Variability of the Red River plume in the Gulf of Tonkin from stochastic modeling and cluster analysis

Duy Tung Nguyen

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Duy Tung NGUYEN

Le 11 mai 2022

Etude du panache du Fleuve Rouge dans le Golfe du Tonkin à partir d'une analyse en clusters et de simulations d'ensemble

---

Ecole doctorale : **SDU2E - Sciences de l'Univers, de l'Environnement et de l'Espace**

Spécialité : **Océan, Atmosphère, Climat**

Unité de recherche :
LEGOS - Laboratoire d'Etudes en Géophysique et Océanographie Spatiale

Thèse dirigée par
Pierre DE MEY et Thanh NGO DUC

Jury

M. Guillaume CHARRIA, Rapporteur
M. Vincent ECHEVIN, Rapporteur
Mme Joanna STANEVA, Examinatrice
M. Xavier DURRIEU DE MADRON, Examinateur
M. Pierre DE MEY, Directeur de thèse
M. Thanh NGO-DUC, Co-directeur de thèse
Mme Nadia AYOUB, Co-directrice de thèse
Mme Rosemary MORROW, Présidente

Abstract

This thesis aims at better understanding the variability of the Red River plume in the Gulf of Tonkin (GOT) in the mid and far field area using a numerical modeling approach. Understanding the plume variability and the fate of the delta waters is of primary importance for an in-depth knowledge and a better prediction capacity of the ocean circulation and hydrology in the GOT, for an improved management of coastal waters and monitoring of the coastal ecosystems.

In this thesis, the SYMPHONIE numerical model is configured with a high-resolution variable grid and realistic forcings relying on the configuration of V. Piton (2019). It is run over 7 years, from 2010 to 2016 with daily outputs. Then, the model is assessed using several observational data and shows its good skill in most of the comparisons. The model configuration and verification provide the foundation for two subsequent studies below.

In the first study, the model results over a 6-year (2011-2016) period are used to examine the daily to interannual variability of the Red River plume and of three rivers whose mouths are nearby. The plume is identified with the help of the simulated passive tracers. The results show that the temporal evolution of the surface coverage of the plume is correlated with the runoff variations, but other processes, such as winds and tides, are involved. Using a K-means unsupervised machine learning algorithm, the main patterns of the plume and their evolution in time are classified in four clusters, analyzed and linked to different environmental conditions. In winter, the plume is narrow and sticks along the coast most of the time due to the downcoast current and northeasterly wind. In early summer, the southwest monsoon wind makes the plume flow offshore. The plume reaches its highest coverage in September after the peak of runoff. Vertically, the plume thickness also shows seasonal variations. In winter, the plume is mixed over the whole water depth, while in summer, the plume can be detached both from the bottom and the coast. The plume can deepen offshore in summer, due to strong wind (in May, June) or specifically due to a recurrent eddy occurring near 19°N (in August). Finally, the receiving basins of the different river systems in the GOT are identified, providing a general picture of the regions where the transport of materials from the river to the ocean is expected to have the greatest impact, for example in case of anthropogenic chemical substances leaked to the river. This first part was published in 2021.

The clustering analysis above shows that whatever the cluster, the plume is strongly affected by the wind. Therefore, the robustness of these results to the wind fields used to force the ocean model needs to be assessed. In the second study, I use an ensemble of simulations to assess the model response to perturbations added to the wind forcing. The sensitivity of the simulation presented in the first part is statistically evaluated by calculating the spread and the distribution of the variables of interest from an ensemble of 50 members. Due to computing and memory constraints, this study is performed over a short period, from June to August 2015, corresponding to the high runoff season. Firstly, the error of the forcing wind is estimated by comparing it with a satellite product. Then, its impact onto the model is assessed for surface and subsurface variables. For sea surface height, the average uncertainty is higher in the northern area, while for the sea surface temperature and salinity, it is higher near the Vietnamese coast and the Red River delta. Vertically, the uncertainty is highest at the surface for salinity and at the sub surface for temperature. The sensitivity of the river plume is then analyzed. The

spread of the plume area is highest in August, which is the same time as when the plume area reaches its peak. The clustering analysis shows some cluster shifts between different members of the ensemble, but the cluster that is most likely to occur is still the one from the reference simulation (with unperturbed wind). These limited changes suggest that the results of the cluster analysis of the reference simulation in the first study are indeed robust to the wind forcing errors. The uncertainty in plume thickness varies over time, typically less than 2m and can peak at 4m (for a total thickness of 10m).

Résumé

Cette thèse vise à mieux comprendre la variabilité du panache du Fleuve Rouge dans le Golfe du Tonkin (GOT) dans la zone proche de l'embouchure et plus au large, en utilisant la modélisation numérique. Comprendre la variabilité du panache et le devenir des eaux du delta est d'une importance capitale pour une connaissance approfondie et une meilleure capacité de prédiction de la circulation océanique et de l'hydrologie dans le GOT, ainsi que pour une meilleure gestion des eaux côtières et surveillance des écosystèmes côtiers.

Dans cette thèse, une configuration du modèle SYMPHONIE est mise en place avec des forçages réalistes et une grille à haute-résolution variable, sur la base de la configuration de V. Piton (2019). Il est exploité sur 7 ans, de 2010 à 2016 avec des sorties quotidiennes. Ensuite, le modèle est évalué à l'aide de plusieurs données d'observation et montre sa bonne habileté dans la plupart des comparaisons. La configuration et la vérification du modèle constituent la base de deux études ultérieures ci-dessous.

Dans la première étude, les résultats du simulation sur une période de 6 ans (2011-2016) est réalisée pour étudier la variabilité journalière à interannuelle du panache du Fleuve Rouge et de trois rivières dont l'embouchure est voisine. Le panache est identifié à l'aide de traceurs passifs injectés dans la simulation. Les résultats montrent que l'évolution temporelle de la surface couverte par le panache est corrélée avec les variations du débit, mais que d'autres processus, tels que les vents et les marées, sont impliqués. En utilisant un algorithme d'apprentissage automatique non supervisé (K-means), les principaux régimes du panache et leur évolution dans le temps sont classifiés et analysés selon quatre clusters, puis liés à différentes conditions environnementales. En hiver, le panache est étroit et reste la plupart du temps le long de la côte en raison du courant côtier et du vent de nord-est. Au début de l'été, le vent de la mousson du sud-ouest fait s'écouler le panache vers le large. Le panache atteint sa plus grande couverture en septembre, après le pic du débit. Sur la verticale, l'épaisseur du panache montre également des variations saisonnières. En hiver, le panache est mélangé sur toute la colonne d'eau, alors qu'en été, le panache peut être détaché à la fois du fond et de la côte. Le panache peut s'approfondir au large en été, en raison de vents forts (en mai, juin) ou spécifiquement en raison d'un tourbillon récurrent se produisant près de 19°N (en août). Enfin, les bassins recevants des différents systèmes fluviaux dans le GOT sont identifiés, fournissant une image générale des régions où le transport de matériaux du fleuve vers l'océan sont susceptibles d'avoir le plus grand impact, par exemple dans le cas où des substances chimiques anthropogéniques seraient déversées dans le fleuve. Cette première partie a fait l'objet d'une publication en 2021.

L'analyse en clusters ci-dessus montre que, quel que soit le cluster, le panache est fortement affecté par le vent. Par conséquent, la robustesse de ces résultats aux champs de vent utilisés pour forcer le modèle océanique doit être évaluée. Dans la deuxième étude, j'utilise un ensemble de simulations pour évaluer la réponse du modèle aux perturbations ajoutées au vent forçant. La sensibilité de la simulation présentée dans la première partie est évaluée statistiquement en calculant la dispersion et la distribution des variables d'intérêt à partir d'un ensemble de 50 membres. En raison des contraintes de calcul et de mémoire, cette étude est réalisée sur une courte période, de juin à août 2015, correspondant à la saison de fort débit. Tout d'abord, l'erreur sur le vent forçant est estimée par comparaison avec un produit satellitaire.

Ensuite, son impact sur le modèle est évalué pour les variables de surface et de subsurface. Pour la hauteur de la surface de la mer, l'incertitude moyenne est la plus élevée dans le nord du bassin, tandis que pour la température et la salinité de surface, elle est plus élevée près de la côte vietnamienne et du delta du fleuve Rouge. Sur la verticale, l'incertitude est la plus forte à la surface pour la salinité et en sub-surface pour la température. J'analyse ensuite la sensibilité du panache de la rivière. La dispersion de la surface du panache est maximale en août, qui est aussi la période où la surface du panache est la plus grande. L'analyse en clusters montre quelques changements de clusters entre les différents membres de l'ensemble, mais le cluster le plus susceptible de se produire est toujours celui de la simulation de référence (avec le vent non perturbé). Ces changements limités suggèrent que les résultats de l'analyse clusters de la simulation de référence dans la première étude sont effectivement robustes aux erreurs de forçage du vent. L'incertitude sur l'épaisseur du panache varie dans le temps, elle est généralement inférieure à 2m et peut culminer à 4m (pour une épaisseur totale de 10m). Enfin, l'ensemble est vérifié en utilisant les jeux d'observations disponibles.

Acknowledgement

Doing a PhD is a special journey. My PhD is even more "special" because of COVID. The epidemic has affected us a lot from the beginning of 2020 (when I was a second-year student) until now (mid-2022). However, time has passed and we have learned that if we stand together, nothing can stop us from moving forward.

During this doctoral journey, I have received tremendous help and support from many people and organizations. First of all, I would like to express my deep gratitude to my supervisor, Pierre De Mey-Frémaux. This PhD could not be started without your help. You guided me from the beginning, you supported me when I applied for the fellowship, and you shaped my study.

I want to express my particular and deepest gratitude to my co-supervisor, Nadia Ayoub, who supported me day and night. She was the person to follow my every step along my PhD journey, not only at work but also in my social life. I will remember all the memories I had with you and your family. It was a pleasure for me to be your student.

I am extremely grateful to Thanh-Ngo Duc, my co-supervisor. Discussions with you about the atmospheric conditions helped me a lot to strengthen the method used in my study. Thank you for all your valuable advice not only during a year in Vietnam but also when I was in France.

I am also indebted to the other members of the research team. A big thank you to Patrick Marsaleix for your help in setting up the models and for all the problems related to the models. Also, you are the one who suggested me to try the passive tracer in the study. I am also grateful to Florence Toublanc for all the discussions throughout the thesis. Finally, thank you for introducing me to Python 5 years ago.

I could not have undertaken this journey without Sylvain Ouillon, who initiated the link between me and my supervisor. I believe that the connections you have made have proven to be strong and undeniable!

I would like to express my sincere thanks to the people on the thesis committee, Marine Herrmann and Anna Rubio. Although we only met once a year, I appreciated your advice and suggestions. I think your recommendations shaped the method I used in this study.

A big thank you to my friends in the lab, Violaine, Simon, Julia, Adélaide, Arne, Quentin, Micael, ... Thanks for the time we spent together. I will miss the international meal very much!

Words cannot express my gratitude to my Vietnamese friends in Toulouse: anh Việt, anh Đa, Bích, Hạnh, chị Huệ, anh Thái, chị Ngọc, Thái con. The party we had almost every week has colored my social life. I will remember the memories we have together. I am also grateful to the people of USTH: chị Trà, anh Sơn, anh Bình, anh Hiền, anh Thường, anh Tuấn, anh Long, anh Quân, anh Kiên admin. I spent an unforgettable year with you in Vietnam. Special thanks to my friends, who I could always count on when I was stressed: Tuấn, Đức Anh, Hải dớ, Thu vb, Hoàng híp.

The last word, I want to give it to my family. My mother and father didn’t have the chance to go to college. They started working before they were 18 and they have continued to work until now. They worked tirelessly so that I could have enough money to go to college. No words can express my gratitude for their sacrifices.

I would also like to thank the French Embassy in Vietnam and IRD (ARTS program) for funding this thesis. This work used the high performance computing resources of CALMIP (University of Toulouse III Paul Sabatier) and HILO (Hanoi University of Science and Technology). Thanks to the authors (Manh Cuong Tran, Florence Lyard, Sylvain Ouillon) who provided the observational data to validate the model.

Table of content

List of figures

List of tables

List of acronyms

GOT	Gulf of Tonkin
HF	High Frequency
NR	Northern Rivers
RR	Red River
SR	Southern Rivers
SSH	Sea Surface Height
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
Std	Standard deviation

CHAPTER I: INTRODUCTION

***

I.1. This work

The study of river plumes is of great interest to the coastal oceanographic community because of the significant impact that plumes can have not only on local hydrology near the mouth, but also on ocean dynamics and ecosystems across the shelf. They also carry natural and anthropogenic contaminants that enter the ocean along with river waters ; the monitoring of river plumes is thus essential for water quality management policies. All of these issues are both more critical and intensified by the fact that coastal areas near river mouths (through estuaries, deltas, lagoons) are generally highly populated: anthropogenic pressure is higher on inland and coastal waters, while water quality and ecosystem health issues impact a large population. This is particularly true in tropical and subtropical deltas where sea level rise, land subsidence, flooding, and salinity intrusion are among the major threats in the context of global change (e.g. Szabo et al., 2016).

The Gulf of Tonkin is bathed by the waters of the Red River, which reach the ocean after passing through a complex network of channels and distributaries forming the Red River delta. The apex of the delta is located about 50km upstream from the capital Ha Noi. The delta is home to a population density of about 1000 inhabitants/km^2, and represents an area of intense agricultural and industrial activities. Understanding the physical and biogeochemical functioning of the complex system constituted by the Red River delta and its plume is essential to better face the challenges posed by global change.

In this context, the main objective of this work is to provide a better understanding of the variability of the Red River (RR) plume in the Gulf of Tonkin (GOT) at different scales.

This thesis work takes place within the framework of the LOTUS project. LOTUS is a joint Vietnam-France project, which aims to better understand the coastal ocean dynamics under the influence of the atmospheric variability and of the delta. This is a crucial knowledge to evaluate natural hazards impacts, to address environmental issues and to set-up sustainable coastal water monitoring systems. The work done within LOTUS involves Vietnamese researchers from several universities and institutions (USTH, HUS, IMER, IO, …) as well as French researchers from IRD, CNRS and French universities. In the framework of the LOTUS project, several studies in the GOT have been conducted. In addition to my work, two other PhD studies deal with the ocean dynamics in the GOT. In the PhD thesis of V. Piton (2019) at LEGOS (Toulouse, France), a high-resolution configuration in the Gulf of Tonkin and downstream part of the Red River delta has been set-up to study tides with the 2D spectral TUGO model and basin-scale circulation with the 3D Symphonie model. The PhD thesis of C. Tran (ongoing) at LOG (Wimereux, France) aimed at better understanding the circulation in the Gulf of Tonkin from tidal to annual scale using the High Frequency (HF) radars installed near the coast.

This section will start with the definition of what a river plume is and the scientific and practical interests of their study. It continues with a description of the regional context of the Gulf of Tonkin and Red River delta, as well as an account of previous studies in the area. Finally, we

announce what will be our objectives and approaches in this work, and the outline of this manuscript.

I.2 River plume

I.2.1 What is a river plume?

A river plume can be defined in a general way as the region of the coastal ocean where its properties and dynamics are affected by the river runoff (Horner-Devine et al. 2015). Though the river runoff is small compared to the whole ocean water volume, it can impact both the physics and biogeochemistry of the coastal ocean depending on the discharge, the properties of the ocean area (bathymetry, bottom roughness) and external forcing (air-sea fluxes, open ocean influence).

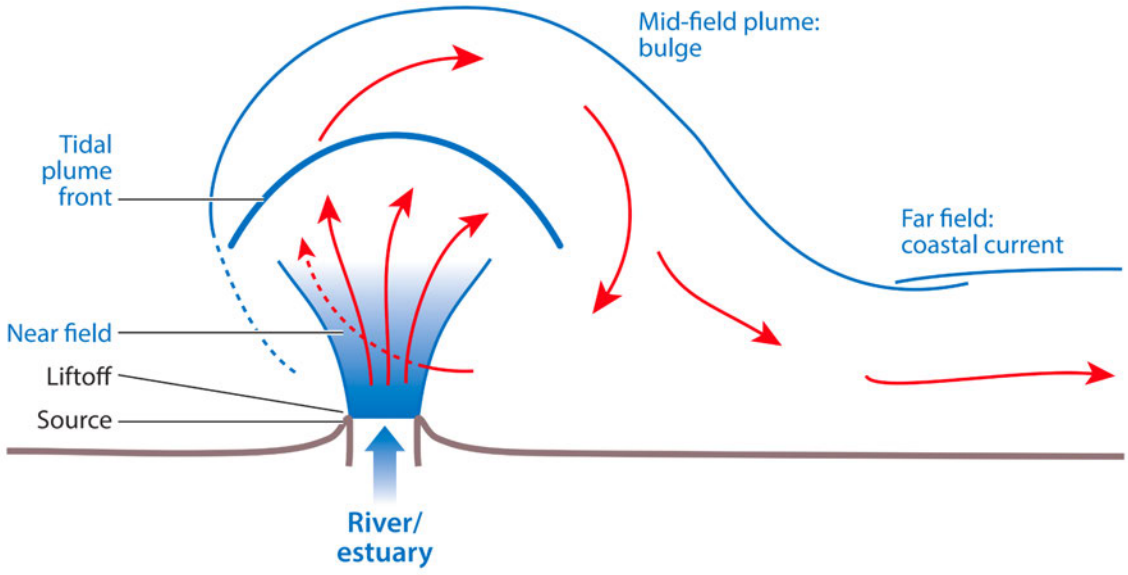

Figure I.1 Concept of a prototypical plume with all the dynamical regions, from Horner-Devine et al. (2015).

The dynamics of a prototypical plume has been studied by several authors (Fong & Geyer. 2002, Hetland. 2005, Horner-Devine et al. 2006). In a simplified coastal numerical simulation, with relatively high discharge and narrow river mouth, Hetland (2005, 2010) divides the prototypical plume into two parts: the near field where the outflow is critical and the advection and shear mixing dominate the dynamics; and the far field where the earth's rotation, wind, and background flow dominate the evolution of the plume. Horner-Devine et al. (2009), in their conceptual model, and Horner-Devine et al. (2015), in their review, describe the Columbia river in terms of four water masses: source water, the near field, the mid-field, and the far field (**Figure I.1**). The source water is the area before the liftoff (the point where the buoyant outflow detaches from the bottom). The near field starts from the liftoff and forms the buoyant layer. In this area, the water flows like a jet. Therefore, the limit of this area is where the supercritical flow ends. Then, in the mid-field, the plume is affected by both the momentum of the river flow and the effect of Earth rotation. Under low wind conditions, it can form a recirculating bulge region. Finally, when the plume loses the momentum of the initial runoff, and is driven mostly

by buoyancy, earth's rotation, wind and ambient currents, but is still distinguished with the ambient water, it is the far field plume.

Depending on the runoff, the characteristics of the river mouths, combined with the atmospheric and oceanic conditions, the structures of the river plume are different. Not every single river plume has all the features of the prototypical plume as described above. Garvine (1995) suggests a classification of the river plume, based on Kelvin number, equal to the ratio of the width of the river mouth to the Rossby radius of deformation. Plumes with high Kelvin number (>>1) are considered large scale plumes, while small scale plumes are characterized by small (<<1) Kelvin numbers. Huq (2009) shows that the bulge is formed if the estuary is narrow compared to the Rossby radius (i.e Kelvin number < 1).

Horner-Devine et al. (2015) divide the plume types into 6 general categories. The prototypical plume (A), as described above, is characterized by a large runoff, narrow river mouth, and its dynamics is affected sharply by Earth's rotation. The non rotational plume (B), in contrast, is less impacted by Earth's rotation due to small runoff or the river mouth is located near the Equator. Wide estuary plume (C), as its name implies, has a wide mouth and small runoff. Inside the river, the runoff flows in one side of the river due to Earth's rotation. The fourth type, angled inflow plume (D), has a river with a small angle to the coast. In delta plume (E), the runoff flows through several river mouths and has a joint plume from different mouths. Finally, region of freshwater influence (ROFI), appears in shallow marginal seas ; the plume has a strong interaction with the bottom.

I.2.2 Why do we need to study the river plume?

Rivers are vectors of exchange between the continent and the ocean. They bring freshwater that impacts the buoyancy of the coastal waters and therefore the coastal dynamics; the rivers impact is sometimes very local, sometimes at the basin scale (as for example in the case of the Amazon). Rivers carry matter and sediments that influence the coastal morphology (bathymetry, coastline). They also carry chemical species either in dissolved or particulate forms (for instance in sediments); some of them are natural while others have an anthropogenic origin from agriculture and industrial activities. Both influence the water quality and the ecosystems; for instance loadings of nutrients (nitrogen, phosphorus …) mainly stemming from watersheds with intensive farming can lead to algal bloom. At last, river flows bring along other kind of contaminants such as plastic debris.

Therefore, there is a need to better understand the fate of the river water from the estuaries to the ocean, especially in the mid-field and far field. Using a fuzzy c-mean clustering, Delaval et al. (2021) classify wind and Rhone River discharge into 6 scenarios. From that, simulations have been done in each scenario to identify the plume pattern in different forcing conditions to predict the main Rhone River plume patterns and consequences in case of radionuclides accidental releases. Using numerical simulation, Menesguen et al. (2018) identify the source of eutrophication in the Bay of Biscay and the English Channel and design a strategy of nutrient load reduction within European institutions directives framework.

As shown above, understanding the variability of the river plume in different scales is the first step toward the study of the dispersion of the possible contaminations and toward the design of

strategies for monitoring and managing the water quality and the health of ecosystems. These are particularly crucial issues in densely populated areas, such as many deltaic regions of Southeast Asia, including the Gulf of Tonkin.

I.3 The Gulf of Tonkin and the Red River

The Gulf of Tonkin (hereafter GOT) (16.9°N – 21.9°N, 105.6°E – 110°E) is a small shelf sea located east of Vietnam and south of China **(Figure I.2A)**. It is a shallow area with depth less than 100m **(Figure I.2B)**. It connects with the open ocean through the Hainan strait (with depth approximately 40 - 50m) and south of Hainan (with depth approximately 80 - 100m).

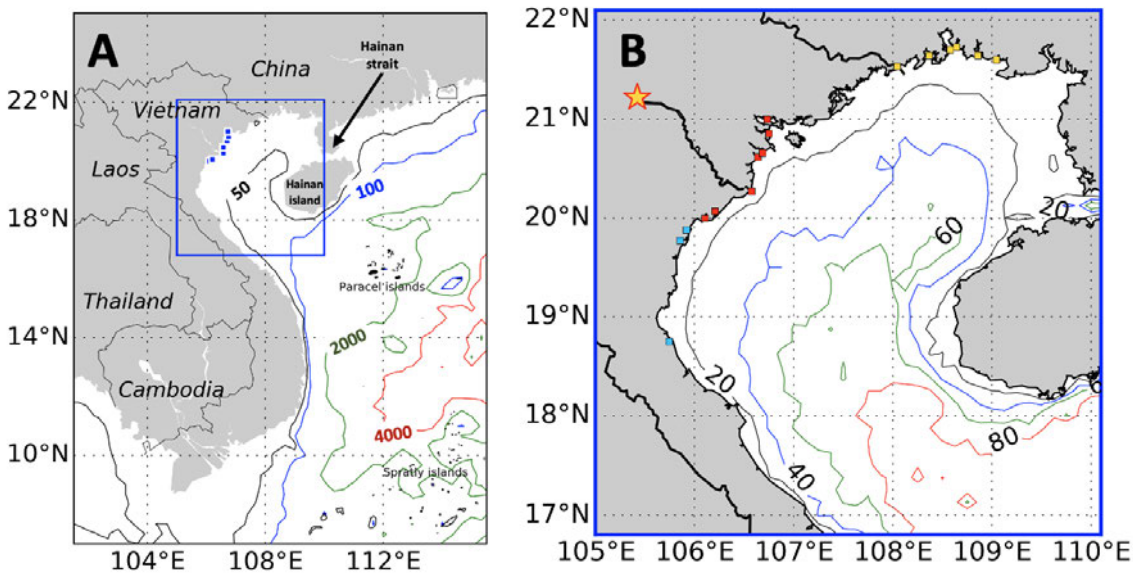

Figure I.2 **(A)** Location of the Gulf of Tonkin; the blue dots indicate the mouths of the Red River delta. Contour lines indicate the bathymetry of the area. **(B)** Same as A, but zoom to the blue box. The star indicates the location of Son Tay hydrological station at the apex of the delta, while the square boxes show the river mouths of the Red River (red square) and of other rivers to the south (light blue) and to the north (yellow).

I.3.1 Atmospheric conditions

The wind conditions in GOT depend on the monsoon. In the winter (October to February), the wind direction is from the northeast, with the velocity from 4 - 8m/s (**Figure I.3**). The wind is strongest in December, with the maximum value of 8m/s in the center of the gulf. The summer monsoon is shorter, from June to August, with less intense wind (2 - 6m/s). Between the two monsoons is the transition period with easterly wind. The wind is weakest in September, with velocity less than 2 m/s over the whole gulf. Due to the monsoon, the river plume undergoes a seasonal variation which will be presented in Chapter 4.

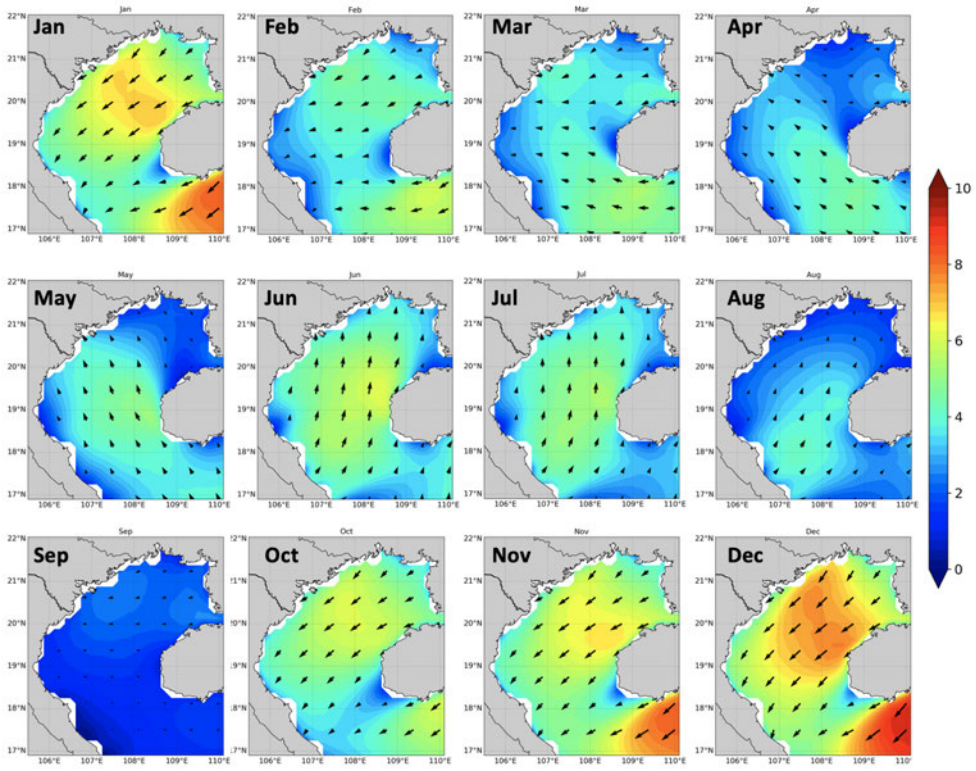

Figure I.3 Monthly averages of wind direction (arrow) and velocity (m/s, in colors) in GOT, calculated from ECMWF analysis in the period from 2009 to 2018.

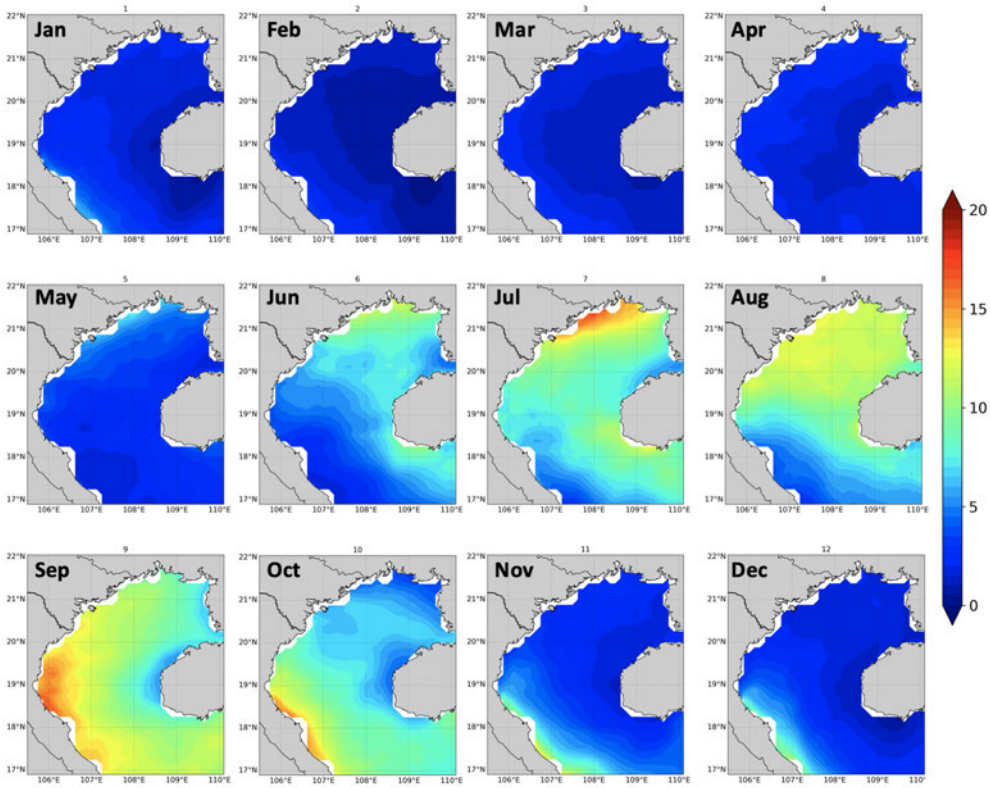

Figure I.4 Daily averaged precipitation in different months, calculated from ECMWF analysis in the period from 2009 to 2018. Units are mm/day.

The monthly mean precipitation in the GOT is shown in **Figure I.4**. From January to May, it is less than 5mm/day, with higher values located near the Vietnamese coast. Then, from June to August, the precipitation pattern changes with the highest values located in the north of the gulf. Over most of the area, the precipitation ranges from 5 - 10mm/day, with a peak of around 15mm/day in July. In September, the pattern changes again with the maximum value located near the Vietnamese coast. The precipitation is mostly higher than 10mm/day, with the local maximum of more than 15 mm/day from 18 to 20N. The high precipitation from July to September is expected to significantly impact the sea surface salinity, with a possible ambiguity when trying to identify the low salinity water of the river plume as discussed in Chapter 4. From October, the precipitation decreases again. In December, it is less than 5 mm/day with a local maximum of 8 mm/day near the coast between 17 - 19N.

I.3.2 Circulation

The ocean circulation of the Gulf of Tonkin (hereafter GOT) has been examined by several authors (e.g. Ding et al., 2013; Gao et al., 2014; Rogowski et al., 2019; for a recent review see Piton et al., 2021). Using model results and observational data, they all agree that the winter circulation at the gulf scale is cyclonic and driven by wind. However, in summer, the circulation is still a hot debate. Based on model results and 10-m current measurements, Ding et al. (2013) claimed that the flow in the northern Gulf of Tonkin features a cyclonic circulation. By conducting different numerical experiments, they also showed that this cyclonic circulation is induced by tide and can be accelerated by the wind. Gao et al. (2013) showed that the current in the southern gulf is anti-cyclonic and driven by the general South China Sea circulation. Using a 3D numerical model, HF radars and in situ data, Rogowski et al. (2019) showed that the current pattern is complex in the north of the gulf. Meanwhile, there is a strong surface water flow from the gulf to the South of Hainan island in June and strong Southward Vietnamese coastal flow in September in the south. From high-resolution model and pluri-annual simulations, Piton et al. (2021) found a basin-scale anticyclonic circulation in summer. They also suggest that the surface circulation is mainly ageostrophic, as a consequence of the monsoon wind forcing, except along the Vietnamese coast where the southward coastal current has a dominant density-driven component.

The flow through Hainan Strait during the summer monsoon is also discussed in several papers but is still an open question. Shi et al. (2002) and Chen et al. (2019) share the same opinion that the transport through Hainan Strait is inflow (or westward or flow into the Gulf) while Gao et al. (2013) showed that the mean current flows out of the gulf in the late spring and summer and into the gulf in other seasons. They also noted that the daily variations of the current are very strong so that the daily flow direction could vary independently of the season. Zhu et al. (2014) showed that the residual current is inflow in spring 2013 and controlled by tidal rectification. The link between the flow direction in Hainan Strait and the circulation in GOT is studied by Wu et al. (2008). They showed that in the summer, if the flow through Hainan Strait is inflow, the circulation is mainly cyclonic and vice versa.

I.3.3 Tides

Tides in the South China Sea in general and GOT in particular have been studied for many years. Wyrtki (1961) shows that 4 main constituents (K1, O1, M2, S2) are enough to explain

the tidal signal in the South China Sea. Until recently, there were only a few studies about the tidal characteristics in the South China Sea in general (Cai et al. 2006, Fang et al. 1999) and in the Hainan Strait (Chen et al., 2009). More recently, using a high resolution model with a horizontal resolution of 4.5 km and satellite and in situ observations, the tidal dynamics in the GOT has been examined by Nguyen et al. (2014). The result shows that the diurnal constituents (K1 and O1) are predominant in the GOT. The largest amplitudes are located in the north of the gulf, with amplitudes of more than 90 and 100 cm, respectively (**Figure I.5**). South of 18N, the amplitude decreases to under 30 cm for both diurnal constituents. M2 and S2 have similar patterns, with higher valuse in the west, along the coast of Vietnam, and the northeast coast of China. While the maximum value of M2 is 70 cm, the maximum value of S2 is 20 cm.

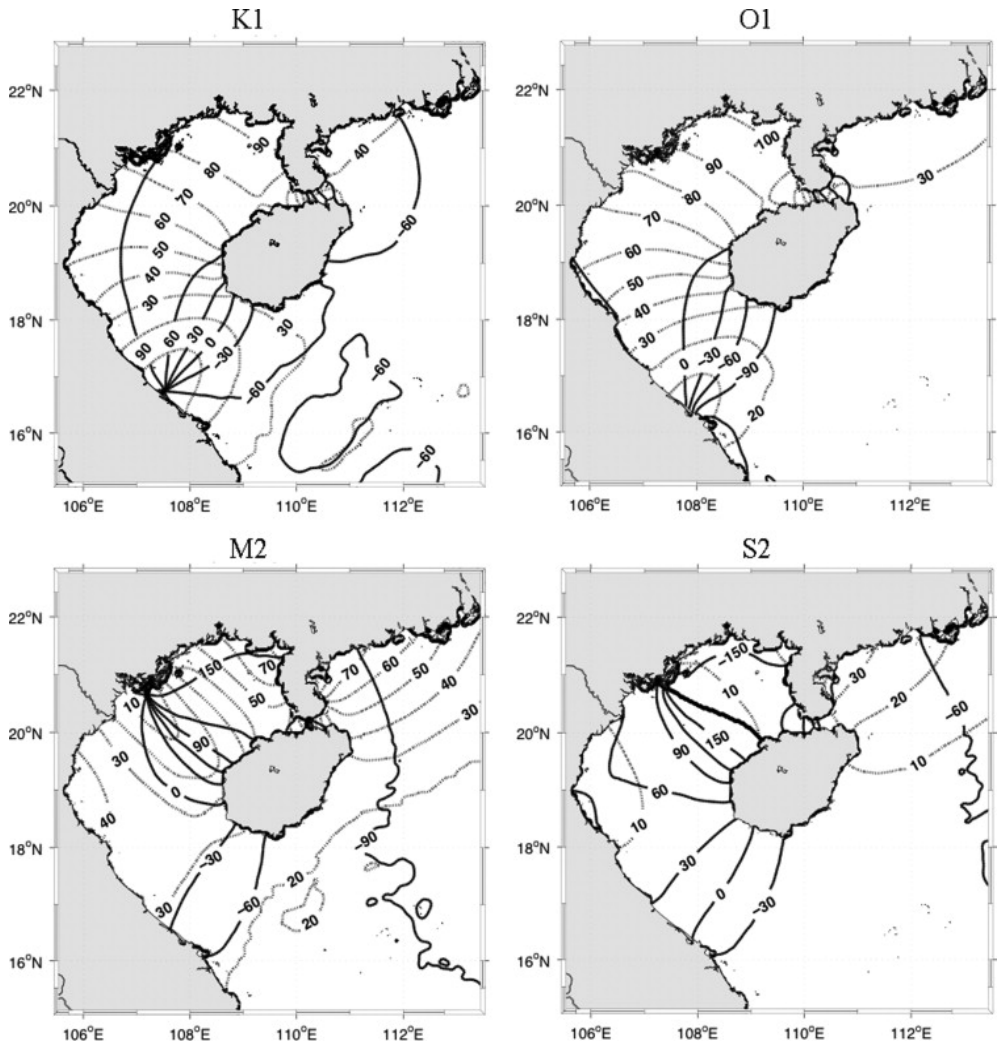

Figure I.5 Co-tidal charts for the constituents of (a) K1, (b) O1, (c) M2, and (d) S2, adapted from Minh et al. (2014). (Solid line: phase-lag in degree, dashed line: amplitude in cm).

Nguyen et al. (2014) also identify areas where large tidal currents induce a strong vertical mixing: in the Hainan Strait, along the western coast of Hainan Island and around the Leizhou Peninsula. The intense mixing results in a significant surface cooling in the model. These areas are also consistent with areas of high Chl-a concentrations as depicted by satellite ocean color observations.

Using a tidal model (T-UGOm), accurate bathymetry and coastline datasets built from from digitized nautical charts, soundings, intertidal DEM, and GEBCO bathymetry and satellite images, Piton et al. (2020) provide an update of the tidal amplitude and phase for the main constituents with significantly lower errors for O1, K1, M2 and S2 (except for the phase of M2) with respect to altimetric and tide gauges observations. Remaining errors are imputed to uncertainties in bathymetry and details of the coastline. Piton et al. (2020) also highlights the impact of the choice of the bottom roughness value; however, regionalizing the roughness length to represent the variability of seabed composition and morphology, only slightly improved the accuracy.

I.3.4 The Red River delta

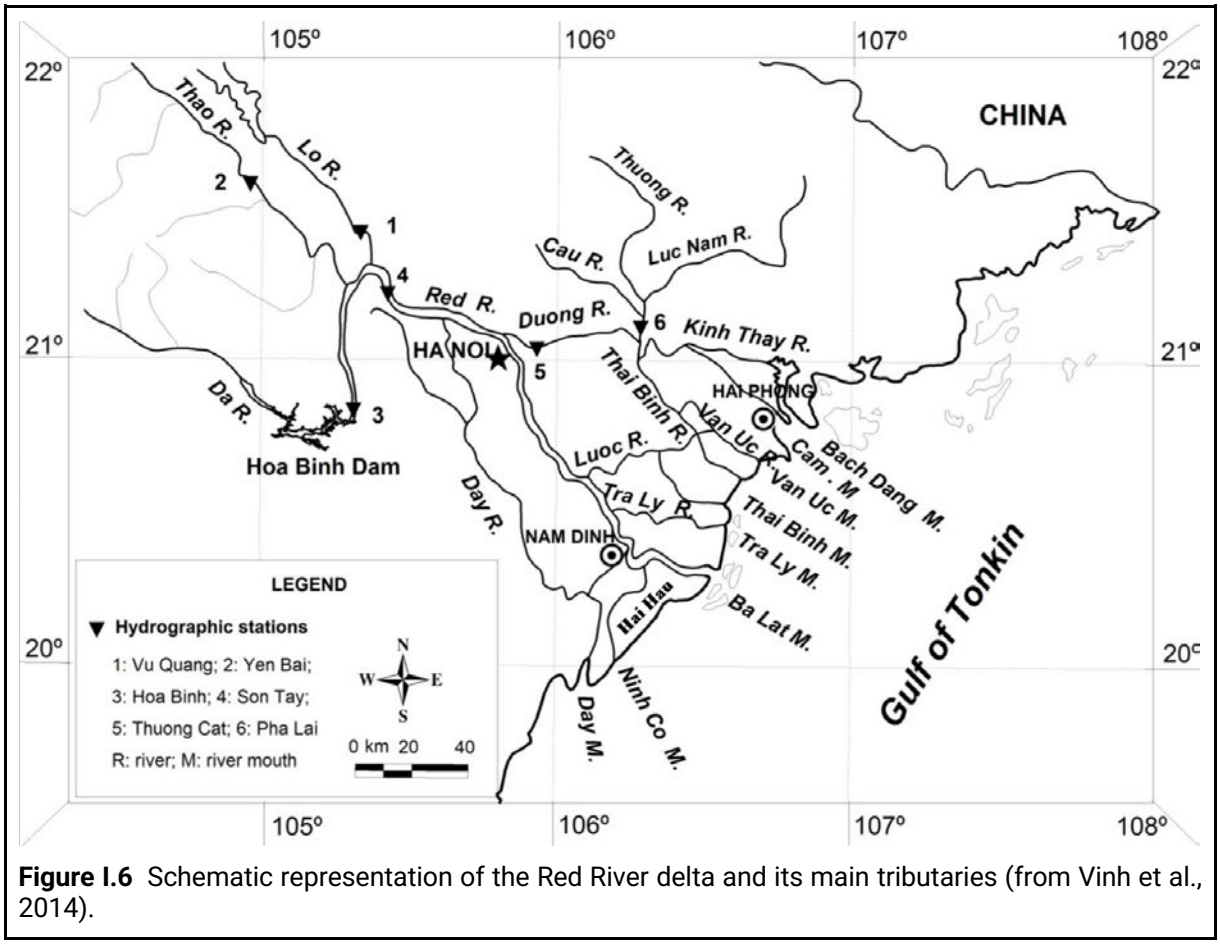

Figure I.6 Schematic representation of the Red River delta and its main tributaries (from Vinh et al., 2014).

The Red River (RR) is the second-largest river system in Vietnam after the Mekong, and is the main river that feeds the gulf (**Figure I.6**). It originates in the Yunnan Chinese province and owes its name to the color of its waters which carry iron-rich sediments (Vinh, 2018). The RR system is formed by 3 tributaries: Da, Thao and Lo Rivers which join upstream the city of Son Tay (about 150 km from the sea). A few kilometers upstream of Ha Noi, they split into the Red River and the Thai Binh river, which in turn split into a network of connected distributaries: the Red River delta. A complex system of channels has also been built in the delta for irrigation and drainage purposes.Daily RR runoff, obtained from the National Hydro–Meteorological Service (NHMS) of Vietnam at Son Tay hydrological station, shows that the mean runoff from 2008 - 2018 is approximately 3000 m3/s, accounting for more than 60% of the total discharge to the gulf. The runoff undergoes a strong seasonal cycle with high discharge during the rainy

season (from June to September) and low discharges in the dry season (October to May). Interannual variability affects the hydrological regime as well, but regulation intervenes to redistribute water in the channels and tributaries (see for instance Luu et al., 2010).

The sediment transport from the Red River varies greatly at seasonal and interannual time scales. According to Lu et al. (2015), the sediment load at Son Tay in 2008, 2009, and 2010 are 30, 12.8, and 10 Mt/yr, respectively. Wei et al. (2021) show that the mean sediment load from 2008 - 2013 is 11.6 Mt/yr.

The RR delta (RRD) is formed from 11 provinces (Prime Minister, 2013), which corresponds to a total population of 22.5 millions (Vietnam General Statistic Office. 2020) and an area of 21 200 km2. The average population density of the RRD is around 1060 inhabitants/km2, which is nearly four times higher than the national average for Vietnam. The very high population density exerts a strong pressure on the rivers and their environment. For instance, large contaminant spills are expected, due to farming and industrial activities.

One of the pollution problems is the emergence of microplastics in the ocean. Microplastics are plastic particles ranging from 1 to 5000 μm long (Frias and Nash, 2019). The measurement made by Strady et al. (2021) showed that the microplastics concentration in Red River is around 2.3 items/m3, while in the ocean, it is 0.4, 0.7 and 0.8 items/m3 in Cua Luc, Do Son and Ha Long Bays, respectively. Better understanding the variability of the river plume can help us to identify the pathway of the microplastics from river to ocean.

South of the RR, there are 3 more rivers: Ma, Yen, Lam. The summed discharges of these rivers and RR can account for around 90% of the total runoff in the GOT.

I.3.5 Red River Plume

In spite of its importance, to date, studies focusing on the RR plume are still scarce.

van Maren (2007) studies the water and sediment dynamics inside the Ba Lat mouth and its near field plume. The measurements at 5km inside the river mouth and at 2 km seaward, performed in the low and high discharge period in 2000 show that the temperature of the fresh river plume was usually 1–3° lower than ambient seawater. The freshwater plume (with salinity less than 6) is separated from the ambient water at 3km from the river mouth. He also suggests that in the dry season, the plume is subcritical, characterised by strong coastal current, while in the wet season, the plume is supercritical with large offshore spreading.

Using 2-years numerical simulation, Gao et al. (2013) showed that the coastal plume was found near the northern and western coasts of the gulf in winter while spreading eastward and offshore in summer. They suggest that the difference between the 2 seasons due to the monsoon wind conditions, which is downwelling favorable in the winter and upwelling favorable in the summer. The 10 – years simulation of Rogowski et al. (2019) show similar results : the southwesterly wind direction which is prominent during the summer monsoon is the main mechanism preventing the downcoast advection of the RR plume.

Using the CTD data located from 2 - 25 km from the coast in 2 years 2015-2016, Rogowski et al. (2019) and Tran et al. (2021) tried to examine the vertical structure of the RR plume from the stratification of salinity in different time of the year. In September, the plume thickness is 5m and extend beyond the survey line (25km from the coast). In the two following months, the

plume is deeper and detaches from the coast. In January, the plume is well mixed, due to the cooling of the coastal water. It is also the time when the runoff is lowest. The low stratification appears until March. In April, the thin plume is observed on the surface. There is no plume signature in June and July, suggesting that the river plume is advected offshore under southwesterly wind conditions.

Figure I.7 shows a Landsat-8 image, with the presence of the large spread of a water region with different optical properties near the coast created by the runoff from the RR. It is an example of the signature of the river plume.

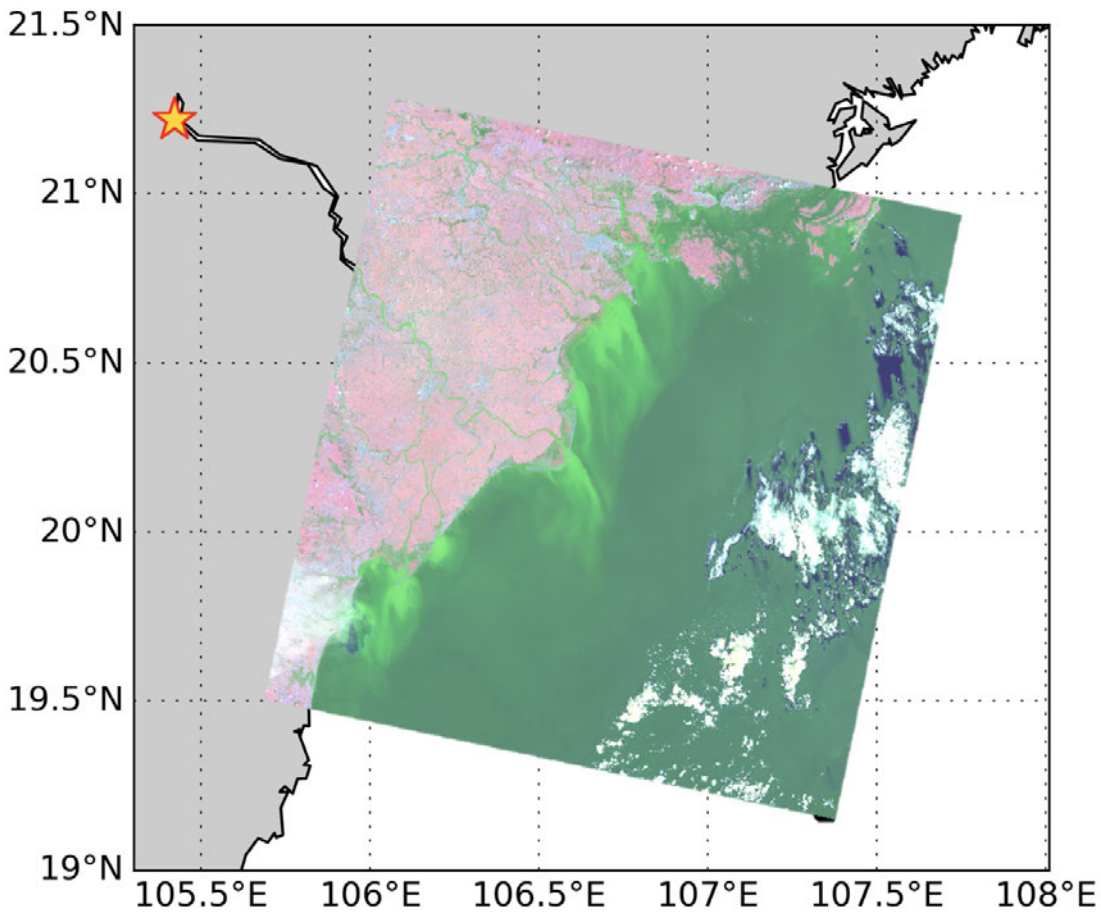

Figure I.7 Signature of the plume from the Red River, as depicted in Landsat-8 data on 25 August 2014 using band 5, 3 and 6. The star indicates the location of Son Tay hydrological station. Landsat-8 image courtesy of the U.S. Geological Survey. Landsat Scene Identifier: LC81260462014268LGN01.

To my knowledge, the existing studies still have some limitations to describe and explain the overall picture of the RR plume variability. In the study of van Maren, D. S. (2007), the role of wind is not examined because the wind velocity is very low during the measurement. The simulation of Gao et al. (2013) is run with climatological mean runoff, therefore cannot show the interannual variability of the plume. The simulation of Rogowski et al. (2019) has aspatial resolution of 6.5km, which is likely insufficient to resolve the dynamics associated with the RR plume.

I.4 Objectives and approaches

In this study, the main objective is to provide a better understanding of the variability of the RR plume in the GOT in different scales. The main scientific questions which we wish to address are the following:

1. How can high-resolution numerical simulations reproduce the Gulf of Tonkin dynamics and the Red River plume variability at different scales in the study years?
2. Which criteria can be used to identify the RR plume in the GOT from the model outputs?
3. How can we formally characterize the sequence of regimes which the RR plume goes through in the study years? Can we link those regimes to specific processes and forcing functions?
4. How can we assess the sensitivity of model variables and in particular on the variability of the river plume, to the main sources of uncertainty, in particular the atmospheric forcing uncertainties?

To address those questions, we will follow an approach involving several tools:

1. We will set up and run an upgraded version of the model configuration created as part of the thesis of V. Piton.
2. The river plume will be identified using different methods, allowing us to select the most appropriate one given our purposes.
3. The RR plume regimes will be identified and linked to physical processes with the help of K-means clustering analysis.
4. We will run an ensemble of simulations assessing the model response due to perturbations of the wind forcing. We will analyze the variability of the resulting uncertainties over space and time of the ocean variables. The K-means clustering approach will be used on ensemble members, giving access to an assessment of the robustness of the results of 3. above.

I.5 Outline of the thesis

The structure of the thesis is as follows. Chapter 1 presents an overview of the river plume and some past studies of the river plume in the different regions. Then, the physical characteristics and an overview of the past studies in GOT is presented. In Chapter 2, the SYMPHONIE model, model configurations, data sources and the unsupervised learning algorithm used to classify the main pattern of the plume (K-means) are described. The model is then evaluated against several observational data sets, including HF radar current data from the results of C. Tran thesis, in Chapter 3. The impact on the dynamics of a new parameterization of the light attenuation in the water column is also presented.

Chapter 4 contains an article published as a part of this thesis. To identify the plume in the GOT, several methods are compared which allows us to select the most appropriate one given our purposes. On that basis, the river plume variability at different time scales is examined with the help of K-means clustering analysis, illustrating the effect of key physical processes. Then, a discussion on the plume classification and a description of the receiving basins from the different river systems in the GOT are presented. The end of this chapter provides some further discussions on the model boundary conditions and on the sensitivity of the clustering algorithm.

Chapter 5 aims at assessing the robustness of the model in general, and of the river plume estimate of Chapter 4 in particular, to the uncertainty of the wind forcing. To do that, an ensemble simulation approach is used. Firstly, the uncertainty of the wind forcing is assessed by comparing it to satellite data. A pseudorandom combination of bivariate wind EOFs is then used to perturb the wind field within that uncertainty. The sensitivity of the main model variables to the perturbations is then assessed by calculating the ensemble spread and the distributions of the variables of interest. Then, the sensitivity of the river plume to the stochastic forcing is presented. To assess the effectiveness of the ensemble configuration at reproducing the error, the ensemble is verified using the available observational datasets.

Finally, Chapter 6 summarizes and gives overall conclusions and future perspectives.

CHAPTER II: METHODOLOGIES AND DATA

***

Most of the investigations, tools and approaches in this Ph.D. thesis use numerical modeling. In this chapter, I will first describe the principle of the SYMPHONIE model. Then, the configuration of the study area is presented. Finally, a list of the available observational data which are used for model validation is provided.

II.1 SYMPHONIE model

SYMPHONIE is a numerical model that solves the primitive, Boussinesq, hydrostatic equations of the ocean circulation (Marsaleix et al. 2006, 2008) on a curvilinear bipolar (Bentsen et al. 1999) Arakawa C–grid with sigma or hybrid vertical levels. SYMPHONIE has been developed to model the coastal circulation from regional scales to shelf seas and nearshore areas. In particular, SYMPHONIE allows to accurately represent tides, intertidal areas, and the oceanic response to high-frequency atmospheric forcing. It can run on highly variable polar or bipolar grids in order to refine the resolution locally, for instance at river mouths or inside estuaries. It has been extensively used in the Mediterranean Sea to study a wide range of physical processes, from the large scale circulation, to the deep convection in the western basin, dense water cascading, river plume, impact of storms, wave-induced currents, sediment transport and impact of the circulation on the biogeochemistry (see https://sirocco.obs-mip.fr). It has also been used in the Bay of Biscay and Gironde estuary, Gulf of Guinea, northern Pacific ocean and southeast asian seas. Since the PhD of V. Piton, it has been implemented in the Gulf of Tonkin. This section is devoted to the description of the model equations and numerical implementation.

II.1.1 Governing equations

In a Cartesian coordinate system, assuming that the origin O corresponds to the rest level of the sea surface, Ox, Oy are the horizontal axes and Oz is the vertical ascending axis. The velocity components u, v and w (corresponding to Ox, Oy and Oz axes) are calculated from the continuity and momentum equations:

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0$$	Eq. 2.1
$$\frac{\partial u}{\partial t}+\frac{\partial uu}{\partial x}+\frac{\partial vu}{\partial y}+\frac{\partial wu}{\partial z}-fv=\frac{1}{\rho_0}\frac{\partial p}{\partial x}+\upsilon\Delta^2 u+\frac{\partial}{\partial z}\left(K_m\frac{\partial u}{\partial z}\right)$$	Eq. 2.2
$$\frac{\partial v}{\partial t}+\frac{\partial uv}{\partial x}+\frac{\partial vv}{\partial y}+\frac{\partial wv}{\partial z}+fu=\frac{1}{\rho_0}\frac{\partial p}{\partial y}+\upsilon\Delta^2 v+\frac{\partial}{\partial z}\left(K_m\frac{\partial v}{\partial z}\right)$$	Eq. 2.3

with:

f: Coriolis parameter

ρ_0: reference water density

$p(x, y, z, t)$: pressure

υ: kinematic viscosity

Δ^2: horizontal squared Laplacian operator

$K_m(x, y, z, t)$: vertical coefficient of eddy diffusivity

The pressure, is given by the hydrostatic hypothesis:

$$p(z) = \int_z^\eta g\rho dz$$	Eq. 2.4

with:

g : acceleration due to gravity,

$\rho(x, y, z, t)$: density

$\eta(x, y, t)$: free surface elevation anomaly with respect to the surface elevation at rest

η is a prognostic variable derived from the depth-averaged component of the current $(\bar{u}, \bar{v})$:

$$(\bar{u}, \bar{v}) = \frac{1}{H}\int_{-h}^{\eta}(u, v)dz$$

using the following equation:

$$\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x}(H\bar{u}) + \frac{\partial}{\partial y}(H\bar{v}) = 0$$	Eq. 2.5

with:

H = h + η is the water column thickness at time t

h: water column thickness at rest above the bottom topography

The water density, ρ, is computed from temperature, salinity and pressure, using a non-linear equation of state (Jackett et al. 2006). Temperature and salinity variations are calculated by:

$$\frac{\partial T}{\partial t} + \frac{\partial uT}{\partial x} + \frac{\partial vT}{\partial y} + \frac{\partial wT}{\partial z} = \frac{\partial}{\partial z}(K_h \frac{\partial T}{\partial z}) + \frac{1}{\rho_0 C_p}\frac{\partial I_s}{\partial z}$$	Eq. 2.6
$$\frac{\partial S}{\partial t} + \frac{\partial uS}{\partial x} + \frac{\partial vS}{\partial y} + \frac{\partial wS}{\partial z} = \frac{\partial}{\partial z}(K_h \frac{\partial S}{\partial z})$$	Eq. 2.7

where $I_s$is the part of the solar radiative forcing that penetrates to depth z, C_p is the water heat capacity. The horizontal diffusion fluxes for temperature and salinity are not explicitly simulated; instead, the diffusivity of the numerical scheme for the advection is used (QUICKEST scheme). The eddy diffusivity coefficient is determined by the turbulence closure scheme.

II.1.2 Turbulence closure scheme

Coefficients of eddy diffusivity (K_h) and viscosity (K_m) are computed using turbulence closure schemes. In this study, the K-ε turbulence closure scheme is used, with the implementation described in Michaud et al. (2012).

The vertical eddy diffusivity K_h and viscosity K_m are defined as:

$$K_h = \sqrt{2E_k} l_k S_h$$	Eq. 2.8
$$K_m = \sqrt{2E_k} l_k S_m$$	Eq. 2.9

with:

E_k: Turbulent Kinetic Energy (TKE)

l_k: mixing length

S_h, S_m: stability functions (Canuto et al. 2001)

E_k and l_k are calculated by solving the following equations:

$$\frac{\partial E_k}{\partial t} + \frac{\partial u E_k}{\partial x} + \frac{\partial v E_k}{\partial y} + \frac{\partial w E_k}{\partial z} = \frac{\partial}{\partial z}(K_m \frac{\partial E_k}{\partial z}) + P + B - \varepsilon$$	Eq. 2.10
$$\frac{\partial \varepsilon}{\partial t} = \frac{\partial}{\partial z}(\frac{K_m}{\sigma_k}\frac{\partial E_k}{\partial z}) + \frac{\varepsilon}{E_k}(c_1 P + c_3 B - c_2 \varepsilon)$$	Eq. 2.11

With ε is the dissipation rate of the turbulent kinetic energy while P is the production of turbulent kinetic energy and B the buoyancy flux. $c_0 = 0.5544$, c_1=1.44 , $c_2 = 1.92$. $c_3 = 1$ if $B \geq 0$ and $c_3 = -0.52$ otherwise (Warner et al., 2005). σ_k = 1.3 is the Schmidt number for dissipation.

The relationship between the turbulent length, l_k , E_k and ε, writes:

$$l_k = \frac{c_0^3 E_k^{3/2}}{\varepsilon}$$	Eq. 2.12

P and B are calculated as:

$$P = K_m \left[\left(\frac{\partial u}{\partial z}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2 \right]$$	Eq. 2.13
$$B = \frac{g}{\rho} K_h \frac{\partial \rho}{\partial z}$$	Eq. 2.14

Eq. 2.13 expresses that the production of TKE is due to the vertical shear of the horizontal current.

II.1.3 Boundary conditions

II.1.3.1 Surface boundary

At the surface, the vertical flux of turbulent kinetic energy ($F_z = K_m \frac{\partial E_k}{\partial z}$) can be expressed, following Craig and Banner (1994), as:

$$F_z = 100\left(\frac{\tau_s}{\rho}\right)^{\frac{3}{2}}$$	Eq. 2.15

where τ_s is the wind stress computed by the bulk formula using the ECMWF variables (wind velocity, temperature and relative humidity).

The surface current condition affected by the wind stress τ_s is determined as follows:

$$\rho_0 K_m\left(\frac{\partial u}{\partial z}, \frac{\partial v}{\partial z}\right) = \left(\tau_{sx}, \tau_{sy}\right)$$	Eq. 2.16

The pressure is forced by the atmospheric pressure at the surface provided by ECMWF.

The heat flux at the ocean-atmosphere boundary is computed from 4 components.

$$K_m \frac{\partial T}{\partial z} = Q_S + Q_L + Q_E + Q_H$$	Eq. 2.17

With:

- Q_S: solar radiation
- Q_L: longwave radiation
- Q_E: latent heat flux (proportional to the evaporation)
- Q_H: sensible heat flux

Surface boundary condition for salinity:

$$K_m \frac{\partial S}{\partial z} = (E_S + Pr_S)S$$	Eq. 2.18

with E_S is the evaporation rate and Pr_S is the precipitation rate.

The solar radiation, the downwelling longwave radiation and the precipitation are provided by ECMWF analyses. The longwave radiation from the surface, turbulent heat fluxes and

evaporation are computed using atmospheric variables from ECMWF analyses, the sea surface temperature simulated by SYMPHONIE and the bulk formulae of Large and Yeager (2004).

II.1.3.2 Bottom boundary

At the bottom, the boundary condition approximation for $E_k$corresponds to the equilibrium between energy production and dissipation (Michaud et al., 2012) , with:

$$E_{k_{z=-h}} = \frac{\|\|\tau_b\|\|}{\rho_0\sqrt{2^{0.5}c_0^3 S_m}}$$	Eq. 2.19

where $\overrightarrow{\tau_b}$is the bottom stress, calculated by:

$$\overrightarrow{\tau_b} = \rho_0 C_D \|\|\overrightarrow{V_b}\|\|\overrightarrow{V_b} = \rho_0 u_*^2$$	Eq. 2.20

with:

C_D: non-dimensional bottom drag coefficient.

$\overrightarrow{V_b}$: bottom velocity

u_* : friction velocity

For the drag coefficient, we use a quadratic parametrization as a function of the bottom roughness (z_0) as follows:

$$C_D = (\frac{\kappa}{log(\frac{z_1}{z_0})})^2$$	Eq. 2.21

with κ= 0.4 is the Von Karman coefficient, z_1 is the thickness of the first layer above the bottom.

The flux of salinity and temperature is equal to 0:

$$K_m(\frac{\partial T}{\partial z}, \frac{\partial S}{\partial z}) = 0$$	Eq. 2.22

II.1.3.3 Lateral boundaries

At a closed boundary (continent or island), the velocity components are equal to 0.

The open-boundary conditions are described in Marsaleix et al. (2006) and Toublanc et al. (2018, their appendix). At an open ocean boundary, the equations write differently for the

barotropic and baroclinic velocities. Here, the barotropic velocities are defined as the depth-averaged velocities, while the baroclinic ones are the anomalies with respect to the vertical mean. For simplicity, we suppose that the domain of integration is 1D (along x) with open boundaries at x=0 and x=L:

For the barotropic current $(\bar{u}, \bar{v})$:, the Flather type condition is applied to the surface elevation η:

$\eta - \eta_f = -\sqrt{\frac{H}{g}}\left(\bar{u} - \overline{u_f}\right)$ at x = 0	Eq. 2.23a
$\eta - \eta_f = \sqrt{\frac{H}{g}}\left(\bar{u} - \overline{u_f}\right)$ at x = L	Eq. 2.23b

where the f index indicates the forcing (large scale circulation and tidal circulation).

The tangential transport at the boundaries is computed using the Neumann condition:

$\frac{\partial H\left(\bar{v} - \overline{v_f}\right)}{\partial x} = 0$ at x = 0 and x = L	Eq. 2.24

The transport component normal to the boundary is deduced from the continuity equation, from η (Eq. 2.23) and from the tangential transport (Eq. 2.24).

For the baroclinic currents (u, v), a radiative condition is applied on the difference between the SYMPHONIE variables (Φ) and the forcing variables (Φ_f):

$\frac{\partial\left(\phi - \phi_f\right)}{\partial t} + C_c\frac{\partial\left(\phi - \phi_f\right)}{\partial x} - \frac{\phi - \phi_f}{\tau_R} = 0$ at x = 0	Eq. 2.25a
$\frac{\partial\left(\phi - \phi_f\right)}{\partial t} - C_c\frac{\partial\left(\phi - \phi_f\right)}{\partial x} - \frac{\phi - \phi_f}{\tau_R} = 0$ at x = L	Eq. 2.25b

with

ϕ: SYMPHONIE variables (u,v)

ϕ_f: forcing variables

C_c: phase velocity of baroclinic waves.

τ_R: restoring time scale

Both the barotropic and baroclinic velocities are relaxed towards the external forcing within a sponge-layer in order to reduce possible reflection of the outgoing flow. In the sponge layer, the relaxation time scale decreases from a fixed value at the boundary to zero at the most internal point of the sponge-layer inside the integration domain.

For the scalar variables (T, S), the open boundary conditions use a hybrid scheme depending on the flow direction. If the flow is incoming, the temperature and salinity (T, S) are deduced from the forcing (T_f, S_f) of a larger scale model; a relaxation to (T_f, S_f) is also applied in the sponge layer with a time scale depending on the forcing current amplitude. In the case of outflow, these variables are issued from the simulated domain.

II.1.3.4 Boundary condition at the river input point

At the river input point, the river boundary current is consistent with the runoff of the river as in Reffray et al (2004), while the vertical shape of the current is given by a gradient condition allowing the current to adjust by itself to the dynamics of the river downstream (rather than arbitrarily prescribing a possibly inappropriate vertical profile). The boundary condition is applied in two steps. Step 1 is given by the above–mentioned gradient condition, leading to a provisional solution of the current:

$\frac{\partial u*}{\partial x} = 0$	if current in downstream direction of the river	Eq. 2.26
$u^* = 0$	otherwise	

where x is the direction of the river flow and u^* a provisional solution of the current. Step 2 ensures that the total flow is consistent with the expected runoff from the river. In practice, the river boundary current is proportional to u^*, i.e. $u = \alpha u^*$ where α is a constant used to satisfy the constraint on the prescribed river runoff:

$$\alpha \Delta y \int_{-h}^{\eta} u(z)^* \, dz = F \qquad \text{Eq. 2.27}$$

where Δy is the cellbox width at the river input point and F the river runoff.

II.1.4 Tides

The tidal signal is made of different constituents corresponding to different frequencies. The tidal surface elevation and barotropic currents can be written as follows (see for instance Pairaud et al., 2008):

$$\eta(\lambda, \varphi, t) = \sum_k f_k \eta_{0,k}(\lambda, \varphi) cos\left(\omega_k (t - t_0) + V_{0,k} + u_k - G_{\eta,k}(\lambda, \varphi)\right) \qquad \text{Eq. 2.28a}$$

$$u(\lambda,\varphi,t) = \sum_k f_k u_{0,k}(\lambda,\varphi) cos\left(\omega_k(t-t_0) + V_{0,k} + u_k - G_{u,k}(\lambda,\varphi)\right)$$	Eq. 2.28b
$$v(\lambda,\varphi,t) = \sum_k f_k v_{0,k}(\lambda,\varphi) cos\left(\omega_k(t-t_0) + V_{0,k} + u_k - G_{v,k}(\lambda,\varphi)\right)$$	Eq. 2.28c

where η_0, u_0, v_0 and G_η, G_u, G_v are the amplitudes and phase lags for sea surface elevations and currents of the tidal constituent k at the longitude λ and the latitude φ. The nodal factors f_k and u_k are considered as constants (Doodson, 1927), ω is the tidal frequency and V_0 is a constant related to the reference time t_0.

In coastal models, tides are prescribed at the open boundaries, usually as harmonic constituents (amplitude and phase) for the main frequencies in the areas of study (the tidal signal is then reconstructed at each time step using Eq. 25). The prescribed fields come from global atlases (such as FES2014) or dedicated simulations (as discussed for instance in Toublanc et al., 2018).

In SYMPHONIE, the tidal forcing also consists of astronomical and loading potentials (respectively Π_A and Π_L) (see Pairaud et al., 2008, their appendix). The tidal potential due to astronomical effects Π_A (Hendershott, 1972) is taken into account as follows:

$$\Pi_A = (1+k_2+h_2) f a \times \left(v_0 \frac{1-3sin^2(\varphi)}{2} + v_1 sin(2\varphi) + v_2 cos^2(\varphi)\right) \times cos\big(\omega(t-t_0)\big) + \nu\lambda + V_0 + u$$	Eq. 2.29

with:

k_2, h_2: Love numbers

a: the equilibrium amplitude of the tidal constituents considered,

v_0, v_1, v_2, and ν: parameters depending on the type of tide. In case of long period tide $v_0 = 1$ and $\nu = 0$, of diurnal tide $v_1 = 1$ and $\nu = 1$, of semi diurnal tide $v_2 = 1$ and $\nu = 2$.

The loading potential Π_L (LSA potential) is expressed as:

$$\Pi_L = f\, \Pi_0(\lambda,\varphi) cos(\omega(t-t_0) + G_\Pi(\lambda,\varphi) + V_0 + u)$$	Eq. 2.30

The hydrostatic pressure force in the horizontal momentum equation is determined by:

$$-\frac{\nabla p}{\rho_0} = -\frac{g}{\rho_0}\nabla \int_z^{\eta} (\rho - \rho_0) dz' - g\nabla(\eta - \Pi_A - \Pi_L) \qquad \text{Eq. 2.31}$$

A wetting and drying scheme is used to represent intertidal areas.

