

단일 태양광 발전소의 운영형 발전량 예측 정확도 향상 실증 연구

(An Empirical Study on Improving Operational PV Power Forecasting Accuracy for a Single Solar Power Plant)

임인선*, 손경우*, 노순국**

(In Seon Lym, Kyoung Woo Son, Sun-Kuk Noh)

요약

본 연구는 단일 발전소 환경에서 모델 구조의 복잡도 보다 입력 예보자료의 품질과 예측 지평선별 정보 가용성 설계가 예측 성능에 더 큰 영향을 미치는지를 검증하였다. 이를 위해 설비용량 120 kW급 단일 태양광 발전소를 대상으로 인버터 실측자료, 기상예보자료, 천문 정보 및 다중 수치예보자료를 통합한 운영형 발전량 예측 파이프라인을 구축하였다. 관측 발전출력은 청천 조건에서 기대되는 발전량을 기준으로 정규화한 청천지수로 변환하여 예측 대상으로 설정하였다. 실험 결과, 제안한 파이프라인은 0-3시간 단기 예측 구간에서 MAE 5.87 kW, 3-24시간 선행예측 구간에서 MAE 8.01 kW의 성능을 보였다. 또한 지평선별 기준선 대비 각각 RMSE skill 0.11 및 0.47의 개선을 나타냈다. 본 연구 결과는 다중 수치예보자료가 확보된 단일 발전소 환경에서는 모델 구조의 복잡도보다 입력 예보자료의 품질, 학습자료와 추론자료의 정합성, 예측 지평선별 정보 가용성 설계가 예측 정확도에 더 중요하게 작용할 수 있음을 보여준다.

■ 중심어 : 태양광 발전량 예측 ; 청천지수 ; XGBoost ; 다중 수치예보 ; 컨포멀 예측 ; 발전량 예측제도

Abstract

This study verified whether the quality of input forecast data and the design of information availability by prediction horizon have a greater impact on prediction performance than the complexity of the model structure in a single power plant environment. To this end, an operational power generation forecast pipeline was constructed for a single 120 kW solar power plant by integrating actual inverter data, meteorological forecast data, astronomical information, and multiple numerical weather forecast data. Observed power output was converted into a clear-sky index, normalized based on the power generation expected under clear conditions, and set as the prediction target. Experimental results showed that the proposed pipeline demonstrated performance with an MAE of 5.87 kW in the 0-3 hour short-term forecasting interval and an MAE of 8.01 kW in the 3-24 hour leading forecasting interval. Furthermore, it showed improvements of RMSE skill of 0.11 and 0.47, respectively, compared to the baseline for each horizon. The results of this study demonstrate that in a single power plant environment where multiple numerical weather forecast data are available, the quality of input forecast data, the consistency between training and inference data, and the design of information availability per prediction horizon may play a more significant role in prediction accuracy than the complexity of the model structure.

■ keywords : PV forecasting ; clear-sky index ; XGBoost ; multi-NWP ; conformal prediction ; policy-aware evaluation

I. 서론

최근 신재생에너지 보급이 확대됨에 따라
태양광 발전량 예측의 중요성은 지속적으로

* 정회원, 조선대학교 일반대학원 과학기술융합학과 박사과정

** 종신회원, 조선대학교 자유전공학부 & 과학기술융합학과 (대학원)

이 논문은 2026년도 조선대학교 실용화프로젝트 학술연구비의 지원을 받아 수행된 연구임

접수일자 : 2026년 07월 01일

게재확정일 : 2026년 07월 20일

교신저자 : 노순국 e-mail : nsk7078@chosun.ac.kr

증가하고 있다. 태양광 발전은 기상 조건과 일사량의 변화에 따라 출력 변동성이 크게 나타나는 전원으로, 발전량 예측의 정확도는 계통 운영의 안정성, 전력거래의 효율성, 전력중개사업자의 정산 리스크 관리에 직접적인 영향을 미친다. 특히 국내 재생에너지 발전량 예측제도는 전력시장운영규칙의 개정에 따라 예측오차율 산정 방식, 정산단가, 적용 대상 및 시행 시점 등이 변경될 수 있다. 따라서 발전량 예측모델의 성능 평가는 단순한 오차 지표의 비교에 그치지 않고, 실제 운영 및 정산제도 적용 가능성을 고려한 검토와 함께 이루어질 필요가 있다[1,2].

단일 태양광 발전소 환경에서 발전량 예측 성능 향상의 주요 요인은 사용되는 딥러닝 모델 구조의 확장에만 있는 것이 아니라, 학습자료와 추론자료의 정합성, 예측 대상의 적절한 정의, 입력 예보자료의 품질, 그리고 예측 지평선별 정보 가용성 설계에 있다. 이에 따라 본 연구는 단일 태양광 발전소의 운영 환경에서 모델 구조의 복잡도 증가보다 입력 예보자료의 품질과 예측 지평선별 가용 정보 설계가 발전량 예측 정확도 향상에 더 큰 영향을 주는지를 확인하고자 한다.

이를 검증하기 위해 본 연구는 동일한 입력자료와 동일한 학습·검증·시험 분할 조건에서 여러 예측 모델의 성능을 비교하고, 입력자료 그룹별 기여도와 예측 지평선별 기준선 대비 성능을 함께 분석한다. 또한 최근 실측 정보를 활용할 수 있는 0-3시간 단기 예측 구간과 이를 활용하기 어려운 3-24시간 선행 예측 구간을 구분하여, 운영 조건에 부합하는 예측 지평선 분리 전략의 타당성을 검토한다.

II. 관련연구

태양광 발전량 예측에 관한 선행연구는 물리 기반 모델과 데이터 기반 모델을 결합한 하이브리드 접근, pvlib 기반의 청천일사 산정,

트랜스포머 계열 시계열 예측모델, 그래디언트 부스팅 및 앙상블 기법, 그리고 컨포멀 예측을 활용한 확률 예측 방법으로 확장되어 왔다[3-11].

