일사량 데이터 관리를 위한 미래 변화 추이 예측

Estimation of Future Trend for Solar Radiation Data Management

오인배*, **안윤애****, **이봉근*****
주성대학 인터넷정보과*, 충주대학교 전기전자및정보공학부**, 충북대학교 전기전자및컴퓨터공학부***

In-Bae Oh(iboh@jsc.ac.kr)*, Yoon-Ae Ahn(yeahn@cjnu.ac.kr)**,
Bong-Keun Lee(Bong9065@hanmail.net)***

요약

일사량 데이터의 측정값은 시간의 변화에 따라 변경되는 특성을 갖기 때문에 측정되는 원시 데이터의 양이 상당히 방대해진다. 따라서 일사량 데이터의 이력 정보를 체계적으로 저장 및 관리할 수 있는 데이터베이스의 구축과 일사량 데이터의 향후 변화 추이에 관한 예측 기법의 연구가 필요하다. 이 논문에서는 일사량 데이터의 이력정보를 저장하기 위한 데이터 저장구조를 제시하고, 시계열 분해법을 적용한 일사량 데이터의 변화 추이 예측 기법을 제안한다. 아울러 국내 20개 도시에서 측정된 데이터를 토대로 한 실험 결과를 제시한다.

■ 중심어 : | 일사량 데이터 | 미래 변화 추이 |

Abstract

Measured values of solar radiation data have a characteristic that they change almost by the minute, so original data can be massive. Therefore, we need to construct a database which stores and manages history data of solar radiation data systematically. A study of an estimation method of the future change trend is also required. In this paper, we present a data structure in order to store history data of solar radiation data and propose an estimation method for the change trend of solar radiation that applies to a time-series decomposition method. Also, we present the results of experiments based on measured data from 20 domestic cities in Korea.

■ keyword : | Solar Radiation Data | Future Change Trend |

I. Introduction

A renewable energy source is a kind of useful future energy. Examples are solar energy, wind force, small hydro-power, biomass, and so on. To know all about renewable energy sources, energy sources should be observed sequentially and periodically. The observed data can be stored in a database and managed so that it can be used in the future. Renewable energy source data has a characteristic that the measured values can change very frequently related to various surroundings, especially such as the seasons [1][2]. Renewable energy sources can be classified as data that has a time characteristic. When these

접수번호 : #070814-004
접수일자 : 2007년 08월 14일
심사완료일 : 2007년 10월 22일
교신저자 : 안윤애, e-mail : yeahn@cjnu.ac.kr

characteristics are considered [3][4], a history database needs to be constructed where users can systematically store and manage massive amounts of history data representing renewable energy. Also, a study to develop an estimation method of the change trend of renewable energy data is needed.

The development and distribution of a renewable energy management system have been going ahead outside of Korea. Significant efforts are being made to make the data servable through the Internet[5-7]. There is not much recognition of the importance of renewable energy data. Currently, a system providing systematic history management of this kind of data and a data searching service through the Internet is not available yet in Korea. Also, the systems which are based on GIS such as some existing meteorological information systems or land management systems have constructed current databases and are in use. However, these systems do not provide various operations and data searching functions for the management of history data. It is not able to operate on measured renewable energy data automatically by these systems [8][9]. The data related to renewable energy can be extremely large because new data is added according to the criterion of adding changing measured data. Unfortunately, there is no effective method to manage this massive data yet. In this paper, we use a storage structure for a history database to manage solar radiation data measured from 20 cities, so far, in Korea. Temporal history database schema is appropriate to manage solar radiation data which have the temporal characteristics[3][4]. We, also, propose an estimation method for the change trend of solar radiation data by using the described model. For this method, we applied a time-series decomposition method. This method considers the time characteristic of solar radiation data using statistical estimation methods. Using this method, we show how future energy data could be effectively used by estimating the trend of changes of energy data.

This paper is composed as follows. First, in Section 2, we show characteristics of time-series data analysis which is used for the estimation of the change trend of solar radiation data. In Section 3, we describe a database structure for the history management of solar radiation data. We, then, describe an estimation method for the change trend of the data by applying time-series decomposition. In Section 4, we discuss the contents of an evaluation analysis and finally we draw conclusions in Section 5.

II. Related Work

Most general approaches among estimation methods involving statistical methods are qualitative estimation methods and quantitative methods [10-13]. Qualitative method can be decided by subjective factors like experience or opinions of internal/external experts of the organization, so it is also called subjective method or conjecturable method. A quantitative method eliminates most experts' subjective factors because they cannot be precisely measured. A quantitative method uses objective data from the past for the attributes that are desired to be estimated. That is why it is also called an objective method. These quantitative methods are divided into time-series estimations and the relations of cause and effect estimations[13-15].

