

# Application of The KNN Algorithm in Determining the Orientation of The Probability Area Containing The Ship Position by GPS Systems on Hai Phong Coastal Area

Thai Duong Nguyen<sup>#1</sup>, Trong Duc Nguyen<sup>#2</sup>

*<sup>#1</sup>Vietnam Maritime University, Haiphong city, Vietnam  
E-mail: <sup>#1</sup>nguyenthaiduong@vimaru.edu.vn*

---

**Abstract**—The need for navigation for the maritime industry is also very urgent. Today, most commercial boats have receivers and positioning signal receivers, giving an accuracy of several meters. For vessels arriving at, or passing through small canals, the ability to locate with an accuracy of less than 1 meter is necessary. In these cases, the use of DGPS technology is necessary. The determination of ship location and navigation depends on global satellite navigation systems, mainly GPS systems. In maritime practice, the position of the specified vessel is considered the most probable position and will be the center of the circle of probability area containing the ship position. However, the probability area containing the ship position is scalar, the radius of error of the circle of probability area depends on many factors, such as the deviation of geodetic system, the accuracy of the chart. Therefore, the determination of the most probable position with the highest accuracy is a quite complex problem. The demand for data processing is also greater; Machine Learning is thus contributing to solving this problem. In the framework of the article, the authors propose the application of the KNN algorithm to determine the orientation of the probability area containing the ship position with the most probable positions. The objective of this study is to improve efficiency and safety in maneuvering and navigation for sea vessels and testing for Hai Phong coastal area.

**Keywords**—KNN algorithm; GPS systems; ship position; maritime industry.

---

## I. INTRODUCTION

Currently, there are 3 global positioning systems in the world: NAVSTAR (Navigation Signal Timing and Ranging) system is commonly known with a shorter name "Global Positioning System (GPS)"; GLONASS system (Russia's Globalnaya Navigatsionnaya Sputnilovaya Sistema - Global Navigation Satellite System) and the European Commission's Galileo system [1]-[3].

Global positioning satellite systems operate on the principle of using distance measurements from satellites to receivers based on the assumption of time synchronization between satellites and receivers [4]. Global positioning satellites continuously transmit signals that contain the satellite's unique code, information about the position and velocity of the satellite at the time of signal transmission. To determine their position, the receiver must receive satellite signals. Receiving satellite signals for civilian applications is free; continuous seven days a week, 24 hours a day; not influenced by the weather [5].

Maritime has, currently, been moving towards the integration of electronic navigation [6], [7]. The safety of ships has become simple but there are many risks arising

from carelessness, ignorance of the officers and the limitations of electronic equipment. The issues of locating and navigating vessels depend on global navigation satellite systems, mainly GPS systems [8].

GPS is designed and managed by the US Department of Defense, but the US Government allows people to use it for free since 1980, regardless of nationality [9]. GPS works in all weather conditions, anywhere on the earth, 24 hours / day. There is no subscription fee or payment for setting up GPS [3].

Today's GPS receivers achieve extremely high accuracy, thanks to the multi-channel design that works in parallel. Parallel 12-channel receivers (of Garmin) quickly lock on the satellites when they are turned on and they maintain this connection. Certain conditions of the atmosphere and other sources of error can affect the accuracy of GPS receivers. GPS receivers have an average accuracy within 15m [10].

With a high-precision location detection function, GPS receivers are indispensable for seafarers. The location data from the GPS receiver will be paired with MF / HF information devices, Inmarsat satellite information to serve the position of the ship to shore during the maritime operation as well as in emergency situations [9].

Accordingly, the International Maritime Organization IMO requires all vessels over 300 tons and passenger ships to install electronic navigation systems on international sea routes according to the provisions of the International Convention on Human Life Safety (SOLAS) [7].

GPS navigation system has become an ideal maritime navigation tool. The accuracy of navigation requirements varies from a few meters (on the beach, at the pier and in the river) to a few hundred meters (navigation on the way) [11]. Accurate GPS positioning using both pseudo-random measurement and measurement of carrier phase can lead to navigation of boats on the river and coastal without buoys[12].

