# Using Machine Learning to Assess Solar Radiation Data Quality

Braedyn McBroom[1,2], Christopher Redmond[2], Dan Regier[2]

1 Department of Geosciences, University of Arkansas, 2 Kansas Mesonet, Kansas State University

## GOAL: Create a tool that assesses the integrity of solar radiation sensor data.

Reliable solar radiation measurements are crucial for various applications, including estimating evapotranspiration for agricultural water management, particularly as the Ogallala aquifer continues to deplete. These measurements also support vital decision-making tools, such as animal comfort indices and fire danger calculations.

The Kansas Mesonet, which uses Apogee CS 300 and CS 301 Silicon Pyranometers**,** relies on accurate solar radiation sensor data to provide these insights. With the recent addition of cameras at select Mesonet stations (**Figure 1**), the opportunity arises to use machine learning to evaluate solar radiation sensor performance. By comparing expected shortwave radiation under clear sky conditions against measured solar radiation readings (Allen 1996), this research aims to develop a simple quality assessment tool.

### Kansas Mesonet stations with cameras (June 2024)




Land cover data from the Kansas Biological Survey

Figure 1

### Cloudy



### Sunny

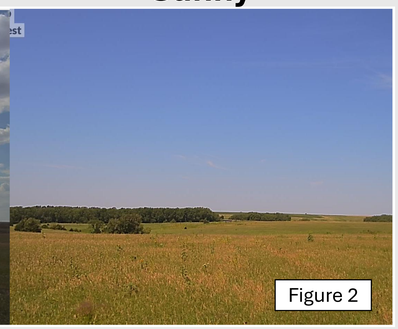


Figure 2

## Step 1: Build a machine learning model

- Download images to train the model:
  - 9AM on the 1st of each month, 10AM on 2nd, …, 3PM on the 7th
  - For stations with less than 1 year of images, download extra
- Sort images in to “sunny” and “cloudy” subfolders
- Scikit-learn Python library – supervised machine learning algorithm
  - 80% of images used to train, 20% used to test
  - Model attempts 20 combinations saving the best performing pair

## Step 2: Find sunny days with the model

- From the date a station began archiving images:
  - Apply the model to the 9AM, 12PM, & 3PM images
  - If an result is cloudy at one of the images, move to the next day without continuing
  - A day is “sunny” if it is sunny at all three hours

| Date | 9AM | 12PM | 3PM |
|---|---|---|---|
| 05/01/2024 | Cloudy | | |
| 05/02/2024 | Sunny | Sunny | Sunny |
| 05/03/2024 | Sunny | Sunny | Cloudy |
| 05/04/2024 | Sunny | Cloudy | |
| … | … | … | … |
| 01/12/2025 | Sunny | Sunny | Cloudy |

## Step 3: Measure the sensors’ accuracies

- On sunny days:
  - Collect hourly observed solar radiation from sunrise to sunset
  - Calculate expected clear sky solar radiation for the same hours
  - Compare observed and expected solar radiation with statistics like root mean square error (RMSE), absolute percent error, Pearson’s correlation coefficient, and Wald’s t-test (example in **Figure 3**)

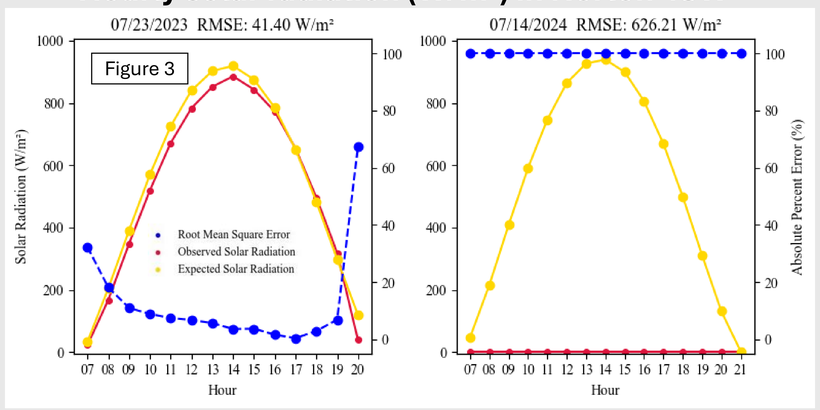
## Step 4: Flask-app website (QR Code)

- Create website using Python, the machine learning model, a server with Ubuntu, and HTML
- Displays station images, model prediction (sunny or cloudy), Mesonet observations, and sunny day trends
- http://164.92.124.250/

