DUST DETECTION IN SATELLITE DATA USING CONVOLUTIONAL NEURAL NETWORKS (CNN)

CyberTraining: Big Data + High-Performance Computing + Atmospheric Sciences

Changjie Cai[1], Jangho Lee[2], Yingxi Rona Shi[3], Camille Zerfas[4]

[1]Dept. of Occupational and Environmental Health, University of Oklahoma

[2]Dept. of Atmospheric Sciences, Texas A&M University

[3]USRA and NASA GSFC Climate and Radiation Laboratory

[4]School of Mathematics and Statistics, Clemson University

Research assistant: Pei Guo[5], Faculty mentor: Zhibo Zhang[6]

[5]Department of Information Systems, University of Maryland Baltimore County

[6]Department of Physics, University of Maryland Baltimore County

WHY DO WE WANT TO TRACK DUST IN THE ATMOSPHERE?

- Dust poses large impacts on Earth environments and human society.
- It influence Earth radiation budget by absorbing and reflecting solar energy.
- It alters the hydrology cycle by interacting with cloud and precipitation formations.
- Dust storm lowers the air quality & visibility and causes large economic loses.
- Air-borne dust particles has been linked to respiratory illnesses, such as asthma, meningitis, and others.
- It can travel for a long distance such as across oceans and continents.
- Dust is also an important mineral source that fertilize the Earth delicate ecosystem.
- Global-warming might impacts dust source region, which are mostly earth's arid and semi-arid lands.

PREVIOUS STUDIES

- Dust detection algorithms have been studied for many years.
- They have used several different types of satellite data: TOMS, OMI, METEOSAT, AVHRR, GOES-VISSR, and MODIS.
- Earlier studies relied on physical-based algorithms, but these algorithms have several limitations: the optical properties of dust might be hard to detect, and the algorithm may miss dust due to complicated observing conditions, or extremely thin or thick dust loading.
- Because of these shortcomings, recent attention has shifted to statistical methods.

SATELLITE DATA USED

CALIOP

- CALIOP can resolve aerosol vertical distribution, and its polarization capability allows it to identify particles shapes. This makes the sensor ultra sensitive to dust particles.
- CALIOP cloud and aerosol layer product is used in this study. The layer product has a horizontal resolution of 1/3km, 1km, and 5 km. Here we uses 5 km product. The CALIOP identified layer feature is used as benchmark for detecting dust in this study.

