Reducing the error in estimates of the Sunda Strait currents by blending HF radar currents with model results

Subekti Mujiasih*[a,b], Dwi Hartanto[b], Jean-Marie Beckers[a] and Alexander Barth[a]

[a]GeoHydrodynamics and Environment Research (GHER), Department of Astrophysics, Geophysics and Oceanography (AGO), University of Liège, Sart-Tilman B5a, 4000 Liège, Belgium

[b]Agency for Meteorology, Climatology, and Geophysics of the Republic of Indonesia (BMKG)

ACCEPTED MANUSCRIPT

Highlights

- Model accuracy can be improved by using a blending method, ETKF, with HF radial velocity and an optimal representativity error from an independent validation
- Improved model analysis can be obtained either from the original HF radar data or from HF radar data obtained at a different site
- Every site has a different optimal representativity error
- The root mean-square errors of the currents are modulated by the strength of the winds

Abstract

The examination of currents by merging model results and radial velocity High-Frequency (HF) radar data has been undertaken in the Sunda Strait, which links Sumatera Island and Java Island, involving two sites (Anyer and

*Corresponding author.

Email address : mujiasih.subekti@uliege.be (S.Mujiasih), subekti.mujiasih@bmkg.go.id

Present address : GeoHydrodynamics and Environment Research (GHER), Department of Astrophysics, Geophysics and Oceanography (AGO), University of Liège, Sart-Tilman B5a, 4000 Liège, Belgium.

Permanent address : Agency for Meteorology, Climatology, and Geophysics of the Republic of Indonesia (BMKG), Regional Office III, Jl. Raya Tuban, Kuta 80361, Badung, Bali, Indonesia

Preprint submitted to the Continental Shelf Research on September 10, 2020

First revised preprint resubmitted to the Continental Shelf Research on April 04, 2021

Second revised preprint resubmitted on July 19, 2021

Accepted on July 22, 2021.

© 2021. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/

For citation, please refers to the final version in https://doi.org/10.1016/j.csr.2021.104512

Labuan) and using the Ensemble Transform Kalman Filter (ETKF). Dependent validation involved the data used during model analysis while independent validation utilised observations from a different site. These validations are needed to obtain an optimal representativity error, which has the lowest averaged root mean square (RMS) over time and is appropriate for all sites. Moreover, we evaluated the optimal representativity error with the relative error reduction and the associated skill score metric. The results show that the model analysis for both independent and dependent validation have better results than the model with no blending. Interestingly, independent validation has a smaller RMS than the model with no blending, although it is still greater than the dependent validation. The best results were obtained from model analysis of all sites, with 0.2704 m/s from Anyer and 0.4459 m/s from Labuan being the representativity error value. However, they have a pattern in the RMS error over time series. It is necessary to consider the factor such as winds that would modulate the magnitude of radial velocity.

Keyword Radial velocity, HF radar, Sunda Strait, Sumatera, Java, Indonesia

ACCEPTED MANUSCRIPT

1 Introduction

Economic activity at sea requires high-quality marine weather information to prevent or reduce the risk of loss. It is a big challenge for researchers to improve the accuracy of both the marine model and the technology of observations. During the last two decades, the research of combining models and observations, has become an important research area and still leaves many unsolved questions. Models are based on mathematical equations that describe physical conditions and are solved numerically. The advantage of the model is the fact that they can produce output for the past, present and future, at high temporal resolution, potentially covering a large area if enough central processing unit (CPU) power is available together with a high spatial resolution. In contrast, model output can have uncertainty, which can be substantial, while observations typically have lower uncertainties than the model. On the other hand, the effort to make observations, retrieval, data collection and maintenance of the observation equipment itself all have a huge cost. The combination of models and observations can be a solution to improve the accuracy of marine forecasting.

To improve the accuracy of an ocean model, ideally one would need evenly distributed and continuously available observations. An observation type, which has these characteristics, is High-Frequency (HF) radar. In the last decade, one of the growing research areas in the field of oceanography is the incorporation of ocean models and HF radars. HF radar is reliable in capturing spatial ocean surface phenomena with a high coverage (Paduan & Washburn, 2012), such as wave (Orasi et al., 2018) and surface currents (Abascal et al., 2012; Kim et al., 2008; Kohut et al., 2004; Paduan & Washburn, 2012; Solabarrieta et al., 2014; Yaremchuk & Sentchev, 2009) and, in particular, tidal currents (Tian et al., 2015) and tsunami (Lipa, Barrick, et al., 2012; Lipa et al., 2011; Lipa, Isaacson, et al., 2012). Indirectly, HF radar also gives information about winds (Kirincich, 2016; Lana et al., 2016; Orasi et al., 2018). In practice, HF radar can be used for managing hazard risks, such as navigation safety at ports and docks, controlling pollution, sedimentation when dredging, tsunami warning, monitoring positions of cold fronts in open ocean and sea breeze fronts in particular locations (Heron et al., 2016). The high spatial coverage attracts researchers to merge these data with the hydrodynamic model.

There are two methods of combination, namely blending and data assimilation. Blending is a method of combination between the model and observation data to obtain the best estimation at time k that we call a "model analysis". Almost similar to blending, data assimilation obtains the best estimation at forward time $k+1$, $k+2$ and so on, that we call "a model forecast". Some blending research has been carried out, such as aiming to produce nowcasts (present and future events) of the surface velocity by filtering using a Codar HF radar System with a Natural Mode Analysis method and gap-free nowcasts (Lipphardt Jr et al., 2000), estimating Lagrangian transport using the Wera system (Berta et al., 2014) and the analysis of tidal hindcast (past and present events) from radial currents, also using the Wera system (Stanev et al., 2015). A further method is data assimilation, which does not only produce the model analysis but also forecasting (Barth et al., 2008; Breivik & Satra, 2001; Ren et al., 2016; Vandenbulcke et al., 2017; Xu et al., 2014; Yu et al., 2012). Data assimilation methods in the body of literature include the Ensemble Kalman Filter (Barth et al., 2008; Breivik & Satra, 2001), the Variational method (Yu et al., 2012), the Optimal Interpolation (Xu et al., 2014), the Lewis assimilation scheme based on shearing stress (Lewis et al., 1998) in (Ren et al., 2016), and the Ensemble Transform Kalman Filter (ETKF) (Vandenbulcke et al., 2017) and the Physical-space Statistical Analysis System (PSAS) scheme, followed by the physical principle (Paduan & Shulman, 2004), and the Kalman Filter applied on HF radial velocity (Shulman & Paduan, 2009).

Studies of HF radar in a strait region (Strait of Gibraltar) have also been previously undertaken and described in the literature (Soto-Navarro et al., 2016). Identifying characteristic surface currents in a strait is not an easy task because of the narrow shape and the difficulty of deploying instruments due to the high degree of economic activity. HF radar, however, has the capability to map surface currents remotely using shore-based equipment. That capability enabled our measurements in the Sunda Strait as well as those in the Straits of Gibraltar (Soto-Navarro et al., 2016). Soto-Navarro et al. (2016) compared the Autonomous Measurement, Prediction and Alert System in the Bay of Algecira (SAMPA is the Spanish acronym) model output against 3 sites of HF radar at the Strait of Gibraltar using statistical metrics such as variance, complex correlation, veering angle, scalar correlation and root mean square error. The period of the used data was February 2013 – September 2014. The parameters used were zonal and meridional velocity components. Their study shows the existence of currents that are stronger than for other areas when currents flow out from the strait. It appeared from monthly mean velocity in February, May, August and November. The pattern also occurred in the assimilation result in the channel between Xuejiadao Island and Xiaomaidao Island (Qingdao, China, on the western coast of the Yellow Sea) (Xu et al., 2014), and in the mean surface circulation pattern from HF radar for 2016–2017 in the Gibraltar Strait (Lorente et al., 2019). We analyse the similarity of this pattern in the Sunda Strait.

Since the 1960s, the Sunda Strait has been receiving the attention of marine researchers notably regarding oceanographic conditions (Amri et al., 2014; Jumarang & Ningsih, 2013; Koropitan et al., 2006; Li et al., 2018; Novico et al., 2015; Oktavia et al., 2011; Pariwono, 1999; Potemra et al., 2016; Rahmawitri et al., 2016; Sandro et al., 2014; Susanto et al., 2016; Wyrtki, 1961). They conducted research using various data such as vessel observations data (Wyrtki, 1961), ship drift (Pariwono, 1999), in situ observation (Amri et al., 2014), ADCP (Li et al., 2018; Novico et al., 2015; Susanto et al., 2016), Princeton Ocean Model (Koropitan et al., 2006), satellite (Rahmawitri et al., 2016; Sandro et al., 2014), geostropic currents derived from tides-gauges (Oktavia et al., 2011), the Nucleus for European

ACCEPTED MANUSCRIPT

Modelling of the Ocean – Ocean Parallelise (NEMO-OPA) Model (Rahmawitri et al., 2016), and HYCOM (Potemra et al., 2016). However, to the best of our knowledge, there are only a few studies focusing specifically on the variability of currents.

The Sunda Strait became a focus of attention because of the presence of HF radar at that location, which was previously used for tsunami detection, but can also be used to better understand the surface circulation. The existence of currents data inspires us to test the merging with the ensemble method so that it allows us to bring the model closer to the observation. With respect to that reason, previous research and the lack of availability of observational currents spatially, we propose a blending method using the Ensemble Transform Kalman Filter (ETKF) (Bishop et al., 2001) and the explanation (Vetra-Carvalho et al., 2018) to provide the best estimation of surface currents in spatial distribution.

This paper is presented in five sections. Section 1 contains the background of this research and a review of previous studies. Section 2 discusses data, methods and steps for processing the data. Section 3 displays the results such as the comparison of a model without blending, a model with blending and the observation; the selection of an optimum representativity error, and the assessment of the usage of the optimum representativity error for producing the best estimation for all sites. Section 4 provides a discussion on topics such as the comparison between the performance of the previous study and the present study and also the fluctuation of RMS signal which obtained by the optimum representativity error. Section 5 concludes.

2 Material and Methods

2.1 Material

In this section, the research domain, data, methods used and steps of data processing are explained. The research area chosen is the Sunda Strait (see Fig. 1), which is created by using the M_Map application (Pawlowicz, 2020). The Sunda Strait connects the Indian Ocean and the Java Sea, and the northeastern part is a narrow channel and shallow, in contrast to the southwestern part which is wide and steep. Three types of data were used in this study. The first type is the output from Copernicus - Marine Environment Monitoring service (CMEMS) model, which has hourly-mean zonal u and meridional velocity component v during one year, 01 September 2013 until 31 August 2014, from the global ocean 1/12 degrees physics analysis and with the forecast updated daily. This model does not include tides. The second type was HF radar Coastal Ocean Dynamics Application Radar (CODAR) SeaSonde radial velocity for 23 September 2013 09 UTC until 22 December 2013 01 UTC from both sides of the strait (Anyer and Labuan; see Fig. 1), which have a "measured antenna pattern" an hourly temporal resolution, 20-60 km of spatial range, 3 km of range resolution, 5 degrees of angular resolutions and spatial resolutions, and 11.5-14 MHz of frequency. Measured antenna pattern means the pattern of antenna, which is adjusted with respect to the environment of the specified site (Paduan et al., 2006). All radial velocity data have metadata, which contain all information of that data representation. One sample of the data is the type of pattern. All data we have are using a "measured" type of antenna pattern. Time series data availability of each site are described in Fig. 2 and spatial data availability in Fig. 3. Limited measurement

ACCEPTED MANUSCRIPT

at Labuan, due to the energy supply from a solar panel, was sometimes unstable. The third data type was hourly wind speed at 10 metres from the meteorological station at Serang Banten, which is part of BMKG. Winds were used for comparing the signal pattern of the blending result.

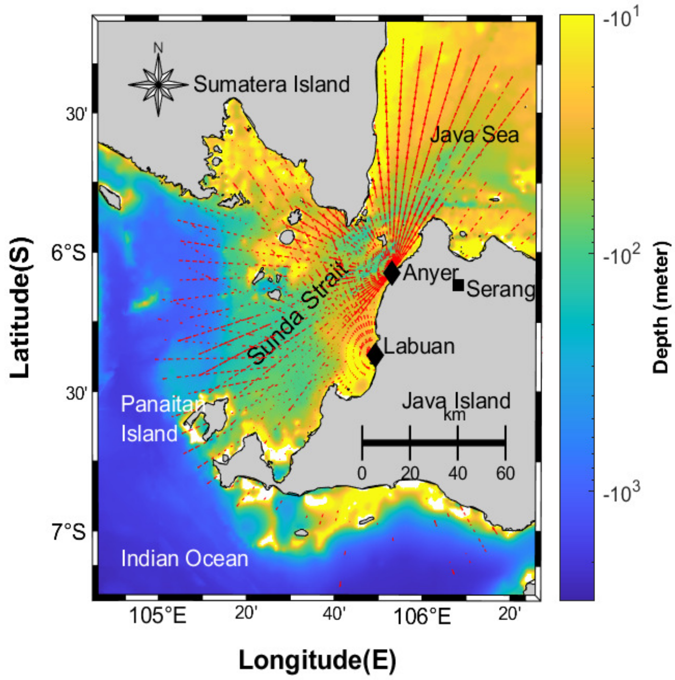

Fig. 1. Research domain.(Credit: Bathymetry from the General Bathymetric Chart of the Oceans (GEBCO)) (Group, 2020)

