

## 제주도 일단위 풍력발전예보 모형개발을 위한 군집분석 및 기상통계모형 실험

김현구\* · 이영섭<sup>1)</sup> · 장문석

한국에너지기술연구원 풍력발전연구센터, <sup>1)</sup>동국대학교 통계학과  
(2010년 7월 31일 접수; 2010년 8월 24일 수정; 2010년 9월 10일 채택)

### Cluster Analysis and Meteor-Statistical Model Test to Develop a Daily Forecasting Model for Jeju-do Wind Power Generation

Hyun-Goo Kim\*, Yung-Seop Lee<sup>1)</sup>, Moon-Seok Jang

*Wind Energy Research Center, Korea Institute of Energy Research, Daejeon 305-343, Korea*

<sup>1)</sup>*Department of Statistics, Dongguk University, Seoul 100-715, Korea*

(Manuscript received 31 July, 2010; revised 24 August, 2010; accepted 10 September, 2010)

#### Abstract

Three meteor-statistical forecasting models - the transfer function model, the time-series autoregressive model and the neural networks model - were tested to develop a daily forecasting model for Jeju-do, where the need and demand for wind power forecasting has increased. All the meteorological observation sites in Jeju-do have been classified into 6 groups using a cluster analysis. Four pairs of observation sites among them, all having strong wind speed correlation within the same meteorological group, were chosen for a model test. In the development of the wind speed forecasting model for Jeju-do, it was confirmed that not only the use a wind dataset at the objective site itself, but the introduction of another wind dataset at the nearest site having a strong wind speed correlation within the same group, would enhance the goodness to fit of the forecasting. A transfer function model and a neural network model were also confirmed to offer reliable predictions, with the similar goodness to fit level.

**Key Words :** Wind power forecasting, Cluster analysis, Meteor-statistical forecasting model, Jeju-do

#### 1. 서론

전력망에 정보기술을 접목함으로써 전력공급자와 소비자가 양방향으로 실시간 정보를 교환하여 에너지 효율을 최적화하며 새로운 부가가치를 창출하는 차세대 전력망을 스마트그리드(smartgrid; 지능형전력망)라고 한다. 특히 전력생산 특성이 매우 다양한 신재생

에너지원을 효과적으로 수용하기 위해서는 스마트그리드가 적용은 필수적이라고 하겠다.

차제에 지식경제부는 제주도를 스마트그리드 실증단지 지정하여 기술개발 및 실용화에 전력을 경주하고 있다. 제주도의 우수한 풍력자원으로 인해 제주도에서 풍력발전이 차지하는 비중은 매우 높다. 즉 2010년 9월말 현재 건설 중인 시설을 포함하여 설비용량은 99 MW이다. 이 때문에 제주도 스마트그리드에서 풍력발전을 수용함에 있어 중요한 기술요건 중 하나는 풍력발전예보이다.

\*Corresponding author : Hyun-Goo Kim, Wind Energy Center, Korea Institute of Energy Research, Daejeon 305-343, Korea  
Phone: +82-42-860-3376  
E-mail: hyungoo@kier.re.kr

풍력발전예보(wind power forecasting)는 예측시간에 따라 단기(시간단위 이하), 중기(일단위), 장기(주단위 이상)로 구분된다. 단기에보는 일반적으로 풍력발전기의 출력인 전력생산량에 대한 통계모형으로 구성되며, 중장기에보는 중규모 기상변화가 고려되어야 하므로 수치기상모델과 같은 물리모형으로 구성된다. 국내에서는 단기에보의 경우 김 등(2006a)이 제주도 월령에 대한 예보모형을 개발하여 KIER-Forecaster로(김 등, 2006b) 특허등록된 바 있으며(김 등, 2008a) 장기에보의 경우 김 등(2010a)이 기상청 MOS 예측값을 풍력예보에 적용한 사례가 있다. 그러나 현재까지 이들 극소수의 논문이 발표되었을 뿐 아직 국내 연구 개발은 초기단계에 있다.

