

# 대용량 학습 데이터를 갖는 태양광 발전 시스템의 확률론적 모델링

## Probabilistic Modeling of Photovoltaic Power Systems with Big Learning Data Sets

조현철\*† · 정영진\*

Hyun Cheol Cho\*† and Young Jin Jung

\*울산과학기술대학교 전기전자공학부

† School of Electrical and Electronic Engineering, Ulsan College

### 요 약

태양광 발전 시스템의 해석적 모델링은 시스템의 동특성을 예측하거나 고장검출 및 진단 등과 같은 고급 공학 기술에 중요하게 적용할 수 있어 최근 많은 각광을 받고 있다. 본 논문은 대용량 학습 데이터를 갖는 태양광 발전 시스템에 대한 확률론적 모델링을 제시한다. 우선 태양광 일사량과 온도 입력 변수에 대한 태양광 시스템의 출력 전력과의 입출력 함수관계를 정의한다. 이 함수관계를 바탕으로 세 확률변수(일사량, 온도, 전력)에 대하여 조건부 확률 식으로 표현한다. 조건부 확률 분포 추정치는 대용량 데이터 시스템에 적합한, 전체 표본 데이터 수 대비 관련 변수의 경우의 수에 대한 비율로 나타내었다. 추정한 확률분포를 통해 평균값 이론을 적용하여 시스템의 출력을 추정하게 된다. 본 논문에서 제안한 모델링 기법은 두 태양광 발전 단지의 사례 연구를 통해 성능을 검증하였다.

**키워드** : 태양광 발전 시스템, 확률 통계, 모델링, 온라인 학습, 파라미터 추정

### Abstract

Analytical modeling of photovoltaic power systems has been receiving significant attentions in recent years in that it is easy to apply for prediction of its dynamics and fault detection and diagnosis in advanced engineering technologies. This paper presents a novel probabilistic modeling approach for such power systems with a big data sequence. Firstly, we express input/output function of photovoltaic power systems in which solar irradiation and ambient temperature are regarded as input variable and electric power is output variable respectively. Based on this functional relationship, conditional probability for these three random variables(such as irradiation, temperature, and electric power) is mathematically defined and its estimation is accomplished from ratio of numbers of all sample data to numbers of cases related to two input variables, which is efficient in particular for a big data sequence of photovoltaic powers systems. Lastly, we predict the output values from a probabilistic model of photovoltaic power systems by using the expectation theory. Two case studies are carried out for testing reliability of the proposed modeling methodology in this paper.

**Key Words** : Photovoltaic Power Systems, Probabilistic Statistics, Modeling, Online Learning, Parameter Estimation

## 1. 서 론

접수일자: 2013년 8월 14일

심사(수정)일자: 2013년 9월 9일

게재확정일자 : 2013년 9월 16일

† Corresponding author

본 논문은 중소기업청에서 지원하는 2013년도 산학연 협력 기업부설연구소 지원사업(No.C0016179)의 연구수행으로 인한 결과물임을 밝힙니다.

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

시스템 모델링은 실제 시스템의 입출력 함수 관계를 수학적 또는 해석적 방법을 통해 표현하는 기법을 말한다. 이러한 기술은 제어시스템, 고장검출 및 진단 등과 같은 고급 응용기술에 적용되고 있으며 향후에도 이러한 분야의 기술 개발은 계속될 것으로 전망된다. 특히 동적 시스템의 고장 검출이나 진단 기술과 같은 응용분야에서 시스템 모델링은 아주 필수적인 기술부분으로서 전체 기술의 성능을 크게 좌우하기도 한다.

태양광 발전 시스템은 태양광으로부터 전력을 생산하는 신재생 에너지 시스템으로 현재 각광을 받고 있다. 많은 국가에서 태양광 시스템에 관련된 다양한 분야에 기술개발을 실시하고 있으며, 고효율의 태양광 셀, 고성능 전력변환장치, 고장검출 및 고장위치 진단을 포함한 고급 모니터링 시스템 등과 같은 것들을 들 수 있다. 앞서 언급한 바와 같이

태양광 시스템의 고장진단 및 검출에서 모델링은 진단 성능과 직접적으로 연관되는 매우 중요한 기술로서 간주된다.