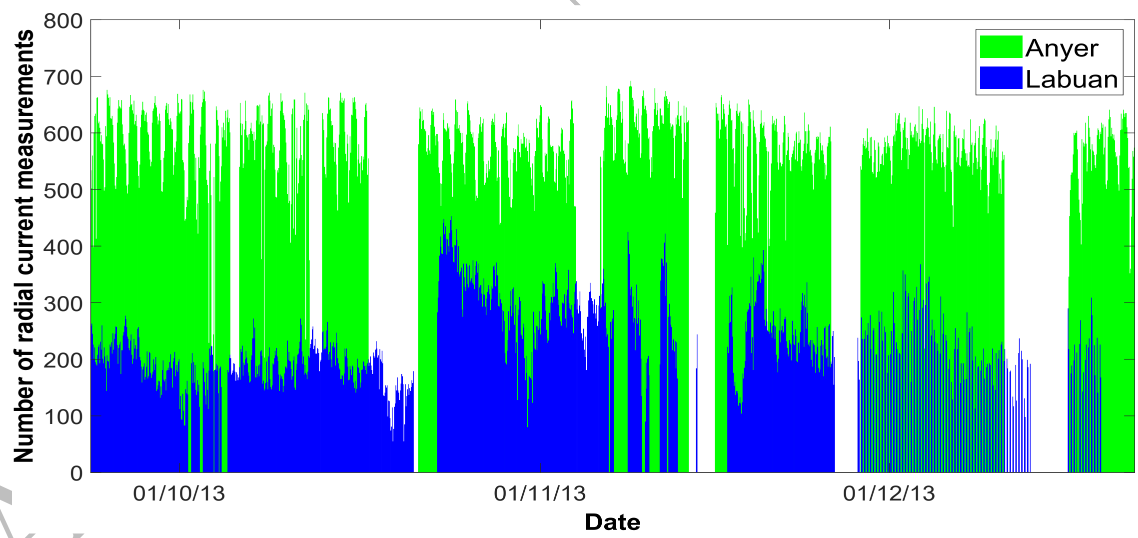

Fig. 2. Time series of the availability of radial currents at each site 23 September - 22 December 2013

Fig. 3 shows the radial velocity's spatial availability illustrated by the percentage of data coverage for each grid. The percentage is an average of the total number of vectors over 2153 hours. Meanwhile, the total number of vectors is a sum of radial vectors since 23 September 09UTC – 22 December 2013 01UTC (over 2153 hours). The minimum percentage of data coverage in the Anyer site is 8.5462% or equal to 184 points, at 105.2377^{o}E, 5.5680^{o}S. The maximum percentage of data coverage in the Anyer site is 78.6345% or equal to 1693 points, at 105.8318^{o}E, 6.1373^{o}S. The minimum percentage of data coverage in the Labuan site is 0.0929% or equal to two points, at 105.7941^{o}E, 6.3621^{o}S.

The maximum percentage of data coverage in the Labuan site is 68.7413% or equal to 1480 points, at 105.3961^{o}E, 6.2651^{o}S. Overall, the Anyer site has better availability than the Labuan site. The data availability is about 40% – 100% of data coverage or about 800 – 1600 points. In contrast with the Labuan site, high availability is only in the west and the center. The other areas are mostly below 40% of data coverage or below 800 points.

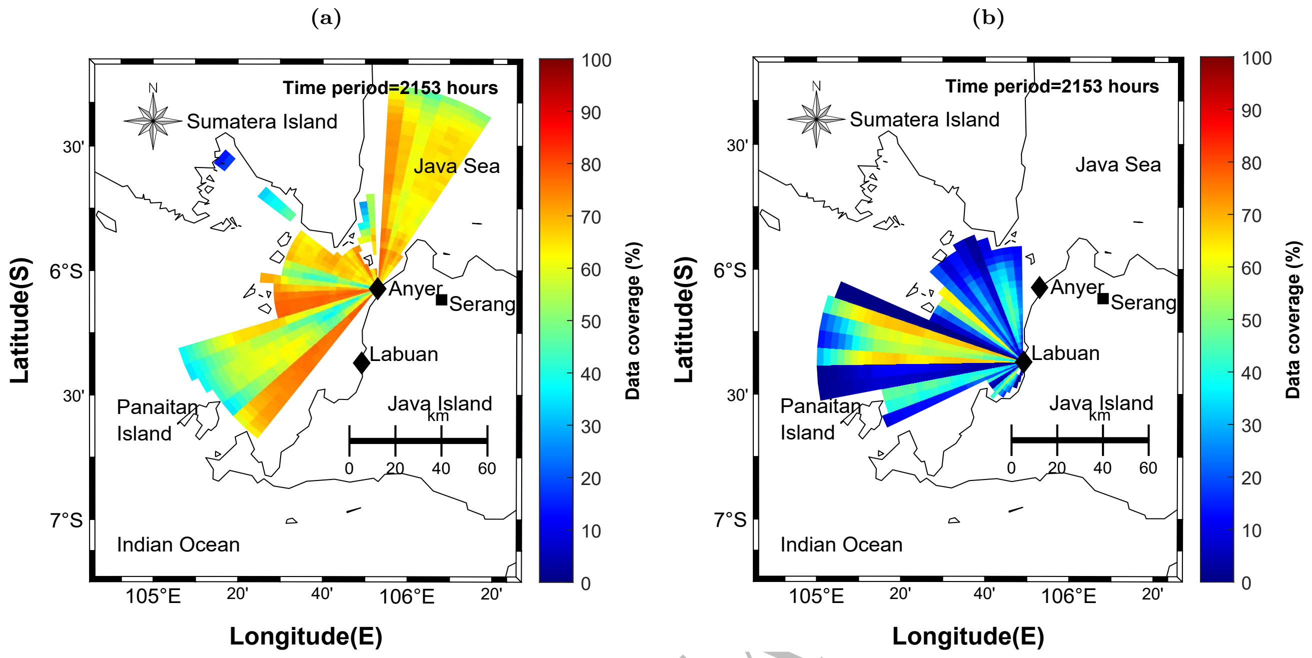

Fig. 3. Percentage of data coverage at each site 23 September - 22 December 2013 at the Anyer site (a) and the Labuan site (b)

ACCEPTED MA

2.2 Method

2.2.1 Background

Our study focuses on applying a blending method to obtain the estimates of currents in the Sunda Strait (linking Sumatera Island and Java Island, Indonesia) (BIG, 2017). The variability of the current in the Sunda Strait flows mostly from the Java Sea towards the Indian Ocean the whole year (Pariwono, 1999; Rahmawitri et al., 2016; Wyrtki, 1961). Rahmawitri et al. (2016) implicitly noted that sea surface height (SSH) in the Java Sea is higher than that of the Indian Ocean throughout the year except for November-January. Surface currents are related to SSH via the geostrophic equilibrium. Nevertheless, there are different exceptional months, the currents flow towards the reverse directions such as in March, August, and October (Pariwono, 1999), and November-January (Rahmawitri et al., 2016). Notwithstanding, Amri et al. (2014) found currents coming from the Java sea which appeared along the west coast of Banten (Carita, Labuan until Tanjung Lesung) flowing southwestward and being deflected northeastward around Panaitan Island. Moreover, Oktavia et al. (2011) conclude that geostrophic currents variation is indirectly influenced by winds in the Sunda Strait. However, with respect to the speed of currents from the previous research papers in the literature, they vary depending on data availability and type. Generally, the maximum speed of currents was about 2.63 m/s on 18 October 2012 at 16:30 Western Indonesian Time (WIB) at the narrow channel in the northeastern part of the Sunda Strait (Novico et al., 2015).

In comparison to previous research such as that undertaken by Lipphardt Jr et al. (2000), Berta et al. (2014), Stanev et al. (2015), and Shulman and Paduan (2009), there are some similarities and differences regarding our investigation. The similarity of the present study with that of Lipphardt Jr et al. (2000), is that we applied blending to combine HF radar and a model. In addition, we used the same HF CODAR but our study is different with respect to method, HF radar data and output. Lipphardt Jr et al. (2000) used the Natural Mode Analysis method for blending and HF surface currents for blending input, and their output was a nowcasting result (a short few hours of forecasting). Meanwhile, in the present study we utilised the ETKF method, HF radial velocity and we also produced a model analysis. One difference with respect to previous research in the literature is that our model analysis is valid for time k, in that, it is not valid for time $k+1$, $k+2$, and so on. Berta et al. (2014) assimilated Lagrange transport (trajectory) from a model, HF radar and drifter using the LAgrangian Variational Analysis (LAVA) method in the Ligurian Sea (between Italian Riviera and Isle of Corsica). Those authors used trajectory objects that were superimposed with surface currents from HF radar plus a model, model analysis and forecasting. While we utilised surface current objects only and compared them to model analysis. Shulman and Paduan (2009) had a similarity with ours, namely the usage of HF radial velocity and the Kalman Filter scheme. However, we used one variant of the Kalman Filter, ETKF. The differences are that they used a numerical model with tides and validated the forecast output from their assimilation. We only focused on a blending method to see the benefit and examined whether we can produce the analyzed model from only one site. Berta et al. (2014) produced blended and forecasting output compared to drifter and HF radar itself, while the present study had a focus on the blending process only. Hence, the similarity is only in terms of an analysis output. Compared to other previous research such as that by Stanev et al. (2015), we also used radial velocity for blending. Another similarity is the use of Kalman filtering for blending. However, in the present work we utilised an ensemble variant of the Kalman filter, the ETKF (Bishop et al., 2001). The difference is that those authors considered tides in simulation because the research area was dominated by tides. While in the present research we used one year of an ensemble model, 3 months of HF radial velocity of CODAR SeaSonde and also tides were not considered. The present study proposes another way to validate by using cross validation of each site, namely independent and dependent validation and estimates of an optimal representativity error to obtain the best analysis for all sites; besides, the previous study applied a blending for shallow water areas. In the present study, we applied a blending of not only strait, which is relatively shallow with about 0 - 100 metres depth, but also of the continental shelf area with more than 200 metres depth (see Fig. 1) because the Sunda Strait borders with the Indian Ocean at the southwestern part. The other difference is that Stanev et al. (2015) have used an acoustic Doppler current profiler (ADCP) for validating blending results, which is not only analysis but is also hindcast, nowcast and forecast. For reducing computational cost, the state vector was decomposed into eigenvectors and eigenvalues. The decomposition method used by those authors is the Empirical Orthogonal Function (EOF), whereas in the present study we use the Singular Value Decomposition (SVD). The analysis covariance matrix used by Stanev et al. (2015) is based on the model state matrix, while in the present study we use the decomposition of the inverse transformation matrix, which originates from the model state perturbation matrix in observation space and the innovation matrix. In our study we use the representativity error which is included in the observation error covariance matrix (R). The resulting model analysis was validated independently relative to radial velocity from different sites (independent validation). As a result, we obtained one representativity error for every site, which gives the lowest average root mean square error (RMS). All model analyses

ACCEPTED MANUSCRIPT

have lower RMS than models without blending, while in Stanev et al. (2015) the representation error was represented by multiplication of the observational error covariance by a factor 25.

In the present study, the best estimation could be achieved by testing the sensitivity of the representativity error to averaged root mean square error (RMS). The testing can show the optimal representativity error, which gives the smallest averaged RMS. In addition, this method can also provide a larger coverage of the best estimation, which is not only limited to the HF radar domain but also goes beyond the models. Hence, the fluctuation of surface currents outside HF radar coverage can be made to clearly appear. To carry out this, we blend the Copernicus Marine Environment Monitoring Service (CMEMS) model with the HF radar radial velocity of each site. Radial velocity from one site is blended with the CMEMS model to create a full map of surface currents and then this current map is compared to the originally used data (dependent validation) and to the HF radar of other site (independent validation). The optimal representativity error for all sites was obtained by cross validation. The improvement is shown by error reduction and skill score after blending. If it is significant, it becomes very interesting to continue with the next step, such as data assimilation, to produce a forecast. The other benefit is that we will have the model analysis (the best estimation) in the same abundance as for the HF radar data; in that, the more HF radar that is involved, the more model analysis that will be produced. We treat a long period of CMEMS model as an ensemble model with which to provide the more representative model analysis with consideration of: the more ensemble members there are, the more accurate the model analysis will be. However, in the present study the ensemble model remains constant over time. We may conclude that the novelties of this study are the usage of HF surface currents, a blending method to obtain the currents estimates from one site of HF radar, and the optimal representativity error.

2.2.2 Preprocessing

The hourly CMEMS model data from the September 1, 2013 to August 31, 2014 representing 8761 time instances are used. In the following, it is assumed that this time variability of the model can be used as proxy of the error covariance (Stanev et al., 2015). Those authors considered winds, which is a time-variant parameter, while in our study, we do not have a true ensemble simulation, hence we also use the time variability as a proxy. However, this approximated ensemble variability keeps constant over time. This approach had been implemented for assimilating altimetry data and ocean model using the Singular Evolutive Extended Kalman (SEEK) filter with a time-independent error sub-space scheme (Brasseur et al., 1999).

Before the HF radial velocity is used in the blending, the data are preprocessed in 3 stages. We deleted bad data such as incomplete coordinate data, vector data, which were not in the sea and not used for total vector and also non-calculable data. Next, we detected and removed outliers by using the scaled Median Absolute Deviation (MAD), which detects elements that have value more than three times the scaled Median Absolute Deviation (MAD) from the median. The scaled MAD is defined as Eq. (1) for a random variable vector A with N scalar observation with Eq. (2) for c coefficient. The last term in Eq. (2) usually uses a value of 1.4826. This method was introduced by R.Hampel (1974) as mentioned by (Leys et al., 2013). In our case, outliers are removed during time series N since September 23, 2013 05 UTC - April 1, 2014 04 UTC. Afterwards, we removed the periodic tides effect on the radial velocity using the T_Tide application (Pawlowicz et al., 2002). The package contains a formula to remove specific frequencies of

ACCEPTED MANUSCRIPT

tides, which explicitly removes the variability of the signal at tidal frequencies but leaves any other frequencies in our dataset. Tidal signal was removed since CMEMS in this study has not considered tides. Hence, the radial velocity to be used also needs to be removed from the effect.

$$\mathrm{MAD} = c(\mathrm{median}\,|A_i - \mathrm{median}(A)|) \tag{1}$$

where i=1,2,....N and

$$c = -1/\sqrt{2}(\mathrm{erfc}^{-1}(\frac{3}{2})) \tag{2}$$

2.2.3 Processing

The method that is used in this study is the Ensemble Transform Kalman Filter (ETKF). It is a variant of the Ensemble Kalman Filter that was first introduced by Evensen (1994), which is a development of the Kalman filter method (Kalman, 1960). The method inverses the observational error covariance matrix $\mathbf{R}$ so that it can be easily identified. For an explanation regarding calculating ETKF the reader is referred to the user manual: Sangoma Package (Vetra-Carvalho et al., 2018).

