OceanRoads: Fishing Area Separation and Relation Discovery through VMS Trace Analysis

Feng Hong[1], Yong Zhen[1], Yuan Zong[1], Yuan Feng[1]*, Haiguang Huang[2], Hao Wang[3], Zhongwen Guo[1]

1. College of Information Science and Engineering, Ocean University of China, Qingdao, China
2. Wenzhou Ocean and Fishery Vessel Safety Rescue Information Center, Wenzhou, China
3. Faculty of Engineering and Natural Sciences, Norwegian University of Science and Technology, Aalesund, Norway
∗ Corresponding Author: Yuan Feng, Ocean University of China, fengyuan@ouc.edu.cn

Abstract—**Vessel Monitoring Systems (VMS) provide trajectory data of fishing vessels to discover the spatial-temporal pattern of fishing activities. Although previous researches exhibit the fishing resource distribution in space and time, the information on vessel sailing activities traveling from one fishing region to another are lost. These kinds of activities indicate the link relationships among fishing regions. Moreover, the pattern accuracy depends on the precision of fishing activity recognition, which cannot be testified by the VMS data itself. The major challenge to find precise fishing regions and sailing roads linking the regions is the chicken-and-egg conundrum. Because there are no pre-existed routes in oceans like on land to define the blocks. This paper proposes an iterative style scheme to calculate the fishing regions and sailing roads between regions simultaneously, which applies the sailing segments between the coarse recognized fishing activities. We takes the VMS traces of 34 otter trawls in the East China Sea as the research object ranging from April 1st, 2014 to June 30th, 2016. The results are calculated with the time period of one quarter, which not only indicate the precise separation of fishing regions , but also reveal the link relationships between regions. This leads to the precise fishing region and roads recognitions, which will help the fishery researchers for further analysis on VMS trace data.**

I. INTRODUCTION

Vessel Monitoring Systems (VMS) are originally designed to enforce and control vessel sailing. It records the sailing information of the vessel, including the position, heading, speed, and date, etc. The exchange of such information can keep the vessel safe during sailing. As the fishing vessels are deployed with VMS systems, a large amount of trajectory data has been collected, which brings a new opportunity for fishing researches to study the spatial-temporal pattern of fishing activity and impact.

For example, Mills et. al. [1] and Mullowney et al. [2] calculated the places where fishing activities happened and create the distribution of fishing regions/resources. Combining fishing regions with temporal information, Fonseca et al. [3] conducted the analysis of dynamics of fisheries. There is, however, some potential error in fishing region recognition, for the VMS trace set itself cannot testify the accuracy of fishing region recognition.

Although above researches exhibit the fishing resource distribution in space and time, the information on vessel sailing activities traveling from one fishing region to another are lost. These kinds of activities indicate the link relationships among fishing regions. Making a metaphor to land scenarios,
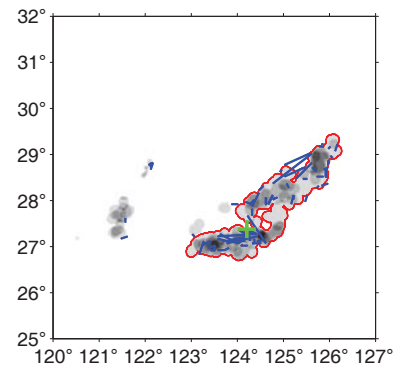

Fig. 1: Coarse recognition results on fishing and sailing activity. Gray color labels the fishing region and the darkness indicates the density. Blue line segment indicates the sailing activity between two fishing regions for a vessel. The red curve indicates that one fishing region are directly separated by image processing.

these links or channels are roads in the oceans. As roads can help to indicate blocks, the links here can further make the separation of the fishing regions clear i.e. testify the fishing region recognition result. Besides, the link relationship among precise fishing regions may disclose more information behind fishing activities.

The major challenge to find precise fishing regions and sailing roads linking the regions is the chicken-and-egg conundrum here. Because there are no pre-existed routes in oceans like on land to define the blocks. Meanwhile, the distribution of fishing grounds nowadays are originated from coarse statistics.