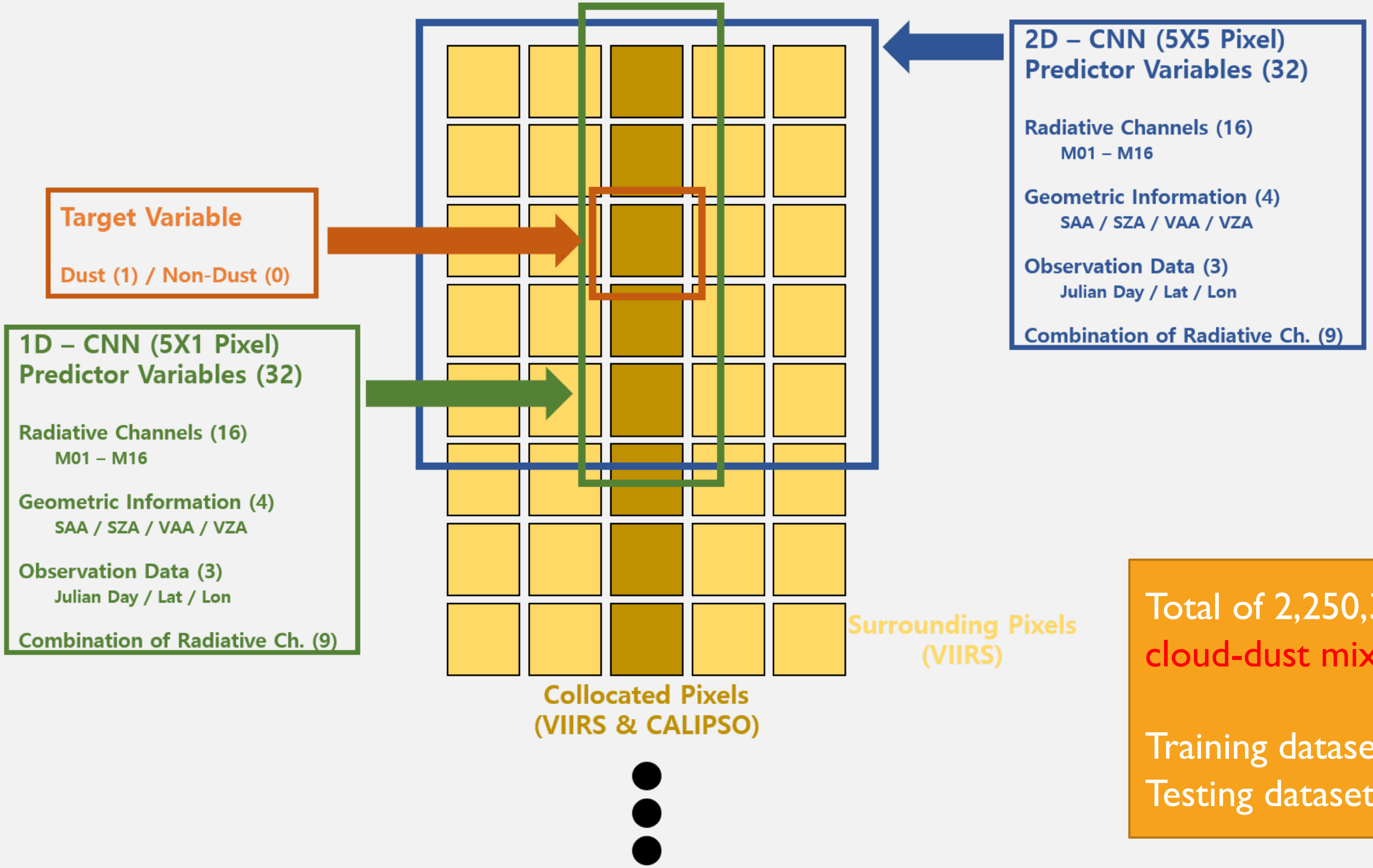
VIIRS

- The VIIRS sensor has 16 M bands with 750 meter native resolutions from 412 nm to 12 micron, and 5 I bands with 375 meter resolution.
- The sensor is designed to retrieve aerosols using split window methos over various spectrum data.
- Compared with CALIOP, VIIRS has big advantage of spatial coverage.

DATA DESCRIPTION (1/3)

- Next we show a diagram of the dataset we used in the 1D and 2D CNN models.
- The collocated CALIPSO and VIIRS satellite data was used. The model data was extracted form each observation using a 1*5 moving window.
- Additional 4 (2 on left, 2 on right) tracks of VIIRS data was added. The model data was extracted form each observation using a 5*5 moving window.

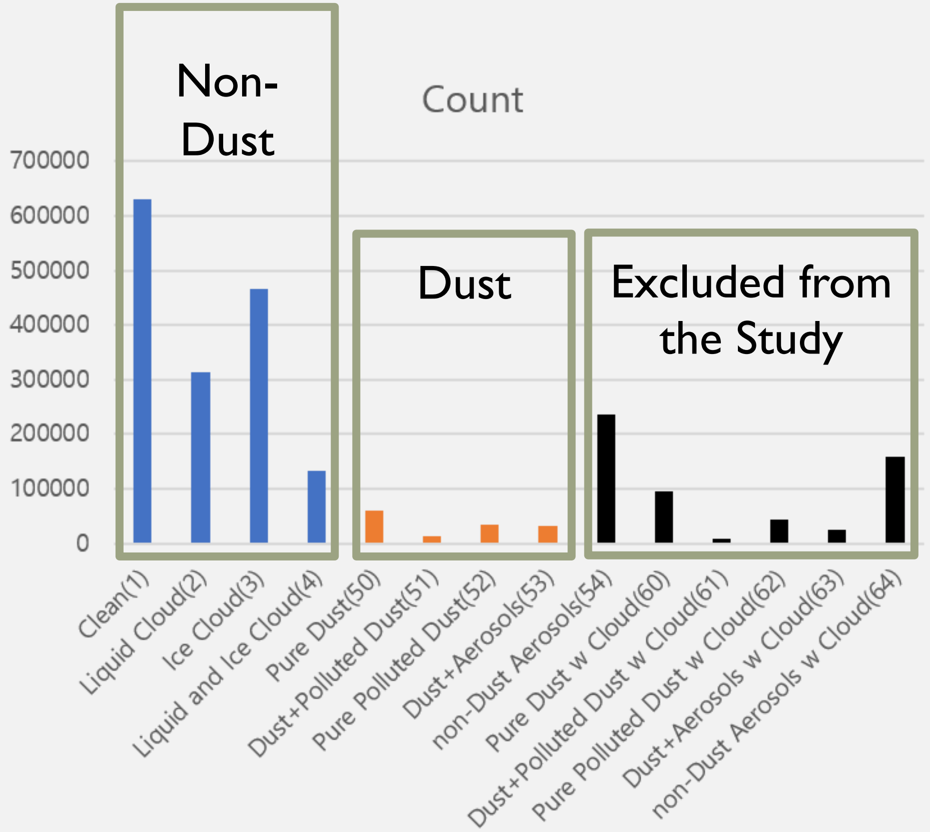
DATA DESCRIPTION (2/3)