Regarding the Ensemble Transform Kalman Filter, there are four general formulas including the updated ensemble mean $\bar{\mathbf{x}}^a$ as Eq. (3), the analysis error covariance matrix $\mathbf{P}^a$ as Eq. (4), Kalman gain $\mathbf{K}$ as Eq. (5), the analysis ensemble $\mathbf{X}^a$ as Eq. (6)

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{K}(\mathbf{y} - H(\bar{\mathbf{x}}^f)) \tag{3}$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f \tag{4}$$

where $\mathbf{K}$ is Kalman gain, which is defined as

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R})^{-1} \tag{5}$$

where $\mathbf{R}$ is an observational error covariance matrix, $H(..)$ in Eq. (3) is a linear observation operator for scalar form. $\mathbf{H}$ Eq. (4) is an observation operator in the forecast ensemble mean (matrix form). The observation operator contains transformation values from model grid to observations grid; with the analysis ensemble given by

$$\mathbf{X}^a = \bar{\mathbf{X}}^a + \mathbf{X}'^a \tag{6}$$

Where $\bar{\mathbf{X}}^a = (\bar{\mathbf{x}}_1^a, \bar{\mathbf{x}}_2^a, ..., \bar{\mathbf{x}}_N^a) \in \mathcal{R}^{nxN}$. $\bar{\mathbf{X}}^a$ is the ensemble analysis mean. While $\mathbf{X}'^a$ is the ensemble analysis perturbation; superscript $(.)^a$ and $(.)^f$ denote analysis and forecast, respectively.

Besides formula in Eq. (4), the initial error covariance matrix $\mathbf{P}$ can be calculated from covariance around the mean $\bar{\mathbf{x}}$ at the time index $k = 0$ by using

ACCEPTED MANUSCRIPT

$$\mathbf{P}^{a,(0)} = \frac{1}{N-1}\sum_{j=1}^{N}(\mathbf{x}_j^{a,(0)} - \bar{\mathbf{x}})(\mathbf{x}_j^{a,(0)} - \bar{\mathbf{x}})^T \tag{7}$$

or

$$\mathbf{P}^{a,(0)} = \frac{\mathbf{X}'^{a,(0)}(\mathbf{X}'^{a,(0)})^T}{N-1} \tag{8}$$

where $j = 1,...,N$ is an ensemble member index, N is the total number of the ensemble. The subscript T denotes transpose. Because we aim to obtain an analysis, so we omit the time index $k = 0$, thus equation of the analysis ensemble error covariance Eq. (8) can be written as

$$\mathbf{P}^{a} = \frac{\mathbf{X}'^{a}(\mathbf{X}'^{a})^T}{N-1} \tag{9}$$

Based on the derivation by Vetra-Carvalho et al. (2018), Eq. (4) and Eq. (9) can be obtained by using the ensemble perturbation matrix in observation space $\mathbf{S}$, the innovation covariance matrix $\mathbf{F}$, and the transformation matrix $\mathbf{TT}^T$ and the ensemble forecast perturbation $\mathbf{X}'^f$ as in the following equations.

$$\mathbf{X}'^{a}(\mathbf{X}'^{a})^T = \mathbf{X}'^{f}(\mathbf{I} - (\mathbf{S}^T(\mathbf{SS}^T + (N-1)\mathbf{R})^{-1}\mathbf{S}))(\mathbf{X}'^{f})^T \tag{10}$$

$$= \mathbf{X}'^{f}(\mathbf{I} - \mathbf{S}^T\mathbf{F}^{-1}\mathbf{S})(\mathbf{X}'^{f})^T \tag{11}$$

$$\mathbf{S} = \mathbf{H}\mathbf{X}'^{f} \tag{12}$$

$$\mathbf{F} = \mathbf{SS}^T + (N-1)\mathbf{R} \tag{13}$$

$$\mathbf{I} - \mathbf{S}^T\mathbf{F}^{-1}\mathbf{S} = \mathbf{TT}^T \tag{14}$$

In this study, the ETKF (Bishop et al., 2001) method is used to transform matrix $\mathbf{TT}^T$ as explained in (Vetra-Carvalho et al., 2018) by using the Sherman-Morrison-Woodbury identity (Golub & Loan, 1996), the scaled forecast ensemble observation perturbation matrix $\tilde{\mathbf{S}}$ (Livings, 2005) and Singular Value Decomposition (SVD)(Vetra-Carvalho et al., 2018). It is more efficient to inverse the observation error covariance matrix $\mathbf{R}$ since a diagonal matrix $\mathbf{R}$ is often a reasonable assumption. In data assimilation, the observation error covariance matrix $\mathbf{R}$ is assumed in diagonal because of the following reason. It is difficult to estimate a non-diagonal element of matrix $\mathbf{R}$, which is largely unknown. If we had a non-diagonal element, the process would be much slower and highly debatable on estimating and putting values in the non-diagonal element. In addition, it is a very challenging or expensive computation to include a non-diagonal component of matrix $\mathbf{R}$. Then, we simply used a robust method such as a diagonal matrix. Next, we substitute $\mathbf{F}$ in Eq. (13) to Eq. (14), then we used the Sherman-Morrison-Woodbury identity (Golub & Loan, 1996) as in Eq. (15) to obtain Eq. (16)

ACCEPTED MANUSCRIPT

$$\mathbf{A}\mathbf{B}^T(\mathbf{C}+\mathbf{B}\mathbf{A}\mathbf{B}^T)=(\mathbf{A}^{-1}+\mathbf{B}^T\mathbf{C}^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{C}^{-1} \tag{15}$$

with $\mathbf{A}=\mathbf{I}$, $\mathbf{B}=\mathbf{S}^T$, $\mathbf{C}=(N-1)\mathbf{R}$

$$\mathbf{T}\mathbf{T}^T=(\mathbf{I}+\frac{1}{N-1}\mathbf{S}^T\mathbf{R}^{-1}\mathbf{S})^{-1} \tag{16}$$

In ETKF, the innovation covariance matrix can be solved by computing an eigenvalue decomposition of the matrix $\mathbf{T}\mathbf{T}^T$ as in Eq. (16). However, as noted by Livings (2005), to avoid the floating point, rounding errors can produce an asymmetric matrix $\mathbf{T}\mathbf{T}^T$, in fact, Eq. (13) is symmetric. Hence, Livings (2005) introduced the scaled forecast observation ensemble perturbation matrix $\tilde{\mathbf{S}}$ as per Eq. (17)

$$\tilde{\mathbf{S}}=(\frac{1}{\sqrt{N-1}})\mathbf{R}^{-1/2}\mathbf{S} \tag{17}$$

we will obtain

$$\mathbf{T}\mathbf{T}^T=(\mathbf{I}+\tilde{\mathbf{S}}^T\tilde{\mathbf{S}})^{-1} \tag{18}$$

Now we can perform SVD to compute $\mathbf{T}\mathbf{T}^T$ efficiently. SVD is used to preserve accuracy (Livings, 2005) and it is a technique to decompose any size of matrix so that it can be processed more easily. SVD produces 3 matrices, namely two orthogonal matrices ($\mathbf{U}_T$ in size (m x m) and $\mathbf{V}_T^T$ in size (n x n)) and diagonal matrix $\mathbf{\Sigma}_T$ with size (m x n) with positive values. The last matrix of SVD contains a singular value according to its singular vector sequence. This singular value plays the biggest role in the variation of the data as a whole, and is stored in the first order of the diagonal matrix $\mathbf{\Sigma}_T$.

$$\tilde{\mathbf{S}}^T=\mathbf{U}_T\mathbf{\Sigma}_T\mathbf{V}_T^T \tag{19}$$

Next step, substitute Eq. (19) to Eq. (18), and because $\mathbf{U}$ and $\mathbf{V}_T$ are orthogonal matrices, hence $\mathbf{V}_T^T\mathbf{V}_T=\mathbf{I}$ and $\mathbf{U}_T\mathbf{U}_T^T=\mathbf{I}$, where $\mathbf{I}$ is Identity matrix, hence, $(\mathbf{U}_T\mathbf{U}_T^T)^{-1}=(\mathbf{U}_T\mathbf{U}_T^T)^T=\mathbf{U}_T\mathbf{U}_T^T$. Hence, we have $\mathbf{T}\mathbf{T}^T$ in another form as Eq. (20)

$$\mathbf{T}\mathbf{T}^T=\mathbf{U}_T(\mathbf{I}+\mathbf{\Sigma}_T\mathbf{\Sigma}_T^T)^{-1}\mathbf{U}_T^T \tag{20}$$

Returning to the ensemble analysis perturbation matrix in Eq. (11), we can replace Eq. (14) which is inside Eq. (11) by Eq. (20), so that we have Eq. (21). We can also take root in Eq. (21) becoming Eq. (22), so that we obtain

ACCEPTED MANUSCRIPT

$$\mathbf{X}'^a(\mathbf{X}'^a)^T = \mathbf{X}'^f\mathbf{U}_T(\mathbf{I}+\mathbf{\Sigma}_T\mathbf{\Sigma}_T^T)^{-1}\mathbf{U}_T^T(\mathbf{X}'^f)^T \tag{21}$$

$$\mathbf{X}'^a = \mathbf{X}'^f\mathbf{U}_T(\mathbf{I}+\mathbf{\Sigma}_T\mathbf{\Sigma}_T^T)^{-1/2}\mathbf{U}_T^T \tag{22}$$

Note that $\mathbf{U}_T^T$ is necessary at the end of the right hand side, so that the ensemble perturbation $\mathbf{X}'^a$ has a zero mean. After that, we can calculate the Kalman gain in Eq. (5) by using derivation of Eq. (12), Eq. (13), Eq. (14), the Sherman-Morrison-Woodbury Identity Eq. (15), Eq. (17) and Eq. (19), so that we obtain

$$\mathbf{K} = (\frac{1}{\sqrt{N-1}})\mathbf{X}'^f\mathbf{U}_T(\mathbf{I}+\mathbf{\Sigma}_T^T\mathbf{\Sigma}_T)^{-1}\mathbf{\Sigma}_T\mathbf{V}_T^T\mathbf{R}^{-1/2} \tag{23}$$

Hence, the updated ensemble mean found in Eq. (3) can be changed by substituting Eq. (23) into Eq. (3)

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + (\frac{1}{\sqrt{N-1}})\mathbf{X}'^f\mathbf{U}_T(\mathbf{I}+\mathbf{\Sigma}_T^T\mathbf{\Sigma}_T)^{-1}\Sigma_T\mathbf{V}_T^T\mathbf{R}^{-1/2}(\mathbf{y}-H(\bar{\mathbf{x}}^f)) \tag{24}$$

With regard to the derivation of ETKF, the needed input data are the forecast ensemble $\mathbf{X}^f$, the ensemble perturbation matrix in observation space $\mathbf{S} = (\mathbf{H}\mathbf{X}'^f)$, the observational error covariance matrices $\mathbf{R}$ and the observation $\mathbf{y}^o$. The analysis ensemble can, therefore, be computed by using Eq. (22) and Eq. (23), and Eq. (24).

The following are data processing stages of obtaining two outputs: the model radial velocity and the model analysis radial velocity. The model radial velocity is output without passing the ETKF process. Meanwhile, the model analysis radial velocity is output with passing the ETKF process. The first output only gets through step (i). The second output uses step (ii)-(xi). We would apply two running schemes, involving not only per HF site (first scheme) but also all HF sites simultaneously (the second scheme). The first scheme aims to find the optimal representativity error for each site. Whereas, the second scheme aims to produce the best estimation of surface currents using the optimal representativity error. Note that the second scheme is simulated after the first scheme. All outputs would be compared to the HF radial velocity to understand the impact of the blending process. The flow of processing can be seen in Fig. A.1 (see Appendix. A).

2.2.4 Post-Processing

The further procedure was a validation process against data used in the analysis (dependent validation) and against withheld data (independent validation) in detail as Table. A.1 (see Appendix. A). In this study, for instance, the dependent validation means that the blended model Anyer (BMA) compared to observations from Anyer itself or the blended model Labuan (BML) compared to observations from Labuan itself. Whereas, the independent validation means that the blended model Labuan for Anyer (BMLA) compared to observations from the Anyer site or the blended model Anyer for Labuan (BMAL) compared to observations from the Labuan site.

ACCEPTED MANUSCRIPT

The validation result was indicated by the root of the mean squared error (MSE) Eq. (25) (Murphy & Epstein, 1989). For every date, the RMS error of the model and the observations are computed (averaging over all coordinates); this time series of RMS errors are averaged over time and the result is the averaged RMS.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2 \tag{25}$$

f is forecast vector, o is observations vector. In this study, f is the blended model (the model analysis) or original model without blending, whereas the observations o are the radial velocity observations from each of the sites. The perfect score for this metric is 0 (which is only possible if the observations have no noise).

Notwithstanding, we examined the blended model (the analysis model) using two metrics, namely the relative error reduction (RER) Eq. (26), and the associated skill score (SS) Eq. (27) (Murphy & Epstein, 1989).

$$RER = \left(\frac{RMS_{originalCMEMSmodel} - RMS_{blendedmodel}}{RMS_{originalCMEMSmodel}}\right) \tag{26}$$

In this study, the RMS of the original CMEMS model was computed by the averaged RMS of the model without blending. The RMS blended model was computed by the averaged RMS of the blended model, such as independent validation and dependent validation. The perfect score for this metric is 1. The greater the reduction, the better the estimation. Our colour scheme is explained in the following result section.

$$SS = 1 - \left(\frac{MSE_{forecast}}{MSE_{ref}}\right) \tag{27}$$

In this study, $MSE_{forecast}$ was computed by square of the averaged RMS of the blended model, MSE_{ref} computed by square of the averaged RMS of the model without blending. The perfect score for this metric is 1, which means that the model would approach observations. Our colour scheme is explained in the following result section.

