

온도를 변수로 갖는 단기부하예측에서의 TAR(Threshold Autoregressive) 모델 도입

이경훈 · 이윤호 · 김진오  
한양대학교 전기공학과

Introduction of TAR(Threshold Autoregressive) Model for Short-Term Load Forecasting including Temperature Variable

Kyung-Hun Lee · Yunho Lee · Jin-O Kim  
Dept. of EE, Hanyang University

**Abstract** - This paper proposes the introduction of TAR(Threshold Autoregressive) model for short-term load forecasting including temperature variable. TAR model is a piecewise linear autoregressive model. In the scatter diagram of daily peak load versus daily maximum or minimum temperature, we can find out that the load-temperature relationship has a negative slope in lower regime and a positive slope in upper regime due to the heating and cooling load, respectively. In this paper, daily peak load was forecasted by applying TAR model using this load-temperature characteristic in these regimes. The results are compared with those of linear and quadratic regression models.

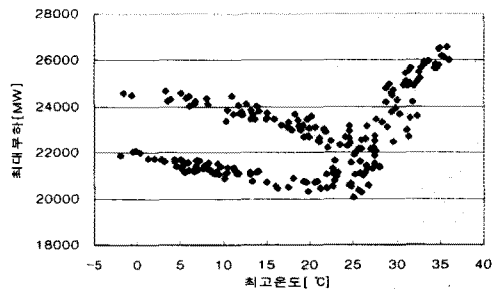
1. 서 론

전력부하 예측의 정확도 향상은 발전 비용의 감소와 전력 공급의 신뢰도 향상의 결과로 나타나기 때문에 장기 및 단기간의 전력 부하의 예측은 전력 계통의 경제적인 운용의 측면에서 매우 중요하다. 특히 그 결과는 전력 계통의 제어 및 전력조류계산의 입력 자료로 사용되기 때문에 지금까지 다양한 방식에 의한 예측 기법이 제안되었다. 이러한 기존의 방법으로는 크게 시계열 방법(2)과 인공지능 기법의 신경회로망(4) 등이 있다. 시계열 방법, 특히 자기회귀(AR : Autoregression), 통합혼합모형(ARIMA : Autoregressive Integrated Moving Average) 등은 예측 모형을 설정하는데 있어서 자기 변수만을 고려하므로 예측 모형으로서 단순하고 선형모델이다. 또한 임계값을 기준으로 좌우 대칭이 아닌 온도와 부하의 관계를 2차 함수 모델에 적용하면 좋은 결과를 얻을 수 없기 때문에 본 논문은 온도를 변수로 갖는 단기부하예측을 위하여 비선형 문제를 다루는 TAR(Threshold Autoregressive) 모델을 도입하였다. 겨울철의 난방부하, 여름철의 냉방부하의 영향으로 온도와 부하의 관계는 임계값을 기준으로 부(負)의 상관관계와 정(正)의 상관관계를 나타낸다. 이러한 부하와 온도의 상관관계를 이용하여 일별 최고 및 최저온도와 최대부하의 관계를 임계값을 기준으로 구분하는 TAR 모델을 도입하여 회귀분석 하였으며 이를 선형 모델, 2차 함수 모델과 비교하여 TAR 모델 도입의 필요성을 검증하였다. 또한 실제 예측 단계에서는 연간부하성장률을 고려한 단순한 데이터 보정만으로도 예측력을 높일 수 있었다.