A time-series in the time-series analysis method means the actual values from the past are arranged by regular time intervals in order like daily, weekly, monthly, quarterly, and yearly. A time-series is used to estimate a future change trend when there exists data for years or even a decade, or the trend is obvious and rather stable. However, time-series analysis completely depends on the data from the past, so a future change trend could be decided by the pattern that happened before. Time-series is more useful for short-term estimation than for long-term estimation. There is a premise that the real values which happen in some period are consisted of trend movement, circulating change, seasonal variation, and chance variation(irregular variation).

A time-series decomposition method is divided into trend analysis, seasonal analysis, circulating factors analysis, and chance factors analysis [10][13-15]. Trend analysis grasps what trend was in the past data. It connects every point after graphing past data with time and decides the trend of past data by using eye-measurement or a statistical method if it has a straight line or parabola. A seasonal index in seasonal factors analysis shows how the value of a time-series changes from the trend by the seasonal factors. A seasonal index is presented by percentages of the trend. There is an index for each season. There are 12 indices in the case of monthly data and 4 in quarterly data. When it is time to decide a seasonal index by using past data, the trend analysis method applies the index to the estimation value predicted to get an estimation value which is season-adjusted. Circulation change in circulation factors analysis cannot be measured by the same method as seasonal variation, but by monthly, quarterly, and yearly data. In the case of using yearly data, it is supposed to include only trend factors and circulation factors, so it is not necessary to get rid of seasonal factors or irregular variations. We can say that a chance variation in chance factors analysis happens by totally unpredictable factors. So, chance variation (irregular variation) in time-series means a variation that cannot be explained by trend movement, seasonal variation, and circulating change.

With a consideration of the time characteristic that the renewable energy source has we present an estimation for the change trend for renewable energy source data. This estimation uses a time-series decomposition method. To apply the time-series decomposition method, we need a structured history database to store measured data from the past which we also propose.

III. An Estimation of the Change Trend of Solar Radiation

Solar radiation data is measured and stored by hour, day, month, and so on, so data newly measured yearly is added newly due to character of data that has a temporal feature. That's how solar radiation data can be very large according as time changes. For the data used in this paper, we did sampling monthly data (amount of 215 pages of A4 size) among solar radiation data measured for 17 years from 1982 to 1998 from 20 areas in the country and input into the database in direct. Therefore, there is some difficulty to present the total size of solar radiation data quantitatively. Storing these massive data simply causes the result that it takes a long time to search for a storage and data. In order to solve this problem, data measured for certain period are stored in the database as a way to manage massive data. Also to propose an estimation function for changes of future energy data, we do modeling history data and propose a method that estimates past measured value or future changes by using solar radiation data stored in the past.

1. Solar Radiation Data

Data feature dealt in this paper has similar features to data applied to temporal database, so we define schema as a structure which apply valid time among temporal data modeling[3][4]. The solar radiation data we describe in this paper has basically 12 different types[1][2]. General attribute information has a different time cycle according to the type of solar radiation data. Because of this, it is not possible to store all solar radiation data with only one relation schema. The data is composed of 4 different relation schemas which cover the 12 types of solar radiation data[8][9] as shown in [Table 1].

Table 1. Type of data schema

Schema	Data Type
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

The following are 4 of the schemas that construct solar radiation data. Schema_1 stores the attribute data related to total solar radiation on a horizontal surface(ilsa), percentages of sunshine(ilcho), cloudiness(cloud), insulation(rate), the number of the clear days(clean), temperature(temp), relative humidity(humi). It has the same structure as [Table 2].

Table 2. 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

[Table 2] shows the attribute values stored in the attribute history relation of schema_1. The effective time cycle of schema_1 uses a unit of a month. Schema_2 stores the attribute data related to solar radiation out of atmosphere. It has a structure such as indicated in [Table 3].

Table 3. schema_2

vtstart	vtend	out_ilsa
1997-01-01	1997-01-01	3571.7

[Table 3] shows the attribute values stored in the attribute history relation of schema_2. The effective time cycle of schema_2 uses a unit of a day. Schema_3 stores the attribute data related to solar radiation for clear days(clean_radi), and elements of total solar radiation(g_radi_element). It has a structure as shown in [Table 4].

Table 4. 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

[Table 4] shows the attribute values stored in the attribute history relation of schema_3. The effective time cycle of schema_3 uses a unit of an hour. Schema_4 stores the attribute data related to azimuth slope radiation(slope_radi). It has a structure as shown in [Table 5]. Azimuth slope radiation data is divided into 6 azimuths of east(E), west(W), south(S), north(N), southeast(SE), and northeast(NE).

