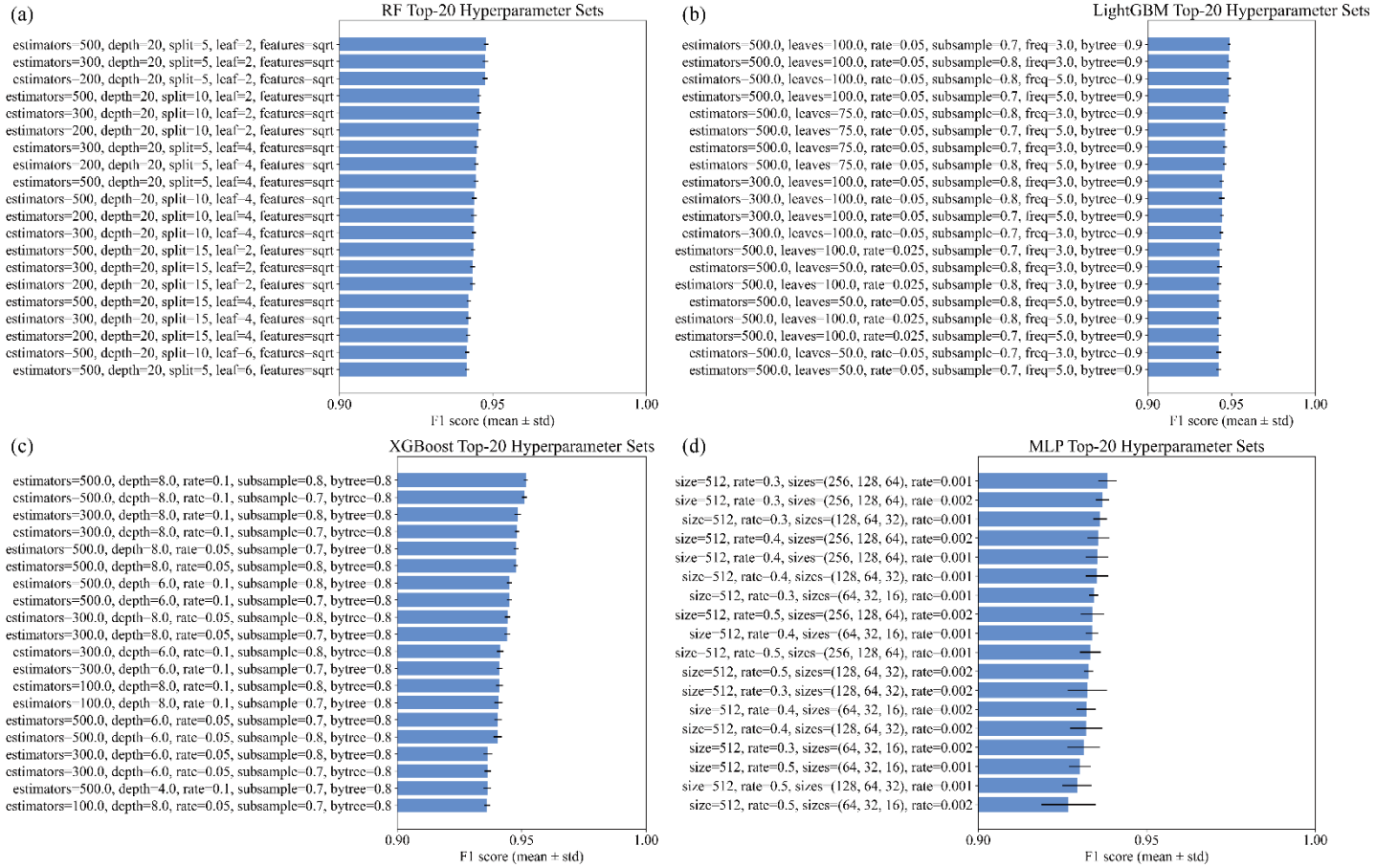


Figure S1. The F1 score of top-20 hyperparameter sets by RF (a), LightGBM (b), XGBoost (c) and MLP (d) with grid search method and five-fold stratified cross-validation.

