Project # 20-026

Improve Cloud Modeled by WRF using COSP and Generative Adversarial Network

Prepared for

Air Quality Research Program (AQRP)
The University of Texas at Austin

Elena McDonald-Buller
Project Manager
Texas Air Quality Research Program

Bright Dornblaser
TCEQ Project Liaison

By

Principal Investigator, Zheng Lu
Graduate Student, Kai Lyu
Department of Atmospheric Sciences, Texas A&M University

QA Requirements: Audits of Data Quality: 10% Required
Report of Findings – Required in Final Report

September 2021

TABLE OF CONTENTS

List of Figures

List of Tables

List of Acronyms

ACM	Asymmetric Convective Model (Scheme)
CF	Cloud Fraction
CFMIP	Cloud Feedback Model Intercomparison Project
COD	Cloud Optical Depth
COSP	CFMIP Observation Simulator Package
DL	Deep Learning
ECWMF	European Centre for Medium-Range Weather Forecasts
FNL	(NCEP) Final
GAN	Generative Adversarial Network
GBM	Grenier-Bretherton-McCaa (Scheme)
LWP	Liquid Water Path
ML	Machine Learning
MODIS	Moderate Resolution Imaging Spectroradiometer
NAM	North American Mesoscale Analysis
NASA	The National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NOAA	National Oceanic and Atmospheric Administration
PBL	Planetary Boundary Layer
WRF	Weather Research and Forecasting
YSU	Yonsei University (Scheme)

Acknowledgement

The preparation of this report (Project No. 20-026) was funded by a grant from the Texas Air Quality Research Program (AQRP) at The University of Texas at Austin through the Texas Commission on Environment Quality (TCEQ). The findings, opinions and conclusions are the work of the authors and do not necessarily represent findings, opinions, or conclusions of the AQRP or the TCEQ.

Executive Summary

Cloud fields play an important role in regulating the formation, transportation, and lifetime of gas and particulate pollutants through processes like aqueous phase chemistry and wet-removal. It is desirable to accurately simulate cloud fields at mesoscale scale using regional numerical models for the air quality prediction application. However, the numerical models, like Weather Research and Forecasting (WRF) model, commonly experience a well-known issue of "too few and too bright cloud" and produce biased cloud fields. To evaluate modeled cloud fields, "apple-to-apple" comparison against satellite observation (cloud products from NASA MODIS in our study) is particularly important; therefore, the modeled cloud fields are further processed by the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP) in our study. Recently, several studies adopted novel deep learning (DL) techniques, including feature detection, short-term prediction, etc. The goal of our study is to improve the WRF modeled cloud fields using a DL tool called Generative Adversarial Network (GAN). The two objectives of this study are:

(1) To conduct a series of WRF simulations as well COSP analysis to find an optimal combination of physics suite and reanalysis input for modeling clouds fields over Texas.

(2) To train a GAN model over the time series of modeled cloud fields so that the macro- and microphysical properties of modeled clouds are more accurate compared to the observations.

Methodology

We conduct WRF simulation with different combination of the microphysical schemes (the Morrison scheme, the Thompson scheme, and the WSM6 scheme), the PBL schemes (the YSU scheme, the GBM scheme, and the ACM2 scheme), and reanalysis inputs (the NAM reanalysis, the FNL reanalysis, and the ECWMF reanalysis datasets). Total 27 WRF cases are conducted for year 2018, and further processed by the COSP calculation. By comparing against NASA MODIS cloud products and ranking the performances of each cases, we select the optimal configuration and further conducts long-term simulations (2005-2020). The daily modeled cloud fields and daily MODIS cloud products are both composed as false-color images with cloud fraction (CF) as the blue channel, liquid water path (LWP) as the red channel, and cloud optical depth (COD) as the green channel. We use the data from years 2005-2019 for training GAN and the data from year 2020 for evaluating GAN. We finally examine the performance of GAN by analyzing GAN-generated cloud fields in the evaluation dataset day-by-day and calculating the statistics of GAN-generated cloud fields.

WRF Simulations

1). We find that, among 27 cases, the choice of the microphysics scheme accounts for the largest variation in cloud field for almost all months of the year, except during June and July, when the choice of PBL scheme is more important.

2). WRF model significantly underestimates CF compared to observation, especially during summer. The cases with the Morrison microphysics scheme perform relatively better in summer compared to other cases.

3) The optimal case that we select is the case with the YSU PBL scheme, the Morrison microphysics scheme, and the NAM reanalysis. This case ranks the first place in terms of simulating CF and domain-averaged LWP based on NMB (normalized mean bias) metric and ranks in high places based on the spatial correlation coefficient metric.

GAN Training

1) GAN can change the texture of modeled cloud fields by adding fine scale features, including gaps/breakings in cloud decks and feathery-like cloud cells so that the modeled cloud fields look more like observed ones.

2) GAN can improve the modeled cloud fields associated with the frontal system by adjusting the location and the width of the frontal system and reducing modeled COD.

3) GAN can substantially improve the CF simulation, decrease NMB to almost zero and increase spatial correlation coefficient by 0.2 throughout the entire 2020.

4) GAN sometimes fails to simulate localized deep convection systems during the summer. GAN also fails to reproduce cloud fields associated with hurricanes because of very few of such examples in long-term simulations.

Recommendations for future work

1) We plan to increase sample number of cloud fields by introducing geo-stationary satellite observations, like NOAA GOES-East. For multiple years simulation, we can only obtain 4728 training samples, which is considered as a very small sample size in DL studies. By comparing against geo-stationary satellite observation, we can generate training datasets multiple times in a day. We believe this could substantially benefits the performance of GAN.

2) We plan to conduct GAN training again but replacing COD fields with IWP (ice water path) fields. Based on our analysis, we find that modeled COD is very biased compared to the

MODIS observation, probably because of inaccurate assumption of cloud droplet number concentration/effective size in COSP.

3) We would like to ask how we can use GAN-generated cloud fields in aid of WRF model simulation. Here we propose a novel method for such purpose: we can firstly run WRF model and get first-guess of cloud fields, which is further fed into GAN and get pseudo-observed cloud fields. Such cloud fields can be assimilated into the second round of WRF simulation.

1. Introduction

This document provides the final report for the Texas Air Quality Research Program (AQRP) project 20-026, "Improve cloud modeled by WRF using COSP and generative adversarial network". The project principal investigator (PI) is Dr. Zheng Lu (Texas A&M University). The AQRP project manager is Dr. Elena McDonald-Buller at the University of Texas, Austin. The project liaison for the Texas Commission on Environmental Quality (TCEQ) is Dr. Bright Dornblaser. This work targets the research area as listed in Texas Air Quality Research Program (AQRP) guideline: "Meteorological Inputs for Modeling" with the goal "to support scientific research related to Texas air quality, in areas of …meteorology and air quality modeling."

1.1 Background

Cloud fields play an important role in regulating the formation, transportation, and lifetime of gas and particulate pollutants [e.g. Liang and Jacob, 1997; Gurciullo and Pandis, 1997; Fan et al., 2004]. The radiation field altered by cloud controls the photolysis reaction associated with ozone formation [Faust, 1994]. Photolysis rate is dependent on many factors that can influence solar actinic flux. Many previous studies have shown that the vertical distribution of aerosol and cloud, and their optical properties can have significant impacts on photolysis rate [e.g. Liao et al., 1999; Lefer et al., 2003; Tie et al., 2003; 2005; Liu et al., 2006]. In general, actinic flux and thereby photolysis rates are reduced below aerosol or cloud layer due to their extinction. On the other hand, over bright cloud the strong cloud reflection can increase the photolysis rate. The impacts on photolysis rate can in turn influences the photochemistry of ozone. Using a photochemical box model driven by airborne measurement from the TRACE-P mission, Lefer et al. [2003] showed that during the TRACE-P mission the net photochemical effect of clouds and aerosols was a large decrease in photochemical O3 production in the boundary.

The interactions between particulate matters suspended in the air – or atmospheric aerosols with cloud fields are complicated and extremely important for climate as well as air quality application [Rosenfeld, et al., 2014, Fan et al., 2016; Seinfeld et al., 2016]. The cloud droplets must be nucleated from aerosol particles, which are referred as cloud condensation nuclei (CCN) if activated. Heavy pollution condition in metropolitan areas can enhance cloud droplet number concentration (CDNC). For fixed amount of liquid water mass, higher CDNC leads to smaller cloud droplets, which can reduce precipitation efficiency. As the precipitation falling, the raindrops can wash off the aerosol particles below the clouds. Aerosol embedding inside cloud droplets and wet removal of aerosols by raindrops, referred to as in-cloud and below cloud scavenging, represent important sink terms of atmospheric aerosols. Therefore, to a large extent, the atmospheric aerosols can control cloud microphysical and macrophysical properties, and vice versa.

Aqueous chemistry is another reason that cloud is important for air quality application. For example, SO2 mass can be efficiently transferred to sulfate aerosols via cloud processing [Wine et al., 1989; Feingold and Kreidenweis, 2002]. To sum up the discussion, as an input to the air quality models, accurate representation of cloud fields, including their macro and microphysical properties by model is essential for the air quality prediction application.

Modeled clouds are often too bright (high cloud brightness) and too few (low cloud fraction) compared to satellite observation [e.g. Otkin et al., 2008; Thompson et al., 2016]; however, the "general pictures" of cloud fields can be well captured by the meso-scale weather prediction models, for instance, convective frontal clouds associated with cyclonic-frontal system; or large decks of cumulus clouds over a large area when atmosphere is stable. For example, WRF model is widely used in simulating the meteorology and cloud fields that are essential for air quality prediction. WRF model parameterizations has been shown to lead to accurate simulations of southeast Texas mesoscale circulations [Ngan et al., 2013]. This indicates that over a relatively large area, the characteristic of modeled clouds is reasonable statistically and can be "adjusted" to match the observations. The direct comparison between modeled and observed cloud fields are like "apple-to-orange" comparison, because of different sampling rate. To facilitate the so-called "apple-to-apple" comparison, we must firstly use the tool called COSP [Bodas-Salcedo et al., 2011; Zhang et al., 2019].

As a new technique, the machine learning and deep learning (ML/DL) tools have not been widely used in geoscience, but have shown great potentials [e.g. Figure 1.1 obtained from Reichstein et al., 2019]. One of the advantages of ML/DL is that is data-driven – in other words, the more data we feed into the tools, the more accurate the results will be. The goal here is to get the clouds right in a period of high ozone concentration – This is challenging, for example, we need to get the properties of diffuse, fair weather cumulus right, and we need to eliminate false positive signals too. With satellite observation, we have large amount of satellite data available for training the ML/DL tools. In this work, we will use a ML/DL tool called Generative Adversarial Network (GAN) to "adjust" modeled cloud fields and add fine features in modeled cloud fields [Goodfellow et al., 2014].

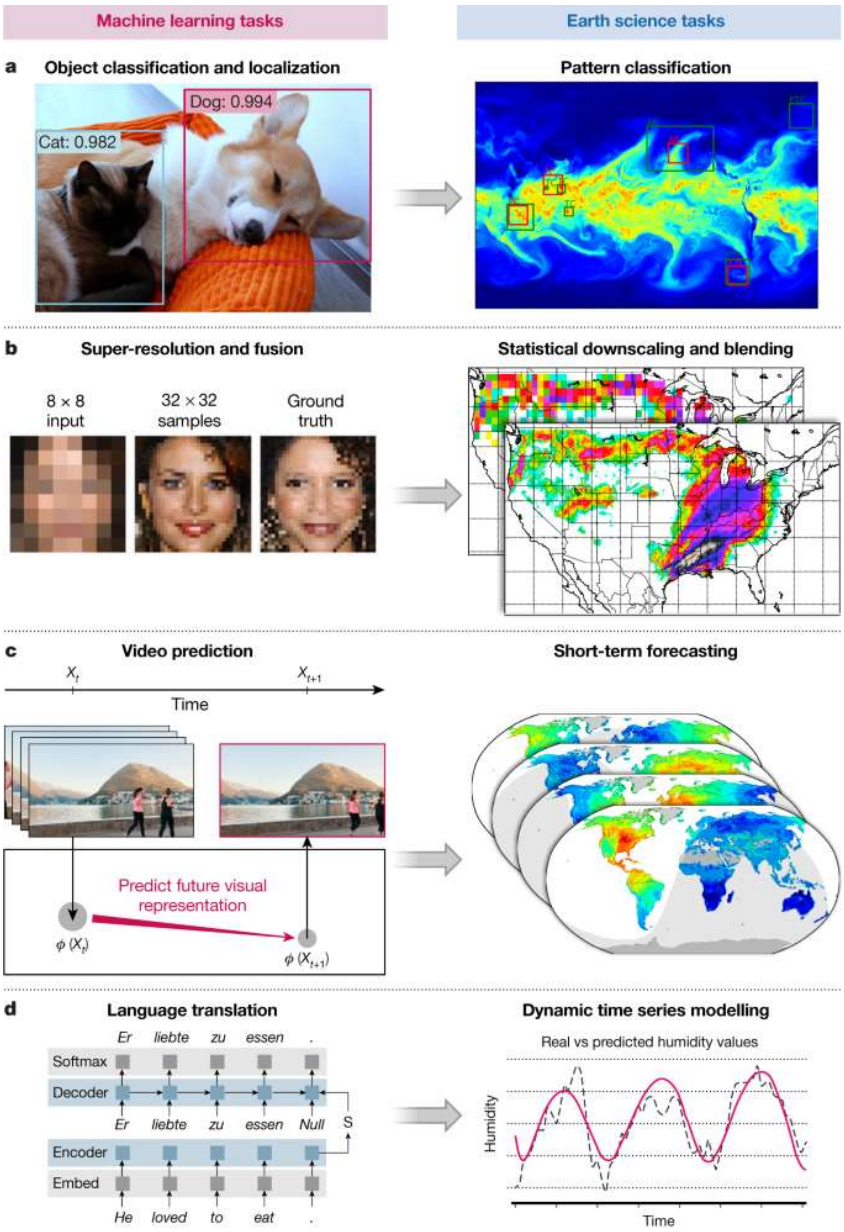


Figure 1.1 Figure is obtained from [Reichstein et al., 2019] demonstrating the potential capability of ML/DL.

1.2 Project Objectives

The overarching goal of this study is to improve the WRF modeled cloud fields using DL tool GAN. The two objectives of this study are:

(1) To conduct a series of WRF simulations as well COSP analysis to find an optimal combination of physics suite and reanalysis input for modeling clouds fields over Texas.

(2) To train a GAN model over the time series of modeled cloud fields so that the macro- and microphysical properties of modeled clouds are more accurate compared to the observations.

The methodology, including model configuration and data will be discussed in section 2. The results obtained from this study will be presented in sections 3 and 4. The audit of data will be discussed in section 5, followed by conclusion in section 6.

2. Methodology

2.1 MODIS cloud product

The cloud products from the Moderate Resolution Imaging Spectroradiometer (MODIS) on board the NASA Terra and Aqua satellites [Levy et al., 2009] play a central role in our study since both the evaluation of model performance and training of GAN heavily relies on them. In our study, we used level 2 MODIS products with spatial resolution of 1km (for cloud liquid water path and cloud optical depth products) and 5km (for cloud fraction products) downloaded from NASA website: https://ladsweb.modaps.eosdis.nasa.gov/.

2.2 WRF model configuration

In this study, we use Weather Research and Forecasting model (WRF) [*Skamarock & Klemp*, 2008] to generate cloud fields, that can be used in air quality forecasting application. The domain will be set up with Texas in the center. Figure 2.1 shows the simulation domains with nesting configuration. Both outer and inner domains have 256 (west-east) by 256 (south-north) grids. The horizontal resolutions are 12 km and 4 km for outer and inner domains, respectively.

In the latest version of WRF model (after V3.9), a suite of physical packages is specifically recommended for simulations over CONUS (CONtinental U. S.). Namely, they are new Thompson microphysics scheme [Thompson et al., 2008], modified Tiedtke scheme for cumulus parameterization [Tiedtke, 1989]; Mellor-Yamada-Janjić TKE scheme for boundary layer scheme (PBL) [Janjić, 1994]; RRTMG radiation scheme for both shortwave and longwave radiation calculation [Iacono et al., 2008]; and unified Noah land-surface model [Koren et al., 1999]. For CONUS application, the initial and boundary conditions (IC and BC) of model is often driven by 6-hourly 12-km North American Mesoscale Analysis [e.g. Li et al., 2008].

However, this physics suite as well as reanalysis input may not be optimal for Texas application and/or cloud field simulations. Based on our previous research experience [Lu and Sokolik, 2013; Lu and Sokolik, 2017; Lu et al., 2018], two physics packages, namely microphysical scheme and PBL scheme, are extremely important physics packages that affect cloud simulation. The selection of re-analysis data also strongly affects large-scale dynamic and resulting cloud deck patterns. Therefore, here we run several groups of one year of simulations with different combination of physics packages and reanalysis datasets, the candidate of which are shown in Table 1.

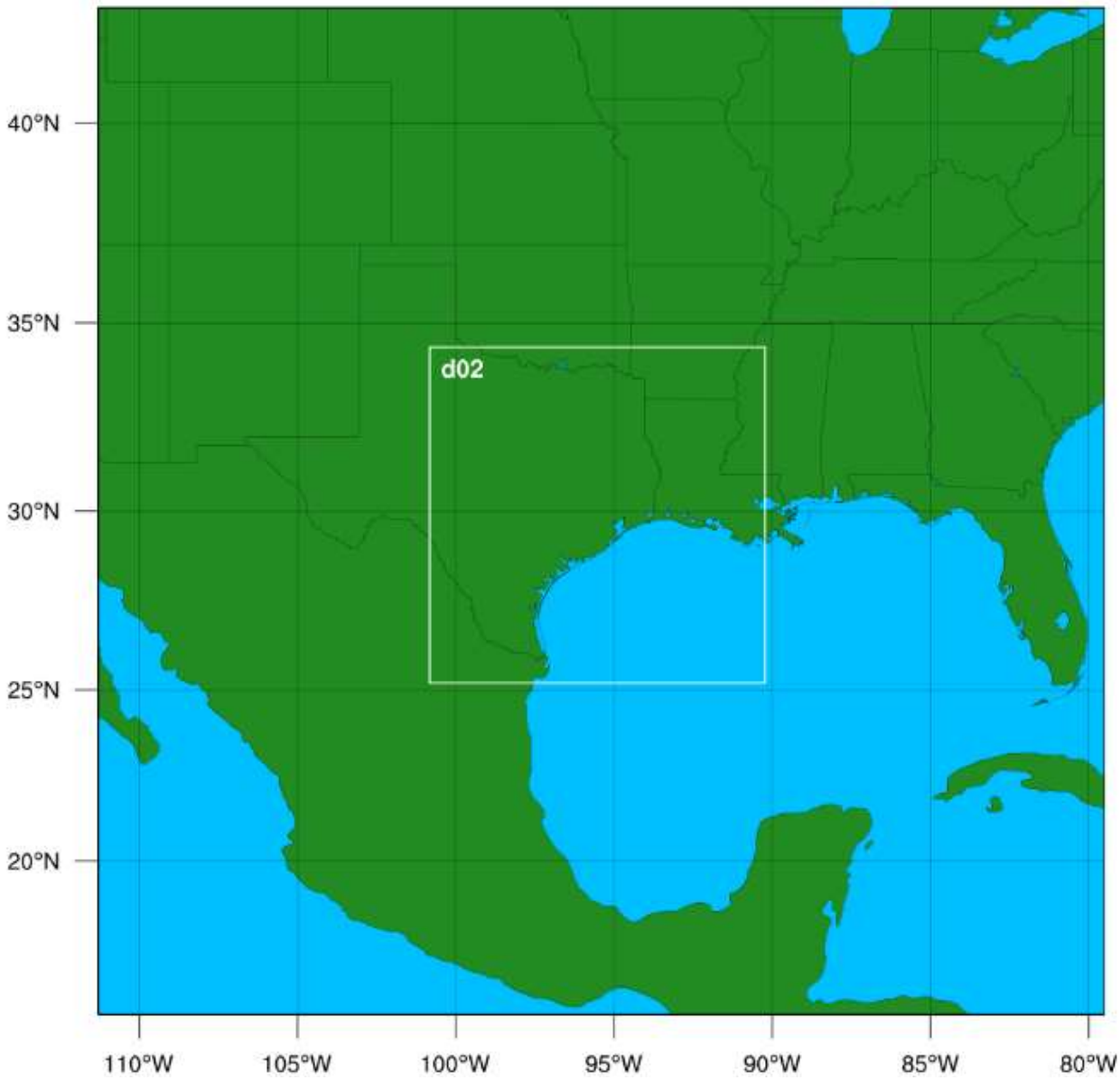


Figure 2.1 Outer domain and inner domain setup with Texas in center

Table 2.1 Physics packages and reanalysis-data used for WRF simulation

Physical parameterization scheme		Acronym	Reference
Cumulus	Multiscale Kain-Fritsch	msKF	Zheng et al. [2016]
Microphysics	1.5-moment 6-class Thompson	Thompson	Thompson et al. [2008]
	2-moment 6 class Morrison	Morrison	Morrions et al. [2009]
	WRF Single-Moment 6-class scheme	WSM6	Hong et al.[2006a]
PBL	Asymmetric Convective Model PBL	ACM	Kolling et al. [2013]
	Yonsei University scheme	YSU	Hong et al. [2006b]
	Grenier-Bretherton-McCaa scheme	GBM	Bretherton et al. [2004]
Reanalysis input	North American Mesoscale Analysis	NAM	Rogers et al. [2009]
	NCEP final (FNL)	FNL	NCEP [2000]
	ECMWF	ECMWF	Gibson et al. [1997]

