

이력 데이터를 이용한 대체에너지원 추정 기법

An Estimation Method of Renewable Energy Resources Using History Data

오 인 배(In-Bae Oh)¹⁾ 안윤애(Yoon-Ae Ahn)²⁾

요약

대체에너지원 데이터는 그 측정값이 시시각각으로 변경되는 특성을 갖기 때문에 시간의 변화에 따라 발생하는 원시 데이터의 양이 상당히 방대하다. 따라서 대체에너지원 데이터의 이력 정보를 체계적으로 저장 및 관리할 수 있는 이력 데이터베이스의 구축과 대체에너지원 예측 시스템에 대한 연구가 필요하다. 이 논문에서는 대체에너지원 데이터의 과거 측정값에 대한 추정 기능의 지원을 통해 시시각각으로 변화하는 데이터를 체계적으로 관리하기 위한 대체에너지원 추정 기법을 제안하였다.

Abstract

A renewable energy resource data has the characteristic that its measurement value changes every moment. So the amount of primitive data which is produced by the change of time are vast. Therefore researches are needed for the construction of history database which can save and manage vast amount of history information of renewable energy resource data systematically, the estimation system of renewable energy resources. In this paper, to solve these problems, the estimation method of renewable energy resources is suggested. The method makes it possible to estimate momentarily changing data of the past systematically.

논문접수 : 2004. 11. 20.

심사완료 : 2004. 12. 15.

1) 정회원 : 주성대학 인터넷정보과 부교수

2) 정회원 : 청주과학대학 컴퓨터학과 전임강사

1. 서론

태양 에너지, 풍력, 소수력, 바이오매스 등과 같은 대체에너지원 데이터는 시간에 따라 그 측정값이 시시각각으로 변경되는 특징을 갖고 있으며, 다양한 공간적인 주변 환경 요소와 밀접하게 관련되기 때문에 실세계에서 시간의 변화에 따라 발생하는 원시 데이터의 양이 매우 커지게 된다[9][10]. 이와 같은 특징들을 고려할 때 대체에너지원 데이터는 시공간(spatiotemporal) 데이터로 분류할 수 있으며 [2][5], 대체에너지원 데이터의 방대한 이력 정보를 체계적으로 저장 및 관리할 수 있는 이력 데이터베이스의 구축과 대체에너지원 데이터에 대한 과거 측정값을 추정하는 기능을 제공하는 예측 방법에 대한 연구가 필요하다.

지금까지 국외에서는 대체에너지원 정보 관리 시스템의 개발 및 보급이 활발히 추진되고 있으며, 이에 대한 정보를 인터넷으로 서비스하기 위해 많은 노력을 하고 있다[1][3][4]. 그러나 아직 국내에서는 대체에너지원 데이터의 중요성에 대한 인식 및 이들 데이터에 대한 체계적인 이력 관리 및 인터넷을 통한 데이터 검색 서비스가 효과적으로 제공되지 않고 있다. 또한, 기상 정보 시스템이나 국토 관리 시스템과 같이 GIS를 기반으로 하는 시스템들은 현재 데이터베이스를 구축하여 활용 단계에 있지만 이력 데이터 관리를 위해 다양한 연산 및 데이터 검색 기능을 제공하지 않으며, 측정된 대체에너지원 데이터들을 자동으로 처리하지 못하고 있다[7][8]. 특히, 대체에너지와 관련된 데이터는 시간의 변화에 따라 측정값이 변경되는 특성으로 인하여 측정된 데이터들이 새롭게 추가되기 때문에 데이터의 양이 커지게 된다. 따라서 이와 같은 대용량 데이터를 관리하기 위한 효율적인 방법이 제공되지 못하고 있다. 이 논문에서는 현재까지 국내 20개 도시에서 측정된 일사량 데이터와 33곳의 지역에서 측정된 풍력 데이터를 체계적으로 관리하기 위한 저장구조를 모델링하고, 이를 처리하기 위한 대

체에너지원 데이터의 추정 기법을 제안한다. 이를 위해 대체에너지원 데이터가 갖는 시간적 특성을 고려한 통계적 예측 기법 중에서 시계열 분해법을 적용하여 과거의 측정값을 추정하는 기능을 통해 대용량 데이터를 효율적으로 관리하는 방법을 제안한다.

이 논문의 전체 구성은 다음과 같다. 먼저, 2장에서는 대체에너지원 데이터의 과거 측정값을 추정하기 위해 사용한 시계열 분석법의 특징을 살펴본다. 3장에서는 대체에너지원 이력 데이터의 효율적인 관리를 위한 데이터베이스 구조를 기술한다. 4장에서는 시계열 분해법을 이용한 대체에너지원 데이터의 추정 방법을 기술한다. 5장에서는 평가 분석 내용을 기술하고, 마지막으로 6장에서는 결론을 맺는다.

2. 관련연구

시계열 데이터 분석 방법 중에서 시계열 분해법(time series decomposition method)은 추세분석, 계절요인 분석, 순환요인 분석, 우연요인 분석과 같은 네 가지 유형의 분석 방법으로 구분한다. 각각의 분석 방법에 대한 특징들은 다음과 같다[6][13][14][15].

