Data-driven approximation of downward solar radiation flux based on all-sky optical imagery using machine learning models trained on DASIO dataset.

Vasilisa Koshkina, Mikhail Krinitskiy,
Nikita Anikin, Mikhail Borisov

Motivation

Flux estimation in literature:

- physics-based modeling (needs detailed physical properties of clouds and aerosols)
- parameterizations-based modeling (needs only standard cloud properties)

Motivation:

- cheaper downward shortwave (SW) radiation flux estimation
- investigation of flux dependence on structural characteristics of clouds

Goal:

Improve accuracy of existing parameterizations of downward SW radiation fluxes

Assumption:

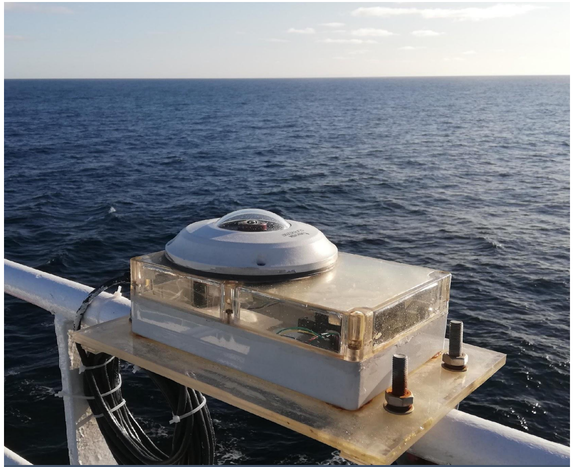
An all-sky photo contains complete information about downward SW radiation

Originality:

Use machine learning (ML) models for approximating downward SW radiation flux

Equipment

radiometer
KIPP&ZONEN CNR-1



cloud-camera
«SAIL CLOUD V.2»

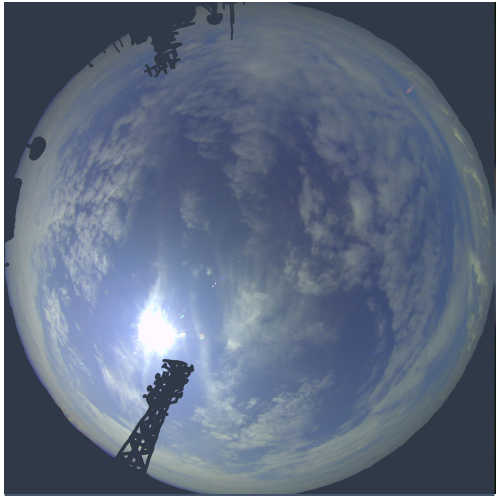
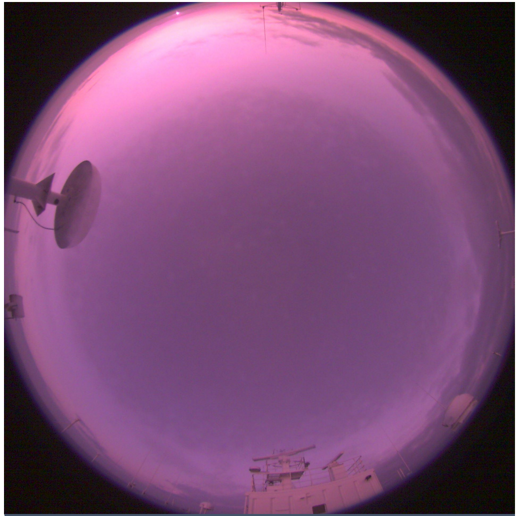