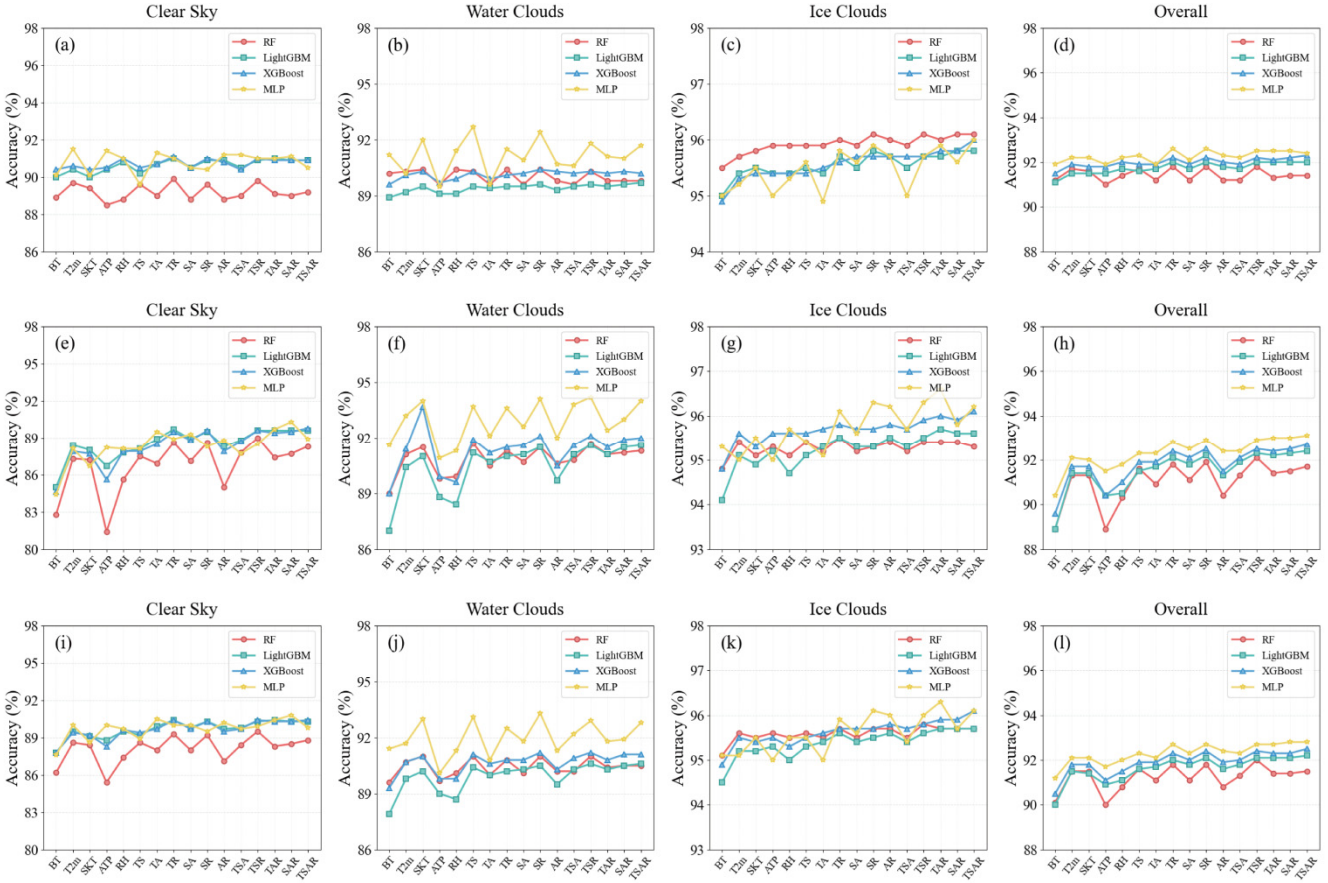


Figure S2. The accuracy of difference models with meteorological factors by RF, LightGBM, XGBoost and MLP under clear sky, water clouds and ice clouds and overall detection in daytime (top), nighttime (middle) and all day (bottom)

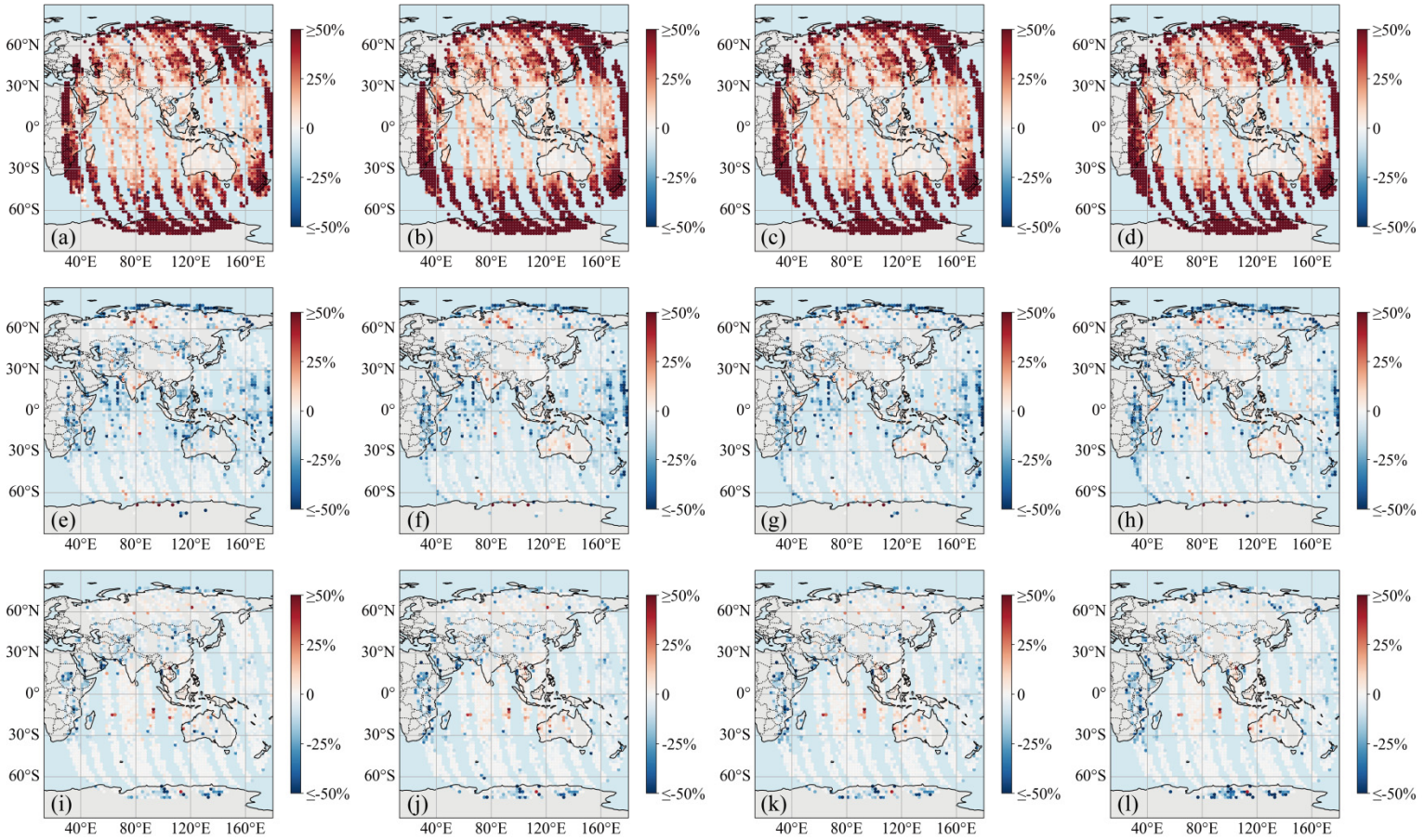


Figure S3. Accuracy differences in detection of clear sky (a, b, c, d), water clouds (e, f, g, h) and ice clouds (i, j, k, l) between TRAS model by RF, LightGBM, XGBoost, MLP and FY-4A L2 cloud mask products during daytime.

