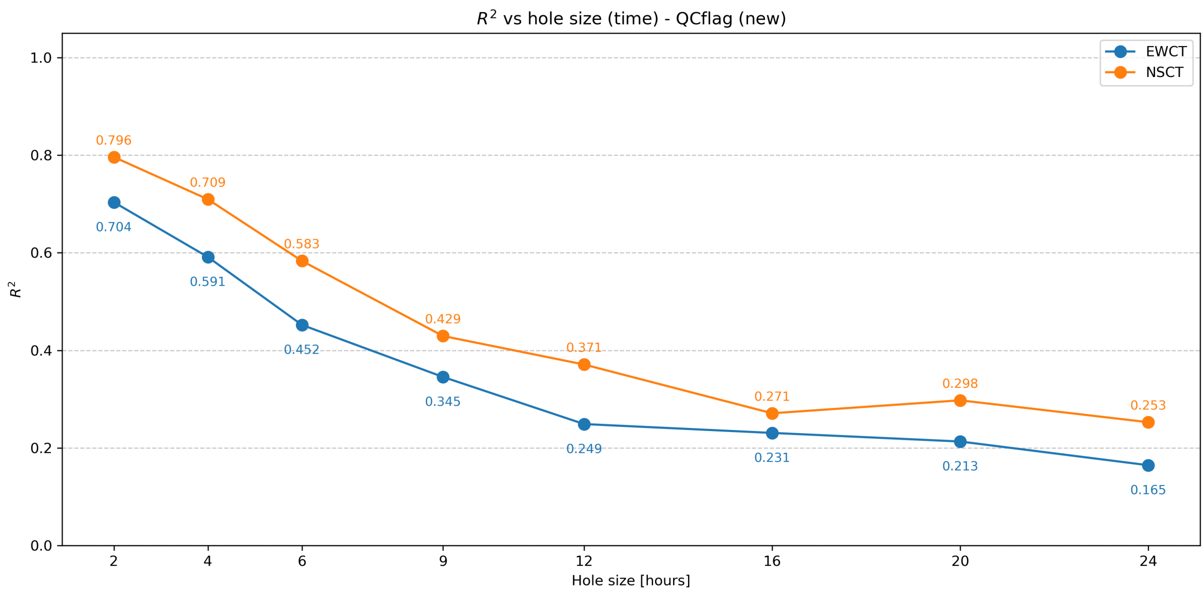

Figure 4.5: Coefficient of determination (R^2) between the true and the gap-filling-reconstructed velocity values, as a function of the duration of the artificial temporal hole (2, 4, 6, 9, 12, 16, 20, 24 hours), selected from the percentile analysis of the natural gap-duration distribution. Blue: eastward component (EWCT); orange: northward component (NSCT); values are annotated at each point. R^2 decreases monotonically with hole duration, from 0.704 (EWCT) / 0.796 (NSCT) at 2 hours to 0.165 / 0.253 at 24 hours, with the steepest drop around the original 6-hour limit of the method.

Method: once the gap exceeds the 6-hour window, reconstructed values depend more on the spatial filling stage instead of real temporal information. The ocean state likely changes significantly compared to the last valid observation, so the linear interpolation becomes less capable of following the true flow evolution. Throughout all the tested durations, the NSCT component shows a higher R^2 than EWCT, suggesting that the northward velocity component behaves in a more predictable or less noisy way in the study area.

Spatial case. Figure 4.6 presents similar results from the spatial artificial-hole experiment. Here, R^2 is shown as a function of the hole size, defined by the number of grid points making up the artificial gap, for the eight sizes selected, inspired by the results of subsection 4.1 .3 (2, 5, 10, 20, 50, 100, 200, and 500 grid points). Like in the temporal case, the reconstruction skill decreases as the hole size increases, but the decline is much steadier, especially for small and medium hole sizes. For a 2-point hole, R^2 equals 0.921 for EWCT and 0.939 for NSCT, staying above 0.86 for both components up to 20 grid points. The decline becomes more evi-