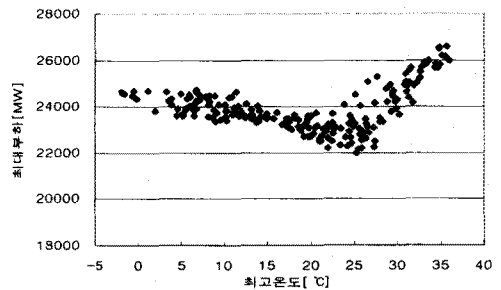
2. 본 론

전력부하의 변동요인 중 기상의 변화는 매우 불규칙적이며 비선형의 특성을 보인다. 특히 온도는 습도나 일조 시간 등의 다른 기상요인에 비해 전력부하의 변동에 더

많은 영향을 미친다. 따라서 부하예측에 있어서 더 좋은 예측 결과를 기대하기 위해서는 부하와 온도의 관계를 규명하는 것이 급선무이다. 그림 1은 출발시점이 서로 다른 1년 동안의 일별 최고온도와 일별 최대부하의 관계이다. 각 데이터는 특수성을 띠는 주말과 특수일을 제외하였으며 그 수는 (a)의 경우 221개, (b)의 경우 210개이다. 그림에서 임계값을 기준으로 그 이상의 영역은 여름철 부하로 그 특성이 온도에 대해 정의 상관관계로 나타나는 반면, 그 이하인 영역은 각각 다르다. 즉, 출발시점을 1월부터 잡은 그림 1의 (a)의 경우는 뚜렷하게 부의 두 가지 선형관계가 나타나고, 출발시점을 7월부터 잡은 (b)의 경우는 하나의 부의 선형관계가 나타난다. 따라서 본 논문은 모델을 단순화하고 예측력을 높이기 위해 그림 1의 (b)와 같이 예측 모델 기간인 1년을 7월부터 다음 해 6월까지로 설정하였다.



(a) 1994년1월1일~12월31일



(b) 1994년7월1일~1995년6월31일

그림 1. 최대부하와 최고온도의 관계

2.1 TAR(Threshold Autoregressive) 모델

TAR(Threshold Autoregressive) 모델은 임계값을 기준으로 데이터를 구분하여 각각을 다른 회귀 모형으로 취급한다. 일반화된 다변수 TAR 모델의 형태는 다음 식 (1)과 같다.

### 3. 사례 연구

$$Y_t = \begin{cases} a_0 + (a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_k Y_{t-k}) \\ \quad + (b_1 X_{t-1} + b_2 X_{t-2} + \dots + b_l X_{t-l}) \\ \quad , \text{ if } X_{t-d} < A \\ c_0 + (c_1 Y_{t-1} + c_2 Y_{t-2} + \dots + c_k Y_{t-k}) \\ \quad + (d_1 X_{t-1} + d_2 X_{t-2} + \dots + d_l X_{t-l}) \\ \quad , \text{ if } X_{t-d} \geq A \end{cases} \quad (1)$$

여기서, A : 임계값(threshold value)  
d : 지연 파라미터(delay parameter)  
X<sub>t-d</sub> : 임계기준 변수(threshold variable)

임계값을 기준으로 두 변수간의 비선형 관계를 규명하기 위해서 일반화된 TAR 모델을 적용한다. TAR 모델은 단지 한 개의 변수로 이루어진 시계열을 다루는 반면, 일반화된 TAR 모델은 다변수 시계열을 다루며 상태 변수를 포함한다. TAR 모델의 각 파라미터를 추정할 때 지연 파라미터 d, 임계값 A를 어떻게 결정할 것인가 문제가 되는데 이러한 값들을 계산하기 위해서 일반적으로 TONG[1]의 방법을 따른다. 또한 임계값을 기준으로 그 이하인 영역을 하위영역(Lower Regime)이라 하고, 그 이상인 영역을 상위영역(Upper Regime)이라고 한다. 다음 식은 본 논문에서 제시하는 일별 최대부하를 종속변수, 최고 및 최저 온도를 독립변수로 갖는 TAR 모델이다.

$$L_{peak} = \begin{cases} a_{0,L} + a_{1,L}T & , \text{ if } T < A \\ a_{0,U} + a_{1,U}T & , \text{ if } T \geq A \end{cases} \quad (2)$$