1. 데이터 및 점검

그림 1은 본 연구에서 구축하고 검증한 운영형 태양광 발전량 예측 파이프라인의 전체 구성을 나타낸다. 입력자료는 인버터에서 수집한 발전량 실측자료, 발전소 현장의 지상 일사계 관측자료, 기상청 국지예보모델 자료, 천문·시간 정보 및 Open-Meteo API를 통해 수집한 다중 수치예보자료로 구성하였다.

수치예보(Numerical Weather Prediction, NWP)는 대기의 물리방정식을 수치적으로 계산하여 미래의 기상 상태를 예측하는 방법을 의미한다. 본 연구에서는 기상청의 국지예보시스템인 LDAPS(Local Data Assimilation and Prediction System) 자료와 Open-Meteo Historical Forecast API에서 제공하는 ECMWF, GFS 및 ICON 계열의 예보자료를 사용하였다[12, 17]. LDAPS는 한반도 주변의 국지적인 기상 현상을 예측하기 위해 기상청에서 운영한 고해상도 국지 수치예보시스템이다. Open-Meteo는 단일 예보모델이 아니라 여러 수치예보모델의 자료를 통합하여 제공하는 API 서비스이며, 본 연구에서는 과거 각 시점에 제공되었던 예보자료를 확보하기 위해 Historical Forecast API를 이용하였다.

본 연구에서는 입력자료의 성격과 추론 시점에서의 가용성을 기준으로 입력 특성을 BASE, LDAPS, Open-Meteo, SI 및 AR의 다섯 그룹으로 구분하였다. 이 가운데 BASE, SI 및 AR은 본 연구에서 입력 특성을 구분하기 위해 정의한 그룹명이다. BASE는 태양 위치, 시간 주기 및 청천 조건을 나타내는 기본 특성 그룹이고, SI(Site Irradiance)는 발전소 현장의

지상 일사계에서 측정한 일사 관측자료 그룹이다. AR(Autoregressive)은 최근 발전량과 청천지수의 시차값 및 이동통계량으로 구성된 자기회귀 특성 그룹이다.

그림 1에 표시된 GK2A 기반 구름 나우캐스트 자료는 서버의 운영 워크플로에서 단기 예측 확장 입력으로 검토된 항목이다. 그러나 본 논문의 실증 분석에 사용된 주요 특성 집합에는 포함되지 않았으므로, 연구 범위와의 구분을 위해 점선 박스로 표시하였다.

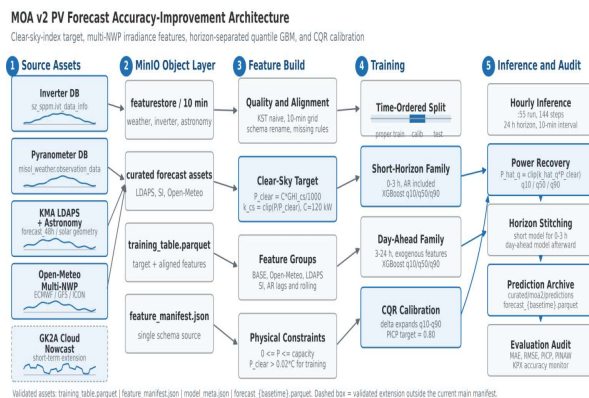


그림 1. 발전량 예측 정확도 향상 실증 아키텍처

가. 입력 피처와 정제 규제

사용한 입력자료는 자료의 성격과 추론 시점의 가용성을 기준으로 BASE, LDAPS, Open-Meteo, SI, AR의 다섯 개 그룹으로 구분하여 표 1에 나타내었다. BASE 그룹은 태양 위치, 시간 주기, 청천 조건 정보를 포함하는 기본 특성으로 모든 예측 구간에서 사용하였다. LDAPS 그룹은 기상청 수치예보자료를 기반으로 한 예보 특성이며, Open-Meteo 그룹은 ECMWF, GFS, ICON 계열의 다중 수치예보 자료를 활용한 입력 특성이다. SI 그룹은 지상 일사계 관측자료를 의미하며, AR 그룹은 최근 발전량 및 청천지수 이력을 이용한 자기회귀 특성으로 구성하였다.

표 1. 입력 피처 그룹 및 추론 가용성

그룹	대표 특성	수	추론 가용성
BASE(기본-천문 특성)	solar_elevation/azimuth/zenith, sin/cos, clearsky_ghi	8	필수
LDAPS(기상청 국지예보)	ldaps_ghi, cloud_low/mid/high/total, temp/rh/wind/precip	12	선택
Open-Meteo (다중 수치예보)	om_ghi, mean/std/ecmwf/gfs/icon, DNI/DHI, cloud, om_kt	9	필수 (운영 모델)
SI(현장 지상일사)	SI, si_ratio	2	선택 (가용률 제한)
AR(자기회귀 특성)	k_lag1/6, k_roll6, power_lag1/6	5	단기만

입력자료의 정제 과정에서는 우선 자료원별 시간 기준을 통일하였다. 원천 자료에는 협정세계시와 한국표준시가 혼재할 수 있으므로, 모든 시각 정보는 한국표준시 기준으로 변환한 뒤 동일한 시간 형식으로 정렬하였다. 이후 발전량, 기상예보, 천문 정보 및 다중 수치예보 자료를 10분 단위 시간 해상도로 맞추고, 동일한 예보 유효 시각을 기준으로 병합하였다. 이 과정에서 자료별로 상이하게 정의된 원시 변수명은 학습 및 추론 단계에서 동일하게 사용될 수 있도록 표준 변수명으로 변환하였다.

학습자료와 추론자료의 정합성을 확보하기 위해 예보자료의 처리 기준을 명확히 하였다. Open-Meteo 기반 다중 수치예보 자료는 과거 시점에 실제 발행되었던 예보값을 제공하는 Historical Forecast API를 통해 수집하였다[12].

각 시점의 입력자료는 해당 시점에서 실제로 가용하였던 예보값을 기준으로 구성하였다. LDAPS와 Open-Meteo 자료는 예보 유효시각을 기준으로 정렬한 뒤 10분 단위로 리샘플링 및 보간하여 학습 테이블에 통합하였다. 따라서 Open-Meteo 입력자료의 기여도는 사후적으로 관측자료에 맞추어 얻은 효과가 아니라, 실제 운영 시점에서 활용 가능한 예보자료의 품질에 따른 효과로 해석할 수 있다.