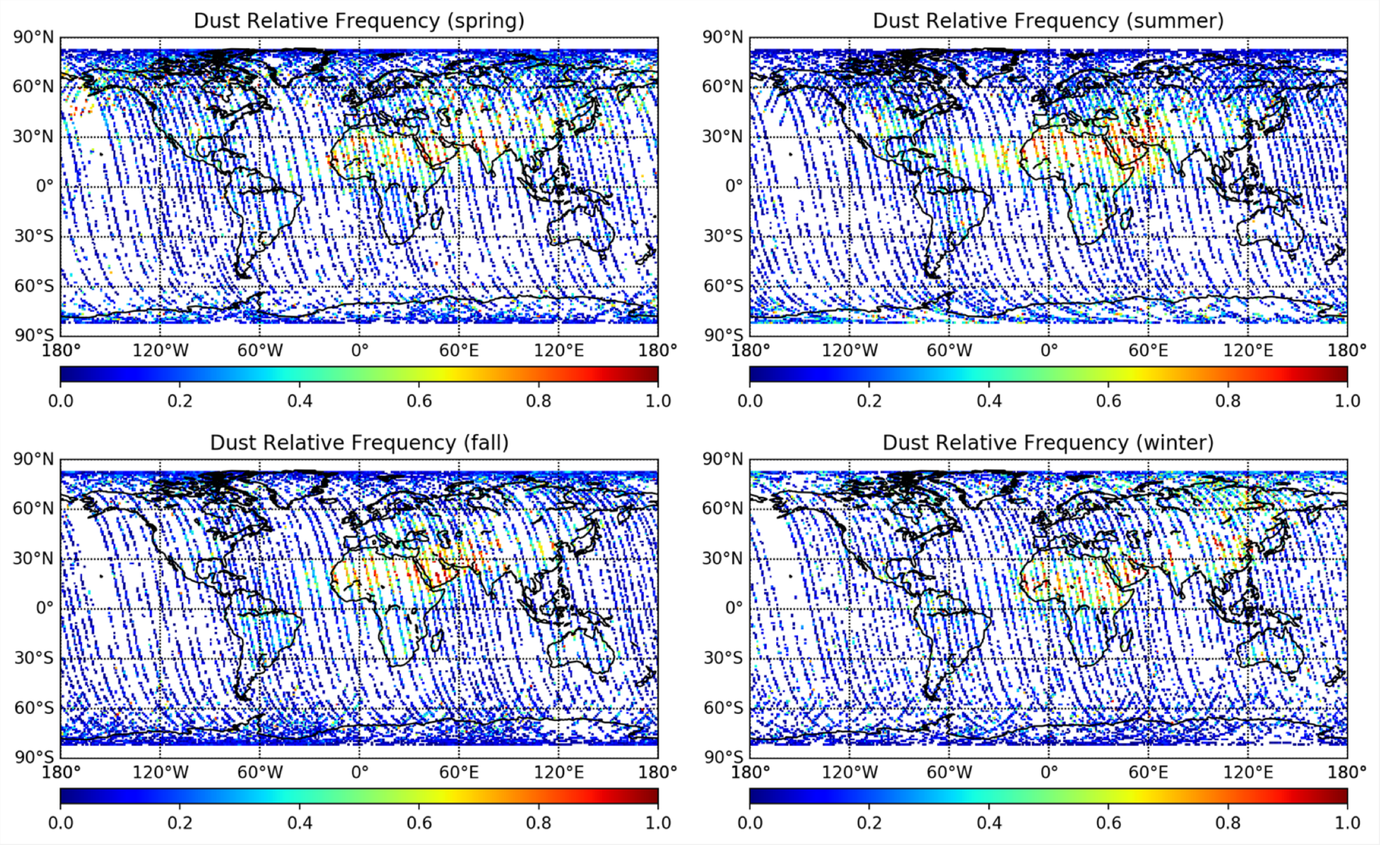


Total of 2,250,339 dataset with removing cloud-dust mixture

Training dataset: 1,500,026 samples
Testing dataset: 750,013 samples

DATA DESCRIPTION (3/3)