추세분석(trend analysis)은 과거의 데이터 속에 어떠한 추세가 있는지를 파악하기 위해서는 과거의 데이터를 시간에 따라 그래프 형태로 표시한다. 이때 표시되는 각 점을 연결하고, 목측법(free hand method)이나 통계적 방법을 이용하여 과거 데이터의 추세가 직선 형태인지, 아니면 포물선 형태인지를 결정할 수 있다. 만일, 데이터의 추세가 직선 형태라면 직선 추세선을 구한 다음 하단 부분을 연장함으로써 과거의 측정값을 추정할 수 있다. 계절요인 분석(seasonal component analysis)에서 계절 지수(seasonal index)는 시계열의 값이 계절적 요인에 의하여 추세로부터 변동하는 정도를 나타낸다. 여기에서 지수는 추세의 퍼센트로 표현하므로 각 계절에 하나의 지수가 있게 된다. 따라서 월별 데이터의 경우에는 12개, 분기별

데이터의 경우에는 4개의 지수를 갖는다. 아울러 과거의 데이터를 이용하여 계절 지수를 결정하면 이를 예측된 추정 값에 적용하여 계절 조정된 예측 값을 얻을 수 있다. 이때, 계절 지수를 계산하는 방법 중에서 가장 많이 사용되는 것은 중심이동평균비율법(ratio to centered moving average method)이다. 순환요인 분석(cyclical component analysis)에서 순환 변동은 계절 변동과 동일한 방법으로 측정할 수는 없지만 월별, 분기별, 그리고 연도별 데이터를 이용하여 측정할 수 있다. 또한 연도별 데이터를 사용할 경우에는 시계열 데이터 속에 추세 요인과 순환요인만이 포함되어 있다고 가정하므로 계절 변동이나 불규칙 변동을 제거할 필요가 없다. 우연요인 분석(random component analysis)에서 우연 변동은 전혀 예측할 수 없는 요인에 의하여 일어난다고 볼 수 있다. 그러므로 시계열의 불규칙 변동은 추세 변동, 계절 변동, 순환 변동에 의해서 설명될 수 없는 변동이다.

이 논문에서는 대체에너지원 데이터가 갖는 시간적 특성을 고려하여 시계열 데이터 분석 기법 중에서 시계열 분해법을 적용하여 대체에너지원 데이터의 과거 추정값을 추정하는 기능을 지원함으로써 대용량 데이터를 효율적으로 관리하기 위한 방법을 제안한다.

3. 대체에너지원 데이터의 저장구조

대체에너지원 데이터의 이력 정보를 체계적으로 저장 및 관리하기 위한 데이터베이스의 구조를 실제 사용되는 일사량 데이터와 풍력 데이터로 구분하여 기술한다.

3.1 일사량 데이터

일사량 데이터는 기본적으로 모두 12가지 유형이 존재한다[9][10]. 이와 같은 유형의 데이터가 갖는 일반 속성 정보는 실제 일사량 관련 수치 값들이 서로 약간씩 다르게 구성되어

있다. 또한 일사량 데이터의 유형에 따라 서로 다른 시간의 주기를 갖는다. 이로 인해 하나의 릴레이션 스키마로 모든 일사량 데이터를 저장할 수 없기 때문에 (표 1)과 같이 서로 다른 4개의 릴레이션 스키마를 구성하여 12가지 유형의 일사량 데이터를 관리한다[7][8].

(표 1) 일사량 데이터의 스키마 유형

구분	데이터 유형
schema_1	ilsa, ilcho, cloud, rate, clean, temp, humi, wind
schema_2	out_radi
schema_3	clean_radi, g_radi_element
schema_4	slope_radi

(표 1)은 일사량 데이터베이스를 구성하는 서로 다른 4가지 유형의 릴레이션 스키마를 구분한 것이다. schema_1은 수평면 전일사량(ilsa), 일조율(ilcho), 운량(cloud), 일사율(rate), 청명일수(clean), 기온(temp), 상대습도(humi), 바람(wind)과 관련된 속성 데이터를 저장하며, (표 2)와 같은 구조를 갖는다.

(표 2) 일사량 : schema_1

vtstart	vtend	ilsa	ilcho	cloud	rate	clean	temp	humi	wind
1996-01-01	1996-01-31	1978	612	36	65	14	-47	69	13
1996-12-01	1996-12-31	1586	541	38	57	12	-18	78	9
1997-01-01	1997-01-31	1851	495	38	63	14	-58	75	13
1997-12-01	1997-12-31	1414	490	42	61	10	-9	79	12

(표 2)는 schema_1의 속성 이력 릴레이션에 저장된 속성 값을 나타낸다. schema_1의 유효 시간 주기는 월(month) 단위로 사용된다. schema_2는 대기권밖 일사량(out_radi)과 관련된 속성 데이터를 저장하는 구조이며, (표 3)과 같은 구조를 갖는다.

(표 3) 일사량 : schema_2

vtstart	vtend	out_ilsa
1997-01-01	1997-01-01	3571.7
1997-01-31	1997-01-31	4388.0
1997-12-01	1997-12-01	3699.7
1997-12-31	1997-12-31	3559.9

(표 3)은 schema_2의 속성 이력 릴레이션에 저장된 속성 값을 나타낸다. schema_2의 유효 시간 주기는 일(day) 단위로 사용된다. schema_3은 청명 일사량(clean_radi), 전일사량 성분(g_radi_element)과 관련된 데이터를 저장하기 위한 구조이며, (표 4)와 같은 구조를 갖는다.

