

## 유류화물 항만물동량 예측모형 개발 연구\*

김정은\*\* 오진호\*\*\* 우수한\*\*\*\*

### An introduction of new time series forecasting model for oil cargo volume

Kim, Jung-Eun Oh, Jin-Ho Woo, Su-Han

#### Abstract

Port logistics is essential for Korea's economy which heavily rely on international trade. Vast amounts of capital and time are consumed for the operation and development of ports to improve their competitiveness. Therefore, it is important to forecast cargo volume in order to establish the optimum level of construction and development plan. Itemized forecasting is necessary for appropriate port planning, since disaggregate approach is able to provide more realistic solution than aggregate forecasting. We introduce a new time series model which is Two-way Seasonality Multiplied Regressive Model (TSMR) to forecast oil cargo volume, which accounts for a large portion of total cargo volume in Korea. The TSMR model is designed to take into account the characteristics of oil cargo volume which exhibits trends with short and long-term seasonality. To verify the TSMR model, existing forecasting models are also used for a comparison reason. The results show that the TSMR excels the existing models in terms of forecasting accuracy whereas the TSMR displays weakness in short-term forecasting. In addition, it was shown that the TSMR can be applied to other cargoes that have trends with short- and long-term seasonality through testing applicability of the TSMR.

*Key words: Forecasting, time series, oil cargo volume*

▷ 논문접수: 2018. 02. 26. ▷ 심사완료: 2018. 03. 25. ▷ 게재확정: 2018. 03. 27.

\* 『본 논문은 중앙대학교 무역물류학과 석사논문을 요약 재정리 한 것임을 밝힙니다.』

\*\* 중앙대학교 무역물류학과 석사, 제1저자, kjungeun93@gmail.com

\*\*\* 중앙대학교 무역물류학과 박사과정, 공동저자, jamiroquai25@naver.com

\*\*\*\* 중앙대학교 국제물류학과 교수, 교신저자, shwoo@cau.ac.kr

## I. 서론

항만은 물류의 시작점으로 국가 경제 전반에 영향을 미치고 있는 주요 기간산업 시설이다(고용기 외, 2008). 따라서 정부는 항만물동량 예측을 기반으로 10년 단위의 항만 개발계획을 수립하며, 이를 통해 항만별 발전방향을 제시하고 있다(김정훈, 2008). 항만개발을 위한 항만시설의 정비 및 확충에는 막대한 자본과 시간이 소요되므로 항만 개발 계획은 정확한 물동량 예측모형을 기반으로 해야한다(Woo et al., 2012).

물동량 예측 방법은 여러 외생변수(환율, 경제성장률, 외교적 문제 등)를 고려하는 모형과, 기존의 시계열 데이터를 기반으로 미래를 예측하는 모형으로 구분된다(김종길, 2011). 항만 물동량의 변화는 외생변수의 영향을 배제할 수 없으므로 물동량 예측에 인과분석이 많이 사용된다. 하지만 이를 위해선 각각의 외생변수들에 대한 예측이 필요하며, 이 과정에서 예측의 오차범위가 커질 수 있다(민경창 외, 2013). 시계열 분석의 경우 외생변수를 고려할 수 없다는 단점이 있지만, 인과분석에 비해 우수한 예측력을 보여준다(방지영, 2015).

기존에는 주로 컨테이너 화물을 대상으로 물동량 예측 연구가 이루어졌다. 컨테이너 중심의 항만 개발 계획이 중점을 이루고 있으며, 벌크 화물 물동량이나 취급에 관한 관심도가 낮은 것이 그 이유이다. 하지만, 벌크화물은 항만에서 고부가가치를 창출할 수 있는 주요 화물로(전기정 외, 2013), 2016년까지 전체물동량의 약 30%를 차지하고 있다(해양수산부, 2016). 또한 컨테이너 화물과 달리 벌크화물의 경우 그 품목에 따라 처리 설비에 차이가 있으므로 컨테이너 화물과 벌크화물 물동량에 대한 예측이 병행되어야 더욱 효율적인 항만 운영이 가능해질 것이다. 따라서 본 논문은 우리나라 유류화물 물동량 예측에 있어 기존 시계열 분석모형을 검토하고 새로운 예측모형을 제시하는 것을 목적으로

한다. 특히 유류화물 물동량 예측은 기존 예측 모델로는 유의하거나 정확성이 높은 예측에 대한 어려움이 따르는 만큼 이를 고려한 새로운 예측 모델의 개발은 의의가 대단히 크다.

본 논문에서 새롭게 제시된 모형(TSMR : two way seasonality multiplied regressive model)은 시계열 데이터의 단기, 중장기의 주기적 특성을 승법 반영한 모형이다. 적합성과 예측력, 일반화 가능성을 기존 시계열 분석 모형들과의 비교를 통해 검증하고자 한다. 그리고 향후 유류화물만이 아닌 다양한 화물에서의 예측 방법으로서 활용 가치가 크다.

## II. 선행연구

물동량 예측 모형에 관한 연구는 새로운 예측 모형을 만들거나 모형 간 비교 검증을 하는 방식으로 이루어진다. 기존 예측 모형들 간 신뢰도와 정확도를 비교하여 데이터의 특성에 맞는 모형, 혹은 기존 모형의 단점을 보완한 새로운 모형을 제시한다. 물동량 예측 모형에 관한 선행연구들은 외생변수를 고려하는 인과모형(전찬영·송주미, 2007; Van et al., 2012; 김현석·장명희, 2015)과 과거의 데이터를 기반으로 미래를 예측하는 시계열 모형(신창훈 외, 2008; 조찬혁, 2008; Xie et al., 2013; Xiao et al., 2014)으로 구분할 수 있다.

전찬영·송주미(2007)는 회귀분석과 인공신경망모형의 우수성을 20개의 수출입품목을 통해 비교 및 분석하였다. 그 결과로 '상승지속형' 데이터는 회귀분석, '상승 둔화형'과 '정체형 또는 불규칙형'의 경우에는 인공신경망모형이 우수함을 주장하였다.

전찬영·송주미(2007)의 연구는 기존 대다수의 연구에서 활용되었던 두 가지 모형을 비교 분석하였지만, 신창훈 외(2008)와 Xie et al. (2013)의 연구는 기존의 모형과, 그 모형들을 결합한 하이브리

드 모형을 비교 분석하였다. 신창훈 외(2008)는 ARIMA와 인공신경망모형(ANN), 그리고 이 둘을 합친 하이브리드 모형을 국내 5개 항만 물동량을 활용하여 비교 분석하였다. 분석 결과 예측 모형의 적합도는 하이브리드 모형이, 모형의 예측력은 ARIMA모형이 가장 우수하다고 도출되었다. 하이브리드 모형의 경우 자료의 증감추세가 있을 때 예측력이 떨어지는 ANN모형의 단점으로 인해 ARIMA보다 예측력이 저하된 것으로 저자는 판단하였다.

Xie 외(2013)는 LSSVR(Least squares support vector regression)과 SARIMA(Seasonal autoregressive integrated moving average), SD(Seasonal decomposition), CD(classical decomposition)를 각각 결합

며, 윈터스 모형의 정확성이 더 뛰어난을 증명하였다. 이에 새로운 추정기법뿐만 아니라 비교적 단순하다고 여겨졌던 윈터스 모형도 염두에 두고 물동량 예측이 진행되어야 함을 주장했다.

신창훈·정수현(2011)은 ARIMA모형, ANN모형, ARIMA와 ANN을 결합한 하이브리드 모형에 대한 비교 분석을 실시하였고, 국내 컨테이너 물동량 데이터에 대해 ANN모형과 하이브리드 모형의 예측성이 뛰어난을 보였다. 하지만 하이브리드 모형의 유효성 검증 결과를 통해 하이브리드모형 또한 원자료에 손실 및 왜곡을 일으킬 수 있다는 한계점을 발견하였다.

