Ocean Dynamics

Improved statistical prediction of surface currents based on historic HF-Radar observations

--Manuscript Draft--

Manuscript Number:	
Full Title:	Improved statistical prediction of surface currents based on historic HF-Radar observations
Article Type:	Special Issue - SAR 2011
Keywords:	Surface current prediction, HF-Radar, search and rescue, Monterey Bay, CA
Corresponding Author:	Sergey Frolov, Ph.D. Monterey Bay Aquarium Research Institute Moss Landing, CA UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Monterey Bay Aquarium Research Institute
Corresponding Author's Secondary Institution:	
First Author:	Sergey Frolov, Ph.D.
First Author Secondary Information:	
All Authors:	Sergey Frolov, Ph.D.
	Jeffrey Paduan, Ph.D.
	Michael Cook
	James Bellingham, Ph.D.
All Authors Secondary Information:	
Abstract:	Accurate short-term prediction of surface currents can improve efficiency of search-and-rescue operations, oil-spill response, and marine operations. We developed a linear statistical model for predicting surface currents (up to 48 hours in the future) based on a short time-history of past HF-radar observations (past 48 hours) and a forecast of surface winds. Our model used empirical orthogonal functions (EOFs) to capture spatial correlations in the HF-radar data and used linear autoregression model to predict the temporal dynamics of the EOF coefficients. We tested the developed statistical model using historical observation of surface currents in Monterey Bay, California. The predicted particle trajectories separated from particles advected with HF-radar data at a rate of 4.4 km/day. The developed model was more accurate than existing statistical model (drifter separation of 5.5 km/day) and circulation model (drifter separation of 8.9 km/day) for the same area. When the wind forecast was not available, the accuracy of our model degraded slightly (drifter separation of 4.9 km/day), but was still better than existing models. We found that the minimal length of the HF-radar data required to train an accurate statistical model was one to two years, depending on the accuracy desired. Our evaluation showed that the developed model is accurate, is easier to implement and maintain than existing statistical and circulation models, and can be relocated to other coastal systems that have a sufficient history of HF-radar observations.

Title: Improved statistical prediction of surface currents based on historic HF-Radar observations

Authors: Sergey Frolov[1*], Jeffrey Paduan[2], Michael Cook[2], James Bellingham[1]
* -- contact author (frolovs@mbari.org)
1 –Monterey Bay Aquarium Research Institute, 7700 Sandholdt Rd., Moss Landing, CA, 95039
2--Naval Postgraduate School, Code OC/Pd, Monterey, CA, 93943

Abstract:

Accurate short-term prediction of surface currents can improve efficiency of search-and-rescue operations, oil-spill response, and marine operations. We developed a linear statistical model for predicting surface currents (up to 48 hours in the future) based on a short time-history of past HF-radar observations (past 48 hours) and a forecast of surface winds. Our model used empirical orthogonal functions (EOFs) to capture spatial correlations in the HF-radar data and used linear autoregression model to predict the temporal dynamics of the EOF coefficients. We tested the developed statistical model using historical observation of surface currents in Monterey Bay, California. The predicted particle trajectories separated from particles advected with HF-radar data at a rate of 4.4 km/day. The developed model was more accurate than existing statistical model (drifter separation of 5.5 km/day) and circulation model (drifter separation of 8.9 km/day) for the same area. When the wind forecast was not available, the accuracy of our model degraded slightly (drifter separation of 4.9 km/day), but was still better than existing models. We found that the minimal length of the HF-radar data required to train an accurate statistical model was one to two years, depending on the accuracy desired. Our evaluation showed that the developed model is accurate, is easier to implement and

maintain than existing statistical and circulation models, and can be relocated to other coastal systems that have a sufficient history of HF-radar observations.

Keywords: Surface current prediction, HF-Radar, search and rescue, Monterey Bay, CA

1 Introduction

Knowledge of surface currents is essential in search and rescues operations, oil spill response, and marine operations. Several observing and modeling systems provide such capability. These include: measurements of surface currents using High Frequency (HF) coastal radars (Barrick et al. 1977), prediction of surface currents using primitive equation models (Breivick and Sætra 2001; Shulman and Paduan 2009), current prediction based on tidal harmonics (Egbert and Erofeeva 2010), and current prediction based on geostrophic balance between wind stress and the Coriolis force (Lagerloef et al. 1999). Of these systems, HF-Radar is the only system that can directly measure surface currents over a large portion of the coastal ocean (up to 200 km offshore).

HF-Radar estimates radial surface current velocities by measuring Doppler shift in the gravity waves that move towards or away from the radar antenna. Using an array of antennas with overlapping lines of sights these radial velocities are combined into a field of vector current. The measured surface velocity is an average velocity over the surface layer of the ocean that varies, depending on the frequency of the radar system, between 0.3 to 2.5 meters. Extensive comparisons of the HF-Radar current measurements against drifting buoys and upward looking ADCP estimated that 50% of the time the radar measurement errors are lower than 7 cm/s (Paduan and Rosenfeld 1996).

One disadvantage of HF-Radar application is that it can only provide information about past conditions. In contrast, marine operations, search and rescue operations, and oil spill response often require forecast of future currents. Two fundamental approaches to such forecasts exist: (1) assimilating HF-Radar currents into physics-based models of the ocean circulation (Paduan and Shulman 2004; Breivick and Sætra 2001; Shulman and Paduan 2009), or (2) using empirical models to forecast future currents based on a short time history of past observations (O'Donnell et al. 2012; Garfield and Paduan 2009; Almeida 2008). In this paper, we present an empirical method for predicting HF-Radar currents.

The proposed prediction method is an extension of our previous work on statistical emulators of physics-based models (Frolov 2007; Frolov et al. 2009; van der Merwe et al. 2007). In this paper, we train the emulators not on simulations of the ocean circulation, but on historical HF-Radar observations of the surface currents. We will further refer to emulators as empirical models in this paper. Our method operates in two steps: (1) we capture the spatial complexity of the field of interest (surface currents) using empirical orthogonal function (EOF) decomposition of a long historic dataset, and (2) we train a compact statistical model that emulates the dynamics of EOF coefficients. In addition to initial conditions from HF-Radar observations, our model incorporates predicted wind stress from the regional atmospheric model.

