

계절 ARIMA 모형을 이용한 여객수송수요 예측: 중앙선을 중심으로

Forecasting Passenger Transport Demand Using Seasonal ARIMA Model - Focused on Joongang Line

김범승*

Beom-Seung Kim*

Abstract This study suggested the ARIMA model taking into consideration the seasonal characteristic factor as a method for efficiently forecasting passenger transport demand of the Joongang Line. The forecasting model was built including the demand for the central inland region tourist train (O-train, V-train), which was opened to traffic in April-, 2013 and run in order to reflect the recent demand for the tourism industry. By using the monthly time series data (103) from January-, 2005 to July-, 2013, the optimum model was selected. The forecasting results of passenger transport demand of the Joongang Line showed continuous increase. The developed model forecasts the short-term demand of the Joongang Line.

Keywords : Joongang Line, Passenger Transport Demand, Seasonal ARIMA Model, O-train, V-train

초 록 본 연구는 중앙선의 여객수송수요를 효율적으로 예측하기 위한 방법으로 계절성 요인을 고려한 ARIMA 모형을 제안하였다. 특히, 최근의 관광수요를 반영하기 위하여 2013년 4월 개통되어 운행되고 있는 중부내륙권 관광전용열차(O-train, V-train)의 수요를 포함하여 예측모형을 구축하였다. 이를 위하여 2005년 1월부터 2013년 7월까지의 월별 시계열 데이터(103개)를 사용하여 최적의 모형을 선정하였으며 예측결과 중앙선의 여객수송수요는 지속적으로 증가할 것으로 나타났다. 구축된 모형은 중앙선의 단기수요를 예측하는데 활용이 가능하다.

주요어 : 중앙선, 여객수송수요, 계절 ARIMA 모형, O-train, V-train

1. 서 론

적정한 수준의 철도 장래 수요예측은 경영계획의 수립, 투자 의사 결정 등에 있어 중요한 역할을 한다. 따라서 잘못된 수요예측에 따른 과잉투자 또는 기회비용의 손실 등은 경영상의 문제를 발생시킨다. 기존의 철도 수요예측 연구에서는 중·장기 예측을 위한 전통적인 4단계 모형, 단기 예측을 위한 시계열 분석 및 직접수요모형을 주로 사용하였다[1-3]. 이 중 단기 예측을 위한 ARIMA 모형은 적은 자료만으로 빠른 시간 내에 정확도가 높은 장래 수요예측이 가능하다[4,5]. ARIMA 모형의 적용에 관한 연구를 살펴보면 Kim과 Hwang [4]은 ARIMA 모형을 이용하여 제주공항 여객 수요를 예측하였고, Kwon과 Lee[5]는 ARIMA 모형을 이용하여 범용적으로 쓸 수 있는 지역간 철도수송수요 예측모형을 구축하였다. 또한 Lee *et al.*[6]은 ARIMA 모형을 이용하여 전력수요를 효율적으로 예측하여 활용하고자 하였고, Hur *et al.*[7]은 다변량 시계열 ARIMA 모형을 이용하여 항공 수요를 예측하였다. 이처럼 ARIMA 모형은 여러 분야에서 다양하게

응용되고 있다. 무엇보다도 각 분야의 특성에 맞는 최적 모형의 개발이 필요하다. 본 연구에서는 중앙선을 대상으로 최근의 관광수요(O-train, V-train)를 포함한 여객수송수요를 효율적으로 예측하기 위한 방법으로 계절 ARIMA 모형을 제안하였다. 본 연구는 2장에서 연구자료 및 ARIMA 모형을 설명하고, 3장에서 모형의 구축 및 검증에 관하여 논하고, 4장에서는 결론 및 향후 연구계획을 제시한다.

2. 연구자료 및 계절 ARIMA 모형

2.1 연구대상 및 연구자료

본 연구를 위한 시간 및 공간적인 범위는 중앙선이다. 중앙선은 청량리~경주(373.8km)까지의 노선으로 우리나라 5대 간선중의 하나이지만 청량리역과 원주역을 제외하면 인구 30만 명 이상의 배후도시가 없는 대표적인 지방 중소도시 연계노선이다[8]. 이는 근본적으로 철도를 이용할 수 있는 인구규모가 작으며 인프라(선로 등)의 개선(복선전철화 및 고속화 등)을 위한 투자 의사 결정시 새로운 수익상품(관광상품 등)의 개발과 함께 적정수준의 장래 수요에 대한 효율적인 예측이 선행되어야 한다. 본 연구에서는 최근의 관광수요가 반영된 자료를 기반으로 수요예측 모형을 구축하고자 하였다. 본 연구를 위하여 구축한 자료는 2005년 1월부터

*Corresponding author.