Our research presents an iterative scheme to calculate the fishing regions and sailing roads between regions simultaneously. The proposed scheme includes two phases: coarse calculating and iterative calculating. During the first phase, we recognize the fishing and sailing activities from VMS data set by applying our previous research [4]. Then we locate the coarse fishing regions by statistics on all the regions where fishing activities exist. At the same time, we collect all the sailing segments between two fishing activities for the same vessel. The fishing regions are the candidate blocks and the sailing segments are potential roads.

Figure 1 shows an example the coarse recognition result of fishing regions and sailing segments on the VMS traces of the second quarter, 2014. The darkness of the region represents the density of the region being fishing. The blue line segments indicate the possible roads between fishing areas. It shows that many segments are located inside the region. This may come from the recognition error of fishing activity, which further leads to calculation errors on fishing region and sailing segments. So the research question is: should we split the fish region further to fit the roads links between regions? Or should we just eliminate the links inside one fishing region as regarded as they are induced from the fishing activity recognition error?

The second phase answers the above question with its iterative design. We consider the difference between the adjacent two fishing density layer on the initial results like Fig. 1. There are three kinds of relationships on fishing regions between two layers. The first is that a new region appears on the lower fishing density layer. The second is that a region expands its boundary from the higher layer to the lower one. For these two types, it is reasonable to treat the regions as a separate area with the boundary indicated by the lower layer. The third, however, is complex, where the separate regions in the higher layer are aggregated as a whole region in the lower one. Here whether the separate regions should be aggregated as one or keep separated are the research questions. We try to apply all the information of regions and segments to answer the question.

The key intuition here is to let the vessels vote. A line segment indicates that a vessel do not treat the current region as a whole one once. We then define the number of line segments inside one region as the negative vote number. On the contrary, the line segment between two regions are treated as positive vote. Note that both the region and the line segment come from the untestified fishing and sailing activity recognition. And the number of vessels is called as the number of voters, which fished inside this region. Then we can compare the vote number between the separate regions and the aggregated one. We will further separate on this fishing region if the latter is higher, or keep the separated regions in the high density layer unchanged.

After ascertain the separation among fishing regions by vote calculation region by region, we get the fishing cluster results based on the whole trace set. We then combine the segments between the clusters into one line segment linking the related cluster centers, whose width indicates the density relationship between one fishing region and another. Thicker the line, more vessels are sailing from one to another, more important the relationships between them are. Moreover, the line segments inside each cluster are deleted, for they are judged to be wrong segments, tracing back to the activity recognition error.

We take the VMS traces of 34 otter trawls in the East China Sea as the object. The traces are recorded from April 1st, 2014 to June 30th, 2016 with a total records number of 2,140,288 and 271.81 MB in total. Specifically, each vessel has 62,949.65 on average with a maximum of 144,413 records
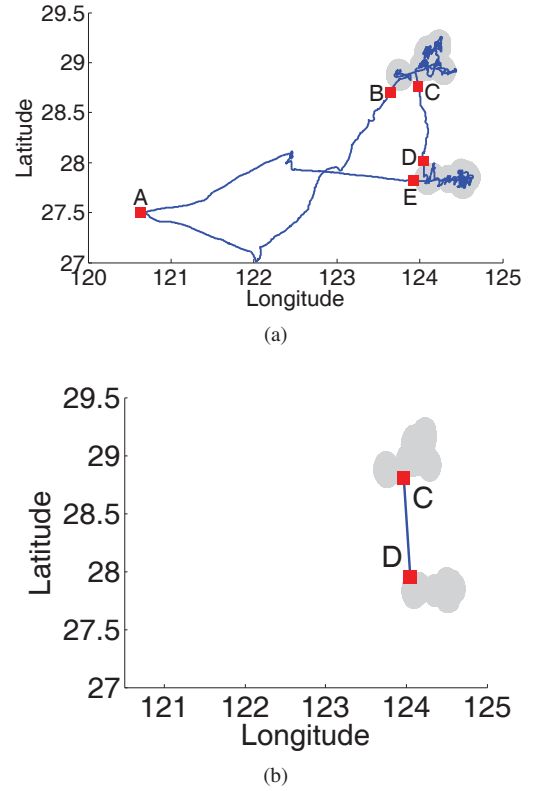

(a)

(b)

Fig. 2: Recognition of fishing and sailing activities in one voyage of one vessel (a) original result (b) after abtracting

and a minimum of 15,799. Size of each vessel file is 8.24 MB on average with a maximum of 17.47 MB and a minimum of 1.91 MB. Besides, the traces are recorded by China Beidou Satellite system, giving them a 60-second time resolution. Our trace set has the property of fine resolution, comparing to AIS system that records data per 2 hour in general.

