

# 주성분 분석기법을 이용한 선박의 연료소비 예측에 관한 연구

김영롱\* · 김구종\*\* · † 박준범

\*한국해양대학교 대학원 석사과정생, \*\*해양안전심판원 조사관, † 한국해양대학교 항해학부 교수

## A Study on the Prediction of Fuel Consumption of a Ship Using the Principal Component Analysis

Young-Rong Kim\* · Gujong Kim\*\* · † Jun-Bum Park

\*Student, Graduate School of Korea Maritime and Ocean University, Busan 49112, Korea

\*\*Investigator, Korea Maritime Safety Tribunal, Sejong 30121, Korea

† Professor, Div. of Navigation Science, Korea Maritime and Ocean University, Busan 49112, Korea

**요 약** : 최근 선박의 배기가스 규제가 강화되면서 연료소비량을 저감하기 위한 많은 방안들이 검토되고 있다. 그중에서도 선박으로부터 수집한 데이터를 활용하여 연료소모량을 예측하는 기계학습 모델을 개발하고자 하는 연구가 활발히 수행되고 있다. 하지만 많은 연구들이 학습 모델의 주요 변수 선정이나 수집데이터의 처리 방법에 대한 고려가 미흡하였으며, 무분별한 데이터의 활용은 변수 간의 다중공선성 문제를 야기할 수도 있다. 본 연구에서는 이러한 문제점을 해결하기 위하여 주성분 분석을 이용하여 선박의 연료소비를 예측하는 방법을 제시하였다. 13K TEU 컨테이너 선박의 운항데이터에 주성분 분석을 수행하였으며, 추출한 주성분으로 회귀분석을 수행하여 연료소비 예측모델을 구현하였다. 평가용 데이터에 대한 모델의 설명력은 82.99%이었으며, 이러한 예측모델은 항해 계획 수립 시 운항자의 의사결정을 지원하고 항해 중 에너지 효율적인 운항상태 모니터링에 기여할 수 있을 것으로 기대된다.

**핵심용어** : 선박 연료소비량, 에너지 효율, 주성분 분석, 예측모델, 운항데이터

**Abstract** : As the regulations of ship exhaust gas have been strengthened recently, many measures are under consideration to reduce fuel consumption. Among them, research has been performed actively to develop a machine-learning model that predicts fuel consumption by using data collected from ships. However, many studies have not considered the methodology of the main parameter selection for the model or the processing of the collected data sufficiently, and the reckless use of data may cause problems such as multicollinearity between variables. In this study, we propose a method to predict the fuel consumption of the ship by using the principal component analysis to solve these problems. The principal component analysis was performed on the operational data of the 13K TEU container ship and the fuel consumption prediction model was implemented by regression analysis with extracted components. As the R-squared value of the model for the test data was 82.99%, this model would be expected to support the decision-making of operators in the voyage planning and contribute to the monitoring of energy-efficient operation of ships during voyages.

**Key words** : Ship Fuel Consumption, Energy Efficiency, Principal Component Analysis, Prediction Model, Ship Operational Data

### 1. 서 론

최근 해상 운송에 의한 배기가스 배출규제가 강화되고 있어 선박의 연료소모량 관리가 해운 사회의 주요한 관심사 중 하나로 자리 잡고 있다. 국제해사기구(International Maritime Organization;IMO)의 해양환경보호위원회(Marine Environment Protection Committee;MEPC)에서는 에너지효율 설계 지수(Energy Efficiency Design Index;EEDI)를 도입하여 선박의 설계 및 건조 시부터 탄소가스 배출을 개량 및 감축하고자 하고 있으며, 운항 중인 선박에 대해서는 에너지효율 운항 지수(Energy Efficiency Operational Index;EEOI) 및 선박 에너지

효율 관리 계획(Ship Energy Efficiency Management Plan;SEEMP)을 적용하여 해상에서 선박 운항자에 의한 에너지효율 관리 및 개선을 독려하고 있다 (IMO, 2009; 2012a; 2012b). 그뿐만 아니라 선박의 크기, 종류 및 운항 노선 등에 따라 차이가 있을 수 있지만, 선박의 연료비용은 전체 운항 경비의 약 50~60%를 차지하는 것으로 알려져 있어 선박을 운영하는 해운회사들도 자발적으로 운영 절차를 개발하여 관리 비용을 절감하고 시장경쟁력을 강화하기 위한 노력들을 하고 있는 중이다(Stopford, 2009).

빅데이터의 수집과 저장 그리고 통신 기술의 발달로 인하여 선박의 운항데이터를 활용하여 에너지효율을 향상시키고

† Corresponding author : 종신회원, jbpark@kmou.ac.kr 051)410-4233

\* 정회원, yr\_kim@kmou.ac.kr

\*\* 종신회원, master114@korea.kr

자하는 연구들이 많아지고 있다. 특히 선박의 실시간 운항상태와 해역의 기상정보와 같은 여러 조건을 활용하여 소모되는 연료량을 예측할 수 있는 학습모델을 개발하고자 하는 시도들이 있었으며, 이를 위하여 실험식이나 통계적인 분석뿐만 아니라 기계학습 알고리즘을 접목하기도 하였다. Pedersen and Larsen(2009)은 선박의 적하상태에 따른 선속, 풍향, 풍속, 해수온도, 대기 온도 및 기상예보 데이터를 인공지능경망으로 학습하여 추진출력을 예측하고자 하였으며, Kim et al.(2017)은 컨테이너선박의 운항데이터에 부분최소제곱회귀를 적용하여 선박의 연료소비패턴을 파악하는 방법을 제안하였다. 또한 Yan et al.(2018)은 항해구간에 따른 최적의 선속을 결정하고자 k-means 군집 알고리즘을 적용하여 선박의 항로를 분할하였으며, 기존의 실험식을 바탕으로 연안페리선의 진저항 및 환경 요소의 영향을 계산하여 에너지효율 최적화 모델을 정립하였다. 이러한 선박 운항데이터를 바탕으로 구현한 예측모델들은 상당한 예측 성능을 보이기도 하였으나, 적절한 분석이나 처리 없이 무분별하게 데이터를 활용하는 경우 변수들 간의 높은 상관성으로 인해 다중공선성(Multicollinearity) 문제가 발생할 수 있다.