나. 학습 대상과 모델

정제된 학습자료는 발전출력을 직접 예측하는 방식이 아니라, 청천 조건을 기준으로 정규화한

지수를 예측하는 방식으로 구성하였다. 발전출력을 그대로 예측 대상으로 사용할 경우 태양 고도, 계절 변화, 일사 조건에 따른 구조적 변동이 모두 포함된다. 이러한 변동은 기상 조건에 따른 발전량 감쇠 효과를 분리하여 학습하는 데 제약으로 작용할 수 있다.

이에 따라 청천 조건에서 기대되는 발전량을 기준으로 실제 발전량을 정규화한 청천지수를 학습 대상으로 설정하였다. 청천지수는 실제 발전량이 청천 조건에서의 기준 발전량에 비해 어느 정도의 수준으로 나타나는지를 의미한다. 이와 같은 변환을 통해 발전량의 절대 규모보다 구름, 강수, 습도 등 기상 조건에 따른 상대적 출력 변화가 학습 과정에서 보다 명확하게 반영되도록 하였다.

청천지수 타깃은 다음과 같이 정의하였다. 먼저 청천 조건에서 기대되는 기준 발전량 (P_{clear})를 설비용량 (C)와 청천 전천일사량 (GHI_{cs})를 이용하여 산정하였다. 이후 실제 발전량 (P)를 (P_{clear})로 나누어 청천지수 (k)를 정의하고, 모델이 예측한 청천지수 ($\hat{k}$)를 다시 (P_{clear})와 곱하여 예측 발전량 ($\hat{P}$)로 복원하였다.

$$P_{clear} = C \frac{GHI_{cs}}{1,000} \quad (1)$$

$$k = clip\left(\frac{P}{P_{clear}}, 0, 1.3\right) \quad (2)$$

$$\hat{P} = clip(\hat{k} \cdot P_{clear}, 0, C) \quad (3)$$

식(1)은 기준발전량, 식(2)는 청천지수, 식(3)은 예측발전량을 나타내었다. 여기서 C 는 발전소 설비용량이며, 본 연구에서는 $C = 120kW$ 이다. GHI_{cs} 는 pvlib의 Ineichen 청천일사 모형으로 산출한 청천 조건의 전천일사량이다. P 는 실제 관측 발전량, k 는 청천지수, $\hat{k}$ 는 모델이 예측한 청천지수, $\hat{P}$ 는 복원된 예측 발전량을 의미한다. $clip(x, a, b)$ 는 값 (x)를 하한 (a)와 상한 (b) 사이로 제한하는 함수이다. 본 연구에서는 야간 및 저고도 구간에서 분모 불안정을 방지하기

위해 $P_{clear} \leq 0.02C$ 인 행을 제외하였고, 청천지수 k 는 $[0, 1.3]$ 범위로 제한하였다.

예측모델은 XGBoost 기반 분위수 회귀모형으로 구성하였다. 분위수 회귀모형은 중앙 예측값뿐만 아니라 하위 분위와 상위 분위의 예측값을 함께 산출할 수 있으므로, 발전량 예측의 불확실성을 표현하는 데 적합하다. 본 절차에서는 하위 분위, 중앙 분위, 상위 분위에 해당하는 모델을 각각 학습하였으며, 중앙 분위 예측값은 대표 발전량 예측값으로 활용하고 하위 및 상위 분위 예측값은 예측구간을 구성하는 데 사용하였다.

예측구간의 경험적 신뢰도를 보정하기 위해 conformalized quantile regression을 적용하였다. 학습 자료와 별도로 분리한 보정 자료에서 실제값이 예측구간을 벗어나는 정도를 산출하고, 이를 바탕으로 예측구간의 보정량을 계산하였다. 산출된 보정량은 하위 및 상위 분위 예측값에 반영되어 최종 예측구간을 형성한다.

그림 2는 이러한 학습 및 보정 절차를 나타낸다. 정제된 입력자료는 청천지수 형태의 학습 대상으로 변환되고, 예측 지평선에 따라 학습자료가 구성된 뒤 분위수 회귀모형을 통해 예측값과 예측구간이 산출된다. 이후 보정자료를 이용한 예측구간 보정을 거쳐 최종 발전량 예측값과 불확실성 범위가 도출된다.

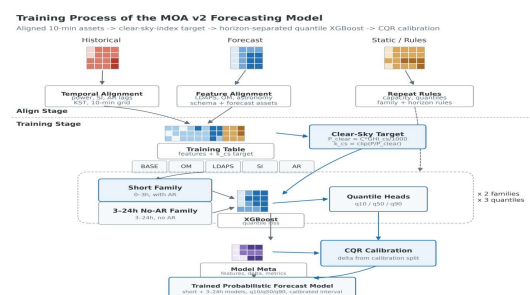


그림 2. AI 모델 학습 과정

다. 추론과 제도 대응

추론 단계에서는 매시 55분을 기준으로 향후 24시간의 발전량 예측값을 10분 간격으로

생성하였다. 이에 따라 한 번의 추론 과정에서 총 144개 시점의 예측값이 산출된다. 매시간 수행되는 추론은 기상예보자료의 갱신 상황을 반영하고 발전소의 예측 상태를 지속적으로 점검하기 위한 운영 절차이다.

이와 별도로 발전량 예측제도 대응을 위한 실증에서는 매일 10시 이전과 17시 이전에 각각 발전량 예측을 수행하여 하루 두 차례 예측 결과를 제출하였다. 각 제출 결과는 해당 제출 마감시각 이전에 시스템에서 확보할 수 있었던 기상예보자료, 천문 정보 및 최근 발전량 자료만을 이용하여 생성하였다. 따라서 10시 제출 결과에는 10시 이후 갱신된 예보자료를 사용하지 않았으며, 17시 제출 결과 역시 17시 이후에 확보된 자료를 포함하지 않았다.