3 Result

We obtained the results from two schemes. The first scheme's outputs are illustrated in Fig. 4, Fig. 5, Fig. 6, Fig. 7. Meanwhile, the second scheme's outputs are in Table. 1, Fig. 8, Fig. 9e and Fig. 9f, Fig. 10c, and Fig. 12. Fig. 4a and Fig. 4b show the sensitivity of the averaged RMS relative to the representativity error ϵ_{rep} of the Anyer site and the Labuan site, respectively. For every date, the RMS error of the model and the observations are computed (averaging over all coordinates). This time series of RMS errors is averaged over time. The blue colour shows comparison between the original model and the observation. The red colour represents a dependent validation, that is, the blended model obtained from the observation itself. The green colour represents independent validation, that is, the blended model obtained from another site. In Fig. 4a the blue colour indicates the sensitivity of the averaged RMS of CMEMS for Anyer relative to the representativity error ϵ_{rep}. The green colour indicates the sensitivity of the averaged RMS of BMLA relative to the representativity error ϵ_{rep}. The red colour indicates the sensitivity of the averaged RMS of BMA relative to the representativity error ϵ_{rep}. While in Fig. 4b, the blue colour indicates the sensitivity of the

ACCEPTED MANUSCRIPT

averaged RMS of CMEMS for Labuan relative to the representativity error ϵ_{rep}. The green colour indicates the sensitivity of the averaged RMS of BMAL relative to the representativity error ϵ_{rep}. The red colour indicates the sensitivity of the averaged RMS of BML relative to the representativity error ϵ_{rep}. Both figures show the averaged RMS of the blended model (red colour and green colour) is smaller than the averaged RMS of the original model (blue colour). It means that the blended model which resulted from blending process is better than the original model. Notwithstanding, the averaged RMS of dependent validation (red colour) is decreasing, as the representativity error ϵ_{rep} is equal to zero. This validation ensures that the blending process is working properly and has been well-examined, because validation of the blended model which is obtained from its own observation should be the smallest error in the representativity error ϵ_{rep}, equal to zero, otherwise the larger the representativity error ϵ_{rep}, the worse the error becomes. One would have expected that the RMS of the red curve is the smallest as the representativity error ϵ_{rep} approaches zero. The small increase of this RMS error in the rounding errors is because the matrices involved in the blended model (the analysis) become ill-conditioned if the representativity error ϵ_{rep} approaches to zero. In theory, the optimal representativity error ϵ_{rep} of dependent validation would be achieved in the representativity error ϵ_{rep} equal to zero, otherwise, it becomes worse when the representativity error ϵ_{rep} equal to unlimited value, in which red line is approaching blue line. We displayed a red line in order to make sure that the dependent validation worked properly, and as well as theory. In Fig. 4, it was achieved by the representativity error ϵ_{rep} equal to nearly zero, namely 0.0603 m/s at the Anyer site and 0.0222 m/s at the Labuan site. Nevertheless, in general the dependent validation has been fulfilled. The representativity error ϵ_{rep} values were not zero because of rounding errors occurring when we inverted the representativity error ϵ_{rep} matrix. Meanwhile, the blended model using the observations of the other site (independent validation), such as BMAL and BMLA, gives higher the averaged RMS than the blended model using its own observations (dependent validation), such as BMA and BML, however, the independent validation's result is still smaller than that of the model without blending, such as CMEMS for Anyer and CMEMS for Labuan. It means that the blended model from independent validation is still considerable. By validating independently, we obtained that each site has own the representativity error ϵ_{rep} with the smallest for the averaged RMS. As a result, we have two representativity errors ϵ_{rep}, each of which have the smallest value for the averaged RMS from the independent validation.

ACCEPTED MANUSCRIPT

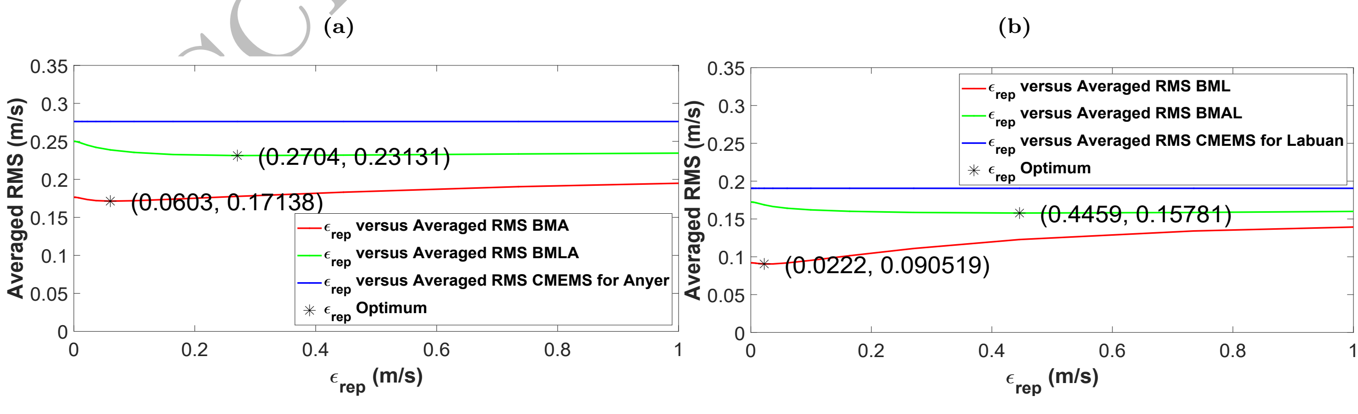

Fig. 4. The sensitivity of the averaged RMS relative to the representativity error ϵ_{rep} at the Anyer site (a) and the Labuan site (b)

To understand the variation of the RMS for each optimum representativity error ϵ_{rep} from the independent validation, we show Fig. 5. We compare the RMS signal of the model without blending, the RMS signal of the blended model from independent validation, and each observation site's standard deviation. The figures illustrated that the blended model's RMS signals are generally lower than the standard deviation signals. It means that the obtained blended models are acceptable.

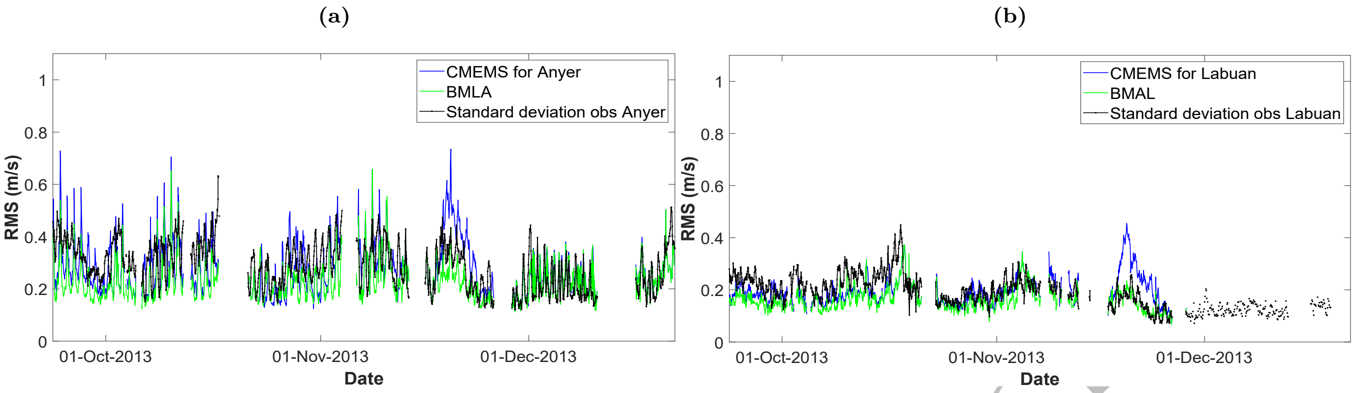

Fig. 5. Variation of RMS signal in time series (a) Anyer (b) Labuan

Moreover, we presented Fig. 6 and Fig. 7 to understand the differencies of the spatial structure of the RMS among the original model, the dependent validation, and the independent validation for each ϵ_{rep} for the Anyer site and the Labuan site. We took three different ϵ_{rep} for each site referring to the result of Fig. 4. They represent the optimum ϵ_{rep} for dependent validation, the optimum ϵ_{rep} for independent validation and the ϵ_{rep} closes to 1, respectively. The three representativity errors for the Anyer site are 0.0603 m/s, 0.2704 m/s, and 1 m/s. Meanwhile, the three representativity errors for the Labuan site are 0.0222 m/s, 0.4459 m/s, and 1 m/s. Overall, the significant error reduction is obtained by the spatial RMS of dependent validation. This result makes sense because the blended models were validated by their observation. Nevertheless, high errors are still happening in the narrow part about 0.25 m/s – 0.35 m/s either the spatial RMS in the Anyer site or the Labuan.

The best spatial RMS of the Anyer site using dependent validation is achieved by 0.0603 m/s of representativity error as in Fig. 6b. It has the most extensive area of the lowest error (0.05 m/s - 0.15 m/s), namely area in the north of Panaitan Island and the Java Sea. In comparison, the best spatial RMS of the Anyer site using independent validation is achieved by 0.2704 m/s of representativity error as in Fig. 6f. It has same signatures, however the errors are slight higher (0.1 m/s - 0.15 m/s) than those of dependent validation. The area which potentially contributes to the biggest error in the Anyer site is that the narrow channel.

The best spatial RMS of the Labuan site using dependent validation is achieved by 0.0222 m/s of representativity error as in Fig. 7b. It has the most extensive area of the lowest error (< 0.1 m/s), namely near the antenna. Besides, it is dominated by errors ranging 0.05 m/s - 0.15 m/s . Whereas, Fig. 7c and Fig. 7d still have big errors (0.2 m/s - 0.35 m/s) in the center and at the edge of the coverage. In comparison, the best spatial RMS of the Labuan site using independent validation is achieved by 0.4459 m/s of representativity error as in Fig. 7f. Almost similar to Fig. 7b, it has the most extensive area of the lowest error (0.05 m/s - 0.15 m/s), namely near the antenna. However, big errors such as > 0.2 m/s are still happening in the center, at the edge of the coverage and at the narrow channel. The areas that potentially contribute to the Labuan site's biggest error are in the narrow channel and in the north of Panaitan

ACCEPTED MANUSCRIPT

Island.

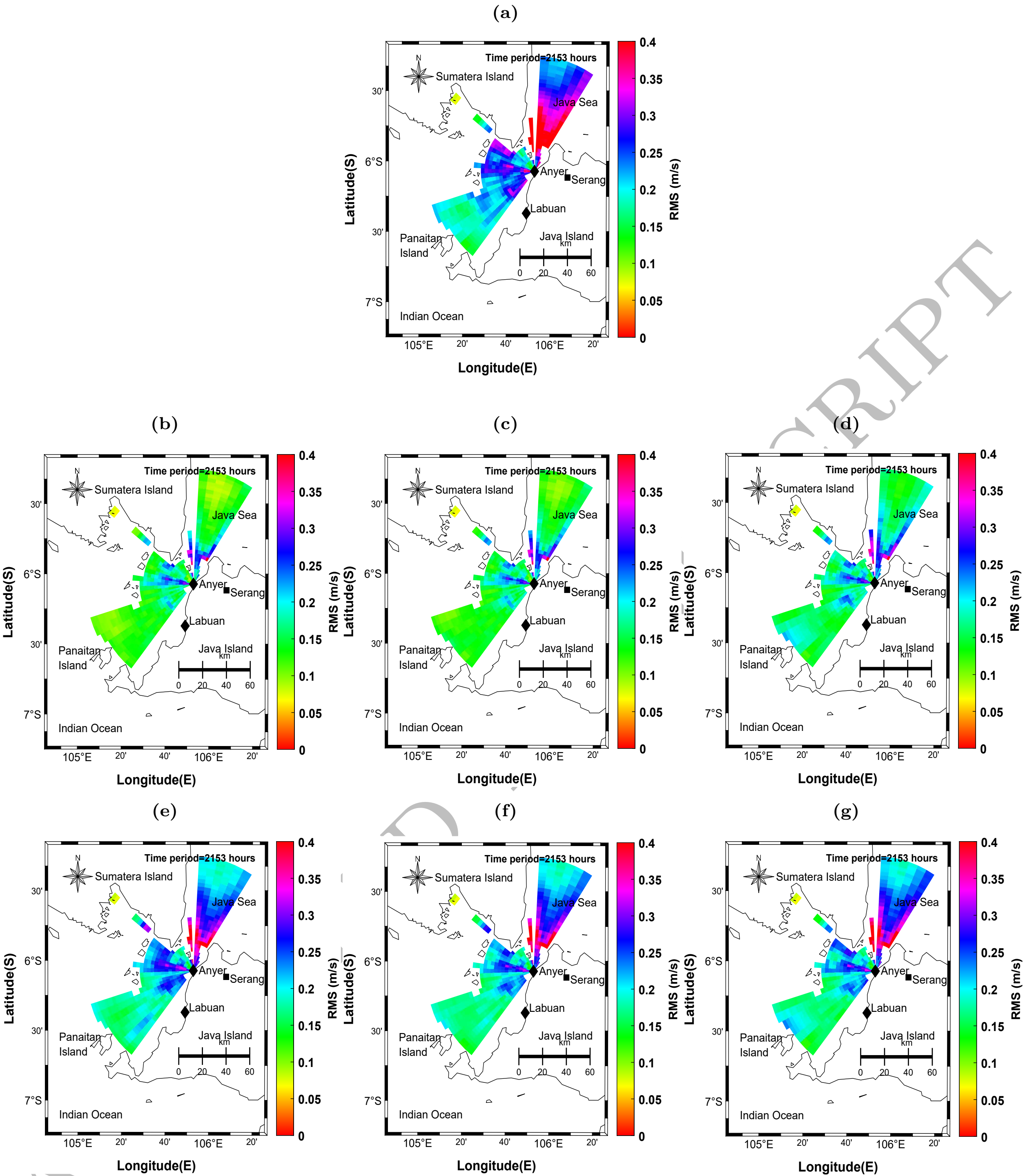

Fig. 6. Comparison of spatial rms - Anyer site 23 September - 22 December 2013. (a) CMEMS for Anyer. (b)(c)(d) are dependent validation outputs (BMA). (b) ϵ_{rep}=0.0603 m/s (c) ϵ_{rep}=0.2704 m/s (d) ϵ_{rep}=1 m/s (e)(f)(g) are independent validation outputs (BMLA). (e) ϵ_{rep}=0.0603 m/s (f) ϵ_{rep}=0.2704 m/s (g) ϵ_{rep}=1 m/s