본 연구에서는 이러한 문제점을 해결하고 효율적인 데이터 분석 및 처리를 수행하기 위하여 주성분 분석(Principal Component Analysis; PCA)을 활용하고자 한다. 주성분 분석은 전체 데이터의 분산을 대표하는 서로 독립적인 성분을 추출해주는 기법이기 때문에 다중공선성 문제를 해결하는 데 효과적이며, 많은 데이터를 실시간으로 송수신할 필요가 있는 선박과 같은 환경에서 주성분으로 변환된 데이터를 활용하면 상대적으로 적은 양의 데이터를 송수신하여도 효과적인 결과를 얻을 수 있을 것으로 판단된다. 따라서 실제 13K TEU 컨테이너 선박으로부터 수집한 운항데이터를 사용하여 주성분 분석을 수행하고, 추출한 주성분을 회귀 모형에 적용하여 선박의 연료소비 예측모델을 구현한다. 결과적으로 얻은 모델에서 구한 예측 결과를 실제 관측데이터와 비교·분석함으로써 모델의 효용성을 검토하고자 한다.

## 2. 선박 운항데이터의 처리

### 2.1 데이터 처리 절차

선박으로부터 수집한 운항데이터로부터 연료소비 예측모델을 구현하기 위하여 Fig. 1과 같은 절차에 따라 연구를 수행하였다. 수집단계에서는 원시 데이터를 취득하며, 전처리 단계에서는 원시 데이터를 통합, 정제, 변환 및 축소과정을 통해서 분석에 적합한 상태가 되도록 처리한다. 데이터 통합 과정은 원시 데이터를 취합하여 분류하고 통일된 형식의 파일 형태로 변환한다. 정제 단계에서는 데이터의 특성을 고려하여 이상치 및 결측값 등을 통계기법을 이용하여 정제한다. 변환 단계에서는 기존의 변수를 조합하여 새로운 변수로 변환하거나 새로

운 속성을 가진 데이터로 변형하여 불필요한 변수의 개수를 줄이고 계산 복잡성을 해소한다. 데이터 축소는 주성분 분석을 수행하여 고차원의 데이터를 저차원의 데이터로 축소한다. 분석단계에서는 회귀분석을 수행하여 선박의 연료소비 예측 모델을 개발하며, 마지막으로 평가단계에서는 학습에 사용하지 않은 데이터 세트를 활용하여 선박의 연료소비를 예측함으로써 모델의 성능을 파악하고자 한다.

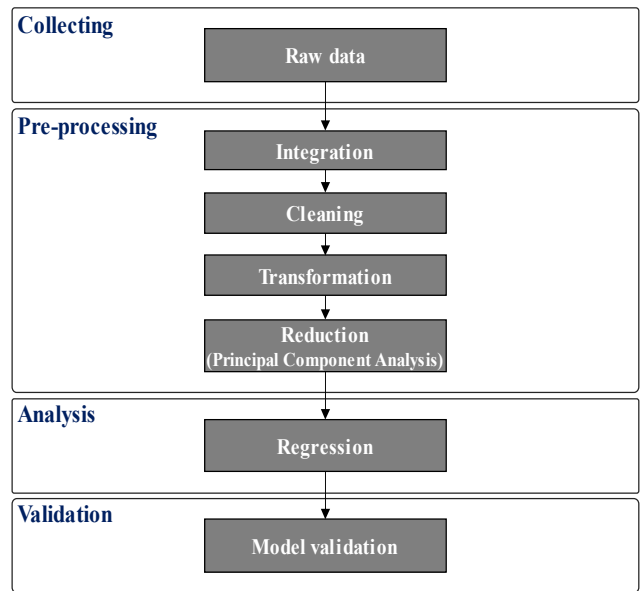


Fig. 1 Flow chart of the study

### 2.2 대상 선박 및 수집 데이터

본 연구에서는 Table 1과 같은 제원을 가진 13K TEU급 컨테이너선을 대상으로 연구를 수행하였으며, 2014년 1월부터 약 6개월간 대상 선박의 감시 제어 시스템(Alarm Monitoring and control System; AMS)으로부터 1분 간격으로 운항데이터를 수집하였다. 선박 운항자의 의사결정을 지원할 수 있는 연료소비 예측모델을 개발하기 위하여 수집된 데이터 중 실질적으로 본선에서 조정 가능한 운항변수, 선박의 침로, 속도, 흘수, 트림 그리고 기상조건에 대한 변수 등을 Table 2와 같이 선택하였으며, 선박의 운항과 직접적인 관련이 있는 주기관의 연료소모량에 대해서만 분석을 수행하였다. 해당하는 기간 동안 대상 선박은 Fig. 2와 같은 항로를 운항하였으며, 선박의 운항 중 에너지효율을 분석하기 위하여 항내 전속 전진 시 주기관의 분당 회전수(M/E RPM)인 50 미만의 데이터 세트는 연구에서 제외하였다. 약 6개월 간의 운항데이터를 항해중(대양), 입출항 및 정박(항내), 결측 구간으로 구분한 결과 각각 3개월 17일, 29일, 1개월로 분류할 수 있었다.

Table 1 Principal particulars of the target ship

Particular	Target Ship
LOA [m]	360.0
LBP [m]	345.0
Breadth [m]	46.0
Moulded depth [m]	28.0
Summer draft [m]	15.5
Deadweight [ton]	142,000.0
Displacement [ton]	185,000.0

Table 2 Data list collected from the target ship

No.	Variable	Unit	Remark
1	M/E RPM		
2	Speed of the ground	knot	
3	Speed through water	knot	
4	Course of the ground	degree	
5	Ship's Heading	degree	Compass direction in which the ship's bow is pointed
6	True wind speed	knot	
7	True wind direction	degree	The direction from which wind originates
8	Rudder angle	degree	
9	Mean draft	meter	
10	Trim	meter	(+)/(-) : Trim by the stern/head
11	Displacement	ton	
12	Wetted surface area	m <sup>2</sup>	
13	Shaft power	kW	
14	M/E fuel consumption	ton/h	

2.3 데이터 처리

선박의 특성상 진동이나 충격 또는 외부 환경적인 요인 등에 의하여 일반적인 작동 범위를 벗어나는 이상값이 관측될 수 있으며, 기기의 고장, 정비 또는 안전상의 이유 등으로 인하여 취득 데이터 구간이 일부 결측되는 경우도 발생한다. 이처럼 실제 경향을 벗어나는 데이터가 포함되거나 불연속적인 데이터가 발생한 구간은 추후 분석 결과에 영향을 미칠 수 있으므로 데이터의 특성에 대한 충분한 이해를 바탕으로 다음과 같은 방법을 적용하여 선박으로부터 수집한 원시데이터를 적절히 처리하였다.

