Scientific Journals
Maritime University of Szczecin

Zeszyty Naukowe
Akademia Morska w Szczecinie

2012, 32(104) z. 2 pp. 157–161

2012, 32(104) z. 2 s. 157–161

Sample size determination and estimation of ships traffic stream parameters

Agnieszka Puszcz, Lucjan Gucma

Maritime University of Szczecin, Faculty of Navigation, Institute of Marine Traffic Engineering
70-500 Szczecin, ul. Wały Chrobrego 1–2, e-mail: a.puszcz@am.szczecin.pl

Key words: ships traffic flow, traffic parameters estimation, sample size determination

Abstract

The paper presents problem of ships traffic flow parameters estimation and the impact of sample size (ships track based on AIS data) on estimation of general population parameters. Too many trials for each estimated point lead to prolonged study, to few trials can no ensure the desired accuracy and reliability of inference. Passages of ships under the Øresund Bridge connecting Sweden and Denmark were analyzed. The conclusions can be used as directives for researches on traffic flows on open and restricted waters.

Introduction

One of the main problems posed by marine traffic engineering is to determine the optimum parameters of new constructed or modernized parts of the waterways. Depending on the type of waterway parameters that can be obtained are for example: lane width or diameter of the turning circle. These parameters are usually determined in one of two methods: a cheaper but less accurate analytical method, or the more expensive and more accurate simulation method. Previous paper [1] presents how utilize historical AIS data to explore the existing manoeuvre pattern in the area, estimate parameters describing traffic flow and used for the construction of a new method of determining a ship manoeuvring area.

Number of measurements, referred to in the statistical literature as sample size, has a significant impact on the results of analyses of stochastic processes, and, by the same token, on the results of statistical evaluation of the quality or efficiency of a process [2]. In analyses of statistical parameters of processes, correct evaluation of sample size is important since it determines the reliability and effectiveness of process evaluation.

Number of AIS data needed to determine the parameters of traffic is sufficient in most cases (over 200 samples in the series). However, in several cases, for certain types of ships and for a specific area, the number of data is limited. An example might be Polish ports, where the number of ships calling ports is much smaller than the other major European ports. Creating a model of traffic flows in the Baltic Sea also can not overlook these areas. The problem therefore is to establish sufficient minimum number of samples.

Analyzed area – The Øresund Bridge

The paper presents the results of researches for the Danish straits – crossing under the Øresund Bridge, which connects Denmark and Sweden. Flintrännan is the navigational channel east of the island Saltholm crossing the Øresund Bridge. Flintrännan is marked with fixed beacons and has a width of 370 m and a limited height of 55 m through the navigation span between the main pylons. A small channel from the port of Malmö named Trindelrenden also crosses the Øresund Bridge. The channel has a width of 100 m through the navigation span and a limited height of 40 m. Specific considerations concerning safety in regard to the ship traffic passing the Øresund bridge are made. Besides the marking a number of protective islands are established with the purpose of avoiding ship collisions with the piers closest to the navigation channel.

Figure 1 shows the navigation conditions near Øresund Bridge.

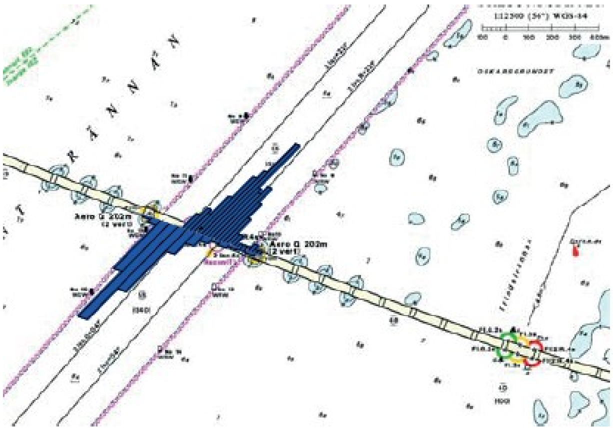

Fig. 1. Navigational conditions in the vicinity of the Øresund bridge with ship traffic distribution (passenger vessel, winter months)

Estimation of ships traffic stream parameters

Estimation is a statistical method aimed at obtaining the optimal model parameter estimates based on available data. Usually, these data are the results of experiments carried out in selected samples. The model parameters point estimation allows to determine the numerical values [3].

One of the most important parameters of ships traffic flow in a given area is spatial distribution. Keeping the desired trajectory is a common feature of individuals on restricted waters. Knowledge about distribution of ships distance to the danger allows to calculate accident probability, and to determine the safety level in the area. Other parameters of traffic streams: traffic intensity, density, velocity and direction of movement. The paper focuses on the spatial distribution of vessels navigating in the Danish straits under the Øresund Bridge.