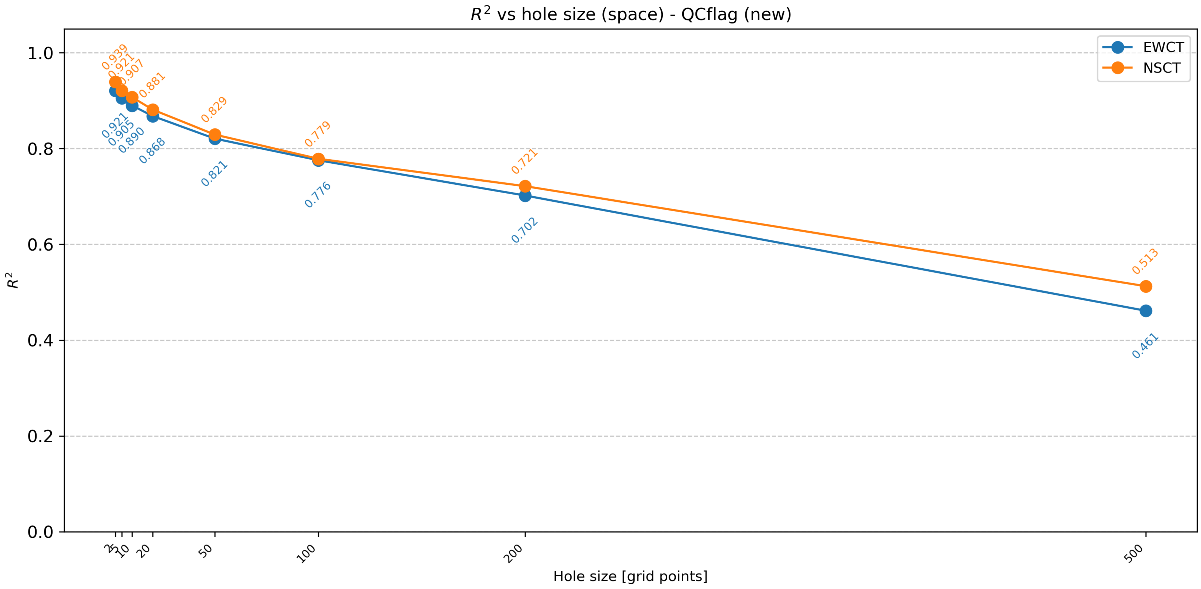

Figure 4.6: Coefficient of determination (R^2) between the true and the gap-filling-reconstructed velocity values, as a function of the size of the artificial spatial hole (2, 5, 10, 20, 50, 100, 200, 500 grid points), selected from the percentile analysis of the natural gap-size distribution. Blue: eastward component (EWCT); orange: northward component (NSCT); values are annotated at each point. R^2 decreases with hole size, from 0.921 (EWCT) / 0.939 (NSCT) at 2 points to 0.461 / 0.513 at 500 points, with a gradual decline up to 100 points followed by a sharper deterioration at 200-500 points.

dent after 50 points, with R^2 dropping to 0.821 for EWCT and 0.829 for NSCT at 50 points, and further down to 0.776 for EWCT and 0.779 for NSCT at 100 points. For the two largest hole sizes tested, performance worsens significantly , reaching $R^2 = 0.702$ for EWCT and 0.721 for NSCT at 200 points, and finally $R^2 = 0.461$ for EWCT and 0.513 for NSCT at 500 points. This trend reflects how the spatial filling stage works. It reconstructs missing values by averaging nearby points inward from the edges of each gap. For small holes, most reconstructed points are close to the gap boundary, benefiting from spatially correlated valid neighbors. For larger holes, more reconstructed points are far from any valid observation, so their values come from multiple averaging iterations. This causes them to lose connection with the actual small-scale velocity structure. In this case, the NSCT component consistently outperforms EWCT across all hole sizes, resembling the behavior seen also in the temporal experiment.

5. Discussion

This chapter discusses the results presented in chapter 3 (Quality Control) and chapter 4 (Gap-filling), organizing the discussion so as to follow, as closely as possible, the same chronological order in which the two parts of the methodology were introduced. The chapter begins with a discussion of the quality control results, including the failures of the former QC configuration, the specific contributions of the new MEDF test, and the benefits achieved by combining the recalibrated CSPD and VART thresholds with MEDF as opposed to the original pipeline. The discussion then proceeds to the gap-filling technique, first evaluating the temporal reconstruction skill, then the spatial reconstruction skill, and then comparing the two stages directly, and turning this comparison into a pragmatic, R^2-based criterion for determining the most suitable interpolation mode for a specific gap. The chapter closes with an oceanographic interpretation of the spatial reconstruction results, relating the size of the artificial holes tested to the characteristic scales of mesoscale motion expected in the study area.

QC results. Before discussing the gap-filling results, it is helpful to revisit the findings from the quality control analysis in chapter 3. The old QCflag configuration (Table 3.1) was not conservative enough. The mask in Figure 3.1 shows vectors with unrealistic magnitudes and directions along the western Ligurian coastline, along with several suspect vectors scattered throughout the domain. These vectors consistently passed through the entire old QC pipeline. This issue stems from the loose CSPD and VART thresholds (1.2 m/s each) used in the original configuration. These thresholds only removed the most extreme outliers and left many unflagged vectors that were physically implausible. In the revised configuration, these thresholds were not simply lowered but re-derived from an independent, physically grounded reference, rather than relying on the conventional values adopted in the original pipeline.

The newly introduced MEDF test (Figure 3.12) targets a different problem that the CSPD and VART filters, even after recalibration, could not fully detect. It addresses local spatial inconsistencies, where a vector disagrees significantly with its nearby neighbors without exceeding a specific speed or temporal-derivative threshold. Because the MEDF threshold of 0.3 m/s was set below the 99.73rd-percentile value of 0.4336 m/s derived from the drifter dV_{2h} distribution (subsection 3.2.4), the test flags a larger portion of the dataset than the strict three-sigma drifter-based criterion. This helps to more effectively identify additional questionable vectors, especially the denser group in the northern part of the domain that the recalibrated CSPD and VART tests would still miss.

Combining the recalibrated CSPD and VART thresholds with the newly introduced MEDF test led to the improvements seen in the new overall QCflag filter (Table 3.2). A comparison of Figure 3.13 and Figure 3.1 shows that the new configuration nearly cleans the area between Portofino and La Spezia. This area previously let many suspect vectors through under the old QCflag configuration. Most of the questionable vectors along the western Ligurian coast are also removed. Figure 3.14 further confirms that the new overall QCflag filter maintains the same general spatial pattern of the flagged data, while being systematically more conservative, resulting in a higher percentage of "bad data" flags across the domain. This outcome highlights that the improvement comes from a combination of two recalibrated tests and the new spatial-consistency test, each addressing a different type of measurement issue.

