Tampere University

Jere Mäkinen

# IMPROVING THE MEMBERSHIP FUNCTIONS OF A FUZZY HYDROMETEOR CLASSIFIER IN AN X-BAND WEATHER RADAR SYSTEM

## Parameter adjustments and performance validation using ground-based forward scatter sensor observations

Master's thesis
Faculty of Engineering and Natural Sciences
Examiners: Prof. Ari Visa
PhD Dirk Klugmann
September 2023

# ABSTRACT

Jere Mäkinen: Improving the membership functions of a fuzzy hydrometeor classifier in an X-Band weather radar system
Master's thesis
Tampere University
Master's Programme in Science and Engineering
September 2023

---

A weather radar is a remote sensing device that measures rain by transmitting a high-energy electromagnetic signal into the atmosphere and receiving echoes that scatter back to the radar. The weather targets are referred to as hydrometeors which is a generic name for any water or ice particle in the atmosphere. Modern radar technology, especially the so-called dual-polarization technique, enables measuring a large scale of parameters describing hydrometeors' size, shape, and orientation. Based on this information, it is possible to classify the hydrometeors into classes such as rain, wet snow, dry snow, or hail.

This procedure is typically done using algorithms based on fuzzy logic. Fuzzy logic is an extension of classical logic. It is capable of modeling the logical "middle ground" that is not included in classical logic by allowing truth values that are something between true and false. Membership functions are part of fuzzy systems that transform the crisp input values into fuzzy truth values. This work presents the basics of fuzzy logic and how it can be utilized in solving a classification problem.

However, weather radars operate using different frequency bands in their transmitted signal. The frequency bands are denoted with letters S, C, and X, listing from the lowest frequency to the highest. The parameters of the membership functions are dependent on the used frequency band of the radar. The aim of this work is to adjust the parameters of Vaisala's fuzzy hydrometeor classification algorithm for X-band based on the old C-band specific parameters. The adjustments made in this work are based on literature references that describe the polarimetric differences of the different frequency bands and similar adjustment processes that have been carried out before for different algorithms.

The performance of the algorithm after the parameter adjustments is studied by comparing real weather data from an X-band weather radar and a ground-based forward scatter sensor. Analysis is also supported by visual and quantitative comparison of the data with the old and the adjusted parameters. All in all, five different raining events were included in the analysis.

The results of the analysis show that after the adjustment, the number of snow bins incorrectly classified as liquid rain was significantly decreased and the algorithm behavior was more consistent in detecting hail and graupel.

Keywords: weather radar, fuzzy logic, algorithm, classification

# TIIVISTELMÄ

Jere Mäkinen: Parametrien säätäminen sadetyypin luokittelussa X-taajuusalueen säätutkassa sumean logiikan algoritmissa
Diplomityö
Tampereen yliopisto
Teknis-luonnontieteellinen DI-ohjelma
Syyskuu 2023

---

Säätutka on etämittausjärjestelmä, joka mittaa sadetta tutkaa ympäröivässä osassa ilmakehää. Tutka lähettää korkeatehoisen elektromagneettisen signaalin ilmakehään ja vastaanottaa kaikuja, jotka siroavat ilmakehässä olevista kohteista takaisin tutkalle. Moderni tutkateknologia mahdollistaa monipuolisen tiedon keräämiseen ilmakehän sääpartikkeleista eli ns. hydrometeoreista. Erityisesti kaksoispolarisaatiotekniikka mahdollistaa tiedon keräämisen hydrometeorien koosta, muodosta ja orientaatiosta. Tämän tiedon avulla hydrometeoreita on mahdollista luokitella eri sadeluokkiin kuten vesisade, lumisade tai raekuuro.

Sadetyypin tunnistus tyypillisesti toteutetaan hyödyntäen sumeaa logiikkaa. Sumea logiikka on klassisen logiikan laajennus. Se sallii myös totuusarvot, jotka ovat jotain toden ja epätoden väliltä. Jäsenfunktioiden avulla mitatut arvot voidaan muuttaa sumean logiikan mukaisiksi totuusarvoiksi, jotka kuvaavat sitä, kuinka paljon objekti kuuluu tiettyyn luokkaa mitatun ominaisuuden perusteella.

Säätutkia toimii kolmella eri taajuusalueella. Taajuusalueita, lueteltuna matalimmasta taajuudesta korkeimpaan, merkitään kirjaimilla S, C ja X. Jäsenfunktioiden parametrit riippuvat tutkan taajuusalueesta, sillä erilaiset signaalit siroavat eri tavalla mittatavista kohteista. Tämän työn tavoitteena on säätää parametreja Vaisalan sadetyypintunnistusalgoritmissa toimimaan taajuusalueella X. Parametreihin tehdyt muutokset perustuvat kirjallisuuskatsaukseen.

Tämä työ esittelee sumean logiikan perusteet ja erityisesti, kuinka sitä voidaan hyödyntää luokitteluongelmien ratkaisussa. Lisäksi esitellään menetelmä, jonka avulla algoritmin suorituskykyä voidaan arvioida vertaamalla sen tuloksia maanpinnalla olevan sääsensorin kanssa. Suorituskyvyn arviointiin kuuluu myös tulosten visuaalinen ja kvantitatiivinen vertailu PPI-kuvista.

Tulosten arviointi osoittaa, että parametrien muutosten jälkeen algoritmin suorituskyky oli parantunut. Vähemmän lumisadedatapisteitä luokiteltiin virheellisesti vesisateeksi muutosten jälkeen. Algoritmi myös toimi johdonmukaisemmin rakeiden ja lumirakeiden tunnistamisessa.

Avainsanat: säätutka, sumea logiikka, algoritmi, luokitteluongelma

# PREFACE

This work is a master's thesis written for Tampere University and Vaisala Oyj. I want to thank Vaisala for providing this interesting topic and the tools that made this master's thesis work possible. Especially, I want to thank Juha Salmivaara, Dirk Klugmann, and Pekka Puhakka for sharing their insights, ideas, and comments. Also, I want to thank Christian Malino for helping with setting up the system that was used to collect the data used in this work. From Tampere University, I want to thank Professor Ari Visa for great guidance during the master's thesis process. It has been a pleasure to work with all of you.

I have really enjoyed the last five years that I have spent studying at Tampere University. The city and the school have offered a great place to live and grow up. Thanks to all of the teachers and fellow students I've met along the way for making this journey as memorable as it has been.

Special thanks also go out to my aunt's family who generously offered me a place to stay during my visits to Vaisala's Vantaa office from Tampere. Last but not least, I want to thank my family and my partner Niina for all of their support during the thesis process and the entire period of my studies. I could not have done it without you.

Tampere, 7th September 2023

Jere Mäkinen

# CONTENTS

# ABBREVIATIONS AND NOTATIONS

| | |
|---|---|
| $A$ | A classical set |
| $\widetilde{A}$ | A fuzzy set |
| $\chi(x)$ | A membership function defining a classical set |
| CSU | Colorado State University |
| FMI | Finnish Meteorological Institute |
| $h$ | Altitude |
| $h_{\mathsf{ML}}$ | Melting level height |
| $K_{\mathsf{dp}}$ | Specific differential phase, range derivative of the differential phase |
| $\mu(x)$ | A fuzzy membership function |
| NSLL | National Severe Storms Laboratory |
| $\varphi_{\mathsf{dp}}$ | Differential phase, phase difference between the horizontal and vertical signals of dual-polarization radar |
| PPI | Plan position indicator, a type of polar data visualization that has the radar antenna in the center of the figure with the distance from the radar shown |
| RHI | Range height indicator, a cartesian type of radar data visualization that shows a given property as a function of range and altitude. |
| $\rho_{\mathsf{hv}}$ | Cross correlation ratio, correlation coefficient between horizontal and vertical received echoes |
| SNR | Signal to noise ratio |
| UTC | Coordinate Universal Time |
| $Z_{\mathsf{dr}}$ | Differential reflectivity, ratio between the horizontal and vertical reflectivity |
| $Z_{\mathsf{h}}$ | Horizontal reflectivity, property of the target that describes how much of the energy transmitted through the the horizontal polarization channel is reflected back to the radar. |
| $Z_{\mathsf{v}}$ | Horizontal reflectivity, property of the target that describes how much of the energy transmitted through the the vertical polarization channel is reflected back to the radar. |

# 1. INTRODUCTION

A weather radar is a remote sensing device that is used to measure weather, especially rain. Radar measurement is based on emitting an electromagnetic signal and receiving echoes that scatter back from objects in the signal's way [13]. Of course, the strength of the received echo, often referred to as reflectivity, can be measured, but modern weather radars have polarimetric capabilities that enable measuring a lot of additional information about the measured targets [14]. This, in turn, enables the creation of advanced data products such as estimated rainfall, attenuation corrected reflectivity, and rain type detection. Rain type detection, or hydrometeor classification, is the main focus of this work.

Hydrometeor is a common term for any water or ice particle in the atmosphere formed as a result of condensation [31]. The hydrometeor classification can be implemented using a classification scheme based on fuzzy logic. Fuzzy logic is an extension of classical logic. Unlike classical logic, fuzzy logic allows for truth values that are something in between true or false [27]. It is an effective tool to solve classification problems that include possibly noisy measurements and overlapping characteristics between different classes [12][33]. Both of these aspects are present in the hydrometeor classification as will be illustrated in this work.

However, weather radars operate using three different frequency bands and the parameters of a fuzzy hydrometeor classifier are dependent on the used frequency. The frequency bands are denoted with letters S, C, and X, listing from the lowest frequency to the highest. The need for adjustments is caused by the different scattering behaviors of different frequency signals.

In Vaisala's weather radar systems, the hydrometeor classification is handled by an algorithm called the HydroClass. The public methods that are the basis of the HydroClass were originally developed for S-band radars [10][21]. Vaisala has recently added a new X-band radar WRS400 into its radar offering, which causes a need to adjust the parameters of the HydroClass for an X-band system. HydroClass parameters have previously been adjusted for C-band radars, and these parameters work as the starting point of this study.

The aim of this work is to adjust the parameters of the HydroClass algorithm and improve the performance of an X-band weather radar system's hydrometeor classification. Also, the aim is to study the changes caused by the parameter adjustments and evaluate the quantity and the significance of the changes in the results.

The adjustments are made based on literature references as similar adjustment processes have been carried out before for other fuzzy hydrometeor classifiers. For example, Dolan and Rutledge studied hydrometeor classification with X-band weather radars in their article *A theory-based hydrometeor identification algorithm for X-band polarimetric radars* [3] and Al-Sakka et. al describe the differences in the classification schemes between the different frequency bands in their article *A new fuzzy logic hydrometeor classification scheme applied to the French X-, C-, and S-band polarimetric radars* [20].

Verifying the performance of a weather radar is not an easy task due to the nature of the measurement. The measurement volumes observed by a radar are large and it is impossible to replicate them in a controlled environment. Therefore, the performance validation must be done based on real weather data measured by a WRS400 radar located at Vaisala headquarters in Vantaa. In this work, a few simple methods to analyze the performance of the HydroClass are presented. The results of the HydroClass are compared to data from a ground-based forward scatter sensor, Vaisala FD70. This provides an accurate reference point, but it considers only a tiny portion of the radar's measurement volume. Therefore, the mentioned method is supported by counting the occurrences of different hydrometeor types given by HydroClass using the old and the adjusted parameters in large measurement volumes. The data is also inspected visually.

This work is divided into the following chapters. Chapter 2 provides the theoretical background for the thesis. This includes an introduction to the basic operating principle of a weather radar, an overview of fuzzy logic focusing on solving classification problems, and a description of the HydroClass algorithm. In the final section of chapter 2, the phenomena that cause the need for adjusting the parameters are introduced. Chapter 3 introduces the methodology and the data set used in the performance validation. This also includes a short description of the devices that were used to collect the data. In chapter 4, the parameter adjustments and the results from the performance analysis are presented. The conclusions from the analysis and possible ideas for further studies and improvements are presented in chapter 5. Chapter 6 summarizes the key content and findings of this work.

# 2. THEORETICAL BACKGROUND

Even though the basic working principle of a radar is rather simple it is a really complex device. As the first part of the theoretical background, a brief introduction to weather radar measurement is presented. The focus is on the polarimetric data moments that modern weather radars are capable of providing. These in addition to the reflectivity of the targets in the radar's measurement volume, provide ways to analyze the weather conditions in a versatile and accurate way [14]. Measuring the polarimetric properties enables the creation of advanced data products such as attenuation-corrected reflectivity, rainfall estimation, and hydrometeor classification, the main topic of this work.

Fuzzy logic is a regularly used, if not the most frequently used, tool to solve the problem of hydrometeor classification [20]. Polarimetric data moments are key inputs for classification. In some methods, temperature or melting level height data are used as inputs as well to include additional information about the surrounding conditions [15]. All in all, the polarimetric capabilities of modern weather radars and fuzzy logic are the key to accurate hydrometeor classification.

HydroClass is the fuzzy logic based method that Vaisala uses for hydrometeor classification. In general, weather radars operate on three different frequency bands, S, C, and X. The parameters of the HydroClass algorithm are dependent on the frequency band that the radar is operating on or, in other words, the wavelength of the transmitted signal. The motivation of this work is the phenomena that cause different scattering phenomena in different frequency bands. The aim of this work is to adjust the parameters of the HydroClass algorithm to improve its performance in an X-band weather radar system. The start point of the adjustment process is the HydroClass parameters adjusted for C-band radars. The adjustments are needed because the scattering of the transmitted signal is different based on the wavelength of the signal.

This chapter provides the theoretical background for precipitation classification by discussing the mathematical basis of the HydroClass algorithm and the public methods that the HydroClass is based on. The basics of the weather radar and the data moments measured by a dual-polarization weather radar are introduced. Finally, the phenomena that cause the need for the frequency band specific adjustments are discussed.

## 2.1 Weather radar

Next, the basics of weather radar are briefly introduced. Dual-polarization and especially the capabilities provided by it are also discussed. The content of this section is based on Skolnik's *Radar Handbook* [14] and Rinehart's *Radar for meteorologists* [16].

### 2.1.1 A Brief introduction and an overview to weather radars

The basic working mechanism of radar is to emit a high frequency electromagnetic signal that scatters from objects in the signal's way [13]. Some of the scattered signal is received and detected back at the radar. Based on the time delay and the frequency shift of the returned signal, the distance, speed, and direction of the encountered object can be determined. The radar measurement can be summarized in the following steps.

1. The radar sends a pulse of electromagnetic energy into the atmosphere.
2. Some of the transmitted signal encounters object(s) on its way and the energy of the signal scatters in many directions.
3. Some of the scattered signal, also referred to as echo, returns back to radar and is received by the antenna.
4. The echo is amplified and moved into signal processing.
5. The signal processor processes the data and the data products are computed.

The summarized steps are based on reference [14]. Figure 2.1 shows a simplified block diagram of radar with the most critical components.

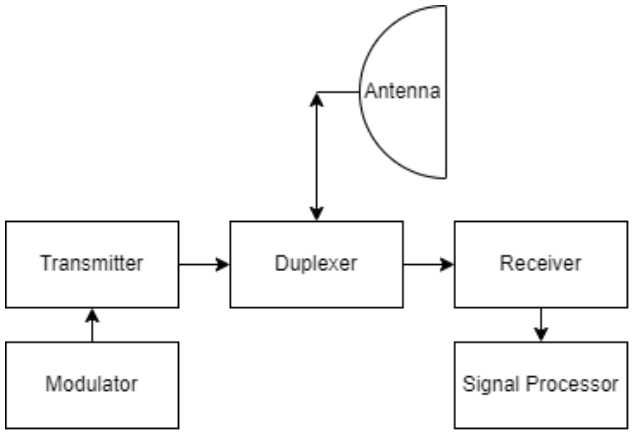



***Figure 2.1.*** *A simplified block diagram of a radar. Figure is based on reference [16].*

The transmitter sends out the signal with a suitable waveform. A typically used transmitter option in weather radar applications is the magnetron transmitter. A magnetron transmitter is a vacuum tube device that generates microwave signals by applying a strong magnetic field to accelerate electrons emitted from a heated cathode, causing them to

spiral in resonant cavities within the anode. This spiraling motion, known as cyclotron resonance, produces amplified microwave radiation. Another option is to use solid-state transmitters. They generate signals using semiconductor devices, like transistors, without relying on vacuum tubes. These transmitters manipulate and amplify electrical currents through solid-state materials, converting them into high-frequency signals. Solid-state transmitters offer advantages such as efficiency, reliability, and compact size compared to traditional vacuum tube-based systems [14].

The modulator's purpose is to switch the transmitter on and off and provide the right waveform for the pulse to be transmitted. In other words, it tells the transmitter when to transmit and for how long. The waveform selection is a topic that is largely omitted in this work but generally, a common waveform could be a series of narrow, rectangular-like pulses. A typical order of magnitude for the duration of a pulse is between 0.1 and 10µs. The pulse duration affects the range resolution of the radar measurement. Solid-state transmitters require the usage of longer pulses to compensate for their lower peak power compared to magnetron transmitters. Similar range resolution can be maintained by using pulse compression techniques [14]. The rate at which the radar transmits is called the pulse repetition frequency (PRF). The PRF and the pulse duration together limit the maximum range of the measurement. The echoes from the transmitted pulse must have time to travel back to the radar before the transmission of the next pulse.

The antenna, which is the most visible and recognizable part of radar, is the device that directs the transmitted signal into the atmosphere. Directing the transmitted signal is done to concentrate the power and at the same time enable determining the direction of the detected targets. The directed transmitted energy is called a beam. The antenna also collects the echoes. The antenna design, shape, and size in combination with the wavelength of the used signal determine the width of the radar beam and the resolution to separate targets in angle. A typical beam width in weather radar applications is $1^\circ$. The antenna can rotate with respect to both the horizontal and vertical axes. The horizontal angle that the antenna is pointed at is called the elevation angle and the vertical angle is called the azimuth angle. The azimuth is reported with respect to North. So, azimuth of $0^\circ$ means that the radar is pointed North and $90^\circ$ means the radar is pointed East, and so on.

The receiver is the component that detects, amplifies, and down converts the signals collected by the antenna. These echoes are often very weak, which means that the receiver must be very sensitive. From the receiver, the echo signal is directed to the signal processor. The signal processor's role is to carry out different computational procedures, such as quality control, calibration, and data processing. The role of digital signal processors has increased in weather radars in recent years. They enable fast and comprehensive computation of different data products. Some of these data products are presented in the next chapter.

The duplexer protects the receiver from the high power emitted by the transceiver. The receiver is designed to detect echoes whose power is several orders of magnitude smaller compared to the power of the transmitted signal. Therefore, the receiver could be damaged if it is exposed to the transmitted power. The duplexer prevents this from happening by blocking the transmitted signal from getting to the receiver. On the other hand, it also directs the received echoes to the receiver rather than the transmitter. All in all, the duplexer allows for the use of the same antenna for transmitting and receiving.

Weather radar can be run using different kinds of measurement schemes, but a typical way to operate a weather radar is a so-called volume scan. In a volume scan, the antenna rotates full $360^\circ$ about the vertical axis. The elevation angle stays constant during the full rotation. This process is repeated using multiple different elevation angles to measure a big volume of the atmosphere surrounding the radar.

### 2.1.2 Data moments measured by a polarimetric weather radar

The signal transmitted by a weather radar is polarized. This means that the electromagnetic signal only vibrates in a specific orientation. Modern weather radars transmit and receive two orthogonal polarizations, horizontal and vertical. This does require some changes to the radar hardware. For example, the radar must have separate transmitting and receiving channels for each polarization. Transmitting a dual-polarized signal is not a trivial task and the theoretical background of the technique must be omitted in this work. However, a good overview of simultaneous horizontal and vertical transmission and its repercussions are given in an article by Scott et. al *The Use of Simultaneous Horizontal and Vertical Transmissions for Dual-Polarization Radar Meteorological Observations* [22].

Dual polarization gives significantly more information about the observed precipitation compared to using only one polarization. The additional information originates from comparing and combining information from received echoes from horizontal and vertical channels. This enables a better understanding of properties such as scatterers' shape, size, phase, and orientation. Next, some of the most common polarimetric data moments are introduced. The following definitions are from Merrill Skolnik's Radar Handbook [14].

Reflectivity describes how much of the energy is reflected back from the target. In weather radar's case, hydrometeors in the measured volume are the targets. Reflectivity is usually denoted by $Z$ and its magnitude is reported in decibels. Polarimetric radars can measure reflectivity from both horizontal and vertical channels. Reflectivity values from horizontal and vertical channels are denoted by $Z_\mathrm{h}$ and $Z_\mathrm{v}$ respectively. In this work, reflectivity refers to horizontal reflectivity $Z_\mathrm{h}$ if not mentioned otherwise.

Differential reflectivity is a ratio of the horizontal reflectivity and the vertical reflectivity. It is defined by

$$Z_{\text{dr}} = Z_{\text{h}} - Z_{\text{v}}, \tag{2.1}$$

where $Z_{\text{dr}}$ is the differential reflectivity. $Z_{\text{h}}$ and $Z_{\text{v}}$ are in decibel units. Positive $Z_{\text{dr}}$ indicates more prominent horizontal echoes, which is common for large raindrops. A negative value, on the other hand, indicates more prominent vertical echoes, which may occur in hail or graupel.

The cross-correlation ratio also describes the relation between the horizontal and vertical echoes as it is the correlation coefficient between horizontal and vertical echoes. The cross-correlation ratio is computed using the formula for the Pearson correlation coefficient

$$\rho_{\text{hv}} = \frac{\text{Cov}(Z_{\text{h}}, Z_{\text{v}})}{\sigma_{\text{h}}\sigma_{\text{v}}}, \tag{2.2}$$

where $\rho_{\text{hv}}$ is the cross-correlation ratio, $\sigma_{\text{h}}$ and $\sigma_{\text{v}}$ are standard deviations of reflectivity in horizontal and vertical polarization respectively, and $\text{Cov}(Z_{\text{h}}, Z_{\text{v}})$ is the covariance of horizontal and vertical reflectivity. If the target is precisely spherical, $\rho_{\text{hv}} = 1$ as the horizontal and vertical echoes would be equal. On the other hand, if there is no correlation at all then, $\rho_{\text{hv}} = 0$. Generally, all meteorological targets are very close to a spherical shape, but higher values indicate uniform precipitation areas and lower values more mixed hydrometeor types, such as melting snow, wet snowflakes, or airborne debris. Significantly lower values indicate noise or ground clutter targets such as buildings.

Differential phase is the phase difference between horizontally and vertically polarized signals given by

$$\varphi_{\text{dp}} = \varphi_{\text{h}} - \varphi_{\text{v}}, \tag{2.3}$$

where $\varphi_{\text{dp}}$ is differential phase and $\varphi_{\text{h}}$ and $\varphi_{\text{v}}$ are horizontal and vertical phases respectively. Closely related to the differential phase, the specific differential phase is the rate of change of the phase difference between the horizontal and vertical channels. It is given by

$$K_{\text{dp}} = \frac{d\varphi_{\text{dp}}}{dr}, \tag{2.4}$$

where $K_{\text{dp}}$ is specific differential phase and $r$ is range. Effectively, $K_{\text{dp}}$ is the range derivative of measured $\varphi_{\text{dp}}$. Greater horizontal shift results in a positive $K_{\text{dp}}$ value while

greater vertical shift results in a negative $K_{\text{dp}}$. As $\varphi_{\text{dp}}$ is often regarded as a data moment prone to noise, $K_{\text{dp}}$ is computed over a longer range window rather than between just two successive range bins. $K_{\text{dp}}$ is a really useful data moment in many applications because it is not affected by propagation attenuation and it is related to the intensity of the precipitation [30].

One additional data moment that is worth mentioning in this context is signal to noise ratio which is given by

$$\text{SNR} = 10 \log_{10} \left( \frac{P_{\text{signal}}}{P_{\text{noise}}} \right), \tag{2.5}$$

where SNR is signal to noise ratio in decibel units and $P_{\text{signal}}$ and $P_{\text{noise}}$ are the power of the radar return signal and power of the background noise in the received echo respectively. It is a measure comparing the level of the desired signal to background noise. It can be used as a threshold in processing the data in order to exclude noisy data points.

Based on the polarimetric data moments, it is possible to create more advanced data products such as hydrometeor classification or rainfall estimation. In this work, the focus is on hydrometeor classification and especially Vaisala's HydroClass algorithm. The polarimetric data moments act as inputs for the algorithm. Classification performed by the HydroClass algorithm is based on fuzzy logic. Next, we will take a brief look into fuzzy logic and how it can be used to solve a classification problem.

## 2.2 Fuzzy logic

Fuzzy logic can be described as an extension of traditional logic. Traditionally, logic defines truth value as a binary variable that can have a value of 1 or 0, true or false. In fuzzy logic, a continuous transition between truth values 0 and 1 is allowed. In a way, introducing fuzzy logic represents a transition from asking the question, 'Does $x$ belong to class $y$?' to asking, 'To what extent does $x$ belong to class $y$?' [27]

In this section, basic fuzzy logic concepts, such as fuzzy sets, membership functions, and inference, are introduced focusing on a viewpoint of a classification problem. A general framework for implementing a fuzzy logic based classification algorithm is presented and finally, the strengths and weaknesses of fuzzy systems are discussed.