Random variable was defined as the location of the vessel in relation to the axis of the track. The data obtained was transformed by means of the middle line method used for the bends and straight sections of the fairway based on AIS data [4]. Based on data about the course, waterline and geometric center of waterline are calculated coordinates of extreme points of the vessel (right and left), then their distances to the track axis (the axis of reference).

In order to determine the spatial distribution of vessels special gate was established, where ships position is registered. Distance between ship and channel axis is calculated to determine spatial distribution.

Figures 2 and 3 show an example of a spatial distribution, derived from the empirical data at a gate situated under the bridge. On the X-axis a value of zero correspondents to the middle of the axis. A positive value for X means that the vessel sails more to the starboard side.

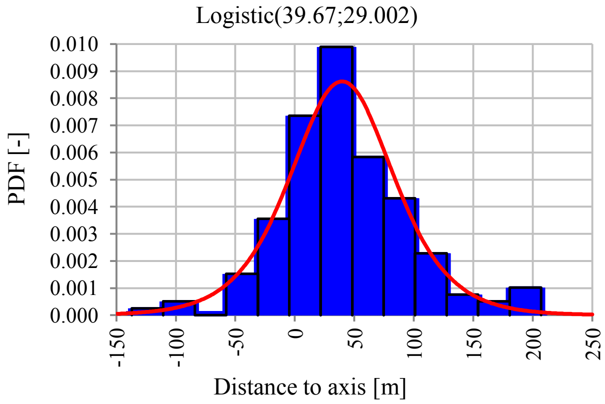

Fig. 2. Spatial distribution over the waterway, northbound traffic. Cargo ships, winter months

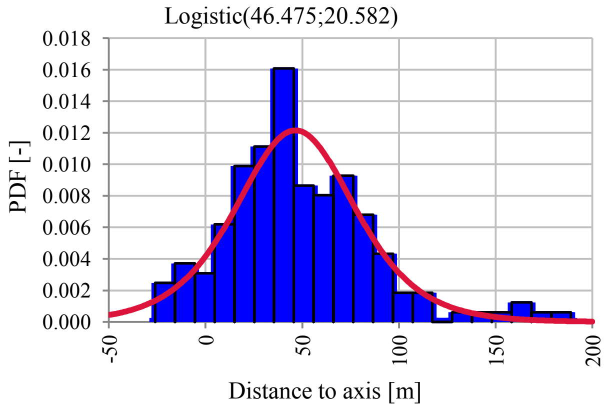

Fig. 3. Spatial distribution over the waterway, northbound traffic. Passenger ships, winter months

A mathematical model of traffic flow is usually defined as a normal distribution [5, 6]. Nowadays, logistic distribution is used increasingly, which is confirmed by the study [1] where the probability density function [7]:

$$f(x) = \frac{e^{-(x-\mu)/s}}{s\left(1+e^{-(x-\mu)/s}\right)^2} \qquad (1)$$

where:

μ – continuous location parameter;
s – continuous scale parameter;
$x \in (-\infty; +\infty)$.

Cumulative Function:

$$\Phi(x) = \frac{1}{1+e^{-(x-\mu)/s}} \qquad (2)$$

Table 3 presents the statistical parameters of distributions for two types of ships (cargo and passenger) heading north and south in the summer and winter months.

In compare to the previous study [6], where calculations were carried out for one general population, the difference is:

- 20 meters for mean value for cargo vessels heading south (lower value for the current study);
- the mean value is comparable for passenger ships heading south;
- the mean values for both ships type (cargo ships and passenger) are greater by about 10 meters in present studies;
- the standard deviation values are comparable only for passenger ships heading north.

Different values of distributions parameters are caused of lack of distinction between types of vessels and seasons in earlier studies.

Sample size determination

Aim of the studies is to determine the minimum sample size of ships manoeuvres based on AIS data for the vessel traffic flow parameters estimation. The minimum sample size can be determined by two methods:

I. Method – increasing the number of trials to stabilize the sample parameters.

II. Method – determining sample size for estimating the mean, when the variance of the general population is unknown, which consists of collecting n' trials and calculating the value of the minimum sample size using equation [8]:

$$n = \frac{z_\alpha^{\ 2} s^2}{d^2} \tag{3}$$

where:

- z_α – value read from standard normal distribution table $N(0,1)$ depend on confidence level of coefficient $(1 - \alpha)$;
- d – the maximum error;
- s – standard deviation for pilot series of n' trials;
- n – sample size.

Formula (3) can be used also for the non-normal distribution by selecting appropriately large pilot sample [8].

The first step to estimate the correct number of measurements repetitions is to determine the level of confidence and maximum error. In this studies assumed significance level $\alpha = 0.05$ and the maximum error of $d = 10$ m and 20 m.

Researches were conducted for a series of crossing under the Øresund Bridge, (straight section of fairway) registered by the AIS station. Passenger and cargo vessels in various hydro meteorological conditions were taken into account.