(표 4) 일사량 : schema_3

vtstart	vtend	clean	rd	diffuse	direct	rt
1997-01-07	1997-01-08	27.91	0.034	24.5	20.2	0.026
1997-01-08	1997-01-09	129.76	0.076	55.3	60.7	0.066
1997-01-09	1997-01-10	244.93	0.110	80.4	106.4	0.107

(표 4)는 schema_3의 속성 이력 릴레이션에 저장된 속성 값을 나타낸다. schema_3의 유효 시간 주기는 시간(hour) 단위로 사용된다. schema_4는 방위별 경사면 일사량(slope_radi)과 관련된 데이터를 저장하며, (표 5)와 같은 구조를 갖는다. 방위별 경사면 일사량 데이터는 E(동), W(서), S(남), N(북), SE(남동), NE(북동) 방향에 대하여 여섯 개의 방위로 구분된다.

(표 5) 일사량 : schema_4

vtstart	vtend	angle	orient	slope_ilsa
1997-01-01	1997-01-31	0	E	1746
1997-01-01	1997-01-31	3	E	1754
1997-01-01	1997-01-31	87	E	953
1997-01-01	1997-01-31	90	E	1060
1997-01-01	1997-01-31	0	NE	1746
1997-01-01	1997-01-31	3	NE	1670
1997-01-01	1997-01-31	87	NE	0
1997-01-01	1997-01-31	90	NE	0
1997-12-01	1997-12-31	0	E	1449
1997-12-01	1997-12-31	3	E	1458
1997-12-01	1997-12-31	87	E	798
1997-12-01	1997-12-31	90	E	878
1997-12-01	1997-12-31	0	NE	1449
1997-12-01	1997-12-31	3	NE	1383
1997-12-01	1997-12-31	87	NE	0
1997-12-01	1997-12-31	90	NE	0

(표 5)는 schema_4의 속성 이력 릴레이션에 저장된 속성 값을 나타낸다. schema_4의 유효 시간 주기는 월 단위로 사용된다.

3.2 풍력 데이터

풍력 데이터의 릴레이션 스키마는 하나로 구성되며, 데이터 속성은 16가지로 세분화된다 [7][8][11][12]. 풍력 데이터의 스키마 유형을 살펴보면 (표 6)과 같다.

(표 6) 풍력 데이터의 스키마 유형

구분	데이터 유형
풍력 데이터	inner_temp, volt, wind30, dev30, wind15, dev15, wind, dev, wind1, dev1, actinometry, atmospheric_press, dev2, dev3, humidity, dev4

(표 6)은 풍력 데이터의 데이터베이스를 구성하는 릴레이션 스키마의 일반 속성을 나타낸 것으로 내부온도(inner_temp), 전압(volt), 30m 풍속(wind30), 30m 표준편차(dev30), 15m 풍속(wind15), 15m 표준편차(dev15), 풍향(wind), 표준편차(dev), 풍향1(wind1), 표준편차1(dev1), 일사량(actinometry), 표준편차2(dev2), 대기압(atmospheric_press), 표준편차3(dev3), 습도(humidity), 표준편차4(dev4)로 구성된다.

풍력 데이터베이스는 하나의 스키마로 16가지의 풍력 데이터 속성을 동일한 이력 릴레이션에서 관리하며, 그 내용은 (표 7)과 같다.

(표 7) 풍력 : schema

vtstart	vtend	inner_temp	volt	wind30	dev30	wind15	dev15
2000-08-25-01-00	2000-08-25-01-59	21.2	126	1.03	0.91	0.97	0.87
2000-08-25-02-00	2000-08-25-02-59	20.8	126	0.73	0.67	0.68	0.64
2000-08-26-01-00	2000-08-25-01-59	20.3	11.7	0.24	0.54	0.18	0.50
2000-08-26-02-00	2000-08-25-02-59	20.3	11.7	0.38	0.59	0.25	0.54

① (표 7)은 풍력 데이터의 속성 이력 릴레이션의 예이다. 여기에서 각 속성은 유효시간의 시작시간(vtstart), 유효시간의 종료시간(vtend), 내부온도(inner_temp), 전압(volt), 30m 풍속(wind30), 30m 표준편차(dev30), 15m 풍속(wind15), 15m 표준편차(dev15)를 나타내며, 유효 시간 주기는 시간(hour) 단위로 사용된다.

4. 대체에너지원 추정 기법

지난 수십년간 측정된 대체에너지원 데이터의 양은 기하급수적으로 증가되어 왔다. 즉, 대체에너지와 관련된 데이터는 시간별, 일별, 월별 등으로 측정하여 저장하기 때문에 시간 속성을 갖는 데이터의 특성으로 인하여 매년 측정된 데이터들이 새롭게 추가되기 때문에 데이터의 양은 시간의 변화에 따라 매우 커지게 된다. 따라서 이러한 대용량 데이터들을 단순히 저장하는 것은 저장 공간을 낭비하는 결과를 초래할 뿐만 아니라 데이터 검색 시간도 문제가 될 수 있다. 이와 같은 문제를 해결하기 위해 대용량 데이터를 관리하기 위한 방안으로 일정기간 측정된 데이터를 데이터베이스에 저장한 후, 기존의 저장된 데이터를 이용하여 과거 측정값을 추정하는 기능을 통해 대용량의 대체에너지원 데이터를 관리하는 방법을 제안한다.