photo with mask

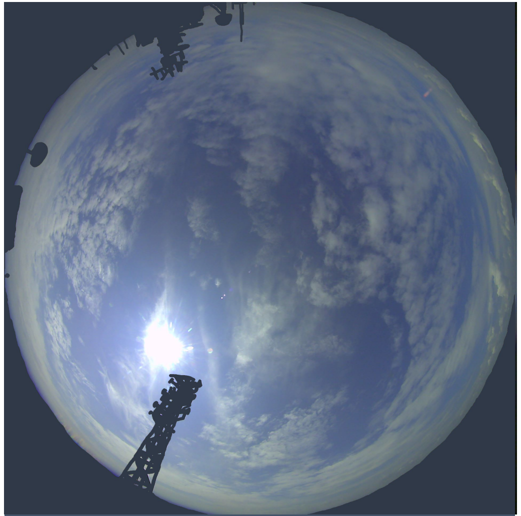
more than 7’000 €

less than 1’500 €

Pre-processing


“Bad” photo at <5 W/m^2

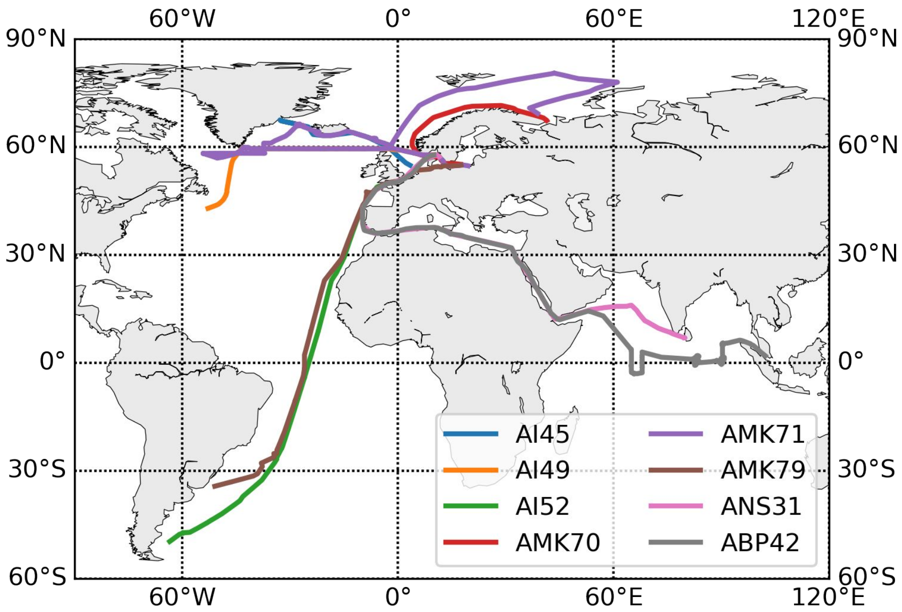

Mask


“Good” photo with mask

Data

- over 2'000'000 all-sky photo from ocean surface:
  The Dataset of All-Sky Imagery over the Ocean (DASIO)[1]

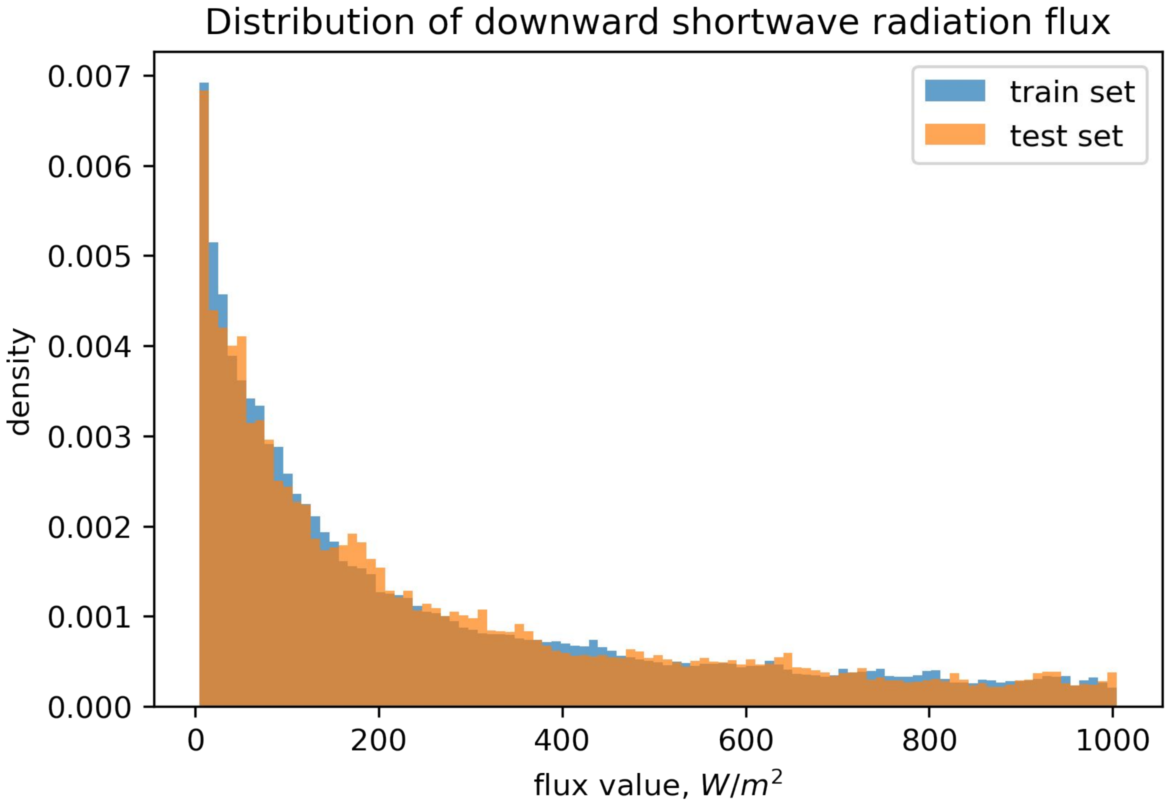
- resolution 1920*1920
- auto white balance
- auto brightness adjustment
- rare images considered outliers
- RGB-channels



