

# 정밀 일사량 추정 전략

공간적 상관관계와 물리적 특성을 활용한 센서리스 모델

Team CHIDIES | 2025 태양광 발전량 추정 경진대회

KAIST 산업및시스템공학과 김수민

KAIST 건설및환경공학과 김채린

## Agenda

---

- 배경 및 목적
- 핵심 인사이트
- 모델 개발 로드맵
- 모델별 상세 내용 및 LB Score
- Discussion

## 배경: 센서리스(Sensorless) 추정의 필요성

---

- 소규모 태양광 발전소는 비용 문제로 일사량계나 모듈 온도 센서 같은 고가 장비 설치가 어려우나, 일사량 데이터는 발전 효율 관리와 전력망 안정화에 필수적임
- (핵심 목표) 환경 센서가 없는 발전소의 실시간 일사량(Target)을 인접 기상 관측소와 주변 발전소 정보를 활용해 정밀하게 추정하는 '가상 센싱(Virtual Sensing)' 기술 개발

# 데이터 분석의 3가지 관점 (Time, Weather, Space)

①

시간

## 주기성과 해상도의 불일치

- 일사량은 태양의 일주 운동과 계절 변화에 따른 명확한 주기성을 가짐
- 기상 데이터는 1시간 단위인 반면, 예측 타겟(일사량)은 5분 단위임
- 적절한 결측치 보간을 통한 해상도 불일치 해결이 필요함

②

기상

## 변동성의 원인

- 구름, 습도, 시정 등은 일사량을 감소시키는 요인임
- 기존 연구처럼 이론상 최대 일사량에 기상 조건에 따른 '감쇠율(Attenuation)'을 적용하는 물리적 도메인 지식을 반영해야 함

③

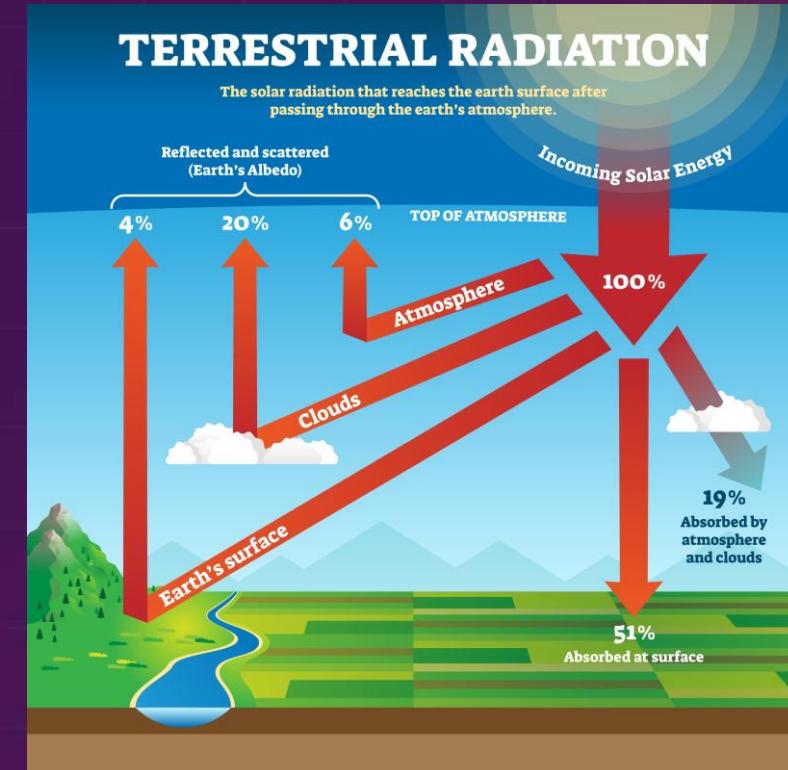
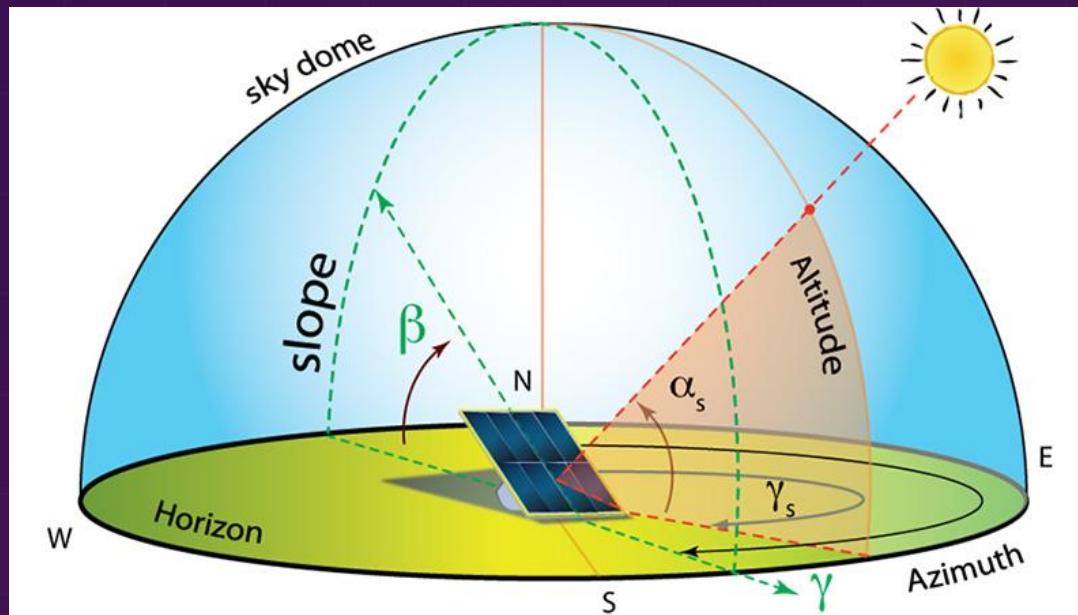
공간