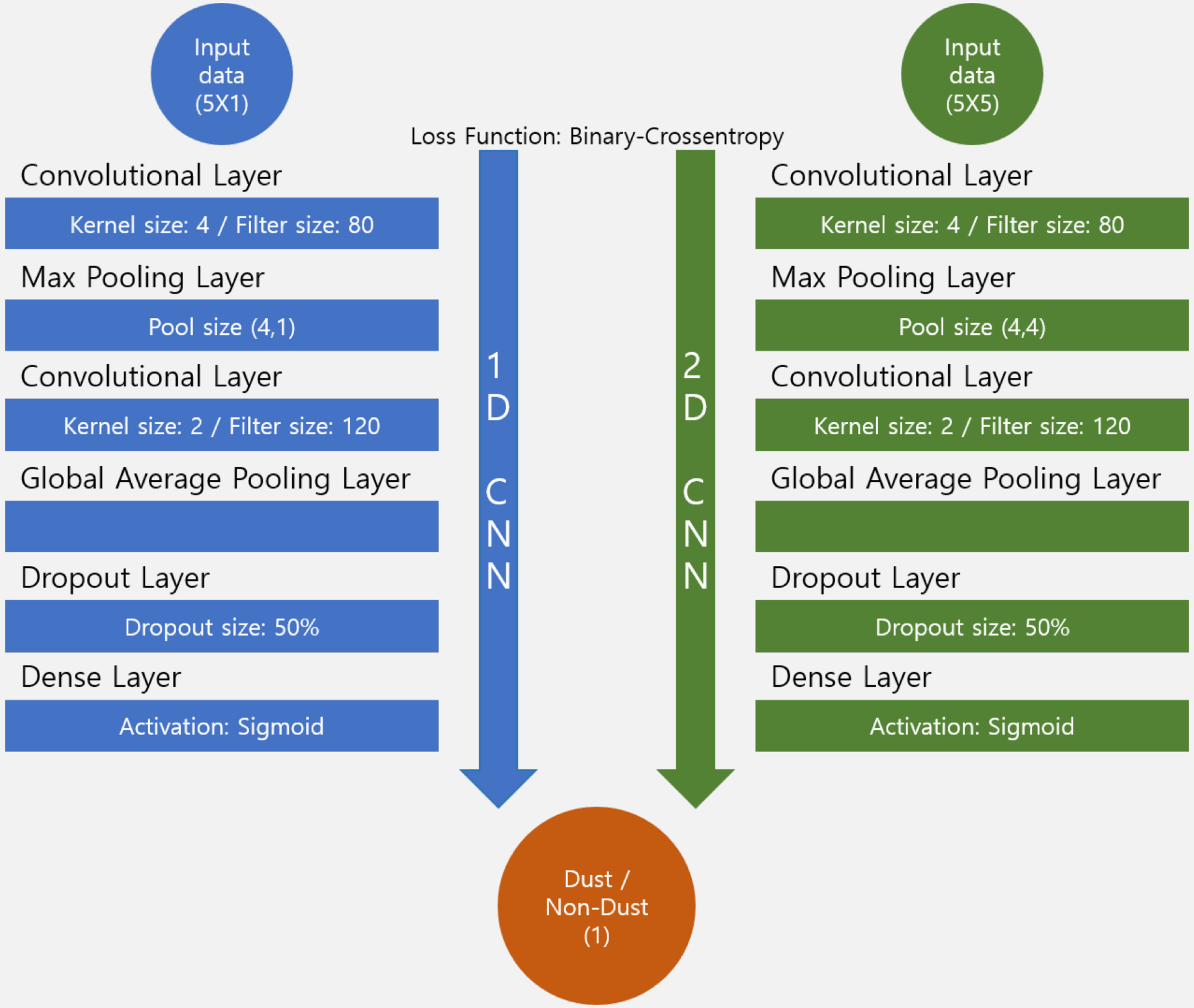




CNN MODEL DESCRIPTION (1/2)

- Loss function: binary cross-entropy. Activation function (for the dense layer): Sigmoid.
- The Adam optimizer is used to control the learning rate.
- Eight models were created: 1D16V, 1D20V, 1D23V, 1D32V, 2D16V, 2D20V, 2D23V, 2D32V
- 16V: Only considers radiative channel as its input variable.
- 20V: Considers radiative channel, and geometric data.
- 23V: Considers radiative channel, geometric, and observation data.
- 32V: Considers radiative channel, geometric information, observation, and combinations of radiative channels.
- All models were trained for 100 epochs with batch size of 2048.