태양광 발전 시스템 모델링은 태양의 일사량, 태양광 셀의 표면 온도 등과 같은 주어진 입력 변수에 대하여 전력과 같은 발전 시스템의 출력을 추정하는 과정을 일컫는다 [1,2]. 이러한 연구에는 일반적으로 미분방정식과 같은 전통적인 수학적 도구를 이용하여 태양광 발전 시스템의 동특성을 표현하는 방법이 있으며, 신경회로망이나 퍼지 추론 기법 등과 같은 소프트 연산 방법을 동원하여 태양광 시스템의 비선형 모델링을 실시한 예도 있다[3,4,5]. 특히 최근에는 확률 및 통계학 이론을 적용하여 태양광 시스템의 확률론적 모델을 개발한 논문도 발표되고 있다[6]. 이러한 확률론적 태양광 시스템 모델은 입력 변수를 확률변수로 간주하여 확률통계의 평균 값 이론을 적용하여 출력 값을 추정하는 방식으로 모델링을 실시하였다. 따라서 복잡한 확률통계의 수학적 이론을 이용하였으므로 해석적 모델로는 성능이 우수하다 할지라도 실시간 구현이나 다른 응용기술에 접목하기에는 복잡한 공학적 단계를 거치는 단점이 있다.

본 논문은 실시간 구현에 매우 효과적이며 상당히 많은 학습 데이터를 포함하고 있는 태양광 발전 시스템의 새로운 확률 모델링 기법을 제안한다. 우선 태양광 시스템의 입출력 확률변수를 정의한 후 이러한 변수에 대한 조건부 확률식을 표현한다. 이 입출력 조건부 확률은 기존의 확률 이론을 통해 결합(joint) 확률 식으로 다시 표현된다. 결합 확률 분포 추정은 전체 데이터 수에 대하여 해당 변수에 대한 경우의 수의 비율로 연산된다. 마지막으로, 확률분포에 대한 평균값 이론을 적용하여 태양광 시스템의 출력 값을 추정하게 된다. 본 논문에서 제안한 태양광 시스템의 모델링 기법에 대한 타당성 검증을 위하여, 휴엔스 사에 설치되어 있는 3KW급 태양광 발전 시스템에 본 논문에서 제안한 모델링 기술을 접목하여 출력을 추정하였으며, 실제 출력 값과의 오차를 비교 및 분석을 실시하였다.

## 2. 태양광 발전 시스템 모델

태양광 발전 시스템은 태양 에너지로부터 전기적 출력을 생성하는 신재생 에너지 시스템으로서 산업체 또는 연구소에서 많은 기술개발이 이루어지고 있다. 일반적으로 태양광 발전 시스템의 출력은 태양의 일사량과 태양광 모듈의 주위 온도와 같은 외적 환경적 변수에 의해 그 크기가 결정된다. 이러한 개념은 그림 1에서 나타낸 바와 같이, 태양의 일사량과 온도를 태양광 시스템의 입력 변수로, 생산 전력을 출력 변수로 정의할 수 있는 함수관계로 표현할 수 있다.



그림 1. 태양광 발전 시스템의 입출력 함수관계

Fig. 1. Functional relationship of the photovoltaic power systems

한편, 태양의 일사량과 태양광 모듈의 주위 온도는 결정론적(deterministic) 변수가 아닌 확률(stochastic) 변수로 간

주할 수 있다. 따라서 태양광 발전 시스템의 출력이 이 두 확률 입력 변수에 대한 조건부 확률(conditional probability)을 갖는 출력 변수로 나타낼 수 있다. 우선, 이 확률 변수를 정의하기 위하여 일사량 변수를  $G$ , 온도 변수를  $T$ , 출력 변수를  $X$ 로 놓으면 조건부 확률은  $p(X|G, T)$ 와 같이 주어진다. 여기서 이 세 확률변수를 이산(discrete) 변수로 정의하기 위하여 표본 공간(sample space)을 다음과 같이 가정한다.

