

A statistical procedure of analyzing container ship operation data for finding fuel consumption patterns

Kyung-Jun Kim^a · Su-Dong Lee^a · Chi-Hyuck Jun^{a,1} · Kae-Myoung Park^b · Sang-Su Byeon^c

^aDepartment of Industrial and Management Engineering, Pohang University of Science and Technology; ^bKorean Register of Shipping; ^cHyundai Ocean Service CO., LTD

(Received July 05, 2017; Revised September 02, 2017; Accepted October 08, 2017)

Abstract

This study proposes a statistical procedure for analyzing container ship operation data that can help determine fuel consumption patterns. We first investigate the features that affect fuel consumption and develop the prediction model to find current fuel consumption. The ship data can be divided into two-type data. One set of operation data includes sea route, voyage information, longitudinal water speed, longitudinal ground speed, and wind, the other includes machinery data such as engine power, rpm, fuel consumption, temperature, and pressure. In this study, we separate the effects of external force on ships according to Beaufort Scale and apply a partial least squares regression to develop a prediction model.

Keywords: ship operation efficiency, Beaufort Scale, PLS regression

1. 서론

선박의 배기가스 관련 국제환경규제의 강화로 인해, 선박 운항의 에너지효율 제고의 필요성이 대두되고 있는 상황이다. 실제로, 국제해사기구(International Maritime Organization; IMO)는 2016년 이후 선박 배출 통제지역(emission control area; ECA)에서 기존배출 규제의 80%의 질소산화물 감축을 목표로 하는 규제를 채택하였고, EU는 대규모 산·학·연 공동프로젝트 등을 통해 친환경 선박 기술 개발 및 환경규제에 범국가적으로 공동대응 중이다. 또한, 일본 조선 업계의 경우 2014년 4월 공동 연구 플랫폼으로서 기술 연구개발을 전문으로 하는 Maritime Innovation Japan Corporation (MIJAC)를 설립하고, 이를 통해 조선업계 기업, 선박 기자재 기업, 일본해사협회, 일본정책 투자은행 등이 참여하여 조선, 해운, 해양 등과 관련하여 에너지 효율 제고, 친환경 조선, 해양 기술 개발 등의 연구를 수행하고 있다. 국제환경규제 강화뿐만 아니라, 국제경기 침체의 장기화로 인한 해운업계의 불황을 타개하기 위해 연료 및 유지보수 비용 절감 등을 통한 다각적인 경쟁력 향상의 방안이 필요하다.

This research was a part of the project titled 'Development of TCS system on ECO-Ship technology', funded by the Ministry of Oceans and Fisheries, Korea.

¹Corresponding author: Department of Industrial and Management Engineering, Pohang University of Science and Technology, 77 Cheongam-ro, Nam-Gu, Pohang, Gyeongbuk 37673, Korea.

E-mail: chjun@postech.ac.kr

최근에 빅데이터의 수집, 저장 및 처리기술의 발달로 선박의 실시간 운항정보 및 기기상태 모니터링 데이터 활용을 위한 시스템이 주요 선진국의 해운업계를 시작으로 보급 및 확산되는 추세이다. 본 연구에서는 특정 선박에서 수집된 운항데이터에 partial least squares (PLS) 회귀분석 (Geladi와 Kowalski, 1986; Jun, 2012)을 적용하여 선박의 연료효율과 관련하여 현재 운항 상태의 적정여부를 판단하고, 환경요인에 대한 선박의 운항을 탐색하고자 한다.

선박 운항 에너지효율 관리 방법으로 여러 방법이 존재하며, 항해성능관리(voyage performance management), 선체 및 프로펠러 조건관리(hull and propeller condition management), 그리고 선박시스템관리(ship system management) 측면으로 분류할 수 있다 (American Bureau of Shipping, 2013). 본 연구에서는 항해성능관리를 집중적으로 다루며, 이는 선박 운항 관련 변수(운행 및 환경 변수 등)를 관리함으로써 에너지효율을 최대화하는 것을 말하며, 항해속도최적화, 최적항로, 트림(trim)/드래프트(draft) 최적화 등을 포함한다.

선박 운항 에너지효율 표준 개발을 위해서는 선행적으로 에너지효율 관련 변수 선정이 중요하다고 알려져 있으며, 여러 연구에서 에너지효율 관련 변수를 제시하였다. Armstrong (2013)은 선박의 속도, 선박, 날씨 등을 언급하였으며, Ballou (2013)은 선박 종류, 화물, 운항 거리 및 시간, 재항시간을 언급하였고, Moon과 Woo (2014)는 재항시간, 선박 종류, 항해 시간, 선박의 속도, 그리고 Ando (2015)는 선박의 속도, 바람 속도, 파도 높이, 파랑 주기, 날씨 등을 에너지효율 관련 변수로 제시하였다.

위의 관련 변수들과 에너지효율 간의 관계 분석을 통해 선박 운항 에너지효율 표준을 개발한 다양한 연구들이 존재한다. Kwon과 Kim (2005)는 해상에서의 선속 손실량 산정을 위한 약산식을 제안하였고, 이는 Beaufort Number와 체적 배수량 등을 고려하여 해상의 파도, 바람, 그리고 표면조도 등에 기인된 선속의 손실량을 개발하였다. Armstrong (2013)은 선박 운행 속도와 연료 소모량 간의 관계를 분석하였고, Ando (2015)는 Armstrong (2013) 보다 한 단계 더 나아가, 바다 상태 별 선박 운행 속도와 연료 소비율 간의 관계를 분석하였다. 또한, Moon과 Woo (2014)는 선박 종류 별 재항시간과 비용 효과 및 이산화탄소 배출량 간의 관계를 회귀분석과 민감도 분석을 활용하여 분석하였다.