Van et al.(2012)은 시스템다이내믹스와 인과분석

표 1. 물동량 예측 모형에 관한 선행연구

연구자	분석 대상 및 연구내용	방법론
전찬영·송주미(2007)	수출입품목 각각 10개	회귀분석과 인공신경망모형
신창훈 외(2008)	부산·광양·인천·울산·마산항 물동량	ARIMA, 인공신경망 모형, 하이브리드 모형(ARIMA, ANN)
Xie et al.(2013)	상해·선전항 컨테이너 물동량	LSSVR&SARIMA, LSSVR&SD, LSSVR&CD
조찬혁(2008)	동해항 시멘트 물동량	Holt-Winter, ARIMA
신창훈·정수현(2011)	국내 항만의 컨테이너 물동량	하이브리드 모형(ARIMA, ANN)
Van et al.(2012)	Le havre항의 장기 물동량	시스템다이내믹스, 인과분석
Xiao et al.(2014)	상해·닝보항 항만 물동량	TF-DPSO(transfer forecasting model guided by discrete particle swarm optimization)
민경창·하현구(2014)	단기 컨테이너 물동량	SARIMA
김현식·장명희(2015)	한국의 수출입 물동량	VAR(vector autoregressive), Bayesian VAR
정성환·강경우(2015)	국내 총 화물물동량	통상최소자승모형, 부분조정모형, 축소된 자기회귀분포시차모형, 벡터자기회귀 모형, 시간변동계수모형

한 하이브리드 모형을 상해, 선전항 물동량을 통해 검증하였다. 분석결과 각각의 모형들은 데이터의 계절적인 특성을 예측에 반영하지 못하였으나, 하이브리드 모형들은 계절성이 있는 시계열 데이터에서 더욱 강점을 드러내는 것으로 나타났다.

조찬혁(2008)은 동해항의 시멘트 물동량으로 윈터스 모형과 ARIMA 모형의 예측력을 비교하였으

를 활용하여 Le havre항의 물동량을 예측하였고, 이를 통해 항만 건설 계획을 위한 장기 물동량 예측 가이드라인을 제시하고자 하였다. 항만 물동량의 인과분석을 위해선 일반적으로 원유 가격, 지역별 무역 패턴, 세계화 등이 고려된다. 하지만 이와 같은 변수들은 2100년까지의 물동량을 예측하는 데 너무 세부적인 사항이라 고려하지 않고, GDP와 항

만 물동량 처리 실적만을 변수로 사용하였다.

Xiao et al.(2014)은 AC(analog complexing)와 입자 무리 최적화 알고리즘(particle swarm optimization algorithm)을 함께 고려한 TF-DPSO(transfer forecasting model guided by discrete particle swarm optimization) 모형을 제안하였고, ARIMA, ENN(Elman network), AC와 비교를 통해 모형을 검증하였다. 상해항과 Ningbo항의 물동량으로 모형을 검증한 결과 TF-DPSO가 상해항의 물동량 예측에서 MAE, MAPE, RMSE값이 모두 우수한 것으로 나타났다.

민경창·하현구(2014)는 ARIMA와 SARIMA의 비교분석을 실시하였다. 기존의 물동량 예측 연구에서는 월별데이터를 주로 사용하였지만, 민경창·하현구(2014)의 연구에서는 분기별 데이터가 분석에 활용되었다. 분석 결과 계절적 특성이 뚜렷한 분기별 자료에 높은 예측 정확도를 갖는 것으로 나타났다.

김현석·장명희(2015)는 환율과 운임지수가 물동량에 미치는 영향에 대해 VAR 모형과 Bayesian VAR 모형으로 분석하였으며, 두 모형의 예측력을 비교하였다. MAFE(Mean Absolute Forecast Error), RMSFE(Root Mean Squared Forecast Error), RMSPEE(Root Mean Squared Percentage Forecast Error)로 예측력을 비교한 결과 Bayesian VAR모형의 예측력이 더 뛰어난 것으로 나타났다. 시계열 분석의 경우 차분변환의 과정에서 데이터의 손실문제가 발생하는 반면, Bayesian VAR모형은 이러한 차분과정 없이 시계열 자료의 불안정성과 변동성을 반영할 수 있어 시계열분석의 한계를 보완했다고 할 수 있다. 하지만 김현석·장명희(2015)의 연구에서는 명목환율과 물가지수만을 변수로 반영하였기에 물동량에 영향을 미치는 다양한 요인들이 변수로써 더 고려되어야 할 것으로 보인다.

정성환·강경우(2015)는 통상최소자승모형(OLS), 부분조정모형(PA), 축소된 자기회귀분포시차모형

(READLM), 벡터자기회귀모형(VAR), 시간변동계수모형(TVP)을 활용하여 국내 총 화물물동량을 예측하였다. 분석결과 자료의 변동성이 큰 경우 시간변동계수모형, 점진적 추세를 갖는 경우 벡터자기회귀모형이 적절한 것으로 나타났다.

### III. 실 증 분 석

기존 물동량 예측에서 사용되어 왔던 ARIMA 모형, 윈터스가법, 단순계절모형과 새로운 물동량 예측모형으로서 제시하고자 하는 TSMR 모형과의 비교를 통하여 TSMR 모형의 우수성과 활용가능성을 논하고자 한다.

ARIMA모형은 단일지표에 대한 과거 추세를 반영하여 예측하는 모형으로 다양한 계절유형 뿐만 아니라 비계절 시계열에도 적용이 가능하다. 하지만 분석에 사용되는 시계열 데이터가 안정화되지 않았을 경우 유효한 예측치를 도출해낼 수 없다.

윈터스 가법은 시계열 데이터가 추세·계절·불규칙 성분들의 합으로 구성되어 있다는 가정을 기반으로 한다. 또한 시계열 데이터의 계절적 변동이 시간의 흐름에 관계없이 일정하다고 가정한다. 단순계절 모형은 시계열 데이터에 추세가 없고 일정한 계절효과를 갖는 다는 가정을 기반으로 한다.

예측 모델로서 새롭게 제시되는 TSMR 모형은 시계열 데이터가 추세·단기계절·중장기 계절 성분들의 곱으로 구성되어있다고 가정한다. 여기서 단기계절성분과 중장기 계절 성분은 시계열 데이터의 자체적인 특성에 따라 달라진다.

예측 모형들의 적합도를 판단하기 위해 오차분석을 실시했으며, 오차분석의 기준으로는 MAPE(mean absolute percentage error), RMSE(root mean square error), MAD(mean absolute deviation), bias, TS(tracking signal)를 사용하였다. TSMR 모형은 새롭게 제안되는 모형이므로 오차분석과 더불어

어 예측력 및 일반화 검증도 실시하였다. 수식에서 A는 실측치, F는 예측치, E는 오차를 의미한다.

$$MAPE_n = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

MAPE는 10%이하일 경우 해당 모형이 높은 정확도를 가졌다는 것을 의미한다 (Lewis, 1982). MAPE는 상대적 측정방법으로 실제 수요와 예측치 간 오차를 백분율로 똑같이 가중하므로, 다른 관측수를 지닌 예측 모형과의 비교가 가능하다. 따라서 수요예측모형 검증에 폭넓게 사용되고 있다.

