Real-Time Quality Control of Current Velocity Data on Individual Grid Cells in WERA HF Radar

Roberto Gomez, Thomas Helzel, Leif Petersen,
Matthias Kniephoff
Helzel Messtechnik GmbH
Kaltenkirchen, Germany
gomez@helzel.com

Clifford R. Merz, Yonggang Liu, Robert H. Weisberg
College of Marine Science
University of South Florida
St. Petersburg, Florida, United States
cmerz@usf.edu

***Abstract*—Ocean surface current data provided by shore-based High Frequency (HF) radar show sporadically non-realistic current vectors in some areas, particularly at the outer edges of the radar range. The present paper describes a quality control (QC) procedure which evaluates the data of an array-type HF radar system in near real-time and assigns a quality value on every single grid-cell measurement. The quality value can then be used to identify non-realistic measurements and remove them from the obtained current maps. This procedure was implemented on a pair of coastal radar systems of the University of South Florida during a measuring period of 9 months. It was observed that the method is very effective in removing occasional non-realistic measurements. In general, the WERA QC procedure has to be tuned to the specific local conditions of the system for optimal results and to have the right trade-off between correct outlier identification and false alarm rate according to the user requirements.**

Keywords—remote sensing; currents; ocean radar; quality control.

I. INTRODUCTION

The array type ocean radar WERA (WEllen RAdar) [1] is a shore based remote sensing system to monitor ocean surface currents, waves and wind direction. The measurement range of these systems can reach out to more than 200 km for current mapping or more than 100 km for wave measurements, depending on the operating frequency.

Publications about the results from systems installed all over the world have proved the accuracy of these systems [2, 3]. The accuracy of the current mapping was tested by means of comparisons between the measured data of ADCPs or drifters with the ocean radar system. The results from various studies typically shows an excellent correlation factor of >0.9 [3].

The temporal availability of ocean surface current data provided by these systems is very high. Nevertheless the resulting current maps show from time to time non-realistic current vectors in some areas, in particular at the outer edges of the range. There are many reasons for such measurements to appear, including radio frequency interference, power-line harmonics interference, ionosphere clutter and ship echoes. There have been many efforts in reducing the effects of such phenomena in the data. Many of these techniques have already been included in the processing software of WERA [4].

However, not all of the 'affected' data are completely removed, which calls for better quality control (QC) in the measured data. Many users of the WERA community have already developed their own QC procedures. One example is [5], where general QC guidelines for oceanographic data are outlined without referring specifically to data from a High Frequency (HF) radar. Other examples are [2] and [6], where QC procedures are employed offline (not in real-time) for WERA HF radar data. In [7], an interesting approach using only past data to evaluate the quality of the measured value is explored, but not yet available as standard tool in the WERA system.

The objective of this work is to implement and evaluate a QC procedure to identify and remove outliers in the near real-time processing on WERA HF radar.

II. CHARACTERISTIC OF OCEAN RADAR CURRENT DATA

Each ocean radar provides maps of radial current velocity which are derived from acquired data that are integrated over an ocean area (in range and azimuth direction), in depth and time. The integration time for array type systems is typically 5 to 15 minutes for current mapping. The integration in depth is about 1 m, depending on the operating frequency. The integration in range direction (beam direction) is given by the allocated operating bandwidth and it does not change with the distance from the radar. In contrast to range, the integration area in azimuth increases with increasing distance from the radar as this parameter is defined by the beam width which is defined by the length of the array (e.g. 6 degrees for a 16 antenna array). The resulting integration area is a segment of an annulus [8].

For most applications, data from two ocean radar systems are combined to generate maps of current velocity vectors given as u and v components (eastward and northward respectively). The measured radial current data from each radar site are transformed to a common and pre-defined Cartesian grid [9]. The resulting total current data are derived in an overlapping area.

III. QC PROCEDURE AND OUTLIER IDENTIFICATION

Although most applications will deliver *u* and *v* current velocity components combined from two HF radars, the QC procedure is applied on each radial velocity data in Cartesian coordinates. The quality values obtained are then stored in a database.