그림 3은 발전량 예측모델의 추론 절차를 나타낸다. 추론 시점에서 가용한 기상예보자료, 천문 정보, 다중 수치예보자료 및 최근 실측자료를 입력으로 구성한 뒤, 예측 지평선에 따라 서로 다른 입력 조건을 적용한다. 0-3시간 단기 예측 구간에서는 최근 발전량 및 청천지수 이력에 기반한 자기회귀 특성을 함께 사용한다. 반면 3시간 이후의 선행예측 구간에서는 최근 실측값을 안정적으로 활용하기 어렵기 때문에, 자기회귀 특성을 제외하고 외생 예보자료를 중심으로 예측을 수행한다.

전력거래소(KPX) 정확도는 모델 간 비교와 운영 성능 점검을 위한 평가 지표로 사용하였다. 다만 실제 정산 적용 여부는 전력시장 운영 규칙의 개정에 따라 달라질 수 있으므로, 제출 시각, 변경 제출 가능 여부, 정산단가, 오차율 산정 방식, 적용 대상 및 유효 시간 기준은 운영 시점의 최신 KPX 전력 시장 운영 규칙과 관련 공고를 기준으로 재확인해야 한다[1][2].

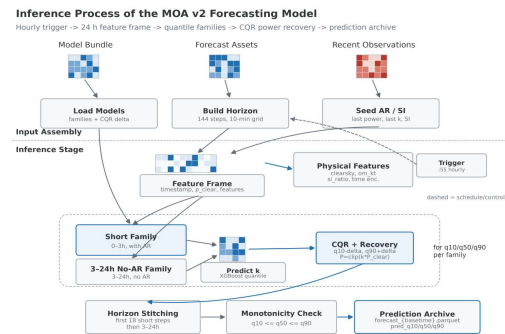


그림 3. AI 모델 추론 과정

III. 실험 및 결과분석

1. 입력 그룹 ablation

입력자료 그룹별 기여도를 분석하기 위해 BASE 그룹을 공통 입력으로 고정하고, LDAPS, Open-Meteo, SI, AR 그룹의 포함 여부에 따른 16개 입력 조합을 구성하였다. 각 조합은 동일한 학습·검증·시험 분할 조건에서 학습하였으며, 입력 그룹의 추가 여부에 따른 예측 성능 변화를 비교하였다.

각 입력 그룹의 평균 한계 기여도 $\Delta(g)$ 는 식(4)에 나타내었고, 특정 입력 그룹 g 를 추가하기 전후의 MAE 차이를 모든 가능한 입력 조합에 대해 평균하여 산출하였다. 여기서 g 는 기여도를 평가할 입력 그룹이며, S_g 는 입력 그룹 g 를 제외한 나머지 입력 그룹들로 구성 가능한 모든 부분 조합의 집합이다. N_g 는 해당 부분 조합의 개수이다. 본 연구에서는 선택 입력 그룹을 LDAPS, Open-Meteo, SI, AR의 네 개로 설정하였으므로, 각 입력 그룹에 대해 총 8개의 조합을 비교하였다. $\Delta(g)$ 가 양수이면 해당 입력 그룹을 추가했을 때 평균적으로 MAE가 감소했음을 의미한다.

$$\Delta(g) = \frac{1}{N_g} \sum_{S \in S_g} [MAE(BASE + S) \quad (4)$$

$$- MAE(BASE + S + \{g\})]$$

$$S_g = \{S | S \subseteq G - \{g\}\}, N_g = |S_g| = 2^{G-1} \quad (5)$$

양수는 해당 그룹 추가 시 오차 감소를 뜻한다. 본 ablation은 AR을 포함한 16개 조합(단기 패밀리, q50 점예측 MAE)에서 계산하였으므로 AR의 기여(+3.93 kW)는 0-3시간 단기 구간에 한정된 양이며, AR을 쓸 수 없는 3-24시간 구간의 지평선별 ablation은 후속 과제로 둔다.

그림 4는 16개 입력 조합별 MAE를 비교한 결과이다. 이러한 결과는 최근 실측자료를 활용한 자기회귀 특성이 0-3시간 단기 예측의 정확도 향상에 크게 기여했음을 보여준다. 또한 Open-Meteo 기반 다중 수치예보자료가 기존 LDAPS 자료보다 더 큰 평균 기여도를 보였다는 점에서, 입력 예보자료의 품질과 다양성이 발전량 예측 성능에 중요한 요인임을 확인할 수 있다. 기상 일사계 자료는 물리적으로 유효한 관측 자료이나, 본 실험에서는 이력자료의 비영관측 가용률이 낮아 평균적인 성능 기여가 제한적으로 나타났다.

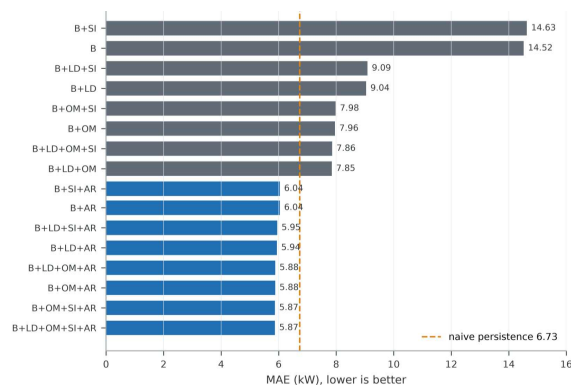


그림 4. 16개 피쳐 조합별 MAE

2. 모델 용량 대조(4-모델 패널)

동일한 입력자료, 데이터 분할 및 청천지수 기반 예측 대상 조건에서 XGBoost, LightGBM, 분위수 MLP, LSTM의 네 가지 예측모델을 비교하였다. 표 2에서 제시한 바와 같이, 네 모델의 MAE는 0-3시간 단기 예측 구간에서 5.826 - 5.938 kW, 3-24시간 선행 예측구간에서 7.910 - 8.006 kW의 범위에 분포하였다. 이는

모델 간 MAE 차이가 각각 약 1.9%와 1.2% 수준에 머물렀음을 의미한다.