$$RMSEn = \sqrt{\frac{1}{n} \sum_{t=1}^n E_t^2}$$

RMSE는 실제 관측값과 예측치의 편차(오차)제곱의 평균을 구한 다음 그 제곱근을 구하는 방법으로, 돌출 값(outliers)에 민감하게 반응한다. 값이 작을수록 모형이 상대적으로 잡음요소를 적게 포함하고 있다는 것을 의미하므로 적합도가 좋은 모형이라 할 수 있다(정동빈, 2009)

$$MAD_n = \frac{1}{n} \sum_{t=1}^n A_t$$

MAD는 오차 절대값의 평균이므로 오차의 방향은 고려하지 않고 오차의 크기만을 측정한다. 따라서 오차의 상대적 크기를 알 수 없다는 단점이 있다 (Meindl, P., & Chopra, S., 2001).

$$bias_n = \sum_{t=1}^n E_t$$

$$TS_t = \frac{bias_t}{MAD_t}$$

오차의 상대적 크기를 알 수 없는 MAD의 단점을 보완할 수 있는 통계량이 bias와 TS이다. bias는

오차의 합이므로 이를 통해 과다 추정 여부를 알 수 있다. 0에 가까울수록 적정한 수준에서 예측이 되었다는 것을 의미한다. TS는 그 값이 -6보다 작을 경우 과소 추정(underforecasting), 6보다 클 경우 과다 추정(overforecasting)되었다는 것을 의미한다. 하지만 분석 대상의 시계열 데이터가 급감, 급증하는 경우 값이 커지는 것이 불가피하므로 ‘-6 < TS < 6’의 범위가 절대적이라고 할 수는 없다 (Meindl, P., & Chopra, S., 2001).

분석에 사용된 데이터는 2005년 1월부터 2017년 4월까지의 월별 전국 유류화물 물동량이며 해운항만물류정보센터를 통해 자료를 수집하였다. 이 시기의 유류화물 물동량은 전반적으로 증가하고 있으며, 물동량 변동폭은 시간이 지남에 따라 커지는 양상을 보이고 있다. 유류화물 물동량은 원유(역청유류), 석유, 석유정제품, 석유가스 및 기타 가스류를 포함하고 있다. 한국해양수산개발원(2014)은 석유정제시설, LNG생산시설, 발전소가 위치한 항만의 물동량 처리실적으로 향후 물동량을 예측하는 Bottom-up방식과 총량을 예측하고 각 항만에 배분하는 Top-down 방식을 함께 사용하고 있다. 본 논문에서는 Bottom-up방식처럼 유류화물 처리시설이 있는 항만을 기준으로 물동량을 예측하기보다는 유류화물 물동량 데이터 자체의 특성을 반영하고자 항만을 특정 짓지 않고 전국 유류화물 물동량을 분석에 활용하였다.

## 1. ARIMA

ARIMA 모형으로 데이터를 분석하기 위해선 데이터의 안정화가 선행되어야 한다. 데이터가 안정화되었다는 것은 모든 시차의 계수가 신뢰 한계 안에 위치하고, Ljung-box 유의수준이 0.05보다 크다는 것을 의미한다. 잔차의 Ljung-box 유의 수준이 0.05보다 클 때, 잔차가 백색잡음으로부터 독립적이며, 적합한 모형을 추정한 것이라 판단한다.(안경모, 이광우, 2005; 한광중, 2015).

표 2. 차분, 계절차분의 자기상관함수 및 유의수준

구분	원 데이터		차분		계절차분			
			1		1		2	
시차	자기상관	유의수준	자기상관	유의수준	자기상관	유의수준	자기상관	유의수준
1	.734	.000	-.643	.000	.133	.116	-.021	.811
2	.785	.000	.351	.000	.325	.000	.278	.007
3	.668	.000	-.145	.000	.172	.000	.030	.018
4	.612	.000	-.001	.000	.249	.000	.166	.008
5	.563	.000	-.031	.000	.126	.000	.003	.018
6	.532	.000	-.051	.000	.209	.000	.134	.013
7	.524	.000	.015	.000	.269	.000	.222	.002
8	.514	.000	-.068	.000	.068	.000	-.007	.004
9	.545	.000	-.045	.000	.266	.000	.239	.000
10	.598	.000	.196	.000	-.002	.000	-.098	.000
11	.546	.000	-.329	.000	.225	.000	.220	.000
12	.663	.000	.521	.000	-.311	.000	-.544	.000

주) 유의수준은 Ljung-box 통계량의 유의수준을 의미

가정되는 기본 프로세스는 잡음으로부터 독립, 근사 카이제곱 근사값을 기준으로 함

분석에는 월별 데이터가 사용되었으므로, 시차는 12로 설정하였다. <표 2>에 나타난 모든 시차의 유의수준이 .000이므로 차분, 계절차분, 자연로그 변환 등을 통해 데이터를 안정화하고자 하였다.

차분 1의 경우 1,2,11,12 시차의 계수가 신뢰한계를 벗어났으며, 시차 1의 계수가 -.643, 시차 12의 계수가 .521으로 이미  $\pm 0.5$ 를 넘어서 과대 차분(Over-differencing)이 되었음을 알 수 있다. 그러므로 더 이상의 차분은 무의미하였다.

계절차분 1의 결과를 보면, 데이터가 어느 정도 안정화 된 것을 알 수 있다. 시차 2와 12의 계수가 신뢰한계를 가장 많이 벗어났지만, 그 수준이 각각 .325와 -.311에 머물렀으므로 과잉 차분되지 않았다. 따라서 계절 차분을 한 차례 더 실시하였다. 그 결과 5개 시차의 계수가 여전히 신뢰 수준을 벗어나고 있었고, 시차 12의 경우에는 계수가 -.544로

과잉 차분이 되었다. 또한 0.05이상의 Ljung-box유의 수준을 갖는 시차 역시 1개밖에 존재하지 않았다. 따라서 더 이상의 차분과 계절차분은 데이터의 안정화에 기여할 수 없음을 알 수 있었다.

차분과 계절 차분의 결과 단일 차분으로는 데이터의 안정화가 불가능하다고 판단되어 자연로그변환과 차분을 병행하였다. 자연로그변환만 한 결과 안정화 이전의 기존 데이터의 자기상관함수와 큰 차이가 없었다. 모든 시차의 계수가 신뢰한계를 넘었으며, Ljung-box유의 수준도 .000에 머물렀다.

<표 3>에서 알 수 있듯, 자연로그변환과 차분1을 병행한 결과 Ljung-box유의수준은 .000에 머물렀으며, 5개 시차에서 신뢰한계 수준을 넘어섰고, 그 중 2개 시차에서 계수가 -.640와 -.500로 과잉 차분되었다.

표 3. 자연로그변환 후 자기상관함수

구분	자연로그변환							
	-		차분1		계절차분1		계절차분2	
시차	자기상관	유의수준	자기상관	유의수준	자기상관	유의수준	자기상관	유의수준
1	.737	.000	-.640	.000	.146	.085	-.001	.994
2	.789	.000	.350	.000	.299	.000	.250	.019
3	.673	.000	-.157	.000	.163	.000	.032	.044
4	.625	.000	.022	.000	.222	.000	.146	.028
5	.571	.000	-.044	.000	.086	.000	-.025	.052
6	.540	.000	-.053	.000	.188	.000	.130	.040
7	.533	.000	.018	.000	.255	.000	.225	.006
8	.522	.000	-.059	.000	.059	.000	.003	.010
9	.547	.000	-.050	.000	.241	.000	.227	.001
10	.596	.000	.191	.000	.002	.000	-.073	.002
11	.545	.000	-.316	.000	.205	.000	.205	.000
12	.655	.000	.500	.000	-.338	.000	-.557	.000

주) 유의수준은 Ljung-box 통계량의 유의수준을 의미  
가정되는 기본 프로세스는 잡음으로부터 독립, 근사 카이제곱 근사값을 기준으로 함

자연로그변환과 계절차분을 병행하였을 때 6개 시차가 여전히 신뢰한계 수준을 넘고 있으며 .05이상의 Ljung-box유의수준은 1개 시차에서만 발견되었다. 하지만  $\pm 0.5$  수준을 넘어선 계수가 존재하지 않아 계절차분을 한 번 더 실시하였다. 그 결과 적절한 유의 수준은 2개 시차에서 발견되었고, 신뢰한계를 벗어난 시차는 3개로 줄어들었다. 그러나 시차 12의 계수가 -.557로 나타나 더 이상의 차분은 무의미한 것으로 도출되었다.