### 2.2.1 Fuzzy sets and membership functions

Before diving into fuzzy classifiers we must define some basic concepts of fuzzy logic. The definitions and theorems presented in this chapter are from reference [24].

Even though the topic at hand is fuzzy logic, let us first discuss sets as they are understood in classical logic. These sets will be referred to as classical sets in this work to separate them from their fuzzy counterparts. When it comes to classical sets, an element either belongs to the set or it doesn't. A mapping that defines belonging to a given set is presented below.

**Definition 2.1.** *Let $A$ be a classical subset of $X$. Then function $\chi_A : X \to \{0, 1\}$, where $\chi_A$ is membership function of set $A$ that defines if element $x \in X$ belongs to set $A$. Function $\chi_A$ is of the form*

$$\chi_A(x) = \begin{cases} 1 & x \in A, \\ 0 & x \notin A. \end{cases} \tag{2.6}$$

Figure 2.2 presents a visualization of a classical set.

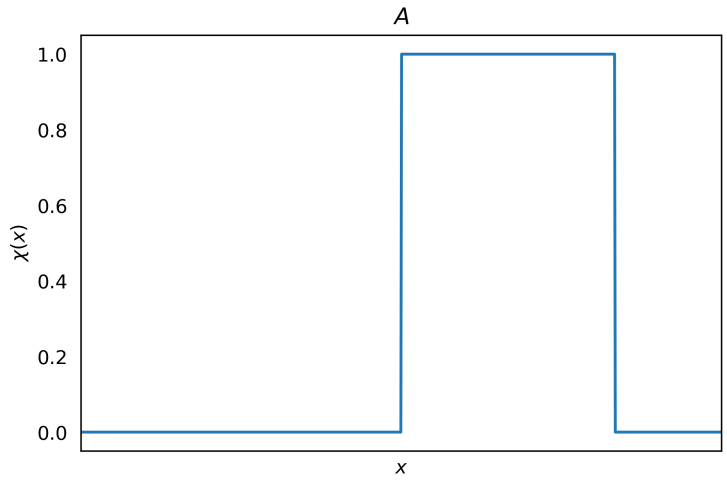



***Figure 2.2.*** *An example of a classical set.*

As previously mentioned, in fuzzy logic, an element may have a degree of belonging to a set that is less than 1 but greater than 0. Definition 2.1 mentions membership functions. Generally, membership functions describe the degree to which a given input $x$ belongs to a given set $A$, and they are a concept that is heavily tied to fuzzy sets. Membership functions are used to convert crisp measured values into fuzzy values that are related to the chosen output values.

**Definition 2.2.** *Let $\widetilde{A}$ be a fuzzy subset of $X$. Then a mapping given by $\mu_{\widetilde{A}} : X \to [0, 1]$ is the membership function of the set $\widetilde{A}$.*

Unlike classical sets, where the membership function always has the form presented in 2.1, a membership function of a fuzzy set may have countless different forms. Figure 2.3 shows some typical examples of shapes that fuzzy sets may have.

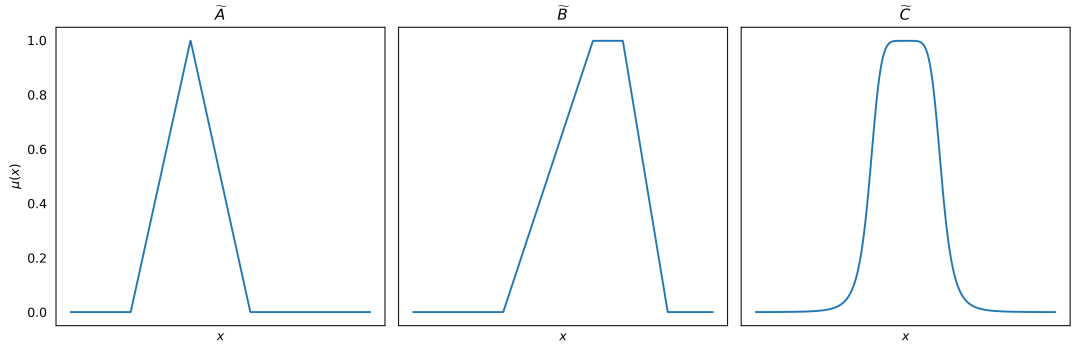



***Figure 2.3.*** *Examples of fuzzy sets.*

By no means does figure 2.3 include all possible shapes that a fuzzy set could have. The shape of a fuzzy set doesn't have to be symmetrical or even convex as the examples presented here are. Of course, fuzzy sets can be multi-dimensional as well, but for the reason of simple visualization, the examples in 2.3 were all chosen to be 1-dimensional. One way to view the shape of a fuzzy set defined by a membership function is by three properties: core, support, and boundary. Next, we will give the definition for these properties.

**Definition 2.3.** *Let $\widetilde{A}$ be a fuzzy subset of $X$ and $\mu_{\widetilde{A}}(x)$ is its membership function. A membership function's core, support, and boundary are defined as follows.*

*Core:*

$$\widetilde{A}_{core} = \{x \mid \mu_{\widetilde{A}}(x) = 1\}$$

*Support:*

$$\widetilde{A}_{support} = \{x \mid \mu_{\widetilde{A}}(x) > 0\}$$

*Boundary:*

$$\widetilde{A}_{boundary} = \{x \mid 0 < \mu_{\widetilde{A}}(x) < 1\}$$

To summarize the content of definition 2.3, the core is the region where membership is full, the support is the region where membership to at least some nonzero extent exists, and the boundary is the region where membership is not full but exists to at least some extent. Therefore, the support includes both the core and the boundary regions. Figure 2.4 further illustrates these properties.

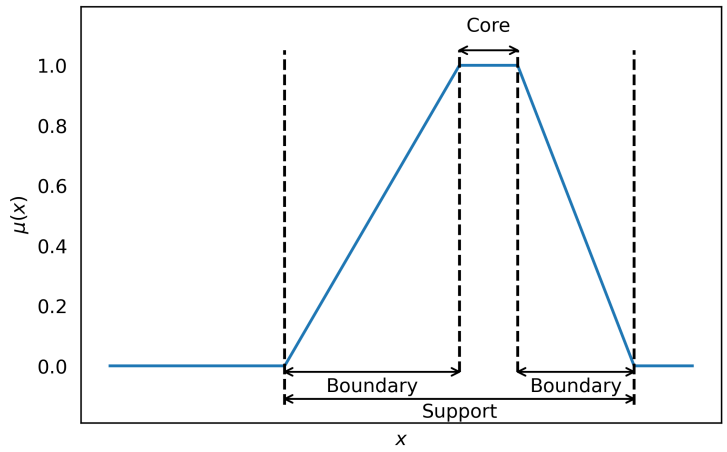



***Figure 2.4.*** *Example visualization of a membership functions core, support, and boundary.*

As fuzzy logic is an extension of classical logic, it is consistent that almost all of the properties, operations, and theorems that are used with classical sets can also be defined for fuzzy sets. For example, operations like union, intersection, and complement are defined for fuzzy sets similarly to how they are defined for classical sets.[24] Figure 2.5 presents an example of the mentioned operations for fuzzy sets $\widetilde{A}$ and $\widetilde{B}$ that were originally presented in figure 2.3. It is extremely important to have these operations defined because it allows combining fuzzy sets with one another and in turn, creating fuzzy systems like fuzzy classifiers [27].

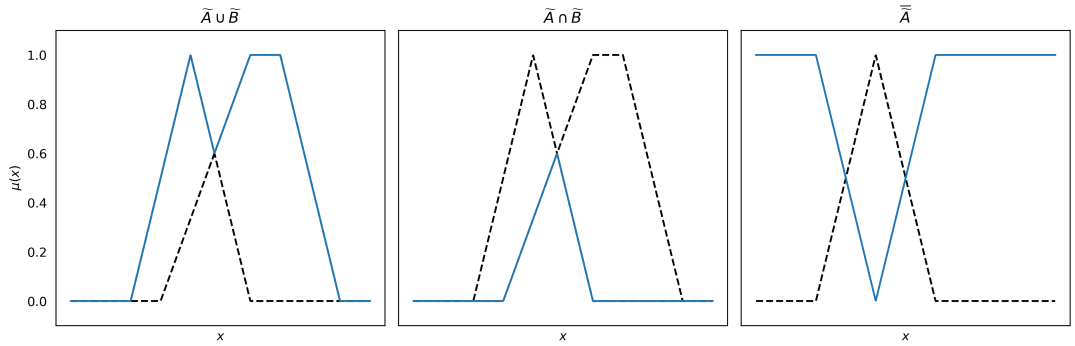



***Figure 2.5.*** *Examples of set operations union, intersection, and complement on fuzzy sets. The blue line represents the result of operations while the dashed black line shows the original fuzzy sets.*

Sets visualized in figure 2.3 follow the formulas given in definition 2.4.

**Definition 2.4.** *Let $\widetilde{A}$ and $\widetilde{B}$ be fuzzy sets of $X$ and $\mu_{\widetilde{A}}(x)$ and $\mu_{\widetilde{B}}(x)$ be the corresponding membership functions. Then set operations complement, union, and intersection are defined by the following membership functions.*

*Complement:*

$$\mu_{\overline{\widetilde{A}}}(x) = 1 - \mu_{\widetilde{A}}(x)$$

*Union:*

$$\mu_{\widetilde{A}\cup\widetilde{B}}(x) = \max(\mu_{\widetilde{A}}(x), \mu_{\widetilde{B}}(x))$$

*Intersection:*

$$\mu_{\widetilde{A}\cap\widetilde{B}}(x) = \min(\mu_{\widetilde{A}}(x), \mu_{\widetilde{B}}(x))$$

Definitions in 2.4 stand for classical sets too. Classical sets could be seen as a special case of a fuzzy set as the membership function of a classical set given in definition 2.1 also matches the definition of fuzzy membership function.

As the operations are similar it is clear that important properties of sets such as commutativity, associativity, and transitivity are the same for fuzzy sets too.

**Theorem 2.1.** *Let $A, B$ and $C$ be classical subsets of $X$. Then, properties of commutativity, associativity, and transitivity are defined as follows.*

*Commutativity*

$$A \cup B = B \cup A,$$
$$A \cap B = B \cap A.$$

*Associativity*

$$A \cup (B \cup C) = (A \cup B) \cup C,$$
$$A \cap (B \cap C) = (A \cap B) \cap C.$$

*Transitivity*

*If $A \subseteq B \subseteq C$, then $A \subseteq C$.*

*The properties and their definitions are identical for fuzzy subsets $\widetilde{A}$, $\widetilde{B}$, and $\widetilde{C}$.*

However, there are also differences in the set operations of classical and fuzzy sets. To be precise, there is particularly one theorem that doesn't hold for fuzzy sets and that is the excluded middle law.

**Theorem 2.2.** *Let $A$ be a classical subset of $X$. Also, let $\phi$ be an empty set*

$$\phi = \{\}.$$

*Then, the excluded middle law states that*

$$A \cup \overline{A} = X,$$
$$A \cap \overline{A} = \phi.$$

Effectively, theorem 2.2 states that any statement is either true or false and there is no so-called middle ground. Figure 2.6 shows a visualization of the theorem 2.2.

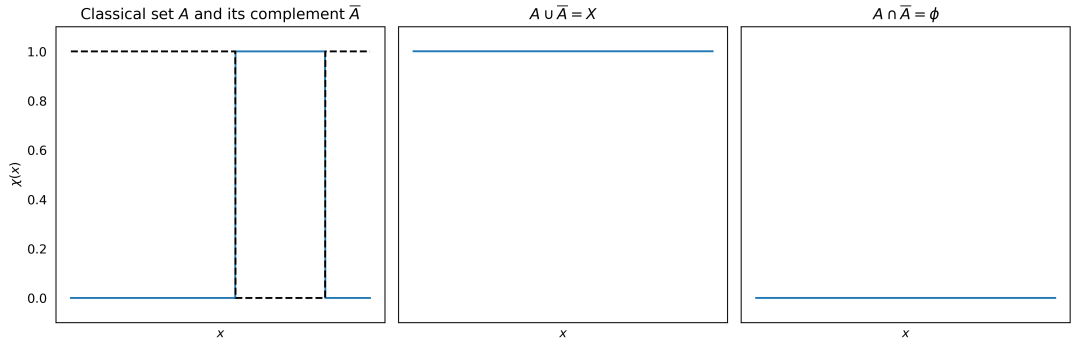



***Figure 2.6.*** *Visualization of the excluded middle law.*

It has already been established in this work that this is not the case when it comes to fuzzy logic. Including the "middle ground" is the core essence of fuzzy logic. For example, in the rightmost plot in figure 2.5 it is clear that the fuzzy set $\widetilde{A}$ and its complement are overlapping. Therefore, elements can belong to both sets at the same time. This is a clear contradiction with theorem 2.2. This is even more evident from figure 2.7. The figure shows the intersection of a fuzzy set and its complement which is clearly not empty.

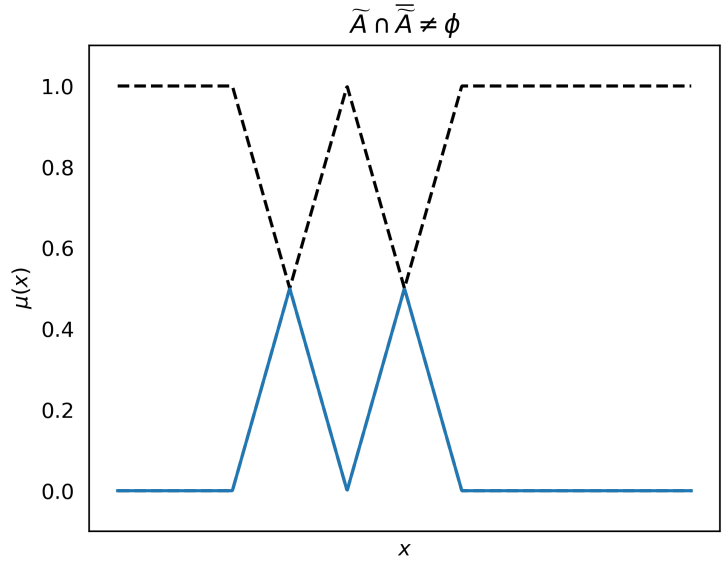



***Figure 2.7.*** *Visualization of the excluded middle law not holding for fuzzy sets. The original fuzzy set $\widetilde{A}$ and its complement are visible as a black dashed line.*

The similarities between the properties of classical and fuzzy sets go beyond what is introduced in this work. However, the main conclusion from the introduced properties and operations is that fuzzy sets can mostly be handled using the same tools as classical sets, but there are some key differences that make fuzzy sets more useful in some cases compared to tools that classical logic can offer. Fuzzy logic can handle statements that are not exclusively true or false. Membership functions are at the core of fuzzy logic as they are the component that models this "fuzziness". In the next section, membership functions' usage is discussed further as a general method for fuzzy logic based classification is introduced.

### 2.2.2 Rule base and fuzzy logic based classification

Fuzzy logic based classifiers combine information from multiple variables or multiple properties of the input object, for example, polarimetric properties of echoes measured by a weather radar. Combining information from multiple variables with fuzzy sets allows modeling human expert-like decision making where different properties are analyzed on how similar they are to a typical object in a given class. In order to achieve this, the crisp measurements have to be converted into fuzzy values that describe the mentioned similarity. This is what the membership functions are used for [27]. Combining information from multiple variables requires constructing inference and aggregation rules [10]. Finally, fuzzy outputs have to be transformed into crisp class labels.

Let us first introduce a basic structure for implementing a fuzzy logic based classifier and then discuss each step in more detail. Generally, a fuzzy classifier includes the following steps.

1. Identification of input variables and output values
2. Fuzzification of the input variables
3. Creating a rule base
   (a) Inference
   (b) Aggregation
4. Defuzzification of the results

The presented structure is created based on references [10] and [12]. Figure 2.8 further illustrates the flow of a general fuzzy classifier.

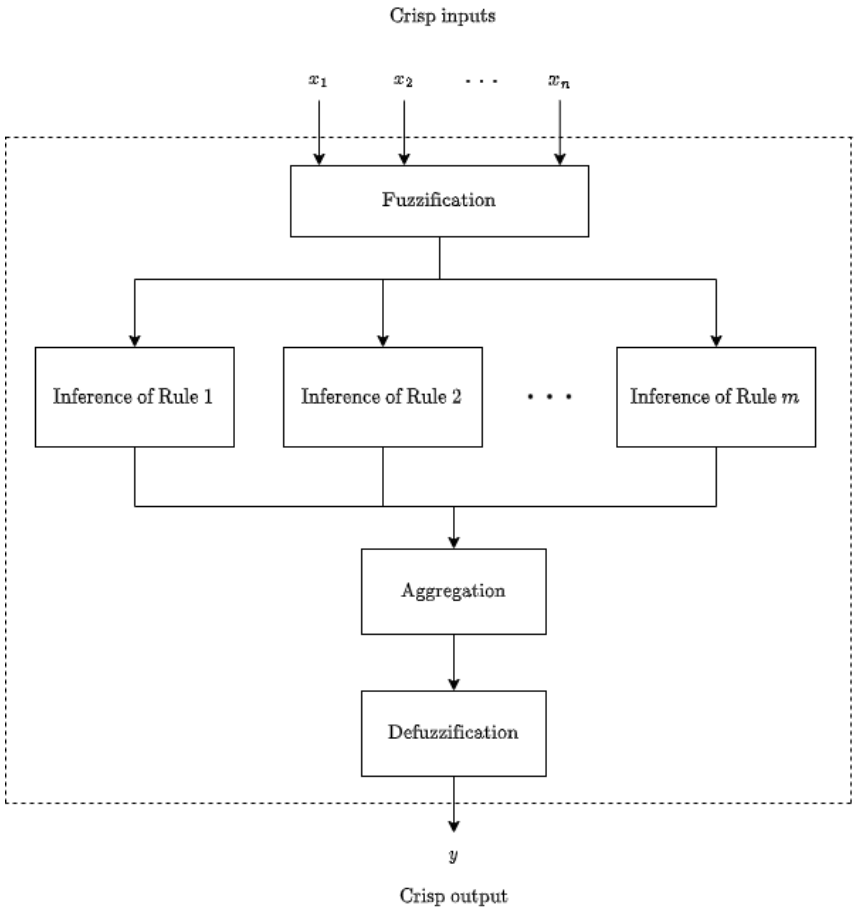



***Figure 2.8.*** *Block diagram of a general fuzzy classifier. Adapted from reference [10].*

First of all, the relevant input variables for solving the problem must be identified. In the process of choosing the input variables, one must consider what variables are available and which variables provide significant additional information for the classification. In

cases where fuzzy logic is typically used the properties related to output classes are overlapping. In other words, different classes have partly similar properties. Therefore, enough input variables are needed in order to discriminate between classes and for the classification to be accurate. Having too many input variables or output classes is not good either because that can rapidly increase the number of membership functions needed and even though fuzzy logic systems are relatively easy to implement this can make the model hard to maintain. As every pair of input variable and output value needs their own membership function, the size of the model can easily become too large.

The process of fuzzification was already briefly touched on in the previous section. It is the process of converting crisp measured values to fuzzy values using membership functions [24]. Therefore, this phase also includes formalizing the membership functions. Selecting suitable membership functions can be tricky and it requires extensive knowledge of a given topic [12]. Membership functions are often based on previous research such as observed properties and distributions [18]. Each input variable and output class combination requires its own membership function. Therefore, a system that takes $m$ input variables and has $n$ output values needs $n \cdot m$ membership functions. This further illustrates the need to consider the number of inputs and outputs carefully.

The fuzzification of the crisp inputs provides $n$ times $m$ fuzzy values that describe to what degree input belongs to a given class based on a given variable. To combine the information from all variables inference is needed. In practice, this means a set of IF-THEN rules. A fuzzy rule system is presented below in definition 2.5.

**Definition 2.5.** *Let $x$ be the object to be classified by the fuzzy classifier and $i = 1, \ldots, n$ be the possible output classes. Furthermore, let $x_j,\ j = 1, \ldots, m$ be input variables of a fuzzy classifier system, $\widetilde{B}_i$ be fuzzy outputs of the classifier, and $\widetilde{A}_{ij}$ be fuzzy subsets of $X_j$. Then a fuzzy rule system is defined as follows.*

*Rule 1: IF $x_1$ is in $\widetilde{A}_{11}$ and $\cdots$ and $x_m$ is in $\widetilde{A}_{1m}$ THEN $x$ is in $\widetilde{B}_1$*
*Rule 2: IF $x_1$ is in $\widetilde{A}_{21}$ and $\cdots$ and $x_m$ is in $\widetilde{A}_{2m}$ THEN $x$ is in $\widetilde{B}_2$*
$\ldots$
*Rule n: IF $x_1$ is in $\widetilde{A}_{n1}$ and $\cdots$ and $x_m$ is in $\widetilde{A}_{nm}$ THEN $x$ is in $\widetilde{B}_n$*

*Membership functions $\mu_{ij}(x)$ describe to what extent does input variable $x_j$ belong to fuzzy set $\widetilde{A}_{ij}$.*

Fuzzy logic allows input $x$ to belong to multiple outputs $\widetilde{B}_i$. Therefore, a way is needed to define the strength of each rule. Maybe the simplest way rule strength could be defined is a weighted sum.

$$\text{RS}_i = \sum_{j=1}^{m} w_j \mu_{ij}(x_j), \tag{2.7}$$

where $\text{RS}_i$ is the rule strength of rule $i$ and $w_j$ is the weight given to $j$th statement in the rules IF part. This is only one possibility for computing the rule strengths. The formula can be a product, a sum, a hybrid of the two, or possibly even something else. The product method minimizes the possibility of absurd classifications as even one clear out of range value will make sure that the corresponding absurd class is not chosen. On the other hand, the additive method provides more control to determine how much each input variable affects the classification [9]. Again, just like in membership function selection, expertise is needed in order to pick the right formula. Some variables are often better indicators for classification than others and then it makes sense to emphasize them in the rule strength computation.

An important note about rule strength is that the values should not be confused with probabilities. The value is not the probability of input belonging to output $\widetilde{B}_i$. It is a reference value and it enables comparison between different output sets. Also, depending on the chosen formula, the rule strength value is not even necessarily limited to interval $[0, 1]$.

The process of obtaining the final conclusions from IF-THEN rules is called aggregation [24]. The simplest way to achieve this is to choose the greatest value of the computed rule strength. This so-called maximum method can be written as

$$y = \arg\max_i(RS_i). \tag{2.8}$$

The final part of a fuzzy classification process is defuzzification. This means converting the fuzzy output of the rule base to a crisp output value. In other words, it returns the class label that corresponds to the aggregation result.

In conclusion, fuzzification is a transition from crisp values to a fuzzy system and defuzzification is a transition back to crisp values. On the other hand, membership functions and the rule base determine the model's operational logic. Chapter 2.3 will provide an example of the classification process as the hydrometeor classification scheme is introduced, but next the strong and weak points of the fuzzy logic are discussed.

### 2.2.3 Strengths and weaknesses of fuzzy systems

Before starting the implementation of a fuzzy logic classifier, one should decide if fuzzy logic is a suitable tool for the problem at hand. Fuzzy logic is a great tool when input values may be noisy or imprecise and the rules that tie input values to outputs may be vague [33]. Both of these aspects are presented in the case of hydrometeor classification too. The nature of weather radar measurement makes it so that the measurements are quite prone to noise. There are targets such as tall buildings, trees, bugs, pollen, or debris in measurement volume that may significantly affect the measured values. Also, many of the measured polarimetric data moments are prone to errors as their value scale are rather small to start with. For example, differential reflectivity $Z_{\mathrm{dr}}$ in weak rain typically gets values that are very close to zero. This value is prone to noise, in the case of decreased signal-to-noise ratio. The vagueness of the rules and output classes is also an apparent issue in the hydrometeor classification problem. Rules that state how typical a given input value is for a given class are often overlapping between different precipitation types. Also, classes, such as light and moderate rain or wet and dry snow, are rather vague and may be hard to tell apart from one another even by a human observer.

But maybe the most significant aspect that makes fuzzy logic an especially useful tool in hydrometeor classification is that implementing a model based on traditional logic or statistics would be exponentially harder than a fuzzy logic system. For example, implementing a Bayesian model or a decision tree classifier would require coming up with strict thresholds for different classes but that is not possible because properties of a given hydrometeor type are not exclusive to it and therefore the properties are overlapping. Another option could be a statistical model but that is significantly more challenging and laborious to implement in comparison to a fuzzy system [10]. Also, fuzzy logic classifiers have historically proven to be quite effective in this task [9][20].

Fuzzy systems have their problems as well. First of all, determining the membership functions can be a challenging task and requires expertise in the field of the given classification problem. Unlike some machine learning methods, fuzzy systems do not have memory and because of that, they lack the ability to learn and adjust themselves. Therefore, hybrid models that combine fuzzy logic with machine learning methods are becoming more common. This transition is visible in recent literature about hydrometeor classification as well. For example, references [2] and [15] describe methods in which a fuzzy classifier is used in combination with clustering methods, and reference [10] describes a method in which a fuzzy classifier is adjusted using neural networks. The methods for analyzing the stability of the model are not as straightforward as for many other common classification methods. Lastly, the term 'fuzzy logic' might be misleading to some [12]. Despite the word fuzzy being related to a lack of precision, there is nothing fuzzy in the method. Instead, it is all built on a firm mathematical foundation as shown in this chapter.