Diebold - Mariano 검정 결과 일부 모델 쌍에서는 통계적으로 유의한 차이가 확인되었다[15]. 그러나 예측오차 차이의 절대 규모는 설비용량 대비 약 0.1% 내외로 나타나, 실제 운영 의사결정 측면에서 그 차이는 제한적인 것으로 판단된다. 따라서 본 연구 대상과 같이 다중 수치예보자료가 확보된 단일 발전소 환경에서는 모델의 구조적 차이보다 입력 예보자료의 품질과 예측 지평선별 정보 가용성 설계가 예측 정확도에 더 중요하게 작용한 것으로 나타났다.

표 2. 동일 입력·분할에서의 4-모델 패널 성능

모델	지평선	MAE	RMSE	CRPS (3분위근사)	Skill (RMSE)	PICP (보정)
XG Boost	단기	5.871	8.458	3.449	+0.111	0.756
LightGBM	단기	5.850	8.464	3.434	+0.110	0.767
MLP	단기	5.938	8.494	3.467	+0.107	0.771
LSTM	단기	5.826	8.450	3.414	+0.112	0.782
XG Boost	3 - 24h	8.006	11.462	4.584	+0.467	0.781
LightGBM	3 - 24h	7.966	11.466	4.561	+0.466	0.791
MLP	3 - 24h	7.910	11.343	4.545	+0.472	0.754
LSTM	3 - 24h	7.945	11.596	4.528	+0.460	0.771

표 2-1. Diebold - Mariano 검정 결과(손실=절대오차, HLN 보정, 양측)

비교쌍	지평선	DM 통계량	p-value
XGBoost vs 기준선	단기	-21.01	<0.001
XGBoost vs 기준선	3 - 24h	-44.71	<0.001
LightGBM vs XGBoost	단기	-2.32	0.021
MLP vs XGBoost	단기	+2.74	0.006
LSTM vs XGBoost	단기	-1.94	0.052 (n.s.)
LightGBM vs XGBoost	3 - 24h	-2.54	0.011
MLP vs XGBoost	3 - 24h	-3.27	0.001
LSTM vs XGBoost	3 - 24h	-1.56	0.118 (n.s.)

손실함수는 절대오차, 양측검정이며 다중 horizon 예측의 오차 자기상관을 고려해 Harvey - Leybourne - Newbold(HLN) 소표본 보정을 적용하였다[15][16]. 음의 통계량은 앞 모델이 더 우수함을 뜻한다. 모든 모델은 지평선-적합 기준선 대비 $p < 0.001$ (Bonferroni 보정에도 유의)로 우수했고, 모델 간 차이는

일부 유의하나 절대 크기는 설비 용량의 0.1% 내외로 운영상 무의미하다.

3. 지평선 분리 성능

시험자료 8,017개 행을 대상으로 예측 지평선별 성능을 평가하였다. 0-3시간 단기 예측구간에서 단기 모델은 MAE 5.87 kW, RMSE 8.46 kW를 기록하였으며, 3-24시간 선행 예측구간에서는 자기회귀 특성을 제외한 모델이 MAE 8.01 kW, RMSE 11.46 kW를 보였다.

성능 지표에서 NMAE는 설비용량이 아니라 평가 자료의 평균 실측 발전량을 기준으로 정규화하였다. 여기서 y 는 평가 대상 시점의 실제 발전량이며, $\text{mean}(|y|)$ 는 평가 자료에서 실제 발전량 절대값의 평균을 의미한다. 따라서 NMAE는 평균적인 실제 발전 규모에 비해 예측오차가 어느 정도인지를 나타내는 지표이다.

$$NMAE = \frac{MAE}{\text{mean}(|y|)} \quad (6)$$

따라서 설비용량 기준 정규화(예: $5.87/120 = 0.049$)와 값이 다르며, KPX 정확도는 별도로 설비용량 기준 정규화를 사용한다.

$$Acc_Q = 100 - 100 \cdot \frac{1}{Q} \sum_{t \in Q} \frac{|f_t - a_t|}{C} \quad (7)$$

KPX 정확도는 설비용량 기준 평균 절대 오차율을 100점 기준에서 차감하여 산정 하였다. 여기서 f_t 는 시점 t 의 예측 발전량, a_t 는 시점 t 의 실제 발전량, C 는 발전소 설비용량, Q 는 정확도를 평가하는 시점 집합을 의미한다. 정확도 값이 클수록 설비용량 대비 예측오차가 작다는 것을 뜻한다.

예측 지평선에 따라 활용 가능한 정보가 다르므로, 기준선 또한 구간별로 구분하여 설정하였다. 단기 예측 구간의 기준선은 최근 청천지수를 유지하여 향후 시점의 청천 기준 발전량에 적용하는 smart persistence로

설정하였으며, 이 기준선의 성능은 MAE 6.80 kW, RMSE 9.51 kW로 나타났다.

이는 입력 조합 실험에서 비교 기준으로 사용한 naive persistence, 즉 직전 발전량을 그대로 유지하는 방식과는 구분된다. 3-24시간 선행 예측 구간에서는 전일 동일 시간대의 청천지수를 활용하는 day-lag persistence와 과거 평균 패턴에 기반한 climatology를 기준선으로 설정하였다.

기준선 대비 성능 개선 정도를 RMSE skill이라 하면

$$Skill_{RMSE} = 1 - \frac{RMSE_{\text{model}}}{RMSE_{\text{baseline}}} \quad (8) \text{로 평가한}$$

결과,

단기 모델은 smart persistence 대비 약 0.11의 개선을 보였으며, 3-24시간 선행 예측 모델은 day-lag persistence 대비 약 0.47의 개선을 나타냈다. 또한 Diebold - Mariano 검정 결과, 모든 모델은 해당 지평선의 기준선과 비교하여 통계적으로 유의한 차이를 보였다($p < 0.001$).

한편 1-step persistence는 최근 관측값을 즉시 활용할 수 있는 매우 짧은 예측 구간에서는 유효한 기준선이 될 수 있으나, 3-24시간 선행예측 구간에서는 해당 시점의 관측값을 사용할 수 없기 때문에 적절한 비교 기준으로 보기 어렵다. 또한 day-lag persistence의 성능이 climatology보다 낮게 나타난 것은 태양광 발전량이 전일 동일 시간대의 패턴만으로 설명되기 어렵고, 일별 기상 변동성의 영향을 크게 받는다는 점을 시사한다.