Figures 4–7 present the mean and standard deviation as a function of 5 to 150–170 of ships movements (northbound and southbound, winter and summer). In both cases, the minimum sample size is determined to $n \approx 20$ measurements in the series – parameters on average level (error $d = 20$ m), or $n \approx 55$ measurements in the series – well stabilized parameters (error $d = 10$ m).

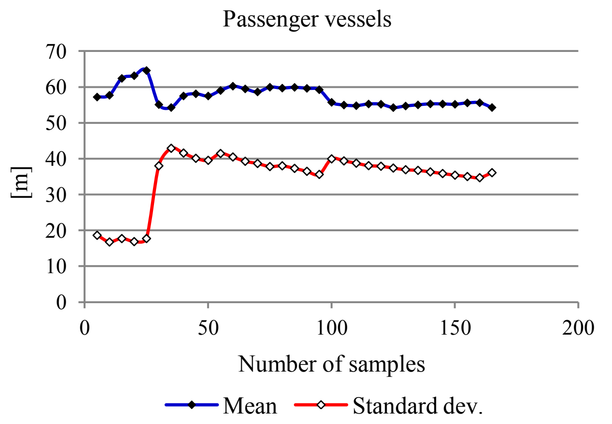

Fig. 4. Mean and standard deviation in relations to the sample size. Passenger ships, southbound, winter months

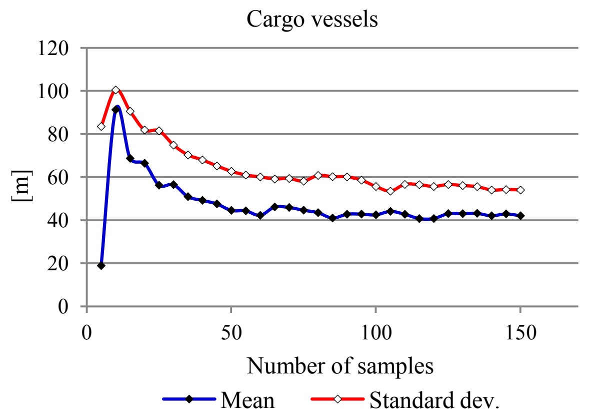

Fig. 5. Mean and standard deviation in relations to the sample size. Cargo vessels, northbound, winter months

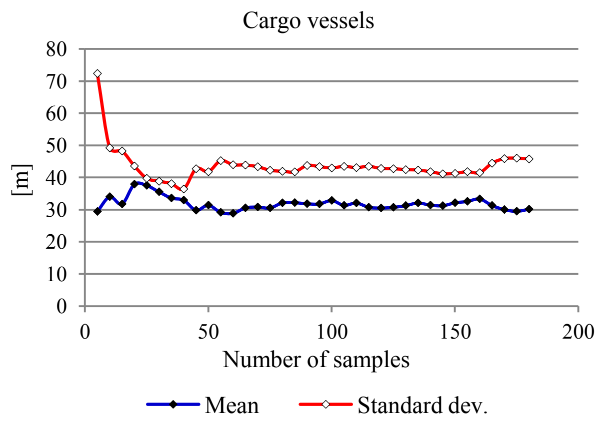

Fig. 6. Mean and standard deviation in relations to the sample size. Cargo vessels, southbound, winter months

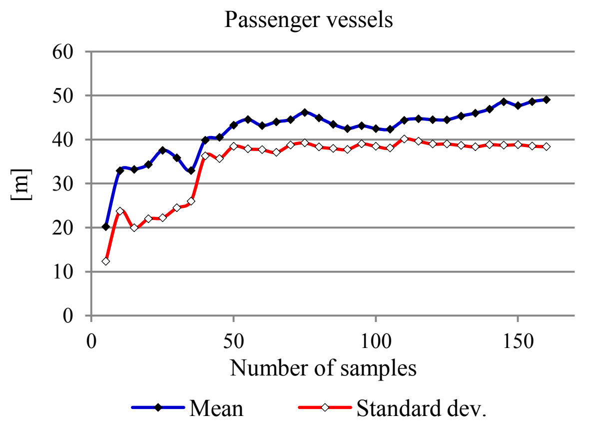

Fig. 7. Mean and standard deviation in relations to the sample size. Passenger vessel, northbound, winter months

Table 1. Sample size for a series of ships crossing determined by method I

Series	Sample size	
	d = 10 m	d = 20 m
1 cargo N 1	55	30
2 cargo N 2	50	35
3 cargo S 1	60	10
4 cargo S 2	40	15
5 pas. N 1	50	20
6 pas. N 2	40	15
7 pas. S 1	70	10
8 pas. S 2	35	10

where: 1 – summer months; 2 – winter months; N – northbound vessel, S – southbound vessel.