The dynamical core used in this study is the Advanced Research WRF (ARW), which is based on the fully compressible, non-hydrostatic Euler equations, with a terrain-following eta coordinate. For outer domain, we will use multiscale Kain-Fritsch cumulus parameterization, the performance of which is tested over Texas area [Zheng et al., 2016]. As for 4km domain, the cumulus convective scheme can be turned off since the WRF model can explicitly resolve the vertical motion. The microphysics scheme deals with very complex mass/number transfer between different hydrometeors through microphysical processes (see an example in Figure 2.2). The Thompson scheme is considered as 1.5-moment because that the mass mixing ratio and number concentration of cloud droplet is predicted, while other only the mass mixing ratios of other five hydrometeors are considered. In contrast, the Morrison scheme is considered as two-moment scheme because that the mass mixing ratios and number concentrations of five hydrometeor species: cloud droplets, cloud ice, snow, rain, and graupel are all predicted. The scheme that predicts two moments can better capture the radii of hydrometeors. Using each species own number distribution and mass-diameter assumptions, a fully consistent effective radii is computed in the microphysics body of code and subsequently passed to the RRTMG interface code. Since the usual size of rain drops and graupel particles is far larger and the number density far lower than the other three species, rain and graupel were neglected in the radiation treatment as is currently done within all WRF radiation schemes at this time. WRF Single-Moment 6-class scheme (WSM6) is another commonly used microphysics scheme [Hong et al., 2006a]. The advantage of WSM6 scheme is its computational efficiency since only one moment of each hydrometeor is tracked. However, it should be noted

that the performance of two-moment scheme is shown to be better compared to the one-moment scheme in terms of cloud simulation [Morrison et al., 2009].

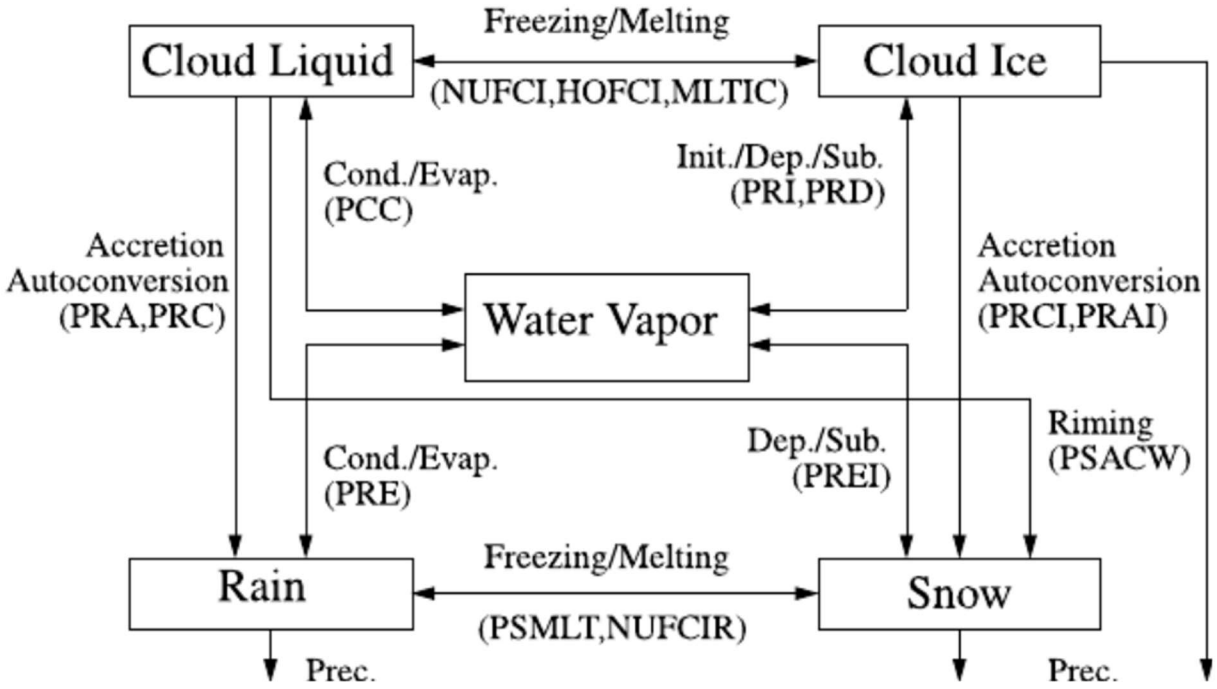


Figure 2.2 Diagram of microphysical processes between water vapor and hydrometeors, including cloud liquid (droplet), cloud ice, rain, and snow (also graupel) considered in WRF microphysics scheme (Figure from [Morrison et al., 2003]).

The impact of three PBL parametrizations has also been addressed. The options tested to describe vertical subgrid-scale fluxes due to eddy transport in the atmosphere were the first-order closure PBL schemes Yonsei University (YSU, used in the control run), the Asymmetrical Convective (AC) model version 2 (ACM2) scheme, and the Grenier–Bretherton–McCaa (GBM) PBL scheme.

The YSU scheme is a non-local model and considers the fluxes implicitly, through a parametrized non-local term, and uses a counter gradient term in the eddy-diffusion equation. It is based on the Medium Range Forecast (MRF) model PBL scheme and improves it with an explicit treatment of the entrainment. The Asymmetric Convective Model 2 (ACM2) (Pleim 2007) uses a combination of local (first-order) and non-local transport that switches off smoothly to local eddy diffusion in stable environments. At unstable conditions, local transport is used for subsidence, while upward fluxes are modeled combining local eddy diffusion with a non-local approach that computes the transition probability between non-consecutive levels. The latter approach can represent rising thermals much larger that the grid spacing. The combination between local and transilient approaches is weighted with a parameter that depends on stability. According to Goehn

et al. (2015), one of the advantages of the GBM scheme is that reductions to the stratocumulus cloud deck owing to vertical mixing are found to be well handled for vertical grid spacing at or smaller than 15 mb. This could be relevant for depicting the impact of stratocumulus clouds on buoyancy preceding potentially severe convection.

2.3 COSP package

Direct comparison between model outputs with satellite observations is challenging because of different spatiotemporal sampling of clouds; however, with the aid of COSP, the comparison becomes possible [Bodas-Salcedo et al., 2011; Zhang et al., 2019]. One big strength of COSP is to facilitate "apple-to-apple comparison of observed cloud data and model-simulated cloud as shown in the example in Figure 2.3.

Modeled vertical profiles of temperature, humidity, hydrometeor mixing ratios, cloud optical thickness and emissivity (a function of cloud water content and particle size), as well surface temperature at satellite overpassing time are feed into COSP. In addition, other cloud properties, which are not directly provided in WRF outputs, are also required by the MODIS simulator. To overcome this lack of data, the same assumptions made by the radiation scheme used for simulations (CAM3) were applied to compute the rest of the required inputs. The cloud droplet effective radius in the Morrison scheme and the Thompson scheme are directly calculated because of both mass and number concentrations for cloud droplet are predicted. Same for the cloud ice particle effective radius in the Morrison scheme. The cloud droplet effective sizes in the WSM6 scheme were computed as proposed by Kiehl *et al.* (1994) and the ice particle effective radius in the Thompson scheme and the WSM 6 scheme following the function of temperature as proposed by Kristjánsson *et al.* (2000). The short-wave radiative properties for liquid clouds follow a generalization of the expression used by Slingo (1989), and for ice clouds they are based on the results of Ebert and Curry (1992). For long-wave radiation, clouds are considered grey bodies with emissivity that depend on condensed water path, cloud phase and cloud particle sizes (Collins *et al.*, 2004).

Firstly, the vertical profiles of model grids are broken into sub-columns to commensurate satellite pixels. Next, vertical profiles of sub-columns are passed to several instrument simulators, which apply models to simulate the radiance signals received by each sensor. Finally, statistical modules gather output from all instrument simulators, and build pseudo-cloud fields that can be directly compared to observations. In our study, we examined the following fields of cloud: **cloud liquid water path** (LWP, in kg m^{-2}); **cloud fraction** (CF, in %) and **cloud optical thickness** (COT, unitless).

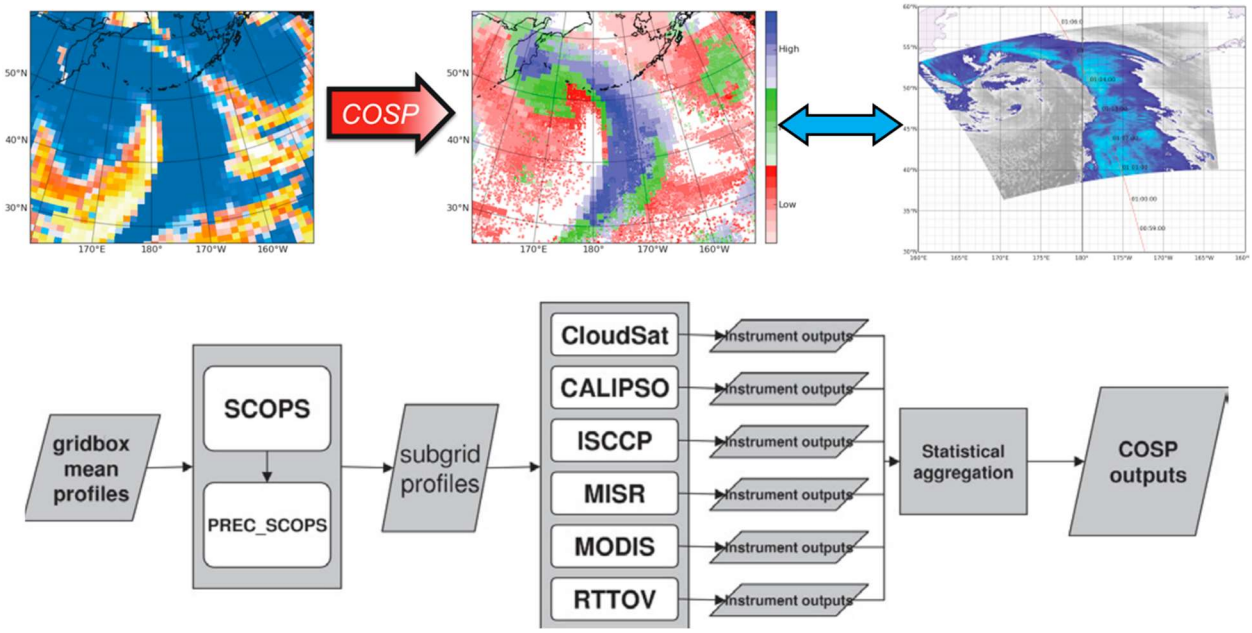


Figure 2.3 Upper panel: example of COSP: modeled cloud fields are converted to pseudo-satellite observations, then compared against actual satellite observations [modified from COSP webpage on https://climatedataguide.ucar.edu/]; Lower panel: flow chart of COSP, adopted from Bodas-Salcedo [2011]

2.4 Generative adversarial network (GAN)

Generative adversarial networks (GANs) are a type of deep learning technique [Goodfellow et al., 2014] that is commonly used in many areas (e.g. super-resolution application that can enhance the details of images). A GAN contains two neutral networks (NN), a generator and a discriminator. The purpose of the generator is to generate fake samples of data/image and tries to "fool" the discriminator. The discriminator on the other hand tries to distinguish the real and fake samples — in other words, two NNs try to compete each other and play zero-sum game. The GANs are formulated as a mini-max game, where the discriminator is trying to minimize its reward V:

$$min_G max_D \mathrm{V(D, G)} = E_{x\sim p_{data}}[\log D(x)] + E_{z\sim p_z}[\log(1 - D(G(z)))]$$

,where x is satellite observed images of LWP, CF, or CTH, and COT, z is COSP simulation outputs of LWP, CF, and COT.

We consider the 2D cloud properties (LWP, CF, and COT) as different layers of one "image" and apply only one GAN model training. To prepare input data and target data for GAN training, we run multiple years of WRF simulations with the optimal configuration, feed vertical profiles of variables into COSP, which generate pseudo-observed CF, LWP, and COT, as input data for the generator to generate fake cloud fields. Target or real fields is simply the corresponding observed MODIS CF, LWP, and COT fields.

Figure 2.4 shows the workflow of GAN training, which contains two parts. In the first part, only discriminator is trained as the network is only forward propagated. The discriminator is trained on target data (observed cloud fields) for n epochs and see if it can correctly predict them as real. Also, in this part, the discriminator is also trained on the fake generated cloud fields from the generator and see if it can correctly predict them as fake. In the second part, the generator is trained while the discriminator is idle. After the discriminator is trained by the generated fake cloud fields of the generator, we can get its predictions and use the results for training the generator and get better from the previous state to try and fool the discriminator. The above method is repeated for a few epochs and then manually check the fake cloud fields how it seems compared to target cloud fields.

Figure 2.4 also shows the architecture of two deep NNs. The generator has this "encoder-decoder" structure. The encoder part of the model is comprised of convolutional layers that use a 2×2 stride to downsample the input source "image" down to a bottleneck layer. The decoder part of the model reads the bottleneck output and uses transpose convolutional layers to upsample to the required output image size. Both encoder and decoder use ReLU activation function. The Adam optimizer will be used in training [Kingma & Ba, 2014].

The code that we used and modified is called "pix2pix" [Isola et al., 2017], which is written in Python with DL/ML package PyTorch V1.4 (https://phillipi.github.io/pix2pix/). The pix2pix code is widely used in many image-to-image translation problems, such as "Labels to Street Scene", "Black-White to Color", "Aerial to Map", and "Edges to Photo" applications. The examples are shown in Figure 2.5.

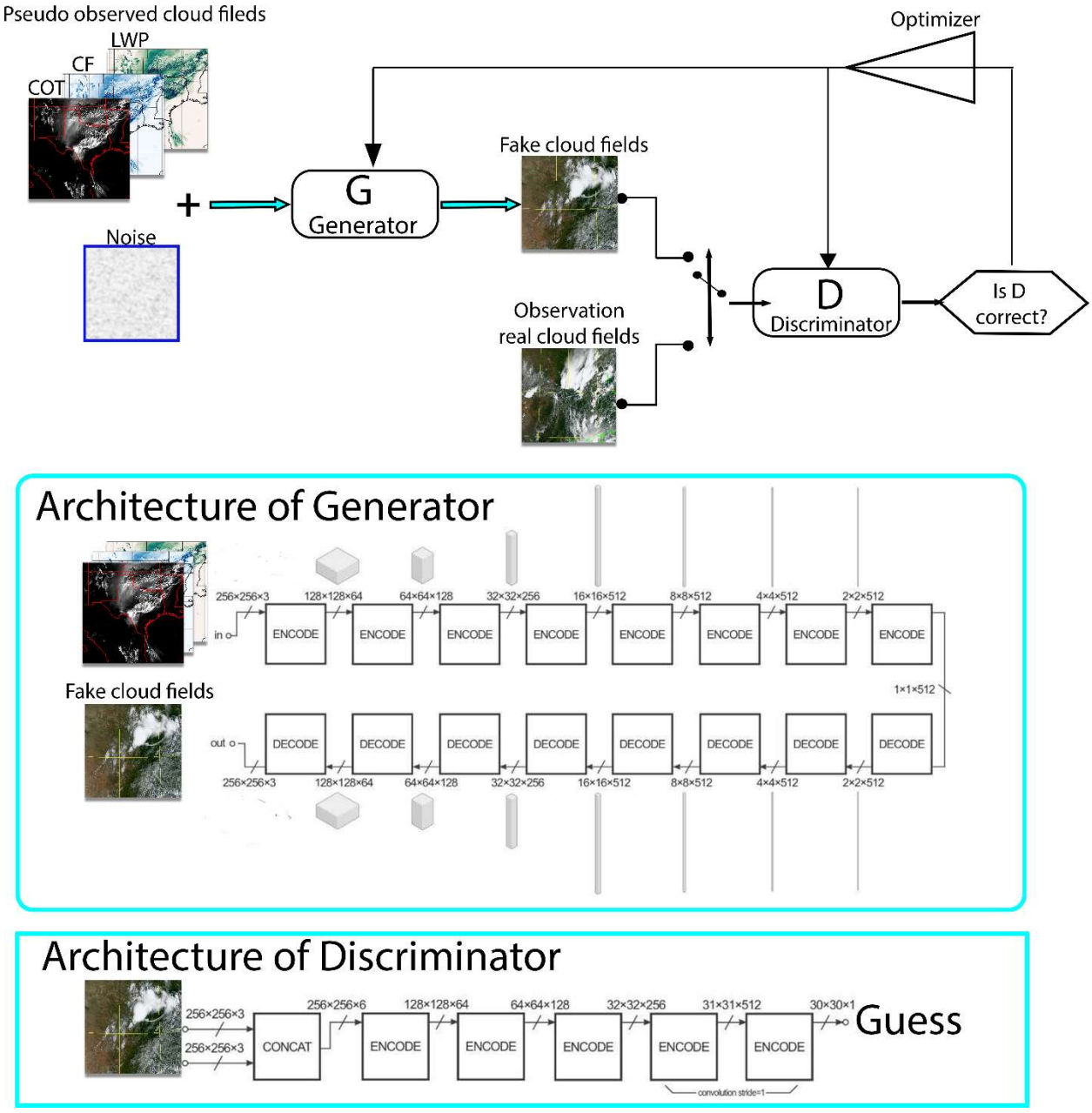


Figure 2.4 Workflow of GAN training and the architecture of generator

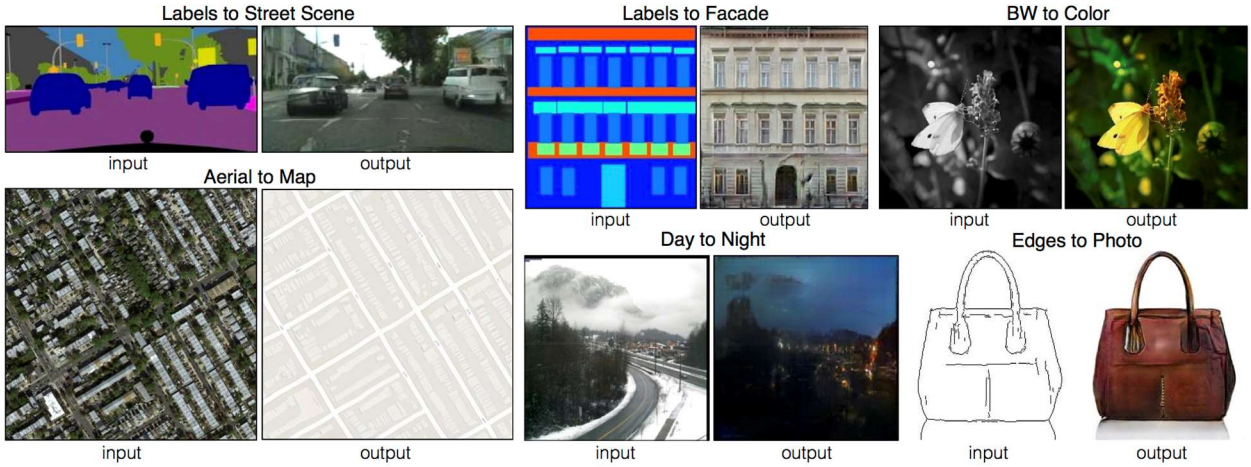


Figure 2.5 Application examples for image-to-image translation problem package "pix2pix". (Image downloaded from https://phillipi.github.io/pix2pix/).