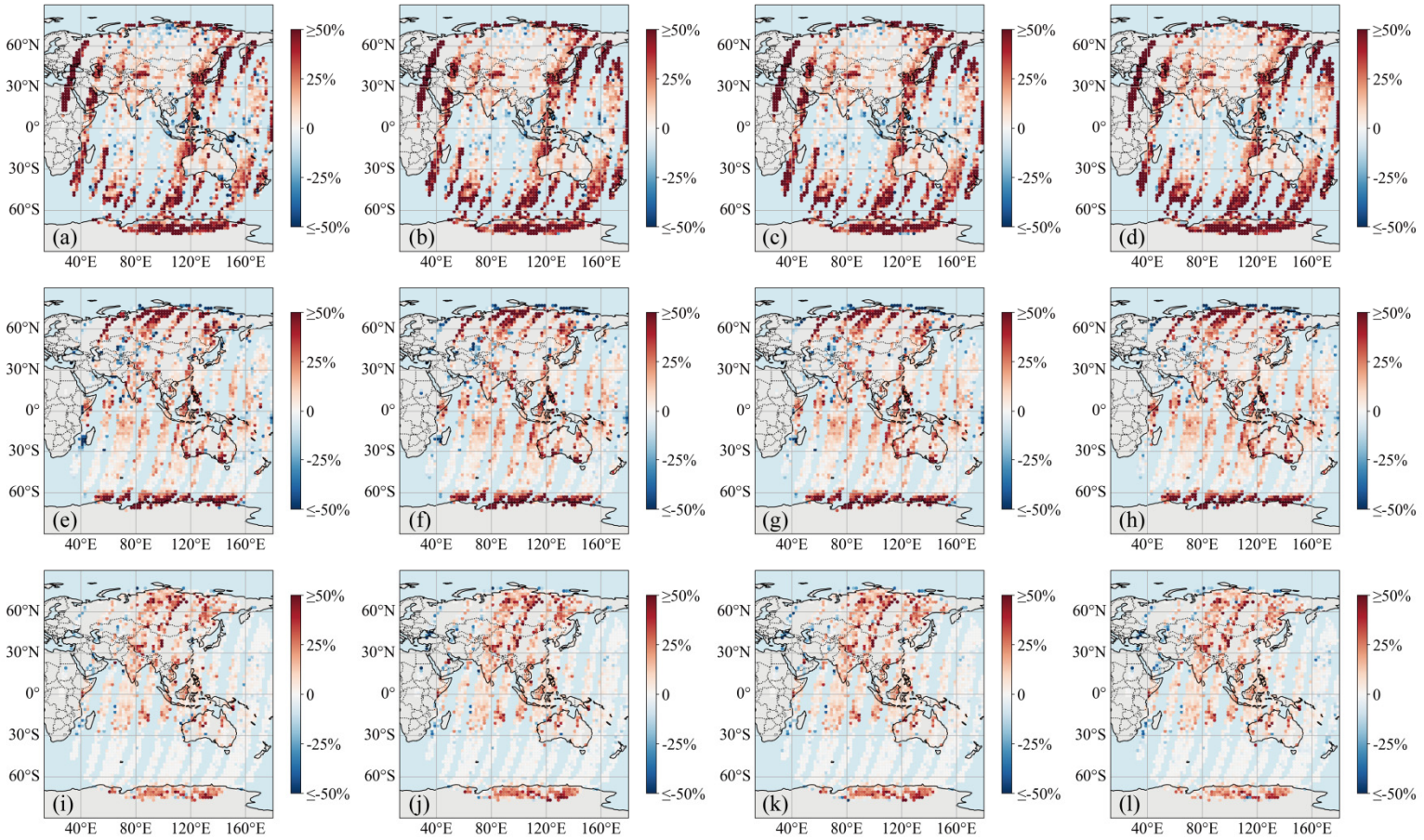


Figure S4. Accuracy differences in detection of clear sky (a, b, c, d), water clouds (e, f, g, h) and ice clouds (i, j, k, l) between TRAS model by RF, LightGBM, XGBoost, MLP and FY-4A L2 cloud mask products during nighttime.

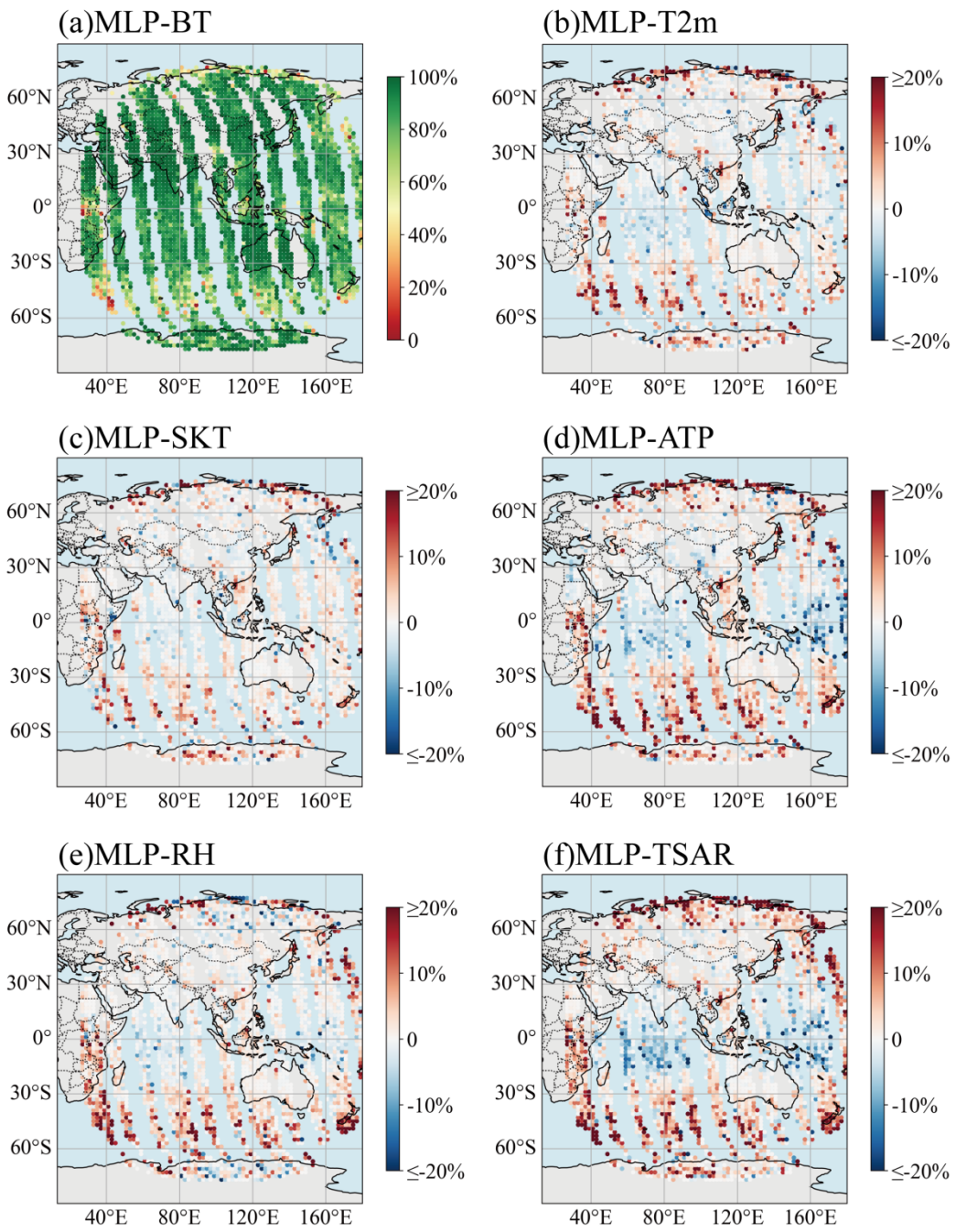


Figure S5. Distribution of daytime clear sky detection accuracy for (a) BT model, and the accuracy differences between (b) T2m, (c) SKT, (d) ATP, (e) RH, (f) TSAR and BT by MLP.