본 연구의 구성은 다음과 같다. 2절에서는 본 연구에서 사용하는 분석 절차 및 각 방법론들에 대해 고찰하며, 3절에서는 운항데이터 분석을 통해 선박 운항효율 예측모델에 대한 실험결과를 요약한다. 마지막으로 4절의 결론으로 이어진다.

2. 통계적 분석 절차

본 연구에서 제안하는 운항데이터의 분석 절차는 Figure 2.1과 같으며, 각 단계에서 진행하는 내용과 방법론들은 아래와 같다.

선박의 특성상 외력의 영향을 많이 받으며, 외력에 따라 에너지를 소모하는 양이나 비율 또한 크게 달라지게 된다. 이에 Step 1에서는 외력에 대한 영향을 운항 중에 발생하는 바람의 속도를 통해 얻으며, 이를 보퍼트 풍력계급(Beaufort Scale; BFS)을 통해 표현한다. BFS는 바람의 속도를 기준으로 BFS0부터 BFS12까지 총 13단계로 나눈 계급을 말하며, 이는 실해역에서 선박의 운항성능 분석/평가를 하거나 항로를 설정할 때 매우 중요한 정보이다. BFS에 대한 정보는 Table 2.1과 같다.

선박의 운항데이터에 속하는 각 변수들은 선박의 에너지 소모와 선형 관계가 아닌 특정 비선형 함수로 표현될 수 있는 물리적인 관계를 갖는다. 이를 반영하기 위해 Figure 2.1의 Step 2에서와 같이 Curve Fitting을 통해 종속변수에 대한 독립변수의 파생변수를 생성하는 과정을 거친다. Curve Fitting을 통해 얻은 독립변수의 파생변수를 사용함으로써, 종속변수와의 선형관계를 증가시켜 개선된 예측모형을 얻을 수 있다. 독립변수의 파생변수에 대한 예시로, 독립변수 $x = [1, 2, 3]^T$, 종속변수 $y =$

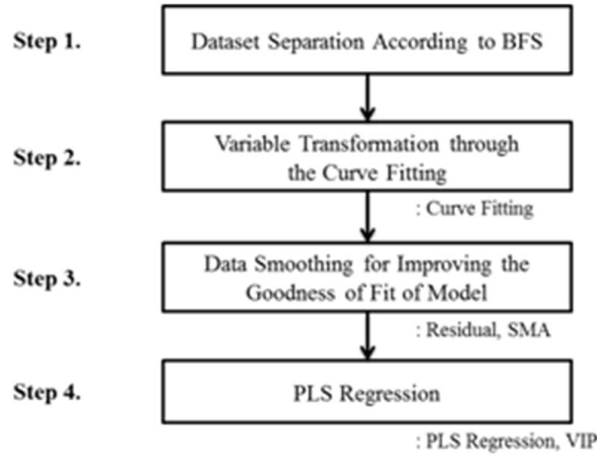


Figure 2.1. Statistical analysis procedure. BFS= Beaufort Scale; SMA = simple moving average; PLS = partial least squares; VIP = variable importance in projection.

Table 2.1. Information of the Beaufort Scale (BFS)

BFS	Wind speed (m/s)		Strength
	Over	Under	
BFS0	0.0	0.3	Calm
BFS1	0.3	1.6	Light Air
BFS2	1.6	3.4	Light Breeze
BFS3	3.4	5.5	Gentle Breeze
BFS4	5.5	8.0	Moderate Breeze
BFS5	8.0	10.8	Fresh Breeze
BFS6	10.8	13.9	Strong Breeze
BFS7	13.9	17.2	Near Gale
BFS8	17.2	20.8	Gale
BFS9	20.8	24.5	Strong Gale
BFS10	24.5	28.5	Storm
BFS11	28.5	32.7	Violent Storm
BFS12	32.7		Hurricane

$[2.4630, 18.1994, 134.4763]^T$ 가 주어졌을 때 이 둘의 상관계수는 0.9154이다. 하지만 여기서 독립변수와 종속변수는 $y = f(x) = (1/3) \exp(2 * x)$ 라는 비선형 관계를 갖고 있으며, 해당 $f(x)$ 에서 상수를 제외한 $\exp(2 * x)$ 를 통해 종속변수와 선형 관계인 파생변수를 생성하게 되면 1.0000의 상관계수를 갖게 된다. 즉 파생변수 $x_{new} = \exp(2 * x)$ 를 기존의 독립변수와 대체함으로써 예측모형의 성능을 향상시킬 수 있다.

운항데이터의 경우, 외력에 의해 순간적으로 관측치가 비정상적으로 측정될 수 있는 동시에 선박의 운항 특성상 그 변동이 매우 심하기 때문에, Step 3에서는 모형의 R^2 향상을 위한 데이터 평활화(data smoothing)을 진행한다. 이를 위해, 회귀모형의 잔차(residual)와 단순이동평균(simple moving average; SMA)을 적용한다.

종속변수의 실제 값 y 와 예측 값 \hat{y} 의 차이인 회귀모형 잔차 $r(= y - \hat{y})$ 에 대한 정규분포 가정을 통해 신

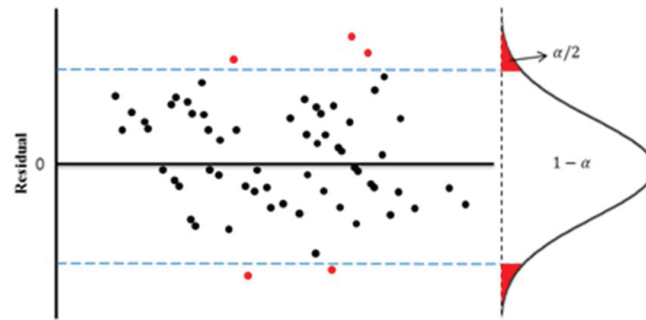


Figure 2.2. Example of outlier elimination.

외수준(significance level)을 나타내는 값인 특정 α 를 설정하여, Figure 2.2와 같이 그 정도를 벗어나는 관측치들을 이상치(outlier)로 간주하고 제거한다. 이를 통해 예측모형의 R^2 를 향상시킨다.