$$X = \{x_1, x_2, \dots, x_n\}, G = \{g_1, g_2, \dots, g_m\}, T = \{t_1, t_2, \dots, t_l\} \quad (1)$$

여기서 조건부 확률  $p(X|G, T)$ 은 잘 알려진 베이시안(Bayesian) 이론[7]을 이용하면 다음과 같이 다시 나타낼 수 있다.

$$p(X|G, T) = \frac{p(G, T|X)p(X)}{p(G, T)} = \frac{p(X, G, T)}{p(G, T)} \quad (2)$$

식 (2)의 분모항인  $p(G, T)$ 는 일사량  $G$ 와 온도  $T$ 의 결합(joint) 확률을 의미하며 확률 공리(axiom)에 의해 다음과 같은 조건을 갖는다.

$$\sum_{i=1}^m \sum_{j=1}^l p(g_i, t_j) = 1 \quad (3)$$

또한 식 (2)의 분자항인  $p(X, G, T)$ 도 마찬가지로 일사량  $G$ , 온도  $T$  및 출력  $X$ 의 세 가지 확률 변수에 대한 결합 확률을 나타내며 마찬가지로

$$\sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^l p(x_i, g_j, t_k) = 1 \quad (4)$$

와 같은 조건이 주어진다.

## 3. 확률분포추정

식 (2)의 조건부 확률  $p(X|G, T)$ 을 구하기 위하여 분자항과 분모항의 결합 확률 분포를 각각 알아야 한다. 태양광 발전 시스템처럼 학습 데이터의 양이 많은 경우 확률분포 추정기법을 이용하여 확률 분포 값을 결정할 수 있다. 현재까지 다양한 목적에 걸맞은 여러 가지 확률분포 추정 알고리즘이 발표되었으며 여러 공학적 및 과학적 분야에 성공적으로 적용되어 오고 있다[8]. 본 논문은 가장 일반적이며 간단한 방법인 전체 표본 수에 대한 경우의 수를 연산하여 확률 분포 값을 구하는 방법론을 이용한다. 이러한 기법은 연산 알고리즘이 매우 간단하며 가장 잘 알려진 가우시안(Gaussian) 등과 같은 특정한 확률분포함수를 사전에 지정하지 않기 때문에 실 데이터 적용에 매우 효과적이다. 특히 대용량 학습 데이터의 확률 분포 추정에 매우 유리한 특징을 가지고 있다.

식 (2)의 분모항인 일사량  $G$ 와 온도  $T$ 의 결합 확률 분포는 전체 표본 수와 변수  $G$ 와  $T$ 가 동시에 발생하는 경우의 수의 비율로 나타낼 수 있으며 수학적으로

$$p(G, T) \approx \frac{N_{G,T}}{N} \quad (5)$$

와 같다. 여기서  $N$ 은 전체 표본 데이터의 수,  $N_{G,T}$ 는 확률 변수  $G$ 와  $T$ 가 동시에 발생한 경우의 수를 나타낸다. 다음으로 식 (2)에서 분자항의 결합 확률도 마찬가지로 다음과 같이 표현된다.

$$p(X, G, T) \approx \frac{N_{X,G,T}}{N} \quad (6)$$

여기서  $N_{X,G,T}$ 는 세 확률변수  $X, G, T$ 가 동시에 발생하는

경우의 수를 나타낸다. 식 (5)와 식 (6)을 식 (2)에 동시에 대입하면 조건부 확률은 다음과 같이 간단하게 표현된다.

$$p(X|G, T) \approx \frac{N_{X,G,T}}{N_{G,T}} \quad (7)$$

식 (7)에서 알 수 있듯이 일사량  $G$ 와 온도  $T$ 가 주어졌을 때 출력 변수  $X$ 의 확률은  $G$ 와  $T$ 가 동시에 발생하는 경우의 수와  $G, T, X$ 가 동시에 발생하는 경우의 수에 대한 비율을 연산함으로써 구할 수 있다.