By default, the WERA QC recognizes 4 different levels of quality:

- Quality value 1: Is the best quality level that can be achieved.
- Quality value 2: Medium quality level.
- Quality value 3: Deficient quality, but not yet considered as an outlier.
- Quality value larger than 3: Non-realistic measurement that should be considered as an outlier.

The QC procedure consists of performing a series of tests on the measured value on each grid cell in the latest radial current map obtained (near real-time). The tests evaluate the measured data in two different perspectives: general evaluation of measured values and evaluation in time domain. Before applying the tests, each measured value is assumed to have a quality of 1. Failing a quality test is penalised by adding an integer value to the quality level of the measurement. At the end, each grid cell of the actual measurement will have a quality level number that starts from 1, with quality levels greater than 3 considered to be outliers and therefore may be optionally removed when creating maps as seen in Figs. 1, 2 and 3. This option can be as well exploited to obtain cleaner current velocity vector maps in eastward and northward components when combining radial velocities from two or more radars, by omitting outliers on all combined radial current velocity data sets.

It is important to note that no data is deleted during the QC procedure. The data are just individually marked with integer numbers to indicate the quality level of each measured value.

A. General Evaluation of Measured Values

For the general evaluation of measured values, the program verifies that the current velocity falls into an expected normal range.

Additionally, on WERA HF radars, each radial current velocity value measured is accompanied with an accuracy value. The accuracy is obtained from a statistical analysis of the samples collected on each grid-cell measurement and considers signal-to-noise, variance, and number of samples [9]. This accuracy value is itself an indicator of quality of the measurement [2] and therefore one of the tests consists of classifying the value under different quality levels.

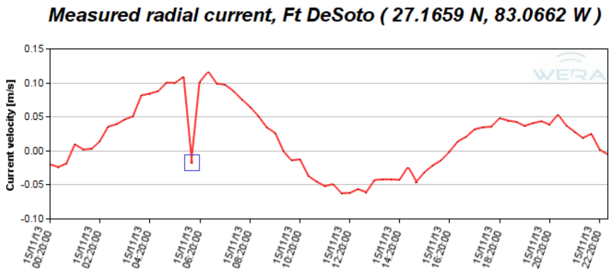

Fig. 1. Example of an outlier on Nov. 15th 2013 at 6:00.

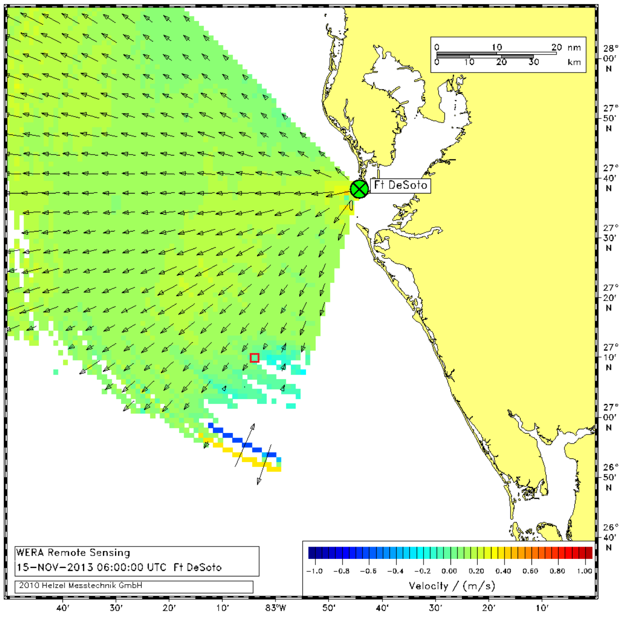

Fig. 2. Original map of radial current velocity.