2.5 Procedure

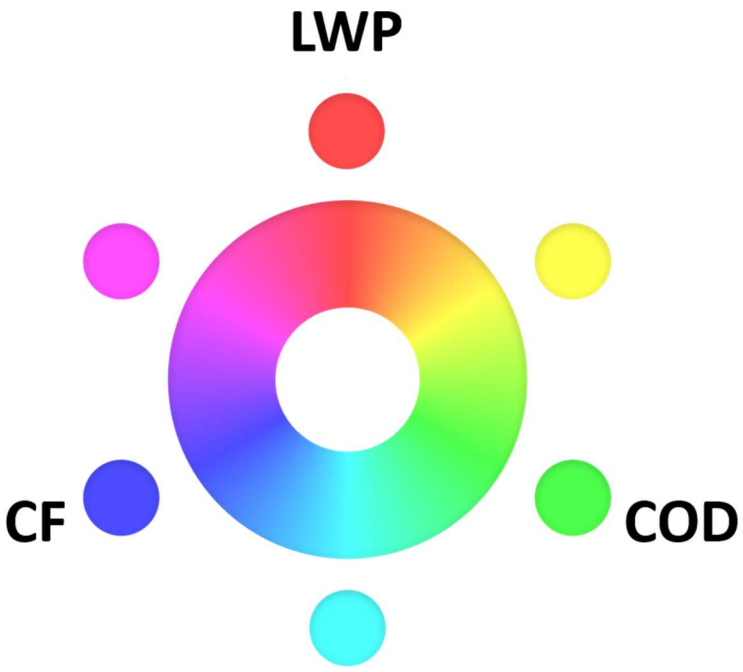
Totally 3×3×3=27 groups of simulations for entire year 2018 will be performed. For each case, we feed the model outputs into the COSP, and examine the GOSP outputs of LWP, CF, and COT. Please note that the cloud fields that are compared against satellite observations are all from COSP. Usually for each day, MODIS will generate two snapshots (granules) of 2D cloud fields over Texas (10:30 and 13:30 local time). Here we select the one observed by the Aqua satellite in local afternoon evaluate the performance of all 27 groups of WRF simulations. Model performance during each snapshot will be evaluated using the spatial correlation coefficient (Pearson correlation of the fields, r. In addition, we also added normalized mean bias (NMB), which is calculated as:

$$NMB = \frac{1}{N}\frac{\sum_{i=1}^{N}(M_i - O_i)}{\sum_{i=1}^{N} O_i}$$

N is the total number of COSP grids, M_i and O_i are the modeled and observed cloud field values over the grids. The objective of using NMB as metric is to evaluate whether the WRF model systematically under- or over-estimates the cloud water amount and cloud fractions over the domain. The correlation coefficient is to evaluate the spatial pattern of cloud simulations. For each snapshot, we rank the performances of 27 experiments in simulating cloud field variables based on NMB and r and score them. The experiment that achieves the highest score will be considered as the optimal configuration.

After optimal combination of physics packages is selected, we use this physics suite with reanalysis data and conduct multiple years of simulation from 2005 to 2020 (when the reanalysis data available) with the same domain setup. The COSP outputs from these simulations will be used in training a generative adversarial network (GAN). A well-trained GAN is expected to 1) adjust large-scale cloud distributions. 2) We can generate the fine features associated with modeled cloud decks, for example, adjust cloud fractions associated with diffuse fair-weather cumulus. 3) We can improve the accuracy of modeled cloud so that COT and LWP become much closer to the observations.

We plan to train three cloud fields at the same time. Therefore, for both modeled cloud variables and MODIS cloud product, we compose false-color image from LWP, COD, and CF fields, which corresponds to the R(ed), G(reen), and B(lue) channels in the figure. The color level of each channel is from 0 to 255, corresponding to cloud field values from zero to maximum. Please note that for LWP and COD, the value levels are firstly linearly distributed from 0 to 220 with relatively small intervals, then linearly distributed from 221 to 255 with relatively large intervals. An example of this composed false-color image from MODIS observations is shown in Figure 2.6. The region with cyan color indicates high CF and COD values, but low LWP values. Similarly, the region with magenta color represents high CF and LWP values, but low COD values. The regions with the white color indicate that all three values are high. It is extremely rare to see yellow color in the composed figures, because once MODIS can retrieve LWP or COD, CF will be automatically assigned values. We also prepare Python script that converts the pixels of false-color image of clouds to gridded values of CF, LWP, and COD in netcdf format.



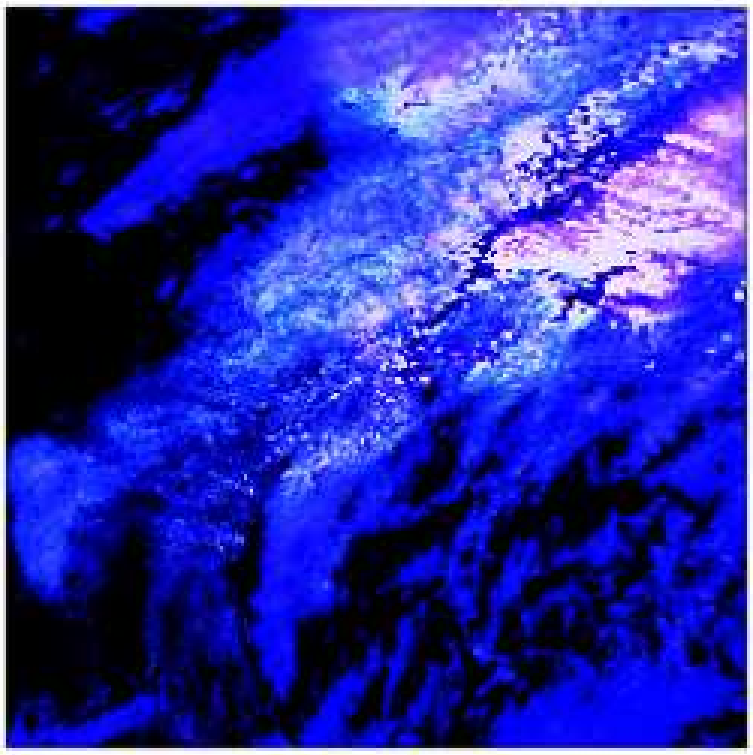


Figure 2.6 (a) RGB color scheme for false color image. (b) an example of false color image for MODIS cloud products.

3. Identify the Optimal Configuration of WRF Simulation

3.1 Simulation statistics

The simulations were conducted on the NCAR supercomputer Cheyenne with 288 CPU cores. Each one-year case costs about 4,000 core hours, which is equivalent of running case for 14 hours. Each one-year case produces the outputs with total size of 360-440 G byte (the restart files excluded). The COSP outputs for each one-year case is about 80 G byte (inner domain only).

3.2 Examine the importance of each physics packages.

In this section, we examine the variance among 27 cases, and identify the relative importance of the microphysics scheme, the PBL scheme, and the reanalysis inputs. Firstly, we use the downwelling shortwave fluxes (SW↓@SFC) at the surface during 12:00pm local time as a proxy for the cloud fields. The SW↓@SFC is an ideal proxy because it represents the synergistic effects of clouds. Lower the value of SW↓@SFC indicates higher cloudiness in the inner domain. In Figure 3.1a, we grouped all 27 cases into 3 groups according to the microphysics schemes (the Morrison, the Thompson, the WSM6 schemes) used in the cases, and plotted the time series of mean values of 9 cases during December 2018. The same analyses are performed again in Figure 3.1b and c, but for examining the importance of PBL schemes and the reanalysis inputs, respectively. We can find that the curves in Figure 3.1a are more widespread compared to Figure 3.1b-c, indicating that the microphysics schemes playing a more important role in terms of explaining the variance of cloud field simulation than the other the choice of PBL scheme and reanalysis inputs for December 2018. Similarly, we perform the same analysis for all months of 2018 and calculate the monthly averaged standard deviation for three physics packages. The result shown in Figure 3.2 indicates that PBL scheme plays a more important role in June and July in terms of determining the cloud fields, while for the other months, the microphysics scheme is more crucial. During the summer, the PBL variance within a day is relatively larger compared to the other seasons. The coupling-decoupling process associated with the transition from stratocumulus clouds to cumulus clouds over the Gulf of Mexico is critical since they are the major cloud types during the summer season. While in other months, the cloudiness is more likely associated with large-scale dynamics, like the frontal-cyclone system. In this case, the choice of microphysics schemes is more important.

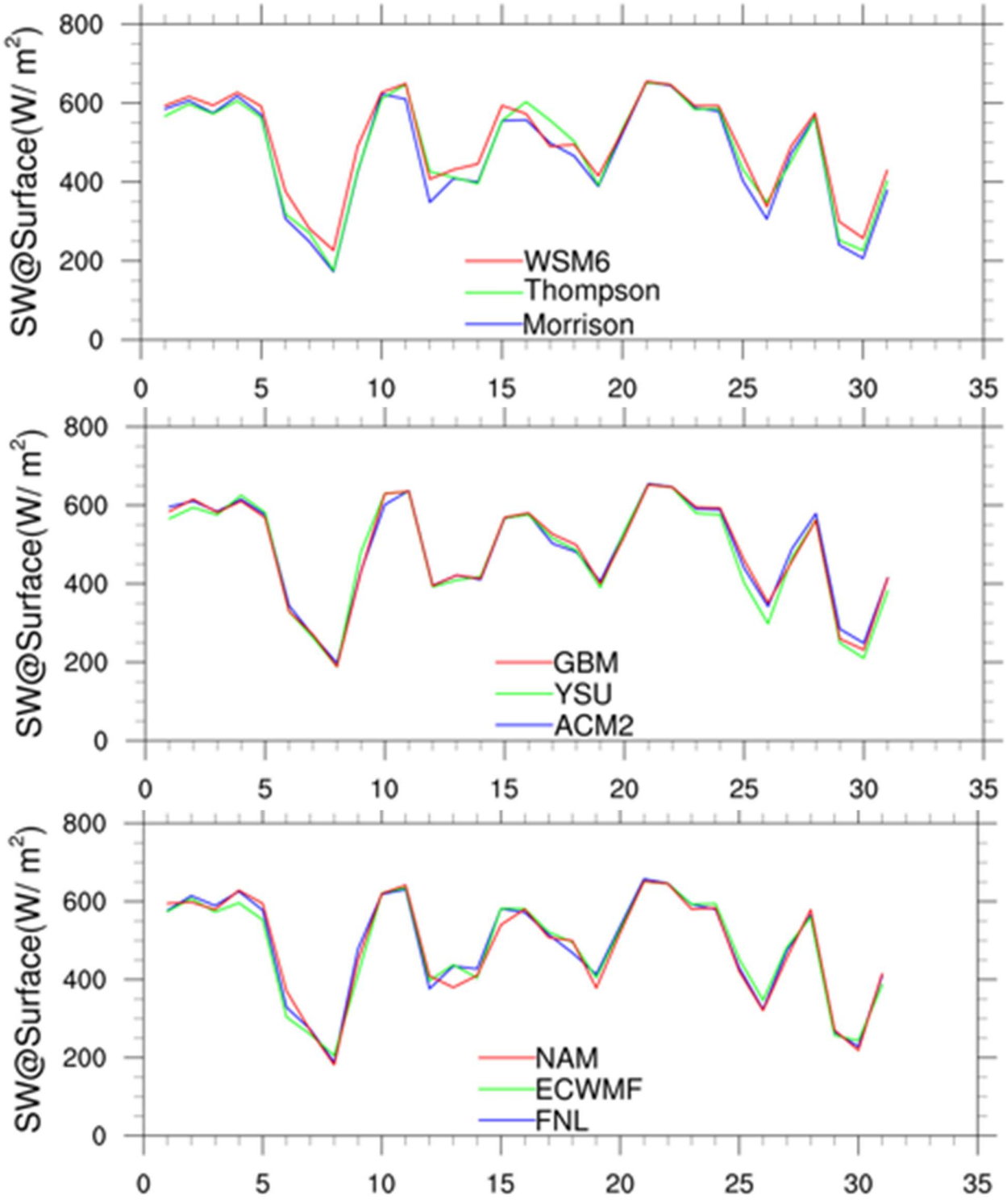


Figure 3.1 Daily SW↓@SFC at noon during 2018 December. Each curve represents the mean values of 9 cases.

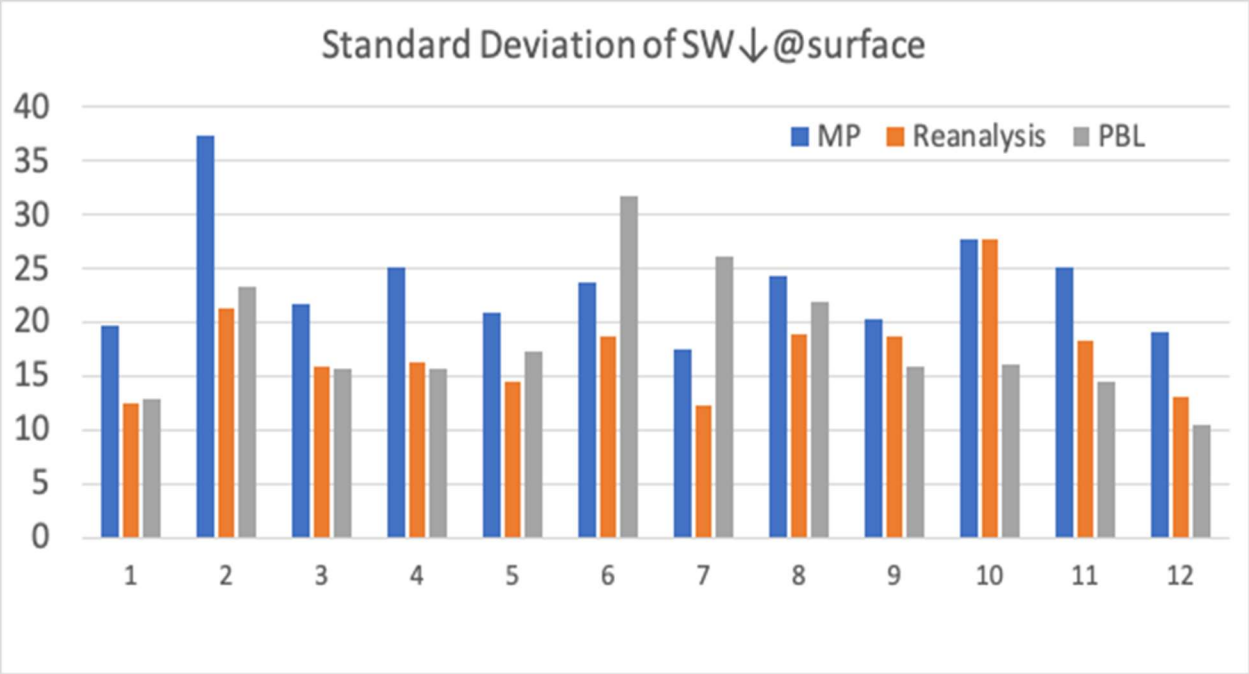


Figure 3.2. Monthly averaged standard deviation of SW↓@SFC as functions of three factors, namely microphysics scheme, PBL scheme, and re-analysis data.

In Figure 3.3, we examine the averaged LWP fields of four seasons that are modeled by 27 cases during the local noon. In Match-April-May season as shown in Figure 3.3a, we find that the cases with the Thompson scheme produce higher LWP values, especially for the cases with YSU and GBM PBL schemes. These high values of LWP are associated with the frontal-cyclone systems that impacted Texas. The cases with ACM PBL scheme and/or WSM6 scheme produce the lowest LWP values and fails to predict enough cloudiness in the domain (very small CF). The performances of different cases differentiate from each other significantly during the summer period (June-July-August). Some cases, especially the cases with the WSM6 and the Thompson microphysical schemes, failed to reproduce enough clouds. The clouds over land seem to be very sparse, while the CF values over the Gulf of Mexico are very low. Out of 27 cases, the cases with the YSU PBL scheme and the Morrison microphysics scheme predicts high CF over the sea, and reasonable amount of CF over the land. In the fall season (September-October-November) and winter season (December-January-February), we found that the cases with the Thompson scheme predicts highest LWP that are associated with the large-scale dynamics system, followed by the Morrison scheme, then by the WSM6 scheme.

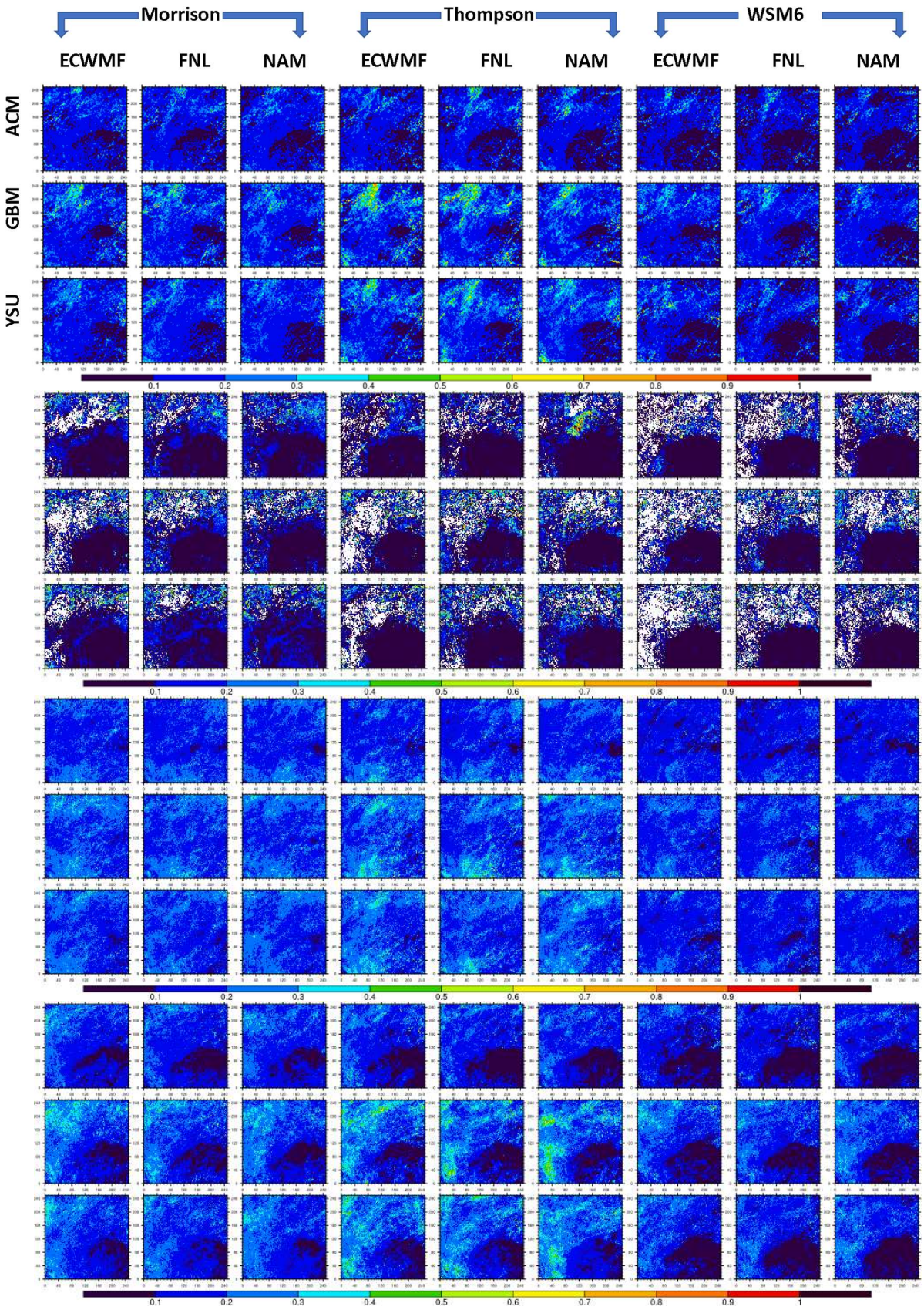

Morrison
Thompson
WSM6
ECWMF
FNL
NAM
ECWMF
FNL
NAM
ECWMF
FNL
NAM
ACM
GBM
YSU
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Figure 3.3. Averaged LWP fields of each season (from top to bottom: MAM, JJA, SON, and DJF of 2018) as modeled by the 27 cases.

