

Gap Filling of the Coastal Ocean Surface Currents from HFR Data: Application to the Mid-Atlantic Bight HFR Network

ERICK FREDJ

Department of Computer Sciences, Jerusalem College of Technology, Jerusalem, Israel, and Department of Marine and Coastal Sciences, Rutgers, The State University of New Jersey, New Brunswick, New Jersey

HUGH ROARTY, JOSH KOHUT, MICHAEL SMITH, AND SCOTT GLENN

Department of Marine and Coastal Sciences, Rutgers, The State University of New Jersey, New Brunswick, New Jersey

(Manuscript received 25 March 2015, in final form 1 December 2015)

ABSTRACT

High-frequency radar (HFR) surface current data are an increasingly utilized tool for capturing complex dynamics of coastal ocean systems worldwide. The radar is uniquely capable of sampling relevant temporal and spatial scales of nearshore processes that impact event response activities and basic coastal ocean research. HFR is a shore-based remote sensing system and is therefore subject to data gaps, which are predominately due to environmental effects, like increased external noise or low signal due to ocean surface conditions. Many applications of these surface current data require that these gaps be filled, such as Lagrangian numerical models, to estimate material transport and dispersion. This study introduces a new penalized least squares regression method based on a three-dimensional discrete cosine transform method to reconstruct hourly HFR surface current data with a horizontal resolution of 6 km. The method explicitly uses both time and space variability to predict the missing value. Furthermore, the method is fast, robust, and requires relatively low computer memory storage. This paper evaluates the method against two scenarios of common data gaps found in HFR networks currently deployed around the world. The validation is based on observed surface current maps along the mid-Atlantic coast of the United States with specific vectors removed to replicate these common gap scenarios. The evaluation shows that the new method is robust and particularly well suited to fill a more common scenario with complete data coverage surrounding an isolated data gap. It is shown that the real-time application of the method is suitable for filling data gaps in large oceanography datasets with high accuracy.

1. Introduction

The coastal ocean is an intricate system that forms the boundary between the land and the deep ocean. This environment consists of tightly linked chemical and biological processes that coexist in a causal relationship with complicated flow dynamics. As the water depth decreases, physical forcing shifts from density gradients to turbulent mixing and frictional forcing along the

surface, bottom, offshore, and inshore boundaries (Robinson and Glenn 1999). In addition, tidal oscillations interacting with low-frequency features along the offshore boundary contribute to the complexity of the shelf dynamics that govern the exchange between the coast and the deep ocean (Magnell et al. 1980). Wind forcing is a large component in coastal ocean flow and can quickly change the dynamics, resulting in the generation of large wave disturbances greater than or of the same magnitude as the underlying low-frequency current. High-frequency radars (HFRs) are commonly used to observe and classify these complicated processes through hourly two-dimensional maps of surface currents.

HFR systems are one technology deployed along the coast to remotely measure the complex surface current dynamics over these highly variable seas. In the

Denotes Open Access content.

Corresponding author address: Erick Fredj, Department of Computer Sciences, Jerusalem College of Technology, 21 Havaad Haleumi Street, P.O. Box 16031, 91160 Jerusalem, Israel.
E-mail: erick.fredj@gmail.com

Mid-Atlantic Bight (MAB), a network of over 40 land-based radar sites provides hourly maps of surface ocean currents in support of oceanographic research and applications ranging from developing offshore wind energy (Seroka et al. 2013), pollution and storm response, and U.S. Coast Guard search and rescue (Roarty et al. 2010). These radars can reliably measure currents from a few kilometers off the coast out to 200 km offshore through a large range of weather and ocean conditions (Fig. 1). The shore-based antenna approach provides continuous temporal and broad spatial surface current observations, enabling the delivery of data in real time. Nearly every application of ocean monitoring requires, to some extent, measurements of surface current velocity maps.

While the coastal deployment of these networks provides some great advantages in setup, maintenance, cost, and access, the remote sensed nature of the measurement leads to sporadic gaps in data coverage in both time and space. Each coastal site within an HFR network uses a radio signal backscattered off the ocean surface to estimate the velocity component in the direction of the antenna. Data from overlapping sites are then geometrically combined to provide a two-dimensional surface current map over time. Throughout the community two primary algorithms are used to combine individual site radial component maps into total vector current maps, unweighted least squares (UWLS; Lipa and Barrick 1983), and optimal interpolation (OI; Kim et al. 2007, 2008). Gaps in the final surface current map are therefore dependent on the coverage of each remote site that feeds the combined product. Many research products and applications require that these data gaps be filled. For example, to predict the material transport, the standard approach is to run a Lagrangian numerical model. Lagrangian applications provide an understanding of transport in complex surface current fields (Peacock and Haller 2013). Traditionally, Lagrangian applications track the trajectories of individual particles determined by time-evolving spatial current fields. Assuming that the velocity field is observed for times t over a finite interval $[t_1, t_2]$, the existence of missing values in HFR observation poses a major obstacle.

Several techniques have been used to fill the gaps in either the UWLS or OI derived total vector maps. These are implemented using covariance derived from normal mode analysis (Lipphardt et al. 2000), open-boundary modal analysis (OMA) (Kaplan and Lekien 2007), and empirical orthogonal function (EOF) analysis (Beckers and Rixen 2003; Alvera-Azcárate et al. 2005); and using idealized or smoothed observed covariance (Davis 1985). A comparison of these methods was given by Yaremchuk and Sentchev (2009), who proposed to

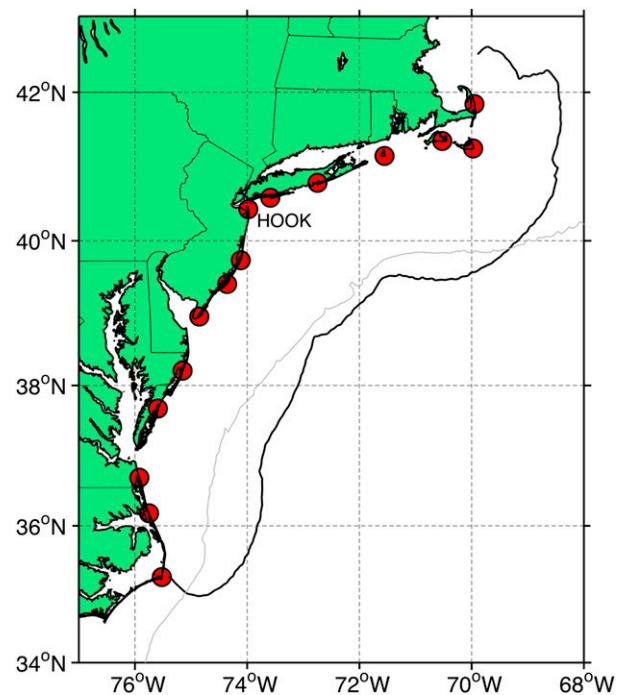


FIG. 1. Map showing the location of the HF radar stations used to construct the MARACOOS surface current maps. The 70% data coverage contour for 2012 (black) marks the best coverage domain that is utilized by the DCT-PLS algorithm to fill data gaps, and the 100-m isobath (gray) are also shown.

add a cost function with the terms penalizing grid-scale variability in the divergence and vorticity fields. However, the mapping methods mentioned above are statistical techniques; therefore, their performance depends on the accuracy of the covariance used for interpolating the HFR data both in space and time. Moreover, present mapping techniques often do not make full use of the dynamical information from the observations.

The goal of the present study is to design an HFR interpolation algorithm capable of filling data gaps in near-real time over the regional scales of a coastal network. To do that we apply a penalized least squares (PLS) regression as a real-time solution to fill gaps in the total vector surface current estimates from an HFR network as a postprocessing step on the derived total vector fields from either the UWLS or OI approach. PLS regression is based on a three-dimensional discrete cosine transform (DCT) (Garcia 2010). The method has been successfully applied to a global soil moisture product derived from Earth observation satellites (Wang et al. 2012). This method is introduced specifically to fill gaps as a required step in many postprocessing real-time applications, including particle trajectories, search and rescue, and spill tracking.

In practice, the occurrence of small data gaps due to environmental factors are more frequent than the larger dropouts due to significant hardware failure or power and communication disruptions at individual radar stations. The highly nonrandom occurrence of missing values in HFR observations challenge their interpretation, since the possible causes include—but are not limited to—geometry of antenna setup, sea state, radio frequency interference, and instrumentation failure. This paper introduces the DCT-PLS technique to HFR gap filling and evaluates it against common gap scenarios observed in regional HFR networks. The paper is organized as follows. In the next section we describe the method and the HFR network used in the evaluation. Section 3 describes the gap-filling results and evaluation. We then discuss these results and implications for application of the method across similar regional networks deployed around the world in section 4.

