UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE POSGRADO EN CIENCIAS DE LA TIERRA
CIENCIAS ATMOSFÉRICAS

EVALUATION OF THE pySTEPS NOWCAST MODELS USING DATA FROM THE QUERÉTARO WEATHER RADAR AND NASA'S GOES-16 SATELLITE OVER CENTRAL MEXICO

TESIS

QUE PARA OPTAR POR EL GRADO DE
DOCTORA EN CIENCIAS DE LA TIERRA

PRESENTA:
MTRA. DIANA ADRIANA ISLAS FLORES

TUTOR PRINCIPAL:
DR. ADOLFO MAGALDI HERMOSILLO, ENES JURIQUILLA, UNAM

COMITÉ TUTOR:
DR. ARTURO QUINTANAR ISAÍAS, ICAyCC, UNAM
DR. DANIEL BRAUSE MICHELSON, ECCC, CANADA

México, CDMX. Abril, 2024

Jurado:

Dr. Arturo Quintanar Isaías	*Presidente*
Dr. David Kenton Adams	*Vocal*
Dr. Adolfo Magaldi Hermosillo	*Secretario*
Dra. Rosario Romero Centeno	*Suplente*
Dr. Tomeu Rigo Rivas	*Suplente*

Acknowledgment

Thank you to my tutor Dr. Adolfo Magaldi for your guidance and support throughout the whole project. I could not have completed this project without it.

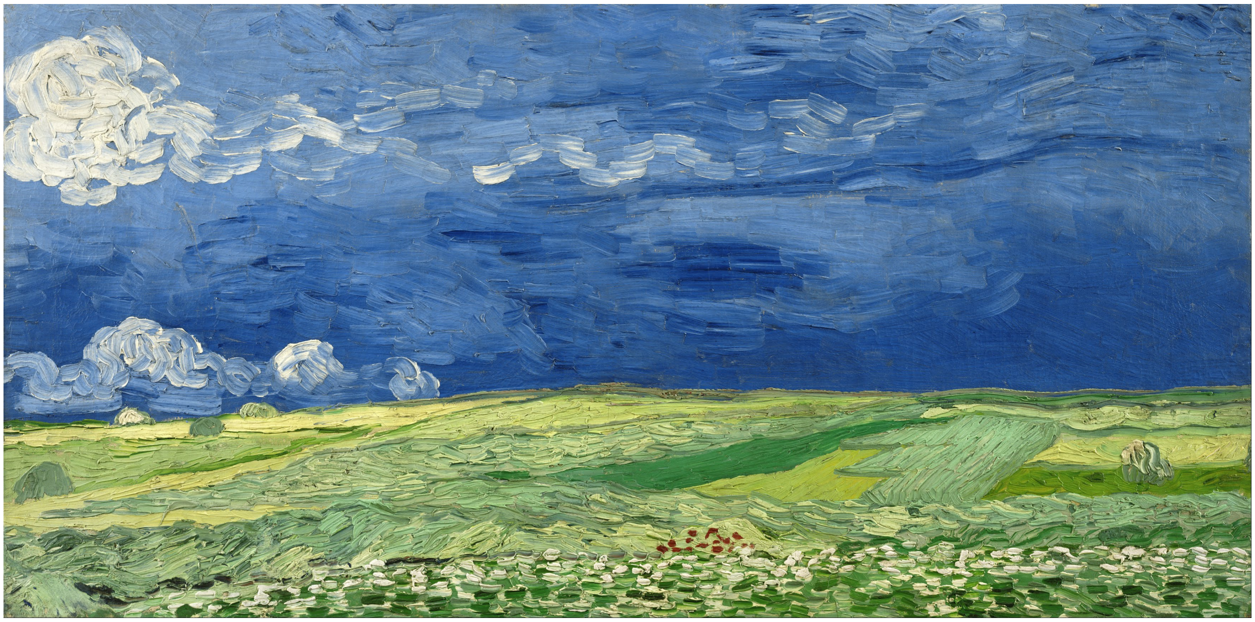
Thank you to Dr. Arturo Quintanar and Dr. Daniel Michelson for being part of the tutor committee and for the advise and suggestions you gave me.

Thank you to the jury for reading, correcting and helping improve this thesis.

I would like to express my gratitude to CONACHyT for the financial support and the administration and staff of the Posgrado en Ciencias de la Tierra for helping me deal with the university's bureaucracy.

Finally, thank you to my mom, dad, and my big sister Luisa for the support during this development of this work.

I dedicate this to my mom, my dad, my sister Luisa, little Atticus Mateo, Sony, and Sargento. I love you all.



Wheatfield Under Thunderclouds by Vincent van Gogh

Declaro conocer el Código de Ética de la Universidad Nacional Autónoma de México, plasmado en la Legislación Universitaria. Con base en las definiciones de integridad y honestidad ahí especificadas, aseguro mediante mi firma al calce que el presente trabajo es original y enteramente de mi autoría. Todas las citas de, o referencias a la obra de otros autores aparecen debida y adecuadamente señaladas, así como acreditadas mediante los recursos editoriales convencionales.

Abstract

In this work, an evaluation of the use of nowcast models with satellite and radar data was conducted for the region of central Mexico. The main objective was to characterize the precipitation that occurs in the region using the Queretaro weather radar data and the GOES-16 satellite data, and to use this analysis to evaluate the nowcast models provided by the *pySTEPS* python library. Before using the *Geostationary Operational Environment Satellite-16 (GOES-16)* satellite data, a preliminary quality control evaluation of the satellites cloud mask was performed using ceilometer and rain gauge data from the UNAM's RUOA network. The results showed the cloud mask had a probability of detection (PoD) of up to 95% and a minimum probability of false detection (PoFD) of 20%. Moreover, the mask had a 99% accuracy in detecting clouds when rainfall was present. Afterwards, the characterization of the rainfall that occurs in the region was carried out with the following satellite products: the ABI Cloud-top height (ACH), the ABI Cloud-top phase (ACTP) and the spectral channels 13 and 15 used to calculate the Split-Window Difference (SWD). The results showed that the SWD served as a good indicator of rainfall, with the average SWD decreasing from 2.5 K to 1.99 K in the 60 minutes prior to rainfall detection. The cloud-top height and the cloud-top phase indicated possible growth in clouds, however, it was difficult to distinguish deep convective clouds and high altitude cirrus clouds. From this analysis, a SWD threshold value of 2.5 K as a marker for possible convective development was established. Finally, for the nowcast assessment, three nowcast models were used from pySTEPS: Extrapolation, S-PROG and STEPS. The former two models proved the most accurate and yielded similar average results; STEPS had significantly less accurate results. The models performed accurately with the SWD data as input, having an average PoD reaching above 70% and a false alarm rate (FAR) of approximately 40% or less for the fifteen minute prediction. The nowcasts were less accurate with radar rain rate, with the PoD being below 70% and FAR reaching 40%, and the predicted intensity resulted in an average mean absolute error (MAE) above the rain rate average. However, the forecasts improved as precipitation events developed, and were best with lower lead times. Overall, it was determined that the SWD can be used as an indicator of future rainfall, and the nowcast models provided by pySTEPS can be used with GOES-16 satellite and radar data available without issues for the region of central Mexico, providing useful 15-minute forecast. Finally, the methodologies developed in this study have the potential to be operationally implemented in the future.

En este trabajo se realizó una evaluación del uso de modelos nowcast con datos satelitales y de radar para la región del centro de México. El objetivo principal fue caracterizar la precipitación que ocurre en la región utilizando los datos del radar meteorológico de Querétaro y los datos del satélite GOES-16, y utilizar este análisis para evaluar los modelos nowcast proporcionados por la biblioteca de Python *pySTEPS*. Antes de utilizar los datos del satélite *GOES-16*, se realizó una evaluación preliminar de control de calidad de la máscara de nubes del satélite utilizando datos de ceilómetros y pluviómetros de la red RUOA de la UNAM. Los resultados mostraron que la máscara de nube tenía una probabilidad de detección (PoD) de hasta el 95% y una probabilidad de detección falsa (PoFD) mínima de 20%. Además, la máscara tenía una precisión del 99% a la hora de detectar nubes cuando llovía. Posteriormente se realizó la caracterización de las precipitaciones que se presentan en la región con los siguientes productos satelitales: la altura de tope de nube, la fase de tope de nube y los canales espectrales 13 y 15 utilizados para calcular el Split-Window Difference (SWD). Los resultados mostraron que la SWD sirvió como un buen indicador de la precipitación, con una disminución promedio de la SWD de 2.5 K a 1.99 K en los 60 minutos previos a la detección de lluvia. La altura del tope de las nubes y la fase del tope de las nubes indicaron un posible crecimiento de las nubes; sin embargo, fue difícil distinguir las nubes convectivas profundas y los cirros de gran altitud. A partir de este análisis, se estableció un valor umbral de SWD de 2.5 K como marcador de un posible desarrollo convectivo. Finalmente, para la evaluación del nowcast, se utilizaron tres modelos de nowcast de pySTEPS: Extrapolación, S-PROG y STEPS. Los dos primeros modelos demostraron ser los más precisos y arrojaron resultados promedio similares; STEPS tuvo resultados significativamente menos precisos. Los modelos funcionaron con precisión con los datos SWD como entrada, con un PoD promedio superior al 70% y una tasa de alarmas falsas (FAR) de aproximadamente el 40% o menos para la predicción de quince minutos. Los pronósticos de nowcast fueron menos precisos con la intensidad de lluvia del radar, con un PoD por debajo del 70% y FAR alcanzando el 40%, y la intensidad prevista resultó en un error absoluto medio (MAE) promedio por encima del promedio de la intensidad de lluvia. Sin embargo, los pronósticos mejoraron a medida que se desarrollaron los eventos de precipitación y fueron mejores con pronósticos más cortos. En general, se determinó que el SWD se puede usar como indicador de lluvias futuras, y los modelos de nowcast proporcionados por pySTEPS se pueden usar con datos de radar y satélite GOES-16 disponibles sin problemas para la región del centro de México, lo que proporciona predicciones de 15-minutos útiles. Finalmente, las metodologías desarrolladas en este estudio tienen el potencial de implementarse operativamente en el futuro.

Contents

List of Figures

List of Tables

Acronyms

ABI *Advance Baseline Imager*

ACC accuracy or fraction of correct results

ACH ABI Cloud-top height

ACM ABI cloud mask

ACTP ABI Cloud-top phase

ALCF Automatic LIDAR and Ceilometer Framework

DATing Thunderstorm Detection and Tracking

FA false alarm ratio equivalent to PoFD

FAR false alarm rate

GOES-16 Geostationary Operational Environment Satellite-16

GOES-R Geostationary Operational Environment Satellites R series

HSS Heidke Skill Score

KSS Kuipers Skill Score

MAE mean absolute error

ME mean error

PoD probability of detection

PoFD probability of false detection

RMSE root mean standard error

RUOA UNAM's Atmospheric Observational Network

S-PROG Spatial-Prognosis

STEPS Short-Term Ensemble Prediction

SWD Split-Window Difference

Chapter 1

Introduction

The ability to locally forecast convective storms is very important, especially in flood prone regions and cities, such as the central region of Mexico. These short-term forecasts, which last 6 hours or less, are called nowcasts and there are several methodologies for developing these forecast using remote sensing methods (Sun et al., 2014). These include: data extrapolation, statistical analysis, and the use of numerical weather prediction models, among others (Browning and Collier, 1989; Sun et al., 2014). During the twentieth century, the most popular methodology for nowcasting convective storms and other phenomena with a short lifetime was extrapolation from radar and satellite images (Browning and Collier, 1989). Although these methods had their flaws and limitations, they have improved with the development of data analysis techniques, including the advancement of the sensors on board satellites that have better temporal and spatial resolution and the widespread use of the dual-polarization radars.

One region in particular that is ideal for the use and implementation of nowcast models is central Mexico. This region experiences a high frequency of convective storms during the rainy season that runs from late May to early October (Jauregui and Romales, 1996). It also contains several urban areas, such as the city of Santiago de Queretaro and Mexico City, that can benefit from short-term rain predictions to mitigate flooding damage. The region is covered by three radars (the Queretaro radar, the Cerro Catedral radar, and the Mexico City radar) and by NASA and NOAA's Geostationary Operational Environment Satellites R series (GOES-R). However, nowcast have not been thoroughly explored or used in central Mexico. Thus, this study explored the use of radar and satellite data to assess the viability of using nowcast models in the region of central Mexico.

The investigation of this work is divided in three main sections, with each section building upon the results of the previous sections. The first part involved evaluating the accuracy of the GOES-R cloud mask over Mexico using ceilometers. The second part of the study was the characterization of the precipitation in central Mexico with GOES-R and radar data. And lastly, the third and main part of the study was the evaluation of the nowcast models provided by pySTEPS over central Mexico using satellite and radar data. This thesis is divided in the following way: Chapter 2 includes background on the past research, methodologies, data and tools used in this study. Chapters 3, 4 and 5 present the three main investigations and results of this study. Finally, Chapter 6 is a summary and the conclusions of the work.

1.1 Motivations: Why is this work important?

Although nowcasts have been thoroughly studies and are currently in operational used in various parts of the world, they are a tool that have not been explored, either in a research or operational capacity, for the region of Mexico. The short term forecast that a nowcast can provide as the potential benefit of an early alarm for strong precipitation for the country's large metropolitan areas and other regions where flash flooding can occur, such as the Metropolitan Area of the Valley of Mexico where Mexico City is. The motivation of this study is to start the investigation into a possible operational implementation of nowcast in the region, by evaluating the performance the pySTEPS library with meteorological radar data available in the region of central Mexico. Additionally, a big obstacle that Mexico has is its lack of a full weather radar coverage, with only 70% of the country covered by a weather radar. Therefore, a further goal was to see if other types of meteorological data sets, specifically the GOES-16 data, can be used in the nowcast models as well, since the GOES-R satellites have good coverage of Mexico.

1.2 Objective

The first objective of this study is to characterize the rainfall that occurs in the central region of Mexico using weather radar data and weather satellite data. The second objective is to evaluate the use of nowcast models using the radar and satellite data available in central Mexico.

1.3 Hypothesis

The hypothesis of this study is that the data available in the region of Central Mexico, which include polarimetric Doppler radar data and GOES-16 satellite data, is reliable and useful in characterizing the precipitation that occurs in the region. Furthermore, these data sets are adequate for the use in nowcast models in the region.

Chapter 2

Background Research

This chapter presents a summary of the background research of the data, methods and tools used in this study.

2.1 Geostationary Operational Environmental-R Satellite Series

One of the main sources of data for this study was NASA and NOAA's Geostationary Operational Environmental R series or GOES-R, that were first launched November 2016. The GOES-R series is set to be made up of four satellites, all equipped with a variety of instruments pointed towards the earth and the sun. The four satellites composing the series are: the GOES-16 or GOES-EAST (launched November 2016), GOES-17 or GOES-WEST (launched March 2018), GOES-18 (launched March 2022, not yet fully operational) and GOES-T set to launch in 2024 (Benz et al., 2019). The instruments onboard each satellite are:

- → **Extreme Ultraviolet and X-ray Irradiance Sensors** is used to understand and monitor the solar irradiance in the upper atmosphere.
- → **Solar Ultraviolet Imager** is a telescope that can observe the sun in the extreme ultraviolet wavelength range.
- → **Space Environment In-Situ Suite** are sensors that monitor charge particle fluxes in the magnetosphere.

→ **Geostationary Lightning Mapper** is a near-infrared optical transient detector that can detect the presence of lightning.

→ ***Advance Baseline Imager*** **(ABI)** is imaging radiometer and is the primary instrument of the GOES-R series.

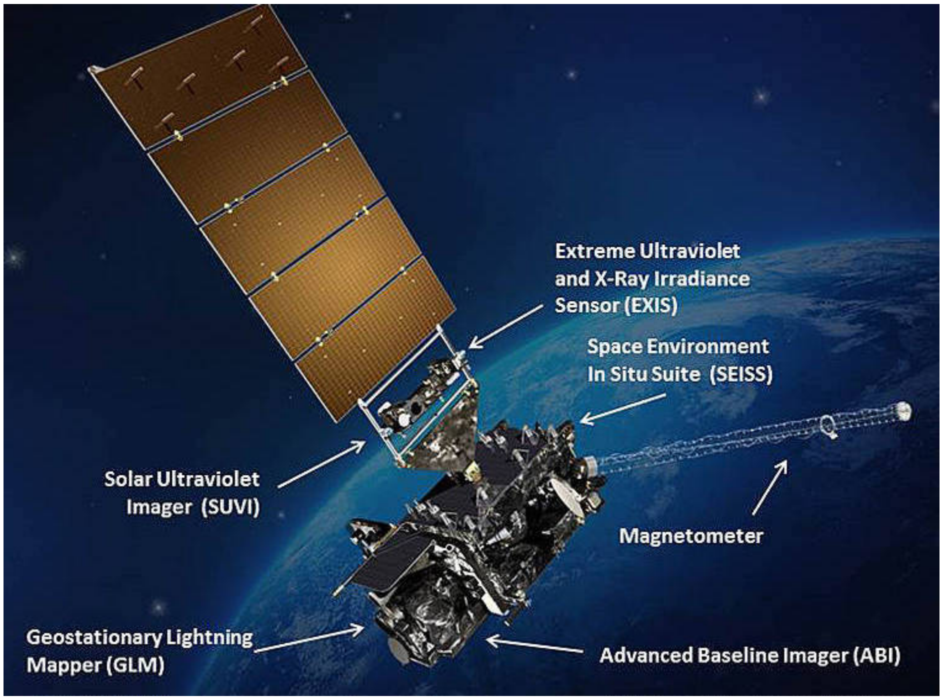


Figure 2.1: Diagram of a satellite from the GOES-R series and each of its main instruments (source: Benz et al., 2019).

2.1.1 Advance Baseline Imager

The main imaging instrument of the GOES-R series is the Advance Baseline Imager (ABI) instrument. This is an imaging radiometer that can take measurements in 16 different spectral bands: two in the visible range of the electromagnetic spectrum, four in the near-infrared and ten in the infrared. This wide range of spectral bands provides scientist with measurements needed to observer several phenomena occurring on Earth's atmosphere, including severe weather events and forest fires (Schmit et al., 2017; Benz et al., 2019). Furthermore, researchers can also use the band measurements to obtain several other derived products.

The ABI can provide measurements in three different coverage areas, in both North and South America: Mesoscale, Continental/Pacific US and Full Disk. Mesoscale has the smallest coverage, a 1000 km by 1000 km section, that is mainly used for monitoring regional conditions, and can be positioned anywhere within the Full disk coverage area. Mesoscale has a spatial resolution of 0.5 km by 0.5 km[1] and gives an image every 60 seconds. Continental (GOES-16)/Pacific (GOES-17) US coverage area is a 3000 km (N-S) by 5000 km (E-W) rectangle that covers part of Canada, the entire continental United States, and a large part of México. PACUS also covers the Hawaii Islands. These two coverage modes have a resolution of 2 km by 2 km and temporal resolution of 5 minutes. Finally, Full disk covers nearly the entire American continent (North and South America), has a spatial resolution of 2 km by 2 km, and provides an image every 10 minutes (Schmit et al., 2017; Benz et al., 2019).

The data provided by ABI is divided into three levels, the first is the raw data to be processed. The second is the reflectance or reflectivity measurements from any of the 16 ABI bands, which are called Level 1 products. These products have the spatial and temporal resolution of the coverage areas mentioned previously. The third type of products are the Level 2 products. These are obtained applying algorithms to the Level 1 data and other auxiliary data sets to obtain various physical quantities such as a cloud mask, stability indexes and cloud top temperature/phase. The spatial resolution of the Level 2 products depend on the coverage area available and the algorithm used to obtain the product (Schmit et al., 2017; Benz et al., 2019).

In this study, the Level 1 and Level 2 data from the ABI instrument of the GOES-16 satellite (Continental US) was used.

2.1.2 Split-Window difference

Although the GOES-R satellites provide several Level 2 products that can be helpful to obtain a description of the atmospheric conditions at a given time, one useful variable not directly available is the *Split-Window Difference* or SWD. The SWD has been useful in the measurement of water vapor and other variables using satellite data for many decades, and new generation of GOES has the capability of obtaining the SWD as well as obtaining better measurements of water vapor in the atmosphere (Lindsey et al., 2014, 2018; Grasso et al., 2020).

The SWD is the difference in the reflectance or brightness temperature measured at the 10.33 μm wavelength (called the “clean-window”) and at the 12.3 μm wavelength (called the “dirty-window”). The difference in these two spectral channels is useful in describing the

---

[1] the resolution can vary depending on the band, some have a minimum resolution of 1 km by 1 km.

amount low-level water vapor in the atmosphere because the water vapor absorbs a large amount of energy in the "dirty-window" and very little energy in the "clean-window", in other words the difference in energy measured between the two is related to the amount of water vapor in the atmosphere. The SWD is largest when there is little low-level water vapor in the atmosphere and the difference in the reflectance between the dirty window and clean window is highest. The SWD falls in value with increasing water vapor, when the absorption of energy by water vapor in the dirty window decreases the difference in the reflectance between the dirty and clean windows (Chesters and Uccellini, 1983).

Although there are other variables to consider that can affect the accuracy of the low-level water vapor measurements obtained through the SWD (e.g. Chesters and Uccellini, 1983; Schroedter-Homscheidt et al., 2008), this study uses the simple form of the variable (equation 2.1). This choice was made because this region lacks other instruments available, such as ground-base weather stations, that cover the whole area of study and can provide the necessary variables to correct the SWD measurements (such as surface temperature). For this reason, one of the secondary objectives of this study is to only use satellite and radar data in order to make any future implementations more accessible.

$$SWD = T_{B,Dirty} - T_{B,Clean} \tag{2.1}$$

2.2 Radar and Quantitative Precipitation Estimate

One of the many variables that can be obtained using weather radar data is the precipitation rate and accumulation. These values are important pieces of information when studying the climate and weather patterns of a region, specially if it is prone to severe weather and flooding. However, there are several obstacles that have to be resolved in order for the precipitation estimations to be as accurate as possible. This work will focus on describing the procedures taken and the problems encountered with the available solutions for obtaining the best rainfall rates and accumulations using weather radar data.

2.2.1 Reflectivity and rainfall rate relation

A weather radar's main variable of information is the measured reflectivity, or Z. The reflectivity is the amount of transmitted power returned by the target. Its relationship with the rainfall rate (R) has been studied extensively for the past several decades. There are several derivations of the relationship between R and Z, and one of the earliest and simplest form of this relation is given in equation 2.2 (Marshall et al., 1947; Hannesen, 2001).

$$Z = aR^b \tag{2.2}$$

With a and b being constants that depend on several factors such as drop size distribution (Marshall and Palmer, 1948).

There are several issues and obstacles needed to solve in order to obtain an accurate estimation of rainfall rates, mainly relating to the accuracy and errors associated with the Z measurements. Although there are several derivations of the $Z-R$ relationship, this relation depends on the drop-size distribution as well as the drop speed because of the reflectivity's sensitivity to these parameters (Magaldi, 2013) for all cases. As a result, this dependency can create errors in the estimated rainfall rate if the wrong precipitation type is considered. There are other issues that can caused error in the rainfall rate estimates obtained though the $Z-R$ relation which need to be considered beforehand, such as selecting the Z value which will give a more accurate precipitation estimate (i.e. choosing between mean reflectivity factor and the Z-distribution of the pulse volume if it is possible to obtain when inhomogeneous beam filling occurs), or choosing the correct interpolation methodology when switching from a polar to a Cartesian grid (Hannesen, 2001).

In addition to the errors from the $Z-R$ relationship, there are other factors that could affect the estimated rainfall rates obtained from the reflectivity. These include radar calibration, propagation effects, clutter and other non-meteorological echoes, and attenuation of the radar beam (Magaldi, 2013). Mitigation or elimination of the latter two issues are discussed in detail in the next sections. Regarding radar calibration issues, there are several methodologies that have improved radar calibration procedures to limit the possible errors that calibration issues can cause (Meischner, 2005). With respect to the propagation effects, these include anomalous propagation caused by varying vertical temperature and humidity atmospheric profiles, and can cause errors by creating clutter. This is because any variation of the vertical profiles of the atmosphere from their standard values causes changes in the refraction index of the atmosphere, which leads to the bending of the radar beam towards the ground, creating the clutter echoes (Magaldi, 2013). Studies have explored this issue and found several algorithms which remove echoes affected by anomalous propagation (e.g. Fornasiero et al., 2005; Dixon, 2007; Magaldi, 2013) and further exploration of removal of clutter echoes is discussed in the next section.

2.2.2 Radar data clutter removal

Because the rainfall estimates derived from radar data depend on the reflectivity given by the data, it is important to process and correctly filter this data in order to obtain the most accurate precipitation rates. An important part of this process is the removal of all echoes that are not associated with meteorological phenomenas. The non-meteorological echoes that are caused by the interaction of the radar beam with fixed targets are called clutter and are caused by two main factors: the blocking of the beam by the fixed targets or by anomalous propagation of the radar beam as discussed previously (Magaldi, 2013). Clutter in radar echoes can be classified into three categories: land clutter, sea clutter and airborne clutter. Land clutter are echoes from topographical and geological features, as well as buildings and other man made structures. Sea clutter is mostly generated by ships and the sea surface, and airborne clutter results from airplanes and living targets such as insects or birds (Magaldi, 2013).

An important aspect of the clearing of clutter from meteorological radar echoes is making sure that no hydrometeor and precipitation signals and echoes are not removed in the process. In general, there are six overall categories of clutter removal algorithms as described by Joss and Lee (1995), Meischner et al. (1997), and Magaldi (2013):

1. **Mapping:** This methodology involves the creation of clutter maps obtained with clear day echoes and subsequently used to eliminate the clutter in echoes to be evaluated.
2. **Doppler:** The analysis of Doppler velocities to remove any possible clutter with zero radial velocity.
3. **Statistical:** The removal of clutter signals that give non-coherent signal statistics.
4. **Interpolation:** The use of surrounding information to fill any gaps left by the removal of the clutter.
5. **Polarization:** The use of polarimetric radars to detect polarization signatures to filter clutter from any meteorological data.
6. **Multi-source:** The removal of clutter by using information from other sources.

While all these methodologies have their drawbacks and limitations, it has been observed that the best results are obtained with a combination of these algorithms (Meischner et al., 1997).

One particular technique for removing clutter that is often used is the one developed by Gabella and Notarpietro (2005). They created an algorithm which takes advantage of the fact that non-stationary echoes and anomalous propagation echoes have large spatial variability (larger than weather echoes). They developed a filter that analyzes the spatial correlation of the pixels to determine whether or not they are clutter by using threshold minimum values, as well as a 'compactness test'. Their results showed good clutter removal and improved meteorological echoes when this algorithm was applied after the application of another clutter removal technique. Thus, they concluded that this particular method should be used after the application of other techniques to obtain better and more optimal results in clutter removal.

2.2.3 Radar data attenuation correction

Another obstacle that is confronted when trying obtaining precipitation estimates from radars is attenuation. Radar beam attenuation is the loss of energy of the radar beam due to the absorption or scattering of the electromagnetic waves of the beam as it interacts with the atmosphere between the radar and the target. The attenuation of the radar beam can cause problems in the estimation and analysis of precipitation values derived from radar data.

One of the main causes of attenuation is heavy rainfall and hail. This type of attenuation can be negligible for S-band radars, but can produce severe loss in beam intensity for short wave radars (C-band and X-band). There have been several methods developed to try to mitigate the errors that can arise from the attenuation of the radar beam. The first to attempt to correct the attenuation in radar precipitation calculations was by Hitschfeld and Bordan (1954), where they calculated imperially the attenuation and the Path-Integrated Attenuation using the radar reflectivity in order to improve rainfall rates. However, they found that this particular approach created significant errors in the calculation of the rain rates if the calibration errors were not kept at an absolute minimum. Another way of addressing errors due to attenuation was the use of other sources of information in conjunction with radar reflectivity. One example of this is the use of the "surface reference technique" by Delrieu et al. (1997). Their study tested the use of the mountain return to correct the effects of attenuation in the region of the city of Marseille, France. They compared the mountain return reflectivity during dry and wet conditions to estimate the attenuation due to rain. Their results showed promise in mitigating and correcting attenuation errors. Other studies have used microwave links to recover the reflectivity lost during strong rainfall events and improve rainfall estimates (Rahimi et al., 2006; Cummings et al., 2009; Bianchi et al., 2013;

Troemel et al., 2014). For example, the study done by Troemel et al. (2014), where they used commercial microwave radio links with a C-band polarimetric radar in southern Germany.

With the development of polarimetric radars, new methodologies for eliminating attenuation errors in rainfall estimates were developed (Testud et al., 2000; Vulpiani et al., 2008). Such as in Vulpiani et al. (2008), who used variables given by the polarimetric radars to improve rainfall rates. More specifically, they calculated the PIA with the differential phase using the relationship between the specific differential phase with the specific attenuation. Since these variables depend on temperature and raindrop size distributions, their algorithm also involved Bayasian classification scheme to determine the best rain regime and improve the rainfall rain results.

Recently, one technique that has been often used to eliminate attenuation errors due to rainfall was developed by Kraemer and Verworn (2008). Their method used Hitschfeld and Bordan (1954)'s empirical relationship between the reflectivity and attenuation ($A = cZ^d$ with A being the attenuation and Z the reflectivity) and created an algorithm that obtained the best parameters (d and c) which give stable correction results. Although Kraemer and Verworn (2008) used attenuation references measurements from microwave links to find the most optimal pair of parameters that varied with each scan, they established an algorithm using an iterative search procedure to correct the attenuation without the need of any reference measurements. This involved maintaining the parameter d constant and the parameter c varied until finding the right value that give stable results (instability was defined by Kraemer (2008) as any corrected reflectivity values surpassing 59 dB). This algorithm was tested by Jacobi and Heistermann (2016) alongside the technique developed by Harrison et al. (2000) and a modified version of Kraemer and Verworn (2008)'s methodology, where they introduced a PIA threshold maximum limit as another instability criteria. Their results showed that Kraemer and Verworn (2008)'s technique allowed for a reduction in the error of the rainfall estimations over a wide range of rainfall situations (moderate to high rainfall). And these results improved when adding the PIA threshold instability criteria.

2.2.4 Polarimetric radar variables

In recent decades, the increased use of dual-polarization Doppler radars have further improved the rain rate estimates obtained with radar data. This is because the double-axis polarization of these radars can give different measurements that improve certain calculations. For example, the differential reflectivity, Z_{DR}, which is the difference between the measured reflectivity in the horizontal polarization plane and the vertical polarization plane, can have different values for large, oblate drops than for smaller drops. This difference in Z_{DR} between raindrop size can help distinguish between different types of precipitation that have the same rainfall rate (Hannesen, 2001). Another useful variable is the differential propagation phase, K_{DP}, or the difference in propagation phase between the horizontal and vertical polarization planes, because it is immune to attenuation problems (Wang and Chandrasekar, 2010).

Several studies have investigated the use of polarimetric radar variables to improve the rainfall rates and accumulated values obtained. As previously discussed, Vulpiani et al. (2008) used polarimetric radar measurements to improve attenuation corrections. Moreover, in their study, Saltikoff and Nevvonen (2011) found that dual-polarization radars have a better capability to distinguish between different types of hydrometeors and precipitation, enabling them to detect the difference between precipitation and non-precipitation echoes. Several studies have also observed that variables given by dual-polarization radars can be used to improve rain rates. Wang and Chandrasekar (2010) used K_{DP} because of its lack of need of the attenuation correction and an absolute calibration of the radars in used, and the rainfall rate estimates obtained from K_{DP} are less sensitive to the drop size distribution that is considered in the calculations. However, although they did observe promising results using this variable, there were still some issues that caused inaccuracies in the results, such as sensitivities and biases in light rain. Other studies have also explored which variable is best for rainfall rate estimation. Matrosov et al. (2013) explored rainfall rates using both the horizontal reflectivity (${Z_H}^2$) and K_{DP} using a X-band dual-polarization radar over Boulder, Colorado during convective rainfall with hail. They found that K_{DP} had agreed with surface precipitation data and Z_H overestimated precipitation values mainly due to the fact that it cannot distinguish between rain and hail echoes (K_{DP} is immune to the presence of hail). Furthermore, Thurai et al. (2017) calculated the rainfall rate using K_{DP}, Z_H, Z_{DR}, and the specific attenuation (A_H) using an X-band dual-polarization radar. They also found that K_{DP} gave the best results when compared with a rain gauge and disdrometer data. Finally, Simpson and Fox (2018) used K_{DP}, Z_H, Z_{DR} and a combination of Z_H and Z_{DR} and of Z_H and K_{DP} to calculate rainfall rate with an X-band dual polarization radar over Missouri, USA. Their results show that the combination of Z_H and Z_{DR} gave the best results when compared to ground precipitation data.

Considering these studies, the development of dual-polarization radar networks have improved the rainfall rates. However, there are problems that need to be address and taken care

---

[2] reflectivity measured in the horizontal polarization plane

of in order to obtain the best and most accurate results (Thurai et al., 2017). These problems can include biases due to radar characteristics (e.g. beam intensity, pulse rate), attenuation of radar signal and sensitivities to variables such as drop-size distribution and precipitation type (e.g. light rain, hail) (Saltikoff and Nevvonen, 2011; Wang and Chandrasekar, 2010; Thurai et al., 2017).

2.3 Nowcast models with satellite and radar data

Nowcasts have been in used in many regions for several decades. Before discussing the work done in this field, it is important to understand the basic workings of a nowcast system. In Figure 2.2, a basic flow diagram illustrates how a nowcast works. First, the initial satellite and radar data is assessed and the required information is extracted from each set. This can be satellite data that analyses convective development and Quantitative Precipitation Estimates from radar. Next, the two data sets are merged into a common grid (spatial and temporal) in order to obtain consistent results. Lastly, the final algorithm or extrapolation method is applied to obtain the short term forecast. As mentioned before, this is a very basic description of how a nowcast is obtained. The complexity of the nowcast creation depends on the type of satellite/radar data used, how it is used or evaluated, as well as the extrapolation method or algorithm used to obtain the nowcast.

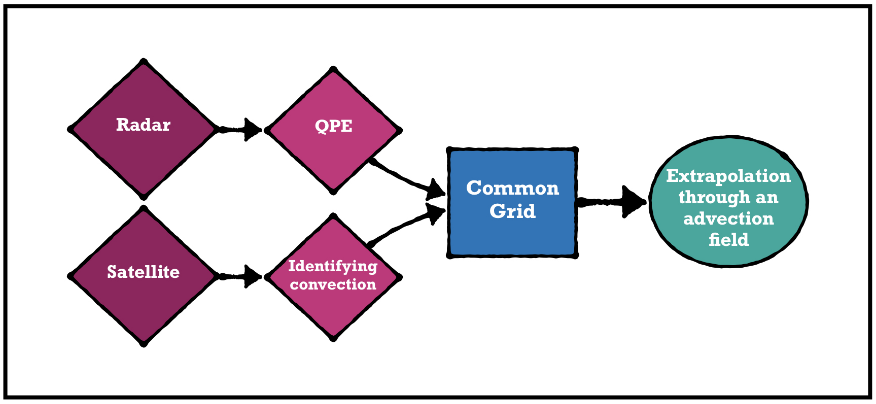


Figure 2.2: Simple flow diagram of how a basic nowcast system works.

For many decades, nowcasts have mainly been developed using analysis and extrapolation of radar images and data. Nowcasting models that mainly utilize radar data have been extensively studied and implemented—e.g., Seed (2003); Hering et al. (2004); Bowler et al. (2006); Ruzanski et al. (2011); Mandapaka et al. (2012); James et al. (2018); Foresti et al. (2019); Pulkkinen et al. (2020). One example in particular is the Python pySTEPS library that offers several nowcast models designed specifically for radar data (Pulkkinen et al., 2019; pySTEPS, 2021). These pySTEPS nowcast models have been used in many studies around the world to evaluate their forecasting skill and to compare them with other types of nowcast models, using radar rain rate and reflectivity data (Pulkkinen et al., 2020; Cuomo and Chandrasekar, 2021; Niu et al., 2021; Saadi et al., 2021; van der Kooij et al., 2021; Han et al., 2022; Cambier van Nooten et al., 2023; Ritvanen et al., 2023) and other types of data sources such as microwave links (Imhoff et al., 2020), and blended NWP data (Imhoff et al., 2023). However, the use of radar data techniques has had its limitations, especially because it is necessary for rain to occur in order to obtain any data for nowcasting development, which can shorten the time of forecast. Regardless, these types of nowcast have been improved by the development of better radar products, such as dual-polarization measurements, and incorporation of satellite data.

Continuing with the work done in nowcast development, recently, studies have explored convective initiation using satellite data to incorporate into nowcasts. Zinner et al. (2008) developed a tracking and nowcasting algorithm called Cumulonim Bus TRacking And Monitoring, that uses Meteosat-8 SEVIRI [3] data. This algorithm uses threshold values from the SEVIRI infrared channels to detect any cumulus clouds, assigning different levels depending on how developed the convection is when it was detected. Afterwards, it tracks and predicts (through extrapolation) where the systems will be heading. Siewert et al. (2010) analyzed the incorporation of satellite data along with radar echoes to predict convective initiation. In addition to the threshold values of 35 dBz in radar echoes, they used data from the Meteosat second generation (MSG-2) to explore convective initiation. They used a cloud masking technique to detect clouds and a 'box-averaging' method to follow cloud development with the satellite data. Their results indicated that the integration of the MSG analysis showed skill in detecting region of convective initiation and false alarms were reduced. Moreover, they were able to apply their analysis in a region with scarce radar data, although further research to validate or reinforce their conclusions was needed. Mecikalski et al. (2010a,b) also used MSG-2 data to explore cumulus cloud development through the exploration of infrared and visible fields. They used the satellite data to obtain sets of critical values/intervals of the

---

[3] Spinning Enhanced Visible and Infrared Imager

brightness temperature at various spectral bands to obtain and describe various cloud properties (such as cloud depth, cloud-top glaciation and updraft velocity). With these values and fields, they explored the possible implementation of a ‘threshold-based scoring’ algorithm (using principal component analysis with current atmospheric observations to find the best field that can describe future cloud growth) to possibly predict cumulus growth.

Continuing, Sieglaff et al. (2011) used GOES-12 satellite data to explore convective initiation prior to the radar precipitation echoes. They applied cloud-type and cloud-mask algorithms to asses cloud conditions over their area of study. With this data, they developed an algorithm that uses a ‘box-averaging’ approach to capture the period from the first signal of cloud vertical growth through time of cloud-top glaciation/storm maturation. With this, they were able to categorize satellite pixels according to their probability of convective initiation/development. They found that this algorithm works best in region where severe weather was expected (they observed higher probability of detection and lower false alarm rates in these regions). Additionally, they were able to extend the nowcast lead-time to approximately an hour, but causing a higher rate of false alarms. Other studies have used satellite-derived precipitation data as inputs in different types of nowcast models (Rivolta et al., 2006; Hill et al., 2020; Kumar et al., 2020; Ehsani et al., 2022). For example, Hill et al. (2020) used the convective rainfall rate and the rapidly developing thunderstorm convective warning products provided by the European Organization for the Exploitation of Meteorological Satellites Nowcasting Satellite Applications Facility in tropical Africa. The authors of the study found the satellite products to provide skillful nowcast of the precipitation of the area for lead times of up to 90 minutes. Finally, as previously mentioned, Lindsey et al. (2018) showed an example of observation of low-level water vapor maximums before the development of convection using the ‘split-window’ technique with GOES-R. As the authors pointed out in their conclusions, these observed maximums can be used as indicators of possible convection which can possibly be implemented in nowcast.

In conclusion, the incorporation of satellite data in the development of nowcast is important, especially in areas where radar data is scarce or non-existent, but count with satellite coverage. The studies described in the previous paragraphs show the potential improvements in nowcasting when incorporating satellite data along with radar data or using satellite data alone. Their analysis demonstrated several new methodologies that can possibly improve convective initiation prediction (Mecikalski et al., 2010a,b; Siewert et al., 2010; Sieglaff et al., 2011; Lindsey et al., 2018), lower false alarm rates (Siewert et al., 2010; Sieglaff et al., 2011), and increase the lead-time of nowcast (Sieglaff et al., 2011; Hill et al., 2020). However, their methodologies and algorithms need further testing in order to fully understand their potential and limitations and assess how they can be implemented in an operational manner and in real-time.

2.4 A forecast tool based on remote sensing data: pySTEPS

Through out the decades, several nowcasting models using various methodologies have been developed and tested worldwide, from extrapolation of data to the use of machine learning algorithms to obtain possible forecast with any given initial variables. One particular source of nowcast models is called *pySTEPS* (Pulkkinen et al., 2019; pySTEPS, 2021), which is a publicly available and community maintained python library and uses the extrapolation technique of Lagrangian persistence to develop precipitation nowcasts.

Th Lagrangian persistence model used by pySTEPS, based on the work of Zawadzki et al. (1994), extrapolates the radar echoes through time along a derived stationary motion field maintaining the intensity constant (Pulkkinen et al., 2019). Additionally, pySTEPS also includes more sophisticated nowcasts, with added elements of spatial decomposition and probabilistic factors to account for uncertainties and errors in the precipitation and motion fields (Pulkkinen et al., 2019).

The motion fields obtained with *pySTEPS* can be calculated using five different methods (pySTEPS, 2021):

1. **Constant** is a method that estimates the motion field by maximizing the correlation between two consecutive images (pySTEPS, 2021).
2. **Lucas-Kanade** is a method based on the work of Lucas and Kanade (1981) and Bouguet (2001).
3. **Variational Echo Tracking** is a global variation echo-tracking method developed by Laroche and Zawadzki (1995) and used in Germann and Zawadzki (2002).
4. **Proesman** is the anisotropic diffusion approach described in Proesmans et al. (1994).
5. **DARTS** is the a spectral approach method used in the DARTS nowcasting model (Ruzanski et al., 2011).

In relation to the nowcast models, pySTEPS has eight built-in options (pySTEPS, 2021), included among these are:

Extrapolation is a simple extrapolation of the precipitation field through the estimated motion fields calculated using the input data.

Spatial-Prognosis (S-PROG) is a method developed by Seed (2003). In summary, the authors describe this nowcast as advection-based spectral decomposition model that takes advantage of the scale-dependence of the properties of rain fields integrating an autoregressive model within each cascade field.

Short-Term Ensemble Prediction (STEPS) is the nowcast model described by Bowler et al. (2006). The STEPS nowcast uses the S-PROG method with an added element of uncertainty in the calculation of the precipitation fields in the form of stochastic perturbations in the precipitation and motion fields.

Autoregressive Nowcasting Vertically Integrated Liquid is the nowcast developed by Pulkkinen et al. (2020). This model uses the decomposition techniques of S-PROG with the vertically integrated liquid calculated from radar reflectivity fields instead of precipitation fields.

In addition to the nowcast models, *pySTEPS* also includes several functions that can help visualize and validate the nowcast results (pySTEPS, 2021).

As mentioned in Section 2.3, nowcast models from the pySTEPS library have been evaluated and compared to other models in many studies around the world (Pulkkinen et al., 2019; Cuomo and Chandrasekar, 2021; Niu et al., 2021; Saadi et al., 2021; van der Kooij et al., 2021; Han et al., 2022; Cambier van Nooten et al., 2023; Ritvanen et al., 2023; Imhoff et al., 2020, 2023). For example, in a study describing and assessing the *pySTEPS* library, Pulkkinen et al. (2019) concludes that the models are reliable but the reliability decreases for increasing rainfall intensity thresholds and lead time. Moreover, the method used to calculate the motion fields does not greatly affect the quality of the nowcast. Finally, the model can provide a skillful nowcast of up to 1 hour, although it depends on the intensity and spatial scales.

2.4.1 DATing Module

The pySTEPS Python library includes two tracking features. The first one utilizes the Lucas-Kanade or LK tracking function of the OpenCV module for tracking storm clusters in each image (Pulkkinen et al., 2019). The second one is the Thunderstorm Detection and Tracking (DATing) module that follows the Thunderstorms Radar Tracking algorithm (Hering et al., 2004; Pulkkinen et al., 2019) used by MeteoSwiss with modifications. The DATing module algorithm involves detecting clusters by analyzing the data matrix of an image. Then, using the pySTEPS Lucas-Kanade advection function, it generates the motion fields with that image and the previous image, and conducts an extrapolation through advection to obtain a prediction of the next image. Afterwards, it compares this prediction with the observed next image, determines if any of the clusters detected appear in both images with a correlation analysis, and labels the clusters for tracking (pySTEPS, 2021).

Chapter 3

Investigation Part I: Evaluation of the GOES-16 clear sky mask using ceilometers

Satellite products follow a hierarchical scheme based on the correct identification of clouds, therefore an evaluation of the cloud mask is fundamental to establish quality control over satellite products. This is especially important for NASA and NOAA's GOES-R satellite series. Because of its relatively recent launch and start of operation, there has not been an extensive amount of research done to assess its products. Many studies (Ackerman et al., 2008; Joro et al., 2010; Letu et al., 2014; Salek and Szabo-Takacs, 2019) have used secondary ground-based data sets, for example, ceilometer data, to evaluate satellite cloud products such as the cloud mask. And although the GOES-16 Clear Sky Mask has been evaluated by Jimenez (2020) by comparing it with the cloud mask provided by the CALIPSO[1] and SEVIRI[2] retrievals in the continental US, Jimenez (2020) only used satellite data and did not include ground-based data and excluded the region of Mexico.