Table 5. schema_4

vtstart	vtend	angle	orient	slope_ilsa
1997-01-01	1997-01-31	0	E	1746

2. The Trend Analysis of Future Estimation Value

Some related researches focus on storing and managing (offering a searching service) measured data mostly, but don't offer an estimation function [13-15][21][22]. Therefore, we consider temporal feature that solar radiation data has and propose a method that manages history information systematically by estimating future changes of solar radiation data by using trend analysis and seasonal element analysis of time-series decomposition method.

In Section 3.1, we explain that the solar radiation data is stored in a database by regular time intervals such as monthly, quarterly, and yearly. The stored data exists for years and the trend is comparatively obvious. This solar radiation data has the characteristic of data that is used in time-series analysis[16-18]. Therefore, the solar radiation data is suitable for prediction of the future change trend by a time-series decomposition method.

In order to estimate the change trend for future solar radiation data, we describe the change trend estimation method of future solar radiation data through a monthly estimation function by using trend analysis and seasonal elements analysis of a time-series decomposition method from time-series data analysis. The process of trend analysis of solar radiation data can be explained briefly as follows in [Fig. 1].

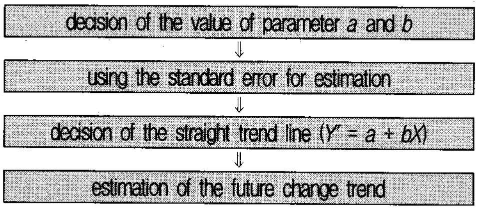


Fig. 1. An outline diagram for trend analysis of future estimation value

To know what trend is in the measured value from the past, measured values (Y-axis) from the past are marked on the graph where times are the horizontal axis (X-axis). After connecting each dot, if it uses an eye-measurement or statistical method, it can decide whether trend of measured value is a form of straight line or index or parabola. If the trend of solar radiation data is a straight line form, straight trend line ($Y'=a+bX$) should be found. If it is an index form, then the trend line of an index function form, ($Y'=abX$), should be found, or the trend line of secondary function form ($Y'=a+bX+cX^2$) should be found if the trend line is in the form of a parabola[16-18].

The trend of each dot (degree of scattering or scatter diagram) presented past values of solar radiation data on the graph in this paper has a straight line form such as [Fig. 2]. A future change trend can be estimated by extending the upper part of the straight line after the straight trend line is found through trend analysis of measured values from the past. When a trend line is straight, that means there is a steady increase (or decrease) of measured average value of each period. The expression of a straight trend line (Y′) is as follows.

$$Y' = a + bX \tag{1}$$

Y' = a measured value of the special period
a = intercept of the Y-axis when X is 0
b = gradient of the straight line
X = 1, 2, 3, 4, 5, ...
(for example, a number of the year)

The most important elements that decide a straight trend line are a and b. Parameter a and b should be decided by using values from the past; to do so, the method of least squares should be used[19][20]. The method of least squares decides a trend line to minimize the total sum of the deviation squares between each of the actual values and corresponded estimation values. The expressions that calculate parameters a and b, are as follows.

$$a = \frac{\sum Y - b\sum X}{n} \tag{2}$$

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

Table 6. The percentages of sunshine in the Daejeon area

Month Year	1	2	3	4	5	6	7	8	9	10	11	12	Sum
1993	474	564	514	575	534	357	325	293	605	615	347	505	5708
1994	566	575	538	551	514	516	605	606	643	561	601	556	6832
1995	634	696	488	630	540	538	329	500	465	589	677	566	6652
1996	612	670	496	658	630	271	520	491	619	576	451	646	6640
1997	583	695	662	592	500	581	379	546	660	710	535	495	6938

We, in our work described in this paper, randomly selected measured values of five years' period (1993 ~ 1997) out of actual measured values of a 17 years' period (1982 ~ 1998) of solar radiation data and used as a sample data. [Table 6] shows the actual data ($kcal/m^2$) of percentages of sunshine for 5 years in the Daejeon area.

A scatter diagram that shows the actual data presented in [Table 6] as a graph is the same as in [Fig. 2]. If we see the data, it shows a straight trend line.

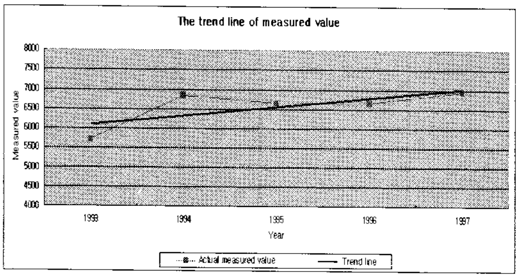


Fig. 2. The trend line of the Daejeon area

[Table 7] is a variable to calculate parameters, *a* and *b*, *Y*(*TCSR*) presents the real data and *XY* shows the actual data (*Y*-axis) for time (*X*-axis). *X* shows a number for the year (*X*-axis).