Determining minimum sample size by method II for level of confidence $\alpha = 0.05$, pilot sample size $n' = 50$ and variability error $d = 10$ m and $d = 20$ m is possible to obtain the minimum number for the corresponding series (Tab. 2).

Table 2. Sample size for a series of ships crossing determined by method II

Series	Sample size	
	d = 10 m	d = 20 m
1 cargo N 1	136	34
2 cargo N 2	148	37
3 cargo S 1	39	10
4 cargo S 2	66	17
5 pas. N 1	69	17
6 pas. N 2	56	14
7 pas. S 1	32	8
8 pas. S 2	59	15

Analyzing results can be stated that:

1. The minimum sample sizes determined by method II are comparable to the sample sizes defined by method I in the series with maximum error $d = 20$ m.
2. The minimum sample sizes determined by the second method are greater the sample size specified by I method.
3. Series 1 and 2 showed a large deviation for both methods for maximum error $d = 10$ m.

Conclusions

The presented analysis of the issue proves that there exists a minimum sample size, which can ensure correct estimation of the spatial distribution of ships traffic for a specified level of confidence. The discussed methods of estimating sample size do not give unequivocal numerical values yet they provide approximate values. It must be pointed out that the choice of a method for determining the minimum sample size depends on the goal that a statistical analysis is aiming to achieve.

The minimum sample size in traffic flow studies based on AIS data are comparable for both methods in the series with maximum error $d = 20$ m. Sample sizes determine by II method are greater for a series with well stabilized parameters.

Minimum sample sizes for cargo vessels are greater than for passenger ships. This results from ferries movements which usually carry out on the same routs. A comparison of the effect of sample size n on the values of basic distribution parameters of a process is presented in table 3.

Column 2 shows mean values of the observed ships position in relations to the axis of the track for eight successive series. In columns 3 and 4, mean values of the observed dimension deviation have been tabulated for the same conditions, but for samples of size $n = 20$ and $n = 50$. Column 5 contains standard deviation values for all units of the series, and columns 6 and 7, the corresponding estimators of standard deviation determined for samples of $n = 20$ and $n = 50$ units.

Table 3. A collation of the determined statistical parameters of the distribution for different sample sizes n

Sample No.	Mean value			Standard deviation		
	n max	n = 20	n = 50	n max	n = 20	n = 50
1	42.94	42.20	52.62	49.80	76.10	60.14
2	42.00	61.45	44.60	54.00	81.90	62.68
3	34.90	45.50	40.84	54.53	38.83	32.33
4	30.10	37.90	31.40	45.70	43.50	41.75
5	52.30	37.90	47.64	38.80	32.53	42.84
6	49.00	34.40	43.24	38.38	22.23	38.45
7	56.60	65.50	65.36	33.20	30.39	29.14
8	52.40	63.15	57.50	36.111	16.88	39.56

The goodness of fit is first determined by performing a Chi-square test (χ_2). This test determines the degree of agreement between the empirical distribution and the theoretical distribution. Logistic distribution describes well the distribution of the analyzed variables.

On the basis of obtained results can be stated that sample size for estimation of standard deviation and mean value suggest that the sample size for estimation of the average and standard deviation was sufficient and the obtained results with available data can be considered highly reliable.

It should be noted that the distribution applies solely to the AIS registered ship traffic. Thus, the leisure boats and fishing boats will be modeled separately to form a complete ship traffic distribution.

References

1. PUSZCZ A., GUCMA L.: Towards the Model of traffic Flow on the Southern Baltic Based on Statistical Data. Advances in Marine Navigation and Safety of Sea Transportation TransNav'2011, Gdynia 2011.
2. FRANKLIN L.A.: Sample size determination for lower confidence limit for estimating process capability indices. Computers & Industrial Engineering, 36, 1999.
3. STANISZ A.: Przystępny kurs statystyki. Tom I, II. StatSoft, Kraków 2007.
4. GUZIEWICZ J., ŚLĄCZKA W.: Methods for determining the manoeuvring area of the vessel used in navigating simulation studies. VII MTE Conference, Szczecin 1997.
5. GUCMA L.: Width of a ship's swept path – a safe manoeuvre criterion defined with the tolerance intervals concept. Research bulletin, Maritime University, Szczecin 2000.
6. The Royal Danish Administration of Navigation and Hydrography, The Danish Maritime Authority and The Swedish Maritime Administration. Navigational safety in the Sound between Denmark and Sweden (Øresund), 2006.
7. TUMIDAJSKI T., SARAMAK D.: Metody i modele statystyki matematycznej w przeróbce surowców mineralnych. Wydawnictwo AGH, Kraków 2009.
8. BALICKI A., MAKAĆ W.: Metody wnioskowania matematycznego. Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2004.