Temporal reconstruction skill. The results from the temporal artificial-hole experiment (Figure 4.5) show a rapid and steady decline in R^2 as the gap duration increases. R^2 starts above 0.70 for both velocity components at 2 hours and falls to below 0.27 at 24 hours. This pattern aligns with the design of the temporal stage of the Progressive Rim Gap-Filling Method, which uses a linear, persistence-based interpolation that works within a maximum window of 6 hours. During this time, the assumption that the flow changes approximately linearly between the two valid surrounding observations is supported by the data, as seen in the acceptable R^2 values recorded up to the 6-hour mark. After this point, the reconstructed values begin to increasingly diverge from the actual development of the current field.

Taken as a whole, the temporal reconstruction results show the strong dependence of the method performance on gap duration, which indicates a limited capacity to predict surface currents when no observations are available during long periods of time (exceeding a few hours). The duration of the gap is just one part of the larger problem of filling the gap. The following analysis of the spatial reconstruction experiment provides a complementary viewpoint, in which the method's skill is considered as a function of the growing size of the missing area, as well as the degree to which the spatial coherence of the surface current field can compensate for the absence of direct observations.

Spatial reconstruction skill. The spatial artificial-hole experiment (Figure 4.6) shows a different trend. R^2 stays high, above 0.86 for both components, for hole sizes up to 20 grid points and only starts to decline more noticeably beyond 50 points, eventually reaching values around 0.46-0.51 at the largest size tested (500 points). Compared to the temporal case, the decrease in reconstruction skill with larger hole sizes is much more gradual, especially in the small-to-medium size range that represents most gaps in the dataset (subsection 4.1 .3). This behavior matches the working principle of the spatial stage, which fills missing values by averaging the nearest valid neighboring grid points. Since the HFR-TirLig grid has a spatial resolution of 2 km and surface current structures in the study area typically extend over larger scales, nearby grid points remain well correlated with the missing ones even with a moderately sized hole, allowing for accurate reconstruction of local flow structures.

Oceanographic interpretation of the spatial reconstruction results. The gradual decline in spatial reconstruction skill observed, even for the largest artificial holes tested, can also be understood in the context of the characteristic length scales of mesoscale motion in the study area. Using glider observations, Schuckmann et al. 2020 report a first-mode Rossby radius of deformation of about 10-15 km in the northwestern Ligurian Sea. This finding aligns with the 5-12 km range calculated for the entire Mediterranean basin by Grilli et al. 1998. Mesoscale structures, such as eddies and meanders, typically extend over spatial scales that are two to three times larger than the local Rossby radius. For this study area, this means structures with a radius of about 20-45 km. A 500-grid-point hole, the largest size considered in the spatial artificial-hole experiment, occupies an area similar to a square of about 23 by 23 grid points. Given the 2 km horizontal resolution of the HFR-TirLig grid (section 2.4), this translates to a side length of about 46 km, a scale that matches the radius of the mesoscale eddies expected in this region. Seeing it this way, the relatively good reconstruction skill still achieved for the largest holes tested (R^2 of 0.46-0.51) is not surprising. Even a gap of this size can be smaller than, or similar to, a single coherent eddy or current-related structure. Thus, the valid observations near the hole may still hold a meaningful degree of spatial coherence with the missing area.

Comparison between the two stages. Overall, these results suggest that the spatial filling stage of the Progressive Rim Gap-Filling Method performs significantly better than the temporal stage, both in terms of the achieved R^2 values and in the durability of that performance as gap size increases. While the temporal stage shows noticeable degradation within its own operational range, the spatial stage maintains high reconstruction accuracy over a much broader portion of its tested range, with a sharp drop in performance limited to the largest and least representative hole sizes. From a methodological perspective, this indicates that whenever a data gap is present, prioritizing the spatial filling stage may provide more reliable reconstructed velocity estimates, especially for gaps lasting several hours or more.

Besides this general conclusion in favor of the spatial stage, the two sets of R^2 curves from the artificial-hole experiments (Figure 4.5 and Figure 4.6) offer a practical, quantitative criterion to select the correct interpolation mode for a given gap, instead of relying on a predetermined preference for one stage or the other. Consider, for example, that at a certain time step we have a gap of 200 consecutive grid points and we have valid observations at the previous and next hourly time steps. In this situation, the gap can be closed either by spatial interpolation viewing it as a 200-point spatial hole, or by temporal interpolation considering it as a 2-hour temporal hole at each affected grid point. The results in section 4.2 show that the 2-hour temporal R^2 (0.704 for EWCT, 0.796 for NSCT) is higher than the 200-point spatial R^2 (0.702 for EWCT, 0.721 for NSCT). In this case, temporal interpolation is expected to do a better job of reconstructing the missing values than spatial interpolation. As the example illustrates, the R^2 curves obtained in this study are used as a decision tool to select the most suitable interpolation mode for each gap depending on its spatial and temporal dimension, rather than applying the two stages in a fixed sequential order regardless of the gap's features.