1) 수집된 운항 데이터 중 결측 구간에 대해서는 해당하는 데이터 세트를 제거하였다. 본 연구에서는 취득 데이터의 양이 상대적으로 많았으며, 결측값이 특정한 항해구간이 아니라 무작위로 발생하였기 때문에 일부 결측 구간을 제외하더라도 데이터를 분석하는 데 문제가 없었다. 하지만 수집된 데이터의 양이 한정적이거나 동적인 특성에 민감한 데이터의 경우 또는 특정한 운항 구간에서 결측이 많이 발생하는 경우는 적절한 전처리가 필요하며 각별한 주의가 필요할 것으로 보인다.

2) 매뉴얼에 명시된 기기의 정상 작동 범위를 크게 벗어나거나 일정 시간 동안 변수의 관측값이 일정하게 유지되는 경우는 데이터 수집 기기의 오류로 간주하여 해당하는 데이터 세트를 제거하였다.

3) 운항변수 중 선박의 기관출력과 연료소모량의 선형 관계를 활용하여 이상값을 식별하였다. 두 변수의 회귀분석에 의한 예측값과 관측값의 차가 정규분포를 이룬다는 가정하에 평균으로부터 3 표준편차를 벗어난 범위의 관측값을 이상값으로 판단하여 원시 데이터로부터 제거하였다. Fig. 3은 기관출력에 따른 연료소모량의 산점도를 나타낸 것으로써 실선은 두 변수의 선형 회귀선, 원은 회귀선을 기준으로 3 표준편차 이내의 값, 별표는 3 표준편차 이외의 값에 해당한다. Fig. 4는 Fig. 3에서 이상값으로 식별된 데이터 중에서 샘플링 지점(Sampling point)의 관측값을 분석한 것이다. 시간에 따른 기관 출력과 연료소모량의 변화를 볼 때 샘플링 지점에서 과도한 값이 관측된 것을 알 수 있다. 전체 데이터 중 결측값을 제외한 항해 구간은 3개월 17일이며, 2)와 3)의 이상값 처리 과정을 수행한 후의 데이터는 항해 구간의 98%에 해당한다.



Fig. 2 Operational route of the target ship

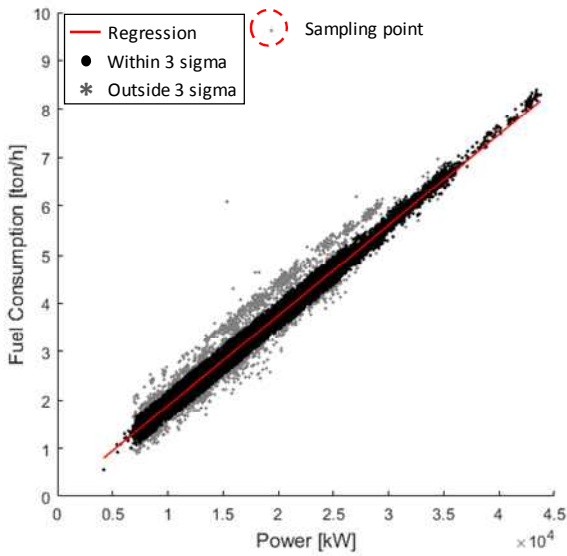


Fig. 3 Outlier detection by the relationship between engine power and fuel consumption

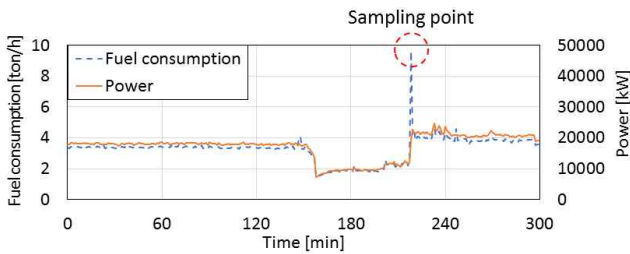


Fig. 4 Time series data of sampling point identified as an outlier

4) 데이터에 포함된 잡음을 처리하기 위하여 중앙값 필터를 적용하였다. 중앙값 필터는 관측값의 주변값들을 오름 또는 내림차순으로 정렬하여 해당하는 관측값을 정렬의 중앙값으로 대체하는 방법이며, 이는 과도한 잡음 신호를 처리하고 데이터의 전반적인 추세를 표현하는 데 효과적이다(Pratt, 2007). Figs. 5-6은 선박의 운항데이터인 평균흡수와 트림에 중앙값 필터를 적용한 결과를 예시로 나타낸 것이다. 파선이 필터를 적용하기 전의 원시 데이터이며 실선으로 나타낸 부분이 10분 간격으로 중앙값 필터를 적용한 결과이다. 평균흡수나 트림의 경우 실제로 선박의 동요 주기나 외력 등에 의하여 지속적으로 값이 변화하는 특성을 가지나 취득 데이터가 1분 간격의 평균값이 아닌 취득 시점의 순간 데이터로 수집되었기 때문에 데이터가 전반적으로 불안정하며 일정한 추세변화를 나타내는데 어려움이 존재하였다. 이러한 이유로 중앙값 필터를 적용하여 데이터의 잡음 및 경향성에 대한 문제점을 보완하고자 하였다.

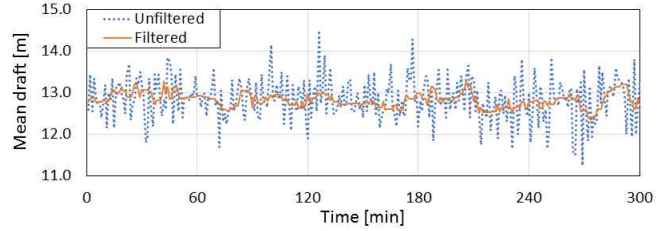


Fig. 5 Time series data of mean draft filtered by median filter

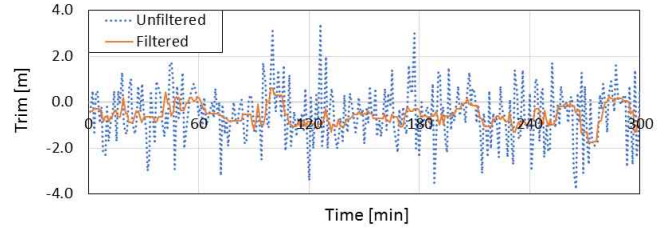


Fig. 6 Time series data of trim filtered by median filter

5) 기존의 변수를 조합하거나 관계를 활용하여 종속변수에 물리적으로 더욱 의미가 있는 변수로 변환하였으며, 불필요한 변수의 개수를 줄여 계산 복잡성을 해소시킬 수 있었다. 즉, 선박의 속력과 선수 방위 그리고 진풍속과 진풍향의 관계로부터 식 (1), (2)와 같이 선박이 조우하는 바람의 상대 방위와 상대 속도를 계산함으로써 연료소모량에 대한 바람 성분의 영향을 직접적으로 반영하였다. 또한 동일한 각도의 경우 좌현과 우현에서 바람의 영향이 같기 때문에 선수방위를 기준으로 0-180도까지의 상대적인 방위를 고려하였다.