ACCEPTED MANUSCRIPT

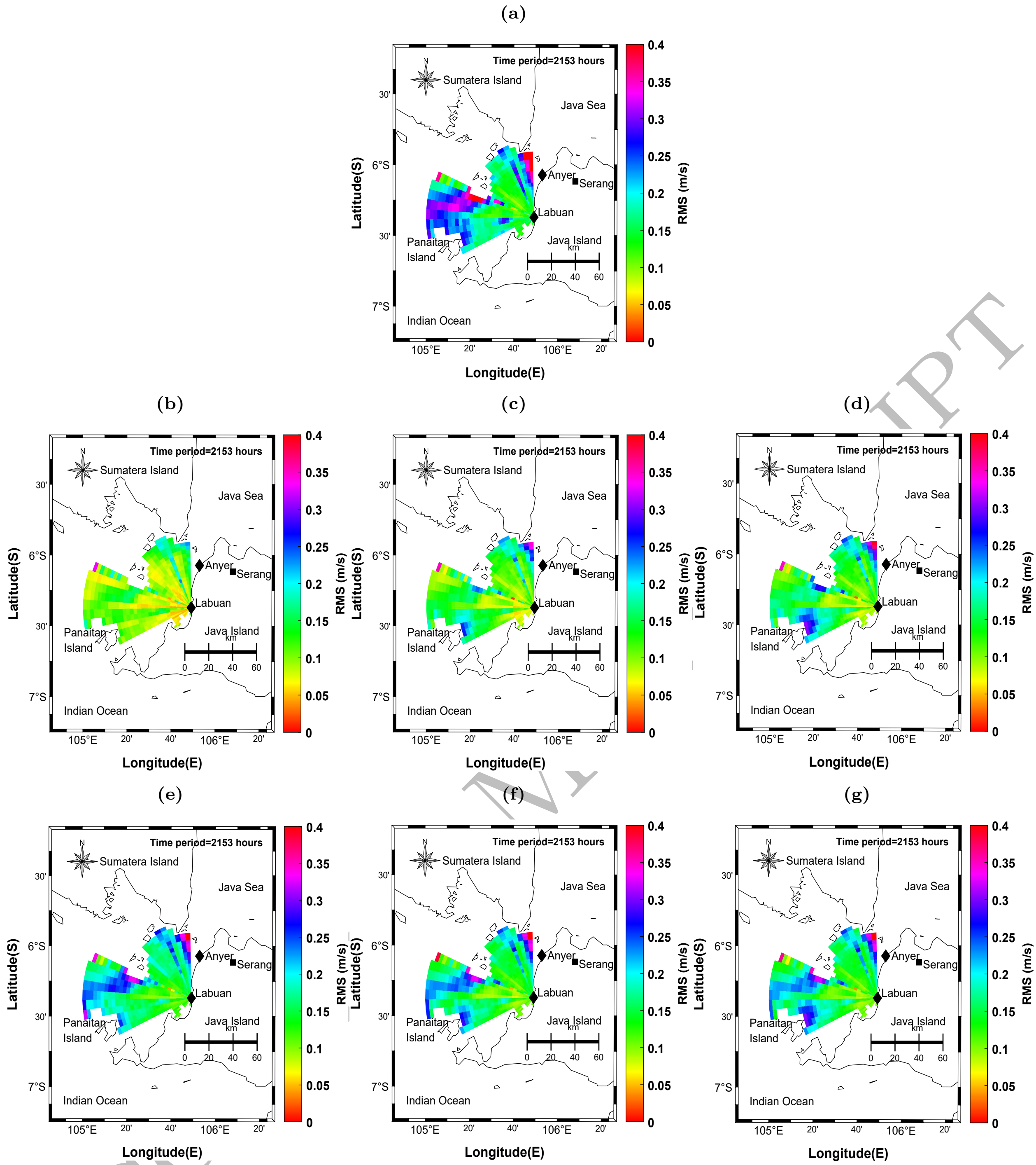

Fig. 7. Comparison of spatial rms - Labuan site 23 September - 22 December 2013. (a) CMEMS for Labuan. (b)(c)(d) are dependent validation outputs (BML). (b) ϵ_{rep}=0.0222 m/s (c) ϵ_{rep}=0.4459 m/s. (d) ϵ_{rep}=1 m/s. (e)(f)(g) are independent validation outputs (BMAL). (e) ϵ_{rep}=0.0222 m/s (f) ϵ_{rep}=0.4459 m/s (g) ϵ_{rep}=1 m/s

Referring to the first scheme result above, we still have 2 optimal values of the representativity error ϵ_{rep} from each site, namely 0.2704 m/s and 0.4459 m/s, respectively, which were from independent validation. It means that every site has an optimal representativity error ϵ_{rep}. We cannot avoid this reality. However, we tried to utilize them optimally to obtain the best possible blended model for all sites. So, we involved the observations from the two sites and the two representativity errors simultaneously in the second scheme. We added a column containing 0.2704 m/s of ϵ_{rep} into the Anyer site and treated them as a vector. Likewise, we added a column containing 0.4459 m/s of ϵ_{rep} into the Labuan site. Then, we concatenated the two vectors into one vector. The result of the second scheme is illustrated in Table. 1, and Fig. 8.

The averaged RMS shown in Table. 1 describes the comparison of the averaged RMS, the relative error reduction (RER), and the associative skill score (SS) for each of the second scheme case studies. One can see that the outputs after blending (BMALL, BMALL for Anyer, and BMALL for Labuan) are better than the output before blending (CMEMS). Besides, compared to the first scheme, it is clear that the second scheme significantly reduced error which is indicated by 0.1583 m/s from the averaged RMS of BMALL. The averaged RMS of CMEMS in the second scheme (0.2416 m/s) is between CMEMS for Anyer and CMEMS for Labuan in the first scheme (0.2761 m/s and 0.1905 m/s). Independent validation in Fig. 4a shows that the averaged RMS of BMLA at 0.2704 m/s of the representativity error ϵ_{rep} is 0.2313 m/s. Meanwhile, independent validation in Fig. 4b shows that the averaged RMS of BMAL at 0.4459 m/s of the representativity error ϵ_{rep} is 0.1578 m/s. Compared to these results, the averaged RMS of BMALL for Anyer (0.1775 m/s) is better than BMLA (0.2313 m/s). On the contrary, the averaged RMS of BMALL for Labuan got worse (0.1578 m/s to 0.2148 m/s). Nevertheless, in general, the second scheme improved the error.

Table 1: Comparison of the averaged RMS of the second scheme's output

The outputs of 2^{nd} scheme	The averaged RMS (m/s)	RER	SS
CMEMS	0.2416		
BMALL	0.1583	0.3448	0.5707
BMALL for Anyer	0.1775	0.2653	0.4602
BMALL for Labuan	0.2148	0.1109	0.2095

Fig. 8 shows fluctuations of the RMS signal of the second scheme in time series and the standard deviation. It shows that BMALL has a significantly reduced error. The maximum RMS signal of CMEMS is about 0.6577 m/s, while the maximum RMS signal of BMALL is about 0.3575 m/s.

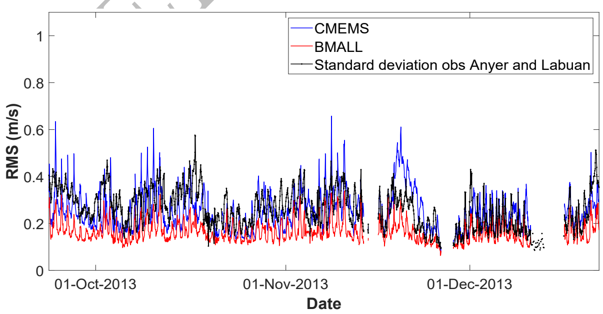

Fig. 8. Comparison of the RMS signal of the second scheme among CMEMS, BMALL and the standard deviation obs Anyer and Labuan

Once the best blended model (BMALL) was obtained, then we compare it against the observation and the original model (without blending). We used one of the sample dates, namely 20 November 2013 at 0100 UTC as in Fig. 9, which consists of observations radial velocity (Fig. 9a and Fig. 9b), radial velocity from the CMEMS (Fig. 9c and Fig. 9d) and radial velocity of the BMALL (Fig. 9e and Fig. 9f). The legend of the figure shows positive (red colour) and negative values (blue colour). A positive value means that radial velocity moves towards the HF radar site, while a negative value means that radial velocity moves away from HF radar site.

Radial velocity in Fig. 9c and Fig. 9d show a stronger velocity than the velocity in Fig. 9a and Fig. 9b. Generally,

ACCEPTED MANUSCRIPT

radial velocity in Fig. 9c and Fig. 9d were dominated by (-0.5 m/s) up to 0.5 m/s. Meanwhile, radial velocity in Fig. 9a and Fig. 9b were about (-0.3 m/s) up to 0.3 m/s. After blending, the CMEMS experiences a significant optimisation as shown in Fig. 9e and Fig. 9f. Those radial velocities show a similar distribution of values to radial velocity in Fig. 9a and Fig. 9b.

If the radial velocity from two sites is combined, we will obtain the total velocity. We use the same date as in Fig. 9. We can compare total velocity of the blended model in Fig. 10c against the observation total velocity (Fig. 10a) and the CMEMS (Fig. 10b). However, in this study the observation total velocity in Fig. 10a was taken directly from HF radar data. We did not combine all radial velocity sites (Fig. 9a and Fig. 9b). Note that the tides effect has been removed from it. Fig. 10a shows that currents were distributed only at the Sunda Strait. Weak currents of about 0 - 0.3 m/s were distributed at the eastern part and the northeastern part. Meanwhile, strong currents were located at the northern part of Panaitan Island with velocity at approximately > 0.5 m/s.

The HF radar coverage does not reach the Java Sea and the Indian Ocean, which is different to the original model with its currents being distributed over all areas with the strongest currents, which are elongated diagonally from the Java Sea until the Indian Ocean. We have characteristics of currents that are not only in the Sunda Strait, but also in the Java Sea and the Indian Ocean. However, the original model is only an estimation. After combining the radial velocity of the CMEMS and the observation, we have a new pattern of currents as in Fig. 10c. It shows all areas having values of currents. Nevertheless, currents speed have significantly decreased except in the northeastern part of the Sunda Strait near the Java Sea, which is > 0.4 m/s, while other areas are generally below 0 - 0.4 m/s, including the Indian Ocean, which is mostly following the speed of currents in the original model. The strong currents at the northern part of Panaitan island and at the centre of the Sunda Strait in the original model are weakening. This is happening not only because of the difference in the speed of the currents but also because of the direction effect of the currents. The direction of the currents near Panaitan Island in the observation move towards the north, whereas the direction of the currents in the original model, generally, moves towards the northeast. In general, all figures show the same direction of the currents, from the Indian Ocean towards the Java Sea. In addition, we involved a monthly mean from November 2013 as shown in Fig. 11, to analyze our blended model. The figure shows generally, that the speed of the currents in the north of Panaitan Island tends to be stronger than in other areas. Our blended model (BMALL) is similar to the monthly mean total speed from the observation.

ACCEPTED MANUSCRIPT

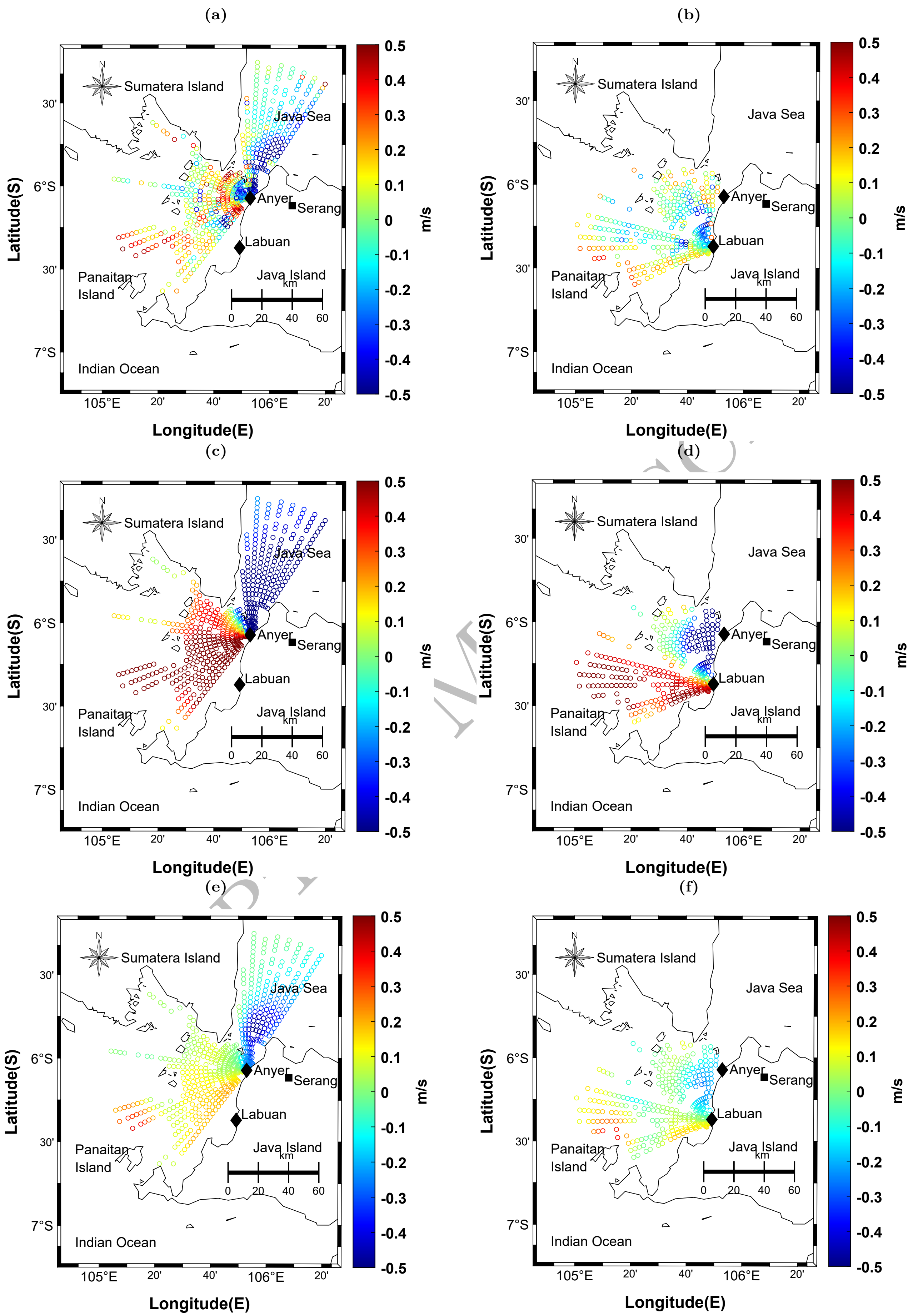

Fig. 9. Comparison of radial velocity all sites 20-Nov-2013 01:00 UTC. (a) Radial velocity observations Anyer (b) Radial velocity observations Labuan. (c) (d) CMEMS (e) BMALL for Anyer (f) BMALL for Labuan