운항데이터의 관측치는 선박에서 10분마다 측정되는 시계열 데이터인 동시에 외력에 의해 측정되는 관측치의 값의 변동 폭이 굉장히 넓은 데이터이다. 이러한 불규칙한 변동을 완화하기 위해, Figure 2.1의 Step 3에서 이상치를 제거하는 것과 더불어 SMA를 적용하며, 이에 따른 수식은 아래와 같다.

$$SMA_{t-(T-1)} = \frac{(x, y)_t + (x, y)_{t-1} + \cdots + (x, y)_{t-(T-1)}}{T}, \quad (2.1)$$

여기서 $x_i \in \mathbb{R}^k$ 는 k 개의 독립변수로 이루어진 i 번째 관측치, $y \in \mathbb{R}$ 는 종속변수, t 는 현재의 기준 시점을 나타내며, T 는 이동평균을 낼 타임윈도우(time window)를 의미한다. 이에 따라 최종적으로 $t - (T - 1)$ 개의 SMA가 적용된 새로운 데이터들을 얻게 된다. 타임윈도우의 경우 너무 낮은(높은) 값으로 설정을 하게 되면, 원래 데이터의 불규칙한 변동이 완화되지 않는(원래 데이터의 성질을 잃어버리는) 문제가 발생할 수 있어, 적절한 값을 설정하는 것이 중요하다.

분석 절차 중에서 마지막인 Step 4에서는 PLS 회귀분석을 통해 선박의 운항효율 예측모형을 수립한다. 이는 독립변수들 간의 중복성을 고려하여 종속변수를 가장 잘 설명하는 최적의 잠재변수들을 찾음과 동시에 중요한 영향을 미치는 독립변수들을 variable importance in projection (VIP)를 통해 선택하기 위함이다. PLS 회귀분석의 모형에 대한 수식은 아래와 같다.

n 개의 관측치와 k 개의 독립변수가 주어졌을 때, $n \times k$ 데이터행렬 \mathbf{X} , 종속변수벡터 \mathbf{y} 에 대하여 동시에 A 개의 잠재변수 t_a ($a = 1, \dots, A$)로 설명할 수 있는 모형을 고려할 수 있으며, 그 관계식은 아래와 같다. 즉, A 는 잠재변수 수로서 독립변수 수 k 보다 작다.

$$\mathbf{X} = t_1 \mathbf{p}_1^T + t_2 \mathbf{p}_2^T + \cdots + t_A \mathbf{p}_A^T + \mathbf{E}, \quad (2.2)$$

$$\mathbf{y} = t_1 b_1 + t_2 b_2 + \cdots + t_A b_A + \mathbf{f}. \quad (2.3)$$

여기서 \mathbf{p}_a 는 \mathbf{X} 에 대응하는 로딩(loading)을 나타내며, \mathbf{E} 와 \mathbf{f} 는 각 모델에 해당하는 오차항이다. 이에 대한 계수 및 잠재변수 값들을 아래와 같은 단계를 통하여 산출할 수 있고, A 개의 잠재변수를 구할 때까지 반복한다.

단계 1. \mathbf{X} 의 각 열벡터를 종속변수로 \mathbf{y} 를 독립변수로 하는 단순회귀모형으로부터 \mathbf{y} 의 기울기를 추정하여 가중치 $\mathbf{w} = (w_1, \dots, w_k)^T$ 를 산출하고 이의 크기가 1이 되도록 조정한다.

$$w_j = \frac{\mathbf{x}_j^T \mathbf{y}}{\mathbf{y}^T \mathbf{y}}, \quad j = 1, \dots, k; \quad \mathbf{w} \leftarrow \frac{\mathbf{w}}{\sqrt{\mathbf{w}^T \mathbf{w}}}. \quad (2.4)$$

단계 2. \mathbf{t} 를 다음과 같이 산출한다.

$$\mathbf{t} = w_1\mathbf{x}_1 + \cdots + w_k\mathbf{x}_k = \mathbf{X}\mathbf{w}. \quad (2.5)$$

단계 3. \mathbf{t} 를 \mathbf{X} 에 회귀시켜 로딩 벡터 \mathbf{p} 를 다음과 같이 구한다.

$$\mathbf{p} = \frac{\mathbf{X}^T\mathbf{t}}{\mathbf{t}^T\mathbf{t}}. \quad (2.6)$$

단계 4. 로딩 벡터 \mathbf{p} 의 크기를 반영하여 \mathbf{t} 와 \mathbf{w} 를 조정된 뒤 \mathbf{p} 의 크기를 1이 되도록 조정한다.

$$\mathbf{t} \leftarrow \mathbf{t}\sqrt{\mathbf{p}^T\mathbf{p}}, \quad \mathbf{w} \leftarrow \mathbf{w}\sqrt{\mathbf{p}^T\mathbf{p}}, \quad \mathbf{p} \leftarrow \frac{\mathbf{p}}{\sqrt{\mathbf{p}^T\mathbf{p}}}. \quad (2.7)$$

단계 5. \mathbf{t} 를 \mathbf{y} 에 회귀시킬 때, 계수 b 를 산출한다.

$$b = \frac{\mathbf{y}^T\mathbf{t}}{\mathbf{t}^T\mathbf{t}}. \quad (2.8)$$

단계 6. \mathbf{X} 와 \mathbf{y} 를 다음과 같이 수정 후, A 개의 잠재변수를 얻을 때까지 단계 1부터 과정을 반복한다.

$$\mathbf{X} \leftarrow \mathbf{X} - \mathbf{t}\mathbf{p}^T \quad (\mathbf{x}_j \leftarrow \mathbf{x}_j - p_j\mathbf{t}, j = 1, \dots, k), \quad \mathbf{y} \leftarrow \mathbf{y} - b\mathbf{t}. \quad (2.9)$$