The validation of the GOES-R products over Mexico is important and necessary because large areas of the country do not contain many or any instruments that monitor weather conditions. The GOES-16 satellite has complete coverage of Mexico, giving an opportunity to improve atmospheric observations over the country. Therefore, verifying that this instrument and the algorithms used to process its data give an accurate representation of the atmospheric conditions of the area is an important step for advancing atmospheric and

---

[1] Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation

[2] Spinning Enhanced Visible and Infrared Imager

climate research and weather monitoring in the region. In this work, the GOES-16 Clear Sky Mask product developed by Heidinger and Straka (2013) was validated over the region of central and northern Mexico, using three group-based Vaisala CL31 ceilometers from the UNAM's Atmospheric Observational Network (RUOA) (RUOA; Peralta et al., 2016): the UNAM, JQRO, and ERNO ceilometers. In addition, one important part of this evaluation was to assess the processing of the ceilometer data itself. The ceilometer data was processed using two different algorithms: the algorithm provided by the manufacturer Vaisala (Vai, 2015), and an algorithm developed by Kuma et al. (2021) named ALCF. Because two different algorithms were used, it was important to see if the results given by each were consistent with one another.

3.1 GOES-16 Clear Sky Mask data

The main satellite product used was the ABI cloud mask (ACM) due to its use in many other algorithms to obtain other products (such as cloud height and cloud top temperature). This is a GOES-R level 2 product labeled "Clear sky mask", referred to as "cloud mask" in this work, available from April 2017 to the present. In brief, this product is obtained by applying an algorithm that analyzes each pixel (and the pixels around it) in various spectral ABI bands to evaluate if it has the right physical characteristics that describe cloudy conditions (Heidinger and Straka, 2013). Although the algorithm categorizes each pixel into four categories (clear, probably clear, probably cloudy and cloudy), the cloud mask gives a binary final results: 0 for clear and 1 for cloudy. The cloud mask product has a 2km by 2km horizontal resolution and the algorithm has an accuracy of 87% (Heidinger and Straka, 2013).

3.2 Ceilometer data

The main source for the ceilometer data was the Atmospheric Observational Network of the UNAM (or RUOA for its initials in Spanish; Peralta et al., 2016). The ROUA ceilometer network consist of the three stations shown in Figure 3.1: ERNO (northernmost station), JQRO (second northernmost station) and UNAM (southernmost station). This network uses the Vaisala CL31 Ceilometer, which has been previously used to evaluate satellite products (Joro et al., 2010; Salek and Szabo-Takacs, 2019) and measures three cloud layers simultaneously with a pulsed diode laser LIDAR. The instrument has a measurement range of 0.0 to

7.6 km, a reporting cycle of 2 to 120 seconds (programable feature; 2 seconds for the ROUA network) and a reporting resolution of 5 meters (Vai, 2020).

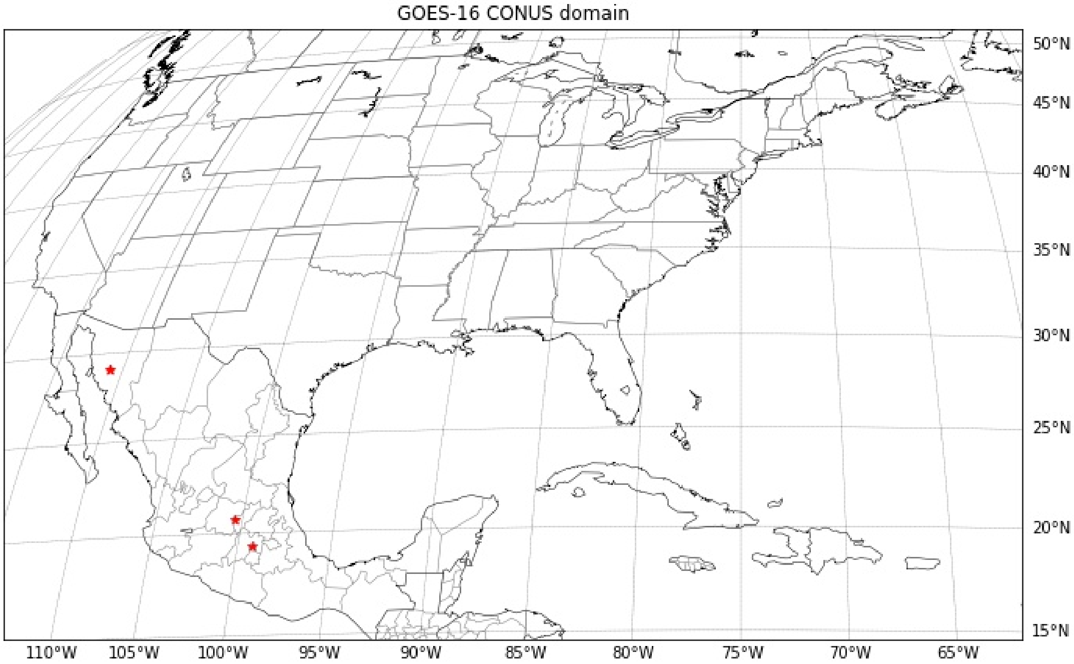


Figure 3.1: Map of the GOES-16 Continental US (CONUS) coverage with the three ceilometer RUOA stations (red stars).

The ceilometer data was processed by two different algorithms. The first algorithm, provided by Vaisala, reads the backscattering data measured by the instrument, filters noise and determines whether they represent cloud, no-cloud or other possible alternatives such as fog or aerosol content. Its output is divided into several sets of variables that include backscattering data, cloud base height in three levels, and 'detection status' (whether or not the ceilometer detects significant backscattering) every 15 seconds (Vai, 2015). For this study, the main information used was the 'detection status' and the data is referred to as "Vaisala Data".

The second algorithm used was the Automatic LIDAR and Ceilometer Framework (ALCF) developed by Kuma et al. (2021) of the University of Canterbury (New Zealand) and the National Institute of Water and Atmospheric Research (NIWA). This algorithm reads the backscattering data measured by the instrument, filters any possible noise within the signal and uses a backscattering threshold values to determine if the data is a cloud or not. If it is greater than the threshold value, it is considered a cloud. The ALCF algorithm provides backscatter data, cloud base height, and cloud mask data, among other variables (ALC, 2020). Because these ceilometers give measurements every 2 seconds, there was a lot of noise in the data that needed to be removed in order to obtain good results with ALCF. For this algorithm, the noise was removed by processing, averaging, and obtaining values every 5 minutes. This data is referred to as “ALCF” ceilometer data.

3.3 Rain gauge data

To further support and solidify the results, it is helpful to used another type of data set and compare the results. Furthermore, it is also necessary to assess the ceilometer data to have assurance the measurements of these instruments are accurate and can give an accurate assessment of the cloud mask. This evaluation was done using rain gauge data from the same meteorological stations where the ceilometers used in this study are found. This data set comes from a Texas Electronics tipping-bucket rain gauge (TR-525M). This particular rain gauge has a resolution of 0.1 mm and an uncertainty of 1% with a 50 mm per hour accuracy. They can report measurements per hour or minute (TE-, 2020), in the case of this study, a 1 minute resolution was used.

3.4 Methodology: differences in resolutions

The main objective of this chapter was to compare the GOES-16’s cloud mask with the results from the ceilometer. However, there are two main obstacles to sort before performing the comparison. The first is the difference in spatial resolution and the second is the difference in temporal resolution.

3.4.1 Spatial resolution

The biggest issue that can add uncertainty and error to the final results is the difference in spatial resolution. The ceilometer gives information in one single point in space (the location of the ceilometer). On the other hand, the satellite provides its data in a grid, with one data point covering a 2 km by 2 km area (called a pixel). The best way to compare these two data sets is by comparing the ceilometer data with the cloud mask data of the pixel that covers the ceilometer.

For the GOES-16 satellite ABI products, the coordinates of each data point in the data matrix provided correspond to the center of the pixel, therefore, the pixel that covers each ceilometer station is the pixel whose center is closest to the each ceilometer station (referred to as "closest pixel" from this point forward). This was done by converting the (x,y) coordinates given by the satellite to geographic (latitude-longitude) coordinates. The Python *pyproj* library was used to convert the coordinates and define the satellite projection with the basic satellite characteristics (altitude, sub-satellite point, inclination, eccentricity). A nearest-neighbor approach was applied to find the closest pixel to each station. First, an area around the station was searched for any satellite pixels within it. This area was initially 0° in diameter with the station in the center (meaning only considering the latitude/longitude of the station) and was increased by 0.02° ($\approx$ 2 km.) in diameter until one satellite pixel center was found. This pixel would be considered the closest pixel to the ceilometer station. The closest pixel value obtained will change depending on the GOES-16 scan strategy or the GOES-16 product used. For the cloud mask grid, the results obtained from this method can be observed in Figure 3.2. Given the methodology used, all stations are within 2 km away from the closest pixel.

In addition to comparing the data of the ceilometer with the cloud mask data of the closest pixel, the ceilometer data was also compared to the average cloud mask of the data of the 3 by 3 pixel or (or 6 km by 6 km) grid with the closest pixel in the center, to further try to decrease the error due to lack of consistent spacial resolution between the two data set. By using this larger pixel area, it was assured that the whole spatial coverage of the ceilometer is covered by the satellite data. It should be noted that this methodology of comparing the ceilometer data with the 3 by 3 pixel grid and the closest pixel cannot eliminate the possibility of cases where small clouds that can only be detected by one instrument, but it will decrease the uncertainty from the lack of complete spatial resolution synchronization.

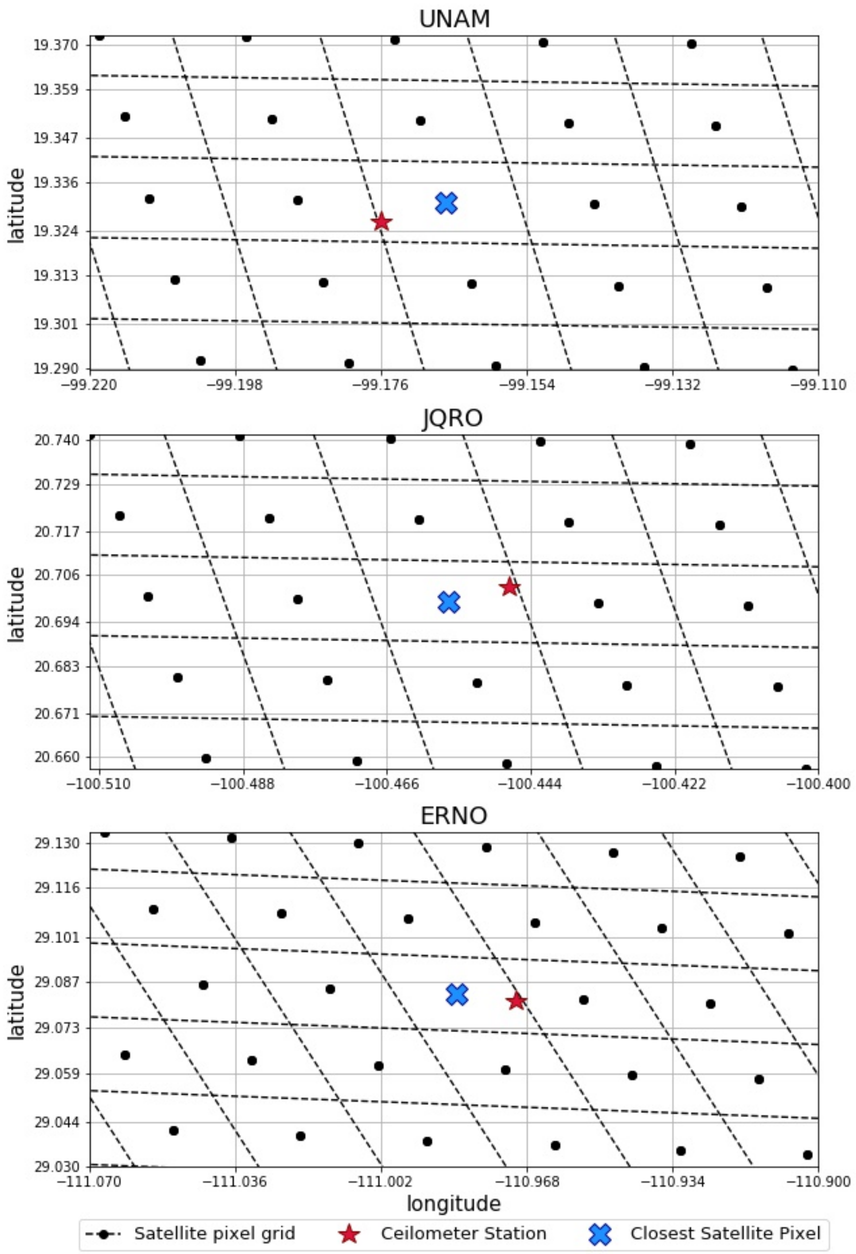


Figure 3.2: Satellite grid with pixel center (dashed lines with black dots) and the location of the ceilometer station (red star) and the nearest pixel (blue ×).

3.4.2 Temporal resolution

The difference in temporal resolution was the second issue that added uncertainty to the overall results. The ceilometer's two processing algorithms result in a temporal resolution of 15 seconds for the Vaisala data and 5 minutes for the ALCF data set. GOES-16 Continental

US coverage has a resolution of 5 minutes. The main concerns were if there was any offset between the intervals, especially the larger time intervals, and which data value was to be used for the comparison (meaning whether to use the value at a given time or an average value over a period of time).

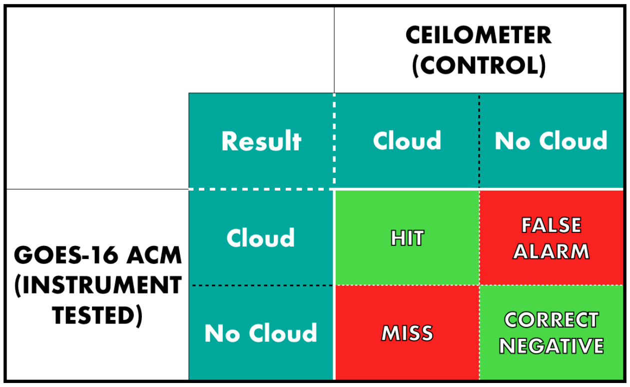
To try to reduce the possible uncertainties and errors that can arise from the difference in temporal resolutions, a data resampling technique (using the Python *Pandas* Library) was used to resample the data at 5, 10 and 15 minutes intervals. In other words, the mean or maximum value of every 5 min/10 min/15 min interval was calculated using the data that falls within the given interval, starting from 0 hours, 0 minutes, 0 seconds of day 1. The interval of 5 minutes was picked because this corresponds to the resolution of both data sets (considering the ALCF algorithm) and the bigger time intervals were used to assess the cloud persistence. To put it differently, since the data is not synchronized temporally, the case where a cloud could be present at the time of the satellite measurement but not at the time of ceilometer measurement cannot be ruled out, and thus this adds to the uncertainty of the results. To reduce this error, a larger interval is analyzed; if a cloud is present throughout this interval, both instruments will have measurements of it. However, like previously said, this did not completely eliminate the error due to lack of temporal synchronization because it cannot take into consideration small clouds that pass fast through the area of the cloud-mask measurement.

3.5 Methodology: contingency table analysis

The evaluation of the cloud mask was based on the work of Joro et al. (2010), who used a contingency table analysis to asses the two different cloud masks[3] with ceilometer data over Finland.

---

[3] the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) Meteorological Products Extraction Facility cloud mask and the Nowcasting Satellite Application Facility (SAFNWC) cloud masks provided by Météo-France



		CEILOMETER (CONTROL)	
	Result	Cloud	No Cloud
GOES-16 ACM (INSTRUMENT TESTED)	Cloud	HIT	FALSE ALARM
	No Cloud	MISS	CORRECT NEGATIVE

Figure 3.3: Example of contingency table used in this study to compare ceilometer data with the GOES-16 satellite cloud-mask (ACM) data. Green squares mark positive results and red squares mark negative results.

The first step was obtaining the contingency tables (Figure 3.3). For this evaluation, this table included the number of times both the satellite and ceilometer indicated the presence of a cloud (*Hit*), the number of times the satellite indicated cloud presence but the ceilometer did not (*False Alarm*), the number of times the satellite and the ceilometer did not detect the presence of a cloud (*Correct Negative*), and the number of times the satellite did not detect a cloud but the ceilometer did (*Miss*). From these contingency tables, the following statistical measurements were obtained: bias, false alarm ratio, false alarm rate, and the hit rate. These measurements were picked because they have been used by previous evaluation/validation studies (Jolliffe and Stephenson, 2003; Barnes et al., 2009; Joro et al., 2010). The *bias* is the ratio of cloud frequency to the observed cloud frequency, indicating whether the satellite overestimates (above 1) or underestimates (below 1) cloud values (Wilks, 2006), and is calculated by

$$BIAS = \frac{hits + false\ alarms}{hits + misses} \tag{3.1}$$

The FAR is the fraction of forecast hits that were false (false alarms) (Wilks, 2006) and is given by

$$FAR = \frac{false\ alarms}{hits + false\ alarms} \tag{3.2}$$

The PoFD is the probability of false detection (Wilks 2006)

$$PoFD = \frac{false\ alarms}{false\ alarms + correct\ negatives} \tag{3.3}$$

And the PoD is described as the amount of correct forecast compared to the total amount of forecast events (Wilks, 2006) and is calculated with

$$PoD = \frac{hits}{hits + misses} \tag{3.4}$$

Finally, two simple skills tests were used as well: the Kuipers Skill Score (KSS) or True Skill Statistic, which measures the ability to separate a Cloud result from a No Cloud result (Woodcock, 1976):

$$KSS = PoD - PoFD \tag{3.5}$$

and the Heidke Skill Score (HSS) that measures the increase improvement of the forecast over the standard forecast (obtaining a correct result by chance; Wilks, 2006) and is given by

$$HSS = 2 \times \frac{hits \times c.n. - f.a. \times misses}{(hits + misses) \times (misses + c.n.) + (hits + f.a.) \times (f.a. + c.n.)} \tag{3.6}$$

with c.n. being correct negatives and f.a. being false alarms

3.5.1 Comparison sets

The contingency table analysis was made with four different comparison sets:

Set 1: comparison between the data of a 3×3 pixel satellite grid surrounding each ceilometer station and the ALCF ceilometer data.

Set 2: comparison between the 3×3 pixel satellite grid data with the Vaisala ceilometer data.

Set 3: comparison between the data from the pixel whose center is closest to the ceilometer station and the ALCF data.

Set 4: comparison between the closest pixel data with the Vaisala ceilometer data.

In addition, the comparisons were also done for each resampling interval discussed in Section 3.4.2, giving a total of 12 contingency table analysis. Finally, each of the 12 comparisons was also done by time-of-day. These times were determined using the solar angle and are:

Day: solar altitude angle higher than 10°.

Twilight: solar altitude angle between 0^o and 10^o, including both values.

Night: negative solar altitude angle.

The period of analisis was from January 1st, 2019 to December 31st, 2019.

3.6 Results

3.6.1 Automatic LIDAR and Ceilometer Framework data cloud-base height correction

Because two different ceilometer data processing algorithms were used in this study, the Automatic LIDAR and Ceilometer Framework (ALCF) algorithm and the Vaisala algorithm, it was important to analyze the differences in the resulting cloud-detection data they both gave as output. Figure 3.4 shows the percentage of occurrence of the cloud-base heights for the three ceilometer stations for the ALCF data, Vaisala data, and ALCF data without cloud-base heights below 200 m. For the Vaisala data, the cloud-base heights (of the three given) used in this analysis was the lowest cloud-base heights. The results of the ALCF and the Vaisala data show a discrepancy in the processing method since the corresponding graphs for each method did not match, and in particularly, there was an increased number of low-level clouds for ALCF data, especially clouds with base heights below 200 m. This discrepancy can be illustrated even more if the cloud-base height below 200 m of the ALCF data were filtered out. The distribution of the cloud-base heights is more consistent between both algorithms, signaling a problem with the measurements of ALCF data below this height. This large amount of low level cloud-base heights is considered an error because cloud-base heights tend to be 1 km minimum in the region.

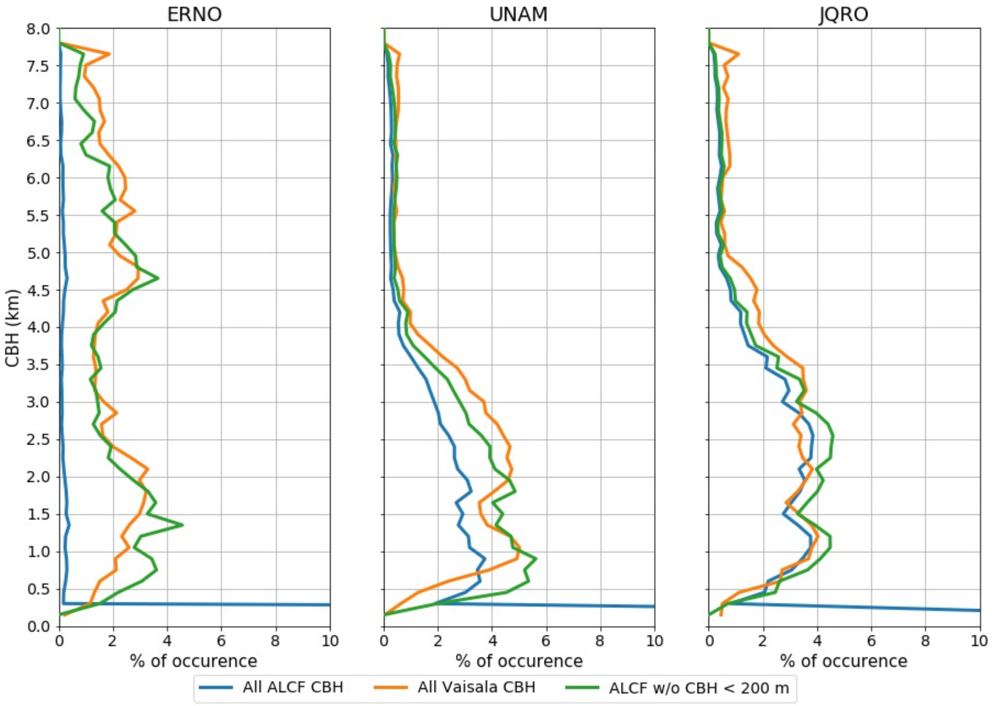


Figure 3.4: Percentage of occurrence of the cloud-base height (CBH) for each station from January 1st to December 31st, 2019, for the ALCF data (blue), the Vaisala data (orange) and the ALCF data without cases with cloud-base heights below 200 m (green). Left graph is for ERNO, middle graph is for UNAM and right graph is for JQRO.

The second analysis performed was the creation of contingency tables comparing both sets of ceilometer data, similarly to the table in Figure 3.3. To consider possible problems in distinguishing low-level clouds, the analysis was done comparing the complete sets of Vaisala and ALCF data, as well as only comparing the cases where the ALCF cloud-base heights was 200 m or above with the corresponding Vaisala data. The results can be observed in Table 3.1. For JQRO and UNAM, there were more positive results (correct negatives and hits) than negative results (false alarms and misses) by almost 60%. These positive results also increased when limiting the cloud-base height minimum to 200 m. This increase was slight for JQRO (between 0% and 2%) and more significant for UNAM (between 6% to 10%). For ERNO, the negative cases outnumbered the positive cases by 11% when considering all cases (approximately 45% positive results and 56% negative results). However, the number of positive cases increased to more than 80% when the minimum cloud-base height of 200 m was applied, indicating that the issues in distinguishing low-level clouds with the ALCF method affects this station more than the other two stations.

	Resamp.	**All**		**≥ 200 m**	
		Pos	**Neg**	**Pos**	**Neg**
ERNO	**5 min**	0.442	0.558	0.884	0.116
	10 min	0.445	0.555	0.849	0.151
	15 min	0.447	0.553	0.821	0.179
UNAM	**5 min**	0.816	0.184	0.897	0.103
	10 min	0.808	0.192	0.880	0.120
	15 min	0.801	0.199	0.867	0.133
JQRO	**5 min**	0.811	0.189	0.835	0.165
	10 min	0.769	0.231	0.790	0.210
	15 min	0.741	0.259	0.760	0.240

Table 3.1: Normalized number of positive and negative results of the contingency tables comparing the ALCF data and Vaisala data for each category indicated per station. "≥ 200 m" indicates the results for the cases with cloud base height of 200 m or above. "Resamp" indicates the resampling interval.

Expanding on these results, the scatter plots and Pearson R correlation coefficients between the lowest cloud-base heights of the Vaisala and ALCF data were obtained for the exact same date-time. Figure 3.5 shows the results of the scatter plots of each station, where although there appears to be a linear pattern in the data, a large amount of small cloud-base heights in the ALCF data can be clearly seen (near the y-axis). This is further evidence that the problems distinguishing low level clouds only affects the ALCF ceilometer because a large amount of ALCF cases with low cloud-base heights correspond to Vaisala results of higher clouds. The correlation results of these plots and the correlation comparing the cloud-base heights of both data sets with the minimum limit of 200 m are found in Table 3.2. The results indicate that the Pearson's R correlation increased when the low cloud-base heights were eliminated and the relationship was close to linear, as expected since both methods processed the same base ceilometer raw data. To guarantee the problem was with ALCF and not Vaisala data, the same analysis was carried out but eliminating Vaisala data below 200 m. However, there was very little change for all stations when this was done, indicating these cases do not have a significant impact on the overall results.

	ERNO	**UNAM**	**JQRO**
All	0.343	0.619	0.641
200 m	0.894	0.809	0.840

Table 3.2: Pearson R correlation values between Vaisala lowest cloud-base heights data and ALCF cloud-base heights data for all three ceilometer stations. "All" indicates comparing all values of the same date-time, "200 m" only compares cases with cloud-base heights of 200 m or more of the same data-time.

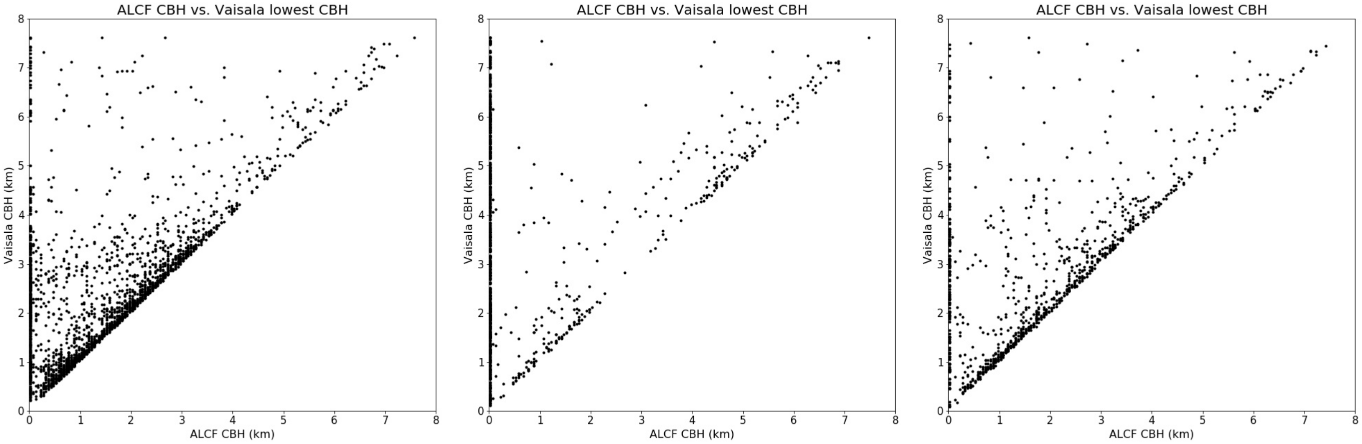


Figure 3.5: Scatter plot of ALCF cloud-base heights data and Vaisala lowest cloud-base heights data for the three ceilometer stations. Left graph is for UNAM, middle graph is for ERNO and right graph is for JQRO.

The previous results show a clear discrepancy between both algorithms for cloud-base heights below 200 m for the ALCF. The results point to a clear flaw in the ALCF algorithm that is not present in the Vaisala algorithm, which is resulting in low cloud-base heights that are not physically possible in the region. When reviewing the documentation and scripts available for the algorithms, the likely cause for this issue is the effect of atmospheric aerosols in the boundary layer on the backscattering signal read by the instrument, especially since the RUOA ceilometers are primarily used to observe the boundary layer. The Vaisala algorithm has the ability to provide information regarding the boundary layer (Morris, 2012), which gives an indication that overall, the algorithm is able to distinguish between clouds and aerosol content backscatter signals. On the other hand, ALCF's analysis of the backscattering data to detect clouds (the removal of noise and the use of a threshold value to determine what signals are clouds and no clouds) cannot distinguish between the backscattering signals resulting from high aerosol concentration and the signal resulting from clouds. Although the

ALCF algorithm's backscattering threshold value that determines what is cloud and no-cloud can be modified by the user, the predetermined value was used in this study. Further analysis can be done to obtain the best backscattering threshold value for each ceilometer used, however it was determined that the use of the data processed by the Vaisala algorithm, and the fact that both data sets gave consistent results when removing all cases with cloud-base heights below 200 m for the ALCF data, was sufficient to obtain acceptable results for this study. Thus, the Vaisala processed data and the ALCF processed data with cloud-base heights of 200 m and above were used in the GOES-16 cloud mask validation, but with more weight given to the Vaisala processed data results.

3.6.2 Contingency table results between the satellite cloud mask and ceilometer data

The results of the contingency tables of all comparison sets[4] for the different resampling intervals are shown on Figures 3.6, 3.7 and 3.8. The graphs show the relative number of positive results (hits and correct negatives) and negative results (false alarms and misses). The results show that the amount of negative results are all below 40% for all three stations. For UNAM, between 17% to 23% are negative results; for JQRO, it ranges from 20% to 37% and for ERNO, the negative results make up around 27% to 35% of all. In terms of the statistics, PoD had the highest value for all stations and comparison sets. This value ranged from 0.6 to 0.95, with the highest values obtained for UNAM comparison set 2 (0.95) and the lowest for JQRO comparison set 3 (0.63). For PoFD, it did not exceed 0.5, with the lowest being for UNAM comparison set 3 (0.27) and the highest being for JQRO comparison set 1 (0.48). FAR did exceed 0.5 for some comparison/resampling cases, falling between 0.3 and 0.55. However, it is not higher than PoD in any case. The skill scores show a moderate skill for the cloud mask. The station with the best skill score was UNAM, with scores between 0.5 and 0.65. For JQRO and ERNO, the comparison sets that used ALCF data had low skills scores 0.3, and for the other comparison sets that used Vaisala data, the skills scores were around 0.5. Additionally, both KSS and HSS were consistent with each other. Finally, the last variable (not shown) is bias. The bias was above 1 for all except comparison set 3 of UNAM (which was close to 1). The station with the highest biases was ERNO and UNAM had the lowest. In all stations, comparison set 3 had the lowest bias,

---

[4] Set 1 compares data from the 3×3 pixel satellite grid data and the ALCF data, Set 2 compares data between the 3×3 pixel satellite grid with the Vaisala data, Set 3 compares the closest-pixel satellite data with the ALCF data and Set 4 compares the closest-pixel satellite data with the Vaisala data

while comparison set 2 had the highest. Furthermore, the bias decreased with increasing resampling interval, with the biggest differences occurring for ERNO (a difference of around 0.3), and the lowest being for UNAM (around 0.1).

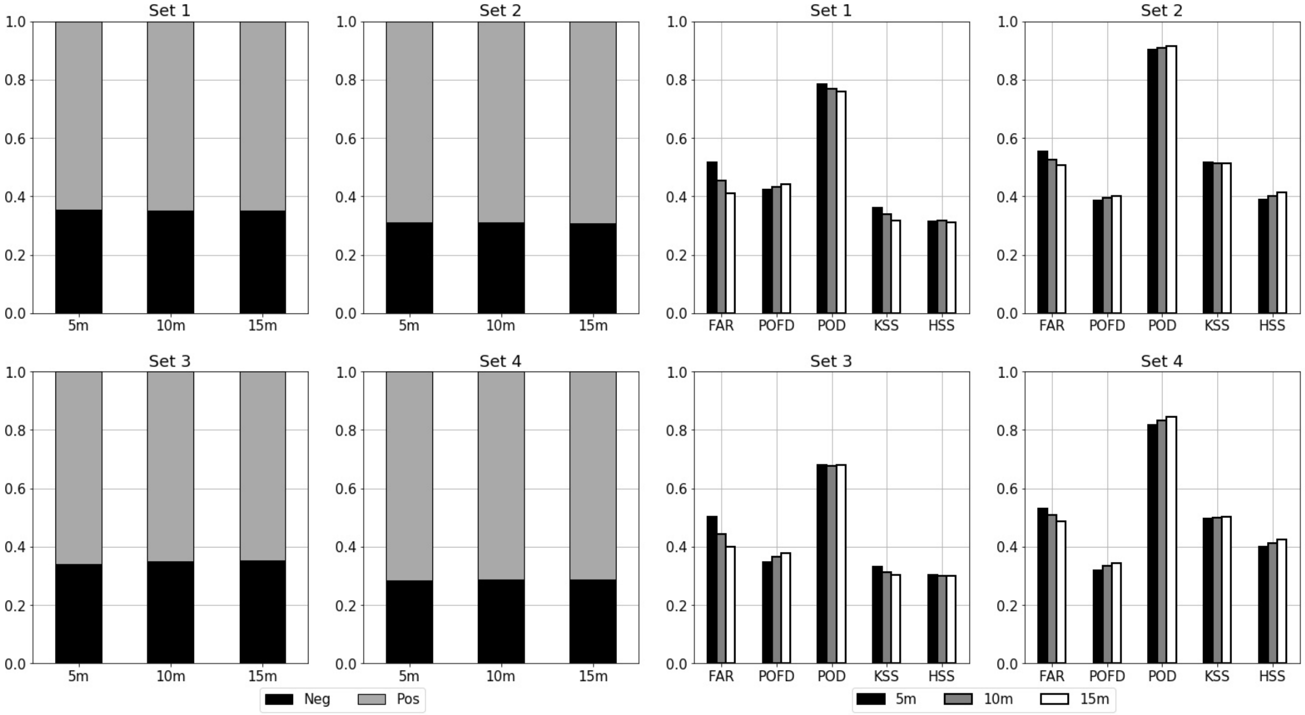


Figure 3.6: Contingency table results for ERNO. The set of bar graphs on the left show the positive (sum of hits and correct negatives) and negative (sum of misses and false alarms) normalized results for all comparison sets and resampling intervals. The set of graphs on the right shows the statistics results for all comparison sets and resampling intervals per station. The comparison sets are as follow: Set 1 compares data from the 3×3 pixel satellite grid data and the ALCF data, Set 2 compares data between the 3×3 pixel satellite grid with the Vaisala data, Set 3 compares the closest-pixel satellite data with the ALCF data and Set 4 compares the closest-pixel satellite data with the Vaisala data.

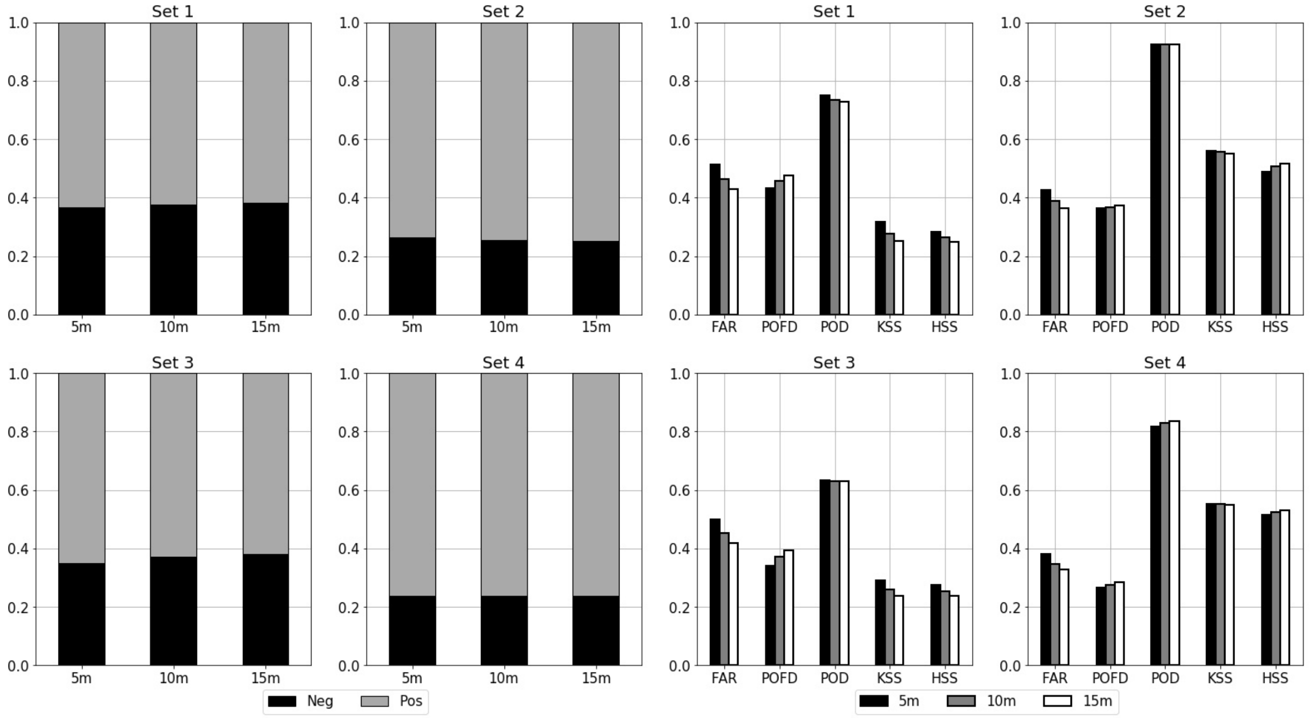


Figure 3.7: Contingency table results for JQRO (same as Figure 3.6).

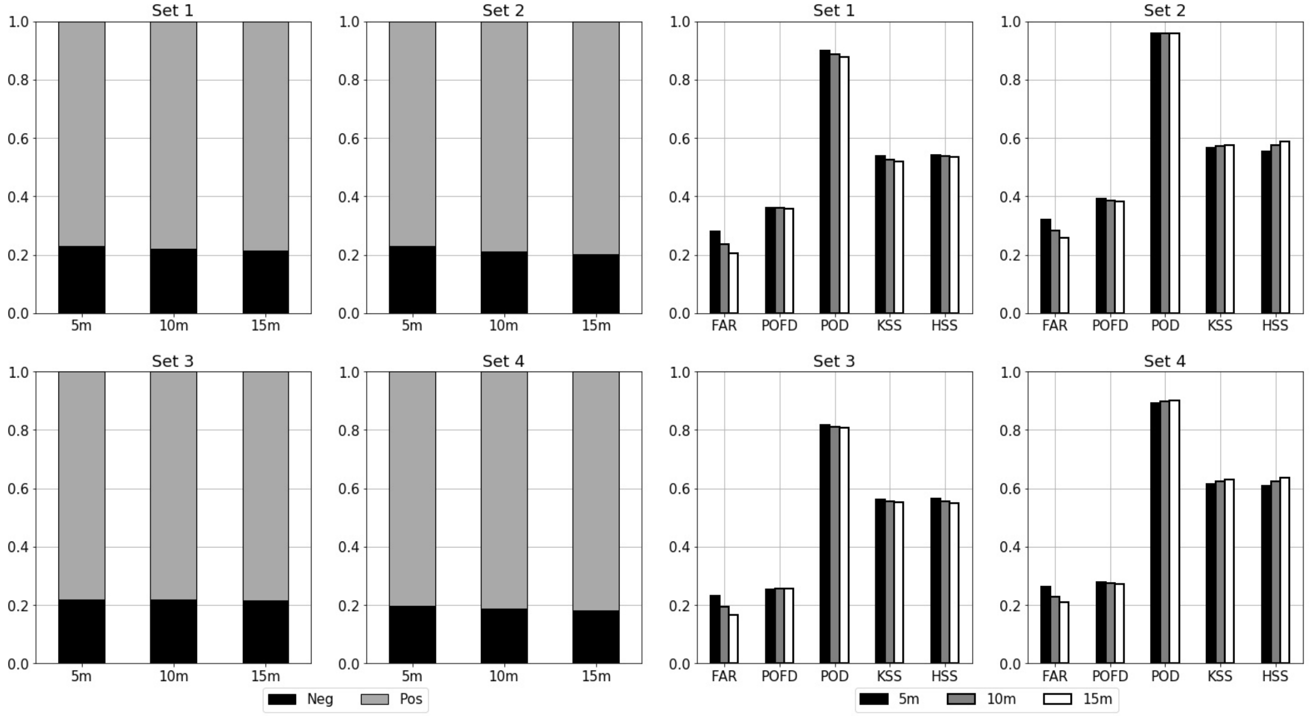


Figure 3.8: Contingency table results for UNAM (same as Figure 3.6).

Overall, the best results were obtained using Vaisala processed data due to the fact that this ceilometer data set is the most consistent (see previous section). The station with the best results was UNAM. In terms of the spatial variation of the GOES satellite data (3x3 grid surrounding the station and individual closest pixel to the station), the UNAM station obtained the best results using the individual pixel (comparison set 4 had the overall best results), however, JQRO and ERNO obtained better results with the 3x3 pixels. The variation between the different resampling intervals is not consistent in terms of positive/negative results, variable, comparison set and station to determine with certainty which interval is best. The biggest difference is observed for the variable bias, was discussed earlier, and FAR (5 min interval is bigger by 0.1 than the result for the 15 min interval) and it is smaller for PoD. When only considering skill scores, the comparison sets using Vaisala data have better scores for the 15 min resampling interval and the comparison sets using ALCF data have better scores for the 5 minute resampling interval, although the difference is less than 0.04. When observing positive and negative results, the highest positive results percentage was for 15 min interval when using Vaisala data. However, for JQRO and ERNO, the differences are very small. When using ALCF data, 5 min resampling interval had slightly better results for JQRO and ERNO, and almost no variation between all intervals for UNAM station.

Figures 3.9, 3.10, and 3.11 show the results for different time of the day intervals for all stations. The figures show that the best overall results were for the cases at “night”, which had the highest PoD, skill scores, and a bias nearest to 1 for ERNO, UNAM, and the Vaisala comparison sets of JQRO, followed by “day” and then “twilight”. Additionally, the behavior of the results for all comparison set were consistent with each other. The fact that “night” had some of the best results is somewhat unexpected since at nighttime, the cloud mask algorithm loses the visible band range, and in preliminary accuracy estimates of the algorithm (performed before the launch of the satellite using a proxy-cloud mask obtained using CALIPSO[5] and SEVIRI[6] data in place of GOES-16 data), the PoD was slightly less for nighttime and “False clear detection” (miss) was the highest, although “false cloud detection” (false alarm) was lower at nighttime than daytime (Heidinger and Straka, 2013). The difference in the results can be explained by the fact that the analysis done by Heidinger and Straka (2013) compares the proxy GOES-16 cloud mask obtained with the GOES-16 algorithm with the CALIPSO satellite cloud mask. While these are two different

---

[5] Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation

[6] Spinning Enhanced Visible and Infrared Imager

satellites, they give similar data set. On the other hand, this study is comparing the data from two different instruments, with very different spatial resolutions, and different limitations and advantages. While the GOES-16 satellite is more limited at night due to the loss of the visible spectrum, the ceilometers are less likely to be affected by atmospheric aerosols and precipitation at night in the region of study. Considering this, the improvement of the results at night could be due to the improved atmospheric conditions at night that can cause less flawed data with the ceilometer (especially for the ALCF data). Lastly, an important point to consider in the analysis of these results is the number of cases for each time-of-day, with "night" having the greatest number of cases and twilight having the least.