여기서, L<sub>peak</sub> : 일별 최대부하[MW]  
T : 일별 최고 및 최저온도[°F]  
a<sub>0,L</sub>, a<sub>1,L</sub>, a<sub>0,U</sub>, a<sub>1,U</sub> : 하위영역(Lower Regime), 상위영역(Upper Regime)의 회귀계수

온도 T를 화씨온도로 정한 이유는 2차 함수 모델과 비교할 때 온도가 0°C 이하인 경우를 고려하기 위해서이다. 또한 본 논문은 부하와 온도의 상관관계를 밝히는 것에 초점을 두고 부하와 온도에 대해 규칙성을 띠지 않는 주말이나 특수일의 데이터를 제외하였으므로, 과거의 시계열 Y<sub>t-i</sub> 또는 Y<sub>t-k</sub>까지 모델에 반영한다면 변수간 간격이 일정하지 않게 되어 신뢰성 있는 모델을 기대할 수 없다. 따라서 일별 최고 및 최저온도만을 독립변수로 정하였다.

#### 2.2 임계값(Threshold Value) 선정

TAR 모델을 수립하기 위해서는 먼저 오차를 가장 작게 하는 임계값 A를 찾아내야 한다. 먼저 n개의 임의의 임계값을 선택하여 예측치와 실제치의 오차가 가장 작도록 하는 함수를 구하는 기법인 최소자승법(OLS : Ordinary Least Squares)을 이용하여 모든 n에 대하여 하위영역과 상위영역 각각의 계수를 구한 다음, 모든 n에 대하여 RSS1과 RSS2, TRSS와 Pooled SE를 계산한다. 여기서 RSS1과 RSS2는 각각 하위영역과 상위영역의 잔차제곱합(RSS : Residual Sum of Squares), TRSS는 RSS1과 RSS2의 합이며 Pooled SE(Pooled Standard Error)는 다음 식 (3)과 같다 [3].

$$Pooled\ SE = (TRSS/\text{자유도})^{0.5} \quad (3)$$

마지막 단계에서는 최소 Pooled SE값을 갖는 임계값 A를 선정한다.

#### 3.1 각 모델의 결과 비교

1994년 7월부터 1995년 6월까지의 부하와 온도 데이터 중 특수성을 띠는 주말과 특수일의 데이터를 제외한 일별 최대부하 데이터와 등가화된 최고 및 최저온도 데이터의 관계를 TAR 모델을 이용하여 회귀분석 하였다. 관찰치 수는 210개이고, 온도의 등가화는 지역별 수요구성비(서울:0.51, 부산:0.22, 대구:0.13, 대전:0.09, 광주:0.05)를 가중평균하여 산출하였다. 임계값 선정 절차에 따라 독립변수를 일별 최고온도로 했을 경우는 14°C로 나타났다.

이러한 임계값에 따라 식 (2)와 같은 형태의 TAR 모델을 수립하였으며 비교를 위한 선형과 2차 함수 모델의 형태는 다음 식 (4), (5)와 같다.

$$L_{peak} = a_0 + a_1T, \quad \forall T \quad (4)$$

$$L_{peak} = a_0 + a_1T + a_2T^2, \quad \forall T \quad (5)$$

각 모델의 결과인 검정 통계량을 표 1에 정리하였다. 잔차들의 자기상관 정도를 검정하는 통계량인 DW (Durbin-Watson)값에 있어서 TAR 모델은 어떤 경우든 2에 가까운 값을 갖으며, 이는 다른 모델과 비교하여 잔차들의 독립성이 더 우수함을 입증한다.

SE(Standard Error)는 최고온도, 최저온도 모두 TAR 모델의 경우가 가장 작게 나타나고, TAR 모델 중에서는 상위영역, 즉 여름철에 SE가 상대적으로 크게 나타난다.