## 2.3 HydroClass

HydroClass is Vaisala's hydrometeor classification algorithm. HydroClass is a combination of public methods, all of which are based on a fuzzy logic approach. The three main parts of the algorithm are MeteoClassifier, PreClassifier, and PrecipClassifier [29]. MeteoClassifiers is based on a fuzzy classification method developed by the Colorado State University (CSU) [10]. The implementation follows the revised version of the CSU method described in reference [9]. PreClassifier and PrecipClassifier are both based on a method developed by the National Severe Storms Laboratory (NSSL) [21].

The methods mentioned above are all different when it comes to input data moments, membership functions, and use cases within HydroClass. Next, the components of HydroClass and the methods that they are based on are introduced in more detail.

### 2.3.1 MeteoClassifier

MeteoClassifier is part of the HydroClass algorithm that performs the actual hydrometeor type classification. Implementation closely follows the method originally developed at the Colorado State University [9]. However, the original CSU method uses 10 final output classes while Vaisala's implementation uses only six. The original model has different classes for small and large hail and drizzle and rain while MeteoClassifier only has one class for hail and one class for liquid form rain.

The hydrometeor classes that Meteoclassifier use are 'Rain', 'Wet snow', 'Dry Snow', 'Graupel', 'Hail', and 'Rain&Hail'. However, inputs that fall under class 'Rain&Hail' are reported as 'Hail' in the final output. Finally, there is a 'No met' class which is given for input with no weather signal. This means either a non-meteorological target like a building or a bird or some other non-meteorological target.

Polarimetric data moments reflectivity ($Z_{\mathrm{h}}$), differential reflectivity ($Z_{\mathrm{dr}}$), specific differential phase ($K_{\mathrm{dp}}$), and cross-correlation coefficient ($\rho_{\mathrm{hv}}$) are the inputs of MeteoClassifier. In addition, observation altitude ($h$) and melting level height ($h_{\mathrm{ML}}$) are used as inputs. The melting layer is the altitude interval in which the frozen hydrometeors melt to liquid form and the top of this layer is called melting level [4]. The air temperature is a useful variable in hydrometeor classification and it is typical that classification algorithms take it into account in some way. For example, some methods may use temperature data from soundings as an additional input [34]. The CSU method and the MeteoClassifier use the observation altitude and the melting level height to achieve the inclusion of the temperature data [10].

Membership functions for all inputs and classes have the same form. The formula for membership functions is

$$\mu_{ij}(x_j; m_{ij}, a_{ij}, b_{ij}) = \frac{1}{1 + \left( \left( \frac{x_j - m_{ij}}{a_{ij}} \right)^{b_{ij}} \right)^2}, \tag{2.9}$$

where $x$ is any of the input variables $Z_{\text{h}}, Z_{\text{dr}}, \rho_{\text{hv}}, K_{\text{dp}}$ or $h$. Parameters $m, a$, and $b$ define the shape of the membership function and they are different for each pair of input variable and output class. The parameter $m$ is the middle point of the membership functions core, the parameter $a$ determines the width of the core, and the parameter $b$ determines the slope of the boundary region as described by definition 2.3.

Membership functions corresponding to $Z_{\text{h}}$ and $\rho_{\text{hv}}$ follow exactly the form presented in 2.9. Figures 2.9 and 2.10 present visualizations of these membership functions.

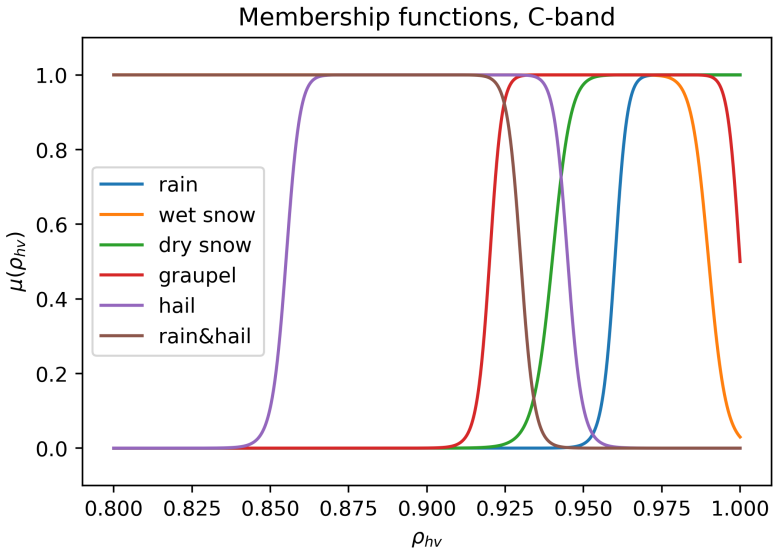



***Figure 2.9.*** *MeteoClassifier's membership functions for $\rho_{hv}$.*

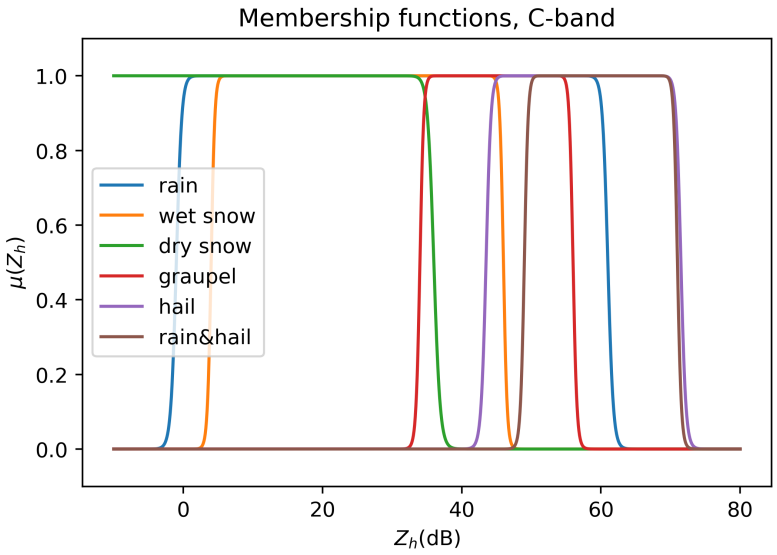



***Figure 2.10.*** *MeteoClassifier's membership functions for $Z_h$.*

The other input variables require a two-dimensional version of the membership function. For $Z_{\text{dr}}$ and $K_{\text{dp}}$, this is achieved by defining the parameters $m, a$, and $b$ as functions of reflectivity $Z_h$. So for these two input variables

$$m = m(Z_{\text{h}}),\ a = a(Z_{\text{h}}),\ \text{and}\ b = b(Z_{\text{h}}). \tag{2.10}$$

In the MeteoClassifier, the functions presented in 2.10 are fourth degree polynomials. Two dimensional membership functions are needed because $Z_{\text{h}}$ and $Z_{\text{dr}}$, or $Z_{\text{h}}$ and $K_{\text{dp}}$, are not independent [10]. This is clearly illustrated in figure 2.11 which presents a $Z_{\text{dr}}, Z_{\text{h}}$ -scatter plot from real measurement data from a rain event.

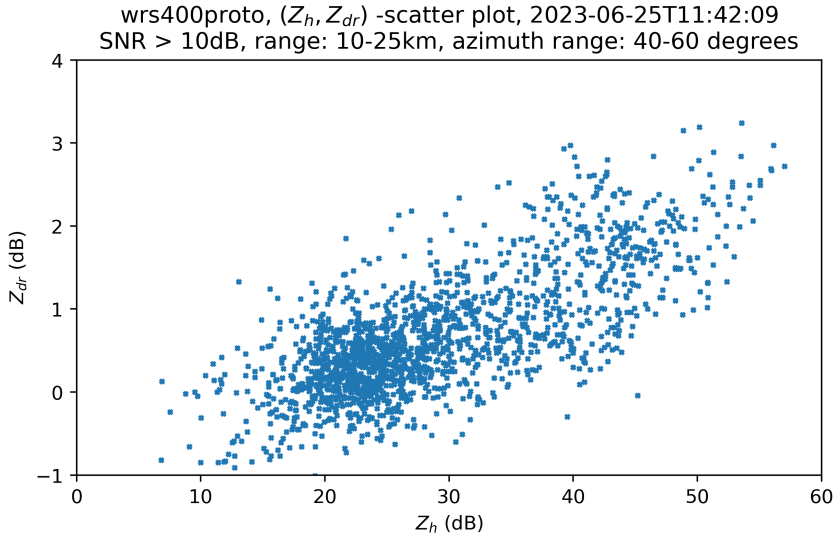



***Figure 2.11.*** *A scatter plot of measured $Z_{dr}$ and $Z_h$*

Figure 2.12 shows the two dimensional membership functions. The figures show the area where the truth value is above $0.5$. The tone of color illustrates the truth value. The darker the color the lower the truth value.

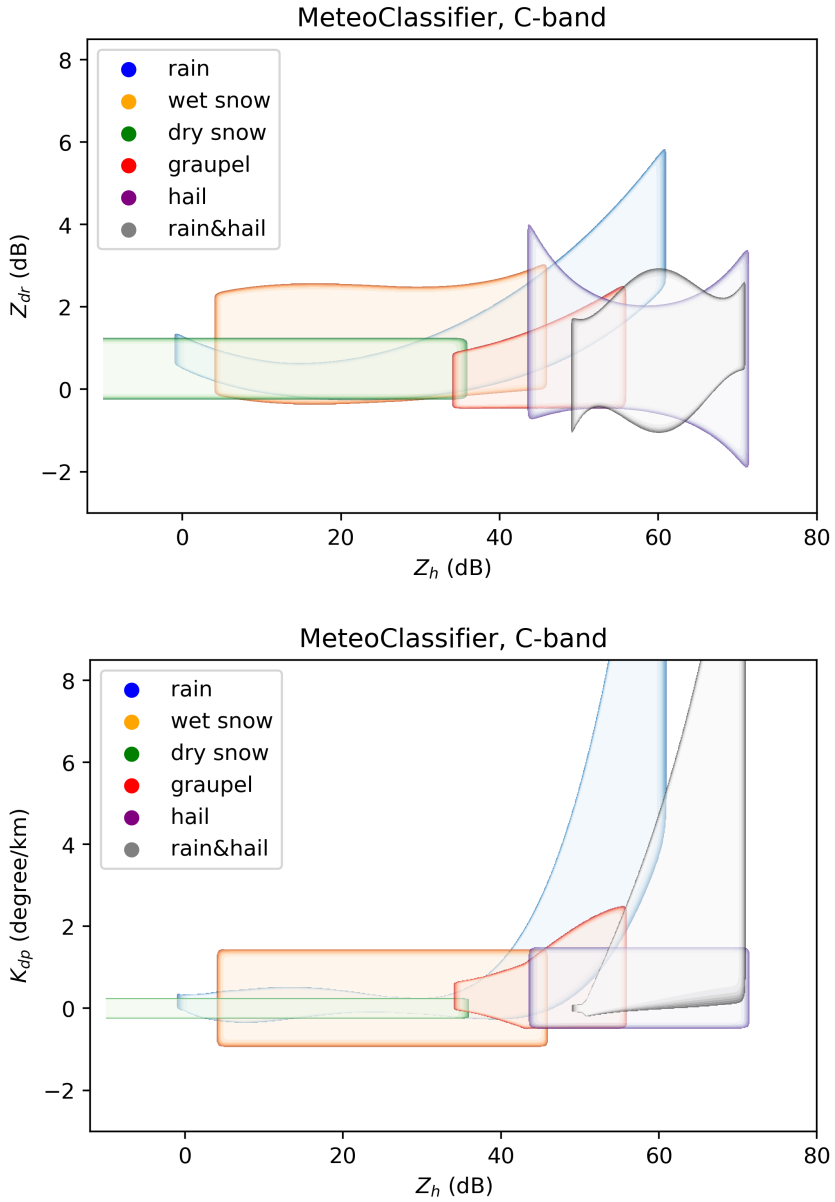



***Figure 2.12.*** *MeteoClassifier's two dimensional membership functions.*

The membership functions for altitude $h$ are also two-dimensional, but parameters are defined as functions of melting level height $h_{\text{ML}}$. So, $h_{\text{ML}}$ is not used in the same way as the other input variables, but it defines parameters for altitude's membership functions. Figure 2.13 presents membership functions for altitude when $h_{\text{ML}} = 2.5$ km. These functions shift linearly as the melting level height is changed.

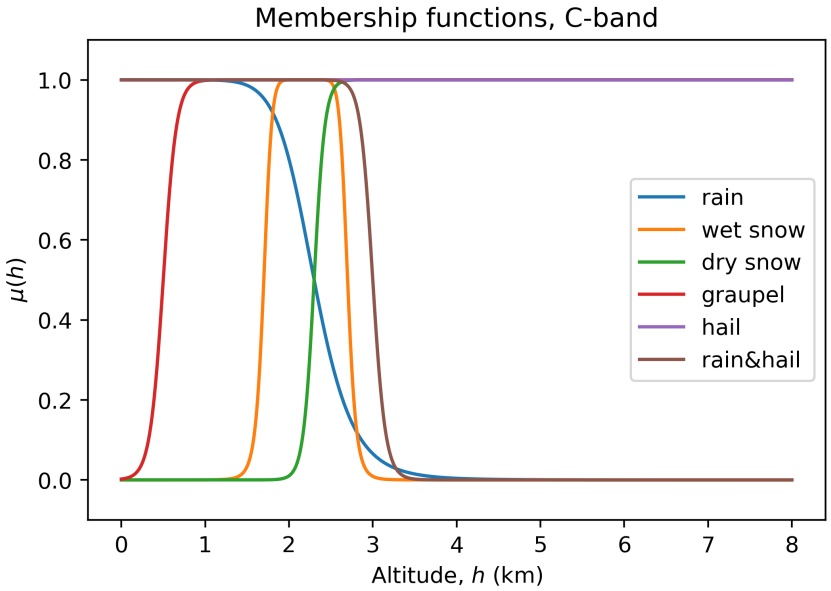



***Figure 2.13.*** *MeteoClassifier's membership functions for altitude,* $h_{ML} = 2.5$*km.*

Rule strengths are computed using summation and product hybrid formula adapted from the CSU method [9]. Coefficients in the formula are different for warm and cold seasons. Cold season is defined as the time when $h_{\text{ML}} < 0$. Rule strengths are computed using the formula

$$\text{RS}_{i,\text{warm}} = \mu_i(Z_\text{h}) \times \mu_i(h) \times \mu_i(Z_\text{dr}) + 0.5 \times \mu_i(K_\text{dp}) + 0.5 \times \mu_i(\rho_\text{hv})$$
$$\text{RS}_{i,\text{cold}} = \mu_i(Z_\text{h}) \times 0.7 \times \mu_i(h) \times \mu_i(Z_\text{dr}) + 0.3 \times \mu_i(K_\text{dp}) + 0.5 \times \mu_i(\rho_\text{hv}).$$

The cold season formula is applied whenever a negative melting layer height is given as an input. Aggregation and defuzzification are performed using the simple maximum rule presented in 2.8.

### 2.3.2 PreClassifier and PrecipClassifier

PreClassifier and PrecipClassifier are different versions of the hydrometeor classification method developed in the National Severe Storms Laboratory (NSSL) in Oklahoma [21]. The original NSSL method's membership functions were adjusted and performance was validated with data from the Joint Polarization Experiment (JPOLE). JPOLE took place from the spring of 2002 until the summer of 2003 on a testbed radar located in Oklahoma. It aimed to test and prove the capabilities of polarimetric weather radars [19].

The two classifiers have different purposes within HydroClass. PreClassifier, as the name suggests, works as a first check and quality control to detect range bins that contain signals from meteorological scatterers. Output classes of PreClassifier are GC/AP (Ground Clutter/Anomalous Propagation), BIO or biological targets, and METEO or meteorological scatterers. Only METEO range bins are then passed on to be classified by the other methods. On the other hand, PrecipClassifier's role is to determine the severity of the rain. Its output classes are light, moderate, heavy rain, and large drops.

Input variables for both PreClassifier and PrecipClassifier are reflectivity $Z_{\mathsf{h}}$, cross-correlation coefficient $\rho_{\mathsf{hv}}$, differential reflectivity $Z_{\mathsf{dr}}$ and differential phase $\varphi_{\mathsf{dp}}$. Differential phase is not used as is but the input variable is so called texture parameter of $\varphi_{\mathsf{dp}}$. To obtain the texture parameter, data is averaged over a running average window and then the smoothed estimates are subtracted from the original values. This results in the standard difference. Effectively, this means the formula

$$\mathsf{TX}(x) = \frac{\sum_{i=1}^{n} x_i}{n} - x, \tag{2.11}$$

where $n$ is the number of range bins included in the running average window. The texture parameter of the reflectivity $Z_{\mathsf{h}}$ is also used as an input. So, $Z_{\mathsf{h}}$ is used in two ways.

PreClassifier uses trapezoidal membership functions. Their formula is defined by

$$\mu_{ij}(x_j; a_{ij}, b_{ij}, c_{ij}, d_{ij}) = \begin{cases} 0 & \text{if } x_j \leq a_{ij} \\ \frac{x_j - a_{ij}}{b_{ij} - a_{ij}} & \text{if } a_{ij} < x_j < b_{ij} \\ 1 & \text{if } b_{ij} \leq x_j \leq c_{ij} \\ \frac{d_{ij} - x_j}{d_{ij} - c_{ij}} & \text{if } c_{ij} < x_j < d_{ij} \\ 0 & \text{if } x_j \geq d_{ij} \end{cases} \tag{2.12}$$

where $x_j$ is the input variable and $a_{ij}, b_{ij}, c_{ij}$, and $d_{ij}$ are the parameters are different for each input variable and output class combination. These parameters determine the shape of the function as presented in figure 2.14.

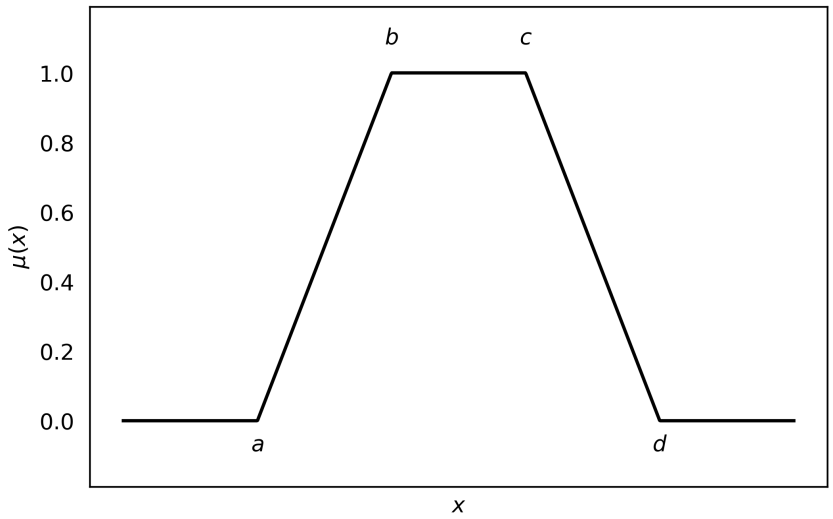



***Figure 2.14.*** *An example trapezoidal curve illustrating the meaning of parameters* $a, b, c,$ *and* $d$*.*

Similarly to the MeteoClassifier, $Z_{\text{dr}}$ requires two dimensional membership functions where parameters $a, b, c,$ and $d$ are functions of $Z_h$. PreClassifier's $Z_{\text{dr}}$ membership functions are presented in figure 2.15.

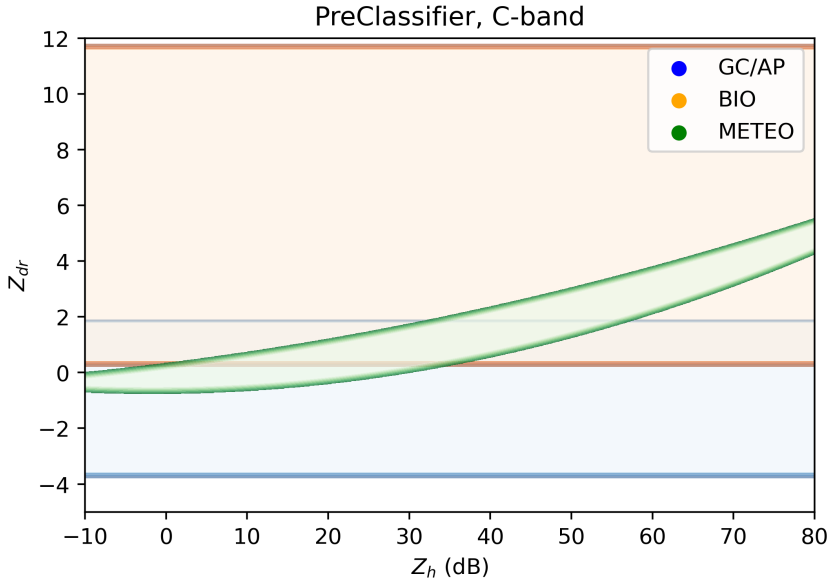



***Figure 2.15.*** *PreClassifier's membership functions for* $Z_{dr}$*.*

The other input variables have one dimensional membership functions following formula 2.12. The membership functions for $\rho_{\text{hv}}$ and $Z_{\text{h}}$ are shown in figure 2.16.

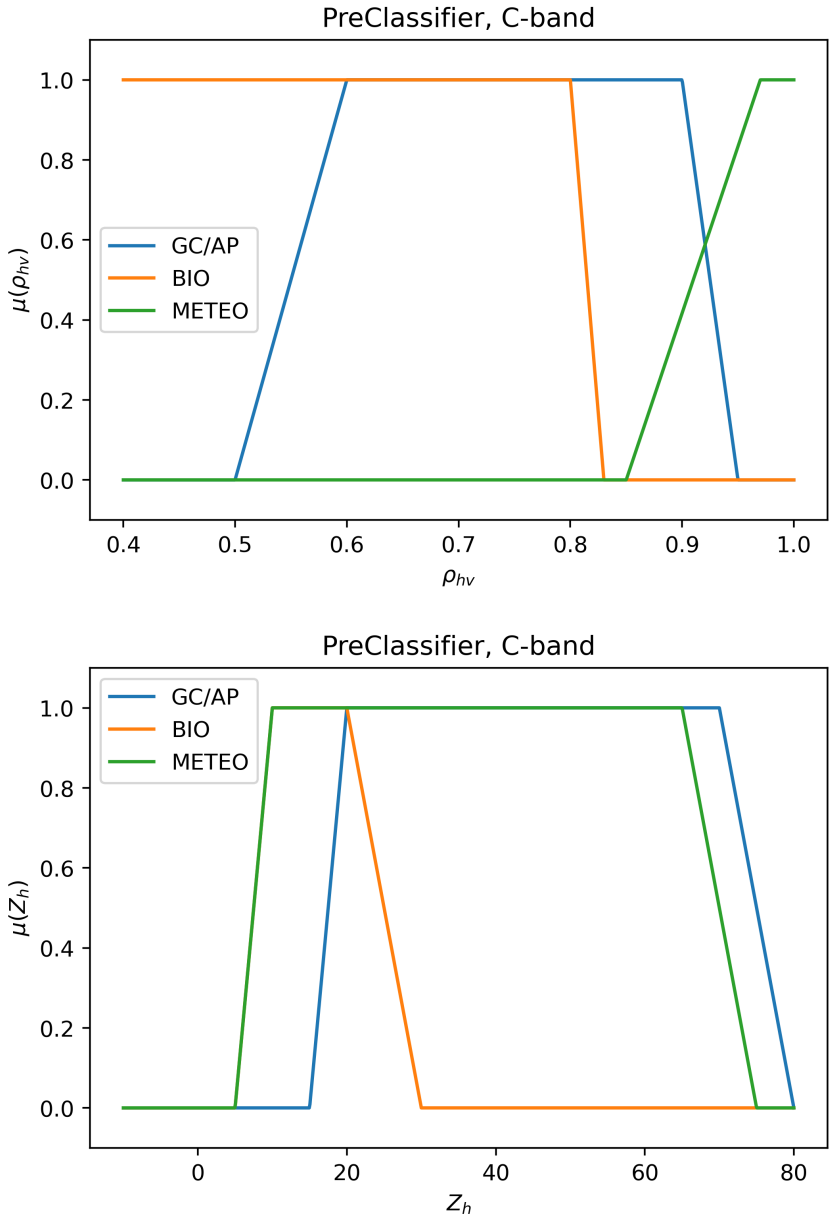



***Figure 2.16.*** *PreClassifier's membership functions for $\rho_{hv}$ and $Z_h$.*

The membership functions for both texture parameters are shown in figure 2.17.