2. Methods

a. DCT-PLS gap-filling method applied to HFR data

According to Ohlmann (2007), a typical 2D instantaneous HFR velocity field can be expressed from HF radar total vectors as

$$\mathbf{V}(t, x) = \mathbf{V}_{\text{total}}(t, x) + \mathbf{V}'(t, x),$$

where $\mathbf{V}_{\text{total}}(t, x)$ is the total HF radar velocity, which is an average over time (t) and space (x); and $\mathbf{V}'(t, x)$ is largely a nondeterministic subgrid-scale velocity component that is not necessarily uniform in space and time.

In the study here, we introduce for the first time a DCT-PLS method applied to HFR data processing. The DCT-PLS method was originally proposed by Garcia (2010, 2011), and we adapt it here for the purpose of filling data gaps of HFR data for real-time and post-processing. We now give an introduction of the DCT-PLS algorithm. For more details on the mathematics of the method, the reader is referred to Garcia (2010).

1) AUTOMATIC SMOOTHING WITH THE DCT-PLS METHOD

The proposed method based on the penalized least squares approach, combined with the DCT, allows for automatic smoothing of multidimensional data that may include outlying and missing data. Let us define the 2D HFR $m \times n$ velocity field $\mathbf{V} = (u, v)$ surface current (shown in the equation above), where u is the zonal (east/west) component and v is the meridional (north/south) component. First, we assume that the HFR data are corrupted by noise only (no outliers, no missing

data). Following Garcia (2010) the smoothed velocity field $\hat{\mathbf{V}}$ can be expressed as

$$\hat{\mathbf{V}} = \text{IDCT2}[\Gamma \circ \text{DCT2}(\mathbf{V})],$$

where DCT2 and IDCT2 denote the type II 2D DCT and its inverse (IDCT), respectively; and \circ stands for the element-wise product. The filtering matrix Γ is defined by

$$\Gamma_{kl} = \left\{ 1 + s \left[4 - 2 \cos \frac{(k-1)\pi}{m} - 2 \cos \frac{(l-1)\pi}{m} \right] \right\}^{-1},$$

where the subscript (k, l) refers to the position in the 2D HFR current field and the parameter s is a positive scalar that controls the degree of smoothing. An unsuitable selection for s causes under- or oversmoothed velocities; as s increases, the smoothness of $\hat{\mathbf{V}}$ also increases. For small values of s , the value of $\hat{\mathbf{V}}$ will be dominated by noise. The value of the parameter s is determined by minimizing the generalized cross validation (GCV) score method introduced by (Craven and Wahba 1978). The GCV criterion makes the DCT-PLS method fully automated.

2) EFFECT OF THE SMOOTHNESS PARAMETER

Our goal is to find the best estimated $\hat{y}(t)$ from the observed value $y(t)$ for the $t = t_1, t_2, \dots, t_n$ predictor, where the index n refers to the number of predictors.

The technique is to minimize F to balance the infidelity of the data measured by the residual sum of square $\text{RSS} = \|y - \hat{y}\|^2$ and a penalty functional $P(s)$ evaluated by a square second-order difference derivative $s\|D\hat{y}\|^2$. The procedure is known as smoothing splines (see Wahba 1990):

$$F(s) = \text{RSS} + P = \|y - \hat{y}\|^2 + s\|D\hat{y}\|^2.$$

The parameter s specifies the “constant” number $s \times n$ of neighboring observation points used to calculate the local fits. Thus, the bandwidth s determines the degree of smoothness of the fitted data: Choosing too small an s value leads to undersmoothing, whereas selecting an s too large may result in oversmoothing of the fitted data that ignores local features of the data, as shown in Fig. 2. In the frequency domain, increasing the smoothness parameter s reduces the low-pass-filter bandwidth. Extremely large values of s , which cause the loss of high-frequency components, may happen with turbulence or high shear flow.

The tuning parameter s controls the amount of regularization, so choosing a good value of the tuning parameter is crucial. We can question what might be considered a good choice of a tuning parameter.

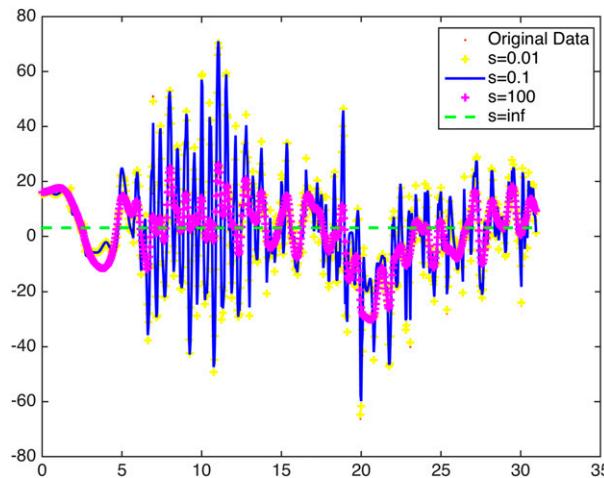


FIG. 2. Smoothness vs original HFR data from 1 Jan 2012 for given s values 0.01, 0.1, 100, and ∞ . Increasing the smoothness control parameter removes the high frequency. The mean HFR velocity is obtained for $s \rightarrow \infty$.

A common solution to select the optimal value of s is to use the cross-validation (CV) procedure. The classical concept of CV consists of splitting the dataset into a *train set* and a *test set* $\{t(i), y_i, i = 1, \dots, n\}$. We fit the model on the train set and test its predictive performance on the test set. By testing the model on a different dataset than the one used for training, we avoid overfitting.

There are many ways to split the initial set dataset into parts like this. One possibility is to remove one sample to form the train set and to put this one sample into the test set. This is called leave-one-out (LOO) cross validation. With N samples, we obtain N sets of train and test sets. The cross validated is the average performance on all these set decompositions.

The expected prediction error rate is defined as $E\{F[y_0, \hat{y}(t_0; s, S_{-1})]\}$, where the CV score has to be taken with respect to new data (t_0, y_0) from the same source and all possible subsets minus one S_{-1} . Those prediction estimators are also linear in the observation $\hat{y}(t_0; s, S_{-1}) = H(s)y(t_0)$, where $H(s)$ provides a measure of leverage. The average leverage is by definition in the range of $[0, 1]$; weak smoothing occurs if the average leverage is close to 1, while oversmoothing appears when the average leverage is 0. A naive approach of the problem selection of s is to select s equal to the value s_0 , thus minimizing the CV score:

$$\min(E\{F[y_0, \hat{y}(t_0; s, S_{-1})]\}) = \min\left\{\frac{1}{n} \sum_i \frac{[y_i - \hat{y}(t_i; s, S_{-1})]^2}{[1 - H_i(s)]^2}\right\}.$$

We note that for each positive $s \geq 0$, there exists a *unique* $P(s)$ that optimizes the normalized mean-square error

(NRMSE), so that the minimum of $F(s)$ is also the solution to the problem known as fair optimizer.

Craven and Wahba (1978) derived an alternative CV criterion. The idea is to replace the weights factor by their average value. This leads to a GCV criterion, which is quickly calculated using

$$\begin{aligned} & \min(E\{F[y_0, \hat{y}(t_0; s, S_{-1})]\}) \\ &= \min\left(\frac{\sum_{i=1}^n [y_i - \hat{y}(t_i; s, S_{-1})]^2/n}{\{1 - \text{Tr}[H(s)]/n\}^2}\right). \end{aligned}$$

Furthermore, Garcia (2010) reduced the computational complexity and increased the speed of the GCV scores by evenly spacing the data.

Thus, \hat{y} can be expressed by

$$\hat{y} = \text{IDCT}[\Gamma \times \text{DCT}(y)],$$

where Γ is given by $\Gamma_{i,i} = (1 + s\{2 - 2 \cos[(i-1)\pi/n]\}^2)$ and $\Gamma_{i,j} = 0$ if $i \neq j$.

However, the smoothing with the minimization of the GCV score has no clear relation to the smoothing parameter and the gap-filling result in time or space. If the variance of the magnitude of the HFR data is great, then an oversmoothing might occur even with an extremely small smoothing parameter (see Fig. 4). Similarly, in Fig. 2, when a smaller smoothing pattern (10^{-2}) is used, there is no relation to the gap filling. Both figures demonstrate that there is no correlation between the smoothing parameter and the actual smoothing achieved.