CNN MODEL DESCRIPTION (2/2)



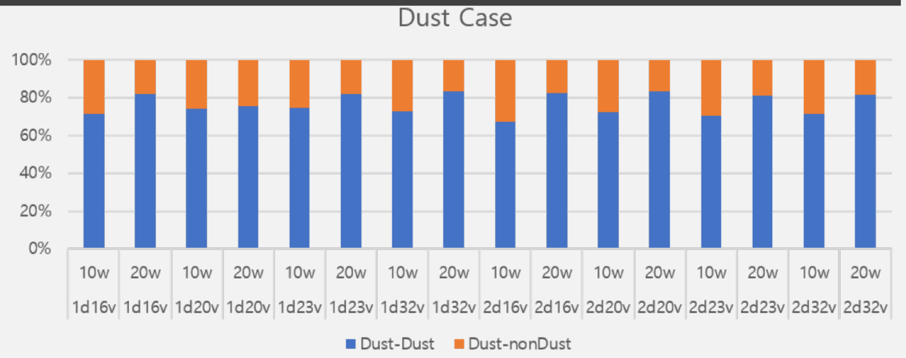
MODEL EVALUATION
1. INPUT DIMENSION SENSITIVITY (1/2)

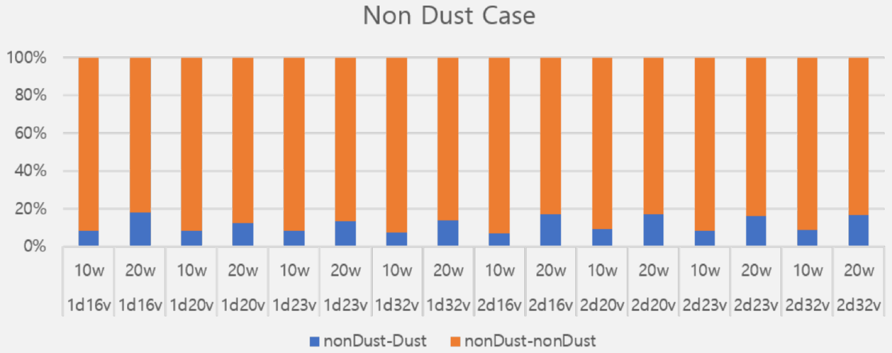
- To test the input dimension sensitivity of the model, 8 different models were tested (1D16V, 1D20V, 1D23V, 1D32V, 2D16V, 2D20V, 2D23V, 2D32V) in 2 different weights (10w, 20w)

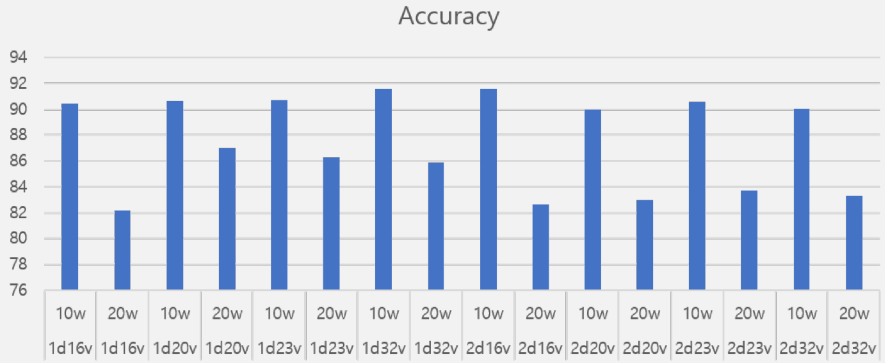
	CALIOP Dust	CALIOP Non-Dust
CNN DUST	Dust-Dust	nonDust-Dust
CNN NON-DUST	Dust-nonDust	nonDust-nonDust

MODEL EVALUATION
I. INPUT DIMENSION SENSITIVITY (2/2)

- The accuracy of the model is more dependant on weight selection than dimension selection, and did not differ a lot from each other.