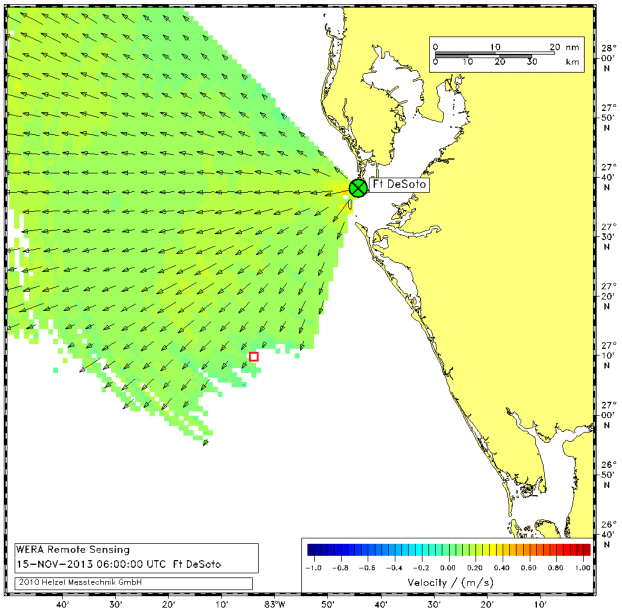

Fig. 3. Same map of radial current velocity with outliers removed.

B. *Evaluation in Time Domain*

Because of the integration effects described in chapter II, all data are smoothed and thus even an extreme event on the ocean should not cause an abrupt change of measured value in the same grid cell from one measurement to the next in time. Based on this assumption, the QC procedure verifies the consistency of the measured value by looking at the past values in the same grid cell.

A first test consists of verifying that the derivative (change in radial current velocity measurement with respect to the last measured, divided by time elapsed) lies between allowed limits.

An additional test consists of verifying that the measurement is near an expected value according to a time tendency. This is done using past measured data to curve-fit a 3rd order polynomial and estimate a value for the actual measurement using extrapolation, as described in [7]. The difference between the estimated value and the measured value is used as an indicator of quality.

It is important to note that the time domain tests will not include past values which were already flagged as outliers. Also, if the measurement immediately before is not available or is already marked as outlier, no tests in time-domain will be performed.

C. *Configurable Parameters*

The QC procedure employed on the WERA HF radar can be configured and optimized to the local behaviour of the data. The parameters which can be configured are the following:

- Minimum expected current velocity, units in m/s.
- Maximum expected current velocity, units in m/s.
- Upper threshold for accuracy value, units in m/s.
- Lower threshold for accuracy value, units in m/s.
- Maximum derivative of velocity, units in m/s·hr (maximum change of current velocity per hour).
- Amount of hours to look back in time for gathering measured data to estimate an expected value.
- Minimum number of measurement points required to estimate an expected value.
- Maximum allowed number of consecutive missing data points or outliers when gathering measured data to estimate an expected value.
- Maximum allowed difference between expected value and measured value.

All these parameters are valid for both stations and for the whole coverage area.

IV. CASE OF STUDY

The QC procedure was implemented on a pair of WERA HF radar systems of the University of South Florida during a measuring time period of 9 months (from May 2013 to January 2014). The systems are part of the real-time Coastal Ocean Monitoring Prediction System (COMPS) for the West Florida Shelf region [10-12]. A discussion of the evolution of the USF HF Radar Network on the West Florida Shelf can be found in reference [13]. Both radar sites, referred as Fort DeSoto and Venice, operate between 12.275 – 13.2 MHz with 12 receive (RX) antennas and an integration time of 15 minutes for each measurement data set. Their angular field of view is ±60° from boresight with a maximum measurement range of 150 km. Each of the systems provides a map of radial current velocity every 20 minutes (3 times per hour on minute 00, 20 and 40). Fig. 4 shows the position, field of view and range of each of the radars.

All data measured from both systems were stored into a database and processed in near real-time with the QC procedure described in chapter 3. Resulting quality values were stored back in the database. The QC parameters used for the system are listed in Table 2 and were empirically tuned using data from one month from both stations.