## 정보의 원천 이동

- 본 과제는 센서가 없는 발전소의 '현재' 일사량을 예측하는 문제임
- 기상 데이터는 1시간 단위이며, 기상 관측소의 위치가 불명확하나, 지리적으로 인접한 발전소들은 정확한 좌표와 5분 단위 실측 일사량을 제공하고 있음

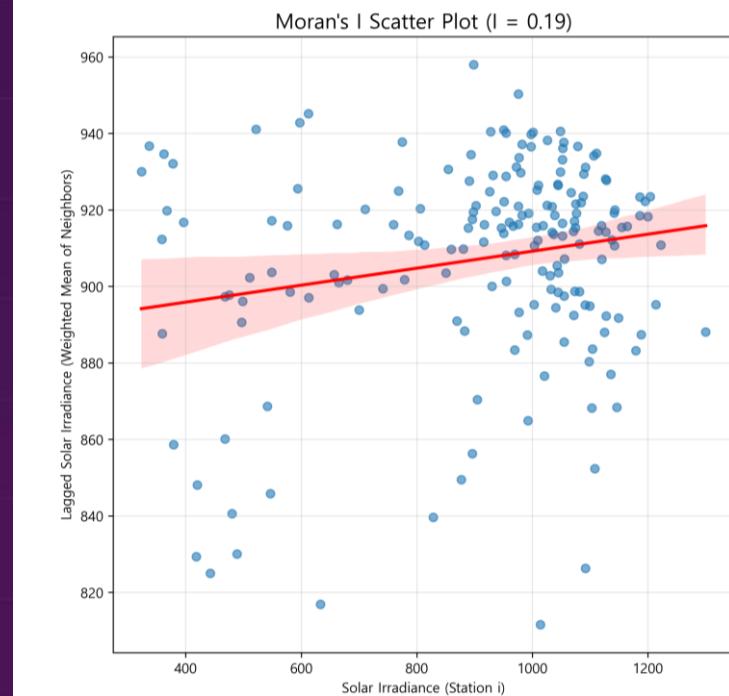
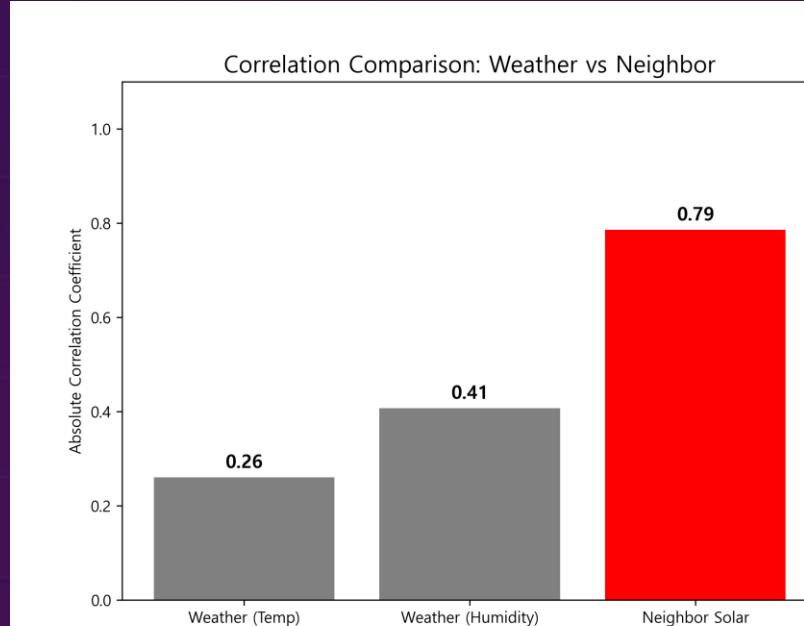
## (핵심 인사이트 1) 데이터의 물리적 특성을 반영하여 태양광이 차단되는 정도인 감쇠율을 예측하는 문제로 치환하여 접근

