



Cumulo : A Dataset for Learning Cloud Classes

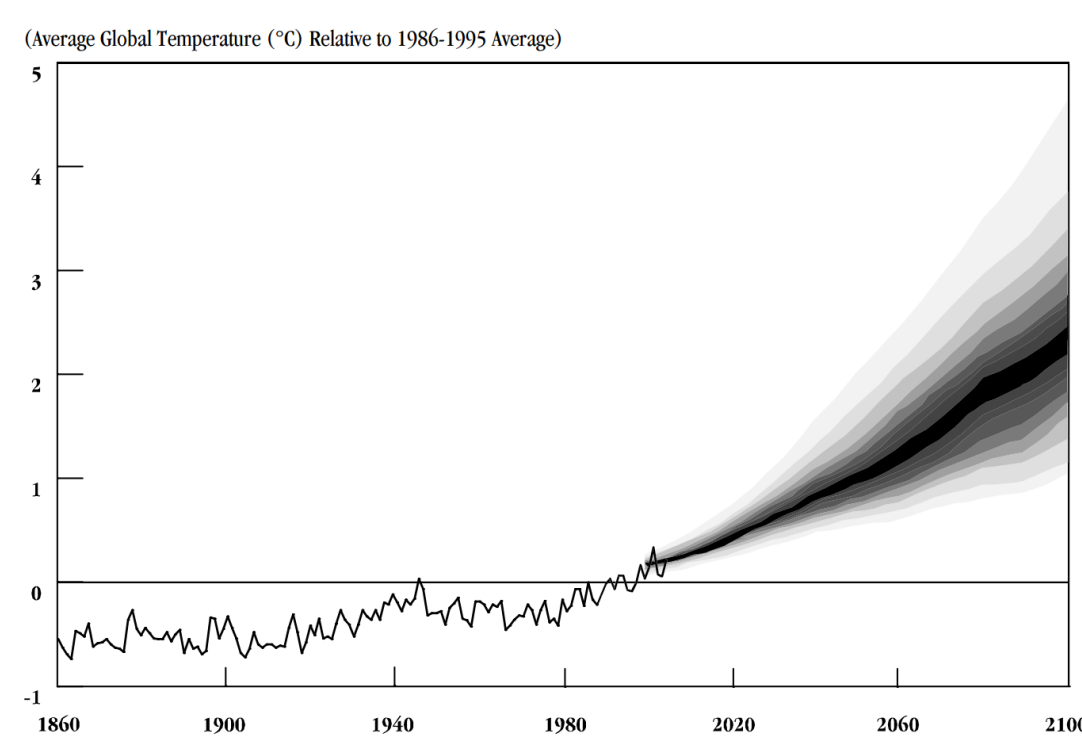
Valentina Zantedeschi Fabrizio Falasca Alyson Douglas
 Richard Strange Matt J. Kusner Duncan Watson-Parris

ESA Frontier Development Lab 2019 - Atmospheric Phenomena and Climate Variability challenge