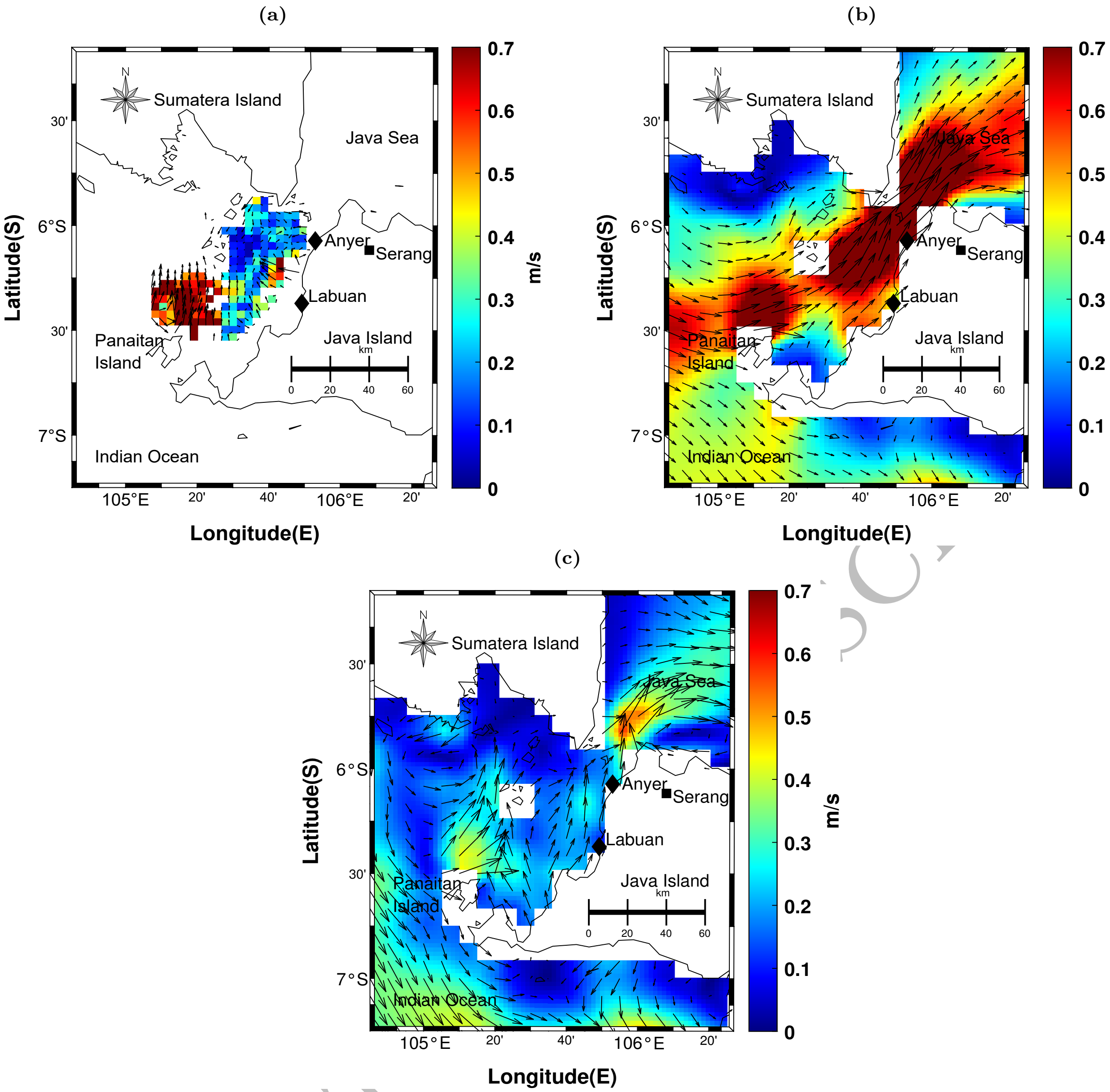

Fig. 10. Comparison of total velocity 20-Nov-2013 01:00 UTC. (a) Observations. (b) CMEMS (c) BMALL

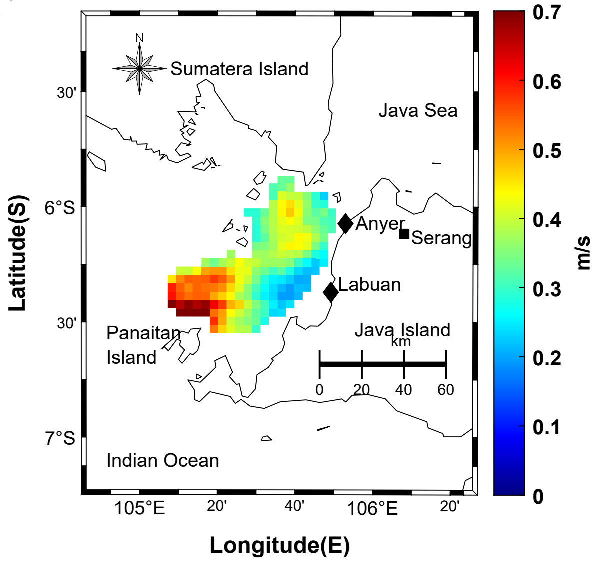

Fig. 11. Monthly mean total speed November 2013

ACCEPTED MANUSCRIPT

4 Discussion

We aim at analysing the effect of multiplication of R by 25 using formula Eq. (26) and Eq. (27) to obtain RER and SS of RMS of u and v from Table 3 page 277 of Stanev et al. (2015). The result shows RER for RMS of u and RMS of v are 0.043 m/s and 0.0178 m/s. While SS of RMS u and v are 0.085 m/s and 0.035 m/s, respectively. RER and SS of our study contain a value of one digit after the decimal point, while their result contained a value of two digits after the decimal point. It turned out that our proposed method of adding the representativity error to spatial quality and temporal quality as Eq. (A.3) (see Appendix. A) could give a better improvement.

Some questions remain regarding Fig. 8, which shows periodic fluctuation with one peak every day or a diurnal cycle. The other point is that they have a very high RMS value for the CMEMS (blue colour). Hence, we aimed at analyzing the RMS signal by comparing it to wind speed 10 meters from the meteorological station at Serang near the HF radar site. We did not compare our data to tides, because we have already removed the tides effect in an earlier step. Operationally, the winds observation is running 24 hours per day. There are zero value data each day because the winds are calm during the night until early morning.

To investigate the similarity between the RMS signal and the winds signal exactly, we applied Low Pass Filter (LPF) with 36 hours of period or frequency pass (1/36) (cycles/hour) and 60 dB of stopband attenuation for all radial data (September 23, 2013 09UTC – December 22, 2013 01UTC). We did not utilize the RMS signal directly. On the other hand, we use the mean radial velocity without tides (mean radial_detide filtered) because RMS in Fig. 8 is based on an averaged RMS over all grids at a particular time. Hence, we treat the same to radial velocity. We averaged radial velocity from all grids for a specific time, so we have mean velocity values every time. However, the data have some missing data that should be concerned in the LPF process. We used a moving median method with 24 hours of a time window to fill the missing values. Here, we only displayed LPF outputs from September 23, 2013 09UTC – November 06, 2013 to easily see the fluctuation in Fig. 12. It shows that some peaks of mean radial (red colour) follow the ridge of winds, for instance, 0 – 20th day (September – October 2013). Nevertheless, the rise is not precisely in the exact location with the winds signal. The radials signal appears after the winds signal. The other diurnal cycle is still happening such as 30th – 40th day even though it does not continue persistent with the previous cycle. Besides, we analyze the mean radial_detide filtered (LPF filtered output) for full data period (September 23, 2013 09UTC – December 22, 2013 01UTC) using a periodicity graph as Fig. 13a. It shows that there is still a diurnal cycle (7 cycles/week), although the magnitude is weak. This way shows us that the winds still positively impact ocean currents in the Sunda Strait. In addition, although the tides have been removed from the radial signal, it still shows a diurnal fluctuation. It makes sense because we use the T_Tide package, which contains a formula to remove specific frequencies of tides, which explicitly removes the variability of the signal at tidal frequencies but leaves any other frequencies in our dataset. We apply the periodicity also for analyzing the nature of winds as Fig. 13b. The winds signal is a diurnal cycle. A similar trend of winds effect to currents has been obtained by Oktavia et al. (2011), which concluded that geostrophic currents variation is indirectly influenced by winds in the Sunda Strait. Those authors calculated monthly averaged geostrophic currents from 4 tide-gauge stations (Ciwandan, Panjang, Tanjung Lesung and Kota Agung) and sea surface height of satellite for March 2008 – February 2009 using formula of the differences

ACCEPTED MANUSCRIPT

in sea level between two stations at a distance 1^o.

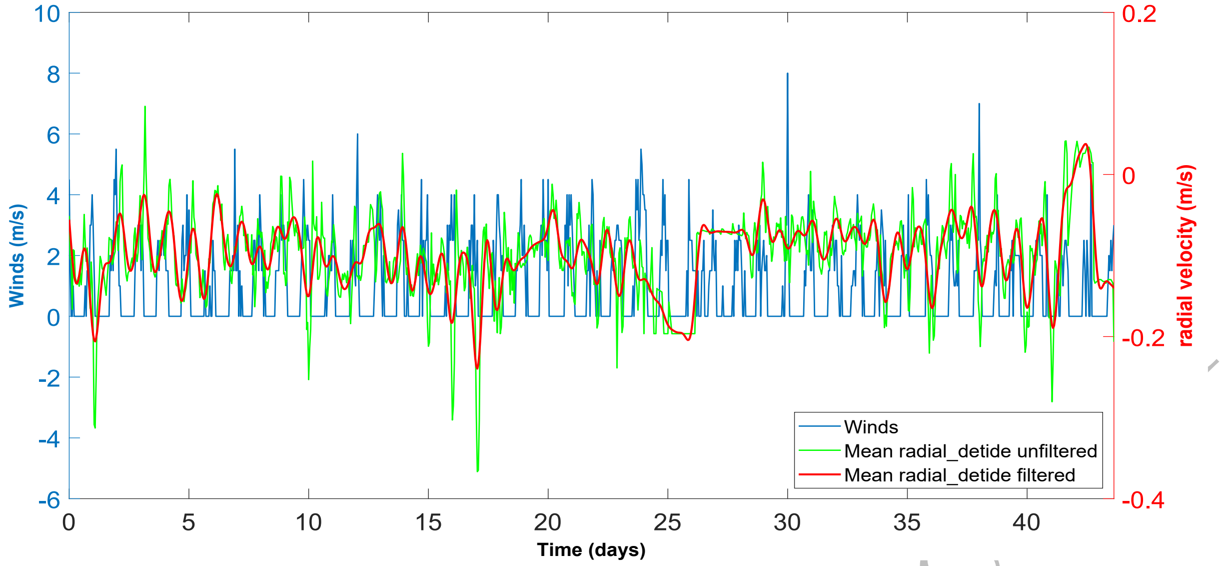

Fig. 12. Comparison of mean radial velocities and winds

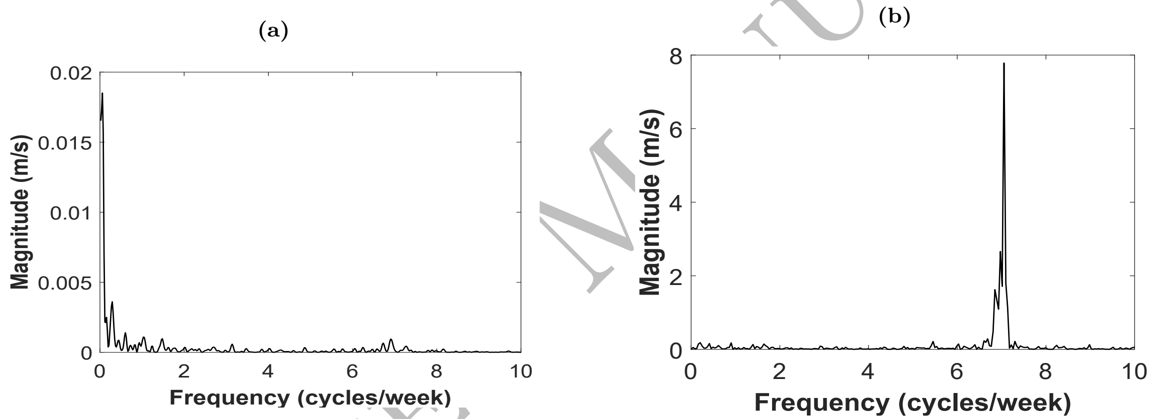

Fig. 13. Comparison of periodicity (a) Mean radial velocities of Anyer site (b) Winds

Theoretically, currents can be produced by winds at the ocean surface (wind-driven circulation), density differences (thermohaline circulation), and tides (NOAA, 2020). As we noted in the preprocessing section, we have already removed the tides effect from the radial velocity, hence the velocity could be due to the first two processes. Note that, HF radar can capture currents only at the surface until 2 meters of depth (Rubio et al., 2017). Considering the Hf radar capability, the produced currents are at the surface, which is predominantly affected by winds. In addition, the magnitudes of the winds are strengthened by the narrow channel. The effect of the wind can be explained in the following way. The Sunda Strait has two channels, one of which is narrow in the northeast near the Java Sea, the depth is shallow about 50 meters, and the other is wide in the southwest near the Indian Ocean. It makes sense if the Anyer signal is affected by wind because HF radar at the Anyer site is located near the narrow channel of the Sunda Strait. Typically, winds leaving a narrow channel have stronger wind than the surrounding environment. We conclude that the strong currents are due to the strong winds leaving the strait. Hence, in our case, the RMS signal obtained via the radial velocity calculation is strongly modulated by winds leaving the strait. The strong winds generated radial currents that affected the HF radar significantly, which appears in the BMALL (the model analysis) on 20 November

ACCEPTED MANUSCRIPT

2013 at 0100 UTC as can be seen in Fig. 10c. Based on this, the wind influences the magnitude and the frequency of high magnitudes of radial velocity.