3) REPLACEMENT OF THE OUTLYING DATA WITH THE DCT-PLS METHOD

The remote sensed nature of the HFR data can lead to spurious vectors. These outliers are commonly the result of a low signal-to-noise ratio due to either a weak return signal often near the outer edges of the coverage or heightened noise due to external sources. Neglecting these outlying HFR current vectors can affect the smoothed HFR field. To overcome the outlying data, the DCT-PLS method uses a robust procedure that is almost not influenced by the outliers. Garcia (2010) expressed the robust procedure of the DCT-PLS method as

$$\hat{\mathbf{V}} = \text{IDCT2}\{\Gamma \circ \text{DCT2}[\mathbf{W}^{\text{bs}} \circ (\mathbf{V} - \hat{\mathbf{V}}) + \hat{\mathbf{V}}]\}.$$

The method uses the initial current velocity ($\hat{\mathbf{V}} = \mathbf{V}$) to generate successive weights coefficients from bisquare weights (\mathbf{W}^{bs}) of the residual current ($\mathbf{V} - \hat{\mathbf{V}}$) until $\hat{\mathbf{V}}$ matches reasonably well $\mathbf{V}_{\text{total}}$. We note that the “robust” procedure discussed here avoids the

weakness related to the linear model of the least square techniques.

4) DEALING WITH MISSING VALUES AND MASKS

To utilize the algorithm, we first have to define the best coverage domain. This was defined as the grid points within the 70% data coverage from Fig. 1, which was taken for the year 2012. Seventy percent was chosen, as this was the highest contour that covered the entire domain of the MAB. This best coverage is a compromise of the data. Any data gaps within the best coverage domain will be filled; however, if the real-time measurement extends beyond the best coverage domain, then the algorithm will remove it. So, it is critical that an appropriate domain be chosen as input to the algorithm. In the presence of missing data, the corresponding weights coefficient is zero, while an arbitrary value is assigned to the velocity. The DCT-PLS method easily deals with missing and masked data by introducing a second weight matrix \mathbf{W}^m , defined by zero if the velocity located in (k, l) is masked or missing and 1 otherwise,

$$\hat{\mathbf{V}} = \text{IDCT2}\{\Gamma \circ \text{DCT2}[\mathbf{W}^m \circ \mathbf{W}^{\text{bs}} \circ (\mathbf{V} - \hat{\mathbf{V}}) + \hat{\mathbf{V}}]\}.$$

This is the equation of the robust DCT-PLS method used for HFR data.

b. HFR

HFR systems deployed along the coast use transmitted radio signals (3–30 MHz) scattered off the ocean surface to calculate radial components of the total surface velocity at a given location (Barrick et al. 1977). Peaks in the backscattered signal are the result of an amplification of a reflected wave, at grazing incidence, by surface gravity waves with a wavelength equal to half that of the transmitted signal (Crombie 1955). The frequency of the backscattered signal will be Doppler shifted depending on the velocity of the scattering surface. Using linear wave theory, the phase speed of the surface waves can be separated from the total frequency shift, leaving only that shift due to the surface current component in the direction of the antenna (Barrick et al. 1977). The radar software isolates the strongest sea echo returns from the Bragg scattering and uses that portion of the radar spectra to calculate radial current velocities.

Over a given time period, sites along the coast generate radial maps of these component vectors with resolutions on the order of 1–6 km in range and 5° in azimuth (Barrick and Lipa 1997; Teague et al. 1997). The HF radar sites in the Mid-Atlantic Regional Association Coastal Ocean Observing System (MARACOOS) network are all

TABLE 1. Typical characteristics of long, medium, and standard range HF radar systems.

System type	Radio frequency (MHz)	Range (km)
Long range	4–6	180
Medium range	12–14	80
Standard range	24–26	30

SeaSonde direction-finding systems manufactured by CODAR Ocean Sensors (Barrick 2008; Roarty et al. 2010). The direction-finding radars use a three-element receive antenna mounted on a single post to determine the direction of the incoming signals. Since the antenna can resolve only the component of the current moving toward or away from the site, information from at least two sites must be geometrically combined to generate total surface current maps.

The MARACOOS HF radar network consists of 43 SeaSonde-type radars (Fig. 1), 17 of which are long range, 18 of which are standard range, and 8 of which are medium range. Table 1 provides the typical characteristics of the different types of systems. For the long-range systems utilized in this study, the radar cell is defined by a range resolution (Δs) of 6 km and an azimuthal resolution ($\Delta\Theta$) of 5°. The maximum range is on the order of 200 km, and the effective bearing angles of individual radars are different due to the coastline geometry. The focus of this study will be the broader coverage provided by the lower-frequency long-range network (Fig. 1).

Each site collects hourly measurements of the radial component of the surface current and wave conditions within a footprint local to the antenna. A suite of CODAR software programs processes the received radar signals to generate the hourly radial current files at each site. Further processing is used to combine the radials from two or more sites to produce total current velocity maps. The existence of a total vector solution depends strongly on the bearing angle diversity of the radial velocities within a search radius at each vector grid point. Since at least two radial velocities from different radar sites are required for a vector solution, the regions with overlapping radar range cells from multiple radar sites have better data coverage through time. The regional radial-to-total processing is accomplished using an OI adaptation developed by Kim et al. (2008) with the MATLAB HFR community toolbox, HFR_Progs (Kohut et al. 2012; Kim et al. 2008). For this method, we used an asymmetric search area stretched parallel to the isobath direction and consistent with the length scales of the currents in the region (Beardsley and Boicourt 1981; Kohut et al. 2004). For quality assurance (QA), we require that both u and v component

uncertainty be less than 60% of the expected variance (Kohut et al. 2012). Each remote site was operated with the quality assurance/quality control (QA/QC) recommendations from the MARACOOS operators and the Radiowave Operators Working Group (ROWG) community (Kohut et al. 2012). These are the same data provided to the national HF radar server at the National Oceanic and Atmospheric Administration (NOAA) National Data Buoy Center (<http://hfradar.ndbc.noaa.gov>). Every hour the available radial velocities are combined into a single total vector map on the national network 6-km grid (Terrill et al. 2006). A total vector was generated only if at least three radial velocities from at least two remote sites were available to the combination algorithm.

c. The Mid-Atlantic Bight study site

For our study we used the MAB as a natural laboratory, as it has an extensive coastal HFR network that supports both research and applications that depend on reliable surface current data delivery. The seasonal forcing cycles drive significant variability in the physical environment of the MAB. Water masses originating from the watershed, deep ocean, and northern latitudes collide in the waters off New Jersey. Ocean fronts, relatively narrow zones that separate these different water types, are important both because of the role they play in ocean dynamics and because they mark some water mass boundaries. Their dynamical importance in the coastal ocean stems from their association with strong currents, such as the equatorward jet observed at the shelfbreak front off the east coast of North America (Loder et al. 1998; Ullman and Cornillon 1999), and with the strong vertical velocities that often occur in coastal regions (Barth et al. 2005; Houghton and Visbeck 1998).

From events lasting several hours to days on through interannual and decadal scales, the variability of the currents helps define the structure of the marine ecological system. The physical structures within the MAB are characterized by transport pathways and strong hydrographic and velocity gradients that vary in space and time. On longer scales of seasons to years, circulation patterns drive persistent cross- and along-shelf transport pathways (Kohut et al. 2004; Dzwonkowski et al. 2009; Gong et al. 2010). On shorter scales of days to weeks, upwelling and strong coastal storms can disrupt or enhance these patterns (Kohut et al. 2006; Dzwonkowski 2009).

d. HFR gap scenarios

The gap-filling method was tested for two scenarios commonly observed in HFR-derived surface current

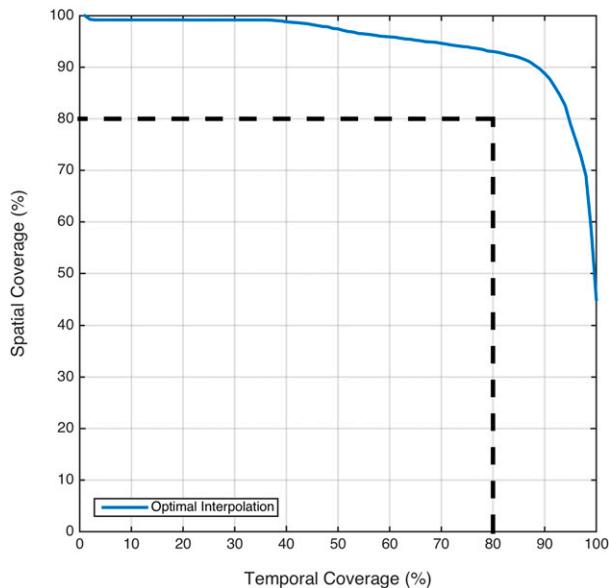


FIG. 3. The ratio of spatial and temporal coverage of the MARACOOS surface current maps for 2012 (blue line). The data delivery target of the network for 80% spatial coverage at least 80% of the time (dashed black line).