ARIMA모형을 활용하기 위해선 데이터의 안정화가 선행되어야 하지만, 본 논문에서 사용한 데이터는 <표 2>, <표 3>에서 알 수 있듯, ARIMA모형에 적합한 안정화가 불가능하였다. 차분과 계절차분을 함께 고려할 수도 있지만, 이 경우 결과 해석의 어려움이 있으므로 진행하지 않았다(한광중, 2015). 그러므로 ARIMA 모형이 많은 연구에서 효율적으로 활용되었지만, 전국 유류화물 물동량 분석에는 적합하지 않은 모형으로 판명되었다.

## 2. 윈터스가법

윈터스 가법 모형의 결과는 <표 4>와 같다. 윈터스 가법 모형은 데이터의 89%를 설명하며, MAPE 값이 3%수준이므로 정확도가 매우 높다. 전체적인 bias는 음수 이므로 윈터스 가법 모형을 사용할 경우 실제보다 적은 수준으로 예측이 이루어졌다. 15개의 TS가 6을 초과하지만, 총 데이터가 148개이므로 적은 수준이며, 최댓값 역시 9에 머물러 오차가 크지 않았다. 또한 Ljung-Box 통계량의 유의수준은 0.05보다 크면 백색잡음으로부터 독립적이라 할 수 있는데, 윈터스 가법의 경우 유의수준이 .156으로 .05를 넘어서고 있어 분석에 적합한 모형이었다. 하지만 모형 모수의 유의수준이 알파를 제외하고 0.05보다 크게 나타나 통계적으로 유의미하지 않았다.

표 4. 윈터스 가법 모형 오차분석 결과

구분		값
모형 통계	Ljung-Box 유의수준	.156
모형 적합	$R^2$	
	.894	
	MAPE	
	0.033	
	RMSE	
	1,259,606	
	MAD	
	1,007,024	
	bias	최댓값
		8,862,048
최솟값		
-5,494,898		
TS	전체	
	-5,014,935	
	최댓값	
	9	
최솟값		
-6		
>6 의 개수		
15		
<-6 의 개수		
1		
모형 모수	알파(수준)	
	0	
	감마(추세)	
0.999		
델타(계절)		
1		

표 5. 단순계절 모형 오차분석 결과

구분			값
모형 통계	Ljung-Box 유의수준		0
모형 적합	$R^2$		0.679
	MAPE		0.055
	RMSE		2,164,428
	MAD		1,710,321
	bias	최댓값	6,977,971
		최솟값	-49,100,381
		전체	-46,745,013
	TS	최댓값	4
		최솟값	-29
		>6 의 개수	0
<-6 의 개수		94	
모형 모수	알파(수준)		0
	감마(추세)		-
	델타(계절)		-

### 3. 단순계절모형

〈표 5〉는 단순계절 모형의 적합성 결과를 정리한 것이다. 단순계절 모형의 모형모수는 모두 분석에 적합한 수준으로 도출되었다. MAPE 역시 5% 수준으로 모형의 정확도가 높음을 알 수 있으며, TS의 최댓값은 6이하로 매우 적절한 수준에 머무르고 있다. 하지만,  $R^2$  값이 0.67이므로 모형의 설명력이 다소 떨어지며, Ljung-Box 유의수준은 0.05보다 작아 백색잡음으로부터 독립적이지 못했다. 또한 -6보다 작은 TS가 94개 존재하고, 최솟값이 -29로 적절한 수준과 다소 차이가 있었다.  $R^2$ 과 TS 모두 절대적인 기준이 있는 것은 아니지만,  $R^2$  값과 Ljung-Box 유의수준을 보았을 때 단순계절모형은 분석에 적합하지 않았다.

### 4. TSMR

#### 1) 모형 개요

TSMR 모형에서는 물동량의 단기, 중장기 특성을 모두 예측에 반영함으로써 모형의 정확도를 높이고자 하였다. TSMR모형의 표기 방법은 아래와 같으며, 분석절차는 다음과 같다. TSMR모형의 표기에는 이동평균의 주기, 단기 계절 주기, 장기 계절 주기를 포함하여 시계열 분석 특성을 한 번에 알 수 있도록 하였다. 〈표 6〉에서의  $t$ 는 시차,  $st$ (short term)는 단기,  $lt$ (long term)는 장기를 의미한다.

$$TSMR(m,s,l)$$

$m$  = 이동평균 주기

$s$  = 단기 계절 주기

$l$  = 장기 계절 주기



표 6. TSMR 분석 절차

계절지수산출 (Compute Seasonal Index)	이동평균(Moving Average)
	계절지수 값( $siv_t$ ) = 실측치( $D_t$ ) / 이동평균값( $MA_t$ )
	단기계절지수( $SI_{st}$ )= $SIV_t, SIV_{t+s}, SIV_{t+2s}, \dots$ 의 평균 중장기계절지수( $SI_{lt}$ )= $SIV_t, SIV_{t+1}, SIV_{t+2l}, \dots$ 의 평균
시계열의 계절성 제거 (Deseasonalize)	실측치( $D_t$ ) / 단기계절지수( $SI_{st}$ ) / 중장기계절지수( $SI_{lt}$ )
회귀분석(Regression)	$F_t = ax + b$ 독립변수 : $t$ 종속변수 : 계절성이 제거된 시계열(Deseasonalized $D_t$ )
예측(Forecasting)	$F_t * \text{단기계절지수}(SI_{st}) * \text{중장기계절지수}(SI_{lt})$

TSMR모형의 분석은 총 4단계로 이루어진다.

첫 번째 계절지수 산출 단계에서는 이동평균을 통해 데이터의 변동 폭을 줄이고, 이동평균의 결과로 실측치를 나누어 계절지수 값을 계산한다. 단기계절지수( $SI_{st}$ )는  $t, t+s, t+2s, \dots$  시점의 계절지수 값들의 평균으로 산정한다.

장기계절지수( $SI_{lt}$ )도 단기계절지수( $SI_{st}$ )와 마찬가지로  $t, t+1, t+2l, \dots$  시점의 계절지수 값들의 평균으로 산정한다.

두 번째, 시계열의 계절성 제거 단계에서는 실측치를 단기계절지수( $SI_{st}$ )와 장기계절지수( $SI_{lt}$ )로 나눈다. 단기, 중장기적 계절성의 제거를 통해 시계열의 추세를 좀 더 명확하게 파악할 수 있다.

세 번째 단계에서는 시차  $t$ 를 독립변수로, 계절성이 제거된 시계열 데이터를 종속변수로 한 회귀분석을 실시한다.

네 번째 단계에서는 회귀분석을 마친 데이터( $F_t$ )에 단기계절지수( $SI_{st}$ )와 장기계절지수( $SI_{lt}$ )를 곱하여 계절성을 복구한다.