이 논문에서는 대체에너지원 데이터가 갖는 시간적 특성을 고려하여 시계열 데이터 분석 방법 중에서 시계열 분해법의 추세분석과 계절요소 분석 방법을 이용한 과거 측정값의 월별 추정 기능을 통해 대체에너지원 데이터의 이력 정보를 체계적으로 관리하는 방법을 제안한다. 여기에서 대체에너지원 데이터에 대한 과거 측정값을 추정하기 위해 필요한 추세분석 과정을 개략적으로 기술하면 (그림 1)과 같다.

직선 추세선을 구하기 위해 최소 자승법을 이용하여 매개변수 a와 b값 결정

직선 추세선으로부터 분산을 측정하기 위해 추정의 표준오차 사용

직선 추세선($Y' = a + bX$) 결정

직선 추세선의 하단 부분을 연장하여 과거의 측정값을 추정

(그림 1) 과거 측정값의 추세분석 과정

4.1 과거 측정값의 추세 분석

이 논문에서는 대체에너지원 데이터에 대한 과거의 측정값을 그래프상에 표현한 각 점들(산포도)의 추세가 (그림 2)와 같이 직선 형태를 갖기 때문에 과거 측정값의 추세 분석을 통해 직선 추세선을 구한 다음 해당하는 직선의 하단 부분을 연장함으로써 과거의 측정값을 추정할 수 있다. 여기에서 추세선이 직선이라는 것은 각 기간에 있어서 평균 측정값의 증가(또는 감소)가 일정하다는 것을 의미한다. 직선 추세선(Y')의 표현식은 (식 1)과 같다.

$$Y' = a + bX \quad (\text{식 1})$$

Y' 는 특정 기간(예를 들면, 연도 X)의 측정값, a 는 X 가 0일때의 Y 축 절편, b 는 직선의 기울기, X 는 1, 2, 3, 4, 5(예를 들면, 연도 번호) 등의 값을 갖는다. 여기에서 과거의 측정값을 이용하여 매개변수 a 와 b 값을 결정해야 하는데, 이를 위해서는 최소자승법(least squares method)을 사용한다[13][14][15]. 최소자승법이란, 각각의 실제 값과 그에 상응하는 예측 값과의 편차 자승의 총합계가 최소가 되도록 추세선을 결정하는 방법이다. 이때, 매개변수 a 와 b 값을 계산하는 표현식은 (식 2), (식 3)과 같다.

$$a = \frac{\sum Y - b \sum X}{n} \quad (\text{식 2})$$

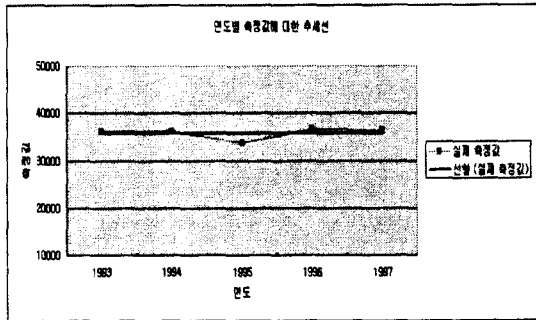
$$b = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} \quad (\text{식 3})$$

이 논문에서는 대체에너지원 데이터에 대한 17년간(1982년~1998년)의 실제 측정값 가운데 임의로 5년간(1993년~1997년)의 측정값을 선택하여 샘플 데이터로 활용하였다. (표 8)은 5년간 춘천 지역의 수평면 전일사량에 대한 실제 측정값(kcal/m²)을 나타낸다.

(표 8) 춘천지역의 수평면 전일사량 (5년간 실제 측정값)

년	1	2	3	4	5	6	7	8	9	10	11	12	합계
1993	161	230	330	367	4529	3956	3999	3381	3729	2919	1455	1426	36297
1994	167	253	337	394	4126	4216	4032	3591	3438	2231	1663	1436	36267
1995	186	205	244	402	4215	4255	2790	2919	3194	2525	2040	1548	33874
1996	197	269	309	439	4439	3376	3605	3731	3614	2384	1796	1586	36700
1997	185	256	343	402	3306	4204	3913	3882	3263	2833	1632	1414	36321

(표 8)과 같은 실제 측정값을 그래프 상으로 나타낸 산포도(scatter diagram)는 (그림 2)와 같다.



(그림 2) 춘천 지역의 측정값에 대한 추세선

(표 9)는 매개변수 a와 b값을 계산하기 위한 변수 값으로 Y(TCSR)는 실제 측정값을 나타내며, XY는 시간(X축)에 대한 실제 측정값(Y축)을 나타낸다. 이때, X는 연도(X축)에 대

한 번호를 나타낸다.

(표 9) 춘천지역의 추세선 예측을 위한 측정값

연도	X	Y=TCSR	XY	X ²	Y ²
1993	1	36297	36297	1	1317472209
1994	2	36267	72534	4	1315295289
1995	3	33874	101622	9	1147447876
1996	4	36700	146800	16	1346890000
1997	5	36321	181605	25	1319215041
합계	15	179459	538858	55	6446320415

먼저, (식 2)와 (식 3) 그리고 (표 9)를 통해 매개변수 a와 b값을 구하는 과정은 다음과 같다. 이때 직선 추세선은 Y' = a + bX와 같이 표현된다. 또한 5년간의 측정값을 이용한 추세선을 계산하기 때문에 n의 값은 5를 갖는다. 아울러, X는 시작하는 연도를 기준으로 값을 지정한다.