중장기에보의 경우, 수치기상예보와 연계되어야 하기 때문에 대규모 전산 인프라, 수치기상모델 운용을 위한 전문인력, 신뢰도 유지를 위한 지속적인 모니터링과 피드백이 요구된다(김 등, 2010b). 풍력발전예보에서 요구하는 풍속예측이란 풍력단지 내 개별 풍력발전기 허브높이에서의 풍속을 말하며, 중규모 수치기상모델로는 공간해상도 100 m 이하의 매우 조밀한 국소규모 해석에는 상당한 기술적 한계가 존재한다. 반면 통계모형은 중장기의 경우 다소 부정확할지라도 모형구축과 운용이 매우 간편하다는 장점이 있다.

본 연구의 목적은 풍력발전예보에 대한 수요가 증대되고 있는 제주도를 대상으로 일단위 중기, 월단위 장기 예보모형을 개발함에 있어 기상통계모형의 적용 가능성을 확인함과 아울러 모형실험을 통하여 효과적인 예보모형을 선별함에 있다. 기상통계모형 실험에 있어서는 제주도 기상관측자료의 군집분석(cluster analysis)에 의한 국지풍계 분류를 참조하여 동일풍계 내에서 지리적으로 가장 가까운 기상관측소 4쌍을 선택하고 전이함수모형(transfer function model), 시계열자기회귀모형(time-series autoregressive model), 신경망모형(neural networks model)에 의한 교차예측(cross prediction)을 실시하여 예측 정확도를 평가하였다.

참고로 본 연구에서 동일풍계 내에 있는 기상관측지점 두 개의 자료를 이용하는 것은, Vaas와 Kim(2008)의 제주도 풍속예보 연구사례에서 확인한 바와 같이 예측지점의 자체의 단일 기상자료만으로 예측하는 것

보다 기상학적 상관성이 높은 인근지점 기상자료를 복합적으로 활용함으로써 예측 정확도의 대폭적인 향상이 기대되기 때문이다.

## 2. 자료 및 방법

### 2.1. 기상관측자료

기상통계모형 실험을 위한 제주도 기상관측자료로는 김 등(2008b)이 선행연구를 통하여 풍력자원 평가에 적합하도록 재가공한 제주도 풍력자원 데이터베이스를 이용하였다. 그리고 종속변수로 풍속을 고려되 일단위 평균풍속으로 환산하여 기상통계모형 실험을 수행하였다. 실험에 사용된 기상관측자료는 이상값 및 결측값에 대해 보정된 2000년부터 2005년까지의 6년간 자료이며, 이 중 5년간의 자료는 모형구축에, 나머지 1년간의 자료는 모형평가에 사용하였다.

### 2.2. 군집분석

군집분석은 군집특성을 알 수 없는 개체들로부터 변수의 유사성(similarity)에 대한 거리를 구하여 가까운 거리에 있는 개체들을 집단화시키는 통계적 방법이다. 기상관측자료의 군집분석은 기상요소의 유클리디안 거리(Euclidean distance)를 이용하는 K-평균 군집분석을 적용하였다(정 등, 2009).

우리나라는 복잡한 지형과 복합적인 풍계로 인하여 공간적 측면에서 바람의 일률성이 현저하게 낮기 때문에 인근의 기상관측지점이라고 하더라도 동일 풍계가 아닌 경우가 많다(김 등, 2008b; 정 등, 2009). 이러한 풍계특성을 반영한 신중한 기상관측자료 선택을 통해서만 신뢰성 높은 풍력예보모형을 구축할 수 있다.

### 2.3. 기상통계모형

#### 2.3.1. 전이함수모형

전이함수모형은 입력계열이 원인이고 출력계열이 결과인 경우에 적용되는 시계열(time-series) 통계모형이다. 전이함수모형의 기본형태는 다음 식과 같다.

$$Y_t = \mu + \sum_i \nu_i(B) X_{i,t} + N_t \quad (1)$$

위 모형식이 의미하는 바는 출력계열( $Y_t$ )을 입력계열( $X_t$ )로 설명함에 있어 입력계열의 현재 및 과거 모

든 정보를 이용하여 설명하려는 것이다.

