DISSERTATION

3D CLOUD NOWCASTING: MERGING EXTRAPOLATION WITH MACHINE LEARNING

Submitted by

Matthew P. King

Department of Atmospheric Science

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2025

Doctoral Committee:

Advisor: Steven D. Miller

Jason M. Apke
Christian D. Kummerow
Susan C. van den Heever
Imme Ebert-Uphoff

ABSTRACT

3D CLOUD NOWCASTING: MERGING EXTRAPOLATION WITH MACHINE LEARNING

Short-term prediction of clouds (0-3 hours), or cloud nowcasting, is critical for both civilian and military operations, ranging from solar energy management to intelligence gathering. Despite the capabilities of contemporary numerical weather prediction (NWP) models, nowcasting methods based on near real-time observations (i.e. satellite imagery) hold operational value due to their relative computational efficiency and accuracy for short-term applications. However, successful cloud nowcasting over large- to global-domains must overcome three challenges simultaneously: 1) trajectory accuracy, 2) computational efficiency, and 3) cloud dissipation and formation. In this dissertation we propose a framework that merges a computationally efficient, cloud extrapolation method with NWP fields and ML to better solve all three challenges.

First, a commonly used nowcasting approach involves using two or more satellite images to retrieve the apparent cloud motions, or the optical flow (OF), which can be used to extrapolate the future location of those features. However, such approaches generally assume that the OF fields remain static with time, which is a limitation when applied to complex, piecewise cloud fields observed by satellites. This study introduces a new OF-based nowcast method that adapts a computer vision technique for image interpolation, commonly referred to as warping, to account for temporal changes to OF fields derived from infrared satellite imagery. With the use of satellite-based cloud property retrievals, we demonstrate that competitively accurate three-dimensional cloud nowcasting can be efficiently implemented using full-disk geostationary imagery that keeps pace with current-generation image scan cadences.

Second, to address cloud dissipation, we propose a method that leverages OF-based nowcasting to generate ML training labels that designates cloud pixels in an initial cloud field for removal with time. Utilizing a dual U-Net approach, one to predict dissipation and the other timing, we demonstrate a proof-of-concept for low-level cloud dissipation that is integrated into an OF-based nowcast framework.

Lastly, to address cloud formation, a post-process approach is implemented that combines OF-based nowcasts and NWP relative humidity fields into a U-Net designed to provide probabilistic predictions for cloud. We demonstrate that this approach further improves cloud nowcasting skill where OF-based nowcast approaches fail to predict cloud. Together, these innovations yield a framework that merges OF, NWP, and ML to produce a computationally efficient and skillful approach for three-dimensional cloud nowcasting on a global scale.

ACKNOWLEDGEMENTS

First, I would like to acknowledge and thank the Air Force Institute of Technology (AFIT) for sponsoring my PhD program and providing the opportunity to continue my service as an Air Force officer as part of the Atmospheric Sciences faculty. Next, I would like to thank Brig Gen Douglas Wickert, Lt Col Kevin Bartlett (ret), Lt Col Robert Wacker (ret), and Dr. Robert Loper, for encouraging me to pursue a PhD and helping me through the selection processes. I would also like to thank my advisor Dr. Steven Miller for being patient and encouraging despite my steep learning curve as a rusty academic after 6 years away from coding and school. Also, I want to thank Dr. Jason Apke for all the time and effort he devoted to me as a mentor and informal advisor through the grueling road that is a PhD. Additionally, I want to thank Dr. Imme Ebert-Uphoff who took the time to give me a crash course in ML and provided priceless feedback and encouragement through my months of trial and error. And of course, I want to thank all my committee members who have been incredibly understanding of my short military mandated timeline. Finally, I would like to thank all the members of the Miller Group for the enlightening discussions, advice, and laughter through these last three years.

Lastly, I want to thank my wife and my daughters who have had to make sacrifices for the family business (military service). Each move has grown harder, but you all have persevered despite the difficulties of being a military family. I love you all with all my heart! And to Kayli in particular, thanks for all the encouragement and sage advice that only a loving wife can impart. I love you!

This research is funded by the Office of Naval Research (ONR) Multidisciplinary University Research Initiative (MURI) Grant N00014-16-1-2040, ONR Broad Agency Announcement (BAA)

Award N00014-21-1-2112, ONR OVERCAST project under award N0001424C2214, and the National Aeronautics and Space Administration (NASA) New Investigator Program (NIP) Grant N80NSSC21K0919.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

POSITIONALITY STATEMENT

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

CHAPTER 1: INTRODUCTION

1.1. Motivation

For the United States in the Cold War era, overhead reconnaissance became a key component in maintaining intelligence superiority against the Soviet Union. Development of overhead reconnaissance ranged from reconnaissance balloons, high altitude aircraft, and of course satellites (Muszyński-Sulima 2023). Systems during the Cold War, however, required cloud-free-line-of-sight (CFLOS) – a direct line between an observer and a target that is absent of clouds – to capture intelligence via photography. Famously, during the Cuban missile crisis, a Corona spy satellite launched specifically to surveil Cuba was unable to provide intelligence on the presence of nuclear missiles due to insufficient image resolution and extensive cloud coverage (Caddell 2016).

Despite modern warfare and intelligence gathering technologies today, the presence of clouds can still impact the battlespace. Impacts can include aircraft hazards such as icing and reduced visibility (Noh et al. 2017; Haynes et al. 2022), usability of directed energy weapons (Burley et al. 2019), or ineffective intelligence gathering from weather sensitive sensors in the air or in space (Golemboski 2001). Because of the impact of clouds, the U.S. Navy Office of Naval Research (ONR) initiated the Optical Variability Evaluation of Regional Cloud Asymmetries in Space and Time (OVERCAST) project with the goal of producing a global 3D representation of clouds in the atmosphere and using it to inform CFLOS calculations and cloud nowcasts in the 0 -3 hour forecast timeframe (Noh et al. 2024), among other applications.

Today's operational geostationary satellites equipped with high resolution and rapid image refresh rates, enable increased potential for accurate and timely cloud nowcasts. Coupled with the advent of graphic processing units (GPUs), dense (every image pixel) calculations of brightness

motions, also known as the "optical flow," can now provide a highly accurate and computationally efficient approximation of cloud motions (Apke et al. 2016, 2018; Apke and Mecikalski 2021). Additionally, the recent breakthroughs in ML (Berthomier et al. 2020; Kellerhals et al. 2022; Yu et al. 2023) provide new potential to improve upon cloud extrapolation methods, particularly when it comes to cloud formation and dissipation. Until recently, DoD operational cloud nowcasting predicted cloud fractions on coarse global grids using global numerical weather prediction (NWP) derived wind fields (Storch and McDonald 2001). To date, there are no studies that have taken advantage of both optical flow (OF) and ML for 3D cloud field nowcasting. Thus, this dissertation seeks to improve the current state of DoD operations as it pertains to OVERCAST's requirements for global cloud nowcasts, exploring how the accuracies and efficiencies of OF when applied to satellite imagery can be merged with ML, and ultimately provide a means to generate large- to global-domain cloud nowcasts that can keep pace with today's advanced, space-based, geostationary imagers.

1.2. Background

On their own NWP models provide reasonably accurate forecasts; however, often NWP model accuracy can underperform short-term forecasts due to inaccuracies of model initial conditions as well as what is often referred to as the "model spin-up problem" which describes the issue where models initially exhibit inaccurate values before model fields have had time to adjust towards a dynamical and physical balance (e.g. Benjamin et al. 2016). Such accuracy problems early in model outputs drive end-users and stakeholders to instead use nowcasting methods based on observations, such as persistence or extrapolation of the currently observed cloud distribution, which perform better than models for short-term forecasts (Vendrasco et al. 2020). Persistence, otherwise known as Eulerian persistence, is simply keeping the most recent observation as the

prediction (Germann and Zawadzki 2002). Extrapolation, sometimes referred to as Lagrangian persistence, assumes that the state of each air parcel is constant and therefore all change to the object of the nowcast (precipitation or clouds) are due only to the background flow (Prudden et al. 2020). Frequently mentioned in nowcasting literature, particularly in the context of radar-based precipitation nowcasting, extrapolation methods estimate motion fields based on two or more past observations, and use inferred motions to advect the initial observations forward in time (Prudden et al. 2020).

Although often superior in skill in the near-term, extrapolation (advection) methods are certainly not without their drawbacks. According to Bowler et al. (2006), the most significant errors from such methods are from: 1) errors in the motion field; 2) not accounting for the evolution of the motion field; 3) not accounting for Lagrangian evolution. One of the early examples of cloud nowcasting that attempted to solve some of these sources of error is the United States Air Force's Advect Cloud model which predicted short-range cloud movements globally (Storch and McDonald 2001). Similarly, as intended with OVERCAST, Advect Cloud was initialized using a global cloud analysis based on satellite data from geostationary and polar orbiting platforms as well as surface observations. The model used trajectories computed from NWP wind data to account for the motion field and its evolution. Importantly, to model the Lagrangian evolution of cloud coverage percentage, the model advected a moisture parameter known as condensation pressure spread (CPS) and used vertical pressure displacement to update the CPS value and subsequently the cloud coverage percentage empirically.

Since the 1960s, wind fields referred to as atmospheric motion vectors (AMVs) have been retrieved from meteorological satellite imagery of cloud and water vapor drift (Menzel 2001). AMVs are derived using OF retrievals which was originally defined in computer vision literature

by Horn and Schunck (1981) in their seminal paper as the "distribution of apparent velocities of movement of brightness patterns in an image." Operational AMVs specifically would use a patch matching method which would in simple terms match target areas between images using either cross-correlation or least squares methods (Fortun et al. 2015 and references within). Recently, Apke et al. (2022) showed that a dense optical flow (DOF) retrieval method can outperform the patch matching-based AMVs when compared to ancillary wind retrieval datasets (Daniels et al. 2020). OF at the high spatial and temporal resolutions of today's advanced imagers, such as the Geostationary Operational Environmental Satellite (GOES)-R series Advanced Baseline Imager (ABI; Schmit et al. 2017) and the Himawari Advanced Himawari Imager (AHI; Bessho et al. 2016), may also allow for new accuracy gains in cloud nowcasting, especially in regions where model wind accuracies break down.

Although advances in OF may provide a means to improve extrapolation nowcasts of cloud distribution, OF does not provide a means to predict the cloud formation and dissipation. Breakthroughs in deep learning of artificial neural networks over the last few decades have spurred the renewed use of ML methods in scientific applications. Consequently, nowcasting research has also begun to embrace ML in efforts to improve nowcast skill beyond what is possible with Lagrangian persistence assumptions. Many recent publications have utilized convolutional neural networks (CNNs) and other similar deep learning networks (Hong et al. 2017; Berthomier et al. 2020; Ionescu et al. 2021; Yu et al. 2023) and have found some improvements to nowcast skill. Such systems, however, demand that ML learns both evolution and advection, which could be significantly simplified if combined with advection-based systems.

1.2.1. Advection Nowcasting

As previously mentioned, the Air Force Advect Cloud model served as an early method for nowcasting cloud cover globally. Unlike more recent approaches, Advect Cloud did not utilize extrapolation on satellite observations to estimate a motion field, but rather informed advection calculations based on wind speeds from a NWP model. Advection from Advect Cloud was calculated on multiple levels thereby allowing for the advection of cloud coverage percentage in 3D in hourly increments. This approach provided useful forecasts for the DoD; however, it was computationally costly due to its adherence to the NWP-like approach rather than a more observational approach which ultimately limited its resolution to synoptic scale cloud features.

Guillot et al. (2012) compared persistence to similar displacement nowcasting techniques over the complex terrain of Utah and southwest Wyoming. They used Terra MODIS (~1030 local time crossing) cloud masks to initialize a cloud nowcast and Aqua MODIS (~1330 local time crossing) cloud masks to verify the nowcast over forecast timeframes of 2 to 4.5 hours. The study found that using radiosonde observations to apply a displacement of the initial cloud field could not outperform persistence (the most computationally simple cloud nowcast).

Miller et al. (2018) implemented another advection method to predict short-term solar insolation for renewable energy production where NWP winds were used to advect groups of clouds based on similar spatial and optical properties retrieved from the Clouds from the Advanced Very High Resolution Radiometer (AVHRR) – Extended (CLAVR-x). Their method took advantage of near-real time geostationary imagery, but used a NWP forecast wind field which was evolved by using linear interpolation between available NWP field times. Clouds were grouped based on retrieved cloud property criteria and were displaced based on the winds associated with their cloud top heights. Although they implemented a technique to evolve the advection wind field,

their assumption of Lagrangian persistence prohibited any changes in cloud formation or dissipation.

Alternatively, Descombes et al. (2014) utilized a simplified version of the Weather Research and Forecasting (WRF) model to advect cloud fraction observations using model winds at different vertical levels. Their method, referred to as Multi-sensor Advection Diffusion nowcast (MADCast) eventually was incorporated into WRF-Solar, a version of the WRF model specifically tailored for solar energy applications, to create MAD-WRF which would infer the presence of clouds based on relative humidity from cloud mask, cloud top height, and cloud base height satellite retrievals to create a 3D cloud analysis which could be blended with the cloud physics of the WRF-Solar model thereby enhancing the accuracy of the model during its short-range forecasts from 0 to 6 hours (Jiménez et al. 2022). This approach again would break down where model winds are inaccurate.

1.2.2. Extrapolation and Optical Flow-based Nowcasting

Rather than relying on NWP models for motion fields, nowcasting via extrapolation or OF involves estimating an advection field from recent observations and then using that field to advect the observations based on the Lagrangian persistence assumption (Prudden et al. 2020). The first challenge with OF nowcasting methods is the estimation of the advection field. Early methods sought a single displacement vector that maximized the cross-correlation coefficient between sequential observations(Prudden et al. 2020). One such method in radar nowcasting by Austin and Bellon (1974) would use an initial guess based on the center of gravity of each image. Efforts from Rinehart and Garvey (1978) and Tuttle and Foote (1990) showed that using cross-correlation between arrays or “boxes” between sequential radar images could provide reasonable estimates of

non-uniform motion on radar data (though such methods suffer similar problems to patch matching-based AMVs which can be overcome with DOF retrieval-based methodologies).

Despite the lack of DOF methods in operational AMV retrievals, such methods have actually been applied in the computer vision community for decades and originate from Horn and Schunck (1981). Their method utilized calculus of variations, hence known as "variational" optical flow (VOF) retrieval, to minimize a penalty integral which was a function of the input images and a guessed OF field (initially stationary motions) that is then iteratively updated based on assumptions of brightness constancy and flow smoothness that follow subjective human vision interpretation. Many improvements have been made to VOF algorithms which have focused on updating algorithms and penalties when assumptions of brightness constancy and flow smoothness fail.

Improvements to capture displacements larger than a single pixel (Anandan 1989; Brox and Malik 2011), preserve motion discontinuities (Black and Anandan 1996), handle motions caused by illumination changes (Brox et al. 2004), relax unrealistic flow smoothness assumptions (Corpetti et al. 2006; Zimmer et al. 2011), and use of non-local image information to enhance flow retrieval accuracy (Sun et al. 2014) have furthered the capabilities of modern DOF algorithms over the decades. Sundaram et al. (2010) has shown that the implementation of dense optical flow computed with parallel processing of graphic processing units (GPUs) decreases computation time significantly which paved the way for Apke et al. (2022) to show that such implementations with GPUs allow for VOF algorithms to have sufficient computational efficiency to keep pace with the high spatial and temporal resolution of today's satellite instruments such as ABI and AHI to produce products in near-real time.

Although VOF algorithms have improved significantly over the last four decades, applying 2D optical flow fields to a 3D structure of clouds is not a straightforward task. For radar nowcasting,

Prudden et al. (2020) explain that once an advection field has been determined, often numerical advection is achieved via a semi-Lagrangian scheme in which the Lagrangian persistence assumption is maintained and trajectories are calculated either forward to a target time or backward from a target time, the latter being the most commonly used. Despite radar derived data often having a varying component of height as a function of the radial distance from the radar, advection for radar nowcasting can reasonably be assumed to be a 2D advection problem. Optical flow calculated on meteorological satellite imagery is, in contrast, a piece-wise field due to the vertical extent of observed clouds and surface features which must be advected differently.

One example of advection nowcasting with piece-wise fields is the EUMETSAT Extrapolated Imagery Product or EXIM (Jann 2017). To perform the advection step after a calculation of AMVs, EXIM assumes that the calculated displacement field remains constant with respect to time over the extent of its nowcast iterations. Thus, to determine the new position for each pixel (at $t + T$) where t is the initial time of the nowcast and T is a user defined time interval, a new pixel position is determined by assuming the movement approximated at time t remains constant. For a second time step, however, (at $t + 2xT$), the movement at the new pixel position is assumed to be the movement observed at time t for that new pixel location; consequently, time-related changes of the advection field are not accounted for in EXIM advection calculations and the method is essentially an example of linear extrapolation of the OF field.

Many other documented studies and research have similar approaches to applying optical flow fields derived from satellite imagery with the same assumption of a static advection field with time. Nonnenmacher and Coimbra (2014) use an optical flow field to calculate a streamline that will determine an area of cloud that will likely drift to a target location assuming a constant flow field with time. Lorenz et al. (2004) used a relatively simple optical flow-derived motion field that

scaled cloud displacement based on the time between the observation and forecast which could not outperform persistence unless the forecast was smoothed. Urbich et al. (2018) compared short-term forecasts of effective cloud albedo using two different methods of optical flow algorithms, one from Farnebäck (2003), which used polynomial expansion to approximate the neighborhood of two initial frames (which struggles to produce accurate DOF fields in the absence of brightness texture), and the second from Zach et al. (2007) which used a variational formulation based on total variation regularization and L1 norm (TV-L1), and found that the TV-L1 method produced better short-term forecast results. Urbich et al. (2018) however, applied the same motion vectors recursively to attain a desired forecast timeframe. In a study by Kosmopoulos et al. (2020), short-term forecasts of downwelling surface solar irradiation were evaluated using the same optical flow methods applied by Urbich et al. (2018), but motion vectors were applied on cloud optical thickness (COT) and cloud modification factor (CMF). Special care to explain the optical flow calculations was taken, but little was explained on how motion was applied to COT and CMF which implies that the same motion initially calculated was applied throughout the forecast timeframes. Nielsen et al. (2021) and Kellerhals et al. (2022) both apply OF algorithms as benchmarks to compare different ML approaches and found OF approaches to have lower skill though these methods do not modify OF motions for successive time steps. In fact, Kellerhals et al. (2022) states that "optical flow methods…lack the ability to take into account advection patterns seen across multiple time steps and across multiple cloud evolution sequences."

Similarly, Yu et al. (2023) compared two ML models to predict cloud top height (CTH) and compared skill to the Real-time Optical flow by Variational methods for Echoes of Radar (ROVER) algorithm proposed by Woo and Wong (2017). Yu et al. (2023) found that their ML model approaches outperformed persistence and ROVER, but ROVER, designed for radar data

and not the piecewise fields of satellite imagery, again assumes a constant flow field with time. Their assessment to optical flow approaches mirrored the view of Kellerhals et al. (2022) in that they did not see a reasonable way to update future flow fields using the optical flow method, an idea cited originally from Shi et al. (2015). Such results imply that OF retrieval is not enough alone for cloud nowcasting and must have updates applied either to the flow field or the advected cloud field to improve the accuracy of a cloud nowcast.

1.2.3. Machine Learning Nowcasting

Most of the attempts of ML for cloud nowcasting have involved some version of a convolutional neural network (CNN) – a technique which emerged from the study of the brain's visual cortex and use convolutional kernels or filters to aid in recognizing patterns in images (Géron 2023). Springing from the progress of artificial neural networks, CNNs like other deep learning methods on data by minimizing a loss function via gradient descent and back-propagation through the network (Géron 2023). Many of the ML networks in cloud nowcasting are some form of recurrent neural networks (RNN) which are designed to extract patterns from a scalar time series (Bansal et al. 2023). However, due to problems with training RNNs, many networks have been equipped with long-short-term memory (LSTM) architectures which are able to ensure memory from early time steps are carried to much later time steps (Bansal et al. 2023). The LSTM concept has also been expanded to images with the use of convLSTM architectures which can operate on image sequences to extract spatial features across time or spatiotemporal features (Bansal et al. 2023).

One of the first attempts at applying ML to the cloud nowcasting problem was from Hong et al. (2017) who used a convolutional sequence-to-sequence network that included two RNNs using an encoder-decoder framework to predict satellite images based on the input of observed satellite

images. Although Hong et al. (2017) did not utilize a benchmark such as persistence to compare with ML methods, it was determined that architecture choices, particularly skip connections, played a large role in the accuracy of predictions. Berthomier et al. (2020) attempted to use multiple ML architectures to nowcast cloud cover and found that a U-Net architecture, a specific form of a CNN, performed the best against other CNNs, RNNs, and networks with LSTMs as well as EUMETSAT's EXIM tool, and a European high resolution NWP model. Ionescu et al. (2021) adapted Google's Xception architecture of depthwise separable two-dimensional convolutions to predict future satellite imagery over Romania and found better skill over other ML methods, except RNNs which require significantly more data to train properly. As mentioned earlier, Nielsen et al. (2021) and Kellerhals et al. (2022) compared ML methods and showed outperformance of OF methods. Nielsen et al. (2021) used an Autoencoder ConvLSTM based architecture detailed by Shi et al. (2015) and a multistate dynamic generative adversarial network (MD-GAN) model based on the architecture from Xiong et al. (2018) on cloud-labeled satellite images and found overall better skill than persistence from both ConvLSTM and MD-GAN with ConvLSTM slightly outperforming MD-GAN. Kellerhals et al. (2022) applied a convolutional gated recurrent unit network (ConvGRU), similar to the architecture described by Shi et al. (2015), which took in 2D cloud fields at multiple timesteps and output forecast 2D cloud fields. Kellerhals et al. (2022) found that the often "blurry image" problem characteristic of convolution-based ML approaches was mitigated with the use of the structural similarity index measure (SSIM) as a loss function in model training. Another key finding by Kellerhals et al. (2022) was that their best version of their ConvGRU appeared to have learned advection of clouds, but had not discovered ways to account for cloud formation or dissipation. In their recent work Yu et al. (2023) nowcasted CTH data utilizing two neural networks, ConvLSTM and TrajGRU, which the latter was proposed by Shi et

al. (2017) as an improvement to ConvLSTM in precipitation nowcasting. Using mean square error (MSE), peak signal to noise ratio (PSNR), and Pearson correlation coefficient (PCC), Yu et al. (2023) showed TrajGRU trained using SSIM outperformed the ConvLSTM model and persistence. Interestingly, despite the performance results and the training using SSIM for their TrajGRU model, their published results show qualitatively a great deal of smoothing compared to their ground truth and persistence plots.

1.3. Research Goals and Approach

Cloud nowcasting has three challenges that have yet to be overcome simultaneously: 1) trajectory accuracy, 2) computational efficiency, and 3) cloud formation and dissipation. Optical flow can compute accurate trajectories due to solving for motion from the most recent observations, however, projecting those motions forward has been difficult due to the evolving dynamical field. ML approaches may provide computational efficiency and the potential for uncovering new statistical relationships related to cloud formation and dissipation. NWP models have the capability to predict cloud formation and dissipation based on the current knowledge and approximations of the underlying physics of the atmosphere but are too computationally expensive to match the high cadence of today's remote sensing platforms. Previous cloud nowcasting attempts have not yet merged the capabilities of optical flow, ML, and NWP to overcome all three cloud nowcasting challenges at the same time. Thus, the goal of this research is to explore possible avenues that make the merging of all three capabilities a reality.

OVERCAST places unique constraints on this research and the cloud nowcasting problem. First, nowcast approaches used must have the capability to operate over large- to global domains at a cadence (10-min) and spatial resolution (2 km) that matches full-disk infrared bands in current generation geostationary imagers. Second, nowcasts must have the capability to operate through

the full diurnal cycle. This requires the use of infrared imagery rather than visible imagery which has higher spatial resolution. Lastly, nowcasts of 3D cloud fields and cloud properties, not just single-variable imagery, are required. This constraint is required to enable the computation of CFLOS and the determination of electromagnetic spectrum limitations for DoD applications. Ultimately, these constraints limit nowcast approaches that can predict the new location of cloud properties, such as CTHs, rather than methodologies that can only predict new imagery.

Consequently, rather than take an approach that relies solely on ML to solve all of the problems related to 3D cloud nowcasting, a physics-informed ML (Kashinath et al. 2021; Meng et al. 2025) approach is utilized in this research. When using a physics-informed ML approach, physical principles and domain knowledge are applied to reduce the complexity of a task or problem thereby reducing the complexity required of ML. A reduction of ML complexity reduces the need for training data, accelerates training processes, and leads to a final ML solution that is generally more physically consistent, scientifically sound, and more transparent and interpretable (Kashinath et al. 2021).

Thus, rather than design a complex model to solve the advection of clouds as well as their evolution over time, the approach taken here is modular in that advection, dissipation, and formation are approached separately. This enables the ability to gain insights into each phenomenon either by using a physical solution (i.e. optical flow) or by constraining the problem to specific phenomenon such as cloud dissipation or formation.

1.4. Research Questions

To discover how to merge the benefits of optical flow, ML, and NWP, in the chapters that follow this dissertation seeks to answer the following questions which include relevant hypotheses:

1) To what extent does accounting for temporal changes in a piecewise OF field improve predictions of 3D cloud movements?

   Hypothesis: if the OF field is updated to account for time related changes for each nowcast time step, then there will be an increase in cloud nowcast skill when compared to a linear extrapolation method in which the OF field is assumed constant with respect to time, which is a common approach in OF-based nowcast applications. This hypothesis is examined by adapting a temporal interpolation method, commonly referred to as warping, which includes simple occlusion reasoning to update the OF field iteratively to better account for temporal changes against a nowcast method that applies OF with simple linear extrapolation.

   Hypothesis: if the same apparent motion observed from cloud top height (CTH) is also applied to cloud base heights (CBH), then the resulting 3D cloud nowcast will be more accurate than a method that applies estimated wind speeds for individual levels. This hypothesis is examined by comparing 3D cloud nowcasts that have been created by linking CTH and CBH with the same apparent motion field against 3D cloud nowcasts that applies advection of cloud layers at multiple levels based on NWP estimated wind velocities.

2) To what extent do observed changes in cloud properties and coarse-resolution NWP relative humidity (RH) fields predict future cloud dissipation in the short-term nowcast timeframe (0 – 3 hours)?

   Hypothesis: if there is predictive information from a series of OF-based cloud nowcasts of cloud properties and NWP RH values, then ML will provide measurable skill in predicting short-term cloud dissipation. This hypothesis is examined by implementing a U-Net architecture that is trained on truth labels generated by identifying what initial cloud pixels

in a OF-based nowcast must dissipate in order to account for the observed dissipation of clouds.

Hypothesis: if short-term changes of nowcasted cloud properties observed across time provide more predictability than NWP RH fields due to the recency of the nowcasts, then an ablation of a ML model that contains nowcasts as input versus an ablation containing NWP RH fields will show more dissipation skill. This hypothesis is examined with an ablation study on the U-Net architecture to determine which inputs (nowcast or NWP) impart more predictability.

3) To what extent does the extrapolation of clouds and modeled environmental information improve predictability of cloud formation?

   Hypothesis: if the context of where clouds advect and NWP RH predictions provide cloud formation predictability, then ML will provide measurable skill on where clouds are likely to exist and therefore likely to form. This hypothesis is examined using OF-based nowcasts as well as temporally interpolated NWP RH fields to specific timeframes as input into a U-Net to provide a probabilistic prediction of cloud to inform where clouds should be added to a nowcast.

   Hypothesis: if NWP RH fields contribute to cloud formation more than extrapolated cloud nowcasts, then an ablation of a ML model that contains NWP RH fields will provide better cloud classification than an ablation that contains only extrapolated cloud nowcasts. This hypothesis is examined by training ablations of a U-Net to determine which inputs, nowcasts or NWP RH, impart more predictability to clouds.

The rest of the dissertation is organized as follows. Chapter 2 addresses the use of temporal interpolation integration into OF-based nowcasting and the assumption of connecting CTHs and

CBHs for 3D cloud nowcasting. Entitled “Nowcasting 3D Cloud Fields Using Forward Warping Optical Flow”, the work in chapter 2 is under review with the *Journal of Atmospheric and Oceanic Technology*. Chapter 3 addresses cloud dissipation by leveraging OF-based nowcasting to generate cloud dissipation labels. Entitled “Machine Learning Prediction of Short-Term Low-Level Cloud Dissipation Using Optical Flow Labeling”, the work in chapter 3 will be submitted to the *Journal of Geophysical Research: Machine Learning and Computation*. Chapter 4 addresses cloud formation by utilizing ML to produce probabilistic predictions of cloud by combining NWP RH fields with OF-based cloud nowcasts as inputs. Entitled “Probabilistic Prediction of Cloud Formation Using Optical Flow-based Nowcasting and Numerical Weather Prediction”, the work in chapter 4 will be submitted to the *Journal of Geophysical Research: Machine Learning and Computation*. Finally, Chapter 5 concludes with a summary of findings and future work.

CHAPTER 2: NOWCASTING 3D CLOUD FIELDS USING FORWARD WARPING OPTICAL FLOW

2.1. Introduction

It has been established that OF retrieval alone is not enough to match or exceed the capabilities of today's ML architectures, namely due to its inability to both capture changes in the cloud motion field with time and changes in cloud morphology (Berthomier et al. 2020; Nielsen et al. 2021; Kellerhals et al. 2022; Yu et al. 2023). However, as will be shown in this chapter, it is possible to provide a simple modification to the OF field to vastly improve its nowcasting capabilities without adding significant computational expense. Such a modification, which is termed "forward warping" here, can help to address the first research question, and apply an advanced optical flow retrieval that preserves the piecewise observations of cloud fields observed from satellites. This chapter introduces the concept of OF and detail its application to the nowcasting of 3D cloud distributions.

2.1.1. Optical Flow

Many cloud nowcasting methodologies require up-front OF retrievals to function properly. As previously mentioned, objective OF retrieval methods originate from the seminal work of Horn and Schunck (1981) in which they pose the problem of determining OF by way of a penalty (or "energy") function to be minimized based on assumptions of the flow field behavior and consistency of brightness with time. Their method utilizes calculus of variations and hence are commonly referred to as "variational" retrieval techniques, to resolve fully dense (i.e. at every image pixel) motions over a given sequence of images. Over the decades, OF algorithms have been improved where assumptions of brightness constancy and flow smoothness fail to handle

challenges such as illumination changes, deformations, and motion discontinuities (see Fortun et al. 2015 and references within).

Today, the state-of-the-art OF algorithms often are based on data-driven machine learning (ML) approaches (e.g. Teed and Deng 2020). ML approaches are computationally efficient and highly accurate; however, their failure points or principal inaccuracies are difficult to diagnose due to their design complexity and lack of transparency and interpretability. It is for this reason that variational OF approaches remain relevant and popular in Earth Sciences research such as for wind retrieval (Stettner et al. 2019; Apke et al. 2022), tracking outflow boundaries (Apke et al. 2020), and deriving sea ice motions (Petrou and Tian 2017). Given that such variational approaches are fully dense, interpretable, and flexibly tunable, they are also optimal for applications such as cloud nowcasting.

2.1.2. Cloud Nowcasting from Satellite Imagery

Most short-term (0-3 hours) cloud and precipitation-nowcasting methodologies use some form of observations from satellite or radar imagery to predict the near-term future state, typically with a few assumptions on morphology. An Eulerian persistence assumption, or simply persistence, is the most basic approach to nowcasting clouds and precipitation, where the most recent observations are used for the prediction (Germann and Zawadzki 2002). In contrast, extrapolation assumes that all change is due only to the background flow (referred to as the "Lagrangian persistence assumption"), meaning initial observations are advected forward in time based on an advection field estimated either from NWP winds interpolated across time (Miller et al. 2018), or from a constant flow field estimated from OF retrievals (Bowler et al. 2006). Lagrangian persistence is heavily used for radar-detected precipitation nowcasting, often as a benchmark to compare with more complex prediction architectures that accommodate feature formation and

decay (Bechini and Chandrasekar 2017; Pulkkinen et al. 2019). When applied to cloud nowcasting with satellite imagery and products, most studies find Lagrangian persistence underperforms compared to alternate methods that take advantage of more complex nowcasting routines, such as those based on ML which despite their complexity have not solved cloud evolution (Berthomier et al. 2020; Nielsen et al. 2021; Kellerhals et al. 2022; Yu et al. 2023).

With the now dominating presence of ML in atmospheric sciences, many new ML-based methods and approaches have become ubiquitous. On a global scale, ML efforts have been successful in incorporating physics-based constraints and have achieved unprecedented accuracy and speed in medium-range weather forecasting such as GraphCast (Lam et al. 2023) and NeuralGCM (Kochkov et al. 2024). Of course, such models are trained on relatively coarse resolution (e.g. 31 km) reanalysis datasets which do not capture the 3D cloud structures needed for CFLOS computations, limiting their usefulness for projects like OVERCAST. Furthermore, CFLOS computations require predictions that preserve the physical relationships between multi-variate datasets predicted (i.e., pixels with predicted cloud-top heights must also have cloud bases, optical depths, and geometric thicknesses), which is still very much an open challenge in ML-based prediction research.

In the context of fine resolution applications, recent improvements in nowcasting skill for precipitation using radar imagery have emerged when using ML-based methods. Ravuri et al. (2021) introduced a conditional generative adversarial network (cGAN) that used past twenty minutes of radar frames as input to generate ensembles of future precipitation over the next 5 to 90 minutes. Their Deep Generative Model of Rain or DGMR provided sharp, probabilistic radar nowcasts outperforming some leading baselines in radar nowcasting with some struggle in predicting heavy precipitation at long lead times. Yang and Yuan (2023) used a customized multi-

scale framework that used a flexible attention module, a mechanism that allows a model to dynamically focus on the most relevant parts of an input, to capture features across multiple spatial and temporal scales to improve 1-hr nowcasts of vertically integrated liquid from the storm event imagery data set or SEVIR. Their implementation of their attention module along with a customized loss function that penalized pixel-wise errors, spatial errors and temporal errors enabled them to improve various deep learning methods to include DGMR.

Ha and Lee (2023) incorporated the use of a cGAN to update an OF-based nowcast of radar data which was derived by using the average of flow fields estimated for different valid times. They used linear extrapolation to advect their radar data with their final and constant OF field and then relied on the cGAN to provide nonlinear changes to the radar data such as growth and decay. They found that their method could outperform the use of linear extrapolation with OF. Ha and Lee (2024) also used multiple OF algorithms with various levels of strengths and weaknesses to linearly extrapolate future radar images. Those new images were then used in a linear regression model to create a new radar image prediction that served as an input to a U-Net that produced a final radar nowcast prediction. They found that their approach outperformed linear extrapolation using single OF algorithms but found that the U-Net was producing blurred results when compared to the output of their linear regression model.

Leinonen et al. (2023) utilized a recurrent-convolutional neural network to predict lightning, hail, and heavy precipitation probabilistically to a lead time of 60 min. They found using Shapley values that despite the use of lightning detection, satellite imagery, and NWP data as inputs into their neural network, radar products where the most important predictors for lightning, hail, and heavy precipitation. Kim et al. (2024) employed a deep neural network that utilized NWP-supplied atmospheric variables to include divergence at 925 hPa and total column water vapor. They found

their network, when trained using a loss function based on critical success index and false alarm ratio, could outperform radar echo extrapolation.

Despite all these successes in the radar nowcasting domain, applying the same or similar methods for satellite imagery and cloud nowcasting is exceptionally challenging. Radar and rain-rate nowcasting is a single-variable, 2D advection problem, as methods are typically designed to predict future precipitation characteristics detected from either a constant-elevation remapped grid or a plan position indicator product. However, cloud motion fields observed from Earth-viewing satellite imagery are piecewise continuous, containing discontinuities around cloud edges and in multi-layer cloud scenes that must be reconciled (e.g. Figure 2.1; Apke and Mecikalski 2021).

Because of the vertical variation of satellite-observed cloud top heights, motions that are retrieved from satellite imagery include motions at varying altitudes. These characteristics to OF-derived motions from satellite imagery present a formidable challenge when applying satellite derived cloud motions to cloud nowcasts. As evident in Figure 2.1, significant temporal changes can occur to a satellite viewed motion field over the course of three hours and thus makes a constant motion field an insufficient assumption for nowcasting applications. Specifically, the motion field in Figure 2.1b is drastically different from the motion field observed 3 hours earlier in Figure 2.1a. When overlaying the original motion field over the imagery valid 3 hours later (Figure 2.1c), it becomes clear that assuming a constant motion field with time will neglect the motion of clouds into areas without clouds. The difference between the final motion field and the initial motion field (Figure 2.1d) shows the drastic errors that are introduced when assuming a constant motion field with respect to time, particularly over areas where clouds move over clear regions. Furthermore, as discussed in Lam et al. (2023) and Kochkov et al. (2024), multi-variate nowcasting

when training a model to learn predictions of multiple variables can induce a loss of conventional physical properties that exist in nature between such variables.