Tel.: +82-2-3149-3421, E-mail : bluedav@korail.com

©The Korean Society for Railway 2014

<http://dx.doi.org/10.7782/JKSR.2014.17.4.307>

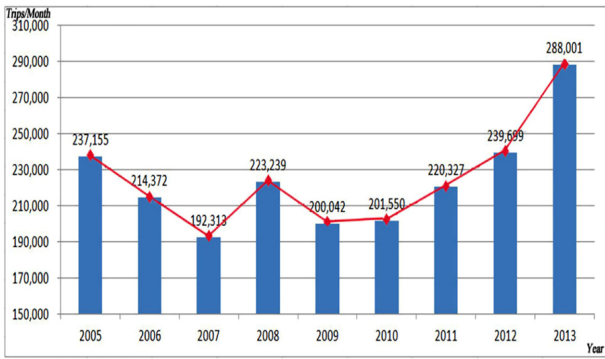


Fig. 1 The passenger transportation number of people by monthly average on an annual basis

2013년 7월(103개월)까지의 주운행선이 중앙선이고 정기권 실적을 포함한 월별 여객열차의 수송인원이다. 여기서 여객 수송인원은 열차종별이 새마을과 무궁화인 열차로 수송된 수 요이다. Fig1은 연도별 월평균 여객수송실적이다. 2005년 청량리 ~ 용문(현 용산 ~ 용문 전동차 운행) 복선전철 개통 운 행 이후 2007년까지 감소, 2008년 소폭상승 하였다가 2009 년 이후 지속적으로 증가하는 추세를 보이고 있으며 특히 2013년 큰 폭의 증가를 보이고 있다. 이는 최근의 관광수요 (O-train, V-train)¹가 반영된 것으로 추정된다. 본 연구에서 는 모형의 식별 및 추정을 위하여 2005년 1월 ~ 2013년 7 월까지의 자료(103개월), 모형의 검증을 위하여 2013년 8월 ~ 2013년 11월까지의 자료(4개월)를 활용하여 계절 ARIMA 모형을 이용한 수요예측모형을 구축하였다.

2.2 계절 ARIMA 모형

시계열이 일정한 계절주기를 가지고 변할 때 사용하는 계 절 ARIMA 모형은 시계열의 구성성분들이 확률적이거나 다 른 성분들과 상관이 있는 경우에 사용할 수 있다. 시계열 Z_t 가 계절주기가 s 인 $ARIMA(p,d,q)(P,D,Q)_s$ 과정을 따를 때 식 (1) 형태의 모형으로 표현한다[2,9]. 여기서, B 는 후행연 산자(Backshift Operator), $\phi(B)$ 와 $\Phi(B^s)$ 는 각각 비계절적 및 계절적 AR연산자, $\theta(B)$ 와 $\Theta(B^s)$ 는 각각 비계절적 및 계절 적 MA연산자, d 와 D 는 각각 비계절적 및 계절적 차분 차 수, p 와 P 는 각각 비계절적 및 계절적 AR차수, q 와 Q 는 각 각 비계절적 및 계절적 MA차수, δ 는 평균에 관련된 상수이 다. 본 연구에서는 계절 ARIMA 모형을 이용하여 수요예측 모형을 구축하였다.

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \delta + \theta_q(B)\Theta_Q(B^s)\varepsilon_t, \varepsilon_t \sim iid N(0, \sigma_\varepsilon^2) \quad (1)$$

¹중부내륙권 관광전용열차: O-train의 O는 'One'의 약자, 순환을 상징 하는 모양, 중부내륙 3도(강원, 충북, 경북도)를 하나(One)로 잇는 순환 열차 의미, V-train의 V는 'Valley(협곡)'의 약자, 동시에 협곡의 모양을 의미, 협곡전용열차