전이함수모형의 목적은 전이함수 모수  $\nu_i(B)$ 의 추정, 입력계열과 독립성을 가지는 잡음과정  $N_t$ 의 설정에 있다. 그러나 전이함수는 다음 식과 같이 무한개의 모수를 가지므로, 이를 유한개의 모수를 갖는 다항식의 분수형태로 나타낼 수 있다. 즉,

$$\nu_i(B) = \sum_{j=0}^{\infty} \nu_{i,j} B^j = \left[ \frac{\omega_i(B)}{\delta_i(B)} \right] B^{b_i} \quad (2)$$

전이함수모형을 설정하는 단계로는, 입력계열의 사전 백색화, 출력계열의 변환, 전이함수 가중값의 추정을 통하여 최종적으로 전이함수 모수를 추정하여 모형식을 완성하게 된다.

전이함수모형에 대한 상세한 설명은 백(2006)을 참고하기 바란다.

### 2.3.2. 시계열자기회귀모형

시계열회귀모형을 이용한 분석방법은 사실상 전이함수모형의 특별한 경우로서 자료에 대하여 독립변수로 시간, 계절성을 설명하는 가변수 또는 순환요인을 설명하는 삼각함수 등을 사용하여 회귀모형을 적합시키는 것을 말한다.

회귀분석을 이용하여 미래의 풍속을 예측함에는 몇 가지 난점이 있다. 첫째로 풍속은 시간변동성이 매우 크기 때문에 선형적합선을 사용하는 회귀분석으로는 풍속예보의 모형화에 제한적일 수밖에 없다. 둘째로 회귀분석에서 미래풍속에 대한 예측은 독립변수의 예측이 선행되어야 한다는 현실적 모순이 존재한다. 마지막으로 회귀분석에서는 오차항이 서로 독립이라는 가정이 전제되어야 하는데 시계열 자료에는 본질적으로 자기상관이 존재하므로 이러한 가정을 위배하게 된다.

본 연구에서는 이러한 문제들을 효과적으로 해결하기 위하여 예측하고자 하는 지점의 풍속을 독립변수로 두고 샘플링 과정을 통하여 모형을 적합하였다.

### 2.3.3. 신경망모형

두뇌 신경생리로부터 영감을 받아 개발된 신경망 모형은 Fig. 1과 같이 은닉마디(hidden units)라고 불리는 독특한 구성요소에 의하여 일반적인 통계모형과 구별되어진다. 은닉마디는 인간 신경세포를 모형화한

것으로써, 각 은닉마디는 입력변수들의 결합을 수신하여 목표변수에 전달한다. 이때 결합에 사용되는 계수를 연결강도(synaptic weight)라고 부르며, 활성화함수는 입력값을 변환하고 이를 입력으로 사용하는 다른 은닉마디로 출력하게 되는 구조로 모형이 구성된다.

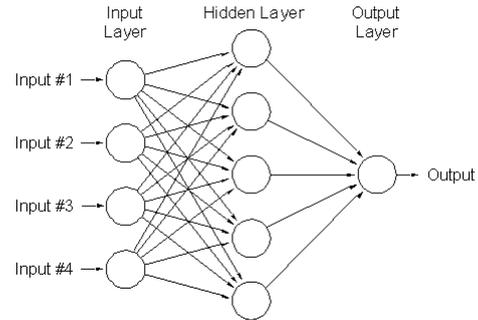


Fig. 1. Neural network structure of a MLP (Multi-Layer Perception) model.

## 3. 결과 및 고찰

### 3.1. 기상관측자료 군집분석

제주도 기상관측지점 20개소에 대한 풍속을 대상으로 군집분석을 실시하였다. Fig. 2는 군집분석에 의한 기상관측지점의 상관성을 보여주는 수목도(dendrogram)이고, Fig. 3은 6개의 군집으로 분류된 기상관측지점을 심볼로 표시한 지도이다. Fig. 3에서 점선으로 연결된