Map of cruises of the research vessels

[1] Krinitskiy et al. "On the Generalization Ability of Data-Driven Models in the Problem of Total Cloud Cover Retrieval." Remote Sensing. 2021; 13(2):326.

Target values distribution



$$w_i \sim d_i$$

$$w_i^* = (w_i - 1) \cdot \alpha + 1$$

Methods

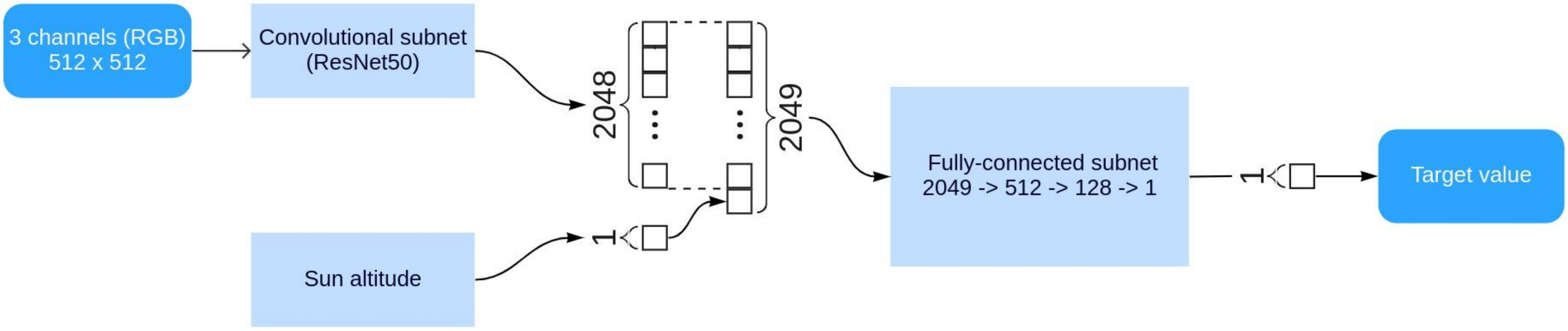
- **Classic ML models** (applied to pre-processed data):
  - linear models: *Linear Regression*
  - ensemble models: *Random Forest* and *Gradient Boosting*

163 features
max, min values
mean, variance
skewness, kurtosis
percentile set
sun altitude

Model	Best parameters
Linear Regression	-
Gradient Boosting	max_depth=10, n_estimators=500
Random Forest	max_depth=32, n_estimators=289

Methods

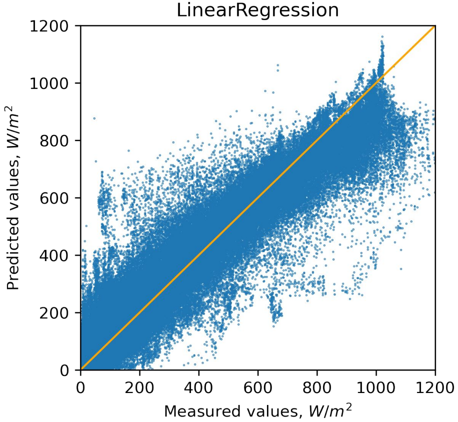
- **End-to-end ML approach** (applied directly to photo):
  - *Convolutional Neural Network (CNN)* with heavy images augmentation (brightness, gaussian noise)