표 1. 검정 통계량에 의한 비교

		선형	2차 함수	TAR	
				하위	상위
최고 온도	R <sup>2</sup>	0.0318	0.5755	0.6156	0.7524
	DW	0.164	0.494	1.85	1.96
	SE	923.763	613.11	395.353	559.163
	Pooled SE	923.763	613.11	445.029	
최저 온도	R <sup>2</sup>	0.0621	0.6058	0.5609	0.7599
	DW	0.168	0.468	1.74	2.28
	SE	909.194	590.87	416.372	601.427
	Pooled SE	909.194	590.87	488.141	

#### 3.2 각 모델의 예측 결과

부하예측 모델은 연간부하성장률이 반영된 데이터를 통해 수립되었다. 예측 기간은 1995년 7월부터 1996년 6월까지이며 마찬가지로 주말과 특수일을 제외하여 예측일수는 219일이다.

TAR 모델의 효용성을 검증하기 위하여 예측 부하의 실제 부하에 대한 평균 오차는 식 (6)과 같이 일별 평균 오차 절대값의 합을 평균하여 구한 절대 평균 백분율 오차(MAPE : Mean Absolute Percentage Error, [%])로 정의하였다[4].

$$MAPE[\%] = \frac{1}{N} \sum_{i=1}^N \frac{|Actual(i) - Forecast(i)|}{Actual(i)} \times 100 \quad (6)$$

여기서, N : 예측 날짜의 개수

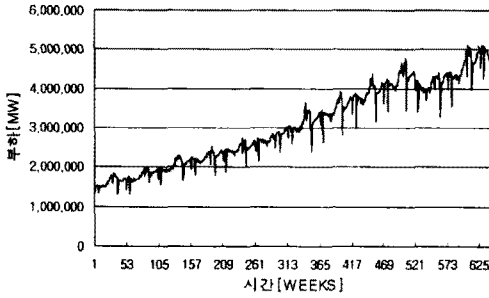


그림 2. 부하의 증가 추세

그림 2는 1988년부터 2000년 5월까지 12년 이상 일주일을 한 단위로 부하의 총합을 나타낸 것으로 IMF 시기를 제외한 매년 대략 10[%] 정도씩 증가해왔다. 이러한 추세는 인구의 증가와 경제 성장률의 상승에 기인한다.

따라서 1년 동안의 부하데이터를 통해 얻은 예측모델을 그대로 이용하여 다음 1년을 예측한다면 좋은 결과를 기대할 수 없기 때문에 연간 부하의 성장률을 고려하여 부하데이터를 보정하기 위하여 다음 식 (7), (8)을 사용하였다.

$$L_y = b_0 + b_1 L_{y-1} \quad (7)$$

$$L_{peak} = b_1 L_{peak} \quad (8)$$

여기서,  $L_y$  : y년의 연간 총부하량(MW)

$L_{peak}$  : 연간부하성장률에 따라 보정된 일별 최대부하(MW)

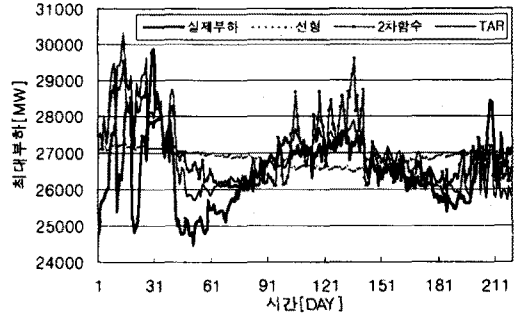
예측하고자 하는 기간이 1995년 7월부터이므로 식 (7)의 y는 1994년까지 설정하였다. 회귀계수  $b_1$ 을 연간부하성장률로 취급하여 식 (8)을 식 (2), (4), (5)에 대입함으로써 연간부하성장률을 고려한 새로운 모델들을 구할 수 있다. 보정된 일별 최대부하를 이용한 예측 모델의 결과를 표 2에 나타내었고, Pooled MAPE는 상, 하위영역 각각의 MAPE를 각각의 예측치 개수에 따라 가중평균한 값이다.