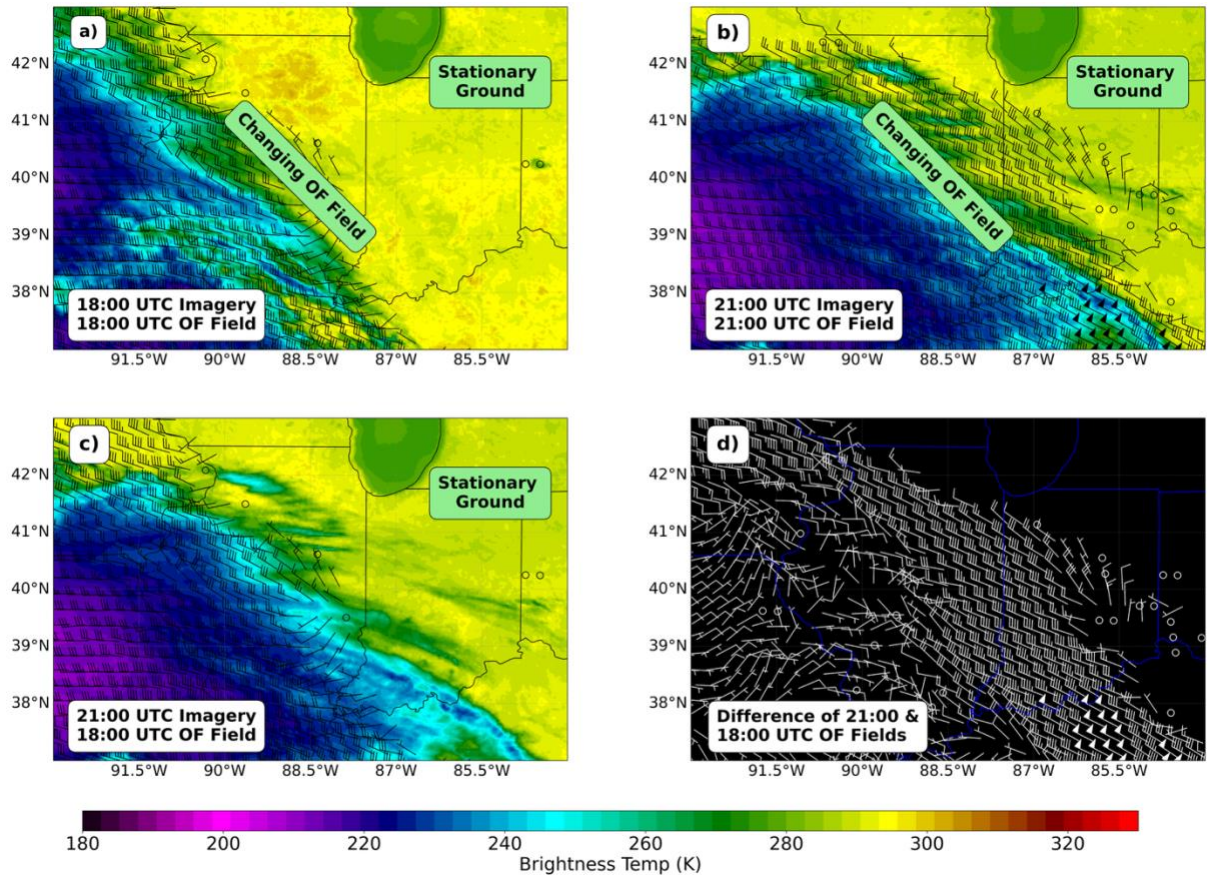

Figure 2.1: a) GOES-16 brightness temperature imagery showing OF-derived winds represented by wind barbs (m/s) valid at 18:00 UTC on 25 April 2024 over Illinois and Indiana. b) GOES-16 brightness temperature imagery valid at 21:00 UTC with 21:00 UTC OF-derived winds. c) GOES-16 brightness temperature imagery valid at 21:00 UTC with the 18:00 UTC OF-derived winds. d) Difference of 21:00 UTC OF field and 18:00 UTC OF field. Pixels with wind speeds less than 1 m/s have been removed and only every 10th pixel wind speed is plotted to avoid clutter.

2.1.3. Temporal Interpolation and Warping

Similar challenges are encountered by the computer vision community when performing tasks such as video frame interpolation and novel view synthesis (e.g. Hani et al. 2020; Dong et al. 2023). In video frame interpolation, two images scanned from the same viewpoint at different times must be blended to synthesize the appearance of a temporally intermediate scanned frame. The same challenge holds for novel view synthesis, except the two images are scanned simultaneously from two different viewpoints, and the intermediate frame represents a view between the perspectives where the images were scanned.

Many techniques benefit from estimation of an intermediate flow field to realistically forward- and backward-extrapolate each scanned image while preserving layers within the images (Jiang et al. 2018). Such an estimation can be made by simple warping based on an optical flow estimate (Baker et al. 2011), or even by using image-to-image ML techniques (e.g. Huang et al. 2022). The two extrapolated images are then blended into the resulting frame, typically with some form of reasoning to accommodate occlusions (where pixels are only visible in one image due to overlap of features; Herbst et al. 2009). Apke et al. (2025) demonstrate how this scheme can be used to correct for cloud-morphology-based discrepancies between different imagers scanning at different times in order to improve downstream large-domain multi-instrument cloud product mosaics while notably preserving multi-variate physical relationships needed for CFLOS computation.

Nowcasting essentially represents half of the frame interpolation problem in that, rather than determining an intermediate flow field, a flow field is only extrapolated forward. To the best of our knowledge, forward-flow estimation and warping have yet to be employed for the problem of cloud nowcasting. When paired with an advanced OF retrieval algorithm, such estimation could provide both a powerful baseline to compare against complex ML architectures, as well as novel,

computationally inexpensive predictors for ML to simplify architectures applied to cloud nowcasting. In this chapter we develop a new technique, Improved Nowcasting via Optical-flow Warping (I-NOW) and evaluate its performance against traditional Eulerian and Lagrangian persistence methods in complex, real-world cloud scenarios.

2.2. Data and Methods

2.2.1. Cloud Data

The backbone of OVERCAST's 3D cloud analyses are the 2D pixel-level cloud retrievals obtained from multiple satellite sensors (Noh et al. 2022a, 2024). In this dissertation, cloud retrieval products are processed from Clouds from the Advanced Very High Resolution Radiometer (AVHRR) – Extended or CLAVR-x (Heidinger et al. 2012; Walther and Heidinger 2012). CLAVR-x is a research development framework for the National Oceanic and Atmospheric Administration (NOAA) Enterprise Cloud Algorithms (available online at: https://cimss.ssec.wisc.edu/cspp/clavrx_v2.0.shtml) originally designed for AVHRR sensors on NOAA's Polar-orbiting Operational Environmental Satellite (POES) constellation, but has since extended to other satellite radiometers.

In the modern version of CLAVR-x, the cloud algorithms utilizing various channel combinations apply not only to polar-orbiting satellites (including the current-generation of NOAA's Joint Polar-orbiting Satellite System; JPSS; Goldberg et al. 2013) but also to images onboard geostationary platforms such as the GOES (Geostationary Operational Environmental Satellites) Advanced Baseline Imager (ABI; Schmit et al. 2017) series as well as international imagers on platforms such as European Meteosat and Japanese Himawari satellites (Miller et al. 2018). For OVERCAST, the individual 2D cloud data fields from multiple sensors are transformed into a 3D global grid system (currently 0.02 x 0.02 degrees and 500-m vertical

intervals up to 20 km). In this chapter, however, vertical intervals are based on GFS pressure levels for the sake of faster computation and easier comparison with selected baselines.

Whereas CLAVR-x provides numerous cloud properties, the 3D cloud state for initial and "truth" conditions in this study relies on just three: 1) Cloud Mask, 2) Cloud Top Height (CTH), and 3) Cloud Base Height (CBH). Algorithms for each of these retrievals are contained in their respective algorithm theoretical basis documents (Heidinger and Straka 2020; Heidinger et al. 2020; Noh et al. 2022b). Of special note with using CBH is that, as explained by Noh et al. (2022b), height estimation is conducted under the assumption of the estimation of the uppermost cloud layer, and is thus optimal for a single-layer cloud. Consequently, multilayer cloud scenes may provide the possibility for incorrect cloud states, particularly for the lowest cloud height. To correct for some of these inaccuracies, NOAA's Global Forecast System (GFS) (available online at: https://registry.opendata.aws/noaa-gfs-bdp-pds) fields of cloud base heights (specifically "low cloud bottom level", "middle cloud bottom level", and "high cloud bottom level") were utilized to define where gaps may be present between CTH and CBH.

To test the skill of the I-NOW nowcasts, rectilinear grids were defined with 0.02-degree horizontal grid-spacing and a vertical dimension of pressure using GFS pressure values from 1000 to 100 hPa (specifically 1000, 975, 950, 925, 900, 850, 800, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 150, and 100 hPa). Values within the grid were defined as 0 for "no cloud" and 1 for "cloud." Clouds were assumed to be present based on the CLAVR-x supplied probabilistic cloud mask flag of either "probably cloudy" or "cloudy." The nearest-neighbor (NN) method was used to place cloud data onto the grid while GFS data were bilinearly interpolated due to coarse grid spacing. Latitude and longitude locations for CTH and CBH were parallax corrected using formulas defined in Miller et. al (2018). Because holes in the cloud data were created from

applying parallax correction, an outside-in strategy that averaged the nearest grid values was used to “fill” the resulting holes.

In forming our evaluation dataset, 1000 nowcasts were randomly selected using GOES-16 data from 2024. Out of the 1000 nowcasts 9 had instances where CLAVR-x data exhibited abrupt changes in cloud retrievals (due likely to a channel drop out) causing commensurately abrupt changes in skill for all methods for a single time step. The 9 cases were consequently omitted from the final validation to avoid introducing cloud retrieval algorithm discrepancies to the nowcast evaluations thereby leaving a total of 991 cases. 3D grids with horizontal dimensions of 1024 x 1024 were used to calculate nowcasts for these cases, and grids of half that size and centered within the 1024 x 1024 grid (with dimensions of 512 x 512) were used as validation. This partitioning is to allow for a fair comparison with the GFS-based advection method which requires establishing a 3D cloud grid first. This partitioning also enabled clouds present initially within the larger 1024 x 1024 grid to advect into the centered validation grid. To be clear, this is not updating the spatial resolution of the grid but rather relegating the validation grid domain to half the size to accommodate for clouds that advect into the validation domain.

Grid locations were randomly selected based on a central latitude and longitude point that was between -50° and 50° latitude and -100° and -40° longitude. Grid location selection was filtered based on the solar zenith angle for the four corners of the 1024 x 1024 grids for all potential nowcast valid times. Grids with any corner containing solar zenith angles greater than or equal to 85° (approaching the day/night terminator of 90°) were omitted to avoid any abrupt changes in CLAVR-x retrievals due to the lack of daytime visible imagery (i.e., algorithmic change) while still ensuring sufficient sample data coverage. Internal logic in CLAVR-x retrievals uses an 82° threshold to begin using NWP data for cloud water path, which is a main input for CBH retrievals.

However, grid selection filtering using the 1024 x 1024 grid avoided potential biases to the 512 x 512 validation grid.

2.2.2. Optical Flow Algorithm

The optical flow algorithm used in this study is the variational OF method available from the Optical flow Code for Tracking, AMV and Nowcasting Experiments (OCTANE; Apke 2023). Apke et al. (2022) explains that OCTANE resolves DOF $\boldsymbol{U} = [u, v]^T$ at every image pixel Ω located in space at $\boldsymbol{x} = [x, y]^T$ and in time t and $t + \Delta t$ by minimizing a penalty function $E(\boldsymbol{U})$ (see equation 2.1). The algorithm solves this minimization rapidly with GPU-based parallel processing (Sundaram et al. 2010) that variationally penalizes optical flow guess deviations from assumptions of brightness constancy (BC), gradient constancy (GC), and smoothness constraint (SC). The results of the processing are realistically smooth motions that track features just like human vision does, with two tunable constants (or hyperparameters) to control the importance of gradient constancy (λ) and smoothness (α) relative to brightness constancy. The algorithm minimizes the penalty function, equation 2.1, where the constraints are encapsulated in Charbonnier robust functions $\rho_d\,(r) = \rho_g(r) = \rho_s(r) = \sqrt{r^2 + \epsilon^2}$, by solving a system of linear equations iteratively with a coarse-to-fine optimization scheme. Further details of the OF algorithm are outlined by Apke et al. (2022).

$$E(\boldsymbol{U}) = \sum_{i,j \in \Omega} \left[\rho_d\left(BC_{i,j}\right) + \lambda\rho_g\left(GC_{i,j}\right) + \alpha\rho_s\left(SC_{i,j}\right)\right] \tag{2.1}$$

Full-disk 10.3 μm longwave infrared GOES-16 ABI imagery, which updates every 10 minutes, were used to retrieve OF fields. To calculate the OF field for a nowcast, both the latest image and the image valid 10 minutes prior are provided to OCTANE to retrieve an optical flow estimate. To optimize λ and α, 1-hour CTH nowcasts were calculated for systematically varied combinations

of λ and α across the year of 2022 of available full-disk 10.3 μm longwave infrared GOES-16 imagery starting at 15:00 UTC which ensured majority amount of the full-disk imagery was in daylight. To be clear, nowcasts used for optimization encompassed an entire full-disk as nowcasts needed to cover global domains and be generalizable to all latitudes and regions as necessitated by OVERCAST requirements. The root-mean square error (RMSE) between the CLAVR-x CTH data valid at 16:00 UTC and the 1-hour CTH nowcast was minimized to find the optimal combination. In instances where there was no value for CTH, whether for observed or forecast cloud, values were set to zero. Accounting for available GOES and CLAVR-x data, a total of 346 days of CTH data were utilized for optimization. RMSE was averaged across the 346 days and the λ and α associated with the lowest averaged RMSE value were utilized for the final nowcast case studies. Figure 2.2 shows the averaged RMSE in meters with respect to varying λ and α values. Optimized values of λ and α were determined to be 0.3 and 3.0, respectively.

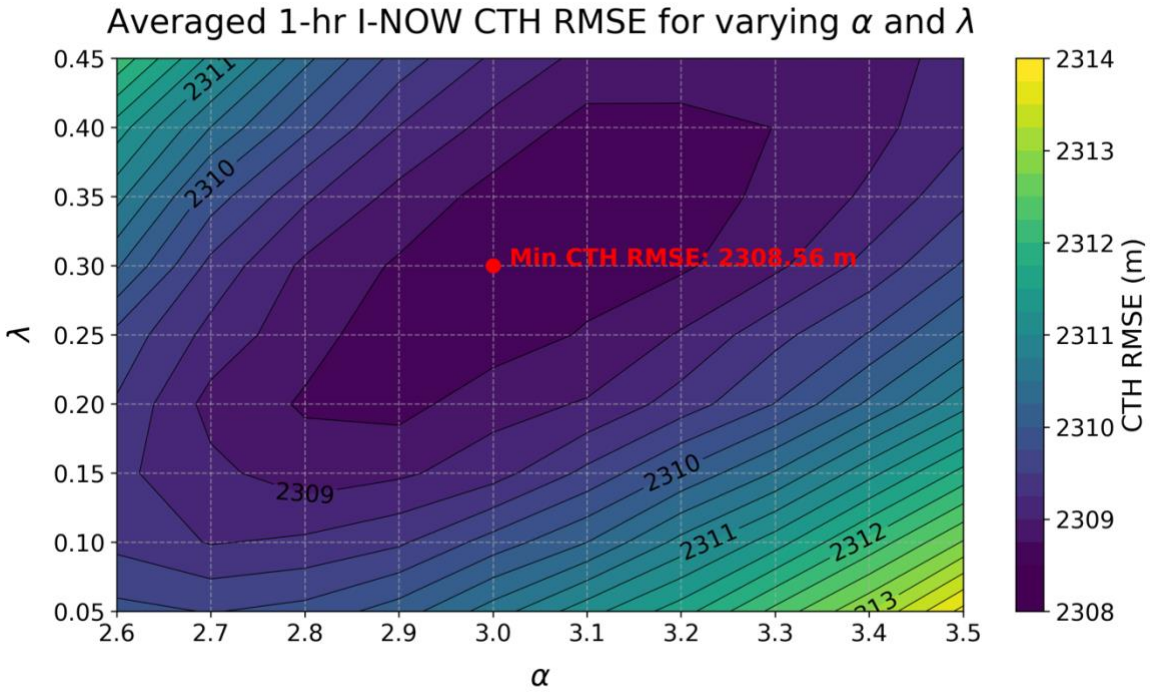

Figure 2.2: Averaged CTH RMSE in meters of 1-hour full-disk nowcasts across 346 days from 2022 with initial start time of 15:00 UTC using varying values for λ and α.

Despite optimizing λ and α values, the minimum RMSE remained over 2km. It may be argued that this value is excessively high (operational requirements in the United States often demand accuracy for CTH diagnostic products to within ~1 km after all; Heidinger et al. 2020). However, it is important to realize that the single value of RMSE also incorporates the error induced by cloud evolution, namely dissipation and formation. Consider the event where a high cloud dissipates, the remaining cloud will induce a pixel error equal to the difference between the height at 0 m above mean sea level (AMSL) and the height of the originally observed cloud. New high cloud formation, likewise, would result in a similar error magnitude. Additionally, cloud trajectory errors which create a slight displacement of nowcasted clouds from validation observations, result in a double penalty error in which clouds may have been correctly predicted, but placed in a slightly incorrect location. It is for this reason that validation includes an area-based validation metric which is discussed later in this section.

Table 2.1 provides OCTANE specific settings used in this study. These settings follow values used by Apke et al. (2022) with the exception of λ, α, and normalization brightness ranges. Normalization maximum and minimum brightness were set to 185.5698 and -1.6443 $mW\ m^{-2}\ sr^{-1}\ (cm^{-1})^{-1}$ respectively which are based on the maximum and minimum radiance values in full-disk 10.3 μm longwave infrared GOES ABI imagery.

Table 2.1: Settings for OCTANE OF algorithm.

OCTANE Algorithm Settings	
Parameter	Value
Coarse-to-fine pyramid levels	4
Inner iterations	3
Conjugate gradient iterations	30
Pyramid scale factor	0.5
λ (gradient constancy constant)	0.3
α (smoothness constraint constant)	3
Normalization max brightness	185.5698
Normalization min brightness	-1.6443

2.2.3. Warping and Nowcasting Method

Although much past work has improved OF algorithm accuracy, it is equally important in satellite and cloud nowcasting applications of OF to account for time-related changes of the piecewise flow field. The I-NOW method is the combination of the OCTANE OF retrieval algorithm with a forward extrapolation approach adapted from the frame interpolation (or "warping") method of Baker et al. (2011). The warping method is unique in the nowcasting literature, so we explain in some detail the four-step process that is implemented on a retrieved OF field $\mathbf{u}_o(x)$:

1) First, we forward-warp the flow $\mathbf{u}_o$ to time t to give $\mathbf{u}_t$where:

$$\mathbf{u}_t\left(\text{round}\left(x + t\mathbf{u}_0(x)\right)\right) = \mathbf{u}_o(x) \tag{2.2}$$

The round operator in equation 2.2 ensures that pixel displacements are calculated to specific pixels rather than between pixels. This results in a new OF field that is valid at time t based on the displacement provided by the initial OF field. However, the new OF field will have pixels that either do not have a OF field value, described here as "holes", or pixels multiple OF field values, described here as "occlusions".

2) Fill any holes in $\mathbf{u}_t$ using an outside-in strategy that requires looping through the flow field array and filling values based on the average of valid pixels surrounding holes. This process is done iteratively until all holes in the flow field array are filled.
3) Account for occlusions: Because forward warping the flow of a divergent field can result in multiple values arriving at the same pixel, an appropriate flow value must be selected. Satellite imagery makes this determination tractable, as radiometrically colder pixels or higher CLAVR-x height assignments can be assumed to be clouds residing at higher altitudes, and are thus selected to be the pixels to be retained at forward-warped locations. If nowcasting CLAVR-x data (such as CTH or CBH), flow from pixels with the higher CTH were retained at forward-warped locations (note: memory of lower cloud pixels is not retained). In the event imagery is nowcasted, pixels that are radiometrically cooler are retained at forward-warped locations. Note that this differs from the occlusion reasoning in Baker et al. (2011) which was OF-based as there was a lack of information on layer context in their imagery.
4) Interpolate pixels from the latest observed image to the new location: For nowcasting satellite imagery this interpolation can be done via the bilinear method based on the negative of the displacement resulting from $\mathbf{u}_t$ such that $x_0 = x - t\ \mathbf{u}_t(x)$ where x_0 is the location of the pixels that will be advected forward from the latest observed image. For discrete

fields such as CTH and CBH in this study, nearest neighbor interpolation can be used to ensure numerical dispersion is limited which would be represented as $x_0 = \text{round}(x - t\,\mathbf{u}_t(x))$.

This method is repeated with step 1 updating the previously updated flow field $\mathbf{u}_t$ and step 4 always referencing the x_0 location in the last observed image, not the pixels of the latest nowcast iteration. Always referencing the last observed image avoids the problem of stationary/locked pixels where the magnitude of displacements for a single time step are insufficient to move pixels beyond their original pixel location. Thus, for the second iteration of this method, x_0 in step 4 becomes $x_0 = x - 2t\,\mathbf{u}_t(x)$. Since full-disk images are available every 10 minutes, nowcast calculations using this warping method were made with 10-minute increments out to 3 hours. Ultimately, rather than assuming a fixed flow field with respect to time as utilized in past approaches, this method better handles the piecewise nature of the satellite imagery OF field and its time-related changes.

2.2.4. Imagery Nowcasts and ML Comparison

OVERCAST mandates unique DoD requirements for nowcasting which makes it necessary to predict CLAVR-x cloud property retrievals rather than imagery. However, because I-NOW is utilizing an image interpolation technique that was originally based on creating new images, the algorithm can easily be used to create predicted infrared brightness temperature imagery (e.g. Figure 2.3). I-NOW uses a Lagrangian persistence assumption such that pixels do not change or evolve but are rather simply translated to a new location. The one exception is that divergence captured in the optical flow retrieval allows for expansion of cloud tops in subsequent nowcast predictions via Step 2 in Section 2.2.3. Evidence of this expansion behavior can be found in the larger convective clouds (e.g. green encircled areas in Figure 2.3d) over the tropic latitudes over

the eastern Pacific, South America, and over the Atlantic just east of Brazil. Another important detail is that the assumption of Lagrangian persistence is quite apparent over South America (near 20° S latitude) where clear-sky brightness temperatures remain warm over the land surface (e.g. blue encircled area in Figure 2.3c and Figure 2.3d) when compared to the cooler brightness temperatures observed at 21:00 UTC. Additionally, smaller areas of unpredicted cloud dissipation (e.g. purple encircled areas in Figure 2.3c and Figure 2.3d) and unpredicted cloud formation (e.g. red encircled areas in Figure 2.3c and Figure 2.3d) are also evident of the Lagrangian persistence assumption.

Overall, I-NOW satisfies the top-line objective of OVERCAST to compute a 3D cloud nowcast that can be run on full-disk imagery at a cadence shorter than the 10-min iterations of full-disk imagery observations (i.e. keeping up with real-time processing). To assess efficiency, computation of I-NOW was timed based on the 100 full-disk nowcast examples predicting imagery. The average computational time for deriving OF from two full-disk images and performing the forward warping to create the arrays required for plotting the full-disk image predictions was 2.43 minutes. Computation was executed on a high-performance Rocky Linux 9.3 Blue Onyx machine equipped with Dual Intel® Xeon® Gold 6348 processors and 2.0 TiB of RAM. OF computation was completed using only a fraction (~8.35 GB) of the available memory on one NVIDIA RTX A6000 GPU.

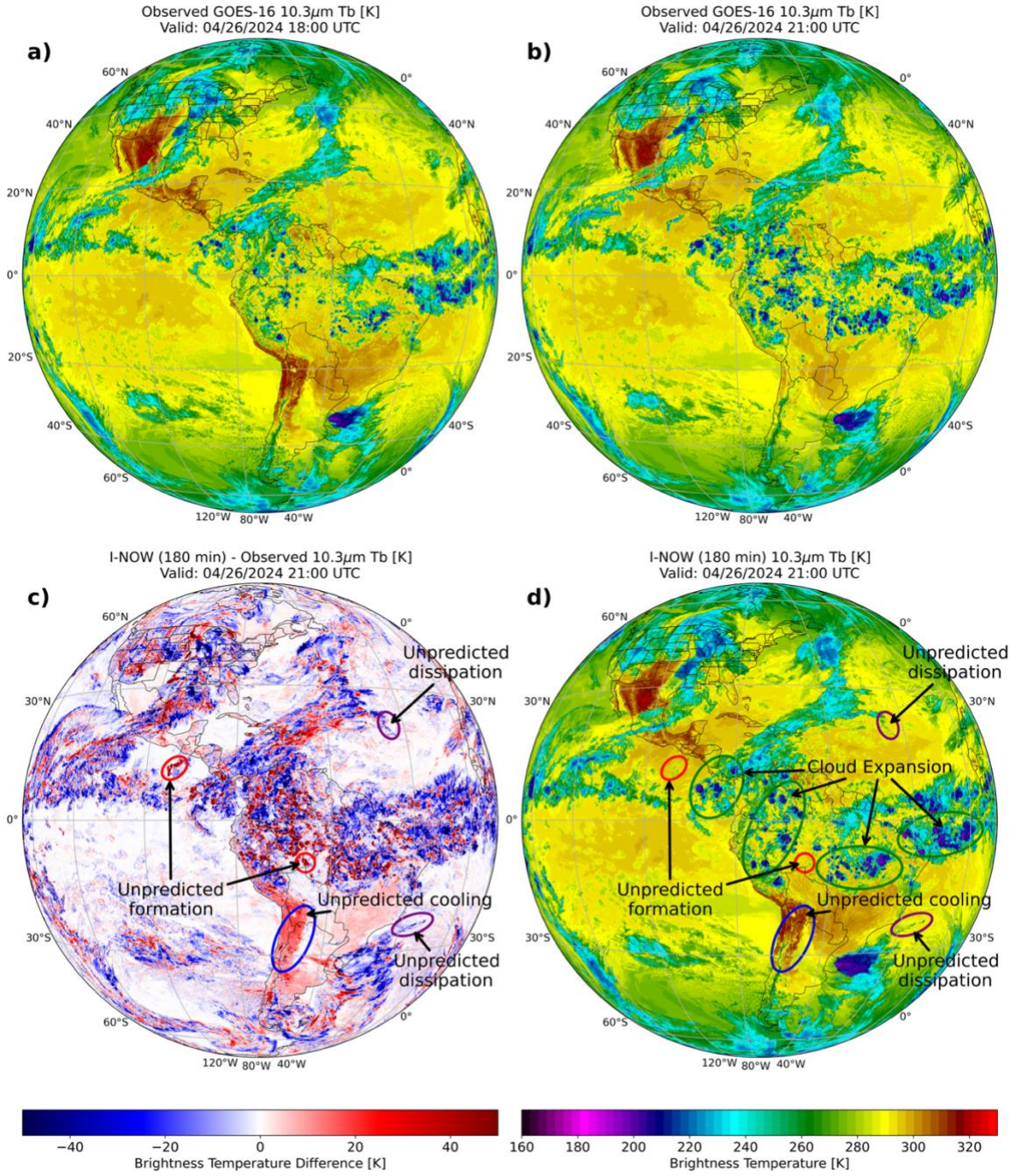

Figure 2.3: a) Observed GOES-16 10.3 μm brightness temperatures (Tb) valid at an initial nowcast time of 18:00 UTC on 26 April 2024; b) Observed GOES-16 10.3 μm brightness temperatures valid at 21:00 UTC on 26 April 2024; c) Difference of I-NOW 3-hour nowcast and observed imagery of GOES-16 10.3 μm brightness temperatures valid at 21:00 UTC on 26 April 2024 with annotated areas of unpredicted dissipation (purple), formation (red), and cooling (blue); d) I-NOW 3-hour nowcast valid at 21:00 UTC on 26 April 2024 derived only from images valid at 17:50 UTC and 18:00 UTC with annotated areas of unpredicted dissipation, formation, and cooling as well as areas of cloud top expansion (green) due to the retrieved divergence in the OF retrieval.

Much of the published work using ML approaches address satellite-based nowcasting within the context of new image prediction in form of a single variable such as an imagery channel (Kellerhals et al. 2022), cloud cover (Berthomier et al. 2020), or cloud top height (Yu et al. 2023). Often ML studies provide OF-based baseline approaches that fail to outperform persistence. Extrapolated Imagery or EXIM (Jann 2017) is an operational satellite nowcasting algorithm that is used operationally by EUMETSAT's Satellite Application Facilities and was used by Berthomier et al. (2020) as a comparison to a U-Net predicting cloud cover over France. Berthomier et al. (2020) showed that EXIM failed to outperform persistence based on average mean square error metrics for cloud cover beyond 60 minutes with a U-Net providing considerable outperformance. Similarly, Yu et al. (2023) used a convolutional-long-short-term-memory (ConvLSTM) and trajectory-gated-recurrent-unit (TrajGRU) ML models and found their baseline OF approach could not outperform persistence beyond 40 minutes. Thus, evaluations against persistence continue to be an important benchmark for contemporary cloud nowcasting research.

Because OF-based approaches are often dismissed for their lack of robustness, even when compared to simple persistence, we include here a brief aside to show a comparison of I-NOW brightness temperature predictions against a baseline U-Net approach outlined by Chase et al. (2025). Chase et al. (2025) provides the framework for implementing diffusion nowcasting for brightness temperatures and uses a U-Net as a comparison baseline to various diffusion models — generative models that iteratively refine random noise into structured outputs by reversing a gradual noising process — primarily for single-variable, infrared brightness temperature imagery nowcasting. We use this same U-Net which was trained on GOES-16 brightness temperature data from 2023 to make our comparison of I-NOW brightness temperature nowcasts.

To conduct the experiment, 1000 nowcast cases from 2024 were selected at random by selecting 100 nowcast initial timeframes from GOES-16 full-disk imagery and then selecting 10 (512 x 512) rectilinear grids of 0.02° latitude and longitude resolution from each of those full-disk cases. This gave a total of 1000 nowcast cases to make the comparison between the U-Net and I-NOW, however, because of missing GOES-16 full-disk imagery for 3 nowcast timeframes, 30 cases were omitted from final skill calculations which resulted in a final 970 total cases.

Figure 2.4 shows the best U-Net prediction when compared to I-NOW out of the 970 nowcasts examples from the year 2024 using GOES-16 10.3 μm IR brightness temperature imagery. Two details are evident in the example. First, I-NOW shows clearly that it is not designed for changes in brightness temperature values but rather shifts locations of brightness temperature pixels based on the initial apparent motion observed between the 17:30 and 17:40 UTC images. The U-Net on the other hand provides changing brightness temperatures but has smoother predictions (i.e. loss of spatial sharpness) with each nowcast iteration that are detrimental to CFLOS computation.

Figure 2.5 provides the RMSE validation for the prediction comparison and shows that the U-Net outperforms I-NOW using RMSE for the entirety of this example with I-NOW only showing outperformance of persistence to the 160-minute valid time. Despite the best-case example for the U-Net, Figure 2.6 shows that when validating all 970 nowcasts, I-NOW outperforms on average both persistence and the U-Net at the 3-hour valid time. The U-Net does show slight outperformance against I-NOW in RMSE until 110 minutes, but this improvement comes at the cost of smoother cloud fields and potential ML hallucinations — outputs generated by a model that appear plausible but are not grounded in the actual input data or reality — which can be difficult to explain. The most important fact, however, is that unlike past studies which include an OF-based nowcast baseline that struggle to outperform persistence within the first hour, I-NOW

shows that when temporal changes to the OF field are accounted for and estimated into the future, nowcast accuracy can meet and exceed the capabilities of newer and more complex ML approaches as well as persistence.

To be clear, the goal of this section is not to dismiss ML approaches for satellite-based cloud nowcasting. On the contrary, ML methods such as diffusion models hold great potential to revolutionize cloud nowcasting capabilities. However, DoD operations often require timely solutions in data sparse situations or data denied environments. Specifically for OVERCAST, requirements of high cadence nowcasts over large domains based on explainable approaches currently prohibits the use of more advanced diffusion-based methods as in Chase et al. (2025). Here, it is demonstrated that an observational approach based on OF should not so quickly be dismissed but instead should be used alongside ML methods.

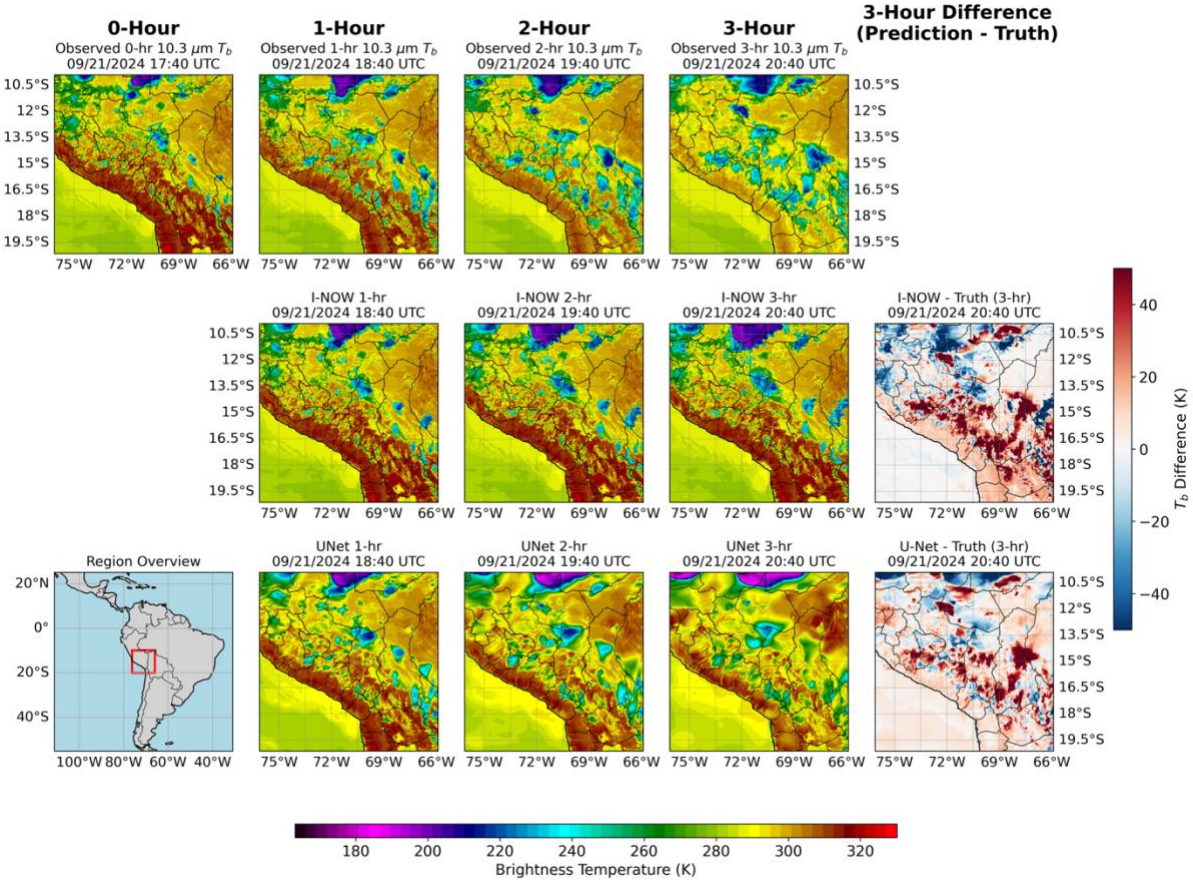

Figure 2.4: Best example of U-Net outperformance compared to I-NOW for GOES-16 *10.*3 μm IR brightness temperature imagery. Nowcasts from both I-NOW and the U-Net are displayed for 1-hour, 2-hour, and 3-hour valid times as well as the difference plots from the observed 3-hour truth and 3-hour nowcasts valid at 20:40 UTC on 21 Sept 2024.