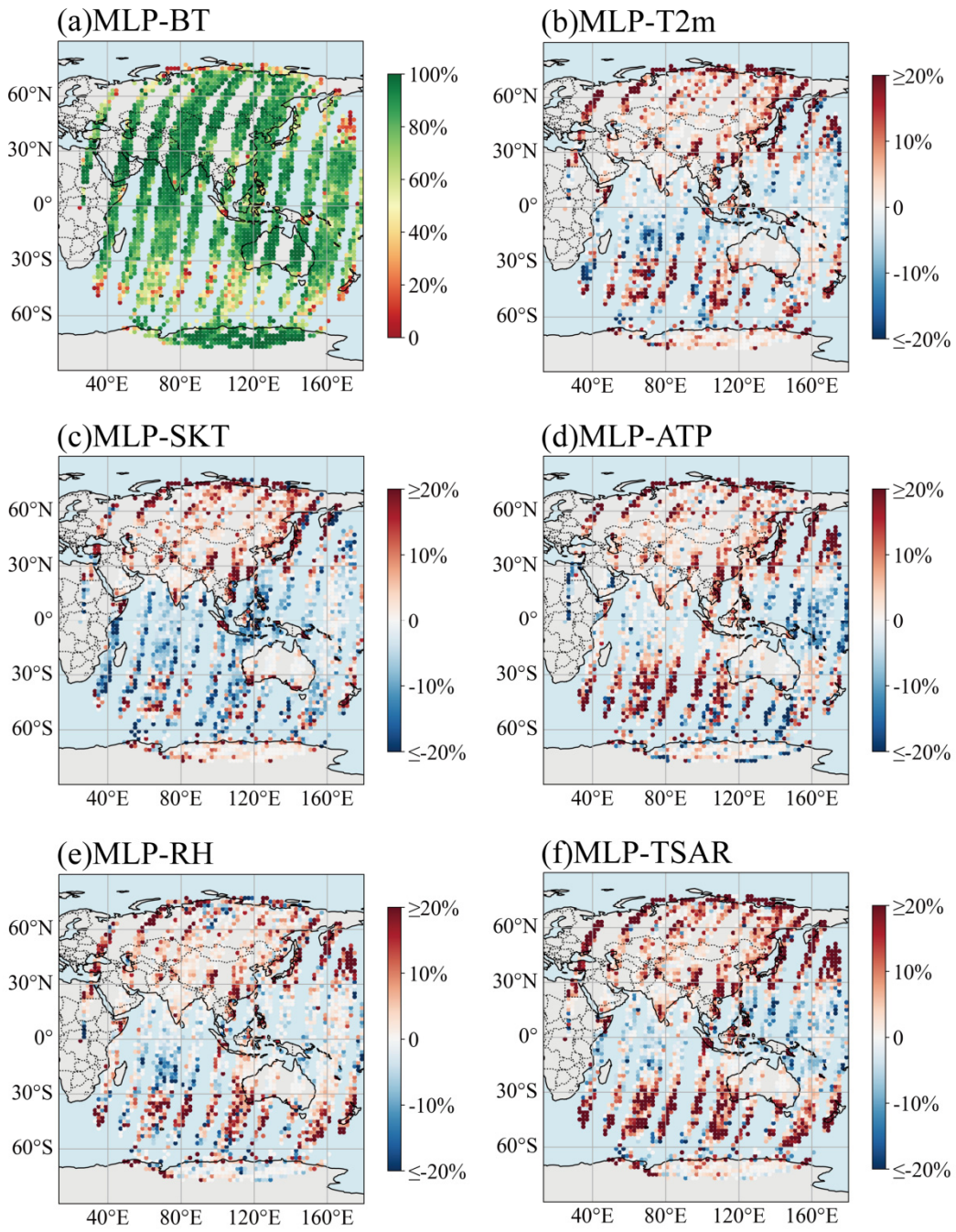


Figure S6. Distribution of nighttime clear sky detection accuracy for (a) BT model, and the accuracy differences between (b) T2m, (c) SKT, (d) ATP, (e) RH, (f) TSAR and BT by MLP.