표 3. 예측 지평선별 모델 및 기준선 성능 비교

모델/기준선	MAE	RMSE	NMAE/PICP
단기 0-3h(AR)	5.87	8.46	0.186 / 0.756
3-24h(무AR)	8.01	11.46	0.254 / 0.781
naive persistence (발전량 지속)	6.73	9.72	0.213
청천 물리	29.21	37.44	0.926
smart persistence (청천지수, 단기 기준선)	6.80	9.51	-
day-lag persistence(3-24h 기준선)	15.34	21.49	-
climatology (3-24h 보조 기준선)	14.12	17.77	-

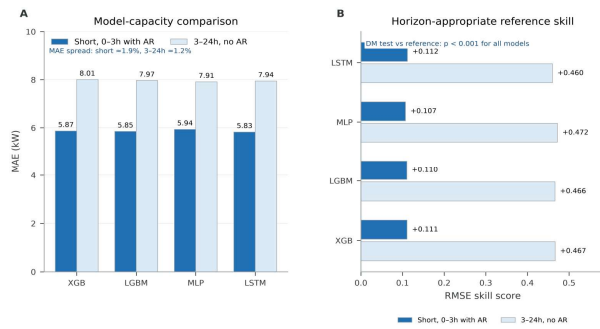


그림 5. 4-모델 용량 대조와 지평선-적합 기준선 skill

4. 확률예측 보정과 예측구간

정적 conformalized quantile regression(CQR)을 적용한 결과,

$$PICP = \frac{1}{n} \sum_{i=1}^n I(L_i \leq y_i \leq U_i) \quad (9)$$

$$|ACE| = |PICP - (1 - \alpha)| \quad (10)$$

$$PINAW = \frac{1}{n} \sum_{i=1}^n \frac{U_i - L_i}{C} \quad (11)$$

$$W_i = (U_i - L_i) + \frac{2}{\alpha} [(L_i - y_i)_+ + (y_i - U_i)_+] \quad (12)$$

$$Winkler = \frac{1}{n} \sum_{i=1}^n W_i \quad (13)$$

여기서 (L_i) 와 (U_i) 는 각각 i 번째 시점의 예측구간 하한과 상한이며, (y_i) 는 실제 발전량, (n) 은 평가자료의 표본 수를 의미한다. 지시함수 $I(L_i \leq y_i \leq U_i)$ 는 실제값 y_i 가 예측구간 $[L_i, U_i]$ 안에 포함되면 1, 그렇지 않으면 0의 값을 갖는다. 따라서 PICP는 전체 평가자료 중 실제값이 예측구간에 포함된 비율을 나타낸다.

(α) 는 유의수준이며, $(1-\alpha)$ 는 명목 커버리지를 의미한다. PINAW는 평균 예측구간 폭을 설비용량 (C) 로 정규화한 값이다. Winkler 점수는 예측구간 폭과 실제값이 구간 밖으로 벗어났을 때의 벌점을 함께 고려하는 지표이며, $((x)_+ = \max(x, 0))$ 으로 정의된다.

예측구간 포함률(PICP)은 0-3시간 단기 예측 구간에서 0.692에서 0.756으로, 3-24시간 선행 예측구간에서 0.697에서 0.781로 개선되었다. 다만 두 구간 모두 명목 커버리지 0.80에는

도달하지 못하였다. 이는 보정 자료와 시험자료 사이에 예측오차 또는 보정 점수의 분포 차이가 존재할 가능성을 시사한다.

이러한 한계를 보완하기 위해 동일한 분위수 예측 모델에 대해 일 단위 reset CQR과 온라인 적응형 컨포멀 추론(adaptive conformal inference, ACI)을 후처리 방식으로 적용하였다[13]. 표 4는 정적 CQR, 일 단위 reset CQR, 온라인 ACI의 예측구간 보정 성능을 예측 지평선별로 비교한 결과이며, 그림 6은 각 보정 방법에 따른 포함률과 구간폭의 변화를 시각적으로 나타낸 것이다.

표 4. 예측구간 보정 방법별 성능 비교(명목 80%, q10-q90)

지평선	보정기법	PICP	ACE	PINAW	Winkler
단기	정적 CQR	0.7563	0.0437	0.1541	27.224
단기	일단위 reset	0.7722	0.0278	0.1594	27.218
단기	온라인 ACI	0.7749	0.0251	0.1614	27.237
3-24h	정적 CQR	0.7813	0.0187	0.2144	34.152
3-24h	일단위 reset	0.7745	0.0255	0.2135	34.169
3-24h	온라인 ACI	0.7786	0.0214	0.2137	33.983

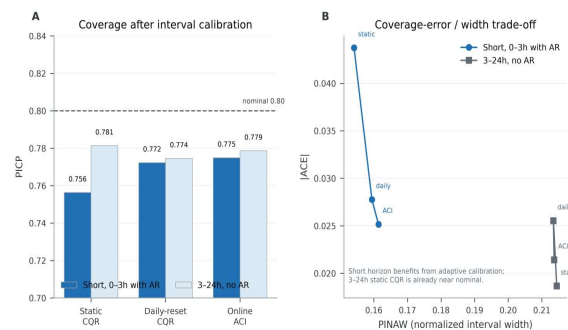


그림 6. 정적 CQR, 일 단위 reset CQR 및 온라인 ACI의 예측구간 보정 비교

표 4와 그림 6에서 확인할 수 있듯이, 0-3시간 단기 예측 구간에서는 온라인 ACI를 적용할 경우 PICP가 0.756에서 0.775로 개선되었으며, 절대 커버리지 오차(|ACE|)도 감소하였다. 이는 단기 예측구간에서 시간에 따른 오차 분포 변화가 상대적으로 크게 나타나며, 이에 대해 적응형 보정 방식이 일정한 개선 효과를 제공할 수 있음을 의미한다. 반면 3-24시간 선행

예측구간에서는 정적 CQR의 PICP가 이미 0.781로 명목 커버리지에 근접하였기 때문에, 적응형 보정에 따른 추가 개선 폭은 제한적으로 나타났다.