MAPE는 각 모델별로 비교할 때 TAR 모델에서 가장 작게 나타난다. 또한 그림 3은 각 모델별 예측 결과인데, TAR 모델은 다른 모델에 비해 실제부하에 가장 근접하는 결과를 나타내지만 여름철, 즉 주말과 특수일을 제외한 7월부터 9월에 해당하는 1~60여 일은 오차가 크게 나타난다. 이는 여름철 부하가 온도 외에도 다른 기상요인에 의해 크게 변동하기 때문이다.

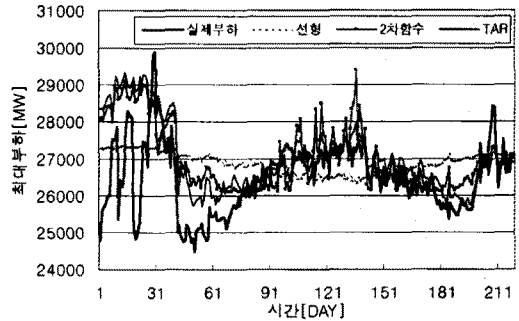
상위영역에서 즉 여름철 부하에 대해 좀더 정확한 예측 결과를 얻기 위해서는 다른 기상요인을 독립변수에 추가하여 새로운 예측 모델을 개발해야 한다고 사료된다.

표 2. MAPE에 의한 비교(%)

		선형	2차 함수	TAR	
				하위	상위
최고 온도	MAPE	3.2595	3.0340	1.7213	5.4881
	Pooled MAPE	3.2595	3.0340	2.7533	
최저 온도	MAPE	3.3158	3.0111	1.4799	4.7829
	Pooled MAPE	3.3158	3.0111	2.7317	



(a) 최고온도-최대부하



(b) 최저온도-최대부하

그림 3. 부하 예측 결과

#### 4. 결 론

본 논문은 임계값을 기준으로 상위영역과 하위영역에서 각각 정(正)의 상관관계와 부(負)의 상관관계를 보이는 부하와 온도의 특성을 이용하여 일별 최대부하와 일별 최고 및 최저온도의 관계를 TAR 모델을 도입하여 회귀분석 하였으며 이를 선형 모델, 2차 함수 모델과 비교하였다. 검정 통계량과 MAPE 모두에서 TAR 모델은 다른 모델에 비해 우수한 결과를 나타내었고 이로써 TAR 모델 도입의 필요성을 검증하였다. 실제예측 단계에서는 연간부하성장률을 고려한 데이터 보정만으로도 예측력을 높일 수 있었다. 향후 예측모델에 온도 외의 다른 기상요인이나 자기변수를 독립변수로 추가하는 방법, 혹은 연간부하성장률을 정확하게 반영할 수 있는 방법 등의 개발을 통해 예측모델을 수립한다면 더 좋은 예측결과를 기대할 수 있을 것으로 사료된다.

#### [참 고 문 헌]

- [1] Tong, H., "Threshold Models in Nonlinear Time Series Analysis", Lecture Notes in Statistics 21, Berlin: Springer Verlag, 1983
- [2] M. T. Hagan and S. M. Behr, "The Time Series Approach to Short-Term Load Forecasting", IEEE Trans. on Power Systems, Vol. 2, No. 3, pp.785~791, 1987
- [3] Yunho Lee, "Nonlinear Dynamic Models: Applications of Chaos Theory and Threshold Autoregressive Model to Agricultural Prices", Thesis, Texas A&M University, 1998
- [4] M. Daneshdoost, M. Lotfalian and G. Bumroongit, "Neural Network with Fuzzy Set-Based Classification for Short-Term Load Forecasting", IEEE Trans. on Power Systems, Vol. 13, No. 4, pp.1386~1391, 1998