Table 7. The measured value for the estimation of the trend line

Year	X	Y=TCSR	XY	X^2	Y^2
1993	1	5708	5708	1	32581264
1994	2	6832	13664	4	46676224
1995	3	6652	19956	9	44249104
1996	4	6640	26560	16	44089600
1997	5	6938	34690	25	48135844
Sum	15	32770	100578	55	215732036

* *T*(Trend fluctuation), *C*(Cyclical fluctuation), *S*(Seasonal fluctuation), *R*(Random fluctuation)

The process of calculating parameters, *a* and *b*, through [Expression 2], [Expression 3], and [Table 7] is as follows. The straight trend line is presented as $Y'=a+bX$. We calculate the trend line by using the data of 5 years, so the value of *n* is 5. *X* designates the values as a standard of the starting year (*X=1, 2, 3, 4, 5, 6*).

$$a=\frac{32770-226.8(15)}{5}=5873.6$$

$$b=\frac{5(100578)-15(32770)}{5(55)-(15)^2}=226.8$$

By using *a* and *b*, we can calculate a yearly estimate of the measured values for the past 5 years and an estimate of the future change trend for the next year (1998) by reference to [Expression 1].

$$Y'_i = 5873.6 + 226.8X \ (X = 1,2,3,4,5,6,...)$$

$$Y'_{1993} = 5873.6 + 226.8(1) \fallingdotseq 6100$$

$$Y'_{1994} = 5873.6 + 226.8(2) \fallingdotseq 6327$$

$$Y'_{1995} = 5873.6 + 226.8(3) \fallingdotseq 6554$$

$$Y'_{1996} = 5873.6 + 226.8(4) \fallingdotseq 6781$$

$$Y'_{1997} = 5873.6 + 226.8(5) \fallingdotseq 7008$$

$$Y'_{1998} = 5873.6 + 226.8(6) \fallingdotseq 7234$$

(annual future estimation value)

Through this, a yearly estimate for the past 5 years, the actual measured value, and a yearly estimate of 1998, were finally decided and are the same as [Table 8]. The yearly estimate is useful in deciding a monthly estimate for future solar radiation data.

Table 8. The actual measured value and future estimation value

Year	X	Actual measured value (Y=TCSR)	Future estimation value(T)
1993	1	5708	6100
1994	2	6832	6327
1995	3	6652	6554
1996	4	6640	6781
1997	5	6938	7008
1998	6	–	7234

In [Table 8], we can see that the actual yearly measured value in 1998 could be more or less than 7234 through [Fig. 2]. The measured value is related to the many dots out of the trend line in [Fig. 2]. If all dots were on the trend line and the actual measured value was 7234, there would not be any error in the estimation. However, it is not possible to estimate 100% for sure, so the estimated standard error has to be used to measure variance from the trend line for the real measured value.

3. A Monthly Estimation of Future Data

[Fig. 2] is a presentation of measured values of each year with a trend line by using the measured values of each month of [Table 6]. The measured values used to decide a trend line include seasonal factors, so if you want to estimate a future change trend based on this trend line, you have to adjust these estimates up and down. The process of computing seasonal indices by using the percentages of the centered moving average method is described with an example in [Table 9] which uses the monthly solar radiation data of [Table 6].

Table 9. Specific monthly seasonal index

Year	Month	Measured value (Y=TCSR)	Monthly moving average	Centered moving average(TC)	Monthly seasonal index(SR)
1993	1	474			
	2	564			
	3	514			
	4	575			
	5	534			
	6	357			
	7	325	475.6667	479.5	0.677789
	8	293	483.3333	483.7917	0.605633
	9	605	484.25	485.25	1.24678
	10	615	486.25	485.25	1.267388
	11	347	484.25	483.4167	0.717807
	12	505	482.5833	489.2083	1.03228
1994	1	566	495.8333	507.5	1.115271
	2	575	519.1669	532.2083	1.080404
	…	…	…	…	…

The procedure of percentages of centered moving average method to estimate future monthly estimates by using measured values from the past is done with 6 steps as follows.

- Step 1 involves computing the monthly moving average: After computing the monthly moving average measured solar radiation for the last 5 years, we locate it in the center of 12 months' intervals. For instance, we designate a value, (474 + ⋯ + 347 + 505) ÷ 12 = 475.6667, as July, 1993. By repeating this again and again, the monthly moving average of [Table 9] can be computed.