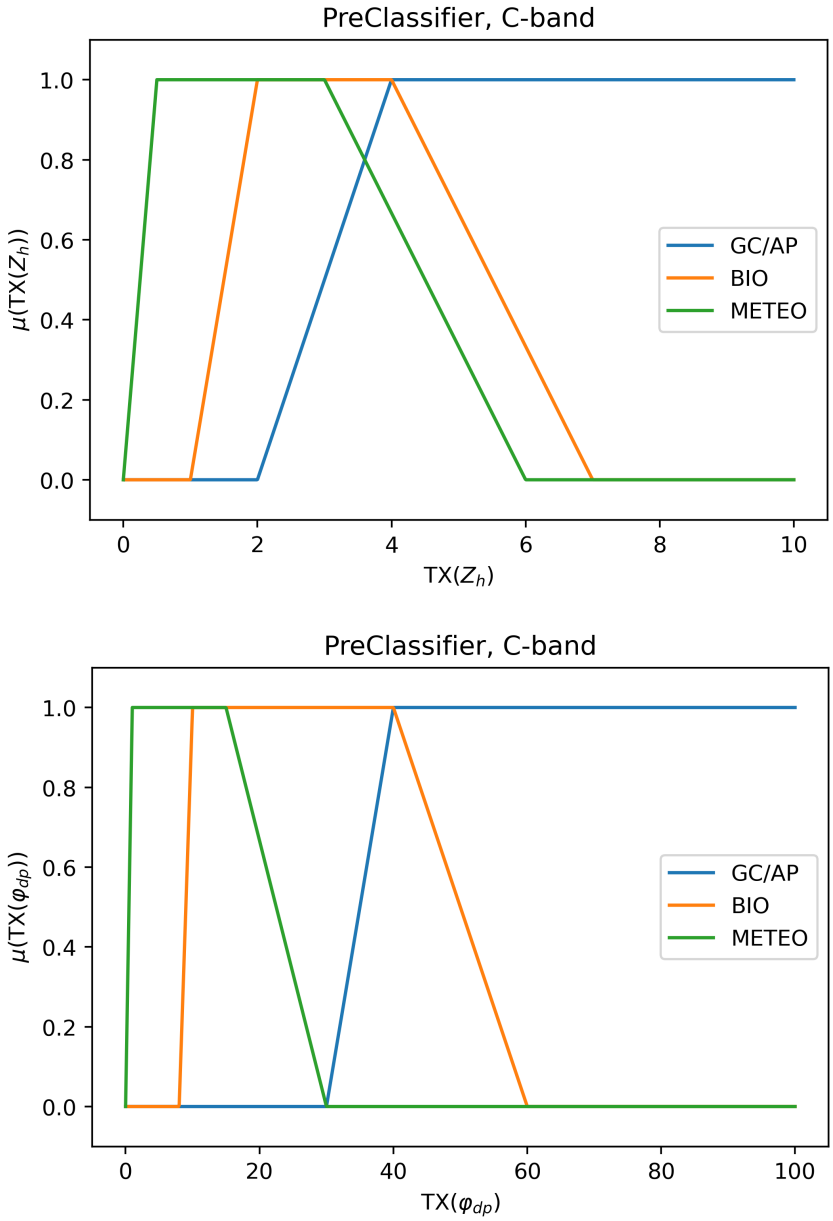



***Figure 2.17.*** *PreClassifier's membership functions for the texture parameters.*

PrecipClassifier classifies the members of the 'Meteo'-class into different classes based on the intensity of the precipitation. The current implementation of PrecipClassifier uses the same membership functions for $\rho_{\mathrm{hv}}$, $\mathrm{TX}(\varphi_{\mathrm{dp}})$, and $\mathrm{TX}(Z_{\mathrm{h}})$ as the 'Meteo'-class of the PreClassifier. Figure 2.18 presents the membership functions for reflectivity $Z_{\mathrm{h}}$.

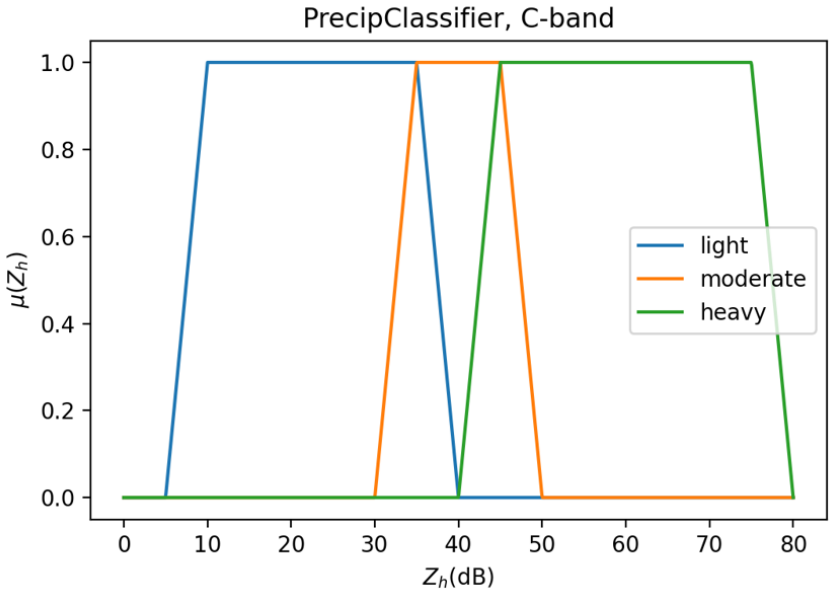



***Figure 2.18.*** *PrecipClassifier's membership functions for $Z_h$.*

Both PreClassifier and PrecipClassifier use simple additive rule strength formula

$$\text{RS}_i = \sum_{j=1}^{5} \mu_{ij}(x_j), \tag{2.13}$$

where $x_j$ is one of the input variables. Again, aggregation and defuzzification are performed using the simple maximum rule presented in 2.8.

## 2.4 Adjusting parameters for X-band radar

There are three different frequency bands that weather radars operate on, S-, C-, and X-band. Table 2.1 shows key figures of the frequency bands. The actual operational frequency and wavelength of an individual radar are chosen within the presented ranges. Typically, radar networks consist of S- and C-band radars. S-band's frequency is generally good for detecting rain because of the used frequency [14]. A lower frequency would produce weaker echoes from rain particles and a higher frequency would make the system prone to attenuation effects.

C-band is a bit more vulnerable to attenuation but the significantly smaller size of the device compared to S-band often makes it a more viable and cheaper option. X-band radars' usage is limited due to decreased measurement range and increased attenuation effects compared to S- and C-band [6]. But they are really useful for filling gaps in S- and C-band radar networks [18]. Due to their size, X-band radars can also be used in mobile radar applications.

***Table 2.1.*** *Weather radar bands and the corresponding frequency bands and wavelengths. Table is created based on reference [16].*

| Band Designation | Nominal Frequency | Nominal Wavelength |
|---|---|---|
| S | 2-4 GHz | 15-8 cm |
| C | 4-8 GHz | 8-4 cm |
| X | 8-12 GHz | 4-2.5 cm |

Both the CSU and NSSL methods were originally developed for S-band systems [10][21]. Over the years, various papers have been published on efforts to modify the methods to work on C- and X-band systems, and the efforts have been successful. The differences of the three frequency bands have been discussed on a theoretical level on multiple occasions [1][3][18]. Hydrometeor classification methods have also been implemented and field tested many times on X-band systems [8][15][20][34].

In literature references, the optimization of parameters is based on simulations that utilize observed drop size distributions. Based on the simulations, one can then create theoretical distributions for measured polarimetric variables on different frequency bands [3]. These distributions do indeed differ from one another which is quite logical as the wavelength of the signal does affect how it scatters from a target.

Generally, the two aspects that create the need for changing the parameters for different frequency bands are so-called resonance effects and already mentioned attenuation [18]. At shorter wavelengths, resonance and attenuation effects play a more significant role than they do in an S-band system. Next, these two phenomena are discussed in more detail.

### 2.4.1 Resonance effects

The following introduction of radar targets' scattering properties is based on reference [16]. The backscattering cross-sectional area is the size a target appears to the radar. Therefore, it also quantifies the amount of energy scattered back to the radar from a target. The backscattering cross-sectional area of a radar target is dependent on the shape, material, and size of the target. In addition, it is also a function of the wavelength of the radar observing the target. It is often reasonable to assume that meteorological targets are spheres. Of course, radar is not measuring individual targets but volume that includes numerous targets. For spherical targets that are large compared to the wavelength of the radar, the cross-sectional area is the same as the target's geometric areas. That is determined by the formula

$$A = \pi r^2, \tag{2.14}$$

where r is the radius of the target. A target is considered large if $D/\lambda > 10$. $D$ is the diameter of the target and $\lambda$ is the wavelength. However, if the size of the sphere is small compared to the wavelength a different formula has to be used. This region is called the Rayleigh region and the backscattering cross-sectional area can be computed using

$$A = \frac{\pi^5 |K|^2 D^6}{\lambda^4}, \tag{2.15}$$

where $|K|^2$ is a parameter related to the complex index of refraction of the material. A target can be considered small if $D/\lambda < 0.1$. A large number of targets that are measured by a weather radar fall into the Rayleigh region.

But, some targets are also in the region between the two mentioned regions. Determining the backscattering cross-sectional area in this region is complicated. This intermediate region is called Mie or resonance region. In this region, the backscattering cross-sectional area can actually decrease as the target's size increases. The size of the backscattering cross-sectional area oscillates approaching the behavior described in equation 2.14. Figure 2.19 presents a visualization of this phenomenon.

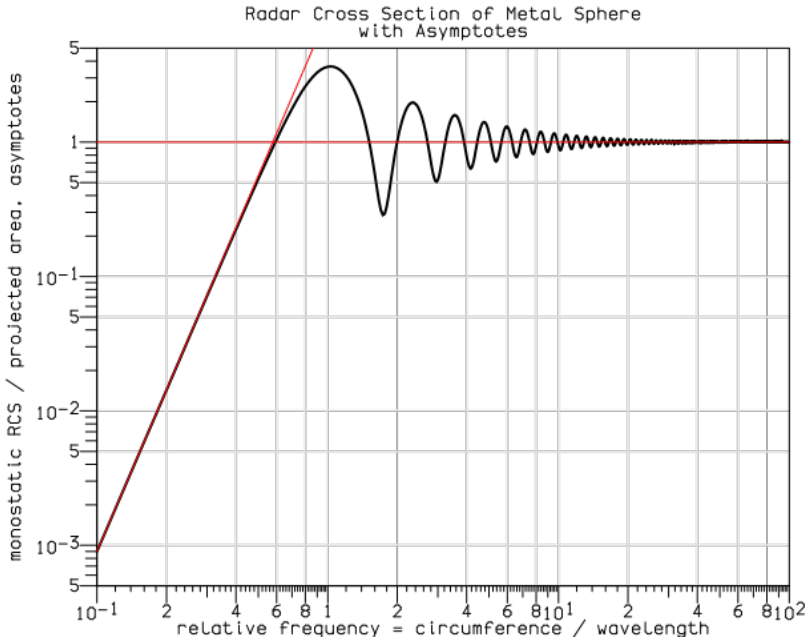



***Figure 2.19.*** *Backscattering cross-sectional area for a metal sphere as a function of relative frequency. The figure is based on reference [25] from open figure bank [32].*

It should be noted that figure 2.19 presents the scattering phenomenon for a metal sphere. Hydrometeors consist of water which makes the scattering dynamics even more complex as the parameter $K$ for hydrometeors is dependent on the signal's frequency [16]. For a metal object, this parameter can be considered to be a constant.

Because the scattering regions are tied to the wavelength used by the radar, the radars operating on different frequency bands encounter resonance effects of different magnitudes measuring different sized targets. This has to be taken into account in the membership functions. The physical background of the phenomenon is largely omitted in this work. But in short, the differences in the resonance effects for hydrometeors are caused by the differences in the imaginary part of the dielectric constant of water. The imaginary part of the dielectric constant is greater in the X-band than in the C-band, which causes the polarimetric data moments measured by an X-band to system behave closer to S-band [18].

The resonance effects affect the precipitation classes that include the largest particles which are liquid rain and hail. Their size can fall outside of the Rayleigh region [16][31]. Ryzhkov and Zrnic discuss and study the difference in radar polarimetry at different frequency bands in the article the *Radar polarimetry at S, C, and X bands, comparative analysis and operational implications* [18]. The analysis was based on simulated fields of the polarimetric data moments. Based on the study, the cross-correlation ratio at the C-band could drop significantly below $0.98$ in liquid rain while in the X-band $0.98$ can safely be used as a limit value. The Mie scattering phenomenon can also cause noticeable higher $Z_{\text{dr}}$ values in C-band than in S- or X-band systems. The same phenomenon can also cause negative $K_{\text{dp}}$ values in C-band. All in all, the resonance effects are more pronounced in the C-band compared to the X-band.

When it comes to hail, based on simulations that aim to model the described scattering mechanics, the membership functions for classifying hail should be adjusted so that the support region is narrowed for $Z_{\text{dr}}$ and widened for $K_{\text{dp}}$ in X-band systems compared to C-band [20]. In addition, the membership function for $\rho_{\text{hv}}$ can be modified to allow higher values in the X-band than in C-band. The modifications that are made in this work are presented in chapter 4.1.

### 2.4.2 Attenuation

As electromagnetic radiation passes through a medium, a part of the signal's power is lost due to absorption and scattering. Power loss caused by absorption refers to the energy of the electromagnetic signal that is absorbed into the particles of the atmosphere. Power loss caused by scattering refers to the energy of the transmitted signal that scatters away from the wanted transmission direction. These two phenomena together are called attenuation. The description of the attenuation phenomenon given in this chapter is based on reference [16]. The amount of attenuated power is dependent on the material of the object and its properties.

There is always some level of attenuation as the signal transmitted by the radar travels through the atmosphere. However, the attenuation effects caused by air, clouds, or even snow are so insignificant that they can typically be neglected. The attenuation caused by liquid rain, on the other hand, must be taken into account.

The strength of the attenuation is dependent on the signal's frequency. The higher the frequency is the stronger the attenuation is as well. Therefore, in X-band systems attenuation effects are more intense compared to C- or S-band. For example, in the revised version of the CSU method described in reference [9], middle point parameter $m$, defined in formula 2.9, is set to $m = 50$ for heavy rain while the same parameter is set to $m = 47.5$ in X-band system. All in all, the support regions, defined in definition 2.3, are extended in the case of the liquid rain classes of the revised CSU method.

Because of the attenuation, the reflectivity $Z_{\mathrm{h}}$ limits of membership functions have to be loosened. There are also methods to correct the attenuation effects in the reflectivity data. Attenuation correction in dual polarization systems is based on differential phase $\varphi_{\mathrm{dp}}$ and specific differential phase $K_{\mathrm{dp}}$. These data moments are used because they are not dependent on radar calibration accuracy [30].

### 2.4.3 Adjustment process

The adjustment of the HydroClass' parameters in this work is based on a literature review. The start point of the process is the current factory settings for the Vaisala C-band weather radar WRM200. The parameters to be adjusted are the membership function parameters $m, a$, and $b$ in MeteoClassifier and parameters $a, b, c$, and $d$ in PrecipClassifier. As mentioned before, in the current implementation only the reflectivity value separates the different classes in PrecipClassifier. Therefore that is the only part of the PrecipClassifier that is adjusted in this work. Although, it should be mentioned that according to the references, there are other differences as well [2]. Different polarimetric characteristics are reported from the rain with different intensities [3][20]. For example, typically $Z_{\mathrm{dr}}$ reaches higher values in heavy precipitation than in light precipitation. This was also visible in figure 2.11 that showed $Z_{\mathrm{dr}}$ as the function of reflectivity $Z_{\mathrm{h}}$. PreClassifier's membership functions don't require adjusting as the targets that are identified, like clutter targets and biological targets, have very different characteristics compared to meteorological targets.

All references used in the optimization use either the trapezoidal formula 2.12 or the CSU formulation 2.9 in their methods' membership functions. When it comes to the range of output classes, references are a little different from HydroClass as they often include more classes. Therefore, the parameters in literature references cannot be used as such but the results have to be adapted to fit the HydroClass classification scheme using the general guidelines that are apparent in all adapted references. In chapter 4, the adjusted

membership functions are presented.

After the parameters have been adjusted based on the literature references, the next step is to verify that the algorithm does indeed work better after the adjustments. The next chapter introduces the data and methods that are used in the performance validation of the X-band adjusted HydroClass.

# 3. PERFORMANCE VALIDATION

Performance validation of the adjusted HydroClass is done by comparing results from classification based on initial C-band specific parameters to classification done using new adjusted parameters for the X-band system. Weather radar data used in the analysis is from the Vaisala WRS400 polarimetric weather radar located in Vantaa, Finland. To have a reference point that is not based on a radar measurement, in situ, ground measurements from a Vaisala forward scatter sensor FD70 are used. Visual comparison of the results is also performed based on PPI and RHI plots.

Next, the data set and the validation process are described. Instruments that were used to measure the data are also briefly introduced as well.

## 3.1 Data

The data used in the analyses of this work was collected using a WRS400 X-band weather radar and an FD70 forward scatter sensor, both located in Vantaa. All in all, four different weather cases are included in the analysis to study how the HydroClass algorithm behaves in different scenarios.

### 3.1.1 Weather radar WRS400

Vaisala WRS400 is a polarimetric X-band weather radar using solid state transmitter technology. The radar used in this work is located at the rooftop of Vaisala headquarters in Vantaanlaakso, 12km north of Helsinki. It is located 35 meters above sea level. The radar on top of the Vaisala headquarters is shown in figure 3.1. Northwest and southwest of the radar, there are trees that block sectors of the radar scan. This is visible as blank sectors in visual examples presented in chapter 4.3.

***Figure 3.1.*** *WRS400 on the roof of Vaisala headquarters in Vantaa.*

The measurement task used to collect the data is a volume scan. This describes a measurement scheme where the radar rotates full 360 degrees using multiple elevation angles. The lowest elevation angle of the volume scan is 1 degree and the highest is 10 degrees. Range resolution is 150m meters and the signal transmitted by the radar has a beam width of $1^{\circ}$. The measuring task was run on a four-minute interval.

***Figure 3.2.*** *3D model of X-band weather radar WRS400.*

In general, validating the performance of a weather radar is not an easy task. Because of the nature of weather radar measurement, it is practically impossible to validate per-

formance in a closed and controlled environment. The validation has to be done based on available real weather. Performance can then be analyzed by, for example, analyzing distributions of the measured data moments and comparing measured values with observations made using other weather radars and ground sensors. In this case, the performance validation is done by comparing the classification given by the WRS400 at Vantaanlaakso to in situ observations made using an FD70 forward scatter sensor.

### 3.1.2 Forward scatter sensor FD70

Vaisala FD70 is a forward scatter sensor that is capable of accurately classifying detected hydrometeors. The sensor transmits a very thin light sheet, which in combination with high sample frequency allows for analysing each droplet individually. FD70 can measure size, velocity, and other scattering properties like intensity [7]. With this data, precipitation type can be determined.

***Figure 3.3.*** *3D model of a FD70 forward scatter sensor.*

Because of the above-mentioned properties, FD70 provides good in situ observations that can be used as a reference point for WRS400 and HydroClass. However, one has to keep in mind that FD70 measures and observes the precipitation on the ground level while the weather radar measures the atmosphere tens, hundreds, or even thousands of meters above ground depending on the distance from the radar. So there is a significant mismatch in the measurement volumes. The specific FD70 sensor that was used in this work is located at Helsinki-Vantaa airport, approximately 6.65 kilometers from WRS400 in Vantaanlaakso at $42^{\circ}$ azimuth. Therefore, the center of the radar beam is approximately

150 meters above ground at the lowest elevation angle of the radar's volume scan. Of course, there is a huge difference in the sizes of the measurement volumes too. It is reasonable to assume that conditions at that height are similar to those on ground level but in some cases, wind and temperature differences could cause differences in the weather conditions observed by the two sensors.

### 3.1.3 Data set

The data set of this work includes 5 different weather events. Each of them with different weather conditions in order to analyze the HydroClass' performance in different scenarios. The events are described in table 3.1. The times presented in the table are all in Universal Coordinate Time (UTC).

***Table 3.1.*** *Summary of the test cases.*

| Time frame (UTC) | Melting layer height | Event description |
|---|---|---|
| 2022-08-28<br>16:00-17:00 | $\approx 3.2$ km | Hail and heavy rain |
| 2023-03-27<br>02:00 – 08:00 | $< 0$km | Snowing |
| 2023-03-27<br>18:00 – 23:15 | $< 0$km | Snowing, mixed with frozen and liquid precipitation |
| 2023-04-26<br>18:00 – 23:15 | $\approx 1.7$ km | Moderate and light rain |
| 2023-06-25<br>11:00 – 13:00 | $\approx 2.5$ km | Short and intense raining event including some graupel |

Melting level height data is from a website hosted by the University of Wyoming [28]. The website includes data from numerous sounding sites. The data used in this work is from a sounding site in Tallinn, which is the closest available site from WRS400 in Vantaanlaakso. It is located approximately $100$ km south of the Vantaanlaakso radar site. The sounding data is from midnight so there is a possibility that the circumstances in the atmosphere have changed between the radar and sounding measurements.

The first and last test cases listed in table 3.1 are used to visually compare the hail and graupel detection with the X-band adjusted parameters compared to the old C-band parameters. Data from the Finnish Meteorological Institute's operational radar at Vihti is also used in the visual comparison. This can be assumed to be a reliable comparison point for as the data is from a governing agency and the data is also used for, for example, creating weather forecasts and warnings in Finland. In general, FMI's data is open for anyone [26]. However, the hydrometeor classification data is only available for the past

6 days so the data used in this work cannot be retrieved from the site anymore. For the same reason, the FMI's data from the hail event on 2022-08-28 could not be used but the event was reported by the Finnish media. For example, *Helsingin Sanomat* reported large hail in Myllypuro, Helsinki [5].

The three other test cases are used in a time series comparison between the FD70 sensor and the WRS400 X-band weather radar. The methods used for both the visual and the time-series comparisons are presented next.

## 3.2 Methods

In this work, the X-band adjusted HydroClass is analyzed using two methods: a comparison with FD70 time series data and a visual inspection and comparison of the results.

### 3.2.1 Time series comparison between the FD70 and the HydroClass

HydroClass is compared with FD70 hydrometeor classification by plotting time series data from both sensors. From radar data, the range bin directly above the sensor is picked out and its hydrometeor class is reported.

Generally, algorithms that consider each bin individually are prone to misclassifications of single bins [9]. These misclassifications can be caused by individual noisy measurements and therefore they don't necessarily mean that the algorithm is not working correctly. These misclassified bins are easy to pick out by a human eye from PPI plots, for example. This is, of course, a significant weakness of these kinds of methods but it is not the main aim of this work to tackle this problem. However, adjusting the membership functions does decrease the number of these misclassifications too. To minimize the effect of these kinds of misclassifications in the analysis, a $3 \times 3$ window around the sensor is picked out instead of a single bin.

The range bins are weighted differently in the analysis depending on their distance from the middle range bin which is the one above the FD70 sensor. The weights are computed using the formula

$$w_{ij} = \begin{cases} \frac{1}{{d_{ij}}^{\alpha}} & \text{if } d_{ij} > 0 \\ 1 & \text{if } d_{ij} = 0, \end{cases} \tag{3.1}$$

where $d_{i,j}$ is the distance between the middle point of the range bin and the middle point of the center range bin and $\alpha$ is a parameter that controls the strength of the weight scaling. The formula 3.1 is based on reference [23]. As we are dealing with a $3 \times 3$ window, $i, j \in \{1, 2, 3\}$ and for center bin $i = 2$ and $j = 2$. In this work, the control

parameter is set to $\alpha = 0.2$. The value was chosen because it offers a good balance to the weights. Figure 3.4 shows how a slightly different control parameter value would alter the weights.

Using the multiplicative inverse of the distance as the starting point of the weight computation, allows us to take the surrounding bins into account in the analysis while giving more weight on the center bin and the bins closest to it. In this case, as the center bin is $6.6$ km from radar, this means the bins that are at the same range but at adjacent azimuth angles.

The weights are then collected to a one dimensional vector $v$ in which each index corresponds to an output class. The elements of the vector are computed by

$$v_k = \sum_{i=1}^{3} \sum_{j=1}^{3} w_{ij} \cdot [c_{ij} = k], \tag{3.2}$$

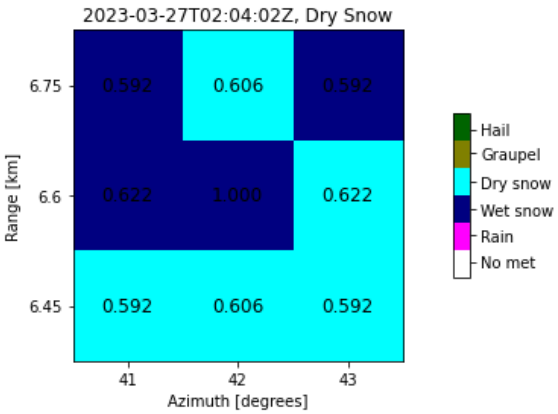
where $c_{ij}$ is the MeteoClassifier class given to the range bin. In other words, $c_{ij} \in \{$'No Met', 'Rain', 'Wet Snow', 'Dry Snow', 'Graupel', 'Hail'$\}$ and

$$[c_{ij} = k] = \begin{cases} 1 & \text{if } c_{ij} = k \\ 0 & \text{if } c_{ij} \neq k. \end{cases} \tag{3.3}$$

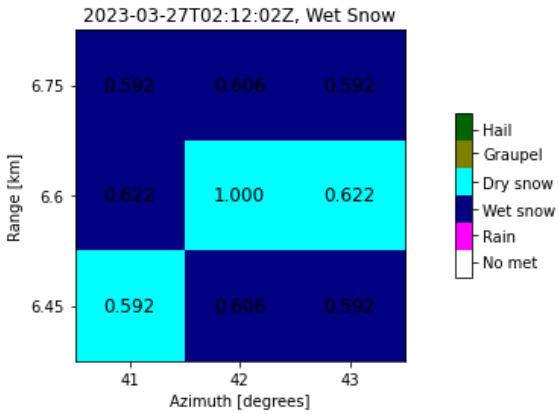
The final output class given to the window is then determined by simply picking out the maximum value from the vector $v$ and the corresponding class. This is given by

$$c_{\text{window}} = \arg\max_k (v_k). \tag{3.4}$$

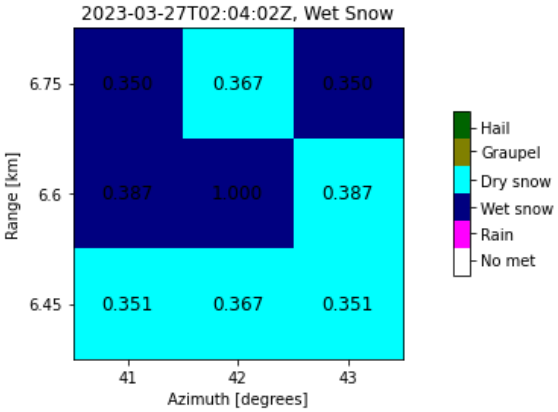
Figure 3.4 presents examples of how the method works in action. Examples presented in figure 3.4 are from real weather data measured by the WRS400 in Vantaanlaakso.