MOTIVATION

clouds are the greatest source of global warming uncertainty



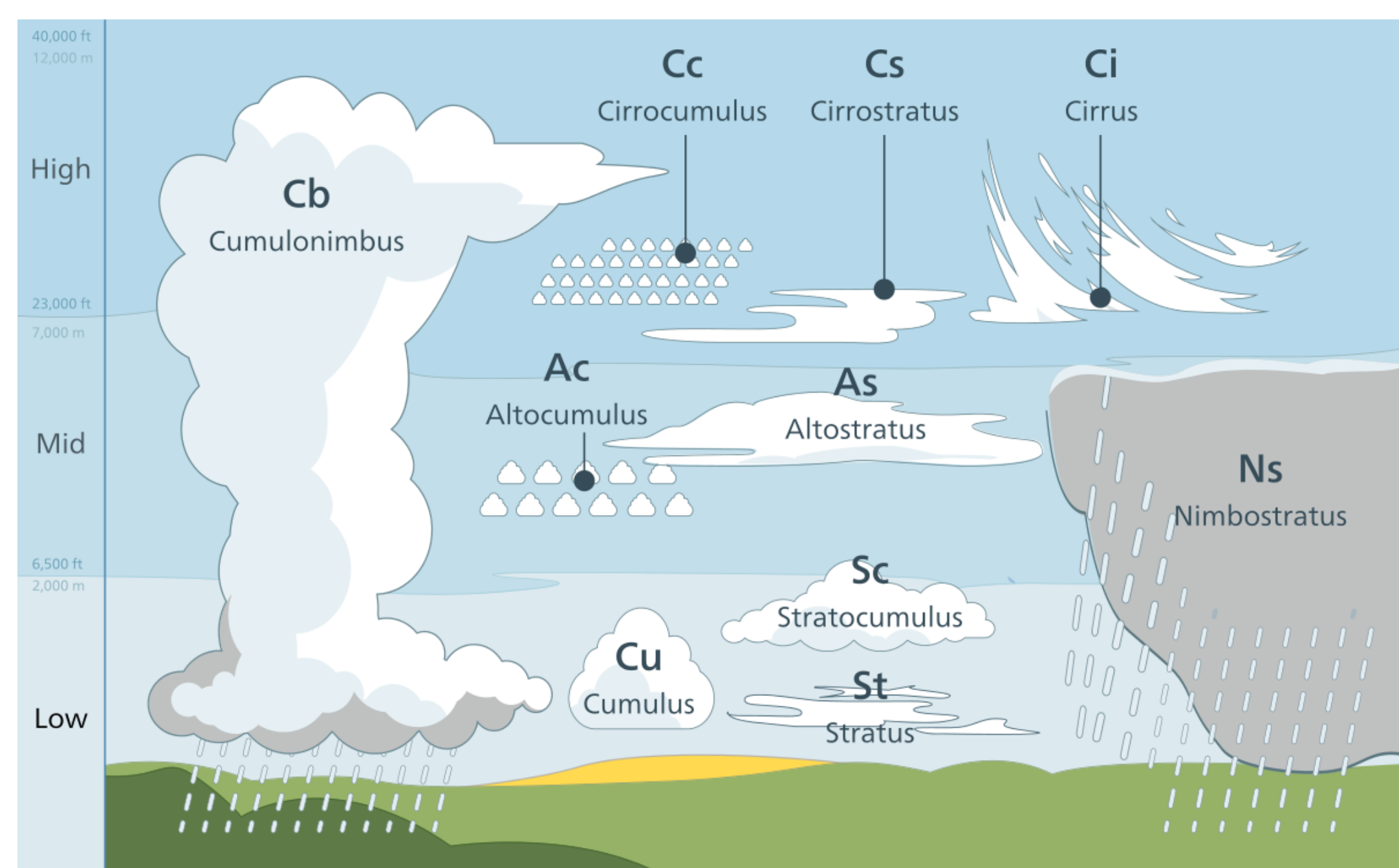
- precipitation
- radiative budget
- feedbacks

To reduce uncertainty

1. classify clouds into types: **Machine Learning**
2. study cloud spatiotemporal variability: **Climate Science**

WMO CLOUD TYPES

cloud types can be distinguished based on their physical properties, their spatial organization and their micro-physical characteristics



coarse classification based on cloud altitude, thickness and precipitation index

Cumulo provides supervision for the following cloud classes:

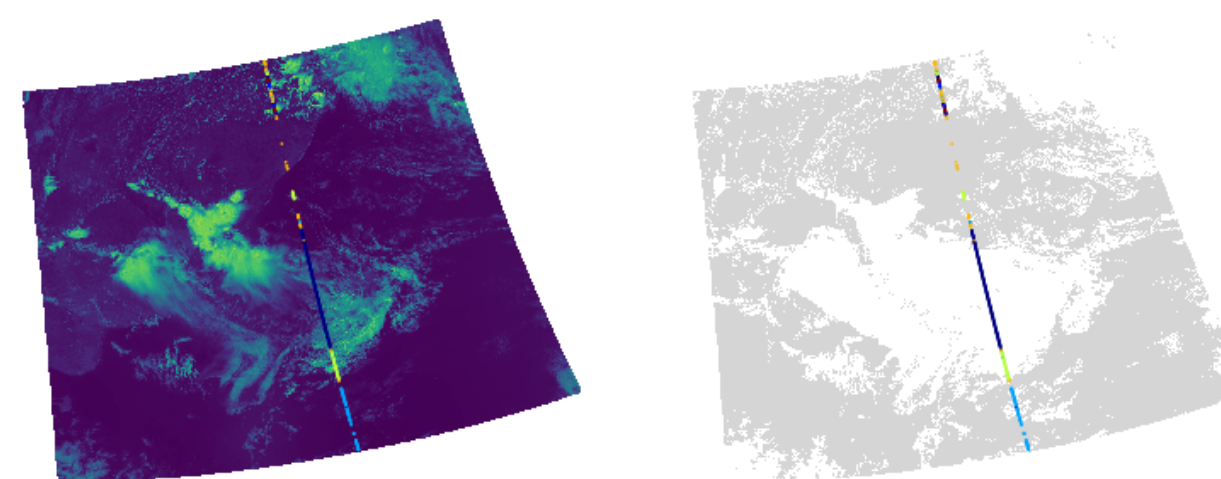
World Meteorological Organization Name	Proportion
Cirrus and Cirrostratus (Ci)	30.68%
Altostratus (As)	16.02%
Altostratus (Ac)	9.53%
Stratus (St)	1.84%
Stratocumulus (Sc)	27.53%
Cumulus (Cu)	6.02%
Nimbostratus (Ns)	7.40%
Deep Convection (Dc)	0.96%