Positioning needs for fieldwork in ocean physics can also be met by GPS. The complementary carrier phase measurement gives the correct boat speed, which is necessary in studying ocean currents[13].

Methods of data processing contribute greatly to eliminating measurement errors. People focus on the following two solutions: In processing GPS data, people are interested in effecting the measured values to eliminate errors, in which the most effective one is the measured value between time points [14]. Signal reception of a satellite, quadratic effect is the measurement effect between satellites and third-order signals that are measured between ground points. Use the rank order to have a base line solution that contains at least one measure of progress over time. The problem of noise filtering is a complex technique in processing satellite data, over time people have come up with more complete filters so that there is only random noise in the measurement [15].

However, depending on specific conditions, the accuracy and reliability requirement of GPS positions are different. Depending on the type of GPS receiver, the manufacturer may recommend radius of distance root mean square (DRMS). However, in fact, the radius of error depends on

many geodetic factors such as geodetic system error and nautical chart accuracy. There has been some research to evaluate the actual error of GPS location [16]. However, they only determine the radius of the errors circled. The disadvantage of these researches is that the probability area containing the ship position is undirected, which may affect the captain's decision in some special cases such as in coastal areas with less reliable survey data [17]. Evidently, the marine safety management system is incomplete. As a result, it may cause many serious accidents because of the captain's inaccurate judgment about the vessel's position [18]. The paper focuses on calculating the orientation of the probability area containing the position of the ship identified by GPS receivers in coastal areas of Vietnam. The research results contribute to improving the reliability and updating to ensure maritime safety in Hai Phong coastal area.

## II. MATERIAL AND METHODS

### A. Method of determining vessel position by GPS tracking equipment

In actual operation, the highest probability vessel position is considered to be the vessel position and tend to be the center of the circular error probability containing the vessel position (Figure 1)[15]. However, there is no reliable basis for making a precise estimate of the highest probability vessel position. Besides that, the circular error probability containing the vessel position is undirected. Then, when actual vessel position is on the boundary of circular error probability containing the vessel position will be very different to the vessel position determined by normal methods [8]. In reality, the position distribution determined by the GPS system is very diverse and the circular error probability is difficult to identify [19].

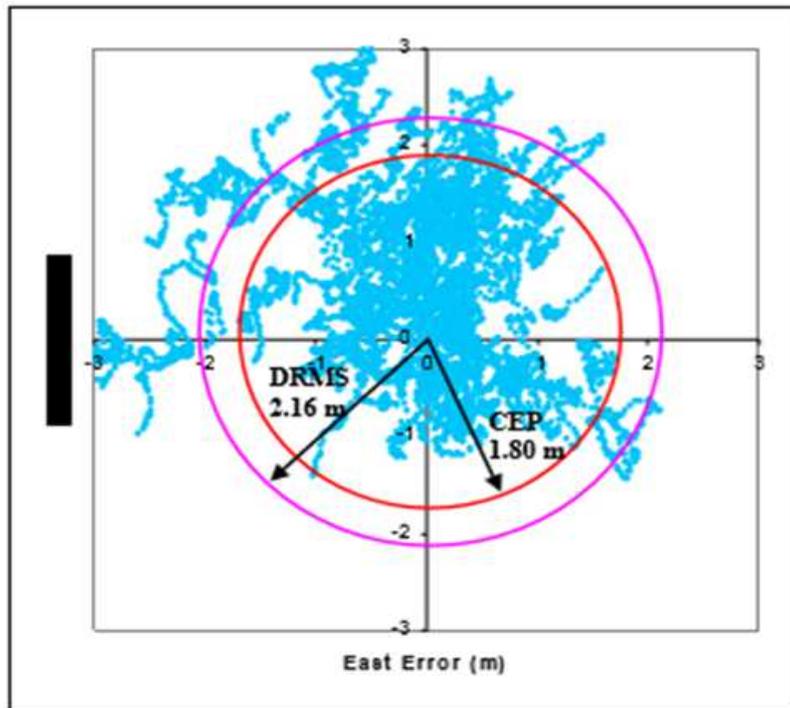


Fig. 1. The vessel position in circular error probability [15]