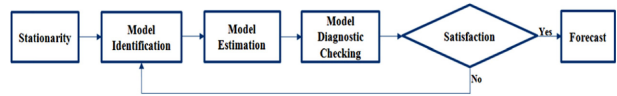


Fig. 2 The building procedure of Seasonal ARIMA Model

계절 ARIMA 모형의 전반적인 구축절차를 살펴보면 Fig. 2와 같으며 내용은 다음과 같다[9]. 첫 번째로 시계열 자료에 대하여 일정한 주기로 나누었을 때, 각 주기에 해당하는 평균과 분산이 일정한 정상성(Stationarity)을 만족시켜 주어야 한다. 만일 시계열 자료의 평균과 분산이 비정상적일 경우, 각각 변수변환 및 차분을 취해 정상성을 만족시킨다. 두 번째로 모형의 식별단계에서는 시계열 자료 내 관측값들 사이에 존재하는 상관관계를 측정하기 위하여 자기상관함수(Autocorrelation function, ACF)와 편자기상관함수(Partial Autocorrelation function, PACF)를 이용하여 $ARIMA(p,d,q)$ 모형을 구성하는 자기회귀(Autoregressive, AR) 요소인 p 와 이동평균(Moving Average, MA)요소인 q 를 임시적으로 결정한다. 세 번째로 모형의 추정단계에서는 식별단계에서 선택한 모형의 계수들을 정확히 추정한다. 네 번째로 모형의 진단단계에서 추정된 모형이 통계적으로 적절하지 여부를 결정하며 최적의 모형을 찾을 때까지 모형의 식별·추정·진 단단계를 반복하여 실행한다. 마지막으로 최종모형이 구축 되면 이를 이용하여 장래 수요를 예측한다.

3. 모형의 구축 및 검증

3.1 정상성 및 계절성 존재여부 점검

Fig. 3의 중앙선의 여객수송인원의 시계열 그래프를 살펴 보면 2010년 3월 이후 급격히 감소했다가 다시 증가하는 추세를 보이며 평균 참조선을 기준으로 일정한 패턴을 보이지 않고 있기 때문에 평균이 정상적이지 않음을 알 수 있다. 따라서 비계절적인 차분을 실시하여 평균을 정상적으로 만들 필요가 있다.

Fig. 4의 비계절적인 1차 차분 후 시계열 도표를 살펴보면 Fig. 3과는 달리 더 이상 감소하거나 증가하는 추세를 보이지 않으며 시계열 자료의 평균이 정상적으로 되었음을 시각적으로 확인할 수 있다.

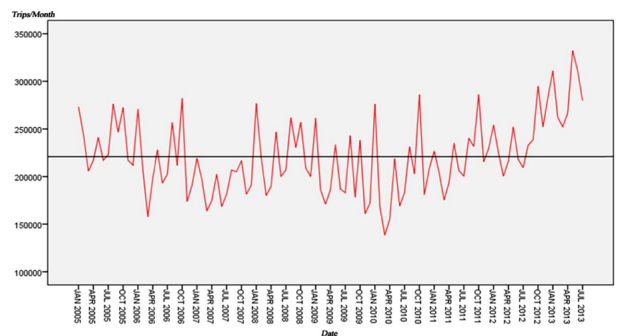


Fig. 3 The time series graph for passenger transportation number of people

또한 Fig. 5에서 1차 비계절적 차분된 자기상관함수를 살펴보면 더 이상 증가하는 추세가 나타나지 않음을 알 수 있다. 그리고 1차 비계절적 차분된 자기상관함수 중에서 몇 개의 스파이크가 신뢰한계선 밖으로 튀어나온 것을 볼 수 있다. 특히, 시차 12와 시차 24의 자기상관계수의 스파이크가 신뢰한계선을 넘어 돌출되어 있고 시차 12와 시차 24의 자기상관계수가 서서히 감소하므로 계절적 차분($D=1$)이 필요함을 알 수 있다.

Fig. 6에서 여객수송인원의 비계절적 및 계절적 1차 차분 그래프를 살펴보면, 평균 0을 중심으로 일정하고, 분산도 시간의 흐름에 따라 일정한 모습을 보인다. 이는 차분후 시계열 자료의 평균과 분산이 정상적인 분포를 나타내고 있음을 알 수 있다.