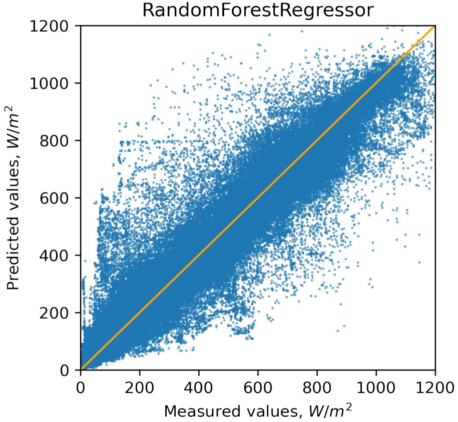


Structure of CNN

Value mapping diagrams

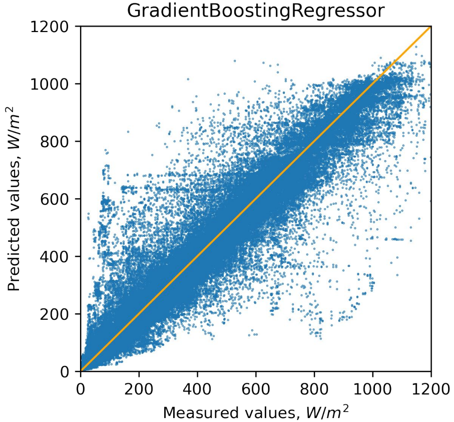
RMSE = 84 W/m^2

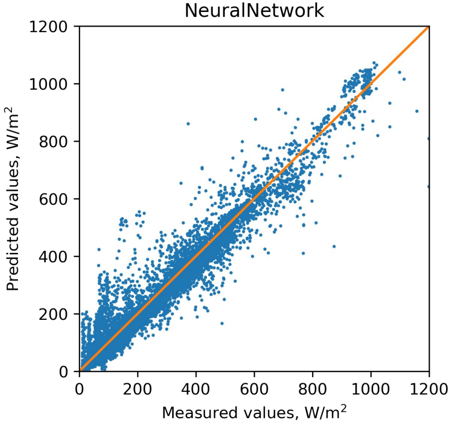




RMSE = 62.1 W/m^2

RMSE = 53.5 W/m^2





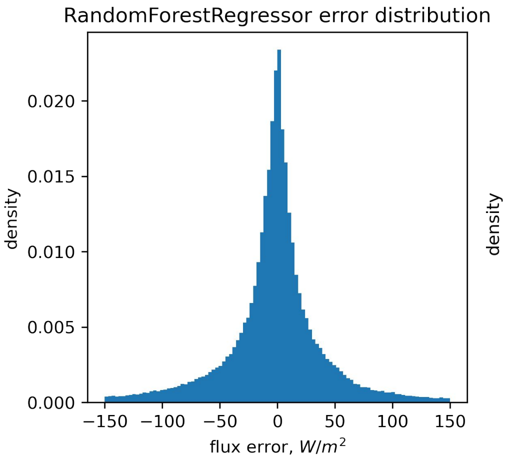
RMSE = 39.2 W/m^2

Results

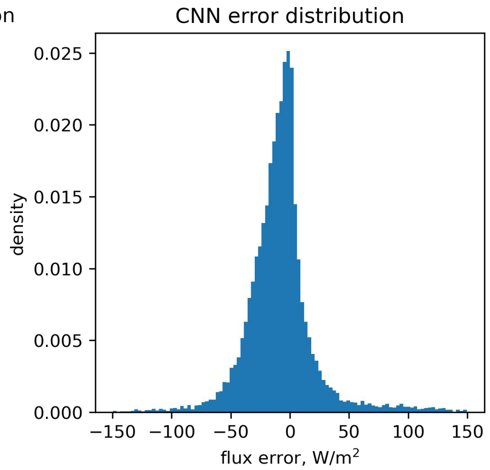
Model	RMSE, W/m^2
Linear Regression	84
Gradient Boosting	53.5
Random Forest	62.1
CNN	**39.2**

Parameterization	RMSE, W/m^2
Dobson–Smith	78.2 (38 – 116)
LVOAMKI[1]	61.9 (26 – 115)

Classic ML approach



End-to-end CNN approach



[1] Aleksandrova et al. "An improvement of parametrization of short-wave radiation at the sea surface on the basis of direct measurements in the Atlantic." Russian Meteorology and Hydrology, 2007. № 4 (32). c. 245–251.

Conclusion

- One may estimate downward SW radiation flux directly using all-sky imagery of clouds
- Quality of **CNN** (RMSE = 39.2 W/m^2) is better compared to existing parameterization LVOAMKI (RMSE = 61.9 W/m^2);
- Convolutional Neural Networks are capable of estimating downward SW radiation flux based on clouds visible structure

Technical means

NVIDIA DGX Station V100