The model computes an online harmonic analysis on the sea surface elevation and depth-averaged currents. A first run is performed to compute the tides. Then the result of the harmonic analysis from this first run is used to detide the daily averages of the sea surface elevation and horizontal current that we use in our studies.

II.1.5 Discretization

II.1.5.1. Spatial discretization

The SYMPHONIE model equations are solved using the finite difference method (Marsaleix et al., 2008) on the Arakawa staggered C-grid (Arakawa and Suarez, 1983) (**Figure II.1**). Horizontal velocity components (u, v) are defined at the center of the edges of the grid, whereas the scalar components (temperature, salinity, surface elevation) are defined at the center of the grid cell. The vertical velocity and kinetic energy are calculated at the center of the grid edges at each vertical level.

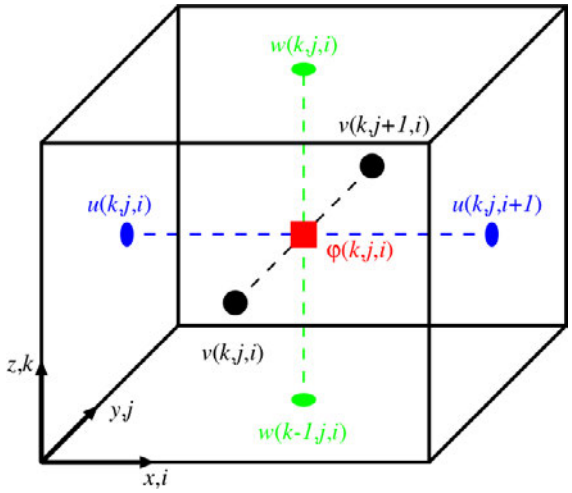

Figure II.1 The Arakawa staggered C-grid (Piton, 2019). The i, j, k indices correspond to x, y, z direction, respectively. φ represents the scalar fields. Turbulent fields are defined at the same positions as the vertical component of velocity.

For the vertical grid, SYMPHONIE provides several options. In this study, due to the shallow bathymetry in the GOT, the terrain-following 'sigma' coordinates system is used. It allows us to model with high vertical resolution the coastal shallow area where the plume spreads while saving the computing resources in the deep area.

The vertical advection is computed using a centered scheme. Horizontal advection and diffusion of temperature and salinity are computed by the QUICKEST scheme (Leonard, 1979). Horizontal advection and diffusion of momentum are each computed with a fourth order centered biharmonic scheme as in Damien et al (2017). The biharmonic viscosity of momentum is calculated according to a Smagorinsky-like formulation derived from Griffies and Hallberg (2000).

II.1.5.2. Temporal discretization

The forward-backward (F-B) time-stepping scheme is used for the temporal discretization of the equations. In this explicit time-stepping methods, the variable value at time t+1 is deduced by the variable value at time t:

$F' = F^t + \Delta t f(F^t)$	Eq. 2.32a
$F^{t+1} = F^t + \Delta t f(F')$	Eq. 2.32b

The F-B time-stepping application and chronology of variables in SYMPHONIE are described in more detail in Marsaleix et al. (2019).

The model uses a time-splitting method with a double loop on time: the model solves first the external mode (SSH and barotropic horizontal velocities) then solves the internal mode (baroclinic velocities, tracers and SSH). The external mode is solved with a very small time step consistent with fast gravity waves (such as tides).

II.2 Model configuration

In this study, the general configuration has been initially set-up for the PhD of V. Piton (defended in 2019) at LEGOS. I have then updated the version of the code, tested a different vertical advection scheme in order to be able to run the model without any runoff for a sensitivity test, used a non variable parameter for the penetration of light and run with passive tracers. In II.2.1, I will first describe the general configuration. In II.2.2 I present a detailed description of the river setup and of the way passive tracers are introduced to follow the river waters. II.2.3 presents a configuration for a sensitivity test with light attenuation. Finally, the list of simulations is available in II.2.4.

II.2.1 General configuration

The model domain covers the GOT area, with 429 * 600 grid points. Due to the variable horizontal grid, the coastal area near the Red River (RR) mouths has a fine horizontal resolution of 300m, while near the open boundary, the grid size can increase to 4500m (**Figure II.2**). The vertical discretization consists of 20 sigma levels. As the coastal area adjacent to the RR delta is characterized by a complex topography, with many islands and islets, a considerable effort has been devoted to the construction of the model bathymetry and shorelines and is described in Piton et al. (2020). In particular, the bathymetry is reconstructed from GEBCO 2014 combined with other sources and field surveys. The bottom drag coefficient follows a logarithmic law (Blumberg and Mellor, 1987), depending upon the bottom roughness length

(z_0). Using the ROMS model, Minh et al. (2014) showed that the tidal error for O1 and K1 is lowest with z_0 around 1 mm. Piton et al. (2020), in their sensitivity study using a 2D configuration of the T-UGOm model, shows that the sum of errors for the 4 main tidal constituents (O1, K1, M2, S2) is lowest with constant z_0 equal to 0.9 mm. Furthermore, the error can be even more reduced with spatially varying z_0. However, these results do not hold in the 3D SYMPHONIE configuration: in Piton (2019), she confirmed that using z_0 = 1 mm gives a better solution in simulating the tide in. Therefore, in this study, z_0 is set up equal to 1 mm.

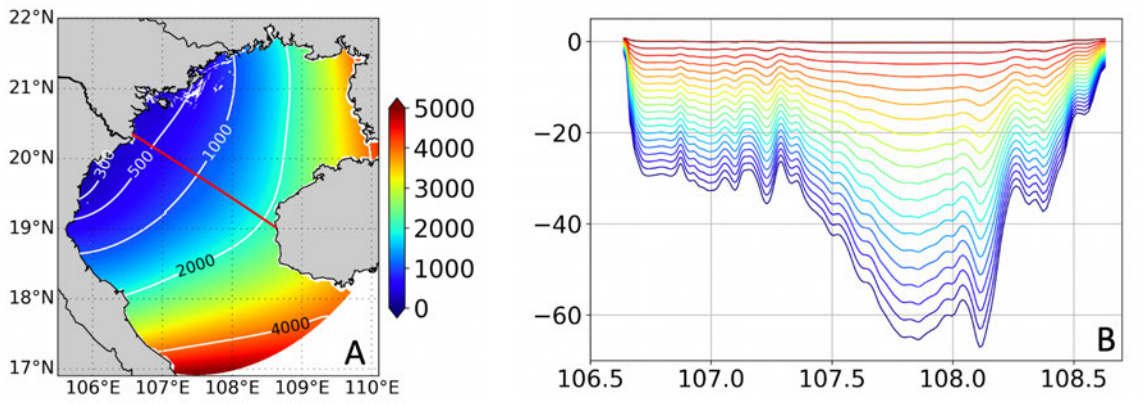

Figure II.2 **(A)**: The grid size of GOT_REF (m). **(B)**: Sigma vertical layers coordinate along the red transects in **(A)**.

At the open boundary, tidal surface elevation and current at K1, O1, P1, Q1, K2, M2, N2, S2, M4 frequencies from the tidal atlas FES2014 (Lyard et al., 2021) are taken into account. The model is also forced by daily averages of sea surface height (SSH), 3D zonal velocity (u), meridional velocity (v), temperature (T) and salinity (S) fields, from the global analysis (hereafter 'OGCM') produced by Mercator-Océan International and provided by Copernicus Marine Systems (CMEMS) at a resolution of 1/12°. CMEMS T, S fields are adjusted to recover consistency with tidal physics before being used at the open-boundary conditions of the model (please refer to the Appendix of the article in Chapter 4 for more information). The sponge-layer is 30-point wide; the relaxation time scales at the boundaries is 1 day and 0.1 day for the baroclinic and barotropic velocities respectively.

At the surface, boundary conditions are provided by an atmospheric model and fluxes of momentum, heat and freshwater are computed using the bulk formulae of Large and Yeager (2004). Operational ECMWF analyses (with a spatial resolution of 1/8°) are used to provide 3–hourly wind, precipitation, solar energy, atmospheric temperature, dew-point temperature, surface pressure. The land points of the atmospheric variables are not masked during the interpolation processes.

As the present study focuses on the fate of continental water into the coastal ocean, a specific effort has been deployed on the river runoffs implementation; section II.2.2 is dedicated to its description.

The model is run from 2010 to 2016, starting from the ocean state condition as provided by the global OGCM on 01/01/2010. The time step is set at 2 minutes. All analyses in this thesis are conducted during 2011–2016 (i.e., following a 1–year spinup). The model outputs include daily averaged variables as well as instantaneous fields every 12 hours. Unless otherwise specified,

both components of the current are detided based on an online harmonic analysis (as described in II.1.4).

Table II.1: General characteristics of the reference simulation (GOT_REF)

Name	**Features**
Grid points	429 * 600
Grid size	300m (near the coast), 4500m (near the boundary)
No. of vertical layers	20
Boundary conditions	global OGCM
Atmospheric conditions	ECMWF analysis (1/8°, 3h)
Tidal forcing	FES2014 (K1, O1, P1, Q1, K2, M2, N2, S2, M4)
Light attenuation	spatially variable (~1m near the coast and 15m offshore)
Bottom roughness	1 mm
Timestep	2 minutes
Period of study	Jan 1, 2010 – Dec 31, 2016

II.2.2 River configuration

II.2.2.1 River runoff data

In the GOT, there are 3 main sources of river runoff: from **RR**, from other 3 rivers (Ma, Yen, Lam) at the south of the Red River delta (hereafter referred to as the 'southern rivers', or **SR**) , and from 6 rivers at the north of the Red River delta (hereafter referred to as the 'northern rivers', or **NR**). The location of these rivers are shown in **Figure II.3**.

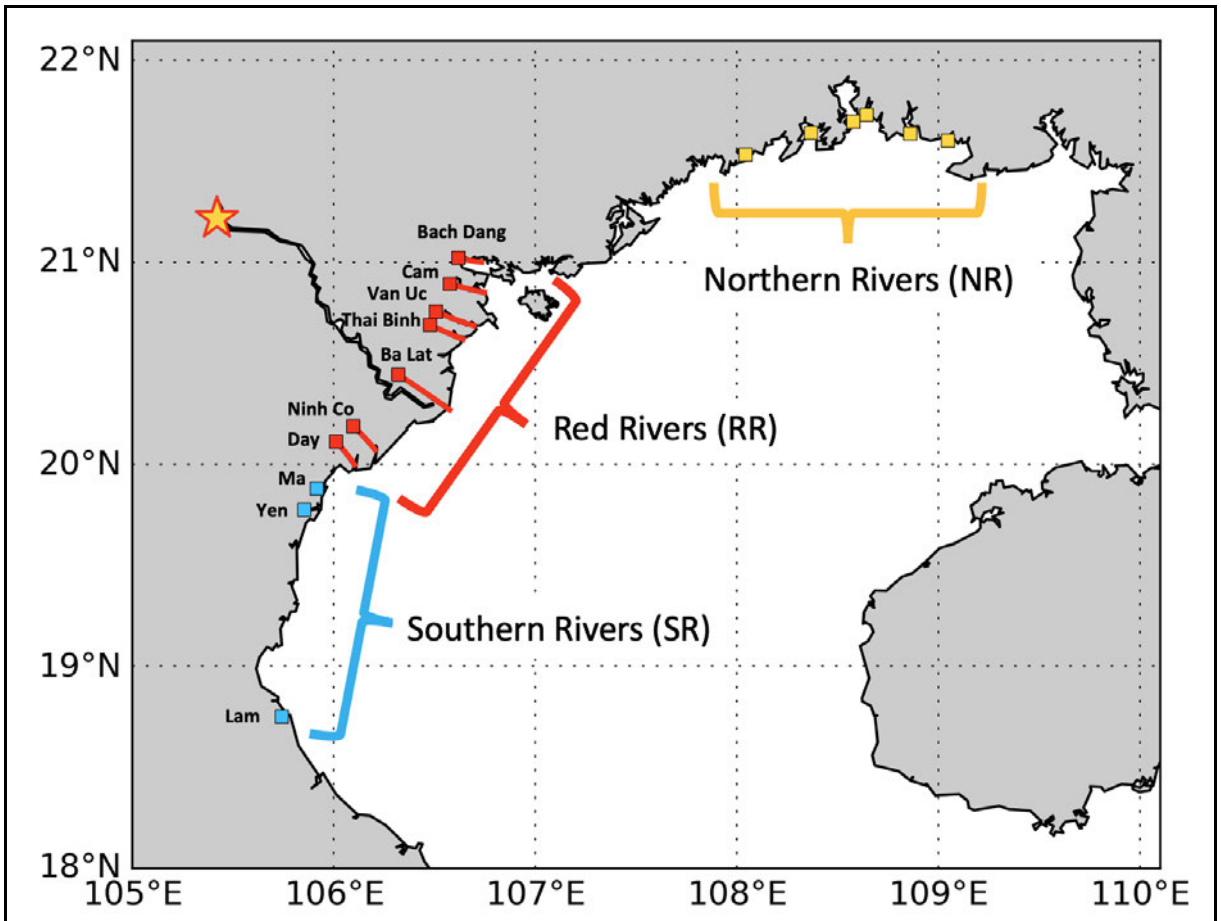

Figure II.3 The river input points of different rivers: Red River (red), southern rivers (cyan), northern rivers (yellow). Red lines illustrate the channels used for the RR system. The river runoff flows from the most upstream point of the channel. Note: the length of the red lines in this figure does not represent the real lengths of the channels.

The daily RR runoff is obtained from the National Hydro-Meteorological Service (NHMS) of Vietnam at Son Tay hydrological station which is located at the apex of the RR delta. For the southern rivers, since the daily runoffs were not available, the monthly climatological runoffs from NHMS is prescribed. The monthly runoffs of the northern rivers are obtained from the data given by Gao et al. (2013).

Figure II.4 shows the discharge from 3 river systems. In general, the discharge from the RR is the largest. On average, the runoff of the Red River system alone accounts for 60% of the total runoff to the gulf. In detail, the average discharge rate for the Red River system in low (December, January, February) and high (July, August, September) discharge periods equal 1632 m^3/s and 4959 m^3/s, respectively. It also shows interannual variations. While the peak in 2012 is approximately 15 000 m^3/s, the peak in 2011 is just over 6 000 m^3/s.

For the SR, runoff in high and low discharge periods is 365 and 2043 m^3/s, respectively. For the NR, they are 164 m^3/s and 1103 m^3/s. Since these values are monthly averaged climatological data, it does not have interannual variability and can be a source of uncertainty when assessing the plume area.

The ratio between high and low discharge seasons for RR is 3, while it is 5.6 for the SR and 6.7 for the NR. The lower ratio for the Red River may be explained by water flow regulation by

several hydrological dams upstream. On average, the total runoff from RR and SR accounts for approximately 90% of the total runoff in the GOT.

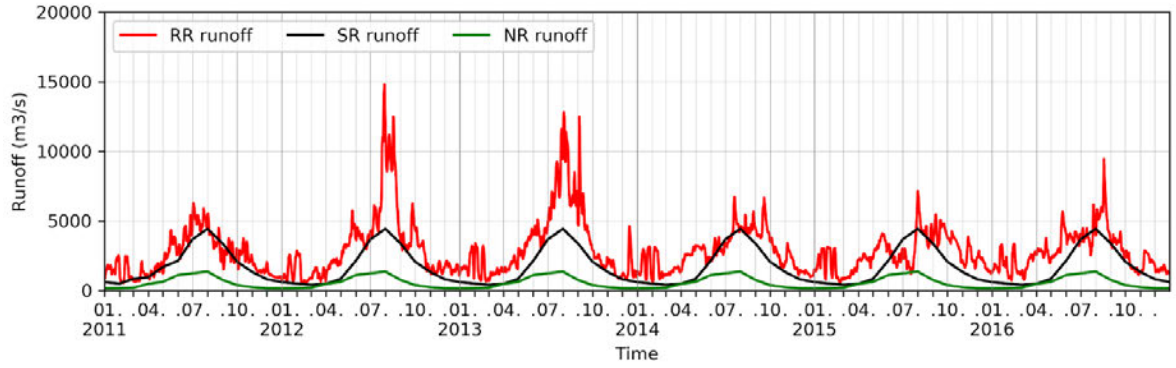

Figure II.4 Daily Red River discharge in the model configuration, from National Hydro-Meteorological Service data (red curve), monthly climatological discharge of the Southern Rivers (black curve), monthly climatological discharge of the Northern Rivers (green curve). Units are m^3/s.

II.2.2.2 River configuration

In most modeling studies to date, the RR was set up with only one mouth (or input grid point) using monthly or annual climatological runoff because more realistic runoff at the hydrological stations was unavailable (Ding et al. 2013, Gao et al. 2013). In this study and in Piton et al. (2021), the river condition is configured as realistically as possible.

Using the river discharge at the upstream and the water level at the river mouths, Vinh et al. (2014) attempted to model the Red River network and calculated the distribution of water discharge from Son Tay hydrological station into different river mouths. The results showed that the runoff is not equi-distributed. In this model configuration, the discharge is distributed across the seven distributaries based on their results: Bach Dang (7%), Cam (13.2%), Van Uc (14.5%), Thai Binh (6.4%), Ba Lat (30.3%), Ninh Co (5.6%), Day (23.0%). These 7 mouths lie along 150km of the coastline, and are expected to represent a better distribution of the plume than a single input point.

Furthermore, each river mouth is connected to a channel in the model to allow the water to enter the coastal ocean with realistic salt and temperature properties and realistic stratification. The length of the channels is chosen to exceed the saltwater intrusion, which is approximately 30 km from the mouth (Nguyen et al. 2020). In detail, the length for each channel is: Bach Dang (57 km), Cam (78 km), Van Uc (221km), Thai Binh (115km), Ba Lat (120km), Ninh Co (72km), Day (72km). At the most upstream point of the channel (input point), the salinity of the river flow is set to 0. The temperature varies seasonally from 17°C (in February) to 29°C (in August) to be consistent with observational data from a river in the north of Vietnam.

In the model, the channel can be configured automatically using some parameters (basic mode) or can be defined in each grid point (advanced mode). In the basic mode, the channel is configurable by the grid length (L), grid width (W) and the height (H) at the most upstream point (**Figure II.5**).

The entrance of the channel (downstream) is connected with a grid point on the ocean. The elevation of the river bed with respect to the reference level (z=0) at each grid point along the

channel is linearly interpolated from the depth of the downstream point and H. We do not define the depth of the channel, but we assume that there is no overflow. It means that the water height only depends on the discharge and the width of the channel. Similarly, the width of each grid point along the channel is interpolated from W and the width of the ocean point. The length is configured similarly. The channel is straight, and it does not represent the shape of the distributary.

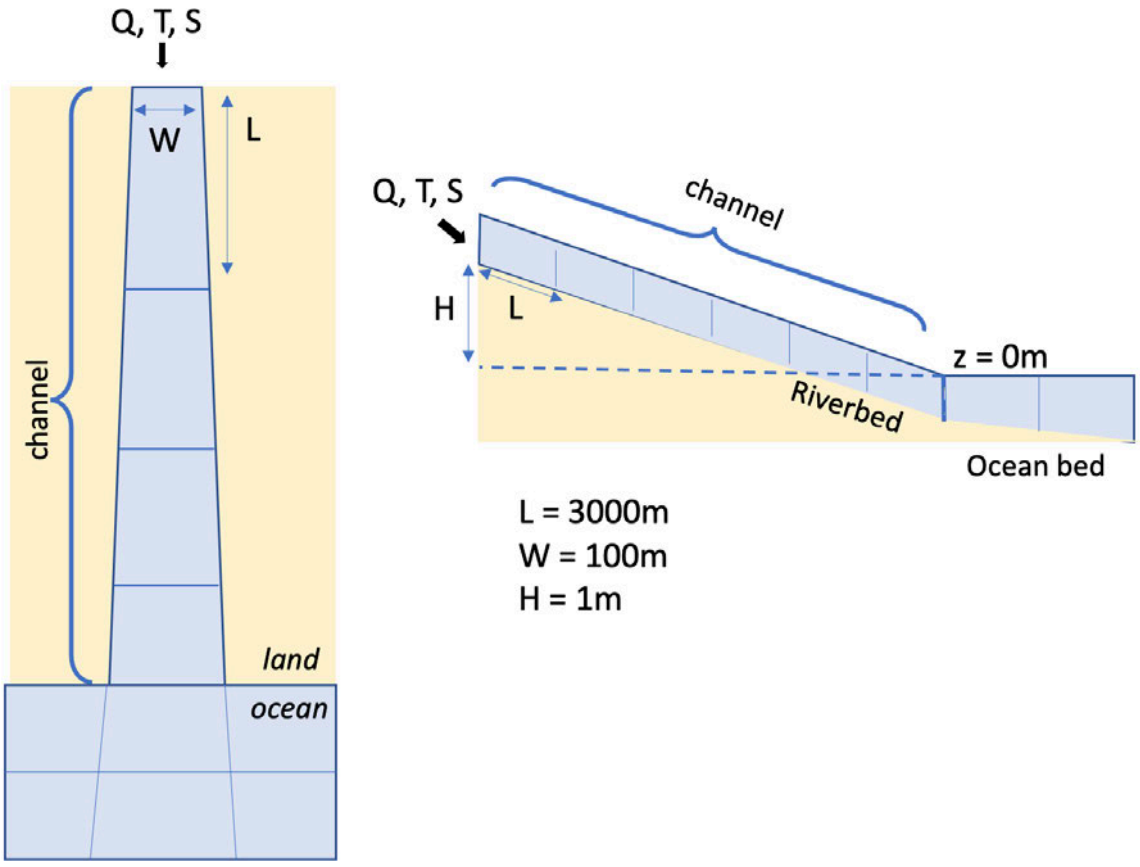

Figure II.5 The scheme and configuration of a channel in the model.

Figure II.6 illustrates the salinity profile in the Ninh Co channel, which has the lowest percentage of runoff over the 7 channels, on 03 March 2015. In the simulation with tide, at 0h, the salinity is present at around 35km from the river mouth (**Figure II.6A**), while at 12h, the propagation is 10km less (**Figure II.6B**), with the contour line of 0.0001. In the simulation without tide, the salinity intrusion seems more stable within 1 day. The salinity in both time frames shows strong stratification at around 25km from the river mouth. It also shows the changes of SSH over the day in GOT_REF, while in GOT_NOTIDE the change in SSH is much less significant. The important thing is that in both cases, with and without tide, the salinity at the river mouth is not zero and the implementation of the channel can help create better salinity mixing at the river mouth.

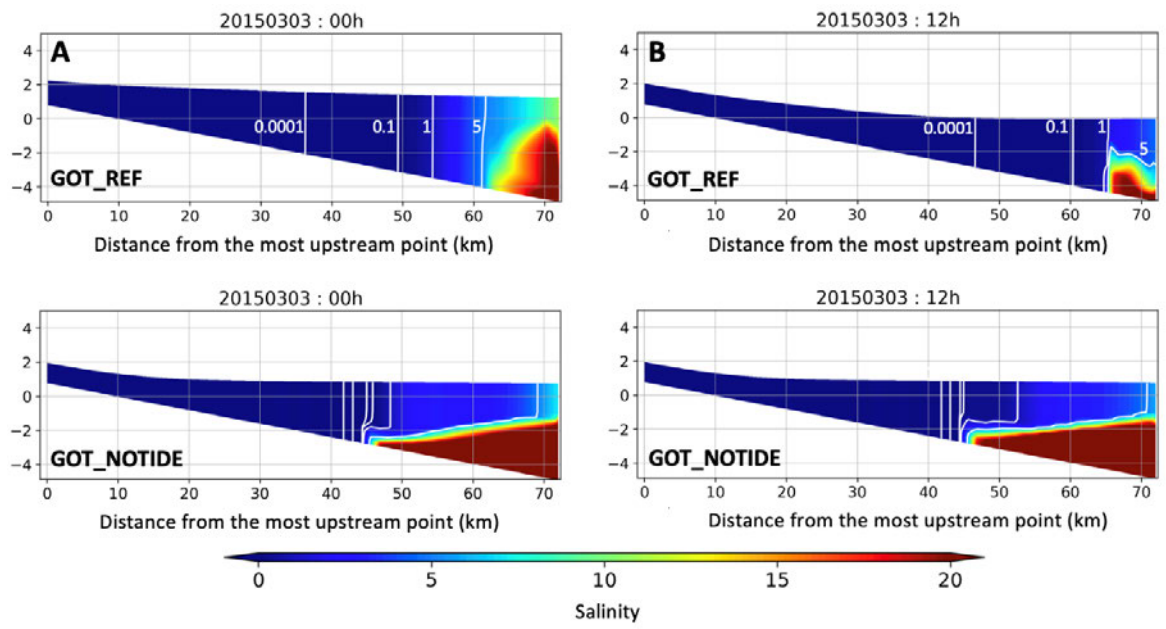

Figure II.6 Salinity sections from GOT_REF in Ninh Co on 03 March 2015, at 0h (A), 12h (B). (C, D): same as (A, B) but in GOT_NOTIDE

II.2.2.3 Passive tracer configuration

There are several ways to identify the plume. It can be detected using the difference between the salinity of the riverine water and the ambient water (e.g. Burla et al. (2010), MacCready et al. (2009) for Columbia river, Falcieri et al. (2014) for Po river), or using the satellite data thanks to the difference in temperature, turbidity, salinity, of the riverine water (e.g. Brando et al. 2015, Hopkins et al. 2013).

In GOT, the precipitation has seasonal variations, as shown in **Figure I.4**, with large precipitation along the coast during the high discharge period. Therefore, using surface salinity to identify the river plume could be biased by the freshening due to precipitation. In order to simulate the pathways of the river water into the GOT, in this study, passive tracers are used. Tracers act as dyes, i.e., they do not affect the dynamics. This approach has been used in several studies. Jia et al. (2019) use a tracer to simulate the pathway of the Connecticut River in the summer under different atmospheric conditions. The distribution of tracer in different areas allows to identify the progression of the river water at the tidal time scale. Gong et al. (2019) use several tracers to examine the plume-to-plume interaction in the Pearl River plume in the winter, when the runoff is highest. The results show that different plumes can affect each other and the existence of a plume can limit the transport of others. Wang et al. (2014) use a passive tracer to study the variability of the Copper River plume by calculating the evolution of the plume volume over time and plume occurrence frequency in different periods.

This is the first time that tracers are used for a pure dynamical study with SYMPHONIE. In total, there are 3 tracers (or three colors) meant to distinguish the inputs from the different river systems. The first tracer is added to the runoff of all the Red River distributaries at constant concentration (100 arbitrary unit/m^3). The second one is added at the mouths of the rivers in the

south and the third one at the mouths of the rivers in the north of RR, with the same concentration (100 arbitrary unit/m^3). Since tracers are injected with the runoff, they are also present in the 1 year of spin-up. The boundary conditions for the passive tracer (which is put directly into the river water) are applied in the same way as the salinity, except that the incoming concentration is always zero at the lateral boundary.

II.2.3 Light attenuation parameterization

II.2.3.1 Light penetration in the first layers of the ocean

In order to model correctly the ocean dynamics, using a suitable heat flux parameterization is very important. The change of heat flux can lead to change of temperature in the ocean and then change the mixing. The change of mixing can then impact the variations of the river plume. Therefore, an appropriate heat flux can improve the model result.

One of the parameters controlling heat fluxes is the shortwave attenuation depth (hereafter SWAD or L2). It characterizes the ability of the shortwave solar radiation f to go through water depth. This depth depends on the wavelength of the irradiance and the optical properties of the water. Short wavelengths can go further down, and clear water will let the light go further down.

From the available data, Jerlov (1968) has constructed the irradiance transmittance of 10 water types and 16 different wavelengths. Paulson and Simpson (1977) then proposed a method that uses only 'shortwave' and 'longwave' irradiances. They processed the data of Jerlov and provided a set of data with "shortwave attenuation length", "longwave attenuation length", "percentage of shortwave", "percentage of longwave".

The SYMPHONIE model uses the method proposed by Paulson and Simpson (1977) for parameterizing the light attenuation. By default, the shortwave attenuation length follows the type I of seawater classification (Jerlov, 1968) with a unique value of 23 m over the whole domain. However, the optical properties vary over space and time due to several factors (IOGC, 2000; Saulquin et al., 2013): biological activity (mainly the Chlorophyll concentration of phytoplankton), inorganic particulate suspended material (e.g. from sediment resuspension or from dust deposition) and coloured dissolved organic matter (CDOM) that can be produced locally or originate from rivers. In the coastal ocean with sediment transport from rivers, the SWAD is lower than in the open ocean. It also varies over time due to the temporal variation of the sediment transported to the ocean. Previous studies (e.g. Murtugudde et al. (2002) for the open ocean, Maraldi et al. (2013) in the Bay of Biscay/North Atlantic) have evidenced the large sensitivity of simulated temperature and stratification to spatially varying SWAD. We propose to investigate the sensitivity to this parameter in our simulation with the aim to improve the representation of the temperature field with respect to the available observations

II.2.3.2 The light attenuation schemes

The assumption of light penetration from the surface to depth is modeled, following Kraus and Businger (1994), as follows:

$$I_{(z)} = Q_S * \sum_{i=1}^{n} R_i \times e^{-\frac{z}{L_i}}$$	Eq. 2.33

With:

- $I_{(z)}$ Penetrative solar radiation at depth z
- Q_S solar radiation at the surface, assumed to be made of n radiations i
- R_i percentage of radiation i, with $\sum_i R_i = 1$
- z depth, with the axis at the surface, positive downward
- L_i penetration length for the radiation i

This equation allows a fraction of radiation (infrared) to be absorbed near the surface while another fraction (blue) can penetrate deeper (Kraus and Businger, 1994). In some cases, to simplify the model, n = 1 is used (Murtugudde et al., 2002). It means that all the radiations unite and share the same attenuation length. In some other cases, to specifically characterize the penetration of different radiations, n = 3 or more can be used (Sweeney et al., 2005). Typically, n = 2 is used for many studies (Maraldi et al., 2013; Ling et al., 2015; Zhou et al., 2015).

In SYMPHONIE, n = 2 is chosen. The original equation becomes:

$$I_{(z)} = Q_S * \left(R \times e^{-\frac{z}{L1}} + (1 - R) \times e^{-\frac{z}{L2}}\right)$$	Eq. 2.34

with:

- R: percentage of red and near-infrared in the solar radiation (58%)
- $L1$: penetration depth of red and near-infrared (set = 0.35m)
- $L2$: penetration depth of visible and ultraviolet.

In Eq. 2.31, the first part represents the percentage of energy which comes from the longwave irradiance. Due to low energy, this irradiance can penetrate just a few cm down from the surface and the longwave attenuation length (L1) is set equal to 0.35m for water type I (Paulson and Simpson, 1977).

The latter part of the equation corresponds to the irradiance from the short-wave visible light and lower band, like ultraviolet. It is called shortwave attenuation length (L2), or sometimes "Shortwave attenuation depth" (SWAD). SWAD varies over space and time depending on the different optical properties of the water. To deal with it, Jerlov (1968) attempted to characterize L2 in different ocean areas with different water types based on their optical clarity. For example, in water type I, the attenuation length is equal to 23m – this is the parameter value currently used in the Piton et al (2021) configuration. However, since the study area is characterized by strong seasonal activity of river transport near the coast, the use of spatially-varied L2 is expected to produce better results than using a single value.

L2 can be estimated from satellite products. The term "diffuse attenuation coefficient for the downwelling photosynthetically available radiation" (Kd(PAR) or k_{PAR}), which is proportional to L2, is used to mention the ability of light to go through depth.

$$L2 = SWAD = PAR = \frac{1}{k_{PAR}}$$	Eq. 2.35

To calculate k_{PAR}, the k_{D490} (diffuse attenuation coefficient at 490 nm) is generally used (Morel et al. 2007, Pierson et al. 2008, Wang et al. 2009, Zhou et al. 2015). k_{D490} is the attenuation coefficient of diffuse downwelling irradiance at 490 nm (Pierson et al. 2008), a parameter that indicates the turbidity of the water column. In other words, it indicates the ability of the light at 490 nm to penetrate in the water column. Higher k_{D490} value means smaller attenuation depth, and lower clarity of ocean water. k_{D490} is acquired by processing data from ocean color satellite sensors such as the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) or the Moderate Resolution Imaging Spectroradiometer (MODIS) (Wang et al. 2009).

Morel et al. (2007) proposed an algorithm to calculate the k_{PAR} from k_{D490} which is applied to optimize the Iberia-Biscay-Ireland configuration described by Maraldi et al. (2013). But this algorithm is valid only to water case I where chlorophyll concentration controls optical properties. For other regions where optical properties are dominated by CDOM, other algorithms are also developed (Wang et al. 2009, Pierson et al. 2008).

II.2.3.3 Estimation of attenuation length in Gulf of Tonkin

Saulquin et al. (2013) propose an alternative algorithm to calculate k_{PAR} in the turbidity-prevailing region. They suggest 2 different equations for low and high turbidity water:

$$k_{PAR} = \frac{4.6051 \times K_{D490}}{6.07 \times K_{D490} + 3.2} \; if \; K_{D490} \leq 0.115 m^{-1}$$	Eq. 2.36a
$$k_{PAR} = 0.81 \times {K_{D490}}^{0.8256} \; if \; K_{D490} > 0.115 m^{-1}$$	Eq. 2.36b

In this equation set, the first one corresponds to the clear water and the second to the turbid water.

To calculate L2 for the GOT region, we use the monthly means of k_{D490} for 9 years (2003-2011) from the GLOBCOAST [1]project (Jamet and Dessailly, 2016) and distributed by IFREMER. k_{D490} is calculated using the data from the MEdium Resolution Imaging Spectrometer (MERIS) sensor, installed on Envisat-1 environmental research satellite, using the method described in Jamet et al. (2012). **Figure II.7** shows an example of k_{D490} in March and September, 2010. In general, k_{D490} is higher in March, both at the center of the gulf and in the coastal areas. In September, near the RR, the area with high value of k_{D490} is advected

[1] https://sextant.ifremer.fr/Donnees/Catalogue#/metadata/2b192671-1129-409c-adae-e2335759f290

further offshore. In both months, k_{D490} is high near Hainan Island, indicating that there is a source of high turbidity water in both seasons. The high turbidity near the Hainan island in **Figure II.7** can come from its small rivers or from the vertical mixing created by tides (as suggested for instance by Nguyen et al., 2014). **Figure III.2** (in Chapter 3), shows a tongue of cold water near the Hainan island, which is consistent with such hypothesis (see discussion in III.2.1.2).

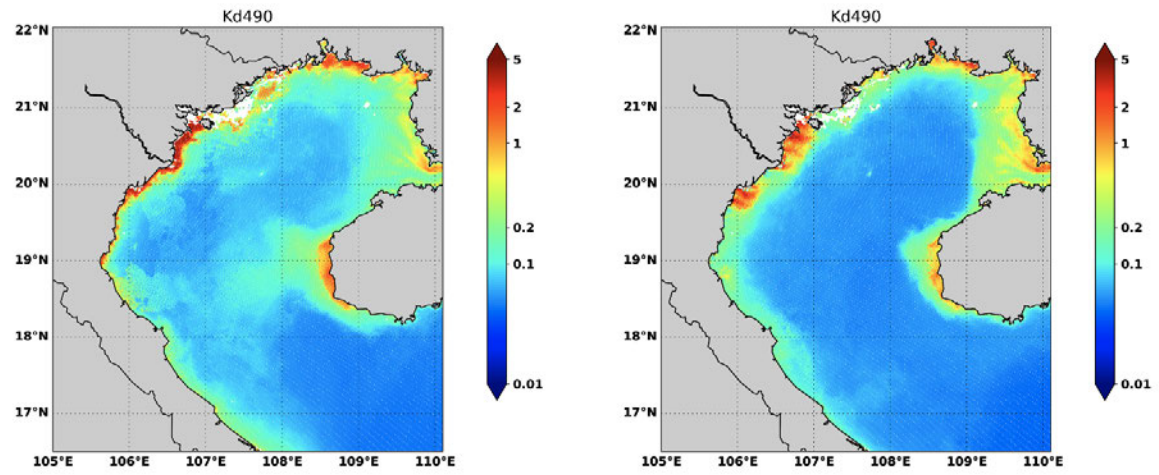

Figure II.7 An example of monthly averaged K_{D490}, downloaded from Ifremer, in **(A)**: March 2010, and in **(B)**: September 2010. Lower values means that the ocean is clearer. Units in m^{-1}.

k_{D490} is interpolated into the simulation grid. Then, k_{PAR} is calculated from k_{D490} by using Eq. 2.36 and then averaged annually, as shown in **Figure II.8**. The calculated L2 here is much shorter than in the clear ocean water (14m maximum compared to 23m) with the presence of regional variability. In the coastal area with river activity (bathymetry < 20m), L2 can fall under 4m while in the deepest area it can increase up to 14m. This spatially varying L2 will be used in the reference simulation. The sensitivity of this change will be assessed in Chapter 3.

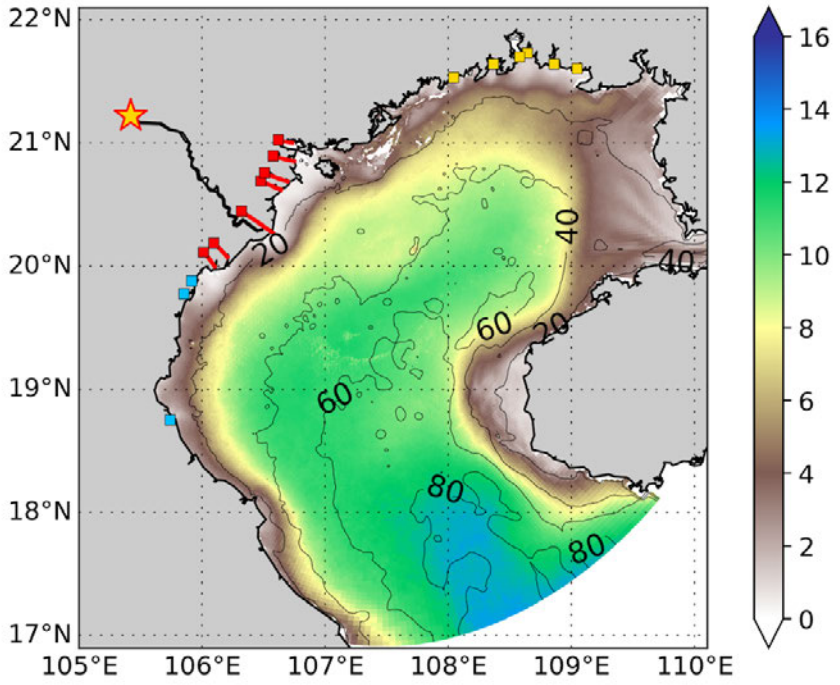

Figure II.8 L2 (m) calculated from KD490 (in color). The isolines show the bathymetry.

II.2.4 Carrying out the numerical simulations

The workflow regarding the run and analysis of the simulation is as follows:

- Run the model
  - Set up the environments (since the model has been run on different machines: HILO at USTH, NUWA at OMP, CALMIP at University of Toulouse).
  - Update the model source in case of using new functions and/or change the parameters that are hard-coded in the source.
  - Compile the model.
  - Update the model parameters depending on the purposes.
  - Run the model
  - Repeat the whole process with each simulation.
- Assess the model with observational data
  - Gridded data: interpolate the model results to the observational data grid if the spatial resolution of the data is lower than the model result.
  - Data on points: find the location of the model grid points which are nearest to the location of the data; usually the model-data comparison is made for several neighboring model grid points in order to account for representativity errors due for instance to errors in bathymetry.
  - Do the calculations

Table II.2 lists all the multiyear simulations used for the analyses in the next chapters. The simulation contains all the settings as described in II.1, II.2, and with spatially varying L2, as in II.3, called GOT_REF. Three other simulations are performed over the same period:

– GOT_23 is a sensitivity test that we used to assess the impact of the new L2 parameterization.

– a twin simulation without river forcing (GOT_NORIV) to assess the impact of the river runoff on the coastal circulation; all other forcings and parameters being the same as in GOT_REF.

– a twin simulation without tides (GOT_NOTIDE) to assess the impact of tides on the main patterns of the river plume variability; all other forcings and parameters being the same as in GOT_REF.

Besides, two ensemble simulations are performed to assess the impacts of the uncertainty of atmospheric forcing on the ocean variables in general and on the river plume in particular. The details of the ensembles will be given in Chapter 5.

Table II.2: List of simulations

Name	Features	Computing specifications
GOT_REF	Reference configuration	Each simulation is run with 168 CPUs on NUWA. CPU time to run GOT_REF: 21 000 hours Disk space for GOT_REF: 2 500 GB. Similar specs for GOT_23, GOT_NORIV, GOT_NOTIDE
GOT_23	Same as GOT_REF but with L2 = 23m	
GOT_NORIV	Same as GOT_REF but there is no river.	
GOT_NOTIDE	Same as GOT_REF but without tide.	
ENS_REF	2 sets of the ensemble simulations. Details will be presented in Chapter 5.	Each simulation is run with 180 CPUs on CALMIP. CPU time to run ENS_REF: 75 000h Disk space for ENS_REF: 8 600 GB.
ENS_0.3		

To configure these simulations above, several preliminary experiments had been carried out. I spent my first 9 months of my PhD in France, where I ran the model and tested the new L2 on the local cluster at the Observatoire Midi-Pyrénées (NUWA system). The result of this new parameterization will be assessed in Chapter 3. Then, I came back to Vietnam for 1 year. At that time, we decided to work with an updated version of the code with respect to the version of Piton (2019), which was ~3 years old. However, the connection to France was very slow so I decided to configure and re-run all new simulations on the local cluster at USTH (HILO system) in Vietnam, and redo all the analysis. At that stage, the plume was identified not by using tracers but by using the SSS. When I came back to France for the last year of PhD, we realized that using SSS to calculate the plume area may lead to an overestimation of the plume which will be discussed in Chapter 4. Therefore, we decided to use the passive tracer to get better results. It again took lots of time to rerun all the simulations with two different boundary conditions of the passive tracer. We also ran some experiments on the regional computing center (CALMIP) in Toulouse. In Table II.2 the list may look very concise, but in the end, more than 20 full time tests were performed which consumed more than 650 000 CPU hours, let alone other small tests. It is equal to around 74 years if we only use 1 CPU.

II.3 Observational data sets used to evaluate the simulations

Before analyzing the results, the model is validated with various observational data, including the salinity and temperature data from CTD and satellite, the tide from tidal gauges and satellite altimetry, and the near surface current from the HF radars. In this section, I will present the description of these data. The model validation itself is presented in the next Chapter.

II.3.1 In situ data

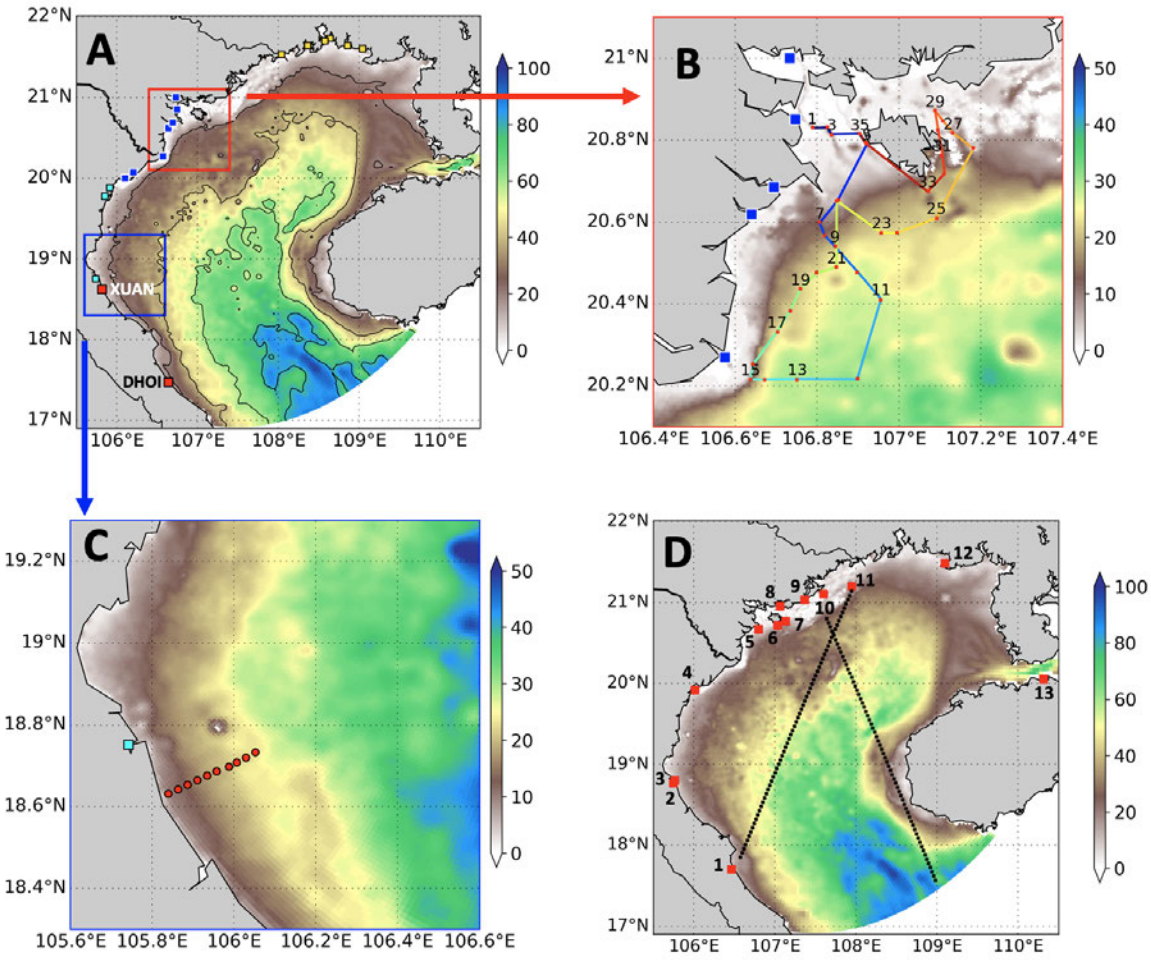

Figure II.9 **(A)** Model domain with bathymetry (m, in color). Blue boxes show the location of river mouths of Red River. Cyan and yellow boxes show the location of other river mouths in the south and north of RR, respectively. XUAN and DHOI are the locations of the HF radars. **(B)**: Location of VITEL CTD stations. **(C)** Location of CFO CTD stations (Red dots). **(D)** Tidal data. Red dots with black numbers: Location of the tide gauges from which historical data of tidal constituents are available. Station 2 and 3 stay nearly at the same location but measurements were made over different periods of time. Black dots: Location of the tracks from the T/P and Jason series altimetric missions from which tidal constituents are estimated.

We use temperature (T) and salinity (S) profiles measured at 35 CTD stations (**Figure II.9B**) during the VITEL cruise which took place in July 2014 (Ouillon, 2014). We also use CTD measurements of T, S acquired repeatedly at 10 stations along a 25 km cross–shelf section (**Figure II.9C**) by Vietnamese and US teams from Center for Oceanography (CFO) and Oregon

State University (Rogowski et al., 2019). This dataset, hereinafter referred to as CFO data, consists in 20 timeframes between September 2015 and July 2016.

We also compare the model-simulated tidal amplitude and phase with historical tidal measurements from 13 stations (whose location is indicated in **Figure II.9D**). The dataset stems from the International Hydrographic Organization (https://www.iho.int/).

II.3.2 High–Frequency radar measurements

We use surface velocity data from the high–frequency radar (hereafter HFR) system based on two antennas located at 18.62°N (XUAN site) and 17.47°N (DHOI site) (Figure II.8A) along the coast and operated by the Center for Oceanography, Vietnam Administration of Sea and Islands (CFO, VASI). The data consists of daily maps of zonal and meridional components of the surface current, over the year 2015, built and provided by Tran et al. (2021). As explained in Rogowski et al. (2019), the summer coverage is lesser than in winter because of low sea state conditions; therefore uncertainties on the interpolated velocities are larger in summer. Comparisons with in–situ measurements over a 12–day period indicate a mean bias of 3 cm/s and an RMS difference of 10 cm/s (Tran et al., 2021); we use these values as rough estimates of the data uncertainties. HFR data is representative of currents at 2.4m below the surface (Tran et al., 2021); as a consequence, for the model assessment, we estimate the model current at 2.4m before interpolating it over the HFR grid.

II.3.3 Altimetric data

Tidal constituents computed from satellite altimetric data provide a rich dataset to evaluate the simulated tides offshore. Along–track amplitudes and phases are calculated from the long time series of sea surface height obtained from 20 years of satellite altimetry (TOPEX/Poseidon, Jason 1–2) by using harmonic analysis. We use the dataset described in Lyard et al. (2021). Location of altimetry tracks is shown in **Figure II.9D**.

The daily L4 product (hereafter AVISO) distributed by the Copernicus Marine Services is used to evaluate the modeled sea surface height (SSH) . It is created with respect to a twenty-year mean from several altimeter missions: Jason-3, Sentinel-3A, HY-2A, Saral/AltiKa, Cryosat-2, Jason-2, Jason-1, T/P, ENVISAT, GFO, ERS1/2. The data can be assessed from the link below:

https://resources.marine.copernicus.eu/product-detail/SEALEVEL_GLO_PHY_L4_MY_008_047/INFORMATION

For this assessment, the sea level anomaly (SLA) is computed by subtracting the mean SSH over the modeling period (2011-2016) from the daily SSH of the AVISO. Similarly, the mean simulated SSH is subtracted is removed from the daily simulated SSH.

II.3.4 Other satellite data

SMOS (Soil Moisture and Ocean Salinity) L3 v04 provides global sea surface salinity (hereafter SSS) with a resolution of 0.25 degree every 4 days. More information about the SMOS product and data availability can be found here:

https://www.catds.fr/Products/Available-products-from-CEC-OS/CEC-Locean-L3-Debiased-v5

OSTIA (Operational SST and Ice Analysis) use data from several satellites to create and provide global sea surface temperature (hereafter SST) in a regular grid at 0.054-degree resolution. The

data is available every day. More information about OSTIA can be found here: https://podaac.jpl.nasa.gov/dataset/UKMO-L4HRfnd-GLOB-OSTIA

II.4 K-means clustering analysis

Many different methods exist to analyze the main characteristics of time series, such as statistical, spectral and classification methods. Clustering analysis is a method to discover the structure of data. For example, one has a set of data, each data is described by a set of features. The objective is to group the data into clusters so that the elements within a cluster have a high degree of "natural association" among themselves while the clusters are "relatively distinct" from one another (Anderberg, 1973). It was used in many areas, from life science (e.g. biology), earth science (e.g. remote sensing), to computer science (e.g. pattern recognition).

There are several algorithms for clustering analysis. For example, Hande et al. (2012) use K-means algorithm to identify 4 wind regimes over Macquarie Island. Weidle et al. (2013) applied Complete Linkage algorithm to assess two strategies of generating initial perturbations in a Limited Area Model Ensemble Prediction System. Delaval et al. (2021) use Fuzzy clustering to identify six different hydroclimatological trends from wind and runoff data, then use the results to force a model to examine the variability of the river plume in 6 different conditions. The Self-Organizing Maps method is used to describe the river plume patterns by Vaz et al. (2018) for the Tagus River and Falcieri et al. (2014) for the Po River.

Among different algorithms, K–means clustering analysis (KMA) is a popular unsupervised learning method (Hastie et al., 2001) that allows to classify objects (observations, model outputs) into different groups, given a measure of dissimilarity. The aim of this method is to iteratively identify clusters by their centroid (the means) then minimize the distance between each member of the cluster and the centroid of that cluster. It was first proposed by MacQueen (1967) and is used widely in ocean science. Yi et al. (2013) used this algorithm to identify spatial distribution of the eddies in the South China Sea. Hänninen et al. (2021) applied KMA to define the 4 oceanographically relevant groups characterized by polymer type, size and particle type. Solabarrieta et al. (2015) used it to examine the relationship between wind and surface circulation in the Bay of Biscay. They found that most of the current patterns are related to a specific wind pattern in the study area. Chen et al. (2017) identified the area of Pearl river plume by applying the K–means clustering to a summer climatological turbidity image. Sonnewald et al. (2019) used it to classify dynamically similar regions of the global ocean based on the barotropic vorticity.

In this study, we use KMA to identify the main patterns of the plume and their temporal variations using daily model outputs. To do this, we use the KMA as implemented in the scikit–learn library coded in python (Pedregosa et al. 2011).

The detail process of the KMA is shown in **Figure II.10**. Firstly, the river plume surface area is identified using a threshold of tracer concentration. Then, it is converted to a binary format, i.e. each point of the model grid is set to 1 (0) for the tracer concentration higher (lower) than this threshold. This is similar to what Falcieri et al. (2014) did for their EOF analysis of the river plume based on SSS. Finally, the KMA is applied to this binary description of the plume.

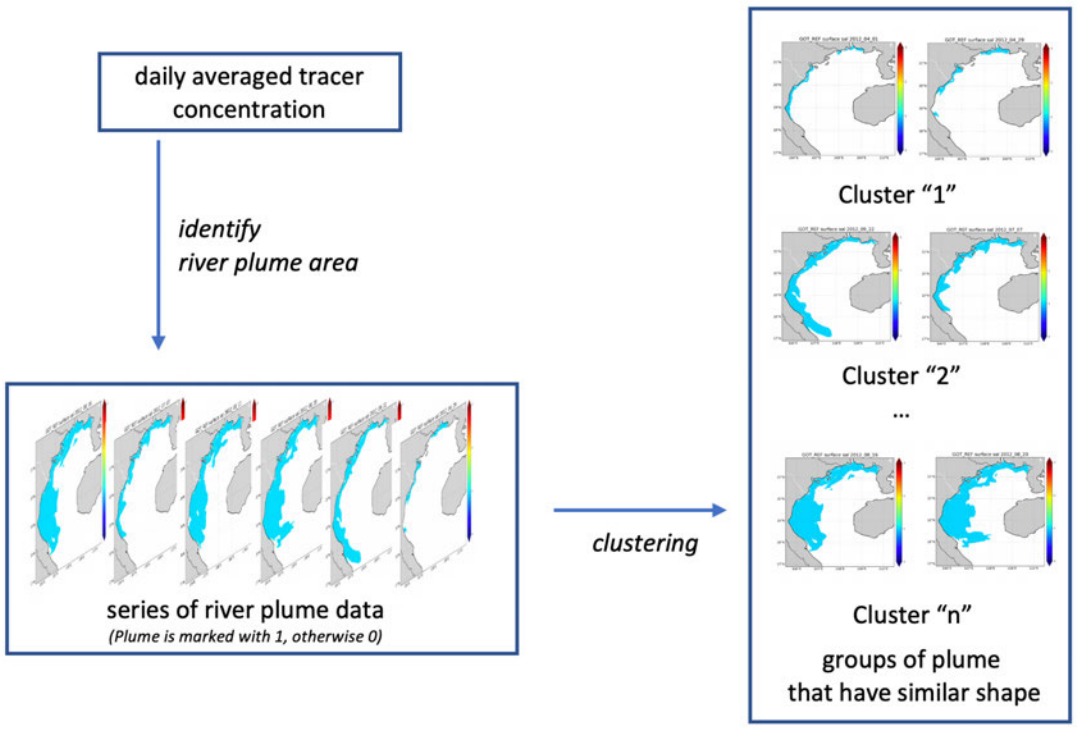

Figure II.10 Process of clustering the river plume in GOT.

In KMA, the number of clusters is defined by the analyst. Therefore, one problem is that which is the appropriate number of clusters. In some studies, the number of clusters is the authors' choice without further explanation (Chen et al. 2017), or it is chosen after a sensitivity test with the number of clusters (Solabarrieta et al. 2015, Hande et al. 2012), or with the help of some indicators (Hänninen et al. 2021, Sonnewald et al. 2019, Yi et al. 2013). Anderberg (1973) and Guanche et al. (2014) share the same view that the number of clusters must be decided by the users according to their experiences because the results of the clustering analysis is linked with the structure of the data (e.g. the variability of the plume should be explained by the variability of the related forcing conditions). In this study, the result of the KMA, as well as the explanation of choice on the number of clusters, will be presented in Chapter 4.

CHAPTER III: MODEL ASSESSMENTS AND CALIBRATIONS

***

III.1 Introduction

To assess the model's performance, we compared it with various datasets (as described in Chapter 2). These include satellite data (altimetry for tide and sea surface anomaly, OSTIA for sea surface temperature, SMOS for sea surface salinity), in situ data (VITEL, CFO, both with temperature and salinity profiles), and HF radar (for surface current). We also compare our results with the OGCM (the global circulation model) to emphasize the added value of our high resolution regional model compared to a global model. Some of the comparisons have been done by Violaine Piton in her thesis (HF radars, OSTIA, SMOS) with a different version of SYMPHONIE and different light penetration parameterization. Therefore, before doing any analysis, we will assess the model with these data again. Most of the assessments have been published in the article in Chapter 4. These include the assessment of tides with respect to tide gauges and along-track altimetric data, of surface currents by comparing with HF radar data, and of temperature and salinity profiles close to the delta with respect to in situ data (VITEL CTD). Here I only introduced the remaining assessments that have not been published.

III.2 Assessment of simulated temperature and salinity in GOT_REF

III.2.1 Comparison with satellite data

III.2.1.1 Comparison with SMOS data

Figure III.1 shows the comparison between GOT_REF and SMOS sea surface salinity (SSS) data (please refer to Chapter 2, section II.3.4 for more information about SMOS). In the winter, from December to January, the space-time average of SMOS SSS is 32.78. OGCM (the global simulation that we use as boundary condition) and GOT_REF show mean values of 32.90 and 32.95, respectively: the mean bias between the data and models is therefore small. In all cases, there is a low salinity region near the Hainan strait, which is probably created by the runoff of the Pearl River in the north (outside the GOT_REF domain). However, in SMOS, this low salinity water extends to the south (**Figure III.1A**), while in OGCM and GOT_REF it extends to the north along the coast and is connected with the other river plumes (**Figure III.1C, E**). The pathway shown in OGCM and GOT_REF is more consistent with the cyclonic current in the winter described in previous studies (Ding et al., 2013; Gao et al., 2014; Rogowski et al., 2019; Piton et al., 2021). It is also possible that the SSS is too high in the OGCM which is used for initial and open-boundary conditions, leading to the overestimated SSS with respect to SMOS data in the center of the basin.

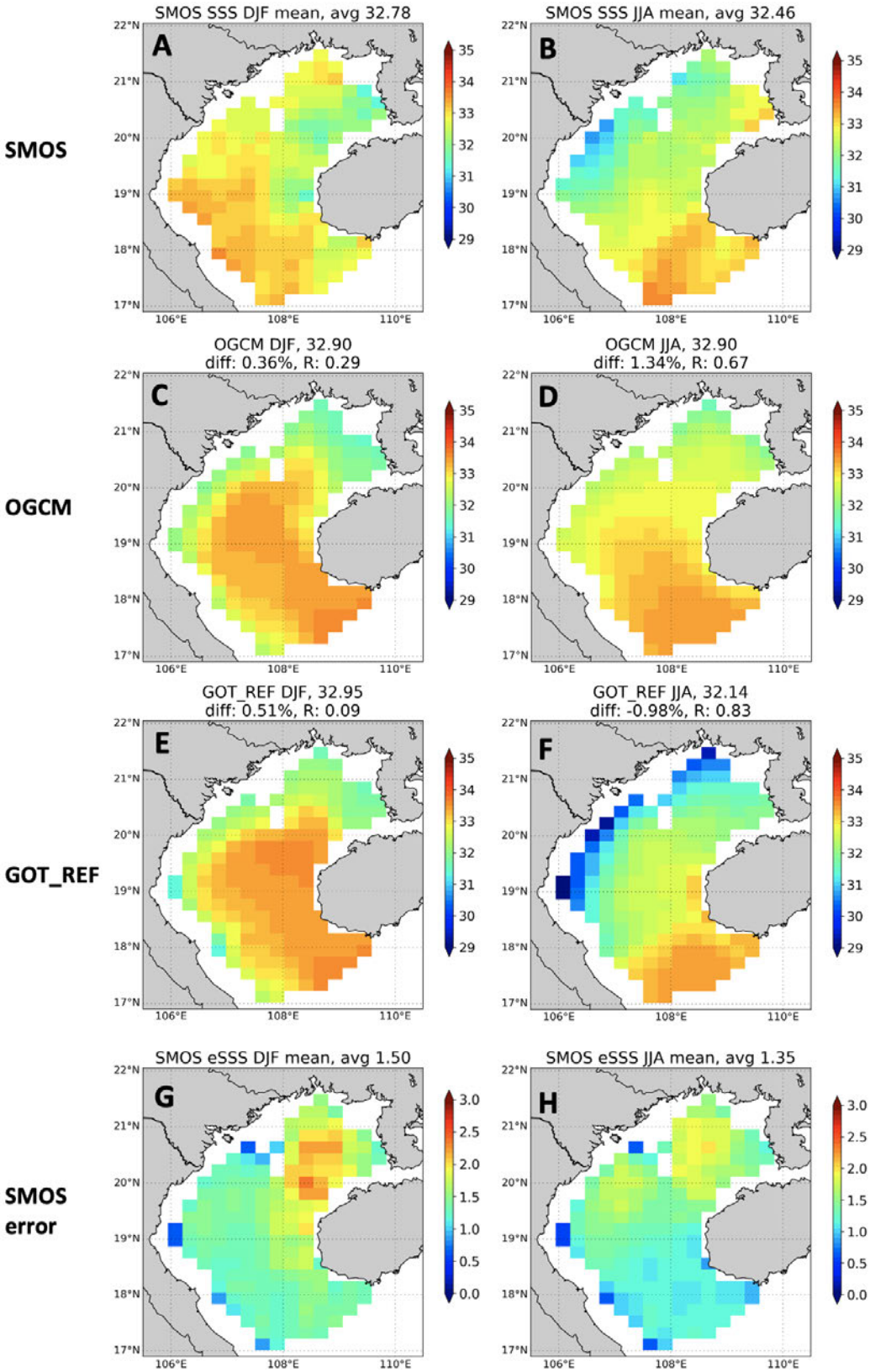

Figure III.1 **(A)**: Sea surface salinity from SMOS satellite, averaged from December to January, from 2011-2016. **(B)**: Same as **(A)** but averaged from June to August. **(C, D)**: Same as **(A, B)** but for OGCM. **(E, F)**: Same as **(A, B)** but for GOT_REF. **(G, H)**: mean error of SMOS, calculated from the same period as **(A, B)**.

In summer, the mean SSS values for SMOS, OGCM and GOT_REF are 32.46, 32.9, and 32.4, respectively. SMOS shows the low salinity pattern created by the Red River runoff and higher salinity near the southern boundary. This SSS gradient is well represented in GOT_REF, while in the OGCM, the salinity is the lowest in the north of the gulf and the low salinity area due to the Red River and other Vietnamese rivers runoff is not represented. Following a discussion with the MERCATOR-Océan team who produced the global OGCM, we know that the Red River runoff is misrepresented in this configuration, leading to an overestimation of the SSS along the Vietnamese coast.

In GOT_REF, the area associated with the plume shows salinity values that are much lower than those of SMOS. This could indicate an overestimation of the runoff or of the plume offshore and southward extension. However, we believe that this is likely due to increased errors in the SMOS observations in this area instead, since the core of the plume is close to the coast where SMOS data are unavailable. The accuracy of the measurements at the closest point to the coast may not be as good as further offshore. We were surprised by the error patterns provided with the SSS maps (**Figure III.1 G, H**); these error maps are empirically derived from the observed spatial SSS variability and might be underestimated close to the coast where the SSS data are partially extrapolated (J. Boutin, pers. comm.).

III.2.1.2 Comparison with OSTIA data

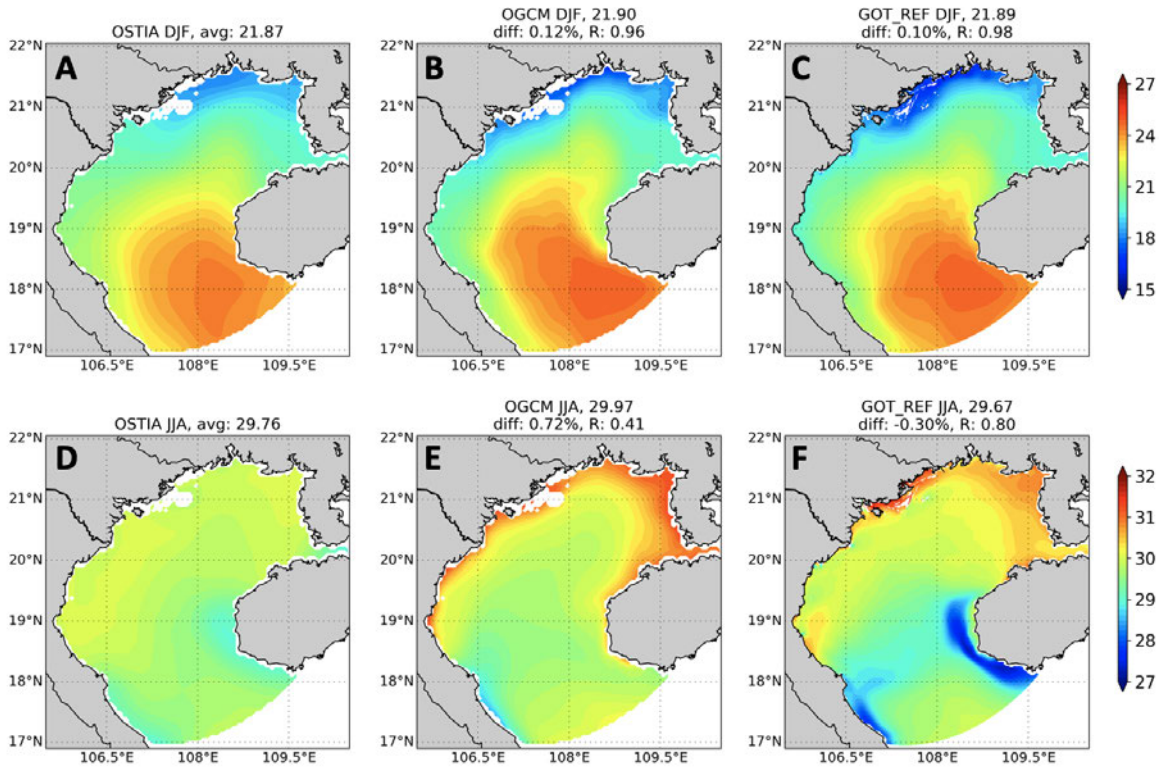

Figure III.2 **(A)**: OSTIA sea surface temperature, averaged from December to January in the period of 2011-2016. **(B)**: Same as **(A)** but for OGCM. **(C)**: Same as **(A)** but for GOT_REF. **(D, E, F)**: Same as **(A, B, C)** but calculated from June to August. Unit in °C.

The comparison between GOT_REF and OSTIA data is shown in **Figure III.2** (please refer to Chapter 2, section II.3.4 for more information about the OSTIA product). In the winter, the

temperature in OSTIA increases southward, from ~19°C in the Chinese coast to 25°C near the southern boundary (**Figure III.2A**). Both OGCM and GOT_REF show the similar pattern, but the values are slightly underestimated in the north (**Figure III.2B, C**).

In the summer, the spatial variations in OSTIA are smaller. Over the whole area, the temperature varies around 29-30°C, with lower values near the Vietnamese coast south of 18°N and south of Hainan island (**Figure III.2D**). Compared to OGCM, GOT_REF simulates lower temperatures, which is more consistent with the pattern in OSTIA. In detail, the low temperature area near Hainan island is presented. It seems to be created by tides, as suggested by Nguyen et al. (2014) and Piton et al. (2020). Indeed, this pattern does not appear in the GOT_NOTIDE (not shown). However, the temperature in that area is underestimated by around 2°C, suggesting that the tidal mixing may be overestimated. Overall, performance of GOT_REF is still better than the OGCM, as shown with higher spatial correlation (0.80 compared to 0.41, corresponding for **Figure III.2F** and **Figure III.2E**).

III.2.1.3 Comparison with SLA

The daily averaged SLA from AVISO data and from GOT_REF is shown in Figure III.3. In the winter (from December to February), the SLA increases from the north (~3 cm) to the south (~13 cm). Near the RR coast, the SLA is around 6 cm (**Figure III.3A**). Both OGCM and GOT_REF show the increase from the north to the south, but GOT_REF has lower SLA than OGCM. Compared to AVISO, GOT_REF shows better results in the general spatial pattern, especially near the RR coast (**Figure III.3C**). The difference is located mostly near the southern boundary, where it reaches around 13cm for AVISO and 11cm for GOT_REF.

In the summer, (from June to August), the SLA shows a reverse spatial pattern, with higher value at the north (-3 cm) and lower value at the south (-16 cm) (**Figure III.3D**). GOT_REF shows closevalues at the north, but near the southern boundary it is overestimated by ~ 1 cm (**Figure III.3F**). Near the RR coast, around 20N, GOT_REF shows a similar spatial variability than in the AVISO data. Around 19N, GOT_REF is underestimated by ~ 4 cm, but it is much better than OGCM (underestimated by ~ 8cm).

The AVISO data used in this study is computed from multi-year along-track data, and interpolated to the global grid map. Therefore, it can contain errors far from the tracks of the altimeter. Besides, altimetric data are expected to be less accurate in coastal areas because of instrumental errors and larger uncertainties in the wet tropospheric and geophysical corrections (tides, surface waves). In the future, it would be more relevant to use the along-track data.