Asymmetry between vertical and horizontal components. In both experiments, we find that in both temporal and spatial cases, the northward component (NSCT) always has a higher R2 than the eastward component (EWCT) for almost all tested hole dimensions. This asymmetry between the two velocity components is not unique to this dataset. Corgnati, Maristella Berta, et al. 2024 also found greater agreement between HFR-derived and drifter-derived velocities for the northward (v) component compared to the eastward (u) component when validating total velocity fields for the same HFR-TirLig network, especially in the sub-domain sampled by the radar stations along the distinctly north-south oriented Tuscan coastline. A fully conclusive explanation for this behavior was not found in that study either, but the convergence between the two independent analyses supports the pattern observed here, suggesting that the higher reconstruction skill obtained for NSCT may reflect a broader characteristic of how the region's velocity components are captured by this HFR network, rather than an artifact specific to the Progressive Rim Gap-Filling Method evaluated in this thesis.

Bias correction based on the regression slope. Another observation from the true-versus-reconstructed comparison in Figure 4.4 is the consistent deviation of the fitted regression line from the ideal 1:1 relationship. As mentioned in the description of that figure, the slope of the fitted line is consistently smaller than one in all artificial-hole experiments. This results in a direct and measurable effect on the reconstructed velocity values. A regression slope of, for example, 0.8 indicates that a true velocity of 1 m/s is typically reconstructed as about 0.8 m/s. In other words, the gap-filling process leads to a systematic underestimation of their magnitude, regardless of whether the values are positive or negative. This pattern aligns with the smoothing effect of interpolation-based reconstruction methods.

Since the regression slope represents the average fraction of the true velocity magnitude preserved by the reconstruction, dividing each reconstructed value by the corresponding slope coefficient would, statistically speaking, bring the reconstructed field closer to the true field. Following the earlier example, a reconstructed value of 0.8 m/s divided by a slope of 0.8 would give a corrected value of 1 m/s, restoring the original magnitude that interpolation smoothed out. It is important to emphasize that this correction would not retrieve the true value at any specific grid point or time step. The regression slope reflects an average bias across all reconstructed points rather than a precise, point-by-point relationship. Further refinement of this idea comes from the fact that the regression slope is not fixed; it changes depending on the size of the artificial hole. The extent of smoothing introduced by interpolation is expected to grow with the size of the gap being filled. From this perspective, the slope correction discussed earlier does not need to be a single, overall multiplicative factor. Instead, it could be customized for the specific time or space dimension of the reconstructed gap. This would involve using a slope value estimated separately for each hole size from the related artificial-hole experiment. formalizing and testing this size-dependent correction is beyond the scope of the present work, but it is a natural and practically motivated direction for the future development of the Progressive Rim Gap-Filling Method.

6. Conclusion

This thesis covered two key methods identified as necessary for effectively using high-frequency radar (HFR) surface current data from the HFR-TirLig network. These methods are the Quality Control (QC) of the total velocity field and the reconstruction of missing data through gap-filling. Both methods originate from a shared concern: the reliability and completeness of HFR-derived current fields directly affect how useful the data is for scientific and societal applications, including Search and Rescue (SAR) operations (Corgnati, Maristella Berta, et al. 2024; Lorente, Aguiar, et al. 2022).

6.1 Summary of the Work

The first part of the thesis (Chapter 3) focused on the revision of the QC pipeline applied to the HFR-TirLig total velocity dataset. Starting from the limitations of the original filter configuration, which was shown to leave a non-negligible number of physically inconsistent vectors in the dataset, particularly along the western Ligurian coastline and in the offshore southeastern portion of the domain, the thresholds of the existing CSPD and VART filters were tightened, and a new spatial-derivative filter, MEDF, was introduced to identify vectors that are locally inconsistent with their neighboring flow. The combined effect of these changes was propagated to the overall QCflag filter, which was shown to retain the general spatial pattern of the original configuration while being considerably more conservative, removing a larger fraction of suspect vectors without discarding an excessive amount of valid signal. This revised QC configuration is a necessary foundation for the second part of the work, since any gap-filling procedure applied to data containing residual non-physical vectors risks propagating those errors through the interpolated field.

The second part of the thesis (Chapter 4) developed and validated the gap-filling method named Progressive Rim Gap-Filling Method, an iterative, data-driven procedure that does not rely on a coastline geometry or a pre-computed modal basis, and which combines a temporal linear-interpolation stage, bounded to a 6-hour persistence window, with a spatial nearest-neighbor averaging stage. The original implementation of the method was reorganized into a modular, split architecture that separates the temporal and spatial stages explicitly, improving both the clarity and the testability of the procedure. To assess the reconstruction skill of the method in a controlled and reproducible way, an artificial-hole validation framework was designed and implemented: natural gaps in the quality-controlled dataset were first characterized statistically, in terms of both their temporal duration and spatial extent, in order to select realistic hole sizes for the validation experiments; artificial holes of these sizes were then inserted into otherwise complete portions of the dataset, using a region-growing algorithm for the spatial case and a border-validity check to ensure that generated holes were bordered by valid data. The reconstruction skill was quantified through the coefficient of determination R^2, separately for the temporal and spatial artificial-hole experiments and for the eastward (EWCT) and northward (NSCT) velocity components.

