

# Machine Learning Application for Improving Cloud Detection and Phase Determination over Sunglint Regions for Geostationary Satellites

Cecilia Wang<sup>1</sup>, William L. Smith, Jr.<sup>2</sup>, David Painemal<sup>2</sup>, Sarah Bedka<sup>1</sup>, Rabindra Palikonda<sup>1</sup>, Thad Chee<sup>3</sup>, Douglas Spangenberg<sup>1</sup>

<sup>1</sup> Analytical Mechanics Associates, Inc., Hampton VA <sup>2</sup>NASA Langley Research Center, Hampton VA <sup>3</sup>ADNET Systems, Inc., Hampton VA  
cecilia.fleeger@nasa.gov



## Introduction

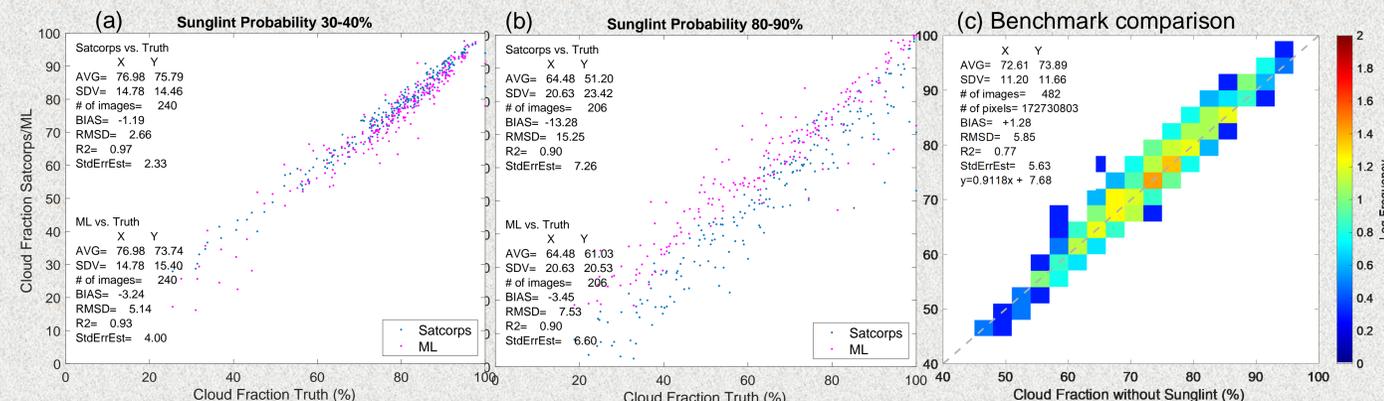
- Satellite-derived cloud properties are useful for a wide range of weather and climate applications.
- Retrieving accurate cloud properties from satellite imagers in sunglint regions is challenging due to the low contrast between cloud and ocean surface reflectance, leading primarily to the under detection of low-level clouds which can affect the accuracy of other level-3 derived cloud properties.
- Traditional cloud masking approaches are tuned to help minimize sunglint errors, but errors remain large.
- The primary objective of this study is to improve cloud detection accuracies in sunglint using machine learning (ML) applied to Geostationary satellite data.

## Summary

- A legacy cloud detection approach is significantly affected by sunglint, particularly when the sunglint probability exceeds 40%.
- We successfully mitigated sunglint-induced artifacts in cloud mask detection across all GEO satellites by employing a sunglint ML algorithm.
- Validation of the algorithm, based on cloud fraction analyses, demonstrated a substantial improvement in detection accuracy, reducing the missing cloud detection rate from over 20% to less than 5%.
- This algorithm is being integrated into the CERES Edition 5 and SatCORPS package to enhance operational processing for GEO satellites.

## Validation Through Cloud Fraction Analyses

- We compared geo-located cloud fractions derived from the standard product and the corresponding ML algorithm-enhanced product, with the standard product from another satellite free of sunglint affect.
- The nearest neighbor distance method is applied to ensure the quality of geo-collocation.