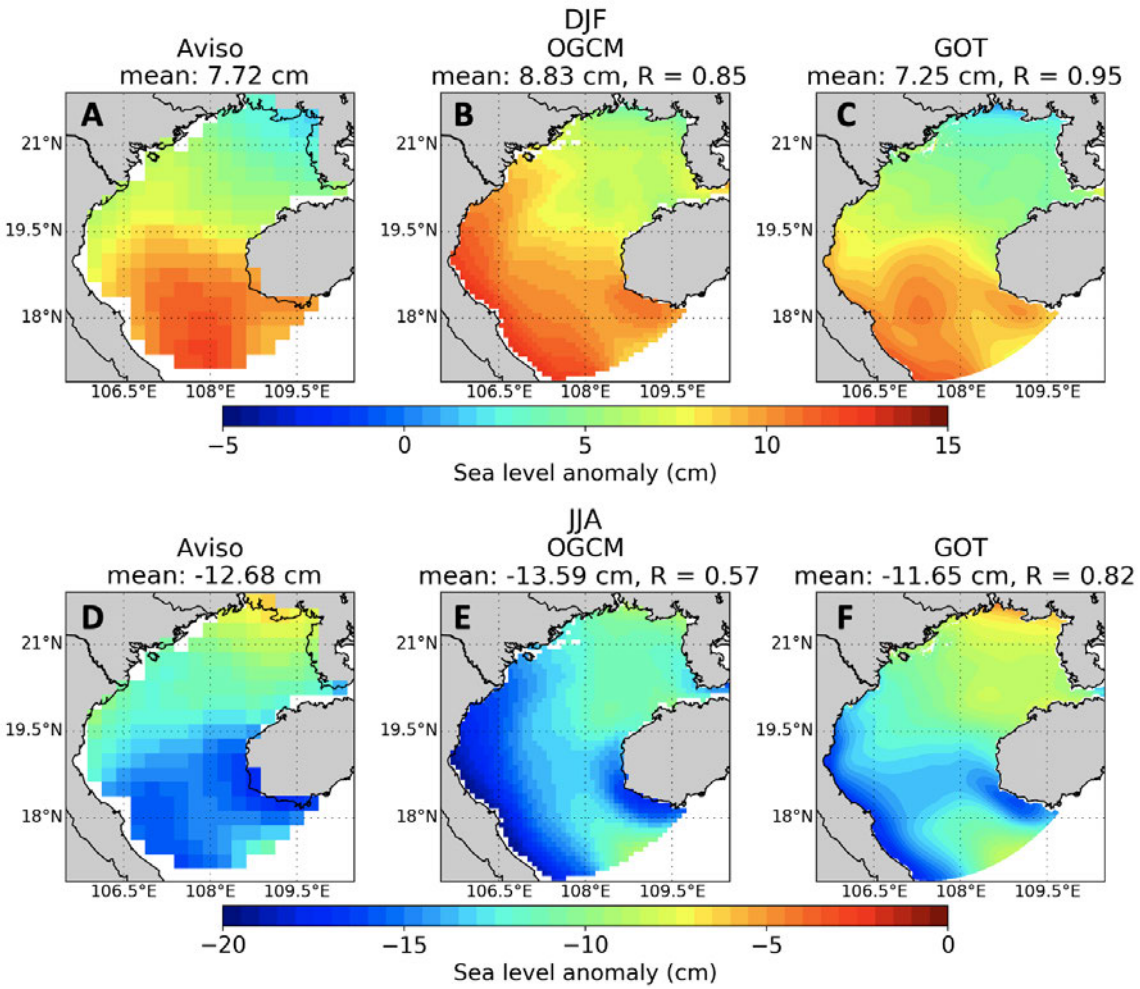

Figure III.3 (A): Daily mean SLA computed from December to February in the period 2011-2016. (B): same as (A) but for OGCM. (C): same as (A) but for GOT_REF. (D, E, F): same as (A, B, C) but computed from June to August.

III.2.2 Temperature and salinity profiles from in situ measurements

We compare vertical T, S sections from CFO profiles (**Figure II.5**) and the corresponding GOT_REF sections for four dates representative of different discharge and monsoon conditions (**Figure III.3**). The description of the CFO data can be found in the Chapter 2.

In September, S shows a strong stratification, with values lower than 27 at the surface increasing to over 33 at 5m depth. The model shows the same pattern, with very low bias (0.2 in absolute value) and high spatial correlation (R = 0.91). T is stratified as well. For both T and S, the model shows a slightly stronger stratification, resulting in more saline/colder values in the deepest part of the section. In January, S is mixed over the whole depth with an average salinity of 30 in the first 20km from the coast and is weakly stratified offshore. The model underestimates S in the first 5 km (S ~28) and does not reproduce the stratification offshore. The same behavior is found for T. In April, S is stratified again throughout the section, but S is higher than in September by 1.4 on average. The model shows a similar pattern (R = 0.91), but S is overestimated by 1.2 on average. Similarly, the correlation is high for T (R = 0.96), but the model overestimates it by 2.2°C. Since the strength of the stratification of the GOT_REF and the data is similar and the

bias is similar over the whole water column, this overestimation can be related to the uncertainty of the boundary. In July, S decreases significantly throughout the section in both data and model. In the data, the surface layer is about 10m deep, with a mean salinity of ~32 and a minimum value obtained at about 15km from the coast. The model surface salinity is more uniform: the overall correlation is 0.62. In contrast, the model temperature matches well with the data (bias = 0.3°C).

Over all 20 stations available, the mean biases (absolute value) between data and model are 0.61 and 0.71°C for S and T respectively, and the mean correlations are 0.81 and 0.72. In April and July, the comparisons suggest that the model runoff is not high enough (as already suggested by the comparison with the VITEL data). In April, the temperature and salinity are higher in the model in the deepest region of the section, which can be due to a bias in the salinity and temperature imposed at the open boundary condition in the south. The vertical mixing seems underestimated also. Overall, despite having biases in some days, the model reproduces accurately the seasonal conditions even in the very shallow area.

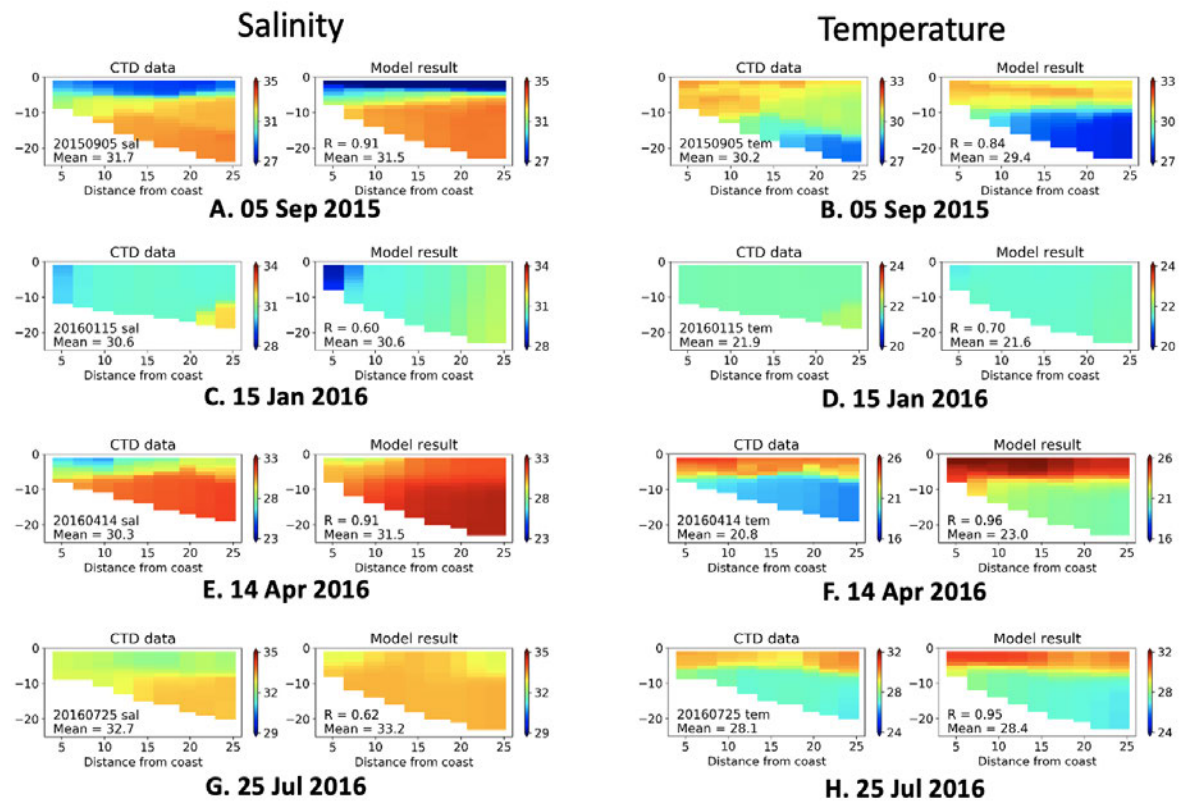

Figure III.4 **(A)**: Vertical salinity sections from CFO CTD data (left) and model result (right) on 05 September 2015. **(B)**: same but for temperature. **(C, D), (E, F), (G,H)**: same as **(A, B)** but for 15 January 2016, 14 April 2016, and 25 July 2016, respectively.

III.3 Assessment of the new L2 parameterization

In the initial configuration of Piton et al. (2021), the L2 parameter (shortwave attenuation length) is equal to 23m all over the domain (for more information please refer to Chapter 2, section II.2.3). This is a standard value in the open ocean. However, it should be noted that in the coastal area where the sediment load transported from the river is high, L2 should be shorter than in the open ocean. Therefore, I tried to improve the model accuracy using a new L2 parameterization,which leads to values lower than 4m near the coast to around 14m in the deepest area, as described in Chapter 2. In this section, we will first assess its effect on the temperature and salinity distribution and then check if it can produce better results by comparing them with observational data.

III.3.1 Effect on temperature

We compare the spatially averaged bottom and surface temperature. Indeed, since GOT_23 has a longer L2, it should simulate higher (resp. lower) temperature at the sub-surface and bottom (resp. surface) compared to GOT_REF.

At the bottom, the temperature exhibits a seasonal variation. In the area where the depth is less than 30m, the temperature varies from around 17°C in February to around 30°C in September (**Figure III.5A**). GOT_23 shows higher temperatures indeed, but the difference between the two runs also has a seasonal variation. From November to February, the difference is very small (< 0.1°C) because the water is fully mixed. From April, the difference increases and reaches the maximum in August (0.8°C on average).

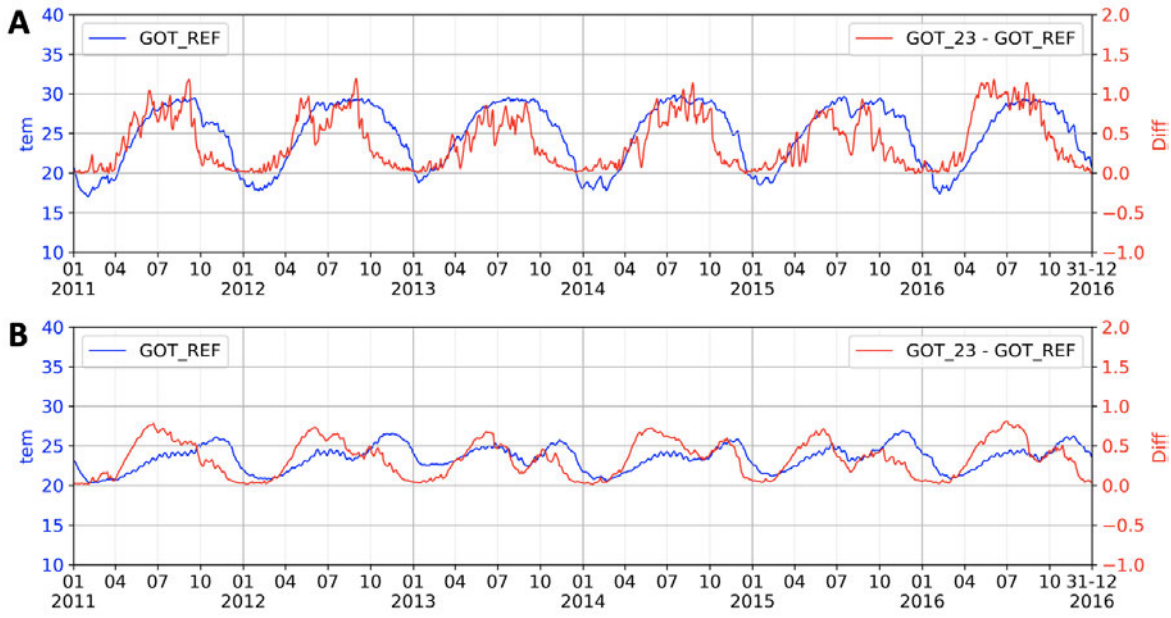

Figure III.5 (A): Bottom temperature (°C) of GOT_REF, spatially averaged where the depth is less or equal to 30m (blue line, left axis); difference between GOT_23 and GOT_REF (red line, right axis). **(B)**: same as **(A)** but the averages are computed over the region deeper than 30m.

In the area deeper than 30m, the seasonal variation of the bottom temperature is smaller, from ~20 to 27°C (**Figure III.5B**). The difference is again very low from November to February, and it reaches the maximum value in June (0.69°C on average). The seasonality of the impact

of L2 on the temperature may be directly due to the seasonal variation of the solar radiation itself; the latter is the largest from April to July (Fig. III9A).

At the surface, the impact of the new L2 value on the temperature is similar between the shallow and deeper regions (not shown) so we just show the comparison of the SST averaged over the whole domain (Fig. III.6). The temperature is lowest in February (~20°C) and highest in August (~30°C). At the surface, most of the time, GOT_23 has a higher temperature than GOT_REF, as expected. From November to January, the difference is less than 0.1°C. Then, the difference increases in March. In April and May, the SST of GOT_23 is 0.3°C lower than the SST of GOT_REF. Overall the impact of the new L2 is larger at the bottom than at the surface.

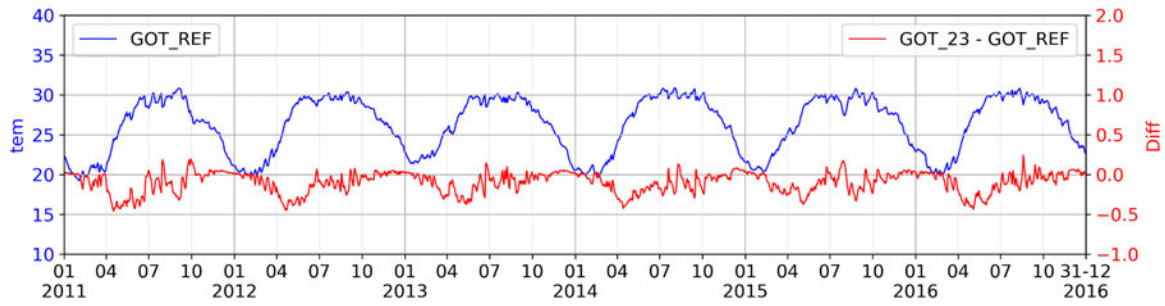

Figure III.6 Surface temperature (°C) of GOT_REF over the whole area (blue line, left axis); difference between GOT_23 and GOT_REF (red line, right axis).

Next, we check the impact of the new L2 on the mixed layer depth (MLD). Here, the MLD is defined as the depth at which the density of seawater is 0.03kgm^{-3} greater than surface values (Zhou et al. 2015). We computed the mean MLD from December to February and from April to September over the period 2011-2016. In GOT_23, the MLD ranges from 2m near the coast to around 8m in the deep region in summer (**Figure III.7A**), while it reaches 30 to 50m in the deep region in winter (not shown).

It should be noted that with a higher value of L2, the mixed layer should be deeper due to the deeper penetration of the irradiance (Zhou et al. 2015). **Figure III.7** shows that in both runs the MLD has the same spatial distribution, with slightly higher MLD in the GOT_23. This is also true in winter (not shown) although the difference between GOT_REF and GOT_23 can reach larger values (10m locally). However, there is no direct link between the change of MLD and the change of L2: while L2 in GOT_REF is less than 4m in both the Red River coastal area and near the Hainan strait, the difference of MLD between GOT_REF and GOT_23 is less than 0.2m in the west, while it can reach more than 1m near the Hainan strait (**Figure III.7C**). This is different from the result of Zhou et al. (2015). In their study at the global scale, the MLD difference has the same regional patterns as the L2 difference. We propose the following explanation for this different behavior: the impact of L2 on MLD is not significant in the regions influenced by the river freshwater discharge where the MLD is mostly controlled by the low surface salinity.

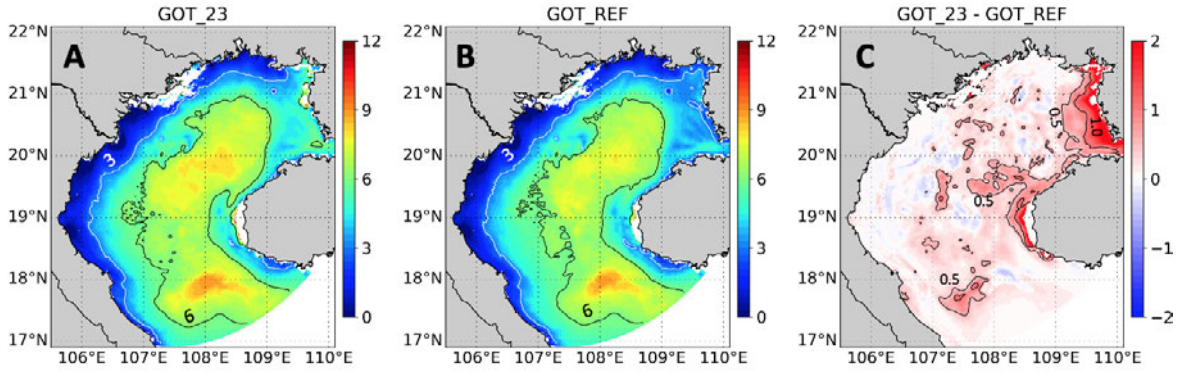

Figure III.7 **(A)**: Mixed layer depth of GOT_23, calculated from the mean density profile from April to September in the period 2011-2016. **(B)**: same as **(A)** but for GOT_REF. **(C)**: the difference between **(A)** and **(B)**. Units in m.

Finally, to check the performance of the new L2, we compared the bias and correlation of GOT_23 and GOT_REF with respect to the VITEL and CFO CTD data. Table III.1 shows that the new L2 improves both the mean bias and the mean of the spatial correlation. Using the new L2 values, the mean bias decreased from 0.82 to 0.61°C for VITEL, and from 0.89 to 0.81°C for CFO. The correlations (R) are improved as well. The comparison with VITEL shows the increase in correlation from 0.57 to 0.71°C, and from 0.57 to 0.74°C for CFO. **Figure III.8** shows that GOT_REF indeed produces a better stratification compared to GOT_23 with respect to VITEL profiles. Looking into more details Table III.1, we find that the new L2 decreases the maximum biases for both VITEL and CFO, but it also increases the absolute values of the minimum biases.

Since there is no CTD data available offshore, we can not assess the effect of the new L2 in the deep area. However, the comparison with VITEL and CFO shows that the new L2 produces better results in the shallow area (with bathymetry less than 30m), which is the area where the plume is mostly located.

Table III.1 Comparison of GOT_REF and GOT_23 with VITEL and CFO temperature data. Unit in °C

Data source	Configuration	Mean absolute bias	Min bias	Max bias	Mean R
VITEL data	**GOT_REF**	**0.61**	**-1.28**	**2.25**	**0.71**
	GOT_23	0.82	-1.14	3.14	0.57
CFO data	**GOT_REF**	**0.81**	**-1.02**	**1.34**	**0.74**
	GOT_23	0.89	-0.57	1.78	0.57

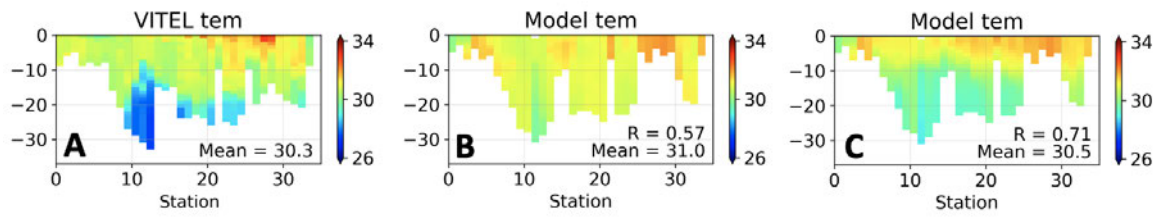

Figure III.8 (A): Temperature profile from VITEL CTD data. Temperature profile from GOT_23 **(B)** and from GOT_REF **(C)**. Unit in °C.

III.3.2 Effect on heat flux

The total heat flux at the atmospheric-oceanic boundary is made up of four components: shortwave solar radiation, longwave radiation, latent and sensible heat fluxes. The ability of the shortwave radiation to go through water depth is characterized by the shortwave attenuation length (L2). Higher attenuation length allows the irradiance to penetrate deeper and vice versa. In GOT_23, since L2 is higher, the solar component of the heat flux can penetrate deeper, therefore the surface temperature is lower than in GOT_REF. However, since the air-sea heat fluxt depends on the SST, a change in SST also leads to a change in longwave, sensible and latent components. **Figure III.9** shows the mean of the different heat flux components over the whole domain in different months, calculated from 2011-2016.

In GOT, the solar flux has seasonal variation, with a minimum value of less than 125 Wm^{-2} in December and the highest value of more than 250 Wm^{-2} in May (**Figure III.9A**). From October to January, the net heat flux is negative for both GOT_REF and GOT_23 (the surface of the ocean is cooling) and the difference of the two simulations is less than 1%. The maximum heat loss occurs in December (-176 Wm^{-2} for GOT_REF). From February to September, the heat flux is positive for both simulations. While the solar flux reached its maximum in May, the maximum net heat gain occurs in April. In May, the net heat flux decreased from 143 Wm^{-2} (GOT_23) to 129 Wm^{-2} (GOT_REF) (10%).

The latent, sensible and longwave heat fluxes from GOT_23 and GOT_REF are shown in **Figure III.9B**. Among the fluxes, the longwave flux has the smallest variation over time. In GOT_REF, the minimum heat loss due to longwave radiation is 35 Wm^{-2} (in March) and maximum is 53 Wm^{-2} (in December). The difference between GOT_23 and GOT_REF is mostly smaller than 1 Wm^{-2}, except in April and May (~ 2 Wm^{-2}). For the sensible heat flux, the maximum heat loss is 42 Wm^{-2} (in December) while the minimum is in May (6 Wm^{-2}). As with the longwave flux, the difference between GOT_23 and GOT_REF is generally smaller than 1 Wm^{-2} except from April to June. The maximum difference occurs in May (2 Wm^{-2}). The latent heat flux contributes the most to the difference between GOT_REF and GOT_23. The maximum difference is 10 Wm^{-2}, in May. In GOT_REF, the minimum latent heat loss is in April (59 Wm^{-2}) and the maximum is in December (202 Wm^{-2}).

As a conclusion, we find that the impact of the new L2 parameterization on the heat flux at the air-sea interface is negligible at the basin scale, which is consistent with the weak impact on

SST. Maps of the seasonal differences between GOT_23 and GOT_REF do not show significant local differences either (not shown).

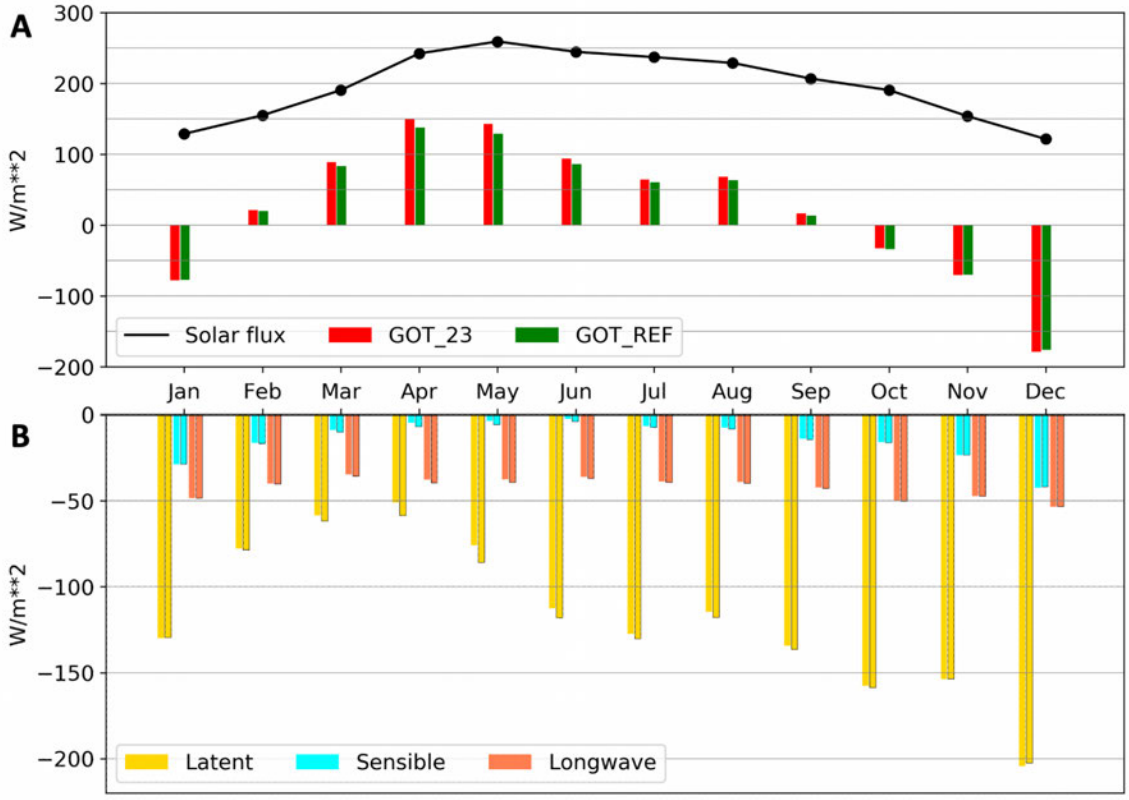

Figure III.9 **(A)** Total heat fluxes in W/m^2 from GOT_23 and GOT_REF, calculated from 2011-2016, averaged over the whole domain. Black line shows the solar flux (solar flux is the same for GOT_REF and GOT_23). **(B)**: Latent, sensible and longwave heat flux from GOT_23 (left column) and GOT_REF (right column). Positive (resp negative) values mean a warming (resp cooling) of the ocean..

III.3.3 Effect on salinity

The RMS of the differences between the daily mean salinity of GOT_REF and GOT_23 over 2011 - 2016 are shown in **Figure III.10** for different depths. The maximum change of salinity is at the surface, reaching more than 1 near the RR coast, and then decreasing quickly with depth. At 20m, the difference is less than 0.2. There is no significant difference between GOT_REF and GOT_23.

The impact of the new L2 on salinity can be due to several reasons. Firstly, it can result from a change in vertical mixing caused by the change of temperature. Comparisons of density profiles close to the coast (not shown) indeed suggest a change in the strength of stratification (even though the MLD does not differ much) but we still need to do more detailed and quantitative analysis to conclude about this effect. Secondly, it can be linked with changes in evaporation induced by changes on SST. However, the impact of L2 on the latent heat flux (the latent heat flux is proportional to the evaporation in the bulk formulae that we use to compute the air-sea heat fluxes) is very small as discussed in the previous section. Besides, the spatial pattern of the impact on the latent heat flux (not shown) is not consistent with the pattern of the impact on the SSS (**Figure III.10**). At last, we assume that small perturbations triggered by the

change of L2 could be amplified in the coastal area because of flow instabilities (for instance at the front created by the plume), leading to the observed *misfit* between GOT_REF and GOT_23. This *misfit* could be seen as the signature of internal variability of the system.

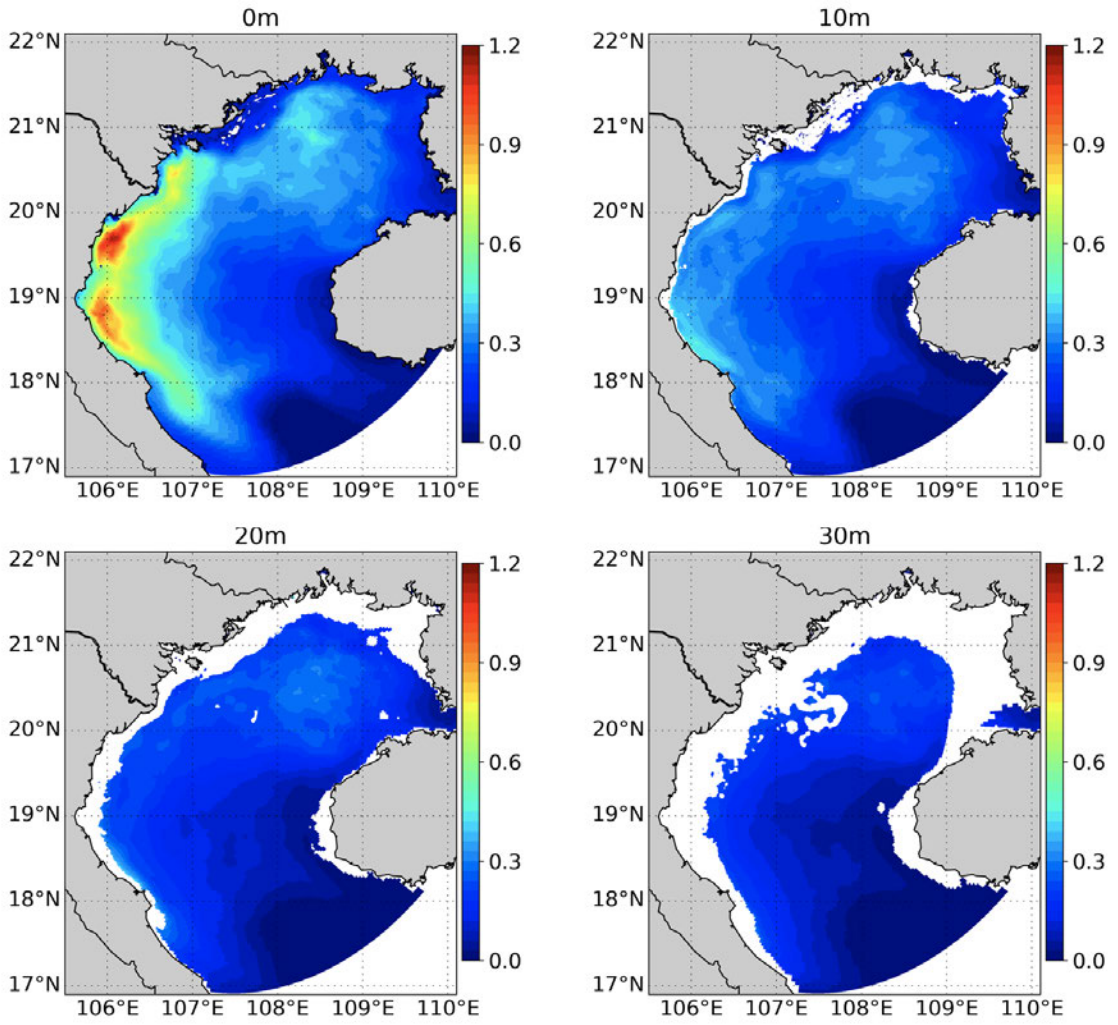

Figure III.10 RMS of the difference between daily salinity fields from GOT_REF and from GOT_23 at different depths computed over the whole period 2011-2016.

III.4 Conclusion

III.4.1 Model assessment

This chapter presents some results of the model assessment that are not included in Nguyen-Duy et al. (2021) (see Chapter 4).

The model is compared to satellite and in-situ data. SMOS, OSTIA and AVISO products derived from satellite data are used to assess the surface salinity, temperature and large-scale sea level anomaly at seasonal time scales. While the model's surface temperature shows good correlation with the observed data, some misfits between the observed and simulated SSS patterns are found in winter. However, this difference can be due to the error of the SMOS product itself rather than the model error. The quality of SMOS data so close to the coast is questionable here.

The comparison with CFO and VITEL in-situ data shows that the model's skills in representing the T,S vertical gradients as well as their seasonal evolution in the coastal region are reasonably good. In spring, we observe a significant bias in both T and S with respect to the CFO profiles; this might come from errors in the large scale field prescribed at the open boundary condition.. Aside from this bias, nothing indicates a major deficiency in the model or configuration.

However, the comparison with VITEL and CFO salinity profiles suggest a surface freshwater deficit. As suggested in Nguyen-Duy et al. (2021) the latter could result from an underestimation of the Red River runoff or an overestimation of the mixing at the river mouths (or in the estuaries) due for instance to uncertainties in tides.

The comparison with HF radar currents, which is not presented in this chapter but in the article of Nguyen-Duy et al. (2021) included in Chapter 4, shows that the monthly variability of the surface currents is consistent with the observations in terms of spatial patterns. In terms of velocity amplitude, significant misfits reaching locally 20 cm/s are observed. They could be due to model errors because of uncertainties in wind forcing or to errors in the HF radar products. In the future, in order to avoid uncertainties due to the HF radar data interpolation onto a regular grid, the model currents should be compared directly to radial velocities.

The evaluation of tides by comparison with altimetric and tide gauges measurements (Chapter 4) for the four main components (O1, K1, M2 and S2) indicates an accurate representation of the tidal signature in sea surface height in the center of the basin (complex error inferior to 3.5 cm). Close to the Red River delta, the comparison with coastal tide gauges leads to larger errors (10.2 cm for K1 and O1) as expected in this area where uncertainties in bathymetry remain. When comparing the model results with the VITEL CTD profiles, we indeed observed some discrepancies between the deepest CTD depth and the one in the model neighboring points (as illustrated in Fig. III.8). Such discrepancies led us to systematically compare the VITEL profiles with several neighboring points of the model field to check the robustness of our comparisons (not shown).

In this study, we used altimetric along-track tidal constituents and seasonal maps of sea level anomaly. In future studies, sea level anomalies along the track should be explored to estimate the geostrophic component of the coastal current. Depending on the quality of the data so close

to the continent, they could also be used to assess the sea level variations in the estuaries and at the river mouth and to analyze the tides, river flow and plume interactions there.

III.4.2 Impact of the new parameterization of the light penetration

In this SYMPHONIE configuration, the spatially varying light penetration depth is calculated from satellite ocean color data to better represent the area under high sediment transport conditions. The model results show that with lower attenuation length, the surface temperature will increase up to 0.5°C locally, while the sub-surface and bottom temperature will decrease (1°C maximum). With lower L2, the MLD is slightly decreased also. The maximum change of MLD happens near the Hainan Strait where there is no river discharge. Due to the increase of the surface temperature, the total heat fluxes can be decreased by more than 10% in spring; the impact of L2 on the heat fluxes is not significant in fall and winter. The heat loss is mostly due to the latent heat flux and the maximum loss occurs in May. Compared to the CTD data, using the spatially varying L2 can create better results than using a single value.

In the framework of this study, we do not use a biogeochemical model. Since the change in temperature and stratification can affect the biological activity, we expect that a change in the L2 can lead to a change in the primary production and the improvement of L2 parameterization could be more meaningful. Also, L2 can vary over time. The calculation with KD490 shows that, in offshore areas, L2 is lowest in the early winter (November, December, January) and highest in the summer. However, in the coastal area, L2 near HaLong bay (near the location of VITEL CTD data) shows the opposite pattern with the highest value in winter and lowest value in summer. Since the difference is large (more than 30%), we expect it can make a significant difference locally in the simulation with high temporal frequency. Therefore, future work could be included to adapt the model to a L2 that varies both spatially and temporally and to investigate its impacts on biogeochemical aspects.

Overall, despite small biases, the model reproduces rather satisfying the conditions for all of the variables that have been checked, and we consider that the model results can be used for further analyses.

CHAPTER IV: VARIABILITY OF RIVER PLUME IN THE GULF OF TONKIN

***

IV.1 Introduction

The Gulf of Tonkin is a shelf sea located east of Vietnam and south of China. In spite of its importance, to date, studies focusing on the plume in the Gulf of Tonkin are still scarce. Therefore, there is a need to better understand the variability of the Red River plume at different scales. In chapters 2 and 3, the method and the validation of the model have been examined. In this chapter, we will attempt to examine the plume variability in the mid-field and far-field regions.

In section IV.2, an article (Nguyen-Duy et al., 2021) published as part of this thesis is reproduced. In this paper, firstly, we introduce a short summary of the study area and model configuration, as written in Chapter 1 and Chapter 2. Then, the model is validated with other datasets that have not been introduced in Chapter 3. Finally, the variability of the Red River plume is presented using the model results and K-means clustering analysis.

During the review process of the paper above, the reviewers raised some questions. Not all of the answers led to us modifying the paper but they are interesting concerns and worth looking at. Also, some sensitivity experiments are performed after the paper is published. These discussions will be presented in section IV.3.

IV.2 Paper published in Frontiers in Marine Science

IV.2.1 Summary of the paper

In the paper, the study area and model configuration is described as a short version of what is written in Chapter 1 and Chapter 2. Then, the model is validated with several assessments. Firstly, the first baroclinic Rossby radius of deformation (Rd) is calculated. This is the first time that the Rd is calculated for our area. Then, we assess the general circulation from our model with results from other studies. The mean surface current in the winter (from November to March) is dominated by the gulf–scale cyclonic gyre. From April to August, there is a weak cyclonic gyre in the center of the gulf and an anticyclonic gyre in the south of the gulf. In September and October, the coastal current is southward again from 21°N to the southern boundary where it exits the gulf. This result is comparable with the past studies from several authors.

The next assessments are the comparisons with the satellite and in-situ data. The comparisons with tidal gauges and altimetry data show that the model result performs well in general in representing the sea level amplitude and phase for the main four components (O1, K1, M2 and S2), though still have some errors when compared to the tidal gauges in the delta. These errors can be explained by the uncertainty of the model bathymetry and is comparable with other studies in this area. Compared with HF radar data, the model shows good results, though still has some misfits.

The next section aimed at explaining the variability of the Red River plume. The river runoff is dyed with the passive tracer to better identify the river plume. From several experiments, the plume is defined as the area where the tracer concentration is equal or higher than 7 arbitrary units/m^3. Then, this criteria is used to calculate the surface plume area. The results show that the plume area follows the same variations as the river runoff, albeit with a time lag. In summer, when the runoff is high, the peak of the plume area occasionally reaches more than 40 000 km^2 (about 27% of the GOT area), while in winter, it sometimes falls below 2 000 km^2. Due to the interannual variability of the runoff and atmospheric conditions, the plume area varies significantly between different years.

To identify the main spatial patterns of variability, K-means clustering method is applied to daily scenes of the plume that were calculated as described above. The cluster analysis identifies the plume regimes and their periods of occurrence without having to pre-define these periods as one would do when computing seasonal averages for instance. We also apply the classification method to the run without tides to evaluate the impact of tides.

From sensitivity experiments, we have chosen to classify the plume patterns into 4 clusters. The first cluster occurs from November to March. In this cluster, the plume is narrow and contained within the inner shelf (bottom depth < 20m), due to low runoff and downwelling wind (winter monsoon) with an intense downcoast current. The second pattern is mainly observed in the transition period of the monsoon (April, May, October) and is wider than the first pattern; both the relaxation of the winter monsoon and the weaker coastal current allow the plume to spread further offshore than in cluster 1. The third pattern occurs when the summer monsoon is the strongest and the runoff increases. The wind is upwelling favorable; the plumes are advected northward and detach locally from the coast. It is the only pattern where the plumes from the

different rivers are disconnected. The final pattern coincides with the highest runoff period and the summer monsoon relaxation. The plumes are connected again in a pool of low salinity waters with the largest coverage, spreading both offshore and southward.

The vertical variability of the plume is examined as well. In winter, the plume is usually mixed over the whole water depth. When the summer monsoon arrives at the end of March, the plume starts to detach from the bottom and spreads offshore as a surface buoyant layer. The plume can detach from the coast near 19°N when a coastal upwelling develops. It may deepen offshore also. In particular, the plume is strongly deepened in the middle of August when trapped in a recurrent eddy near 19°N.

Finally, we tried to identify the “receiving basin” from different rivers. In this study, we define the "receiving basin" for different rivers from the passive tracer content with the same threshold as defined in section 4 (concentration > 7 unit/m^3), but this time we apply it to each river system. The results show that in the low runoff period, the receiving basins from different rivers are very narrow and elongated along the coast, while in the high runoff period, receiving basins of Red River and Southern River are connected.

IV.2.2 Variability of the Red River Plume in the Gulf of Tonkin as Revealed by Numerical Modeling and Clustering Analysis

frontiers in Marine Science

ORIGINAL RESEARCH
published: 12 November 2021
doi: 10.3389/fmars.2021.772139

Variability of the Red River Plume in the Gulf of Tonkin as Revealed by Numerical Modeling and Clustering Analysis

Tung Nguyen-Duy[1,2], Nadia K. Ayoub[1], Patrick Marsaleix[1], Florence Toublanc[1], Pierre De Mey-Frémaux[1], Violaine Piton[1,3], Marine Herrmann[1,2], Thomas Duhaut[1], Manh Cuong Tran[4,5] and Thanh Ngo-Duc[2]*

[1] LEGOS, Université de Toulouse, CNES, CNRS, IRD, UPS, Toulouse, France, [2] LOTUS Laboratory, University of Science and Technology of Hanoi, Vietnam Academy of Science and Technology, Hanoi, Vietnam, [3] Ecological Engineering Laboratory, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [4] Laboratory of Oceanology and Geosciences, CNRS UMR 8187, Université du Littoral Côte d'Opale, University of Lille, IRD, Wimereux, France, [5] Center for Oceanography, Vietnam Administration of Seas and Islands, Hanoi, Vietnam

OPEN ACCESS

Edited by:
Alexander Yankovsky, University of South Carolina, United States

Reviewed by:
Xavier Carton, Université de Bretagne Occidentale, France
Yan Jia, University of Connecticut, United States

***Correspondence:**
Tung Nguyen-Duy
nguyenduytung@hotmail.com

Specialty section:
This article was submitted to Coastal Ocean Processes, a section of the journal Frontiers in Marine Science

Received: *07 September 2021*
Accepted: *13 October 2021*
Published: *12 November 2021*

Citation:
Nguyen-Duy T, Ayoub NK, Marsaleix P, Toublanc F, De Mey-Frémaux P, Piton V, Herrmann M, Duhaut T, Tran MC and Ngo-Duc T (2021) Variability of the Red River Plume in the Gulf of Tonkin as Revealed by Numerical Modeling and Clustering Analysis. Front. Mar. Sci. 8:772139. doi: 10.3389/fmars.2021.772139

We study the daily to interannual variability of the Red River plume in the Gulf of Tonkin from numerical simulations at high resolution over 6 years (2011–2016). Compared with observational data, the model results show good performance. To identify the plume, passive tracers are used in order to (1) help distinguish the freshwater coming from different continental sources, including the Red River branches, and (2) avoid the low salinity effect due to precipitation. We first consider the buoyant plume formed by the Red River waters and three other nearby rivers along the Vietnamese coast. We show that the temporal evolution of the surface coverage of the plume is correlated with the runoff (within a lag), but that the runoff only cannot explain the variability of the river plume; other processes, such as winds and tides, are involved. Using a K-means unsupervised machine learning algorithm, the main patterns of the plume and their evolution in time are analyzed and linked to different environmental conditions. In winter, the plume is narrow and sticks along the coast most of the time due to the downcoast current and northeasterly wind. In early summer, the southwesterly monsoon wind makes the plume flow offshore. The plume reaches its highest coverage in September after the peak of runoff. Vertically, the plume thickness also shows seasonal variations. In winter, the plume is narrow and mixed over the whole water depth, while in summer, the plume can be detached both from the bottom and the coast. The plume can deepen offshore in summer, due to strong wind (in May, June) or specifically to a recurrent eddy occurring near 19°N (in August). This first analysis of the variability of the Red River plume can be used to provide a general picture of the transport of materials from the river to the ocean, for example in case of anthropogenic chemical substances leaked to the river. For this purpose, we provide maps of the receiving basins for the different river systems in the Gulf of Tonkin.

Keywords: Red River, river plume, coastal ocean modeling, K-means, clustering analysis, passive tracers, unsupervised learning

INTRODUCTION

River plume can be defined in a general way as the region of the coastal ocean where its properties and dynamics are affected by the river runoff (Horner-Devine et al., 2015). Though the river runoff is small compared to the whole ocean water volume, it can impact both the physics and biogeochemistry of the coastal ocean depending on the discharge, the properties of the ocean area (bathymetry, bottom roughness) and external forcing (air-sea fluxes, open ocean influence). Furthermore, rivers carry sediments and anthropogenic contaminants from agriculture and industrial activities. Therefore, there is a need to better understand the fate of the river water from the estuaries to the ocean. It is the first step toward the study of the dispersion of the possible contaminations and toward the design of strategies for monitoring and managing the water quality and the health of ecosystems. These are particularly crucial issues in densely populated areas, such as many deltaic regions of Southeast Asia, including the Gulf of Tonkin.

The Gulf of Tonkin (16.9°N – 21.9°N, 105.6°E – 110°E) is a small shelf sea located east of Vietnam and south of China with depth less than 100 m (**Figure 1A**). It is a meso-tidal region, dominated by diurnal constituents (K1, O1) as described for instance by Nguyen et al. (2014) and Piton et al. (2020). The ocean circulation of the Gulf of Tonkin (hereafter GOT) has been examined by several authors (e.g., Ding et al., 2013; Gao et al., 2014; Rogowski et al., 2019; for a recent review see Piton et al., 2021). Using model results and observational data, they all agree that the winter circulation at the gulf scale is cyclonic and driven by wind. However, in summer, the circulation is not so well explained. Wu et al. (2008) showed that the flow through Hainan Strait has an impact on the circulation in summer: if it is inflow, the circulation is mainly cyclonic and vice versa. From high-resolution model pluri-annual simulations, Piton et al. (2021) found a basin-scale anticyclonic circulation in summer. They also suggest that the surface circulation is mainly ageostrophic, as a consequence from the monsoon wind forcing, except along the Vietnamese coast where the southward coastal current has a dominant density-driven component.

One of the expected drivers of the dynamics is the freshwater input from the continent. In the GOT, several rivers feed the gulf. The main one is the Red River (hereafter RR) system. It is formed by 3 tributaries that connect at Son Tay (**Figure 1B**) and then split again into several distributaries. On average, the RR's runoff accounts for more than 60% of the total runoff in GOT (**Figure 1C**).

In spite of its importance, to date, studies focusing on the RR plume are still scarce. Gao et al. (2013) showed that the coastal plume was found near the northern and western coasts of the gulf in winter while spreading eastward and offshore in summer. Rogowski et al. (2019) analyzed the mean seasonal circulation and suggested that the southwesterly wind direction which is prominent during the summer monsoon is the main mechanism preventing downcoast advection of the RR plume. However, that study did not attempt to explain further the temporal variations of the RR plume. In this study, we propose to take the analysis a step further and examine the plume variability in the mid-field and far-field regions (as defined by Horner-Devine et al., 2015) in more detail, using high-resolution simulations and an unsupervised learning method.

The objectives of this study are (1) to propose a method to identify the RR plume in the GOT from the model outputs, (2) to describe its development and characterize its variability at different scales and (3) to attempt to describe the physical processes at work. To do this, we use numerical simulations combined with cluster analysis. In section "Methods, Model, and Data," the model configurations, data sources and the unsupervised learning algorithm used to classify the main pattern of the plume (K-means) are described. The model is then evaluated against several observational data sets in section "General Circulation and Model Assessment." In section "River Plume in GOT: Identification and Variations of Area," several methods to identify the plume in the GOT are compared which allows us to select the most appropriate one given our purposes. On that basis, the river plume variability at different time scales is examined in section "Variability of the RR Plume," illustrating the effect of key physical processes. A discussion on the plume classification and a description of the receiving basins from the different river systems in the GOT are presented in section "Discussion," followed by conclusions in section "Conclusion."

METHODS, MODEL, AND DATA

SYMPHONIE Model and General Configuration

SYMPHONIE is a numerical model that solves the primitive, Boussinesq, hydrostatic equations of the ocean circulation (Marsaleix et al., 2006, 2008) on a curvilinear bipolar (Bentsen et al., 1999). Arakawa C-grid with regular sigma vertical levels. The QUICKEST scheme is used for advection and diffusion of tracers (Neumann et al., 2011), while horizontal advection and diffusion of momentum are respectively computed with a 4th order centered and a bi-harmonic scheme, and vertical advection of momentum by a 2nd order centered scheme (Damien et al., 2017). The k-epsilon turbulence closure scheme is implemented as in Michaud et al. (2012).

Our reference configuration, hereafter called **GOT_REF**, is an update of the configuration of Piton et al. (2021). The model domain covers the GOT area. Due to the variable horizontal grid, the coastal area near the RR mouths has a fine horizontal resolution of 300 m, while near the open boundary, the grid size can increase to 4500 m (**Figure 2A**). As the coastal area adjacent to the RR delta is characterized by a complex topography, with many islands and islets, a considerable effort has been devoted to the construction of the model bathymetry and shorelines and is described in Piton et al. (2020). In particular, the bathymetry is reconstructed from GEBCO 2014 combined with other sources and field surveys. The bottom drag coefficient follows a logarithmic law (Blumberg and Mellor, 1987), depending upon the bottom roughness length which is set to 1 mm. The parametrization for the solar penetration depth (as described for instance by

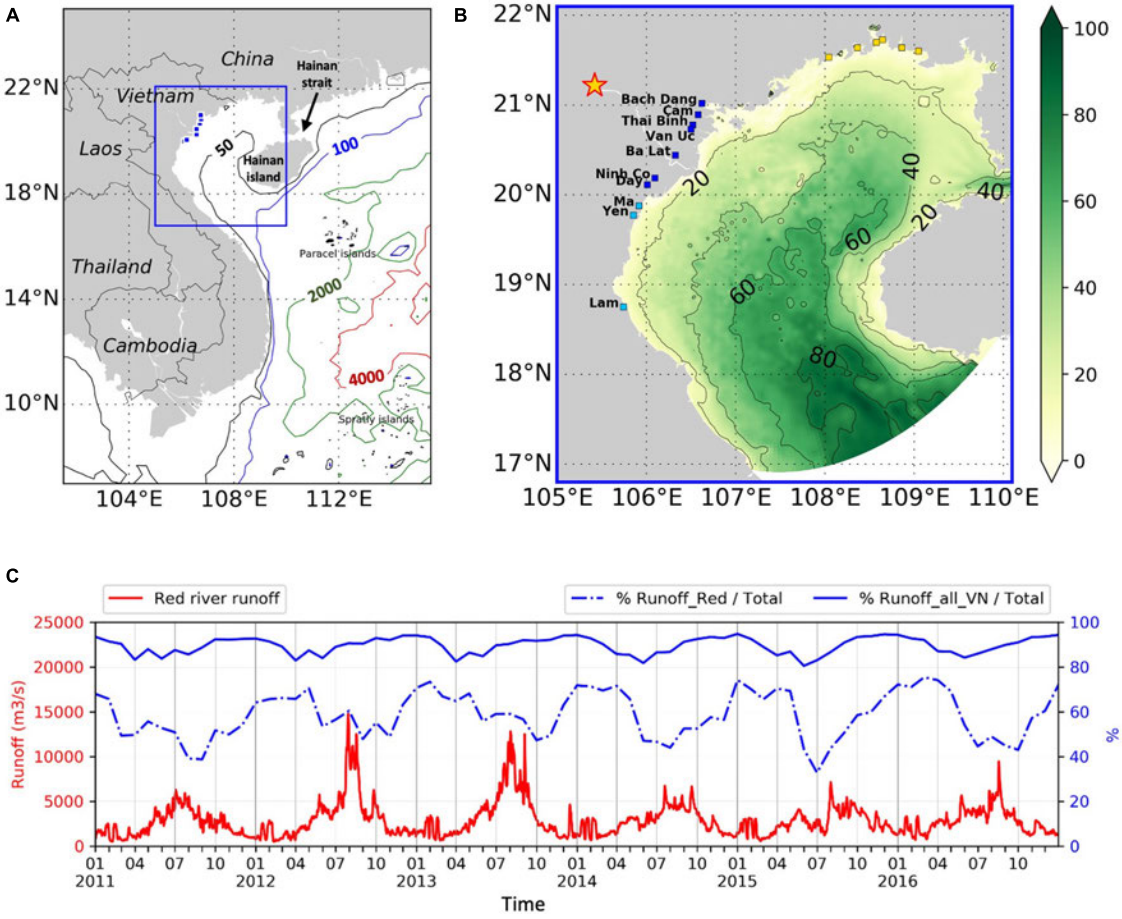

FIGURE 1 | (A) Location of the Gulf of Tonkin. Blue dots show the location of the river mouths of the Red River system. Contour lines indicate the bathymetry of the area. **(B)** Model domain with bathymetry. The star indicates the location of Son Tay hydrological station at the apex of the delta. Blue boxes show the most upstream points where the runoffs flow to the model. Cyan and yellow boxes show the location of other river mouths in the south and north of RR, respectively. **(C)** Daily Red River discharge in the model configuration, from National Hydro-Meteorological Service data (red curve, left y-axis). Blue dashed line: Percentage of Red River runoff compared to total runoff (sum of runoff from Red River, other rivers in Vietnam, rivers in China). Blue line: percentage of all rivers in Vietnam (including Red and other rivers in the GOT) compared to total runoff. Blue lines are referred to the right y-axis.

Maraldi et al., 2013) distinguishes the red and near-infrared radiations which are absorbed in surface layers (e-folding length scale: l = 0.35 m), and shorter wavelengths (mostly visible and ultraviolet) which penetrate deeper (e-folding length scale from less than 4 m along the coastline up to 15 m in the deeper region of the GOT).

Lateral boundary conditions, as described in Toublanc et al. (2018), allow the model to be nested into a larger scale model. At the open boundary, tidal surface elevation and current at K1, O1, P1, Q1, K2, M2, N2, S2, M4 frequencies from the tidal atlas FES2014 (Lyard et al., 2021) are taken into account as in Pairaud et al. (2008). The model is also forced by daily averages of sea surface height (SSH), 3D zonal velocity (u), meridional velocity (v), temperature (T) and salinity (S) fields, from the global analysis (hereafter 'OGCM') produced by Mercator-Océan International and provided by Copernicus Marine Service (CMEMS) at a resolution of 1/12°. CMEMS T, S fields are adjusted to recover consistency with tidal physics before being used at the open-boundary conditions of the model (**APPENDIX A**).

At the surface, boundary conditions are provided by an atmospheric model and fluxes of momentum, heat and freshwater are computed using the bulk formulae of Large and Yeager (2004). Operational ECMWF analyses (with a spatial resolution of 1/8°) are used to provide 3-h wind, precipitation, solar energy, atmospheric temperature, dew-point temperature, surface pressure.

As the present study focuses on the fate of continental water into the coastal ocean, a specific effort has been deployed on the

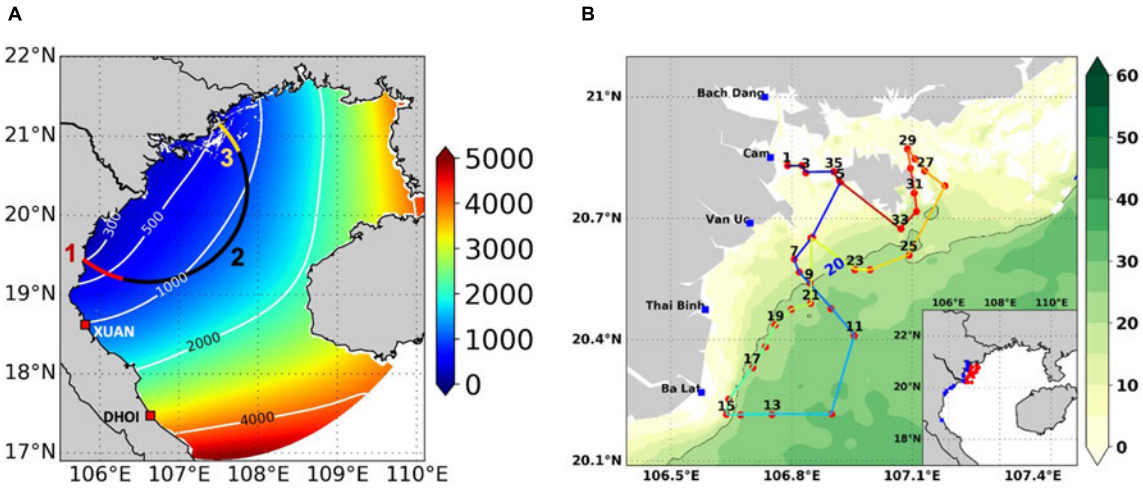

FIGURE 2 | (A) Grid cell size (m). Red squares show the location of HFR antennas. (1, 2, 3): Sections used to calculate the tracer transport. VITEL **(B)** CTD stations location. Red dots indicate the station locations, while blue points indicate the river mouths' locations.

river runoffs implementation; section "River Configurations" is dedicated to its description.

The model is run from 2010 to 2016, starting from the ocean state condition as provided by the global OGCM on 01/01/2010. The time step is set at 2 min. Further analyses are calculated during 2011–2016 (i.e., following a 1-year spinup). The model outputs include daily averaged variables as well as instantaneous fields every 12 h. Unless otherwise specified, both components of the current are detided based on an online harmonic analysis. A summary of the general configuration is available in **Table 1**.

Two other simulations are performed over the same period:

- a twin simulation without river forcing (GOT_NORIV) to assess the impact of the river runoff on the coastal circulation; all other forcings and parameters are the same as in GOT_REF.
- a twin simulation without tides (GOT_NOTIDE) to assess the impact of tides on the main patterns of the river plume variability; all other forcings and parameters are the same as in GOT_REF.

River Configurations

In most modeling studies to date, the RR was set up with only one mouth (or input grid point) using monthly or annual climatological runoff because more realistic runoff at the hydrological station was unavailable (Ding et al., 2013; Gao et al., 2013). In this study and in Piton et al. (2021), the river condition is configured as realistically as possible. Firstly, the delta is represented taking into account seven input grid points representing the mouths of the main RR delta distributaries (Bach Dang, Cam, Van Uc, Thai Binh, Ba Lat, Ninh Co., Day) (**Figure 1B**). Secondly, the daily RR runoff is obtained from the National Hydro-Meteorological Service (NHMS) of Vietnam at Son Tay hydrological station which is located at the apex of the RR delta. The discharge is distributed across the seven distributaries based on the results of Vinh et al. (2014) (Bach Dang (7%), Cam (13.2%), Van Uc (14.5%), Thai Binh (6.4%), Ba Lat (30.3%), Ninh Co (5.6%), Day (23.0%)). Furthermore, each river mouth is connected to a channel in the model to allow the water entering the coastal ocean with realistic salt and temperature properties and realistic stratification. The length of the channels (35–45 km, depending on the channel) is chosen to exceed the saltwater intrusion, which is approximately 30 km from the mouth (Nguyen Thi Hien et al., 2020). The results from GOT_REF confirm that the salty water never reaches the end of the channel, even in the low discharge period. The river runoff is converted into a vertically sheared current at the most upstream point of the channel (**APPENDIX B**); there the salinity of the river flow is 0 and the temperature varies seasonally from 17°C (in February) to 29°C (in August).

TABLE 1 | General characteristics of GOT_REF.

Name	Features
Grid points	429 * 600
Grid size	300 m (near the coast), 4500 m (near the boundary)
No. of vertical layers	20
Boundary conditions	global OGCM
Atmospheric conditions	ECMWF analysis (1/8°, 3 h)
Tidal forcing	FES2014 (K1, O1, P1, Q1, K2, M2, N2, S2, M4)
Light attenuation	Varying by space (~1 m near the coast and 15 m offshore)
Bottom roughness	1 mm
Timestep	2 min
Time	2010 – 2016

Other Vietnamese rivers (Ma, Yen, Lam) at the south of the Red River delta (hereafter referred to as the 'southern rivers') are also taken into account (**Figure 1B**). As daily runoffs were not available to us, we prescribe monthly climatological runoffs from NHMS. At the north of the Red River delta, 6 rivers (hereafter referred to as the 'northern rivers') are accounted for (**Figure 1B**), based on the data given by Gao et al. (2013). In general, the runoff of the Red River system alone accounts for 60% of the total runoff to the gulf, while adding other rivers runoff in Vietnam accounts for around 90% of the total runoff (**Figure 1C**). In detail, the average discharge for the Red River system in low (December, January, February) and high (July, August, September) discharge period equals to 1632 m^3/s and 4959 m^3/s, respectively. For the southern rivers, this value is 365 and 2043 m^3/s. For the northern rivers, it is 164 m^3/s and 1103 m^3/s. It is clearly shown that for the Red River, the ratio between high and low discharge seasons is only 3 times, this ratio is 5.6 times for the southern rivers and 6.7 times for the northern rivers. The lower ratio of the Red River can be due to the presence of several hydrological dams upstream.

In order to simulate the pathways of the river water into the GOT, passive tracers are used. Tracers act as dyes, i.e., they do not affect the dynamics. In total, there are 3 tracers (or three colors) meant to distinguish the inputs from the different river systems. The first tracer is added to the runoff of all the Red River distributaries at constant concentration (100 arbitrary unit/m^3). The second one is added at the other rivers in the south and the third one at the rivers in the north of Red River, with the same concentration (100 arbitrary unit/m^3). Since tracers are injected with the runoff, they are also submitted to the 1 year of spin-up.

Observational Data Sets Used to Evaluate the Simulation

In situ Data

We use temperature (T) and salinity (S) profiles measured at 35 CTD stations (**Figure 2B**) during the VITEL cruise which took place in July 2014 (Ouillon, 2014). We further use CTD measurements of T, S acquired repeatedly at 10 stations along a 25 km cross-shelf section by Vietnamese and US teams from Center for Oceanography (CFO) and Oregon State University (Rogowski et al., 2019). This dataset, hereinafter referred to as CFO data, consists in 20 timeframes between September 2015 and July 2016.

We compare the simulated tidal amplitude and phase with historical tidal measurements from 13 stations (whose location is indicated in **Figure 3**). The dataset stems from the International Hydrographic Organization[1].

High-Frequency Radar Measurements

We use surface velocity data from the high-frequency radar (hereafter HFR) system based on two antennas located at 18.62°N (XUAN site) and 17.47°N (DHOI site) (**Figure 2A**) along the coast and operated by the Center for Oceanography, Vietnam Administration of Sea and Islands (CFO, VASI). The data consists of daily maps of zonal and meridional components of the surface current, over the year 2015, built by Tran et al. (2021). The radial velocity measurements are gap-filled, interpolated onto a 6-km rectangular grid and detided as described by Tran et al. (2021). As explained in Rogowski et al. (2019), the summer coverage is lesser than in winter because of low sea state conditions; therefore uncertainties on the interpolated velocities are larger in summer. Comparisons with *in situ* measurement over a 12-day period indicate a mean bias of 3 cm/s and an RMS difference of 10 cm/s (Tran et al., 2021); we use these values as rough estimates of the data uncertainties. HFR data is representative of currents at 2.4 m below the surface (Tran et al., 2021); as a consequence, for the model assessment, we estimate the model current at 2.4 m before interpolating it over the HFR grid.

Altimetric Data

Tidal constituents computed from satellite altimetric data provide a rich dataset to evaluate the simulated tides offshore. Along-track amplitudes and phases are calculated from the long time series of sea surface height obtained from satellite altimetry (TOPEX/Poseidon, Jason 1–2) by using harmonic analysis. We use the dataset described in Lyard et al. (2021).

K-Means Unsupervised Learning Algorithm for Time Series Pattern Analysis

Many different methods exist to analyze the main characteristics of time series, such as statistical, spectral and classification methods. In terms of classification, there are several methods. For example, Self-Organizing Maps method is used to describe the river plume patterns by Vaz et al. (2018) for the Tagus River and Falcieri et al. (2014) for the Po River. Here, we apply another method, K-means, to analyze not only the plume patterns but the associated forcing conditions.

K-means clustering analysis (KMA) is a popular unsupervised learning method (Hastie et al., 2001) that allows to classify objects (observations, model outputs) into different groups, given a measure of dissimilarity.

The aim of this method is to iteratively identify clusters by their centroid (the means) then minimize the distance between each member of the cluster and the centroid of that cluster. The procedure of this method is made of six steps.

Step 1: Choose the number of clusters to compute.
Step 2: Allocate random numbers as the centroids of the clusters.
Step 3: Calculate the distance from each member to the centroids, using the Euclidian distance (see **APPENDIX C** for more details).
Step 4: compare these distances, then assign the member to the cluster corresponding to the minimum distance.
Step 5: for each cluster, re-calculate the centroid based on the mean of all members which belong to that cluster.
Step 6: repeat step 3 to 5 until the clusters no longer change.

This method has been used in several past studies in coastal oceanography. Solabarrieta et al. (2015) used it

[1]https://www.iho.int/

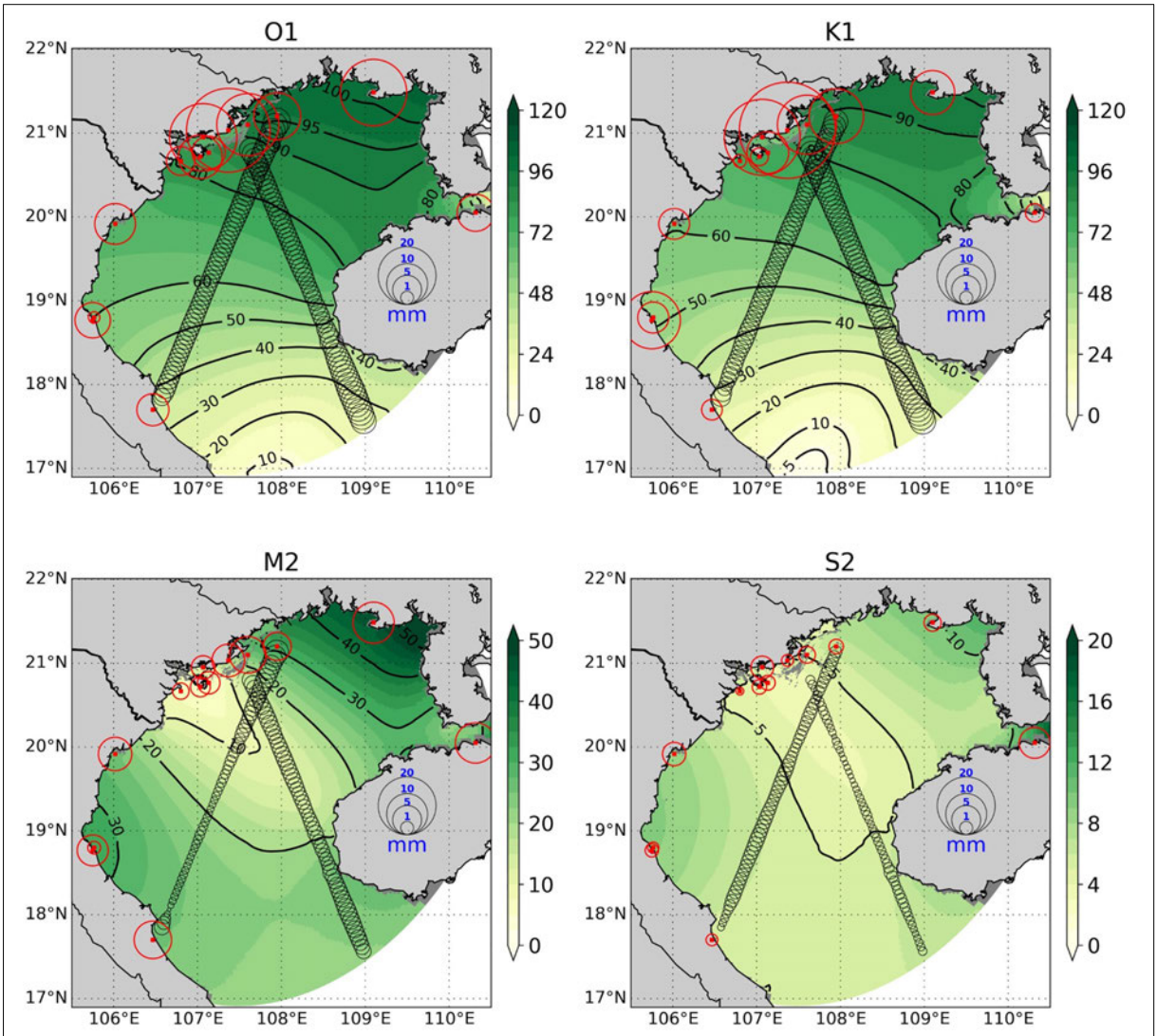

FIGURE 3 | Amplitude of the tidal constituent (in colors, cm) from GOT_REF. Each figure shows the complex error (mm), calculated from the amplitude and phase, between model result and observational data for 4 main tidal constituents (O1, K1, M2, S2). Black circles show errors compared to altimetry data, while red circles indicate errors compared to the tidal gauges data.

to examine the relationship between wind and surface circulation in the Bay of Biscay. They found that most of the current patterns are related to a specific wind pattern in the study area. Chen et al. (2017) identified the area of Pearl river plume by applying the K-means clustering to a summer climatological turbidity image. Sonnewald et al. (2019) used it to classify different regions based on the barotropic vorticity.

In this study, we use KMA to identify the main patterns of the plume and their temporal variations using daily model outputs. To do this, we use the KMA as implemented in the scikit-learn library coded in python (Pedregosa et al., 2011).

In the KMA method, the number of clusters is not fixed automatically, it is chosen depending on the application. The number of clusters should lead to an easy interpretation of the classification, since the objective is to reduce the dimension of the ensemble of scenes to analyze, without eliminating too much variability. In this study, we choose 4 clusters. The method to identify the number of clusters is described in **APPENDIX C**.

We applied KMA to the "masked river plume." Firstly, plumes are identified using tracers (as described in section "River Plume in GOT: Identification and Variations of Area"). Then, for a given day (a scene), each grid point where the plume is considered to be present is masked with 1, while other points are masked with 0.

The results of the classification are shown and analyzed in section "Variability of the RR Plume."

GENERAL CIRCULATION AND MODEL ASSESSMENT

The main objectives of this section are to provide a general description of the patterns of the seasonal circulation at the scale of the GOT from GOT_REF, to assess the consistency of model results with previous studies (or evidence their specificities) and to provide a qualitative assessment by comparing to observations. As a description of the general circulation and model verification (in particular using satellite data) have been provided recently by Piton et al. (2021) from a very similar configuration, we do not provide a detailed analysis. We introduce the section by a discussion of the first Rossby radius, continue with the basin-scale circulation, evaluation of tidal elevation and with the comparison to *in situ* temperature and salinity data and surface current velocity from high-frequency radars in the area of the RR plume.