$$b = \frac{5(538858) - 15(179459)}{5(55) - (15)^2} = 48.1$$

$$a = \frac{179459 - 48.1(15)}{5} = 35747.5$$

a와 b값을 이용하여 과거 5년간의 실제 측정된 값에 대한 연간 추정 값과 과거(1992년도)의 측정값에 대한 연간 추정 값은 (식 1)을 참조하여 다음과 같이 계산할 수 있다.

$$Y'_i = 35747.5 + 48.1X \quad (X=0,1,2,3,4,5)$$

$$Y'_{1992} = 35747.5 + 48.1(0) = 35748$$

(연간 과거 추정 값)

$$Y'_{1993} = 35747.5 + 48.1(1) = 35796$$

$$Y'_{1994} = 35747.5 + 48.1(2) = 35844$$

$$Y'_{1995} = 35747.5 + 48.1(3) = 35892$$

$$Y'_{1996} = 35747.5 + 48.1(4) = 35940$$

$$Y'_{1997} = 35747.5 + 48.1(5) = 35988$$

최종적으로 결정되는 과거의 5년간에 대한

실제 측정값과 연간 추정 값 그리고 과거(1992년도)의 측정값에 대한 연간 추정 값은 (표 10)과 같다. 여기에서 과거에 대한 연간 추정 값은 현재 저장되어 있지는 않지만 추후 과거의 대체에너지원 데이터에 대한 월별 추정 값을 결정하는데 유용하게 사용한다.

(표 10) 춘천지역의 실제값과 과거 추정값

년도	X	실제 측정값(Y=TCSR)	연간 추정 값(T)
1992	0	-	35748
1993	1	36297	35796
1994	2	36267	35844
1995	3	33874	35892
1996	4	36700	35940
1997	5	36321	35988

(표 10)에서 제시한 1992년도의 실제 연간 측정값은 35748보다 많을 수도 있고, 적을 수도 있다는 것을 (그림 2)를 통해 쉽게 짐작할 수 있다. 이와 같은 사실은 (그림 2)에서 많은 점들이 추세선 밖에 있는 것과 관련이 있다. 만일 모든 점이 추세선 상에 있고, 1992년도의 실제 연간 측정값이 35748이 된다면 예측에 대한 오차는 없을 것이다. 그러나 100% 정확하게 추정할 수는 없기 때문에 실제 측정값의 추세선으로부터 분산(dispersion)을 측정하기 위하여 추정의 표준오차(standard error of the least squares estimate)를 사용해야 한다.

4.2 과거 측정값의 월별 추정

앞에서 제시한 (그림 2)는 (표 8)의 월별 측정값을 이용하여 연도별 측정값을 추세선과 함께 표현한 것이다. 한 가지 특이한 상황은 많은 점들이 추세선 밖에 있다는 것이다. 또한 추세선을 결정하기 위해 사용된 측정값은 계절적 요인을 포함하기 때문에 이와 같은 추세선에 근거하여 과거의 측정값을 추정하기 위해서는 이러한 예측 값을 상하로 조정해야만 한다.

여기에서 계절 지수는 추세의 퍼센트로 표현하기 때문에 각 월마다 하나의 지수를 갖게 된다. 따라서 월별 데이터의 경우 12개의 지수가 존재하게 된다. 아울러 과거의 측정값을 이용하여 계절 지수를 결정하면 이것을 예측된 추세 값에 적용하여 월별 조정 추정 값을 계산할 수 있다. 또한 중심이동평균비율법을 이용하여 계절지수를 계산하는 과정은 (표 8)의 월별 대체에너지원 데이터의 측정값을 이용한 (표 11)을 예로 기술한다.

(표 11) 특정 월별 계절지수

연도	월	측정값 (Y=TCSR)	월별 이동평균	중심이동평균(TC)	월별계절 지수(SR)
1993	1	1617			
	2	2305			
	3	3305			
	4	3676			
	5	4529			
	6	3965			
	7	3999	3025.5	3028	1.320674
	8	3381	3030.5	3040.167	1.11211
	9	3729	3049.833	3052.875	1.221472
	10	2919	3055.917	3067	0.951744
	11	1455	3078.083	3061.292	0.47529
	12	1426	3044.5	3054.958	0.466782
1994	1	1677	3065.417	3066.792	0.546826
...

기존의 저장된 측정값을 이용하여 과거의 월별 측정값을 추정하기 위한 중심이동평균비율법의 절차는 6단계의 과정으로 이루어지며, 그 과정은 다음과 같다.

- 월별 이동평균 값을 계산
- 월별 중심 이동평균 값을 계산
- 각 월별 특정 계절지수를 계산

각 월별 특정 계절지수를 연도별로 평균 처리

$$(SR)_{10,1993} = \frac{Y_{10,1993}}{CMA_3} = \frac{TCSR}{TC} = \frac{2919}{3067} = 0.951744$$

계절 조정된 월별 평균 측정값을 계산

계절 조정된 월별 추정 값을 계산

(그림 3) 과거 측정값의 월별 중심이동평균비율법

② (그림 3)에서 제시한 바와 같이 과거의 측정값을 추정하기 위한 6단계의 중심이동평균비율법을 단계적으로 기술하면 다음과 같다.

③

• **단계1** : 과거 5년간 측정된 대체에너지원 데이터의 월별 이동평균 값을 계산하여 12개월의 한 가운데 위치에 지정한다.

• **단계2** : 두 개의 월별 이동평균 값을 이용하여 재차 단계1의 과정과 같은 방법으로 중심이동평균(CMA: centered moving average) 값을 계산하여 두 개의 월별 이동평균 사이에 지정한다.