The results, discussed in detail in Chapter 5, showed that the two stages of the Progressive Rim Gap-Filling Method behave differently as the size of the missing-data region increases. The temporal stage is accurate for short gaps but its performance deteriorates rapidly once the gap duration approaches and exceeds the 6-hour persistence limit built into the method, consistent with the assumption of quasi-linear flow evolution over short intervals. The spatial stage, by contrast, maintains a high reconstruction skill over a much wider range of hole sizes, reflecting the fact that mesoscale structures in this study area have a radius on the order of 20-45 km, corresponding to the size of a 500-grid-point hole. Taken together, these results indicate that the spatial filling stage is the more robust of the two components of the method, and that the overall reconstruction skill of the Progressive Rim Gap-Filling Method is primarily limited by the temporal stage when gaps of long duration are encountered.

6.2 Significance of the Results

The revised QC process and the validated gap-filling method create a clear processing chain that transforms the HFR-TirLig dataset from raw radar output to a continuous, quality-assured current product ready for operational use. Additionally, the artificial-hole framework developed in this study stands as a methodological contribution on its own. It is rooted in the statistical analysis of the dataset's natural gap distribution rather than using arbitrary hole sizes. This approach offers a validation standard that accurately reflects the missing-data conditions faced by the HFR-TirLig network, and it could be reused to assess different gap-filling methods on the same dataset.

6.3 Limitations

The main limitation of this work is that validating the gap-filling method relies solely on artificial holes placed within otherwise good data. While this method allows for controlled, quantitative evaluation of reconstruction skill, it does not confirm that the reconstructed fields match independent, in-situ observations of ocean surface circulation.

6.4 Future Work

Considering this, a straightforward extension of the current validation framework is to compare the gap-filled velocity fields against independent Lagrangian observations from the drifter campaigns mentioned in chapter 2. This would use a method similar to that applied in assessing the OMA method for the same network (Corgnati, Maristella Berta, et al. 2024). Such a comparison, which involves real rather than artificial gaps, would provide a complementary test of the Progressive Rim Gap-Filling Method's practical reconstruction skill. Targeted drifter deployment could also help clarify questions raised in this thesis, such as why the northward velocity component (NSCT) shows systematically higher R^2 values than the eastward component (EWCT) in both the temporal and spatial artificial-hole experiments (section 4.2). A denser, more targeted in-situ dataset would enable investigation into whether this asymmetry reflects a true oceanographic feature of the study area (possibly related to the predominant orientation of the boundary current system described in section 2.1) or if it arises from the methodological aspects of the gap-filling procedure itself.

Another direction involves systematically comparing the Progressive Rim Gap-Filling Method with other documented gap-filling techniques, such as open-boundary modal analysis, already evaluated for the HFR-TirLig network (Corgnati, Maristella Berta, et al. 2024). This would help characterize the strengths of a purely data-driven, geometry-independent method versus those that use boundary geometry.

Beyond these two extensions of the validation framework, there are several other directions that could expand the impact and applicability of this work. From an operational point of view, the revised QC pipeline and the Progressive Rim Gap-Filling Method developed here could be proposed to the wider European HF radar community. This would follow the networking and standardization efforts already promoted through events like the biennial European HFR conferences. Offering the method beyond the HFR-TirLig network would allow other operators to adopt a validation standard based on their own datasets, promoting uniform QC and gap-filling practices across the European HFR network.

In a similar operational vein, a dedicated production line could be created to deliver the gap-filled current product directly to end users. For instance, there could be a direct interface with the Geographic Information System (GIS) used by the Italian Coast Guard. This would allow the continuous, quality-assured current fields produced by this method to be integrated directly into existing maritime safety and Search and Rescue workflows, rather than remaining solely as a research product.

6.5 Closing Remarks

In conclusion, by improving the QC of the total velocity dataset and providing a validated, low-cost gap-filling procedure along with a reusable assessment framework, this work aids in producing continuous, quality-assured current fields necessary for various applications. These include Search and Rescue operations, maritime safety and broader environmental monitoring of a coastal region that is significant ecologically and socially.

Bibliography

Alberola C Millot Claude, Font J (1995). "On the seasonal and mesoscale variabilities of the Northern Current during the PRIMO-0 experiment in the western Mediterranean-sea". In: *Oceanologica Acta* 18.2, pp. 163–192. URL: https://archimer.ifremer.fr/doc/00096/20770/.

Angeletti, Lorenzo, Giorgio Castellan, Paolo Montagna, Alessandro Remia, and Marco Taviani (Aug. 2020). "The "Corsica Channel Cold-Water Coral Province" (Mediterranean Sea)". In: *Frontiers in Marine Science* 7, pp. 1–15. DOI: 10.3389/fmars.2020.00661.

Astraldi, M. and G. P. Gasparini (1992). "The seasonal characteristics of the circulation in the north Mediterranean basin and their relationship with the atmospheric-climatic conditions". In: *Journal of Geophysical Research: Oceans* 97.C6, pp. 9531–9540. DOI: 10.1029/92JC00114.

Astraldi, M., G. P. Gasparini, G. M. R. Manzella, and T. S. Hopkins (1990). "Temporal variability of currents in the eastern Ligurian Sea". In: *Journal of Geophysical Research: Oceans* 95.C2, pp. 1515–1522. DOI: 10.1029/JC095iC02p01515.

Astraldi, M. et al. (1995). "Climatic fluctuations, current variability and marine species distribution - a case-study in the Ligurian sea (north-west Mediterranean)". In: *Oceanologica Acta* 18.2, pp. 139–149. URL: https://archimer.ifremer.fr/doc/00096/20768/.