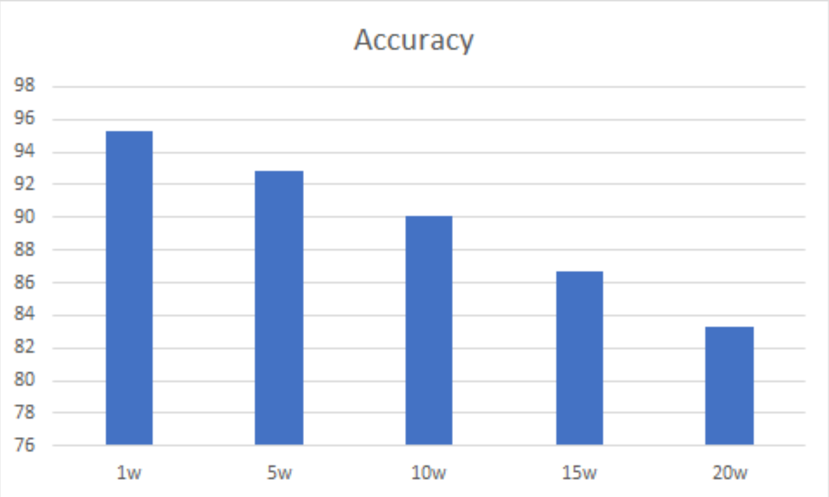


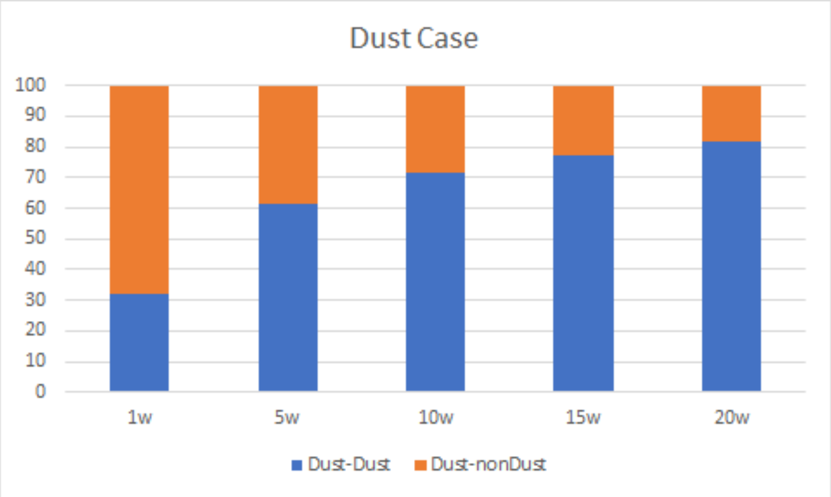
MODEL EVALUATION
2. WEIGHT SENSITIVITY (1/2)

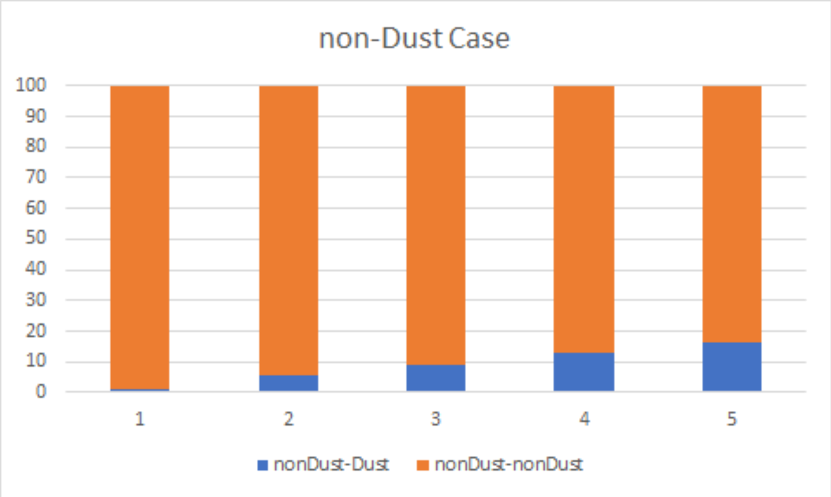
- Since there is a bias of numbers in data samples between dust:non-dust (1:11), appropriate weight selection is crucial on setting up the model.
- For example, when the model weight is set to 10:1, the model will count predicting dust correctly 10 times more valuable then predicting the non-dust correctly.
- Model weight sensitivity test is done with 2D32V model, which is the most complex model.

MODEL EVALUATION
2. WEIGHT SENSITIVITY (2/2)

- Even the accuracy is the lowest in 20w model, the accuracy of predicting dust as dust is the highest in 20w model, with 83%
- Compared with 1w, which have 32% accuracy of predicting dust and 98% of predicting non-dust, the higher weighted models tend to predict dust far better then lower weighted models (51%), while prediction performance of non-dust case decrease slightly (15%)