maps. Based on a 7-yr dataset (MARACOOS; <http://maracoos.org/>), the hourly coverage of the regional HFR network in the Mid-Atlantic Bight is characterized based on both spatial and temporal coverage. The operational data coverage goal of the network is to provide at least 80% spatial coverage 80% of the time. In this metric, the percentage of spatial coverage is the proportion of grid points within the data footprint beyond the 15-m isobath and within 150 km of the coast with measured data. The measurements within the 15-m isobath are excluded because the deep-water wave assumption that the radar utilizes is no longer valid at our operating frequency of 5 MHz. The points beyond 150 km are excluded, as this is the maximum range of the radar stations during nighttime interference. The temporal coverage can be variable between hours to a time frame of years. This linked spatial temporal metric describes the typical coverage observed across the network over our study period, January–December 2012 (Fig. 3). Figure 3 shows that over much of the year, small spatial gaps of less than 20% of the complete data footprint are more common than larger gaps (>40%) observed during significant hardware or communication disruptions. These smaller data dropouts are isolated areas of the data footprint due to local environmental factors. The larger gaps observed less frequently are due to more significant issues that remove one or more remote sites from the network. In this analysis we define two scenarios that reproduce each of these situations. These

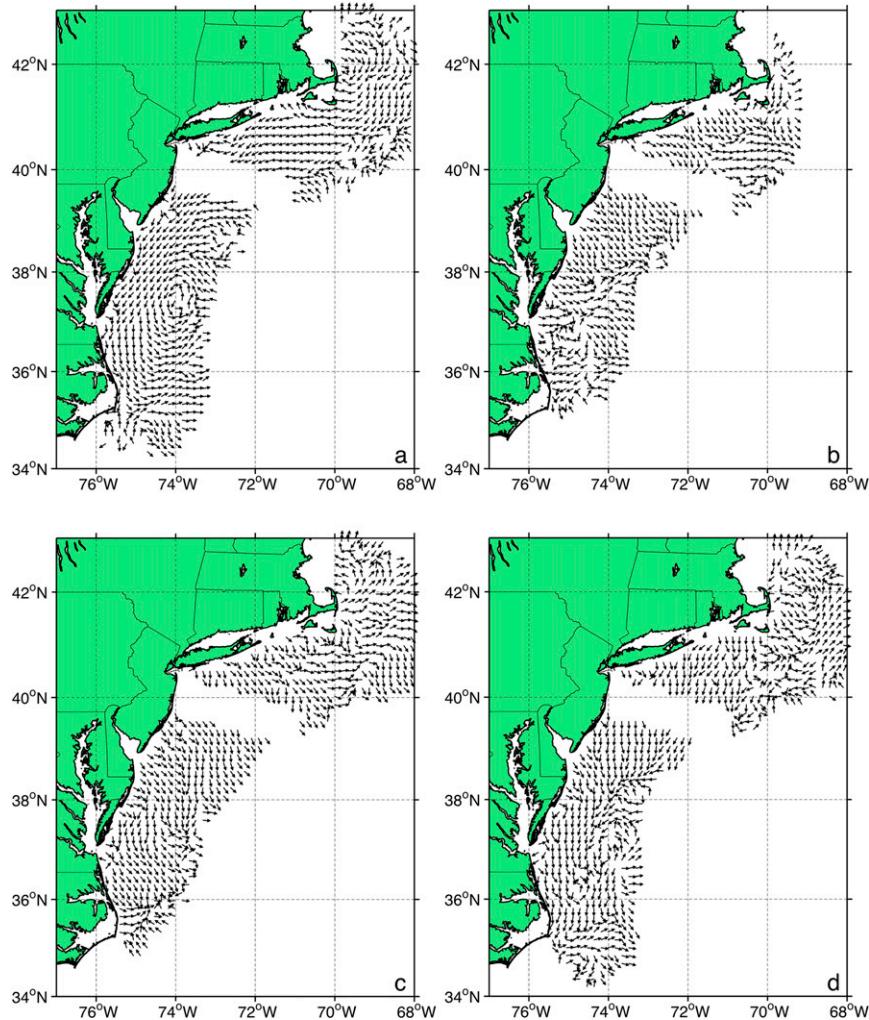


FIG. 4. Surface current maps showing artificial gaps under scenario 1 for the (a) winter, (b) spring, (c) summer, and (d) fall test periods.

more detrimental gaps will typically reduce the coverage by at least 40%.

1) SCENARIO 1

The first scenario tested replicates a major hardware or communication disruption that effectively removes at least one site from the network. Observed gaps under this scenario can be best described as a gap that extends along the coast from the shore to the offshore edge of the coverage, effectively splitting a single data footprint into two. This is very uncommon and is primarily due to a disruption in either the real-time communication link or a hardware failure. The result is a gap that stretches from the coast out to the edge of the coverage (Fig. 4). The size of the band with no data depends on the site spacing and the number of sites that are not reporting data. For the purposes of this

analysis, we are simulating a loss in contributing radials from a single site in Sandy Hook, New Jersey, near the apex of the MAB in the vicinity of the approaches to New York Harbor.

2) SCENARIO 2

The second gap scenario tested replicates more common situations in which each site is contributing radial vectors, but there is a reduction in the number of radial data from one or more of the sites (Fig. 5). These dropouts could be due to a number of environmental factors. The most common cause is an increase in external noise that lowers the signal-to-noise ratio and therefore limits the range a detectable signal can be used to determine radial velocity (Barrick 1971). For the long-range system, this is more common during local nighttime hours, when the ionosphere effects

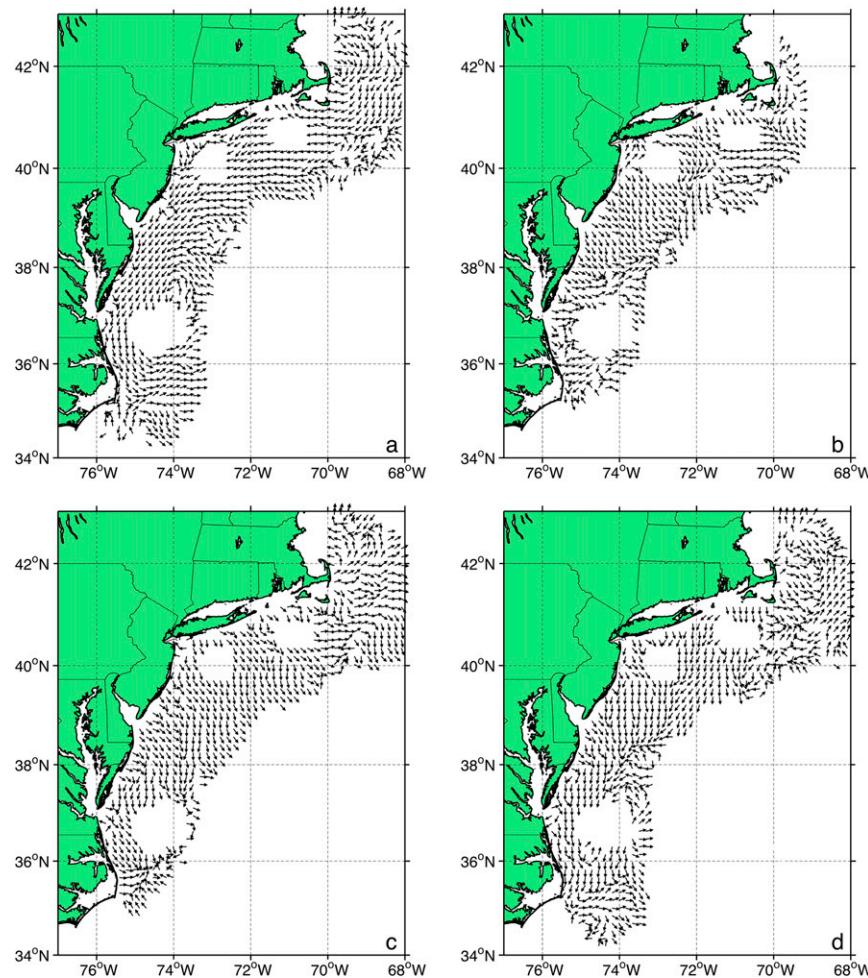


FIG. 5. Surface current maps showing artificial gaps under scenario 2 for the (a) winter, (b) spring, (c) summer, and (d) fall test periods.

increase the range at which a given site receives external noise. Additional environmental factors like local wind and waves could also reduce coverage. These reductions in coverage from sites contributing radials are manifested in the total vector maps as isolated holes in the coverage. The size and location of the gaps depend on the location and magnitude of the reduction of coverage from each individual site. To replicate this in our evaluation, we chose three holes, approximately 30–50 km in diameter, that simulate reduction in coverage from a site in the south, central, and northern regions of the MAB coverage. Based on our analysis of the 7-yr (2007–13) dataset in the MAB coastal radar network, scenario 1 occurs less than 20% of the time with gaps and the smaller, more isolated gaps of scenario 2 represented by any of three gaps shown in Fig. 5 occur 80% of the time with gaps (Fig. 3). This analysis will quantify the accuracy of estimated

vectors from our DCT-PLS method for each of these scenarios.