### B. Calculate the Probability Area Containing the Ship Position

The vessel positions determined by the GPS tracking equipment are operated on the nautical chart [19]. The error of the specified position is carried out experimentally: At the same time, the actual vessel position is  $(\varphi, \lambda)$  and vessel position by GPS is  $(\varphi_1, \lambda_1)$ . The difference between actual and GPS position:

Latitude error:

$$\Delta\varphi_1 = \varphi - \varphi_1 \quad (1)$$

Longitude error:

$$\Delta\lambda_1 = \lambda - \lambda_1 \quad (2)$$

Conducting monitoring to determine the position of the ship ( $t \leq 20$  times) in a coastal area, applying the average square error method to receive the following error value:

Latitude root mean square error:

$$\Delta\varphi = \pm \sqrt{\frac{\sum_{i=1}^{20} (\Delta\varphi_i)^2}{20-1}} \quad (3)$$

Longitude root mean square error:

$$\Delta\lambda = \pm \sqrt{\frac{\sum_{i=1}^{20} (\Delta\lambda_i)^2}{20-1}} \quad (4)$$

Absolute error (99.97%):

$$m_\varphi = 3\Delta\varphi \text{ and } m_\lambda = 3\Delta\lambda \quad (5)$$

Latitude area from 30°N to 30°S, the scale-up is not large, the radius of the circular error probability containing the vessel position determined by GPS tracking equipment can be approximated according to the formula [8]:

$$M = \sqrt{(3\Delta\varphi)^2 + (3\Delta\lambda)^2} \quad (6)$$

At the high latitude area, the scale-up is large, the radius of the circular error probability containing the vessel position determined by GPS tracking equipment can be approximated according to the formula:

$$M = \sqrt{(3\Delta\varphi)^2 + k(3\Delta\lambda)^2} \quad (7)$$

k is the coefficient indicating the increase in the ratio of the progressive latitude of Mercator projection, calculated according to the formula:  $k = \frac{\varphi}{D}$  [18].

$$D = a \ln \operatorname{tg} \left( \frac{\pi}{4} + \frac{\varphi}{2} \right) \frac{1 - e \sin \varphi}{1 + e \sin \varphi} \quad (8)$$

Because  $k < 1$ , in reality, we may use the formula (7) to increase the safety level of the probability area containing the vessel location. The actual vessel position  $(\varphi, \lambda)$  can be equal arithmetic average measurement results, calculated according to the formula [20]:

$$\varphi = \sum_{i=1}^n \varphi_i \text{ and } \lambda = \sum_{i=1}^n \lambda_i \quad (9)$$

### C. Accuracy evaluation of vessel position in circular error probability

#### Accuracy evaluation – 2D mode

Distance Root Mean Squared Method – DRMS evaluate the accuracy of GPS position in 2D mode:

$$\text{DRMS} = \sqrt{\sigma_x^2 + \sigma_y^2} \quad (10)$$

$\sigma$  stands for average error in two directions (x, y)

Circular Error Probability Method – CEF: is the radius of a circle, whose boundary is expected to include the vessel position of 50% of the rounds. For example, if the radius of the circle is 5 meters, then the probability of the vessel's actual position is inside the circle of 5m radius is 50%. The radius of a circle in which 95% of the probability contains the actual vessel position is often referred as R95 [21].

The following table introduce a few ways to calculate the accuracy of 2D GPS locations and their probabilities:

TABLE I  
2D MODE ACCURACY EVALUATION [22]

Methods	Formula	Probability	Definition
DRMS	$\sqrt{\sigma_x^2 + \sigma_y^2}$	65%	The square root of the total square of the mean error
2DRMS	$2\sqrt{\sigma_x^2 + \sigma_y^2}$	95%	Equal 2 times of DRMS
CEP – R50	$0.62\sigma_y + 0.56\sigma_x$ and $\frac{\sigma_x}{\sigma_y} > 0.3$	50%	The radius of a circle has a position probability of 50%
CEP - R95	$R(0.62\sigma_y + 0.56\sigma_x)$ (R=2.08, when $\frac{\sigma_x}{\sigma_y} > 0.3$ )	95%	The radius of a circle has a position probability of 95%.