CUMULO DATASET

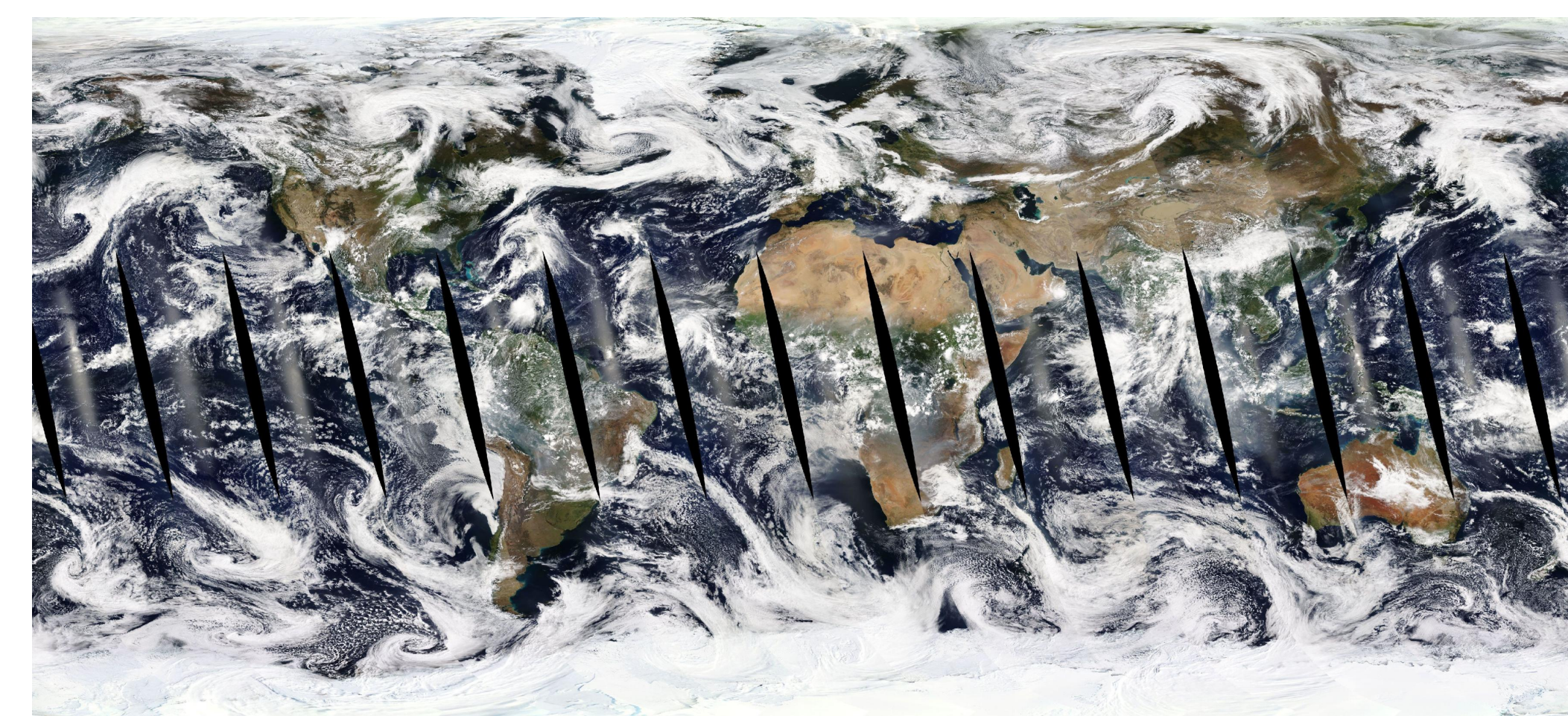
We introduce a new cloud classification dataset to the machine learning community.