Several important differences exist between our prediction method and the empirical prediction methods of O'Donnell et.al. (2012), and Garfield and Paduan (2009). Both O'Donnell et.al. (2012), and Garfield and Paduan (2009) developed their methods for predicting currents at each individual grid point, without taking into account

information about surface currents at neighboring grid points that can inform the prediction about propagating ocean fronts and eddies. In contrast, our method incorporates spatial correlations during the EOF pre-processing step. Both O'Donnell et.al. (2012) and Garfield and Paduan (2009) developed their models as a two-step process, where the predicted currents are a combination of tidal currents, predicted with the harmonic model, and low-pass currents, predicted using weighted averages of low-pass signal during the last few days. In contrast, our model directly learns the tidal signal, low-pass signal, and their interactions from data. Finally, the prediction system by O'Donnell et.al. (2012) requires weekly re-training of the model. In contrast, our model is trained only once.

We test the developed prediction system using a five-year-long dataset (01/01/2006-10/30/2010) of HF-radar observations in Monterey Bay, CA. To evaluate the predictive skill of the developed system, we used two error statistics: root mean square (RMS) error between the predicted and observed currents, and separation between drifters advected with predicted and observed currents. To understand how the accuracy of the developed system compares to the accuracy of existing operational systems in Monterey Bay, we compared the error statistics of our model with the error statics of the empirically-based prediction model of Garfield and Paduan (2009) and with the data-assimilative JPL-ROMS circulation model (Chao et al. 2009).

2 Background on the circulation in Monterey Bay

Several authors (Paduan and Cook 1997; Paduan and Rosenfeld 1996) used HF-radar data to provide an extensive description of circulation patterns in Monterey Bay. Paduan and Cook (1997) showed that the circulation can be divided into three equally

important parts: (1) low-frequency (days to weeks) circulation due to changes in regional wind patterns, (2) semi-diurnal tidal circulation, and (3) circulation due to diurnal sea breeze.

The low-frequency circulation is driven by intensification, relaxation, and reversal of predominant equator-ward winds. (See Figure 1.e for the timeseries of along-shore winds and Figure 1.a for the mean circulation field). When upwelling winds dominate, a strong (~0.1-0.2 m/s) equator-ward jet develops across the mouth of the Monterey Bay (Figure 1.b). During upwelling, circulation inside of the Bay is characterized by weakening of the circulation leeway of Santa Cruz Mountains and a bay-wide counter-clockwise retentive eddy. When upwelling winds relax, circulation over the entire region becomes weak and confused (Figure 1.c). The only remaining coherent pattern is the offshore expansion of the counter-clockwise eddy that was previously trapped inside of the Monterey Bay by a strong upwelling jet. During infrequent periods of strong winter storms, the direction of the flow inside and outside of the bay becomes pole-ward, with a strong coastal jet developing inside of the Monterey Bay (Figure 1.d).

Tidal circulation in Monterey Bay is dominated by barcolinic tides (Paduan and Cook 1997; Rosenfeld et al. 2009). Intensity of surface currents associated with baroclinic (internal) tides depends on the time-varying density structure of the ocean and on the bathymetry. The map of tidal ellipses computed from the HF-radar data (Figure 4 in Paduan and Cook (1997)) shows that M2 tidal velocities vary from almost zero over the deep-waters of the canyon to 0.25 m/s at the head of the canyon.

The diurnal circulation in the Bay is dominated by the sea breeze (Paduan and Cook 1997). Figure 4 in Paduan and Cook (1997) shows high coherence of the diurnal

currents across the entire bay. The circulation ellipses are oriented consistent with direction of the Salinas valley that serves as conduit for marine air entering inland. The diurnal circulation is strongest in the middle of the bay (~0.20 m/s), and decays offshore and in proximity of the land boundary.

3 Methods

3.1 Training and testing datasets

To train and test the developed surface current prediction system, we used a five-year-long dataset (01/01/2006-10/30/2010) of HF-radar currents in Monterey Bay California. Figure 2 shows the configuration of the standard-range HF-Radar network in Monterey Bay. The spatial resolution of the dataset was 3 km and the temporal resolution was 1 hour. To fill-in gaps in the current field due to poor radar returns, the HF-Radar currents were interpolated using Objective Mapping Analysis (OMA; (Kaplan and Lekien 2007)).

To improve the predictive skill of the forecast, we experimented with incorporating the following extraneous forcing variables as an input to the prediction system:

1) Wind stress from the Navy's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS; (Doyle et al. 2009)). The dataset covered the period from January 2006 to December 2010 and had 3 km resolution in Monterey Bay. The temporal resolution was hourly, with the 48-hour forecast issued twice daily.

2) Harmonic prediction of tidal elevations with the regional tidal model of Egbert and Erofeeva (2010). Tidal elevations were predicted based on 8 tidal constituents for 133 points evenly distributed through the HF-Radar domain.

Similar to surface current vectors, we used the EOF pre-processing before incorporating wind stress and tidal data into our empirical model of surface currents.

3.2 Linear autoregressive prediction model

Consider the following linear system that describes the evolution of the surface currents in the ocean:

$$x_{k+1} = \mathbf{A}x_k + \mathbf{B}w_k \quad (1),$$

where $x \in \mathbb{R}^{l_x}$ is the state variable consisting of the vertically concatenated u_{HF} and v_{HF} components of the surface currents at each grid point of the domain; $w \in \mathbb{R}^{l_w}$ is the forcing vector consisting of extraneous forcings, such as u_{wind} and v_{wind} components of the wind stress; **A** and **B** are the state and input matrices; and k is the time index.

Using a least-square training procedure, we are interested in finding such $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ that will minimize the mismatch error ε between observed x and predicted $\tilde{x}$ surface currents:

$$x_{k+1} = \tilde{x}_{k+1} + \varepsilon = \tilde{\mathbf{A}}x_k + \tilde{\mathbf{B}}w_k + \varepsilon$$
$$J(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}) = \sum_{i=1}^{N} \|\varepsilon_i\|_2^2 = \sum_{i=1}^{N} \|x_i - \tilde{x}_i\|_2^2 \quad (2),$$

where N is the number of training samples.

Our previous work (Frolov 2007; Frolov et al. 2009; van der Merwe et al. 2007) showed that it is possible to improve the accuracy of the prediction system (Eq. 1) and the numerical properties of the least-square training algorithm (Eq. 2) by incorporating the following modifications to (Eqs. 1 and 2).