- Step 2 involves computing the monthly centered moving average: By using two of monthly moving averages, we compute centered moving average in the same way as Step 1 and set it between two monthly moving averages. For instance, we designate a value, CMA_1 = (475.6667 + 483.3333) ÷ 2 = 479.5, as July, 1993. The value 479.5 means the adjusted monthly measured value. In this way, we compute the value of centered moving average row of [Table 9].

- Step 3 involves computing the specific seasonal index of each month: To separate the effect of the seasonal variation for each month, we divide the actual measured value into a centered moving average which corresponds to the value and compute. For instance, specific monthly indices of September and October in 1993 can be computed by the process as follows. Here $(SR)_{9,1993}$ shows the specific seasonal index of September, 1993.

$$(SR)_{9,1993} = \frac{Y_{9,1993}}{CMA_1} = \frac{TCSR}{TC} = 1.24678$$

$$(SR)_{10,1993} = \frac{Y_{10,1993}}{CMA_2} = \frac{TCSR}{TC} = 1.267388$$

The computed result through this process is a specific seasonal index which is except centered moving average from the actual measured value and shows a value of specific seasonal index row of [Table 9].

The processing step 1 to step 3 for the ratio to centered moving average method which computes the monthly future estimation value using the past measured value is as follows [Algorithm 1].

Algorithm 1. Ratio to centered moving average method(Step 1 ~ Step 3)

```
Algorithm Pre_EstimationValue(Y, Y′, n)
Input : Y(Real measured value), Y′(Annual estimation value),
        n(Number of measured year)
Output : SR(Specific seasonal index)
Begin
  For i = 1, 12*(n-1)+1
    k1, k2 ← i+7;
    For j = k1, i+12
      M_M_A[k1] ← sum(Y[j]) / 12;
      // Computation of monthly moving average
    End For
  End For
  For i = k2, 12*n
    TC[i] ← (M_M_A[i] + M_M_A[i+1]) / 2;
    // Computation of centered moving average
    SR[i] ← Y/TC[i];
    //Computation of a specific seasonal index
  End For
  Return SR; //Return the specific seasonal index
End
```

We input to the Pre_EstimationValue module of the [Algorithm 1] a real measured value and then compute the monthly moving average of the measured data. Next this module computes the centered moving average using the two monthly moving averages. Finally this module computes a specific seasonal index to remove the effect about the seasonal variation. At this time, the chance variation which is included in the result value must be excluded and the monthly estimation value is calculated.

- Step 4 involves the average processing of a specific seasonal index for each month by year: We get rid of the chance variation by taking an average specific seasonal index of each month by year and separate seasonal effect. For instance, seasonal indices of September of 1993 (S_9) are computed by the process as follows. Here the value in denominator is the value of finding an average of 4 years between 1993~1996 and refers to [Table 10].

$$S_9 = \frac{((SR)_{9,1993} + (SR)_{9,1994} + (SR)_{9,1995} + (SR)_{9,1996})}{4}$$

$$= \frac{(1.24678 + 1.102916 + 0.844431 + 1.105522)}{4}$$

$$= \frac{4.299649}{4} = 1.074912$$

The seasonal index for each month to be decided is the same as [Table 10]. What you have to consider is that the sum of seasonal indices of each month should be the value 12 exactly nonetheless, it could be bigger or smaller than 12. Therefore, you need to adjust the seasonal index of each month up and down. Here, it is smaller than 12, so it has to be adjusted upward. For instance, the seasonal index of July can be computed by the expression as follows. The adjusted averages are in [Table 10]. This presents the seasonal index for each month.

We have already explained that the yearly measured value in 1998 of the estimated solar radiation data is 36036 by using a trend line. We use this value and compute the measured value of each month in 1998. The process to do so is as follows.

- Step 5 involves computing the monthly average measured value with adjusting seasonal factors: The

Table 10. Monthly seasonal index

Year \ Month	1	2	3	4	5	6	7	8	9	10	11	12	Sum
1993							0.677789	0.605633	1.24678	1.267388	0.717807	1.03228	
1994	1.115271	1.080404	0.983846	1.00885	0.926961	0.909719	1.057384	1.044753	1.102916	0.960274	1.021096	0.941442	
1995	1.093103	1.23386	0.883658	1.153846	0.981224	0.971265	0.594489	0.906755	0.844431	1.066707	1.215258	1.029637	
1996	1.119768	1.209113	0.885451	1.162459	1.132924	0.492727	0.941816	0.88956	1.105522	1.021126	0.811273	1.146576	
1997	1.021985	1.22602	1.159623	1.023926	0.851426	0.994083	0.68427						
Sum	4.350127	4.749397	3.912578	4.349081	3.892535	3.367794	3.271478	3.446701	4.299649	4.315495	3.765434	4.149935	
Mean	1.087532	1.187349	0.978145	1.08727	0.973134	0.841949	0.8178695	0.861675	1.074912	1.078874	0.941359	1.037484	11.9676
Modified Mean	1.090481	1.190569	0.980797	1.090218	0.975772	0.844231	0.8200871	0.864012	1.077827	1.081799	0.943911	1.040297	12

monthly average measured value with the adjusting seasonal value is computed by dividing 7234, yearly measured value in 1998, into 12. That is, 7234/12 = 602.8.