Hyperspectral Satellite Images (MODIS Aqua)
 +
Cloud Class Tracks (CloudSat + CALIPSO)

- 105,120 images of 1354×2030 pixels
- near global, daily coverage (1 km^2 resolution)
- 1 km-wide tracks and vertical information
- aligned to image features



Data is available for the whole year 2008 and is stored in NetCDF format.



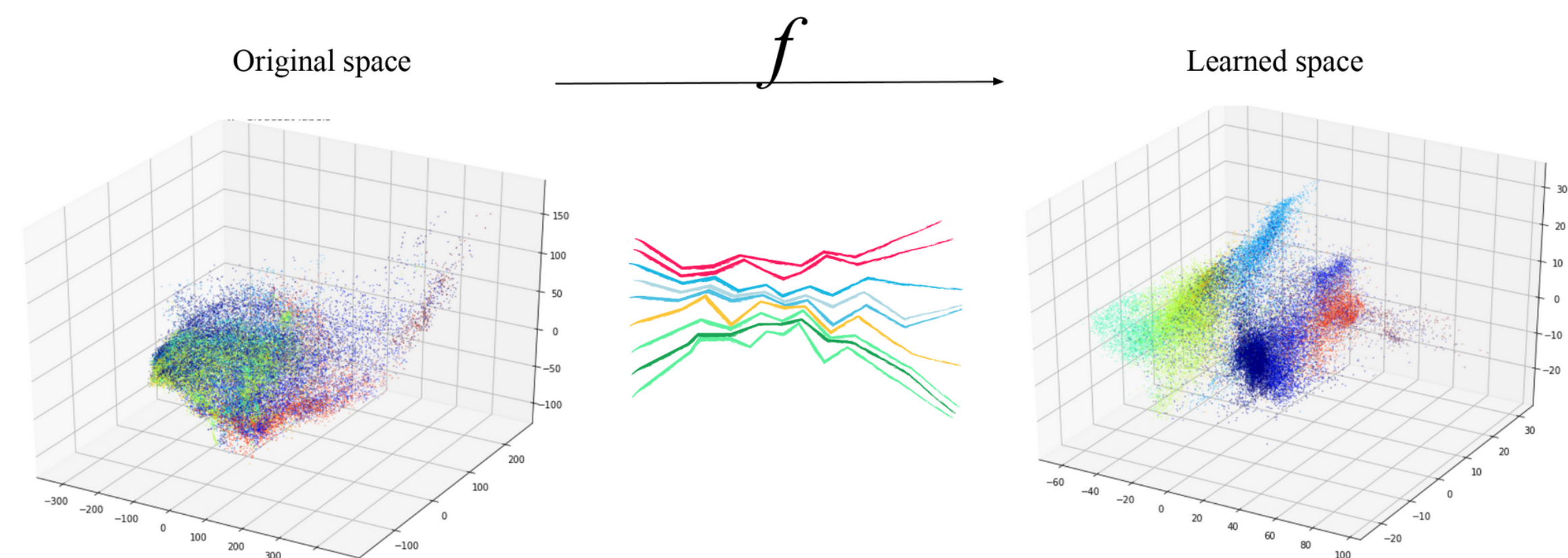
Source	Name/Description	Index	Primary Use
MODIS	shortwave visible (red)	1	land/shadow/cloud/aerosols boundaries
	shortwave near infrared	2	land/shadow/cloud/aerosols boundaries
	longwave thermal-infrared	20-23	surface/cloud temperature
	shortwave near infrared	26	Cirrus clouds water vapor
	longwave thermal-infrared	27	water vapor
	longwave thermal-infrared	29	cloud properties
MODIS Cloud Mask	cloud mask	33-36	cloud detection
	cloud layer type		target classes
MODIS Cloud Product	liquid water path		physical validation
	cloud optical thickness		physical validation
	cloud effective radius		physical validation
	cloud particle phase		physical validation
	cloud top pressure		physical validation
	cloud top height		physical validation
	cloud top temperature		physical validation
	cloud effective emissivity		physical validation
	surface temperature		physical validation