#### D. Accuracy evaluation – 3D mode

Similar to 2D accuracy evaluation, there are many ways to evaluate the accuracy in 3D mode with different probabilities. Evaluate accuracy methods in 3D mode is the same as 2D mode by expanding one more direction - vertical. Spherical error probable (SEP) in 3D mode corresponds to

the circular error probable in 2D mode. Besides that, mean radial spherical error (MRSE) in 3D mode corresponds to distance root mean squared (DRMS) in 2D mode. The following table shows a few ways to calculate the accuracy of 3D GPS locations and their probability [20]:

TABLE II  
3D MODE ACCURACY EVALUATION [21]

Methods	Formula	Probability	Definition
SEP – R50	$0.51(\sigma_x + \sigma_y + \sigma_z)$	50%	The radius of the sphere has position probability of 50%.
MRSE	$\sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}$	61%	The square root of the total square of the mean error
SEP – R90	$0.833(\sigma_x + \sigma_y + \sigma_z)$	90%	The radius of the sphere has position probability of 90%.
SEP – R99	$1.121(\sigma_x + \sigma_y + \sigma_z)$	99%	The radius of the sphere has position probability of 99%.

As mentioned in the previous section, with the complex distribution on the circular error probability, determining the highest probability vessel position with the highest accuracy is not a simple matter. In the context of this paper, the authors propose the application of the KNN algorithm to solve the problem.

#### E. K-nearest neighbors (KNN) algorithm

The K - Nearest Neighbors is known as an algorithm to classify a given data point based on the feature similarity between the objects with their nearest neighbors. To classify an object, the nearest neighbor (K) of the object is defined. The distance between the object to be classified with all neighboring objects will be calculated. On that basis, the class of the object will be determined by the subclass of the nearest neighbor K [2, 21]. The distance between the object to be classified with the neighbor K of the object (d) can be determined by many methods such as Manhattan, Euclid, Minkowski, and Chebyshev for problems with real numbers input; Hamming distance for problems with binary input; Cosine function for classification problems, etc. In the context of this paper, to determine the vessel orientation in the circular error probability, Euclid function will be used to calculate the distance d:

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \quad (11)$$

$(x_1, y_1)$  stands for coordinates of the object to be determined – the center of the circular error probability,

$(x_2, y_2)$  stands for coordinates of possible vessel positions determined by GPS.

#### F. The process determines the vessel orientation in the circular error probability

The steps of the algorithm:

*Step 1:* Locate the vessel at a fixed position 100 times continuously. Put the specified positions on the nautical chart.

*Step 2:* Determine the probability area containing the Vessel positions (circle with the smallest radius covering all positions) according to the KNN algorithm

For each vessel's position, we calculate the distance from this position to the remaining positions according to Euclid function, then storing the distances into the data array (array A- 100 elements). The total distance from one position to the remaining positions is calculated based on the sum of the elements of A. We repeat the process with the next elements, then collect these datas to the array B (100 elements) that stores the total distance from one position to the remaining positions.

Select the first 20 elements in array B, then construct a circle covering 20 positions corresponding to the selected 20 elements. The center of the circle is the position corresponding to the first element in array B (the smallest element), circle radius is the twentieth element in array B (the largest element in the first twenty elements)

*Step 3:* Determine the direction: construct a straight line passing at least one point in the circular error probability so that the total distance from the remaining points to the straight line is minimal according to the method of defining the delta line:

For each of the 20 locations selected, find two positions closest to them based on the data stored in array A (Step 2),

Build the delta line equation through these positions so that the distance from the remaining two positions to the straight line is the shortest,

Calculate the total distance from the remaining positions (17 positions) to the line just defined.

Repeat the above steps with the remaining positions (19 positions), storing the values in the array C (20 elements),

Arrange array C in ascending order, the focus line of probability that is defined as the line corresponding to the first value in array C (smallest). Algorithm testing is carried out on the data set obtained from coastal areas with the radius of circular error probability of about 0.8 nautical miles.