유류화물 물동량은 약 12개월 단위로 단기적인 계절성을 보이고 있으며, 30개월 단위로 저점과 고점이 반복되고 있다. 따라서 분석에는 TSMR(9,12,30) 모형이 활용되었다. 회귀식은  $y = 25168386 * x + 74345.62$  이었으며 12개월, 30개월 단위의 계절지수는 <표 7>, <표 8>과 같다. 평균에 보다 많은 지수들이 고려되는 12개월 단위 계절지수에 비해 30개월 단위 계절지수의 분산이 더 크다. 그러므로 30개월 단위의 계절지수를 예측에 포함시킴으로써 시계열 데이터에 중장기적으로 나타나는 큰 변동을 효과적으로 반영시킬 수 있었다.

표 7. 12개월 단위 계절지수

시차	계절지수	시차	계절지수
1	1.088	7	1.000
2	0.973	8	0.965
3	1.062	9	0.972
4	0.938	10	1.011
5	1.002	11	0.976
6	0.949	12	1.053

표 8. 30개월 단위 계절지수

시차	계절지수	시차	계절지수
1	1.060	16	0.965
2	0.950	17	1.020
3	1.022	18	0.963
4	0.969	19	1.049
5	0.990	20	0.974
6	0.997	21	1.013
7	1.020	22	0.979
8	0.970	23	0.960
9	1.000	24	1.032
10	0.977	25	1.028
11	0.992	26	0.962
12	0.998	27	0.998
13	1.055	28	0.995
14	0.985	29	0.984
15	1.037	30	1.017

## 2) 오차분석

ARIMA 모형의 경우 분석을 위한 선행 조건이 충족되지 않아 분석이 불가했고, 따라서 오차분석 결과가 도출되지 않았다. 윈터스 가법은 모형통계와 모형모수, 단순계절 모형은 모형 통계가 적합하지 않았으나, TSMR모형 검증에 위한 분석이 가능하였기에 오차분석 결과를 도출하였다. 즉, 윈터스 가법모형, 단순계절 모형 모두 유통화물 물동량 예측에는 적합하지 않았지만, TSMR모형의 검증을 위해

적합도를 비교하였다. 윈터스 가법 모형과 단순계절 모형의 적합도를 비교한 결과는 <표 9>와 같다.

TSMR 모형의 설명력은 윈터스 가법 모형보다는 약 5%p 낮지만, 단순계절 모형에 비해 약 17%p 높은 85%로 나타났다. MAPE를 기준으로 세 모형 모두 매우 높은 정확도를 갖고 있으며, TSMR모형이 4%로, 윈터스 가법 모형보다는 높지만, 단순계절 모형에 비해 1%p이상 낮게 도출되었다. TSMR의 RMSE값은 단순계절 모형보다 약 70%, MAD값은 약 73% 수준에 머물러 낮은 수준을 유지했다. 적정수준의 TS 범위를 넘어선 값들의 개수는 단순계절 모형이 94개, TSMR이 85개, 윈터스가법 모형이 16개로 TSMR 모형에서 과다 혹은 과소 추정되는 예측치가 비교적 많은 것으로 나타났다. TSMR 모형에서 bias와 TS의 최댓값이 단순계절 모형에 비해 큰 것으로 나타났다. 또한 TS의 최댓값이 단순계절 모형과 비교해 약 4배정도 차이가 나고 있으며 6을 넘어서는 값의 개수 또한 월등히 많았다. 이는 TSMR이 단순계절 모형보다 과다 추정하는 경우가 많다는 것을 의미한다. 하지만 TSMR에서 -6 이하의 값을 가지는 TS의 개수는 32개, 6이상의 값을 가지는 TS는 53개로 모형 자체적으로는 과소 추정된 경우가 많았다. 즉, TSMR은 단순계절 모형에

표 9. 모형별 오차분석 결과 비교

구분		TSMR	윈터스 가법 모형	단순계절 모형
모형통계	Ljung-Box 유의수준	-	.156	0
모형적합	$R^2$	0.846	.894	0.679
	MAPE	0.042	0.033	0.055
	RMSE	1,521,274	1,259,606	2,164,428
	MAD	1,259,586	1,007,024	1,710,321
	bias	최댓값	20,337,861	8,862,048
		최솟값	-15,548,676	-5,494,898
		전체	4,612,206	-5,014,935
	TS	최댓값	16	9
		최솟값	-14	-6
		>6 의 개수	53	15
		<-6 의 개수	32	1
모형모수	알파(수준)	-	0	0
	감마(추세)	-	0.999	-
	델타(계절)	-	1	-

비해서는 예측치가 과다 추정되었다고 볼 수 있지만, TSMR모형 자체만을 보았을 때는 전반적으로 과소 추정되고 있다는 것을 알 수 있다.

TSMR의 모형 적합도를 윈터스 가법 모형, 단순 계절 모형과 비교한 결과 TSMR모형은 전반적으로 우수한 결과를 보였으나, 과다추정의측면에서 취약하고 윈터스가법 모형에 비해서는 모형적합도가 낮은 것으로 드러났다.

### 3) 예측력 검증

예측 결과와 실제 데이터와의 비교를 통하여 정확도에 대한 검증이 필요하다. 이를 위해서 2015년 이후의 실측 데이터를 삭제하고 2001년 1월부터 2014년 12월까지 월별데이터로 2015년 1월부터 2017년 4월까지의 물동량을 예측하고, 실측치와 비교하였다.

예측결과는 <표 10>과 같다. 전반적으로 TSMR의 예측치는 실측치보다 낮은 수준에서 이뤄졌다. 이는 앞서 모형 적합성 검증에서 도출된 결과와 상통한다. MAPE는 총 28개월 중 2개월에서만 10%를 넘는 수치가 나타났으며, 나머지 26개월의 MAPE는 모두 10%이하를 유지하였다. 이는 28개의 예측치 중 26개는 매우 높은 정확도로 예측되었고, 2개는 적절한 수준에서 예측되었다는 것을 의미한다. TS는 2015년 1월부터 2016년 1월까지의 매우 적절한 범위 내에 존재했다. 하지만 16년 2월부터 적정 범위를 넘어서기 시작했으며 마지막 예측치인 2017년 4월에 적정범위를 가장 많이 벗어난 값이 나타났다.

앞서 예측된 결과에 대한 적합도 결과는 <표 11>과 같다. TSMR 모형의 설명력은 84%로 높은 수준이었으며, MAPE는 4.6%로 높은 정확도를 유지했다. 이전의 모형 적합도 검증에서 적정수준을 넘어서는 TS의 개수는 148개 예측치 중 77개로 약 52%였으나, <표 11>에서는 28개 예측치 중 15개인 약 53%로 그 비중이 증가했다.

표 10. TSMR 모형의 예측력 검증 결과

연도	월	MAPE	bias	TS
'15	1	0.013	474,163	1
	2	0.079	-2,128,487	-1
	3	0.006	-1,921,075	-2
	4	0.002	-1,855,364	-2
	5	0.007	-2,096,852	-3
	6	0.009	-1,804,238	-3
	7	0.098	1,353,247	1
	8	0.093	-1,832,989	-1
	9	0.014	-2,293,450	-2
	10	0.046	-3,894,364	-3
	11	0.036	-5,142,016	-4
	12	0.027	-6,164,434	-5
'16	1	0.013	-6,681,927	-6
	2	0.105	-10,469,648	-8
	3	0.062	-12,900,318	-9
	4	0.005	-13,046,145	-10
	5	0.083	-16,112,439	-11
	6	0.097	-19,553,870	-13
	7	0.019	-18,869,406	-13
	8	0.015	-18,405,136	-13
	9	0.040	-19,816,766	-14
	10	0.084	-23,001,617	-15
	11	0.076	-25,693,827	-16
	12	0.001	-25,660,238	-17
'17	1	0.045	-27,490,383	-18
	2	0.105	-31,267,971	-19
	3	0.064	-33,872,384	-20
	4	0.046	-35,501,814	-21

ARIMA 모형의 경우 오차분석에서와 마찬가지로 데이터가 안정화 되지 않아 유효한 분석 예측치를 도출할 수 없었다. 윈터스 가법 모형은 모형모수 중 델타 값이 0.98로 0.05를 넘어서 통계적으로 유의미하지 않았다. 모형적합도를 보더라도  $R^2$ 과 MAPE는 유의미한 수준이지만, bias와 TS값은 TSMR모형에 매우 뒤처지는 것으로 나타났다. 단순 계절 모형도 윈터스 가법 모형과 마찬가지로 모형모수에서 0.05를 넘는 값이 나와 통계적으로 유의미하지 않았다. 또한 MAPE, RMSE, MAD, bias, TS 모두 TSMR에 비해 부정적으로 나타났으며 적게는 2배, 많게는 3배까지 차이가 났다.