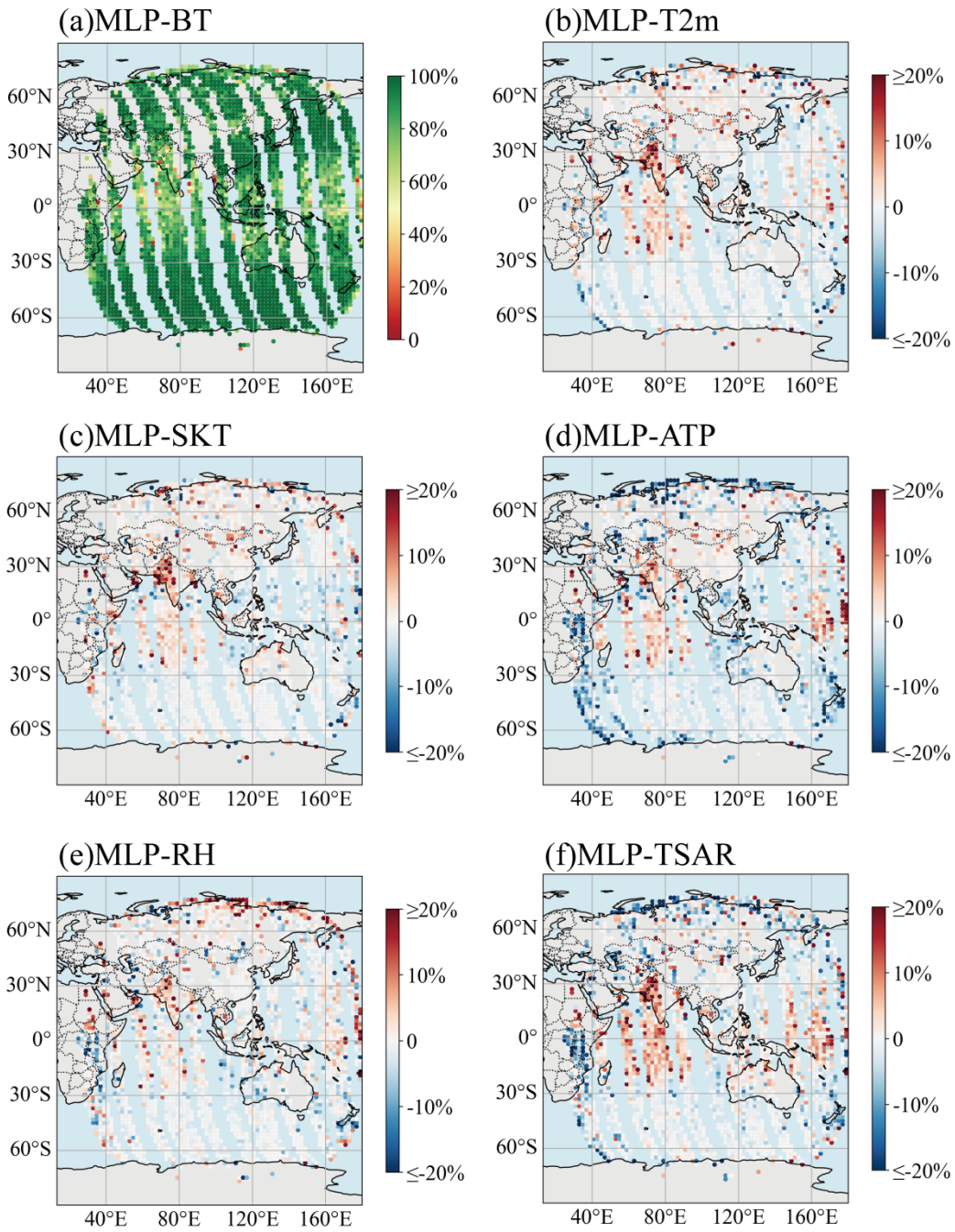


Figure S7. Distribution of daytime water cloud detection accuracy for (a) BT model, and the accuracy differences between (b) T2m, (c) SKT, (d) ATP, (e) RH, (f) TSAR and BT by MLP.

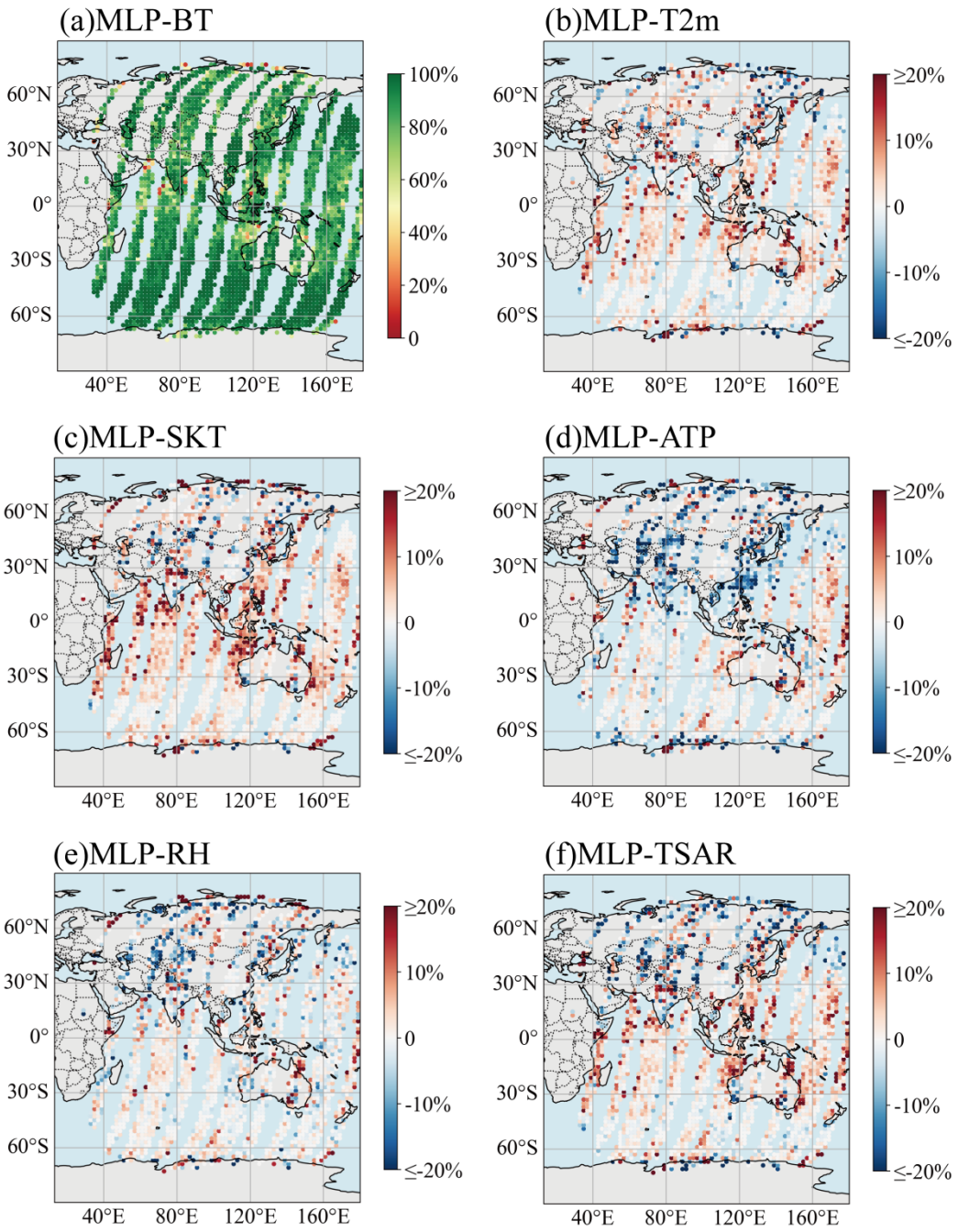


Figure S8. Distribution of nighttime water cloud detection accuracy for (a) BT model, and the accuracy differences between (b) T2m, (c) SKT, (d) ATP, (e) RH, (f) TSAR and BT by MLP.