First Baroclinic Rossby Radius of Deformation

The first baroclinic Rossby radius of deformation (Rd) sets the scale of mesoscale baroclinic instabilities and, as such, should be resolved by the numerical circulation models (see for instance Greenberg et al., 2007). In river plume dynamics, Rd in the near-field or far-field (coastal current) is used to estimate the relative impact of Coriolis force on the plume dynamics with respect to other forcing (runoff, ambient current). For instance, Yankovsky and Chapman (1997) showed that for surface-advected plumes, the offshore extension is more than four Rossby radii.

In this study, Rd is computed from the mean seasonal field of temperature and salinity from our simulation over 2011–2016. The calculation is made by solving a Sturm-Liouville problem as described for instance in Chelton et al. (1998). We use the method developed by F. Lyard at LEGOS for free surface vertical modes of the internal pressure and described for instance in Nugroho (2017). At depths larger than 60 m, the Rossby radius decreases from 10 to 13 km in April–September (**Figure 4A**) to 4–7 km in October–March (**Figure 4B**). In shallower areas, it decreases from 3–8 to 1–5 km. Such a seasonal variation is expected as the stratification is stronger in April–September with higher air temperature. A large spatial variability is also observed; by construction it is correlated to the bathymetric variations. In particular, a striking feature is the eastward extension of the shelf, between the Bach Long Vi island and the coast; the Rd does not exceed 4 km there. Close to the coast in both summer and winter monsoon periods, it can drop to less than 1 km. Moreover, the Rd spatial variability also reflects some patterns of the circulation. For instance, a strong mixing occurs west of Hainan, mostly due to tides, as suggested in the literature (e.g., Nguyen et al., 2014; Piton et al., 2020) and in agreement with the local maximum of the K1 and O1 current in our simulation (**Supplementary Figure 1**). As a consequence, the Rd is only a few hundred meters west of Hainan Island. On the other hand, the RR runoff leads to an increase of stratification along the Vietnamese coast south of 21°N which has a clear signature in a locally larger Rd (∼4–5 km) in the October–March period.

Assuming that the effective resolution is roughly between six and ten times its grid resolution, we can evaluate the model's ability to resolve Rd. Rd is well resolved in the northwest quarter of our domain where the mesh is well refined in summer, and along the Vietnamese coast between 19 and 21°N in winter. However, if we consider that the scales of interest are about 4 times the Rossby radius, then the model resolution is sufficient between 18 and 21°N.

General Circulation

Mean Surface Circulation

The current in the GOT has been studied with mainly a focus on the northern region, due to the lack of observational data in the south (Wu et al., 2008; Ding et al., 2013; Gao et al., 2014). Recently, thanks to the deployment of a high-frequency radar system, the surface circulation in the southwestern area has been documented (Rogowski et al., 2019; Tran et al., 2021). However, the variability of the general surface and subsurface circulation remains relatively undocumented over the whole gulf and more specifically along the Vietnamese coasts.

Figure 5A shows the mean surface current from November to March from GOT_REF for the 2011–2016 period. The circulation in this winter monsoon period is dominated by the gulf-scale cyclonic gyre, as reported by several authors (Ding et al., 2013; Gao et al., 2014; Rogowski et al., 2019; Piton et al., 2021). The coastal current along Vietnam originates from the merging at 20°N of the westward flow from Hainan Strait and a coastal current from the northernmost shelf. It flows downcoast down to the southern boundary of the domain. It is the dominant feature of the surface circulation with a mean amplitude of about 30 cm/s. South of Hainan Island, an inflow from the South China Sea is simulated in the deepest part of the area with a decreasing intensity from November to March. This inflow is deflected westward and joins the coastal current south of 19°N, creating a small-scale cyclonic gyre as in Ding et al. (2013).

In April, the summer monsoon sets in and lasts until August. The southward coastal current weakens and becomes intermittent (**Figure 5B**). When present, it is deflected eastward at ∼18°N; it then splits into 2 branches. The first one forms a cyclonic circulation in the central basin. The second branch feeds an intense current south of Hainan that flows out from the gulf. This corresponds to the circulation scheme described by Gao et al. (2014) and Rogowski et al. (2019), but, in our simulation, the current variability is large at daily and interannual time scales.

In September and October (**Figure 5C**), the coastal current is southward again from 21°N to the southern boundary where it exits the gulf. It reaches its maximum amplitude (∼35 cm/s) and width. Besides, the circulation is characterized by two inflows with large velocity (∼30 cm/s): one from Hainan Strait and the other from the southern boundary in the deep region.

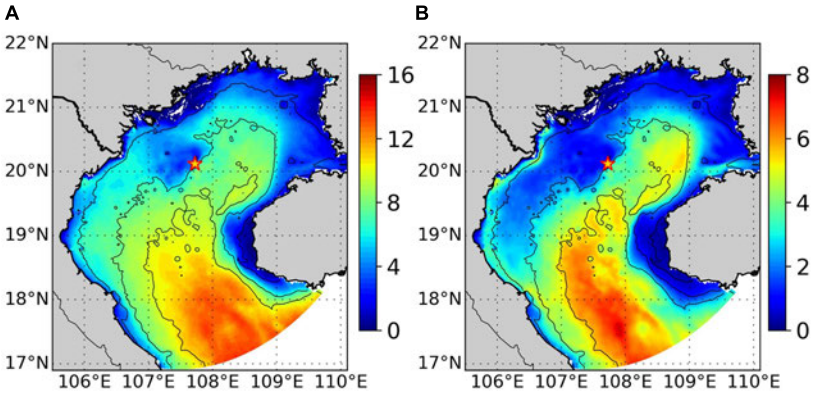

FIGURE 4 | (A,B) The first baroclinic Rossby radius of deformation (km), over the period 2011 – 2016, computed from the temperature and salinity profiles averaged from April to September **(A)** and October to March **(B)**. Contour lines indicate the bathymetry. Star shows the location of Bach Long Vi island. *Note: Different color bar scales.*

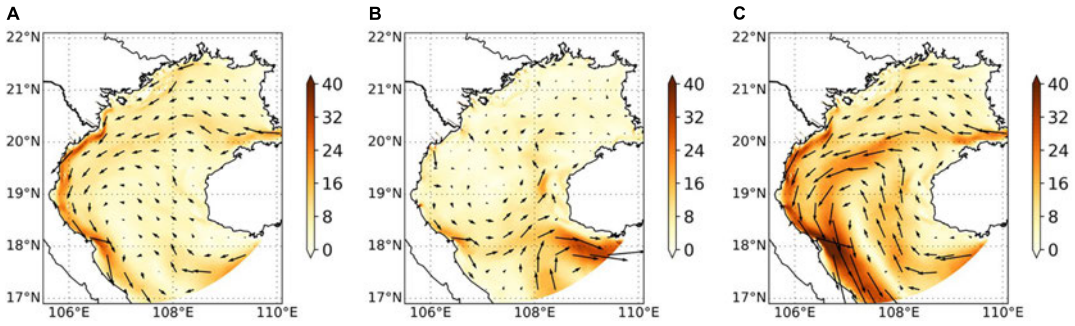

FIGURE 5 | Means of the detided daily surface current (cm/s) simulated by GOT_REF. **(A)** November – March average. **(B)** April – August average. **(C)** September – October average.

Horizontal Transport

The horizontal transport displays a similar seasonal cycle as the surface current. From September to March, the general circulation is cyclonic, with a large downcoast transport over the shelf in the west. From April to August, the transport inside the gulf weakens significantly. The main specific patterns are a cyclonic gyre in the central area.

The flow direction during the summer monsoon through Hainan Strait is discussed in several papers (Shi et al., 2002; Gao et al., 2013; Chen et al., 2017). In our simulation, the flux is inflow from September to April and an outflow during the summer monsoon. It varies between −0.18 Sv and +0.15 Sv at 110 E, with significant daily and interannual variability (not shown). Our results however are consistent with the situation described by Wu et al. (2008); in case of inflow the circulation is mainly cyclonic, while in case of outflow the overall circulation is anticyclonic.

Besides, the gulf is fed all year long by an inflow from the southern boundary, which is maximum in summer monsoon period. From September to December, it enters through the east and either penetrates northward in the gulf or is deflected westward and joins the coastal current. The rest of the year it enters through the center, is deflected eastward and generates a large outflow south of Hainan Island.

Model Assessment

Evaluation of Tides

The simulated amplitudes in terms of sea surface elevation for the four main tidal constituents (O1, K1, M2, S2) are shown in **Figure 3**. O1 and K1 have the same amplitude distribution. The amplitude is largest in the north (100 cm and 95 cm for O1 and K1, respectively) and decreases to its lowest value at the south of the gulf, near the boundary. M2 has two peaks of amplitude. One peak is located in the north (50 cm) and another peak is located at 19°N (30 cm) along the Vietnamese coast. The lowest amplitude is located near RR mouths. These patterns are similar to the ones found by Nguyen et al. (2014) and Piton et al. (2020). We also examined the tidal currents (see **Supplementary Figure 1**): the O1 and K1 tidal currents are the strongest west and south of Hainan Island and in the Hainan Strait (more than 50 cm/s). Near the Vietnamese coast, they are weaker and range from 10 to 20 cm/s. They reach a local maximum (∼ 30 cm/s) along the coast at ∼18°N, 106.5°E. M2 tidal currents vary between 10 and

20 cm/s, with a local maximum close to the Ba Lat mouth. S2 currents have a similar spatial distribution but smaller amplitude (<4 cm/s close to the RR mouths).

The simulated tides show remarkable comparisons to along-track altimetric data (**Figure 3**): the root mean squares of the complex errors (i.e., the model-data misfits) over the domain are as low as 2.6 cm, 2.8 cm, 1.3 cm, 0.7 cm for O1, K1, M2 and S2, respectively. These values correspond to 5.0%, 6.0%, 7.4%, and 12.9% of the signal. The model-data misfits are homogeneous in the center of the basin. They are the largest near the coast as expected due to the larger uncertainty of both the altimetry data near the coast and due to the uncertainty of model bathymetry in the very shallow area. These values are comparable to those obtained by Piton et al. (2020) with the T-UGOm tidal model over the same domain (2.6 cm, 3.5 cm, 3.0 cm, and 1.2 cm for O1, K1, M2 and S2, respectively) and to those of Nguyen et al. (2014).

Compared to historical tidal gauges data, the complex error is larger: 18.3, 20.9, 6.2, and 2.3 cm for O1, K1, M2, S2, respectively (**Figure 3**). There are several reasons for these higher errors. Firstly, the observation period for the tidal gauges data that are available is relatively short (15 days to 1 year) and may not be long enough to be accurate and representative of 'mean' tides. Secondly, the largest errors are observed in the area of the RR delta where the extremely complicated coastline with the presence of thousands of small islands makes the accurate representation of tides a challenge for such a configuration. Compared to Piton et al. (2020), GOT_REF performs better for M2 (6.8 cm) and S2 (3.5 cm), but gives poorer results for K1 (10.2 cm) and O1 (10.2 cm). Future work will be dedicated to tuning the properties of the channels (see section "River Configurations") to better simulate the propagation of the tidal wave in the delta.

Comparisons With High-Frequency Radar Currents

Figure 6 shows the data from the HFRs and from GOT_REF in different seasons in 2015, as well as the maps of correlation between simulated and observed u and v over the whole time series (1 year).

In January, the main observed coastal current is southward, with an amplitude of ∼ 30 cm/s at the coast and decreasing offshore. The model shows the same current direction, but its amplitude is underestimated by 20 cm/s (**Figures 6A,B**). In April, the coastal current is less intense, but the northward current at 107–107.5°E is stronger. In GOT_REF, both the southward and northward circulations are observed, with a persistent underestimation of the coastal current regarding the HFR data (**Figures 6C,D**). In May, the coastal current reverses north of 19°N in both the observations and simulation. At around 18°N, the simulated downcoast current is deflected eastward consistently with the HFR observations. However, both at 19°N and 18°N the simulated current is stronger than in the HFR data (**Figures 6E,F**). In July, HFR data show a nearly closed circulation. The coastal current flows southward at 19°N, then rotates at 18°N and flows northward and finally joins the coastal current at 19°N again. In GOT_REF, the circulation shows a more complex pattern, with a northward current (of ∼20 cm/s) at 19°N and a large temporal variability (**Figures 6G,H**). In October, HFR data show a similar current pattern as in January, with a strong southward coastal current (∼40 cm/s). The model shows very good consistency, with the strongest velocity near 18°N (**Figures 6I,J**).

The temporal correlation between simulated and observed daily fields of u and v is higher for u offshore (∼0.5) (**Figure 6K**), while the correlation for v is better near the coast (>0.6) (**Figure 6L**). The larger misfits between the model and data at ∼19°N seem to come from a larger spatial variability of the simulated u component (not shown) than the observed one. In particular, as discussed in section "Variability of the RR Plume," this area is characterized by some eddy activity in summer which may not be resolved by the HFR observations or may be out of phase with the observations.

Temperature and Salinity Profiles From *in situ* Measurements

Figure 7 shows the salinity and temperature profiles from the VITEL campaign and the GOT_REF simulation. Although it is a very shallow area, the bathymetry of the model is quite accurate compared to the depth of the CTD data. The largest bathymetry misfit is reached at stations 26 to 30 where the model is too shallow by ∼10 m. Since the campaign took place in summer (July), there is a strong stratification in both salinity and temperature profiles. At the surface, the observed salinity can be as low as 15, while at the bottom it is around 34. The mean value at all points is 28.7. The model performance is good, with a correlation R = 0.88 computed from all profiles, but the mean value is overestimated by 0.7. The model overestimates the surface salinity at stations located in the shallowest area. This suggests that the river runoff is underestimated or that the simulated mixing in the estuaries is too strong, possibly due to local errors on the tidal representation.

The observed temperature goes up to 33°C at the surface and 27°C at the bottom, with a mean value of 30.3°C. The mean bias of the model is low (0.2°C), although the temperature is overestimated at both the surface and bottom at some stations. Overall, R is 0.72.

The comparison with the T, S profiles from the CFO campaigns is not described in detail as we get similar results to those described by Piton et al. (2021) in a similar configuration of the model. For all the 20 sections available, the mean biases (absolute value) between data and model are 0.61 and 0.71°C for S and T respectively, and the mean correlations 0.81 and 0.72. The comparisons suggest that the model runoff is not high enough (as already suggested by the comparison with the VITEL data), or the background salinity provided by the boundary condition is too high. The vertical mixing seems underestimated also. Overall, despite small biases, the model reproduces accurately the seasonal conditions even in the very shallow area.

RIVER PLUME IN GOT: IDENTIFICATION AND VARIATIONS OF AREA

This section will first review some past studies on plume identification then explains our method to identify the plume

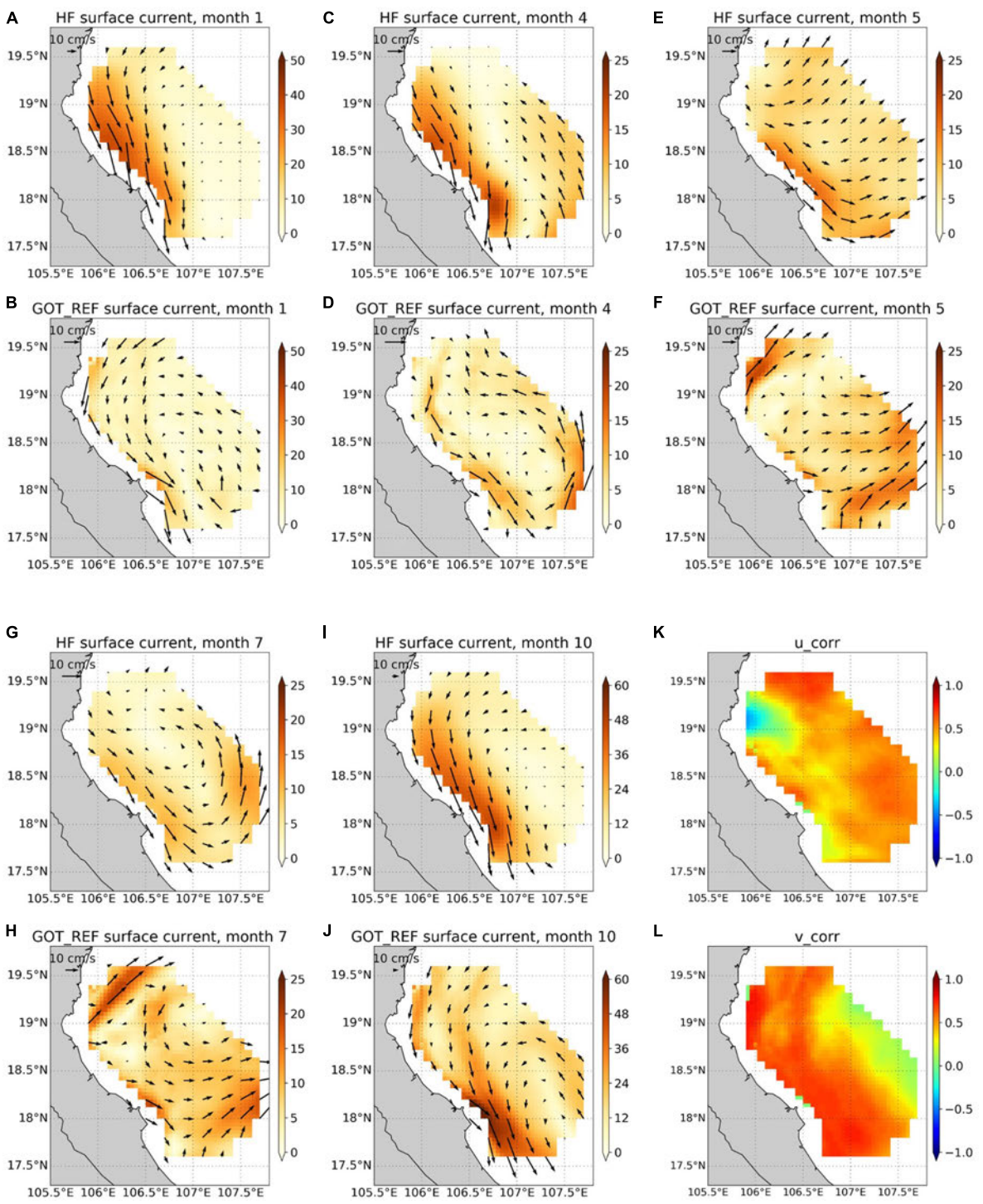

FIGURE 6 | Surface current from HFR data **(A,C,E,G,I)** and model simulations **(B,D,F,H,J)** for different months of 2015 (Jan, Apr, May, Jul, Oct) (cm/s). Temporal correlation of u and v **(K,L)** computed from daily currents. *Note: Different color bar scales.*

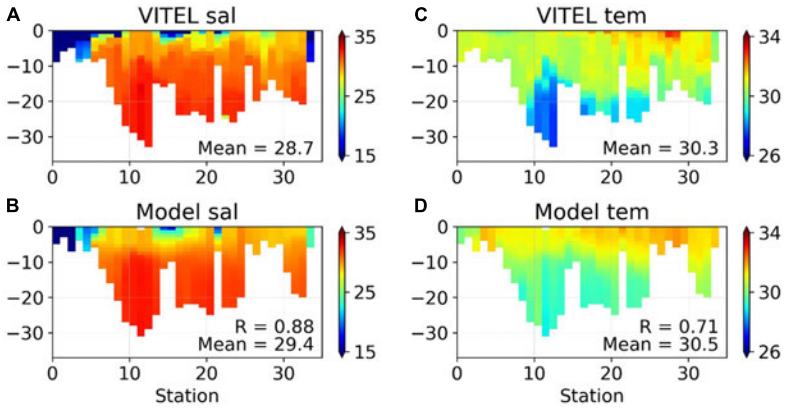

FIGURE 7 | Profiles of salinity from VITEL data **(A)** from GOT_REF **(B)**. **(C,D)** Same but for temperature.

locally in the GOT. Then, the resulting variations of the plume area in the period 2011–2016 are analyzed.

A Brief and Non-exhaustive Review of Past Studies

There are no clear criteria to identify a river plume in general. Usually, such criteria involve sea surface salinity (hereafter SSS) because the river plume has a lower surface salinity than the surrounding ambient waters. However, due to different environmental conditions, the reference salinity is difficult to set in a general enough manner. In some studies, the authors may choose a suitable value based on their knowledge and experience; however, the choice is seldom justified in a rigorous way. This may lead to a situation where different authors choose different values for the same plume, depending on the observational or modeling approach and the objectives. For example, for the Columbia river plume, Liu et al. (2009) use SSS = 29 to detect the plume, while Burla et al. (2010) use the value of SSS = 28 and in MacCready et al. (2009) the plume is identified using SSS < 31.

Other authors defined other criteria to identify the plume. Otero et al. (2008) use the highest horizontal gradient of mixed layer depth to classify the river plume area in the Northwest Iberia and then use the SSS value of 35.6 that coincides with this maximum gradient. Another approach to identify river plumes is based on ocean color satellite images that reveal turbid water masses rich in Chlorophyll, suspended matter, or colored dissolved organic matter (e.g., Chen et al. (2017) for the Pearl River).

To our knowledge, there have been very few studies dedicated to the RR plume, so there is no published reference ambient salinity value. Rogowski et al. (2019) locate the RR plume using monthly means of ocean color maps from MODIS. In tropical areas, the presence of clouds is however a limitation to the use of optical products to investigate the sea surface variability at time scales shorter than a month.

Stratification Index and SSS Threshold

To identify the RR plume, different methods are in consideration in this paper. Firstly, we apply a method based on the Stratification Index (hereafter SI), assuming that the river water forms a buoyant layer over the ocean water. This index was first proposed by Hansen and Rattray (1966) for estuary classification. It is defined as the relative difference between the surface and bottom salinity (eq. 1). In this paper, we tested this method with a threshold of 0.1. If SI equals or exceeds 0.1, the water column is defined as belonging to the river plume. The time series of the daily plume area for all the rivers in GOT based on SI is shown in **Figure 8A**. As expected, the river plume area varies with the river runoff. In July–August (high runoff period), the plume area reaches its highest value, while it is lowest in December–January (low runoff period).

$$SI = \frac{salinity_{bottom} - salinity_{surface}}{vertical\ mean\ salinity} \quad (1)$$

However, there is one major drawback to the SI method. Vertical salinity sections indicate that the plume in the GOT consists of 2 zones. One is the very shallow coastal area where the plume is mixed over the entire water column; in the second one, farther offshore, the plume is detached from the bottom and stays on top of the high salinity water. SI being based on the difference between the salinity at the top and the bottom of the water column, it cannot detect the plume zone in the shallow area where it is totally mixed. In the low runoff period, as we will show in section "Variability of the RR Plume," the downcoast current and the northeasterly wind make the plume stick to the coast; in such a case, the whole plume is well-mixed and SI is not helpful since the calculated area is close to zero (**Figure 8A**).

To deal with that problem, in a second step, we considered using SI as a proxy to find the suitable SSS that can be used as the plume indicator. After several experiments, we selected the value of SSS = 30. In high runoff period, this criterion defines a plume with a similar pattern offshore to the one deduced from the SI

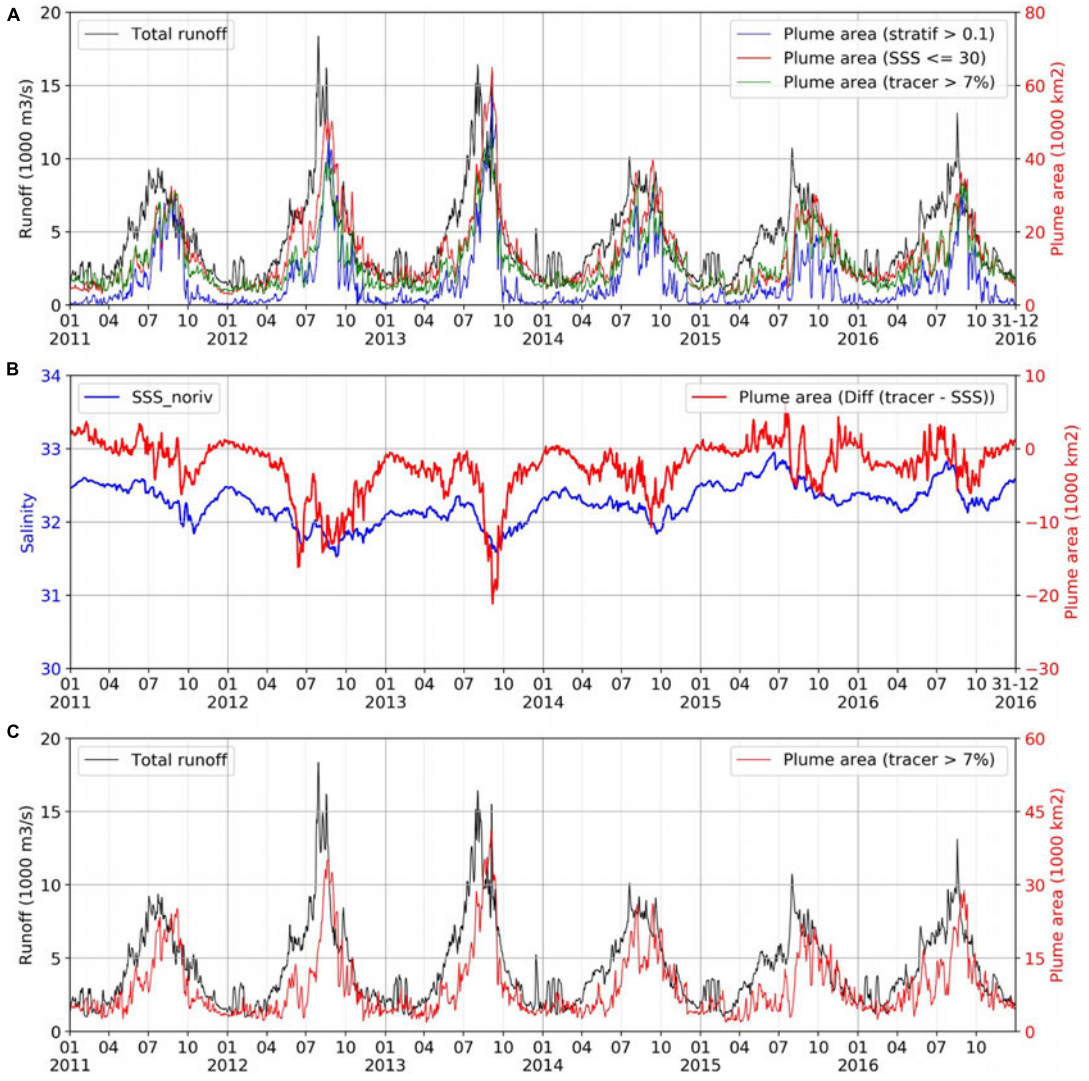

FIGURE 8 | (A) Total runoff (in m^3/s) and plume area (in km^2) for all the rivers in the Gulf of Tonkin. The plume area is calculated using the Stratification Index (blue), SSS (red) and tracer concentration (green). **(B)** Spatially averaged SSS over the whole domain (in blue) from the simulation without river (GOT_NORIV), and difference of river plume area between SSS and tracer methods (in red). **(C)** Daily runoff and plume area for the Red River and Southern rivers.

criterion. In the low runoff period, this discriminant value also captures the narrow plume area that runs along the coast (**Figure 8A**).

Using SSS, the envelope of river plumes created by all the rivers in the GOT was generally identified. However, this "stratification-based SSS" method has in turn two drawbacks. Firstly, it cannot distinguish the plumes created by the different rivers that feed the GOT. Secondly, the SSS criterion may be slightly biased at seasonal time scales by the effect of precipitation. Indeed, the precipitation has seasonal variations due to the monsoon system. The average precipitation from January to March is 1.8 mm/day, while it is 5 times higher (9.4 mm/day) from August to October (calculated from ECMWF daily precipitation over 2011–2016).

Passive Tracers

As an alternative method, we use passive tracers that behave in the model as any buoyancy-free particle or passive chemical from the rivers (as explained in section "River Configurations"). This is particularly relevant to identify the coastal area influenced by the river input of contaminants and to distinguish the freshwater input from other sources than the river such as advection by coastal current and other coastal runoffs [see for instance

Wang et al. (2014) for the Copper River in the Gulf of Alaska] or to study mixing between the plume and ambient waters [e.g., Vlasenko et al. (2013) for the Columbia River]. As for the SSS criteria, the choice of a tracer concentration threshold to define the plume is author-dependent. In Vlasenko et al. (2013), the river plume is identified as the area where the concentration exceeds 10 units/m^3 while in Wang et al. (2014), the threshold is 5 units/m^3. In our study, several experiments led to identify the river plume as the area where the concentration exceeds 7 unit/m^3, which best fits with the area identified by the criterion on SSS in the low runoff period (dry season).

As indicated in section "Observational Data Sets Used to Evaluate the Simulation," 3 tracers were added, respectively to the RR distributaries, the southern rivers and the northern rivers. **Figure 8A** shows the plume area for all the rivers in the GOT calculated by summing the three tracers' concentrations. The plume areas computed from the SSS and tracer criteria have a similar trend in both the low and high runoff periods. However, in the high runoff period, especially during the peaks in 2012 and 2013, the plume area calculated from the tracer is much lower than from the SSS. This difference can go up to over 20,000 km^2 (**Figure 8B**). **Figure 8B** also shows the SSS of the twin simulation without any river runoff (GOT_NORIV). Even with no river, the SSS shows significant variations between 31 and 33. The low sea surface salinity often happens in summer, which is in the rainy season. It is minimum in 2012 and 2013, which are also the years when the runoff reaches the highest peaks (**Figure 8A**), hinting at a correlation between local precipitation and runoff in both years. The SSS drop in the GOT_NORIV run has a similar trend than the difference of plume area calculated by the two methods, which suggests that this difference could be indeed due to precipitation.

In conclusion, for our objectives, the method based on the passive tracers seems the most appropriate to help identify the RR plume among the three methods investigated in this study. It can capture the plume in both the highly stratified zone and the shallow mixed zone. Besides, it can distinguish the plumes created by different rivers (as discussed in section "Discussion").

Temporal Variability of the Plume Area

South of 20°N, the RR plume is quickly joined by the plumes of the southern rivers (Ma, Yen and Lam, **Figure 1**), creating a unique buoyant plume which extends southward along the coast most of the year. In this section and in the following we therefore analyze the variability of the resulting plume from the RR and the three southern rivers together. Section "Discussion" will discuss the receiving basins by distinguishing the RR from the other rivers, thanks to the multi-tracer approach.

Figure 8C shows the evolution of the cumulated runoff from the Red, Lam, Yen and Ma rivers and the corresponding plume area from 2011 to 2016. As expected, the plume area follows the same variations as the river runoff, albeit with a time lag. In summer, when the runoff is high, the peak of the plume area occasionally reaches more than 40,000 km^2 (about 27% of the GOT area), while in winter, it sometimes falls below 2,000 km^2. Due to the interannual variability of the runoff, the plume area varies significantly between different years. The river plume area peak in 2013 (41,000 km^2) is nearly twice as large as the peak in 2015 (22,000 km^2). Both the total discharge and the plume area undergo a strong variability at shorter time scales of a few days as well.

Figure 8C also shows two more characteristics of the plume variability. Firstly, a higher runoff does not ensure a larger river plume area. In 2012, the runoff peak is higher than in 2013, while the plume extension in 2013 is larger. Similarly, although the rainy seasons in 2014 and 2015 are characterized by similar runoffs, the plume area in 2015 is much lower than in 2014. Secondly, there is a time lag between the runoff peak and the peak of the plume area. This lag is also described in Rogowski et al. (2019). Depending on the year, it can be up to 1 month (in 2013). In 2015 and 2016, the plume area is relatively small until July and reaches its peak value rapidly within a few days. In short, both the time lag and plume area do not appear directly correlated to the runoff intensity (i.e., a higher runoff will not necessarily create a larger plume area about 1 month later). Such a variability evidences the fact that river runoff intensity is not the only factor driving the plume size and variability. Wind variability and its impact on the surface circulation and mixing is likely to be another driver, as we show in the next section.

Figure 8C displays as well some interannual variability in both the runoff and the plume area variations. The years 2015–2016 were marked by an intense El Nino event. Generally, El Nino events lead to decreased rainfall, drought and saltwater intrusion in deltas in Vietnam (Sutton et al., 2019). **Figure 8C** does not evidence a drastic drop in the 2015–2016 runoff with respect to 2014 and 2011. However, the SSS signal in the GOT_NORIV run is significantly different in 2015–2016 with respect to the previous years (**Figure 8B**): the mean SSS stays above 32 and the seasonal signal is inverted with a maximum SSS in summer. El Nino is likely to influence the coastal circulation, through alteration of wind and air-sea fluxes of freshwater and heat. The specific influence of ENSO on the coastal plume will be investigated in future studies as we focus in this paper on seasonal and shorter time scales.

VARIABILITY OF THE RR PLUME

Spatial Patterns of the Plume

In the previous section, the river plume in GOT has been identified using passive tracers. We have documented the time variations of the total surface area and compared them to the runoff variations. Now, the seasonal variability of the plume spatial patterns is examined using KMA.

Figure 9 shows the spatial distribution of frequency of occurrence of the plume within each of the 4 clusters calculated from the plume area from 2011 to 2016, and the temporal evolution of the clusters, revealing both seasonal and interannual variations. The main variables of interest (SSS, wind, surface current, SSH) are averaged over the period corresponding to each cluster and are shown in **Figure 10**.

The first cluster usually occurs from October to March and appears on 992 days (45.3%) in 6 years. In this cluster, the plume is very narrow and is mostly confined to the shallow area (depth $<$ 20 m). 75% of the time that the plume occurs,

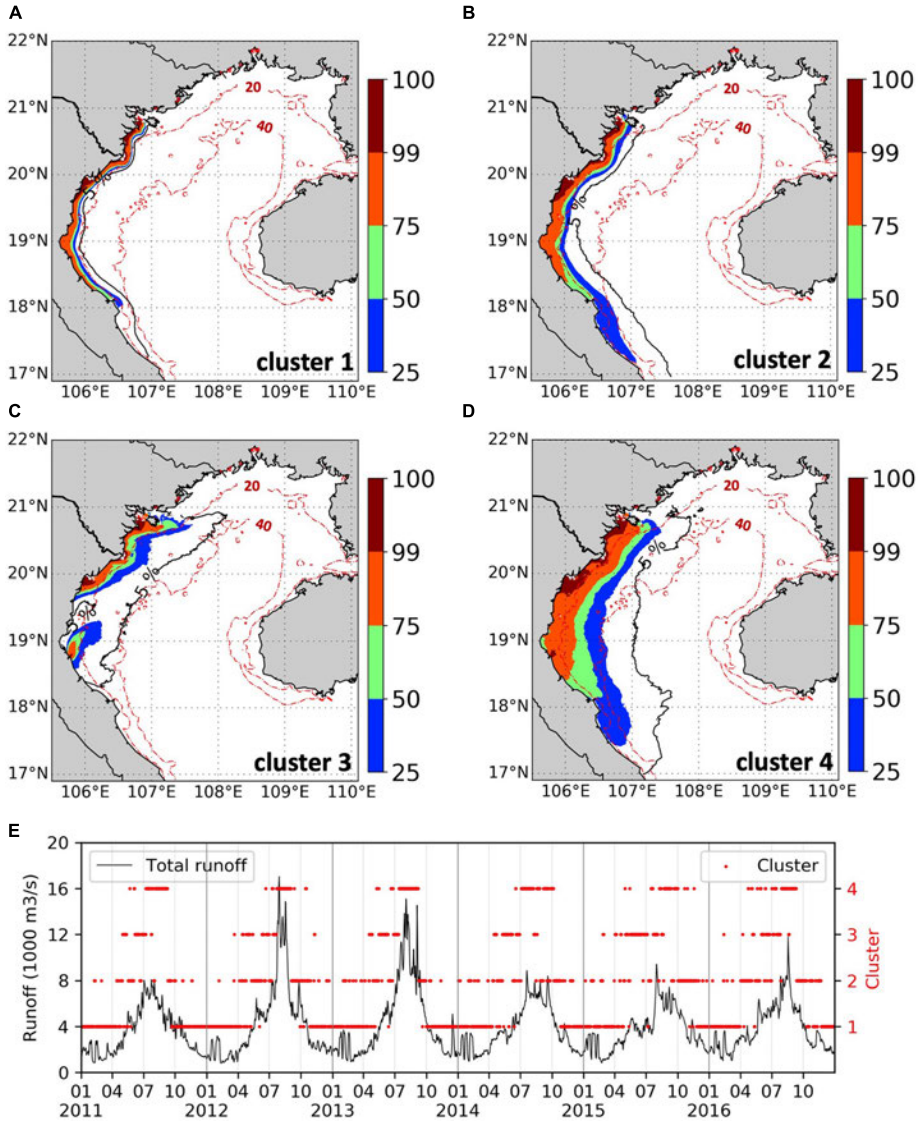

FIGURE 9 | Frequency of occurrence of the plume within cluster 1 **(A)**, cluster 2 **(B)**, cluster 3 **(C)**, cluster 4 **(D)**. For instance, for cluster 1: at a given point, a frequency of 25 means that over the whole period when cluster 1 is present, the plume is present at this point 25% of the time. **(E)** Temporal distribution of each cluster (in red) and the runoff from Red and Southern rivers (in black).

at 19°N, it extends to 105.8E (∼20 km from the coast) while 25% of the plume extends to 105.85E (∼25 km from the coast) (**Figure 9A**). If the runoff (or other forcing) undergoes strong fluctuations, the plume area should also vary. In this cluster, the difference between the spatial occupancy of 25% and 75% is small: it means that the forcing condition is relatively stable. The simultaneous occurrence of several conditions can explain the relatively small extent of the plume in this cluster. Firstly, it happens in the lowest runoff period (average discharge over the cluster equal to 2197 m^3/s). Secondly, the wind is strong and from the north-east (winter monsoon, **Figure 10A**), which is downwelling-favorable. At the GOT scale, the mean wind velocity is 6.4 m/s; it is stronger than in any other cluster, even near the coast (it reaches ∼ 5.7 m/s on average inside the red box shown on **Figure 10A**). The coastal surface current is southward, with a mean speed larger than 20 cm/s between 20.5 and 17.5°N (**Figure 10E**); consistently with the strong monsoon wind, the current also reaches its largest intensity in cluster 1. All these conditions favor the low salinity (SSS < 30) water to be confined to the coast and to extend southward

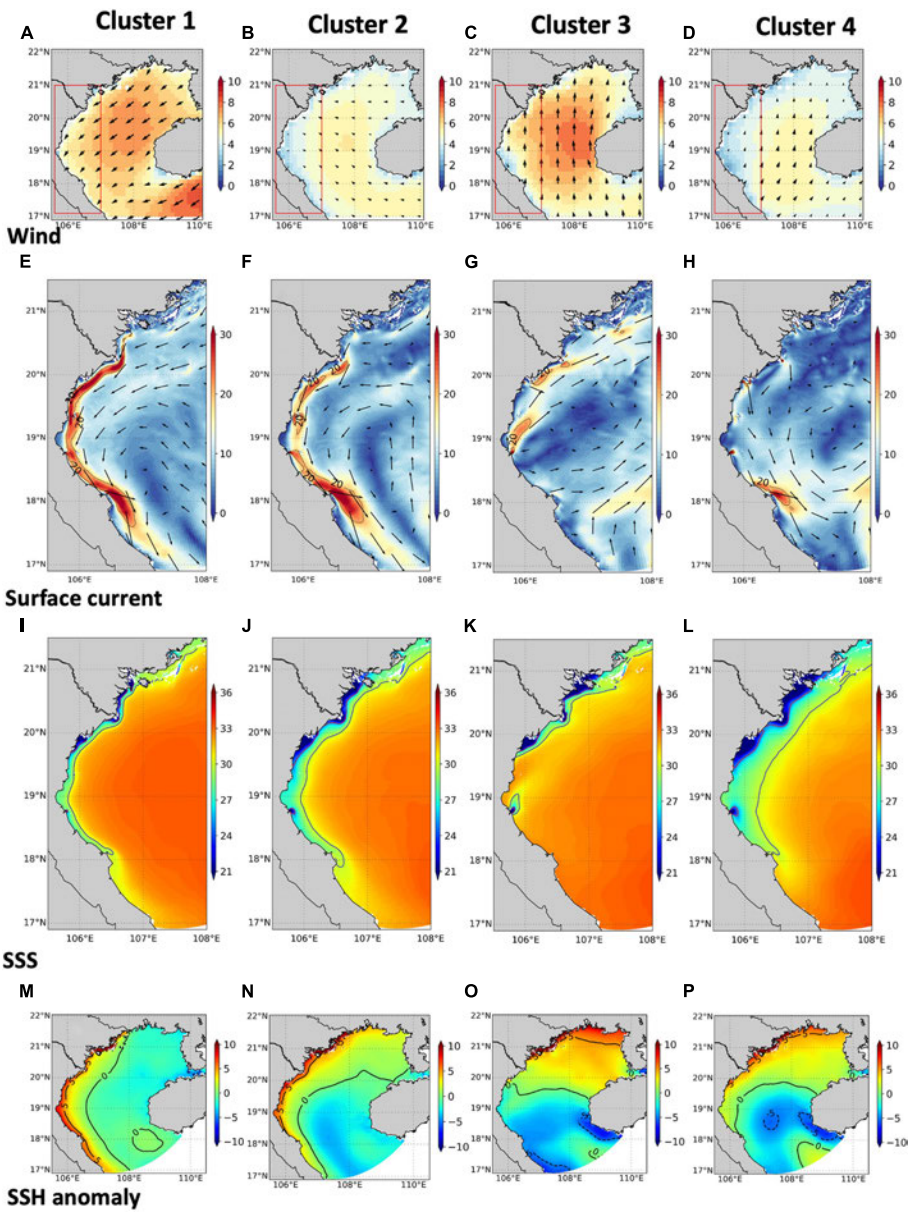

FIGURE 10 | (A–D) Wind conditions (m/s) corresponding to the 4 clusters. **(E–H)** and **(I–L)** Same as **(A–D)** but for surface current (cm/s) and SSS. **(M–P)** SSH anomaly with respect to the mean over the basin for each cluster (cm).

all along the coast (**Figure 10I**). We also observe a cross-shore gradient of SSH with an elevation larger than 5 cm with respect to the basin-averaged SSH (**Figure 10M**). Here, the coastal current is fed by buoyant river waters which contribute to the geostrophic component of the current. Indeed, in the GOT_NORIV simulation, the surface current in cluster 1 is 10 to 15 cm/s weaker; the SSH is 1 to 4 cm lower (not shown). Similarly, Piton et al. (2021) estimate that the geostrophic contribution to the downcoast surface current reaches up to 60% in December-February.

The second cluster appears on and off throughout the year, with a slightly greater rate of occurrence in April and September (usually before the occurrence of cluster 3 and after the one of cluster 4), and accounts for 518 days (23.6%). It happens mostly in late spring and early winter, i.e., during the seasonal transition of the monsoon. This suggests that cluster 2 represents

a transition regime for the river plume. In this cluster, the river plume extends both further offshore and further southward compared to cluster 1. 75% of the time, the plume extends to 105.85°E (∼25 km from the coast) while 25% of the time, it extends to 106°E (∼40 km from the coast). The low salinity strip defined by the 30 isohaline is about twice as wide as in cluster 1 (**Figure 10J**). 50% of the time, the plume reaches 18°N, which is the same as in cluster 1, while 25% of the time, it extends as south as 17.1°N (**Figure 9B**). The higher spatial coverage of the plume area in this cluster can be explained by the higher runoff (4020 m^3/s) than in cluster 1 (2197 m^3/s). Also, in this cluster, the wind direction is from the southeast, therefore different with respect to cluster 1 (**Figure 10B**). The wind speed is weak, with a mean speed of 4.3 m/s near the coast (4.7 m/s at the gulf scale). The downwelling effect is relaxed and the light plume water can spread seaward. The coastal current is still flowing southward but is weaker than in cluster 1. As in cluster 1 and cluster 4 (see below), it is locally intensified, around 106.5°E, 18°N. The tidal current is intensified there as well (**Supplementary Figure 1**), probably because this is an area where the shelf is thinning and the coastline draws a cape. When present, the coastal current may be intensified there, through the same processes as for the tidal current. Another assumption is that interactions between tides and the coastal current occur there and result in an intensification of the local circulation.

The third cluster, which happens primarily in June and July (315 days, 14.4%), has a different shape with respect to clusters 1 and 2 (**Figure 9C**). In this cluster, the plume is advected northward and seaward. This is also the cluster in which the river plumes from the various rivers are disjoint. It corresponds to the period when the runoff increases with the mean value equal to 4631 m^3/s and the summer monsoon wind reaches its strongest intensity. The mean gulf-scale wind speed is 5.7 m/s (5.6 m/s inside the box). It blows northeastward therefore corresponding to upwelling conditions. The surface coastal current is not present anymore south of ∼18.7°N; at 18°N the surface current is oriented seaward, with weak speed (∼10–15 cm/s). North of 18.7°N, the coastal current is reversed, flowing northward, with a weaker speed than in cluster 2 (∼15 cm/s), but locally intensified by the river runoffs. Although the average runoff is larger than for cluster 2, the very low salinity strip (SSS < 25) is smaller than in cluster 2, suggesting that strong mixing is taking place, either vertically due to wind-induced turbulent kinetic energy or horizontally. **Figure 9C** shows that 5% of the time, the plume reaches the 40 m isobath, and indeed the salinity over this area is decreased with respect to cluster 2, supporting the hypothesis of lateral mixing and/or stirring. The involved processes are assumed to be at daily timescales or shorter (inertial, tidal time scales) and are discussed in section "Variability of the Plume Dynamics at Daily Time Scales."

The fourth cluster exhibits the largest spatial coverage and occurs mostly in August and September (367 days, 16.7%), that is at the peak of the high runoff season (the mean runoff over this cluster reaches 7336 m^3/s). In this cluster, the plume extends the farthest offshore. At 19°N, it can extend to more than 100 km from the coast. 75% of the plume extends to ∼55 km from the coast, while 25% of the plume extends to 100 km from the coast. 5% of the plume can extend as far east as 107.0°N, which is 145 km from the coast (**Figure 9D**). In this cluster, the mean wind pattern is still upwelling-favorable, i.e., capable of driving the plume offshore (**Figure 10D**). However, the wind speed is less intense: 4.5 m/s at the gulf scale and 4.0 m/s inside the box only. It is the period with the lowest wind speed and the weakest coastal current; the latest is downcoast again. Its speed does not exceed 10 cm/s (except around 106.5°E, 18°N where it is locally accelerated up to more than 20 cm/s as in clusters 1 and 2, **Figure 10H**). The plume is surface-advected, which is verified by the analysis of the plume thickness in the next section. The large runoff during the August–September period is responsible for the very low salinity (SSS < 25) simulated along the coast between 21°N and 19.5°N. As for cluster 3, the salinity over the whole shelf in the western half of the GOT is decreased in this cluster, with respect to the previous one, indicating a larger extent of the river water influence than the one formalized by the plume definition that we adopted in section "River Plume in GOT: Identification and Variations of Area." In other words, even highly diluted in the ocean waters, the river water is still influencing the ocean SSS. The comparison between the reference run and the run without runoff (GOT_NORIV) confirms that the runoff influences the SSS all over the gulf in clusters 2, 3 and 4 (**Supplementary Figure 2**), with the largest influence in cluster 4 (more than 1 west of 107°N and north of ∼20.5°N). The SSH shows a small elevation between the center of the basin and the Vietnamese coast, consistently with a weak downcoast circulation (**Figure 11P**). In GOT_NORIV, the SSH east of 107°E is decreased by 1 to 2.5 cm with respect to REF (not shown), while the surface current only differs from the REF run locally and the differences are smaller than for cluster 1 (and to a lesser extent for cluster 2), therefore suggesting a weak contribution of the riverine waters to the circulation in late summer.

We estimate the export of riverine waters by computing the daily fluxes of dye concentration (for the RR and southern rivers) through three sections (**Figure 2A**). We then average the fluxes over the periods of the cluster to obtain a mean daily flux for each cluster. We found that the highest offshore flux (section 2) occurs in cluster 3 (∼402,000 unit/s), while the largest downcoast flux (section 1) is in cluster 2. Overall the largest exports are across section 1 in clusters 2 (473,532 unit/s) and 4 (431,408 unit/s). The flux across section 3 is much smaller in all the clusters; it is the largest (northward) in cluster 3 (76,000 unit/s).

Plume Thickness

We define the plume thickness as the maximum depth at which water is found with a tracer concentration greater than 7 unit/m^3. It is computed daily. **Figure 11** shows the time evolution of plume thickness and temperature at 5 m along the 19°N section. For the sake of clarity, we show years 2013 and 2014 only, representative of high and low discharge conditions respectively, but similar conclusions would be drawn for the other years (**Supplementary Figure 3**).

As we have seen in the previous section, from September to March, the plume is narrow and elongated along the coast due to the winter monsoon downwelling winds. During that period the plume is bottom attached, i.e., it is filling the whole water

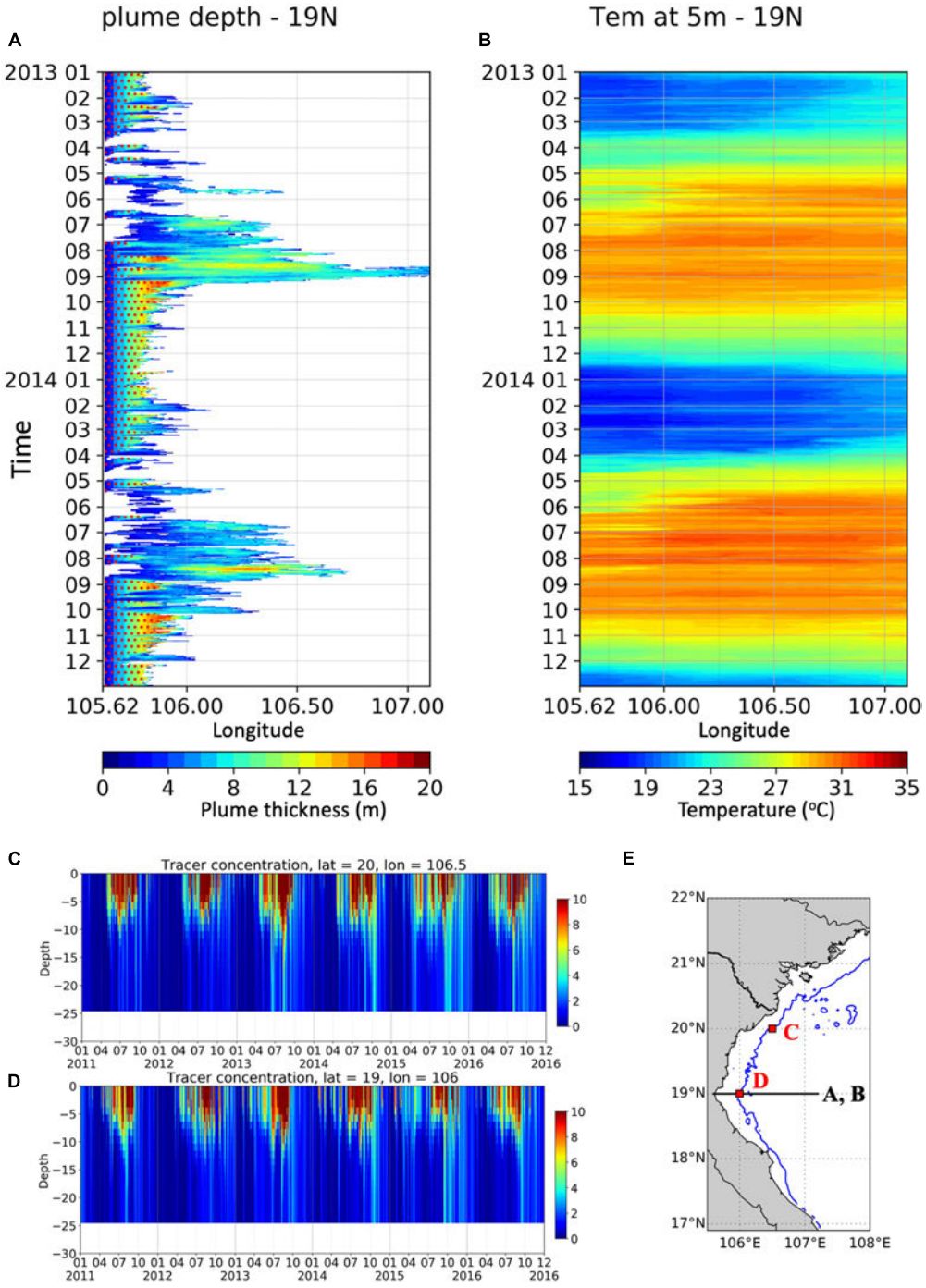

FIGURE 11 | (A) Plume thickness (m) at 19°N section from 2013 to 2014. Red points indicate the area where the plume is present over the whole depth. **(B)** Temperature (°C) at 5 m at the same section as **(A)**. **(C,D)** Tracer concentration over time at two coastal points. The location of the section and the points are shown in panel **(E)**.

column with a thickness of the order of 10 m at 20°N and 15 m at 19°N (**Figure 11A**). Several factors can explain such a thickness: first the downwelling winds tend to create a convergent flow onshore; secondly the vertical mixing may be enhanced by the bottom-generated turbulence resulting from the strong coastal current (Wiseman and Garvine, 1995). Wiseman and Garvine (1995) also suggest that onshore Ekman flow may result in dense ocean water overriding buoyant plume water, therefore creating instabilities and mixing.

From May on, the monsoon changes to southwesterly wind, driving the plume northward and detaching it from the coast and from the bottom. We observe that events where the plume detaches from the coast at 19°N coincide with the temperature at the coast colder than seaward (**Figure 11B**). This supports the assumption that the southerly monsoon winds generate a coastal upwelling. The offshore extension of the low-temperature signal is less than 40 km. This summer upwelling was also simulated by Gao et al. (2013) for 2007, and is likely to be responsible for the detachment of the plume from the coast. When detached, the plume thickness increases from ∼4 m onshore to about 10 m seaward. Sometimes in May-August, the plume is thickening offshore to 12–15 m (**Figure 11A**). That deepening may be due to strong winds events that increase the mixing offshore in the upwelling region or to the increase of river discharge.

In mid-August, we observe a strong plume deepening. In 2013, the plume is still attached to the coast, contrary to 2014 (**Figure 11A**). This difference can be due to two reasons. Firstly, the runoff in 2013 is higher than in 2014. Secondly, the wind in 2014 is stronger than in 2013 (not shown). However, both deepening events appear to be linked with a seasonally recurrent anti-cyclonic eddy developing near 19°N, which is examined in more detail in the next section.

We also look into detail at the temporal evolution of the tracer concentration over the water column at some points near the coast (**Figures 11C–E**). After a peak discharge in summer, from October until November or December (depending on the year), the tracer is rapidly mixed down to the bottom (∼25 m) at a low concentration (3–5 unit/m^3). Then, from January to March, the concentration gets close to zero, except for some sporadic events of a few days.

Variability of the Plume Dynamics at Daily Time Scales

The cluster analysis allowed to identify clear seasonal patterns of the plume variability, but the analysis of the plume thickness variability highlighted a strong variability at daily time scales as well. In particular, the plume is observed to spread seaward in spring and summer, which is synonymous of export of fresh water and riverine materials from the coastal zone to the interior of the GOT. Those sporadic events are represented by the 5% of plume occurrence in **Figure 9**. **Figures 12A–C** illustrates some of these events for three dates in June 2012. They show that the river waters spread within the whole basin, with a low concentration though. They extend as far east as Hainan coasts. As the forcings (wind, runoff, tides) are also strongly variable, it is not possible to disentangle their respective impacts. Besides, the plume seems to be shaped by small-scale circulation patterns over the whole shelf, suggesting that mesoscale or submesoscale activity strongly influences the horizontal transport as well as the dilution.

In particular, at 19°N, we observe a recurrent anticyclonic eddy as mentioned in the previous section. **Figures 12D,E** shows the relative surface vorticity and the plume thickness on 11 and 15 August 2014. The eddy is depicted by the minimum of negative surface vorticity. It appears clearly that the plume is deepening at the center of the eddy, with a difference of thickness reaching 8 m between the edge and the center for instance on 15 August 2014. In our simulation, this eddy generally happens in August, when the wind direction is southwest, which is upwelling-favorable, and the runoff is high. Then, this eddy disappears when the coastal southward current develops again and/or the wind is not favorable anymore. Its lifetime varies from a few days to ∼15 days; its diameter is about 50 km, i.e., much larger than the first Rossby radius. There is no mention about such an eddy from the HFR analysis (Rogowski et al., 2019; Tran et al., 2021); the availability and resolution of the HFR measurements may be too limited to depict such a pattern. The formation mechanism of this eddy is unclear. The comparison with GOT_NORIV shows that most of the time, without the river, some vorticity gradients can be depicted, suggesting a weak anticyclonic structure which vanishes on a shorter time scale (for instance in 2013, 6 days in GOT_NORIV instead of 8 days in GOT_REF) (**Supplementary Figure 4**). The anticyclone seems strongly connected to the Lam river plume. Bottom topography may be another forcing factor. Indeed a small island is present at 106°N and is likely to influence the river flow. Both the buoyant Lam river plume and the island may generate instabilities of the coastal flow that could lead to the eddy development.

Impact of Tides

The influence of tides in shaping the far field plume is investigated by comparing the reference simulation GOT_REF with the simulation without tides (GOT_NOTIDE). We first compute the total area of the plume from all the Vietnamese rivers (including the RR). On average, the plume in GOT_REF is 5% larger than in GOT_NOTIDE, with the largest differences found in the high discharge period (not shown). Significant differences (∼10%) are observed during the summer of 2011 and 2012, with the plume being larger in the GOT_REF run; over the remaining period, the differences are weaker.

We performed a KMA analysis on the GOT_NOTIDE outputs for 4 clusters; the resulting clusters are very close to the ones from GOT_REF, both in terms of spatial structures and temporal distribution (**Supplementary Figure 5**). The main differences are found in clusters 3 and 4, that is during the summer monsoon and high runoff period. Without tides, the plume spreads more to the north. In cluster 4, the plume spreads a bit further to the south without tides.

Besides, of equal importance are the tidal effect on vertical mixing (already inside the estuaries) and frontal activities (Guarnieri et al., 2013). The impact of tides on the vertical mixing is investigated by comparing the tracer concentration between

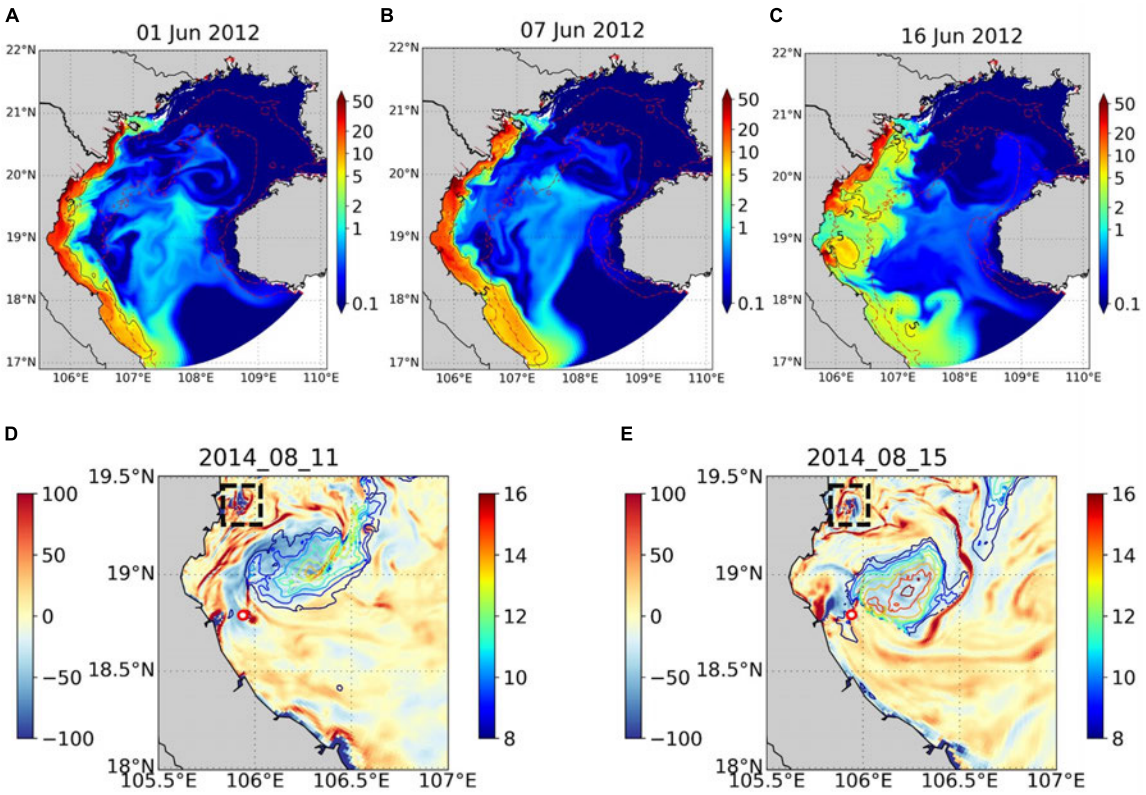

FIGURE 12 | (A–C) Surface tracer concentration on 01 June 2012 **(A)**, 07 June 2012 **(B)**, and 16 June 2012 **(C)**. **(D,E)** Relative surface vorticity ($10^{-6}s^{-1}$) (contour fill, left color bar) and plume thickness (m) (line, right color bar) on 11 and 15 August 2014. The black box and red circle indicate the location of two islands.

GOT_REF and GOT_NOTIDE (**Figure 13**). The tidal influence differs depending on the area: along the coast, just north of the mouth (close to ∼21°N), it is larger without tides at both surface and bottom (>4 unit/m^3). Everywhere else, the bottom concentration is larger with tides (>0.5 unit/m^3). At the surface, the concentration is smaller at the mouths (>4 unit/m^3) and larger in the plume (0.5–1 unit/m^3) in the run with tides. Over the shelf, the impact is small; in clusters 3 and 4 we observe a larger surface concentration (>0.5–1 unit/m^3) in the run without tides at ∼ 107–108°E and ∼ 20°N. These differences are small but significant with regard to the concentrations in GOT_REF (**Figure 13**): the values are 5–50 unit/m^3 (resp. 1–20 unit/m^3) in the plume and 0.5–10 unit/m^3 (resp. 0.1–1 unit/m^3) over the shelf west of 108°E at surface and at the bottom, respectively.

We interpret the results as follows: tides enhance the vertical mixing, which explains the larger concentration at the bottom in GOT_REF. However, tides also enhance the export of riverine waters offshore at the mouth, hence the slightly larger surface concentration values in GOT_REF in the plume and the smaller concentration close to the mouth. The plume reaches the small coastal area north of the RR mouths only in the absence of tides as mentioned previously, so the surface and bottom concentrations are larger in GOT_NOTIDE.

DISCUSSION

In this section, we summarize our findings, presenting them with paradigms found in the literature on plume dynamics (classification) or on environmental applications (receiving basins).

Classification

Several attempts have been made in the past to classify buoyant coastal runoffs with the objective of deriving general dynamical properties and to compare different plume systems. Garvine (1995) introduced the Kelvin number (K) which is defined as the ratio of the typical cross-shore scale of the buoyant runoff (usually taken as the estuary mouth width) over the local internal Rossby radius. K measures the importance of rotation: for $K < 1$, the influence of earth rotation on the plume dynamics is not significant and the river outflow forms energetic jets whose direction is controlled by local bathymetry and coastline (Hetland and Hsu, 2013). When $K > 1$, the Coriolis effect is dominant and the plume flows downcoast forming a coastal current; for large K, the downcoast flow may be insensitive to wind-driven motion in the opposite direction (Wiseman and Garvine, 1995).

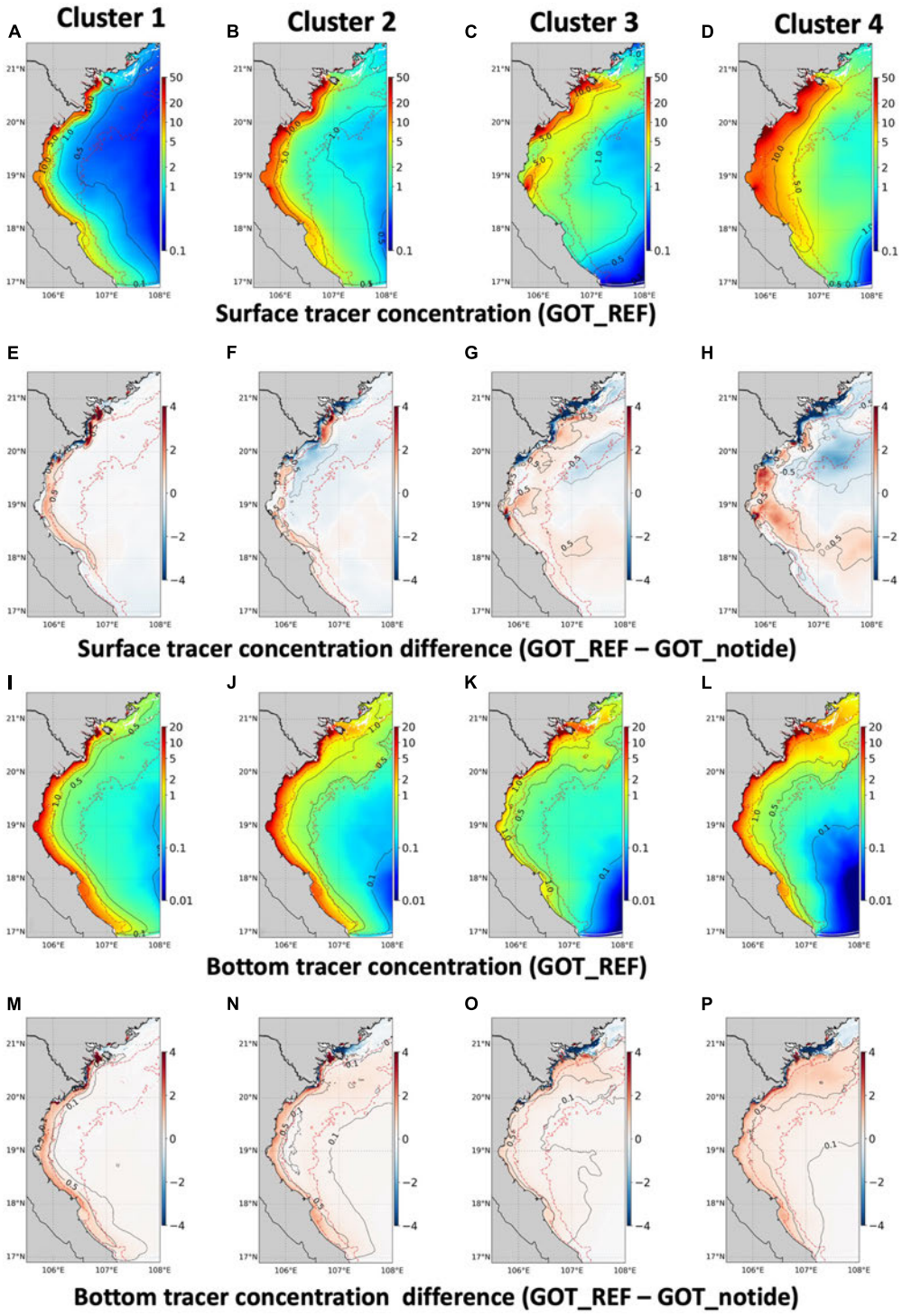

FIGURE 13 | (A–D) Maps of mean surface tracer concentration correspond for 4 clusters. **(E–H)** Difference of surface tracer concentration between GOT_REF and GOT_NOTIDE correspond for 4 clusters. **(I–L)** Same as **(A–D)** but at the bottom. **(M–P)** Same as **(E–H)** but at the bottom.