3.3 Identify the optimal configuration of WRF simulation

In this section, we compared modeled cloud properties of 27 cases against the Aqua MODIS observation, and rank their performances. In Figure 3.4, we shows the time series of daily CF (unit: %) modeled or measured in local afternoon (15 LST). We find that model signicantly underestimate CF. As discussed in the previous section, the understimation of CF is especially significant during the summer because of too quick decoupling process between marine statocumulus-topped boundary layer with underlying sea surface. The largest discrepancy between observation and simulation in CF can be as high as 60~80%. By average, the CF during the winter and spring is high. The variation of CF of each day is also large because the cloudiness in domain is controlled by large-scale dynamics. Model can generally capture these highs and lows, however, CF is still underestimated. For example, the CF modeled by 27 cases sometimes reaches nearly zero, while according to MODIS observation, there are still more than 20% of CF in the domain.

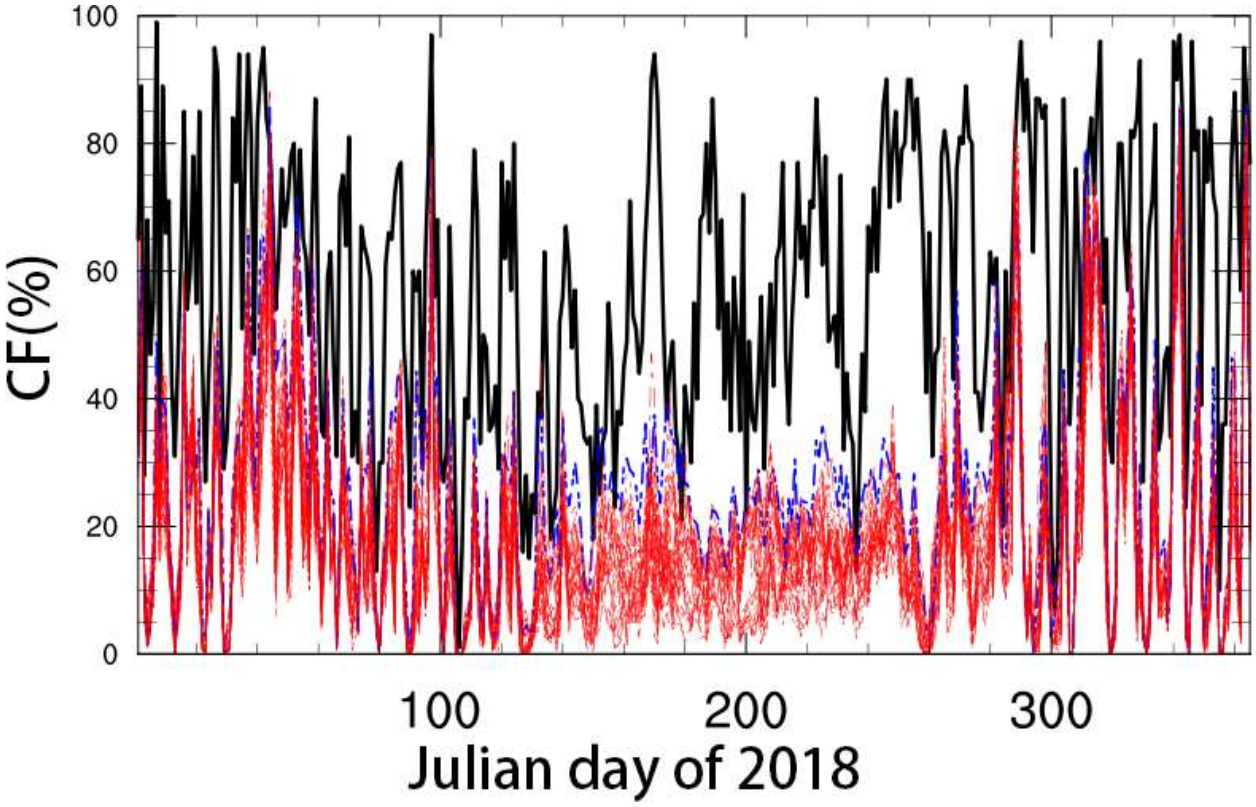


Figure 3.4 Daily cloud fraction (%) of year 2018. Black line represents daily MODIS retrieved cloud fraction. Blue line represents WRF simulation of YSU_MORR_NAM case, while red lines represent all other WRF simulation cases.

Table 3.1 lists five cases with the top performance is simulating CF based on the metrics of normalized mean bias (NMB). Four cases with YSU PBL scheme predicts less biased CF compared to many other cases. The case that ranks the first place is the one with the YSU PBL scheme, the Morrison microphysics scheme, and the NAM reanalysis inputs (blue line in Figure 3.4). However, the case still underestimates the CF over Texas by an enormous amount of 53.6%. We also examine the performance of each cases by calculating the spatial correlation coefficient. The case with the ACM2 PBL scheme, the Morrison microphysics scheme, and the NAM reanalysis correlates with the MODIS observation with the correlation coefficient as high as 0.61 (n=365). (The case ranks the fifth place by the metric of NMB.) For the aforementioned four cases with the YSU PBL scheme, the spatial correlation coefficients of two cases with the NAM reanalysis input apparently performs much better than the two cases with the FNL reanalysis inputs. The YSU_MORR_NAM case ranks the eighth out of total 27 cases with the correlation coefficient of 0.56.

Table 3.1 Ranking of model cases in terms of simulating CF.

Ranking	Normalized mean bias (NMB)	Ranking by correlation
1st	YSU_MORR_NAM (-53.6%)	8th (0.56)
2nd	YSU_THOM_NAM (-57.1%)	7th (0.56)
3rd	YSU_MORR_FNL (-58.3%)	21st (0.53)
4th	YSU_THOM_FNL (-60.1%)	26th (0.51)
5th	ACM_MORR_NAM (-61.7)	1st (0.61)

Figure 3.5 shows the time series of domain-averaged cloud liquid water path (LWP) as observed by MODIS Aqua and modeled by 27 cases for local afternoon (15 LST) of 2018. During the spring and winter, model performed very well in simulating domain-averaged LWP in comparison to MODIS observation. Model successfully captures the magnitudes of the peaks of LWP, except several occasions. However, considering underestimated CF, model undoubtedly overestimates the in-cloud LWP. During the summer, the domain-averaged LWP is underestimated, but in-cloud LWP is likely reasonably simulated. This is the well-known issue of cloud modeling – “too few too bright”.

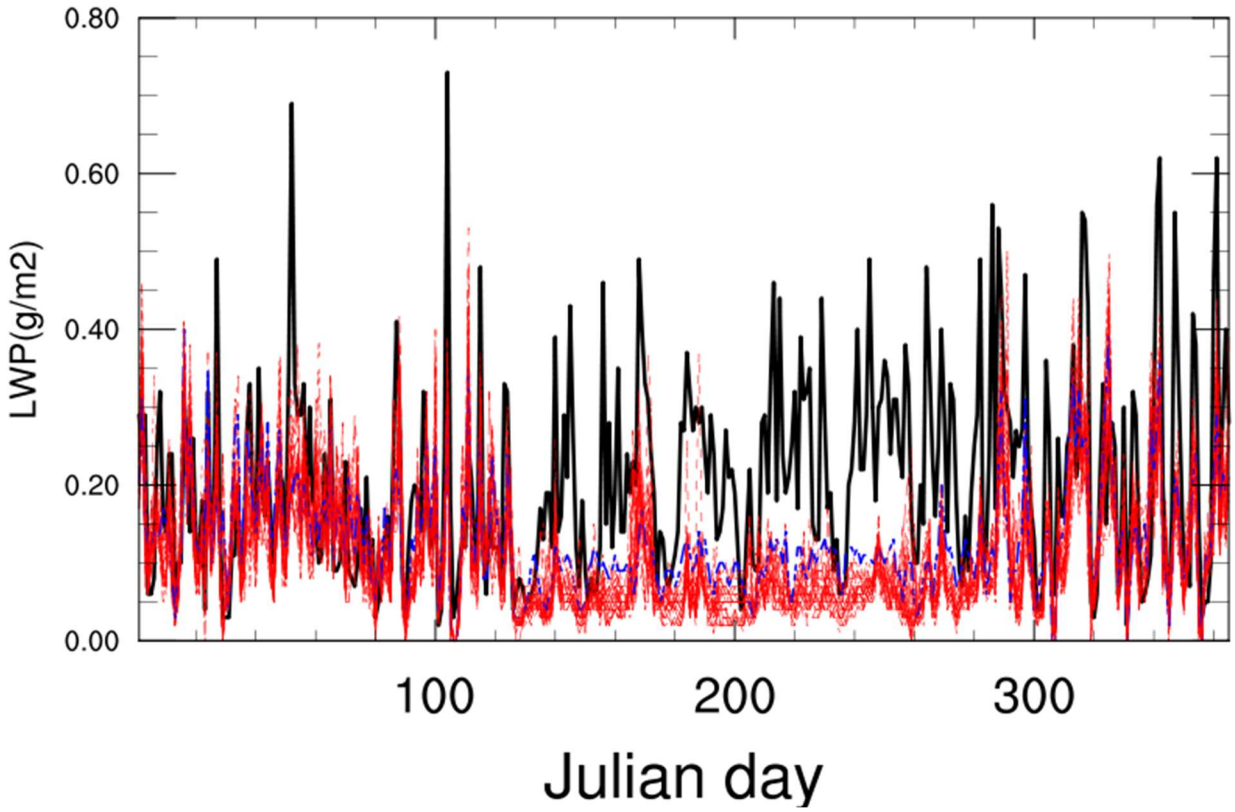


Figure 3.5 Daily cloud liquid water path (kg/m^2) of year 2018. Black line represents daily MODIS retrieved cloud fraction. Blue line represents WRF simulation of YSU_MORR_NAM case, while red lines represent all other WRF simulation cases.

Table 3.2 list the five cases with top performance in terms of simulating domain-averaged LWP based on the metrics of NMB. By average, these five cases, which all feature the YSU PBL scheme, underestimated domain-averaged LWP by about 18%-20%. The case YSU_MORR_NAM still ranks the first place with NMB of -18.2%. This case also produces LWP fields highly correlated with observed LWP fields with a correlation coefficient of 0.47, while all other four cases perform relatively poorly.

Table 3.2 Ranking of model cases in terms of simulating LWP

Ranking	Normalized mean bias (NMB)	Ranking by correlation
1st	YSU_MORR_NAM (-18.2%)	3rd (0.47)
2nd	YSU_THOM_ECW (-19.1%)	26th (0.37)
3rd	YSU_THOM_NAM (-19.2%)	10th (0.44)
4th	YSU_MORR_FNL (-19.3%)	20th (0.42)
5th	YSU_THOM_FNL (-20.5%)	22nd (0.41)

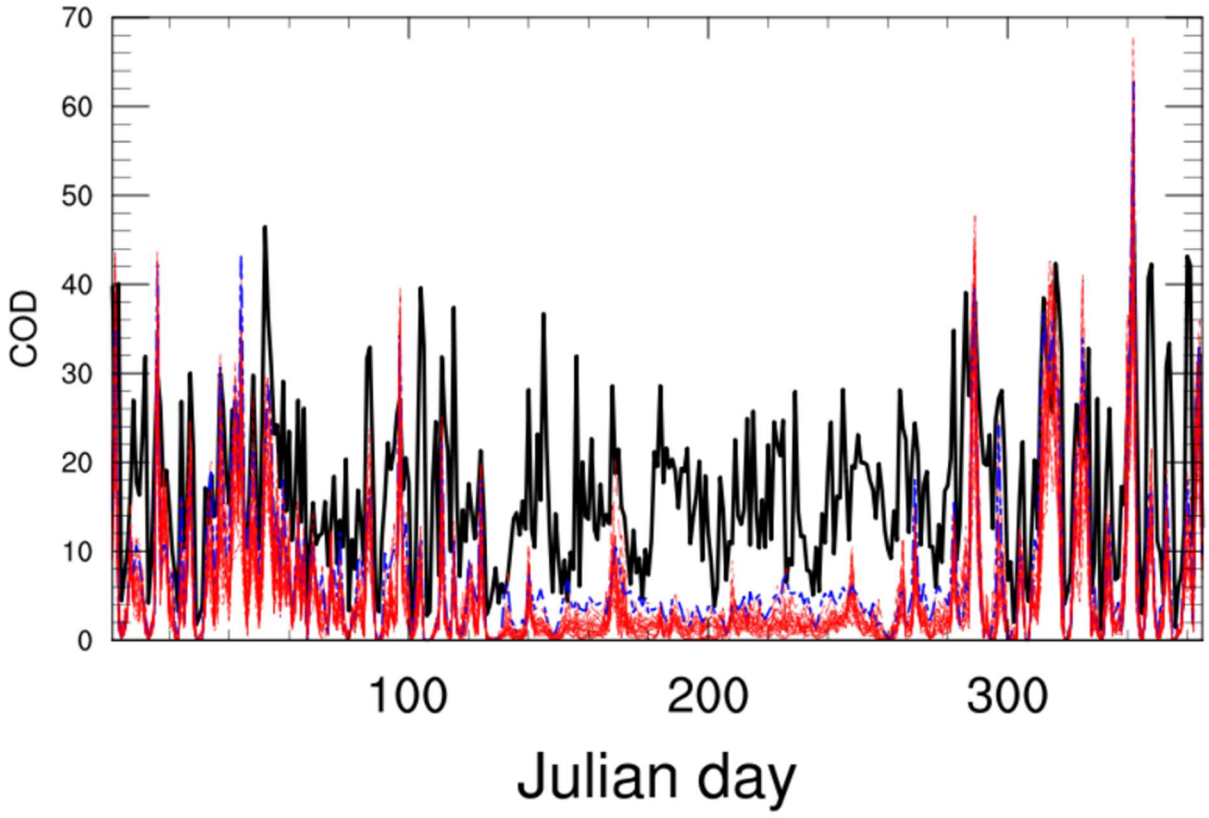


Figure 3.6 Daily cloud optical depth (COD) of year 2018. Black line represents daily MODIS retrieved cloud fraction. Blue line represents WRF simulation of YSU_MORR_NAM case, while red lines represent all other WRF simulation cases.

Figure 3.5 shows the time series of domain-averaged cloud liquid water path (LWP) as observed by MODIS Aqua and modeled by 27 cases for local afternoon (15 LST) of 2018. The comparison between modeled and observed COD is very similar to the LWP comparison (too few too bright) as shown in Figure 3.6. It should be noted that the YSU_MORR_NAM case does not place in the first place, but the eighth place in the COD simulation.

To sum up, we select YSU_MORR_NAM case as the optimal configuration. We conduct long term simulation of WRF using this configuration from year 2005 to year 2020. Unfortunately, the NAM reanalysis input is not available or corrupted for years 2007 and 2011. (After contacting the NOAA scientist, we are informed that these data can not be updated or replaced.) We will use the rest simulations for training GAN.

3.4 WRF simulation and COSP Deliverables

We will archive the WRF simulation outputs from multiple case simulations, and long-term simulations in netcdf format. We will also archive the COSP inputs/outputs for both multiple case simulations and long-term simulations. We will also archive the code used for COSP calculations. Considering the large file size, we only save the WRF outputs for inner domain.

4. Conduct the GAN training

4.1 GAN statistics

We use totally 4728 daily modeled and observed cloud field pairs from year 2005 to 2019 (excluding year 2005 and 2007) to train GAN; and use 353 cloud field pairs from year 2020 to evaluate our training (some pairs are removed because of poor MODIS retrieval quality). The training is conducted for 200 epochs with initial learning rate of 0.002.

4.2 GAN fed with multiple cloud fields

In this section, we firstly examine the examples of GAN results in comparison to WRF/COSP cloud fields and MODIS retrievals, then followed by the statistical analysis.

Figure 4.1 shows the WRF/COSP modeled, GAN outputted, and MODIS observed cloud fields for Jan 10, 2020. In the WRF/COSP modeled cloud field, we see pixels with blueish pixels, which most likely corresponds to ice-phase clouds; pixels with cyan color, indicating model predict relatively higher COD than LWP compared to MODIS observation. In MODIS observed cloud filed, we see both cyan and magenta colors. For those pixels with magenta color, MODIS reports relatively higher LWP values than COD values. In Figure 4.1, we find that GAN outputted cloud field agrees well with the MODIS observation. Firstly, GAN CF is increased to almost overcasting after performing GAN, which agrees well with the MODIS observation. Secondly, the WRF model certainly pick up signal of the frontal system that affects Texas; however, the position, intensity, and microphysical properties of modeled frontal system do not agree well with the observation. After performing GAN, the GAN frontal system agrees much better with observation. Finally, the GAN can produce small feathery details associated with the frontal system that can not be directly modeled.