5 Conclusion

Based on the results described in the previous section, we showed that blending HF radar reduced the error of the model. Another satisfying result occurred when we blended two sites separately and validated each of them through independent validation. The result shows independent validation giving a lower error than the model without blending, even though it is still higher than with dependent validation.

Dependent validation can be used for any various data with the condition the data is obtained from the other such as the blended model versus model or the blended model versus own observations. On the other hand, independent validation should use an independent real or independent actual data to prove whether a blending process is useful or not in reducing error. Independent validation would have the optimal representativity error ϵ_{rep} when the averaged RMS is the lowest. We used two sites separately, hence we have two optimal representativity errors ϵ_{rep} from each site, namely 0.2704 m/s and 0.4459 m/s, which means that every observation has its representativity error and contributes to form a model analysis.

Involvement of the two sites of observation and the two optimum representativity error ϵ_{rep} produced the best blended model which has the smallest of possible error and can be applied operationally.

Applying the value yields a completed spatial distribution of surface currents, which is the strongest in the narrow part and a lower currents in the surrounding area.

This study can also illustrate how HF radar data from a single site can be used to obtain total currents with the help of a model as long as the model has a realistic variability.

Assessing treatment of R shows that the addition of a representativity error to R could be another way to reduce the error rather than multiplication R by a specific value.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research could not have been accomplished without the research funding from the Indonesia Endowment Fund for Education (LPDP), Ministry of Finance of the Republic of Indonesia with contract number : PRJ-2378/LPDP.3/2016, HF radar and winds data from the Agency for Meteorology, Climatology, and Geophysics of the Republic of Indonesia (BMKG), and hardware, software and assistance facilities from the University of Liège. We are also grateful for the material and nonmaterial support from BMKG Regional Office III Denpasar, Bali. This study has been conducted using E.U. Copernicus Marine Service Information. We also thank the anonymous reviewers for their valuable feedback and comments.

Accepted Manuscript

Appendix A Flow of processing

(i) The model radial velocity is calculated. Zonal u and meridional v velocity components of the CMEMS model were interpolated based on radial velocity HF radar coordinates. They were then calculated altogether using the bearing of HF radial velocity θ to obtain the model radial velocity using Eq. (A.1). The outputs of the first scheme are the original model Anyer and the original model Labuan. Meanwhile, the output of the second scheme is the original model of all sites.

$$U = u(-\sin(\theta)) + v(-\cos(\theta)) \tag{A.1}$$

(ii) The model ensemble $\mathbf{X}$ is formed. The zonal u and meridional v velocity components of the CMEMS model from different time instances are assembled into the model ensemble $\mathbf{X}$. Grid cells corresponding to points on land are excluded from the state vector. Hence, the total number of ensemble members N is 8761. Theoretically, the more ensemble members there are, the more accurate the analysis ensemble will be.

(iii) The ensemble mean $\bar{\mathbf{X}}$ is calculated. It can be calculated by averaging all state vectors from the initial ensemble $\mathbf{X}$ over N ensemble members.

(iv) The ensemble perturbation $\mathbf{X}'$ is calculated. All ensemble members $\mathbf{X}$ are subtracted by the ensemble mean $\bar{\mathbf{X}}$.

(v) The forecast ensemble $\mathbf{X}^f$ is calculated. It can be calculated by summating the ensemble perturbation $\mathbf{X}'$ and member of model state $\mathbf{X}$ at time k.

(vi) The observation part of ensemble members $\mathbf{HX}^f$ is computed. It is using the forecast ensemble $\mathbf{X}^f$, which is interpolated based on HF radial velocity and calculated using Eq. (A.1).

(vii) In the next step, the forecast ensemble observation perturbation matrix $\mathbf{S}$ or $\mathbf{HX}'^f$ is calculated. Beforehand, the observation part of ensemble members $\mathbf{HX}^f$ from the previous step was averaged by N yielding $\mathbf{H}\bar{\mathbf{X}}^f$. The subtraction of the observation part of ensemble members $\mathbf{HX}^f$ by $\mathbf{H}\bar{\mathbf{X}}^f$ obtained the forecast ensemble observation perturbation matrix ($\mathbf{S}$).

(viii) The observation $\mathbf{y}_k^o$ every k time instance until N size of the ensemble was defined. These variable values are taken from radial velocity HF radar itself, but only those in the sea are selected.

(ix) The observation error covariance matrix $\mathbf{R}$ is determined. As is known, observations value $\mathbf{y}^o$ consists of real observation (that unknown exactly) and observation error (ϵ_o) as Eq. (A.2). Observation error comes from 3 sources, namely instrument noise, forward model error and representativity error.

$$\mathbf{y}^o = H(\mathbf{x}) + \epsilon_o \tag{A.2}$$

In this study, the observation error covariance matrix $\mathbf{R}$ was the sum of instrument error and representativity error, which was made in the form of a diagonal matrix. The related SeaSonde instrument has 4 ordered products,

ACCEPTED MANUSCRIPT

namely the radial velocity from spectra, the Short-Time Radials, the Final Output Radial, and the Total Vector. The Short-time Radials are a merged list of radial velocity from spectra, which are within the same range and bearing and in the same time interval 10 minutes (for a standard range type of Seasonde). The Final Output Radial is calculated from a merged collection of Short-time radials over 5 degree and the configured time. Total velocity is a combination of radial velocity from at least two sites of HF Radar. We used spatial quality $\mathbf{Q}_s$ and temporal quality $\mathbf{Q}_t$ from the Final Output Radial category to determine $\mathbf{R}$. Actually, the spatial quality and the temporal quality were from the Short-Time Radials, the second category. Spatial quality is the standard deviation of the radial velocity list in the Short-Time Radials (every 10 minutes) at the same grid and the same interval time. In comparison, temporal quality is the standard deviation of the radial velocity list at the same grid across the Short-Time Radials over one hour. Then, there are several merged Short-Time Radials to calculate the standard deviation. We can conclude that the Final Output Radial calculation originates from the Short-Time Radials over one hour. The spatial quality could be due to horizontal shear. Temporal quality could be due to the change of the current pattern over time (CODAR, 2009, 2013). Hence, the instrument error in this study was from standard deviation of radial velocity measurement for both spatial quality $\mathbf{Q}_s$ and temporal quality $\mathbf{Q}_t$. Representativity errors (ϵ_{rep}) have various values, which were tested between 0 and 1. Note that it is for the first scheme. In contrast, the second scheme uses the optimum ϵ_{rep} from the first scheme's result. The error in this context is associated with radial velocity in order that they have the same unit in m/s. In the present study, we apply a quadratic function for each component, so that the observational error covariance matrix $\mathbf{R}$ is the sum of squares of spatial quality $\mathbf{Q}_s$, temporal quality $\mathbf{Q}_t$, and representativity error ϵ_{rep} as Eq. (A.3).

$$\mathbf{R}_{ii} = \mathbf{Q}_{s\,i}^2 + \mathbf{Q}_{t\,i}^2 + \epsilon_{\text{rep}}^2 \tag{A.3}$$

Where i is an index of a grid cell, which has non-zero elements. $\mathbf{R}$ was transformed into a sparse array following length of matrix $\mathbf{Q}_s$ or $\mathbf{Q}_t$. The unit of $\mathbf{R}$ is m^2/s^2.

(x) The analysis ensemble $\mathbf{X}^a$ and the analysis ensemble mean $\bar{\mathbf{x}}^a$ are computed. After this, all variables are available, such as forecast ensemble $\mathbf{X}^f$, the ensemble perturbation matrix in observation space $\mathbf{S}$, observational error covariance matrix $\mathbf{R}$, and observations $\mathbf{y}^o$, then we can use these in the ETKF equations, such as Eq. (12), Eq. (17), Eq. (19), Eq. (22), Eq. (23), Eq. (24), $\mathbf{y}_k^o$ from step viii, Eq. (A.3), which are available in Sangoma package (Vetra-Carvalho et al., 2018).

(xi) Last but not least, the analysis ensemble is re-interpolated based on the position of the coordinates of radial velocity. The re-interpolation result is needed to validate the model analysis against the observation in the same grid. We chose the analysis ensemble mean $\bar{\mathbf{x}}^a$ for re-interpolating the model analysis radial velocity according to coordinates of radial velocity on each site. The term "the model analysis" refers to the definition of the best estimation resulted in time k. To maintain simplicity, we use the term "the blended model" to represent "the model analysis", as Fig. A.1. The outputs of this process were summarized in Table. A.1.

ACCEPTED MANUSCRIPT

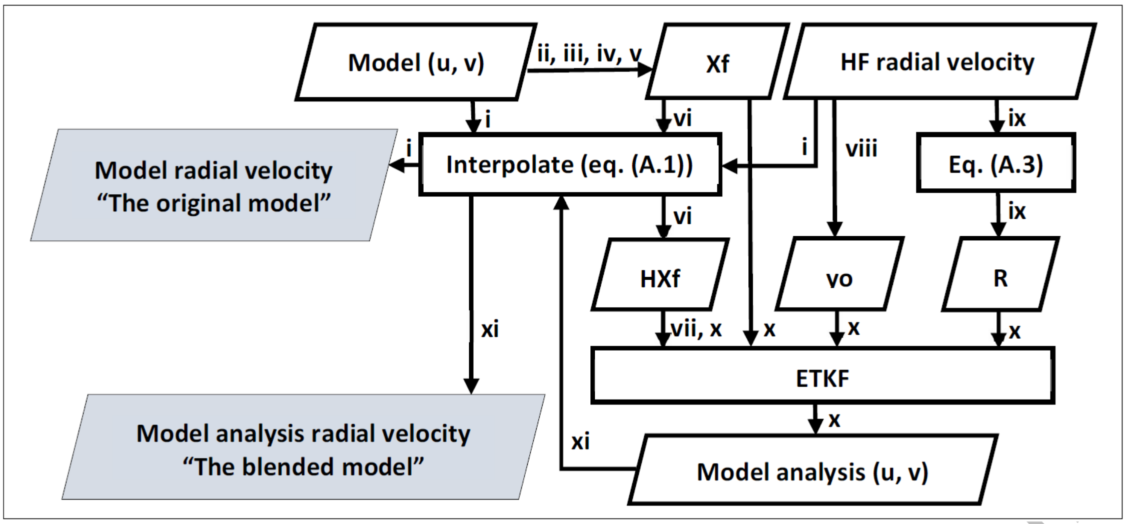

Fig. A.1. Flow chart of processing

Table A.1: Definition of the output term used

Scheme	Process type	Category of output	The output term used	Validation	Abbreviation
Scheme 1: observations per site	without ETKF	The model radial velocity	The original model Anyer		CMEMS for Anyer
			The original model Labuan		CMEMS for Labuan
	with ETKF, ϵ_{rep} =0-1	The model analysis radial velocity	The blended model Anyer	Dependent	BMA
			The blended model Anyer for Labuan	Independent	BMAL
			The blended model Labuan	Dependent	BML
			The blended model Labuan for Anyer	Independent	BMLA
Scheme 2: observations of all sites simultaneously	without ETKF	The model radial velocity	The original model of all sites		CMEMS
	with ETKF, the optimum ϵ_{rep} from scheme 1	The model analysis radial velocity	The blended model of all sites	Dependent	BMALL
			The blended model of all sites for Anyer	Dependent	BMALL for Anyer
			The blended model of all sites for Labuan	Dependent	BMALL for Labuan

References

Abascal, A. J., Castanedo, S., Fernandez, V., & Medina, R. 2012. Backtracking drifting objects using surface currents from high-frequency (HF) radar technology. *Ocean Dynamics*, *62*(7), 1073–1089. doi: 10.1007/s10236-012-0546-4

Amri, K., Priatna, A., & Suprapto. 2014. Oceanographycal Characteristic and Phytoplankton Abundance in Sunda Strait in East Monsoon. *Bawal*, *6(1) April*, 11–20.