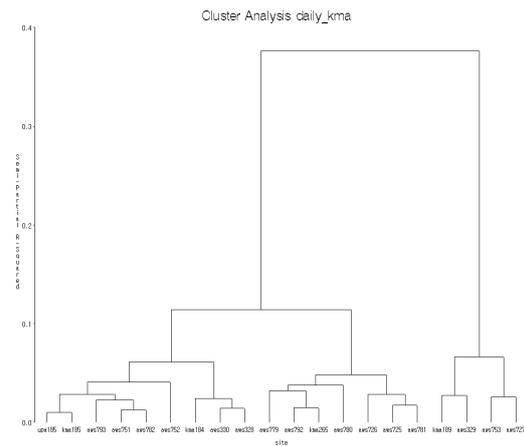


Fig. 2. Dendrogram of a K-mean clustering of daily wind speed of 20 meteorological observation sites in Jeju.

4쌍의 기상관측지점은 각각 상이한 풍계영역 내에서 풍속요소의 상관도가 높을 뿐 아니라 지리적으로도 가장 가까이 위치한 지점들을 선택한 것이다. 즉, 선홍2리[aws751]↔성판악[aws782], 하원[aws330]↔중문[aws328], 성산포[kma265]→가시[aws792], 그리고 우도[aws725]↔구좌[aws781]이다. 여기서 [ ]에 표기된 기호는 기상청 분류코드로, kma는 기상관측소, aws는 자동기상관측소를 의미하며, ↔는 교차예측을, →는 단방향 예측을 의미한다.

### 3.2. 기상통계모형 실험

3가지 기상통계모형에 대해서 4쌍의 기상관측자료의 일단위 풍속예측에 대한 실험을 수행하였다. 풍속예측은 한 가지 조합을 제외하고는 교차예측을 실시하였기 때문에 총 21가지 실험이 수행되었다. 지면관계 상 각각의 실험 중 대표적인 경우에 대해서만 실험과정과 결과를 설명하기로 한다.

실험결과로부터 각 지점의 풍속예보에는 어떤 모형을 적합한지를 판별하기 위한 기준으로 AIC, SBC를 검토하였다.

AIC(Akaike's information criterion)는 적합모수의 수가 커짐에 따라 로그 최소값은 음수로 작아지는 경향이 있으므로 최소 AIC기준은 서로 반대방향으로

작용하는 전자와 후자를 절충시킴으로써 최적모형을 선택하게 되는 것이다. 결국 AIC는 음의 작은 값을 가질수록 모형의 적합도가 우수함을 의미한다(Akaike, 1974). SBC(Schwartz's bayesian criterion)은 AIC와 유사한 기준 통계량으로 음의 작은 값을 가질수록 모형이 잘 적합되었음을 의미한다(Schwarz, 1978).

#### 3.2.1. 전이함수모형 실험

우도[aws725]→구좌[aws781] 실험에 앞서 입력계열의 분산특성을 파악하기 위해 Fig. 4와 같이 풍속 시도표를 작성하였다. 시도표를 보면 추세는 보이지 않으나 주기성을 가지며 분산이 불안정한 형태를 나타내므로 분산 안정화가 필요함을 확인할 수 있다. 분산 안정화를 위해 입력계열 풍속에 대해 로그변환(logarithmic transformation)을 실시하였다. 참고로 분산특성이 와이블(Weibull) 형태인 풍속자료의 경우 로그변환을 통하여 정규분포로 변환함으로써 통계모형의 적합도를 향상시킬 수 있었다(이 등, 2006).

로그변환한 후의 시도표인 Fig. 5를 보면 풍속의 분산은 변환 전보다 안정화되었으며 다소 주기성은 보이는 하지는 않지만 계절성에 의한 변동이 평활화 되었으므로 정상적인 통계분포 형태라고 판단할 수 있다.



Fig. 3. Clustering result of the meteorological observation sites classified into the 6 groups (symbol) and the 4 site pairs for meteor-statistical model test (dashed lines).

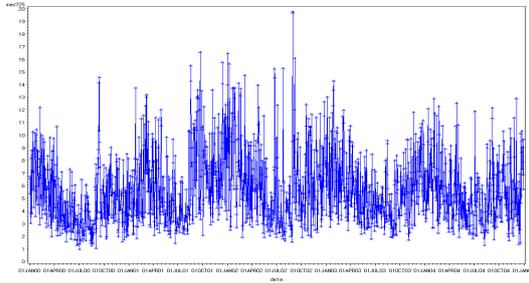


Fig. 4. Time-series plot of daily wind speed at [aws725].