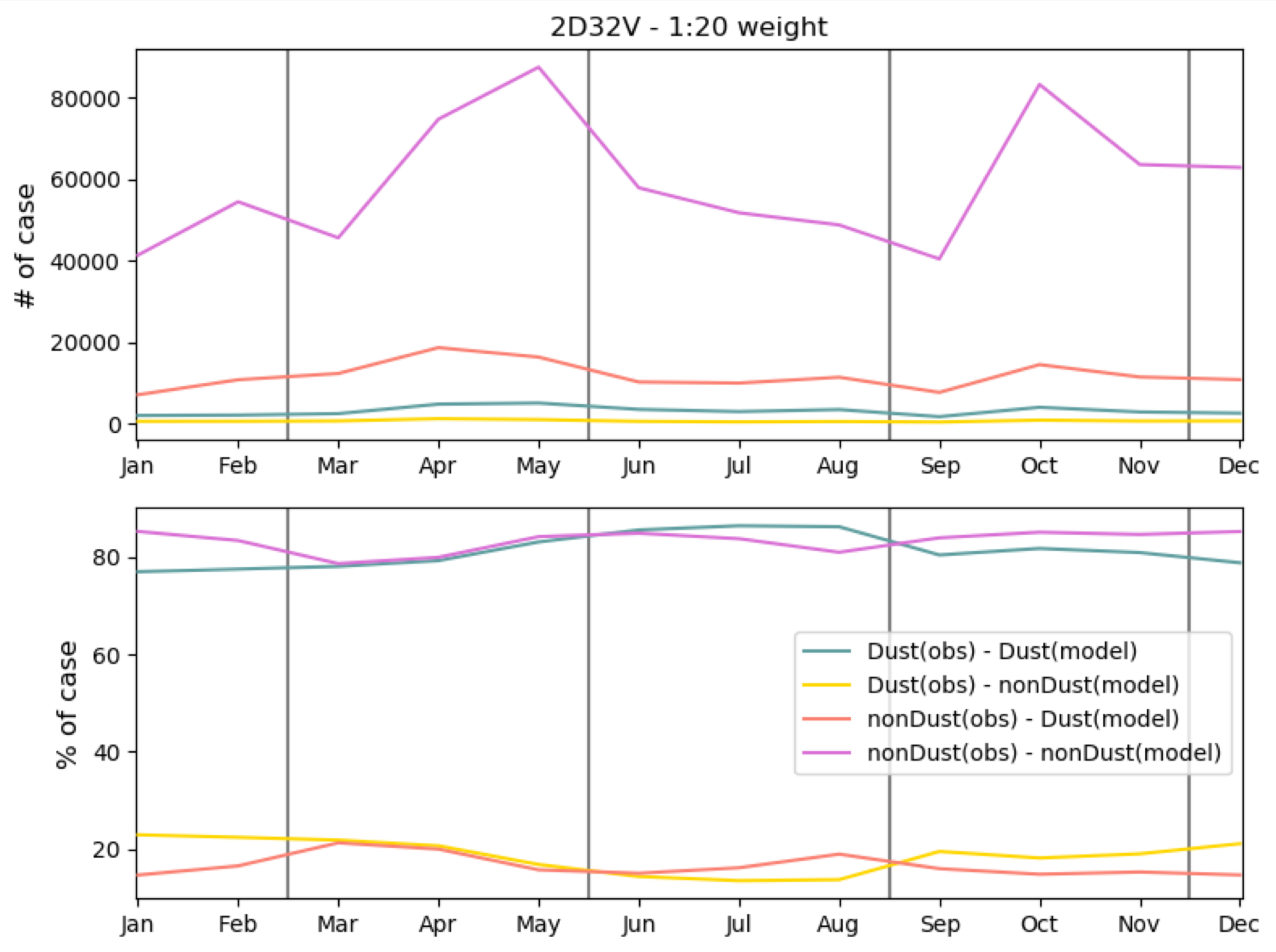






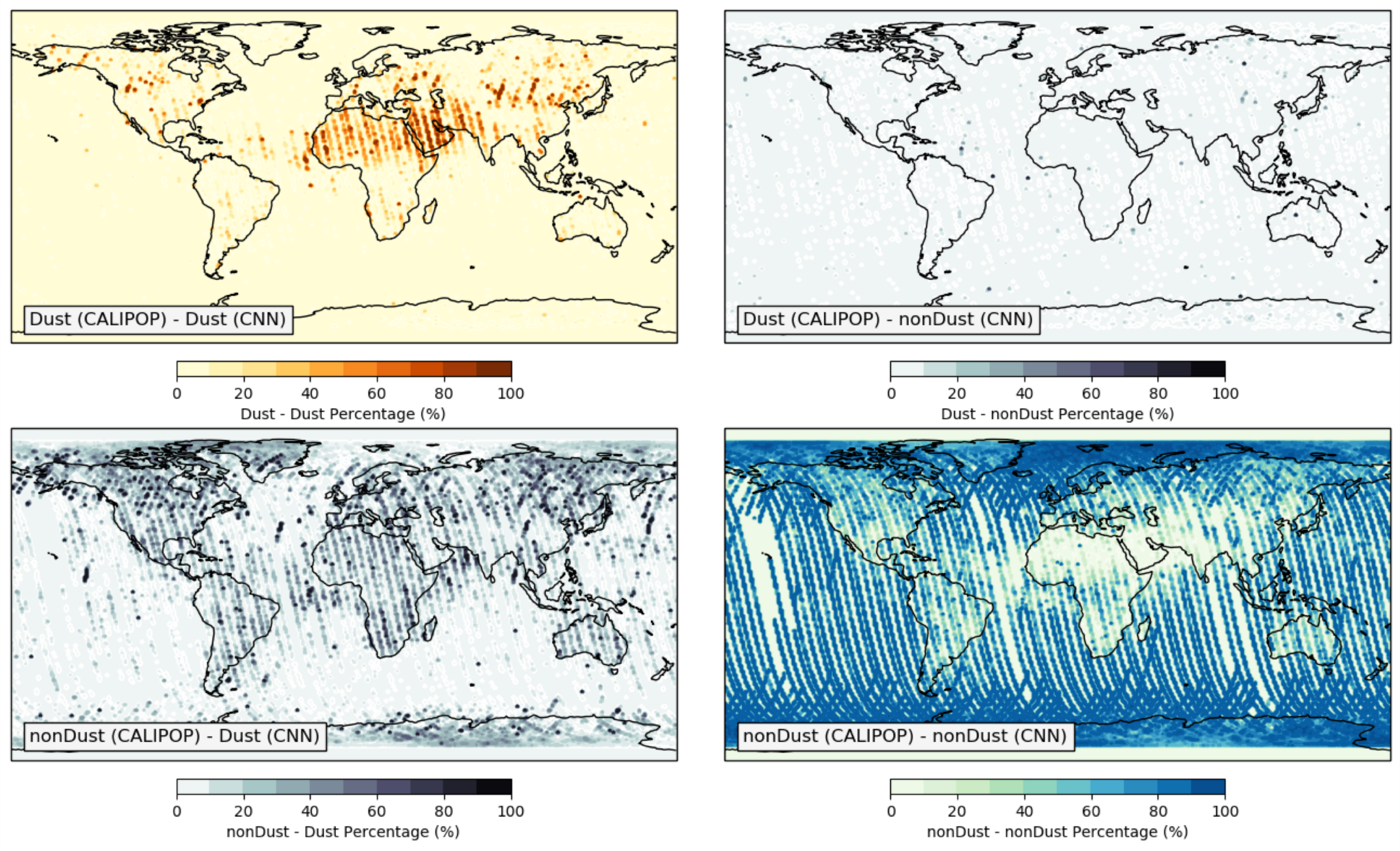
MODEL EVALUATION
3. TEMPORAL SENSITIVITY

- 2D32V20w model was selected for temporal sensitivity test.
- The accuracy is fairly constant throughout the season.



MODEL EVALUATION
4. SPATIAL SENSITIVITY

- Spatial dependency on performance of the model was also tested.
- Accuracy of dust-dust is high over Northern Africa (Sahara) and Central China (Gobi).
- Otherwise, the accuracy is low since the number of observed dust is close to zero.
- Error of predicting dust as non-dust is fairly low throughout the globe. However, the error of predicting non-dust as dust is relatively higher.
- The error is high over the land area, especially northern Canada and polar region. Accuracy of predicting non-dust as non-dust is very high throughout the globe, except for regions with lot of dust occurrence (Sahara, Gobi).



MODEL EVALUATION
5. PERFORMANCE COMPARED TO EXISTING MODELS (1/2)