#### 4. 모델링 검증

태양광 발전 시스템 모델링은 실제 시스템 입력 변수를 모델에도 동일하게 인가하여 모델의 출력을 연산하여 실제 시스템의 출력을 잘 추종하도록 하는 목적을 갖고 이에 대한 개념도를 그림 2에 도시하였다. 다시 말해 이 두 출력을 근거로 하여 출력 오차를 연산하여 모델링의 성능을 검증할 수 있다. 본 논문에서 정의하는 출력 오차 변수  $\zeta$ 는 다음과 같다.

$$\zeta = \sqrt{(X - \hat{X})^2} \quad (8)$$

여기서  $\hat{X}$ 는 태양광 시스템 모델에서 산출되는 출력 추정 변수를 나타낸다. 본 논문에서 구한 확률 모델로부터 출력  $X$ 에 대한 추정식은 평균값 이론을 적용하여 다음과 같이 표현할 수 있다.

$$\hat{X} = E[X|G, T] \quad (9)$$

여기서  $E[\cdot]$ 는 평균값 연산자를 나타낸다. 확률 이론을 통해 출력변수  $X$ 에 대한 조건부 확률 분포를 이용하여 식 (9)의 평균값은 다음과 같이 표현된다.

$$\hat{X} = \sum_{i=1}^n x_i p(X|G, T) \quad (10)$$

한편, 식 (7)의 확률 식을 식 (10)에 대입하면

$$\hat{X} = \sum_{i=1}^n x_i \left( \frac{N_{X,G,T}}{N_{G,T}} \right) \quad (11)$$

가 된다. 최종적으로, 식 (11)을 식 (8)에 대입하면 출력 오차 식은 다음과 같이 주어진다.

$$\zeta = \sqrt{\left( X - \sum_{i=1}^n x_i \left( \frac{N_{X,G,T}}{N_{G,T}} \right) \right)^2} \quad (12)$$

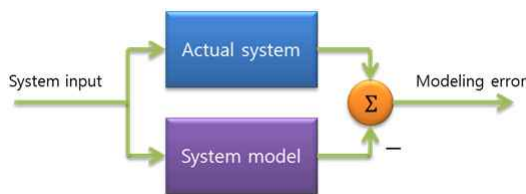


그림 2. 태양광 시스템의 모델링 검증  
Fig. 2. A model testing of photovoltaic systems

#### 5. 사례연구 : 실시간 실험

본 논문에서 제안한 태양광 발전 시스템 모델링 기법의 타당성을 검증하기 위하여 휴엔스 사에 설치되어 있는 3kW급 용량의 태양광 시스템을 이용하였다. 그림 3은 실제 설

치된 태양광 모듈을 보여준다. 이 태양광 모듈의 표면에 일사량 센서와 온도 센서를 각각 부착하여 그 신호는 전기 접속반으로 전송되어 진다. 그림 4는 접속반의 실제 모습을 보여준다. 접속반에는 기상관측 센서들의 데이터 신호를 전자통신을 이용해 사용자 컴퓨터로 전송을 위한 4개의 센서 박스가 설치되어 있다. 기상 검출용 센서박스는 220V 전원이 사용되며 계통과 연계된 인버터의 2차측에서 전원을 공급한다. 센서박스와 인버터는 각각 고유 아이디를 부여하여 RS-485 통신의 해당 노드에 분배가 가능토록 하였다. 태양광 모듈은 각각 250W의 12개가 모두 직병렬로 연결되어 있으며 2개의 채널의 스트링 구조를 갖도록 하였다.