$$A = \sqrt{W^2 + V^2 + 2WV\cos\alpha} \quad (1)$$

$$\beta = \arccos\left(\frac{W\cos\alpha + V}{A}\right) \quad (2)$$

여기에서  $V$ 는 선박의 속력,  $W$ 는 실제 바람의 속력,  $\alpha$ 는 선박의 선수방위를 기준으로 계산한 실제 바람의 방향,  $A$ 는 겉보기 바람의 속력,  $\beta$ 는 겉보기 바람의 방향이다.

기관출력은 식 (3), (4)와 같이 주기관의 분당 회전수, 토크 등으로 계산할 수 있기 때문에 중복적인 변수의 사용을 방지하기 위하여 기관 출력과 분당 회전수 중 하나를 제외하는 것이 바람직할 것으로 판단하였다. 기관 출력의 경우 선박 운항자가 항해 중 즉각적으로 확인하기에는 용이하지 못하며 통상적으로 주기관의 분당 회전수를 이용하여 선박의 속력을 조정하기 때문에 주기관의 분당 회전수를 활용하였다.

$$\omega = \frac{2\pi RPM}{60} \quad (3)$$

$$\begin{aligned} \text{Shaft power} &= T w & (4) \\ &= \frac{2\pi T RPM}{60} \end{aligned}$$

여기에서  $RPM$ 은 주기관의 분당 회전수,  $w$ 는 프로펠러축의 각속도,  $T$ 는 토크이다.

선행 연구에서는 단위 시간당 연료소모량을 종속변수로 하여 예측모델을 생성하는 경우가 많았으나, 항해사 입장에서 보면 주어진 외부환경에서 선박이 일정 거리를 이동하는 동안 소모되는 연료량을 지표로 제공해야 에너지 효율적인 운항상태를 파악하고 연료소비를 저감하는 데 도움이 될 것이다 (Petersen et al., 2012; Kim et al., 2017). 따라서 식 (5)와 같이 단위 항해 거리당 연료소모량을 계산하여 선박의 연료효율을 판단하고자 하였다.

$$\text{Fuel consumption rate} = \frac{\text{Fuel consumption per hour}}{\text{Navigation distance per hour}} \quad (5)$$

Table 3은 데이터 정제과정 및 변수 변환을 수행한 후의 각 운항변수에 대한 기술통계량을 나타낸 것이다. 예측모델의 독립변수는 Table 3의 1~10번에 해당하는 주기관의 분당 회전수, 대지속력, 대수속력, 상대 풍속, 상대 풍향, 타각, 평균흘수, 트림, 배수량, 선체 침수표면적이며, 종속변수는 Table 3의 11번인 단위 항해 거리당 연료소모량이다.

Table 3 Descriptive statistics of operational variables

No.	Variable Name	Min	Max	Average	Standard deviation
1	M/E RPM	50.0	90.0	63.5	7.5
2	Speed of the ground [knot]	8.0	22.4	14.7	2.0
3	Speed through water [knot]	7.4	21.8	14.5	1.9
4	Relative wind speed [knot]	0.0	72.9	19.6	10.2
5	Relative wind direction [degree]	0.0	180.0	43.1	43.5
6	Rudder angle [degree]	0.0	36.1	1.6	2.4
7	Mean draft [meter]	11.3	15.7	14.3	0.9
8	Trim [meter]	-2.3	2.2	-0.1	0.6
9	Displacement [ton]	123466.4	184009.6	163225.1	13240.0
10	Wetted surface Area [m <sup>2</sup> ]	13108.0	14514.3	14047.8	292.9
11	Fuel consumption rate [ton/mile]	0.1	0.5	0.2	0.1

6) 각 변수는 일반적으로 서로 다른 단위체계를 사용하고 있기 때문에 추후 회귀분석에 의한 계수를 추정하더라도 종속 변수에 대한 독립변수들의 영향력을 파악하기 어려운 경우가 발생한다. 이를 방지하기 위해 식 (6)과 같은 z값 변환을 수행하여 데이터를 표준화하고자 하였으며, 표준화된 변수  $z_i$ 는 평균이 0이고 표준편차가 1인 특징을 갖는다.

$$z_i = \frac{x_i - \mu}{\sigma}, \forall i = 1, 2, \dots, p \quad (6)$$

여기에서  $x_i$ 는 각 변수의  $i$ 번째 관측값,  $\sigma$ 는 표준편차,  $\mu$ 는 평균,  $z_i$ 는 표준화된 값,  $p$ 는 각 변수의 관측값 개수이다. 본 연구에서  $p$ 는 6개월간 1분 간격으로 취득한 데이터에 해당한다.

### 3. 주성분 분석 기반의 선박 연료소비 예측모델

#### 3.1 주성분 분석

주성분 분석은 변수들 사이의 분산-공분산 관계를 이용하여 변수들의 선형결합으로 나타낼 수 있는 주성분을 찾는 방법이다. 전체 데이터의 분산을 대표할 수 있는 주성분을 추출하기 때문에 데이터의 차원을 감소시켜 해석을 용이하게 하는 역할을 하며 서로 독립적인 새로운 변수들로 주성분을 구성하여 정보 손실의 최소화 및 다중공선성이 의심되는 회귀 모형에도 응용할 수 있다 (Pearson, 1901). 다음은 주성분 분석을 통하여 데이터의 차원을 축소하는 방법을 나타낸 것이다.