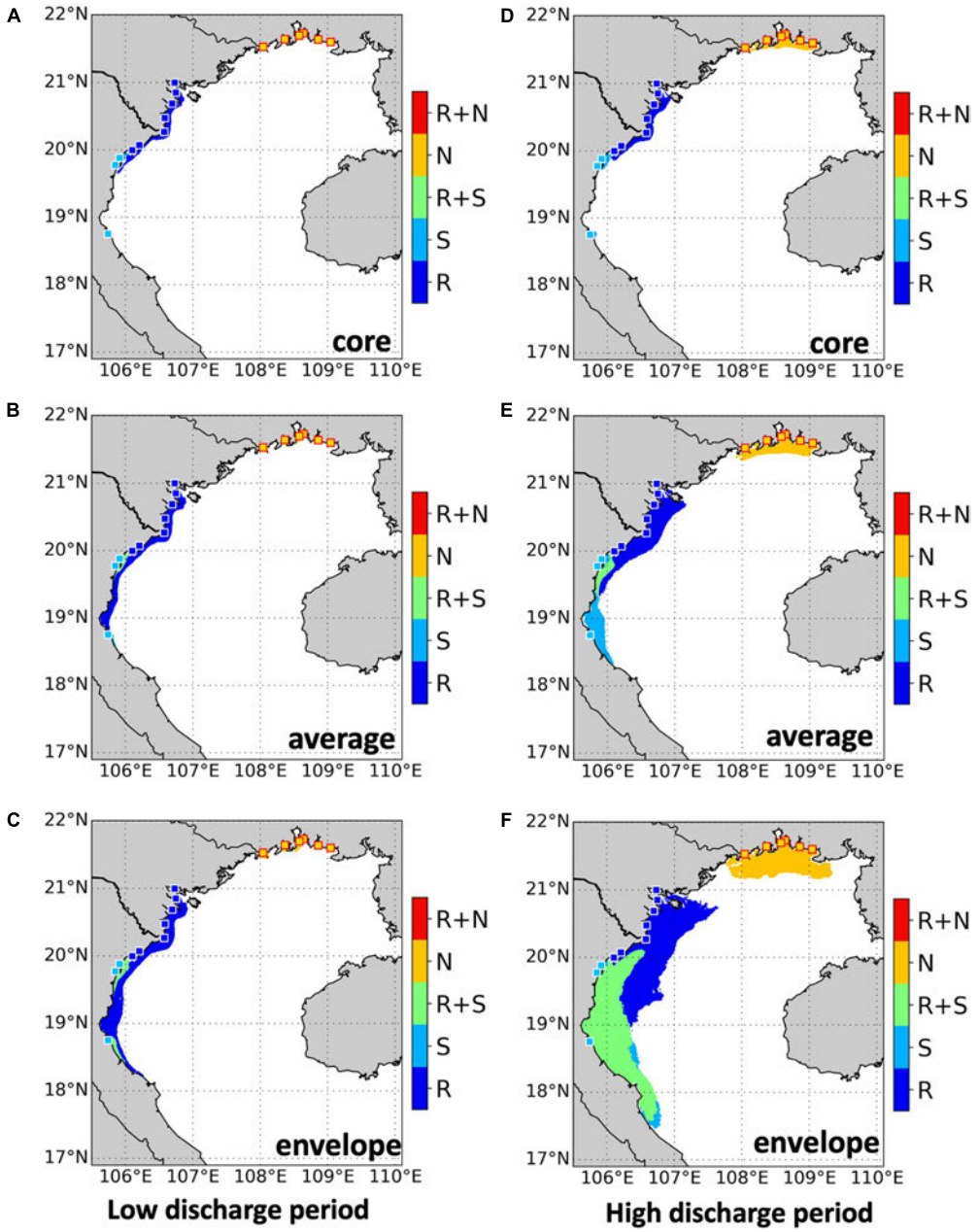

FIGURE 14 | The "receiving basins" from different rivers. The "core of the receiving basin" is the area where the plume is present at least 90% of the time **(A)**, the "average receiving basin" is the area where the plume is present at least 50% of the time **(B)**, "the envelop of the receiving basin" is the area where the plume is present at least 10% of the time **(C)**, all in low runoff period. **(D–F)** Same as **(A–C)** but in high runoff period. The location of the rivers is indicated by little boxes with the same color as the corresponding receiving basins. Explanation: (R): Red River plume, (S): plume of the rivers south of the RR delta, (N): the rivers north of the RR delta, (R+S): both R and S are present, (R+N): both R and N are present.

Very close to the RR mouths, the internal Rossby radius is between 1.5 and 5 km. Given the typical width of the distributaries mouths in our configuration (from 0.6 to 2.5 km), the Kelvin numbers vary between 0.5 and 1. Therefore in theory, bulges could be formed and the impact of the Coriolis force should be small. Close to the mouth, the currents for each cluster

do not show evidence of bulge formation. The outflow from the Cam river is very much constrained by the bathymetry and coastline. At the Van Uc and Ba Lat mouths, the outflow forms a kind of jet. In clusters 1 and 2, the plume is deflected downcoast and merges into the strong (wind-generated) coastal current (not shown). At the mouth of the Lam river, we identified daily scenes where the outflow has the shape of a bulge with a recirculating current (not shown). There the Kelvin number is estimated around 0.35, so a recurrent bulge might possibly be formed.

However, as in other delta systems, the typical length scale of the plume, and therefore the Kelvin number K, is not defined unambiguously, according to whether one considers the individual mouths or the system created by the merging plumes. Indeed, the plumes of the Red River distributaries (plus the Ma and Yen rivers) interact and merge in all clusters except cluster 3. Therefore, the overall fate of the Red River is likely to be better described by considering the system as a whole, as suggested by the 'menagerie' of plumes summarized by Horner-Devine et al. (2015; their **Figure 5**).

The Receiving Basin of Different Rivers in the GOT

As described in section "Observational Data Sets Used to Evaluate the Simulation," the introduction of several tracers allows us to follow the pathway of different river systems: Red River, the rivers south of the RR delta (SR) and the rivers north of the RR delta (NR). This information is helpful for many practical purposes and environmental applications (e.g., contaminant dispersion, water quality). Mapping the 'receiving basin' of a river is a useful tool to estimate the zone of influence of dissolved or particulate terrestrial inputs, with the aim to design evaluation strategies of the coastal ocean ecological status and to build scenarios for the preservation or restoration of vulnerable areas. This is done, for instance, by Menesguen et al. (2018) in a numerical study to identify the source of eutrophication in the Bay of Biscay and the English Channel and to design a strategy of nutrient load reduction within European institutions directives framework. In this study, we define the "receiving basin" for different rivers from the passive tracer content with the same threshold as defined in section "River Plume in GOT: Identification and Variations of Area" (concentration > 7 unit/m^3), but this time we apply it to each river system. The calculation is made for the high runoff/summer monsoon period (July–August–September, hereafter JAS) and for the low runoff/winter monsoon season (December–January–February, hereafter DJF) from 2011 to 2016 (**Figure 14**). The choice to work with 'classical' seasons, and not on the clusters of section "Variability of the RR Plume," is motivated by the concern to provide maps that can be easily used by the readers of this article. For each season, the "core of the receiving basin" is the area where the plume is present at least 90% of the time, the "average receiving basin" where the plume is present at least 50% of the time, "the envelope of the receiving basin" where the plume is present at least 10% of the time. **Figure 14** also shows areas where the cores/averages/envelopes of the receiving basins for two different river systems superimpose.

In the low runoff period (DJF, **Figures 14A–C**), the receiving basins are very narrow and elongated along the coast. While the envelope of the RR receiving basin can extend to approximately 18.2°N and mix with the plume of SR, its core appears mostly at the river mouths. In all cases, the receiving basins of NR is not detected with this choice of threshold for the tracer concentration (see below).

In the high runoff period (JAS, **Figures 14D–F**), the envelopes of the receiving basins of RR and SR are connected. South of 19°N, the plumes of RR and SR are fully mixed and their envelopes are indistinguishable. North of 19°N, the eastward extension of the RR envelope exceeds the one of the SR envelope and extends as eastward as 107.5°E. The receiving basin of NR appears north of 21°N but does not connect with the RR one. The average basins show different characteristics than the envelopes: while the SR influence can reach 18.3°N, the RR basin does not extend further south than 19.3°N.

In **Figure 14**, there is no connection between the RR and NR. However, it does not mean that they do not interact with each other. If we consider a lower threshold to identify the plume (e.g., 3 unit/m^3), the envelopes of NR and RR are connected both in low and high discharge periods (not shown). This suggests that the NR waters are strongly and quickly diluted. This also means that depending on the objectives of the study, and therefore on the choice of the tracer concentration threshold, the basins will be connected or not.

As written before, the information about receiving basins can help in case of water quality monitoring or contamination leakage. This analysis leads to 2 concluding remarks. Firstly, the freshwater along the Vietnamese coasts tends to flow southward and to exit the gulf. However, the freshwater from all the rivers tends to dilute to under 7% before it leaves the domain. This means that in case of contaminant leakage, the contaminants tend to mix and affect the gulf's water quality locally mostly. Secondly, the RR average receiving basin extends further southward and in a wider coastal strip than the SR basin during the low runoff season, while the SR basin extends further southward and further eastward (south of 19.5°N) than the RR basin during the high runoff period (**Figures 14B,E**). In both cases though, the RR runoff is higher than the SR runoff. These contrasted situations highlight the importance of wind and coastal circulation on shaping the receiving basin and consequently, on the fate of terrestrial inputs depending on their origin (SR or RR rivers systems).

CONCLUSION

We presented a comprehensive study of the plume formed by the waters from the Red River and three nearby rivers in the GOT using numerical simulation, in a realistic high-resolution configuration over the period 2011–2016. Compared to various observational data, the model shows good results. We then compare several methods to identify the river plume in the study area. We found that identification through passive (dye-like) tracers is preferable since it allows to distinguish the runoff influence from the precipitation one and to distinguish

the runoff from different rivers as well. The runoff shows large seasonal and interannual variability. It usually reaches the highest value in August and lowest in February, in phase with the monsoon system. At the surface, the plume area shows similar variations. However, the plume area reaches its peak in September, which is about 1 month later than the peak of runoff.

To identify the main spatial patterns of variability, we apply a clustering method to daily scenes of the model outputs where the plume is defined from the tracer concentration (>7 arbitrary unit/m^3). The cluster analysis identifies the plume regimes and their period of occurrence without having to pre-define these periods as one would do when computing seasonal averages for instance. Besides it allows identifying transition regimes (e.g., cluster 2).

The plume pattern of the first cluster is characteristic of the November–March period: the plume is narrow and contained within the inner shelf (bottom depth < 20 m), due to low runoff and downwelling wind (winter monsoon) with an intense downcoast current. The second pattern is mainly observed in the transition period of the monsoon (April, May, October) and is wider than the first pattern; both the relaxation of the winter monsoon and the weaker coastal current allow the plume to spread further offshore than in cluster 1. The third pattern occurs when the summer monsoon is the strongest and the runoff increases. The wind is upwelling favorable; the plumes are advected northward and detach locally from the coast. It is the only pattern where the plumes from the different rivers are disconnected. The final pattern coincides with the highest runoff period and the summer monsoon relaxation. The plumes are connected again in a pool of low salinity waters with the largest coverage, spreading both offshore and southward. Overall, we found that this hydrological system with multiple source points and a strong seasonal variability of along-shore wind is similar to the one for instance described by Kourafalou et al. (1996) in the Southern Atlantic Bight along the US coast.

The vertical variability of the plume is examined as well. In winter, the plume is usually mixed over the whole water depth. When the summer monsoon arrives at the end of March, the plume starts to detach from the bottom and spreads offshore as a surface buoyant layer. The plume can detach from the coast near 19°N when a coastal upwelling develops. It may deepen offshore also. In particular, the plume is strongly deepened in the middle of August when trapped in a recurrent eddy near 19°N.

The cluster analysis therefore evidenced at least three regimes. In winter regime, the plumes of all the Vietnamese rivers are connected in a well-mixed freshwater strip which is contained to the coast; the downcoast current is likely to act like a barrier preventing exchanges of riverine water and material cross-shore toward the center of the gulf. In early summer, the upwelling winds lead to a reverse circulation, where the plumes are advected northward and slightly offshore; the plumes are partly disconnected from each other. In the late summer regime, the wind relaxes and the large amount of freshwater due to the annual runoff peak spreads at the surface, reaching the maximal offshore and southward extension.

The impact of tides on the shape of the plume has been explored using a dedicated simulation without tides; we found that the general shape of the plume is not significantly influenced by tides, except in the Ha Long Bay area during the high runoff period (clusters 3 and 4). The main forcings of the far-field plume are the wind and basin-scale circulation. However, tides impact the vertical structure of the plume: they enhance the vertical mixing hence the riverine water concentration in the bottom layer. They also lead to a larger export offshore or the riverine water at the surface close to the mouth during the high discharge season. The influence of tides is obviously not limited to the far field; first, tides determine the buoyancy of the river water entering the ocean. Then, in the near and mid-field area, the bottom friction induced by tides is likely to impact the local dynamics and mixing, therefore influencing the plume properties in the far-field as well. An in-depth analysis of the tidal processes, of the tides-river flow interactions and of their impact on the velocity, shear and stratification in the estuaries and at the mouths was beyond the scope of this paper. It will be the topic of a future study aimed at better understanding the shape and dynamics of the outflow in the near and mid fields. The dynamics at the mouth is also highly important for sediment dynamics close to the mouth area.

Small-scale patterns of the coastal circulation, such as the recurrent eddy observed at 19°N in August, are also involved in the dilution and fate of the riverine water in the ocean. Submesoscale features due to instabilities within the plume or at the fronts may also develop and influence the mixing as found in other systems (e.g., Horner-Devine et al., 2015; Ayouche et al., 2021 this issue). Such small-scale processes should be examined in dedicated studies.

At last, this study provides some reference information about the river plume variability, showing in particular the role of the wind. In such a context, ensemble simulations will be useful to assess the sensitivity of our results due to the uncertainty of the forcing conditions (wind and runoff).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article (i.e., the model simulations) will be made available by the authors without undue reservation.

AUTHOR CONTRIBUTIONS

TuN-D, NA, PM, and PDM-F contributed to the conception and scientific design of the study. PM provided the model. VP, MH, TD, and PM developed the previous GOT configuration used for this analysis. MT provided the HFR data. TuN-D, NA, and FT performed most of the analyses. TuN-D wrote the first draft of the manuscript. NA wrote sections of the manuscript. All authors actively contributed to manuscript revision, read, and approved the submitted version.

FUNDING

TuN-D Ph.D. thesis is funded by the "Programme d'Excellence" of the French Embassy of Vietnam and by IRD ARTS program. This work is a contribution to the LMI LOTUS project (http://lotus.usth.edu.vn) and to the SWOT COCTO-FO project funded by CNES (French Spatial Agency). This work has used HPC resources from CALMIP/France (grant 2020-p1119) and HILO at the University of Science and Technology of Hanoi (USTH). Support from IRD, CNES, CNRS, and USTH is gratefully acknowledged.

ACKNOWLEDGMENTS

CFO CTD data used in this study was collected by the Vietnam Center for Oceanography (CFO, VASI) in the framework of the United States–Vietnam collaboration project: "Gulf of Tonkin Circulation study" (NICOP N62909-15-1-2018), financed by the Office of Naval Research (ONR). The VITEL (VIetnam TELédétection remote sensing) campaign has been funded by CNES; we are grateful to S. Ouillon for providing the T/S data. The historical tidal gauges dataset stems from the International Hydrographic Organization (https://www.iho.int/, last access: November 2, 2019) and is available upon request at https://www.admiralty.co.uk/ukho/tidal-harmonics (last access: November 2, 2019). The altimetric tidal constituents were provided to us by F. Lyard (LEGOS). We also acknowledge F. Lyard for the use of his code to compute the Rossby Radius and for his expertise on the model's bathymetry field. This study has been conducted using ECMWF product (www.ecmwf.int) and E.U. Copernicus Marine Service Information (product: GLOBAL_ANALYSIS_FORECAST_PHY_001_024, available at https://resources.marine.copernicus.eu/?option=com_csw&view=details&product_id=GLOBAL_ANALYSIS_FORECAST_PHY_001_024). Many thanks to C. Nguyen and D. Gazen for the help with the model and computing facilities at LAERO (Toulouse). The SYMPHONIE code and the code to compute the Rossby radius are distributed by the SIROCCO French national service website (https://sirocco.obs-mip.fr/), hosted by the Observatoire Midi-Pyrénées, Toulouse, France.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.772139/full#supplementary-material

REFERENCES

Ayouche, A., Charria, G., Carton, X., Ayoub, N., and Theetten, S. (2021). Nonlinear processes in the Gironde river plume (North-East Atlantic): instabilities and mixing. *Front. Mar. Sci.* 8:701773. doi: 10.3389/fmars.2021.701773

Bentsen, M., Evensen, G., Drange, H., and Jenkins, A. D. (1999). Coordinate transformation on a sphere using conformal mapping. *Mon. Weather Rev.* 127, 2733–2740.

Blumberg, A. F., and Mellor, G. L. (1987). "A description of a three–dimensional coastal ocean circulation model," in *Three-Dimensional Coastal Ocean Models*, ed. N. S. Heaps (Washington, DC: American Geophysical Union), doi: 10.1029/CO004p0001

Burla, M., Baptista, A. M., Zhang, Y., and Frolov, S. (2010). Seasonal and interannual variability of the Columbia river plume: a perspective enabled by multiyear simulation databases. *J. Geophys. Res.* 115:C00B16. doi: 10.1029/2008JC004964

Chelton, D. B., deSzoeke, R. A., Schlax, M. G., El Naggar, K., and Siwertz, N. (1998). Geographical variability of the first baroclinic rossby radius of deformation. *J. Phys. Oceanogr.* 28, 433–460.

Chen, Z., Gong, W., Cai, H., Chen, Y., and Zhang, H. (2017). Dispersal of the Pearl river plume over continental shelf in summer. *Estuar. Coast. Shelf Sci.* 194, 252–262. doi: 10.1016/j.ecss.2017.06.025

Csanady, G. T. (1982). *Circulation in the Coastal Ocean*. Dordrecht: Springer, 281. doi: 10.1007/978-94-017-1041-1

Damien, P., Bosse, A., Testor, P., Marsaleix, P., and Estournel, C. (2017). Modeling postconvective submesoscale coherent vortices in the Northwestern Mediterranean Sea. *J. Geophys. Res. Oceans* 122, 9937–9961. doi: 10.1002/2016JC012114

Ding, Y., Chen, C., Beardsley, R. C., Bao, X., Shi, M., Zhang, Y., et al. (2013). Observational and model studies of the circulation in the Gulf of Tonkin, South China Sea. *J. Geophys. Res. Oceans* 118, 6495–6510. doi: 10.1002/2013JC009455

Falcieri, F. M., Benetazzo, A., Sclavo, M., Russo, A., and Carniel, S. (2014). Po river plume pattern variability investigated from model data. *Cont. Shelf Res.* 87, 84–95. doi: 10.1016/j.csr.2013.11.001

Gao, J., Shi, M., Chen, B., Guo, P., and Zhao, D. (2014). Responses of the circulation and water mass in the Beibu Gulf to the seasonal forcing regimes. *Acta Oceanol. Sin.* 33, 1–11.

Gao, J., Xue, H., Chai, F., and Shi, M. (2013). Modeling the circulation in the Gulf of Tonkin, South China Sea topical collection on the 4th international workshop on modelling the ocean in Yokohama, Japan 21–24 May 2012. *Ocean Dyn.* 63, 979–993. doi: 10.1007/s10236-013-0636-y

Garvine, R. W. (1995). A dynamical system for classifying buoyant coastal runoffs. *Cont. Shelf Res.* 15, 1585–1596.

Greenberg, D. A., Dupont, F., Lyard, F. H., Lynch, D. R., and Werner, F. E. (2007). Resolution issues in numerical models of oceanic and coastal circulation. *Cont. Shelf Res.* 27, 1317–1343. doi: 10.1016/j.csr.2007.01.023

Guarnieri, A., Pinardi, N., Oddo, P., Bortoluzzi, G., and Ravaioli, M. (2013). Impact of tides in a baroclinic circulation model of the Adriatic Sea. *J. Geophys. Res. Oceans* 118, 166–183. doi: 10.1029/2012JC007921

Hansen, D. V., and Rattray, M. (1966). New dimensions in estuary classification. *Limnol. Oceanogr.* 11, 319–326. doi: 10.4319/lo.1966.11.3.0319

Hastie, T., Tsbshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer, 533.

Hetland, R. D., and Hsu, T.-J. (2013). "Freshwater and sediment dispersal in large river plumes," in *Biogeochemical Dynamics at Large River–Coastal Interfaces: Linkages with Global Climate Change*, eds T. S. Bianchi, M. A. Allison, and W.-J. Cai (New York, NY: Springer), 55–85.

Horner-Devine, A. R., Hetland, R. D., and MacDonald, D. G. (2015). Mixing and transport in coastal river plumes. *Annu. Rev. Fluid Mech.* 47, 569–594. doi: 10.1146/annurev-fluid-010313-141408

Kourafalou, V. H., Lee, T. N., Oey, L.-Y., and Wang, J. D. (1996). The fate of river runoff on the continental shelf: 2. Transport of coastal low–salinity waters under realistic wind and tidal forcing. *J. Geophys. Res.* 101, 3435–3455. doi: 10.1029/95JC03025

Large, W. G., and Yeager, S. (2004). *Diurnal to Decadal Global Forcing for Ocean and Sea–Ice Models: The Data Sets and Flux Climatologies,* NCAR Technical Note, NCAR/TN-460+STR. Boulder, CO: National Center for Atmospheric Research. doi: 10.5065/D6KK98Q6

Liu, Y., MacCready, P., and Hickey, B. M. (2009). Columbia river plume patterns in summer 2004 as revealed by a hindcast coastal ocean circulation model. *Geophys. Res. Lett.* 36:L02601. doi: 10.1029/2008GL036447

Lyard, F. H., Allain, D. J., Cancet, M., Carrère, L., and Picot, N. (2021). FES2014 global ocean tide atlas: design and performance. *Ocean Sci.* 17, 615–649. doi: 10.5194/os-17-615-2021

MacCready, P., Banas, N. S., Hickey, B. M., Dever, E. P., and Liu, Y. (2009). A model study of tide– and wind–induced mixing in the Columbia river estuary and plume. *Cont. Shelf Res.* 29, 278–291. doi: 10.1016/j.csr.2008.03.015

Maraldi, C., Chanut, J., Levier, B., Ayoub, N., De Mey, P., Reffray, G., et al. (2013). NEMO on the shelf: assessment of the Iberia–Biscay–Ireland configuration. *Ocean Sci.* 9, 745–771. doi: 10.5194/os-9-745-2013

Marsaleix, P., Auclair, F., and Estournel, C. (2006). Considerations on open boundary conditions for regional and coastal ocean models. *J. Atmos. Ocean. Technol.* 23, 1604–1613. doi: 10.1175/JTECH1930.1

Marsaleix, P., Auclair, F., Floor, J. W., Herrmann, M. J., Estournel, C., Pairaud, I., et al. (2008). Energy conservation issues in sigma–coordinate free–surface ocean models. *Ocean Model.* 20, 61–89. doi: 10.1016/j.ocemod.2007.07.005

Menesguen, A., Dussauze, M., and Dumas, F. (2018). Designing optimal scenarios of nutrient loading reduction in a WFD/MSFD perspective by using passive tracers in a biogeochemical–3D model of the English Channel/Bay of Biscay area. *Ocean Coast. Manage.* 163, 37–53. doi: 10.1016/j.ocecoaman.2018.06.005

Michaud, H., Marsaleix, P., Leredde, Y., Estournel, C., Bourrin, F., Lyard, F., et al. (2012). Three–dimensional modelling of wave–induced current from the surf zone to the inner shelf. *Ocean Sci.* 8, 657–681. doi: 10.5194/os-8-657-2012

Neumann, L. E., Šimůnek, J., and Cook, F. J. (2011). Implementation of quadratic upstream interpolation schemes for solute transport into HYDRUS–1D. *Environ. Model. Softw.* 26:11. doi: 10.1016/j.envsoft.2011.05.010

Nguyen, N. M., Marchesiello, P., Lyard, F., Ouillon, S., Cambon, G., Allain, D., et al. (2014). Tidal characteristics of the Gulf of Tonkin. *Cont. Shelf Res.* 91, 37–56. doi: 10.1016/j.csr.2014.08.003

Nguyen Thi Hien, A., Vu Minh Cat, B., and Roberto Ranzi, C. (2020). "Upstream effects on salinity dynamics in the Red River Delta," in *APAC 2019*, eds N. Trung Viet, D. Xiping, and T. Thanh Tung (Singapore: Springer), 1439–1443.

Nugroho, D. (2017). *The Tides in a General Circulation Model in the Indonesian Seas.* Ph.D. thesis. Toulouse: Univ. Paul Sabatier 3.

Otero, P., Ruiz-Villarreal, M., and Peliz, A. (2008). Variability of river plumes off Northwest Iberia in response to wind events. *J. Mar. Syst.* 72, 238–255. doi: 10.1016/j.jmarsys.2007.05.016

Ouillon, S. (2014). *VITEL–NORTH Cruise, RV Alis*. doi: 10.17600/14004100

Pairaud, I. L., Lyard, F., Auclair, F., Letellier, T., and Marsaleix, P. (2008). Dynamics of the semi–diurnal and quarter–diurnal internal tides in the Bay of Biscay. Part 1: barotropic tides. *Cont. Shelf Res.* 28, 1294–1315. doi: 10.1016/j.csr.2008.03.004

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit–learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.

Piton, V., Herrmann, M., Lyard, F., Marsaleix, P., Duhaut, T., Allain, D., et al. (2020). Sensitivity study on the main tidal constituents of the Gulf of Tonkin by using the frequency–domain tidal solver in T–UGOm. *Geosci. Model Dev.* 13, 1583–1607. doi: 10.5194/gmd-13-1583-2020

Piton, V., Herrmann, M., Marsaleix, P., Duhaut, T., Ngoc, T. B., Tran, M. C., et al. (2021). Influence of winds, geostrophy and typhoons on the seasonal variability of the circulation in the Gulf of Tonkin: a high–resolution 3D regional modeling study. *Reg. Stud. Mar. Sci.* 45:101849. doi: 10.1016/j.rsma.2021.101849

Reffray, G., Fraunié, P., and Marsaleix, P. (2004). Secondary flows induced by wind forcing in the Rhône region of freshwater influence. *Ocean Dyn.* 54, 179–196. doi: 10.1007/s10236-003-0079-y

Rogowski, P., Zavala-Garay, J., Shearman, K., Terrill, E., Wilkin, J., and Tran, H. L. (2019). Air-sea-land forcing in the Gulf of Tonkin: assessing seasonal variability using modern tools. *Oceanography* 32, 150–161. doi: 10.5670/oceanog.2019.223

Shi, M., Chen, C., Xu, Q., Lin, H., Liu, G., Wang, H., et al. (2002). The role of Qiongzhou Strait in the seasonal variation of the South China Sea circulation. *J. Phys. Oceanogr.* 32, 103–121. doi: 10.1175/1520-04852002032<0103:TROQSI<2.0.CO;2

Solabarrieta, L., Rubio, A., Cárdenas, M., Castanedo, S., Esnaola, G., Méndez, F. J., et al. (2015). Probabilistic relationships between wind and surface water circulation patterns in the SE Bay of Biscay. *Ocean Dyn.* 65, 1289–1303. doi: 10.1007/s10236-015-0871-5

Sonnewald, M., Wunsch, C., and Heimbach, P. (2019). Unsupervised learning reveals geography of global ocean dynamical regions. *Earth Space Sci.* 6, 784–794. doi: 10.1029/2018EA000519

Sutton, W. R., Srivastava, J. P., Rosegrant, M., Thurlow, J., and Sebastian, L. (2019). *Striking a Balance, Managing El Niòo and la Niòa in Vietnam Agriculture,* RepNo 132068. Washington, DC: World Bank.

Toublanc, F., Ayoub, N. K., Lyard, F., Marsaleix, P., and Allain, D. J. (2018). Tidal downscaling from the open ocean to the coast: a new approach applied to the Bay of Biscay. *Ocean Model.* 124, 16–32. doi: 10.1016/j.ocemod.2018.02.001

Tran, M. C., Sentchev, A., and Nguyen, K. C. (2021). Multi-scale variability of circulation in the Gulf of Tonkin from remote sensing of surface currents by high-frequency radars. *Ocean Dyn.* 71, 175–194. doi: 10.1007/s10236-020-01440-x

Vaz, N., Rodrigues, J. G., Mateus, M., Franz, G., Campuzano, F., Neves, R., et al. (2018). Subtidal variability of the Tagus river plume in winter 2013. *Sci. Total Environ.* 627, 1353–1362. doi: 10.1016/j.scitotenv.2018.01.325

Vinh, V. D., Ouillon, S., Thanh, T. D., and Chu, L. V. (2014). Impact of the Hoa Binh dam (Vietnam) on water and sediment budgets in the Red river basin and delta. *Hydrol. Earth Syst. Sci.* 18, 3987–4005. doi: 10.5194/hess-18-3987-2014

Vlasenko, V., Stashchuk, N., and McEwan, R. (2013). High–resolution modelling of a large–scale river plume. *Ocean Dyn.* 63, 1307–1320. doi: 10.1007/s10236-013-0653-x

Wang, Y., Xue, H., Chai, F., Chao, Y., and Farrara, J. (2014). A model study of the Copper river plume and its effects on the northern Gulf of Alaska topical collection on the 5th International Workshop on Modelling the Ocean (IWMO) in Bergen, Norway 17–20 June 2013. *Ocean Dyn.* 64, 241–258. doi: 10.1007/s10236-013-0684-3

Wiseman, W. J., and Garvine, R. W. (1995). Plumes and coastal currents near large river mouths. *Estuaries* 18:509. doi: 10.2307/1352368

Wu, D., Wang, Y., Lin, X., and Yang, J. (2008). On the mechanism of the cyclonic circulation in the Gulf of Tonkin in the summer. *J. Geophys. Res. Oceans* 113, 2–11. doi: 10.1029/2007JC004208

Yankovsky, A. E., and Chapman, D. C. (1997). A simple theory for the fate of buoyant coastal runoffs. *J. Phys. Oceanogr.* 27, 1386–1401.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

APPENDICES

APPENDIX A: POST–PROCESSING OF THE CMEMS T,S FIELDS

On the continental shelves, the turbulence generated by the friction of the tidal currents on the bottom is generally significant. Near the bottom, the temperature and salinity stratification are in principle homogenized over a layer that can be up to several tens of meters, highly depending on the strength of the tides. CMEMS fields provide boundary conditions for temperature, salinity, surface elevation and currents associated with non–tidal processes. Because tides are not taken into account in the CMEMS simulation, the stratification is unfortunately insufficiently mixed near the bottom. As the errors induced on T and S are potentially important, we make a modification to the CMEMS T,S fields before introducing them at the open boundaries of our model. This modification consists in homogenizing T and S near the bottom, over a thickness deduced from the expected theoretical height of the tidally induced turbulent bottom layer. This height is estimated from the friction velocity associated with the tidal current, i.e. $D = 0.1u^*/f$ (Csanady, 1982) where f is the Coriolis parameter. The friction velocity, u^*, is calculated from the tidal harmonics of the depth–averaged current provided by FES2014, and a drag coefficient of $1e^{-3}$.

APPENDIX B: RIVER BOUNDARY CONDITIONS

In this appendix, we describe the way the inflow of river water is implemented at the most upstream point of the channel (if any) or directly at the river mouth into the ocean in case there is no channel. To simplify, we will call it the "river boundary condition at the river input point". (Note that for a given river, in case there is no channel, there can be several input points to represent a river mouth wider than the actual grid size; in that case, the boundary condition as described below is easily adapted.)

At the river input point, the river boundary current is consistent with the runoff of the river as in Reffray et al (2004), while the vertical shape of the current is given by a gradient condition allowing the current to adjust by itself to the dynamics of the river downstream (rather than arbitrarily prescribing a possibly inappropriate vertical profile). The boundary condition is applied in two steps. Step 1 is given by the above–mentioned gradient condition, leading to a provisional solution of the current:

$$\frac{\partial u^*}{\partial x} = 0 \qquad \text{if current in downstream direction of the river}$$

$$u^* = 0 \text{ otherwise}$$

where x is the direction of the river flow and $u*$ a provisional solution of the current. Step 2 ensures that the total flow is consistent with the expected runoff from the river. In practice, the river boundary current is proportional to u^*, i.e. $u = \alpha\, u^*$, where α is a constant used to satisfy the constraint on the prescribed river runoff:

$$\alpha \, \Delta y \int_{-h}^{\eta} u^*(z) dz = F$$

where Δy is the cellbox width at the river input point, h the bottom depth, η the sea surface height and F the river runoff.

Appendix C: Equation of the distance and procedure to identify the number of clusters.

The distance from a member (X) to the centroid (A) is calculated using Euclidean distance. The distance d(X,A) is computed for a given variable v defined at every model grid point i (i=1,N).The equation of the Euclidean distance is:

$$d\,(X, A) = \sqrt{\sum_{i=1}^{N} (X_i - A_i)^2}$$

where

Xi: value of variable v at grid point i of the member X

Ai: value of variable v at grid point i of the centroid of cluster A

N: the total number of grid points

As explained in section 2.4, the variable v at point i takes only two values: Xi=1 if the plume is present, Xi=0 otherwise.

In the K–means method, the number of clusters has to be chosen a priori. The number is usually a trade–off between the need to reduce the dimension of the ensemble of scenes to be analyzed and the search for meaningful distinct clusters. The number of clusters can be set subjectively (e.g. Solabarieta et al., 2015), based on a priori knowledge of the studied process (here the plume variability), or can rely on information criteria (Sonnewald et al., 2019).

In this study, since KMA is based on the distance to the cluster centroid, in order to evaluate how many clusters are suitable, we compute the sum of all the distances between the members and the centroids. The more clusters there are, the smaller the total distance will be decreased.

We applied the KMA for 8 cases, from 2 to 9 clusters. The total distance and the distance reduction are shown in Figure C1. We can clearly see that the passage from 2 to 3 clusters induces a significant reduction of the distance (more than 2×10^6). Similarly, from 3 to 4 clusters, the distance reduction is still large. However, from 4 to 5 clusters, the distance reduction is decreased significantly ($\sim 0.8 \times 10^6$). It means that using 4 clusters is suitable in terms of the total distance.

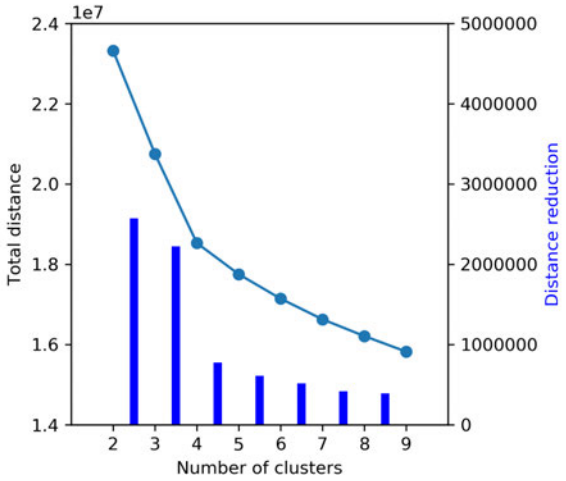

Figure C1: total distance as a function of the number of clusters (line – linked to left axis) and amount of distance reduction (bar – linked to right axis). The distance reduction is calculated by subtracting the total distance of (i) cluster and (i+1) clusters.

To double check, we run KMA for two cases: 4 clusters and 5 clusters. The objective is to verify if using 5 clusters can lead to a new unique pattern, or if it just adds a pattern that is similar to one of the 4 existing patterns (obtained when we compute 4 clusters only). Let's call the patterns of the 4–cluster analysis A1, A2, A3, A4; and the patterns of 5–cluster analysis B1, B2, B3, B4, B5.

The results show that cluster A3 has similar pattern and temporal distribution as cluster B3. It is the same remark for clusters A1, A4, B1 and B4.

The difference is mainly visible in cluster A2. When we use 5 clusters, cluster B2 and cluster B5 are similar to cluster A2, but cluster B2 has a little bit less offshore extension and cluster B5 has a little more offshore extension than A2 (Figure C2). In short, using 5 clusters does not create a new different pattern compared to using 4 clusters.

For all these above reasons, in this study, we decided to work with 4 clusters.

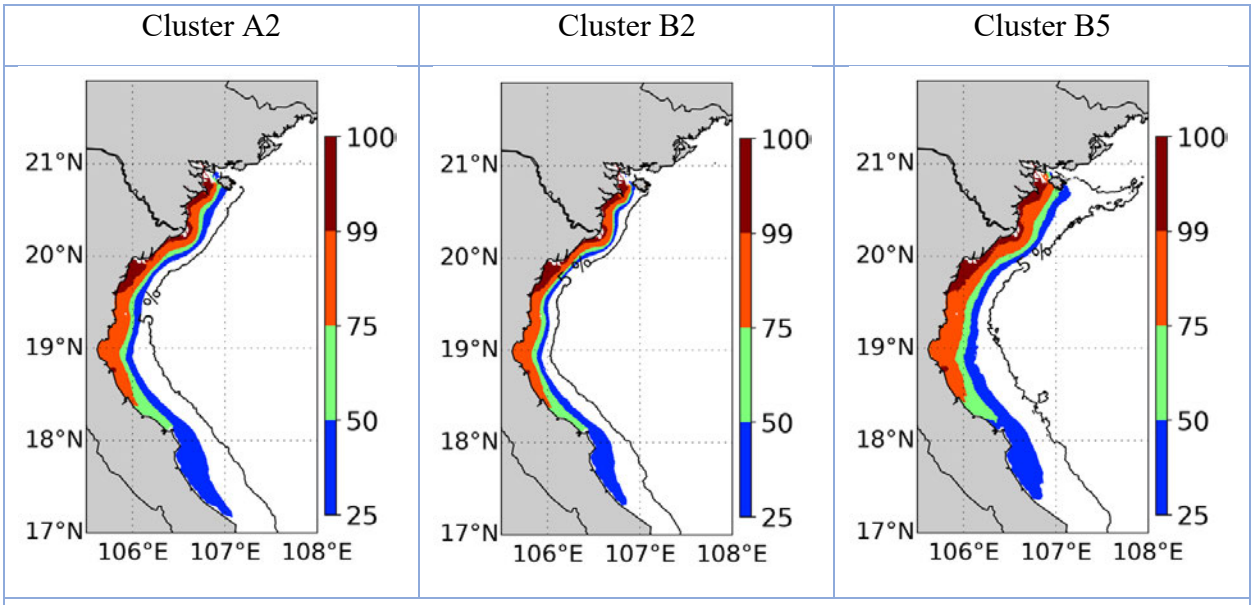

Figure C2: cluster 2 in 4–clusters analysis (cluster A2), cluster 2 in 5–clusters analysis (cluster B2) and cluster 5 in 5–clusters analysis (cluster B5)

Supplementary materials

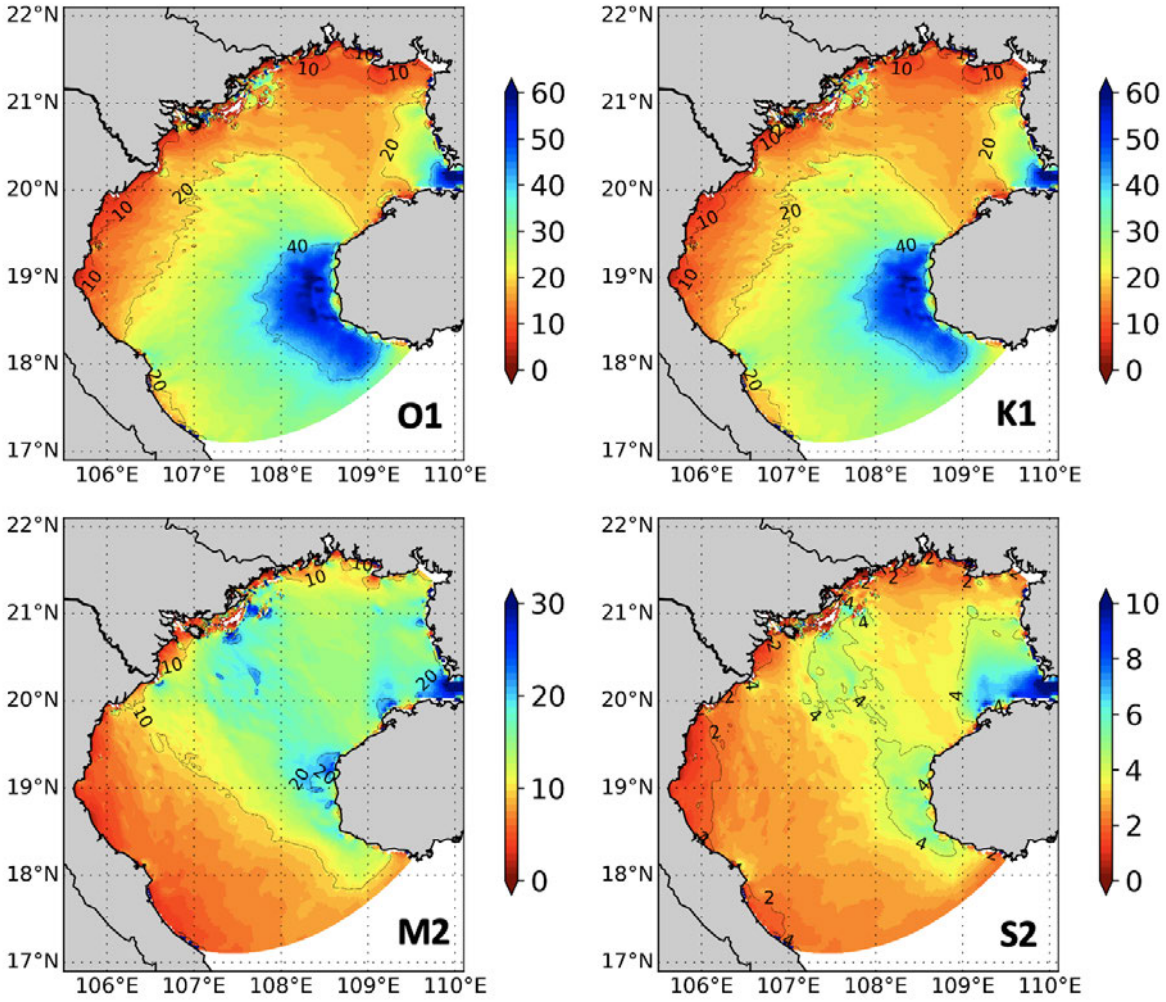

Figure S1: Maps of tidal current (cm/s) for 4 main constituents (O1, K1, M2, S2) from GOT_REF.

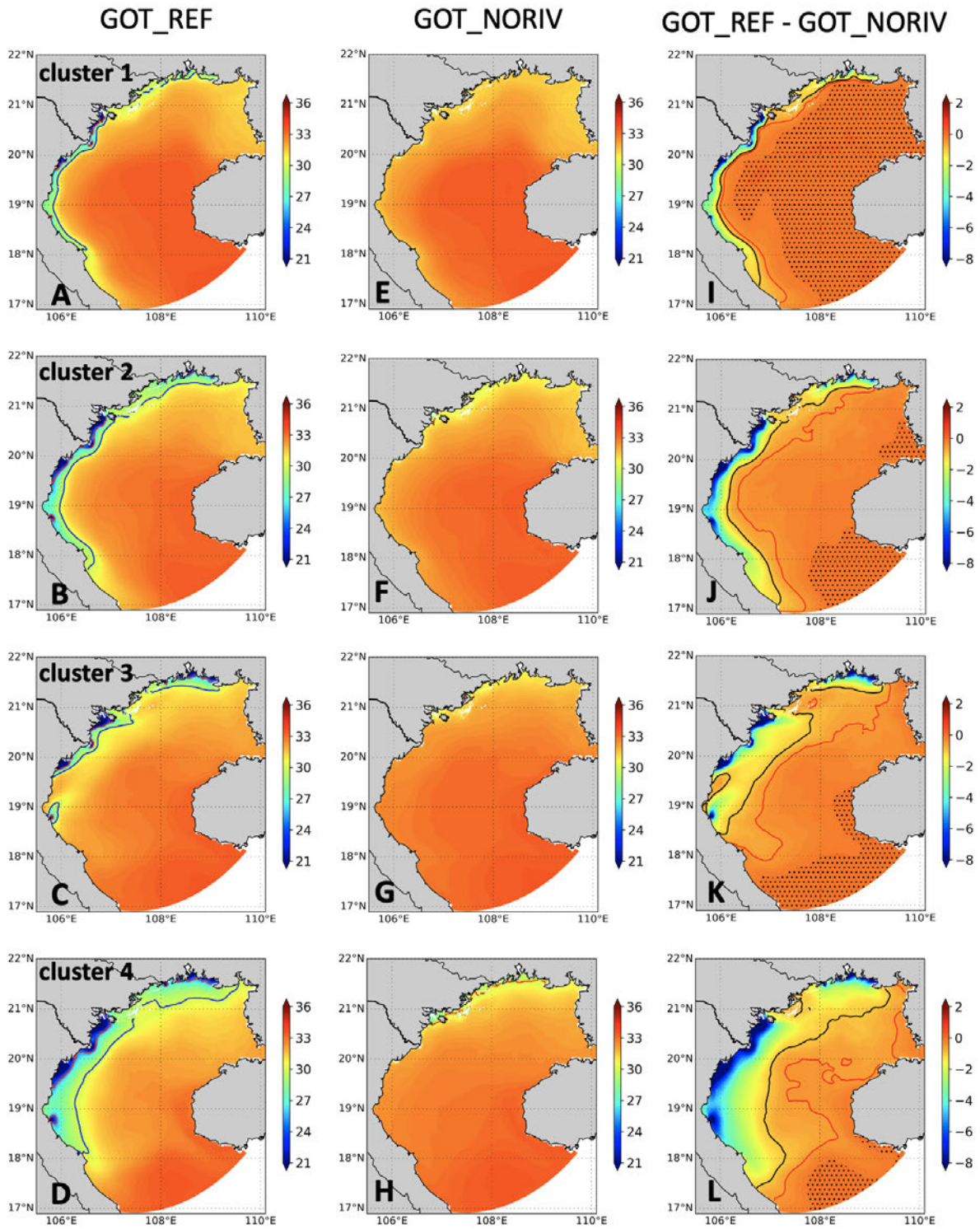

Figure S2: (A, B, C, D): Average salinity for GOT_REF for each cluster. Black contour indicates the 30 isohalines. (E, F, G, H): Same as (A, B, C, D) but for GOT_noriv. (I, J, K, L): Difference between GOT_REF and GOT_NORIV. Red and black contour lines indicate the difference in SSS of 0.5 and 1, respectively. Black dots show the area where the SSS difference is less than 0.1.

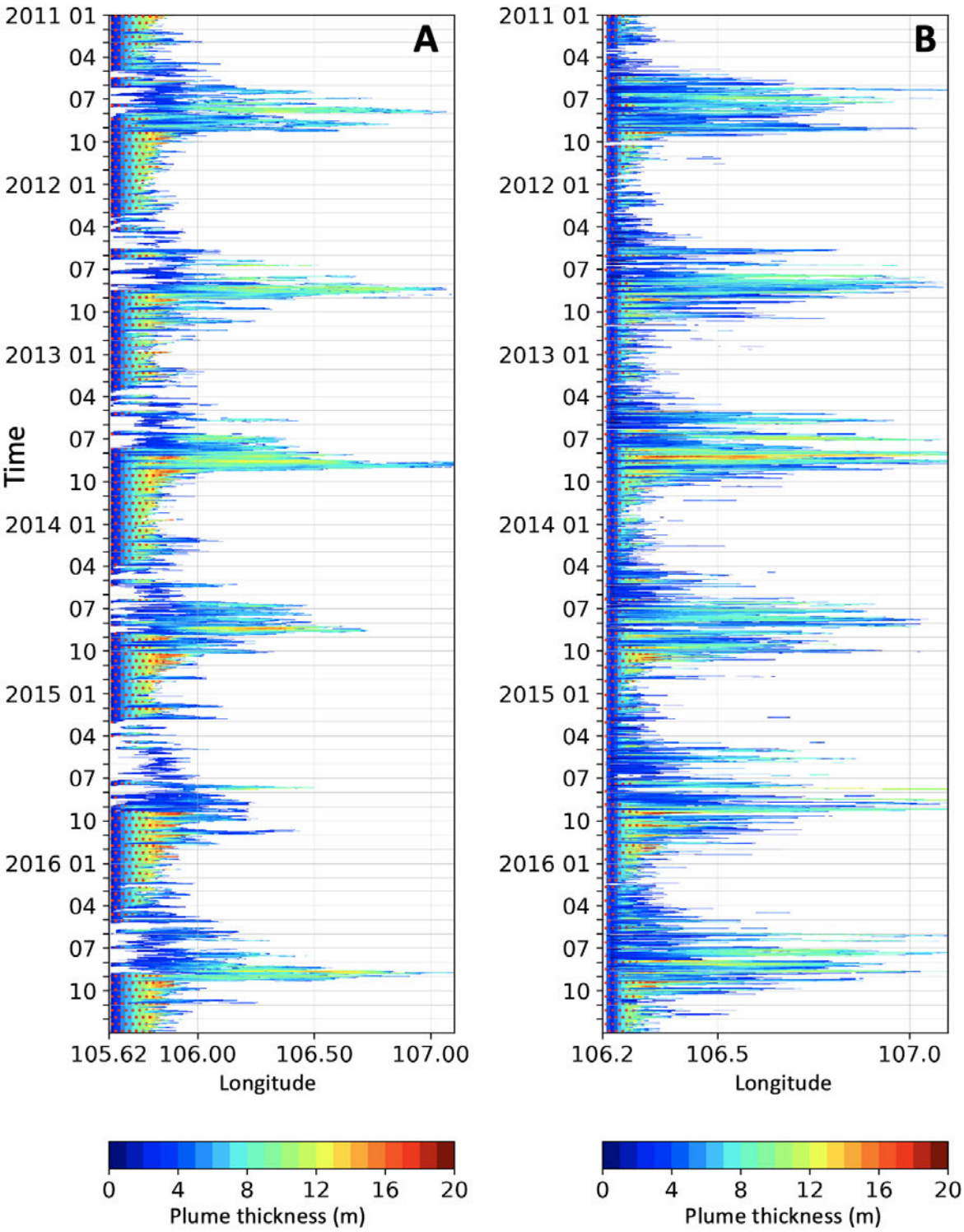

Figure S3: Plume thickness at 19 N **(A)** and at 20N **(B)** from 2011 – 2016. Red points indicate the area where the plume is present over the whole depth.

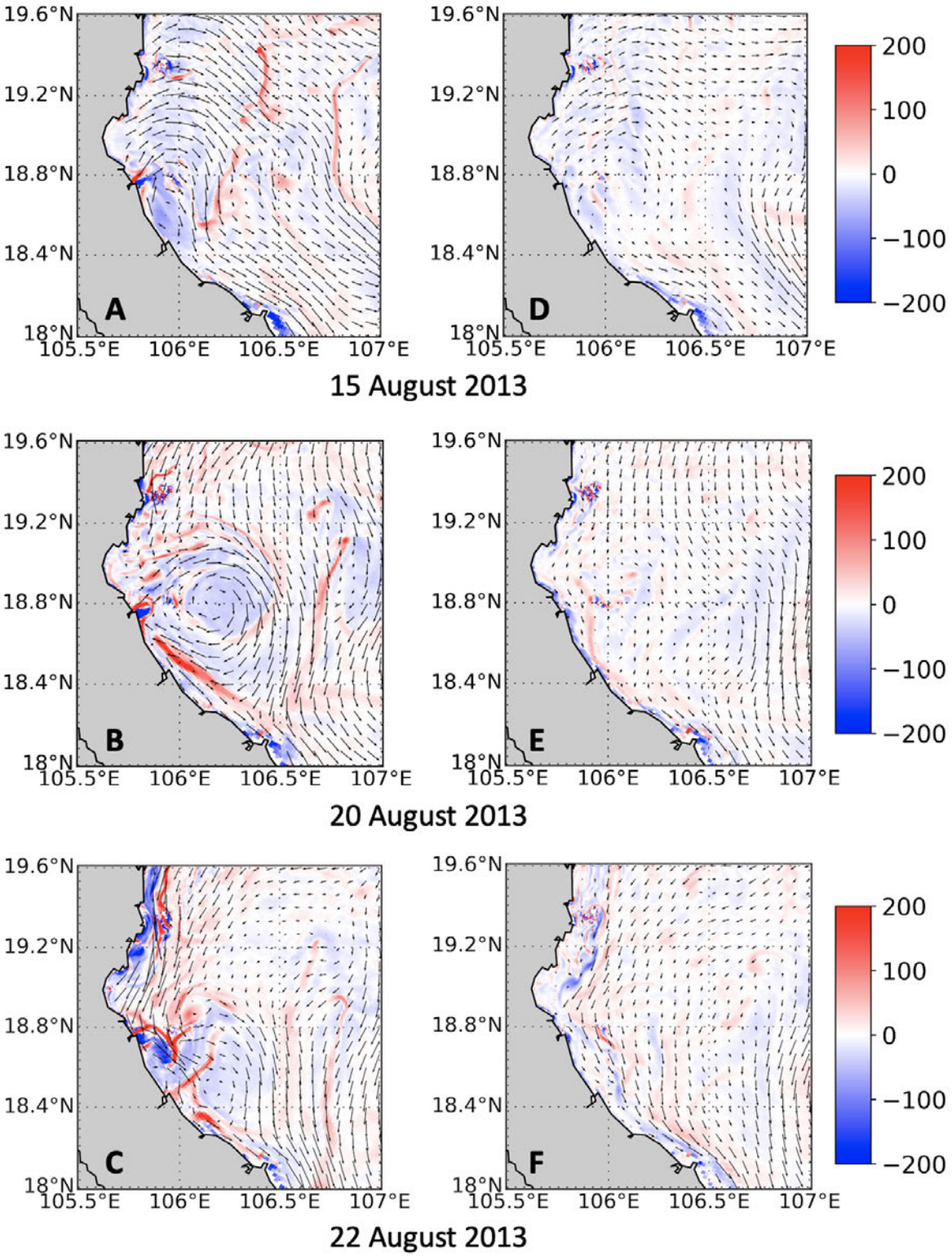

Figure S4: **(A, B, C)**: Relative surface vorticity ($10^{-6}s^{-1}$) (contour fill, right color bar) and surface current direction (arrow) in GOT_REF on 15, 20, and 22 August 2014, respectively. **(D, E, F)**: same as **(A, B, C)** but for GOT_noriv.

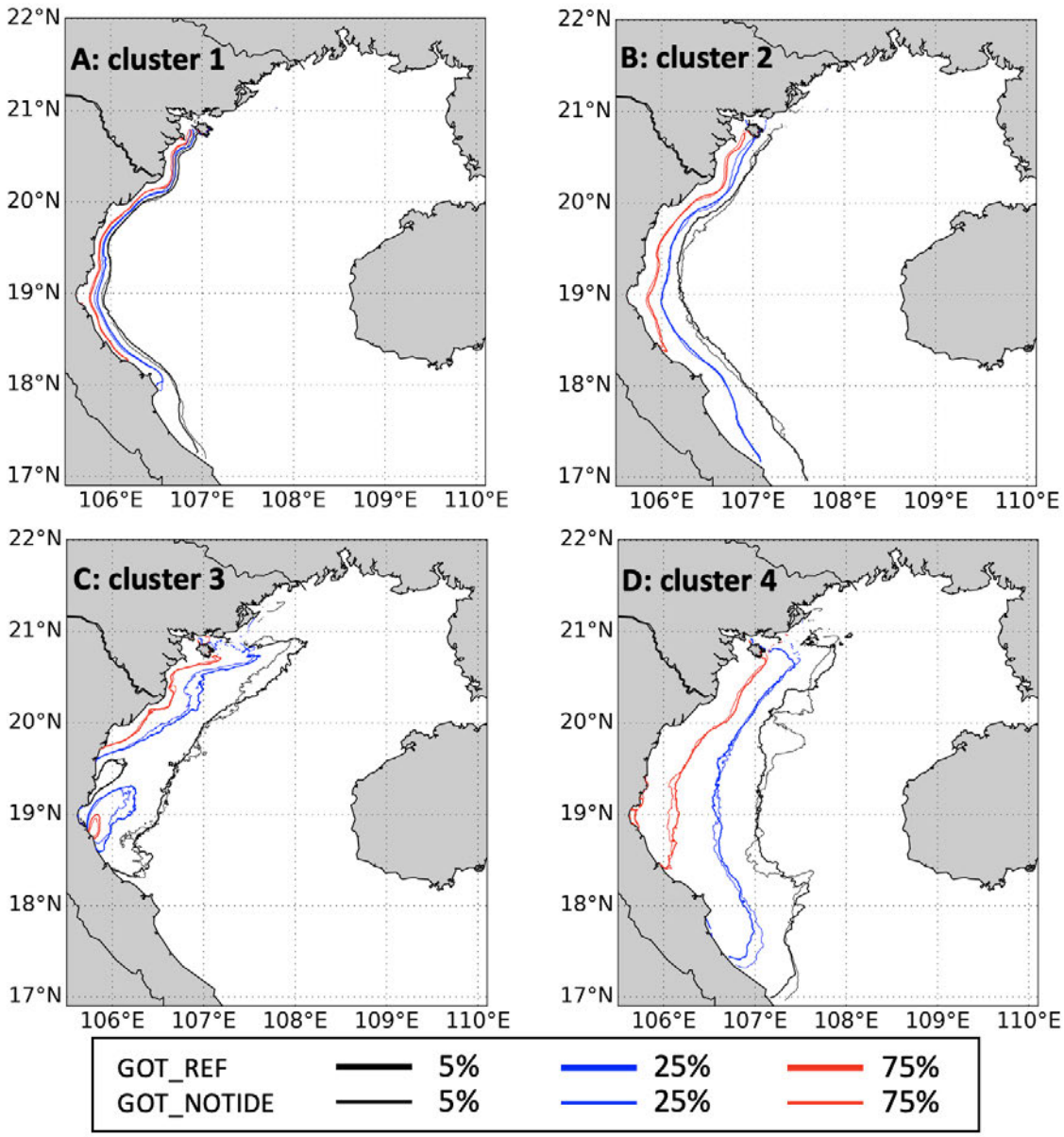

Figure S5: frequency of occurrence of the plume within cluster 1 **(A)**, cluster 2 **(B)**, cluster 3 **(C)**, cluster 4 **(D)** for GOT_REF (thick lines) and GOT_notide (thin lines) at 5%, 25% and 75%. For instance, for cluster 1: at a given point, a frequency of 25 means that over the whole period when cluster 1 is present, the plume is present at this point 25% of the time.

IV.3 Further discussions

IV.3.1 The sensitivity of the clusters due to initial state and the reproducibility

In K-means clustering analysis, there are 3 problems that need to be addressed to have a reliable result. The first question relates to the number of clusters. This is addressed in the Appendix 3 in section IV.2. Since the K-means initial conditions consist of random numbers, the second problem is: "are the results sensitive to the choice of initial random numbers?", and the third question is: "if the initial condition is random, are the results reproducible or not?" In this section, I will address these 2 problems.

To produce results in section IV.2, in the initialization process, the first centroid of the cluster is taken "randomly" from the daily scenes of the plume area. Then, the clustering algorithm is run for 10 times, and the best result is kept. To check the robustness of the result, the process above is repeated 100 times. The results are shown in **Figure IV.15**. Most of the time, the clustering result is stable. A change of cluster is observed in less than 1% of the study period (15 days over 2192 days). Also, the change only happens 2-3 times over 100 times of run, with the maximum change being 5 times. Therefore, the result of the clustering analysis, in this study, can be considered robust and reproducible.

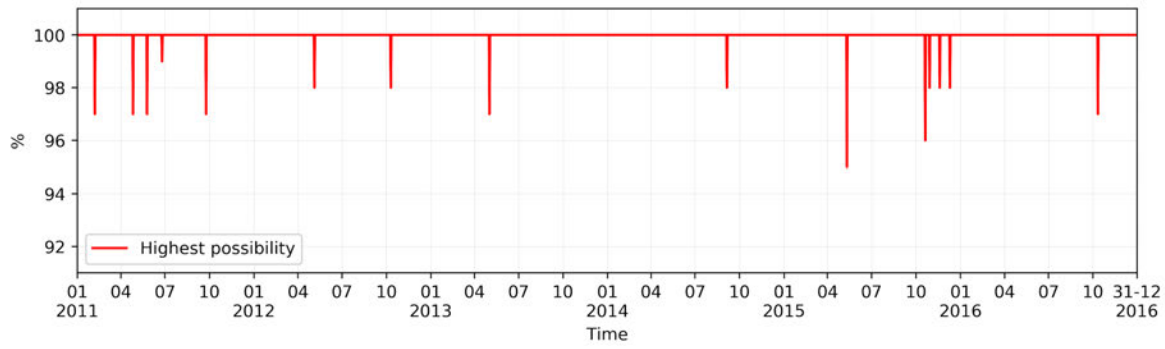

Figure IV.15 Highest frequency of occurrence (%). For example, 95% means that when the clustering analysis is run for 100 times (each time with 10 random conditions and the best result is kept), the plume pattern is attributed to the same cluster for 95 times, and to another cluster 5 times.

IV.3.2 The sensitivity of the tracer concentration near the boundary due to the sponge layer configuration

When looking at the surface tracer concentration on some days especially in the period of high runoff, it can be noticed that the tracer concentration shows a strong gradient near the boundary (**Figure 4.16A**). This strong gradient can be attributable to the sponge layer characteristics. Therefore, another twin simulation of GOT_REF, called GOT_S10, is created. The only difference between GOT_REF and GOT_S10 is that the sponge layer of GOT_REF contains 30 grid points, while GOT_S10 has 10 points.

Figure IV.16 shows the surface concentration of GOT_REF and GOT_S10 on 15 August 2011. In both cases, near the Southern boundary, the tracer concentration shows a sharp decrease to the South, but in GOT_S10, the area with low concentration is smaller. It means that the

thickness of the sponge layer may affect the size of the low-concentration area near the boundary. However, it is not the only reason that creates this low tracer concentration area. In fact, it is due to the imperfection of the boundary condition and the dynamics of this region. In the model, the inflow tracer concentration is set to 0 at the open boundary. Therefore, if a tracer flows out of the model, it never gets replenished. In the summer, there is a recirculation at the south of the domain, as described in the Section 3 (Model assessment) of the paper above. A part of this recirculation is located outside of the model domain. Therefore, on one side, it brings water from the boundary (with tracer concentration = 0) into the model, but on the other side, it takes the tracer out of the domain. Therefore, the area of this recirculation is always filled with the water with tracer concentration = 0, which creates a very low concentration area near the southern boundary. This problem does not happen in the Hainan strait boundary because in the high runoff period the water direction is outflow.

Another test with the boundary condition is implemented. This time, the boundary condition of the tracer is equal to the concentration of the last point of the domain, and the other setting is the same as GOT_REF. However, this time, since all of the inflow water has the concentration of that last point, it leads to the accumulation of the tracer concentration inside the model and the results are more unrealistic.

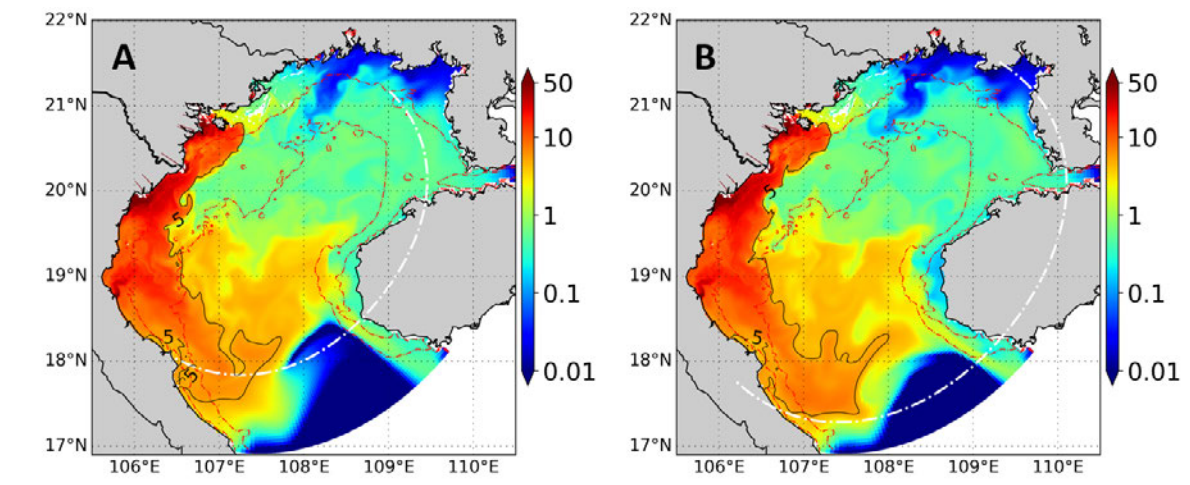

Figure IV.16 (A) Surface tracer concentration of GOT_REF (sponge layer = 30 grid points) on 15 August 2011. White dashed line shows the limit of the sponge layer. (B): Surface concentration at the same time but for a GOT_S10 (sponge layer = 10 points).

III.4 Limitations due to the open boundary conditions

In GOT_NORIV (the simulation without any river), the same open-boundary forcing (same OGCM) is applied as in GOT_REF. Therefore, if there is an inflow from the boundary, and this inflow has a lower salinity than the salinity in the model, the low salinity water from the OGCM (created by river runoff in OGCM configuration) can flow into GOT_NORIV and create the uncertainty in the salinity of the GOT_NORIV.

To check if it happens or not, we looked at some sections at the coast near the southern boundary and found that the current is mostly southward, except between July and September, when an inflow can occur (**Figure III.9**). In the periods of inflow, we checked the salinity at some points along the section and found that the salinity from the OGCM is sometimes smaller than in

GOT_REF. In these cases, some “unreal” freshwater can be indeed introduced into the domain so it can create uncertainty in the NORIV configuration. However, these differences are small (< 0.5 in all cases). This is the limitation for all the configurations without river if they use the same boundary condition as the reference configuration. To reduce this “unreal” freshwater, in the future, we can create another NORIV configuration with a larger domain, and use it as the boundary for the current NORIV configuration. Still, this can only reduce but cannot remove totally the unreal water flow into the GOT domain, since some unreal water can flow into the big NORIV.

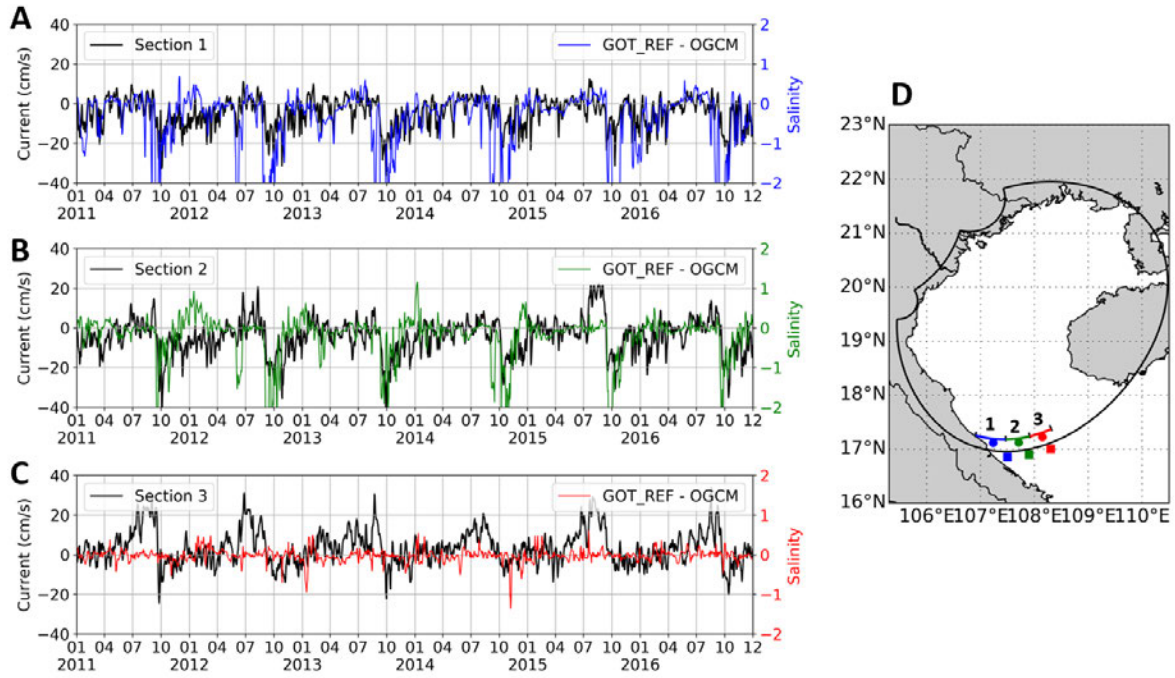

Figure IV.17 **(A)**: Black line: depth-averaged current of section 1 (positive means northward). Blue line: Difference of depth-averaged salinity between GOT_REF and OGCM at blue point. **(B)**: same as **(A)** but for section 2 and green point. **(C)**: same as **(A)** but for section (3) and red point. The locations of section 1, 2, 3 and blue, green, red points are shown in **(D)**. The salinity is taken at the circle for GOT_REF and at the square for OGCM. If the black line is positive and the colored line is negative, the unreal freshwater will be introduced to the model.

IV.4 Conclusion

In this chapter, we examine the variability of the Red River plume. The result has been published in the paper above. In summary, the river plume is first identified from a threshold of passive tracer concentration. This criteria is used to calculate the surface plume area. Then, we applied the K-means clustering method to daily scenes of the plume calculated above to identify the main plume spatial patterns. From sensitivity experiments, we have chosen to classify the plume patterns into 4 clusters, each linked with different forcing conditions. The vertical variability of the plume is examined as well, with different characteristics in different seasons. Finally, we tried to identify the "receiving basin" from different rivers.

The paper also contains maps of the first baroclinic Rossby radius of deformation for two seasons estimated from the model. To our knowledge this is the first regional estimate of this quantity at such a resolution in the GoT After the paper review, we did some further sensitivity experiments. Firstly, we tested the stability of the K-means clustering to the initial conditions. The result shows that our results in the paper can be considered as robust. Then, we checked the boundary conditions and found that the largeness of the low concentration area near the boundary in the summer is linked with the parameterization of the width of the sponge layer. Finally, we examined the current and salinity near the boundary condition of the simulation without rivers and found that some low salinity water can be introduced into the domain in some day summers.

While the model's performance is almost good (compared to the data as in Chapter 3 and in the paper), these discussions above show that it still has some limitations at the boundary condition. This limitation can then lead to some uncertainties of the tracer concentration near the boundary. Beside the boundary conditions, there are other sources of errors. As shown in the clustering analysis, the plume pattern is partly controlled by the wind condition. Therefore, the errors in wind forcing may limit the conclusions of the paper regarding the plume regimes as identified by the clusters. One of the objectives of the next Chapter will be to study the impact of the wind forcing errors to the variability of the river plume.

CHAPTER V: ENSEMBLE MODELING OF THE RED RIVER PLUME IN THE GOT

***

V.1 Introduction

First, let us summarize what has been done so far. Using numerical results combined with observational data, the variability of the RR plume has been examined by several authors. Gao et al. (2013) and Rogowski et al. (2019) attempted to explain the plume variability in the summer and winter. The variability of the RR and southern rivers plume is further examined by Nguyen-Duy et al. (2021) using a high resolution, multiple sources cross validated bathymetry, pluri-annual 3D model. As shown in Nguyen-Duy et al. (2021), the wind plays an important role in the variability of the plume. Different wind regimes can create different plume variabilities, both in terms of surface coverage and thickness of the plume. A cluster analysis was performed in order to identify the main pattern of the Red River plume over the period 2011-2016. The authors found that in all 4 general patterns, the plume is strongly affected by the wind.