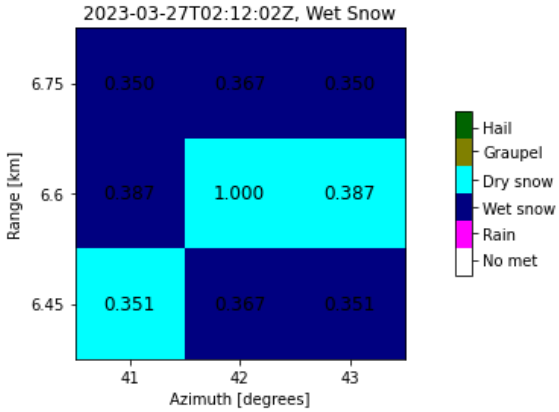

a) $\alpha = 0.1$, center result overturned



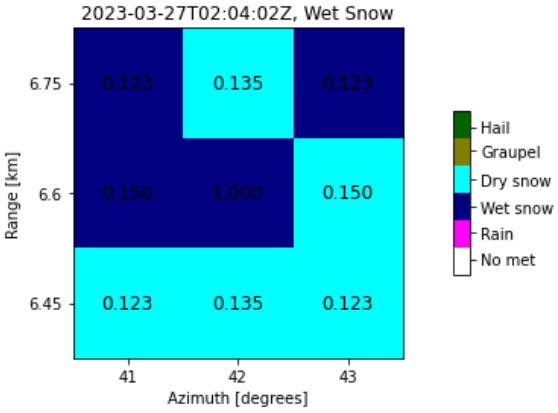

b) $\alpha = 0.1$, center result overturned



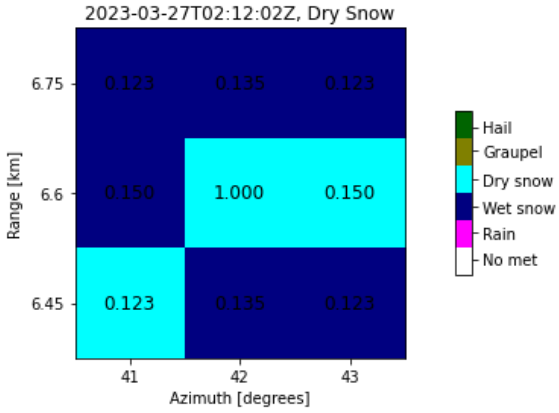

a) $\alpha = 0.2$, center result overturned




b) $\alpha = 0.2$, center result not overturned




a) $\alpha = 0.4$, center result not overturned




b) $\alpha = 0.4$, center result not overturned

***Figure 3.4.*** *Examples of classifying a $3 \times 3$ window. The final output class of the window is shown on top of the figure. The weights of individual bins are also shown.*

In the analysis, the results from HydroClass with C-band parameters are compared with HydroClass with X-band adjusted parameters. The hydrometeor classification data from the FD70 forward scatter sensor acts as the ground truth to which both HydroClass results are compared. Results from the described analysis are presented in chapter 4.2.

### 3.2.2 Radar data visualizations

Two types of visualizations are presented in this work: PPI and RHI plots. A PPI or a plan position indicator plot is a map-like visualization showing the radar in the middle with the range increasing outwards. The values of data moments are indicated by colors. The other visualization type is an RHI or a range height indicator. This type of visualization shows distance on the horizontal axis and height on the vertical axis. So, a PPI is a polar visualization and a RHI is a cartesian visualization. Chapter 4.3 presents many examples of both visualization types.

Visual inspection of the data allows a qualitative comparison between the old parameters and the new adjusted parameters. We can also see if the new parameters produce results that are more in line with expected hydrometeor types. For example, aspects such as whether the melting layer is visible in the data, indicating a region where frozen hydrometeor types, mainly dry snow, transition to wet snow and eventually to liquid rain, can be analyzed. We can also observe the individual rain pixels in the two cold season test cases mentioned in table 3.1 and investigate if they form a clear cluster or if they are just individual misclassifications.

To add a quantitative aspect to the analysis, the numbers of different hydrometeor types appearing in the PPI figures are reported as well. This will give an illustration of how significant the change is with respect to the number of different classifications. However, it should be noted that this analysis doesn't show if the changes are for the better as we are missing a reliable source of ground truth data for the entire volume measured by the radar. By visually inspecting the data we can evaluate if the results seem realistic or not. For example, a coherent region of one hydrometeor type is certainly more realistic than individual pixels here and there.

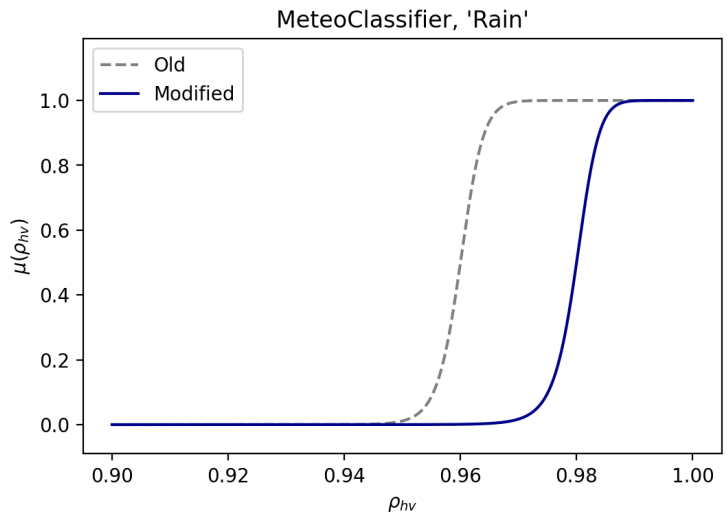
# 4. RESULTS

In this chapter, the results from the data and analysis methods described in the previous chapter are presented. The actual changes that were made to the membership functions of the MeteoClassifier and the PrecipClassifier based on the theory presented in 2.4 are also presented.

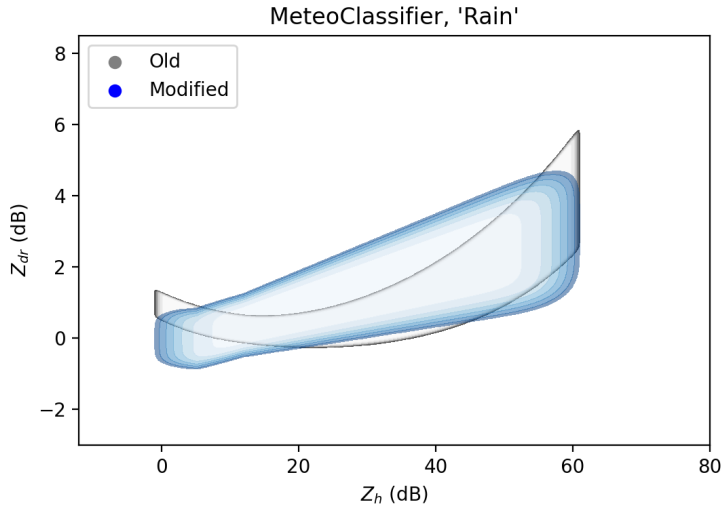
## 4.1 Changes to the membership functions

Due to the phenomena described in chapter 2.4, changes are made to the MeteoClassifier's membership functions of the 'Rain' and 'Hail' classes. Both of these hydrometeor types may have large particles that land outside of the Rayleigh region and therefore be affected by the resonance effects. The modified membership functions are presented next to the initial functions to highlight the changes. The two dimensional functions are presented in a similar fashion as in chapter 2.3. In other words, the figures show the area where the truth value is above $0.5$ and the tone of color illustrates the truth value. The darker the color the lower the truth value.

Figure 4.1 presents the changes made to the 'Rain' class' membership functions. The changes are in line with the observations from the literature mentioned in chapter 2.4. The support region of the $\rho_{\text{hv}}$ membership function is narrowed. Similarly, the highest values included in the fuzzy set of $Z_{\text{dr}}$ are lower at the higher end of the reflectivity spectrum. The boundary region is also a bit wider than before.



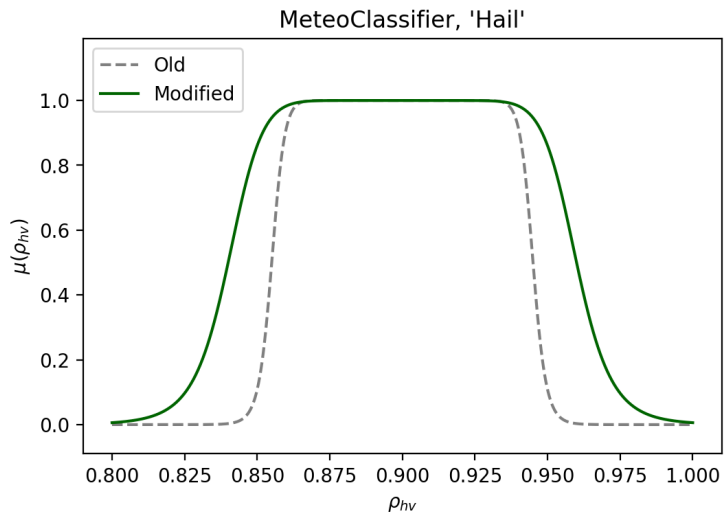

a) Membership functions for $\rho_{\text{hv}}$



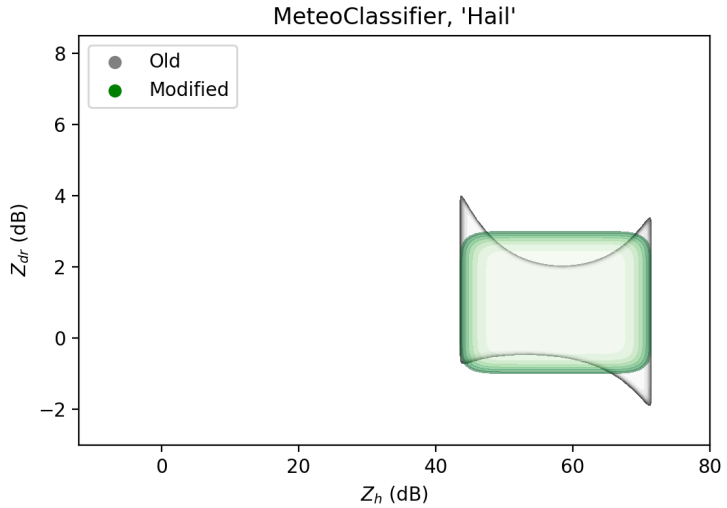

b) Membership functions for $Z_{\text{dr}}$

***Figure 4.1.*** *The adjusted membership function for Meteoclassifier's 'Rain'-class*

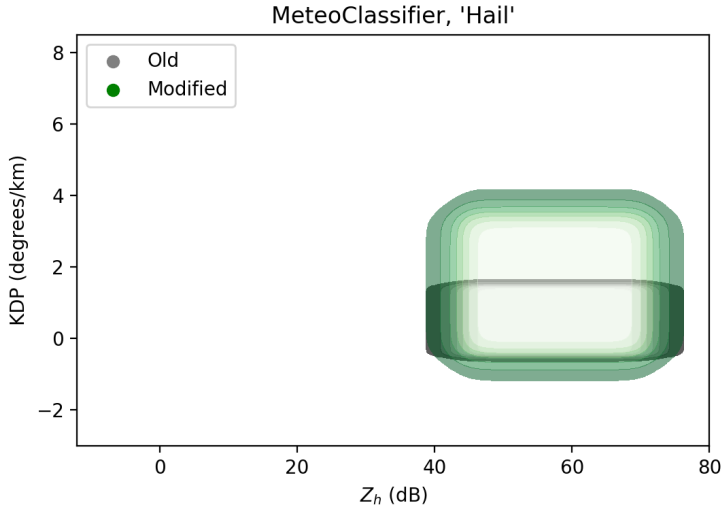
Figure 4.2 presents the changes made to the 'Hail' class' membership functions. Again, the changes are made based on the aspects noted in chapter 2.4. The support region of the $\rho_{\text{hv}}$ membership function is widened to allow higher values than with the old parameters. The $Z_{\text{dr}}$ membership function is less curved and the support region for $K_{\text{dp}}$ is widened. The adjusted $Z_{\text{dr}}$ membership function is now more in line with literature because most of the references used in this work use constant support region for hail [2][8][34]. The modification also leaves the highest and the lowest values of the initial membership functions' support region outside of the support region.




a) Membership functions for $\rho_{\text{hv}}$



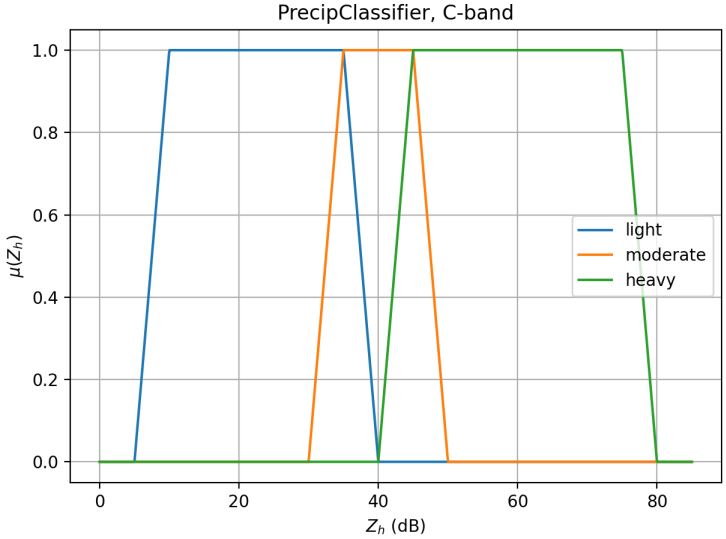

b) Membership functions for $Z_{\text{dr}}$



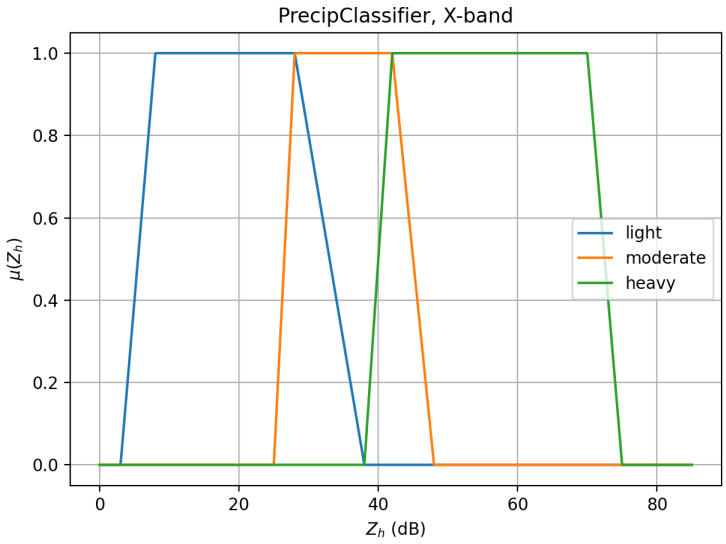

c) Membership functions for $K_{\text{dp}}$

***Figure 4.2.*** *The adjusted membership function for the Meteoclassifier's 'Hail'-class.*

Due to the attenuation effects, the PrecipClassifiers $Z_{\mathrm{h}}$ membership functions are shifted towards the lower end of the spectrum. Figure 4.3 presents the changes made to the PrecipClassifier's membership functions next to the old membership functions.




a) Old membership functions




b) Adjusted membership functions

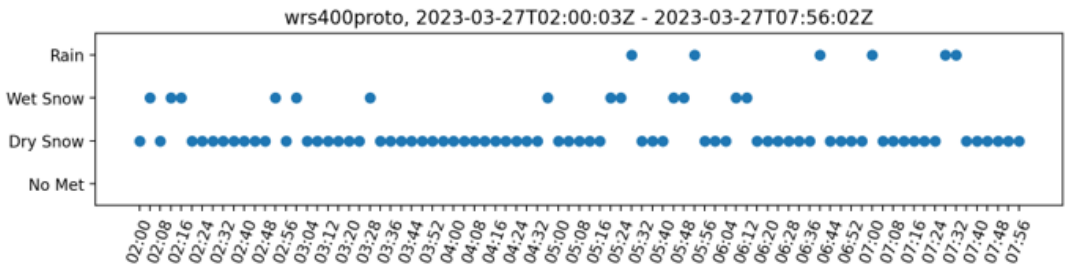
***Figure 4.3.*** *The adjusted membership functions for the PrecipClassifier.*

As the PrecipClassifier classification is just based on differences in the reflectivity values, these changes are not analyzed further. The focus is on the changes in the MeteoClassifier. Next, the results from the MeteoClassifier after the adjustments are presented.

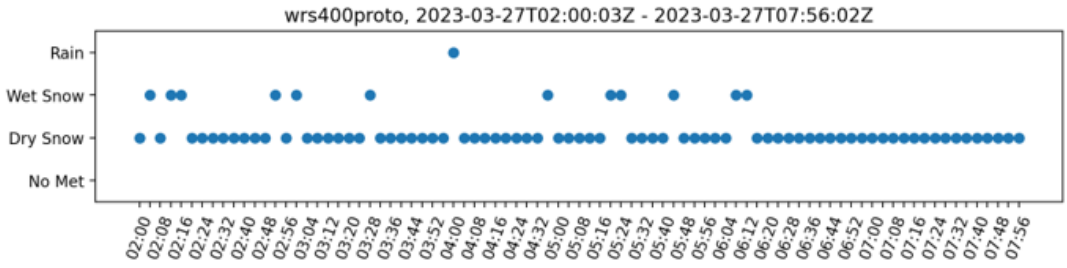
## 4.2 Results from the time series comparison

The data from three test cases, 2023-03-27 (UTC: 02:00 - 08:00), 2023-03-27 (UTC: 18:00 - 23:15), and 2023-04-26 (UTC: 18:00 - 23:15), are used in order to analyze hydrometeor classification results from the HydroClass algorithm and the FD70 forward scatter sensor. From the FD70, the rain type reported by the sensor is shown. For the HydroClass algorithm, the reported class is determined by the method described in chapter 3.2.

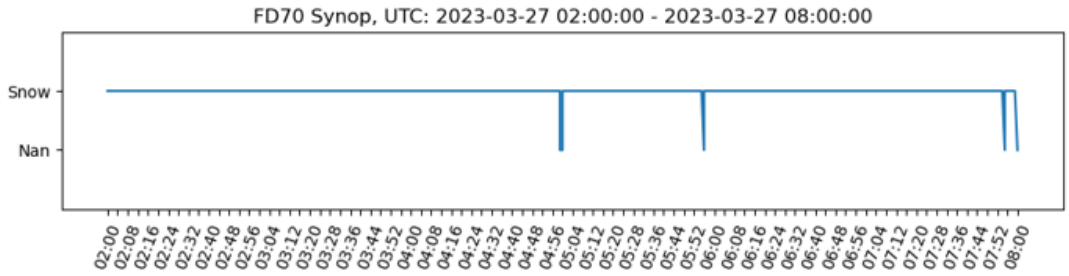
Figure 4.4 shows the results from the test case 2023-03-27, UTC: 02:00 - 08:00. The ground truth measurement of the FD70 shows snow for the whole time interval.




a) HydroClass with old parameters




b) HydroClass with modified parameters



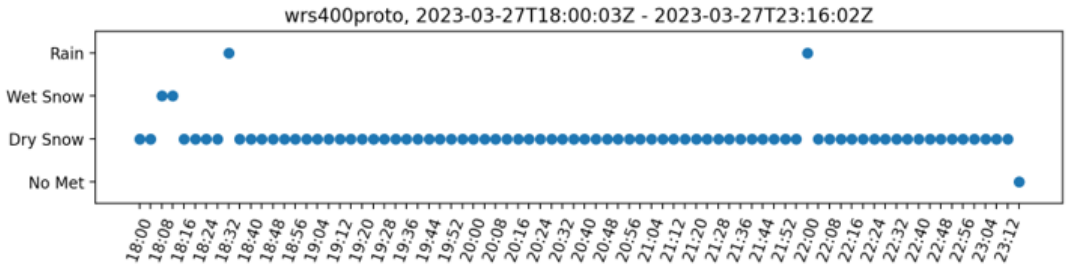

c) FD70 classification results

***Figure 4.4.*** *Classification results for the test case 2023-03-27, UTC: 02:00 - 08:00.*

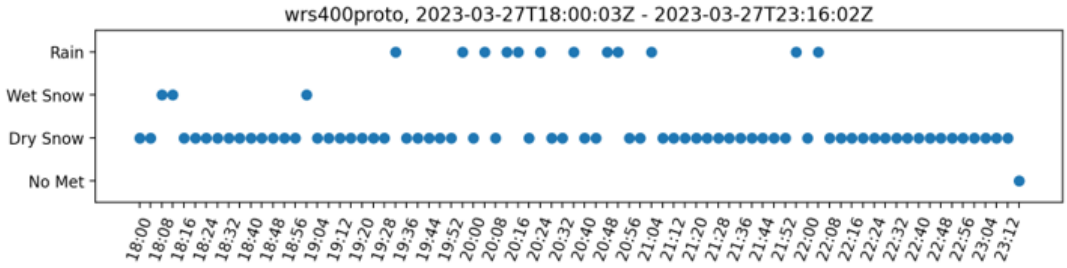
FD70 doesn't discriminate between wet and dry snow, but the data clearly shows that there was no liquid rain at any point. The results from the old C-band parameters show

six misclassified points where dry snow is reported as rain. The results from the modified parameters show a significant decrease in the number of misclassifications as there is only one of them in this case.

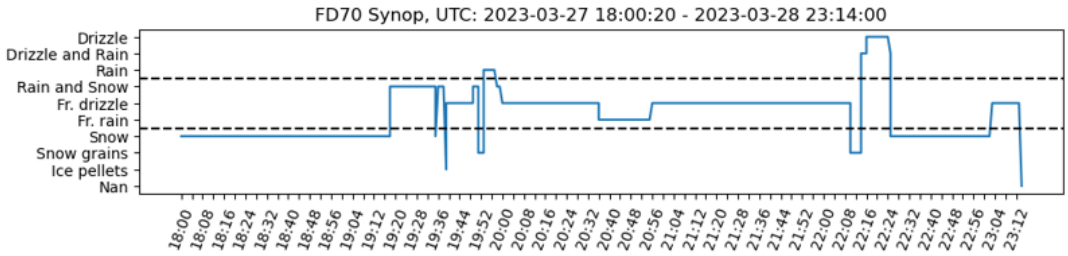
The second test case, 2023-03-27, UTC: 18:00-23:15, includes various different hydrometeor types reported by the FD70 sensor. This includes, for example, frozen drizzle, rain and snow mixture, and rain in liquid form.




a) HydroClass with old parameters




b) HydroClass with modified parameters



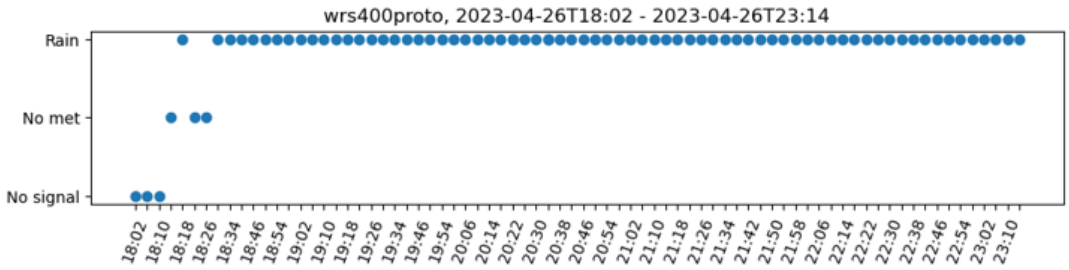

c) FD70 classification results

***Figure 4.5.*** *Classification results for the test case 2023-03-27, UTC: 18:00 - 23:15. The dashed lines in the FD70 plot separate solid, frozen, and liquid hydrometeor classes from each another.*

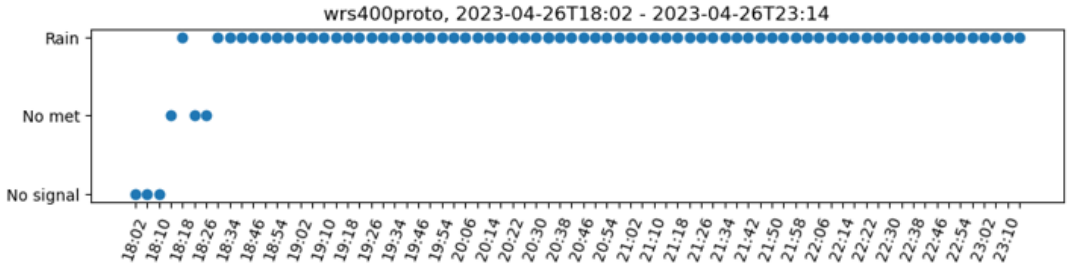
This is a difficult case for the HydroClass algorithm as there is no class that would exactly correspond to the frozen hydrometeor types. Figure 4.5 presents the results from this case. HydroClass with the old parameters doesn't seem to react to the frozen hydrometeors at all and classifies them all as dry snow apart from one point. Also, there is

one misclassified point at the start of the interval which appears with the old parameters but not with the modified parameters. The results from the modified parameters clearly highlight the time interval where FD70 reported the mixed hydrometeor types as there 12 points classified as rain. None of them land in into the time interval where FD70 reported standard snow. All in all, the results have improved in this case too. Human observer looking at the HydroClass results could notice that the precipitation is not just dry snow but with the old parameters such observation would not be possible.