- 일사량 데이터는 무작위 시계열이 아니라 지구의 공전과 자전에 의해 결정되는 이론적 최대 일사량이라는 물리적 상한선이 존재함
- 구름이나 습도, 미세먼지 등으로 인해 태양광이 얼마나 차단되는지에 대한 감쇠율을 예측하는 문제로 치환함



## (핵심 인사이트 2) 기상 정보보다 인접한 발전소의 일사량이 더 중요한 정보를 가지고 있음

- 기상 정보는 1시간 단위로 주어지며, 기상 관측소와 얼마나 떨어져 있는지 알 수 없음
- 가까운 곳에 있는 발전소는 거의 동일한 기상 환경(구름 등)을 공유하므로 인접 발전소의 실측값이 기상 예보보다 훨씬 강력한 정보원이 됨
- Moran's I 분석 결과 0.19로 양의 상관관계를 보이며 일사량 데이터가 공간적으로 군집되어 있음을 확인함



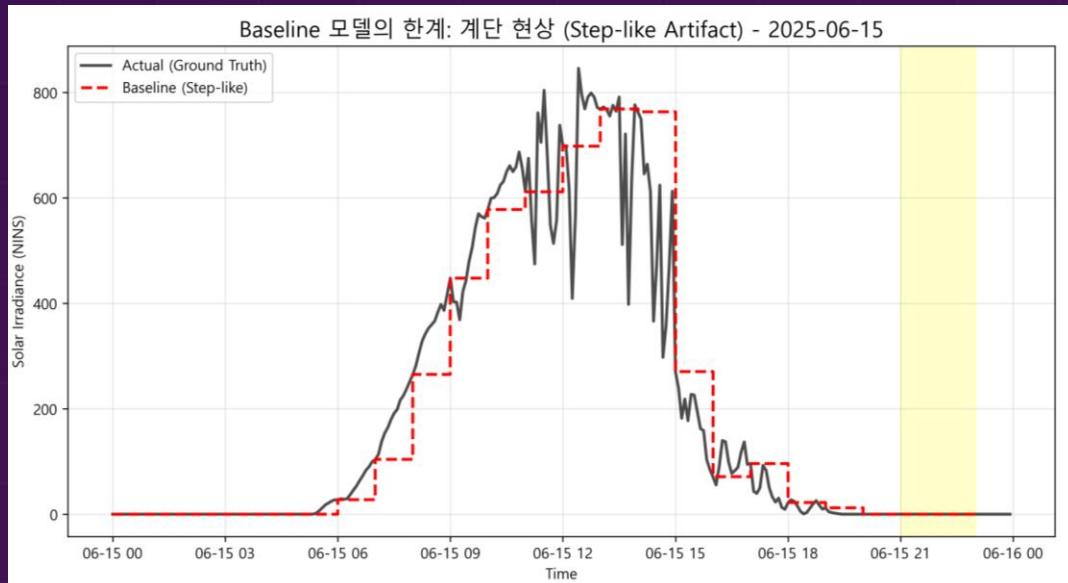
## (로드맵)