Hence, applying the proposed iterative scheme, we calculate the separation of fishing region from the fishing density figure like Fig. 1. The results are calculated with the time period of one quarter. These new results not only indicate the precise separation of fishing regions based on the sailing activities of vessels, but also reveal the link relationships i.e. roads between regions. Therefore, the knowledge of the fisherman has been implicitly applied to determine the fishing regions and linking roads at the same time.

II. MATERIALS AND METHODS

A. Data

The original data used in this research is the VMS traces of 34 otter trawls in the East China Sea as the object, recorded by Zhejiang Province Ocean and Fisheries Bureau. The traces are recorded from April 1st, 2014 to June 30th, 2016. Besides, the traces are recorded by China Beidou Satellite system, giving them a 60-second time resolution. Our trace set has the property of fine resolution, comparing to AIS system which records data per 2 hour in general. This interval is far shorter
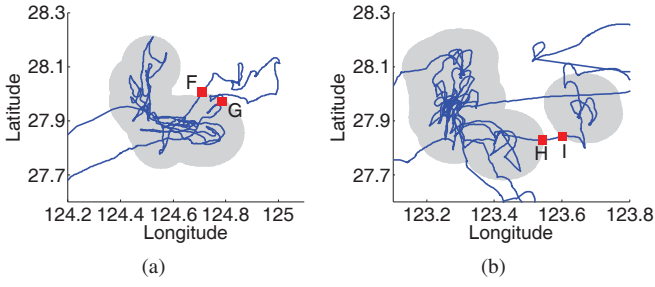
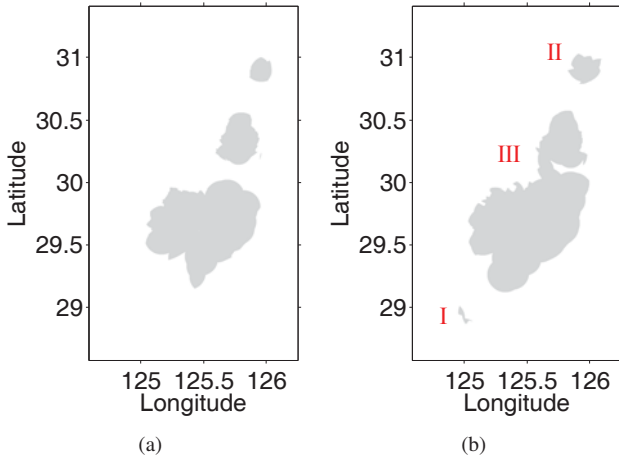Fig. 3: Two types of redundant segments

Fig. 4: Example of adjacent density layers of fishing regions of the third quarter, 2015

Fig. 5: Illustration on the problem to combine or separate fishing regions with the data of 1st quarter, 2016 (a) α density layer (b) adjacent lower density layer

than that used in previous researches. So we eliminate the work of interpolation on VMS data. However, due to transmission anomaly, equipment abnormal shutdown and other factors, our data are still with a small amount of outliers and missing records.

B. Methods

The proposed scheme includes two phases: coarse calculating and iterative calculating. During the first phase, we recognize the fishing and sailing activities from VMS data set by applying our previous research [4]. The results are the fishing and sailing activity distinguishing for all voyages of all vessels. Figure 2a shows one result for a voyage of one vessel. Here the region forms one fishing area which covers the trajectory between points B and C. Identically, another fishing region covers the trajectory between points D and E. The trajectory linking points C and D is the sailing segment. We further carry out segments abstraction and eliminating redundant segments as follows.