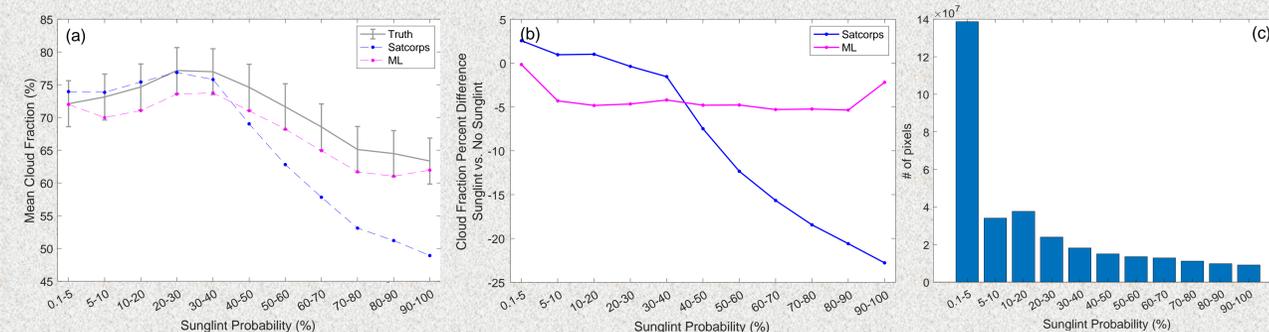
**Figure 2:** Cloud fraction of SatCORPS standard product vs. sunglint ML-enhanced product compared to truth cloud fraction for sunglint probability 30-40% (a), and 80-90% (b); Benchmark cloud fraction comparison (sunglint probability 0.1-10%) from both satellites (c).

## Datasets and Methods

- **GOES-16 and GOES-18 GCC SatCORPS Products:** Full disk Level 1b cloud products at 3 km resolution for October 2024.
- **Cloud Properties Retrieval:** NASA SatCORPS algorithm, integrated with the ML module.
- **Composite Optimal ML Training Datasets:** Generated by the ML pre-processing package, utilizing product files from the preceding 4-hour period.

- **GEOS-5 Forward Processing (FP) Products:** Provided by NASA's Goddard Space Flight Center (GSFC) Global Modeling and Assimilation Office (GMAO).
- **AI/ML methodology:** The K-Nearest Neighbor (KNN) approach leverages the infrared (IR) radiances to extrapolate cloud detections beyond sunglint-affected spatial and temporal domains into sunglint regions.

- **Panel (a):** Sunglint doesn't affect cloud detection significantly if the sunglint probability < 40% as biases and standard deviations are within the range of the ones from the truth dataset (Panel d).
- **Panel (b):** For sunglint probability between 80% and 90%, the SatCORPS standard cloud fraction exhibits a bias of -13% and a standard deviation of 15%. In contrast, the sunglint ML-enhanced cloud fraction aligns closely with the truth dataset, demonstrating improved accuracy.
- **Panel (c):** Validation against the truth dataset shows an uncertainty of  $\pm 6\%$ , attributed to cloud shadows and the parallax effect caused by differing viewing geometries between two satellites.



**Figure 3:** Mean cloud fraction from the SatCORPS standard package (blue), the sunglint ML-enhanced algorithm (magenta) vs. the truth dataset (grey). **Panel (a):** Cloud fraction as a function of sunglint probability; **Panel (b):** Cloud fraction percentage deviation from the truth cloud fraction; **Panel (c):** Histogram of valid pixels distributed across each sunglint probability category.