Barth, Alexander, A. Alvera-Azcárate, Rixen Michel, and J.-M Beckers (2005). "Two-way Nested Model of Mesoscale Circulation Features in the Ligurian Sea". In: *Progress In Oceanography* 66, pp. 171–189. DOI: 10.1016/j.pocean.2004.07.017.

Berta, M., P.-M. Poulain, R. Sciascia, A. Griffa, and M. G. Magaldi (2021). *CARTHE drifters deployment within the DDR20 - 'Drifter Demonstration and Research 2020' experiment in the NW Mediterranean Sea*. [Data set]. Seanoe. doi:10.17882/85161.

Berta, M., R. Sciascia, M. G. Magaldi, and A. Griffa (2020). *Drifter deployment in the framework of the IMPACT (Port impacts on marine protected areas: transnational cooperative actions) project*. [Data set]. Seanoe. doi:10.17882/72369.

Breivik, Oyvind, Arthur Allen, Christophe Maisondieu, and M. Olagnon (Nov. 2012). "Advances in Search and Rescue at Sea". In: *Ocean Dynamics* 63. DOI: 10.1007/s10236-012-0581-1.

Bushnell, Michael and Heather Worthington (2022). *Manual for Real-Time Quality Control of High Frequency Radar Surface Currents (HFR QARTOD Manual Update Final-1b)*. IOOS QARTOD. URL: https://cdn.ioos.noaa.gov/media/2022/07/HFR_QARTOD_Manual_Update_Final-1b.pdf (visited on 06/02/2026).

Choukroun, Severine, Owen Stewart, Luciano Mason, and Michael Bode (Dec. 2024). "Larval dispersal predictions are highly sensitive to hydrodynamic modelling choices". In: *Coral Reefs* 44, pp. 1–13. DOI: 10.1007/s00338-024-02563-z.

CNR-ISMAR (2016). *Near Real Time Surface Ocean Velocity by HFR-TirLig Network*. European HFR Node. DOI: 10.57762/35S0-EE87. URL: https://doi.org/10.57762/35S0-EE87.

Corgnati, Lorenzo (2024). *Recommendation Report 2 on Improved Common Procedures for HFR QC Analysis*. Deliverable 5.14. Version 2.0. JERICO-NEXT, WP5 Data Management, p. 17.

Corgnati, Lorenzo, Maristella Berta, Zoi Kokkini, Carlo Mantovani, Marcello Magaldi, Anne Molcard, and Annalisa Griffa (July 2024). "Assessment of OMA Gap-Filling Performances for Multiple and Single Coastal HF Radar Systems: Validation with Drifter Data in the Ligurian Sea". In: *Remote Sensing* 16, p. 2458. DOI: 10.3390/rs16132458.

Corgnati, Lorenzo, Carlo Mantovani, Annalisa Griffa, Maristella Berta, Pierluigi Penna, Paolo Celentano, Lucio Bellomo, Daniel Carlson, and Raffaele D'Adamo (May 2018). "Implementation and Validation of the ISMAR High-Frequency Coastal Radar Network in the Gulf of Manfredonia (Mediterranean Sea)". In: *IEEE Journal of Oceanic Engineering* PP, pp. 1–22. DOI: 10.1109/JOE.2018.2822518.

Cowen, Robert and Su Sponaugle (Jan. 2009). "Larval Dispersal and Marine Population Connectivity". In: *Annual review of marine science* 1, pp. 443–66. DOI: 10.1146/annurev.marine.010908.163757.

Di Maio, Antonia, Martin Mathew, and Roberto Sorgente (Aug. 2016). "Evaluation of the search and rescue LEEWAY model in the Tyrrhenian Sea: A new point of view". In: *Natural Hazards and Earth System Sciences* 16, pp. 1979–1997. DOI: 10.5194/nhess-16-1979-2016.

Doronzo, Bartolomeo, Michele Bendoni, Stefano Taddei, Angelo Boccacci, and Carlo Brandini (July 2025). "A Combined HF Radar and Drifter Dataset for Analysis of Highly Variable Surface Currents". In: *Data* 10, p. 115. DOI: 10.3390/data10070115.

Enrichetti, Francesco, Marzia Bo, Carla Morri, Monica Montefalcone, Margherita Toma, Giorgio Bavestrello, Leonardo Tunesi, Simonepietro Canese, Michela Giusti, Eva Salvati, Rosa Maria Bertolotto, and Carlo Nike Bianchi (2019). "Assessing the environmental status of temperate mesophotic reefs: A new, integrated methodological approach". In: *Ecological Indicators* 102, pp. 218–229. DOI: 10.1016/j.ecolind.2019.02.028.

European HFR Node (2026a). *HFR-TirLig*. URL: https://www.hfrnode.eu/networks/hfr-tirlig/ (visited on 06/02/2026).

— (2026b). *HFR-TirLig*. European HFR Node. URL: https://www.hfrnode.eu/networks/hfr-tirlig/ (visited on 07/07/2026).

— (2026c). *High Frequency Radar Technology*. URL: https://www.hfrnode.eu/hf-radar-technology/ (visited on 07/07/2026).

Fang, Y.-C., T. J. Weingartner, R. A. Potter, P. R. Winsor, and H. Statscewich (2015). "Quality Assessment of HF Radar-Derived Surface Currents Using Optimal Interpolation". In: *Journal of Atmospheric and Oceanic Technology* 32.2, pp. 282–296. DOI: 10.1175/JTECH-D-14-00123.1.

