Machine-Learning Snowfall Rate Retrieval

Sungmin Park
Mentor: Yulan Hong

Objectives

Estimate snowfall rate (SFR) from passive-microwave observations using machine-learning methods.

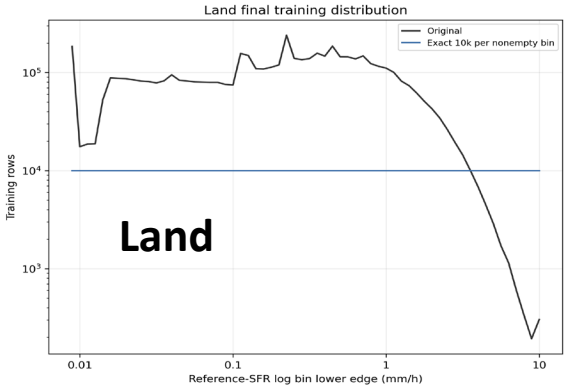
Models and Data

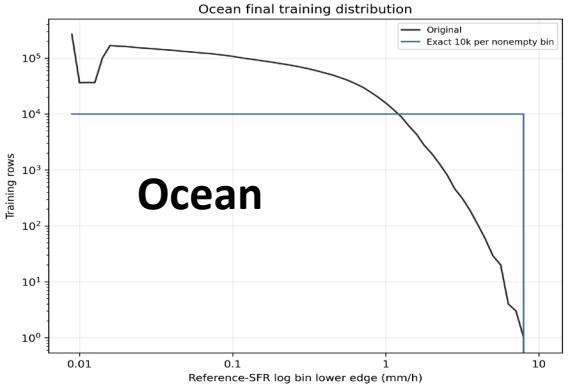
Satellite / scope: NPP ATMS; global retrieval
Reference SFR:
- Land: CPR+ERA5 + ST4
- Ocean: CPR+ERA5

ML models: XGBoost (primary), LightGBM, residual MLP, and TabFM
Feature sets: surface-specific Top 30

Data and Predictors

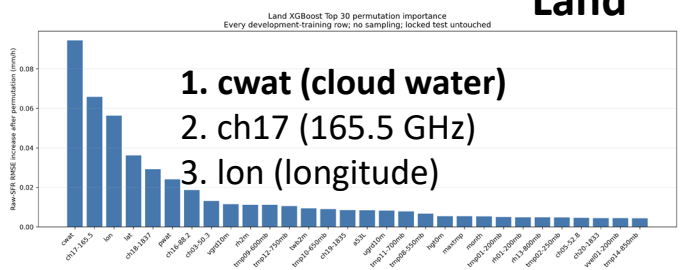
SFR distributions





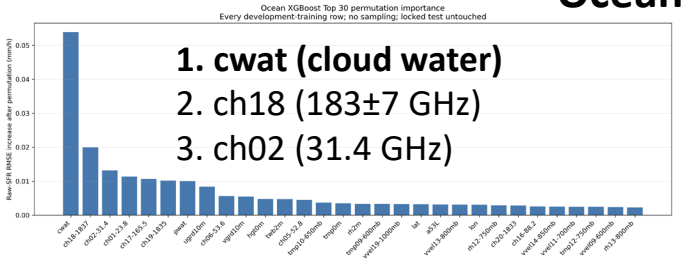
Feature importance

Land



1. cwat (cloud water)
2. ch17 (165.5 GHz)
3. lon (longitude)

Ocean



1. cwat (cloud water)
2. ch18 (183±7 GHz)
3. ch02 (31.4 GHz)

EARTH SYSTEM SCIENCE INTERDISCIPLINARY CENTER

Model Statistics

Sungmin Park
Mentor: Yulan Hong

Evaluation Protocol

- Held-out years: 2013, 2016, 2020, and 2022
- ST4 uses 2016/2020/2022; CPR includes all four years
- Feature ranking and resampling use training rows only
- Test metrics: correlation, mean bias, and RMSE