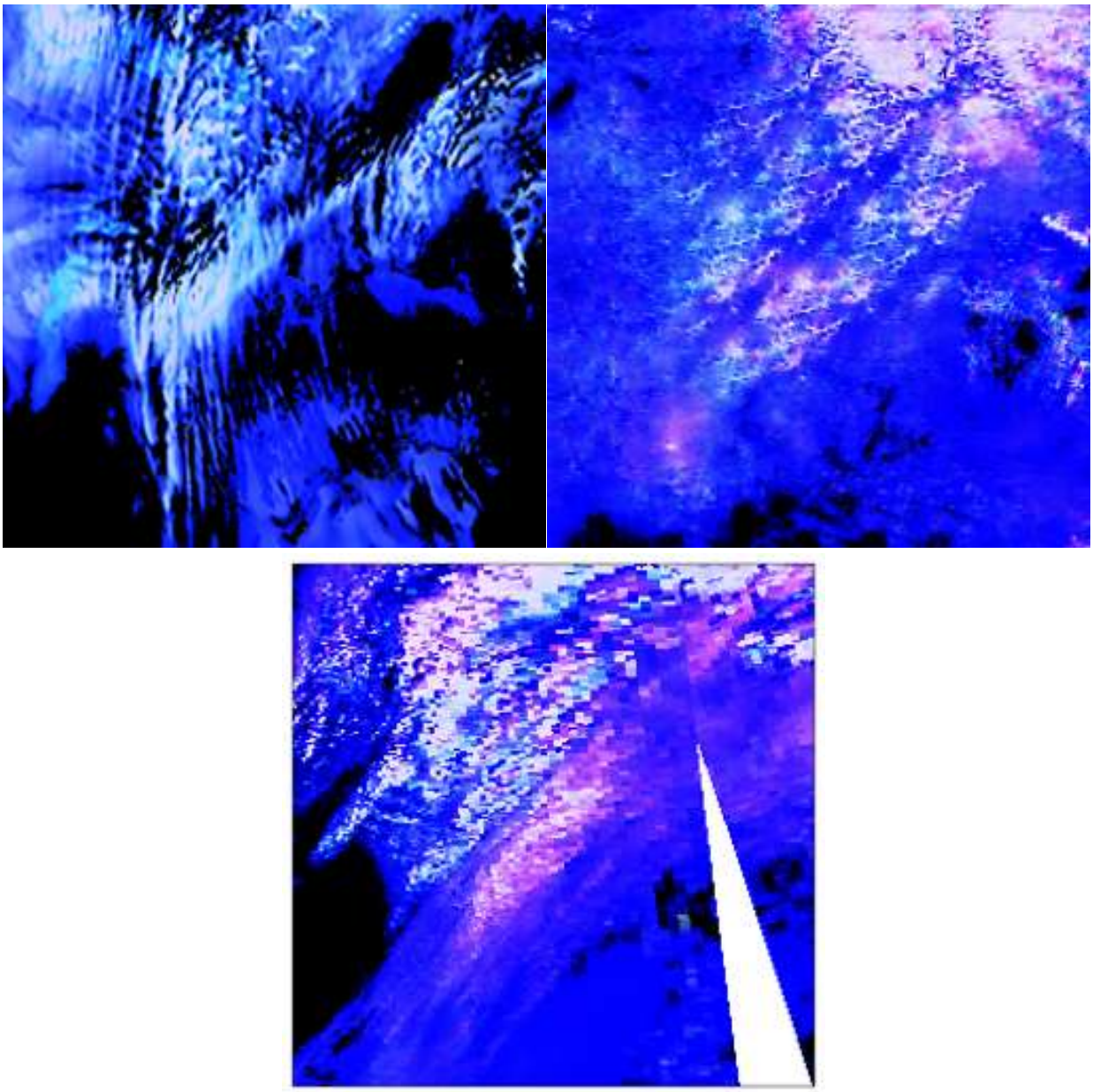


Figure 4.1 Cloud fields of Jan 10, 2020. Top-left: WRF/COSP simulated cloud fields (three-channel composed fields of CF[B], LWP[R], and COD(G)). Top-right: GAN outputs. Bottom: MODIS retrievals.

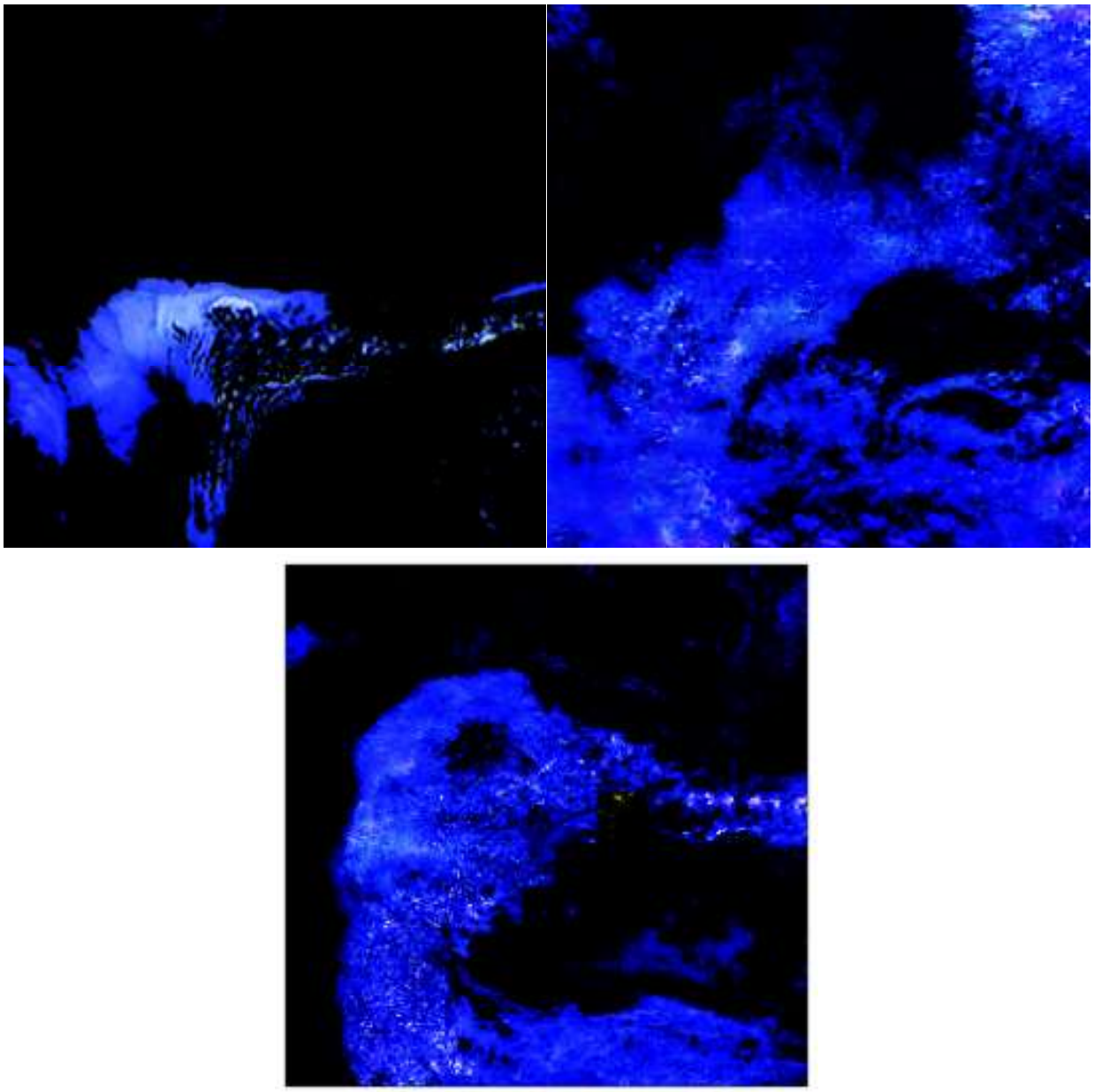


Figure 4.2 Same as Figure 4.1, but for Feb. 7, 2020.

Figure 4.2 shows the cloud fields for Feb. 7, 2020. Compared to the MODIS observation, WRF/COSP modeled CF is significantly underestimated. The domain is mainly filled with thin ice-phase clouds because of most pixels are in blue color. GAN successfully generates cloud fields that are very close to observation, without increasing LWP and COD fields. The location of breaking of thin cloud deck is also successfully predicted by GAN. GAN also reproduces fine features that are associated with observed cloud field, while in WRF/COSP simulation, these are completely missing.

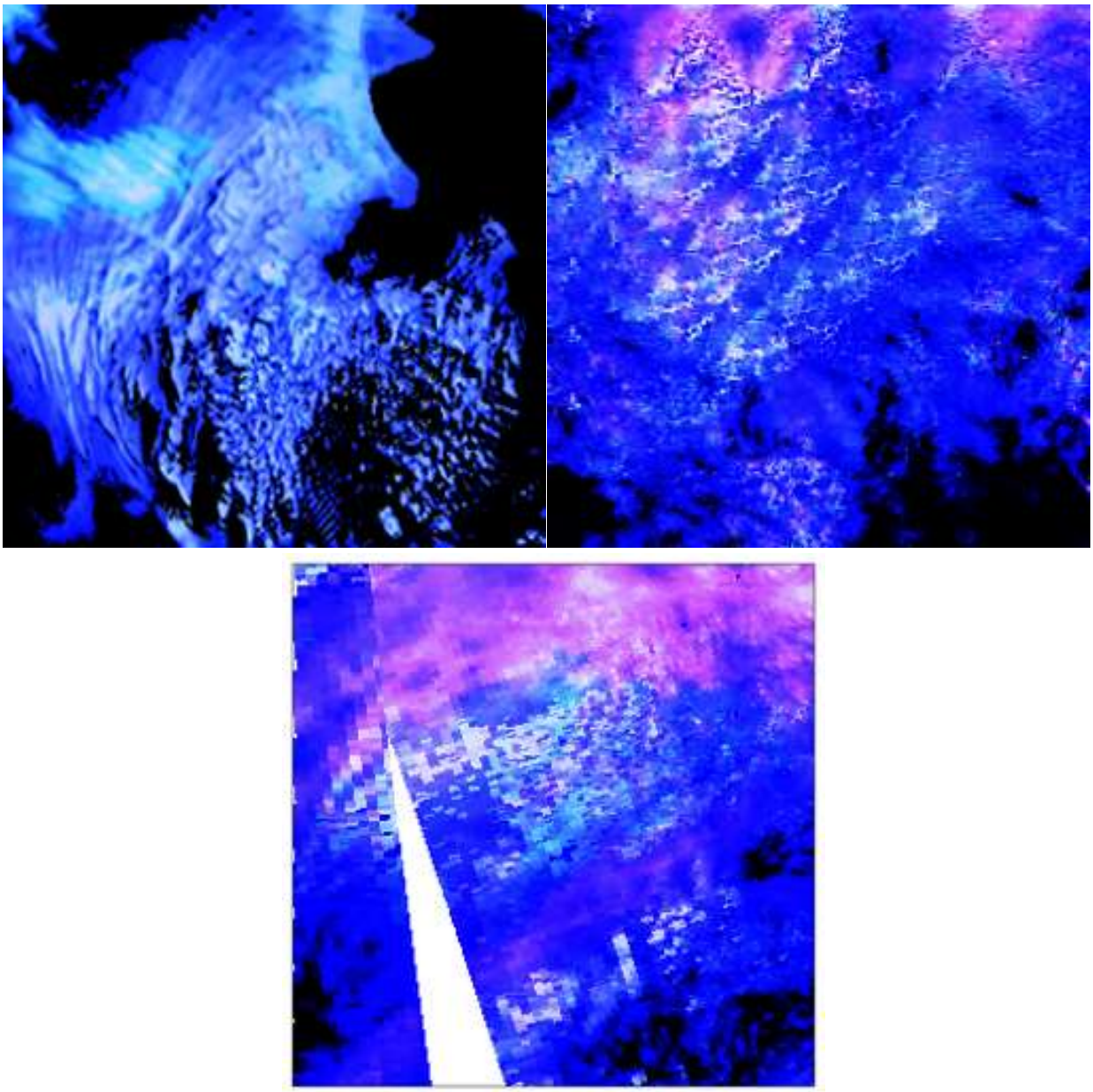


Figure 4.3 Same as Figure 4.1, but for Feb. 23, 2020.

Figure 4.3 shows the cloud fields for Feb. 23, 2020. As shown in the WRF/COSP modeled cloud fields, large aera is in cyan color, indicating model tends to relatively over predicts COD field. This overestimation is very like due the fact that model predicts smaller cloud droplets than realism. Such cyan pixels are removed after applying GAN. In addition, GAN adds more magenta pixels to match the observations. The WRF/COSP predicts too few but too bright clouds over the Gulf of Mexico. GAN transfers those pixels into color magenta. The location of high LWP and associated magnitude agree well with the observation. The most impressive of improvement of using GAN is that it generates observed feathery-like fine features.

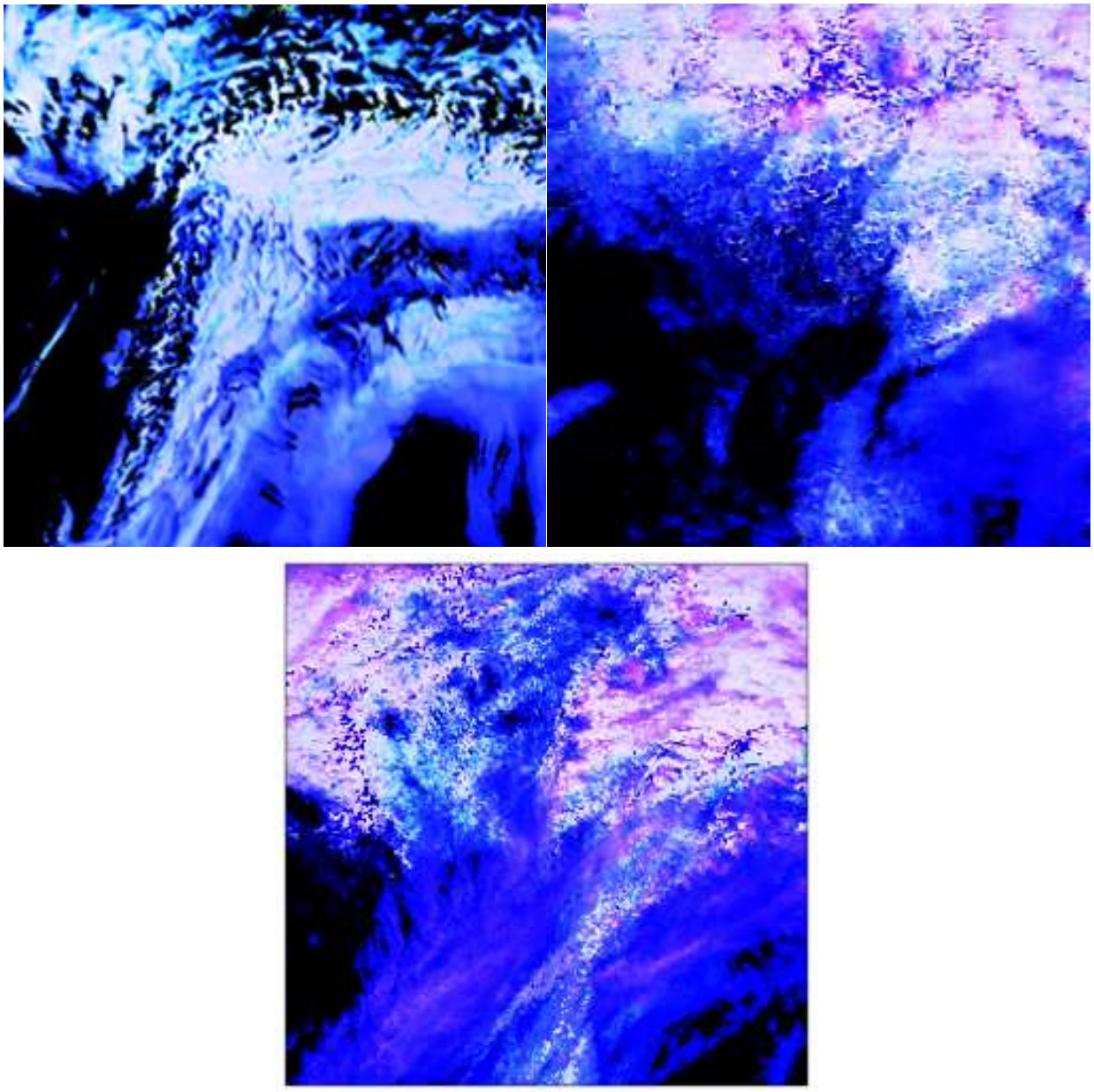


Figure 4.3 Same as Figure 4.1, but for Mar. 4, 2020.

For Mar. 4, 2020, the WRF/COSP predicts a very strong frontal system that affects the Texas as shown in Figure 4.4. Many grid boxes/pixels appear color of cyan and white, indicating heavy cloud water that is contained by this frontal system. GAN again reduces the COD (green color) and increases the LWP (red color) so that many pixels appear magenta color. GAN does capture the frontal system feature; however, the stretch of the frontal system is as long as observed and WRF/COSP modeled frontal system.

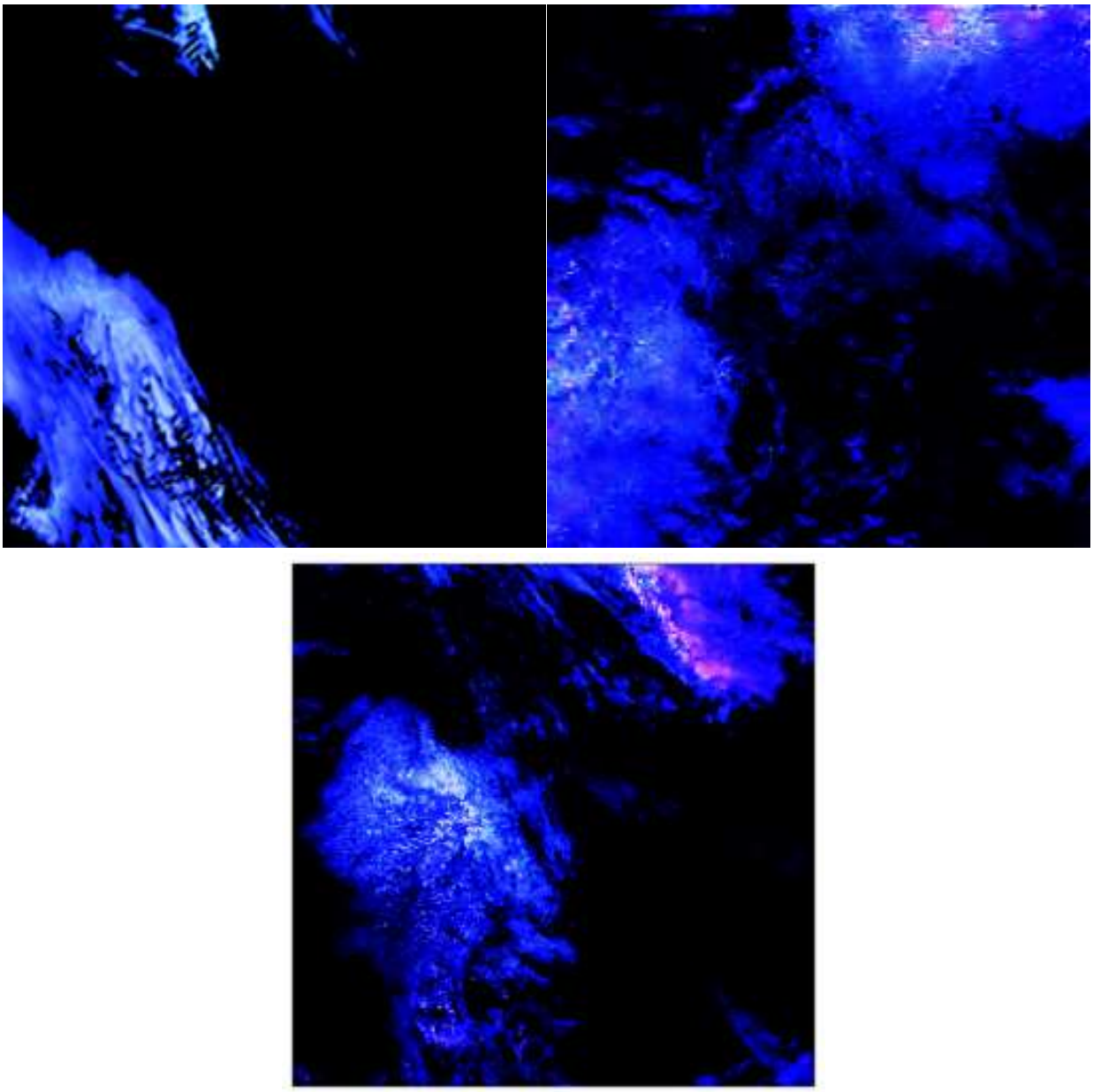


Figure 4.4 Same as Figure 4.1, but for May 7, 2020.

The performance of GAN is very good again for May 7, 2020. The WRF/COSP predicts one cloud deck and the tip of a frontal system in the domain. However, as seen in the MODIS observation, there should be two large cloud decks – one associated with the frontal system, and one cloud deck behind the frontal system. GAN transform WRF/COSP modeled cloud field to the one very close to the observation. The high LWP (e.g. magenta color) in the center of the frontal system, the spatial coverage of the frontal system, as well as the location of frontal system are all very reasonably reproduced. In the other cloud deck, GAN generates the center of the cloud deck with high COD but relatively low LWP. In addition, the GAN also generates small-scale cumulus

between the cloud decks. This is very impressive, since the WRF model predicts zero cloudiness in this region.