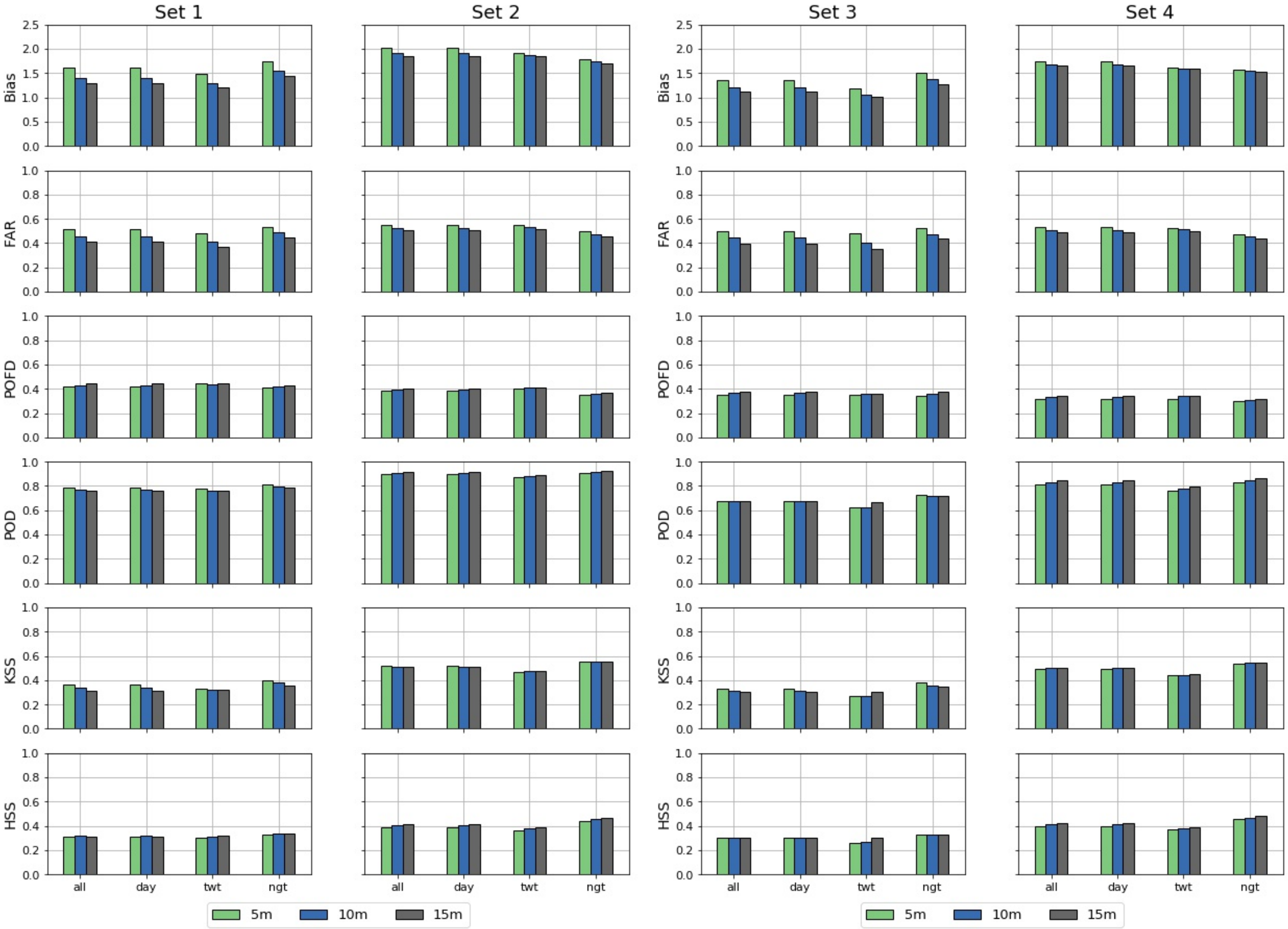


Figure 3.9: Results of the statistics from the contingency table results for all comparison sets and resampling intervals divided by time-of-day for ERNO station. All = all day, day = daytime cases, twt = twilight cases and ngt = nighttime cases.

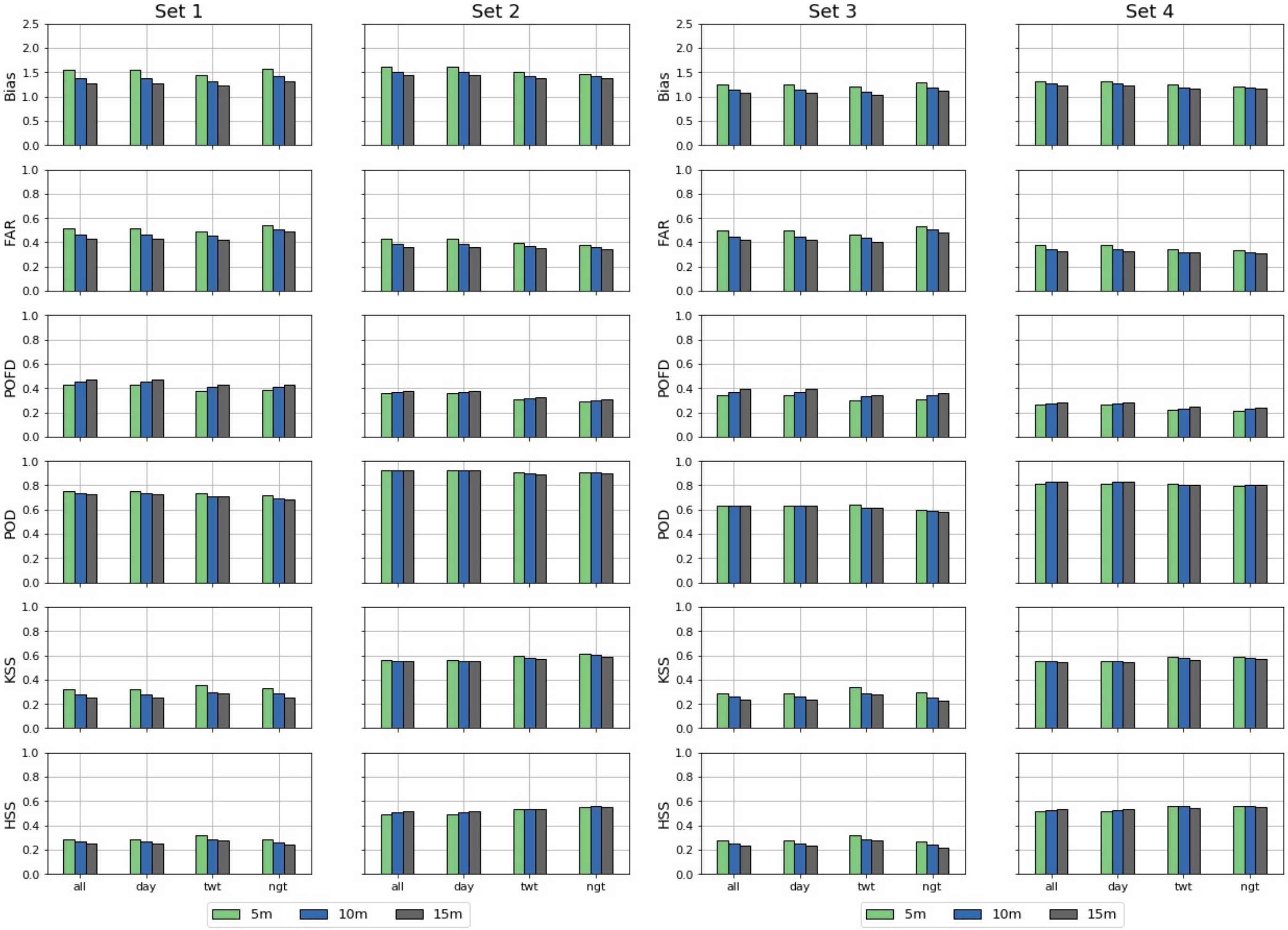


Figure 3.10: Same as 3.9 for JQRO station.

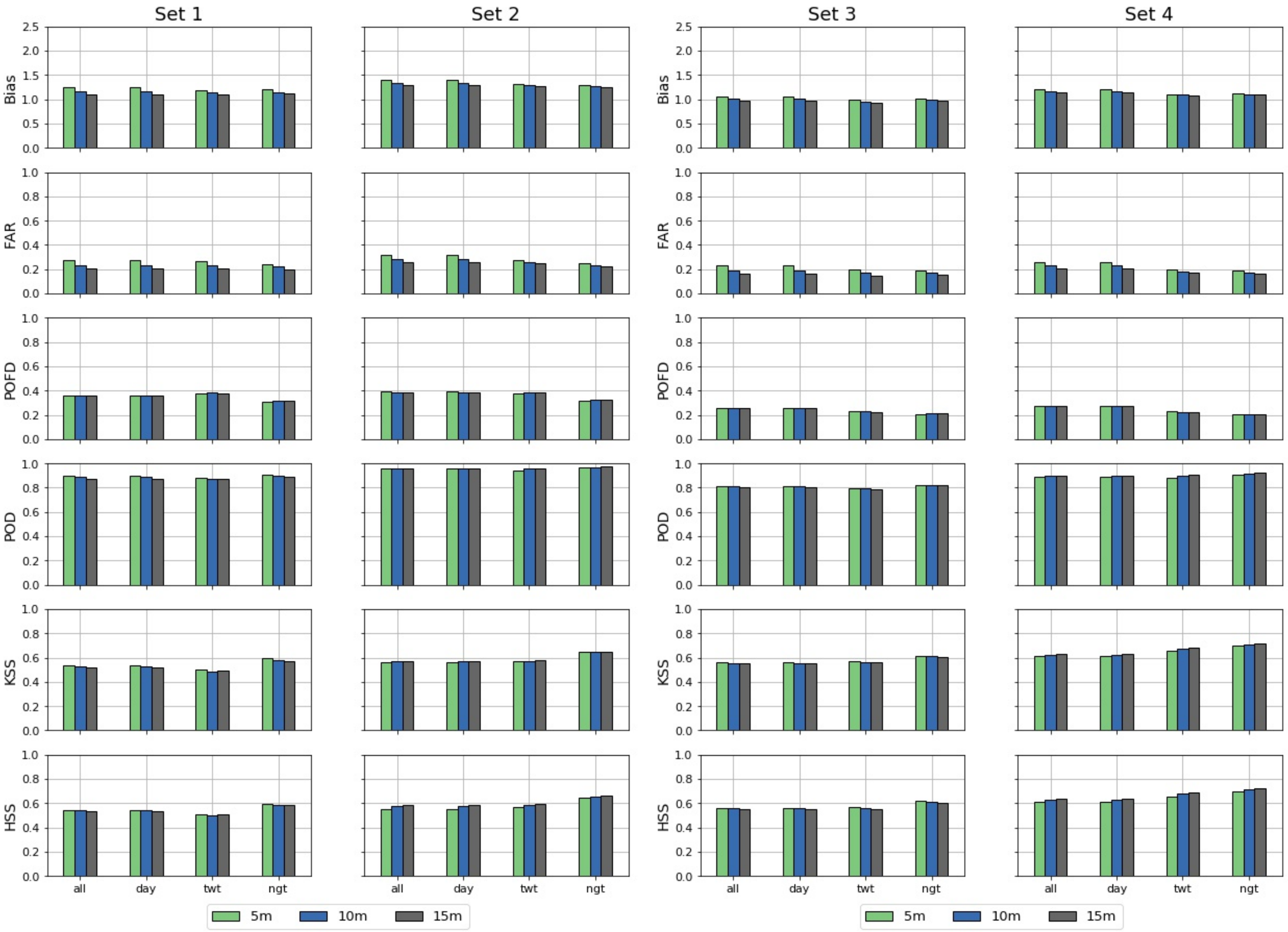
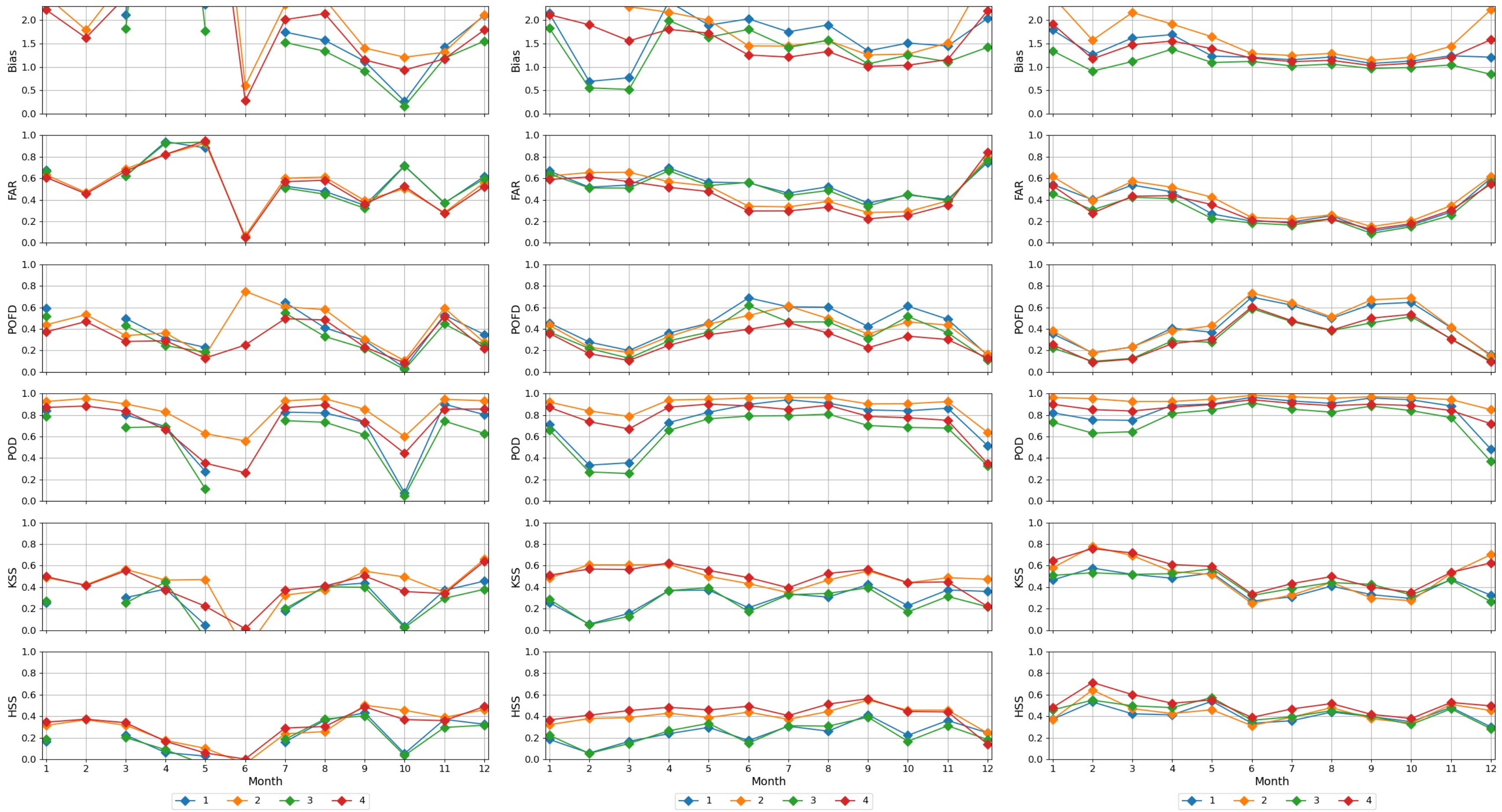


Figure 3.11: Same as 3.9 for UNAM station.

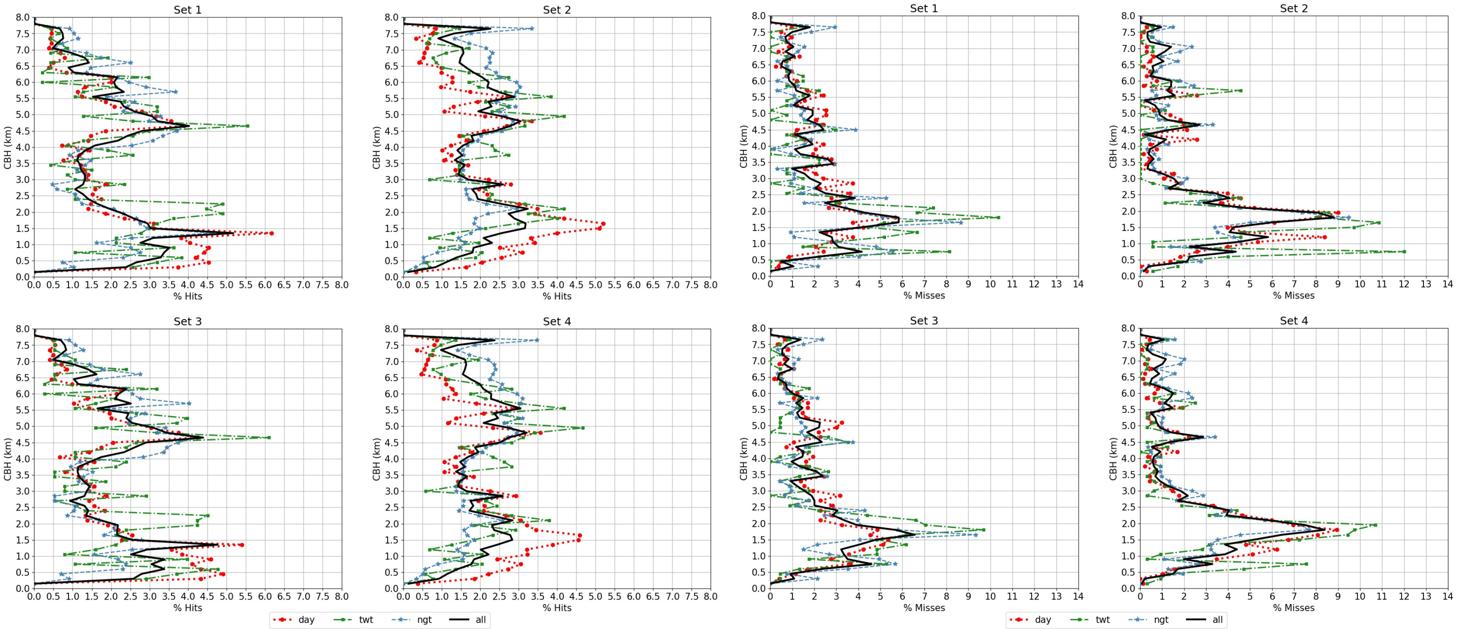
When observing the statistical results per month of Figure 3.12 there are no clear patterns. First, the data for ERNO is not very reliable since there is a large amount of missing data for the initial months. This is due to the fact that this station has the most data removed when applying the 200 m filter to the cloud-base height of the ALCF data. When observing the Vaisala comparison sets, there is no clear pattern because of sudden changes in several variables in the months of May, June and October. On the other hand, UNAM had more complete and consistent results throughout the year. FAR, bias, and both skill scores (although to a lesser extent) are high at the beginning of the year, then start to decrease from April to September, with a subsequent increase the rest of the year. The opposite occurs for PoD and PoFD. This patter is most likely due to the rainy season in Mexico for UNAM and JQRO, which runs from late May to early October. During this period of the year, more convective clouds occur, creating more cases and more robust statistics. It should be noted that both January and December had several days missing (for the ceilometer), which explains the inconsistent and, in some cases, lower accuracy of the cloud mask during these months in comparison to the rest of the year. JQRO had similar results, with the exception of February and March, which had low bias, PoD and skill scores for the ALCF data.



(a) (b) (c)

Figure 3.12: Results of the statistics for all comparison sets for the 5-minute resampling interval per month for (a) ERNO, (b) JQRO and (c) UNAM.

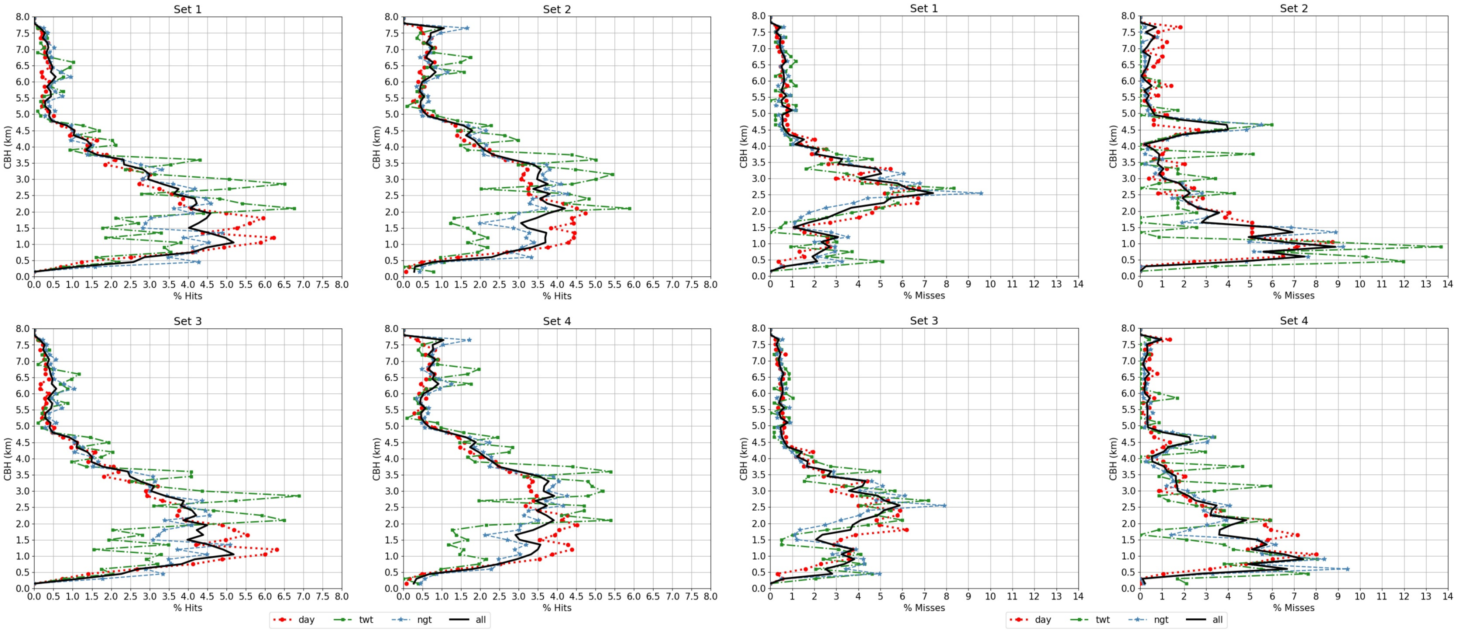
On Figures 3.13, 3.14 and 3.15, the percentage of hits and misses with respect to the cloud-base height is presented for each station at a 5 min resampling interval. The graphs are divided into time-of-day as well as for all day. In general, the pattern observed is similar to the cloud-base height distribution observed in Figure 3.4. The results show that for UNAM, low-level cloud-base heights had the highest hit percentage (all hit percentage results are below 7%) and decreased around 3,000 m and reaching its minimum around 4,000 m. There was also a small dip around 1,500 m. Below 2,000 m, the night cases had the highest hit percentages and day cases had the lowest. Between 2,000 m and 4,000 m, the night cases have the lowest, and the day and twilight cases have similar numbers above night cases. For the misses at UNAM, the behavior was similar, although it is skewed by the high percentage values of the twilight cases. It was observed that, in general, for all stations, twilight cases have a more chaotic behavior in all graphs because it has the lowest number of cases (less than 10% of the total). Nevertheless, it can be observed that the misses percentage is highest for low cloud-base heights, especially for twilight cases, followed by night cases and then day cases. Additionally, for the Vaisala comparison sets, the misses increase during the day for high cloud-base heights (>6,000 m). For JQRO, the behavior of both hits and misses percentages were similar to those of UNAM, however, day cases had the highest hit percentages for low-level cloud-base heights and the decrease started around 4,000 m. The misses percentages were lower for low cloud-base heights than in the UNAM station, there was also a peak between 4,000 m and 5,000 m for the Vaisala comparison sets and a dip between 1,000 m and 2,000 m for the twilight and night cases for the ALCF comparison sets. For ERNO, the behavior of the misses percentages had a similar general behavior (high for low cloud-base heights), however, the data is much more noisy. For the hits percentages, the behavior differs for the mid-cloud-base heights values (between 2,000 m to 4,500 m), where the percentages decrease and then increase when surpassing the 4,500 m to values similar to those of the low cloud-base heights.



(a) Hits

(b) Misses

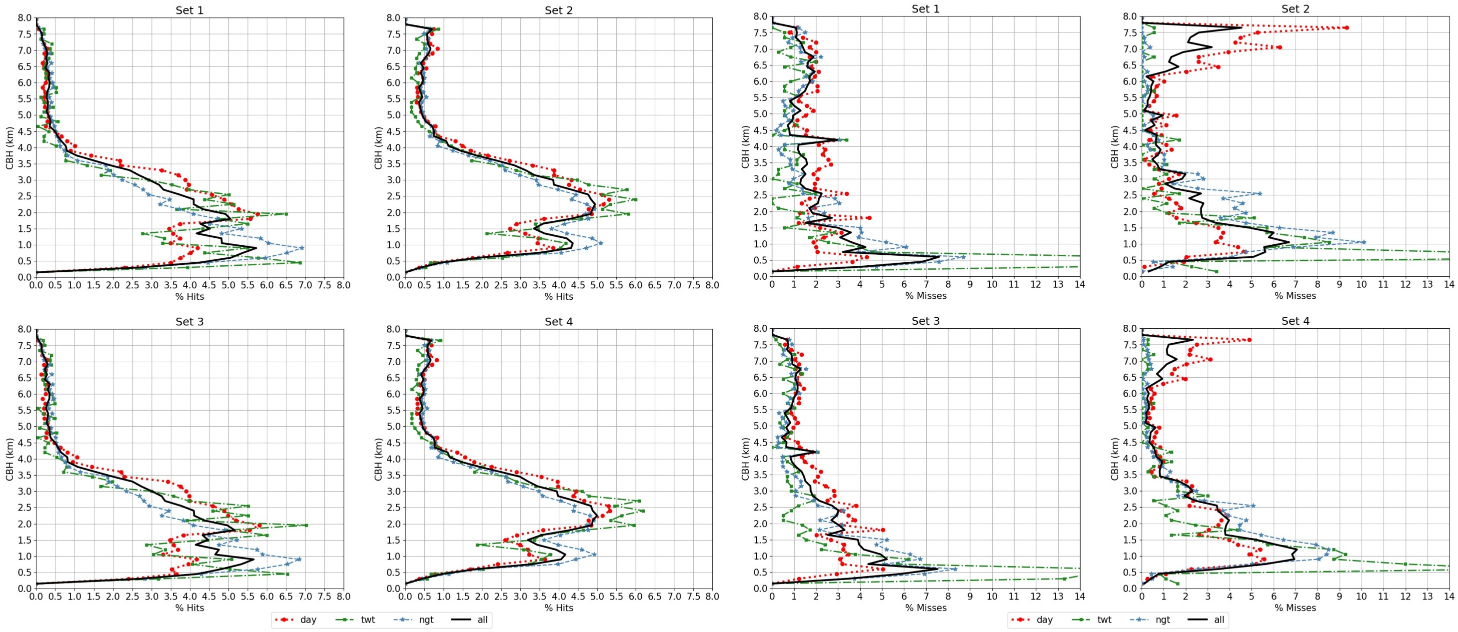
Figure 3.13: Percentage of (a) hits and (b) misses per 250 m interval of cloud-base heights at a 5-minute resampling interval divided into time-of-day as well as all day for ERNO.



(a) Hits

(b) Misses

Figure 3.14: Same as 3.13 for JQRO station.



(a) Hits

(b) Misses

Figure 3.15: Same as 3.13 for UNAM station.

3.6.3 Using rainfall data to evaluate ceilometer data

The assessment of the ceilometer data with the rain gauge data was done using the same contingency table study. This involved all of the Vaisala and ALCF data sets, as well as the ALCF dataset with a minimum limit of the cloud-base height of 200 m. The rainfall data is given every minute while the ALCF data is given every 5 minutes and the Vaisala data every 15 seconds. For this reason, resampling intervals of 1-minute and 5-minutes were used. For the Vaisala data, the 1 minute resampling interval was only used with two different definitions of a cloud reading and no-cloud reading. The first definition of cloud measurement within each resampling interval is that the average of the the whole interval has to be above 0.0 (meaning at least one reading within the interval is cloud) and the second definition is that the average of the interval has to be above 0.5, which indicates that at least 50% of the readings had to be cloud within the interval.

For this contingency table analysis, although correct negatives and false alarms were calculated, they do not give any information on the accuracy of the ceilometers. False alarms indicate that the ceilometer detected a cloud but there was no rain measured at the same interval, something that can occur. And correct negatives indicate that there was no rain and no cloud detected. This is a positive result, however, it does not give information on whether or not the ceilometer can correctly detect a cloud or correctly detect no-cloud (clear sky) on days without rain. Thus, only cases where there was a rain measurement (hits and misses) can be used to asses the ceilometers with the rain gauge data.

On Table 3.3, the percentage of misses and hits (only considering rainfall cases) are presented for all comparison data sets. It is clear that for Vaisala, the ceilometer only missed less than 3% of the time, and the best results (only slightly) occurred when considering a cloud measurement being a measurement of the whole sampling interval being greater than 0. For the ALCF data, the results improved by 2% from those obtained using Vaisala data for UNAM and ERNO, but for JQRO, the hit percentage decreases by 20%. A possible reason for this large amount of change in JQRO is the amount of cases used to calculate these percentages. For the Vaisala data, there were 4,221 cases for UNAM, 1,221 for JQRO and 2,430 for ERNO, which is a good sample population for this study. For the ALCF data, UNAM had 1,997 cases for the 5 minute resampling, ERNO had 1,060 and JQRO only had 672 cases. And for the 1 minute resampling, these cases decreased by around half for all stations. These numbers indicate that Vaisala had almost twice as many cases to work with, mainly due to the different temporal resolution of both data set, which give more accurate statistics than ALCF. For JQRO, the amount of cases used with the ALCF data was far fewer than all other comparison sets, which can affect the overall results and can explain the large variation between the Vaisala results and ALCF results. Using resampled and the limited ALCF data lead to the elimination of around 92% of all possible rainfall cases available for JQRO, consequently producing less accurate statistical measurements. Nonetheless, the results do indicate that the ceilometers can detect the presence of clouds when rainfall occurs with high accuracy, an important results since ceilometers have been known to fail during rainfall.

		Vaisala		**ALCF**	
		>0.0	**>0.5**	**5 min**	**1 min**
UNAM	**Hit%**	99.1	98.2	99.2	99.4
	Miss%	0.9	1.8	0.8	0.6
JQRO	**Hit%**	99.8	99.5	79.2	78.5
	Miss%	0.2	0.5	20.8	21.5
ERNO	**Hit%**	97.9	97.0	99.3	99.2
	Miss%	2.1	3.0	0.7	0.8

Table 3.3: Results of the hits and miss percentages (only considering cases where rain was measured) of the contingency tables between rainfall data and ceilometer data. For the Vaisala data, ">0.0" indicates that a cloud hit was defined as a measurement bigger than 0.0 and ">0.5" indicates that the cloud measurement had to be a measurement bigger than 0.5. For the ALCF data, "5 min" indicates a 5 minute resampling and "1 min" indicates a 1 minute resampling.

3.6.4 Using rainfall data to evaluate the satellite cloud mask

Like the ceilometer data, a contingency table analysis was done with the cloud mask using the rain gauge data. Because these two data sets also had different temporal resolutions (5 minutes for the cloud mask and 1 minute for the rainfall data), two resampling intervals of 1 minute and 5 minutes were used. For the spatial resolution, the pixel whose center is closest to the meteorological station was used for the comparison. Additionally, two different minimum average rainfall within the 5 minute interval were set, above 0.0 mm (1 reading within the interval indicated rain) and above 0.06 mm (more than half the readings within the 5 minute interval were of rain).

The overall results of this analysis can be observed in Table 3.4. Like the analysis between the ceilometer and the rainfall data, only the hits and misses (cases where rainfall occurred) can give an insight into the accuracy of the cloud mask. These results clearly show the cloud mask is very precise, with an accuracy of more than 99%. It should be noted that, like in the analysis between the ceilometer data and rain gauge data, the number of cases analyzed is as large as the number of total cases analyzed in this study, with some stations only reporting less than 500 cases of rain. Regardless, it is enough to conclude that the cloud mask can accurately detect clouds when rainfall is occurring. Additionally, for the 15 total misses, the measured precipitation was below 0.2 mm, which can suggest that its accuracy decreases only slightly with light rainfall.

		1 min	**5 min**	
			>0.00	**>0.06**
UNAM	**Hit%**	99.7	99.8	100.0
	Miss%	0.3	0.2	0.0
JQRO	**Hit%**	100.0	99.8	100.0
	Miss%	0.0	0.2	0.0
ERNO	**Hit%**	99.4	99.6	100.0
	Miss%	0.6	0.4	0.0

Table 3.4: Same as Table 3.3 for the contingency results between the cloud mask and the rainfall data. These were done with two resampling intervals ("1 min" for 1 minute resampling and "5 min" for the 5 minute resampling). For the 5 minute resampling data, ">0.00" means that the rainfall within the interval had to average above 0.0 mm for the whole interval to be considered a rainfall measurement and ">0.06" means the average had to be above 0.06 mm for the whole interval to be considered a rainfall measurement.

3.7 Conclusions

From the results presented in the previous two sections, it is clear that the cloud mask has a good accuracy. For all comparison sets and resampling intervals, the number of positive results were above 60%, the values of PoD were above 0.6, with the highest values reaching 0.95, the PoFD values were below 0.5, and the FAR values were below 0.55. The skill score KSS, which calculates the difference between PoD and PoFD, was above 0.2 for all, indicating that, in general, PoD is higher than PoFD by at least 0.2.

The best results were obtained for the UNAM station, which had a PoD higher than 0.8 and a PoFD below 0.3, with its comparison set 2 having the highest overall PoD of 0.95. The results using Vaisala processed data were better than those obtained using ALCF data. This variation between the two ceilometer data sets can be partly explained by the low cloud-base heights issues with the ALCF data. The results obtained comparing the 3 by 3 satellite pixel grid around the ceilometer station with the ceilometer data were better for PoD than the ones calculated for the pixel whose center was closest to the station; however, the skill scores indicate that the individual pixel did better. The nighttime results were better than those at any other time-of-day and the whole day, although it should be noted that the number of cases for each time-of-day varied, with nighttime having more cases, and the ceilometers are less affected at night by atmospheric aerosols and precipitation. In terms of resampling intervals, there were cases where the 5 min interval gave better results than the larger intervals and cases were the opposite occurred, as well as inconsistent patterns with each variable, which makes it hard to determine which interval was best. For example, for the most part PoFD was higher for the 15 min interval, however, there were cases where PoFD had no change between intervals or was slightly higher for the 5 min interval. Regardless, for the majority of the cases, the difference in the results between the three sampling intervals varied from 0.0 to 0.05, with the exception of the FAR, which tended to vary by almost 0.1. Finally, throughout the year, the best results were obtained in the summer months, between May and October, coinciding with the rainy season of Mexico.

In summary, when only considering Vaisala data that had more consistency and reliability, the cloud mask had an accuracy between 69% - 85%, a probability of detection of 0.81 - 0.95, a probability of false detection of 0.37 - 0.40, and a skill score of around 0.39 - 0.62.These overall results are consistent with the results obtained with different instruments, in particular, by Jimenez (2020), who used CALIPSO retrievals to assess the Clear Sky mask of the GOES-16 satellite and found accuracy levels of 86% overall. With these results and the >99% precision in detecting clouds when precipitation is occurring, it can be concluded that the *GOES-16 cloud mask has a good overall cloud detection accuracy.*

Chapter 4

Investigation Part II: Characterization of rainfall using GOES-16 satellite products

One of the main objectives of this study was to characterize the precipitation that occurs in the Central Mexico region using GOES-16 satellite data, and now that one of the fundamental Level 2 products of the GOES-16 satellite has been validated for Mexico, it can be safely assumed that the rest of the Level 2 products that use the cloud mask as a base are representative of the atmospheric conditions in Mexico. The precipitation characterization analysis involved evaluating the rain events with the reflectance values given by the GOES-16 satellite, in order to see if there are any patterns in the satellite data that occur before the development of rainfall. This characterization was done using Queretaro weather radar data for the precipitation data.

4.1 GOES-16 satellite data

The GOES-16 satellite has various Level 1 and Level 2 products. However, several Level 2 products were not available during rainfall events for the period of analysis. For example, because the dominating precipitation regime in the region during the rainy season is a convective precipitation regime, one potentially helpful product available is the "Derived Stability Indexes". These include indexes such as the Convective Available Potential Energy that give an insight into the stability of the atmosphere and can be used as indicators of the possibility of convective development. However, as mentioned before, these were not available before or during rainfall events. Figure 4.1 shows an example of the GOES-16 Convective Available Potential Energy values with a weather radar echo during a rainfall event, clearly showing the disappearance of the satellite values over areas of precipitation measurements.

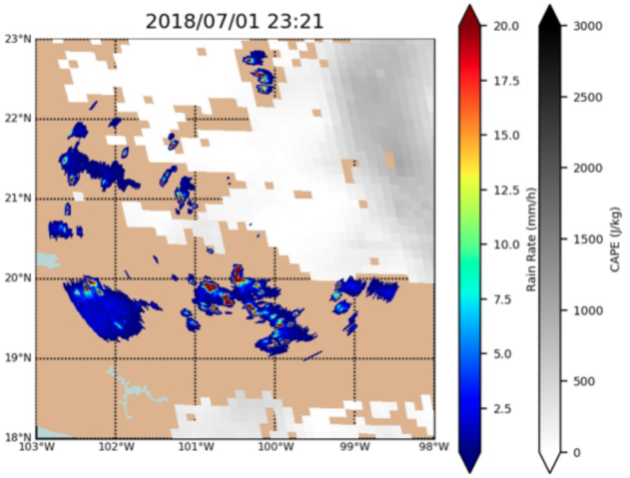


Figure 4.1: Example image of GOES-16 "Convective Available Potential Energy (CAPE)" values (gray-color scale) with Queretaro radar rain rate (jet-color scale) for the event occurring July 1st, 2018 at 23:21Z.

For this reason, most of the Level 2 products were not useful for this rainfall analysis. Only two Level 2 products were found to give information regardless of the presence of rainfall during the period of analysis: "cloud-top height" and "cloud-top phase". The GOES-16 cloud-top height product or ACH is a level 2 product that measures the cloud top height of each pixel marked as cloud. It is derived using the ABI Cloud Height algorithm with three of the 16 spectral channels (14, 15 and 16 with wavelength of 11.2 μm, 12.3 μm, 13.3 μm respectively) and uses the satellite's cloud mask as a base (Heidinger, 2012). The product has a 10 km by 10 km grid in the Continental US GOES-16 mode and can provide data during the day and night.

The GOES-16 ACTP determines the cloud-top phase of a cloud pixel. The product algorithm uses the cloud mask and the ABI channels 9, 10, 14, 15 to determine if a pixel is cloud or not and determines which of the four cloud phases best describes the pixel (Pavolonis, 2010). The result the product gives is a number between 0 to 5; 0 being clear sky and 1 through 4 for the cloud phase categories, which are: warm liquid water phase (1), supercooled liquid water phase (2), mixed phase (ice and liquid water) (3), and ice phase (4). The results marked with 5 are of unknown cloud phase. This product has the same spatial resolution (2 km by 2 km) and grid as the Continental US GOES-16 mode and is also available day and night.

The other products used in this analysis were the Level 1 ABI channel 13 (10.3 μm wavelength) and channel 15 products (12.3 μm wavelength). These two channels are the 'dirty' and 'clean' windows (respectively) used to calculate the SWD with equation 2.1 (Section 2.1.2) to indirectly measure low-level water vapor in the atmosphere (CIMSS, 2018a,b; Schmit et al., 2017).

4.2 Queretaro Radar

The Queretaro meteorological radar, operated by the Queretaro State Water Commission (CEA, 2022), is a Doppler, "C" band radar with pulse compression (O'Hora and Bech, 2005), located at 20.7802° Lat, $-100.5504°$ Lon, approximately 27 km northeast of the city of Santiago de Queretaro, Queretaro, Mexico. It has a range of 239.75 km and a beam width of 1°. The data from this radar was cleaned using the Gabella methodology (Gabella and Notarpietro, 2005) and processed with the "Wradlib" Python library (Heistermann et al., 2013). The rain rate was calculated using the Z-R relation described in equation 2.2 with the parameters $a = 220$, and $b = 1.6$. After performing a visual quality control of the radar data, a total of 5 days were removed due to faults, from July 2nd, 17:42Z to July 5th, 17:42Z, July 16th, 2:57Z to July 17th, 14:07Z, and August 16th starting at 2:02Z.

Figure 4.2: Picture of the Queretaro weather radar run by the CEA (source: CEA, 2022).

4.3 Methodology

The first step to obtaining a characterization of the precipitation in the region was to obtain a single common grid to work with both data sets. This radar-satellite common grid was also an essential component in later parts of the study that involved the use of nowcast models.

It was determined that the radar data was to be interpolated into the GOES-16 satellite grid because the satellite grid has the largest resolution of the both types of data sets and it is a uniform, rectangular grid (for the area in question). The radar data was interpolated using the *griddata* function with the nearest-neighbor interpolation method of the *scipy* library of Python. A visualization of the final radar-satellite common grid over the Queretaro radar coverage area can be observed in Figure 4.3 for the 2 km by 2 km GOES-16 Continental US grid and in Figure 4.4 for the 10 km by 10 km GOES-16 ACH grid. The radius of the grid is the radar radius/range plus the resolution of each grid (10 km or 2 km), guaranteeing that all of the radar range was completely covered with the grid.

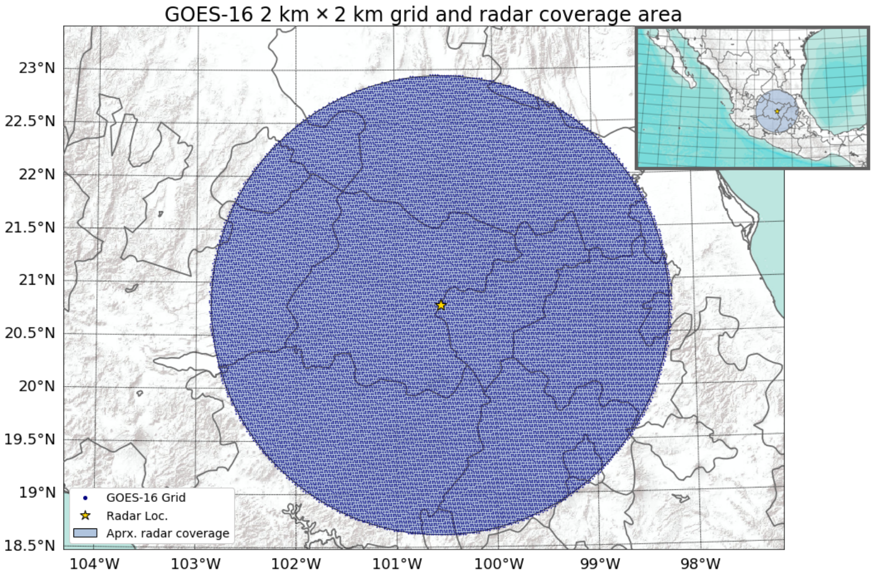


Figure 4.3: GOES-16 satellite 2 km by 2 km grid used as radar-satellite common grid over the Queretaro radar coverage area. Grid is marked by the pixel centers. Basemap sources: Esri, USGS, NOAA (ESRI, 2020).

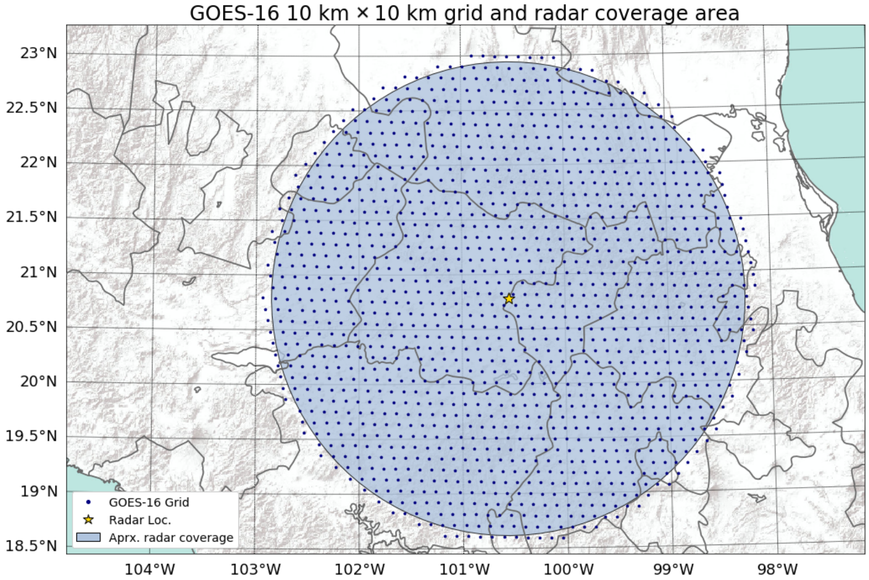


Figure 4.4: GOES-16 satellite 10 km by 10 km grid used as radar-satellite common grid over the Queretaro radar coverage area. Grid is marked by the pixel centers. Basemap sources: Esri, USGS, NOAA (ESRI, 2020).

The characterization analysis involved comparing the radar rain rate values interpolated into the radar-satellite common grid with the values of the cloud-top height, SWD and the cloud-top phase, for each individual pixel in the grid. Furthermore, it included observing how the values of each of the satellite products used changed in the time prior to the measurements of the radar rain rate, as well as comparing the values of the satellite products when there is rainfall with the values when there is no rainfall. Finally, the period of analysis was 47 days, from July 1st 2018 to August 16th, 2018, which falls within the rainy season of the region (which runs from late May to early October).

4.4 Results

4.4.1 Precipitation characterization using the satellite cloud-top height

Figure 4.5 shows an example of the ABI Cloud-base height (ACH in the figures) and the radar rain rate over the Queretaro radar coverage area, in their respective grids. From these and other cases not shown, it can be observed that the deeper clouds coincide with the occurrence of a high rain rate. Continuing, in order to obtain a more general comparison between the cloud-top height and the radar rain rate, a scatter plot was obtained depicting the values of each pixel (denoted a case) in the radar-satellite common grid for the whole period of study, and it can be found in Figure 4.6. The plot does not show a clear pattern between both variables, although a large cluster of data points was found between the cloud-top height of 10 km and 17.5 km, consistent with altitudes of cumulus and cumulonimbus clouds. Additionally, the majority of points with high rain rates occur within this interval of cloud-top height. The average value for the cloud-top height for all cases was 11.29 km.