- Step 6 involves computing the monthly estimate with adjusting seasonal factors: By using the monthly seasonal index, the monthly estimate with the adjusting seasonal factor is computed. The computing process applies to rules as below. The final result is the same as [Table 11].

Table 11. The monthly estimation value

Section	Monthly mean value	Monthly seasonal index	Monthly estimation value(Y')
Y'_1	602.8	1.090481	657
Y'_2	602.8	1.190569	718
Y'_3	602.8	0.980797	591
Y'_4	602.8	1.090218	657
Y'_5	602.8	0.975772	588
Y'_6	602.8	0.844231	509
Y'_7	602.8	0.8200871	495
Y'_8	602.8	0.864012	521
Y'_9	602.8	1.077827	650
Y'_{10}	602.8	1.081799	652
Y'_{11}	602.8	0.943911	569
Y'_{12}	602.8	1.040297	627

As seen in [Table 11], the estimate of January in 1998 is 657, and the one of February, 1998 is 718. The one for December, 1998 is 627. So, we can confirm that the yearly estimate for the year 1998 is 7234 when adding all monthly estimates.

The processing of step 4 to step 6 for the ratio to centered moving average method which computes the monthly future estimation value using the past measured value is as follows [Algorithm 2].

Algorithm 2. Ratio to centered moving average method(Step 4 ~ Step 6)

```
Algorithm Next_EstimationValue(SR,EstimationValue,n)
Input : SR(Specific seasonal index),
        EstimationValue(Annual estimation value),
        n(Number of measured year)
Output : Y'(Monthly estimation value)
Begin
  For i = 1, 12
    For j = 0, n-1
      SS = sum( SR[i+j*12] )
    End For
    S[i] = SS / n;//Monthly seasonal index average
    SS' = sum(S[i]);
    //Sum of monthly seasonal index average
  End For
  For i = 1, 12
    S'[i] = S[i]*12 / SS';
    //Monthly seasonal index modified average
  End For
  M_A = EstimationValue/12;
  //Monthly average measured value
  For i = 1, 12
    Y'[i] = M_A * S'[i]; //Monthly estimation value
  End For
  Return Y'; //Return monthly estimation value
End
```

The Next_EstimationValue module of [Algorithm 2] calculates the annual average value of the specific seasonal index. Also we can compute by the use of the monthly average measured value and the monthly seasonal index, the monthly estimation value modified by season.

IV. Evaluation Analysis

The estimation standard error of the future change trend estimation method by applying the time-series decomposition method has been evaluated and analyzed by comparing to the actual measured values of solar radiation data. The solar radiation data used in the evaluation analysis is the real measured data from 20 cities in Korea. In order to confirm the accuracy of the estimates for the actual measured values, a 95% confidence interval, which is generally and widely used in statistical methodologies, was applied. In the process of estimating the future trend change for solar radiation data, it is not possible to have 100% accuracy, for sure, so an estimated standard error has to be used. The standard error for estimation[16][17] can be computed by an expression such as [Expression 4] and to compute the standard error for the estimation, [Table 8], presented in Section 3.2, is used.

$$S_Y = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}} \qquad (4)$$

The computing process for estimation standard error for the future change trend in 1998 about Daejeon area in [Table 8] is as follows. In order to get a confidence interval for the actual measured value, if $n<30$, a t-distribution table is used. On the other hand, if $n>30$, a normal distribution table is used. Here, the value of n is 6 and [Expression 5] is used.

$$Y' \pm t^{(\frac{\alpha}{2}, n-2)} S_Y \qquad (5)$$

In [Expression 5] $\frac{\alpha}{2}$ means probability value which is out of the confidence interval of both sides in the t-distribution table, and n-2 means two of the degrees of freedom are lost. This is because two variables of X and Y use the sampling value. Therefore, the value of 95% of confidence interval on the t-distribution table computes the degree of freedom as 3 and the value 3.182 will be chosen if $\frac{\alpha}{2}$ is pointed 0.025. If the computing standard error (S_Y) for the estimation through [Expression 4], the value comes 1315.39. The computing process is as follows.