- Performance of the model has been compared to the existing model of VIIRS aerosol detection product (ADP), which is developed by NOAA VIIRS team and is considered as state of art dust/smoke detection product that is based on physics. The product was evaluated using collocated ADP and CALIOP data. The metric used accuracy (Acc), Probability of Correct Detection (POCD), and Probability of false detection (POFD) which is defined as follows:
  - 1) ACC = (TP+TN)/(TP+TN+FP+FN)
  - 2) POCD = (TP)/(TP+FP)
  - 3) POFD = (FN)/(TP+FN)
    - where TP, FN, FP, FN each represents true positive, true negative, false positive, false negative

MODEL EVALUATION
5. PERFORMANCE COMPARED TO EXISTING MODELS (1/2)

- 1) ACC = (TP+TN)/(TP+TN+FP+FN)
- 2) POCD = (TP)/(TP+FP)
- 3) POFD = (FN)/(TP+FN)
  - where TP, FN, FP, FN each represents true positive, true negative, false positive, false negative

	S-NPP	CNN-1d32v10w
ACC	92.3	91.6
POCD	92.5	72.6
POFD	19.8	27.3

THANK YOU!

This study is supported by NSF award (#1730250) for CyberTraining: DSE: Cross-Training of Researchers in Computing, Applied Mathematics and Atmospheric Sciences using Advanced Cyberinfrastructure Resources.

We thank the UMBC High Performance Computing Facility (HPCF) for their hardware and maintainance.