1. To reduce the dimensionality of the training problem, we used an EOF dimension reduction technique:

$$x^s = \mathbf{\Pi}_x (x - \bar{x})$$
$$w^s = \mathbf{\Pi}_w (w - \bar{w}) \quad (3),$$

where $x^s \in \mathbb{R}^{r_x}$ and $w^s \in \mathbb{R}^{r_w}$ are the reduced state and forcings vectors (vectors of EOF coefficients); $\mathbf{\Pi}_x \in \mathbb{R}^{r_x \times l_x}$ and $\mathbf{\Pi}_w \in \mathbb{R}^{r_w \times l_w}$ are the EOF dimension-reduction operators with r_x and r_w modes retained; and $\bar{x}$ and $\bar{w}$ are the mean state and forcing vectors. We computed the dimension reduction operators $\mathbf{\Pi}_x$ and $\mathbf{\Pi}_w$ based on the training data. Prior to the computation, we dimensionalized the state and the forcing vectors by dividing the u and v components by their standard deviations:

$$x = \begin{bmatrix} u_{\mathrm{HF}} / \sigma_{u_{\mathrm{HF}}} \\ v_{\mathrm{HF}} / \sigma_{v_{\mathrm{HF}}} \end{bmatrix}; w = \begin{bmatrix} u_{\mathrm{wind}} / \sigma_{u_{\mathrm{wind}}} \\ v_{\mathrm{wind}} / \sigma_{u_{\mathrm{wind}}} \end{bmatrix} \quad (4).$$

One normalization coefficient σ was computed across all spatial locations.

2. To better capture the temporal evolution of the system and, hence, to improve the prediction accuracy, we introduce augmented vectors *X, W,* and $\tilde{X}$ for initial conditions, forcings, and predicted states.

$$X_k = \begin{bmatrix} x_k^s \\ \vdots \\ x_{k-n}^s \end{bmatrix}; W_k = \begin{bmatrix} w_{k+m}^s \\ \vdots \\ w_{k-n}^s \end{bmatrix}; \tilde{X}_{k+1} = \begin{bmatrix} \tilde{x}_{k+m}^s \\ \vdots \\ \tilde{x}_{k+1}^s \end{bmatrix} \quad (5).$$

For past states and forcings, we used 9 lags: -48, -36, -24, -18, -12, -6, -3, -1, and 0 hours. For predicted states and future forcings, we used 4 lags: 1, 3, 6, and 12 hours. Including more time-embedded states did not significantly improve the prediction accuracy.

3. To capture temporal correlations in each of the time-embedded vectors *X, W,* and $\tilde{X}$, we used a second EOF decomposition:

$$X^s = \mathbf{\Pi}_{x2}\mathbf{T}_{x2}(X - \bar{X})$$
$$W^s = \mathbf{\Pi}_{w2}\mathbf{T}_{w2}(W - \bar{W}) \quad (6),$$

where $\mathbf{\Pi}_{x2} \in \mathbb{R}^{r_{x2} x (n*r_x)}$ and $\mathbf{\Pi}_w \in \mathbb{R}^{r_{w2} x (n+m)*r_w}$ are the EOF dimension-reduction operators with r_{x2} and r_{w2} modes retained to capture 99.9% of the variance, and $\mathbf{T}_{x2}$ and $\mathbf{T}_{w2}$ are the diagonal matrices that normalize each component of the time embedded vectors X and W to have unit variance. We use this normalization to improve the numerical properties of the training algorithm. We do not normalize or apply the secondary dimension reduction to the output vector $\tilde{X}$.

4. Finally, to reduce over-fitting of the matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ to training data, we introduced a regularized cost function that penalizes large weights in matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$:

$$J(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}) = \sum_{i=1}^{N} \left\| X_i^s - \tilde{X}_i^s \right\|_2^2 + \lambda \left\| \begin{bmatrix} \tilde{\mathbf{A}}(:) \\ \tilde{\mathbf{B}}(:) \end{bmatrix} \right\|_2^2 \quad (7),$$

where λ is the regularization parameter that is fit using a cross validation procedure (Frolov et al. 2009).

Using the dimension reduction operators (Eq. 3), time-embedded vectors (Eq. 5), and secondary dimension reduction operators (Eq. 6), we can expand the prediction system (Eq. 2) as follows:

$$\begin{bmatrix} \tilde{x}_{k+m}^s \\ \vdots \\ \tilde{x}_{k+1}^s \end{bmatrix} = \tilde{\mathbf{A}}\mathbf{\Pi}_{x2}\mathbf{T}_{x2} \begin{bmatrix} \mathbf{\Pi}_x(x_k - \bar{x}) \\ \vdots \\ \mathbf{\Pi}_x(x_{k-n} - \bar{x}) \end{bmatrix} + \tilde{\mathbf{B}}\mathbf{\Pi}_{w2}\mathbf{T}_{w2} \begin{bmatrix} \mathbf{\Pi}_w(w_{k+m} - \bar{w}) \\ \vdots \\ \mathbf{\Pi}_w(w_{k-n} - \bar{w}) \end{bmatrix} \quad (8).$$

The output of the prediction system can then be reconstructed as:

$$\tilde{x}_{k+1} = \mathbf{\Pi}_x^T \tilde{x}_{k+1}^s + \bar{x} \quad (9).$$

We implement the prediction system (Eq. 8) and the least square fitting procedure (Eq. 7) using the Netlab© package (Nabney 2004)—an open source network training package for Matlab©.

3.3 Error metrics

4 Results

To characterize the accuracy of the developed system, we trained a series of empirical models with varying inputs. The parameters of each model are summarized in Table 1.

[Table 1 here]

We used two error metrics to evaluate the accuracy of the trained models. (1) An Eulerian RMS error between predicted and observed velocity fields. (2) A separation error between two Lagrangian particles that were advected with predicted and observed currents. Lagrangian particles were seeded every 3 days at each grid-point of the domain. We used an Euler integration method with a timestep of one hour. Reducing the timestep by the factor of 4 did not alter the results significantly.

We evaluated the prediction accuracy for forecast horizons up to 48 hours in the future. The 48-hour forecast was generated as a sequence of hourly forecasts, where the output of the previous forecast was fed back as initial conditions for the next forecast cycle.

4.1 Prediction accuracy

To evaluate the prediction accuracy of the developed system, we trained the empirical model E-HF-W on four years of data (1/1/2006-12/31/2009) and evaluated its performance on ten months of data that were not seen in training (1/1/2010-11/1/2010).

The statistics of the errors are shown in Figure 3. In the first 6 hours, the Eulerian RMS error increased sharply to 0.09 m/s and leveled off at ~0.1 m/s for the 48-hour prediction. The Lagrangian drifter separation increased continuously at a rate of 180 m per hour. The separation between simulated drifters was 4.4 km after 24 hours and, 8.8 km after 48 hours. The timeseries of RMS and drifter separation errors (Figure 4) showed weak correlation (~0.5) with the mean flow speed, indicating an increase in prediction error during periods of energetic flows. Correlations with the along-shore winds were lower (~0.2). However, visual inspection (Figure 4) suggested that periods of increased error coincided with the periods of stronger winds.