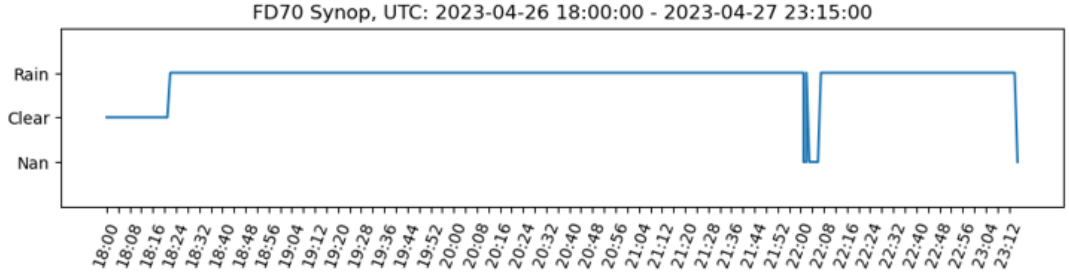
The last test case, 2023-04-26, UTC: 18:00 - 23:15, is a long-lasting light and moderate raining event. In this case, there is no difference between the old and modified parameters. Even the old parameters manage to classify this event correctly. The results are shown in figure 4.6.




a) HydroClass with old parameters



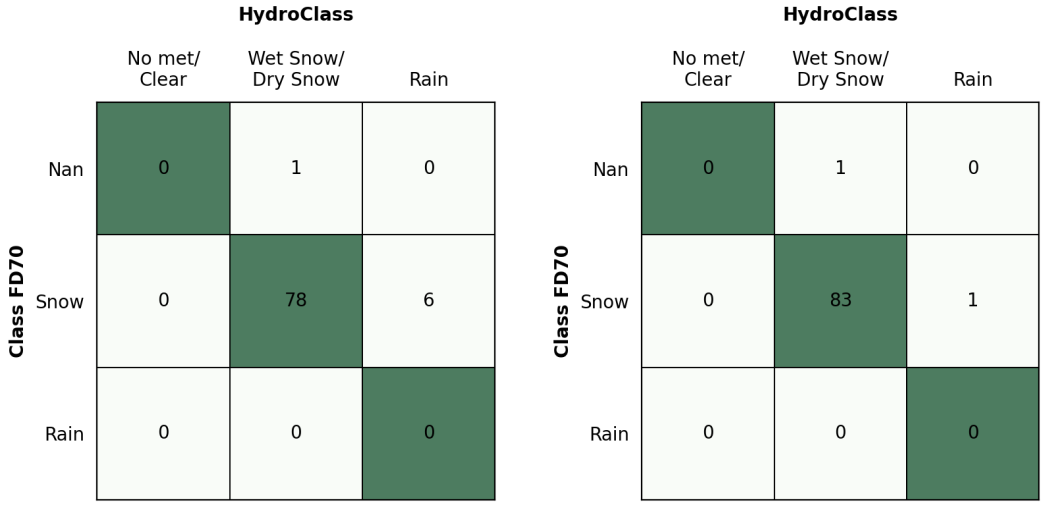

b) HydroClass with modified parameters




c) FD70 classification results

***Figure 4.6.*** *Classification results for the test case 2023-03-27, UTC: 18:00 - 23:15.*

Figure 4.7 presents how well the HydroClass matches the classification given by the FD70 for the test case from 2023-03-27, UTC: 02:00 - 08:00. For the FD70, the value chosen for the comparison is an average over three data points with time stamps closest to the time of the corresponding radar measurement. The values that fall into the diagonal axis are correct classifications and outside of the diagonal are misclassifications and false alarms. This shows that with the old parameters there are 7 points outside of the diagonal while there are only 2 such points with the modified parameters.




a) HydroClass with old parameters b) HydroClass with modified parameters

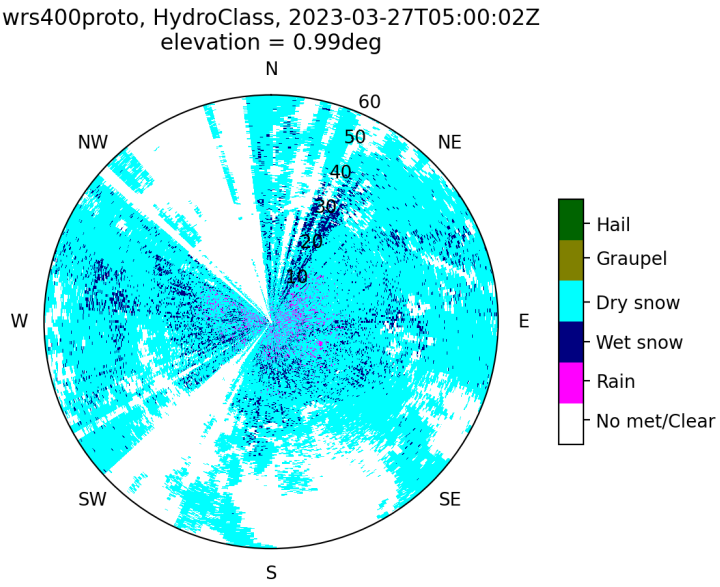
***Figure 4.7.*** *Classification comparison table with the FD70 for a) old parameters and b) X-band adjusted parameters.*

For the other two test cases similar mode of visualization doesn't provide meaningful information. For test case from 2023-04-26, UTC: 18:00 - 23:15, both parameters provide the same result. For the test case from 2023-03-27, UTC: 18:00 - 23:15, HydroClass doesn't have corresponding classes to the classification for some of the classes, such as freezing drizzle or freezing rain, provided by the FD70, which makes matching the results practically impossible.

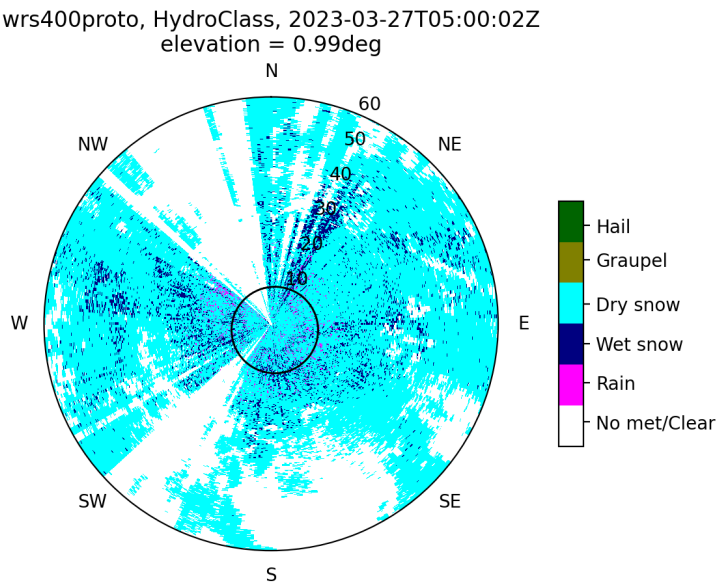
All in all, the modified parameters decreased the number of misclassifications in the snowing events without decreasing the performance in the case of the raining event. However, this analysis is focusing only on a small measurement volume. Therefore, we will next take a look at the visual comparisons that show a bigger picture of how the modifications affect the HydroClass' results.

## 4.3 Visual comparison

In this chapter, we will present PPI and RHI plots to illustrate the changes in the results after the parameter adjustments. From PPI plots, the occurrences of different hydrometeor types are counted and compared between the old and the modified parameters. The first visual example is from the snowing event on 2023-03-27. Figure 4.8 shows results from HydroClass using the old C-band specific parameters and the results from the modified parameters for X-band.



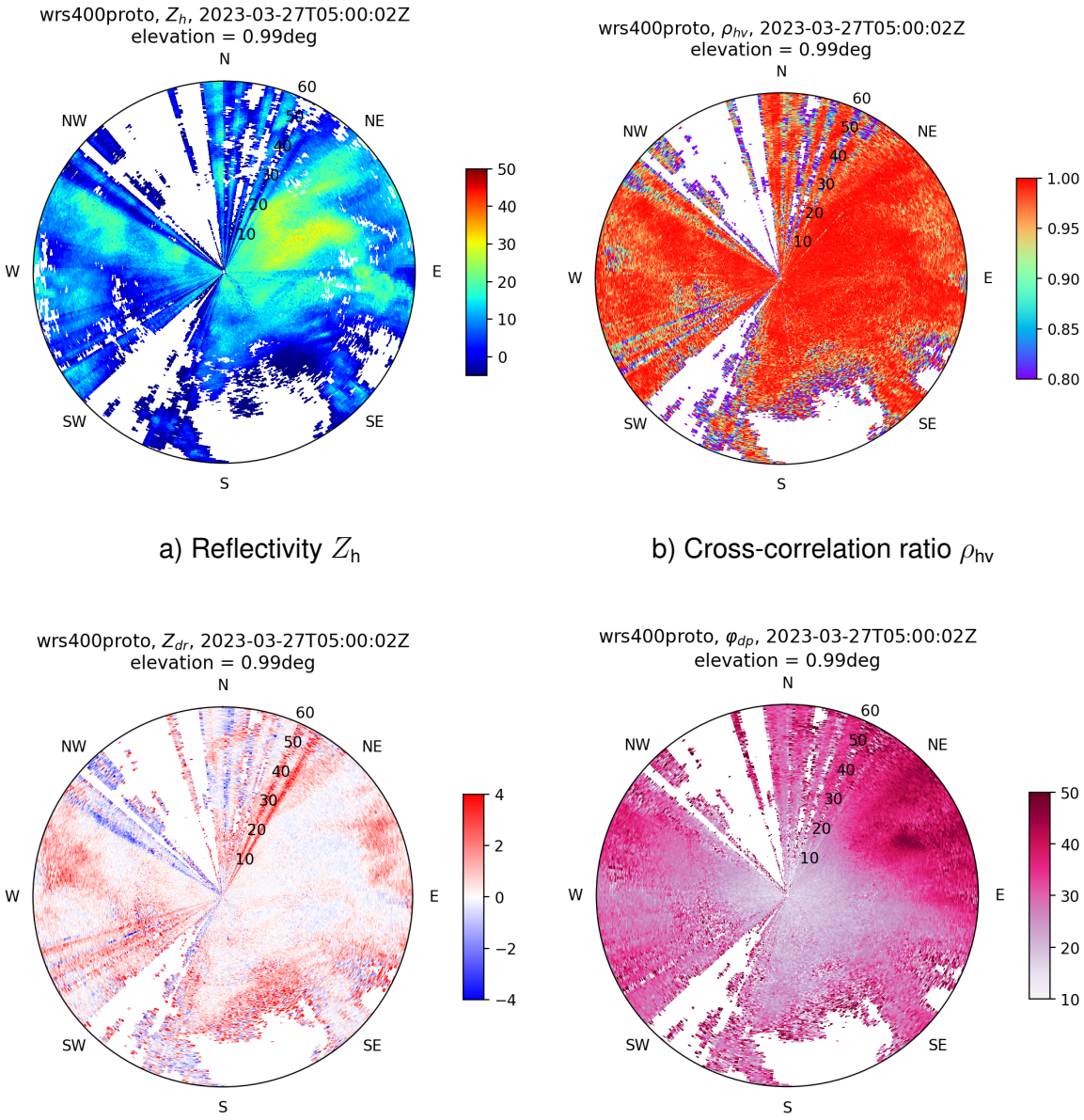

a) HydroClass with old parameters




b) HydroClass with modified parameters

***Figure 4.8.*** *PPI examples of the HydroClass a) with the old C-band parameters and b) with the X-band adjusted parameters, 2023-03-27, UTC: 05:00*

Figure 4.9 shows PPI plots from the polarimetric data moments corresponding to the results presented in figure 4.8.




a) Reflectivity $Z_{\mathrm{h}}$

b) Cross-correlation ratio $\rho_{\mathrm{hv}}$

c) Differential reflectivity $Z_{\mathrm{dr}}$

d) Differential phase $\varphi_{\mathrm{dp}}$

***Figure 4.9.*** *PPI examples of the polarimetric data moments, 2023-03-27, UTC: 05:00*

Figure 4.10 presents the HydroClass distribution from the PPI figure presented in figure 4.8. The noisiest measurements are left out of the analysis so data is thresholded so that points with signal-to-noise ratio below 10 dB are excluded.

The number of 'Rain'-pixels dropped from 4041 to 2479 while the number of 'Dry Snow'-pixels increased from 71844 to 73426. The change is visually apparent in the area marked by the black circle in figure 4.8 b). It is expected that some dry snow pixels get incorrectly classified as rain because their membership functions are overlapping as one can see from chapter 2.3. Our only ground truth measurement for this case comes from the FD70 sensor which reported snow on this time period. The FD70 obviously doesn't cover the entire measurement volume shown in figure 4.8 but it is reasonable to assume that weather conditions at its close proximity are similar. Therefore, we can say that it is quite probable that the decrease in the number of 'Rain'-pixels is a change for the better.

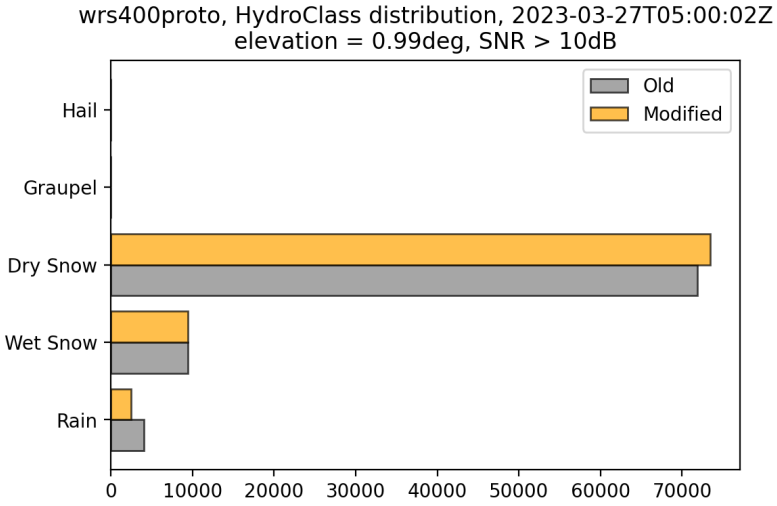
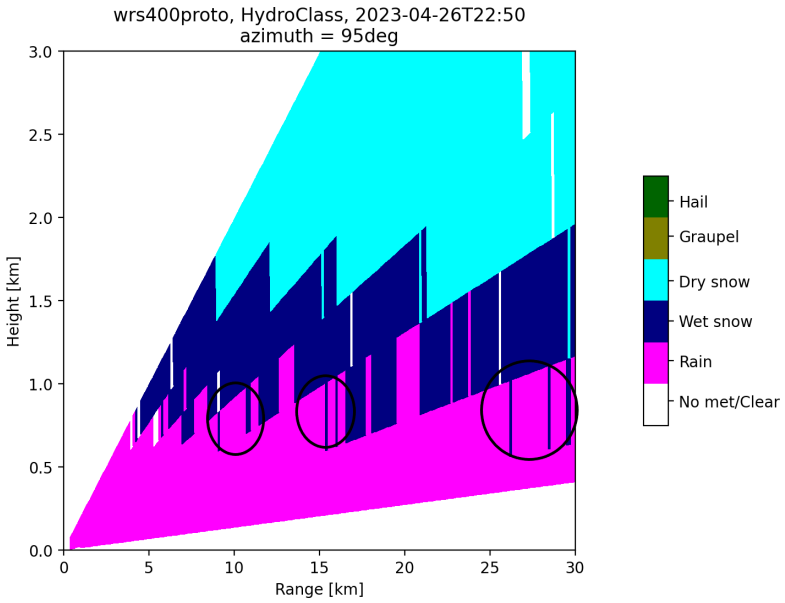
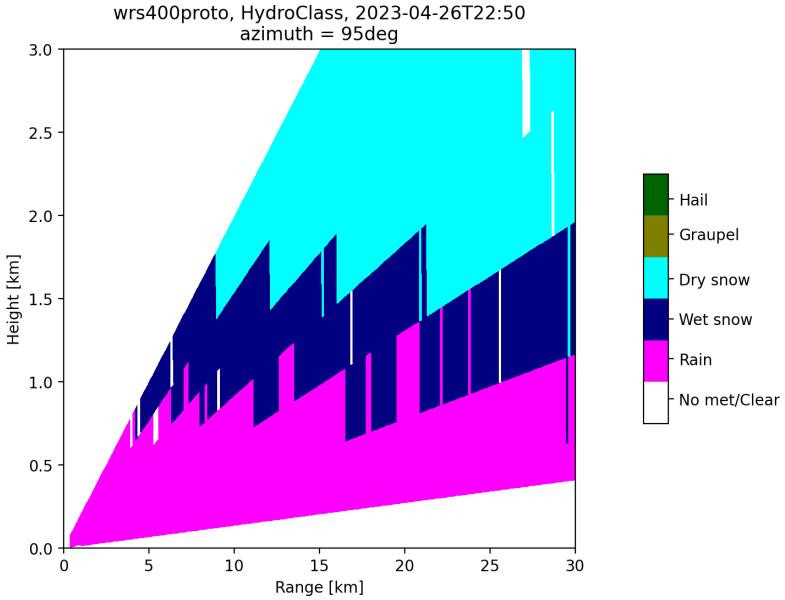



***Figure 4.10.*** *Distribution of different hydrometeor types reported by the HydroClass, 2023-03-27, UTC: 05:00*

The next examples are from the rain event in 2023-04-26. The results show how the modified parameters cause the result to be more consistent at the melting level height. Figure 4.11 shows RHI plots from the HydroClass results with the old and the modified parameters.




a) Hydrometeor classification with the old parameters

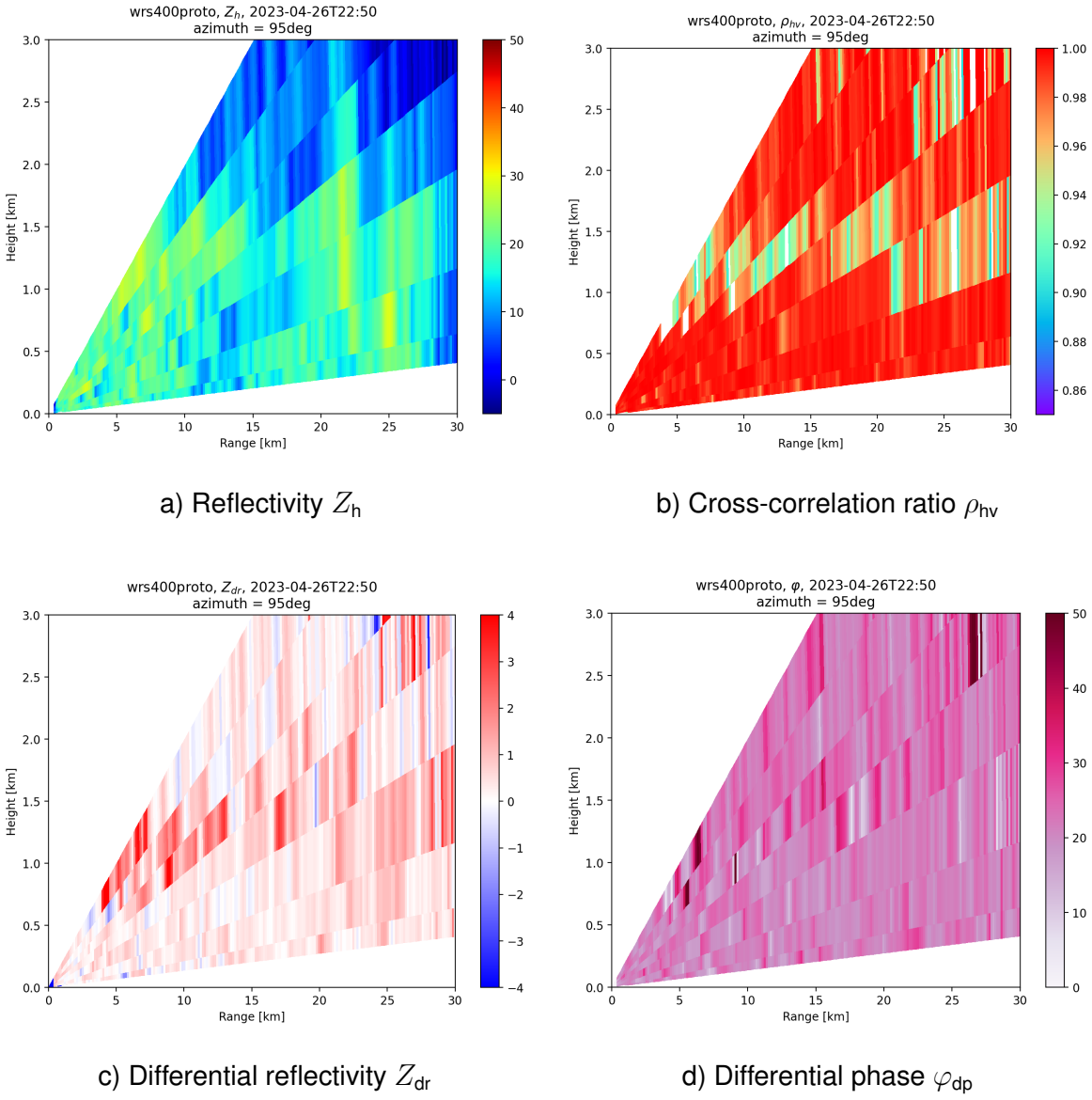



b) Hydrometeor classification with the modified parameters

***Figure 4.11.*** *RHI examples of the HydroClass a) with the old C-band parameters and b) with the X-band adjusted parameters, 2023-04-26, UTC: 22:50*

In plot a) of figure 4.11 some 'Wet snow' pixels appear even in the measurement volume which can be expected to be purely rain. These points are marked by black circles. The melting layer seems to be partly leaking into the liquid rain layer. This phenomenon is not as visible in plot b) which shows the results from the modified parameters. However, RHI plots do generally distort the shape of the weather events, especially in the vertical axis due to the nature of this visualization.