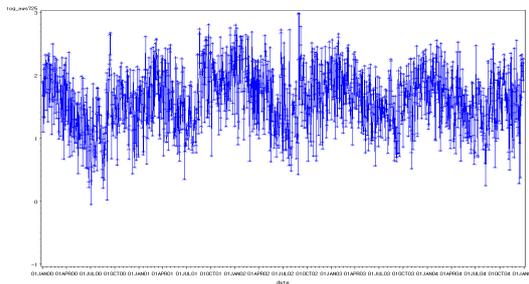


Fig. 5. Time-series plot of daily wind speed at [aws725] after a logarithmic transformation.

① ARIMA 식별

입력계열인 우도[aws725] 풍속이 정상 시계열인지를 확인하고 모수를 구하기 위해 자기상관함수(ACF; Auto-Correlation Function)와 편자기상관함수(PACF; Partial Auto-Correlation Function)를 구하여 모형을 식별해 보았다. ACF는 거의 모든 시차에서 유의한 상관관계가 보이나, PACF는 시차1에서 가장 유의한 상관관계가 보이며 시차3에서도 유의한 상관관계가 보인다. 백색잡음에 대한 자기상관검정 결과를 살펴본 결과, 로그변환된 [aws725] 풍속시계열은 정상성을 만족함을 최종 확인하였다.

② ARIMA 추정

ACF와 PCAF를 가지고 모형적합을 한 결과, 입력계열 [aws725]에 대한 모형은 다음과 같다.

$$[aws725] = \exp\left(\frac{1 - C_1B^2}{1 + C_2B - C_3B^2 - C_4B^3}\right) a_t \quad (3)$$

단  $C_1=0.99945$ ,  $C_2=0.49119$ ,  $C_3=0.92377$ , 그리고  $C_4=0.41779$ 이다. 이 모형의 모수에 대한 추정량은 모두 유의한 값을 나타내며,  $AIC=1933$ ,  $SBC=1955$ 로

산출되었다. 포트만토 검정(portmanteau test) 결과, 모형은 타당하며 잔차의 ACF와 PACF도 좋은 것으로 나타났다.

③ 전이함수 추정

최종적으로 구성된 우도[aws725]→구좌[aws781] 풍속예측을 위한 전이함수모형은 다음과 같다.

$$\log([aws781]_t) = \log([aws725]_t) + \frac{1 + C_5B - C_6B^2 - C_7B^3}{1 - C_8B^2} a_t \quad (4)$$

$$\log([aws781]_t) = C_8 \log([aws781]_{t-2}) + \log([aws725]_t) - C_8 \log([aws725]_{t-2}) + a'_t \quad (5)$$

단  $C_5=0.3866$ ,  $C_6=0.8125$ ,  $C_7=0.22138$ , 그리고  $C_8=0.97199$ 이다.

이 모형을 해석하면 구좌의 풍속은 자신의 2일전 풍속 그리고 우도의 현재 풍속, 2일전 풍속의 지배적 영향을 받음을 알 수 있다. 입력계열 [aws725]와 잔차계열의 상관관계 검정을 위한 포트만토 검정에서도 상호간의 독립성이 보장되므로 전이함수모형적합이 잘 이루어졌음을 확인하였다. 이 모형은  $AIC=-323$ ,  $SBC=-296$ 으로 나타났다.

전이함수모형에 의한 구좌[aws781] 예측풍속의 시도표를 Fig. 6에 도시하였다. 예측풍속과 함께 그려진 95% 신뢰구간을 벗어나는 시점이 없는 것으로 보아서도 모형은 매우 적절히 구축되었음을 시각적으로 판단할 수 있다.