Abstracting segments: It is not useful for the detailed trajectory inside the fishing regions for further steps. We only keep the fishing density value and boundary for fishing areas. The sailing segments are representative for the relationship between two fishing regions. However, the detailed trajectory
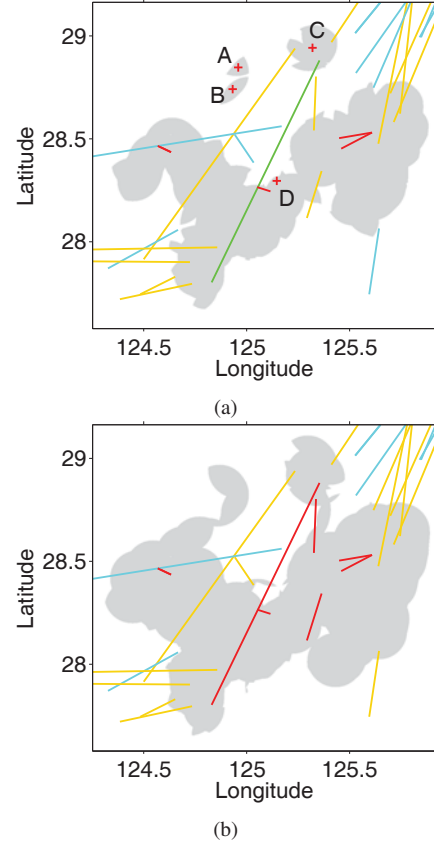
of sailing segment will bring in too many details. So we abstract the sailing trajectories to line segments. Figure 2b shows the abstracting results of the trajectory in Fig. 2a, in which segment CD indicates the sailing segment between two fishing regions.

Eliminating redundant segments: Without logbook or other records, we can just recognize fishing activities approximately. Actually, shape and scale of fishing regions can hardly be confirmed for uncertainty of sailing. When applying abstraction, there are two types of segments needed to be further removed. The first kind is those segments with two end points linked at the same fishing region, like the example shown in Fig. 3a. These segments may be the errors due to fishing activity recognition, so we eliminate such segments. The second type is some short segments between two fishing regions, as the example in Fig. 3b. Such segment may come from misclassifying one fishing regions into two regions. We set a distance threshold for 2 nautical miles and eliminate the segments below this threshold.

When adding all the results from all voyages like in Fig. 2, we got the results after abstraction and elimination, shown as Fig. 1. Blue segments indicate all the sailing segments of the 2rd quarter in 2014, and shaded regions form a heat map of overlapping fishing regions. The heat represents the fishing
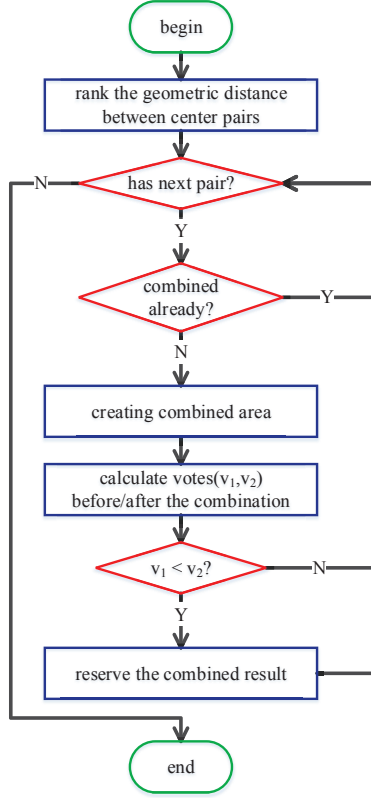
Fig. 6: Flow chart of combination and voting algorithm

density here. Fishing regions connect to each other and there are a mount of segments, which means many vessels have at least twice independent fishing activities. In other words, we need to separate the fishing region into parts according to the relevant fishing and sailing activities, so that the analysis can be specific and efficient. Hence, we can reveal fishing vessels behaviors and the fishery resources distribution to a more precise level by analyzing fishing regions separately.

The second phase of proposed scheme takes the results in Fig. 1 as the input. We take the difference between adjacent fishing density layer into consideration. An example of two adjacent fishing density layers is shown Fig. 4. The example comes from the results of the third quarter in 2015. Here we choose the density layer of $alpha = 1/4$ and the one whose density is just below α. There are three kinds of relationships on fishing regions between two layers. Firstly, a new region appears on the lower density layer as the region I shown in Fig. 4b. The second is that a region expands its boundary from the higher layer to the lower one, as the region II in Fig. 4b. For these two types, it is reasonable to treat the regions as a separate area with the boundary indicated by the lower layer. The third, however, is complex, where the separate region in the higher layer are aggregated as a whole region in the lower one, as the region III in Fig. 4b comparing to Fig. 4a. It is a nontrivial question whether this region III should be treated as one or separated as two. Here is the segment information which should be applied.