3. Results

a. Gap-filling results

First, we verify that the new automatic gap-filling method discussed in this paper is appropriate for HFR data gap filling. To do this the DCT-PLS-filled vectors were evaluated over time at grid points in the northern MAB (Fig. 6a). The data coverage during January 2012 and the location of our two analysis points are shown in Fig. 6a. The coverage shows high data returns over the continental shelf with reduced coverage along the edge of the data footprint well offshore near one of the analysis points. The DCT-PLS algorithm was applied to the entire spatial dataset over the month of January to

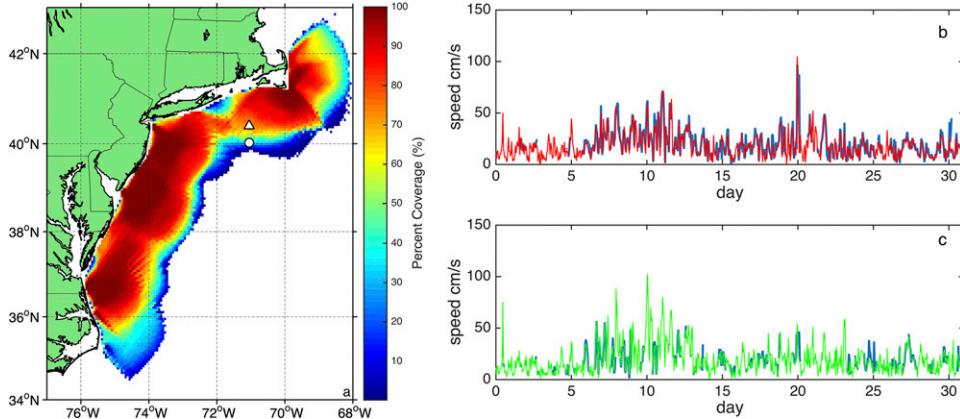


FIG. 6. (a) HFR data coverage over January 2012. The location of our two test sites with 39% (white circle) and 76% (white triangle) are shown. (b) Time series of HFR observations (blue) and the corresponding DCT-PLS model reconstructions (red) for the 76% coverage (white triangle). (c) Time series of HFR observations (blue) and the corresponding DCT-PLS model reconstructions (green) for the 39% coverage.

fill some of these data gaps. The two test sites fall along the same line of longitude and originally possessed 39% and 76% temporal data coverage. We chose these two points to quantify the impact of the gap-filling algorithm over the month. In Figs. 6b and 6c, we show two time series for our selected points in which the algorithm filled the temporal gaps with information from the grid surrounding these locations with higher temporal coverage over the month. The more complete time series of the DCT-PLS-filled values are shown in red and green for the two test sites, respectively. The method does a good job of filling gaps in the time series while retaining the integrity of the data in the surrounding regions without gaps. In a spatiotemporal dataset, the spatially continuous gaps can be temporally intermittent, or vice versa as in shown in Fig. 6. Here the method takes advantage of the spatial and temporal data provided by the HFR to fill gaps in time.

The method was also tested against varied levels of noise in the input data. Specifically, the DCT-PLS method was analyzed on the HFR field with additive Gaussian noise with a variance of $(0.1 \times V_{\max})^2$. The results of this test are shown in Fig. 7. An amount of 0%–50% (using an increment of 5%) of clustered missing data were included within the original HFR data field from 1300:00 UTC 8 January 2012 using random Poisson distribution, and 100 Monte Carlo simulations per configuration were performed.

The performance of the methods is evaluated by using the NRMSE. The NRMSE remained relatively low (<28%) even with 50% of additional missing vectors and was mostly influenced by the additive noise.

Although this case represents an artificial HFR velocity field, it clearly illustrates that the DCT-PLS method can efficiently deal with a large percentage of clustered missing data. In conclusion, these results demonstrate that the DCT-PLS method is highly robust to clustered missing data.

b. Comparison between DCT-PLS and OMA methods

In practice, hardware and environmental factors lead to gaps in HFR-derived surface current maps. In such cases, local interpolations often fail over gap scenarios highlighted in Figs. 4 and 5. As part of our DCT-PLS evaluation, we computed interpolated vectors across the large data gap due to one or two site outages within the MAB network with both the DCT-PLS and OMA methods during autumn (scenario 1; Fig. 4d). We implemented the OMA in a way that could be run across the entire domain in a real-time mode to address potential gaps across the entire domain.

The OMA was performed with the OpenMA toolbox developed by Kaplan and Lekien (2007). The application of OMA to hourly current data is carried out in several steps. First, modes are generated on a specific domain with a continuous boundary. Next, the modes are typically interpolated on the total current grid. The next step is to fit data to the modes. This can be done with either radial current measurements or total currents. After the fits the OMA currents are ready to be used. We applied the OMA method to the MARACOOS domain hourly sampling on a uniform grid with 6 km \times 6 km intervals. The fits were performed using minimum spatial scales of 6 km (all modes) on the total current

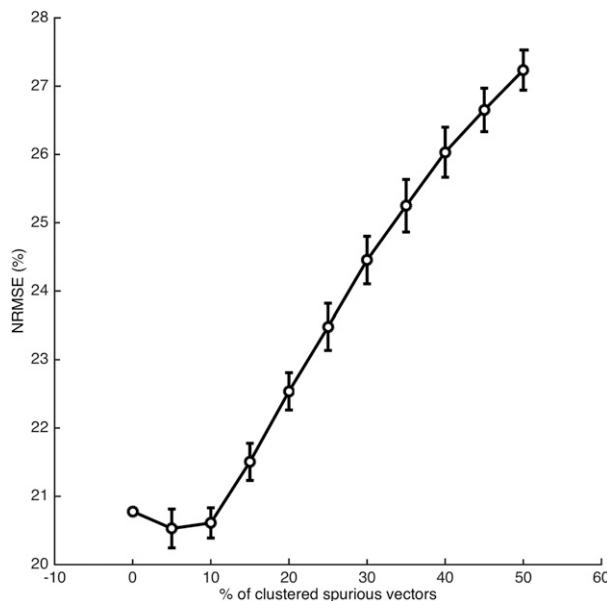


FIG. 7. HFR data postprocessed with the DCT-PLS method: NRMSE (between the postprocessed and original velocity fields) as a function of the percentage of Gaussian noise with a standard deviation of 1% of the maximum velocity.

measurements based on the OpenMA toolbox default value of 200 modes. We acknowledge that the 200 modes fall short of the theoretical \sim 6000 total modes, at least 3000 Dirichlet modes and 3000 Neumann modes, needed to resolve features approaching the grid resolution over our domain. Given the computing constraints and our intention to use the OMA as an alternative to benchmark the DCT-PLS method in a real-time data delivery setting, the available OMA tools will fail to produce this large number of modes. So, we had to reduce the number of modes to the toolbox default of 200. The OMA method has two primary input parameters: the spatial length scale L , which defines the number of modes used for the interpolation; and the diffusion parameter “ κ ,” which penalizes the magnitude of the modes. The parameters used in our application of OMA were $L = 6$ km and $\kappa = 10^{-4}$.

We investigated the reconstruction of the missing data performance of both algorithms on the fall scenario 1 and analyzed the reconstruction of the current patterns within the data gap (Fig. 8). A visual comparison showed that for this scenario, the DCT-PLS method performed as well as and across much of the domain better than the OMA interpolated vectors. The velocity pattern of the DCT-PLS interpolated vectors better replicated the patterns of the removed vectors across much of the gap and were more realistic compared to the OMA velocities. Table 2 presents the RMS error statistics for the vector magnitude and direction

comparison between these two methods and the withheld vectors. We caution the reader that the quality of the OMA interpolation is very dependent on the number of modes selected. Our intention in this paper is to see whether the new DCT-PLS application is comparable to the OMA application that has been more widely applied to HFR gap filling over our entire domain as a real-time tool. This required us to reduce the number of modes to the toolbox default value of 200. Therefore, the OMA-derived fields will not be able to resolve the finer spatial scales. In general the DCT-PLS method had smaller RMS errors in both scenarios across our four seasonal test periods. In the OMA formulation, the number of modes is proportional to $(D/L)^2$ (see Kaplan and Lekien 2007), where D is the horizontal size of the domain and L is the spatial length scale introduced previously. To achieve a better reconstruction of the more spatially complex current fields with OMA, we must increase the number of modes by reducing $L = 2$ –3 km, which will require an increased κ . This optimization of the OMA for our specific region and data gap is beyond the scope of this study. In addition, both the OMA and DCT-PLS methods did not accurately represent the small-scale features of the HFR velocity field, especially in scenario 1.