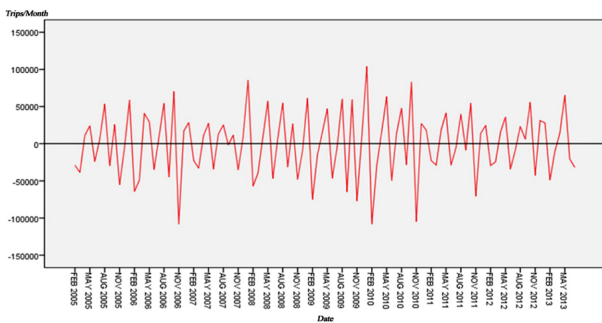


Fig. 4 The primary difference non-seasonal time series graph

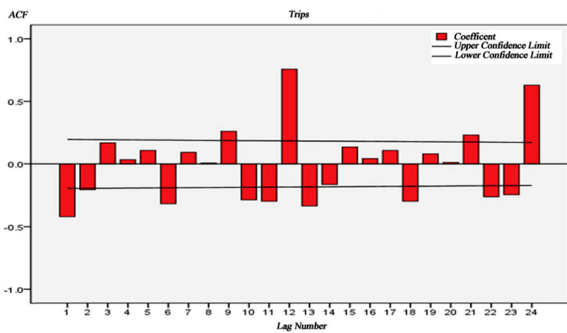


Fig. 5 The primary difference non-seasonal autocorrelation function

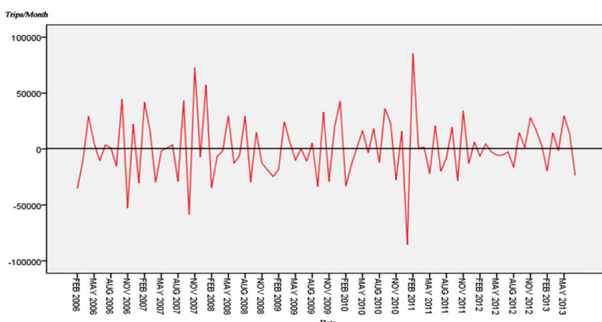


Fig. 6 The primary difference non-seasonal and seasonal time-series graph

3.2 모형의 식별

자기상관함수는 자기회귀(AR)의 경우 이론적으로 지수함수 또는 사인곡선 형태로 서서히 0으로 감소하는 형태이며, 이동평균(MA)의 경우에는 이동평균 차수에 해당하는 시차에서 두드러진 스파이크가 나타나고, 이 시차 이후에 해당하는 자기상관계수가 모두 0으로 절단되며, 혼합된 경우에는 0을 향해 서서히 감소하는 형태를 보인다[4,9]. Fig. 7에서 자기상관함수의 시차 1에서 뚜렷한 스파이크를 발견할 수 있으며 그 이후 단기시차에서 서서히 0으로 절단되는 형태이므로, 비계절적 부분을 $MA(1)_{12}$ 모형으로 고려할 수 있다. 마찬가지로, 자기상관함수의 시차 12에서 뚜렷한 스파이크를 발견할 수 있고 시차 24에서 0으로 절단하는 형태이므로 $MA(1)_{12}$ 모형을 고려할 수 있다. 또한 비계절적 및 계절적 1차 자기상관함수는 계절시차 12에서 유의적인 값을 갖고 있다. 따라서 본 연구에서는 Model1: $ARIMA(0,1,1)(0,1,1)_{12}$, Model2: $ARIMA(0,1,1)(0,1,0)_{12}$, Model3: $ARIMA(0,1,0)(1,1,1)_{12}$, Model4: $ARIMA(0,1,0)(1,1,0)_{12}$, Model5: $ARIMA(0,1,0)(0,1,1)_{12}$ 의 5가지 모형을 고려하였다. 이때 통계적으로 유의하지 않은 상수항은 모형에서 제외한다. 본 연구에서는 적절한 모형의 차수결정을 위해 자기상관함수와 편자기상관함수를 이용하고, 그 후에 ARIMA 모형을 객관적으