• **단계3** : 각각의 월별로 계절 변동의 영향을 분리하기 위해 실제 측정값을 그에 해당하는 중심이동평균 값으로 나누어 특정 월별 계절지수(specific month seasonal index)를 계산한다. 여기에서 $(SR)_{7, 1993}$ 은 1993년도 7월에 대한 특정 계절지수를 나타낸다.

$$(SR)_{7,1993} = \frac{Y_{7,1993}}{CMA_1} = \frac{TCSR}{TC} = \frac{3999}{3028} = 1.320674$$

$$(SR)_{8,1993} = \frac{Y_{8,1993}}{CMA_2} = \frac{TCSR}{TC} = \frac{3381}{3040.167} = 1.11211$$

$$(SR)_{9,1993} = \frac{Y_{9,1993}}{CMA_3} = \frac{TCSR}{TC} = \frac{3729}{3052.875} = 1.221472$$

• **단계4** : 각 월별 특정 계절지수를 연도별로 평균하여 우연변동을 제거하고 계절 영향을 분리시킨다. 그 이유는 특별히 우연변동 요인을 고려할 필요가 없기 때문이다.

$$S_6 = \frac{((SR)_{6,1994} + (SR)_{6,1995} + (SR)_{6,1996} + (SR)_{6,1997})}{4} = \frac{(1.39518 + 1.509847 + 1.104441 + 1.396021)}{4} = \frac{5.405489}{4} = 1.351372$$

$$S_7 = \frac{((SR)_{7,1993} + (SR)_{7,1994} + (SR)_{7,1995} + (SR)_{7,1996})}{4} = \frac{(1.320674 + 1.330693 + 0.986694 + 1.185157)}{4} = \frac{4.823218}{4} = 1.205805$$

여기에서 한 가지 유의해야 할 사항은 각 월별 계절지수의 합은 정확하게 12라는 값을 나타내야 함에도 불구하고 경우에 따라서는 12보다 크거나 작은 값을 가질 수 있기 때문에 각 월별 계절지수를 상하로 조정해야 한다는 점이다. 여기에서는 12보다 작기 때문에 상향 조정해야 한다.

$$\text{수정평균값} = \frac{\text{월별평균값} \cdot 12}{\text{월별평균값의합}} = \frac{1.205805 \cdot 12}{11.98077} = 1.20774$$

이와 같이 계산된 결과인 수정 평균값은 (표 12)에 제시한 내용과 같으며, 이것은 각 월별 계절지수를 나타낸다.

(표 12) 월별 계절지수

년월	1	2	3	4	5	6	7	8	9	10	11	12	합계
1993							1.320	1.112	1.220	0.950	0.473	0.466	
1994	0.546	0.824	1.091	1.291	1.361	1.391	1.331	1.181	1.161	0.761	0.561	0.481	
1995	0.645	0.730	0.881	1.431	1.501	1.501	0.981	1.021	1.091	0.851	0.681	0.521	
1996	0.670	0.981	1.001	1.421	1.441	1.101	1.181	1.231	1.191	0.781	0.601	0.531	
1997	0.527	0.841	1.141	1.331	1.091	1.391							
합계	2.390	3.301	4.131	5.511	5.411	6.540	4.824	4.554	4.673	3.352	2.332	2.017	
평균	0.597	0.828	1.031	1.378	1.351	1.351	1.208	1.131	1.160	0.839	0.580	0.504	1.980
추정	0.598	0.828	1.031	1.381	1.351	1.351	1.201	1.141	1.170	0.840	0.580	0.504	
평균	0.467	0.509	0.505	1.181	1.406	1.511	1.741	1.479	2.151	2.566	2.211	1.175	12

또한 추세선을 이용하여 추정한 대체에너지원 데이터의 1992년도 연간 추정 값이 35748이라는 것은 (표 10)에서 제시하였다. 이러한 값을 이용하여 1992년도에 대한 각각의 월별 추정 값을 계산하기 위한 과정은 다음과 같다.

• **단계5** : 1992년도 연간 추정 값인 35748을 12개월로 나누어 계절 조정된 월별 평균 추정 값을 계산한다. 즉, 그 값은 35748 / 12 = 2979와 같다.

• **단계6** : 다음과 같은 규칙에 월별 계절지수를 이용하여 계절 조정된 월별 추정 값을 계산한다.

$$Y'_{1-12} = \text{월별평균추정값} * \text{월별계절지수} = \text{월별추정값}$$

(표 13) 계절 조정된 월별 추정 값

구분	월별평균추정값	월별계절지수	월별추정값(Y')
Y'1	2979	0.598467	1783
Y'2	2979	0.826509	2462
Y'3	2979	1.034505	3082
Y'4	2979	1.381186	4115
Y'5	2979	1.356406	4041
Y'6	2979	1.353541	4032
Y'7	2979	1.20774	3598
Y'8	2979	1.141479	3400
Y'9	2979	1.170215	3486
Y'10	2979	0.840556	2504
Y'11	2979	0.58422	1740
Y'12	2979	0.505175	1505

따라서 (표 13)을 살펴볼 때 1992년도 1월의 추정되는 추정값은 1783, 그리고 12월의 추정되는 추정값은 1505가 된다는 것을 추정할 수 있다. 결국, 계산된 월별 추정 값을 합산하면 1992년도 연간 추정 값이 35748이라는 것을 확인할 수 있다.