To illustrate the behavior of simulated drifters, we plotted trajectories of virtual drifters deployed at the location of M0, M1, and M2 moorings. We plotted these trajectories for a period of high (January 21 and April 22, 2010) and low (February 22 and August 05, 2010) RMS error (Figure 5).

4.2 Sensitivity to the forcing functions

To study how the prediction accuracy changes as a function of forcing inputs to the empirical model, we trained two new models:

- In **E-HF-W-Tide**, we added tidal elevations predicted by the regional model of Egbert and Erofeeva (2010) as an additional input to our base empirical model E-HF-W.
- In **E-HF**, we removed wind stress forcing from the base empirical model E-HF-W. Model E-HF was essentially an unforced system that predicts future surface currents based on past conditions.

See Table 1 for further details on the configuration of each empirical model.

When we compared the error statistics of the three empirical models (Figure 6), we found that our base model E-HF-W (line marked with dots) performed similarly to the model E-HF-W-Tide (line marked with circles). In fact, the two lines were almost indistinguishable on Figure 6. This finding suggests that our autoregressive model was able to independently learn the tidal variability in the system and did not require an external tidal prediction model.

When we compared the performance of the base empirical model E-HF-W with the empirical model E-HW, we found that including wind stress forecast improved surface current prediction for longer forecast times (greater than 6 hours for the RMS error criterion and greater than 12 hours for the drifter separation criterion). For 24 hour prediction the RMS error decreased from 0.1 m/s (E-HW) to 0.09 m/s (E-HF-W) and the Lagrangian separation error decreased from 4.9 km to 4.4 km.

4.3 Sensitivity to the length of the training set

To determine the minimum training length for the empirical model, we trained our base model on a sequence of progressively longer training sets (from 0.25 years to 4 years). We tested these empirical models on the same time interval (1/1/2010-11/1/2010). The test interval was independent of the training data. Figure 7 shows that errors were lower for the longer training sets. A one to two year dataset was required to train an empirical model with accuracy comparable or better than the model of Garfield and Paduan (2009) (see section 4.4).

4.4 Comparisons with existing operational models

How does the prediction accuracy of our base surrogate E-HF-W compare to the accuracy of existing surface current prediction systems? At the time of this publication,

two such systems provided operational forecasts of surface currents in the Monterey Bay area:

1) A JPL-ROMS circulation model (Chao et al. 2009) that assimilated observations of satellite surface temperature, some HF-Radar observations, and profiles of salinity and temperature from moorings and gliders.
2) And an empirical surface current model of Garfield and Paduan (2009) that was trained on the HF-Radar observations for Monterey Bay.

We computed simulated drifter separation errors for all three models for the month of October 2010, when the outputs of the models overlapped. The results of the comparison (Figure 8) show that our empirical model E-HF-W had the lowest error of all three models. After 24 hours, the separation errors were 3.8 km/day for the empirical model E-HF-W, 5.5 km/day for the empirical model of Garfield and Paduan (2009), and 8.9 km/day for the JPL-ROMS circulation model.

Comparisons of the prediction errors between empirical models and the JPL-ROMS circulation model should be taken with a note of caution. This comparison is likely to slightly favor empirical models that were trained to mimic HF-Radar currents exactly. In reality, HF-radar measurements are not error free, and a more fair comparison would compare predicted currents with independent measurement of the surface circulation, such as surface drifters. Unfortunately no surface drifters were deployed in Monterey Bay during the period of this study.

5 Summary and discussion

A novel empirical modeling framework was developed that enabled prediction of ocean surface currents based on the past history of HF-Radar observations and an

optional forecast of wind stress. The developed system was trained on a historic dataset of HF-Radar observations and provides forecasts of surface currents up to 48 hours in the future.

We tested the developed system in Monterey Bay, CA. Surface currents in Monterey Bay are equally driven by low-pass variation in offshore winds, daily sea breeze, and semi-diurnal baroclinic tides. We trained our empirical model using four years of hourly HF-Radar data, and tested the performance of the model on 10 months of data that was not seen in training. The Eulerian RMS errors rose sharply to 0.09 m/s for the 6-hour-prediction and leveled off at 0.1 m/s for the 48-hour-prediction. The Lagrangian drifter separation increased continuously at a rate of 180 m per hour. After 24 hours the separation between simulated drifters was 4.4 km and, 8.8 km after 48 hours. The timeseries of errors showed that the errors were higher during periods of energetic flow.

Through a series of sensitivity studies, we determined that incorporation of the wind stress improved prediction of the surface currents beyond the 12 hour forecast horizon. However, incorporating an external tidal model did not improve the forecast, suggesting that the autoregressive formulation of our model is capable of learning the tidal variability directly. We found that a dataset of at least 1 to 2 years in duration was required to train a prediction system with accuracy exceeding existing empirical prediction systems.

The developed model compared favorably to existing operational forecasts of surface currents in Monterey Bay. For the one month when data for all models were available, the divergence between the simulated drifters was lowest for the developed

empirical model (3.8 km in 24 hours), higher (5.5 km in 24 hours) for the existing empirical model of Garfield and Paduan (2009), and highest (8.9 km in 24 hours) for the data-assimilative JPL-ROMS circulation model. We attribute improved accuracy of our model, as compared to the empirical model of Garfield and Paduan (2009), to (a) incorporating wind stress prediction in our model and (b) better capability to capture spatial and temporal correlations in the data using the two EOF pre-processing steps in our model. We explain the low accuracy of the circulation model because it (a) assimilated only four daily HF-Radar snapshots, as compared to hourly snapshots in our empirical model and (b) it was fitting the HF-Radar data as a least-square compromise between fitting the model forecast, observations of salinity and temperature profiles, and observations of sea surface temperature from the satellite. These results suggest that empirical models may provide a better forecast of surface current conditions than the circulation models that are designed to predict the entire three-dimensional circulation and the hydrography field.

Acknowledgments: We are grateful to the technicians and funding agencies that over the years supported the collection of HF-radar measurements in the Monterey Bay area. The open access to the datasets used in this paper (HF-radar, COAMPS, and JPL-ROMS forecasts) was funded by the Central and Northern California Ocean Observing System. We specifically thank Chris Edwards and James Doyle for facilitating access to COMAPS model files. This work was supported by the ONR project (N00014-10-1-0424).

Tables

Table 1: Description of trained empirical models.