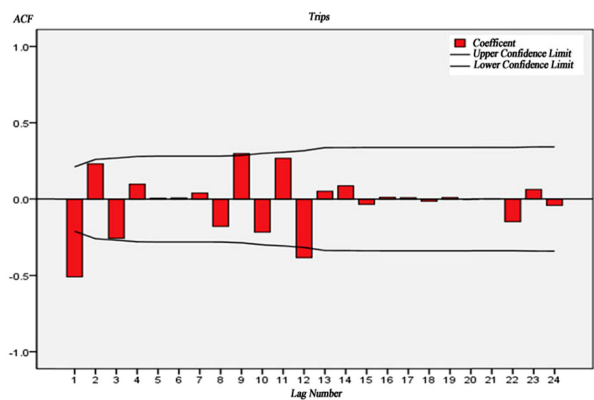


Fig. 7 The primary difference non-seasonal and seasonal autocorrelation function

Table 1. Model identification

ARIMA Model	Stationary R square	Normalized BIC
Model1: $(0,1,1)(0,1,1)_{12}$	0.598	19.548
Model2: $(0,1,1)(0,1,0)_{12}$	0.269	20.220
Model3: $(0,1,0)(1,1,1)_{12}$	0.252	20.332
Model4: $(0,1,0)(1,1,0)_{12}$	0.126	20.398
Model5: $(0,1,0)(0,1,1)_{12}$	0.267	20.223

로 식별하기 위하여 정규화된 BIC(Normalized Bayesian Information Criterion) 방법을 사용하였다[9]. 다음 Table 1을 살펴보면 Model1의 BIC 값이 19.548으로 가장 작으므로 Model1을 우선적으로 최적모형으로 고려하되 다른 Model들의 경우도 함께 살펴보기로 한다.

3.3 모형의 추정

Table 2에서 볼 수 있듯이 모형의 추정단계에서 우선적으로 고려한 Model1을 제외한 Model2 ~ Model5의 상수항의 경우 95% 신뢰수준에서 통계적으로 유의하지 않았고 Model3을 제외한 다른 Model들의 추정된 계수들은 모두 99% 신뢰수준에서 통계적으로 유의한 것으로 나타났다.

3.4 모형의 진단

모형의 진단은 추정된 모형이 관측된 시계열을 통계적으로 잘 적합시키고 있는지를 확인하는 것이다. 모형을 진단하는 방법으로 잔차분석을 주로 이용하는데 통계적 적절성

을 점검하는데 있어 가장 중요한 검증은 백색잡음(White Noise)의 독립성 가정 여부이다. 본 연구에서는 주로 사용하는 모형 진단 방법 중 Ljung-Box 통계량을 이용하였다 [4,9]. Table 3의 Ljung-Box 통계량을 살펴보면 Model1, Model2의 경우 유의확률이 각각 $0.552 > 0.05$, $0.309 > 0.05$ 이므로 “잔차는 자기상관이 없다”는 귀무가설이 기각되지 않기 때문에 잔차에는 자기상관이 없다. 따라서 “백색잡음항은 독립이다”라는 귀무가설을 채택하게 된다. 전체적으로 종합하여 판단하면 모형의 식별단계의 BIC 값, 모형의 추정단계의 t-통계량, p-Value 및 모형의 진단단계의 Ljung-Box 통계량을 고려해 볼 때 Model1을 최적모형으로 선정할 수 있다. 또한 계절성을 포함한 단순 지수평활법, Winters(승법) 방법의 시계열 모형과 비교를 위하여 Fig. 8, Table 4와 같이 예측력과 모형 통계량을 고려하였다. 각 비교 모형들의 예측력에 대한 우의를 비교하기 위하여 최근 대략 10% 자료인 10개(2012년 10월부터 2013년 7월까지)의 관측값을 제거한 후 각 모형에 근거하여 제거된 10개 시점의 예측값을 산출하였다. Fig. 8을 살펴보면 관측값에 가까운 개수를 고

Table 2. Model estimation

		Estimate	SE	t	P(=Sig.)	
Model1 (0,1,1)(0,1,1) ₁₂	Constant	793.938	322.896	2.459	0.016	
	Difference	1	-	-	-	
	MA	Lag 1	0.614	0.086	7.125	0.000
	Seasonal difference	1	-	-	-	
	MA,Season	Lag 1	0.795	0.186	4.284	0.000
			Estimate	SE	t	P(=Sig.)
Model2 (0,1,1)(0,1,0) ₁₂	Constant	1073.401	1048.084	1.024	0.309	
	Difference	1	-	-	-	
	MA	Lag 1	0.580	0.088	6.562	0.000
	Seasonal difference	1	-	-	-	
		Estimate	SE	t	P(=Sig.)	
Model3 (0,1,0)(1,1,1) ₁₂	Constant	0.003	0.004	0.617	0.539	
	Difference	1	-	-	-	
	AR,Season	Lag 1	0.198	0.195	1.013	0.314
	Seasonal difference	1	-	-	-	
	MA,Season	Lag 1	0.988	4.244	0.233	0.816
		Estimate	SE	t	P(=Sig.)	
Model4 (0,1,0)(1,1,0) ₁₂	Constant	0.003	0.009	0.373	0.710	
	Difference	1	-	-	-	
	AR,Season	Lag 1	-0.368	0.099	-3.706	0.000
	Seasonal difference	1	-	-	-	
		Estimate	SE	t	P(=Sig.)	
Model5 (0,1,0)(0,1,1) ₁₂	Constant	615.694	892.893	0.690	0.492	
	Difference	1	-	-	-	
	Seasonal difference	1	-	-	-	
	MA,Season	Lag 1	0.834	0.202	4.122	0.000