변수가  $p$ 개, 샘플의 수가  $n$ 개로 이루어진 데이터  $X(x_1, x_2, \dots, x_p)$ 에 주성분 분석을 수행하여 새로운 변수  $D(d_1, d_2, \dots, d_p)$ 를 생성한다고 가정하면, 식 (7), (8)과 같이 입력 벡터와 평균 벡터의 차를 통해  $X$ 의 공분산 행렬 (Covariance matrix)을 구할 수 있다. 식 (9)와 같이 고유벡터의 정의에 의하여 공분산 행렬의 고유값(Eigen value,  $\lambda$ )과 고유 벡터(Eigen vector,  $A$ )를 구하며, 여기서 고유벡터는 샘플 데이터의 분산이 가장 큰 방향을 나타내고 각각의 고유값은 해당하는 방향으로 축을 변환했을 때 분산의 크기를 나타낸다. 새로운 변수  $D$ 는 식 (10)과 같은 관계를 가지며, 고유값  $\lambda$ 를  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ 와 같이 크기 순서대로 정렬하고 대응하는 고유벡터를 구하면  $D$ 의 요소인  $k$ 번째 주성분  $d_k$ 는 식 (11)과 같이 나타낼 수 있다. 즉, 주성분 분석은 Fig. 7과 같이 원데이터  $X$ 의 분산을 최대화해주는 최적의 축으로 투영하여 데이터의 차원을 그보다 낮은 차원의 공간으로 축소하는 것을 의미한다.

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (7)$$

$$C = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T \quad (8)$$

$$\det(C - \lambda E) = 0 \quad (9)$$

$$D = AX \quad (10)$$

$$d_k = \sum_{j=1}^p a_{kj} x_j = a_{k1} x_1 + a_{k2} x_2 + \dots + a_{kp} x_p \quad (11)$$

큰 고유값을 가지는 주성분은 입력 데이터와의 상관관계가 상대적으로 높은 특징을 갖는 것을 의미하며 설명력이 높은 주성분만을 선택하여 차원축소를 수행할 필요가 있다. 주성분의 개수를 선정하기 위하여 통상적으로 고유값의 크기, 전체 변동에 대한 공헌도, 스크리 도표와 같은 지표들이 활용되며, 선행 연구에서는 고유값의 크기 0.7이상, 전체 변동에 대한 공헌도는 80~90%, 고유값의 감소 추세가 완만해지는 시점 등을 기준으로 제안한 바 있다(Cattell, 1966; Jolliffe, 1982).

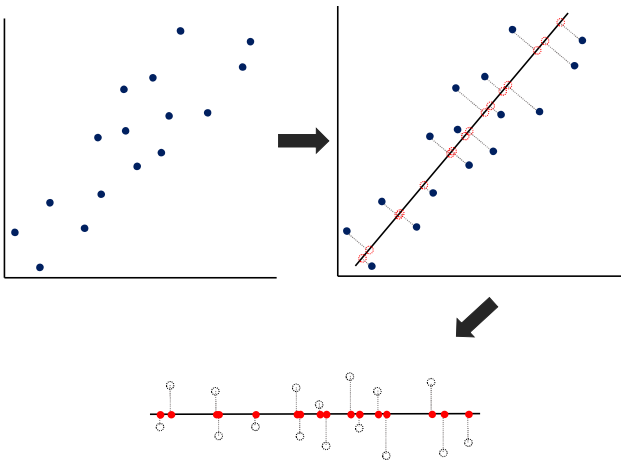


Fig. 7 Linear transformation using principal component analysis

### 3.2 주성분 분석을 활용한 운항 변수 간의 관계 파악

선박의 운항변수와 연료소비 간의 관계를 파악하기 위하여 독립변수와 종속변수를 포함한 전체 운항데이터에 대한 주성분 분석을 수행하였다. Fig. 8은 주성분 분석을 수행한 결과로 주성분 개수에 따른 고유값과 누적 분산값을 나타낸 것이다. 제4주성분까지의 고유값은 각각 4.691, 2.499, 1.503, 1.135이며 5번째 주성분부터 그 값이 확연히 떨어지는 것을 알 수 있다. 또한 제4주성분까지의 누적분산비율이 전체 분산의 89.3%를 설명할 수 있으므로 4번째 주성분까지를 추출하여 분석하였다.

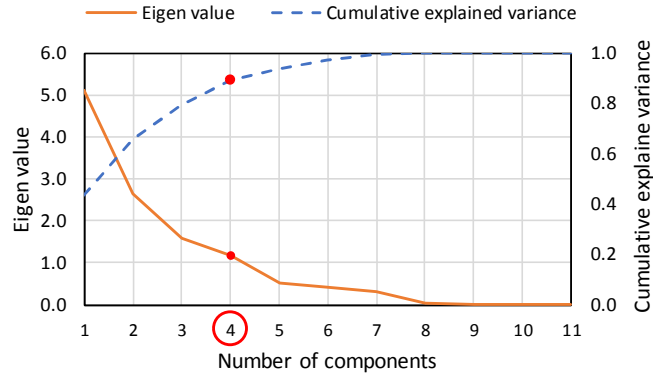


Fig. 8 Eigen values and cumulative variances corresponding to the number of components(all variables)

Fig. 9는 주성분 분석을 수행하여 얻은 각 변수들의 주성분 점수를 원의 크기와 명도를 활용하여 이해하기 쉽도록 나타낸 것이다. 가장 바깥쪽의 원이 제1주성분을 구성하고 있는 변수들의 주성분 점수이며, 가장 안쪽이 제4주성분에 해당한다. 원의 크기는 주성분 점수의 절대값에 비례하며, 0을 기준으로 점수가 양으로 증가할수록 명도가 밝아지고 음으로 증가할수록 명도가 어두워진다.

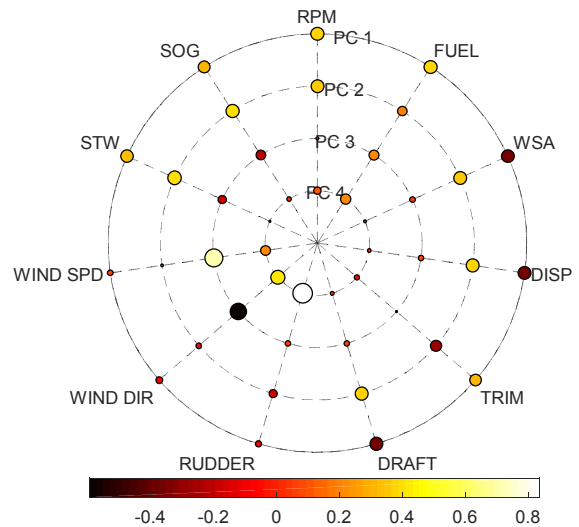


Fig. 9 Score plot of each variable according to the principal component

Fig. 9의 주성분 분석으로부터 파악한 선박의 운항 변수와 연료소비 간의 관계는 다음과 같다.