Challenges

1. Supervision is available only 1 every 1354 pixels => **weakly-labelled data**
2. Pixels can be annotated with multiple types of clouds => **multi-labelled data**
3. Many cloud classes are underrepresented => **class imbalance**
4. Some channels are available only at daytime => **missing data**
5. Many cloud types have sub-types => **unsupervised learning**

BASELINE PERFORMANCE ANALYSIS

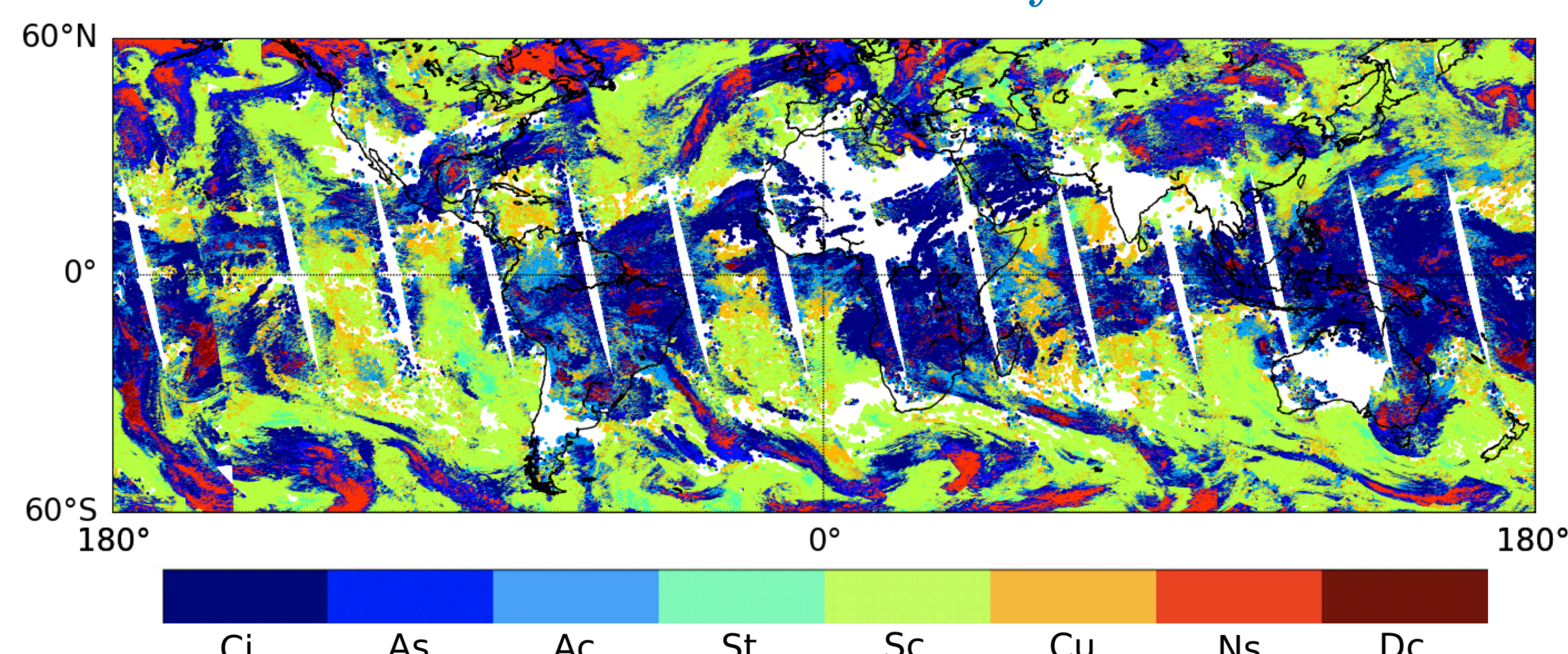
Semi-supervised cloud classification using a **Hybrid Invertible Residual Network** [1] on 3x3 daytime tiles sampled from January 2008.