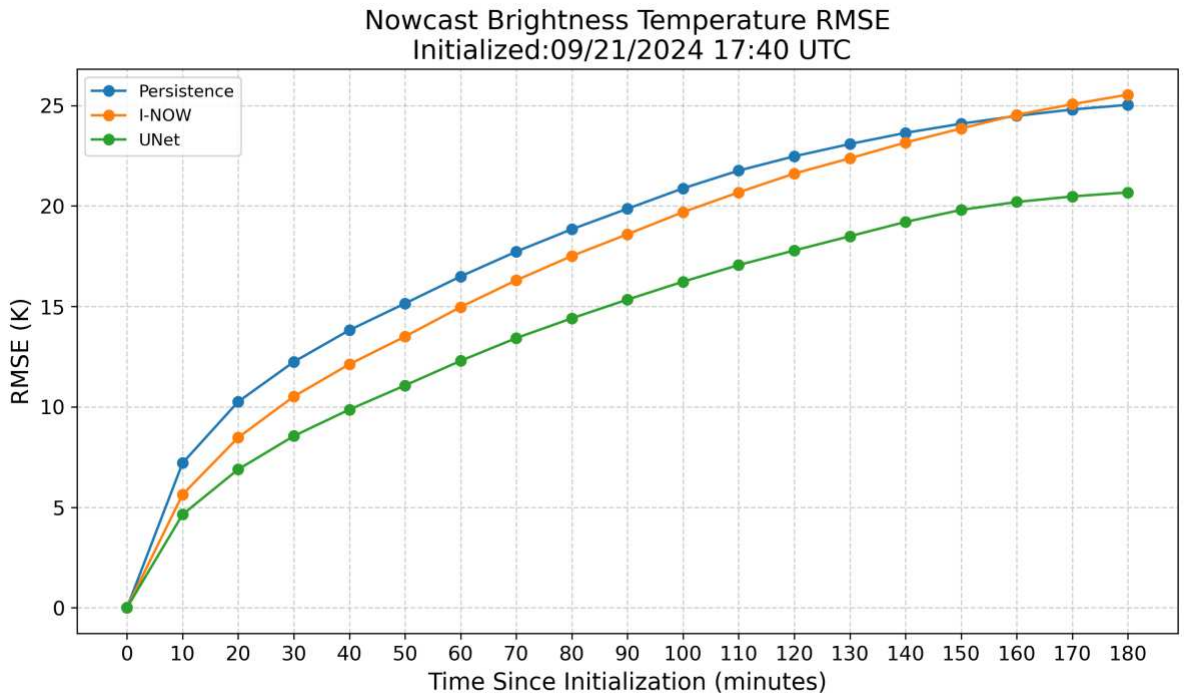

Figure 2.5: RMSE validation on for 21 Sept 2024 best example of U-Net brightness temperature prediction vs I-NOW.

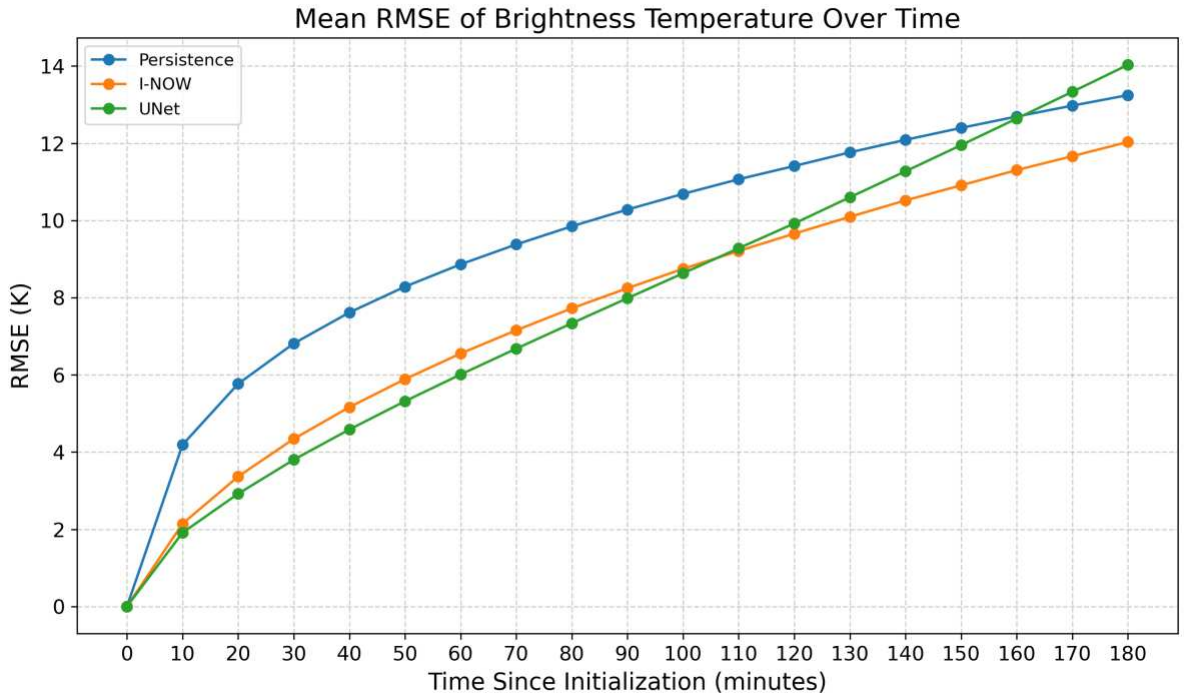

Figure 2.6: Mean RMSE results for all 970 brightness temperature nowcasts.

2.2.5. Baseline Methods

Three different baseline methods were applied to quantify the effective skill of I-NOW 3D cloud nowcasts. The first and simplest was the above-mentioned method of Eulerian persistence in which the initial 3D cloud field is assumed to be the forecasted field. Although a crude and simplistic assumption, persistence is often difficult to outperform in very short timeframes as clouds often revisit locations of previous clouds in nature, and slight shifts in the forecast cloud field can produce large errors when computing statistics on the pixel scale (Guillot et al. 2012). Additionally, considering that past studies have indicated that typical OF-based cloud nowcasting approaches struggle to outperform persistence (Berthomier et al. 2020; Yu et al. 2023), it is appropriate to include persistence as a baseline method for comparison.

The second baseline is the OF nowcast *without* the forward warping method, mirroring the commonly used approach with OF where the motion field is assumed to be static with respect to time and a linear forward extrapolation method is used to advect Lagrangian persistent observations forward. Using this method allowed for a comparison of the added skill that the forward warping method of I-NOW provides within the context of the piecewise OF field retrieved from satellite imagery.

The third baseline method is a simple advection method based on GFS model wind data. Using the 3D cloud grid previously described, horizontal advection at each pressure level is calculated using a semi-Lagrangian method similar to that detailed by Gitro et al. (2018) for the Advected Layer Precipitable Water (ALPW) product. Winds were linearly interpolated from NOAA's archived GFS model data across space and time. This third baseline method, although computationally expensive, allows for clouds at varying heights to advect in different directions which provides a reasonable comparison to the optical flow method which assumes CBH to advect

with the same flow as CTH. We select these rather simple baselines for comparison as they represent where current DoD capabilities stand to produce such high-resolution CFLOS nowcasts (though no operations currently produce them to the standards pursed by the OVERCAST project).

Figure 2.7 provides an overview of how each nowcast method, including I-NOW and the selected baselines, construct 3D cloud field nowcasts from CLAVR-x retrievals of CTH and CBH. Each method results in a 3D cloud field valid at 10-minute increments that can be validated against a 3D cloud field constructed from observed CLAVR-x retrievals.

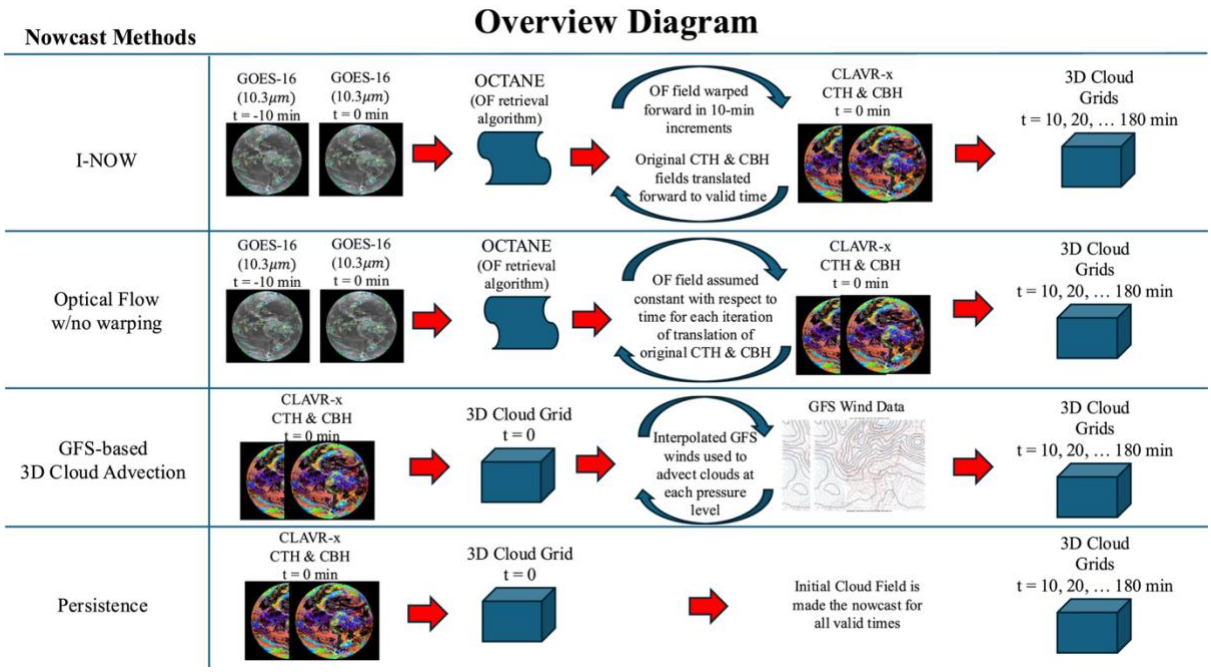

Figure 2.7: Overview diagram illustrating the how each nowcast method employs various data to produce 3D cloud grids.

2.2.6. Validation Metrics

As previously discussed in the subsection on cloud data, CLAVR-x retrievals of CTH and CBH were utilized to create 3D cloud grids using GFS pressure levels for the vertical coordinate. Thus, any nowcast method tested in this study supplies a 3D cloud grid in which cloud is designated as a 1 and no cloud is designated as 0 for each GFS pressure level. This binary assignment allowed for a comparative analysis of all methods using metrics that required a binary deterministic forecast for each grid point within a 3D gridded domain.

Four skill metrics were used to evaluate nowcast skill: 1) Probability of Detection (POD), 2) Critical Success Index (CSI), 3) False Alarm Ratio (FAR), and 4) Fraction Skill Score (FSS). The POD, CSI, and FAR metrics, as defined by Wilks (2011), use a 2x2 contingency table composed of parameters (*A, B, C, D*), where *A* is the number of forecast events (i.e., cloudy pixels) which were observed, *B* is the number of forecast events which were not observed, *C* is the number of forecast non-events (i.e., clear sky pixels) which were observed, and *D* is the number of forecast non-events which were not observed (not used here). The POD, CSI, and FAR metrics are defined as:

$$POD = \frac{A}{(A+C)} \quad (2.3) \qquad CSI = \frac{A}{(A+B+C)} \quad (2.4) \qquad FAR = \frac{B}{(A+B)} \quad (2.5)$$

Because POD, CSI, and FAR are interrelated, they are often plotted using a single performance diagram that allows for simultaneous depiction of all three metrics in a single diagram (see Figure 2.8). A subtle change in the metrics is that FAR is depicted as Success Ratio (SR) which is simply $1 - FAR$ where a SR value of 1 would denotes that all cloudy forecasts were correct and a value of 0 would denote no cloudy forecasts were correct. Additionally, a performance diagram allows for the depiction of forecast bias, which is simply the ratio of POD to SR in which a value of 1

denotes that the POD is the same as SR indicating neither over forecasting (value greater than 1; or high-bias) or under forecasting (value lower than 1; or low-bias).

The FSS, as defined by Roberts and Lean (2008), was initially used for validating gridded precipitation forecasts. Their reason for using FSS over traditional skill scores like POD, CSI, and FAR was to avoid the so-called "*double penalty phenomenon*" where a forecast is penalized twice for having a correct forecast at a slightly misplaced location. FSS is specifically beneficial for this study as the metric describes accuracy of the cloud fraction over a given area which is directly related to the accuracy of a CFLOS nowcast. FSS requires binary fields of observed and forecast data which are calculated based on a chosen threshold. Conveniently, grids in this study are already binary with 0 defined as "no cloud" and 1 defined as "cloud." FSS then involves generating fractions from those binary fields. For every grid point, a fraction of surrounding points within a given square of length n that have a value of 1 is calculated. This is described in the equations 2.6 and 2.7:

$$O_{(n)}(i,j) = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n} I_o\left[i+k-1-\frac{(n-1)}{2}, j+l-1-\frac{(n-1)}{2}\right] \qquad (2.6)$$

$$F_{(n)}(i,j) = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n} I_f\left[i+k-1-\frac{(n-1)}{2}, j+l-1-\frac{(n-1)}{2}\right] \qquad (2.7)$$

in which $O_n(i,j)$ is the resultant field of observed fractions for a square of length n obtained from the observed binary field I_o and $F_{(n)}(i,j)$ is the resultant field of forecast fractions obtained from the forecasted binary field I_f. The value of i goes from 1 to N_x, which is the number of columns in the grid, and the value of j goes from 1 to N_y, which is the number of rows.

Next, the mean square error (MSE) and a reference (MSE_{ref}) were calculated from the fractions. According to Roberts and Lean (2008) the MSE_{ref} can be thought as the largest possible MSE that

can be obtained from the forecast and observed fractions. MSE and MSE_{ref} were calculated by the following:

$$MSE_{(n)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[O_{(n)i,j} - F_{(n)i,j} \right]^2 \tag{2.8}$$

$$MSE_{(n)ref} = \frac{1}{N_x N_y} \left[\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} O_{(n)i,j}^2 + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} F_{(n)i,j}^2 \right] \tag{2.9}$$

Finally, the FSS was calculated from equations 2.8 and 2.9 as:

$$FSS_{(n)} = \frac{MSE_{(n)} - MSE_{(n)ref}}{MSE_{(n)perfect} - MSE_{(n)ref}} = 1 - \frac{MSE_{(n)}}{MSE_{(n)ref}} \tag{2.10}$$

In this study, a 3x3 sliding window (0.06° x 0.06°, or ~6 x 6 km) was used to calculate FSS over the 512 x 512 rectilinear grid defined for validation. To be clear, the box slides by 1 pixel so that each grid point has a calculated fraction. For instances when the 3x3 sliding window overlapped beyond the edge of the validation grid, values within the sliding window that fell outside the grid were set to zero.

For all four skill metrics, nowcast performance was evaluated on each vertical level in the validation domain so that each skill score depicted the skill for the entire 3D validation grid.

2.3. Results and Discussion

2.3.1. Overall Results

Figure 2.8 shows the averaged 3D cloud validation results for all 991 nowcasts from 2024 where the left panel is the performance diagram and the right panel is the plot of FSS. In the performance diagram a perfect nowcast would have a point at the top-right corner. Thus, nowcasts that are closer to the top-right corner are an objectively better prediction. Furthermore, points along the

diagonal have minimal bias (i.e. neither over- nor underpredict). To reduce clutter in the performance diagram, only the 1-hour, 2-hour, and 3-hour data points for each method are included.

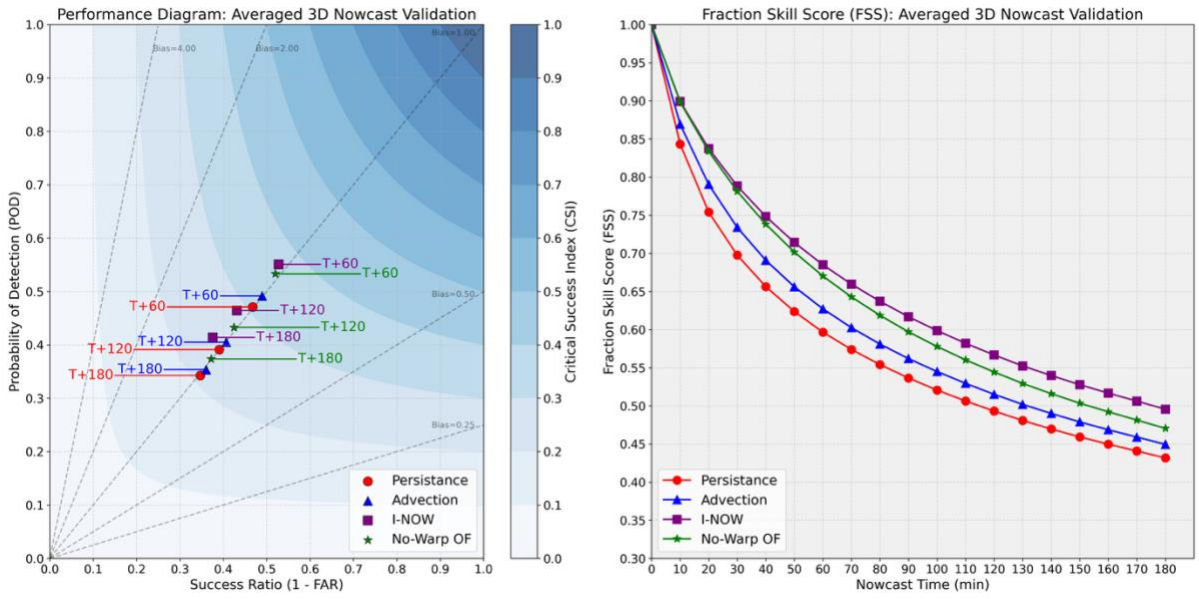

Figure 2.8: Averaged 3D nowcast validation scores across nowcast increments for all 991 randomly selected nowcasts from 2024 using persistence (red), advection (blue), I-NOW (purple) and no-warp OF (green) methods. The performance diagram (left) provides the combination of POD, CSI, and SR (1 – FAR) where perfect scores for POD, CSI, and SR are 1 and no skill is 0. Therefore, an ideal prediction on the performance diagram would be a data point in the top right corner where POD, CSI, and SR would all equal 1. The plot for FSS (right) provides the skill score for each nowcast prediction timeframe out to 3 hours. A perfect score for FSS is 1 while a no skill score is 0.

The I-NOW method maintained superior averaged skill for POD, CSI, and FSS throughout all nowcasting timeframes from 10 minutes out to 3 hours. Such improvements are, in fact, statistically significant ($p << 0.001$) when evaluated with a Wilcoxon signed-rank test (e.g. Rosner et al. 2006). Further, CSI and FSS improvements between I-NOW and no-warp methods increased with time, indicating that the warping method indeed accounts for missed time-related changes to the flow field using Lagrangian persistence.

In contrast, the SR is slightly lower (higher FAR) at the end of the nowcast in the I-NOW method vs the no-warp method which is indicated by the increase in bias of I-NOW from the 1-hour, 2-hour, and 3-hour nowcast timeframes. We postulate this lower SR relates to cloud dissipation, which occurs in nature but was not parameterized in this version of the I-NOW method. In the no-warp method, clouds get removed along convergence boundaries near sharp motion discontinuities. Because the flow field for the no-warp OF method remains constant, clouds that initially flow in the direction of discontinuities merge and subsequently create artificial cloud dissipation. This unintended dissipation does cause substantial decrease in POD. Hence, despite the relatively higher bias and lower SR, the I-NOW method produces the highest overall skill.

Interestingly, all the computational nowcast methods outperformed persistence in POD, CSI, and FSS metrics, with the computationally more expensive advection method performing worse than the OF methods. The advection method likely performs poorly because the accuracy of GFS winds is lower compared to OF retrievals, which are obtained right at the beginning of the nowcast period. This finding is significant, since the OF-based methods are not only more accurate representations of the actual cloud motions but are also computationally more efficient than the NWP-based advection method which requires advection calculations at multiple levels. Consequently, the efficiency of the OF-based methods enables near real-time computation over significantly larger and finer-scale domains.

Often in cloud nowcasting, validation is performed based on the 2D results of data rather than 3D cloud fields (e.g. Jann 2017; Berthomier et al. 2020). Thus, for consistency with past research, the same validation metrics were calculated for the resulting 2D cloud mask of the nowcasts. Figure 2.9 presents the results for cloud mask validation, showing even better performance of the I-NOW method for POD, CSI, and FSS when compared against with the baseline methods. Again,

we observe the bias increase with nowcast timeframe for I-NOW; however, I-NOW maintains higher skill (further to the top right of the performance diagram) relative to the same nowcast timeframes for all other baselines methods. The no-warp OF method maintained a bias closer to 1, but failed to exceed I-NOW in POD, CSI, or FSS. This performance is once again indicative of the no-warp OF method's tendency to reduce clouds in areas where there are flow discontinuities. Again, because the flow field for the no-warp OF method remains constant, clouds that initially flow in the direction of discontinuities merge and subsequently create artificial cloud dissipation. Like in the 3D validation, the POD, CSI, and FSS indicate that I-NOW outperforms all other techniques out to three hours.

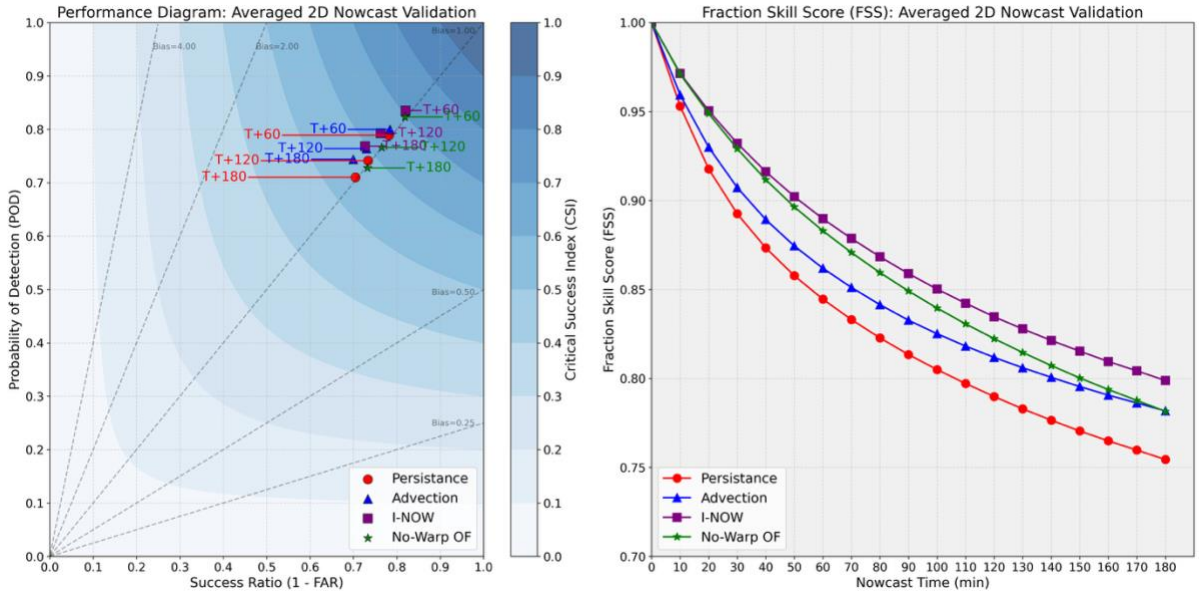

Figure 2.9: Same as Figure 2.8 now for cloud mask validation scores.

To determine if the outperformance of I-NOW was statistically significant when compared to the no-warp OF method, a Wilcoxon signed-rank test was conducted for the differences between the 3-hour metrics of POD, CSI, FAR, and FSS. All p-values for 3D and 2D validation scores were less than 0.01 except for the 2D FAR metric. Thus, apart from the cloud mask FAR metric,

I-NOW's superior performance is statistically significant when compared to a method that utilizes only linear extrapolation based on a constant flow field. To further demonstrate why this performance is better, in sections to follow we proceed to show each method applied to independent and illustrative case studies from 2023.

2.3.2. Case Study 1 – (16 January 2023, 16:00 UTC)

Figure 2.10 shows the cloud nowcasting results for the first nowcast case study, from 16 January 2023. This case study lies over the Pacific Ocean, off the west of the coast of Chile. The plots were made from the 512 x 512 verification grids with colors indicating the pressure level of the cloud top. Warmer colors (green and yellow) indicate lower clouds, while colder colors (blue) indicate higher clouds. Although the nowcasts were computed and verified in 10-minute increments, only the 1-hour, 2-hour, and 3-hour nowcasts are provided for the sake of brevity. Additionally, for the ease of comparisons, the difference of each prediction method and the truth valid at the 3-hour prediction timeframe are also provided where red indicates a higher-pressure value from the prediction compared to the truth (lower cloud prediction) and blue indicates a lower pressure value from the prediction compared to the truth (higher cloud prediction).

The most prominent feature of the observations (shown in the top row of Figure 2.10) is the distribution of the high clouds that move from the southwest to the northeast towards the center of the domain. Both the advection method (2nd row) and the I-NOW method (3rd row) appear to handle this cloud movement well in that the predictions retain the structure that was observed at their respective validation times. The no-warp OF method (bottom row), however, does not retain the same structure, especially as the forecast period increases. This drop-off in performance is due to the merging of clouds in the areas of flow discontinuities.

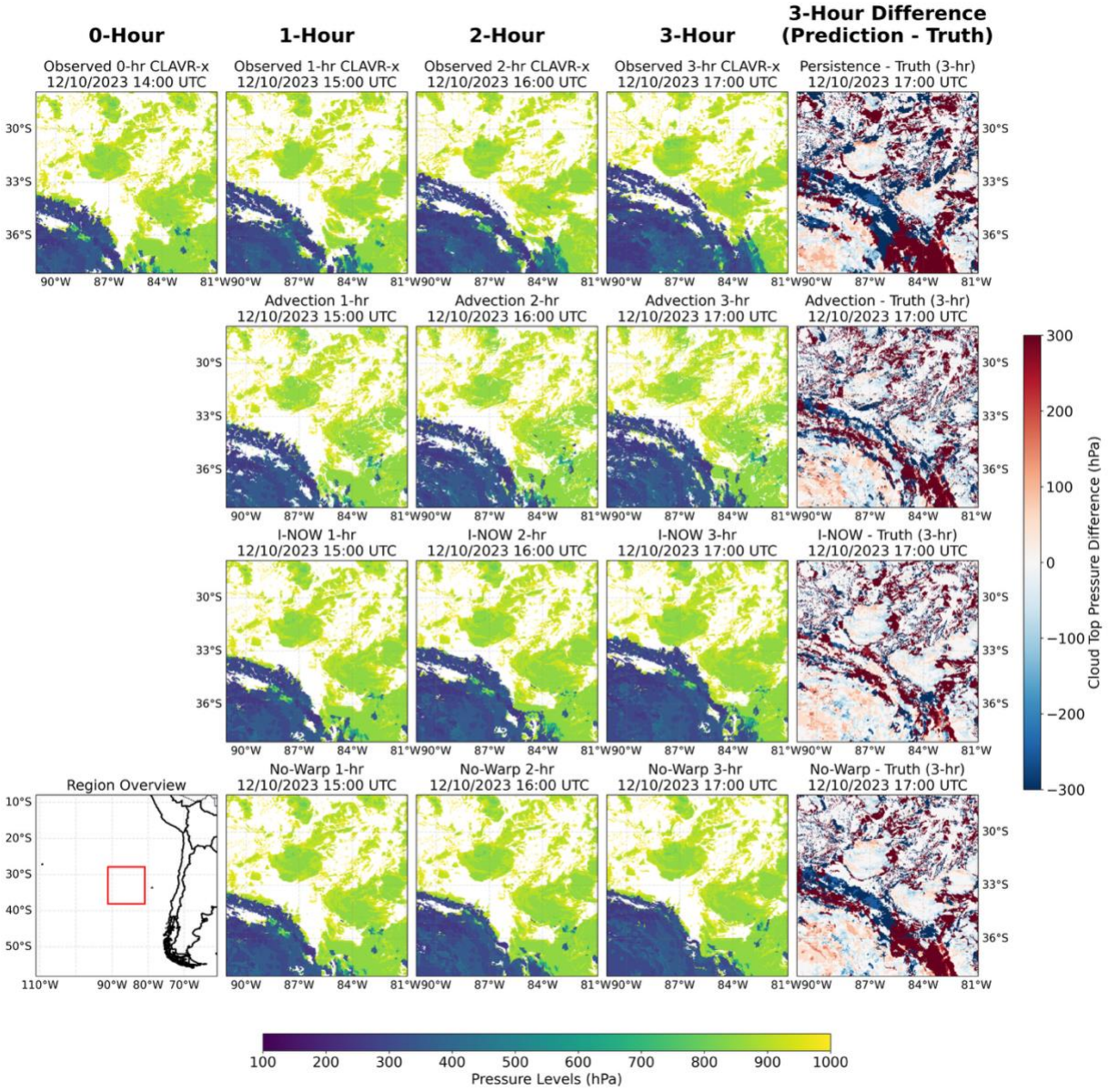

Figure 2.10: 16 January 2023, 16:00 UTC case study over the open Pacific, west of the coast of Chile. The domain is roughly 1000 km on each side. The persistence method can be seen in the truth plot in the upper-left corner with the rest of the row indicating the 1-hour, 2-hour, and 3-hour observed state of the CLAVR-x cloud top pressures as indicated in the bottom color bar. The GFS-based wind advection method is the second row, the I-NOW method is the third row, and the no-warped OF method is the last row. The difference between each prediction method and the truth observed at the 3-hour prediction timeframe is shown in the last column on the right with the right-side color bar indicating the difference in cloud top pressure between the prediction and truth.

Another result to note is from the GFS-based wind advection method (2nd row). Starting in the 1-hour plot and continuing into the subsequent predictions, cloud-free areas or cloud gaps begin to appear in the 2D forecast cloud field. These gaps are due to the different wind velocities, with respect to the vertical pressure coordinate, that are applied to the 3D cloud field. Clouds that extend through multiple vertical levels are advected independently at each level using winds that are acquired from the respective GFS pressure level which ultimately causes cloud structures to be pulled apart. This is not observed in the actual CLAVR-x observations nor are they observed in OF-based methods due to the assumption that CBHs will move in lock-step with the motion of their corresponding CTHs.

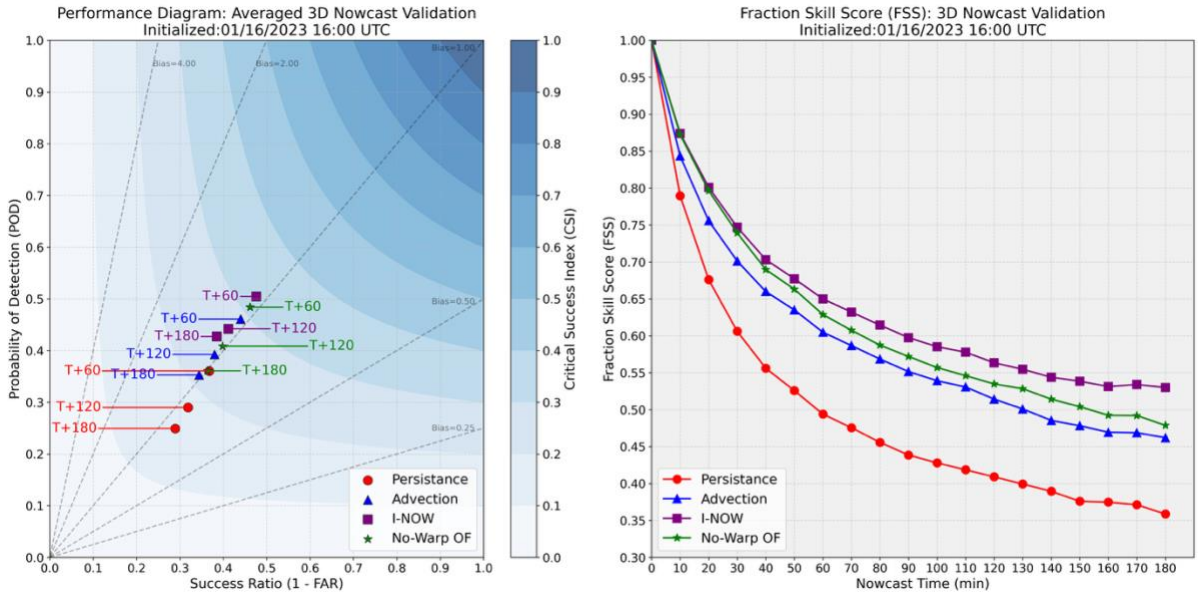

Figure 2.11: 3D validation results for 16 January 2023, 16:00 UTC case study.

The 3D validation for this case study is presented in Figure 2.11. The results show superior I-NOW performance over the other methods across all metrics (POD, CSI, and FSS). Similar to previous results, I-NOW's bias again increased slightly at later prediction timeframes.

Noteworthy, however, is the significant superior forecasting skill for I-NOW in comparison to all the other methods with increasing prediction timeframes across all other measures of statistical performance. This performance advantage for I-NOW is particularly evident in the FSS plot.

The straight-forward cloud mask validation (following previous study benchmarks) is provided in Figure 2.12. Here the I-NOW method significantly outperforms the other baselines out to 3 hours for the POD, CSI, and FSS metrics. Again, bias is slightly elevated for I-NOW when compared with the other methods which reinforces the likelihood that the lack of cloud dissipation is a key contributor to reduced skill in the current version of I-NOW. Despite this shortcoming, the dramatically improved POD, CSI, and FSS throughout the 3-hour nowcast, especially in comparison to persistence, gives cause for optimism that I-NOW may be further innovated to overcome its limitations in high-bias and produce an overall superior nowcast.

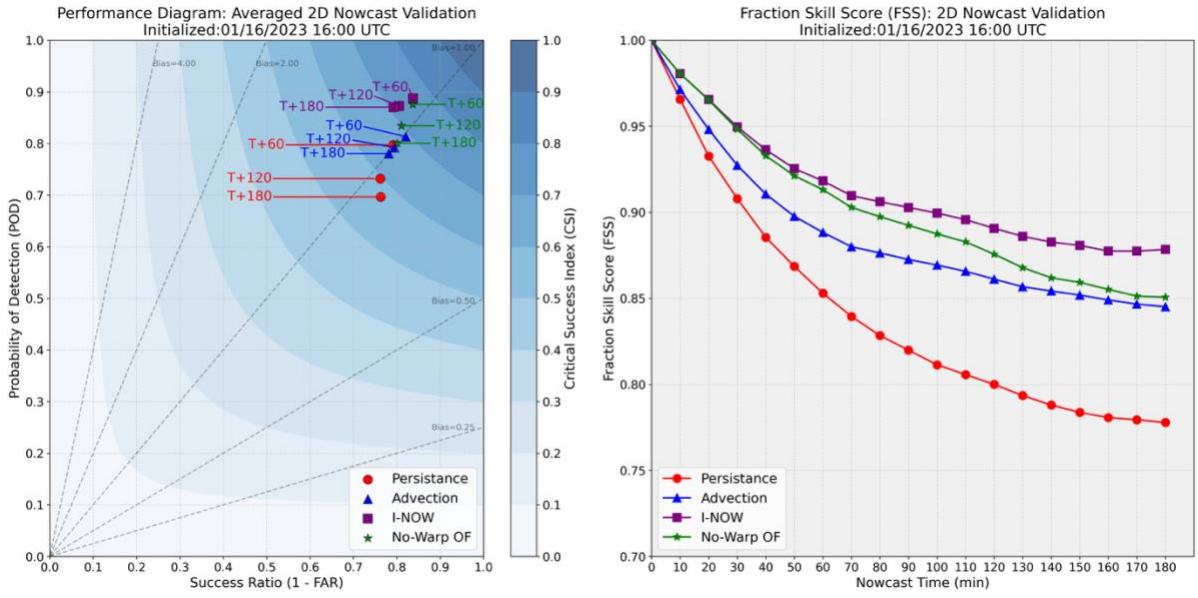

Figure 2.12: Cloud mask validation results for 16 January 2023, 16:00 UTC case study.