Figures

Figure 1: Climatology of HF-Radar currents. (a) Mean flow-field for all data (01/01/2006-10/30/2010). (b) Mean flow field during upwelling-favorable winds ($V_{alongshore}$ <-0.5 m/s). (c) Mean flow field during relaxation-favorable winds ($V_{alongshore}$ between -0.5 and 0.5 m/s). (d) Mean flow field during downwelling-favorable winds

($V_{alongshore}$>0.5 m/s). (e) Timeseries of along shore winds (positive north) that were used to segment the flow field in panels (b-d). Winds were rotated 30 degrees to the left to orient wind direction with the shoreline direction.

Figure 2: Configuration of the HF-radar installation for Monterey Bay and locations of M1 and M2 moorings.

Figure 3: Average error for empirical model E-HF-W. (a) Eulerian RMS error averaged over the entire domain. (b) Separation error between simulated drifters. Errors were computed for a test period (1/1/2010-11/1/2010).

Figure 4: Time series of 24-hour ahead prediction errors (a-b), average water speed (c), and (d) wind speeds at mooring M1. All time series are for a test period (1/1/2010-11/1/2010). Vertical lines marked with roman numerals mark the periods shown in Figure 5.

Figure 5: Trajectories of drifters advected with HF-Radar currents (blue) and with the currents predicted by E-HF-W model. Selected period correspond to vertical lines marked with roman numerals in Figure 4.

Figure 6: Average errors for empirical models forced with different inputs. (a) Eulerian RMS error. (b) Separation error between simulated drifters. Errors were computed for a test period (1/1/2010-11/1/2010).

Figure 7: Dependence of average errors on the length of the training set. (a) Eulerian RMS error. (b) Separation error between simulated drifters. Errors were computed for the test period from 1/1/2010 to 11/1/2010.

Figure 8: Average separation between simulated drifters for three models. Errors were computed for the test period from 10/4/2010 to 10/30/2010.

References

Almeida F (2008) The influence of wind on HF radar surface current forecasts. Naval Postgraduate School, Monterey, CA

Barrick DE, Evans MW, Weber BL (1977) Ocean Surface Currents Mapped by Radar. Science 198 (4313):138-144

Breivick Ø, Sætra Ø (2001) Real-time assimilation of HF radar currents into a coastal ocean model. Journal of Marine Systems 28:161-182

Chao Y, Li ZJ, Farrara J, McWilliams JC, Bellingham J, Capet X, Chavez F, Choi JK, Davis R, Doyle J, Fratantoni DM, Li P, Marchesiello P, Moline MA, Paduan J, Ramp S (2009) Development, implementation and evaluation of a data-assimilative ocean forecasting system off the central California coast. Deep-Sea Res Pt Ii 56 (3-5):100-126. doi:10.1016/j.dsr2.2008.08.011

Doyle JD, Jiang QF, Chao Y, Farrara J (2009) High-resolution real-time modeling of the marine atmospheric boundary layer in support of the AOSN-II field campaign. Deep-Sea Res Pt Ii 56 (3-5):87-99. doi:10.1016/j.dsr2.2008.08.009

Egbert GD, Erofeeva S (2010) OSU Tidal Inversion software, 2010 solution for the U.S. West Coast.

Frolov S (2007) Enabling technologies for data assimilation in a coastal-margin observatory. Ph.D. thesis, Oregon Health & Science University Portland, OR

Frolov S, Baptista AM, Leen TK, Lu Z, van der Merwe R (2009) Fast data assimilation using a nonlinear Kalman filter and a model surrogate: an application to the Columbia River estuary. Dynamics of Atmosphere and Oceans 48 (1-3):16-45. doi:doi:10.1016/j.dynatmoce.2008.10.004

Garfield N, Paduan JD (2009) Delivery and Quality Assurance of Short-Term Trajectory Forecasts from HF Radar Observations. The Coastal Response Research Center,

Kaplan DM, Lekien F (2007) Spatial interpolation and filtering of surface current data based on open-boundary modal analysis. J Geophys Res-Oceans 112 (C12). doi:10.1029/2006JC003984

Lagerloef GSE, G.Mitchum, R.Lukas, P.Niiler (1999) Tropical Pacific near-surface currents estimated from altimeter, wind and drifter data. Journal of Geophysical Research 104:23,313-323,326

Nabney IT (2004) NETLAB: Algorithms for Pattern Recognition. Advances in Pattern Recognition Springer,

O'Donnell J, Ullman D, Edwards CA, Fake T, Allen A (2012) The Operational Prediction of Circulation and Lagrangian Trajectories in the Coastal Ocean. Ocean Dynamics (this issue)

Paduan JD, Cook M (1997) Mapping surface currents in Monterey bay with CODAR-type HF Radar. Oceanography 10 (2)

Paduan JD, Rosenfeld L (1996) Remotely sensed surface currents in Monterey Bay from shorebased HF radar (Coastal Ocean Dynamics Application Radar). Jounral of Geophysical Research 101 (C9):20,669-620,686

Paduan JD, Shulman I (2004) HF radar data assimilation in the Monterey Bay area. J Geophys Res-Oceans 109 (C7). doi:10.1029/2003JC001949

Rosenfeld L, Shulman I, Cook M, Paduan J, Shulman L (2009) Methodology for a regional tidal model evaluation, with application to central California. Deep-Sea Res Pt Ii 56 (3-5):199-218. doi:10.1016/j.dsr2.2008.08.007

Shulman I, Paduan JD (2009) Assimilation of HF radar-derived radials and total currents in the Monterey Bay area. Deep-Sea Res Pt Ii 56 (3-5):149-160. doi:10.1016/j.dsr2.2008.08.004

van der Merwe R, Leen TK, Lu Z, Frolov S, Baptista AM (2007) Fast Neural Network Surrogates for Very High Dimensional Physics-based Models in Computational Oceanography. Neural Networks. doi:doi:10.1016/j.neunet.2007.04.023