TABLE I. QC TESTS AND THEIR RESPECTIVE QUALITY PENALTY

Test	Increment on quality value
Current velocity values within expected range	Within range: 0 Out of range: +2
Current velocity accuracy value below thresholds	Below lower threshold: 0 Lower threshold exceeded: +1 Upper threshold exceeded: +2
Absolute value of derivative below threshold	Below threshold: 0 Threshold exceeded: +3
Difference between 3rd order extrapolated value and measured value	Below threshold: 0 Not enough points for reliable extrapolation: +1 Threshold exceeded: +2

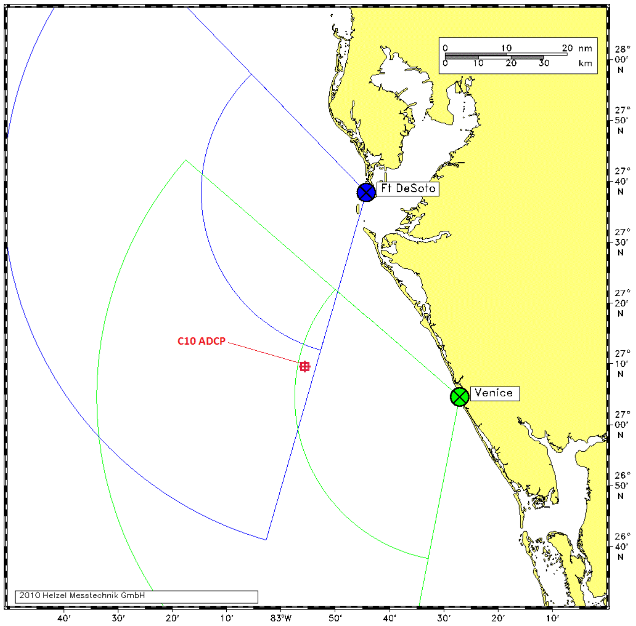

Fig. 4. Position and range of USF WERA systems used for case of study.

TABLE II. QC CONFIGURATION PARAMETERS USED IN CASE OF STUDY

Parameter	Value
Minimum expected current velocity	-0.45 m/s
Maximum expected current velocity	0.45 m/s
Upper threshold for accuracy	0.07 m/s
Lower threshold for accuracy	0.025 m/s
Maximum derivative of velocity	0.35 m/s·hr
Amount of hours to look in the past	10 hours
Minimum number of measurements	25
Maximum consecutive missing data points	2
Maximum difference between expected value and measured value	0.1 m/s

Additionally, hourly data measured from an Acoustic Doppler Current Profiler (ADCP) is available from May 24th to September 3rd. The ADCP is called C10 and its location is marked on Fig. 4 (27° 9' 36" N, 82° 55' 48" W). It lies in the field of view of both radars, approximately 55 km from Fort DeSoto at an angle of -56° from boresight, and 48 km from Venice at an angle of +30° from boresight [2, 14]. Data provided by the ADCP is the eastward and northward current velocity components (*u* and *v*) at a depth of 4 m.

V. METHODOLOGY FOR EVALUATION OF QC PROCEDURE

To evaluate the performance of the QC procedure, it is required to have an alternative method to determine which measurements are outliers. This method should be at least more reliable and robust as the one used in real-time, so that it could be used as reference. Then it would be possible to compare the "reliably-identified" outliers with the outliers flagged by WERA to verify if the QC procedure actually identified them correctly. Two ways are proposed for such task.

A. Low Pass Filter on the Measured Data as Reference

One approach is to select several grid cells in the coverage area with different distance to the shore. Then, apply a low pass filter to the radial current velocity time series measured by the radar to smooth the data and reveal spikes which will be then considered to be outliers. This method is assumed to be more robust and reliable than what it is used in the WERA QC real-time procedure since it uses also data from the future.

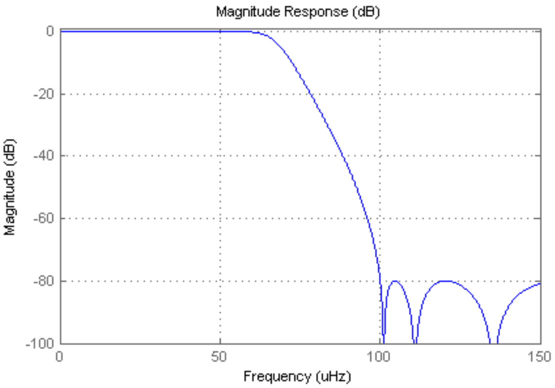

Fig. 5. Low pass filter used to smooth the data.