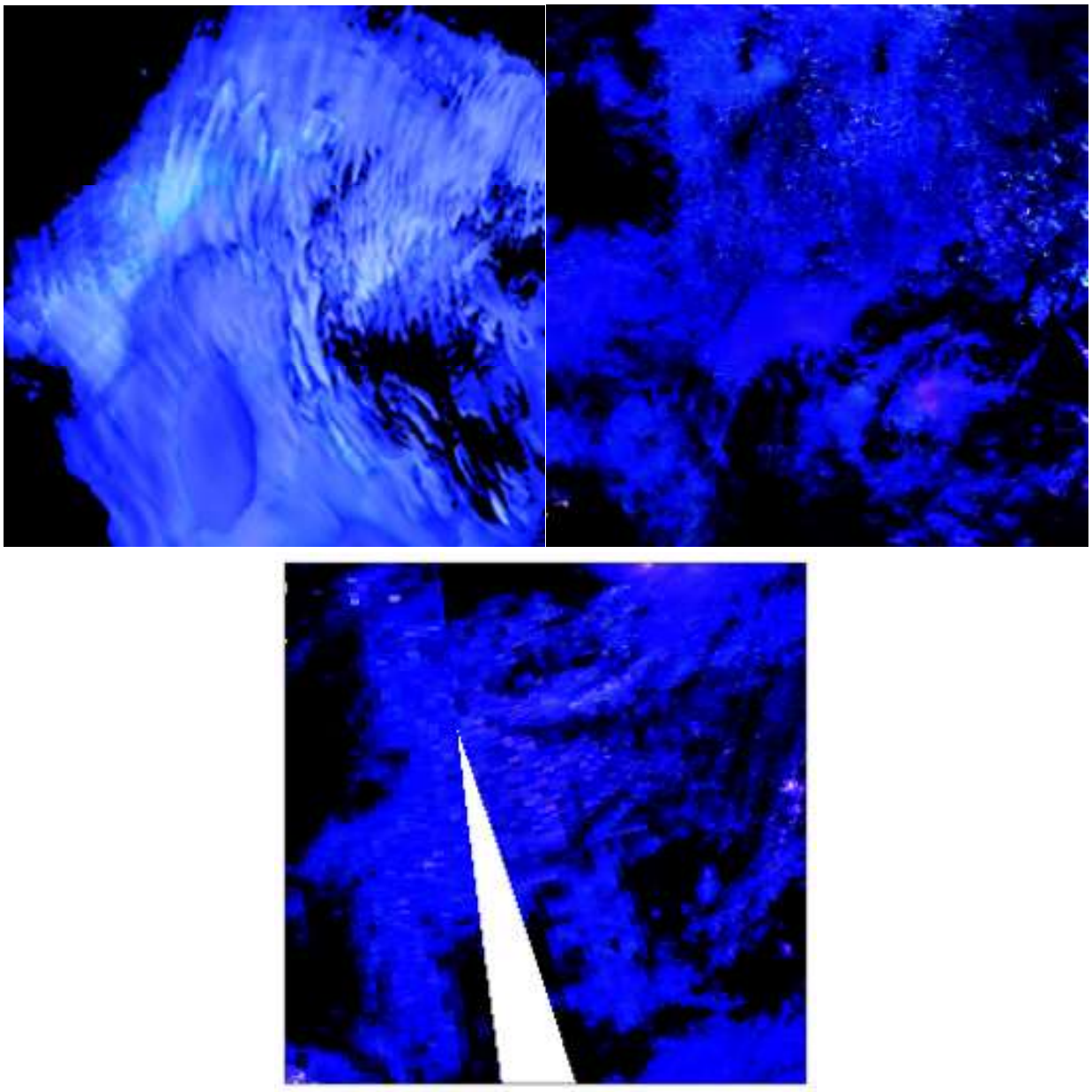


Figure 4.5 Same as Figure 4.1, but for May 22, 2020.

The modeled CF on May 22, 2020 is not underestimated as usual as shown in Figure 4.5. On contrary, model predicts higher CF compared to the MODIS observation because of lack of gaps in the cloud deck. GAN successfully reduces the COD of the cloud deck and produces enough gaps/breakings in the cloud deck.

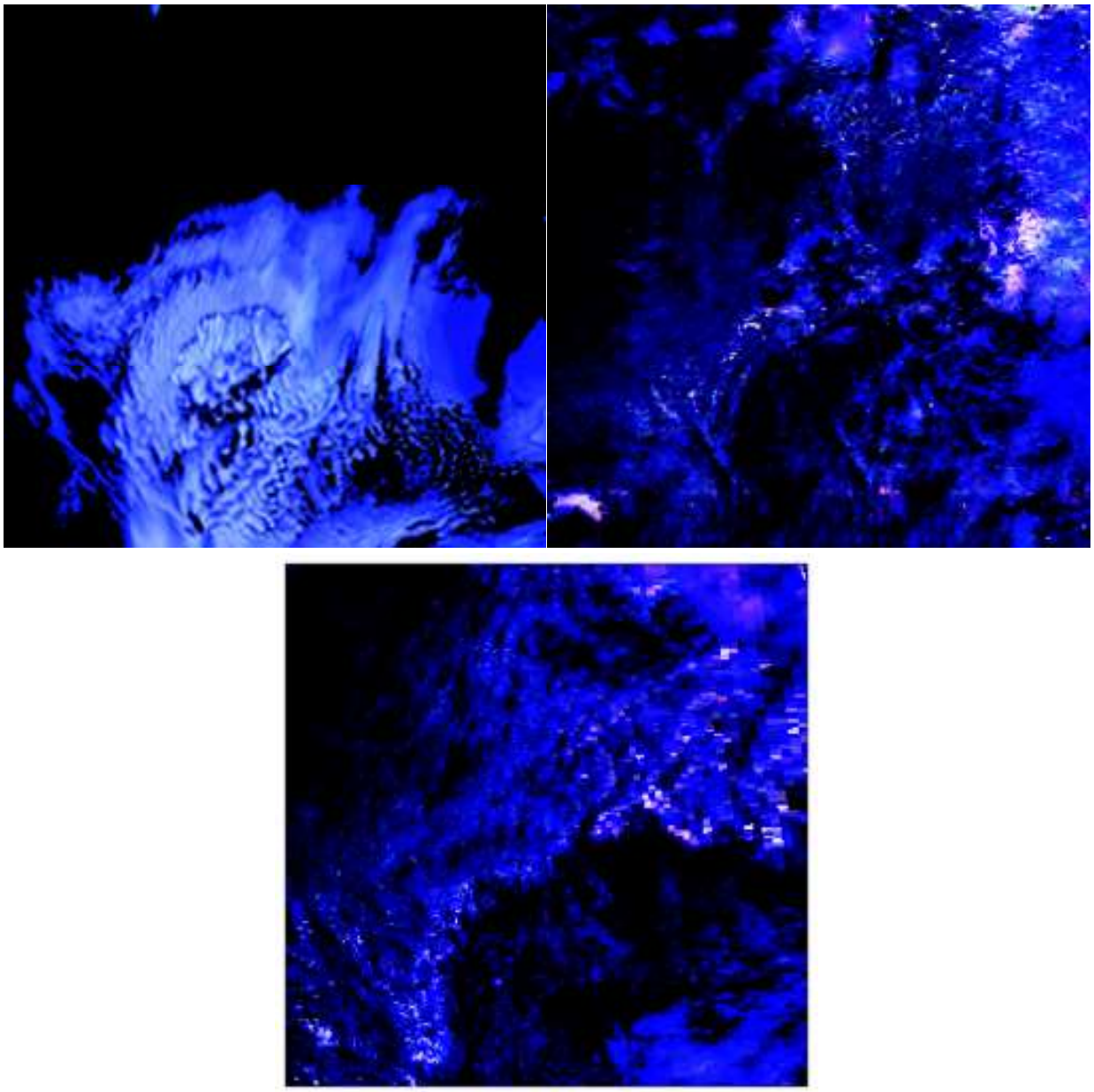


Figure 4.6 Same as Figure 4.1, but for June 4, 2020.

During the summer, WRF model usually predicts small clusters of cloud cells over the Gulf of Mexico. This feature can be interpreted as marine stratocumulus. Figure 4.6 shows such example on June 4, 2020. By examining the MODIS observation, we find that, during this day, the (stratocumulus) cloud decks mostly break into cumulus over the ocean. While over the land, we see significant amount of cloud cells, some of which has high LWP and COD values. All these features are fully captured by the cloud fields generated by GAN as shown in Figure 4.6.

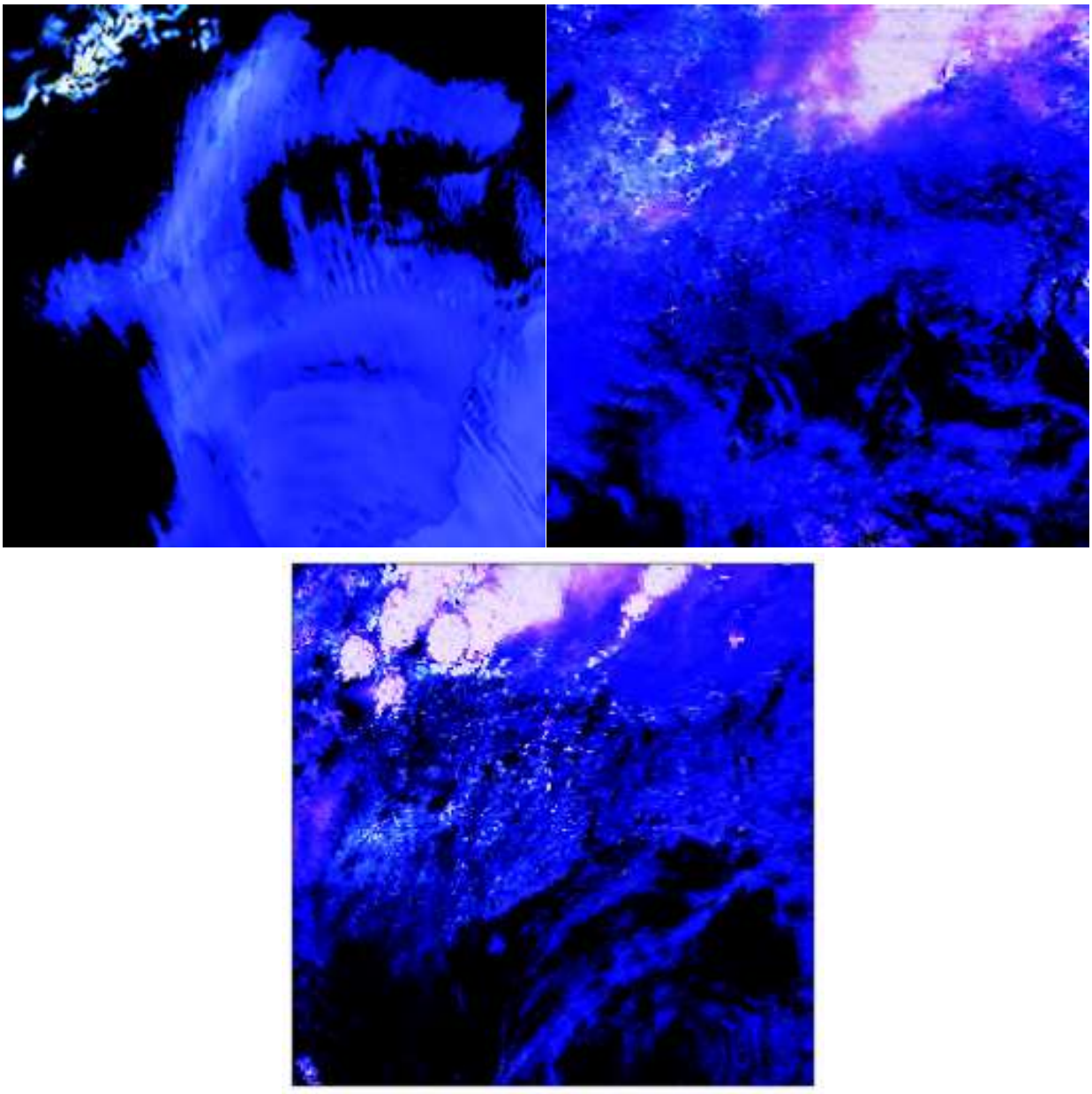


Figure 4.7 Same as Figure 4.1, but for Sept. 1, 2020.

One common issue of WRF simulating frontal system is that the horizontal extent (width) is usually underestimated as shown in an example of Sept. 1, 2020 in Figure 4.7. GAN transforms the modeled frontal system with larger width, adjusts the location of the frontal system, and reduces the biased high COD (by removing green color). In the domain, the cloud decks/cells with gaps and breakings over the Gulf of Mexico are also well captured by GAN.

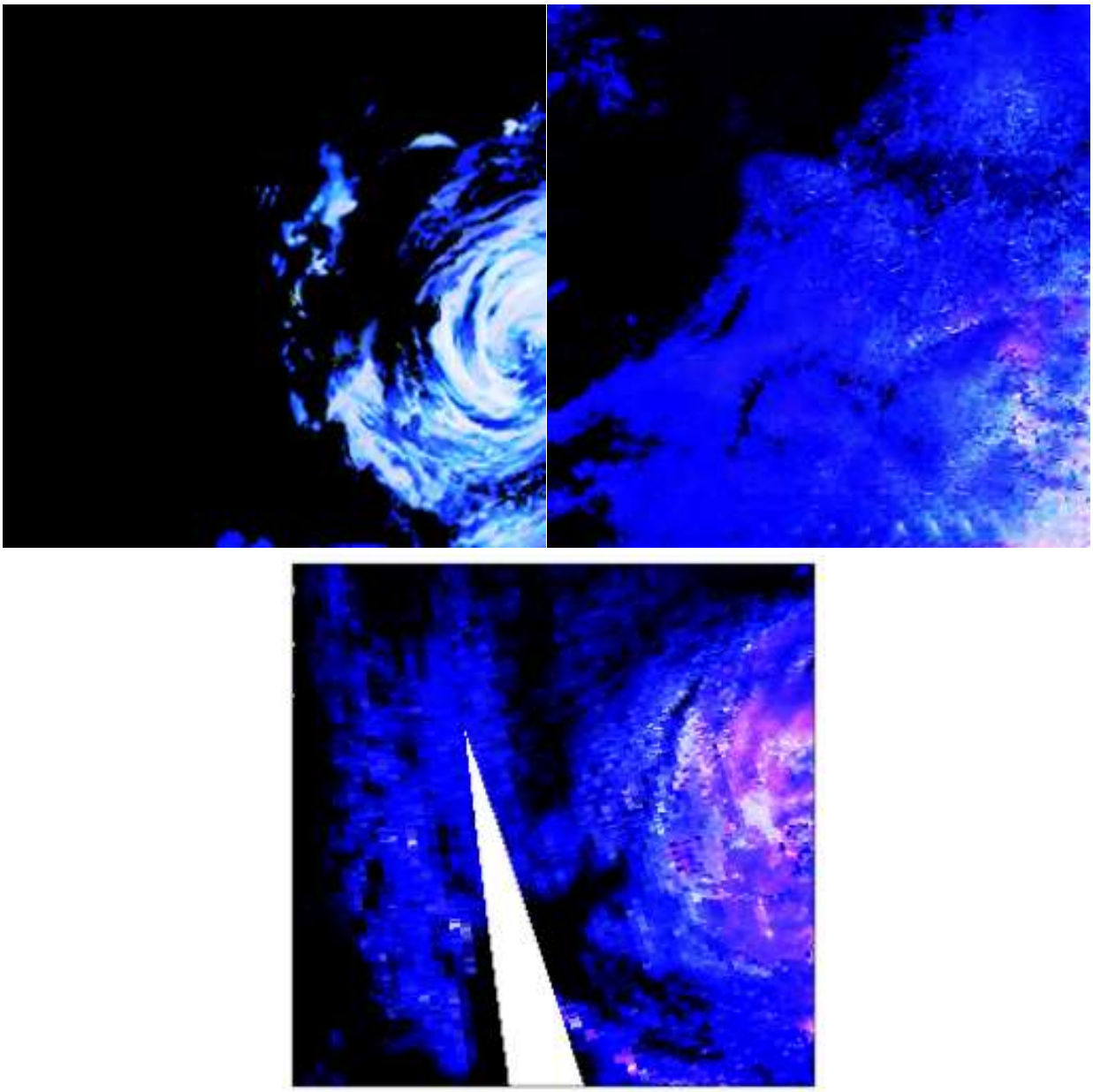


Figure 4.8 Same as Figure 4.1, but for June 7, 2020.

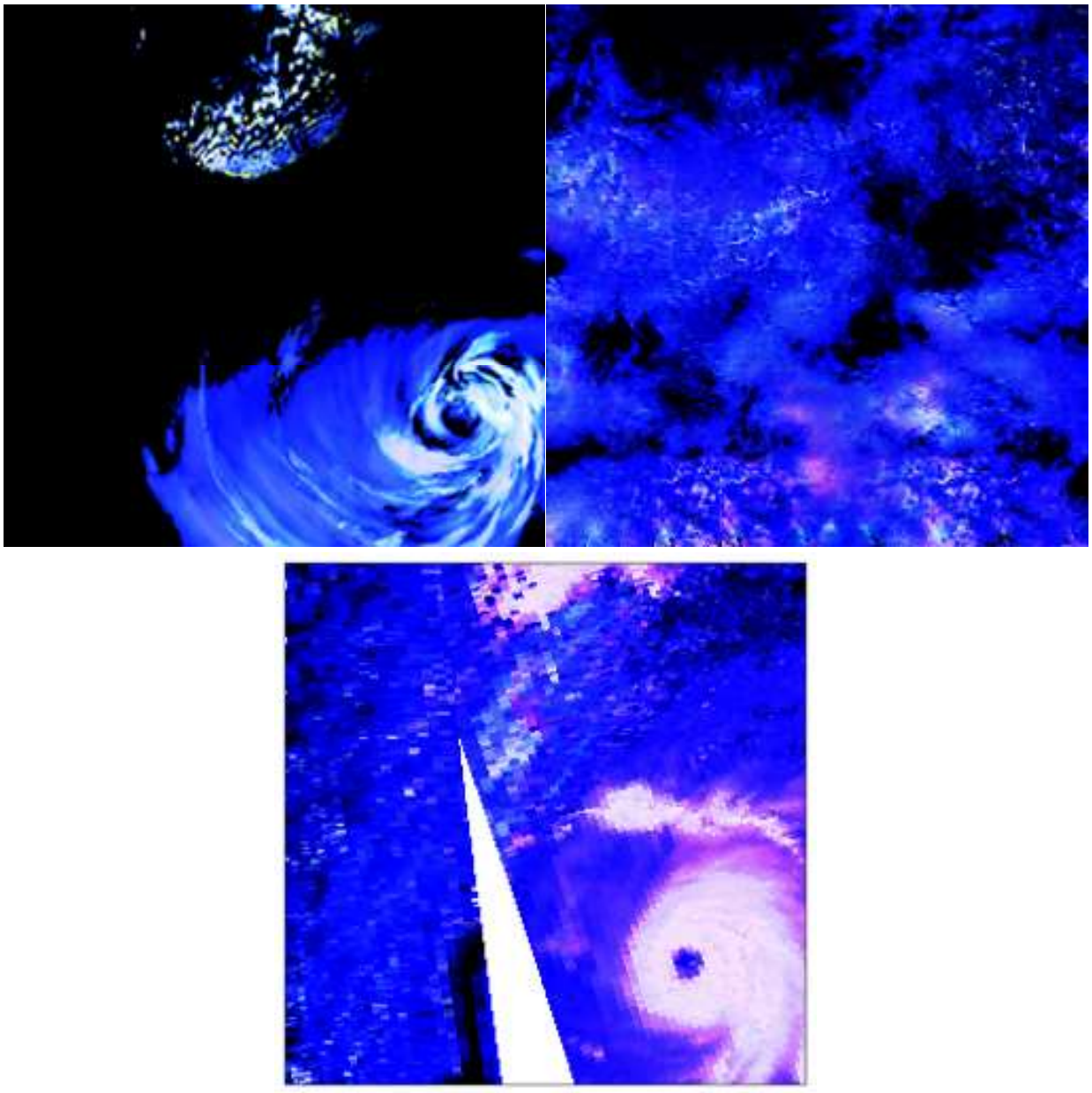


Figure 4.9 Same as Figure 4.1, but for Aug. 26, 2020.