Figure V.1 displays the tracer concentration, SSS, and wind conditions, corresponding to cluster 3 and cluster 4, but averaged from June to August 2015 instead of over the whole period (2011-2016) as presented in Chapter 4. Though averaged over a much shorter period with respect to the published article, the figure still shows the characteristic behavior of the plume under different wind conditions. Under strong upwelling favorable wind, the plumes from the different rivers are disconnected, and advected more to the north (Figure V.1A), while they form a single plume when the upwelling wind is less intense (Figure V.1D). Secondly, regarding the plume thickness, in winter, when the wind is downwelling favorable, the plume sticks to the coast and reaches the bottom. In summer, due to monsoon, the wind is upwelling favorable and even if the runoff is three times higher than in winter, the plume is thinner and advected offshore.

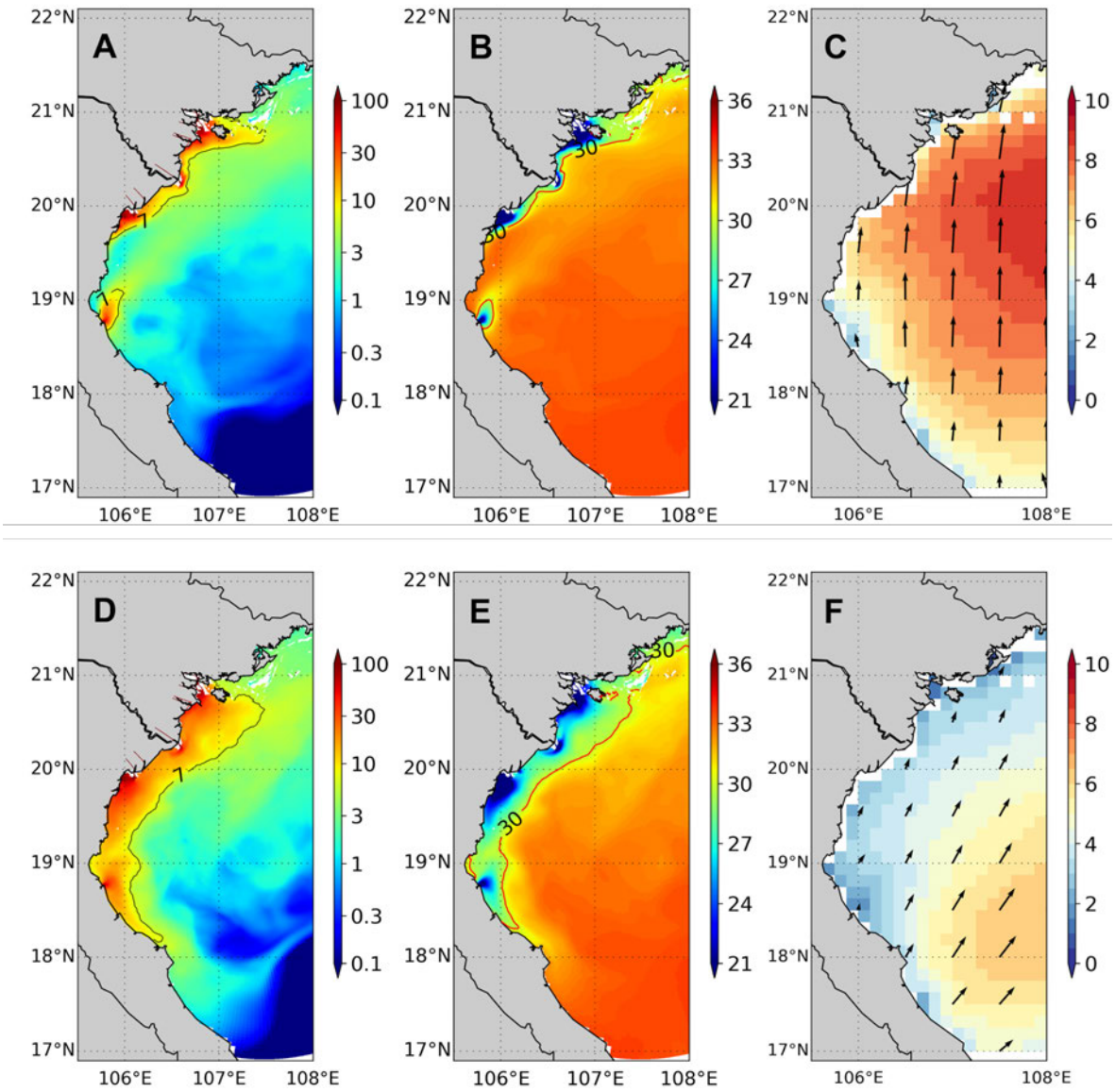

Figure V.1 **(A):** Surface tracer concentration (arbitrary unit/m^3), averaged from daily concentration, from June to August 2015 for the days when the plume is classified into cluster 3. **(B):** Same as **(A)** but for SSS. **(C):** Same as **(A)** but for 10m wind from ECMWF (in m/s). **(D, E, F)**: same as **(A, B, C)** but averaged for the days when the plume is classified into cluster 4. The contour lines in (A, D) show the area where the surface concentration is equal or higher than 7 unit/m^3, which is the threshold to be identified as the river plume.

The work of Nguyen-Duy et al. (2021), as other studies on the river plume in other areas, rely on numerical modeling. However, the model results are known within some uncertainties. The uncertainties can come from the model itself (physical assumptions, parameterizations, imperfect bulk formula, …), from the model configuration (prescribed coastline and bathymetry, parameters, …) or from the forcing (atmospheric forcing, lateral boundary forcing, …). Understanding the uncertainties and their sources can help us better assess the model results in different scenarios, and quantifying the uncertainties is a necessary first step toward data assimilation. One way to estimate model uncertainties is to use one or several ensembles of

simulations, to compute statistics from the ensembles, such as spread, quantiles, and to analyze the probabilistic distribution. Ensemble methods in oceanography were introduced about three decades ago for data assimilation purposes (e.g. Evensen 1994). Since then, they have been used as well for uncertainties estimates or budgets in open ocean (e.g. Andreu-Burillo et al., 2002; Lucas et al., 2008) or coastal ocean studies (e.g. Auclair et al., 2003; Kim et al., 2011), and generalized to coupled models such as physical-biogeochemical ones (e.g. Vervatis et al. 2021a). Probabilistic modeling is also implemented in numerical weather prediction systems, such as ECMWF, while ensembles of models are used to assess seasonal to interannual predictions (e.g. Kirtman et al., 2014, Herrmann et al. 2021).

As said before, the work of Nguyen-Duy et al. (2021) in Chapter IV has revealed the role of wind on the variability of the plume. However, until now, the robustness of these results has not yet been assessed. In particular, since the plume variability is strongly influenced by the wind, as evidenced by the cluster analysis, the impact of uncertainties in the surface forcing may be critical. More generally, the coastal and shelf sea dynamics respond mainly to the wind action, sometimes combined with the effect of tides and river runoff. For instance, by perturbing several sources of errors (atmospheric forcing, model parameterizations), Vervatis et al. (2021) suggest that in the Bay of Biscay, the main source of error is the wind forcing. This is also the conclusion of other studies, for example Jordà et al. (2010) in the Catalan sea or Auclair et al. (2002) for the Gulf of Lions. Therefore, this study aims at assessing the robustness of the Chapter IV results and conclusions with respect to an important source of uncertainties, namely the wind forcing. This in turn translates into the following three questions:

1. What is the space and time variability of the ensemble spread of the surface ocean variables ? In particular, does the model sensitivity to wind uncertainties depend on the runoff regime ?

2. Can we identify the physical processes at work to explain the sensitivity and its variability ?

3. Is the cluster analysis of Nguyen-Duy et al (2021) robust with respect to the simulated wind uncertainties ?

Besides, the ensemble approach raises several methodological questions, three of which we address in this study:

- How to generate wind perturbations that are indeed representative of ECMWF wind errors ?
- How (by means of which metrics) do we define the sensitivity to wind uncertainties with an ensemble?
- Is the generated ocean ensemble significantly representative of model errors ?

In order to do that, we will run an ensemble of simulations and assess the model response to perturbations added to the wind forcing. This category of methods has been used by several authors (Ayoub et al. 2015, Ghantous et al. 2020, Jordà et al. 2010, Quattrocchi et al. 2014, Vandenbulcke & Barth, 2015, among many others). Firstly, we will try to assess the uncertainty of the wind forcing. In some studies (Quattrocchi et al. 2014, Vervatis et al. 2016, Ghantous et al., 2020), the perturbations of the wind forcing are based on the author's assumption on the

wind uncertainty (amplitude, temporal and spatial structures). In contrast, Ayoub et al. (2015) estimated the error of the wind forcing by comparing it to a satellite product and then used the EOFs of the difference between the two products to create the perturbation. In this study, we will first attempt to estimate the amplitude of the uncertainty of the wind forcing by comparing the ECMWF product with a satellite product. In a second step, perturbations will be expressed and calculated as a pseudorandom linear combination of bivariate wind EOFs. Then, we will examine the response of the model to those uncertainties, at the surface and in the subsurface layer. We will revisit all the results that have been introduced in Nguyen-Duy et al. (2021), that is the surface plume area, the plume thickness and the stability of the plume pattern in terms of cluster classification.

We will also attempt to conduct an empirical assessment of the ensemble, that is to check if the estimated uncertainties from the ensemble are consistent with the model error deduced from model-observation misfits for the available observational data sets (as discussed for instance in Vervatis et al 2021a,b). To do so, we use rank histograms. This method allows us to assess the compatibility of the distribution of the ensemble with the distribution of the observational data.

The period of study is June to August 2015. It corresponds to the rainy season with high runoff and is a period when the wind goes through highly variable regimes.

This chapter's outline is as follows. In section V.2, the method to create the wind perturbations is described. The results from the ensemble simulation are presented and examined in section V.3. The summary and conclusion will be delivered in section V.4.

V.2 Method

V.2.1 Numerical model and general configuration

The SYMPHONIE model (Marsaleix et al., 2006, 2008) is used in the same configuration (called GOT_REF) as described in Chapter II. The simulation runs from a save point on 04 April 2015 of the run which started originally from 01 January 2010. The perturbation starts from 01 May 2015 on. We then allow for a 1-month spin-up time. All the analyses are carried out from 01 June 2015 to 31 Aug 2015.

V.2.2 Ensemble generation

The ensemble is generated by perturbing the ECMWF wind forcing. The perturbations are meant to represent uncertainties on the ECMWF wind analyses. Those uncertainties being unknown, we need to make some assumptions to build the perturbations. Several methods are found in past studies, as reported by Ayoub et al. (2015) for instance. Here, we assume that the space and time structure of the wind uncertainties are the same as the space and time structure of the wind variability as in Le Hénaff et al. (2009). This allows us to build the perturbations from the EOFs of the wind variability as described below.

Let's call W the 2D vector of wind velocity at 10m, composed of zonal U and meridional V components. We have:

$$W(x,y,t) = \bar{W}(x,y) + W'(x,y,t) \qquad \text{Eq. 5.1}$$

With $\underline{W}$ being the time-mean wind velocity and W' the variability of the wind.

We then form the Matrix of Samples (MoS) from the individual samples of W'. The wind EOFs are calculated by decomposing the MoS using SVD analysis. We end up with the EOF decomposition of the wind:

$$W'(x,y,t) = \sum_{k=1}^{n} \phi_k(x,y) * \theta_k(t) \qquad \text{Eq. 5.2}$$

With:

ϕ_k: bivariate (U,V) spatial component of EOF k (also called EOF k) – a nondimensional 2D vector field ("left" singular vectors)
θ_k: temporal component (also called "amplitude") of EOF k – a dimensional 1D time series ("right" singular vectors)

The fraction of variance explained by EOF k, called P_k, is calculated as:

$$P_k = \frac{(s_k)^2}{\sum_{k=1}^{n} (s_k)^2} \qquad \text{Eq. 5.3}$$

where s_k is the k-th singular value.

The wind is perturbed by multiplying each of the summing operands in Eq. 5.2 by a factor (1+σ), where σ is a pseudorandom number drawn from a Gaussian distribution, with a zero mean and a pre-defined standard deviation ε (assumed independent of the rank of the EOF). This choice is made as we assume that the variance of the uncertainties represent about ε^2*100% of the field variance.

The perturbed wind condition writes:

$$W_p(x,y,t) = W(x,y,t) + \sum_{k=1}^{n} \phi_k(x,y) * \theta_k(t) * \sigma_{\boldsymbol{k}}\ (\boldsymbol{t})$$	Eq. 5.4

Values of σ are pseudo-randomly drawn for each EOF rank and each member, and are updated every 5 days, which is equal to the value used in the Bay of Biscay in the previous studies (Quattrochi et al. 2014, Ghantous et al. 2020); in between, they are linearly interpolated. In this study, we also update the pseudorandom numbers every 5 days, as it reflects well the synoptic time-scale over the region (e.g. tropical cyclones, cold surges).

The perturbations are computed at the time step of the ECMWF fields, that is every 3 hours. They are then linearly interpolated at every model time step and spatially interpolated onto the model grid.

Another parameter of the ensemble simulation is the number of members. We ran our simulations with a total of 50 members, divided into 25 pairs. In each pair, the absolute value of the perturbations are the same, but the signs are opposite. Therefore, at any time, the ensemble average of perturbations is 0 (but the ocean response can have nonzero ensemble mean due to nonlinearities in the model).

To sum up, to create the wind perturbations, we will need the wind EOFs and pseudorandom numbers. In the next 2 sections, we will present the choices we will apply for both of them .

V.2.3 EOF wind analysis

Figure V.2 shows the spatial and temporal components of the first 5 bivariate (U,V) EOFs and the fraction of variance explained by the first 12 EOFs, calculated from 3-hour ECMWF wind from 01 May to 31 August 2015 over the GOT domain. The EOF decomposition is computed using SVD (see Eq. 5.2) as implemented in the NumPy library[2] in Python. From 19 - 26 June, there is a tropical cyclone in the gulf. It is an extreme event and not in the scope of this study, therefore we decided not to use winds in that period to calculate the EOFs. It means that in that period, we chose to set the perturbation to 0 (therefore, in the ensemble simulations, the model will run with unperturbed wind forcing).

[2] https://numpy.org

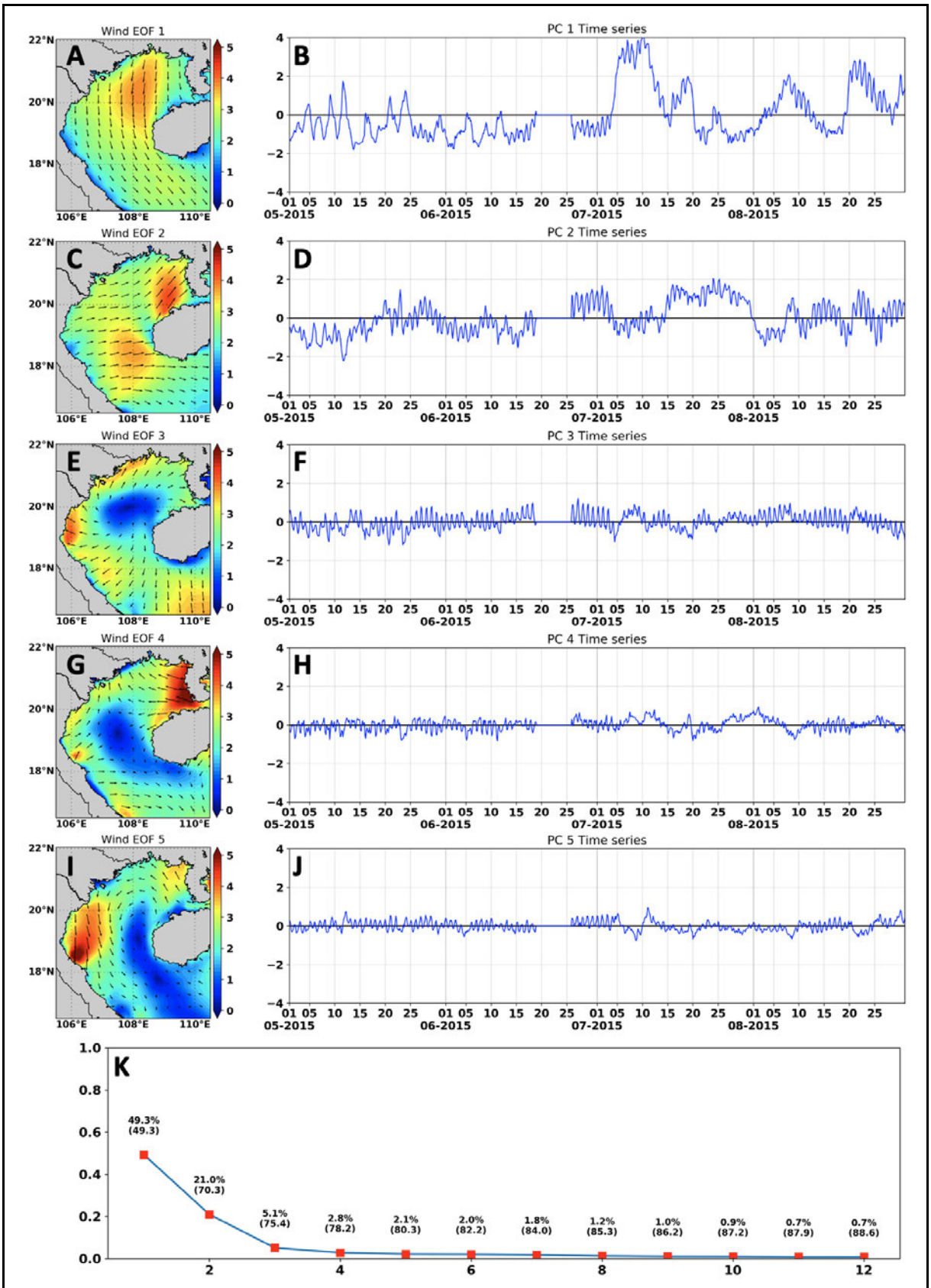

Figure V.2 **(A)** The spatial component of EOF 1. The colors and arrows present the surface wind velocity and direction, respectively. **(B)** The temporal component of EOF 1. **(C, D)**, **(E, F)**, **(G, H)**, **(I, J)**: same but for EOF 2, 3, 4, and 5, respectively. **(K)**: Fraction of variability explained by the first 12 EOFs. The upper number is the ratio of the EOF variance to the total variance, while the lower number in the bracket is the cumulative variance.

EOF 1 is the main contribution to the wind, explaining 49.3% of the total variance. Its spatial component shows a wind mostly in the N-S direction, following the shape of the gulf. The velocity is lower near the coast and reaches its highest values near the center - north of the gulf. The temporal components in May and June show negative values, which means that the wind direction is mostly southerly. From July, the wind direction reverses with positive values being more observed. EOF 2 accounts for 21% of the total variance, representing mostly a zonal wind pattern with the highest velocity located at the north of Hainan Island. EOF 3 explains 5.1% of the total variance, with weakest values at the center of the gulf (near 20°N), and highest at the coast.

The wind perturbations are built from Eq. (5.4) with a limited number of EOFs. Here, the first 5 EOFs are kept to reach at least 80% off the total variance. After EOF 5, the percentage of accumulated variance does not change much (from EOF 5 to EOF 12 it only increases from 80.3% to 88.6%). EOF 4 shows a very strong wind near Hainan Strait. We suspect that this pattern may not realistic, therefore, in this study, 4 EOFs are used (1, 2, 3, 5) with total variance ~ 77.5%.

V.2.4 Random error generation

Another challenge of the perturbation generation is the choice of the amplitude of random errors. Quattrocchi et al. (2014), Vervatis et al. (2016) and most recently, Ghantous et al. (2020), in their stochastic studies in the Bay of Biscay, create the ensemble with a Gaussian pseudorandom number generator with standard deviation $\varepsilon = 0.3$, which assumes that the standard deviation of uncertainty on the wind velocity is 30% of the standard deviation of the wind. However, this value is not necessarily general as it is based on the assumptions of the authors for a specific situation. Pasmans and Kurapov (2019) estimate ε from satellite and in situ observations of wind and from a Bayesian Hierarchical Model; in their case, ε differs according to the rank of the EOFs. They found that in their area of study, the value 30% is usually an underestimation, especially for high modes. In this study, we propose to estimate ε, using ASCAT satellite data as well but in a much simpler approach and without introducing modal dependency.

The ASCAT product distributed by EUMETSAT (EUMETSAT, OSI SAF, 2021) contains level 2 wind data, retrieved from the Advanced Scatterometer (ASCAT) on MetOp-B satellite, at 12.5 km sampling resolution (with an effective resolution of 25 km). It is available daily and can be downloaded freely from the JPL website: https://podaac.jpl.nasa.gov/dataset/ASCATB-L2-Coastal . The data are downloaded in the period from 01 May to 31 August 2015. The ASCAT User Manual indicates that 'the accuracy is better than 2 ms^{-1} in wind component standard deviation with a bias of less than 0.5 ms^{-1} in wind speed'.

Before doing the calculation, the data are processed through several steps. Firstly, since the data have different data tracks on different days, they are interpolated onto the ECMWF grid. Then, the data that have less than 50% spatial coverage are filtered out; over the 123 days of this period, only 72 days (~58.5%) are used for further calculation. Finally, all the missing values of the ASCAT are mirrored to ECMWF data values to make the two datasets homogenous.

Then, the spatially-varying daily mean bias between the ECMWF and ASCAT is calculated, and the temporal standard deviation (hereafter std) of that bias is shown in Figure V.3A (for U

component) and Figure V.3D (for V component). For both U and V, the std of the bias is higher near the Red River coast, reaching around 3ms^{-1} for U and 3.5 ms^{-1} for V (note that these values are larger than the accuracy of the ASCAT product). The temporal standard deviation of the ECMWF wind is shown in Figure V.3B (for U) and Figure V.3E (for V). As for the bias, the maximum variance of ECMWF V is higher than U. However, the maximum variability in U is shown at the south of the gulf (~3.7 ms^{-1}), while for V it is at the north of the gulf (~ 5 ms^{-1}). The ratio between the standard deviation of bias and standard deviation of ECMWF is shown in Figure V.3C and V.3F. For both U and V, this ratio is higher at the Red River coast and can reach a maximum local value higher than 1. On average, this ratio (a proxy for ε) is around 0.8 for U and 0.7 for V.

Another way (not shown) of calculating the standard deviation has been considered also. Instead of calculating the ratio between the two maps of std(bias) and std(ECMWF), then taking the spatial mean, we just calculate a single value for the std(bias) and std(ECMWF) over time and space and then take their ratio. This time, the ratio is 0.8, 0.6, and 0.6, with respect to U, V and using simultaneously U and V.

In conclusion, these comparisons show that it seems reasonable to choose the value of ε within the range [0.6-0.9]. However, (1) if the perturbation is too large, the perturbed wind can go too far from the reference dataset and may totally change the reference pattern, and (2) we did not take into account the ASCAT measurement error in calculating the ratios. As a compromise, we choose ε= 0.6.

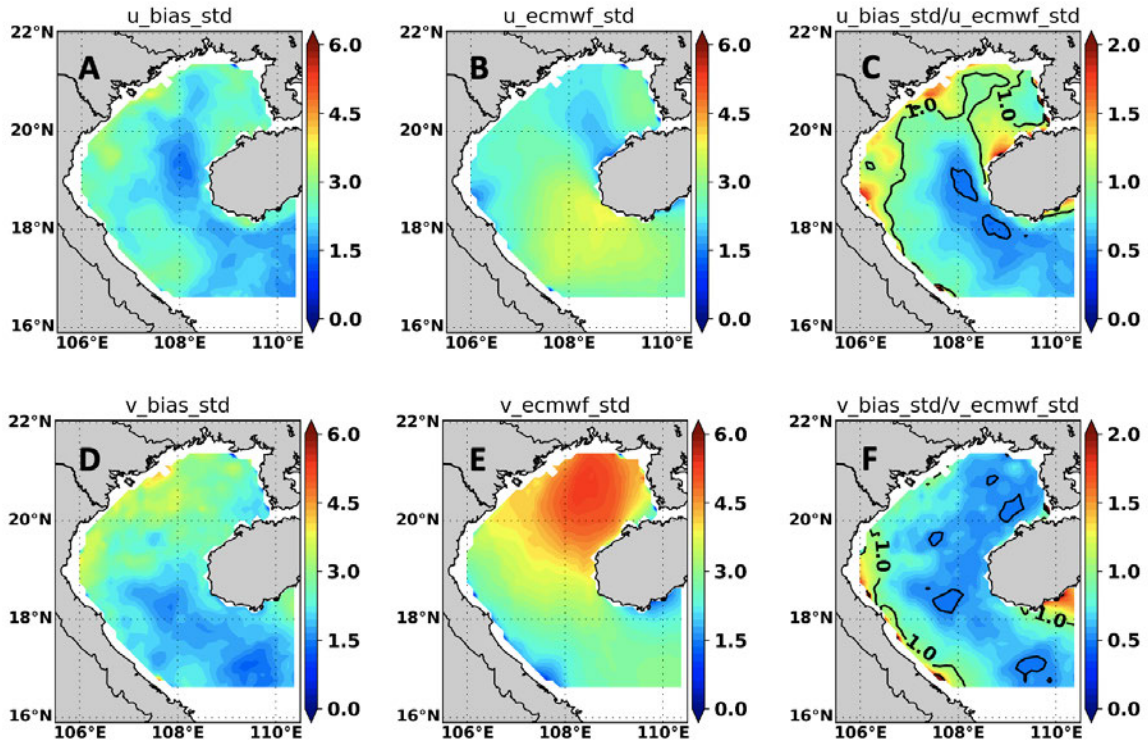

Figure V.3 (A): Standard deviation of the bias between ASCAT and ECMWF, for U component (ms-1). (B): Standard deviation of daily mean ECMWF wind for U component (ms-1). (C): ratio between (A) and (B). (D, E, F): same as (A, B, C) but for V component.

To further assess the appropriateness of this choice, we examine what we could have obtained with another method. In the absence of information from observations, we can use wind fields from another year or from a typical year. In that case, the perturbations do not mimic the wind uncertainties; instead, the sensitivity of the ocean solution is assessed with respect to other 'wind solutions' (in the space of the ECMWF model's solutions).

Here we compute the daily mean wind fields from 10 years of ECMWF wind field (hereafter ECMWF_MEAN). The bias is computed between the year 2015 and ECMWF_MEAN and then we do the similar calculation as we do with ASCAT, as if ECMWF_MEAN were observational data. In this case, the interannual variability of the wind is considered as the uncertainty of the wind. Such an approach is often adopted in climate sciences: to evaluate the skill of a prediction scheme, the prediction is often compared to the climatological value in order to evaluate if the forecast gives some added values compared to the simple climatological mean.

Figure V.4A shows the temporal standard deviation of the bias for U between the daily ECMWF in 2015 and ECMWF_MEAN. The standard deviation is lower near the RR coast (1.7 ms^{-1}) than at the south (3 ms^{-1}). The ratio between the standard deviation of the bias and standard deviation of the wind is shown in Figure V.4C. It is lower in the south and the center of the gulf (~0.9) and higher in the north (~ 1 to 1.05).

The standard deviation of the bias for V is higher than U, ranging from 3 ms^{-1} at the south of the gulf to 5 ms^{-1} at the north. The ratio between std(bias) and std (ECMWF) varies slightly, ranging from 0.9 to 1.0 in almost all areas.

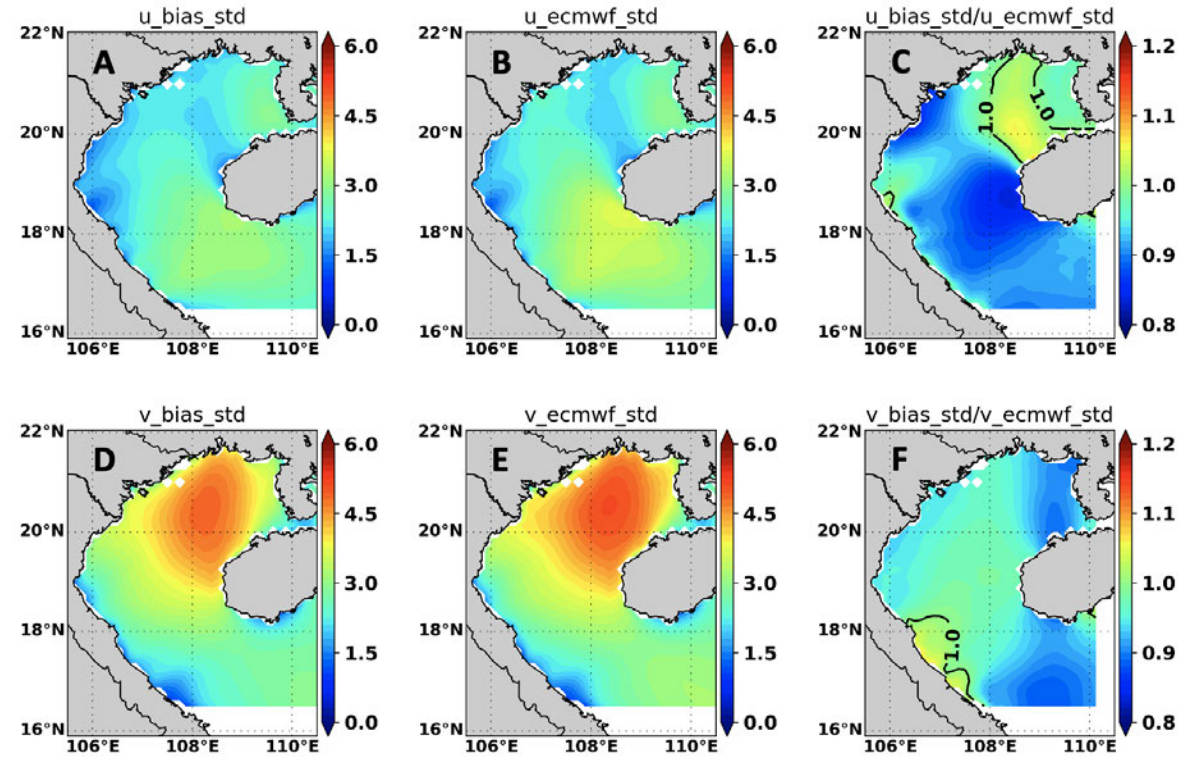

Figure V.4 (A): Standard deviation of the bias between the climatological ECMWF mean and ECMWF, for U component. (B): Standard deviation of daily mean ECMWF for U component. (C): ratio between (A) and (B). (D, E, F): same as (A, B, C) but for V component. Units for **(A)**, **(B)**, **(D)**, **(E)**: ms-1.

These calculations confirm that the choice of ε should be higher than 0.3, the value chosen in other studies. For the reason given earlier, we take ε =0.6. In the following we however present some results from an ensemble generated with ε =0.3 as a reference.

V.2.5 List of simulations

The simulation without wind perturbation is called GOT_REF. There are two ensembles. The first ensemble, called ENS_REF, which uses $\varepsilon = 0.6$. Another ensemble, called ENS_0.3, has the same configuration as ENS_REF, but with $\varepsilon = 0.3$. The wind perturbations are computed *online* during the SYMPHONIE simulation.

Table V.1: List of simulations

Ensemble simulation	Number of members	ε
GOT_REF	1	0
ENS_REF	50	0.6
ENS_0.3	50	0.3

V.3 Results

In this section, we will assess the impact of the wind perturbation to the ocean. The assessments are mostly based on the ensemble spread, which is computed as the standard deviation along the ensemble with respect to the ensemble mean.

V.3.1 Ensemble spread of the standard model variables

V.3.1.1 Surface variables

V.3.1.1.1 Wind stress

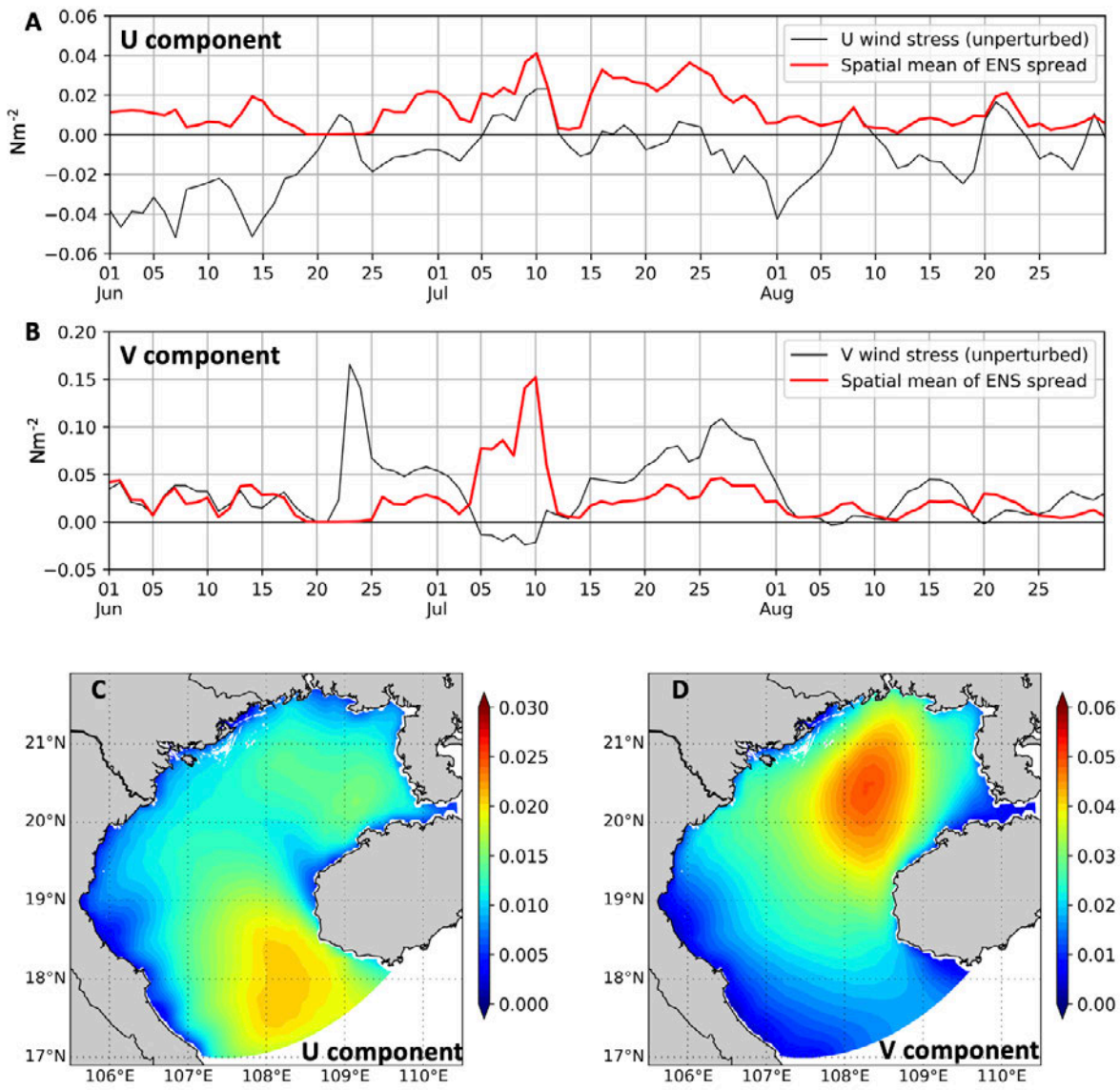

Figure V.5 (A): Spatial mean of unperturbed wind stress, for U component (black line) and of ensemble spread of wind stress (red line) from ENS_REF. **(B)**: Same as **(A)** but for V component. **(C)**: mean of ensemble spread of wind stress over the period, for U component (Nm^{-2}). **(D)**: same as **(C)** but for V component.

The spatial mean of unperturbed wind stress and ensemble spread of wind stress for both components from ENS_REF are shown in **Figure V.5 A, B**. Most of the time, for the U

component, the absolute value of unperturbed wind stress is less than 0.02 Nm^{-2}, except the first half of June and on 01 August. For the V component, the unperturbed wind has two significant peaks: the first peak is the storm mentioned before, from 23 - 25 June. In this period, the maximum wind stress can reach 0.17 Nm^{-2}. The second peak is observed from 20 July to 31 July. In this period, the maximum wind stress can reach around 0.11 Nm^{-2}, which is lower than during the storm, but this event lasts longer than the storm. For both components, the spatial mean of ensemble spread is less than 0.03 Nm^{-2} most of the time, except for the period of 05-11 July. The peak of the spread of wind stress in this period is consistent with the peak of EOF 1, suggesting that the spread is sensitive to the particular pattern of EOF 1. However, in another peak of EOF 1, such as from 21 - 25 August, the correspondence is weaker, suggesting more complex effects.

Figure V.5 C, D shows the mean of ensemble spread of wind stress over the period, for U and V components, respectively. While the spread of the U component is higher in the south of the gulf, while the spread of V component is higher in the north. This is consistent with the standard deviation of the U and V components of ECMWF (**Figure V.4 B, E**).

V.3.1.1.2 SSH

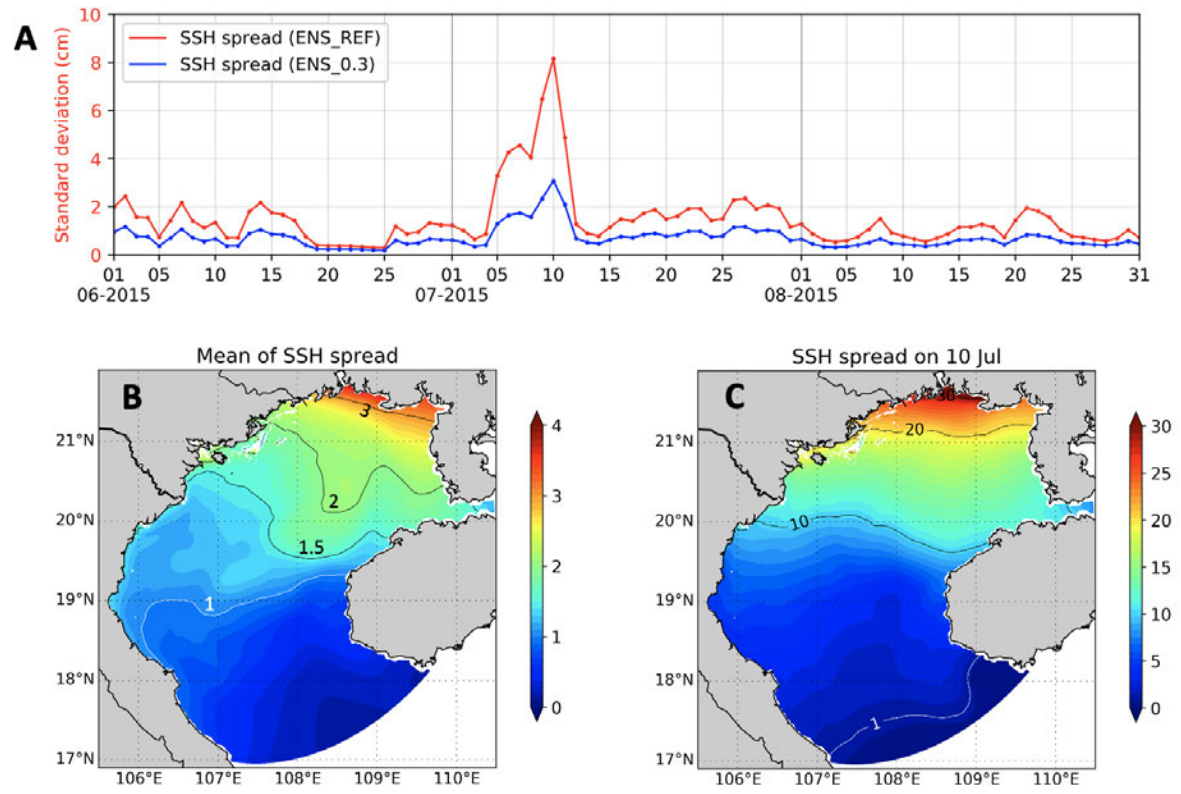

Figure V.6 (A): Temporal evolution of daily spatial mean of the ensemble spread of sea surface height (SSH) for ENS_REF and ENS_0.3. **(B)**: mean of daily SSH ensemble spread of ENS_REF over the whole period, excluding the period from 01 - 15 July. **(C)**: SSH spread of ENS_REF on 10 July 2015. Units are cm.

Figure V.6A shows the temporal evolution of the spatial mean of the ensemble spread of SSH for ENS_REF and ENS_0.3. Most of the time, the SSH spread < 2cm (for ENS_REF) and < 1 cm (for ENS_0.3). Overall, the mean spread of ENS_REF and ENS_0.3 are 1.48 and 0.71 cm,

respectively. It is found that the spread of SSH generally follows the trend of the wind stress spread, with a peak occurring on July 10.

The time-mean of the ensemble spread is shown in **Figure V.6B**. In order to avoid the average being contaminated by the much larger values reached from 01 to 15 July, this period is not taken into account into the average The spread of the SSH increases from less than 0.5 cm near the southern open boundary to over 3 cm in the North. Near the RR coast, the spread is from 1 - 1.4 cm. It is not clear yet if the slight increase of the SSH spread at the coast is due to the plume or to the direct wind effect onto the coastal sea level (wind surge).

The spatial pattern of ensemble spread on July 10 shows a north-south gradient as well (**Figure V.6C**) but stronger, with the smallest spread value at the south (around 1 cm) and largest at the north (30 cm). This coincides in time with the maximum of EOF 1 (wind perturbations mostly along the N-S direction, translating into possible water piling up/entrainment). Interestingly, it has been known from previous studies (eg Nguyen et al., 2014) that the diurnal tides are resonant which leads to their amplification in the northern GOT. We could assume that a large wind stress perturbation with a time scale of about 1 day (such as those on July 10) would generate a perturbation of the sea level that would be resonant and amplified with a similar structure as the diurnal tides (see for instance Fig.3 of Nguyen-Duy et al (2021) in Chapter 4). This assumption has not been verified yet.

V.3.1.1.3 Surface current

Figure V.7 shows the temporal evolution of the spatial mean of the ensemble spread of the surface current for ENS_REF. Most of the time, the spread of U and V components is similar, around ~ 4 cm s^{-1}, except for the period of 5 - 11 July (mostly the meridional currents), which is the same time when the spread of wind and SSH increases. Some other higher spread events are 08 August and 21 August, which correspond also to high SSH spread.

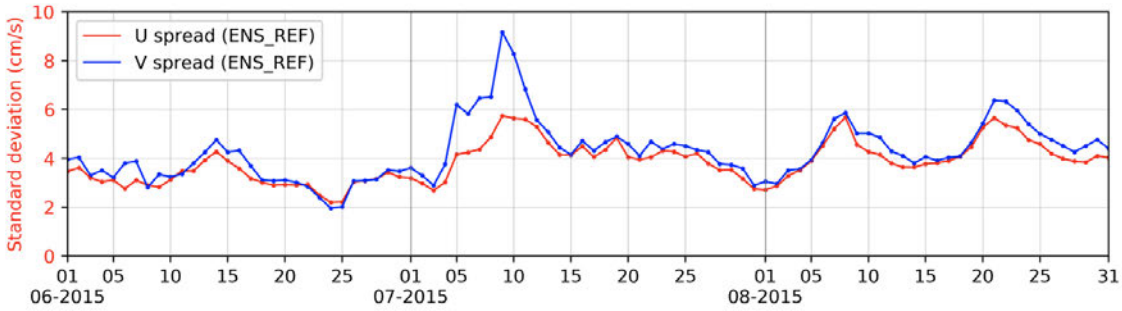

Figure V.7 Temporal evolution of daily spatial mean of the ensemble spread of surface current for ENS_REF (cm s^{-1}).

The ensemble spread for 09 July is shown in **Figure V.8 A, B**. On this day, the spread of U is highest near the RR coast (more than 20 cm s^{-1}), and the spread of V is highest near the west of Hainan Island (more than 30 cm s^{-1}). The high spread along the coast in V component suggests that the wind perturbation has an impact on the coastal current created by the river runoff.

Near the RR mouths, the maximum value of spread is more than 20 cm s^{-1} for both components (**Figure V.8 C, D**), which is even larger than the ensemble mean (5 cm s^{-1} for U and 10 cm s^{-1} for V component - **Figure V.8 E, F**).

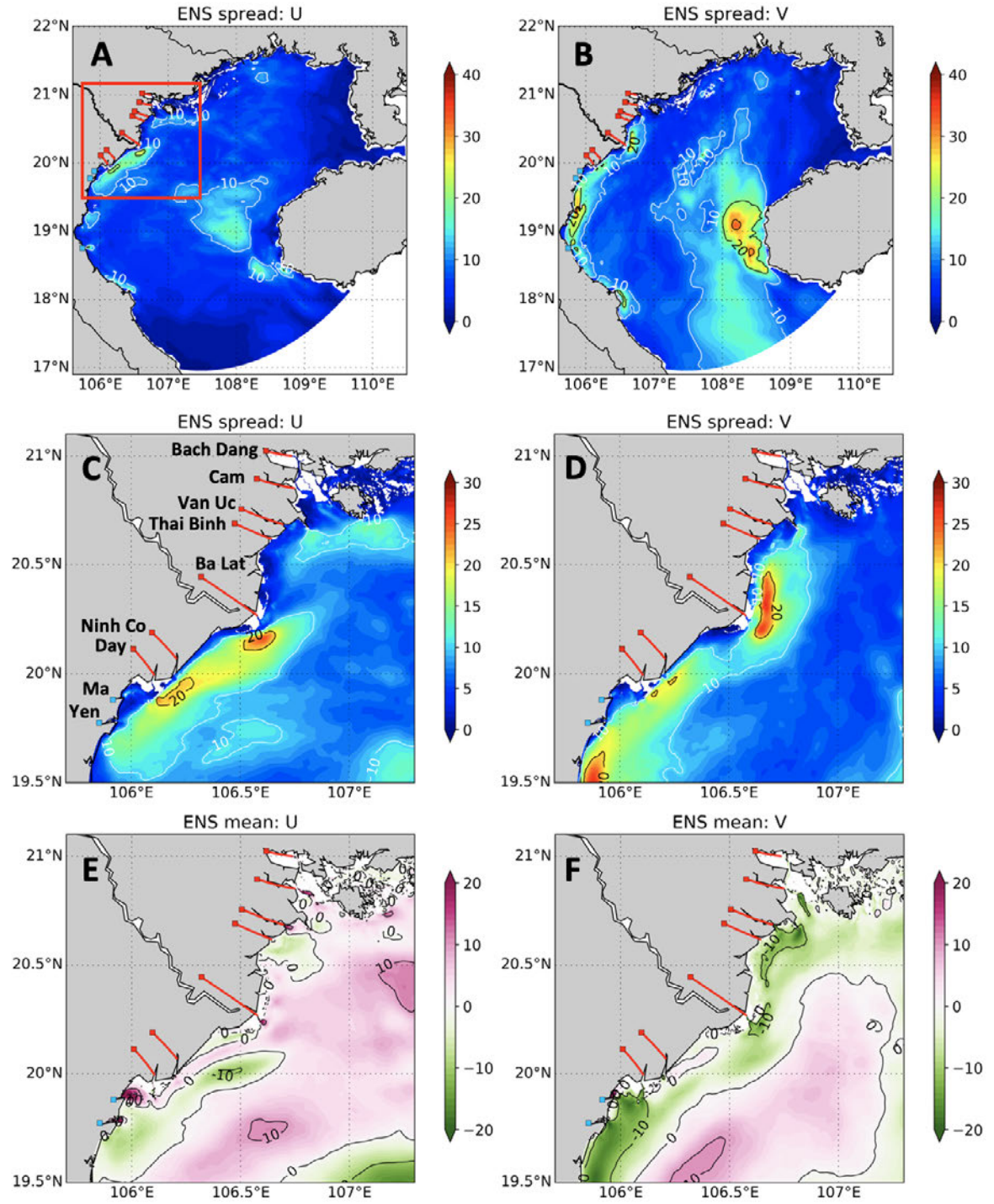

Figure V.8 Ensemble spread of surface current on 09 July 2015, for U component **(A)** and V component **(B)**. **(C, D)**: same as **(A, B)** but zoom on the red box. **(E, F)**: same as **(C, D)** but for ensemble mean. Units are cm s-1.

V.3.1.1.4 SST

The ensemble spread of SST for ENS_REF and ENS_0.3 is shown in **Figure V.9**. The mean spread for ENS_REF and ENS_0.3 is 0.33 and 0.19°C, respectively. As for SSH, but slightly lagged, there is a peak of SST spread on 11 July. The spatial pattern of the spread on this day is shown in **Figure V.9B**. The maximum spread is near the southern boundary, close to Hainan island, where it reaches up to ~2.5°C. In the unperturbed simulation, the temperature undergoes a strong mixing near the Hainan island due to the tide, which leads to a cooling at the surface in summer (**Figure V.9C**). When the wind changes here, it changes the horizontal advection and leads to the displacement of the low temperature water.

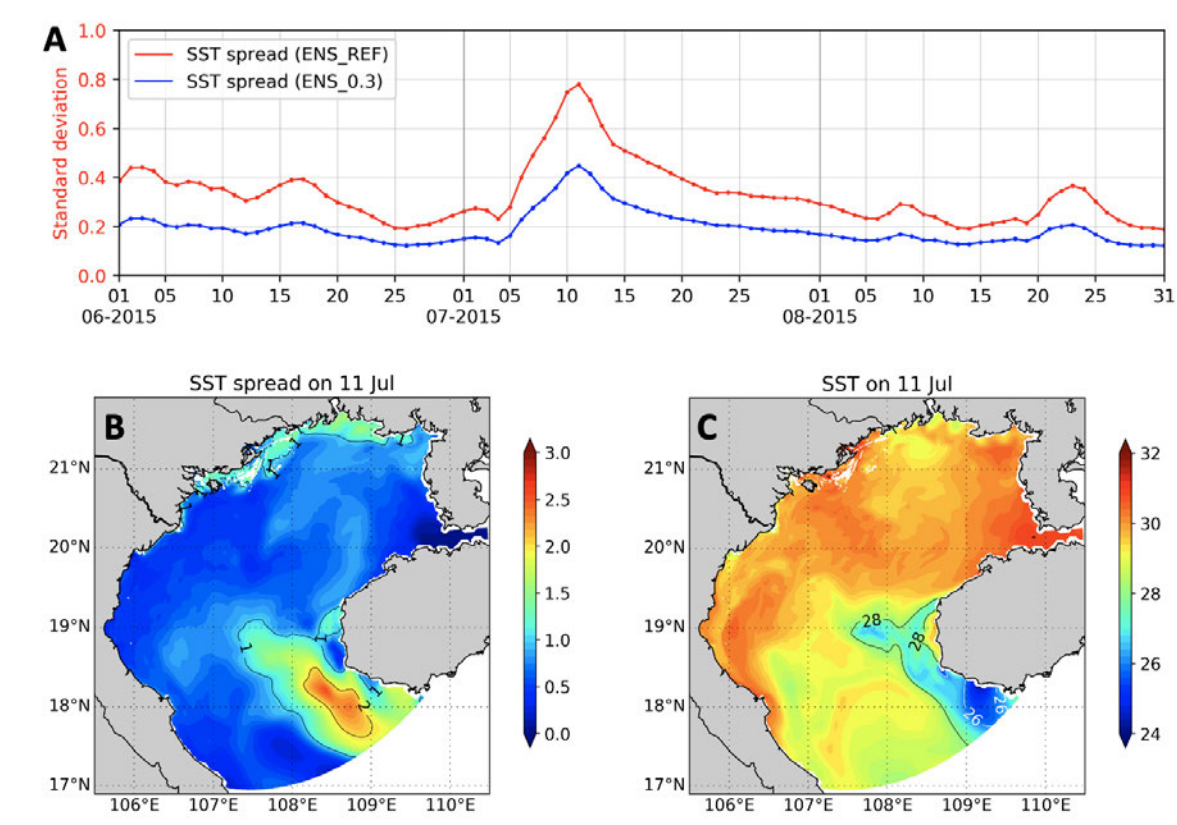

Figure V.9 **(A)**: Temporal evolution of daily spatial mean of the ensemble spread of sea surface temperature (SST) for ENS_REF and ENS_0.3. **(B)**: SST spread of ENS_REF on 11 July 2015. **(C)**: SST of GOT_REF on the same day as (B). Units are °C.

Near the RR mouths, the spread is around 0.9 - 1.2°C (**Figure V.10A**). The higher spread near the RR coast can be due to the difference between the temperature of the river and the ocean water, as suggested by the low temperature near the mouths of GOT_REF in **Figure V.10B**. In the model configuration, the highest temperature for the riverine water is 29°C, while as seen in **Figure 9B**, the surface temperature near the RR coast can reach more than 30°C. The measurements made by van Maren (2007) in 2000 show that the temperature of the fresh river plume was usually 1–3°C lower than ambient seawater.

Near Ma River, the river configuration is the same, however since the spread of the surface current is very small (**Figure 8C, D**) therefore the spread of the SST near this river mouth is also small.

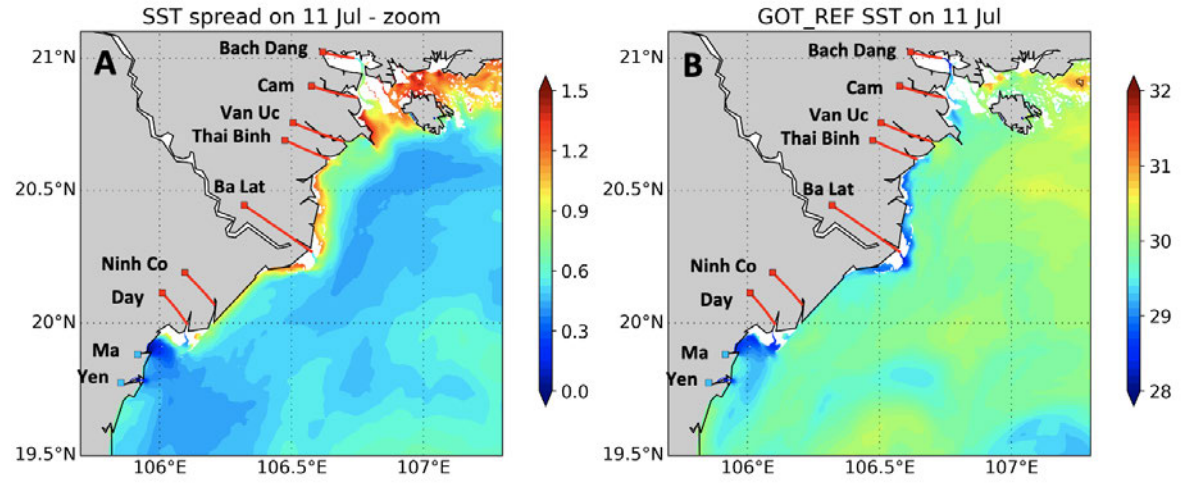

Figure V.10 (A): SST spread of ENS_REF on 11 July 2015, but zoom to RR coast. **(B)**: SST of GOT_REF on the same day. Units are °C.

Figure V.11A shows the temporal variation of the SST spread at 20°N. The high spread tends to occur where the horizontal gradient of the SST is high (**Figure V.11B**). This gradient can be created by the upwelling in June and early July, as suggested by Nguyen-Duy et al. (2021) and can be seen here in the signature of the temperature at the surface, 5 m and 10 m of GOT_REF (**Figure V.11 B, C, D**). High spread can occur also when the wind and surface current spreads are large (July 8 - 12). The largest spread occurs near the coast on 23 August, but there is no signature of the upwelling in the SST. Instead, it coincides with a local maximum of spread in SSS (not shown). As said above, it can be due to the difference between the temperature of the river and ocean water (**Figure V.11 D**).

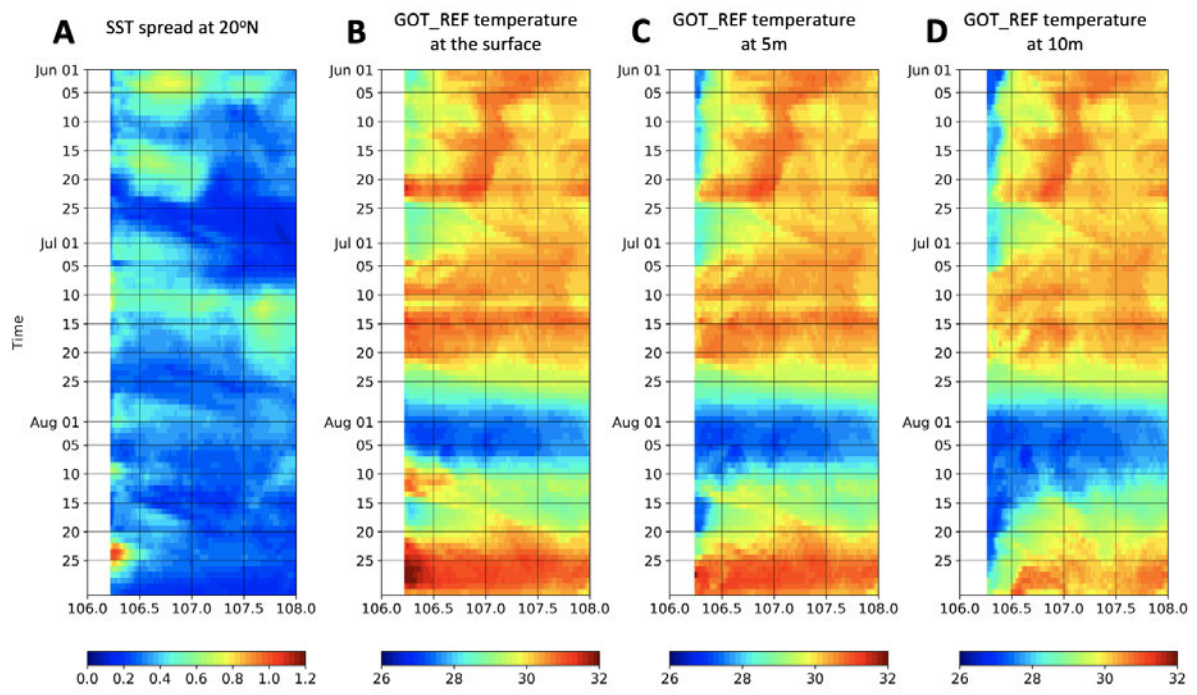

Figure V.11 (A): Hovmuler diagram of the SST spread at 20°N as a function of longitude and time. **(B, C, D)**: GOT_REF temperature at the surface, 5m, and 10m, respectively. Units are °C.

V.3.1.1.5 SSS

In ENS_REF, the spatial mean ensemble spread increases from June to August. In June, it ranges from 0.15 to 0.25, then increases to 0.38 in mid July before decreasing to 0.22 at the end of the month (**Figure V.12A**). In August, it reaches the peaks of 0.52 on 11th and 24th. The temporal mean of spread is 0.30. Over the whole period, the ensemble spread of ENS_0.3 is always 0.1 lower than ENS_REF with the temporal mean spread equal to 0.21. Since the ensemble spread of the wind stress does not show an increase in August, the increase of the SSS in August can be linked to the larger and more variable RR plume. This is confirmed when compared to the ensemble spread of the surface tracer concentration (**Figure V.12B**).

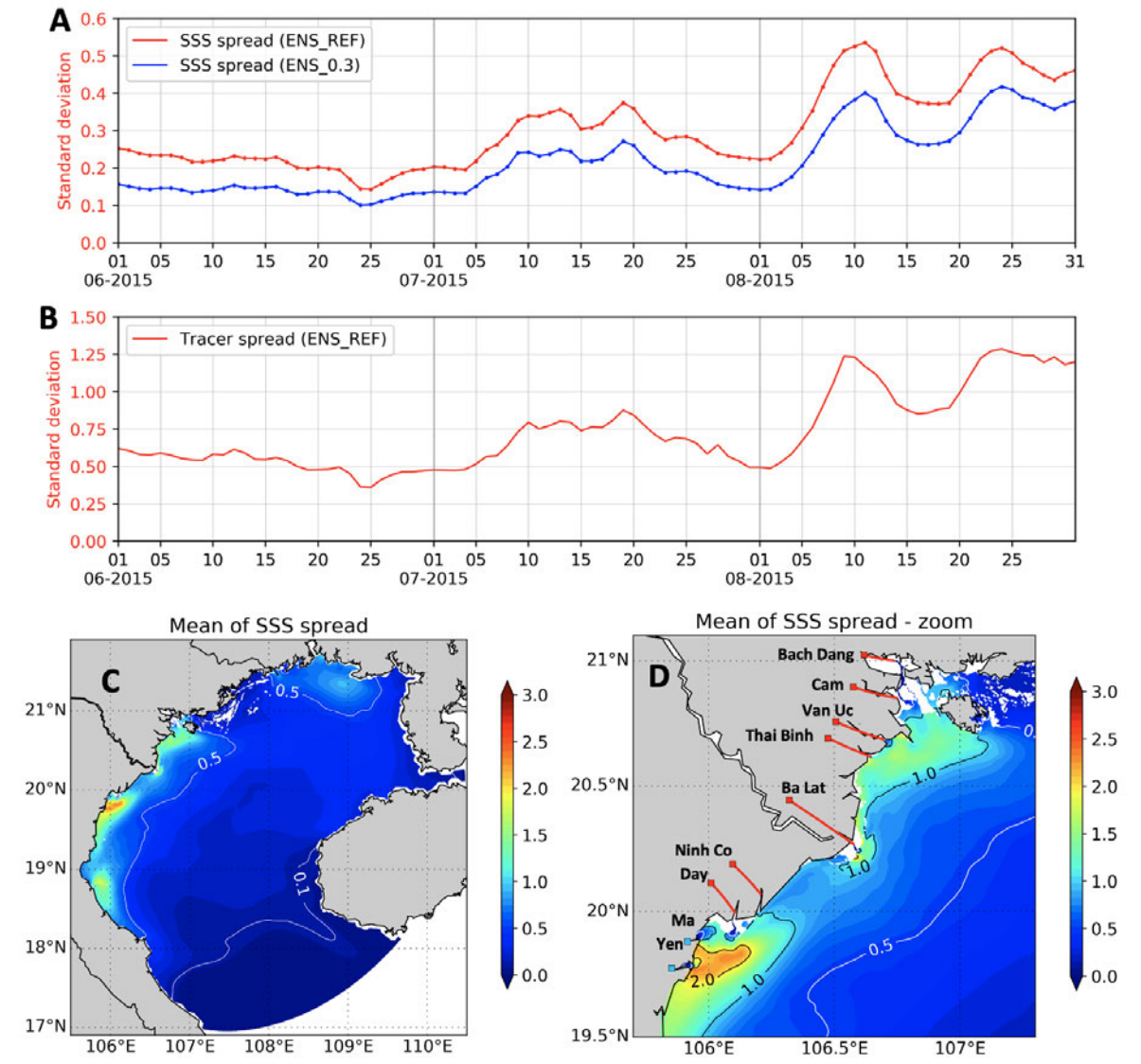

Figure V.12 (A): Temporal evolution of daily spatial mean of the ensemble spread of sea surface salinity (SSS) for ENS_REF and ENS_0.3. **(B)**: Same as **(A)** but for surface tracer concentration of ENS_REF. Time-mean of daily SSS spread of ENS_REF over the whole period **(C)** and zoom to the RR coast **(D)**.

Figure V.12C shows the spatial distribution of the ensemble spread on 11 August when the spread is the highest. It shows that the spread is less than 1 in almost all areas, except near the coast: where the RR plume is present, it can reach 4. The SSS spread over the whole period shows the same pattern, with a high spread near the RR coast (**Figure V.12C**). Near the southern boundary, the SSS spread is less than 0.1. Values higher or equal to 0.5 are only reached near the coast near the river mouths.

Figure V.12D shows the spread near the RR mouths, where the spread is the highest. At the river mouth, the spread is very low due to the homogeneity of the riverine water properties which are set by the runoff and the tides. Downstream, the spread increases and can reach 2 near the Yen river. The high spread near the rivers can be due to 2 reasons. Firstly, wind perturbation can change the wind-driven current at the surface, and then change the horizontal advection. Secondly, it can have an impact on the vertical mixing. In that case, the uncertainties on the plume should also be significant at the subsurface. In the next section, the subsurface spread will be examined.

V.3.1.2 Ensemble spread at the sub-surface

The vertical section at 19°N of the ensemble spread for temperature on 11 August is shown in **Figure V.13A**. The maximum spread reaches 1.5°C at around 30m depth. The same section of GOT_REF shows that the maximum spread occursat the maximum gradient depth (base of the mixed layer) (**Figure V.12B, C**) consistently with the findings of Andreu-Burillo et al. (2002) and Ayoub et al. (2015). The change of wind can have both impacts: at the surface and at the sub surface. However, since the horizontal gradient of the temperature is small, even if there are some horizontal displacements, they do not have any signature on the change of the temperature.

For salinity, the maximum spread occurs near the surface (**Figure V.13D**), where the salinity is low due to the runoff. Similar to the temperature, the maximum spread should occur where the gradient is highest. For salinity, the horizontal gradient is high where the river plume is present, therefore the spread is highest at the surface. The vertical gradient of salinity is small except for the area strongly influenced by the runoff. Here, the condition is reversed to the temperature. Even if the wind perturbation changes the vertical mixing, the displacement of the water mass with similar salinity does not create any signature on the salinity ensemble spread.

The vertical distribution of the spread for the salinity and temperature shows that the perturbations of the wind can have an impact both at the surface and subsurface. Depending on the variables, the spread may be higher at the surface or at the subsurface.

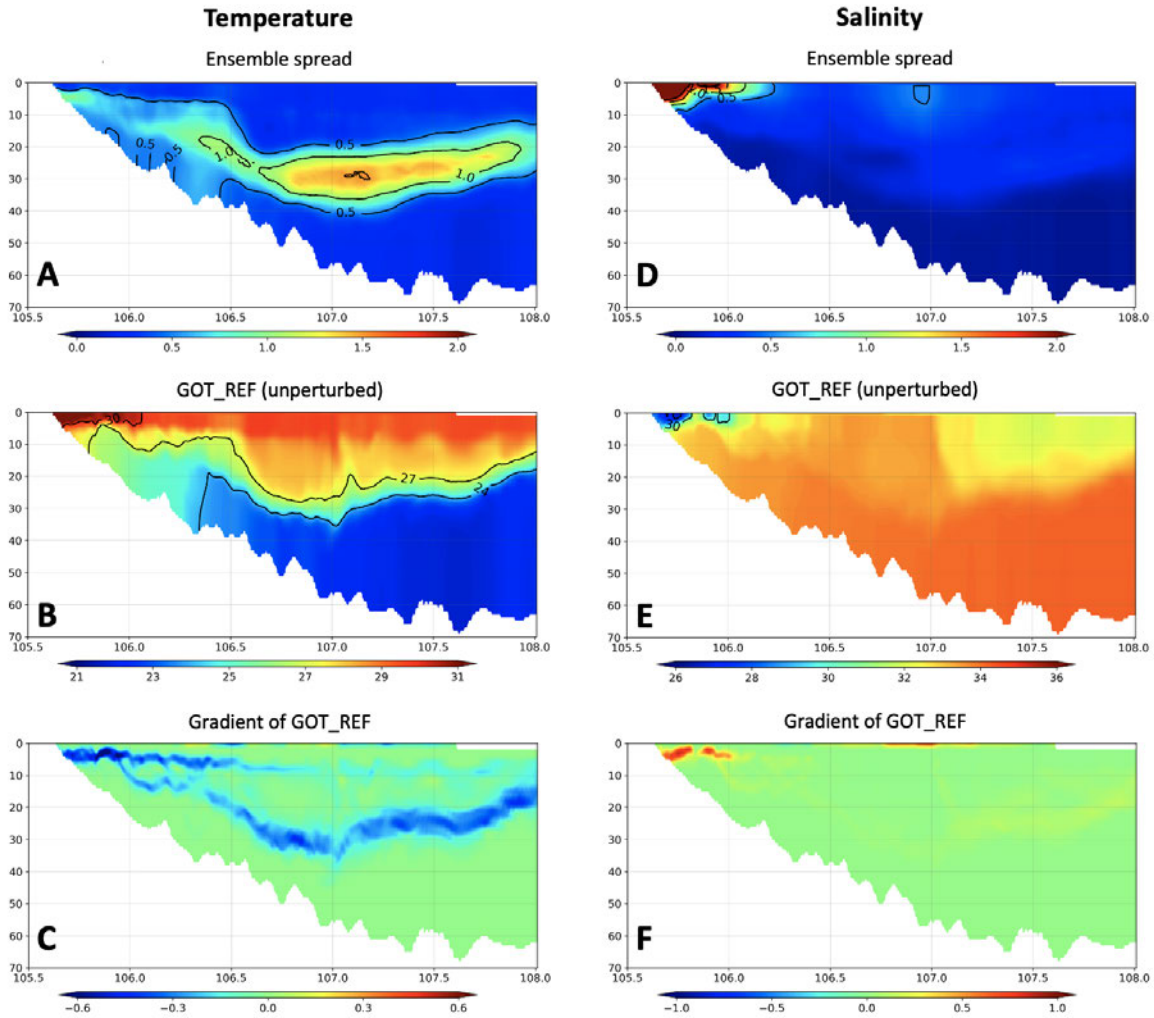

Figure V.13 (A): Zonal ensemble spread of temperature at 19°N on 11 August 2015 (°C). **(B)**: Vertical section on the same day off GOT_REF (°C). **(C)**: vertical gradient of **(B)** (°C/m). **(D, E, F)**: same as **(A, B, C)** but for salinity.

V.3.2 Ensemble spread of the plume

V.3.2.1 Plume area

Figure V.14A shows the spread of the plume area and the ratio between the spread and the plume area in GOT_REF, from the beginning of the ensemble simulation (01 May 2015). The runoff of the RR and southern rivers, as well as the plume area in GOT_REF, is shown in **Figure V.14B**. On average, the spread of the plume area is around 1452 km^2, which is equal to 18.8% of the plume area in GOT_REF. The peak of plume spread is from 20 - 20 August at more than 3600 km^2, but the percentage of the spread reached the maximum value on 02 June (more than 50%). The relationship between the spread and the GOT_REF plume area is not clear, except in the end of August where both the spread and the GOT_REF plume area reach their peaks. Interestingly, the members have a different behavior from August 13-15: in some members the plume area continues increasing, while in others, it decreases more or less (as in GOT_REF) and, in some cases, stays below their August 15 value until the end of the run. Similarly to the spread of the salinity and temperature, in general, the high spread of the plume can be due to the change of the vertical mixing and surface current.