Figure 5 plots the same kind of scenarios with the data of the first quarter of 2016. There are four separate fishing regions, labeled with A, B, C and D in Fig. 5a. And there is only one whole fishing region in Fig. 5b. The sailing segments are also depicted with different colors. Blue labels the segment whose two end points are both not inside any fishing region. Yellow labels the segment whose one end resides in the fishing region. Green labels the segment with both ends in two different region. And red labels the segment with both ends in the same region.

We apply the segments of different colors as the voter to determine whether the four regions should be combined or not. The blue segment relates no fishing regions, so we treat it as abstaining from voting. The yellow segment represents that a vessel sailing out from one fishing region, so it supports that there is one fishing region. We treat it as one affirmative vote. Identically, the green segment supports two regions at its both end, so there is one affirmative vote for each region. In contrast, the red segment indicates that there is one vessel who conduct fishing, sailing and fishing again in one region. This tells that the vessel does not take the region as a whole fishing region. We count two negative votes here for the related region. Then we can compare the votes before and after region combination. If the first is higher, we keep the regions separated as the higher density layer. Otherwise, we take the combined results. Therefore, the sailing behaviors of the vessels are used to decide the combination regions i.e. the knowledge of the fisherman has been implicitly applied to determine the fishing regions.

There is still one problem of how to combine the separate regions. For example, there are 4 regions in Fig. 5a with only one in Fig. 5b. We design an algorithm for the combination and voting comparison process as shown in Fig. 6. We first sort the geometric distance between the region centers in the ascend order. Then each two regions will be considered whether to combine according to the order. We first create one combined region for these two. Then we can compare the votes before and after the combination and make a decision.

Meanwhile, compared with Fig. 5b, region A and B are not separated between them and even among C and D. So where is the boundary of the combination between A and B? Fig. 5b does not provide any clue to answer this question. We present a eclipse cover method to decide the boundary of the combination of two parts. For example, the two nearest region centers are of A and B in Fig. 7a . We link the two region center of A and B and extend this line to the shortest of the boundaries. The center of the eclipse is taken as the middle point of the segment between the two region center. Then the eclipse can be created with above conditions as depicted in Fig. 7a . The eclipse area are further intersected with the fishing regions of the adjacent lower density layer. The intersection is used to eliminate some non-fishing regions created by eclipses. The combination result of A and B is shown in Fig. 7b.

Now we compare the votes between and after the combination. For this example, the vote are both 0, for no segments
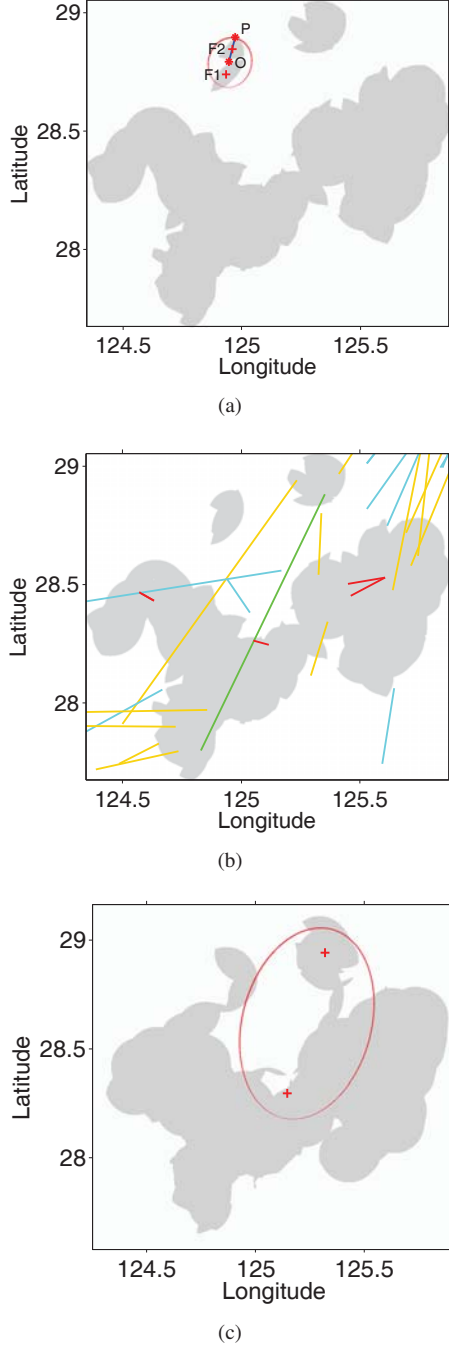
(a)

