Toronto Metropolitan University

ADVANCED SOLAR POWER FORECASTING

Accurate solar forecasting without irradiance data
First validated approach to remove the most expensive input while maintaining accuracy.
MAE 0.033 | R^2 0.693

LOWER COST

WIDER ACCESS

SAME ACCURACY

Peer-reviewed; presented at the Winter Simulation Conference (WSC 2024). Recognized by Innovation Boost Zone (TMU) Esch Competition — Stage 1 Winner

ACCURATE SOLAR FORECASTING WITHOUT SOLAR IRRADIANCE DATA

Project Overview

Solar energy is becoming a core pillar of clean power, but to use it well we must know how much electricity a site will produce in the next hour or day. Our project delivers accurate, scalable solar-power forecasts using only widely available data (weather, location, and time). By engineering the right features and combining strong machine-learning models, we match the accuracy of conventional approaches while removing their cost and complexity. The result: forecasting that is affordable to deploy anywhere reliable weather data exists — supporting grid operators, solar farms, and emerging markets alike.

Problem Statement

Most forecasting methods treat solar irradiance data — the measured sunlight reaching the surface (W/m^2) — as indispensable. To obtain it, operators depend on pyranometers and other ground sensors (purchase, calibration, cleaning, replacement) and/or satellite irradiance products (specialized instruments, licenses, latency). This infrastructure is expensive, maintenance-heavy, uneven in coverage, and difficult to scale across new sites. Because irradiance has long been seen as the "gold-standard" input, many projects absorb these costs or forgo high-quality forecasting altogether.

Solution

We replace irradiance inputs with a purpose-built ML pipeline that learns sunlight patterns from meteorological, geographic, and temporal signals. The workflow includes temporal/cyclic feature engineering, solar-geometry features (elevation/azimuth), smoothing and polynomial interactions, feature selection (RFE), and an ensemble of LightGBM and XGBoost combined by a meta-model. This carefully orchestrated set of techniques closes the information gap left by irradiance, delivering the same forecasting accuracy at much lower cost and with global scalability. Validated on real-world data from 12 U.S. DoD sites (25 regions) and peer-reviewed (WSC 2024), our approach makes high-quality solar forecasting sensor-free, accessible, and ready to deploy.

WHY IRRADIANCE DATA IS EXPENSIVE

What it is: Sunlight at the surface (GHI/DNI/DHI) that most forecasts rely on.

Satellites: Imagers/radiometers → irradiance derived, not direct. Needs cloud/aerosol models, paid data feeds, heavy compute, and has latency.

Ground sensors: Pyranometer (GHI), pyrheliometer + sun-tracker (DNI), shaded/POA sensors, datalogger/comms, power, mounting.

Upkeep: Frequent cleaning, desiccant changes, tracker alignment, firmware, calibration every 12–24 months, QA/QC and downtime handling.

Cost snapshot: Site capex often \$10k–\$30k+ (tracker setups drive cost) + ongoing calibration, field visits, data/storage → scales to six-figure lifetime for fleets.