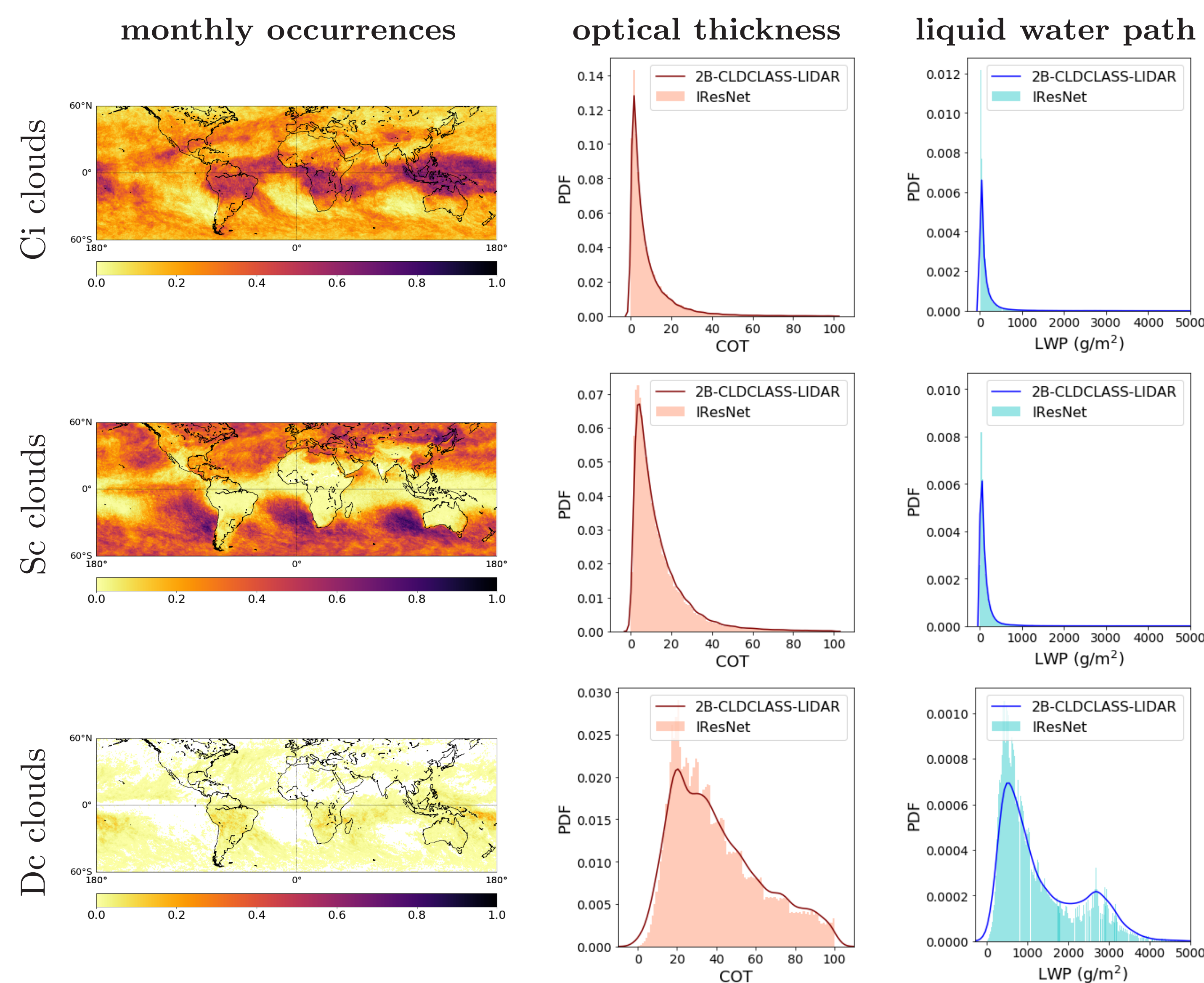
Standard Evaluation Metrics

	Ci	As	Ac	St	Sc	Cu	Ns	Dc	Mean
Accuracy (%)	81.30	84.50	88.29	97.73	88.90	92.40	90.92	98.84	90.36
F1 score	0.68	0.43	0.45	0.58	0.80	0.40	0.47	0.58	0.55
IoU index	0.52	0.28	0.29	0.41	0.66	0.25	0.31	0.41	0.39

Classification of January 18



Climate-based Evaluation



REFERENCES

[1] Nalnick Eric, Matsukawa Akihiro, Teh Yee Whye, Gorur Dilan and Lakshminarayanan Balaji "Hybrid models with deep and invertible features" ICML 2019