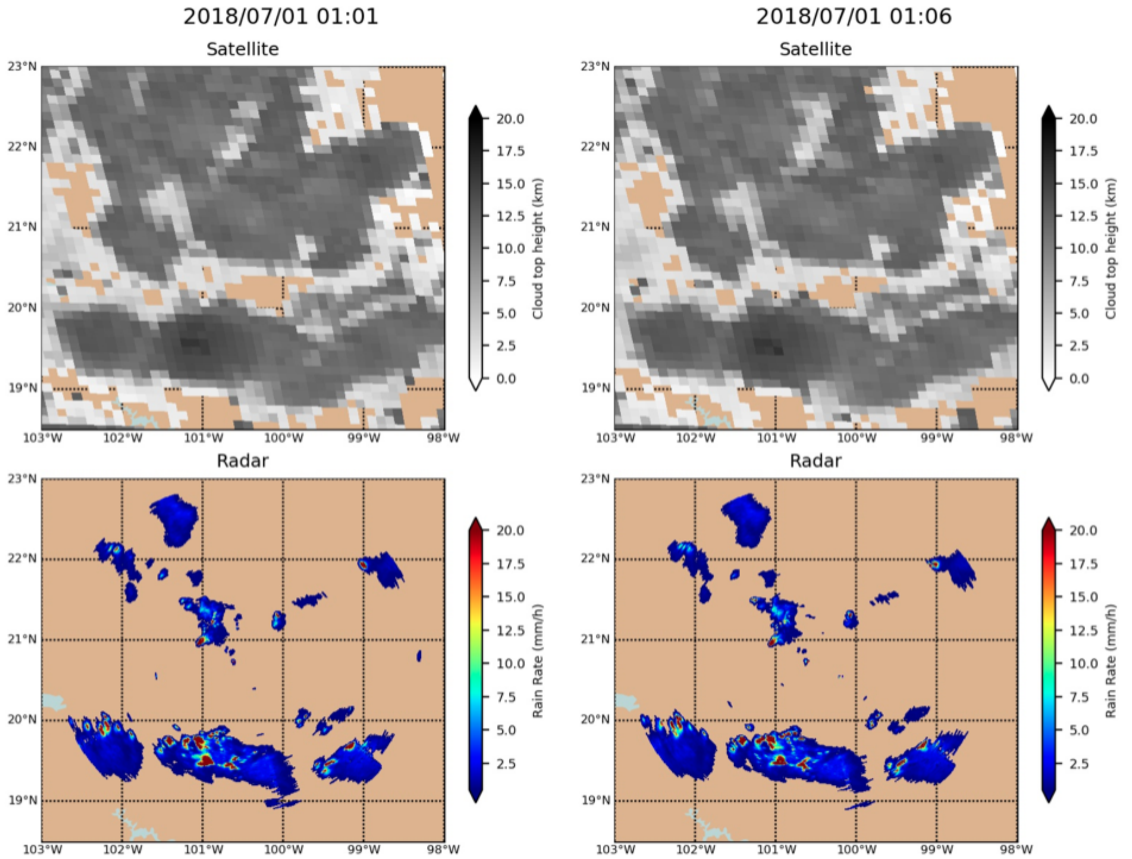


Figure 4.5: Example images of the cloud-top height values (top) and the Queretaro radar rain rate (bottom) for the event occurring July 1st, 2018 at 01:01Z (left) and July 1st, 2018 at 01:06Z (right).

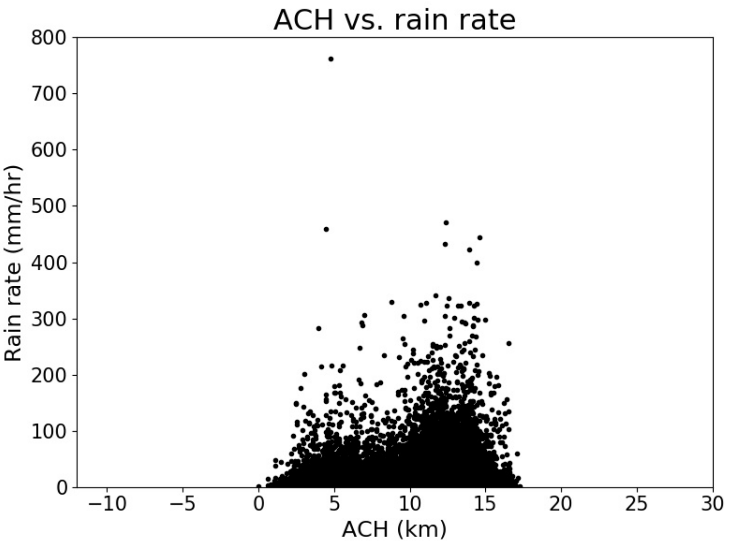


Figure 4.6: Scatter plot of the values of rain rate and cloud-top height (ACH) that occur in each pixel in the radar-satellite common grid over the study area during the entire period of analysis.

These results did not give a complete insight into the possible relationship between these two variables. Thus, to expand this analysis, Figure 4.7 shows the average cloud-top height for each of the offset times for all cases and for cases with rain rates that fell within certain intervals. Meaning, the values of the rain rate were compared with the values of the cloud-top height of a given amount of minutes (offset time) prior to the rain rate values. The offset times were from 5 to 90 minutes in intervals of 5 minutes and from 120 to 240 minutes in intervals of 30 minutes. This graph uses negative offset times to indicate the time prior to precipitation. Considering Figure 4.7, it can be observe that the average cloud-top height increased with increasing rain rate values. Additionally, the average cloud-top height decreased with increasing offset time for all intervals. In other words, the cloud-top height on average increased in the time prior to the development of rainfall. The biggest increases occurred for the highest rain rate values, with cases with rain rate of 200 mm/hr or higher rising 2.6 km from the average value observed 90 minutes prior to rainfall. On average, the average cloud-top height increased 2.5 km from the value observed 240-minute before rainfall occurs and 0.5 km from the average observed 90-minute prior to the development of rain. In addition, Figure 4.8 shows the change in the average cloud-top height per 5 minutes for each offset interval. For rain rates above 5 mm/hr, the sharpest increases occurred between the offset times of 85-minute and 45-minute, indicating that these are the times when, on average, the cloud-top height grows the most prior to the development of this type of rainfall. Moreover, for the highest cloud-top height, the values in Figures 4.7 and 4.8 were noisier than for smaller rain rates. This is due to the fact that the statistical arithmetic average (or mean) is not robust or resistant and is affected by the sample size used. For this analysis, the amount of cases decreased significantly with higher rain rates, with the distribution of cases per rain rate interval in percentage presented in Figure 4.9 and the total cases analyzed being 593429. From these percentages, rain rates higher than 100 mm/hr have less than 0.11% of the cases, or approximately 700 cases, and all other intervals have more than 2,200 cases. Moreover, the average was further affected by the fact that the high rain rates of 200 mm/hr or more had a larger distribution of rain rate ($\geq$ 200 mm/hr) because they were divided into larger intervals, whereas lower rain rates were distributed into intervals less than 100 mm/hr in length. This fact and the smaller sample size caused a higher likelihood of outliers occurring in these intervals (something that is observed in the scatterplots of Figure **??**), affecting the averages obtained.

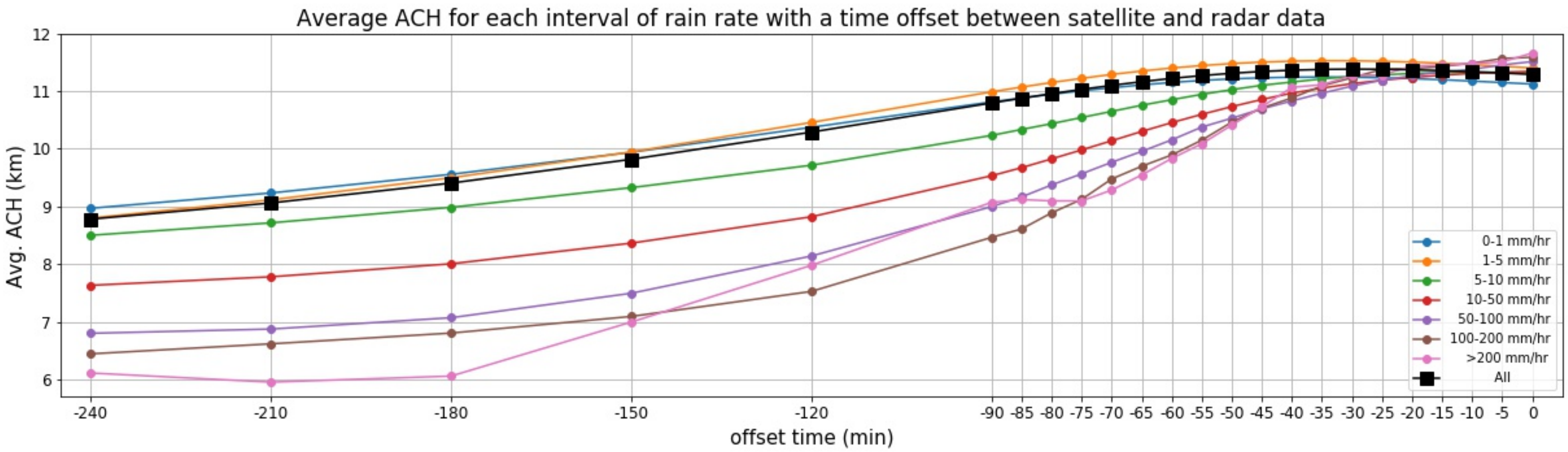


Figure 4.7: Average cloud-top height (ACH) for offset times between satellite and radar data for the various intervals of rain rate intensity indicated in the legend. Satellite data is ahead of radar data by the offset time indicated. Negative offset time values indicate time prior to development of rainfall.

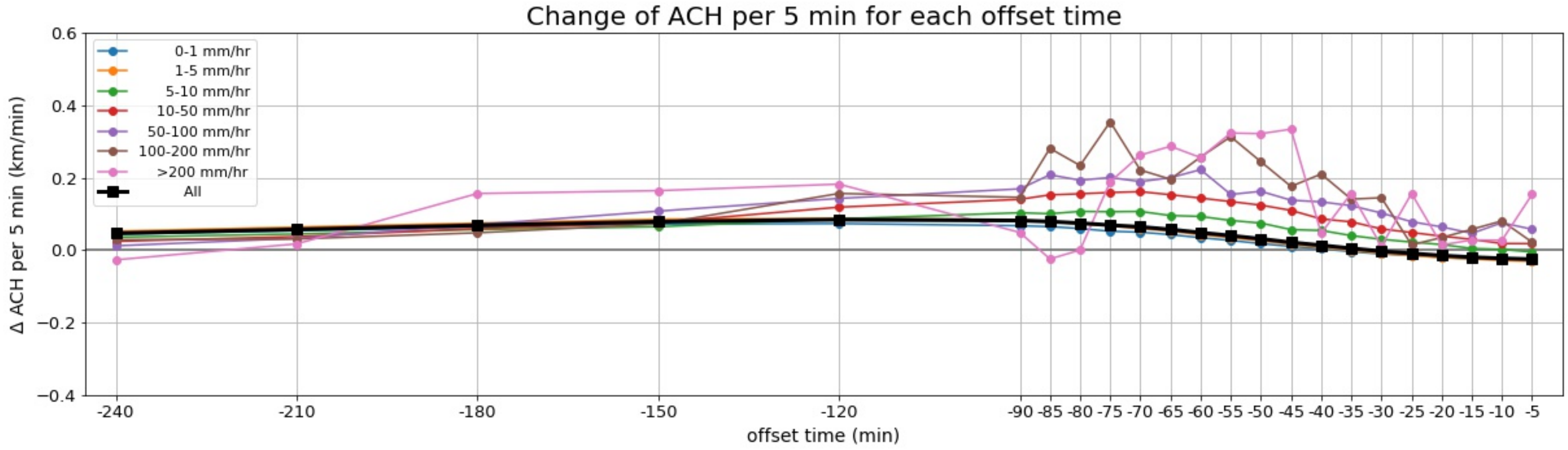


Figure 4.8: Change in average cloud-top height per 5 minutes for the same offset times and rain rate intervals of Figure 4.7.

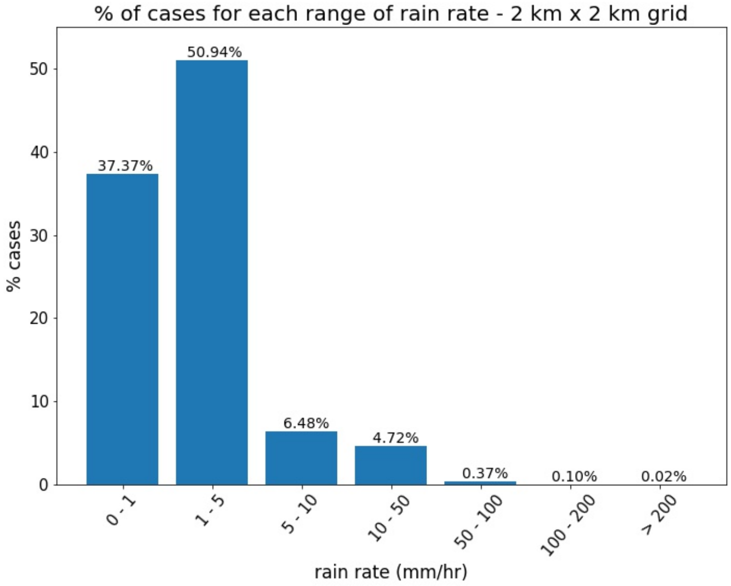
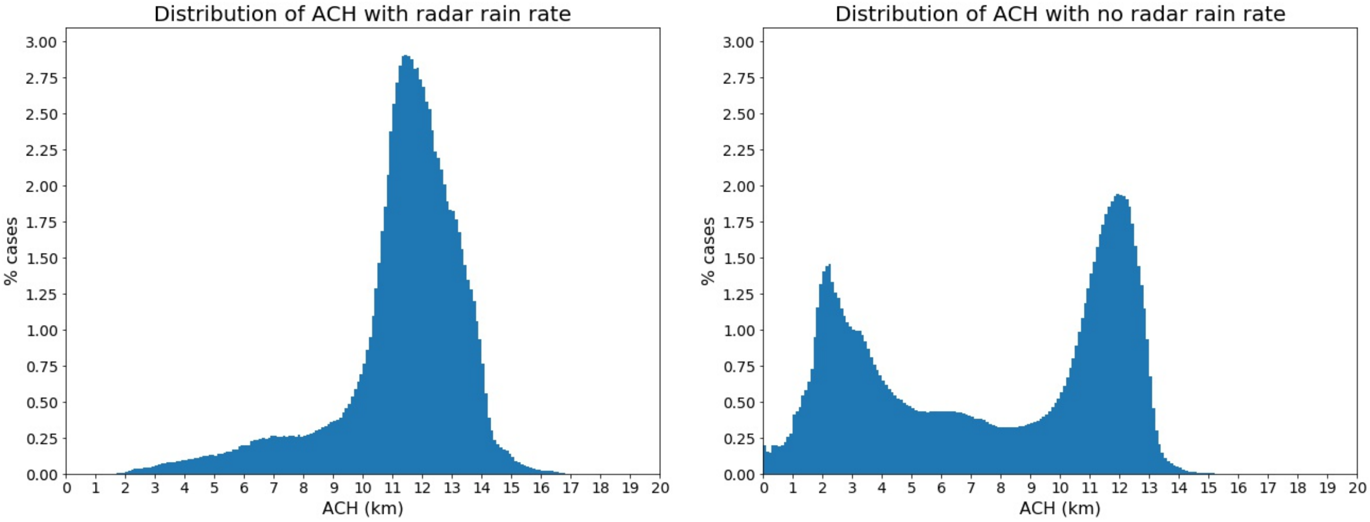


Figure 4.9: Distribution of the percentage of cloud-top height cases that fall in each rain rate interval indicated.

Finally, Figure 4.10 is the distribution of cloud-top height for pixels with rain rate (a) and for pixels without rain rate (b). Figure 4.10a shows that for the vast majority of rain rate pixels, the cloud-top height fell between 10 km and 13.5 km, with the highest percentage being 2.9% occurring at 11.4 km. On the other hand, for non-rain rate pixels Figure 4.10b, the peak cloud-top height fell between 10.5 km to 13 km with a maximum of 1.9% at 11.9 km, and a second clear peak of cases occurring around 1.5 km - 3.5 km with a maxima of 1.45% at 2.2 km. The average for non-rain rate pixels was 7.66 km. This points to the fact that for around half the non-rain rate pixels, cloud-top height is lower than for rain-rate pixels. One final note on the comparison between the cloud-top height and non-rain rate pixels, the methodology used does not take into consideration the fact that some non-rain rate pixels can be part of a bigger system where there is only rainfall in certain parts and not in the whole system itself. Additionally, the cloud-top height does not give any information on the depth of the cloud, which causes difficulty in distinguishing hight-altitude shallow clouds, and deep convective clouds. Considering this can explain, in part, the large amount of non-rain rate pixel cases with high cloud-top height.

Overall, the results suggest that for around one third of the time, cloud-top height tends to be around 8 km - 9 km lower when there is no precipitation present. Furthermore, the average cloud-top height tends to grow by 0.5 km at 90 minutes prior to the occurrence of rain, with the fastest growth occurring between 85 to 45 minutes prior to the development of rainfall for rain rates of 10 mm/hr or higher, suggesting a growth in the depth of the clouds prior to the development of rainfall, consistent with the behavior of the growth of cumulus clouds prior to convective precipitation. However, the characterization of precipitation using only the cloud-top height product is very limited due to the fact some factors cannot be determined with this variable alone (such as type of cloud).



(a) (b)

Figure 4.10: Distribution of cloud-top height cases (a) with radar rain rate and (b) without radar rain rate.

4.4.2 Precipitation characterization using the satellite split-window difference

When comparing the pixel-by-pixel results of the radar rain rate and the GOES-16 Split-Window Difference (SWD), Figure 4.11 shows a relationship between both variables, with the majority of rain rate cases having a positive SWD, with extreme rain rates (of 100 mm/hr or above) showed an SWD between 0 K to 5 K, with an average of 1.99 K.

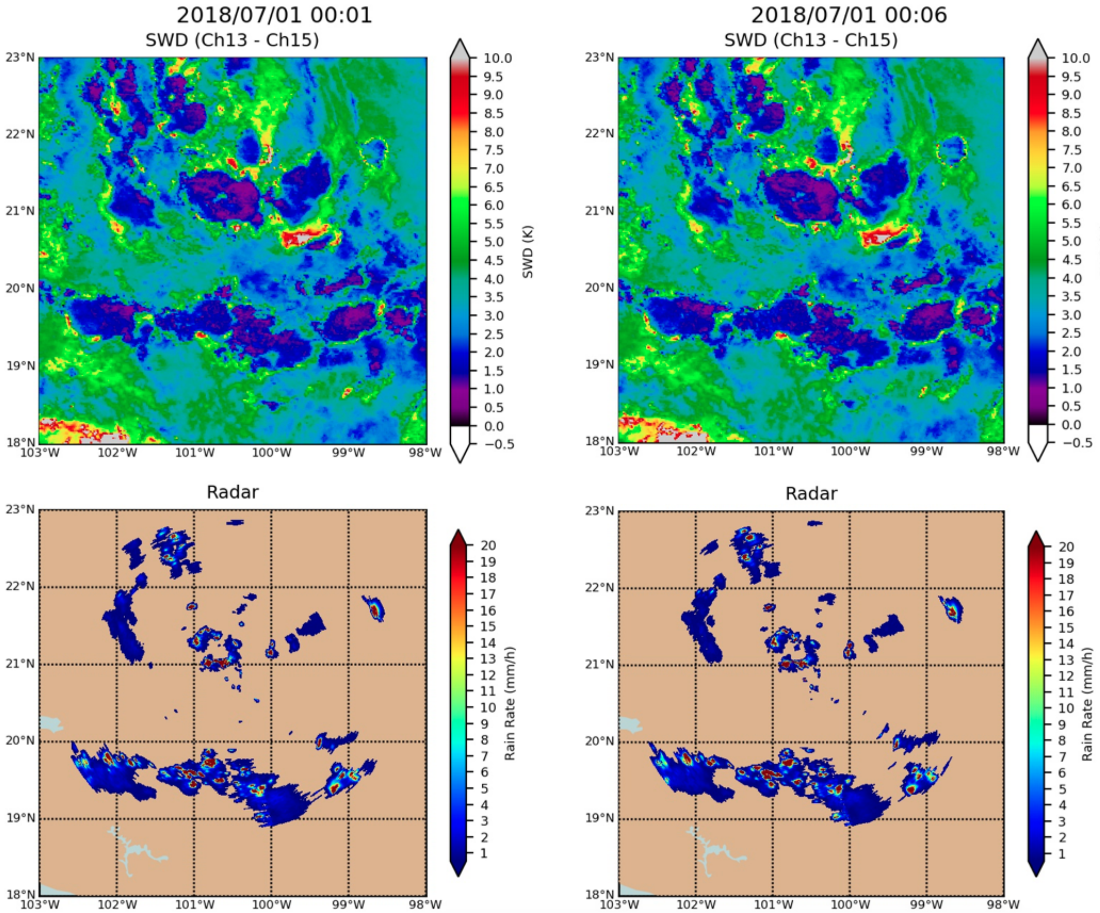


Figure 4.11: Example images of the SWD values (top) and the Queretaro radar rain rate (bottom) for the event occurring July 1st, 2018 at 00:01Z (left) and July 1st, 2018 at 00:06Z (right).

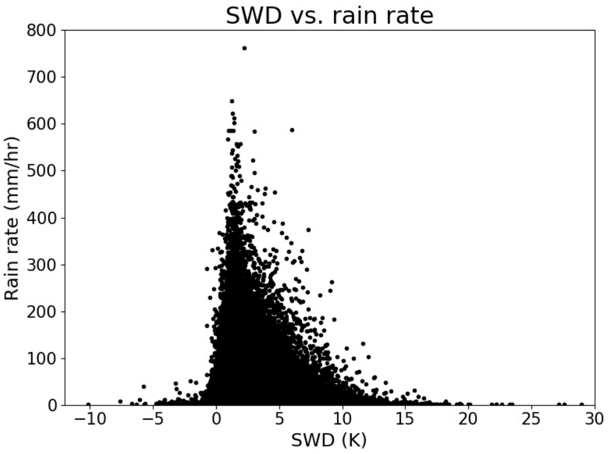


Figure 4.12: Scatter plot of the values of rain rate and SWD that occur in each pixel in the radar-satellite common grid over the study area during the period of analysis.

Continuing, an analysis was conducted with different offset times (comparing radar data with satellite data that occurred at the defined offset time prior) to obtain information regarding how the SWD behaves during convective development. Figure 4.13 shows the average SWD at different offset times for different intervals of rain rate. For all cases, the average SWD increased with offset time, or in other words, the average SWD decreased when getting closer in time to the rainfall measurement. The average change between the 0-minute and 240-minute offset times was 1.89 K; and between the 0-minute and 90-minute offset times was 0.81 K. This change increased with increasing rain rate values, with the highest change observed for the rain rate interval of > 200 mm/hr, equal to 3.13 K for the 0-minute to 240-minute offset times, and 2.53 K for the 0-minute to 90-minute offset times. Furthermore, this graph also shows that the average SWD decreased with increasing rain rate for offset times of 20-minute or less, and the opposite was true for offset times of 25-minute or higher. This decrease in SWD preceding rain events was also observed by Grasso et al. (2020), where a storm case study analyzed the SWD derived from the GOES-16 data, revealing that the SWD dropped below 3 degrees Celsius just before and during the onset of rainfall. Figure 4.14 shows the change in average SWD per 5-minutes for all offset times. In general, the change was negative, indicating that the SWD decreased before the development of rainfall. The biggest decrease occurred for high intensity precipitation between the 5-minute offset time and the 90-minute offset time. The values also formed a valley between the 75-minute and 25-minute offset times, with the maximum change occurring between the 50-minute and 45-minute offset times. The average change that occurred during this valley was of 0.6 K per 5-minutes for all cases. Lastly, the percentage of cases in each interval of rain rate used for Figure 4.13 is similar to those of Figure 4.7 with a variation of less than 1% for the total cases of 1,4900,868. All intervals had more than 2100 cases for the analysis.

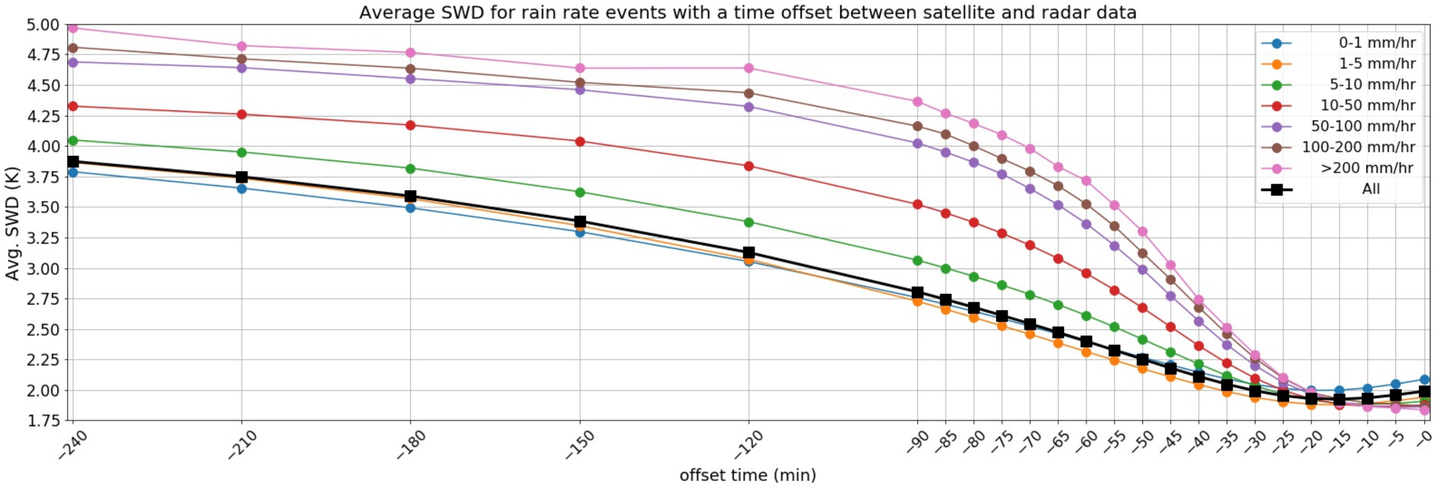


Figure 4.13: Average SWD for offset times between satellite and radar data for the various intervals of rain rate indicated in the legend. Satellite data is ahead of radar data by the offset time indicated. Negative offset time values indicate time prior to development of rainfall.

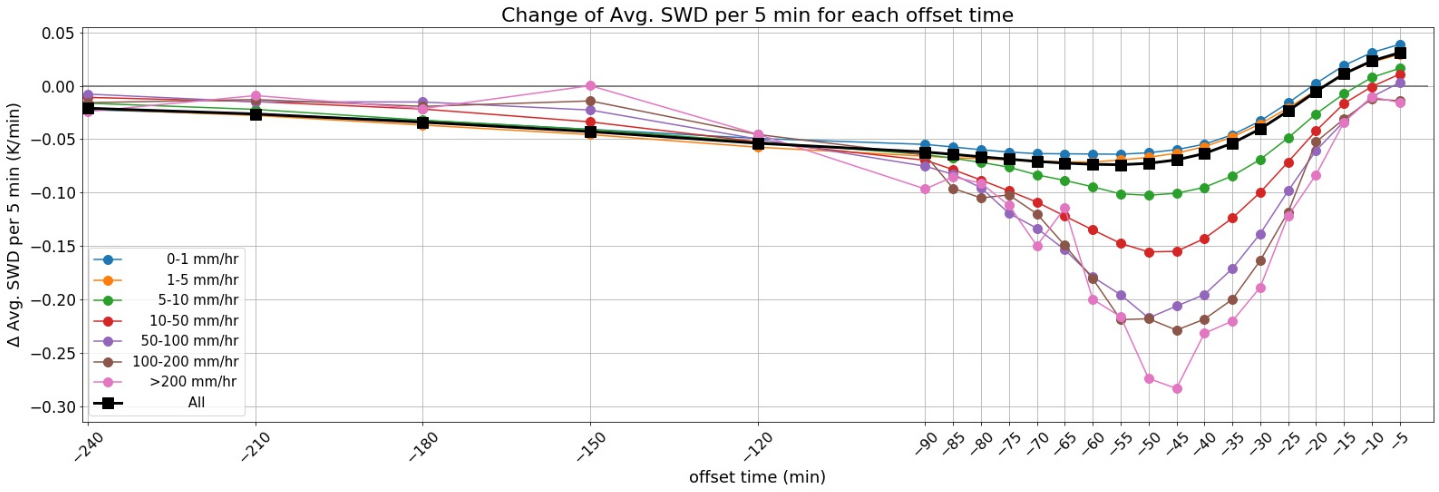
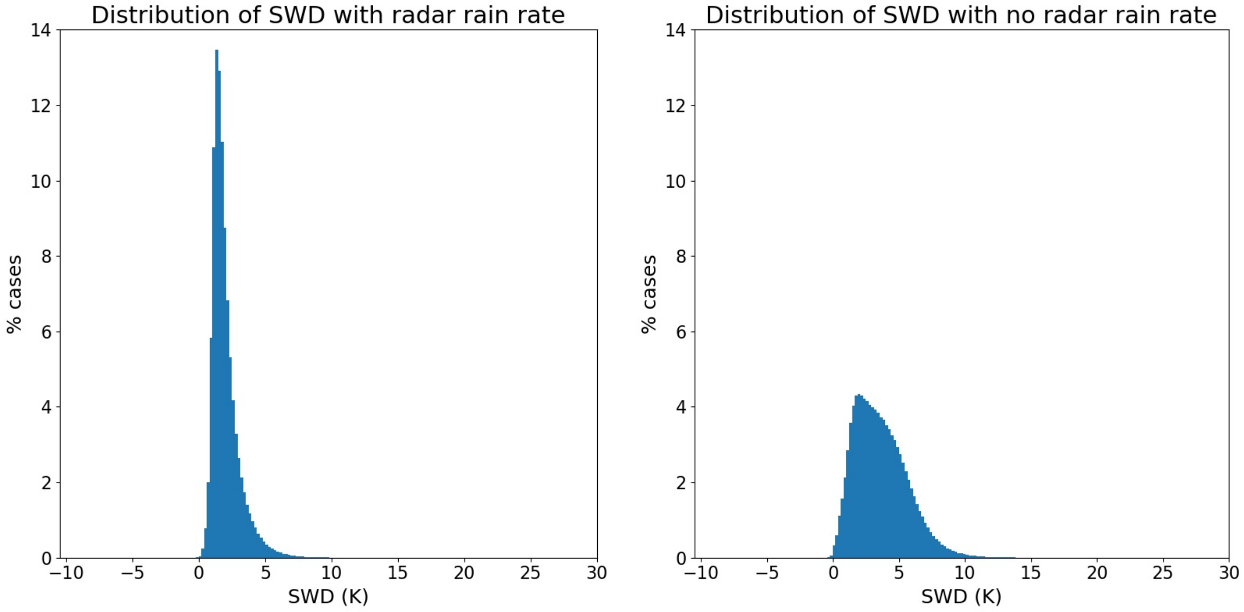


Figure 4.14: Change in average SWD per 5 minutes for the same offset times and rain rate intervals of Figure 4.13.

Finally, Figure 4.15 shows the distributions of the SWD for pixels with rain rate (a) and pixels without rain rate (b). Figure 4.15a supports the observations made with Figure 4.18 that most cases had an SWD between 0 K - 5 K. The interval with the highest percentage of cases, 13.5%, was that of 1.23 K - 1.44 K. For non-rain rate cases, Figure 4.15b shows the majority of cases fell between 0 - 10K, with an increase in the number of cases between 5 K - 10 K, in comparison to rain rate cases. The higher percentage of cases, 4.1%, occurred for the SWD between 1.86 K - 2.07 K. The average for non-rain cases was 3.62 K. The main differences observed between the no rain rate cases distribution and the rain rate cases distribution are the number of cases above and below 2.5 K. For non-rate rate cases, 37.2% of cases were below 2.5 K and for rain rate cases there were 82.2%, indicating that SWD below 2.5 K tended to occur along with precipitation.



(a)

(b)

Figure 4.15: Distribution of SWD cases (a) with radar rain rate and (b) without radar rain rate.

Considering all the results in this section, the relationship between rain rate and SWD is an inverse one. The SWD tends to decrease with the presence of rain rate and intensity of rain rate. On average, the SWD tends to decrease the most between 75 minutes to 25 minutes prior to the onset of rain, with a drop of 0.6 K for all intensities of rain rate. This change in SWD prior to the measurement of rainfall is consistent with observations made regarding the convective precipitation regime. Adams et al. (2013) observed an increase in the vertical water vapor of the atmosphere in the hours prior to the occurrence of convective precipitation.

For this case, the increase of low-level water vapor marked by the decrease of SWD can be indicative of the increase in vertical water vapor associated with the development of convective precipitation.

4.4.3 Precipitation characterization using the satellite cloud-top phase

Figure 4.16 shows the ABI Cloud-top phase (ACTP) and the radar rain rate. From the images, it can be observed that the cloud-top phase that occurs with rain rate is phase 4 or ice phase. Because the results are a set of discrete data, a scatter plot would not be helpful when trying to obtain a general relationship between the two variables like the previous analysis with cloud-top height and SWD. Thus, a distribution was obtained (Figure 4.17a) of the cloud-top phase results for all rain rate pixels or cases (in the radar-satellite common grid). The Figure shows that 92.3% of cases were the ice phase (phase 4), 4.2% were of mixed phase (phase 3), 1.85% were supercooled water (phase 2) and 1.4% were warm water (phase 1). Only around 0.2% were clear sky pixels (phase 0) and there were less than 0.01% of unknown (phase 5). Figure 4.17b shows the same distribution of cloud-top phase but for cases with no rain rate. The differences between Figure 4.17a and 4.17b were clear, with only 42.3% of the cases being ice phase and almost 29.5% being clear sky pixels for Figure 4.17b. A sharp increase in the percentage of warm water phase cases was also observed, growing to 22.4% in Figure 4.17b.

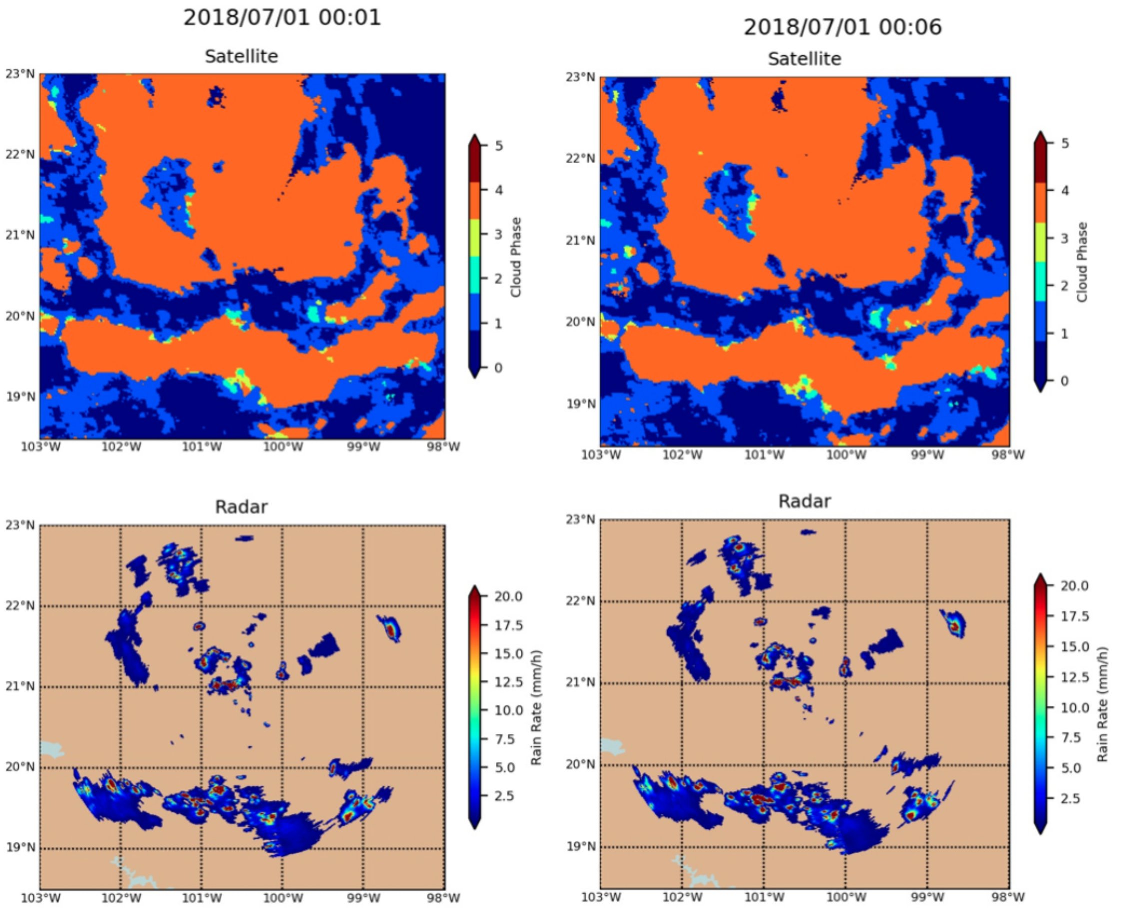
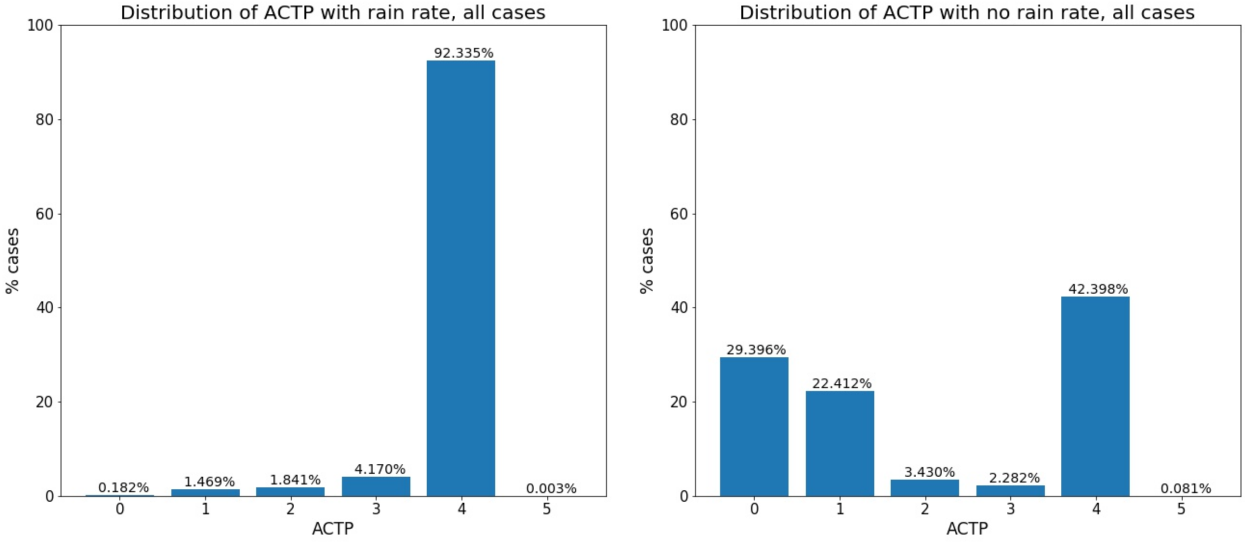


Figure 4.16: Example images of the cloud-top phase values (top) and the Queretaro radar rain rate (bottom) for the event occurring July 1st, 2018 at 00:01Z (left) and July 1st, 2018 at 00:06Z (right).



(a) (b)

Figure 4.17: Distributions of the cloud-top phase (ACTP) results for (a) cases with rain rate and (b) cases with no rain rate.

Similar to the previous two satellite products, an analysis was performed with different offset times. Figure 4.18 displays the plot of the percentage of the distributions for each cloud-top phase result for the various offset times analyzed. This graph shows that the percentage of ice phase results decreased with offset time. In other words, the amount of ice phase cases increased prior to the onset of rainfall. The increase was of around 32.0% from 240-minute to 0-minute offset time, and of 11.0% from 90-minute to 0-minute offset time. The other cloud-top phase result that had a significant change was the warm water phase. The graph shows that this particular phase decreased prior to the onset of rainfall by 24.2% in 240-minutes and by 9.0% in 90 minutes. Clear sky cases were higher with high offset time, but they were 6.2% or less of the total, and they decreased to less than 1.0% with 75-minute offset time or less. These results could suggest the possible growth of clouds prior to the onset of precipitation. The increase of ice phase cases and decrease of warm water phase cases can mean the deepening of clouds to heights where the temperature is low enough for water vapor condense and turn to ice. Additionally, the slight increase in the percentage of mixed phase cases up until 25 minutes prior to the development of rain and then decreasing, suggesting that there were cases that could have turned from warm to mixed phase to ice phase in the 240 minutes prior to the development of rain.

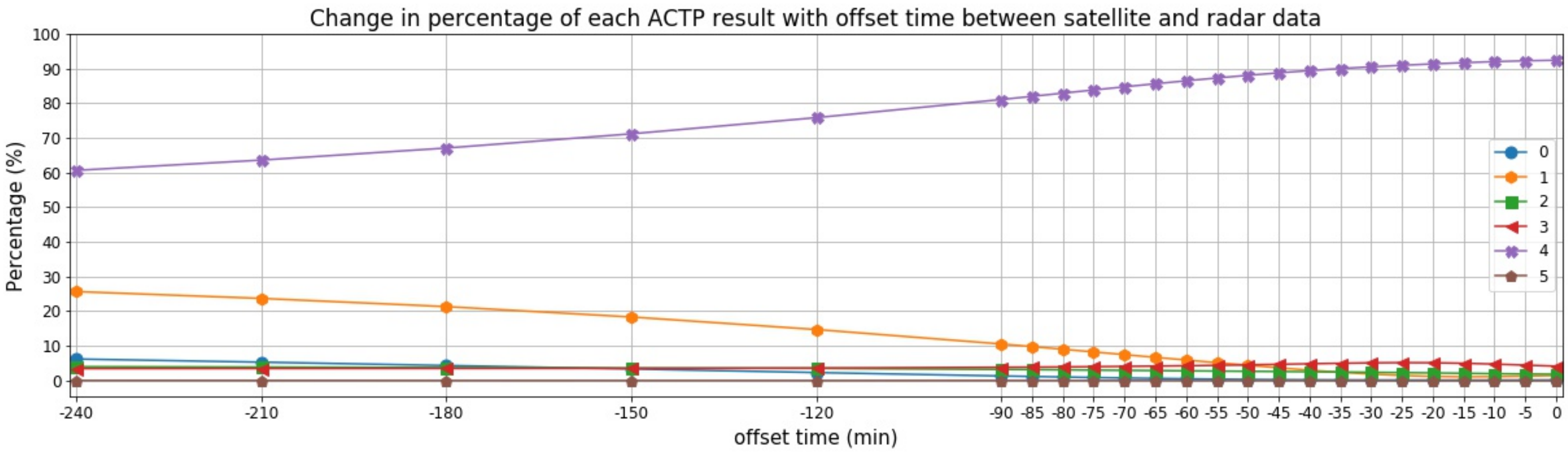


Figure 4.18: Percentage of each cloud-top phase result for each offset time indicated. Satellite data is ahead of radar data by the offset time indicated. Negative offset time values indicate time prior to development of rainfall..

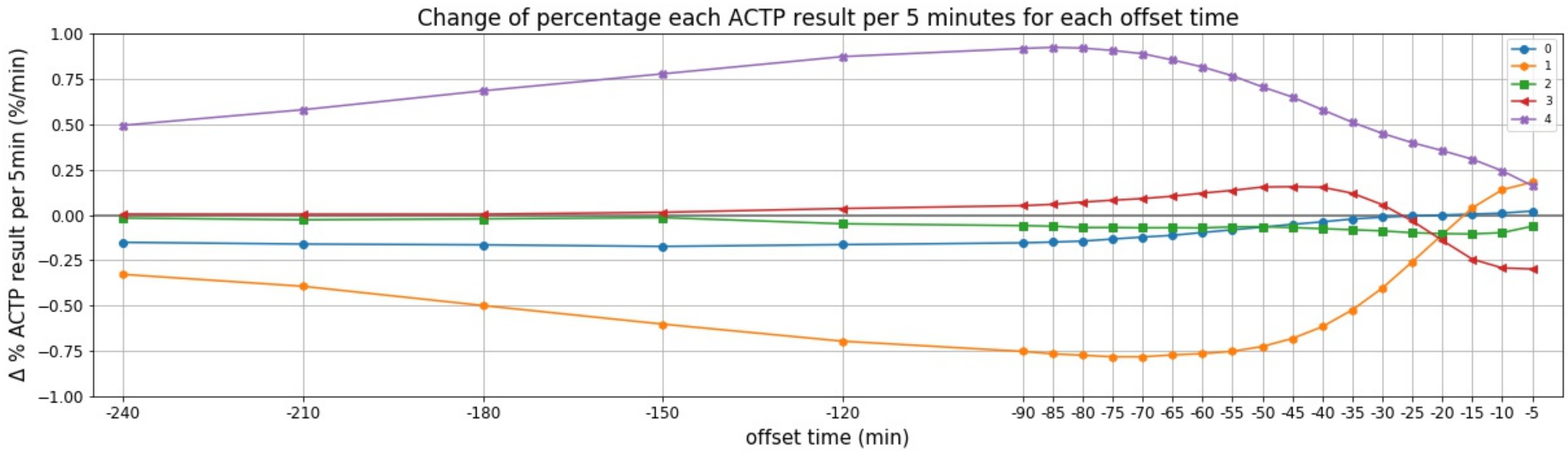


Figure 4.19: Change in percentage of each cloud-top phase result per 5 minutes for the same offset times as 4.18.