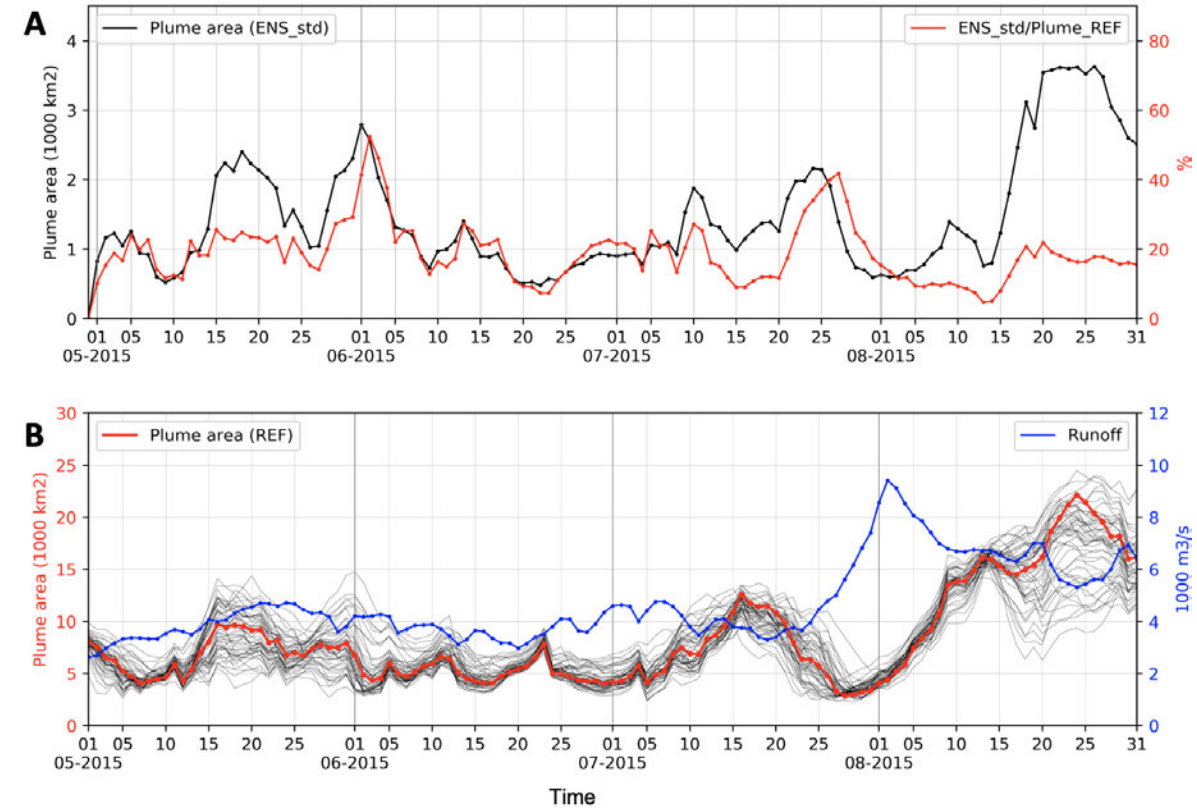

Figure V.14 (A): Ensemble spread of the plume area of ENS_REF in km2 (black line, left axis). Ratio between the ensemble spread and the plume area from GOT_REF (red line, right axis). **(B)**: plume area (in km^2) from GOT_REF (red line, left axis). Plume area from all the members of the ensemble (black line, left axis). Daily runoff from the Red River and southern rivers (blue line, right axis in m^3/s).

V.3.2.2 Clustering analysis

As in Chapter IV, to assess the uncertainty of the plume pattern, we classify the daily plume pattern of each member of the ensemble into 4 clusters, based on the *reference pattern classification* in Nguyen-Duy et al. (2021). In other words, we use the results from 6-year cluster analysis as the training data for this study.

To do that, the daily plume pattern of each member is identified by the threshold as described in Chapter IV (>=7 unit/m^3). Then, the distances between this daily pattern and the centroids of the 4 reference clusters described in Nguyen-Duy et al. (2021) are computed. Finally, the daily pattern for each member of the ensemble is attributed to the cluster which has the minimum distance.

We calculated a simple Maximum Likelihood estimate by choosing, for each date, the cluster that occurred the most frequently across the ensemble (denoted as MAX estimate). **Figure V.15** shows that the MAX clusters are equal to the reference clusters (REF on the figure), except on 21 July when the most likely cluster changes from 4 to 3. As a consequence, we estimate that within our experimental conditions the findings of Nguyen-Duy et al. (2021) are yet unchallenged and robust with respect to wind perturbations.

In GOT_REF, in this summer period, cluster 3 occurs the most (54/92 days) and also is the most stable cluster, in the following sense: across the ensemble, we do not observe many changes from cluster 3 to other clusters; if it changes, it mostly changes to cluster 1 (5 July) or cluster 2 (5 August).

Cluster 1 happens in just 2 days (06 and 07 July), and fluctuates across the ensemble between clusters 1 and 2. In Nguyen-Duy et al. (2021), cluster 1 is described as the pattern that occurs in the lowest runoff period, so it is not expected to occur much in the rainy season in July. Its fluctuation between cluster 1 and 2 when submitted to the wind perturbations suggests that the plume pattern on 06 and 07 July in GOT_REF is on the edge of cluster 1 and cluster 2 and just closer to cluster 1. In that case, the attribution to cluster 1 may not be as robust as the one to cluster 3.

The plume pattern appears to be the least robust to the wind perturbations in 3 periods: 10-22 July, 5-12 August, and 21-28 August. It is the period where we can see at least 3 different clusters occurring in different members on the same day. The high variability of the plume patterns in those periods is consistent with the large spread of the sea surface salinity, suggesting that the runoff could be another major source of uncertainty along with the wind. We will come back to this in the conclusions.

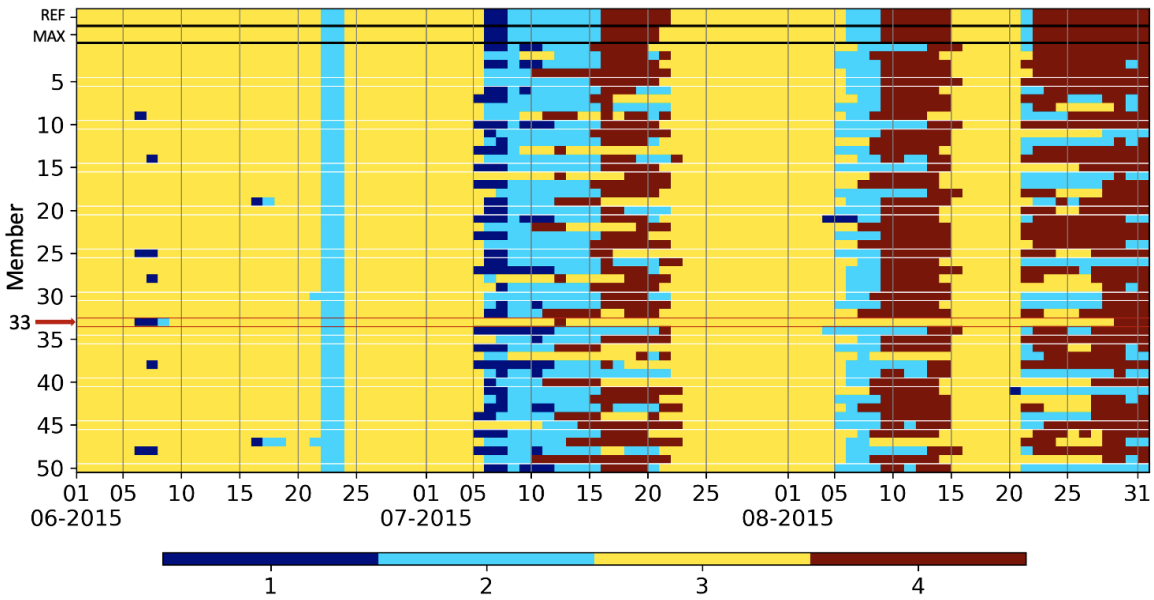

Figure V.15 Classification of the daily plume patterns in June-July 2015 across the ENS_REF ensemble, using the reference clusters from Nguyen-Duy et al. (2021) and Chapter IV. **REF** shows the same classification for the GOT_REF run. **MAX** is the cluster number with the MAXimum likelihood, defined as the most frequent cluster draw within the 50 samples shown underneath.

Though **Figure V.15** shows that most of the time, amongst 50 members, the most frequent cluster classification is the reference cluster classification, it also shows that sometimes the wind perturbation can significantly change the regime of the plume. For example, in member 33, the plume is classified as cluster 3 most of the time, though the reference is cluster 1 (e.g. 6-7 July), cluster 2 (e.g. 8 - 15 July), cluster 4 (e.g. 16 - 21 July).

To confirm that the wind perturbation can indeed change the plume regime, and clarify what cluster 3 is able to capture in member 33, here we analyze two short periods.

From 5 - 10 July, member 33 is labeled as cluster 3. The daily mean surface concentration shows that the plume is advected northward, and the plume from Ma river is disjoint from the plume from other rivers, which is indeed the typical characteristic of reference cluster 3 (**Figure V.16**). North of 20N, the wind is stronger and favors a northward advection, which explains the offshore transport of the plume. South of 20N, the wind is weaker and the direction changes to northerly, however, the current is northeastward so the plume from Ma river is advected offshore. In member 34, which belongs to cluster 1, the plume is thin and attached to the coast, which is also the typical shape of cluster 1. The strong northeasterly and westerly winds in the north and south of the GOT respectively trigger the southward coastal current, which explains why the tracer is trapped near the coast.

The similar situation happens again from 15 - 20 July (**Figure V.17**). In member 33, the surface plume is advected northward, consistent with wind forcing and Ekman transport, while in member 34, the coastal wind is weak and the current is southward, which explains the thin plume attached to the coast. The shape of the plume in member 34 is similar to the typical shape of cluster 2.

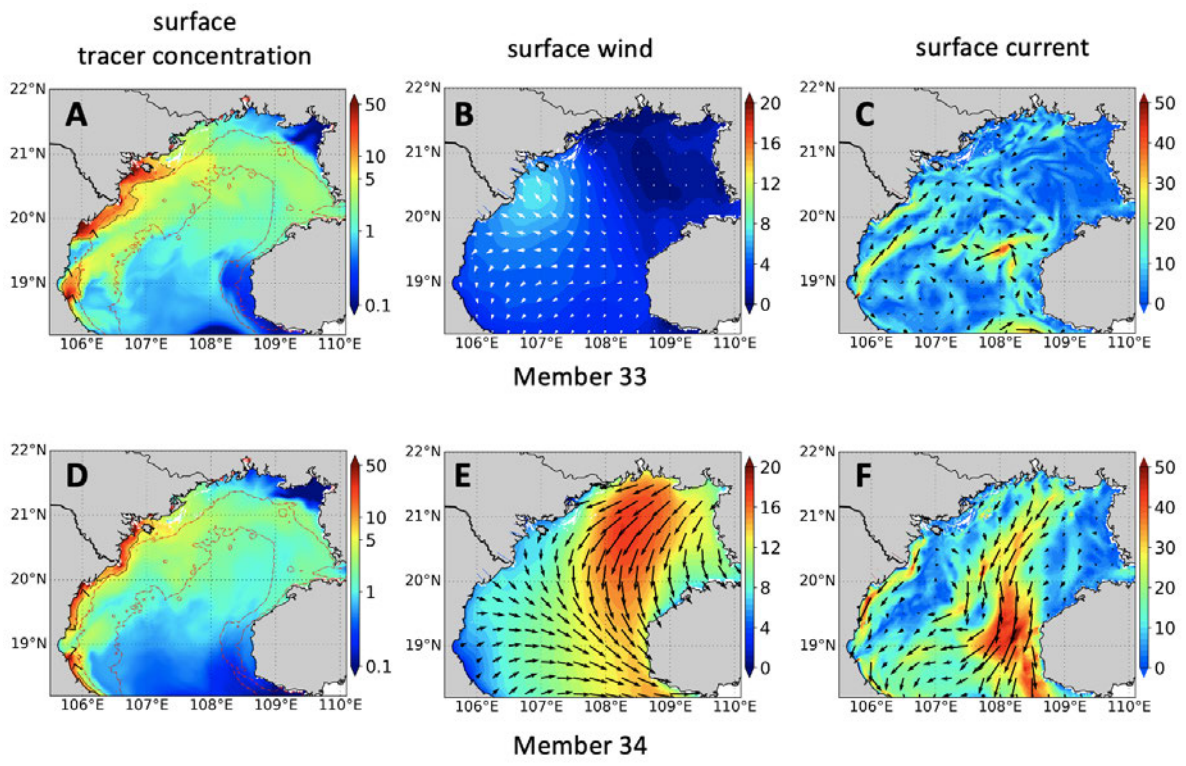

Figure V.16 (A): daily mean surface concentration (unit.m^{-3}), computed from 5 - 10 July for member 33. **(B)**: same as **(A)** but for surface wind (m.s^{-1}). **(C):** same as **(A)** but for surface current (cm.s^{-1}). **(D, E, F)**: same as **(A, B, C)** but for member 34.

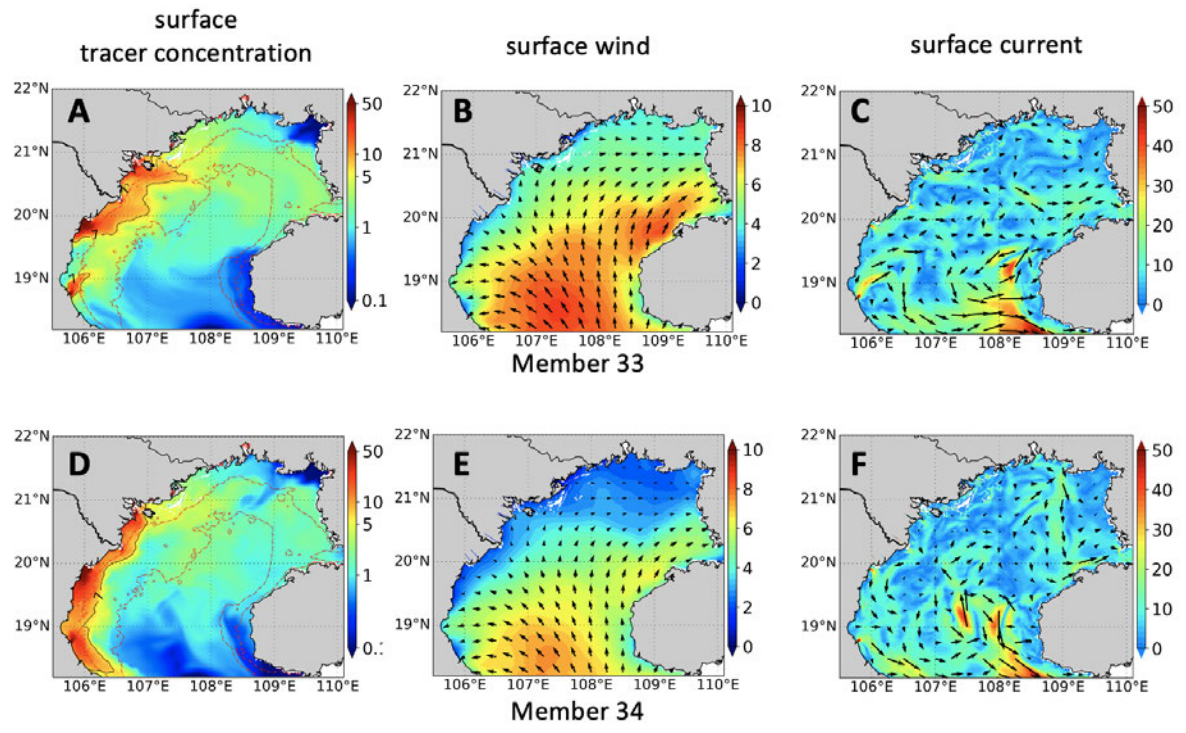

Figure V.17: same as **Figure V.16** but computed from 15 - 20 July.

V.3.2.3 Plume thickness

The plume thickness at 19°N in GOT_REF is shown in **Figure V.18A** as a function of time. It is computed in the same way as in Nguyen-Duy et al. (2021) (the base of the plume is the depth where the concentration is equal to 7 units/m^3). In June, the plume is about 2 m deep and detached from the coast. Then, the thickness increases and the plume extends further offshore, reaching the maximum thickness of around 10m on 20-25 July. In August, the plume thickness varies from 2 to 7m.

The spread of the plume thickness is shown in **Figure V.18B**. Due to the wind perturbation, the plume is not only changed vertically but also changed horizontally. Depending on the perturbation, the plume can stick nearer to the coast or extend further offshore. Due to the horizontal displacement of the plume, it is complicated to assess the spread of the ensemble thickness. After August 20, east of 106.3°N, the spread is nearly equal to 0 not because the plume is stable across the ensemble, but due to the fact that only the plume of 1-2 members can reach that longitude (**Figure V.18C**). It means that the presence of the plume in these points is actually unstable across the ensemble, which is consistent with the high spread of the plume area analysis above.

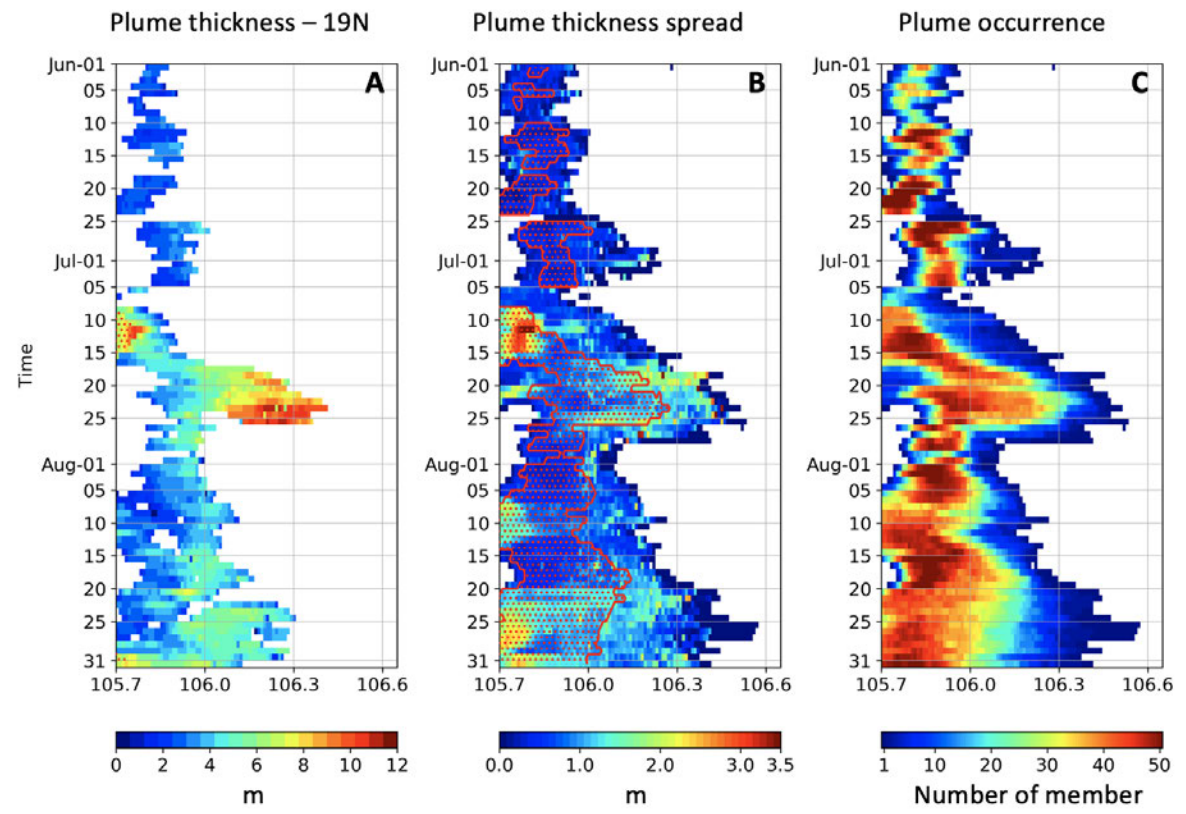

Figure V.18 (A): Plume thickness in GOT_REF (m) at 19°N as a function of time. Red dots represent the area where the plume reaches the bottom. **(B)**: spread of plume thickness (m) at the same section. Red dots indicate the area where the plume is present in at least 30 members of the ensemble. **(C)**: the number of members for which the plume is present and is used to compute the spread.

To better estimate the spread of the plume thickness, we look at the points at which the plume is present in at least 30 members (dotted red in **Figure V.18B**). In June, the spread is less than 1m. The peak of the spread happens from 09-13 July (> 3m), which is the same time of the largest spread of wind stress, suggesting that the change of the wind condition also has a considerable impact on mixing. Other high values of the spread occuralso in the period of 5-12 August and 21-28 August at the same times as the large variations of the plume pattern (**Figure V.15**), and the same time of peaks in the surface current spread (**Figure V.7A**) and in the SSS spread (**Figure V.12A**). In Section V.3.3 we analyze in more details the response to the wind perturbations in an attempt to explain such a variability.

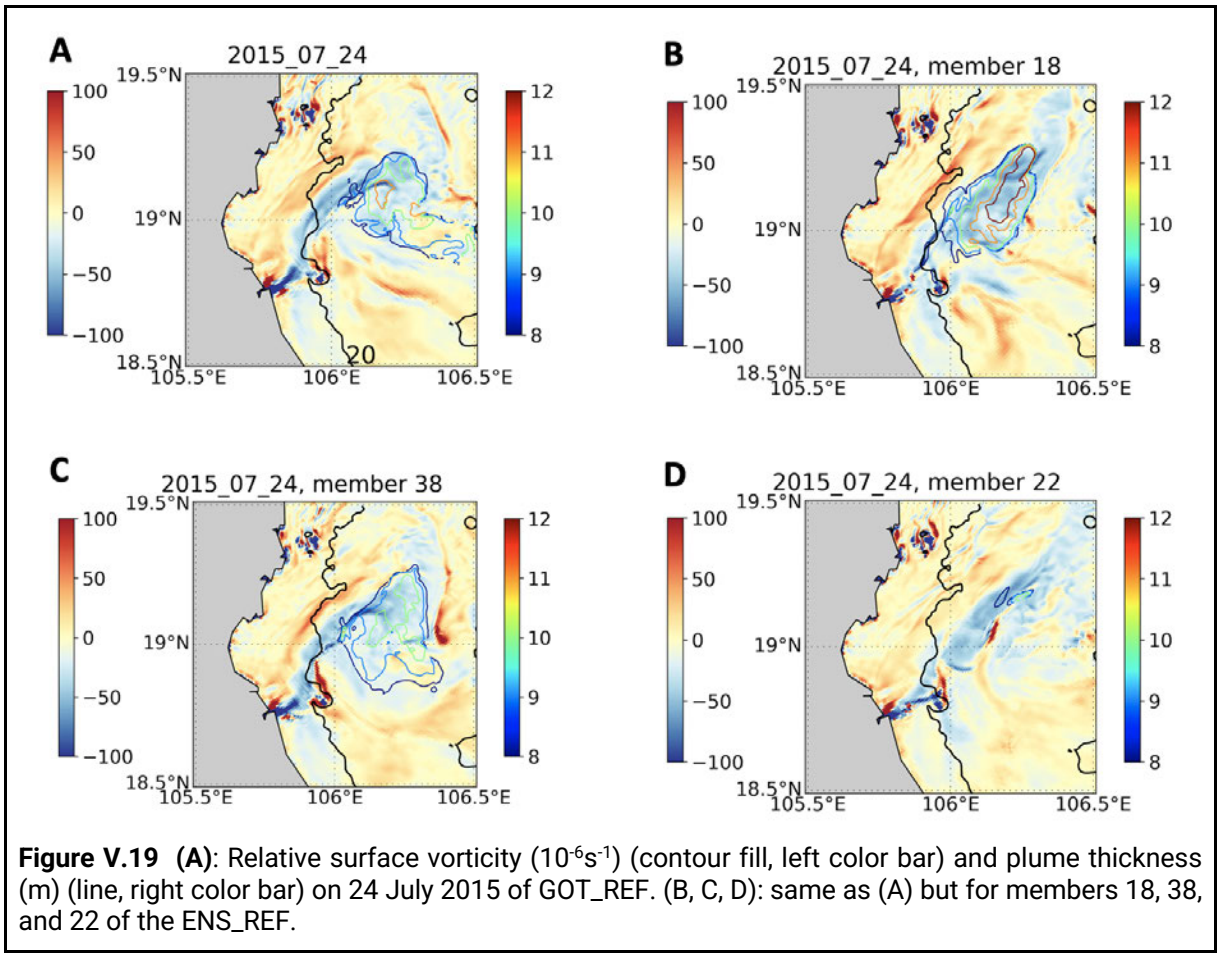

Figure V.19 (A): Relative surface vorticity ($10^{-6}s^{-1}$) (contour fill, left color bar) and plume thickness (m) (line, right color bar) on 24 July 2015 of GOT_REF. (B, C, D): same as (A) but for members 18, 38, and 22 of the ENS_REF.

In the period of 20-25 July, the plume in GOT_REF reaches the highest thickness, increasing from 3m near the coast to 10m offshore. The deepening of the plume is associated with an eddy (**Figure V.19A**), such as the one described in Nguyen-Duy et al. (2021). The map of relative surface vorticity on 24 July 2015 shows the signature of an anticyclonic eddy, and the plume is deepening inside. In this period, the spread is around 1-2m. Our first idea is that the processes causing or involved in the eddy have a critical impact on the mixing of the plume and the wind perturbations do not change much the mixing inside the plume. However, the plume thickness computed from each member of the ENS_REF shows that the wind perturbation can alter the current, change the shape of the eddy and finally modify the plume thickness. **Figures V.19 B, C, D** show the relative surface vorticity and plume thickness on the same day but from 3 members of the ENS_REF. Under different wind conditions, the plume can be either thicker

(**Figure V.19B**), thinner (**Figure V.19C**), or mostly disappear (**Figure V.19D**). Another interesting thing is that the eddy in GOT_REF is located in the area where the plume is present in less than 20 members of ENS_REF, which means it is unstable.

V.3.3 Event analysis

In this section, we will examine the impact of the wind perturbation on the plume variability in some special events.

V.3.3.1 27 July

From 25 - 30 July, both the wind stress of GOT_REF and the spread of wind stress is high (**Figure V.5B**). It is also the time when the ratio between the spread of the plume area and the unperturbed plume area is high (**Figure V.14A**).

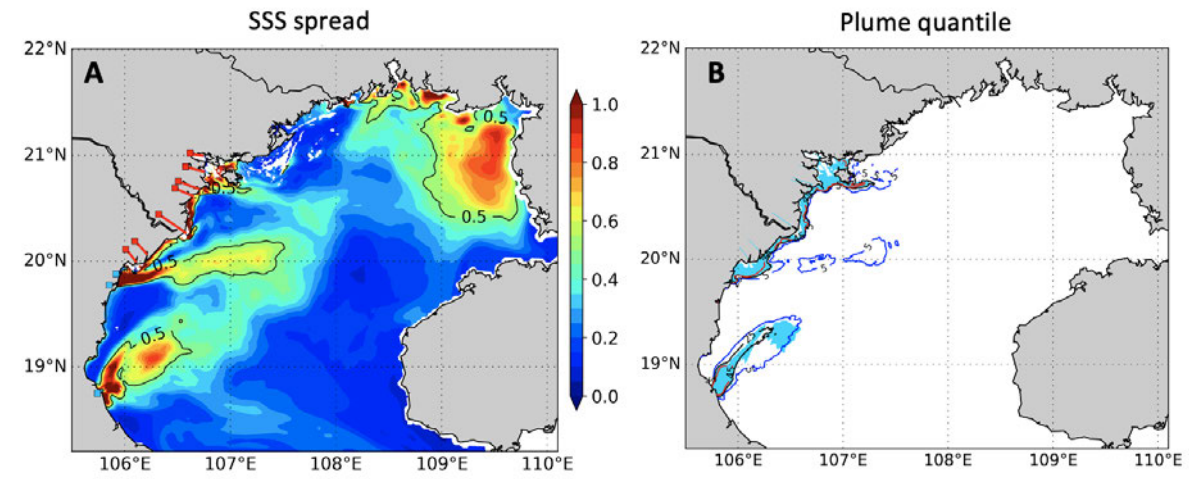

Figure V.20 (A): The spread of the SSS on 27 July 2015. (B): the quantile of plume occurrence on the same day. The contour lines show the area where the plume occurs in 5 (blue), 25 (black) and 45 (red) times over 50 members of ENS_REF. Cyan color shows the plume area from GOT_REF.

Figure V.20A shows the ensemble spread of SSS on 27 July 2015. The spread is highest near the RR mouth (>1) as expected, but it is also high in the northeast of the gulf which can be due to the runoff from northern rivers. The spread is high offshore, but is low near the coast from 19.0 - 19.8°N, indicating that the wind may be upwelling favorable in most members.

The quantile of plume occurrence is shown in **Figure V.20B**. Along the coast from 20 - 21°N, the plume seems to be stable across the ensemble, as demonstrated by the small difference between the 5-member and 45-member plume occurrence. Along the latitude of 20N, the plume spread is large. For half of the members, the plume can extend to 106.3°E, while in 5 members it can reach around 107.4°E, which is more than 100km from the coast. A large spread is also observed near 19°N.

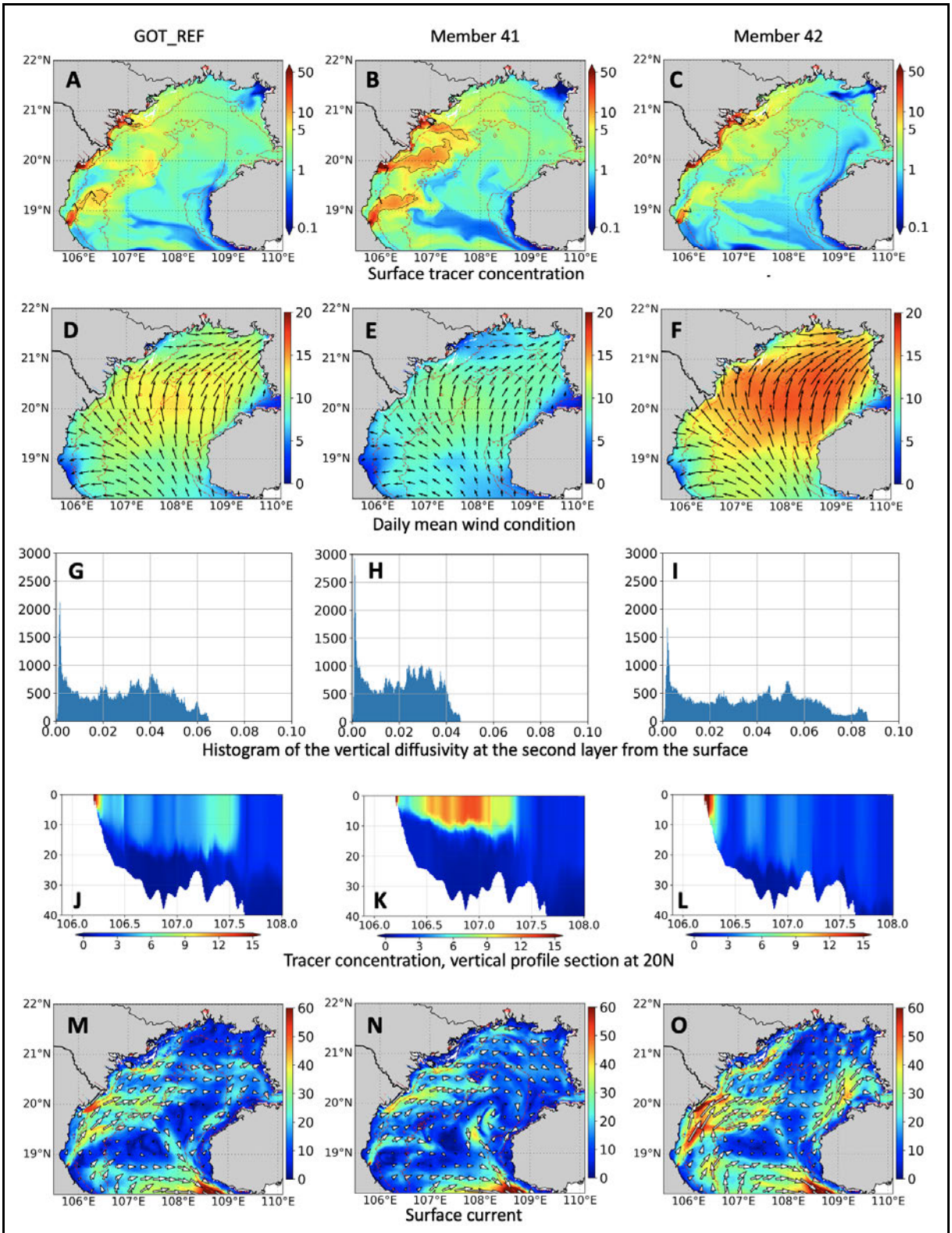

Figure V.21 Surface tracer concentration of GOT_REF **(A)**, the member with the largest plume area **(B)**, the member with the smallest surface plume area **(C)** on 27 July (arbitrary unit/m^3). **(D, E, F)**: Daily mean wind conditions (m/s). **(G, H, I)**: Vertical diffusivity at the second layer from the surface (in m^2/s). **(J, K, L)**: Vertical profiles of tracer concentration at 20°N (arbitrary unit/m^3). **(M, N, O)**: Surface current (cm/s).

Figure V.21 A, B, C shows the surface tracer concentration of GOT_REF, of the member with the largest plume area (member 41) and with the smallest plume area (member 42) on 27 July. In all 3 simulations, the passive tracer tends to extend offshore, however, the surface tracer concentration of member 41 is the highest and member 42 is the lowest.

The daily wind condition corresponding to each simulation is shown in **Figure V.21 D, E, F**. The wind patterns are similar, which is southeasterly in the south and southwesterly in the north of the gulf. Under upwelling favorable wind, the change of wind speed can have 2 types of impacts.

Firstly, it changes the mixing. A stronger wind increases the vertical mixing, therefore the tracer at the surface is decreased and the tracer concentration at the sub-surface is increased.. Indeed, the histogram of vertical diffusivity shows that the mixing is stronger under the stronger wind (**Figure V.21 G, H, I**). It reaches its highest (lowest) value in the strongest (weakest) wind condition. The vertical profiles of the tracer concentration confirm that the tracer can mix deeper in member 42 (which has strongest wind) to more than 30m, while in member 41 (which has weakest wind) the tracer is mixed in less than 15m (**Figure V.21 J, L**).

Secondly, the wind perturbation also changes the surface circulation. Near the RR coast, the surface current is eastward and northeastward in all cases (**Figure V.21 M, N, O**). However, the current is the strongest in member 42 in which the wind is the strongest. Under stronger current, the plume passive tracer is advected further and surface concentration is more diluted. The concentration of tracer can eventually fall below the threshold of the plume.

V.3.3.2 20 August

From 15 - 20 August, the river plume shows a large spread (**Figure V.14A**), although the spread of the wind stress is small (**Figure V.5A**). However, it is the period where the runoff is high.

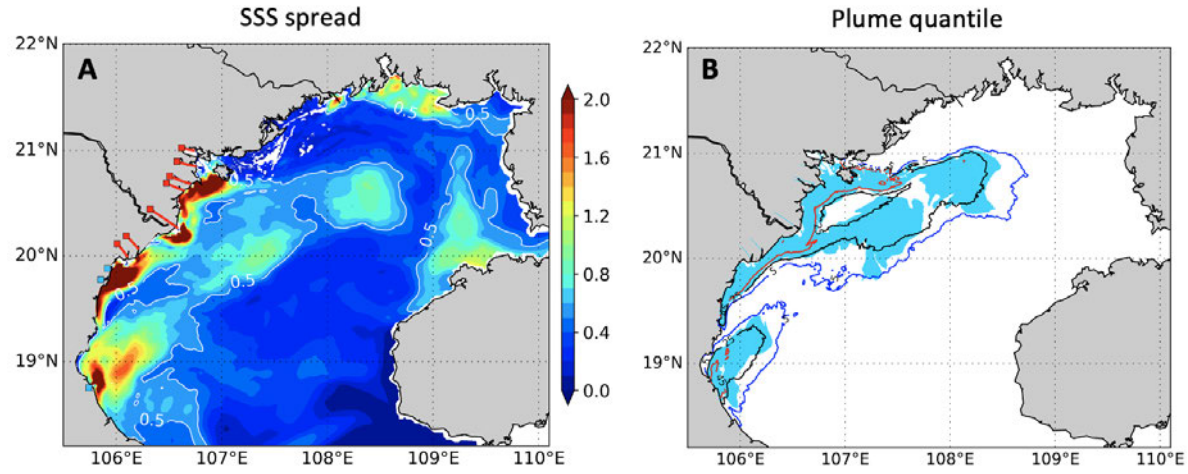

Figure V.22 (A): The spread of the SSS on 20 August 2015. (B): the quantile of the plume occurrence on the same day. The contour lines show the area where the plume occurs in 5 (blue), 25 (black) and 45 (red) times over 50 members of ENS_REF. Cyan color shows the plume area from GOT_REF.

The spread of SSS across the members of ENS_REF on 20 August are presented in **Figure V.22A**. The spread of the SSS is large offshore, indicating the large spread of the plume. Indeed, the quantile of the plume occurrence shows that the variance of the plume is large, as indicated by the large difference between the 5-member and 45-member plume occurrence (**Figure V.22B**), especially between 20°N and 21°N.

Figure V.23 A, B, C shows the surface tracer concentration of the GOT_REF, the member with the largest plume area (member 40) and the one with the smallest plume area (member 39) on 20 August.

In member 39, the small plume area is explained by the strongest wind, around 6 - 7 m/s near the coast between 20°N and 21°N (**Figure V.23F**), similar to the event of the 27 July, and also due to the strong downcoast current (**Figure V.23I**).

In GOT_REF, the coastal wind is the weakest (1 - 3 m/s) between 20°N and 21°N (**Figure V.23D**), however, the plume area does not reach the maximum like for the event on 27 July. Member 40, which has the in-between wind speed (3 - 5 m/s, **Figure V.23E**), has the largest plume area. The largest plume area in member 40 can be explained by the strong offshore transport current between 20 - 21°N (**Figure V.23H)**. It means that the runoff plays an important role on the spread of the plume. On July 27, when the runoff is low, the increase of the current due to the wind perturbation can dilute the plume, leading to a smaller plume area. However, on August 20, when the runoff is high, the increase of the current can help the plume to be advected more offshore and increase the plume area.

Using only the wind condition on August 20, it is hard to assess the effect of wind on mixing. In the first 15m, member 39, which has the strongest wind, seems to have the strongest mixing,

as shown by the vertical profile of concentration at 20°N (**Figure V.23L**). However, from 30m and below, the tracer concentration of member 40 is higher (**Figure V.23K**). It means that the dynamic on a day does not only depend on the current conditions, but also on the past conditions, which have been changed due to previous wind perturbations. Member 39 and 40 have been generated by paired, opposite-sign perturbations, but have totally different surface currents and tracer concentrations. The changes do not happen in just one day. Therefore, in the future, the impact of the wind perturbation could be assessed with enhanced benefit over a longer period.

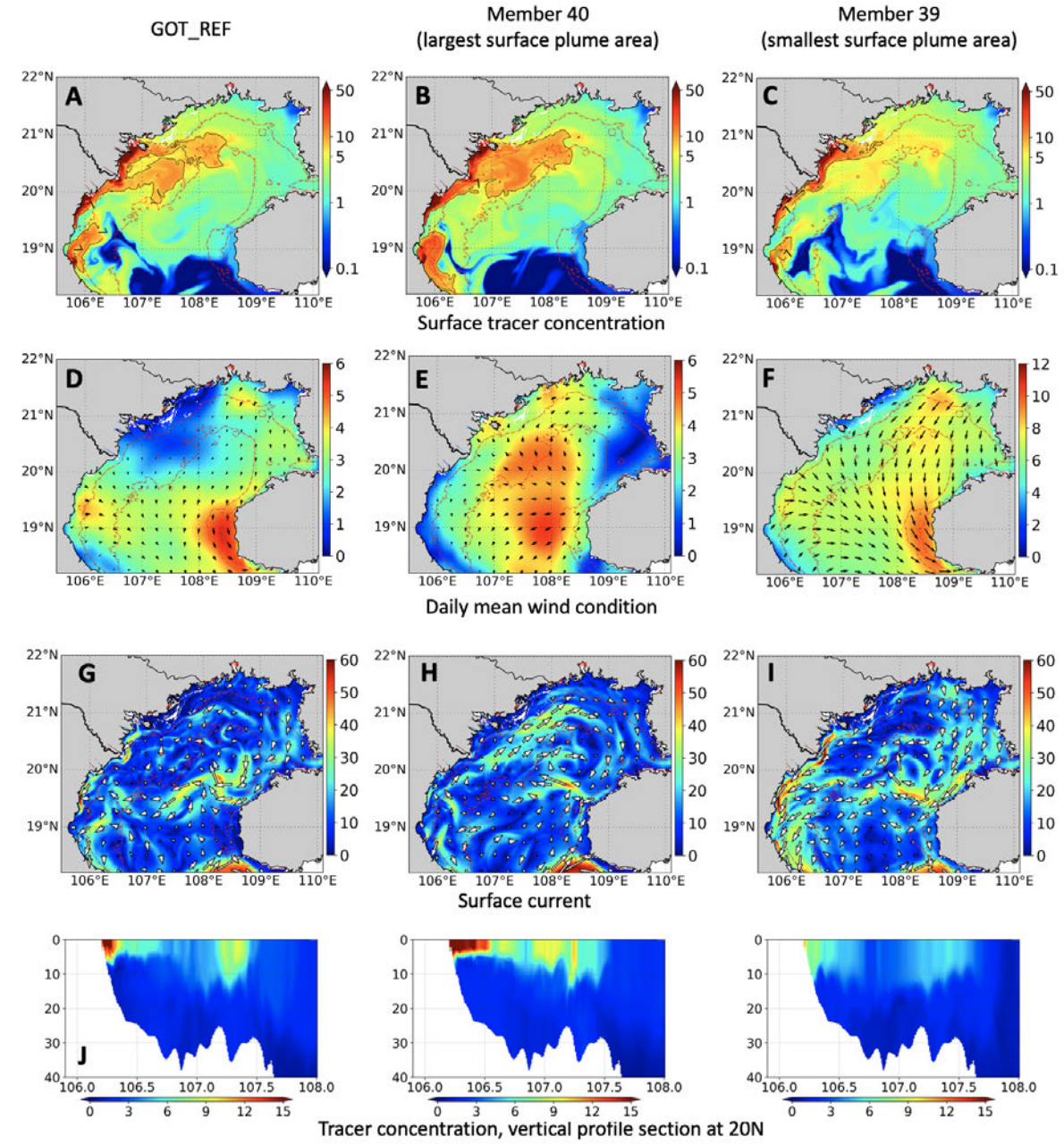

Figure V.23 Surface tracer concentration of GOT_REF **(A)**, the member with largest plume area **(B)**, the member with smallest surface plume area **(C)** on 27 July (arbitrary unit/m^3). **(D, E, F)**: Vertical profiles of tracer concentration at 20N (arbitrary unit/m^3). **(G, H, I)**: daily mean wind conditions (m/s). **(J, K, L)**: surface current (cm/s). Please note that the color scales for **D, E, F** are different.

V.3.4 Empirical ensemble verification

V.3.4.1 Rank histograms

To assess the ensemble with respect to observations, rank histograms (hereafter RH) are used. This method allows us to assess the compatibility between the distribution of the ensemble and the distribution of the observational data. It is also called the Talagrand diagram.

The principle of this method is as follows. In the ensemble simulation, the probability density function (pdf) of a variable X is estimated from N values Xi (i = 1,...,N) with N being the number of members of the ensemble. Ranking these values in ascending order defines N+1 intervals (ranks). The observational data Yk (k = 1,...,P) are then placed on the suitable rank based on their values. We use the following convention: if Yk is smaller than all ensemble members, it is assigned as rank 1, reversely, if it is larger than all ensemble members, it is placed at rank N+1. The histogram of the position of Yk with respect to the Xi 's defines a measure of the statistical consistency of the ensemble with respect to the observations: If Yk is an independent realization of the same pdf which has produced the Xi 's, it will be statistically indistinguishable from the Xi 's, and will therefore fall with equal frequency 1/(N + 1) in each of the intervals defined by the Xi 's (Talagrand et al. 1997).

The flat rank histogram indicates that the distribution of the observation is well represented by the ensemble (**Figure V.24**). Other than that, a rank histogram with a U-shape indicates that the spread of the ensemble is not large enough (the ensemble is underdispersive), while a dome shape means that the ensemble spread is too large (overdispersive ensemble). The figure also shows biased ensembles. Note that types can be mixed, e.g. underdispersive and high-biased. Another remark is that the presence of observational errors will modify the observational pdf and can make conclusions misleading.

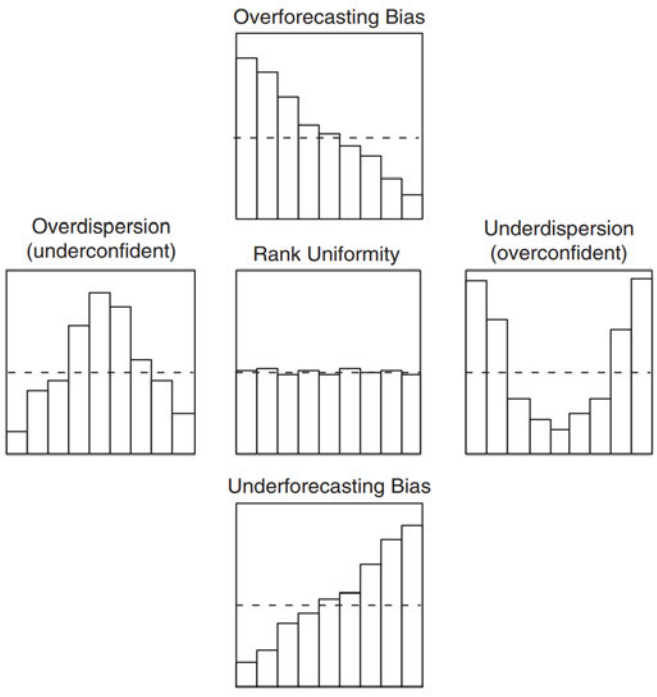

Figure V.24 Example verification rank histograms for ensembles with 8 members. Perfect rank uniformity (also called histogram flatness) is indicated by the dashed lines. Adapted from Wilks (2005).

Rank histograms are often used in ensemble/stochastic modeling and/or data assimilation. It can be used to empirically assess the ensemble, e.g. in Vervatis et al (2021b) and Yan et al. (2015) with respect to the physical (SST, SSH) and biological variables (phytoplankton functional types, chlorophyll). Santana-Falcón et al. (2020) use this method to assess the performance of the free-run and assimilated ensembles with respect to chlorophyll observations from SeaWiFS on selected regions. Van Velzen et al. (2016) use rank histograms to assess the performance of the Ensemble Kalman Filter (EnKF) data assimilation on SSH in different time scales and different numbers of members of the ensembles.

Here, we compute the rank histograms for SST and for surface currents, focusing on the plume area. In accordance with the usual practice for the calculation of rank histograms, the observations are not perturbed.

To compute the rank histograms of SST from ENS_REF and OSTIA data, firstly, for each day in June-August, the ensemble SST is spatially interpolated onto the OSTIA grid. Then, at each point of the area between 105.5 - 107.5E, 18.5N - 21.5N (**Figure V.25A**), OSTIA data are attributed to the suitable rank according to the procedure indicated above.

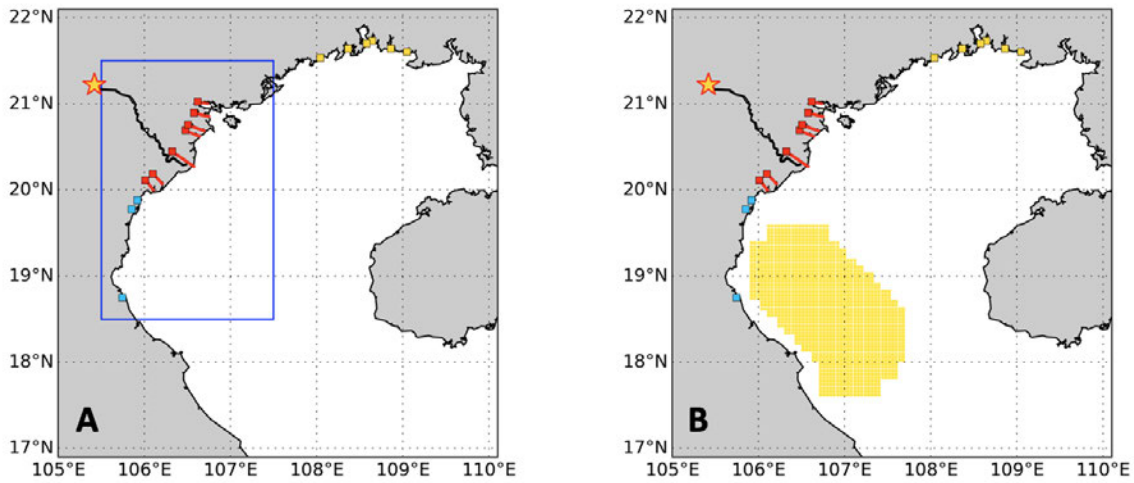

Figure V.25: **(A)**: Area to compute the rank histogram for SST (blue box). **(B)** Area to compute the rank histogram for surface currents (yellow-filled polygon).

The results are shown in **Figure V.26**. Often in the period, the histograms appears U shaped, with more values at rank 1 and rank 51, suggesting that the ensemble is underdispersive when this happens. However, in generating the ensemble, we only took into account one error process (surface wind), while it is not the only error process at play, on one hand, and on the other hand the OSTIA observations may contain errors.

The histograms typology is found to vary with time. In June, the ensemble is quite consistent with the OSTIA data, as shown with the relative flatness of the histograms, except on 3 days at the end of the month when it is low-biased (many data project onto class 51). The ensemble is the most low-biased from around July 25 to August 08. This inconsistency may be not related to the wind uncertainty but to the limitations of the river runoff configuration in the model.

First, the inconsistency happens when the runoff reaches its peak (**Figure V.14**). Second, in the model configuration, the maximum temperature of the river runoff is 29°C (see section II.2.2.2), while the SST in summer can reaches over 30°C on average (**Figure III.2**).

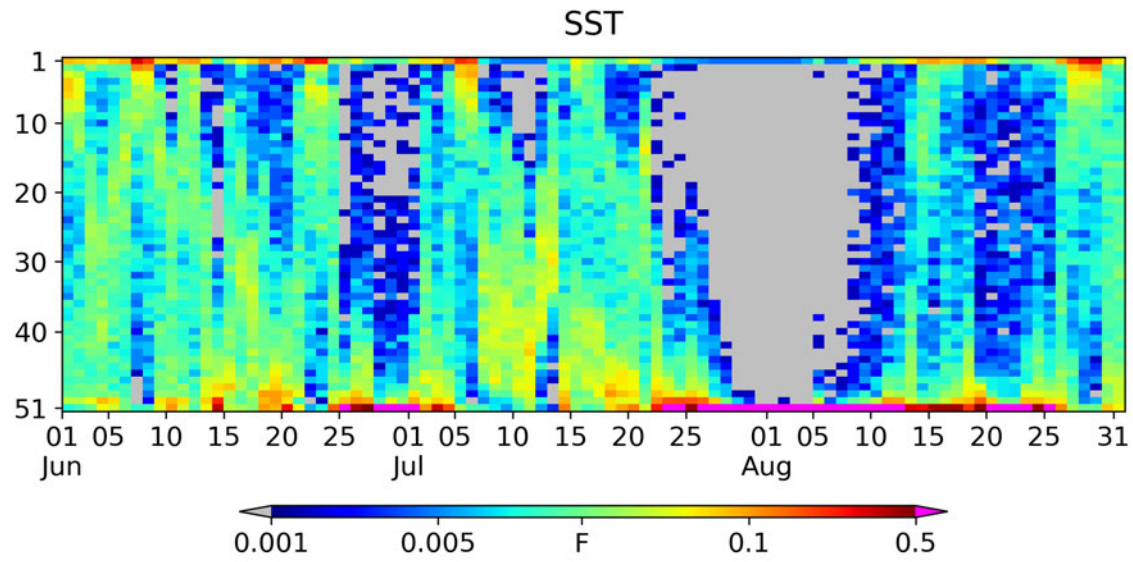

Figure V.26 Rank histograms of ENS_REF SST versus OSTIA data over the box, using 51 ranks (Rank 1 and 51 are open rank). F defines the histogram flatness. The y-axis shows the rank. Gray (pink) color shows the pdf that is lower (higher) than the color scale.

The rank histograms for the two components of the surface current versus the HF radar (hereafter HFR) data are shown in **Figure V.27**. The RH is computed over all points of the grid of the HFR data (**Figure V.25B**).

In general, the ensemble is found to be underdispersive, although results vary with time. For the U component, it is U-shaped in June, but it is high-biased in mid July and low-biased in middle and late August. For the V component, the ensemble appears low-biased in the first half of June, then the histograms have a U shape most of the time.

For both SST and surface currents, the histograms flatness seems to be more pronounced in July 5 - 15, which is the same time as the large spread of wind stress. This could mean that increasing the perturbation may improve the performance of the ensemble at describing the observed variance in statistical terms.

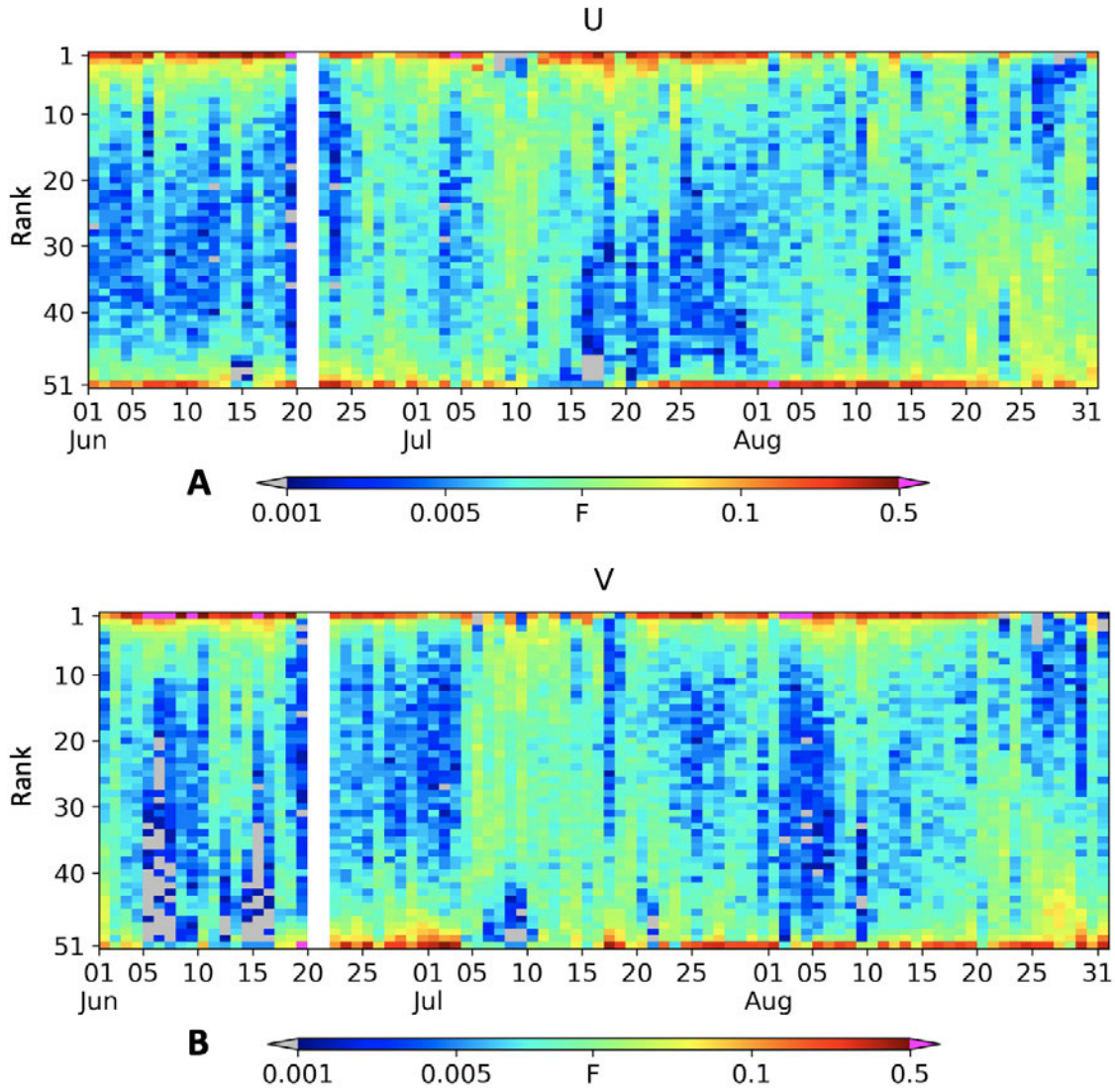

Figure V.27: Rank histograms of ENS_REF versus HF data, for U component **(A)** and V component **(B)**. The y-axis shows the rank. There is no data on 20 and 21 June.

V.3.4.2 Uncertainties budget

Another way to assess the relevance of the ensemble in simulating the model errors is to evaluate the skill of the spread at reproducing model errors. This method is based on the fact that we have two estimates for the model errors: the ensemble spread and, on the other hand, the misfits between the model and the observations. Of course the comparison is only valid within the observational errors. This test is useful in particular to assess the uncertainty budget in ensemble-based data assimilation methods.

In concrete terms, we first compute the spatial mean of the ensemble spread for each (U, V) component over the HF radar domain. We also compute the difference at each radar grid point between the ensemble-mean current and the observed one (for each component); we then compute the RMS of these differences over the radar domain (hereafter 'the RMSE').

Figure V.27 shows the temporal evolution of the spatial mean of ensemble spread and the ensemble-mean RMSE with respect to the HF radar current data. For the U component, most

of the time, the RMSE is around two times larger than the ensemble spread. On average, the temporal mean of RMSE and ensemble spread are 8.05 cm/s and 4.08 cm/s, respectively. The ensemble spread varies from 2 - 6 cm/s and increases gradually from June to August. The RMSE ranges from 4 - 15 cm/s (except the period of the tropical storm), with a peak in late July and early August. However, the difference between the two curves stays well within the observational error estimates: considering the end July/beginning of August period, taking $25cm^2/s^2$ for the ensemble variance in U and $100cm^2/s^2$ for the MSE of U, we end up with a variance budget of observational errors of $100\text{-}25=75cm^2/s^2$, i.e. a st.dev. of 8.66cm/s, which is compatible with the HFR observational error estimates. Indeed, Tran et al. (2021) estimate that the RMSE of HFR data with respect to the current measured at a mooring during 14 days in October 2015 is 12 cm/s for zonal current and 10 cm/s for meridional current (as shown in Chapter IV).

For the V component, the mismatch between the RMSE and ensemble spread is smaller. On average, the temporal mean of RMSE and ensemble spread are 7.16 cm/s and 4.76 cm/s, respectively. Similarly to the surface current rank histograms (Fig. V.26), the ensemble spread is increased and well matched with the RMSE from 5 - 15 July, which is the period of large wind stress spread, consistently with the result that the current is sensitive to the large wind perturbations. As for the U component, the difference between both curves is compatible with the estimated HFR observational errors, although the curves are closer to each other this time, and therefore the resulting observational error variance budget is smaller. This is in fact consistent with the smaller RMSE of the V component as estimated by Tran et al. (2021).

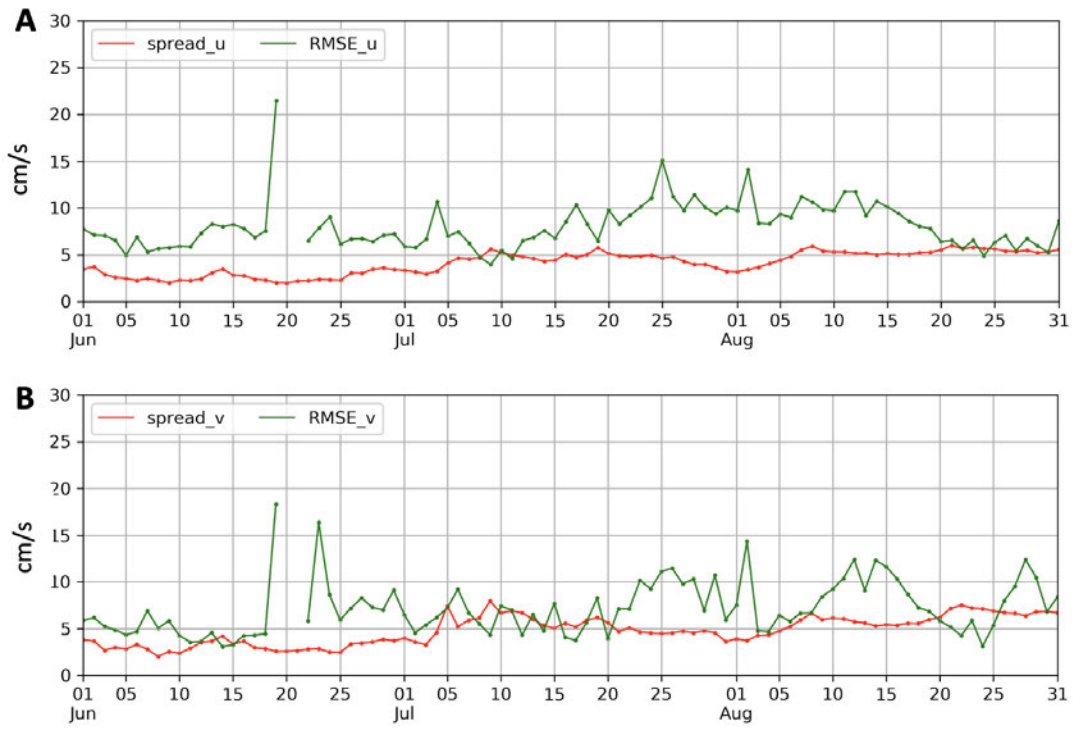

Figure V.27: (A): Spatial mean of ensemble spread (red line) of ENS_REF and spatial mean of RMSE between HF data and ensemble mean of ENS_REF, for U component. (B): same as (A) but for V component. Units are cm/s

V.4 Summaries and conclusions

In this Chapter, ensemble simulations are used to assess the impact of an important source of uncertainties, namely the wind uncertainty, on standard model variables in general and on the variations of the Red River plume in particular. In particular, we aim at assessing the robustness of the Chapter IV results and conclusions with respect to wind forcing uncertainties. In addition, we tried to address the following more general questions with the ensemble approach:

1. what is the space and time variability of the ensemble spread of the surface ocean variables ?

2. can we identify the physical processes at work to explain the sensitivity ?

One preliminary methodological question was “how to generate wind perturbations that are indeed representative of ECMWF wind errors ?”. By comparing the daily ECMWF data to the ASCAT data and ECMWF climatological data, the uncertainty of the ECMWF wind in this area is estimated as 60% of the wind variability. The wind perturbations are then created as linear combinations of bivariate wind EOFs weighted using a Gaussian pseudorandom number generator with standard deviation equal to the uncertainty value estimated above. The ensemble is run with 50 members, perturbed by pairs of opposite perturbations so that the ensemble average of perturbations is zero.

The ECMWF to ASCAT comparison also indirectly answers another methodological question, “is the generated ocean ensemble significantly representative of model errors ?”, insofar as we calibrated in this manner the stochastic forcing. But the question also needs of course to be addressed for the ocean variables (see below).

Another preliminary methodological question was “how (by means of which metrics) do we define the sensitivity to wind uncertainties with an ensemble? In terms of metric definition?”. In the adopted approach, the analysis of the ensembles in terms of sensitivity consisted mainly in calculating the ensemble standard deviation (the “spread”), as the ocean response of ocean variables to the stochastic forcing, and the statistical distributions of ensemble samples.

The assessments are mostly based on the ensemble spread, which is computed as the standard deviation along the ensemble with respect to the ensemble mean. First, we assess the spread of the model variables. The results show that the SSH spread is less than 2 cm most of the time. The spatial mean spread of the SSH over the whole period decreases southerly. The spread increases locally near the coast where the river plume is present. The peak of the spread happens at the same time as the peaks of the wind stress, which is also the peak of the EOF1, indicating that the result may be sensitive to the EOF analysis.

The surface current shows the peaks of the ensemble spread mostly at the same time as the peaks of SSH spread, although the links between both would have to be further clarified. The high spread along the coast in V component suggests that the wind perturbation has an impact on the coastal current created by the river runoff.

The temporal evolution of the spatial mean spread of SST shows the same time pattern as the SSH. However, unlike the SSH, the maximum spread of SST is highest near the south of Hainan island. We expect that the spread of the wind leads to the change of the mixing (which is already high due to tidal effect in that area) and then leading to the displacement of the water mass that

has a high temperature gradient. Over the whole period, the mean spread is highest over the Red River coast, which can be related to the difference between temperature of the riverine water or the upwelling present in this area. The SSS, as expected, has high spread near the Red River.

The vertical section at 19°N is investigated to assess the vertical structure of the spread of temperature and salinity. The temperature spread is highest around 20-30m deep, where the vertical gradient is highest (base of the mixed layer; as in Andreu-Burillo et al, 2002, and Ayoub et al., 2015), a clear confirmation of the fact that the uncertainty of the wind affects vertical mixing. On salinity, the spread is highest near the surface in the vicinity of the sanity gradient due to the runoff.

Next, the spread of the Red River plume is assessed. At the surface, the spread of the plume area is highest in August, but it does not follow the same trend as the SSS. The maximum spread of the plume area happens at the same time as the maximum of the plume area, suggesting that our results might be sensitive to the uncertainty of the river runoff as well.

The clustering analysis was carried out per-member using as a reference the clusters of GOT_REF, as computed in Nguyen Duy et al. (2021). Results evidence some changes of clusters across the ensemble, but almost all the time, the pattern that is most likely to occur is still the reference cluster. As a consequence, we estimate that within our experimental conditions the findings of Nguyen-Duy et al. (2021) are yet unchallenged and robust with respect to wind perturbations.

Finally, the spread of the plume thickness shows that the plume is not only changed vertically but also changed horizontally. In June, the spread is less than 1m. The peak of the spread happens from 09-13 July (> 3m), which is the same time as for the largest spread of wind stress, suggesting that the change of the wind condition also has a considerable impact on mixing. The plume can deepen inside the eddy, as discussed in Nguyen-Duy et al. 2021. The wind perturbation can alter the current, change the shape of the eddy and finally modify the plume thickness.

The spread of the plume for some specific days is also examined. The stronger wind increases the vertical mixing. In particular, stronger upwelling favorable wind can increase the current and help the plume to be advected further. In two members which have in-pair perturbations, the current conditions can be totally different in both direction and value, which is due to the past conditions, especially in GOT since the wind can change very fast. Therefore the impact of the wind perturbation should be assessed over a period (not instantaneously).

We would like to finish with several remarks and limitations of our approach.

First, we found out that the wind difference patterns obtained when matching ECMWF data with ASCAT data were not quite the same as the patterns generated by our base of EOFs, even if the error scales seemed similar. We did not take into account the spatial structure of ASCAT errors when doing that comparison, because we do not know those errors. As a consequence, we limited ourselves in this study to using that comparison to get a very rough estimate of ECMWF error variances, at the risk of overestimating them and of structures of smaller scales than those actually taken into account, and did not try to infer anything from the difference patterns themselves.

Second, in this study, we limited the source of stochastic perturbations to the surface wind only. Although several studies, cited in the introduction of this Chapter, agree on the wind being an essential, if not dominant, source of ocean uncertainties, we could have considered other sources. For instance, we became aware during this study that the runoff could be another major source of uncertainty impacting the RR plume, along with the wind. As another example, we could not fully clarify the apparent partial correlation between the spread of surface currents and the SSH spread – in order to better explore that link, atmospheric pressure would need to be perturbed in a consistent way with the wind, in particular since the GoT is a shelf sea and semi-enclosed sea, so it should be sensitive to water displacements as, e.g., a low-pressure center passes through. One way to do this comparison in a meaningful way would be to use multivariate ECMWF ensembles, perturb both the wind and the pressure, and perhaps add a wave model to better model the wind effects on currents. We did not have the time to explore those two directions, but it will need to be done.

CHAPTER VI: CONCLUSIONS AND PERSPECTIVES

***

VI.1 Conclusion

Gulf of Tonkin is a shallow region (depth <60m) of the southwest Pacific, under the influence of the Red River (hereafter RR) discharge, of a strong atmospheric variability (monsoons, typhoons) and of tides. The Red River delta has a total population of 22.5 millions and an average population density of 1060 inhabitants/km^2, which is nearly four times higher than the national average for Vietnam. The very high population density exerts a strong pressure on the rivers and their environment.

This work is part of the LOTUS project. LOTUS is a joint Vietnam-France project, which aims to better understand the coastal ocean dynamics under the influence of atmospheric variability and of the delta. In the PhD thesis of V. Piton (2019) at LEGOS (Toulouse, France), a high-resolution configuration in the Gulf of Tonkin and downstream part of the Red River delta has been set-up to study tides with the 2D spectral TUGO model and basin-scale circulation with the 3D SYMPHONIE model. In this study, the main objective was to provide a better understanding of the variability of the RR plume in the Gulf of Tonkin at different scales using an updated version of the configuration with SYMPHONIE.