Table 3. Model diagnostic checking

Ljung-box Q statistics			
ARIMA Model	Statistics	df	P(=Sig.)
Model1(0,1,1)(0,1,1) ₁₂	14.632	16	0.552
Model2(0,1,1)(0,1,0) ₁₂	34.255	17	0.309
Model3(0,1,0)(1,1,1) ₁₂	48.243	16	0.000
Model4(0,1,0)(1,1,0) ₁₂	69.621	17	0.000
Model5(0,1,0)(0,1,1) ₁₂	54.780	17	0.000

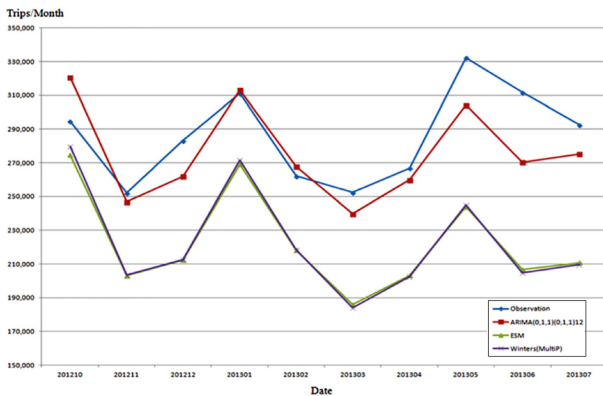


Fig. 8 Ability of forecast

려하면 구축한 모델이 우위에 있음을 알 수 있다. 또한 Table 4와 같이 모형적합 통계량의 관점과 Ljung-Box 통계량을 비교해 볼 때 구축 모형이 우위에 있다고 할 수 있다. 그리고 구축 모형의 실제 예측력 검증에 위하여 취득 가능한 자료 4개(2013년 8월부터 2013년 11월까지)의 실측값과 예측값을 비교하여 검증한 결과 RMSE 값이 16146.079로 다른 모형에 비하여 우위에 있음을 확인 할 수 있으며 Fig. 9와 같이 실측값에 근접하고 있음을 알 수 있다. 따라서 구축 모형이 다른 시계열 모형보다 우위에 있으며 중앙선 여객수송수요를 예측하는데 적합한 모형이라고 할 수 있다.

3.5 장래수요의 예측

최적 모형으로 선정된 Model1: ARIMA(0,1,1)(0,1,1)₁₂를

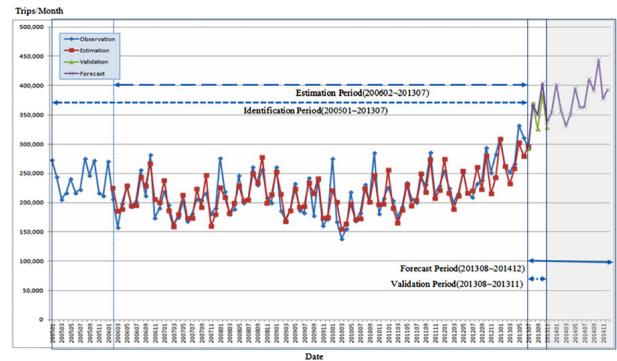


Fig 9. The future demand forecast for passengers

이용하여 2013년 8월부터 2014년 12월까지의 중앙선의 여객수송수요를 예측하였다. 그 결과는 Fig 9와 같으며 최소치 350,268명 최고치 444,006명으로 연평균 381,445명, 2013년 대비 연평균 91,624명(1.32%) 증가하는 것으로 예측되었고 2014년 12월까지 꾸준히 증가하는 추세를 보이고 있다.