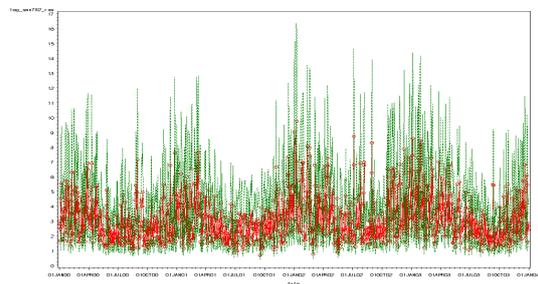


Fig. 6. Time-series plot of predicted wind speed at [aws781] with 95% confidence interval.

3.2.2. 시계열자기회귀모형 실험

중문[aws328]→하원[aws330] 실험에서는 하원의 현재 풍속[aws330]<sub>t</sub>는 하원의 어제 풍속[aws330]<sub>t-1</sub>과 중문의 현재 풍속[aws328]<sub>t</sub>, 어제 풍속[aws328]<sub>t-1</sub>을 독립변수로 설정하여 회귀모형을 구축하였다.

도출된 하원[aws330] 풍속의 시계열회귀모형은 다음과 같다.

$$[aws330]_t = 0.4727 + 0.2494 [aws330]_{t-1} + 1.3396 [aws328]_t - 0.2839 [aws328]_{t-1} \quad (6)$$

참고로 이 모형은 AIC=-180이고 SBC=-157이다.

3.2.3. 신경망모형 실험

선홍2리[aws751]→성판악[aws782] 실험에서 구축된 신경망은 Fig. 6과 같다. 즉, 성판악 현재 풍속[aws782]<sub>t</sub>는 성판악 1일전 풍속[aws782]<sub>t-1</sub>, 2일전 풍속[aws782]<sub>t-2</sub>, 3일전 풍속[aws782]<sub>t-3</sub>, 그리고 선홍2리 현재 풍속[aws751]<sub>t</sub>, 1일전 풍속[aws751]<sub>t-1</sub>, 2일전 풍속[aws751]<sub>t-2</sub>을 입력변수로 신경망이 구축되었다. 참고로 이 모형은 AIC=-485, SBC=-366이다.

일반적으로 신경망모형은 예측성능이 우수함에 비하여 Fig. 7에서도 알 수 있듯이 모형이 매우 복잡하게 엮여져 있기 때문에, 모형의 특성에 대한 이해 및 변수 간 인과관계에 대한 해석은 반대로 매우 어렵다는 단점이 있다.

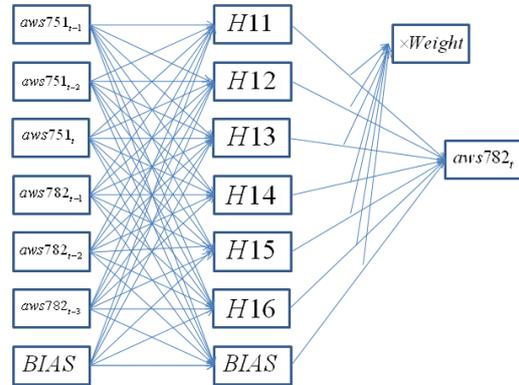


Fig. 7. MLP neural network structure established for [aws751]→[aws782] case.

3.3. 기상통계모형 비교평가

세 가지 풍속예보모형 즉, 전이함수모형, 자기회귀모형, 신경망모형을 모형 평가척도 중 AIC와 SBC로 비교한 결과를 Table 1에 정리하였다.

종합적으로 판정하여 본다면 시계열회귀모형은 1가지 경우에서만 가장 우수한 예측결과를 보였으며 신경망모형은 3가지, 전이함수모형도 3가지 경우에서 타 모형에 비해 우수한 예측결과를 보였다.