그림 5는 본 연구에 적용한 태양광 발전 시스템의 모니터링에 대한 전체 시스템 구성도를 보여준다. 태양광 시스템의 신호 전송 및 처리를 위하여 NI사의 A/D컨버터를 이용하였으며 통신방식은 이더넷을 사용하였다. RS-485 통신은 모니터링 컴퓨터 앞단에 RS-232/485 컨버터, 인버터, 센서박스의 순서로 연결하였다. 센서박스의 노드는 직접 LED창을 감시하여 설정하였으며 인버터의 기본값은 0으로 설정하였다. 인버터와 같은 시스템에 문제가 발생할 경우 인버터가 자체적으로 알람표시를 하지만 일부 어레이(모듈의 집합체)에서 발생하는 문제는 표시되지 않았다.

그림 6은 모니터링 컴퓨터 화면에 나타나는 내용들을 보여준다. 이 모니터링 시스템을 통해 태양광 시스템에 대한 환경에 대한 데이터를 포함하여 출력 전압 및 전류 등에 대한 정보를 알 수 있으며 자동적으로 데이터가 저장되고 있다. 그림 7은 이러한 데이터들의 화면에 나타난 그래프를 보여주고 있다. 실시간으로 계속해서 데이터를 수집하고 모니터링 화면에 나타내어 주고 있는 것을 볼 수 있다. 그림 8은 저장된 데이터를 엑셀 프로그램을 통해 보여주고 있다.

본 사례연구에 사용한 이 발전 시스템의 학습 데이터는 2013년 4월부터 6월까지 약 90일 동안 관측한 것으로서 일사량, 온도 및 출력 전력에 대하여 각각 8,000개의 데이터 수를 모델링에 적용하였다. 그림 9는 이 관측 데이터의 파형을 각각 보여준다. 이러한 연속시간 데이터 집합의 이산화(discretization)을 위하여 데이터의 범위를 결정하여야 하며, 일사량 크기의 범위는 0에서  $120 \times 10 [W/m^2]$ , 온도는  $-10$ 에서  $70 [^{\circ}C]$ , 출력 전력은 0에서  $350 \times 10 [W]$ 로 한 후 1씩 증가하는 이산 데이터를 만들기 위하여 식 (1)에서  $n = 351, m = 121, l = 81$ 로 각각 결정하였다. 따라서 일사량과 온도 변수에 대한 결합 확률 분포의 수는 9,801개 ( $= 121 \times 81$ )이며 세 변수 일사량, 온도, 출력 전력변수에 대한 결합 확률 분포 수는 총 3,440,151개 ( $= 121 \times 81 \times 351$ )가 된다. 우선 그림 9의 데이터 집합에 대하여 본 논문에서 제안한 확률 분포 추정 알고리즘을 통해 확률 값을 결정하였다. 제안한 추정 알고리즘은 Matlab(c) 소프트웨어를 이용하여 구성하였으며 연산 과정에 대한 순서도를 그림 10에 나타내었다. 다음으로, 태양광 발전 시스템의 500개의 새로운 일사량 및 온도 값에 대하여 추정한 확률 분포 값을 이용하여 식 (11)의 평균 값 연산식을 통해 출력 값을 추정하여 실제 시스템의 출력 값과 비교 분석하였으며 그 결과 파형을 그림 11에 도시하였다. 또한 이 두 파형에 대한 식 (12)의 오차를 연산하여 파형으로 나타내었으며 그림 7과 같다. 그림 12의 오차 파형에서,  $50 [W]$  이상의 오차 범위는 갖는 경우가 있지만 대부분의 경우  $50 [W]$  이하의 범위를 갖는 것을 볼 수 있다. 평균 오차의 경우  $28 [W]$ 로서 전체적으로 만족스러운 것으로 알 수 있다.