## Example Station – Norton 4SW

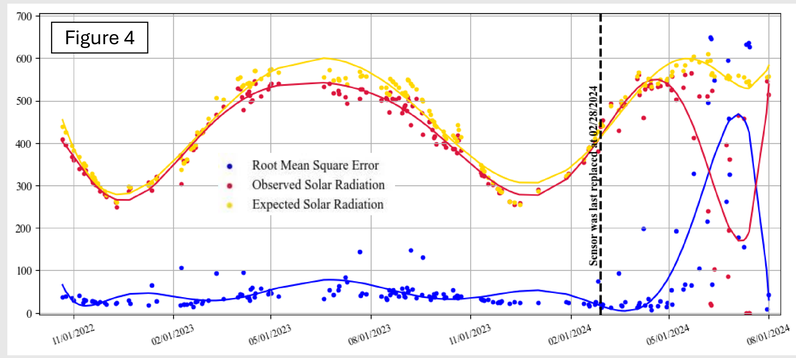
For most stations that had a sensor replacement since image archival began, relationships between expected and measured solar radiation on clear sky days did not significantly change. However, in Norton 4SW, significant discrepancies were identified where sensor readings showed incorrect zero values under sunny conditions after the pyranometer was replaced in February 2024 (see **Figures 3, 4, and 5**).

### Hourly solar radiation (W/m²) in Norton 4SW



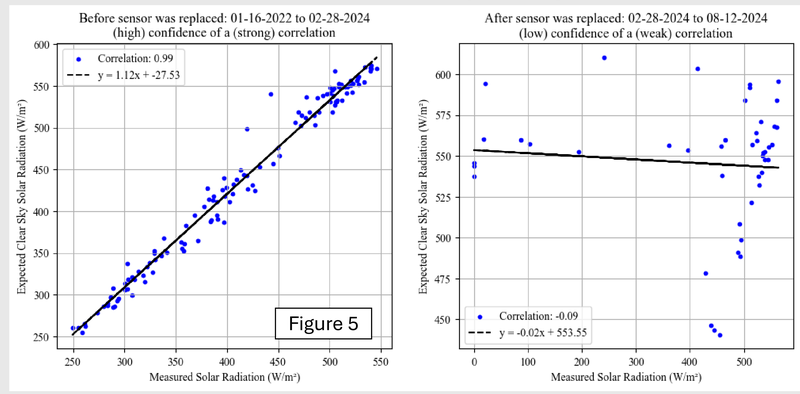

Both July 23rd, 2023 and July 14th, 2024 appeared to be sunny throughout the entire day. However, the RMSE of observed & expected solar radiation in 2024 was 15x larger than in 2023, suggesting the new sensor may be falsely reading zero.

### Time series of solar radiation (W/m²) in Norton 4SW




After the pyranometer was replaced, the frequency of large errors increased, showing that the new sensor likely had a defect.

### Correlation of solar radiation (W/m²) in Norton 4SW




Before the sensor was replaced, expected and measured solar radiation were highly correlated (R=0.99). Afterwards, they were not (R=-0.09).

## Limitations

- Usefulness of website is reliant on the machine learning model accuracy
  - Varies across stations due to camera angle, haziness, ground cover, etc.
  - Much less accuracy for stations with newly installed cameras
- Small complications such as calibration or defects are not easily detected
- Time-consuming, nonautomatic image sorting & new station adding
- Lacks a simple result

## Usefulness

- Machine learning model can be deployed for all stations with cameras
  - Performs well at identifying clear skies and snow-covered ground
- Clear sky solar radiation measurements compare well with the calculated clear sky solar radiation, usually
- Website can effectively identify major issues, such as faulty sensors

## Sources

Allen, R. G., 1996: Assessing integrity of weather data for reference evapotranspiration estimation. *Journal of Irrigation and Drainage Engineering*, (**2**), 97-106.

Kansas Biological Survey, 2017: 2015 Land Cover – Level 1. Accessed August 2024, https://kars-geoplatform-ku.hub.arcgis.com/pages/kansas-land-cover

Kansas Department of Agriculture, 2023: Kansas Agriculture. Accessed August 2024, https://www.agriculture.ks.gov/kansas-agriculture.

Kansas Mesonet, 2021-2024: Kansas Mesonet Historical Data. https://mesonet.k-state.edu/rest/