- 1) 제1주성분 : 선박 주기판의 분당 회전수가 증가하면 선박의 대지속력과 대수속력이 증가하며 단위 항해 거리당 연료소모량이 증가한다. 즉, 선속이 높을수록 연료효율이 감소하는 경향이 있다.
- 2) 제2주성분 : 선박의 평균흘수가 증가하면 배수량 및 선체 수선면 하부의 표면적이 증가하고 단위 항해 거리당 연료소모량이 증가한다. 이는 흘수 증가로 인하여 선박의 저항이

증가하기 때문에 연료효율을 감소시키는 역할을 하는 것으로 판단된다. 또한 평균홀수가 증가하면 트림이 감소하는 경향이 있으며 이는 홀수 증가분을 보완하기 위한 선박에서의 조정으로 보인다.

3) 제3주성분 : 상대 풍향이 감소하면 상대 풍속이 증가하고 단위 항해 거리당 연료소모량이 증가한다. 상대 풍향과 상대 풍속은 바람과 선박의 벡터 합성으로 계산된다. 따라서 상대 풍향의 감소는 선박이 강한 선수 방향 바람을 조우하거나 고속으로 항해하는 경우에 해당하며, 이는 연료효율을 감소시키는 역할을 한다.

4) 제4주성분 : 선박에서 사용하는 타각이 커질수록 단위 항해 거리당 연료소모량은 증가한다. 선박이 운항 중 타를 사용하게 되면 타판에 작용하는 항력에 의해서 속도를 저하시키는 역할을 하며, 이는 연료효율을 감소시킨다.

이러한 주성분 분석을 활용한 변수 간의 상관관계는 센서의 오류 식별에도 활용이 가능하다. 각 운항변수들의 시계열 데이터를 비교·분석하여 주성분 분석으로부터 파악한 유의미한 경향을 벗어나는 구간은 수집된 데이터의 이상으로 의심해볼 필요가 있다(Perera & Mo, 2016).

3.3 주성분 분석을 활용한 회귀분석 모델 구현

연료소비 예측모델에 사용될 입력변수를 추출하기 위하여 종속변수인 연료소모율을 제외한 나머지 독립변수들에 대하여 주성분 분석을 수행하였다. 전체 운항데이터 세트 중 70%를 예측모델의 구현을 위한 학습용 데이터, 나머지 30%는 모델의 성능을 평가하기 위한 평가용 데이터로 활용하였다. Fig. 10은 주성분의 개수에 따른 고유값과 누적 분산값을 나타낸다. 제3주성분까지는 고유값이 각각 3.315, 2.617, 1.038이며, 제4주성분부터는 0.621로 감소하는 폭이 다소 줄어드는 것을 알 수 있다. 누적 분산 비율을 고려하면 제3주성분은 전체의 약 83.0%, 제4주성분까지는 약 90.4%를 설명할 수 있어 운항 데이터를 충분히 설명하기 위하여 본 연구에서는 제4주성분까지를 예측모델의 입력변수로 추출하였다.

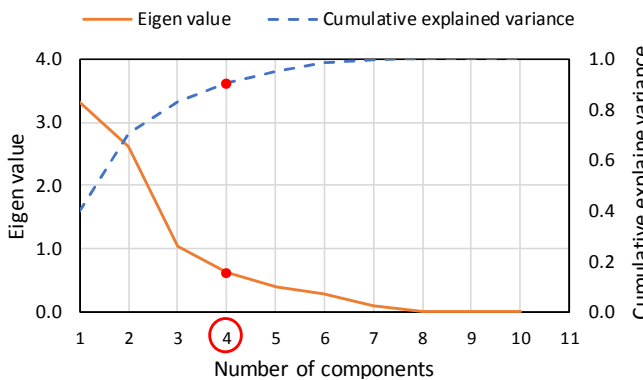


Fig. 10 Eigen values and cumulative variances corresponding to the number of components (independent variables only)

제4주성분까지의 각 주성분에 대한 운항변수들의 주성분점수를 Table 4와 같이 나타내었다. 각 주성분을 구성하는 변수들을 분석해보면 제1주성분은 주기관 분당 회전수, 대지속력, 대수속력이 0.453, 0.436, 0.440의 주성분 점수를 가지고 있어 선박의 속도에 영향을 미치는 추진성분을 대표하는 것을 알 수 있다. 제2주성분은 평균홀수, 트림, 배수량 및 침수표면적의 주성분 점수가 0.417, -0.215, 0.420, 0.431이며 선박의 하중과 관련된 인자를 의미한다. 제3주성분은 상대 풍속과 상대 풍향의 주성분 점수가 0.836, -0.486으로 선박에 작용하는 외력 성분을 나타내고 있음을 알 수 있다. 마지막으로 제4주성분은 타각성분을 의미한다. Fig. 11은 이러한 주성분의 구성을 도식화하여 나타낸 것이다.

Table 4 Principal component scores of each variable

	PC 1	PC 2	PC 3	PC 4
M/E RPM	0.453	0.323	0.126	0.141
Speed of the ground	0.436	0.379	-0.111	0.033
Speed through water	0.440	0.386	-0.041	0.027
Relative wind speed	-0.007	-0.019	0.836	0.406
Relative wind direction	0.022	-0.074	-0.486	0.454
Rudder angle	-0.048	-0.127	-0.173	0.771
Mean draft	-0.352	0.417	-0.012	0.058
Trim	0.144	-0.215	0.068	0.035
Displacement	-0.354	0.420	-0.012	0.056
Wetted surface Area	-0.372	0.431	0.000	0.070