4. 결론 및 향후계획

중앙선은 청량리역과 원주역을 제외하면 인구 30만 명 이상의 배후도시가 없는 대표적인 지방 중소도시 연계노선으로서 근본적으로 철도를 이용할 수 있는 인구규모가 작다. 따라서 인프라(선로 등)의 개선(복선전철화 및 고속화 등)을 위한 투자외에 결정시 새로운 수익상품(관광상품 등)의 개발과 함께 적정수준의 장래수요에 대한 효율적인 예측이 선행되어야 한다. 본 연구는 중앙선의 여객수송수요를 효율적으로 예측하기 위한 방법으로 계절성 요인을 고려한 ARIMA 모형을 제안하였다. 특히, 최근의 관광수요를 반영하기 위하여 2013년 4월 개통되어 운행되고 있는 중부내륙권 관광전용열차(O-train, V-train)의 수요를 포함하여 예측모형을 구축하였다. 이를 위하여 2005년 1월부터 2013년 7월까지의 월별 시계열 데이터(103개)를 사용하여 최적의 모형을 선정하였으며 예측결과 중앙선의 여객수송수요는 지속적으로 증가할 것으로 나타났다. 향후 각 노선별 특성을 반영한 ARIMA 모형의 구축에 관한 연구를 진행 할 계획이다.

Table 4. Comparison of model statistics

Model Statistics		ARIMA(0,1,1)(0,1,1) ₁₂	ESM	Winters(MultiP)
Model Fit	Stationary R Square	0.598	0.483	0.506
	R Square	0.817	0.761	0.781
	RMSE	16801.332	19790.415	18495.526
	MAPE	5.965	6.829	6.259
	MaxAPE	19.744	22.688	19.575
	MAE	13200.406	14714.252	13919.936
	MaxAE	48473.135	56189.600	51591.820
	Normalized BIC	19.548	19.936	19.786
Ljung-Box Q	P(=Sig.)	0.552	0.347	0.450

후 기

본 연구는 2013년도 한국철도학회 추계학술대회 발표 논문을 수정·보완하여 작성하였으며 2013년도 코레일 정보기술단 ICT 연구회 활동의 일환으로 진행되었습니다.

References

- [1] H.S Kim, D.H Yun, S.D Lee (2012) Travel behavior analysis for short-term KTX passenger demand forecasting, *Journal of the Korean Statistical Society*, 19(1), pp. 183-192.
- [2] K.H Kim, H.S Kim (2011) KTX passenger demand forecast with intervention ARIMA model, *Journal of the Korean Society for Railway*, 14(5), pp. 470-476.
- [3] H.J Kim, C.M Jung (2010) Development of a direct demand estimation model for forecasting of railroad traffic demand, *Proceeding of the Conference of the Korean Society for Railway*, 14(5), pp. 2166-2178.
- [4] K.B Kim, K.S Hwang (2012) A study on the demand forecasting and efficient operation of Jeju national airport using seasonal ARIMA model, *Journal of the Korea Academia-Industrial cooperation Society*, 13(8), pp. 3381-3388.
- [5] J.M Lee, Y.J Kwon (2011) A study on dynamic change of transportation demand using seasonal ARIMA model, *Journal of Korean Society of Transportation*, 29(5), pp. 139-155.
- [6] J.S Lee, H.G. Shon, S. Kim (2013) Daily peak load forecasting for electricity demand by time series models, *The Korean Journal of Applied Statistics*, 26(2), pp. 349-360.
- [7] N.K Hur, J.Y Jung, S. Kim (2009) A study on air demand forecasting using multivariate time series models, *Journal of Korean Statistical Society*, 22(5), pp. 1007-1017.
- [8] C.H Choi (2013) Suggestions for improvement of connectivity between railway stations in small and medium-sized cities based on a survey focused on the joongang line, *Journal of the Korean Society for Railway*, 16(1), pp. 52-58.
- [9] D.B Jung (2009) *Time series demand forecasting I(SPSS/PASW)*, Hannarae Publishing Co., Naengcheon B/D, 88 Kyonggidae-ro, Seodaemun-gu, Seoul, 120-050, Korea, pp. 23-112.

접수일(2014년 2월 4일), 수정일(2014년 6월 16일),
게재확정일(2014년 8월 5일)

Beom-Seung Kim: bluedav@korail.com

Korea Railroad Information Technology Office, Railroad B/D 416,
432 Chengpa-ro, Jung-gu, Seoul, 100-162, Korea