**(Step 00 & 01) bfill 방식은 정각부터 다음 정각 직전까지 같은 값을 예측하므로 낮은 성능을 보이기 때문에 선형 보간을 통해 데이터의 연속성을 회복시킴**

### Step 00

- (LB Score) 63.53
- (아이디어) 제공된 기상 정보만을 사용한 간단한 LightGBM 모델
- (전처리) 기상 데이터의 결측치를 bfill (backward fill)로 단순하게 채움
- (한계) 1시간 단위 기상 데이터로 인해 5분 단위 예측값이 계단 형태로 나타나는 등(e.g., 1시 5분부터 2시까지 예측값이 동일) 명확한 한계를 보임



### Step 01

- (LB Score) 53.83
- (아이디어) 기상 데이터는 급격히 변하기보다 **연속적으로 변할 것이라는 가정** 하에, 발전소별로 기상 데이터의 결측치를 **선형 보간(Linear Interpolation)**하여 정보의 손실을 줄임
- (개선점) 단순 bfill에 비해 시간의 흐름을 더 자연스럽게 반영하여 성능이 크게 향상됨

Step 1 MAE  
**53.83**

(▼ 9.7 개선)

## (Step 02) 일사량을 직접 예측하는 대신, 실제 일사량과 이론 일사량의 비율로 정의되는 감쇠율을 예측하는 모델을 개발함

### Step 02

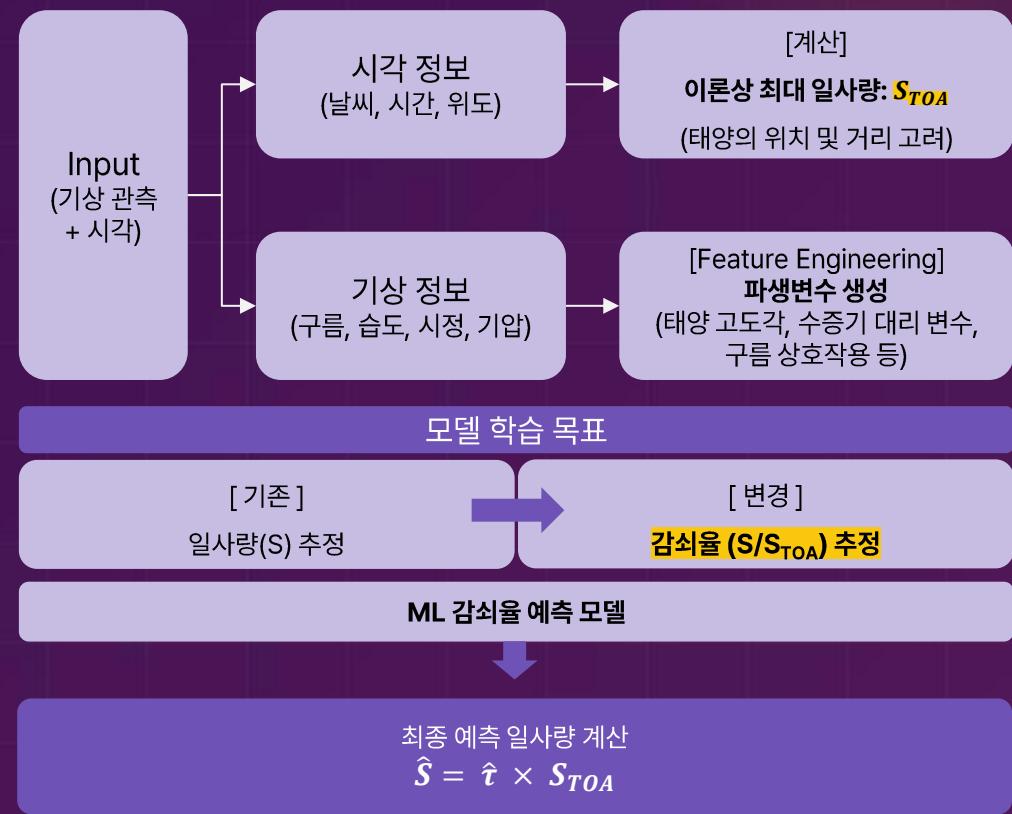
- (LB Score) 51.43
- (아이디어) 일사량을 직접 예측하는 대신, **실제 일사량 / 이론상 최대 일사량으로 정의되는 감쇠율(Attenuation Rate)을 예측**
- (이론상 최대 일사량) 위도, 경도, 시간을 통해 계산된, 대기 영향이 없을 때의 최대 일사량으로 모델은 기상 조건이 이 감쇠율에 어떤 영향을 미치는지 학습하게 되어, 물리적으로 더 타당한 예측을 수행함
- (피처 엔지니어링) 태양 고도각, 주기적 시간 특성(sin/cos) 등 고급 특성을 추가함