이렇게 얻은 새로운 잠재변수들을 통해, 종속변수 \mathbf{y} 에 대한 예측식의 형태는 기존의 다중선형회귀분석과 유사하다.

$$\hat{\mathbf{y}} = \hat{b}_1\mathbf{t}_1 + \cdots + \hat{b}_A\mathbf{t}_A = \mathbf{T}\hat{\mathbf{b}}. \quad (2.10)$$

VIP는 원래의 각 독립변수가 실질적으로 종속변수에 얼마나 영향을 미치는 지를 나타내는 독립변수의 중요도 척도이며, 이는 PLS에서 각 독립변수가 잠재변수에 얼마나 영향을 주며 다시 이러한 잠재변수들이 종속변수에 어떤 영향을 미치는 지를 동시에 고려한 값이다. j 번째 독립변수에 대한 VIP는 아래와 같이 정의한다.

$$\text{VIP}_j = \sqrt{\frac{k}{\text{SSR}} \sum_{a=1}^A \text{SSR}_a \left(\frac{w_{aj}}{\|\mathbf{w}_a\|} \right)^2}, \quad j = 1, \dots, k \quad (2.11)$$

이때, w_{aj} 는 독립변수 \mathbf{f}_j 의 잠재변수 \mathbf{t}_a 에 대한 기여도, SSR_a 는 a 번째 잠재변수가 종속변수 \mathbf{y} 를 설명하는 기여도로서 회귀제곱 합으로 구할 수 있다. $\|\mathbf{w}_a\|$ 는 $\sum_j w_{aj}^2$ 의 값이 1이 되도록 하는 정규화 값이다.

위의 정의에 따르면, 아래와 같은 성질이 성립하므로,

$$\sum_{j=1}^k \text{VIP}_j^2 = k. \quad (2.12)$$

독립변수들의 VIP 제곱의 평균은 1임을 알 수 있다. 따라서, VIP가 1보다 큰 독립변수를 유의한 독립변수로 간주하는 것이 일반적이다.

지금까지의 분석을 이용하여 선박의 운항을 탐색 하기 위해, BFS 별로 PLS 회귀분석의 VIP를 통해 얻어진 중요독립변수들에 대해 외력의 차이를 데이터 상에서 확인할 수 있는 지에 대한 분석을 시각화를 통해 진행한다.

Table 3.1. Information of independent variables and target

Type	Name	Description	Unit
Independent variable	LWS	Longitudinal water speed	knots
	LGS	Longitudinal ground speed	knots
	POWER	Engine power	kW
	RPM	Revolutions per minute	rpm
	Trim	Slope of the ship	<i>m</i>
	Draft	Depth of the ship under water	<i>m</i>
	Displacement	Displacement	<i>t</i>
	Wetted surface	Wetted surface of the ship	<i>m</i> ²
	Propeller immersion	Propeller immersion	%
	Wind resistance	Wind resistance	kN
Target	SFOC	Specific fuel oil consumption	g/kWh
	FOC	Fuel oil consumption	l/h, kg/h

Table 3.2. Component ratio of the data according to Beaufort Scale (BFS) and direction

East-Bound	Observation	Ratio (%)	West-Bound	Observation	Ratio (%)
BFS0	391	0.92	BFS0	186	0.44
BFS1	972	2.29	BFS1	617	1.45
BFS2	2,619	6.16	BFS2	1,912	4.50
BFS3	5,011	11.78	BFS3	3,862	9.09
BFS4	7,680	18.06	BFS4	6,062	14.26
BFS5	7,208	16.95	BFS5	6,996	16.46
BFS6	6,777	15.94	BFS6	8,328	19.60
BFS7	5,095	11.98	BFS7	6,881	16.19
BFS8	3,304	7.77	BFS8	5,732	13.49
BFS9	2,477	5.83	BFS9	1,693	3.98
BFS10	795	1.87	BFS10	225	0.53
BFS11	188	0.44	BFS11	3	0.01
BFS12	4	0.01	BFS12	0	0.00
Total	42,521	100.00	Total	42,497	100.00

3. 운항데이터 분석 결과

이 장에서는 운항데이터 분석을 통해 선박 운항효율 예측모델과 선박의 운항의 탐색에 대한 실험결과를 요약한다.

3.1. 데이터

본 연구에서 사용된 운항데이터는 2014년 2월부터 2016년 7월까지 약 2년 6개월 동안 6,800TEU 컨테이너선에서 얻어진 것으로, 독립변수의 총 개수는 917개이며 관측치는 10분마다 측정되는 시계열 데이터이다. 이 중에서 Sea-Going 항해 구간에 대한 데이터를 이용하였으며, 선박은 항해할 때 환경조건에 따라 성능이 다르게 나타나며, 동쪽으로 갈 때의 조류, 기상조건이 서쪽으로 갈 때와는 사뭇 다르기 때문에 항해 방향에 따라 동쪽으로 항해하는 East-Bound와 서쪽으로 항해하는 West-Bound로 구분하였다. 이론적 이상치를 제거한 운항데이터의 관측치 수는 East-Bound에서는 42,521개, West-Bound에서는 42,497개로 이루어져 있다.

본 연구에서는 변수선택법(feature selection)을 적용하지 않고, Table 3.1과 같이 전문가들의 의견을 반

Table 3.3. Results of curve fitting

Target	Independent variable									
	LWS	LGS	RPM	Draft	Displacement	Wetted surface	Trim	Wind resistance	POWER	Propeller immersion
SFOC	P	P	P	L	L	L	L	L	P	L
	L	0.2618	0.2427	0.2766	-	-	-	-	0.3054	-
<i>R</i> -squared	E	0.3153	0.3152	0.3272	-	-	-	-	0.4811	-
	P	0.3326	0.3388	0.3496	-	-	-	-	0.4946	-
<i>b</i>		-0.88	-0.69	-1.54	-	-	-	-	-0.62	-
FOC	E	E	E	L	L	L	L	L	L	L
	L	0.4293	0.4154	0.5368	-	-	-	-	-	-
<i>R</i> -squared	E	0.5959	0.5636	0.6186	-	-	-	-	-	-
	P	0.5887	0.5585	0.6026	-	-	-	-	-	-
<i>b</i>		0.12	0.12	0.03	-	-	-	-	-	-

L = linear; E = exponential; P = power.