In conclusion, when a large data gap is present, the DCT-PLS method with RMS differences between 3.5 and 18.9 cm s^{-1} and 14.4 and 204.3 cm s^{-1} for the vector magnitude and phase, respectively, is better than the OMA with RMS differences between 8.6 and 31.2 cm s^{-1} and 19.9° and 191° for the vector magnitude and phase, respectively. These are lower averages on average because of the robust statistical ability of DCT-PLS to estimate the current within the gap. Based on this basic evaluation, the DCT-PLS method is comparable to the OMA method, and in many regions of our test scenario it produces more realistic interpolated vectors. Since the DCT-PLS method does not require any preprocessing, it is also more computationally efficient to run on large HFR networks like that deployed in the MAB. More work is needed to quantify the differences and similarities of these two methods and others in filling a variety of gaps in HFR networks. The details of the comparison between the DCT-PLS method introduced in this manuscript is discussed in more detail in the following section.

c. Synthetic data validation of the DCT-PLS method

The evaluation of the interpolated fields is organized into tests that replicate typical gap scenarios observed in the coastal networks deployed around the world (Lipphardt et al. 2000; Paduan and Rosenfeld 1996). The

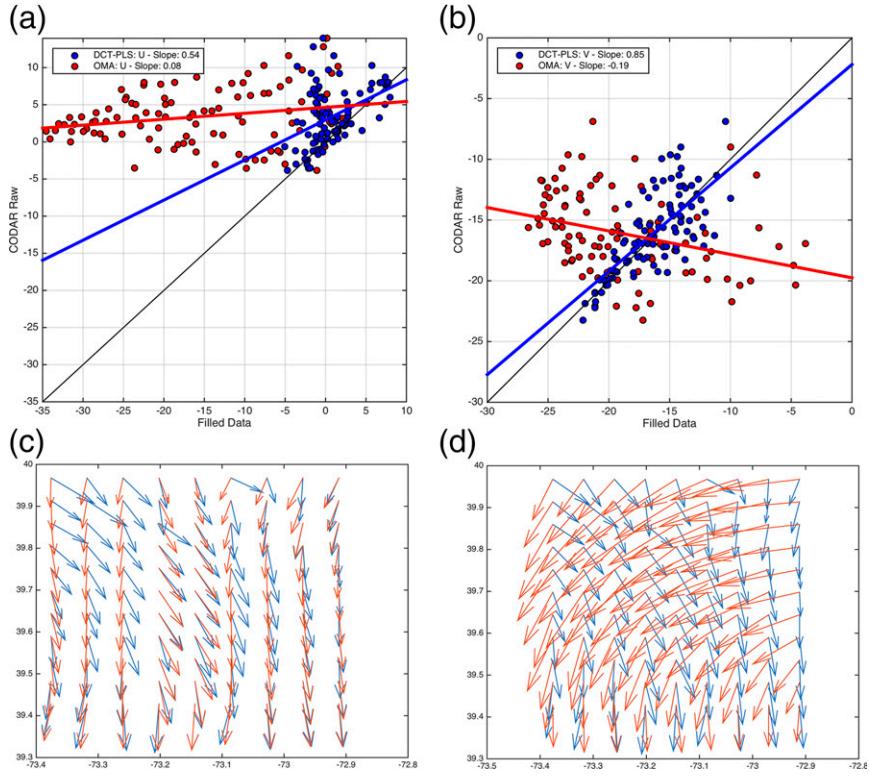


FIG. 8. Scatterplots comparing the estimated velocities with the removed observations for the DCT-PLS (blue) and OMA (red) methods for the (a) east and (b) north velocity components for scenario 1. Vector maps showing the CODAR observations (blue) and the filled values (red) for the (c) DCT-PLS and (d) OMA.

challenge we had in designing the evaluation of the method was to artificially define the gaps so that we could use the withheld data as truth. The size of the gaps in each scenario was chosen based on the analysis described in Fig. 3. Since the gaps represented in our two scenarios do occur in the spatial time series, we could not consistently identify observations to remove and use as truth throughout the entire time series. As an alternative, we identified four maps with complete coverage that represent the range of spatial complexity observed in the maps over our 7-yr time series (Dzwonkowski et al. 2009; Dzwonkowski 2009; Gong et al. 2010). During the windier better mixed months of the fall and winter, the maps tend to be more uniform compared to the shorter decorrelation scales observed during the calmer months of the spring and summer. These hourly current maps sampled in each season provide the consistent ground truth needed for our evaluation and the variability in the flow fields representative of the entire time series.

For scenarios 1 and 2, we evaluated these four velocity fields by comparing the interpolated vectors to those removed within each gap. The comparison between the

removed vectors and the predicted values from our method for each scenario is shown in Fig. 9. The scatter shows a stronger agreement between the predicted currents and the observed under scenario 2 representing the more common occurrence of small isolated data gaps. Under this scenario the method performed well with slopes for all four time periods above 0.7 for both the u and v components. The slopes less than one indicate that, on average, the filled-in values were slightly less than the observed velocities. For the less frequent gap scenario 1, the method does not perform as well with slopes below 0.35 and increased variance. The comparison statistics between the removed and predicted vectors across each of these scenarios are shown in Table 2. For scenario 1, the RMS error between the DCT-PLS predicted and removed vector magnitudes across the four time periods range from 3.4 to 18.9 cm s^{-1} . This variability across the time periods tested is shown in Fig. 10. The four time periods represent a range in the characteristics of the flow surrounding the gap. They were chosen to represent the typical structure observed throughout the year in the MAB (Gong et al. 2010). The lowest correlation in

TABLE 2. RMS error between the DCT-PLS and OMA estimated velocities and the removed observations over each scenario and season.

Scenario 1				
	DCT-PLS magnitude (cm s ⁻¹)	DCT-PLS direction (°)	OMA magnitude (cm s ⁻¹)	OMA direction (°)
Winter	18.9	204.3	31.3	191.0
Spring	5.3	41.4	8.7	94.4
Summer	6.1	14.4	8.6	20.0
Fall	3.5	30.8	12.8	104.9
Scenario 2				
	DCT-PLS magnitude (cm s ⁻¹)	DCT-PLS direction (°)	OMA magnitude (cm s ⁻¹)	OMA direction (°)
Winter	14.4	105.1	34.5	169.8
Spring	5.2	66.7	11.6	91.5
Summer	8.4	28.5	23.1	35.0
Fall	9.1	76.7	12.2	110.8

the winter is characterized by broad scatter with slopes close to zero for both the u and v components (Fig. 10). The highest correlation occurred in the summer with a slope closer to 1, particularly in the north/south component (0.82).

In contrast to the large range of values seen in scenario 1, the correlation of the interpolated vectors in scenario 2 was more consistent. Similarly, the scatterplots all show a more concentrated distribution along a line closer to the target 1:1 line (Fig. 11). The exception was the fall test, when the slopes for both components fell below 0.5. In the winter, the correlation was the highest observed at 0.95 with slopes for both components above 0.7. The relatively high winter RMS differences reported in Table 2 compared to the other seasons tested over scenario 2 are due to the small number of points above the 1:1 line (Fig. 11a). Because of the faster currents in this winter scenario, these points bias the RMS difference statistics high compared to the majority of filled values in this test that fall on the 1:1 line.

4. Discussions and conclusions

In this study we introduced an efficient automated DCT-PLS method for filling data gaps in the HFR ocean spatiotemporal dataset applied to the MARACOOS domain. The procedure explicitly utilizes both spatial and temporal information to derive the statistical model and to predict the missing values.

The evaluation highlights the sensitivity of the gap-filling method to the vectors surrounding the gaps. In our analysis we chose two scenarios to replicate the conditions typically observed in coastal networks operating around the world. The band scenario is a less common occurrence in which either a communication or hardware failure causes a gap in the coverage that stretches from the coast to the outer edge of the coverage. In this scenario we saw a large range in the accuracy of the interpolated vectors. Since this scenario by definition does not have observed vectors surrounding the gap, the quality of the interpolated vectors is dependent

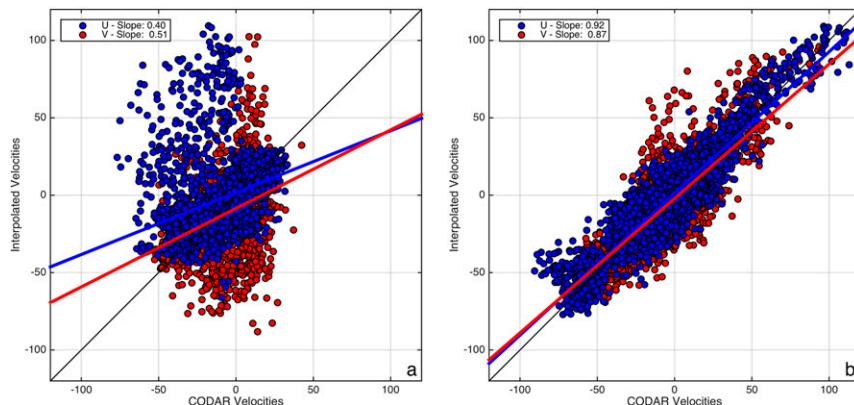


FIG. 9. Scatterplots comparing the estimated velocities with the removed observations for the east (blue) and north (red) components in the gaps under (a) scenario 1 and (b) scenario 2 for all the seasonal tests.