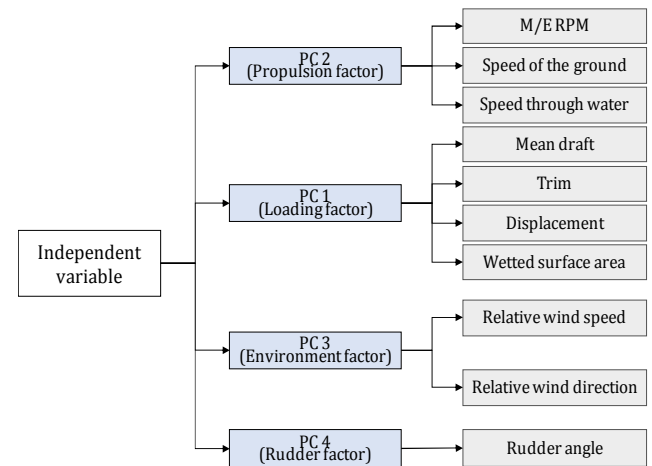


Fig. 11 Analysis of principal components constituting independent variables

주성분 분석을 통해 10개의 운항변수를 선박의 추진성분, 하중성분, 외력성분 및 타각성분을 설명해주는 4개의 주성분

으로 축약하였고, 이를 이용하여 식 (11)과 같은 회귀분석 모델을 구현하였다.

$$\begin{aligned}
 \text{Fuel consumption rate} = & 0.0160PC1 + 0.0107PC2 \quad (11) \\
 & + 0.0119PC3 + 0.0187PC4 \\
 & + 0.2058
 \end{aligned}$$

### 3.4 예측성능 평가 결과

예측모델의 평가 기준은 식 (12), (13)과 같은 평균 제곱근 오차와 조정 결정계수를 적용하였다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

$$Adjusted R^2 = 1 - \frac{SSE \times (n-1)}{SST \times (n-d-1)} \quad (13)$$

여기에서  $y_i$ 는  $i$ 번째 종속변수의 관측값,  $\hat{y}_i$ 는  $i$ 번째 종속변수의 예측값,  $\bar{y}$ 는 관측값의 평균,  $n$ 은 관측값의 개수,  $d$ 는 모델에 사용된 독립변수의 개수,  $SST$ 는 총제곱합,  $SSE$ 는 오차제곱합이다.

Moore and Kirkland(2007)에 의하면 회귀분석의 결정계수 값이 0.7 이상인 경우 추정된 모형이 주어진 자료를 강력하게 설명할 수 있으며, 통상적으로 자연과학이나 공학 분야에서는 0.7 이상의 결정계수를 추천하고 있다. 식 (12)와 식 (13)의 기준을 적용하여 전체 평가용 데이터에 대한 예측모델의 정확도를 계산한 결과 평균 제곱근 오차와 조정 결정계수는 각각 0.0211 및 0.8299이었으며, 이는 해당하는 회귀모델이 관측값의 상당 부분을 잘 설명해주고 있음을 의미한다. 실제 선박의 연료소비 경향에 대한 모델의 예측성능을 분석하기 위하여 평가용 데이터 중 10일 동안을 Fig. 12와 같이 예시로 나타내었다. 그래프의 실선 부분이 선박의 실제 연료소모율, 점선 부분이 주성분 회귀모델로 예측한 연료소모율이며, 해당하는 기간 동안 선박에서 소비한 총 연료량을 Table 5에서 비교하였다.

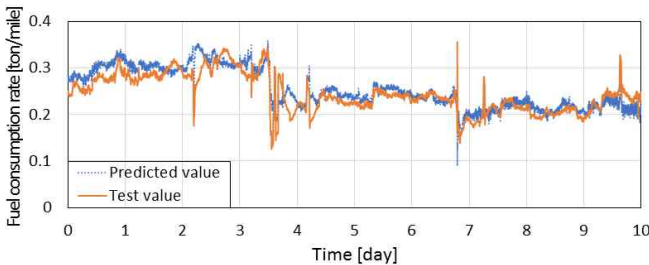


Fig. 12 Prediction accuracy of the model for 10 days of test data

Table 5 Fuel consumption of the target ship in 10 days

Total distance	Actual fuel consumption	Predicted fuel consumption	Average error
3704 miles	917.5 tons	958.6 tons	4.1 tons/day

Fig. 12에서 10일 동안의 선박 연료소모율 변화를 참고하면 주성분 회귀모델에 의한 예측값이 실제 선박의 연료소비 경향을 추종하고 있는 것을 확인할 수 있다. 해당하는 기간 동안 대상 선박의 주기관에서 소모한 연료량은 917.5톤이었으며, 예상 연료량은 958.6톤으로 일평균 4.1톤 정도의 예측 오차가 발생하였으나, 전반적인 연료소비 경향의 예측이 가능하기 때문에 항해 계획 수립 또는 항해 중 선박의 에너지 효율적인 운항상태를 유지하기 위한 목적에는 충분히 적용할 수 있을 것으로 사료된다. 향후 다른 차원 축소 방법이나 인공지능망과 같은 기계학습 방법을 추가로 고려한다면 모델의 예측 정확도를 보완할 수 있을 것으로 판단된다.

다양한 기계학습의 발달로 인하여 적절한 분석 없이 무분별하게 데이터를 입력 시켜 원하는 결과를 얻을 경우, 비슷한 물리적인 의미를 가지는 입력 변수 간의 다중공선성을 야기할 수 있고 회귀 모형의 부적절한 계수 추정 등을 발생시킬 수 있다. 결과적으로 이러한 문제들은 분석 결과의 신뢰성을 떨어뜨리며 예측 성능을 저하시킨다. 하지만 주성분 분석 기반의 예측모델은 서로 상관성이 없거나 독립적인 새로운 변수들로 주성분을 구성하기 때문에 정보 손실을 최소화하면서 다중공선성의 문제를 회피할 수 있으며, 전체 데이터의 분산을 대표하는 주성분을 추출함으로써 데이터의 차원을 감소시키기 때문에 선박이나 육상에서의 데이터 송수신 및 저장에도 효과적일 것으로 판단된다.

## 4. 결론 및 향후 과제

본 연구에서는 13K TEU급 컨테이너 선박으로부터 약 6개월간 수집한 운항데이터를 활용하여 선박의 연료소비를 예측하고자 하였다. 데이터의 통합, 정제, 변환 및 축소 과정을 통해서 데이터의 전처리를 수행하였으며, 주성분 회귀분석을 통해 예측모델을 구현하였다. 연구 결과를 요약하면 다음과 같다.

1) 선박의 운항 특성을 고려한 적절한 데이터 정제를 수행하기 위하여 변수 간의 선형 관계, 중앙값 필터, 주성분 분석과 같은 방법을 활용하였다.