Step 2 MAE

**51.43**

(▼ 2.4 개선)

$$S = S_{TOA} \times \tau$$



## (Step 03 & 04) 공간적 자기상관성에 집중하여 예측 대상 발전소와 가까운 학습 발전소를 찾아 값을 예측하는 모델을 개발함

### Step 03 Simple K-nearest Neighbors 최근접 이웃 모델

- (LB Score) 48.74
- (아이디어) "같은 시간, 가까운 위치의 발전소는 비슷한 일사량을 가질 것이다." 라는 공간적 자기상관성에 집중하여 모델 개발  
=> 예측 대상 발전소와 가장 가까운 K개의 학습 발전소를 찾아, 그들의 일사량 값을 (거리를 고려하여) 가중 평균하여 예측함
- (결과) 기상 정보를 전혀 사용하지 않았음에도 불구하고, 이전 모델들보다 월등히 높은 성능을 기록함으로써 공간 정보가 이 대회의 핵심임을 확인함

### Step 04 Local Inverse Distance Weighting, IDW 역거리 가중 모델

- (LB Score) 45.35
- (아이디어) 최근접 이웃 모델을 개선한 모델로, 단순 평균이 아닌 거리가 가까울수록 더 높은 가중치를 부여하는 IDW 방식을 사용함  
=> 가중치  $\propto 1 / (\text{거리}^p)$
- (최적화) 이웃의 수(k)와 거리 가중치 지수(p)를 조절하며 최적의 조합을 탐색함

Step 3 MAE

**48.74**

Step 4 MAE

**45.35**

(▼ 3.39 개선)

## (Step 05) 기상 및 시간 특성만으로 일사량 추세를 예측하고 이때 발생하는 잔차를 공간적 특성으로 보간한 모델을 개발함

### 1. Trend Model

공간 정보를 제외한 기상/시간 특성만으로  
LGBM 모델을 학습시켜 일반적인  
Trend를 예측

### 2. Residual IDW

잔차를 계산 >> 이때 잔차는 모델이  
예측하지 못한 공간적 요인에 의한 오차로  
해석

### 3. Final Correction

전반적 추세에 국지적 오차를 더해 최종 예  
측값을 산출  
 $\text{예측} = \text{Trend} + \text{Local Error}$

Step 5 MAE

**46.21**

Trend Model이 충분히 튜닝되지 않아 순수 공간 모델(Step 4)보다 성능이 낮음

(Step 6) 역거리 가중 모델과 잔차 보간 모델을 3:1 비율로 합성한 모델을 개발함으로써 최종 성능 MAE 44.93을 달성함

Step 04  
Local Inverse Distance Weighting, IDW  
역거리 가중 모델

Step 05  
Regression with IDW Residuals  
잔차 보간 모델

3

1

Step 06  
Final Ensemble  
최종 모델

Step 6 MAE

**44.93**

(최종)



총 29.3% 성능 향상 달성

## 1. 기상 트렌드 모델 고도화

- 현재 기본적인 기본적인 LightGBM 사용
- 컴퓨팅 자원의 한계로 기상 트렌드 모델 성능을 더 끌어올리지 못함
- → **Autogluon 같은 최신 AutoML 프레임워크** 사용하여 트렌드 예측 성능 향상 가능

## 2. 지구통계학적 모델 도입

- IDW는 결정론적 방법으로 데이터의 분산 구조 등 통계적 특성을 반영하지 못함
- Regression Kriging 같은 **geostatistics 방법**으로 고도화 할 수 있음
- 배리오그램(Variogram) 분석을 통해 거리별 상관관계 감소 패턴을 정밀 모델링하여 예측 불확실성 감소

## 3. 시공간 통합 모델

- 현재는 데이터의 시간+기상 특성과 공간 특성을 분리하여 반영
- 발전소를 노드로, 지리적 관계를 엣지로 정의하여 **Spatiotemporal-GNN 같은 그래프 기반 딥러닝 모델**을 도입  
→ e.g. 현재의 모델은 가까운 거리의 이웃만 반영하는데, 과거의 북서쪽 이웃의 일사량 반영하도록 어텐션 학습 가능

# EoD

Q & A

Team CHIDIES | 2025 태양광 발전량 추정 경진대회