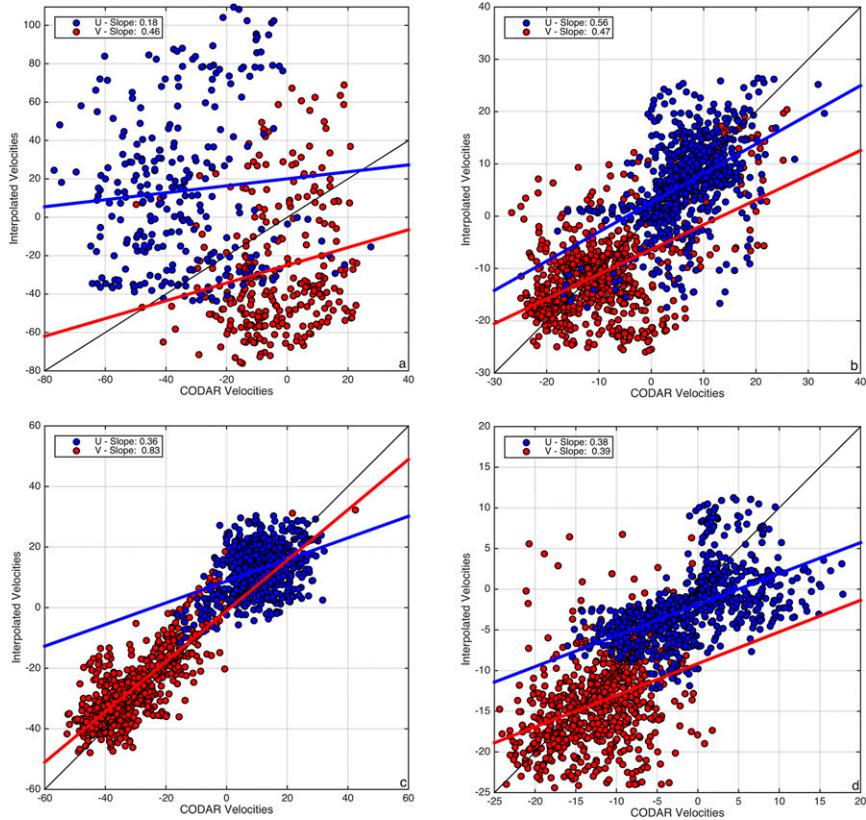


FIG. 10. Scatterplots comparing the estimated velocities with the removed observations for the east (blue) and north (red) components in the gaps under scenario 1 for the (a) winter, (b) spring, (c) summer, and (d) fall test periods. Note that the velocity scales of each panel change as they are optimized for the range of the input data.

on the spatial structure of the flow on either side of the data gap. For those times when the flow was uniform and flowing along the gap, the comparison was quite good with a correlation of 0.7. If the flow was not uniform or flowing mostly across the band, then the lack of vectors nearshore and offshore of the band reduced the quality of the interpolated vectors. This is most evident in the wintertime image with flow around the band moving mostly across the band.

Scenario 2 tested gaps that are much more typical in regional networks. Under this scenario the gaps are smaller and isolated within complete coverage. They occur when environmental conditions reduce the range of individual coastal sites. Under this scenario the comparison on average was much better. Unlike the band scenario, observed currents that informed the interpolation method surrounded these gaps. With information surrounding the gap, the method performed better. The flow characteristics did impact the quality of the interpolated vectors with the highest correlation observed when the flow was largely uniform across the gap. As the complexity of the flow reached scales

equivalent to the size of the gap, the correlation dropped.

The user, however, should be aware of some limitations of the automatic gap-filling procedure. The method was tested as a gap-filling solution to a real-time HFR data stream. Consequently, the GVC criterion was applied for the fully automated smoothing algorithm. Therefore, good results are expected for a Gaussian noise with zero mean and constant variance (scenario 2). [Garcia \(2011\)](#) and [Wahba \(1990\)](#) reported that the GVC criterion is fairly well adapted to non-Gaussian noise and nonhomogeneous variances. Additionally, the GVC criterion may cause problems when the area of missing data size is large with incomplete surrounding data coverage (scenario 1). Under these conditions, the automated application of the method may lead to poorly predicted vectors. In this case, the best smoothing parameter will need to be determined manually based on the specific gap location and size. As a consequence, the efficiency of the automated gap filling depends specifically upon the original data and on the properties of the additive noise, as shown above.

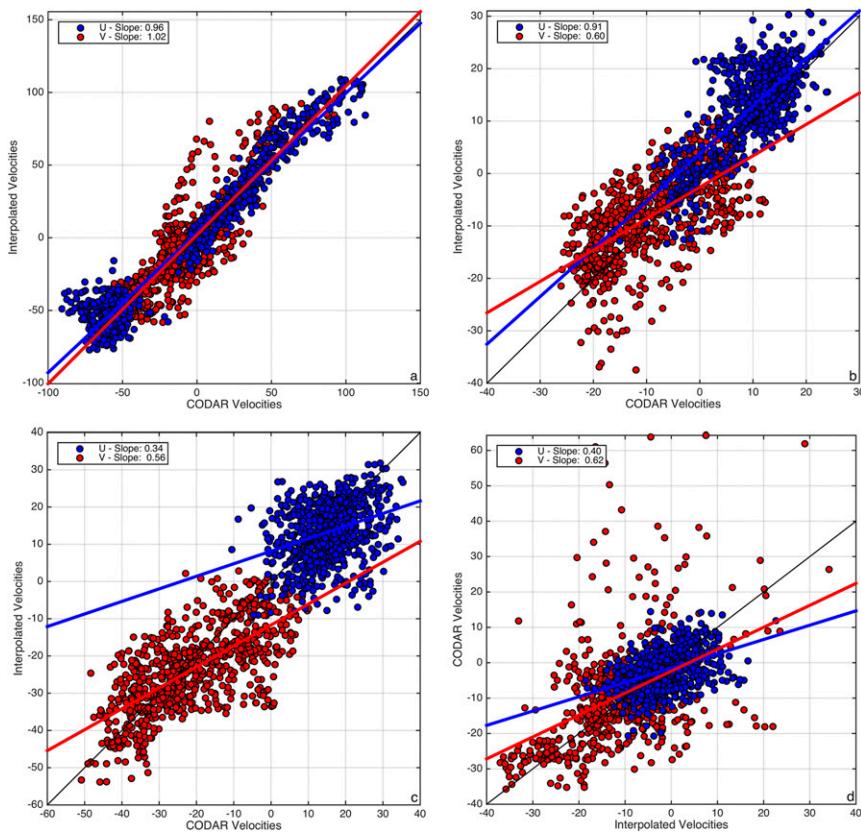


FIG. 11. Scatterplots comparing the estimated velocities with the removed observations for the east (blue) and north (red) components in the gaps under scenario 2 for the (a) winter, (b) spring, (c) summer, and (d) fall test periods. Note that the velocity scales of each panel change as they are optimized for the range of the input data.

We have evaluated the DCT-PLS method for filling gaps inherent in HFR real-time data streams. The method is shown to be a robust solution for the most common gap scenarios characterized as holes, approximately 30–50 km in diameter, in the data coverage with observations completely surrounding the gap. Under the less common scenario in which more significant outages can remove entire sites from a coastal network, the effectiveness of the method depends on the characteristics of the surrounding flow. Individual HFR network operators will need to assess the scales of variability in their operating area to determine the optimal way to apply this method in either a real-time or post-processed application.

Acknowledgments. This study was supported by a grant from the Israel Science Foundation and the Taiwan Ministry of Science and Technology (EF). The HFR dataset was supported through NOAA Award NA11NOS0120038, “Towards a Comprehensive Mid-Atlantic Regional Association Coastal Ocean Observing System (MARACOOS),” and the National Ocean Service (NOS), National Oceanic

and Atmospheric Administration (NOAA) NOAA-NOS-IOOS-2011-2002515 / CFDA: 11.012, “Integrated Ocean Observing System Topic Area 1: Continued Development of Regional Coastal Ocean Observing Systems.” We also acknowledge the time and advice provided by Dr. Bruce Lipphardt and three reviewers, all of whom helped to revise the manuscript into its present form.

REFERENCES

Alvera-Azcárate, A., A. Barth, M. Rixen, and J. M. Beckers, 2005: Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: Application to the Adriatic Sea surface temperature. *Ocean Model.*, **9**, 325–346, doi:[10.1016/j.ocemod.2004.08.001](https://doi.org/10.1016/j.ocemod.2004.08.001).

Barrick, D. E., 1971: Theory of HF and VHF propagation across the rough sea: 1. The effective surface impedance for a slightly rough highly conducting medium at grazing incidence. *Radio Sci.*, **6**, 517–526, doi:[10.1029/RS006i005p00517](https://doi.org/10.1029/RS006i005p00517).