Table 3.1 provides descriptions on the abbreviated names of independent variable and target.

영한 10개의 독립변수와 2개의 종속변수를 활용해 분석을 진행한다. 2개의 종속변수는 각각 specific fuel oil consumption (SFOC)와 fuel oil consumption (FOC)로 서로 독립적이지 않으며, SFOC의 단위는 g/(kW×hour)로 단위일당 연료소모량을, FOC의 단위는 liter/hour로 연료소모량을 나타낸다.

3.2. 환경요인을 고려한 운항데이터 구분

외력에 의해 데이터가 받는 영향을 파악하기 위해 BFS를 기준으로 운항데이터를 구분하였고, 그 결과는 Table 3.2와 같다.

3.3. Curve Fitting을 통한 파생변수 생성

선택된 10개의 독립변수와 종속변수 간의 일대일 함수 관계를 확인하며, 변수 간 관계함수로는 아래와 같이 linear, exponential, 그리고 power 함수를 사용한다.

$$\begin{aligned} \text{linear: } f(x) &= a \times x + b, \\ \text{exponential: } f(x) &= a \times \exp(b \times x), \\ \text{power: } f(x) &= a \times x^b. \end{aligned}$$

각 독립변수와 종속변수에 대해, 가장 높은 *R*-squared 값을 갖는 함수를 최종 함수로 선택한다. 이에 대한 결과는 Table 3.3과 같다.

3.4. 모형의 R² 향상을 위한 데이터 필터링

운항데이터의 경우, 외력에 의해 순간적으로 관측치가 비정상적으로 측정될 수 있기 때문에, 이를 방지하기 위해 회귀모형의 잔차와 SMA를 적용한다. 이에 앞서 모든 독립변수는 curve fitting을 통해 얻어진 선형화된 파생변수를 사용한다.

회귀모형 잔차의 경우, 회귀모형으로부터 실제 종속변수의 값과 회귀모형에 의해 추정된 값의 차이인 잔차를 계산하여, 이 값이 신뢰구간을 벗어나는 관측치를 이상치(outlier)로 간주하고 제거하여 예측모형의 R²를 향상시킨다. 여기서, 유의수준은 α = 0.5로 설정한다.

Table 3.4. VIP through the PLS regression

	LWS	LGS	RPM	Draft	Displacement	Wetted surface	Trim	Wind resistance	POWER	Propeller immersion
East-Bound	1.38	1.29	1.52	0.38	0.39	0.39	0.48	0.55	1.60	0.37
West-Bound	1.47	1.40	1.50	0.47	0.47	0.47	0.25	0.50	1.53	0.41

VIP = variable importance in projection; PLS = partial least squares.

Table 3.1 provides descriptions on the abbreviated names of independent variable.

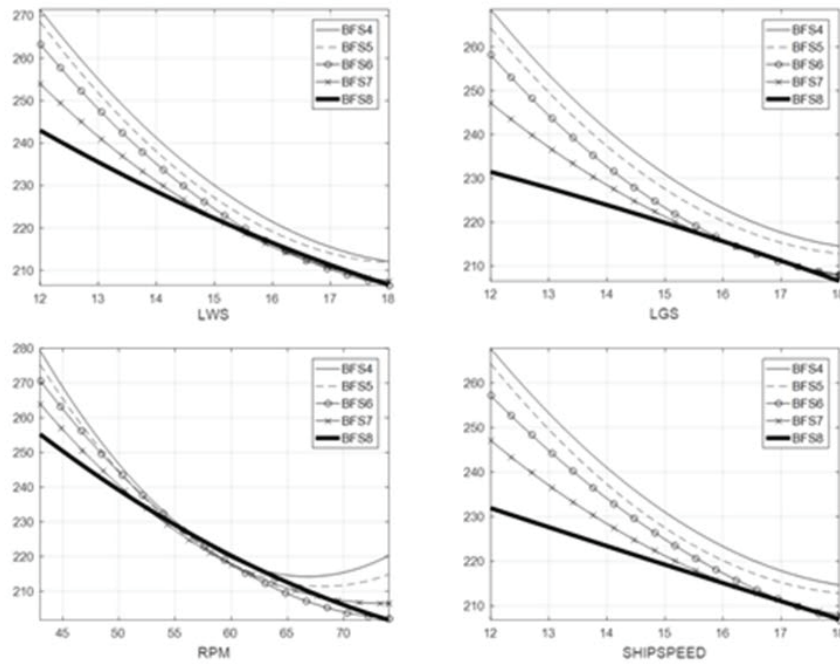


Figure 3.1. Results of the fitting with (East-Bound & SFOC). BFS = Beaufort Scale; SFOC = specific fuel oil consumption; LWS = longitudinal water speed; LGS = longitudinal ground speed; RPM = revolutions per minute.

SMA에서 정해주어야 할 값인 타임윈도우는 $T = 24$ 로 설정하며, 이는 관측치를 1시간 단위(즉, $T = 6, 12, 18, 24, 30, \dots$)로 증가시키다가 Table 3.3에서 구한 추정된 값인 b 를 변화시키지 않는 최대값이다.