It was decided to pick the grid cells that are located in front of each station (center beam, perpendicular to RX antenna array) at distances of 30, 50, 75 and 100 km away from the shore. This would result in a total of 8 data sets, 4 from each WERA system. The radial current velocity data and their respective quality values provided on these grid cells during the 9 month period will be used to determine how good the QC procedure is.

The methodology to be used on each of the 8 data sets to identify outliers is basically a zero-phased low pass digital filter. The filter is a 10th order low pass Chebyshev type II filter set to 100 µHz as cut frequency and -80 dB as stop band. The filter processes the data in the forward and reverse direction to obtain zero-phase shift and double the order.

The proposed methodology steps are the following:

1) Since the filter requires a continuous stream of data with no missing points the 9-month-long time series is split at points where a gap of data of more than 4 hours is found. This will result in subsections of data which will be filtered separately. Splitting the data is required to overcome periods of either no received data or periods of minimal data return, for example for the outer ranges (75 km and 100 km) where the data availability may be limited.

2) All remaining points where no data are available are filled using a 3rd order polynomial curve-fit interpolation. The curve-fit uses a moving window of 10 hours with the point to be filled being in the middle (5 hours in the past and 5 hours in the future). The only reason for interpolating these points is to make possible the application of the filter without introducing a significant deviation to the tendency of the time series. Interpolated points are actually not used as reference or compared to the filtered reference for outlier identification.

3) Now that subsections of continuous data streams are available, we can proceed to filter the data. The only exception is those subsections of data with less than 30 points since is the minimum data points required for applying the filter (3 times the order). These short subsections of data periods will be left out of the analysis as will be clear in Table 3 presented in the next chapter.

4) All measured data from WERA that was not left out of the analysis is compared to the filtered reference. A difference of more than 0.1 m/s is then considered to be an outlier. A threshold of 0.1 m/s was used to be consistent with the parameters of the WERA QC procedure (see Table II).

5) Outliers identified are compared to the quality value assigned by the WERA real-time QC procedure. If the quality value is greater than 3 it is counted as confirmed outlier identification, otherwise as false outlier identification.

B. Using ADCP measured Data as Reference

A second approach is to use the data measured from the moored buoy mounted C10 ADCP as the reference. In this case, only the WERA measurements at minute 00 will be considered, since the ADCP provides hourly averaged data. The hourly averaged ADCP data, consists of a 6-minute average of 360 pings, one ping per second, starting at minute 00.

It is important to note that the location of the ADCP is at the very edge of the visible field of view of Fort DeSoto radar (-56° from boresight) and additionally a small island is located in between. For these reasons, a strong variability in the data measured by Fort DeSoto site coming from the ADCP coordinates is expected and the evaluation of the QC procedure may not be under the same circumstances as other points. Considering the above, only data provided from the Venice radar will be used for this approach.

To use this approach, the radial current velocity as seen from Venice radar needs to be calculated from the eastward and northward components provided by the ADCP. Afterwards, the WERA measurements that deviate by more than 0.1 m/s from the ADCP measurement will be considered outliers. Data points provided by ADCP where no WERA measurement is available will not participate in the analysis.

VI. RESULTS

A. Using Low Pass Filter as Reference

A time series example of the low pass filter methodology to evaluate the QC procedure is shown in Fig. 6. Table III shows the results of this methodology.

From Fig. 6, it can be seen that the filter does a very good job in smoothing the data and setting the reference, from which outliers will be identified.

From Table III, it is observed that the method is at all times identifying more than half the outliers present in the data, and in one case even identified 100% of the outliers found.

This methodology for evaluating the QC procedure is particularly useful since no in-situ instrument data is required. Furthermore, for the mere objective of identifying outliers or spikes in the time-series without involving the verification of accuracy of measurements, the filtered signal appears to be very effective.