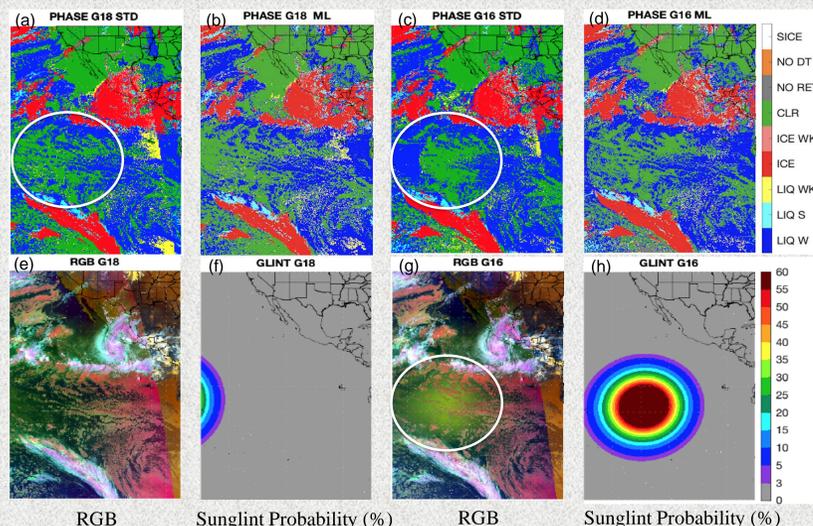
- SatCORPS standard cloud fraction exhibits a low bias when sunglint probability exceeds 40%, with the bias increasing proportionally to the sunglint probability, reaching a maximum of over 20% missing cloud coverage.
- In contrast, the cloud fraction derived using the sunglint ML algorithm shows minimal impact from sunglint, with a deviation of approximately 5% from the truth cloud fraction, well within the  $\pm 6.0\%$  uncertainty range of the truth dataset.

## References

- [1] Minnis, P., L. Nguyen, R. Palikonda, P. W. Heck, D. A. Spangenberg, D. R. Doelling, J. K. Ayers, W. L. Smith, Jr., M. M. Khaiyer, Q. Z. Trepte, L. A. Avey, F.-L. Chang, C. R. Yost, T. Chee, S. Sun-Mack, "Near-real time cloud retrievals from operational and research meteorological satellites", Proc. SPIE 7107, Remote Sensing of Clouds and the Atmosphere XIII, 710703 (13 October 2008); <https://doi.org/10.1117/12.800344>
- [2] Trepte, Q. Z., P. Minnis, S. Sun-Mack, C. R. Yost, Y. Chen, Z. Jin, G. Hong, F.-L. Chang, W. L. Smith, Jr., K. Bedka, T.L. Chee, 2019: Global cloud detection for CERES Edition 4 using Terra and Aqua MODIS data. IEEE Trans. Geosci. Remote Sens., vol. 57, no. 11, pp. 9410-9449, Nov. 2019, doi: 10.1109/TGRS.2019.2926620.

## Acknowledgments

This research was supported by the NASA Clouds and the Earth's Radiant Energy System (CERES) program and a Satellite Needs Working Group initiative to enhance the use and availability of NASA satellite data products to support Earth System Model development.



**Figure 1:** Example of observations w/o sunglint on October 21, 2024. **Top Panels:** Scene ID (or cloud phase) determined with the standard algorithm (STD) and the ML-enhanced (ML) for GOES-16 and GOES-18 satellites. **Bottom Panels:** Corresponding RGB imagery and sunglint probability maps from the two satellite observations.

- **Comparison of GOES-16 and GOES-18 cloud phases:**
  - **Panel c:** (GOES-16 standard cloud phase) is affected by sunglint (Panel h & g), resulting in missing and false cloud detections within the sunglint region (highlighted by white circle).
  - **Panel a:** (GOES-18 standard cloud phase), free of sunglint (Panel f & e), serves as the reference for truth data.
- **Validation Approach:**
  - GOES-16 and GOES-18 full-disk cloud products (October 2024) are used to validate the algorithm, as sunglint does not simultaneously affect both satellites in most scenarios.
  - Sunglint probabilities are categorized into 11 bins.
  - Cloud fraction is derived from valid pixels within each sunglint probability category.