3.5. 실험 결과

운항데이터를 BFS 별로 나누었을 때, 가장 많이 운항을 하고 관심이 있는 구간인 BFS 4-8 구간을 분석하였고, PLS 회귀분석을 통해 산출된 VIP 값의 평균값에 대한 결과는 Table 3.4와 같다. VIP를 통해 얻은 주요변수로 선택된 독립변수들을 이용해 BFS 별로 종속변수 SFOC, FOC와의 관계가 구분되어 나타나는지를 파악하게 위해 그 관계를 추정하였고, 주요독립변수는 Table 3.4와 같이 선박의 속도를 나타내는 변수들인 LWS, LGS, RPM가 전반적으로 종속변수에 가장 많은 영향을 미치는 변수들로 선택되었다. 추가적으로 배의 속도인 SHIPSPEED를 사용하였고, 이에 대해 추정된 결과는 Figure 3.1-3.4와 같다.

주요독립변수와 종속변수와의 관계를 추정한 결과, 연료소모량을 나타내는 FOC의 경우에는 주요독립

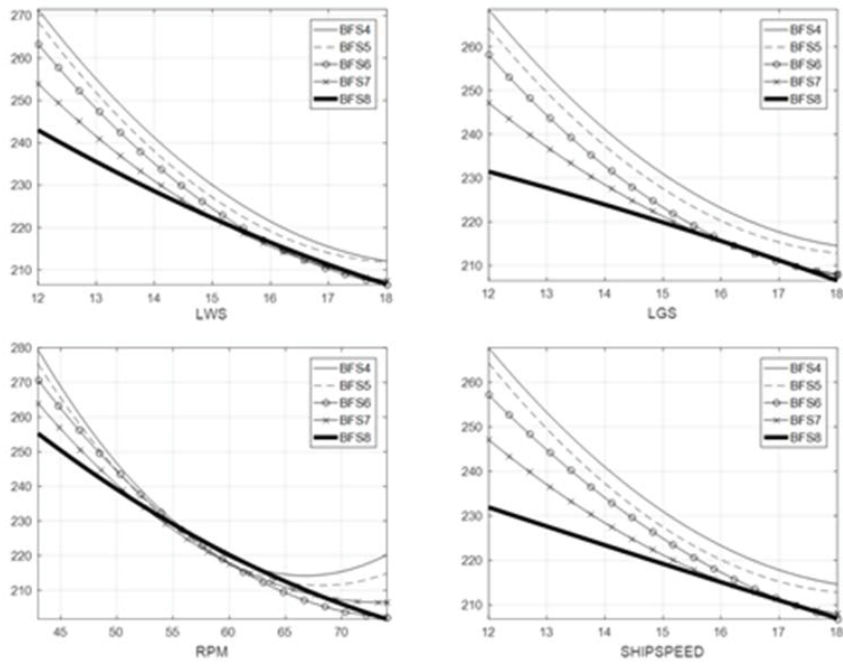


Figure 3.2. Results of the fitting with (West-Bound & SFOC). BFS = Beaufort Scale; SFOC = specific fuel oil consumption; LWS = longitudinal water speed; LGS = longitudinal ground speed; RPM = revolutions per minute.

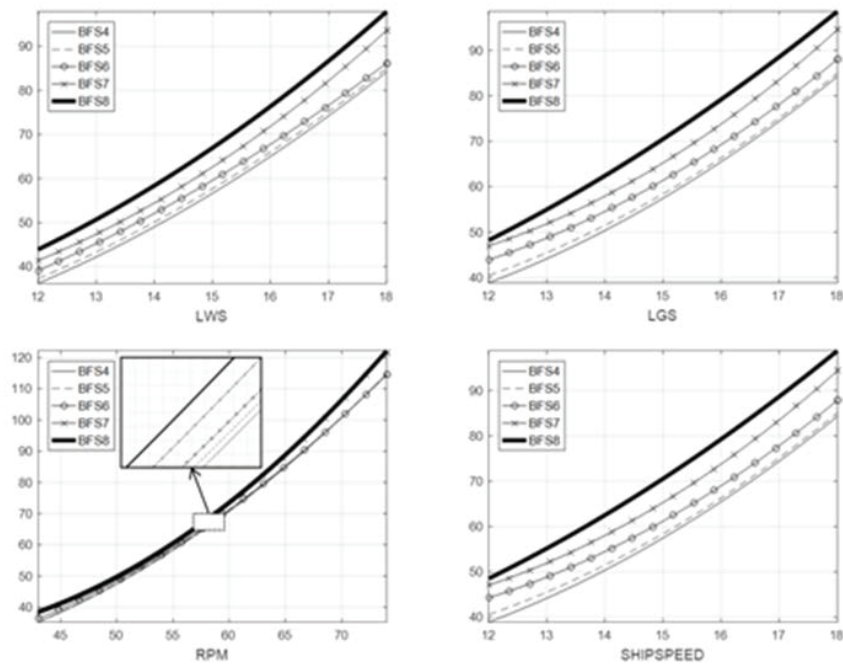


Figure 3.3. Results of the fitting with (East-Bound & FOC). BFS = Beaufort Scale; SFOC = specific fuel oil consumption; LWS = longitudinal water speed; LGS = longitudinal ground speed; RPM = revolutions per minute.