Fredj, Erick, Hugh Roarty, Josh Kohut, Michael Smith, and Scott Glenn (Feb. 2016). "Gap Filling of the Coastal Ocean Surface Currents from HFR Data: Application to the Mid-Atlantic Bight HFR Network". In: *Journal of Atmospheric and Oceanic Technology* 33, p. 160222131645008. DOI: 10.1175/JTECH-D-15-0056.1.

Gauci, Adam, Aldo Drago, and John Abela (2016). "Gap Filling of the CALYPSO HF Radar Sea Surface Current Data through Past Measurements and Satellite Wind Observations".

In: *International Journal of Navigation and Observation* 2016.1, p. 2605198. DOI: 10.1155/2016/2605198.

Grilli, Federica and Nadia Pinardi (1998). “The computation of Rossby radii of deformation for the Mediterranean Sea”. In: *MTP news* 6.4, pp. 4–5.

Hernández-Carrasco, Ismael, Lohitzune Solabarrieta, Anna Rubio, Ganix Esnaola, Emma Reyes, and Alejandro Orfila (Mar. 2018). “Impact of HF radar currents gap-filling methodologies on the Lagrangian assessment of coastal dynamics”. In: *Ocean Science Discussions*, pp. 1–34. DOI: 10.5194/os-2018-26.

Jones, Geoffrey, Glenn Almany, Russ GR, Peter Sale, Steneck RR, Madeleine van Oppen, and Bette Willis (June 2009). “Larval retention and connectivity among populations of corals and reef fishes: History, advances and challenges”. In: *Coral Reefs* 28, pp. 307–325. DOI: 10.1007/s00338-009-0469-9.

Jones, Geoffrey, Maya Srinivasan, and Glenn Almany (Sept. 2007). “Population Connectivity and Conservation of Marine Biodiversity”. In: *Oceanography* 20, pp. 100–111. DOI: 10.5670/oceanog.2007.33.

Laws, Kenneth E. (2001). “Measurements of Near Surface Ocean Currents Using HF Radar”. Ph.D. dissertation. University of California, Santa Cruz.

Lorente, Pablo, E. Aguiar, Michele Bendoni, Maristella Berta, Carlo Brandini, Alejandro Caceres Euse, Fulvio Capodici, Daniela Cianelli, Giuseppe Ciraolo, Lorenzo Corgnati, V. Dadic, Bartolomeo Doronzo, A. Drago, Dylan Dumas, Pierpaolo Falco, Maria Fattorini, Adam Gauci, Roberto Gomez, Annalisa Griffa, and V. Cardin (June 2022). “Coastal high-frequency radars in the Mediterranean -Part 1: Status of operations and a framework for future development”. In: pp. 761–795.

Lorente, Pablo, Silvia Piedracoba, Marcos Sotillo, and Enrique Alvarez Fanjul (Jan. 2019). “Long-Term Monitoring of the Atlantic Jet through the Strait of Gibraltar with HF Radar Observations”. In: *Journal of Marine Science and Engineering* 7, p. 16. DOI: 10.3390/jmse7010003.

Ludwig, Wolfgang, Egon Dumont, Michel Meybeck, and Serge Heussner (2009). “River discharges of water and nutrients to the Mediterranean and Black Sea: Major drivers for ecosystem changes during past and future decades?” In: *Progress in Oceanography* 80.3, pp. 199–217. DOI: 10.1016/j.pocean.2009.02.001.

Mantovani, Carlo, Lorenzo Corgnati, Jochen Horstmann, Anna Rubio, Emma Reyes, Céline Quentin, Simone Cosoli, Jose Luis Asensio, Julien Mader, and Annalisa Griffa (2020). “Best Practices on High Frequency Radar Deployment and Operation for Ocean Current Measurement”. In: *Frontiers in Marine Science* 7. DOI: 10.3389/fmars.2020.00210.

Millot, Claude (1999). “Circulation in the Western Mediterranean Sea”. In: *Journal of Marine Systems* 20.1, pp. 423–442. DOI: 10.1016/S0924-7963(98)00078-5.

Paduan, Jeffrey and Hans Graber (Jan. 1997). “Introduction to High-Frequency Radar: Reality and Myth”. In: *Oceanography* 10, pp. 36–39. DOI: 10.5670/oceanog.1997.18.

Paduan, Jeffrey and Libe Washburn (Dec. 2011). “High-Frequency Radar Observations of Ocean Surface Currents”. In: *Annual review of marine science* 5. DOI: 10.1146/annurev-marine-121211-172315.

Planes, Serge, Geoffrey Jones, and Simon Thorrold (May 2009). “Larval dispersal connects fish populations in a network of marine protected areas”. In: *Proceedings of the National*

Academy of Sciences of the United States of America 106, pp. 5693–7. DOI: 10.1073/pnas.0808007106.