There are also some examples of the "poor" performance of GAN if we examine results day-by-day. In Figure 4.8 and 4.9, we show the cloud fields associated with two hurricanes that hit Texas during the summer of 2020. The performance of model is impressive in terms of capturing locations and intensity of hurricanes. Although model again overestimates the COD associated with clouds (i.e. cyan colors in modeled cloud fields vs. magenta colors in observed cloud fields). Model also fails to reproduce overcasting cloudiness surrounding the hurricanes. GAN does produce such overcasting cloud feature; however, no hurricane feature is preserved after applying GAN. This is not surprising because the performance of GAN greatly depends on the sample. The

times that hurricanes hit Texas each year are only handful. With such small training samples, the failure of GAN in reproducing hurricane feature is guaranteed.

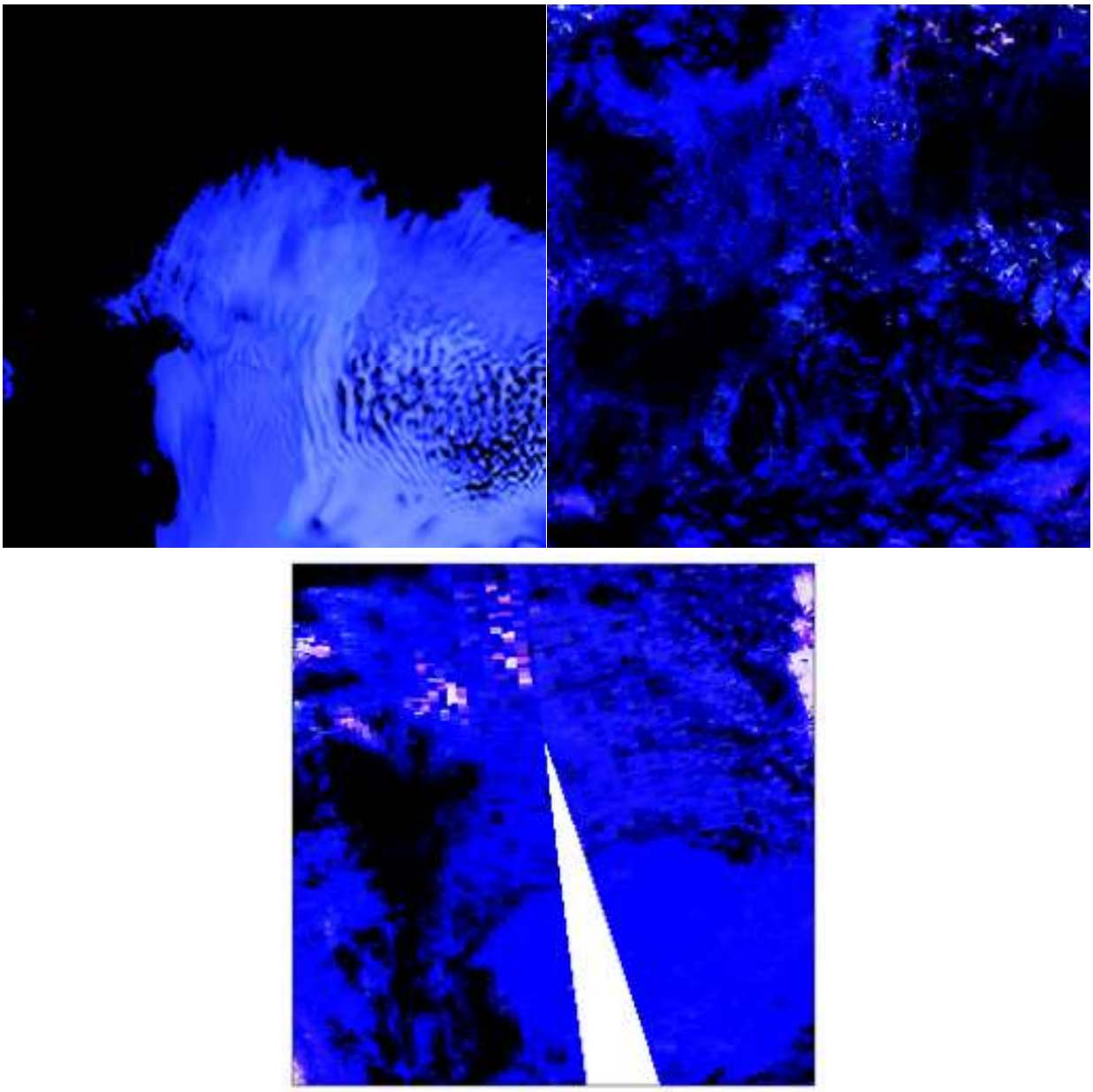


Figure 4.10 Same as Figure 4.1, but for July 3, 2020.

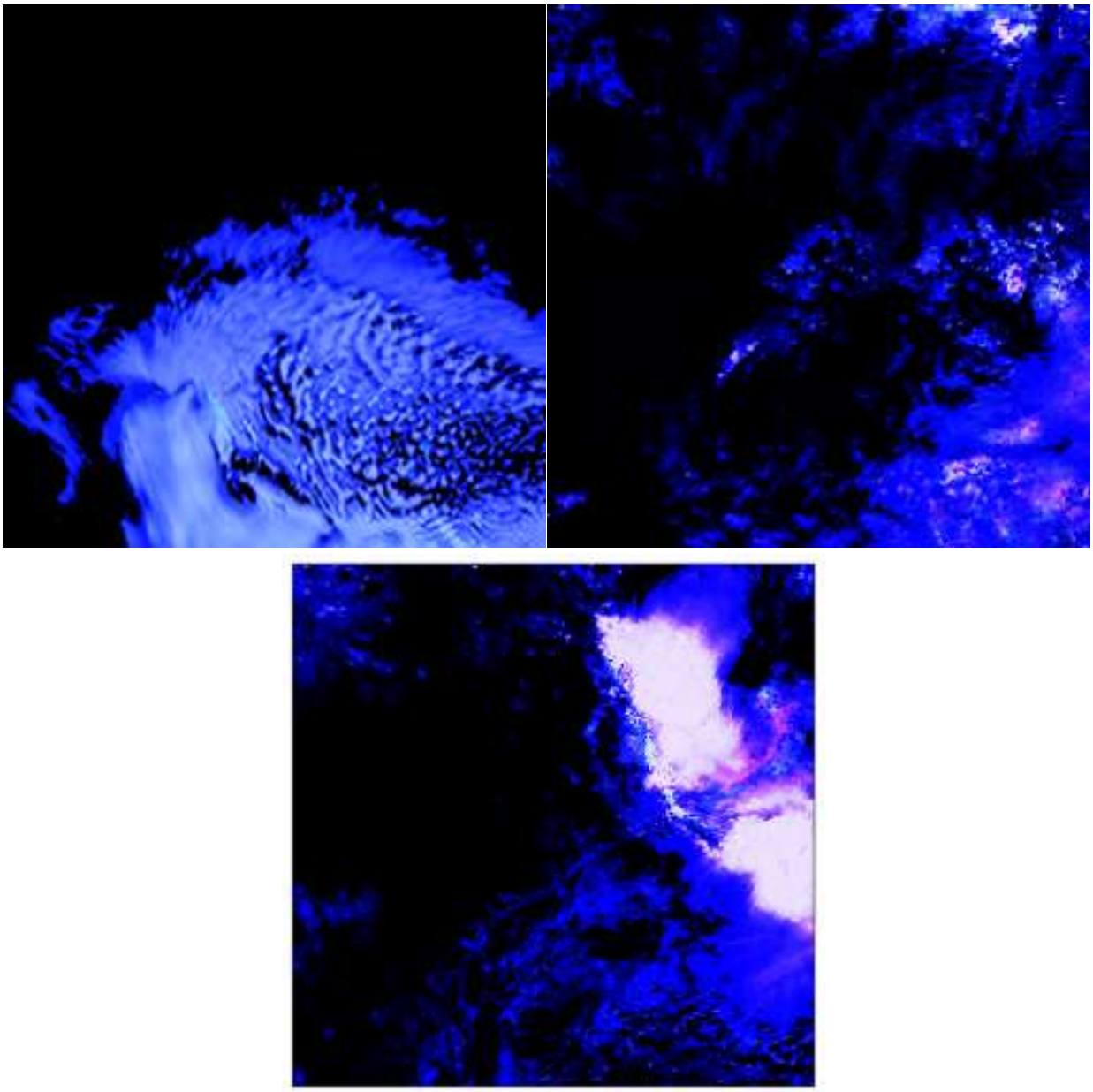


Figure 4.11 Same as Figure 4.1, but for July 4, 2020.

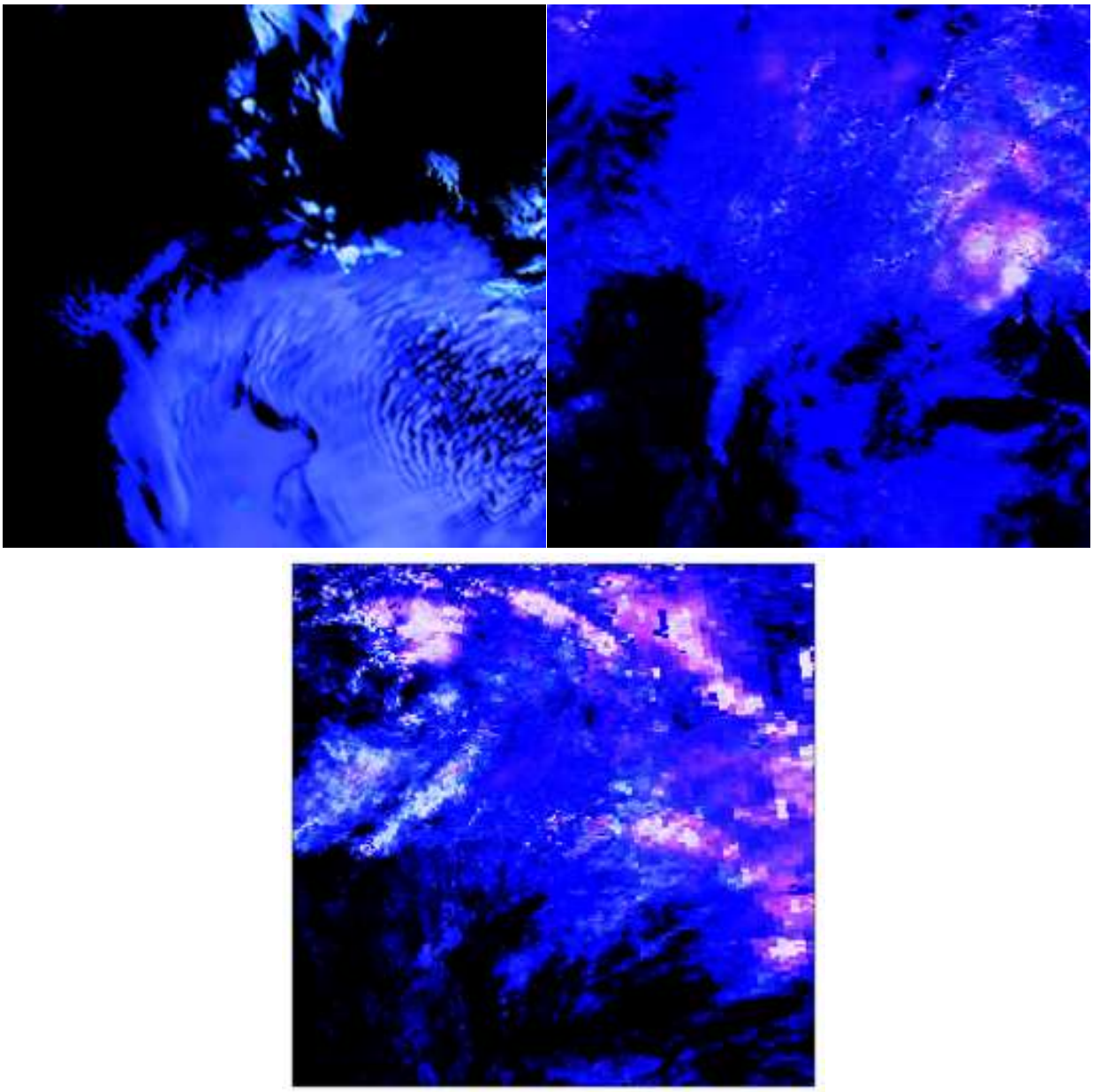


Figure 4.12 Same as Figure 4.1, but for July 6, 2020.

Another big challenge for GAN is associated with localized deep convection. Shown in Figures 4.10, 4.11, and 4.12 are cloud fields during July, for which WRF model predicts very similar cloud fields (stratocumulus over the Gulf of Mexico, cloud decks stretching from coast to some part of inland) in these three days, especially on July 3 and July 4, 2020. However, according to MODIS, the realistic cloud fields are very different in these three days. On July 4, 2020, very strong localized deep convection system was already in its mature stage as observed by the MODIS. GAN fails to pick up this deep convection system at all. For July 6, however, the model does predict some features associated with localized deep convection. GAN can generate a similar feature compared to the MODIS observation, although the number of such features is still

underestimated. This indicates that GAN's performance still significantly depends on WRF's performance. Once model simulation can catch some features that related to deep convection, GAN can pick up the signal and reproduce these features very well. The examples shown in Figures 4.8-4.12 highlights the importance of sample number.

We extract CF, LWP, and COD values from the three channels of false-color image of GAN outputs, and examine the performance of GAN by conducting statistical analysis. In Table 4.1 and Table 4.2, we compare both the WRF/COSP cloud fields and the GAN cloud fields to the MODIS cloud fields and evaluate the performance using the metrics of averaged daily NMB and daily correlation coefficient. GAN significantly improves the CF simulation – compared to WRF/COSP outputs, the magnitude of NMB reduces from -37% to only -0.05%. The correlation coefficient is significantly enhanced from 0.49 to 0.69 (n = 353, several days' results are removed from the training set because of poor retrieval quality of these MODIS retrievals). We find that the NMB between the LWP field in GAN outputs and observations becomes a little worse – drops from -40% to -45%. Given the fact that we are examining domain-averaged value instead of in-cloud value and the CF simulation is much improved by GAN, this decrease in NMB indicates LWP are more reasonably distributed in cloudiness. NMB associated with COD changes in sign – from overestimation to underestimation. By performing GAN, we can solve the famous "too few too bright" issue. Both correlation coefficients of LWP and COD are much improved after applying GAN.

Table 4.1 Averaged daily NMB between WRF/COSP simulations and the MODIS observations as well as between GAN outputs and the MODIS observations.

NMB	WRF/COSP	GAN
CF	-37%	-0.05%
LWP	-40%	-45%
COD	16%	-35%

Table 4.2 Averaged correlation coefficient between WRF/COSP simulations and the MODIS observations as well as between GAN outputs and the MODIS observations.

Correlation Coefficient	COSP	GAN
CF	0.49	0.69
LWP	0.16	0.20
COD	0.17	0.20

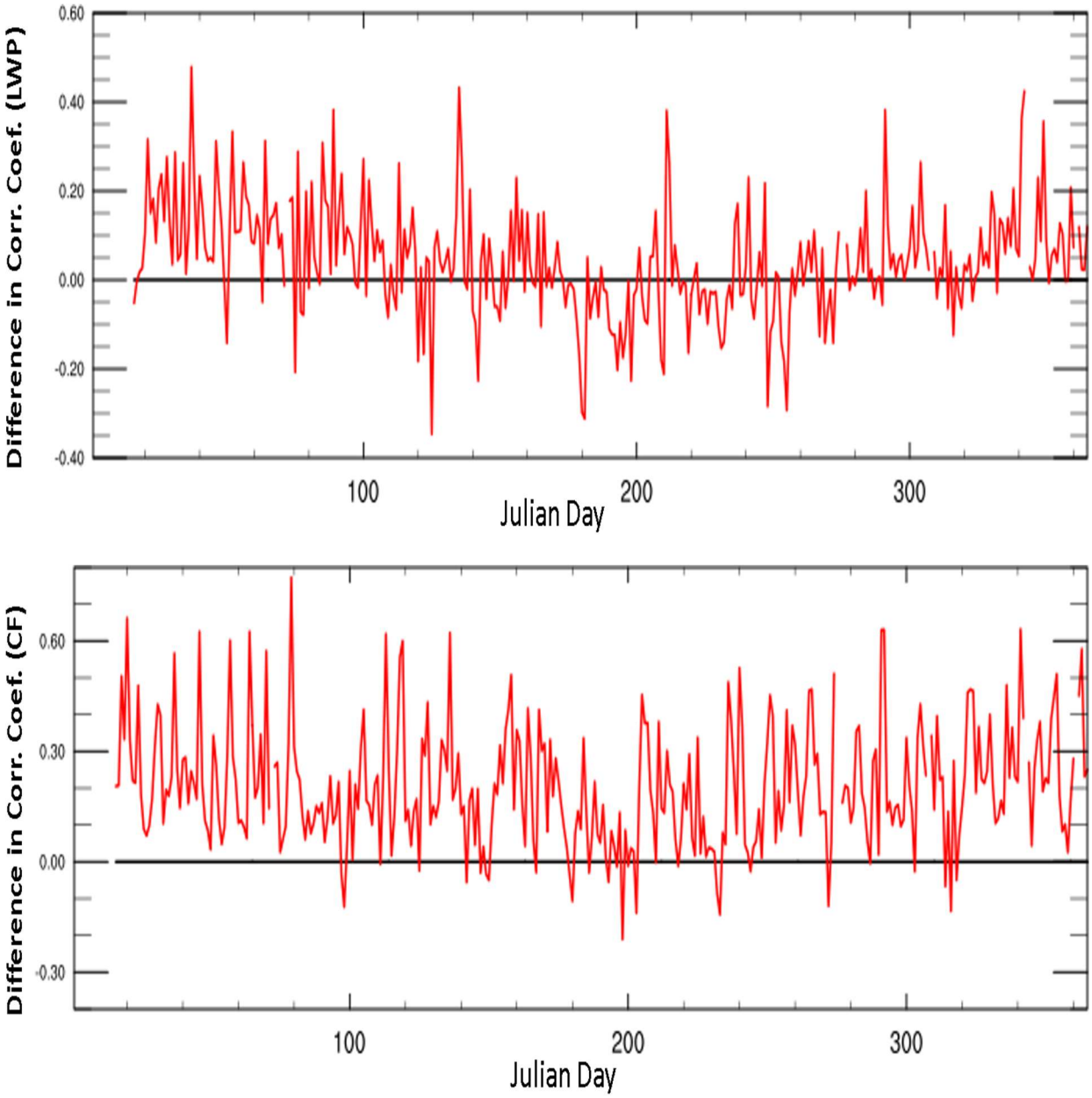


Figure 4.13 Time series of the differences in spatial correlation coefficient between GAN-MODIS pair and WRF/COSP-MODIS pair. Top: for LWP field. Bottom: for CF field.

In Figure 4.13, we show the time series of the difference in daily spatial correlation coefficients between GAN-MODIS pair and WRF/COSP-MODIS pair. For LWP, the spatial correlation coefficients are mostly improved by 0.1 during spring and winter. No such improvement can be seen for summer. The spatial correlation coefficients for CF are almost improved in all days after applying GAN. The increases in the correlation coefficients are about 0.2 by average.

4.3 GAN fed with one cloud field

Here we introduce a study that is not originally proposed: we feed GAN with only one cloud field LWP and examine if the cloud field simulation can be improved. Here we conduct the training of LWP with different level of color red (0-255) for years 2005-2019, and test GAN in simulating LWP fields of 2020. We find that GAN fed with LWP only improves the simulation of texture of cloud fields, but the magnitudes (NMB) and correlation coefficients of LWP fields are not improved or somehow become even worse. This indicates that GAN is less sensitive to one channel. We plan to conduct additional training with different hyperparameters, like increasing epoch number and learning rate and see if this could improve the performance of GAN.