4. 결론

- (1) 제주도 풍력발전예보 모형을 개발함에 있어 군

Table 1. Comparison AIC and SBC among the predictions of three meteor-statistical models

Case	Correlation factor	Evaluation measure	Transfer function	Time-series autoregression	Neural network
Seonheul 2 → Seongpanak	0.88	AIC	254	-434	-485
		SBC	281	-401	-336
Seongpanak → Seonheul 2	0.87	AIC	254	-620	-651
		SBC	281	-587	-531
Hawon → Joongmoon	0.87	AIC	-2870	-974	-924
		SBC	-2849	-950	-804
Joongmoon → Hawon	0.81	AIC	-99	-181	-179
		SBC	-78	-157	-59
Seongsanpo → Gasi	0.82	AIC	-528	-1028	-1119
		SBC	-491	-995	-961
Woodo → Gujwa	0.82	AIC	-323	-9	-227
		SBC	-296	-11	-72
Gujwa → Woodo	0.82	AIC	-150	367	194
		SBC	-128	387	349

집분석을 통하여 동일풍계에 위치하며 기상학적 상관도가 높은 참조자료를 보조적으로 활용함으로써 예보 정확도를 향상시키는 방법을 개발하였다.

(2) 일단위 풍속예보에 적합한 기상통계모형을 선정하기 위하여 전이함수모형, 자기회귀모형 그리고 신경망모형을 실험하였으며 그 결과 신경망모형과 전이함수모형의 적합도가 높은 것을 확인하였다. 그러나 신경망 분석은 모형의 복잡성에 비하여 전이함수모형 대비 월등한 예측 정확도를 제공하지는 못한 반면, 전이함수모형은 모형 해석을 통하여 기상학적 통계특성을 이해하기 쉽다는 장점이 있다.

(3) 본 연구에서 적용한 전이함수모형은 기존의 신경망 분석이나 일반 선형회귀모형에서는 고려하지 못한 시계열 자료들 간의 자기상관을 고려하였으며, 예측지점 자신의 자기상관뿐만 아니라 동일풍계 내 기상학적 상관성이 높은 참조지점의 시계열 자료까지도 동시에 고려함으로써 모형의 적합도를 대폭 향상시킬 수 있었다.

**감사의 글**

본 연구는 한국에너지기술연구원의 부처임무사업 지원을 받아 수행되었습니다. 자료분석을 위해 수고를 아끼지 않은 김덕현 군에게 감사를 포함합니다.

**참 고 문 헌**

김현구, 이영섭, 장문석, 경남호, 2006, 월령단지 풍력발전 예보모형 개발에 관한 연구, 한국태양에너지학회지, 26(2), 27-34.

김현구, 이영섭, 장문석, 경남호, 2006, 풍력발전 예보시스템 KIER-Forecaster의 개발, 한국신재생에너지학회지, 2(2), 37-43.

김현구, 경남호, 장문석, 2008, 풍속패턴 분류를 통한 풍력발전 예보방법, 특허등록번호 10-0802402.

김현구, 장문석, 이은정, 2008, 제주도 풍력자원 데이터베이스 구축을 위한 기상통계분석, 한국환경과학회지, 17(6), 591-599

김경보, 박윤희, 박정근, 고경남, 허종철, 2010, 기상청 MOS 예측값 적용을 통한 풍력 발전량 예측 타당성 연구, 한국태양에너지학회 논문집, 30(2), 46-53.

김경보, 양경부, 박윤희, 문창은, 박정근, 허종철, 2010, 제주지역 풍력발전량 실시간 감시 시스템 구축에 관한 연구, 한국태양에너지학회 논문집, 30(3), 25-32.

백경원, 2006, 이변수 시계열의 전이함수모형 분석, 석사학위논문, 연세대학교.

이영섭, 김현구, 박종석, 김희경, 2006, 변수변환을 통한 포항지역 미세먼지의 통계적 예보모형에 관한 연구, 한국대기환경학회지, 22(5), 614-626.

정우식, 이화운, 박종길, 김현구, 김은별, 최현정, 김동혁, 김민정, 2009, 남한지역 풍력자원 평가 및 바람지도 구축을 위한 바람권역 분류, 한국환경과학회지, 18(8), 899-901.

Akaike, H., 1974, A New Look at Statistical Model Identification, IEEE Transactions on Automatic Control, AU-19, 716-722.

Schwarz, G., 1978, Estimating the Dimension of a Model, Annals of Statistics, 6, 461-464.

Vaas, F., Kim, H. G., 2008, Application of Neural Network for Long-Term Correction of Wind Data, J. Korean Soc. for New and Renewable Energy, 4(4), 23-29.