표 11. 모형별 예측력 비교

구분		TSMR	윈터스 가법	단순 계절
모 형 적 합	$R^2$	0.845	0.878	0.877
	MAPE	0.046	0.08	0.096
	RMSE	2,084	310	3,681
	MAD	1,652	2,885	3,475
	b i a s	최댓값	1,353	-7,447
		최솟값	-35,502	-80,784
		전체	-35,502	-80,784
	T S	최댓값	1	-2
		최솟값	-21	-28
		>6 의 개수	0	0
		<-6 의 개수	15	23
모 형 모 수	알파(수준)	-	0	0
	감마(추세)	-	0.013	-
	델타(계절)	-	0.98	1

주) RMSE, MAD, bias의 단위 : 천

## 4) 일반화 검증

대표적으로 유류화물 물동량 예측에 TSMR 모형이 활용되었지만, 다양한 화물에 쓰일 수 있는지에 대한 범용성 차원에서의 검증이 필요하다. 이를 증명하기 위해 유류화물 이외에 철강, 유연탄, 기계류 화물에도 모형을 적용하였다. 분석대상으로는 유류 화물과 마찬가지로 2005년 1월부터 2017년 4월까지 총 148개월의 월별 물동량 데이터가 사용되었다. 일반화 검증에 사용된 3가지 품목 모두 데이터가 안정화가 이루어지지 않아 ARIMA모형을 활용한 분석은 불가능하였다.

철강의 시계열 분석에는 TSMR(9,12,40), 윈터스 승법, 단순계절 모형이 활용되었으며, 모형별 오차 분석 결과는 <표 12>와 같다.

모형 적합도의 경우 윈터스 승법모형과 단순계절 모형의 결과가 TSMR모형보다 더 적합하게 도출되었다. 하지만, 윈터스 승법과 단순계절 모형 모두

표 12. 일반화 검증(철강)

구분		TSMR (9,12,40)	윈터스승 법	단순 계절
모 형 적 합	$R^2$	0.832	0.88	0.89
	MAPE	0.067	0.058	0.056
	RMSE	709	587	573
	MAD	538	465	449
	b i a s	최댓값	11,121	3,512
		최솟값	-10,693	-2,374
		전체	754	2,905
	T S	최댓값	26	9
		최솟값	-21	-6
		>6 의 개수	68	6
		<-6 의 개수	54	1
모 형 모 수	알파(수준)	-	0	0
	감마(추세)	-	0.957	-
	델타(계절)	-	0.055	1

주) RMSE, MAD, bias의 단위 : 천

모형 모수에서 0.05를 넘는 값이 있었으므로 통계적으로 유의미하지 않았다.

TSMR모형 자체의 적합도만 보았을 때는 설명력이 80%를 넘고, MAPE도 10% 이하로 유지되고 있었다. 다만 적정수준을 넘어서는 TS의 개수가 전체 데이터의 약 82%로 높은 수준이었다. 세 모형 모두 실측치보다 낮은 수준의 예측치가 도출되었다.

유연탄의 시계열 분석에는 TSMR(9,12,50), 윈터스 가법, 단순계절 모형이 사용되었으며, 오차분석을 비교한 결과는 <표 13>과 같다.

모형 적합도의 관점에서 윈터스 가법과 단순계절 모형이 TSMR보다 우수한 결과를 보였으나, 두 모형 모두 모수가 적합하지 않아 통계적으로 유의미하지 않았다.  $R^2$ 과 MAPE를 비교 하였을 때 TSMR모형이 두 모형에 비해서는 결과가 좋지 않았다. 하지만 절대적인 기준에서는 TSMR의 예측력은 우수한 수준인 것으로 판단된다. TS값을 보았을 때

표 13. 일반화 검증(유연탄)

구분		TSMR (9,12,50)	윈터스 가법	단순 계절
모 형 적 합	$R^2$	0.854	0.891	0.885
	MAPE	0.07	0.062	0.063
	RMSE	763	658	675
	MAD	614	521	536
	b I a s	최댓값	13,978	2,617
		최솟값	-14,495	-6,261
		전체	1,408	-528
	T S	최댓값	33	5
		최솟값	-25	-14
		>6 의 개수	61	0
		<-6 의 개수	54	21
모 형 모 수	알파(수준)	-	0	0
	감마(추세)	-	0.999	-
	델타(계절)	-	0.995	1

주) RMSE, MAD, bias의 단위 : 천

TSMR 모형은 과대 추정의 경향을 보였으며, 전체 데이터 중 약 77%의 데이터가 적절한 수준의 TS범위를 넘어선 것으로 나타났다. 나머지 두 모형은 과소추정의 경향이 두드러졌다.

기계류는 TSMR(5,12,70), 윈터스가법, 단순계절 모형으로 분석하였으며 결과는 <표 14> 와 같다.

윈터스가법과 단순계절 모형의 설명력은 95%로 매우 높게 나왔으며, MAPE는 4%로 매우 높은 정확성을 가지고 있는 것으로 나타났다. 하지만 철강, 유연탄의 분석에서와 마찬가지로 모형모수의 유의확률이 0.05를 넘어섰으므로 통계적으로 유의하지 않았다. 단순계절 모형은 -6이하의 TS값이 118개로 과소추정 경향이 두드러졌다. 윈터스 가법모형은 TS의 최댓값이 9, 최솟값이 -6이었으며 적정 범위를 넘어선 TS의 개수가 8개밖에 나오지 않아 세 모형 중에서 TS를 기준으로 했을 때 가장 우수한 모형이었다. TSMR은 나머지 윈터스 가법모형에 비해 적정 범위를 넘어서는 데이터가 약 11배가 많

표 14. 일반화 검증(기계류)

구분		TSMR (9,12,50)	윈터스 가법	단순 계절
모 형 적 합	$R^2$	0.882	0.958	0.956
	MAPE	0.07	0.04	0.042
	RMSE	938	554	573
	MAD	754	433	459
	b I a s	최댓값	16,978	4,070
		최솟값	-230,904	-2,292
		전체	2,045	-603
	T S	최댓값	22	9
		최솟값	-34	-6
		>6 의 개수	51	7
		<-6 의 개수	38	1
모 형 모 수	알파(수준)	-	0	0
	감마(추세)	-	1	-
	델타(계절)	-	1	0.285

주) RMSE, MAD, bias의 단위 : 천

았지만, 전체에서 차지하는 비중은 60%로 타 품목에 비해 낮은 수준을 유지했다.