—, 2008: 30 years of CMTC and CODAR. *CMTC 2008: IEEE/OES 9th Working Conference on Current Measurement Technology*, IEEE, 131–136, doi:[10.1109/CCM.2008.4480856](https://doi.org/10.1109/CCM.2008.4480856).

—, and B. J. Lipa, 1997: Evolution of bearing determination in HF current mapping radars. *Oceanography*, **10**, 72–75, doi:[10.5670/oceanog.1997.27](https://doi.org/10.5670/oceanog.1997.27).

—, M. W. Evens, and B. L. Weber, 1977: Ocean surface currents mapped by radar. *Science*, **198**, 138–144, doi:10.1126/science.198.4313.138.

Barth, J. A., S. D. Pierce, and T. J. Cowles, 2005: Mesoscale structure and its seasonal evolution in the northern California Current System. *Deep-Sea Res. II*, **52**, 5–28, doi:10.1016/j.dsr2.2004.09.026.

Beardsley, R. C., and W. C. Boicourt, 1981: On estuarine and continental-shelf circulation in the Middle Atlantic Bight. *Evolution of Physical Oceanography*, B. A. Warren and C. Wunsch, Eds., MIT, 198–233.

Beckers, J. M., and M. Rixen, 2003: EOF calculations and data filling from incomplete oceanographic datasets. *J. Atmos. Oceanic Technol.*, **20**, 1839–1856, doi:10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2.

Craven, P., and G. Wahba, 1978: Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.*, **31**, 377–403, doi:10.1007/BF01404567.

Crombie, D. D., 1955: Doppler spectrum of sea echo at 13.56 Mc./s. *Nature*, **175**, 681–682, doi:10.1038/175681a0.

Davis, R. E., 1985: Objective mapping by least squares fitting. *J. Geophys. Res.*, **90**, 4773–4778, doi:10.1029/JC090iC03p04773.

Dzwonkowski, B., 2009: Surface current analysis of shelf water in the central Mid-Atlantic Bight. Ph.D. thesis, University of Delaware, 178 pp.

—, J. T. Kohut, and X.-H. Yan, 2009: Seasonal differences in wind-driven across-shelf forcing and response relationships in the shelf surface layer of the central Mid-Atlantic Bight. *J. Geophys. Res.*, **114**, C08018, doi:10.1029/2008JC004888.

Garcia, D., 2010: Robust smoothing of gridded data in one and higher dimensions with missing values. *Comput. Stat. Data Anal.*, **54**, 1167–1178, doi:10.1016/j.csda.2009.09.020.

—, 2011: A fast all-in-one method for automated post-processing of PIV data. *Exp. Fluids*, **50**, 1247–1259, doi:10.1007/s00348-010-0985-y.

Gong, D., J. T. Kohut, and S. M. Glenn, 2010: Seasonal climatology of wind-driven circulation on the New Jersey Shelf. *J. Geophys. Res.*, **115**, C04006, doi:10.1029/2009JC005520.

Houghton, R. W., and M. Visbeck, 1998: Upwelling and convergence in the Middle Atlantic Bight Shelfbreak Front. *Geophys. Res. Lett.*, **25**, 2765–2768, doi:10.1029/98GL02105.

Kaplan, D. M., and F. Lekien, 2007: Spatial interpolation and filtering of surface current data based on open-boundary modal analysis. *J. Geophys. Res.*, **112**, C12007, doi:10.1029/2006JC003984.

Kim, S. Y., E. Terrill, and B. Cornuelle, 2007: Objectively mapping HF radar-derived surface current data using measured and idealized data covariance matrices. *J. Geophys. Res.*, **112**, C06021, doi:10.1029/2006JC003756.

—, —, and —, 2008: Mapping surface currents from HF radar radial velocity measurements using optimal interpolation. *J. Geophys. Res.*, **113**, C10023, doi:10.1029/2007JC004244.

Kohut, J. T., S. M. Glenn, and R. J. Chant, 2004: Seasonal current variability on the New Jersey inner shelf. *J. Geophys. Res.*, **109**, C07S07, doi:10.1029/2003JC001963.

—, H. J. Roarty, and S. M. Glenn, 2006: Characterizing observed environmental variability with HF Doppler radar surface current mappers and acoustic Doppler current profilers: Environmental variability in the coastal ocean. *IEEE J. Oceanic Eng.*, **31**, 876–884, doi:10.1109/JOE.2006.886095.

Kohut, J., H. Roarty, E. Randall-Goodwin, S. Glenn, and C. Lichtenwalner, 2012: Evaluation of two algorithms for a network of coastal HF radars in the Mid-Atlantic Bight. *Ocean Dyn.*, **62**, 953–968, doi:10.1007/s10236-012-0533-9.

Lipa, B. J., and D. E. Barrick, 1983: Least-squares methods for the extraction of surface currents from CODAR cross-loop data: Application at ARSLOE. *IEEE J. Oceanic Eng.*, **OE-8**, 226–253, doi:10.1109/JOE.1983.1145578.

Lipphardt, B. L., A. D. Kirwan, C. E. Grosch, J. K. Lewis, and J. D. Paduan Jr., 2000: Blending HF radar and model velocities in Monterey Bay through normal mode analysis. *J. Geophys. Res.*, **105**, 3425–3450, doi:10.1029/1999JC900295.

Loder, J. W., B. Petrie, and G. Gawarkiewicz, 1998: The coastal ocean off northwestern North America: A large-scale view. *The Global Coastal Ocean: Regional Studies and Syntheses*, A. R. Robinson and K. H. Brink, Eds., The Sea—Ideas and Observations on Progress in the Study of the Seas, Vol. 11, John Wiley and Sons, 105–133.

Magnell, B. A., S. L. Spiegel, R. I. Scarlet, and J. B. Andrews, 1980: The relationship of tidal and low-frequency currents on the north slope of Georges Bank. *J. Phys. Oceanogr.*, **10**, 1200–1212, doi:10.1175/1520-0485(1980)010<1200:TROTAL>2.0.CO;2.

Ohlmann, C., P. White, L. Washburn, E. Terrill, B. Emery, and M. Otero, 2007: Interpretation of coastal HF radar-derived surface currents with high-resolution drifter data. *J. Atmos. Oceanic Technol.*, **24**, 666–680, doi:10.1175/JTECH1998.1.

Paduan, J. D., and L. Rosenfeld, 1996: Remotely sensed surface currents in Monterey Bay from shore-based HF radar (Coastal Ocean Dynamics Application Radar). *J. Geophys. Res.*, **101**, 20 669–20 686, doi:10.1029/96JC01663.

Peacock, T., and G. Haller, 2013: Lagrangian coherent structures: The hidden skeleton of fluid flows. *Phys. Today*, **66**, 41–47, doi:10.1063/PT.3.1886.

Roarty, H. J., and Coauthors, 2010: Operation and application of a regional high-frequency radar network in the Mid-Atlantic Bight. *Mar. Technol. Soc. J.*, **44**, 133–145, doi:10.4031/MTSJ.44.65.

Robinson, A. R., and S. M. Glenn, 1999: Adaptive sampling for ocean forecasting. *Naval Res. Rev.*, **51**, 28–38.

Seroka, G., J. Kohut, L. Palamara, S. Glenn, H. Roarty, L. Bowers, and R. Dunk, 2013: Spatial evaluation of high-resolution modeled offshore winds using estimated winds derived from a network of HF radars. *Proc. Oceans—San Diego, 2013*, San Diego, CA, IEEE, 1–5.

Teague, C. C., J. F. Vesecky, and D. M. Fernandez, 1997: HF radar instruments, past to present. *Oceanography*, **10** (2), 40–43, doi:10.5670/oceanog.1997.19.

Terrill, E., and Coauthors, 2006: Data management and real-time distribution in the HF-Radar National Network. *OCEANS 2006*, IEEE, 1–6, doi: 10.1109/OCEANS.2006.306883.

Ullman, D. S., and P. C. Cornillon, 1999: Satellite-derived sea surface temperature fronts on the continental shelf off the northeast U.S. coast. *J. Geophys. Res.*, **104**, 23 459–23 478, doi:10.1029/1999JC900133.

Wahba, G., 1990: Estimating the smoothing parameter. *Spline Models for Observational Data*. Society for Industrial Mathematics, SIAM, 45–65, doi: 10.1137/1.9781611970128.ch4.

Wang, G., D. Garcia, Y. Liu, R. de Jeu, and A. J. Dolman, 2012: A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations. *Environ. Modell. Software*, **30**, 139–142, doi:10.1016/j.envsoft.2011.10.015.

Yaremchuk, M., and A. Sentchev, 2009: Mapping radar-derived sea surface currents with a variational method. *Cont. Shelf Res.*, **29**, 1711–1722, doi:10.1016/j.csr.2009.05.016.