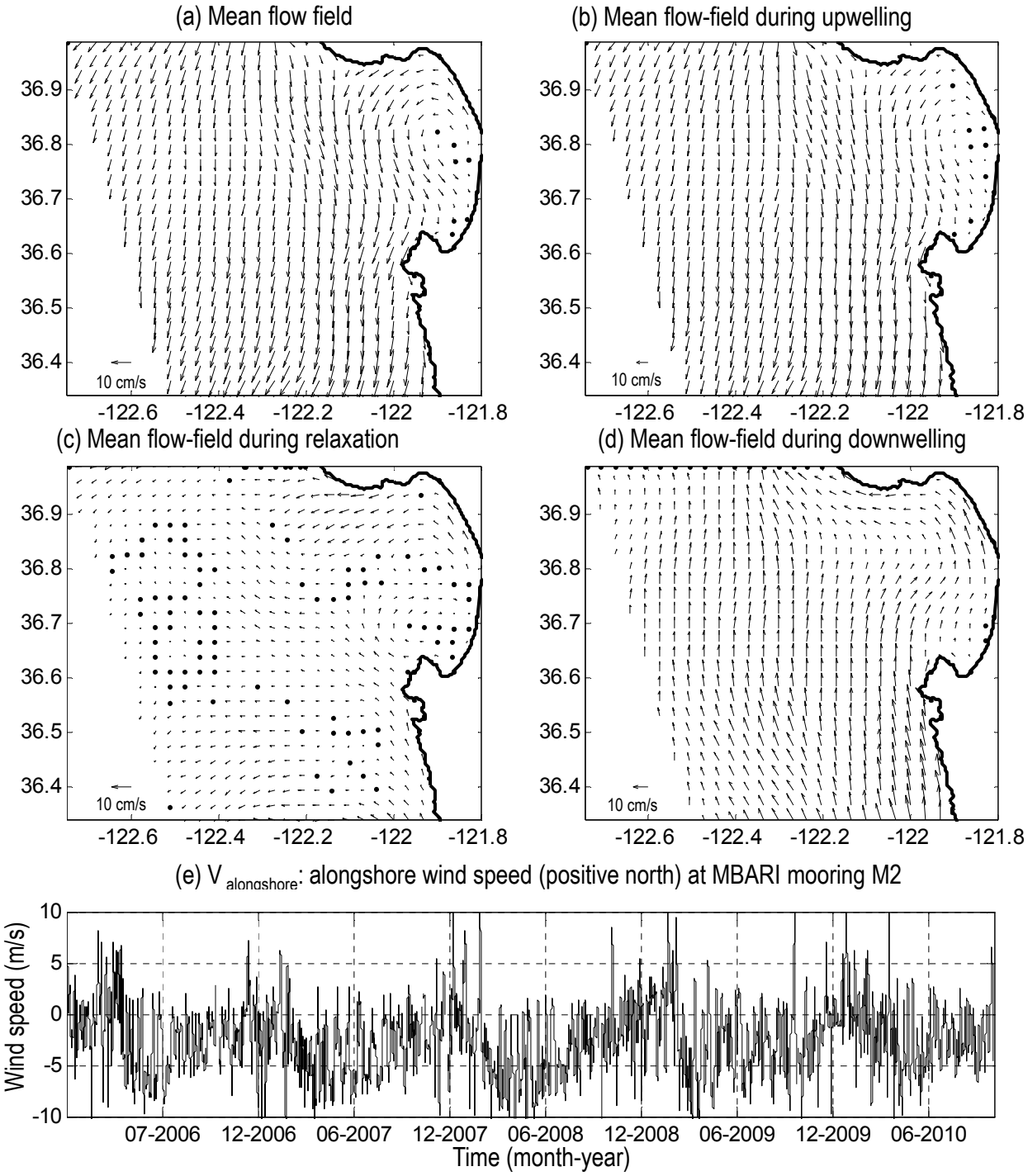

Figure 1: Climatology of HF-Radar currents. (a) Mean flow-field for all data (01/01/2006-10/30/2010). (b) Mean flow field during upwelling-favorable winds ($V_{alongshore}$ <-0.5 m/s). (c) Mean flow field during relaxation-favorable winds ($V_{alongshore}$ between -0.5 and 0.5 m/s). (d) Mean flow field during downwelling-favorable winds ($V_{alongshore}$ >0.5 m/s). (e) Timeseries of along shore winds (positive north) that were used to segment the flow field in panels (b-d). Winds were rotated 30 degrees to the left to orient wind direction with the shoreline direction.

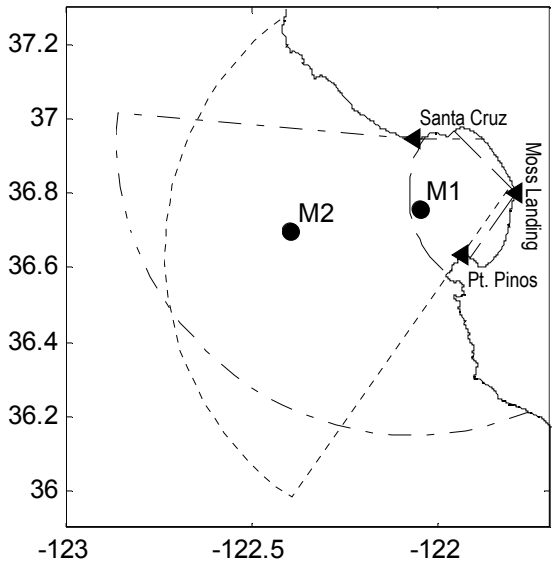

Figure 2: Configuration of the HF-radar installation for Monterey Bay and locations of M1 and M2 moorings.

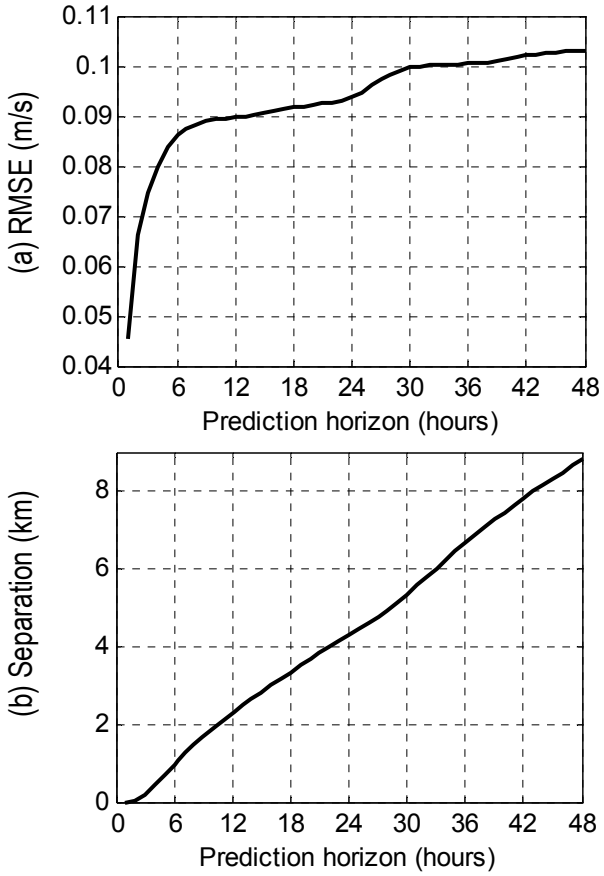

Figure 3: Average error for empirical model E-HW-W. (a) Eulerian RMS error averaged over the entire domain. (b) Separation error between simulated drifters. Errors were computed for the test period from 1/1/2010 to 11/1/2010.