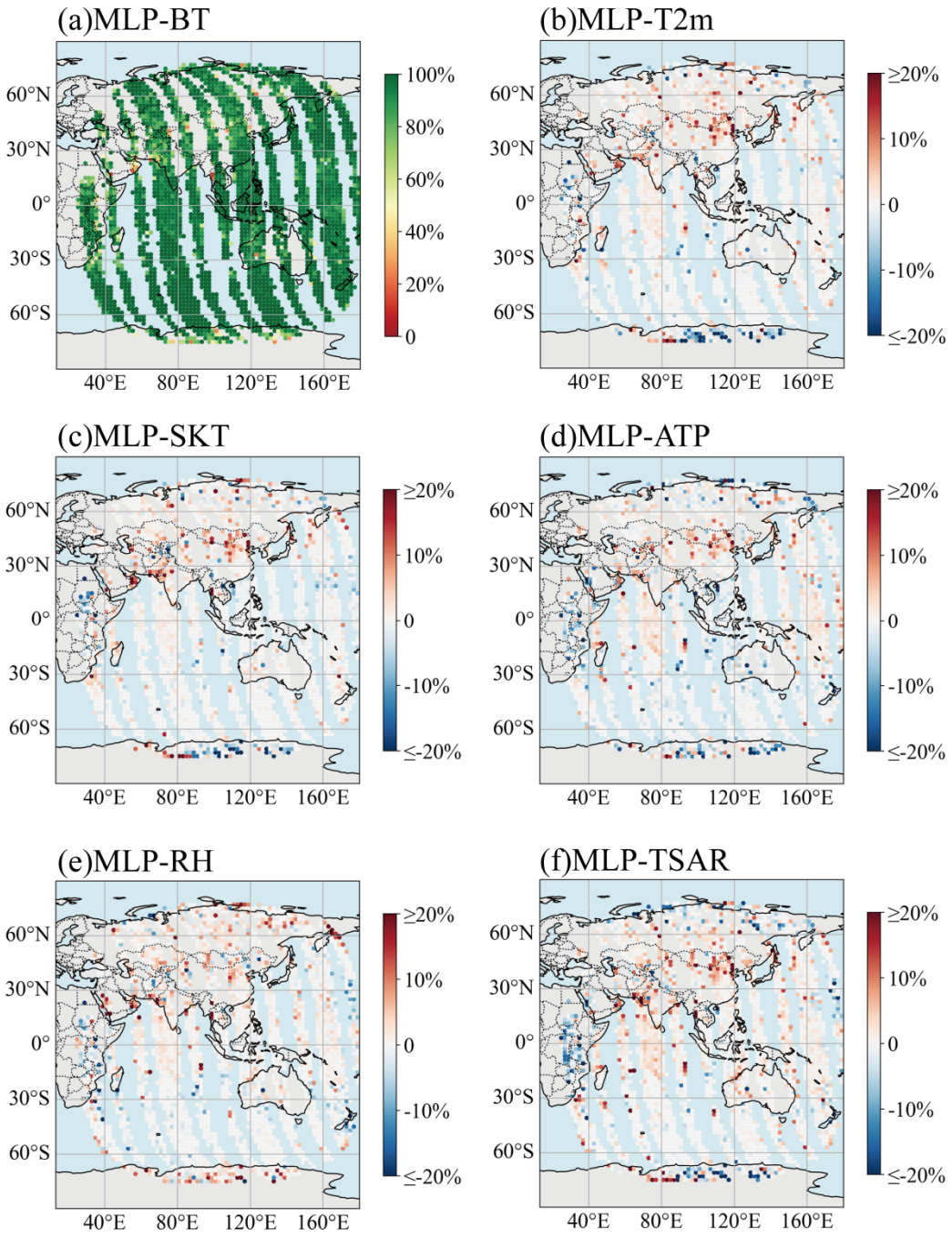


Figure S9. Distribution of daytime ice cloud detection accuracy for (a) BT model, and the accuracy differences between (b) T2m, (c) SKT, (d) ATP, (e) RH, (f) TSAR and BT by MLP.

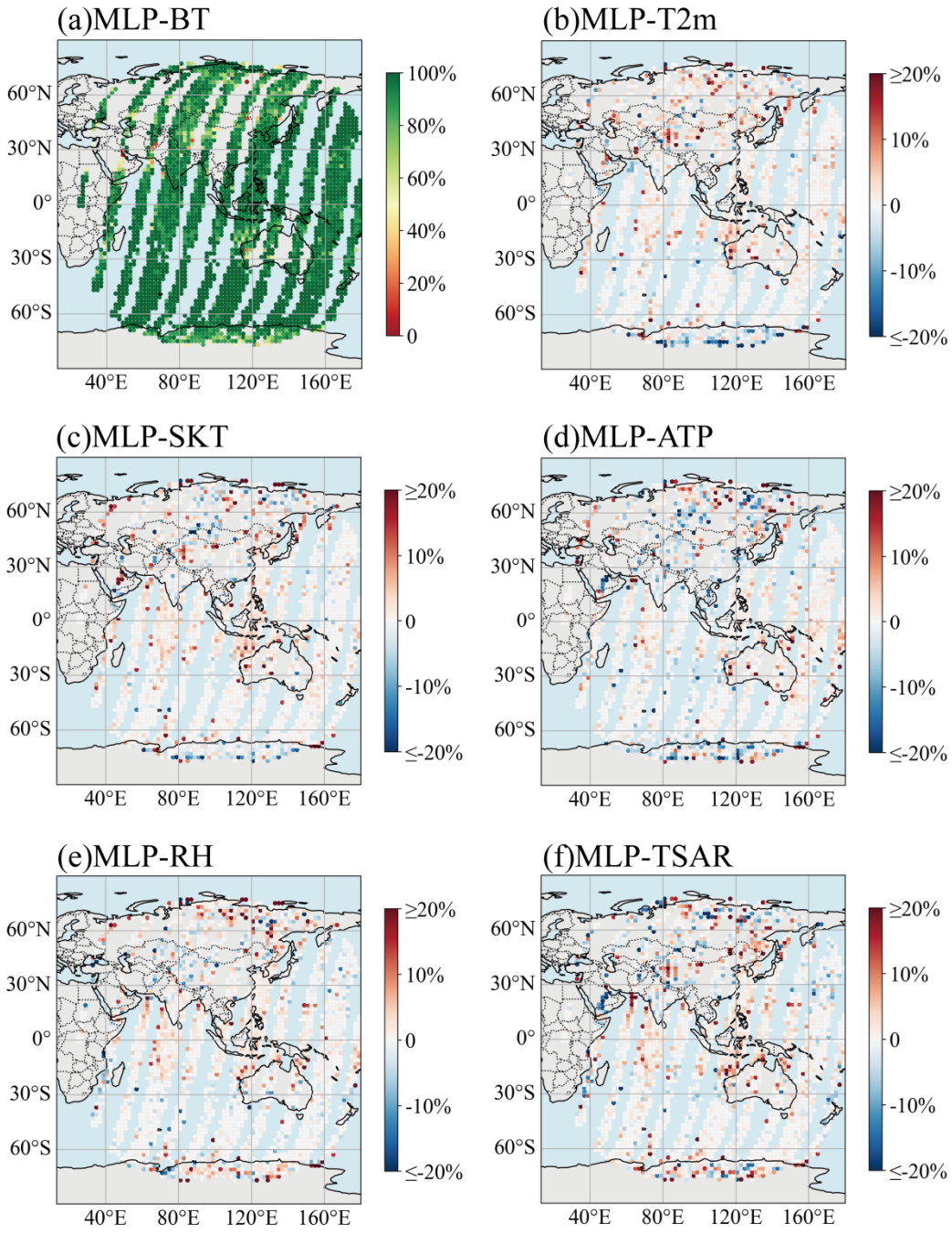


Figure S10. Distribution of nighttime ice cloud detection accuracy for (a) BT model, and the accuracy differences between (b) T2m, (c) SKT, (d) ATP, (e) RH, (f) TSAR and BT by MLP.