그림 3. 태양광 발전 시스템 외형

Fig. 3. Appearance of the photovoltaic power system

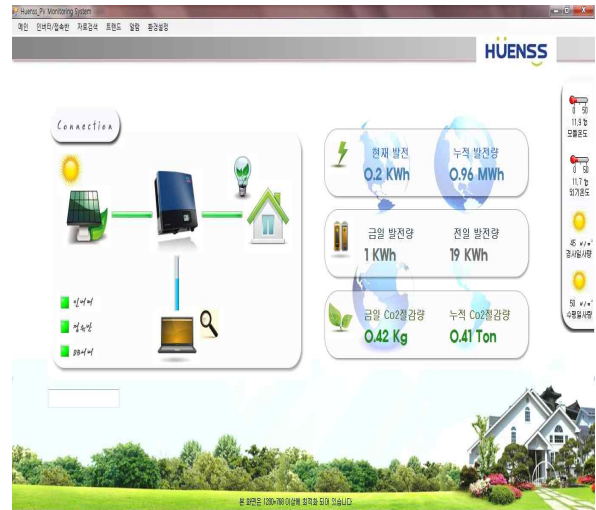


그림 6. 태양광 시스템 모니터링 화면

Fig. 6. A monitoring display of the photovoltaic power system



그림 4. 태양광 발전 시스템의 접속반

Fig. 4. Connection parts of the photovoltaic power system

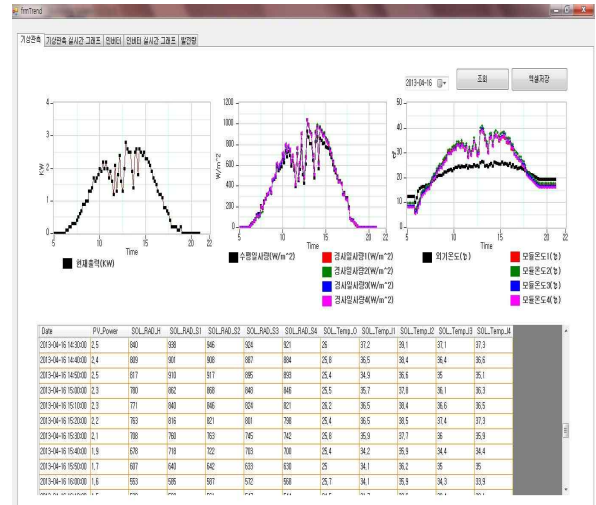


그림 7. 기상관측 및 태양광 출력 화면

Fig. 7. Weather measurement and photovoltaic power monitoring

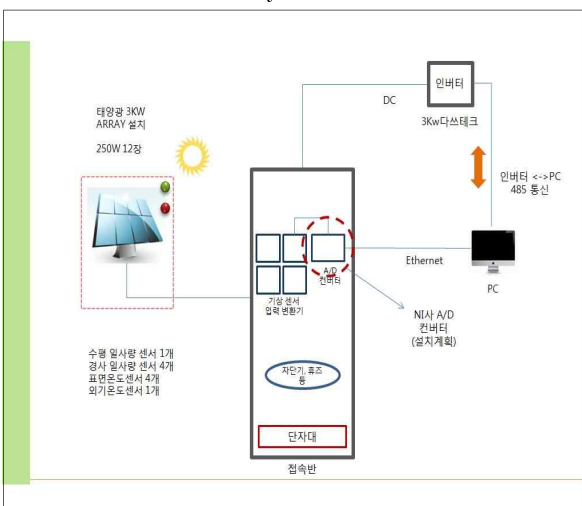


그림 5. 태양광 발전 시스템의 모니터링 구성도

Fig. 5. A monitoring framework of the photovoltaic power system