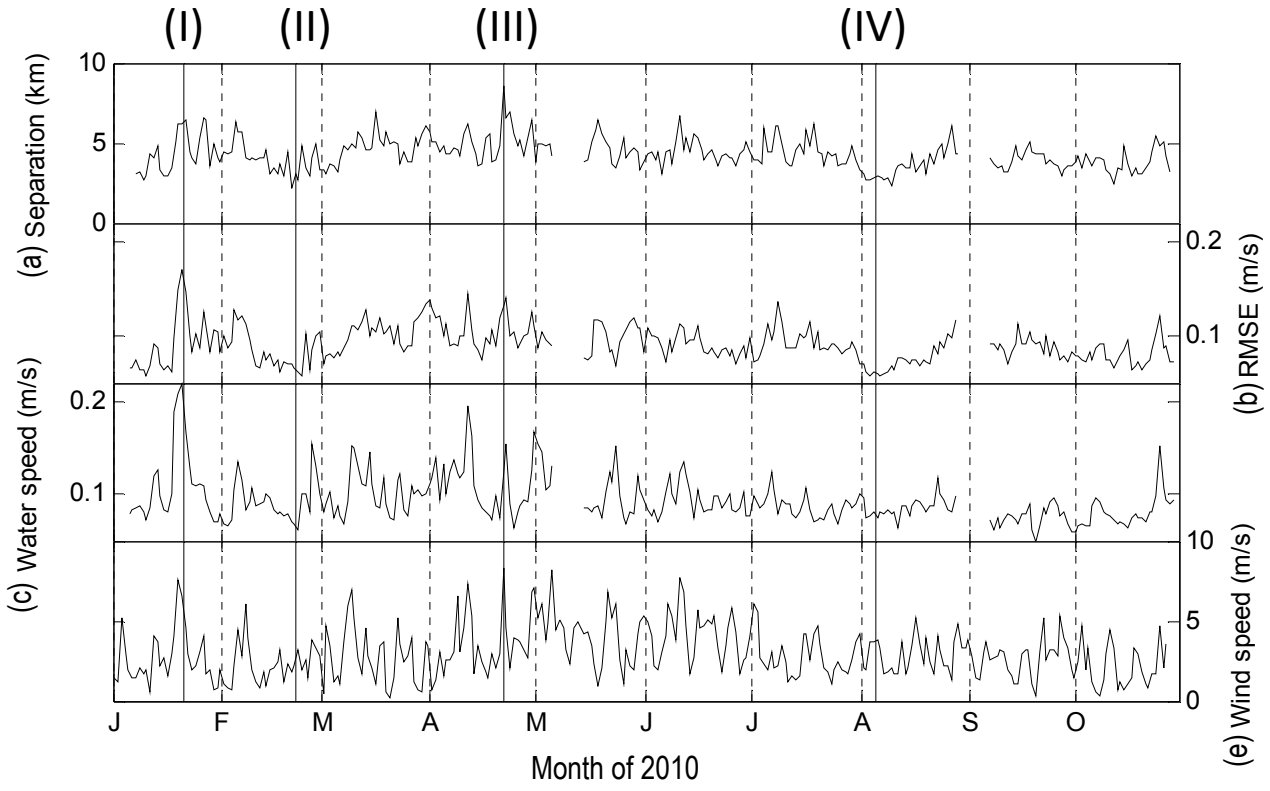

Figure 4: Time series of 24-hour ahead prediction errors (a-b), average water speed (c), and (d) wind speed at mooring M1. All time series are daily-averages and are computed for a test period (1/1/2010-11/1/2010). Vertical lines marked with roman numerals mark the periods shown in Figure 5.