5. 평가 분석

이 논문에서 제시한 과거의 추정값에 대한 추정 표준오차를 대체에너지원 데이터의 실제 측정값과 평가 분석하였다. 그리고 대체에너지원 데이터의 실제 측정값에 대한 추정 값의 정확성을 확인하기 위해 일반적으로 통계적 방법에서 많이 사용하는 95% 신뢰구간을 적용하였다.

대체에너지원 데이터에 대한 과거 측정값을 추정하는 과정에서 100% 정확하게 추정할 수는 없기 때문에 추정의 표준오차를 사용해야만 한다. 추정에 대한 표준오차는 (식 4), (식 5), 그리고 (식 6)과 같은 표현식을 통해 계산할 수 있다.

$$S_y = \sqrt{\frac{\sum Y^2 - a \sum Y - b \sum XY}{n-2}} \quad (\text{식 4})$$

혹은

$$S_y = \sqrt{\frac{\sum (Y - Y')^2}{n-2}} \quad (\text{식 5})$$

(표 9)의 춘천 지역에 대한 1992년도 과거 측정값에 대한 추정 표준오차를 계산하는 과정은 다음과 같다. 아울러, 실제 측정값의 신뢰 구간을 구하기 위해서는 n<30인 경우, t-분포표를 이용하며, n>30인 경우, 정규 분포표를 이용한다. 여기에서 n값은 5를 나타내며, (식 6)을 사용한다.

$$Y' \pm t \left(\frac{\alpha}{2}, n-2 \right) S_y \quad (\text{식 6})$$

(식 6)에서 $\alpha/2$ 는 t 분포표에서 양쪽 부분의 신뢰구간을 벗어난 확률 값, 그리고 $n-2$ 는 두 개의 자유도가 상실되는 것을 의미하는 데, 이것은 X, Y 의 두 변수가 표본치를 사용하기 때문이다. 따라서 t -분포표에서 95% 신뢰구간의 값은 자유도를 3으로 계산하고, $\alpha/2$ 를 0.025로 지정하면 3.182 값이 선택된다. 따라서 (식 4)나 (식 5)를 통해 추정에 대한 표준오차 값을 계산하면 그 값은 1315.39가 된다. 그 계산 과정은 다음과 같다.

$$S_r = \sqrt{\frac{6446320415 - (35747.5 \cdot 179459) - (48.1 \cdot 538858)}{5-2}}$$

$$= 1315.39$$

$$\alpha = 1 - 0.95 = 0.05$$

$$\text{즉, } \frac{\alpha}{2} = 0.025, \quad t^{(0.025, 3)} = 3.182$$

$$Y' \pm t^{(\frac{\alpha}{2}, n-2)} S_r = 35748 \pm (3.182)(1315.39)$$

$$= 35748 \pm 4186 (\therefore 31562 \sim 39934)$$

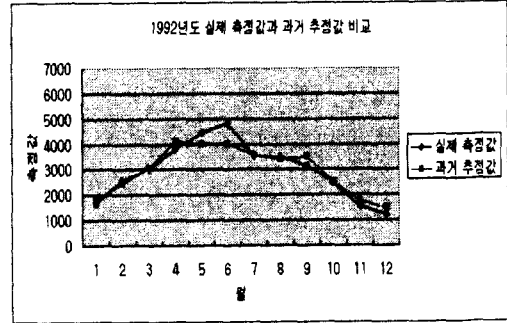
따라서 1992년도의 연간 측정값(Y'), t -분포표, 그리고 추정의 표준오차 값을 이용한 1992년도의 연간 추정 값의 95% 신뢰 구간은 35748 ± 4186 이 된다. 즉, 1992년도의 실제 연간 추정 값이 31562부터 39934 사이에 있을 확률은 95%라고 볼 수 있으며, 그 추정 값은 (표 14)와 같다.

(표 14) 1992년도 월별실제측정값과 과거추정값

구분	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
측정값	1672	2585	2998	3817	4472	4823	3555	3490	3134	2475	1538	1237
추정값	1783	2462	3082	4115	4041	4032	3598	3400	3486	2504	1740	1505
신뢰 구간	1434 2132	2113 2811	2733 3431	3766 4464	3692 4390	3683 4381	3249 3947	3051 3749	3137 3835	2155 2853	1391 2089	1156 1854

(표 14)에서 월별 신뢰구간은 월별 추정 값 ± 349 (연간 추정 표준오차 범위인 4186을 12로 나눈 값)가 된다. 이와 같은 실제 측정값과 과거 추정 값에 대한 비교를 찍은선 그래프로 표

현하면 (그림 4)와 같다.



(그림 4) 1992년도 월별 실제 측정값과 과거 추정 값(95% 신뢰구간)

(그림 4)에서 제시하는 그래프상의 의미를 분석할 때 특히 사항으로 6월과 같이 월별 실제 측정값과 과거 추정 값을 비교할 때 신뢰구간을 크게 벗어나는 경우는 대체에너지원 데이터의 계절적인 요인 중에서 우연 변동 요인이나 불규칙 변동 요인이 크게 작용한 기간으로 추정할 수 있다.