However, it is worthy to mention that as the WERA QC procedure identifies confirmed outliers it also marks some measurements as outlier, which after post-processing were observed to be good measurements (around 60% in average, of all outliers marked by WERA QC). False outlier identification percentage can be diminished by decreasing the difference threshold value, but this will result in a lower confirmed outlier identification percentage.

B. Using ADCP Data as Reference

Table IV shows the results of evaluating the QC procedure using the ADCP data as a reference. As observed, the percentage of confirmed outlier identification is very low in comparison with the results using the filtered measurements as reference.

However, when comparing the data provided by the ADCP and the data provided by WERA additional factors are indirectly playing a role in the results obtained [2]. The top bin of the ADCP velocity measurement is at a depth of 4 m. On the other hand, the effective sensed centroid depth of the WERA HF Bragg scatter is 0.9m for the frequency of operation. Furthermore, a WERA and an ADCP have quite different typical accuracy values [2], and the rms difference between hourly WERA and in-situ current measurement is reported to be between 6 and 20 cm/s for different regions [2, 15]. Since, as mentioned in the introduction, the WERA QC procedure is aimed to identify outliers in the data, it is suspected that the inherent difference in measurements between ADCP and WERA may render the threshold of 10 cm/s to be inadequate. It was observed that if the threshold is relaxed to a value of 17 cm/s the percentage of confirmed outlier identification increases to 77%, but at an expense of increasing the WERA false outlier identification to 64%.

Fig. 7 shows a time series example of the WERA data compared to the data measured by the ADCP (labeled as C10). A difference between both signals is constantly observed, reaching 15 cm/s at some points.

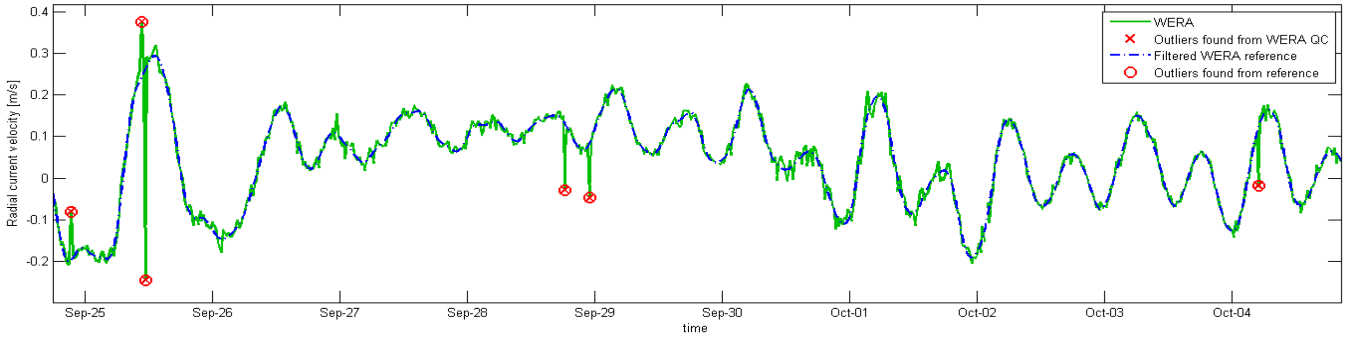

Fig. 6. Identification of outliers and comparison with WERA flagged outliers. Data from Fort DeSoto radar on grid cell 75 km away from shore.

TABLE III. RESULTS OF EVALUATION OF WERA QC USING FILTERED MEASURED DATA AS REFERENCE

	Fort DeSoto WERA system				**Venice WERA system**			
Distance from shore	***30 km***	***50 km***	***75 km***	***100 km***	***30 km***	***50 km***	***75 km***	***100 km***
Available measurement points	18525	18444	17539	14997	19155	18564	16234	11091
Measurement points left out of analysis	25	25	58	238	0	0	107	445
Confirmed outlier identification (QC confirmed/total from ref.)[a]	85.71% (24/28)	71.43% (15/21)	70.73% (29/41)	83.87% (104/124)	100% (5/5)	72.22% (13/18)	76.92% (30/39)	62.5% (45/72)