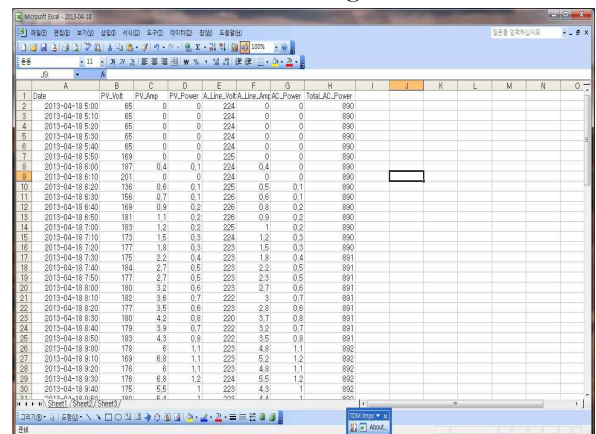


그림 8. 태양광 시스템의 데이터 저장화면

Fig. 8. Data saving display of the photovoltaic system



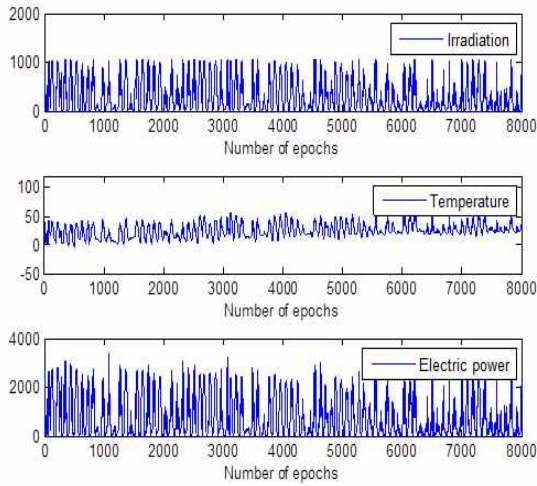


그림 9. 모델링에 사용된 관측 데이터 파형  
 Fig. 9. Waveforms of the measured data in modeling

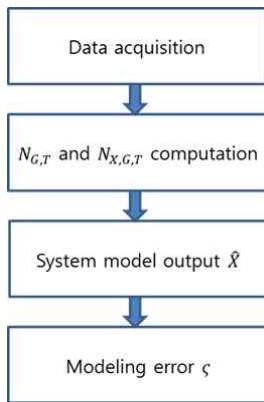


그림 10. 모델링 연산과정의 순서도  
 Fig. 10. A flowchart of computational procedure in modeling

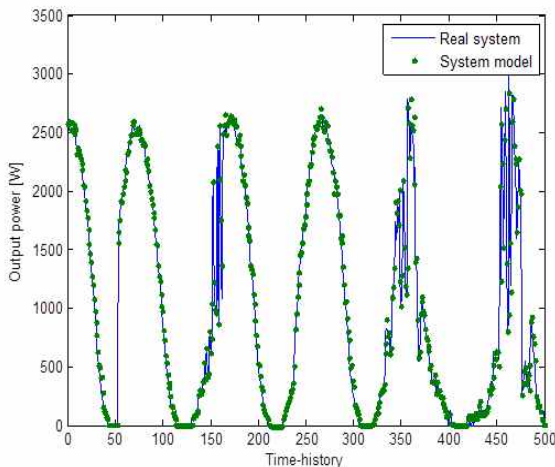


그림 11. 실제 시스템과 시스템 모델의 출력 값의 비교 파형  
 Fig. 11. Comparative waveforms of an actual system and a system model