2.3.3. Case Study 2 – (15 July 2023, 17:00 UTC)

The second case study covers part of northern Brazil and southern Venezuela, offering a good opportunity to evaluate performance in the face of significant convective activity as well as cloud dissipation. Results for this case are shown in Figure 2.13. Low-level cumulus clouds in the initial conditions appear as scattered areas of green, while higher-level cloud tops from convective activity appear as dark blue. The hourly observations (top row) show that much of the low-level cumulus dissipates as time progresses while convective development grows much of the higher cloud tops. Due to the Lagrangian persistence assumptions in the nowcast methods, none of the low-level clouds are dissipated in the predictions. However, I-NOW (3rd row) does show some divergence-related spread in the convective areas as indicative of the "growing" areas of the dark blue cloud tops. This nowcast-based cloud growth should not be described as true cloud formation, however, as the growth is not due to the creation of new cloud in the prediction, but rather is the stretching or warping of clouds that are already present based on the divergence of the OF retrievals. While the cloud may spread, any vertical development in the cloud-top heights would not be captured.

Another significant performance insight gleaned from this case study is the poor quality of results from the GFS-based wind advection method (2nd row). The assumption that clouds at different levels will advect at the wind speeds given from the model works poorly in this case, particularly for the convection located over the southern portion, just east of the center of the grid box. The different wind velocities at varying heights of that cloud structure are evident by the mid-level clouds (dark green) moving west as the higher-level clouds in dark blue move east. This behavior pulls the initial cloud structure apart creating an unrealistic 3D cloud representation.

Thus, the assumption used with OF nowcasting that CTH and CBH move together with the same flow is proven to work better in this case.

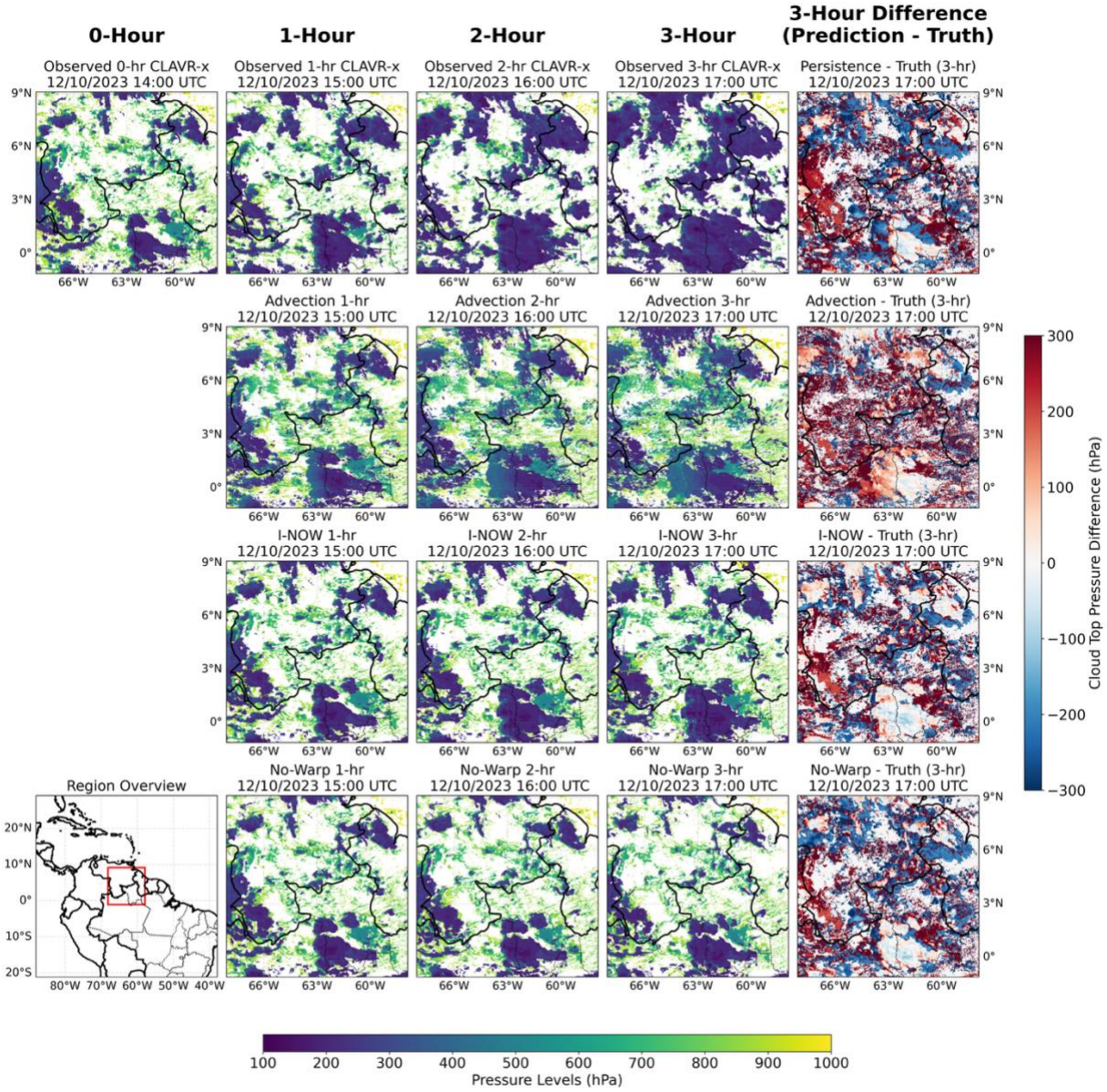

Figure 2.13: As in Figure 2.10, but for a case study on 15 July 2023, 17:00 UTC over northern Brazil and southern Venezuela.

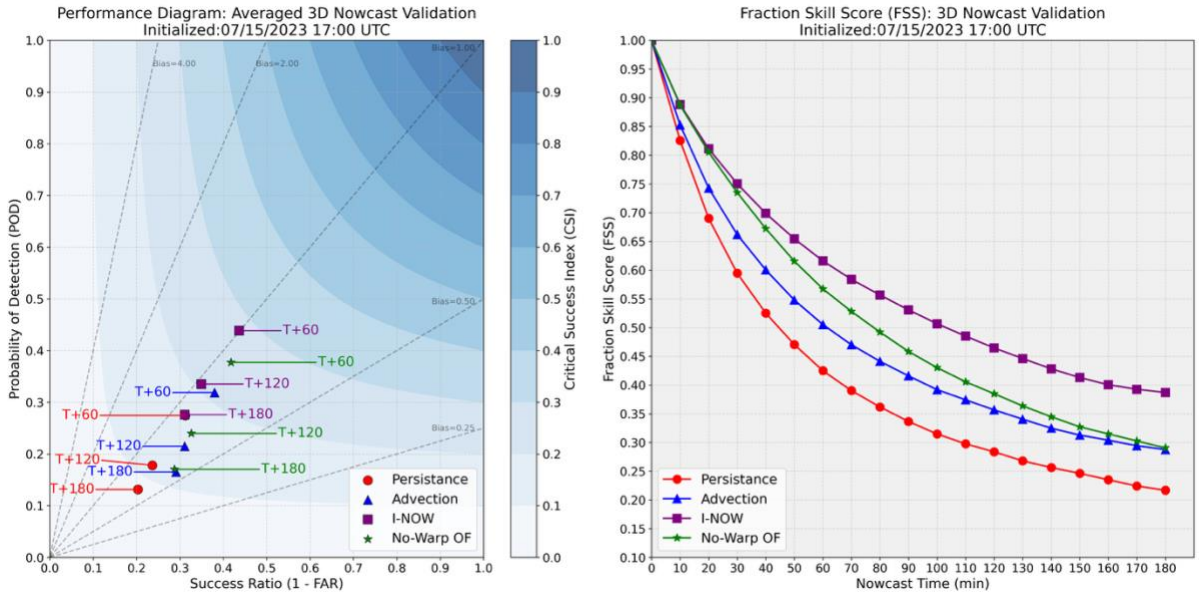

Figure 2.14: As in Figure 2.11, 3D validation results for 15 July 2023, 17:00 UTC case study.

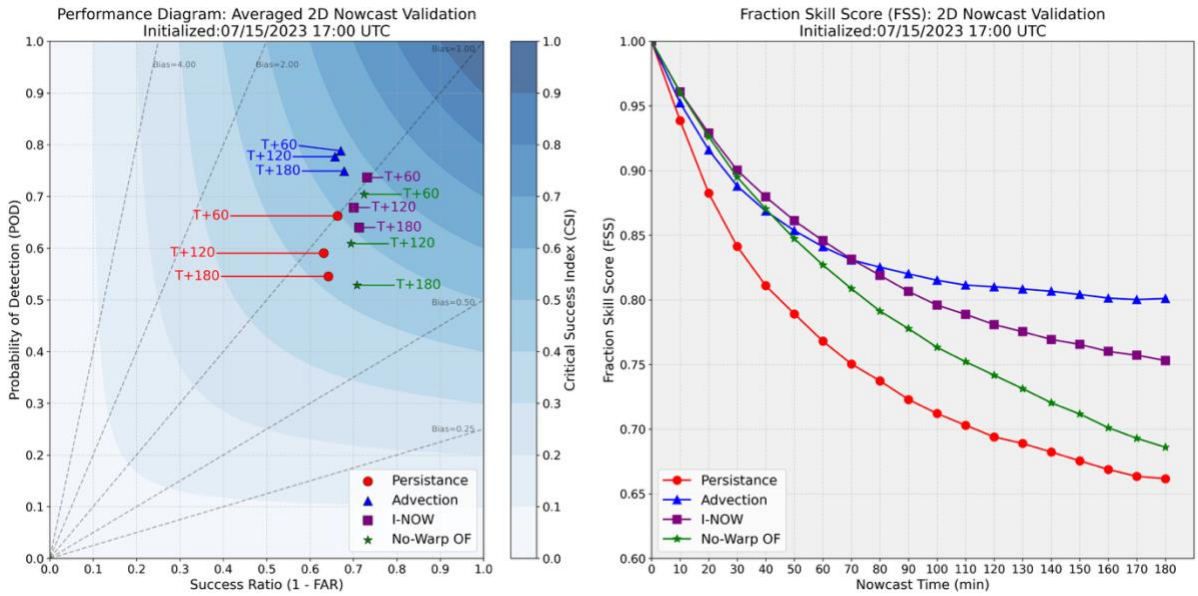

Figure 2.15: As in Figure 2.12, cloud mask validation results for 15 July 2023, 17:00 UTC case study.

The 3D nowcast validation for this second case study is shown in Figure 2.14. Again, results show superior validation for the I-NOW method for POD, CSI, and FSS mtrics, with lower bias than the previous case study. However, this performance is in stark contrast to the cloud mask results in Figure 2.15 which show surprisingly superior results for the advection method in the latter half of the nowcast period for POD, CSI, and FSS. Despite these seemingly better 2D results, the advection method exhibits a significant increase in cloud mask bias when compared to the other methods. This observation agrees with visual inspection of Figure 2.13 where the advection method ultimately fills more of the 2D cloud mask validation grid and thus allows for higher cloud mask POD, CSI, and FSS at the expense of creating many more false alarms. The fact that the advection method is only able to outperform the other methods for 2D validation is indicative of its prediction being more accurate for cloud mask for the wrong reasons. The independent advection at each pressure level creates an unrealistic depiction of the 3D cloud field which allows for I-NOW to have a higher skill score for 3D clouds. Ultimately, in the application of a 3D cloud-free-line-of-sight calculation, the I-NOW nowcast would be more accurate and reinforces that the keeping CBHs tightly coupled with CTHs is a useful assumption in 3D cloud nowcasting.

2.3.4. Case Study 3 – (10 December 2023, 14:00 UTC)

The last case study, located over western Argentina and central Chile, is shown in Figure 2.16. This case offers a good evaluation of an environment containing multiple cloud layers. The observations (top row) show three main noteworthy features. The first feature is the presence of cirrus clouds, in dark blue, moving through the area towards the southeast. The second feature is the low-level cloud deck (yellow) on the western edge of the grid and well off the coast of Chile. This low-level cloud deck, although difficult to discern in individual images, is moving towards the north, and is partially obscured by the cirrus clouds which becomes evident as the cirrus clouds

move towards the southeast. The last feature of note is the low to mid-level clouds (green) east of the Chilean and Argentinian border, some of which are mostly stationary due to the influence of mountainous terrain and others that are moving towards the south.

The advection method (2nd row) upon first inspection appears to handle the different cloud layers qualitatively well. The cirrus moves towards the southeast and the clouds east of the Chilean and Argentinian border are retained below the cirrus. However, the single-layer cloud assumption becomes evident when a hole (first subplot; second row of Figure 2.16) in the low-level cloud deck emerges in the first hour of the advection method nowcast. Located on the western part of the grid, this hole in the low-level clouds was created when the cirrus moved toward the southeast revealing that the initial state of the cloud field was not aware of the low-level clouds underneath the cirrus. This analysis points to an issue of the initial state rather than the advection method itself, but ultimately negatively impacts the advection method validation. The CLAVR-x algorithm does provide an 'overlap' cloud classifier, designed to identify such "cirrus over low-cloud' situations. Thus, it may be possible to introduce such information a priori to mitigate such scenarios, but this was not attempted in the current version of I-NOW or the baselines methods, and is a topic relegated to future research.

One issue with the OF-based nowcasts (3rd and 4th rows) is apparent over the southwest portion of the domain, where the low-level clouds offshore are visible among the higher cirrus. Rather than creating holes as observed with the GFS wind-based advection method, the OF methods utilize an incorrect flow retrieval due to the presence of multiple layers. Consequently, the low-level clouds move along with the cirrus towards the coast when the lower-level clouds should move towards the north.

As seen in the first case study, the no-warp OF method (bottom row) does not continue to move cirrus clouds towards the southeast, as these clouds move into an area of flow discontinuity over the Chilean and Argentinian border. The I-NOW method (3rd row), on the other hand, does continue to move the cirrus clouds. However, the occlusion reasoning eliminates the clouds below the cirrus unlike the advection method (2nd row) which retains the clouds. Overall, the case reveals mixed performance and exposes current limitations with the current version of I-NOW as well as opportunities for iterative improvement to the method in future research.

Despite the observed limitations, the I-NOW method still outperformed the other methods statistically. Figure 2.17 shows the performance diagram and FSS for 3D validation, and Figure 2.18 shows the same for cloud mask validation. In particular, the no-warp OF method shows a steady decrease in POD, CSI, and FSS when compared to the I-NOW method as the nowcast extends towards later times, due in large part to the no-warp method's stagnation of the cirrus clouds as they approached flow discontinuities. Additionally, in contrast to the other case studies, bias is relatively low for the cloud mask validation for all the methods in this final case with the lowest SR values associated with the advection method. This finding of low bias matches visually what is observed in Figure 2.16 in that few clouds appear to dissipate, and the advection method is penalized for the hole it produced in the low-level cloud field.

Ultimately, considering the superior performance of I-NOW, this case shows again that accounting for the time related changes to the OF field by warping and applying relatively simple occlusion reasoning enables for a timely and skillful 3D cloud nowcast.

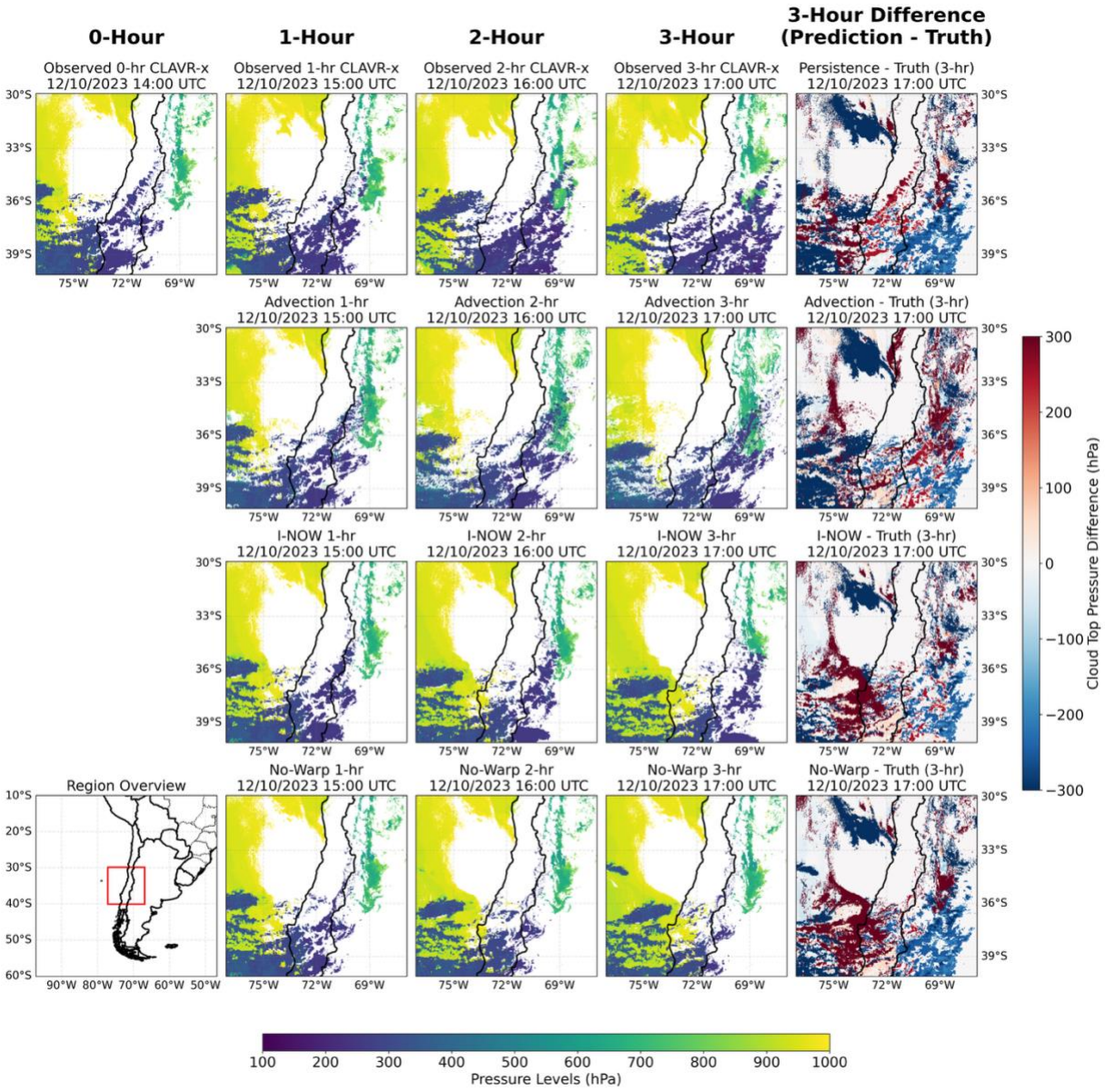

Figure 2.16: As in Figure 2.10, but for a case study on 10 December 2023, 14:00 UTC over western Argentina and central Chile.

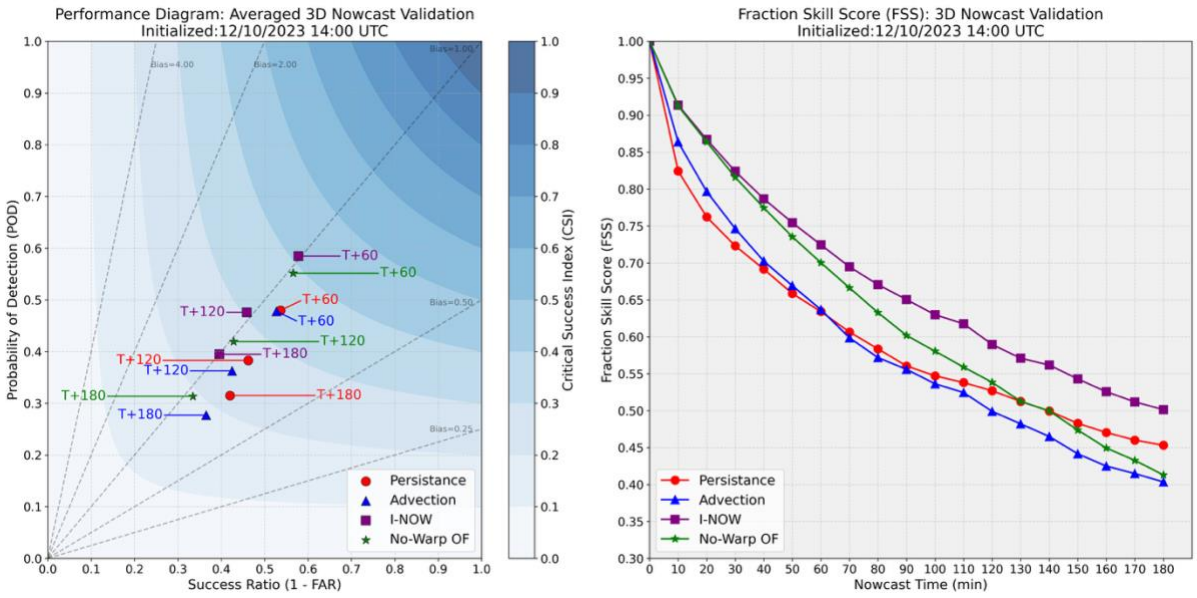

Figure 2.17: As in Figure 2.11, 3D validation results for 10 December 2023, 14:00 UTC case study.

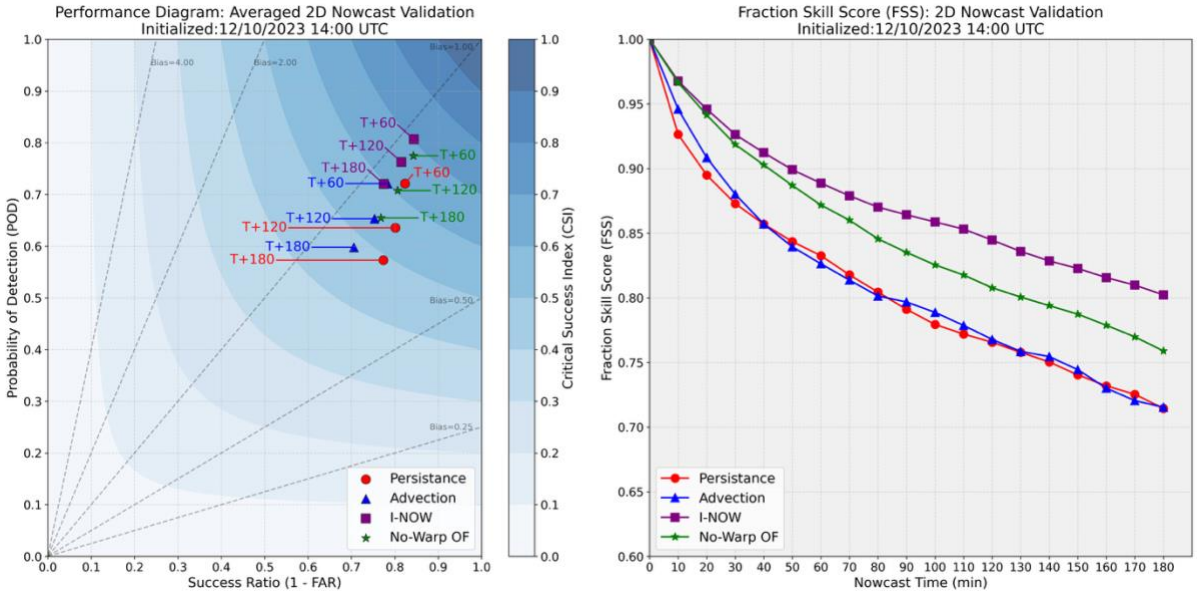

Figure 2.18: As in Figure 2.12, cloud mask validation results for 10 December 2023, 14:00 UTC case study.

2.4. Conclusions

Cloud nowcasting by way of satellite imagery has often utilized OF to extrapolate the future location of clouds using the assumption of a constant flow field with respect to time. However, since the retrieved OF field derived from satellite imagery is piecewise and often contains flow discontinuities that rapidly change, the advantages of a constant flow field assumption quickly decline with time. In this study, a method commonly utilized in the computer vision community for the temporal interpolation of image frames, referred to as warping, was adapted to better capture time related changes in OF fields retrieved from meteorological satellite imagery to nowcast 3D cloud fields constructed from CLAVR-x data.

This chapter posed the following science question: To what extent does accounting for temporal changes in a piecewise OF field improve predictions of 3D cloud movements? Two hypotheses followed this question: 1) if the OF field is updated to account for time related changes for each nowcast time step, then there will be an increase in cloud nowcast skill when compared to a linear extrapolation method in which the OF field is assumed constant with respect to time, which is a common approach in OF-based nowcast applications; 2) if the same apparent motion observed from cloud top height (CTH) is also applied to cloud base heights (CBH), then the resulting 3D cloud nowcast will be more accurate than a method that applies estimated wind speeds for individual levels.

In regard to the first hypothesis, it was determined that using temporal interpolation method of warping in conjunction with simple occlusion reasoning, satellite imagery and retrieved cloud properties can be better extrapolated forward while accounting for cloud occlusions and piecewise OF fields when compared to linear extrapolation methods assuming a constant flow field. Specifically, it was determined that accounting for the temporal changes to the OF field using the

I-NOW method, nowcast skill improvement was statistically significant compared with methods using traditional linear extrapolation and no occlusion reasoning.

For the second hypothesis it was found that within the context of 3D cloud nowcasting, the assumption of utilizing the same flow field for both CTH and CBH, was indeed more accurate than utilizing NWP winds at multiple levels.

Overall, the OF-based nowcast method using a warping technique based on temporal interpolation techniques, I-NOW, was found to outperform three baseline methods including persistence, multilayer NWP-based (GFS) wind advection, and OF-based nowcasting utilizing a constant flow field and linear extrapolation based on results from a sample of 991 randomly selected cases from 2024. Good performance on multivariate prediction to infer 3D cloud states here is also indicative that similar cloud motions inferred from infrared satellite imagery may generalize well to other datasets where motions are captured in infrared fields (i.e. dust or thick smoke plumes; e.g. Grasso et al. 2021). Thus, an end user could use the I-NOW approach to predict products beyond those produced by the CLAVR-x algorithm.

Additionally, three deep-dive case study analyses from 2023 demonstrate I-NOW performance against traditional baselines and illustrate real-world meteorological scenarios where the different approaches (including I-NOW) struggle. Moreover, this study compared I-NOW in the task of satellite brightness temperature imagery prediction against a baseline U-Net and showed that by applying OF warping in a manner that better handles flow discontinuities and occlusions OF-based predictions can outperform newer methods of image prediction.

Although I-NOW provided superior statistical and qualitative performance, future work is required to address some of the method's outstanding weaknesses, enumerated as follows:

i) I-NOW does not account for acceleration-related changes in motion within the OF field, which is mainly an issue in rotational and divergent scenes. One potential improvement along these lines could be to include taking multiple OF retrievals prior to a nowcast initial time to estimate higher-order changes in the OF field, to then update flow values across time. A key challenge to this approach is determining a method to include the higher-order changes to the warped OF field in each warping iteration. A possible approach could include estimating higher order changes by taking the difference between optical flow retrievals from prior times and the latest OF retrieval and adding those changes to each warp iteration. Experimentation, of course would be required to determine if such an approach would better account for the time related changes to the OF field.

ii) The simple occlusion reasoning in I-NOW prohibits the retention of multiple cloud layers when occlusions occur. A possible improvement would involve applying the warped OF method on low, middle, and high cloud layers separately which would allow some "memory" of clouds obscured by occlusions and may result in further improvements to the skill scores observed here (at the expense of increases needed in computation power or time).

iii) As illustrated in the final case study, OF retrievals can, at times, fail when OF algorithm constraints cause smoothing or cross-contamination of flow estimations across multiple cloud layers. Integrating cloud-top heights in OCTANE, specifically in the smoothness constraint (to mitigate cross-layer flow smoothing in the retrieval phase), could improve the retrieved motions for I-NOW-based applications. Additionally, using imagery with a higher temporal resolution (i.e. 5-min or 1-min) may provide better

accuracy for OF retrievals. However, testing on this question was not implemented in this study due to the OVERCAST requirement for global nowcasts which consequently necessitated investigated methods to be applied to full-disk 10-min imagery cadences.

iv) Cloud nowcasting techniques generally struggle to overcome three challenges simultaneously: 1) trajectory accuracy, 2) computational efficiency, and 3) cloud formation and dissipation. Here, I-NOW can reasonably achieve both trajectory accuracy and computational efficiency. Cloud formation and dissipation are, of course, unsolved outside of the context of NWP models which are unable to provide trajectory accuracy and computational efficiency within the realm of nowcast timeframes. ML shows promise in solving all three challenges simultaneously, but also can struggle at resolving advection without over-smoothing output solutions (Nielsen et al. 2021; Kellerhals et al. 2022).

Given the positive results observed in this initial development, it naturally follows that efforts should be devoted to combining ML with I-NOW to better resolve cloud formation and dissipation regions. Dissipation, in particular, would help to reduce I-NOW's cloud mask FAR statistics. The next chapter is devoted to leveraging the I-NOW algorithm to facilitate ML labeling for dissipation prediction, while the formation problem is relegated to a later chapter.

CHAPTER 3: MACHINE LEARNING PREDICTION OF SHORT-TERM LOW-LEVEL CLOUD DISSIPATION USING OPTICAL FLOW LABELING

3.1. Introduction

Cloud cover can significantly disrupt a wide range of operations and practical applications, including solar energy generation (2008), commercial and general aviation (Herman and Schumacher 2016), and military battlespace awareness (Dolinar and Nachamkin 2024). Short-term cloud forecasts (0 – 3 hours), or nowcasts, are especially valuable when NWP models are limited by coarse spatial resolution or stale initialization times, making them less accurate than simple persistence methods. In such situations, satellite-based nowcasting techniques offer a skillful alternative.

A common satellite-based nowcasting approach involves the use of extrapolation of the motion of clouds from successive images, an example of which was presented in the previous chapter. Such approaches use OF, which as defined by the seminal computer vision paper on the topic, is the distribution of apparent velocities of movement of brightness patterns in an image (Horn and Schunck 1981). With satellite imagery sequences, the OF field provides the apparent motion of clouds at a given point in time which enables the extrapolation of future cloud positions. However, as shown in the previous chapter, OF-based nowcasting is inherently limited to advecting existing clouds and cannot account for cloud formation or dissipation which are key aspects of cloud evolution that remain an open challenge in cloud nowcasting (Schütz et al. 2024; Sim et al. 2024).

With the stunning progress of Machine Learning (ML) methods in other applications today, many studies naturally look to data-driven approaches, typically using deep-learning-based convolutional neural networks (CNNs) or related architectures to predict future cloud fields or imagery from sequences of satellite images (Berthomier et al. 2020; Nielsen et al. 2021; Kellerhals

et al. 2022; Yu et al. 2023). This approach, although seemingly intuitive, requires ML models to attempt to learn patterns that predict cloud advection, formation, and dissipation simultaneously. However, such models are limited by their reliance on purely image-based input; they have limited data to predict clouds entering from outside their domains, and they often struggle to detect dissipation or formation due to insufficient contextual information such as moisture, vertical motion, or other atmospheric dynamical parameters (e.g. Berthomier et al. 2020; Kellerhals et al. 2022; Partio et al. 2024). Furthermore, to minimize loss function penalties, ML-based regression predictions tend to become overly smoothed (i.e., losing spatial detail and sharpness), especially as the forecast timeframe increases (Kellerhals et al. 2022; Partio et al. 2024).

In recent work, diffusion has become a new avenue to avoid smoothing and gain better predictability of cloud evolution. Chase et al. (2025) utilized diffusion models, which learn to generate data by gradually denoising random noise, to predict infrared satellite imagery which were more skillful and sharper than a U-Net model. Chen et al. (2025) also utilized diffusion models to predict imagery containing convective clouds and showed superior performance against both persistence and an OF-based method. Although such approaches do appear to provide promising results through their increased complexity, there is often a cost to comprehensibility or model interpretability (Goethals et al. 2022) and an increased requirement for training data (Chen et al. 2025), as well as increased computational needs for training and inference (Chase et al. 2025).

Thus, this chapter posits that by narrowing the ML prediction task, improvements to cloud evolution predictions will be more easily achieved. Specifically, we show that ML can skillfully predict low-level cloud dissipation when advection is removed from the prediction task. This removal is realized by using the OF-based nowcast defined in the previous chapter which consistently outperforms persistence to create a labeled dataset that classifies which clouds in an

initial cloud field will dissipate and in what timeframe. To support the predictions, we include OF-based nowcasts of CTH and CGT initiated from prior timeframes as well as NWP-supplied RH fields to provide ML with atmospheric context. Finally, we design the architecture to make predictions for dissipation as a classification task to update the OF-based nowcast method rather than by direct prediction to avoid the common problem in ML-based techniques of spatial smoothing.

To guide the research for the problem of cloud dissipation, the following science question is posed: to what extent do observed changes in cloud properties and coarse-resolution NWP relative humidity (RH) fields predict future cloud dissipation in the short-term nowcast timeframe (0 – 3 hours)? This question is accompanied by two hypotheses: 1) if there is predictive information from a series of OF-based cloud nowcasts of cloud properties and NWP RH values, then ML will provide measurable skill in predicting short-term cloud dissipation; 2) if short-term changes of nowcasted cloud properties observed across time provide more predictability than NWP RH fields due to the recency of the nowcasts, then an ablation of a ML model that contains nowcasts as input versus an ablation containing NWP RH fields will show more dissipation skill.

3.2. Data and Methods

3.2.1. Overview

The work in this chapter is built upon the OF-based nowcast method presented in the previous chapter, Improved Nowcasting via Optical flow Warping, referred to hereafter as I-NOW. Like any OF-based nowcasting method, I-NOW is limited to the advection of existing clouds. To provide cloud dissipation capability to I-NOW, two U-Net models were designed to predict cloud dissipation of low-level (i.e., 0 – 3000m AMSL) clouds. The first U-Net was trained to predict which clouds in an initial cloud field would dissipate between the initial nowcast time and 3 hours.

The second U-Net was trained to predict, based on the simulated output of the first U-Net, at which time those clouds would dissipate in 10-minute increments. The resulting dual U-Net prediction of cloud dissipation was then used within I-NOW to dissipate clouds based on their assigned 10-minute dissipation increment. Figure 3.1 provides an overview of the dual U-Net setup in which the prediction of the dissipation U-Net is included as an input into the timing U-Net to ultimately predict which clouds observed in the initial cloud field should be dissipated and at which time step they should be dissipated.

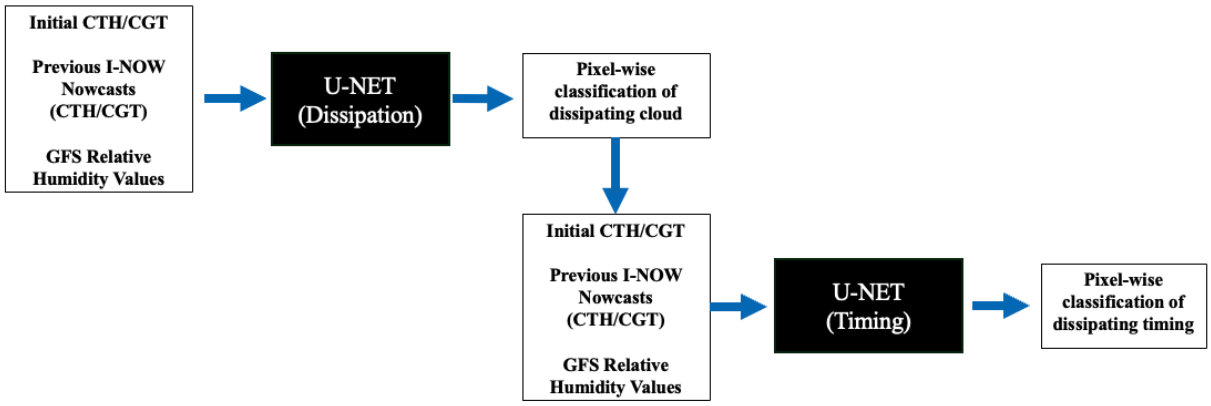

Figure 3.1: Overview of ML setup to predict pixel-wise classification of cloud dissipation to supplement OF-based nowcasting.

By design, I-NOW always extrapolates clouds from the initially observed cloud field while incrementally updating the OF field with each iterative time step. This avoids the problem of stationary/locked pixels where the magnitude of displacements for a single time step are insufficient to move pixels beyond their original pixel location. Furthermore, this design also enables the direct comparison of the original cloud field with the observed cloud field at a later timeframe, thereby allowing for the determination of which cloud pixels should be designated as dissipated and at which timeframe for ML labeling. At inference the resulting prediction classifies

individual pixels as either constant or dissipated at a particular timeframe. This ultimately allows cloud pixels to be reclassified as a clear or "no-cloud" pixel at the appropriate time in a nowcast.