TSMR모형의 설명력은 세 가지 품목 모두 80%보다 높았으며, MAPE 역시 유류화물의 정확도보다는 낮지만 모두 10%미만으로 상당히 높은 정확도를 가지고 있었다. 하지만 적절한 TS 범위를 벗어나는 예측치가 철강 122개, 유연탄 115개, 기계류 89개로 77개였던 유류 물동량에 비해 높게 나타났다. 또한 유류 물동량 예측치의 경우 전반적으로 과소 추정되는 양상을 보였으나, 철강, 유연탄, 기계류 모두 과다 추정된 예측치의 개수가 많았다.

TSMR 모형의 월별 예측치는 적절한 과소 혹은 과대 추정될 확률이 높으나, 전체적인 관점에서 모형의 설명력이나 정확도는 매우 우수한 수준이었다. 따라서 TSMR 모형은 미시적인 월별 예측보다는 1년 이상의 거시적인 예측에 활용되는 것이 적합하다고 판단된다.

표 15. 물동량 예측치

(단위 : 톤)

연도	월	TSMR예측치	연도	월	TSMR예측치	연도	월	TSMR예측치
2017	1		2021	1	45,327,583	2025	1	49,668,317
	2			2	38,024,746		2	39,958,171
	3			3	43,648,654		3	46,854,428
	4			4	36,024,603		4	39,394,612
	5	35,758,490		5	40,694,205		5	43,054,976
	6	35,054,490		6	36,451,136		6	41,111,921
	7	38,593,784		7	41,950,658		7	44,416,289
	8	33,409,029		8	37,621,158		8	40,815,960
	9	36,297,929		9	39,510,473		9	42,479,406
	10	35,843,989		10	39,744,040		10	43,198,511
	11	35,466,066		11	37,721,808		11	42,448,761
	12	38,616,987		12	43,853,065		12	46,164,565
	총계	289,040,765		총계	480,572,129		총계	519,565,916
2018	1	40,862,896	2022	1	45,158,896	2026	1	50,444,216
	2	34,919,541		2	37,952,546		2	42,308,966
	3	39,247,767		3	42,910,955		3	48,557,297
	4	34,003,761		4	37,961,029		4	40,068,281
	5	36,930,227		5	40,159,244		5	45,253,505
	6	35,242,986		6	39,359,773		6	40,527,439
	7	39,342,870		7	43,324,070		7	46,633,241
	8	35,505,971		8	37,495,492		8	41,812,679
	9	37,756,390		9	40,728,719		9	43,904,332
	10	36,591,400		10	40,210,483		10	44,155,697
	11	37,414,057		11	39,777,768		11	41,901,261
	12	38,207,341		12	43,302,262		12	48,702,885
	총계	446,025,207		총계	488,341,239		총계	534,269,800
2019	1	43,058,972	2023	1	45,810,654			
	2	35,902,268		2	39,139,151			
	3	40,711,162		3	43,980,860			
	4	34,882,594		4	38,096,226			
	5	36,584,927		5	41,366,004			
	6	37,313,939		6	39,467,648			
	7	39,248,219		7	44,049,591			
	8	35,485,175		8	39,745,219			
	9	37,166,921		9	42,255,374			
	10	38,608,633		10	40,942,924			
	11	36,970,238		11	41,854,611			
	12	41,309,570		12	42,733,096			
	총계	457,242,620		총계	499,441,358			
2020	1	44,526,068	2024	1	48,149,365			
	2	35,828,356		2	40,138,256			
	3	42,020,189		3	45,505,107			
	4	35,337,023		4	38,982,151			
	5	38,627,973		5	40,876,144			
	6	36,891,939		6	41,682,126			
	7	39,864,913		7	43,833,897			
	8	36,640,643		8	39,623,129			
	9	38,141,321		9	41,492,578			
	10	38,794,486		10	43,093,383			
	11	38,128,512		11	41,256,375			
	12	41,474,094		12	46,089,552			
	총계	466,275,518		총계	510,722,063			

#### IV. 유류화물 물동량 예측

2017년 4월부터 10년간의 물동량 예측 결과는 <표 15>와 같다. 유류 물동량은 2020년에 4.6억 톤, 2024년에 처음으로 연간 5억 톤의 물동량을 처리할 것으로 예측되었다. 2018년 1월에 처음으로 예측치가 4천만 톤을 넘어섰으며, 2026년까지 매년 1월의 물동량이 가장 높은 것으로 예측되었다. 2021년부터 1년 중 5개월의 예측치가 4천만 톤을 넘어섰다. 2022년부터는 한해의 절반 이상의 예측치가 4천만 톤을 넘을 것으로 예측되었다. 2026년 1월에는 처음으로 월별 물동량이 5천만 톤을 넘을 것으로 예측되었다.

TSMR모형의 적합성과 예측력, 일반화 가능성을 살펴 본 결과 적합성과 예측력은 기존에 활용되고 있던 시계열 분석 모형에 비해 예측력이 상대적으로 높은 것으로 나타났다. 따라서 ARIMA모형과 윈터스 모형이 분석에 적합하지 않고, 단기적인 계절성과 중장기적인 주기성을 동시에 보이는 시계열 데이터의 예측에 기존 모형의 대안으로 TSMR을 고려할 수 있을 것으로 보인다.

TSMR은 외생변수를 고려하지 않고 과거의 데이터만을 활용해 미래를 예측했으므로 물동량이 단기간에 큰 폭 상승하는 경우는 없었다. 오히려 단기적으로는 과소 추정된 경우가 더 많았다. 하지만, 계절지수를 승법 적용했기 때문에 물동량 증가의 폭이 점차적으로 커지는 양상을 보였고, 2020년 이후 5년에 약 5백만 톤씩 물동량이 증가할 것으로 전망되었다.

본 논문의 분석결과를 종합해보면 TSMR은 물동량 예측을 위하여 사용된 ARIMA, 윈터스 모형, 단순계절에 비하여 예측력이 우수한 것으로 나타났다. 물동량 예측 모델로서 ARIMA가 주로 활용되고 있지만, ARIMA의 제한된 활용범위를 보완하고 오차를 줄인 한 단계 진전된 모델인 TSMR을 활용하여 더 정확한 미래 예측이 가능하다. 그리고 이의

화물들의 예측 모델로서도 활용이 가능함에 따라 다양한 분야에서도 활용 가능한 만큼 의의가 크다.

#### V. 결 론

항만개발 계획은 물동량의 흐름을 기반으로 한다. 물동량 예측으로 항만 개발계획의 방향과 항만 시설 기능 재배치, 시설 증대 등의 정책이 결정된다. 이러한 정책들은 막대한 시간과 자본의 투자를 필요로 하며, 국민 경제의 전반적인 부분에 영향을 주고 있다. 이 점에서 물동량 예측 방식의 비교 분석과 적합도, 예측력, 정확성에 대한 확인 및 검증은 매우 중요하다. 기존의 연구에서는 물동량 예측의 대상으로 컨테이너 화물 이외에 개별적인 품목을 설정한 경우가 드물었다. 하지만, 유류화물은 우리나라 총 물동량 중 약 30%를 차지하는 품목으로 컨테이너 화물을 제외하고는 가장 큰 비중을 차지하고 있다. 또한 유류화물 처리 시설의 건설에 막대한 자본이 투자되고 있다. 따라서 본 연구에서는 유류화물의 특성을 감안한 새로운 모형을 제시하고자 하였다.

물동량은 외생변수에 많은 영향을 받기 때문에 물동량 예측에 인과분석이 활용되기도 한다. 하지만 인과분석의 경우 외생변수의 예측이 선행되어야 하므로 잘못된 변수 설정이나 예측으로 오차의 폭이 커질 위험이 있다. 시계열 분석은 외생변수를 고려하지 못한다는 단점을 갖고 있다. 하지만 본 연구에서는 외생변수로 인한 물동량의 변화보다는 과거 시계열 데이터의 단기, 중장기적 특성을 반영해 정확도 높은 예측치를 얻고자 하였다.