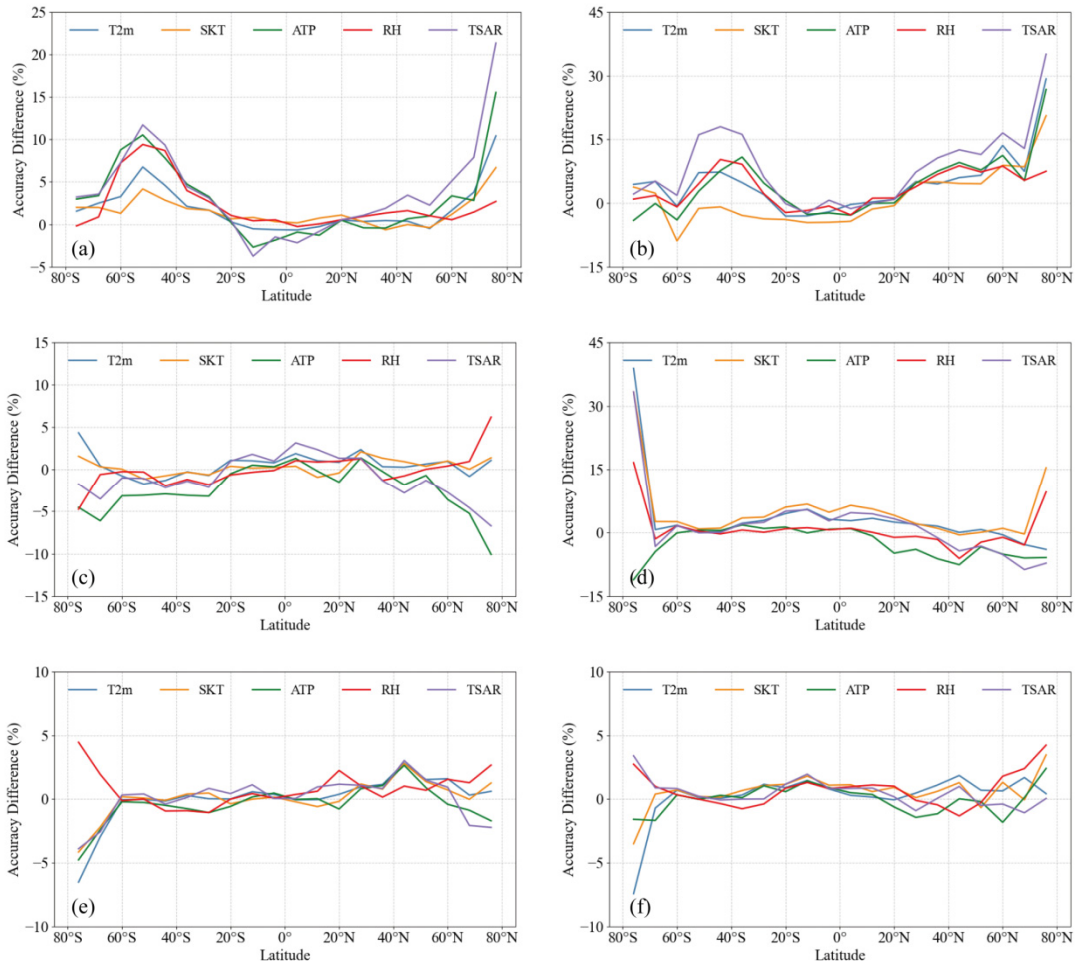


Figure S11. Latitude dependence of accuracy differences in clear sky (a, b), water cloud (c, d) and ice cloud (e, f) detection between T2m, SKT, ATP, RHP, TRAS and BT during daytime (left) and nighttime (right) by MLP.

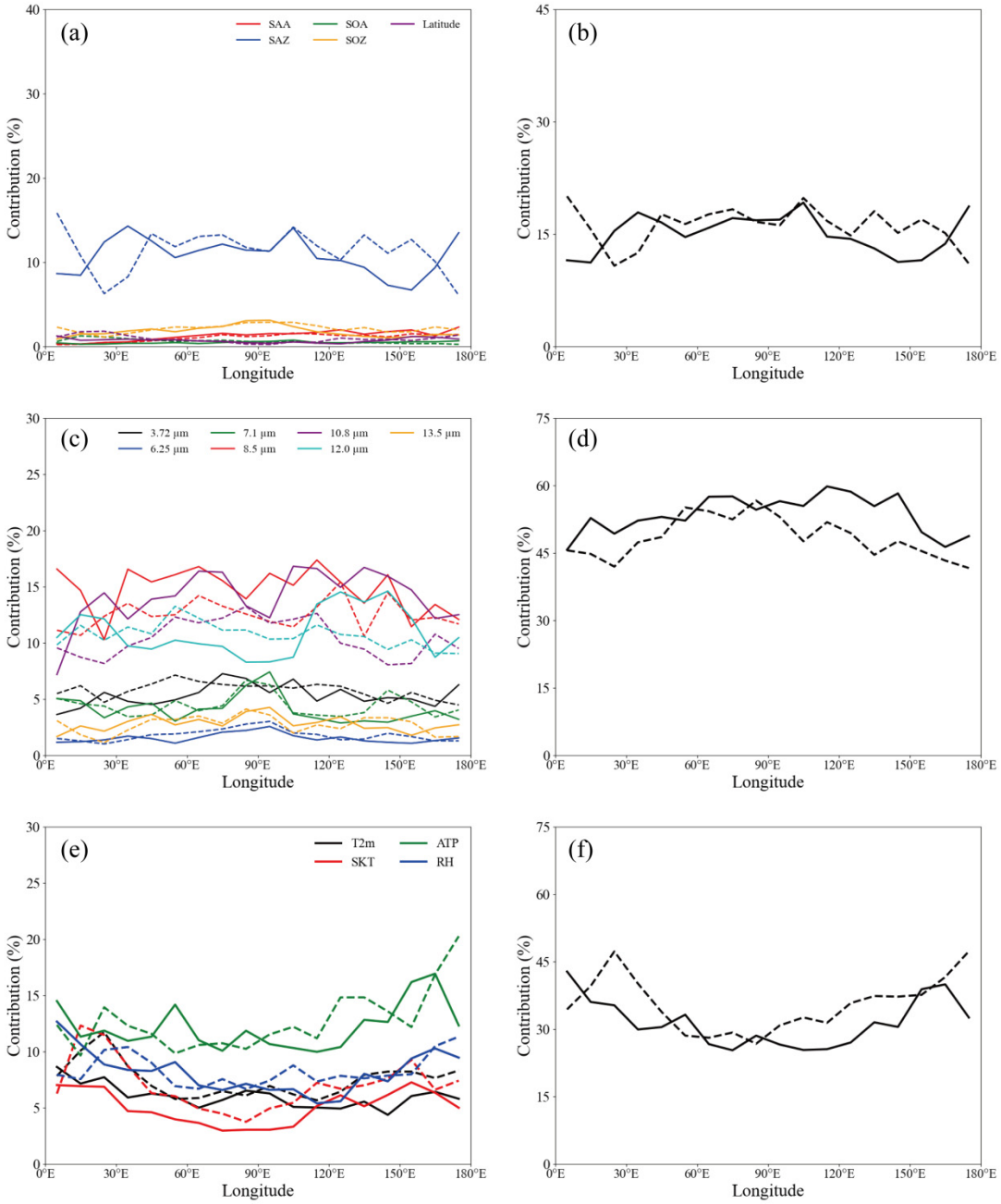


Figure S12. Longitudinal variation of contributions to the MLP-TSAR model from geometry parameters (a) and combined geometry parameters (b), BTs (c) and combined BTs (d) and individual meteorological factors (e) and combined meteorological factors (f). Solid lines and dashed lines represent daytime and nighttime, respectively. Curves represent latitudinal averages. Solid and dashed lines represent daytime and nighttime, respectively (SHAP analysis performed on the model using the 2020 dataset with a sample size of 10,000).