$$S_Y = \sqrt{\frac{215732036-(5873.6*32770)-(226.8*100578)}{5-2}}$$

$$= 665.6$$

$$\alpha = 1 - 0.95 = 0.05 \qquad \text{that is } \frac{\alpha}{2} = 0.025$$

$$t^{(0.025,3)} = 3.182$$

$$Y' \pm t^{(\frac{\alpha}{2}, n-2)} S_Y = 7234 \pm (3.182)(665.6)$$

$$= 7234 \pm 2118$$

(confidence interval : 5116 ~ 9352)

By using the yearly measured value (Y') of 1998, a t-distribution table, and standard error of estimation, the 95% confidence interval of the yearly estimation of 1998 becomes 36036 ± 4186. The probability of the actual yearly measured value of 1998 that should be between 31850 and 40222 could be 95% and the estimates are in [Table 12].

Table 12. The monthly actual measured value and the estimation value

Month / Section	1	2	3	4	5	6	7	8	9	10	11	12
Measured value	469	523	592	428	463	330	323	250	489	565	632	629
Estimation value	657	718	591	657	588	509	495	521	650	652	569	627
Confidence interval	480 ~ 834	541 ~ 895	414 ~ 768	480 ~ 834	411 ~ 765	332 ~ 686	318 ~ 672	344 ~ 698	473 ~ 827	475 ~ 829	392 ~ 746	450 ~ 804

The monthly confidence interval in [Table 12] becomes the monthly estimated value ± 349 which is the value divided 4186, the yearly estimated standard error range, into 12. If the comparison of the actual measured value and estimate can be presented by a graph of broken line, it will be the same as in [Fig. 3].

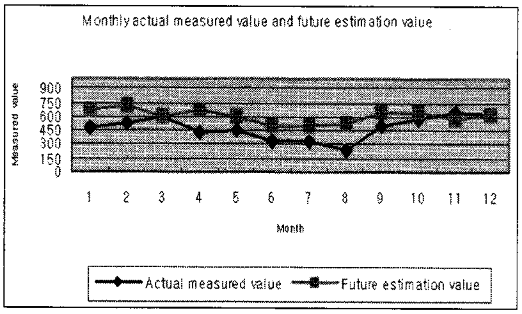


Fig. 3. The monthly actual measured value and the estimation value

V. Conclusions

Solar radiation data has a time attribute. So, the amount of data becomes larger as time goes by. In order to solve this problem of handling this continually increasing amount of data, the data measured for some time in the past should be stored in a database and there should be a function that estimates the change trend of this solar radiation data by using this past data.

In this paper, we proposed how to store and manage this solar radiation data by using data measured in 20 cities, so far, in Korea. We considered a temporal property of the solar radiation data, then used a data schema with valid time of the temporal database. Temporal history database schema is appropriate to manage solar radiation data which have the temporal characteristics. We proposed a method that estimates the future change trend by using data measured from the past. To estimate the change trend of solar radiation data, we applied a time-series decomposition method.

Also the accuracy of the estimating method has been evaluated by a comparative analysis of the estimated standard error for the change trend estimating result and the actual measured data. We know that by using the estimation method, proposed in this paper, using solar radiation data with an application of a time-series decomposition method, one of many statistical estimation methods, we should be able to effectively use the energy source data that was observed for decades and stored in a history database.

The estimation method for solar radiation data, proposed by this paper, could be useful in constructing a fundamental database which enables us to provide analysis information not only of solar radiation data, but also of various types related to renewable energy sources such as the wind force, biomass, small hydro-power, geothermal energy, ocean energy, etc. In the near future, we hope to have underway more studies of this estimation method of renewable energy source data as a data mining technique.

Also, we plan to further development of an estimation algorithm technique which considers time complexity so that data can be processed with more fast speed. However, there is no start of comparative study between estimation technique presented in this paper and other techniques yet. Hopefully having a comparative analysis with other techniques, we can draw a better way.

References

[1] T. K. Lee. et al., "Study of Solar Radiation and Standardization of Engineering Insolation Data in Korea(II)," Ministry of Commerce Industry and Energy Final Report, 1999.

[2] Y. S. Yang, "Study of Solar Radiation and Standardization of Engineering Insolation Data in Korea," Korea Institute of Energy Research, 1998.

[3] K. H. Ryu and Y. A. Ahn, "Application of Moving Objects and Spatio-temporal Reasoning," Time Center TR-58, 2001.

[4] 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.

[5] Fsec, "Automated Field Data Management and Quality Assurance," Florida Solar Energy Center, 2003.

[6] B. A. Wielicki et al., "CERES(Clouds and the Earth's Radiant Energy System) Validation Plan Overview," Technical Document, Release 4, 2000.