### III. RESULT AND DISCUSSIONS

Figure 2 illustrates the results obtained when locating the vessel position at Hon Dau; the time of locating is 5 minutes apart. The positions are randomly distributed and located in the circular error probability with radius  $M = 0.015$  nautical

miles. The positions tend to be concentratedly distributed in the direction of  $072^{\circ} - 252^{\circ}$ . The position number 16 has the smallest total distance to other locations. The position number 5 has the highest distance to the direction of  $072^{\circ} - 252^{\circ}$ .

The results of the analysis and evaluation received above show that the average shift of the quasigeoid surface through

the Hon Dau tidal test station compared to the global quasigeoid reached the value of approximately 0.4 m. The error of determining this quantity depends mainly on the accuracy of the global height value quasigeoid calculated according to the gravity model of the Earth. Currently, for Vietnamese territory, it has the same numerical value as in many different regions of the world, 0.2 m.

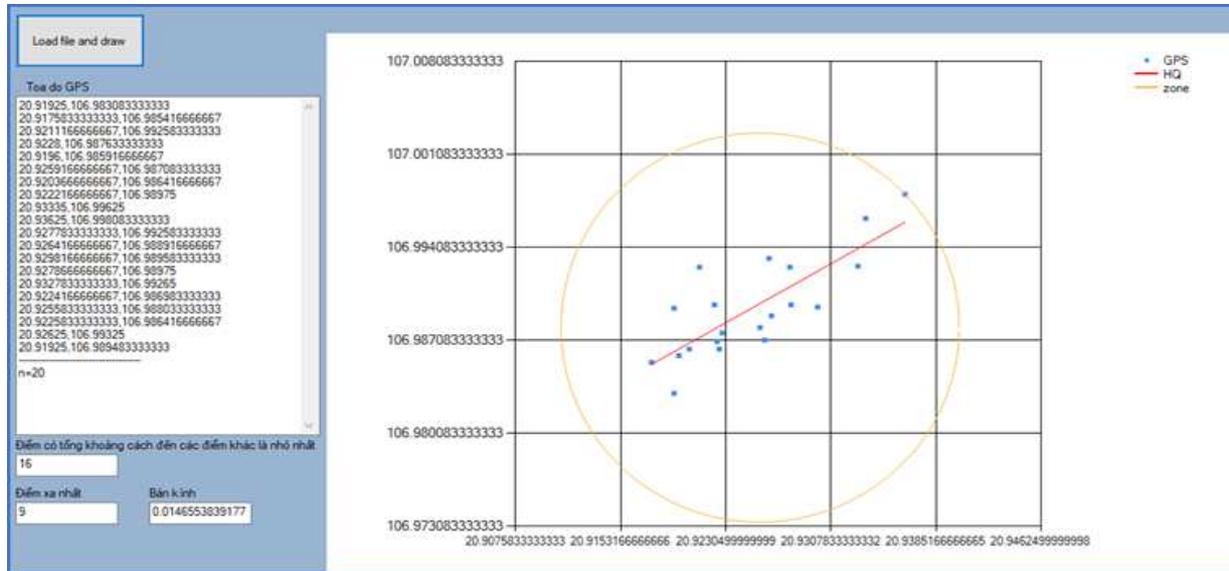


Fig. 2. GPS location in Hon Dau area - 05 minutes

Figure 3 illustrates the results obtained when locating the vessel position at Hon Dau; the time of locating is 30 minutes apart. The positions are randomly distributed and located in the circular error probability with radius  $M = 0.016$  nautical miles. The positions tend to be concentrated

distributed in the direction of  $071^{\circ} - 251^{\circ}$ . The position number 16 has the smallest total distance to other locations. The position number 8 has the highest distance to the direction of  $071^{\circ} - 251^{\circ}$ .