3.2.2. I-NOW

As a reminder to the reader, I-NOW utilizes the variational OF method available from the Optical flow Code for Tracking, AMV, and Nowcasting Experiments (OCTANE; Apke et al. 2022). With the use of two successive satellite images, OCTANE resolves dense OF at every image pixel by minimizing a penalty function that variationally penalizes optical flow guess deviations from assumptions of brightness constancy, gradient constancy, and smoothness constraint with two tunable hyperparameters to control the importance of gradient constancy and smoothness (Apke et al. 2022). The previous chapter provided the specific hyperparameters for 10.3 μm GOES ABI imagery which minimizes the root mean square error (RMSE) of 1-hour CTH full-disk nowcasts.

The signature feature to I-NOW that yields improvement upon past applications of OF-based cloud nowcasting is its account of time-related changes of the piecewise flow field that are inherent to satellite imagery OF retrievals. The previous chapter showed that assuming a constant motion field to nowcast satellite imagery introduces unnecessary inaccuracies to extrapolate cloud movements. I-NOW, instead, utilizes an adapted frame interpolation or "warping" method originally presented by Baker et al. (2011) to first extrapolate the future state of the OF field to a future timeframe. The extrapolated OF field is then used to advect clouds from the original cloud field forward to the desired time step with simple occlusion reasoning that retains the highest CTHs, or coldest radiometric pixels if nowcasting imagery, in the event multiple trajectories end at the same pixel location.

3.2.3. CLAVR-x & GFS NWP data

In this chapter, CTH and CGT retrievals from GOES-16 imagery, processed from Clouds from the Advanced Very High Resolution Radiometer (AVHRR) – Extended or CLAVR-x (Heidinger et al. 2012; Walther and Heidinger 2012) were utilized as ML input data. These data are assumed to be truth data when developing the labeled dataset for ML training. However, CLAVR-x retrieval algorithms change methods when daytime visible imagery is not available. Thus, in order to avoid any significant differences in the truth data, instances where solar zenith angles exceeded 75 degrees were omitted in the training, validation, and test datasets.

NWP relative humidity fields were also utilized in ML input data. Specifically, NOAA's Global Forecast System (GFS) (available online at https://registry.opendata.aws/noaa-gfs-bdp-pds) was leveraged to provide atmospheric context to potential cloud dissipation. Because the GFS archive contains only NWP data for 3-hour intervals, relative humidity values for pressure levels were interpolated in time by way of a semi-Lagrangian advection method using the model's native resolution of 0.25° latitude and longitude. Winds forecasted from the model were used to advect relative humidity values to a desired valid time that may exist between any 3-hour valid model times both forward from the earlier time and backward from the later time. The interpolations were then weighted based on their distance with respect to their initial advection timeframe and the desired interpolated timeframe.

After temporal interpolation of the model data, relative humidity values were then interpolated onto a 0.02° resolution 512 x 512 rectilinear latitude and longitude grid for input into the U-Nets. It is natural to question the use of the GFS model due to its coarse resolution and parameterized physics rather than a more advanced and higher resolution mesoscale model. Since OVERCAST

requires nowcasts on large- to global-domains, it is necessary to use a readily available global model that could easily be utilized for any area over the globe.

Additionally, there are two reasons for the use of RH values specifically rather than a combination of numerous variables from GFS. First, with limited CLAVR-x data available on archive for OVERCAST and limited resources, including more input variables increases the risk that an optimal ML solution will not be achieved. Secondly, a channel importance analysis of GFS variables when applying the task of cloud classification with a simple random forest approach showed that RH values as having the highest contributor to cloud classification. This agrees with the study by Sim et al. (2024) in that RH was the highest contributor to the short-term prediction of ocean fog. Although other variables from GFS may provide additional benefit, for the sake of keeping the ML setup simplified, only RH values are utilized.

3.2.4. Optical Flow Labeling

Because I-NOW always references the original cloud field to extrapolate cloud movement for each iterative nowcast time step, creating a static ML label for cloud dissipation was relatively straightforward. I-NOW was utilized to nowcast full-disk CLAVR-x retrievals of CTH using 10.3 μm GOES-16 imagery to retrieve OF fields. Nowcast iterations were set to 10-minute increments which coincided with the 10-minute increments of GOES-16 full-disk imagery as well as the CLAVR-x retrievals. For each I-NOW nowcast increment, the original array indices of the initial cloud field were incremented along with the values of CTH. This allowed for the position of the original cloud field pixels to be referenced with future positions of clouds in observed CLAVR-x CTH data.

For any given nowcast pixel of CTH, if the observed CLAVR-x data indicated that no cloud was observed, then the pixel in the original cloud field location was designated as dissipated at the

observed CLAVR-x valid time. To avoid noise in the labels, any pixels that were at first designated for dissipation and were found to match with an observed CTH pixel in CLAVR-x at a later timeframe were classified as constant cloud. Labeling consisted of 20 different classes. In order from 0 – 19: where index 0 = constant cloud, indices 1-18 = 10-minute increment of cloud dissipation, and index 19 = no cloud in the original cloud field. The labels were made with classes encoded as integer values. This labeling strategy enabled the use of sparse categorical cross-entropy as the loss function rather than categorical cross-entropy which requires more memory during training. Figure 3.2 provides a visual representation of an example dissipation truth label.

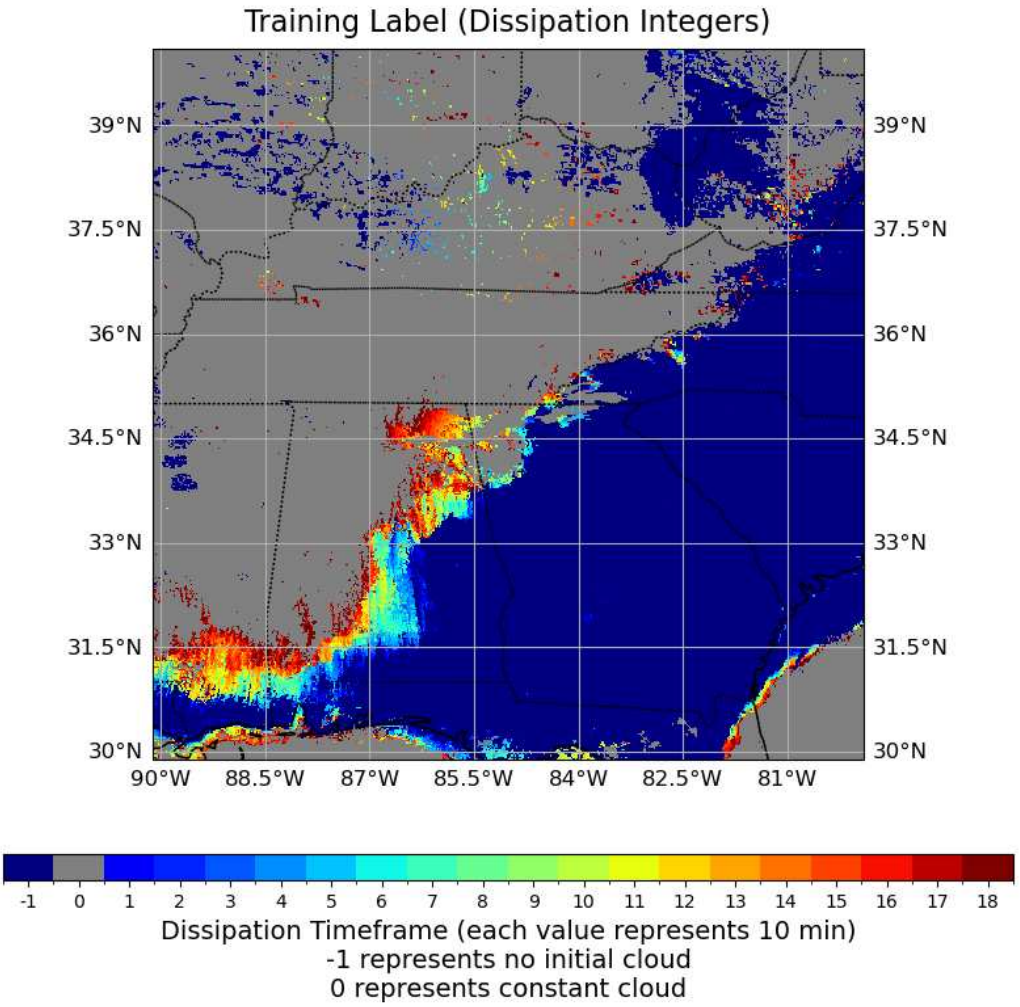

Figure 3.2: Visual representation of a dissipation training label representing the timeframes at which cloud pixels have been determined to have dissipated. Values of -1 represent where there

is no cloud present in the initial cloud field (-1 is shown here for ease of viewing rather than 19). Values of 0 represent clouds that remain persistent through the 3-hour nowcast. Values 1 – 18 represent the timeframe in 10-min increments at which a cloud pixel has been observed to have dissipated.

3.2.5. Model Architecture

With the use of Tensorflow, the two U-Nets were designed using similar architecture as the original U-Net developed by Ronneberger et al. (2015) where input data is contracted in an encoder section and expanded in a decoder section with connections (i.e. skip connections) between the encoder and decoder at each respective level. Graphical representations of the two U-Net architectures are shown in Figure 3.3 and Figure 3.4. Starting with 64 filters, the number of convolutional filters was doubled through each level of the encoder and halved for each level through the decoder. Both U-Nets for this study have the same architecture except for the number of input channels, number of output classes, and a variation of forced class predictions based on the contents of the first channel of the input tensor. Rectified Linear Unit or ReLU (Krizhevsky et al. 2012) was utilized for all activations in 2D convolutions with the exception of the last 2D convolution which was used simply to obtain the number of channels required for the number of desired classes. Softmax (Terven et al. 2025) was used as the last activation function to determine the appropriate cloud dissipation class for each pixel. Softmax for a given pixel is expressed as:

$$\hat{p}_{i,j} = \frac{\exp(z_{i,j})}{\sum_{k=1}^{C} \exp(z^{i,k})} \tag{3.1}$$

where i is the index to a specific pixel-wise classification, j is the index for class (i.e. constant cloud, dissipating cloud, constant cloud), $z_{i,j}$ is the logit or raw model value produced by the model prior to the activation function, k refers to a specific class, C is the total number of classes, and $\hat{p}_{i,j}$ is the softmax or normalized probability that pixel i belongs to class j.

Sparse categorical cross-entropy (SCCE) was used as the loss function which differs from categorical cross-entropy in that the class labels are encoded as integers rather than one-hot vectors. SCCE is crucial here, as using one-hot encoding can be memory-intensive for large numbers of classes (Terven et al. 2025). Where n is the number of pixels, summing over all pixels, SCCE can be expressed as:

$$SCCE = -\frac{1}{n}\sum_{i=1}^{n}\log(p_{i,k}) \tag{3.2}$$

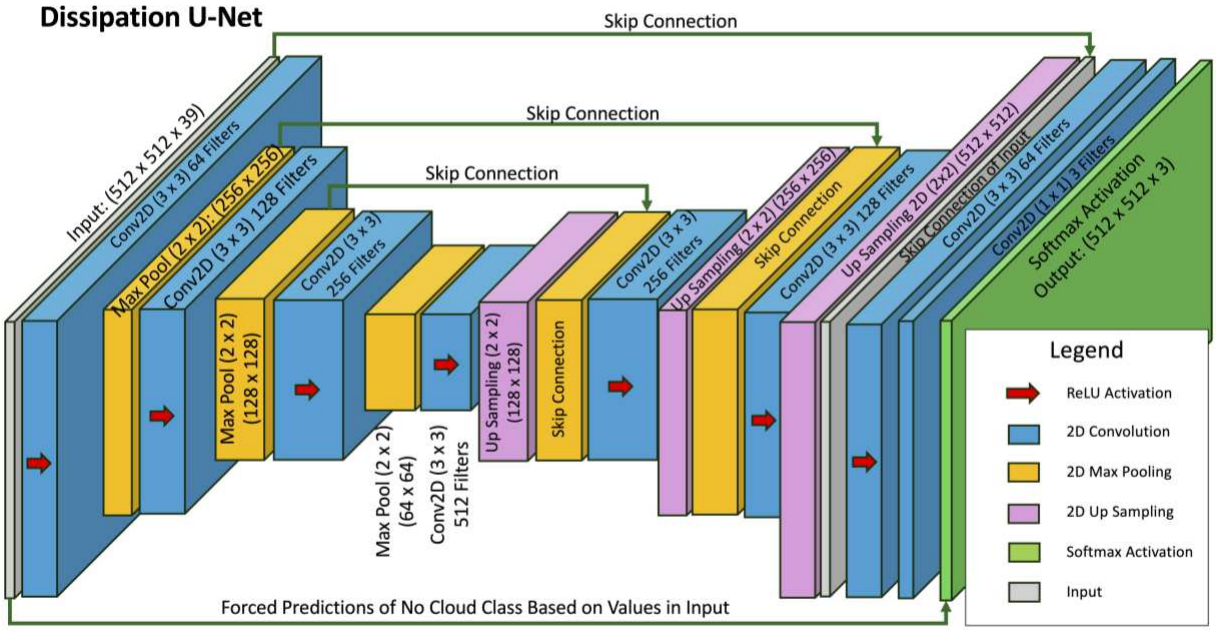

Figure 3.3: Architecture of dissipation U-Net. Softmax is used to predict 3 classes: 0 = constant cloud, 1 = dissipated cloud, 2 = no cloud. The blocks indicate the output of each layer. Each block is labeled with its size. Additionally, each block is also labeled with the operation (layer) that leads to it.

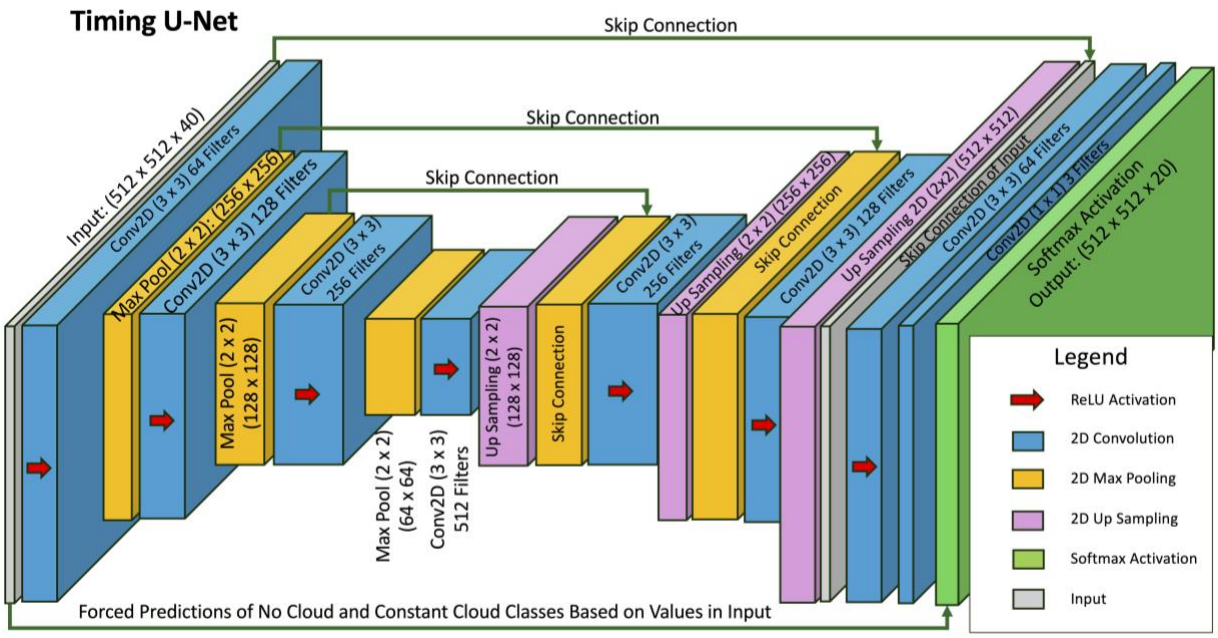

Figure 3.4: Architecture of timing U-Net. Output from the dissipation U-Net is added as an additional channel to the timing U-Net which increases the input channels to 40. Softmax is used to predict 20 classes: 0 = constant cloud, 1 - 18 = 10-minute increment of dissipation, 19 = no cloud. Blocks are annotated in the same manner as Figure 3.3.

The reason for the use of two U-Nets stemmed from the difficulty in training a single model for the dissipation task. Initial attempts with training a single U-Net architecture would consistently result in nearly all dissipating cloud to be classified for 10-minute dissipation. Thus, rather than constructing more intricate architecture or increasing trainable parameters in a single model, the tasks of classifying dissipation and timing were separated.

The first U-Net, the so-called Dissipation U-Net (Figure 3.3), was trained to only predict which cloud pixels will dissipate within the next 3 hours. The training labels were adapted during training to convert any pixel designated for dissipation to just one class, thereby providing a truth label dataset that consisted of just three classes: constant cloud, dissipated cloud, and no cloud. Because initial attempts to train the Dissipation U-Net showed that it was struggling to learn that values representing no cloud in the initial CTH field were always classified as no cloud in the output, the architecture was adapted to force the prediction of no cloud based on the CTH channel input.

The second U-Net, the so-called Timing U-Net (Figure 3.4), was trained to only predict the timing of dissipation rather than which pixels would dissipate. During training, simulated output from the Dissipation U-Net was placed in the input tensor of the Timing U-Net based on the truth from the dataset itself rather than using the output from the trained Dissipation U-Net. The addition of the dissipation prediction increased the input channels from 39 to 40 when compared to the Dissipation U-Net. Using the simulated output within the training dataset allowed for both U-Nets to be trained independently. Like the Dissipation U-Net, initial attempts to train the Timing U-Net showed that the architecture struggled to flawlessly learn the correct predictions associated with no cloud pixels and constant cloud pixels from the simulated dissipation input. Thus, both classes, constant cloud and no cloud classes, were forced predictions based on the prediction associated with the Dissipation U-Net.

3.2.6. Model Inputs

The inputs for both U-Nets consisted of initial CTH and CGT values, I-NOW nowcasts, and GFS relative humidity values. I-NOW nowcasts were initiated every 10 minutes from 10 to 50 minutes prior of the initial timeframe targeted for the prediction. Thus, if the UNets were tasked with predicting dissipation for a nowcast initiated at 18:00 UTC, the I-NOW nowcasts would be initiated at 17:50, 17:40, 17:30, 17:20, and 17:10 UTC respectively. The I-NOW nowcasts provided nowcasts of CTH and CGT, however, the difference between the initial CGT and the nowcast CGT was placed into the input rather than simply the CGT nowcast prediction. The intent for this step was for the U-Nets to learn connections between the change in observed CGT from successive nowcasts to predict dissipation and timing.

GFS relative humidity values were used in the input, but only at the following levels: 1000, 975, 950, 925, 900, 850, 800, 750, and 700 hPa. These pressure levels were included due to the

focus of the prediction on low-level cloud dissipation ranging from 0 to 3000m AMSL. GFS inputs consisted of 27 channels, where 9 channels were GFS values valid at the initial timeframe, 9 channels were valid 3 hours after the initial timeframe, and 9 channels that were the difference between the previous 18 channels. To remove advection impacts in the data, GFS values valid 3 hours after the initial timeframe were advected backwards in time using the model's own wind velocities. Doing so enabled the U-Nets to have a static (with respect to time) representation of the changes in relative humidity to correspond with the static prediction of cloud dissipation.

3.2.7. Training, Validation, and Testing Split

Available CLAVR-x data utilized in this study supported research for the ONR OVERCAST project, which seeks to create a global 3D representation of clouds in the atmosphere to inform cloud nowcasts (Noh et al. 2024). Due to updates in CLAVR-x retrieval data for CGT, consistently available data began on 10 Apr 2023. Consequently, CLAVR-x data from 10 Apr 2023 to 30 Dec 2024 was utilized for training, validation, and testing.

To avoid cross-contamination of data between training, validation, and testing datasets, specific days were randomly selected to be part of each dataset. Because only daytime data were selected based on solar zenith angle, each dataset remained unique. Total individual dates numbered 611, with splits of 80% for training (488 days), 10% for validation (61 days), and 10% for testing (62 days). The GOES-16 full-disk ABI imagery and subsequently CLAVR-x data were available at 10-minute intervals. The satellite imagery provided 16,390 potential nowcast initial times for training, 2,037 for validation, and 2,101 for testing after filtering on solar zenith angles. Of the available nowcast initial times, at random 1000 were selected for training, 100 for validation, and 100 for testing.

Although I-NOW nowcasts and CLAVR-x data were available as full-disk arrays, limited computational resources prohibited training on full sized arrays. Thus, data were interpolated using bilinear interpolation onto rectilinear latitude and longitude grids at 0.02° resolution with 512 x 512 dimensions. Ten rectilinear grids for each nowcast timeframe were selected based on a random central latitude/longitude coordinate within each full-disk nowcast timeframe. The corners of the grids were utilized to determine if any grid met the solar zenith angle constraint through the entire 3-hour nowcast timeframe. Considering selected nowcast initial times and selected grids, in total there were 10,000 training samples, 1,000 validation samples, and 1,000 test samples.

3.2.8. Inference Approach

Due to I-NOW using full-disk sized arrays in operation, using 512 x 512 bilinear interpolated grids to denote cloud dissipation was inefficient during inference (i.e. using the U-Nets to make predictions for use in I-NOW). Additionally, limited computational resources prohibited the input of full-sized arrays into the U-Nets directly. Thus, rather than interpolating data onto rectilinear grids after the U-Nets were trained, the full-sized arrays during inference were simply applied as input to the U-Nets in patches by moving across the arrays from left to right starting from the top left of the arrays. Because NWP data were also required in the input, GFS data were interpolated onto the GOES-16 grid directly.

Despite training on rectilinear grids, the predictions from the U-Nets perform well on the native GOES grid and provide meaningful skill for cloud dissipation. This approach ultimately allows for simplified implementation ideal for the OVERCAST operational environment.

3.2.9. Evaluation Metrics

In Chapter 2, 3D cloud grids were built based on CLAVR-x CTH and CBH retrievals. However, in this chapter, where cloud dissipation prediction is the goal, constructing a 3D cloud grid and

applying parallax correction is unnecessary as the 2D array from the CTH retrieval provides the relevant data on whether a cloud pixel should be dissipated or not. Thus, after using the U-Nets to predict cloud dissipation at inference, updated I-NOW nowcasts are evaluated on essentially the resulting cloud mask from CLAVR-x retrievals.

Metrics used in Chapter 2, POD, CSI, FAR, and FSS are again used here in Chapter 3 to evaluate the original I-NOW algorithm, I-NOW with the added ML dissipation predictions, and persistence. POD, CSI, and FAR are provided within the context of a performance diagram, and FSS is provided in an additional subplot for each nowcast timeframe.

3.3. Results

3.3.1. Validation Data Results

Before evaluating the combination of the dual U-Net prediction and I-NOW on test data, evaluation was first performed on the validation dataset to determine if tuning was needed or if and the predictions were satisfactory when compared to I-NOW without dissipation. Figure 3.5 shows the averaged results of I-NOW with the ML-enabled dissipation compared with persistence and the original I-NOW nowcast algorithm for the validation dataset. What is evident from Figure 3.5 is that I-NOW with the ML-enabled dissipation performs poorly on the validation dataset compared to the original I-NOW algorithm. In the performance diagram (Figure 3.5, left), I-NOW with dissipation increased its SR (decreased FAR) but at the expense of lower POD, resulting in a low bias (under forecasted cloud). For FSS, (Figure 3.5, right) it is clear that dissipation in fact decreased the skill of the nowcast when compared to the original I-NOW algorithm as the average FSS for the ML enabled dissipation version failed to exceed I-NOW for all nowcast timeframes.

The dissipation U-Net was tasked to predict three different classes, namely constant cloud (0), dissipation cloud (1), or no cloud in the initial cloud field (2). Because there are three classes rather

than just two, the probability by which a cloud could be classified as dissipated may well be less than 50% when the maximum probability value is selected. Thus, in this application, it is important to explicitly define a probability or confidence threshold by which we accept the prediction of a dissipated cloud before applying the prediction to the timing U-Net. To remedy the initial underperformance on the validation dataset, varying confidence thresholds from the dissipation U-Net were evaluated through inference and the resulting skill scores were compared to the original I-NOW algorithm.

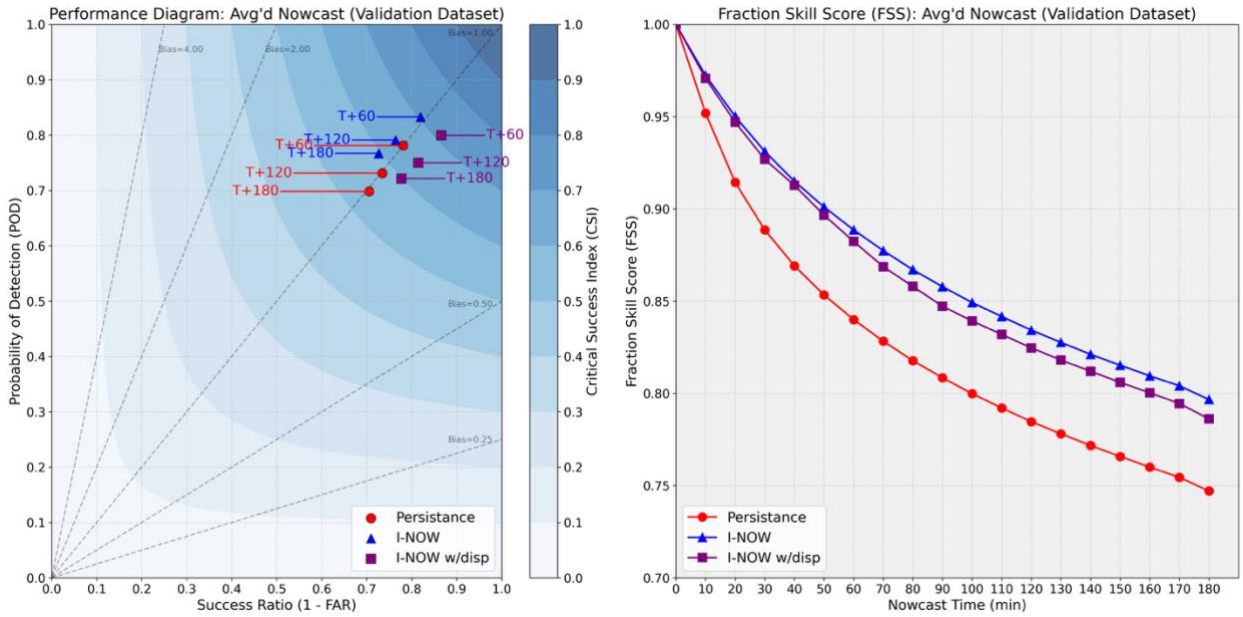

Figure 3.5: Validation dataset results of the combination of the dual U-Net cloud dissipation prediction and I-NOW algorithm. The performance diagram (left) shows the skill of persistence (red), I-NOW (blue), and I-NOW with ML enabled dissipation (purple) for 1-hour, 2-hour, and 3-hour nowcast timeframes. The subplot of FSS (right) shows the skill of persistence (red), I-NOW (blue), and I-NOW with ML enabled dissipation (purple) for all 10-minute nowcast timeframes.

Figure 3.6 shows the differences between the resulting metrics (POD, CSI, FAR, and FSS) of I-NOW with dissipation with varying confidence thresholds and the original I-NOW algorithm for the validation dataset. It is evident that with increasing confidence, there is a trade-off between increased false alarms and increased CSI and FSS. Additionally, there is a steady increase in CSI

and FSS for all thresholds in the initial third of the nowcast which indicates a possible reduction in cloud dissipation predictability after around an hour. Of the thresholds examined, the 70% threshold is the only threshold by which I-NOW with dissipation can maintain a FSS higher than the original version of the I-NOW algorithm. Interestingly, the CSI has a positive impact for nearly all the examined thresholds which is in stark contrast to FSS which had mostly negative impacts later in the nowcast timeframes except for the 70% threshold. This incongruity between CSI and FSS is likely due to the smoothing of cloud edges after clouds have dissipated despite attempts to structure the architecture to avoid the impact of smoothing.

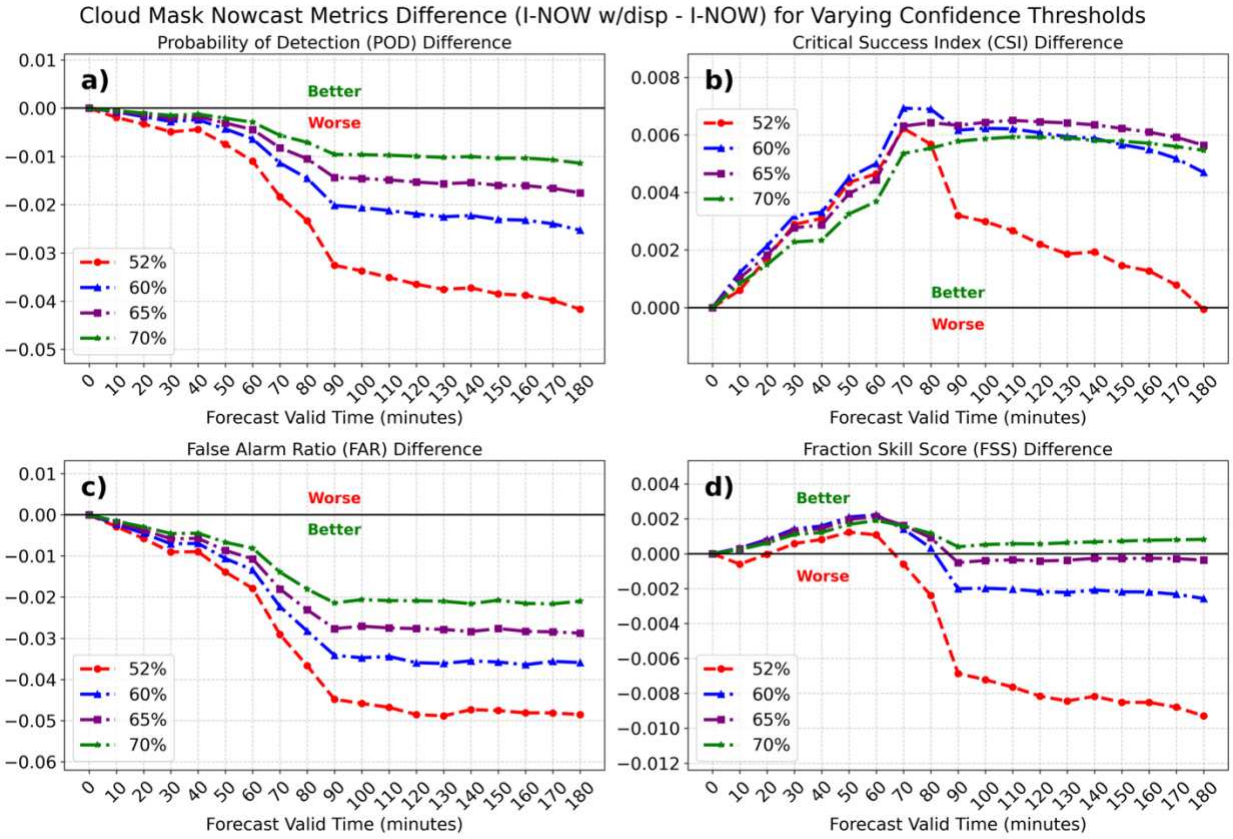

Figure 3.6: Difference of cloud mask nowcast metrics between varying dissipation confidence thresholds of I-NOW with dissipation and original I-NOW algorithm; a) POD difference; b) CSI difference; c) FAR difference, and d) FSS difference. Only the 70% confidence threshold for dissipation outperforms the original I-NOW algorithm for CSI and FSS.

3.3.2. Test Data Results

Figure 3.7 shows the averaged nowcast results on the test dataset using the 70% confidence threshold for dissipation compared to persistence and the original I-NOW algorithm. The results for the test dataset reveal a relatively high bias for all three methods; however, there is a definite decrease in bias for I-NOW with dissipation. This is clear evidence that there is an improvement to I-NOW using the U-Nets for dissipation. Additionally, there are small increases to CSI and FSS compared to the original I-NOW algorithm.

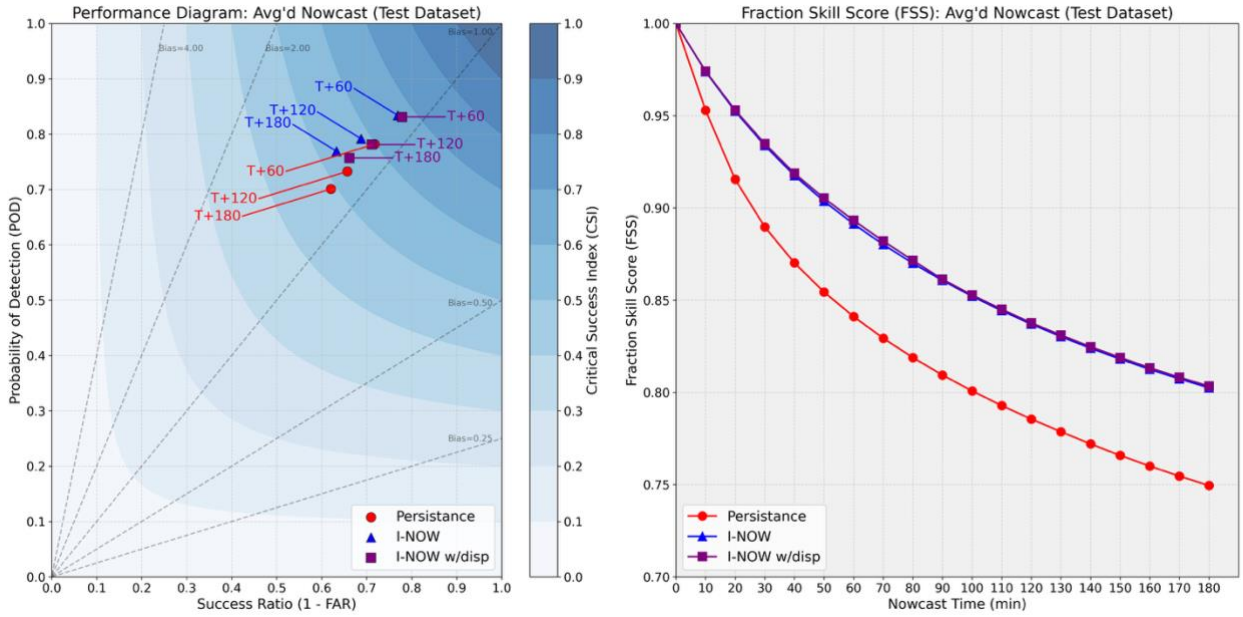

Figure 3.7: As in the Figure 3.5, for results for the test dataset using the 70% confidence threshold for dissipation for I-NOW with dissipation.

Table 3.1 shows the results of Wilcoxon signed-rank tests for POD, CSI, FAR, and FSS for the 3-hour nowcast timeframe. For POD, CSI, and FSS the Wilcoxon signed-rank test used the alternative hypothesis that I-NOW with dissipation was greater than the original I-NOW algorithm and for FAR the Wilcoxon signed-rank test used the alternative hypothesis that I-NOW with

dissipation was significantly less than the original I-NOW algorithm. Although the average results as displayed in Figure 3.7 show only a small average increase for CSI and FSS compared to the original I-NOW algorithm, the Wilcoxon signed-rank test results in Table 3.1 indicate that the increase is statistically significant (i.e. $p << 0.01$). In contrast, POD shows a p-value of 1.0 indicating that, with 100% statistical certainty, POD does not increase when utilizing I-NOW with dissipation. However, the decrease in POD is traded for a statistically significant decrease in FAR which provides ultimately a lower nowcast bias. Thus, objectively, there is a statistically significant improvement for I-NOW with ML enabled dissipation when compared to the original I-NOW algorithm.

Table 3.1: Wilcoxon Signed-Rank Test Results for the difference of I-NOW with dissipation using the 70% confidence threshold for dissipation compared to the original I-NOW algorithm for the nowcast metrics (POD, CSI, FAR, and FSS) computed on the test dataset.