[7] L. M. Murphy, J. Brokaw, J. Pulaski, and K. McComack, "The National Alliance of Clean Energy Business Incubators," NREL/K-720-28724, 2000.

[8] I. B. Oh, Y. A. Ahn, K. H. Ryu, and K. D. Kim, "Design of Solar Radiation Energy Data Management System," Journal of Korea Information Processing Society, Vol.10-D, No.3, pp.531-540, 2003.

[9] I. B. Oh, Y. A. Ahn, W. T. Kim, K. H. Ryu, and K. D. Kim, "Design of a History Data Management System for the Renewable Energy Resources," Journal of Korea Information Processing Society, Vol.10-A, No.6, pp.757-768, 2003.

[10] A. C. Harvey, "Forecasting, structural time series models and the Kalman filter," Cambridge University Press, 1989.

[11] A. C. Harvey, "Time Series Models," Philip Allan, 1981.

[12] D. C. Montgomery and L. A. Johnson, "Forecasting and Time Series Analysis," McGraw-Hill, 1976.

[13] B. A. Wielicki et al., "CERES(Clouds and the Earth's Radiant Energy System) Validation Plan Overview," Technical Document, Release 4, 2006.

[14] D. Anderson, "The Meteorological Service of Canada," Annual Report, pp.4-25, 2001.

[15] G. O. P. Obasi, "Reducing Vulnerability to Weather and Climate Extremes," Switzerland, World Meteorological Organization, 2002.

[16] C. H. Kim, "Predicting Science - Theory and Practice," Sungkyunkwan University Press, 1985.

[17] W. C. Kim, et al., "Revised Edition - Introduction of Statistics," Youngjimunwhasa, 1994.

[18] B. G. Park and J. S. Jung, "New Statistics," Kyungjinsa, 1991.

[19] S. H. Lim, J. R. Choi, and C. D. Kim, "Product·Operation Management," Samyoungsa, 2003.

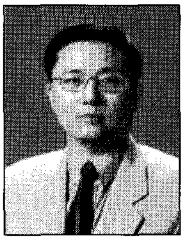
[20] B. T. Kim, S. N. Kim, and G. H. Cha, "Product Management Theory of Global Era," Myungkyungsa, 1999.

[21] K. D. Kim and J. H. Jeong, "Information Strategy Planning for GIS based Management System Development with New Renewable Energy Resource Information," Proceedings of the Korean Society of Remote Sensing, 2005.

[22] J. H. Hwang, Y. J. Jeong, Y. A. Ahn, K. H. Ryu, K. D. Kim, and S. G. Ahn, "Spatio-temporal Representation and 3D Visualization for Renewable Energy Data," Proceedings of the Korean Association of Geographic Information Studies, pp.216-225, 2001.

저 자 소 개

오 인 배(In-Bae Oh) 정회원



- 1987년 2월 : 한남대학교 컴퓨터공학과 (공학사)
- 1989년 2월 : 건국대학교 컴퓨터공학과 (공학석사)
- 2004년 2월 : 충북대학교 전자계산학과 (이학박사)
- 1989년 ~ 1992년 : (주)LG-히다찌 해외 S/W 개발실 근무
- 1992년 3월 ~ 현재 : 주성대학 인터넷정보과 부교수

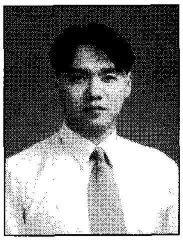
<관심분야> : 시공간 데이터베이스, 모바일 데이터베이스, 이동객체 데이터베이스, 가상현실

안 윤 애(Yoon-Ae Ahn) 정회원

- 1992년 2월 : 한남대학교 컴퓨터공학과 (공학사)
- 1996년 2월 : 충북대학교 전자계산학과 (이학석사)
- 2003년 2월 : 충북대학교 전자계산학과 (이학박사)
- 2003년 3월 ~ 2006년 2월 : 청주과학대학 컴퓨터과학과 조교수
- 2006년 3월 ~ 현재 : 충주대학교 전기전자 및 정보공학부 조교수

<관심분야> : 시공간 데이터베이스, LBS, 상황인식, 모바일 S/W, 임베디드 응용

이 봉 근(Bong-Keun Lee) 정회원



- 1997년 2월 : 한남대학교 컴퓨터공학과 (공학사)
- 1999년 2월 : 한남대학교 컴퓨터공학과 (공학석사)
- 2000년 3월 ~ 현재 : 충북대학교 전기전자 컴퓨터공학부 박사과정

<관심분야> : 시공간 데이터베이스, 데이터마이닝 및 스트림 처리, 데이터베이스 보안, 상황인식