Figure 4.12 shows RHI plots of the corresponding polarimetric data moments.




a) Reflectivity $Z_{\mathrm{h}}$

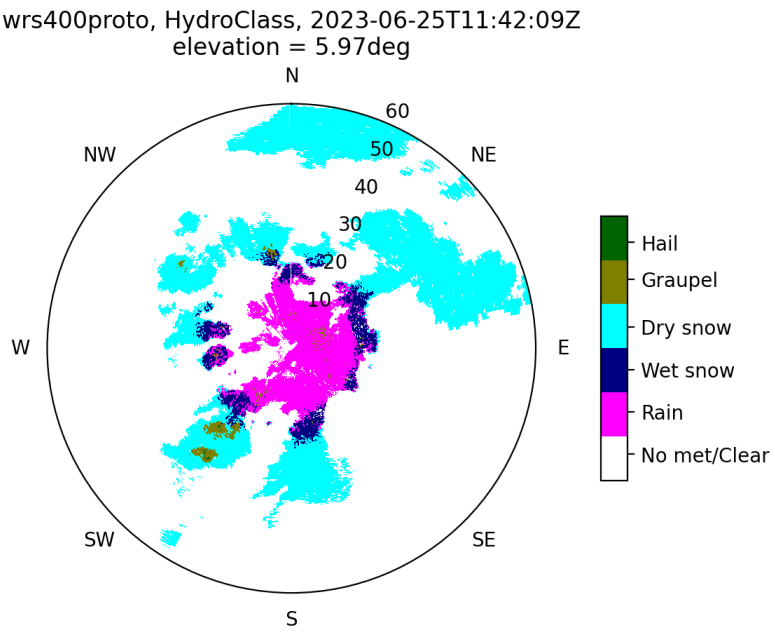
b) Cross-correlation ratio $\rho_{\mathrm{hv}}$

c) Differential reflectivity $Z_{\mathrm{dr}}$

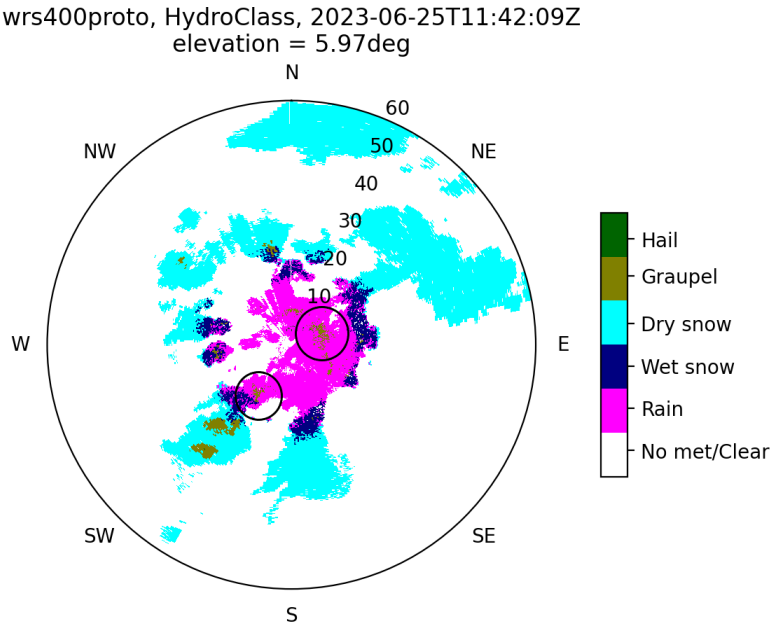
d) Differential phase $\varphi_{\mathrm{dp}}$

***Figure 4.12.*** *RHI examples of the polarimetric data moments, 2023-04-26, UTC: 22:50*

Figure 4.15 shows PPI visualizations from the heavy raining event on 2023-06-25. The event also included some lightning and graupel. These examples are from an elevation angle of $6^{\circ}$.



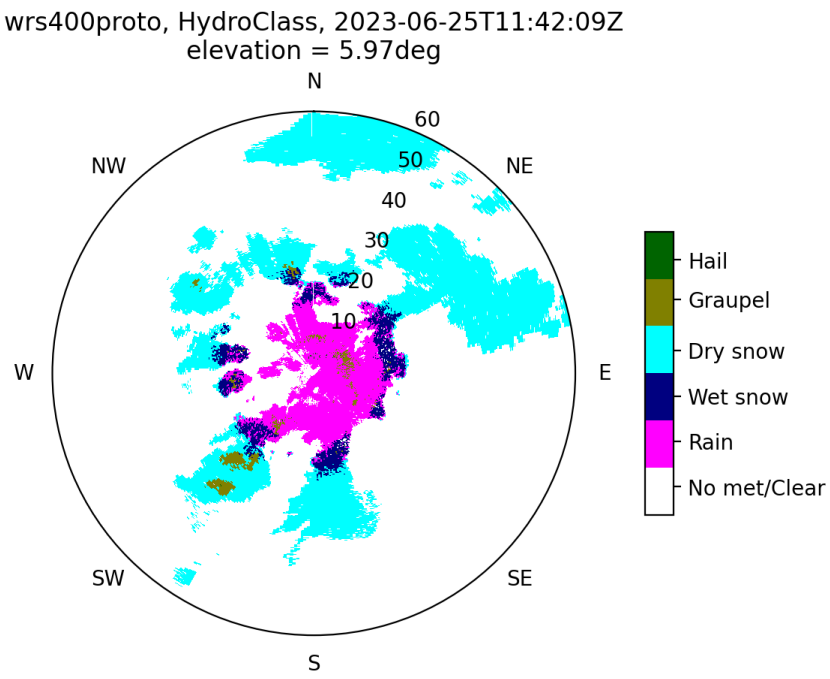

a) HydroClass with old parameters



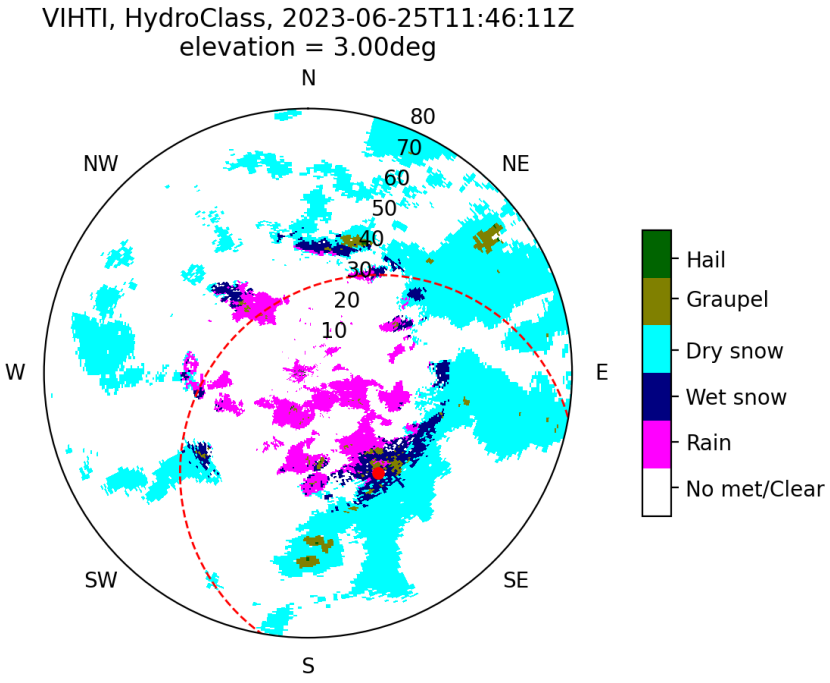

b) HydroClass with modified parameters

***Figure 4.13.*** *PPI examples of the HydroClass a) with the old C-band parameters and b) with the X-band adjusted parameters, 2023-06-25, UTC: 11:42*

The graupel bins are more apparent in the classification provided by the modified parameters. Again, these areas are marked by black circles. To verify the correctness of this observation a PPI plot from approximately the same time from FMI's weather radar in Vihti is shown in figure 4.14.



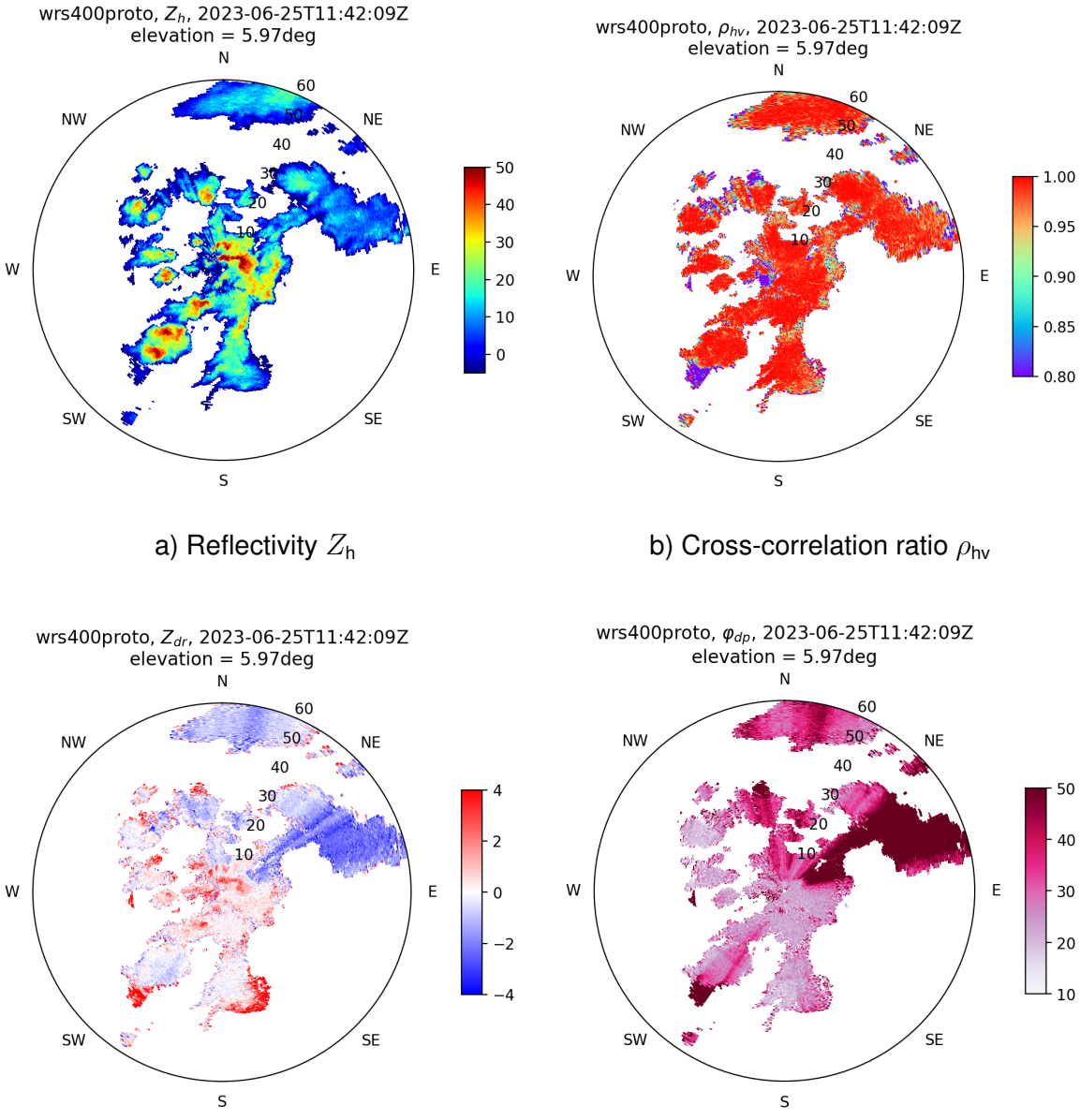

a) HydroClass with modified parameters




b) FMI Vihti hydrometeor types

***Figure 4.14.*** *The reported hydrometeor types from FMI's weather radar in Vihti a) and the HydroClass results with the X-band adjusted parameters b). The area covered by WRS400 in figure 4.13 is marked by the red dashed line.*

Because the WRS400 in Vantaanlaakso and FMI's radar in Vihti are located in different places and the plots are from different elevation angles there are some clear differences in the PPI visualizations from different radars. This is because the radars' beams are at different altitudes at different ranges. However, the graupel areas, both the one almost exactly above and the one to the southwest direction of the WRS400, are visible in both figures 4.13 and 4.14. The improved performance caused by the modified parameters shows that the graupel areas, especially the one above the WRS400 radar, are more coherent in plot b) of figure 4.13 than their counterparts in plot a) of the same figure. These areas are marked by the black circles in figure 4.13 b). Figure 4.15 shows PPI plots of the polarimetric data moments corresponding to the plots in figure 4.13.




a) Reflectivity $Z_{\mathrm{h}}$

b) Cross-correlation ratio $\rho_{\mathrm{hv}}$

c) Differential reflectivity $Z_{\mathrm{dr}}$

d) Differential phase $\varphi_{\mathrm{dp}}$

***Figure 4.15.*** *PPI examples of the polarimetric data moments, 2023-06-25, UTC: 11:42*

Figure 4.16 shows the distribution of hydrometeor types in PPI plot 4.17. The horizontal axis showing the number of occurrences is set to a logarithmic scale to make the changes in ’Hail’ visible as their number is significantly smaller than other classes.

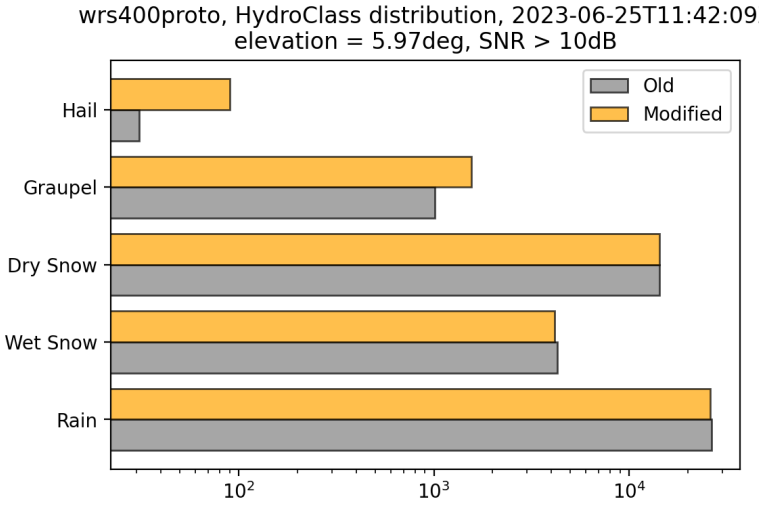
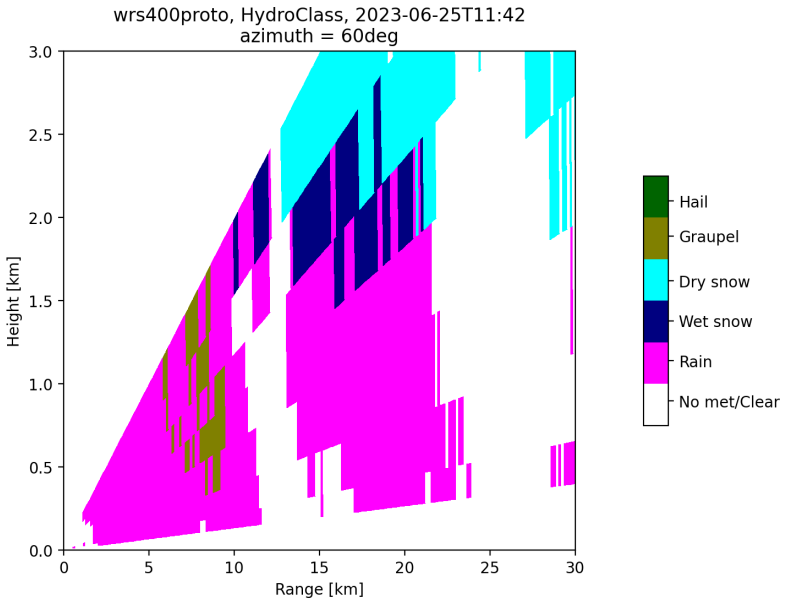


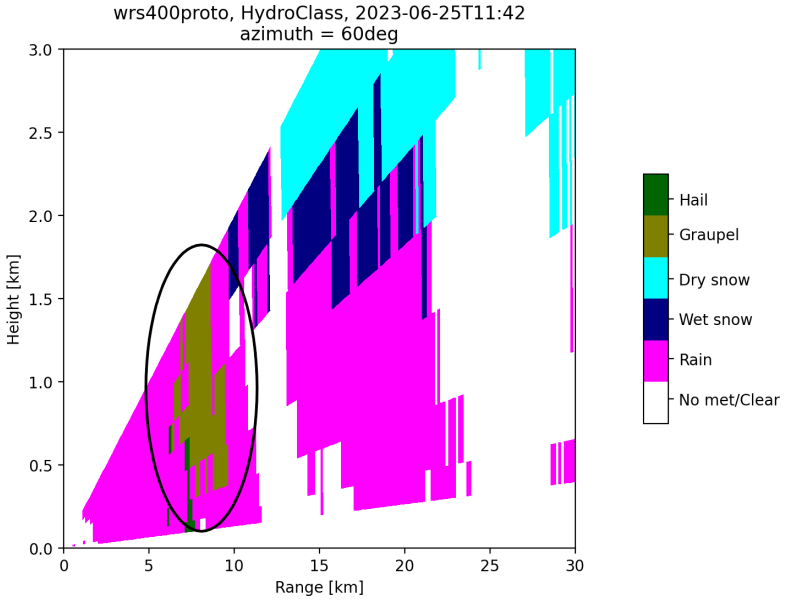

***Figure 4.16.*** *Distribution of different hydrometeor types reported by the HydroClass, 2023-06-25, UTC: 11:42*

The number of bins classified as ’Graupel’ and ’Hail’ has increased from 1012 to 1552 and from 31 to 90 respectively. These changes in combination with the observations from the FMI data indicate improved performance in graupel and hail detection.

Figure 4.17 shows RHI plots from the same heavy rain event as in figure 4.13. These visualizations show the same phenomenon that was already observed before graupel range bins appeared in a more coherent cluster with the modified parameters compared to the old parameters. In this case, the modified parameters behave in a significantly more expected way than the old parameters. It is highly unlikely that a graupel cell is split in between by liquid rain in a way shown in figure 4.17 a). This is probably the example where the more consistent and expected behavior and classification of the new parameters are most visible.



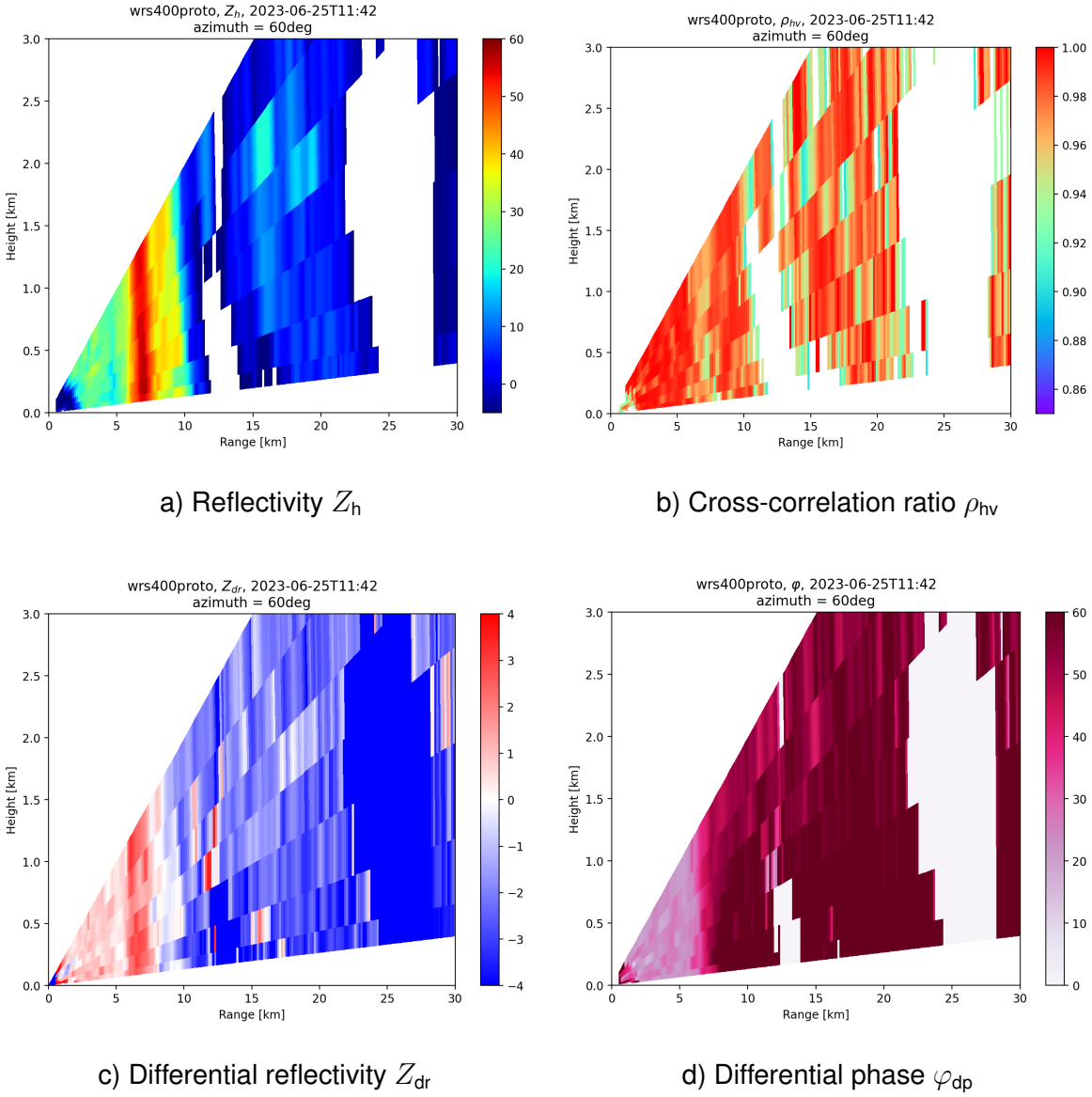

a) HydroClass with old parameters




b) HydroClass with modified parameters

***Figure 4.17.*** *RHI examples of the HydroClass a) with the old C-band parameters and b) with the X-band adjusted parameters, 2023-06-25, UTC: 11:42*

Figure 4.18 shows RHI plots corresponding to the HydroClass results presented in figure 4.17. The figure shows that there was a really strong weather echo with reflectivity $Z_{\mathrm{h}}$ of approximately 50 dB in combination with cross-correlation $\rho_{\mathrm{hv}}$ values too low to be liquid rain [18]. This further supports that the modified parameters perform better than the old parameters in this case too.




a) Reflectivity $Z_{\text{h}}$

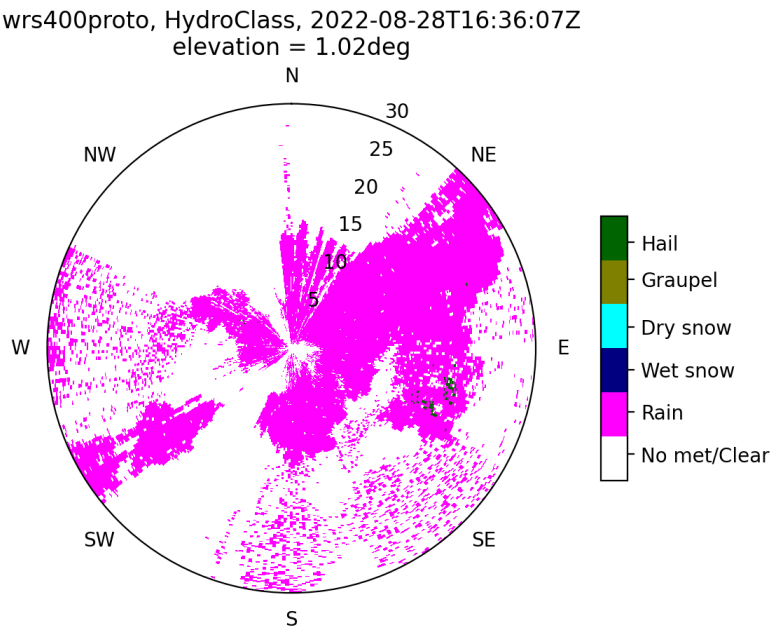
b) Cross-correlation ratio $\rho_{\text{hv}}$

c) Differential reflectivity $Z_{\text{dr}}$

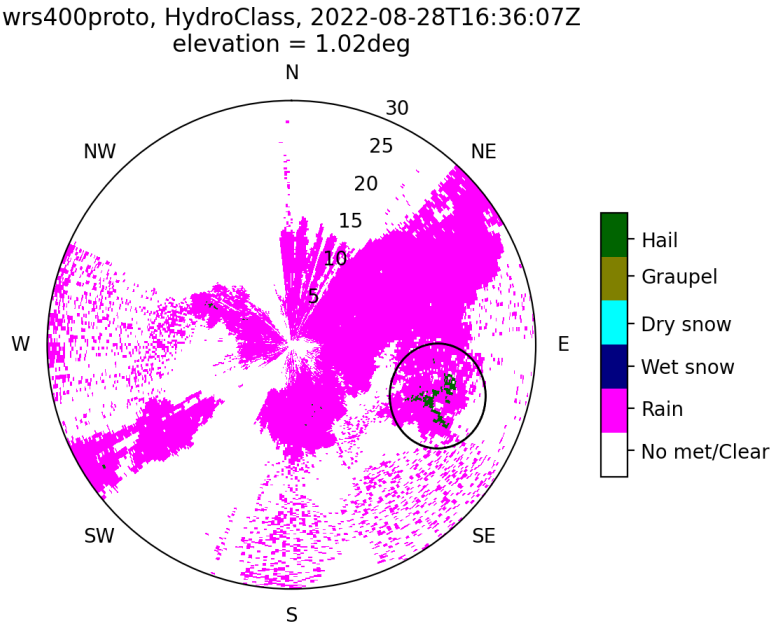
d) Differential phase $\varphi_{\text{dp}}$

***Figure 4.18.*** *RHI examples the polarimetric data moments, 2023-06-25, UTC: 11:42*

The final example of the MeteoClassifier presents data from a heavy rain and hail event on 2022-08-28. Myllypuro, which is the area where the hail was reported, is located approximately 12km from WRS400 in Vantaanlaakso at an azimuth angle of approximately $120°$. Figure 4.19 shows the hydrometeor classification results from this event. The group of hail range bins falls on top of the Myllypuro area.