(b)

(c)

Fig. 7: Combination and voting process illustration (a) creating covering eclipse for A and B (b) the result after solving combination A and B (c) covering eclipse and intersection for B and D

related to A and B. In this condition, we take the combination result as these two are the nearest. Then we consider the regions A and C. We find that A and C are still separated, when applying intersection between the combined eclipse and the lower layer. So A and C will keep separated directly without counting the votes. Then the results are the same regarding B and C or B and D.

Next the cover eclipse of combining region C and D is illustrated in Fig. 7c. The intersection between the covering eclipse and the low density layer is shown as the shaded region inside the eclipse. Now we should count the votes before and after the combination of C and D, which are 9 and 5 respectively. So we keep C and D separated. A and D are further calculated with above process. The voting result is still support separation. Therefore, the final separating results for this quarter is just the one shown in Fig, 7b. Then we can count the segment number between two regions. A line segment will be used to represent all the segments between two regions with its width as the segment number, which indicates the road between these two regions.

After applying the algorithm in Fig. 6 to all quarter coarse results and calculating the roads, we can achieve the goal to separate regions and roads on the basis of sailing segments.

III. RESULTS

Figure 8 shows the final results of fishing regions and sailing roads of the original dataset. Comparing the figures of the different quarters, it shows that there are significant differences among them. The roads in the fourth quarter has the most complex shapes, the longest length and the widest width. This may illustrate that the fishermen's activity are most aggressive in fourth quarter. The second complex ones are the first quarters, which present some similarity with the fourth quarter. This may come from the similarity in the fishing resource distribution. More importantly, it may be because the holiday seasons are in the middle of these two quarters. So the fisherman are trying to get more fishery production by sailing longer distance and fishing more areas.

On the contrary, there are three fishing off season across the second and third quarter. Hence, the fishing roads are of shorter distance and the fishing region are of less areas. Meanwhile, the shapes of fishing road are more like each other in the same quarter than the ones in different quarters. Therefore, more information and conclusion can be supported with both the fishing region and road results.