Figure 4.19 shows the change of the percentages of each cloud-top phase result per 5 minutes for the all offset times. This graph shows that the clear sky phase decreased until reaching 30-minute offset time, then the change from this point forward to smaller offset times was almost 0. For the warm water phase, it decreased throughout the the majority of the period of time, with the largest decrease occurring between 90-minute and 50-minute offset time. For offset times of 15-minute and less, the warm water phase increased. The supercooled water phase decreased throughout the whole period. However, the change was small especially for offset times higher than 120-minute. The mixed phase's change was also very small for offset times of 150-minute or higher. For offset times between 120-minute and 25-minute, the mixed phase increased with a peak occurring at 45-minute offset time. For offset times smaller than 25-minute, the mixed phase cases started to decrease. Finally, the percentage of ice phase results increased throughout the whole interval, with the largest increase occurring between 90-minute and 70-minute offset time. Similar to Figure 4.18, the changes showed in this graph, especially the decrease in warm water phase cases, the increase and decrease of mixed phase and the increase of ice phase, also indicate that within the 240 minutes prior to the development of rainfall, clouds tended to grow to heights were water freezes, transitioning from warm cloud phase to mixed phase to ice phase. Additionally, the magnitud of the changes observed in Figure 4.18 also varied with rain rate intensity (not shown). The magnitud of the change in the percentages of each phase increased with increasing rain rate intensity, but the general pattern observed was the same as Figure 4.18 for all intensities.

Table 4.1 shows the percentages of the various cloud-top phase phase results from the distributions similar to Figure 4.17a but for different intervals of rain rates. The overall pattern in each distribution was the same, with over 91% of the cases being in the ice phase. The interval with the highest percentage of the ice phase, 95% was the rain rates of 200 mm/hr or higher, and dropped with decreasing rain rate to around 91.4% until reaching the interval 1 mm/hr to 5 mm/hr, where the percentage increased to 93.0%. However, the cases decreased to 91.9% for lower rain rates. All other cloud-top phase results have the opposite behavior, where they increased with decreasing rain rate. The second highest cloud-top phase results was the mixed phase, with cases falling between 2.0% - 5.0% and the interval with the highest clear sky cases was for the lowest rain rate interval at 0.2%.

rain rate (mm/hr)	**Phase 0**	**Phase 1**	**Phase 2**	**Phase 3**	**Phase 4**	**Phase 5**
0 - 1	0.221%	1.468%	2.057%	4.396%	91.855%	0.003%
1 - 5	0.162%	1.352%	1.699%	3.858%	92.926%	0.003%
5 - 10	0.164%	1.910%	1.817%	4.735%	91.371%	0.003%
10 - 50	0.128%	2.089%	1.749%	4.979%	91.053%	0.002%
50 -100	0.058%	1.847%	1.390%	4.259%	92.441%	0.005%
100 - 200	0.075%	1.888%	1.155%	3.254%	93.622%	0.007%
>200	0.045%	1.920%	0.848%	2.232%	94.955%	0.000%

Table 4.1: Percentage of cases of each cloud-top phase phase for the rain rate intervals indicated.

In summary, the overall results show that before the onset of rainfall, there was a change in phase for a large quantity of cases, around 29% from warm water phase to ice phase. These changes, along with the changes seen in the supercooled phase and mixed phase cases, indicate the possible growth of clouds to heights with temperatures low enough for water vapor to freeze, which can be representative of cumulus clouds growing and deepening prior to the occurrence of precipitation. Notwithstanding, like with cloud-top height, these results cannot distinguish between high-altitude cirrus clouds and deep convective clouds.

4.4.4 General Diurnal Cycle

Considering the previous results, it was clear the most helpful satellite variable of those analyzed with the radar rain rate was the Split-Window Difference (SWD), thus a general idea of the diurnal cycle was obtained using the SWD and the radar rain rate. This was done by calculating the average of both variables over the entire common grid (called "area average") for each image available in the analysis time frame and observing how it varied throughout the day. Moreover, the radar rain rate average ($\overline{RR}$) was calculated by replacing NaN values with 0 to have a consistent methodology and results with the SWD area averages ($\overline{SWD}$) values. This change did not affect the overall patterns of $\overline{RR}$, only decreased the magnitude of the values.

Figure 4.20 shows the running mean for all days available for the $\overline{SWD}$ and the $\overline{RR}$, as well as the hourly average of both. The mean values show that rainfall typically commenced after 11 hr local time, with a maximum around 17 hr local time. On the other hand, the $\overline{SWD}$ tended to increase at around 6 hr local time and its maximum occurs at 12 hr local time. The behavior of $\overline{RR}$ is consistent with the observed diurnal cycle of the convective precipitation in various regions during the rainy season (Jauregui and Romales, 1996; Magana et al., 2003; Ochoa et al., 2015). The graphs also highlight the inverse relationship between $\overline{SWD}$ and the presence of precipitation, with a noticeable decrease in $\overline{SWD}$ coinciding with an increase in $\overline{RR}$. Additionally, the rise in SWD observed around 6 hr local time corresponds to the sunrise, indicating a decrease in low-level water vapor levels with the presence of sunlight. It is important to emphasized that both $\overline{SWD}$ and $\overline{RR}$ were averages over the entire region, considering all areas where there is no rain rate as 0.0 mm/hr. Consequently, the observed maximum and minimum $\overline{SWD}$ values exceeded those reported in Section 4.4.2.

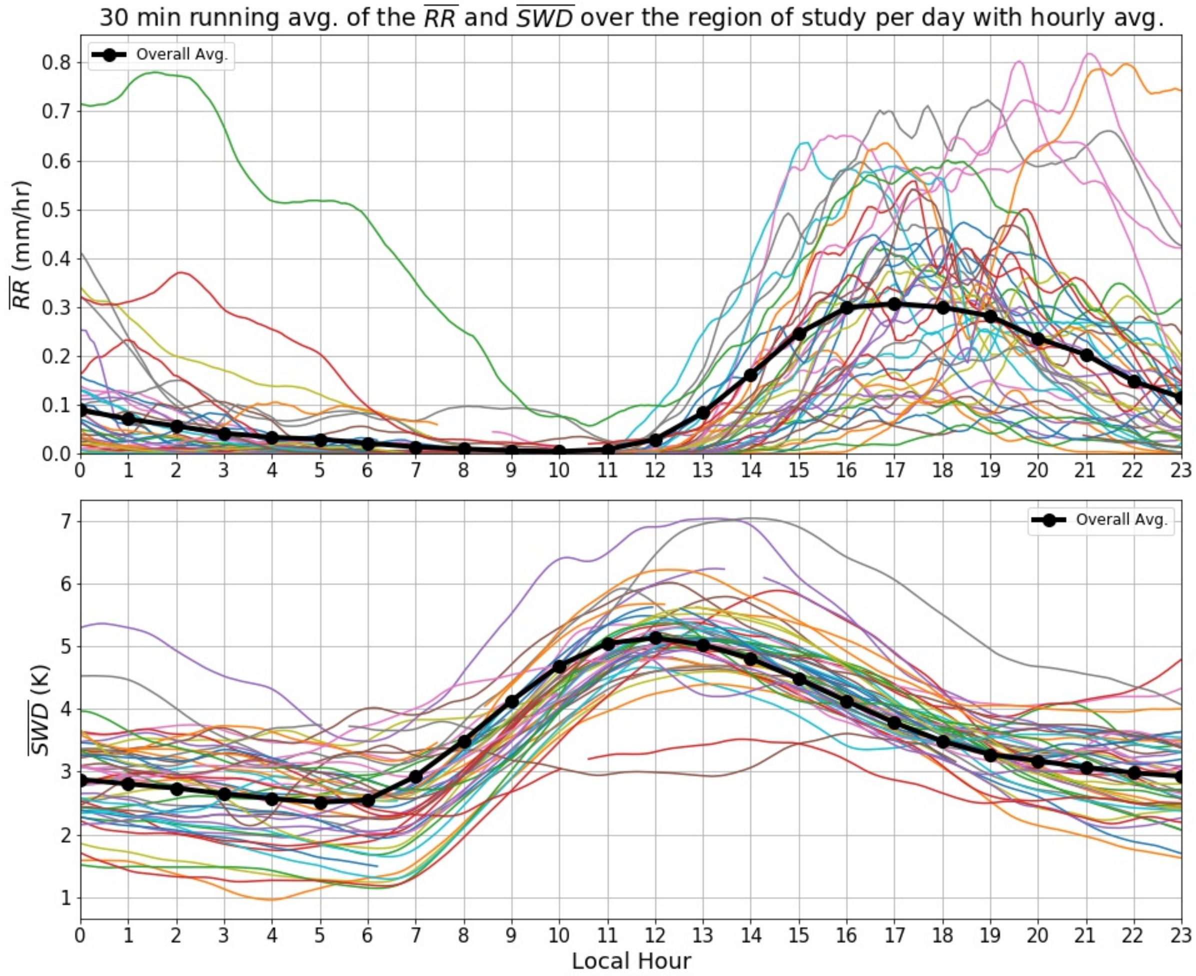


Figure 4.20: The 30-minute running mean of the area average of the radar rain rate ($\overline{RR}$) and SWD ($\overline{SWD}$) per day for all available days with the hourly average (black line).

4.5 Conclusion

In conclusion, an analysis of three out of the four variables used to characterize the radar rain rate in central Mexico revealed noteworthy relationships, summarized in Table 4.2.

In regards to the cloud-top height and cloud-top phase (cloud-top height and cloud-top phase respectively), the changes observed in these two satellite products give a small insight into the type of precipitation that occurs in this region by indicating cloud growth. During the rainy season, which includes the period of analysis for this study, convective precipitation is the most common type of precipitation in the region. The changes in the average cloud-top height prior to the measurement suggest that, on average, clouds tend to grow prior to rainfall, characteristic of convective precipitation. The variations in the average percentage of the various cloud phases given by the cloud-top phase prior to precipitation also suggest a growth of clouds, from shallow "warm-water" cumulus clouds to deep "ice-phase" cumulonimbus clouds. However, the cloud-top height and cloud-top phase were unable to discern between high-altitude clouds and the cumulonimbus clouds. Consequently, these variables proved less effective for identifying potential convective activity areas.

Conversely, while the Split-Window Difference (SWD) displayed a weak relationship with radar rain rate, it exhibited a substantial alteration preceding rainfall, dipping to approximately 2 K when rainfall was present and falling below 2.5 K 60 minutes before rain rate measurements, indicating that an increase in low-level water vapor occurs before the development of rainfall, characteristic of convective precipitation (Adams et al., 2013). These findings established 2.5 K as a threshold value signaling potential convective development, a metric utilized in evaluating the nowcast model.

Variable	**Avg. w/rain rate**	**Avg. w/o rain rate**
ACH	11.29 km	7.66 km
SWD	1.99 K	3.62 K
ACTP	N/A	N/A

Table 4.2: Summary of results of the characterization of the precipitation using radar rain rate and the GOES-16 satellite products.

Chapter 5

Investigation Part III: Evaluation of the *pySTEPS* nowcast models with GOES-16 and Queretaro radar data

The primary aim of this study was to assess the applicability of nowcast systems in the central Mexico region and determine the feasibility of implementing such forecasting methodologies leveraging the available data in the area. The key source of nowcast models for this investigation was the pySTEPS Python library. As outlined in Chapter 2.4, pySTEPS encompasses various nowcast models and a range of tools to validate these forecasts. This chapter entails the methodology description and the outcomes derived from evaluating the pySTEPS nowcasts using data from GOES-16 satellite and the Queretaro radar.

5.1 Why the Split-Window Difference?

Chapter 4 showed that the satellite products used in the characterization of precipitation had a weak relationship with precipitation, especially because some products were not able to distinguish between different types of clouds, such as deep convective clouds and high altitude cirrus clouds. Although this weak relationship was also observed with the split-window difference or SWD for, a change in the SWD was observed between 90 and 60 minutes prior to precipitation occurring, indicative of the increase in low-level water vapor that occurs before the development of convective precipitation. In other words, the decrease of the SWD below 2.5 K indicates the likelihood of rainfall within 60 minutes for the region. Figure 5.1 shows the SWD field limited to values between 0 and 2.5 K and radar rain rate of one of the events analyzed, 8:07 Z, 09/08/2018. This Figure highlights how the SWD values can indicate potential areas of convective development one hour before and after the rain event started, and in some cases 2 hours prior. However, this example also shows that not all SWD areas below 2.5 K developed into precipitation, suggesting that while the SWD can overestimate potential precipitation areas. Regardless, it remains a valuable tool for identifying potential convective development ahead of rainfall. These observations are consistent with findings from previous studies utilizing satellite data, including GOES-16 data (Lindsey et al., 2014, 2018; Grasso et al., 2020). The SWD data can be used as input in nowcast models to predict potential precipitation areas before the initial radar measurements of rainfall. Once the precipitation event commences, radar data becomes more reliable for predicting rain intensity, while the SWD can aid in indicating the possible direction of precipitation alongside radar data, though not its intensity due to the lack of a clear relationship between rain intensity and SWD. Ultimately, the combined use of SWD with nowcast models facilitates short-term predictions of potential precipitation areas before the first radar measurements, followed by the use of radar observations to forecast storm behavior.

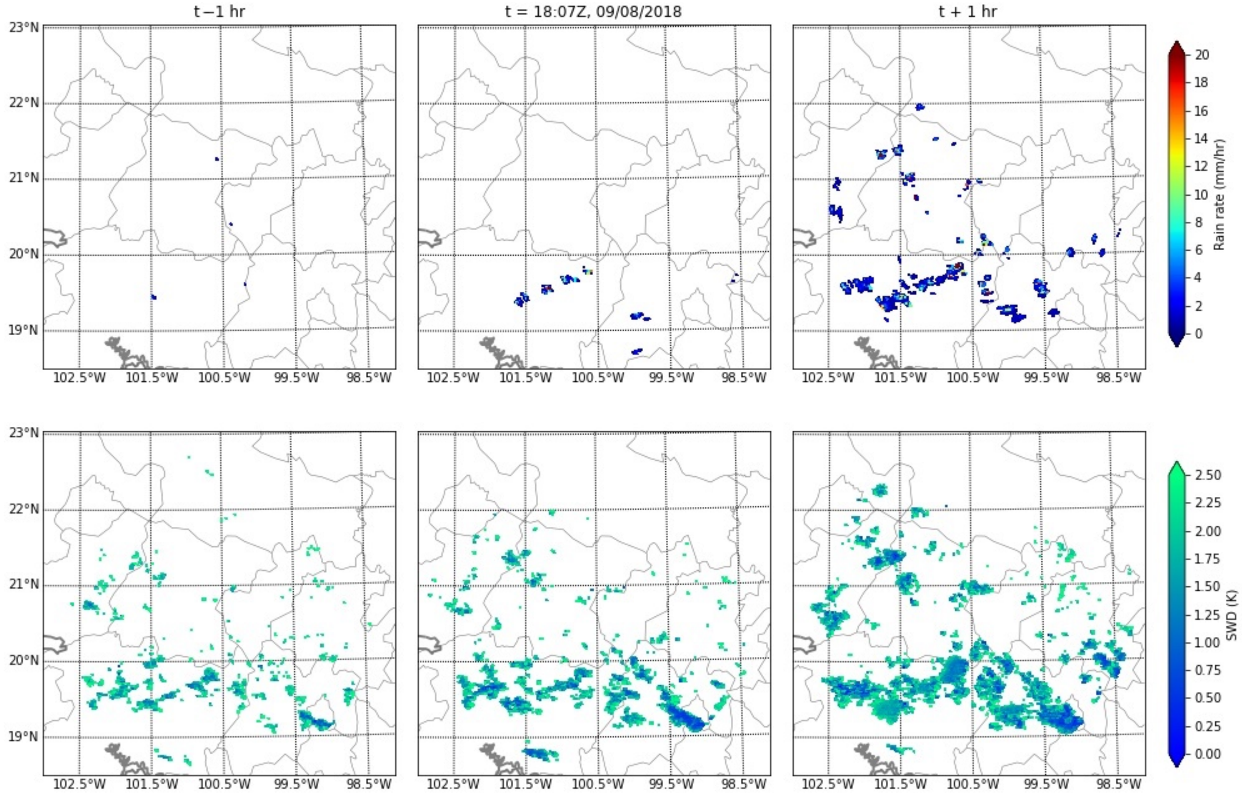


Figure 5.1: Rain rate echoes from the Queretaro radar (top images) and SWD field below 2.5 K from the GOES-16 satellite (bottom images) for the event starting at 18:07Z, 09/08/2018. Left images are for one hour before start of event, middle images are of start of event and right images are of 1 hour after start of event.

5.2 Data preparation

Two main data sets were used in this part of the study: Queretaro radar rain rate and the SWD calculated with the GOES-16 satellite data. However, pySTEPS is mainly designed to work with radar data. For this reason, before actually inputing any satellite data into the nowcast models of pySTEPS, it was necessary to prepare the data sets in order for pySTEPS to work correctly and efficiently with the data at hand. Furthermore, it was also important to verify that pySTEPS also worked properly with the Queretaro radar rain rate data. Before inputting the SWD values to be used by the nowcast models or advection functions in pySTEPS, the SWD field provided by the satellite had to be modified. It was observed that the algorithms developed by pySTEPS work best when the input data matrix is made up of NaN or missing values with clusters or groups of data points (like the data matrix provided by most radars). However, the satellite provides a data value in every single pixel of the grid. Thus, the SWD field had to be masked to those areas where development of rainfall is possible. The 2.5 K threshold SWD value obtained in Section 4.4.2 was used to filter the SWD field, by eliminating all values of SWD above 2.5 K and below 0 K (since there were very few cases with negative SWD values). Additionally, the values had to be multiplied by 100 to amplify the changes that occurred in the field. With the filtered SWD field, the nowcast models and the Lucas-Kanade flow field function were used and results were obtained without issue. From this point forward, any reference to "all SWD field" or "all of the SWD field" or "the SWD field" refers to this filtered SWD field, unless otherwise specified.

As for the radar data, the same processing methodology described in Section 4.2 was also done to obtain the rain rate for this part of the study. The radar rain rate data used in pySTEPS was the data interpolated into the radar-satellite common grid (see Section 4.3 and Figure 4.3) due to the fact that pySTEPS works with rectangular grids, such as the radar-satellite common grid, and cannot work properly with polar grids, such as the original radar data.

5.3 Events

Another obstacle to tackle before evaluating the nowcast models was determining the events used in the evaluation. The period of analysis was the same as Chapter 4, from July 1st 2018, to August 16th, 2018. This allowed the events to be obtained using the 30-minute running mean of the area average of the radar rain rate, presented in Figure 4.20 in Chapter 4. The start of each event was defined as the occurrence of a local minimum in the data after 10 hr local time and before 20 hr local time, a period where the graph shows the majority of the events tended to start. To obtain the events efficiently and computationally fast, the second derivative test[1] was used to locate the local minimum. After locating he local minimum, the start of the event was considered the time when the average precipitation reaches a minimum of 0.002 mm/hr after the local minimum. This was done to eliminate any possibility that the local minimum was caused by noise in the radar data and to mark the start of the event when moderate rain is starting to be detected and not when small

---

[1] the second derivative test tells where a function has a local minimum: where the first derivative is 0 (inflection point) and the second derivative is greater than 0.

isolated light rain events start. This methodology resulted in a total of 36 events, 23 for July and 13 for August. This methodology can be considered efficient because, assuming an event per day like Figure 4.20 suggests, the process picked 36 or 43 possible events, resulting in an 83% efficiency.

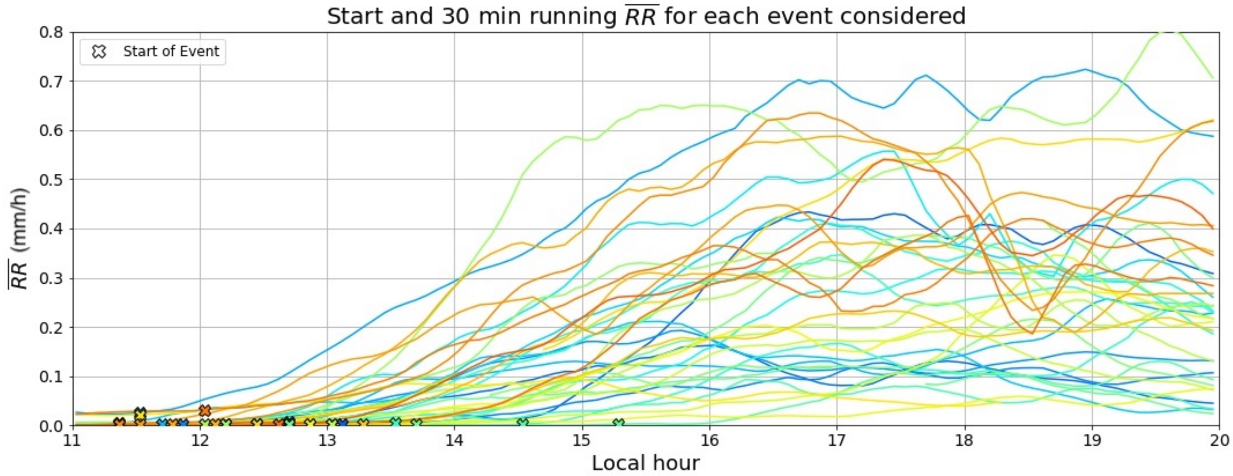


Figure 5.2: The 30-minute running mean of the area average of the radar rain rate ($\overline{RR}$) of the events considered for the nowcast evaluation with the start of each event.

5.4 Methodology

For the SWD nowcast assessment, the nowcasts used were those provided by the pySTEPS library of Python (see Section 2.4). The pySTEPS library includes several types of nowcasts, however, only three were used for this study: Extrapolation, S-PROG, and STEPS. Extrapolation is the simplest nowcast method available in pySTEPS, which is an extrapolation method through advection. In other words, it simply moves the given data through a specified motion field without modifying its magnitude or intensity. The Spectral-Prognosis or S-PROG nowcast, described by Seed (2003), is an extrapolation through advection nowcast method with an element of spectral decomposition in the spatial field. Finally, Short-Term Ensemble Prediction or STEPS is a nowcast system that not only includes the spatial scale spectral decomposition, but it also adds stochastic elements in the intensity and motion fields to quantify some uncertainty and randomness in the evolution of rainfall. These three nowcast methods were chosen partly because they do not use too much computational power. Additionally, because the nowcasts were going to be used with a satellite variable and not radar data, it was important to know how they would modify the intensity of the input variable. Extrapolation and S-PROG were chosen because they do not modify the intensity of the data, only the position and the spatial scales (in the case of S-PROG), while STEPS was chosen to assess how small modifications that were done when adding the uncertainty element in the model would affect the output when using other variables not associated with the radar.

Although all motion fields calculation functions available through pySTEPS were tested, the Lucas-Kanade method was used to obtain the motion fields for all nowcast methods because it used an adequate amount of computational power and produced a good representation of the motion field. The evaluation of these methods used the discrete verification statistics provided in the pySTEPS library, which include the Probability of Detection (PoD; Equation 3.4 in Section 3.5), False Alarm rate (FAR; Equation 3.2 in Section 3.5), false alarm ratio equivalent to PoFD (FA) (Equation 3.3 in Section 3.5), accuracy or fraction of correct results (ACC), bias (Equation 3.1 in Section 3.5), and the HSS (Equation 3.6 in Section 3.5). The Extrapolation and S-PROG models give one prediction per lead time, and STEPS gives an ensemble of predictions per lead time, each with different stochastic elements, with their size determined by the user. For this study, the STEPS model results were made up of an ensemble of 24 members, and the evaluation of the model was done with the average forecast of the ensemble in each lead time.

In addition to using pySTEPS's nowcast models, the DATing module described in Section 2.4.1 was also used to obtain SWD clusters that could lead to possible convective development. The DATing module was used in this study to try to eliminate individual pixels or groups of a few pixels of SWD data by picking out the clusters, assuming that these most likely lead to strong multi-cell convective storms. The following DATing tracking parameters were observed to yield the best results for the SWD data and were consequently used for this study: (units $\times$ K $\times$ 100)

- minimum threshold value = 100,
- maximum threshold value = 250,
- minimum difference between two maximums for object distinction = 60,
- minimum cluster size = 10 (pixels),
- minimum of maximum values = 240,
- minimum distance between clusters = 5 (pixels).

Finally, the evaluation was done for a period of time of 1 hour and 35 minutes before the start of a precipitation event and 4 hours and 55 minutes after the start of a precipitation event. This time frame was used in the DATing module to obtain the SWD clusters for each event. Every single image or data matrix available in this 6 hour and 35 minute period around the start of the event, for all precipitation events the occurred during the 48-day analysis, was used as input for the nowcasts. In addition, the nowcasts used a bigger common grid than the one used in Chapter 4 in order to evaluate how coverage area affected the results. This area can be observed in Figure 5.3.

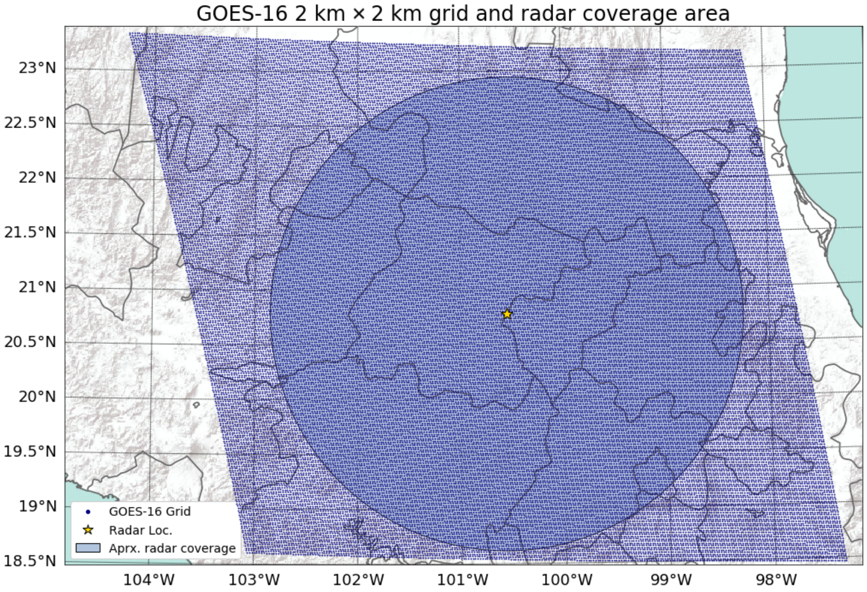


Figure 5.3: GOES-16 satellite 2 km by 2 km grid used as radar-satellite common grid for the nowcast models evaluation over the Queretaro radar coverage area. Grid is marked by the pixel centers. Basemap sources: Esri, USGS, NOAA (ESRI, 2020).

5.5 Results

The results of the evaluation statistics of the nowcast models are presented as a comparison between those obtained using all the SWD fields and those obtained using the SWD clusters. To maintain consistency, the same comparison was done for the radar rain rate data, comparing the results of all the rain rate fields and the results of the rain rates inside the SWD clusters. Finally, the evaluating statistics were obtained comparing the 15-minute forecast results with the actual observation unless otherwise specified.

5.5.1 Evaluation of nowcast models using the SWD data

Starting with the evaluation of the nowcast models using the SWD field data, Figures 5.4 and 5.5 show the results of the evaluating statistics obtained for all three nowcast models, in each 5 minute interval within the period of analysis of every event, using all the SWD field (Figure 5.4) and using only the SWD clusters (Figure 5.5). General observations were that for Extrapolation and S-PROG, all statistics were generally higher than for STEPS. And a large majority of events have consistent behavior throughout the period of analysis, especially when using the SWD clusters. Additionally, the general patterns observed differ between using all the SWD field and using the SWD clusters: all the SWD field tended to have little variability throughout the period of analysis, with some events increasing in accuracy, while using the SWD clusters had low accuracy at the beginning but quickly improved after 20 minutes.

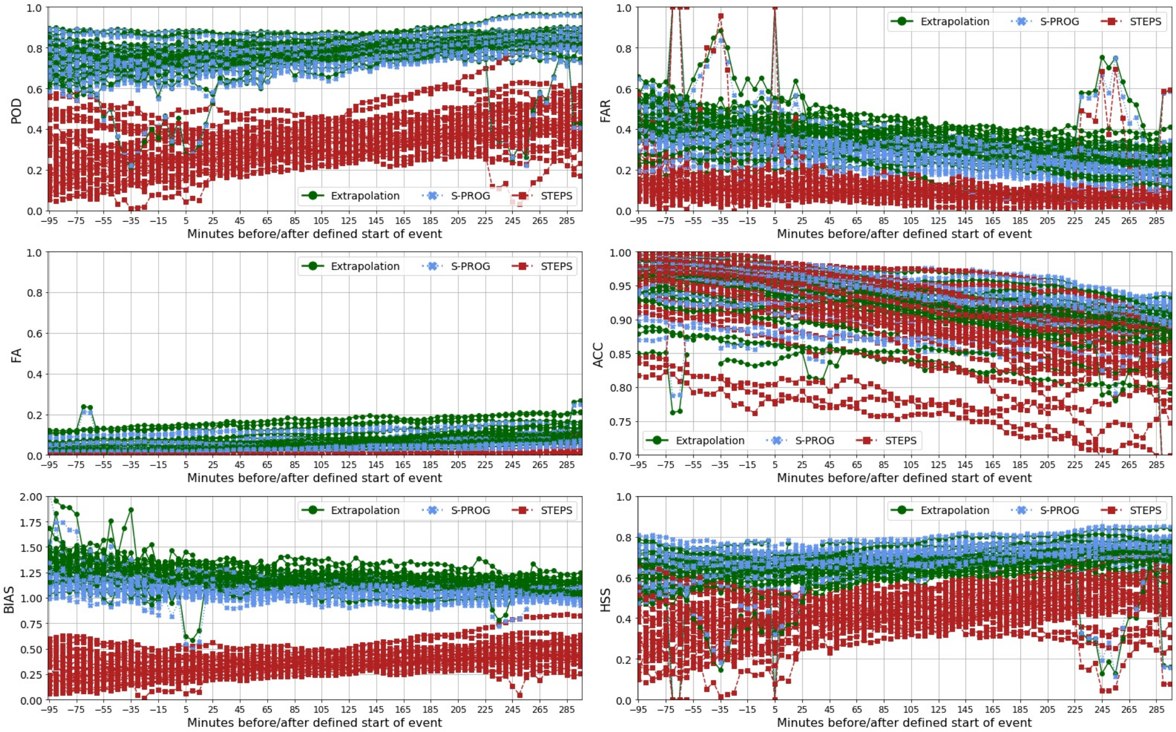


Figure 5.4: Graph of the statistics (clockwise starting from the top left graph: PoD, FAR, ACC, HSS, BIAS and FA) obtained in the evaluation of each nowcast model with respect to the time before or after the defined start of all events, using all the SWD field as input.

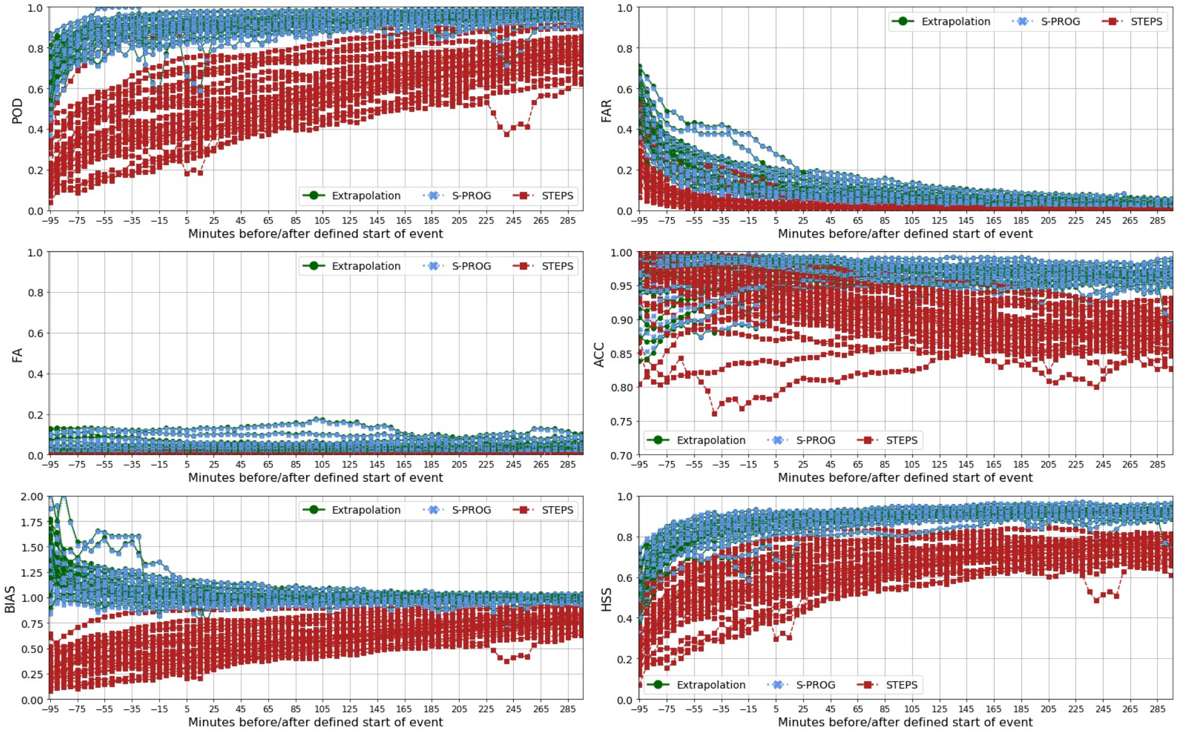


Figure 5.5: Graph of the statistics (clockwise starting from the top left graph: PoD, FAR, ACC, HSS, BIAS and FA) obtained in the evaluation of each nowcast model with respect to the time before or after the defined start of all events, using SWD clusters as input.

Figure 5.6 shows the scatter plots of the results of all statistics used to evaluate the models for all cases and all three nowcast models, comparing the results between using the entire filtered SWD field (labeled as ‘Using all SWD field’ on the graphs) and using only the detected clusters (marked ‘SWD clusters’). The scatter plots show that in general, using SWD clusters gives better overall results for the majority of cases. The PoD, FAR, and FA results show that the Extrapolation and S-PROG models are accurate and give similar results, with approximately 93% of cases with PoD of 70% or higher, 71% of the cases had a FAR below 0.3, and 68% of the cases had a HSS at or above 70%. While the STEPS model results were less accurate, with 85% of the PoD cases were below 0.7 and 99% of the FAR cases were below 0.3.

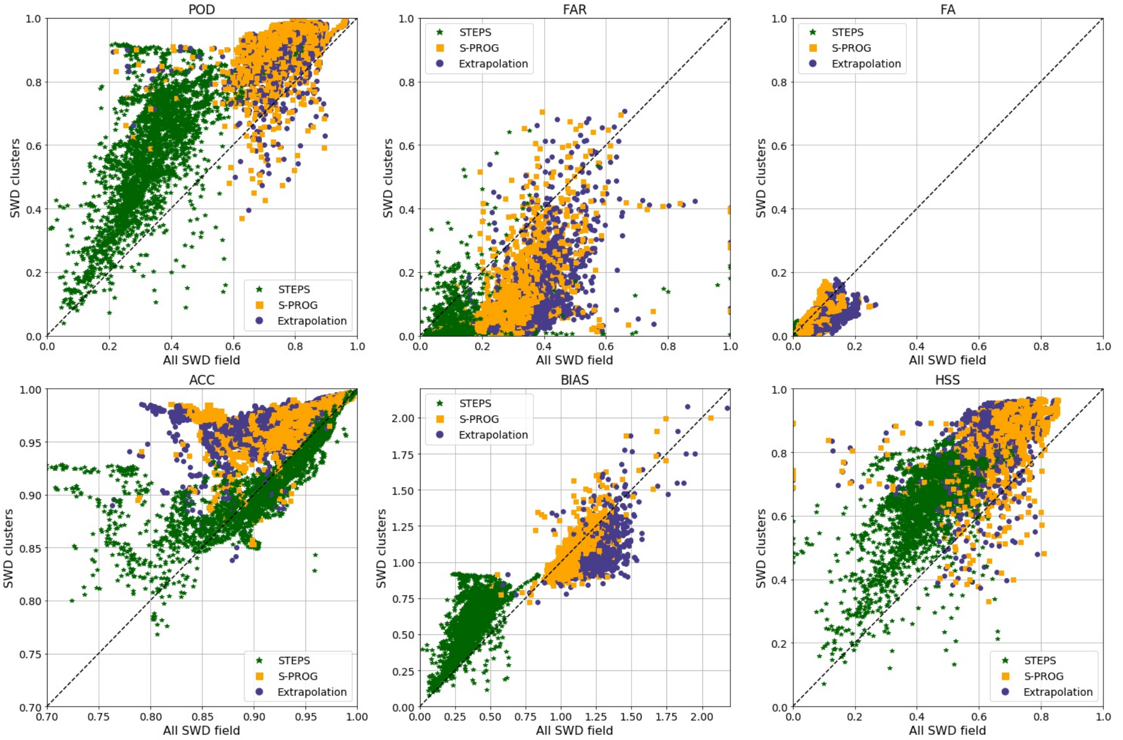


Figure 5.6: Scatter plot of the statistics (clockwise starting from the top left graph: PoD, FAR, FA, HSS BIAS and ACC) obtained in evaluating each nowcast model when using All SWD field vs. SWD clusters as inputs for the whole period of analysis.

Figures 5.7 and 5.7 present the averages of all the statistics results for each 5-minutes interval within the time frame the events were analyzed, and the difference between the average results obtained using the SWD field and SWD clusters. The graphs illustrate both the average SWD clusters results and the average all the SWD results improved with time. These graphs show that on average, using SWD cluster fields gave better results than using all the SWD field for all but the first 10 minutes of the period analyzed, and Extrapolation and S-PROG results were similar and better than STEPS. It should be noted that the ACC and FA are greatly affected by the large quantities of NaN values which are considered correct or true negatives. On average, the SWD data used in this analysis had 85% of NaN values (minimum of 40%) in an image and the radar data had 98% (minimum 85%). The evaluation of the nowcast considered correctly forecasting NaN values as correct or true negatives, which leads to the skewness of any statistics that heavily depend on the correct negatives to be close to 0 or 1. For this reason, the results for PoD and FAR were given more weight and consideration. Although several nowcast models that used satellite data (Rivolta et al., 2006; Zinner et al., 2008; Mecikalski et al., 2010a,b; Sieglaff et al., 2011; Lebedev et al., 2019; Ribeiro et al., 2019; Hill et al., 2020; Kumar et al., 2020; Ehsani et al., 2022), there have been no prior studies employing SWD satellite data as inputs in the pySTEPS nowcast models. While some studies have utilized the optical flow fields available in pySTEPS with satellite data (Burton et al., 2022) and not the nowcast models themselves, certain studies have evaluated S-PROG and Extrapolation pySTEPS nowcast models with radar data, using some of the same dichotomous statistics. Pulkkinen et al. (2020), Cuomo and Chandrasekar (2021), and Niu et al. (2021) conducted evaluations of the two models in the USA (former two) and China, respectively, with radar data for precipitation events. Their PoD and FAR results closely resemble those obtained with the SWD data, with values between 0.6 and 0.75 for average PoD and values between 0.2 and 0.4 for average FAR for lead times near 15 minutes. However, because Pulkkinen et al. (2020), Cuomo and Chandrasekar (2021), and Niu et al. (2021) involved radar data rather than satellite data, caution should be exercised when comparing these results with those obtained using the SWD.

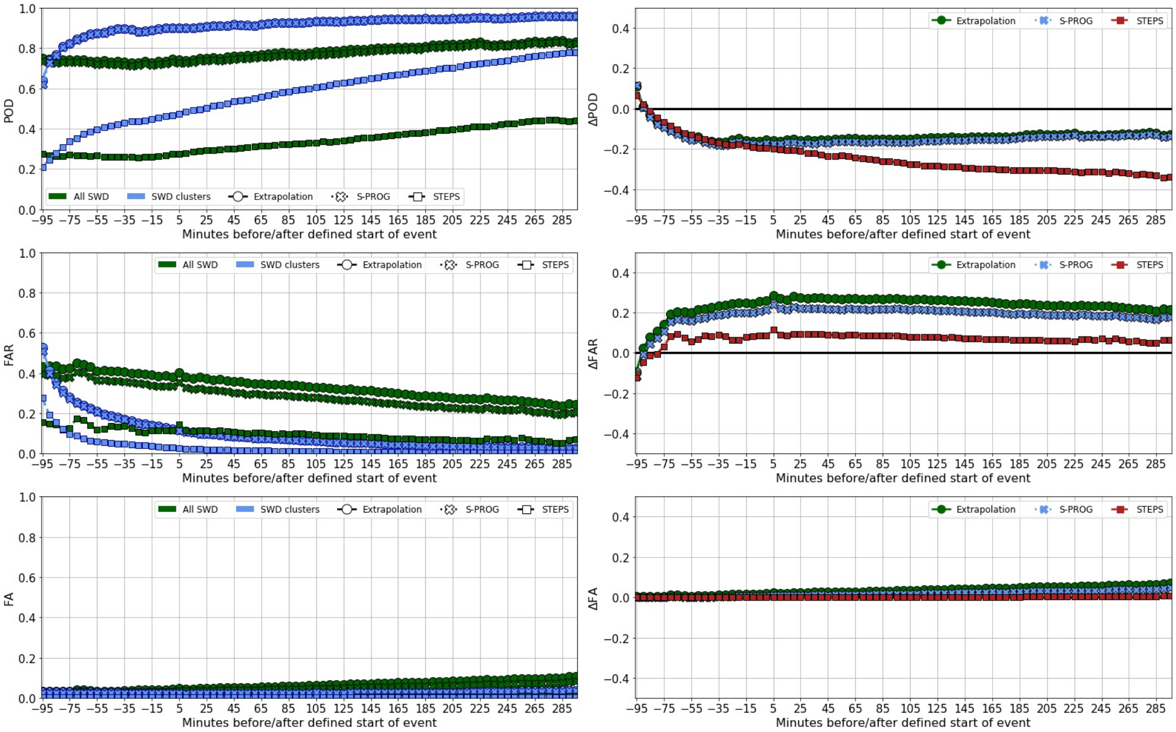


Figure 5.7: Left Column: graphs of the average (per nowcast model) of the statistics (from top row to bottom row) PoD, FAR, and FA obtained in the evaluation of each nowcast model with respect to the time before or after the defined start, using both all the SWD field and the SWD clusters as input. Right column: Difference of the results of the left column between using all the SWD and using the SWD clusters (“All SWD field” minus “SWD Clusters”) for each interval of time indicated.

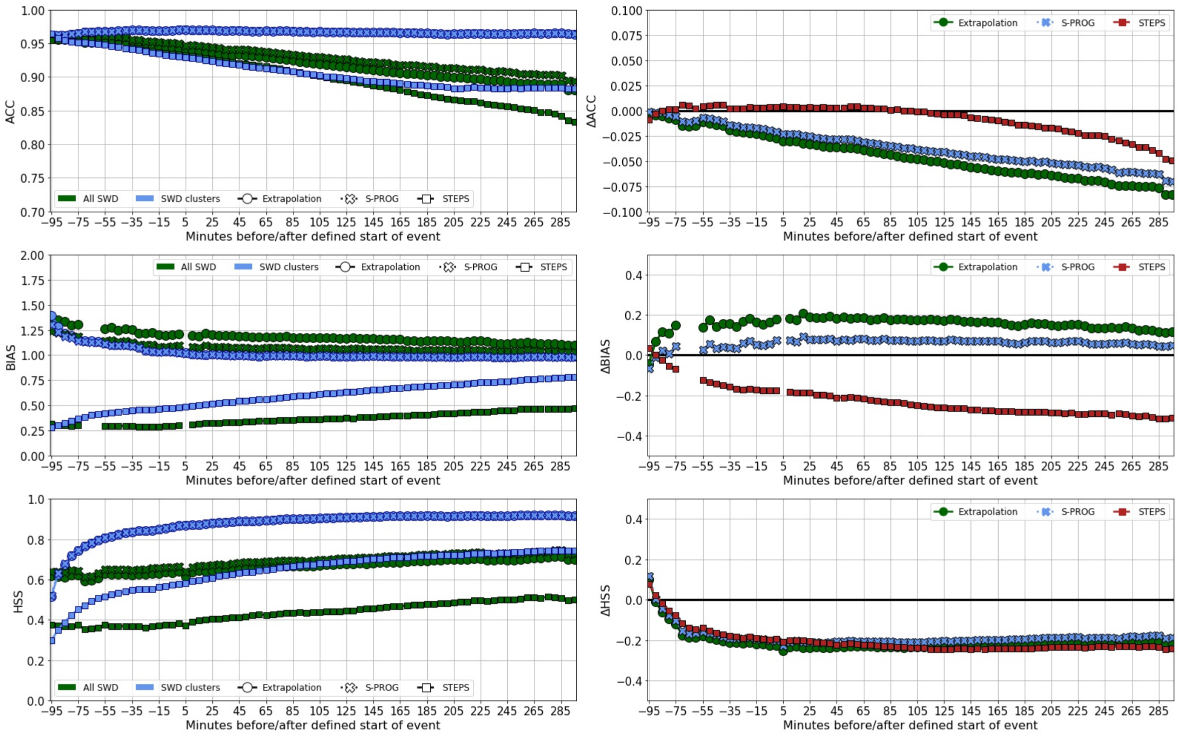


Figure 5.8: Same as 5.7 but for the statistics ACC, BIAS and HSS.