확률 예측 성능은 예측구간 포함률뿐만 아니라 CRPS, WIS, pinball loss 등 확률 예측의 정합도를 함께 고려하여 평가하였다[14]. q10, q50, q90의 세 분위만으로는 continuous ranked probability score(CRPS)의 근사가 제한적일 수 있으므로, q05, q25, q50, q75, q95의 다섯 분위 예측값을 추가로 산출하여 확률예측 지표를 재계산하였다. 그 결과 5분위 기반 CRPS는 단기 예측 구간에서 3.94, 3 - 24시간 선행 예측구간에서 5.25로 나타났다. 이는 세 분위 기반 근사값보다 크게 산출되었으나, 더 많은 분위 정보를 반영한다는 점에서 확률분포 평가에는 보다 적절한 결과로 해석된다.

표 5. 확률예측 정식 지표 비교

지평선	CRPS (5분위)	WIS (raw→ CQR)	평균 pinball	IS90 (CQR)	PICP90 (raw→ CQR)	PICP50 (raw→ CQR)
단기	3.941	3.749→ 3.734	1.875	33.2	0.791→ 0.863	0.439→ 0.465
3 - 24h	5.245	4.939→ 4.916	2.470	40.1	0.780→ 0.873	0.425→ 0.484

표 5는 CRPS와 함께 weighted interval score(WIS), 평균 pinball loss, interval score 및 중앙 50%·90% 예측구간 포함률을 제시한 결과이다. 중앙 90% 예측구간의 포함률은 구간별 CQR 보정 후 단기 예측 구간에서 0.791에서 0.863으로, 3 - 24시간 선행예측 구간에서 0.780에서 0.873으로 개선되었다. 반면 중앙 50% 예측구간의 포함률은 0.44 - 0.48 수준에 머물러 상대적으로 과소 포함되는 경향을 보였다.

이러한 결과는 단일 보정량을 모든 예측구간에 동일하게 적용하는 방식만으로는 각 분위 구간의 포함률을 충분히 조정하기 어렵다는 점을 보여준다. 따라서 태양광 발전량과 같이 시간에 따른 오차 분포 변화가 존재하는

자료에서는 구간별 보정 또는 적응형 보정 절차를 함께 고려할 필요가 있다. 또한 본 연구의 자료 분할은 시간순 구조를 가지므로, 컨포멀 예측의 교환가능성 가정이 완전하게 충족된다고 보기 어렵다. 따라서 본 절의 확률예측 지표는 이론적 보장보다는 실제 운영 조건에서 관찰된 경험적 확률예측 성능으로 해석하는 것이 타당하다

5. 시간적 교차검증(robustness)

단일 학습·시험 분할에 따른 결과의 우연성을 완화하기 위해 expanding-window 방식의 6-폴드 rolling-origin 교차검증을 수행하였다. 표 6은 각 폴드의 평균 성과 표준편차를 요약한 결과이며, 그림 7은 폴드별 성능 변화를 시각적으로 나타낸 것이다.

표 6. Rolling-origin 교차검증 결과

지평선	MAE (평균±std)	RMSE	CRPS	Skill (RMSE)	PICP (보정)
단기	5.59±0.71	8.06±1.14	3.31±0.43	+0.113±0.022	0.775±0.053
3 - 24h	7.81±0.79	11.20±0.82	4.68±0.71	+0.499±0.069	0.763±0.067

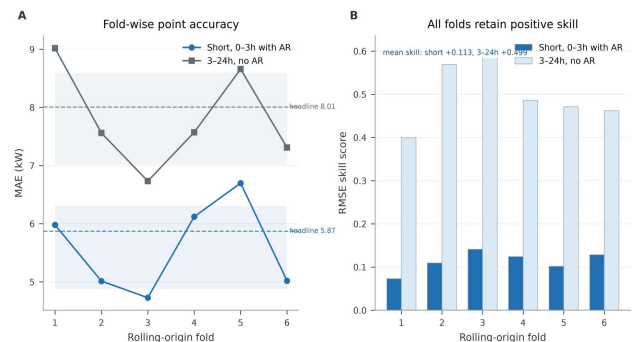


그림 7. 6-폴드 rolling-origin 교차검증 결과

교차검증 결과, RMSE skill은 0 - 3시간 단기 예측구간에서 0.113 ± 0.022 , 3 - 24시간 선행 예측구간에서 0.499 ± 0.069 로 나타났다. 모든 폴드에서 RMSE skill이 양의 값을 유지하였다는 점은 제안한 예측 파이프라인이 특정 시험 구간에만 의존하지 않고, 기준선 대비 일관된 개선 효과를 보였음을 의미한다.

단일 분할에서 산출된 주요 MAE는 0 - 3시간 단기 예측구간에서 5.87 kW, 3 - 24시간 선행 예측구간에서 8.01 kW였다. 이는 rolling-origin 교차검증에서 확인된 평균 MAE 범위인 단기 5.59 ± 0.71 kW, 3 - 24시간 선행 예측 7.81 ± 0.79 kW 내에 위치한다. 따라서 단일 시험 구간의 결과는 교차검증 결과와 크게 벗어나지 않으며, 전체 평가 결과를 대표하는 범위 안에 있는 것으로 판단된다.

IV. 결론

본 연구는 설비용량 120 kW급 단일 태양광 발전소를 대상으로 발전량 예측 정확도 향상을 위한 운영형 예측 파이프라인의 성능을 실증적으로 검증하였다. 실제 운영 데이터 자산을 기반으로 학습자료와 추론자료의 정합성을 점검하고, 청천지수 기반 예측 대상 설정, 다중 수치예보자료 보강, 예측 지평선 분리, 예측구간 보정 절차를 결합하여 단일 발전소 환경에서의 예측 성능을 평가하였다.

실험 결과, 제안한 파이프라인의 평균절대오차는 0 - 3시간 단기 예측 구간에서 5.87 kW, 3 - 24시간 선행예측 구간에서 8.01 kW로 나타났다. 두 예측 구간 모두 지평선별 기준선보다 향상된 성능을 보였으며, Diebold - Mariano 검정을 통해 이러한 성능 차이가 통계적으로 유의함을 확인하였다.