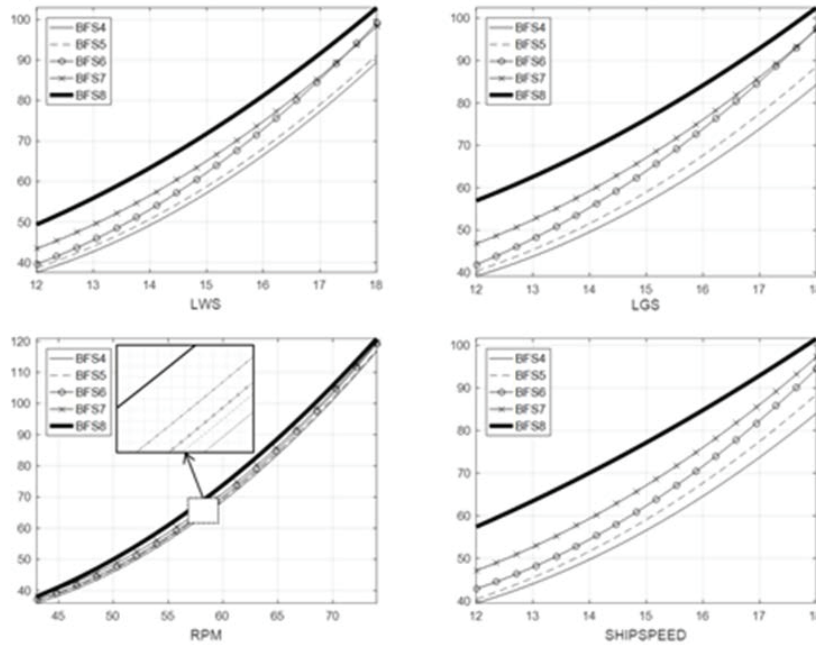


Figure 3.4. Results of the fitting with (West-Bound & FOC). BFS = Beaufort Scale; SFOC = specific fuel oil consumption; LWS = longitudinal water speed; LGS = longitudinal ground speed; RPM = revolutions per minute.

Table 3.5. Results of the PLS regression with SFOC

		East-Bound					
		SMA unapplied			SMA applied		
Statistical measure		G1	G2	T	G1	G2	T
R^2	Mean	0.6680	0.6216	0.6895	0.8378	0.7891	0.8343
	Std.	0.0087	0.0100	0.0081	0.0028	0.0109	0.0005
RMSE	Mean	13.4609	11.0120	12.4503	8.0779	6.5866	7.8349
	Std.	0.1241	0.2422	0.2100	0.0690	0.1861	0.0455
SFOC	Error ratio (%)	8.45	10.69	8.44	6.25	9.25	6.33
			West-Bound				
		SMA unapplied			SMA applied		
Statistical measure		G1	G2	T	G1	G2	T
R^2	Mean	0.4768	0.4792	0.5278	0.7884	0.7148	0.7660
	Std.	0.0168	0.0121	0.0138	0.0128	0.0128	0.0113
RMSE	Mean	22.6614	8.4743	18.7094	10.3306	4.6729	10.2239
	Std.	1.1422	0.4403	0.9824	0.3533	0.1013	0.4356
Error ratio (%)		6.035	4.57	5.45	4.58	3.67	4.53

PLS = partial least squares; SFOC = specific fuel oil consumption; SMA = simple moving average; BFS = Beaufort Scale; G1 = BFS 4-6; G2 = BFS 7-8; T = BFS 4-8.

변수들이 증가할수록 증가하는 방향으로 움직이는 것을 알 수 있고, 환경요인인 BFS 별로 운항의 반응들이 확연히 구분되어 나타나는 것을 확인할 수 있었다. 이에 반해, 단위일당 연료소모량을 나타내는 SFOC의 경우에는 주요독립변수가 증가할수록 감소하는 관계를 갖고 있지만, 일부 결과에서 BFS 별로

Table 3.6. Results of the PLS regression with FOC

Statistical measure		East-Bound					
		SMA unapplied			SMA applied		
		G1	G2	T	G1	G2	T
R^2	Mean	0.9160	0.8659	0.9000	0.9548	0.9456	0.9582
	Std.	0.0019	0.0069	0.0025	0.0007	0.0021	0.0013
RMSE	Mean	2.3355	2.3000	2.4078	1.5926	1.1581	1.4247
	Std.	0.0192	0.0359	0.0243	0.0189	0.0362	0.0252
FOC	Error ratio (%)	7.07	6.74	6.79	5.41	4.54	4.60
	Statistical measure		West-Bound				
SMA unapplied			SMA applied				
G1			G2	T	G1	G2	T
R^2	Mean	0.8818	0.9436	0.8842	0.9498	0.9869	0.9534
	Std.	0.0038	0.0045	0.0044	0.0011	0.0013	0.0021
RMSE	Mean	2.3739	2.0543	2.1578	1.3429	0.9344	1.2264
	Std.	0.0312	0.0745	0.0320	0.0096	0.0444	0.0118
FOC	Error ratio (%)	7.42	6.42	6.34	4.62	3.19	4.19

PLS = partial least squares; FOC = fuel oil consumption; SMA = simple moving average;
BFS = Beaufort Scale; G1 = BFS 4-6; G2 = BFS 7-8; T = BFS 4-8.

운항의 반응들이 겹치거나 구분이 어려운 구간이 존재하는 것을 확인할 수 있었다.

운항데이터의 회귀분석 또한 BFS 4-8 구간을 분석하였고, 외력을 나타내는 BFS가 비슷한 구간은 그룹을 지어 결과를 나타내었다. 그룹1은 G1으로 표기하며 BFS 4-6 구간의 데이터들을 함께 사용한다. 마찬가지로, 그룹2는 G2로 표기하였으며 BFS 7-8 구간의 데이터들을 함께 사용하며, 전체데이터를 나타내는 표기인 T는 BFS 4-8 모든 구간의 데이터를 함께 사용함을 의미한다.

이러한 운항데이터에 PLS 회귀분석을 적용한 결과는 종속변수 SFOC는 Table 3.5, 종속변수 FOC는 Table 3.6과 같다. 운항데이터의 회귀분석 결과, 종속변수에 대한 그룹들인 G1, G2, T 간에는 East-Bound와 West-Bound 별로도 서로 다른 성능을 보여주었으며, 운항데이터에 SMA를 적용하는 경우, 회귀분석의 성능이 눈에 띄게 향상하는 것을 확인할 수 있었다.