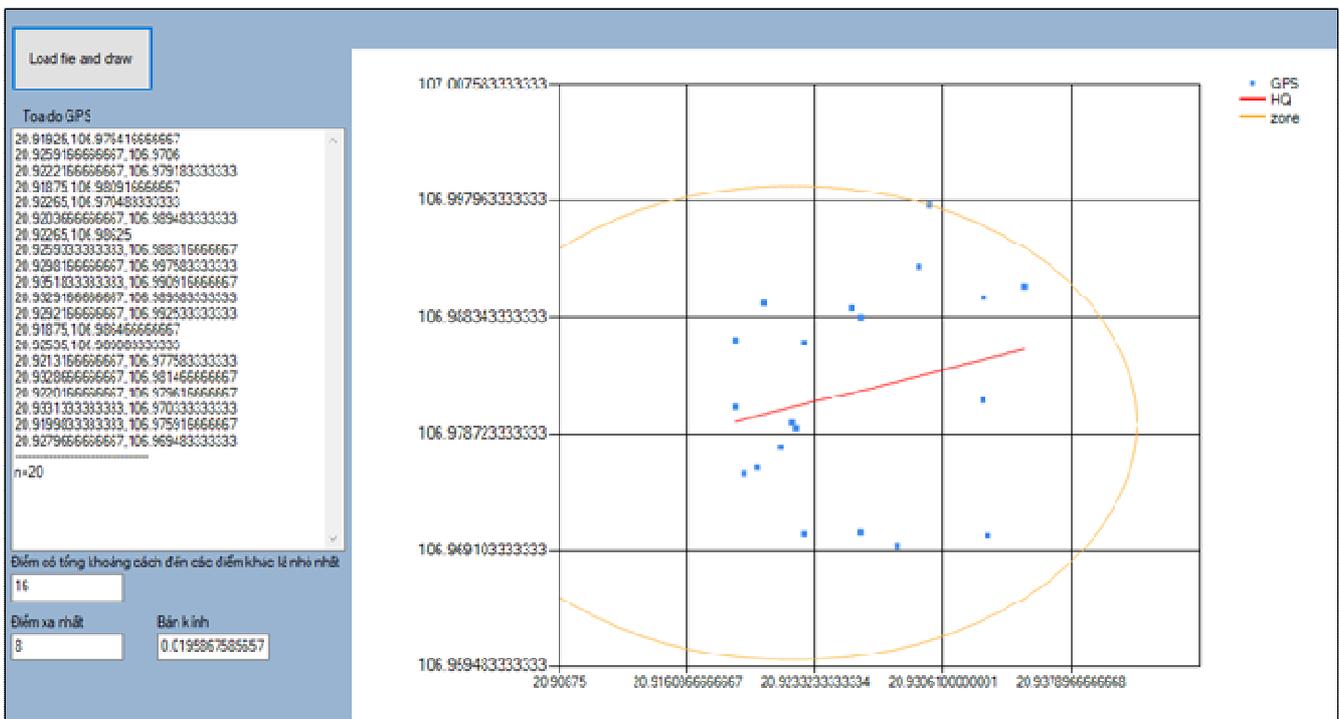


Fig. 3. GPS location in Hon Dau area - 30 minutes

To determine the position of the ship, the true height of the celestial body must be determined, recording the observation time, train speed, direction, azimuth to the

celestial bodies, adjusting the celestial height from the previous time to the time points later.