Although a general idea of the best nowcast models and the best type of SWD data input was observed in Figures 5.6, 5.7 and 5.8, the results do not point to a possible explanation. Therefore, visual images of a few of the events were obtained to understand these results. Figures 5.9, 5.10, and 5.11 display an example of the 15-minutes forecast for an event that started at 14/07/2018 20:32Z. The images show that the STEPS nowcast increased the number of false alarms in the borders of all SWD data clusters (big and small), thus explaining its low overall results in Figures 5.7 and 5.8. The Extrapolation and S-PROG provided similar results. When observing the images of the SWD clusters (not shown), the same observations regarding STEPS were made. Additionally, because the SWD clusters obtained from the DATing module did not include the smaller SWD areas, the errors that arose from not predicting the movement of these small SWD clusters were not considered. These errors, although small, were large in quantity, and therefore, help explain why using the entire SWD field resulted in overall less accurate forecasts. It should be noted that the DATing algorithm does not fully detect all large clusters, and this can cause missed areas of precipitation (the missed clusters were not taken into consideration when evaluating the forecasts).

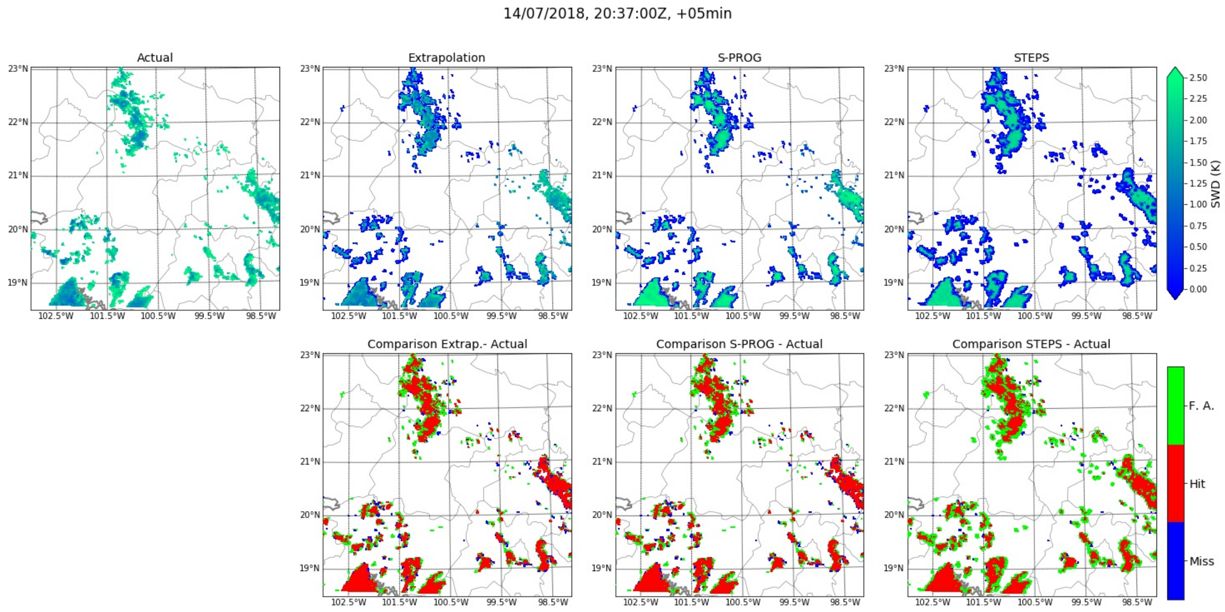

14/07/2018, 20:37:00Z, +05min

Figure 5.9: Images of the observed SWD field and the 5-min forecast SWD field obtained using the three nowcast models indicated (top images) and the comparison between the observed data and the forecast data (bottom images) for the input image occurring at 14/07/2018 20:32Z.

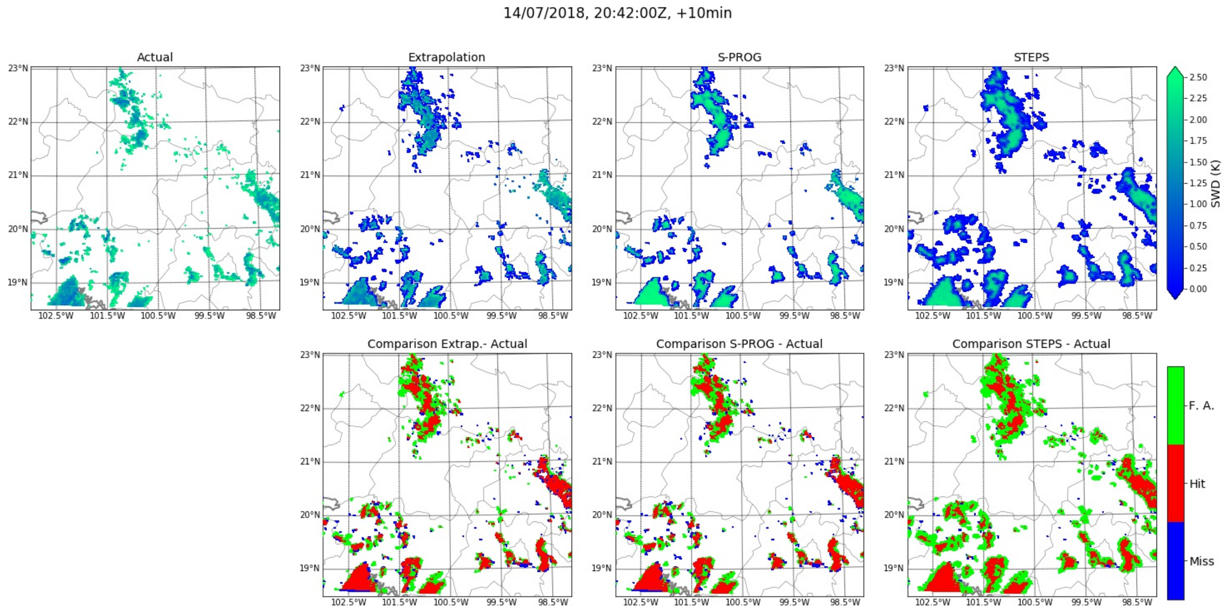

14/07/2018, 20:42:00Z, +10min

Figure 5.10: Same as Figure 5.9 for the 10-min forecast.

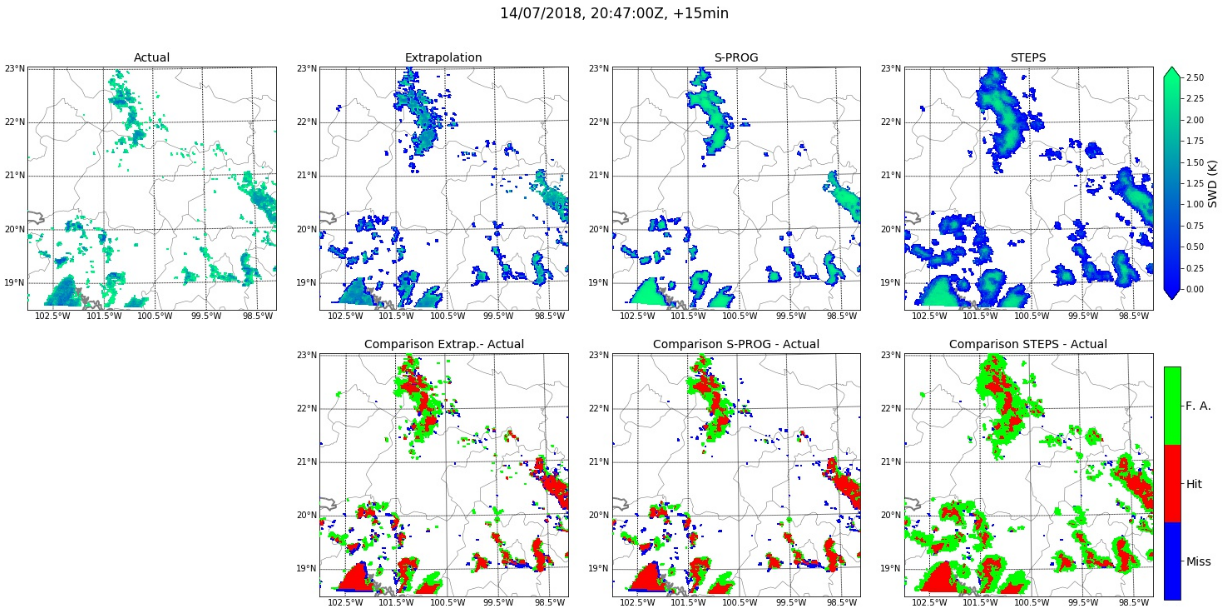


Figure 5.11: Same as Figure 5.9 for the 15-min forecast.

5.5.2 Evaluation of nowcast models using Queretaro radar data

The same evaluation of the nowcast models presented in the previous section was also done using the Queretaro rain rate fields as input (called all rain rate field). Figures 5.12 show all the statistics results per 5 minute interval in the period of analysis for all events, for the results obtained using all rain rate field and and Figure 5.13 for the results acquired from the rain rate inside the SWD clusters. These graphs revealed a lack of consistency in the behavior within the hour after the start of the event, as well as low forecast skill. However, the forecast skill improves as the events evolve. In addition, the Extrapolation and S-PROG models give better results than STEPS.

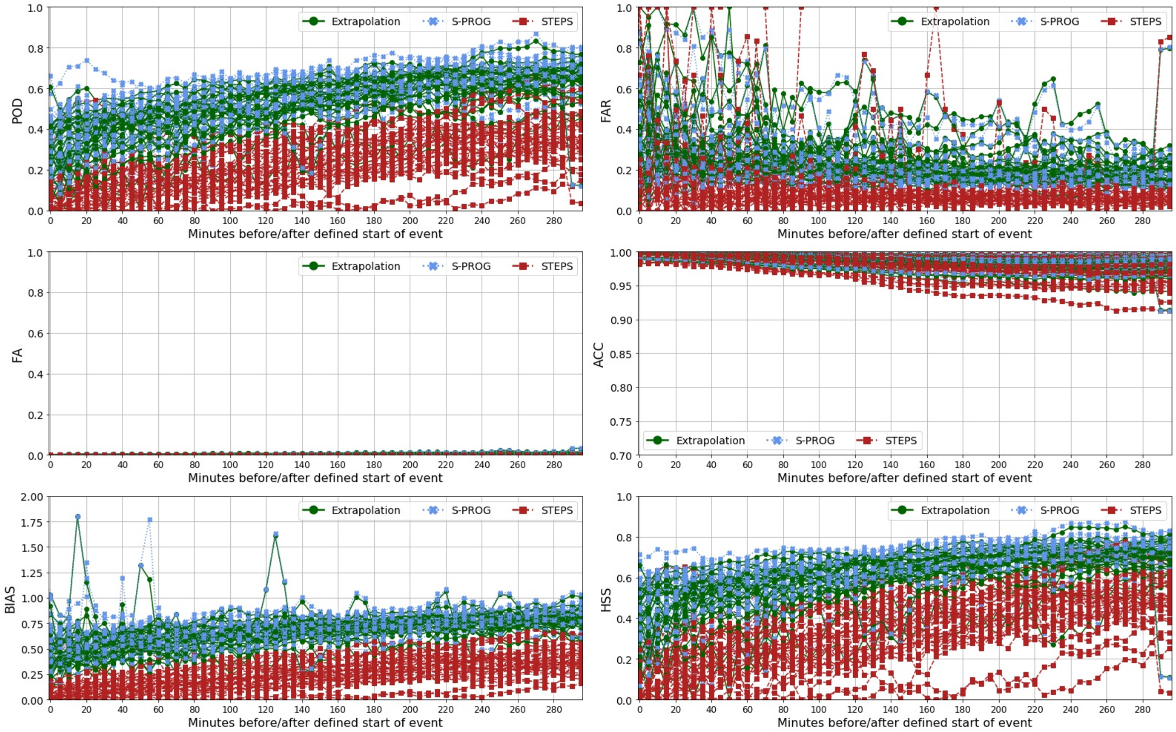


Figure 5.12: Graph of the statistics (clockwise starting from the top left graph: PoD, FAR, ACC, HSS, BIAS and FA) obtained in the evaluation of each nowcast model with respect to the time before or after the defined start of all events, using “All Rain Rate (RR) field” as input.

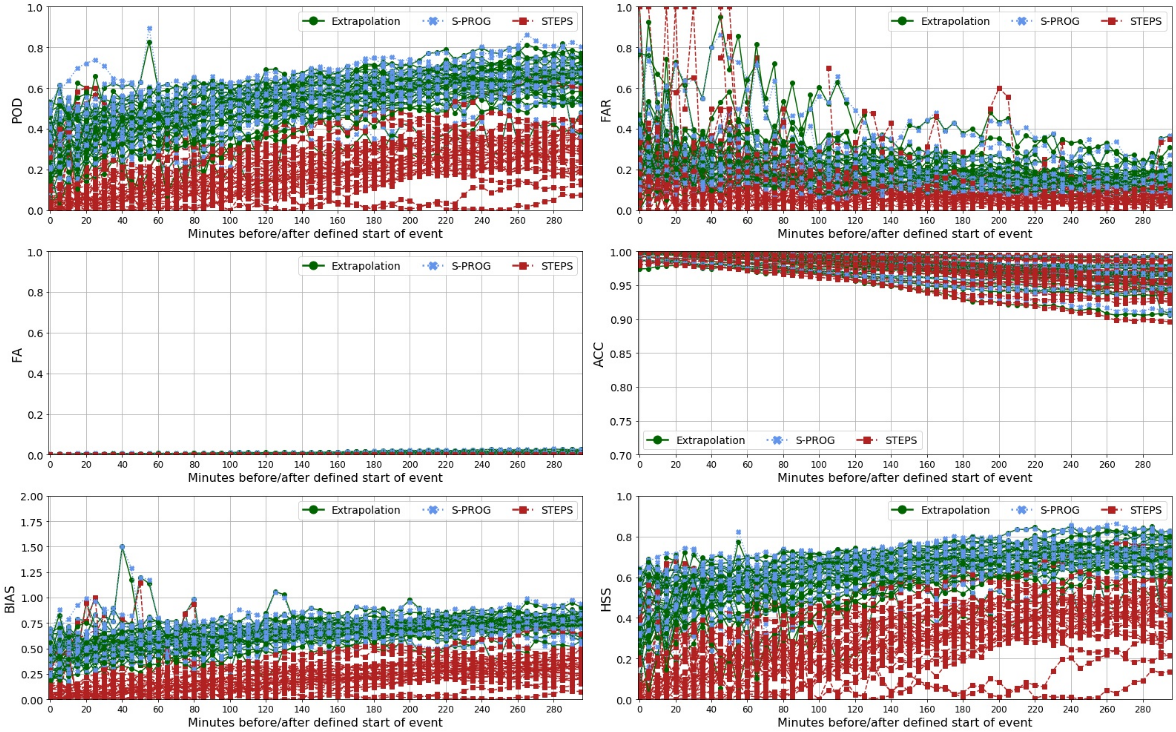


Figure 5.13: Graph of the statistics (clockwise starting from the top left graph: PoD, FAR, ACC, HSS, BIAS and FA) obtained in the evaluation of each nowcast model with respect to the time before or after the defined start of all events, using "Rain Rate (RR) in SWD clusters" as input.

Figure 5.14 shows the scatter plot results for all cases using the two models comparing radar rain rate data (called rain rate field) and (only using) the rain rate that fell within the SWD clusters detected (referred as rain rate in clusters). The results for the PoD and HSS suggest that using the rain rate was not as accurate as using the SWD. For Extrapolation and S-PROG, the PoD, 93% of the cases fell below 0.7 and 58% fell between 0.7 and 0.5. For FAR, 87% of the cases were below 0.3. For STEPS, all cases of PoD were less than 0.7 and 97% of FAR results were below 0.3. For many cases, the difference between using "rain rate in clusters" and "All rain rate field" was small, since most values fell around the 45° angle line passing though the origin, also known as the identity line.

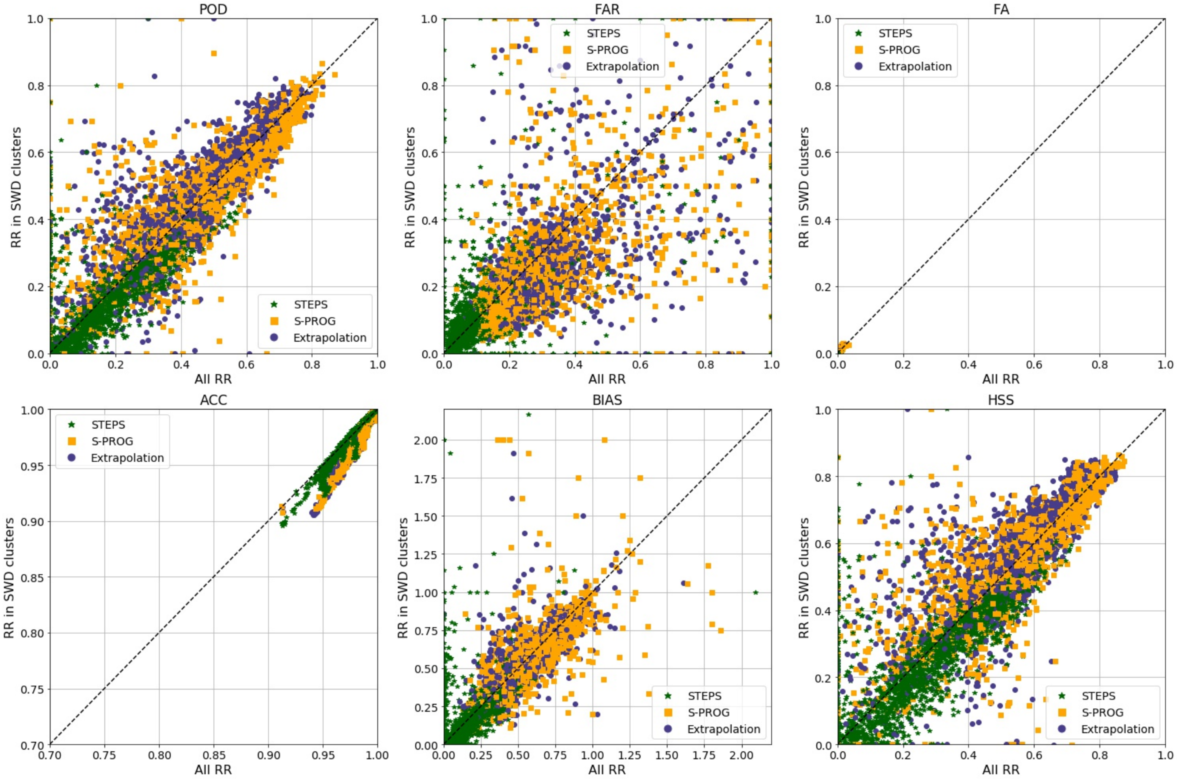


Figure 5.14: Scatter plot of the statistics (clockwise starting from the top left graph: PoD, FAR, FA, HSS BIAS and ACC) obtained in evaluating each nowcast model when using “All Rain Rate field” vs. “Rain rate in SWD clusters” as inputs for the period of analysis.

Figures 5.15 and 5.16 show the average of the statistical results for the 15-minute forecast for each 5-minute interval after the defined start of the events. These graphs show that, on average, using all the rain rate field resulted in similar forecast as using the rain rate in clusters. Additionally, for Extrapolation, S-PROG and STEPS using all the rain rate field, the PoD tended to increase on average by almost 0.4 in the 295 minutes after the start of the event, and FAR decreased by 0.2 in the same period of time. STEPS for the rain rate in the SWD clusters saw an higher increase of approximately 0.6 in 295 minutes. The results for the rain rate field are not as accurate as those obtained by previous works. Pulkkinen et al. (2020) evaluated the models for 10 events in the Texas, USA region and obtained PoD values of 0.6 and 0.75 for Extrapolation and S-PROG respectively, and FAR values between 0.3 and 0.4 for both models. Cuomo and Chandrasekar (2021) used 80 events from Texas and Colorado, USA, and radar reflectivity with S-PROG and Extrapolation, which resulted in average PoD of approximately 0.75 and average FAR of 0.2 for the 15-minute lead time. Lastly, Niu et al. (2021) used the Extrapolation and S-PROG models with radar reflectivity data from China for over 100 events. They found that for short lead times (12 and 24 minutes), the POD reached 0.75 and the FAR was below 0.2. However, the PoD values were closer to those observed by Pulkkinen et al. (2020), Cuomo and Chandrasekar (2021), and Niu et al. (2021) as the events developed.

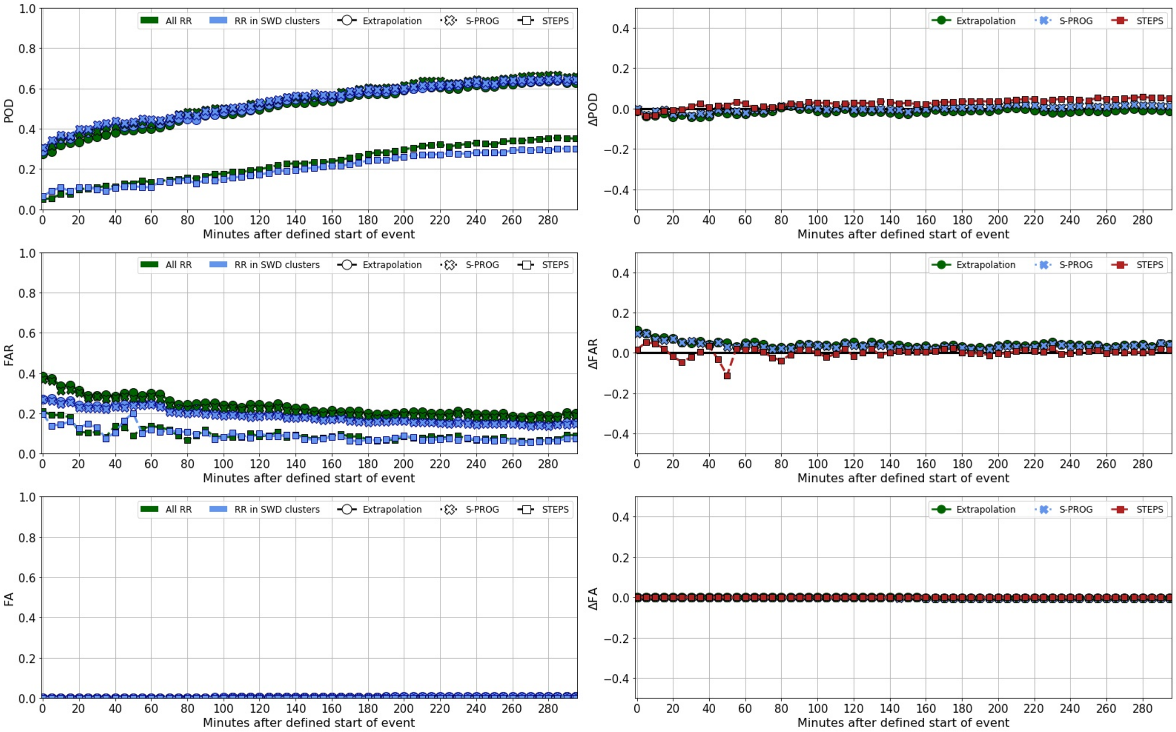


Figure 5.15: Left Column: graphs of the average (per nowcast model) of the statistics (from top row to bottom row) PoD, FAR, and FA obtained in the evaluation of each nowcast model with respect to the time before or after the defined start, using both all the rain rate field and the rain rate in SWD clusters as input. Right column: Difference of the results of the left column between using all the rain rate and using the rain rate in SWD clusters ("All rain rate (RR)" minus "RR in SWD Clusters" for each interval of time indicated.

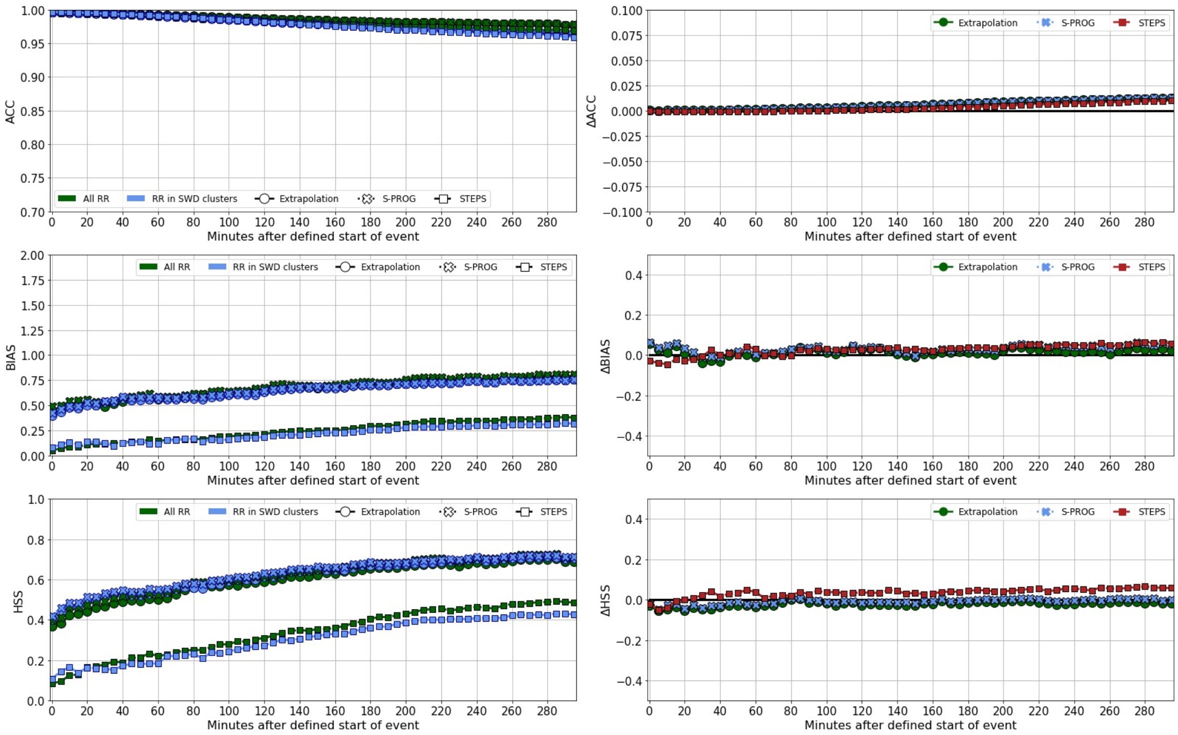


Figure 5.16: Same as 5.7 but for the statistics ACC, BIAS and HSS.

An image analysis was also carried out for the forecast using the same examples as those used with the SWD data. The example can be observed in Figures 5.17, 5.18 and 5.19 for the input image occurring at 14/07/2018 20:32Z, These showed similar observations as those made with the SWD data: a large amount of false alarms for the STEPS model, and similar results between the Extrapolation and S-PROG models. Additionally, a simple visual analysis of the images point to the fact that all the models tended to overestimate the intensity of rain rate.

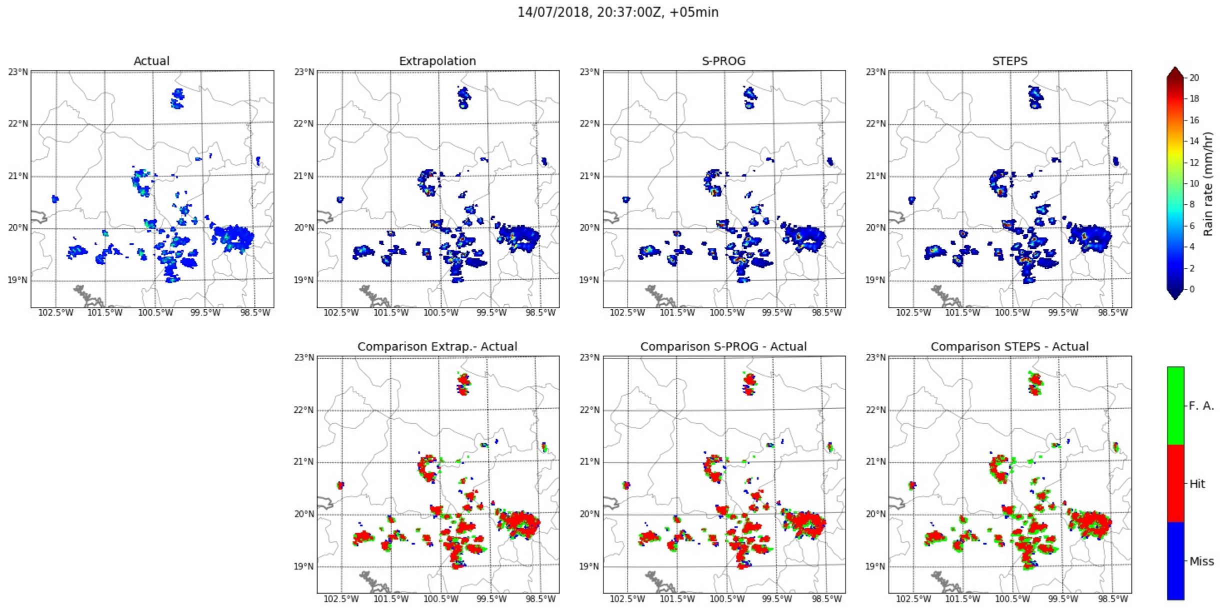


Figure 5.17: Images of the observed rain rate field and the 5-min forecast rain rate field obtained using the three nowcast models indicated (top images) and the comparison between the observed data and the forecast data (bottom images) for the input image occurring at 14/07/2018 20:32Z.

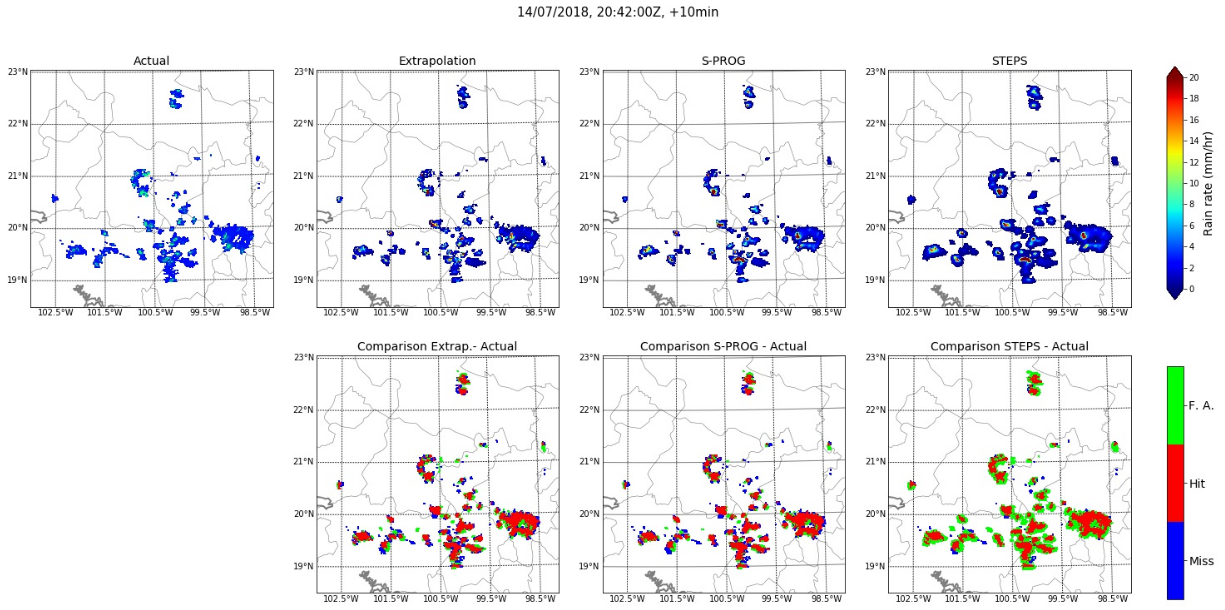


Figure 5.18: Same as 5.17 for the 10-min forecast.

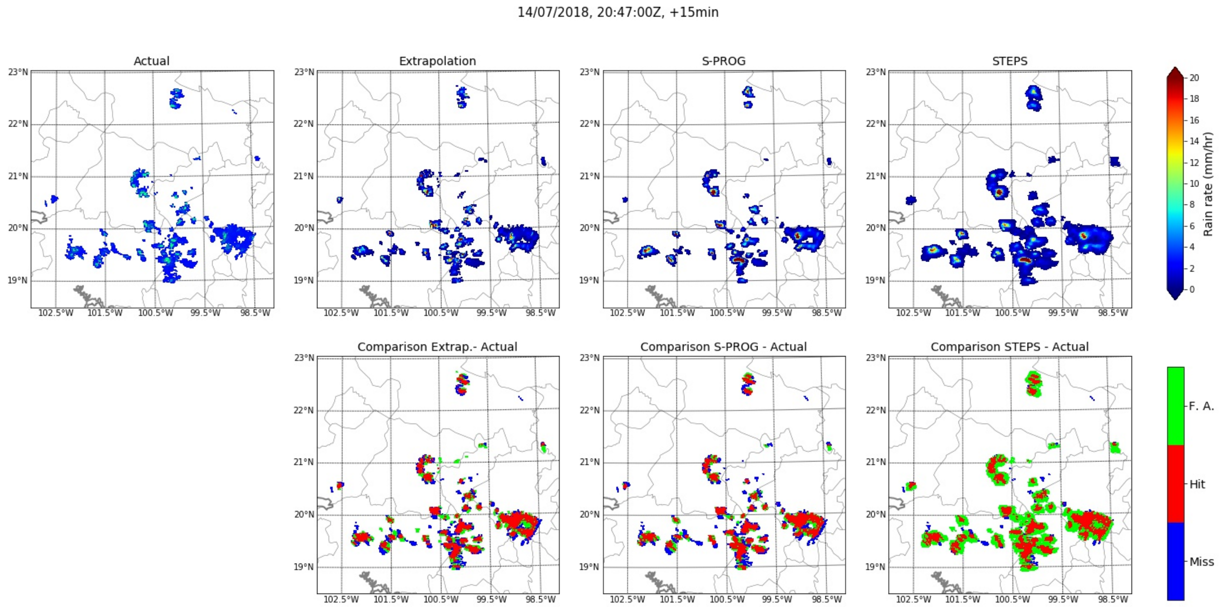


Figure 5.19: Same as 5.17 for the 15-min forecast.

In general, as mentioned before, these results were not as accurate as those obtained using SWD data, in part because there was a greater amount of SWD data and the SWD clusters were larger, increasing the likelihood of hits. However, FAR results were curious because they were slightly better for the rain rate results than for using the SWD field. The reason for this inconsistency is due to how the nowcast process SWD and rain rate data. Figures 5.9 - 5.11 and 5.17 - 5.19 can give an idea of this process. Both nowcast models tend to create the false alarms and misses around the edges of the variable clusters in an image. But the models tend to create a larger quantity of false alarms for SWD than for rain rate, most likely due to the fact that the models are not made for forecasting this particular variable. The variation in the number of false alarms created by the models with each type of data lead to a variation in the statistics that depend on the false alarms, such as FAR. For this reason, the average FAR for the SWD was around 0.1 higher than for the radar rain rate.

5.5.3 Evaluation of nowcast models limiting the Queretaro radar data

The processing methods used to clean the radar data were not one hundred percent precise and, therefore, a significant quantity of noise in the data set was still used in this study. One way to limit or clean some of this noise was by using SWD data. This was partly done in the previous section when presenting the results using the rain rate that falls inside SWD clusters, however, further analysis were done. The nowcast model was also evaluated with radar data by limiting the radar data using the entire SWD data field. The limitations used were the following: 1) only using the rain rate values that fell within the SWD field; 2) using the rain rate field with the SWD motion field; and 3) using the rain rate values that fell within the SWD field with the SWD motion field.

Figure 5.20 and 5.21 show the differences in average PoD and FAR, respectively, obtained when using the entire rain rate field (denoted in this section as “All rain rate”), the rain rate values inside the SWD clusters (denoted in this section as ”rain rate clusters”), and the additional three limitations previously described for the 15-minute forecast obtained in every 5-minute interval of the period of study. For these graphs, positive values indicate that using “All rain rate” yielded the highest results for the variable in question. In general, there was little difference in the modified rain rate fields and“All rain rate”. For PoD, using All rain rate resulted in higher PoD less than 0.1 for all cases except the comparison with ”rain rate clusters” in the first hour of the event. For FAR, the difference was higher than PoD, at 0.1 or below, but the results favored using “All rain rate”, with the exception of comparing All rain rate the ”rain rate clusters”, the FAR was higher for “All rain rate”.

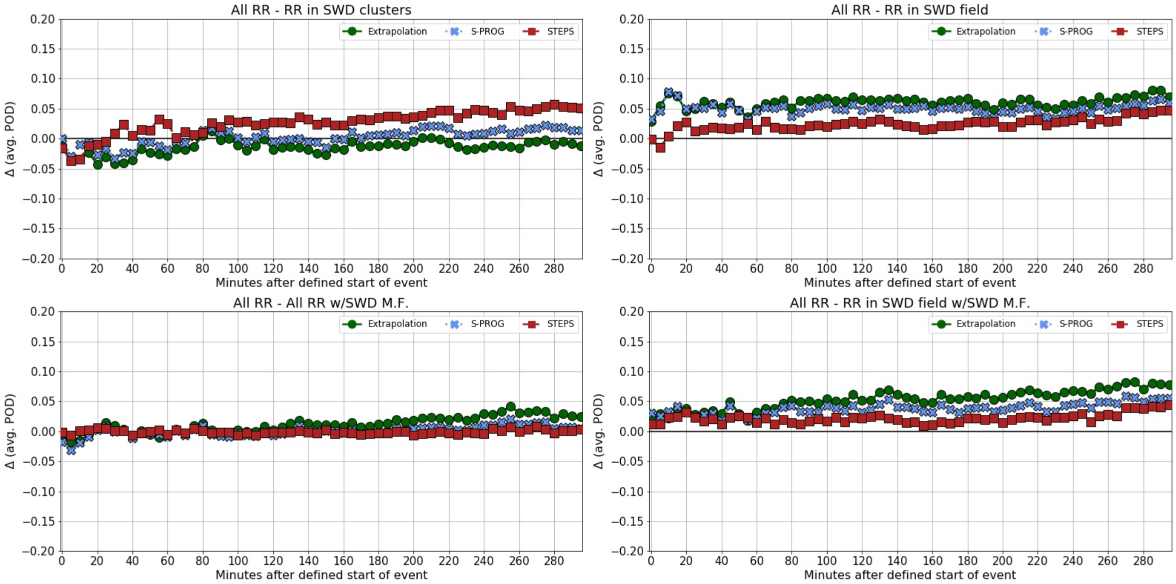


Figure 5.20: Graph of the differences in average POD for the 15-min forecast using the image of each 5-minute interval after the defined start of the event for rain rate data. The difference is between using all the rain rate field (All RR) and the rain rate inside the SWD clusters (top left), the rain rate inside the SWD field (top right), the rain rate using the SWD motion field (bottom left), and the rain rate inside the SWD field with the SWD motion field (bottom right).

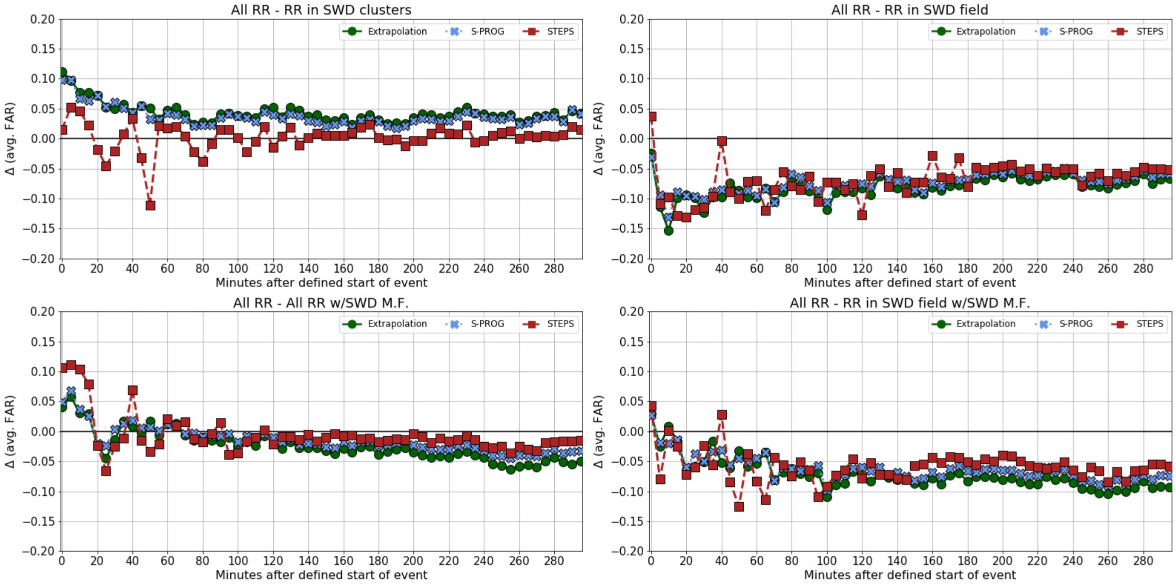
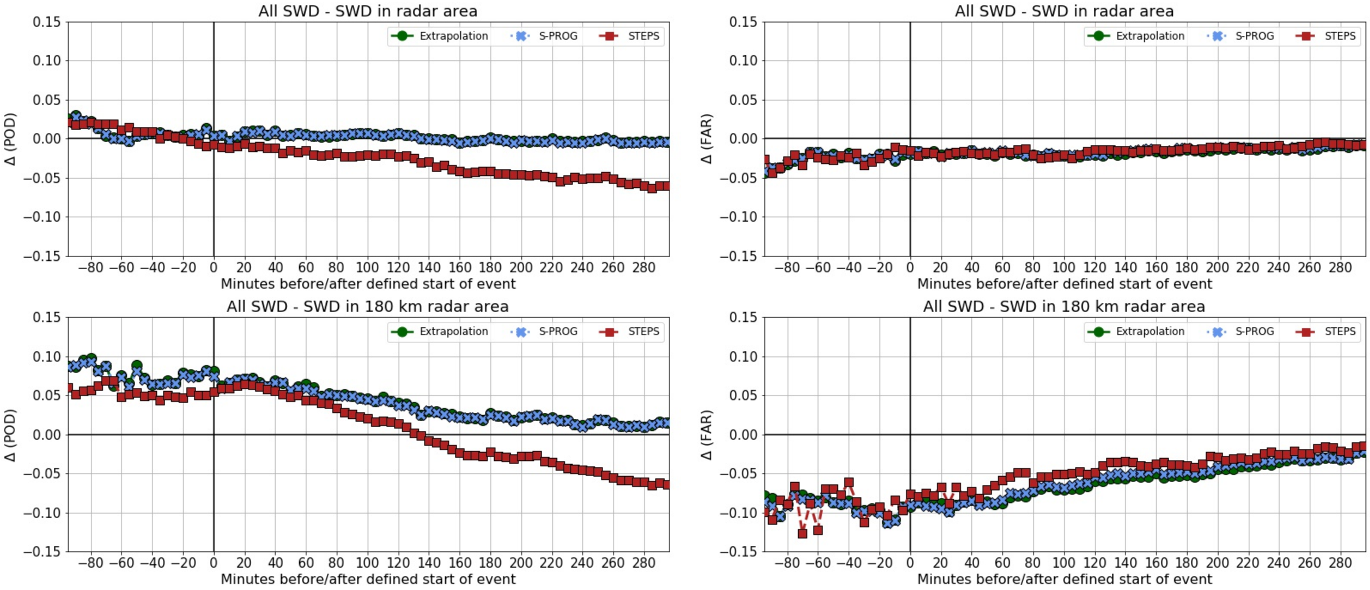


Figure 5.21: Same as 5.20 but for the average FAR.

5.5.4 Evaluation of nowcast models with variation in coverage area

The previous results showed that the nowcasts gave less accurate results when radar data was used as input. This might be because the radar has some limitations in the coverage area. In other words, the radar can only potentially “see” within the radar coverage area, and there are factors that cause the radar to not be able to acquire data from all its coverage area. In order to see if this limitation had any effect on the overall results, the nowcast models were evaluated by applying two limits to the area. The first limit was to limit the data to the radar coverage area (denoted “all radar area”; Figure 4.3 in Chapter 4); and the second limit was to draw a 180 km radius circle around the radar (where the radar has the best coverage), denoted as “180 km radar area”.