[a] (Outliers found by both WERA QC and from reference / Total outliers found from reference)

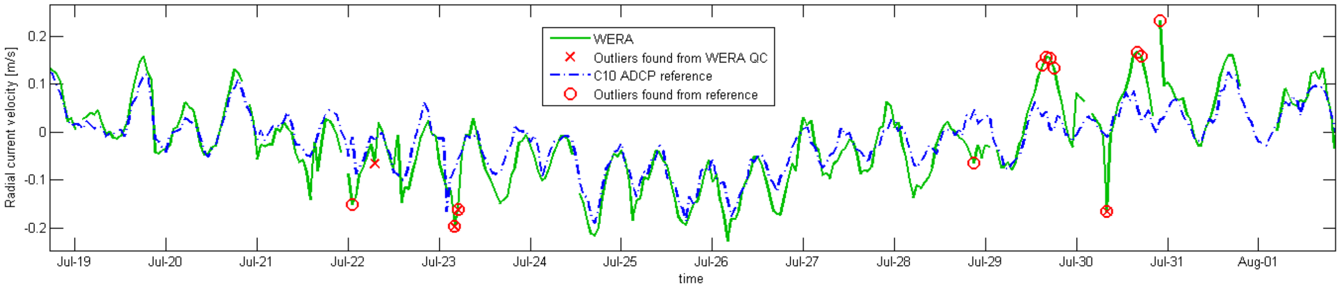

Fig. 7. Identification of outliers using C10 ADCP data as reference and comparison with WERA flagged outliers. Data from Venice radar on grid cell co-located with the ADCP.

TABLE IV. RESULTS OF EVALUATION OF WERA QC USING ADCP AS REFERENCE

ADCP available measurement points	2470
Points where WERA data is not available	154
Confirmed outlier identification (QC confirmed/total from ref.)[a]	20.99% (17/81)

[a] (Outliers found by both WERA QC and from reference / Total outliers found from reference)

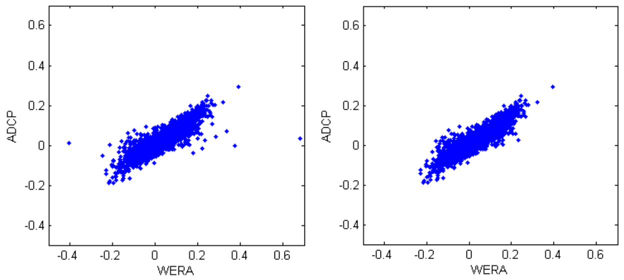

Fig. 8. WERA vs ADCP measurements plotted using all data (left) and without confirmed outliers (right). An obvious removal of outliers is observed.

If the WERA and ADCP data are compared, a correlation coefficient of 0.87 is obtained. However, after removing the outliers identified in real-time the correlation increases to 0.90. This indicates an overall performance improvement utilizing the WERA QC procedure of outlier removal. When comparing the scatter plots (see Fig. 8), the removal of outliers in the data becomes apparent.

VII. CONCLUSIONS AND FUTURE WORK

The QC procedure described in this paper appears to be a very helpful tool in removing the majority of the outliers which sporadically occur in the data measured by the WERA system. An advantage of this method is that it is employed in near real-time and is readily available to be installed on all WERA systems.

In general, the WERA QC procedure has to be tuned to the specific local conditions of the system for optimal results. It is always a question as to the optimal setting of the QC procedure parameters in order to obtain the correct trade-off between confirmed outlier identification and false outlier identification.

Current efforts are being focused in applying similar methods to identify outliers in wave measured data and wind direction data (additional data provided in WERA systems). Also, tools are being developed to provide data with different levels of quality. This way, the QC procedure parameters may be set to be very relaxed in order that a preliminary cleaning of data with low false outlier identification rates (but also low confirmed outlier identification rates) is available. At the same time, reliable data with only the best quality levels (quality level 1, therefore high confirmed outlier identification rates) would also be available for the most demanding applications (e.g. data assimilation).