이러한 결과는 본 연구 대상과 같이 다중 수치예보자료가 확보된 단일 발전소 환경에서는 모델 구조의 복잡도뿐만 아니라 입력 예보자료의 품질, 학습자료와 추론자료의 정합성, 예측 지평선별 정보 가용성 설계가 예측 성능을 결정하는 중요한 요인임을 보여준다. 또한 CQR과 적응형 컨포멀 보정을 적용함으로써 점예측 성능과 함께 예측구간의 경험적 신뢰도를 평가할 수 있었다. 6-폴드 rolling-origin 교차검증에서도 제안한

파이프라인이 모든 폴드에서 기준선보다 일관되게 우수한 성능을 보였다.

다만 본 연구는 특정 지역의 단일 태양광 발전소를 대상으로 수행되었으므로, 발전소 규모, 설비 구성 및 기상 조건이 다른 환경에 연구 결과를 일반화하는 데에는 한계가 있다. 향후에는 위성 기반 일사량 및 구름 정보를 단기 예측 입력으로 활용하고, 지상 일사계 자료를 추가로 확보하며, 다수 발전소를 대상으로 모델의 전이 가능성을 검증할 필요가 있다.

감사의 글

“이 논문은 2026년도 과학기술정보통신부의 재원으로 과학기술사업화진흥원의 지원을 받아 수행된 연구임(1711198118, 지역과학기술성과실용화지원사업 (조선대학교))”

REFERENCES

- [1] 한국전력거래소, 「전력시장운영규칙(2025.2.12 시행)», <https://www.kpx.or.kr/> (accessed Jan., 20, 2026).
- [2] 한국전력거래소, 「전력시장운영규칙 현행 전문」 (2026), <https://marketrule.kpx.or.kr/> (accessed Jan., 20, 2026).
- [3] L. de Oliveira Santos, T. AlSkaif, G. C. Barroso, and P. C. M. de Carvalho, “Photovoltaic power estimation and forecast models integrating physics and machine learning: A review on hybrid techniques,” *Solar Energy*, vol. 284, Article 113044, Dec., 2024.
- [4] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, “pvlib python: A python package for modeling solar energy systems,” *Journal of Open Source Software*, vol. 3, no. 29, Article 884, Sep., 2018.
- [5] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785 - 794, San Francisco, CA, USA, Aug., 2016.
- [6] XGBoost Developers, “XGBoost Parameters: Learning Task Parameters-reg:quantileerro

- r」, <https://xgboost.readthedocs.io/> (accessed Jan., 20, 2026).
- [7] Y. Romano, E. Patterson, and E. J. Candès, “Conformalized quantile regression,” *Advances in Neural Information Processing Systems 32: Proceedings of the 33rd Annual Conference on Neural Information Processing Systems*, pp. 3543 - 3553, Vancouver, BC, Canada, Dec., 2019.
- [8] Y. Renkema, L. Visser, and T. AlSkaif, “Enhancing the reliability of probabilistic PV power forecasts using conformal prediction,” *Solar Energy Advances*, vol. 4, Article 100059, Apr., 2024.
- [9] V. Suresh, B. Sri Revathi, and J. M. Guerrero, “A non-parametric adaptive conformal inference based probabilistic hour-ahead solar PV power forecasting method,” *Scientific Reports*, vol. 16, no. 1, Article 11730, Apr., 2026.
- [10] V. Suresh, “Benchmarking transformer variants for hour-ahead PV forecasting: PatchTST with adaptive conformal inference,” *Energies*, vol. 18, no. 18, Article 5000, Sep., 2025.
- [11] G. Rohini, T. Mariprath, S. Al Bustanji, and I. Zaitsev, “A stacked Gradient Boosting - XGBoost ensemble with ridge meta-learner for accurate short-term solar PV power forecasting in smart grids,” *Scientific Reports*, vol. 16, no. 1, Article 16871, Apr., 2026.
- [12] Open-Meteo, 「Historical Forecast API」, <https://open-meteo.com/> (accessed Jan., 20, 2026).
- [13] I. Gibbs and E. J. Candès, “Adaptive conformal inference under distribution shift,” *Advances in Neural Information Processing Systems 34: Proceedings of the 35th Annual Conference on Neural Information Processing Systems*, pp. 1660 - 1672, Dec., 2021.
- [14] T. Gneiting and A. E. Raftery, “Strictly proper scoring rules, prediction, and estimation,” *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359 - 378, Mar., 2007.
- [15] F. X. Diebold and R. S. Mariano, “Comparing predictive accuracy,” *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 253 - 263, Jul., 1995.
- [16] D. I. Harvey, S. J. Leybourne, and P. Newbold, “Testing the equality of prediction mean squared errors,” *International Journal of Forecasting*, vol. 13, no. 2, pp. 281 - 291, Jun., 1997.
- [17] 기상청, 「수치모델 국지예보모델(IDAPS) 수치예보 자료」, <https://apihub.kma.go.kr/> (accessed Jul., 20, 2026).

저 자 소 개



임인선(정회원)

2004년 목포대학교 정보통신공학/신소재 학사 졸업.
2025년 조선대학교 과학기술융합학과 석사 졸업.
2026년 조선대학교 과학기술융합학과 박사 재학.

<주관심분야 : AIXOps, 인공지능응용, 시계열 데이터 분석>



손경우(정회원)

2013년 일본 東北大學校 전기통신공학 석사 졸업
2019년 전남대학교 행정학과 / 행정학 박사 졸업
2026년 조선대학교 과학기술융합학과 박사 재학

<주관심분야 : 인공지능응용, 모빌리티 자율주행 >



노순국(종신회원)

2000년 조선대학교 대학원 전자공학 공학박사
2002년 ~ 2004년 전북대학교 BK기금교수
2004년 ~ 2011년 호남대학교 이동통신공학과 조교수
2012년 ~ 2018년 조선아공대학교 전자과 조교수
2018년 ~ 현 재 조선대학교 자유전공학부 & 과학기술융합학과 (대학원) 부교수

<주관심분야 : 무선이동통신, 전파전파, IoT시스템, 인공지능응용, 신재생에너지 등>