예측에는 2005년 1월부터 2017년 4월까지의 월별 데이터가 사용되었다. 유류화물의 시계열 분석에는 ARIMA모형, 윈터스 가법모형, 단순계절 모형과 본 논문에서 새롭게 제안된 TSMR모형이 고려되었다. 기존 모형들과의 비교 결과 TSMR모형의 적

합도는 우수한 것으로 나타났다. 예측력 검증 결과 물동량이 과소 추정되는 경향이 두드러졌으나, 다른 품목에 적용해본 결과 모든 시계열 데이터에 대해 과소 추정되지는 않았다. 모형 검증에는  $R^2$ , MAPE, RMSE, MAD, bias, TS가 사용되었으며, bias와 TS에서 적정수준을 벗어난 이상 값이 측정되었다. 하지만 bias와 TS 경우 모형의 결격에 대한 명확한 기준이 있는 것이 아니므로 시계열 분석 모형으로써 TSMR은 충분히 활용 가능하다. 다만 분석 결과를 활용하는 데 있어 단기적인 물동량보다는 중장기적인 물동량의 흐름을 파악하는 것에 중점을 두는 것이 바람직 할 것으로 보인다.

TSMR 모형은 단기와 중장기의 주기적 특성과 추세가 있는 시계열 데이터 분석 시 기존 ARIMA 모형과 윈터스모형, 단순계절 모형의 새로운 대안이 될 수 있을 것으로 판단된다. ARIMA모형의 경우 단기적인 예측에 주로 사용되므로, 중장기적 예측에 적합한 TSMR모형과 상호보완적인 활용이 가능할 것이다.

## 참고문헌

- 고용기 · 김은지 · 신정용 · 김태호(2008), 인천항의 수출 적 컨테이너화물 물동량 추정에 관한 연구, 한국항만경제학회, 제24권 제3호, 57-77.
- 김정훈(2008), 시계열 모형을 이용한 부산 북항의 물동량 예측, 한국항만경제학회, 제24권 제2호, 1-17.
- 김종길(2011), SD기법과 ARIMA를 적용한 컨테이너물동량 예측에 관한연구, 인천대학교 박사학위논문
- 김현석, 장명희(2015), VAR모형과 Bayesian VAR모형의 물동량 예측력 비교 연구, 해운물류연구, 제86권(-), 449-466.
- 민경창 · 전영인 · 하현구(2013), 계절성 ARIMA 모형을 이용한 항공화물 수요 예측: 인천국제공항발 유럽 항공노선을 중심으로, 대한교통학회지, 제31권 제3호, 3-18
- 민경창 · 하현구(2014), SARIMA모형을 이용한 우리나라 항만컨테이너물동량 예측, 대한교통학회, 제32권 제6호, 600-614.
- 방지영(2015), 음력효과를 반영한 물류 서비스 산업의 택배 물동량 수요예측, 한양대학교 석사학위논문
- 신창훈 · 정수현(2011), 항만물동량 예측력 제고를 위한 ARIMA 및 인공신경망모형들의 비교 연구, 한국항해항만학회지, 제35권 제1호, 83-91.
- 신창훈 · 강정식 · 박수남 · 이지훈(2008), 하이브리드 ARIMA-신경망 모델을 통한 컨테이너물동량 예측에 관한 연구, 한국항해항만학회지, 제32권 제1호, 81-88.
- 안경모 · 이광우(2015), ARIMA Intervention Model을 이용한 한국인 관광객의 태국여행수요예측에 관한 연구, 한국호텔외식관광경영학회, 제14권 제4호, 273-288
- 전기정 · 정태원 · 한종길(2013), 인천항 벌크화물 유치를 위한 개선방안에 관한 연구, 해운물류연구, 제80권(단일호), 1035-1054.
- 전찬영 · 송주미(2007), 인공신경망모형의 항만물동량 예측 적용에 관한 연구, 해운물류연구, 제53권(-), 65-82.
- 정동빈(2009), SPSS(PASW) 시계열수요예측 I.
- 정성환 · 강경우(2015), 개량경제 모형간 국내 총화물물동량 예측 정확도 비교 연구, 대한교통학회, 제33권 제1호, 61-69.
- 조찬혁(2008), 동해항 시멘트 물동량의 추정에 관한 연구, 물류학회지, 제18권 제1호, 33-53.
- 한광중(2015), SPSS 활용 미래 예측과 시계열 분석 : 미래 예측 및 시계열 자료 분석에 필요한 기법들을 설명. 백산출판사.
- 한국해양수산개발원(2014), 2015년 품목별 항만물동량 예측보고서
- 해양수산부(2016) 제3차 전국 항만기본계획 수정계획 (2016-2020).
- Lewis, C. D.(1982). *International and business forecasting methods*. Butter-Worths, London.
- Meindl,P. & Chopra,S.(2001). *Supply Chain Management : Strategy, Planning And Operation*, 5/e. Pearson Education India.
- Van Dorsser, J.C.M., Wolters, M., & vanWee,B.(2012). *A Very Long Term Forecast of the Port Throughput in the Le Havre - Hamburg Range up to 2100*. EJTIR, 12(1), 88-110.
- Xiao,J., Xiao,Y., Fu,J., & Lai,K.K.(2014). *A transfer forecasting model for container throughput*



- guided by discrete PSO. Journal of Systems Science and Complexity, 27(1),181-192.*
- Xie, G., Wang, S., Zhao, Y., & Lai, K. K. (2013). Hybrid approaches based on LSSVR model for container throughput forecasting : a comparative study. *Applied Soft Computing, 13(5), 2232-2241.*
- Woo, S. H., Pettit, S., Beresford, A., & Kwak, D. W. (2012). Seaport research: A decadal analysis of trends and themes since the 1980s. *Transport Reviews, 32(3), 351-377.*

## 유류화물 항만물동량 예측모형 개발 연구

김정은 · 오진호 · 우수한

### 국문요약

우리나라의 경제발전은 무역을 주축으로 하고 있어 항만을 통한 물류가 필수적이다. 항만의 운영과 개발을 위해 막대한 자본과 시간이 투자되고 있으며 항만은 국가 경제 전반에 영향을 미치고 있다. 따라서 사회·경제적 손실을 방지하기 위해선 적정수준의 개발계획이 중요하다. 항만시설 계획은 항만 물동량 예측을 기반으로 수립되므로, 정확한 물동량 예측이 선행되어야 한다. 더불어 항만에서는 품목별로 취급 방식이 다르므로 품목별 예측이 이루어져야 구체적인 시설계획이 가능하다. 따라서 컨테이너 화물이나 항만 전체 물동량에 대해 주로 예측했던 선행 연구들과는 달리 본 논문에서는 전체 물동량에서 큰 비중을 차지하고 있는 유류화물을 분석 대상으로 설정하였다. 단기, 중장기의 주기적 특성과 추세를 갖고 있는 유류화물 물동량을 효율적으로 예측하고자 새로운 예측모형인 TSMR을 개발하였다. TSMR모형의 검증에 위해 기존의 시계열 모형들과 비교분석을 진행하였으며 ARIMA모형의 경우 물동량 데이터가 안정화되지 않아 유효한 결과를 산출할 수 없었다. 윈터스 가법, 단순계절모형과 비교하였을 때 단기적인 예측에는 다소 취약하였으나, TSMR모형의 전반적인 적합도와 예측력은 우수한 것으로 나타났다. 또한 철강, 유연탄, 기계류의 물동량 분석결과 TSMR모형의 일반화 가능성도 충분한 것으로 나타났다.

주제어: 유류화물, 항만물동량, 시계열 예측, 물동량 예측