Future work may include performing additional tests based on spatial analysis of the measurement, which could diminish the dependency on reliable past data. It is also of interest to implement the possibility of tuning different QC parameters for each radar site separately or for special locations in the coverage area. This would allow furthering optimizing the performance of the QC procedure.

ACKNOWLEDGMENT

USF COMPS program support was through NOAA via the U.S. IOOS Office. The USF HF-radar network was a contribution by the NOAA funded Southeast Coastal Ocean Observing Regional Association (SECOORA). WERA equipment additions were through USF internal R&D funds. The success of the seagoing operations is largely due to the technical expertise of J. Law and J. Donovan, and D. Mayer assisted with the ADCP data analysis. This is CPR Contribution # 33.

REFERENCES

[1] Gurgel, et al., "Wellen Radar(WERA), a new ground-wave based HF radar for ocean remote sensing", Coastal Engineering Vol. 37, NOS. 3-4, ISSN 0378-3839, pp. 219...234, August 1999.

[2] Liu,Y.,Weisberg, R.H., Merz, C.R., "Assessment of CODAR SeaSonde and WERA HF radars in mapping surface currents on the West Florida Shelf", Journal of Atmospheric and Oceanic Technology (resubmitted with minor revision), 2014.

[3] Cochin, Thomas, Mariette, and Gurgel. "SURLITOP experiment in West Brittany (France): Results and validation". 6th intern. Radiowave Oceanography Workshop (ROW-6), Hamburg, Germany, May 2006.

[4] Gurgel, Barbin, Schlick. "Radio Frequency Interference Suppression Techniques in FMCW Modulated HF Radars", Proc. of IEEE/OES Oceans '07 Europe, Aberdeen, Scotland, UK, June 2007.

[5] Howard. "U.S. IOOS Program Office Quality Assurance of Real-Time Ocean Data Project", Oceans 2012, Virginia, U.S.A., October 2012.

[6] Heron, Prytz. "The Data Archive for the Phased Array HF Radars in the Australian Coastal Ocean Radar Network" Oceans 2011 IEEE, Santander, Spain, June 2011.

[7] Gurgel, Schlick, Voulgaris, Seeman, Ziemer. "HF Radar Observations in the German Bight: Measurements and Quality Control". Currents, Waves and Turbulence Measurements (CWTM), 2011.

[8] Helzel, Kniephoff, "Software Beam Forming for Ocean Radar WERA Features and Accuracy", IEEE Oceans Conference Proceedings, Seattle, 2010.

[9] Barth, et al., "Ensemble perturbation smoother for optimizing tidal boundary conditions by assimilation of High-Frequency radar surface currents – application to the German Bight" Ocean Sci., 6, 161-178, doi:10.5194/os-6-161-2010, 2010.

[10] Merz, C.R. "An overview of the coastal ocean monitoring and prediction system (COMPS)". In MTS/IEEE Oceans 2001, MTS 0-933957-28-9, 2001.

[11] Weisberg, R.H., Barth, A., Alvera-Azcárate, A., Zheng, L. "A coordinated coastal ocean observing and modeling system for the West Florida Shelf", Harmful Algae, 8, 585-598, 2009.

[12] Weisberg, R.H., Liu, Y., Mayer, D.A. "West Florida Shelf mean circulation observed with long-term moorings", Geophysical Research Letters, 36, L19610, doi:10.1029/2009GL040028, 2009.

[13] Merz, C.R., Weisberg, R.H., Liu, Y., "Evolution of the USF/CMS CODAR and WERA HF Radar Network", IEEE/Oceans 2012, 1-5, doi:10.1109/OCEANS.2012.6404947, 2012.

[14] Liu, Y., Weisberg, R.H., Shay, L.K., "Current patterns on the West Florida Shelf from joint Self-Organizing Map analyses of HF radar and ADCP data", Journal of Atmospheric and Oceanic Technology, 24(4), 702–712, 2007.

[15] Essen, Gurgel, Schlick. "On the accuracy of current measurements by means of HF radar". IEEE Journal of Oceanic Engineering, Vol. 25, NO. 4, October 2000.