Nevertheless, for some certain days, the performance of GAN is reasonable or better than GAN fed by three cloud fields. Here we provide some examples. Figure 4.14 shows the LWP fields of July 6, 2020. Compared to GAN fed with three cloud fields shown in Figure 4.12, GAN fed with LWP only predicts better localized deep convection signals, in terms of the sizes of cloud cells and magnitude of LWP values.

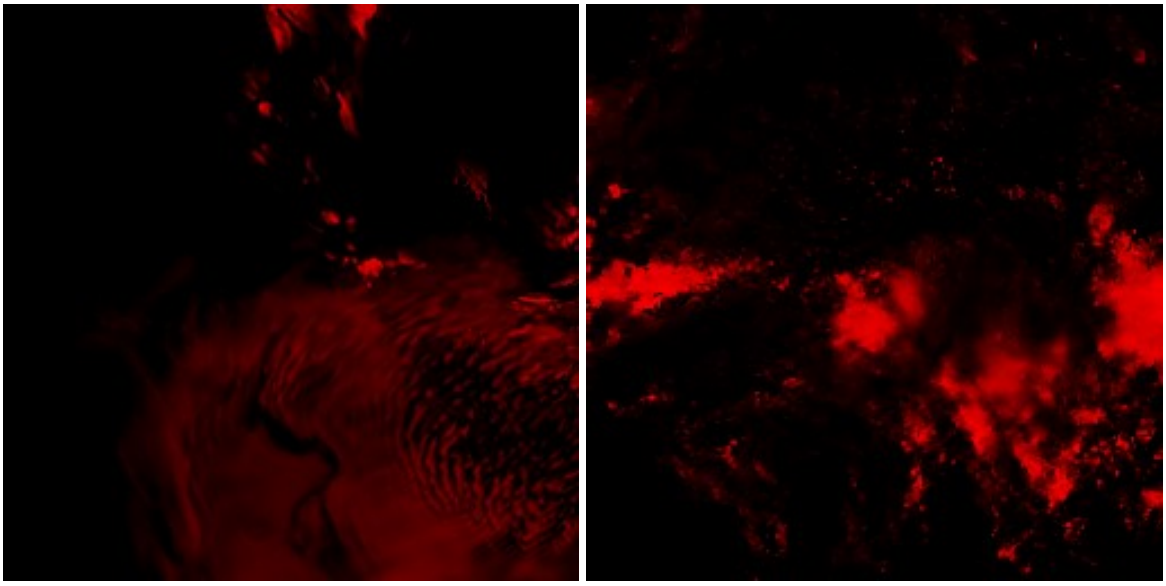


Figure 4.14 Cloud field of July 6, 2020. Top-left: WRF/COSP simulated cloud field of LWP Top-right: GAN outputs. Bottom: MODIS retrievals.

4.4 GAN deliverables

We will archive GAN code and array that contains weights in digital form. We will archive all inputs (figures of cloud fields of WRF/COSP and of MODIS) and outputs (figures of cloud fields generated by GAN). We will also archive the netcdf files containing cloud fields that are extracted from the figures of GAN outputs.

5. Audit Data Quality

5.1 Coding

The WRF code that used for simulation is directly downloaded from NCAR website (https://www2.mmm.ucar.edu/wrf/users/) without any code changes. The COSP data that we used is directly downloaded from Github website (https://github.com/CFMIP/COSPv2.0) without any code changes. The GAN code is also directly downloaded from Github website (https://phillipi.github.io/pix2pix/) with only minor changes (e.g. epoch number and figure input dimensions). To calculate the statistics, we use the default functions/routines in the NCAR command language (NCL, https://www.ncl.ucar.edu/).

5.2 Input and output datasets

The reanalysis input data are all downloaded from the default websites. Zheng Lu double-checked the quality of all input datasets and examine whether data has any corruption or inconsistence. For example, Zheng Lu finds that many data of 2007 NAM reanalysis are missing on the server, and 2011 NAM reanalysis data are corrupted with wrong vertical levels. Zheng Lu also double checked the quality of more than 10% of WRF outputs, focusing on examining whether the thermodynamic and cloud fields look reasonable. Zheng Lu also examine more than 10% of COSP outputs, focusing on checking whether any results have unreasonable cloud property outputs. Kai Lyu prepared all the inputs of false color images of cloud fields for GAN training and test. Zheng Lu and Kai Lyu examined all input figures by naked eyes since GAN training heavily depends on the quality of inputs.

6. Conclusion and Future Plan

The goal of this study is to improve the WRF modeled cloud fields using DL tool GAN. We firstly conduct WRF simulation with different combination of microphysical schemes, PBL schemes, and reanalysis inputs. Total 27 cases are conducted for year 2018. The WRF outputs are further processed through the COSP calculation. WRF/COSP modeled cloud fields are thoroughly compared against the MODIS observations. By ranking the performances of each cases, we select the optimal configuration and further conducts long-term simulations (2005-2020). The daily modeled cloud fields and daily MODIS observed cloud fields are both composed as false-color images, which are further inputted into GAN training (data from years 2005-2019) and evaluation (data from year 2020). We finally examine the performance of GAN by analyzing daily cloud fields and calculating the statistics of GAN-generated cloud fields.

The main findings on WRF simulations are summarized as following:

1). When we examine the standard deviation of modeled downwelling shortwave flux at surface (a proxy of cloud field), we find that the choice of the microphysics scheme accounts for the largest variation among different cases for almost all months of the year, except during June and July, when the choice of PBL scheme is more important.

2). WRF model significantly underestimates CF compared to observation, especially during summer, very like due to quick decoupling process between sea surface and stratocumulus-topped boundary layer. The performance of different cases varies from season to season. Usually, the cases with the Thompson scheme yield higher LWP, while the cases with WSM6 microphysics scheme predicts lowest LWP and CF. The cases with the Morrison scheme perform relatively better in summer.

3) Several modeling cases with the YSU PBL scheme performs better in simulating CF and LWP than other cases based on statistical metrics of NMB and correlation coefficient. The optimal case that we select is the case with the YSU PBL scheme, the Morrison microphysics scheme, and the NAM reanalysis. This case ranks the first place in terms of simulating CF and domain-averaged LWP based on NMB metric.

The main findings on GAN performance are summarized as following:

1) GAN can change the texture of modeled cloud fields by adding fine scale features, including gaps/breakings in cloud decks and feathery-like cloud cells.

2) GAN can improve the modeled cloud fields associated with the frontal system by adjusting the location and the width of the frontal system and reducing modeled COD.

3) WRF simulation often omits the overcasting thin cloud decks that accompany the frontal system, while GAN can often compensate proper thin clouds in the domain.

4) Based on the statistical analysis, we found that GAN can substantially improve the CF simulation – for example, the magnitudes of NMB between GAN cloud fields and MODIS observation is lower than 0.1%. In addition, the spatial correlation coefficient between GAN and MODIS increases by 0.2 compared to the original WRF/COSP cloud fields.

5) The LWP and COD simulation are somewhat improved, especially in terms of spatial correlation coefficient.

6) The performance of GAN for summer clouds is relatively poor because the WRF model sometimes fails to simulate localized deep convection systems that occur in the summer. GAN also fails to properly reproduce cloud fields associated with hurricanes because of very few of such examples in long-term simulations.

7) GAN fed with only one cloud variable (LWP) looks problematic at current stage but may be improved by tuning hyper-parameters.

The recommendations for future work are as following:

1) We plan to increase sample number of cloud fields by introducing geo-stationary satellite observations, like NOAA GOES-East. We know that the performance of DL technique heavily relies on sample number. For multiple years simulation, we can only obtain 4728 traininga samples, which is considered as a very small sample size in DL studies. By comparing against geo-stationary satellite observation, we can generate training datasets multiple times in a day. We believe this could substantially benefits the performance of GAN.

2) We plan to conduct GAN training again but replacing COD fields with IWP (ice water path) fields. Based on our analysis, we find that modeled COD is very biased compared to the MODIS observation, probably because of inaccurate assumption of cloud droplet number concentration/effective size in COSP. It is very desirable to properly simulate IWP, since it is very important for estimating radiative energy budget.

3)The most important application of our study is for cloud prediction purpose. Once the cloud fields are predicted, we can input those fields into GAN and get a more accurate picture of future cloud fields. We would like to ask how we can use GAN-generated cloud fields in aid of WRF model simulation. Here we propose a novel method for such purpose: we can firstly run WRF model and get first-guess of cloud fields, which is further fed into GAN and get

pseudo-observed cloud fields. Such cloud fields can be assimilated into the second round of WRF simulation (like White et al., [2018] but with pseudo-observed cloud fields).

7. References

Bodas-Salcedo, A., et al. (2011), COSP: Satellite simulation software for model assessment, Bull. Amer. Meteor. Soc., 92, 1023– 1043.

Bretherton, C. S., McCaa, J. R., & Grenier, H. (2004). A new parameterization for shallow cumulus convection and its application to marine subtropical cloud-topped boundary layers. Part I: Description and 1D results. *Monthly weather review*, *132*(4), 864-882.

Cohen, A. E., Cavallo, S. M., Coniglio, M. C., & Brooks, H. E. (2015). A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern US cold season severe weather environments. Weather and forecasting, 30(3), 591-612.

Collins, W. D., Rasch, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., ... & Dai, Y. (2004). Description of the NCAR community atmosphere model (CAM 3.0). *NCAR Tech. Note NCAR/TN-464+ STR*, *226*, 1326-1334.

Ebert, E. E., & Curry, J. A. (1992). A parameterization of ice cloud optical properties for climate models. *Journal of Geophysical Research: Atmospheres*, *97*(D4), 3831-3836.

Fan, S. M., Horowitz, L. W., Levy, H., & Moxim, W. J. (2004). Impact of air pollution on wet deposition of mineral dust aerosols. Geophysical research letters, 31(2).

Fan, J., Wang, Y., Rosenfeld, D., & Liu, X. (2016). Review of aerosol–cloud interactions: Mechanisms, significance, and challenges. Journal of the Atmospheric Sciences, 73(11), 4221-4252.

Feingold, G., & Kreidenweis, S. M. (2002). Cloud processing of aerosol as modeled by a large eddy simulation with coupled microphysics and aqueous chemistry. Journal of Geophysical Research: Atmospheres, 107(D23), AAC-6.

Gibson, J. K., Kallberg, P., Uppala, S., Hernandez, A., Nomura, A., & Serrano, E. (1997). ERA description. ECMWF Re-Analysis Project Report Series 1, ECMWF. Reading, UK, 77.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative Adversarial Networks. Retrieved from https://arxiv.org/abs/1406.2661v1

Grenier, H., & Bretherton, C. S. (2001). A moist PBL parameterization for large-scale models and its application to subtropical cloud-topped marine boundary layers. Monthly weather review, 129(3), 357-377.

Gurciullo, C. S., & Pandis, S. N. (1997). Effect of composition variations in cloud droplet populations on aqueous - phase chemistry. Journal of Geophysical Research: Atmospheres, 102(D8), 9375-9385.

Hong S,-Y, Y. Noh, and J. Dudhia (2006a). A new vertical diffusion package with an explicit treatment of entrainment processes. Mon. Wea. Rev., 134, 2318–2341, https://doi.org/10.1175/MWR3199.1.

Hong, S. Y., & Lim, J. O. J. (2006b). The WRF single-moment 6-class microphysics scheme (WSM6). Asia-Pacific Journal of Atmospheric Sciences, 42(2), 129-151.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. Journal of Geophysical Research, 113, D13103. https://doi.org/10.1029/2008JD009944

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

Janjić, Z. I. (1994). The Step-Mountain Eta Coordinate Model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. Monthly Weather Review, 122(5), 927–945. https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. ArXiv:1412.6980[Cs]. Retrieved from http://arxiv.org/abs/1412.6980

Kiehl, J. T., Hack, J. J., & Briegleb, B. P. (1994). The simulated earth radiation budget of the National Center for Atmospheric Research community climate model CCM2 and comparisons with the Earth Radiation Budget Experiment (ERBE). *Journal of Geophysical Research: Atmospheres*, *99*(D10), 20815-20827.

Kolling, J. S., Pleim, J. E., Jeffries, H. E., & Vizuete, W. (2013). A multisensor evaluation of the Asymmetric Convective Model, version 2, in southeast Texas. Journal of the Air & Waste Management Association, 63(1), 41-53.

Koren, V., Schaake, J., Mitchell, K., Duan, Q.-Y., Chen, F., & Baker, J. M. (1999). A parameterization of snowpack and frozen ground intended for NCEP weather and climate models.

Journal of Geophysical Research, 104(D16), 19,569 – 19,585. https://doi.org/10.1029/1999JD900232

Kristjánsson, J. E., Edwards, J. M., & Mitchell, D. L. (2000). Impact of a new scheme for optical properties of ice crystals on climates of two GCMs. *Journal of Geophysical Research: Atmospheres*, *105*(D8), 10063-10079.

Lefer, B. L., R. E. Shetter, S. R. Hall, J. H. Crawford, and J. R. Olson (2003), Impact of clouds and aerosols on photolysis frequencies and photochemistry during TRACE-P: 1. Analysis using radiative transfer and photochemical box models, J. Geophys. Res., 108(D21), 1125, doi:10.1029/2002JD003171.

Levy, R. C., Remer, L. A., Tanre, D., Mattoo, S. and Kaufman, Y. J. (2009). Algorithm for remote sensing of tropospheric aerosol over dark targets from MODIS: Collections 005 and 051: Revision 2, MODIS Algorithm Theoretical Basis Document for the MOD04_L2 Product.

Li, Y., Pickering, K. E., Barth, M. C., Bela, M. M., Cummings, K. A., & Allen, D. J. (2018). Evaluation of parameterized convective transport of trace gases in simulation of storms observed during the DC3 field campaign. Journal of Geophysical Research: Atmospheres, 123, 11,238–11,261. https://doi.org/10.1029/2018JD028779

Liang, J., & Jacob, D. J. (1997). Effect of aqueous phase cloud chemistry on tropospheric ozone. Journal of Geophysical Research: Atmospheres, 102(D5), 5993-6001.

Liao, H., Y. L. Yung, and J. H. Seinfeld (1999), Effects of aerosols on tropospheric photolysis rates in clear and cloudy atmospheres, J. Geophys. Res., 104(D19), 23697–23707, doi:10.1029/1999JD900409.

Liu, H., J. Crawford, R. Pierce, P. Norris, S. Platnick, G. Chen, J. Logan, R. Yantosca, M. Evans, and C. Kittaka (2006), Radiative effect of clouds on tropospheric chemistry in a global three-dimensional chemical transport model, J Geophys Res, 111(20,303).

Lu Z., X. Liu, Z. Zhang, C. Z., K. Meyer., C. Rajapakshe., C. Wu., Z. Yang., J. E. Penner. (2018). Biomass smoke from southern Africa can significantly enhance the 4 brightness of stratocumulus over the southeastern Atlantic Ocean. Proceedings of the National Academy of Sciences of the United States of America. 115 (12) 2924-2929. DOI:10.1073/pnas.1713703115.

Lu Z., and I.N. Sokolik. (2017). Examining the impact of smoke on clouds and precipitation under different fire regimes: A case study of Yakutsk wildfires of 2002, J. Geophys. Res. Atmos., 122, 12,765–12,785. DOI: 10.1002/2017JD027001.

Lu, Z., and I. N. Sokolik. (2013). The effect of smoke emission amount on changes in cloud properties and precipitation: A case study of Canadian boreal wildfires of 2007. J. Geophys. Res. Atmos., 118. DOI:10.1002/2013JD019860.

Morrison, H., Shupe, M. D., & Curry, J. A. (2003). Modeling clouds observed at SHEBA using a bulk microphysics parameterization implemented into a single-column model. *Journal of Geophysical Research: Atmospheres*, *108*(D8).

Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Monthly weather review*, *137*(3), 991-1007.

NCEP (National Centers for Environmental Prediction). (2000. NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, https://doi.org/10.5065/d6m043c6.

O'Neill, L.W., S.K. Esbensen, N. Thum, , R.M. Samelson, and D. B. Chelton, 2010b: Dynamical analysis of the boundary layer and surface wind responses to mesoscale SST perturbations. *J. Climate*, **23**, 559–581, doi:10.1175/2009JCLI2662.1.

Otkin, J. A., & Greenwald, T. J. (2008). Comparison of WRF model-simulated and MODIS-derived cloud data. Monthly Weather Review, 136(6), 1957-1970.

Pleim, J. E. (2007). A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *Journal of Applied Meteorology and Climatology*, *46*(9), 1383-1395.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743), 195-204.

Rogers, E., and Coauthors (2009). The NCEP North American Mesoscale Modeling System: Recent changes and future plans. Preprints, 23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction, Omaha, NE, Amer. Meteor. Soc., 2A.4.

Rosenfeld, D., Sherwood, S., Wood, R., & Donner, L. (2014). Climate effects of aerosol-cloud interactions. Science, 343(6169), 379-380.

Seinfeld, J. H., Bretherton, C., Carslaw, K. S., Coe, H., DeMott, P. J., Dunlea, E. J., ... & Kraucunas, I. (2016). Improving our fundamental understanding of the role of aerosol−cloud interactions in the climate system. Proceedings of the National Academy of Sciences, 113(21), 5781-5790.

Slingo, A. (1989). A GCM parameterization for the shortwave radiative properties of water clouds. *Journal of Atmospheric Sciences*, *46*(10), 1419-1427.

Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for Weather Research and Forecasting applications. Journal of Computational Physics, 227(7), 3465–3485. https://doi.org/10.1016/j.jcp.2007.01.037

Song, Q., D. B. Chelton, S. K. Esbensen, N. Thum, and L. W. O'Neill, 2009: Coupling between sea surface temperature and low-level winds in mesoscale numerical models. *J. Climate*, **22**, 146–164, doi:10.1175/2008JCLI2488.1.

Thompson, G., Field P.R., Rasmussen R.M., and Hall W.D. (2008). Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. Mon. Wea. Rev., 136, 5095–5115, https://doi.org/10.1175/2008MWR2387.1.

Thompson, G., Tewari, M., Ikeda, K., Tessendorf, S., Weeks, C., Otkin, J., & Kong, F. (2016). Explicitly-coupled cloud physics and radiation parameterizations and subsequent evaluation in WRF high-resolution convective forecasts. Atmospheric Research, 168, 92-104.

Tie, X., S. Madronich, S. Walters, R. Zhang, P. Rasch, and W. Collins (2003), Effect of clouds on photolysis and oxidants in the troposphere, J. Geophys. Res., 108(D20), 4642, doi:10.1029/2003JD003659.

Tiedtke, M., (1989). A Comprehensive mass flux scheme for cumulus parameterization in large-scale models. Mon. Wea. Rev., 117, 1779–1800, https://doi.org/10.1175/1520-0493(1989)117,1779: ACMFSF.2.0.CO;

White, A. T., Pour-Biazar, A., Doty, K., Dornblaser, B., & McNider, R. T. (2018). Improving cloud simulation for air quality studies through assimilation of geostationary satellite observations in retrospective meteorological modeling. Monthly Weather Review, 146(1), 29-48.

Zhang, Y., Xie, S., Lin, W., Klein, S. A., Zelinka, M., Ma, P.-L., Rasch, P. J., Qian, Y., Tang, Q., & Ma, H.-Y. (2019). Evaluation of clouds in version 1 of the E3SM atmosphere model with satellite simulators. Journal of Advances in Modeling Earth Systems, 11, 1253– 1268. https://doi.org/10.1029/2018MS001562

Zheng, Y., K. and co-authors (2016). Improving high-resolution weather forecasts using the Weather Research and Forecasting (WRF) Model with an updated Kain–Fritsch scheme. Mon. Wea. Rev., 144, 833–860, https://doi.org/10.1175/MWR-D-15-0005.1.