(i) 선형 관계를 나타내는 선박의 기관출력과 연료소모량의 회귀선으로부터 3 표준편차를 벗어나는 관측값을 이상값으로 식별하였으며, 식별한 값의 시계열 분석을 통해 이러한 방법의 유효성을 확인할 수 있었다.

(ii) 중앙값 필터를 사용하여 데이터에 포함된 잡음을 처리하고 일정한 경향성의 데이터로 정제하였다. 특히 선박의 동요 주기나 외력 등에 의하여 지속적으로 값이 변화하는 데이터는 수집 간격이 길어질수록 일정한 추



세변화를 나타내기에 부적절한 경우가 있었으며, 이를 해결하기 위하여 중앙값 필터를 적용하였다.

(iii) 주성분 분석을 활용하여 운항 변수와 연료소비 간의 관계를 파악하였으며, 이는 수집한 원시데이터의 이상값을 식별하는 방법으로도 사용될 수 있다.

2) 예측모델을 구현하기 위하여 독립변수에 대한 주성분 분석을 수행하였으며, 다음과 같은 4개의 주성분을 추출하였다. 제1주성분은 주기관 분당 회전수, 대지속력, 대수속력과 같은 선박의 추진 성분을 의미하였다. 제2주성분은 평균홀수, 트림, 배수량 및 침수표면적과 같은 선박의 하중 성분을 나타내었다. 제3주성분은 선박이 조우하는 상대 풍속과 상대 풍향과 같은 외력 성분, 제4주성분은 타각성분을 의미하였다.

3) 평가용 데이터에 대한 주성분 회귀모델의 예측 성능은 평균 제곱근 오차 0.0211, 조정 결정계수 0.8299이었다. 10일 동안 선박의 연료소비를 예측한 결과 하루 평균 4.1 ton 정도의 연료소모량의 차이가 발생하였으며, 선박의 전반적인 연료 소비 경향을 예측하기에는 적합할 것으로 사료된다.

주성분 분석을 통한 예측모델은 다중공선성 문제를 극복할 수 있으면서도 차원을 효율적으로 감소시켜 주기 때문에 선박과 육상에서의 빅데이터 송수신 및 저장에도 효과적일 것으로 판단되며, 항해 계획을 수립하거나 선박의 운항 중 에너지 효율적인 상태 모니터링 및 예측에 적용하여 선박 운항자의 의사결정을 지원하기 위한 에너지효율 최적화 시스템 개발에 활용될 수 있을 것으로 사료된다. 또한 운항데이터의 처리와 분석에 사용된 여러 접근 방법들은 예지 정비 및 안전 시스템 개발과 같은 다양한 분야에도 확장 가능할 것으로 판단된다. 다만, 본 연구에서는 선박에서 관측한 풍향과 풍속 데이터를 외부 환경 조건으로 활용하였는데 파랑 스펙트럼, 유의파고, 파의 주기 등과 같은 추가적인 데이터를 취득할 수 있다면 연료소비에 대한 외력의 영향을 더욱 정교하게 반영할 수 있을 것으로 판단된다. 또한 본 연구에서는 한 척의 컨테이너 선박을 대상으로 연구를 수행하여 일반적인 적용가능성에 대한 한계가 존재하였으며, 다양한 선종과 운항 상태에 대한 데이터가 축적된다면 고도화된 통합 연료소비 예측모델을 수립할 수 있을 것이라고 판단된다. 향후 주성분 분석뿐만 아니라 다른 차원감소 방법에 의한 변수 선정에 대해서도 검토할 예정이며, 모델의 정확도 향상을 위하여 기계학습 방법도 적용할 예정이다.

## References

- [1] Cattell, R. B.(1966), "The Scree Test For The Number of Factors", *Multivariate behavioral research*, Vol. 1, No. 2, pp. 245-276.
- [2] IMO.(2009), "Guidelines for voluntary use of the ship energy efficiency operational indicator (EEOI)", MEPC 1/Circ. 684.
- [3] IMO.(2012a), "Guidelines on survey and certification of the energy efficiency design index (EEDI)", MEPC 63/23/Add.1, Annex 10.
- [4] IMO.(2012b), "Guidance for the development of a ship energy efficiency management plan (SEEMP)", MEPC 59/24/Add.1, Annex 19.
- [5] Jolliffe, I. T.(1982), "A Note on the Use of Principal Components in Regression", *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Vol. 31, No. 3, pp. 300-303.
- [6] Kim, K. J., Lee, S. D., Jun, C. H., Park, K. M. and Byeon, S. S.(2017), "A Statistical Procedure of Analyzing Container Ship Operation Data for Finding Fuel Consumption Patterns", *Korean Journal of Applied Statistics*, Vol. 30, No. 5, pp. 633-645.
- [7] Moore, D. S. and Kirkland, S.(2007), *The Basic Practice of Statistics*, Vol. 2, WH Freeman Press.
- [8] Pearson, K.(1901), "Principal Components Analysis", *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Vol. 6, No. 2, p. 559.
- [9] Pedersen, B. P. and Larsen, J.(2009), "Prediction of Full-Scale Propulsion Power using Artificial Neural Networks", *Proceedings of the 8th international conference on computer and IT applications in the maritime industries (COMPIT'09)*, Budapest, Hungary, pp. 10-12.
- [10] Perera, L. P. and Mo, B.(2016), "Marine Engine Operating Regions under Principal Component Analysis to evaluate Ship Performance and Navigation Behavior", *IFAC-PapersOnLine*, Vol. 49, No. 23, pp. 512-517.
- [11] Petersen, J. P., Jacobsen, D. J. and Winther, O.(2012), "Statistical modelling for ship propulsion efficiency", *Journal of marine science and technology*, Vol. 17, No. 1, pp. 30-39.
- [12] Pratt, W. K.(1991), *Digital Image Processing*, John Wiley and Sons Press.
- [13] Pukelsheim, F.(1994), "The Three Sigma Rule", *The American Statistician*, Vol. 48, No. 2, pp. 88-91.
- [14] Stopford, M.(2009), *Maritime Economics*, 3rd Ed, Routledge Press.
- [15] Yan, X., Wang, K., Yuan, Y., Jiang, X. and Negenborn, R. R.(2018), "Energy-efficient shipping: An application of big data analysis for optimizing engine speed of inland ships considering multiple environmental factors", *Ocean Engineering*, Vol. 169, pp. 457-468.

Received 26 July 2019

Revised 26 August 2019

Accepted 9 September 2019