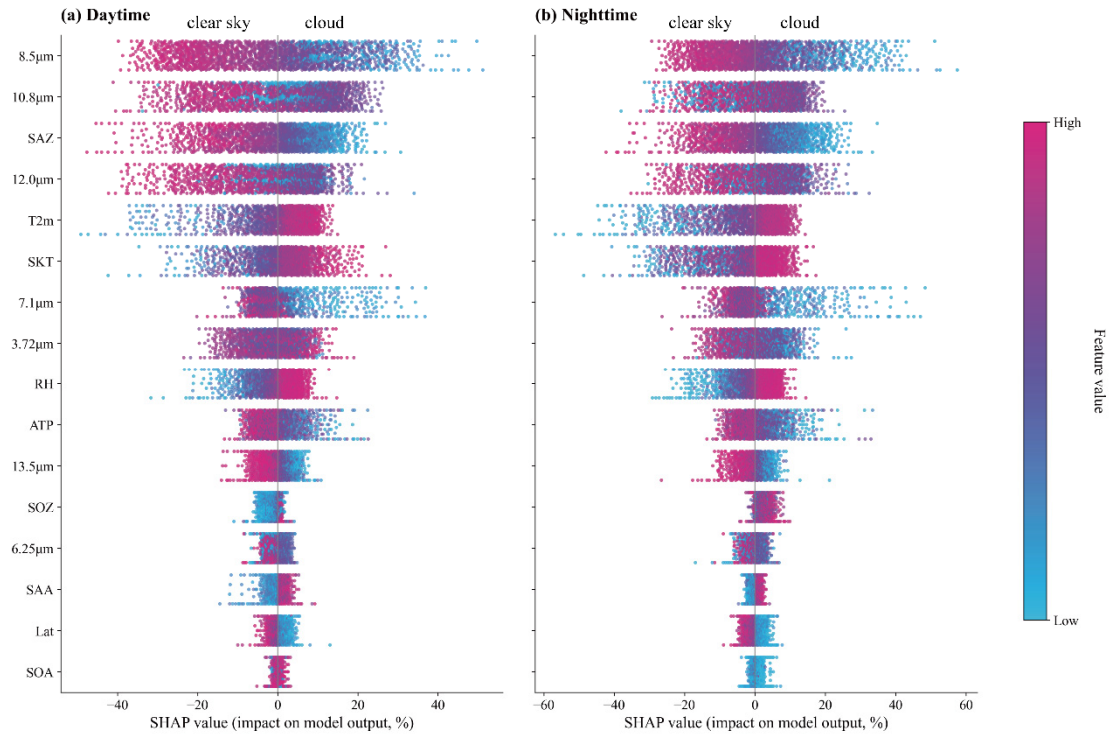


Figure S13. This figure presents the feature importance analysis results based on the SHAP (SHapley Additive exPlanations) method, illustrating the contribution of each input feature to the model predictions under (a) daytime and (b) nighttime conditions, respectively. The vertical axis lists the model input features in descending order of importance, including infrared channel brightness temperatures (8.5μm, 10.8μm, 12.0μm, 7.1μm, 3.72μm, 13.5μm, and 6.25μm), geometric parameters (SAZ: satellite zenith angle; SOZ: solar zenith angle; SAA: satellite azimuth angle; SOA: solar azimuth angle; Lat: latitude) and meteorological factors (T2m: 2-meter air temperature; SKT: skin temperature; ATP: atmospheric temperature profile; RH: relative humidity profile). The horizontal axis represents the SHAP values, indicating the magnitude of each feature's impact on the model output (expressed as a percentage). Positive values indicate that the feature increases the model prediction of cloud, while negative values indicate the feature increases the model prediction of clear-sky. Each point in the figure represents an individual sample, with the color indicating the original value of the corresponding feature: pink denotes high values, while blue denotes low values. The color distribution allows for intuitive interpretation of the relationship between feature values and their directional impact on the model (SHAP analysis performed on the model using the 2020 dataset with a sample size of 10,000).

Table S1. The sample size counts for different categories (clear sky, cloud, water cloud, ice cloud).

Dataset	Clear sky	Cloud	Water cloud	Ice cloud
Training set	654849	1008287	472312	535975
Validation set	163713	252072	117860	134212
Testing set	799752	1293773	607781	685992

Table S2. Hyperparameter configurations for the four machine learning models.

Model	Parameter	Value
RF [43]	Number of estimators	500
	Maximum tree depth	20
	Minimum samples for node split	5
	Minimum samples per leaf node	2
	Maximum features per split	√n (square root)
LightGBM [44]	Number of estimators	500
	Number of leaves	100
	Learning rate	0.05
	Feature fraction	0.9
	Bagging fraction	0.7
	Bagging frequency	3
XGBoost [45]	Number of estimators	500
	Maximum tree depth	8
	Learning rate	0.1
	Subsample ratio	0.8
	Column sample ratio per tree	0.8
MLP [46,47]	Hidden layer architecture	256 → 128 → 64 neurons
	Dropout rates	0.3
	Optimizer	Adam (Learning rate= 0.001)
	Batch size	512
	Maximum epochs	100

Table S3. Statistical significance matrix.

	Accuracy (%)	RF	LightGBM	XGBoost	MLP	FY-4A CLM
RF	91.5	-	L ($p<0.05$)	L ($p<0.05$)	L ($p<0.05$)	W ($p<0.05$)
LightGBM	92.2	W ($p<0.05$)	-	L ($p<0.05$)	L ($p<0.05$)	=
XGBoost	92.5	W ($p<0.05$)	W ($p<0.05$)	-	L ($p<0.05$)	=

MLP	92.8	W (p<0.05)	W (p<0.05)	W (p<0.05)	-	W (p<0.05)
FY-4A CLM	83.1	L (p<0.05)	=	=	L (p<0.05)	-

Matrix description:

'W (p<0.05)': Row model significantly outperforms column model

'L (p<0.05)': Row model is significantly inferior to column model

'=': No significant difference

'-': diagonal