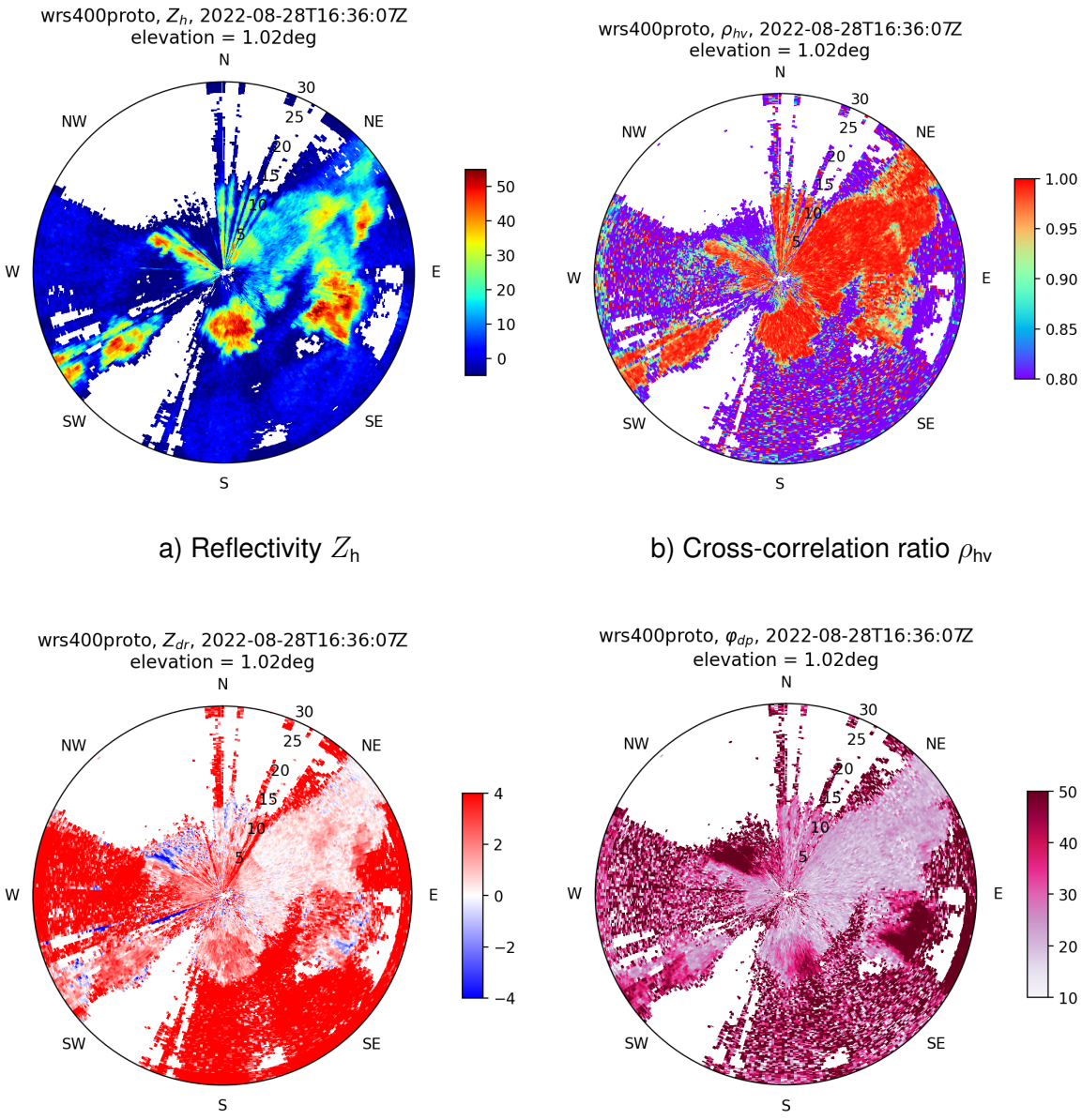

a) HydroClass with old parameters




b) HydroClass with modified parameters

***Figure 4.19.*** *PPI examples of the HydroClass a) with the old C-band parameters and b) with the X-band adjusted parameters, 2022-08-28, UTC: 16:36*

Again, the modified parameters provide a coherent group of hail bins while the old parameters show a more sparse group of hail bins. Figure 4.20 shows PPI plots of the polarimetric parameters corresponding to the MeteoClassifier results shown in figure 4.19.




a) Reflectivity $Z_h$

b) Cross-correlation ratio $\rho_{hv}$

c) Differential reflectivity $Z_{dr}$

d) Differential phase $\varphi_{dp}$

***Figure 4.20.*** *PPI examples of the polarimetric data moments, 2022-08-28, UTC: 16:36*

The polarimetric parameters show an area with strong $Z_h$, relatively low $\rho_{hv}$, and small, even negative, $Z_{dr}$. This is a combination that is not typically seen with liquid rain which further supports that the classification shown in figure 4.20 is reasonable and that the performance is enhanced with the parameter modification [18].

Figure 4.16 shows the hydrometeor type distribution in the PPI plot 4.19. Again, the vertical axis is set on a logarithmic scale to showcase the change in the number of hail bins.

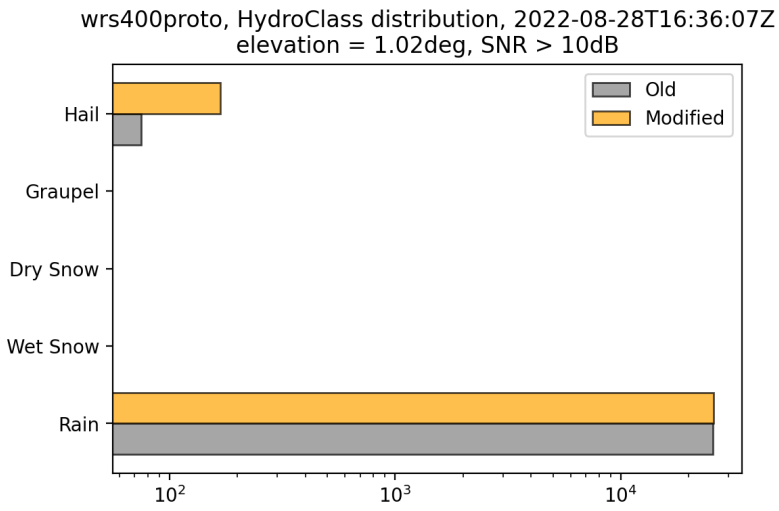



***Figure 4.21.*** *Distribution of different hydrometeor types reported by the HydroClass, 2022-08-28, UTC: 16:36*

The number of 'Hail'-pixels increased from 75 to 168. This in combination with the fact that the hail cell that appeared in figure 4.19 a) was sparse and not coherent, indicates that the parameter modification increased the accuracy of hail detection.

All in all, the different examples illustrate the improved performance with the adjusted parameters for the X-band system. It can be seen in the decreased misclassification of snow bins and a more coherent behavior in graupel and hail events. However, it must be noted that the quantitative analysis doesn't necessarily mean increased behavior. But in combination with the visual inspection, it is likely that the performance has indeed improved.

# 5. CONCLUSIONS AND DISCUSSION

The aim of this work was to adjust the parameters of the HydroClass algorithm and validate the improved performance after the adjustments. In this chapter, we will discuss if this goal was met and what could possibly be done in order to enhance the performance and the analysis methods even further.

## 5.1 Conclusions from the analysis results and study methods

The performance validation process done in this work can be divided into two parts: time series comparison with the FD70 forward scatter sensor and visual comparison. Based on the time series comparison we were able to analyze the HydroClass' performance in long-lasting snow and raining events and in a case where a snowing event includes also mixed hydrometeor types. A summary of the time series comparison with the FD70 is shown in table 5.1.

***Table 5.1.*** *Summary of the FD70 time series comparison*

| **Hydrometeor type** | **2023-03-27 UTC: 02:00 – 08:00 Snow** | | **2023-03-27 UTC: 18:00 – 23:15 Snow, mixed hydrometeors** | | **2023-06-26 UTC: 18:00 – 23:15 Rain** | |
|---|---|---|---|---|---|---|
| | Old | Modified | Old | Modified | Old | Modified |
| Rain | 6 | 1 | 2 | 12 | 74 | 74 |
| Wet Snow | 13 | 12 | 2 | 3 | 0 | 0 |
| Dry Snow | 66 | 72 | 76 | 65 | 0 | 0 |

The visual comparison offered a good general view of what changed in the classification after the parameter adjustments which was supported by simple quantitative analysis of the hydrometeor type distributions. A summary of the quantitative analysis of the PPI cases is shown in table 5.2

***Table 5.2.*** *Summary of the hydrometeor type distribution in PPI examples.*

| **Hydrometeor type** | **2022-08-28 UTC: 16:36** | | **2023-03-27 UTC: 05:00** | | **2023-06-25 UTC: 11:42** | |
|---|---|---|---|---|---|---|
| | Old | Modified | Old | Modified | Old | Modified |
| Rain | 23554 | 23738 | 4041 | 2479 | 26426 | 26054 |
| Wet Snow | 0 | 0 | 9471 | 9451 | 4298 | 4146 |
| Dry Snow | 0 | 0 | 71844 | 73426 | 14294 | 14289 |
| Graupel | 0 | 0 | 0 | 0 | 1012 | 1552 |
| Hail | 75 | 168 | 0 | 0 | 31 | 90 |

The first conclusion that can be made from the results of the analysis is that even with the old parameters, HydroClass performs well in regular snowing and rain events. Especially in the test case from 2023-04-26, which included a long-lasting rain event, the old and the adjusted parameters performed identically in the time series comparison. However, the visual inspection showed that with the modified parameters the behavior of the classification algorithm in the melting layer was more consistent with the expected behavior. In the test case from 2023-03-27 which included a long-lasting snowing event, the parameter adjustments did significantly reduce the number of snow bins misclassified as liquid rain. But even the old parameters performed relatively well overall in this case.

In the test case from 2023-03-28 with mixed hydrometeor types, the old parameters did not seem to react to the changes in the hydrometeor types. In the time series from the HydroClass using the adjusted parameters, the shift is clearly visible. The time period that included frozen drizzle and snow and rain mixture is visible as there are more bins classified as rain. HydroClass doesn't have an output class that clearly corresponds to these obscure hydrometeor types but if a human observer were to look at the data from HydroClass from this time period they should be able to notice that the event did not consist of dry snow only. That would be impossible from the classifications produced by the old parameters. This case did also show that the adjustments to the parameters did not just reduce the number of rain classifications in snowing events but actually caused the algorithm to perform more accurately compared to the ground observations.

The hail and graupel cases from 2022-08-28 and 2023-06-25 showed a more consistent behavior with the adjusted parameters than with the old parameters. The graupel and hail areas appeared to be much more uniform. This is well illustrated in figure 4.17. In the visual comparison with FMI's data, the modified parameters showed a great match to the graupel areas reported by the FMI's operational weather radar while in the old parameters' case, some graupel bins could even be interpreted as individual misclassifications rather

than a clear graupel cell. The improvement in performance is evident in this case too.

All in all, it is clear that the performance of the HydroClass in the X-band system improved with the parameter modifications. The performance was especially better in snowing events and in hail and graupel detection. A narrower support in $\rho_{hv}$ membership function and a slightly revised $Z_{dr}$ membership function of the 'Rain ' class reduced the number of snow bins incorrectly classified as liquid rain. They also improve the performance in graupel cases. The membership functions for the 'Graupel'-class were not changed but the graupel detection still improved. This is due to the changes in the 'Rain' class. The modifications made to the 'Hail'-class' membership functions resulted in a more consistent behavior of the algorithm in detecting hail cells.

A weakness of the used methods in this work is that they don't really provide a numerical value on how big the improvement is. The time series comparison with the FD70 is a good method in the sense that it is based on reliable ground truth data measured by the forward scatter sensor. However, the time series comparison is focused only on a small part of the radar measurement volume. On the other, while the visual comparison considers larger parts of the measurement volume, it does lack the capability to quantify and verify the degree of improvement. Counting the classification results and comparing them between old and modified parameters does tell how much the results did change, but it doesn't tell if the change was in the correct direction. Additional conclusions have to be drawn from the visual inspection on how coherent and expected the classification result seems to be. Keeping these aspects in mind, we can say that the aim of this work, improving and validating HydroClass' performance in an X-band weather radar system, was clearly achieved but it is impossible to say exactly how big the improvement is.

To improve the accuracy of this analysis method it would be beneficial to have a network of forward scatter sensors. This would provide more reference points and increase the probability of a suitable weather event occurring at the location of the sensor. For example, we did not have data from a test case in which a hail or graupel event would have taken place on top of the FD70 sensor at the airport so it was not possible to analyze time series data from such event. Of course, human observers could be utilized as well but that is quite laborious and probably not as accurate as the classification provided by the FD70. Having a network of FD70 sensors in an area covered by a radar network would provide a great set of data points and an opportunity for a thorough analysis of the HydroClass' performance. This would also provide a meaningful sample size for the analysis method presented in 4.7.

Another weakness of the study methods in this work is that the reference values are from a ground-based measurement and we don't have similar observations from higher altitudes. Of, course getting such data is not an easy task. It would require some sort of aircraft with a suitable sensor or human observer on board. Maybe in the future, it could be possible to

utilize drones in this manner. FMI's operational radars do offer data from higher altitudes as well but in that case, matching the measurement volumes is problematic due to radars observing the atmosphere from different locations using different directional angles.

Even though the visual comparison doesn't offer a numerical accuracy value as a result either, they clearly do still add some value to the analysis. They illustrate how the algorithm behaves in different scenarios and it enables a simple and easy-to-understand comparison between setups.

## 5.2 Improvement suggestions and possible topics for further studies

The literature review and the adjustment process also brought up some ideas for the development of the HydroClass and further study topics related to the topic of this work. First of all, the analysis of selecting a 3$\times$3 window rather than a single range bin showed that including information about the surrounding bins does improve the classification results even in this really simple implementation. This takeaway could be developed further either by including information on the HydroClass classification of the surrounding bins or the underlying polarimetric data moments. For example, clustering methods could be used to achieve this like in the revised version of the CSU method [2].

Even though the parameter adjustments significantly reduced the number of snow pixels incorrectly classified as rain there are some still visible in the results. This is at least partly because the polarimetric characteristics of dry snow and light to moderate rain are pretty similar as we can see from the figures 2.9, 2.10, and 2.12. The melting layer height is already used as an input which does take care of this issue at temperatures close to 0°C but it doesn't solve the issue for clearly freezing conditions. This is because the membership functions of the melting level height are identical for all cases $h_{\mathrm{ML}} < 0$ km. Therefore, the air temperature would be a good indicator that could separate the two classes. A simple solution for that could be integrating a thermometer into the radar that would measure the air temperature at the radar site. This could then be used in order to reject liquid rain classifications in clearly freezing conditions.

Machine learning is a trend that has been on the rise in solving different classification problems during recent decades. There are also some studies where different machine learning methods have been used for hydrometeor classification. For example, k-means clustering and support vector machines have been used in combination with a fuzzy logic approach [2][15][17]. A convolutional neural network has also been used in order the classify hydrometeors without fuzzy logic [11]. Out of these cases, the clustering approaches in combination with fuzzy logic seem to have been the most successful. This could be a possible addition to be considered in the future if Vaisala's hydrometeor classification

scheme is to be updated.

Another observation from hydrometeor classification algorithms presented in the literature reference is that many algorithms do not use two dimensional membership functions for any other class than rain [2][21][34]. This makes these systems simpler and easier to manage. On the other hand, there seems to be a consensus in that for rain the two dimensional membership functions are necessary. This is caused by the physical properties of liquid water compared to ice [16].

Many hydrometeor classification schemes have more output classes compared to those in the HydroClass. For example, there are different classes for different sized hail and graupel [34]. However, this is not necessarily a weakness of the HydroClass. Instead, it is a question of prioritizing the accurate classification of fewer classes rather than the inclusion of more classes. Also, increasing the number of classes would also make maintaining the system harder and of course, require adding a multitude of new membership functions.

All in all, the study process showed that fuzzy logic is a great tool for hydrometeor classification. This is supported by the results shown in this work and by the overview of the literature on the topic. As the visualizations of the membership functions presented in this work show, the characteristics of different hydrometeors are often overlapping. It would be hard to model this using classical logic or some other method. Fuzzy logic enables discriminating different precipitation classes despite their often overlapping characteristics.

When it comes to further study ideas, of course, it would be possible to collect more data from different weather conditions and analyze time series data from those cases too. A good further study would also be to analyze how sensitive the algorithm is to changes in the measured values. For example, how does a slightly incorrect input melting level height affect the results or what kind of effects does a bias in one of the polarimetric parameters have? The methods presented in this work provide good tools for these further study topics if they were to be carried out.

A logical future study topic related to hydrometeor classification is adjusting the other algorithms that use the polarimetric data moments in order to create advanced data products. For example, attenuation correction and melting layer detection are issues for which Vaisala provides algorithmic solutions. They should be adjusted for the X-band system as well.

# 6. SUMMARY

Fuzzy logic is a great tool for hydrometeor classification. Its capabilities allow for modeling the so-called logical middle ground. The main idea of fuzzy logic is to expand the set of allowed truth values to include any real value between 0 and 1. This means that an object may belong to multiple alternate classes to some extent. Membership functions are part of fuzzy systems that model this "fuzziness". They convert the measured values to fuzzy values, which makes them an essential part of any fuzzy system. Vaisala's hydrometeor classification algorithm HydroClass is also based on fuzzy logic.

Modern weather radars measure the polarimetric properties of the hydrometeors. These variables are the key to hydrometeor classification as their values differ between different hydrometeor types. Although the characteristics of different hydrometeor types are overlapping when it comes to some of the polarimetric properties. Therefore, the classification is based on multiple polarimetric variables to discriminate between classes.

The polarimetric characteristics detected by the radar are dependent on the frequency of the radar's transmitted signal. This is caused by different scattering effects that are dependent on the ratio of the transmitted signal's wavelength and the size of the target. The most significant scattering phenomena, in this case, are resonance effects and attenuation. Because of these phenomena, the parameters of the membership functions must be adjusted for each frequency band. In this work, the modifications were made based on literature references of previously carried out adjustment processes.

As shown in this work, the performance of a hydrometeor classification algorithm can be evaluated by comparing it to the results given by a ground-based forward scattering sensor. To counter misclassifications caused by noisy measurements, a $3 \times 3$ range bin window from the radar measurement surrounding the ground-based sensor can be picked out instead of an individual bin on top of the ground-based sensor. Using this method it was possible to show that a more accurate classification result in snowy conditions was achieved with the adjusted parameters compared to the initial parameters. On the other hand, visual inspection showed a more consistent behavior in graupel and hail detection.

In the future, the performance of HydroClass could be analyzed in a more varied set of rain cases and the robustness of the method could be investigated further too. Vaisala's other algorithms based on polarimetric variables should also be adjusted.

# REFERENCES


[1] L Baldini, E Gorgucci, and V Chandrasekar. “Hydrometeor classification methodology for C-band polarimetric radars”. In: *Proceedings of ERAD*. Vol. 62. 66. 2004.

[2] Renzo Bechini and V Chandrasekar. “A semisupervised robust hydrometeor classification method for dual-polarization radar applications”. In: *Journal of Atmospheric and Oceanic Technology* 32.1 (2015), pp. 22–47.

[3] Brenda Dolan and Steven A Rutledge. “A theory-based hydrometeor identification algorithm for X-band polarimetric radars”. In: *Journal of Atmospheric and Oceanic Technology* 26.10 (2009), pp. 2071–2088.

[4] *Glossary of Meteorology*. American Meteorological Society. URL: https://glossary.ametsoc.org/wiki/Melting_layer.

[5] “Golfpallon kokoiset rakeet jättivät jälkeensä tuhoa Helsingissä”. In: *Helsingin Sanomat* (2022). URL: https://www.hs.fi/kaupunki/art-2000009034596.html.

[6] Eugenio Gorgucci and V Chandrasekar. “Evaluation of attenuation correction methodology for dual-polarization radars: Application to X-band systems”. In: *Journal of Atmospheric and Oceanic Technology* 22.8 (2005), pp. 1195–1206.

[7] Dirk Klugmann and Lasse Kauppinen. “FD70-A Tool for Supporting Satellite Weather Observations”. In: *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*. IEEE. 2022, pp. 7471–7474.

[8] Takeharu Kouketsu et al. “A hydrometeor classification method for X-band polarimetric radar: Construction and validation focusing on solid hydrometeors under moist environments”. In: *Journal of Atmospheric and Oceanic Technology* 32.11 (2015), pp. 2052–2074.

[9] Sanghun Lim, Venkatachalam Chandrasekar, and Viswanathan N Bringi. “Hydrometeor classification system using dual-polarization radar measurements: Model improvements and in situ verification”. In: *IEEE Transactions on Geoscience and Remote Sensing* 43.4 (2005), pp. 792–801.

[10] Hongping Liu and V Chandrasekar. “Classification of hydrometeors based on polarimetric radar measurements: Development of fuzzy logic and neuro-fuzzy systems, and in situ verification”. In: *Journal of Atmospheric and Oceanic Technology* 17.2 (2000), pp. 140–164.

[11] Yuping Lu and Jitendra Kumar. “Convolutional neural networks for hydrometeor classification using dual polarization Doppler radars”. In: *2019 International Conference on Data Mining Workshops (ICDMW)*. IEEE. 2019, pp. 288–295.

[12] T MUNAKATA and Y JANI. “Fuzzy systems : an overview”. eng. In: *Communications of the ACM* 37.3 (1994), pp. 69–76. ISSN: 0001-0782.

[13] Eni G Njoku. *Encyclopedia of remote sensing*. eng. The encyclopedia of earth sciences series. Springer, 2014. ISBN: 9780387367002.

[14] *Radar handbook*. eng. 3rd ed. McGraw-Hill's AccessEngineering. Maidenhead: McGraw-Hill Professional, 2008. ISBN: 0-07-162113-X.

[15] Jean-François Ribaud, Luiz Augusto Toledo Machado, and Thiago Biscaro. “X-band dual-polarization radar-based hydrometeor classification for Brazilian tropical precipitation systems”. In: *Atmospheric Measurement Techniques* 12.2 (2019), pp. 811–837.

[16] Ronald E. Rinehart. *Radar for Meteorologists*. Rinehart Publications, 1997. ISBN: 0-9658002-0-2.

[17] Nicoletta Roberto et al. “A support vector machine hydrometeor classification algorithm for dual-polarization radar”. In: *Atmosphere* 8.8 (2017), p. 134.

[18] Alexander Ryzhkov and Dusan Zrnic. “9R. 3 RADAR POLARIMETRY AT S, C, AND X BANDS COMPARATIVE ANALYSIS AND OPERATIONAL IMPLICATIONS”. In: (2005).

[19] Alexander V Ryzhkov et al. “The Joint Polarization Experiment: Polarimetric rainfall measurements and hydrometeor classification”. In: *Bulletin of the American Meteorological Society* 86.6 (2005), pp. 809–824.

[20] Hassan Al-Sakka et al. “A new fuzzy logic hydrometeor classification scheme applied to the French X-, C-, and S-band polarimetric radars”. In: *Journal of Applied Meteorology and Climatology* 52.10 (2013), pp. 2328–2344.

[21] T Schuur et al. “Observations and classification of echoes with the polarimetric WSR-88D radar”. In: *Report of the National Severe Storms Laboratory, Norman, OK* 73069 (2003), p. 46.

[22] Richard D Scott, Paul R Krehbiel, and William Rison. “The use of simultaneous horizontal and vertical transmissions for dual-polarization radar meteorological observations”. In: *Journal of Atmospheric and Oceanic Technology* 18.4 (2001), pp. 629–648.

[23] Donald Shepard. “A two-dimensional interpolation function for irregularly-spaced data”. In: *Proceedings of the 1968 23rd ACM national conference*. 1968, pp. 517–524.

[24] S.N. Sivanandam. *Introduction to Fuzzy Logic using MATLAB*. eng. 1st ed. 2007. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. ISBN: 1-280-70031-9.

[25] Merrill I. Skolnik. *Introduction to radar systems*. eng. 3rd ed. New York (NY): McGraw-Hill, 2001. ISBN: 0-07-290980-3.

[26] *The Finnish Meteorological Institute's open data*. Available online. URL: https://en.ilmatieteenlaitos.fi/open-data.

[27] Esko. Turunen. *Mathematics behind fuzzy logic*. eng. Advances in soft computing. Heidelberg: Physica-Verlag, 1999. ISBN: 3-7908-1221-8.

[28] *University of Wyoming, Department of Atmospheric Science, Atmospheric Soundings*. Available online. URL: https://weather.uwyo.edu/upperair/sounding.html.

[29] *User Guide, Dual Polarization, IRIS and RDA*. Access restricted. Vaisala. 2021.

[30] Yanting Wang and V Chandrasekar. “Algorithm for estimation of the specific differential phase”. In: *Journal of Atmospheric and Oceanic Technology* 26.12 (2009), pp. 2565–2578.

[31] Neil C Wells. *The Atmosphere and Ocean: A Physical Introduction*. eng. 3. Aufl. Vol. 5. Advancing weather and climate science. Hoboken: Wiley, 2011.

[32] *Wikimedia Commons, Radar cross section of metal sphere from Mie theory*. Available online. URL: https://commons.wikimedia.org/wiki/File:Radar_cross_section_of_metal_sphere_from_Mie_theory.svg.

[33] Lotfi A. Zadeh. “Is there a need for fuzzy logic?” eng. In: *Information sciences* 178.13 (2008), pp. 2751–2779. ISSN: 0020-0255.

[34] Chuanhong Zhao et al. “An improved hydrometeor identification method for X-band dual-polarization radar and its application for one summer Hailstorm over Northern China”. In: *Atmospheric research* 245 (2020), p. 105075.