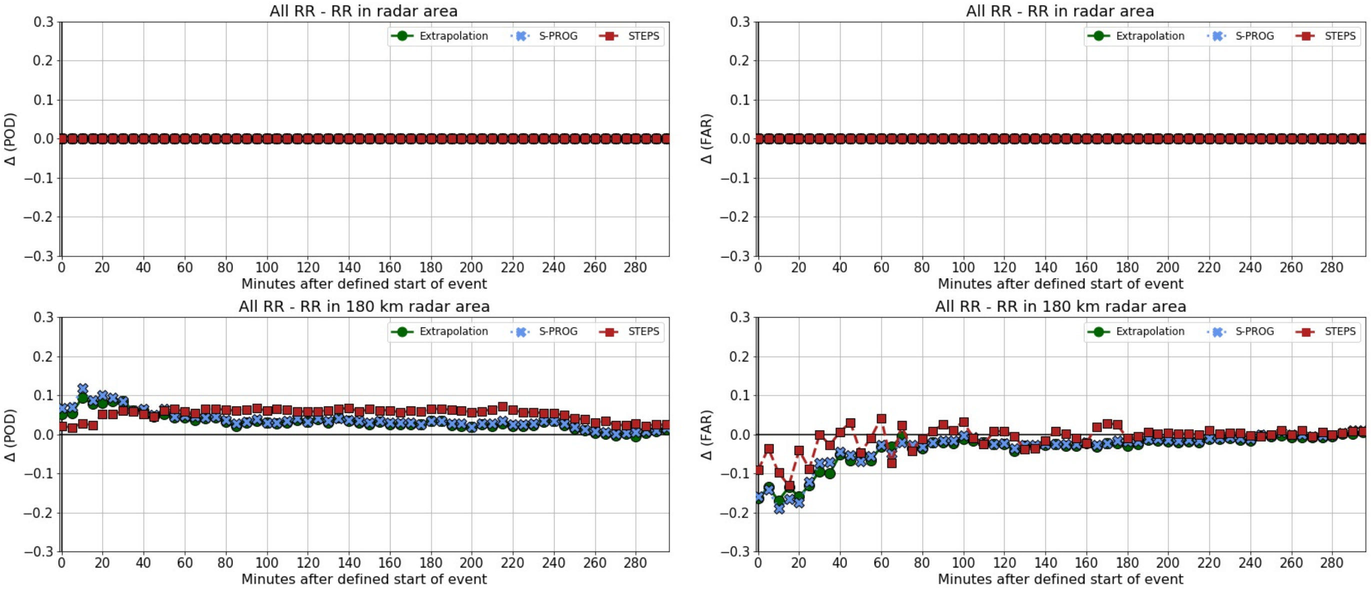


(a) (b)

Figure 5.22: Graph of the differences in (a) average PoD and (b) average FAR for the 15-min forecast using the image of each 5-minute interval after the defined start of the event for rain rate data. The difference is between using all the SWD field and the SWD in the radar area (top), and between all the SWD field and the SWD in the 180 km radar area (bottom).

Figure 5.22 shows the differences in the average PoD and the average FAR between using the entire SWD field and using only the SWD in all radar area (top graph of both images); and using only the SWD in the 180 km radar area (bottom graph in both figures). For the comparison between all the SWD field and the SWD within the radar area, the results were less than 0.025 better than with all the SWD field, and the highest change was between 95- and 75-minutes prior the start of the event. The exception is for STEPS’s PoD result, where the PoD in the radar area is higher after the start of the event, reaching around 0.5 difference by the end of the period of analysis. On the other hand, when comparing with the SWD in 180 km radar area, the statistics were lower by almost 0.0.5 before the start of the event, but the difference decreased by the end of the time of analysis. The rate of change in STEPS PoD results was higher than the other nowcast starting 60 minutes after the start of the event, with the PoD in the 180 km radar area being higher after 120 minutes from the start of the event.

For radar rain rate, Figure 5.23 shows the differences in the average PoD and FAR between using the entire rain rate field, using the rain rate in all radar area, and the rain rate in the 180 km radar area. As expected, there was no difference between the entire rain rate field and the rain rate in all radar area because this limit had already been established for the radar data. However, it is important to note that this signifies that the models are not predicting any rain outside the area of coverage. As for the 180 km radar area, the results for both PoD and FAR were better when using all the radar field in the first 60 minutes after the start of the event analysis period, but the improvement decreased to less than 0.05 for the rest of the period, when the events were theoretically more developed and there was more area with rainfall. This indicates that with more real data, the results improve.



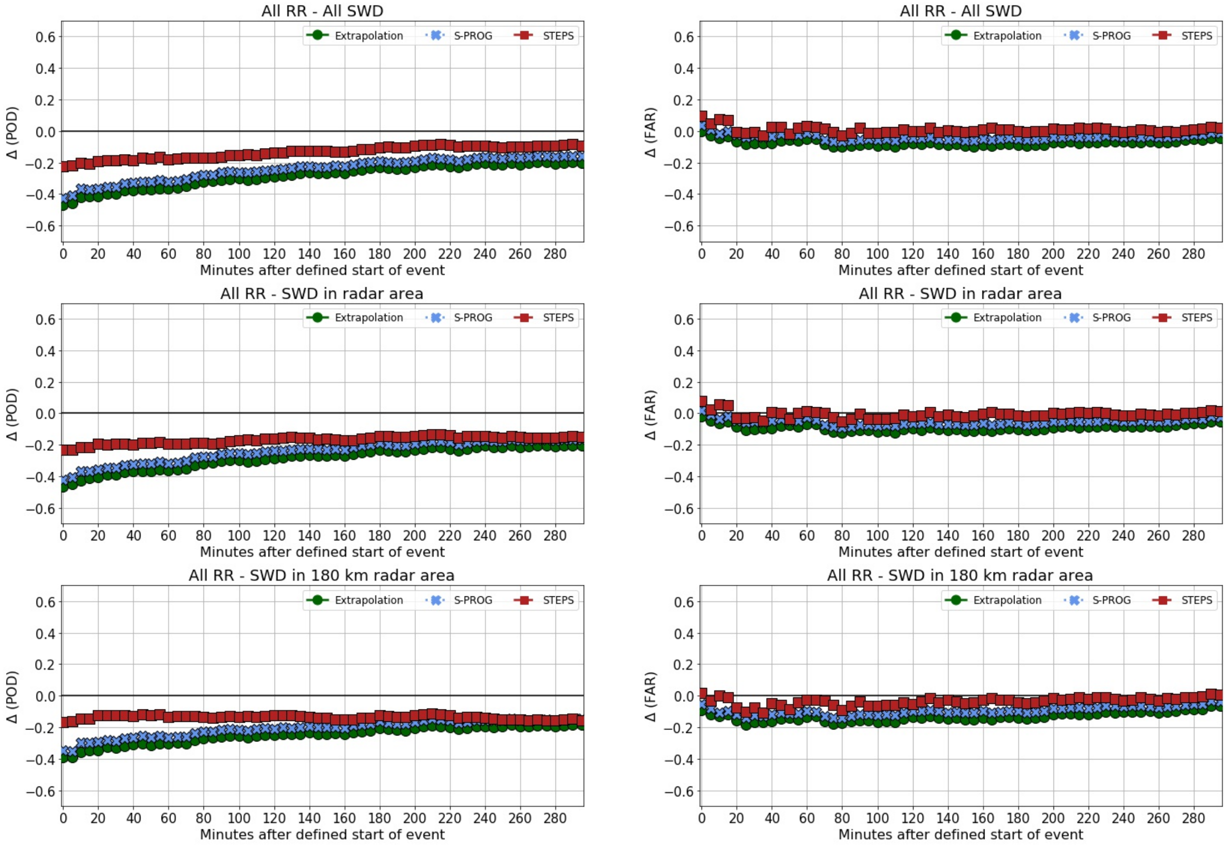
(a)

(b)

Figure 5.23: Graph of the differences in (a) average PoD and (b) average FAR for the 15-min forecast using the image of each 5-minute interval after the defined start of the event for rain rate data. The difference is between using all the RR field and the RR in the radar area (top), and between all the RR field and the RR in the 180 km radar area (bottom).

Finally, Figure 5.24 shows similar graphs where the differences between the average PoD and average FAR can be observed between using the rain rate field and the entire SWD field, the SWD in all radar area and the SWD in the 180 km radar area. For all comparison, the rain rate field had smaller POD and FAR values. The POD tended to be more than 0.2 higher for the SWD fields, with the difference being smallest for STEPS, and the FAR was at most 0.1 higher for the SWD fields, with the smallest difference resulting with STEPS. The variation in the difference between the rain rate field results and the each of the SWD fields compared was very small, reaching a maximum of 0.05 at the beginning of the event and decreasing to almost 0 by the end.

These results suggest that the lack of even coverage between the SWD and radar data has an effect on the overall results, and that the differences decreased as the events evolved and grew. Meaning, the size of the events could play a significant role in the overall results of the evaluation of the nowcast models. On the other hand, it can also be observed that the effect of the size of coverage area of the input on the overall results is not enough to explain why the radar results are less accurate than the SWD results.



(a) (b)

Figure 5.24: Same as Figure 5.22 but showing the difference between using all the RR field and all the SWD field, between the RR field and the SWD in the radar area, and between the RR field and the SWD in the 180 km radar area. (a) is for average PoD and (b) is for average FAR.

5.5.5 Relationship between NaN pixels in input image and and statistics results

In the previous section, the results showed that the nowcast model skill decreased with decreasing coverage, it was better when using SWD data, and increased with the size or development of the events. One of the main reasons proposed for this was the fact that the radar rain rate data contained far fewer data values within each image than the SWD data, causing outliers to have a significant effect on the results. This particular hypothesis was analyzed in this section using a fraction of the number of pixels with real data (non-NaN) for each of the variables for all cases to see if there was any correlation with the results.

Figure 5.25 displays a scatter plot showcasing the log_{10} of the fraction of pixels in each image that contain real data (pixel fraction) and the PoD for each of the two nowcast models divided into SWD and radar rain rate categories. It also illustrates their linear correlation. All datasets exhibit a visually linear and positive correlation with the log_{10} of the pixel fraction. For SWD data, the concentration of data points was higher in the PoD values above 0.5, yet some outliers were notably distant from the main data cluster. In radar data, the values remained below 0.8, with a more scattered distribution around the correlation line. Unlike SWD data, radar data did not display many outliers far from the primary data cluster.

Figure 5.26 illustrates a similar pattern but for FAR. In this case, there appears to be a negative linear correlation with the log_{10} of pixel fraction. SWD data is closer to the correlation line with a few outliers, whereas radar data is more widely dispersed, especially for pixel fraction values below 0.003 (approximately 186 pixels). The observed linear correlation in the graphs is supported by the Pearson's R (R), Spearman's Rank (ρ) and Kendall tau (τ) values outlined in Table 5.1. The table showcases R, ρ, and τ values, denoting the correlation between each statistic studied (except bias) and the log_{10} of the pixel fraction, or specifically the pixel fraction (only for ACC and FA). All statistics exhibit a high correlation with the pixel fraction, except for the HSS for SWD, which shows a moderate correlation, and FAR for STEPS's rain rate results, which presents a moderately weak correlation.

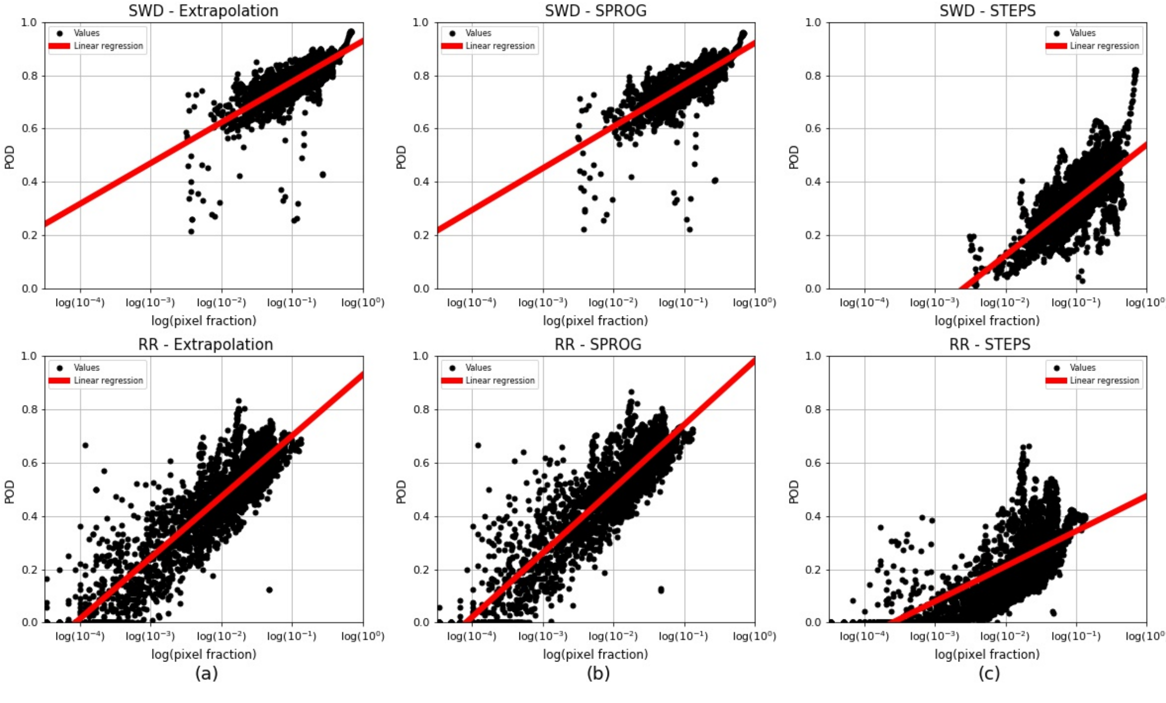
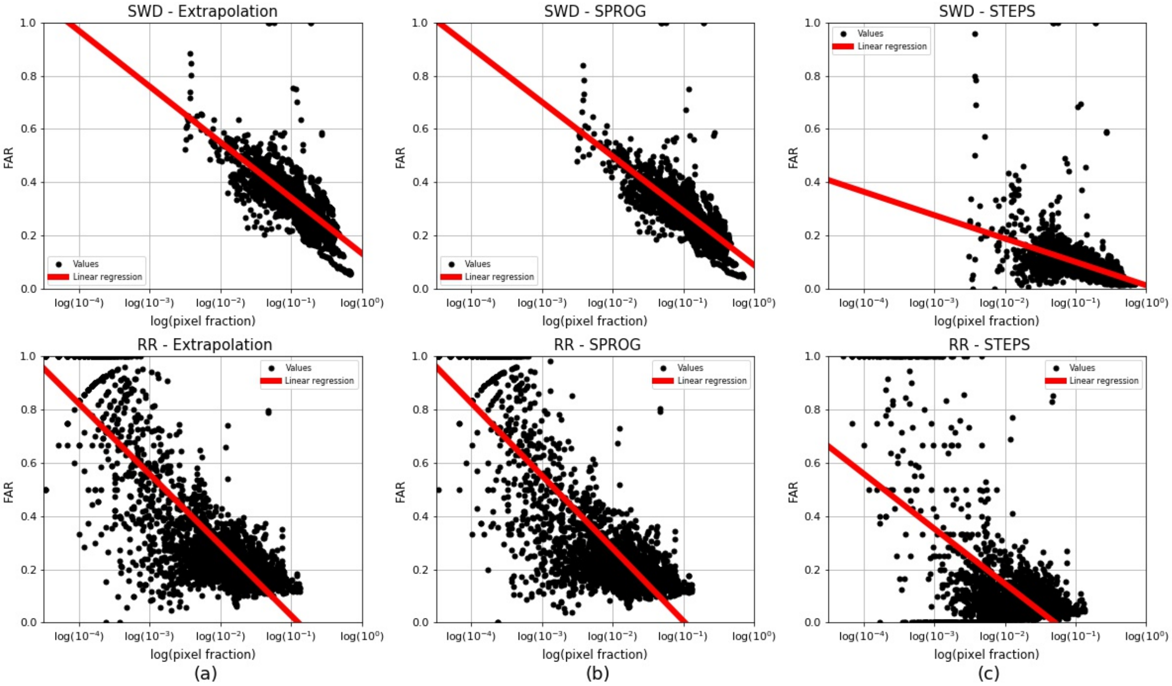


Figure 5.25: Scatter plot of the $log_{10}(pixel\ fraction)$ and the PoD for SWD (top graphs) and rain rate (RR; bottom graphs) results for (a) Extrapolation, (b) S-PROG, and (c) STEPS (pixel fraction is the fraction of the non-Nan pixels).



(a) (b) (c)

Figure 5.26: Same as 5.25 for FAR.

Statistics	Model	Pearson's R		Spearman's Rank		Kendall Tau	
		SWD	RR	SWD	RR	SWD	RR
POD*	Extrapolation	0.77	0.91	0.83	0.88	0.64	0.70
	S-PROG	0.78	0.91	0.84	0.87	0.65	0.70
	STEPS	0.70	0.77	0.67	0.84	0.49	0.65
FAR*	Extrapolation	-0.79	-0.84	-0.83	-0.73	-0.65	-0.56
	S-PROG	-0.79	-0.84	-0.84	-0.73	-0.65	-0.55
	STEPS	-0.44	-0.60	-0.60	-0.23	-0.44	-0.17
FA	Extrapolation	0.89	0.93	0.96	0.97	0.83	0.86
	S-PROG	0.88	0.92	0.95	0.97	0.82	0.85
	STEPS	0.79	0.70	0.92	0.90	0.77	0.72
ACC	Extrapolation	-0.80	-0.96	-0.94	-0.98	-0.81	-0.88
	S-PROG	-0.81	-0.95	-0.94	-0.98	-0.81	-0.87
	STEPS	-0.92	-0.97	-0.98	-0.98	-0.88	-0.89
HSS*	Extrapolation	0.55	0.91	0.60	0.87	0.44	0.70
	S-PROG	0.60	0.91	0.67	0.87	0.50	0.68
	STEPS	0.54	0.81	0.47	0.84	0.34	0.64

Table 5.1: Results of the correlation coefficients between the fraction of total pixels with real data and the statistics results for all nowcast models. Asterisk ($*$) marks the use of the $log_{10}(pixel\ fraction)$ for the correlation analysis (pixel fraction is the fraction of the non-Nan pixels).

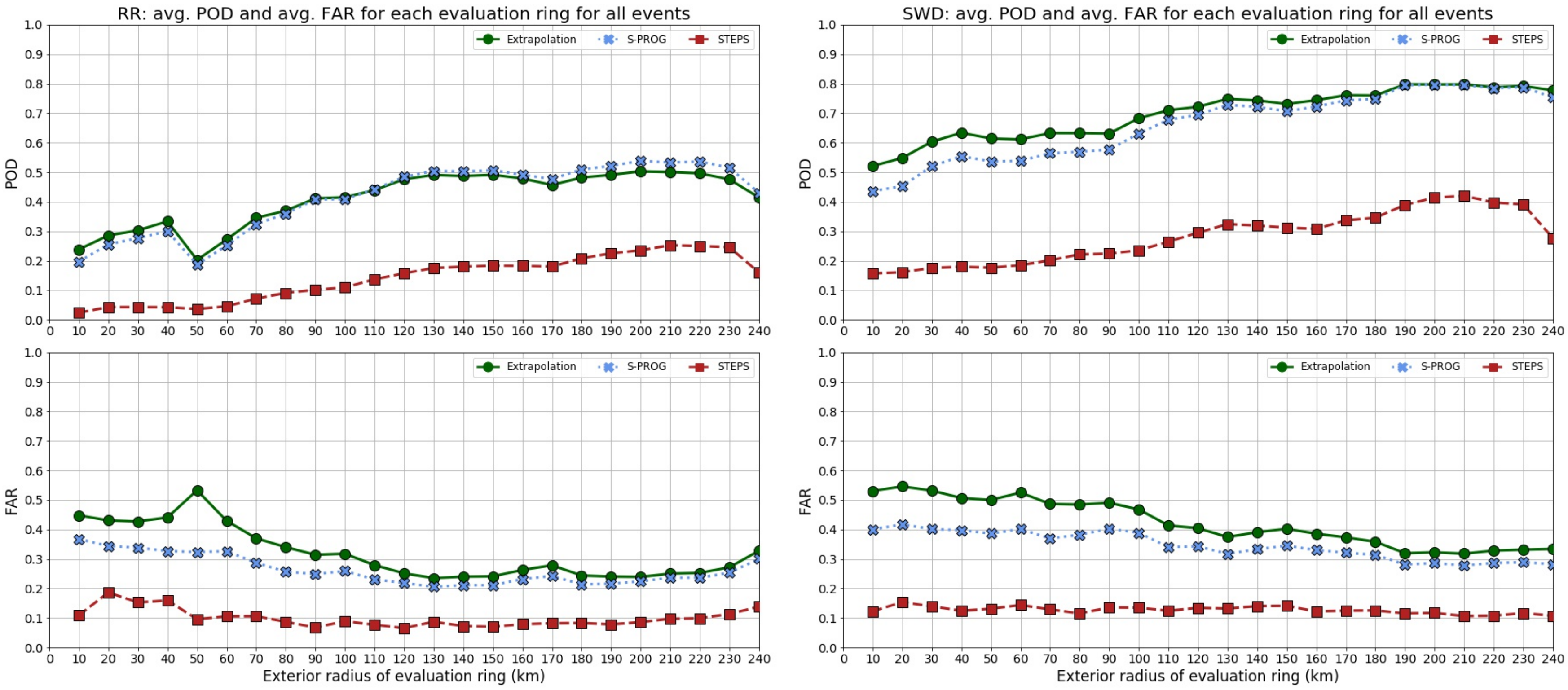
The results of Figures 5.25 and 5.26 and Table 5.1 demonstrated the existence of a relationship between the nowcast evaluation statics and the amount of real data (non-NaN) pixels in an image, with the nowcast having better results overall with a higher number of real data pixels in the input image. Notwithstanding, some points from the results need to be addressed. The first is that the ACC and the FA exhibited a direct relationship with the pixel fraction itself, not the log_{10} of the pixel fraction. This is because, as mentioned previously, the ACC and FA heavily depend on the correct negatives, which include the generally large amount of correctly predicted NaN values. Thus, the variables heavily depend on the NaN values in the images, which is basically one minus the pixel fraction.

The second point to address is the discrepancy in the relationship between the variables and the pixel fraction when using the SWD and using the radar data. The difference can be observed in the variation of coefficient values of Table 5.1 for each type of data, and more clearly in Figure 5.26. The inconsistency arises from the tendency of the nowcasts to generate more false alarms for SWD than for radar data because they are not design to predict the SWD. This causes a difference in the behavior of the statistics that depend on the false alarms when using SWD data and radar rain rate data.

5.5.6 Deeper evaluation of nowcast models with radar data

In general, the results obtained using radar rain rate data were less favorable than those obtained using SWD data. While it was shown in Section 5.5.5 that the amount of real data points in the image used as input influences the results, there are other factors that could potentially have an effect on the nowcast skill results. One of these factors is the radar's ability to see or accurately measure the entire region covered by the radar range.

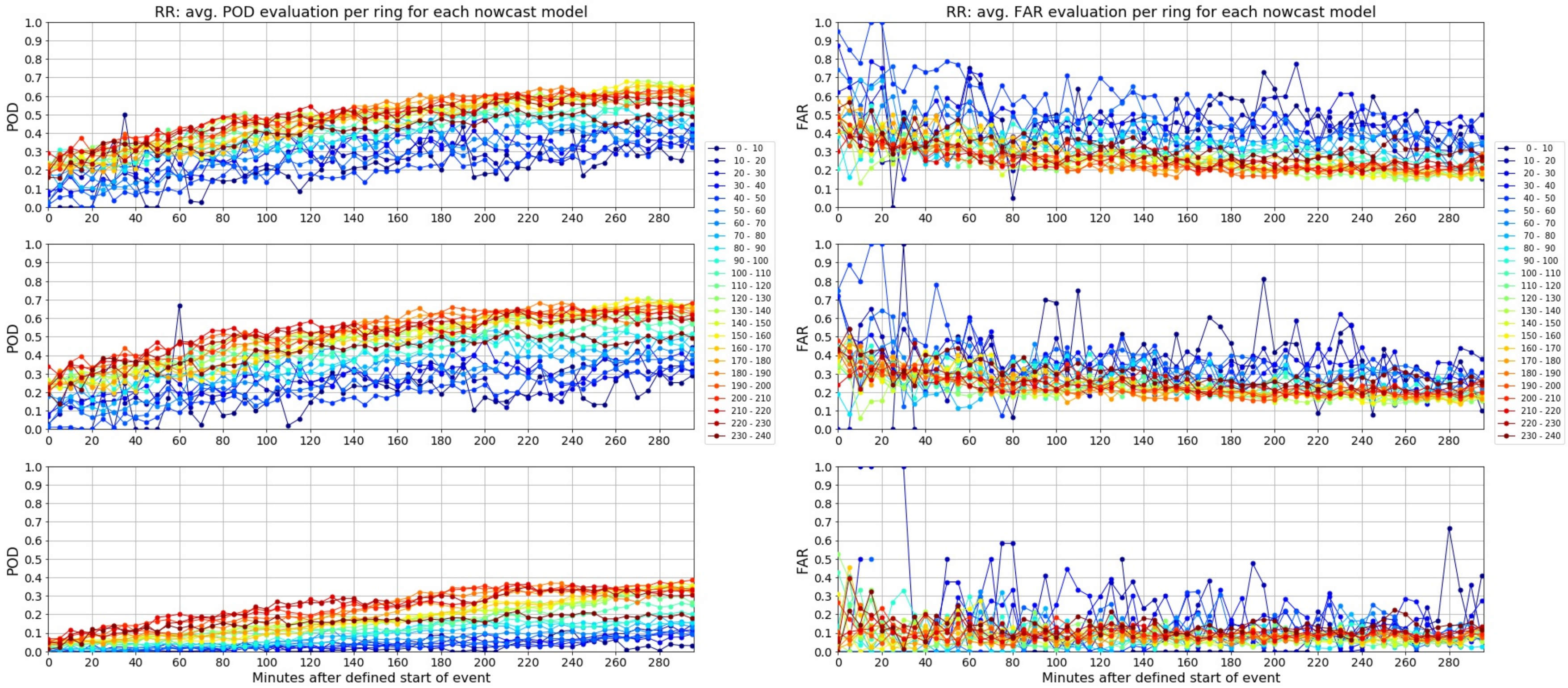
An initial assessment was to observe how the nowcast skill varied with radar range. This was done by dividing the radar coverage area into rings, 10 km in width, and obtaining the evaluation statistics for each ring. Figure 5.27a shows the average PoD and FAR per outer ring radius. For PoD, the average value was around 0.2 near the radar, and increased gradually until reaching 40 km, when the PoD decreased to slightly less than the initial value and continued with a gradual increase up to 0.5 at 120 km. This value was more or less constant up to 220 km, then there was slight decrease to 0.4 in PoD for the outer-most radii. FAR the opposite behavior, starting near 0.4 - 0.45 and maintaining these values until 50 km, where a peak occurred and followed by gradual decrease to around 0.2 - 0.3 at 120km, remaining constant up to 220 km, and ending with a slight increase. Additionally, there was larger difference between SPROG and extrapolation nowcast of around 0.05 for all radii up to 170 km. Figure 5.28 show the average PoD and FAR for rain rate per ring per time after the start of the events, for each nowcast model. These graphs also show that the PoD was low for rings closest to the radar and increased as the distance away from the radar increases, and the opposite occurs for FAR. Additionally, the general average behavior is observed for all distances: the PoD tended to increase with time and the FAR decreased with time. In general, STEPS results were similar but approximately 0.2 lower.



(a)

(b)

Figure 5.27: Average PoD (top) and average FAR (bottom) for (a) radar rain rate results and (b) SWD results per outer radius of the evaluation rings for all nowcast models.



(a)

(b)

Figure 5.28: The change of (a) average PoD and (b) average FAR for the rain rate over the period of analysis for each evaluation ring. Top graphs are for Extrapolation model, middle graphs are for the S-PROG model and bottom graphs are for the STEPS model.

The graphs indicate consistent performance of the nowcast models between 100 km and 220 km, encompassing the 180 km radius where the radar has optimal visibility without obstructions, as detailed in Section 5.5.4. However, the accuracy of the nowcast diminishes both closer to and farther away from the radar. The reduced accuracy near the radar might be attributed to fewer pixels within these ranges, resulting in lower skill values, as previously observed. Conversely, the farthest radii encompassed the largest number of pixels, but radars typically experience reduced accuracy at greater ranges, leading to fewer pixels capturing actual rain rate measurements and subsequently lowering forecast skill.

5.5.7 Evaluation of 30-minute and 60-minute forecast

The nowcast evaluations of the previous sections use the 15-min forecast because it is considered the shortest lead time that could be used in an operational manner. However it would be more ideal to use a larger lead time, particularly for the SWD. The reason for this is that it was observed that the SWD decreased below 2.5 K 60 minutes prior to the development of rainfall, leading to conclusion that the SWD has potential as a marker for future precipitation. For this reason, an analysis was also done for 30- and 60-min forecast by comparing their results with those of the 15-min forecast.

Figure 5.29 illustrates the average SWD PoD and FAR for the analysis period, considering the 15-, 30-, and 60-minute forecasts for each nowcast model. The graphs distinctly display that the 15-minute forecast exhibited greater accuracy compared to the 30- and 60-minute forecasts. In the case of Extrapolation and S-PROG, the PoD differences between the 15- and 30-minute forecasts were 0.15 at the beginning of the evaluation period, decreasing to 0.1 by the end. When comparing the 15-minute and 60-minute forecasts, the PoD was over 0.3 higher for the longest lead time, diminishing to just above 0.2 by the evaluation period's conclusion. In terms of FAR, it was lower for the 15-minute forecast by 0.1 or less compared to the 30-minute forecast and by 0.2 or less for the 60-minute forecast, with the differences decreasing over time.

In the case of STEPS, the differences in PoD between the 15-minute forecast and the other two lead times increased over time, particularly for the 60-minute forecast, rising from 0.25 to 0.4 throughout the entire analysis period. While the STEPS FAR did not vary considerably between the 15- and 30-minute forecasts, it exhibited more erratic behavior at the beginning of the analysis period for the 60-minute forecast.

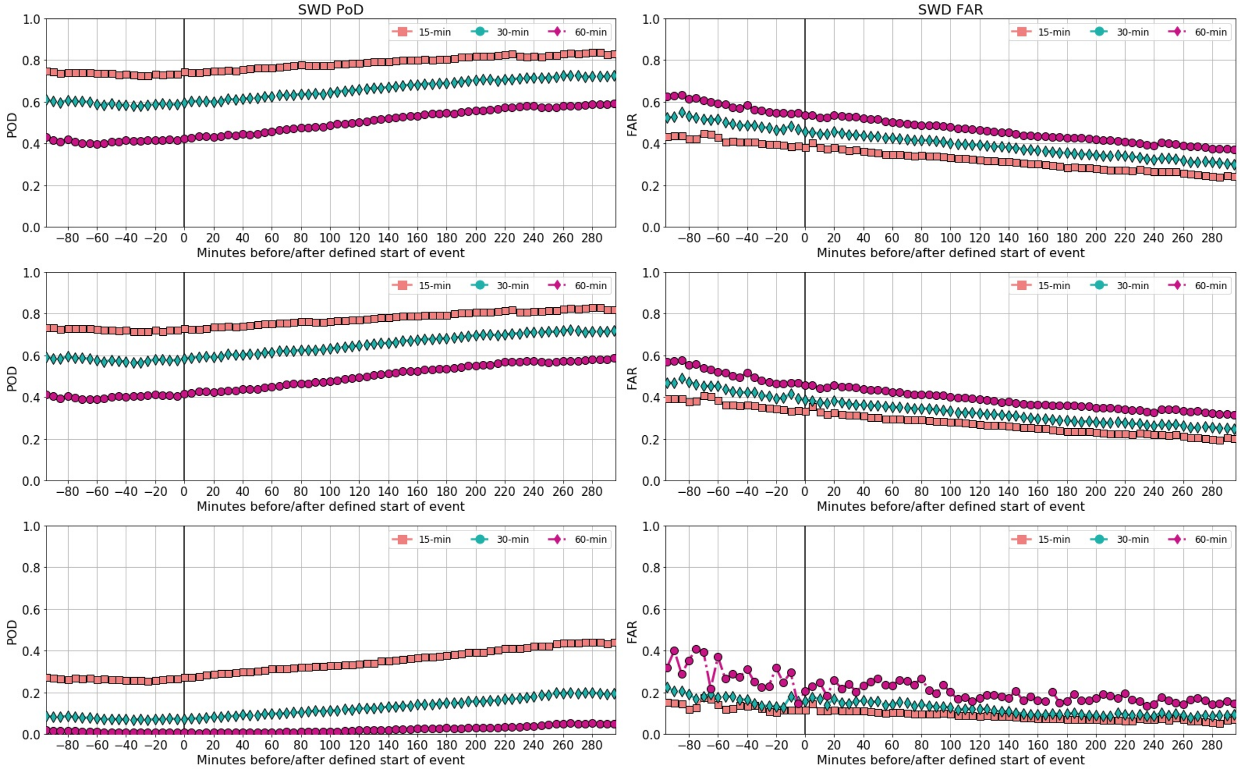


Figure 5.29: Average PoD and FAR for the SWD per nowcast model for the 15-, 30- and 60-min forecast. Top graphs are for Extrapolation, middle graphs for S-PROG and bottom graphs for STEPS.

Figure 5.30 shows the rain rate average PoD and FAR for the three lead times analyzed during the period of evaluation. The changes were slightly lower to those observed for the SWD, varying by 0.05 or less, and the change was more constant thought the whole time period. The only exception was for the STEPS results, where the difference in PoD decreased over time, and the FAR results for the 60-min forecast were very unstable for the majority of the period of analysis. The general decrease in skill of the models with respect to the lead time was also observed by Pulkkinen et al. (2020), Cuomo and Chandrasekar (2021), and Niu et al. (2021) for the S-PROG and the Extrapolation models.

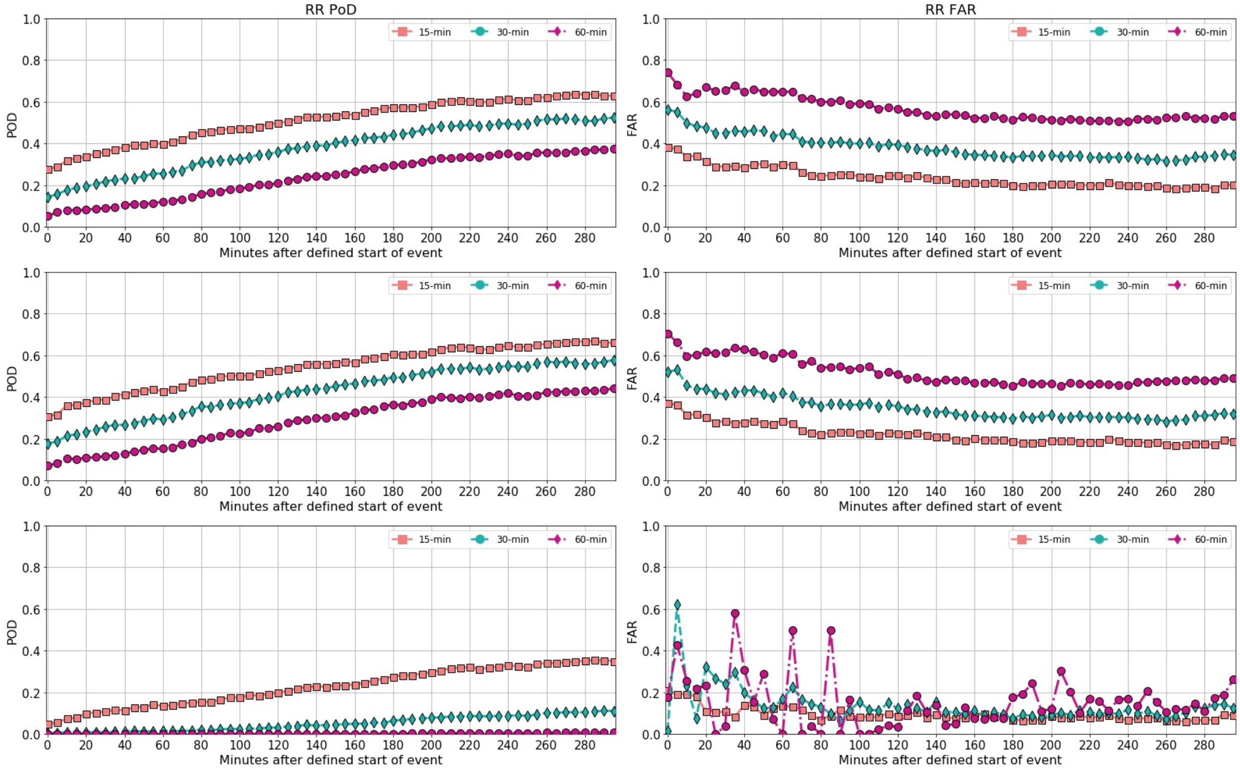


Figure 5.30: Same as Figure 5.29 but for rain rate.

5.5.8 Evaluation of the intensity of rain rate forecast by the nowcast models

The previous sections presented the nowcast evaluations using dichotomous verification statistics, which were used because the intensities of the satellite products are not expected to be correctly predicted by the nowcast models developed for radar data. Nevertheless, to complete the evaluation of the nowcast models for the rain rate, a brief analysis on how the nowcast models predict the intensity of the rain rate provided by the Queretaro weather radar was conducted. This section only evaluated the Extrapolation and S-PROG models since they were the most skilled models according to the dichotomous statistics results.

Figure 5.31 shows the average mean error (ME), the MAE and the root mean standard error (RMSE) for each model, all obtained through the pySTEPS functions, for the 15-min forecast obtained every five minutes for the 295 minutes after the start of the events. The ME indicates that the models tended to underestimate (on average) the rain rate for the first 165 minutes. The MAE showed that the mean magnitude of the errors was generally below 4.5 mm/hr for Extrapolation and decreased to around 3 mm/hr at the end of the period of analysis, indicating that the errors in intensity tended to decrease as the events grew and evolved. A similar behavior was observed for S-PROG, but the errors were around 0.5 mm/hr higher. Finally, the RMSE for Extrapolation started around 8 mm/hr, and increased to approximately 11 mm/hr at 160 minutes after the start of the events, followed by a decrease to around 9 mm/hr by the end of the analysis. A similar pattern was observed for S-PROG with an increase of 0.5 mm/hr. The results in comparison with the work of Pulkkinen et al. (2020) are far better. Pulkkinen et al. (2020) analyzed 10 events in Texas and obtained an MAE of between 11 mm/hr and 12 mm/hr on average. Additionally, Cuomo and Chandrasekar (2021) tested both S-PROG and Extrapolation among other models and found that the MAE was higher for S-PROG than Extrapolation as well, although Cuomo and Chandrasekar (2021) results were obtained comparing radar reflectivity and were presented as an average for various minimum reflectivity thresholds.

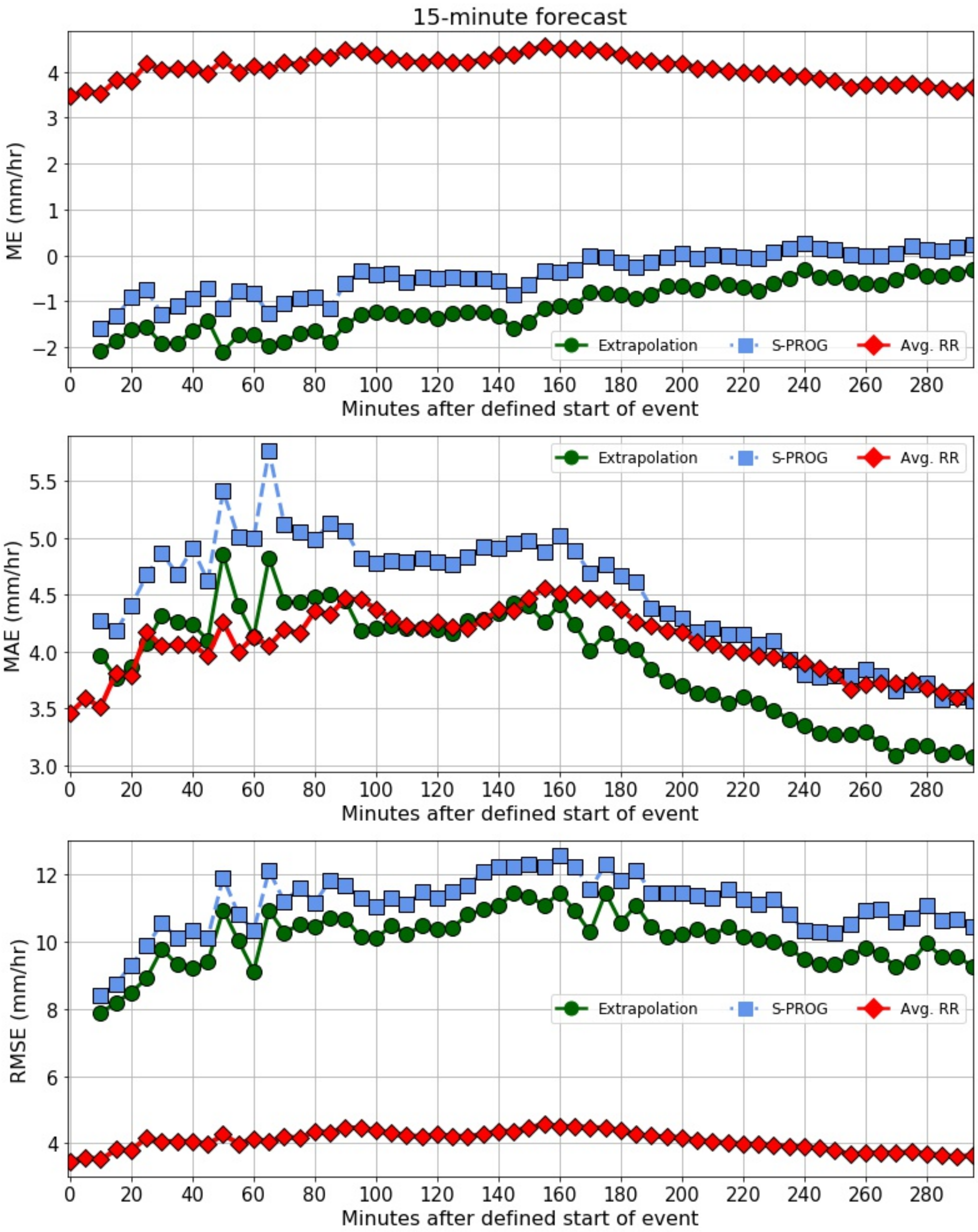


Figure 5.31: Average ME (top), MAE (center) and RMSE (bottom) for each 15-min forecast, obtained using the images in every 5-minute interval as input image, for the Extrapolation and S-PROG nowcast in the 295 minutes after the start of the events. The red line indicates the average rain rate observed in that interval.

Figure 5.32 shows the difference in the average of the mean rain rate for the Extrapolation, S-PROG and the observed rain rates. It is clear from these images that the models underestimated rain rate, between 2.0 mm/hr and 1.2 mm/hr for Extrapolation, and between 1.4 mm/hr and 0.6 mm/hr for S-PROG.

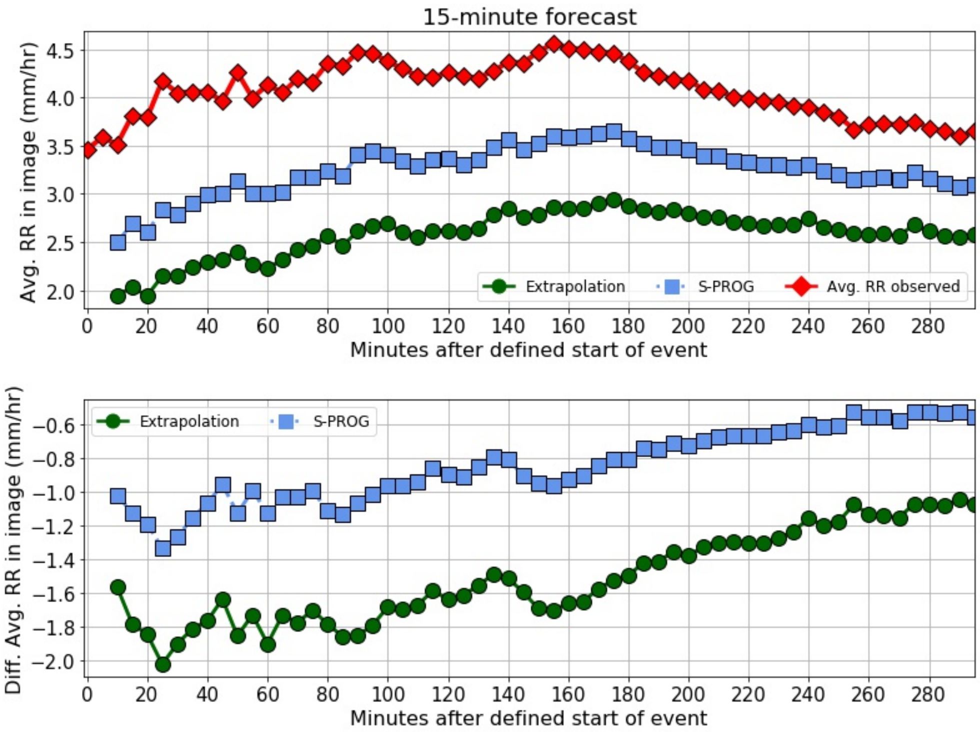


Figure 5.32: Average mean rain rate for the output image of the 15-min forecast, obtained using the images in every 5-minute interval as input image, for the Extrapolation and S-PROG nowcast in the 295 minutes after the start of the events; and the average mean rain rate observed in the interval.