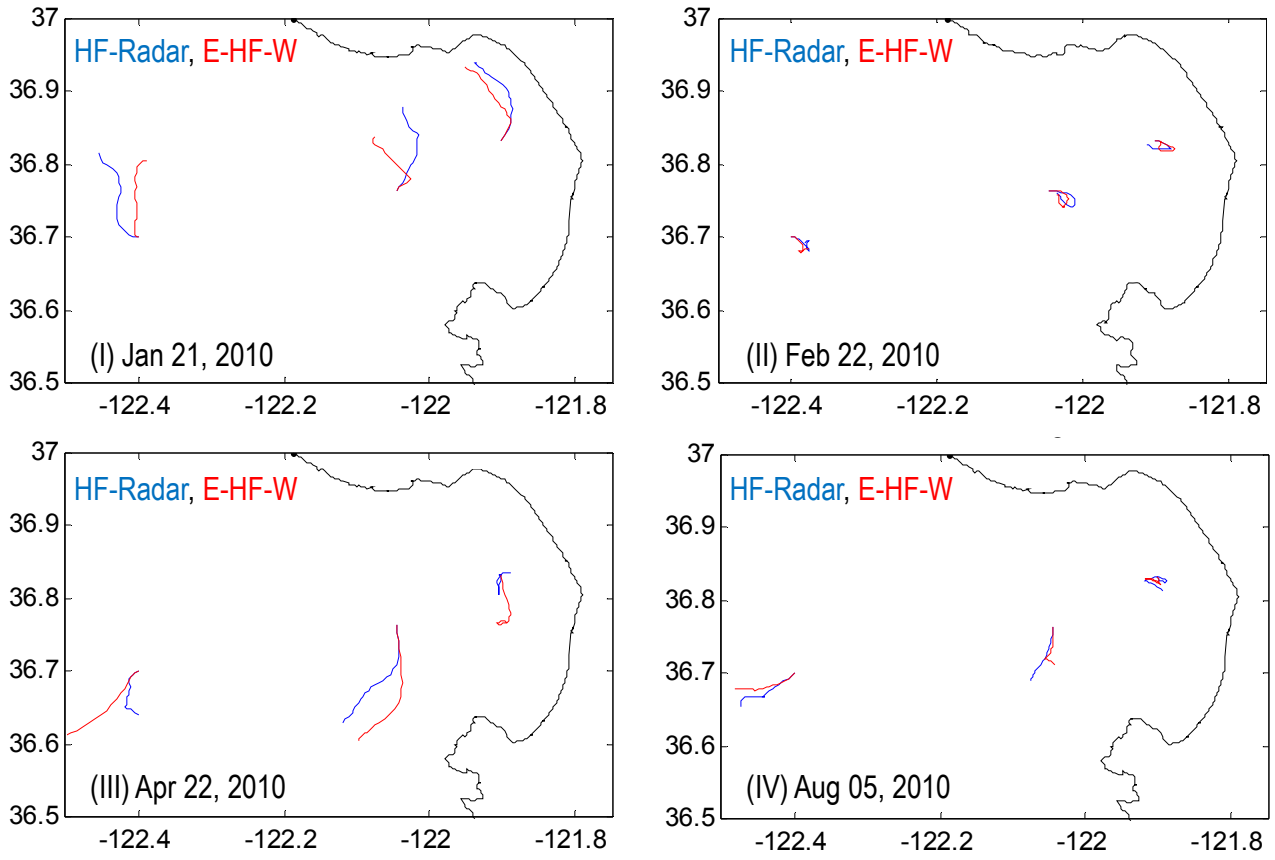

Figure 5: Trajectories of drifters advected with HF-Radar currents (blue) and with the currents predicted by E-HF-W model (red). Selected period correspond to vertical lines marked with roman numerals in Figure 4.

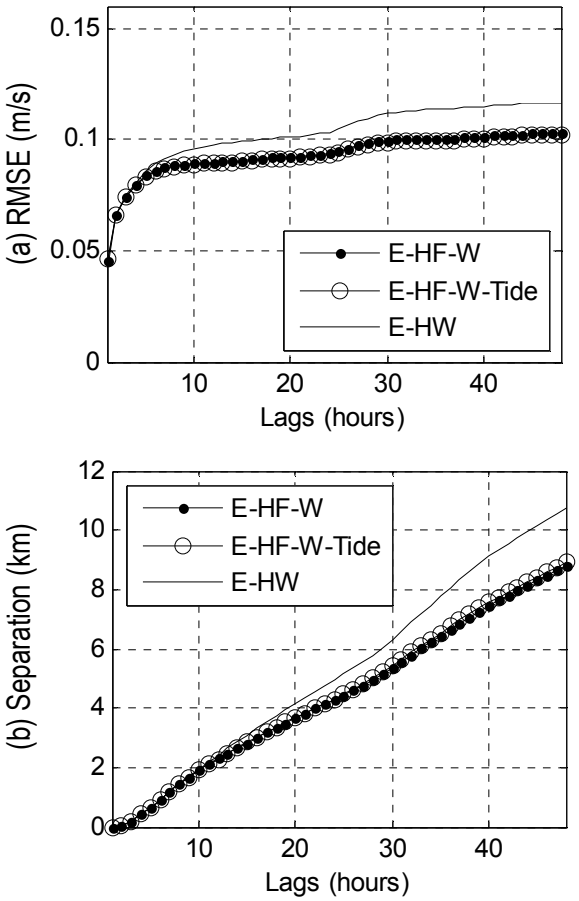

Figure 6: Average errors for empirical models forced with different inputs. (a) Eulerian RMS error. (b) Separation error between simulated drifters. Errors were computed for the test period from 1/1/2010 to 11/1/2010.

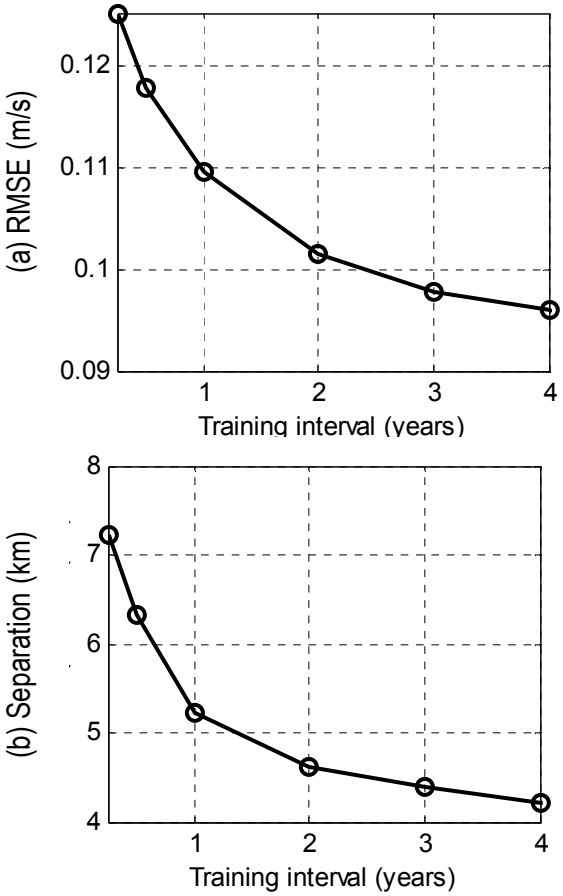

Figure 8: Dependence of average errors on the length of the training set. (a) Eulerian RMS error. (b) Separation error between simulated drifters. Errors were computed for the test period from 1/1/2010 to 11/1/2010.

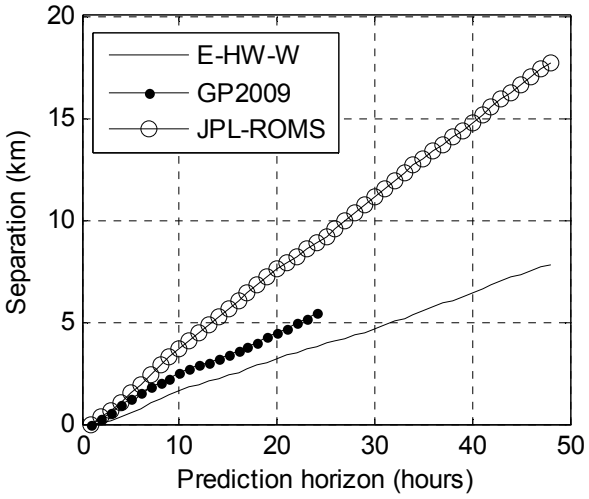

Figure 7: Average separation between simulated drifters for three models. Errors were computed for the test period from 10/4/2010 to 10/30/2010.

Table 1: Description of trained empirical models

Experiment ID	Emulator inputs: # of EOFS (% variance)		
	HF-radar	Wind stress	Tide
E-HF-W	50 (94%)	40 (99%)	---
E-HF	50 (94%)	---	---
E-HF-Tide	50 (94%)	40 (99%)	3 (97%)