Reyes, E., E. Aguiar, M. Bendoni, M. Berta, C. Brandini, A. Cáceres-Euse, F. Capodici, V. Cardin, D. Cianelli, G. Ciraolo, L. Corgnati, V. Dadić, B. Doronzo, A. Drago, D. Dumas, P. Falco, M. Fattorini, M. J. Fernandes, A. Gauci, R. Gómez, A. Griffa, C.-A. Guérin, I. Hernández-Carrasco, J. Hernández-Lasheras, M. Ličer, P. Lorente, M. G. Magaldi, C. Mantovani, H. Mihanović, A. Molcard, B. Mourre, A. Révelard, C. Reyes-Suárez, S. Saviano, R. Sciascia, S. Taddei, J. Tintoré, Y. Toledo, M. Uttieri, I. Vilibić, E. Zambianchi, and A. Orfila (2022). "Coastal high-frequency radars in the Mediterranean – Part 2: Applications in support of science priorities and societal needs". In: *Ocean Science* 18.3, pp. 797–837. DOI: 10.5194/os-18-797-2022.

Roarty, Hugh, Thomas Cook, Lisa Hazard, Doug George, Jack Harlan, Simone Cosoli, Lucy Wyatt, Enrique Alvarez Fanjul, Eric Terrill, Mark Otero, John Largier, Scott Glenn, Naoto Ebuchi, Brian Whitehouse, Kevin Bartlett, Julien Mader, Anna Rubio, Lorenzo Corgnati, Carlo Mantovani, and Stephan Grilli (May 2019). "The Global High Frequency Radar Network". In: *Frontiers in Marine Science* 6. DOI: 10.3389/fmars.2019.00164.

Roarty, Hugh, Teresa Updyke, Laura Nazzaro, Michael Smith, Scott Glenn, and Oscar Schofield (May 2024). "Real-time quality assurance and quality control for a high frequency radar network". In: *Frontiers in Marine Science* 11. DOI: 10.3389/fmars.2024.1352226.

Röhrs, Johannes, Ann Sperrevik, Kai Christensen, Göran Broström, and Oyvind Breivik (May 2015). "Comparison of HF radar measurements with Eulerian and Lagrangian surface currents". In: *Ocean Dynamics*. DOI: 10.1007/s10236-015-0828-8.

Rubio, Anna, Julien Mader, Lorenzo Corgnati, Carlo Mantovani, Annalisa Griffa, Antonio Novellino, Céline Quentin, Lucy Wyatt, Johannes Schulz-Stellenfleth, Jochen Horstmann, Pablo Lorente, Enrico Zambianchi, Michael Hartnett, Carlos Fernandes, Vassilis Zervakis, Patrick Gorringe, Angélique Melet, and Ingrid Puillat (2017). "HF Radar Activity in European Coastal Seas: Next Steps toward a Pan-European HF Radar Network". In: *Frontiers in Marine Science* 4. DOI: 10.3389/fmars.2017.00008.

Sammari, C., C. Millot, and L. Prieur (1995). "Aspects of the seasonal and mesoscale variabilities of the Northern Current in the western Mediterranean Sea inferred from the PROLIG-2 and PROS-6 experiments". In: *Deep Sea Research Part I: Oceanographic Research Papers* 42.6, pp. 893–917. DOI: 10.1016/0967-0637(95)00031-Z.

Schuckmann, Karina von et al. (2020). "Copernicus Marine Service Ocean State Report, Issue 4". In: *Journal of Operational Oceanography* 13.sup1, S1–S172. DOI: 10.1080/1755876X.2020.1785097.

Sciascia, Roberta, Katell Guizien, and Marcello G Magaldi (Sept. 2022). "Larval dispersal simulations and connectivity predictions for Mediterranean gorgonian species: sensitivity to flow representation and biological traits". In: *ICES Journal of Marine Science* 79.7, pp. 2043–2054. DOI: 10.1093/icesjms/fsac135.

Sciascia, Roberta, Marcello Magaldi, and Anna Vetrano (Dec. 2018). "Current reversal and associated variability within the Corsica Channel: The 2004 case study". In: *Deep Sea Research Part I: Oceanographic Research Papers* 144. DOI: 10.1016/j.dsr.2018.12.004.

Staaterman, Erica and Claire B. Paris (June 2014). “Modelling larval fish navigation: the way forward”. In: *ICES Journal of Marine Science* 71.4, pp. 918–924. DOI: 10.1093/icesjms/fst103.

Stewart, Robert H. and Joseph W. Joy (1974). “HF radio measurements of surface currents”. In: *Deep Sea Research and Oceanographic Abstracts* 21.12, pp. 1039–1049. DOI: https://doi.org/10.1016/0011-7471(74)90066-7.

Swearer, Stephen, Eric Treml, and Jeffrey Shima (Aug. 2019). “A Review of Biophysical Models of Marine Larval Dispersal”. In: pp. 325–356. ISBN: 9780429026379. DOI: 10.1201/9780429026379-7.

Virtanen, Pauli, Ralf Gommers, Travis Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan Walt, Matthew Brett, Joshua Wilson, K. Millman, Nikolay Mayorov, Andrew Nelson, Eric Jones, Robert Kern, Eric Larson, and Yoshiki Vázquez-Baeza (Feb. 2020). “SciPy 1.0: fundamental algorithms for scientific computing in Python”. In: *Nature Methods* 17, pp. 1–12. DOI: 10.1038/s41592-019-0686-2.

Zhao, Hong, Fan Min, and William Zhu (Jan. 2013). “Test-Cost-Sensitive Attribute Reduction of Data with Normal Distribution Measurement Errors”. In: *Mathematical Problems in Engineering* 2013. DOI: 10.1155/2013/946070.