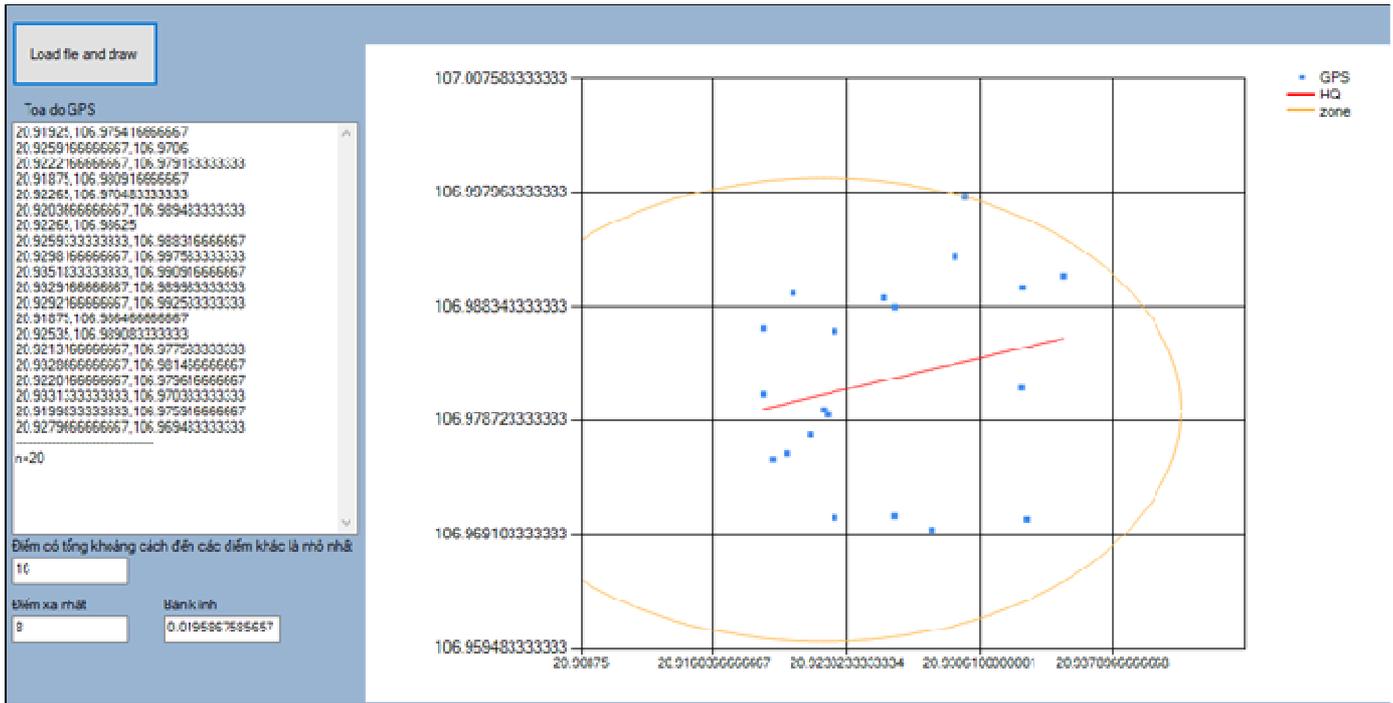


Fig. 4. GPS location in Hon Dau area - 60 minutes

Figure 4 illustrates the results obtained when locating the vessel position at Hon Dau; the time of locating is 60 minutes apart. The positions are randomly distributed and located in the circular error probability with radius  $M = 0.015$  nautical miles. The positions tend to be concentrated distributed in the direction of  $071^\circ - 251^\circ$ . The position number 16 has the smallest total distance to other locations. The position number 8 has the highest distance to the direction of  $071^\circ - 251^\circ$ .

To sum up, theoretically, the vessel orientation is determined by the GPS tracking equipment in Hon Dau area is  $071^\circ - 251^\circ$ .

#### IV. CONCLUSION

The paper has studied the theoretical basis for determining and evaluating the error of vessel position determined by GPS in the Haiphong coastal area. The theoretical background and literature review for determining the vessel orientation and the circular error probability was built based on the KNN algorithm. On that basis, the author conducted the practical examination in Hon Dau, Haiphong coastal area. As explained above, vessel position and orientation determined by GPS tracking equipment in Haiphong coastal area is  $071^\circ - 251^\circ$ . The actual results confirm the reliability of this method. However, it is necessary to conduct additional surveys on the area and increase the numbers of vessel locations to improve the reliability of the method. The ultimate goal is to complete this method and to make a recommendation for the circular error probability and vessel orientation based on locations from GPS tracking equipment in Haiphong coastal area, Vietnam coastal area and other places in the world.