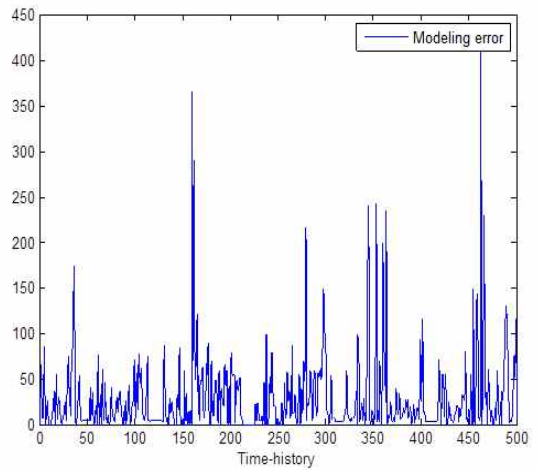


그림 12. 모델링 오차 파형  
 Fig. 12. Modeling error waveform

### 6. 결론

본 논문은 대용량 학습 데이터를 갖는 태양광 발전 시스템의 확률론적 모델링 방법론을 제시하였다. 태양광 시스템의 입력 변수인 태양광의 일사량과 태양광 모듈의 표면 온도에 대한 출력 전력의 조건부 확률 추정은 전체 데이터 수와 해당 변수의 경우의 수에 대한 비율로 결정하였다. 이러한 확률 분포 값은 평균 값 이론에 적용하여 실제 시스템의 출력을 추정하는 사용되었다. 제한한 확률론적 모델링 기법의 타당성 검증에 위하여 실제 산업체에 설치된 태양광 발전 시스템을 이용한 사례 연구를 실시하였으며 실제 시스템 출력과 비교 분석을 실시하여 그 신뢰성을 검증하였다. 하지만 오프라인으로 학습을 하지 못한 관측 데이터의 값에 대해서는 상대적으로 큰 오차 범위를 갖는 것으로 분석되었으나 평균적으로 전체적인 오차의 크기는 허용할 수 있는 범위에 있어 결과적으로 만족스러운 모델링 성능을 나타내었다. 향후 연구로는 모델링 오차의 값을 줄이기 위한 방향으로 연구가 더 진행되어야 할 것이며, 오프라인에서 경험하지 못한 학습 데이터를 온라인 모델링 방식을 통해 오차의 범위를 줄이고자 한다.

### References

- [1] Hyun C. Cho, Kwan H. Lee, "Fault detection algorithm of photovoltaic power systems using stochastic decision making approach," *Korean Signal & Systems Journal*, vol. 12, no. 3, pp. 212-216, 2011.
- [2] D. Sera, R. Teodorescu, "Photovoltaic module diagnostics by series resistance monitoring and temperature and rated power estimation," *Annual Conference of IEEE*, pp. 2195-2199, 2008.
- [3] K.-H. Chaoa, S.-H. Hob, M.-H. Wanga, "Modeling and fault diagnosis of a photovoltaic system," *Electric Power Systems Research*, vol. 78, no. 1, pp. 97 - 105, 2008.

- [4] A. Mellit, S. Kalogirou, "ANFIS-based modelling for photovoltaic power supply system: A case study," *Renewable Energy*, vol. 36, no. 1, pp. 250-258, 2011.
  - [5] C. S. Chin, A. Babu, W. McBride, "Design, modeling and testing of a standalone single axis active solar tracker using MATLAB/Simulink," *Renewable Energy*, vol. 36, no. 11, pp. 3075-3090, 2011.
  - [6] Hyun C. Cho, "A study on dynamic modeling of photovoltaic power generator systems using probability and statistics theories," *The Trans. of the Korean Institute of Electrical Engineers*, vol. 61, no. 7, pp. 1007-1013, 2012.
  - [7] F. V. Jensen, T. D. Nielsen, *Bayesian networks and decision graphs*, Springer, 2007.
  - [8] Hyun Cheol Cho, Sami M. Fadali, Kwon S, Lee, "Online parameter estimation and convergence property of dynamic Bayesian networks," *Int. Journal of Fuzzy Logic and Intelligent Systems*, vol. 7, no. 4, pp. 285-294, 2007.
- 

## 저 자 소 개



### 조현철(Hyun Cheol Cho)

1997년 : 부경대학교 제어계측공학과 공학사  
1999년 : 동아대학교 전기공학과 공학석사  
2006년 : 미국 네바다 주립대학교 전기공학과 공학박사  
2006년~2009년 : 동아대학교 NRL 선임연구원  
2009년~현재 : 울산과학기술대학교 전기전자공학부 부교수

관심분야 : 자동제어, 신재생에너지, 고장진단, 신경회로망  
Phone : +82-10-9211-0351  
E-mail : hcjo@uc.ac.kr



### 정영진 (Young Jin Jung)

2003년 : 연세대학교 전기공학과 공학사  
2009년 : 서울대학교 전기공학과 공학박사  
2009년~2012년 : LG Display 선임연구원  
2012년~현재 : 울산과학기술대학교 전기전자공학부 조교수

관심분야 : Solar-cell, photonics  
Phone : +82-10-8422-9243  
E-mail : yjjung@uc.ac.kr