6. 결론

대체에너지원 데이터는 시간 속성을 갖는 데이터의 특성으로 인하여 매년 측정된 데이터들이 새롭게 추가되기 때문에 데이터의 양은 시간의 변화에 따라 커지게 된다. 이와 같은 문제를 해결하기 위해 일정 기간 측정된 데이터를 데이터베이스에 저장한 후, 기존의 저장된 데이터를 이용하여 대체에너지원 데이터의 과거 측정값을 추정하는 기능이 필요하다.

이 논문에서는 현재까지 국내 20개 도시에서 측정된 일사량 데이터와 33곳의 지역에서 측정된 풍력 데이터를 체계적으로 저장 및 관리하고, 과거 측정값을 추정하기 위한 대체에너지원 데이터의 추정 기법을 제안하였다. 대체에너지원 데이터에 대한 과거 측정값을 추정하기 위해 대체에너지원 데이터가 갖는 시간적인 특성을 고려한 통계적 예측 기법 중에서 시계열

분해법을 적용하였다. 과거의 측정값에 대한 추정 표준 오차를 대체에너지원 데이터의 실제 측정값과 비교 분석함으로써 추정 기법의 정확성을 평가하였다. 이러한 평가 분석 결과를 토대로 이 논문에서 제안한 통계적 예측 기법 중에서 시계열 분해법을 적용한 대체에너지원 추정 기법은 기존의 시스템에서 지원하지 못하는 과거 측정값을 추정하는 기능을 통해 다양한 서비스 제공이 가능한 장점을 제공한다.

이 논문에서 제안한 대체에너지원 데이터의 추정 기법은 일사량 및 풍력 데이터뿐만 아니라 바이오매스, 소수력, 지열, 해양 에너지 등과 같은 대체에너지원 데이터와 관련된 다양한 형태의 분석 정보를 제공할 수 있는 기본 데이터베이스 구축에 유용하게 활용될 것이다. 향후에는 데이터 마이닝 기법을 적용한 대체에너지원 데이터에 대한 예측 기법의 연구가 추가적으로 필요하다.

참고문헌

- [1] Fsec, "Automated Field Data Management and Quality Assurance", Florida Solar Energy Center, <http://logger.fsec.ucf.edu/met>, 2003.
- [2] K. H. Ryu and Y. A. Ahn, "Application of Moving Objects and Spatiotemporal Reasoning", TimeCenter TR-58, 2001.
- [3] B. A. Wielicki et al., "CERES(Clouds and the Earth's Radiant Energy System) Validation Plan Overview", Technical Document, Release 4, <http://bsrn.ethz.ch>, 2000.
- [4] L.M.Murphy, J.Brokaw, J.Pulaski, K.McCormack, "The National Alliance of Clean Energy Business Incubators", N R E L / B K - 7 2 0 - 2 8 7 2 4 , <http://www.nrel.gov>, 2000.
- [5] M. Erwig, R. H. Guting, M. Schneider and M. Vazirgiannis, "Spatio-Temporal Data Types: An Approach to Modeling and Querying Moving Objects in Databases", Chorochronos Technical Report CH-97-8, 1997.
- [6] A. C. Harvey, "Forecasting, structural time series models and the Kalman filter", Cambridge University Press, 1989.
- [7] 오인배, 안윤애, 류근호, 김광득, "일사량 에너지 데이터 관리 시스템 설계," 한국정보처리학회 논문지, 제10-D권 제3호, pp. 531-540, 2003.
- [8] 오인배, 안윤애, 김원태, 류근호, 김광득, "대체에너지원 이력 데이터 관리 시스템 설계", 한국정보처리학회 논문지, 제10-A권 제6호, pp. 757-768, 2003.
- [9] 이태규외 6명, "국내 일사량 분석·평가 및 데이터 표준화 연구", 산업자원부 최종보고서, 1999.
- [10] 양윤섭, "국내 일사량 분석·평가 및 데이터 표준화 연구", 한국 에너지기술연구원, 1998.
- [11] 김건훈외 5명, "풍력발전 유망지역의 풍력 자원 정밀조사", 한국에너지기술연구원, 1997.
- [12] 박경호, "풍력 발전 유망 지역의 풍력 자원 정밀조사(II)", 한국에너지기술연구소 연구보 고서 KIER-961125, 1996.
- [13] 김우철외 7명, "改訂版 統計學概論", 英志文化社, 1994.
- [14] 박병기, 정종식, "新 統計學", 經進社, 1991.
- [15] 김치홍, "豫測科學 - 그理論과實際 -, 成均館大學校出版部, 1985.



오 인 배

e-mail : iboh@jsc.ac.kr

1987년 한남대학교 컴퓨터공
학과(공학사)

1989년 건국대학교 대학원
컴퓨터공학과(공학석사)

2004년 충북대학교 대학원
전자계산학과(이학박사)

1989년~1992년 (株)LG-HITACHI Ltd. 해외
S/W 개발실 근무

1992년~현재 주성대학 인터넷정보과 부교수
관심분야 : 시공간 데이터베이스, 이동객체 데
이터베이스, 지리정보 시스템, 가상현실(Virtual
Reality) 등

안 윤 애

e-mail : yeahn@cjnc.ac.kr



1993년 한남대학교 컴퓨터공
학과(공학사)

1996년 충북대학교 대학원
전자계산학과(이학석사)

2003년 충북대학교 대학원
전자계산학과(이학박사)

2003년~현재 청주과학대학
컴퓨터과학과 전임강사

관심분야 : 시공간 데이터베이스, 이동객체 데
이터베이스, 위치기반 시스템, 지리정보 시스
템, 지식기반 시스템 등