IV. CONCLUSION

This paper proposes an iterative style scheme to calculate the fishing regions and sailing roads between regions simultaneously, which applies the sailing segments between the coarse recognized fishing activities. The knowledge of the fisherman has been implicitly applied to determine the fishing regions and linking roads at the same time. This leads to the precise fishing region and roads recognitions, which will help the fishery researcher for further analysis on VMS trace data.
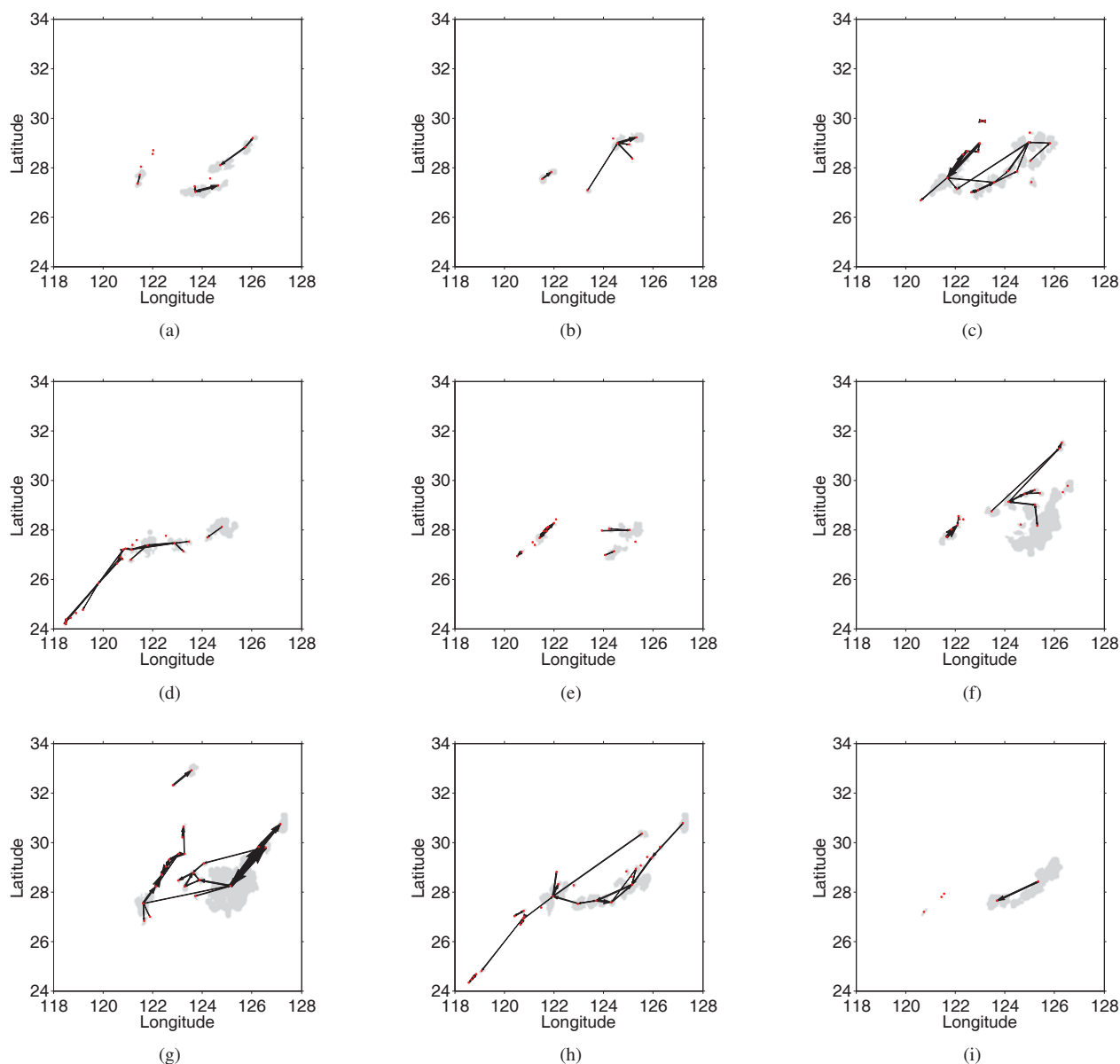
Fig. 8: Fishing regions and roads recognition results across all quarters (a) 2nd quarter 2014 (b) 3rd quarter 2014 (c) 4th quarter 2014 (d) 1st quarter 2015 (e) 2nd quarter 2015 (g) 3rd quarter 2015 (g) 4th quarter 2015 (h) 1st quarter 2016 (j) 2nd quarter 2016

REFERENCES

[1] Mills, C. M., Townsend, S. E., Jennings, S., Eastwood, P. D., and Houghton, C. A. Estimating high resolution trawl fishing effort from satellite-based vessel monitoring system data. ICES Journal of Marine Science, 64: 248-255, DOI: 10.1093/icesjms/fsl026. 2007.

[2] Mullowney, D. R., and Dawe, E. G., Development of performance indices for the Newfoundland and Labrador snow crab (Chionoecetes opilio) fishery using data from a vessel monitoring system. Fisheries Research, 100: 248-254, DOI: 10.1016/j.fishres.2009.08.006. 2009.

[3] Fonseca, T., Campos, A., Afonso-Dias, M., Fonseca, P., and Pereira, J.. Trawling for cedhalopods off the Portuguese coast?fleet dynamic and landings composition. Fisheries Re- search, 92: 180-188, DOI: 10.1016/j.fishres.2008.01.015. 2008.

[4] Yuan Zong, Haiguang Huang, Feng Hong, Yong Zhen and Zhongwen Guo. Recognizing Fishing Activities via VMS Trace Analysis Based on Mathematical Morphology, IEEE/MTS Techo-Ocean 2016 conference, Kobe, Japan, Oct. 2016.

[5] Hartigan, J. A.; Wong, M. A. Algorithm AS 136: A k-Means Clustering Algorithm. Journal of the Royal Statistical Society, Series C. 28 (1): 100-108. 1979.