Wilcoxon Signed-Rank Test Results (p-values)	
POD (alternative hypothesis: Greater)	1.0
CSI (alternative hypothesis: Greater)	3.19×10^{-89}
FAR (alternative hypothesis: Less)	1.11×10^{-152}
FSS (alternative hypothesis: Greater	1.85×10^{-16}

3.3.3. Case Study: 7 January 2025

Figure 3.8 shows a case study comparing the original I-NOW algorithm to I-NOW with dissipation with starting initial conditions on 7 January 2025, 15:10 UTC. As a reminder, the training, validation, and testing datasets were from 10 April 2023 to 30 December 2024. Therefore the case study was not in the training data and was not utilized to adjust any weights or biases in the U-Nets.

The case study shows a 3-hour period during which a low-level stratus cloud deck gradually dissipates over Mississippi and Alabama. The 3-hour differences show clearly that I-NOW with dissipation has fewer stratus clouds than the original I-NOW algorithm and is a good indication that the combination of the U-Nets with I-NOW is indeed dissipating clouds. Because the focus here was not on cloud formation, both versions of I-NOW miss cloud formation that occurs over West Virgina. Additionally, because the focus was on low-level cloud dissipation, the mid- to high-level clouds that move from west Kentucky to east Kentucky and dissipate through the nowcast period are not nowcasted to dissipate but rather remain persistent in both nowcasts which is to be expected. Furthermore, an issue with the OF retrieval, similar to what was observed in the third case study in Chapter 2, is observed over northern Georgia where low-level clouds are nowcasted by both methods to advect to the southeast due to the trajectories of cirrus clouds that initially were observed over northern Georgia.

.

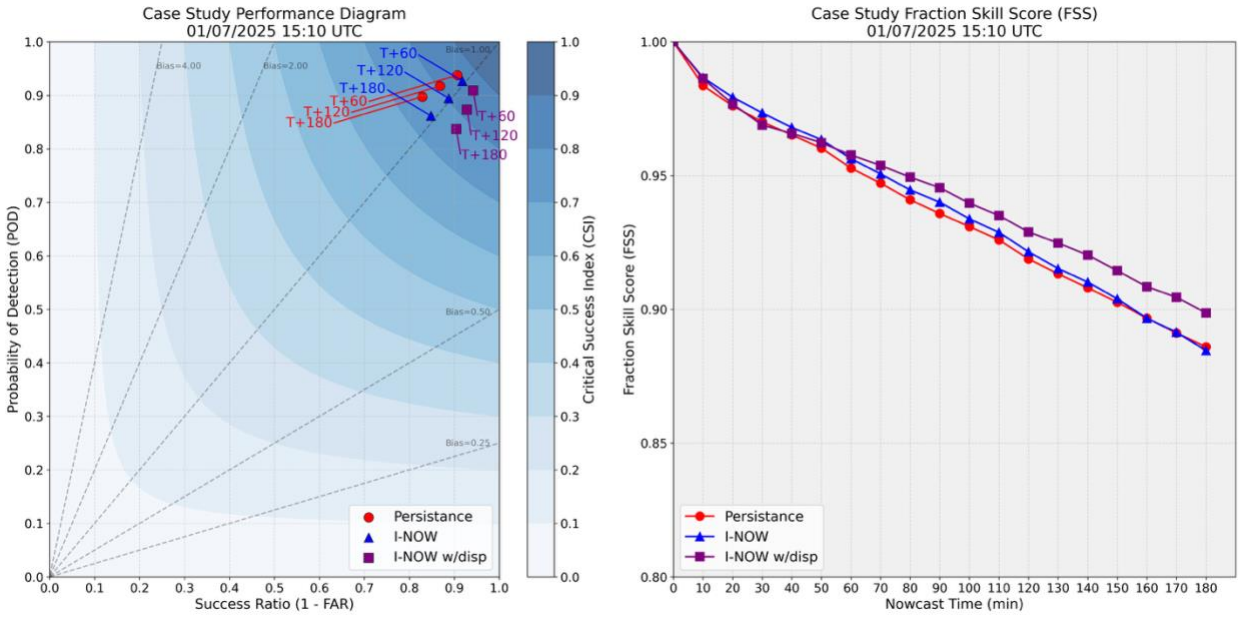

Figure 3.9 shows the case study metrics via performance diagram and FSS. Interestingly for the cloud mask metrics in the performance diagram, persistence has a high bias, the original I-NOW no bias, and I-NOW with dissipation a low bias. All three methods show relatively good skill as they are in the general top right of the performance diagram. Additionally, it appears this case is indeed a difficult scenario in which to outperform persistence as the CSI scores for persistence are higher than the original I-NOW algorithm. In the subplot of FSS, I-NOW and persistence have very close scores. I-NOW with dissipation ultimately has a higher cloud mask FSS which increases beyond both persistence and the original I-NOW algorithm starting at the1-hour nowcast timeframe. Also, its superior performance in FSS increases as the nowcast progresses to the 3-hour nowcast timeframe. This trend is a very clear indicator that dissipating clouds in low-level stratus or fog scenarios is a powerful augmentation for I-NOW.

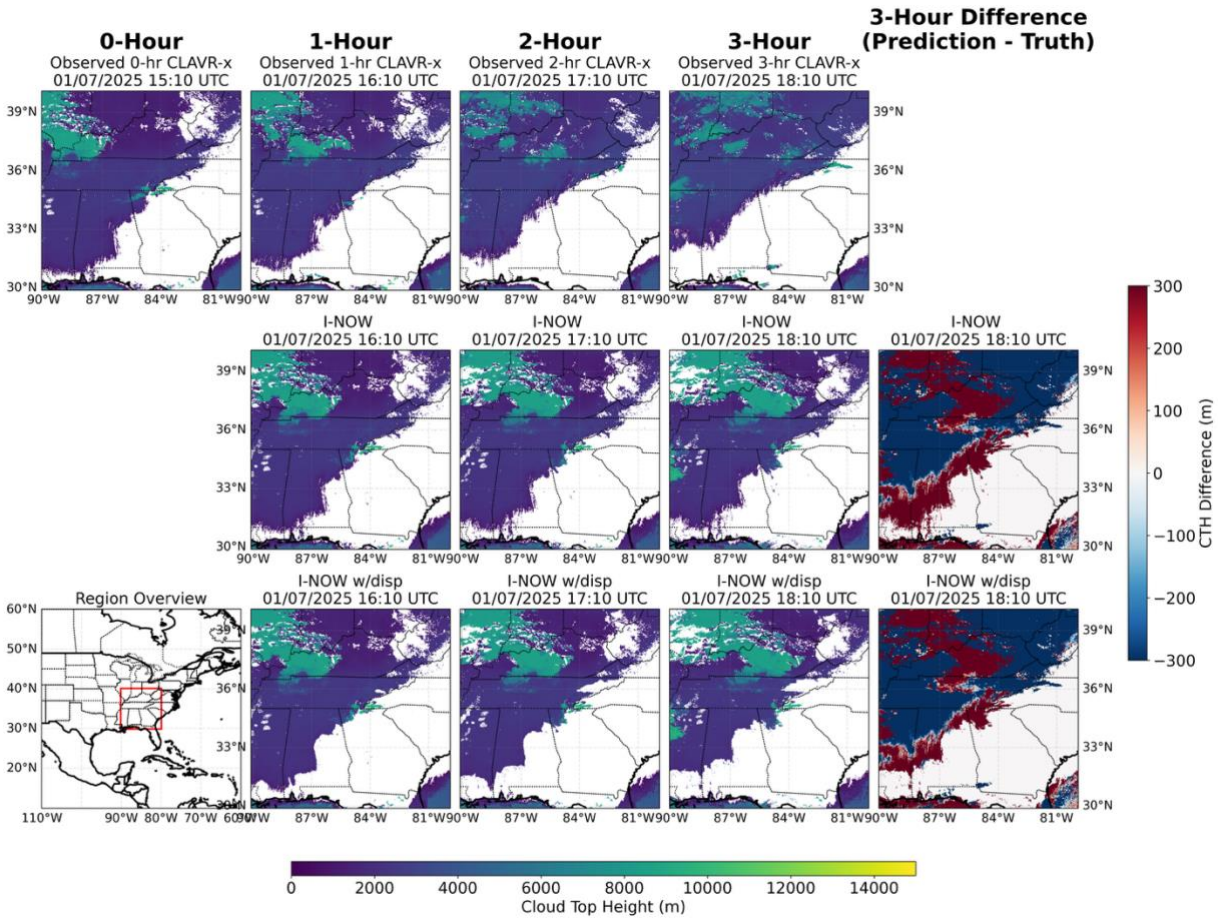

Figure 3.8: Case Study example of I-NOW with dissipation and original I-NOW algorithm from 7 January 2025. The top row shows the observed CLAVR-x CTH at the initial timeframe of 15:10 UTC and 1-hour, 2-hour, and 3-hour nowcast timeframes. The second row shows the CTH nowcasts for the original I-NOW algorithm. The last row shows I-NOW with dissipation along with a subplot on the bottom left indicating the region of the nowcast. The last column on the right shows the 3-hour difference between the original I-NOW algorithm and the observed CLAVR-x CTH and the difference of I-NOW with dissipation and the observed CLAVR-x CTH. Values of observed and nowcast CTHs for difference calculations that were initially "nans" or not a number were set to zero to better show the difference between the nowcasts and CLAVR-x observations.

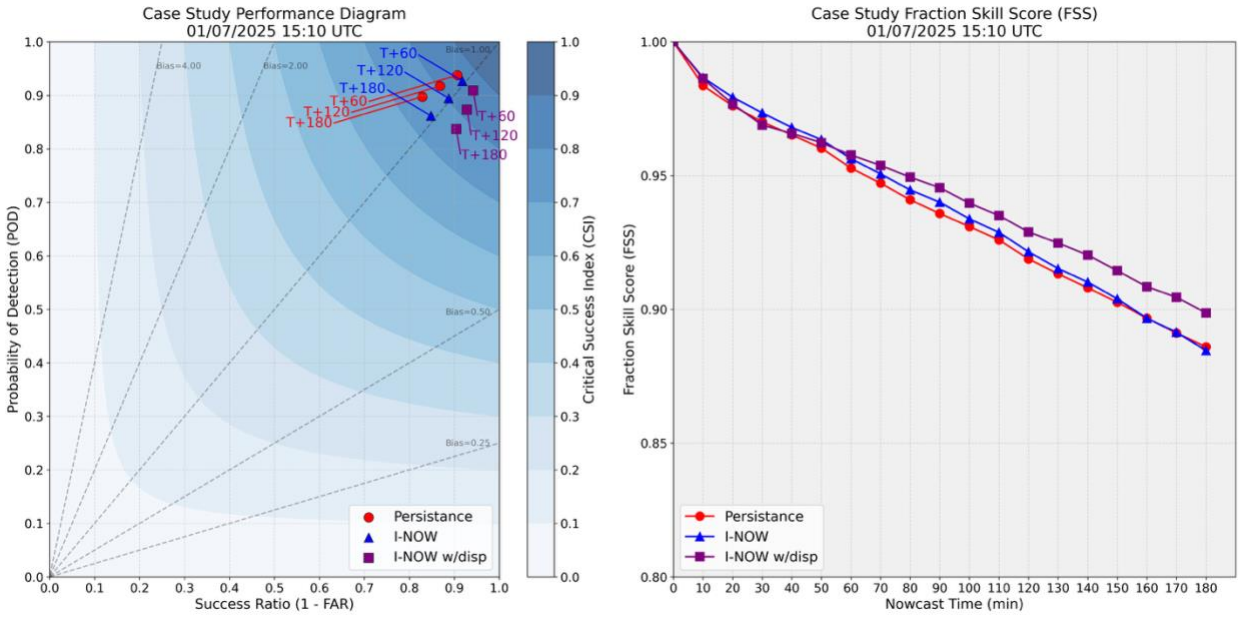

Figure 3.9: Cloud mask metrics for 7 January 2025 case study showing performance diagram (left) and FSS (right).

3.3.4. Ablation Study

To gain insights from the dissipation learned from ML, an ablation study – an experimental analysis where components or inputs of a model are systematically removed or modified to assess their impact on performance – was performed on the dissipation U-Net in which two additional models were trained on the training data but with varying inputs. The first ablation model used the initial CTH and CGT values along with the I-NOW nowcast inputs from 10-, 20-, 30-, 40-, and 50-minute initial timeframes before the initial start time of the nowcast. The second ablation model used the initial CTH and CGT values along with the GFS RH values.

To compare the original trained dissipation U-Net with the ablation models, a simple calculation of intersection over union (IoU; Everingham et al. 2010) was utilized for each class, namely constant cloud (class 0), dissipating cloud (class 1), and no initial cloud (class 2). In IoU, the area

or total of a class where the truth and the predictions overlap are divided by the area or total of both the truth and predictions. The resulting ratio indicates how skillfully a class is predicted.

Table 3.2 shows the results of the ablation study on the dissipation U-Net. It can be concluded that the initial conditions which all three models contain (out of necessity due to the forced predictions in the architecture for class 2) have the most impact on the dissipation class (class 1) due to the small difference between ablation 1 and ablation 2. However, it is also apparent that the I-NOW inputs impart more predictability than the GFS RH inputs. This behavior is likely due to the coarse resolution (both spatially and temporally) of GFS relative to the high resolution of the CLAVR-x retrievals.

Although contributions of I-NOW nowcasts and GFS relative humidity values are relatively small, the ablation study does indicate that there is a positive contribution when they are integrated into the same model based on the higher IoU score for class 1 for the original U-Net.

Table 3.2: Dissipation U-Net ablation study results.

Dissipation U-Net Ablation Study Results			
	Class 0	Class 1	Class 2
Dissipation U-Net	.891	.371	1.0
Ablation 1 (I-NOW inputs)	.891	.366	1.0
Ablation 2 (GFS inputs)	.889	.359	1.0

3.4. Conclusions

This chapter sought to build upon the I-NOW nowcast algorithm in the task of low-level cloud dissipation. By leveraging I-NOW's features of Lagrangian persistence and extrapolation of initial cloud conditions, two U-Nets were trained independently to predict cloud dissipation and timing across a 3-hour nowcast timeframe. Training was enabled by utilizing OF-based nowcasts to create labels for cloud dissipation based on initial conditions of cloud fields.

The following science question was posed in this chapter: To what extent do observed changes in cloud properties and coarse-resolution NWP relative humidity (RH) fields predict future cloud dissipation in the short-term nowcast timeframe (0 – 3 hours)? Two hypotheses followed this question: 1) if there is predictive information from a series of OF-based cloud nowcasts of cloud properties and NWP RH values, then ML will provide measurable skill in predicting short-term cloud dissipation; 2) if short-term changes of nowcasted cloud properties observed across time provide more predictability than NWP RH fields due to the recency of the nowcasts, then an ablation of a ML model that contains nowcasts as input versus an ablation containing NWP RH fields will show more dissipation skill.

In regard to the first hypothesis, it was determined that through a dual U-Net architecture a observed changes in cloud properties and coarse-resolution NWP relative humidity (RH) fields predict future cloud dissipation with a relatively small but statistically significant improvement to nowcast skill.

For the second hypothesis it was determined through an ablation study that the combination of OF-based nowcasts across time and GFS RH fields do impart predictability to cloud dissipation with the OF-based nowcasts providing more predictability. However, the initial CTH and CGT conditions imparted most of the predictability of cloud dissipation.

Although it was demonstrated that OF-based nowcasts using the Lagrangian persistence assumption can enable labeling for cloud dissipation, the work here is only a proof-of-concept. Additional work is required to improve cloud dissipation and is enumerated as follows:

i) To impart more skillful gains for OF-based nowcasts and augmentation through ML dissipation, multiple cloud levels (i.e. mid- and high-level) also need to be addressed. This requires equipping I-NOW or similar methods to better account for multilayer clouds and training additional models that can account for different types of clouds and levels which have often have different physical processes that determine dissipation.

ii) Training data here were limited due to the recency of the available CLAVR-x retrieval database associated with OVERCAST. More training data raises the potential for much better ML model predictions with added training data.

iii) Training time and limited computational resources were major constraints in this study as both limited the ability to tune hyperparameters of the U-Nets. Additionally, because training was performed independently on both models for the sake of reducing training time, the timing U-Net likely introduced prediction variations that were not constrained due to the variations of the dissipation U-Net when its predicted classes were input into the timing U-Net. With enough time and computational resources, the two models can be better fine-tuned.

iv) Although GFS model fields have coarse resolution, they did show some useful predictability for cloud dissipation. Thus, as NWP and even ML-based NWP emulators become more robust and quicker to produce predictions at higher resolutions, it is worth investigating using model inputs to improve on cloud dissipation predictions.

CHAPTER 4: PROBABILISTIC PREDICTION OF CLOUD FORMATION USING OPTICAL FLOW-BASED NOWCASTING AND NUMERICAL WEATHER PREDICTION

4.1. Introduction

Although the previous chapters demonstrated improvements to OF-based cloud nowcasting by accounting for temporal changes in the OF field (Chapter 2) and incorporating cloud dissipation (Chapter 3), forecasting cloud formation remains an inherently (and, perhaps, the most) difficult challenge. Cloud formation, particularly when associated with low visibility and low ceilings, remains among the top weather-related conditions coincident with aviation accidents (Gultepe et al. 2019). For DoD operations, cloud formation can close windows of opportunity, whether for intelligence gathering purposes (Golemboski 2001) or for weapons targeting that is limited by CFLOS (Burley et al. 2019).

Clouds arise from a complex synergy of atmospheric dynamics, thermodynamics, microphysics, radiative processes, and chemical processes. OF-based cloud nowcasting methods, on the other hand, rely on the presence of existing clouds to extrapolate their future positions, offering little ability to deterministically predict the emergence of new clouds in previously clear regions.

Recent advances in ML, particularly diffusion models, have shown promise in satellite-based image nowcasting applications (e.g. Chen et al. 2025; Chase et al. 2025). From a short sequence of images, these techniques have demonstrated an ability to render realistic evolutions of cloud patterns, including the formation of new, realistic-appearing cloud structures. However, such models often trade interpretability and increased computational requirements for gains in predictive skill, posing potential limitations in constrained operational environments. U-Nets (e.g. Kellerhals et al. 2022; Yu et al. 2023) have frequently been utilized in nowcast approaches,

however, when used to make direct nowcast predictions U-Nets often capture formation at the expense of cloud texture and consistent bias.

Arguably, NWP models provide the best representation of the complex microphysics and nonlinear relationships occurring in the atmosphere that are necessary to anticipate the future formation of clouds. NWP models such as NOAA's Rapid Refresh (RAP) and High-Resolution Rapid Refresh (HRRR) attempt to simulate these coupled processes through physics-based calculations (Benjamin et al. 2016), and evolve the environmental state (including prognostic variables for clouds) with time, but require substantial data assimilation and high-performance computing to produce timely, operationally relevant forecasts.

Depending on the scale and strength/organization of the dynamics, NWP models may perform better (strong/synoptic-scale) or worse (weak/mesoscale) in terms of generating clouds at the right places and at the right times. To first order, the extrapolated cloud field from OF-based cloud nowcasts may provide a better representation of the near-term cloud state vis-à-vis a NWP model – especially in the early stages of model spin-up and in situations of weak dynamical forcing. Thus, we consider here the problem of nowcasting cloud formation, and to what extent basic information on the environmental state (e.g. modeled relative humidity) and the extrapolated cloud field can provide useful insight to this problem.

Specifically, this chapter seeks to answer the following question: to what extent does short-term extrapolation of clouds and modeled environmental information improve predictability of cloud formation? This question is accompanied by two hypotheses: 1) if the context of where clouds advect and NWP RH predictions provide cloud formation predictability, then ML will provide measurable skill on where clouds are likely to exist and therefore likely to form; 2) if NWP RH fields contribute to cloud formation more than extrapolated cloud nowcasts, then an ablation of a

ML model that contains NWP RH fields will provide better cloud classification than an ablation that contains only extrapolated cloud nowcasts.

In the previous chapter, a dual U-Net architecture was utilized to predict low-level cloud dissipation to inform the I-NOW OF-based cloud nowcasting method. This approach was successful in that pixels in the initial cloud field could be designated to dissipate at specific nowcast timeframes thereby reducing false alarms and increasing nowcast skill metrics (i.e. CSI and FSS). This chapter seeks to further improve upon those nowcast results by integrating cloud formation predictions to the nowcast where the original I-NOW algorithm predicts cloud free areas.

In DoD contexts, it is critical to consider scenarios in which communication, computational, or observational capabilities may be degraded or denied (Pence 2021). Consequently, rather than rely on computationally intensive approaches such as high-resolution NWP or diffusion-based ML models, we incorporate readily available coarse-resolution GFS global model fields (for relative humidity) and a computationally efficient U-Net model to enhance OF-based nowcasting with probabilistic cloud formation predictions.

This chapter is structured as follows: Section 4.2. describes the data and methods for combining OF-based nowcasts and GFS relative humidity parameters within a U-Net architecture. Section 4.3 presents U-Net test results and a case study on OF-based nowcast integration. Section 4.4 concludes the chapter and identifies areas for future investigation.

4.2. Data and Methods

4.2.1. Overview

Building upon the work showcased in the previous chapters, this chapter attempts to leverage the predictions of I-NOW along with GFS relative humidity fields to determine probabilities of cloud cover. Unlike in previous chapters where nowcasts were evaluated at 10-minute increments,

this chapter evaluates the probabilities determined for a single nowcast timeframe. To be clear, I-NOW nowcasts and temporally interpolated GFS data specific for a particular nowcast timeframe (i.e. 1-hour or 2-hours) are utilized as inputs to ML. The expected output of the ML model are probabilities, ranging from 0 to 1, denoting the likelihood of the existence of a cloud for a particular grid point.

Figure 4.1: Conceptual representation of ML task for cloud probabilities prediction

Full-disk GOES-16 data from CLAVR-x CTH and CGT data were again utilized in I-NOW to produce full-disk nowcasts. The same available CLAVR-x dataset from Chapter 3, 10 April 2023 to 30 December 2024, was utilized for training, validation, and testing. Full-disk grids of both nowcasts and observed CTH and CGT data were bilinearly interpolated onto 512 x 512 pixel rectilinear grids with 0.02° resolution in latitude and longitude. Parallax correction was applied to CTH and CGT values (both nowcasts and observations) using the equations documented by Miller et al. (2018) in the same manner applied in Chapter 2. Grids that contained solar zenith angles exceeding 75° were omitted from the training, validation, and test datasets to again avoid abrupt changes in CLAVR-x retrievals due to a change of retrieval methods from the lack of visible reflectance imagery.

GFS RH values from pressure levels 1000 to 100 hPa were utilized in the ML model input. Just as in Chapter 3 where archived GFS data were only available in 3-hour forecast increments, RH

values were temporally interpolated across time using semi-Lagrangian advection for each pressure level to the nowcast valid timeframe.

The truth data used to train the ML model was the resulting binary cloud field (0 for no cloud and 1 for cloud) based on the parallax corrected locations of the CTH data. This allowed the ML model to predict clouds as a binary classification task where predicted values from the model would range between 0 and 1 where a value of 1 would represent a 100% probability of a cloud and a value of 0 would represent a 0% probability of cloud. Ultimately, with the ML model providing a probabilistic output for the presence of cloud an end-user is provided insight of the likelihood of the existence of clouds and can thereby utilize that probability to weight operational decisions such as targeting or intelligence gathering.

The original I-NOW algorithm, as presented in Chapter 2 (i.e. no ML augmentation), was utilized to produce nowcasts of CTH and CGT valid at 1-hour and 2-hours. Nowcasts were executed at 10-minute increments, meaning the OF field was updated at 10-minute increments, but only the resulting nowcast fields valid at 1-hour and 2-hours were saved for use in the ML model input. The reason for only utilizing 1-hour and 2-hour nowcasts is due to the time required to prepare training data. To be clear, a separate model to predict clouds at the 1-hour and 2-hour timeframes is necessary for this approach. Including every 10-minute increment would require training 18 separate models each 10-minute increment of the nowcast was to be augmented from 10 to 180 minutes. Thus, for this experiment and purposes of forming a proof-of-concept, we limit cloud predictions to only the 1-hour and 2-hour timeframes.

4.2.2. Model Architecture

The U-Net architecture from Chapter 3 was adapted here for the cloud probability task (see Figure 4.2). The key differences between the Chapter 3 U-Net and the new configuration were the

number of input channels, number of output channels, and last layer activation. Input channels, when combining CTH, CGT, and GFS relative humidity values, totaled 23 where the first two channels were the I-NOW nowcasts (CTH and CGT) and the last 21 channels were the RH values from GFS interpolated to the same nowcast timeframe. The output prediction of the U-Net was an array that was 512 x 512 in size and contained only one channel. To arrive at a prediction that provided a probability-like output, the last layer activation function was updated to a sigmoid activation function. The sigmoid activation function, is mathematically expressed as:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{4.1}$$

This sigmoid activation function has values that range between 0 and 1 allowing for implementation in binary classification tasks (Géron 2023). Here, the 50% threshold was used to determine whether the prediction denoted cloud or no cloud. Said another way, a grid point value from the array was predicted to be cloud (no cloud) if it was greater than or equal to (less than) 0.5.

To train the U-Net, binary cross-entropy (BCE;Terven et al. 2025) was utilized as the loss function. The binary cross-entropy loss for a single pixel can be express as:

$$L(y_i, \hat{p}_i) = -[y_i \cdot \log(\hat{p}_i) + (1 - y_i) \cdot \log(1 - \hat{p}_i)] \tag{4.2}$$

where $\hat{p}_i$ is the predicted probability output and y_i is the true label (0 or 1).

It follows that the average BCE loss which is minimized over a dataset of n grid points can be expressed as:

$$BCE_{avg} = \frac{1}{n}\sum_{i=1}^{n} L(y_i, \hat{p}_i) \tag{4.3}$$

To assess whether I-NOW or GFS inputs provide better predictability of clouds and cloud formation, an ablation of the U-Net was performed in which the same architecture was utilized but

with fewer inputs. The second U-Net trained with only I-NOW inputs and a third U-Net trained with only GFS inputs. This ablation exercise enabled the examination of whether short-term extrapolation of cloud cover better enhanced predictability of clouds and new cloud formation versus coarse-resolution global-scale GFS relative humidity fields.

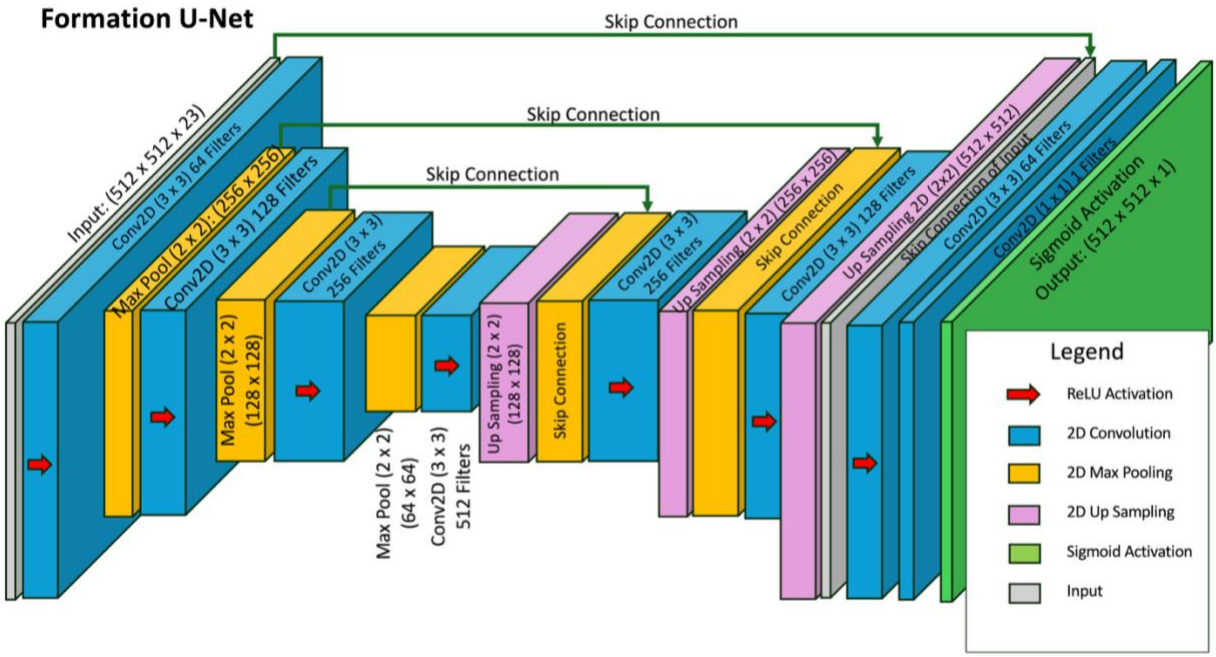

Figure 4.2: U-Net Architecture for cloud probabilities task.

4.2.3. Training, Validation, and Testing Split

To avoid cross-contamination of data between training, validation, and testing datasets, specific days were randomly selected to form unique and independent datasets. Again, as was done in the previous chapter, only daytime data were selected (less than 75° solar zenith angle) to avoid differences in CLAVR-x retrieval algorithms between day and night.

The training, validation, and testing split was done separately for training a U-Net for 1-hour and 2-hour predictions. Consequently, the total number of individual dates used for training,

validation, and testing were slightly different for 1-hour and 2-hours. For the 1-hour datasets, total individual dates numbered 627. Those dates were split based on ~80% for training (500 days), ~10% for validation (63 days), and ~10% for testing (63 days). For the 2-hour datasets, total individual dates numbered 362. Those 362 dates were split based on ~80% for training (289 days), ~10% for validation (36 days), and ~10% for testing (37 days). The reason for the difference of available days was due to the solar zenith angle filtering. For a 1-hour nowcast, there were more days in which a randomly selected grid would have a solar zenith angle that did not exceed 75 °.

As in Chapter 3, 1000, 100, and 100 nowcast initial timeframes were randomly selected for training, validation, and testing respectively. From those timeframes, which represented full-disk nowcasts, 10 rectilinear, 512 x 512, 0.02° resolution grids were randomly selected based on a center latitude/longitude value that remained within 100°W to 40° W longitude and 50° S to 50° N latitude. This grid selection ultimately resulted in the 1-hour and 2-hour datasets to each consisting of 10,000 samples for training, 1,000 samples for validation, and 1,000 samples for testing.

4.2.4. Evaluation metrics

To evaluate the ML model predictions, performance diagrams and area under the curve (AUC) of receiver operating characteristic (ROC) curves were utilized (see Figure 4.5 and Figure 4.6). The performance diagrams followed the same approach and metrics as done in Chapters 2 and 3, where POD, SR (1 – FAR), and CSI are required. ROC curves are calculated from the relationship between the probability of false detection (POFD, occasionally called False Alarm Rate) and POD (Chase et al. 2022). Just as POD, CSI, and FAR were defined using a 2 x 2 contingency table as mentioned in Chapter 2, POFD is also described with a contingency table. When using the

straightforward terms of true positive (correct cloud prediction), false positive (incorrect cloud prediction), and false negative (incorrect no-cloud prediction), POD and POFD are defined as:

$$POD = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{4.4}$$

$$POFD = \frac{False\ Positive}{True\ Positive + False\ Positive} \tag{4.5}$$

AUC of ROC curves provide an area by varying threshold probability from 0% to 100% rather than selecting a specific probability threshold to which a prediction decision can be determined. This results in a curve in which the area underneath can be calculated to discern predictive skill. An area of 0.5 is equivalent to a random classifier, whereas an area that is closer to 1.0 is an objectively better classifier.

4.3. Results

4.3.1. Case Study Results

Prior to investigating the results of the 1-hour and 2-hour U-Nets and their ablations from the test dataset, we revisit the 7 Jan 2025 case study presented in the previous chapter to examine how the ML models may be forming clouds and to what extent they improve upon the version of I-NOW equipped with dissipation. To be clear, the 1-hour and 2-hour U-Nets are used in this case study to update the final prediction 1-hour and 2-hour predictions of I-NOW with dissipation, thereby creating a final version I-NOW prediction equipped with cloud dissipation and formation capability.

Figure 4.3 shows the 1-hour and 2-hour results of the original I-NOW nowcast, I-NOW with dissipation, and I-NOW with dissipation and formation from the 1-hour and 2-hour U-Nets for the

case study. As a reminder, this case study lies outside of the training, validation, and testing datasets.

To produce this case study, the original I-NOW 1-hour and 2-hour nowcasts along with the appropriate temporal interpolated GFS RH values were input into each respective U-Net. CTH pixels were filled using an outside-in averaging strategy in which pixels designated to have cloud formation took on the properties of the average of the surrounding pixels. Cloud-free pixels were designated for cloud formation where the resulting probability fields indicated that a cloud should be formed (i.e. probability of 50% or higher).

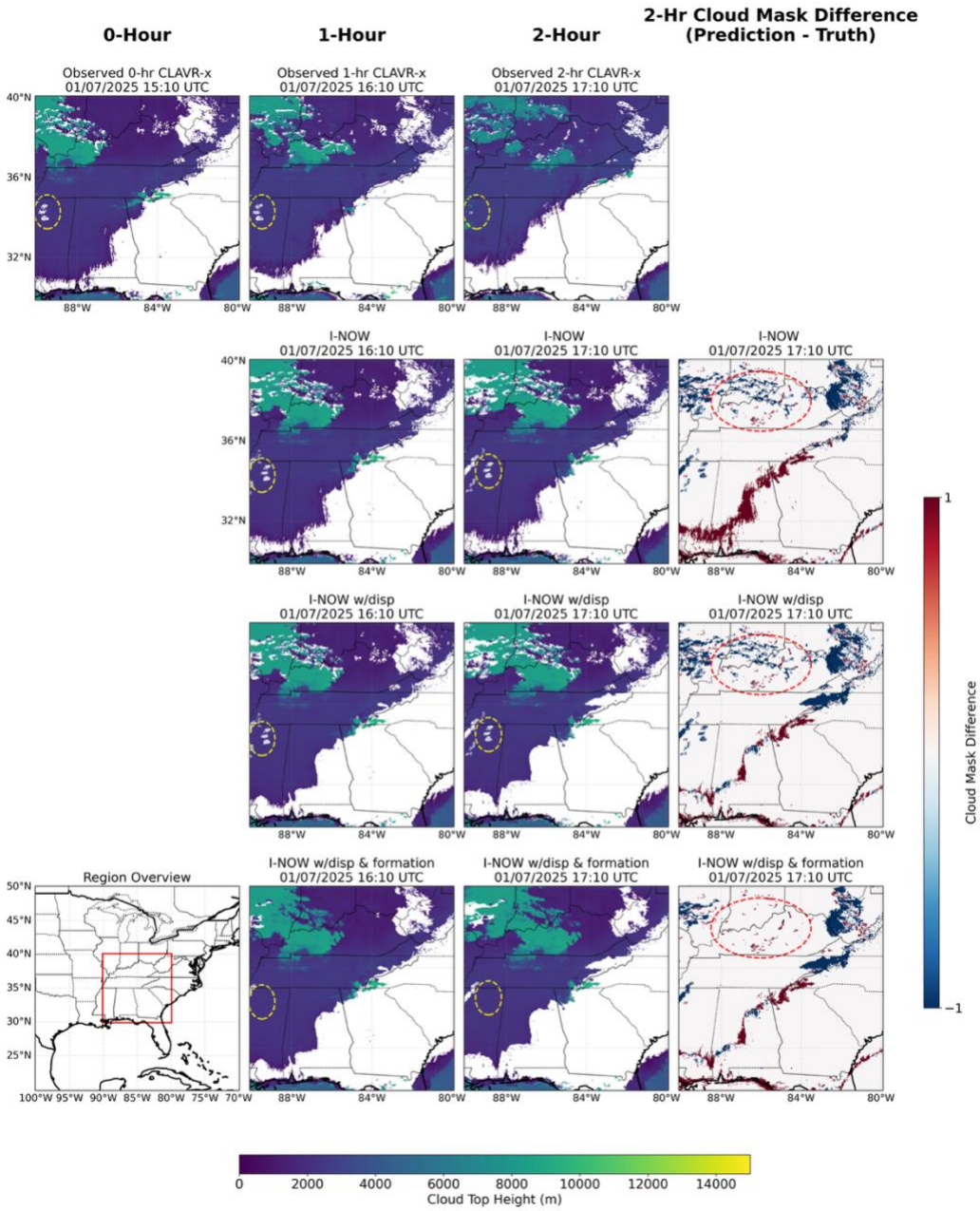

Figure 4.3: 7 Jan 2025 case study which includes observed CLAVR-x CTH (top row), original I-NOW algorithm (2nd row), I-NOW with dissipation (3rd row), and I-NOW with dissipation and formation (bottom row). 2-hour cloud mask differences are presented in the last column (right) where blue indicates the nowcast had no cloud, but cloud was observed and red indicates the nowcast had cloud when no cloud was observed.