#### REFERENCES

- [1] G. Xu and Y. Xu, GPS: theory, algorithms and applications. Springer, 2016.
- [2] Y. Huang and Q. Zhang, "Identification of anomaly behavior of ships based on KNN and LOF combination algorithm," in AIP Conference Proceedings, 2019, vol. 2073, no. 1, p. 020090: AIP Publishing.
- [3] X. Li, X. Zhang, X. Ren, M. Fritsche, J. Wickert, and H. Schuh, "Precise positioning with current multi-constellation global navigation satellite systems: GPS, GLONASS, Galileo and BeiDou," ScientificReports, vol. 5, p. 8328, 2015.
- [4] J. T. Pisz and B. H. Inouye, "GPS gate system," ed: Google Patents, 2016.
- [5] K.-H. Yang, X. Wang, S. Liu, and L.-L. Zhao, "Precision analysis on result of GPS pseudo-range point positioning and differential positioning," in Mechatronics And Manufacturing Technologies-Proceedings Of The International Conference (Mmt 2016), 2017, p. 3: World Scientific.
- [6] H. Greidanus, M. Alvarez, C. Santamaria, F.-X. Thoorens, N. Kourti, and P. Argentieri, "The SUMO ship detector algorithm for satellite radar images," Remote Sensing, vol. 9, no. 3, p. 246, 2017.
- [7] J. Bhatti and T. E. Humphreys, "Hostile control of ships via false GPS signals: Demonstration and detection," NAVIGATION: Journal of the Institute of Navigation, vol. 64, no. 1, pp. 51-66, 2017.
- [8] K. Jansen, N. O. Tippenhauer, and C. Pöpper, "Multi-receiver GPS spoofing detection: error models and realization," in Proceedings of the 32nd Annual Conference on Computer Security Applications, 2016, pp. 237-250: ACM.
- [9] L. Deng, H. Guo, N. Xiros, and M. Yu, "A research on roll angle calculations based on IMU/GPS compass for ships," in 2016 IEEE/ION Position, Location and Navigation Symposium (PLANS), 2016, pp. 976-980: IEEE.
- [10] T. B. Kiland and J. S. Gray, The Military GPS: Cutting Edge Global Positioning System. Enslow Publishing, LLC, 2016.
- [11] M. H. Assaf, E. M. Petriu, and V. Groza, "Ship track estimation using GPS data and Kalman Filter," in 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2018, pp. 1-6: IEEE.

- [12] B. Jigena et al., "Improving the learning process in the subject of Basic Maritime Training using GPS and Google Earth as useful tools," 2016.
- [13] F. Mazzarella, V. F. Arguedas, and M. Vespe, "Knowledge-based vessel position prediction using historical AIS data," in 2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2015, pp. 1-6: IEEE.
- [14] Š. Pedišić, "Navigation instruments and equipment," University of Zadar. Maritime Department. Division of Nautical Studies., 2016.
- [15] P. Rieth and U. Stählin, "Update of digital maps and position-finding," ed: Google Patents, 2016.
- [16] Y. Shi, C. Shen, H. Fang, and H. Li, "Advanced control in marine mechatronic systems: A survey," *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 3, pp. 1121-1131, 2017.
- [17] J. Xiong, L. Shu, Q. Wang, W. Xu, and C. Zhu, "A scheme on indoor tracking of dynamic ship positioning based on distributed multi-sensor data fusion," *IEEE Access*, vol. 5, pp. 379-392, 2016.
- [18] X. Yang, H. Sun, X. Sun, M. Yan, Z. Guo, and K. Fu, "Position detection and direction prediction for arbitrary-oriented ships via multitasking rotation region convolutional neural network," *IEEE Access*, vol. 6, pp. 50839-50849, 2018.
- [19] S. D. Ilčev, *Global mobile satellite communications theory: for maritime, land, and aeronautical applications*. Springer, 2016.
- [20] P. K. Gaikwad and S. Pawar, "Implementation of real-time GPS receiver system for providing navigation based services and SMS tracking," in 2015 International Conference on Industrial Instrumentation and Control (ICIC), 2015, pp. 630-634: IEEE.
- [21] A. L. Duca, C. Bacciu, and A. Marchetti, "A K-nearest neighbor classifier for ship route prediction," in *OCEANS 2017-Aberdeen*, 2017, pp. 1-6: IEEE.
- [22] A. Deep, M. Mittal, and V. Mittal, "Application of Kalman Filter in GPS Position Estimation," in 2018 IEEE 8th Power India International Conference (PIICON), 2019, pp. 1-5: IEEE.