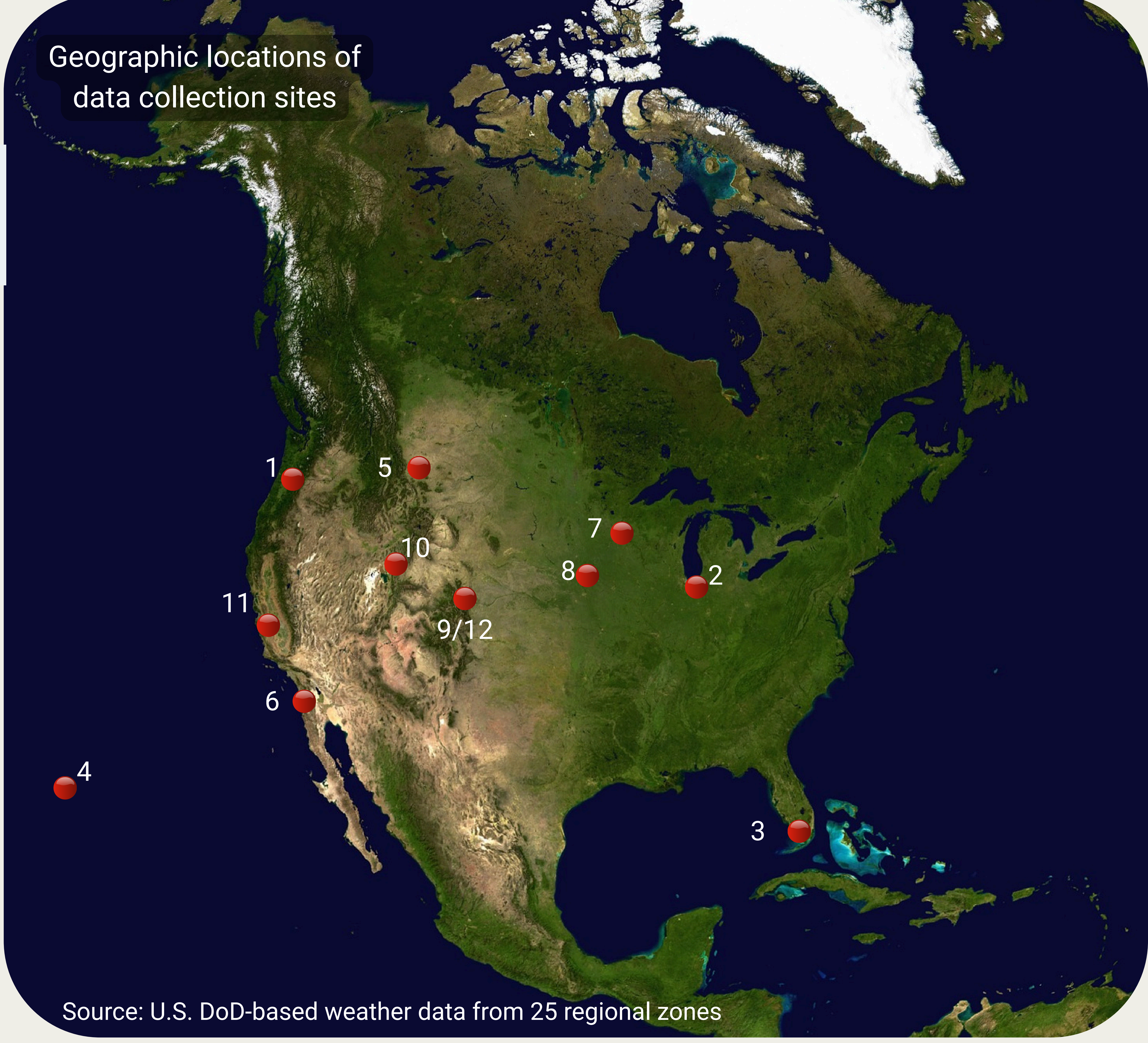
Our edge: We forecast without irradiance sensors, using widely available weather + geographic + time features—same accuracy, far lower cost & complexity.

REAL-WORLD VALIDATION ACROSS STRATEGIC U.S. LOCATIONS

Data collected from 12 geographically diverse U.S. Department of Defense installations.

Real-world conditions.
Real performance.
Sensor-free forecasting

tested and proven.



Toronto Metropolitan University

OUR NOVEL MACHINE LEARNING ARCHITECTURE

A unique combination of feature engineering, solar geometry, and ensemble modeling — purpose-built for low cost, sensor-free forecasting.

Input Data

Weather forecasts, date/time info. Sources: Government APIs or weather services.

STEP 01

Feature Engineering

Rolling averages, smoothing, polynomial terms, and solar position

STEP 02

Feature Selection

Recursive Feature Elimination (RFE) to retain the most relevant features.

STEP 03

Base Learners

Two parallel models: LightGBM and XGBoost.

STEP 04

Meta-Model

Linear Regression combines outputs of base models to improve robustness.

STEP 05

Output

Accurate solar power prediction — scalable, sensor-free

STEP 06

Toronto Metropolitan University

TRADITIONAL VS. OUR SOLAR FORECASTING MODEL A CLEAR LEAP FORWARD

Aspect	Traditional Model	Our ML-based Model
Solar Irradiance Sensors	**Required** (Satellite+Ground)	**Not Needed**
Infrastructure Cost	**High** (Sensors, Ground Stations)	**Minimal**
Scalability	**Limited** by Hardware	**Globally Scalable**
Forecasting Accuracy	Good	**Same Accuracy**
Setup Time	**Complex**	**Instant**
Maintenance	**Ongoing** Sensor Upkeep	**Zero Sensor Maintenance**