Finally, similar to the dichotomous results, the forecast was less accurate in predicting rain rate intensity for the 60- and 30-min forecast. Figure 5.33, shows the average ME, MAE and RMAE for the 15-, 30- and 60-min forecast per model. The ME generally increased with lead time for both models. On the other hand, the MAE oscillated more with higher lead times in the first 80 minutes after the start of the event, with the 60-min forecast at times being lower than the 15-min forecast. Afterwards, the 30-min forecast had generally higher MAE than the other forecast for both models, and the 60-min forecast had MAE similar or higher than the 15-min forecast. S-PROG had generally higher MAE than Extrapolation for all three lead times. RMSE for Extrapolation was similar between the 15-min and 30-min forecast, but the 60-min forecast had lower RMSE in the first 120 minutes. For S-PROG, the RMSE was also lower for the 60-min forecast in the first 70 minutes, however, the 15-min forecast had the lowest RMSE afterwards. The results showed that the 60-min forecast on average can best predict the intensity of the precipitation at the beginning of the events, and as the events grew and evolved, the 15-min forecast became the most accurate. However, the errors obtained for all lead times are still close to or higher than the observed average rain rate. In general, the observed increase in the statistics with lead time was also observed by Pulkkinen et al. (2020); Cuomo and Chandrasekar (2021).

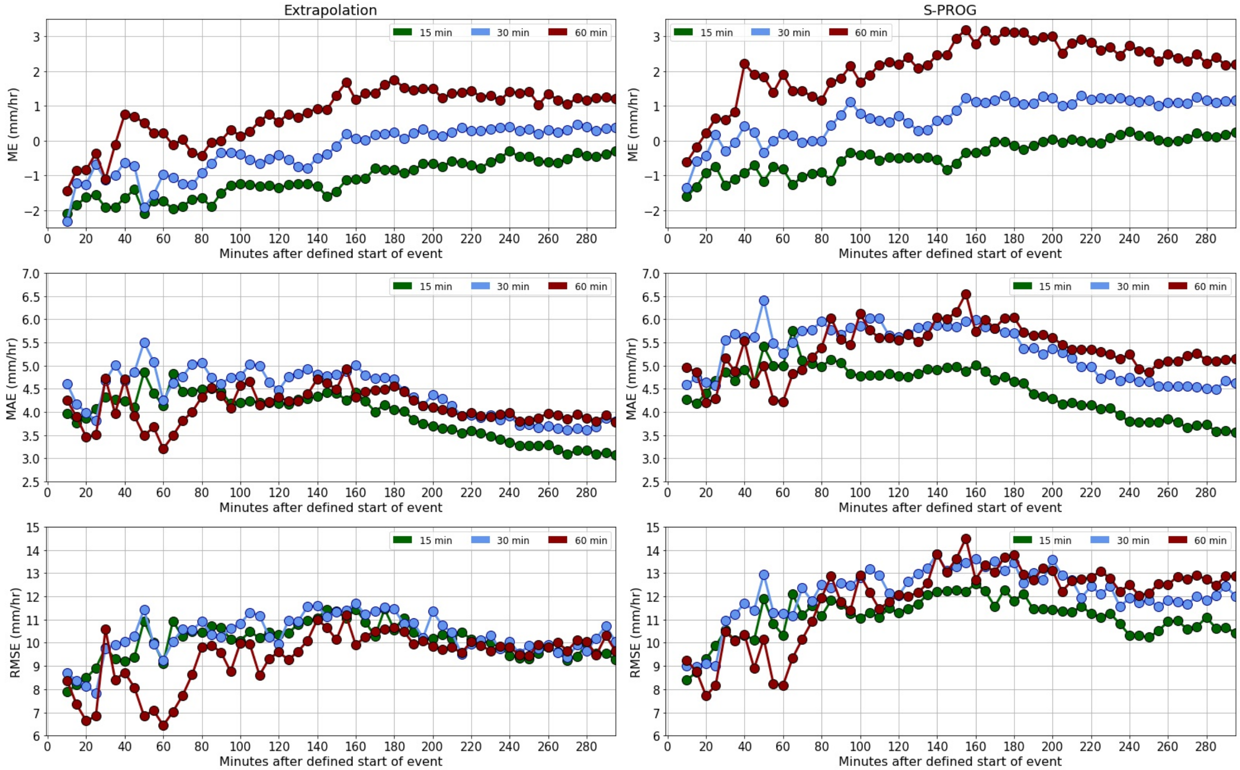


Figure 5.33: Average ME (top), MAE (center) and RMSE (bottom) for the 15-, 30-, and 60-min forecasts, obtained using the images in every 5-minute interval as input image, for the Extrapolation and S-PROG nowcast in the 295 minutes after the start of the events. The red line indicates the average rain rate observed in that interval.

5.6 Discussion

The results of the evaluation of nowcast models suggest that the best models to use with the GOES-16 SWD data are the Extrapolation and S-PROG models, while STEPS resulted in a large number of false alarms, and lower probability of detection (PoD) values. The most likely explanation is the addition of the stochastic element to both the advection field and the input variable fields added by the model. Additionally, the results showed that STEPS overestimated the development of precipitation around the actual precipitation clusters, decreasing the overall accuracy of the predictions. However, why this occurs is not clear. This could be due to the fact that the prediction evaluated from STEPS was the average of all ensemble members, but further analysis with individual members must be performed in order to test this hypothesis, as well as to analyze how well the models predict the intensity of the rainfall.

In general, the 15-minute forecast resulted in good accuracy, with the accuracy on average being above 60% within the 90 minutes before and two hours after the start of a precipitation event. Additionally, the DATing module from pySTEPS can also be used to detect and track large clusters of the SWD data, albeit with some missed medium-size clusters. And because of how it was developed and how it functions, the results of the 15-minute forecast from the nowcast models when using the detected and tracked SWD clusters were far better than when using the entire SWD field available. As mentioned in the results, the pySTEPS library was used with satellite data with a primary purpose of implementing the optical flow algorithms (Burton et al., 2022). But the SWD data PoD and false alarm rate (FAR) results are consistent with previous studies (Pulkkinen et al., 2020; Cuomo and Chandrasekar, 2021; Niu et al., 2021), that used radar data for the evaluations of the Extrapolation and S-PROG models. However, a direct comparison cannot be made, given the different characteristics between SWD and rain rate.

As previously mentioned, the 15-minute forecast is considered the shortest lead time that could have some future operational use, and the results obtained with this forecast were considered good. However, because of the observation made in Section 4.4.2 of a decrease of SWD, representative of an increase in low-level water vapor, being a possible marker of future convective precipitation occurring within one hour, it is more ideal to use larger lead times. Considering this result, a 30-minute or a 60-minute forecast would be even more useful when using the SWD. For this reason, the evaluation analysis was also performed using the 30-minute forecast and the 60-minute forecast. The general patterns observed with the 15-minute forecast were also present with the 30- and 60-minute forecast. This includes the differences in the accuracy between using all the SWD field below 2.5 K and only using the detected and tracked SWD clusters, the similar results between the Extrapolation and S-PROG models, and the STEPS model having the least accurate results for both rain rate and SWD. However, the 30-minute forecast was on average 15% less accurate than the 15-minute forecast and the 60-minute forecast was around 30% less accurate, as expected since larger lead times have less accurate forecast. Although these forecasts were less accurate than the 15-minute forecast, they can still be useful. Applying the methodologies established in this study, the nowcasts have the potential to predict the areas of possible convective precipitation one hour prior to the event with an accuracy of around 40% and 30 minutes prior with an accuracy near 50%. Nevertheless, further research is necessary to enhance and refine the forecasts.

In addition, although it was observed that the SWD fell below 2.5 K around 60 to 90 minutes before the rainfall event, not all areas of SWD that fell below 2.5 K developed into precipitation. Meaning, not all areas of high low-level water vapor developed into convective precipitation. These observations indicate that the SWD is not enough to precisely detect an area of future rain development. This study only used the SWD because it lacked any other tools and sources of data that could provide information on the atmospheric condition of the entire region of study. As mentioned in the Introduction, Mexico lacks a nationwide rain gauge system and a comprehensive meteorological station network, with the existing ones being very scatter throughout the country to provide the necessary spatial resolution for these nowcast systems to function efficiently. Moreover, various other satellite products were used in the characterization of the precipitation of the region in Chapter 4, and many of the products were not available for the region during the time of analysis or were missing data in areas of precipitation. The ones that were available did not have a clear relationship with rain rate or the development of rainfall like the SWD did. Regardless, once the data sets for the other GOES-16 products become available for the region, future research can use these satellite products to further examine the precipitation that occurs in the region with the radar data, as well as determine more precise locations for future precipitation along with the SWD.

Continuing with SWD, the evaluation statistics used were purposely chosen as dichotomous evaluators, only analyzing whether the forecast correctly predicted the presence of the SWD and disregarding the intensity of the forecast SWD. This was done because the models were designed to predict radar derived rain rate and reflectivity, rather than satellite products. As highlighted in Section 4.4.2, the SWD decreases before the measurement of rain. Because of this, if the models are predicting a general dissipation of rainfall by decreasing the rain rate, they can cause an apparent increase in the convection development when used with SWD values. It is for this reason, all other nowcasts available in the pySTEPS library were not used for this work. The Extrapolation and S-PROG models yielded satisfactory results due to the absence of factors in their algorithms that directly modify the intensity of the input variable, such as a dissipation factor. Additionally, S-PROG and Extrapolation yielded similar results, indicating that the spectral decomposition of the spatial scale used in S-PROG does not improve or modify the position forecast when using SWD, although a modification in the intensity was observed in the SWD forecast. On the other hand, STEPS, which modifies the intensity of the input variable, resulted in less accurate forecast, both in position and intensity.

Regarding the radar rain rate data, for the convective precipitation that occurs during the rainy season, the Extrapolation and S-PROG models demonstrated satisfactory performance, showing minimal disparity between using the entire rain rate field and solely considering the data within the SWD detected and tracked clusters. However, the STEPS model resulted in far less accurate forecast. These results were consistent with Cuomo and Chandrasekar (2021), but they observed a smaller difference between the three models. In addition, there was a significant increase in the nowcast skill as the events developed. In general, the forecast obtained using the radar rain rate data was also less accurate than the forecast from SWD data, with the PoD being between 0.2 and 0.4 smaller for rain rate data. Conversely, the FAR was around 0.1 smaller for rain rate data. The slightly better results obtained for FAR were mainly due to the nowcast models tendency to create more false alarms around the SWD data than the rain rate data, effecting statistics that depend false alarms, including FAR. Comparing the findings of this study with the earlier works of Pulkkinen et al. (2020), Cuomo and Chandrasekar (2021) and Niu et al. (2021), the PoD and FAR values did not significantly deviate for developed events. Although the PoD values were significantly worse than those observed in both studies at the beginning of the events, as the events developed, the PoD improved, coming within approximately 0.1 of the values observed in the other studies. Additionally, it is also important to note that the studies did not use the same threshold rain rate value that indicate rain and non-rain conditions.

Regardless, it is clear that the models were less skilled in predicting rain rate than the SWD. This discrepancy can be partly attributed to the higher presence of real or non-NaN SWD data per data matrix compared to the rain rate data. As demonstrated in Section 5.5.5, a positive correlation exists between the quantity of pixels with real data and most of the statistics used. This correlation arises due to the increased impact of any outliers on the results in the instance of a reduced number of actual data points within an input image. Throughout the majority of event analyzed, the SWD data greatly outnumbered the rain rate data, contributing to the superior statistical results observed for the SWD. This was mainly because the areas where there was a high quantity of low-level water vapor (SWD below 2.5 K) were larger in comparison to the size of the convective systems that occurred in the region, which were often less than 50 km in size. In addition, the observation that the forecasts's skill improved with time can be attributed to the fact that the size of the precipitation events grew with time, increasing the non-NaN pixels in the input image. It was calculated that an image with at least 420 pixels (out of 58,880 total pixels) with rain rate values above the set noise threshold of 0.5 mm/hr was necessary to obtain a skilled forecast. Another reason the radar rain rate results were not as good as the SWD results is the radar's lack of consistent accuracy throughout the radar coverage area, whether this is due to blockages from mountains or radio antennas or attenuation. For the Queretaro weather radar, most of the echo blockage occurs beyond 180 km radius. The effects of this were seen with the statistics with respect to the distance away from the radar were obtained, and the PoD and FAR showed the nowcast models' skill changed with the distance. In particular, the statistics were lower for distances close to the radar and at the edge of the radar coverage area. While the area between had consistent results on average for all three models evaluated.

Finally, in regard to the evaluation of how the nowcast predict the intensity of the rain rate, the results were not very favorable. This part of the analysis was performed only for the Extrapolation and S-PROG models since the STEPS forecasts were the least skilled with the dichotomous statistics. The mean absolute error (MAE) showed the average error was less than 5.5 mm/hr, the root mean square error (RMSE) was above 8 mm/hr and the Extrapolation nowcast had smaller average values than S-PROG. However, the MAE and RMSE values were higher than the average observed mean rain rate (excluding 0.0 mm/hr measurements), although MAE were similar to those of observed by Cuomo and Chandrasekar (2021) and far better than the results of Pulkkinen et al. (2020), whose MAE was twice as high. Additionally, the errors decreased as the events developed and grew. Comparing the average mean rain rate of the models and the averaged observed measurements, the results showed the models underestimated the rain rate. These results were not expected or ideal, especially because it was assumed that the spectral decomposition of spatial scales and autoregression of the S-PROG would result in better intensity results than the simple Extrapolation through advection model, as observed in Seed (2003). However, when considering all other results, the values obtained were more consistent with other previous studies (Pulkkinen et al., 2020; Cuomo and Chandrasekar, 2021; Niu et al., 2021) for developed events, or approximate 120 minutes after the start of the events. Nonetheless, because the average mean rain rate was underestimated in a consistent manner for both models point to a possible path for resolution to the issue. Thus, it is crucial to continue studying these nowcast models for this region, to assess why these models are not performing well in forecasting the intensity of precipitation and how they could be improved.

Chapter 6

Summary and conclusions

Due to the large amount of results obtained in this work, Figure 6.1 presents a flowchart of all the important results obtained in this study.

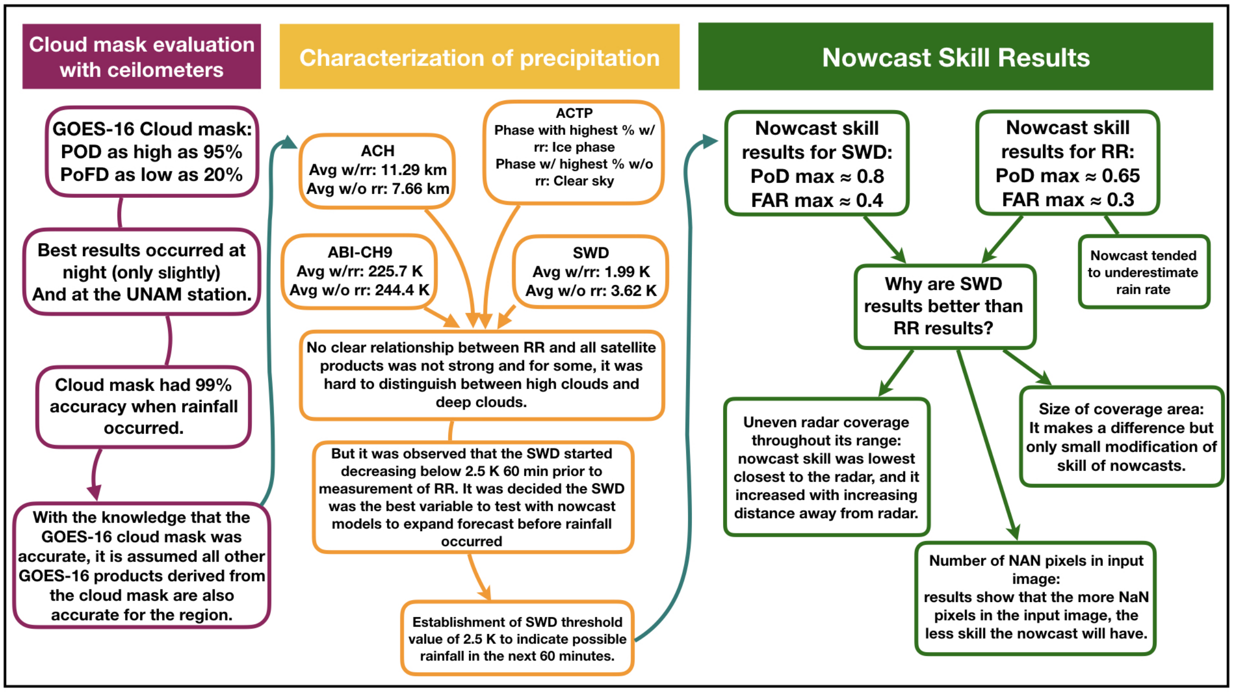


Figure 6.1: Flowchart of the key results obtained in the three main investigations of this study.

The implementation of operational nowcast systems has brought great benefit in precipitation-prone regions around the world. However, the advantages that these short-term forecast provide have not been enjoyed by Mexico due to the lack of investigation of such systems. It is for this reason that this study explored the use of nowcasts by evaluating the precipitation that occurs in the region of central Mexico, using satellite and radar data, as well as assessing the skill of various nowcast models for the region. The investigation mainly used the GOES-16 satellite data, specifically the Level 1 ABI channel 13 and 15 data, the Level 2 Clear-Sky mask (ACM), cloud-top height (ACH) and cloud-top phase (ACTP) data, and the Queretaro weather radar rain rate data.

The first step taken was a quality control evaluation of the GOES-16 cloud mask, a product that uses Level 1 products and is the base of several Level 2 products. The evaluation used the RUOA network's ground ceilometers and pluviometers to evaluate the cloud mask with various methodologies to decrease or measure the uncertainties associated with the differences in spatial and temporal resolution between the satellite data and the ground-base instruments' data. The results showed when considering the most consistent and reliable ceilometer data, the cloud mask has an accuracy between 69%-85%, a probability of detection of 0.81 - 0.95, a probability of false detection of 0.37 - 0.40 and a skill score of around 0.39 - 0.62, and additionally, the cloud mask had a $> 99\%$ precision in detecting clouds when precipitation is occurring This leads to the conclusion that the cloud mask has good overall accuracy in cloud detection. These results allowed the use of the GOES-16 satellite products without having any questions or doubts regarding their reliability.

The next step was to obtain the characterization of the precipitation that occurs in the study region using the GOES-16 satellite products, with the Queretaro radar data. The GOES-16 products were the cloud-top height, cloud-top phase and the Split-Window Difference (SWD) calculated from the ABI channel 13 and 15. The results found that the SWD or split-window difference, on average, had a value of 1.99 K during rainfall. Furthermore, it was observed that, on average, the SWD tended to fall from around 3.8 K to 2.28 K in the 240 minutes prior to the measurement of rainfall and from 2.5 K to 2.28 K in the 60 minutes prior to the measurement of rainfall. This change in SWD showed how the low-level water vapor tended to increase prior to the development of convective precipitation. These results lead to the establishment of the threshold value of 2.5 K minimum, that indicates a possible development of rainfall within the next 60 minutes. For the cloud-top height or cloud height, the average with the presence of radar rain rate was 11.29 km and without radar rain rate was 7.66 km. In addition, this variable tended to increase around 0.5 km in the 90 minutes prior to the occurrence of rain. The cloud-top phase or cloud-top phase results showed a change in phase from warm water phase to ice phase for around 29% of the area, with the occurrence of rain rate. And, they also showed the growth and deepening of clouds to heights with temperatures low enough for water vapor to freeze, prior to the measurement of rainfall. Although the cloud-top height and cloud-top phase observed behaviors associated with the development of convective precipitation that occurs in the region of study, they could not distinguish between high-altitude cirrus clouds and deep convective clouds by themselves, making them less ideal for further use in the evaluation of the nowcast models.

The third and final part of the study was the evaluation of nowcast models in the region. Using the 2.5 K threshold established, masked SWD fields were used in three nowcast models provided by pySTEPS (Extrapolation, S-PROG, and STEPS) to evaluate how this variable performed when used with nowcast models developed to be used with radar data, using the entire SWD field and only using the large clusters detected when using the pySTEPS DATing module. The results showed that the S-PROG and extrapolation models, both of which did not directly modify the intensity of the input data, performed better than the STEPS, the model evaluated that did modified input data. The performance also improved on average as the start of the rainfall event got closer and as the event developed. Additionally, the DATing module algorithm did miss several clusters, which made the overall results of the evaluation with the SWD clusters less reliable than when using all SWD data available.

When evaluating the same models but using radar data, the results showed less accurate forecast than those obtained with SWD data, however the skilled improved with the development and growth of the precipitation events. This was partly caused by the fact that the radar did not have consistent accuracy over all its coverage area, and the SWD fields had more real data (non-NaN values) throughout the field. A decrease in the amount of real data available led to more pronounced effects of the errors and outliers on the statistics used. Furthermore, the models' ability to accurately predict the intensity of rain rate was not high, with the models underestimating average rainfall and Extrapolation model having more skill in predicting intensity than the S-PROG model.

Finally, in the evaluation of the same models with both data sets but expanding the forecast time to 30 minutes and 60 minutes, the results were consistent with those obtained for the 15-minute forecasts, but they were less accurate. For the 30-minute forecast, the accuracy was around 20% lower and the 60-minute forecast accuracy was approximately 40% lower.

In conclusion, the extrapolation and S-PROG models give a skilled 15-minute, 30-minute, and 60-minute prediction with both radar and satellite data for the region of central Mexico during the rainy season. Additionally, it was observed that the SWD can be used to detect areas of possible future convective precipitation. However, the SWD alone was not enough to have high precision in the prediction of rainfall, and the addition of other measurements, such as wind or surface measurements, is necessary to improve the detection of areas of possible rainfall. Lastly, the methodologies used in this work can be used as a foundation to explore the use of other GOES-16 satellite products in the nowcast models and functions provided by the pySTEPS library. Further research to this study includes the improvements of the predictions by the nowcast models, especially for the satellite products, as well as the exploration of using other satellite products, once they become available, in the nowcast models. Nevertheless, the operational implementation of these models can still be achieved with the help of the methodologies used in this study, if the satellite and radar data is measured and processed in less than 10 minutes.

Bibliography

1. (2015). *User's Guide: Vaisala Ceimeter CL31*. Vaisala Oyj.
2. (2020). *Ceimeter CL31 for cloud height detection Data Sheet*. Vaisala Oyj.
3. (2020). *Documentation: ALCF output*. Automatic Lidar and Ceilometer Framework.
4. (2020). *Texas Technology Rain Gauge Tipping Bucket TR-525M Rainfall Sensor Spec Sheet*. Texas Electronics Inc.
5. Ackerman, S. A., Holz, R. E., Frey, R., Eloranta, E. W., Maddux, B. C., and McGill, M. (2008). Cloud Detection with MODIS. Part II: Validation. *Journal of Oceanic Technology*, 25:1073 – 1086.
6. Adams, D. K., Gutman, S., Holub, K., and Pereira, D. (2013). GNSS observations of deep covective time scales in the Amazon. *Geophysical Reaserch Letters*, 40(11):2818–2823.
7. Barnes, L. R., Schultz, D. M., Gruntfest, E. C., Hyden, M. H., and Benight, C. C. (2009). CORRIGENDUM: False alarm rate or false alarm ratio? *Weather and Forecasting*, 24(1452 - 1454).
8. Benz, A., Chapel, J., Jr., A. D., Birckenstead, B., Tillier, C., Nilsson III, W., Colon, J., Dence, R., Jr., F. H., and Campbell, P. (2019). *GOES-R Series Data Book*. Lockheed Martin and NASA GOES-R Series Program Office, Greenbelt, Maryland.
9. Bianchi, B., van Leeuwen, P. J., Hogan, R. J., and Berne, A. (2013). A variation approach to retrieve rain rate by combining information from rain gauges, radars, and microwave links. *Journal of Hydrometeorology*, 14:1897 – 1909.
10. Bouguet, J.-Y. (2001). Pyramidal implementation of the affine Lucas Kanade feature tracker description of the algorithm. *Intel Corp*, 5(4).

11. Bowler, N. E., Pierce, C. E., and Seed, A. W. (2006). STEPS: a probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Quarterly Journal of the Royal Meteorological Society*, 132(620):2127 – 2155.

12. Browning, K. A. and Collier, C. (1989). Nowcasting of precipitation systems. *Rev. Geophysics*, 27(3):345 – 370.

13. Burton, R. R., Blyth, A. M., Cui, Z., Groves, J., Lamptey, B. L., Fletcher, J. K., Marsham, J. H., Parker, D. J., and Roberts, A. (2022). Satellite-based nowcasting of West African mesoscale storms has skill at up to 4-h leat time. *Weather and Forecasting*, 37:445–455.

14. Cambier van Nooten, C., Schreurs, K., Wijnands, J. S., Leijnse, H., Schmeits, M., Whan, K., and Shapavalova, Y. (2023). Improving precipitation nowcasting for high-intensity events using deep generative models with balanced loss and temperature data: a case study in the Netherlands. *Artificial Intelligence for the Earth Systems.*

15. CEA (2022). Radar. url: https://www.ceaqueretaro.gob.mx/clima-y-pronostico/radar.

16. Chesters, D. and Uccellini, L. W. (1983). Lower-level water vapor fields from the VISSR Atmospheric Sounder (VAS) "Split Window" channels. *Journal of Climate and Applied Meteorology*, 22:725 – 743.

17. CIMSS (2018a). ABI Band 13 (10.3 μm) - Quick Guide. https://cimss.ssec.wisc.edu/goes/OCLOFactSheetPDFs/ABIQuickGuide_Band13.pdf.

18. CIMSS (2018b). ABI Band 15 (12.3 μm) - Quick Guide. https://www.star.nesdis.noaa.gov/goes/documents/ABIQuickGuide_Band15.pdf.

19. Cummings, R. J., Upton, G. J. G., Holt, A. R., and Kitchen, M. (2009). Using microwave links to adjust the radar rainfall field. *Advances in Water Resources*, 32(7).

20. Cuomo, J. and Chandrasekar, V. (2021). Use of deep learning for weather radar nowcasting. *Journal of Atmospheric and Oceanic Technology*, 24:1241–1261.

21. Delrieu, G., Caoudal, S., and Creutin, J. D. (1997). Feasibility of using mountain return for the correction of ground-based X-Band weather radar data. *Journal of Atmospheric and Oceanic Technology*, 14:368 – 385.

22. Dixon, M. (2007). Proceedings, 33rd Conference on Radar Meteorology. In *Intelligent mitigation of normal propagation and anomalous propagation clutter*. A. M. Society, ed.

23. Ehsani, M. R., Zarei, A., Gupta, H. V., Barnard, K., Lyons, E., and Behrangi, A. (2022). NowCasting-Nets: representation learning to mitigate latency gap of satellite precipitation products using convolutional and recurrent neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–21.

24. ESRI (2020). World_Terrain_Base [basemap]. Scale: 1:1m. "World Terrain Base". https://www.arcgis.com/home/item.html?id=c61ad8ab017d49e1a82f580ee1298931.

25. Foresti, L., Sideris, I. V., Nerini, D., Beusch, L., and Germann, U. (2019). Using a 10-year radar archive for nowcasting precipitation growth and decay: a probabilistic machine learning approach. *Weather and Forecasting*, 34:1547 – 1569.

26. Fornasiero, A., Alberoni, P., Amorati, R., Ferraris, L., and Taramasso, A. (2005). Effects of propagation conditions on radar beam-ground interaction: impact on data quality. *Advances in Geosciences*, 2:201 – 208.

27. Gabella, M. and Notarpietro, R. (2005). Ground clutter characterization and elimination in mountainous terrain. *ERAD*, pages 305 – 311.

28. Germann, U. and Zawadzki, I. (2002). Scale-dependence of the predictability of precipitation from continental radar images. Part I: description of the methodology. *Monthly Weather Review*, 130(12):2859 – 2873.

29. Grasso, L., Bikos, D., Dostalek, J. F., Wu, T.-C., Hilburn, K., Szoke, E., and Torres, J. (2020). Application of the GOES-16 Advance Baseline Imager: Morphology of a preconvective environment on 17 April 2019. *Electronic Journal of Severe Storms Meteorology*, 15(2):1–24.

30. Han, L., Zhang, J., Chen, H., Zhang, W., and Yao, S. (2022). Towards the predictability of a radar-based nowcasting system for different precipitation systems. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5.

31. Hannesen, R. (2001). Quantitative precipitation estimation from radar data—a review of current methodologies. *MUSIC European Commission Project, Deliverable*, 4(31).

32. Harrison, D. L., Driscoll, S. J., and Kitchen, M. (2000). Improving precipitation estimates from weather radar using quality con- trol and correction techniques. *Meteorological Applications*, 6:135 – 144.

33. Heidinger, A. (2012). Algorithm Theoretical Basis Document: ABI Cloud Height. Technical report, NOAA NESDIS Center for Satellite Applications and Research.

34. Heidinger, A. and Straka, W. C. (2013). *Algorithm Theoretical Basis Document: ABI Cloud Mask*. NOAA NESDIS Center for Satellite Applications and Research, 3 edition.

35. Heistermann, M., Jacobi, S., and Pfaff, T. (2013). Technical note: An open source library for precessing weather radar data (wradlib). *Hydrology and Earth System Sciences*, 17:863 – 871.

36. Hering, A. M., Morel, C., Galli, G., Ambrosetti, P., and Boscacci, M. (2004). In Proceedings of the Third ERAD together with the COST 717 Final Seminar. In *Nowcasting thunderstorms in the alpine region using a radar based adaptive thresholding scheme*, pages 206 – 211.

37. Hill, P. G., Stein, T. H. M., Roberts, A. J., Fletcher, J. K., Marsham, J. H., and Groves, J. (2020). How skillful are nowcasting satellite applications facility products for tropical Africa? *Meteorological Applications*, 27.

38. Hitschfeld, W. and Bordan, J. (1954). Errors inherent in the radar measurement of rainfall at attenuating wavelengths. *Journal of Meteorology*, 11(58 - 67).

39. Imhoff, R. O., Cruz, L. D., Dewettinck, W., Brauer, C. C., Uijlenhoet, R., van Heeringen, K. J., Velasco-Forero, C., Nerini, D., Ginderachter, M., and Weerts, A. H. (2023). Scale-dependent blending of ensemble rainfall nowcasts and numerical weather predictions in the open-source Pysteps library. *Quarterly Journal of the Royal Meteorological Society*, 149(753):1335–1364.

40. Imhoff, R. O., Overeem, A., Brauer, C. C., Leijnse, H., Weerts, A. H., and Uijlenhoet, R. (2020). Rainfall nowcasting using commercial microwave links. *Geophysical Reaserch Letters*, 47.

41. Jacobi, S. and Heistermann, M. (2016). Benchmarking attenuation correction procedures for six years of single-polarized C-band weather radar observations in South-West Germany. *Geomatics, Natural Hazards and Risk*, 7(6):1785 – 1799.

42. James, P. M., Reichert, B. K., and Heizenreder, D. (2018). NowCastMIX: Automatic integrated warnings for severe convection on nowcasting time scales at the German Weather Service. *Weather and Forecasting*, 33:1413 – 1433.

43. Jauregui, E. and Romales, E. (1996). Urban effects on convective precipitation in México City. *Atmospheric Environment*, 30(20):3383 – 3389.

44. Jimenez, P. A. (2020). Assessment of the GOES-16 Clear sky mask product over the contiguous USA using CALIPSO retrievals. *Remote Sensing*, 12(1630).

45. Jolliffe, I. T. and Stephenson, D. B. (2003). *Forecast Verification: A Practitioners Guide in Atmospheric Science*. Wiley.

46. Joro, S., Hyvärinen, O., and Kotro, J. (2010). Comparisons of satellite cloud masks with ceilometer sky conditions in Southern Finland. *Journal of Applied Meteorology and Climatology*, 49:2508 – 2526.

47. Joss, J. and Lee, R. (1995). The application of radar-gauge comparisons to operational precipitation profile corrections. *Journal of Applied Meteorology*, 34:2612 – 2630.

48. Kraemer, S. (2008). *Quantitative radar data processing for rainfall forecasting and urban drainage*. PhD thesis, Gottfried Wilhelm Leibniz University, Hannover, Germany.

49. Kraemer, S. and Verworn, H.-R. (2008). 11th International Conference on Urban Drainage. In *Improved C-band radar data processing for real time control of urban drainage systems*, Edinburgh, Scotland, UK.

50. Kuma, P., MacDonald, A. J., Morgenstern, O., Querel, R., Silber, I., and Flynn, C. J. (2021). Ground-based lidar processing and simulator framework for comparing models and observations (ALCF 1.0). *Geoscientific Model Development*, 14:43 – 72.

51. Kumar, A., Islam, T., Sekimoto, Y., Mattmann, C., and Wilson, B. (2020). Convcast: an embedded convolutional LSTM based architecture for precipitation nowcasting using satellite data. *PLoS ONE*, 15(3).

52. Laroche, S. and Zawadzki, I. (1995). Retrievals of horizontal winds from single-Doppler clear-air data by methods of cross correlation and variational analysis. *Journal of Atmospheric and Oceanic Technology*, 12(4):721 – 738.

53. Lebedev, V., Ivashkin, V., Rudenko, I., Ganshin, A., Molchanov, A., Ovcharenki, S., Grokhovetskiy, R., Bushmarinov, I., and Solomentsev, D. (2019). Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '19). In *Precipitation Nowcasting with Satellite Imagery*, pages 2680 – 2688, New York, NY, USA. Association for Computing Machinery.

54. Letu, H., Nagao, T. M., Nakajima, T. Y., and Matsumae, Y. (2014). Method for validating cloud mask obtained from satellite measurements using ground-based sky camera. *Applied Optics*, 53:7523 – 7533.

55. Lindsey, D. T., Bikos, D., and Grasso, L. (2018). Using GOES-16 Split Window Difference to detect a boundary prior to cloud formation. *Bulletin of the American Meteorological Society*, pages 1541 – 1544.

56. Lindsey, D. T., Grasso, L., Dostalek, J. F., and Kerkmann, J. (2014). Use of the GOES-R Split-Window Difference to diagnose deepening low-level water vapor. *Journal of Applied Meteorology and Climatology*, 53:2005 – 2016.

57. Lucas, B. D. and Kanade, T. (1981). Proceedings of the 1981 DARPA Imaging understanding workshop. In *An iterative image registration technique with an application to stereo vision*, pages 121 – 130.

58. Magaldi, A. V. (2013). *Synergies in rain detection using weather radar and SEVIRI based NWC-SAF products.* PhD thesis, University of Barcelona, Barcelona, Spain.

59. Magana, V., Perez, J., and Mendez, M. (2003). Diagnosis and prognosis of extreme precipitation events in the Mexico City Basin. *Geofisica Internacional*, 41(2):247–259.

60. Mandapaka, P. V., Germann, U., Panziera, L., and Hering, A. M. (2012). Can lagrangian extrapolation of radar fields be used for precipitation nowcasting over complex alpine orography? *Weather and Forecasting*, 27:28 – 49.

61. Marshall, J. S., Languille, R. C., and Palmer, W. (1947). Measurement of rainfall by radar. *Journal of Meteorology*, 4:186 – 192.

62. Marshall, J. S. and Palmer, W. (1948). The distribution of raindrops with size. *Journal of Meteorology*, 5:165 – 166.

63. Matrosov, S. Y., Cifelli, R., and Gochis, D. (2013). Measurements of Heavy Convective Rainfall in the Presence of Hail in Flood-Prone Areas Using an X-Band Polarimetric Radar. *Journal of Applied Meteorology and Climatology*, 52:395 – 407.

64. Mecikalski, J. R., MacKenzie Jr., W. M., Koenig, M., and Muller, S. (2010a). Cloud-top properties of growing cumulus prior to convective initiation as measured by Meteosat second generation. Part I: Infrared fields. *Journal of Applied Meteorology and Climatology*, 49:521 – 534.

65. Mecikalski, J. R., MacKenzie Jr., W. M., Koenig, M., and Muller, S. (2010b). Cloud-top properties of growing cumulus prior to convective initiation as measured by Meteosat second generation. Part II: Use of visible reflectance. *Journal of Applied Meteorology and Climatology*, 49:2544 – 2558.

66. Meischner, P. (2005). *Weather radar: principles and advanced applications.* Springer.

67. Meischner, P., Collier, C., Illingworth, A., Joss, J., and Randeu, W. (1997). Advance weather radar systems in Europe: The COST 75 Action. *Bulletin of the American Meteorological Society*, 78(7):1411 – 1430.

68. Morris, V. R. (2012). Vaisala Ceilometer (VCEIL) Handbook. Technical report, Atmospheric Radiation Measurement, U.S. Department of Energy Office of Science.

69. Niu, D., Huang, J., Zang, Z., Xu, L., Che, H., and Tang, Y. (2021). Two-stage spatiotemporal context refinement network for precipitation nowcasting. *Remote Sensing*, 13:4285.

70. Ochoa, A. C., Quintanar, A. I., Raga, G. B., and Baumgardner, D. (2015). Changes in intense precipitation events in Mexico City. *Journal of Hydrometeorology*, 16:1804–1820.

71. O'Hora, F. and Bech, J. (2005). Operational use of pulse compression in weather radar. In *Proceedings of the 32nd Conference on Radar Meteorology of the American Meteorolofical Society*, Alburquerque, N. M.

72. Pavolonis, M. J. (2010). *GOES-R Advanced Baseline Imager (ABI) Algorithm Theoretical Basis Document for Cloud Type and Cloud Phase*. NOAA NESDIS Center for Satellite Applications and Research.

73. Peralta, O., Adams, D., Castro, T., Grutter, M., and Varela, A. (2016). Mexico's university network of atmospheric observatories. *EOS*, 97.

74. Proesmans, M., van Gool, L., Pauwels, E., and (1994), A. O. (1994). *Lecture Notes in Computer Science*, volume 801, chapter Determination of optical flow and its discontinuities using non-linear diffusion, pages 294–304. Springer Berlin Heidelberg.

75. Pulkkinen, S., Chandrasekar, V., von Lerber, A., and Harris, A. M. (2020). Nowcasting of convective rainfall using volumetric radar observations. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1 – 15.

76. Pulkkinen, S., Nerini, D., Perez-Hortal, A., Velasco-Forero, C., Germann, U., Seed, A. W., and Foresti, L. (2019). pySTEPS: an open-source Python library for probabilistic precipitation nowcasting (v1.0). *Geosci. Model Dev.*, 12(10):4185 – 4219.

77. pySTEPS, D. (2021). pySTEPS references, release 1.4.1. Technical report.

78. Rahimi, A. R., Holt, A. R., Upton, G. J. G., Kraemer, S., Redder, A., and Verworn, H.-R. (2006). Attenuation calibration of an X-band weather radar using a microwave link. *Journal of Atmospheric and Oceanic Technology*, 23:395 – 405.

79. Ribeiro, B. Z., Machado, L. A. T., Huaman Ch., J. H., Biscaro, T. S., Freitas, E. D., Mozer, K. W., and Goodman, S. J. (2019). An evaluation of the GOES-16 Rapid Scan for nowcasting in Southeastern Brazil: Analysis of a severe hailstorm case. *Weather and Forecasting*, 34:1829–1848.

80. Ritvanen, J., Harnist, B., Aldana, M., Makinen, T., and Pulkkinen, S. (2023). Advection-free convolutional neural network for convective rainfall nowcasting. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:1654–1667.

81. Rivolta, G., Marzano, F. S., Coppola, E., and Verdecchia, M. (2006). Artificial neural-network technique for precipitation nowcasting from satellite imagery. *Advances in Geosciences*, 7:97–103.

82. Ruzanski, E., Chandrasekar, V., and Wang, Y. (2011). The CASA nowcasting system. *Journal of Atmospheric and Oceanic Technology*, 28(5):640 – 655.

83. Saadi, M., Furusho-Percot, C., Belleflamme, A., Troemel, S., Kollet, S., and Reinoso-Rondinel, R. (2021). Comparison of three radar-based precipitation nowcast for the extreme July 2021 flooding event in Germany. *Journal of Hydrometeorology*, 24:1241–1261.

84. Salek, M. and Szabo-Takacs, G. (2019). Comparison of SAFNWC/MSG satellite cloud type with Vaisala CL51 ceilometer-detected cloud base layer using the sky condition algorithm and Vaisala BL-View software. *Atmosphere*, 10(316).

85. Saltikoff, E. and Nevvonen, L. (2011). First experiences of operational use of a dual-polarization weather radar in Finland. *Meteorologische Zeitschrift*, 20(3):323 – 333.

86. Schmit, T. J., Griffith, P., Gunshor, M. M., Daniels, J. M., Goodman, S. J., and Lebair, W. J. (2017). A Closer Look at the ABI on the GOES-R Series. *Bulletin of the American Meteorological Society*, 98(4):681 – 698.

87. Schroedter-Homscheidt, M., Drews, A., and Heise, S. (2008). Total water vapor column retrieval from MSG-SEVIRI split window measurements exploiting the daily cycle of land surface temperatures. *Remote Sensing of Enviromnet*, 112:249 – 258.

88. Seed, A. W. (2003). A dynamic and spatial scaling approach to advection forecasting. *Journal of Applied Meteorology*, 42(3):381 – 388.

89. Sieglaff, J. M., Cronce, L. M., Feltz, W. F., Bedka, K. M., Pavolonis, M. J., and Heidinger, A. K. (2011). Nowcasting convective storm initiation using satellite-based box averaged cloud-top and cloud-type trends. *Journal of Applied Meteorology and Climatology*, (110 - 126).

90. Siewert, C. W., Koenig, M., and Mecikalski, J. R. (2010). Application of Meteosat second generation data towards improving the nowcasting of convective initiation. *Meteorological Applications*, 17:442 – 451.

91. Simpson, M. J. and Fox, N. I. (2018). Dual-polarized quantitative precipitation estimation as a function of range. *Hydrology and Earth System Sciences*, 22:3375 – 3389.

92. Sun, J., Xue, M., Wilson, J. W., Zawadzki, I., Ballard, S. P., Onvlee-Hooimey, J., Joe, P., Baker, D. M., Li, P., Golding, B., Xu, M., and Pinto, J. (2014). Use of NWP for nowcasting convective precipitation: Recent progress and challenges. *Bulletin of the American Meteorological Society*, 95:409 – 426.

93. Testud, J., Le Bouar, E., Obligis, E., and Ali-Mehenni, M. (2000). The rain profiling algorithm applied to polarimetric weather radar. *Journal of Atmospheric and Oceanic Technology*, 17:332 – 356.

94. Thurai, M., Mishra, K. V., Bringi, V. N., and Krajewski, W. F. (2017). Initial results of a new composite-weighted algorithm for dual-polarized X-band rainfall estimation. *Journal of Hydrometeorology*, 18:1081 – 1100.

95. Troemel, S., Simmer, C., and Chwala, C. (2014). Using microwave backhaul links to optimize the performance of algorithms for rainfall estimation and attenuation correction. *Journal of Atmospheric and Oceanic Technology*, 31:1748 – 1760.

96. van der Kooij, E., Schleiss, M., Taormina, R., Fiorenelli, F., Lugt, D., van Hoek, M., Leijnse, H., and Overeem, A. (2021). Nowcasting heavy precipitation over the Netherlands using 13-year radar archive. a machine learnign approach. In *In Proceedings of the EGU Gen. Assem. Conf. 2021*.

97. Vulpiani, G., Tabary, P., and Chatelet, J. P. D. (2008). Comparison of advance radar polarimetric techniques for operational attenuation correction at C-band. *Journal of Atmospheric and Oceanic Technology*, 25:1118 – 1135.

98. Wang, Y. and Chandrasekar, V. (2010). Quantitative Precipitation Estimation in the CASA X-band Dual-Polarization Radar Network. *Journal of Atmospheric and Oceanic Technology*, 27:1665 – 1676.

99. Wilks, D. S. (2006). *Statistical Methods in the Atmospheric Sciences*. Academic Press, Cambridge Massachusetts, 2 edition.

100. Woodcock, F. (1976). The evaluation of Yes/No forecast for scientific and administrative purposes. *Monthly Weather Review*, 104(10):1209–1214.

101. Zawadzki, I., Morneau, J., and Laprise, R. (1994). Predictability of precipitation patterns: An operation approach. *Journal of Meteorology*, 33:1562 – 1571.

102. Zinner, T., Mannstein, H., and Tafferner, A. (2008). Cb-TRAM: Tracking and monitoring severe convection from onset over rapid development to mature phase using multi-channel Meteosat-8 SEVIRI data. *Meteorology and Atmospheric Physics*, 101:191 – 210.