Overall Statistics

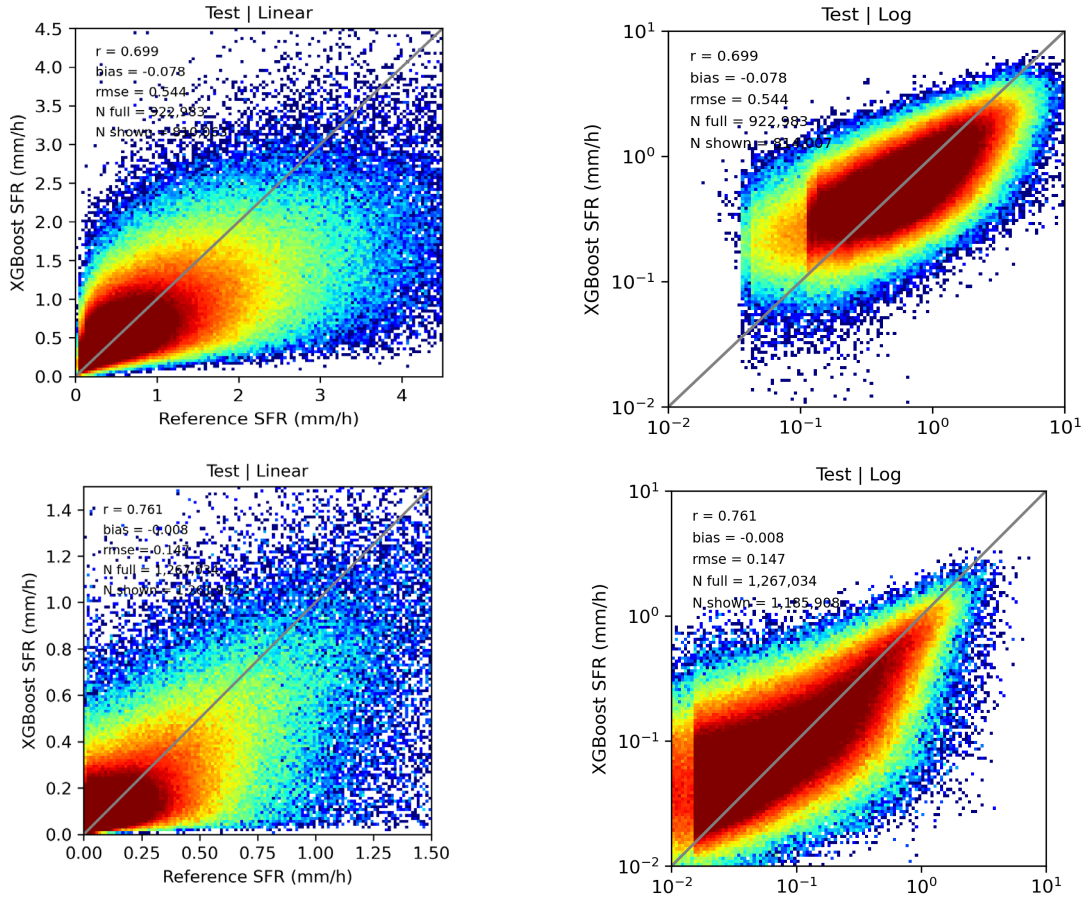
- ST4 land: r=.699 | bias=-.078 | RMSE=.544
- CPR land: r=.657 | bias=-.005 | RMSE=.184
- CPR ocean: r=.761 | bias=-.008 | RMSE=.147

Good overall agreement on independent held-out data

Remaining limitation: high SFR is underestimated

EARTH SYSTEM SCIENCE INTERDISCIPLINARY CENTER

Held-Out Predictions



Snowfall Cases and Conclusions

Sungmin Park
Mentor: Yulan Hong

January 26, 2026 Case

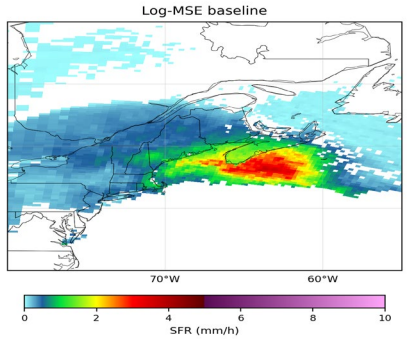
- Apply land/ocean XGBoost by surface type
- Compare baseline with exact-uniform resampling
- Evaluate the remaining heavy-SFR underestimation

Conclusions

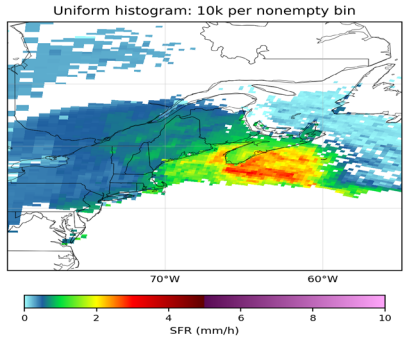
- Baseline XGBoost performs best overall
- Uniform resampling increases intensity but worsens held-out r/RMSE and adds positive bias
- Heavy-SFR underestimation remains the main limitation

NPP 2026-01-26 Case

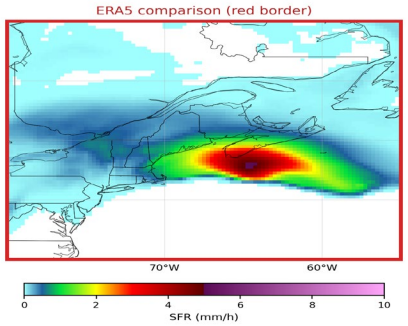
Baseline XGBoost



Uniform-resampled XGB



ERA5 comparison



EARTH SYSTEM SCIENCE INTERDISCIPLINARY CENTER