The right column of Figure 4.3, which shows the 2-hour cloud mask difference (prediction – truth), is annotated with a red-hashed oval indicating the area where the 2-hour U-Net has clearly produced cloud formation. Areas of the difference plot that are blue (negative values) indicate where the nowcast predicted no cloud, but cloud was observed. Areas of the difference plot that are red (positive values) indicate areas where the nowcast predicted cloud, but cloud was not observed. Thus, blue areas that appear in the top and middle difference plots, but do not appear in the bottom difference plot are areas where cloud was successfully formed by the U-Net.

In contrast, the large area of blue in the difference plot over the border of Kentucky and West Virginia shows a large area where the U-Net failed to produce cloud. This is evidence that there is a definite limit to which the extrapolated cloud field in combination with GFS RH fields can offer predictive information on cloud formation. Likely when there are larger cloud free areas in the extrapolated field, the GFS RH field lacks the information to inform the U-Net that there is likely cloud formed in the area.

Another prominent feature in the case study is indicated in the observed CLAVR-x CTH and nowcast fields with a hashed yellow oval. In the initial observed subplot (top left), there are three clear areas in the CTH field. These three clear areas are an error in CLAVR-x retrievals due possibly to a missing imagery channel. By the 2-hour timeframe, the error is no longer observed in the retrieval (top right). Evidence that the cloud free areas are in fact a retrieval error is observed in the original I-NOW nowcast (second row) and the I-NOW nowcast with dissipation (third) row. Because there is indeed cloud in those areas which has a non-zero velocity, the nowcast algorithm extrapolates those clear areas to move towards the east. In both the 1-hour and 2-hour I-NOW plots with dissipation and formation, those cloud free areas have been designated for formation and have

been filled accordingly. Although not actually cloud formation, the U-Nets successfully correct the nowcast to rightly indicate there should be cloud in those areas.

An objective evaluation of the resulting cloud masks from the case study using a performance diagram and FSS are presented in Figure 4.4. In the performance diagram, it is clear the U-Nets have had a positive impact on the nowcast skill for both 1-hour and 2-hour timeframes when compared to the original I-NOW algorithm and I-NOW with dissipation since I-NOW with dissipation and formation achieves both higher CSI, POD, and a bias closer to 1.0. For the FSS plot, a similar indication of increased skill is evident as both the 1-hour and 2-hour timeframes indicate an increase in FSS compared to the original I-NOW algorithm and I-NOW with dissipation. In fact, the 2-hour timeframe has a greater increase in FSS than the 1-hour timeframe when compared to the previous versions of I-NOW.

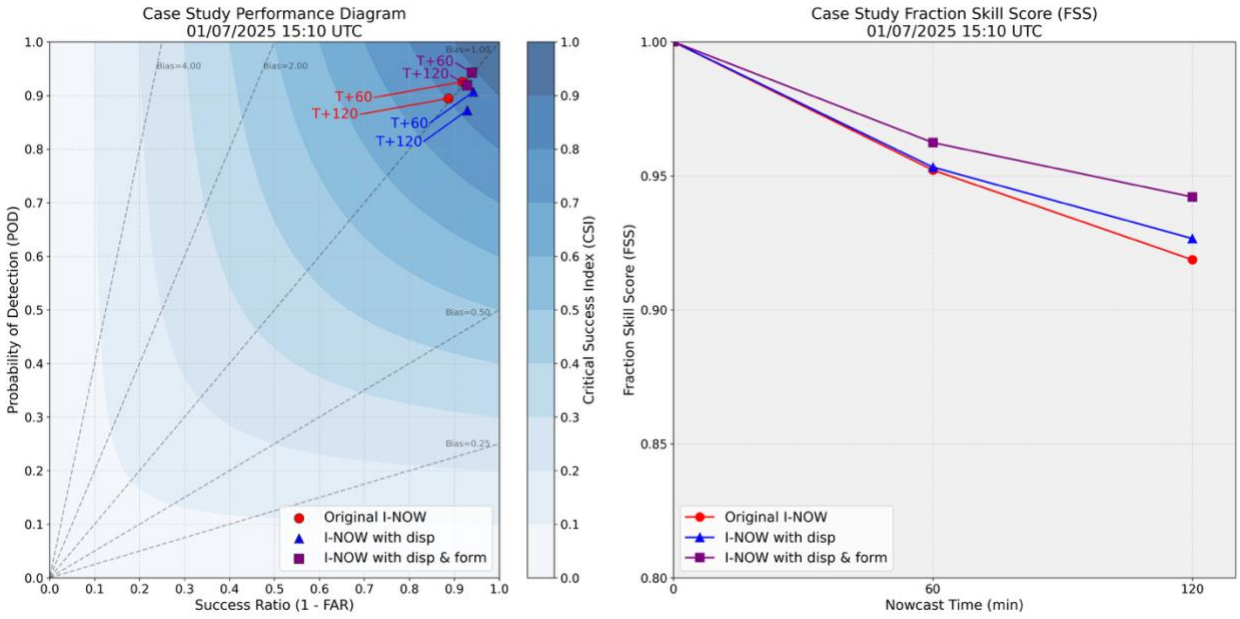

Figure 4.4: Cloud mask results from 7 Jan 2025 case study. Performance diagram (left) provides POD, SR, and CSI results for original I-NOW algorithm (red circles), I-NOW with dissipation (blue triangles), and I-NOW with dissipation and formation (purple squares). FSS (right) is presented for 1-hour and 2-hour nowcast timeframes for original I-NOW algorithm (red circles), I-NOW with dissipation (blue triangles), and I-NOW with dissipation and formation (purple squares).

4.3.2. 1-Hour Test Data Results

With the above case study indicating increased skill from the trained U-Nets, the results of the models and their ablations on the test dataset can provide insight into how the inputs are impacting the predictability of clouds and cloud formation. The following sections analyze the performance of the U-Nets on the test datasets using performance diagrams and AUC of ROC curves. Specifically, the analysis separates prediction performance for all clouds (cloud prediction), and prediction performance for clouds in cloud-free I-NOW prediction pixels (cloud formation) which were used to enhance the I-NOW predictions in the above case study.

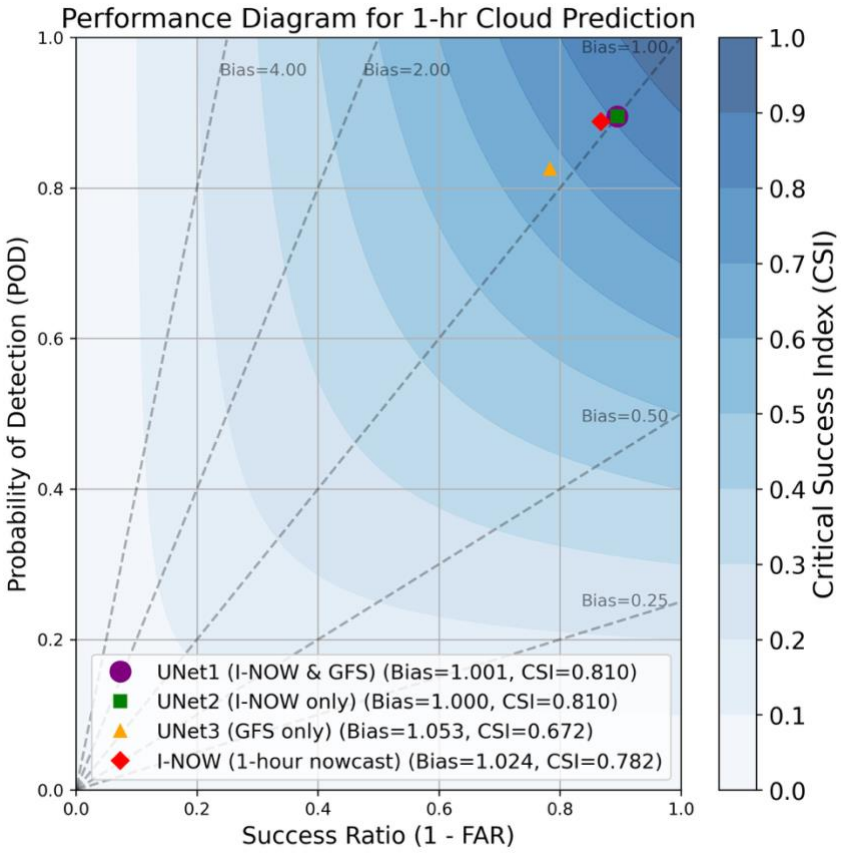

Figure 4.5: Performance diagram for 1-hour cloud prediction for U-Net, its ablations, and the original version of I-NOW. Bias and CSI values are included in the legend. The full U-Net is plotted in blue, the first ablation with only I-NOW inputs is plotted in orange, the second ablation with only GFS inputs is plotted in green, and the original 1-hour I-NOW cloud prediction is plotted in red.

Figure 4.5 shows the 1-hour nowcast performance diagram results on the test dataset of the U-Net and its ablations along with the nowcast of I-NOW. Interestingly, the full U-Net and its ablation version with only I-NOW inputs have nearly identical performance on the diagram with CSI of .810 and biases near or at 1.0. This performance is an improvement to the original I-NOW algorithm which has a higher bias and lower CSI. The ablation version with only GFS inputs shows a lower CSI and a higher bias than all the other methods. This is a clear indication that the I-NOW nowcast is providing more predictive information to the U-Net for the 1-hour nowcast timeframe.

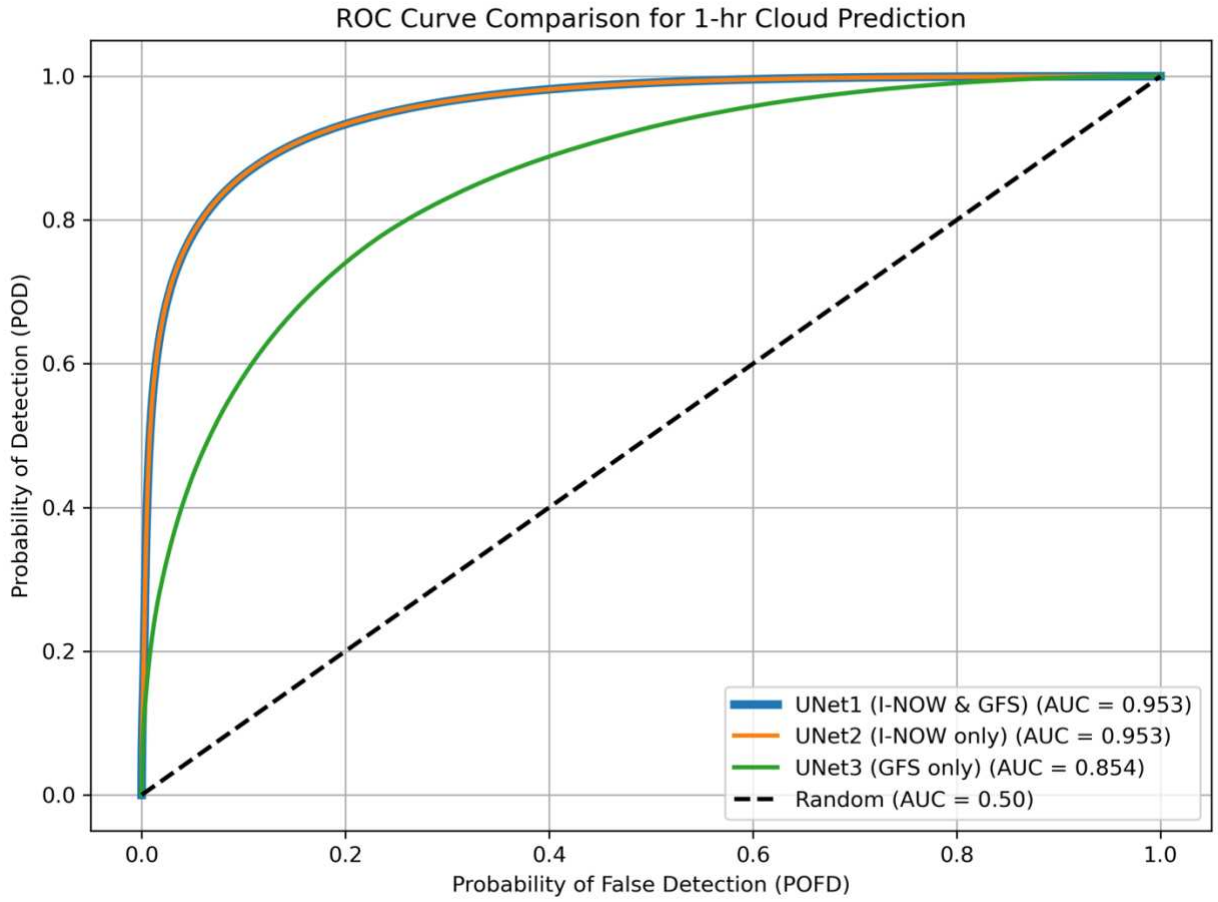

Figure 4.6: AUC of ROC curve for 1-hour cloud prediction for the full U-Net and its ablations. The full U-Net is plotted in blue, the first ablation with only I-NOW inputs is in orange, the second ablation is in green, and the plot of a random classifier is in black. The I-NOW prediction is not included as its deterministic cloud prediction does not allow the varying of probability thresholds.

Figure 4.6 provides the AUC of the ROC curve for the different 1-hour nowcast U-Net versions. Again, the full U-Net and the ablation with only I-NOW inputs have the same prediction skill with AUC values of .953. In contrast the GFS only U-Net only achieves an AUC value of .854. All versions, however, show clear distinction and superior performance to a random classifier.

To determine skill on cloud formation, the grid points in which I-NOW did not predict cloud were isolated and the evaluation performed in Figure 4.5 and Figure 4.6 was repeated. This isolated the prediction evaluation to the areas where cloud formation may have occurred. Figure 4.7 shows the 1-hour nowcast performance diagram results for the grid points where I-NOW did not have a cloud in its prediction. Rather than a large distinction in skill scores, there is a clear trade-off between underprediction (low bias) of cloud formation (i.e. full U-Net and U-Net with only I-NOW inputs) and overprediction (high bias) of cloud formation (i.e. U-Net with only GFS inputs). There is, however, slightly higher CSI scores for the full U-Net and the U-Net with only I-NOW inputs versus the U-Net with only GFS inputs which does indicate that the extrapolation of cloud movements is providing context on some areas where clouds may likely form.

Figure 4.8 shows the AUC of ROC curves for isolated clear I-NOW grid points. The figure shows a clearer indication of skill than the performance diagram with the full U-Net (both I-NOW and GFS inputs) and the U-Net with only I-NOW inputs. In particular, the version with only I-NOW inputs performed with nearly identical AUCs, but clearly lower than corresponding curves in Figure 4.6. for all grid points.

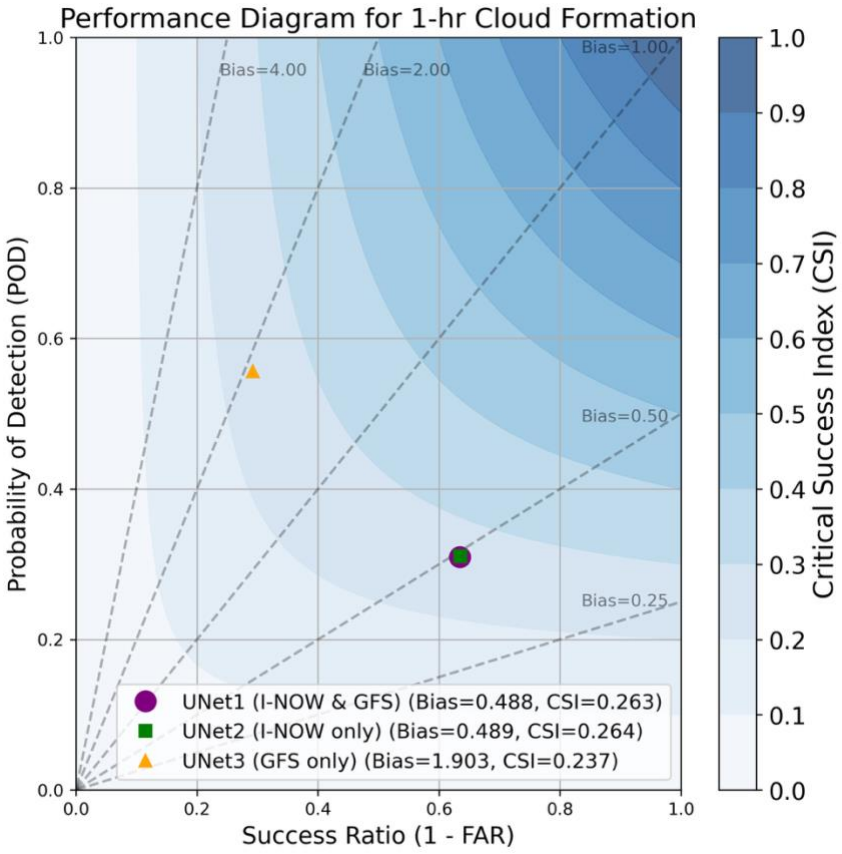

Figure 4.7: As in Figure 4.5, but for grid points where I-NOW nowcast had no-cloud.

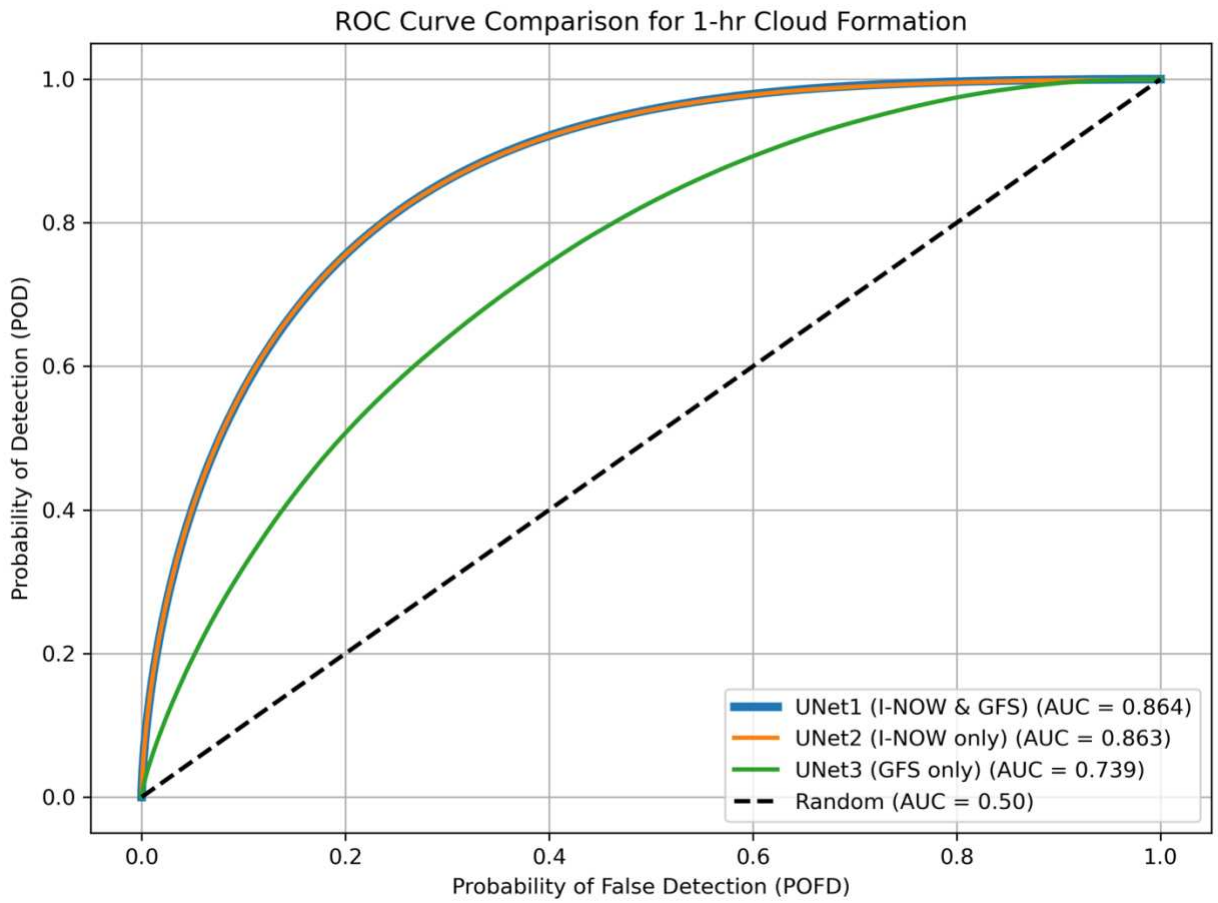

Figure 4.8: As in Figure 4.6, but for grid points where I-NOW nowcast had no-cloud.

4.3.3. 2-Hour Test Data Results

Following the analysis for the 1-hour predictions, the analysis for the 2-hour predictions was conducted in the same manner. Figure 4.9 and Figure 4.10 show the performance diagram and AUC of ROC curves for all grid points, and Figure 4.11 and Figure 4.12 show the performance diagram and AUC of ROC curves for no cloud I-NOW grid points respectively.

The performance diagram for all grid points indicates much of the same results as the 1-hour predictions in that the full U-Net and the U-Net with only I-NOW inputs are nearly identical, and both versions outperforming the original I-NOW 2-hour nowcast. The U-Net with GFS inputs

again underperformed versus the original I-NOW, but less so than in the 1-hour case (comparing with Figure 4.5.

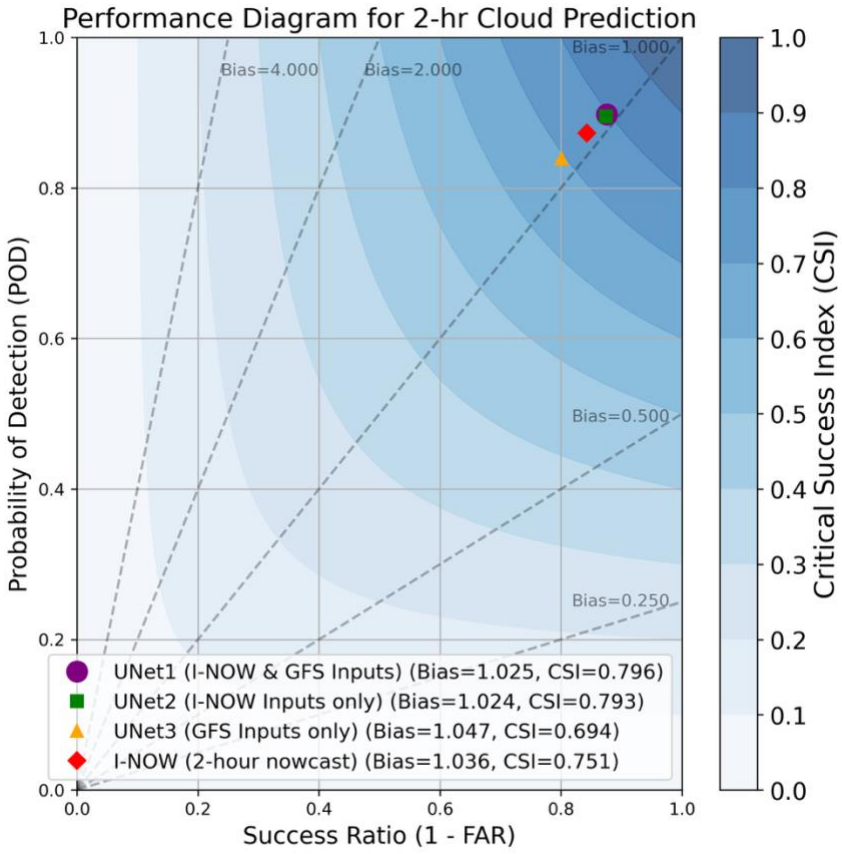

Figure 4.9: As in Figure 4.5, but for 2-hour results.

The AUC of the ROC curves in Figure 4.10 show a similar result. However, the U-Net with only I-NOW inputs shows slightly inferior performance to the full U-Net version. Additionally, while the 2-hour predictions show a reduction in AUC for both the full U-Net and the U-Net with only I-NOW inputs, the U-Net with only GFS inputs increases its AUC from 1-hour to 2-hours indicating that the spin-up time of the GFS may be contributing to lower short-term prediction skill. Likely with a farther nowcast prediction timeframe (i.e. $\geq$ 3 or more hours, beyond the

OVERCAST nowcasting window), predictions from GFS only inputs will eventually outperform predictions from extrapolations.

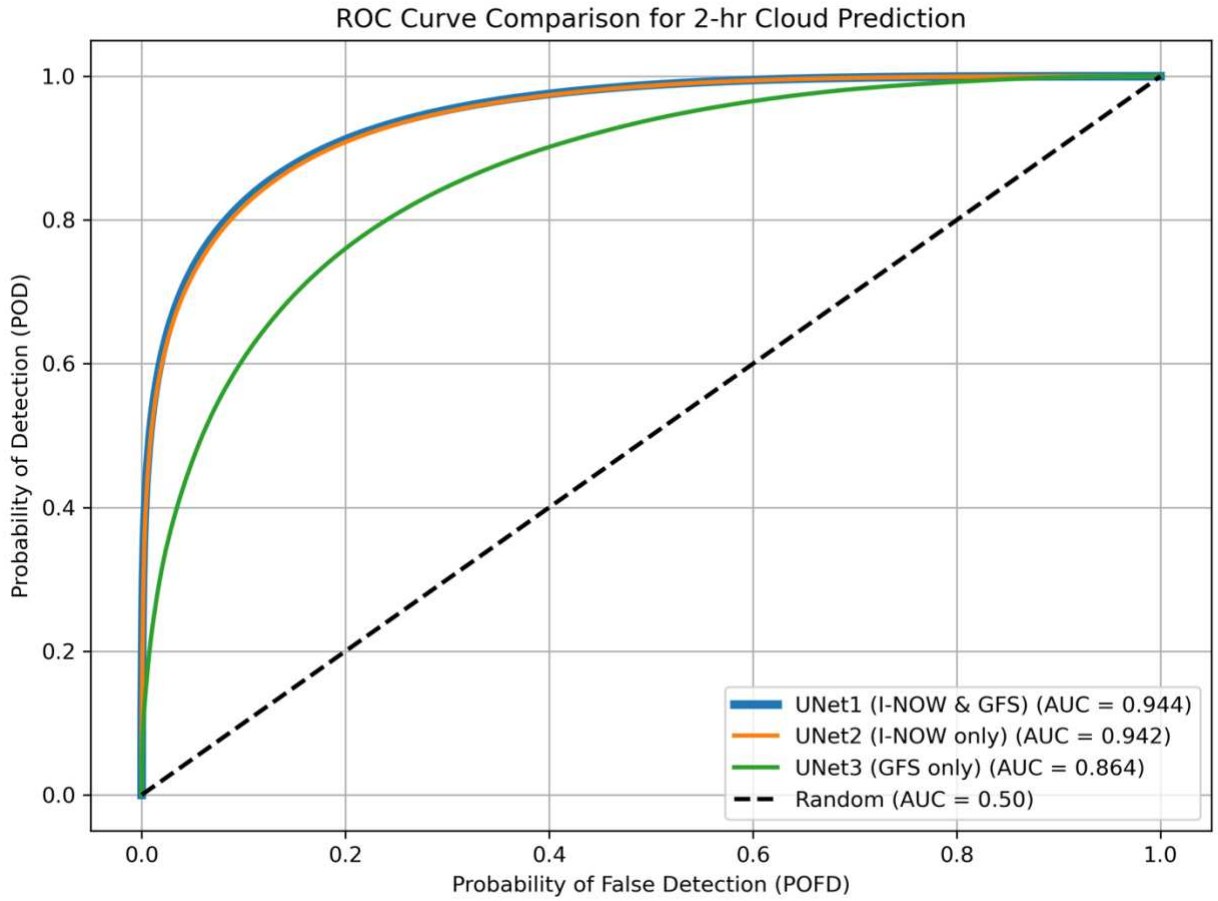

Figure 4.10: As in Figure 4.6, but for 2-hour results.

The 2-hour performance diagram for isolated grid points (Figure 4.11) shows similar results to the 1-hour results with a clear trade-off of overprediction with the coarse resolution GFS inputs and underprediction with extrapolation inputs. However, CSI scores remained higher for I-NOW only inputs compared to GFS only inputs. In contrast to the 1-hour results, the full U-Net showed increased skill for cloud formation due to the addition of the GFS inputs.

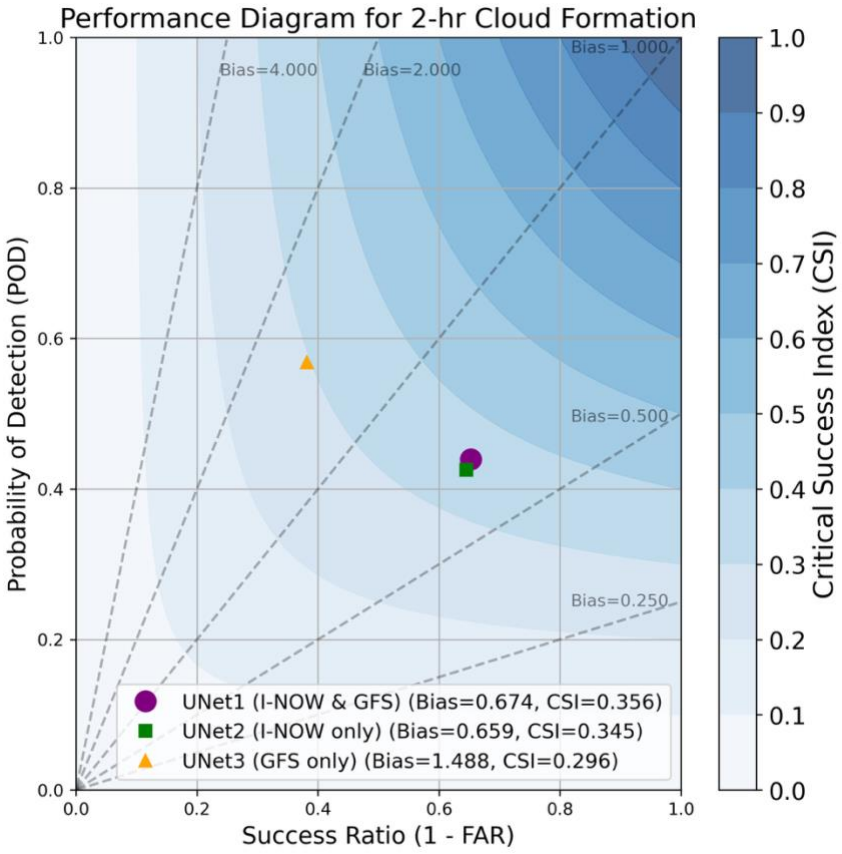

Figure 4.11: As in Figure 4.7, but for 2-hour results.

Similarly, the AUC of the ROC for isolated grid points (Figure 4.12) also indicates an increase in contribution from GFS inputs with higher AUC for the full U-Net verses the U-Net with only the I-NOW inputs.

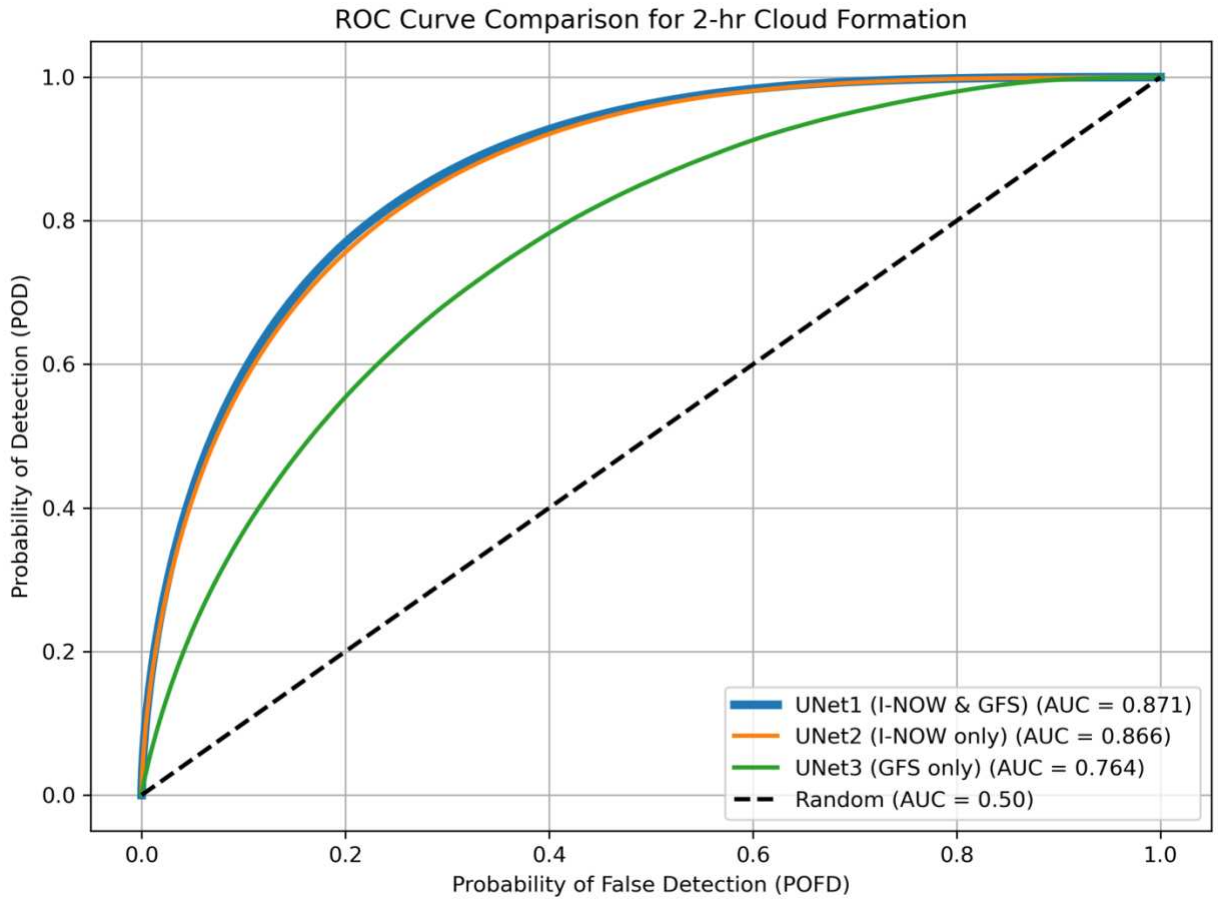

Figure 4.12: As in Figure 4.8, but for 2-hour results.

4.4. Conclusions

Although cloud extrapolation does not lend itself to straightforward prediction of cloud formation, this chapter sought to leverage OF-based cloud nowcasts along with readily available coarse, global-scale NWP data to partially resolve the cloud formation shortfall that all extrapolation- or OF-based cloud nowcasting methods inherently have. Specifically, it was shown that the Improved Nowcasting via Optical flow Warping or I-NOW method for cloud nowcasting presented in Chapter 2 in combination with temporally interpolated GFS RH fields could increase

the skill of cloud prediction as well as new cloud formation when utilized as input into a ML model (i.e. U-Net).

The following science question was posed in this chapter: to what extent does the extrapolation of clouds and modeled environmental information improve predictability of cloud formation? Two hypotheses followed this question: 1) if the context of where clouds advect and NWP RH predictions provide cloud formation predictability, then ML will provide measurable skill on where clouds are likely to exist and therefore likely to form; 2) if NWP RH fields contribute to cloud formation more than extrapolated cloud nowcasts, then an ablation of a ML model that contains NWP RH fields will provide better cloud classification than an ablation that contains only extrapolated cloud nowcasts.

For the first hypothesis it was determined that the combination of extrapolated clouds and NWP RH fields do provide some information, but large cloud free areas do not provide enough context for ML to learn where clouds are to form.

For the second hypothesis it was determined that I-NOW nowcasts of CTH and CGT both for 1-hour and 2-hour timeframes provided more predictability to clouds and new cloud formation than temporally interpolated GFS relative humidity fields. However, there is an indication that as nowcast timeframes extend beyond 2-hours, GFS relative humidity fields become more important in predicting clouds and cloud formation. This is likely because of model spin-up time and the increasing inaccuracies of nowcast methods at longer prediction timeframes.