4. 결론

총 10개의 독립변수들 중에서, 종속변수인 SFOC와 FOC에 실질적으로 유의한 영향을 미치는 독립변수는 LWS, LGS, RPM, POWER와 같이 선박의 속도와 직접적으로 관련이 있는 변수들이었다. 외력을 나타내는 변수인 Wind Resistance의 경우, 종속변수에 크게 영향을 주는 변수로 보이지는 않지만, 바람의 속도로 BFS를 구분해 나타내었기 때문에 외력을 운항데이터에 반영해주었다고 볼 수 있다. 결과적으로, 종속변수가 FOC인 경우에 R-squared와 RSME 측면에서 가장 좋은 결과를 보이는 것을 알 수 있었고, 종속변수가 SFOC인 경우는 산식에 역의 곱으로 POWER가 이용이 되는 동시에 운항데이터의 특성상 데이터의 관측 범위의 변동이 심하여 SFOC의 변동 역시 심화되어, 정확한 예측모델을 수립하지 못한 것으로 보인다.

BFS 별로 독립변수와 종속변수의 관계를 추정한 결과인 Figure 3.1-3.4과도 마찬가지로, 종속변수가 FOC인 경우가 SFOC인 경우에 비해 BFS에 대한 분명한 영향의 차이를 결과를 얻어낼 수 있음을 알 수 있었다. 이를 통해, 선박이 BFS 별로 어느 정도의 에너지를 소모하여야 정상적인 운항을 하고 있는지를 제시해 줄 수 있다. 또한, BFS에 따른 그룹 별 데이터에서 항해 방향에 따른 East-Bound, West-

bound와 데이터 평활화를 위한 SMA의 적용에 따른 데이터의 구분도 PLS 회귀분석의 결과를 통해 서로 다른 성능을 나타내는 것으로 나타나, 상황에 따른 데이터 구분이 분석에서 중요한 요소인 것으로 보인다.

본 연구에서는 BFS 별로 선박의 운항데이터를 구분지어 분석하였으며, 독립변수와 종속변수 간의 상관관계를 파악 및 반영한 예측모델을 수립하였다. 기존의 연구에서는 종속변수 SFOC, FOC와 독립변수 LWS, LGS, SHIPSPEED, RPM과의 관계에 대해, RPM의 변화량과 SFOC, FOC의 관계는 공장에서 엔진을 만들 때 나오는 ship test 결과만이 있을 뿐이고 실해역에서 수집된 데이터를 통해 분석된 상관관계는 없었다. 또한 이 때 해상상태의 등급을 의미하는 BFS 별로 그 상관관계를 산출한 데이터는 전무하다고 볼 수 있다. 따라서, 분석결과가 산업현장에서는 매우 중요한 실전 테스트 분석결과로 활용될 수 있으며, 구체적으로 BFS 별, 독립변수 별로 실해역에서 선박의 운항성능 분석/평가를 하거나 항로를 설정하는 데 중요한 내용으로 활용할 수 있다.

향후에는 정성적인 독립변수 선택뿐 만이 아니라 변수선택법을 활용한 정량적인 독립변수 선택을 적용하고자 한다. 또한 예측모델의 경우, PLS 회귀분석 외에 더 고도화된 다양한 예측모델들을 적용 및 비교하고자 한다.

References

- American Bureau of Shipping (2013). *Ship Energy Efficiency Measures: Status and Guidance*, ABS, Houston.
- Ando, H. (2015). *How We Tackle IoT of Ship: Data Utilization and Standardization*, *International Seminar on Practical Use of Maritime Big Data*, MTI, Tokyo.
- Armstrong, V. N. (2013). Vessel optimisation for low carbon shipping, *Ocean Engineering*, **73**, 195–207.
- Ballou, P. J. (2013). Ship energy efficiency management requires a total solution approach, *Marine Technology Society Journal*, **47**, 83–95.
- Björck, Å. (1996). *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia.
- Geladi, P. and Kowalski, B. R. (1986). Partial least-squares regression: a tutorial, *Analytica Chimica Acta*, **185**, 1–17.
- Jun, C. H. (2012). *Data Mining Techniques and Applications*, Hannarae, Seoul.
- Kwon Y.-J. and Kim D. Y. (2005). A research on the approximate formulae for the speed loss at sea, *Journal of Ocean Engineering and Technology*, **19**, 90–93.
- Moon, D. S. H. and Woo, J. K. (2014). The impact of port operations on efficient ship operation from both economic and environmental perspectives, *Maritime Policy & Management*, **41**, 444–461.

연료 소비 패턴 발견을 위한 컨테이너선 운항데이터 분석의 통계적 절차

김경준^a · 이수동^a · 전치혁^{a,1} · 박개명^b · 변상수^c

^a포항공과대학교 산업경영공학과, ^b(사)한국선급, ^c현대해양서비스(주)

(2017년 07월 05일 접수, 2017년 09월 02일 수정, 2017년 10월 08일 채택)

요약

본 연구는 컨테이너선의 연료 소비 패턴의 발견을 위해 운항데이터 분석의 통계적 절차를 제안한다. 우리는 현 시점의 연료 소비를 발견하기 위해 연료 소비에 영향을 미치는 변수들을 파악하는 동시에 예측 모델을 개발 및 적용하는 것을 목적으로 한다. 선박의 데이터는 크게 운항데이터와 기기데이터로 분류할 수 있으며, 운항데이터는 항로, 항해 정보, 대수속도, 대지속도, 바람과 같은 외력에 대한 정보 등이 있고, 기기데이터는 엔진출력, RPM, 연료 소모량, 기기들의 온도 및 압력 등이 있다. 본 연구에서, 우리는 선박에 미치는 외력의 영향을 Beaufort Scale (BFS)을 기준으로 구분한 후에 PLS 회귀분석을 통한 예측 모델을 개발하였다.

주요용어: 선박 운항 효율, 보퍼트 풍력계급, PLS 회귀분석

이 논문은 2016년 해양수산부 재원으로 한국해양과학기술진흥원의 지원을 받아 수행된 연구임(친환경선박 TCS 시스템 구축).

¹교신저자: (37673) 경북 포항시 남구 청암로 77, 포항공과대학교 산업경영공학과.

E-mail: chjun@postech.ac.kr