Barth, A., Alvera-Azcárate, A., & Weisberg, R. H. 2008. Assimilation of high-frequency radar currents in a nested model of the West Florida Shelf. *Journal of Geophysical Research: Oceans*. doi: 10.1029/2007JC004585

Berta, M., Bellomo, L., Magaldi, M. G., Griffa, A., Molcard, A., Marmain, J., . . . Taillandier, V. 2014. Estimating Lagrangian transport blending drifters with HF radar data and models: Results from the TOSCA experiment in the Ligurian Current (North Western Mediterranean Sea). *Progress in Oceanography*, *128*, 15–29. doi: 10.1016/j.pocean.2014.08.004

BIG. 2017. Map of the Republic of Indonesia scale 1:5000.000, Edition 2017, *Badan Informasi Geospasial (BIG)*. (available online :https://www.big.go.id/uploads/content/Layanan%20Publik/Download/Peta%20NKRI/

ACCEPTED MANUSCRIPT

20190828_NKRI_Skala_5_juta_Versi_Inggris.pdf accessed on 11/07/2021)

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. 2001. Adaptive sampling with the ensemble transform Kalman filter. *Monthly Weather Rev.*, *129*(3), 420–436.

Brasseur, P., Ballabrera-Poy, J., & Verron, J. 1999. Assimilation of altimetric data in the mid-latitude oceans using the singular evolutive extended kalman filter with an eddy-resolving, primitive equation model. *Journal of Marine Systems*, *22*(4), 269-294.

Breivik, Ø., & Satra, Ø. 2001. Real time assimilation of HF radar currents into a coastal ocean model. *Journal of Marine Systems*, *28*(3-4), 161–182. doi: 10.1016/S0924-7963(01)00002-1

CODAR. 2009. CODAR's SeaSonde Radial Suite Release 6 Remote Operator's Manual.

CODAR. 2013. SeaSonde Current Vector Uncertainties The Start of QA/QC, CODAR News September 2013, available online at http://www.codar.com/news_09_4_2013.shtml.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistic. *Journal of Geophysical Research*, *99*(C5), 10,143-10,162.May 15,1994.

Golub, G. H., & Loan, C. F. V. 1996. *Matrix computation.* The Johns Hopkins University Press, 3rd edition.

Group, G. C. 2020. GEBCO 2020 Grid. doi:10.5285/a29c5465-b138-234de053-6c86abc040b9.

Heron, M., Gomez, R., Weber, B., Dzvonkovskaya, A., Helzel, T., Thomas, N., & Wyatt, L. 2016. Application of HF Radar in Hazard Management. *International Journal of Antennas and Propagation*, *2016*. doi: 10.1155/2016/4725407

Jumarang, M. I., & Ningsih, N. S. 2013. Transpor Volume Massa Air Di Selat Sunda Akibat Interaksi Enso, Monsun dan Dipole Mode. In *Semirata 2013 fmipa universitas lampung.* Retrieved from http://jurnal.fmipa.unila.ac.id/semirata/article/view/768/588

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering, Transactions of the ASME*, *82*(1), 35–45. doi: 10.1115/1.3662552

Kim, S. Y., Terrill, E. J., & Cornuelle, B. D. 2008. Mapping surface currents from HF radar radial velocity measurements using optimal interpolation. *Journal of Geophysical Research: Oceans*, *113*(10), 1–16. doi: 10.1029/2007JC004244

Kirincich, A. 2016. Remote sensing of the surface wind field over the coastal ocean via direct calibration of HF radar backscatter power. *Journal of Atmospheric and Oceanic Technology*, *33*(7), 1377–1392. doi: 10.1175/JTECH-D-15-0242.1

Kohut, J. T., Glenn, S. M., & Chant, R. J. 2004. Seasonal current variability on the New Jersey inner shelf. *Journal of Geophysical Research C: Oceans*, *109*(7), 1–16. doi: 10.1029/2003JC001963

Koropitan, A. F., Hadi, S., & Radjawane, I. M. 2006. Three-Dimensional Simulation of Tidal Current in Lampung Bay:Diagnostic Numerical Experiments. *Remote Sensing and Earth Sciences*, *3*(September). Retrieved from http://dx.doi.org/10.30536/j.ijreses.2006.v3.a1205

Lana, A., Marmain, J., Fernández, V., Tintoré, J., & Orfila, A. 2016. Wind influence on surface current variability in the Ibiza Channel from HF Radar. *Ocean Dynamics.* doi: 10.1007/s10236-016-0929-z

Lewis, J. K., Shulman, I., & Blumberg, A. F. 1998. Assimilation of Doppler radar current data into numerical ocean

ACCEPTED MANUSCRIPT

models. *Continental Shelf Research*, *18*(5), 541–559. doi: 10.1016/S0278-4343(98)00006-5

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. doi: 10.1016/j.jesp.2013.03.013

Li, S., Wei, Z., Susanto, R. D., Zhu, Y., Setiawan, A., Xu, T., ... Fang, G. 2018. Observations of intraseasonal variability in the Sunda Strait throughflow. *Journal of Oceanography*, *74*(5), 541–547. doi: 10.1007/s10872-018-0476-y

Lipa, B., Barrick, D., Diposaptono, S., Isaacson, J., Jena, B. K., Nyden, B., ... Kumar, T. S. 2012. High frequency (HF) radar detection of the weak 2012 Indonesian tsunamis. *Remote Sensing*. doi: 10.3390/rs4102944

Lipa, B., Barrick, D., Saitoh, S. I., Ishikawa, Y., Awaji, T., Largier, J., & Garfield, N. 2011. Japan tsunami current flows observed by HF radars on two continents. *Remote Sensing*, *3*(8), 1663–1679. doi: 10.3390/rs3081663

Lipa, B., Isaacson, J., Nyden, B., & Barrick, D. 2012. Tsunami arrival detection with high frequency (HF) radar. *Remote Sensing*, *4*(5), 1448–1461. doi: 10.3390/rs4051448

Lipphardt Jr, B. L., Kirwan Jr, A. D., Grosch, C. E., Lewis, J. K., & Paduan, J. D. 2000. Blending HF radar and model velocities in Monterey Bay through normal mode analysis. *Journal of Geophysical Research*, *105*(1999), 3425–3450.

Livings, D. 2005. *Aspects of the Ensemble Kalman Filter* ([Master's thesis]). University of Reading, UK.

Lorente, P., Piedracoba, S., Sotillo, M. G., & álvarez Fanjul, E. 2019. Long-term monitoring of the Atlantic jet through the strait of gibraltar with HF radar observations. *Journal of Marine Science and Engineering*, *7*(1), 1–16. doi: 10.3390/jmse7010003

Murphy, A. H., & Epstein, E. S. 1989. Skill Scores and Correlation Coefficients in Model Verificationf. *Monthly Weather Review*, *117*(March 1989), 572–581. doi: https://doi.org/10.1175/1520-0493(1989)117<0572:SSACCI>2.0.CO;2

NOAA. 2020. What's the difference between a tide and a current? National Ocean Service website, https://oceanservice.noaa.gov/facts/tidescurrents.html, 04/01/20.

Novico, F., Astawa, I. N., Sinaga, A., & Ali, A. 2015. Seafloor Morphology Influences on Current Condition in a Sunda Strait Bridge Project Using Numerical Model. *Bulletin of the Marine Geology*, *30*(2), 55–66. doi: 10.32693/bomg.30.2.2015.75

Oktavia, R., Pariwono, J. I., & Manurung, P. 2011, dec. Sea Level Variation and Geostrophic Current of The Sunda Strait Based on Tidal and Wind Data in Year 2008. *Jurnal Ilmu dan Teknologi Kelautan Tropis*, *3*(2), 127–152. doi: 10.29244/jitkt.v3i2.7827

Orasi, A., Picone, M., Drago, A., Capodici, F., Gauci, A., Nardone, G., ... Alonso-Martirena, A. 2018. HF radar for wind waves measurements in the Malta-Sicily Channel. *Measurement: Journal of the International Measurement Confederation*, *128*(January), 446–454. doi: 10.1016/j.measurement.2018.06.060

Paduan, J. D., Kim, K. C., Cook, M. S., & Chavez, F. P. 2006. Calibration and validation of direction-finding high-frequency radar ocean surface current observations. *IEEE Journal of Oceanic Engineering*, *31*(4), 862–875. (available online http://codar.com/images/about/2006_Paduan_JOE.pdf accessed on 17/06/2021)

ACCEPTED MANUSCRIPT

Paduan, J. D., & Shulman, I. 2004. HF radar data assimilation in the Monterey Bay area. *Journal of Geophysical Research C: Oceans*, *109*(7), 1–17. doi: 10.1029/2003JC001949

Paduan, J. D., & Washburn, L. 2012. High-Frequency Radar Observations of Ocean Surface Currents. *Annual Review of Marine Science*, *5*(1), 115–136. doi: 10.1146/annurev-marine-121211-172315

Pariwono, J. I. 1999. *Kondisi Oseanografi Perairan Pesisir Lampung* (Tech. Rep.). Proyek Pesisir Publication,Technical report (TE- 99/12-I).Coastal resources Center,University of Rhode Island.Jakarta, Indonesia.28 halaman.

Pawlowicz, R. 2020. "M_Map: A mapping package for MATLAB", version 1.4m, *Computer software*, available online at www.eoas.ubc.ca/~rich/map.html.

Pawlowicz, R., Beardsley, B., & Lentz, S. 2002. Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE. *Computers and Geosciences*, *28*(8), 929–937. doi: 10.1016/S0098-3004(02)00013-4

Potemra, J. T., Hacker, P. W., Melnichenko, O., & Maximenko, N. 2016. Satellite estimate of freshwater exchange between the Indonesian Seas and the Indian Ocean via the Sunda Strait. *Journal of Geophysical Research: Oceans*, *119*, 5098–5111. doi: 10.1002/2015JC011618

Rahmawitri, H., Atmadipoera, A. S., & Sukoraharjo, S. S. 2016. Circulation and Current Variability in The Sunda Strait Waters. *Jurnal Kelautan Nasional*, *11*(No.3 Desember 2016), 141–157. doi: 10.15578/jkn.v11i3.6115

Ren, L., Nash, S., & Hartnett, M. 2016. Forecasting of surface currents via correcting wind stress with assimilation of high-frequency radar data in a three-dimensional model. *Advances in Meteorology*. doi: 10.1155/2016/8950378

R.Hampel, F. 1974. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, *69*(346), 383-393. doi: 10.1080/01621459.1974.10482962

Rubio, A., Mader, J., Corgnati, L., Mantovani, C., Griffa, A., Novellino, A., ... Puillat, I. 2017. HF Radar activity in European coastal seas: Next steps toward a Pan-European HF Radar network. *Frontiers in Marine Science*, *4*(JAN), 1–20. doi: 10.3389/fmars.2017.00008

Sandro, R., Arnudin, Tussadiah, A., Utamy, R. M., Pridina, N., & Afifah, L. N. 2014. Study of wind, tidal wave and current potential in sunda strait as an alternative energy. In *Energy procedia* (Vol. 47, pp. 242–249). Elsevier B.V. doi: 10.1016/j.egypro.2014.01.220

Shulman, I., & Paduan, J. D. 2009. Assimilation of HF radar-derived radials and total currents in the Monterey Bay area. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *56*(3-5), 149–160. doi: 10.1016/j.dsr2.2008.08.004

Solabarrieta, L., Rubio, A., Castanedo, S., Medina, R., Charria, G., & Hernández, C. 2014. Surface water circulation patterns in the southeastern Bay of Biscay: New evidences from HF radar data. *Continental Shelf Research*, *74*, 60–76. doi: 10.1016/j.csr.2013.11.022

Soto-Navarro, J., Lorente, P., Álvarez Fanjul, E., Carlos Sánchez-Garrido, J., & García-Lafuente, J. 2016, mar. Surface circulation at the Strait of Gibraltar: A combined HF radar and high resolution model study. *Journal of Geophysical Research: Oceans*, *121*(3), 2016–2034. doi: 10.1002/2015JC011354

Stanev, E. V., Ziemer, F., Schulz-Stellenfleth, J., Gurgel, K.-W., Seemann, J., & Staneva, J. 2015. Blending Surface Currents from HF Radar Observations and Numerical Modeling: Tidal Hindcasts and Forecasts. *Journal of Atmospheric and Oceanic Technology*, *32*(2), 256–281.

ACCEPTED MANUSCRIPT

Susanto, D., Wei, Z., Adi, R., Zhang, Q., Fang, G., Fan, B., ... Setiawan, A. 2016. Oceanography Surrounding Krakatau Volcano in the Sunda Strait, Indonesia. *Oceanography*. doi: 10.5670/oceanog.2016.31

Tian, Y., Wen, B., Tan, J., & Li, Z. 2015. HF radar observation of tidal current over the Beibu Gulf, South China Sea. *Wuhan University Journal of Natural Sciences*, *20*(No.1), 055–058. doi: 10.1007/s11859-015-1058-0

Vandenbulcke, L., Beckers, J. M., & Barth, A. 2017. Correction of inertial oscillations by assimilation of HF radar data in a model of the Ligurian Sea. *Ocean Dynamics*, *67*(1), 117–135. doi: 10.1007/s10236-016-1012-5

Vetra-Carvalho, S., van Leeuwen, P. J., Nerger, L., Barth, A., Altaf, M. U., Brasseur, P., ... Beckers, J. M. 2018. State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems. *Tellus, Series A: Dynamic Meteorology and Oceanography*, *70*(1), 1–38. doi: 10.1080/16000870.2018.1445364

Wyrtki, K. 1961. *Physical Oceanography of the Southeast Asian waters* (Tech. Rep.). UC San Diego Naga Report Vol.2. The University of California Scripps Institution of Oceanography La Jolla, California.

Xu, J., Huang, J., Gao, S., & Cao, Y. 2014. Assimilation of high frequency radar data into a shelf sea circulation model. *Journal of Ocean University of China*. doi: 10.1007/s11802-014-2224-2

Yaremchuk, M., & Sentchev, A. 2009. Mapping radar-derived sea surface currents with a variational method. *Continental Shelf Research*, *29*(14), 1711–1722. doi: 10.1016/j.csr.2009.05.016

Yu, P., Kurapov, A. L., Egbert, G. D., Allen, J. S., & Kosro, P. M. 2012. Variational assimilation of HF radar surface currents in a coastal ocean model off Oregon. *Ocean Modelling*, *49-50*, 86–104. doi: 10.1016/j.ocemod.2012.03.001

ACCEPTED MANUSCRIPT