Ultimately, with the practicality of both OF-based extrapolation methods and NWP as inputs to ML for cloud nowcasting, it follows that there exist natural pathways for future research, as captured in the following observations and associated questions:

1) First, we applied an averaging method that took the properties of surrounding cloud to fill cloud free areas in the case study; however, this may or may not be a good assumption when extended to a 3D cloud field. Future work would be required to validate whether this averaging method translates nowcast skill gains to 3D validation.
2) Second, with increasing prediction timeframes, NWP becomes more important as Lagrangian persistence assumptions fail to predict cloud evolution, so at what timeframe does NWP, even at coarse resolutions, become more predictive of clouds versus OF or extrapolation-based nowcast methods?
3) Third, with the rapid development of ML-based NWP emulators, could such emulators provide more accurate cloud nowcasts when combined with OF-based cloud nowcasts?
4) Finally, because only coarse resolution NWP RH fields were utilized here to provide environmental context for ML, higher resolution NWP may provide more predicative information for ML to learn. This would be an additional reason to run higher resolution NWP models for larger domains.

CHAPTER 5: CONCLUSION

5.1. Summary

Successful cloud nowcasting must overcome three challenges simultaneously: 1) trajectory accuracy, 2) computational efficiency, and 3) cloud formation and dissipation. Traditional forms of OF-based cloud nowcasting offer accurate initial trajectories and computational efficiency but lack the ability to solve cloud formation and dissipation and often quickly diminish in accuracy with time. ML methods, although promising, have yet to solve all three challenges without sacrificing prediction sharpness, interpretability, or computational efficiency. NWP models offer the best representation of the known physics of cloud formation and dissipation, but due to their computational costs, long periods between updates, and spin-up inaccuracies they often fail to provide more accurate predictions than observational-based nowcasts in the short-term (0-3 hours). Thus, this dissertation sought to develop a cloud nowcasting method that merged OF-based nowcasting, NWP data, and ML together to better solve all three challenges.

In seeking to find a better solution for cloud nowcasting, this dissertation explored three science questions:

1) To what extent does accounting for temporal changes in a piecewise OF field improve predictions of 3D cloud movements?

   Hypotheses: 1) if the OF field is updated to account for time related changes for each nowcast time step, then there will be an increase in cloud nowcast skill when compared to a linear extrapolation method in which the OF field is assumed constant with respect to time, which is a common approach in OF-based nowcast applications; 2) if the same apparent motion observed from cloud top height (CTH) is also applied to cloud base

heights (CBH), then the resulting 3D cloud nowcast will be more accurate than a method that applies estimated wind speeds for individual levels.

2) To what extent do observed changes in cloud properties and coarse-resolution NWP relative humidity (RH) fields predict future cloud dissipation in the short-term nowcast timeframe (0 – 3 hours)?

   Hypotheses: 1) if there is predictive information from a series of OF-based cloud nowcasts of cloud properties and NWP RH values, then ML will provide measurable skill in predicting short-term cloud dissipation; 2) if short-term changes of nowcasted cloud properties observed across time provide more predictability than NWP RH fields due to the recency of the nowcasts, then an ablation of a ML model that contains nowcasts as input versus an ablation containing NWP RH fields will show more dissipation skill.

3) To what extent does the extrapolation of clouds and modeled environmental information improve predictability of cloud formation?

   Hypotheses: 1) if the context of where clouds advect and NWP RH predictions provide cloud formation predictability, then ML will provide measurable skill on where clouds are likely to exist and therefore likely to form; 2) if NWP RH fields contribute to cloud formation more than extrapolated cloud nowcasts, then an ablation of a ML model that contains NWP RH fields will provide better cloud classification than an ablation that contains only extrapolated cloud nowcasts.

Chapter 2 introduced the Improved Nowcasting via Optical flow Warping or I-NOW method that utilized a temporal interpolation approach inspired by the advances from the computer vision community to better account for the temporal changes of retrieved OF fields. It was determined that using temporal interpolation in conjunction with simple occlusion reasoning, satellite imagery

and retrieved cloud properties could be better extrapolated forward while accounting for occlusions and piecewise OF fields. With regard to the first hypothesis, it was determined that when accounting for the temporal changes to the OF field using the I-NOW method, nowcast skill improvement was statistically significant over methods using traditional linear extrapolation and no occlusion reasoning. Likewise, for the second hypothesis it was determined for the context of 3D cloud nowcasting, the assumption of utilizing the same flow field for both CTH and CBH, was more accurate than utilizing NWP winds at multiple levels.

Chapter 3 described a dual U-Net approach to predict which low-level clouds observed in an initial cloud field would dissipate during the nowcast timeframe (0-3 hours) and at which specific 10-minute increment they would dissipate. This solution leveraged I-NOW to create ML training labels based on its OF-based nowcasts. In addressing the first hypothesis to the second science question, it was demonstrated that statistically significant improvement to nowcast skill can be achieved by utilizing nowcasts across time as well as GFS RH values as input to ML. In response to the second hypothesis of the same question, it was determined that the combination of OF-based nowcasts across time and GFS RH fields do impart predictability to cloud dissipation with the OF-based nowcasts providing more predictability. However, the initial CTH and CGT conditions imparted most of the cloud dissipation predictability.

Chapter 4 investigated the use of probabilistic prediction of clouds to augment I-NOW output with cloud formation capabilities. For the third science question, it was determined that nowcast skill could indeed be increased by adding clouds to the nowcast based on probabilities produced by ML using both OF-based nowcasts and GFS RH values. However, regarding the final hypothesis, it was determined that OF-based nowcasts imparted more predictability than GFS RH values. Despite that result, it is important to note that with longer prediction timeframes NWP data

contributes incrementally more to the probability prediction and at some point is expected to exceed the contribution possible from OF-based nowcasts.

Together, these chapters demonstrate that integrating temporal interpolation and data-driven cloud dissipation and cloud formation prediction into an OF-based framework meaningfully advance short-term cloud nowcasting. The methods developed here show that it is possible to achieve better balance among computational efficiency, trajectory accuracy, and cloud evolution by merging OF-based nowcasting, NWP, and ML while limiting the need to sacrifice interpretability and prediction sharpness. While further work is needed to extend these methods to broader operational contexts and enhance the accuracy of ML components, the findings of this dissertation provide a solid foundation for future developments in operationally relevant cloud nowcasting.

5.2. Limitations

Despite the breakthroughs presented here, there are key limitations to the methodology as presented. Many of the limitations were previously mentioned, but we briefly summarize them below for the sake of consolidation.

First, I-NOW does not account for acceleration-related changes within the OF field. This becomes an issue in rotational and divergent scenes which the temporal interpolation and OF hole filling cannot fully remedy as currently implemented. One potential improvement could include taking multiple OF retrievals prior to a nowcast initial time to estimate higher-order changes in the OF field and then update flow values across time. As previously mentioned in Chapter 2, the challenge to this approach is determining how to include the higher-order changes to the warped OF field in each warping iteration. One potential approach could include estimating higher order

changes by taking the difference between optical flow retrievals from prior times and the latest OF retrieval and adding those changes to each warp iteration.

OCTANE, the OF algorithm driving the I-NOW method, does have issues resolving flow in areas of multiple cloud layers as evidenced in the last case study of Chapter 2. Potentially, integrating CTH into the OCTANE algorithm could help in resolving the retrieval shortfall in multilayer scenes.

Another limitation is that I-NOW in its current form does not retain multiple cloud layers. To remedy this, multiple layers (i.e. low, middle, and high cloud layers) could be warped separately which would allow some "memory" of clouds obscured by occlusions. Such computations could be performed given I-NOW's computational efficiency over large- to global-domains as required by OVERCAST.

For the ML components of the framework, a key limitation was the limited amount of available data from CLAVR-x retrievals. Continued curating of the current OVERCAST database of CLAVR-x data is necessary to make the amount of training data more robust. It is expected, however, that as work for OVERCAST continues, more data will become available and will enable ML to incrementally improve in accuracy.

Lastly, although some cloud could be removed or added to account for inaccuracies in the I-NOW nowcasts, ML failed to produce cloud predictions where there are large cloud-free areas and at times missed key areas of cloud dissipation. Potentially, ML architecture could be adapted to better resolve cloud prediction by increasing the size of the convolutional filters in the U-Nets. However, it may be more fruitful to incorporate higher resolution NWP fields which may provide more accurate environmental parameters. Considering that the coarse resolution of GFS likely

limited predictability, higher resolution model data may provide more accurate parameters and therefore better ML predictions.

5.3. Future Work

The nowcasting framework presented in this dissertation lays the foundation for a range of follow-on investigations. With regards to I-NOW, it was implemented using full-disk imagery out of necessity for OVERCAST research requirements. However, it would be expected that imagery at higher cadences will likely provide more accurate OF retrievals and subsequently more accurate nowcasts. If this is proven, it could bolster reasoning for higher temporal resolution satellite imagery in future missions. For ML dissipation, low-level clouds were the focus in this dissertation. However, dissipation of mid- and high-level clouds remains unresolved. Additional work will be required to solve dissipation for varying cloud types, particularly for higher clouds which typically are composed of ice. The differences in the microphysical processes of ice clouds may cause variations in cloud dissipation predictability (whether dissipation happens in nature or due to detectability issues in CLAVR-x) when compared to liquid water-based clouds and may offer a more difficult challenge within the context of cloud nowcasting.

Lastly, with the advances of ML and even NWP, the inclusion of higher resolution and more accurate environmental fields appears to be a key avenue for investigation. Given NWP continues to be developed and run at higher resolutions and larger domains with more robust physics, ML may find better predictability in cloud evolution if provided with more accurate environmental contexts. Additionally, as ML-based NWP emulators become more comparable with and efficient than physics-based NWP models, their emulations may provide timely environmental data to better account for cloud evolution for cloud nowcasts.

REFERENCES

Anandan, P., 1989: A computational framework and an algorithm for the measurement of visual motion. *Int J Comput Vision*, **2**, 283–310, https://doi.org/10.1007/BF00158167.

Apke, J. M., 2023: JasonApke/OCTANE. Accessed 13 January 2024, https://github.com/JasonApke/OCTANE.

——, and J. R. Mecikalski, 2021: On the Origin of Rotation Derived from Super Rapid Scan Satellite Imagery at the Cloud-Tops of Severe Deep Convection. *Mon. Wea. Rev.*, **149**, 1827–1851, https://doi.org/10.1175/MWR-D-20-0209.1.

——, J. R. Mecikalski, and C. P. Jewett, 2016: Analysis of Mesoscale Atmospheric Flows above Mature Deep Convection Using Super Rapid Scan Geostationary Satellite Data. *J. Appl. Meteorol. Climatol.*, **55**.

——, J. R. Mecikalski, K. Bedka, E. W. McCaul, C. R. Homeyer, and C. P. Jewett, 2018: Relationships between Deep Convection Updraft Characteristics and Satellite-Based Super Rapid Scan Mesoscale Atmospheric Motion Vector–Derived Flow. *Mon. Wea. Rev.*, **146**, 3461–3480, https://doi.org/10.1175/MWR-D-18-0119.1.

——, K. A. Hilburn, S. D. Miller, and D. A. Peterson, 2020: Towards objective identification and tracking of convective outflow boundaries in next-generation geostationary satellite imagery. *Atmos. Meas. Tech.*, **13**, 1593–1608, https://doi.org/10.5194/amt-13-1593-2020.

——, Y. J. Noh, and K. Bedka, 2022: Comparison of Optical Flow Derivation Techniques for Retrieving Tropospheric Winds from Satellite Image Sequences. *J. Atmos. Oceanic Technol.*, **39**, 2005–2021, https://doi.org/10.1175/JTECH-D-22-0057.1.

——, J. Tobin, Y. J. Noh, C. White, and J. M. Haynes, 2025: Using Optical Flow Temporal Interpolation of Satellite Imagery to Assist Multi-Sensor Global Cloud Product Composites. *J. Atmos. Oceanic. Technol. In Review*.

Austin, G. L., and A. Bellon, 1974: The use of digital weather radar records for short-term precipitation forecasting. *Quarterly Journal of the Royal Meteorological Society*, **100**, 658–664, https://doi.org/10.1002/qj.49710042612.

Baker, S., D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, 2011: A Database and Evaluation Methodology for Optical Flow. *Int. J. Comput. Vision*, **92**, 1–31, https://doi.org/10.1007/s11263-010-0390-2.

Bansal, A. S., Y. Lee, K. Hilburn, and I. Ebert-Uphoff, 2023: Leveraging spatiotemporal information in meteorological image sequences: From feature engineering to neural networks. *Environmental Data Science*, **2**, e31, https://doi.org/10.1017/eds.2023.26.

Bechini, R., and V. Chandrasekar, 2017: An Enhanced Optical Flow Technique for Radar Nowcasting of Precipitation and Winds. *J. Atmos. Oceanic Technol.*, **34**, 2637–2658, https://doi.org/10.1175/JTECH-D-17-0110.1.

Benjamin, S. G., and Coauthors, 2016: A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh. *Monthly Weather Review*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Berthomier, L., B. Pradel, and L. Perez, 2020: Cloud Cover Nowcasting with Deep Learning. *2020 Tenth Int. Conf. on Image Process., Theory, Tools, and Appl. (IPTA)*, 1–6, https://doi.org/10.1109/IPTA50016.2020.9286606.

Bessho, K., and Coauthors, 2016: An Introduction to Himawari-8/9— Japan's New-Generation Geostationary Meteorological Satellites. *Journal of the Meteorological Society of Japan. Ser. II*, **94**, 151–183, https://doi.org/10.2151/jmsj.2016-009.

Black, M. J., and P. Anandan, 1996: The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields. *Computer Vision and Image Understanding*, **63**, 75–104, https://doi.org/10.1006/cviu.1996.0006.

Bowler, N. E., C. E. Pierce, and A. W. Seed, 2006: STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Q. J. Roy. Meteor. Soc.*, **132**, 2127–2155, https://doi.org/10.1256/qj.04.100.

Brox, T., and J. Malik, 2011: Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **33**, 500–513, https://doi.org/10.1109/TPAMI.2010.143.

——, A. Bruhn, N. Papenberg, and J. Weickert, 2004: High Accuracy Optical Flow Estimation Based on a Theory for Warping. *Computer Vision - ECCV 2004*, T. Pajdla and J. Matas, Eds., *Lecture Notes in Computer Science*, Berlin, Heidelberg, Springer, 25–36, https://doi.org/10.1007/978-3-540-24673-2_3.

Burley, J. L., S. T. Fiorino, B. J. Elmore, and J. E. Schmidt, 2019: A Remote Sensing and Atmospheric Correction Method for Assessing Multispectral Radiative Transfer through Realistic Atmospheres and Clouds. *J. Atmos. Oceanic Technol.*, **36**, 203–216, https://doi.org/10.1175/JTECH-D-18-0078.1.

Caddell, J., 2016: Corona over Cuba: The Missile Crisis and the Early Limitations of Satellite Imagery Intelligence. *Intelligence & National Security*, **31**, 416–438, https://doi.org/10.1080/02684527.2015.1005495.

Chase, R. J., D. R. Harrison, A. Burke, G. M. Lackmann, and A. McGovern, 2022: A Machine Learning Tutorial for Operational Meteorology. Part I: Traditional Machine Learning. *Weather and Forecasting*, **37**, 1509–1529, https://doi.org/10.1175/WAF-D-22-0070.1.

——, K. Haynes, L. V. Hoef, and I. Ebert-Uphoff, 2025: Score-based diffusion nowcasting of GOES imagery, https://doi.org/10.48550/arXiv.2505.10432.

Chen, H., and Coauthors, 2025: Skillful Nowcasting of Convective Clouds With a Cascade Diffusion Model, https://doi.org/10.48550/arXiv.2502.10957.

Corpetti, T., D. Heitz, G. Arroyo, E. Mémin, and A. Santa-Cruz, 2006: Fluid experimental flow estimation based on an optical-flow scheme. *Exp Fluids*, **40**, 80–97, https://doi.org/10.1007/s00348-005-0048-y.

Daniels, J., W. Bresky, S. Wanzong, A. Bailey, and C. Velden, 2020: Atmospheric Motion Vectors Derived via a New Nested Tracking Algorithm Developed for the GOES-R Advanced Baseline Imager (ABI). *2020 EUMETSAT Meteor. Satellite Conf.*

Descombes, G., D. Auligne, H.-C. Lin, D. Xu, S. Schwartz, and F. Vandenberghe, 2014: Multi-sensor Advection Diffusion nowCast (MADCast) for cloud analysis and short-term prediction, https://doi.org/10.5065/D62V2D37.

Dolinar, E. K., and J. E. Nachamkin, 2024: Perspectives on Cloud Prediction, Postprocessing, and Verification for DoD Applications. *Bulletin of the American Meteorological Society*, **105**, E1034–E1041, https://doi.org/10.1175/BAMS-D-24-0077.1.

Dong, J., K. Ota, and M. Dong, 2023: Video Frame Interpolation: A Comprehensive Survey. *ACM Trans. Multimedia Comput. Commun. Appl.*, **19**, 78:1-78:31, https://doi.org/10.1145/3556544.

Everingham, M., L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, 2010: The Pascal Visual Object Classes (VOC) Challenge. *Int J Comput Vis*, **88**, 303–338, https://doi.org/10.1007/s11263-009-0275-4.

Farnebäck, G., 2003: Two-Frame Motion Estimation Based on Polynomial Expansion. *Image Analysis*, J. Bigun and T. Gustavsson, Eds., Vol. 2749 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 363–370, https://doi.org/10.1007/3-540-45103-X_50.

Fortun, D., P. Bouthemy, and C. Kervrann, 2015: Optical flow modeling and computation: A survey. *Comput. Vision Image Understanding*, **134**, 1–21, https://doi.org/10.1016/j.cviu.2015.02.008.

Germann, U., and I. Zawadzki, 2002: Scale-Dependence of the Predictability of Precipitation from Continental Radar Images. Part I: Description of the Methodology. *Mon. Wea. Rev.*, **130**, 2859–2873, https://doi.org/10.1175/1520-0493(2002)130<2859:SDOTPO>2.0.CO;2.

Géron, A., 2023: *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems*. Third edition. O'Reilly 834pp.

Gitro, C. M., and Coauthors, 2018: Using the Multisensor Advected Layered Precipitable Water Product in the Operational Forecast Environment. *J. Operational Meteor.*, **06**, 59–73, https://doi.org/10.15191/nwajom.2018.0606.

Goethals, S., D. Martens, and T. Evgeniou, 2022: The non-linear nature of the cost of comprehensibility. *Journal of Big Data*, **9**, 30, https://doi.org/10.1186/s40537-022-00579-2.

Goldberg, M. D., H. Kilcoyne, H. Cikanek, and A. Mehta, 2013: Joint Polar Satellite System: The United States next generation civilian polar-orbiting environmental satellite system. *J. Geophys. Res. Atmos.*, **118**, 13,463-13,475, https://doi.org/10.1002/2013JD020389.

Golemboski, J., 2001: Analysis of Cloud-Free Line-of-Sight Probability Calculations. M.S. thesis, Dept. of Engineering Physics, Air Force Institute of Technology, 63pp., https://scholar.afit.edu/etd/4616.

Grasso, L., and Coauthors, 2021: Satellite imagery and products of the 16–17 February 2020 Saharan Air Layer dust event over the eastern Atlantic: impacts of water vapor on dust detection and morphology. *Atmospheric Measurement Techniques*, **14**, 1615–1634, https://doi.org/10.5194/amt-14-1615-2021.

Guillot, E. M., T. H. Vonder Haar, J. M. Forsythe, and S. J. Fletcher, 2012: Evaluating Satellite-Based Cloud Persistence and Displacement Nowcasting Techniques over Complex Terrain. *Wea. Forecasting*, **27**, 502–514, https://doi.org/10.1175/WAF-D-11-00037.1.

Gultepe, I., and Coauthors, 2019: A Review of High Impact Weather for Aviation Meteorology. *Pure Appl. Geophys.*, **176**, 1869–1921, https://doi.org/10.1007/s00024-019-02168-6.

Ha, J.-H., and H. Lee, 2023: Enhancing Rainfall Nowcasting Using Generative Deep Learning Model with Multi-Temporal Optical Flow. *Remote Sensing*, **15**, 5169, https://doi.org/10.3390/rs15215169.

——, and ——, 2024: A Deep Learning Model for Precipitation Nowcasting Using Multiple Optical Flow Algorithms. *Wea. Forecasting*, **39**, 41–53, https://doi.org/10.1175/WAF-D-23-0104.1.

Hani, N., S. Engin, J.-J. Chao, and V. Isler, 2020: Continuous Object Representation Networks: Novel View Synthesis without Target View Supervision. *Adv. Neural Inf. Process. Syst.*, Vol. 33 of, Curran Associates, Inc., 6086–6099.

Haynes, J. M., Y. J. Noh, S. D. Miller, K. D. Haynes, I. Ebert-Uphoff, and A. Heidinger, 2022: Low Cloud Detection in Multilayer Scenes Using Satellite Imagery with Machine Learning Methods. *J. Atmos. Oceanic Technol.*, **39**, 319–334, https://doi.org/10.1175/JTECH-D-21-0084.1.

Heidinger, A. K., and W. Straka, 2020: NOAA Enterprise Cloud Mask Algorithm Theoretical Basis Document.

——, A. T. Evan, M. J. Foster, and A. Walther, 2012: A Naive Bayesian Cloud-Detection Scheme Derived from CALIPSO and Applied within PATMOS-x. *J. Appl. Meteor. Climatol.*, **51**, 1129–1144, https://doi.org/10.1175/JAMC-D-11-02.1.

——, Y. Li, and S. Wanzong, 2020: Enterprise AWG Cloud Height Algorithm (ACHA) Algorithm Theoretical Basis Document.

Herbst, E., S. Seitz, and S. Baker, 2009: Occlusion Reasoning for Temporal Interpolation using Optical Flow.

Herman, G. R., and R. S. Schumacher, 2016: Using Reforecasts to Improve Forecasting of Fog and Visibility for Aviation. *Weather and Forecasting*, **31**, 467–482, https://doi.org/10.1175/WAF-D-15-0108.1.

Hong, S., S. Kim, M. Joh, and S. Song, 2017: PSIque: Next Sequence Prediction of Satellite Images using a Convolutional Sequence-to-Sequence Network, https://doi.org/10.48550/arXiv.1711.10644.

Horn, B. K. P., and B. G. Schunck, 1981: Determining optical flow. *Artif. Intell.*, **17**, 185–203, https://doi.org/10.1016/0004-3702(81)90024-2.

Huang, Z., X. Shi, C. Zhang, Q. Wang, K. C. Cheung, H. Qin, J. Dai, and H. Li, 2022: FlowFormer: A Transformer Architecture for Optical Flow. *Comput. Vision – ECCV 2022*, S. Avidan, G. Brostow, M. Cissé, G.M. Farinella, and T. Hassner, Eds., Vol. 13677 of *Lecture Notes in Computer Science*, Cham, Springer Nature Switzerland, 668–685, https://doi.org/10.1007/978-3-031-19790-1_40.

Ionescu, V.-S., G. Czibula, and E. Mihuleţ, 2021: DeePS at: A deep learning model for prediction of satellite images for nowcasting purposes. *Procedia Computer Science*, **192**, 622–631, https://doi.org/10.1016/j.procs.2021.08.064.

Jann, A., 2017: *User Manual for the Extrapolated Imagery Processor of the NWC/GEO: Science Part*. EUMETSAT NWCSAF, https://www.nwcsaf.org/AemetWebContents/ScientificDocumentation/Documentation/GEO/v2016/NWC-CDOP2-GEO-ZAMG-SCI-UM-EXIM_v1.0.pdf.

Jiang, H., D. Sun, V. Jampani, M.-H. Yang, E. Learned-Miller, and J. Kautz, 2018: Super SloMo: High Quality Estimation of Multiple Intermediate Frames for Video Interpolation. *2018 IEEE/CVF Conf. on Comput. Vision and Pattern Recognition*, 9000–9008, https://doi.org/10.1109/CVPR.2018.00938.

Jiménez, P. A., J. Dudhia, G. Thompson, J. A. Lee, and T. Brummet, 2022: Improving the cloud initialization in WRF-Solar with enhanced short-range forecasting functionality: The MAD-WRF model. *Solar Energy*, **239**, 221–233, https://doi.org/10.1016/j.solener.2022.04.055.

Kashinath, K., and Coauthors, 2021: Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **379**, 20200093, https://doi.org/10.1098/rsta.2020.0093.

Kellerhals, S. A., F. De Leeuw, and C. Rodriguez Rivero, 2022: Cloud Nowcasting with Structure-Preserving Convolutional Gated Recurrent Units. *Atmosphere*, **13**, 1632, https://doi.org/10.3390/atmos13101632.

Kim, W., C.-H. Jeong, and S. Kim, 2024: Improvements in deep learning-based precipitation nowcasting using major atmospheric factors with radar rain rate. *Computers & Geosciences*, **184**, 105529, https://doi.org/10.1016/j.cageo.2024.105529.

Kochkov, D., and Coauthors, 2024: Neural general circulation models for weather and climate. *Nature*, **632**, 1060–1066, https://doi.org/10.1038/s41586-024-07744-y.

Kosmopoulos, P., and Coauthors, 2020: Short-Term Forecasting of Large-Scale Clouds Impact on Downwelling Surface Solar Irradiation. *Energies*, **13**, 6555, https://doi.org/10.3390/en13246555.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2012: ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, Vol. 25 of, Curran Associates, Inc.

Lam, R., and Coauthors, 2023: Learning skillful medium-range global weather forecasting. *Science*, **382**, 1416–1421, https://doi.org/10.1126/science.adi2336.

Leinonen, J., U. Hamann, I. V. Sideris, and U. Germann, 2023: Thunderstorm Nowcasting With Deep Learning: A Multi-Hazard Data Fusion Model. *Geophysical Research Letters*, **50**, e2022GL101626, https://doi.org/10.1029/2022GL101626.

Lorenz, E., A. Hammer, and D. Heinemann, 2004: Short term forecasting of solar radiation based on satellite data. *EUROSUN2004 (ISES Europe Sol. Congress)*, https://www.researchgate.net/publication/267971482_Short_term_forecasting_of_solar_radiation_based_on_satellite_data.

Meng, C., S. Griesemer, D. Cao, S. Seo, and Y. Liu, 2025: When physics meets machine learning: a survey of physics-informed machine learning. *Mach. Learn. Comput. Sci. Eng*, **1**, 20, https://doi.org/10.1007/s44379-025-00016-0.

Menzel, W. P., 2001: Cloud Tracking with Satellite Imagery: From the Pioneering Work of Ted Fujita to the Present. *Bull. Amer. Meteor. Soc.*, **82**, 33–48, https://doi.org/10.1175/1520-0477(2001)082<0033:CTWSIF>2.3.CO;2.

Miller, S. D., M. A. Rogers, J. M. Haynes, M. Sengupta, and A. K. Heidinger, 2018: Short-term solar irradiance forecasting via satellite/model coupling. *Sol. Energy*, **168**, 102–117, https://doi.org/10.1016/j.solener.2017.11.049.

Muszyński-Sulima, W., 2023: Cold War in Space: Reconnaissance Satellites and US-Soviet Security Competition. *European journal of American studies*, **18**, https://doi.org/10.4000/ejas.20427.

Nielsen, A. H., A. Iosifidis, and H. Karstoft, 2021: CloudCast: A Satellite-Based Dataset and Baseline for Forecasting Clouds. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **14**, 3485–3494, https://doi.org/10.1109/JSTARS.2021.3062936.

Noh, Y. J., and Coauthors, 2017: Cloud-Base Height Estimation from VIIRS. Part II: A Statistical Algorithm Based on A-Train Satellite Data. *J. Atmos. Oceanic Technol.*, **34**, 585–598, https://doi.org/10.1175/JTECH-D-16-0110.1.

——, and Coauthors, 2022a: A Framework for Satellite-Based 3D Cloud Data: An Overview of the VIIRS Cloud Base Height Retrieval and User Engagement for Aviation Applications. *Remote Sens.*, **14**, 5524, https://doi.org/10.3390/rs14215524.

——, S. Miller, C. J. Seaman, J. M. Haynes, Y. Li, A. K. Heidinger, and M. Kulie, 2022b: Algorithm Theoretical Basis Document (ATBD), Enterprise Algorithm Working Group (AWG) Cloud Base Algorithm (ACBA) Version 2.3.

——, and Coauthors, 2024: Building Unified Global 3D Cloud Data from Multiple Satellites for Advancing Weather and Climate Research. *2024 IEEE Int. Geosci. and Remote Sens. Symp. (IGARSS)*, 7-12 July 2024, Athens, Greece, 1280–1283, https://doi.org/10.1109/IGARSS53475.2024.10642296.

Nonnenmacher, L., and C. F. M. Coimbra, 2014: Streamline-based method for intra-day solar forecasting through remote sensing. *Sol. Energy*, **108**, 447–459, https://doi.org/10.1016/j.solener.2014.07.026.

Papachristopoulou, K., and Coauthors, 2024: Effects of clouds and aerosols on downwelling surface solar irradiance nowcasting and short-term forecasting. *Atmospheric Measurement Techniques*, **17**, 1851–1877, https://doi.org/10.5194/amt-17-1851-2024.

Partio, M., L. Hieta, and A. Kokkonen, 2024: CloudCast -- Total Cloud Cover Nowcasting with Machine Learning, https://doi.org/10.48550/arXiv.2410.21329.

Pence, S., 2021: Fighting as Intended: The Case for Austere Communications. *Joint Force Quarterly*, **102**, 4–13.

Petrou, Z. I., and Y. Tian, 2017: High-Resolution Sea Ice Motion Estimation with Optical Flow Using Satellite Spectroradiometer Data. *IEEE Trans. on Geosci. and Remote Sens.*, **55**, 1339–1350, https://doi.org/10.1109/TGRS.2016.2622714.

Prudden, R., S. Adams, D. Kangin, N. Robinson, S. Ravuri, S. Mohamed, and A. Arribas, 2020: A review of radar-based nowcasting of precipitation and applicable machine learning techniques, https://doi.org/10.48550/arXiv.2005.04988.

Pulkkinen, S., D. Nerini, A. A. Pérez Hortal, C. Velasco-Forero, A. Seed, U. Germann, and L. Foresti, 2019: Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1.0). *Geosci. Model Dev.*, **12**, 4185–4219, https://doi.org/10.5194/gmd-12-4185-2019.

Ravuri, S., and Coauthors, 2021: Skilful precipitation nowcasting using deep generative models of radar. *Nature*, **597**, 672–677, https://doi.org/10.1038/s41586-021-03854-z.

Rinehart, R. E., and E. T. Garvey, 1978: Three-dimensional storm motion detection by conventional weather radar. *Nature*, **273**, 287–289, https://doi.org/10.1038/273287a0.

Roberts, N. M., and H. W. Lean, 2008: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Ronneberger, O., P. Fischer, and T. Brox, 2015: U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W.M. Wells, and A.F. Frangi, Eds., *Lecture Notes in Computer Science*, Cham, Springer International Publishing, 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.

Rosner, B., R. J. Glynn, and M.-L. T. Lee, 2006: The Wilcoxon signed rank test for paired comparisons of clustered data. *Biometrics*, **62**, 185–192, https://doi.org/10.1111/j.1541-0420.2005.00389.x.

Schmit, T. J., P. Griffith, M. M. Gunshor, J. M. Daniels, S. J. Goodman, and W. J. Lebair, 2017: A Closer Look at the ABI on the GOES-R Series. *Bull. Amer. Meteor. Soc.*, **98**, 681–698, https://doi.org/10.1175/BAMS-D-15-00230.1.

Schütz, M., A. Schütz, J. Bendix, and B. Thies, 2024: Improving classification-based nowcasting of radiation fog with machine learning based on filtered and preprocessed temporal data. *Quarterly Journal of the Royal Meteorological Society*, **150**, 577–596, https://doi.org/10.1002/qj.4619.

Shi, X., Z. Chen, H. Wang, D.-Y. Yeung, W. Wong, and W. Woo, 2015: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, https://doi.org/10.48550/arXiv.1506.04214.

——, Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W. Wong, and W. Woo, 2017: Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model, https://doi.org/10.48550/arXiv.1706.03458.

Sim, S., J. Im, S. Jung, and D. Han, 2024: Improving Short-Term Prediction of Ocean Fog Using Numerical Weather Forecasts and Geostationary Satellite-Derived Ocean Fog Data Based on AutoML. *Remote Sensing*, **16**, 2348, https://doi.org/10.3390/rs16132348.

Stettner, D., C. Velden, R. Rabin, S. Wanzong, J. Daniels, and W. Bresky, 2019: Development of Enhanced Vortex-Scale Atmospheric Motion Vectors for Hurricane Applications. *Remote Sens.*, **11**, 1981, https://doi.org/10.3390/rs11171981.

Storch, S., J., and D. McDonald G., 2001: Improvements to Air Force Cloud Forecast Model, Advect Cloud. *11th Conf. on Satell. Meteor. and Oceanogr.*

Sun, D., S. Roth, and M. J. Black, 2014: A Quantitative Analysis of Current Practices in Optical Flow Estimation and the Principles Behind Them. *Int J Comput Vis*, **106**, 115–137, https://doi.org/10.1007/s11263-013-0644-x.

Sundaram, N., T. Brox, and K. Keutzer, 2010: Dense Point Trajectories by GPU-Accelerated Large Displacement Optical Flow. *Computer Vision – ECCV 2010*, K. Daniilidis, P. Maragos, and N. Paragios, Eds., *Lecture Notes in Computer Science*, Berlin, Heidelberg, Springer, 438–451, https://doi.org/10.1007/978-3-642-15549-9_32.

Teed, Z., and J. Deng, 2020: RAFT: Recurrent All-Pairs Field Transforms for Optical Flow. *Comput. Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., Springer International Publishing, 402–419, https://doi.org/10.1007/978-3-030-58536-5_24.

Terven, J., D.-M. Cordova-Esparza, J.-A. Romero-González, A. Ramírez-Pedraza, and E. A. Chávez-Urbiola, 2025: A comprehensive survey of loss functions and metrics in deep learning. *Artif Intell Rev*, **58**, 195, https://doi.org/10.1007/s10462-025-11198-7.

Tuttle, J. D., and G. B. Foote, 1990: Determination of the Boundary Layer Airflow from a Single Doppler Radar. *Journal of Atmospheric and Oceanic Technology*, **7**, 218–232, https://doi.org/10.1175/1520-0426(1990)007<0218:DOTBLA>2.0.CO;2.

Urbich, I., J. Bendix, and R. Müller, 2018: A Novel Approach for the Short-Term Forecast of the Effective Cloud Albedo. *Remote Sens.*, **10**, 955, https://doi.org/10.3390/rs10060955.

Vendrasco, E. P., L. A. T. Machado, B. Z. Ribeiro, E. D. Freitas, R. C. Ferreira, and R. G. Negri, 2020: Cloud-Resolving Model Applied to Nowcasting: An Evaluation of Radar Data Assimilation and Microphysics Parameterization. *Wea. Forecasting*, **35**, 2345–2365, https://doi.org/10.1175/WAF-D-20-0017.1.

Walther, A., and A. K. Heidinger, 2012: Implementation of the Daytime Cloud Optical and Microphysical Properties Algorithm (DCOMP) in PATMOS-x. *J. Appl. Meteor. Climatol.*, **51**, 1371–1390, https://doi.org/10.1175/JAMC-D-11-0108.1.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences*. 3rd ed. Academic Press 676pp.

Woo, W., and W. Wong, 2017: Operational Application of Optical Flow Techniques to Radar-Based Rainfall Nowcasting. *Atmosphere*, **8**, 48, https://doi.org/10.3390/atmos8030048.

Xiong, W., W. Luo, L. Ma, W. Liu, and J. Luo, 2018: Learning to Generate Time-Lapse Videos Using Multi-Stage Dynamic Generative Adversarial Networks, https://doi.org/10.48550/arXiv.1709.07592.

Yang, S., and H. Yuan, 2023: A Customized Multi-Scale Deep Learning Framework for Storm Nowcasting. *Geophysical Research Letters*, **50**, e2023GL103979, https://doi.org/10.1029/2023GL103979.

Yu, Z., Z. Tan, S. Ma, and W. Yan, 2023: Nowcast for cloud top height from Himawari-8 data based on deep learning algorithms. *Meteor. Appl.*, **30**, e2130, https://doi.org/10.1002/met.2130.

Zach, C., T. Pock, and H. Bischof, 2007: A Duality Based Approach for Realtime TV-L 1 Optical Flow. *Pattern Recognition*, F.A. Hamprecht, C. Schnörr, and B. Jähne, Eds., Vol. 4713 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 214–223, https://doi.org/10.1007/978-3-540-74936-3_22.

Zimmer, H., A. Bruhn, and J. Weickert, 2011: Optic Flow in Harmony. *Int J Comput Vis*, **93**, 368–388, https://doi.org/10.1007/s11263-011-0422-6.