The main scientific questions which I attempted to address are the following:

1. How can high-resolution numerical simulations reproduce the Gulf of Tonkin dynamics and the Red River plume variability at different scales in the study years?
2. Which criteria can be used to identify the RR plume in the GoT from the model outputs?
3. How can we formally characterize the sequence of regimes which the RR plume goes through in the study years? Can we link those regimes to specific processes and forcings ?
4. How can we assess the sensitivity of model variables and in particular on the variability of the river plume, to the main sources of uncertainty, in particular the atmospheric forcing uncertainties?

To address those questions, we adopted the following tool-based **approaches**:

1. We set up and run an upgraded version of the SYMPHONIE model configuration created as part of the thesis of V. Piton. A reference simulation (GOT_REF) is run with realistic forcing conditions over a 7-year period (2010-2016). Additionally, a twin simulation without river forcing (GOT_NORIV) is performed to assess the impact of the river runoff on the coastal circulation and a twin simulation without tides (GOT_NOTIDE) is run to evaluate the impact of tides on the main patterns of the river plume variability.
2. The river plume was identified using different methods, allowing us to select the most appropriate one given our purposes. The selected method is based on passive tracers injected in the rivers.
3. The RR plume regimes were identified and linked to physical processes with the help of K-means clustering analysis (KMA).

4. We ran an ensemble of simulations assessing the model response due to perturbations of the wind forcing. We analyzed the variability of the resulting uncertainties over space and time of the ocean variables. The K-means clustering approach was used on ensemble members, providing an assessment of the robustness of the results of 3. above.

Using the tool described above, combined with the observational data, we obtained the main results below:

1. Model assessment

The model results from GOT_REF have been assessed by comparing with available observational data: satellite data (altimetry for tide, OSTIA for sea surface temperature, SMOS for sea surface salinity), in situ data (VITEL, CFO, both datasets with temperature and salinity profiles), and HF radar (for surface current). We also compared our results with the OGCM (the global circulation model) to emphasize the added value of our high resolution regional model compared to a global model. Overall, despite some biases, the model reproduced well the seasonal conditions even in the very shallow area. The comparison with the CTD data near the coast shows an overestimation of the surface salinity, therefore we suspect that the runoff may be underestimated.

We also discussed the typical horizontal length scale (first baroclinic Rossby radius) computed from the simulated density field and described the general circulation.

2. RR plume identification

To identify the RR plume, different methods were taken into consideration. Firstly, we applied a method based on the Stratification Index (hereafter SI), assuming that the river water forms a buoyant layer over the ocean water. The second method consisted in identifying the plume from the sea surface salinity (SSS). Thirdly, we used passive tracers that behave in the model as any buoyancy-free particle or passive chemical from the rivers. Several experiments led to an identification of the river plume as the area where the tracer concentration exceeds 7 unit/m^3, which best fits with the area identified by the criterion on SSS in the low runoff period (dry season).

The maps of the plume area show that south of 20°N, the RR plume is quickly joined by the plumes of three rivers south of the delta (Ma, Yen, Lam, hereafter the 'southern rivers'), creating a unique buoyant plume which extends southward along the coast most of the year. Therefore, we analyzed the variability of the resulting plume from the RR and the three southern rivers together. We first computed the plume surface area (in km^2) and found that the plume area follows the same variations as the river runoff, albeit with a time lag. Due to the interannual variability of the runoff, the plume area varies significantly between different years. Both the total discharge and the plume area undergo a strong variability at shorter time scales of a few days as well.

3. RR plume variability

The temporal variation and the regimes of the plume surface patterns.

The variation of daily spatial patterns of the plume is examined using KMA. The first cluster usually occurs from October to March, and is the most frequent cluster. In this cluster, the plume is very narrow and is mostly confined to the shallow area (depth < 20 m). The second cluster appears on and off throughout the year, with a slightly greater rate of occurrence in April and September which is during the seasonal transition of the monsoon. In this cluster, the river plume extends both further offshore and further southward compared to cluster 1. The third cluster happens primarily in June and July. In this cluster, the river plumes from the various rivers are disjoint. The fourth cluster exhibits the largest spatial coverage and occurs mostly in August and September. In this cluster, the plume extends the farthest offshore. Each cluster is associated with different forcing conditions (runoff, wind, current) that explain its spatial and temporal variability.

The thickness of the plume.

From September to March, the plume is narrow, elongated along the coast and attached to the bottom due to the winter monsoon downwelling winds. From May on, the monsoon changes to southwesterly wind, driving the plume northward and detaching it from the coast and from the bottom. In that case, the plume is surface advected with a thickness of 2-3 to 7-8 m. We observe that events where the plume detaches from the coast at 19°N coincide with the temperature at the coast colder than seaward. This supports the assumption that the southerly monsoon winds generate a coastal upwelling locally. Sometimes in May-August, the plume is thickening offshore to 12-15 m. That deepening may be due to strong winds events that increase the mixing offshore in the upwelling region or to the increase of river discharge.

In mid-August, we observed a strong plume deepening over the shelf (106-106.5°E). The deepening events appear to be linked with a seasonally recurrent anti-cyclonic eddy developing near 19°N. In our simulation, this eddy generally occurs in August, when the wind direction is from the southwest, which is upwelling–favorable, and the runoff is high. Then, this eddy disappears when the coastal southward current develops again and/or the wind is not upwelling-favorable anymore. The formation mechanism of this eddy is still unclear. In GOT_NORIV, some vorticity gradients can be depicted, suggesting an anticyclonic structure, but its amplitude is much weaker than in the reference simulation and it vanishes on a shorter time scale. The anticyclone seems strongly connected to the Lam river plume. Bottom topography may be another forcing factor since the eddy is located in the vicinity of an island.

Impacts of tides.

The influence of tides in shaping the far field plume is investigated by comparing the reference simulation GOT_REF with the simulation without tides (GOT_NOTIDE). On average, the plume in GOT_REF is 5% larger than in GOT_NOTIDE, with the largest differences found in the high discharge period. The KMA analysis on the GOT_NOTIDE shows that the 4 clusters are very close to the ones from GOT_REF, both in terms of spatial structures and temporal distribution.

However, tides impact the vertical structure of the plume: they enhance the vertical mixing hence the riverine water concentration in the bottom layer. They also lead to a larger export offshore of the riverine water at the surface close to the mouth during the high discharge season.

As this study focuses on the far field, we have not looked into details at the impact of tides on the water properties and dynamics at the mouth of the rivers, through the tidal-induced bottom friction. This should be done in future work.

4. Model response due to perturbations of the wind forcing

By comparing the daily ECMWF fields to the ASCAT data and ECMWF climatological mean, the uncertainty of the ECMWF wind in this area was estimated as 60% of the wind variability. The wind perturbations were then created as linear combinations of bivariate wind EOFs weighted using a Gaussian pseudorandom number generator with standard deviation equal to the uncertainty value estimated above. An ensemble of SYMPHONIE simulations was run with 50 members, perturbed by pairs of opposite perturbations so that the ensemble average of perturbations is zero.

The assessments are mostly based on the ensemble spread, which is computed as the standard deviation along the ensemble with respect to the ensemble mean.

The spread of the model variables.

The results show that the SSH spread is less than 2 cm most of the time. The mean spread of the SSH over the whole period decreases from the north to the south. It increases locally near the coast where the river plume is present. The peak of the spread occurss at the same time as the peaks of the wind stress spread, which is also the peak of the EOF1, and corresponds to a reversal of the monsoon over a few days..

The surface currents are highly sensitive to the wind perturbations, as expected. In particular, the high spread along the coast in V component suggests that the wind perturbation has an impact on the coastal current created by the river runoff. The coastal current can have opposite direction between two members at daily time scales, while the spread can exceed the current amplitude of the ensemble mean.

The temporal evolution of the spatial mean spread of SST shows the same time pattern as the SSH. However, unlike the SSH, the maximum spread of SST is highest near the south of Hainan island. We expect that the wind perturbations lead to changes in vertical mixing (which is already high due to tidal effect in that area) and then leading to the displacement of the water mass that has a high temperature gradient. Over the whole period, the mean spread is highest over the Red River coast, which can be related to the difference between temperature of the riverine water or the upwelling present in this area. The SSS, as expected, has high spread near the Red River delta.

The vertical section at 19°N is investigated to assess the vertical structure of the spread of temperature and salinity. The temperature spread is highest around 20-30m deep, where the vertical gradient is highest (base of the mixed layer; as in Andreu-Burillo et al, 2002, and Ayoub et al., 2015), a clear confirmation of the fact that the uncertainty of the wind affects vertical

mixing. On salinity, the spread is highest near the surface in the vicinity of the sanity gradient due to the runoff.

The spread of the Red River plume.

At the surface, the spread of the plume area is highest in August, but it does not follow the same trend as the SSS. The maximum spread of the plume area happens at the same time as the maximum of the plume area, suggesting that our results might be sensitive to the uncertainty of the river runoff as well.

The clustering analysis was carried out per-member using as a reference the clusters of GOT_REF, as computed in Nguyen Duy et al. (2021). Results evidence some changes of clusters across the ensemble, but almost all the time, the pattern that is most likely to occur is still the reference cluster.

Finally, the spread of the plume thickness shows that the plume is not only changed vertically but also changed horizontally. In June, the spread is less than 1m. The peak of the spread happens from 09-13 July (> 3m), which is the same time as for the largest spread of wind stress, confirming the impact on vertical mixing. The plume can deepen inside eddies, as found near 19°N and as discussed in Nguyen-Duy et al. 2021. The wind perturbation can alter the current, change the shape of the eddies and finally modify the plume thickness.

The spread of the plume for some specific days is also examined. The distribution over the ensemble of the vertical mixing coefficient evidences the increase of mixing in members with stronger wind. The mixing can in turn lead to a decrease of the tracer concentration at surface; if the concentration decreases below the threshold chosen to identify the plume, then the plume is not identified anymore, while a thicker mass of water is influenced by the river input. This suggests that using two conditions to identify the plume (one based on the surface concentration and the other one of the thickness of the water column impacted by the riverine waters) could be more robust to uncertainties in the wind forcing. Stronger winds also favor a spreading of the plume further offshore. However, disentangling the wind effect on vertical mixing and on advection is not straightforward. First, because the horizontal transport is also driven by the plume buoyancy, at least in the mid field (Horner-Devine et al., 2015). Second, because the ocean conditions at a given day result from the cumulative effect of forcing over a ‘certain’ period before (maybe several days). This should be taken into consideration when comparing the plume properties between different members at a given date, but as the daily variability of the wind is quite high in the GOT, such an approach is quite complex.

The ensemble has been assessed empirically using the SST and HFR data. In both cases, the ensemble is underdispersive, with more value fall in the first and last rank of the rank histogram. However, the error budget analysis shows that the difference between the ensemble spread and the ensemble-mean RMSE is compatible with the HFR observational error estimates.

Finally, to conclude, we estimate that within our experimental conditions the findings of Nguyen-Duy et al. (2021) are yet unchallenged and robust with respect to wind perturbations.

VI.2 Perspective

1. Model configurations and assessments

The assessment of the model representation of the plume dynamics evidences a lack of systematic high-resolution observations of the plume area, except for the limited area covered by the HF radar. The contribution of present altimetric measurements from the Jason satellite series and Sentinel-3 satellites should be investigated, but their spatial coverage (limited to the satellite nadir) may still be an issue. The future SWOT mission, with wide-swath altimetry (e.g. Fu et al., 2012) will represent a large amount of additional data which are expected as a precious contribution to both the delta and plume observation.

In this study, the river configuration is set up as realistic as possible, however, in the future, further improvements can be made.

The first improvement can be related to the runoff data. In this study, the total runoff in the Red River is totally the runoff data from the Son Tay hydrological station. However, there are other sources of the runoff, from Cau, Thuong, and Luc Nam river, that were not taken into account in this study because the discharge data was not available to us. The Web portal of Bac Giang (2016) indicates that the total runoff from 3 rivers is 7.52 billion m^3/year, which is equal to around 238 m3/s while Vinh et al. (2014) estimates a different value of 133.5m^3/s. These values are equivalent to around 4 - 7% more of the total runoff in the RR. Also, the ratio of the runoff between the high discharge and low discharge period for the southern rivers is higher than RR. This can be explained by the fact that the RR runoff is controlled by dams. However, recently, some dams were also constructed in the southern rivers (e.g. Trung Son dam commissioned in 2017). This should be considered if the simulation period is extended in the future.

The second improvement can be the better distribution of the runoff over the RR mouth. In this study, the RR runoff is distributed throughout different distributaries using the annually averaged percentage of discharge made by Vinh et al. (2014). However, as indicated in their study, this percentage also has seasonal variation. For example, in Van Uc river, the discharge percentage for dry and rainy season is 14.1% and 14.6%, while for Cam, it is 9.9% and 9.4%, respectively. Therefore, another ensemble study could be useful to assess the impact of the uncertainty of the discharge rate and the distribution percentage on the plume variability.

2. RR plume variability

Warrick and Farnsworth (2017) suggested that more studies on interactions among different plumes are needed in general. As shown by Gong et al. (2019), a plume can have an impact on another plume, therefore the interaction between plumes is important for environmental management. With different passive tracers, the interaction between different rivers in the RR plume can be examined and can be a promising topic to explore.

To identify the main regime of the surface river plume, K-means clustering method is used in this study. Though 4 clusters is enough to describe the main pattern, it is shown that the standard deviation of the SSS inside cluster 3 and cluster 4 is higher than inside cluster 1 and cluster 2, suggesting that the forcing conditions in clusters 3 and 4 are more unstable. A nested clustering inside these clusters can be done to further explore the plume pattern at the smaller scale.

The impact of tides on the shape of the plume has been explored, however, an in-depth analysis of the tidal processes, of the tides-river flow interactions and of their impact on the velocity, shear and stratification in the estuaries and at the mouths is still an open question. It will be the topic of a future study aimed at better understanding the shape and dynamics of the outflow in the near and mid fields. The dynamics at the mouth is also highly important for sediment dynamics close to the mouth area.

The dilution of river waters due to stirring by eddies or mixing could be explored further than what we started in this study. Actually, hourly outputs of currents from the GOT_REF and GOT_noriv simulations have been used by Cuong Manh Tran and Alexei Sentchev (LOG) to compute Lagrangian diagnostics (such as diffusivity) and compared them with diagnostics from HF radar currents. An article by Cuong Manh Tran is in preparation and should be submitted in March 2022 ('Assessment of relative dispersion in the Gulf of Tonkin using numerical modeling and HF radar observations'); I am co-author of this article.

The study of mixing due to instabilities at the front of the plume (eg Horner-Devine et al., 2015) or within the plume (such as those described by Ayouche et al. 2021 in the Gironde River plume for instance) was beyond the scope of this work. However, such submesoscale processes should be quantified in dedicated studies to better understand the dynamics at work in the Red River plume and to better anticipate the modelling and observational needs for an accurate and reliable monitoring and forecasting system of the plume. Similarly, the impact of surface waves (swell or sea state) on the mixing and their interactions with the current within the plume are other processes that would deserve further work.

3. Model response due to perturbations of the wind forcing

In this study, we did not take into account the spatial structure of ASCAT errors, because we do not know those errors in the GOT region. As a consequence, we limited ourselves in this study to using the comparison between ECMWF and ASCAT to get a very rough estimate of ECMWF error variances. Future studies should explore methods to extract more information from the satellite wind to generate the perturbations.

We also limited the source of stochastic perturbations to the surface wind only. Although several studies, cited in the introduction of this Chapter, agree on the wind being an essential, if not dominant, source of ocean uncertainties, we could have considered other sources. For instance, we became aware during this study that the runoff could be another major source of uncertainty impacting the RR plume, along with the wind. As another example, we could not fully clarify the apparent partial correlation between the spread of surface currents and the SSH spread – in order to better explore that link, atmospheric pressure would need to be perturbed in a consistent way with the wind, in particular since the GoT is a shelf sea and semi-enclosed sea, so it should be sensitive to water displacements as, e.g., a low-pressure center passes through. One way to do this comparison in a meaningful way would be to use multivariate ECMWF ensembles, perturb both the wind and the pressure, and perhaps add a wave model to better model the wind effects on currents. We did not have the time to explore those two directions, but it will need to be done.

Due to the constraint in computing resources, in this study, the ensemble has 50 members. This value is based on our experience in previous studies with the similar configurations(for example

Ghantous et al. (2021) with 50 members, Quattrocchi et al. (2014) with 40 members, Vervatis et al. (2021) with 40 members). Therefore, a sensitivity test on the impact of the ensemble size on the spread estimate should be done in the future.

In addition to the analysis of the ensemble spread, another approach to characterize the ensemble could be based on the clustering analysis. This would allow us to assess the pattern variability across the ensemble, for example, clustering across the members for a given day to assess the general regime of the model response.

We could also compute ensemble covariances in order to better characterize the spatial structures of the model uncertainties and to identify the relationships between the uncertainties on different variables (see for instance Echevin et al., 2000).

Finally, ensemble empirical assessment can be carried out to assess the realisticness of the ensemble.

CONCLUSION GÉNÉRALE

***

Le Golfe du Tonkin est une région peu profonde (profondeur < 60m) du Pacifique sud-ouest, soumise à l'influence des apports d'eau douce du Fleuve Rouge (par la suite FR), d'une forte variabilité atmosphérique (moussons, typhons) et des marées. Le delta du Fleuve Rouge compte une population totale de 22,5 millions d'habitants et une densité de population moyenne de 1060 habitants/km^2, soit près de quatre fois plus que la moyenne nationale du Vietnam. La très forte densité de population exerce une forte pression sur les rivières et leur environnement.

Ce travail s'inscrit dans le cadre du projet LOTUS. LOTUS est un projet conjoint Vietnam-France, qui vise à mieux comprendre la dynamique océanique côtière sous l'influence de la variabilité atmosphérique et du delta du Fleuve Rouge. Dans la thèse de V. Piton (2019) au LEGOS (Toulouse, France), une configuration haute résolution du Golfe du Tonkin et de la partie avale du delta du Fleuve Rouge a été mise en place pour étudier les marées avec le modèle spectral 2D TUGO et la circulation à l'échelle du bassin avec le modèle 3D SYMPHONIE. Dans cette thèse, l'objectif principal est de fournir une meilleure compréhension de la variabilité du panache du FR dans le Golfe du Tonkin à différentes échelles en utilisant une version mise à jour de la configuration avec SYMPHONIE.

Les principales questions scientifiques auxquelles nous avons tenté de répondre sont les suivantes :

1. Comment les simulations numériques à haute résolution peuvent-elles reproduire la dynamique du Golfe du Tonkin et la variabilité du panache du Fleuve Rouge à différentes échelles au cours des années étudiées ?
2. Quels critères peuvent être utilisés pour identifier le panache du Fleuve Rouge à partir des sorties du modèle ?
3. Comment pouvons-nous caractériser formellement la séquence de régimes que le panache du Fleuve Rouge traverse au cours des années d'étude ? Pouvons-nous lier ces régimes à des processus et des forçages spécifiques ?
4. Comment pouvons-nous évaluer la sensibilité des variables du modèle et en particulier celles caractérisant le panache aux principales sources d'incertitude, en particulier les incertitudes du forçage atmosphérique ?

Pour répondre à ces questions, nous avons adopté les approches suivantes basées sur différents outils :

1. Nous avons mis mis à jour et amélioré la configuration de SYMPHONIE créée dans le cadre de la thèse de V. Piton. Une simulation de référence (GOT_REF) est produite dans des conditions de forçage réalistes sur une période de 7 ans (2010-2016). En outre, une simulation jumelle sans forçage fluvial (GOT_NORIV) est mise en place pour évaluer l'impact de l'apport d'eau des rivières sur la circulation côtière et une simulation jumelle

sans marées (GOT_NOTIDE) est réalisée pour évaluer l'impact des marées sur les principaux modes de variabilité du panache fluvial.

2. Le panache de la rivière a été identifié à l'aide de différentes méthodes, ce qui nous a permis de sélectionner la plus appropriée compte tenu des objectifs. La méthode sélectionnée est basée sur des traceurs passifs qui sont injectés dans les rivières.
3. Les régimes du panache du FR ont été identifiés et liés aux processus physiques à l'aide d'une analyse en clusters K-means (KMA).
4. Nous avons réalisé un ensemble de simulations pour évaluer la réponse du modèle aux perturbations du forçage du vent. Nous avons analysé la variabilité des incertitudes résultantes dans l'espace et le temps des variables océaniques. L'approche de clustering K-means a été utilisée sur les membres de l'ensemble, donnant accès à une évaluation de la robustesse des résultats du point 3. ci-dessus.

En utilisant les outils décrits ci-dessus, combinés aux observations, nous avons obtenu les principaux résultats ci-dessous :

1. **Evaluation du modèle**

La simulation GOT_REF est évaluée en la comparant aux observations disponibles : données satellitaires (altimétrie pour la marée, OSTIA pour la température de surface de la mer, SMOS pour la salinité de surface de la mer), données in situ (VITEL, CFO, avec des profils de température et de salinité dans ces deux jeux de données), et radar HF (pour le courant de surface). Nous comparons également nos résultats avec l'OGCM (le modèle de circulation global utilisé aux frontières ouvertes) pour déterminer la valeur ajoutée de notre modèle régional à haute résolution par rapport à un modèle global. Globalement, malgré quelques biais, le modèle reproduit bien les conditions saisonnières, même dans la zone très peu profonde. La comparaison avec les données CTD près de la côte montre une surestimation de la salinité de surface, donc nous soupçonnons que les débits prescrits peuvent être sous-estimés.

Nous discutons également l'échelle de longueur horizontale typique (premier rayon de Rossby barocline) calculée à partir du champ de densité simulé et décrivons la circulation générale.

2. **Identification du panache du Fleuve Rouge**

Pour identifier le panache du FR, différentes méthodes sont envisagées. Tout d'abord, nous appliquons une méthode basée sur l'indice de stratification (SI), en supposant que l'eau du fleuve forme une couche qui flotte au-dessus de l'eau de l'océan. La deuxième méthode consiste à identifier le panache à partir de la salinité de surface de la mer (SSS). Dans la troisième, nous utilisons des traceurs passifs qui se comportent dans le modèle comme des particules sans flottabilité ou des produits chimiques passifs provenant des rivières. Plusieurs expériences ont permis d'identifier le panache fluvial comme la zone où la concentration du traceur dépasse 7

unités/m^3, ce qui correspond le mieux à la zone identifiée par un critère en SSS dans la période de faible débit (saison sèche).

Les cartes du panache montrent qu'au sud de 20°N, le panache du FR est rapidement rejoint par les panaches de trois rivières situées au sud du delta (Ma, Yen, Lam, appelées par la suite "les rivières du sud"), créant un panache flottant unique qui s'étend vers le sud le long de la côte la plupart de l'année. Par conséquent, nous analysons la variabilité du panache résultant du FR et des trois rivières du sud ensemble. Nous calculons d'abord la surface du panache (en km^2) et trouvons que la surface du panache suit les mêmes variations que le débit, bien qu'avec un décalage dans le temps. En raison de la variabilité interannuelle du débit, la surface du panache varie significativement entre les différentes années. Le débit total et la zone du panache subissent également une forte variabilité à des échelles de temps plus courtes de quelques jours.

3. Variabilité du panache du Fleuve Rouge

Les variations temporelles et les régimes du panache en surface

La variabilité saisonnière du panache est examinée à l'aide de l'analyse en K-means et de scènes quotidiennes du panache. Le premier cluster se produit habituellement d'octobre à mars et est le cluster le plus fréquent. Dans ce groupe, la forme du panache est très étroite et le panache est principalement confiné à la zone peu profonde (profondeur < 20 m). La faible différence entre les différentes formes que revêt le panache dans les différentes scènes associées au cluster 1 signifie que la condition de forçage est relativement stable. Ce cluster est associé à la période de débit le plus faible, sous un fort vent de nord-est (mousson d'hiver), qui favorise le downwelling. Le courant de surface côtier est orienté vers le sud et atteint sa plus grande intensité dans le cluster 1, ce qui est cohérent avec le vent. Toutes ces conditions favorisent le confinement des eaux faiblement salées du panache (SSS <30) sur la côte et leur extension vers le sud tout au long de la côte.

Le deuxième cluster apparaît par intermittence tout au long de l'année, avec un taux d'occurrence légèrement plus élevé en avril et en septembre (généralement avant l'occurrence du cluster 3 et après celle du cluster 4), ce qui correspond à la transition saisonnière de la mousson. Ceci suggère que le cluster 2 représente un régime de transition pour le panache. Dans ce cluster, le panache fluvial s'étend à la fois plus au large et plus au sud que dans le cluster 1. Dans le troisième cluster, qui se produit principalement en juin et juillet, le panache a une forme différente de celle des clusters 1 et 2; les panaches des différentes rivières sont disjoints. Le quatrième cluster se produit principalement en août et en septembre, c'est-à-dire au plus fort de la saison de fort débit. Le panache présente alors la plus grande couverture spatiale et s'étend le plus loin au large.

Chaque cluster est associé à différentes conditions de forçage (débit des rivières, vent, courant) qui expliquent leur variabilité spatiale et temporelle.

L'épaisseur du panache.

De septembre à mars, le panache est étroit, étendu le long de la côte et 'attaché au fond' (bottom-attached) en raison des vents de downwelling de la mousson d'hiver. A partir de mai, la mousson

se transforme en vent de sud-ouest, poussant le panache vers le nord et le détachant de la côte et du fond. Dans ce cas, le panache est advecté en surface avec une épaisseur de 2-3 à 7-8 m. Nous observons que les événements où le panache se détache de la côte à 19°N coïncident avec la température à la côte plus froide qu'en mer. Ceci soutient l'hypothèse que les vents de mousson du sud génèrent localement un upwelling côtier. Parfois, en mai-août, le panache s'épaissit au large jusqu'à 12-15 m. Cet approfondissement peut être dû à des vents forts qui augmentent le mélange au large dans la région d'upwelling ou à l'augmentation du débit des rivières.

A la mi-août, nous observons un fort approfondissement du panache sur le plateau (106-106.5°E). Ces événements d'approfondissement semblent être liés à un tourbillon anticyclonique saisonnier récurrent se développant près de 19°N. Dans notre simulation, ce tourbillon se produit généralement en août, lorsque la direction du vent est du sud-ouest, ce qui est favorable à l'upwelling, et que le débit est élevé. Ensuite, ce tourbillon disparaît lorsque le courant côtier vers le sud se développe à nouveau et/ou que le vent n'est plus upwelling-favorable. Le mécanisme de formation de ce tourbillon n'est toujours pas clair. Dans GOT_NORIV, des gradients de vorticité peuvent être représentés, suggérant une faible structure anticyclonique, d'amplitude beaucoup plus faible que dans la simulation de référence et avec une durée de vie plus courte. L'anticyclone semble fortement connecté au panache de la rivière Lam. La topographie du fond peut être un autre facteur de forçage puisque le tourbillon est situé à proximité d'une île.

Impacts des marées.

L'influence des marées sur le panache dans le far-field est étudiée en comparant la simulation de référence GOT_REF avec celle sans marées (GOT_NOTIDE). En moyenne, le panache dans GOT_REF est 5% plus grand que dans GOT_NOTIDE, avec les plus grandes différences trouvées dans la période de haut débit. L'analyse KMA sur la simulation GOT_NOTIDE montre que les 4 clusters sont très proches de ceux de GOT_REF, tant en termes de structures spatiales que de distribution temporelle.

Cependant, les marées ont un impact sur la structure verticale du panache : elles augmentent le mélange vertical et donc la concentration d'eau d'origine fluviale dans la couche inférieure. Elles conduisent également à une plus grande exportation au large de l'eau fluviale à la surface près de l'embouchure pendant la saison de haut débit.

Comme cette étude se concentre sur le far-field, nous n'avons pas examiné en détail l'impact des marées sur les propriétés hydrologiques et sur la dynamique à l'embouchure des rivières, par le biais de la friction du fond induite par les marées. Ceci devrait être fait dans un travail futur.

4. Réponse du modèle due aux perturbations du forçage du vent

En comparant les champs quotidiens ECMWF aux données ASCAT et aux moyennes climatologiques de l'ECMWF, l'incertitude du vent ECMWF dans cette zone est estimée à 60%

de la variabilité du vent. Les perturbations du vent sont ensuite créées comme des combinaisons linéaires d'EOFs bivariés de vent, pondérées à l'aide d'un générateur de nombres pseudo-aléatoires gaussien dont l'écart-type est égal à la valeur de l'incertitude estimée ci-dessus. Un ensemble de simulations SYMPHONIE est réalisé avec 50 membres, perturbés par des paires de perturbations opposées de sorte que la moyenne des perturbations de l'ensemble soit nulle.

L'analyse des ensembles en termes de sensibilité a consisté principalement à calculer l'écart-type de l'ensemble par rapport à la moyenne de l'ensemble (la "dispersion") des variables océaniques et les distributions statistiques des échantillons de l'ensemble.

La dispersion de différentes variables du modèle

Les résultats montrent que la dispersion de la hauteur de la surface de la mer (SSH) est inférieure à 2 cm la plupart du temps. La dispersion moyenne de la SSH sur l'ensemble de la période diminue vers le sud, mais augmente aussi localement près de la côte vietnamienne dans la zone du panache. Le pic de la dispersion en SSH se produit en même temps que le pic principal de la dispersion de la tension de vent. Il s'agit de la periode où l'EOF1 atteint sa plus forte amplitude et correspond à une renverse de la mousson sur quelques jours.

Les courants de surface sont très sensibles aux perturbations du vent, comme on pouvait s'y attendre. En particulier, la forte dispersion le long de la côte de la composante V suggère que la perturbation du vent a un impact sur le courant côtier créé par les eaux fluviales. Le courant côtier peut avoir une direction opposée entre deux membres à des échelles de temps quotidiennes, tandis que la dispersion peut dépasser l'amplitude du courant de la moyenne d'ensemble.

La moyenne spatiale de la dispersion de la température de surface de la mer (SST) montre les mêmes variations temporelles que la SSH. Cependant, contrairement à la SSH, le maximum de dispersion de la SST se produit près du sud de l'île de Hainan. Nous pensons que les perturbations du vent entraînent une modification du mélange (qui est déjà élevé en raison de l'effet de marée dans cette zone) et conduit ensuite au déplacement de la masse d'eau qui présente un fort gradient de température. Sur l'ensemble de la période, l'écart moyen est plus élevé sur la côte Vietnamienne, ce qui peut être lié d'une part aux déplacements du gradient de température entre les eaux fluviales et les eaux océaniques côtières et d'autre part à un upwelling présent dans cette zone. La SSS, comme prévu, présente une forte dispersion près du delta du Fleuve Rouge.

Nous évaluons la structure verticale de la dispersion de la température et de la salinité à partir de sections verticales à 19°N. La dispersion de la température est la plus élevée autour de 20 - 30m de profondeur, là où le gradient vertical est le plus élevé à la base de la couche de mélange (comme dans Andreu-Burillo et al. 2002 et Ayoub et al., 2015). Ceci est une confirmation claire du fait que l'incertitude du vent affecte le mélange vertical. Sur la salinité, l'écart est le plus élevé près de la surface, à proximité du gradient de salinité dû aux eaux des rivières.

Dispersion des caractéristiques du panache.

En surface, la dispersion de la surface de la zone du panache est la plus élevée en août, mais elle ne suit pas la même tendance que la dispersion en SSS. La dispersion est maximale quand

la surface du panache est elle-même maximale, ce qui suggère une sensibilité significative de nos résultats aux incertitudes sur les débits également.

Une analyse en clusters est effectuée sur chaque membre, en prenant pour référence les clusters de GOT_REF, calculés et présentés dans Nguyen-Duy et al. (2021). Les résultats mettent en évidence certains changements de clusters dans l'ensemble, mais presque tout le temps, le cluster le plus susceptible de se produire reste celui comme dans le run de référence.

Enfin, la dispersion de l'épaisseur du panache montre que le panache n'est pas seulement impacté verticalement mais aussi horizontalement par les perturbations en vent. En juin, la dispersion est inférieure à 1m. Le pic de la dispersion se produit du 09 au 13 juillet (> 3m), ce qui correspond à la plus grande dispersion de la tension de vent, confirmant l'impact des perturbations en vent sur le mélange vertical. Le panache peut s'approfondir au centre de tourbillons, comme dans ceux simulés près de 19°N et discutés dans Nguyen-Duy et al. (2021). Les perturbations du vent peuvent modifier le courant, changer la forme du tourbillon et enfin modifier l'épaisseur du panache.

La dispersion d'ensemble du panache est également examinée. pour certains jours spécifiques. La distribution sur l'ensemble du coefficient de mélange vertical met en évidence l'augmentation du mélange dans les membres avec un vent plus fort. Le mélange peut à son tour conduire à une diminution de la concentration du traceur à la surface ; si la concentration diminue en dessous du seuil choisi pour identifier le panache, alors le panache n'est plus considéré comme présent, alors qu'une masse d'eau plus épaisse est influencée par l'apport fluvial. Ceci suggère que l'utilisation de deux conditions pour identifier le panache (l'une basée sur la concentration en surface et l'autre sur l'épaisseur de la colonne d'eau impactée par les eaux fluviales) pourrait être plus robuste aux incertitudes du forçage en vent.

Des vents plus forts favorisent également une extension du panache plus au large. Cependant, il n'est pas simple de démêler l'effet du vent sur le mélange vertical et sur l'advection. Premièrement, parce que le transport horizontal dépend aussi de la flottabilité du panache, au moins dans le mid-field (Horner-Devine et al., 2015). Deuxièmement, parce que les conditions océaniques à un jour donné résultent de l'effet cumulatif du forçage sur une " certaine " période antérieure (peut-être plusieurs jours). Cela devrait être pris en compte lors de la comparaison des propriétés du panache entre différents membres à une date donnée, mais comme la variabilité quotidienne du vent est assez élevée dans le GOT, une telle approche est assez complexe à mettre en oeuvre.

Finalement, en conclusion, nous estimons que dans nos conditions expérimentales, les résultats de Nguyen-Duy et al. (2021) ne sont pas remis en cause et sont robustes par rapport aux incertitudes sur le vent.

REFERENCES

Anderberg, M. R. (Ed.). (1973). Probability and mathematical statistics: A series of monographs and textbooks. In Cluster Analysis for Applications. Academic Press. doi:10.1016/B978-0-12-057650-0.50027-2

Andreu-Burillo, I., G. Caniaux, M. Gavart, P. De Mey & R. Baraille (2002). Assessing ocean-model sensitivity to wind forcing uncertainties. Geophys. Res. Letters, 29(18), 5.1-5.4.

Arakawa, A., and M. J. Suarez, Vertical differencing of the primitive equations in sigma coordinates, Monthly Weather Review, 111, 34–45, 1983.

Auclair, F., Marsaleix, P., & De Mey, P. (2003). Space-time structure and dynamics of the forecast error in a coastal circulation model of the Gulf of Lions. Dynamics of Atmospheres and Oceans, 36(4), 309–346. doi:10.1016/S0377-0265(02)00068-4

Ayoub, N. K., Lucas, M., & De Mey, P. (2015). Estimating uncertainties on a Gulf Stream mixed-layer heat budget from stochastic modeling. Journal of Marine Systems, 150, 66–79. doi:10.1016/j.jmarsys.2015.04.008

Ayouche, A., Charria, G., Carton, X., Ayoub, N.K., Theetten, S., (2021). Non-Linear Processes in the Gironde River Plume (North-East Atlantic): Instabilities and Mixing. Front. Mar. Sci. 8, 810. doi:10.3389/fmars.2021.701773

Bentsen, M., Evensen, G., Drange, H., Jenkins, A.D., 1999. Coordinate transformation on a sphere using conformal mapping. Mon. Weather Rev. 127, 2733–2740.

Blumberg, A.F. and Mellor, G.L. (1987). A Description of a Three–Dimensional Coastal Ocean Circulation Model. In Three–Dimensional Coastal Ocean Models, N.S. Heaps (Ed.). doi:10.1029/CO004p0001

Brando, V. E., Braga, F., Zaggia, L., Giardino, C., Bresciani, M., Matta, E., Bellafiore, D., Ferrarin, C., Maicu, F., Benetazzo, A., Bonaldo, D., Falcieri, F. M., Coluccelli, A., Russo, A., & Carniel, S. (2015). High-resolution satellite turbidity and sea surface temperature observations of river plume interactions during a significant flood event. Ocean Science, 11(6), 909–920. doi:10.5194/os-11-909-2015

Burchard H., Bolding L. (2001) Comparative analysis of four second-moment turbulence closing models for the oceanic mixed layer, J. of Physical Oceanogr.,31 (8), 1943-1968, doi:10.1175/1520-0485(2001)031<1943:CAOFSM>2.0.CO;2

Burla, M., Baptista, A. M., Zhang, Y., & Frolov, S. (2010). Seasonal and interannual variability of the Columbia River plume: A perspective enabled by multiyear simulation databases. Journal of Geophysical Research, 115(C2). doi:10.1029/2008JC004964

Cai, S., Long, X., Liu, H., & Wang, S. (2006). Tide model evaluation under different conditions. Continental Shelf Research, 26(1), 104–112. doi:10.1016/j.csr.2005.09.004

Canuto, V. M., Howard, A., Cheng, Y., & Dubovikov, M. S. (2001). Ocean Turbulence. Part I: One–Point Closure Model—Momentum and Heat Vertical Diffusivities. Journal of Physical Oceanography, 31(6), 1413–1426. doi:10.1175/1520-0485(2001)031<1413:OTPIOP>2.0.CO;2

Chelton, D. B., deSzoeke, R. A., Schlax, M. G., El Naggar, K., & Siwertz, N. (1998). Geographical Variability of the First Baroclinic Rossby Radius of Deformation, Journal of Physical Oceanography, 28(3), 433–460.

Chen, C., Li, P., Shi, M., Zuo, J., Chen, M., & Sun, H. (2009). Numerical study of the tides and residual currents in the Qiongzhou Strait. Chinese Journal of Oceanology and Limnology, 27(4), 931–942. doi:10.1007/s00343-009-9193-0

Chen, B., Xu, Z., Ya, H., Chen, X., & Xu, M. (2019). Impact of the water input from the eastern Qiongzhou Strait to the Beibu Gulf on Guangxi coastal circulation. Acta Oceanologica Sinica, 38(9), 1–11. doi:10.1007/s13131–019–1472–2

Chen, Z., Gong, W., Cai, H., Chen, Y., & Zhang, H. (2017). Dispersal of the Pearl River plume over continental shelf in summer. Estuarine, Coastal and Shelf Science, 194, 252–262. doi:10.1016/j.ecss.2017.06.025

Craig P. D., Banner M. L. (1994) Modeling wave-enhanced turbulence in the ocean surface layer. J. of Phys. Oceanogr., 24(12), 2546-2559, doi:10.1175/1520-0485(1994)024<2546:MWETIT>2.0CO;2.

Csanady, G.T. (1982) Circulation in the Coastal Ocean, Springer, 281 pages, doi: 10.1007/978-94-017-1041-1

Damien, P., Bosse, A., Testor, P., Marsaleix, P., and Estournel, C. (2017), Modeling Postconvective Submesoscale Coherent Vortices in the Northwestern Mediterranean Sea, J. Geophys. Res. Oceans, 122, 9937– 9961, doi:10.1002/2016JC012114.

Dang, T. H. et al. (2010) 'Long-term monitoring (1960-2008) of the river-sediment transport in the Red River Watershed (Vietnam): Temporal variability and dam-reservoir impact', Science of the Total Environment. Elsevier, 408(20), pp. 4654–4664. doi:10.1016/j.scitotenv.2010.07.007.

Delaval, A., Duffa, C., Pairaud, I., & Radakovitch, O. (2021). A fuzzy classification of the hydrodynamic forcings of the Rhone River plume: An application in case of accidental release of radionuclides. Environmental Modelling & Software, 140, 105005. doi:10.1016/j.envsoft.2021.105005

Ding, Y., Chen, C., Beardsley, R. C., Bao, X., Shi, M., Zhang, Y., Lai, Z., Li, R., Lin, H., & Viet, N. T. (2013). Observational and model studies of the circulation in the Gulf of Tonkin, South China Sea. Journal of Geophysical Research: Oceans, 118(12), 6495–6510. doi:10.1002/2013JC009455

Doodson A. T. (1927) The analysis of tidal observations. Philosophical Transactions of the Royal Society of London, 227, 223-279.

Echevin V., De Mey P. and G. Evensen, . 2000. Horizontal and vertical structure of the representer functions for sea surface measurements in a coastal circulation model. Journal of Physical Oceanography 30: 2627–2635

EUMETSAT, OSI SAF 2021. ASCAT Wind Product User Manual, SAF/OSI/CDOP/KNMI/TEC/MA/126, Version 1.17.l

Evensen G., 2003: The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dynamics, 53, 343–367, doi:10.1007/s10236-003-0036-9.

Falcieri, F.M., Benetazzo, A., Sclavo, M., Russo, A., and Carniel, S. (2014) Po River plume pattern variability investigated from model data. Cont. Shelf Res., 87, 84-95, doi:10.1016/j.csr.2013.11.001

Fang, G., Kwok, Y.-K., Yu, K., & Zhu, Y. (1999). Numerical simulation of principal tidal constituents in the south china sea, gulf of tonkin and gulf of thailand. Continental Shelf Research, 19(7), 845–869. doi:10.1016/S0278-4343(99)00002-3

Fong, D. A., & Geyer, W. R. (2002). The alongshore transport of freshwater in a surface-trapped river plume*. Journal of Physical Oceanography, 32(3), 957–972. doi:10.1175/1520-0485(2002)032<0957:TATOFI>2.0.CO;2

Frias, J. P. G. L., & Nash, R. (2019). Microplastics: Finding a consensus on the definition. Marine Pollution Bulletin, 138, 145–147. doi:10.1016/j.marpolbul.2018.11.022

Fu, L.-L., Alsdorf, D., Morrow, R., Rodriguez, E., & Mognard, N. (2012). SWOT: The Surface Water and Ocean Topography Mission: Wide-Swath Altimetric Measurement of Water Elevation on Earth.

Gao, J., Xue, H., Chai, F., & Shi, M. (2013). Modeling the circulation in the Gulf of Tonkin, South China Sea Topical Collection on the 4th International Workshop on Modelling the Ocean in Yokohama, Japan 21–24 May 2012. Ocean Dynamics, 63(8), 979–993. doi:10.1007/s10236–013–0636–y

Gao, J., Shi, M., Chen, B., Guo, P., & Zhao, D. (2014). Responses of the circulation and water mass in the Beibu Gulf to the seasonal forcing regimes. Acta Oceanologica Sinica, 33(7), 1–11. doi:10.1007/s13131–014–0506–6

Gao, J., Zhijun, D., Xuefei, M., Zhenpeng, G., Wen, W., Hualiang, X, Shushi, L. (2015). Interference of natural and anthropogenic forcings on variations in continental freshwater runoff from the Red River (Vietnam) to sea, Quaternary International, Volumes 380–381, 2015, Pages 133–142, ISSN 1040–6182, doi:10.1016/j.quaint.2015.01.007.

Garvine, Richard W. (1995), A dynamical system for classifying buoyant coastal runoffs, Continental Shelf Research, Volume 15, Issue 13, 1995, Pages 1585–1596, ISSN 0278–4343

Gaspar P., Gregoris Y., Lefevre J.M. (1990) A simple eddy kinetic energy model for simulations of the oceanic vertical mixing: tests at station Papa and long-term upper ocean study site. J.of Geophys. Res., 95,16179–16193.

Ghantous, M., Ayoub, N.K., De Mey-Frémaux, P., Vervatis, V., & Marsaleix, P. (2020). Ensemble downscaling of a regional ocean model. Ocean Modelling, 145, 101511. doi:10.1016/j.ocemod.2019.101511

Gong, W., Chen, L., Chen, Z., & Zhang, H. (2019). Plume-to-plume interactions in the Pearl River Delta in winter. Ocean & Coastal Management, 175, 110–126. doi:10.1016/j.ocecoaman.2019.04.001

Greenberg, David A., Frédéric Dupont, Florent H. Lyard, Daniel R. Lynch, Francisco E. Werner (2007), Resolution issues in numerical models of oceanic and coastal circulation, Continental Shelf Research, Volume 27, Issue 9, 2007, Pages 1317–1343, ISSN 0278–4343, doi:10.1016/j.csr.2007.01.023

Guanche, Y., Mínguez, R., & Méndez, F. J. (2014). Autoregressive logistic regression applied to atmospheric circulation patterns. Climate Dynamics, 42(1), 537–552. doi:10.1007/s00382-013-1690-3

Guarnieri, A., Pinardi, N., Oddo, P., Bortoluzzi, G., and Ravaioli, M. (2013), Impact of tides in a baroclinic circulation model of the Adriatic Sea, J. Geophys. Res. Oceans, 118, 166–183, doi:10.1029/2012JC007921.

Hande, L. B., Siems, S. T., & Manton, M. J. (2012). Observed trends in wind speed over the southern ocean. Geophysical Research Letters, 39(11). doi:10.1029/2012GL051734

Hänninen, J., Weckström, M., Pawłowska, J., Szymańska, N., Uurasjärvi, E., Zajaczkowski, M., Hartikainen, S., & Vuorinen, I. (2021). Plastic debris composition and concentration in the arctic ocean, the north sea and the baltic sea. Marine Pollution Bulletin, 165, 112150. doi:10.1016/j.marpolbul.2021.112150

Hansen, Donald V. Rattray, Maurice, (1966), New dimensions in estuary classification, Limnology and Oceanography, 11, doi: 10.4319/lo.1966.11.3.0319.

Hastie, T., Tsbshirani R., Friedman J., (2001). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 533 pp.

Hendershott, M. C., The Effects of Solid Earth Deformation on Global Ocean Tides, Geophysical Journal International, Volume 29, Issue 4, October 1972, Pages 389–402, doi:10.1111/j.1365-246X.1972.tb06167.x

Herrmann, M., Nguyen-Duy, T., Ngo-Duc, T., & Tangang, F. (2021). Climate change impact on sea surface winds in Southeast Asia. International Journal of Climatology, joc.7433. doi:10.1002/joc.7433

Hetland, R. D. (2005). Relating river plume structure to vertical mixing. Journal of Physical Oceanography, 35(9), 1667–1688. doi:10.1175/JPO2774.1

Hetland, R. D. (2010). The effects of mixing and spreading on density in near-field river plumes. Dynamics of Atmospheres and Oceans, 49(1), 37–53. doi:10.1016/j.dynatmoce.2008.11.003

Hetland, RD, Hsu T–J., (2013). Freshwater and sediment dispersal in large river plumes. In Biogeochemical Dynamics at Large River–Coastal Interfaces: Linkages with Global Climate Change, ed. TS Bianchi, MA Allison, W–J Cai, pp. 55–85. New York: Springer

Hopkins, J., Lucas, M., Dufau, C., Sutton, M., Stum, J., Lauret, O., & Channelliere, C. (2013). Detection and variability of the Congo River plume from satellite derived sea surface temperature, salinity, ocean colour and sea level. Remote Sensing of Environment, 139, 365–385. doi:10.1016/j.rse.2013.08.015

Horner-Devine, A. R., Fong, D. A., Monismith, S. G., & Maxworthy, T. (2006). Laboratory experiments simulating a coastal river inflow. Journal of Fluid Mechanics, 555, 203. doi:10.1017/S0022112006008937

Horner-Devine, A. R., Jay, D. A., Orton, P. M., & Spahn, E. Y. (2009). A conceptual model of the strongly tidal Columbia River plume. Journal of Marine Systems, 78(3), 460–475. doi:10.1016/j.jmarsys.2008.11.025

Horner-Devine, A. R., Hetland, R. D., & MacDonald, D. G. (2015). Mixing and transport in coastal river plumes. Annual Review of Fluid Mechanics, 47(1), 569–594. doi:10.1146/annurev-fluid-010313-141408

Huq, P. (2009). The role of kelvin number on bulge formation from estuarine buoyant outflows. Estuaries and Coasts, 32(4), 709–719. doi:10.1007/s12237-009-9162-z

IOCCG (2000). Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex, Waters. Sathyendranath, S. (ed.), Reports of the International Ocean-Colour Coordinating Group, No. 3, IOCCG, Dartmouth, Canada.

Jamet, C. and Dessailly, D. (2016) GlobCoast Monthly mean of diffuse attenuation coefficient (Kd). doi:10.12770/2b192671-1129-409c-adae-e2335759f290.

Jamet, C., Loisel, H., & Dessailly, D. (2012). Retrieval of the spectral diffuse attenuation coefficient Kd(Λ) in open and coastal ocean waters using a neural network inversion. Journal of Geophysical Research: Oceans, 117(C10). doi:10.1029/2012JC008076

Jerlov, N. G. (1968) Optical oceanography, Limnology and Oceanography. Elsevier. doi:10.4319/lo.1968.13.4.0731.

Jia, Y., & Whitney, M. M. (2019). Summertime Connecticut river water pathways and wind impacts. Journal of Geophysical Research: Oceans, 124(3), 1897–1914. doi:10.1029/2018JC014486

Jordà, G., & De Mey, P. (2010). Characterization of error dynamics in a 3D coastal model of the Catalan sea using stochastic modelling. Continental Shelf Research, 30(5), 419–441. doi:10.1016/j.csr.2009.12.013

Kim S., Samelson R.M., Snyder C., (2011). Toward an uncertainty budget for a coastal ocean model, Mon. Wea. Rev., 139, doi: 10.1175/2010MWR3352.1

Kirtman, B., D. Min, J. Infanti, J. Kinter, D. Paolino, Q. Zhang, H. van den Dool, S. Saha, M. Mendez, E. Becker, P. Peng, P. Tripp, J. Huang, D. DeWitt, M. Tippett, A. Barnston, S. Li, A. Rosati, S. Schubert, M. Rienecker, M. Suarez, Z. Li, J. Marshak, Y. Lim, J. Tribbia, K. Pegion, W. Merryfield, B. Denis, and E. Wood, (2014): The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. Bull. Amer. Meteor. Soc., 95, 585–601, doi:10.1175/BAMS-D-12-00050.1.

Kraus, E. B. and Businger, J. A. (1994) Atmosphere-ocean interaction. Oxford University Press.

Kourafalou, V. H., Lee, T. N., Oey, L.–Y., and Wang, J. D. (1996), The fate of river runoff on the continental shelf: 2. Transport of coastal low–salinity waters under realistic wind and tidal forcing, J. Geophys. Res., 101(C2), 3435– 3455, doi:10.1029/95JC03025.

Large, W. G., & Yeager, S. (2004). Diurnal to decadal global forcing for ocean and sea–ice models: The data sets and flux climatologies. doi:10.5065/D6KK98Q6

Le Hénaff, M., De Mey, P., Marsaleix, P., (2009). Assessment of observational networks with the Representer Matrix Spectra method-application to a 3D coastal model of the Bay of Biscay. Ocean Dyn. 59, 3-20

Leonard, B.P. (1979). A stable and accurate convection modelling procedure based on quadratic upstream interpolation, Comput. Meth. Appl. Mech. Eng. 19, 59 (1979).

Ling, T., Xu, M., Liang, X.-Z., Wang, J. X. L., & Noh, Y. (2015). A multilevel ocean mixed layer model resolving the diurnal cycle: Development and validation: MULTILEVEL OCEAN MIXED LAYER MODEL. Journal of Advances in Modeling Earth Systems, 7(4), 1680–1692. https://doi.org/10.1002/2015MS000476

Liu, Y., MacCready, P., & Hickey, B. M. (2009). Columbia River plume patterns in summer 2004 as revealed by a hindcast coastal ocean circulation model. Geophysical Research Letters, 36(2), 1–6. doi:10.1029/2008GL036447

Lu, X. X., Oeurng, C., Le, T. P. Q., & Thuy, D. T. (2015). Sediment budget as affected by construction of a sequence of dams in the lower Red River, Viet Nam. Geomorphology, 248, 125–133. doi:10.1016/j.geomorph.2015.06.044

Lucas M., N. Ayoub, B. Barnier, T. Penduff and De Mey, P., 2008. Stochastic study of the temperature response of the upper ocean to uncertainties in the atmospheric forcing in an Atlantic OGCM, Ocean Modell., Vol.20, 1, 90-113, doi: 10.1016/j.ocemod.2007.07.006.

Luu, T. N. M., Garnier, J., Billen, G., Orange, D., Némery, J., Le, T. P. Q., Tran, H. T., & Le, L. A. (2010). Hydrological regime and water budget of the Red River Delta (Northern Vietnam). Journal of Asian Earth Sciences, 37(3), 219–228. doi:10.1016/j.jseaes.2009.08.004

Lyard, F. H., Allain, D. J., Cancet, M., Carrère, L., and Picot, N. (2021): FES2014 global ocean tide atlas: design and performance, Ocean Sci., 17, 615–649, doi:10.5194/os–17–615–2021

MacCready, P., Banas, N. S., Hickey, B. M., Dever, E. P., & Liu, Y. (2009). A model study of tide– and wind–induced mixing in the Columbia River Estuary and plume. Continental Shelf Research, 29(1), 278–291. doi:10.1016/j.csr.2008.03.015

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, 5.1, 281–298

Maraldi, C. et al. (2013) NEMO on the shelf: Assessment of the Iberia-Biscay-Ireland configuration, Ocean Science, 9(4), pp. 745–771. doi:10.5194/os-9-745-2013.

Marsaleix, P., F. Auclair, and C. Estournel. " Considerations on Open Boundary Conditions for Regional and Coastal Ocean Models", Journal of Atmospheric and Oceanic Technology 23, 11 (2006): 1604–1613, doi:10.1175/JTECH1930.1

Marsaleix, P., Francis Auclair, Jochem Willem Floor, Marine Julie Herrmann, Claude Estournel, Ivane Pairaud, Caroline Ulses, Energy conservation issues in sigma–coordinate free–surface ocean models, Ocean Modeling, Volume 20, Issue 1, 2008, Pages 61–89, doi:10.1016/j.ocemod.2007.07.005.

Marsaleix, P., Michaud, H., Estournel, C., 2019. 3D phase-resolved wave modelling with a non-hydrostatic ocean circulation model. Ocean Modelling, 136, 28-50. doi:10.1016/j.ocemod.2019.02.002

Menesguen, A., Dussauze M., Dumas F. (2018). Designing optimal scenarios of nutrient loading reduction in a WFD/MSFD perspective by using passive tracers in a biogeochemical–3D model of the English Channel/Bay of Biscay area. Ocean & Coastal Management, 163, 37–53. doi:10.1016/j.ocecoaman.2018.06.005

Michaud, H., Marsaleix, P., Leredde, Y., Estournel, C., Bourrin, F., Lyard, F., Mayet, C., & Ardhuin, F. (2012). Three–dimensional modelling of wave–induced current from the surf zone to the inner shelf. Ocean Science, 8(4), 657–681. doi:10.5194/os–8–657–2012

Morel, A., Huot, Y., Gentili, B., Werdell, P. J., Hooker, S. B., & Franz, B. A. (2007). Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach. Remote Sensing of Environment, 111(1), 69–88. doi:10.1016/j.rse.2007.03.012

Murtugudde, R. et al. (2002) ‘Effects of penetrative radiation of the upper tropical ocean circulation’, Journal of Climate, 15(5), pp. 470–486. doi:10.1175/1520-0442(2002)015<0470:EOPROT>2.0.CO;2.

Neumann, L.E., Šimůnek J., Cook, F.J. (2011). Implementation of quadratic upstream interpolation schemes for solute transport into HYDRUS–1D, Environmental Modelling & Software, Vol.26, 11, doi:10.1016/j.envsoft.2011.05.010.

Nguyen M. N., Marchesiello P., Lyard F., Ouillon S.,Cambon G., Allain D. & Van Uu, D. (2014). Tidal characteristics of the Gulf of Tonkin. Continental Shelf Research, 91, 37–56. doi:10.1016/j.csr.2014.08.003

Nguyen-Duy T., Ayoub N.K., Marsaleix P., Toublanc F., De Mey-Frémaux P., Piton V., Herrmann M., Duhaut T., Tran M.C. and Ngo-Duc T. (2021) Variability of the Red River Plume in the Gulf of Tonkin as Revealed by Numerical Modeling and Clustering Analysis. Front. Mar. Sci. 8:772139. doi:10.3389/fmars.2021.772139

Nguyen Thi Hien, A., Vu Minh Cat, B., & Roberto Ranzi, C. (2019). Upstream effects on salinity dynamics in the Red River Delta. In APAC 2019 (pp. 1439–1443). Springer Singapore. doi:10.1007/978-981-15-0291-0_194

Nugroho, D. (2017). The Tides in a general circulation model in the Indonesian Seas. PhD manuscript, Univ. Toulouse 3 Paul Sabatier (https://tel.archives-ouvertes.fr/tel-01556796v2)

Otero, P., Ruiz–Villarreal, M., & Peliz, A. (2008). Variability of river plumes off Northwest Iberia in response to wind events. Journal of Marine Systems, 72(1–4), 238–255. doi:10.1016/j.jmarsys.2007.05.016

Ouillon, S. (2014). VITEL–NORTH cruise, RV Alis, doi:10.17600/14004100.

Pasmans I, Kurapov A.L., (2019) Ensemble of 4DVARs (En4DVar) data assimilation in a coastal ocean circulation model, Part I: Methodology and ensemble statistics, Ocean Modelling, Volume 144, doi:10.1016/j.ocemod.2019.101493.

Pairaud, I. L., Lyard F., Auclair F., Letellier T., Marsaleix P. (2008). Dynamics of the semi–diurnal and quarter–diurnal internal tides in the Bay of Biscay. Part 1: Barotropic tides, Continental Shelf Research, 28, 1294–1315. doi:10.1016/j.csr.2008.03.004

Paulson, C. A. and Simpson, J. J. (1977) 'Irradiance Measurements in the Upper Ocean', Journal of Physical Oceanography, 7(6), pp. 952–956. doi:10.1175/1520-0485(1977)007<0952:imituo>2.0.co;2.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit–learn: Machine Learning in Python. Journal of Machine Learning Research, 12.

Phan, H. M., Ye, Q., Reniers, A. J., & Stive, M. J. (2019). Tidal wave propagation along The Mekong deltaic coast. Estuarine, Coastal and Shelf Science, 220, 73-98. doi:10.1016/j.ecss.2019.01.026

Pierson, D. C. et al. (2008) 'Relationship between the attenuation of downwelling irradiance at 490 nm with the attenuation of PAR (400 nm-700 nm) in the Baltic Sea', Remote Sensing of Environment. Elsevier, 112(3), pp. 668–680. doi:10.1016/j.rse.2007.06.009.

Piton (2019), Du Fleuve Rouge au Golfe du Tonkin: dynamique et transport sédimentaire le long du continuum estuaire-zone côtière, PhD thesis, University of Toulouse III, France.

Piton, V., Herrmann, M., Lyard, F., Marsaleix, P., Duhaut, T., Allain, D., & Ouillon, S. (2020). Sensitivity study on the main tidal constituents of the Gulf of Tonkin by using the

frequency-domain tidal solver in T-UGOm. Geoscientific Model Development, 13(3), 1583–1607. doi:10.5194/gmd-13-1583-2020

Piton, V., Herrmann, M., Marsaleix, P., Duhaut, T., Ngoc, T. B., Tran, M. C., Shearman, K., & Ouillon, S. (2021). Influence of winds, geostrophy and typhoons on the seasonal variability of the circulation in the Gulf of Tonkin: A high–resolution 3D regional modeling study. Regional Studies in Marine Science, 45, 101849. doi:10.1016/j.rsma.2021.101849

Prime Minister (2013), Decision No 795/QĐ-TTg on Approving the master plan for socio-economic development in the Red River Delta region to 2020 (in Vietnamese).

Reffray, G., Fraunié P. and Marsaleix P., (2004). Secondary flows induced by wind forcing in the Rhône region of freshwater influence. Ocean Dynamics, 54, 179–196. doi :10.1007/s10236–003–0079–y

Rogowski, P., Zavala–Garay, J., Shearman, K., Terrill, E., Wilkin, J., & Tran, H. L. (2019). Air–Sea–Land Forcing in the Gulf of Tonkin: Assessing Seasonal Variability Using Modern Tools. Oceanography, 32, 150–161. doi:10.5670/oceanog.2019.223

Santana-Falcón, Y., Brasseur, P., Brankart, J. M., & Garnier, F. (2020). Assimilation of chlorophyll data into a stochastic ensemble simulation for the North Atlantic Ocean. Ocean Science, 16(5), 1297–1315. https://doi.org/10.5194/os-16-1297-2020

Saulquin, B. et al. (2013) Estimation of the diffuse attenuation coefficient K dPAR using MERIS and application to seabed habitat mapping, Remote Sensing of Environment., 128, pp. 224–233. doi:10.1016/j.rse.2012.10.002.

Shi, M., Chen, C., Xu, Q., Lin, H., Liu, G., Wang, H., Wang, F., & Yan, J. (2002). The role of Qiongzhou Strait in the seasonal variation of the South China Sea circulation. Journal of Physical Oceanography, 32(1), 103–121. doi:10.1175/1520–0485(2002)032<0103:TROQSI>2.0.CO;2

Solabarrieta, L., Rubio, A., Cárdenas, M., Castanedo, S., Esnaola, G., Méndez, F. J., Medina, R., & Ferrer, L. (2015). Probabilistic relationships between wind and surface water circulation patterns in the SE Bay of Biscay. Ocean Dynamics, 65(9–10), 1289–1303. doi:10.1007/s10236–015–0871–5

Sonnewald, M., Wunsch, C., & Heimbach, P. (2019). Unsupervised learning reveals geography of global ocean dynamical regions. Earth and Space Science, 6, 784– 794. doi:10.1029/2018EA000519

Strady, E., Dang, T. H., Dao, T. D., Dinh, H. N., Do, T. T. D., Duong, T. N., Duong, T. T., Hoang, D. A., Kieu-Le, T. C., Le, T. P. Q., Mai, H., Trinh, D. M., Nguyen, Q. H., Tran-Nguyen, Q. A., Tran, Q. V., Truong, T. N. S., Chu, V. H., & Vo, V. C. (2021). Baseline assessment of microplastic concentrations in marine and freshwater environments of a developing Southeast Asian country, Viet Nam. Marine Pollution Bulletin, 162, 111870. doi:10.1016/j.marpolbul.2020.111870

Sutton, W.R., Srivastava J.P., Rosegrant M., Thurlow J., Sebastian L., (2019) Striking a balance, Managing El Niño and la Niña in Vietnam agriculture, RepNo 132068, World Bank, Washington D.C.

Sweeney, C., Gnanadesikan, A., Griffies, S. M., Harrison, M. J., Rosati, A. J., & Samuels, B. L. (2005). Impacts of shortwave penetration depth on large-scale ocean circulation and

heat transport. Journal of Physical Oceanography, 35(6), 1103–1119. doi:10.1175/JPO2740.1

Szabo, S., Brondizio, E., Renaud, F.G. et al. Population dynamics, delta vulnerability and environmental change: comparison of the Mekong, Ganges–Brahmaputra and Amazon delta regions. Sustain Sci 11, 539–554 (2016). https://doi.org/10.1007/s11625-016-0372-6

Talagrand, O., Vautard R., and Strauss B., 1997. Evaluation of probabilistic prediction systems. Proceedings, ECMWF Workshop on Predictability. ECMWF, 1–25.

Toublanc, F., Ayoub, N. K., Lyard, F., Marsaleix, P., & Allain, D. J. (2018). Tidal downscaling from the open ocean to the coast: A new approach applied to the Bay of Biscay. Ocean Modelling, 124, 16–32. doi:10.1016/j.ocemod.2018.02.001

Tran, M. C., Sentchev, A., & Nguyen, K. C. (2021). Multi–scale variability of circulation in the Gulf of Tonkin from remote sensing of surface currents by high–frequency radars. Ocean Dynamics, 71(2), 175–194. doi:10.1007/s10236–020–01440–x

Vandenbulcke, L., & Barth, A. (2015). A stochastic operational forecasting system of the Black Sea: Technique and validation. Ocean Modelling, 93, 7–21. doi:10.1016/j.ocemod.2015.07.010

van Maren, D. S. (2007). Water and sediment dynamics in the Red River mouth and adjacent coastal zone. Journal of Asian Earth Sciences, 29(4), 508–522. https://doi.org/10.1016/j.jseaes.2006.03.012

van Velzen, N., Altaf, M. U., & Verlaan, M. (2016). OpenDA-NEMO framework for ocean data assimilation. Ocean Dynamics, 66(5), 691–702. https://doi.org/10.1007/s10236-016-0945-z

Vaz, Nuno, João G. Rodrigues, Marcos Mateus, Guilherme Franz, Francisco Campuzano, Ramiro Neves, and João Miguel Dias. (2018). “Subtidal Variability of the Tagus River Plume in Winter 2013.” Science of the Total Environment 627: 1353–62. doi:10.1016/j.scitotenv.2018.01.325.

Vervatis, V. D., De Mey-Frémaux, P., Ayoub, N., Karagiorgos, J., Ghantous, M., Kailas, M., Testut, C.-E., & Sofianos, S. (2021a). Assessment of a regional physical–biogeochemical stochastic ocean model. Part 1: Ensemble generation. Ocean Modelling, 160, 101781. doi:10.1016/j.ocemod.2021.101781

Vervatis, V. D., De Mey-Frémaux, P., Ayoub, N., Karagiorgos, J., Ciavatta, S., Brewin, R. J. W., & Sofianos, S. (2021b). Assessment of a regional physical–biogeochemical stochastic ocean model. Part 2: Empirical consistency. Ocean Modelling, 160, 101770. doi:10.1016/j.ocemod.2021.101770

Vietnam General Statistic Office (2020), Completed result of the 2019 Viet Nam population and housing census, Statistical Publishing House.

Vinh, V. D., Ouillon, S., Thanh, T. D., & Chu, L. V. (2014). Impact of the Hoa Binh dam (Vietnam) on water and sediment budgets in the Red River basin and delta. Hydrology and Earth System Sciences, 18(10), 3987–4005. doi:10.5194/hess–18–3987–2014

Vinh, V.D (2018) Suspended sediment dynamics in Red River distributaries and along the Red River delta: focus on estuarine processes and recent balances. Ph D thesis, University of Science and Technology of Hanoi.

Vlasenko, V., Stashchuk, N., & McEwan, R. (2013). High–resolution modelling of a large–scale river plume. Ocean Dynamics, 63(11–12), 1307–1320. doi:10.1007/s10236–013–0653–x

Wang, M., Son, S. and Harding Jr., L. W. (2009) 'Retrieval of diffuse attenuation coefficient in the Chesapeake Bay and turbid ocean regions for satellite ocean color applications', Journal of Geophysical Research: Oceans, 114(C10). doi:10.1029/2009JC005286.

Wang, Y., Xue, H., Chai, F., Chao, Y., & Farrara, J. (2014). A model study of the Copper River plume and its effects on the northern Gulf of Alaska. Ocean Dynamics, 64(2), 241–258. doi:10.1007/s10236-013-0684-3

Warrick, J. A., & Farnsworth, K. L. (2017). Coastal river plumes: Collisions and coalescence. Progress in Oceanography, 151, 245–260. doi:10.1016/j.pocean.2016.11.008

Web portal of Bac Giang (2016), Water resources, https://vietyen.bacgiang.gov.vn/xuat-ban-thong-tin/-/asset_publisher/vYGFBWdWN3jE/content/tai-nguyen-nuoc , accessed on 10 Feb 2022 (in Vietnamese)

Wei, X., Sauvage, S., Ouillon, S., Le, T. P. Q., Orange, D., Herrmann, M., & Sanchez-Perez, J.-M. (2021). A modelling-based assessment of suspended sediment transport related to new damming in the Red River basin from 2000 to 2013. CATENA, 197, 104958. doi:10.1016/j.catena.2020.104958

Weidle, F., Wang, Y., Tian, W., & Wang, T. (2013). Validation of strategies using clustering analysis of ecmwf eps for initial perturbations in a limited area model ensemble prediction system. Atmosphere-Ocean, 51(3), 284–295. doi:10.1080/07055900.2013.802217

Wilks, D. (2005), Statistical Methods in the Atmospheric Sciences, Academic Press.

Wiseman, W.J., Garvine, R.W. (1995). Plumes and coastal currents near large river mouths. Estuaries 18, 509. doi:10.2307/1352368

Wu, D., Wang, Y., Lin, X., & Yang, J. (2008). On the mechanism of the cyclonic circulation in the Gulf of Tonkin in the summer. Journal of Geophysical Research: Oceans, 113(9), 2–11. doi:10.1029/2007JC004208

Wyrtki K., 1961: Scientific Results of marine Investigations of the South China Sea and the Gulf of Thailand 1959-1961. NAGA report – Volume 2.

Yankovsky, A. E., & Chapman, D. C. (1997). A Simple Theory for the Fate of Buoyant Coastal Runoffs, Journal of Physical Oceanography, 27(7), 1386–1401.

Yan, Y., Barth, A., Beckers, J. M., Candille, G., Brankart, J. M., & Brasseur, P. (2015). Ensemble assimilation of ARGO temperature profile, sea surface temperature, and altimetric satellite data into an eddy permitting primitive equation model of the North Atlantic Ocean: ENSEMBLE ASSIMILATION TO AN OCEAN MODEL. Journal of Geophysical Research: Oceans, 120(7), 5134–5157. https://doi.org/10.1002/2014JC010349

Yi, J., Du, Y., Wang, X., He, Z., & Zhou, C. (2013). A clustering analysis of eddies' spatial distribution in the South China Sea. Ocean Science, 9(1), 171–182. doi:10.5194/os-9-171-2013

Zhou, X., Marsland, S. J., Fiedler, R., Bi, D., Hirst, A. C., & Alves, O. (2015). Impact of different solar penetration depths on climate simulations. Tellus A: Dynamic Meteorology and Oceanography, 67(1), 25313. doi:10.3402/tellusa.v67.25313

Zhu, X.–H., Ma, Y.–L., Guo, X., Fan, X., Long, Y., Yuan, Y., Xuan, J.–L., and Huang, D. (2014), Tidal and residual currents in the Qiongzhou Strait estimated from shipboard ADCP data using a modified tidal harmonic analysis method, J. Geophys. Res. Oceans, 119, 8039– 8060, doi:10.1002/2014JC009855.